ISSN: 1859-2171
e-ISSN: 2615-9562
TNU Journal of Science and Technology 208(15): 19 - 26

HOÀN THIỆN CÁC VÙNG PHÁ HỦY HÌNH DẠNG BẤT KỲ TRONG ẢNH SỬ DỤNG KIẾN TRÚC MẠNG THẶNG DƯ VÀ NHÂN CHẬP TỪNG PHẦN

Lê Đình Nghiệp[1], Phạm Việt Bình[2], Đỗ Năng Toàn[3], Hoàng Văn Thi[4]
[1]Trường Đại học Hồng Đức,
[2]Trường Đại học Công nghệ thông tin & Truyền thông – ĐH Thái Nguyên,
[3]Viện Công nghệ thông tin – ĐH Quốc gia Hà Nội, [4]Sở giáo dục và Đào tạo Thanh Hóa

TÓM TẮT

Ngày nay, các giải thuật dựa trên học sâu cho bài toán hoàn thiện ảnh (image inpainting) đã thu được kết quả tốt khi xử lý các vùng mất mát thông tin có hình dạng vuông hoặc các hình phổ dụng. Tuy nhiên, vẫn thất bại trong việc tạo ra các kết cấu hợp lý bên trong vùng bị phá hủy do thiếu các thông tin xung quanh. Trong nghiên cứu này, bắt nguồn từ giải thuật học thặng dư được dùng để dự đoán các thông tin bị mất trong vùng bị phá hủy, thuận lợi cho tích hợp các đặc trưng và dự đoán kết cấu, chúng tôi đề xuất mạng nhân chập từng phần thặng dư cải tiến dựa trên kiến trúc mã hóa và giải mã U-net để lấp đầy vùng bị phá hủy bảo toàn kết cấu không chỉ với các hình dạng phổ dụng mà còn cho các hình dạng bất kỳ. Các thí nghiệm dựa trên định tính và định lượng đều cho thấy mô hình đề xuất có thể giải quyết các vùng bị phá hủy có hình dạng bất kỳ và đạt hiệu suất thực thi tốt hơn các phương pháp inpainting trước đó.

Từ khóa: *inpainting ảnh; mặt nạ không phổ dụng; mặt nạ bất kỳ; mạng thặng dư; thị giác máy tính; nhân chập từng phần;*

Ngày nhận bài: 11/9/2019; Ngày hoàn thiện: 18/9/2019; Ngày đăng: 03/10/2019

IMAGE INPAINTING FOR ARBITRARY HOLES USING CUSTUMIZED RESIDUAL BLOCK ARCHITECTURE WITH PARTIAL CONVOLUTIONS

Le Dinh Nghiep[1], Pham Viet Binh[2], Do Nang Toan[3], Hoang Van Thi[4]
[1]Hong Duc University,
[2]University of Information and Communication Technology - TNU,
[3]Institute of Information Technology - VNU, [4]Thanh Hoa Department of Education and Training

ABSTRACT

Recently, learning-based algorithms for image inpainting achieve remarkable progress dealing with squared or regular holes. However, they still fail to generate plausible textures inside damaged area because there lacks surrounding information. In this paper, motivated by the residual learning algorithm which aims to learn the missing information in corrupted regions, thus facilitating feature integration and texture prediction we propose Residual Partial Convolution network (RBPConv) based on *e*ncoder and decoder U-net architecture to maintain texture while filling not only regular regions but also random holes. Both qualitative and quantitative experimental demonstrate that our model can deal with the corrupted regions of arbitrary shapes and performs favorably against previous state-of-the-art methods.

Keywords: *generative image inpainting; irregular mask; residual network; computer vision; arbitrary mask; partial convolution.*

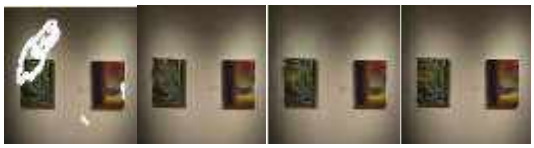
Received: 11/9/2019; Revised: 18/9/2019; Published: 03/10/2019

* Corresponding author. *Email: ledinhnghiep@hdu.edu.vn*

1. Giới thiệu

Inpainting ảnh là một tiến trình lấp đầy hay khôi phục lại các vùng bị mất mát thông tin hay vùng bị phá hủy (gọi là vùng đích) trong ảnh bằng cách tổng hợp từ các vùng không bị phá hủy hay các ảnh gốc khác (gọi là vùng nguồn). Inpainting được sử dụng trong rất nhiều ứng dụng thực tế như: loại bỏ các đối tượng không mong muốn ra khỏi ảnh, khôi phục các vùng ảnh bị phá hủy, hoàn thiện các vùng bị che khuất, khử nhiễu. Mặc dù đã được nghiên cứu trong nhiều thập niên qua, inpainting ảnh vẫn là một bài toán mở và khó trong lĩnh vực đồ họa và thị giác máy tính do tính mơ hồ không rõ ràng và độ phức tạp của ảnh tự nhiên. Nói chung, kết quả ảnh inpainting phải thỏa mãn yêu cầu về bảo toàn cấu trúc ngữ nghĩa tổng thể và kết cấu chi tiết.

Các phương pháp inpainting cổ điển dựa trên khuyếch tán [1] [2] hay lấy mẫu [3] [4] [5] [6] đều sử dụng ý tưởng thẩm thấu các thông tin về cấu trúc và kết cấu trong từ các vùng nguồn vào trong các vùng đích. Với các cách tiếp cận này tiến trình inpainting ảnh được thực hiện theo từng bước từ rìa vùng đích vào trong. Vì vậy, kết quả của bước sau phục thuộc rất nhiều vào độ chính xác của bước trước đó, lỗi thẩm thấu sẽ xuất hiện nếu như việc khôi phục thất bại ở một bước nào đó thì kết cấu tổng thể cũng như chi tiết sẽ bị sai lệch (hình 1b).



(a) (b) (c) (d)

Hình 1. *Một số kỹ thuật inpainting. (a) ảnh với vùng cần hoàn thiện. (b) Ảnh hoàn thiện dựa trên lấy mẫu PathMach [6]. (c) Ảnh hoàn thiện dựa trên mạng Context Encoder [7]. (d) Ảnh gốc*

Bên cạnh đó quá trình tìm kiếm lân cận gần nhất có chi phí thời gian lớn. Cách tiếp cận này hiệu quả khi có thể tìm thấy các mẫu ảnh với đầy đủ sự tương quan về mặt trực quan tuy nhiên sẽ thất bại nếu không tìm thấy mẫu ảnh tương tự trong cơ sở dữ liệu. Hơn nữa, các phương pháp này có thể cần đến cơ sở dữ liệu mẫu bên ngoài làm giảm phạm vi của ảnh cần hoàn thiện.

Ngược lại với các phương pháp truyền thống dựa trên lấy mẫu sử dụng các đặc trưng xung quanh vùng trống trong ảnh hoặc từ tập mẫu chọn trước, các giải thuật dựa trên mạng nhân chập học sâu (Deep Convolution Neural Network (DCNN)) cũng đã được đề xuất để học các đặc trưng dùng cho dự đoán các phần mất mát thông tin dựa trên tập dữ liệu huấn luyện. Lợi ích từ dữ liệu huấn luyện lớn, các phương pháp dựa trên DCNN đưa kết quả inpainting với ngữ nghĩa hợp lý hơn. Tuy nhiên, một số phương pháp dựa trên DCNN thường hoàn thiện các vùng mất mát thông tin bằng cách thẩm thấu các đặc trưng nhân chập của các vùng xung quanh thông qua một tầng kết nối đầy đủ, làm cho kết quả inpainting đôi khi thiếu các chi tiết kết cấu tốt và có vết mờ (hình 1c).

Một giới hạn khác của các kỹ thuật inpainting trước đây là chỉ tập trung trên các vùng trống hình chữ nhật và giả thiết nó thường được đặt ở xung quanh trung tâm của ảnh [7] [8] [9]. Những giới hạn này có thể dẫn đến tình trạng quá khớp trên các vùng trống hình chữ nhật và giới hạn ứng dụng của các mô hình này trong thực tế. Một vài nghiên cứu [10] [11] gần đây đã mở rộng hình dạng của mặt nạ vùng trống với các khuôn dạng phổ dụng như hình chữ nhật, hình thoi, hình elip… và đặt chúng ở các vị trí ngẫu nhiên trong ảnh. Tuy nhiên nghiên cứu cũng chưa thu được kết quả tốt trên tập mặt nạ này. Dựa trên tập mặt nạ với hình dáng và đường kẻ đa dạng có được từ nghiên cứu [12], kết hợp với phép nhân chập từng phần nghiên cứu [13] cho kết quả inpainting tương đối tốt trên tập mặt nạ không phổ dụng này.

Để gia tăng tốc độ, hiệu suất thực thi cũng như kết quả inpainting, chúng tôi đề xuất một mô hình kiến trúc mạng DCNN sử dụng các khối residual kết hợp với nhân chập từng phần được giới thiệu lần đầu trong [13] nhằm

gia tăng khả năng trừu tượng hóa đặc trưng tạo ra các ảnh hoàn tiện tốt hơn. Các thực nghiệm được thực hiện trên các tập dữ liệu chuẩn cho thấy phương pháp đề xuất thu được hiệu suất cao và thời gian xử lý nhanh, bền vững với các thể loại mặt nạ khác nhau với cả hình dạng thông dụng, không thông dụng, hoặc mặt nạ bất kỳ khi so sánh với các phương pháp trước đó. Các phần tiếp theo của bài viết này được tổ chức như sau: trước hết các nghiên cứu liên quan được trình bày trong phần 2; kiến trúc mô hình đề xuất với các khối Residual cải tiến kết hợp với nhân chập từng phần được giới thiệu trong phần 3; môi trường thực nghiệm và các kết quả được trình bày trong phần 4; cuối cùng, kết luận được đưa ra trong phần 5.

2. Các nghiên cứu liên quan

Các cách tiếp cận không dựa trên mạng học sâu thường được chia thành hai loại: phương pháp dựa trên khuyếch tán [14] [1] [2] và phương pháp dựa trên lấy mẫu [6] [15] [16] [17] [18]. Các phương pháp dựa trên khuyếch tán thường lấp đầy các vùng đích chỉ dựa trên việc thẩm thấu các thông tin bề mặt từ của vùng xung quanh chúng. Phương pháp này chỉ có thể xử lý các vùng trống hẹp trong ảnh có sự biến thiên về kết cấu và màu sắc nhỏ. Chúng thất bại trong việc tổng hợp các nội dung ngữ nghĩa do các thông tin chỉ đến từ các lân cận của nó và như vậy không thể giải quyết trường hợp vùng trống kích thước lớn. Các phương pháp dựa trên lấy mẫu chia nhỏ vùng đích thành các vùng trống nhỏ và nỗ lực tìm các vùng tương tự hoặc có liên quan đến các vùng này sau đó lắp ghép chúng vào vùng trống nhỏ tương ứng. Các phương pháp này có thể tổng hợp cho kết quả tương đối mượt và chấp nhận được nếu như giải thuật tham lam dùng để xác định ưu tiên của mảnh ghép tốt, nhưng chi phí tính toán là rất lớn. Khắc phục nhược điểm này PatchMatch [6] đề xuất một giải thuật tìm kiếm mẫu xấp xỉ nhanh cho kết quả khá tốt, tuy nhiên việc hoàn thiện ảnh sẽ thất bại nếu không tìm thấy mẫu ghép có độ so khớp cao và vẫn chưa đủ nhanh cho các ứng dụng thời gian thực. Một giới hạn khác của các cách tiếp cận này là không tạo ra được các cấu trúc chi tiết vì chúng chỉ xử lý trên bề mặt cục bộ mức thấp và không thể thu nhận các thông tin ngữ nghĩa ở mức cao.

Hình 2. *Kiến trúc mô hình đề xuất*

Gần đây, các cách tiếp cận dựa trên mạng DCNN thu được nhiều kết quả vượt trội trong lĩnh vực inpainting ảnh với các vùng đích có kích thước lớn [7] [19] [10] [9] [20]. Các phương pháp trong cách tiếp cận này cải thiện kết quả inpainting bằng cách sử dụng các thông tin ngữ nghĩa trong ảnh. Một trong các nghiên cứu đầu tiên dựa trên DCNN cho bài toán inpainting là Context Encoder [7], sử dụng một kiến trúc mã hóa – giải mã (encoder-decoder) để lấp đầy vùng trống, đồng thời bổ sung thêm hàm loss đối kháng (adversarial loss) trong pha huấn luyện để nâng cao chất lượng trực quan của ảnh hoàn thiện. Mặc dù Context Encoder hiệu quả trong việc đạt được cấu trúc tổng thể và ngữ nghĩa của ảnh, nhưng chỉ với kiến trúc mạng chuyển tiếp đơn các kết cấu chi tiết tốt vẫn không được sinh ra. Sau khi các mạng đối kháng sinh (generative adversarial networks (GAN)) được giới thiệu trong nghiên cứu [21], các nghiên cứu sau đó dựa trên GAN như [22] [23] [24] [20] [11] hoàn thiện vùng đích dựa trên lớp ngữ nghĩa của vùng nguồn đưa ra kết quả hợp lý hơn về mặt trực quan. Nghiên cứu [25] bổ sung thêm hàm loss cấu trúc nhằm duy

trì tái cấu trúc của cạnh. Zhang và các cộng sự [26] chia tiến trình lấp đầy vùng trống này thành nhiều pha, qua mỗi pha kích thước của vùng trống giảm dần tạo ra kết quả khá tốt. Tuy nhiên kích thước của vùng trống bị giới hạn là các vùng hình vuông hoặc oval. Chúng không thể xử lý với các vùng trống khác hoặc các mặt nạ với kích thước đa dạng. Lui và các cộng sự [13] sử dụng các phép nhân chập từng phần (partial convolution) trong đó phép nhân chập chỉ dựa trên các điểm ảnh chắc chắn nhằm giảm thiểu tác động gây ra bởi sự khác biệt phân bố giữa vùng mặt nạ và vùng ngoài mặt nạ. Phương pháp này ngoài việc sử dụng các mặt nạ hình dạng phổ dụng còn có thể áp dụng cho các mặt nạ không phổ dụng được sinh ra trong nghiên cứu [12] dựa trên ước lượng ảnh mặt nạ giữa hai khung ảnh liên tiếp trong video.

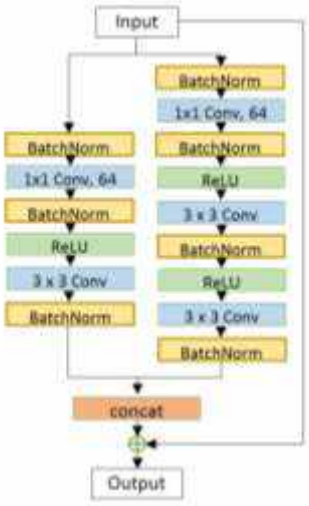
Hiện nay, các mạng DCNN đạt được hiệu suất thực thi rất cao trong nhận dạng và phân loại ảnh. Đặc biệt là mạng ResNet [27] có tác động to lớn đến sự phát triển của mạng nhân chập học sâu. Với khối cấu trúc được thiết kế hiệu quả tạo ra mạng có kiến trúc sâu hơn, khắc phục được vấn đề mất mát gradient tại pha huấn huyện [27]. Ngoài ra các khối residual còn chứa các kết nối nhanh (short-cut) cho kết quả tốt hơn với cả hiệu suất và thời gian thực thi. Các ưu điểm của kiến trúc residual được nghiên cứu cải tiến đưa vào mô hình đề xuất nhằm gia tăng kết quả inpainting ảnh.

3. Mô hình đề xuất

Mô hình đề xuất RBPconv của chúng tôi cho bài toán inpainting trên kiến trúc nền U-net kết hợp với các khối Residual cải tiến và phép nhân chập từng phần. Sơ đồ tổng quát của mô hình được minh họa trong hình 2.

3.1. Khối Residual

Kiến trúc của các khối residual cải tiến được minh họa trong hình 3. Khối này được chia thành 2 khối con. Đầu tiên bộ lọc nhân chập kích thước 1x1 được áp dụng cho mỗi khối con trong kiến trúc hình tháp với mục đích làm giảm số chiều của bản đồ đặc trưng trước khi áp dụng bộ lọc thông dụng 3x3. Điều này giúp cho số chiều của bản đồ đặc trưng, giảm chi phí tính toán. Ví dụ đầu ra của tầng trước (là đầu vào của tầng hiện tại) là 100x100x128 đi qua tầng nhân chập hiện tại cho đầu ra là 100x100x256 sau khi nhân chập với mặt nạ 3x3 với 256 kênh (stride =1, pad=2), thì các tham số sẽ là 128x3x3x256 = 294912. Nếu đầu ra của tầng trước đi qua tầng nhân chập kích thước 1x1 với 64 kênh trước và sau đó nhân chập với mặt nạ 3x3, 256 kênh thì kết quả vẫn là 100x100x256, nhưng tham số nhân chập giảm xuống 128x1x1x64 + 64x3x3x256=155648, tức là giảm gần 2 lần.



Hình 3. *Kiến trúc khối residual cải tiến*

Một khối con chứa một tầng nhân chập 3x3 khối còn lại chứa hai tầng 3x3 (hình 3). Các đặc trưng cục bộ của hai khối này với kích thước khác nhau được tập hợp và nối lại với nhau. Kết nối short-cut được áp dụng trực tiếp giữa đầu vào và đầu ra ngăn chặn mất gradient trong mạng học sâu. Các kết nối short-cut được chứng minh trong nghiên cứu [27] không làm gia tăng thêm các tham số cũng như độ phức tạp chi phí tính toán.

3.2. Partial Convolution

Khái niệm về nhân chập từng phần được đề xuất lần đầu trong nghiên cứu [13] áp dụng cho bài toán inpainting với các vùng trống

không phổ dụng đã thu được kết quả khả quan. Nhân chập từng phần có thể được suy ra bằng các mặt nạ và có được tái chuẩn hóa chỉ dựa trên các điểm ảnh hợp lệ. Gọi W là trọng số của bộ lọc nhân chập và b là độ lệch chuẩn tương ứng. X là các giá trị đặc trưng trong cửa sổ trượt hiện tại, M là mặt nạ nhị phân tương ứng. Nhân chập từng phần tại mỗi vị trí được biểu diễn như sau:

$$x' = \begin{cases} W^T(X \odot M)\frac{1}{sum(M)} + b, sum(M) > 0 \\ 0, \qquad\qquad ng\square\square c\ l\square i \end{cases} \quad (1)$$

Trong đó $\odot$ biểu diễn phép nhân từng phần tử tương ứng của hai ma trận. Có thể thấy rằng, các giá trị tính được chỉ phụ thuộc vào vùng ngoài mặt nạ. Nhân chập từng phần có ảnh hưởng tốt hơn nhân chập chuẩn khi xử lý chính xác với các mặt nạ kích thước bất kỳ. Khác với bài toán phân loại ảnh hay dò tìm đối tượng trong đó tất cả các điểm ảnh của ảnh đầu vào là hợp lệ, bài toán inpainting lại có nhiều điểm ảnh không hợp lệ nếu bị rơi vào vùng bị phá hủy hay các vùng trong mặt nạ. Các giá trị điểm ảnh của vùng mặt nạ thông thường được đặt là 0 hoặc 1. Tận dụng các ưu điểm của phép nhân chập từng phần này, mô hình đề xuất thay thế phép nhân chập chuẩn ở tất cả các tầng nhân chập bằng phép nhân chập từng phần.

Ngoài ra, theo sau mỗi phép nhân chập từng phần là cơ chế phát sinh và cập nhật mặt nạ tự động cho các tầng nhân chập tiếp theo như là một phần của mạng chuyển tiếp. Nếu như phép nhân chập có thể ước định đầu ra của nó trên ít nhất một giá trị đầu vào hợp lệ thì vị trí này được đánh dấu là hợp lệ. Điều này có thể được biểu diễn bởi công thức:

$$m' = \begin{cases} 1\ n\square u\ sum(M) > 0 \\ 0, \quad ng\square\square c\ l\square i \end{cases} \quad (2)$$

3.3. Kiến trúc mô hình

Nghiên cứu của chúng tôi bắt nguồn từ mô hình kiến trúc mạng encoder-decoder. Tuy nhiên để tăng tốc độ huấn luyện, chúng tôi đề xuất sử dụng các khối residual thay vì các tầng nhân chập thông thường cho các lớp ở giữa mạng này. Tại các mức đặc trưng thấp, cả tầng nhân chập đơn giản và tầng nhân chập phức tạp đều cho kết quả tương tự nhau [28]. Do đó tại tầng nhân chập thứ nhất, các mặt nạ 3x3x64 được sử dụng để thu được bản đồ đặc trưng mức thấp 64 chiều. Sau đó các khối residual được thiết lập cho các tầng nhân chập. Sự thay thế này làm gia tăng nhiều hiệu suất thực thi của mạng.

Trong mô hình kiến trúc mạng của chúng tôi, tương tự như kiến trúc mạng sử dụng trong [13] sử dụng kiến trúc mạng encoder-decoder với tổng cộng 16 tầng trong đó 8 tầng trong phần encoder và 8 tầng trong phần decoder tương ứng. Phần encoder được dùng để học các đặc trưng ảnh, đây cũng chính là một tiến trình mô tả đặc tính của các ảnh. Phần Decoder là một tiến trình khôi phục và giải mã các đặc trưng đã học tạo ra ảnh thực. Trong nhiều trường hợp, các thông tin được cung cấp bởi các điểm ảnh xung quanh một điểm ảnh được xem xét. U-net [29] sử dụng một kiến trúc mạng gồm 2 phần giảm mẫu (down-sampling) và tăng mẫu (up-sampling). Down-sampling được sử dụng để lấy dần các thông tin môi trường và tiến trình up-sampling trộn các đặc trưng đã học và các thông tin môi trường trong down-sampling để khôi phục các chi tiết.

Trong mô hình đề xuất mỗi tầng nhân chập nguyên bản trong U-net được thay thế là một khối residual cải tiến có kiến trúc trong hình 3. Trong cải tiến này mỗi tầng nhân chập con được theo sau bởi chuẩn hóa batch và hàm kích hoạt. Hàm kích hoạt ReLU được sử dụng cho các tầng encoder và LeakyReLU với alpha=0.2 được sử dụng trong các tầng decoder. Bên cạnh đó, tất cả các tầng nhân chập được thay thế bằng nhân chập từng phần. Zero padding với kích thước 1 được sử dụng để làm cho tất cả các bản đồ đặc trưng có cùng kích thước.

3.4. Hàm loss

Ký hiệu I_{in} là ảnh đầu vào chứa các vùng trống cần hoàn thiện, I_{rec} là ảnh khôi phục qua mô hình mạng, I_{gt} là ảnh chuẩn (grounth

truth). Gọi M là một mặt nạ nhị phân khởi tạo tương ứng với vùng ảnh bị xóa. Các phần tử trong M có giá trị 0 nếu điểm ảnh đó bị phá hủy và 255 cho các điểm ảnh còn lại. Khi đó để so sánh sự khác biệt giữa hai cấu trúc ảnh khôi phục và ảnh gốc trong hàm Loss cấu trúc sử dụng chuẩn L_1 được định nghĩa như sau:

$$\mathcal{L}_{rec} = \left\| M \odot (I_{gt} - I_{rec}) \right\|_1 \quad (3)$$

Hàm loss về trực quan (perceptual loss) dùng để đo sự khác biệt về trực quan và ngữ nghĩa giữa hai ảnh được định nghĩa tương tự như trong [30]:

$$\mathcal{L}_{per}^{\Phi,j}(I_{rec}, I_{gt}) = \frac{1}{C_j H_j W_j} \left\| \phi_j(I_{rec}) - \phi_j(I_{gt}) \right\|_1 \quad (4)$$

Trong đó $\phi_j(I)$ là các bản đồ đặc trưng kích hoạt đầu ra của tầng thứ j của mạng ϕ khi xử lý ảnh I; $\phi_j(I)$ là một bản đồ đặc trưng có kích thước C_j x H_j x W_j. Perceptual loss lần đầu tiên được áp dụng cho bài toán inpainting ảnh trong nghiên cứu [9].

Bên cạnh đó, hàm loss hình dạng (style loss) cũng được sử dụng để loại bỏ các thành phần lạ hình bàn cờ [23], tương tự như perceptual loss, nhưng ma trận tương quan (ma trận Gram) trên mỗi bản đồ đặc trưng được sử dụng và được định nghĩa như sau:

$$\mathcal{L}_{style}^{\Phi,j}(I_{rec}, I_{gt}) = \frac{1}{C_j H_j W_j} \left\| G_j^{\Phi}(I_{rec}) - G_j^{\Phi}(I_{gt}) \right\|_1 \quad (5)$$

Trong đó, $\phi_j(I)$ là một bản đồ đặc trưng mức cao có hình dạng C_j x H_j x W_j, đưa ra một ma trận gram G_j^{Φ} kích thước C_j x C_j và $\frac{1}{C_j H_j W_j}$ là hệ số chuẩn hóa cho tầng thứ j.

Qua các thí nghiệm, chúng tôi thiết lập các trọng số dựa trên kinh nghiệm thu được hàm loss tổng thể như sau:

$$Loss = 20\mathcal{L}_{rec} + \mathcal{L}_{perc}^{\Phi,j} + 100\mathcal{L}_{style}^{\Phi,j} \quad (6)$$

4. Thực nghiệm và kết quả

Trong nghiên cứu này, chúng tôi sử dụng tập mặt nạ tạo ra trong nghiên cứu [13] để kiểm thử mô hình đề xuất và so sánh kết quả với các mô hình khác. Tập mặt nạ huấn luyện này gồm 55.116 mặt nạ và tập kiểm thử gồm 24.886 mặt nạ. Tất cả các mặt nạ và ảnh dùng cho pha huấn luyện và kiểm thử đều có cùng kích thước 256x256. Một số mặt nạ minh họa như trong hình 4.

Hình 4. *Một số mặt nạ*

Để tiện so sánh kết quả thực nghiệm của mô hình đề xuất với các kết quả thực nghiệm của các nghiên cứu gần nhất, trong nghiên cứu này thực nghiệm được tiến hành với tập mặt nạ sinh ra bên trên cho tập dữ liệu Places2 [31]. Tiến trình huấn luyện được thực hiện trên máy chủ *Nvidia Tesla V100* GPU (16GB). Mô hình đề xuất được tối ưu hóa sử dụng giải thuật Adam [32] với tỷ lệ học là 0.0002, kích thước mỗi batch là 16.

Ảnh cần hoàn thiện GLCIC [10] CA [11] PIC [33] RBPConv

Hình 5. *So sánh kết quả của RBPconv với các phương pháp trước đó*

So sánh định tính

Hình 5 biểu diễn các kết quả trực quan của RBPconv so với một vài phương pháp được phát triển gần đây nhất như là GLCIC (Global and Local Consistent Image Completion) [10], CA(Contextual Attention) [11], PIC (Pluralistic Image Completion) [33]. Những kết quả này minh chứng rằng mặc dù không có một mạng tách biệt cho phát sinh cạnh như trong nghiên cứu [33] nhưng ảnh được khôi phục vẫn bảo toàn các cấu trúc hợp lý. Mô hình đề xuất tận dụng kiến trúc residual có thể cập nhật các mặt nạ từng bước và cũng cho phép các bộ lọc nhân chập tự hoàn thiện các đường bao. Hơn nữa trong ảnh hoàn thiện các vết mờ rất cũng ít xuất hiện. Các ảnh tạo ra bởi mô hình RBPConv gần với ground truth hơn các ảnh sinh từ các phương pháp khác. Mặc dù trong một số ít trường hợp có thể xuất hiện vết mờ, nhưng nó lại thích hợp với nền của các vùng xung quanh.

So sánh định lượng

Trong nghiên cứu này, chúng tôi sử dụng các độ đo chất lượng ảnh SSIM (Structural Similarity Index) [34] và PSNR (Peak Signal-to-Noise Ratio) [35] được cài đặt trong bộ Matlab R2017a để đo chất lượng của phương pháp đề xuất với các phương pháp inpainting khác. Các phương pháp so sánh được phát triển trước đó gồm CA(Contextual Attention) [11], PConv (Partial Convolution Unet) [13] và EC (EdgeConnect) [26]. Các giá trị cụ thể được thể hiện trong bảng 1. Để có được số liệu này chúng tôi đã sử dụng các trọng số của các mạng huấn luyện tương ứng có sẵn. Kết quả của PConv được lấy từ bài viết [13] do mã nguồn chưa được nhóm tác giả công bố. Các số liệu thống kê có được sau khi tính toán trên 1.000 ảnh ngẫu nhiên lấy từ tập kiểm thử. Kết quả cho thấy mô hình RBPcov cho hiệu suất thực thi tốt hơn các phương pháp khác.

Bảng 1. *Kết quả định tính (PSNR, SSIM) trên tập dữ liệu Places2 với các phương pháp: CA [11], PConv [13] and EC [23], * nghĩa là giá trị lấy từ bài báo [13]*

	CA	PConv*	EC	RBPConv
PSNR	21.34	24.90	24.65	25.29
SSIM	0.806	0.777	0.857	0.868

5. Kết luận

Trong nghiên cứu này, chúng tôi phát triển một mạng RBPConv cho bài toán inpainting dựa trên các khối residual cải tiến, phép nhân chập từng phần và kiến trúc nền U-net. Các khối residual cải tiến, thành phần chính của mạng RBPconv duy trì sự biểu diễn ảnh độ phân giải cao thích hợp cả cho tái cấu trúc kết cấu và sự hội tụ của mạng. Mô hình RBPconv đề xuất đặc biệt hiệu quả cho việc lấp đầy các vùng trống với hình dạng bất kỳ và kích thước không lớn phù hợp với các mặt nạ sinh ra khi xóa bỏ một đối tượng trong ảnh và thay thế nó bằng đối tượng khác tương ứng về mặt kích thước.

TÀI LIỆU THAM KHẢO

[1]. Bertalmio, M., Vese, L., Sapiro, G. and Osher, S., "Simultaneous structure and texture image inpainting," *IEEE transactions on image processing*, Vol. 12, No. 8, pp. 882-889, 2003.

[2]. Liu, D., Sun, X., Wu, F., Li, S., and Zhang, Y., "Image compression with edge-based inpainting," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 17, No. 10, pp. 1273-1287, 2007.

[3]. Criminisi, A., Perez, P., and Toyama, K., "Object removal by exemplar-based inpainting," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 2, pp. 721-728, 2003.

[4]. Drori, I., Cohen-Or, D., and Yeshurun, H., "Fragment-based image completion," *TOG*, Vol. 22, No. 3, pp. 303-312, 2003.

[5]. N. Komodakis, "Image completion using global optimization," *CVPR*, pp. 442–452, 2006.

[6]. Barnes, C., Shechtman, E., Finkelstein, A., Goldman, D. B., "Patchmatch: A randomized correspondence algorithm for structural image editing," *ACM Transactions on Graphics-TOG*, Vol. 28, No. 3, 2009.

[7]. Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., Efros, A.A., "Context encoders: Feature learning by inpainting," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2536–2544, 2016.

[8]. Yan, Z., Li, X., Li, M., Zuo, W., and Shan, S., "Shift-net: Image inpainting via deep feature rearrangement.," *arXiv preprint arXiv:1801.09392*, 2018.

[9]. Yang, C., Lu, X., Lin, Z., Shechtman, E., Wang, O., Li, H, "High-resolution image

inpainting using multi-scale neural patch synthesis," *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Vol. 1, pp. 3, 2017.
[10]. Iizuka, S., Simo-Serra, E., Ishikawa, H., "Globally and locally consistent image completion," *ACM Transactions on Graphics (TOG),* Vol. 36, No. 4, 2017.
[11]. Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., Huang, T.S., "Generative image inpainting with contextual attention," *arXiv preprint arXiv:1801.07892,* 2018.
[12]. Sundaram, N., Brox, T., and Keutzer, K., "Dense point trajectories by gpu-accelerated large displacement optical flow," *European conference on computer vision*, pp. 438-451, 2010.
[13]. Liu, G., Reda, F. A., Shih, K. J., Wang, T.-C., Tao, A., and Catanzaro, B., "Image inpainting for irregular holes using partial convolutions," *arXiv preprint arXiv:1804.07723,* 2018.
[14]. Bertalmio, M., Sapiro, G., Caselles, V., and Ballester, C., "Image inpainting," *Proceedings of the 27th annual conference on Computer graphics and interactive techniques. ACM Press/Addison-Wesley Publishing Co*, p. 417–424, 2000.
[15]. Darabi, S., Shechtman, E., Barnes,C., Goldman, D. B., and Sen, P., "Image melding: Combining inconsistent images using patch-based synthesis," *ACM Trans. Graph,* 2012.
[16]. Huang, J., Kang, S. B., Ahuja, N. and Kopf, J., "Image completion using planar structure guidance," *ACM Transactions on graphics (TOG),* 2014.
[17]. Sun, J., Yuan, L., Jia, J., Shum, H., "Image completion with structure propagation," *ACM Transactions on Graphics (ToG),* pp. 861–868, 2005.
[18]. Xu, Z., and Sun, J., "Image inpainting by patch propagation using patch sparsity," *IEEE transactions on image processing,* pp. 1153–1165, 2010.
[19]. Liu, P., Qi, X., He, P., Li, Y., Lyu, M. R., and King, I., "Semantically consistent image completion with fine-grained details," *arXiv preprint arXiv:1711.09345,* 2017.
[20]. Yeh, R. A., Chen, C., Lim, T. Y., Schwing, A. G., HasegawaJohnson, M., and Do,M. N., "Semantic image inpainting with deep generative models," *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* pp. 5485–5493, 2017.
[21]. Radford, A., Metz, L., and Chintala, S., "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.
[22]. Isola, P., Zhu, J., Zhou, T., and Efros, A. A., "Image-to-Image Translation with Conditional Adversarial Networks," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1125-1134, 2017.
[23]. Nazeri, K., Eric, Ng., Joseph, T., Qureshi, F., and Ebrahimi, M., "EdgeConnect: Generative Image Inpainting with Adversarial Edge Learning," *arXiv preprint arXiv:1901.00212,* 2019.
[24]. Xiong, W., Lin, Z., Yang, J., Lu, X., Barnes, C., and Luo, J., "Foreground-aware Image Inpainting," *arXiv preprint arXiv:1901.05945,* 2019.
[25]. Huy V. V., Ngoc Q. K. D., and Pérez,P., "Structural Inpainting," *Proceedings of the 26th ACM International Conference on Multimedia (MM '18),* pp. 1948–1956, 2018.
[26]. Zhang, H., Hu, Z., Luo, C., Zuo, W., and Wang, M., "Semantic Image Inpainting with Progressive Generative Networks," *ACM Multimedia Conference on Multimedia Conference*, pp. 1939–1947, 2018.
[27]. He, K., Zhang, X., Ren,S., and Sun, J., "Deep residual learning for image recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778, 2016.
[28]. Zeiler, M. D., and Fergus, R., "Visualizing and understanding convolutional networks," *arXiv:1311.2901,* 2013.
[29] Ronneberger, O., Fischer, P., and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, 2015.
[30]. Johnson, J., Alahi, A., and Fei-Fei, L., "Perceptual losses for real-time style transfer and super-resolution," *European Conference on Computer Vision*, p. 694–711, 2016.
[31]. Mahajan, K. S., Vaidya, M. B., "Image in Painting Techniques: A survey," *IOSR Journal of Computer Engineering,* vol. 5, no. 4, pp. 45-49, 2012.
[32]. Kingma, D. P., Ba, J. L.: Adam, "A method for stochastic optimization," *international conference on learning representations* , 2015.
[33]. Zheng, C., Cham,T., and Cai, J., "Pluralistic Image Completion," *CoRR abs/1903.04227,* 2019.
[34]. Zhou, W., Bovik, A. C., Sheikh, H. R., and Simoncelli E. P., "Image Qualifty Assessment: From Error Visibility to Structural Similarity.," *IEEE Transactions on Image Processing,* vol. 13, no. 4, p. 600–612, 2004.
[35]. Gonzalez, R., and Wood, R. , "Digital Image Processing," *Pearson Edn,* 2009.