

SIÊU PHÂN GIẢI VÀ KHỬ NHIỄU DÙNG TRÍ TUỆ NHÂN TẠO NÂNG CAO CHẤT LƯỢNG ẢNH TRONG NHẬN DẠNG VÂN TAY

SUPER-RESOLUTION AND DENOISE USING ARTIFICIAL INTELLIGENCE FOR IMAGE ENHANCEMENT
IN FINGERPRINT RECOGNITION

Doãn Thanh Bình

TÓM TẮT

Bài báo này đề xuất một giải pháp siêu phân giải và khử nhiễu sử dụng trí tuệ nhân tạo để giải quyết bài toán nâng cao chất lượng ảnh vân tay đầu vào cho hệ thống nhận dạng vân tay. Siêu phân giải và khử nhiễu hình ảnh là hai nhiệm vụ quan trọng trong xử lý hình ảnh nhằm cải thiện chất lượng hình ảnh. Siêu phân giải hình ảnh làm nhiệm vụ ánh xạ hình ảnh có độ phân giải thấp sang hình ảnh có độ phân giải cao trong khi khử nhiễu làm nhiệm vụ khôi phục từ hình ảnh bị nhiễu đầu vào. Bằng các phân tích và kết quả tính toán các tham số chất lượng ảnh xử lý qua thực nghiệm, chúng ta sẽ chứng minh rằng phương pháp được đề xuất là vượt trội so với một số phương pháp hiện có như biến đổi Curvelet, biến đổi wavelet rời rạc (DWT), siêu phân giải rất sâu (VDSR), mạng nơ ron tích chập siêu phân giải (SRCNN).

Từ khóa: Nhận dạng vân tay, siêu phân giải và khử nhiễu.

ABSTRACT

This paper presents a super-resolution and denoising solution using artificial intelligence to solve the problem of improving the quality of input fingerprint images for fingerprint recognition systems. Super resolution and image noise reduction are two important tasks in image processing to improve image quality. Image Super-Resolution is mapping a low-resolution image to a high-resolution image while denoising takes care of restoring from noisy input images. By analyzing and calculating the experimental image quality parameters, we will prove that the proposed method is superior to some existing methods such as Curvelet transform, discrete wavelet transform (DWT), very deep hyper-resolution (VDSR), hyper-resolution convolution neural network (SRCNN).

Keywords: Fingerprint recognition, super-resolution and denoising.

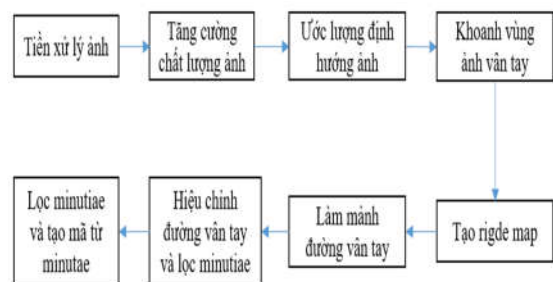
Trường Đại học Điện Lực
Email: binhdt@epu.edu.vn
Ngày nhận bài: 04/5/2021
Ngày nhận bài sửa sau phản biện: 02/7/2021
Ngày chấp nhận đăng: 25/8/2021

1. GIỚI THIỆU

Theo số liệu thống kê về đóng góp của các ngành sinh trắc học thì vân tay đứng vị trí hàng đầu trong các ứng dụng nhận dạng. Nhằm mục đích mô tả, người ta định nghĩa một số hình thức thể hiện của vân tay. Mỗi hình thức thể hiện sẽ có hình thức trích chọn đặc tả vân tay tương

ứng. Vì vậy, căn cứ vào ảnh vân tay chúng ta có thể chọn lựa đặc tính cần trích chọn và đưa ra phương án nhận dạng.

Các nguyên lý nhận dạng vân tay liên quan chặt chẽ đến cách thức tìm ra sự tương đồng giữa ảnh vân tay đầu vào với một bản mẫu trong cơ sở dữ liệu vân tay. Ba phương pháp chính trong nhận dạng vân tay là: dựa vào đối sánh điểm đặc trưng (Minutiae-base matching), sự tương quan (Correlation-base matching) và đặc tính đường vân tay (Ridge feature-base matching). Phương pháp dựa trên cơ sở đối sánh điểm đặc trưng là phổ biến hơn và được ứng dụng rất rộng rãi. Hai vân tay sẽ được so sánh bằng các điểm đặc trưng (minutiae) của các vân tay đó. Giai đoạn thứ nhất này còn gọi là tạo mã, các điểm đặc trưng tối thiểu này tạo thành tập hợp điểm đặc trưng và được lưu trữ trong từng mảng 2D. Giai đoạn thứ hai sẽ tìm vị trí khớp phù hợp giữa vân tay mẫu và ảnh vân tay đầu vào sao cho số các điểm đặc trưng tạo thành cặp tương đồng là lớn nhất. Hai đặc trưng cơ bản của vân tay là "Ridge termination" và "ridge termination".



Hình 1. Quá trình xử lý nhận dạng vân tay

2. XÂY DỰNG MÔ HÌNH

2.1. Siêu phân giải và khử nhiễu

Siêu phân giải là quá trình tạo ra hình ảnh có độ phân giải cao từ hình ảnh có độ phân giải thấp. Để giải quyết bài toán này, người ta sử dụng các phương pháp siêu phân giải hình ảnh dựa trên nội suy do cách thức này đơn giản và nhanh chóng. Các phương pháp này xử lý hình ảnh có độ phân giải thấp bằng kỹ thuật tăng tần số lấy mẫu. Tuy nhiên, điều này gây ra các hiệu ứng làm mờ hình ảnh gốc do các phương pháp như vậy cho kết quả dưới mức tối ưu trong việc khôi phục các chi tiết hình ảnh mẫu.

Mục tiêu của việc khử nhiễu hình ảnh là khôi phục một hình ảnh được khử nhiễu I từ một đối tượng nhiễu K với $K = I + V$ trong đó V là nhiễu (thường được giả định là nhiễu cộng Gaussian). Tham số thường được sử dụng để đo độ nhiễu trong hình ảnh là tỷ số tín hiệu trên nhiễu (PSNR). Trước đây, một số mô hình đã được xây dựng nhằm khử nhiễu hình ảnh như: mô hình Sparse, mô hình Gradient và mô hình Markov.... Mặc dù các phương pháp này cho hiệu suất cao trong việc khử nhiễu, nhưng lại phức tạp về mặt tính toán vì chúng liên quan đến vấn đề tối ưu hóa độ phức tạp trong giai đoạn thử nghiệm [3].

Phương pháp Discriminative Learning (phân biệt các mẫu dữ liệu với nhau) đang được áp dụng gần đây để khắc phục nhược điểm này. Một ví dụ tiêu biểu, sử dụng các phương pháp Discriminative Learning để tách nhiễu ra khỏi hình ảnh nhiễu bằng mạng tích chập CNN. Cách thức này hiện đang được sử dụng rộng rãi để giải quyết vấn đề suy giảm hiệu suất trong các mạng học sâu (Deep Learning) [4].

Hướng đi được nghiên cứu trong bài báo này là xây dựng mạng có khả năng thực hiện siêu phân giải và khử nhiễu hình ảnh được gọi là siêu phân giải khử nhiễu mạng nơ-ron tích chập (SD-CNN) sử dụng phương pháp Residual Learning của mạng tích chập. Thay vì học ánh xạ (mapping learning) đầu cuối, chúng ta huấn luyện mạng để tạo ra dư ảnh. Dư ảnh này là sự khác biệt của hình ảnh đầu vào và hình ảnh gốc (mẫu). Mục đích của hệ thống này là loại bỏ nhiễu và sau đó nâng cao độ phân giải hình ảnh (siêu phân giải).

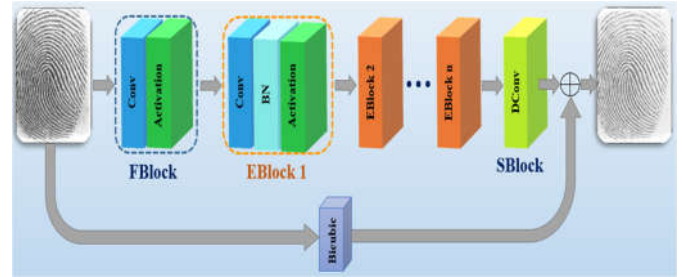
2.2. Kiến trúc mạng đề xuất

Mô hình đề xuất SDCNN là một mạng nơ-ron học sâu gồm 20 lớp. Hàm kích hoạt (activation function) được sử dụng là *tanh*. Hàm tổn thất (loss function) được sử dụng là sự kết hợp giữa sai số bình phương trung bình (MSE) và sai số tuyệt đối trung bình (MAE). Ma trận Kernel có kích thước là 3x3 và giống nhau cho tất cả 20 lớp tích chập. Số lượng ánh xạ tính năng được cung cấp cho lớp chập đầu tiên là 3 tương ứng với các kênh RGB của hình ảnh. Số lượng ánh xạ tính năng được cung cấp cho lớp 2 đến lớp 20 là 64. Đầu ra cuối cùng của mạng có 3 kênh. Tất cả các trọng số được khởi tạo từ phân phối ngẫu nhiên chuẩn. Với kiến trúc như vậy, tổng tham số mạng đề xuất là 672.835 trong đó 670.531 là tham số có thể đào tạo và 2.304 là tham số không thể đào tạo. Trong giai đoạn huấn luyện, đầu vào cho mạng là một hình ảnh nội suy hai chiều 32x32 và đầu ra mục tiêu là dư ảnh 32x32. Mạng có khả năng hoạt động với mọi kích thước hình ảnh đầu vào với trọng số mà nó học được trong quá trình đào tạo. Số lượng tham số có thể đào tạo cho lớp tích chập không phụ thuộc vào kích thước đầu vào [8].

Từ những phân tích trên, bài báo đề xuất kiến trúc của mạng với các khối Fblock tiến xử lý trích xuất thông tin, các khối Eblock khử nhiễu và tăng cường chất lượng ảnh, và khối khôi phục đầu ra Sblock (hình 2).

Quy trình này có thể được trình bày là:

$$I_0 = f(x) \tag{1}$$



Hình 2. Kiến trúc mạng đề xuất SDCNN

2.3. Hàm tổn thất (Hàm Loss - Loss Function)

Hai hàm Loss được xem xét sử dụng để đo độ lệch giữa hình ảnh độ phân giải cao theo dự đoán \hat{I} và thực tế I tương ứng. Hàm đầu tiên là sai số bình phương trung bình (MSE), là hàm Loss được sử dụng rộng rãi nhất để phục hồi hình ảnh chung như được định nghĩa dưới đây:

$$I_{MSE} = \frac{1}{N} \sum_{i=1}^N \|I_i - \hat{I}_i\|_2^2 \tag{2}$$

Tuy nhiên, qua thực nghiệm đã cho thấy rằng đào tạo với hàm Loss MSE không phải là một lựa chọn tốt [6]. Hàm Loss thứ hai có nghĩa là sai số tuyệt đối trung bình (MAE), được xây dựng như sau:

$$I_{MAE} = \frac{1}{N} \sum_{i=1}^N \|I_i - \hat{I}_i\|_1 \tag{3}$$

Để cải thiện hiệu suất, mô hình đề xuất trong bài báo đã thực hiện đào tạo mạng với MAE và sau đó tinh chỉnh theo MSE.

2.4. Đơn vị khử nhiễu và tăng cường chất lượng ảnh

Đơn vị khử nhiễu và tăng cường chất lượng ảnh là một kernel 3x3. Biểu thị các kích thước ánh xạ của lớp thứ i là D_i ($i = 1, \dots, 6$). Mối quan hệ của các lớp tích chập có thể được biểu thị bằng:

$$D_3 - D_1 = D_1 - D_2 = d \tag{4}$$

Trong đó, d biểu thị độ chênh lệch giữa lớp thứ nhất và lớp thứ hai hoặc giữa lớp thứ nhất và lớp thứ ba.

Tương tự, kích thước của các kênh trong mô-đun dưới cũng có mối quan hệ này và có thể được mô tả như sau:

$$D_6 - D_4 = D_4 - D_5 = d \tag{5}$$

Giả sử đầu vào của mô-đun này là I_{k-1} , đầu ra của mô-đun được biểu diễn:

$$P_1^k = C_a(I_{k-1}) \tag{6}$$

Trong đó, I_{k-1} biểu thị đầu ra của khối trước đó đồng thời là đầu vào của khối hiện tại, C_a biểu thị hoạt động tích chập nối tiếp.

Các ma trận feature map với các kích thước $\frac{D_3}{s}$ và đầu vào của lớp tích chập đầu tiên được liên kết với nhau trong khuôn khổ kênh:

$$R^k = C(S(P_1^k, \frac{1}{s}), I_{k-1}) \tag{7}$$

C, S đại diện cho hoạt động ghép nối và hoạt động phân chia. Cụ thể, kích thước của P_1^k là D_3 . Do đó, $S(P_1^k, \frac{1}{s})$ biểu thị rằng các đặc trưng kích thước $\frac{D_3}{s}$ được tìm nạp từ

P_1^k . Ngoài ra, $S\left(P_1^k, \frac{1}{s}\right)$ ghép nối các đặc trưng với B_{k-1} trong khuôn khổ kênh. Mục đích là để kết hợp thông tin trước đó với một số thông tin hiện tại. Chúng ta lấy phần còn lại của thông tin đường dẫn ngắn cục bộ làm đầu vào cho kernel, chủ yếu trích xuất thêm các ánh xạ tính năng đường dài,

$$P_2^k = C_b\left(S\left(P_1^k, 1 - \frac{1}{s}\right)\right) \tag{8}$$

Trong đó, P_2^k , C_b tương ứng là các hoạt động tích chập đầu ra và xếp nối của kernel. Do đó, đơn vị tăng cường có thể được trình bày như sau:

$$P^k = P_2^k + R^k = C_b\left(S\left(C_a(I_{k-1}), 1 - \frac{1}{s}\right)\right) + C\left(S\left(C_a(I_{k-1}), \frac{1}{s}\right), I_{k-1}\right) \tag{9}$$

Trong đó: P^k là đầu ra của đơn vị tăng cường.

2.5. Chuẩn hóa theo Batch (BN)

Chuẩn hoá theo batch được áp dụng cho từng tầng riêng lẻ (hoặc có thể cho tất cả các tầng) và hoạt động như sau: Trong mỗi vòng lặp huấn luyện, tại mỗi tầng, đầu tiên tính giá trị kích hoạt như thường lệ. Sau đó chuẩn hóa những giá trị kích hoạt của mỗi nút bằng việc trừ đi giá trị trung bình và chia cho độ lệch chuẩn. Cả hai đại lượng này được ước tính dựa trên số liệu thống kê của minibatch hiện tại [8].

Ánh xạ đặc tính đầu ra được xác định:

$$y_{ijkt} = w_k \frac{x_{ijkt} - \mu_k}{\sqrt{\sigma_k^2 + \epsilon}} + b_k \tag{10}$$

$$\mu_k = \frac{1}{HWT} \sum_{i=1}^H \sum_{j=1}^W \sum_{t=1}^T x_{ijkt}$$

$$\sigma_k^2 = \frac{1}{HWT} \sum_{i=1}^H \sum_{j=1}^W \sum_{t=1}^T (x_{ijkt} - \mu_k)^2$$

Trong đó, các mảng đầu vào và đầu ra được coi là các tensor 4D để làm việc với một loạt các ánh xạ đặc tính. Các tensors w và b xác định các hằng số nhân và cộng thành phần. T là kích thước batch và:

$$x, y \in R^{H \times W \times K \times T}, W \in R^K, b \in R^K \tag{11}$$

3. THỰC NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ

3.1. Thực nghiệm

Việc thực hiện khử nhiễu và siêu phân giải hình ảnh vân tay bằng mạng đề xuất sẽ tiến hành các bước như sau:

Bước 1: Tạo hình ảnh có độ phân giải thấp từ một hình ảnh tham chiếu có độ phân giải cao và thêm nhiễu.

Bước 2: Thực hiện khử nhiễu và siêu phân giải trên hình ảnh đầu vào bằng cách sử dụng phép nội suy xoắn bậc 3 (bicubic).

Bước 3: Thực hiện khử nhiễu và siêu phân giải trên hình ảnh đầu vào sử dụng mạng nơ ron đề xuất.

Bước 4: Đánh giá trực quan các hình ảnh độ phân giải cao được khôi phục bằng phương pháp đề xuất và các phương pháp hiện có.

Bước 5: Đánh giá chất lượng của hình ảnh khôi phục bằng cách định lượng độ tương tự của hình ảnh với hình ảnh tham chiếu có độ phân giải cao qua các tham số tính toán chất lượng ảnh RMSE, PSNR, Entropi, SC, NIQE [6], SSIM [7].

Cuối cùng, thực hiện đánh giá hiệu suất về thời gian thực thi giữa các phương pháp [1 ÷ 5].

3.2. Đánh giá kết quả

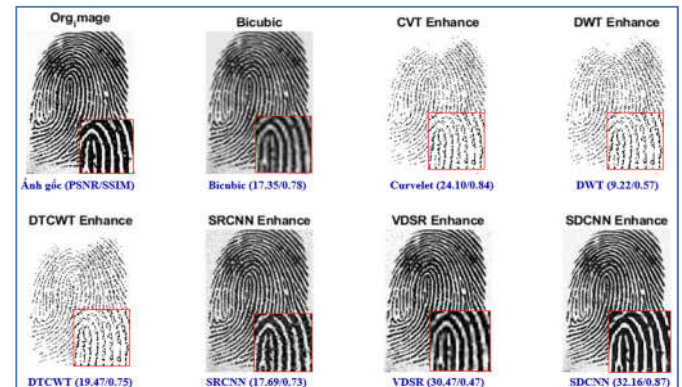
Đầu vào cho mạng đề xuất trong giai đoạn đào tạo là các hình ảnh (patches) có kích thước 32 x 32. Tác giả đã sử dụng thuật toán tối ưu lặp Descent Gradient với tốc độ học là 0.001, số Epoch được lựa chọn là 50.

3.2.1. Đánh giá trực quan

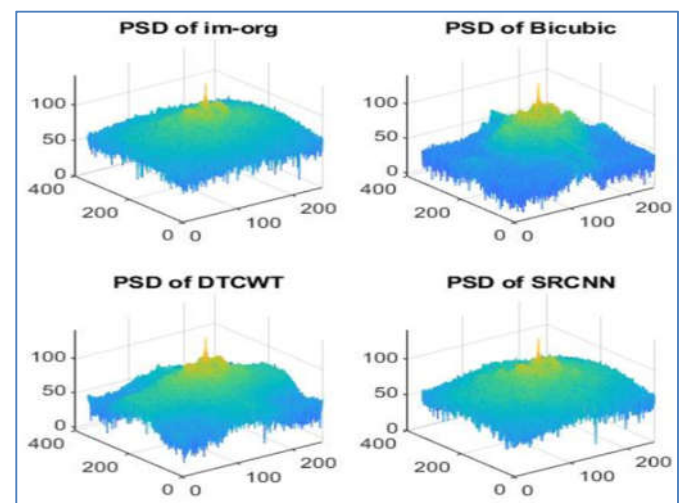
Kết quả đánh giá trực quan hình ảnh vân tay và mật độ phổ năng lượng sau khi sử dụng phương pháp đề xuất được thể hiện như trong hình 3 và 4.

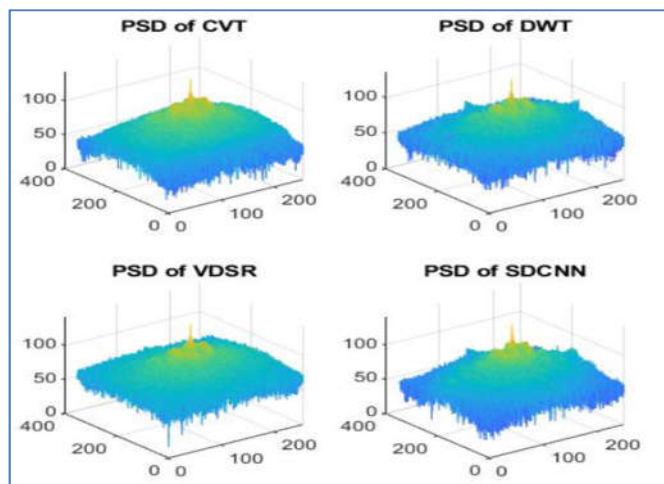
Chúng ta có thể thấy rằng kết quả trực quan theo phương pháp đề xuất thể hiện chất lượng hình ảnh tốt nhất. Ảnh khôi phục với phương pháp đề xuất cho độ nhạy sáng tốt hơn và khả năng quan sát rõ ràng hơn các chi tiết ảnh.

Đối với mật độ phổ năng lượng, ảnh chất lượng thấp sẽ cho một mật độ năng lượng quang phổ phẳng. Do đó, từ các kết quả trên dễ dàng thấy rằng phương pháp đề xuất cho chất lượng xử lý tốt nhất bởi vì PSD càng lớn cho thấy kết quả tăng cường ảnh càng tốt.



Hình 3. Đánh giá trực quan ảnh vân tay





Hình 4. Mật độ phổ năng lượng

3.2.2. Đánh giá các tham số tính toán chất lượng hình ảnh

Các tham số đánh giá chất lượng hình ảnh khi thực hiện bằng phương pháp đề xuất SDCNN với các phương pháp khác được thể hiện như trong bảng 1 với màu xanh chỉ ra mức hiệu suất tốt nhất và màu đỏ chỉ ra mức hiệu suất tốt thứ hai.

Chúng ta có thể thấy rằng có sự cải thiện đáng kể giá trị PSNR đối với nhiễu. Sự cải thiện chất lượng này đáng kể đến mức chúng ta có thể dễ dàng cảm nhận nó bằng mắt thường.

Bảng 1. Thông số đánh giá chất lượng ảnh theo từng phương pháp

Method	RMSE	PSNR	Entropi	SC	NIQE	SSIM	Time
Bicubic	34,5615	17,3590	7,7301	0,8989	18,5838	0,7889	-
CVT	15,8890	24,1089	6,0688	0,9917	19,7447	0,8462	8,2649
DWT	88,1914	9,2223	6,6997	0,8542	28,0550	0,5716	3,1145
DTCWT	27,0834	19,4768	6,0023	0,9003	20,2188	0,7548	3,1622
SRCNN	33,2316	17,6998	7,7082	1,0364	23,9151	0,7357	5,5199
VDSR	10,2414	30,4753	7,4619	0,9496	26,5871	0,4724	2,5928
SDCNN	6,2839	32,1662	8,4809	0,7007	18,6761	0,8774	1,6397

4. KẾT LUẬN

Từ kết quả nghiên cứu và thực nghiệm cho thấy rằng mạng được đề xuất đã chứng minh được tính hiệu quả trong việc khử nhiễu và siêu phân giải các hình ảnh nhận dạng vân tay. Mạng có độ học sâu phù hợp để học các dạng nhiễu từ chính hình ảnh vân tay đầu vào. Một mạng đơn giản không thể thực hiện tốt tác vụ này vì số lượng ít các lớp dẫn tới số lượng tham số có thể huấn luyện ít hơn không đủ khả năng để nắm bắt các mức độ nhiễu và loại nhiễu khác nhau. Do đó, thực hiện tác vụ khử nhiễu mà không có thông tin trước về nhiễu. Trừ khi mạng có một trường tiếp nhận lớn phù hợp nếu không nó sẽ không thể tạo ra mô hình nhiễu. Trong mạng đề xuất SDCNN, trường tiếp nhận lớn này đạt được nhờ một số lượng lớn các lớp tích chập. Vì vậy, mạng đề xuất có thể bắt được các mẫu

nhiều khác ngoài Gaussian và Poisson. Bên cạnh đó, giải pháp đề xuất có thời gian thực thi và thời gian đào tạo tương đối hợp lý, nó cho thấy cơ hội tốt để trở thành một giải pháp cải thiện hình ảnh trong các lĩnh vực xử lý ảnh.

TÀI LIỆU THAM KHẢO

- [1]. O J. Kim, J. K. Lee, K. M. Lee, 2016. *Accurate image super-resolution using very deep convolutional networks*. CVPR, pp. 1646-1654.
- [2]. C. Dong, C. C. Loy, X. Tang, 2016. *Accelerating the super-resolution convolutional neural network*. Computer Vision and Pattern Recognition. pp. 391-407.
- [3]. Zhang, K., W. Zuo, Y. Chen, D. Meng, L. Zhang, 2016. *Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising*. IEEE Transactions on Image Processing, Vol. 26, Number 7, pp. 295-307.
- [4]. C. Dong, C. C. Loy, K. He, X. Tang, 2017. *Image super-resolution using deep convolutional networks*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 38(2), pp. 3142-3155.
- [5]. C. Patvardan, A.K.Verma, C.V. Lakshmi, 2012. *Denoising of Document Images using Discrete Curvelet Transform for OCR Applications*. International Journal of Computer Applications.
- [6]. Mittal, A., R. Soundararajan, A. C. Bovik, 2016. *Making a Completely Blind Image Quality Analyzer*. IEEE Signal Processing Letters, Vol. 22, Number 3, pp. 209-212.
- [7]. Zhou, W., A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, 2004. *Image Quality Assessment: From Error Visibility to Structural Similarity*. IEEE Transactions on Image Processing, Vol. 13, Number 4, pp. 600-612.
- [8]. S. Ioe, C. Szegedy, 2015. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. ArXiv e-prints.

AUTHOR INFORMATION

Doan Thanh Binh

Electric Power University