

TÔ MÀU CHO ẢNH XÁM DÙNG MẠNG NƠ-RON TÍCH CHẬP

DEEP CONVOLUTIONAL NEURAL NETWORK BASED COLORIZATION

Đông Quốc Tranh, Hồ Phước Tiến, Trần Thị Minh Hạnh

Trường Đại học Bách khoa – Đại học Đà Nẵng; hptien@dut.udn.vn

Tóm tắt - Tô màu cho ảnh xám có ý nghĩa quan trọng trong việc phục hồi ảnh cũ và làm cho ảnh sinh động hơn. Gần đây với sự thành công của mạng nơ-ron nhiều lớp trong lĩnh vực thị giác máy tính, mạng nơ-ron tích chập nhiều lớp cũng mang lại những kết quả tích cực đối với bài toán tô màu. Tuy nhiên, các nghiên cứu trước đây thường tập trung vào một tập dữ liệu nào đó và ít quan tâm đến sự tác động của các bộ dữ liệu khác nhau đến kết quả thu được. Bài báo này sẽ xây dựng và đánh giá các mô hình tô màu khác nhau trên nhiều tập dữ liệu, dựa trên các tiêu chí về chất lượng ảnh. Kết quả thực nghiệm thể hiện sự hiệu quả của mô hình mạng residual trong bài toán tô màu. Đồng thời, bài báo còn cho thấy việc lựa chọn tập dữ liệu huấn luyện phù hợp sẽ góp phần xây dựng một mô hình tô màu hiệu quả.

Từ khóa - Tô màu; học sâu; mạng residual; chất lượng ảnh.

1. Giới thiệu bài toán tô màu

Tô màu cho ảnh xám có ý nghĩa quan trọng trong việc phục hồi màu sắc cho các phim, ảnh xưa, giúp tăng cường chất lượng hình ảnh, cũng như tăng độ sinh động. Tô màu là một bài toán không mới với những nhà mỹ thuật, nhưng lại là một bài toán mới trong lĩnh vực khoa học kỹ thuật. Một phương pháp tô màu được xem như một thợ sơn, giúp bức ảnh xám có màu sắc phù hợp và tự nhiên.

Trước đây, việc tô màu cho một vùng R nào đó trong một ảnh xám thường được thực hiện bằng cách tìm trong cơ sở dữ liệu vùng S tương tự với R; màu sắc của vùng S này sẽ được dùng để tô cho vùng R trong ảnh xám. Phương pháp này khá rõ ràng, tuy nhiên chất lượng thường không như mong đợi bởi hai yếu tố. Thứ nhất, cần phải có một phương pháp phân vùng hiệu quả, tuy nhiên đây lại là vấn đề chưa được giải quyết tốt. Thứ hai, việc chọn ra miền tương tự với miền đã cho cũng không phải dễ dàng, bởi rất khó để định lượng sự tương tự hay tương đồng, đồng thời có thể có nhiều màu phù hợp với một vùng cho trước.

Gần đây, sau nhiều thành công nổi bật trong các lĩnh vực nhận dạng hình ảnh, tiếng nói, hay xử lý ngôn ngữ [1, 2], phương pháp học sâu (Deep Learning) cũng mang đến cách tiếp cận mới cho bài toán tô màu ảnh xám. Cụ thể, sau khi được huấn luyện, một mô hình mạng nơ-ron nhiều lớp sẽ sinh ra một ảnh màu từ một ảnh xám ở đầu vào. Tức là, nếu so sánh với phương pháp cũ trên đây thì tất cả các bước như phân vùng, chọn màu được thực hiện một cách tự động bởi mạng nơ-ron.

Nhìn chung, với phương pháp học sâu, mô hình giải quyết bài toán tô màu cho ảnh xám là một mạng nơ-ron tích chập CNN (Convolutional Neural Network), với các lớp nối tiếp nhau như tích chập, kích hoạt phi tuyến, chuẩn hóa, kết hợp với việc lấy mẫu xuống (sub-sampling) và lấy mẫu lên (up-sampling) để đảm bảo ảnh đầu ra cùng kích thước với ảnh đầu vào. Ở [3], để tăng chất lượng của việc tô màu, tác giả đã kết hợp bài toán tô màu - sử dụng mạng CNN như trên - với bài toán phân loại ảnh, dù mục đích chính vẫn là tô màu. Việc phân loại ảnh, được thực hiện đồng

Abstract - Colorization is of great importance in restoring old gray pictures and making them more vivid. Thanks to the recent success of deep neural networks in various problems of computer vision, deep convolutional neural network has also been proposed for colorization and has brought about promising results. Yet previous works usually focus on some dataset and do not consider the influence of training data. This paper builds and evaluates many colorization models on various datasets according to criteria of image quality. The experimental results show interesting performance of residual networks in colorization. Moreover, an appropriate choice of training data may help to build an effective colorization model.

Key words - Colorization; convolutional neural network; residual network; image quality.

thời, sẽ giúp mô hình lấy được thông tin toàn cục về ảnh đầu vào, và từ đó giúp việc tô màu chính xác hơn. Ở [4], việc tô màu được thực hiện trên từng pixel, thông qua việc xử lý một vùng nhỏ xung quanh pixel đó (patch-based).

Một vấn đề quan trọng trong cách tiếp cận sử dụng phương pháp học sâu là sự lựa chọn hàm tổn hao. Nghiên cứu [3] sử dụng tiêu chí MSE (mean square error) kết hợp với hàm tổn hao đối với việc phân loại, ví dụ hàm cross-entropy. Các tác giả ở [5] lại cho rằng, MSE không phản ánh được sự phân bố màu của pixel, và đề xuất dùng hàm tổn hao như là sự khác biệt về histogram màu. Trong khi đó, bài báo [4] dùng cả MSE và histogram.

Một biến thể khác của các phương pháp trên đây là kết hợp chúng với những chỉ dẫn (hint) được cung cấp từ bên ngoài, khi đó ta có mô hình tương tác với người dùng. Bài báo ở [6] sử dụng những chỉ dẫn từ người dùng, như màu sắc và vị trí, để quá trình huấn luyện trở nên thuận lợi hơn và màu sắc thu được tự nhiên hơn.

Các mô hình học sâu trên đây đều nhận ở đầu vào là một ảnh xám và đưa ra hai lớp ứng với hai thành phần màu (thường là a và b trong hệ màu Lab); hai lớp này kết hợp với ảnh xám ban đầu để tạo ra ảnh màu hoàn chỉnh (Hình 1).

Nhìn chung, các mô hình trên cho kết quả tương đối tốt, nhưng phần lớn được huấn luyện với một bộ cơ sở dữ liệu, và chưa đánh giá mức độ ảnh hưởng của đặc trưng của dữ liệu đến kết quả tô màu.

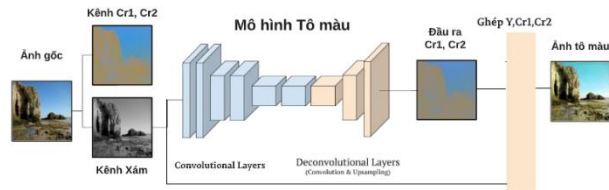
Bài báo này sẽ xây dựng và so sánh nhiều cấu trúc mạng CNN khác nhau đối với bài toán tô màu. Quá trình đánh giá được thực hiện với nhiều cơ sở dữ liệu khác nhau để phân tích ảnh hưởng của các bộ dữ liệu phổ biến đến kết quả thu được. Từ đó, chỉ ra cách thức để cải thiện chất lượng mô hình tô màu, giúp màu sắc trong ảnh thu được trở nên chân thực hơn.

2. Phương pháp tô màu cho ảnh xám sử dụng CNN

2.1. Mô hình

Mô hình CNN cho bài toán tô màu được thể hiện một

cách tổng quát trong Hình 1. Mô hình này gồm hai khối chính, đi từ trái sang phải, là Convolutional Layers (màu xanh nhạt) và Deconvolutional Layers (màu vàng nhạt), tương ứng với hai quá trình Encoder và Decoder. Khối Encoder cho phép biểu diễn những đặc trưng của ảnh xám đầu vào; khối Decoder kết hợp những đặc trưng này để tìm ra màu thích hợp. Chú ý, khối Encoder làm cho kích thước ảnh giảm dần, trong khi khối Decoder làm tăng kích thước và phục hồi kích thước của ảnh ban đầu. Mục tiêu chính là tìm ra hai kênh màu sắc Cr1 và Cr2 (chrominance channels) ứng với ảnh xám ở đầu vào. Có nhiều cách biểu diễn màu sắc, thông qua các hệ màu khác nhau như HSV, YCbCr, Luv, Lab. Trong đó, hai kênh màu lần lượt là H và S, Cb và Cr, u và v, a và b. Các hệ màu này ít nhiều tương tự nhau, và có thể chuyển đổi qua lại, cũng như với hệ RGB. Bài báo này sẽ sử dụng hệ màu Lab, giống như trong các nghiên cứu trước đây. Cụ thể, đầu vào là ảnh xám, thể hiện cường độ sáng L; đầu ra sẽ là hai kênh màu a và b. Hai kênh màu này sẽ kết hợp với cường độ sáng L để tạo thành ảnh màu hoàn chỉnh.

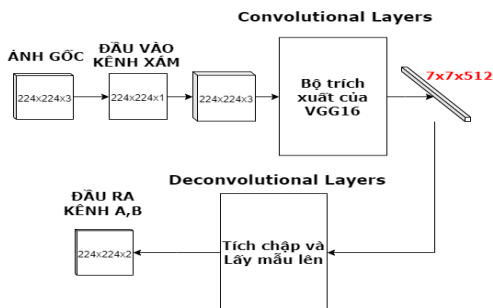


Hình 1. Mô hình mạng học sâu tô màu cho ảnh xám

Không nhằm mục tiêu đánh giá ảnh hưởng của hàm kích hoạt, bài báo này sử dụng hàm kích hoạt ReLU dựa trên các ưu điểm như tốc độ hội tụ nhanh và tính toán đơn giản. Hàm kích hoạt thường sẽ được thực hiện sau phép tính tích chập và chuẩn hóa. Đối với mô hình VGG16, trong phần Convolutional Layers, hàm kích hoạt thực hiện ngay sau phép tính tích chập và không thực hiện chuẩn hóa, nhưng khối Deconvolutional thì có dùng hàm kích hoạt và chuẩn hóa như các mô hình khác. Do đầu vào có phân bố khác nhau nên ảnh hưởng đến tốc độ hội tụ của giải thuật huấn luyện; việc chuẩn hóa giúp giải thuật huấn luyện hội tụ nhanh hơn. Phép chuẩn hóa được thực hiện trên các tập con mini-batch (64 ảnh), dựa vào giá trị trung bình và phương sai của mỗi tập con.

Phần tiếp theo sẽ trình bày các cấu trúc cụ thể trong các khối Convolutional Layers (Encoder) và Deconvolutional Layers (Decoder). Những cấu trúc khác nhau này sẽ dẫn đến các mô hình khác nhau mà sẽ được dùng để phân tích và đánh giá trong mục 3.

2.2. Mô hình kết hợp VGG16



Hình 2. Mô hình kết hợp VGG16

Hình 2 mô tả kiến trúc của mô hình tô màu sử dụng mạng tích chập VGG16. VGG [7] đã thể hiện nhiều ưu điểm trong

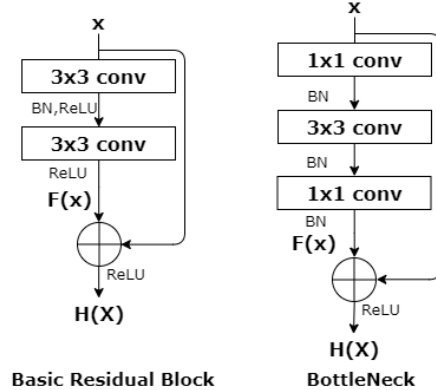
bài toán nhận dạng hình ảnh, và đặc biệt có khả năng trích xuất các thuộc tính của ảnh. Chính điều này khiến cho mạng VGG được sử dụng trong nhiều bài toán khác nhau, ngoài vấn đề nhận dạng. Trong Hình 2, khối Convolutional Layers là khối trích xuất của mô hình VGG16 đã được công bố [7], bao gồm các lớp tích chập liên tiếp nhau, dùng hàm kích hoạt ReLU, và không sử dụng chuẩn hóa. Kích thước đầu ra giảm 2⁵ lần so với kích thước đầu vào. Với đầu vào có kích thước 224x224, đầu ra của khối này có kích thước là 7x7 và bao gồm 512 kênh, tức có dạng (7,7,512).

Khối Deconvolutional Layers gồm các lớp tích chập và lấy mẫu lên (upsampling) để trả về kích thước ban đầu là 224x224, với hai kênh a và b, do đó có dạng (224,224,2).

2.3. Mô hình kết hợp ResNet18

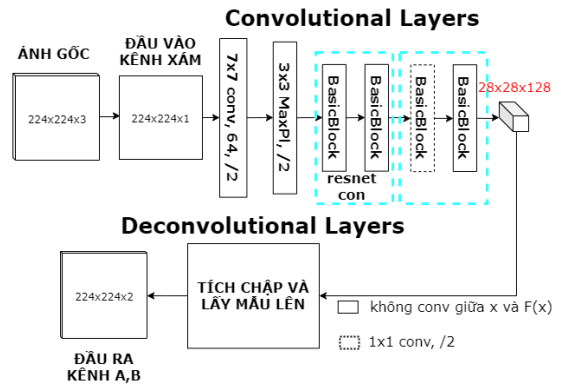
ResNet (residual network) được đề xuất để cải thiện chất lượng bài toán nhận dạng hình ảnh [8]. ResNet18 có cấu trúc gồm nhiều basic residual block (Hình 3), mỗi block thể hiện phép biến đổi thông qua phương trình sau:

$$H(x) = F(x) + x. \tag{1}$$



Hình 3. Cấu trúc Basic Residual Block và Bottleneck

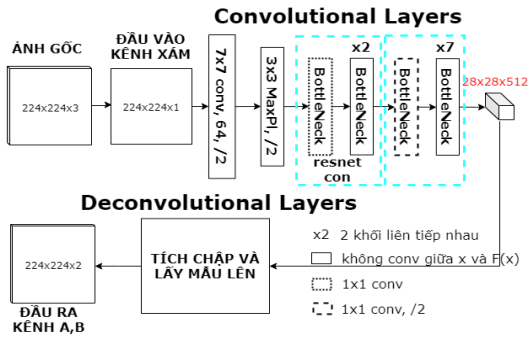
Mỗi block thực hiện các phép tích chập, kích hoạt phi tuyến ReLU và chuẩn hóa BN. ResNet giúp giải quyết vấn đề triệt tiêu gradient (gradient vanishing), và từ đó cho phép huấn luyện mạng tích chập với rất nhiều lớp.



Hình 4. Mô hình kết hợp ResNet18

Hình 4 biểu diễn kiến trúc của mô hình tô màu sử dụng mạng tích chập ResNet18. Khối Convolutional Layers chuyển ảnh xám kích thước 224x224 thành ảnh có kích thước (28,28,512). Khối Deconvolutional Layers chứa các lớp tích chập và upsampling, để đảm bảo đầu ra có kích thước (224,244,2). Trong khối này, lớp chuẩn hóa BN cũng được sử dụng sau mỗi lớp tích chập.

2.4. Mô hình kết hợp ResNet152



Hình 5. Mô hình kết hợp ResNet152

Tương tự như mô hình với ResNet18, kiến trúc của mô hình tô màu sử dụng mạng tích chập ResNet512 được thể hiện trong Hình 5. Khác biệt cơ bản giữa ResNet152 và ResNet18 là số lượng tham số của ResNet152 lớn hơn rất nhiều, và ResNet152 sử dụng các khối BottleNeck (Hình 3). BottleNeck gồm ba lớp tích chập thay vì hai lớp như trong basic residual block của ResNet18.

2.5. Huấn luyện

Với mỗi mô hình tô màu đã trình bày từ Mục 2.2 đến 2.4, ta cần thực hiện quá trình huấn luyện để tìm ra các trọng số của chúng. Ở đây, cũng như trong hầu hết các phương pháp tô màu khác, quá trình học có giám sát

(supervised learning) được sử dụng. Khi đó, hàm tổn hao được định nghĩa như sau:

$$L = \frac{1}{N} \sum_{i=1}^N (\|\hat{a}_i - a_i\|^2 + \|\hat{b}_i - b_i\|^2), \tag{2}$$

với: $\|\hat{a}_i - a_i\|^2 = \sum_{k=1}^K \sum_{l=1}^L |\hat{a}_i(k, l) - a_i(k, l)|^2,$ (3)

$$\|\hat{b}_i - b_i\|^2 = \sum_{k=1}^K \sum_{l=1}^L |\hat{b}_i(k, l) - b_i(k, l)|^2. \tag{4}$$

Trong đó, N là số lượng ảnh. \hat{a}_i, \hat{b}_i là các kênh màu của ảnh thứ i được ước lượng từ mô hình. a_i, b_i là kênh màu tương ứng đã biết trước (ground truth). K, L lần lượt là chiều cao và chiều rộng của ảnh gốc.

Việc cập nhật các trọng số của mạng được thực hiện với giải thuật lan truyền ngược (back-propagation), kết hợp với phương pháp Adam [9], kích thước của batch là 64. Quá trình huấn luyện được thực hiện cho đến khi kết quả thu được ổn định, trên nền tảng PyTorch và môi trường Google Colab. Thời gian huấn luyện có thể lên đến một tuần.

3. Thử nghiệm và kết quả

3.1. Dữ liệu

Quá trình thử nghiệm sử dụng bốn bộ dữ liệu để đánh giá các mô hình tô màu. Các bộ dữ liệu này (trừ tập 3) đều được chia theo tỉ lệ 9:1 (huấn luyện : kiểm tra). Cụ thể như sau:

Ảnh xám	VGG16	ResNet18	ResNet152	Ảnh gốc

Hình 6. Một số kết quả tô màu của VGG16, ResNet18, ResNet152 với Tập 1

Tập 1 (testSet hay testSetPlaces205_resized) có tổng số 41000 ảnh, trong đó 36 901 ảnh cho huấn luyện và 4 099 ảnh cho kiểm tra. Đây là bộ kiểm tra của bộ dữ liệu Places205 [10], với 205 loại ảnh phong cảnh như công viên, biển, núi rừng, nhà thờ, cầu.

Tập 2 (Flowers) [11] có kích thước tương đối nhỏ gồm 4326 ảnh, sau khi loại bỏ một số ảnh chất lượng kém, tập dữ liệu gồm 4319 ảnh về hoa được sử dụng, trong đó có 3888 ảnh huấn luyện và 431 ảnh kiểm tra. Bộ dữ liệu này chứa 5 loại hoa: hoa hướng dương, hoa cúc, hoa bồ công anh, hoa hồng và hoa tulip.

Tập 3 (Object) [12] chứa 15571 ảnh, trong đó có 10552 ảnh huấn luyện và 5019 ảnh kiểm tra. Các đối tượng trong bộ dữ liệu này bao gồm bàn, ghế, hoa, xích đu, con vật, người... với phần nền có nội dung phức tạp. Bộ dữ liệu này không được chia theo tỷ lệ 9:1, mà giữ nguyên như ở [12] để đảm bảo các đặc trưng của nó.

Tập 4 (Blue) tập trung vào các ảnh có màu xanh dương (blue). Đây là bộ dữ liệu được tác giả trích ra từ bộ dữ liệu Places205 [10] nhằm phân tích sự ảnh hưởng của đặc trưng dữ liệu đến việc khôi phục màu sắc. Tập 4 có tổng cộng 44156 ảnh, với 39740 ảnh huấn luyện và 4 416 ảnh kiểm tra. Bộ dữ liệu này chứa các ảnh liên quan đến cảnh như công viên, khán phòng, ảnh nghệ thuật, con đường, lâu đài.

3.2. Tiêu chí đánh giá

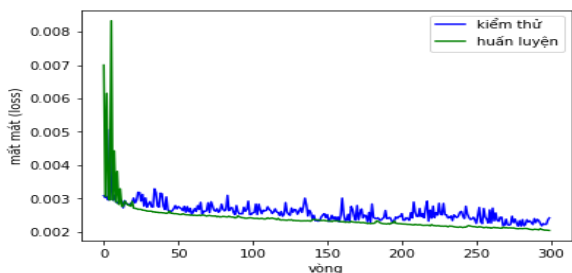
Tiêu chí khách quan: trong bài báo này, PSNR được dùng để đánh giá chất lượng của ảnh màu khôi phục:

$$\text{PSNR} = 10 \log_{10} \left(\frac{\text{MAX}_I^2}{\text{MSE}} \right). \quad (5)$$

Trong đó, MAX_I là giá trị lớn nhất trong ảnh. Ở đây, $\text{MAX}_I = 1$ vì ảnh đã chuẩn hóa trước khi tính toán. MSE là sai số bình phương trung bình giữa ảnh màu gốc và ảnh màu được khôi phục, tương tự như phương trình (2). PSNR càng lớn, chất lượng ảnh khôi phục càng tăng. Trong bài báo này, PSNR liên quan trực tiếp đến hàm mục tiêu (MSE), dùng để tối ưu hóa mô hình.

Tuy nhiên, tiêu chí khách quan không thể phản ánh hết chất lượng ảnh màu. Trên thực tế, ta thường kết hợp với đánh giá chủ quan, tức quan sát bằng mắt người để đánh giá chất lượng ảnh.

3.3. Kết quả



Hình 7. Đồ thị mất mát khi huấn luyện ResNet152 sử dụng bộ dữ liệu testSet (Tập 1)

Mục này sẽ đánh giá chất lượng tô màu của các mô hình được giới thiệu từ Mục 2.2 đến Mục 2.4. Để thuận tiện cho việc trình bày, các mô hình này được đặt tên lần lượt là VGG16, ResNet18, và ResNet152, tương ứng với các kiến trúc của các mạng tích chập được sử dụng. Các kết quả

được trình bày sau khi mô hình đã ổn định. Hình 7 thể hiện kết quả của hàm tổn hao (mất mát) của mô hình ResNet152 đối với tập huấn luyện và kiểm tra của bộ dữ liệu 1 (testSet). Theo hình này, kết quả của hàm tổn hao đối với tập kiểm tra gần như không đổi sau khoảng 300 lần lặp.

3.3.1. So sánh ba mô hình tô màu

Trong mục này, ta so sánh chất lượng tô màu của ba mô hình nêu trên. Bộ dữ liệu testSet (Tập 1) sẽ được sử dụng cho mục đích này, đây cũng là bộ dữ liệu có số lượng ảnh tương đối lớn.

Bảng 1. PSNR của ba mô hình ResNet18, ResNet152, VGG16 sau 301 vòng lặp huấn luyện

	ResNet18	ResNet152	VGG16
Số lượng trọng số	954 722	6 839 554	19 660 418
PSNR (dB)	25,84	26,40	25,13

Theo kết quả ở Bảng 1, mô hình VGG16 phức tạp nhất, nhưng kết quả lại thấp nhất. Mô hình sử dụng Residual Network có vẻ cho kết quả tốt hơn. Ngoài ra, mô hình ResNet152 có độ phức tạp lớn hơn và cho kết quả PSNR tốt hơn ResNet18.

Nhận xét trên cũng phù hợp với những kết quả quan sát ở Hình 6: mô hình ResNet152 cho kết quả với màu sắc tự nhiên nhất.

3.3.2. Ảnh hưởng của các bộ dữ liệu khác nhau

Trong mục này, ta đánh giá ảnh hưởng của các bộ dữ liệu khác nhau đối với cùng một mô hình tô màu. Ở đây, ta chọn mô hình ResNet18 bởi nó thể hiện sự cân bằng tốt nhất giữa chất lượng tô màu và độ phức tạp (ảnh hưởng đến thời gian huấn luyện). Kết quả về PSNR ứng với ba tập dữ liệu (đánh giá trên tập kiểm tra) được cho ở Bảng 2.

Bảng 2. PSNR của mô hình ResNet18 đối với ba bộ dữ liệu

	Tập 1 (testSet)	Tập 2 (Flowers)	Tập 3 (Object)
PSNR (dB)	25,84	22,19	24,90

Từ Bảng 2, ta thấy kết quả tô màu tốt nhất ứng với bộ dữ liệu 1, đây là tập chứa nhiều ảnh cảnh vật. Bộ dữ liệu về các đối tượng (tập 3) cho kết quả thấp hơn một chút. Ta nhận thấy rằng, việc tô màu đối với các ảnh phong cảnh có hiệu quả cao hơn so với các ảnh chứa các đối tượng, vốn thể hiện sự đa dạng và tương đối phức tạp hơn. Bộ dữ liệu về hoa (Tập 2) cho kết quả thấp nhất. Nguyên nhân có thể đến từ sự đa dạng hình dáng và màu sắc của hoa, nhưng cũng có thể do số lượng của tập 2 chưa đủ lớn, nên mô hình tô màu chưa học được một cách hiệu quả.

3.3.3. Ảnh hưởng màu sắc

Từ kết quả của những thí nghiệm trên ta thấy, bộ dữ liệu huấn luyện ảnh hưởng đến kết quả tô màu. Ví dụ, bộ dữ liệu 1 (testSet) cho kết quả tốt với màu xanh lục (đây là màu xuất hiện nhiều trong ảnh phong cảnh), nhưng có vẻ không thật sự tốt với màu xanh dương. Để cải thiện vấn đề này, ta xây dựng bộ dữ liệu mới (tập 4). Bộ dữ liệu này được trích từ bộ huấn luyện của Places205 [10], nhưng tập trung vào các ảnh có màu xanh dương.

Ta sẽ huấn luyện mô hình ResNet152 trên hai tập dữ liệu này. Cụ thể, một mô hình ResNet152 được huấn luyện với 36901 ảnh của tập 1 và một mô hình được huấn luyện với 39740 ảnh của tập 4. Kết quả sau khi huấn luyện, ta có hai phiên bản của ResNet152, ứng với hai bộ dữ liệu huấn luyện khác nhau. Hai phiên bản của ResNet152 này sẽ

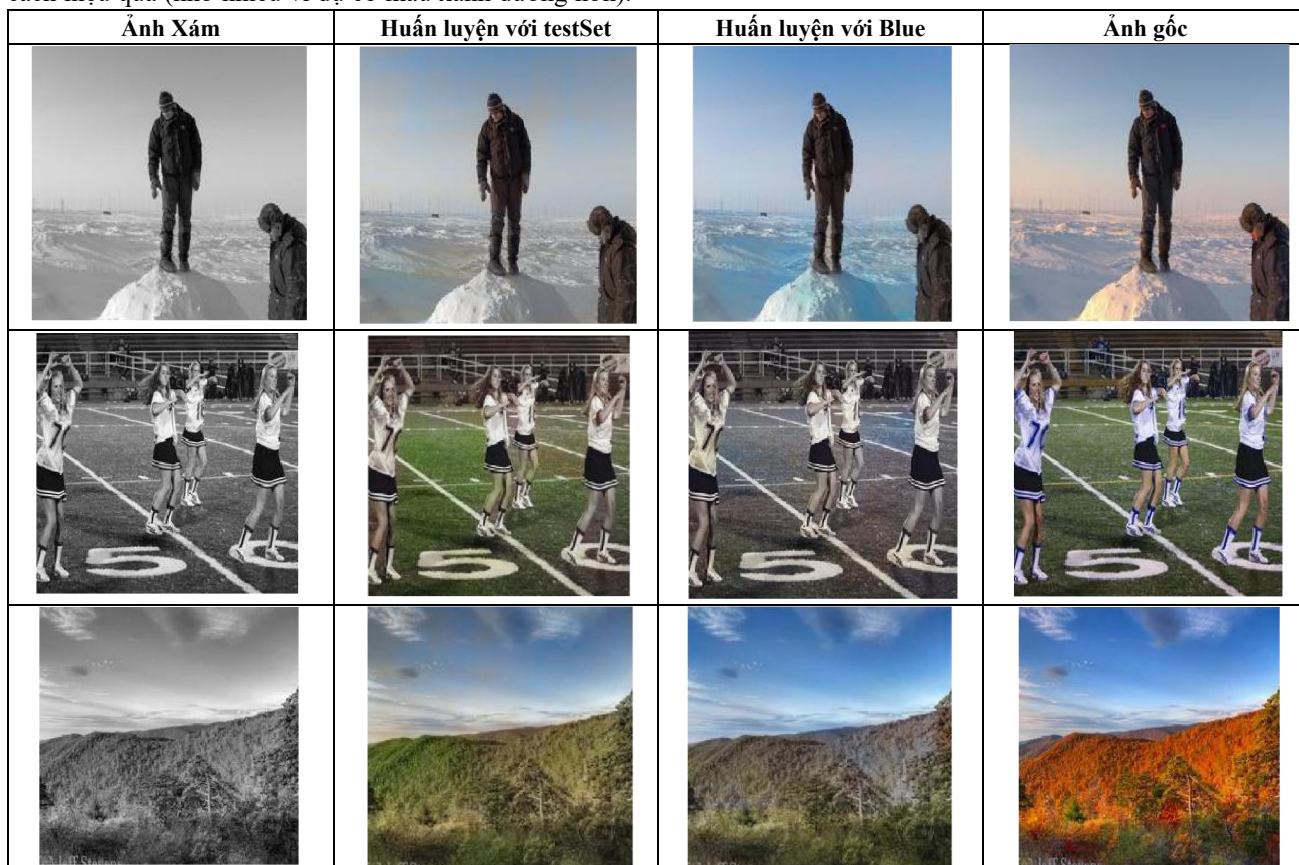
được đánh giá trên tập kiểm tra gồm 8515 ảnh, gồm 4099 ảnh kiểm tra của tập 1 (bộ testSet) và 4416 ảnh kiểm tra của tập 4. Kết quả được tổng kết trong Bảng 3.

Từ Bảng 3, ta thấy mô hình ResNet152 khi huấn luyện bộ dữ liệu Blue cho kết quả tốt hơn. Nguyên nhân là trong trường hợp này, mô hình đã học được màu xanh dương một cách hiệu quả (nhờ nhiều ví dụ có màu xanh dương hơn).

Điều này được kiểm nghiệm trên Hình 8. Cụ thể, bầu trời xanh dương được phục hồi tương đối tốt và tự nhiên.

Bảng 3. PSNR của mô hình ResNet152 sử dụng bộ dữ liệu 1 (testSet) và 4 (Blue) sau 301 vòng lặp huấn luyện

	Tập 1 (testSet)	Tập 4 (Blue)
PSNR (dB)	25,28	26,09



Hình 8. Kết quả của mô hình ResNet152 được huấn luyện với bộ dữ liệu testSet và Blue

4. Kết luận

Bài báo này đã đánh giá việc sử dụng các mô hình mạng nơ-ron tích chập nhiều lớp khác nhau để giải quyết bài toán tô màu cho ảnh xám. Kết quả thực nghiệm cho thấy mô hình Residual Network có khả năng tô màu tương đối hiệu quả, đặc biệt khi sử dụng một mô hình đủ mạnh và với số lượng dữ liệu huấn luyện đủ lớn. Bên cạnh đó, đặc trưng của dữ liệu huấn luyện ảnh hưởng nhiều đến kết quả thu được. Cụ thể, dữ liệu huấn luyện cần có sự phân bố đều về màu sắc. Trong tương lai, ta có thể kết hợp với mô hình tập trung thị giác [13] để việc tô màu có thể hiệu quả hơn và phù hợp với cảm nhận thị giác của con người, cũng như khai thác mối liên hệ giữa màu sắc và mạng nơ-ron nhiều lớp [14].

TÀI LIỆU THAM KHẢO

- [1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning", *Nature*, vol. 521, pp. 436-444, 2015.
- [2] T., Ho-Phuoc, "CIFAR10 to compare visual recognition performance between deep neural networks and humans", *arXiv:1811.07270*, 2018.
- [3] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Let there be Color!: Joint End-to-end Learning of Global and Local Image Priors for Automatic Colorization with Simultaneous Classification", *SIGGRAPH*, 35(4), 2016.
- [4] G. Larsson, M. Maire, and G. Shakhnarovich, "Learning Representations for Automatic Colorization", *ECCV*, 2016.
- [5] R. Zhang, P. Isola, and A. A. Efros, "Colorful image colorization", *ECCV*, 2016.
- [6] R. Zhang, J. Y. Zhu, P. Isola, X. Geng, A. S. Lin, T. Yu, and A. A. Efros, "Real-Time User-Guided Image Colorization with Learned Deep Priors", *SIGGRAPH*, 36(4), 2017.
- [7] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", *ICLR*, 2015.
- [8] K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition", *CVPR*, 2016.
- [9] Diederik P. Kingma and Jimmy Lei Ba, *Adam: A method for stochastic optimization*. 2014. arXiv:1412.6980v9.
- [10] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, "Learning Deep Features for Scene Recognition using Places Database." *Advances in Neural Information Processing Systems 27 (NIPS)*, 2014
- [11] Alexander Mamaev, *Flowers Recognition*, version 2, 2018, <https://www.kaggle.com/alexamamaev/flowers-recognition>
- [12] L. Wang, H. Lu, Y. Wang, M. Feng, D. Wang, B. Yin, and X. Ruan, "Learning to Detect Salient Objects with Image-level Supervision", *CVPR*, 2017.
- [13] T., Ho-Phuoc, "Développement et mise en oeuvre de modèles d'attention visuelle", PhD thesis, Université de Grenoble, 2010.
- [14] V. Buhrmester, D. Münch, D. Bulatov, and M. Arens M., "Evaluating the Impact of Color Information in Deep Neural Networks", *Pattern Recognition and Image Analysis*, LNCS, 2019.

(BBT nhận bài: 10/3/2020, hoàn tất thủ tục phản biện: 15/5/2020)