

# Nâng cao hiệu quả mô hình Convolutional Neural Network nhận diện loài cây qua hình ảnh lá cây sử dụng các lớp Denoising Convolutional Autoencoder và áp dụng Data Augmentation trong quá trình học

TS. Ngô Quốc Việt

Khoa Công nghệ thông tin, Trường Đại học Sư phạm Thành phố Hồ Chí Minh

Vương Lê Minh Nguyên

Sinh viên hệ đại học chính quy khoa CNTT  
trường Đại học Sư phạm Thành phố Hồ Chí  
Minh

Lương Công Tâm

Sinh viên hệ đại học chính quy khoa CNTT  
trường Đại học Sư phạm Thành phố Hồ Chí  
Minh

**Sơ lược** – "Nhận diện loài cây thông qua hình ảnh lá cây một cách tự động, cũng như nhiều chủ đề nhận diện hình ảnh trong lĩnh vực Sinh học, là một chủ đề thu hút được sự quan tâm sâu sắc từ những chuyên gia về Thị giác máy tính (Computer Vision - CV) và Hệ thống thông minh (Intelligence System). Trong các nghiên cứu trước đây, nhiều nhóm nghiên cứu đã ứng dụng Học máy (Machine Learning – ML) cho phương pháp nhận diện loài cây dựa vào những đặc tính của những chiếc lá của chúng như hình dạng, màu sắc, kết cấu, ..., những đặc tính không trùng nhau bởi bất kỳ một loại lá cây nào. Độ chính xác của các phương pháp trên đã được gia tăng đáng kể từ khi Học sâu (Deep Learning – DL) phát triển. Với khả năng mô phỏng cấu trúc hệ thần kinh của con người, các mô hình mạng lưới Convolutional Neural Network (CNN) này có khả năng tự học lấy các đặc trưng của lá cây, từ đó có thể phân biệt được loài cây thông qua hình ảnh của một chiếc lá. Tuy nhiên, độ chính xác của CNN có thể bị ảnh hưởng bởi các yếu tố như các điểm ảnh nhiễu, điều kiện ánh sáng xấu, bị khuất hoặc mất một vài đặc trưng quan trọng của chiếc lá. Để khắc phục vấn đề trên, chúng tôi sẽ xây dựng, thử nghiệm và so sánh một mô hình kết hợp giữa Denoising Convolutional Autoencoder và Convolution Neural Network (DCA-CNN) với một mô hình mạng CNN đơn thuần."

**Keywords** – *plant leaf classification; convolutional neural network; convolutional autoencoder;*

## I. GIỚI THIỆU

### 1) Giới thiệu

Thực vật nói chung và cây cối nói riêng đều nắm giữ một vai trò rất quan trọng trong môi trường tự nhiên, chúng hiện đang bao phủ khoảng 30% bề mặt của Trái đất tính đến năm 2010<sup>1</sup>, là lá phổi tự nhiên với vai trò điều hòa lượng khí O<sub>2</sub> và CO<sub>2</sub>. Trên thế giới hiện nay đã ghi nhận được có hơn 60,065 loài cây [1], nhưng có thể vẫn còn nhiều loài chưa được phát hiện. Với số lượng chủng loại lớn như thế thì việc nhận

dạng chúng là quá phức tạp, thậm chí có một số loài có những đặc điểm gần giống nhau, rất khó để có thể phân biệt được.

Để phân biệt, nhận biết một loài cây ta có thể dựa vào những đặc tính của chúng như thân cây, rễ, nụ, nhị, quả, khu vực sinh trưởng, đặc tính hóa học, hình dáng, màu sắc của hoa, lá cây, ... Trong đó, lá cây là đặc tính được sử dụng nhiều nhất trong bài toán nhận diện loài cây này bởi vì so với những đặc tính khác, lá cây là bộ phận bao phủ trên cây nhiều nhất, dễ thu thập

---

<sup>1</sup> Theo Global Forest Watch.

mẫu nhất, còn ví dụ đối với đặc tính hóa học của cây thì ta phải thu thập trong phòng thí nghiệm, đối với hoa, quả của cây thì chỉ có thể thu thập mẫu vào một khoảng thời gian cụ thể tùy theo từng loại cây, các đặc tính khác cũng có một số khó khăn cụ thể khi thu thập mẫu.

Chủ đề nhận biết loài cây này không phải là một chủ đề quá mới, nó đã được thực hiện bởi rất nhiều phương pháp khác nhau, từ các phương pháp ứng dụng ML, xử lý ảnh (Image Processing) cho tới các phương pháp ứng dụng DL, NN, CNN... Các phương pháp trên đã từng bước cải thiện được độ chính xác khi nhận diện bằng thông tin đặc trưng trên lá cây. Tuy nhiên vẫn còn gặp một số vấn đề khi gặp các trường hợp ảnh kém chất lượng như lệch sáng, nhiễu, lá bị hư hỏng một phần nhỏ.

Nhằm khắc phục những thiếu sót trên, chúng tôi đưa ra một phương pháp kết hợp mô hình Denoising Convolutional Autoencoder với Convolutional Neural Network (DCA-CNN) đồng thời kết hợp sử dụng Data Augmentation trong quá trình học. Để kiểm tra tính khả thi khi sử dụng kết hợp lớp DCA với CNN, chúng tôi sẽ sử dụng cùng một cấu trúc CNN cho 2 mô hình, một mô hình sẽ được liên kết với lớp DCA, sau đó tiến học cho cả 2 cùng học và thực hiện dự đoán trên cùng một train, test sau đó thống kê và đưa ra kết luận.

## 2) Dataset

Flavia [2] là bộ dataset chúng tôi sử dụng để huấn luyện mô hình CNN trong đề tài này. Bộ dataset này gồm 1907 hình ảnh kích thước 1600x1200 của 32 loài cây, tên và danh pháp khoa học của chúng.



## II. CÁC PHƯƠNG PHÁP TRƯỚC ĐÂY

Trong các nghiên cứu trước đây, nhiều nhóm nghiên cứu đã đề xuất rất nhiều phương pháp nhận dạng lá cây sử dụng các thuật toán xử lý ảnh cổ điển như HoG-based, shape-based (contour, texture), SIFT-based để trích xuất đặc trưng từ hình ảnh lá cây và so khớp với dữ liệu có sẵn để nhận dạng và phân loại. Nhược điểm của những phương pháp cổ điển này là quá trình xử lý ảnh nặng nề, đồng thời khả năng khái quát hóa hình ảnh kém do tính không linh động của các thuật toán cổ điển. [4][7][8]

Kể từ khi máy học ra đời, cùng với phát triển mạnh mẽ của các thuật toán máy học trong khoa học dữ liệu, nhiều nhóm nghiên cứu đã tiến hành áp dụng máy học vào nhận dạng và phân loại lá cây qua hình ảnh. Các thuật toán tiêu biểu có thể kể đến đã mang lại hiệu quả cao trên hình ảnh lá là AdaBoost, SVM, đã được áp dụng vào nhận dạng hình ảnh lá cây và mang lại hiệu quả vượt xa các phương pháp cổ điển. Tuy nhiên, các thuật toán máy học vẫn có nhiều nhược điểm lớn. Một trong số đó là quá trình rút trích đặc trưng vẫn phải lệ thuộc vào các phương pháp xử lý ảnh cổ điển, do đó rất nặng nề về mặt tính toán. Ngoài ra, các phương pháp máy học không hỗ trợ tìm kiếm đặc trưng tối ưu tự động mà còn lệ thuộc vào nhiều phương pháp biến đổi dữ liệu đầu vào như giảm chiều dữ liệu (PCA), onehot encode các đặc trưng thuộc dạng nhóm (categorical features), ... [5]

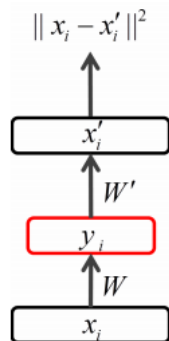
Tiếp nối mô hình mạng neural nhân tạo đơn giản của máy học là sự ra đời của các phương pháp học sâu với mô hình nhiều lớp đã mang đến một cuộc cách mạng lớn đối với thị giác máy tính nói riêng và ngành khoa học dữ liệu nói chung. Mạng neural tích chập (Convolutional Neural Network – CNN), ứng dụng phép tích chập trên một nhân chập kích thước 3x3, 5x5, 7x7,... với bộ weight của nhân chập có thể được cập nhật lại trong quá trình học, được ứng dụng rộng rãi vào

các mô hình nhận dạng phân loại trên ảnh và đạt được độ chính xác đáng kinh ngạc. Một ưu điểm vượt bậc của CNN là có thể tự động tìm kiếm và rút trích đặc trưng mà không cần lệ thuộc vào quá trình tiền xử lý dữ liệu như đối với các mô hình máy học cổ điển. [6]

Kể từ khi được ứng dụng vào thực tế bởi mô hình CNN đơn giản LeNet-5 (1998), hơn một thập kỷ qua đã chứng kiến những bước tiến dài của ngành học sâu qua các dấu ấn nổi bật AlexNet (2012), VGG-16 (2014), các mô hình ứng dụng Inception Module điển hình là GoogLeNet (2015 đến nay). Sự phát triển của các mô hình học sâu đã mở ra rất nhiều hướng phát triển cho chủ đề nhận dạng hình ảnh, đặc biệt là hình ảnh lá cây. Từ năm 2015 đến nay đã có nhiều bài báo công bố về các phương pháp nhận dạng hình ảnh lá cây sử dụng CNN cùng các biến thể của nó. [9][10]

### III. MÔ HÌNH DCA-CNN

Autoencoder là một dạng mô hình ứng dụng của Neural Network được sử dụng để biến đổi dữ liệu thành dạng đơn giản hơn mà ở đó các đặc trưng có thể được rút trích dễ dàng. Autoencoder có cấu trúc như hình bên, với  $x_i$  là điểm đầu vào,  $y_i$  là dữ liệu mã hóa từ  $x_i$  qua ánh xạ  $W$ , để từ đó dựng lại  $x_i'$  gần giống với  $x_i$  ban đầu.  $\|x_i - x_i'\|^2$ , mean square error, là hàm loss cần được tối ưu hóa của dạng mô hình này. [3]



Trong thí nghiệm này, chúng tôi không sử dụng Autoencoder cổ điển mà thay vào đó là Denoising Convolutional Autoencoder (DCA) với các lớp nhân chập thích hợp cho việc mã hóa hình ảnh. DCA sẽ tiến hành mã hóa hình ảnh đầu vào và khôi phục lại hình ảnh đầu ra với cùng kích thước đầu vào, cùng với hiệu quả giảm sự phức tạp của hình ảnh và đồng thời khử nhiễu.

Model: "model_6"		
Layer (type)	Output Shape	Param #
=====		
input_7 (InputLayer)	(None, 384, 512, 3)	0
conv2d_15 (Conv2D)	(None, 384, 512, 16)	1216
max_pooling2d_9 (MaxPooling2)	(None, 192, 256, 16)	0
conv2d_16 (Conv2D)	(None, 190, 254, 32)	4640
conv2d_transpose_10 (Conv2DT	(None, 192, 256, 16)	4624
up_sampling2d_6 (UpSampling2	(None, 384, 512, 16)	0
conv2d_transpose_11 (Conv2DT	(None, 384, 512, 3)	1203
=====		
Total params: 11,683		
Trainable params: 11,683		
Non-trainable params: 0		

Cuối cùng chúng tôi sử dụng một mô hình CNN đã được huấn luyện thông qua phương pháp Data Augmentation (gia tăng dữ liệu bằng các phép biến đổi không làm thay đổi nhãn dữ liệu như xoay, lật, phóng to, thu nhỏ, thêm nhiễu, ...) để phân loại hình ảnh lá cây đầu vào thành một trong 32 loài cây có trong dataset.

Model: "model_1"		
Layer (type)	Output Shape	Param #
=====		
input_1 (InputLayer)	(None, 384, 512, 3)	0
conv2d_1 (Conv2D)	(None, 192, 256, 64)	1792
conv2d_2 (Conv2D)	(None, 96, 128, 64)	36928
max_pooling2d_1 (MaxPooling2	(None, 48, 64, 64)	0
conv2d_3 (Conv2D)	(None, 24, 32, 128)	73856
conv2d_4 (Conv2D)	(None, 12, 16, 128)	147584
max_pooling2d_2 (MaxPooling2	(None, 6, 8, 128)	0
dropout_1 (Dropout)	(None, 6, 8, 128)	0
flatten_1 (Flatten)	(None, 6144)	0
dense_1 (Dense)	(None, 2048)	12584960
dense_2 (Dense)	(None, 256)	524544
dense_3 (Dense)	(None, 32)	8224
=====		
Total params: 13,377,888		
Trainable params: 13,377,888		
Non-trainable params: 0		

Mô hình CNN này sau đó được kết hợp với mô hình Denoising Convolutional Autoencoder (DCA) để giảm overfitting, tăng khả năng khái quát hóa dữ liệu và tăng khả năng dự đoán của mô hình trên ảnh lá cây bị nhiễu.

### IV. HUẤN LUYỆN

Khi huấn luyện mô hình DCA, chúng tôi sử dụng hàm loss là binary cross entropy, chủ động sinh thêm nhiễu dạng phân phối chuẩn (normal distributed random) và nhiễu dạng phân phối đều

(uniform distributed random) để giúp mô hình học được các rút trích, khôi phục đặc trưng từ ảnh nhiễu cũng như ảnh thông thường.

Khi huấn luyện mô hình CNN để phân lớp hình ảnh lá cây, chúng tôi sử dụng các phương pháp Data Augmentation để làm đa dạng tập dữ liệu huấn luyện và xác thực cho mô hình, giúp giảm overfitting và tăng tính khái quát cho mô hình.

## V. KIỂM THỬ MÔ HÌNH VÀ KẾT QUẢ

### 1) Kiểm thử mô hình

Chúng tôi chia bộ dữ liệu thành 2 phần, 85% hình ảnh được sử dụng cho huấn luyện (train – valid set) và 15% hình ảnh dùng cho kiểm thử (test set). Tập train – valid lại tiếp tục được chia thành 2 phần, 85% để huấn luyện mô hình, và 15% còn lại để xác thực tính hiệu quả của quá trình học.

Chúng tôi tiến hành 3 bước kiểm thử mô hình. Ở bước đầu tiên, chúng tôi kiểm tra độ chính xác của mô hình trên tập valid. Bước tiếp theo chúng tôi kiểm tra độ chính xác của mô hình trên tập test. Cuối cùng, chúng tôi kiểm tra tính hiệu quả của mô hình bằng cách áp dụng Data Augmentation và sinh nhiễu chủ động vào tập test để kiểm thử mô hình. Bước cuối cùng này chính là bước quyết định tính hiệu quả và khả thi của mô hình khi áp dụng thực tế.

Ta có kích thước các tập dữ liệu như sau:

Tập train: 1377 hình ảnh

Tập valid: 243 hình ảnh

Tập test: 287 hình ảnh

Tập sinh ra từ test generator: 957 hình ảnh

### 2) Kết quả

Kết quả kiểm thử trên mô hình CNN:

- F1-score của tập valid: 0.93
- F1-score của tập test: 0.96

- F1-score của test generator: 0.48

Kết quả kiểm thử trên mô hình DCA-CNN:

- F1-score của tập valid: 0.91
- F1-score của tập test: 0.94
- F1-score của test generator: 0.74

## VI. KẾT LUẬN

Qua kết quả thu được từ thí nghiệm, ta nhận thấy các mô hình CNN rất dễ dẫn đến overfitting trên tập dữ liệu mà nếu không có phương pháp kiểm thử hợp lý thì khi ứng dụng thực tế, mô hình dự đoán của chúng ta sẽ không đủ khả năng khái quát hóa để đạt độ chính xác cao như khi còn trong thí nghiệm.

Denoising Autoencoder là một mô hình hữu hiệu để giảm overfitting khi kết hợp với các mô hình CNN dùng trong thị giác máy tính nói chung và phân loại thực vật thông qua hình ảnh lá cây nói riêng.

Tất cả các tài nguyên, công việc của chúng tôi có thể được tải về tại đây: <https://github.com/nguyenvlm/PlantLeafClassification>

## REFERENCES

- [1] E. Beech, M. Rivers, S. Oldfield and P. P. Smith, "GlobalTreeSearch: The first complete global database of tree species and country distributions," *Journal of Sustainable Forestry*, vol. 36, pp. 454-489, 2017.
- [2] G. W. Stephan, S. B. Forrest, Y. X. Eric, X. W. Yu, F. C. Yi and L. Chiao, "A Leaf Recognition Algorithm for Plant classification Using Probabilistic Neural Network," *IEEE 7th International Symposium on Signal Processing and*

- [3] Wei Wang, Yan Huang, Yizhou Wang, Liang Wang, "Generalized Autoencoder: A Neural Network Framework," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2014.
- [4] A. . Kadir, L. E. Nugroho, A. . Susanto and P. I. Santosa, "Leaf Classification Using Shape, Color, and Texture Features," *arXiv: Computer Vision and Pattern Recognition*, vol. , no. , p. , 2014.
- [5] X. . Shang and L. A. Chisholm, "Classification of Australian Native Forest Species Using Hyperspectral Remote Sensing and Machine-Learning Classification Algorithms," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2481-2489, 2014.
- [6] J. E. S. Sklan, A. J. Plassard, D. . Fabbri and B. A. Landman, "Toward content-based image retrieval with deep convolutional neural networks," *Proceedings of SPIE*, vol. 9417, no. , p. , 2015.
- [7] D. . Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, p. , 2004.
- [8] X.-Y. . Xiao, R.-X. . Hu, S. . Zhang and X.-F. . Wang, "HOG-based approach for leaf classification," , 2010. [Online]. Available: [https://link.springer.com/chapter/10.1007/978-3-642-14932-0\\_19](https://link.springer.com/chapter/10.1007/978-3-642-14932-0_19). [Accessed 25 12 2019].
- [9] C. . Szegedy, W. . Liu, Y. . Jia, P. . Sermanet, S. . Reed, D. . Anguelov, D. . Erhan, V. . Vanhoucke and A. . Rabinovich, "Going Deeper with Convolutions," *arXiv: Computer Vision and Pattern Recognition*, vol. , no. , p. , 2014.
- [10] K. . Simonyan and A. . Zisserman, "VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION," *arXiv: Computer Vision and Pattern Recognition*, vol. , no. , p. , 2014.