



Tô màu cho ảnh đen trắng với opencv và deep learning Coloring black and while photos with OpenCV and deep learning

Đình Văn Chuyên^{1,*}.

¹ B&K ENGINEERING

Từ khóa:

TÓM TẮT

Tô màu cho ảnh đen trắng với openCV

Với sự phát triển của AI trong những năm gần đây, càng ngày càng có nhiều các phần mềm chỉnh sửa hình ảnh tốt hơn, giúp tiết kiệm thời gian hơn so với sử dụng các công cụ chuyên dụng như Photoshop. Một số ứng dụng hay của AI trong mảng này là tô màu cho ảnh đen trắng (Image Colorization), thêm khung hình cho video (Video frame Interpolation), tăng độ phân giải của các video (Super Resolution),.... Tuy nhiên điều này là một vấn đề khó khăn và thường đòi hỏi phải thực hiện thủ công điều chỉnh để đạt được chất lượng hình ảnh không bị nhiễu. Để giải quyết vấn đề này, phương pháp của mô hình này đưa ra giải pháp tô màu hoàn toàn tự động và đạt chất lượng cao, dựa trên cơ sở dữ liệu ảnh tham chiếu.

Keywords:

ABSTRACT

Image Colorization, Black and White image colorization with opencv

With the development of AI in recent years, there are more and more good image editing software, helping us save more time compared to using specialized tools like Photoshop. Some good applications of AI in this field are adding color to black and white photos (Image Colorization), adding frames to videos (Video frame Interpolation), increasing the resolution of videos (Super Resolution).... However, this is a challenging problem and often requires manual adjustments to achieve noise-free image quality. To address this issue, the paper proposes a fully automatic and high-quality colorization solution, based on a reference image database.

1. Giới thiệu

Việc tô màu hình ảnh là sẽ gán một màu cho mỗi điểm ảnh của ảnh đen trắng. Các phương pháp tô màu được thực hiện thường được chia làm hai nhóm: tô màu bằng phương pháp dựa trên các nét màu cơ bản [1-5] và tô màu dựa trên các mẫu [6],

[6-11]. Các phương pháp tô màu này đòi hỏi người tô phải có khiếu thẩm mỹ, khéo léo... Đặc biệt là tốn rất nhiều thời gian và công sức, đặc biệt là đối với những người mới làm quen và những hình ảnh trắng đen có những chi tiết nhỏ.

Để giảm bớt một phần khó khăn cho người tô màu, T.Welsh, M.Ashikhmin và K.Mueller [11] (tác

* Đình Văn Chuyên. B&K ENGINEERING
Email: chuyen.dinh76@gmail.com

<https://www.doi.org/10.55228/JTST150101>

Ngày nhận bài: 20/08/2025; Ngày nộp bài sửa: 20/10/2025 ; Ngày chấp nhận đăng: 10/12/2025

Ngày xuất bản trực tuyến: 15/01/2026

pISSN: 1859-4263; eISSN: 3030-4261

giả của mô hình chuyển màu sắc vào ảnh đen trắng. Transferring Color to Greyscale Images) đề xuất một phương pháp dựa trên tập ảnh mẫu có màu sắc và chuyển các giá trị sắc độ và độ sáng của ảnh màu sang ảnh đen trắng tương ứng và sau đó được G.Charpiat, M.Hofmann và B.Scholkopf, R.Irony cải tiến bằng phương pháp xử lý trực tiếp trên đa thức (multimodality) và ước lượng cho từng điểm ảnh cần tô màu và phân phối xác suất của tất cả các màu có thể thay đổi để hoàn thiện mô hình tô màu [6], [9]. Phương pháp này lấy thông tin màu từ ảnh tham chiếu tô màu cho ảnh đen trắng. Tuy nhiên, việc tìm kiếm hình ảnh tham chiếu phù hợp gây khó khăn cho người tô màu. A. Y.-S. Chia, S. Zhuo, R. K. Gupta, Y.-W. Tai, S.-Y. Cho, P. Tan, and S. Lin [7], [10] đã đơn giản hoá phương pháp này bằng cách sử dụng dữ liệu hình ảnh trên Internet và đề xuất các chương trình lọc phù hợp để chọn ảnh tham chiếu phù hợp.

Tuy nhiên, cả hai phương pháp trên có những hạn chế nhất định. X. Liu, L. Wan, Y. Qu, T.-T. Wong, S. Lin, C.-S. Le- ung, and P.-A. Heng et al [10] tìm những đối tượng giống nhau trên Internet để sử dụng màu của ảnh tham chiếu tìm được tô cho ảnh đen trắng. Phương pháp này bị giới hạn bởi các vật thể có hình dạng ảnh không thể chia nhỏ ra được (ví dụ như: địa danh). A. Y.-S. Chia, S. Zhuo, R. K. Gupta, Y.-W. Tai, S.-Y. Cho, P. Tan, and S. Lin et al [7] yêu cầu người dùng cung cấp văn bản mô tả và phân loại cho đối tượng, phong nền. Trong thực hành, các tín hiệu phân loại thủ công khó có thể có được vì hình ảnh đen trắng chứa nhiều chi tiết và đối tượng phức tạp (ví dụ: toà nhà, cây cối, oto, con voi ...). Các phương pháp này có cùng hạn chế, hiệu suất phụ thuộc rất nhiều vào hình ảnh tham chiếu được chọn.

Một phương pháp tô màu hoàn toàn tự động được đề xuất để giải quyết hạn chế này. Theo trực giác, một hình ảnh tham chiếu không thể bao gồm tất cả các tình huống có thể xảy ra trong ảnh đen trắng cần tô màu. Do đó, G. Charpiat, M. Hofmann, and B. Schölkopf, A. Y.-S. Chia, S. Zhuo, R. K. Gupta, Y.-W. Tai, S.-Y. Cho, P. Tan, and S. Lin. [6], [7], [9], [11] yêu cầu hình ảnh tham chiếu tương tự, giống với ảnh cần tô màu. Một giải pháp đáng tin cậy hơn là xác định vị trí điểm/điểm ảnh của tập ảnh dữ

liệu tham chiếu, sau đó chuyển thông tin màu đó đến điểm/điểm ảnh của ảnh cần tô màu. Tuy nhiên sẽ sinh ra nhiễu quá lớn khi sử dụng cơ sở dữ liệu lớn và làm chậm và tổn hao tài nguyên của hệ thống xử lý.

Cho đến nay, các kỹ thuật mạng học sâu đã đạt được những thành công to lớn trong việc mô hình hoá các dữ liệu lớn. Nó đã cho thấy khả năng học tập mạnh mẽ thậm chí còn vượt trội hơn con người ở một số mức độ (ví dụ [12]) và các kỹ thuật mạng học sâu đã được chứng minh là rất hiệu quả trên nhiều ứng dụng xử lý hình ảnh và thị giác máy tính bao gồm, phân loại hình ảnh [13], phát hiện người đi bộ [14], [15] siêu phân giải hình ảnh [16], điều chỉnh ảnh [17]. v.v. Sự thành công của các kỹ thuật mạng học sâu thúc đẩy chúng ta khám phá ứng dụng tiềm năng của nó trong bối cảnh của chúng ta. Mục đích của nghiên cứu là xây dựng công thức tô màu hình ảnh như sử dụng mạng neural nhân tạo hồi qui và deep learning. Cơ sở hình ảnh tham chiếu (tập mẫu) bao gồm nhiều đối tượng (như: cây cối, động vật, toà nhà, biển, núi, đồng bằng, cao ốc...v...v) được sử dụng để đào tạo mạng neural. Một số ví dụ được trình bày trong Hình 1. Tốc độ đào tạo rất chậm do tập dữ liệu mẫu lớn. Mô hình sau khi đào tạo có thể sử dụng trực tiếp để tô màu cho ảnh đen trắng một cách hiệu quả nhất. Với phương pháp này đạt yêu cầu tương thích giữa ảnh màu tham chiếu và ảnh đen trắng cần tô.

Gần đây, các nhà nghiên cứu đã chứng minh được rằng ảnh với độ phân giải cao rất hữu ích cho việc giải quyết các vấn đề về thị giác máy tính (như: cải thiện chất lượng hình ảnh [17], phát hiện cạnh của vật thể [18]). Với mục đích tô màu cho ảnh thường kết hợp ngôn ngữ mô tả, bài báo này đề xuất một tính năng mô tả mới để kết hợp nhận thức ngôn ngữ mô tả vào quá trình tô màu hình ảnh dùng tập dữ liệu ảnh mẫu. Cụ thể, tính năng trong mô hình đề xuất đã đào tạo mạng hình ảnh tham khảo từ các trang web khác nhau như có thể thấy trong Hình 1 (b). Mô hình đã được đào tạo và được sử dụng để tô màu nhiều ảnh đen trắng khác nhau trong Hình 1 (a). Màu sắc kết quả thể hiện trong Hình 1 (c) minh chứng tính mạnh mẽ và hiệu quả của phương pháp đề xuất.



Hình 1. Kết quả tô màu tự động. (a) Ảnh lợp vào, (b) Ảnh trong tập dữ liệu mẫu, (c) Kết quả tô màu.

2. Thực trạng của vấn đề tô màu cho ảnh đen trắng

Việc tô màu cho ảnh đen trắng đã được nhiều nhà khoa học nghiên cứu và ứng dụng, có thể phân chia theo một số xu hướng sau.

Tô màu dựa trên các nét màu cơ bản. Đề xuất này là công trình của Levin [2] đã đề xuất một cách tiếp cận hiệu quả đòi hỏi người tô màu phải cung cấp các nét vẽ màu cơ bản trên ảnh đen trắng. Sử dụng hàm tối ưu hoá bình phương nhỏ nhất (least-square optimization) tính toán và tô màu cho phần còn lại của ảnh. Huang et al. [1] phát triển thuật toán phát hiện cạnh thích ứng để giảm hiện tượng tràn màu xung quanh biên của một vùng. Yatziv et al. [5] tô màu các điểm ảnh bằng giá trị điểm ảnh của các nét vẽ màu của người tô. Qu et al. [4] và Luan et al. [3] sử dụng tính năng lân cận điểm để giảm số lượng nét vẽ của người tô.

Tô màu theo ảnh mẫu, không giống phương pháp tô màu dựa trên nét màu, tô màu theo ảnh mẫu sử dụng một ảnh màu làm mẫu tô màu cho ảnh đen trắng. Phương pháp này phân làm hai mô hình chính phụ thuộc vào nguồn tham khảo, gồm:

- Phương pháp điền màu theo ảnh mẫu được cung cấp. Trong phương pháp này yêu cầu cung cấp hình ảnh tham chiếu phù hợp. Với công nghệ xử lý ảnh [19] và công nghệ phối màu [20], Welsh et al. [11] sử dụng cường độ và thống kê của các điểm ảnh lân cận để tìm những điểm ảnh giống nhau trong ảnh tham chiếu sau đó chuyển giá trị tham chiếu vào ảnh cần tô màu. Phương pháp này được cải thiện trong [9] bằng cách tính đến đặc tính liên kết lân cận của điểm ảnh. Charpiat et al. [6] đề xuất một thuật toán tối ưu hoá toàn cục để

tô màu cho một điểm ảnh. Gupta et al. [8] phát triển một phương pháp tô màu dựa trên siêu điểm ảnh để cải thiện tính nhất quán về không gian. Các phương pháp này chất lượng màu phụ thuộc vào hình ảnh tham chiếu. tuy nhiên không có tiêu chuẩn cho hình ảnh tham chiếu do đó tìm được một ảnh tham chiếu phù hợp là một nhiệm vụ khó khăn.

- Tô màu bằng hình ảnh tham chiếu trên Internet. Để giảm bớt khó khăn trong việc tìm kiếm ảnh tham chiếu phù hợp, Liu et al. [10] và Chia et al. [7] sử dụng dữ liệu hình ảnh tham chiếu lấy trên Internet. Liu et al. [10] tính toán phân chia các tính chất vật lý của ảnh dựa vào tập ảnh tham chiếu từ Internet. Phương pháp này đã phân biệt được sự khác nhau về độ chói (illumination) giữa ảnh tô và ảnh tham chiếu, nhưng đồng thời phương pháp này đòi hỏi ảnh chứa các đối tượng/cảnh phải giống nhau để có độ chính xác theo điểm ảnh của ảnh tham chiếu và ảnh được tô màu. Phương pháp này không thể tô màu cho ảnh động có những yếu tố như người, xe... giữa ảnh tham chiếu và ảnh cần tô, vì những yếu tố này bị loại trừ trong quá trình tính toán các tham số vật lý của ảnh. Do đó phương pháp này bị giới hạn ở những ảnh là cảnh vật tĩnh và các đối tượng/ cảnh trong hình.

3. Mô hình đề xuất

3.1. Tổng quan

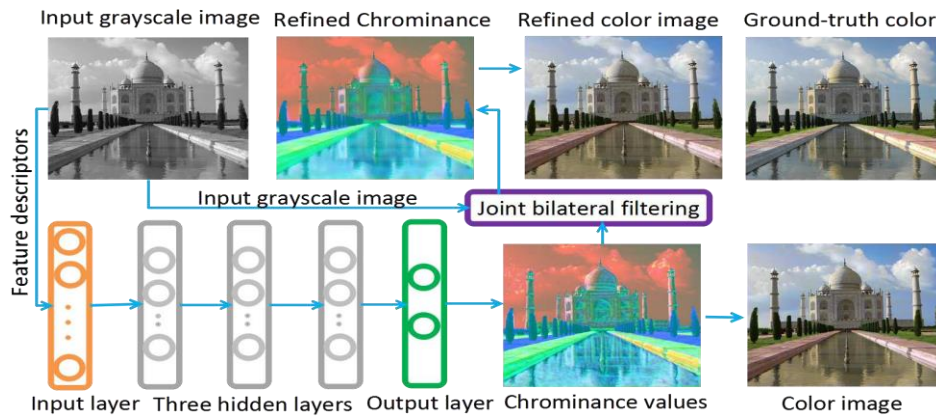
Tổng quan về các phương pháp tô màu được đề xuất, trình bày trong Hình 2. Tương tự như các phương pháp tiếp cận dựa trên mạng neural học sâu khác. Hướng nghiên cứu này đề xuất một phương pháp tô màu đáp ứng hai mục tiêu:

Mục tiêu 1: Đào tạo mạng neural sử dụng một tập dữ liệu lớn các hình ảnh tham chiếu;

Mục tiêu 2: Sử dụng mạng neural sau khi đã được huấn luyện để tô màu cho hình ảnh đen trắng.

Trên cơ sở các mục tiêu nghiên cứu, tác giả đề xuất phương pháp tô màu hình ảnh dựa trên học

tập sâu đầu tiên và thể hiện hiệu quả của nó trên nhiều cảnh khác nhau, đồng thời đi sâu phân tích rõ ràng các mô tả, thông tin, tính năng của hình ảnh, phân biệt rõ ràng các mô tả tính năng từ mức thấp đến cao, đó là chìa khóa thành công của phương pháp tô màu được đề xuất.

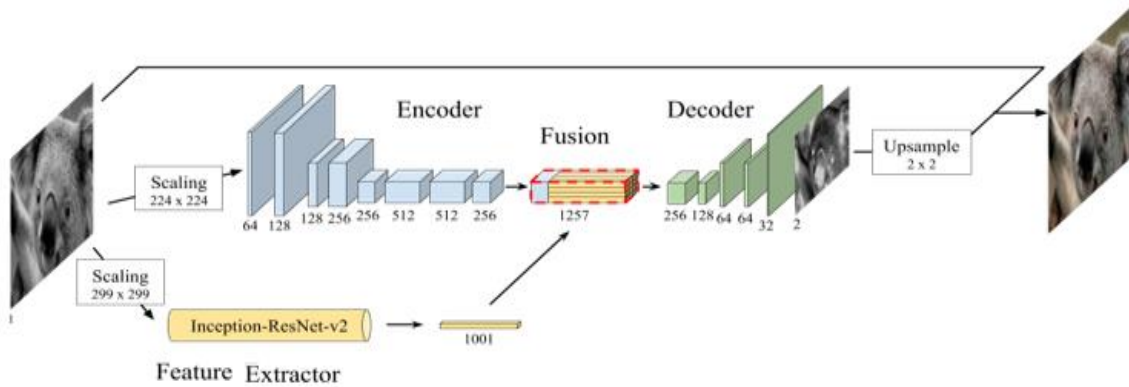


Hình 2. Mô hình tổng quan về phương pháp tô màu đề xuất.

3.1.1. Nội dung của mô hình đề xuất

Nội dung 1: Từ tập ảnh tham chiếu xây dựng thuật toán huấn luyện mạng neural học sâu để sử dụng tập dữ liệu gồm:

- (1) Tập ảnh dữ liệu huấn luyện ImageNet Dataset;
 - (2) Xây dựng và huấn luyện mạng neural;
- Kiến trúc tổng quát của mô hình đề xuất như trên Hình 3.

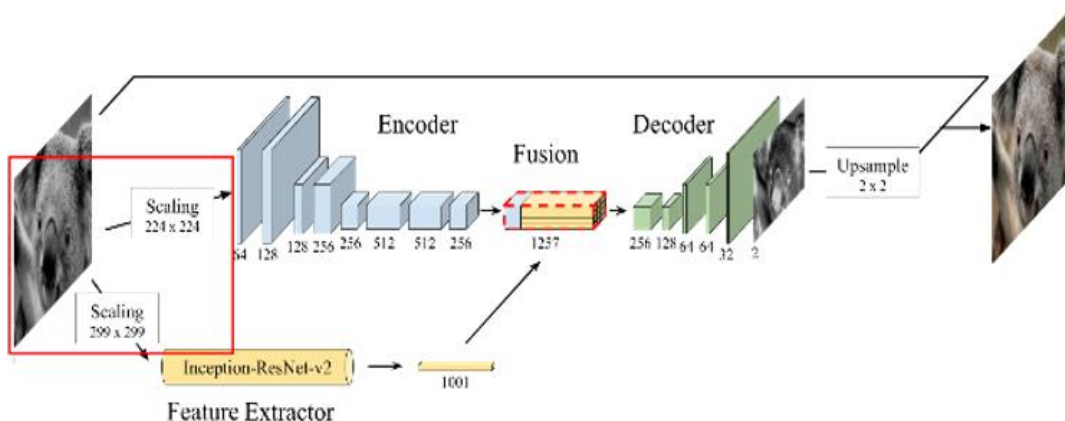


Hình 3. Mô hình đề xuất.

Nội dung chính gồm các phần sau:

- * Cài đặt kích thước đầu vào ảnh huấn luyện:
- Chọn Inception-ResNet-v2 là 299x299;

- Chọn kích thước cho mô hình nghiên cứu là 244x244;
 - Giá trị điểm ảnh trung tâm là [-1, 1].
- Mô hình đặc tính ảnh đầu vào như trên hình 4.



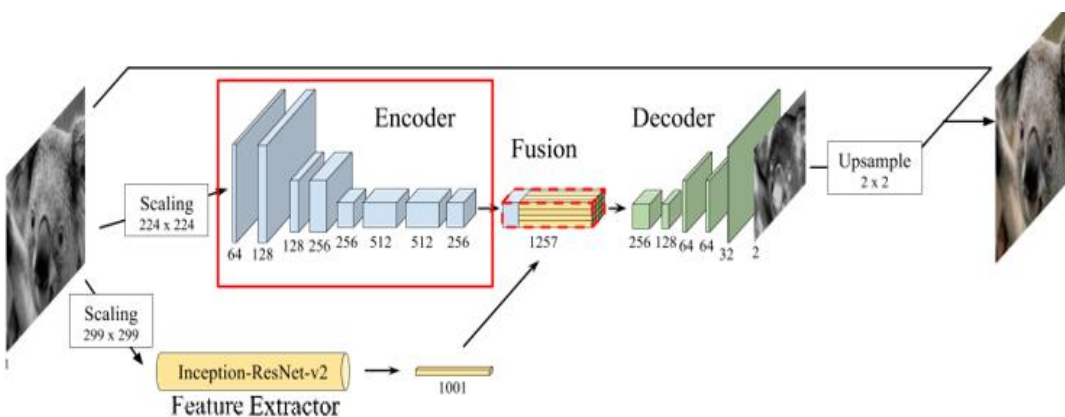
Hình 4. Mô hình đặc tính ảnh đầu vào.

* Mã hoá:

- Input: $H \times W \times 1$ (L component);
- Output: $H/8 \times W/8 \times 512$;
- 8 lớp tích chập kích thước 3×3 .

- Bước 1 và đệm cho dữ liệu kích thước vào;
- Bước 2 giảm một nửa kích thước đầu vào.

Mô hình mã hóa như trên hình 5.

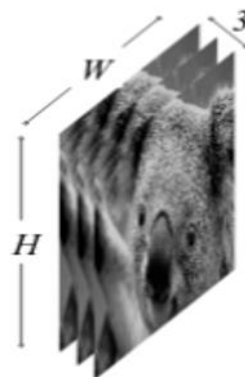


Hình 5. Mô hình mã hóa.

* Trích xuất đặc tính

- Sử dụng mô hình Inception được huấn luyện trước.
- Lặp lại độ chói 3 lần trên cùng ảnh (kích thước cụ thể).

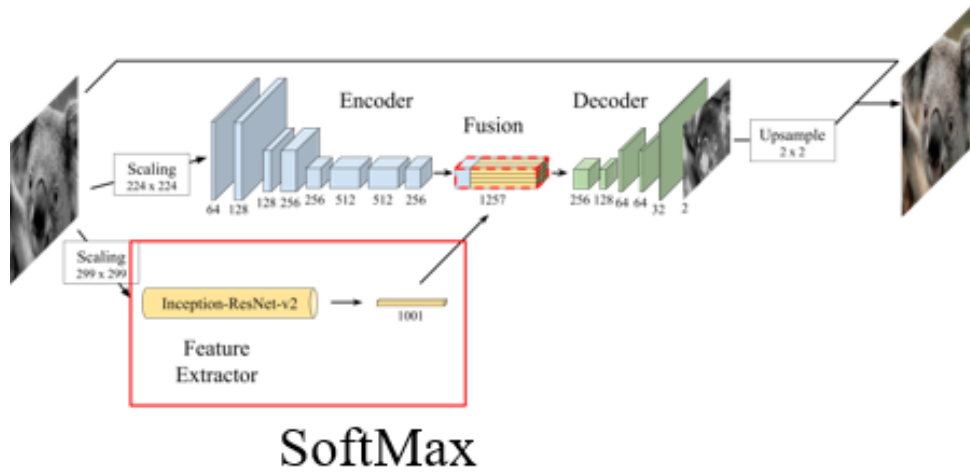
Mô hình ảnh xếp chồng theo kích thước như trên hình 6.



Hình 6. Mô hình ảnh xếp chồng có yêu cầu kích thước.

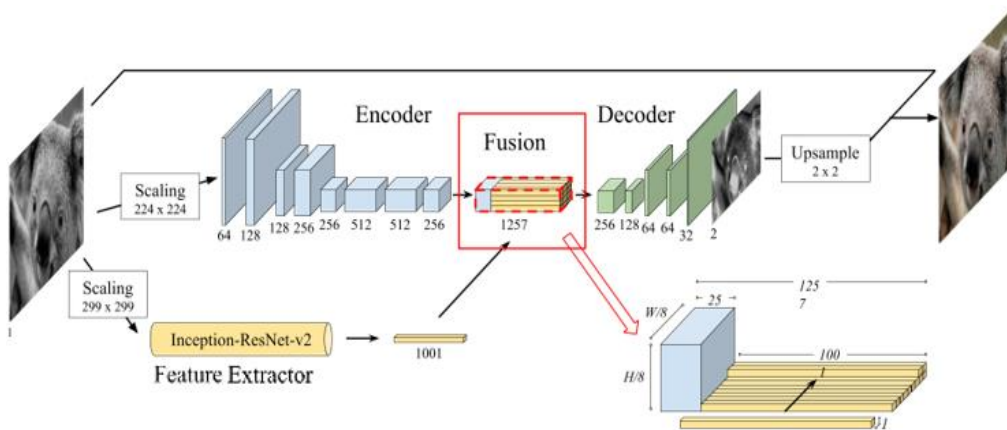
- * Trích xuất các đặc tính từ lớp cuối cùng trước khi sử dụng hàm chức năng SoftMax.

Mô hình trích xuất đặc tính như hình 7.



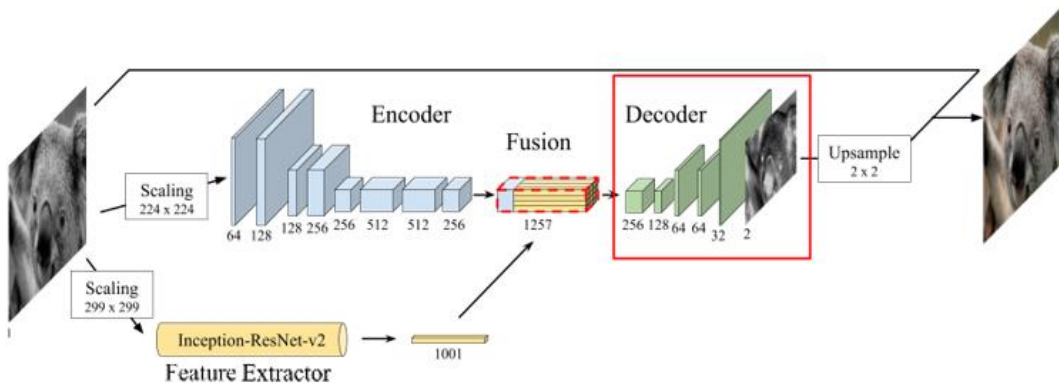
Hình 7. Mô hình trích xuất đặc tính.

- * Liên kết lớp (Fusion layer)
 - Giảm kích thước HW/8² lần;
 - Mã hoá thể tích theo độ sâu của trục;
 - Độc lập với kích thước ban đầu.
- Mô hình liên kết lớp như trên hình 8.



Hình 8. Mô hình liên kết lớp ẩn.

- * Giải mã
 - Kết quả output: H x W x 2 (hai kênh màu a*b).
- Mô hình giải mã như trên Hình 9



Hình 9. Mô hình giải mã.

Nội dung 2: Huấn luyện mạng được thực hiện theo các bước sau:

(1) Sử dụng hàm lỗi bình phương trung bình dự đoán giữa các màu pixel và màu thực của ảnh.

$$(X, \theta) = \frac{1}{2HW} \sum_{k \in \{a,b\}} \sum_{i=1}^H \sum_{j=1}^W (X_{ki,j} - \widehat{X}_{ki,j})^2 \quad (1)$$

(2) Hội tụ theo thuật toán tối ưu Adam. ($\eta=0.001$).

(3) Cố định kích thước hình ảnh đầu vào để cho phép xử lý hàng loạt.

(4) Tập mẫu hơn 1.400.000 ảnh.

(5) Truy cập miễn phí từ ImageNet Data.

(6) Dùng 60.000 ảnh cho huấn luyện mạng.

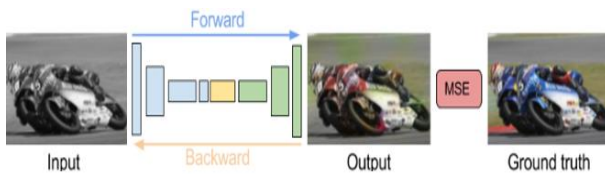
(7) 90% được huấn luyện – 10% validation/test.

(8) Batch size 100 ảnh.

(9) 23 giờ huấn luyện mạng.

(10) Sử dụng mô hình Keras của Tensorflow backend.

Mô hình huấn luyện mạng được mô tả như trên hình 10.



Hình 10. Mô hình huấn luyện mạng.

3.2. Mô hình học sâu (deep learning model) để tô màu hình ảnh.

Nội dung phần này là xây dựng bài toán tô màu hình ảnh, sử dụng thuật toán hồi quy và giải quyết bài toán bằng mô hình mạng neural học sâu (deep neural network).

3.2.1. Thuật toán

A deep neural network (DNN) là một loại mạng thần kinh nhân tạo đặc trưng bằng cách có nhiều lớp giữa các lớp đầu vào và đầu ra. Các mạng này được thiết kế để mô hình hóa các mẫu và biểu diễn phức tạp trong dữ liệu, cho phép chúng thực hiện các tác vụ như phân loại, hồi quy và trích xuất tính năng [21]. Tập dữ liệu đào tạo $\Lambda = \{ \vec{G}, \vec{C} \}$ trong đó \vec{G} tập ảnh đen trắng, \vec{C} là ảnh màu tương ứng, phương pháp của mô hình là dựa trên tiền đề: Tồn

tại một ánh xạ từ Λ , dùng hàm ánh xạ F chuyển đổi ảnh đen trắng sang ảnh màu. Đối với điểm ảnh p trong \vec{G} , đầu ra của hàm F chỉ đơn giản là các kênh a và b của các điểm ảnh tương ứng trong \vec{C} và đầu vào của hàm F là các mô tả các điểm đặc trưng của điểm ảnh p . Điểm đặc trưng được giới thiệu chi tiết trong phần 3.2. Phương trình chuyển đổi ảnh đen trắng sang ảnh màu như sau:

$$c_p = F(\theta, x_p)^2 \quad (2)$$

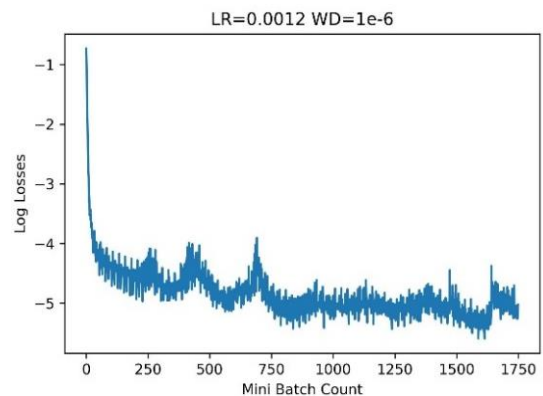
Trong đó: x_p đặc tính được rút trích từ điểm ảnh p ; c_p giá trị độ sáng; θ tham số học của hàm ánh xạ F học được từ Λ .

Giải phương trình bằng hàm cực tiểu bình phương (least squares minimization) để tìm cực tiểu của thông số θ :

$$\operatorname{argmin}_{\theta \in Y} \sum_{p=1}^n \|F(\theta, x_p) - c_p\|^2 \quad (3)$$

Trong đó: n là tổng các điểm ảnh huấn luyện tạo được lấy mẫu từ Λ và Y là hàm không gian của $F(\theta, x_p)$

Kết quả giải ra ta tìm được giá trị hàm mất mát như trên Hình 11.



Hình 11. Giá trị hàm mất mát.

3.2.2. Kiến trúc mạng nơ-ron

Mạng deep neural networks (DNNs) thường bao gồm một lớp vào, nhiều lớp ẩn và một lớp ra. Mỗi lớp có thể bao gồm nhiều nơ-ron khác nhau. Trong mô hình này sử dụng số nơ-ron đầu vào bằng với kích thước của bộ mô tả đặc điểm được trích xuất từ mỗi vị trí điểm ảnh của ảnh đen trắng, lớp đầu ra có hai nơ-ron kênh a và b tương ứng với giá trị màu của điểm ảnh trong ảnh đen trắng. Đặt

số nơ-ron trong lớp ẩn bằng một nửa số nơ-ron đầu vào. Mỗi nơ-ron trong lớp ẩn hoặc lớp đầu ra được kết nối với tất cả các nơ-ron trong lớp tiếp theo và mỗi kết nối được liên kết với một trọng số. Giả sử \mathbf{o}_j^l là nơ-ron đầu ra thứ j trong lớp l , \mathbf{o}_j^l có thể được tính như sau:

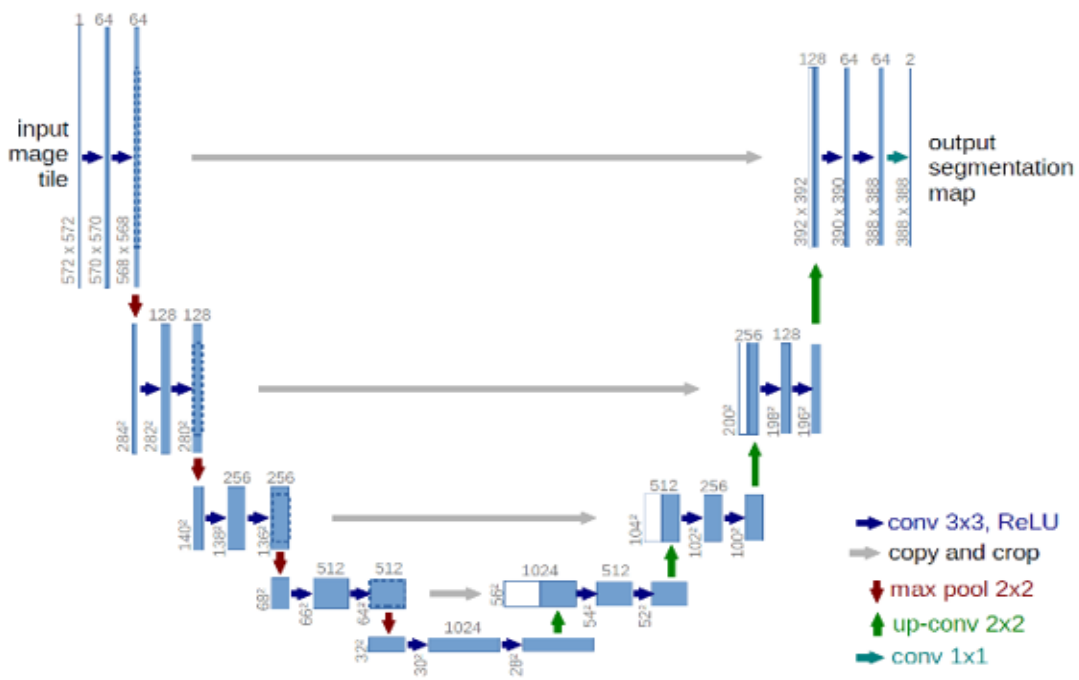
$$\mathbf{o}_j^l = f(w_{j0}^l \mathbf{b} + \sum_{i>0} w_{ji}^l \mathbf{o}_i^{l-1}) \quad (4)$$

Trong đó: w_{ji}^l là trọng số kết nối giữa nơ-ron j^{th} trong lớp l^{th} . Nơ-ron i^{th} trong lớp $(l-1)^{\text{th}}$.

\mathbf{b} là giá trị nền (bias) của nơ-ron, \mathbf{b} là một hằng số ngõ ra.

$f(z)$ là hàm kích hoạt không tuyến tính (như hàm tanh, sigmoid, ReLU[13]).

Nơ-ron ngõ ra của lớp ra là sự liên kết với trọng số của các đầu ra của các nơ-ron trong lớp tiếp theo.



Hình 12. Kiến trúc mạng U-net.

Trong mô hình này sử dụng hàm kích hoạt là hàm ReLU[13] để làm tăng tốc độ hội tụ của quá trình đào tạo mạng. Ở đây áp dụng thuật toán lan truyền ngược để sửa lỗi và điều chỉnh tham số học θ . Kiến trúc mạng nơ-ron được trình bày trong hình 12 và hình 2.

3.3. Diễn giải phần trích chọn đặc trưng

Thiết kế bám theo đặc tính của đối tượng được tô là phương pháp được áp dụng vào các bài toán tô màu hiện nay. Có rất nhiều thông số ảnh hưởng đến quá trình đào tạo mạng như là: (SIFT, SURF, Gabor, Location, Intensity histogram...). Trong mô hình này đúc kết một vài kinh nghiệm thử nghiệm tô màu, giữ lại các đặc điểm tác động đến kết quả tô màu. Phân tích các đặc điểm được áp dụng thành các đặc điểm cấp thấp, trung bình và

cao. Giả sử x_p^L, x_p^M, x_p^H biểu thị các mô tả đặc điểm cấp độ khác nhau được trích xuất tại điểm ảnh p , nối các đặc điểm này lại với nhau thành:

$$x_p = \{x_p^L; x_p^M; x_p^H\} \quad (5)$$

Xây dựng mô tả phần trích, chọn đặc trưng của hình ảnh.

Ý tưởng chính của mạng U-Net gần giống như một auto-encoder, từ một ảnh ban đầu, sử dụng các lớp Conv2D để phân tích và trích chọn đặc trưng, thu nhỏ đưa về lớp kết nối hoàn toàn, sau đó dùng Conv2D Transpose để phóng to lại về kích thước ảnh ban đầu. Việc trích đặc trưng này làm cho mạng neuron thường khá sâu, chính vì vậy tại các lớp cuối, mô hình có thể bị quên đi các đặc trưng từ ảnh ban đầu hoặc đặc trưng được trích ra

từ những lớp ban đầu, vậy nên U-Net sẽ đưa các ảnh từ các lớp ban đầu vào quá trình huấn luyện của các lớp sau như 1 tham số để tránh tình trạng đó.

3.3.1. Tính năng so sánh các vùng lân cận của điểm ảnh (low level patches feature).

* Giá trị điểm ảnh đơn:

- Trong hình ảnh màu, các điểm ảnh có cùng độ sáng (luminance), nhưng có màu khác nhau độ sắc (chrominance);
- Chỉ sử dụng các giá trị độ chói là không đủ để biểu diễn các điểm ảnh chính xác.

* Giải pháp dựa vào tính năng so sánh các điểm ảnh lân cận:

- Sử dụng vùng lân cận có kích thước $\{7 \times 7\}$ xung quanh mỗi điểm ảnh p mô tả đặc tính của giá trị x_p ;
- Nắm được cấu trúc đặc điểm của các điểm ảnh lân cận cho kết quả tốt hơn giá trị của một điểm ảnh đơn;
- Phân biệt được độ chói của các điểm ảnh lân cận.

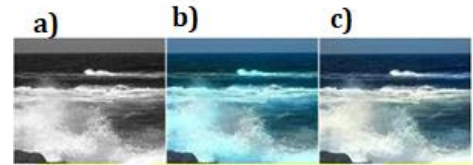
* Ưu điểm

- Vượt trội hơn so với các mô hình truyền thống như: (SIFT, DAISY) trong vùng điểm ảnh mật độ thấp;
- Tránh được sự đồng nhất về màu của các vùng khác nhau trong vùng lân cận (ví dụ: màu của biển giống nhau trên toàn ảnh);
- Tạo được độ nhạy giá trị với các điểm ảnh lân cận.

* Kết quả của phương pháp:

Với phương pháp trích xuất này sẽ tạo ra mẫu không gian cục bộ (được lưu lại trong quá trình so sánh), tạo ra tập dữ liệu mạnh hơn so với sử dụng giá trị điểm ảnh đơn lẻ.

Kết quả đánh giá tính năng so sánh điểm lân cận của điểm ảnh như trên hình 13.



Hình 13. Tính năng so sánh điểm lân cận. (a) Ảnh lớp vào, (b) Lân cận điểm, (c) Tổng hợp tính năng.

3.3.2. Phương pháp cấp trung của mô hình DAISY (Mid-level DAISY feature).

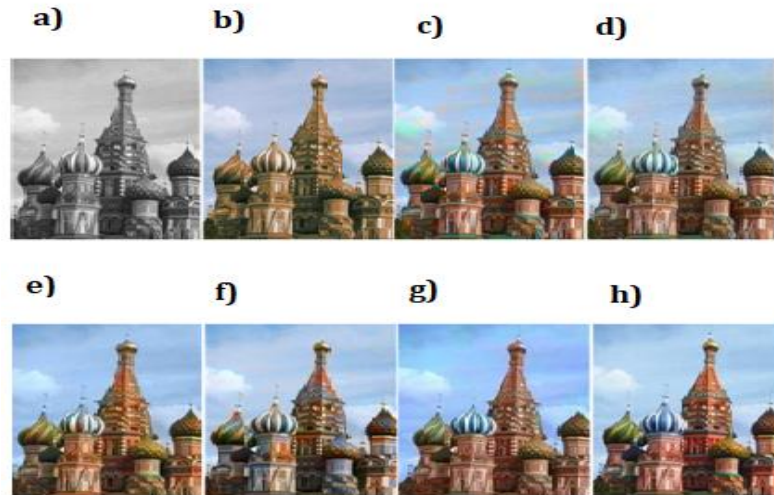
Mô hình DAISY mô tả trực tiếp biến cục bộ để đạt được sự đồng nhất [22]. Không giống như phương pháp so sánh các vùng lân cận của điểm ảnh. Mô hình DAISY tính toán các đặc tính của mỗi điểm ảnh, thay đổi cho nó phù hợp với yêu cầu của các điểm ảnh lân cận.

Nội dung mô hình DAISY trong tô màu gồm:

(1) Đảm bảo được kết cấu và cấu trúc của ảnh

Mô hình DAISY mã hoá các mẫu hình học ở lớp trung gian (cạnh, góc, và các lớp chuyển dịch ngắn), giúp duy trì sự gắn kết cấu trúc trong hình ảnh màu.

Không giống như phương pháp so sánh điểm ảnh lân cận (dựa trên các lân cận điểm ảnh đơn), mô hình DAISY cung cấp một cách trừu tượng hơn về các gradien lân cận điểm ảnh. Hình 14 minh hoạ kết quả có và không có áp dụng mô hình DAISY trong cấu trúc của ảnh.



Hình 14. Tính năng của mô hình DAISY. (a) Ảnh đầu vào, (b) Ảnh không sử dụng mô hình DAISY, (c) Ảnh sử dụng mô hình DAISY, (d) Ảnh sử dụng mô hình DAISY và lọc nhiễu, (e) Mô hình Gupta, (f) Mô hình Irony, (g) Mô hình Chia, (h) Mô hình Liu.

(2) Nhạy cảm với sự thay đổi độ sáng. Mô hình DAISY dựa trên sự thay đổi các hướng của gradien (như SIFT). Mô hình DAISY ít nhạy cảm với sự thay đổi độ chói tuyệt đối, cải thiện tính khái quát tổng thể.

(3) Bổ xung cho phương pháp so sánh điểm ảnh lân cận (Complementary to low-level patches). Với phương pháp so sánh các điểm ảnh lân cận (ví dụ sử dụng kích thước $\{7 \times 7\}$ của các điểm ảnh lân cận trong ảnh đen trắng), trích xuất các đặc tính. Mô hình DAISY cung cấp các tín hiệu theo ngữ cảnh rộng hơn, giúp giải quyết sự không rõ ràng trong kết cấu của các vùng kết cấu không rõ ràng.

(4) Tác động đến sự tô màu của mô hình DAISY.

Các tính năng của mô hình DAISY thường được trích xuất ở nhiều thang đo và được nối thành một mô tả cho mỗi điểm ảnh.

Kết hợp với các tính năng so sánh các điểm ảnh lân cận, giúp sự tô màu sẽ hợp lý hơn, đặc biệt là ở các vùng không đủ độ chói (vd. Bề mặt phẳng với kết cấu không rõ ràng).

(5) Ưu điểm so với tính năng so sánh điểm ảnh lân cận (SIFT).

Mật độ tính toán các điểm ảnh dày hơn. Trích xuất tại mỗi điểm ảnh thay vì với các điểm ảnh lân cận. Hiệu quả: Tối ưu hoá để trích xuất nhanh hơn, làm cho sự điền màu cho ảnh được đầy đủ hơn.

(6) Giới hạn.

Tuy nhiên mô hình DAISY không phù hợp cho với những vùng/ đối tượng có độ đồng nhất thấp, với những đối tượng này sẽ bị giảm hiệu suất tô xung quanh các vùng như thấy trong Hình 14(c). Một cách xử lý khác sẽ được giới thiệu trong phần 3.3.4 để giảm hiện tượng nhiễu và kết quả trong Hình 14(d). Hơn nữa, chúng ta có thể thấy kết quả của bài báo có thể so sánh với kết quả của Liu et al. [10] (Yêu cầu các đối tượng được tô phải có hình dạng cố định không thay đổi hình dạng khi thay đổi kích thước hình học và giống với đối tượng được tham chiếu), và Chia et al. [7] (Yêu cầu phân loại thủ công và xác định các đối tượng nền), mặc dù phương pháp của mô hình là tự động hoàn toàn.

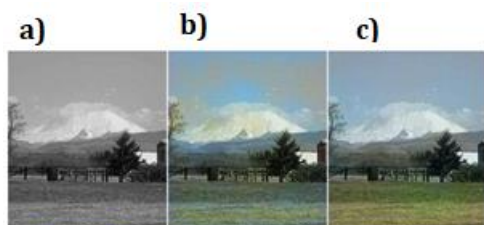
3.3.3. Mô tả đặc tính điểm ảnh

Phương pháp so sánh điểm lân cận (patches) và mô hình DAISY là phần trích xuất bậc thấp và bậc trung, chúng dựa vào cấu trúc hình học của các điểm ảnh lân cận. Các phương pháp hiện đại thường sử dụng các tính năng như vậy để tìm điểm giống nhau giữa ảnh đen trắng và ảnh tham chiếu. Những năm gần đây, các trích xuất bậc cao của hình ảnh đã chứng minh được tầm quan trọng và giá trị của nó trong một số lĩnh vực (ví dụ: cải thiện hình ảnh [17], phát hiện cạnh [18]). Chúng ta xem rằng tô màu là một quá trình nhận thức. Trong bài báo này sẽ trích xuất đặc tính tại mỗi điểm ảnh để miêu tả đặc tính thuộc của điểm ảnh (ví dụ: bầu trời, biển, động vật, con người...).

Sử dụng thuật toán phân tích mới nhất của I.Long [23] để chú thích mỗi điểm ảnh và dán nhãn cho nó, tạo thành tập dữ liệu đầu vào. Tập dữ liệu đầu vào với những đặc tính không rõ ràng ở những vùng biên. Kết quả là, nó được làm mịn lọc bỏ nhiễu bằng cách sử dụng bộ lọc biên tích cực của E.S.Gastal [24] với sự tham chiếu của ảnh đen trắng đầu vào. Một vecto xác suất N chiều sẽ được tính toán tại mỗi vị trí điểm ảnh, trong đó N là tổng số đặc tính và mỗi phần tử là xác suất mà điểm ảnh hiện tại thuộc về đặc tính tương ứng. Vecto xác suất này được sử dụng như là mô tả cấp cao được ký hiệu là x^H .

Hình 15 cho thấy kết quả tô màu có thể thay đổi đáng kể và không có đặc tính rõ ràng.

Việc áp dụng đặc tính rõ ràng giúp giảm đáng kể sự không rõ ràng trong khi huấn luyện mạng. Ví dụ, nếu một điểm ảnh được phát hiện nằm trong vùng bầu trời, chỉ các giá trị màu của bầu trời nằm trong tập dữ liệu màu của bầu trời tham chiếu mới được sử dụng. Do đó, vấn đề tô màu được đơn giản hoá sau khi tích hợp thông tin rõ ràng và kết quả tô màu sẽ trực quan hơn nhiều.

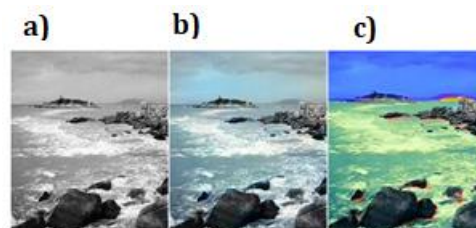


Hình 15. Kết quả tô màu khi các đặc tính của ảnh được mô tả chi tiết. (a) Ảnh đen trắng lớp vào, (b) Sử dụng lân cận điểm và mô hình DAISY, (c) Mô hình DAISY và mô tả đặc điểm của ảnh.

3.3.4. Tinh chỉnh độ sắc

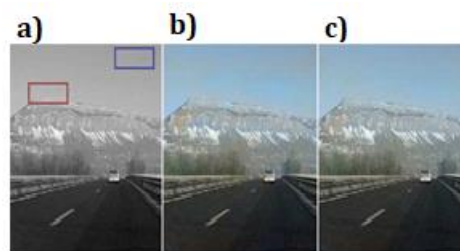
Phương pháp này đề xuất áp dụng hai tính năng so sánh điểm lân cận và mô hình DAISY. Sử dụng tính năng so sánh điểm lân cận để mô tả các vùng đơn giản có cấu tạo đơn giản và mô hình DAISY để mô tả cấu trúc và vùng phức tạp. Tuy nhiên, bài báo chỉ đơn giản là nối hai tính năng thay vì đào tạo ra một sự kết hợp tốt hơn giữa hai phương pháp. Điều này sẽ dẫn đến một số các chi tiết trong ảnh bị ẩn nếu là xung quanh là môi trường có mối tương đồng về màu như là: bầu trời, biển... Mô

hình DAISY tạo ra các giá trị âm đối với những chi tiết này. Hình 16 mô tả đánh giá tính năng tổng quát của các mô hình.

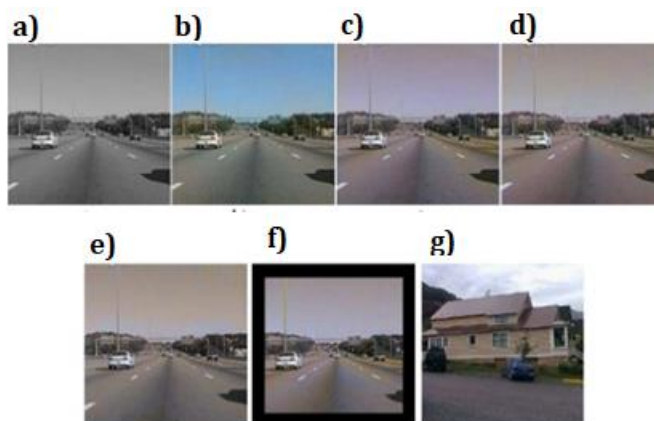


Hình 16. Kết quả đánh giá tính năng tổng quát của mô hình. (a) Ảnh đen trắng lớp vào, (b) Kết hợp giữa các mô hình, (c) Kết hợp các mô hình và thuật toán hiệu chỉnh của nghiên cứu.

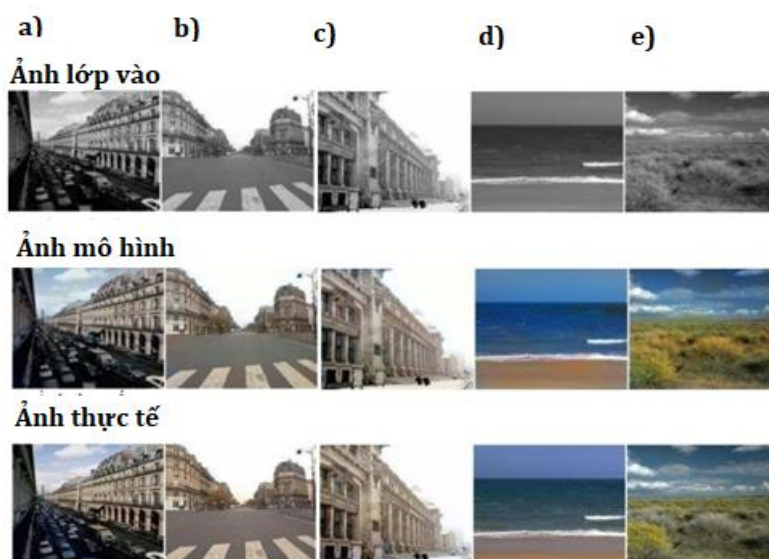
Các đối tượng xung quanh vùng có cấu tạo đơn giản có thể được giảm đáng kể bằng cách sử dụng kỹ thuật lọc (joint bilateral filtering technique [25]). Kỹ thuật này lần đầu tiên được giới thiệu để loại bỏ nhiễu ảnh của một máy ảnh không có đèn Flash bằng phương pháp sử dụng một hình có flash không nhiễu. Vấn đề của mô hình nghiên cứu tương tự, các giá trị độ sắc thu được từ mạng neural được tạo có nhiễu (Tạo ra các hiện tượng không mong muốn) trong khi ảnh đen trắng của đầu vào không có nhiễu. Để đạt chất lượng ảnh được tô, áp dụng bộ lọc (joint bitateral filtering technique) để làm mịn và tinh chỉnh các giá trị độ sắc (giá trị được hiệu chỉnh trong quá trình huấn luyện mạng neural) với kết quả đầu ra mong muốn là ảnh trắng đen của tập mẫu. Hình 17 trình bày kết quả trước và sau khi tinh chỉnh độ sắc. Lưu ý rằng hầu hết các đối tượng có thể nhìn thấy được mà không mong muốn xuất hiện được loại bỏ thành công. Hình 18 và Hình 19 so sánh các phương pháp tô màu khác với phương pháp nghiên cứu "[6], [8], [9], [11]".



Hình 17. Kết quả tinh chỉnh độ sắc bằng bộ lọc. (a) Ảnh đen trắng lớp vào, (b) Ảnh sau khi tô màu, (c) Ảnh màu sau khi qua bộ lọc.



Hình 18. So sánh các phương pháp tô màu. (a) Ảnh lớp vào, (b) Ảnh của mô hình nghiên cứu, (c) Ảnh của Gupta, (d) Ảnh của Irony, (e) Ảnh của Welsh, (f) Ảnh của Charpiat, (g) Ảnh tham chiếu.



Hình 19. So sánh ảnh tô màu với ảnh thực tế. (a) Thành phố, (b) Đường phố, (c) Tòa nhà, (d) Biển, (e) Rừng.

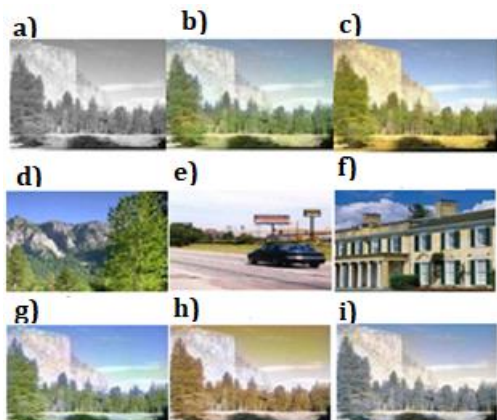
3.4. Đánh giá giữa các phương pháp trình bày và phương pháp sử dụng trong nghiên cứu.

Các thuật toán trước đó [6], [7], [8], [9], [11] thường được sử dụng một ảnh tham chiếu tương tự hoặc là tập dữ liệu hình ảnh tham chiếu tương tự, lấy màu từ các ảnh tham chiếu đó tô màu cho ảnh trắng cần tô [8] là phương pháp dựa trên ví dụ tiên tiến nhất vì nó vượt trội hơn các biên bản khác về hiệu suất và phạm vi ứng dụng. Tuy nhiên, hiệu suất của nó phụ thuộc rất nhiều vào hình ảnh tham chiếu được đưa như minh họa trong Hình 20. R. K. Gupta, A. Y.-S. Chia, D. Rajan, E.S.Ng, and H. Zhiyong có thể thu được kết quả tô màu rất tốt khi sử dụng hình ảnh tham chiếu chứa các đối tượng giống nhau với hình ảnh đen trắng trong tập dữ liệu mẫu. Với hình ảnh tham chiếu khác với hình ảnh được tô, hiệu suất của mô hình khá thấp như thể hiện trong Hình 20.

Để giảm thiểu sự phụ thuộc vào hình ảnh tham chiếu, bài báo sử dụng tập dữ liệu lớn. Trong quá trình tô màu sẽ tìm ảnh trong tập dữ liệu mẫu giống với ảnh đen trắng cần tô và chuyển màu từ ảnh mẫu sang ảnh đen trắng. Đây là lý do tại sao phương pháp này hiệu quả với nhiều hình ảnh đen trắng khác nhau.

Theo trực giác, một hình ảnh tham chiếu không thể bao gồm tất cả các tương ứng phù hợp cho các điểm ảnh trong ảnh đen trắng. Đây là lý do tại sao hiệu suất của [8] phụ thuộc rất nhiều vào hình ảnh tham chiếu phù hợp. Như thể hiện trong Hình 21, sử dụng một vài hình ảnh tham chiếu tương tự có thể cải thiện kết quả tô màu. Tuy nhiên, khi hình ảnh tham chiếu chứa nhiều đối tượng (vd: cửa ra vào, cửa sổ, tòa nhà, xe cộ...v.v), kết quả tô màu trở nên không tự nhiên, mặc dù một số ảnh cần tô giống như ảnh của tập tham chiếu. Điều này có

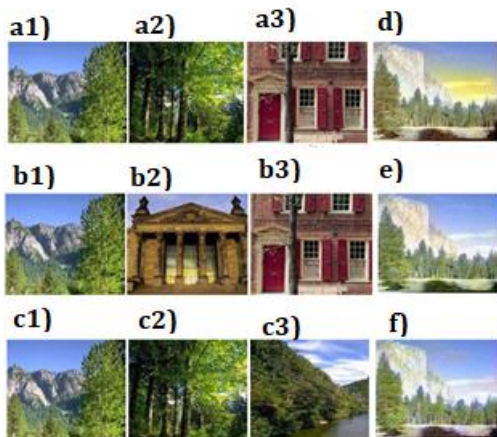
nghĩa là vẫn tồn tại nhiễu nhất định trong quá trình so sánh giữa ảnh cần tô và ảnh tham chiếu.



Hình 20. Sự phụ thuộc vào hình ảnh tham chiếu. (a) Ảnh lớp vào, (b) Ảnh được tô bằng mô hình, (c) Ảnh thật, (d) Ảnh tham chiếu 1, (e) Ảnh tham chiếu 2, (f) Ảnh tham chiếu 3, (g) Ảnh tham chiếu 4 + lọc thông thấp 19dB, (h) Ảnh tham chiếu 5 + lọc thông thấp 17dB, (i) Ảnh tham chiếu 6 + lọc thông thấp 18dB.

Ví dụ, đồng cỏ trong Hình 20(a) được khớp với bức tường trong Hình 21(e), và bầu trời khớp với toà nhà trong Hình 21(f).

Xây dựng kỹ thuật mạng học sâu với tập dữ liệu tham chiếu lớn. Mạng neural học sâu giúp kết hợp các đặc tính khác nhau của một điểm ảnh và tính toán giá trị độ sắc tương ứng. Ngoài ra, các phương pháp hiện đại rất chậm vì phải tìm những điểm ảnh giống nhau từ tập dữ liệu lớn. Khi so sánh mạng neural học sâu được thiết kế tính toán với tập dữ liệu lớn. Mặc dù tốc độ học của mạng chậm khi dữ liệu đầu vào lớn, nhưng mô hình tô màu cho ảnh trắng đen với kích thước màu là {256x256}.



Hình 21. Tập ảnh tham chiếu. (a1,2,3) – (c1,2,3) là các ảnh tham chiếu khác nhau và (d) – (f) là các kết quả tô màu tương ứng.

4. Kết quả thực nghiệm

Mạng neural tô màu được đào tạo trên 60.000 hình ảnh lấy từ cơ sở dữ liệu của trang web ImageNet Dataset [24]. Tập hình ảnh được chia thành 47 mục (vd: toà nhà, xe hơi, biển...v.v). Mạng neural có một lớp đầu vào, ba lớp ẩn và một lớp đầu ra. Theo thí nghiệm của nghiên cứu, sử dụng nhiều lớp ẩn hơn không cải thiện kết quả tô màu.

Lớp vào: 49 neural có đặc tính là so sánh lân cận điểm kích thước [7;7].

32 neural có đặc tính của mô hình DAISY [22] (4 locations, 8 orientations);

47 neural có đặc tính rút trích từ đặc tính của mỗi điểm ảnh đầu vào.

Tổng cộng có 128 neural lớp vào.

Bài báo sử dụng số neural lớp ẩn bằng một nửa số neural ở lớp vào, có 2 neural ở lớp ra (hai neural lớp ra mang đặc tính của độ chói).

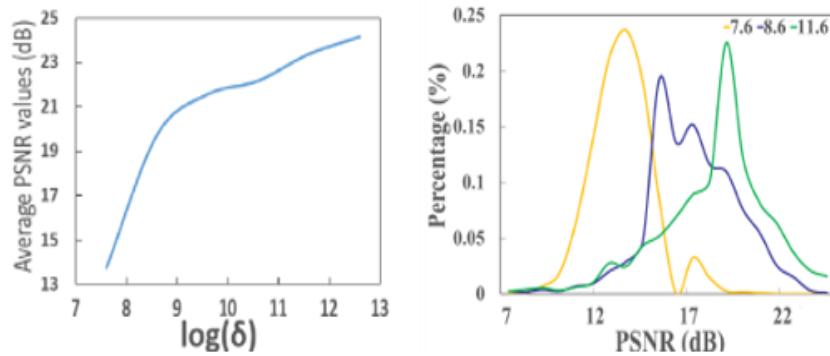
Hình 18 so sánh kết quả tô màu sử dụng mô hình nghiên cứu và các phương pháp khác [6], [8], [9], [11]. Hiệu suất của tô màu của phương pháp này rất cao khi hình ảnh tham chiếu được tối ưu hoá trong tập dữ liệu mẫu (vd: tập dữ liệu mẫu chứa hình ảnh giống như hình đen trắng cần tô). Như ở [6], [8], [9], [11]. Tuy nhiên hiệu suất của mô hình giảm khi ảnh cần tô không giống ảnh của tập dữ liệu mẫu.

Kiểm tra mô hình với kích thước mẫu khác nhau huấn luyện mạng neural học sâu (DNNs), sau đó sử dụng mô hình tô màu cho 1344 ảnh đen trắng.

Ψ : tập dữ liệu huấn luyện.

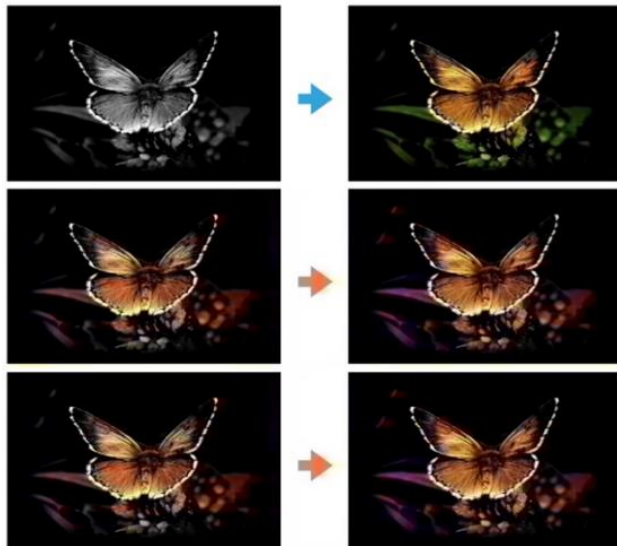
Kích thước mẫu: $\delta = \text{size}(\Psi)$

Hình 22 mô tả ảnh hưởng của kích thước mẫu đến kết quả tô màu, (a) và (b) trình bày giá trị trung bình và phân phối của PSRN bằng cách sử dụng mẫu khác nhau. Các đường cong màu vàng, xanh lơ, xanh lá trong (b) thể hiện phân phối PSRN khi $\log(\delta)$ được gán 7.6, 8.6, 11.6. Chúng ta thấy nếu tập dữ liệu mẫu càng lớn dẫn đến kết quả huấn luyện càng chính xác.



Hình 22. Ảnh hưởng của kích thước mẫu đến kết quả tô màu. (a) PSNR trung bình và (b) PSNR phân phối.

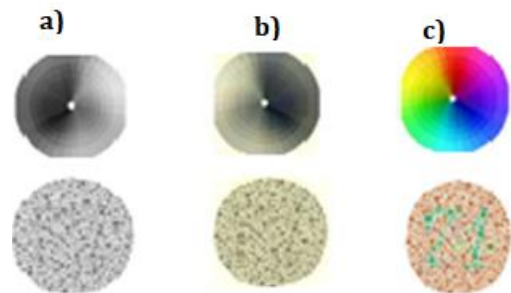
Kết quả tô màu thu được khi áp dụng mô hình nghiên cứu được mô tả trong Hình 23.



Hình 23. Kết quả tô màu của mô hình nghiên cứu.

Hạn chế: Phương pháp này vẫn còn hạn chế do hạn chế về cơ sở dữ liệu hình ảnh tham chiếu, nếu tổng hợp nhiều hình ảnh tham chiếu thì tập dữ liệu mẫu quá lớn như thể hiện trong Hình 1(b).

Điều này là không thể trong thực tế. Ví dụ mô hình được đào tạo trên hình ảnh thực tế do đó không phù hợp với hình ảnh tổng hợp. Cũng không thể phục hồi thông tin màu bị mất do chuyển đổi ảnh màu sang ảnh trắng đen. Hai trường hợp lỗi được trình bày trong Hình 24 là những hạn chế của mô hình nghiên cứu. Mô hình này không phù hợp đối với ảnh kết hợp nhiều cánh và màu và không khôi phục được thông tin bị mất trong quá trình chuyển đổi ảnh từ màu sang ảnh đen trắng.



Hình 24. Hạn chế của mô hình nghiên cứu. (a) Ảnh lớp vào, (b) Ảnh mô hình nghiên cứu, (c) Ảnh thật.

5. Kết luận

Mô hình nghiên cứu trình bày một phương pháp tô màu hoàn toàn tự động, sử dụng mạng neural học sâu, tập dữ liệu ảnh tham chiếu từ ImageNet Dataset để giải quyết bài toán tô màu tự động, giảm thiểu sự phụ thuộc vào người và kỹ năng tô màu.

Tuyên bố không xung đột lợi ích và cam kết bản quyền

Tác giả tuyên bố về sự không xuất hiện những xung đột tiềm ẩn từ nghiên cứu này, và cam kết bài báo chưa từng được công bố trước đây.

Chia sẻ dữ liệu theo yêu cầu

Dữ liệu sẽ không được cung cấp theo yêu cầu.

^{1st} Đình Văn Chuyên. *B&K ENGINEERING*

*Corresponding author: chuyen.dinh76@gmail.com

Tài liệu tham khảo

- [1] Y. C. Huang, Y. S. Tung, J. C. Chen, S. W. Wang, and J. L. Wu, "An adaptive edge detection based colorization algorithm and its applications," in *Proceedings of the 13th ACM International Conference on Multimedia, MM 2005*, 2005, pp.

- 351–354. doi: 10.1145/1101149.1101223.
- [2] A. Levin, D. Lischinski, and Y. Weiss, "Colorization using optimization," in *ACM SIGGRAPH 2004 Papers, SIGGRAPH 2004, 2004*, pp. 689–694. doi: 10.1145/1186562.1015780.
- [3] Q. Luan, F. Wen, D. Cohen-Or, L. Liang, Y.-Q. Xu, and H.-Y. Shum, "Natural image colorization," in *Proc. 18th Eurographics Symp. Rendering (EGSR '07), 2007*, pp. 309–320. doi: 10.2312/EGWR/EGSR07/309-320.
- [4] Y. Qu, T. T. Wong, and P. A. Heng, "Manga colorization," in *ACM Transactions on Graphics, Jul. 2006*, pp. 1214–1220. doi: 10.1145/1141911.1142017.
- [5] L. Yatziv and G. Sapiro, "Fast image and video colorization using chrominance blending," *IEEE Trans. Image Process.*, vol. 15, pp. 1120–1129, May 2006, doi: 10.1109/TIP.2005.864231.
- [6] G. Charpiat, M. Hofmann, and B. Schölkopf, "Automatic image colorization via multimodal predictions," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Verlag, 2008, pp. 126–139, doi: 10.1007/978-3-540-88690-7_10.
- [7] Y. Yang, H. Zhao, L. You, R. Tu, X. Wu, and X. Jin, "Semantic portrait color transfer with internet images," *Multimed. Tools Appl.*, vol. 76, pp. 523–541, Jan. 2017, doi: 10.1007/s11042-015-3063-x.
- [8] R. K. Gupta, A. Y. S. Chia, D. Rajan, E. S. Ng, and H. Zhiyong, "Image colorization using similar images," in *MM 2012 - Proceedings of the 20th ACM International Conference on Multimedia, 2012*, pp. 369–378, doi: 10.1145/2393347.2393402.
- [9] R. Irony, D. Cohen-Or, and D. Lischinski, "Colorization by example," in *EGSR '05: Proceedings of the Sixteenth Eurographics Conference on Rendering Techniques*, Konstanz, Germany, 2005, pp. 201–210. doi: 10.2312/EGWR/EGSR05/201-210.
- [10] Y. Xia, S. Qu, and S. Wan, "Scene guided colorization using neural networks," *Neural Comput. Appl.*, vol. 34, pp. 11083–11096, 2022, doi: 10.1007/s00521-018-3820-z.
- [11] T. Welsh, M. Ashikhmin, and K. Mueller, "Transferring color to greyscale images," in *SIGGRAPH '02: Proceedings of the 29th annual conference on Computer graphics and interactive techniques, 2002*, pp. 277–280, doi: 10.1145/566570.566576.
- [12] K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," *arXiv*, 2015, arXiv:1502.01852 [cs.CV], doi: 10.48550/arXiv.1502.01852.
- [13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.
- [14] L. Tian, M. Li, Y. Hao, J. Liu, G. Zhang, and Y. Q. Chen, "Robust 3-D Human Detection in Complex Environments With a Depth Camera," *IEEE Trans. Multimedia*, vol. 20, no. 9, pp. 2249–2261, Sept. 2018, doi: 10.1109/TMM.2018.2803526.
- [15] X. Zeng, W. Ouyang, and X. Wang, "Multi-stage Contextual Deep Learning for Pedestrian Detection," in *Proc. 2013 IEEE Int. Conf. Comput. Vision (ICCV)*, Sydney, NSW, Australia, 2013, pp. 121–128, doi: 10.1109/ICCV.2013.22.
- [16] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image Super-Resolution Using Very Deep Residual Channel Attention Networks," in *Computer Vision – ECCV 2018*, Munich, Germany, 2018, pp. 294–310, doi: 10.1007/978-3-030-01234-2_18.
- [17] Z. Yan, H. Zhang, B. Wang, S. Paris, and Y. Yu, "Automatic Photo Adjustment Using Deep Neural Networks," *arXiv*, 2014, arXiv:1412.7725 [cs.CV], doi: 10.48550/arXiv.1412.7725.
- [18] S. Zheng, A. Yuille, and Z. Tu, "Detecting object boundaries using low-, mid-, and high-level information," *Comput. Vision Image Understand.*, vol. 114, no. 10, pp. 1055–1067, Oct. 2010, doi: 10.1016/j.cviu.2010.07.004.
- [19] A. A. Efros and W. T. Freeman, "Image quilting for texture synthesis and transfer," in *SIGGRAPH '01: Proc. 28th Annu. Conf. Comput. Graph. Interactive Techn.*, 2001, pp. 341–346, doi: 10.1145/383259.383296.
- [20] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Comput. Graph. Appl.*, vol. 21, no. 5, pp. 34–41, July–Aug. 2001, doi: 10.1109/38.946629.

- [21] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Netw.*, vol. 2, no. 5, pp. 359–366, 1989, doi: 10.1016/0893-6080(89)90020-8.
- [22] E. Tola, V. Lepetit, and P. Fua, "A fast local descriptor for dense matching," in *Proc. 2008 IEEE Conf. Comput. Vision Pattern Recognit. (CVPR)*, Anchorage, AK, USA, 2008, doi: 10.1109/CVPR.2008.4587673.
- [23] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *arXiv*, 2014, arXiv:1411.4038 [cs.CV], doi: 10.48550/arXiv.1411.4038.
- [24] EE. S. L. Gastal and M. M. Oliveira, "Domain transform for edge-aware image and video processing," *ACM Trans. Graph.*, vol. 30, no. 4, art. no. 69, pp. 1–12, July 2011, doi: 10.1145/2010324.1964964.
- [25] G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, H. Hoppe, and K. Toyama, "Digital photography with flash and no-flash image pairs," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 664–672, Aug. 2004, doi: 10.1145/1015706.1015777.