

ỨNG DỤNG MẠNG EFFICIENTNET TRONG PHÂN LOẠI LỚP PHỦ BỀ MẶT

PHẠM LÊ TUẤN⁽¹⁾, ĐỒNG VŨ HÀ⁽¹⁾, LÊ CHÍ THÀNH⁽²⁾
ĐỖ THỊ PHƯƠNG THẢO⁽³⁾, BÙI QUANG THÀNH⁽¹⁾

⁽¹⁾Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội

⁽²⁾Sở Tài nguyên và Môi trường tỉnh Cà Mau

⁽³⁾Trường Đại học Mở - Địa Chất

Tóm tắt:

Sự ra đời của các phương pháp học máy (các thuật toán phân loại, phân mảnh hay các thuật toán tối ưu hóa) hỗ trợ tự động hóa quá trình phân tích ảnh đã tạo tiền đề cho việc nâng cao độ chính xác trong theo dõi diễn biến sự thay đổi trên bề mặt trái đất. Nghiên cứu này sử dụng bộ dữ liệu mẫu EuroSat để huấn luyện một số biến thể mạng EfficientNet, ứng dụng cho phân loại lớp phủ mặt đất theo cảnh (scene-based classification). Kết quả cho thấy độ chính xác phân loại của Mạng EfficientNet B3, B4, và B5 đạt được tương ứng là 97,7%, 97,74%, 97,9% và mô hình được sử dụng để phân loại một số loại hình lớp phủ của mẫu thử nghiệm tại khu vực Đông bắc Bắc bộ, Việt Nam. Sai số này có thể do bộ mẫu huấn luyện không có một số loại hình lớp phủ tương ứng như tại Việt Nam.

Từ khóa: Lớp phủ mặt đất, Mạng EfficientNet, Dữ liệu vệ tinh

1. Giới thiệu

Trong những năm gần đây phương pháp học sâu (deep learning - DL), mạng nơ-ron sâu (deep neural network - DNN) ngày càng phổ biến nhờ vào khả năng vượt trội trong việc nâng cao độ chính xác cho các bài toán xử lý dữ liệu lớn (Wang, Wang, Zhang, Xiang, & Pan, 2017). DL, DNN được xây dựng trên cơ sở (1) có nhiều lượng dữ liệu hơn để huấn luyện các mô hình, đặc biệt cho các bài toán phân loại cần tập mẫu để kiểm định (2) khả năng phân tích xử lý được tăng cường, đặc biệt là việc sử dụng các bộ xử lý đồ họa (GPUs) được tối ưu cho việc giải quyết các bài toán song song hóa hay xử lý hiệu năng cao và (3)

sự xuất hiện các thuật toán nâng cao cho phép DNNs giải quyết các bài toán chuyên sâu với nhiều đầu ra khác nhau, dẫn đến hiệu suất vượt trội so với các kiến trúc truyền thống. Có nhiều mô hình học sâu được xây dựng, ứng dụng trong nhận diện hình ảnh (thị giác máy tính) hay phân loại lớp phủ từ ảnh vệ tinh, ảnh máy bay hay UAV. Các mô hình EfficientNet được xây dựng và thử nghiệm đem lại độ chính xác cao trong các tập dữ liệu mẫu khác nhau. Nghiên cứu này thử nghiệm và đánh giá khả năng ứng dụng các biến thể của mạng EfficientNet trong phân loại lớp phủ mặt đất. Trong đó sử dụng bộ dữ liệu EuroSat (Helber, Bischke, Dengel, & Borth, 2018) để huấn

luyện mô hình, sau đó thử nghiệm tại khu vực Đông bắc Bắc bộ của Việt Nam.

2. Dữ liệu và phương pháp

2.1. Dữ liệu vệ tinh

Bộ dữ liệu EuroSAT được xây dựng từ các ảnh vệ tinh Sentinel-2 (Helber et al, 2018), một hệ thống quan sát Trái Đất do Cơ quan Vũ trụ Châu Âu (ESA) triển khai. Dữ liệu từ vệ tinh Sentinel-2 cung cấp thông tin về mặt đất với độ phân giải cao và đa phổ, cho phép nhận diện và phân loại các đối tượng trên mặt đất. Dữ liệu từ Sentinel-2 được sử dụng rộng rãi cho nhiều ứng dụng như theo dõi môi trường, quản lý tài nguyên, nông nghiệp, và nhận diện biến đổi đất địa. Bộ dữ liệu EuroSAT được tạo ra bằng cách lựa chọn hình ảnh từ Sentinel-2

và gán nhãn cho từng loại đất địa để tạo thành một tập dữ liệu đa dạng và thách thức cho các mô hình nhận dạng hình ảnh. Bộ dữ liệu bao gồm 27.000 tấm ảnh có tọa độ (kích thước 64 x 64 pixels) với 13 kênh phổ (bộ dữ liệu Multiple spectrals) hoặc bộ RGB, bao gồm 10 loại hình lớp phủ được gán nhãn, thu thập tại các nước Châu Âu. Các loại hình chủ yếu bao gồm Rừng, Cây trồng hàng năm, Đường cao tốc, Thảm thực vật thân thảo, Đồng cỏ, Khu dân cư, Sông ngòi, Công nghiệp, Cây trồng lâu năm và Mặt nước (Biển/Hồ). Bộ dữ liệu này đã được sử dụng trong nhiều nghiên cứu, thử nghiệm các mô hình phân loại lớp phủ khác nhau (Pham & Bui, 2021; Tsagkatakis et al., 2019).



Hình 1: Bộ dữ liệu Eurosat RGB với 10 loại hình lớp phủ được xác định. (Helber et al., 2018)

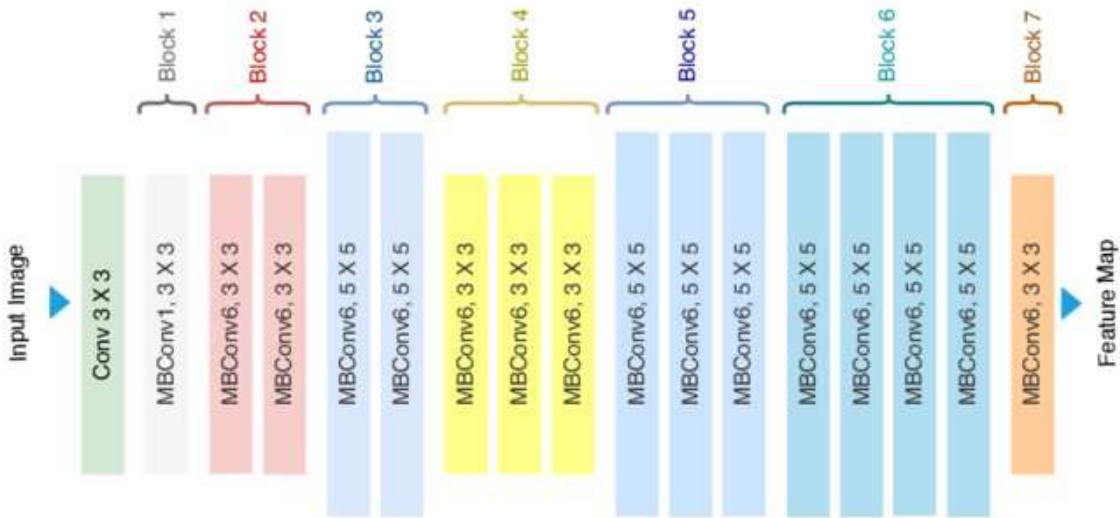
2.2. Các mô hình EfficientNet

EfficientNet là một dòng kiến trúc mạng nơ-ron nổi tiếng trong lĩnh vực thị giác máy tính, được giới thiệu qua bài báo (Tan & Le, 2019). Điểm nổi bật của EfficientNet là khái niệm "compound scaling," trong đó đồng thời tăng kích thước chiều sâu, chiều rộng, và độ phân giải của mô hình để đảm bảo hiệu suất tối ưu. By carefully balancing these dimensions, EfficientNet achieves better

performance while being more computationally efficient than other CNN architectures (Hoang & Jo, 2021; Punuri et al., 2023). Với các biến thể từ B0 đến B7, EfficientNet cung cấp sự linh hoạt cho việc chọn lựa mô hình phù hợp với yêu cầu tài nguyên tính toán và không gian bộ nhớ của ứng dụng cụ thể. Các phiên bản này đã chứng minh khả năng xuất sắc trong nhiều nhiệm vụ, đặc biệt là nhận dạng hình ảnh. Với hiệu suất

cao và khả năng thích ứng đối với điều kiện tài nguyên đa dạng, EfficientNet đã trở thành một lựa chọn phổ biến trong cộng đồng nghiên cứu và ngành công nghiệp. Ứng dụng của nó

không chỉ giới hạn trong thị giác máy tính mà còn mở rộng sang nhiều lĩnh vực khác nhau, đánh dấu đóng góp to lớn trong phát triển và ứng dụng của trí tuệ nhân tạo.



Hình 2: Cấu trúc EfficientNet B0, các biến thể sau này mở rộng từ phiên bản B0 (Ahmed & Sabab, 2021; Tan & Le, 2019). Các mạng EfficientNet thường được huấn luyện trước sử dụng bộ dữ liệu hình ảnh sẵn có. Các tham số của mô hình có thể được sử dụng và được tinh chỉnh (fine-tune) sử dụng EuroSat hoặc các bộ dữ liệu khác

EfficientNet B0: Là phiên bản nhỏ nhất của EfficientNet. Thích hợp cho các ứng dụng có tài nguyên hạn chế, ví dụ như trên thiết bị di động. **EfficientNet B1:** Tăng kích thước so với B0 để cải thiện độ chính xác. Vẫn giữ được tính di động và phù hợp cho nhiều ứng dụng di động. **EfficientNet B2:** Tiếp tục tăng kích thước so với B1. Thích hợp cho các ứng dụng yêu cầu độ chính xác cao hơn một chút. **EfficientNet B3:** Một phiên bản lớn hơn, thích hợp cho các ứng dụng có khả năng tính toán và bộ nhớ nâng cao. **EfficientNet B4-B7:** Các biến thể này tiếp tục tăng kích thước và độ phức tạp của mô hình, dành cho các ứng dụng đòi hỏi độ chính xác cao và có tài nguyên tính toán đủ. Với đối tượng của nghiên cứu là

3. Kết quả và thảo luận.

Với 27.000 tấm ảnh EuroSat (loại RGB) sử dụng trong mô hình, trong số đó 70% được

lựa chọn ngẫu nhiên để huấn luyện (18.900 tấm), 15% dùng để hiệu chỉnh (4050 tấm) và 15% còn lại (không tham gia trong quá trình huấn luyện) dùng để đánh giá độ chính xác phân loại mô hình. Trong nghiên cứu này, các tham số của mạng EfficientNet được huấn luyện trước (pre-trained) với bộ dữ liệu Imagenet (với hơn 14 triệu tấm ảnh của 1000 đối tượng), EuroSat được sử dụng trong quá trình huấn luyện lớp phân loại cuối cùng (dense layer). Các mô hình được huấn luyện với các tham số mặc định và một số tham số đầu vào khác như (1) Batchsize: 64, (2) Learning rate: 0.001, (3) Epoch: 200, (4) thuật toán tối ưu: Adam. (5) Hàm mục tiêu: "categorical_crossentropy". Sau khi huấn luyện các mô hình được đánh giá độ chính xác, kết quả mô hình B5 được thể hiện trong (bảng 1). Độ chính xác mô hình B5 đạt được cao nhất 97,9%.

Bảng 1: Độ chính xác phân loại theo mô hình EfficientNet B5. Tổng số cảnh sử dụng để đánh giá độ chính xác là 4050 (15%), chia đều cho 10 loại hình lớp phủ



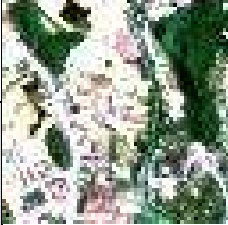
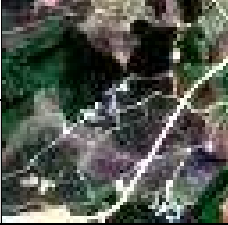


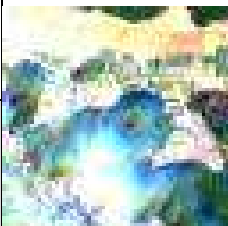
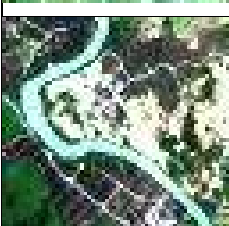
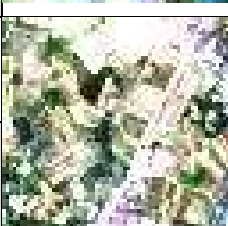
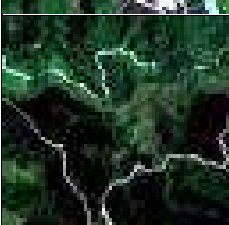
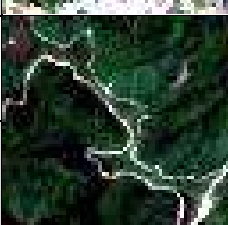

Dữ liệu kiểm tra (Reference data). Mỗi loại hình lớp phủ bao gồm 405 tấm ảnh.												
	Rừng	Cây trồng hàng năm	Đường cao tốc	Thảm thực vật thân thảo	Đồng cỏ	Khu dân cư	Sông ngòi	Công nghiệp	Cây trồng lâu năm	Mặt nước	Tổng hàng	
Rừng	401	0	0	0	0	0	0	1	3	0	405	0,990
Cây trồng hàng năm	0	385	0	0	1	0	0	0	5	0	391	0,985
Đường cao tốc	0	2	401	1	3	0	0	0	0	0	407	0,985
Thảm thực vật thân thảo	2	3	0	390	3	0	0		3	0	401	0,973
Đồng cỏ	0	6	0	0	395	0	0	0	0	0	401	0,985
Khu dân cư	0	0	3	0	0	402	0	5	0	0	410	0,980
Sông ngòi	0	0	0	0	0	0	403	0	0	4	407	0,990
Công nghiệp	0	3	1	8	2	3	0	397	3	0	417	0,952
Cây trồng lâu năm	2	6	0	6	1	0	0	2	391	0	408	0,958
Mặt nước	0	0	0	0	0	0	2	0	0	401	403	0,995
Tổng cột	405	405	405	405	405	405	405	405	405	405		
	0,99	0,95	0,99	0,96	0,98	0,99	1,00	0,98	0,97	0,99		0,979

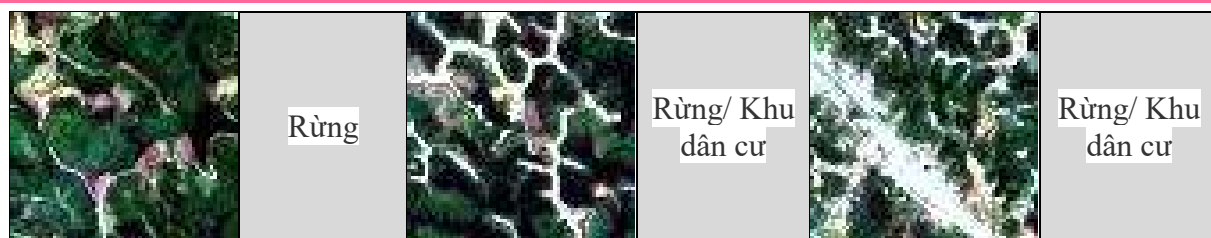
Thông thường các mô hình học sâu hiện nay đã được huấn luyện sẵn với bộ ảnh đa dạng từ các nguồn khác nhau, sau đó người sử dụng vi chỉnh bằng bộ dữ liệu mục tiêu. Nghiên cứu này sử dụng cấu trúc mặc định của mạng EfficientNet, sau đó vi chỉnh tham số sử dụng bộ dữ liệu EuroSat. Mô hình Efficient có thể được huấn luyện từ đầu (các tham số mô hình được khởi tạo ngẫu nhiên) với bộ dữ liệu viễn thám sử dụng, tuy nhiên việc huấn luyện này sẽ mất thêm thời gian. Bên cạnh đó, độ chính xác phân loại có thể được cải thiện nếu tích hợp thêm các mô hình tối ưu trong lựa chọn các tham số như Dropout rate..., hoặc tối ưu cấu trúc lớp dense layer. Bên cạnh đó, độ chính xác phân loại có thể tăng thêm nếu sử dụng bộ dữ liệu EuroSat đa phổ (13 kênh phổ).

Trong các nghiên cứu xây dựng mô hình học sâu phân loại ảnh viễn thám, bộ dữ liệu mẫu chuẩn, có kích thước lớn đóng vai trò quan trọng để huấn luyện mô hình, nhằm nâng

cao độ chính xác và khả năng nhận diện các loại hình lớp phủ đa dạng. Hiện nay có nhiều bộ dữ liệu chuẩn được xây dựng tại các khu vực khác nhau trên thế giới với các đặc điểm địa lý khác nhau, tuy nhiên dữ liệu tại Việt Nam hiện còn hạn chế, chủ yếu là dữ liệu từ các nghiên cứu đơn lẻ, chưa có bộ dữ liệu mẫu thống nhất trên toàn lãnh thổ. Trong nghiên cứu này, bộ dữ liệu mẫu được xây dựng tại Châu Âu, do đó mẫu dạng có thể có sự khác biệt so với mẫu dạng các đối tượng tại Việt Nam (hình dạng, phân bố các đối tượng trên bề mặt trái đất). Trong (hình 3), kết quả phân loại từ mô hình khá tương đồng với kết quả giải đoán bằng mắt. Tuy nhiên có một số hạn chế, đặc biệt có một số mẫu tỷ lệ phân bố của các đối tượng khó có thể lựa chọn chính xác đối tượng chi phối của cảnh ảnh là gì. Thêm vào đó các khu dân cư tại khu vực miền núi thường có diện tích nhỏ, xen kẽ với các đối tượng khác nên cũng khó để phân biệt rõ.

Hình 3: Kết quả phân loại một số mẫu tại Việt Nam sử dụng mô hình EfficientNet B5, với nguồn ảnh đầu vào là Sentinel 2. Hàng trên là kết quả phân loại từ mô hình, hàng dưới là kết quả giải đoán

	Rừng		Rừng		Sông
	Rừng		Rừng		Sông
	Sông		Khu dân cư		Đường
	Sông		Khu dân cư		Thảm thực vật thân thảo/ Khu dân cư
	Khu dân cư		Đường		Đường
	Khu dân cư		Cây trồng lâu năm		Cây trồng lâu năm / Đường
	Cây trồng hàng năm		Khu dân cư		Cây trồng hàng năm
	Cây trồng hàng năm		Khu dân cư (một phần bị mây che phủ)		Cây trồng hàng năm
	Sông		Khu dân cư		Khu dân cư
	Sông		Khu dân cư		Khu dân cư
	Rừng		Rừng		Rừng
	Rừng		Rừng		Rừng
	Rừng		Rừng		Khu dân cư



4. Kết luận

Nghiên cứu này thử nghiệm các mô hình EfficientNet trong phân loại ảnh viễn thám (scene-based classification) sử dụng bộ dữ liệu chuẩn EuroSat (RGB). Kết quả cho thấy độ chính xác sau khi huấn luyện vi chỉnh (fine-tuning) đạt trên 97% cho cả 3 mô hình. Độ chính xác có thể được cải thiện nếu các tham số của mô hình được vi chỉnh (fine tuning) bằng các thuật toán tối ưu hóa, thay vì sử dụng các tham số mặc định sẵn có của mô hình. Bên cạnh đó có thể tích hợp các bộ dữ liệu khác nhau để đa dạng hóa số lượng mẫu phân loại đại diện cho các khu vực địa lý khác nhau. Việc tích hợp này sẽ nâng cao tính ứng dụng của mô hình trong các bài toán phân loại lớp phủ với quy mô rộng lớn, với đặc điểm lớp phủ đa dạng. ○

Dữ liệu sử dụng: Bộ dữ liệu EuroSat được tải từ <https://github.com/phelber/eurosat>

Lời cảm ơn: Nghiên cứu này được tài trợ bởi Bộ Khoa học và Công nghệ, Chương trình phát triển khoa học cơ bản trong lĩnh vực hóa học, khoa học sự sống, khoa học trái đất và khoa học biển giai đoạn 2017 - 2025. Mã số đề tài: ĐTĐL.CN-42/23.

Tài liệu tham khảo

[1]. Ahmed, T., & Sabab, N. H. N. (2021). Classification and understanding of cloud structures via satellite images with efficientnet. *ESS Open Archive*. doi:10.1002/essoar.10507423.1

[2]. Helber, P., Bischke, B., Dengel, A., & Borth, D. (2018, 22-27 July 2018). *Introducing Eurosat: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification*. Paper presented at the IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium.

[3]. Hoang, V. T., & Jo, K. H. (2021, 8-10 July 2021). *Practical Analysis on Architecture of EfficientNet*. Paper presented at the 2021 14th International Conference on Human System Interaction (HSI).

[4]. Pham, V.-D., & Bui, Q.-T. (2021). Spatial resolution enhancement method for Landsat imagery using a Generative Adversarial Network. *Remote Sensing Letters*, *12*(7), 654-665. doi:10.1080/2150704X.2021.1918789

[5]. Punuri, S. B., Kuanar, S. K., Kolhar, M., Mishra, T. K., Alameen, A., Mohapatra, H., & Mishra, S. R. (2023). Efficient Net-XGBoost: An Implementation for Facial Emotion Recognition Using Transfer Learning. *Mathematics*, *11*(3). doi:10.3390/math11030776

[6]. Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv e-prints*, arXiv:1905.11946. doi:10.48550/arXiv.1905.11946

[7]. Tsagkatakis, G., Aidini, A., Fotiadou, K., Giannopoulos, M., Pentari, A., & Tsakalides, P. (2019). Survey of Deep-

Learning Approaches for Remote Sensing Observation Enhancement. *Sensors*, 19(18). doi:10.3390/s19183929

Convolutional Neural Network for Semantic Segmentation in High-Resolution Images. *Remote Sensing*, 9(5). doi:10.3390/rs9050446.○

[8]. Wang, H., Wang, Y., Zhang, Q., Xiang, S., & Pan, C. (2017). Gated

Summary

Application of EfficientNet in surface coating classification

Pham Le Tuan, Dong Vu Ha, Bui Quang Thanh

University of Natural Sciences, Hanoi National University

Le Chi Thanh

Department of Natural Resources and Environment of Ca Mau province

Do Thi Phuong Thao

Mine geological college

The emergence of machine learning methods, including classification algorithms, segmentation techniques, and optimization algorithms, has significantly contributed to automating the image analysis process and improved accuracies in monitoring changes on the Earth's surface. In this study, the Euroset dataset was used to train various variants of EfficientNet models, applied to scene-based classification of land cover. The results show that the classification accuracy of EfficientNet B3, B4, and B5 reached respective values of 97.7%, 97.74%, 97.9%, and the models were used to classify several scenes in the Northeastern part of Vietnam. Some miss-match occurred and it may be attributed to the training dataset lacking certain land cover types that are prevalent in Vietnam.○

Keywords: Landcover, EfficientNet, Sattelite data