

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC MỎ - ĐỊA CHẤT

BÁO CÁO HỌC THUẬT

NĂM HỌC 2025-2026

Học kỳ 1

**ĐỀ XUẤT XÂY DỰNG MÔ HÌNH HỌC SÂU TRÍCH XUẤT DỮ
LIỆU TÒA NHÀ TỪ ẢNH ĐỘ PHÂN GIẢI CAO (TỪ MÁY BAY
KHÔNG NGƯỜI LÁI UAV) ÁP DỤNG TRONG ĐIỀU KIỆN
PHỨC TẠP VÀ ĐA DẠNG Ở VIỆT NAM**

Người báo cáo: TS Phạm Trung Dũng

Đơn vị: Bộ môn Trắc địa Công trình

HÀ NỘI, tháng 12 năm 2025

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC MỎ - ĐỊA CHẤT

BÁO CÁO HỌC THUẬT

NĂM HỌC 2025-2026

Học kỳ 1

**ĐỀ XUẤT XÂY DỰNG MÔ HÌNH HỌC SÂU TRÍCH XUẤT DỮ
LIỆU TÒA NHÀ TỪ ẢNH ĐỘ PHÂN GIẢI CAO (TỪ MÁY BAY
KHÔNG NGƯỜI LÁI UAV) ÁP DỤNG TRONG ĐIỀU KIỆN
PHỨC TẠP VÀ ĐA DẠNG Ở VIỆT NAM**

Phòng KHCN

Bộ môn trắc địa công trình

Người báo cáo

TS. Phạm Trung Dũng

HÀ NỘI, tháng 12 năm 2025

MỤC LỤC

Chương 1	3
XÂY DỰNG QUY TRÌNH TRÍCH XUẤT DỮ LIỆU TỪ ẢNH UAV	3
1.1. Tổng quan.....	3
1.2. Quy trình trích xuất dữ liệu tòa nhà sử dụng mạng học sâu.....	4
2.1. Giai đoạn Chuẩn bị Dữ liệu (Data Preparation)	6
2.2. Giai đoạn Lựa chọn và Huấn luyện Mô hình (Model Training).....	7
2.3. Giai đoạn Đánh giá mô hình (Model Evaluation)	10
2.4. Giai đoạn Hậu xử lý và Vector Hóa (Post-Processing & Vectorization)	14
Chương 2	16
THỰC NGHIỆM XÂY DỰNG QUY TRÌNH TRÍCH XUẤT DỮ LIỆU TÒA NHÀ TRÊN ẢNH UAV TẠI VIỆT NAM.....	16
2.1.1. Thu thập dữ liệu từ thiết bị bay không người lái UAV	16
2.1.2. Tạo mẫu phục vụ quá trình đào tạo sử dụng mạng học sâu.....	20
2.1.3. Tăng cường dữ liệu (Data augmentation)	22
2.1.4. Chia tập dữ liệu.....	25
2.2. Giai đoạn 2- Huấn luyện mô hình	28
2.2.1. Lựa chọn kiến trúc mô hình mạng học sâu.....	28
2.2.2. Lựa chọn backbone.....	30
2.2.3. Huấn luyện mô hình	31
2.3. Giai đoạn 3- Đánh giá mô hình	33
2.3.1. Đo lường hiệu suất mô hình	33
2.3.2. Phân tích kết quả.....	33
2.4. Giai đoạn 4- Hậu xử lý và vector hóa	35
2.4.1. Dự đoán (Prediction)	36
2.4.2. Gán Thông tin Địa lý (Georeferencing)	36
2.4.3. Vector Hóa (Polygonization)	36
2.5. Đề xuất quy trình ứng dụng công nghệ học sâu vào sản xuất bản đồ địa hình và dữ liệu nền địa lý	37
2.6. Đánh giá khả năng thành lập bản đồ và cơ sở dữ liệu nền địa lý sử dụng công nghệ trích xuất tự động AI trên ảnh UAV	41
KẾT LUẬN VÀ KIẾN NGHỊ	44
TÀI LIỆU THAM KHẢO	46

DANH MỤC HÌNH VẼ, ĐỒ THỊ

Hình 1. Quy trình tổng quan trích xuất dữ liệu tòa nhà từ dữ liệu ảnh sử dụng mạng học sâu	5
Hình 2. Giai đoạn 1- Chuẩn bị dữ liệu	7
Hình 3. Giai đoạn 2- Huấn luyện mô hình	9
Hình 4. Giai đoạn 3- Đánh giá hiệu suất mô hình học sâu	14
Hình 5. Giai đoạn 4- Hậu xử lý và vector hóa tòa nhà	15
Hình 6: Sơ đồ quy trình xử lý ảnh UAV	17
Hình 7. Trạm điều khiển mặt đất GNSS (trái); Tiến hành bay UAV (phải)	19
Hình 8. Ảnh trực giao sau khi nhấn trên phần mềm QGIS	20
Hình 9. Các mẫu tiles được chia nhỏ từ một tấm ảnh UAV ortho lớn	21
Hình 10. Các mask tương ứng với các tiles	21
Hình 11. Một mẫu từ tập dữ liệu trong khu vực nghiên cứu trong đó (a)-ảnh trực giao, (b)- là vector của tòa nhà, và (c) là mask được raster hóa tòa nhà	22
Hình 12. Biến đổi hình học	23
Hình 13. Biến đổi phổ ảnh	24
Hình 14. Biến đổi che khuất	24
Hình 15. Mô hình U-net với ảnh đầu vào RGB kích thước 512 x 512 pixel 3 kênh và mask đầu ra kích thước 512 x512 với 1 kênh	29
Hình 16. Kiến trúc mô hình mạng DeepLabV3+(Chen, Zhu et al. 2018).	30
Hình 17. Sơ đồ nối tắt (skip connection) trong ResNet (He, Zhang et al. 2016)	30
Hình 18. So sánh hiệu suất trích xuất tòa nhà ở các khu vực có đặc điểm nhà khác nhau trên bộ Testing Set - HUMG	35
Hình 19: So sánh kết quả dự đoán (đường màu đỏ) và đối tượng thực (màu vàng)	37
Hình 20. Quy trình thành lập bản đồ chuyên đề từ ảnh UAV tích hợp công nghệ học sâu	40

Chương 1

XÂY DỰNG QUY TRÌNH TRÍCH XUẤT DỮ LIỆU TỪ ẢNH UAV

1.1. Tổng quan

Tự động trích xuất và lập bản đồ các đối tượng quan trọng trên bề mặt Trái Đất, như tòa nhà, đường xá, và cây xanh, từ ảnh viễn thám độ phân giải cao (VHR) và ảnh máy bay không người lái (UAV) là một lĩnh vực nghiên cứu cốt lõi trong lĩnh vực bản đồ, Viễn thám và Hệ thống Thông tin Địa lý (GIS). Các phương pháp truyền thống, dựa trên phân loại pixel hoặc thuật toán dựa trên đối tượng (OBIA), thường gặp nhiều hạn chế trong việc xử lý tính phức tạp cao của hình học đô thị và sự biến đổi quang phổ trên quy mô lớn.

Sự xuất hiện của Học sâu (Deep Learning) trong những năm gần đây, đặc biệt là Mạng nơ-ron tích chập (CNN), đã tạo ra một cuộc cách mạng trong lĩnh vực này. CNN cho phép các mô hình tự động học và trích xuất các đặc trưng cấp cao (high-level features) từ dữ liệu ảnh. Phương pháp này được gọi là Phân đoạn Ngữ nghĩa (Semantic Segmentation), nơi mục tiêu là gán nhãn lớp chính xác cho từng pixel trong ảnh, từ đó xác định chính xác ranh giới của đối tượng.

Các kiến trúc mạng tiên tiến như U-Net, U-Net++, DeepLabV3+, PSPNet, và Feature Pyramid Network cùng với các backbone mạnh mẽ như ResNet (ResNet-18, 34, 50, 101, 152) đã trở thành tiêu chuẩn vàng cho mục tiêu trích xuất dữ liệu từ hình ảnh. Các mô hình này không chỉ cải thiện đáng kể độ chính xác so với các phương pháp truyền thống mà còn cho phép xây dựng một quy trình hoàn chỉnh, từ dự đoán pixel thô (Raster) đến tạo ra các lớp dữ liệu vector (Shapefile/GeoJSON) sẵn sàng cho phân tích trên môi trường GIS. Quy trình hoàn chỉnh “End-to-End” này bao gồm các bước không thể thiếu như gán thông tin địa lý (Geo-referencing) và vector hóa (Polygonization), nhằm đảm bảo dữ liệu đầu ra không chỉ chính xác về phân loại mà còn hữu ích về mặt địa lý. Sự phát triển này mở ra khả năng lập bản đồ tự động, cập nhật dữ liệu đô thị nhanh chóng, và hỗ trợ hiệu quả cho quy hoạch và quản lý tài nguyên.

1.2. Quy trình trích xuất dữ liệu tòa nhà sử dụng mạng học sâu

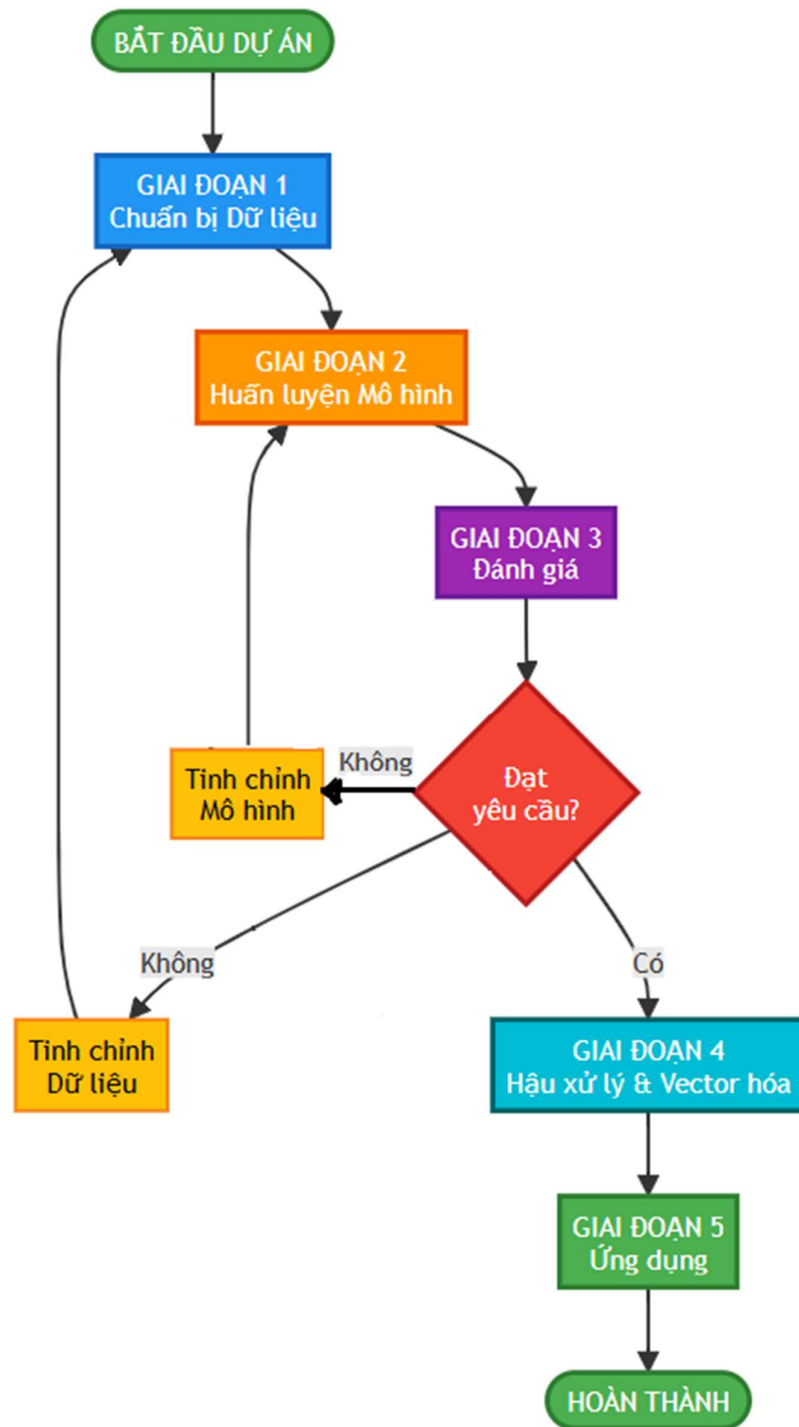
Quy trình trích xuất dữ liệu tòa nhà trên ảnh viễn thám, ảnh UAV có thể được chia thành năm giai đoạn chính như Hình 1. Mục đích của 04 giai đoạn của quy trình có thể tóm tắt như sau:

Giai đoạn 1: Chuẩn bị Dữ liệu, nơi dữ liệu thô được thu thập, gán nhãn, và tăng cường. Dữ liệu sau đó được đưa vào giai đoạn 2.

Giai đoạn 2: Huấn luyện Mô hình để đào tạo mạng học sâu.

Giai đoạn 3: Đánh giá sẽ kiểm tra hiệu suất của mô hình sau khi huấn luyện. Nếu kết quả Không đạt yêu cầu, quy trình sẽ lặp lại, có thể quay về Tinh chỉnh Mô hình (ví dụ: thay đổi kiến trúc, siêu tham số) hoặc quay về Tinh chỉnh Dữ liệu (thu thập thêm/cải thiện nhãn). Nếu kết quả Có đạt yêu cầu, quy trình tiến đến giai đoạn 4.

Giai đoạn 4: Hậu xử lý & Vector hóa, nơi các mask dự đoán (Raster) được làm sạch, gán thông tin địa lý, và chuyển đổi thành định dạng vector GIS (Shapefile).



Hình 1. Quy trình tổng quan trích xuất dữ liệu tòa nhà từ dữ liệu ảnh sử dụng mạng học sâu

2.1. Giai đoạn Chuẩn bị Dữ liệu (Data Preparation)

Đây là bước nền tảng, quyết định chất lượng và độ chính xác của mô hình. Chi tiết từng bước trong giai đoạn chuẩn bị dữ liệu có thể được nêu dưới đây (xem Hình 2):

Thu thập Dữ liệu (Imagery Acquisition): Sử dụng ảnh vệ tinh hoặc ảnh UAV độ phân giải cao (ví dụ: Sentinel-2, Landsat-8, WorldView, hoặc ảnh UAV).

Gắn nhãn (Labeling/Annotation): Tạo ra các **mask (mặt nạ)** nhị phân (binary mask) chính xác cho từng tòa nhà trên ảnh. Trong mask, các pixel thuộc tòa nhà được gán giá trị **1** (hoặc 255), và nền được gán giá trị **0**. Đây là công việc đòi hỏi nhiều thời gian và công sức nhất của toàn bộ quy trình này.

Tăng cường Dữ liệu (Data Augmentation): Áp dụng các phép biến đổi như xoay (rotation), lật (flipping), thay đổi độ sáng (radiometric), hoặc co giãn (geometric) để tăng số lượng dữ liệu huấn luyện và giúp mô hình tổng quát hóa tốt hơn và cũng để tránh hiện tượng quá khớp (overfitting) khi đào tạo mô hình.

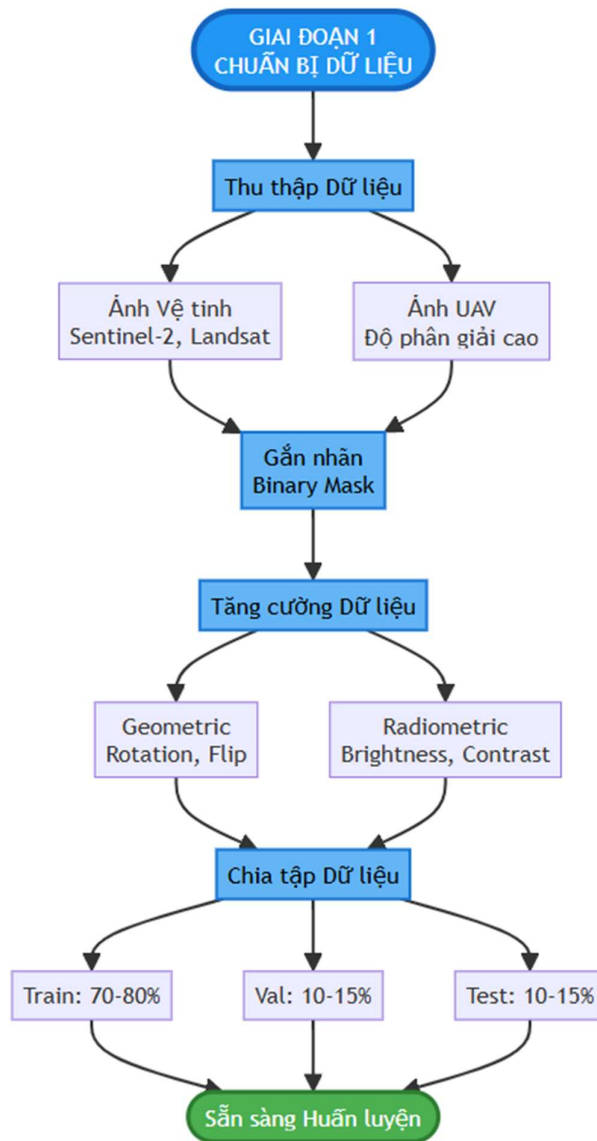
Chia tập Dữ liệu: Chia bộ dữ liệu đã gắn nhãn thành ba phần, bao gồm:

Tập huấn luyện (Training Set): Dùng để đào tạo mô hình.

Tập xác thực (Validation Set): Dùng để theo dõi hiệu suất và điều chỉnh siêu tham số (hyperparameters) trong quá trình huấn luyện.

Tập kiểm tra (Testing Set): Dùng để đánh giá hiệu suất cuối cùng của mô hình.

Việc chia tập dữ liệu thành ba bộ phải đảm bảo tỉ lệ tương ứng giữa ba bộ training, validation và testing set. Về nguyên tắc chia phải đảm bảo số lượng mẫu đủ lớn cho quá trình học để mô hình có khả năng tổng quát hóa đối tượng cần học. Số lượng mẫu trong tập validation và testing set cũng đảm bảo nguyên tắc đủ lớn để có thể đánh giá được độ chính xác dự đoán. Ba bộ dữ liệu này phải thảo mãn độc lập, và tỉ lệ có thể chia từ 50%, 60%, 70%, 80% cho bộ training tùy vào độ lớn của bộ dữ liệu hiện có và mức độ phức tạp của các tòa nhà cần trích xuất.



Hình 2. Giai đoạn 1- Chuẩn bị dữ liệu

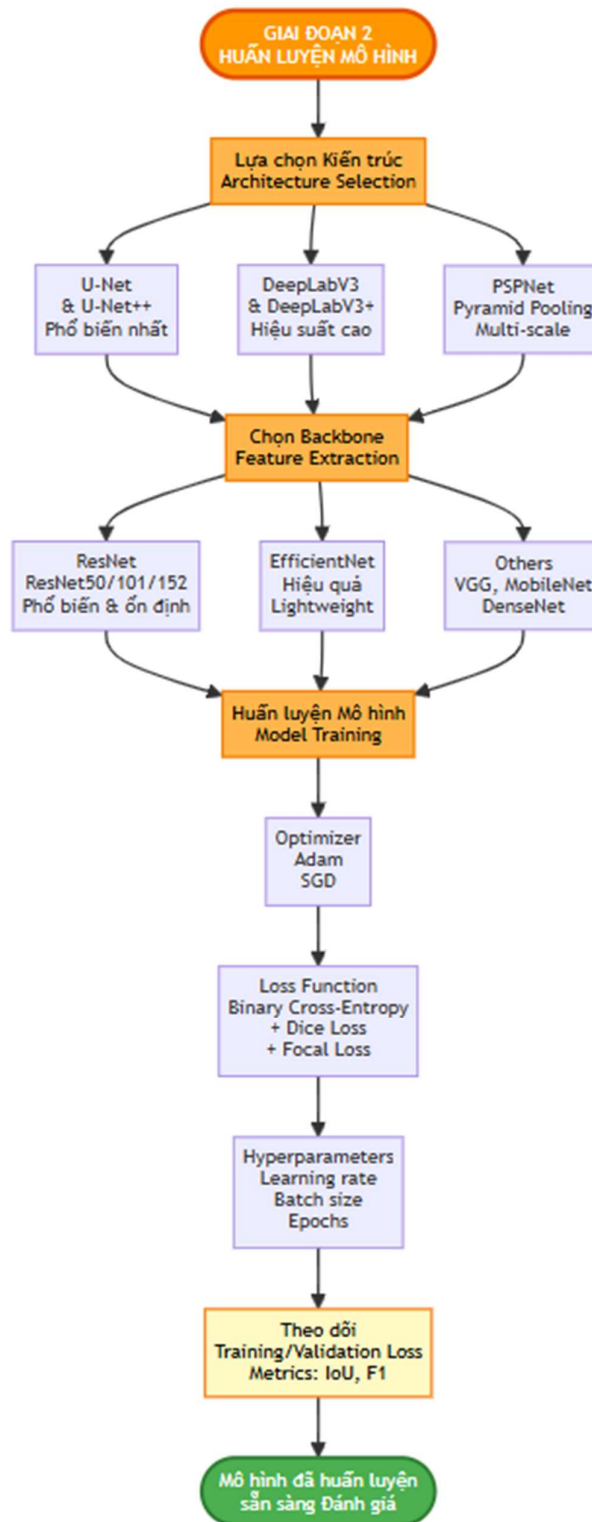
2.2. Giai đoạn Lựa chọn và Huấn luyện Mô hình (Model Training)

Giai đoạn này tập trung vào việc lựa chọn kiến trúc mạng học sâu và tiến hành huấn luyện mô hình. Các bước chính trong giai đoạn 2 (xem Hình 3) như sau:

Lựa chọn Kiến trúc (Backbone Selection): Chọn một kiến trúc mạng phân đoạn ngữ nghĩa phổ biến. Các lựa chọn các mô hình hàng đầu thường là U-Net (và các biến thể như U-Net++), DeepLabV3/DeepLabV3+, hoặc PSPNet.

Lựa chọn Backbone: Quyết định kiến trúc mạng cơ sở (backbone) để trích xuất đặc trưng (feature extraction), ví dụ: ResNet50, ResNet101, ResNet152, hoặc EfficientNet. Việc lựa chọn này phụ thuộc chính vào cấu hình máy tính hiện có và độ chính xác yêu cầu.

Huấn luyện Mô hình: Chạy thuật toán tối ưu hóa (ví dụ: Adam, SGD) để giảm thiểu hàm mất mát (loss function) – thường là Binary Cross-Entropy (BCE) Loss kết hợp với Dice Loss hoặc Focal Loss.



Hình 3. Giai đoạn 2- Huấn luyện mô hình

2.3. Giai đoạn Đánh giá mô hình (Model Evaluation)

Hiệu suất của các thuật toán học máy có thể được đánh giá bằng các chỉ số theo hai cách chính. Thứ nhất, mô hình được tối ưu hóa thông qua một hàm mất mát (loss function). Thứ hai, hiệu suất của nó được xác thực và đánh giá. Trong bài báo này, hiệu suất của mô hình U-Net cho nhiệm vụ trích xuất tòa nhà được đo lường bằng sáu chỉ số chung: độ chính xác (**accuracy**), độ đúng (**precision**), độ nhạy (**recall**), điểm F1 (**F1-score**), chỉ số Jaccard (**IoU**), và hệ số Dice (**Dice coefficient**). Bốn chỉ số đầu tiên được suy ra từ ma trận nhầm lẫn (**confusion matrix**), bao gồm dương tính đúng (**TP**), dương tính sai (**FP**), âm tính sai (**FN**), và âm tính đúng (**TN**). Trong bối cảnh phân đoạn hình ảnh, TP đề cập đến số lượng pixel mà mô hình dự đoán đúng. FP đại diện cho số lượng pixel mà mô hình phân loại sai. TN chỉ ra số lượng pixel nền (âm tính) mà mô hình xác định đúng, và FN là số lượng pixel nền mà mô hình phân loại sai.

Đánh giá hiệu suất của mô hình đã huấn luyện trên tập mẫu kiểm tra (test set) (xem Hình 4). Đo lường Hiệu suất sử dụng các chỉ số tiêu chuẩn cho phân đoạn ngữ nghĩa, gồm có:

Intersection over Union (IoU) / Jaccard Index: Đo lường sự chồng lấn giữa mask dự đoán và mask nhãn thực tế. Đây là chỉ số quan trọng nhất.

Accuracy (Độ chính xác), Precision (Độ đúng), Recall (Độ nhạy), F1-Score dựa trên ma trận Confusion matrix. Dưới đây là phân tích ý nghĩa của ma trận Confusion matrix ứng dụng trích xuất tòa nhà trên ảnh:

1. Accuracy (Độ chính xác)

Độ chính xác là một chỉ số phổ biến trong phân đoạn ngữ nghĩa, được định nghĩa là số lượng pixel được phân đoạn chính xác chia cho tổng số pixel:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

2. Precision (Độ đúng)

Precision được tính là số lượng pixel được phân đoạn chính xác (True Positives) chia cho tổng số pixel được mô hình dự đoán là đối tượng (True Positives + False Positives)::

$$Precision = \frac{TP}{TP + FP}. \quad (2)$$

3. Recall (Độ nhạy)

Recall (Độ nhạy) đo lường tỷ lệ giữa True Positives và tổng của True Positives và False Negatives:

$$Recall = \frac{TP}{TP + FN}. \quad (3)$$

4. F1-Score

F1-score là sự đánh đổi (trade-off) giữa precision và recall, và được tính bằng công thức trung bình điều hòa của chúng:

$$F1 = 2 \left(\frac{Precision \times Recall}{Precision + Recall} \right). \quad (4)$$

5. IoU và Dice Score (Chỉ số ưu tiên)

Bên cạnh bốn chỉ số chung trên, chỉ số Jaccard và Dice được ưu tiên sử dụng trong tác vụ phân đoạn ngữ nghĩa vì chúng thuộc loại chỉ số Intersection over Union (IoU). **Jaccard Index (IoU)** (hoặc Jaccard Index) được định nghĩa là tỷ lệ giữa phần giao (Intersection) và phần hợp (Union) giữa vùng dự đoán (P) và vùng thực tế (A):

$$IoU = \frac{|P \cap A|}{|P \cup A|} = \frac{|P \cap A|}{|P| + |A| - |P \cap A|}. \quad (5)$$

Dice Score (F1-Score Tăng cường)

Giống như Jaccard/IoU, Dice score là một chỉ số đáng tin cậy khác trong phân đoạn ngữ nghĩa, được tính để đo lường sự tương đồng giữa hai mẫu (vùng dự đoán và vùng thực tế):

$$Dice = \frac{2|P \cap A|}{|P| + |A|}. \quad (6)$$

Bảng 1. Ý nghĩa của ma trận confusion matrix trong trích xuất tòa nhà trên ảnh

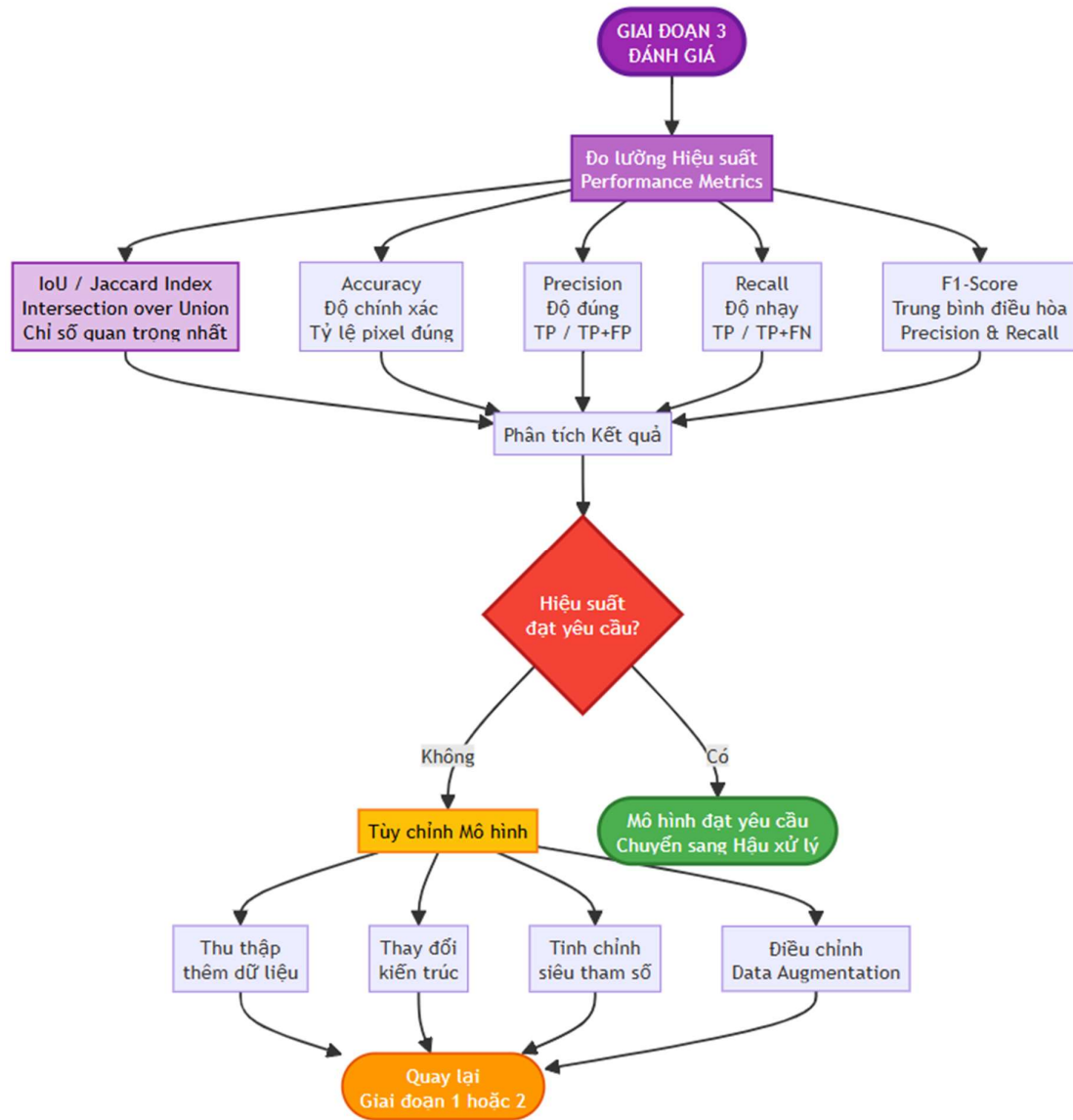
Ký hiệu	Tên gọi (Tiếng Việt)	Tiếng Anh	Ý nghĩa Cụ thể trong Nhận diện Tòa nhà
TP	Dương tính thật	True Positive	Pixel được mô hình dự đoán là Tòa nhà, và thực tế nó là Tòa nhà. (Phát hiện chính xác)
TN	Âm tính thật	True Negative	Pixel được mô hình dự đoán là Nền, và thực tế nó là Nền (đường, cây cối, đất trống). (Loại trừ chính xác)
FP	Dương tính giả	False Positive	Pixel được mô hình dự đoán là Tòa nhà, nhưng thực tế nó là Nền (Ví dụ: Nhận diện sân hoặc bóng đổ là Tòa nhà).
FN	Âm tính giả	False Negative	Pixel được mô hình dự đoán là Nền, nhưng thực tế nó là Tòa nhà. (Ví dụ: Bỏ sót các tòa nhà nhỏ hoặc bị che khuất).

Bảng 2. Các chỉ số đánh giá độ chính xác, ý nghĩa và cách sử dụng trong trích xuất tòa nhà trên ảnh

Chỉ số	Ý nghĩa trong Trích xuất Tòa nhà	Sử dụng
Accuracy (Độ chính xác)	Đo lường tỷ lệ các pixel được phân loại đúng (cả Tòa nhà và Nền) trên tổng số pixel. Thường	Ít được ưu tiên so với IoU/F1. Chỉ dùng để đánh giá hiệu suất tổng thể của pixel, nhưng

	có giá trị cao do sự mất cân bằng lớp (Nền chiếm diện tích lớn).	không đáng tin cậy do hiện tượng mất cân bằng lớp.
Precision (Độ đúng)	Đo lường độ tin cậy của mô hình: Trong số tất cả các pixel là Tòa nhà, có bao nhiêu phần trăm thực sự là Tòa nhà?	Để giảm thiểu FP (tránh gắn nhãn sai cây cối/đường thành tòa nhà). Cần thiết cho các ứng dụng yêu cầu độ tin cậy cao.
Recall (Độ nhạy)	Đo lường khả năng phát hiện: Mô hình đã phát hiện được bao nhiêu phần trăm của tổng số Tòa nhà thực tế trong ảnh?	Để giảm thiểu FN (tránh bỏ sót các tòa nhà nhỏ hoặc bị che khuất). Cần thiết cho ứng dụng lập bản đồ và quản lý tài sản khi mọi đối tượng cần được ghi nhận.
F1-Score	Đo lường sự cân bằng giữa Precision và Recall.	Sử dụng khi cần một mô hình vừa có độ tin cậy hợp lý (Precision) và vừa có khả năng phát hiện tốt (Recall).
IoU (Intersection over Union)	Chỉ số Vàng. Đo lường chất lượng của ranh giới đối tượng: Phần trăm chồng lấn giữa đa giác dự đoán và đa giác thực tế.	Quan trọng nhất trong GIS/Viễn thám. IoU là thước đo chính xác về chất lượng hình học của đa giác tòa nhà được trích xuất.

Tùy chỉnh Mô hình: Nếu hiệu suất không đạt yêu cầu, quay lại Giai đoạn 1 hoặc 2 để thu thập thêm dữ liệu, thay đổi kiến trúc hoặc tinh chỉnh siêu tham số (như tốc độ học, số lượng ảnh đưa vào mô hình mỗi lần batchsize, quán tính momentum, ...).



Hình 4. Giai đoạn 3- Đánh giá hiệu suất mô hình học sâu

2.4. Giai đoạn Hậu xử lý và Vector Hóa (Post-Processing & Vectorization)

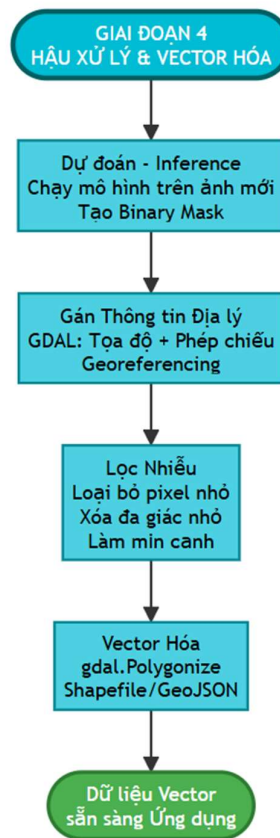
Sau khi có được mask dự đoán cuối cùng, cần chuyển đổi chúng thành dữ liệu vector để quản lý và sử dụng trên môi trường GIS. Chi tiết các bước được miêu tả trong Hình 5 như sau:

Dự đoán (Inference): Chạy mô hình trên ảnh UAV/vệ tinh với các ảnh mới để tạo ra các mask nhị phân đầu ra.

Gán Thông tin Địa lý (Georeferencing): Sử dụng các hàm GDAL để gán thông tin tọa độ và phép chiếu từ ảnh gốc sang mask dự đoán (như hàm `apply_geotransform`).

Lọc Nhiễu (Sieving/Smoothing): Áp dụng các bộ lọc để loại bỏ các pixel nhiễu nhỏ hoặc các đa giác quá nhỏ (ví dụ: các vùng chỉ 1-2 pixel).

Vector Hóa (Polygonization): Sử dụng các công cụ GIS như `gdal.Polygonize` hoặc các thư viện như GDAL/OGR hoặc Rasterio để chuyển đổi các pixel có giá trị 1 (tòa nhà) thành các đối tượng đa giác (Shapefile hoặc GeoJSON).



Hình 5. Giai đoạn 4- Hậu xử lý và vector hóa tòa nhà

Chương 2

THỰC NGHIỆM XÂY DỰNG QUY TRÌNH TRÍCH XUẤT DỮ LIỆU TỌA NHÀ TRÊN ẢNH UAV TẠI VIỆT NAM

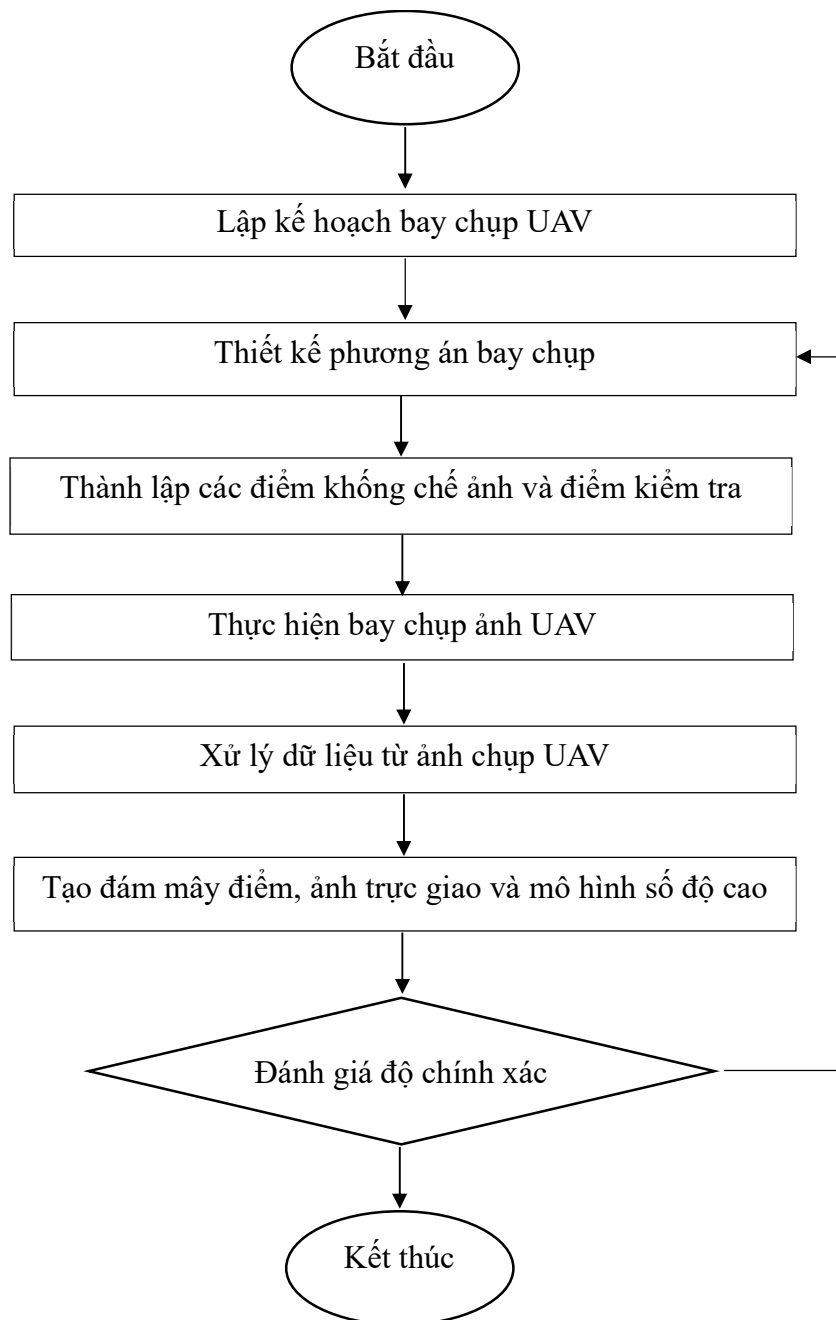
Trích xuất dữ liệu từ ảnh UAV để được các đối tượng cần thiết và thể hiện trên bản đồ địa hình, bản đồ địa chính cần phải thực hiện theo một quy trình chặt chẽ và phù hợp. Mặc dù ở nước ta đã ban hành thông tư số 07/2021 về việc xây dựng bản đồ địa hình tỉ lệ lớn từ dữ liệu UAV tuy nhiên các công tác trích xuất dữ liệu từ ảnh UAV phần lớn vẫn phải thực hiện thủ công. Kế thừa những thành tựu phát triển của cuộc cách mạng trí tuệ nhân tạo mà hiện nay việc trích xuất thông tin các đối tượng trên ảnh đã khả thi và đang là xu hướng. Những lợi ích to lớn từ việc trích xuất dữ liệu tự động như giảm giá thành, tăng năng suất lao động và đáp ứng kịp thời cho sản xuất là lý do để nhiều nước đầu tư nghiên cứu.

Tuy nhiên ở nước ta, công tác trích xuất dữ liệu tự động từ ảnh UAV để thành lập bản đồ và cập nhật dữ liệu vẫn còn là một chủ đề mới. Do đó, việc nghiên cứu để ban hành các hướng dẫn kỹ thuật, quy trình trích xuất dữ liệu tự động từ thiết bị UAV một cách chặt chẽ và phù hợp với điều kiện sản xuất ở nước ta là một hướng đi cần thiết.

2.1. Giai đoạn 1- Chuẩn bị dữ liệu

2.1.1. Thu thập dữ liệu từ thiết bị bay không người lái UAV

Việc thu thập dữ liệu từ ảnh UAV phục vụ cho công tác thành lập bản đồ đã được thực hiện ở nước ta từ nhiều năm qua. Có nhiều nghiên cứu và các hướng dẫn kỹ thuật về quy trình tạo ra ảnh trực giao (ortho images), mô hình số độ cao (DEM), và dữ liệu đám mây điểm (point cloud) từ ảnh chụp UAV (Lê, Lê et al. 2020), (Long, Goyal et al. 2021), (Long and Cường 2019), (Sỹ, Quý et al. 2017), (Cường, Trường et al. 2019). Tuy nhiên, ảnh trực giao, mô hình số độ cao và đám mây điểm từ ảnh chụp UAV có thể tóm tắt lại thông qua các bước chính như sơ đồ Hình 6.



Hình 6: Sơ đồ quy trình xử lý ảnh UAV

1. Lập kế hoạch bay chụp UAV

Dựa trên thông tư 07/2021/TT- BTNMT về “Quy định kỹ thuật thu nhận và xử lý dữ liệu ảnh số từ tàu bay không người lái phục vụ xây dựng, cập nhật cơ sở dữ liệu nền địa lý quốc gia tỷ lệ 1:2.000, 1:5.000 và thành lập bản đồ địa hình tỷ lệ 1:500,

1:1.000. Ngoài ra, xem xét các quy định, quy phạm trong thành lập bản đồ địa hình với các đặc điểm địa lý, kinh tế - xã hội của khu vực thành lập bản đồ. Sau đó tiến hành các công việc sau để lên kế hoạch bay chụp UAV:

2. Thiết kế phương án bay chụp

Dựa trên tọa độ các điểm ranh giới của khu vực khảo sát tiến hành tính chuyển sang tọa độ WGS-84 và xuất ra file KML, KMZ và trực quan hóa khu đo trên ứng dụng phần mềm Google Earth. Tiếp theo, tiến hành xác định chiều cao bay và mức độ độ chồng phủ (dọc và ngang) của ảnh, từ đó tính toán số ca bay với lưu ý đến thời lượng pin/1 ca (trung bình 15phút/1 ca bay đối với Phantom 4 RTK).

3. Thành lập các điểm khống chế ảnh và điểm kiểm tra

Tiến hành tìm kiếm các điểm mốc khống chế trắc địa mặt bằng và độ cao đã tồn tại trong khu vực và đánh giá khả năng sử dụng của các điểm này. Tiếp theo, tiến hành chêm dày hoặc thành lập mới các điểm khống chế và điểm kiểm tra tuân thủ theo các quy phạm hiện hành. Các điểm khống chế và điểm kiểm tra phục vụ công tác bay chụp cần đảm bảo cả về độ chính xác cũng như sự thông thoáng và rõ nét cho mục đích bay chụp UAV.

4. Thực hiện bay chụp ảnh UAV

Sau khi thiết kế ca bay và thành lập các điểm khống chế cũng như các điểm kiểm tra tiến hành bay chụp UAV. Công tác bay chụp UAV cần phải đảm bảo các yêu cầu sau:

- Kiểm tra đầy đủ pin cho thiết bị UAV, bộ phận điều khiển từ xa RC và bộ máy phát tín hiệu GNSS từ trạm base mặt đất.
- Cài đặt trạm Base theo phương pháp PPK để thu thập dữ liệu đo tĩnh tại mốc khống chế tọa độ và độ cao trong khu vực bay chụp.
- Kiểm tra thẻ SD và dung lượng bộ nhớ SD trong thiết bị bay.
- Tiếp theo tiến hành công tác bay:
- Khởi động trạm điều khiển và phần mềm điều khiển thiết bị bay

- Kiểm tra kết nối giữa tay điều khiển và màn hình kiểm soát
- Kiểm tra tín hiệu GNSS và tín hiệu kết nối với trạm điều khiển
- Thực hiện cất cánh thiết bị bay
- Theo dõi quá trình hoạt động của thiết bị bay như: tín hiệu kết nối với điều khiển, mức pin, tốc độ gió và các cảnh báo khác đến khi kết thúc ca bay
- Kiểm tra dữ liệu thu nhận và lưu dữ liệu từ thẻ nhớ sang máy tính hoặc các thiết bị khác.



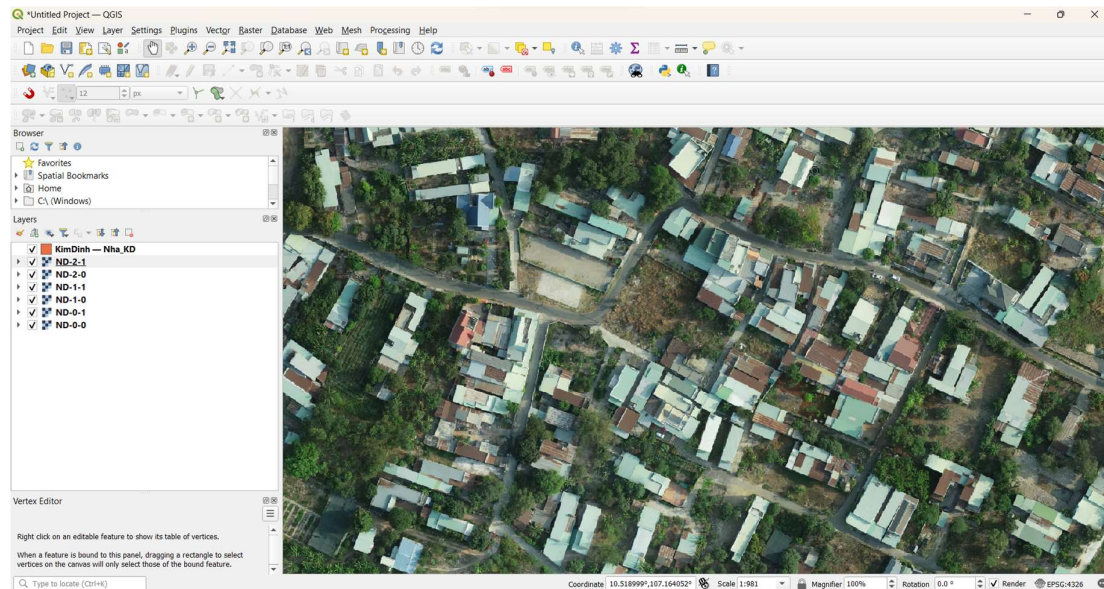
Hình 7. Trạm điều khiển mặt đất GNSS (trái); Tiến hành bay UAV (phải)

5. Xử lý dữ liệu từ ảnh chụp UAV

Ảnh sau khi thu nhận được bằng thiết bị bay không người lái sẽ phải thực hiện nắn ảnh thông qua hai bài toán cơ bản là định hướng trong (định hướng tương đối) và định hướng ngoài (định hướng tuyệt đối). Các ảnh được định hướng tương đối nhờ độ chồng phủ dọc và ngang giữa các tấm ảnh khi chụp. Để đưa tọa độ mô hình về hệ tọa độ trắc địa cần sử dụng các điểm khống chế ảnh mặt đất hoặc sử dụng tọa độ tâm chụp ảnh. Do đó để xử lý dữ liệu từ ảnh chụp UAV cần phải thực hiện hai công việc chính là xác định tọa độ tâm chụp và xử lý ảnh được trình bày chi tiết như sau:

2.1.2. Tạo mẫu phục vụ quá trình đào tạo sử dụng mạng học sâu

Từ ảnh sau khi đã nắn (ảnh trực giao - ortho) sẽ được sử dụng để vector hóa (số hóa) các đối tượng nhà trực tiếp trên ảnh. Có nhiều công cụ cho phép sử dụng để số hóa lớp đối tượng nhà như AutoCad, Arcgis, QGIS...



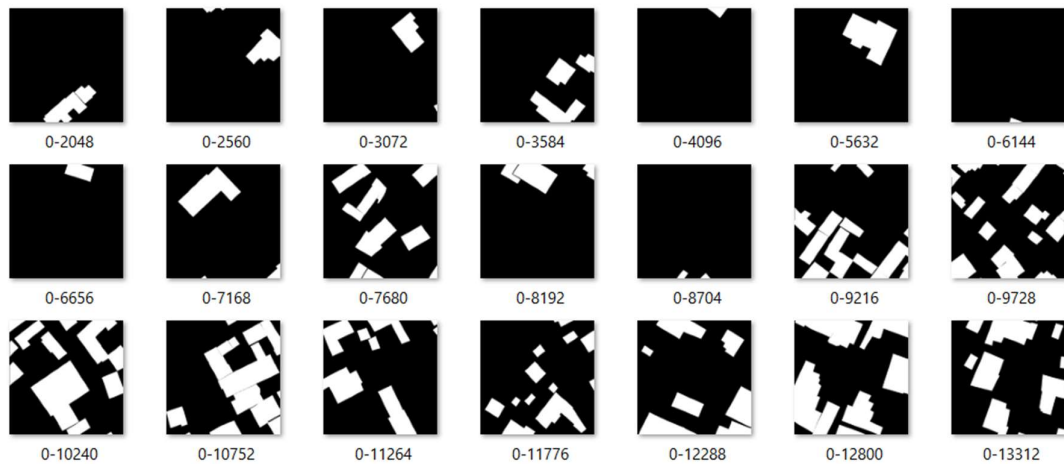
Hình 8. Ảnh trực giao sau khi nắn trên phần mềm QGIS

Sử dụng phần mềm QGIS để vector hóa lớp đối tượng trực tiếp trên ảnh theo các bước sau: Quá trình tạo mẫu cho mô hình học sâu được thực hiện bằng cách chi nhỏ ảnh ortho đầu vào thành nhiều mẫu nhỏ. Vì ảnh UAV sau khi nắn từ nhiều ảnh chụp sẽ có kích thước rất lớn (GiGa Byte như đã thảo luận trong phần ảnh màu) không khả thi cho việc phát triển mô hình học máy sử dụng mạng. Việc đọc toàn bộ hình ảnh vào bộ nhớ và sử dụng nó cho việc phát triển mô hình U-Net là không khả thi. Vì thế, để tăng hiệu suất của quá trình huấn luyện ảnh nắn sẽ được chia thành nhiều mẫu nhỏ gọi là các tiles. Kích thước của tiles phụ thuộc vào cách thiết kế mô hình và thường có kích thước (256 x 256), (512 x 512) hoặc (1024 x 1024) pixel.



Hình 9. Các mẫu tiles được chia nhỏ từ một tấm ảnh UAV ortho lớn

Sau khi đã chia toàn bộ ảnh thành các mẫu nhỏ sẽ thực hiện công việc raster hóa các polygon thành các “mặt nạ - mask” với mức độ xám là 8 bit. Các mask sử dụng để phân biệt đối tượng bởi mã nhị phân. Ví dụ trong Hình 10 phần có đối tượng nhà (tương ứng với màu trắng) được mã là 1 và phần không có đối tượng nhà (tương ứng với màu đen hay là background) có mã là 0.



Hình 10. Các mask tương ứng với các tiles



(a): True ortho- ảnh trực giao (b): Vector đối tượng nhà (c): mặt nạ của tòa nhà

Hình 11. Một mẫu từ tập dữ liệu trong khu vực nghiên cứu trong đó (a)-ảnh trực giao, (b)- là vector của tòa nhà, và (c) là mask được raster hóa tòa nhà.

2.1.3. Tăng cường dữ liệu (Data augmentation)

a. Biến đổi Hình học (Geometric Transformations)

Các biến đổi Hình học là thành phần **quan trọng nhất** trong chiến lược tăng cường. Chúng giúp mô hình chống lại các biến đổi tự nhiên về góc nhìn và hướng xảy ra trong ảnh chụp trên không.

Các phép biến đổi: Bao gồm thay đổi kích thước về độ phân giải chuẩn **512 x 512 pixel**, dịch chuyển vừa phải tối đa 5% kích thước ảnh, thay đổi tỉ lệ trong phạm vi 10%, và xoay giới hạn trong 15 độ.

Phản chiếu: Lật ngang được áp dụng với xác suất 50% và lật dọc với xác suất 30%.



Hình 12. Biến đổi hình học

b. Biến đổi phổ ảnh (Radiometric Transformations)

Các biến đổi này được thiết kế để mô phỏng các biến thể về điều kiện ánh sáng và đặc tính cảm biến, giúp mô hình thích ứng với các môi trường chụp khác nhau.

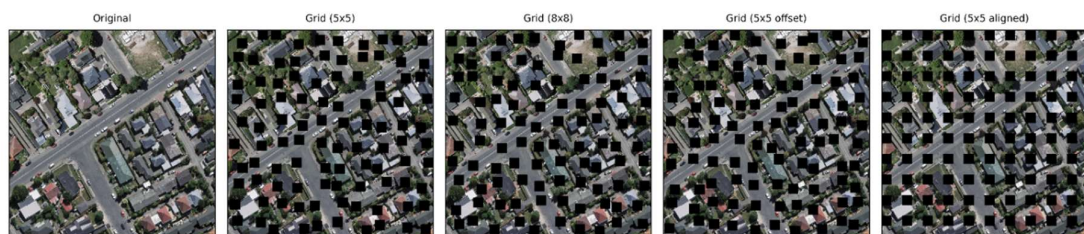
Các phép biến đổi: Bao gồm điều chỉnh độ sáng và độ tương phản ngẫu nhiên trong phạm vi 20%, dịch chuyển nhỏ các kênh RGB lên tới 15 đơn vị cường độ, và thỉnh thoảng xáo trộn kênh (channel shuffling) với xác suất 10%.



Hình 13. Biến đổi phổ ảnh

c. Kỹ thuật Che khuất (Occlusion Techniques)

Kỹ thuật này được đại diện bởi **GridDropout** trong triển khai, mô phỏng các vật cản một phần hoặc các lỗi cảm biến. GridDropout tạo ra một mẫu lưới 5 x 5 và được áp dụng với xác suất 30%.



Hình 14. Biến đổi che khuất

2.1.4. Chia tập dữ liệu

Dữ liệu phục vụ nghiên cứu trong đề án tốt nghiệp này là lớp đối tượng nhà tại một số khu vực dân cư ở cảng biển tại Ngãi Giao, Kim Dinh tỉnh Bà Rịa - Vũng Tàu và khu vực Cam Đức, Cam Ranh tỉnh Khánh Hòa. Lớp đối tượng nhà được số hóa trực tiếp từ ảnh màu UAV (3 kênh phổ) với độ phân giải không gian rất cao từ 5 đến 10cm. Để thực hiện việc đào tạo mô hình học máy, bộ dữ liệu được chia thành ba tập dữ liệu độc lập để phục vụ cho ba mục đích gồm đào tạo (training), đánh giá mô hình (validation), và kiểm chứng thực tế (testing). Sau đây sẽ trình bày cụ thể ba tập dữ liệu này.

Trong nghiên cứu này, Bộ dữ liệu **Trường Đại học Mở - Địa chất Hà Nội (HUMG)** được tạo ra với độ phân giải không gian cao nhằm mục đích kiểm tra hiệu suất của mô hình đề xuất trong môi trường xây dựng địa phương có tính đa dạng và phức tạp cao.

1. Đặc điểm Dữ liệu

Độ phân giải: 0.1m.

Phạm vi: Bao gồm hình ảnh được thu thập trên nhiều tỉnh, đại diện cho các khu vực miền Bắc, miền Trung và miền Nam Việt Nam.

Số lượng tòa nhà: Chứa khoảng trên 40.000 tòa nhà đã được gán nhãn.

Tính đa dạng: Bộ dữ liệu phản ánh một phổ rộng các loại cấu trúc, bao gồm khu công nghiệp, nhà cao tầng hiện đại, nhà ở nông thôn truyền thống và các khu đô thị đông dân cư.

2. Cấu trúc dữ liệu và Phân chia tập dữ liệu

Bộ dữ liệu HUMG bao gồm tổng cộng 6.857 ảnh có kích thước 512x512 pixel. Dữ liệu được chia thành hai tập chính:

Tập Huấn luyện (Training Set): 5.312 ảnh (77%). Việc phân chia tập huấn luyện lớn nhằm mục đích đảm bảo mô hình có thể xử lý được các loại tòa nhà đa dạng.

Tập Xác thực (Validation Set): 1.545 ảnh (23%).

Mục đích: Tập xác thực được sử dụng để theo dõi quá trình học và đánh giá hiệu suất của mô hình, trong khi các siêu tham số (hyperparameters) được giữ cố định

Bảng 3. Đặc điểm nhà trong bộ dữ liệu HUMG sử dụng để kiểm tra (Testing set)

Loại khu vực	Mô tả	Thách thức phân đoạn	Hình ảnh trên ảnh UAV
Khu vực Đô thị: Tòa nhà Hiện đại (48 mẫu)	Kiến trúc hiện đại với hình dạng hình học rõ ràng và kết cấu bề mặt tối thiểu; thường được xây dựng từ bê tông, kính hoặc kim loại.	Các bề mặt đồng nhất lớn thường gây nhầm lẫn với các đối tượng không thấm nước khác như đường sá hoặc vỉa hè.	
Khu vực Đô thị: Đa dạng cao (182 mẫu)	Môi trường xây dựng truyền thống, mật độ dày đặc với sự khác biệt lớn về kích thước, hình dạng và vật liệu.	Mật độ xây dựng cao và các cấu trúc chồng lấn làm tăng lỗi phân loại nhằm giữa các mái nhà liền kề.	

Khu vực Nông thôn: Đa dạng thấp (747 mẫu)	Các khu vực dân cư thưa thớt chứa ít tòa nhà, cô lập giữa các cảnh quan thiên nhiên rộng lớn.	Mật độ xây dựng thấp dẫn đến việc bỏ sót đối tượng (missed detections) và tỷ lệ dương tính sai (false negatives) cao hơn.	
Khu vực Nông thôn: Đa dạng trung bình (209 mẫu)	Môi trường nông thôn nổi bật với các loại và kích thước tòa nhà đa dạng ở mật độ vừa phải.	Tính biến đổi về cấu trúc kiểm tra khả năng tổng quát hóa của mô hình và độ chính xác của ranh giới.	
Khu vực Chuyên biệt: Công nghiệp (159 mẫu)	Các khu phức hợp công nghiệp quy mô lớn với hình dạng đặc trưng và mái nhà rộng.	Các cấu trúc lớn, liền kề nhau khó được xác định ranh giới chính xác, thường dẫn đến phân đoạn không hoàn chỉnh.	
Khu vực Chuyên biệt: Che khuất bởi Cây cối (200 mẫu)	Các khu vực mà thảm thực vật che khuất một phần hoặc toàn bộ tòa nhà.	Tán cây gây nhiễu thị giác, cản trở việc phát hiện chính xác các cấu trúc bị che khuất	

		hoặc chỉ nhìn thấy một phần.	
--	--	---------------------------------	--

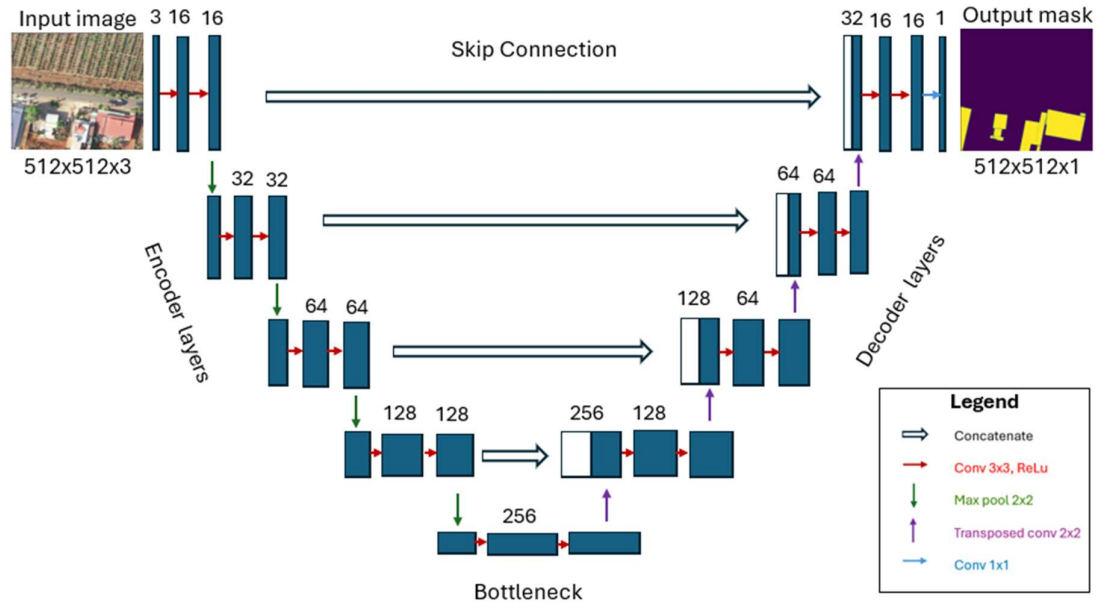
2.2. Giai đoạn 2- Huấn luyện mô hình

2.2.1. Lựa chọn kiến trúc mô hình mạng học sâu

Các mô hình dựa trên CNN là những mô hình học sâu phổ biến hiện nay cho lĩnh vực thị giác máy tính. Trong trích xuất tòa nhà trên ảnh, các mô hình U-Net, DeepLab, PSPNet và FPN là những mô hình đã được đánh giá là có hiệu quả cao. Trong chuyên đề này sẽ tập trung giới thiệu về hai mô hình U-Net và DeepLab trong trích xuất tòa nhà trên ảnh UAV và ảnh viễn thám.

a. Mô hình mạng U-Net

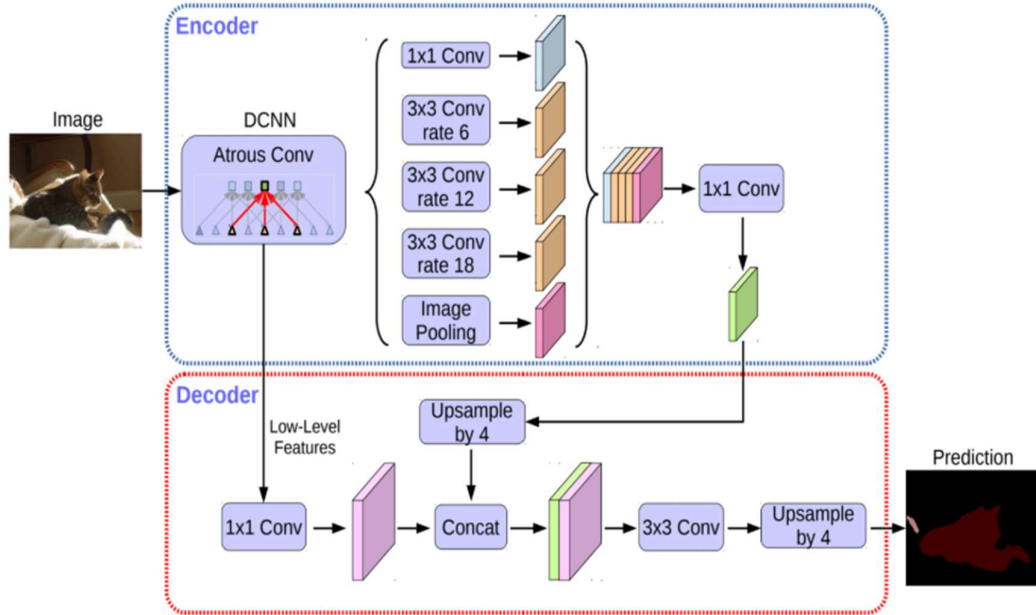
Mạng U-Net, được phát triển bởi Ronneberger, Fischer et al. (2015), là một mô hình phổ biến cho việc phân đoạn ngữ nghĩa hình ảnh (image semantic segmentation) do có độ chính xác và hiệu quả cao. Mạng này là một Mạng nơ-ron tích chập hoàn toàn (FCNN - fully convolutional neural network) sử dụng cả hai luồng thu gọn (contracting path) và mở rộng (expanding path) trong quá trình tích chập. U-Net có cấu trúc đối xứng, tạo thành hình chữ U với các phần mã hóa (encoder) và giải mã (decoder) (xem Hình 15).



Hình 15. Mô hình U-net với ảnh đầu vào RGB kích thước 512 x 512 pixel 3 kênh và mask đầu ra kích thước 512 x 512 với 1 kênh.

b. Mô hình mạng DeepLab

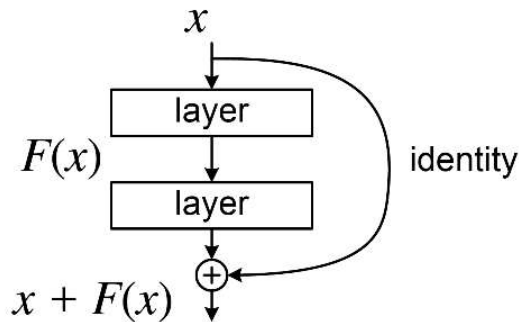
DeepLab là mô hình phân đoạn ngữ nghĩa được phát triển bởi Google Research và được biết đến với khả năng nắm bắt các chi tiết trên ảnh và phân loại đối tượng dựa trên pixel ảnh. Cải tiến chính của mô hình này là sử dụng tích chập giãn nở (atrous convolution hoặc dilation convolution). Phiên bản đầu tiên (V1) của DeepLab được phát triển dựa trên cấu trúc mạng VGG-16 (gồm 16 lớp) bởi nhóm Visual Geometry Group tại đại học Oxford (Yu, Yang et al. 2016) tuy nhiên có thay thế lớp kết nối đầy đủ (Fully connected layer) cuối cùng bởi lớp tích chập giãn nở.



Hình 16. Kiến trúc mô hình mạng DeepLabV3+ (Chen, Zhu et al. 2018).

2.2.2. Lựa chọn backbone

Mạng ResNet (Residual Network) được biết đến là mạng học sâu sử dụng cho các bài toán xác định đối tượng và phân đoạn hình ảnh. Hiện nay có rất nhiều biến thể của mạng ResNet với số lớp khác nhau gồm 18 lớp (ResNet18), 34 lớp (ResNet34), 50 lớp (ResNet50), 101 lớp (ResNet101), hoặc 152 lớp (ResNet152).



Hình 17. Sơ đồ nối tắt (skip connection) trong ResNet (He, Zhang et al. 2016)

Bảng 4: So sánh các Backbone ResNet sử dụng trích xuất tòa nhà trên ảnh

Đặc điểm	ResNet18	ResNet50	ResNet101	ResNet152
Độ Sâu (Số lớp)	18	50	101	152
Số Lượng Tham số (Triệu)	~11.7M	~25.6M	~44.5M	~60.2M
Tốc độ Xử lý	Nhanh nhất	Nhanh	Trung bình	Chậm nhất
Khả năng Trích xuất Đặc trưng	Thấp/Trung bình	Tốt	Rất Tốt	Rất tốt
Yêu cầu Bộ nhớ (RAM/VRAM)	Thấp nhất	Thấp	Trung bình	Cao nhất
Rủi ro Overfitting	Thấp	Thấp/Trung bình	Cao	Rất Cao (Khi dữ liệu ít)
Khuyến nghị cho Trích xuất Tòa nhà	Tốt cho các ứng dụng thời gian thực hoặc dữ liệu nhỏ/vừa.	Lý tưởng cho hầu hết các tác vụ. Cân bằng giữa tốc độ và độ chính xác.	Dùng cho các bộ dữ liệu lớn hoặc yêu cầu độ chính xác cao.	Dùng cho các bộ dữ liệu rất lớn hoặc yêu cầu độ chính xác rất cao.

2.2.3. Huấn luyện mô hình

a. Thiết lập Tính toán (Computational Setup)

Framework và Môi trường: Tất cả quá trình phát triển và huấn luyện mô hình được thực hiện bằng cách sử dụng framework học sâu PyTorch (phiên bản 1.11.0) trong môi trường Google Colaboratory.

Phần cứng: Quá trình tận dụng GPU NVIDIA Tesla T4 với 16 GB bộ nhớ GDDR6. Thiết lập tính toán này cho phép tăng tốc huấn luyện và tạo điều kiện thuận lợi cho việc thử nghiệm với các kiến trúc, khối lượng dữ liệu và chiến lược tăng cường khác nhau.

Kích thước Batch: Huấn luyện mini-batch được tiến hành với kích thước batch là bốn ảnh. Lựa chọn này được đưa ra dựa trên các ràng buộc về bộ nhớ GPU và nhằm mục đích tối đa hóa việc sử dụng phần cứng mà không làm ảnh hưởng đến sự ổn định của quá trình huấn luyện.

b. Tối ưu hóa Mô hình (Model Optimization)

Bộ tối ưu hóa (Optimizer): Quá trình tối ưu hóa mô hình được thực hiện bằng bộ tối ưu hóa AdamW. AdamW được chọn vì khả năng xử lý hiệu quả các mô hình học sâu phức tạp và cơ chế điều chuẩn Weight Decay tích hợp giúp giảm thiểu hiện tượng quá khớp (overfitting).

Tốc độ học (Learning Rate) và Epoch: Tốc độ học tập ban đầu được đặt là 0.001 và được giảm động trong quá trình huấn luyện theo lịch trình Cosine Annealing, đảm bảo sự hội tụ suôn sẻ và ổn định.

Dừng sớm (Early Stopping): Quá trình huấn luyện diễn ra tối đa 100 epochs, với cơ chế dừng sớm được áp dụng dựa trên độ mất mát (validation loss) để tránh quá khớp và tính toán không cần thiết.

c. Hàm Mất mát

Hàm Mất mát (Loss Function): Để tối ưu hóa mô hình, một hàm mất mát kết hợp giữa Binary Cross-Entropy (BCE) loss và hệ số Dice được áp dụng. Phương pháp tiếp cận mục tiêu kép này nhằm mục đích cân bằng giữa độ chính xác phân loại pixel với chất lượng chồng lán của các mask tòa nhà dự đoán và thực tế, qua đó giải quyết

trực tiếp vấn đề mất cân bằng lớp (class imbalance), một thách thức thường gặp trong nhiệm vụ phân đoạn tòa nhà do sự chiếm ưu thế của các pixel nền trong hầu hết các cảnh.

2.3. Giai đoạn 3- Đánh giá mô hình

2.3.1. Đo lường hiệu suất mô hình

Đánh giá độ chính xác phân đoạn tòa nhà được thực hiện trên bộ HUMG với mô hình U-Net được thực hiện với 5 thước đo trong bảng sau:

Bảng 5. Kết quả đánh giá độ chính xác sử dụng mô hình U-Net trên bộ HUMG

Loại Khu vực	Accuracy/ Recall	Precision	F1-score	IoU
Đô thị: Hiện đại	0.960	0.952	0.952	0.925
Đô thị: mật độ đông đúc	0.933	0.936	0.933	0.878
Nông thôn: mật độ thấp	0.916	0.920	0.906	0.873
Nông thôn: mật độ trung bình	0.959	0.950	0.953	0.917
Khu Công nghiệp	0.958	0.949	0.947	0.913
Khu vực che khuất bởi cây	0.938	0.942	0.936	0.890

2.3.2. Phân tích kết quả

Việc phân tích bảng dữ liệu hiệu suất theo từng khu vực cho thấy mô hình của bạn hoạt động khác nhau tùy thuộc vào đặc điểm của môi trường xây dựng và thách thức về hình ảnh. Dữ liệu của bạn bao gồm 5 chỉ số chính (Accuracy, Precision, Recall, F1-score, IoU) được đo trên 6 loại khu vực khác nhau.

Mô hình hoạt động tốt nhất ở các khu vực có hình học rõ ràng và mật độ không quá phức tạp (dễ phân biệt nền và đối tượng):

Đô thị: Hiện đại (Modern Buildings): Đạt IoU cao nhất (0.925). Điều này hợp lý vì các tòa nhà hiện đại thường có hình dạng hình học đơn giản, sạch sẽ, giúp mô hình dễ dàng xác định ranh giới (IoU) và ít bỏ sót (Recall 0.960).

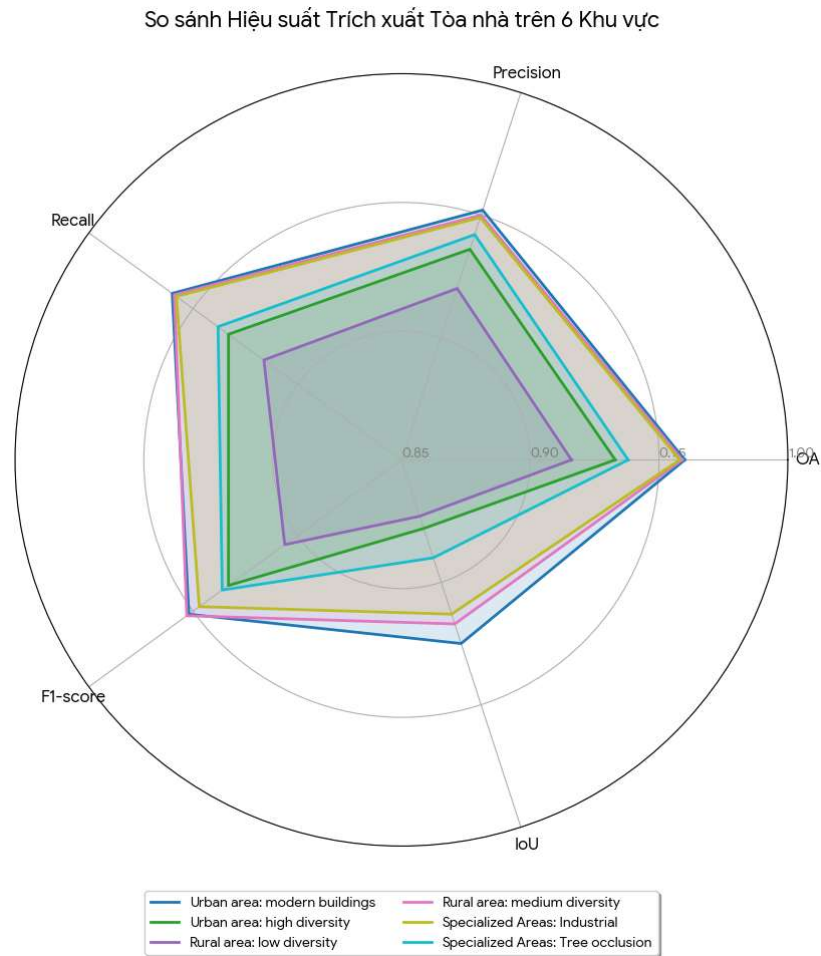
Nông thôn: mật độ nhà ở mức trung bình: Đạt IoU cao thứ hai (0.917) với F1-score là 0.953. Điều này cho thấy mật độ vừa phải và kích thước đa dạng không phải là thách thức lớn khi các đối tượng được phân tách rõ ràng.

Khu công nghiệp: IoU là 0.913. Các khu công nghiệp dù lớn nhưng thường có hình dạng đồng nhất, giúp mô hình dễ dàng phân đoạn (TP cao).

Tuy nhiên, hiệu suất giảm rõ rệt ở hai loại khu vực, bao gồm:

Nông thôn: Đa dạng Thấp: Mặc dù dễ phân biệt, IoU chỉ đạt 0.873 và F1-score là 0.906. Điều này thường do mật độ tòa nhà quá thấp, dẫn đến tỉ lệ False Negatives (FN) cao (bỏ sót các tòa nhà cô lập), làm giảm Recall (0.916) và kéo F1/IoU xuống.

Đô thị: mật độ cao: IoU đạt 0.878. Thách thức ở đây là mật độ dày đặc và cấu trúc chồng lấn. Precision (0.936) và Recall (0.933) gần bằng nhau, cho thấy mô hình vừa gặp lỗi báo sai (FP) do nhầm lẫn giữa các mái nhà liền kề, vừa gặp lỗi bỏ sót (FN) các ranh giới phức tạp.



Hình 18. So sánh hiệu suất trích xuất tòa nhà ở các khu vực có đặc điểm nhà khác nhau trên bộ Testing Set - HUMG

2.4. Giai đoạn 4- Hậu xử lý và vector hóa

Sau khi mô hình học sâu (ví dụ: U-Net hoặc DeepLab) hoàn tất quá trình dự đoán, tạo ra các mask nhị phân thể hiện vị trí các tòa nhà, Giai đoạn Hậu xử lý và Vector hóa trở nên tối quan trọng. Mục tiêu là chuyển đổi các pixel dự đoán thô thành dữ liệu **vector GIS** có thể quản lý và ứng dụng trực tiếp trong các hệ thống quy hoạch đô thị tại Việt Nam. Chi tiết các bước được thực hiện theo trình tự:

2.4.1. Dự đoán (Prediction)

Chạy mô hình học sâu (ví dụ: U-Net với backbone ResNet) đã được huấn luyện trên ảnh UAV mới (bộ mẫu mới) – thường là ảnh có độ phân giải rất cao (ví dụ: 0.1m). Quá trình này tạo ra các mask nhị phân (binary masks) đầu ra, trong đó pixel được phân loại là tòa nhà (giá trị 1 hoặc 255) hoặc nền (giá trị 0).

2.4.2. Gán Thông tin Địa lý (Georeferencing)

Các mask nhị phân cần được định vị chính xác trên bản đồ. Sử dụng các hàm GDAL (Geospatial Data Abstraction Library) để sao chép thông tin tọa độ và phép chiếu (ví dụ: VN-2000 hoặc WGS84). Sử dụng hàm như `apply_geotransform` được dùng để gán thông tin Geotransform và Projection từ ảnh UAV gốc sang mask dự đoán. Nhiệm vụ này nhằm đảm bảo rằng ranh giới tòa nhà trích xuất sẽ nằm đúng vị trí thực tế trên bản đồ nền Việt Nam, phục vụ cho công tác kiểm kê, quản lý đất đai.

2.4.3. Vector Hóa (Polygonization)

Chuyển đổi dữ liệu từ định dạng raster (pixel) sang định dạng vector (đa giác/polygon) để có thể tính toán diện tích, chu vi, và sử dụng trong phần mềm GIS. Sử dụng các công cụ GIS mạnh mẽ như `gdal.Polygonize` hoặc các thư viện như GDAL/OGR để chuyển đổi các pixel có giá trị 1 (tòa nhà) thành các đối tượng đa giác. Kết quả là các Shapefile hoặc GeoJSON cuối cùng chứa các đa giác ranh giới của các tòa nhà đã được trích xuất trên lãnh thổ Việt Nam, sẵn sàng cho các ứng dụng như cập nhật bản đồ quy hoạch đô thị.



Hình 19: So sánh kết quả dự đoán (đường màu đỏ) và đối tượng thực (màu vàng)

2.5. Đề xuất quy trình ứng dụng công nghệ học sâu vào sản xuất bản đồ địa hình và dữ liệu nền địa lý

Dựa trên các kết quả thực hiện thực nghiệm, nghiên cứu đề xuất quy trình thành lập bản đồ chuyên đề tích hợp công nghệ học sâu nhằm ứng dụng vào thực tiễn sản xuất dữ liệu địa lý. Quy trình được thiết kế trong Hình 20 đảm bảo sự liên kết chặt chẽ từ khâu ngoại nghiệp đến sản phẩm cuối cùng. Quy trình thành lập bản đồ chuyên đề từ ảnh UAV tích hợp công nghệ học sâu có thể được tóm tắt thông qua 5 bước sau:

Bước 1: Chuẩn bị và Thu nhận dữ liệu ảnh UAV

Đây là giai đoạn tiền đề, thực hiện theo quy chuẩn ngoại nghiệp:

Thiết kế bay chụp: Xác định khu vực, lựa chọn thiết bị UAV phù hợp và thiết lập thông số bay (độ cao, độ phủ) để đảm bảo độ phân giải thực địa theo thiết kế. (ví dụ độ phân giải 10cm).

Đo đạc không chế ảnh (GCP- ground control point): Bố trí và đo tọa độ các điểm khống chế mặt đất bằng công nghệ GNSS/RTK để đảm bảo độ chính xác hình học cho ảnh số theo Thông tư 07/2021/TT-BTNMT.

Bước 2: Xử lý nội nghiệp và Tạo bình đồ ảnh

Chuyển đổi dữ liệu ảnh thô thành dữ liệu không gian chuẩn:

Xử lý ảnh: Nắn chỉnh hình học, bình sai khối ảnh dựa trên các điểm khống chế mặt đất hoặc sử dụng tâm ảnh xác định bởi GNSS-RTK.

Sản phẩm trung gian: Xây dựng mô hình số độ cao (DEM/DSM) và thành lập bình đồ ảnh trực giao (Orthophoto). Đây là dữ liệu đầu vào trực tiếp cho mô hình học sâu.

Bước 3: Trích xuất đối tượng tự động (AI Inference)

Đây là khâu cải tiến then chốt thay thế cho việc số hóa thủ công: Sử dụng bộ trọng số đã được đào tạo bởi mô hình học sâu trên bộ mẫu chuẩn (đã nêu ở mục 2.5). Theo đó từ ảnh UAV đầu vào, mô hình sẽ tiến hành khoan vùng các tòa nhà trực tiếp trên ảnh.

Bước 4: Biên tập và Hiệu chỉnh dữ liệu

Chuẩn hóa dữ liệu từ AI để đạt tiêu chuẩn bản đồ địa hình:

Nắn chỉnh ranh giới hình học tòa nhà (Regularization): Sử dụng các thuật toán bổ sung để khử hiện tượng "răng cưa" do pixel ảnh, đưa các cạnh tòa nhà về dạng thẳng và vuông góc.

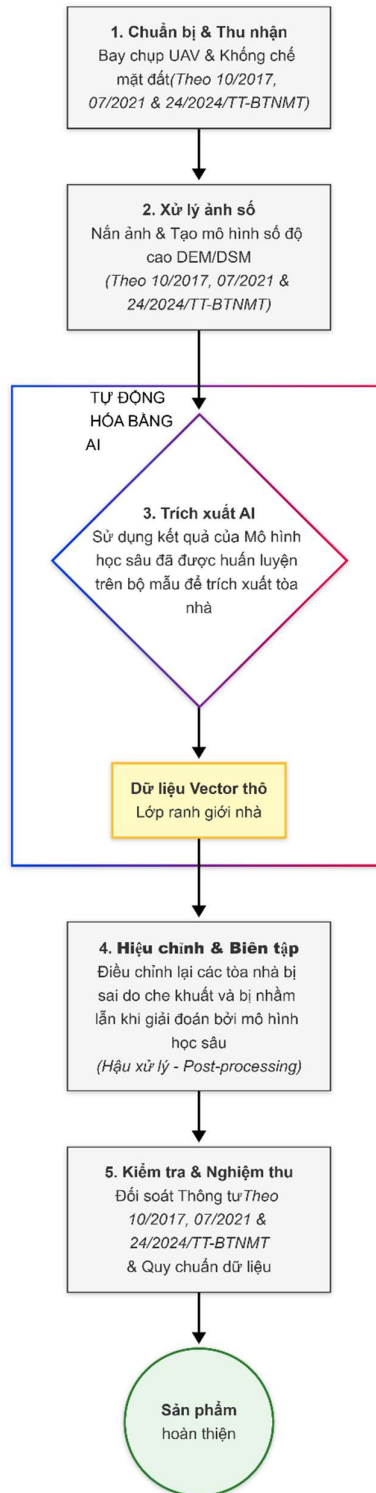
Xử lý thủ công: Kỹ thuật viên đối soát để bổ sung các vị trí bị che khuất bởi bóng đổ, tán cây hoặc xóa bỏ các vùng mô hình nhận diện sai.

Bước 5: Kiểm tra chất lượng và Nghiệm thu

Đảm bảo tính pháp lý và độ chính xác của sản phẩm:

Đối soát tiêu chuẩn: Kiểm tra sai số vị trí mặt bằng đối tượng địa lý theo Thông tư 07/2021/TT-BTNMT, 10/2021/TT-BTNMT.

Hoàn thiện: Trình bày bản đồ chuyên đề, gán thuộc tính và đóng gói cơ sở dữ liệu nền địa lý để nghiệm thu theo quy định của Thông tư 10/2017/TT-BTNMT và 24/2024/TT-BTNMT.



Hình 20. Quy trình thành lập bản đồ chuyên đề từ ảnh UAV tích hợp công nghệ học sâu.

2.6. Đánh giá khả năng thành lập bản đồ và cơ sở dữ liệu nền địa lý sử dụng công nghệ trích xuất tự động AI trên ảnh UAV

Sau khi đã xây dựng và thực hiện các thực nghiệm đánh giá mô hình học sâu trong việc trích xuất đối tượng tòa nhà tại Chương II, việc xác lập khả năng ứng dụng thực tiễn của công nghệ này vào quy trình sản xuất dữ liệu địa lý là một yêu cầu cấp thiết. Để chuyển đổi từ các chỉ số đánh giá của máy học sang các chỉ tiêu kỹ thuật chuyên ngành đo đạc bản đồ, cần có một sự đối soát khách quan dựa trên các quy định pháp lý hiện hành. Mục này sẽ tập trung phân tích mối liên hệ toán học giữa độ chính xác trích xuất diện tích (IoU) và sai số vị trí điểm, đồng thời đánh giá mức độ đáp ứng của dữ liệu ảnh UAV đối với các tiêu chuẩn thành lập bản đồ và cơ sở dữ liệu nền địa lý quốc gia theo quy chuẩn của Bộ Tài nguyên và Môi trường (nay là bộ Nông nghiệp và Môi trường).

Trong thành lập bản đồ địa hình từ ảnh UAV, sai số được đánh giá qua vị trí điểm, trong khi đánh giá độ chính xác mô hình học sâu trong đề tài sử dụng đánh giá thông qua diện tích. Do đó, cần tiến thể thiết lập mối liên hệ giữa sai số vị trí và sai số diện tích thông qua mô hình hình học.

Giả sử tòa nhà có hình vuông có cạnh chiều dài L . Nếu mô hình học sâu trích xuất có sai số biên (Δs) (là độ dịch chuyển giữa ranh giới tòa nhà trong thực tế và tòa nhà được trích xuất theo mô hình), ta có:

Diện tích vùng giao giữa kích thước tòa nhà thực (A) và tòa nhà dự đoán (P) là:

$$A \cap P = (L - \Delta s)^2 \quad (1)$$

Diện tích vùng hợp:

$$A \cup P = L^2 + L^2 - (L - \Delta s)^2 = L^2 + 2L\Delta s - \Delta s^2 \quad (2)$$

Sai số đánh giá mô hình học sâu theo công thức IoU sẽ được tính theo kích thước tòa nhà L và độ lệch Δs như sau:

$$IoU = \frac{|A \cap P|}{|A \cup P|} = \frac{(L - \Delta s)^2}{L^2 + 2L\Delta s - \Delta s^2} \quad (3)$$

Từ đó thiết lập được giá trị độ lệch theo vị trí của tòa nhà từ giá trị độ chính xác của mô hình IoU là:

$$\Delta s \approx \frac{L}{2} \left(\frac{1 - IoU}{1 + IoU} \right) \quad (4)$$

Đối với độ chính xác dự đoán tòa nhà đối với các đặc điểm tòa nhà khác nhau được ở Việt Nam nhìn chung có độ chính xác trên 85%. Vì thế khi áp dụng công thức 7 với giả sử kích thước tòa nhà có chiều có kích thước ($L=10m$) thì độ lệch về vị trí có thể được xác định bởi:

$$\Delta s \approx \frac{10}{2} \left(\frac{1 - 0.85}{1 + 0.85} \right) = 0.4(m)$$

Căn cứ trên các tiêu chuẩn kỹ thuật tại Điều 5 và Phụ lục của Thông tư 07/2021/TT-BTNMT về sai số vị trí mặt bằng đối tượng địa lý, nghiên cứu thực hiện đối soát và đánh giá khả năng ứng dụng thực tế của mô hình như sau:

Về độ chính xác hình học: Đối với bản đồ địa hình tỷ lệ 1:1.000 và 1:2.000, sai số vị trí mặt bằng cho phép ngoài thực địa tương ứng là 0,5m và 1,0m. Kết quả phân tích toán học từ thực nghiệm cho thấy, với chỉ số IoU đạt trên 85%, sai số dịch chuyển biên trung bình được kiểm soát trong ngưỡng 0,4m. Giá trị này hoàn toàn nằm trong giới hạn cho phép, đảm bảo đáp ứng yêu cầu khắt khe về sai số vị trí điểm của bản đồ tỷ lệ 1:1.000.

Về độ phân giải dữ liệu: Theo quy định, kích thước pixel thực địa (GSD) yêu cầu cho bản đồ tỷ lệ 1:1.000 và 1:2.000 lần lượt là 10cm và 20cm. Như vậy, dữ liệu ảnh số từ UAV với độ phân giải 10cm được sử dụng trong nghiên cứu không chỉ tối ưu cho việc huấn luyện mô hình học sâu mà còn đáp ứng đầy đủ tiêu chuẩn đầu vào để thành lập bản đồ địa hình và cập nhật cơ sở dữ liệu nền địa lý quốc gia tỷ lệ 1:1.000.

Mặc dù quy trình trích xuất đối tượng tòa nhà đã được tự động hóa bằng công nghệ học sâu, tuy nhiên, để đảm bảo độ chính xác hình học và tính nhất quán của dữ liệu theo các quy chuẩn kỹ thuật hiện hành, công tác hậu kiểm và hiệu chỉnh là bắt buộc. Do ảnh hưởng của độ phân giải không gian (kích thước pixel), ranh giới đối tượng trích xuất trực tiếp từ mô hình AI thường xuất hiện hiện tượng 'răng cưa' tại các cạnh thẳng. Vì vậy, các đa giác ranh giới tòa nhà cần được chuẩn hóa thông qua thuật toán nắn thẳng (Regularization) để khôi phục tính hình học nguyên bản. Bên cạnh đó, các sai số cục bộ do mô hình nhận diện nhầm hoặc các khu vực tòa nhà bị che khuất bởi bóng đổ và tán cây cần được xử lý thông qua thao tác biên tập thủ công. Sự kết hợp giữa năng lực tính toán tự động của AI và quy trình kiểm soát chất lượng từ con người đảm bảo sản phẩm cuối cùng đạt độ tin cậy cao, đáp ứng đầy đủ các sai số cho phép trong thành lập bản đồ địa hình và cơ sở dữ liệu nền địa lý.

KẾT LUẬN VÀ KIẾN NGHỊ

1. Kết luận

Nghiên cứu về việc xây dựng quy trình tự động trích xuất đối tượng tòa nhà từ ảnh UAV bằng phương pháp học sâu đã hoàn thành các mục tiêu đề ra và đạt được những kết quả quan trọng sau:

Về quy trình công nghệ: Nghiên cứu đã đề xuất thành công quy trình sản xuất dữ liệu địa lý tích hợp học sâu theo mô hình bậc thang. Quy trình này không chỉ có cơ sở khoa học vững chắc mà còn bám sát các tiêu chuẩn kỹ thuật của Thông tư 07/2021/TT-BTNMT, 07/2021/TT-BTNMT và Thông tư 24/2024/TT-BTNMT, tạo tiền đề cho việc chuyển đổi số và tự động hóa trong ngành đo đạc bản đồ.

Về tính chính xác và thực tiễn: Thông qua việc thiết lập mối liên hệ toán học giữa chỉ số diện tích (IoU) và sai số vị trí điểm, nghiên cứu đã chứng minh được rằng với mức độ chính xác $\text{IoU} > 85\%$, sai số vị trí mặt bằng đạt ngưỡng 0,4m. Kết quả này khẳng định khả năng ứng dụng của mô hình trong việc thành lập bản đồ địa hình và cơ sở dữ liệu nền địa lý tỷ lệ 1:1.000 và 1:2.000.

Về hiệu quả sản xuất: Quy trình đề xuất minh chứng tính ưu việt trong việc giải quyết bài toán trích xuất dữ liệu tự động, giúp giảm đáng kể thời gian lao động nội nghiệp, góp phần giảm giá thành, tăng năng suất và đáp ứng kịp thời nhu cầu dữ liệu địa lý hiện nay.

2. Kiến nghị

Để nâng cao hơn nữa độ chính xác và khả năng tổng quát hóa, đưa quy trình vào ứng dụng rộng rãi trong sản xuất thực tế, nghiên cứu đề xuất một số định hướng sau:

Về công tác xây dựng bộ mẫu: Cần mở rộng và chuẩn hóa công tác xây dựng bộ dữ liệu mẫu (Ground Truth), đảm bảo tính đồng nhất và đại diện cho sự đa dạng về đặc điểm hình học, quang phổ của các vùng miền khác nhau tại Việt Nam. Đây là yếu tố then chốt để nâng cao khả năng học của mô hình.

Về tối ưu hóa công nghệ: Cần nghiên cứu tích hợp thêm các thuật toán hậu xử lý (Regularization) nâng cao để tự động hóa hoàn toàn khâu nắn thẳng cạnh đối tượng, giảm thiểu sai số răng cưa do độ phân giải ảnh. Đồng thời, tiếp tục thử nghiệm các kiến trúc học sâu tiên tiến (như các biến thể của Transformer hoặc DeepLabv3+) để cải thiện độ chính xác tại các khu vực có mật độ nhà ở phức tạp hoặc bị che khuất cao.

Về mở rộng đối tượng: Kiến nghị mở rộng quy trình trích xuất tự động cho các đối tượng địa lý khác (giao thông, thủy văn, thực vật) để hoàn thiện hệ thống trích xuất dữ liệu tự động cho cơ sở dữ liệu nền địa lý quốc gia.

TÀI LIỆU THAM KHẢO

- Chen, L.-C., Y. Zhu, G. Papandreou, F. Schroff and H. Adam (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. Proceedings of the European conference on computer vision (ECCV).
- Cường, N. S., T. X. Trường, T. H. Hạnh and Đ. N. J. T. c. K. h. K. t. M.-Đ. c. Vũ, 60 (2019). "Nâng cao chất lượng xây dựng mô hình 3d bằng kết hợp công nghệ bay chụp uav và quét laser mặt đất." 10.
- He, K., X. Zhang, S. Ren and J. Sun (2016). Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition.
- Lê, V. C., T. T. H. Lê and X. C. Cao (2020). "Nghiên cứu lựa chọn vị trí cất cánh cho thiết bị bay không người lái tích hợp gnss động phục vụ đo vẽ thành lập bản đồ địa hình tỷ lệ lớn cho các mỏ lộ thiên= research on optimal takeoff positions of uav integrated gnss-rtk in producing large scale topological maps for open-pit mines."
- Long, N. Q. and C. X. J. T. c. c. n. m. Cường (2019). "Ứng dụng máy bay không người lái (uav) để xây dựng mô hình số bề mặt và bản đồ mỏ lộ thiên khai thác vật liệu xây dựng." 1: 21-29.
- Long, N. Q., R. Goyal, L. K. Bui, C. X. Cuong, L. V. Canh, N. Q. Minh and X. N. J. A. o. M. S. Bui (2021). "Optimal choice of the number of ground control points for developing precise dsm using light-weight uav in small and medium-sized open-pit mine." 66(3).
- Sỹ, M. V., B. N. Quý and P. V. J. T. c. K. h. Đ. đ. v. B. đ. Hiệp (2017). "Nghiên cứu sử dụng dữ liệu ảnh máy bay không người lái (uav) trong thành lập bản đồ địa hình tỷ lệ lớn." (33): 49-57.
- Yu, W., K. Yang, Y. Bai, T. Xiao, H. Yao and Y. Rui (2016). Visualizing and comparing alexnet and vgg using deconvolutional layers. Proceedings of the 33 rd International Conference on Machine Learning.