

PHÂN LOẠI CHẤT LƯỢNG KHÔNG KHÍ TỪ HÌNH ẢNH VỚI HỌC SÂU ÍT MẪU

Phạm Đức Cường

Học viên Công nghệ Bưu chính Viễn thông, Việt Nam

Tóm tắt: Phân loại chất lượng không khí từ hình ảnh đóng vai trò quan trọng trong việc giám sát và cảnh báo về ô nhiễm môi trường, đặc biệt khi cảm biến thường tồn kém và khó lắp đặt, trong khi camera ngày càng trở nên phổ biến và dễ dàng tiếp cận. Nghiên cứu này giới thiệu phương pháp học sâu ít mẫu, với ý tưởng áp dụng Few-Shot Learning - FSL (học ít mẫu) cho các mô hình học sâu phân loại chất lượng không khí. Bằng cách áp dụng FSL, chúng tôi xây dựng những mô hình học sâu có khả năng học hiệu quả từ số lượng nhỏ dữ liệu huấn luyện, đồng thời đạt độ chính xác cao trong phân loại các mức độ chất lượng không khí. Phương pháp đã được thử nghiệm trên bộ dữ liệu Air Pollution Image Dataset from India and Nepal, cho thấy độ chính xác vượt trội so với phương pháp huấn luyện truyền thống. Phương pháp chứng tỏ khả năng ứng dụng FSL trong phân loại chất lượng không khí, đặc biệt trong bối cảnh dữ liệu hình ảnh về chất lượng không khí còn hạn chế.

Từ khóa: Few-shot learning, phân loại chất lượng không khí, trí tuệ nhân tạo.

I. GIỚI THIỆU

Chất lượng không khí là một yếu tố quan trọng ảnh hưởng đến sức khỏe cộng đồng, môi trường và chất lượng cuộc sống. Chất lượng không khí kém có thể gây ra các vấn đề sức khỏe nghiêm trọng như bệnh hô hấp, các vấn đề tim mạch và gia tăng tỷ lệ tử vong. Theo Tổ chức Y tế Thế giới (WHO), ô nhiễm không khí ngoài trời gây ra khoảng 4.2 triệu ca tử vong mỗi năm, chủ yếu liên quan đến bụi mịn PM2.5, trong khi một số báo cáo nêu rõ con số này có thể lên đến 8.8

triệu khi tính cả ô nhiễm từ các nguồn nhân tạo và tự nhiên [1]. Các phương pháp giám sát chất lượng không khí truyền thống thường dựa vào các cảm biến vật lý, vốn đắt đỏ và bị giới hạn về phạm vi bao phủ. Ngoài ra, những cảm biến này cần được bảo dưỡng thường xuyên và đặt tại các vị trí cố định, dẫn đến việc thiếu tính linh hoạt trong việc theo dõi ô nhiễm không khí ở quy mô lớn. Những tiến bộ gần đây trong lĩnh vực thị giác máy tính và học sâu đã mở ra những hướng đi mới cho việc phân loại chất lượng không khí thông qua hình ảnh, cung cấp một giải pháp thay thế có hiệu quả về chi phí và dễ mở rộng.

Các mô hình học sâu, đặc biệt là mạng nơ-ron tích chập (CNN), đã chứng minh có thể áp dụng trong phân loại chất lượng không khí từ hình ảnh. Tuy nhiên, việc huấn luyện các mô hình này đòi hỏi một lượng lớn dữ liệu. Vấn đề này là rào cản lớn trong bài toán phân loại chất lượng không khí, khi mà công việc thu thập đầy đủ dữ liệu và gán nhãn cho các mức chất lượng không khí rất khó khăn. Để khắc phục hạn chế này, nghiên cứu của chúng tôi đề xuất áp dụng FSL [2] vào mô hình học sâu nhằm cải thiện độ chính xác của mô hình học sâu khi huấn luyện với số lượng nhỏ dữ liệu.

Mục tiêu của nghiên cứu là cải thiện độ chính xác trong phân loại chất lượng không khí từ hình ảnh với FSL. Nghiên cứu đã được thử nghiệm trên tập dữ liệu Air Pollution Image Dataset From India and Nepal [3] cho thấy FSL chỉ cần sử dụng một lượng dữ liệu rất nhỏ để huấn luyện các mô hình học sâu mà vẫn đem lại độ chính xác tương đương hoặc hơn khi so sánh với phương pháp truyền thống huấn luyện trên lượng lớn dữ liệu.

Tác giả: Phạm Đức Cường

Email: cuongpd@ptit.edu.vn

Đón toà soạn: 10/2024; Chính sửa: 11/2024; Chấp nhận đăng: 12/2024.

Bảng 1: Phương pháp phân loại chất lượng không khí gần đây

Bài báo	Tập dữ liệu	Mô hình	Phương pháp	Độ chính xác
Utomo et al., 2023	Air Pollution Image Dataset from India and Nepal (loại bỏ phần dữ liệu của Nepal)	Eff-AQI	Kết hợp dữ liệu từ các cảm biến không khí cùng với các đặc điểm hình ảnh liên quan	89.38%
Mahajan et al., 2024	Air Pollution Image Dataset from India and Nepal	VGG16 + mạng nơ-ron tùy chỉnh	Kết hợp trích xuất đặc trưng bằng VGG16 và dự đoán với mạng nơ-ron tùy chỉnh	87.44%

II. NGHIÊN CỨU LIÊN QUAN

Nghiên cứu của Utomo et al., 2023 [4] trình bày một mô hình CNN hiệu quả để phân loại chất lượng không khí tại India. Nghiên cứu này nhấn mạnh tầm quan trọng của việc phát triển các phương pháp tiên tiến nhằm cung cấp thông tin chính xác và kịp thời về chất lượng không khí, một vấn đề ngày càng trở nên nghiêm trọng do tình trạng ô nhiễm môi trường đang ngày càng nghiêm trọng. Mô hình Eff-AQI được thiết kế để giảm thiểu số lượng đặc trưng cần thiết mà vẫn đạt được độ chính xác cao trong phân loại chất lượng không khí. Nghiên cứu sử dụng dữ liệu từ các cảm biến không khí cùng với các đặc điểm hình ảnh liên quan để cải thiện độ chính xác của các dự đoán. Kết quả thử nghiệm 9.56 RMSE, 0.99 R2, 89.92% độ chính xác cân bằng, 89.38% độ chính xác cho thấy mô hình này có khả năng xử lý và phân tích dữ liệu hình ảnh một cách hiệu quả, từ đó cung cấp những đánh giá đáng tin cậy về ứng dụng học sâu trong phân loại chất lượng không khí.

Mahajan et al., 2024 [5] nghiên cứu sử dụng mô hình VGG16 cho việc trích xuất đặc trưng từ hình ảnh, sau đó kết hợp với mạng nơ-ron tùy chỉnh để phân loại chất lượng không khí. Kết quả thử nghiệm mô hình đạt độ chính xác 88.54% trên tập dữ liệu huấn luyện và độ chính xác 87.44% trên tập dữ liệu đánh giá, chỉ ra hiệu quả vượt trội so với các phương pháp truyền thống. Nghiên cứu này không chỉ khẳng định khả năng của công nghệ học sâu mà còn mở ra hướng đi cho các nghiên cứu tương lai nhằm cải thiện độ chính xác của bài toán phân loại chất lượng không khí từ hình ảnh.

Những nghiên cứu trong Bảng 1 chứng minh xu hướng phát triển của các phương pháp ứng dụng mô hình học sâu trong phân loại chất lượng không khí từ hình ảnh.

III. PHƯƠNG PHÁP ĐỀ XUẤT

A. Mạng nơ-ron tích chập

Mạng nơ-ron tích chập (Convolutional Neural Network - CNN) [6] [7] là một trong những kiến trúc học sâu phổ biến nhất được sử dụng trong các bài toán xử lý hình ảnh, đặc biệt là phân loại, nhận dạng và phát hiện đối tượng. CNN có khả năng học và trích xuất các đặc trưng không gian từ dữ liệu đầu vào như hình ảnh thông qua các lớp tích chập (convolutional layers), giúp tối ưu hóa việc nhận diện các đặc trưng như cạnh, góc, hoặc các mẫu phức tạp hơn khi đi sâu vào mạng.

Kể từ khi CNN được giới thiệu, nhiều mô hình CNN tiên tiến đã ra đời, đóng góp quan trọng vào sự phát triển của lĩnh vực học sâu và thị giác máy tính. Các mô hình này không chỉ giúp nâng cao độ chính xác trong các bài toán xử lý hình ảnh, mà còn mở ra những hướng đi mới trong xây dựng và huấn luyện mạng học sâu. Một số mô hình có thể kể đến như Alexnet [8], GoogLeNet [9] VGGNet [10], ResNet [11] và DenseNet [12]. VGGNet, ResNet cùng DenseNet đã trở thành nền tảng cho nhiều nghiên cứu và được ứng dụng trong các lĩnh vực khác nhau.

VGGNet được phát triển theo xu hướng tăng độ sâu của mạng CNN, bằng cách sử dụng các bộ lọc nhỏ kích thước 3x3, kết hợp với nhiều lớp chồng lên nhau. Với 16 hoặc 19 lớp, VGGNet đã minh chứng rằng việc tăng độ sâu của mạng có thể cải thiện hiệu suất phân loại hình ảnh.

ResNet (Residual Network) sử dụng cơ chế "skip connection" (kết nối tắt), khắc phục vấn đề thông tin bị mất dần khi mạng trở nên quá sâu. Điều này cho phép mạng huấn luyện hiệu quả ngay cả khi độ sâu của mạng tăng lên, giảm thiểu hiện tượng vanishing gradient. Các biến thể như ResNet-50 và ResNet-101 đã trở thành

nền tảng cho nhiều nghiên cứu hữu ích về mạng học sâu.

DenseNet cải tiến ý tưởng từ ResNet bằng cách sử dụng các kết nối dày đặc hơn, trong đó mỗi lớp trong mạng kết nối với tất cả các lớp trước đó, thay vì chỉ với một lớp liền kề như trong ResNet. Điều này giúp các lớp sau có thể nhận thông tin từ mọi lớp trước đó, tăng cường khả năng truyền tải và tái sử dụng các đặc trưng được học. Cấu trúc này còn tối ưu trong việc giảm thiểu tình trạng mất mát thông tin khi mạng trở nên sâu hơn, trong khi vẫn duy trì số lượng tham số tương đối thấp nhờ việc chia sẻ đặc trưng giữa các lớp.

B. Few-shot learning

Few-shot learning là một phương pháp học máy mà trong đó mô hình phải học để phân loại các mẫu mới với rất ít dữ liệu huấn luyện. FSL bao gồm các biến thể đặc biệt như one-shot learning (học một mẫu) [13] – phân loại mẫu chỉ với một mẫu cho mỗi lớp, và zero-shot learning (học không mẫu) [14] – xử lý dữ liệu chưa được học bằng cách sử dụng mô tả dữ liệu mà không cần bất kỳ mẫu nào được gán nhãn. Trong bài toán phân loại truyền thống, mô hình thường yêu cầu một lượng lớn dữ liệu được gán nhãn để đạt được hiệu quả cao. Tuy nhiên, FSL tìm cách giảm số lượng dữ liệu cần thiết, giúp mô hình học từ một số lượng ít ví dụ (thường là 1, 5 hoặc 10 ví dụ trên mỗi lớp). Để làm được điều này, FSL thường sử dụng các kỹ thuật như meta-learning (siêu học) [15], siamese networks (mạng so sánh) [16], hoặc prototypical networks (mạng nguyên mẫu) [17].

Meta-learning: Trong meta-learning, mô hình không chỉ học cách phân loại mà còn học cách học hiệu quả từ lượng nhỏ dữ liệu. Một cách tiếp cận phổ biến là sử dụng thuật toán "học cách thích ứng", trong đó mô hình được huấn luyện trên nhiều tập dữ liệu con nhỏ để có thể nhanh chóng thích nghi với các tập dữ liệu mới. Công thức tổng quát của quá trình meta-learning:

$$\theta^* = \arg \min_{\theta} \sum_{T_i \sim p(T)} \mathcal{L}_{T_i}(f_{\theta}) \quad (1)$$

Trong đó: θ là tham số của mô hình cần huấn luyện, θ^* là tham số tối ưu, T_i là các tập dữ liệu nhỏ rút ra từ phân phối $p(T)$, $\mathcal{L}_{T_i}$ là hàm mất

mát cho tập dữ liệu nhỏ T_i , có nhiệm vụ đánh giá hiệu suất của mô hình trên T_i , f_{θ} là mô hình với tham số hiện tại.

Siamese networks: Một cách tiếp cận khác của FSL là sử dụng siamese networks, trong đó mô hình học cách so sánh giữa các cặp mẫu và dự đoán xem chúng có thuộc cùng một lớp hay không. Trong quá trình huấn luyện, mạng được huấn luyện để tính toán độ tương đồng giữa các cặp mẫu đầu vào. Với hai đầu vào x_1 và x_2 , mạng sẽ học một mô hình ánh xạ $f_{\theta}(x)$ để tính toán khoảng cách hay độ tương đồng giữa hai đầu vào:

$$d(x_1, x_2) = |f_{\theta}(x_1) - f_{\theta}(x_2)| \quad (2)$$

Ở đây, $d(x_1, x_2)$ là khoảng cách giữa hai mẫu trong không gian đặc trưng, và mô hình sẽ học cách tối thiểu hóa khoảng cách giữa các mẫu cùng lớp và tối đa hóa khoảng cách giữa các mẫu khác lớp.

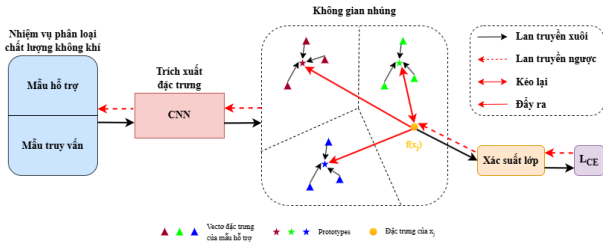
Prototypical Networks: Mô hình Prototypical Networks giả định rằng mỗi lớp có một prototype (mẫu đại diện), là trung bình của các đặc trưng trong lớp đó. Khi gặp một mẫu mới, mô hình sẽ tính khoảng cách giữa đặc trưng của mẫu và prototype của các lớp để dự đoán lớp của mẫu. Công thức của prototypical networks có thể được biểu diễn như sau:

$$c_k = \frac{1}{|S_k|} \sum_{x_i \in S_k} f_{\theta}(x_i) \quad (3)$$

Trong đó, c_k là prototype của lớp k , S_k là tập dữ liệu huấn luyện của lớp k , $f_{\theta}(x_i)$ là mô hình ánh xạ đầu vào x_i sang không gian đặc trưng học bởi mạng. Khoảng cách giữa một mẫu mới và prototype thường được tính bằng khoảng cách Euclidean.

C. Phương pháp đề xuất

Nghiên cứu đề xuất áp dụng FSL dựa trên Prototypical Networks với các mô hình CNN. Phương pháp này giúp các mô hình CNN đạt được độ chính xác cao, ngay cả khi có rất ít dữ liệu huấn luyện cho mỗi mức độ chất lượng không khí. Phương pháp đề xuất được trình bày trong hình 1.



Hình 1: Phương pháp đề xuất

1) *Trích xuất đặc trưng*: Bước đầu tiên là trích xuất đặc trưng, hình ảnh trong tập dữ liệu được đưa qua một mạng CNN để trích xuất đặc trưng.

$$f_{\theta}(x_i) = CNN(x_i) \quad (4)$$

Trong đó, x_i là hình ảnh trong tập dữ liệu, θ là tham số của mạng, CNN là mạng nơ-ron tích chập như VGGNet, ResNet, DenseNet và $f_{\theta}(x_i)$ là vectơ đặc trưng của hình ảnh x_i .

2) *Tính prototype*: Với mỗi lớp k , prototype c_k được định nghĩa là trung bình đặc trưng các hình ảnh trong tập huấn luyện S_k của lớp đó như công thức (3).

3) *Phân loại hình ảnh mới*: Khi gặp hình ảnh mới x_i , mô hình sẽ tính khoảng cách giữa đặc trưng của hình ảnh đó và các prototype của các lớp. Khoảng cách này được tính bằng khoảng cách Euclidean:

$$d_E(f_{\theta}(x_i), c_k) = \|f_{\theta}(x_i) - c_k\| \quad (5)$$

Hình ảnh mới được phân loại vào lớp có prototype gần nhất, tức là lớp k với khoảng cách $d_E(f_{\theta}(x), c_k)$ nhỏ nhất.

4) *Hàm mất mát*: Mô hình được tối ưu hóa bằng cách giảm thiểu hàm mất mát dựa trên khoảng cách giữa đặc trưng của mẫu và prototype của lớp đúng. Nghiên cứu lựa chọn sử dụng hàm mất mát cross entropy loss với khoảng cách Euclidean:

$$\mathcal{L}_{CE} = - \sum_{i=1}^N \log p(y_i = k | x_i) \quad (6)$$

Với N là số lượng mẫu trong tập huấn luyện, $p(y_i = k | x_i)$ là xác suất dự đoán hình ảnh x_i thuộc lớp y_i mà y_i bằng lớp thực tế k của x_i . Xác suất $p(y_i = k | x_i)$ được tính theo công thức của hàm softmax, với C là các lớp trong tập huấn luyện:

$$p(y_i = k | x_i) = \frac{\exp(-d_E(f_{\theta}(x_i), c_k))}{\sum_{k' \in C} \exp(-d_E(f_{\theta}(x_i), c_{k'}))} \quad (7)$$

D. Chỉ số đánh giá

1) *Chỉ số Accuracy*: **Accuracy (Độ chính xác)** là một trong những chỉ số đánh giá hiệu suất cơ bản và phổ biến nhất của các mô hình phân loại. Accuracy biểu diễn tỷ lệ dự đoán đúng của mô hình so với tổng số mẫu đánh giá. Accuracy cao cho thấy mô hình hoạt động tốt trên tập dữ liệu.

Công thức Accuracy:

$$\text{Accuracy} = \frac{\text{Số lượng dự đoán đúng}}{\text{Tổng số mẫu}} \quad (8)$$

2) *Chỉ số F1-Score*: **F1-Score** là một chỉ số đánh giá hiệu suất của mô hình phân loại nhị phân, đặc biệt hữu ích trong các bài toán có dữ liệu mất cân bằng. F1-Score được tính toán dựa trên *Precision* (độ chính xác dương tính) và *Recall* (độ nhạy), nhằm cân bằng giữa việc mô hình có dự đoán đúng các mẫu dương tính và không đưa ra quá nhiều dự đoán sai. Với mô hình phân loại đa lớp, F1-Score được tính bằng trung bình F1-Score riêng lẻ cho mỗi lớp.

Công thức F1-Score của lớp k :

$$F1\text{-Score}_k = 2 \times \frac{\text{Precision}_k \times \text{Recall}_k}{\text{Precision}_k + \text{Recall}_k} \quad (9)$$

Trong đó: Precision_k - Tỷ lệ mẫu thuộc lớp k mà mô hình dự đoán đúng trong tất cả các dự đoán là thuộc lớp k , Recall_k - Tỷ lệ các mẫu thuộc lớp k mà mô hình dự đoán đúng trong tất cả các mẫu của lớp k trong tập dữ liệu.

IV. KẾT QUẢ

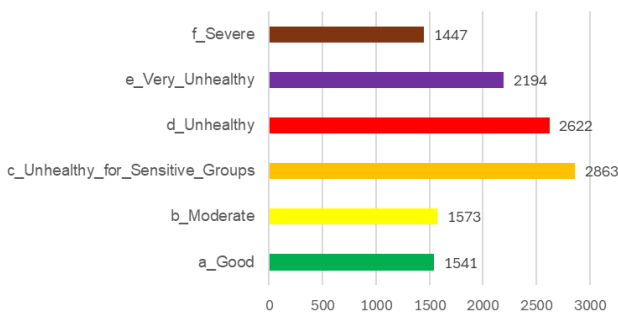
A. Dữ liệu

Air Pollution Image Dataset from India and Nepal (APIN) [3] là một bộ dữ liệu được thu thập với mục đích nghiên cứu và phân loại mức độ ô nhiễm không khí dựa trên hình ảnh. Tập dữ liệu này chứa các hình ảnh được chụp từ hai quốc gia là India và Nepal, với nhiều mức độ ô nhiễm không khí khác nhau. Tập dữ liệu này có thể được sử dụng trong các nghiên cứu thị

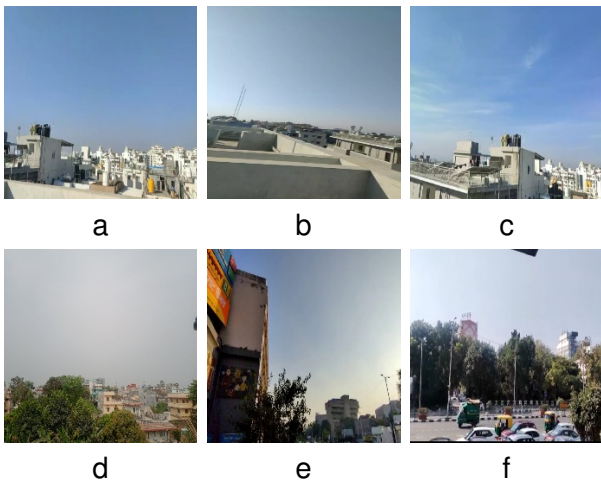
giác máy tính, đặc biệt là trong các bài toán như phân loại chất lượng không khí từ hình ảnh.

Tập dữ liệu APIN bao gồm 12240 hình ảnh định dạng JPG, độ phân giải 224x224 và được chia thành 6 lớp tương ứng các mức chất lượng không khí. Hình 2 thể hiện phân phối dữ liệu từng lớp và hình 3 là hình ảnh ví dụ cho mỗi lớp.

- **a_Good:** Tốt
- **b_Moderate:** Trung bình
- **c_Unhealthy_for_Sensitive_Groups:** Không lành mạnh cho các nhóm nhạy cảm
- **d_Unhealthy:** Không lành mạnh
- **e_Very_Unhealthy:** Rất không lành mạnh
- **f_Severe:** Nghiêm trọng



Hình 2: Phân phối dữ liệu



Hình 3: Hình ảnh ví dụ cho mỗi lớp

B. Kết quả thực nghiệm

Thực nghiệm được tiến hành trên ba mô hình VGG16, ResNet-50 và DensetNet-121 với hai phương pháp huấn luyện. Ba mô hình trên được

lựa chọn vì chúng đạt được sự cân bằng giữa độ chính xác và tốc độ thực thi. Hai phương pháp huấn luyện bao gồm:

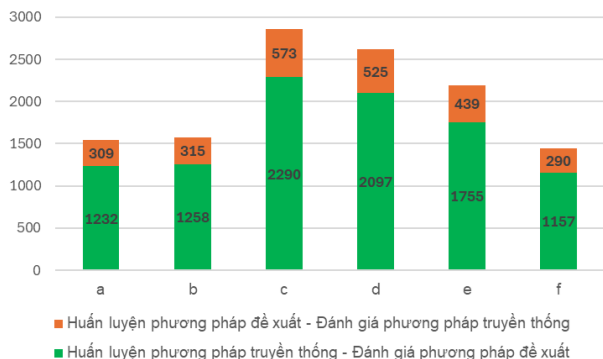
- Huấn luyện truyền thống: huấn luyện theo phương pháp truyền thống. Tập dữ liệu được chia 80% cho huấn luyện - 9789 ảnh và 20% cho đánh giá - 2451 ảnh. Dữ liệu được chia đảm bảo tỉ lệ tương tự cho từng lớp. Accuracy và F1-Score của mô hình được tính toán trên tập dữ liệu đánh giá.
- **Phương pháp đề xuất - huấn luyện với FSL dựa trên prototypical networks kết hợp cross entropy loss:** nghiên cứu sử dụng dữ liệu đánh giá của phương pháp truyền thống cho quá trình huấn luyện - 2451 ảnh và dữ liệu huấn luyện của phương pháp truyền thống cho đánh giá - 9789 ảnh. Quá trình huấn luyện thực hiện trên từng nhiệm vụ, mỗi nhiệm vụ lấy ngẫu nhiên 5 mẫu hỗ trợ và 10 mẫu truy vấn từ mỗi lớp trong tập huấn luyện. Quá trình huấn luyện dự đoán lớp của mẫu truy vấn dựa trên thông tin của mẫu hỗ trợ, sau đó tính toán hàm mất mát dựa trên lớp dự đoán và lớp thực tế của mẫu truy vấn. Giá trị loss này được sử dụng để cập nhật tham số cho mô hình.

Accuracy của mô hình được tính toán bằng trung bình giá trị này trên tất cả các nhiệm vụ đánh giá, mỗi nhiệm vụ lấy ngẫu nhiên 5 mẫu hỗ trợ và 1000 mẫu truy vấn khác mẫu hỗ trợ từ mỗi lớp trong tập đánh giá. Nghiên cứu thực hiện đánh giá với 100 nhiệm vụ - tổng cộng 1000 mẫu * 6 lớp * 100 nhiệm vụ = 600000 mẫu đánh giá và có thể trùng lặp do các nhiệm vụ lựa chọn ngẫu nhiên mẫu truy vấn. Mẫu hỗ trợ được lựa chọn khác nhau trong từng nhiệm vụ. Vì số lượng mẫu truy vấn của mỗi lớp giống nhau trong từng nhiệm vụ đánh giá, F1-Score cho mỗi lớp được nhân với một hệ số w_k trước khi tính trung bình F1-Score của nhiệm vụ. Kết quả F1-Score cuối cùng của mô hình là trung bình các giá trị F1-Score trên tất cả nhiệm vụ đánh giá. Hệ số w_k được định nghĩa là tỉ lệ mẫu của lớp k trong toàn bộ dữ liệu đánh giá.

Bảng III cho thấy kết quả của phương pháp đề xuất đạt Accuracy khoảng 0.93 với cả ba mô hình CNN, cao hơn so với nghiên cứu của

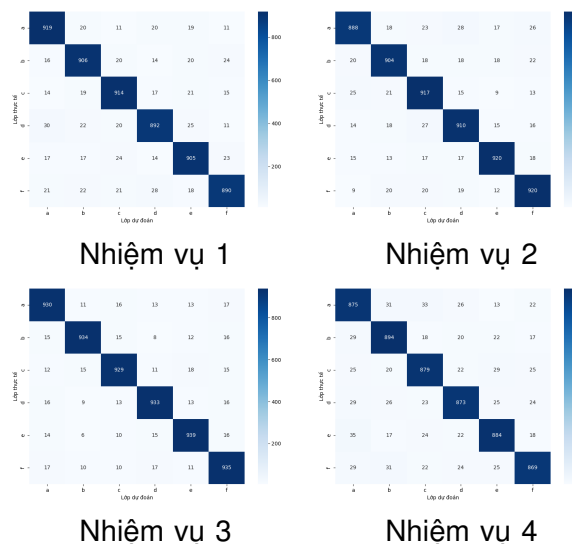
Bảng II: Giá trị hệ số w_k

k	a	b	c	d	e	f
w_k	0.1259	0.1285	0.2339	0.2142	0.1793	0.1182



Hình 4: Phân chia dữ liệu huấn luyện - đánh giá

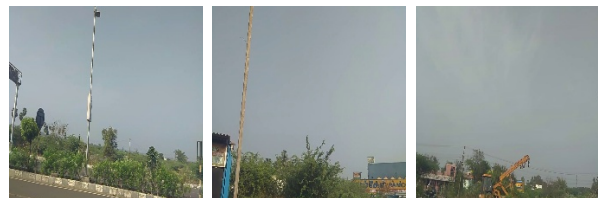
Mahajan et al. [5] - 0.8938 và mô hình Eff-AQI [18] - 0.8744 dù sử dụng số lượng dữ liệu huấn luyện ít hơn nhiều. Kết quả cũng cho thấy các thông số vượt trội về Accuracy cũng như F1-Score khi áp dụng phương pháp đề xuất vào mô hình CNN.



Hình 5: Ma trận nhầm lẫn DenseNet-121-FSL

Hình 5 thể hiện ma trận nhầm lẫn của mô hình DenseNet-121-FSL trong 4 nhiệm vụ đầu tiên. Ma trận nhầm lẫn cho thấy mô hình hoạt động ổn định và cân bằng trên tất cả các lớp, với đường chéo chính đậm và các phần còn lại nhạt hơn, xấp xỉ nhau. Nguyên nhân có thể do bộ dữ liệu có những hình ảnh không khí gần

giống nhau thuộc những lớp chất lượng không khí khác nhau. Điều này gây ra dự đoán sai khi mô hình lựa chọn mẫu hỗ trợ hay mẫu đánh giá vào những hình ảnh này.



Lớp a

Lớp b

Lớp c

Hình 6: Hình ảnh không khí gần giống nhau

Thực nghiệm được tiến hành trên máy tính CPU Intel Xeon E5-2680, GPU NVIDIA GeForce RTX 3060, RAM 32GB cho thấy phương pháp hoàn toàn có thể áp dụng trong thời gian thực với tốc độ thực thi của các mô hình đều khá nhanh. Ngoài ra kết quả còn cho thấy sự tối ưu của những mô hình học sâu hiện đại, khi mà với số lượng tham số ít hơn rất nhiều nhưng độ chính xác mang lại vẫn có phần hơn so với những mô hình trước đó. Cụ thể, bảng IV cho thấy ResNet-50 có số lượng tham số chỉ khoảng 1/5 VGG16, DenseNet-121 có số lượng tham số chỉ khoảng 1/16 VGG16 và tốc độ thực thi của ResNet-50, DenseNet-121 chỉ khoảng một nửa VGG16.

Những điều này minh chứng rằng FSL có hiệu quả rất tốt trong phân loại chất lượng không khí từ hình ảnh.

Bảng III: Kết quả thực nghiệm

Model	Accuracy	F1-Score
VGG16	0.8592	0.8375
ResNet-50	0.8670	0.8398
DenseNet-121	0.8703	0.8471
VGG16-FSL	0.9235	0.9106
ResNet-50-FSL	0.9308	0.9180
DenseNet-121-FSL	0.9392	0.9281

Bảng IV: So sánh thời gian thực thi

Model	Số lượng tham số	Thời gian thực thi
VGG16-FSL	138M	318 ms
ResNet-50-FSL	23M	159 ms
DenseNet-121-FSL	8M	168 ms

V. KẾT LUẬN

Nghiên cứu đã giới thiệu và đánh giá hiệu quả của FSL dựa trên prototypical networks, một phương pháp nâng cao hiệu suất phân loại chất lượng không khí từ hình ảnh dù chỉ sử dụng một lượng dữ liệu hạn chế. Kết quả thực nghiệm trên tập dữ liệu Air Pollution Image Dataset from India and Nepal cho thấy phương pháp đề xuất đạt độ chính xác cao, với chỉ số Accuracy và F1-Score đều tốt hơn so với các phương pháp hiện có. Trong tương lai, hướng nghiên cứu tiếp theo sẽ tập trung vào tối ưu và nâng cao độ chính xác với những mẫu có hình ảnh không khí tương đối giống nhau. Đồng thời thử nghiệm phương pháp với những bộ dữ liệu chất lượng không khí tại Việt Nam, nâng cao giá trị thực tế cho nghiên cứu.

TÀI LIỆU THAM KHẢO

- [1] M. Roser, "Data review: how many people die from air pollution?" *Our World in Data*, 2021, <https://ourworldindata.org/data-review-air-pollution-deaths>.
- [2] A. Parnami and M. Lee, "Learning from few examples: A summary of approaches to few-shot learning," 2022. [Online]. Available: <https://arxiv.org/abs/2203.04291>
- [3] A. Rouniyar, S. Utomo, J. A. A., and P.-A. Hsiung, "Air pollution image dataset from india and nepal," 2023. [Online]. Available: <https://www.kaggle.com/ds/3152196>
- [4] S. Utomo, A. Rouniyar, G. Jiang, C. Chang, K. Tang, H.-C. Hsu, and P.-A. Hsiung, "Eff-aqi: An efficient cnn-based model for air pollution estimation: A study case in india," 09 2023, pp. 165–172.
- [5] A. Mahajan, S. Mate, C. Kulkarni, and S. Sawant, "Predicting lung disease severity via image-based aqi analysis using deep learning techniques," 2024. [Online]. Available: <https://arxiv.org/abs/2405.03981>
- [6] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [7] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," 2015. [Online]. Available: <https://arxiv.org/abs/1511.08458>
- [8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds., vol. 25. Curran Associates, Inc., 2012. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf
- [9] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," 2014. [Online]. Available: <https://arxiv.org/abs/1409.4842>
- [10] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015. [Online]. Available: <https://arxiv.org/abs/1409.1556>
- [11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015. [Online]. Available: <https://arxiv.org/abs/1512.03385>
- [12] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," 2018. [Online]. Available: <https://arxiv.org/abs/1608.06993>
- [13] N. O' Mahony, S. Campbell, A. Carvalho, L. Krpalkova, G. Velasco-Hernandez, S. Harapanahalli, D. Riordan, and J. Walsh, "One-shot learning for custom identification tasks; a review," *Procedia Manufacturing*, vol. 38, pp. 186–193, 01 2019.
- [14] O. A. Soysal and M. Serdar Guzel, "An introduction to zero-shot learning: An essential review," in *2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 2020, pp. 1–4.
- [15] H. Gharoun, F. Momenifar, F. Chen, and A. H. Gandomi, "Meta-learning approaches for few-shot learning: A survey of recent advances," 2023. [Online]. Available: <https://arxiv.org/abs/2303.07502>
- [16] T. Müller, G. Pérez-Torró, and M. Franco-Salvador, "Few-shot learning with siamese networks and label tuning," 2022. [Online]. Available: <https://arxiv.org/abs/2203.14655>
- [17] J. Snell, K. Swersky, and R. S. Zemel, "Prototypical networks for few-shot learning," 2017. [Online]. Available: <https://arxiv.org/abs/1703.05175>
- [18] S. Utomo, A. Rouniyar, H.-C. Hsu, and P.-A. Hsiung, "Federated adversarial training strategies for achieving privacy and security in sustainable smart city applications," *Future Internet*, vol. 15, no. 11, p. 371, 2023.

AIR QUALITY CLASSIFICATION FROM IMAGES USING FEW-SHOT LEARNING

Abstract: Classifying air quality from images plays a crucial role in environmental monitoring and pollution warning, especially when sensors are often expensive and difficult to install, while cameras are becoming increasingly common and accessible. This study introduces a few-shot learning approach, applying Few-Shot Learning (FSL) to deep learning models for air quality classification. By using FSL, we develop deep learning models that can effectively learn from a small amount of training data, while achieving high accuracy in classifying air quality levels. The method was tested on the Air Pollution Image Dataset from India and Nepal, showing superior accuracy compared to traditional training methods. The approach demonstrates the potential of applying FSL in air quality classification, particularly in contexts where image data on air quality is limited.

Keywords: Few-shot learning, air quality classification, artificial intelligence.



Phạm Đức Cường , tốt nghiệp đại học chuyên ngành Công nghệ thông tin năm 2020 và thạc sĩ chuyên ngành Hệ thống thông tin năm 2023, cùng tại Học viện Công nghệ Bưu chính Viễn thông. Hiện nay đang là giảng viên khoa Trí tuệ nhân tạo, Học viện Công nghệ Bưu chính Viễn thông. Lĩnh vực yêu thích bao gồm học máy, xử lý ảnh

và thị giác máy tính.