ISBN: 978-981-94-4198-3
2025 the 15th International Workshop on Computer Science and Engineering (WCSE 2025)
doi: 10.18178/wcse.2025.06.029

Latent Denoising Diffusion Models for Colorization of Historical Portraits Photographs

Jincheng Peng[1,2+], Zhenyu Guo[1,2] , Guoyue Chen[3], Ming Zhou [2]

[1] Beijing Key Laboratory of Millimeter Wave and Terahertz Techniques, Beijing Institute of Technology, Beijing, China

[2] Yangtze Delta Region Academy of Beijing Institute of Technology, Jiaxing, Zhejiang, China

[3]Akita Prefectural University, Department of Information and Computer Science, Akita, Japan

Abstract. In the early stages of photographic technology, due to technical limitations, most photographic images were presented in grayscale. Early color photographs, when exposed to harsh environmental conditions over time, would gradually degrade, leading to the permanent loss of valuable image content. In recent years, with the development of computer-based digital imaging and scanning technologies, it has become possible to restore the colors of historical photographs. However, the color restoration of historical photos requires professional expertise and is a labor-intensive, expensive, and time-consuming task. Therefore, developing automated network models for old photo colorization and restoration holds significant importance. The latest Denoising Diffusion Probabilistic Models (DDPMs), based on probabilistic modeling, analyze the diffusion relationships between pixels in an image while preserving the image's details and structure. These models have shown remarkable performance in various emerging fields of computer vision. This paper proposes an latent denoising diffusion probabilistic network-based automatic colorization method by studying deep learning-based colorization techniques and DDPMs. A VQVAE-based latent variable model is introduced into the forward diffusion process and reverse inference of the DDPM, serving as a transitional bridge during the reconstruction process. This model is capable of converting black-and-white images into color images. Experimental results demonstrate that the latent denoising diffusion probabilistic model achieves optimal colorization performance on facial image datasets. Finally, this colorization technique is extended to the colorization of real historical portraits

Keywords: historical portraits photographs, latent denoising diffusion models, image colorization, VQVAE-model.

1. Introduction

Historical photograph images, as important carriers of digital information in the modern era, serve as an objective reflection of events and are one of the key ways for humanity to understand and perceive the world. Therefore, historical photograph images are especially important in the protection and exploration of cultural heritage. If these historical photos, particularly personal portrait photos, can be colorized, they not only provide a rich visual sensory experience but also offer a more immersive and intimate experience for contemporary audiences[1]. However, during the early stages of photographic technology, photos were typically captured as grayscale images. color photographs from decades ago may have suffered from environmental degradation during preservation, such as cracks, blurring, and color fading. Therefore, restoring the original colors of old photographs is essential. Image processing algorithms based on computer vision techniques are particularly appealing for revitalizing historical images and bringing them back to life.

Early computer colorization methods largely involved manually using color brushes to label specific areas and extend color to local pixel blocks, treating the problem as a local image optimization task. T. Welsh[2] proposed an image color transfer method, adjusting the original image and using pixel brightness and neighborhood standard deviations to find the most matching pixel in a reference image, thereby transferring its color. This method also incorporated image segmentation to improve search efficiency but suffered from miscoloring in pixels with the same brightness but different color values.

+ Corresponding author. Tel.: + 86-17828090682;.
E-mail address: popabe@bit.edu.cn.

In recent years, with the rapid development of deep learning, researchers have leveraged the powerful representation ability of convolutional neural networks (CNNs) to solve various low-level image restoration problems, including image colorization. Cheng et al.[3] proposed a neural network-based grayscale image colorization method that extracts color features from grayscale images using neural networks and trains the network with L2 loss, combined with bilateral filtering to improve results. However, Cheng's method used limited training data and a simple network structure, which significantly restricted the types of images it could handle and the colorization results. Larsson et al.[4] integrated the VGG neural network as a semantic and localization feature extractor in the colorization system, predicting color histograms for each image location. Although these networks can predict color regions based on color distribution, relying solely on CNNs to extract color features leads to the loss of semantic information, resulting in misclassification of domains and unsatisfactory colorization effects.

In recent years, with the rapid development of Generative Adversarial Networks (GANs) composed of generators and discriminators, several studies have proposed GAN-based colorization methods. For instance, a conditional GAN-based colorization method predicts the color distribution of each domain by having the discriminator assess the loss between generated and real images. Deshpande et al.[5] embedded VAE-encoded color modeling into the Gaussian distribution of grayscale images and sampled different colors from the Gaussian distribution. Cao et al.[6] used an unsupervised colorization network based on CGAN to color images, where the generator didn't adopt an encoder-decoder structure but instead introduced random optimization noise into various layers of the fully convolutional network, achieving more realistic results and greater diversity in generated images. However, the introduction of strong noise in each layer increased the randomness, leading to uncontrollable colorization results and poor image generation quality.

Although these methods have achieved satisfactory colorization effects in general image tasks, the lack of large-scale datasets for old photos makes it particularly challenging to process such restoration tasks. Additionally, GANs must satisfy strict Nash equilibrium conditions to avoid issues like gradient explosion and mode collapse. To fundamentally solve the shortcomings of GAN models, Jonathan Ho et al.[7] improved the mathematical computation method of existing diffusion probabilistic models and proposed the Denoising Diffusion Probabilistic Model (DDPM). By defining a Markov chain for the diffusion steps and using controlled noise/denoising steps, DDPMs transform input data and predict new data distributions. Research [8] has shown that the latest DDPMs excel at generating realistic image texture features and have outperformed traditional GANs in image synthesis, image translation, image restoration, image colorization, and other computer vision research areas, leading image processing research into a new direction.

In this study, based on research into existing denoising diffusion probabilistic networks and deep learning colorization algorithms, we have designed a novel end-to-end latent variable denoising diffusion probabilistic colorization model for the color restoration of large-scale historical photos. As color feature mapping in the image's color space is required, we employ a conditional denoising diffusion probabilistic network[9], incorporating a VQ-VAE module that introduces latent variables as a transitional bridge during the colorization process. This model extracts color features of the characters and background from training datasets via forward noise addition. In the reverse inference generation process, starting from random Gaussian noise, the model combines the black-and-white image condition with the Gaussian noise, iteratively refining the reverse process to convert random Gaussian noise into a realistic color image. Using this trained conditional model, we reconstruct the colors of black-and-white images based on facial features from the original training set, thereby colorizing the historical portrait. The VQ-VAE layer we added effectively fills in missing information after colorization and constrains the colorization prediction space, reducing the model's computational cost without affecting the colorization results. This allows for the rapid generation of high-quality images from black-and-white photos while maintaining the style and content consistency with the original input image, making the generated image more natural. Experimental results show that our improved latent denoising diffusion probabilistic network method achieves a more realistic and natural colorization effect on the publicly available CelebFaces Attribute dataset. The colorized images are richer and more realistic, and the trained model was subsequently validated on real-world black-and-white photos taken decades ago, yielding promising results. Future research will build upon this to achieve colorization of video content.

2. Methods

2.1. Denoising Diffusion Probabilistic Model(DDPM)

The Denoising Diffusion Probabilistic Model (DDPM) utilizes a parameterized Markov chain trained via variational inference, consisting of a forward noise diffusion Markov process and a reverse denoising Markov process. It learns a Markov chain such that, after T steps of noise addition, a real image becomes indistinguishable from a Gaussian noise distribution. Conversely, the reverse denoising process, optimized through a U-Net based neural network, learns to minimize the distance between the generated data distribution and the true data distribution. By applying a theoretically grounded step-by-step denoising process over T steps, the model can generate images that closely resemble the true data distribution.

As shown in Fig.1, the forward process of the denoising diffusion model, also known as the diffusion process, refers to the gradual addition of Gaussian noise to image data until it becomes pure Gaussian noise. As T increases, the $X_T$data progressively transforms into a Gaussian noise distribution. Over T time steps, the original image $X_0$is converted into standard Gaussian white noise X_T~N(0, 1).

$$q(x_t \mid x_{t-1}) = N(x_t; \sqrt{1-\overline{\beta}_t}\, x_{t-1}, \beta_t I) \tag{1}$$

The forward process can be defined as:

$$q(x_{1-T}|x_0)=\prod_{t=1}^{T} q(x_t|x_{t-1}) \tag{2}$$

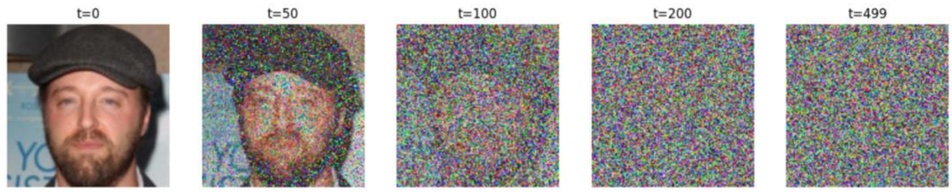

Fig. 1: The forward noise addition process for the denoising diffusion probabilistic model

The reverse process is the inverse of the forward process. Using the trained model to define the parameters of the theoretical formulas, noise is gradually reduced from the image, resulting in a denoised image. By knowing the true data distribution $q(x_{t-1}|x_t)$ at each step of the denoising process and continuously sampling through reverse iteration, Gaussian noise is progressively removed until a complete image $q(x_0)$ is generated. However, directly obtaining distribution$q(x_{t-1}|x_t)$requires access to the entire training dataset, making it difficult to compute exactly.

Therefore, the denoising diffusion model approximates this distribution by constructing a neural network parameterized by θ It is assumed that the distribution $p_\theta(x_{t-1}|x_t)$, representing the reverse generation process, follows a Gaussian distribution, where the mean μ_θ and variance σ_θ are functions of x_t and time step t, leading to the following expression:

$$p_\theta(x_{t-1}|x_t)=N(x_t; \mu_\theta(x_t,t), \sigma_\theta(x_t,t)) \tag{3}$$

By combining Bayes' theorem

$$p_0(x_{t-1}|x_t)=N(x_{t-1}; \mu_\theta(x_t,t), \Sigma_\theta(x_t,t)) \tag{4}$$

Here, μ_θ and σ_θ represent the mean and variance.

2.2. Latent-Variable Denoising Diffusion Probabilistic Colorization Network Architecture

The network structure of the latent variable denoising diffusion probabilistic colorization model is shown in Fig.2. In the generation process of the model, the key idea is to use the latent variable features generated by the VQ-VAE encoder to learn the mapping between the latent code Z of a color image and the conditional grayscale image features. The reverse inference process approximates the generation of a conditionally constrained color image $I_L$based on an input grayscale image. In this latent variable denoising diffusion probabilistic model, a VQ-VAE encoder is used to encode and map the color image data distribution into a low-dimensional latent space, producing a latent variable z, also referred to as the latent code. The forward process of the DDPM (Denoising Diffusion Probabilistic Model) then adds noise to this latent code.

During the reverse inference process, the model uses the input grayscale image as a guiding condition and combines it with random Gaussian noise. Through iterative refinement in the reverse diffusion process, the random Gaussian noise is gradually transformed into a distribution similar to the latent variable distribution of the color image. Finally, the VQ-VAE decoder module is used to reconstruct the latent variable into a distribution close to the original generated image, thus achieving conditional color image generation.

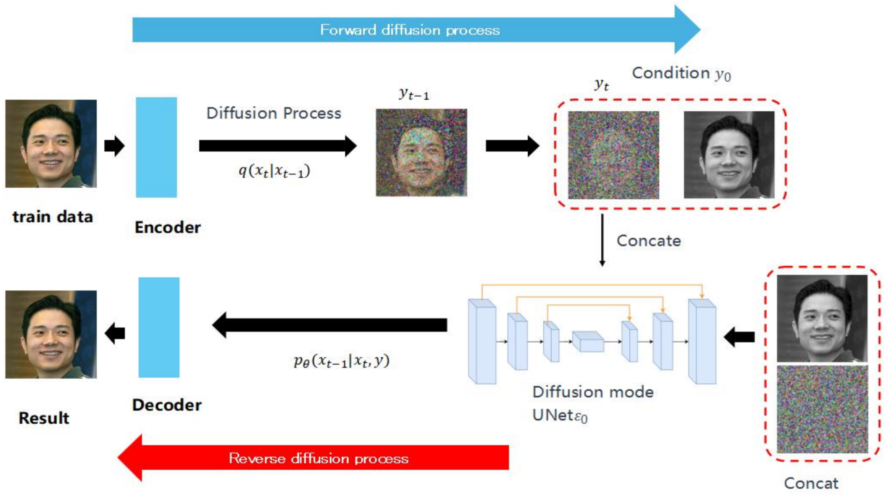

Fig. 2: Architecture of the latent DDPM colorization network

For training, our colorization model uses the L1 loss function as the objective metric to regularize the distance between the generated color image and the ground truth. This constrains the generated image to retain latent information such as texture, shape, and style, similar to the real image.

3. Experiments

3.1. Dataset and Experimental Environment

We adopt the CelebFaces Attributes (CelebA) dataset [10] as the training set for restoring and colorization historical portrait photographs. CelebA contains face images of thousands of celebrities and comes with rich attribute annotations, such as age, gender, ethnicity, facial expression, hair color, and background. Because the images are drawn from publicly available photos on the internet, the collection is large and exhibits diverse, complex color characteristics, making it well-suited for realistic color restoration. Given the computational limits of our lab, we selected 9,000 face images from the original dataset to train our colorization model. Each image is stored in PNG format at a resolution of 256 × 256 pixels. The experimental environment used for this study includes Google Colab with an NVIDIA Tesla T4 GPU. The deep learning framework is primarily based on PyTorch, and the denoising diffusion probabilistic colorization network model is built using the Hugging Face diffusers library. The training parameters are set as follows: batch size is 15, initial learning rate is 0.0002, and the optimization algorithm is Adam. The noise scheduling for the diffusion model uses a cosine scheduling timetable for β_1 and β_2, with the training set to 300 epochs and the total number of time steps for the generative model set to 1,000.

3.2. Comparative Experiments

To evaluate the color-restoration performance of our latent DDPM colorization network, we conducted a comparative study against several publicly available colorization models. As illustrated in Fig. 3, Zhang et al.'s "Colorful Image Colorization" network (ECCV 2016) tends to underestimate color distributions, producing overly muted results with an overall yellow-gray cast.

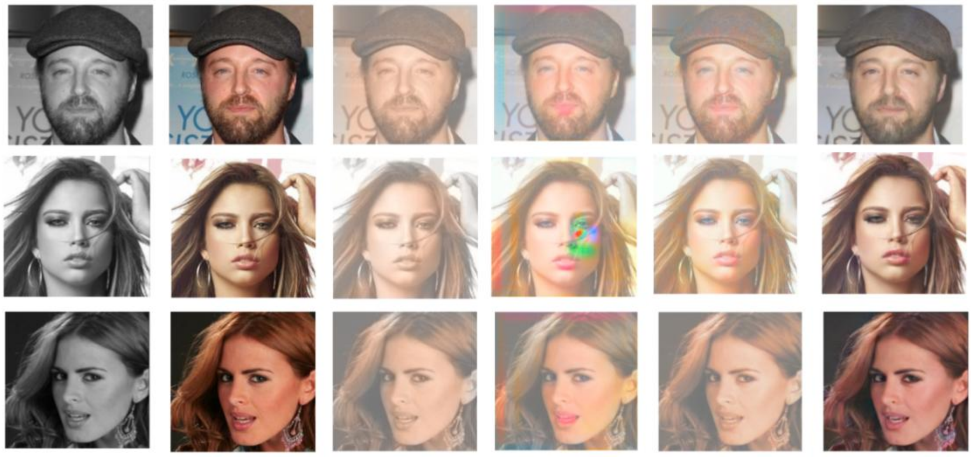

Fig. 3: Visual comparison of different Colorization methods on the CelebFaces Attributes dataset

Pix2Pix and SNGAN both GAN-based approaches perform slightly better than Zhang's method, yet each shows clear drawbacks. Pix2Pix introduces conspicuous blotches and artifacts; for example, the person's teeth in the last row, which should appear white, are incorrectly tinted red. Such issues stem from the need to balance the generator and discriminator during GAN training, which can lead to gradient explosions and mode collapse. SNGAN mitigates these problems by incorporating Spectral Normalization, but its generated images still deviate noticeably from genuine color photographs.

Finally, to assess whether our trained model can colorize real-world portraits taken several decades ago, we selected several images from the Vintage Photo Restoration Collection dataset on Kaggle [11] as test samples. For optimum consistency, each photograph was first converted to grayscale before being processed by our trained colorization network, which then performed inverse color inference. As shown in Fig. 4, the model successfully reconstructs many fine details and recovers plausible, lifelike colors from the original black-and-white photos. Nonetheless, the right-hand example exhibits less satisfactory color prediction most likely because the color-training corpus is still too small, or because some model hyper-parameters remain sub-optimal. Future work will address these limitations and refine the approach accordingly.

Fig. 4: Validation results on vintage black-and-white photographs taken decades ago

4. Conclusions

This paper introduces an latent DDPM based method for colorization historical portrait photographs. By modifying the Denoising Diffusion Probabilistic Model (DDPM), we incorporate a VQ-VAE module into the original conditional DDPM and introduce latent variables that serve as a bridge during the intermediate colorization stages. Experiments demonstrate that the proposed network produces sharp, natural images that more closely match the reference grayscale inputs. Further analysis explored how the number of training epochs affects color diversity, and the final model was applied to genuine black-and-white photographs taken several decades ago, successfully synthesizing plausible color versions. These findings suggest that the proposed deep-learning approach can be widely deployed for restoring old photographs and classic films.

Nonetheless, there remains room for improvement; future work will refine the model to achieve even better color fidelity and extend the technique to automatic video colorization.

5. References

[1] Wan Z, Zhang B, Chen D, et al. Bringing old photos back to life. proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020: 2747-2757.

[2] Welsh T, Ashikhmin M, Mueller K. Transferring color to greyscale images. Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques, 2002: 277-280.

[3] Cheng Z, Yang Q, Sheng B. Deep colorization. Proceedings of the IEEE International Conference on Computer Vision, 2015: 415-423.

[4] Larsson G, Maire M, Shakhnarovich G. Learning representations for automatic colorization. European Conference on Computer Vision, 2016: 577-593.

[5] Deshpande A, Lu J, Yeh M C, et al. Learning diverse image colorization. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 6837-6845.

[6] Cao Y, Zhou Z, Zhang W, et al. Unsupervised diverse colorization via generative adversarial networks. Joint European Conference on Machine Learning and Knowledge Discovery in Databases, 2017: 151-166.

[7] Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems, 2020, 33: 6840-6851.

[8] Dhariwal P, Nichol A. Diffusion models beat GANs on image synthesis. Advances in Neural Information Processing Systems, 2021, 34: 8780-8794.

[9] Saharia C, Chan W, Chang H, et al. Palette: Image-to-image diffusion models. ACM SIGGRAPH 2022 Conference Proceedings, 2022: 1-10.

[10] CelebFaces Attributes (CelebA) Dataset. https://www.kaggle.com/datasets/jessicali9530/celeba-dataset.

[11] Vintage Photo Restoration Collection. https://www.kaggle.com/datasets/marcinrutecki/old-photos.