

# NGHIÊN CỨU, ỨNG DỤNG MẠNG D-LINKNET ĐỂ GIẢI QUYẾT BÀI TOÁN PHÁT HIỆN ĐƯỜNG TỪ ẢNH VỆ TINH WORLDVIEW-3 TRÊN KHU VỰC QUẬN CẦU GIẤY, HÀ NỘI

## (RESEARCH AND APPLICATION OF D-LINKNET NETWORK TO SOLVE THE PROBLEM OF DETECTING ROAD FROM WORLDVIEW-3 SATELLITE IMAGE IN CAU GIAY DISTRICT, HANOI)

Lê Quang Đạo, Nguyễn Nhật Nam, Lưu Quang Thắng, Hà Đức Văn, Hà Minh Cường,

Phạm Bảo Sơn, Bùi Quang Hưng

Trung tâm Công nghệ tích hợp liên ngành Giám sát hiện trường

Trường ĐH Công Nghệ - ĐH Quốc Gia Hà Nội

Email: daolq@fimo.edu.vn

**Abstract:** In this paper, we have outlined the need for solving the road detection problem, from which to compare and find the type of data to be used as Worldview-3 satellite images, to find out the method of success. Very good achievement in the DeepGlobe contest is D-LinkNet. This thesis has collected Worldview-3 and yandex image data for preprocessing data. We have collected and pre-processed Worldview-3 data, deployed and installed D-LinkNet. Application of evaluation method of IoU point of DeepGlobe. In addition, learn how to develop a realistic assessment method rather than a long-distance evaluation. D-LinkNet works well in sparsely populated areas, big roads such as suburbs or new urban areas, resettlement areas. In densely populated urban areas, small alleys are obscured by houses and trees, making it difficult and difficult to detect roads. Some areas with the same spectrum as the road are also mislabeled.

**Keyword:** satellite image; Worldview-3; D-LinkNet; road detection; evaluation methods.

## 1. GIỚI THIỆU

Bản đồ số Vmap nằm trong đề án "Phát triển hệ tri thức Việt số hóa", do Bộ Khoa học và Công nghệ chủ trì, tổng công ty Bưu điện Việt Nam (VNPOST) kết hợp cùng Đại học Quốc gia Hà Nội, Trung ương Đoàn Thanh niên Cộng sản Hồ Chí Minh xây dựng và triển khai. Vmap được ra mắt ngày 01/10/2019 và có 23.4 triệu dữ liệu địa chỉ trên cả nước. Để có được dữ liệu này, 120000 nhân viên bưu điện và thanh niên, đoàn viên đã đi từng khu phố, thôn bản để thu thập thông tin. Các dữ liệu bao gồm thông tin chi tiết của địa điểm và ghi chú. Bên cạnh đó Vmap còn sử dụng dữ liệu lớn từ OpenStreetMap [1].

Bảng 1: Thống kê độ dài đường bộ của OSM và Google tại quận Cầu Giấy

	OSM	Google
Tổng độ dài đường bộ (Km)	283,49	349,32

So sánh dữ liệu đường hiện có của OSM và Google tại khu vực quận Cầu Giấy, ta thấy đối với Google, OSM còn thiếu rất nhiều đường. Bởi vì sử dụng dữ liệu đường phụ thuộc vào người dùng đóng góp nên lượng dữ liệu bị thiếu là điều thường xuyên xảy ra. So với thực tế, trong khu

vực nội thành hoặc khu đông đúc dân cư, Google bị thiếu nhiều ngõ ngách nhỏ; khu vực ngoại thành và các tỉnh thành hay các khu thưa thớt dân, các con đường mới xây dựng chưa được cập nhật. Vì thế, để so sánh với thực tế, dữ liệu mà Vmap đang sử dụng thiếu rất nhiều về cả các ngõ ngách nhỏ và các con đường mới thi công.

Về phần tốc độ cập nhật đường, dữ liệu OSM sử dụng có tốc độ cập nhật rất thấp. Chỉ khi nào người dùng chỉnh sửa thủ công hoặc tải lên bộ dữ liệu thì dữ liệu mới được cập nhật. Khu vực được cập nhật khá hạn chế, đa phần khu vực được cập nhật là khu vực nội thành, nơi có đông đúc dân cư và có nhiều tình nguyện viên. Các khu vực ngoại thành, các tỉnh hoặc vùng sâu, vùng xa ít khi được cập nhật dữ liệu mới dẫn đến các khu vực này thiếu khá nhiều đường.

Điều này gây ra ảnh hưởng không nhỏ đến chất lượng dịch vụ bản đồ:

- Ảnh hưởng đến chức năng dẫn đường, gây ra những gợi ý đường xa hơn thực tế hay thậm chí là đi sai đường (đi vào đường cấm, ngược chiều v..v..)
- Ảnh hưởng đến chức năng nội suy số nhà: Chức năng nội suy tìm số nhà sử dụng dữ liệu đường bộ để có thể nội suy một cách chính xác. Nếu dữ liệu địa chỉ có tồn tại nhưng lại không có dữ liệu đường thì việc nội suy là vô ích vì không thể xác định được cấu trúc dãy nhà.

Từ các ảnh hưởng đến chất lượng dịch vụ bản đồ dẫn đến trải nghiệm không tốt của người dùng hoặc thậm chí là ảnh hưởng đến doanh thu của các doanh nghiệp vận tải sử dụng chính bản đồ đó.

## 2. CÁC NGHIÊN CỨU LIÊN QUAN

Tiếp cận bài toán phân loại điểm ảnh từng lớp với mô hình mạng học sâu có thể chia làm các loại sau dựa trên đặc điểm kỹ thuật của chúng [2]:

1. Fully convolutional networks
2. Convolutional models with graphical models
3. Encoder-decoder based models
4. Multi-scale and pyramid network based models
5. R-CNN based models
6. Dilated convolutional models and DeepLab family
7. Recurrent neural network based models
8. Attention-based models
9. Generative models and adversarial training
10. Convolutional models with active contour models
11. Các mô hình khác

Trong đó phương pháp thứ ba Encoder-Decoder based models đang có những kết quả khả quan. Minh chứng là trong cuộc thi Road Extraction Deep Globe năm 2018. Mạng D-LinkNet thuộc phương pháp này đã đạt giải quán quân, vượt trên kết quả của các mạng khác [3].

## 3. THỰC NGHIỆM

Trong phần này, chúng tôi tiến hành các thực nghiệm về lựa chọn dữ liệu ảnh vệ tinh, xây dựng, thu thập và tiền xử lý dữ liệu, triển khai mạng D-LinkNet và các phương pháp đánh giá.

### 3.1. Lựa chọn dữ liệu ảnh vệ tinh và giải pháp phát hiện đường

Trên thế giới có rất nhiều loại ảnh vệ tinh khác nhau. Mỗi loại ảnh vệ tinh có thông số khác nhau và có nhiều ứng dụng trong nhiều lĩnh vực. Ảnh vệ tinh chia làm 2 loại chính là ảnh radar và ảnh quang phổ. Ảnh radar có lợi thế hình học và không bị cản trở bởi mây hay ánh sáng tuy nhiên sự tương đồng lớn giữa đường và các công trình xây dựng là sự cản trở rất lớn. Ảnh quang phổ chứa nhiều thông tin hơn ảnh radar và có thể tạo bộ dữ liệu dễ dàng hơn và phổ biến hơn nhiều. Theo đó sự lựa chọn ở khóa luận này là ảnh quang phổ.

*Bảng 2: Một số vệ tinh viễn thám thương mại hiện có [4]*

Độ phân giải thấp(>20m)	Độ phân giải cao(2,5 - 20m)	Độ phân giải siêu cao(<2.5m)
Landsat-8	Sentinel-1,2	WorldView-2,3,4
RISAT-1	FORMOSAT-2	Gaofen-1
IRS-1A,!D,P6	VNRedSat-1	QuickBird-2
	SPOT-5	GeoEye-1
	RapidEye	KOMPSAT-3
	Theos	Pleiades-1A

Một trong những thông số quan trọng nhất của ảnh vệ tinh là độ phân giải không gian. Độ phân giải không gian của ảnh vệ tinh là kích thước nhỏ nhất của một đối tượng hay khoảng cách tối thiểu giữa hai đối tượng liền kề có khả năng phân biệt được trên ảnh. Độ phân giải không gian của ảnh vệ tinh cũng có thể hiểu là kích thước của một pixel với hệ quy chiếu trên mặt đất. Thông thường, ảnh vệ tinh có độ phân giải không gian càng cao thì ảnh càng rõ nét và dễ phân biệt các vật thể trong ảnh hơn.

Với ảnh vệ tinh có độ phân giải siêu cao, ảnh sẽ rõ nét hơn và các vật thể trong ảnh sẽ rõ ràng hơn với những ảnh vệ tinh còn lại. Do đặc điểm của bài toán tự động phát hiện đường từ ảnh vệ tinh thuộc dạng bài toán phân loại điểm ảnh (phân loại pixel) nên ảnh càng rõ nét, mỗi con đường chứa càng nhiều điểm ảnh thì khả năng phát hiện được đúng đường càng cao. Vì vậy ảnh vệ tinh sử dụng cho bài toán phát hiện đường từ ảnh vệ tinh nên là loại có độ phân giải siêu cao

*Bảng 3: Bảng so sánh một số vệ tinh độ phân giải cao [4]*

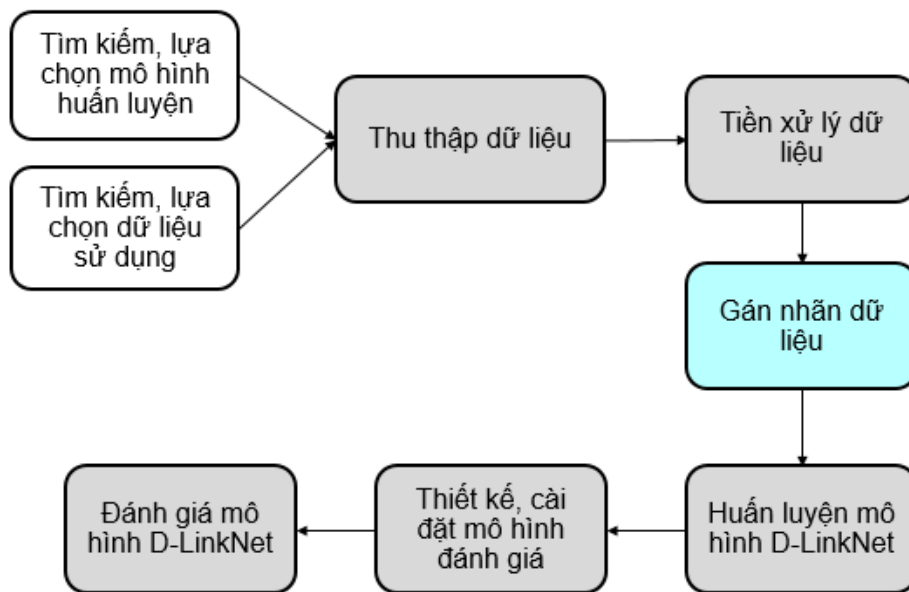
Tên vệ tinh	Số dải phổ	Độ phân giải panchromatic(m)	Độ phân giải phổ khác(m)	Thời gian chụp mới nhất tại Hà Nội
WorldView-3	16	0.31	1.2	2020
GeoEye-1	5	0.46	1.84	2020

WorldView-2	9	0.46	1.84	2020
Pleiades-1A	5	0.5	2	2020
QuickBird-2	5	0.61	2.62	2014

Bảng 3 tìm hiểu và thống kê dữ liệu ảnh vệ tinh trên thế giới có độ phân giải cao và xét chung trong điều kiện độ che phủ mây nhỏ hơn 5%. Dựa vào bảng nhận thấy được ảnh Worldview 3 là ảnh phù hợp nhất vì đáp ứng được tất cả các tiêu chí và nổi trội hơn so với các ảnh khác về độ phân giải và số phổ.

### 3.2. Quy trình ứng dụng mạng D-LinkNet cho khu vực quận Cầu Giấy

Nghiên cứu này đã xây dựng được quy trình ứng dụng mạng D-LinkNet để đánh giá có phù hợp với khu vực Cầu Giấy hay không:



### 3.3. Xây dựng, thu thập và tiền xử lý dữ liệu

Dữ liệu ảnh Worldview-3 được thu thập từ công ty Apollo Mapping. Lựa chọn khu vực cần lấy mẫu bằng cách chọn 2 điểm trái trên và phải dưới trên bản đồ tạo thành shapefile khu vực cần thu thập ảnh. Khu vực được lựa chọn có tọa độ: 21.05536 105.77002 đến 21.00472 105.81327 bao phủ hầu hết các phường thuộc địa phận quận Cầu Giấy.

Ảnh WorldView 3 đã thu thập có độ phân giải 15060x16384. Sau khi cắt nhỏ với độ phân giải 1024x1024 để sử dụng cho công việc huấn luyện thì chỉ được 224 bộ dữ liệu. Bộ dữ liệu này quá ít để huấn luyện và test. Đối với một mạng học sâu, việc có nhiều dữ liệu là rất quan trọng, điều này ảnh hưởng rất nhiều đến kết quả. Từ đây sinh ra vấn đề cần giải quyết là tìm thêm dữ liệu ảnh ngoài ảnh Worldview 3. Tuy nhiên có rất nhiều dịch vụ bản đồ cung cấp dữ liệu ảnh vệ tinh miễn phí, việc lựa chọn dịch vụ nào phụ thuộc vào độ tương đồng với ảnh Worldview 3, độ nghiêng của vật thể, thời gian chụp và độ che phủ mây. Giải pháp đưa ra là sử dụng ảnh Yandex bởi Yandex thỏa mãn toàn bộ các tiêu chí đề ra. Ảnh viễn thám được tải về từ map yandex thông qua phần

mềm SAS Planet. Việc tải ảnh về thông qua việc lựa chọn 2 điểm góc trên trái và góc dưới phải của bản đồ và các thông số độ zoom, hệ quy chiếu. Thông số lựa chọn là độ zoom 21 trên hệ quy chiếu WGS84 / EPSG 4326.

Xảy ra vấn đề là độ phân giải mặt đất band RGB của ảnh Worldview-3 là 1.2m, độ phân giải này quá nhỏ để có thể phát hiện ra những con đường. Pansharpening là phép tăng cường sử dụng ưu điểm của hai loại ảnh là toàn sắc (Pan\_Panchromatic) có độ phân giải không gian cao và ảnh đa phổ (MS\_MultiSpectral) có màu sắc nhưng độ phân giải thấp hơn; kết hợp cả hai loại ảnh sẽ được dữ liệu ảnh mới có màu sắc và độ phân giải cao. Trong nghiên cứu này, chúng tôi sử dụng phương pháp Pansharpening Brovey của phần mềm ENVI để tăng cường chất lượng ảnh RGB Worldview-3.

### 3.4. Triển khai mạng D-LinkNet

Sử dụng Google Colab để triển khai cài đặt .

Môi trường sử dụng là python 2.7.

Thời gian huấn luyện: 18h.

Bộ dữ liệu huấn luyện bao gồm 2243 ảnh.

Nhược điểm của google colab là giới hạn thời gian sử dụng GPU là 12 giờ. Từ đó phải lưu model để sau khi sử dụng hết 12 giờ ta có thể huấn luyện tiếp [4].

Sau khi huấn luyện xong, sử dụng 50 bộ ảnh để test. Kết quả sẽ lấy trung bình trên bộ 50 ảnh này.

### 3.5. Phương pháp đánh giá

Trong cuộc thi DeepGlobe có sử dụng phương pháp đánh giá là điểm IoU qua các pixel [5]. Tuy nhiên, trong thực tế, dữ liệu đường được lưu dưới dạng node và way. Khi đó độ rộng hay số lượng pixel không còn quan trọng nữa. Phương pháp đánh giá qua điểm IoU xuất hiện điểm trừ là trường hợp các pixel được gán nhãn là đường phủ trên khắp các đoạn đường trong ảnh nhưng không phủ rộng kín con đường đó lại không được đánh giá cao hơn trường hợp các pixel được gán nhãn kín các con đường lớn nhưng không được gán nhãn ở các đường nhỏ. Nguyên nhân bởi phương pháp IoU tính điểm dựa trên số lượng pixel trùng khớp. Thực tế cần số lượng đường hơn là độ chính xác đến từng pixel của đường. Một phương pháp đánh giá mới được khóa luận này đề xuất ra là đánh giá dựa trên độ dài của đường.

Xương sống trong phương pháp này toán tử skeleton. Các con đường được trích xuất hình dạng đặc trưng, khi đó các hình ảnh đường không bị phụ thuộc vào độ rộng hay số lượng pixel bao phủ.

Ứng dụng toán tử skeleton vào đánh giá độ dài đường ta làm theo quy trình sau:

Dữ liệu đầu vào: Bao gồm 2 thư mục ảnh nhãn của tập test và nhãn được sinh ra,

- Bước 1: Nạp 2 ảnh đầu vào.
- Bước 2: Tạo ra 3 ảnh nhị phân thể hiện 3 loại đường: đường đã phát hiện khớp với nhãn tập test (ảnh 1), đường đã phát hiện ngoài tập test (ảnh 2), đường của nhãn tập test (ảnh 3).

Phương pháp: Sử dụng các hàm numpy logical bởi các hàm này tính toán với tốc độ rất nhanh và chuyển ảnh về dạng nhị phân là yêu cầu đầu vào của toán tử skeleton.

Xét ví dụ chồng 2 ảnh lên nhau, coi 2 đối tượng đường của 2 ảnh là A và B. Phần A là đường của nhãn tập test, phần B là đường của nhãn được sinh ra. Đối với đường đã phát hiện khớp với nhãn của tập test, ta trích xuất phần giao của 2 đối tượng. Đối với đường đã phát hiện ngoài tập test, trích xuất phần bù của B. Đường nhãn của tập test chỉ cần chuyển về ảnh nhị phân

- Bước 3: Lọc nhiễu: Có nhiều khu vực phát hiện bị nhiễu như các pixel đường nhỏ lẻ, rời rạc, không mang lại nhiều thông tin. Cần loại bỏ các pixel này để tránh ảnh hưởng đến kết quả. Ta dùng bộ lọc trung vị, bởi bộ lọc này hiệu quả với nhiễu dạng salt and pepper noise.
- Bước 4: Sử dụng toán tử skeleton để lấy khung xương của đường
- Bước 5: Tính toán các số liệu:

Sau khi đã có khung xương của đường. Đếm số pixel là đường của 3 ảnh tính được 3 giá trị sau:

- True positive: số pixel đường của ảnh 1
- False positive: số pixel đường của ảnh 2
- False negative: số pixel đường ảnh 3 trừ đi số pixel đường ảnh 1

Với mỗi pixel tương ứng với độ phân giải mặt đất nên chỉ cần đếm số pixel đã thể hiện được độ dài của con đường.

#### 4. KẾT QUẢ THỰC NGHIỆM

*Bảng 4: Kết quả đánh giá theo điểm IoU*

Precision	Recall	F1	IoU
80.32%	72.82%	73.58%	65%

*Bảng 4: Kết quả đánh giá theo độ dài đường*

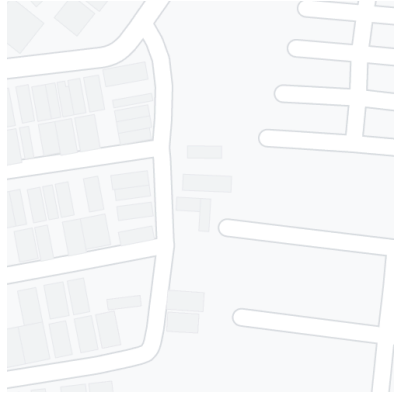
Precision	Recall	F1
57.54%	85.19%	64.74%



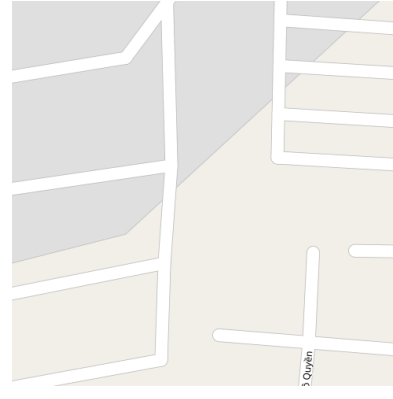
*Hình 1: Ảnh thực tế*



*Hình 2: Nhãn dự đoán*



Hình 3: Đường đã có của Google



Hình 4: Đường đã có của OSM

Dựa vào 4 hình 1, 2, 3, 4 có thể thấy rằng tại ngõ bên trái thứ 3 từ trên xuống, đường bị gián đoạn bởi bóng của tòa nhà đổ trên đường. So sánh với Google và đường đã có sẵn của OSM nhận thấy khu vực bên phải của ảnh, nơi có dân cư đông đúc hơn bên trái, các con đường rất nhỏ, xen kẽ giữa các tòa nhà sát nhau. Điều này gây khó khăn cho mạng D-LinkNet và không phát hiện được đường từ đó.



Hình 5: Ảnh thực tế có sân



Hình 6: Nhãn dự đoán

Dựa vào hình 4.11 và 4.12, đối với trường hợp này, có thể nhận thấy khu vực sân bê tông của kho xưởng có phổ tương đương với đường bộ và kết quả D-LinkNet đã gán nhãn sai cho khu vực sân này. Và cũng trong trường hợp này có thể dễ dàng nhận ra cây xanh che phủ phần lớn diện tích gây ra sự không liên mạch của các con đường được gán nhãn.

Xét ảnh hưởng của phương tiện giao thông trên đường:



Hình 7: Ảnh thực tế có phương tiện



Hình 7: Ảnh nhãn dự đoán

Trong ảnh 4.13 có nhiều phương tiện giao thông di chuyển trên đường, có cả xe khách lớn. Tuy nhiên không ảnh hưởng đến kết quả phân đoạn ngõ nghĩa.

Từ kết quả ta rút được nhận xét như sau:

- D-LinkNet phát hiện đường qua đặc trưng phổ nên những khu vực là mặt đường nhựa không bị che khuất bởi bóng hay nhà cửa có thể phát hiện ra dễ dàng. Chủ yếu là các con đường ngoại thành, khu dân cư thưa hoặc khu đô thị mới có chiều rộng lớn hơn 5 mét.
- Tuy nhiên vì phát hiện qua đặc trưng phổ nên một số khu vực có phổ tương tự đường cũng bị gán nhãn sai là đường. Ví dụ như sân nhà, bãi đỗ xe, v..v..
- Khu vực bị che chắn có phổ khác với phổ của đường nhựa nên khó phát hiện ra
- Đường dễ bị ngắt quãng bởi sự che chắn của cây xanh và nhà cửa.
- Kết quả đạt được chưa được như mong đợi.

Từ kết quả trích xuất đường, ta phải xử lý thêm một quy trình nữa là chuyển đường từ ảnh phát hiện thành dạng node, way, relation để có thể lưu trong cơ sở dữ liệu

## 5. KẾT LUẬN VÀ KIẾN NGHỊ

Chúng tôi đã nghiên cứu tìm ra phương pháp ứng dụng mạng D-LinkNet để phát hiện đường trên ảnh worldview 3 trên khu vực Cầu Giấy. Thu thập và xử lý được dữ liệu ảnh worldview 3 và ảnh Yandex. Bên cạnh đó, chúng tôi đã tìm và thực thi được 2 phương pháp đánh giá theo điểm IoU và đánh giá theo độ dài đường.

Tuy chưa được như mong đợi nhưng kết quả đạt được của nghiên cứu này là tiền đề cho các nghiên cứu tiếp theo để phát triển xây dựng bộ dữ liệu đường cho Vmap.

- Do thực tế ở Việt Nam, các con đường bị che chắn nhiều do cây xanh và nhà cửa, dẫn đến phát hiện đường trong khu vực nội thành gặp nhiều khó khăn.
- Mạng D-LinkNet được tối ưu cho bộ dữ liệu DeepGlobe chứ chưa phải tối ưu cho dữ liệu ở Việt Nam.

Trong tương lai, nghiên cứu hướng đến chỉnh sửa mạng D-LinkNet cho phù hợp điều kiện ở Việt Nam, hiện tại có Facebook đang phát triển và chỉnh sửa mạng D-LinkNet này và cho kết quả khả quan ở Việt Nam. Bên cạnh đó, bổ sung thêm dữ liệu là điều cần thiết (hiện tại chỉ có 2243 ảnh đào tạo) làm tăng độ chính xác cho một mạng học sâu như D-LinkNet...



## LỜI CẢM ƠN

Chúng tôi xin cảm ơn dự án nghiên cứu VT / UD-06 / 16-20 trong Chương trình Khoa học và Công nghệ Vũ trụ Quốc gia đã hỗ trợ kinh phí thực hiện nghiên cứu này.

## TÀI LIỆU THAM KHẢO

- [1] “vmap.vn.” <https://itrithuc.vn/ban-do-itritruc-app-1.html> (accessed Jun. 28, 2020).
- [2] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, “Image Segmentation Using Deep Learning: A Survey.”
- [3] L. Zhou, C. Zhang, and M. Wu, “D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction,” *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2018-June, pp. 192–196, 2018, doi: 10.1109/CVPRW.2018.00034.
- [4] N. N. Nam, “Nghiên cứu, ứng dụng mạng UNet để giải quyết bài toán phát hiện đường từ ảnh vệ tinh Worldview-3 trên khu vực quận Cầu Giấy, Hà Nội,” University of Engineering and Technology, 2020.
- [5] L. Zhou, Z. C. Zhang, Z. M. Wu, and W. R. Zhang, “DeepGlobe Workshop at CVPR 2018,” 2018.