

ỨNG DỤNG KỸ THUẬT HỌC SÂU DEBLURGAN TRONG KHỬ MỜ ẢNH SỐ

APPLICATION OF DEBLURGAN DEEP LEARNING IN DIGITAL IMAGE DE-BLUR

Đào Thụy Ánh, Đường Tuấn Hải

Khoa Công nghệ thông tin, Trường Đại học Kinh tế - Kỹ thuật Công nghiệp

Đến Tòa soạn ngày 29/04/2022, chấp nhận đăng ngày 17/05/2022

Tóm tắt: Học sâu (deep learning) là một phần của học máy (machine learning), còn học máy là thuật ngữ chung để chỉ các phương pháp, chương trình khiến cho máy tính có thể thông minh lên, tạo ra trí tuệ nhân tạo (AI). Khử mờ ảnh bị mờ do chuyển động (motion blur). Dựa trên tiến bộ đáng kể gần đây đã đạt được trong các lĩnh vực liên quan đến siêu phân giải hình ảnh (super-resolution) và xóa bỏ chi tiết thừa trong ảnh (inpainting) bằng cách áp dụng các mạng sinh phân biệt (GANs). Bài báo này trình bày kỹ thuật học sâu deblurGAN - một cách tiếp cận dựa trên các mạng sinh phân biệt có điều kiện và hàm mất mát đa thành phần (multi-component loss function) áp dụng cho bài toán khử mờ ảnh số bị mờ do chuyển động.

Từ khóa: Học sâu, mạng sinh phân biệt (GAN), deblurGAN, hàm mất mát đa thành phần.

Abstract: Deep learning is a part of machine learning. Machine learning is a general term that refers to methods and programs that make computers smarter, and creating artificial intelligence (AI). Image de-blur that are blurred due to motion. Building on the significant progress recently made in the areas related to image super-resolution and remove redundant details in digital photos remove excess detail in the image by applying Generative Adversarial Networks (GANs). This paper presents deblurGAN deep learning technique - an approach based on Generative Adversarial Networks and multi-component loss function applied to the problem of de-blurring digital images that are blurred due to motion.

Keywords: Deep learning, GANs, deblurGAN, Multi-component loss function.

1. GIỚI THIỆU

Học sâu (deep learning) là một phần của học máy (machine learning), còn machine learning là thuật ngữ chung để chỉ các phương pháp, chương trình khiến cho máy tính có thể thông minh lên, tạo ra trí tuệ nhân tạo (AI). Ngành trí tuệ nhân tạo (AI) thường được coi là chính thức ra đời vào năm 1956, và đã trải qua hai chu kỳ “boom-bust” (hay có thể gọi là: sôi sục – nguội lạnh), và hiện đang ở thời kỳ sôi sục mới, chính nhờ vào học sâu, nay đã trở thành phần chính, phần quan trọng nhất của lĩnh vực

học máy. Rất nhiều ứng dụng tin học quan trọng nổi bật gần đây là có công đóng góp của học sâu: dịch vụ khách hàng, *nhận dạng phim ảnh*, chuẩn đoán bệnh, phòng chống tội phạm, xe tự lái, dịch tiếng tự động, thư ký robot... [4] Hầu như trong bất cứ lĩnh vực nào cũng có thể hình dung các công nghệ mới trong đó AI đóng vai trò quan trọng, và AI đó chính là dựa trên học máy, và phần lớn là học sâu. Khử mờ ảnh bị mờ do chuyển động (motion blur)[1] bằng *phương pháp học sâu GAN*. Dựa trên tiến bộ đáng kể gần đây đã đạt được trong các

lĩnh vực liên quan đến siêu phân giải hình ảnh (super-resolution) và xóa bỏ chi tiết thừa trong ảnh (inpainting) bằng cách áp dụng các mạng sinh phân biệt (GANs). GAN được biết đến với khả năng bảo toàn các chi tiết kết cấu trong hình ảnh, tạo ra các giải pháp gần giống với bản sao hình ảnh thực và trông có vẻ thuyết phục. Lấy cảm hứng từ công việc gần đây về siêu phân giải hình ảnh và dịch từ hình ảnh sang hình ảnh của các mạng GAN, chúng tôi coi việc khử mờ là một trường hợp đặc biệt của dịch từ hình ảnh sang hình ảnh. Chúng tôi trình bày deblurGAN - một cách tiếp cận dựa trên các mạng sinh phân biệt có điều kiện và hàm mất mát đa thành phần (multi-component loss function). Bài báo này trình bày 3 vấn đề chính khi sử dụng phương pháp deblurGAN xử lý ảnh số bị mờ do chuyển động.

Đầu tiên, đưa ra một sự mất mát và mô hình học sâu GAN được coi là hiện đại nhất trong việc khử mờ chuyển động, nó nhanh hơn vài lần so với đối thủ cạnh tranh thuộc loại nhanh nhất.

Thứ hai, trình bày một phương pháp dựa trên các quỹ đạo ngẫu nhiên để tạo ra một bộ dữ liệu phục vụ cho đào tạo khử mờ chuyển động theo kiểu tự động từ bộ dữ liệu sắc nét. Việc

kết hợp nó với một bộ dữ liệu hiện có để học cách làm mờ chuyển động sẽ cải thiện kết quả so với chỉ đào tạo về hình ảnh trong thế giới thực.

Cuối cùng, trình bày một bộ dữ liệu mới và phương pháp để đánh giá các thuật toán khử mờ dựa trên cách chúng cải thiện kết quả.

2. MỘT SỐ PHƯƠNG PHÁP KHỬ MỜ ẢNH

2.1. Phương pháp khử mờ ảnh truyền thống

- Thuật toán Richardson-Lucy, hay giải chấp Lucy-Richardson (LR) [6].
- Khử mờ ảnh bằng phương pháp Regularized Filter (RF) [2].

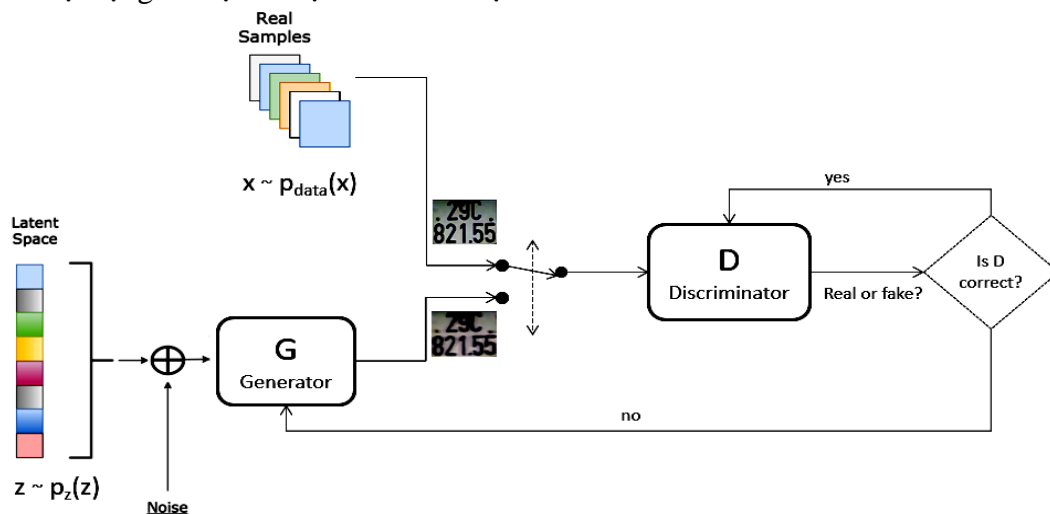
2.2. Phương pháp khử mờ ảnh hiện đại [8]

Người ta phân chia các phương pháp khử mờ ảnh hiện nay làm bốn nhóm chính:

- Maximum a posteriori (MAP);
- Variational Bayesian;
- Edge prediction;
- Power-spectrum methods.

3. KHỬ MỜ ẢNH BẰNG PHƯƠNG PHÁP HỌC SÂU GAN

3.1. Generative Adversarial Network (GAN)



Hình 1. Mô hình hoạt động của GAN

Generative Adversarial Network (GAN): mạng sinh phân biệt - hình thành trên ý tưởng về sự cạnh tranh của hai mạng neural network [3]:

- Discriminative network (mạng phân biệt): có nhiệm vụ cố gắng phân biệt giữa data thật và data giả mạo.
- Generative network (mạng sinh): sinh ra data giả, và mục tiêu là làm sao cho sinh ra được các data giống với thật nhất, làm cho mạng phân biệt không thể phân biệt nổi.

Mô hình hoạt động của GAN xem hình 1

Generator (bộ sinh) lấy một đầu vào ngẫu nhiên và cố gắng sinh ra được mẫu mới của data truyền vào. Như mô tả ở hình trên, generator $G(z)$ lấy input z từ $p_z(z)$, với z là mẫu thuộc phân phối xác suất $p(z)$, được sinh ngẫu nhiên từ latent space, sau đó gán thêm nhiễu (noise). Mẫu mới sinh ra từ $G(z)$ được nạp vào Discriminator network $D(x)$. Công việc của Discriminator network là lấy input từ tập train (real sample) và mẫu mới được sinh ra từ G (generated sample) và xác định xem mẫu nào mới là thật. Mẫu thật x được lấy từ phân phối xác suất $p_{data}(x)$.

$D(x)$ xử lý vấn đề phân loại nhị phân bằng cách sử dụng hàm sigmoid, trả về kết quả khoảng từ 0 đến 1, với xác suất đầu ra càng cao thì khả năng mẫu đó là thật (mẫu lấy từ tập data) càng lớn, và ngược lại. D được huấn luyện để tối đa xác suất gán đúng nhãn cho mẫu mới, đồng thời G lại được huấn luyện để tối thiểu khả năng phát hiện của D , tương đương tối thiểu $\log(1 - D(G(z)))$. Nói cách khác, việc huấn luyện D và G tương ứng với trò chơi minimax giữa hai người cho hàm số [3]:

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log (1 - D(G(z)))]$$

Trong hàm số $V(D, G)$ [7]:

$E_{x \sim p_{data}(x)} [\log D(x)]$ là giá trị kỳ vọng khả năng sample từ phân phối training được D đánh giá dữ liệu thật. Giá trị này càng cao thì khả năng đánh giá dữ liệu training của D càng chính xác.

$E_{z \sim p_z(z)} [\log (1 - D(G(z)))]$ là giá trị kỳ vọng khả năng sample từ G (được G sinh ra từ phân phối p_z) được D đánh giá là dữ liệu giả. Giá trị càng cao khả năng đánh giá dữ liệu sinh của D càng chính xác.

Tổng thể, D cố gắng maximize $V(D, G)$ trong khi G thì ngược lại. Quá trình huấn luyện sẽ hội tụ khi $p_g = p_{data} \Leftrightarrow D(x) = \frac{1}{2} \Leftrightarrow D$ không thể phân biệt được 1 sample là thật hay giả (xác suất đều là 0.5).

Quá trình training của GAN được chia thành phần, thực hiện theo các bước như sau:

Huấn luyện Discriminator network, cố định Generator network (cố định nghĩa là chỉ lan truyền xuôi (feedforward), không lan truyền ngược backpropagation): Discriminator được đào tạo về dữ liệu thực cho n epoch và xem liệu nó có thể dự đoán chính xác chúng là thật hay không. Ngoài ra, trong giai đoạn này, Discriminator cũng được đào tạo về dữ liệu được tạo ra từ Trình tạo và xem liệu nó có thể dự đoán chính xác chúng là giả hay không.

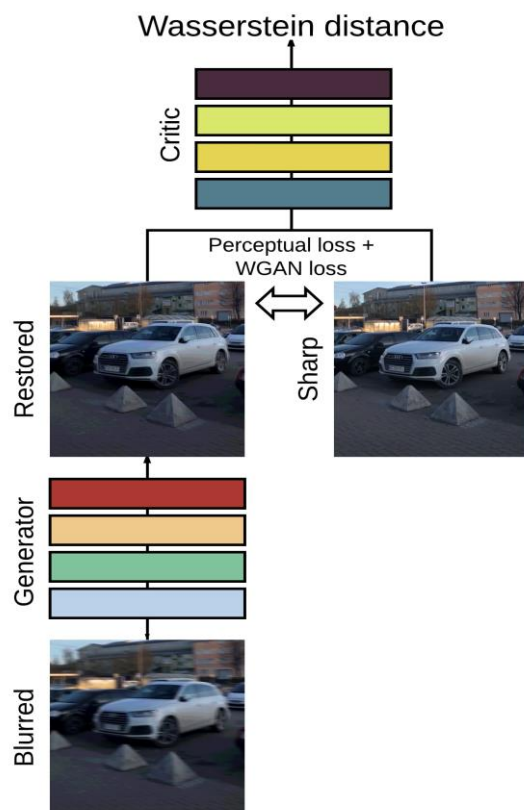
Sau khi Discriminator được huấn luyện bởi dữ liệu giả được tạo ra từ Trình tạo, chúng ta có thể nhận được dự đoán của nó và sử dụng kết quả để huấn luyện Trình tạo và nhận được tốt hơn từ trạng thái trước đó để thử và đánh lừa Trình phân biệt.

2.2. Mô tả thuật toán

Mục tiêu là để khôi phục hình ảnh sắc nét I_S mà chỉ được cung cấp một hình ảnh mờ I_B làm đầu vào, do đó không có thông tin nào về hạt nhân mờ được cung cấp. Việc khử mờ được

thực hiện bằng cách huấn luyện CNN G_{θ_G} , được gọi là *Trình sinh*. Đối với mỗi I_B , nó ước tính hình ảnh I_S tương ứng. Ngoài ra, trong giai đoạn huấn luyện, bài báo giới thiệu *Trình phân biệt* D_{θ_D} và huấn luyện cả hai mạng theo cách đối nghịch [9].

Mạng sinh lấy hình ảnh mờ làm đầu vào và tạo ra của hình ảnh sắc nét ước tính. Mạng phân biệt lấy các hình ảnh được tạo ra và hình ảnh sắc nét để tìm ra độ khác nhau giữa chúng. Tổng mất mát bao gồm mất mát WGAN từ bộ phân biệt và mất mát về cảm quan. Sự mất mát về nhận thức là sự khác biệt giữa các mạng đặc trưng VGG-19 *conv3.3* của hình ảnh sắc nét và hình ảnh được tạo ra. Sau khi huấn luyện thành công, chỉ có trình sinh được giữ lại.



Hình 2. Mô hình đào tạo deblurGAN

Hàm mất mát được xây dựng dựa trên sự kết hợp của mất mát nội dung và mất mát phân biệt [5]:

$$\mathcal{L} = \underbrace{\mathcal{L}_{GAN}}_{adv\ loss} + \underbrace{\lambda \cdot \mathcal{L}_X}_{content\ loss} = total\ loss$$

Mất mát phân biệt, hầu hết các bài báo liên quan đến GAN có điều kiện đều sử dụng vanilla GAN làm hàm mất mát. Gần đây xuất hiện một cách khác là sử dụng LSGAN (Least Squares Generative Adversarial Networks) ổn định hơn và tạo ra kết quả chất lượng cao hơn. Chúng tôi sử dụng WGAN-GP làm chức năng phân biệt, được chứng minh là mạnh mẽ cho sự lựa chọn kiến trúc Trình phân biệt [9]. Mất mát được tính như sau [1]:

$$\mathcal{L}_{GAN} = \sum_{n=1}^N -D_{\theta_D}(G_{\theta_G}(I^B))$$

DeblurGAN được đào tạo mà không có thành phần GAN hội tụ, nhưng tạo ra hình ảnh mịn và mờ.

Mất mát nội dung: Hai lựa chọn cổ điển cho hàm mất mát nội dung là mất mát L1 hoặc MAE, mất mát L2 hoặc MSE trên các pixel thô. Việc sử dụng các chức năng đó làm mục tiêu tối ưu hóa duy nhất dẫn đến các tạo phẩm mờ trên các hình ảnh được tạo do trung bình pixel của các giải pháp có thể có trong không gian pixel. Thay vào đó, chúng tôi đã thông qua mất đề xuất Nhận thức gần đây. Mất mát cảm quan là mất L2 đơn giản, nhưng dựa trên sự khác biệt của hình ảnh được tạo ra và mạng đặc trưng mục tiêu [9]. Nó được định nghĩa như sau:

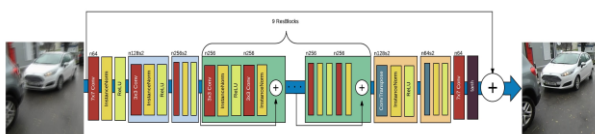
$$\mathcal{L}_X = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I^S)_{x,y} - \phi_{i,j}(G_{\theta_G}(I^B))_{x,y})^2$$

trong đó $\phi_{i,j}$ là mạng đặc trưng thu được từ tích chập thứ j (sau khi kích hoạt) trước lớp maxpooling thứ i trong mạng VGG19, được đặt trước trên ImageNet, $W_{i,j}$ và $H_{i,j}$ là các kích thước của ánh xạ đặc trưng (feature map). Trong bài báo này tác giả sử dụng kích hoạt từ

tích chập VGG3,3. Kích hoạt của các lớp sâu hơn đại diện cho các đặc trưng trừu tượng cao hơn.

Mất mát cảm quan tập trung vào việc khôi phục nội dung chung trong khi mất mát phân biệt tập trung vào khôi phục chi tiết kết cấu. DeblurGAN được đào tạo mà không có mất mát cảm quan hoặc với MSE đơn giản trên pixel thay vào đó không hội tụ về trạng thái có ý nghĩa.

Kiến trúc mạng nơron tích chập (CNN) của trình tạo như sau: Nó chứa hai khối tích chập có sai 1/2, chín khối dư (ResBlocks) và hai khối chập chuyển vị. Mỗi ResBlock bao gồm một lớp tích chập, lớp chuẩn hóa và kích hoạt ReLU. Chuẩn hóa Dropout với xác suất 0,5 được thêm vào sau lớp chập đầu tiên trong mỗi ResBlock. CNN học IR hiệu chỉnh dư cho hình ảnh mờ IB, vì vậy $IS = IB + IR$. Công thức như vậy làm cho đào tạo nhanh hơn và kết quả mô hình tổng quát tốt hơn. Trong giai đoạn đào tạo, xác định một mạng phân biệt D0D, đó là Wasserstein GAN với hình phạt độ dốc (gradient penalty) gọi là WGAN-GP. Kiến trúc của mạng phê bình giống hệt với PatchGAN. Tất cả các lớp chập trừ lớp cuối cùng được theo sau bởi lớp chuẩn hóa InstanceNorm và kích hoạt LeakyReLU với $\alpha = 0 : 2$ [9].



Hình 3. Kiến trúc mạng của trình tạo

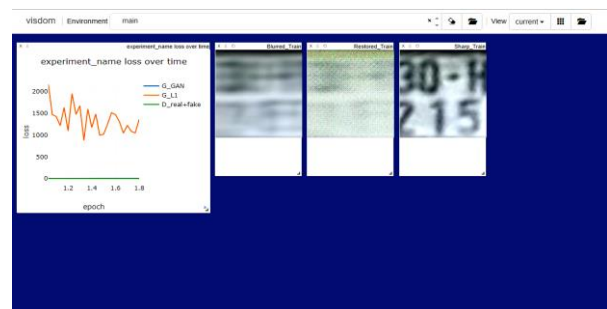
3. KẾT QUẢ THỰC NGHIỆM

Triển khai tất cả các mô hình của mình bằng khung học sâu PyTorch [11]. Việc đào tạo được thực hiện trên một máy tính chuyên xử lý đồ họa có cấu hình GPU Nvidia GeForce GTX Titan X, chip Core(TM) i9-12000K và RAM 32Gb...

Chuẩn bị một bộ dữ liệu ảnh tĩnh của xe máy được chụp tại chỗ đậu xe trong một số trung tâm thương mại lớn. Các hình ảnh được tiền hành tiền xử lý bằng cách tách riêng nội dung biển số xe ra. Sau đó một số ảnh biển số chính diện, không phân biệt màu sắc hay kích thước. Bộ dữ liệu sau đó được xử lý bằng hàm *filter2D* kết hợp với các hạt nhân mờ ngẫu nhiên để tạo ra những ảnh mờ chuyển động tương ứng.

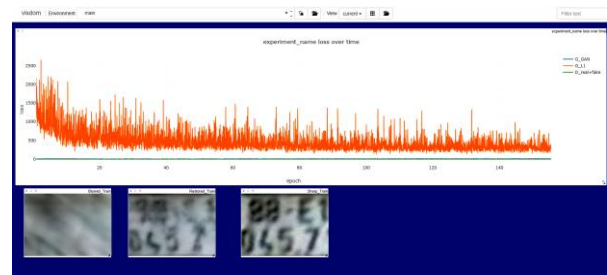
Các tham số cơ bản như: Kiểu GAN: gan_type = wgan-gp (Wasserstein GAN with Gradient Penalty), số lượng ảnh huấn luyện: 600, Số epoch: 300, learning rate: 0.0001, batchSize: 1

Một số hình ảnh thể hiện quá trình huấn luyện. Trong đó biểu đồ thể hiện độ mất mát. 3 ảnh tiếp theo tương ứng là ảnh mờ, ảnh được tạo ra nhờ bộ sinh và ảnh nét.



Hình 4. Hình ảnh trực quan của việc huấn luyện ở epoch 1

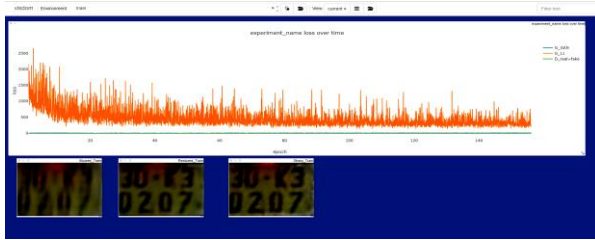
Bắt đầu huấn luyện ở epoch đầu tiên. Mất mát đang còn khá lớn, và bộ sinh chưa thể tạo ra những bức ảnh tốt.



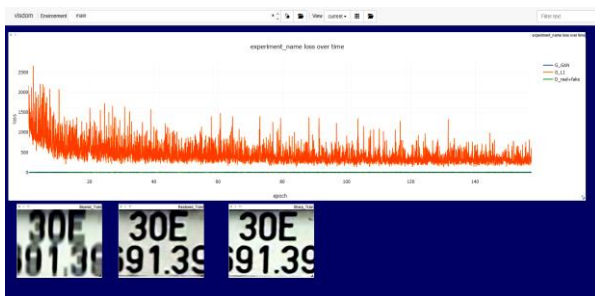
Hình 5. Hình ảnh trực quan của việc huấn luyện ở epoch 150

Đến Epoch thứ 150, mất mát đang có xu

hướng giảm xuống rõ rệt, và bằng cảm mắt thường có thể thấy bộ sinh đã tạo ra sự khác biệt giữa ảnh sinh ra và ảnh mờ truyền vào.

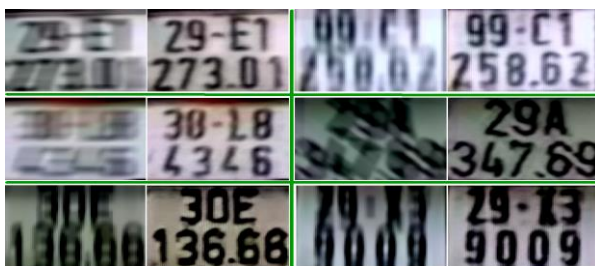


Hình 6. Hình ảnh trực quan của việc huấn luyện ở epoch 160



Hình 7. Hình ảnh trực quan của việc huấn luyện ở epoch 170

Càng về sau, hình ảnh được tạo ra đã được khử mờ rõ rệt. Phương pháp deblurGAN có chỉ số PSNR ổn định và ngang bằng hoặc cao hơn các phương pháp khác. Bên cạnh đó chỉ số SSIM xấp xỉ bằng với một số phương pháp hiện nay. Một số hình ảnh thực tế trong quá trình thử nghiệm:



Hình 8. Ảnh khử mờ 1 số trong quá trình huấn luyện

Bằng mắt thường có thể thấy, ảnh đã được khử mờ khá tốt. Chi tiết được phục hồi đáng kể. Có thể đọc được số một cách rõ ràng.

Với mục đích xem bộ khử mờ có thể sử dụng thực tế hay không, lấy 1 ảnh bị mờ trên mạng internet, sau đó cắt phần có biển số bị mờ ra.



Hình 9. Ảnh thực tế có chứa biển số xe bị mờ

Tiến hành tách biển số xe ra và đưa vào khử mờ.



Hình 10. Ảnh kết quả thử nghiệm

Kết quả cho thấy ảnh được khử mờ tương đối tốt. Mặc dù ảnh được khử mờ chưa được sắc nét như các ảnh mờ nhân tạo ở trên, nhưng kết quả này cũng coi là chấp nhận được.

4. KẾT LUẬN

Trong bài báo này các tác giả đã tìm hiểu và trình bày về phương pháp khử mờ ảnh số dựa trên kỹ thuật học sâu deblurGAN. Đồng thời triển khai mô hình để tiến hành huấn luyện. Kết quả 1 số bước của quá trình thực nghiệm được đưa ra dưới dạng hình ảnh. Bằng các kết quả thu được, các tác giả nhận thấy việc ứng dụng kỹ thuật học sâu GAN vào khử mờ ảnh số do chuyển động thu được kết quả tốt hơn so với nhiều phương pháp hiện nay đang dùng và có thể đưa vào ứng dụng trong các bài toán thực tế ví dụ như nhận dạng biển số phương tiện từ hình ảnh thu được bởi các camera giám sát được trình bày ở trên. Trong tương lai, các tác giả sẽ tập trung nghiên cứu sâu thêm về

deblurGAN để đưa vào áp dụng với những bộ dữ liệu lớn kết hợp với các vòng lặp huấn luyện lớn và những tham số đầu vào hợp lý hơn nhằm áp dụng cho các bài toán khử mờ ảnh số thực tế như nhận dạng văn bản, khuôn mặt... đạt được kết quả tốt hơn.

TÀI LIỆU THAM KHẢO

- [1] Chakrabarti (2016), "*A neural approach to blind motion deblurring*", 'In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)'.
- [2] A.N. Rajagopalan, Rama Chellappa (2014), "*Motion Deblurring Algorithms and Systems*", 'Indian Institute of Technology, University of Maryland, College Park'.
- [3] Generative Adversarial Network. <https://www.geeksforgeeks.org/generative-adversarial-network-gan>
- [4] Ian Goodfellow, Yoshua Bengio, Aaron Courville (2016), "*Deep Learning*", MIT Press.
- [5] K. Bousmalis, N. Silberman, D. Dohan, D. Erhan, and D. Krishnan (2016), "*Unsupervised Pixel-Level Domain Adaptation with Generative Adversarial Networks*", ArXiv e-prints.
- [6] Lucy richardson deconvolution. <https://stargazerslounge.com/topic/228147-lucy-richardson-deconvolution-so-what-is-it/>
- [7] M. Arjovsky, S. Chintala, and L. Bottou (Jan-2017), *Wasserstein GAN*, ArXiv e-prints.
- [8] Wang_Recent progress in image deblurring_1409.6838
- [9] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, Jiří Matas (2018), "*DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks*", 'Ukrainian Catholic University, Lviv, Ukraine'.
- [10] Point spread function. https://en.wikipedia.org/wiki/Point_spread_function
- [11] PyTorch. <https://pytorch.org>

Thông tin liên hệ: **Đào Thụy Ánh**

Điện thoại: 0913554138 - Email: dtanh@uneti.edu.vn

Khoa Công nghệ thông tin, Trường Đại học Kinh tế - Kỹ thuật Công nghiệp.