

**BỘ GIÁO DỤC
VÀ ĐÀO TẠO**

**VIỆN HÀN LÂM KHOA HỌC VÀ
CÔNG NGHỆ VIỆT NAM**

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ



TRỊNH HIỀN ANH

**NGHIÊN CỨU CÁC KỸ THUẬT PHÁT HIỆN, NHẬN DIỆN BẤT
THƯỜNG THIẾT BỊ TRÊN LƯỚI ĐIỆN CAO THẾ TỪ ẢNH THU
NHẬN BẰNG UAV DỰA TRÊN HỌC SÂU**

LUẬN ÁN TIẾN SĨ MÁY TÍNH

Hà Nội – 2026

BỘ GIÁO DỤC
VÀ ĐÀO TẠO

VIỆN HÀN LÂM KHOA HỌC VÀ
CÔNG NGHỆ VIỆT NAM

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

Trịnh Hiền Anh

NGHIÊN CỨU CÁC KỸ THUẬT PHÁT HIỆN, NHẬN DIỆN BẤT
THƯỜNG THIẾT BỊ TRÊN LƯỚI ĐIỆN CAO THẾ TỪ ẢNH THU
NHẬN BẰNG UAV DỰA TRÊN HỌC SÂU

LUẬN ÁN TIẾN SĨ MÁY TÍNH

Ngành: Hệ thống thông tin

Mã số: 9 48 01 04

Xác nhận của Học viện
Khoa học và Công nghệ

Người hướng dẫn 1
(Ký, ghi rõ họ tên)

Người hướng dẫn 2
(Ký, ghi rõ họ tên)

Hà Nội - Năm 2026

LỜI CAM ĐOAN

Tôi xin cam đoan tất cả các nội dung trong luận án: *“Nghiên cứu các kỹ thuật phát hiện, nhận diện bất thường thiết bị trên lưới điện cao thế từ ảnh thu nhận bằng UAV dựa trên học sâu”* là công trình nghiên cứu của riêng tôi, dưới sự hướng dẫn khoa học của TS. Nguyễn Thị Thanh Tân và PGS.TS.Ngô Quốc Tạo.

Luận án sử dụng thông tin trích dẫn từ nhiều nguồn tham khảo khác nhau, các thông tin trích dẫn được ghi rõ nguồn gốc. Các kết quả nghiên cứu của tôi được công bố chung với các tác giả khác đã được sự nhất trí của đồng tác giả khi đưa vào luận án. Các số liệu, kết quả được trình bày trong luận án là hoàn toàn trung thực và chưa từng được công bố trong bất kỳ một công trình nào khác ngoài các công trình công bố của tác giả. Luận án được hoàn thành trong thời gian tôi làm nghiên cứu sinh tại Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam.

Hà Nội, ngày 21 tháng 09 năm 2026

Tác giả luận án

TRỊNH HIỀN ANH

LỜI CẢM ƠN

Lời đầu tiên, tôi xin được bày tỏ lòng biết ơn sâu sắc nhất đến cô TS. Nguyễn Thị Thanh Tân và thầy PGS.TS Ngô Quốc Tạo, thầy/cô đã luôn tận tình chỉ bảo, hướng dẫn tôi trong suốt quá trình định hướng nghiên cứu, phương pháp nghiên cứu, cho đến cách trình bày các bài báo khoa học, các báo cáo chuyên đề và luận án. Thầy/cô luôn động viên những lúc tôi gặp khó khăn trong chặng đường nghiên cứu của mình.

Tôi xin chân thành cảm ơn Ban lãnh đạo Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam, các thầy cô của Học viện Khoa học và Công nghệ đã động viên, giúp đỡ và tạo điều kiện thuận lợi cho tôi trong suốt quá trình thực hiện luận án. Tôi cũng xin cảm ơn các đồng nghiệp trong Viện Công nghệ thông tin, Viện Hàn lâm Khoa học và Công nghệ Việt Nam đã có nhiều đóng góp quý báu giúp tôi hoàn thiện luận án.

Con xin cảm ơn bố, mẹ, anh chị em hai bên gia đình, em xin cảm ơn chồng và hai con những người luôn ở bên, ủng hộ, động viên cho con/em có thời gian, điều kiện tốt nhất để nghiên cứu và hoàn thành luận án.

Hà Nội, ngày 21 tháng 09 năm 2026

Tác giả luận án

Trịnh Hiền Anh

Mục lục

LỜI CAM ĐOAN	I
LỜI CẢM ƠN	II
DANH MỤC TỪ VIẾT TẮT	VI
DANH MỤC BẢNG BIỂU	VIII
DANH MỤC HÌNH ẢNH	IX
MỞ ĐẦU	1
Chương 1. GIỚI THIỆU BÀI TOÁN PHÁT HIỆN VÀ NHẬN DIỆN BẤT THƯỜNG THIẾT BỊ TRÊN LƯỚI ĐIỆN TỬ ẢNH UAV	5
1.1 Giới thiệu bài toán	5
1.1.1 Phát hiện thiết bị trên lưới truyền tải điện cao thế	7
1.1.2 Phát hiện bất thường trên chuỗi cách điện	9
1.1.3 Đặc điểm và thách thức của dữ liệu UAV trong giám sát lưới điện	10
1.2 Các nghiên cứu liên quan	18
1.2.1 Phát hiện thiết bị và bất thường từ ảnh thông thường	19
1.2.2 Phát hiện thiết bị và bất thường từ ảnh UAV	21
1.2.3 Các phương pháp sinh dữ liệu tổng hợp nhằm giải quyết tình trạng khan hiếm mẫu bất thường và mất cân bằng dữ liệu	28
1.2.4 Khoảng trống nghiên cứu	30
1.3 Các nội dung nghiên cứu và kết quả chính của luận án	31
1.3.1 Các câu hỏi nghiên cứu	31
1.3.2 Các nội dung nghiên cứu của luận án	33
1.3.3 Kết quả chính của luận án	36
1.4 Kết luận chương 1	37

Chương 2.	PHÁT HIỆN THIẾT BỊ TRÊN LƯỚI ĐIỆN CAO THẾ TỪ ẢNH UAV	38
2.1	Các nghiên cứu liên quan về phát hiện thiết bị trên lưới điện ..	38
2.1.1	Nghiên cứu phát hiện thiết bị trên lưới điện từ ảnh UAV ...	38
2.1.2	Phát hiện thiết bị trên lưới điện sử dụng mô hình một giai đoạn	38
2.2	Mô hình HVE-YOLO (High Voltage Equipment Detection)	41
2.2.1	Mô hình HVE-YOLO	41
2.2.2	Các cải tiến trong HVE-YOLO	46
2.3	Thử nghiệm	53
2.3.1	Dữ liệu thử nghiệm.....	53
2.3.2	Kịch bản thử nghiệm	56
2.3.3	Kết quả thử nghiệm	62
2.4	Kết luận chương 2.....	64
Chương 3.	PHÁT HIỆN BẤT THƯỜNG TRÊN CHUỖI CÁCH ĐIỆN TỪ ẢNH UAV	66
3.1	Các nghiên cứu liên quan về phát hiện bất thường trên chuỗi cách điện.....	66
3.1.1	Nghiên cứu về bất thường trên chuỗi cách điện	66
3.1.2	Nghiên cứu phát hiện bất thường trên chuỗi cách điện từ ảnh thu nhận bằng UAV.....	68
3.2	Mô hình đề xuất FGS (FPN–GELAN–Diffusion Segmentation)	69
3.2.1	Kiến trúc tổng quan	69
3.2.2	Backbone GELAN	72
3.2.3	Mô hình khuếch tán (Diffusion Model).....	74
3.2.4	Vùng quan tâm và các nhánh dự đoán	76
3.2.5	Giai đoạn hậu xử lý (Post-processing)	77
3.3	Thử nghiệm	78
3.3.1	Thiết lập thử nghiệm	78
3.3.2	Chuẩn bị dữ liệu	78
3.3.3	Kết quả thử nghiệm	80
3.4	Kết luận chương 3.....	84
Chương 4.	SINH DỮ LIỆU ẢNH BẤT THƯỜNG CỦA CHUỖI CÁCH ĐIỆN	86

4.1	Các nghiên cứu liên quan về sinh dữ liệu trong huấn luyện các mô hình học sâu	87
4.1.1	Các phương pháp sinh dữ liệu truyền thống	87
4.1.2	Các phương pháp sinh dữ liệu và học với dữ liệu hạn chế trong phát hiện bất thường	89
4.2	AAPID (Attention-aware Adaptive Prompting for Insulator Diagnosis) – Mô hình Prompt thích nghi dựa trên cơ chế Attention cho chẩn đoán lỗi chuỗi cách điện	91
4.2.1	Tổng quan kiến trúc	91
4.2.2	Cơ chế Prompt thích nghi dựa trên Attention	92
4.2.3	Học biểu diễn Prompt bằng Information Bottleneck	95
4.2.4	Anchor Loss và tối ưu không gian đặc trưng	96
4.2.5	Hàm mục tiêu và phân tích mô hình	97
4.3	Thử nghiệm	97
4.3.1	Môi trường thử nghiệm	98
4.3.2	Dữ liệu thử nghiệm	98
4.3.3	Kết quả thử nghiệm	99
4.4	Kết luận chương 4	104
	KẾT LUẬN VÀ KIẾN NGHỊ	106
	CÔNG TRÌNH CÔNG BỐ	110
	TÀI LIỆU THAM KHẢO	111

DANH MỤC TỪ VIẾT TẮT

Viết tắt	Viết đầy đủ	Ý nghĩa tiếng Việt
HLAT	Hành lang an toàn	Hành lang an toàn lưới điện
HTĐ	Hệ thống điện	Hệ thống truyền tải và phân phối điện
NCS	Nghiên cứu sinh	Học viên nghiên cứu bậc tiến sĩ
TBA	Trạm biến áp	Cơ sở chuyển đổi điện áp trong hệ thống điện
AI	Artificial Intelligence	Trí tuệ nhân tạo
AAPID	Attention-aware Adaptive Prompting for Insulator Diagnosis	Phương pháp prompting thích ứng có nhận thức chú ý cho chẩn đoán cách điện
C2PSA	Cross Stage Partial with Parallel Spatial Attention	Khối CSP kết hợp cơ chế chú ý không gian song song
CNN	Convolutional Neural Network	Mạng nơ-ron tích chập
CSP	Cross Stage Partial	Kết nối từng phần giữa các tầng của mạng nơ-ron tích chập
EVNNPT	National Power Transmission Corporation of Vietnam	Tổng công ty Truyền tải điện Quốc gia
FPN	Feature Pyramid Network	Mạng kim tự tháp đặc trưng
FRP	Fiber Reinforced Polymer	Vật liệu composite sợi gia cường polymer
FSCIL	Few-Shot Class-Incremental Learning	Học gia tăng lớp với số lượng mẫu huấn luyện rất ít
GELAN	Generalized Efficient Layer Aggregation Network	Mạng tổng hợp lớp hiệu quả tổng quát
GSD	Ground Sample Distance	Kích thước mặt đất tương ứng với một pixel trong ảnh số
HAcc	Harmonic Accuracy	Độ chính xác điều hòa
HOG	Histogram of Oriented Gradients	Biểu đồ hướng gradient
HVE	High Voltage Equipment	Thiết bị trên lưới điện cao thế
IB	Information Bottleneck	Nút thắt thông tin

(Còn tiếp ở trang sau)

(Tiếp theo DANH MỤC TỪ VIẾT TẮT)

Viết tắt	Ý nghĩa tiếng Anh	Ý nghĩa tiếng Việt
MAIFI	Momentary Average Interruption Frequency Index	Chỉ số tần suất mất điện thoáng qua trung bình
PAN	Path Aggregation Network	Mạng tổng hợp đường dẫn đặc trưng
PD	Performance Dropping	Mức suy giảm hiệu năng
RCNN	Region-based Convolutional Neural Network	Mạng nơ-ron tích chập dựa trên vùng
RELAN	Residual Efficient Layer Aggregation Network	Mạng tổng hợp lớp dư hiệu quả
RNN	Recurrent Neural Network	Mạng nơ-ron hồi tiếp
RPN	Region Proposal Network	Mạng đề xuất vùng
RQ	Research Question	Câu hỏi nghiên cứu
SMOTE	Synthetic Minority Over-sampling Technique	Kỹ thuật tăng mẫu tổng hợp cho lớp thiểu số
SIFT	Scale-Invariant Feature Transform	Biến đổi đặc trưng bất biến theo tỷ lệ
SSD	Single Shot MultiBox Detector	Bộ phát hiện đa hộp một lần suy luận
SPPF	Spatial Pyramid Pooling Fast	Khối gộp kim tự tháp không gian nhanh
UAV	Unmanned Aerial Vehicle	Thiết bị bay không người lái
ViT	Vision Transformer	Mô hình Transformer cho xử lý ảnh
YOLO	You Only Look Once	Thuật toán phát hiện đối tượng thời gian thực

(Kết thúc DANH MỤC TỪ VIẾT TẮT)

DANH MỤC BẢNG BIỂU

Bảng 1.1: Ứng dụng UAV trong giám sát lưới điện tại các khu vực	12
Bảng 1.2: Khoảng cách an toàn giữa UAV và đường dây truyền tải điện cao thế	16
Bảng 1.3: So sánh hiệu năng của một số mô hình phát hiện đối tượng sử dụng Transformer và mô hình lai CNN–Transformer trong ảnh UAV	27
Bảng 1.4: So sánh các phương pháp phát hiện, nhận diện bất thường thiết bị trên lưới truyền tải điện sử dụng UAV	34
Bảng 2.1: Phân bố số lượng mẫu theo từng lớp thiết bị trong bộ dữ liệu UAV .	57
Bảng 2.2: Bảng so sánh hiệu năng giữa HVE-YOLO và HVE-Mask R-CNN .	63
Bảng 3.1: Phân bố các loại nền trong bộ dữ liệu VNElectric	79
Bảng 3.2: So sánh hiệu năng phát hiện bất thường chuỗi cách điện trên bộ dữ liệu VNElectric	80
Bảng 3.3: Nghiên cứu ablation của mô hình FGS	81
Bảng 4.1: So sánh hiệu năng trên tập dữ liệu CIFAR100 (Thiết lập: 60+8x5- way 5-shot)	101
Bảng 4.2: So sánh hiệu năng trên tập dữ liệu VPI-Fault-2500 (thiết lập: 3+2- way 5-shot)	101
Bảng 4.3: Kết quả nghiên cứu loại bỏ theo độ chính xác trung bình (Average Accuracy $\$A_{avg}\$$) (%)	102

DANH MỤC CÁC HÌNH VẼ, ĐỒ THỊ

Hình 1.1: Các thiết bị trên lưới truyền tải điện	8
Hình 1.2: Các loại chuỗi cách điện	9
Hình 2.1: Sơ đồ cấu trúc khối C3k2	39
Hình 2.2: Quy trình xây dựng các mô hình nhận dạng thiết bị cho lưới điện cao thế	42
Hình 2.3: Quy trình thu thập dữ liệu trên lưới điện cao thế bằng thiết bị bay không người lái	43
Hình 2.4: So sánh backbone YOLOv11 và HVE-YOLO	47
Hình 2.5: So sánh nguyên lý hoạt động của SPD-Conv và Stride Convolution .	48
Hình 2.6: Gán nhãn dữ liệu	55
Hình 3.1: Một số dạng bất thường trên chuỗi cách điện	67
Hình 3.2: Kiến trúc tổng thể của mô hình FGS	71
Hình 3.3: Kiến trúc mô hình GELAN	73
Hình 3.4: Kiến trúc mô hình DiT	75
Hình 3.5: Kết quả phát hiện bất thường chuỗi cách điện trong các điều kiện khác nhau.	82
Hình 4.1: Kiến trúc tổng thể của mô hình AAPID	92
Hình 4.2: Trực quan hóa t-SNE kết quả nghiên cứu loại bỏ trên bộ dữ liệu VPI-Fault-2500.	103

MỞ ĐẦU

TÍNH CẤP THIẾT

Trong bối cảnh nhu cầu điện năng không ngừng gia tăng cùng với yêu cầu ngày càng cao về độ tin cậy và an toàn vận hành của hệ thống điện quốc gia, lưới truyền tải điện cao thế và siêu cao thế giữ vai trò đặc biệt quan trọng trong việc bảo đảm an ninh năng lượng và duy trì sự ổn định của nền kinh tế – xã hội. Các thiết bị trên lưới truyền tải, đặc biệt là chuỗi cách điện, phải hoạt động liên tục trong điều kiện môi trường khắc nghiệt như nhiệt độ cao, độ ẩm lớn, ô nhiễm công nghiệp, gió bão và tác động cơ học kéo dài. Theo thời gian, các yếu tố này làm gia tăng nguy cơ xuất hiện các bất thường như nứt vỡ bát sứ, mất bát cách điện, suy giảm vật liệu hoặc nhiễm bẩn bề mặt. Những hư hỏng này có thể gây phóng điện, làm giảm độ tin cậy cung cấp điện, thậm chí dẫn đến các sự cố nghiêm trọng trên hệ thống truyền tải điện.

Cùng với quá trình chuyển đổi số trong ngành điện lực, nhiều công nghệ giám sát tiên tiến đã được triển khai nhằm nâng cao hiệu quả kiểm tra tình trạng thiết bị trên lưới truyền tải điện. Bên cạnh các phương pháp kiểm tra trực tiếp truyền thống, các đơn vị vận hành đã từng bước ứng dụng camera độ phân giải cao, thiết bị ảnh nhiệt, UAV và các hệ thống giám sát từ xa để hỗ trợ thu thập dữ liệu. Nhờ đó, khả năng tiếp cận hiện trường được cải thiện đáng kể, đặc biệt đối với các tuyến đường dây đi qua khu vực đồi núi, rừng sâu hoặc địa hình phức tạp.

Tuy nhiên, thách thức hiện nay không còn chủ yếu nằm ở khâu thu thập dữ liệu mà chuyển dần sang bài toán xử lý, phân tích và khai thác hiệu quả lượng dữ liệu lớn được tạo ra từ các hệ thống giám sát. Trong thực tế, việc nhận diện các bất thường trên thiết bị vẫn đòi hỏi sự tham gia đáng kể của chuyên gia vận hành, đặc biệt đối với các hư hỏng nhỏ hoặc hiếm gặp. Khi số lượng ảnh và video thu nhận ngày càng lớn, việc rà soát thủ công không chỉ làm gia tăng chi phí vận hành và thời gian kiểm tra mà còn dễ dẫn đến bỏ sót các dấu hiệu bất thường tiềm ẩn, ảnh hưởng đến khả năng phát hiện sớm nguy cơ sự cố trên lưới điện.

Trong số các công nghệ giám sát hiện đại, thiết bị bay không người lái (UAV) đang được xem là một giải pháp đầy tiềm năng nhờ khả năng thu nhận hình ảnh độ phân giải cao từ nhiều góc quan sát khác nhau với chi phí hợp lý. UAV cho phép kiểm tra chi tiết tình trạng thiết bị mà không cần cắt điện hoặc tiếp cận trực tiếp, đồng thời giúp nâng cao hiệu quả giám sát trên phạm vi rộng. Tuy nhiên, khối lượng dữ liệu hình ảnh UAV thu thập được là rất lớn, vượt quá khả năng xử lý hiệu quả bằng các phương pháp thủ công truyền thống. Điều này đặt ra yêu cầu cấp thiết về các giải pháp tự động có khả năng phát hiện thiết bị và nhận diện bất thường một cách nhanh chóng, chính

xác và tin cậy.

Trong những năm gần đây, các phương pháp thị giác máy tính và học sâu đã đạt được nhiều thành tựu nổi bật trong lĩnh vực phân tích ảnh và phát hiện đối tượng. Các mô hình hiện đại như YOLO, Faster R-CNN, RetinaNet, DETR và các kiến trúc dựa trên Transformer đã chứng minh hiệu quả trong nhiều lĩnh vực như giám sát giao thông, nông nghiệp chính xác, quản lý hạ tầng đô thị và kiểm tra công nghiệp. Nhờ khả năng tự động học các đặc trưng biểu diễn từ dữ liệu lớn, các mô hình này đã vượt trội so với các phương pháp xử lý ảnh truyền thống về độ chính xác, khả năng thích nghi và năng lực tổng quát hóa.

Tuy nhiên, dữ liệu UAV trong lĩnh vực điện lực có nhiều đặc thù khiến các phương pháp hiện có chưa thể phát huy đầy đủ hiệu quả. Các thiết bị điện thường xuất hiện dưới nhiều tỷ lệ kích thước khác nhau, được thu nhận từ nhiều góc quan sát và khoảng cách khác nhau, trong khi nền ảnh thường chứa nhiều thành phần gây nhiễu như cây cối, công trình xây dựng, đồi núi hoặc bầu trời. Ngoài ra, sự thay đổi của điều kiện thời tiết, ánh sáng và hiện tượng che khuất làm gia tăng đáng kể độ phức tạp của bài toán phát hiện và nhận diện thiết bị.

Một thách thức quan trọng khác là số lượng mẫu bất thường trong thực tế thường rất hạn chế so với dữ liệu vận hành bình thường. Việc thu thập các trường hợp hư hỏng thiết bị gặp nhiều khó khăn do sự cố xảy ra với tần suất thấp, phân bố không đồng đều và đòi hỏi quá trình kiểm tra, xác nhận bởi các chuyên gia. Điều này dẫn đến hiện tượng mất cân bằng dữ liệu trong quá trình huấn luyện, làm suy giảm khả năng nhận biết các trường hợp bất thường của mô hình học sâu. Bên cạnh đó, nhiều dạng hư hỏng chỉ biểu hiện ở quy mô rất nhỏ hoặc xuất hiện cục bộ trên một bộ phận của thiết bị, khiến việc phân biệt giữa trạng thái bình thường và bất thường trở nên khó khăn hơn.

Ngoài các yêu cầu về độ chính xác, việc triển khai hệ thống trong thực tế còn đòi hỏi khả năng xử lý hiệu quả trên các nền tảng tính toán hạn chế như UAV hoặc các thiết bị biên. Do đó, các mô hình cần đồng thời đáp ứng các yêu cầu về tốc độ suy luận, khả năng tổng quát hóa, hiệu quả sử dụng tài nguyên và độ ổn định trong nhiều điều kiện vận hành khác nhau. Đây là những thách thức có ý nghĩa khoa học và thực tiễn quan trọng đối với việc xây dựng các hệ thống giám sát lưới điện tự động thể hệ mới.

Do đó, việc nghiên cứu các kỹ thuật phát hiện và nhận diện bất thường thiết bị trên lưới điện cao thế từ ảnh UAV dựa trên học sâu mang tính cấp thiết cả về mặt khoa học lẫn thực tiễn. Nghiên cứu không chỉ góp phần nâng cao hiệu quả giám sát và phát hiện sớm hư hỏng thiết bị mà còn thúc đẩy quá trình chuyển đổi số trong ngành điện lực, hướng tới mô hình bảo trì dự báo thông minh (predictive maintenance), giảm

thiếu sự cố và nâng cao độ tin cậy vận hành của hệ thống điện quốc gia.

Xuất phát từ những yêu cầu thực tiễn và ý nghĩa khoa học nêu trên, luận án lựa chọn đề tài:

“Nghiên cứu các kỹ thuật phát hiện, nhận diện bất thường thiết bị trên lưới điện cao thế từ ảnh thu nhận bằng UAV dựa trên học sâu”.

MỤC TIÊU NGHIÊN CỨU

Mục tiêu tổng quát

Đề xuất các mô hình học sâu nhằm nâng cao hiệu quả phát hiện và nhận diện bất thường trên thiết bị lưới truyền tải điện từ ảnh UAV trong điều kiện đối tượng nhỏ, nền phức tạp và dữ liệu mất cân bằng.

Mục tiêu cụ thể

- Nghiên cứu các mô hình phát hiện đối tượng dựa trên học sâu cho ảnh UAV;
- Đề xuất mô hình phát hiện và nhận diện bất thường trên chuỗi cách điện;
- Nghiên cứu kỹ thuật xử lý mất cân bằng dữ liệu nhằm nâng cao hiệu quả huấn luyện;
- Xây dựng và đánh giá thực nghiệm hệ thống trên bộ dữ liệu UAV thực tế.

NỘI DUNG NGHIÊN CỨU

- Nghiên cứu hiện trạng giám sát thiết bị trên lưới truyền tải điện;
- Phân tích các mô hình học sâu trong phát hiện đối tượng từ ảnh UAV;
- Xây dựng và đánh giá bộ dữ liệu thiết bị điện thu nhận bằng UAV;
- Đề xuất các giải pháp nâng cao hiệu quả phát hiện và nhận diện bất thường, bao gồm:
 - Mô hình phát hiện thiết bị trên lưới truyền tải điện;
 - Mô hình phát hiện bất thường trên chuỗi cách điện;
 - Kỹ thuật tăng cường và cân bằng dữ liệu.

ĐỐI TƯỢNG VÀ PHẠM VI NGHIÊN CỨU

Đối tượng nghiên cứu

Với mục tiêu đã đề ra của luận án, đối tượng nghiên cứu của luận án bao gồm:

- Các mô hình học sâu trong phát hiện đối tượng từ ảnh thu nhận bằng UAV;

- Chuỗi cách điện và các dạng bất thường thường gặp trên lưới truyền tải điện;
- Các mô hình phát hiện đối tượng và phát hiện bất thường dựa trên học sâu;
- Các chiến lược xử lý mất cân bằng dữ liệu trong thị giác máy tính.

Phạm vi nghiên cứu

Luận án tập trung vào bài toán phát hiện, nhận diện bất thường thiết bị trên lưới điện cao thế từ ảnh quang học thu nhận bằng UAV trong môi trường thực tế, không mở rộng sang các bài toán dự báo theo chuỗi thời gian dài hạn hoặc mô phỏng hệ thống điện.

ĐÓNG GÓP CỦA LUẬN ÁN

- Xây dựng mô hình cải tiến HVE-YOLO nhằm nâng cao hiệu quả phát hiện các thiết bị trên lưới truyền tải điện từ ảnh UAV;
- Phát triển khung học sâu tích hợp phục vụ bài toán phát hiện và phân loại bất thường trên chuỗi cách điện trong điều kiện đối tượng nhỏ và nền ảnh phức tạp;
- Xây dựng chiến lược tăng cường dữ liệu dựa trên Attention-Aware Prompting nhằm khắc phục tình trạng mất cân bằng dữ liệu trong quá trình huấn luyện.

Các phương pháp đề xuất được đánh giá thông qua thực nghiệm trên dữ liệu thu nhận thực tế trên lưới truyền tải điện cao thế 110kV tại Việt Nam và so sánh với các mô hình tiên tiến, qua đó khẳng định tính hiệu quả và khả năng ứng dụng trong giám sát lưới điện bằng UAV.

CHƯƠNG 1

GIỚI THIỆU BÀI TOÁN PHÁT HIỆN VÀ NHẬN DIỆN BẤT THƯỜNG THIẾT BỊ TRÊN LƯỚI ĐIỆN TỪ ẢNH UAV

Chương này trình bày tổng quan về bài toán phát hiện và nhận diện bất thường thiết bị trên lưới truyền tải điện sử dụng dữ liệu ảnh thu nhận từ thiết bị bay không người lái (Unmanned Aerial Vehicle – UAV). Nội dung chương tập trung phân tích bối cảnh ứng dụng, các thách thức của dữ liệu ảnh UAV và các hướng tiếp cận dựa trên học sâu trong bài toán giám sát lưới điện thông minh.

Bên cạnh việc phân tích nhu cầu tự động hóa công tác kiểm tra thiết bị và xu hướng ứng dụng UAV trong thu nhận dữ liệu phục vụ quản lý vận hành, bảo trì dự báo và chuyển đổi số ngành điện, chương cũng khảo sát và đánh giá các công trình nghiên cứu tiêu biểu trong và ngoài nước liên quan đến phát hiện đối tượng và nhận diện bất thường trên lưới truyền tải điện. Thông qua quá trình phân tích, các hạn chế còn tồn tại của những phương pháp hiện nay như khả năng phát hiện đối tượng kích thước nhỏ, tính thích nghi với dữ liệu UAV, ảnh hưởng của nền cảnh phức tạp và bài toán mất cân bằng dữ liệu được làm rõ.

Trên cơ sở đó, chương xác định động cơ nghiên cứu, xây dựng bài toán khoa học cần giải quyết và định hướng các nội dung nghiên cứu chính của luận án nhằm phát triển các phương pháp phát hiện và nhận diện bất thường thiết bị trên lưới truyền tải điện dựa trên dữ liệu ảnh UAV và học sâu.

1.1 Giới thiệu bài toán

Hệ thống điện là một trong những hạ tầng thiết yếu bảo đảm an ninh năng lượng và sự phát triển bền vững của nền kinh tế - xã hội. Trong đó, lưới truyền tải điện cao áp và siêu cao áp đóng vai trò trung gian kết nối giữa các nhà máy phát điện và các trung tâm phụ tải, cho phép truyền tải điện năng trên phạm vi rộng với độ tin cậy và tính liên tục vận hành cao [1, 2]. Sự ổn định của hệ thống truyền tải điện phụ thuộc trực tiếp vào tình trạng vận hành của các thiết bị trên đường dây, do đó việc giám sát và bảo trì thiết bị là nhiệm vụ quan trọng nhằm bảo đảm an toàn cung cấp điện.

Trong quá trình vận hành lâu dài, các thiết bị trên lưới truyền tải như chuỗi cách điện, dây dẫn, khóa néo, tạ chống rung và các phụ kiện liên kết phải làm việc trong điều kiện môi trường phức tạp, chịu ảnh hưởng của nhiệt độ, độ ẩm, ô nhiễm không khí, rung động cơ học và quá trình lão hóa vật liệu. Các tác động này có thể gây ra nhiều dạng suy giảm hoặc hư hỏng như nứt vỡ cách điện, mất bát sứ, ăn mòn phụ kiện

hoặc suy giảm khả năng cách điện của thiết bị [3]. Nếu không được phát hiện và xử lý kịp thời, các bất thường có thể dẫn đến hiện tượng phóng điện, ngắn mạch hoặc sự cố mất điện trên diện rộng, gây ảnh hưởng nghiêm trọng đến độ tin cậy và an toàn vận hành của hệ thống điện [1].

Nhằm duy trì trạng thái vận hành ổn định của lưới điện, nhiều phương thức kiểm tra và giám sát thiết bị đã được triển khai trong thực tế như kiểm tra trực tiếp ngoài hiện trường, sử dụng trực thăng, robot chuyên dụng, camera giám sát, ảnh nhiệt và các hệ thống cảm biến từ xa [4, 5]. Mặc dù các phương pháp này có khả năng cung cấp thông tin hữu ích về tình trạng thiết bị, chúng vẫn tồn tại những hạn chế nhất định liên quan đến chi phí triển khai, phạm vi bao phủ, mức độ an toàn lao động và khả năng tiếp cận các khu vực địa hình phức tạp. Đặc biệt, đối với các tuyến đường dây kéo dài qua vùng núi cao, rừng sâu hoặc khu vực khó tiếp cận, công tác kiểm tra định kỳ thường đòi hỏi nguồn lực lớn và phụ thuộc đáng kể vào kinh nghiệm của chuyên gia vận hành.

Trong bối cảnh đó, thiết bị bay không người lái (Unmanned Aerial Vehicle – UAV) đã nổi lên như một giải pháp đầy tiềm năng cho công tác kiểm tra lưới điện thông minh. Với ưu điểm về tính cơ động, khả năng tiếp cận linh hoạt và chi phí vận hành thấp, UAV cho phép thu nhận hình ảnh và video có độ phân giải cao từ nhiều góc quan sát khác nhau mà không làm gián đoạn quá trình vận hành của hệ thống điện [2, 3]. Dữ liệu thu nhận từ UAV không chỉ hỗ trợ quá trình quan sát trực quan mà còn tạo tiền đề cho việc phát triển các hệ thống phân tích tự động dựa trên trí tuệ nhân tạo và học máy.

Song song với sự phát triển của công nghệ UAV, lĩnh vực thị giác máy tính và học sâu đã đạt được nhiều thành tựu quan trọng trong các bài toán phân tích ảnh và nhận dạng đối tượng. Các kiến trúc hiện đại như Faster R-CNN, RetinaNet, YOLO và DETR đã chứng minh hiệu quả vượt trội trong việc phát hiện đối tượng trên nhiều miền ứng dụng khác nhau [6–9]. Gần đây, các mô hình dựa trên Transformer tiếp tục mở ra những hướng nghiên cứu mới với khả năng học biểu diễn toàn cục và khai thác hiệu quả mối quan hệ ngữ cảnh giữa các đối tượng trong ảnh. Trong lĩnh vực điện lực, việc kết hợp dữ liệu UAV với các mô hình học sâu đã cho thấy tiềm năng lớn trong việc tự động phát hiện thiết bị, nhận diện bất thường và đánh giá trạng thái vận hành của hệ thống truyền tải điện [5, 10].

Từ góc độ thị giác máy tính, bài toán giám sát lưới điện từ ảnh UAV có thể được chia thành nhiều nhiệm vụ khác nhau như phát hiện đối tượng (object detection), phân đoạn ảnh (image segmentation), phân loại trạng thái thiết bị (classification) và phát hiện bất thường (anomaly detection). Trong số đó, phát hiện đối tượng đóng vai trò nền tảng vì kết quả xác định vị trí và loại thiết bị là cơ sở cho các bước phân tích

chuyên sâu như đánh giá hư hỏng hoặc dự báo nguy cơ sự cố.

Mặc dù các mô hình học sâu hiện đại đã đạt được nhiều kết quả ấn tượng trên các bộ dữ liệu tổng quát, việc áp dụng trực tiếp vào dữ liệu UAV trong lĩnh vực điện lực vẫn gặp nhiều thách thức. Thứ nhất, các thiết bị điện thường chỉ chiếm một tỷ lệ rất nhỏ trong ảnh có độ phân giải cao, làm gia tăng độ khó trong việc trích xuất đặc trưng và định vị đối tượng. Thứ hai, dữ liệu UAV được thu nhận ở nhiều độ cao và góc quan sát khác nhau, dẫn đến sự biến đổi đáng kể về kích thước, hình dạng và hướng xuất hiện của đối tượng. Thứ ba, nền ảnh thường chứa nhiều thành phần gây nhiễu như bầu trời, cây cối hoặc công trình xây dựng, làm giảm khả năng phân biệt giữa đối tượng và môi trường xung quanh [3, 5]. Ngoài ra, số lượng mẫu bất thường trong thực tế thường hạn chế, gây ra hiện tượng mất cân bằng dữ liệu nghiêm trọng và làm suy giảm hiệu quả của các mô hình học sâu trong quá trình huấn luyện.

Xuất phát từ những yêu cầu thực tiễn và các thách thức nêu trên, luận án tập trung nghiên cứu các phương pháp học sâu nhằm giải quyết ba bài toán cốt lõi gồm: (i) phát hiện thiết bị trên lưới truyền tải điện từ ảnh UAV; (ii) nhận diện bất thường trên chuỗi cách điện trong điều kiện môi trường phức tạp; và (iii) nâng cao khả năng học của mô hình trong bối cảnh thiếu hụt dữ liệu bất thường. Các kết quả nghiên cứu kỳ vọng góp phần xây dựng hệ thống giám sát lưới điện thông minh theo hướng tự động hóa, hỗ trợ công tác quản lý tài sản và bảo trì dự báo trong quá trình chuyển đổi số ngành điện lực.

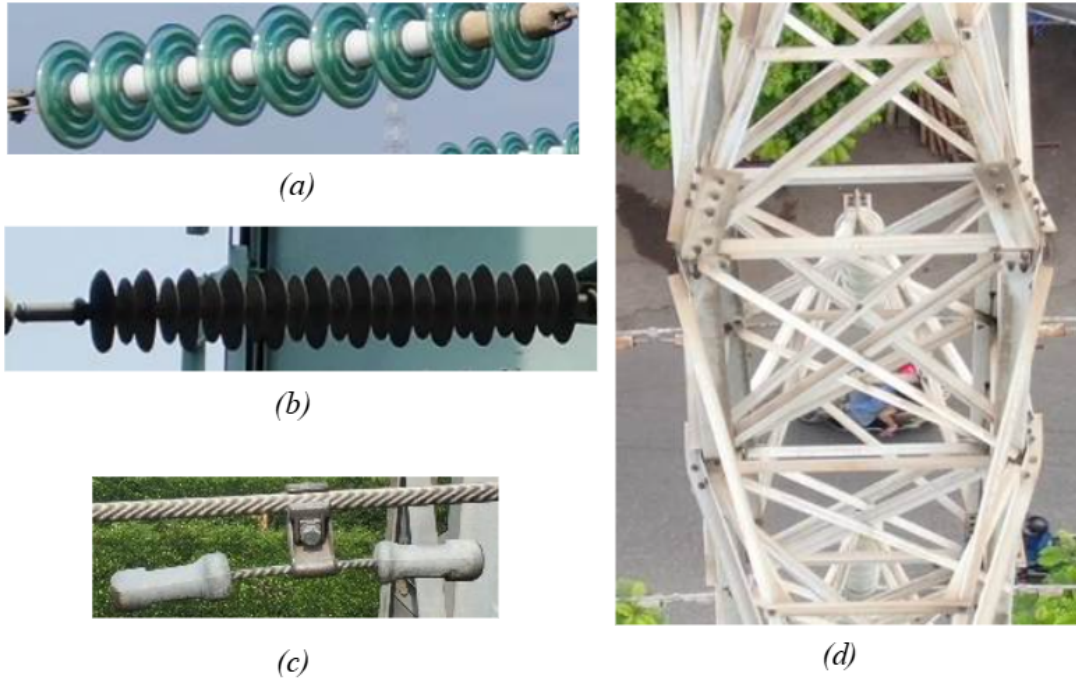
1.1.1 Phát hiện thiết bị trên lưới truyền tải điện cao thế

Sự phát triển nhanh chóng của thiết bị bay không người lái (UAV), kết hợp với những tiến bộ trong lĩnh vực thị giác máy tính và học sâu, đã tạo ra những thay đổi đáng kể trong công tác giám sát và quản lý lưới truyền tải điện. Thay vì phụ thuộc chủ yếu vào các hoạt động kiểm tra thủ công, dữ liệu hình ảnh và video thu nhận từ UAV có thể được khai thác để tự động nhận dạng, theo dõi và đánh giá tình trạng vận hành của các thiết bị trên đường dây truyền tải. Đây được xem là một trong những hướng ứng dụng quan trọng góp phần thúc đẩy quá trình chuyển đổi số và hiện thực hóa mô hình lưới điện thông minh trong ngành điện lực [1, 2].

Trong quy trình phân tích dữ liệu UAV, nhiệm vụ xác định và định vị các thiết bị trên lưới điện đóng vai trò nền tảng đối với toàn bộ hệ thống giám sát tự động. Mục tiêu của bài toán là nhận biết vị trí xuất hiện cũng như phân loại các đối tượng như chuỗi cách điện, dây dẫn, khóa néo, kẹp nối, tạ chống rung, cột điện và các phụ kiện đường dây khác. Kết quả thường được biểu diễn dưới dạng các hộp giới hạn (bounding boxes) hoặc các vùng đối tượng tương ứng, cung cấp thông tin không gian cần thiết

cho các bước phân tích tiếp theo [6, 8, 9].

Đối với hệ thống truyền tải điện, việc nhận dạng chính xác các thiết bị không chỉ hỗ trợ công tác quản lý tài sản và xây dựng cơ sở dữ liệu số mà còn tạo nền tảng cho các ứng dụng phân tích thông minh như đánh giá tình trạng vận hành, nhận diện hư hỏng, phát hiện bất thường và dự báo nguy cơ sự cố. Một số thiết bị điển hình thường xuất hiện trên lưới truyền tải điện cao thế được minh họa trong Hình 1.1.



Hình 1.1. Các thiết bị trên lưới truyền tải điện: (a) chuỗi cách điện thủy tinh; (b) chuỗi cách điện silicon; (c) tạ chống rung; (d) cột thép hình

Mặc dù các mô hình phát hiện đối tượng hiện đại đã đạt được hiệu quả cao trên nhiều bộ dữ liệu tổng quát, việc ứng dụng chúng vào bài toán giám sát lưới điện từ UAV vẫn tồn tại nhiều thách thức đặc thù. Trước hết, các thiết bị trên đường dây thường chỉ chiếm một phần rất nhỏ trong ảnh có độ phân giải cao, gây khó khăn cho quá trình học đặc trưng và định vị chính xác đối tượng. Bên cạnh đó, sự thay đổi về độ cao bay, góc chụp và khoảng cách quan sát làm xuất hiện sự khác biệt lớn về kích thước, hình dạng và hướng của đối tượng trong ảnh. Ngoài ra, nền ảnh thường chứa nhiều thành phần phức tạp như cây cối, địa hình tự nhiên, công trình xây dựng hoặc bầu trời, làm gia tăng mức độ nhiễu và ảnh hưởng đến hiệu quả nhận dạng [3, 5].

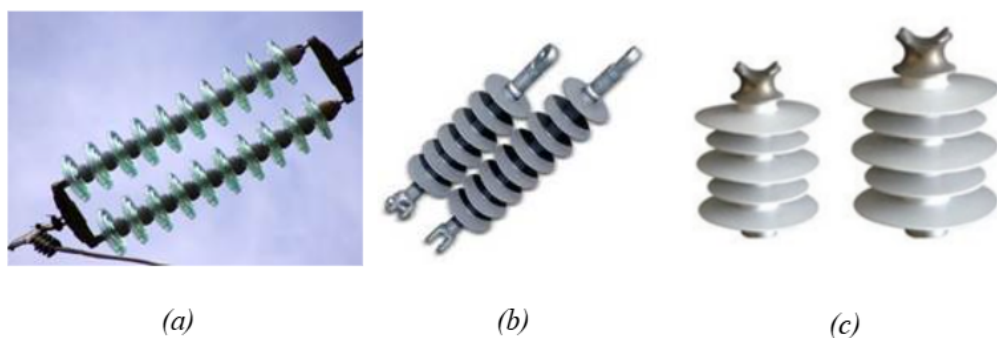
Một đặc điểm quan trọng khác là nhiệm vụ phát hiện thiết bị thường đóng vai trò bước tiền xử lý cho các bài toán giám sát ở mức cao hơn. Để nhận diện các dạng bất thường như nứt vỡ bát cách điện, mất bát cách điện, ăn mòn phụ kiện, dây dẫn hư hỏng hoặc sự xuất hiện của vật thể lạ trên đường dây, hệ thống trước tiên phải xác định chính xác vùng chứa thiết bị cần kiểm tra. Vì vậy, chất lượng của giai đoạn phát

hiện ảnh hưởng trực tiếp đến độ chính xác của các bước phân tích tiếp theo. Điều này đặt ra yêu cầu nghiên cứu các mô hình có khả năng nhận dạng hiệu quả các đối tượng kích thước nhỏ, hoạt động ổn định trong môi trường phức tạp và đáp ứng yêu cầu xử lý thời gian thực, phục vụ triển khai thực tế trên các hệ thống giám sát lưới điện dựa trên UAV.

1.1.2 Phát hiện bất thường trên chuỗi cách điện

Trong hệ thống truyền tải điện cao thế, chuỗi cách điện là một trong những thiết bị quan trọng nhất, có chức năng cách ly điện giữa dây dẫn và kết cấu cột điện, đồng thời chịu tải trọng cơ học của đường dây truyền tải. Chất lượng làm việc của chuỗi cách điện ảnh hưởng trực tiếp đến độ an toàn, độ tin cậy và tính liên tục trong vận hành của hệ thống điện. Trong quá trình khai thác, thiết bị này phải làm việc lâu dài dưới tác động của các điều kiện môi trường khắc nghiệt như nhiệt độ cao, độ ẩm lớn, gió bão, sương muối, ô nhiễm công nghiệp và quá trình lão hóa vật liệu. Những tác động này có thể làm phát sinh nhiều dạng bất thường và hư hỏng, gây suy giảm khả năng cách điện hoặc độ bền cơ học của thiết bị [1, 11].

Về cấu trúc, chuỗi cách điện thường được hình thành từ nhiều đơn vị cách điện liên kết nối tiếp nhau thông qua các phụ kiện kim loại, tạo thành một cấu trúc mô-đun có tính lặp cao. Tùy theo cấp điện áp và điều kiện vận hành, số lượng phân tử trong mỗi chuỗi có thể thay đổi nhằm đáp ứng yêu cầu cách điện và chịu lực của hệ thống. Hiện nay, các loại chuỗi cách điện được sử dụng phổ biến trên lưới truyền tải điện bao gồm chuỗi cách điện thủy tinh, chuỗi cách điện sứ và chuỗi cách điện polymer composite, như minh họa trong Hình 1.2.



Hình 1.2. Các loại chuỗi cách điện: (a) chuỗi cách điện thủy tinh; (b) chuỗi cách điện polymer silicon; (c) chuỗi cách điện sứ

Từ góc độ giám sát vận hành, các bất thường trên chuỗi cách điện có thể xuất

hiện dưới nhiều dạng khác nhau như mất bát cách điện, nứt vỡ bề mặt, ăn mòn phụ kiện kim loại, nhiễm bẩn, phóng điện bề mặt hoặc suy giảm đặc tính cách điện do lão hóa vật liệu. Phần lớn các hư hỏng này phát triển dần theo thời gian và thường khó nhận biết bằng các phương pháp kiểm tra trực quan truyền thống, đặc biệt đối với các tuyến đường dây đi qua khu vực địa hình phức tạp hoặc khó tiếp cận. Nếu không được phát hiện và xử lý kịp thời, các bất thường có thể dẫn đến hiện tượng phóng điện, ngắn mạch hoặc sự cố mất điện, gây ảnh hưởng đến độ tin cậy vận hành của hệ thống truyền tải điện [11].

Sự phát triển của UAV đã tạo điều kiện thuận lợi cho việc thu thập dữ liệu hình ảnh phục vụ giám sát chuỗi cách điện trên quy mô lớn. Tuy nhiên, việc tự động phát hiện bất thường từ ảnh UAV vẫn là một bài toán đầy thách thức. Các đối tượng bất thường thường có kích thước rất nhỏ so với toàn bộ ảnh, xuất hiện với nhiều hình dạng và mức độ hư hỏng khác nhau. Đồng thời, sự thay đổi về góc quan sát, khoảng cách chụp, điều kiện ánh sáng và hiện tượng che khuất một phần bởi dây dẫn hoặc các thành phần kết cấu khác làm gia tăng đáng kể độ khó của bài toán. Bên cạnh đó, số lượng mẫu bất thường trong thực tế thường rất hạn chế so với dữ liệu bình thường, dẫn đến hiện tượng mất cân bằng dữ liệu nghiêm trọng trong quá trình huấn luyện mô hình học sâu.

Dưới góc độ thị giác máy tính, phát hiện bất thường trên chuỗi cách điện có thể được xem là sự kết hợp giữa bài toán phát hiện đối tượng và bài toán nhận diện bất thường. Hệ thống cần vừa xác định chính xác vị trí của chuỗi cách điện trong ảnh, vừa nhận biết các dấu hiệu hư hỏng hoặc trạng thái bất thường của thiết bị. Do đó, hiệu quả của bài toán phụ thuộc đồng thời vào khả năng định vị chính xác đối tượng, trích xuất đặc trưng biểu diễn và học được sự khác biệt giữa các trạng thái bình thường và bất thường.

Trong các hệ thống giám sát lưới điện thông minh, kết quả phát hiện bất thường trên chuỗi cách điện không chỉ phục vụ công tác kiểm tra định kỳ mà còn là cơ sở cho các chiến lược bảo trì dự báo (predictive maintenance), hỗ trợ lập kế hoạch sửa chữa và giảm thiểu nguy cơ sự cố. Vì vậy, nghiên cứu các phương pháp học sâu nhằm nâng cao khả năng phát hiện bất thường trên chuỗi cách điện từ ảnh UAV là một hướng nghiên cứu có ý nghĩa khoa học và giá trị thực tiễn cao đối với ngành điện lực hiện nay.

1.1.3 Đặc điểm và thách thức của dữ liệu UAV trong giám sát lưới điện

UAV như nền tảng thu nhận dữ liệu cho giám sát lưới điện

Sự phát triển của máy bay không người lái (Unmanned Aerial Vehicles – UAV) đã mở ra hướng tiếp cận mới trong kiểm tra và giám sát lưới điện cao thế. So với các

phương pháp kiểm tra truyền thống như tuần tra trực tiếp hoặc sử dụng trực thăng chuyên dụng, UAV có nhiều ưu điểm nổi bật như tính linh hoạt cao, chi phí vận hành thấp, mức độ an toàn cao hơn và khả năng tiếp cận các khu vực địa hình phức tạp [12, 13].

Một trong những ưu điểm quan trọng nhất của UAV là khả năng thu nhận dữ liệu ảnh và video có độ phân giải cao từ nhiều góc nhìn khác nhau. UAV có thể tiếp cận gần các thiết bị trên lưới điện và ghi nhận chi tiết trạng thái của chuỗi cách điện, dây dẫn hoặc phụ kiện treo mà không cần cắt điện. Ngoài ảnh RGB thông thường, UAV còn có thể tích hợp các cảm biến khác như camera nhiệt (thermal camera), camera theo dõi tín hiệu bức xạ tử ngoại (UV camera), LiDAR hoặc hyperspectral sensor nhằm hỗ trợ phát hiện các dạng bất thường khó quan sát bằng mắt thường [4, 14]. Bên cạnh đó, UAV cho phép tự động hóa quá trình thu thập dữ liệu trên diện rộng với tần suất cao hơn nhiều so với phương pháp kiểm tra truyền thống. Điều này tạo ra lượng dữ liệu lớn phục vụ xây dựng các hệ thống phân tích dựa trên trí tuệ nhân tạo và học sâu. Theo khảo sát của Ahmed và đồng nghiệp [13], UAV hiện đang trở thành nền tảng thu nhận dữ liệu quan trọng nhất trong các hệ thống kiểm tra lưới điện thông minh thế hệ mới.

Việc ứng dụng UAV trong ngành điện lực đã phát triển mạnh tại nhiều quốc gia. Tại Hoa Kỳ, các công ty điện lực lớn như Ameren và American Electric Power (AEP) đã triển khai UAV như một phần trong chiến lược chuyển đổi số của hệ thống quản lý tài sản lưới điện. UAV được sử dụng để bay dọc theo hành lang tuyến truyền tải điện nhằm thu thập dữ liệu hình ảnh chi tiết của các thành phần như dây dẫn, chuỗi cách điện, cột điện và các phụ kiện treo. Dữ liệu thu thập từ UAV sau đó được phân tích để phát hiện các dấu hiệu hư hỏng, ăn mòn hoặc bất thường trên thiết bị, qua đó giảm nhu cầu kiểm tra trực tiếp tại hiện trường và nâng cao mức độ an toàn cho nhân viên vận hành. So với phương pháp kiểm tra truyền thống bằng trực thăng hoặc tuần tra thủ công, UAV giúp giảm đáng kể chi phí và thời gian kiểm tra, đồng thời tăng tần suất giám sát hệ thống [15].

Tại châu Âu, UAV đã được tích hợp với các công nghệ trí tuệ nhân tạo nhằm tự động hóa quá trình phân tích dữ liệu kiểm tra lưới điện. Các hệ thống UAV không chỉ thực hiện nhiệm vụ thu thập dữ liệu mà còn được kết hợp với các thuật toán thị giác máy tính để phát hiện các khiếm khuyết trên thiết bị điện như nứt vỡ cách điện, ăn mòn phụ kiện hoặc các điểm phát nhiệt bất thường. Việc kết hợp UAV với các mô hình học sâu giúp nâng cao hiệu quả phát hiện lỗi và giảm sự phụ thuộc vào kiểm tra thủ công của con người. Ngoài ra, UAV còn được sử dụng để khảo sát hành lang tuyến, phát hiện cây cối hoặc vật cản có nguy cơ ảnh hưởng đến an toàn vận hành của đường dây truyền tải điện [16].

Trong khu vực ASEAN, UAV cũng đang được ứng dụng ngày càng rộng rãi trong công tác quản lý vận hành lưới điện. Tại Thái Lan, Electricity Generating Authority of Thailand (EGAT) đã triển khai UAV để kiểm tra các tuyến truyền tải điện có điện áp từ 115 kV đến 500 kV. Các UAV được trang bị camera quang học và cảm biến nhiệt nhằm phát hiện sớm các bất thường trên thiết bị như nứt vỡ chuỗi cách điện hoặc các điểm phát nhiệt bất thường [17].

Tại Malaysia, công ty Tenaga Nasional Berhad (TNB) đã sử dụng UAV để khảo sát các cột điện, chuỗi cách điện và trạm biến áp. Theo các báo cáo kỹ thuật của TNB, việc sử dụng UAV giúp giảm đáng kể thời gian kiểm tra và nâng cao độ chính xác của quá trình đánh giá tình trạng thiết bị. Trong khi đó, tại Indonesia, công ty điện lực quốc gia Perusahaan Listrik Negara (PLN) sử dụng UAV để giám sát các tuyến truyền tải điện ở khu vực địa hình phức tạp hoặc xa xôi, nơi việc tiếp cận bằng phương pháp kiểm tra truyền thống gặp nhiều khó khăn.

Tại Singapore, các hệ thống UAV đã được tích hợp với các mô hình học sâu nhằm tự động phát hiện các bất thường trên lưới điện. Các hệ thống này cho phép phân tích dữ liệu hình ảnh thu thập từ UAV một cách tự động và phát hiện các hư hỏng trên thiết bị điện như chuỗi cách điện bị nứt vỡ hoặc biến dạng cấu trúc cơ khí.

Bảng 1.1 trình bày sự so sánh về tình hình ứng dụng UAV trong giám sát lưới điện tại một số khu vực trên thế giới.

Bảng 1.1. Ứng dụng UAV trong giám sát lưới điện tại các khu vực

Khu vực	Năm	Đơn vị triển khai	Công nghệ sử dụng	Ứng dụng chính	Đặc điểm nổi bật
Hoa Kỳ	2022	Ameren, AEP	UAV + AI + LiDAR	Kiểm tra đường dây truyền tải điện	Tích hợp UAV với hệ thống quản lý tài sản và phân tích dữ liệu tự động [18]
Châu Âu	2021	Công ty điện lực, viện nghiên cứu	UAV + Computer Vision + AI	Kiểm tra thiết bị điện, hành lang tuyến	Kết hợp AI để phát hiện lỗi chuỗi cách điện [19]
Nhật Bản	2022	TEPCO, Kansai Electric	UAV + LiDAR + Image Processing	Giám sát đường dây vùng đồi núi	Kiểm tra lưới điện trong địa hình phức tạp [20]

(Còn tiếp ở trang sau)

(Tiếp theo Bảng 1.1)

Khu vực	Năm	Đơn vị triển khai	Công nghệ sử dụng	Ứng dụng chính	Đặc điểm nổi bật
Trung Quốc	2023	State Grid Corporation of China	UAV + Deep Learning + Thermal Camera	Phát hiện lỗi cách điện, hotspot	Triển khai UAV tự động quy mô lớn [21]
Úc	2022	Energy Queensland	UAV + Thermal Imaging + AI	Kiểm tra thiết bị điện	Giảm chi phí bảo trì và tăng an toàn vận hành [22]
ASEAN	2023	EGAT, TNB, PLN	UAV + Thermal Camera + AI	Kiểm tra chuỗi cách điện	Ứng dụng UAV trong bảo trì quy mô lớn [23]
Việt Nam	2023	EVNNPT, EVNNPC	UAV + Computer Vision + AI	Kiểm tra đường dây 220–500 kV	Bước đầu ứng dụng AI trong giám sát lưới điện [24]
Trung Quốc	2024	State Grid Research Institute	UAV + Edge AI	Phát hiện lỗi thời gian thực	Triển khai UAV tự động tại trạm điện thông minh [25]
Hàn Quốc	2024	KEPCO	UAV + AI + Thermal Camera	Phát hiện quá nhiệt thiết bị	Hỗ trợ bảo trì dự báo [26]
Châu Âu	2025	European Smart Grid Projects	UAV + AI + Digital Twin	Mô phỏng tình trạng thiết bị	Kết hợp Digital Twin tối ưu vận hành [27]
Trung Quốc	2025	State Grid UAV Program	UAV Swarm + AI	Giám sát diện rộng	Ứng dụng đàn UAV (UAV swarm) để kiểm tra lưới điện [28]

(Kết thúc Bảng 1.1)

Tại Việt Nam, việc ứng dụng UAV trong giám sát lưới truyền tải điện đã được triển khai trong những năm gần đây bởi Tổng công ty Truyền tải điện Quốc gia (EVNNPT) và các công ty truyền tải điện khu vực. UAV được sử dụng để kiểm tra hành lang tuyến và tình trạng thiết bị trên các đường dây truyền tải điện 220 kV và 500 kV. Thông qua việc thu thập hình ảnh độ phân giải cao, UAV giúp phát hiện các dấu hiệu hư hỏng trên chuỗi cách điện, phụ kiện treo, dây dẫn và kết cấu cột điện mà không cần tiếp cận trực tiếp tại hiện trường. Việc ứng dụng UAV giúp giảm khối lượng

lao động thủ công và nâng cao hiệu quả công tác quản lý vận hành lưới điện [29].

Theo các báo cáo của ngành điện lực, UAV không chỉ được sử dụng để thu thập dữ liệu hình ảnh mà còn được tích hợp với các hệ thống phân tích dữ liệu nhằm hỗ trợ phát hiện sớm các rủi ro trên lưới điện. Việc sử dụng UAV giúp giảm đáng kể nhân lực cho công tác kiểm tra đường dây và cho phép phát hiện sớm các nguy cơ như dây dẫn võng quá mức, điểm phát nhiệt bất thường hoặc các vi phạm hành lang tuyến [30]. Ngoài ra, UAV còn được sử dụng để khảo sát địa hình, hỗ trợ thiết kế và xây dựng các tuyến truyền tải điện mới.

Bên cạnh các ứng dụng thực tế trong ngành điện lực, nhiều nghiên cứu trong nước cũng đã bắt đầu quan tâm đến việc ứng dụng UAV và các kỹ thuật xử lý ảnh trong giám sát lưới điện. Một số nghiên cứu tập trung vào việc phát hiện các thành phần của hệ thống lưới điện từ ảnh UAV như chuỗi cách điện, dây dẫn hoặc cột điện bằng các mô hình học sâu. Các phương pháp này cho thấy tiềm năng lớn trong việc tự động hóa quá trình phân tích dữ liệu UAV và hỗ trợ phát hiện bất thường trên thiết bị điện.

Mặc dù đã có những bước tiến đáng kể trong việc ứng dụng UAV cho giám sát lưới điện, việc áp dụng các phương pháp trí tuệ nhân tạo trong phân tích dữ liệu UAV tại Việt Nam vẫn còn ở giai đoạn phát triển ban đầu. Phần lớn các hệ thống hiện nay chủ yếu tập trung vào việc thu thập dữ liệu hình ảnh phục vụ kiểm tra trực quan, trong khi các nghiên cứu về tự động hóa phân tích dữ liệu bằng các mô hình học sâu vẫn còn hạn chế. Điều này cho thấy cần thiết phải nghiên cứu các phương pháp phát hiện bất thường trên chuỗi cách điện từ ảnh UAV, nhằm nâng cao mức độ tự động hóa và hiệu quả của hệ thống giám sát lưới điện trong bối cảnh chuyển đổi số của ngành năng lượng.

Những thập kỉ gần đây, sự phát triển của phần mềm và phần cứng vi điện tử, công nghệ định vị và truyền thông đổi mới, cùng những đột phá trong công nghệ vật liệu và năng lượng, nền tảng máy bay không người lái (UAV) đã trở nên phổ biến. UAV được sử dụng với mục tiêu thu thập hình ảnh trong nhiều lĩnh vực như: giao thông [31, 32], quản lý tài nguyên thiên nhiên [33], khảo sát ứng dụng dân dụng [34], quân sự [35], nông nghiệp [36], ngành điện trong nước và thế giới cũng sử dụng thiết bị bay không người lái để phục vụ công tác nghiên cứu, kiểm tra, giám sát [37, 38].

Ứng dụng UAV trong kiểm tra, giám sát lưới điện không chỉ là một giải pháp kỹ thuật hiện đại mà còn là yếu tố then chốt để chuyển đổi số trong công tác quản lý vận hành hệ thống điện, giúp nâng cao độ tin cậy, an toàn và năng suất cho các doanh nghiệp truyền tải điện trong bối cảnh lưới điện ngày càng mở rộng và phức tạp. Để có thể sử dụng trong hoạt động bay giám sát lưới điện cao thế, máy bay không người lái cần đảm bảo một số điều kiện đặc biệt như khả năng chịu đựng sức gió, tải .

Một số nghiên cứu gần đây đã chứng minh hiệu quả của UAV kết hợp với các thuật toán học sâu trong nhiệm vụ nhận diện và phát hiện các thành phần của lưới điện, chẳng hạn như đường dây truyền tải, cột điện, và các bất thường trên thiết bị từ dữ liệu ảnh thu thập bằng UAV [10, 39]. Các nghiên cứu tổng quan cũng chỉ ra rằng UAV cung cấp một giải pháp toàn diện cho giám sát tự động các cơ sở hạ tầng điện lực, với tiềm năng ứng dụng rộng trong các hệ thống kiểm tra, bảo trì và đánh giá trạng thái thiết bị [2, 3, 5]

So với các nền tảng truyền thống, UAV cho phép thu thập dữ liệu video một cách linh hoạt mà không bị giới hạn về địa hình đồng thời khắc phục được tình trạng mất thông tin do điều kiện thời tiết và thời gian quan sát [3]. Với khả năng tiếp cận gần đối tượng quan sát, UAV có thể cung cấp dữ liệu chi tiết, đa góc nhìn, phục vụ hiệu quả cho các bài toán giám sát, kiểm tra và phân tích trạng thái thiết bị trong hệ thống hạ tầng kỹ thuật. Ảnh thu nhận từ thiết bị bay không người lái (Unmanned Aerial Vehicle – UAV) ngày càng được sử dụng rộng rãi trong công tác kiểm tra và giám sát lưới điện truyền tải. So với ảnh vệ tinh và ảnh chụp mặt đất, ảnh UAV có những đặc trưng riêng biệt về độ phân giải, hình học, phân bố kích thước đối tượng và điều kiện môi trường thu nhận. Các đặc trưng này tác động trực tiếp đến thiết kế mô hình học sâu và chiến lược xử lý dữ liệu trong bài toán phát hiện thiết bị điện

Những điểm lưu ý khi sử dụng UAV trong kiểm tra lưới điện

Trong quá trình thu nhận dữ liệu trên lưới điện, để bảo đảm chất lượng hình ảnh, UAV thường được điều khiển tiếp cận gần đối tượng kiểm tra. Tuy nhiên, khoảng cách bay cần được tính toán phù hợp nhằm bảo đảm an toàn vận hành, bao gồm khoảng cách an toàn điện và độ dự trữ an toàn bay.

Khi UAV hoạt động gần đường dây điện cao thế, điện trường và từ trường có thể ảnh hưởng đến cảm biến, mô-đun truyền thông và hệ thống điều khiển của thiết bị bay. Do đó, UAV sử dụng trong kiểm tra lưới điện cần bảo đảm khả năng chống nhiễu điện từ và duy trì khoảng cách vận hành an toàn đối với đường dây truyền tải điện. Bảng 1.2 trình bày khoảng cách an toàn khuyến nghị giữa UAV và các cấp điện áp trên lưới truyền tải điện.

Trong quá trình kiểm tra lưới điện cao thế bằng UAV, khoảng cách an toàn giữa thiết bị bay và dây dẫn là yêu cầu bắt buộc nhằm hạn chế nguy cơ phóng điện, nhiễu điện từ và mất ổn định bay. Theo các khuyến nghị kỹ thuật quốc tế [40, 41] và tiêu chuẩn tại Việt Nam, UAV cần duy trì khoảng cách vận hành phù hợp tương ứng với từng cấp điện áp của đường dây truyền tải điện.[42], UAV cần duy trì khoảng cách tối thiểu từ 3 m đối với đường dây 110 kV, 4 m đối với 220 kV và không nhỏ hơn 6 m đối với lưới 500 kV.

Đặc trưng của ảnh UAV trong giám sát lưới điện

Bảng 1.2. Khoảng cách an toàn giữa UAV và đường dây truyền tải điện cao thế

Cấp điện áp	Điện áp danh định	Khoảng cách tối thiểu (m)	Khoảng cách khuyến nghị (m)	Ghi chú kỹ thuật
Trung thế	≤ 35 kV	≥ 2.0	3–5	Tránh nhiễu điện từ, phù hợp UAV dân dụng
Cao thế	66–110 kV	≥ 3.0	5–7	Phù hợp kiểm tra phụ kiện và chuỗi cách điện
Cao thế	220 kV	≥ 4.0	6–10	Nguy cơ phóng điện tăng cao khi độ ẩm lớn
Siêu cao thế	500 kV	≥ 6.0	10–15	Yêu cầu UAV công nghiệp có khả năng chống nhiễu mạnh
Trạm biến áp	110–500 kV	≥ 5.0	≥ 8	Cần tránh vùng có điện trường và từ trường mạnh

Thứ nhất, ảnh UAV có độ phân giải không gian cao do được thu nhận ở độ cao bay tương đối thấp. Theo Colomina và Molina [43], các hệ thống UAV có khả năng thu nhận dữ liệu với kích thước điểm ảnh mặt đất (Ground Sampling Distance – GSD) ở mức centimet, vượt trội so với nhiều hệ thống ảnh vệ tinh thương mại. Độ phân giải không gian cao cho phép quan sát chi tiết các cấu trúc nhỏ như chuỗi sứ cách điện, kẹp néo, tạ chống rung và các phụ kiện treo trên đường dây truyền tải. Tuy nhiên, đặc điểm này đồng thời làm gia tăng sự biến thiên đa tỉ lệ (multi-scale variation) trong ảnh, khi các đối tượng có thể xuất hiện với kích thước rất khác nhau tùy thuộc vào khoảng cách quan sát và góc chụp. Do đó, các mô hình thị giác máy tính cần được thiết kế với khả năng trích xuất đặc trưng đa tầng nhằm khai thác hiệu quả thông tin ở nhiều mức độ phân giải khác nhau [44, 45].

Thứ hai, ảnh UAV thể hiện sự biến thiên lớn về hình học và phối cảnh. Do UAV có khả năng thay đổi linh hoạt độ cao và góc quay camera, ảnh có thể được thu nhận theo phương xiên hoặc ngang, tạo ra biến dạng phối cảnh và thay đổi hình dạng biểu kiến của đối tượng. Nghiên cứu của Zhang, C. và Kovacs, J. M. [46] chỉ ra rằng tính linh hoạt trong góc chụp là ưu điểm quan trọng của UAV nhưng đồng thời làm tăng độ phức tạp trong xử lý hình học, chuẩn hóa dữ liệu và trích xuất đặc trưng trong các hệ thống thị giác máy tính.

Ngoài ra, sự thay đổi liên tục về khoảng cách quan sát giữa UAV và đối tượng cũng dẫn đến sự biến thiên mạnh về tỷ lệ kích thước của đối tượng trong ảnh. Điều này đặc biệt ảnh hưởng đến các thiết bị trên lưới truyền tải điện, vốn thường có cấu trúc mảnh và kích thước nhỏ so với toàn bộ khung hình. Theo khảo sát của Tang và cộng sự [45], bài toán phát hiện đối tượng trong ảnh UAV cần giải quyết đồng thời ba thách thức chính: (i) kích thước đối tượng nhỏ so với ảnh nền, (ii) mật độ đối tượng cao trong một số kịch bản quan sát và (iii) sự phức tạp của nền ảnh do môi trường tự

nhiên hoặc cấu trúc nhân tạo xung quanh.

Thứ ba, dữ liệu UAV trong giám sát lưới điện thường mang đặc trưng ưu thế đối tượng kích thước nhỏ (small-object dominance) [47]. Các thành phần như chuỗi sứ cách điện, dây dẫn hoặc phụ kiện treo thường chiếm tỷ lệ rất nhỏ trong toàn bộ khung hình. Khi số lượng điểm ảnh đại diện cho đối tượng hạn chế, thông tin đặc trưng có thể bị suy giảm đáng kể qua các tầng tích chập sâu của mạng neuron, dẫn đến giảm khả năng nhận diện chính xác. Tong và cộng sự [48] chỉ ra rằng hiệu năng của các mô hình phát hiện đối tượng suy giảm rõ rệt khi kích thước đối tượng giảm xuống dưới một ngưỡng nhất định, đặc biệt trong các bài toán có nền ảnh phức tạp. Do đó, việc sử dụng các cơ chế biểu diễn đặc trưng đa tỉ lệ và tích hợp thông tin ngữ cảnh không gian được xem là hướng tiếp cận quan trọng nhằm cải thiện khả năng phát hiện các đối tượng nhỏ.

Thứ tư, ảnh UAV trong môi trường lưới điện thường chứa nền cảnh phức tạp và đa dạng. Các đường dây truyền tải điện thường đi qua nhiều loại địa hình khác nhau như đồi núi, rừng cây, khu dân cư hoặc khu công nghiệp. Sự xuất hiện của các yếu tố nền có màu sắc và kết cấu tương đồng với thiết bị điện có thể làm tăng nguy cơ nhầm lẫn trong quá trình phát hiện đối tượng. Nghiên cứu của Liu và cộng sự [49] cho thấy sự phức tạp của nền ảnh là một trong những nguyên nhân chính dẫn đến các lỗi phát hiện sai (false positives) trong các hệ thống kiểm tra lưới điện dựa trên UAV.

Thứ năm, do được thu nhận trong môi trường ngoài trời, ảnh UAV chịu ảnh hưởng mạnh của điều kiện chiếu sáng và các yếu tố khí quyển. Các thay đổi về cường độ ánh sáng, bóng đổ, sương mù hoặc điều kiện thời tiết có thể làm gia tăng biến thiên nội lớp (intra-class variation) của đối tượng trong dữ liệu. Bên cạnh đó, chuyển động của UAV trong quá trình bay có thể gây ra hiện tượng rung hoặc mờ chuyển động, đặc biệt trong điều kiện gió mạnh hoặc khi tốc độ bay cao. Các yếu tố này làm suy giảm chất lượng hình ảnh và ảnh hưởng trực tiếp đến hiệu năng của các mô hình phát hiện đối tượng [45, 47].

Tóm lại, dữ liệu ảnh UAV phục vụ giám sát lưới truyền tải điện được đặc trưng bởi năm yếu tố chính:

- i) độ phân giải không gian cao;
- ii) biến thiên hình học và phối cảnh lớn;
- iii) các đối tượng chiếm tỷ lệ rất nhỏ trong toàn bộ khung hình;
- iv) nền cảnh phức tạp và đa dạng;
- v) sự biến động của điều kiện chiếu sáng cùng nhiều chuyển động.

Các đặc trưng này tạo ra những thách thức đáng kể đối với các hệ thống phát hiện và nhận diện thiết bị điện trên lưới truyền tải. Đồng thời, chúng cũng định hướng cho việc lựa chọn kiến trúc mạng, thiết kế cơ chế biểu diễn đặc trưng đa tỉ lệ và xây dựng các chiến lược tăng cường dữ liệu phù hợp trong các nghiên cứu phát hiện bất thường dựa trên ảnh UAV.

1.2 Các nghiên cứu liên quan

Trong những năm gần đây, cùng với sự phát triển mạnh mẽ của học sâu (Deep Learning) và thị giác máy tính (Computer Vision), bài toán phát hiện thiết bị và nhận diện bất thường trên lưới điện đã trở thành một trong những hướng nghiên cứu được quan tâm trong lĩnh vực kiểm tra và giám sát hệ thống điện. Các công trình nghiên cứu chủ yếu khai thác dữ liệu ảnh kết hợp với các mô hình học sâu nhằm tự động hóa quá trình phát hiện thiết bị, đánh giá tình trạng vận hành và hỗ trợ cảnh báo sớm các sự cố, góp phần nâng cao độ tin cậy và hiệu quả vận hành của hệ thống điện.

Dựa trên nguồn dữ liệu sử dụng, mục tiêu nghiên cứu và phương pháp tiếp cận, các công trình liên quan có thể được phân thành ba nhóm chính.

- **Nhóm thứ nhất** gồm các nghiên cứu phát hiện thiết bị và nhận diện bất thường từ ảnh thông thường, như ảnh chụp cận cảnh, ảnh thu nhận từ mặt đất hoặc trong điều kiện phòng thí nghiệm. Các nghiên cứu này có ưu điểm là chất lượng ảnh cao, đối tượng quan sát rõ ràng và nền ảnh đơn giản, tạo điều kiện thuận lợi cho việc huấn luyện và đánh giá mô hình. Tuy nhiên, khả năng bao quát phạm vi kiểm tra còn hạn chế và chưa phản ánh đầy đủ các điều kiện vận hành thực tế của lưới điện truyền tải;
- **Nhóm thứ hai** tập trung vào các nghiên cứu sử dụng ảnh thu nhận từ thiết bị bay không người lái (UAV). Nhờ khả năng cơ động cao, phạm vi quan sát rộng và chi phí triển khai hợp lý, UAV đã trở thành nền tảng thu thập dữ liệu quan trọng trong các hệ thống kiểm tra lưới điện hiện đại. Tuy nhiên, ảnh UAV cũng đặt ra nhiều thách thức đối với các mô hình phát hiện, bao gồm sự thay đổi về góc quan sát, kích thước đối tượng nhỏ, nền ảnh phức tạp, điều kiện chiếu sáng thay đổi và hiện tượng mất cân bằng dữ liệu giữa các lớp.
- **Nhóm thứ ba** tập trung vào các phương pháp sinh dữ liệu tổng hợp nhằm giải quyết tình trạng khan hiếm mẫu bất thường và mất cân bằng dữ liệu. Khác với hai nhóm nghiên cứu trước chủ yếu tập trung cải tiến mô hình phát hiện, hướng nghiên cứu này khai thác các mô hình sinh dữ liệu để tạo ra các mẫu bất thường hiếm gặp hoặc khó thu thập trong thực tế, từ đó mở rộng tập huấn luyện, cân

bằng phân bố dữ liệu và nâng cao khả năng tổng quát hóa của mô hình. Đây là hướng nghiên cứu mới, đặc biệt phù hợp với các bài toán có phân phối dài đuôi (long-tail distribution), trong đó số lượng mẫu của các lớp bất thường thường rất hạn chế.

Có thể thấy, hai nhóm nghiên cứu đầu chủ yếu hướng tới việc cải tiến kiến trúc mô hình nhằm nâng cao độ chính xác phát hiện, cải thiện khả năng nhận dạng các đối tượng kích thước nhỏ và tối ưu hiệu quả tính toán. Trong khi đó, nhóm nghiên cứu thứ ba tập trung vào việc nâng cao chất lượng dữ liệu huấn luyện thông qua các kỹ thuật sinh dữ liệu và tăng cường dữ liệu. Nhiều nghiên cứu gần đây cho thấy chất lượng và sự đa dạng của dữ liệu huấn luyện có ảnh hưởng trực tiếp đến khả năng học biểu diễn và hiệu năng của các mô hình học sâu, đặc biệt trong các bài toán có dữ liệu mất cân bằng hoặc phân phối dài đuôi.

Trên cơ sở phân loại trên, luận án tiến hành khảo sát và phân tích các công trình nghiên cứu tiêu biểu của từng nhóm nhằm đánh giá những kết quả đã đạt được, làm rõ ưu điểm, hạn chế của các hướng tiếp cận hiện có, đồng thời xác định các khoảng trống nghiên cứu cần tiếp tục giải quyết. Đặc biệt, luận án tập trung phân tích sự thiếu hụt các phương pháp sinh dữ liệu có khả năng khai thác tri thức miền trong lĩnh vực lưới điện truyền tải để tạo ra dữ liệu tổng hợp có chất lượng và có tính hữu ích đối với các bài toán hạ nguồn, như phát hiện bất thường và nhận diện sự cố thiết bị. Những phân tích này là cơ sở khoa học để luận án đề xuất các phương pháp nghiên cứu được trình bày trong các chương tiếp theo.

1.2.1 Phát hiện thiết bị và bất thường từ ảnh thông thường

Phát hiện đối tượng (object detection) là một trong những bài toán cốt lõi của thị giác máy tính, với mục tiêu xác định vị trí và phân loại các đối tượng xuất hiện trong ảnh hoặc video. Khác với bài toán phân loại ảnh chỉ gán một nhãn duy nhất cho toàn bộ ảnh, phát hiện đối tượng đồng thời thực hiện hai nhiệm vụ: (i) định vị đối tượng thông qua hộp giới hạn (bounding box); và (ii) gán nhãn phân loại tương ứng cho từng đối tượng được phát hiện.

Trong thập kỷ gần đây, sự phát triển mạnh mẽ của học sâu đã tạo ra những bước tiến đột phá cho bài toán này. Từ các mô hình mạng nơ-ron tích chập (CNN) đến các kiến trúc dựa trên Transformer, độ chính xác và tốc độ xử lý đều được cải thiện đáng kể. Các phương pháp này hiện được ứng dụng rộng rãi trong nhiều lĩnh vực như xe tự hành, giám sát giao thông, robot thị giác và giám sát hạ tầng kỹ thuật, bao gồm cả hệ thống lưới điện thông minh [50–52].

Quá trình phát triển của các phương pháp phát hiện đối tượng hiện đại có thể được phân thành ba hướng tiếp cận chính [1]:

- Mô hình hai giai đoạn (Two-stage detectors);
- Mô hình một giai đoạn (One-stage detectors);
- Mô hình dựa trên kiến trúc Transformer.

1. Mô hình hai giai đoạn Các phương pháp tiêu biểu gồm R-CNN, Fast R-CNN, Faster R-CNN và Mask R-CNN, đã đặt nền móng cho bài toán phát hiện đối tượng hiện đại [6, 53–55].

R-CNN là một trong những mô hình đầu tiên áp dụng học sâu, sử dụng thuật toán Selective Search để tạo các vùng đề xuất (region proposals), sau đó trích xuất đặc trưng bằng CNN và phân loại bằng SVM. Tuy nhiên, R-CNN có hạn chế lớn về tốc độ do phải xử lý độc lập từng vùng đề xuất.

Fast R-CNN khắc phục bằng cách chia sẻ bản đồ đặc trưng cho toàn ảnh và sử dụng lớp RoI Pooling, giúp giảm đáng kể thời gian tính toán, nhưng vẫn phụ thuộc vào bước tạo vùng đề xuất bên ngoài.

Bước đột phá đến từ Faster R-CNN với việc tích hợp Mạng đề xuất vùng (Region Proposal Network - RPN) vào trong kiến trúc mạng, cho phép mô hình tự sinh vùng đề xuất và huấn luyện end-to-end, từ đó nâng cao cả tốc độ lẫn độ chính xác.

Mask R-CNN mở rộng Faster R-CNN bằng cách thêm nhánh dự đoán mặt nạ phân đoạn (segmentation mask), giúp cung cấp thông tin chi tiết về hình dạng đối tượng, đặc biệt hữu ích khi phân tích cấu trúc hoặc hư hỏng thiết bị.

2. Mô hình một giai đoạn Các kiến trúc như YOLO (You Only Look Once), SSD (Single Shot MultiBox Detector) và RetinaNet được phát triển nhằm tối ưu hóa tốc độ suy luận cho các ứng dụng thời gian thực [8, 52, 56].

YOLO tiếp cận bài toán phát hiện như một bài toán hồi quy trực tiếp: chia ảnh thành lưới và dự đoán đồng thời các hộp giới hạn, độ tin cậy và xác suất lớp cho mỗi ô lưới. Nhờ đó, YOLO đạt tốc độ rất cao. Tuy nhiên, các phiên bản đầu tiên gặp khó khăn với các đối tượng nhỏ hoặc xếp chồng do hạn chế về đặc trưng đa tỉ lệ.

Để khắc phục, các phiên bản cải tiến như YOLOv5, YOLOv7 và YOLOv8 đã tích hợp nhiều cải tiến ở phần mạng xương sống (backbone), neck như PANet, FPN và các cơ chế tăng cường đặc trưng đa tỉ lệ, cơ chế attention, giúp cải thiện đáng kể khả năng phát hiện đối tượng nhỏ [57–59].

SSD sử dụng các bản đồ đặc trưng ở nhiều độ phân giải khác nhau để phát hiện đối tượng ở nhiều kích thước. Trong khi đó, RetinaNet đề xuất hàm mất mát

Focal Loss nhằm giải quyết bài toán mất cân bằng nghiêm trọng giữa foreground và background, đặc biệt hiệu quả với tập dữ liệu có tỷ lệ đối tượng nhỏ cao.

3. Mô hình dựa trên Transformer Gần đây, DETR (DEtection TRansformer) đã mở ra hướng tiếp cận mới khi xem bài toán phát hiện là bài toán dự đoán tập hợp (set prediction) [9]. DETR sử dụng kiến trúc encoder-decoder của Transformer và cơ chế attention để mô hình hóa quan hệ toàn cục giữa các vùng ảnh, loại bỏ hoàn toàn các thành phần thủ công như anchor box hay thuật toán Non-Maximum Suppression (NMS).

Tuy nhiên, DETR tồn tại nhược điểm là thời gian huấn luyện hội tụ chậm và hiệu suất kém với đối tượng nhỏ. Các biến thể cải tiến như Deformable DETR [60] đã giải quyết vấn đề này bằng cách sử dụng deformable attention chỉ tập trung vào một tập nhỏ các điểm lấy mẫu quan trọng, giúp tăng tốc hội tụ và cải thiện độ chính xác trên đối tượng nhỏ.

1.2.2 Phát hiện thiết bị và bất thường từ ảnh UAV

Cùng với sự phát triển mạnh mẽ của học sâu và nền tảng UAV, nhiều nghiên cứu đã tập trung vào bài toán phát hiện thiết bị và nhận diện bất thường trên lưới truyền tải điện. Các công trình hiện đại tập trung giải quyết ba thách thức chính gồm: (i) phát hiện đối tượng kích thước nhỏ, (ii) xử lý nền ảnh phức tạp và (iii) đảm bảo khả năng suy luận thời gian thực trên nền tảng UAV có giới hạn tài nguyên tính toán [50, 51, 61].

Các phương pháp dựa trên mô hình hai giai đoạn

Các mô hình hai giai đoạn (two-stage detectors) là nhóm phương pháp nền tảng trong bài toán phát hiện đối tượng, trong đó tiêu biểu là Faster R-CNN [6]. Phương pháp này hoạt động theo hai bước: trước tiên sinh các vùng đề xuất (region proposals), sau đó thực hiện phân loại và tinh chỉnh vị trí đối tượng.

Nhờ cơ chế tập trung vào các vùng quan tâm, Faster R-CNN cho độ chính xác cao trong nhiều bài toán, đặc biệt phù hợp với dữ liệu UAV có đối tượng nhỏ như chuỗi cách điện hoặc phụ kiện đường dây. Một số nghiên cứu trong giám sát lưới điện cho thấy mô hình này đạt hiệu quả ổn định trong phát hiện thiết bị có kích thước nhỏ và trong điều kiện nền phức tạp [62, 63].

Tuy nhiên, hạn chế lớn của nhóm phương pháp này là chi phí tính toán cao do phải thực hiện hai giai đoạn suy luận tách biệt. Điều này dẫn đến tốc độ xử lý thấp, khó đáp ứng yêu cầu thời gian thực trong các hệ thống UAV. Ngoài ra, hiệu năng của mô hình cũng suy giảm khi đối tượng quá nhỏ hoặc bị che khuất một phần bởi cây cối, dây dẫn hoặc điều kiện quan sát không thuận lợi.

Do đó, mặc dù đạt độ chính xác tốt, các phương pháp hai giai đoạn thường chỉ phù hợp với bài toán phân tích ngoại tuyến hơn là triển khai thời gian thực trên UAV.

Phương pháp phát hiện một giai đoạn

Để cải thiện tốc độ xử lý, các phương pháp phát hiện một giai đoạn (one-stage detectors) đã được nghiên cứu mạnh mẽ, tiêu biểu là họ YOLO (You Only Look Once), SSD (Single Shot MultiBox Detector) và RetinaNet [8, 56]. Khác với các phương pháp hai giai đoạn, các mô hình này thực hiện đồng thời phát hiện và phân loại đối tượng chỉ trong một lần suy luận duy nhất, từ đó giảm đáng kể chi phí tính toán và tăng tốc độ xử lý thời gian thực.

Trong các mô hình trên, YOLO là kiến trúc được sử dụng phổ biến nhất trong lĩnh vực giám sát lưới điện nhờ khả năng cân bằng tốt giữa độ chính xác và tốc độ suy luận. Phiên bản YOLO đầu tiên được Redmon và cộng sự đề xuất với ý tưởng chuyển bài toán phát hiện đối tượng thành bài toán hồi quy trực tiếp bounding box và xác suất lớp [8]. Mặc dù đạt tốc độ xử lý cao khoảng 45 FPS, YOLOv1 chỉ đạt mAP khoảng 63.4% trên VOC2007 và gặp hạn chế lớn trong phát hiện vật thể nhỏ do mỗi ô lưới chỉ dự đoán số lượng đối tượng hạn chế.

Để khắc phục các hạn chế này, YOLOv3 được cải tiến với backbone Darknet-53 kết hợp Feature Pyramid Network (FPN), giúp tăng khả năng học đặc trưng đa tỉ lệ. Liu và cộng sự [64] đã đề xuất mô hình Improved YOLOv3 cho bài toán phát hiện chuỗi cách điện trong ảnh UAV có nền phức tạp. Để tăng hiệu quả phát hiện vật thể nhỏ, tác giả cải tiến YOLOv3 bằng: Dense Connection, Feature Pyramid Fusion, tối ưu anchor box, tăng cường truyền đặc trưng đa tầng. Kết quả thực nghiệm cho thấy mô hình đạt: mAP đạt 93.81%, Recall đạt 90.2%, Precision đạt 92.4%. Đồng thời, nghiên cứu chỉ ra rằng cơ chế feature fusion đa tỉ lệ giúp cải thiện đáng kể khả năng phát hiện chuỗi cách điện kích thước nhỏ trong ảnh UAV có nền nhiễu phức tạp như cây cối hoặc dây dẫn chồng lấn [64].

Sau YOLOv3, Bochkovski và cộng sự đã đề xuất YOLOv4 nhằm tối ưu đồng thời tốc độ và độ chính xác phát hiện đối tượng [65]. YOLOv4 tích hợp nhiều kỹ thuật hiện đại như:

- CSPDarknet giúp giảm redundancy và FLOPs,
- PANet ,
- Mosaic augmentation .

Kết quả trên tập COCO cho thấy YOLOv4 đạt: $AP = 43.5\%$, $AP50 = 65.7\%$, tốc độ khoảng 65 FPS trên Tesla V100.

Công trình của nhóm tác giả Bohdan Bochkovskiy với mô hình YOLOv4 được xem là một trong những bộ phát hiện một giai đoạn đạt sự cân bằng tốt nhất giữa tốc độ suy luận và độ chính xác tại thời điểm công bố [65]. Trong lĩnh vực giám sát lưới điện, YOLOv4 đã được ứng dụng hiệu quả để phát hiện nhiều thành phần quan trọng như chuỗi cách điện, dây dẫn, khóa néo và spacer trong ảnh UAV.

Các nghiên cứu thực nghiệm cho thấy YOLOv4 đặc biệt phù hợp với môi trường UAV nhờ khả năng xử lý tốt vật thể đa tỉ lệ và bối cảnh phức tạp. Cụ thể, nghiên cứu DSMH-YOLOv4 [66] cho bài toán phát hiện cách điện từ UAV cho thấy YOLOv4 có khả năng duy trì độ chính xác cao ngay cả khi đối tượng xuất hiện nhỏ, bị che khuất hoặc nằm trong nền phức tạp như sông, cây cối và kết cấu cột điện cao thế. Đồng thời, kiến trúc mạng kim tự tháp đặc trưng và cơ chế dự đoán đa mức của YOLOv4 giúp cải thiện đáng kể khả năng phát hiện vật thể kích thước nhỏ trong ảnh.

Ngoài ra, nhiều công trình cải tiến dựa trên YOLOv4 đã tiếp tục chứng minh tính hiệu quả của kiến trúc này trong giám sát lưới điện UAV. Ví dụ, các nghiên cứu về YOLOv4-tiny và lightweight YOLOv4 [67] cho thấy mô hình vẫn duy trì tốc độ suy luận thời gian thực trên thiết bị biên (edge devices) trong khi cải thiện khả năng nhận dạng lỗi cách điện nhỏ, nền nhiễu mạnh và điều kiện chiếu sáng phức tạp.

Bên cạnh đó, một số nghiên cứu còn ghi nhận YOLOv4 đạt tốc độ xử lý khoảng 100 FPS trong bài toán phát hiện cách điện từ ảnh UAV, vượt trội so với các mô hình hai giai đoạn như Faster R-CNN nhưng vẫn duy trì độ chính xác cao, cho thấy ưu thế rõ rệt của YOLOv4 trong các hệ thống giám sát lưới điện yêu cầu xử lý thời gian thực.

Tiếp theo, YOLOv5 được phát triển theo hướng tối ưu khả năng triển khai thực tế với cấu trúc nhẹ hơn, hỗ trợ huấn luyện linh hoạt và tốc độ suy luận cao [68]. Các nghiên cứu ứng dụng YOLOv5 trong giám sát lưới điện cho thấy mô hình có khả năng phát hiện hiệu quả: chuỗi cách điện, khóa néo, phụ kiện truyền tải điện, tạ chống rung, spacer.

YOLOv5 tích hợp:

- auto-anchor,
- adaptive image scaling,
- mosaic augmentation,
- lightweight CSP backbone.

Nhờ đó, mô hình đạt: mAP@0.5 khoảng 94–95%, tốc độ suy luận khoảng 40–50 FPS, giảm đáng kể số lượng tham số so với YOLOv4.

Đặc biệt, khả năng thích nghi với dữ liệu UAV có tỉ lệ đối tượng thay đổi mạnh giúp YOLOv5 được sử dụng rộng rãi trong các hệ thống giám sát lưới điện thời gian thực.

Gần đây, YOLOv8 tiếp tục cải thiện kiến trúc bằng cơ chế anchor-free detection và backbone tối ưu hơn [69]. So với các phiên bản trước, YOLOv8:

- giảm độ phức tạp dự đoán anchor box,
- tăng tốc độ hội tụ,
- cải thiện localization accuracy,
- tăng hiệu quả phát hiện vật thể nhỏ.

Các nghiên cứu đánh giá YOLOv8 trên dữ liệu UAV cho thấy mô hình phát hiện hiệu quả:

- chuỗi cách điện,
- dây dẫn,
- phụ kiện lưới điện,
- dị vật trên đường dây.

Kết quả thực nghiệm đạt: mAP@0.5 khoảng 95–96%, Precision trên 94%, Recall khoảng 93%, tốc độ suy luận khoảng 55 FPS trên RTX3090.

Mặc dù đạt hiệu năng cao, các mô hình YOLO hiện nay vẫn tồn tại một số hạn chế quan trọng trong bài toán giám sát lưới điện. Trước hết, phần lớn backbone CNN truyền thống vẫn sử dụng stride convolution hoặc pooling để giảm kích thước feature map. Quá trình downsampling này làm mất nhiều thông tin không gian quan trọng, đặc biệt đối với các lỗi kích thước rất nhỏ như:

- vết nứt nhỏ,
- mẻ bát sứ,
- cháy sém cục bộ,
- vùng ô nhiễm bề mặt.

Ngoài ra, ảnh UAV trong thực tế thường chứa nền phức tạp như cây cối, dây dẫn chồng lấn hoặc nhiễu thời tiết, khiến các mô hình dễ phát sinh false positive và false negative. Mặc dù các cơ chế attention và feature fusion đa tỉ lệ đã cải thiện phần

nào khả năng nhận diện, việc phát hiện các đối tượng cực nhỏ hoặc bị che khuất vẫn là thách thức lớn đối với các mô hình một giai đoạn hiện nay.

Bên cạnh đó, dữ liệu lỗi trong lĩnh vực điện lực thường mất cân bằng nghiêm trọng khi số lượng ảnh bình thường lớn hơn rất nhiều so với ảnh bất thường. Điều này làm mô hình dễ thiên lệch về lớp phổ biến và giảm recall đối với các lỗi hiếm. Các nghiên cứu gần đây bắt đầu tích hợp:

- Wise-IoU,
- focal loss,
- adaptive feature fusion,
- attention mechanism.

Tuy nhiên, việc phát hiện lỗi kích thước siêu nhỏ trong môi trường UAV thực tế vẫn chưa được giải quyết triệt để. Đây cũng là động lực thúc đẩy các nghiên cứu mới tập trung cải tiến backbone, tối ưu downsampling và tăng cường học đặc trưng đa tỉ lệ cho bài toán phát hiện bất thường trên lưới điện thông minh.

Phương pháp dựa trên Transformer và mô hình kết hợp

Trong những năm gần đây, cùng với sự phát triển mạnh mẽ của học sâu, các mô hình dựa trên Transformer đã được nghiên cứu và ứng dụng rộng rãi trong bài toán phát hiện đối tượng từ ảnh UAV. Khác với CNN truyền thống vốn chủ yếu khai thác đặc trưng cục bộ thông qua phép tích chập, Transformer có khả năng mô hình hóa quan hệ toàn cục giữa các vùng ảnh thông qua cơ chế self-attention. Điều này đặc biệt hữu ích trong các bài toán giám sát lưới điện, nơi đối tượng thường có kích thước nhỏ, hình dạng phức tạp và xuất hiện trong môi trường nền nhiễu cao.

Một trong những công trình tiêu biểu là DETR (DEtection TRansformer) do Carion et al. đề xuất [9]. Mô hình này kết hợp backbone CNN với Transformer encoder–decoder nhằm thực hiện phát hiện đối tượng theo hướng end-to-end mà không cần sử dụng anchor boxes hay Non-Maximum Suppression (NMS). Ưu điểm nổi bật của DETR là khả năng học trực tiếp mối quan hệ toàn cục giữa các đối tượng trong ảnh, giúp giảm hiện tượng dự đoán trùng lặp và cải thiện tính tổng quát hóa. Tuy nhiên, DETR tồn tại nhược điểm là tốc độ hội tụ chậm, yêu cầu tập dữ liệu huấn luyện lớn và hiệu quả chưa thực sự tối ưu đối với các vật thể nhỏ trong ảnh UAV. Điều này là do cơ chế attention toàn cục làm suy giảm khả năng tập trung vào các chi tiết cục bộ có kích thước nhỏ.

Để cải thiện hạn chế của Vision Transformer truyền thống, Liu et al. đã đề xuất Swin Transformer [70]. Mô hình này sử dụng cơ chế shifted windows nhằm tính toán self-attention theo từng vùng cục bộ thay vì toàn ảnh, từ đó giảm đáng kể độ phức tạp tính toán. Đồng thời, kiến trúc phân cấp nhiều tầng của Swin Transformer cho phép mô hình biểu diễn đặc trưng đa tỉ lệ tương tự CNN truyền thống. Các thực nghiệm cho thấy Swin Transformer đạt hiệu quả cao trong các tác vụ phát hiện đối tượng và phân đoạn ảnh, đặc biệt trong các môi trường có nền phức tạp và biến đổi lớn về góc nhìn. Tuy nhiên, nhược điểm của Swin Transformer là chi phí huấn luyện vẫn cao hơn đáng kể so với CNN, đồng thời yêu cầu tài nguyên phần cứng lớn khi triển khai trên hệ thống nhúng hoặc UAV edge devices.

Trong lĩnh vực giám sát lưới điện, nhiều nghiên cứu gần đây đã khai thác Transformer nhằm nâng cao khả năng phát hiện các thiết bị điện và lỗi bất thường từ ảnh UAV. Chẳng hạn, công trình TransEffiDet đã kết hợp EfficientDet với cơ chế attention Transformer để cải thiện khả năng phát hiện chuỗi cách điện trong môi trường nền phức tạp. Kết quả thực nghiệm cho thấy mô hình cải thiện đáng kể độ chính xác đối với các vật thể nhỏ và đối tượng bị che khuất một phần. Tuy nhiên, mô hình này vẫn gặp khó khăn khi xử lý dữ liệu mất cân bằng giữa mẫu lỗi và mẫu bình thường, vốn là đặc trưng phổ biến trong dữ liệu giám sát lưới điện thực tế.

Ngoài ra, một số nghiên cứu khác đã tích hợp Vision Transformer vào kiến trúc YOLO nhằm tăng cường khả năng biểu diễn ngữ cảnh toàn cục. Ví dụ, các biến thể như TPH-YOLOv5 và YOLO-Transformer sử dụng attention module trong backbone hoặc neck để cải thiện khả năng phát hiện vật thể nhỏ trong ảnh UAV. Các nghiên cứu này cho thấy Transformer giúp mô hình tập trung tốt hơn vào vùng mục tiêu trong các bối cảnh phức tạp như rừng cây, đồi núi hoặc hệ thống dây dẫn chằng chịt. Tuy nhiên, việc bổ sung attention toàn cục cũng làm tăng số lượng tham số và giảm tốc độ suy luận, gây khó khăn đối với các ứng dụng yêu cầu xử lý thời gian thực.

Do đó, xu hướng nghiên cứu hiện nay tập trung vào các mô hình lai (hybrid CNN–Transformer), trong đó CNN được sử dụng để trích xuất đặc trưng cục bộ và giảm độ phức tạp tính toán, còn Transformer đảm nhiệm việc học quan hệ ngữ cảnh toàn cục. Hướng tiếp cận này tận dụng ưu điểm của cả hai kiến trúc: CNN có khả năng xử lý hiệu quả các chi tiết biên và cấu trúc cục bộ, trong khi Transformer cải thiện khả năng biểu diễn phụ thuộc dài hạn và ngữ cảnh toàn ảnh. Các mô hình hybrid cho thấy hiệu quả vượt trội trong phát hiện vật thể nhỏ, nhận dạng đối tượng bị che khuất và xử lý nền phức tạp trong dữ liệu UAV.

Mặc dù đạt được nhiều kết quả tích cực, các nghiên cứu hiện nay vẫn còn tồn tại một số hạn chế đáng chú ý. Thứ nhất, phần lớn mô hình Transformer yêu cầu tập dữ liệu quy mô lớn để huấn luyện hiệu quả, trong khi dữ liệu UAV cho giám sát lưới

điện thường khó thu thập và gán nhãn. Thứ hai, khả năng phát hiện các đối tượng cực nhỏ như nứt vỡ cách điện, spacer kích thước nhỏ hoặc dây dẫn ở khoảng cách xa vẫn chưa thực sự tối ưu. Thứ ba, nhiều mô hình có độ phức tạp tính toán cao, khó triển khai trên các hệ thống UAV nhưng có tài nguyên hạn chế. Ngoài ra, bài toán mất cân bằng dữ liệu giữa trạng thái bình thường và bất thường vẫn chưa được xử lý triệt để trong phần lớn các nghiên cứu hiện nay.

Những hạn chế trên cho thấy vẫn còn nhiều khoảng trống nghiên cứu quan trọng trong lĩnh vực phát hiện đối tượng từ ảnh UAV cho giám sát lưới điện thông minh. Các hướng nghiên cứu tiềm năng bao gồm phát triển mô hình lightweight CNN–Transformer, cải thiện khả năng phát hiện vật thể cực nhỏ, tối ưu attention mechanism cho xử lý thời gian thực và xây dựng các bộ dữ liệu UAV quy mô lớn mang tính thực tiễn cao hơn.

Bảng 1.3. So sánh hiệu năng của một số mô hình phát hiện đối tượng sử dụng Transformer và mô hình lai CNN–Transformer trong ảnh UAV

Mô hình	Tập dữ liệu	mAP (%)	Precision (%)	Đặc điểm nổi bật
DETR [9]	COCO	42.0	61.1	Khả năng học quan hệ toàn cục tốt, loại bỏ anchor và NMS, tuy nhiên hội tụ chậm và chưa tối ưu cho vật thể nhỏ.
Swin Transformer [70]	COCO	58.7	67.1	Biểu diễn đa tỉ lệ hiệu quả nhờ shifted window attention, cải thiện mạnh khả năng phát hiện trong nền phức tạp.
TPH-YOLOv5 [71]	VisDrone	39.3	67.4	Tích hợp Transformer Head giúp cải thiện phát hiện vật thể nhỏ trong ảnh UAV đông đối tượng.
TransEffiDet [72]	UAV Insulator Dataset	96.4	95.1	Kết hợp EfficientDet và Transformer attention nhằm cải thiện nhận dạng chuỗi cách điện trong nền nhiễu cao.
YOLO-Transformer Hybrid [73]	UAV Powerline Dataset	97.1	96.2	Kết hợp CNN và Transformer để cân bằng giữa tốc độ suy luận và khả năng học ngữ cảnh toàn cục.

Bảng 1.3 cho thấy các mô hình dựa trên Transformer và hybrid CNN–Transformer đạt hiệu quả cao trong bài toán phát hiện đối tượng từ ảnh UAV. DETR mở ra hướng tiếp cận end-to-end mới cho object detection nhưng còn hạn chế đối với vật thể nhỏ và tốc độ hội tụ. Trong khi đó, Swin Transformer cải thiện đáng kể hiệu năng nhờ cơ chế attention phân cấp và biểu diễn đa tỉ lệ.

Đối với dữ liệu UAV, các mô hình hybrid như TPH-YOLOv5 hay YOLO-Transformer Hybrid cho thấy ưu thế rõ rệt trong việc phát hiện vật thể nhỏ và xử lý

nền phức tạp. Đặc biệt, các mô hình kết hợp CNN và Transformer thường đạt độ chính xác cao hơn so với CNN thuần túy nhờ khả năng khai thác đồng thời đặc trưng cục bộ và ngữ cảnh toàn cục.

Tuy nhiên, có thể nhận thấy rằng khi độ chính xác tăng lên, chi phí tính toán và số lượng tham số của mô hình cũng tăng đáng kể. Đây là thách thức lớn trong triển khai thực tế trên UAV hoặc các hệ thống edge computing có tài nguyên hạn chế. Do đó, việc tối ưu giữa độ chính xác, tốc độ suy luận và kích thước mô hình vẫn là hướng nghiên cứu quan trọng trong các hệ thống giám sát lưới điện thông minh.

Các kết quả trong bảng cũng cho thấy xu hướng chuyển dịch từ các mô hình CNN truyền thống sang các kiến trúc lai CNN–Transformer nhằm tận dụng ưu điểm của cả hai phương pháp. Điều này mở ra tiềm năng lớn cho việc phát triển các mô hình lightweight hybrid có khả năng xử lý thời gian thực nhưng vẫn duy trì độ chính xác cao trong môi trường UAV thực tế.

1.2.3 Các phương pháp sinh dữ liệu tổng hợp nhằm giải quyết tình trạng khan hiếm mẫu bất thường và mất cân bằng dữ liệu

Trong các bài toán phát hiện bất thường (Anomaly Detection), đặc biệt đối với dữ liệu ảnh công nghiệp và ảnh UAV kiểm tra lưới điện, số lượng mẫu bất thường thường rất hạn chế do các sự cố hiếm khi xảy ra, chi phí thu thập dữ liệu cao và yêu cầu chuyên gia trong quá trình gán nhãn. Sự mất cân bằng nghiêm trọng giữa lớp bình thường và lớp bất thường làm cho mô hình học sâu có xu hướng thiên lệch về lớp chiếm ưu thế, dẫn đến suy giảm khả năng phát hiện các lỗi hiếm và giảm khả năng tổng quát hóa trên dữ liệu thực tế [75?]. Vì vậy, sinh dữ liệu tổng hợp (Synthetic Data Generation - SDG) đã trở thành một hướng nghiên cứu quan trọng nhằm bổ sung dữ liệu cho lớp thiểu số, cải thiện phân bố dữ liệu và nâng cao hiệu quả huấn luyện của các mô hình học sâu.

Các phương pháp sinh dữ liệu tổng hợp có thể được chia thành ba nhóm chính gồm: (i) các phương pháp lấy mẫu quá mức dựa trên nội suy (oversampling); (ii) các mô hình sinh dựa trên học sâu (deep generative models); và (iii) các mô hình khuếch tán (Diffusion Models). Mỗi nhóm có những ưu điểm và hạn chế riêng, phù hợp với từng loại dữ liệu và yêu cầu ứng dụng khác nhau.

Nhóm thứ nhất là các phương pháp oversampling truyền thống như SMOTE, Borderline-SMOTE, ADASYN, KMeans-SMOTE và DeepSMOTE. Ý tưởng chung của nhóm phương pháp này là tạo ra các mẫu mới thông qua phép nội suy giữa các điểm dữ liệu lân cận thuộc lớp thiểu số thay vì sao chép trực tiếp các mẫu hiện có. Ưu điểm của các phương pháp này là thuật toán đơn giản, tốc độ xử lý nhanh, dễ triển

khai và không yêu cầu huấn luyện mô hình sinh. Tuy nhiên, do quá trình sinh dữ liệu chủ yếu dựa trên nội suy tuyến tính nên các mẫu mới thường thiếu tính đa dạng, khó mô hình hóa các phân bố dữ liệu phi tuyến và dễ làm mờ ranh giới giữa các lớp khi áp dụng cho dữ liệu ảnh có cấu trúc phức tạp [74, 75].

Nhóm thứ hai là các mô hình sinh dựa trên học sâu như Variational Autoencoder (VAE), Generative Adversarial Network (GAN), Conditional GAN (CGAN), CycleGAN và StyleGAN. Các phương pháp này học trực tiếp phân bố xác suất của dữ liệu để sinh ra các mẫu mới có mức độ chân thực và đa dạng cao hơn so với các phương pháp oversampling truyền thống. Trong lĩnh vực thị giác máy tính, GAN đã được ứng dụng rộng rãi để giải quyết bài toán mất cân bằng dữ liệu trong phân loại ảnh, phát hiện đối tượng và phát hiện bất thường. Tuy nhiên, GAN vẫn tồn tại một số hạn chế như quá trình huấn luyện không ổn định, hiện tượng *mode collapse*, khó hội tụ và yêu cầu số lượng dữ liệu huấn luyện tương đối lớn để đạt được chất lượng ảnh sinh tốt [? ?].

Những năm gần đây, Diffusion Models được xem là thế hệ mới của các mô hình sinh dữ liệu nhờ khả năng mô hình hóa phân bố dữ liệu tốt hơn GAN. Thay vì học trực tiếp ánh xạ từ nhiễu sang dữ liệu, Diffusion Models thực hiện quá trình thêm nhiễu và khử nhiễu nhiều bước để tái tạo mẫu dữ liệu, từ đó tạo ra các ảnh có chất lượng cao, đa dạng và ổn định hơn. Nhiều nghiên cứu cho thấy Diffusion Models vượt trội GAN trong các bài toán sinh dữ liệu ảnh và đặc biệt phù hợp với các bài toán phát hiện bất thường do có khả năng sinh các mẫu lỗi hiếm với cấu trúc gần giống dữ liệu thực. Tuy nhiên, nhược điểm của Diffusion Models là thời gian huấn luyện và sinh dữ liệu dài, yêu cầu tài nguyên tính toán lớn và khó triển khai trong các hệ thống thời gian thực [76, 77].

Xu hướng nghiên cứu hiện nay tập trung vào các mô hình lai (hybrid models) kết hợp ưu điểm của nhiều phương pháp khác nhau, chẳng hạn như kết hợp SMOTE với GAN, Diffusion Models với học biểu diễn đặc trưng hoặc sử dụng các mô hình sinh có điều kiện nhằm bảo toàn đặc trưng hình thái của đối tượng. Các phương pháp này cho thấy khả năng cải thiện đáng kể độ đa dạng dữ liệu, giảm hiện tượng mất cân bằng và nâng cao hiệu quả phát hiện bất thường trong các lĩnh vực như y tế, sản xuất thông minh và giám sát hạ tầng truyền tải điện [75?].

Đối với bài toán phát hiện bất thường trên chuỗi cách điện từ ảnh UAV, phương pháp sinh dữ liệu tổng hợp cần đáp ứng đồng thời ba yêu cầu: (i) duy trì đặc trưng hình thái của các dạng hư hỏng; (ii) tạo được sự đa dạng về điều kiện chiếu sáng, góc quan sát và bối cảnh nền; và (iii) không làm thay đổi phân bố thống kê của dữ liệu gốc. Đây cũng là xu hướng nghiên cứu đang được quan tâm nhằm nâng cao độ chính xác và khả năng tổng quát hóa của các mô hình học sâu trong điều kiện dữ liệu bất

thường còn hạn chế.

1.2.4 Khoảng trống nghiên cứu

Trên cơ sở khảo sát các công trình liên quan trong mục 1.2.1, 1.2.2 và 1.2.3 có thể thấy các phương pháp học sâu đã đạt được những tiến bộ đáng kể trong bài toán phát hiện thiết bị và bất thường trên lưới điện. Tuy nhiên, khi triển khai trong điều kiện thực tế của dữ liệu UAV, vẫn còn tồn tại ba khoảng trống nghiên cứu chính chưa được giải quyết triệt để, là cơ sở để xác định các định hướng của luận án.

Hạn chế 1: Chưa xử lý tốt bài toán vật thể nhỏ (Small Object Detection) trong nền phức tạp

Mặc dù các mô hình hiện đại như YOLOv7/v8, DETR đã cải thiện mAP tổng thể lên trên 90%, hiệu năng đối với vật thể nhỏ như tạ chống rung, khóa néo hay khuyết tật nứt bất cứ vẫn suy giảm nghiêm trọng, thường thấp hơn 15-20% so với vật thể lớn như cột điện [3]. Nguyên nhân cốt lõi đến từ ba khía cạnh: (i) **Suy giảm thông tin trong quá trình trích xuất đặc trưng**: các phép toán downsampling liên tiếp trong backbone làm mất thông tin chi tiết của vật thể chỉ chiếm 0,05%-0,3% diện tích ảnh; (ii) **Cơ chế hợp nhất đặc trưng chưa hiệu quả**: các kiến trúc FPN, PANet truyền thống chưa duy trì được thông tin không gian có độ phân giải cao cho vật thể nhỏ, trong khi các cơ chế attention toàn cục lại có độ phức tạp tính toán cao và dễ bị nhiễu bởi nền cảnh; (iii) **Nhiều nền chi phối**: ảnh UAV có nền chứa cây cối, công trình, bầu trời có kết cấu tương tự thiết bị, khiến mô hình khó phân biệt.

Các nỗ lực gần đây như tích hợp Super-Resolution cho vật thể nhỏ [78, 79] hay cơ chế attention không gian [80] mới chỉ cải thiện một phần, chưa giải quyết đồng thời cả ba vấn đề trên trong một khung mô hình thống nhất, đặc biệt là khi khuyết tật nằm chồng lấn lên dây dẫn hoặc bị che khuất một phần.

Hạn chế 2: Chưa có giải pháp hiệu quả cho bài toán mất cân bằng dữ liệu nghiêm trọng (Severe Data Imbalance)

Trong các bộ dữ liệu UAV thực tế như CPLID, IDID hay các bộ dữ liệu thu thập tại Việt Nam, tỷ lệ mẫu bất thường so với mẫu bình thường thường nhỏ hơn 1:20, với một số dạng lỗi hiếm như vỡ do sét đánh, tỷ lệ này có thể lên tới 1:100 [5]. Hiện tượng mất cân bằng này dẫn đến hai hệ quả: mô hình bị thiên lệch về lớp đa số (majority bias) và hàm mất mát (loss function) bị chi phối bởi các mẫu dễ (easy negatives).

Các giải pháp hiện có như điều chỉnh hàm mất mát (Focal Loss, Class-Balanced Loss) hay tăng cường dữ liệu truyền thống (lật, xoay, Copy-Paste) chỉ can thiệp ở mức bề mặt. Focal Loss giúp giảm trọng số mẫu dễ nhưng không tạo ra được thông tin ngữ nghĩa mới cho lớp hiếm. Trong khi đó, Copy-Paste tạo ra các mẫu giả có biên không

tự nhiên, dễ khiến mô hình học các đặc trưng giả (spurious correlation) thay vì bản chất khuyết tật. Do đó, vẫn thiếu một chiến lược tái cân bằng ở mức phân phối dữ liệu (distribution-level) có khả năng sinh ra các mẫu lỗi hiếm đa dạng và chân thực.

Hạn chế 3: Dữ liệu sinh ra thiếu tính hữu ích và chưa gắn với tri thức miền lưới điện (Lack of Usefulness and Domain Knowledge)

Hướng tiếp cận sinh dữ liệu tổng hợp bằng GAN và gần đây là Diffusion Models đã cho thấy tiềm năng, tuy nhiên các nghiên cứu hiện nay chủ yếu tập trung vào tính chân thực về mặt thị giác (visual realism) thông qua các chỉ số FID, KID mà bỏ qua tính hữu ích (usefulness) cho bài toán hạ nguồn (downstream task).

Ba thiếu sót chính có thể chỉ ra: (i) **Thiếu ràng buộc vật lý**: các mô hình sinh chung (general-purpose) như Stable Diffusion thường sinh ra chuỗi cách điện có số lượng bất không hợp lý, cấu trúc kim loại bị biến dạng, hoặc khuyết tật xuất hiện ở vị trí bất khả thi về mặt vật lý; (ii) **Thiếu kiểm soát có điều kiện chính xác**: việc sinh khuyết tật nhỏ, cục bộ đòi hỏi kiểm soát ở mức pixel bằng mask hoặc edge map, nhưng các phương pháp hiện tại chưa duy trì được sự nhất quán giữa khuyết tật sinh ra và bối cảnh nền UAV phức tạp, dẫn đến domain gap lớn; (iii) **Chưa có cơ chế đánh giá vòng kín (closed-loop)**: dữ liệu sinh ra chưa được đánh giá bằng hiệu năng cải thiện thực tế của mô hình phát hiện (ví dụ: tăng bao nhiêu % AP cho lớp hiếm), mà chỉ dừng lại ở đánh giá định tính.

Do đó, khoảng trống lớn nhất hiện nay là thiếu một khung sinh dữ liệu có định hướng tri thức miền (domain-knowledge-guided generative framework), trong đó tri thức chuyên gia về cấu trúc chuỗi cách điện, các dạng lỗi điển hình và đặc thù ảnh UAV được mã hóa thành các điều kiện (prompt, mask, layout) để dẫn dắt mô hình khuếch tán sinh ra dữ liệu không chỉ đẹp mà còn hữu ích và tuân thủ vật lý.

Tổng hợp lại, ba khoảng trống trên cho thấy sự cần thiết của một nghiên cứu tích hợp, vừa cải tiến kiến trúc phát hiện cho vật thể nhỏ, vừa giải quyết mất cân bằng dữ liệu ở mức phân phối thông qua sinh dữ liệu có kiểm soát và gắn với tri thức miền. Đây chính là cơ sở để luận án đề xuất ba định hướng nghiên cứu tương ứng trong mục 1.3.2

1.3 Các nội dung nghiên cứu và kết quả chính của luận án

1.3.1 Các câu hỏi nghiên cứu

Xuất phát từ ba khoảng trống nghiên cứu đã được xác định trong mục 1.2.4, luận án đặt ra ba câu hỏi nghiên cứu chính (Research Questions - RQs) làm định hướng xuyên suốt cho các chương tiếp theo. Mỗi câu hỏi tương ứng với một thách thức cốt lõi và sẽ được trả lời thông qua một đóng góp khoa học cụ thể của luận án.

RQ1: Làm thế nào để nâng cao hiệu quả phát hiện thiết bị kích thước nhỏ trên lưới truyền tải điện trong ảnh UAV có nền phức tạp và biến thiên quan sát lớn?

Câu hỏi này xuất phát từ Hạn chế 1. Mặc dù các mô hình phát hiện hiện đại đạt mAP cao trên vật thể lớn, hiệu năng suy giảm nghiêm trọng đối với các thiết bị như tạ chống rung, khóa néo, hay các bát cách điện đơn lẻ. Vấn đề cốt lõi nằm ở việc mất thông tin chi tiết sau các tầng downsampling và nhiều nền chi phối. Do đó, RQ1 hướng đến việc nghiên cứu: (i) cơ chế trích xuất và hợp nhất đặc trưng đa tỷ lệ nào có khả năng bảo toàn thông tin không gian có độ phân giải cao, và (ii) cơ chế chú ý không gian nào giúp mô hình tập trung vào vùng thiết bị thay vì nền cảnh. Giả thuyết của luận án là việc kết hợp FPN cải tiến với cơ chế GELAN và attention có định hướng sẽ cải thiện đáng kể khả năng phát hiện vật thể nhỏ. RQ1 sẽ được trả lời trong Chương 2 thông qua mô hình HVE-YOLO đề xuất.

RQ2: Làm thế nào để nhận diện chính xác các bất thường cục bộ, kích thước nhỏ trên chuỗi cách điện khi ranh giới giữa trạng thái bình thường và bất thường là mờ và dễ bị che khuất?

Câu hỏi này nhằm giải quyết phần còn lại của Hạn chế 1 ở mức fine-grained. Khác với phát hiện thiết bị, nhận diện bất thường đòi hỏi phân biệt các chi tiết vi mô như vết nứt, vỡ một phần bát sứ hay ăn mòn kim loại. Cách tiếp cận phát hiện độc lập và phân đoạn độc lập hiện nay chưa khai thác được mối quan hệ tương hỗ giữa hai nhiệm vụ. RQ2 đặt ra yêu cầu xây dựng một khung mô hình tích hợp (unified framework) có khả năng vừa định vị chính xác chuỗi cách điện, vừa phân đoạn đồng thời vùng bất thường ở mức pixel để duy trì thông tin không gian chi tiết. Giả thuyết là việc học đa nhiệm (multi-task learning) giữa phát hiện và phân đoạn với cơ chế chia sẻ đặc trưng có chọn lọc sẽ giúp mô hình học được các đặc trưng phân biệt tinh vi hơn. RQ2 sẽ được trả lời trong Chương 3 thông qua khung mô hình FGS (FPN-GELAN-Segmentation) đề xuất.

RQ3: Làm thế nào để giải quyết bài toán mất cân bằng dữ liệu nghiêm trọng và nâng cao tính hữu ích của dữ liệu sinh tổng hợp bằng cách tích hợp tri thức miền lưới điện vào mô hình khuếch tán?

Câu hỏi này trực tiếp giải quyết Hạn chế 2 và Hạn chế 3. Các chiến lược tái cân bằng hiện có ở mức hàm mất mát hay tăng cường truyền thống không tạo ra được thông tin ngữ nghĩa mới cho lớp hiếm. Trong khi đó, các phương pháp sinh dữ liệu hiện tại tuy tạo ra ảnh chân thực nhưng thiếu ràng buộc vật lý và thiếu kiểm soát chính xác vị trí, hình thái khuyết tật. Do đó, RQ3 tập trung vào: (i) làm thế nào để mã hóa tri thức miền (số lượng bát, cấu trúc không gian, các dạng lỗi điển hình) thành các điều kiện có thể kiểm soát được cho mô hình khuếch tán (diffusion model), và (ii) chiến lược huấn luyện kết hợp dữ liệu thực và dữ liệu sinh nào giúp cải thiện thực

sự hiệu năng phát hiện trên tập kiểm tra thực tế (usefulness) thay vì chỉ cải thiện chỉ số FID. Giả thuyết là một cơ chế prompting có nhận thức về khuyết tật (anomaly-aware prompting) kết hợp với ControlNet được dẫn dắt bởi mặt nạ khuyết tật và mô tả vật lý sẽ sinh ra dữ liệu tuân thủ ràng buộc miền và giúp tăng recall cho lớp hiếm. RQ3 sẽ được trả lời trong Chương 4 thông qua phương pháp AAPID (Attention-aware Adaptive Prompting for Insulator Diagnosis) đề xuất.

Tổng hợp lại, ba câu hỏi nghiên cứu trên tạo thành một chuỗi logic khép kín: RQ1 giải quyết bài toán phát hiện ở mức thiết bị, RQ2 giải quyết ở mức bất thường chi tiết, và RQ3 giải quyết bài toán dữ liệu làm nền tảng cho cả RQ1 và RQ2. Việc trả lời ba câu hỏi này sẽ dẫn đến mục tiêu cuối cùng của luận án là xây dựng một hệ thống giám sát lưới điện thông minh có tính tự động hóa, chính xác và khả năng tổng quát hóa cao trong điều kiện vận hành thực tế.

1.3.2 Các nội dung nghiên cứu của luận án

Qua phân tích các nghiên cứu liên quan trong bảng 1.4 có thể thấy rằng các phương pháp phát hiện bất thường trên lưới truyền tải điện sử dụng UAV và học sâu đã có nhiều bước tiến đáng kể trong những năm gần đây. Các mô hình hai giai đoạn như Faster R-CNN [81] cho độ chính xác tương đối cao nhưng tốc độ xử lý còn hạn chế, khó đáp ứng yêu cầu giám sát thời gian thực.

Sự phát triển của các mô hình phát hiện một giai đoạn, đặc biệt là họ YOLO, đã cải thiện đáng kể hiệu năng phát hiện cả về độ chính xác và tốc độ xử lý. Các nghiên cứu gần đây tích hợp thêm cơ chế attention, mạng lightweight hoặc kỹ thuật super-resolution đã nâng mAP lên trên 95% trên một số bộ dữ liệu chuyên biệt, đồng thời cải thiện khả năng phát hiện đối tượng nhỏ trong nền ảnh phức tạp [7, 78, 80]. Song song, một số nghiên cứu đã bắt đầu quan tâm đến bài toán phát hiện bất thường trên thiết bị điện, tập trung vào các dạng hư hỏng điển hình như nứt vỡ, mất bất cách điện hoặc suy giảm vật liệu bằng các mô hình phát hiện đối tượng và phân đoạn ảnh [91, 92].

Tuy nhiên, khi triển khai trong điều kiện thực tế của dữ liệu UAV giám sát lưới truyền tải điện, các phương pháp hiện có vẫn còn tồn tại bốn hạn chế mang tính hệ thống.

Thứ nhất, thách thức về vật thể nhỏ và nền phức tạp. Dữ liệu UAV có đặc thù là độ phân giải cao nhưng đối tượng quan tâm lại chiếm tỷ lệ rất nhỏ, thường chỉ 0,05%–0,3% diện tích ảnh, cùng với sự biến thiên lớn về góc quan sát, khoảng cách và điều kiện chiếu sáng. Các thiết bị như chuỗi cách điện dễ bị hòa lẫn với nền như cây cối, bầu trời hoặc địa hình. Các kiến trúc FPN, PANet truyền thống suy giảm

Bảng 1.4. So sánh các phương pháp phát hiện, nhận diện bất thường thiết bị trên lưới truyền tải điện sử dụng UAV

Năm	Phương pháp	Bộ dữ liệu	mAP (%)	P (%)	R (%)	FPS	Đối tượng / bất thường phát hiện
2021	Faster R-CNN + CNN Feature Extraction [81]	UAV dataset	88.7	90.2	85.6	8	Nứt vỡ cách điện, bám bẩn, hư hỏng
2021	Faster R-CNN / YOLO Benchmark [82]	STN-PLAD	89.5	91.3	87.4	10–15	Cột điện, chuỗi cách điện, damper, spacer
2023	YOLOv4 / YOLOX Comparison [83]	RSIn dataset	92.7	93.1	90.5	22	Bất thường chuỗi cách điện
2023	YOLO + ResNet-18 Hybrid [84]	UAV dataset	94.1	94.8	92.3	28	Nứt vỡ, gãy chuỗi cách điện
2024	YOLOv4 + Attention + Thermal Imaging [85]	Infrared UAV dataset	95.1	95.6	93.8	62	Quá nhiệt thiết bị, bất thường cách điện
2024	IDD-YOLO (YOLO + GhostNet + Attention) [86]	Drone dataset	93.8	94.2	92.7	45	Bất thường chuỗi cách điện
2024	YOLOv8 Real-time Detector [87]	CPLID + Custom dataset	91.4	92.7	89.9	33	Bất thường cách điện
2024	YOLOv8 + Efficient Local Attention [88]	UAV dataset	96.9	96.4	95.1	74	Bất thường chuỗi cách điện
2025	YOLOv8 + Attention + Super-resolution [89]	UAV dataset	96.6	95.8	95.3	30	Nứt vỡ, thiếu bát cách điện, bám bẩn
2025	YOLOv8m + GAM Attention [90]	Power Grid dataset	95.7	95.2	94.5	38	Vật thể lạ, bất thường cách điện

thông tin chi tiết sau các tầng downsampling, dẫn đến giảm đáng kể khả năng định vị chính xác.

Thứ hai, thiếu khung mô hình tích hợp giữa phát hiện thiết bị và nhận diện bất thường chi tiết. Phần lớn các nghiên cứu hiện nay xử lý hai bài toán này một cách độc lập. Trong khi đó, trong thực tế vận hành, bất thường thường xuất hiện cục bộ với kích thước nhỏ và hình thái đa dạng trên từng bát sứ hoặc phụ kiện kim loại. Việc chỉ phát hiện ở mức hộp giới hạn (bounding box) là chưa đủ, đòi hỏi mô hình phải duy trì thông tin không gian ở mức pixel và học được các đặc trưng tinh vi (fine-grained) để phân biệt ranh giới mờ giữa trạng thái bình thường và bất thường.

Thứ ba, thách thức về mất cân bằng dữ liệu nghiêm trọng. Trong các bộ dữ liệu thực tế, tỷ lệ mẫu bất thường so với mẫu bình thường thường nhỏ hơn 1:20, thậm chí 1:100 đối với các lỗi hiếm. Hiện tượng này khiến hàm mất mát bị chi phối bởi lớp đa số, mô hình bị thiên lệch và suy giảm recall trên lớp hiếm. Các giải pháp như Focal Loss [7] hay tăng cường truyền thống như Copy-Paste [93] chỉ can thiệp ở mức bề mặt, chưa tạo ra được thông tin ngữ nghĩa mới cho phân phối dài đuôi (long-tail distribution).

Thứ tư, dữ liệu sinh tổng hợp thiếu tính hữu ích và chưa gắn với tri thức miền. Các phương pháp sinh dữ liệu dựa trên GAN và Diffusion Models gần đây như ControlNet [94] hay AnomalyDiffusion [92] chủ yếu được đánh giá bằng tính chân thực thị giác (FID, KID) mà chưa đánh giá tính hữu ích cho bài toán hạ nguồn. Hơn nữa, các mô hình sinh chung thường tạo ra các mẫu phi vật lý như số lượng bát không hợp lý hoặc khuyết tật ở vị trí bất khả thi, do thiếu ràng buộc về cấu trúc và hình học đặc thù của thiết bị điện.

Từ các phân tích trên, có thể tổng kết ba khoảng trống nghiên cứu chính làm cơ sở xác định các nội dung nghiên cứu của luận án:

- (i) Hiệu quả phát hiện đối tượng kích thước nhỏ trong ảnh UAV có nền phức tạp vẫn còn hạn chế, do suy giảm thông tin trong quá trình trích xuất đặc trưng và thiếu cơ chế chú ý không gian hiệu quả;
- (ii) Dữ liệu bất thường mất cân bằng nghiêm trọng và khan hiếm, trong khi các phương pháp tái cân bằng hiện có chưa giải quyết được ở mức phân phối dữ liệu;
- (iii) Khả năng tổng quát hóa của mô hình đối với các điều kiện vận hành thực tế còn chưa cao, do dữ liệu sinh ra thiếu ràng buộc vật lý và chưa được đánh giá bằng tính hữu ích cho bài toán phát hiện.

Trên cơ sở đó, luận án tập trung vào ba nội dung nghiên cứu chính tương ứng:

Nội dung 1: Nghiên cứu và đề xuất mô hình HVE-YOLO cho phát hiện thiết bị kích thước nhỏ trên lưới truyền tải điện trong ảnh UAV. Nội dung này trả lời RQ1, tập trung cải tiến cơ chế trích xuất đặc trưng đa tỷ lệ và hợp nhất đặc trưng có định hướng để bảo toàn thông tin chi tiết cho vật thể nhỏ.

Nội dung 2: Nghiên cứu và xây dựng khung mô hình tích hợp FGS kết hợp phát hiện và phân đoạn cho nhận diện bất thường chi tiết trên chuỗi cách điện. Nội dung này trả lời RQ2, đề xuất học đa nhiệm giữa phát hiện và phân đoạn với cơ chế chia sẻ đặc trưng có chọn lọc nhằm duy trì thông tin không gian và học đặc trưng fine-grained.

Nội dung 3: Nghiên cứu chiến lược xử lý mất cân bằng dữ liệu thông qua sinh dữ liệu có kiểm soát bằng mô hình khuếch tán có định hướng tri thức miền (AAPID). Nội dung này trả lời RQ3, tập trung mã hóa tri thức miền lưới điện thành các điều kiện kiểm soát (defect mask, prompt, layout) cho mô hình khuếch tán và xây dựng chiến lược huấn luyện kết hợp dữ liệu thực và dữ liệu sinh nhằm nâng cao recall cho lớp hiếm.

Ba nội dung trên tạo thành một quy trình khép kín từ phát hiện thiết bị, nhận diện bất thường chi tiết đến giải quyết bài toán dữ liệu nền tảng, được đánh giá thực nghiệm trên bộ dữ liệu UAV thu thập thực tế nhằm hướng tới xây dựng hệ thống giám sát lưới điện thông minh có tính tự động hóa và khả năng triển khai cao.

1.3.3 Kết quả chính của luận án

Từ các khoảng trống nghiên cứu và mục tiêu đã đặt ra, luận án tập trung phát triển các mô hình học sâu nhằm nâng cao hiệu quả phát hiện thiết bị và nhận diện bất thường trên lưới truyền tải điện từ ảnh UAV. Qua quá trình nghiên cứu và thực nghiệm, luận án đã đạt được một số kết quả chính như sau:

- Xây dựng mô hình cải tiến HVE-YOLO (High-Voltage Power Grid Equipment Recognition) nhằm nâng cao hiệu quả phát hiện các thiết bị trên lưới điện cao thế từ ảnh UAV, đặc biệt cải thiện khả năng nhận dạng các đối tượng kích thước nhỏ như tạ chống rung (damper);
- Phát triển khung mô hình FGS (Feature Pyramid Network – GELAN – Diffusion Segmentation) cho bài toán phát hiện và nhận diện bất thường trên chuỗi cách điện, giúp tăng độ chính xác trong điều kiện nền ảnh phức tạp và nhiễu mạnh;
- Đề xuất phương pháp AAPID (Attention-aware Adaptive Prompting for Insulator Diagnosis) nhằm xử lý hiện tượng mất cân bằng dữ liệu, qua đó nâng

cao khả năng phát hiện các trường hợp bất thường hiếm gặp trong dữ liệu UAV thực tế.

Các phương pháp đề xuất được đánh giá thực nghiệm toàn diện trên bộ dữ liệu UAV thực tế và so sánh với nhiều mô hình tiên tiến. Kết quả thực nghiệm cho thấy các mô hình của luận án đạt hiệu quả cao về độ chính xác, khả năng tổng quát hóa và tính ứng dụng trong bài toán giám sát, kiểm tra thiết bị trên lưới truyền tải điện.

1.4 Kết luận chương 1

Chương 1 đã trình bày cơ sở lý thuyết và bối cảnh nghiên cứu của bài toán giám sát thiết bị trên lưới truyền tải điện cao thế từ ảnh UAV dưới góc nhìn thị giác máy tính. Các đặc trưng của dữ liệu thực tế, bao gồm đối tượng kích thước nhỏ, nền ảnh phức tạp, biến thiên phối cảnh lớn và phân bố lớp mất cân bằng, đã được phân tích nhằm làm rõ tính đặc thù của miền ứng dụng.

Trên cơ sở khảo sát các kiến trúc phát hiện đối tượng hiện đại theo hai hướng tiếp cận một giai đoạn và hai giai đoạn, có thể nhận thấy rằng mặc dù các mô hình học sâu đã đạt được độ chính xác cao trên các bộ dữ liệu chuẩn hóa, việc chuyển giao trực tiếp sang môi trường UAV giám sát lưới điện vẫn tồn tại những hạn chế đáng kể về khả năng tổng quát hóa, tính ổn định và hiệu năng thời gian thực.

Các phân tích lý thuyết và thực tiễn cho thấy ba vấn đề trọng tâm cần được giải quyết: (i) nâng cao khả năng biểu diễn và phát hiện đối tượng kích thước nhỏ trong điều kiện nền phức tạp; (ii) xây dựng cơ chế nhận diện bất thường cục bộ trên chuỗi cách điện với mức độ chi tiết cao; và (iii) xử lý hiệu quả sự mất cân bằng dữ liệu giữa trạng thái bình thường và bất thường.

Từ những vấn đề đặt ra, luận án được định hướng theo ba trục nghiên cứu tương ứng: cải tiến kiến trúc phát hiện cho thiết bị trên lưới truyền tải điện cao thế trong ảnh UAV; phát triển khung mô hình tích hợp cho phát hiện và nhận diện bất thường; và đề xuất chiến lược tăng cường dữ liệu có định hướng nhằm tái cân bằng phân bố huấn luyện. Các nội dung này sẽ được trình bày chi tiết trong các chương tiếp theo của luận án.

CHƯƠNG 2

PHÁT HIỆN THIẾT BỊ TRÊN LƯỚI ĐIỆN CAO THẾ TỪ ẢNH UAV

Công tác đảm bảo hệ thống điện vận hành ổn định, an toàn là điều kiện tiên quyết quan trọng trong việc truyền tải điện bởi điện lưới có ảnh hưởng lớn tới sản xuất và đời sống. Việc kiểm tra, giám sát và bảo trì hệ thống giúp xác định tình trạng thiết bị điện, phát hiện những nguy cơ tiềm ẩn đe dọa tới các thiết bị có trên hệ thống.

Phát hiện thiết bị trong ảnh UAV lưới điện là một bài toán có đặc thù rõ rệt. Thiết bị thường chiếm tỷ lệ diện tích rất nhỏ so với toàn ảnh độ phân giải cao; cấu trúc hình học mảnh và kéo dài; nền cảnh phức tạp gồm cây cối, dây dẫn chằng chịt và biến thiên chiếu sáng mạnh. Những đặc điểm này làm cho các mô hình phát hiện đối tượng hiện đại như Faster R-CNN hay các kiến trúc thuộc họ YOLO khi áp dụng trực tiếp thường suy giảm hiệu năng ở nhóm đối tượng kích thước nhỏ, đặc biệt là trong điều kiện nền ảnh phức tạp cũng như góc nhìn thay đổi.

Trong chương này, dựa trên nền tảng của YOLOv11, nghiên cứu sinh đề xuất mô hình cải tiến là HVE-YOLO chuyên biệt cho bài toán *phát hiện sáu loại thiết bị cốt lõi* trên lưới điện cao thế.

2.1 Các nghiên cứu liên quan về phát hiện thiết bị trên lưới điện

2.1.1 Nghiên cứu phát hiện thiết bị trên lưới điện từ ảnh UAV

Hệ thống lưới điện truyền tải cao thế đóng vai trò huyết mạch trong việc duy trì sự ổn định của an ninh năng lượng quốc gia. Tuy nhiên, do đặc thù vận hành ngoài trời, các thành phần trên lưới thường xuyên chịu tác động tiêu cực từ điều kiện thời tiết khắc nghiệt, dẫn đến nguy cơ hư hỏng và sự cố lây lan. Công tác kiểm tra định kỳ là yêu cầu bắt buộc nhằm đảm bảo an toàn vận hành, tuân thủ các tiêu chuẩn kỹ thuật và tối ưu hóa chi phí bảo trì. Trong những năm gần đây, công nghệ thiết bị bay không người lái (UAV) đã chứng minh được hiệu quả vượt trội trong việc giám sát và thu thập dữ liệu hình ảnh các thành phần trên lưới điện cao thế [95–97].

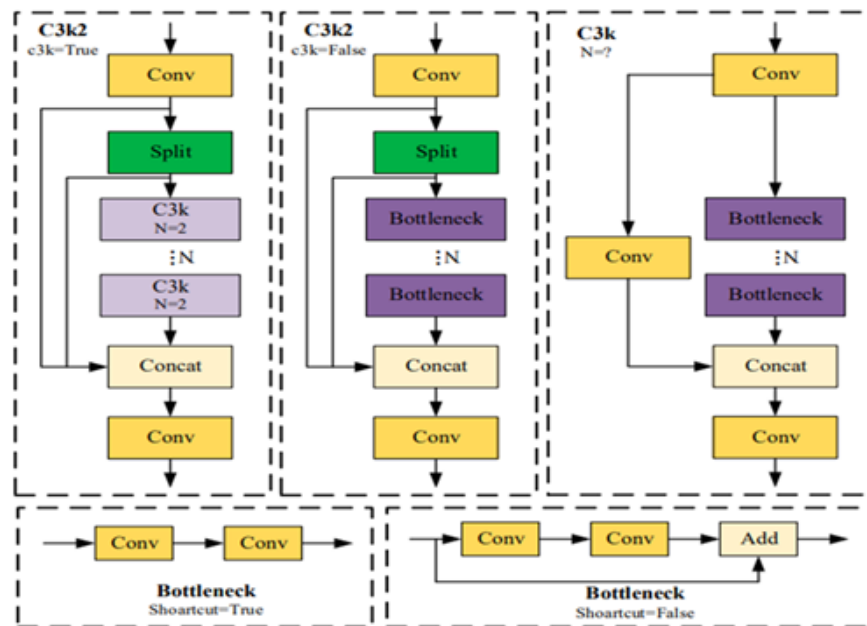
2.1.2 Phát hiện thiết bị trên lưới điện sử dụng mô hình một giai đoạn

Với khả năng phát hiện đối tượng theo thời gian thực, độ chính xác cao và chi phí tính toán tương đối tối ưu, các phiên bản YOLO CT1, CT4 đã được ứng dụng rộng rãi trong nhận dạng chuỗi cách điện, phụ kiện đường dây, điểm phát nhiệt bất thường cũng như

các dạng hư hỏng trên thiết bị điện. Các nghiên cứu gần đây cho thấy YOLO không chỉ cải thiện hiệu quả tự động hóa công tác kiểm tra mà còn góp phần giảm thiểu rủi ro cho nhân viên vận hành và nâng cao độ tin cậy của hệ thống điện.

Kiến trúc backbone Yolo11

Kiến trúc YOLOv11 bao gồm ba thành phần chính: Backbone (trích xuất các đặc trưng đa tỷ lệ), Neck (tổng hợp các đặc trưng từ các tỷ lệ khác nhau) và Head (đưa ra dự đoán về nhãn lớp, điểm chính và hộp giới hạn).



Hình 2.1. Sơ đồ cấu trúc khối C3k2

- **SPPF (Spatial Pyramid Pooling - Fast):** Mô-đun SPPF hỗ trợ trích xuất đặc trưng đa tỷ lệ thông qua cơ chế max-pooling nhiều mức. Nhờ đó, mô hình có khả năng xử lý hiệu quả các đối tượng với kích thước khác nhau trên lưới điện, từ các thiết bị lớn như cột điện, chuỗi cách điện đến các thành phần nhỏ như khóa néo hoặc phụ kiện kết nối. Đặc điểm này đặc biệt quan trọng trong các kịch bản UAV inspection khi khoảng cách và góc chụp thay đổi liên tục.
- **C2PSA (Convolutional Block with Parallel Spatial Attention):** Mô-đun C2PSA tích hợp cơ chế chú ý không gian song song nhằm tăng cường khả năng tập trung của mạng vào các vùng đặc trưng quan trọng. Điều này giúp cải thiện độ chính xác khi phát hiện các thiết bị bị che khuất, có độ tương phản thấp hoặc nằm trong nền cảnh phức tạp như cây cối và địa hình đồi núi. Ngoài ra, việc sử dụng tham số Shortcut=False trong các lớp tích chập giúp hạn chế mất mát đặc trưng cục bộ, trong khi mô-đun ResNet-Attention làm nổi bật các đặc

trung quan trọng để nâng cao khả năng nhận diện đối tượng. Phép nối đặc trưng (Concatenate) giữa các tầng khác nhau cũng góp phần cải thiện khả năng biểu diễn ngữ nghĩa đa mức của mô hình.

- **Khả năng suy luận thời gian thực:** YOLOv11 tiếp tục duy trì ưu điểm nổi bật của họ YOLO là tốc độ suy luận cao, cho phép triển khai trực tiếp trên các hệ thống UAV hoặc edge device phục vụ kiểm tra lưới điện theo thời gian thực. Điều này giúp giảm thời gian kiểm tra thủ công, nâng cao hiệu quả giám sát và hỗ trợ phát hiện sớm các sự cố tiềm ẩn trên đường dây truyền tải.
- **Khả năng học đặc trưng đa tỉ lệ:** Nhờ sự kết hợp giữa Backbone cải tiến và Neck tối ưu, YOLOv11 có khả năng tổng hợp thông tin từ nhiều mức đặc trưng khác nhau. Điều này giúp mô hình duy trì hiệu quả phát hiện trong các tình huống thiết bị xuất hiện ở nhiều khoảng cách và kích thước khác nhau trong ảnh UAV, đồng thời cải thiện độ ổn định khi môi trường quan sát thay đổi mạnh về ánh sáng và góc nhìn.

Hạn chế mô hình

Mặc dù YOLOv11 đạt được những cải tiến đáng kể về độ chính xác và tốc độ suy luận trong các bài toán phát hiện đối tượng thời gian thực, mô hình vẫn bộc lộ một số hạn chế khi áp dụng cho bài toán phát hiện thiết bị trên lưới truyền tải điện cao thế. Một trong những thách thức quan trọng là khả năng nhận diện các đối tượng có kích thước nhỏ như chốt khóa, spacer, damper hoặc các vết nứt nhỏ trên chuỗi cách điện. Theo Khanam và Hussain [98], mặc dù YOLOv11 đã tăng cường khả năng biểu diễn đặc trưng đa tỉ lệ thông qua các mô-đun như C3k2 và C2PSA, quá trình giảm kích thước đặc trưng (downsampling) trong kiến trúc mạng tích chập vẫn có thể làm suy giảm thông tin không gian quan trọng của các mục tiêu nhỏ. Hệ quả là mô hình dễ xảy ra hiện tượng bỏ sót đối tượng hoặc suy giảm độ chính xác phát hiện trong các kịch bản kiểm tra bằng UAV với ảnh đầu vào có độ phân giải lớn và mật độ chi tiết cao.

Bên cạnh đó, ảnh thu nhận từ UAV trong môi trường lưới điện truyền tải thường chứa nền cảnh phức tạp như rừng cây, địa hình đồi núi, khu dân cư hoặc vùng bầu trời có độ tương phản thấp so với thiết bị điện. Theo nghiên cứu của Wang [99], nhiễu nền cùng với sự biến đổi đa tỉ lệ của đối tượng là những yếu tố chính làm suy giảm hiệu quả phát hiện trong các hệ thống kiểm tra đường dây truyền tải sử dụng UAV. Trong nhiều tình huống thực tế, dây dẫn và các phụ kiện điện có cấu trúc mảnh, kích thước nhỏ và màu sắc tương đồng với môi trường xung quanh, khiến mô hình YOLOv11 dễ phát sinh các dự đoán sai (false positive) hoặc định vị đối tượng thiếu chính xác.

Một hạn chế đáng chú ý khác của YOLOv11 là sự phụ thuộc lớn vào tập dữ liệu huấn luyện có chất lượng cao và được gán nhãn đầy đủ. Trong lĩnh vực điện lực, dữ liệu lỗi thực tế thường khan hiếm, tồn tại sự mất cân bằng nghiêm trọng giữa lớp bình thường và lớp lỗi, trong khi chi phí thu thập dữ liệu UAV và gán nhãn chuyên gia tương đối cao. Theo Gao [100], những đặc điểm này có thể dẫn đến hiện tượng quá khớp (overfitting) và làm suy giảm khả năng tổng quát hóa của mô hình khi triển khai trong điều kiện thực tế. Ngoài ra, do YOLOv11 thuộc nhóm mô hình học có giám sát, mô hình gặp khó khăn trong việc nhận diện các dạng lỗi mới hoặc các bất thường chưa từng xuất hiện trong quá trình huấn luyện.

Việc triển khai YOLOv11 trên UAV hoặc các thiết bị edge cũng đặt ra nhiều thách thức về tài nguyên tính toán. Để đảm bảo khả năng phát hiện hiệu quả các đối tượng kích thước nhỏ trên lưới điện, hệ thống thường yêu cầu ảnh đầu vào có độ phân giải cao, chẳng hạn ảnh 4K hoặc kích thước đầu vào lớn. Tuy nhiên, điều này làm gia tăng đáng kể chi phí tính toán, thời gian suy luận cũng như nhu cầu bộ nhớ GPU trong quá trình xử lý [101].

2.2 Mô hình HVE-YOLO (High Voltage Equipment Detection)

Nhằm khắc phục những hạn chế còn tồn tại của các mô hình phát hiện đối tượng truyền thống, đặc biệt là YOLOv11 trong bài toán nhận dạng thiết bị trên lưới truyền tải điện cao thế từ ảnh UAV, nghiên cứu sinh đề xuất mô hình HVE-YOLO (High Voltage Equipment Detection). Mục tiêu chính của mô hình là nâng cao khả năng phát hiện chính xác các thiết bị có kích thước nhỏ, phân bố phức tạp và thường xuất hiện trong môi trường nền nhiễu cao như cây cối, đồi núi, nhà dân hoặc hệ thống dây dẫn chằng chịt.

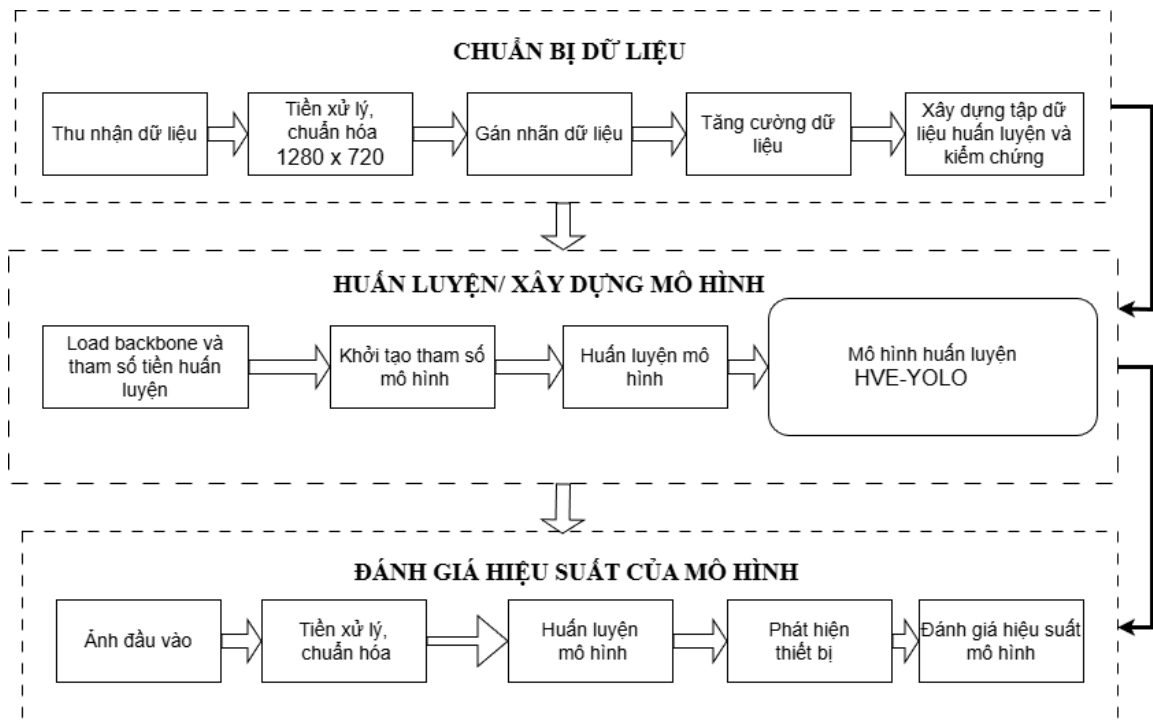
Khác với các bài toán phát hiện đối tượng thông thường, dữ liệu ảnh UAV trong giám sát lưới điện tồn tại nhiều thách thức như thay đổi góc chụp, điều kiện chiếu sáng không ổn định, hiện tượng che khuất một phần và sự mất cân bằng giữa các lớp dữ liệu. Ngoài ra, nhiều thiết bị điện có cấu trúc hình học tương đồng hoặc kích thước rất nhỏ khi quan sát từ khoảng cách xa, làm gia tăng độ khó trong quá trình nhận dạng. Do đó, HVE-YOLO được xây dựng theo hướng tối ưu đồng thời khả năng biểu diễn đặc trưng đa tỉ lệ, tăng cường học ngữ cảnh và duy trì tốc độ suy luận phù hợp với các hệ thống giám sát thời gian thực.

2.2.1 Mô hình HVE-YOLO

Mô hình HVE-YOLO [CT6] được xây dựng nhằm nâng cao hiệu quả phát hiện và nhận dạng thiết bị trên lưới truyền tải điện cao thế từ ảnh UAV. Mô hình có khả năng

phát hiện sáu loại thiết bị cơ bản trên lưới điện, bao gồm: cột đơn thân, cột thép hình, dây dẫn điện, tạ chống rung, chuỗi cách điện thủy tinh và chuỗi cách điện silicon. Đây đều là các thành phần quan trọng trong công tác giám sát vận hành và bảo trì lưới điện truyền tải.

Kiến trúc tổng thể của mô hình được xây dựng theo hướng tối ưu hóa khả năng phát hiện đối tượng nhỏ và nâng cao độ bền vững của mô hình trong các điều kiện môi trường phức tạp. Quy trình xây dựng và huấn luyện mô hình HVE-YOLO được mô tả trong Hình 2.2.



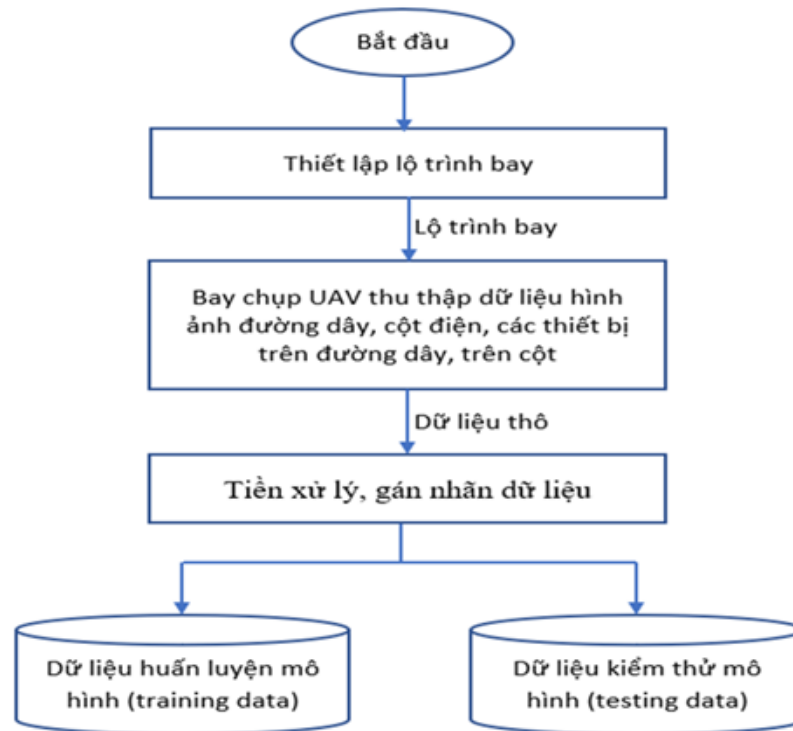
Hình 2.2. Quy trình xây dựng các mô hình nhận dạng thiết bị cho lưới điện cao thế

Toàn bộ quy trình phát hiện thiết bị trên lưới truyền tải điện được chia thành ba giai đoạn chính gồm: (i) chuẩn bị dữ liệu, (ii) xây dựng và huấn luyện mô hình, và (iii) đánh giá mô hình. Các giai đoạn này có mối liên hệ chặt chẽ với nhau nhằm đảm bảo mô hình đạt hiệu quả cao cả về độ chính xác lẫn khả năng tổng quát hóa trong môi trường thực tế.

Chuẩn bị dữ liệu

Giai đoạn chuẩn bị dữ liệu đóng vai trò đặc biệt quan trọng, ảnh hưởng trực tiếp đến khả năng học và hiệu quả phát hiện của mô hình. Dữ liệu được thu thập bằng thiết bị bay không người lái theo quy trình khảo sát thực tế trên lưới điện truyền tải. Việc sử dụng UAV cho phép thu nhận dữ liệu ở nhiều góc nhìn khác nhau, bao quát toàn bộ hành lang tuyến điện và giảm đáng kể chi phí so với phương pháp kiểm tra thủ công hoặc sử dụng trực thăng truyền thống.

Quy trình thu thập dữ liệu được trình bày trong Hình 2.3. Trong quá trình khảo sát, UAV được điều khiển bay dọc theo tuyến truyền tải điện nhằm ghi nhận hình ảnh các thiết bị ở nhiều độ cao, khoảng cách và điều kiện môi trường khác nhau. Điều này giúp tập dữ liệu phản ánh sát hơn các tình huống thực tế trong vận hành lưới điện.



Hình 2.3. Quy trình thu thập dữ liệu trên lưới điện cao thế bằng thiết bị bay không người lái

Sau khi thu thập, dữ liệu ảnh được tiến hành tiền xử lý nhằm chuẩn hóa chất lượng và tăng tính đồng nhất cho tập dữ liệu đầu vào. Các ảnh được chuẩn hóa về kích thước 1280×720 để phù hợp với kiến trúc huấn luyện của mô hình. Đồng thời, các ảnh bị mờ, nhiễu mạnh hoặc chứa quá ít thông tin hữu ích được loại bỏ nhằm giảm ảnh hưởng tiêu cực đến quá trình học.

Tiếp theo, quá trình gán nhãn dữ liệu được thực hiện thủ công bằng công cụ chuyên dụng nhằm xác định chính xác vị trí và lớp của từng thiết bị điện trong ảnh. Để nâng cao tính đa dạng của dữ liệu và giảm hiện tượng overfitting, nhiều kỹ thuật tăng cường dữ liệu được áp dụng như xoay ảnh, lật ảnh, thay đổi độ sáng, thay đổi độ tương phản, Mosaic augmentation và biến đổi phối cảnh. Các kỹ thuật này giúp mô hình cải thiện khả năng nhận dạng trong điều kiện môi trường phức tạp và tăng khả năng tổng quát hóa đối với dữ liệu thực tế.

Cuối cùng, tập dữ liệu được phân chia thành tập huấn luyện và tập kiểm chứng theo tỉ lệ 80% và 20%. Việc phân chia dữ liệu được thực hiện ngẫu nhiên nhưng vẫn đảm bảo tính cân bằng tương đối giữa các lớp đối tượng nhằm tránh hiện tượng thiên

lệch trong quá trình huấn luyện.

Xây dựng và huấn luyện mô hình

Sau giai đoạn chuẩn bị dữ liệu, mô hình HVE-YOLO được xây dựng dựa trên kiến trúc YOLO kết hợp với các cải tiến nhằm nâng cao khả năng phát hiện thiết bị điện kích thước nhỏ và tăng độ ổn định trong môi trường nền phức tạp. Backbone của mô hình được khởi tạo cùng với các tham số tiền huấn luyện nhằm tận dụng tri thức học được từ các tập dữ liệu lớn trước đó, giúp tăng tốc độ hội tụ và cải thiện hiệu quả nhận dạng.

Trong quá trình huấn luyện, các tham số như learning rate, batch size, optimizer và số epoch được thiết lập phù hợp với đặc điểm dữ liệu UAV. Ngoài ra, các chiến lược tối ưu như learning rate scheduling, early stopping và regularization cũng được áp dụng nhằm giảm hiện tượng overfitting và cải thiện khả năng tổng quát hóa của mô hình.

Mô hình HVE-YOLO được huấn luyện theo hướng tối ưu đồng thời độ chính xác và tốc độ suy luận. Điều này đặc biệt quan trọng trong các ứng dụng giám sát lưới điện thời gian thực, nơi hệ thống cần phát hiện nhanh các thiết bị hoặc bất thường để hỗ trợ công tác vận hành và cảnh báo sự cố. Cấu trúc chi tiết của mô hình cùng các cải tiến đề xuất sẽ được trình bày trong các mục tiếp theo của luận án.

Đánh giá mô hình

Sau khi hoàn tất quá trình huấn luyện, mô hình được đánh giá trên tập dữ liệu kiểm chứng nhằm xác định hiệu quả phát hiện đối tượng trong các điều kiện khác nhau. Các chỉ số đánh giá được sử dụng bao gồm Precision, Recall, mAP và tốc độ suy luận FPS. Ngoài ra, nghiên cứu cũng tiến hành phân tích trực quan kết quả phát hiện nhằm đánh giá khả năng nhận dạng đối tượng nhỏ, đối tượng bị che khuất và các trường hợp nền phức tạp.

Kết quả thử nghiệm cho thấy mô hình HVE-YOLO11 đạt độ chính xác phát hiện cao với giá trị mAP đạt 0,972 trên bộ dữ liệu thử nghiệm. Tuy nhiên, cần lưu ý rằng kết quả này được đánh giá trên tập dữ liệu được thu nhận trong điều kiện khoảng cách bay và độ phân giải ảnh phù hợp với yêu cầu giám sát các tuyến đường dây truyền tải điện cao thế hiện nay. Trong thực tế, khi triển khai trên các hệ thống điện siêu cao thế như 220 kV hoặc 500 kV, điều kiện thu nhận dữ liệu có thể thay đổi đáng kể do yêu cầu bảo đảm khoảng cách an toàn điện từ trường giữa UAV và thiết bị điện.

Cụ thể, đối với các tuyến đường dây 110 kV, UAV thường có thể tiếp cận thiết bị ở khoảng cách khoảng 7–8 m để thu nhận hình ảnh chi tiết. Trong khi đó, đối với lưới điện 500 kV, khoảng cách vận hành an toàn thường được khuyến nghị trong khoảng 18–20 m hoặc lớn hơn tùy theo điều kiện khai thác thực tế. Khoảng cách quan sát tăng lên dẫn đến kích thước biểu diễn của thiết bị và các khuyết tật trên ảnh bị thu nhỏ đáng

kê, làm gia tăng độ khó của bài toán phát hiện đối tượng nhỏ. Đồng thời, ảnh hưởng của rung động UAV, điều kiện thời tiết và dao động quang học cũng trở nên rõ rệt hơn, có thể làm suy giảm chất lượng dữ liệu đầu vào.

Do đó, mặc dù kiến trúc HVE-YOLO11 đã chứng minh được khả năng phát hiện hiệu quả trên bộ dữ liệu hiện tại, chưa thể khẳng định mô hình sẽ duy trì nguyên vẹn mức $mAP = 0,972$ khi áp dụng trực tiếp cho dữ liệu thu nhận ở khoảng cách lớn hơn trên các tuyến điện siêu cao thế. Hiệu năng thực tế sẽ phụ thuộc đáng kể vào chất lượng ảnh đầu vào, độ phân giải không gian của đối tượng và khả năng duy trì các đặc trưng chi tiết của thiết bị trên ảnh.

Tuy nhiên, với việc HVE-YOLO11 được thiết kế nhằm tăng cường khả năng trích xuất đặc trưng đa tỷ lệ và cải thiện hiệu quả phát hiện đối tượng nhỏ, mô hình được kỳ vọng vẫn có khả năng duy trì độ chính xác cao nếu quá trình thu nhận dữ liệu được bảo đảm bằng các thiết bị UAV chuyên dụng có hệ thống chống rung ổn định, camera độ phân giải cao và khả năng zoom quang học lớn. Việc sử dụng các cảm biến hình ảnh chất lượng cao giúp duy trì kích thước biểu diễn của thiết bị trên ảnh ngay cả khi UAV phải hoạt động ở khoảng cách an toàn lớn, từ đó hạn chế sự suy giảm hiệu năng phát hiện.

Trong các nghiên cứu tiếp theo, cần tiến hành xây dựng bộ dữ liệu chuyên biệt trên các tuyến truyền tải điện 220 kV và 500 kV với nhiều khoảng cách quan sát khác nhau để đánh giá định lượng khả năng tổng quát hóa của mô hình. Đồng thời, có thể kết hợp các kỹ thuật siêu phân giải ảnh (Super-Resolution), phát hiện đối tượng siêu nhỏ (Tiny Object Detection) và hợp nhất dữ liệu đa cảm biến nhằm duy trì độ chính xác phát hiện trong các điều kiện giám sát từ xa.

Hiệu quả của mô hình được đánh giá bằng các chỉ số như ma trận nhầm lẫn, độ chính xác, độ thu hồi, điểm F1, đường cong độ tin cậy F1, đường cong độ chính xác-độ tin cậy, đường cong độ chính xác-độ thu hồi, đường cong độ thu hồi-độ tin cậy, AP và mAP.

Độ chính xác (Precision) là thông số thể hiện tỉ lệ các dự đoán là đúng (so với ground truth) so với tổng số các dự đoán của model, được tính bằng công thức cụ thể như sau:

$$Precision = TP / (TP + FP)$$

Độ thu hồi (Recall) là thông số thể hiện tỉ lệ dự đoán đúng trên tổng số ground truth, được tính bằng công thức:

$$Recall = TP / (TP + FN)$$

Trong đó, TP (True Positive) là số thiết bị (đối tượng) được phát hiện đúng trên ảnh; FP (False Positive) là số thiết bị phát hiện sai (đối tượng được phát hiện không phải là thiết bị cần quan tâm); FN là số thiết bị không phát hiện được (bỏ sót).

F1-score là trung bình điều hòa giữa Precision và Recall, cho phép cân bằng hai yếu tố này và được sử dụng rộng rãi trong các bài toán có phân bố lớp không đồng đều.

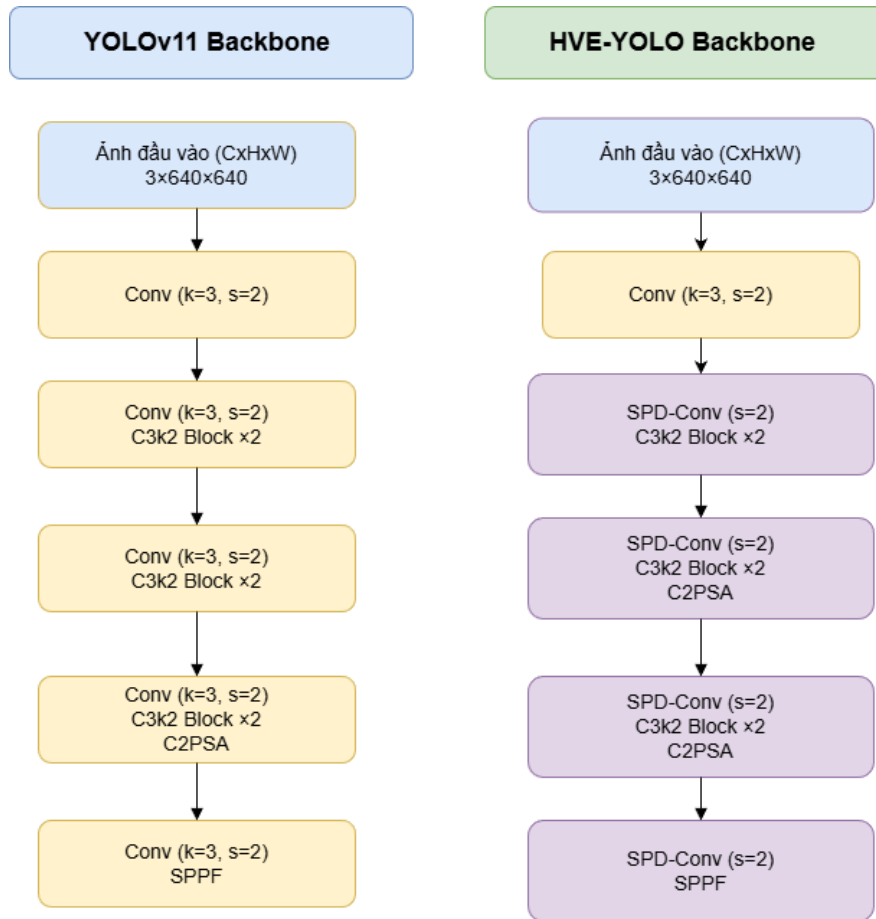
2.2.2 Các cải tiến trong HVE-YOLO

Cải tiến backbone nhằm tăng khả năng phát hiện vật thể nhỏ

Để hạn chế hiện tượng mất mát thông tin không gian trong quá trình giảm mẫu (downsampling), HVE-YOLO đề xuất thay thế một phần phép tích chập stride truyền thống bằng mô-đun SPD-Conv (Space-to-Depth Convolution), như minh họa trong Hình 2.4.

Khác với cơ chế stride convolution chỉ giữ lại một tập con các điểm ảnh theo bước nhảy cố định, SPD-Conv thực hiện tái tổ chức thông tin không gian sang miền kênh trước khi tiến hành tích chập. Cụ thể, quá trình này bao gồm ba bước chính:

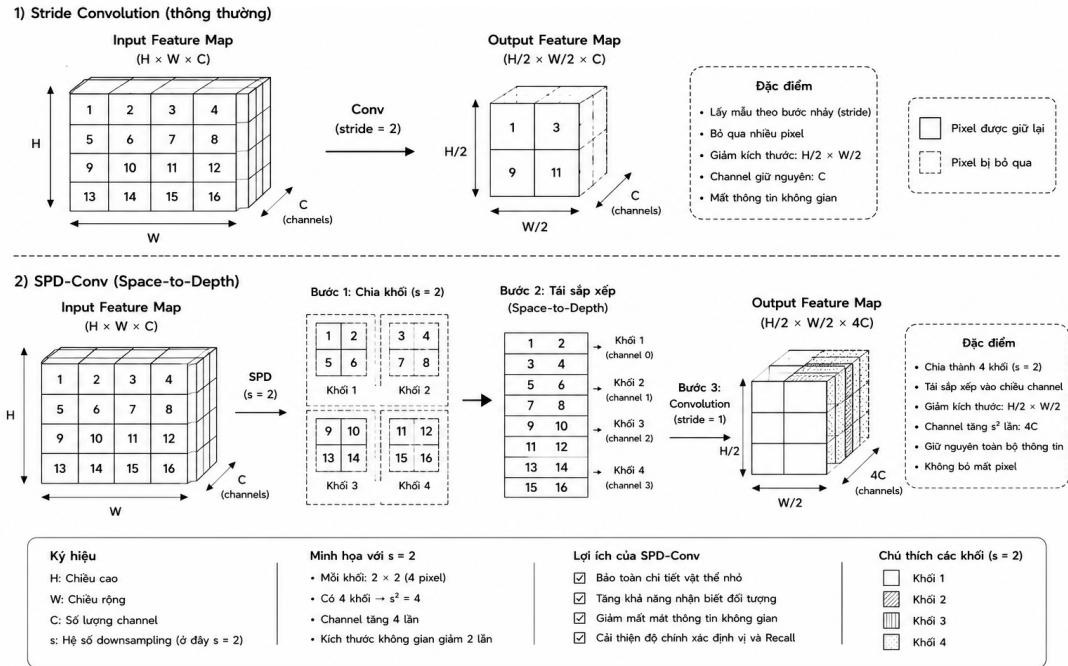
1. Chia feature map đầu vào thành các khối không gian không chồng lấp với kích thước xác định.
2. Tái sắp xếp các giá trị điểm ảnh trong từng khối vào chiều kênh (channel dimension).
3. Thực hiện phép tích chập trên feature map đã được tái cấu trúc.



Hình 2.4. So sánh backbone YOLOv11 và HVE-YOLO

Nhờ cơ chế này, kích thước không gian của bản đồ đặc trưng được giảm xuống nhưng toàn bộ thông tin điểm ảnh ban đầu vẫn được bảo toàn dưới dạng biểu diễn trên các kênh đặc trưng. Điều này khác biệt cơ bản so với stride convolution truyền thống, trong đó một phần thông tin không gian bị loại bỏ trong quá trình lấy mẫu.

Hình 2.5 minh họa sự khác biệt về nguyên lý giữa SPD-Conv và stride convolution.



Hình 2.5. So sánh nguyên lý hoạt động của SPD-Conv và Stride Convolution

Trong phương pháp stride convolution, quá trình giảm kích thước được thực hiện bằng cách lấy mẫu theo bước nhảy, dẫn đến việc loại bỏ nhiều điểm ảnh trung gian. Mặc dù giúp giảm chi phí tính toán và kích thước đặc trưng, cách tiếp cận này có thể gây suy giảm thông tin không gian, đặc biệt đối với các đối tượng nhỏ hoặc có cấu trúc mảnh như dây dẫn, khóa néo và các phụ kiện trên lưới điện truyền tải.

Ngược lại, SPD-Conv thực hiện giảm kích thước theo cơ chế Space-to-Depth, trong đó thông tin không gian được ánh xạ sang chiều kênh thay vì bị loại bỏ. Cụ thể, feature map có kích thước $H \times W \times C$ được chuyển thành $H/2 \times W/2 \times 4C$. Nhờ đó, toàn bộ thông tin ban đầu được bảo toàn trong không gian đặc trưng mới, hạn chế hiện tượng mất mát do lấy mẫu thưa.

Trong bài toán phát hiện thiết bị trên lưới điện từ ảnh UAV, việc bảo toàn thông tin không gian đóng vai trò quan trọng do các đối tượng thường có kích thước nhỏ, phân bố thưa và dễ bị nhiễu bởi nền cảnh phức tạp. Do đó, SPD-Conv giúp tăng cường khả năng biểu diễn đặc trưng của mô hình, từ đó cải thiện hiệu quả phát hiện các đối tượng kích thước nhỏ và nâng cao độ chính xác tổng thể so với phương pháp stride convolution truyền thống.

Bên cạnh đó, SPD-Conv còn cải thiện chất lượng bản đồ đặc tại các tầng nông (shallow layers) của backbone, nơi lưu giữ nhiều thông tin không gian chi tiết. Đây là yếu tố quan trọng đối với các đối tượng nhỏ, vốn phụ thuộc mạnh vào các đặc trưng low-level như biên, kết cấu và hình dạng cục bộ.

Nhờ khả năng bảo toàn thông tin không gian, HVE-YOLO có thể khai thác

hiệu quả các đặc trưng cục bộ, từ đó:

- cải thiện khả năng nhận diện các đối tượng nhỏ bị che khuất một phần,
- nâng cao độ chính xác đối với các thiết bị có độ tương phản thấp so với nền,
- tăng khả năng chống nhiễu trong môi trường nền phức tạp như cây cối, địa hình tự nhiên hoặc khu vực dân cư.

Một ưu điểm quan trọng khác là SPD-Conv không làm gia tăng đáng kể số lượng tham số và độ phức tạp tính toán của mô hình. So với các phương pháp như tăng độ sâu mạng hoặc tăng kích thước ảnh đầu vào, SPD-Conv mang lại cải thiện đáng kể trong phát hiện vật thể nhỏ với chi phí tính toán thấp hơn, đảm bảo cân bằng giữa hiệu năng và hiệu quả.

Do đó, HVE-YOLO vẫn duy trì được các đặc tính quan trọng trong triển khai thực tế, bao gồm: (i) khả năng suy luận thời gian thực, (ii) mức FLOPs hợp lý phù hợp với thiết bị hạn chế tài nguyên, và (iii) khả năng triển khai trên các nền tảng UAV edge AI hoặc GPU nhúng.

CBAM Attention

Nhằm nâng cao khả năng trích xuất và biểu diễn đặc trưng trong môi trường ảnh UAV có nền cảnh phức tạp, luận án tích hợp mô-đun CBAM (Convolutional Block Attention Module) vào kiến trúc backbone của mô hình HVE-YOLO. CBAM là cơ chế attention có cấu trúc gọn nhẹ, cho phép mô hình tăng cường khả năng tập trung vào các đặc trưng quan trọng thông qua việc kết hợp đồng thời attention theo chiều kênh (Channel Attention) và attention theo không gian (Spatial Attention). Nhờ đó, mô hình có thể tự động nhấn mạnh các vùng chứa thông tin hữu ích, đồng thời suy giảm ảnh hưởng của các đặc trưng nền không liên quan trong ảnh đầu vào.

Cấu trúc CBAM bao gồm hai thành phần chính:

- Channel Attention: thực hiện học trọng số trên các kênh đặc trưng nhằm đánh giá mức độ quan trọng của từng kênh trong quá trình biểu diễn thông tin.
- Spatial Attention: thực hiện attention theo miền không gian nhằm xác định các vị trí chứa thông tin quan trọng trên feature map.

Quá trình xử lý của CBAM được thực hiện tuần tự qua hai giai đoạn. Trước hết, mô-đun Channel Attention khai thác thông tin ngữ nghĩa toàn cục thông qua các phép toán Global Average Pooling và Global Max Pooling để xây dựng vector attention theo chiều kênh. Vector attention này được sử dụng để tái hiệu chỉnh trọng số của các

kênh đặc trưng, qua đó tăng cường khả năng biểu diễn đối với các đặc trưng liên quan trực tiếp đến đối tượng mục tiêu.

Tiếp theo, mô-đun Spatial Attention tiếp tục xử lý feature map đã được tinh chỉnh bằng cách xây dựng bản đồ attention theo không gian. Thông qua việc kết hợp thông tin trung bình và cực đại theo chiều kênh, mô-đun này xác định các vùng có khả năng chứa đối tượng và làm nổi bật các vị trí quan trọng trên ảnh đầu vào.

Trong bài toán phát hiện thiết bị điện từ ảnh UAV, dữ liệu đầu vào thường chứa nhiều thành phần nền phức tạp như cây cối, sông nước, công trình xây dựng, kết cấu khung thép hoặc địa hình có độ nhiễu cao. Các yếu tố này làm gia tăng sự tương đồng về đặc trưng giữa đối tượng và nền cảnh, từ đó ảnh hưởng trực tiếp đến khả năng nhận diện của mô hình, đặc biệt đối với các thiết bị điện có kích thước nhỏ hoặc độ tương phản thấp.

Việc tích hợp CBAM vào backbone của HVE-YOLO giúp mô hình tăng cường khả năng tập trung vào các vùng chứa thiết bị điện, đồng thời hạn chế tác động của các đặc trưng nhiễu trong môi trường xung quanh. Nhờ cơ chế attention theo cả chiều kênh và không gian, mô hình có khả năng biểu diễn hiệu quả hơn các đối tượng quan trọng trên lưới điện truyền tải như chuỗi cách điện, tạ chống rung (damper), dây dẫn điện và các phụ kiện kết nối trên đường dây truyền tải.

Ngoài ra, CBAM còn góp phần cải thiện khả năng phát hiện đối tượng trong các trường hợp bị che khuất một phần, thay đổi góc quan sát hoặc chịu ảnh hưởng bởi điều kiện chiếu sáng không đồng đều. Do có cấu trúc tương đối nhẹ, mô-đun này chỉ làm tăng không đáng kể chi phí tính toán nhưng mang lại hiệu quả tích cực trong việc nâng cao độ chính xác phát hiện cũng như khả năng biểu diễn đặc trưng của mô hình HVE-YOLO.

Cải tiến hàm mất mát định vị

Trong bài toán phát hiện thiết bị điện từ ảnh UAV, các đối tượng mục tiêu như chuỗi cách điện, phụ kiện kết nối và dây dẫn điện thường có kích thước nhỏ, hình dạng kéo dài và xuất hiện trong môi trường nền phức tạp. Đối với các đối tượng kích thước nhỏ, chỉ một sai lệch nhỏ trong quá trình hồi quy khung bao cũng có thể làm suy giảm đáng kể giá trị độ giao hội trên hợp (Intersection over Union – IoU), từ đó ảnh hưởng trực tiếp đến độ chính xác phát hiện của mô hình.

Các hàm mất mát dựa trên IoU truyền thống thường gặp hạn chế trong việc xử lý các mẫu khó hoặc các khung bao có chất lượng thấp. Trong nhiều trường hợp, các mẫu ngoại lai có thể tạo ra giá trị gradient lớn, gây mất ổn định trong quá trình huấn luyện, trong khi các mẫu dễ lại chiếm ưu thế trong quá trình tối ưu hóa. Điều này làm

suy giảm khả năng học của mô hình đối với các đối tượng nhỏ hoặc bị che khuất.

Để khắc phục hạn chế trên, luận án sử dụng hàm mất mát Wise-IoU phiên bản WIoU-v3 nhằm thay thế hàm mất mát định vị truyền thống trong mô hình HVE-YOLO. WIoU-v3 sử dụng cơ chế tập trung động phi đơn điệu (*dynamic non-monotonic focusing*), cho phép điều chỉnh linh hoạt mức độ đóng góp của từng mẫu huấn luyện thông qua hệ số khuếch đại gradient. Nhờ đó, mô hình có khả năng tập trung tối ưu vào các mẫu có độ khó trung bình, đồng thời giảm ảnh hưởng của các mẫu quá dễ hoặc quá khó trong quá trình cập nhật trọng số.

Trong mô hình HVE-YOLO, tổng hàm mất mát được xác định như sau:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\text{WIoU-v3}} + \lambda_2 \mathcal{L}_{cls} + \lambda_3 \mathcal{L}_{obj} \quad (2.1)$$

Trong đó:

- $\mathcal{L}_{\text{WIoU-v3}}$ là hàm mất mát định vị khung bao dựa trên Wise-IoU;
- $\mathcal{L}_{cls}$ là hàm mất mát phân loại đối tượng;
- $\mathcal{L}_{obj}$ là hàm mất mát đánh giá khả năng tồn tại của đối tượng;
- λ_1, λ_2 và λ_3 là các hệ số cân bằng nhằm điều chỉnh mức độ đóng góp của từng thành phần hàm mất mát trong quá trình tối ưu hóa.

Trong đó, thành phần hàm mất mát định vị của WIoU-v3 được xác định như sau:

$$\mathcal{L}_{\text{WIoU-v3}} = r \cdot \mathcal{L}_{\text{WIoU-v1}} \quad (2.2)$$

Hệ số tập trung động r được xác định bởi:

$$r = \frac{\beta}{\delta \cdot \alpha^{(\beta-\delta)}} \quad (2.3)$$

với:

$$\beta = \frac{\mathcal{L}_{IoU}}{\overline{\mathcal{L}_{IoU}}} \quad (2.4)$$

Trong các biểu thức trên:

- $\mathcal{L}_{\text{WIoU-v1}}$ là hàm mất mát cơ sở của Wise-IoU;
- $\mathcal{L}_{IoU}$ là giá trị hàm mất mát IoU của mẫu hiện tại;
- $\overline{\mathcal{L}_{IoU}}$ là giá trị trung bình động của hàm mất mát IoU trong quá trình huấn luyện;

- β biểu diễn mức độ ngoại lai của mẫu huấn luyện;
- α và δ là các siêu tham số điều khiển cơ chế tập trung động;
- r là hệ số khuếch đại gradient dùng để điều chỉnh trọng số cập nhật của từng mẫu trong quá trình tối ưu hóa.

Khác với các hàm mất mát IoU truyền thống, WIoU-v3 không tăng đơn điệu trọng số theo độ khó của mẫu mà áp dụng cơ chế tập trung phi đơn điệu (*non-monotonic focusing*). Cơ chế này giúp mô hình hạn chế ảnh hưởng của các mẫu nhiễu hoặc các khung bao có chất lượng quá thấp, đồng thời tập trung tối ưu vào các mẫu có khả năng đóng góp tích cực cho quá trình học.

Mục tiêu tối ưu hóa toàn cục của mô hình được xác định như sau:

$$\Theta^* = \arg \min_{\Theta} \frac{1}{M} \sum_{k=1}^M \mathcal{L}(f_{\Theta}(I_k), \mathcal{G}_k) \quad (2.5)$$

Trong đó:

- Θ là tập tham số của mô hình HVE-YOLO;
- $f_{\Theta}(I_k)$ là kết quả dự đoán của mô hình đối với ảnh đầu vào I_k ;
- $\mathcal{G}_k$ là tập nhãn chuẩn tương ứng;
- M là tổng số mẫu huấn luyện.

Trong quá trình huấn luyện, mô hình thực hiện tối ưu hóa các tham số Θ nhằm cực tiểu hóa tổng hàm mất mát trên toàn bộ tập dữ liệu. Việc sử dụng WIoU-v3 giúp cải thiện đáng kể khả năng hồi quy khung bao đối với các thiết bị điện có kích thước nhỏ hoặc xuất hiện trong điều kiện nền cảnh phức tạp.

Đặc biệt, cơ chế tập trung động của WIoU-v3 giúp HVE-YOLO nâng cao khả năng định vị chính xác trong các trường hợp:

- chuỗi cách điện có kích thước nhỏ;
- phụ kiện kim loại dạng mảnh;
- thiết bị bị che khuất một phần;
- đối tượng xuất hiện dưới góc quan sát phức tạp hoặc trong điều kiện chiếu sáng không đồng đều.

Nhờ đó, mô hình đạt được độ ổn định huấn luyện cao hơn, cải thiện chất lượng hồi quy khung bao và nâng cao hiệu quả phát hiện thiết bị điện trong hệ thống giám sát lưới điện sử dụng UAV.

2.3 Thử nghiệm

Khác với các nghiên cứu đánh giá trên máy chủ hiệu năng cao truyền thống, nghiên cứu này xây dựng môi trường thử nghiệm theo hướng mô phỏng trực tiếp điều kiện triển khai thực tế trên thiết bị bay không người lái (UAV). Các mô hình được triển khai và đánh giá trên nền tảng nhúng NVIDIA Jetson Orin Nano nhằm kiểm chứng khả năng xử lý thời gian thực trong môi trường tính toán giới hạn tài nguyên.

Toàn bộ thuật toán và mô hình được phát triển bằng Python 3 kết hợp với framework PyTorch. Quá trình huấn luyện ban đầu được thực hiện trên máy tính có cấu hình gồm CPU Intel Core i7-9700 với xung nhịp tối đa 4.7 GHz, bộ nhớ RAM 32 GB và hai GPU NVIDIA Tesla K80 12 GB GDDR5 nhằm bảo đảm khả năng xử lý khối lượng dữ liệu lớn trong giai đoạn huấn luyện.

Sau khi huấn luyện, các mô hình được tối ưu hóa và triển khai trên nền tảng NVIDIA Jetson Orin Nano để đánh giá hiệu năng suy luận thực tế trong môi trường edge computing. Để nâng cao tốc độ xử lý và giảm chi phí tính toán, các mô hình được chuyển đổi sang định dạng FP16 và INT8 thông qua công cụ TensorRT với kỹ thuật Post-Training Quantization (PTQ). Việc lượng tử hóa mô hình giúp giảm kích thước bộ nhớ, tối ưu khả năng khai thác tài nguyên phần cứng và cải thiện tốc độ suy luận mà vẫn duy trì độ chính xác ở mức chấp nhận được.

Môi trường thử nghiệm này cho phép đánh giá khách quan khả năng triển khai của mô hình trong các ứng dụng giám sát lưới điện thời gian thực bằng UAV, nơi yêu cầu đồng thời độ chính xác cao, tốc độ xử lý nhanh và mức tiêu thụ tài nguyên thấp.

2.3.1 Dữ liệu thử nghiệm

Dữ liệu thử nghiệm được đánh giá trên tập dữ liệu thu thập từ lưới điện cao áp 110 kV bằng máy bay không người lái DJI Mavic 2 Pro.

Trong chương này, nghiên cứu sinh tập trung vào bài toán nhận dạng các thành phần chính của lưới truyền tải điện cao thế, bao gồm cột điện, dây dẫn và các thiết bị như chuỗi cách điện, tạ chống rung. Các đối tượng này đóng vai trò quan trọng trong vận hành và đảm bảo độ tin cậy của hệ thống điện, do đó việc nhận dạng chính xác chúng từ dữ liệu ảnh UAV là bước nền tảng cho các hệ thống giám sát và phát hiện bất thường tự động.

Về mặt cấu trúc, các đường dây truyền tải điện cao áp thường được thiết kế theo cấu hình ba pha, trong đó mỗi pha được bố trí trên một hoặc nhiều dây dẫn. Trong nhiều tuyến truyền tải thực tế, đặc biệt ở các cấp điện áp từ 110 kV trở lên, các dây dẫn thường được bố trí theo cấu hình 3 hoặc 6 dây dẫn chạy song song dọc theo tuyến đường dây. Cấu hình ba dây tương ứng với trường hợp mỗi pha sử dụng một dây

dẫn đơn, trong khi cấu hình sáu dây thường xuất hiện khi mỗi pha sử dụng bó dây dẫn (bundle conductor) gồm hai dây dẫn song song nhằm tăng khả năng truyền tải công suất và giảm hiện tượng phóng điện vầng quang. Cấu trúc nhiều dây dẫn song song này tạo ra các đặc trưng hình học đặc biệt trong ảnh UAV, đồng thời cũng làm tăng độ phức tạp của bài toán nhận dạng do các dây dẫn thường có kích thước nhỏ, dạng mảnh và kéo dài trong ảnh.

Trong quá trình thu thập dữ liệu, máy bay không người lái được thiết lập ở chế độ bay tự động nhằm đảm bảo quỹ đạo bay ổn định và thu thập dữ liệu một cách hệ thống dọc theo tuyến đường dây. Quy trình bay được thiết kế theo hai giai đoạn chính.

- Ở giai đoạn thứ nhất, UAV bay dọc theo tuyến truyền tải để ghi nhận hình ảnh tổng thể của dây dẫn, các bộ giảm chấn rung (vibration dampers) và các cấu trúc treo dây trên cột điện. Các bộ giảm chấn rung có nhiệm vụ hạn chế dao động của dây dẫn do tác động của gió, do đó việc ghi nhận hình ảnh các thiết bị này giúp bổ sung dữ liệu quan trọng cho bài toán nhận dạng thiết bị trên đường dây.
- Ở giai đoạn thứ hai, khi UAV tiếp cận vị trí từng cột điện, thiết bị bay thực hiện quét ảnh theo quỹ đạo từ trên xuống dưới và từ trái sang phải xung quanh khu vực cột. Cách thức thu thập dữ liệu này giúp ghi nhận đầy đủ hình ảnh của cột điện truyền tải cũng như các thiết bị gắn trên cột, bao gồm chuỗi sứ cách điện, các phụ kiện của chuỗi sứ, kẹp treo và các cấu kiện cơ khí khác. Việc thu thập ảnh theo nhiều góc nhìn khác nhau tại mỗi vị trí cột giúp giảm thiểu khả năng che khuất đối tượng, đồng thời làm phong phú hơn các biến thiên hình học của dữ liệu ảnh.

Khoảng cách từ máy bay không người lái đến đường dây hoặc cột truyền tải được thiết lập trong khoảng 4 m đến 5 m. Khoảng cách này được lựa chọn nhằm đảm bảo cân bằng giữa độ an toàn vận hành của UAV gần hệ thống điện cao thế và khả năng thu nhận hình ảnh có độ phân giải đủ cao để quan sát chi tiết các thiết bị nhỏ như chuỗi sứ cách điện hoặc phụ kiện treo dây. Với khoảng cách bay này, các đối tượng mục tiêu chiếm tỷ lệ điểm ảnh đủ lớn trong khung hình, tạo điều kiện thuận lợi cho quá trình gán nhãn dữ liệu và huấn luyện các mô hình học sâu.

Để đảm bảo tính đa dạng và khả năng tổng quát hóa của bộ dữ liệu, quá trình thu thập ảnh được thực hiện tại nhiều thời điểm khác nhau trong ngày, bao gồm buổi sáng, buổi trưa và buổi chiều. Bên cạnh đó, dữ liệu cũng được ghi nhận trong nhiều điều kiện thời tiết khác nhau, chẳng hạn như trời nhiều mây, mưa phùn, nắng nhẹ hoặc nắng gắt. Ngoài ra, các chuyến bay cũng được tiến hành trong những điều kiện gió khác nhau, từ gió nhẹ đến gió vừa với cấp gió dao động từ mức 1 đến mức 5. Sự đa dạng về điều kiện chiếu sáng và môi trường quan sát giúp bộ dữ liệu phản ánh tốt hơn

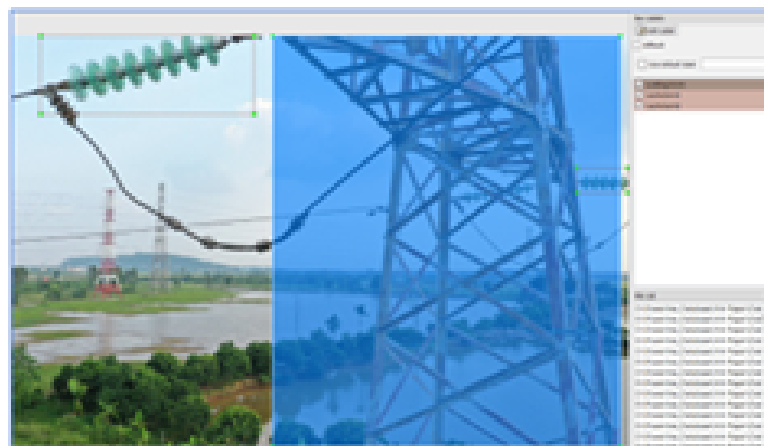
các tình huống thực tế trong quá trình giám sát lưới điện, đồng thời góp phần nâng cao khả năng thích nghi và độ bền vững của các mô hình nhận dạng khi triển khai trong thực tế.

Tổng thể, quy trình thu thập dữ liệu được thiết kế nhằm đảm bảo tính đầy đủ về đối tượng, đa dạng về góc nhìn và phong phú về điều kiện môi trường, qua đó xây dựng một bộ dữ liệu ảnh UAV có chất lượng cao phục vụ cho việc nghiên cứu và đánh giá các phương pháp nhận dạng tự động các thành phần trên lưới truyền tải điện.

Dữ liệu thô sau khi thu thập sẽ được tiền xử lý nhằm loại bỏ các dữ liệu vô nghĩa, dữ liệu trùng lặp, căn chỉnh hình ảnh, loại bỏ các đối tượng gây nhiễu và tăng cường chất lượng ảnh đầu vào. Tiếp theo, dữ liệu được gán nhãn (annotation) theo định dạng *COCO*. Mỗi tệp gán nhãn được lưu dưới dạng một tệp JSON với cấu trúc như sau:

```
{
  "info": info,
  "licenses": [licenses],
  "categories": [category],
  "images": [image],
  "annotations": [annotation]
}
```

Trong đó: *info*: Chứa thông tin về tập dataset; *licenses*: Chứa thông tin về giấy phép của tập dataset; *categories*: Chứa một danh sách tên các đối tượng; *images*: lưu trữ thông tin của các hình ảnh như đường dẫn đến ảnh, tên hình ảnh, ID của ảnh (mỗi ảnh có 1 ID duy nhất); *annotations*: Trường dữ liệu này chứa thông tin về các bounding boxes hoặc các thông tin về segmentation của các đối tượng trong ảnh, Hình 2.6



Hình 2.6. Gán nhãn dữ liệu

Để đảm bảo tính cân bằng giữa các lớp dữ liệu trong bộ dữ liệu huấn luyện, luận án áp dụng kết hợp hai nhóm phương pháp tăng cường dữ liệu. Thứ nhất, các mô hình sinh dữ liệu dựa trên mạng đối kháng sinh [102] (Generative Adversarial Networks – GAN) được sử dụng để tổng hợp thêm các mẫu ảnh cho những lớp có số lượng dữ liệu hạn chế. Thứ hai, các kỹ thuật biến đổi hình ảnh truyền thống [103] (data augmentation) được áp dụng nhằm mở rộng không gian dữ liệu huấn luyện và cải thiện khả năng tổng quát hóa của mô hình.

Bên cạnh phương pháp sinh dữ liệu, luận án cũng áp dụng các kỹ thuật tăng cường dữ liệu dựa trên biến đổi hình học và biến đổi cường độ ảnh, thông qua thư viện Albumentations. Các phép biến đổi này giúp mô phỏng các biến thiên thường gặp trong ảnh UAV như thay đổi góc nhìn, tỉ lệ, điều kiện chiếu sáng hoặc nhiễu cảm biến. Cụ thể, các phép biến đổi được sử dụng bao gồm:

- *RandomScale, Downscale, ShiftScaleRotate*: thay đổi tỷ lệ và thực hiện các phép dịch chuyển, xoay ảnh nhằm mô phỏng sự thay đổi khoảng cách và góc nhìn của UAV đối với đối tượng.
- *HorizontalFlip, VerticalFlip*: lật ảnh theo trục ngang hoặc trục dọc nhằm tăng tính đa dạng về hướng quan sát của đối tượng trong ảnh.
- *GaussNoise, RGBShift, Sharpen*: bổ sung nhiễu Gaussian, thay đổi kênh màu hoặc tăng độ sắc nét nhằm mô phỏng các biến thiên do cảm biến camera và điều kiện môi trường.
- *RandomBrightnessContrast*: điều chỉnh độ sáng và độ tương phản của ảnh để mô phỏng sự thay đổi điều kiện chiếu sáng trong quá trình UAV thu nhận dữ liệu.

Tổng cộng có 51800 hình ảnh, được tổng hợp từ hai nguồn: một bộ sưu tập ban đầu gồm 30500 hình ảnh thực tế được chụp bằng UAV trên lưới truyền tải, và thêm 21300 hình ảnh được tạo ra thông qua các kỹ thuật tăng cường dữ liệu 2.1.

2.3.2 Kịch bản thử nghiệm

Để đánh giá toàn diện hiệu quả của mô hình HVE-YOLO trong bài toán nhận dạng thiết bị trên lưới truyền tải điện từ ảnh UAV, nghiên cứu sinh xây dựng hệ thống kịch bản thử nghiệm theo nhiều góc độ khác nhau, bao gồm: độ chính xác phát hiện, khả năng nhận dạng đối tượng nhỏ, tính ổn định trong điều kiện môi trường phức tạp, hiệu năng xử lý thời gian thực và khả năng triển khai trên nền tảng tính toán nhúng.

Bảng 2.1. Phân bố số lượng mẫu theo từng lớp thiết bị trong bộ dữ liệu UAV

Tên lớp thiết bị	Mẫu huấn luyện	Mẫu kiểm chứng	Tổng cộng
<i>elec-cable</i> (dây dẫn)	16.643	4.851	21.494
<i>slc-insulator</i> (cách điện silicon)	10.316	3.702	14.018
<i>gl-insulator</i> (cách điện thủy tinh)	7.343	1.825	9.168
<i>steel-pole</i> (cột thép)	5.074	1.261	6.335
<i>monopole</i> (cột đơn)	2.935	1.361	4.296
<i>vib-damper</i> (tạ chống rung)	11.473	3.171	14.644
Tổng	53.784	16.171	69.955

Khác với các nghiên cứu chỉ đánh giá trên máy chủ GPU hiệu năng cao, toàn bộ các thí nghiệm trong luận án được triển khai trực tiếp trên nền tảng nhúng **NVIDIA Jetson Orin Nano**, nhằm mô phỏng môi trường tính toán hạn chế thực tế khi mô hình được tích hợp trực tiếp trên UAV phục vụ giám sát lưới điện truyền tải. Việc đánh giá trên nền tảng edge AI cho phép kiểm chứng khả năng triển khai thực tiễn của mô hình trong các hệ thống UAV tự hành có giới hạn về tài nguyên tính toán, bộ nhớ và năng lượng tiêu thụ.

Toàn bộ các mô hình đối chứng và mô hình đề xuất đều được triển khai trong cùng điều kiện phần cứng và phần mềm nhằm đảm bảo tính khách quan trong quá trình đánh giá.

Kịch bản 1: Đánh giá hiệu quả nhận dạng tổng thể

Kịch bản đầu tiên được xây dựng nhằm đánh giá khả năng học đặc trưng tổng quát của mô hình trên bộ dữ liệu UAV thu thập từ hệ thống lưới truyền tải điện cao áp.

Tập dữ liệu được chia theo tỷ lệ: 80% ảnh sử dụng cho huấn luyện; 20% ảnh sử dụng cho kiểm chứng và đánh giá.

Quá trình phân chia dữ liệu được thực hiện ngẫu nhiên nhưng vẫn đảm bảo sự cân bằng tương đối giữa các lớp đối tượng nhằm hạn chế hiện tượng mất cân bằng dữ liệu trong quá trình huấn luyện.

Để đảm bảo tính công bằng trong so sánh, tất cả các mô hình được huấn luyện với cùng: tập dữ liệu; số epoch huấn luyện; kích thước ảnh đầu vào; bộ tối ưu hóa; chiến lược tăng cường dữ liệu.

Trong quá trình huấn luyện và đánh giá, mô hình sử dụng:

- kích thước ảnh đầu vào 640×640 ;
- bộ tối ưu AdamW;
- learning rate khởi tạo 10^{-3} ;
- batch size bằng 16;

- số epoch huấn luyện là 300.

Các chỉ số đánh giá bao gồm: Precision; Recall; mAP@0.5:0.95.

Kịch bản này được sử dụng nhằm đánh giá khả năng nhận dạng tổng thể của mô hình đối với toàn bộ các lớp thiết bị trên lưới truyền tải điện.

Kịch bản 2: Đánh giá khả năng phát hiện đối tượng nhỏ

Trong ảnh UAV, nhiều thiết bị điện như dây dẫn, chuỗi cách điện hoặc tạ chống rung chỉ chiếm một số lượng điểm ảnh rất nhỏ trong khung hình. Đây là nhóm đối tượng khó phát hiện do:

- kích thước nhỏ;
- hình dạng kéo dài;
- dễ hòa lẫn với nền trời hoặc nền thực vật;
- thường xuất hiện ở khoảng cách quan sát xa.

Do đó, nghiên cứu sinh xây dựng một tập kiểm thử chuyên biệt nhằm đánh giá khả năng phát hiện đối tượng nhỏ của mô hình.

Trong kịch bản này:

- các ảnh chứa đối tượng nhỏ được tách riêng thành tập đánh giá độc lập;
- ưu tiên lựa chọn các trường hợp đối tượng bị che khuất một phần hoặc xuất hiện dưới góc nghiêng lớn;
- các đối tượng có diện tích nhỏ hơn 5% diện tích ảnh được xem là đối tượng nhỏ.

Kịch bản này được sử dụng để đánh giá hiệu quả của:

- lớp đặc trưng P2;
- cơ chế attention trong backbone;
- hàm mất mát WIoU-v3;
- chiến lược fusion đặc trưng đa tỉ lệ.

Các chỉ số đánh giá bao gồm:

- Recall đối với đối tượng nhỏ;
- mAP-small;
- độ chính xác định vị bounding box;

- tỷ lệ bỏ sót đối tượng.

Thông qua kịch bản này, luận án đánh giá khả năng cải thiện phát hiện đối tượng nhỏ của HVE-YOLO trong môi trường UAV thực tế.

Kịch bản 3: Đánh giá độ bền vững theo điều kiện môi trường

Để đánh giá khả năng thích nghi của mô hình trong điều kiện vận hành thực tế, nghiên cứu sinh xây dựng tập kiểm thử bao gồm nhiều điều kiện môi trường khác nhau.

Các điều kiện thử nghiệm bao gồm:

- ánh sáng mạnh vào buổi trưa;
- ánh sáng yếu vào sáng sớm hoặc chiều muộn;
- điều kiện trời nhiều mây;
- điều kiện mưa nhẹ hoặc sương mù;
- rung lắc UAV do gió;
- nền ảnh phức tạp gồm cây cối, nhà dân hoặc địa hình đồi núi.

Ngoài ra, tập dữ liệu còn bao gồm các trường hợp:

- hiện tượng mờ chuyển động;
- nhiễu cảm biến;
- thay đổi cân bằng trắng;
- thay đổi độ tương phản;
- thay đổi góc quan sát của UAV.

Mục tiêu của kịch bản này là đánh giá:

- tính ổn định của mô hình;
- khả năng khái quát hóa dữ liệu;
- khả năng chống nhiễu của đặc trưng học sâu;
- độ bền vững khi triển khai trên nền tảng nhúng thực tế.

Các chỉ số đánh giá bao gồm:

- độ suy giảm mAP giữa các điều kiện môi trường;

- độ ổn định Recall;
- độ biến thiên F1-score;
- tỷ lệ suy giảm hiệu năng suy luận trên Jetson Orin Nano.

Kịch bản 4: Đánh giá hiệu năng thời gian thực trên nền tảng nhúng

Kịch bản này được xây dựng nhằm đánh giá khả năng triển khai mô hình trong hệ thống UAV giám sát thời gian thực sử dụng nền tảng NVIDIA Jetson Orin Nano.

Các bài thử nghiệm được thực hiện trực tiếp trên thiết bị nhúng thông qua ba chế độ:

- Intensive Test;
- Parallel Processing Test;
- Stress Test.

Intensive Test được sử dụng nhằm đánh giá tốc độ suy luận cực đại của mô hình khi xử lý liên tục các khung hình đầu vào trong điều kiện tải cao.

Parallel Processing Test đánh giá khả năng xử lý đồng thời nhiều luồng dữ liệu nhằm mô phỏng hệ thống nhiều UAV hoặc nhiều camera hoạt động song song.

Stress Test được thực hiện liên tục trong thời gian dài nhằm đánh giá:

- tính ổn định của FPS;
- mức sử dụng GPU;
- khả năng duy trì hiệu năng;
- hiện tượng suy giảm tốc độ theo thời gian;
- nhiệt độ hệ thống và mức tiêu thụ tài nguyên.

Các chỉ số đánh giá bao gồm:

- FPS trung bình;
- thời gian suy luận trên mỗi ảnh;
- GPU utilization;
- memory consumption;
- độ dao động FPS theo thời gian;

- công suất tiêu thụ năng lượng trên Jetson Orin Nano.

Kịch bản này có vai trò đặc biệt quan trọng trong việc đánh giá khả năng triển khai thực tế của mô hình trên UAV, nơi tồn tại nhiều giới hạn về tài nguyên tính toán và công suất tiêu thụ.

Kịch bản 5: So sánh với các mô hình đối chứng

Để đánh giá hiệu quả của mô hình đề xuất, nghiên cứu sinh tiến hành so sánh trực tiếp giữa HVE-YOLO và các mô hình đối chứng như HVE-Mask R-CNN và các phiên bản YOLO khác.

Các mô hình được huấn luyện và đánh giá:

- trên cùng bộ dữ liệu;
- cùng tập kiểm thử;
- cùng cấu hình Jetson Orin Nano;
- cùng kích thước ảnh đầu vào;
- cùng chiến lược tăng cường dữ liệu.

Các tiêu chí so sánh bao gồm:

- độ chính xác phát hiện;
- khả năng phát hiện đối tượng nhỏ;
- độ chính xác định vị bounding box;
- tốc độ xử lý;
- khả năng suy luận thời gian thực;
- mức sử dụng GPU và bộ nhớ;
- hiệu quả triển khai trên nền tảng edge AI.

Bên cạnh các chỉ số định lượng, nghiên cứu sinh còn thực hiện phân tích định tính thông qua:

- hình ảnh kết quả phát hiện;
- các trường hợp phát hiện sai;
- các trường hợp bỏ sót đối tượng;
- khả năng xử lý nền ảnh phức tạp;

- khả năng phát hiện trong điều kiện rung lắc và ánh sáng không ổn định.

Kịch bản này cho phép đánh giá toàn diện ưu điểm và hạn chế của từng kiến trúc, đồng thời kiểm chứng tính khả thi của mô hình HVE-YOLO khi triển khai trực tiếp trên nền tảng UAV thông minh sử dụng NVIDIA Jetson Orin Nano.

2.3.3 Kết quả thử nghiệm

So sánh kết quả trên mô hình HVE-YOLO và HVE-Mask R-CNN

Dựa trên kết quả thử nghiệm thu được từ bộ dữ liệu ảnh UAV thực tế tại Việt Nam, nghiên cứu sinh thực hiện phân tích đối chứng giữa hai mô hình HVE-YOLO và HVE-Mask R-CNN trên hai phương diện: độ chính xác nhận diện và hiệu năng tính toán.

Phân tích độ chính xác qua các chỉ số đo lường

Kết quả từ Ma trận nhầm lẫn (*Confusion Matrix*) và đường cong F1-Confidence cho thấy HVE-YOLO có ưu thế vượt trội về khả năng nhận diện tổng thể.

Độ chính xác theo lớp: HVE-YOLO đạt độ chính xác cao nhất ở các lớp đối tượng có cấu trúc phức tạp như gl-insulator (97,59%) và vib-damper (97,79%). Đặc biệt, đối với lớp elec-cable (dây dẫn) — vốn là đối tượng khó nhận diện do đặc tính mảnh và dài — HVE-YOLO đạt 93,59%, cao hơn đáng kể so với mức 90,37% của HVE-Mask R-CNN.

Độ ổn định và ngưỡng tin cậy: Đường cong F1-Confidence cho thấy HVE-YOLO đạt giá trị F1 cực đại là 0,95 tại ngưỡng tin cậy 0,647. Trong khi đó, HVE-Mask R-CNN chỉ đạt F1 cực đại 0,851 tại ngưỡng 0,830. Điều này chứng tỏ HVE-YOLO duy trì sự cân bằng tốt hơn giữa độ chính xác (*Precision*) và độ triệu hồi (*Recall*).

Phân tích hiệu năng tính toán trên GPU

Thử nghiệm hiệu năng (GPU Benchmarks) được thực hiện qua ba kịch bản: Kiểm thử cường độ cao (*Intensive Test*), Xử lý song song (*Parallel Processing*) và Kiểm thử áp lực (*Stress Test*).

Tốc độ xử lý: HVE-YOLO thể hiện sự áp đảo về tốc độ với 165,5 FPS trong bài thử nghiệm áp lực, cao hơn gần 70% so với HVE-Mask R-CNN (82,3 FPS). Kết quả này khẳng định kiến trúc một giai đoạn của YOLOv11 được tối ưu hóa cực tốt cho các tác vụ đòi hỏi phản hồi thời gian thực.

Xử lý song song: Đáng chú ý, trong kịch bản xử lý song song nhiều luồng dữ liệu, HVE-Mask R-CNN đạt 98,6 FPS, nhỉnh hơn nhẹ so với 95,6 FPS của HVE-YOLO. Điều này gợi mở rằng kiến trúc hai giai đoạn có khả năng quản lý tài nguyên bộ nhớ GPU hiệu quả hơn khi xử lý các lô dữ liệu (*batches*) đồng thời.

Bảng dưới đây tóm tắt các thông số kỹ thuật và hiệu năng cốt lõi của hai mô hình đề xuất 2.2

Bảng 2.2. Bảng so sánh hiệu năng giữa HVE-YOLO và HVE-Mask R-CNN

Chỉ số so sánh	HVE-YOLO	HVE-Mask R-CNN
Kiến trúc	Một giai đoạn (One-stage)	Hai giai đoạn (Two-stage)
mAP@0.5 (Toàn bộ lớp)	0,972	0,855
F1-Score cực đại	0,950	0,851
Độ triệu hồi (Recall) tổng thể	0,990	0,906
Tốc độ (Intensive Test)	162,9 FPS	75,8 FPS
Tốc độ (Stress Test - 30s)	165,5 FPS	82,3 FPS
Xử lý song song	95,6 FPS	98,6 FPS
Khả năng nhận diện đối tượng nhỏ	Rất tốt (nhờ lớp P2)	Tốt (nhờ FPN)

Kết quả thực nghiệm cho thấy mô hình HVE-YOLO đạt hiệu năng vượt trội so với HVE-Mask R-CNN trên hầu hết các chỉ số đánh giá. Điều này có thể được giải thích bởi một số yếu tố liên quan đến kiến trúc và chiến lược trích xuất đặc trưng của mô hình.

Thứ nhất, kiến trúc YOLOv11 được tối ưu hóa mạnh mẽ cho các tác vụ phát hiện đối tượng thời gian thực. Các khối C3k2 và C2PSA giúp tăng khả năng biểu diễn đặc trưng trong khi vẫn duy trì số lượng tham số ở mức hợp lý. Điều này đặc biệt quan trọng đối với bài toán phát hiện các thiết bị nhỏ trong ảnh UAV.

Thứ hai, việc bổ sung lớp đặc trưng P2 trong mạng kim tự tháp đặc trưng giúp mô hình duy trì độ phân giải cao hơn ở các tầng sâu. Nhờ đó, các đặc trưng không gian của các đối tượng nhỏ như tạ chống rung hoặc chuỗi cách điện được bảo toàn tốt hơn so với các kiến trúc truyền thống.

Ngược lại, mô hình HVE-Mask R-CNN thể hiện ưu thế trong việc phân đoạn chi tiết hình thái đối tượng. Nhánh dự báo mặt nạ cho phép mô hình học được cấu trúc hình học của thiết bị ở mức pixel, từ đó cải thiện độ chính xác định vị trong các trường hợp nền ảnh phức tạp.

Tuy nhiên, chi phí tính toán của kiến trúc hai giai đoạn vẫn cao hơn đáng kể so với các mô hình một giai đoạn. Điều này làm giảm khả năng triển khai trực tiếp trên các hệ thống giám sát thời gian thực sử dụng dữ liệu video UAV.

Tổng hợp các phân tích trên cho thấy HVE-YOLO là lựa chọn phù hợp cho các hệ thống giám sát tự động yêu cầu tốc độ xử lý cao, trong khi HVE-Mask R-CNN phù hợp hơn cho các bài toán phân tích chi tiết cấu trúc thiết bị trong giai đoạn hậu xử lý.

2.4 Kết luận chương 2

Chương 2 đã tập trung nghiên cứu, phát triển và đánh giá hai kiến trúc học sâu tiên tiến là HVE-YOLO và HVE-Mask R-CNN cho bài toán phát hiện và nhận diện các thiết bị trên lưới truyền tải điện 110 kV tại Việt Nam. Các thử nghiệm được thực hiện trên bộ dữ liệu thực tế thu thập bằng UAV trong nhiều điều kiện vận hành khác nhau nhằm đánh giá toàn diện hiệu quả của các mô hình đề xuất.

Kết quả thử nghiệm cho thấy cả hai mô hình đều đạt hiệu năng cao trong bài toán định vị và nhận diện sáu loại thiết bị cốt lõi trên lưới điện. Đặc biệt, mô hình HVE-YOLO đạt giá trị $mAP@0.5$ là 0,972, thể hiện khả năng phát hiện chính xác ngay cả đối với các đối tượng có kích thước nhỏ như tạ chống rung (vib-damper) nhờ cơ chế tăng cường đặc trưng đa tỷ lệ và việc bổ sung lớp đặc trưng P2. Trong khi đó, HVE-Mask R-CNN không chỉ thực hiện tốt nhiệm vụ phát hiện đối tượng mà còn cung cấp thông tin phân đoạn ở mức pixel, tạo điều kiện thuận lợi cho các tác vụ phân tích hình thái và đánh giá chi tiết trạng thái thiết bị.

Về hiệu năng tính toán, HVE-YOLO đạt tốc độ suy luận tối đa 165,5 FPS trên nền tảng GPU, đáp ứng tốt yêu cầu giám sát thời gian thực từ dữ liệu video UAV. HVE-Mask R-CNN mặc dù có tốc độ xử lý thấp hơn nhưng vẫn đảm bảo hiệu quả phát hiện ổn định và mang lại ưu thế trong các bài toán yêu cầu độ chính xác không gian cao. Kết quả thử nghiệm trên bộ dữ liệu thu thập trong điều kiện môi trường đa dạng đã cho thấy khả năng thích nghi và tính tổng quát hóa tốt của các mô hình đối với những thay đổi về ánh sáng, góc quan sát, khoảng cách thu nhận và sự phức tạp của nền ảnh.

Bên cạnh các kết quả đạt được, nghiên cứu vẫn còn một số hạn chế nhất định. Thứ nhất, các thử nghiệm trong chương chủ yếu được thực hiện trên dữ liệu thu nhận từ lưới điện 110 kV; do đó, khả năng duy trì hiệu năng của mô hình khi triển khai trên các hệ thống điện áp cao hơn như 220 kV hoặc 500 kV vẫn cần được kiểm chứng thông qua các bộ dữ liệu thực tế tương ứng. Thứ hai, các mô hình hiện sử dụng chủ yếu dữ liệu ảnh quang học, do đó hiệu quả phát hiện có thể bị ảnh hưởng trong các điều kiện thời tiết bất lợi hoặc đối với các dạng hư hỏng tiềm ẩn chưa biểu hiện rõ trên bề mặt thiết bị. Thứ ba, mặc dù mô hình đã đáp ứng yêu cầu xử lý thời gian thực, các kỹ thuật tối ưu hóa triển khai như lượng tử hóa tham số, cắt tỉa mạng hoặc tối ưu hóa suy luận trên nền tảng phần cứng nhúng vẫn chưa được nghiên cứu một cách hệ thống và là hướng phát triển tiềm năng trong tương lai.

Tóm lại, việc phát hiện và định vị chính xác các thiết bị trên lưới điện là bước nền tảng trong quy trình giám sát tự động bằng UAV. Tuy nhiên, để tiến tới mục tiêu tự động hóa toàn diện công tác quản lý và bảo trì dự phòng, hệ thống cần có khả năng

đánh giá sâu hơn về trạng thái kỹ thuật và mức độ suy giảm của từng thiết bị. Trong số các thiết bị được nhận diện, chuỗi cách điện là thành phần có vai trò đặc biệt quan trọng và thường xuyên phát sinh các dạng hư hỏng trong quá trình vận hành. Vì vậy, các kết quả định vị chuỗi cách điện với độ chính xác trên 96% đạt được trong chương này sẽ là dữ liệu đầu vào quan trọng cho Chương 3, nơi tập trung nghiên cứu các phương pháp học sâu nâng cao nhằm phát hiện và nhận diện các bất thường trên chuỗi cách điện như nứt, vỡ, nhiễm bẩn và các dấu hiệu suy giảm chất lượng khác.

CHƯƠNG 3

PHÁT HIỆN BẤT THƯỜNG TRÊN CHUỖI CÁCH ĐIỆN TỪ ẢNH UAV

Trong Chương 2, luận án đã tập trung nghiên cứu và phát triển các mô hình học sâu phục vụ bài toán phát hiện thiết bị trên lưới điện cao thế từ ảnh UAV. Trên cơ sở phân tích đặc điểm dữ liệu thực tế cũng như yêu cầu triển khai trong hệ thống giám sát lưới điện thông minh, nghiên cứu sinh đã đề xuất mô hình HVE-YOLO nhằm nâng cao khả năng phát hiện và định vị các thiết bị điện trong môi trường ngoài trời phức tạp. Kết quả thực nghiệm cho thấy mô hình đạt hiệu quả phát hiện cao với giá trị $mAP@0.5$ đạt 97%, đồng thời vẫn duy trì được tốc độ xử lý thời gian thực với 165,5 FPS trên GPU. Điều này cho thấy mô hình không chỉ đảm bảo độ chính xác cao mà còn đáp ứng tốt yêu cầu triển khai thực tế trên các hệ thống UAV phục vụ kiểm tra lưới điện.

Những kết quả đạt được ở chương trước đã chứng minh khả năng của các mô hình học sâu trong việc tự động hóa quá trình phát hiện và nhận diện thiết bị trên lưới điện cao thế từ dữ liệu ảnh UAV. Tuy nhiên, phát hiện và định vị thiết bị mới chỉ là bước đầu tiên trong chu trình giám sát tự động. Trong thực tế vận hành, mục tiêu quan trọng hơn của các hệ thống kiểm tra thông minh là đánh giá trạng thái kỹ thuật của thiết bị, phát hiện sớm các dấu hiệu bất thường hoặc hư hỏng tiềm ẩn nhằm hỗ trợ công tác bảo trì dự báo và giảm thiểu nguy cơ sự cố trên lưới điện truyền tải.

Chương này tập trung nghiên cứu bài toán phát hiện bất thường trên chuỗi cách điện từ ảnh UAV thu nhận bằng camera quang học. Nội dung nghiên cứu hướng tới xây dựng mô hình có khả năng học biểu diễn đặc trưng của trạng thái bình thường, từ đó phát hiện hiệu quả các dấu hiệu bất thường trên chuỗi cách điện trong điều kiện dữ liệu hạn chế và môi trường phức tạp. Kết quả nghiên cứu không chỉ góp phần nâng cao hiệu quả giám sát tự động lưới điện truyền tải mà còn tạo tiền đề cho việc phát triển các hệ thống bảo trì thông minh và dự báo sự cố trong tương lai.

3.1 Các nghiên cứu liên quan về phát hiện bất thường trên chuỗi cách điện

3.1.1 Nghiên cứu về bất thường trên chuỗi cách điện

Trong số các thiết bị trên lưới truyền tải điện, chuỗi cách điện là thành phần có tần suất hư hỏng tương đối cao và ảnh hưởng trực tiếp đến độ an toàn vận hành của hệ thống. Các bất thường thường gặp xem hình 3.1 bao gồm:



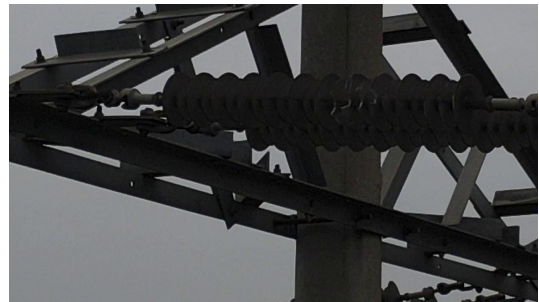
(a) Cách điện thủy tinh mất bát và bám bẩn



(b) Cách điện thủy tinh bám bẩn



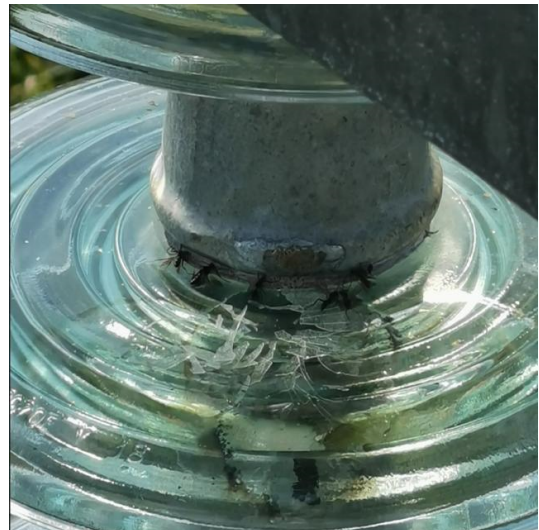
(c) Cách điện silicon rách tán



(d) Cách điện silicon rách tán



(e) Cách điện thủy tinh nứt



(f) Nứt cách điện thủy tinh

Hình 3.1. Một số dạng bất thường trên chuỗi cách điện

- Nứt hoặc vỡ đĩa sứ đối với chuỗi cách điện thủy tinh hoặc sứ, dẫn đến suy giảm khả năng cách điện;
- Hư hỏng hoặc bong tróc lớp vật liệu bọc silicon đối với cách điện polymer;

- Mất hoặc thiếu bất cách điện trong chuỗi cách điện, làm thay đổi phân bố điện trường;
- Nhiễm bẩn bề mặt do bụi công nghiệp, muối biển hoặc ô nhiễm môi trường, có thể dẫn đến hiện tượng phóng điện bề mặt (surface discharge);
- Lệch vị trí hoặc biến dạng cơ học của chuỗi cách điện do tải trọng cơ học hoặc tác động của gió mạnh;
- Hiện tượng cháy xém hoặc dấu vết phóng điện cục bộ trên bề mặt cách điện.

3.1.2 Nghiên cứu phát hiện bất thường trên chuỗi cách điện từ ảnh thu nhận bằng UAV

Backbone YOLOv9

Trong dữ liệu ảnh UAV, các bất thường trên chuỗi cách điện thường có kích thước nhỏ, hình thái đa dạng và dễ bị che khuất bởi nền ảnh phức tạp. Điều này đặt ra yêu cầu cao đối với các mô hình học sâu trong việc trích xuất đặc trưng và phát hiện các dấu hiệu bất thường.

Trong các mô hình phát hiện đối tượng sử dụng YOLO sử dụng mạng Backbone là CSPNet [104] hoặc ELAN [105], cùng với nhiều biến thể khác, để thực hiện các phương pháp tính toán cốt lõi. Trong giai đoạn phát triển ban đầu của dòng YOLO, YOLOv3 đã được cải tiến bằng cách sử dụng các đặc trưng từ FPN [106] hoặc PAN [107] làm đầu vào cho dự đoán. Gần đây, các biến thể YOLO khác đã xuất hiện sử dụng RT DETR [108] dựa trên DETR [109] để trích xuất các đặc trưng đầu vào. Tuy nhiên, một nhược điểm của DETR là yêu cầu dữ liệu đủ tốt và lớn để huấn luyện mô hình nhằm đạt được chất lượng tốt cho mô hình YOLO. YOLOv9 [110] được giới thiệu bằng cách kết hợp YOLOv7 [111] với GELAN [110]. YOLOv9 có thể suy luận từ độ dốc ban đầu đến các lớp mạng rất sâu, do đó giữ lại nhiều thông tin để dự đoán mục tiêu một cách chính xác. GELAN [110] dựa trên mạng khả nghịch để giải quyết vấn đề "thông tin nút thắt cổ chai", điều rất quan trọng trong mạng nơ-ron.

GELAN (Generalized Efficient Layer Aggregation Network) là một kiến trúc mạng nơ-ron tích chập hiện đại được đề xuất nhằm nâng cao hiệu quả trích xuất đặc trưng trong các bài toán thị giác máy tính, đặc biệt là phát hiện đối tượng (object detection). Mô hình GELAN được phát triển dựa trên tư tưởng kế thừa và mở rộng từ các kiến trúc mạng gộp lớp hiệu quả trước đó như ELAN (Efficient Layer Aggregation Network), vốn đã được chứng minh hiệu năng cao trong các mô hình thuộc họ YOLO.

Mục tiêu cốt lõi của GELAN là tăng cường khả năng học biểu diễn đặc trưng của mạng sâu trong khi vẫn giữ được độ ổn định của gradient và hiệu quả tính toán.

Điều này đặc biệt quan trọng trong quá trình huấn luyện các mô hình học sâu có độ sâu lớn, nơi hiện tượng suy giảm gradient và trùng lặp thông tin đặc trưng có thể làm giảm hiệu năng học.

3.2 Mô hình đề xuất FGS (FPN–GELAN–Diffusion Segmentation)

3.2.1 Kiến trúc tổng quan

Mặc dù các vùng quan tâm (Region of Interest – RoI) chứa chuỗi cách điện đã được xác định và tách lọc từ giai đoạn tiền xử lý trong Chương 2, bài toán phát hiện bất thường trên chuỗi cách điện từ ảnh UAV vẫn tồn tại nhiều thách thức khi triển khai trong môi trường thực tế. Nguyên nhân chủ yếu xuất phát từ sự phức tạp của bối cảnh ảnh, sự đa dạng về điều kiện vận hành và đặc điểm hình thái rất nhỏ của các dạng hư hỏng trên thiết bị cách điện.

Trong hệ thống giám sát lưới truyền tải điện bằng UAV, dữ liệu ảnh đầu vào thường chịu ảnh hưởng bởi nhiều yếu tố ngoại cảnh như rung động của thiết bị bay, thay đổi góc quan sát, dao động khoảng cách thu nhận ảnh, điều kiện thời tiết và sự biến thiên cường độ chiếu sáng. Các yếu tố này làm cho chất lượng ảnh đầu vào không ổn định, đồng thời gây ra sự thay đổi đáng kể về phân bố đặc trưng thị giác giữa các mẫu dữ liệu. Điều này dẫn đến khó khăn cho các mô hình học sâu trong việc xây dựng biểu diễn đặc trưng có tính ổn định và khả năng khái quát hóa cao.

Ngoài ra, môi trường nền trong ảnh UAV của hệ thống truyền tải điện thường có mức độ phức tạp lớn. Kết quả khảo sát trên bộ dữ liệu VNelectric cho thấy phần lớn ảnh được ghi nhận trong các khu vực có nhiều thành phần nền như cây xanh, công trình xây dựng, dây dẫn giao cắt hoặc kết cấu bê tông. Các đối tượng này thường tạo ra đặc trưng hình học và kết cấu có tính tương đồng với các vùng cách điện bị nhiễm bẩn hoặc hư hỏng bề mặt. Vì vậy, mô hình rất dễ xuất hiện hiện tượng nhầm lẫn giữa đối tượng cần phát hiện và nhiễu nền, đặc biệt trong các trường hợp độ tương phản ảnh thấp hoặc điều kiện ánh sáng thay đổi mạnh.

Bên cạnh ảnh hưởng của môi trường, các dạng bất thường trên chuỗi cách điện thường có kích thước nhỏ và phân bố không đồng đều. Các lỗi như nứt vi mô, bong tróc lớp phủ silicon hoặc nhiễm bẩn cục bộ chỉ chiếm một vùng diện tích nhỏ trên toàn bộ cấu trúc chuỗi cách điện. Trong các kiến trúc CNN truyền thống, quá trình giảm kích thước đặc trưng liên tục thông qua các lớp pooling hoặc downsampling có thể làm mất các thông tin biên và chi tiết cục bộ quan trọng. Hệ quả là nhiều bất thường giai đoạn sớm không được bảo toàn đầy đủ trong không gian đặc trưng sâu, dẫn đến suy giảm khả năng nhận dạng của mô hình.

Một thách thức khác đến từ sự khác biệt về đặc tính quang học giữa các loại vật

liệu cách điện đang được sử dụng trong hệ thống truyền tải điện tại Việt Nam. Trong thực tế vận hành, cách điện thủy tinh và cách điện gốm được sử dụng đồng thời với tỷ lệ lớn. Trong khi cách điện thủy tinh có khả năng phản xạ ánh sáng mạnh và dễ xuất hiện hiện tượng lóa sáng dưới điều kiện ánh sáng trực tiếp, cách điện gốm lại có xu hướng hấp thụ ánh sáng và tạo ra độ tương phản thấp hơn trong ảnh UAV. Điều này làm tăng mức độ biến thiên của dữ liệu đầu vào và gây khó khăn cho quá trình học đặc trưng tổng quát.

Để giải quyết các vấn đề nêu trên, luận án đề xuất mô hình FGS (FPN–GELAN–Diffusion Segmentation) cho bài toán phát hiện bất thường trên chuỗi cách điện từ ảnh UAV. Kiến trúc đề xuất được xây dựng theo hướng tích hợp nhiều cơ chế học đặc trưng nhằm tăng cường khả năng nhận dạng các bất thường kích thước nhỏ trong điều kiện dữ liệu phức tạp.

Mô hình FGS bao gồm ba thành phần chính là mạng đặc trưng đa tỉ lệ FPN, backbone GELAN và mô hình khuếch tán (Diffusion Model). Đồng thời, hệ thống còn tích hợp nhánh phân đoạn mức điểm ảnh nhằm nâng cao độ chính xác trong việc xác định hình dạng và phạm vi của vùng hư hỏng.

Trong kiến trúc đề xuất, mạng FPN đảm nhiệm vai trò xây dựng biểu diễn đặc trưng đa mức cho ảnh đầu vào. Thông qua cơ chế kết hợp giữa đặc trưng tầng nông và đặc trưng tầng sâu, FPN cho phép mô hình duy trì đồng thời thông tin không gian chi tiết và thông tin ngữ nghĩa mức cao. Điều này đặc biệt quan trọng đối với bài toán phát hiện các bất thường có kích thước nhỏ, nơi mà việc bảo toàn độ phân giải đặc trưng ảnh hưởng trực tiếp đến khả năng nhận dạng.

Backbone GELAN được sử dụng nhằm tối ưu hóa quá trình lan truyền thông tin và tăng cường khả năng học đặc trưng sâu. Khác với các backbone CNN truyền thống, GELAN sử dụng cơ chế tổng hợp đặc trưng đa nhánh nhằm duy trì hiệu quả thông tin gradient giữa các tầng mạng. Nhờ đó, mô hình hạn chế được hiện tượng suy giảm đặc trưng khi chiều sâu mạng tăng lên, đồng thời cải thiện khả năng biểu diễn đối với các cấu trúc bất thường phức tạp.

Một thành phần quan trọng khác của FGS là mô hình khuếch tán được tích hợp trong quá trình tăng cường đặc trưng. Thay vì sử dụng các kỹ thuật khử nhiễu thông thường, mô hình khuếch tán thực hiện học phân bố dữ liệu thông qua quá trình thêm nhiễu và tái tạo lặp. Cơ chế này cho phép mạng học được cấu trúc hình học ổn định của chuỗi cách điện, đồng thời phục hồi các chi tiết bị suy giảm bởi nhiễu môi trường hoặc chất lượng ảnh thấp.

Việc sử dụng Diffusion Model đặc biệt hiệu quả trong các trường hợp ảnh UAV bị ảnh hưởng bởi:

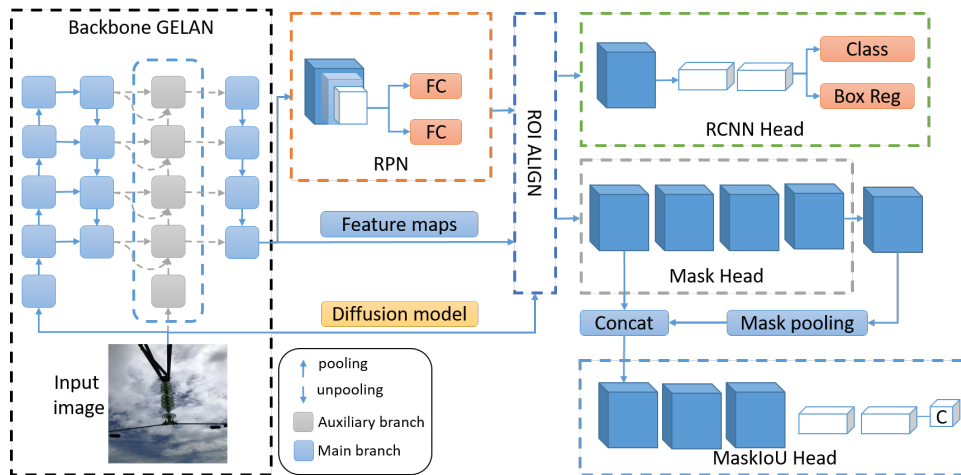
- rung động trong quá trình bay;
- hiện tượng mờ chuyển động;
- thay đổi điều kiện chiếu sáng;
- nhiễu nền từ cây cối hoặc công trình xây dựng.

Thông qua cơ chế tái tạo đặc trưng, mô hình có khả năng duy trì các thông tin hình học quan trọng của chuỗi cách điện, từ đó cải thiện khả năng nhận dạng các bất thường nhỏ hoặc có cấu trúc phức tạp.

Bên cạnh nhiệm vụ phát hiện đối tượng bằng bounding box, mô hình FGS còn mở rộng sang bài toán phân đoạn mức điểm ảnh thông qua việc tích hợp Mask Head và MaskIoU Head. Nhánh phân đoạn cho phép hệ thống không chỉ xác định vị trí bất thường mà còn mô tả chính xác hình dạng của vùng lỗi. Điều này có ý nghĩa thực tiễn quan trọng trong các hệ thống giám sát thiết bị điện, nơi mà mức độ lan rộng của vùng hư hỏng ảnh hưởng trực tiếp đến quá trình đánh giá và lập kế hoạch bảo trì thiết bị.

Mục tiêu tổng quát của mô hình FGS là xây dựng một hệ thống giám sát tự động có khả năng thực hiện đồng thời các nhiệm vụ phát hiện, phân loại và phân đoạn bất thường trên chuỗi cách điện. Đầu vào của hệ thống là các vùng RoI được trích xuất từ giai đoạn trước, trong khi đầu ra bao gồm trạng thái hoạt động của chuỗi cách điện như bình thường (normal), mất bất cách điện (missing) hoặc nhiễm bẩn (polluted).

Phương pháp đề xuất trong luận án kế thừa và phát triển từ các kết quả nghiên cứu đã được công bố trong công trình CT2. Điểm mở rộng trọng tâm của nghiên cứu tập trung vào việc kết hợp mô hình khuếch tán với backbone GELAN và mạng FPN nhằm nâng cao khả năng biểu diễn đặc trưng đối với dữ liệu UAV có mức nhiễu cao và độ biến thiên lớn.



Hình 3.2. Kiến trúc tổng thể của mô hình FGS

Hình 3.2 minh họa cấu trúc tổng thể của hệ thống FGS. Kiến trúc được xây dựng theo hướng tích hợp thống nhất giữa các giai đoạn trích xuất đặc trưng, đề xuất vùng, tăng cường biểu diễn và phân đoạn bất thường.

Trong giai đoạn đầu, backbone GELAN kết hợp với FPN thực hiện trích xuất bản đồ đặc trưng đa tỉ lệ từ ảnh đầu vào. Các đặc trưng này sau đó được chuyển tới mạng đề xuất vùng RPN nhằm xác định các khu vực có khả năng chứa chuỗi cách điện hoặc vùng bất thường tiềm năng.

Song song với quá trình trên, mô hình khuếch tán thực hiện tăng cường và tái tạo đặc trưng nhằm giảm ảnh hưởng của nhiễu nền và phục hồi các cấu trúc chi tiết bị suy giảm. Đầu ra của mô hình khuếch tán được kết hợp với đặc trưng từ RPN thông qua lớp RoI Align để xây dựng biểu diễn RoI có độ phân biệt cao hơn.

Các đặc trưng RoI sau đó được đưa vào RCNN Head để thực hiện phân loại và hồi quy bounding box. Đồng thời, nhánh Mask Head sinh mặt nạ phân đoạn cho từng vùng bất thường, trong khi MaskIoU Head được sử dụng để đánh giá và hiệu chỉnh chất lượng của mặt nạ dự đoán.

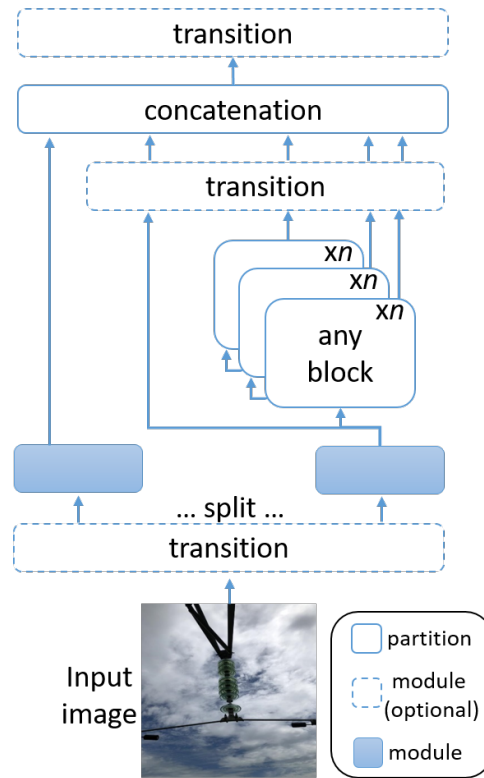
Nhờ sự kết hợp giữa cơ chế trích xuất đặc trưng đa tỉ lệ, tăng cường biểu diễn bằng mô hình khuếch tán và phân đoạn mức điểm ảnh, FGS có khả năng xử lý hiệu quả các bất thường nhỏ trong điều kiện môi trường phức tạp. So với các phương pháp phát hiện truyền thống, mô hình đề xuất không chỉ cải thiện độ chính xác nhận dạng mà còn nâng cao khả năng khái quát hóa và tính ổn định khi triển khai trong thực tế.

3.2.2 Backbone GELAN

Backbone GELAN được sử dụng nhằm tăng cường khả năng học đặc trưng sâu và tối ưu hóa dòng truyền thông tin trong mạng. Kiến trúc này được thiết kế theo hướng cân bằng giữa hiệu năng phát hiện và chi phí tính toán, phù hợp với các hệ thống giám sát thời gian thực sử dụng UAV.

Kiến trúc mô hình GELAN được mô tả trong hình 3.3. Ở giai đoạn đầu, mạng sử dụng các lớp tích chập kết hợp downsampling để xây dựng các bản đồ đặc trưng sơ cấp B1 và B2. Các bản đồ này lưu giữ thông tin biên và cấu trúc hình học cơ bản của chuỗi cách điện. Tiếp theo, các mô-đun RepConv được sử dụng nhằm tăng cường khả năng biểu diễn đặc trưng trong khi vẫn duy trì hiệu quả tính toán.

Một điểm quan trọng của GELAN là sự kết hợp giữa cơ chế chú ý SE (Squeeze-and-Excitation) và phép tích chập chiều sâu DWConv. Cơ chế SE cho phép mô hình học trọng số thích nghi theo từng kênh đặc trưng, từ đó tập trung mạnh hơn vào các vùng thông tin quan trọng liên quan đến bất thường. Trong khi đó, DWConv giúp giảm số lượng tham số và chi phí tính toán mà vẫn duy trì khả năng trích xuất đặc



Hình 3.3. Kiến trúc backbone GELAN được sử dụng trong mô hình đề xuất

trung không gian hiệu quả.

Nhờ cấu trúc phân cấp và cơ chế tổng hợp đặc trưng đa tầng, backbone GELAN có khả năng duy trì tốt các đặc trưng chi tiết của chuỗi cách điện trong suốt quá trình lan truyền sâu. Điều này đặc biệt quan trọng đối với các bất thường nhỏ như vết nứt hoặc bong tróc bề mặt.

Backbone GELAN được tích hợp trong FGS nhằm giải quyết trực tiếp vấn đề trên thông qua cơ chế tổng hợp đặc trưng hiệu quả giữa các tầng mạng. Khác với các backbone CNN truyền thống chỉ truyền thông tin theo chiều sâu tuần tự, GELAN sử dụng cơ chế kết nối đa nhánh và tái sử dụng đặc trưng nhằm bảo toàn tối đa thông tin gradient trong quá trình huấn luyện.

Cải tiến này mang lại ba lợi ích quan trọng.

Thứ nhất, mô hình duy trì được các đặc trưng không gian ở tầng nông trong khi vẫn khai thác được thông tin ngữ nghĩa mức cao từ các tầng sâu. Điều này giúp tăng cường khả năng phát hiện các bất thường nhỏ có hình dạng phức tạp.

Thứ hai, cơ chế lan truyền gradient hiệu quả giúp quá trình huấn luyện ổn định hơn, hạn chế hiện tượng mất gradient hoặc bùng nổ gradient khi mạng có số tầng lớn. Đây là yếu tố đặc biệt quan trọng đối với các mô hình kết hợp nhiều nhánh xử lý như FGS.

Thứ ba, backbone GELAN tối ưu hóa hiệu quả tính toán thông qua việc sử dụng

RepConv và DWConv. So với các phép tích chập chuẩn, DWConv giảm đáng kể số lượng tham số và chi phí FLOPs trong khi vẫn duy trì khả năng biểu diễn đặc trưng mạnh. Điều này giúp mô hình phù hợp hơn với các hệ thống triển khai thực tế yêu cầu xử lý thời gian thực.

Ngoài ra, cơ chế chú ý SE (Squeeze-and-Excitation) được tích hợp trong backbone cho phép mô hình học trọng số thích nghi theo từng kênh đặc trưng. Nhờ đó, mạng có xu hướng tập trung mạnh hơn vào các vùng chứa thông tin bất thường thay vì các thành phần nền không liên quan.

Trong bối cảnh ảnh UAV có độ phức tạp cao, cơ chế này giúp mô hình giảm đáng kể ảnh hưởng của các đặc trưng nhiễu như:

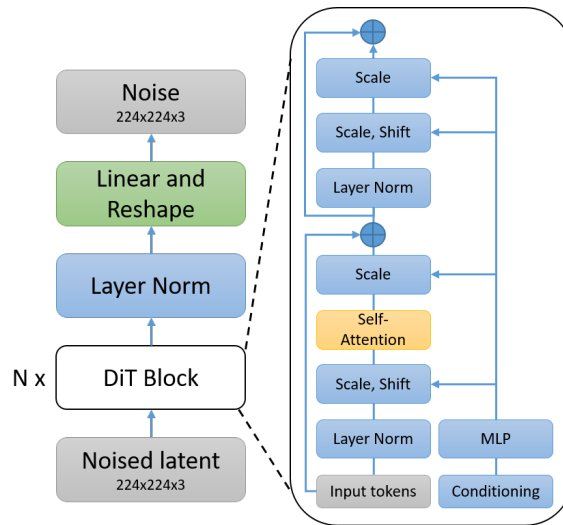
- tán cây có cấu trúc lặp;
- dây dẫn giao cắt;
- phản xạ ánh sáng từ bề mặt kim loại;
- các kết cấu nền có độ tương phản cao.

3.2.3 Mô hình khuếch tán (Diffusion Model)

Một trong những điểm khác biệt quan trọng của FGS so với các mô hình phát hiện bất thường truyền thống là việc tích hợp mô hình khuếch tán (Diffusion Model) vào giai đoạn tăng cường đặc trưng. Trong các mô hình CNN hoặc YOLO thông thường, đặc trưng được trích xuất chủ yếu thông qua các phép tích chập tuần tự. Mặc dù phương pháp này đạt hiệu quả cao đối với các đối tượng có kích thước lớn hoặc đặc trưng rõ ràng, khả năng bảo toàn thông tin chi tiết của các vùng bất thường nhỏ vẫn còn hạn chế. Đối với bài toán phát hiện bất thường trên chuỗi cách điện, các lỗi như nứt vi mô, bong tróc lớp phủ silicon hoặc nhiễm bẩn nhẹ thường chỉ xuất hiện dưới dạng các biến đổi cường độ rất nhỏ trên bề mặt vật liệu. Khi ảnh đầu vào trải qua nhiều tầng giảm mẫu liên tiếp, các đặc trưng này dễ bị triệt tiêu hoặc hòa lẫn với nhiễu nền. Điều này làm suy giảm đáng kể khả năng nhận diện lỗi giai đoạn sớm. Để giải quyết vấn đề trên, mô hình khuếch tán trong FGS không chỉ đóng vai trò khử nhiễu mà còn hoạt động như một cơ chế học phân bố đặc trưng của chuỗi cách điện. Thông qua quá trình thêm nhiễu và tái tạo lặp lại, mô hình học được cấu trúc hình học bất biến của đối tượng, từ đó tăng cường khả năng phục hồi các chi tiết biên và kết cấu nhỏ. Về bản chất, quá trình khuếch tán giúp mạng học cách phân biệt giữa nhiễu ngẫu nhiên và cấu trúc đặc trưng thực sự của cách điện. Điều này đặc biệt quan trọng trong môi trường UAV, nơi ảnh thường bị ảnh hưởng bởi:

- rung động của thiết bị bay;
- thay đổi điều kiện ánh sáng;
- hiện tượng mờ chuyển động (motion blur);
- nhiều nền phức tạp từ cây cối, dây dẫn hoặc công trình xây dựng.

Khác với các bộ lọc làm trơn truyền thống có xu hướng làm mất chi tiết biên, mô hình khuếch tán trong FGS duy trì được các đặc trưng có giá trị phân biệt cao, đồng thời giảm nhiễu không liên quan. Nhờ đó, khả năng phát hiện các bất thường nhỏ được cải thiện đáng kể. Ngoài ra, việc tích hợp Diffusion Model còn giúp tăng cường tính tổng quát hóa của mô hình. Thay vì học trực tiếp trên dữ liệu đầu vào nguyên thủy vốn có sự phân bố không ổn định, mạng học trên không gian đặc trưng đã được chuẩn hóa thông qua cơ chế khuếch tán. Điều này giúp mô hình thích nghi tốt hơn với các điều kiện vận hành khác nhau mà không cần phụ thuộc quá mức vào đặc trưng riêng của từng bộ dữ liệu.



Hình 3.4. Kiến trúc mô hình DiT (Diffusion Transformer) được sử dụng trong mô-đun khuếch tán của hệ thống FGS

Hình 3.4 trình bày kiến trúc DiT (Diffusion Transformer) được sử dụng trong mô hình đề xuất. Khác với các mạng CNN truyền thống chỉ học ánh xạ trực tiếp từ ảnh đầu vào sang đặc trưng đầu ra, mô hình khuếch tán thực hiện quá trình học thông qua hai pha gồm khuếch tán thuận (forward diffusion) và khuếch tán nghịch (reverse diffusion).

Trong pha khuếch tán thuận, nhiễu Gaussian được thêm dần vào dữ liệu đầu vào nhằm phá hủy cấu trúc gốc của ảnh. Quá trình này giúp mô hình học được sự biến thiên của dữ liệu trong không gian đặc trưng. Ngược lại, ở pha khuếch tán nghịch, mô hình học cách tái tạo dữ liệu sạch từ dữ liệu đã bị nhiễu. Đối với bài toán phát hiện

bất thường trên chuỗi cách điện, cơ chế này đặc biệt hữu ích trong việc khôi phục các chi tiết biên và cấu trúc hình học nhỏ vốn dễ bị mất mát trong ảnh UAV.

Giả sử ảnh đầu vào được ký hiệu là I , quá trình thêm nhiễu và khử nhiễu được biểu diễn như sau:

$$F_0 = ReverseDeblurring(F_1) \quad (3.1)$$

$$F_1 = ForwardBlurring(F_0) \quad (3.2)$$

trong đó F_1 biểu diễn đặc trưng sau khi thêm nhiễu và F_0 là đặc trưng được tái tạo sau quá trình khử nhiễu. Sau khi quá trình tái tạo đặc trưng hoàn tất, các vùng quan tâm được kết hợp với đầu ra của mạng RPN để xây dựng đặc trưng RoI:

$$F_R = RoIFeatureExtractor(F_0, O) \quad (3.3)$$

trong đó O là tập các vùng đề xuất từ RPN và F_R là đặc trưng RoI cuối cùng được sử dụng cho các tác vụ phát hiện và phân đoạn.

3.2.4 Vùng quan tâm và các nhánh dự đoán

Trong mô hình đề xuất, lớp RoI Align đóng vai trò trung gian quan trọng trong việc kết hợp thông tin từ nhiều nguồn đặc trưng khác nhau. Lớp này tiếp nhận ba đầu vào chính gồm: đặc trưng từ RPN, đặc trưng từ backbone GELAN và đặc trưng tăng cường từ mô hình khuếch tán. Việc kết hợp đa nguồn đặc trưng giúp mô hình xây dựng được biểu diễn RoI có độ phân biệt cao hơn.

Sau khi trải qua quá trình RoI Align, các đặc trưng được đưa vào RCNN Head để thực hiện hai tác vụ chính gồm phát hiện và phân loại bất thường. Đối với nhiệm vụ phát hiện, đặc trưng RoI được xử lý qua các lớp fully connected nhằm hồi quy vị trí đối tượng và xác định lớp tương ứng.

Đối với nhiệm vụ phân loại, RCNN Head sử dụng hàm softmax để xây dựng phân bố xác suất trên từng lớp bất thường. Kết quả đầu ra cho phép xác định trạng thái của chuỗi cách điện dựa trên xác suất dự đoán cao nhất. Bên cạnh nhánh phát hiện, mô hình còn tích hợp nhánh Mask Head nhằm thực hiện phân đoạn mức điểm ảnh. Nhánh này sinh ra các mặt nạ nhị phân biểu diễn chính xác hình dạng của vùng lỗi. Để nâng cao chất lượng phân đoạn, MaskIoU Head được bổ sung nhằm hồi quy độ tương đồng giữa mask dự đoán và mask chuẩn. Hàm mất mát L2 được sử dụng để tối ưu hóa quá trình hồi quy MaskIoU, giúp cải thiện độ chính xác của mặt nạ phân đoạn cuối cùng.

So với các mô hình CNN hoặc YOLO truyền thống, phương pháp đề xuất có ba đóng góp chính. Thứ nhất, mô hình tích hợp cơ chế khuếch tán nhằm tăng cường

khả năng biểu diễn đặc trưng trong điều kiện ảnh UAV có nhiễu mạnh và độ phân giải thấp. Thứ hai, mô hình xây dựng một kiến trúc hợp nhất đồng thời thực hiện phát hiện, phân loại và phân đoạn trong cùng một hệ thống. Thứ ba, toàn bộ kiến trúc được tối ưu hóa riêng cho bài toán phát hiện bất thường kích thước nhỏ trên chuỗi cách điện. Các cải tiến này giúp mô hình đạt hiệu năng vượt trội so với các phương pháp đối sánh, đồng thời nâng cao khả năng ứng dụng thực tế trong các hệ thống giám sát lưới điện thông minh.

3.2.5 Giai đoạn hậu xử lý (Post-processing)

Sau khi các nhánh RCNN Head, Mask Head và MaskIoU Head hoàn thành quá trình suy luận, hệ thống thu được tập các dự đoán bao gồm vị trí đối tượng, nhãn lớp và mặt nạ phân đoạn tương ứng. Tuy nhiên, các kết quả đầu ra ban đầu vẫn có thể tồn tại các dự đoán trùng lặp, các đối tượng có độ tin cậy thấp hoặc các mặt nạ phân đoạn chưa đạt chất lượng mong muốn. Vì vậy, mô hình FGS tích hợp một giai đoạn hậu xử lý nhằm nâng cao độ chính xác và độ ổn định của kết quả trước khi đưa ra quyết định cuối cùng.

Trong bước đầu tiên, các dự đoán có độ tin cậy thấp được loại bỏ dựa trên một ngưỡng xác suất được xác định thông qua tập xác thực. Quá trình này giúp giảm thiểu các dự đoán sai và hạn chế ảnh hưởng của nhiễu trong ảnh UAV.

Tiếp theo, thuật toán Non-Maximum Suppression (NMS) được áp dụng nhằm xử lý hiện tượng nhiều bounding box cùng dự đoán cho một đối tượng. Thuật toán giữ lại dự đoán có độ tin cậy cao nhất và loại bỏ các vùng dự đoán chồng lấp quá mức. Nhờ đó, hệ thống có thể xác định chính xác vị trí của từng bất thường mà không tạo ra các kết quả dư thừa.

Đối với nhánh phân đoạn, Mask Head tạo ra các bản đồ xác suất biểu diễn khả năng xuất hiện bất thường tại từng điểm ảnh. Các bản đồ này được chuyển đổi thành mặt nạ nhị phân thông qua quá trình ngưỡng hóa để xác định chính xác phạm vi của vùng lỗi. Sau đó, MaskIoU Head được sử dụng để đánh giá chất lượng của từng mặt nạ dự đoán. Các mặt nạ có chất lượng thấp sẽ bị loại bỏ hoặc hiệu chỉnh nhằm đảm bảo độ tin cậy của kết quả phân đoạn cuối cùng.

Sau khi hoàn thành các bước lọc và hiệu chỉnh, thông tin từ nhánh phát hiện và nhánh phân đoạn được hợp nhất để tạo thành kết quả đầu ra cuối cùng của hệ thống. Mỗi bất thường được mô tả thông qua vị trí xuất hiện, loại lỗi, độ tin cậy dự đoán và mặt nạ phân đoạn tương ứng. Kết quả này không chỉ cho phép xác định sự tồn tại của bất thường mà còn cung cấp thông tin chi tiết về hình dạng và phạm vi ảnh hưởng của vùng hư hỏng.

Trong môi trường vận hành thực tế, các kết quả hậu xử lý tiếp tục được sử dụng để tính toán các chỉ số phục vụ đánh giá tình trạng kỹ thuật của chuỗi cách điện, chẳng hạn như diện tích vùng hư hỏng, mức độ nhiễm bẩn hoặc số lượng bát cách điện bị mất. Những thông tin này hỗ trợ hiệu quả cho công tác kiểm tra, bảo trì dự báo và quản lý tài sản lưới điện, góp phần nâng cao độ tin cậy và an toàn vận hành của hệ thống truyền tải điện.

3.3 Thử nghiệm

3.3.1 Thiết lập thử nghiệm

Toàn bộ các thử nghiệm được triển khai trên nền tảng Python 3.10 sử dụng thư viện PyTorch. Hệ thống thử nghiệm được trang bị bộ xử lý Intel Core i9-13900K, bộ nhớ RAM 64 GB và card đồ họa NVIDIA RTX 4090 với 24 GB bộ nhớ đồ họa.

Các mô hình được huấn luyện từ đầu (training-from-scratch) nhằm đảm bảo tính công bằng trong quá trình so sánh. Kích thước ảnh đầu vào được thiết lập ở mức 640×640 pixel. Quá trình huấn luyện kéo dài 50 epoch với bộ tối ưu AdamX và tốc độ học khởi tạo bằng 5×10^{-4} .

Trong năm epoch đầu tiên, cơ chế linear warm-up được áp dụng nhằm ổn định quá trình lan truyền gradient và cải thiện khả năng hội tụ. Sau giai đoạn này, tốc độ học được giảm dần theo lịch trình cosine decay. Để hạn chế sự phụ thuộc quá mức vào dữ liệu tăng cường, toàn bộ kỹ thuật augmentation được vô hiệu hóa trong 10 epoch cuối của quá trình huấn luyện.

3.3.2 Chuẩn bị dữ liệu

Để đánh giá hiệu quả của phương pháp đề xuất, nghiên cứu sử dụng hai bộ dữ liệu bao gồm bộ dữ liệu thực tế VNElectric được xây dựng từ ảnh UAV thu nhận trên hệ thống lưới điện truyền tải 110 kV khu vực miền Bắc Việt Nam và bộ dữ liệu công khai Insulator Defect Detection [112]. Việc sử dụng đồng thời dữ liệu thực tế và dữ liệu công khai cho phép đánh giá cả hiệu năng phát hiện lẫn khả năng tổng quát hóa của mô hình trên các miền dữ liệu khác nhau.

Dữ liệu VNElectric được thu nhận bằng thiết bị bay không người lái (UAV) trong quá trình kiểm tra định kỳ đường dây truyền tải điện. So với các bộ dữ liệu được xây dựng trong môi trường kiểm soát, dữ liệu UAV phản ánh chính xác điều kiện vận hành thực tế với nhiều biến thiên về góc quan sát, độ cao bay, khoảng cách thu nhận và điều kiện môi trường. Các ảnh bao gồm nhiều kịch bản khác nhau như góc chụp từ trên xuống, góc chụp từ dưới lên, góc nghiêng, ảnh cận cảnh và ảnh toàn cảnh của

chuỗi cách điện.

Một trong những thách thức chính của bài toán là các bất thường trên chuỗi cách điện thường chỉ chiếm một vùng rất nhỏ trong ảnh, trong khi nền ảnh lại chứa nhiều thành phần gây nhiễu như cây cối, công trình xây dựng, dây dẫn và các thiết bị điện khác. Ngoài ra, hiện tượng rung lắc UAV, thay đổi điều kiện chiếu sáng, phản xạ ánh sáng trên bề mặt cách điện thủy tinh và che khuất cục bộ làm gia tăng đáng kể độ khó của bài toán phát hiện.

Bộ dữ liệu VNElectric bao gồm tổng cộng 1.596 ảnh được gán nhãn thủ công bởi các chuyên gia vận hành lưới điện. Dữ liệu chứa nhiều trạng thái vận hành khác nhau của chuỗi cách điện, bao gồm cả mẫu bình thường và mẫu bất thường. Trong đó, cách điện thủy tinh chiếm 57.6% tổng số mẫu, trong khi cách điện sứ chiếm 42.4%. Sự đa dạng về chủng loại thiết bị giúp nâng cao tính đại diện của bộ dữ liệu và hỗ trợ đánh giá khả năng thích nghi của mô hình trong các điều kiện thực tế.

Ngoài sự đa dạng về đối tượng, bộ dữ liệu còn bao gồm nhiều loại nền ảnh khác nhau. Phân bố chi tiết được trình bày trong Bảng 3.1. Có thể nhận thấy nền rừng cây và khu vực công trình chiếm tỷ lệ lớn nhất, phản ánh đúng điều kiện vận hành thực tế của hệ thống truyền tải điện tại Việt Nam.

Bảng 3.1. Phân bố các loại nền trong bộ dữ liệu VNElectric

Loại nền	Số lượng	Tỷ lệ (%)
Rừng cây	521	32.6
Công trình xây dựng	487	30.5
Đất nông nghiệp	263	16.5
Thiết bị điện khác	325	20.4
Tổng cộng	1,596	100.0

Để giảm hiện tượng quá khớp và cải thiện khả năng tổng quát hóa, nhiều kỹ thuật tăng cường dữ liệu được áp dụng trong quá trình huấn luyện, bao gồm xoay ngẫu nhiên, lật ngang, thay đổi tỷ lệ, dịch chuyển vị trí và cắt ảnh ngẫu nhiên. Đồng thời, phương pháp lấy mẫu phân tầng được sử dụng nhằm duy trì sự cân bằng giữa các nhóm dữ liệu trong từng tập huấn luyện và kiểm định.

Toàn bộ dữ liệu được chia theo tỷ lệ 3:1:1 tương ứng với tập huấn luyện, tập xác thực và tập kiểm tra. Tập huấn luyện được sử dụng để tối ưu tham số mạng, tập xác thực được dùng để lựa chọn siêu tham số và giám sát quá trình hội tụ, trong khi tập kiểm tra được sử dụng để đánh giá hiệu năng cuối cùng của mô hình.

3.3.3 Kết quả thử nghiệm

Để đánh giá hiệu quả của phương pháp đề xuất, nghiên cứu tiến hành so sánh với kiến trúc YOLOv9-E trên cùng bộ dữ liệu và cùng điều kiện huấn luyện. Các độ đo được sử dụng bao gồm Precision, Recall và mAP@50. Đây là các chỉ số phổ biến trong bài toán phát hiện đối tượng, phản ánh đồng thời khả năng phân loại, phát hiện đầy đủ đối tượng và độ chính xác định vị.

Kết quả thử nghiệm được trình bày trong Bảng 3.2. Mô hình FGS đạt Precision 92.3%, Recall 91.4% và mAP@50 đạt 93.7%, vượt trội so với YOLOv9-E trên tất cả các độ đo đánh giá.

Bảng 3.2. So sánh hiệu năng phát hiện bất thường chuỗi cách điện trên bộ dữ liệu VNElectric

Mô hình	Params (M)	Precision	Recall	mAP@50
YOLOv9-E	57.3	90.6	90.3	92.8
FGS (đề xuất)	103.4	92.3	91.4	93.7

So với YOLOv9-E, mô hình FGS cải thiện 1.7% về Precision, 1.1% về Recall và 0.9% về mAP@50. Mặc dù mức cải thiện về mAP@50 không quá lớn, đây là kết quả đáng chú ý trong bối cảnh YOLOv9-E vốn đã là một trong những kiến trúc phát hiện đối tượng có hiệu năng cao. Điều này cho thấy các thành phần được đề xuất đã bổ sung hiệu quả cho khả năng biểu diễn đặc trưng của hệ thống.

Hiệu quả vượt trội của mô hình FGS đến từ sự kết hợp giữa ba thành phần chính. Thứ nhất, backbone GELAN giúp tăng cường khả năng lan truyền thông tin và tái sử dụng đặc trưng trong quá trình trích xuất đặc trưng sâu. Thứ hai, cấu trúc FPN hỗ trợ hợp nhất thông tin đa tỉ lệ, cho phép mô hình phát hiện hiệu quả các bất thường có kích thước nhỏ xuất hiện ở khoảng cách xa. Thứ ba, mô hình khuếch tán (Diffusion Model) hỗ trợ tái tạo và tăng cường các đặc trưng quan trọng, giúp cải thiện khả năng phân biệt giữa đối tượng mục tiêu và nền ảnh phức tạp.

Ngoài đánh giá định lượng, kết quả quan sát trực quan cho thấy mô hình đề xuất có khả năng phát hiện ổn định trong nhiều điều kiện khác nhau, bao gồm các trường hợp nền ảnh phức tạp, đối tượng bị che khuất một phần hoặc xuất hiện dưới điều kiện ánh sáng bất lợi. Trong khi đó, các mô hình đối sánh thường gặp hiện tượng bỏ sót đối tượng nhỏ hoặc nhầm lẫn với các thành phần nền có đặc điểm thị giác tương đồng.

Nghiên cứu Ablation

Để đánh giá mức độ đóng góp của từng thành phần trong kiến trúc đề xuất, nghiên cứu tiến hành các thí nghiệm ablation bằng cách lần lượt loại bỏ từng mô-đun trong

hệ thống và giữ nguyên các thành phần còn lại.

Kết quả được trình bày trong Bảng 3.3.

Bảng 3.3. Nghiên cứu ablation của mô hình FGS

Cấu hình	Params (M)	Precision (%)	Recall (%)	mAP@50 (%)
Không sử dụng GELAN	99.1	88.4	87.5	89.8
Không sử dụng FPN	91.1	86.2	87.6	88.1
Không sử dụng Diffusion Model	69.4	91.7	90.5	92.0
FGS đầy đủ	103.4	92.3	91.4	93.7

Kết quả trong Bảng 3.3 cho thấy tất cả các thành phần của mô hình đều đóng góp tích cực vào hiệu năng tổng thể của hệ thống.

Trong số các thành phần được khảo sát, FPN là mô-đun có ảnh hưởng lớn nhất. Khi loại bỏ FPN, mAP@50 giảm từ 93.7% xuống còn 88.1%, tương ứng mức suy giảm 5.6%. Kết quả này khẳng định vai trò quan trọng của đặc trưng đa tỉ lệ trong bài toán phát hiện bất thường chuỗi cách điện.

Backbone GELAN cũng đóng góp đáng kể với mức cải thiện khoảng 3.9% mAP@50. Điều này cho thấy khả năng truyền dẫn thông tin hiệu quả và cơ chế tái sử dụng đặc trưng của GELAN giúp nâng cao chất lượng biểu diễn ngữ nghĩa trong toàn bộ hệ thống.

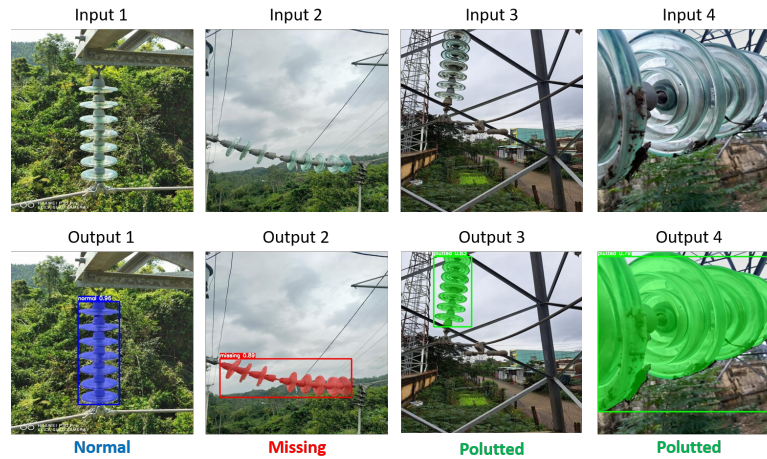
Đối với Diffusion Model, mặc dù mức cải thiện về mAP@50 đạt 1.7%, thành phần này đóng vai trò quan trọng trong việc tăng cường khả năng chống nhiễu và làm nổi bật các vùng đặc trưng quan trọng. Vai trò của Diffusion Model đặc biệt rõ rệt trong các trường hợp nền ảnh phức tạp hoặc điều kiện chiếu sáng không thuận lợi.

Một kết quả đáng chú ý là hiệu năng của mô hình không phụ thuộc vào một thành phần đơn lẻ mà đến từ sự phối hợp giữa các mô-đun. GELAN cung cấp khả năng trích xuất đặc trưng mạnh, FPN hỗ trợ hợp nhất thông tin đa tỉ lệ, trong khi Diffusion Model tăng cường chất lượng biểu diễn đặc trưng. Sự kết hợp này tạo nên một kiến trúc hoàn chỉnh có khả năng xử lý hiệu quả các thách thức của dữ liệu UAV trong môi trường vận hành thực tế.

Phân tích trực quan

Bên cạnh các đánh giá định lượng, nghiên cứu tiến hành phân tích trực quan kết quả phát hiện nhằm đánh giá khả năng hoạt động của mô hình trong các điều kiện thực tế khác nhau.

Hình 3.5 minh họa một số kết quả phát hiện tiêu biểu của mô hình FGS trên bộ dữ liệu VNElectric. Có thể nhận thấy rằng mô hình duy trì khả năng phát hiện ổn



Hình 3.5. Kết quả phát hiện bất thường chuỗi cách điện trong các điều kiện khác nhau.

định trong nhiều điều kiện quan sát khác nhau.

Trong các trường hợp nền ảnh là rừng cây hoặc công trình xây dựng, FGS vẫn xác định chính xác vị trí của chuỗi cách điện và các vùng bất thường liên quan. Điều này cho thấy khả năng phân biệt hiệu quả giữa đối tượng mục tiêu và các thành phần nền có đặc điểm thị giác tương đồng.

Đối với các trường hợp đối tượng xuất hiện ở khoảng cách xa hoặc chỉ chiếm một vùng nhỏ trong ảnh, mô hình vẫn duy trì độ chính xác cao. Kết quả này phản ánh hiệu quả của cơ chế hợp nhất đặc trưng đa tỉ lệ trong FPN, cho phép khai thác thông tin từ nhiều mức phân giải khác nhau.

Một số ví dụ cũng cho thấy khả năng hoạt động ổn định của mô hình trong điều kiện chiếu sáng không đồng đều. Ngay cả khi đối tượng xuất hiện dưới điều kiện ngược sáng hoặc có độ tương phản thấp, mô hình vẫn có thể nhận diện được các vùng bất thường với độ chính xác tương đối cao.

Tuy nhiên, nghiên cứu cũng ghi nhận một số trường hợp thất bại điển hình. Các lỗi chủ yếu xuất hiện trong các tình huống sau:

- Chuỗi cách điện bị che khuất nghiêm trọng bởi cây cối hoặc các thành phần kết cấu khác.
- Hiện tượng phản xạ mạnh trên bề mặt cách điện thủy tinh gây mất thông tin kết cấu.
- Ảnh bị mờ do rung động của UAV hoặc điều kiện thời tiết không thuận lợi.
- Đối tượng bất thường có kích thước cực nhỏ và xuất hiện ở khoảng cách rất xa.

Mặc dù vẫn tồn tại một số trường hợp khó, mô hình FGS nhìn chung thể hiện khả năng phát hiện ổn định hơn so với các phương pháp đối sánh trong phần lớn các điều kiện thực tế.

Thảo luận

Các kết quả thử nghiệm cho thấy mô hình FGS có tiềm năng ứng dụng cao trong các hệ thống giám sát và kiểm tra tự động đường dây truyền tải điện sử dụng UAV. Mô hình đạt được hiệu năng vượt trội trong phát hiện và phân đoạn các bất thường trên chuỗi cách điện, đồng thời duy trì tính ổn định trong nhiều điều kiện môi trường khác nhau. Điều này cho thấy khả năng thích nghi tốt của phương pháp đề xuất đối với các thách thức thường gặp trong thực tế như sự thay đổi điều kiện chiếu sáng, nền ảnh phức tạp, nhiễu môi trường và sự đa dạng về hình thái của chuỗi cách điện.

Kết quả so sánh với các mô hình đối chứng cho thấy việc kết hợp cơ chế hợp nhất đặc trưng đa tỉ lệ và mô hình khuếch tán đã góp phần nâng cao khả năng nhận diện các bất thường có kích thước nhỏ hoặc độ tương phản thấp. Bên cạnh đó, cơ chế phân đoạn ở mức pixel giúp cung cấp thông tin chi tiết hơn về vị trí và phạm vi hư hỏng, tạo điều kiện thuận lợi cho các bước phân tích chuyên sâu, đánh giá tình trạng thiết bị và xây dựng các hệ thống hỗ trợ ra quyết định trong công tác quản lý vận hành lưới điện.

Tuy nhiên, bên cạnh những kết quả đạt được, mô hình FGS vẫn tồn tại một số hạn chế cần được xem xét khi triển khai trong thực tế. Với quy mô khoảng 103,4 triệu tham số, mô hình đòi hỏi tài nguyên tính toán và bộ nhớ tương đối lớn so với các kiến trúc phát hiện đối tượng thời gian thực hiện nay. Điều này dẫn đến sự đánh đổi nhất định giữa độ chính xác và tốc độ suy luận. Mặc dù mô hình phù hợp cho các kịch bản kiểm tra định kỳ hoặc phân tích chuyên sâu sau chuyến bay, việc triển khai trực tiếp trên các nền tảng điện toán biên tích hợp trên UAV vẫn có thể gặp trở ngại về độ trễ xử lý, đặc biệt đối với các thiết bị có năng lực tính toán hạn chế.

Từ góc độ ứng dụng thực tế, đây là sự đánh đổi hợp lý nhằm ưu tiên độ tin cậy của hệ thống phát hiện. Trong lĩnh vực giám sát lưới điện truyền tải, việc bỏ sót các bất thường hoặc đưa ra các cảnh báo sai có thể gây ảnh hưởng trực tiếp đến công tác vận hành và bảo trì thiết bị. Do đó, nghiên cứu ưu tiên nâng cao độ chính xác phát hiện và giảm thiểu các trường hợp dương tính giả, ngay cả khi phải chấp nhận chi phí tính toán cao hơn so với các mô hình được tối ưu hóa cho tốc độ.

Ngoài ra, các thử nghiệm trong nghiên cứu chủ yếu được thực hiện trên dữ liệu ảnh quang học thu nhận từ UAV đối với các tuyến truyền tải điện 110 kV. Khả năng tổng quát hóa của mô hình đối với các hệ thống điện áp cao hơn như 220 kV hoặc 500 kV, nơi khoảng cách quan sát lớn hơn và kích thước biểu diễn của đối tượng trên ảnh nhỏ hơn, vẫn cần được tiếp tục đánh giá. Đồng thời, các dạng bất thường tiềm ẩn chưa biểu hiện rõ trên bề mặt thiết bị cũng là thách thức đối với các phương pháp chỉ sử dụng ảnh quang học.

Trong tương lai, các nghiên cứu tiếp theo có thể tập trung vào việc tối ưu hóa kiến trúc nhằm giảm chi phí tính toán thông qua các kỹ thuật lượng tử hóa tham số, cắt tỉa mạng, tri thức chưng cất (knowledge distillation) và tối ưu hóa suy luận bằng TensorRT trên các nền tảng phần cứng nhúng. Bên cạnh đó, việc tích hợp dữ liệu đa phương thức như ảnh nhiệt, LiDAR hoặc dữ liệu cảm biến chuyên dụng có thể giúp nâng cao khả năng phát hiện các bất thường tiềm ẩn và mở rộng phạm vi ứng dụng của hệ thống trong công tác bảo trì dự báo.

Tổng hợp các kết quả thử nghiệm cho thấy mô hình FGS đã đạt được sự cân bằng hiệu quả giữa độ chính xác phát hiện, khả năng chống nhiễu và tính tổng quát hóa. Mặc dù vẫn còn những thách thức liên quan đến chi phí tính toán và triển khai thời gian thực, phương pháp đề xuất đã chứng minh được tiềm năng ứng dụng trong các hệ thống giám sát lưới điện thông minh dựa trên UAV và tạo nền tảng quan trọng cho các nghiên cứu tiếp theo về phát hiện bất thường và dự báo sự cố thiết bị điện.

3.4 Kết luận chương 3

Chương 3 đã trình bày một cách hệ thống nghiên cứu đề xuất và triển khai mô hình FGS (FPN-GELAN-Diffusion Segmentation) nhằm giải quyết bài toán phát hiện bất thường trên chuỗi cách điện từ ảnh UAV trong môi trường truyền tải điện phức tạp tại Việt Nam. Dựa trên những kết quả thực nghiệm và phân tích định lượng, nghiên cứu rút ra các kết luận trọng tâm sau:

Thứ nhất, việc tích hợp mô hình khuếch tán (Diffusion Model) dựa trên kiến trúc Diffusion Transformer (DiT) đã nâng cao khả năng tái tạo đặc trưng và loại bỏ nhiễu nền phức tạp. Cơ chế này giúp cải thiện đáng kể khả năng phát hiện các bất thường có kích thước nhỏ hoặc độ tương phản thấp, vốn là những trường hợp khó xử lý đối với các mô hình phát hiện truyền thống.

Thứ hai, sự kết hợp giữa khung xương GELAN (Generalized Efficient Layer Aggregation Network) và mạng kim tự tháp đặc trưng FPN đã nâng cao hiệu quả trích xuất đặc trưng đa tỉ lệ, đảm bảo khả năng thích nghi của mô hình đối với sự biến thiên về kích thước, hình thái và vật liệu của chuỗi cách điện trong điều kiện vận hành thực tế.

Thứ ba, việc triển khai đồng thời cơ chế phân đoạn mức pixel thông qua các khối RCNN Head, Mask Head và đặc biệt là MaskIoU Head đã nâng cao độ chính xác định lượng cho hệ thống. Kết quả thực nghiệm trên bộ dữ liệu VNElectric cho thấy mô hình FGS đạt Precision 92,3

Mặc dù đạt hiệu năng cao, mô hình FGS vẫn tồn tại một số hạn chế nhất định. Kiến trúc có số lượng tham số lớn dẫn đến chi phí tính toán và độ trễ suy luận cao hơn

so với các mô hình thời gian thực, gây khó khăn cho việc triển khai trực tiếp trên các nền tảng điện toán biên có tài nguyên hạn chế. Bên cạnh đó, khả năng tổng quát hóa của mô hình trên các hệ thống điện áp cao hơn và các nguồn dữ liệu đa phương thức vẫn cần được tiếp tục nghiên cứu và đánh giá.

Tổng kết lại, cách tiếp cận phân tầng sử dụng kết quả định vị RoI từ Chương 2 làm đầu vào cho mô hình FGS ở Chương 3 đã chứng minh được tính hiệu quả trong việc giảm nhiễu nền, tăng cường độ chính xác phát hiện và nâng cao độ tin cậy của hệ thống giám sát tự động. Những kết quả đạt được không chỉ góp phần nâng cao hiệu quả phát hiện bất thường trên chuỗi cách điện mà còn tạo cơ sở khoa học quan trọng cho việc phát triển các hệ thống hỗ trợ ra quyết định và bảo trì dự báo đối với lưới truyền tải điện thông minh trong các chương tiếp theo của luận án.

CHƯƠNG 4

SINH DỮ LIỆU ẢNH BẤT THƯỜNG CỦA CHUỖI CÁCH ĐIỆN

Sự phát triển nhanh chóng của công nghệ thiết bị bay không người lái (UAV) cùng với những tiến bộ vượt bậc trong lĩnh vực học sâu đã tạo ra những thay đổi đáng kể trong công tác kiểm tra, giám sát và bảo trì hệ thống truyền tải điện. Thay vì phụ thuộc vào các phương pháp kiểm tra thủ công truyền thống với nhiều hạn chế về chi phí, thời gian và mức độ an toàn, UAV cho phép thu nhận dữ liệu hình ảnh từ nhiều góc quan sát khác nhau, bao phủ phạm vi rộng và tiếp cận các vị trí khó quan sát trong điều kiện vận hành thực tế. Kết hợp với các mô hình học sâu hiện đại, dữ liệu ảnh UAV đã trở thành nguồn thông tin quan trọng phục vụ các bài toán phát hiện, nhận diện bất thường và đánh giá tình trạng thiết bị điện.

Trong số các thành phần của hệ thống truyền tải điện, chuỗi cách điện là một trong những thiết bị có vai trò đặc biệt quan trọng nhằm đảm bảo khả năng cách ly điện và duy trì độ tin cậy vận hành của đường dây. Các bất thường xuất hiện trên chuỗi cách điện như nứt vỡ, sứt mẻ, phóng điện bề mặt hoặc nhiễm bẩn nghiêm trọng có thể làm suy giảm khả năng cách điện, dẫn tới các sự cố ảnh hưởng trực tiếp đến an toàn vận hành hệ thống điện. Vì vậy, việc tự động phát hiện và nhận diện các dạng bất thường trên chuỗi cách điện từ ảnh UAV đã trở thành một trong những hướng nghiên cứu nhận được nhiều sự quan tâm trong những năm gần đây.

Các kết quả nghiên cứu được trình bày trong các chương trước cho thấy việc ứng dụng học sâu đã mang lại những cải thiện đáng kể về khả năng phát hiện và nhận diện thiết bị điện cũng như các dạng bất thường trên chuỗi cách điện. Đặc biệt, mô hình phát hiện hai giai đoạn được đề xuất ở Chương 3 đã góp phần nâng cao độ chính xác nhờ tập trung phân tích trực tiếp vùng chứa đối tượng thay vì xử lý trên toàn bộ ảnh UAV. Tuy nhiên, hiệu quả của các mô hình học sâu hiện nay vẫn chịu ảnh hưởng đáng kể bởi chất lượng và sự phân bố của dữ liệu huấn luyện.

Trong thực tế vận hành lưới điện, các trạng thái bất thường thường xuất hiện với tần suất thấp hơn rất nhiều so với trạng thái vận hành bình thường. Nhiều dạng lỗi nghiêm trọng chỉ xuất hiện trong những điều kiện đặc biệt hoặc xảy ra với xác suất rất nhỏ, khiến việc thu thập số lượng lớn mẫu dữ liệu trở nên khó khăn và tốn kém. Bên cạnh đó, một số dạng bất thường mới có thể xuất hiện trong quá trình vận hành nhưng chưa từng được ghi nhận trong các bộ dữ liệu hiện có. Những đặc điểm này dẫn đến hiện tượng mất cân bằng dữ liệu nghiêm trọng giữa các lớp bình thường và các lớp

bất thường, đồng thời làm phát sinh bài toán thiếu dữ liệu đối với các lớp lỗi hiếm.

Sự mất cân bằng dữ liệu không chỉ làm giảm khả năng học đặc trưng của các lớp thiểu số mà còn khiến mô hình có xu hướng thiên lệch về các lớp xuất hiện với tần suất cao trong tập huấn luyện. Hệ quả là nhiều trường hợp bất thường quan trọng có thể không được nhận diện chính xác, làm giảm độ tin cậy của hệ thống giám sát tự động. Đối với các phương pháp học sâu hiện đại, đặc biệt là các mô hình có số lượng tham số lớn, tình trạng thiếu dữ liệu còn làm gia tăng nguy cơ quá khớp và hạn chế khả năng tổng quát hóa khi triển khai trên dữ liệu thực tế.

Để khắc phục những hạn chế trên, một trong những hướng nghiên cứu đang nhận được nhiều sự quan tâm là sinh dữ liệu ảnh bất thường (abnormal image generation). Thay vì phụ thuộc hoàn toàn vào quá trình thu thập dữ liệu thực địa, các kỹ thuật sinh dữ liệu cho phép tạo ra các mẫu bất thường mới có chất lượng cao và đa dạng về ngữ cảnh xuất hiện. Các mẫu dữ liệu được sinh ra không chỉ góp phần cân bằng phân bố dữ liệu giữa các lớp mà còn hỗ trợ mô hình học được các đặc trưng biểu diễn phong phú hơn đối với các dạng lỗi hiếm hoặc các dạng bất thường mới xuất hiện. Đây được xem là một giải pháp tiềm năng nhằm nâng cao hiệu quả huấn luyện các mô hình phát hiện và nhận diện bất thường trong điều kiện dữ liệu hạn chế.

Xuất phát từ những yêu cầu thực tiễn nêu trên, chương này tập trung nghiên cứu bài toán sinh dữ liệu ảnh bất thường cho chuỗi cách điện từ ảnh UAV nhằm hỗ trợ nâng cao hiệu quả của các mô hình phát hiện và nhận diện lỗi. Nội dung chương được tổ chức thành các phần chính như sau: (i) phân tích đặc điểm dữ liệu bất thường trong giám sát thiết bị điện và các nguyên nhân dẫn đến hiện tượng mất cân bằng dữ liệu; (ii) khảo sát các hướng tiếp cận sinh dữ liệu phổ biến dựa trên học sâu; (iii) đề xuất phương pháp sinh dữ liệu ảnh bất thường phục vụ bài toán nhận diện lỗi trên chuỗi cách điện; và (iv) đánh giá thử nghiệm nhằm phân tích mức độ cải thiện của dữ liệu sinh đối với hiệu năng của các mô hình phát hiện và nhận diện bất thường.

4.1 Các nghiên cứu liên quan về sinh dữ liệu trong huấn luyện các mô hình học sâu

4.1.1 Các phương pháp sinh dữ liệu truyền thống

Hiệu quả của các mô hình học sâu phụ thuộc mạnh vào số lượng và chất lượng dữ liệu huấn luyện. Tuy nhiên, trong nhiều bài toán thực tế, đặc biệt là phát hiện bất thường công nghiệp và giám sát thiết bị điện bằng UAV, việc thu thập dữ liệu thường gặp nhiều khó khăn do chi phí cao, điều kiện vận hành phức tạp và sự hiếm gặp của các trạng thái lỗi. Để khắc phục hạn chế này, các kỹ thuật sinh dữ liệu (Data Augmentation) đã

được phát triển nhằm mở rộng tập huấn luyện và cải thiện khả năng tổng quát hóa của mô hình [113].

Các phương pháp tăng cường dữ liệu truyền thống chủ yếu dựa trên các phép biến đổi hình học và quang học của ảnh đầu vào như xoay (Rotation), lật (Flip), tịnh tiến (Translation), thay đổi tỷ lệ (Scaling), cắt xén (Cropping), biến đổi phối cảnh (Perspective Transformation), thay đổi độ sáng, độ tương phản và bổ sung nhiễu ngẫu nhiên [38]. Các phép biến đổi này giúp mô hình học được tính bất biến đối với góc quan sát, vị trí xuất hiện của đối tượng và điều kiện môi trường.

Sau đó, nhiều kỹ thuật tăng cường dữ liệu tiên tiến hơn đã được đề xuất nhằm tạo ra các mẫu huấn luyện có độ đa dạng cao hơn. Mixup [114] tạo dữ liệu mới bằng cách nội suy tuyến tính giữa hai ảnh và hai nhãn tương ứng. CutMix [115] thực hiện việc thay thế một vùng ảnh bằng vùng tương ứng từ ảnh khác, giúp mô hình học được các đặc trưng cục bộ của đối tượng. Trong lĩnh vực phát hiện đối tượng, Mosaic được tích hợp trong YOLOv4 [65], cho phép kết hợp bốn ảnh thành một ảnh huấn luyện duy nhất nhằm gia tăng sự đa dạng về ngữ cảnh và kích thước đối tượng.

Mặc dù đạt hiệu quả tích cực, các phương pháp tăng cường dữ liệu truyền thống chủ yếu tạo ra các biến thể của dữ liệu hiện có và không thể sinh ra các mẫu bất thường hoàn toàn mới. Điều này trở thành hạn chế đáng kể trong các bài toán phát hiện lỗi thiết bị điện, nơi số lượng ảnh lỗi thường rất ít và phân bố dữ liệu mất cân bằng nghiêm trọng.

Nhằm giải quyết vấn đề này, các mô hình sinh dữ liệu dựa trên học sâu đã được phát triển mạnh mẽ trong những năm gần đây. Generative Adversarial Network (GAN) do Goodfellow và cộng sự đề xuất [102] cho phép học phân bố dữ liệu và sinh ra các ảnh mới có độ chân thực cao. Nhiều biến thể như DCGAN [116], Conditional GAN [117] và StyleGAN [118] đã được ứng dụng rộng rãi trong các bài toán tăng cường dữ liệu công nghiệp. Bên cạnh đó, Variational Autoencoder (VAE) [119] cung cấp cơ chế học không gian tiềm ẩn liên tục, giúp sinh dữ liệu mới một cách ổn định hơn.

Gần đây, các mô hình khuếch tán (Diffusion Models) như DDPM [120], Improved DDPM [121] và Latent Diffusion Model [122] đã chứng minh khả năng sinh ảnh có chất lượng vượt trội so với nhiều kiến trúc GAN truyền thống. Các mô hình này đang trở thành hướng nghiên cứu chủ đạo trong sinh dữ liệu nhờ khả năng tạo ra các mẫu đa dạng và có độ chân thực cao, đặc biệt phù hợp với các bài toán dữ liệu ít mẫu.

4.1.2 Các phương pháp sinh dữ liệu và học với dữ liệu hạn chế trong phát hiện bất thường

Trong các hệ thống công nghiệp, dữ liệu bất thường thường xuất hiện với tần suất rất thấp so với dữ liệu bình thường. Điều này dẫn tới hiện tượng mất cân bằng dữ liệu nghiêm trọng và gây khó khăn cho quá trình huấn luyện các mô hình học sâu. Để giải quyết vấn đề này, nhiều nghiên cứu đã tập trung vào việc kết hợp sinh dữ liệu với các phương pháp học ít mẫu (Few-Shot Learning) và học gia tăng (Incremental Learning).

Các nghiên cứu ban đầu chủ yếu sử dụng GAN để sinh thêm dữ liệu bất thường nhằm cân bằng tập huấn luyện. Schlegl và cộng sự [123] đề xuất kiến trúc GAN phục vụ phát hiện bất thường trong ảnh y tế. Sau đó, Akcay và cộng sự [124] phát triển mô hình GANomaly, cho phép học biểu diễn của dữ liệu bình thường và phát hiện các mẫu bất thường thông qua sai lệch trong không gian đặc trưng. Các nghiên cứu này cho thấy dữ liệu tổng hợp có thể cải thiện đáng kể hiệu quả nhận dạng trong điều kiện số lượng mẫu bất thường hạn chế.

Bên cạnh hướng sinh dữ liệu, một nhánh nghiên cứu khác tập trung khai thác tri thức từ các mô hình nền quy mô lớn đã được huấn luyện trước trên tập dữ liệu khổng lồ. Sự xuất hiện của Vision Transformer (ViT) [125] và các mô hình nền thị giác đã mở ra hướng tiếp cận mới trong học gia tăng với dữ liệu hạn chế. Thay vì huấn luyện lại toàn bộ mạng nơ-ron, các phương pháp này chỉ điều chỉnh một số lượng nhỏ tham số bổ sung nhằm tận dụng tri thức có sẵn trong mô hình nền.

Một trong những hướng nghiên cứu nổi bật là Prompt Learning. Ý tưởng cốt lõi của Prompt Learning là giữ nguyên toàn bộ trọng số của backbone đã được tiền huấn luyện và chỉ học thêm các token prompt nhằm điều hướng quá trình trích xuất đặc trưng [126]. Cách tiếp cận này giúp giảm đáng kể số lượng tham số cần tối ưu, đồng thời hạn chế hiện tượng quên thảm họa trong học gia tăng.

Learning to Prompt (L2P) [127] là một trong những công trình tiên phong áp dụng Prompt Learning cho bài toán Class-Incremental Learning. Phương pháp này xây dựng một tập prompt ứng viên (Prompt Pool) và sử dụng cơ chế truy xuất dựa trên độ tương đồng đặc trưng để lựa chọn các prompt phù hợp với từng mẫu đầu vào. Kết quả thử nghiệm trên CIFAR100 và ImageNet-R cho thấy L2P đạt hiệu năng vượt trội so với nhiều phương pháp replay-free trước đó. Tuy nhiên, hiệu quả của mô hình phụ thuộc mạnh vào chất lượng của Prompt Pool. Trong điều kiện Few-Shot Class-Incremental Learning (FSCIL), số lượng mẫu huấn luyện hạn chế khiến các prompt khó học được biểu diễn giàu ngữ nghĩa cho các lớp mới.

Để cải thiện khả năng lưu trữ tri thức, DualPrompt [128] đề xuất kiến trúc gồm hai loại prompt: General Prompt dùng để lưu trữ tri thức chung và Expert Prompt

dùng để lưu trữ tri thức đặc thù của từng nhiệm vụ. Cơ chế này giúp mô hình cân bằng tốt hơn giữa việc duy trì tri thức cũ và học tri thức mới. Tuy nhiên, số lượng Expert Prompt tăng dần theo số nhiệm vụ, làm giảm khả năng mở rộng của mô hình trong các kịch bản học gia tăng dài hạn.

Tiếp theo, CodaPrompt [129] đề xuất xây dựng không gian prompt trực giao nhằm giảm hiện tượng chồng lấn tri thức giữa các nhiệm vụ. Thay vì học các prompt độc lập, CodaPrompt tạo ra một không gian prompt liên tục được cập nhật trong suốt quá trình học. Mặc dù đạt hiệu quả cao hơn L2P và DualPrompt, phương pháp này vẫn gặp khó khăn khi số lượng mẫu của các lớp mới quá ít.

Một hướng nghiên cứu mới hơn là sử dụng cơ chế sinh prompt động từ dữ liệu đầu vào. Attention-aware Self-adaptive Prompt (ASP) [130] được xem là một trong những phương pháp tiêu biểu theo hướng này. Khác với các phương pháp Prompt Learning truyền thống, ASP sử dụng Prompt Encoder để sinh ra các prompt thích nghi dựa trên đặc trưng ảnh đầu vào. Kiến trúc của ASP bao gồm Task-Invariant Prompt (TIP) nhằm lưu trữ tri thức dùng chung và Task-Specific Prompt (TSP) được sinh động cho từng mẫu dữ liệu. Nhờ đó, mô hình có khả năng chuyển giao tri thức hiệu quả hơn giữa các lớp đã học và các lớp mới.

Một đóng góp quan trọng của ASP là tích hợp nguyên lý Information Bottleneck [131] vào quá trình sinh prompt. Thông qua cơ chế này, Prompt Encoder được khuyến khích giữ lại các thông tin liên quan đến nhiệm vụ phân loại và loại bỏ các thành phần dư thừa hoặc nhiễu. Kết quả thử nghiệm trên CIFAR100, CUB200 và ImageNet-R cho thấy ASP đạt độ chính xác cao hơn nhiều phương pháp FSCIL trước đó.

Mặc dù đạt được những kết quả tích cực, các nghiên cứu hiện nay chủ yếu được đánh giá trên các bộ dữ liệu nhận dạng đối tượng tổng quát. Trong bài toán nhận diện bất thường trên chuỗi cách điện, các đặc trưng lỗi như nứt bề mặt, sứt vành, vỡ thân hoặc bám bẩn thường chỉ chiếm một vùng rất nhỏ trong ảnh và có độ tương phản thấp so với nền. Đồng thời, dữ liệu lỗi thường có mức độ mất cân bằng nghiêm trọng và liên tục xuất hiện các dạng bất thường mới trong quá trình vận hành. Những đặc điểm này đặt ra yêu cầu cần xây dựng một cơ chế Prompt Learning có khả năng thích nghi tốt với dữ liệu ít mẫu, duy trì tri thức cũ và đồng thời tối ưu hóa không gian đặc trưng nhằm tăng khả năng phân biệt giữa các lớp lỗi.

Đây cũng chính là khoảng trống nghiên cứu mà luận án hướng tới giải quyết thông qua mô hình AAPID được đề xuất ở phần tiếp theo.

4.2 AAPID (Attention-aware Adaptive Prompting for Insulator Diagnosis) – Mô hình Prompt thích nghi dựa trên cơ chế Attention cho chẩn đoán lỗi chuỗi cách điện

Sử dụng mô hình học sâu truyền thống để giải quyết bài toán phát hiện chuỗi cách điện và nhận diện bất thường đem lại hiệu quả khá cao với độ chính xác đạt trên 90%. Tuy nhiên, trong thực tế kiểm tra vận hành của lưới truyền tải điện, bất thường trên chuỗi cách điện thường có số lượng hạn chế và phân bố không đồng đều. Đặc biệt, khi xuất hiện các dạng bất thường mới, các mô hình học sâu huấn luyện theo cách thông thường dễ gặp phải hiện tượng quên thảm họa (*catastrophic forgetting*) hoặc yêu cầu phải huấn luyện lại toàn bộ mô hình với chi phí lớn.

Để khắc phục các hạn chế này, luận án kế thừa và phát triển hướng tiếp cận Few-Shot Class-Incremental Learning (FSCIL) cho bài toán nhận diện bất thường trên chuỗi cách điện. Trên cơ sở đó, mô hình Attention-aware Adaptive Prompting for Insulator Diagnosis (AAPID) được đề xuất và tích hợp vào luận án như một giải pháp nâng cao cho bài toán phát hiện và nhận diện bất thường trong điều kiện dữ liệu hạn chế.

4.2.1 Tổng quan kiến trúc

Trong bối cảnh bài toán Few-Shot Class-Incremental Learning (FSCIL), mô hình cần đồng thời giải quyết hai thách thức cơ bản gồm: (i) học được các lớp mới chỉ với số lượng rất nhỏ mẫu huấn luyện và (ii) duy trì tri thức đã học trước đó mà không xảy ra hiện tượng quên thảm họa (*catastrophic forgetting*). Để giải quyết bài toán này, phương pháp Attention-aware Adaptive Prompting for Insulator Diagnosis (AAPID) được đề xuất dựa trên nền tảng Vision Transformer (ViT) kết hợp với cơ chế sinh prompt thích nghi và học biểu diễn có giám sát.

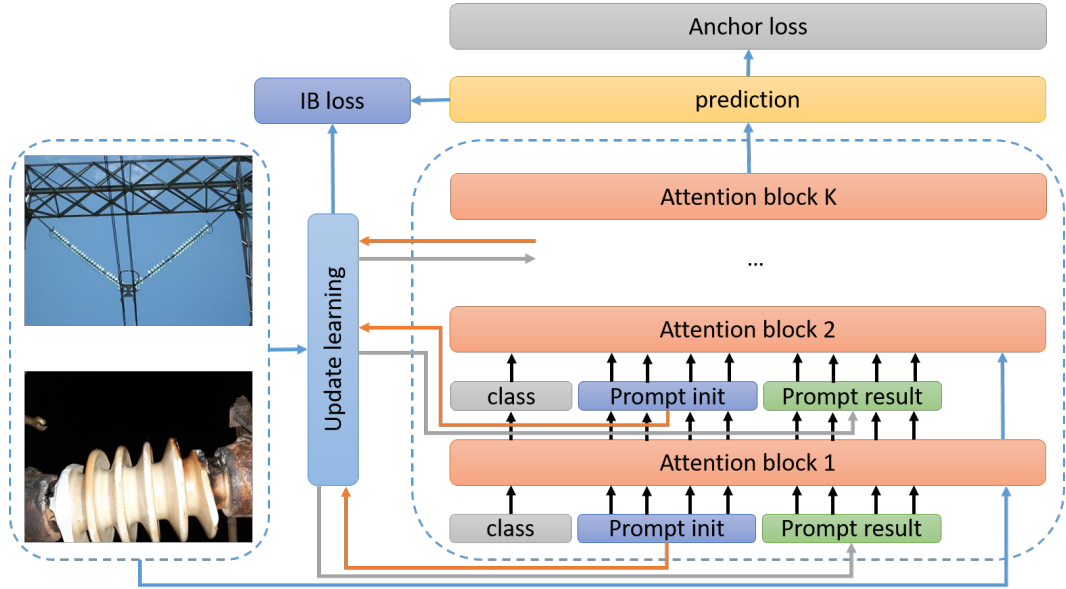
Khác với các phương pháp FSCIL truyền thống thường cập nhật toàn bộ tham số mạng khi xuất hiện lớp mới, AAPID giữ cố định toàn bộ trọng số của bộ mã hóa ViT đã được huấn luyện trước. Trong quá trình học gia tăng, mô hình chỉ tối ưu một số lượng nhỏ tham số liên quan đến prompt và bộ sinh prompt. Cách tiếp cận này giúp tận dụng khả năng biểu diễn mạnh của mô hình nền đồng thời hạn chế nguy cơ quá khớp khi số lượng mẫu huấn luyện rất ít.

Về mặt cấu trúc, kiến trúc AAPID bao gồm bốn thành phần chính:

1. Bộ mã hóa đặc trưng ViT cố định;
2. Prompt bất biến theo nhiệm vụ (*Task-Invariant Prompt – TIP*);

3. Bộ sinh Prompt thích nghi (*Prompt Encoder*);
4. Hàm mất mát kết hợp giữa Information Bottleneck Loss và Anchor Loss.

Các thành phần này phối hợp với nhau nhằm xây dựng không gian đặc trưng có tính phân biệt cao đối với các lớp lỗi mới trong khi vẫn duy trì khả năng nhận biết các lớp đã học.



Hình 4.1. Kiến trúc tổng thể của mô hình AAPID

4.2.2 Cơ chế Prompt thích nghi dựa trên Attention

Bộ mã hóa Vision Transformer cố định

Thành phần cốt lõi của AAPID là bộ mã hóa Vision Transformer được huấn luyện trước trên tập dữ liệu quy mô lớn như ImageNet. Khác với chiến lược fine-tuning toàn bộ mạng thường được áp dụng trong học sâu truyền thống, AAPID khóa toàn bộ tham số của backbone trong suốt quá trình huấn luyện FSCIL.

Giả sử ảnh đầu vào được ký hiệu là:

$$X \in \mathbb{R}^{H \times W \times 3} \quad (4.1)$$

Ảnh được chia thành N patch có kích thước cố định:

$$\mathcal{X} = \{x_1, x_2, \dots, x_N\} \quad (4.2)$$

Mỗi patch được ánh xạ sang không gian đặc trưng thông qua lớp nhúng patch:

$$z_i = E(x_i) \quad (4.3)$$

trong đó $E(\cdot)$ là hàm patch embedding. Tập các patch embedding được biểu diễn dưới dạng:

$$Z_p[z_1, z_2, \dots, z_N] \quad (4.4)$$

Sau khi bổ sung embedding vị trí, chuỗi token được đưa qua K khối Transformer liên tiếp:

$$Z^{(l)} \text{Transformer}_l \left(Z^{(l-1)} \right), \quad l = 1, \dots, K \quad (4.5)$$

với $Z^{(0)}$ là đầu vào ban đầu của ViT. Biểu diễn cuối cùng thu được là:

$$Z^{(K)} \text{Transformer}_K \circ \dots \circ \text{Transformer}_1 \left(Z^{(0)} \right) \quad (4.6)$$

Việc giữ cố định bộ mã hóa mang lại nhiều lợi ích quan trọng. Thứ nhất, nó giúp hạn chế hiện tượng quá khớp trên các tập dữ liệu few-shot. Thứ hai, toàn bộ tri thức nền đã học từ tập dữ liệu quy mô lớn được bảo tồn trong suốt quá trình học gia tăng. Cuối cùng, số lượng tham số cần tối ưu giảm đáng kể, từ đó giúp quá trình huấn luyện ổn định hơn.

Tuy nhiên, khi backbone bị đóng băng, khả năng thích nghi với các lớp mới phụ thuộc chủ yếu vào cơ chế prompt. Đây chính là động lực hình thành hệ thống prompting thích nghi trong AAPID.

Prompt bất biến theo nhiệm vụ

Trong các phương pháp Prompt Learning hiện đại, prompt được xem như một tập tham số học được nhằm điều chỉnh hành vi của mô hình nền mà không cần thay đổi trọng số của backbone.

AAPID sử dụng một tập prompt khởi tạo gọi là Task-Invariant Prompt (TIP), được biểu diễn bởi thành phần “Prompt Init” trong Hình 4.1.

Mục tiêu của TIP là học các đặc trưng mang tính phổ quát, tồn tại ở tất cả các lớp đối tượng. Đối với bài toán chẩn đoán lỗi cách điện, các thông tin này bao gồm:

- Hình dạng tổng thể của chuỗi cách điện;
- Màu sắc vật liệu;
- Đặc điểm cấu trúc hình học;
- Các đặc trưng bề mặt cơ bản.

Giả sử tập prompt gồm M token:

$$P_{TIP} = [p_1, p_2, \dots, p_M] \quad (4.7)$$

Theo thiết kế của AAPID, toàn bộ token trong TIP được khởi tạo cùng một giá trị. Điều này buộc cơ chế attention xử lý các prompt theo cách nhất quán giữa các nhiệm vụ khác nhau, từ đó giúp prompt học được tri thức dùng chung thay vì ghi nhớ đặc điểm của một lớp cụ thể.

Sau khi được ghép với chuỗi patch token, đầu vào của Transformer trở thành:

$$Z = [P_{TIP}; X] \quad (4.8)$$

Trong đó các prompt cùng tham gia vào quá trình self-attention với các token ảnh.

Bộ sinh Prompt thích nghi

Mặc dù TIP có khả năng lưu trữ tri thức chung, thành phần này chưa đủ để phân biệt các loại lỗi khác nhau. Do đó AAPID bổ sung cơ chế Task-Specific Prompt (TSP), được thể hiện trong kiến trúc dưới tên gọi “Prompt Result”.

Khác với các phương pháp như L2P hay DualPrompt, AAPID không học riêng một prompt cho từng lớp. Thay vào đó, mô hình xây dựng một Prompt Encoder để sinh ra prompt thích nghi trực tiếp từ dữ liệu đầu vào.

Giả sử đặc trưng ảnh đầu vào được biểu diễn bởi:

$$F = f(X) \quad (4.9)$$

Prompt Encoder thực hiện phép ánh xạ:

$$P_{TSP} = G(F) \quad (4.10)$$

trong đó:

- $G(\cdot)$ là Prompt Encoder;
- P_{TSP} là prompt thích nghi được sinh ra.

Các prompt này chứa thông tin đặc thù của từng loại lỗi như:

- Nứt bề mặt;
- Sứt vành;
- Vỡ thân;
- Bám bẩn nghiêm trọng.

Nhờ đó mô hình có khả năng học được các lớp mới ngay cả khi chỉ có một số lượng rất nhỏ mẫu huấn luyện.

Cơ chế tương tác giữa Prompt và Self-Attention

Điểm đặc biệt của AAPID là prompt không chỉ được chèn vào đầu mạng mà còn được cập nhật liên tục qua các Attention Block.

Trong mỗi tầng Transformer, các prompt tham gia trực tiếp vào quá trình tính toán attention:

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d}}\right) V \quad (4.11)$$

trong đó:

$$Q = W_Q Z, \quad K = W_K Z, \quad V = W_V Z \quad (4.12)$$

với Z bao gồm cả token ảnh và token prompt.

Nhờ cơ chế này, prompt đóng vai trò như một bộ điều khiển attention, hướng mô hình tập trung vào các vùng ảnh liên quan đến lỗi thay vì các thành phần nền không quan trọng.

Về bản chất, attention lúc này không còn là quá trình học hoàn toàn dựa trên dữ liệu đầu vào mà được dẫn hướng bởi tri thức được lưu trữ trong prompt. Điều này giúp tăng khả năng nhận diện các đặc trưng lỗi có kích thước nhỏ hoặc xuất hiện với tần suất thấp trong tập dữ liệu.

4.2.3 Học biểu diễn Prompt bằng Information Bottleneck

Ràng buộc Information Bottleneck

Để Prompt Encoder sinh ra các prompt có khả năng tổng quát hóa cao, AAPID áp dụng nguyên lý Information Bottleneck (IB).

Ý tưởng cốt lõi của IB là giữ lại thông tin liên quan đến nhãn lớp đồng thời loại bỏ các thông tin dư thừa hoặc nhiễu từ dữ liệu đầu vào.

Hàm mục tiêu được xây dựng như sau:

$$L_{IB} = I(P; X) - \gamma I(P; y) \quad (4.13)$$

trong đó:

- $I(P; X)$ là lượng thông tin giữa prompt và dữ liệu đầu vào;

- $I(P; y)$ là lượng thông tin giữa prompt và nhãn lớp;
- γ là hệ số cân bằng.

Mục tiêu tối ưu là giảm $I(P; X)$ nhưng đồng thời tăng $I(P; y)$. Khi đó prompt chỉ giữ lại những đặc trưng quan trọng liên quan đến việc phân loại lỗi.

Do việc tính toán trực tiếp Mutual Information là khó thực hiện, bài báo sử dụng dạng xấp xỉ biến phân:

$$L_{IB} = \mathbb{E}[-\log P(y|x)] + \beta_{KL} D_{KL}(P(p|x) || r(p)) \quad (4.14)$$

Thành phần thứ nhất là Cross-Entropy Loss nhằm đảm bảo khả năng phân loại chính xác. Thành phần thứ hai là KL Divergence đóng vai trò điều chuẩn, ngăn prompt ghi nhớ quá nhiều thông tin riêng biệt của từng mẫu.

Nhờ đó, Prompt Encoder có khả năng sinh ra các prompt mang tính tổng quát cao, đặc biệt phù hợp với bài toán few-shot learning.

4.2.4 Anchor Loss và tối ưu không gian đặc trưng

Mặc dù Information Bottleneck giúp prompt giàu ngữ nghĩa hơn, không gian đặc trưng vẫn có thể xuất hiện hiện tượng chồng lấn giữa các lớp.

Để tăng cường tính phân biệt của biểu diễn đặc trưng, AAPID sử dụng thêm Anchor Loss.

Đối với mỗi lớp k , mô hình lựa chọn một mẫu đại diện làm anchor:

$$\hat{c}_k = f_{\theta,p}(\hat{x}_k) \quad (4.15)$$

Anchor Loss được xác định dựa trên độ tương đồng cosine:

$$L_{anchor} = 1 - \frac{\hat{c}_k^T f_{\theta,p}(x_k)}{\|\hat{c}_k\| \cdot \|f_{\theta,p}(x_k)\|} \quad (4.16)$$

Hàm mất mát này có tác dụng:

- Thu hẹp khoảng cách giữa các mẫu cùng lớp;
- Tăng khoảng cách giữa các lớp khác nhau;
- Tạo ra các cụm đặc trưng rõ ràng trong không gian embedding.

Nhờ đó, bộ phân loại cuối cùng có thể dễ dàng xác định ranh giới giữa các lớp ngay cả trong điều kiện số lượng mẫu rất hạn chế.

4.2.5 Hàm mục tiêu và phân tích mô hình

Hàm mục tiêu tổng thể

Toàn bộ mô hình được tối ưu thông qua sự kết hợp giữa Information Bottleneck Loss và Anchor Loss:

$$L_{total} = L_{IB} + \lambda L_{anchor} \quad (4.17)$$

trong đó λ là hệ số điều chỉnh mức độ đóng góp của Anchor Loss.

Sự kết hợp giữa hai thành phần này tạo nên cơ chế tối ưu kép, vừa đảm bảo khả năng học biểu diễn ngữ nghĩa của prompt, vừa tối ưu tính phân biệt của không gian đặc trưng.

Đánh giá kiến trúc

Từ góc độ học sâu, AAPID là sự kết hợp giữa ba hướng nghiên cứu hiện đại gồm Vision Transformer tiền huấn luyện, Prompt Learning và Few-Shot Class-Incremental Learning.

Điểm đổi mới quan trọng nhất của kiến trúc nằm ở việc sử dụng Prompt Encoder dựa trên Information Bottleneck để sinh prompt thích nghi thay vì học prompt cố định cho từng lớp. Điều này cho phép mô hình chuyển giao tri thức từ các lớp đã biết sang các lớp mới với số lượng mẫu rất nhỏ.

Đồng thời, Anchor Loss giúp hình thành không gian đặc trưng có tính phân biệt cao, làm giảm đáng kể hiện tượng quên thảm họa trong quá trình học gia tăng. Nhờ vậy, AAPID trở thành một kiến trúc phù hợp cho các bài toán nhận dạng lỗi thiết bị điện trong thực tế, nơi dữ liệu lỗi mới thường hiếm và xuất hiện liên tục trong quá trình vận hành.

4.3 Thử nghiệm

Mục tiêu của phần thử nghiệm là đánh giá toàn diện hiệu quả của phương pháp Attention-aware Adaptive Prompting for Insulator Diagnosis (AAPID) trong bài toán học tăng dần ít mẫu đối với nhận diện lỗi chuỗi cách điện từ ảnh UAV. Các thử nghiệm được thiết kế nhằm kiểm chứng ba khía cạnh quan trọng của mô hình, bao gồm khả năng học các lớp lỗi mới với số lượng mẫu huấn luyện hạn chế, khả năng duy trì tri thức đã học trong quá trình học tăng dần và khả năng xây dựng không gian đặc trưng có tính phân biệt cao giữa các loại lỗi khác nhau.

Bên cạnh việc đánh giá trên bộ dữ liệu chuẩn thường được sử dụng trong cộng

đồng nghiên cứu FSCIL, luận án còn tiến hành thử nghiệm trên bộ dữ liệu thực tế thu nhận từ hệ thống truyền tải điện tại Việt Nam nhằm đánh giá khả năng triển khai của phương pháp trong môi trường ứng dụng thực tế.

4.3.1 Môi trường thử nghiệm

Toàn bộ thử nghiệm được triển khai trên nền tảng PyTorch 2.0 kết hợp thư viện CUDA để tăng tốc quá trình huấn luyện. Mô hình nền được lựa chọn là Vision Transformer ViT-B/16 đã được huấn luyện trước trên tập dữ liệu ImageNet-1K. Việc sử dụng mô hình tiền huấn luyện giúp tận dụng khả năng biểu diễn đặc trưng tổng quát của các mô hình nền tảng thị giác hiện đại, đồng thời giảm đáng kể yêu cầu về số lượng dữ liệu huấn luyện trong miền ứng dụng thiết bị điện.

Trong tất cả các thử nghiệm, ảnh đầu vào được chuẩn hóa theo thống kê của ImageNet và thay đổi kích thước về 224×224 pixel. Các phép tăng cường dữ liệu cơ bản như lật ngang ngẫu nhiên, xoay góc nhỏ và thay đổi độ sáng được áp dụng trong giai đoạn huấn luyện nhằm tăng khả năng tổng quát hóa của mô hình.

Giai đoạn học cơ sở được thực hiện trong 20 epoch với bộ tối ưu SGD và tốc độ học khởi tạo bằng 0,01. Kích thước batch được lựa chọn là 48 dựa trên khả năng xử lý của phần cứng và kết quả khảo sát thử nghiệm ban đầu. Độ dài prompt được thiết lập bằng 3 đối với CIFAR100 và bằng 5 đối với VPI-Fault-2500 nhằm đảm bảo khả năng biểu diễn phù hợp với độ phức tạp của từng bộ dữ liệu.

Trong hàm mất mát tổng hợp của mô hình, hệ số cân bằng λ được lựa chọn thông qua quá trình kiểm chứng chéo trên tập xác thực. Cách tiếp cận này giúp xác định cấu hình tham số tối ưu đồng thời hạn chế hiện tượng quá khớp trong quá trình huấn luyện.

Đối với các lớp tăng dần trong thiết lập FSCIL, chỉ có 5 mẫu của mỗi lớp mới được sử dụng để huấn luyện, trong khi các mẫu còn lại được dùng để đánh giá khả năng tổng quát hóa của mô hình. Thiết lập này mô phỏng sát điều kiện thực tế trong giám sát lưới điện, khi các dạng lỗi mới thường chỉ xuất hiện với số lượng mẫu rất hạn chế.

4.3.2 Dữ liệu thử nghiệm

Để đánh giá toàn diện hiệu quả của phương pháp AAPID trong bài toán học tăng dần ít mẫu đối với nhận diện lỗi chuỗi cách điện, các thử nghiệm được tiến hành trên hai bộ dữ liệu có đặc tính khác nhau, bao gồm một bộ dữ liệu chuẩn được sử dụng rộng rãi trong cộng đồng nghiên cứu FSCIL và một bộ dữ liệu ứng dụng được xây dựng từ dữ liệu UAV thu nhận trên lưới điện truyền tải tại Việt Nam.

Bộ dữ liệu thứ nhất là CIFAR100 [132], bao gồm 60.000 ảnh màu kích thước 32×32 pixel thuộc 100 lớp đối tượng khác nhau. Đây là bộ dữ liệu chuẩn thường được sử dụng để đánh giá các thuật toán học tăng dần ít mẫu do có số lượng lớp lớn và mức độ đa dạng cao. Theo thiết lập phổ biến trong các nghiên cứu FSCIL, 60 lớp đầu tiên được sử dụng làm tập lớp cơ sở để huấn luyện ban đầu, trong khi 40 lớp còn lại được chia thành 8 phiên học tăng dần liên tiếp, mỗi phiên bao gồm 5 lớp mới với 5 mẫu huấn luyện cho mỗi lớp. Thiết lập này cho phép đánh giá đồng thời khả năng tiếp thu tri thức mới và khả năng duy trì tri thức đã học của mô hình trong suốt quá trình học tăng dần.

Bộ dữ liệu thứ hai là VPI-Fault-2500, được xây dựng từ dữ liệu ảnh UAV thu nhận trong quá trình kiểm tra hệ thống truyền tải điện cao thế tại Việt Nam. Bộ dữ liệu bao gồm 2.500 ảnh chuỗi cách điện, được phân loại thành năm nhãn: Normal (bình thường), Surface Crack (nứt bề mặt), Rim Chip (sứt vành), Body Breakage (vỡ thân), và Severe Surface Contamination (nhiễm bẩn bề mặt nghiêm trọng), với 500 ảnh cho mỗi nhãn.

Để mô phỏng bài toán xuất hiện lỗi mới trong quá trình vận hành hệ thống điện, ba lớp Normal, Surface Crack và Rim Chip được lựa chọn làm các lớp cơ sở. Hai lớp lỗi hiếm còn lại là Body Breakage và Severe Surface Contamination được đưa vào một phiên học tăng dần duy nhất dưới dạng bài toán 2-way 5-shot. Trong thiết lập này, mô hình chỉ được cung cấp năm mẫu huấn luyện cho mỗi lớp lỗi mới. Đây là tình huống thường gặp trong thực tế khi các dạng bất thường nghiêm trọng xuất hiện với tần suất thấp và rất khó thu thập đủ số lượng mẫu để huấn luyện theo cách tiếp cận học sâu truyền thống.

4.3.3 Kết quả thử nghiệm

Các thử nghiệm được xây dựng nhằm trả lời ba câu hỏi chính:

- **RQ1:** Phương pháp AAPID có cải thiện hiệu năng tổng thể so với các phương pháp học tăng dần và học tăng dần ít mẫu hiện nay hay không?
- **RQ2:** Mức độ đóng góp của từng thành phần trong kiến trúc AAPID đối với hiệu năng cuối cùng của hệ thống là như thế nào?
- **RQ3:** AAPID có khả năng học được không gian đặc trưng có tính phân biệt cao giữa các lớp lỗi cơ sở và các lớp lỗi mới hay không?

Chỉ số đánh giá

Để đánh giá toàn diện hiệu quả của mô hình, luận án sử dụng ba chỉ số thường được áp dụng trong các nghiên cứu FSCIL.

- *Độ chính xác trung bình (Average Accuracy – Aavg)*: là giá trị trung bình của độ chính xác trên tất cả các lớp qua toàn bộ các phiên học ($t = 0, \dots, T$). Chỉ số này phản ánh hiệu năng tổng thể của mô hình trong suốt quá trình học tăng dần.
- *Mức suy giảm hiệu năng (Performance Dropping-PD)*: được xác định là $A_0 - A_T$, với A_0 là độ chính xác trên các lớp cơ sở sau phiên học đầu tiên, và A_T là độ chính xác trên cùng các lớp đó sau phiên học cuối cùng. Chỉ số này dùng để định lượng mức độ quên thảm họa (catastrophic forgetting) của mô hình.
- *Độ chính xác điều hòa (Harmonic Accuracy – HAcc)*: là giá trị trung bình điều hòa giữa độ chính xác trên các lớp cơ sở (A_0) và độ chính xác trung bình trên các lớp mới (A_n) sau phiên học cuối. Chỉ số này đánh giá mức độ cân bằng giữa lớp mới và khả năng duy trì tri thức của lớp

$$HAcc = \frac{2A_0A_n}{A_0 + A_n} \quad (4.18)$$

trong đó, A_0 là độ chính xác trên các lớp cơ sở, A_n là độ chính xác trung bình trên các lớp mới sau phiên học cuối cùng.

Các phương pháp đối sánh

Để đảm bảo tính khách quan trong đánh giá, luận án tiến hành so sánh AAPID với ba nhóm phương pháp tiêu biểu.

Nhóm thứ nhất bao gồm các phương pháp học tăng dần truyền thống như iCaRL và FOSTER. Đây là các phương pháp đã được sử dụng rộng rãi trong cộng đồng nghiên cứu CIL và đóng vai trò là các đường cơ sở quan trọng.

Nhóm thứ hai gồm các phương pháp FSCIL hiện đại như CEC, FACT và TEEN. Đây là các phương pháp được thiết kế chuyên biệt cho kịch bản học tăng dần ít mẫu, do đó phản ánh trực tiếp mức độ cạnh tranh của phương pháp đề xuất.

Nhóm thứ ba là các phương pháp học tăng dần dựa trên prompt bao gồm L2P, DualP và CodaP. Bên cạnh các phiên bản gốc, luận án còn triển khai các phiên bản cải tiến sử dụng bộ phân loại nguyên mẫu nhằm đánh giá chính xác hơn tiềm năng của hướng tiếp cận dựa trên prompt trong bối cảnh FSCIL.

Kết quả đánh giá tổng thể

Kết quả trình bày trong Bảng 4.1 và Bảng 4.2 cho thấy phương pháp AAPID đạt hiệu năng tốt nhất trên cả hai bộ dữ liệu.

Bảng 4.1. So sánh hiệu năng trên tập dữ liệu CIFAR100 (Thiết lập: 60+8x5-way 5-shot)

Phương pháp	$A_{avg}(\%) \uparrow$	PD($\%$) $\downarrow$	HAcc($\%$) $\uparrow$
iCaRL	78.4	27.2	57.5
Foster	73.9	35.9	11.0
CEC	81.0	19.0	64.1
FACT	78.6	21.7	55.5
TEEN	87.3	8.8	81.2
L2P	71.1	37.0	0.0
DualP	70.6	36.9	0.1
CodaP	72.0	37.4	0.0
DualP+	81.1	8.5	75.3
AAPID (Ours)	89.5	4.8	86.1

Bảng 4.2. So sánh hiệu năng trên tập dữ liệu VPI-Fault-2500 (thiết lập: 3+2-way 5-shot)

Phương pháp	$A_{avg}(\%) \uparrow$	PD($\%$) $\downarrow$	HAcc($\%$) $\uparrow$
iCaRL	85.1	15.4	78.2
Foster	82.6	20.1	65.7
TEEN	91.5	6.2	88.3
DualP+	88.9	8.0	85.1
AAPID (đề xuấtS)	94.2	3.1	92.5

Đối với CIFAR100, AAPID đạt độ chính xác trung bình 89,5%, cao hơn tất cả các phương pháp đối sánh. Đồng thời, mức suy giảm hiệu năng chỉ đạt 4,8%, thấp nhất trong số các phương pháp được khảo sát. Điều này cho thấy cơ chế prompting thích nghi kết hợp với chiến lược học tăng dần đề xuất đã giúp mô hình duy trì hiệu quả tri thức của các lớp đã học trong khi vẫn tiếp thu tốt các lớp mới.

Trên bộ dữ liệu VPI-Fault-2500, ưu thế của phương pháp đề xuất tiếp tục được khẳng định. AAPID đạt $A_{avg} = 94,2\%$ và $HAcc = 92,5\%$, vượt đáng kể so với phương pháp TEEN. Kết quả này cho thấy mô hình có khả năng thích nghi tốt với các dạng lỗi hiếm trong điều kiện dữ liệu hạn chế, đồng thời chứng minh tính khả thi của phương pháp khi triển khai trên dữ liệu UAV thực tế.

Đáng chú ý, mức cải thiện của AAPID trên bộ dữ liệu VPI-Fault-2500 lớn hơn đáng kể so với CIFAR100. Điều này cho thấy cơ chế prompting thích nghi được đề xuất đặc biệt phù hợp với các bài toán công nghiệp, nơi dữ liệu thường có tính chuyên biệt cao và số lượng mẫu lỗi mới rất hạn chế.

Nguyên nhân chính của sự cải thiện này có thể được lý giải bởi ba yếu tố. Thứ nhất, cơ chế prompting thích nghi cho phép mô hình tận dụng hiệu quả tri thức đã học từ mô hình Vision Transformer được huấn luyện trước trên tập dữ liệu lớn. Thứ hai, việc sử dụng Information Bottleneck Loss giúp loại bỏ các thông tin nhiễu từ nền ảnh UAV, từ đó tăng khả năng khái quát hóa của mô hình. Thứ ba, Anchor Loss giúp hình thành các cụm đặc trưng rõ ràng trong không gian biểu diễn, qua đó cải thiện khả năng phân biệt giữa các loại lỗi khác nhau.

Những kết quả này cho thấy phương pháp AAPID không chỉ phù hợp với các thiết lập thử nghiệm trong môi trường nghiên cứu mà còn có tiềm năng ứng dụng cao trong các hệ thống giám sát lưới điện thực tế.

Nghiên cứu loại bỏ-Ablation Study (RQ2)

Nhằm làm rõ vai trò của từng thành phần trong mô hình đề xuất, luận án tiến hành nghiên cứu loại bỏ (ablation study) bằng cách lần lượt loại bỏ các mô-đun cấu thành của AAPID. Kết quả thử nghiệm được tổng hợp trong Bảng 4.3. Kết quả cho thấy, việc loại bỏ bất kỳ thành phần nào cũng đều dẫn đến sự suy giảm hiệu năng theo thước đo A_{avg} trên cả hai tập dữ liệu. Điều này khẳng định tính hợp lý của thiết kế mô hình, đồng thời cho thấy các thành phần trong AAPID có mối quan hệ bổ trợ lẫn nhau. Đặc biệt, trong trường hợp loại bỏ mô-đun Task-result Prompts (w/o TRP), mức suy giảm hiệu năng là lớn nhất, qua đó nhấn mạnh vai trò then chốt của thành phần này trong việc học các lớp mới.

Bảng 4.3. Kết quả nghiên cứu loại bỏ theo độ chính xác trung bình (Average Accuracy $\$A_{avg}\$$) (%)

Phương pháp	CIFAR100	VPI-Fault-2500
AAPID (Full model)	89.5	94.2
w/o Task-init Prompts (TIP)	88.9	93.6
w/o Task-result Prompts (TRP)	87.8	92.1
w/o Anchor Loss	88.7	93.4
w/o IB Loss	88.2	92.9

Nghiên cứu chất lượng (RQ3)

Trả lời cho câu hỏi thứ ba của nghiên cứu (RQ3), ”Phương pháp AAPID có học được một không gian đặc trưng trong đó các lớp lỗi khác nhau (bao gồm cả các lớp cơ sở và các lớp few-shot) được tách biệt rõ ràng hay không?” kỹ thuật t-SNE được sử dụng giảm chiều và trực quan hóa các đặc trưng được trích xuất từ các mẫu kiểm tra (Query Set) trên tập dữ liệu VPI-Fault-2500. Hình 4.2 trình bày sự so sánh trực quan giữa không gian đặc trưng được học bởi phương pháp cơ sở (baseline) và phương pháp AAPID được đề xuất.”



Hình 4.2. Trực quan hóa t-SNE kết quả nghiên cứu loại bỏ trên bộ dữ liệu VPI-Fault-2500.

Không gian đặc trưng của phương pháp cơ sở (cột bên trái): Phương pháp cơ sở tạo ra một không gian đặc trưng rối loạn và kém hiệu quả. Trong biểu đồ của tập Query Set, các điểm dữ liệu thuộc cả năm lớp (được biểu diễn bằng năm màu khác nhau) bị trộn lẫn dày đặc và không thể phân biệt, tạo thành một cụm lớn ở trung tâm. Không quan sát thấy bất kỳ cấu trúc hay sự phân tách rõ ràng nào giữa các lớp. Điều này cho thấy mô hình cơ sở hoàn toàn không học được các đặc trưng có khả năng phân biệt, dẫn đến việc không thể phân biệt giữa các lớp cơ sở và các lớp mới trong bối cảnh few-shot. Kết quả này minh họa rõ ràng hiện tượng quên thảm họa (catastrophic forgetting) cũng như hạn chế của phương pháp cơ sở trong việc học hiệu quả từ dữ liệu hạn chế.

Không gian đặc trưng của AAPID (cột bên phải): Trái ngược hoàn toàn với phương pháp cơ sở, phương pháp AAPID tạo ra một không gian đặc trưng có cấu trúc rõ ràng và được tổ chức tốt. Trong biểu đồ của tập Query Set, năm lớp khác nhau được phân tách rõ rệt thành năm cụm riêng biệt và nhất quán. Mỗi cụm có mật độ cao, đồng nhất về màu sắc và duy trì khoảng cách biên rõ ràng so với các cụm còn lại. Kết quả này cho thấy AAPID không chỉ bảo toàn được tri thức của các lớp đã học trước đó mà còn học hiệu quả các đặc trưng đặc thù của các lớp mới chỉ với một số lượng rất hạn chế mẫu huấn luyện, đồng thời tổ chức chúng một cách có hệ thống trong không gian đặc trưng.

Phân tích trực quan này khẳng định mạnh mẽ câu hỏi nghiên cứu RQ3, cho thấy AAPID thực sự học được một không gian đặc trưng vượt trội, trong đó các lớp

được biểu diễn một cách tách biệt với mức độ phân ly cao. Chất lượng của không gian đặc trưng này là nền tảng quan trọng cho các kết quả định lượng ẩn tượng đã được báo cáo trong các bảng kết quả trước đó.

Thảo luận

Các kết quả thử nghiệm cho thấy AAPID đạt hiệu năng vượt trội trên cả bộ dữ liệu chuẩn và bộ dữ liệu thực tế.

Khả năng duy trì tri thức cũ, thích nghi với các lớp lỗi mới và xây dựng không gian đặc trưng có tính phân biệt cao là những yếu tố chính giúp phương pháp đạt được hiệu quả vượt trội.

4.4 Kết luận chương 4

Chương 4 của luận án đã tập trung phân tích bài toán mất cân bằng dữ liệu trong phát hiện và nhận diện bất thường trên chuỗi cách điện từ ảnh thu nhận bằng UAV – một thách thức phổ biến trong các hệ thống giám sát lưới điện tự động. Trong thực tế vận hành, các trạng thái bất thường thường xuất hiện với tần suất thấp và phân bố không đồng đều so với trạng thái bình thường, dẫn đến hiện tượng mất cân bằng dữ liệu nghiêm trọng. Điều này làm suy giảm khả năng học của các mô hình học sâu truyền thống, đặc biệt đối với các lớp thiểu số và các dạng lỗi hiếm.

Trên cơ sở phân tích các đặc điểm của dữ liệu UAV và các hạn chế của những phương pháp xử lý mất cân bằng dữ liệu hiện có, chương này đã trình bày tổng quan các hướng tiếp cận phổ biến bao gồm phương pháp ở mức dữ liệu, phương pháp ở mức thuật toán và các phương pháp tổ hợp. Mặc dù các phương pháp này có thể cải thiện phần nào hiệu năng mô hình, chúng vẫn gặp nhiều hạn chế khi áp dụng cho bài toán phát hiện bất thường nhỏ trong ảnh UAV với nền phức tạp và số lượng mẫu lỗi hạn chế.

Để khắc phục những hạn chế này, luận án đã đề xuất mô hình Attention-aware Adaptive Prompting for Insulator Diagnosis (AAPID) dựa trên khung học Few-Shot Class-Incremental Learning (FSCIL). Phương pháp đề xuất khai thác khả năng biểu diễn mạnh mẽ của mô hình Vision Transformer đã được huấn luyện trước và kết hợp với cơ chế prompting thích nghi nhằm học hiệu quả các đặc trưng bất thường ngay cả khi số lượng mẫu huấn luyện rất hạn chế. Kiến trúc AAPID bao gồm hai loại prompt chính: prompt bất biến theo tác vụ nhằm học các đặc trưng chung của chuỗi cách điện và prompt đặc thù theo tác vụ được sinh ra một cách thích nghi từ dữ liệu đầu vào. Ngoài ra, mô hình còn tích hợp hai thành phần hàm mất mát quan trọng là Information Bottleneck Loss nhằm tăng khả năng khái quát hóa của prompt và Anchor Loss nhằm

nâng cao khả năng phân biệt trong không gian đặc trưng.

Các thực nghiệm được tiến hành trên hai bộ dữ liệu gồm bộ dữ liệu chuẩn CIFAR100 và bộ dữ liệu thực tế VPI-Fault-2500 thu nhận từ hệ thống UAV giám sát lưới truyền tải điện tại Việt Nam. Kết quả thực nghiệm cho thấy phương pháp AAPID đạt hiệu năng vượt trội so với các phương pháp học tăng dần và học ít mẫu hiện nay trên cả hai bộ dữ liệu. Cụ thể, mô hình đạt độ chính xác trung bình cao hơn, mức suy giảm hiệu năng thấp hơn và độ chính xác điều hòa tốt hơn, cho thấy khả năng duy trì tri thức cũ đồng thời học hiệu quả các lớp lỗi mới với số lượng mẫu rất hạn chế. Kết quả nghiên cứu loại bỏ cũng chứng minh vai trò quan trọng của từng thành phần trong kiến trúc đề xuất, đặc biệt là mô-đun prompt đặc thù theo tác vụ. Bên cạnh đó, phân tích trực quan hóa không gian đặc trưng bằng t-SNE cho thấy phương pháp AAPID học được các biểu diễn đặc trưng có cấu trúc rõ ràng và khả năng phân tách tốt giữa các lớp lỗi.

Nhìn chung, các kết quả đạt được trong chương này khẳng định tính hiệu quả của phương pháp đề xuất trong việc giải quyết bài toán mất cân bằng dữ liệu và học tăng dần trong điều kiện số lượng mẫu hạn chế. Phương pháp AAPID không chỉ cải thiện hiệu năng nhận diện bất thường trên chuỗi cách điện mà còn mở ra một hướng tiếp cận tiềm năng cho các bài toán giám sát thiết bị điện dựa trên ảnh UAV trong thực tế vận hành hệ thống điện.

KẾT LUẬN VÀ KIẾN NGHỊ

Những kết quả nghiên cứu của luận án

Mục tiêu của luận án là nghiên cứu kỹ thuật học sâu nhằm đề xuất giải pháp nâng cao chất lượng phát hiện thiết bị trên lưới truyền tải điện và cảnh báo với thiết bị điện lỗi từ tập ảnh thu nhận bằng UAV. Sau quá trình nghiên cứu, luận án đạt được một số đóng góp khoa học và thực tiễn chính như sau:

- Thứ nhất, đề xuất mô hình HVE-YOLOv11 nhằm phát hiện sáu loại thiết bị đặc thù trên lưới truyền tải điện từ ảnh UAV thực tế, đồng thời đánh giá hiệu quả trên dữ liệu thu thập thực nghiệm.
- Thứ hai, xây dựng khung học sâu tích hợp cho bài toán phát hiện và nhận diện bất thường trên chuỗi cách điện, kết hợp backbone GELAN, mô hình khuếch tán (diffusion) và kiến trúc hai giai đoạn phát hiện–phân đoạn, giúp nâng cao khả năng xử lý trong điều kiện nền phức tạp và đối tượng nhỏ.
- Thứ ba, đề xuất chiến lược tăng cường dữ liệu dựa trên Attention-Aware Prompting, cho phép sinh và chọn lọc dữ liệu theo vùng chú ý, qua đó giảm ảnh hưởng của mất cân bằng dữ liệu giữa trạng thái bình thường và bất thường.

Bên cạnh đó, luận án tiến hành đánh giá thực nghiệm toàn diện thông qua so sánh với các phương pháp hiện đại, nghiên cứu ablation và phân tích định tính, qua đó khẳng định tính hiệu quả, độ ổn định và khả năng ứng dụng thực tiễn của các phương pháp đề xuất trong giám sát lưới điện bằng UAV.

Ý nghĩa và khả năng ứng dụng vào thực tiễn

Để tiếp cận và thực nghiệm, luận án xây dựng được quy trình thu thập, tiền xử lý, gán nhãn và tổ chức bộ dữ liệu ảnh UAV cho lưới truyền tải điện 110kV tại Việt Nam. Bộ dữ liệu này được sử dụng không chỉ để đánh giá mô hình đề xuất mà còn có giá trị tham khảo cho các nghiên cứu liên quan.

Những vấn đề còn hạn chế của luận án

Mặc dù đạt được các kết quả nghiên cứu tích cực, luận án vẫn còn tồn tại một số hạn chế cần tiếp tục nghiên cứu và cải thiện gồm:

Thứ nhất, hiệu năng của mô hình đề xuất phụ thuộc đáng kể vào chất lượng ảnh được UAV thu nhận. Trong các điều kiện khách quan bất lợi như thời tiết xấu, ánh sáng thay đổi đột ngột, yếu sẽ ảnh hưởng tới kết quả phát hiện và nhận diện bất thường trên thiết bị. Đây cũng chính là điểm hạn chế chung của các phương pháp dựa trên ảnh quang học.

Thứ hai, luận án tập trung vào bài toán phát hiện và nhận diện bất thường trên ảnh UAV, dữ liệu đơn nguồn. Luận án chưa khai thác đồng thời các nguồn dữ liệu bổ sung như ảnh nhiệt, dữ liệu LiDAR (Light Detection And Ranging) để tăng thêm khả năng nhận biết sớm các dạng bất thường tiềm ẩn của thiết bị.

Thứ ba, hiện tại số lượng mẫu bất thường trong tập dữ liệu huấn luyện còn hạn chế, điều này có thể ảnh hưởng tới khả năng tổng quát hóa của mô hình đối với một số dạng bất thường hiếm, ít gặp trong thực tế.

Hướng nghiên cứu tiếp theo

Từ những kết quả nghiên cứu đã được thực hiện và các hạn chế đã được chỉ ra, nghiên cứu sinh đề xuất một số hướng nghiên cứu mở rộng như sau:

Thứ nhất, **tích hợp dữ liệu đa nguồn và đa phương thức nhằm nâng cao độ tin cậy của hệ thống trong phát hiện và nhận diện bất thường**. Trong luận án, các mô hình được xây dựng chủ yếu dựa trên dữ liệu ảnh quang học thu nhận từ UAV. Mặc dù dữ liệu ảnh quang học cung cấp nhiều thông tin về hình dạng, màu sắc và trạng thái bề mặt của thiết bị, hiệu quả phát hiện có thể bị ảnh hưởng đáng kể trong các điều kiện môi trường bất lợi như sương mù, mưa lớn, khói bụi, ánh sáng yếu hoặc hiện tượng ngược sáng. Bên cạnh đó, nhiều dạng hư hỏng tiềm ẩn bên trong thiết bị điện thường chưa biểu hiện rõ ràng trên bề mặt và khó có thể được phát hiện chỉ bằng ảnh quang học.

Do đó, một hướng nghiên cứu đầy tiềm năng là kết hợp dữ liệu từ nhiều cảm biến khác nhau như ảnh nhiệt hồng ngoại, dữ liệu LiDAR, cảm biến corona hoặc cảm biến siêu âm để xây dựng các mô hình hợp nhất dữ liệu đa phương thức (multimodal fusion). Trong đó, ảnh nhiệt có khả năng phản ánh các bất thường về phân bố nhiệt và hiện tượng phát nóng cục bộ; dữ liệu LiDAR cung cấp thông tin hình học ba chiều với độ chính xác cao; còn ảnh quang học cung cấp các đặc trưng trực quan về trạng thái bề mặt thiết bị. Sự kết hợp giữa các nguồn dữ liệu này có thể giúp khắc phục các điểm mù do điều kiện thời tiết bất lợi hoặc ánh sáng không thuận lợi, đồng thời hỗ trợ phát hiện sớm các hư hỏng tiềm ẩn bên trong thiết bị trước khi xuất hiện các dấu hiệu hư hỏng quan sát được bằng mắt thường. Đây được xem là một trong những hướng nghiên cứu quan trọng nhằm xây dựng các hệ thống giám sát lưới điện thông minh có độ tin cậy cao trong tương lai.

Thứ hai, **phát triển các mô hình nhẹ và tối ưu hóa cho việc triển khai thời gian thực**. Mặc dù các mô hình được đề xuất trong luận án đạt hiệu năng cao trong bài toán phát hiện bất thường trên chuỗi cách điện, chi phí tính toán vẫn là một thách thức đáng kể khi triển khai trên các nền tảng phần cứng có tài nguyên hạn chế như thiết bị nhúng hoặc thiết bị bay không người lái. Do đó, các nghiên cứu trong tương lai có thể tập trung vào các kỹ thuật tối ưu hóa mô hình như nén mô hình, lượng tử

hóa tham số, chưng cất tri thức hoặc thiết kế các kiến trúc mạng nơ-ron gọn nhẹ. Các hướng tiếp cận này nhằm giảm yêu cầu về tài nguyên tính toán và bộ nhớ trong khi vẫn duy trì độ chính xác ở mức chấp nhận được, qua đó đáp ứng yêu cầu xử lý thời gian thực trong các hệ thống giám sát lưới điện quy mô lớn.

Thứ ba, **mở rộng bài toán phát hiện bất thường sang dự báo sớm sự cố thiết bị**. Các nghiên cứu trong luận án hiện chủ yếu tập trung vào việc phát hiện và nhận diện bất thường trên chuỗi cách điện dựa trên dữ liệu ảnh quang học. Trong tương lai, việc tích hợp dữ liệu đa phương thức có thể mở rộng đáng kể phạm vi các dạng bất thường có khả năng nhận biết, bao gồm hiện tượng phát nhiệt bất thường, phóng điện cục bộ, rò rỉ điện và suy giảm chất lượng vật liệu cách điện. Đặc biệt, dữ liệu ảnh nhiệt và các cảm biến chuyên dụng có khả năng phản ánh những dấu hiệu suy giảm từ giai đoạn rất sớm, khi các biểu hiện hư hỏng bên ngoài chưa xuất hiện hoặc chưa thể quan sát trực tiếp bằng ảnh quang học.

Bên cạnh đó, việc khai thác dữ liệu quan sát theo thời gian kết hợp với các mô hình học sâu chuyên xử lý dữ liệu chuỗi, chẳng hạn như mạng nhớ ngắn hạn dài hạn, mạng biến đổi và mạng tích chập theo thời gian, có thể hỗ trợ dự báo xu hướng suy giảm chất lượng thiết bị cũng như ước lượng nguy cơ xảy ra sự cố trong tương lai. Hướng nghiên cứu này cho phép chuyển đổi từ chiến lược bảo trì phản ứng sang bảo trì dự báo, góp phần nâng cao độ tin cậy, tính an toàn và hiệu quả vận hành của hệ thống điện.

Thứ tư, **nghiên cứu khả năng thích nghi và tổng quát hóa mô hình trên các loại thiết bị, điện áp lưới và điều kiện vận hành khác nhau**. Các nghiên cứu tiếp theo có thể mở rộng phạm vi áp dụng của mô hình sang nhiều loại thiết bị khác trên lưới truyền tải như dây dẫn, phụ kiện đường dây, tạ chống rung, cột điện, máy biến áp hoặc các thiết bị tại trạm biến áp. Đồng thời, cần đánh giá khả năng thích nghi của mô hình trên các hệ thống điện có cấp điện áp khác nhau như 110kV, 220kV và 500kV cũng như trong các điều kiện địa hình, khí hậu và môi trường vận hành đa dạng hơn. Việc nghiên cứu các kỹ thuật học chuyển giao (transfer learning), học thích nghi miền (domain adaptation) hoặc học liên tục (continual learning) sẽ góp phần nâng cao khả năng tổng quát hóa của mô hình trong thực tế.

Cuối cùng, xây dựng hệ thống hỗ trợ ra quyết định thông minh phục vụ công tác quản lý, vận hành và giám sát toàn diện lưới điện. Trên cơ sở các mô hình phát hiện và nhận diện bất thường đã được đề xuất trong luận án, các nghiên cứu tương lai có thể hướng tới phát triển hệ thống hỗ trợ ra quyết định tích hợp, kết hợp kết quả phân tích từ nhiều nguồn dữ liệu khác nhau như ảnh quang học, ảnh nhiệt, dữ liệu LiDAR, dữ liệu vận hành SCADA và tri thức chuyên gia. Hệ thống không chỉ dừng lại ở việc phát hiện bất thường mà còn có khả năng đánh giá mức độ rủi ro, xác

định nguyên nhân tiềm ẩn, dự báo xu hướng phát triển của hư hỏng và đề xuất phương án xử lý phù hợp.

Đây được xem là bước phát triển quan trọng hướng tới xây dựng lưới điện số và lưới điện thông minh, trong đó dữ liệu đa phương thức và trí tuệ nhân tạo đóng vai trò trung tâm trong việc nâng cao mức độ tự động hóa, tối ưu hóa công tác bảo trì và bảo đảm vận hành an toàn, tin cậy cho hệ thống điện.

CÔNG TRÌNH CÔNG BỐ

- CT1 **Trịnh Hiền Anh**, Nguyễn Thị Thanh Tân, Bùi Xuân Hùng, Đoàn Ngọc Duy, 2022, “Phát hiện thiết bị trên lưới truyền tải điện cao thế 110kV sử dụng kỹ thuật học sâu”, Hội thảo Quốc gia lần thứ XXV: Một số vấn đề chọn lọc của công nghệ thông tin và truyền thông, Hà Nội, ngày 8–9/12/2022, tr. 174–179.
- CT2 **Anh Trinh Hien**, A. D. Tran, D. C. Viet, Q. D. T. Thuy và Q. N. Huu, 2025, “Insulator Defect Detection Based on Feature Pyramid Network and Diffusion Model”, *Signal, Image and Video Processing*, Vol. 19, Issue 5, pp. 372–384 [SCIE, Q3].
- CT3 **Anh Trinh Hien**, Tao Ngo Quoc, Thanh-Tan Nguyen Thi, 2025, “Attention-Aware Prompting for Learning New Faults on Insulators with Limited Data”, *Journal of Information Hiding and Multimedia Signal Processing*, Vol. 16, No. 4, December 2025, pp. 1199–1208 [SCOPUS, Q4].
- CT4 **Trịnh Hiền Anh**, Nguyễn Thị Thanh Tân, 2023, “Một phương pháp hiệu quả để phát hiện đường dây, cột điện và các thiết bị trên lưới truyền tải điện cao thế 110kV bằng kỹ thuật học sâu”, *Tạp chí Khoa học và Công nghệ Năng lượng – Trường Đại học Điện lực*, Số 32, ISSN: 1859-4557, tr. 1–11.
- CT5 **Trịnh Hiền Anh**, Nguyễn Thị Thanh Tân, 2023, “Một kỹ thuật tăng cường chất lượng ảnh UAV thu nhận trên hệ thống truyền tải điện cao thế 110kV”, Hội thảo Quốc gia lần thứ XXVI: Một số vấn đề chọn lọc của công nghệ thông tin và truyền thông, Bắc Ninh, ngày 5–6/10/2023, tr. 206–209.
- CT6 Thi-Thanh-Tan-Nguyen, **Anh Trinh Hien**, Tao Ngo Quoc, 2026, ”HVE-YOLO: A multi-scale neural network integrated for high-voltage power grid equipment recognition”, under review

TÀI LIỆU THAM KHẢO

- [1] Zhaoyun Zhang, Delong Cai, and Zhi Zhang. Review on online operation insulator identification and fault diagnosis based on uav patrol images and deep learning algorithms. *Frontiers in Energy Research*, 2022.
- [2] Baoyu Xu, Yilin Zhao, Tao Wang, and Qingning Chen. Development of power transmission line detection technology based on unmanned aerial vehicle image vision. *SN Applied Sciences*, 5:72, 2023.
- [3] Esra Ince. Transmission line inspection with uavs: A review of technologies, applications, and challenges. *International Journal of Energy and Smart Grid*, 9:60–69, 2024.
- [4] Rui Jiang, Dong Wu, and Xiaoqiang Li. A review of uav-based applications for power line inspection. *Electric Power Systems Research*, 189:106795, 2020.
- [5] Bongumsa Mendu and Nhlanhla Mbuli. State-of-the-art review on the application of unmanned aerial vehicles (uavs) in power line inspections: Current innovations, trends, and future prospects. *Drones*, 9(4), 2025.
- [6] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems*, volume 28, 2015.
- [7] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, pages 2980–2988, 2017.
- [8] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *CVPR*, 2016.
- [9] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *European Conference on Computer Vision (ECCV)*, pages 213–229. Springer, 2020.
- [10] Tiago Santos, Tiago Cunha, André Dias, António Paulo Moreira, and José Almeida. Uav visual and thermographic power line detection using deep learning. *Sensors*, 24(17), 2024.

- [11] Y Tian et al. Deep learning-based insulator detection for uav inspection systems: A review. *IEEE Access*, 8:200055–200075, 2020.
- [12] Jaka Katrasnik, Franjo Pernuš, and Boštjan Likar. A survey of mobile robots for distribution power line inspection. *IEEE Transactions on Power Delivery*, 25(1):485–493, 2010.
- [13] S Ahmed et al. Uav-based inspection techniques for transmission lines: A review. *Journal of Intelligent & Robotic Systems*, 111(2):1–28, 2024.
- [14] M Faisal et al. Deep learning and uav-based inspection for smart transmission grids: Recent advances and challenges. *Applied Energy*, 376:124097, 2025.
- [15] Robert Walton. Ameren’s latest drone use demonstrates future of transmission line monitoring. *Utility Dive*, December 2018.
- [16] B. Wang et al. Deep learning for power line inspection: A survey. *IEEE Transactions on Power Delivery*, 2021.
- [17] Electricity Generating Authority of Thailand (EGAT). Uav inspection supports power system reliability. Technical report, Electricity Generating Authority of Thailand, 2024.
- [18] John Smith and Thomas Brown. Drone-based inspection of power transmission lines in the united states. *IEEE Power and Energy Magazine*, 20(3):45–53, 2022.
- [19] Miguel Garcia and Luca Rossi. Uav applications for power line inspection in europe. *Renewable and Sustainable Energy Reviews*, 145:111097, 2021.
- [20] Hiroshi Tanaka and Kenji Yamamoto. Application of uav and lidar for power line inspection in japan. *IEEE Access*, 10:52345–52354, 2022.
- [21] Yong Li and Qiang Zhang. Deep learning-based uav inspection system for power transmission lines. *Electric Power Systems Research*, 220:109315, 2023.
- [22] Peter Wilson and David Clarke. Thermal uav inspection for power grid maintenance in australia. *Energy Reports*, 8:1234–1242, 2022.
- [23] Ahmad Rahman and Siew Lim. Uav-based inspection systems for power grid maintenance in asean countries. *Energy Reports*, 9:3456–3465, 2023.

- [24] Huu Thanh Nguyen and Minh Quang Tran. Application of uav technology for transmission line inspection in vietnam. *Journal of Electrical Engineering and Technology*, 18(5):2345–2354, 2023.
- [25] Hao Zhang and Yong Liu. Intelligent uav inspection system for smart grid based on edge ai. *IEEE Access*, 12:34567–34578, 2024.
- [26] Sung-Ho Kim and Joon Park. Ai-based uav inspection system for power grid maintenance in korea. *Energy Reports*, 10:5678–5686, 2024.
- [27] Luca Rossi and Thomas Meyer. Integration of uav and digital twin technology for smart grid inspection. *Renewable Energy*, 230:120345, 2025.
- [28] Qiang Li and Ming Zhao. Swarm uav-based inspection system for large-scale power transmission networks. *IEEE Transactions on Power Delivery*, 2025.
- [29] EVNNPT. Ứng dụng thiết bị bay không người lái trong kiểm tra đường dây truyền tải điện, 2021.
- [30] Tổng công ty Điện lực miền Bắc (EVNNPC). Ứng dụng uav trong kiểm tra và giám sát vận hành lưới điện. Báo cáo kỹ thuật, 2025.
- [31] Konstantinos Kanistras, Goncalo Martins, Matthew J. Rutherford, and Kimon P. Valavanis. A survey of unmanned aerial vehicles (uavs) for traffic monitoring. In *2013 International Conference on Unmanned Aircraft Systems (ICUAS)*, pages 221–234, 2013.
- [32] Mozhddeh Shahbazi, Jérôme Théau, and Patrick Menard. Recent applications of unmanned aerial imagery in natural resource management. *GIScience & Remote Sensing*, 51:1–27, 06 2014.
- [33] Hazim Shakhatreh, Ahmad H. Sawalmeh, Ala Al-Fuqaha, Zuochao Dou, Eyad Almaita, Issa Khalil, Noor Shamsiah Othman, Abdallah Khreishah, and Mohsen Guizani. Unmanned aerial vehicles (uavs): A survey on civil applications and key research challenges. *IEEE Access*, 7:48572–48634, 2019.
- [34] Doina Bein, Wolfgang Bein, Ashish Karki, and Bharat B. Madan. Optimizing border patrol operations using unmanned aerial vehicles. In *2015 12th International Conference on Information Technology - New Generations*, pages 479–484, 2015.

- [35] Y.A. Pederi and H.S. Cheporniuk. Unmanned aerial vehicles and new technological methods of monitoring and crop protection in precision agriculture. In *2015 IEEE International Conference Actual Problems of Unmanned Aerial Vehicles Developments (APUAVD)*, pages 298–301, 2015.
- [36] Bongumsa Mendu and Nhlanhla Mbuli. State-of-the-art review on the application of unmanned aerial vehicles (uavs) in power line inspections: Current innovations, trends, and future prospects. *Drones*, 9(4), 2025.
- [37] Quang Thắng. Trao đổi kinh nghiệm điều khiển và xử lý dữ liệu thiết bị bay uav phục vụ quản lý vận hành lưới điện truyền tải, 2023. Nguồn: Website Tổng công ty Truyền tải điện Quốc gia (NPT), truy cập 06/12/2025.
- [38] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 1*, NIPS’12, page 1097–1105, Red Hook, NY, USA, 2012. Curran Associates Inc.
- [39] T. N. Karadeniz, S. Ekici, and E. Avci. Deep learning-based detection of power transmission lines using yolov4 and yolov8. *Scientific Reports*, 16:2382, 2026.
- [40] International Electrotechnical Commission. Insulation co-ordination – part 1: Definitions, principles and rules. Technical Report IEC 60071-1, International Electrotechnical Commission, Geneva, Switzerland, 2019. Edition 9.0.
- [41] Institute of Electrical and Electronics Engineers. Ieee guide for rf protection of personnel working in the vicinity of wireless communications antennas on electric utility structures. Technical Report IEEE Std 1654-2017, IEEE, New York, NY, USA, 2017.
- [42] Tổng công ty Truyền tải điện Quốc gia (EVNNPT). *Quy trình tạm thời vận hành thiết bị bay không người lái kiểm tra đường dây truyền tải điện*. Tập đoàn Điện lực Việt Nam (EVN), Hà Nội, Việt Nam, 2020. Lưu hành nội bộ.
- [43] I. Colomina and P. Molina. Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 92:79–97, 2014.
- [44] Tsung-Yi Lin et al. Feature pyramid networks for object detection. In *CVPR*, 2017.

- [45] Guangyi Tang, Jianjun Ni, Yonghao Zhao, Yang Gu, and Weidong Cao. A survey of object detection for uavs based on deep learning. *Remote Sensing*, 16(1):149, 2024.
- [46] C. Zhang and J. M. Kovacs. The application of small unmanned aerial systems for precision agriculture: A review. *Precision Agriculture*, 13(6):693–712, 2012.
- [47] Qihan Feng, Xinzheng Xu, and Zhixiao Wang. Deep learning-based small object detection: A survey. *Mathematical Biosciences and Engineering*, 20(4):6551–6590, 2023.
- [48] K. Tong, Y. Wu, and F. Zhou. Recent advances in small object detection based on deep learning: A review. *Image and Vision Computing*, 97:103910, 2020.
- [49] X. Liu, Y. Zhang, Z. Liu, and J. Wang. Deep learning-based inspection of power transmission components using uav imagery. *IEEE Access*, 8:102746–102760, 2020.
- [50] Zhengxia Zou, Keyan Chen, Zhenwei Shi, Yuhong Guo, and Jieping Ye. Object detection in 20 years: A survey. *Proceedings of the IEEE*, 111(3):257–276, 2023.
- [51] Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu, and Xindong Wu. Object detection with deep learning: A review. *IEEE Transactions on Neural Networks and Learning Systems*, 30(11):3212–3232, 2019.
- [52] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. Ssd: Single shot multibox detector. *arXiv preprint arXiv:1512.02325*, 2015.
- [53] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.
- [54] Ross Girshick. Fast r-cnn. In *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, pages 1440–1448, 2015.
- [55] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 2961–2969, 2017.

- [56] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 2999–3007, 2017.
- [57] Glenn Jocher. Yolov5 by ultralytics. *GitHub*, 2020.
- [58] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *CVPR*, 2023.
- [59] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics yolov8, 2023.
- [60] Xizhou Zhu et al. Deformable detr: Deformable transformers for end-to-end object detection. *ICLR*, 2021.
- [61] Y. Liu et al. A survey on uav-based object detection and applications. *Remote Sensing*, 2024.
- [62] Q. Zhou et al. Deep learning-based object detection: A review of recent advances. *Pattern Recognition*, 2023.
- [63] Y. Liu and X. Wang. Review of deep learning methods for power line inspection. *IEEE Access*, 2023.
- [64] Yaqi Liu, Kaizhu Xu, Jiakun Xu, and Zhijun Wang. A deep learning approach for insulator instance segmentation and defect detection. *Neural Computing and Applications*, 34(9):7253–7269, 2022.
- [65] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*, 2020.
- [66] Yong Liu et al. Dsmh-yolov4-based uav insulator detection method in complex backgrounds. *Electronics*, 12(4):933, 2023.
- [67] Wei Zhang et al. Lightweight yolov4 for real-time insulator defect detection from uav images. *Array*, 12:100084, 2021.
- [68] Ultralytics. Yolov5 implementation and improvements. *GitHub Repository*, 2020.
- [69] J. Wang et al. Yolov8: A comprehensive review and analysis. *arXiv preprint*, 2023.

- [70] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10012–10022, 2021.
- [71] Xiang Zhu, Shilin Lyu, Xu Wang, and Qi Zhao. Tph-yolov5: Improved yolov5 based on transformer prediction head for object detection on drone-captured scenarios. *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pages 2778–2788, 2021.
- [72] Hao Wang, Peng Liu, Rui Zhang, and Xiaodong Chen. Transeffidet: Transformer-based efficientdet for uav insulator detection in complex environments. *Sensors*, 22(18):7045, 2022.
- [73] Jian Li, Ming Zhou, Tao Xu, and Yifan Huang. Hybrid cnn–transformer network for uav-based power line component detection. *IEEE Access*, 11:45672–45685, 2023.
- [74] D. Elreedy and A. Atiya. A comprehensive analysis of synthetic minority oversampling technique (smote) for handling class imbalance. *Information Sciences*, 505:32–64, 2022.
- [75] N. Micheletti et al. Synthetic data generation for imbalanced learning: A comprehensive review. *Machine Learning with Applications*, 21:100650, 2025.
- [76] F.-A. Croitoru et al. Diffusion models in vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.
- [77] Z. Wu et al. Diffusion models for image generation and data augmentation: Recent advances and applications. *IEEE Access*, 13:1–28, 2025.
- [78] H. Zhang, Y. Wang, and J. Liu. Super-resolution-based small object detection for uav inspection of power transmission lines. *IEEE Transactions on Power Delivery*, 38(5):3125–3136, 2023.
- [79] V. D. Tran, T. H. Nguyen, and Q. B. Le. Yolo-sr: Enhancing small insulator defect detection via lightweight super-resolution and attention fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 2341–2350, 2024.
- [80] Yue Zhao, Qiang Li, Zhen Zhang, Sheng Wang, and Xiaohui Li. Saliency-guided attention network for small object detection in uav-based power line inspection. *IEEE Trans. Instrum. Meas.*, 73:1–12, 2024.

- [81] Long Wen, Liang Gao, and Xinyu Li. Deep learning approaches on defect detection in high resolution aerial images of insulators. *Sensors*, 21(4):1033, 2021.
- [82] André Luiz Buarque Vieira-e Silva, Heitor de Castro Felix, Thiago de Menezes Chaves, Francisco Paulo Magalhães Simões, Veronica Teichrieb, Michel Mozinho dos Santos, Hemir da Cunha Santiago, Virginia Adélia Cordeiro Sgotti, and Henrique Baptista Duffles Teixeira Lott Neto. Stn-plad: A dataset for multi-size power line assets detection in high-resolution uav images. *arXiv preprint arXiv:2108.07944*, 2021.
- [83] Kai Zhang et al. Rsin-dataset: Uav-based insulator detection dataset and benchmark. *Drones*, 7(2):125, 2023.
- [84] Jian Liu et al. Hybrid-yolo for classification of insulator defects in transmission lines. *International Journal of Electrical Power & Energy Systems*, 147:108835, 2023.
- [85] Daqi Tian, Jinlin Chen, and Xin Wang. Application research of uav infrared diagnosis technology in intelligent inspection of substations. *Energy Informatics*, 7(1):56, 2024.
- [86] Wei Chen et al. A lightweight insulator defect detection model based on drone images. *Drones*, 8(9):431, 2024.
- [87] Miguel Garcia et al. Aerial inspection of high-voltage power lines using yolov8 real-time object detector. *Energies*, 17(11):2535, 2024.
- [88] Ming Zhao et al. Yolo-ela: Efficient local attention modeling for insulator defect detection. *arXiv preprint arXiv:2410.11727*, 2024.
- [89] Zhu-Ye Xu and Xiao Tang. Transmission line insulator defect detection algorithm based on map-yolov8. *Scientific Reports*, 15(1):10288, 2025.
- [90] Lei Zhang et al. Foreign object detection in high-voltage transmission lines based on improved yolov8m. *arXiv preprint arXiv:2502.07175*, 2025.
- [91] Gongjie Zhang, Kaiwen Cui, Rongliang Wu, Shuang Lu, and Ying Tian. Defect-gan: High-fidelity defect synthesis for high-resolution defect inspection. *IEEE Trans. Ind. Informat.*, 18(11):8063–8072, 2022.

- [92] Teng Hu, Jiangning Zhang, Ran Yi, Yabiao Du, Xiaoshuai Xu, Fang Liu, Guannan Gao, and Chengjie Wang. Anomalydiffusion: Few-shot anomaly image generation with diffusion models. In *Proc. AAAI Conf. Artif. Intell.*, volume 38, pages 8526–8534, 2024.
- [93] Golnaz Ghiasi, Yin Cui, Aravind Srinivas, Rui Qian, Tsung-Yi Lin, Ekin D. Cubuk, Quoc V. Le, and Barret Zoph. Simple copy-paste is a strong data augmentation method for instance segmentation. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pages 2918–2928, 2021.
- [94] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, pages 3836–3847, 2023.
- [95] Daoxing Li, Xiaohui Wang, Jie Zhang, and Zhixiang Ji. Automated deep learning system for power line inspection image analysis and processing. *Global Energy Interconnection*, 6:614–633, 2023.
- [96] Amol Doke and Mahesh Gaikwad. Survey on automated machine learning (automl) and meta-learning. In *2021 International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. IEEE, 2021.
- [97] Pratik Doke and Rahul Singh. A survey on automated machine learning for power grid infrastructure monitoring. *Renewable and Sustainable Energy Reviews*, 145:111054, 2021.
- [98] Rahima Khanam and Muhammad Hussain. Yolov11: An overview of the key architectural enhancements. *arXiv preprint arXiv:2410.17725*, 2024.
- [99] Dezhi Wang, Riqing Song, Minghui Liu, et al. Dcdw-yolov11: An intelligent defect-detection method for key transmission-line equipment. *Sensors*, 26(3):1029, 2026.
- [100] Depeng Gao, Yihan Yin, Han Zhang, et al. Yolov11-based uav foreign object detection for power transmission lines. *Electronics*, 14(18):3577, 2025.
- [101] Huma Hafeez, Matthew Garratt, et al. Yolo11-4k: An efficient architecture for real-time small object detection in 4k panoramic images. *arXiv preprint arXiv:2512.16493*, 2025.
- [102] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *arXiv preprint arXiv:1406.2661*, 2014.

- [103] Agastya Kalra, Guy Stoppi, Bradley Brown, Rishav Agarwal, and Achuta Kadambi. Towards rotation invariance in object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.
- [104] C. Y. Wang, H. Y. Mark Liao, Y. H. Wu, P. Y. Chen, J. W. Hsieh, and I. H. Yeh. Cspnet: A new backbone that can enhance learning capability of cnn. In *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, pages 1571–1580, June 2020.
- [105] L. Xie et al. Elan: Towards generic and efficient elastic training for deep learning. In *2020 IEEE 40th International Conference on Distributed Computing Systems (ICDCS)*, pages 78–88, Singapore, 2020.
- [106] H. Gong, H. Li, K. Xu, and Y. Zhang. Object detection based on improved yolov3-tiny. In *2019 Chinese Automation Congress (CAC)*, pages 3240–3245, Hangzhou, China, 2019.
- [107] L. Deng, H. Li, H. Liu, and J. Gu. A lightweight yolov3 algorithm used for safety helmet detection. *Scientific Reports*, 12(1):1–15, 2022.
- [108] X. Liu, R. Huang, J. Cai, X. Yang, and Z. Xie. Feasibility analysis of rt-detr in the context of foreign object intrusion at railway boundaries. In M. Gong, L. Jia, Y. Qin, J. Yang, Z. Liu, and M. An, editors, *Proceedings of the 6th International Conference on Electrical Engineering and Information Technologies for Rail Transportation (EITRT) 2023*, volume 1138 of *Lecture Notes in Electrical Engineering*. Springer, Singapore, 2024.
- [109] X. Dai, Y. Chen, J. Yang, P. Zhang, L. Yuan, and L. Zhang. Dynamic detr: End-to-end object detection with dynamic attention. In *Proc. IEEE Int. Conf. Comput. Vis.*, pages 2968–2977, 2021.
- [110] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao. Yolov9: Learning what you want to learn using programmable gradient information, 2024. Available: <http://arxiv.org/abs/2402.13616>.
- [111] Y. Wang, H. Wang, and Z. Xin. Efficient detection model of steel strip surface defects based on yolo-v7. *IEEE Access*, 10:133936–133944, 2022.
- [112] Y. Liu, D. Liu, X. Huang, and C. Li. Insulator defect detection with deep learning: A survey. *IET Gener. Transm. Distrib.*, 17(16):3541–3558, 2023.

- [113] Connor Shorten and Taghi Khoshgohfar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 2019.
- [114] Hongyi et al. Zhang. mixup: Beyond empirical risk minimization. In *ICLR*, 2018.
- [115] Sangdoo et al. Yun. Cutmix: Regularization strategy to train strong classifiers with localizable features. In *ICCV*, 2019.
- [116] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2016.
- [117] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv*, abs/1411.1784, 2014.
- [118] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4401–4410. IEEE, 2019.
- [119] Diederik Kingma and Max Welling. Auto-encoding variational bayes. *ICLR*, 2014.
- [120] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *NeurIPS*, 2020.
- [121] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, volume 139 of *Proceedings of Machine Learning Research*, pages 8162–8171. PMLR, 2021.
- [122] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10684–10695. IEEE, 2022.
- [123] Thomas et al. Schlegl. Unsupervised anomaly detection with generative adversarial networks. *IPMI*, 2017.
- [124] Samet et al. Akcay. Ganomaly: Semi-supervised anomaly detection via adversarial training. *ACCV*, 2019.

- [125] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations (ICLR)*, 2021.
- [126] Menglin et al. Jia. Visual prompt tuning. In *ECCV*, 2022.
- [127] Zifeng et al. Wang. Learning to prompt for continual learning. In *CVPR*, 2022.
- [128] Zifeng et al. Wang. Dualprompt: Complementary prompting for rehearsal-free continual learning. In *ECCV*, 2022.
- [129] James Seale et al. Smith. Coda-prompt: Continual decomposed attention-based prompting. In *CVPR*, 2023.
- [130] Chenxi Liu, Zhenyi Wang, Tianyi Xiong, Ruibo Chen, Yihan Wu, Junfeng Guo, and Heng Huang. Few-shot class incremental learning with attention-aware self-adaptive prompt. In *European Conference on Computer Vision (ECCV)*, pages 1–18, 2024.
- [131] Naftali Tishby, Fernando Pereira, and William Bialek. The information bottleneck method. *arXiv preprint physics/0004057*, 2000.
- [132] Zhixiang Chi, Jianfeng Wang, Bo Li, Zhen Li, and Bo Zhang. Metafscil: A meta-learning approach for few-shot class incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14140–14149. IEEE, 2022.