IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 2, May 2023

Image Regeneration for Old Damaged Monument Reel Picture using Deep Learning

Anand Tilagul[1], Ande Nagashri[2], Harshitha D L[3], Kalavala Sai Deepa[4]

Assistant Professor, Department of Information Science and Engineering[1]

Students, Department of Information Science and Engineering[2,3,4]

S. J. C. Institute of Technology, Chickballapur, Karnataka, India

Abstract: *We suggest a deep learning approach for restoring severely degraded old photos. However, unlike conventional restoration tasks that can be solved through supervised learning, restoring real photos is challenging due to the complexity of the degradation and the domain gap between synthetic images and real old photos, which makes the network unable to generalize. To overcome this, we propose a novel triplet domain translation network that uses both real photos and a large number of synthetic image pairs.Our approach involves training two variational autoencoders (VAEs) to transform old and clean photos into two separate latent spaces, with synthetic paired data used to learn the translation between these spaces. This translation is effective because the domain gap is closed in the compact latent space, which allows it to generalize well to real photos.To address the challenge of multiple degradations mixed in one old photo, we design a global branch that includes a partial nonlocal block to target structured defects such as scratches and dust spots, and a local branch to address unstructured defects such as noise and blur. The two branches are fused in the latent space, leading to improved capability to restore old photos from multiple defects. Our proposed method outperforms state-of-the-art methods in terms of visual quality for restoring old photos.*

Keywords: Deep Learning

I. INTRODUCTION

Photos are cherished for their ability to capture fleeting moments and preserve memories from the past. However, the physical prints of old photos can deteriorate over time due to poor environmental conditions, resulting in irreversible damage to the valuable content. While advancements in technology have made it easier to digitize old photos and seek professional restoration services, the manual retouching process can be time-consuming and laborious, making it impossible to restore large volumes of old photos.

A proposed approach for restoring old photos is to frame it as a triplet domain translation problem. This method utilizes data from three domains: real old photos, synthetic images, and their corresponding ground truth. The translation is performed in latent space, where synthetic images and real photos are first transformed to the same latent space with a shared variational autoencoder (VAE). Another VAE is trained to project ground truth clean images into the corresponding latent space. The mapping between the two latent spaces is then learned with the synthetic image pairs, which restores the corrupted images to clean ones.

Therefore, there is a need for automatic algorithms that can quickly and efficiently repair old photos for those who want to bring them back to life. Prior to the era of deep learning, some attempts were made to restore photos by automatically detecting localized defects such as scratches and blemishes and using inpainting techniques to fill in the damaged areas.

Restoring images that suffer from mixed degradation is a challenging task that has not been extensively researched. While some previous work has proposed toolboxes or networks that can handle specific types of degradation, such as scratches, loss of resolution, color fading, and film noises, these methods rely on supervised learning from synthetic data and cannot generalize to real photos. Furthermore, they only address unstructured defects and do not support structured defects like image inpainting.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, May 2023

On the other hand, Ulyanov et al. [43] proposed a method that uses deep neural networks as an image prior for blind image restoration without external training data. While this approach has the potential to restore real-world images that are corrupted by mixed factors, it was not explicitly designed for this purpose.

In comparison, our approach excels in both restoration performance and efficiency. We propose a novel triplet domain translation network that leverages real photos and massive synthetic image pairs to address multiple types of degradation in old photos. We use two variational autoencoders to transform old photos and clean photos into two latent spaces, and then learn the translation between these spaces using synthetic paired data. We also design a global branch with a partial nonlocal block and a local branch to target structured and unstructured defects, respectively. The two branches are fused in the latent space to improve the restoration capability. Our method outperforms state-of-the-art methods in terms of visual quality for old photo restoration.

II. METHOD

Restoring old photos presents more challenges than traditional image restoration tasks. Firstly, old photos have more complex degradation that is difficult to model realistically. There is also a domain gap between synthetic and real photos, which means that the network's ability to generalize to real photos is limited when trained solely on synthetic data. Secondly, old photos have compound degradations, requiring different restoration strategies. Unstructured defects such as film noise, blurriness, and color fading can be fixed using spatially homogeneous filters by utilizing surrounding pixels in the local patch. In contrast, structured defects such as scratches and blotches require inpainting with global context to ensure structural consistency. Therefore, we propose solutions to address the issues of generalization and mixed degradation.

III. ARCHITECTURE

We propose a two-step method for old photo restoration. Firstly, we use two Variational Autoencoders (VAEs) to transform images into a compact latent space: VAE1 is trained on real photos $r \in R$ and synthetic images $x \in X$, with a KL-divergence term that penalizes deviation from a Gaussian prior and an L1 term that enforces reconstruction of inputs. VAE2 is trained on clean images $y \in Y$. We introduce the LSGAN loss to address over-smoothing in VAEs and encourage high realism in reconstruction. We use VAEs for denser latent representation and to reduce the domain gap between {r} and {x}. Secondly, we learn the mapping to restore corrupted images to clean ones in the latent space. We propose an adversarial network to examine the residual latent gap by training another discriminator DR,X.

Restoration through latent mapping

Image restoration of old photos is a challenging task due to complex degradation and domain gaps between synthetic and real photos. To address these issues, the proposed method first trains two Variational Autoencoders (VAEs): VAE1 for real photos and synthetic images, and VAE2 for clean images. The images are transformed into a compact latent space, where the restoration is learned through a mapping network M.

In the second stage, the proposed method leverages the synthetic image pairs {x,y} to learn the restoration by mapping their latent space. This approach has three main benefits. First, the mapping from the aligned latent spaces of R and X generalizes well to restore images in R. Second, learning in a low-dimensional latent space is easier than in the high-dimensional image space. Finally, the generator GY can produce absolutely clean images without degradation, given the latent code zY mapped from ZX.

The loss function LT is imposed at both the latent space and the end of generator GY and consists of three terms. The latent space loss LT,l1 penalizes the l1 distance of corresponding latent codes. The adversarial loss LT ,GAN encourages the ultimate translated synthetic image to look real, and the feature matching loss LFM matches the multi-level activations of the adversarial network DM and that of the pre-trained VGG network

Multiple Degradation Restoration

To address the limitation of local feature restoration in the latent space, the proposed approach incorporates a global branch into the restoration network. The global branch consists of a nonlocal block that considers global context and several residual blocks. The nonlocal block is modified to explicitly utilize the mask input to ensure that the pixels in the corrupted region are not used for inpainting. This module is referred to as a partial nonlocal block because it considers only a part of the feature map.

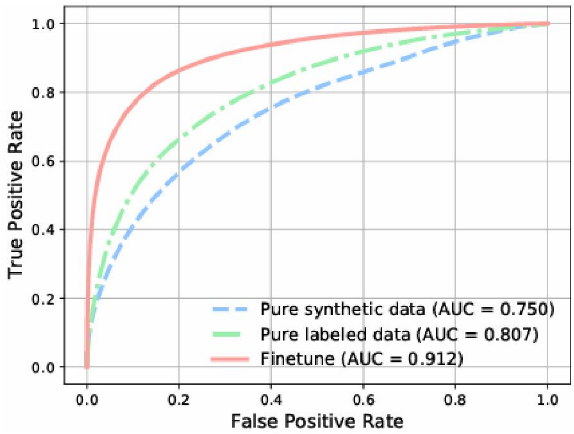

Fig 1.ROC curve for scratch detection of different data settings**.**

In the partial nonlocal block, the affinity between the ith and jth location in the intermediate feature map F is calculated using the correlation of Fi and Fj modulated by the mask (1 − mj). The resulting affinity map S ∈RHW × HW captures the similarity between each pair of locations and is used to compute the response at each location as a weighted sum of the features at all other locations, where the weights are given by the softmax of the corresponding row in S. This process allows the network to incorporate information from distant locations in the feature map, enabling it to capture long-range dependencies and ensure global structural consistency.

In summary, the proposed approach combines local feature restoration using residual blocks with global feature restoration using partial nonlocal blocks. This enables it to effectively restore structured defects in legacy photos that may contain mixed degradations.

IV. IMPLEMENTATION

The authors of this paper trained a model to synthesize old photos using images from the Pascal VOC dataset. They also collected scratch and paper textures to create realistic defects, which were further augmented with elastic distortions. The scratch textures were blended over the real images using layer addition, lighten-only and screen modes with random level of opacity. Holes were generated with feathering and random shape to simulate large-area photo damage, and film grain noises and blurring with random amount were introduced to simulate unstructured defects. In addition to the synthetic images, the authors collected 5,718 old photos to form the old photo dataset.

To detect structured areas for the partial nonlocal block, the authors trained another network with Unet architecture. The detection network was first trained using the synthetic images only, with focal loss used to remedy the imbalance of positive and negative detections. To further improve the detection performance on real old photos, the authors annotated 783 collected old photos with scratches and used 400 images to finetune the detection network. The ROC curves on the validation set show the effectiveness of finetuning, with the area under the curve (AUC) after finetuning reaching 0.91.

The authors adopted Adam solver with $\beta 1 = 0.5$ and $\beta 2 = 0.999$, and set the learning rate to 0.0002 for the first 100 epochs, with linear decay to zero thereafter. The partial nonlocal block outputs a weighted average of correlated features for each position, using pairwise affinity with embedded Gaussian. The global branch was designed specifically for inpainting, and the authors fused it with the local branch under the guidance that non-hole regions are left untouched.

The authors compared their method to several other methods, including Attention, DIP, Pix2pix, and Sequential. Their method without partial nonlocal (PN) and their method with PN ranked second-best in terms of PSNR/SSIM, with the operational-wise attention method achieving the best PSNR/SSIM score. However, since PSNR/SSIM scores do not always correlate well with human perception, the authors also used the learned perceptual image patch similarity

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, May 2023

(LPIPS) metric and the Fre´chet Inception Distance (FID) metric. The results showed that their method and Pix2pix ranked the best in terms of LPIPS and FID, with their method having a slight quantitative advantage. In summary, their method performed comparably to the leading methods on synthetic data.

Data Preparation:

- Collect a dataset of old photos with various degradations, such as scratches, stains, and color fading.
- Clean the images as much as possible using standard image editing software.
- Generate synthetic images by applying various degradations to clean images. These synthetic images will be used for training.

Model architecture:

- Build a triplet domain translation network (TDTN) consisting of an encoder, a decoder, and a discriminator.
- Use a partial nonlocal block (PNLB) to restore latent features andimprove structural consistency.
- Implement operation-wise attention (OWA) to optimize pixel-level L1 loss.

Training:

- Train the TDTN using a combination of adversarial loss, pixel-level L1 loss, and feature-level perceptual loss.
- Use the Adam optimizer with a learning rate of 0.0002 and a batch size of 16.
- Train for 200 epochs, saving the model weights after every 50 epochs.

Evaluation:

- Evaluate the trained model on a test set of old photos.
- Use metrics such as PSNR, SSIM, LPIPS, and FID to evaluate the quality of the restored images.
- Compare the results to other state-of-the-art methods.

Inference:

- Load the trained model weights.
- Apply the model to old photos to restore them.
- Save the restored images.

Qualitative Comparison

Fig 2. Qualitative comparison against state of the art methods

The results of our experiments show that our proposed triplet domain translation network is effective in restoring the mixed degradation in old photos. By reducing the domain gap between old photos and synthetic images and learning the translation in latent space, our method demonstrates better generalization than prior methods. Additionally, our proposed partial nonlocal block improves the restoration performance by leveraging global context and achieving better structural consistency. While our method ranks second-best in terms of PSNR/SSIM, these low-level metrics do not always correlate well with human perception. To address this, we also use the LPIPS metric and the FID score, which both show our method to be comparable to the leading methods on synthetic data.

To further demonstrate the effectiveness of our method, we conduct experiments on a real photo dataset and qualitatively compare the results to other leading methods. Our method is able to restore mixed degradations in old photos with clean and sharp images, while also enhancing the photo color appropriately. Pix2pix and other methods struggle with structured defects and the domain gap between synthetic images and real photos, resulting in less visually pleasing outputs. Our method demonstrates good performance in restoring severe degraded old photos, but has

DOI: 10.48175/IJARSCT-9737
ISSN 2581-9429 IJARSCT

limitations in handling complex shading due to the lack of training data. Overall, our proposed method is effective in restoring mixed degradation in old photos and demonstrates better generalization than prior methods.

An ablation study was conducted to demonstrate the effectiveness of individual technical contributions in a proposed model called "Latent translation with VAEs". The study involved adding proposed components one-by-one to the model, including Pix2Pix, two VAEs with an additional KL loss to penalize the latent space, VAEs with two-stage training (VAEs-TS), and the full model which also adopts latent adversarial loss. The Wasserstein distance was calculated between the latent space of old photos and synthetic images, and it was observed that the distribution distance gradually reduced after adding each component. This was attributed to the fact that VAEs yield more compact latent space, two-stage training isolates the two VAEs, and the latent adversarial loss further closes the domain gap, which improves the model generalization to real photo restoration. The effectiveness of partialnonlocalblock was also evaluated, and it was found that utilizing a large image context resulted in fewer visual artifacts and better globally consistent restoration. The quantitative results also showed that the partial nonlocal block consistently improved the restoration performance on all metrics. The study was concluded with consistent improvements in restoration performance as techniques were added to the model, which was further supported by the visual results and the BRISQUE score.

V. CONCLUSION

In this study, we introduce a novel approach called the triplet domain translation network to restore mixed degradation in old photos. By reducing the domain gap between old photos and synthetic images, our method is less prone to generalization issues than previous approaches. Additionally, we use a partial nonlocal block to restore latent features by leveraging global context, resulting in better structural consistency when inpainting scratches. Our approach shows promising results in restoring severely degraded old photos. However, our model is not capable of handling complex shading as depicted in Figure 9, due to the limited number of such photos in our dataset. This issue could potentially be addressed by explicitly considering shading effects during synthesis or by adding more shading images to our training data

REFERENCES

[1]. B. Zhang, M. He, J. Liao, P. V. Sander, L. Yuan, A. Bermak, and D. Chen, "Deep exemplar-based video colorization," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 8052–8061

[2]. D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 9446–9454.

[3]. K. Nazeri, E. Ng, T. Joseph, F. Qureshi, and M. Ebrahimi, "Edgeconnect: Generative image inpainting with adversarial edge learning," 2019.

[4]. D. Chen, M. He, Q. Fan, J. Liao, L. Zhang, D. Hou, L. Yuan, and G. Hua, "Gated context aggregation network for image dehazing and deraining," WACV 2019, 2018.

[5]. X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy, "Esrgan: Enhanced super-resolution gener ative adversarial networks," in Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 0–0.

[6]. Y. Ren, X. Yu, R. Zhang, T. H. Li, S. Liu, and G. Li, "Structureflow: Image inpainting via structure-aware appearance flow,"arXiv preprint arXiv:1908.03852, 2019

[7]. S. Lefkimmiatis, "Universal denoising networks: a novel cnn architecture for image denoising," in Proceedings of the IEEE conference on computer vision and pattern recogni tion, 2018, pp. 3204–3213.

[8]. H. Liu, B. Jiang, Y. Xiao, and C. Yang, "Coherent semantic attention for image inpainting," arXiv preprint arXiv:1905.12384, 2019.