영상 컬러화를 위한 ColorGAN

ColorGAN: Generative Adversarial Network based Image Colorization

함자 샤피크, 이범식
Hamza Shafiq and Bumshik Lee

조선대학교 정보통신공학과
E-mail: hamzashafiq@chosun.kr, bslee@chosun.ac.kr

요 약

We propose an adversarial learning based approach for diverse image colorization using Encoder-Generator architecture with added noise in latent space. In the proposed method, the noise is generated from a normal distribution and then processed through a series of convolution layers before the latent space feature map. Further, information from encoder layers is passed to the generator for improved image reconstruction. In addition, we employ perceptual loss in conjunction with adversarial loss to improve visual quality. The proposed method outperforms the state-of-the-art approaches.

키워드 : Image Colorization, Deep Learning, Computer Graphics, Generative Adversarial Network

1. Introduction

Image colorization is the task of restoring the colors of a given grayscale image. It offers a wide range of applications, including legacy photo/video colorization, cartoon colorization, and false color detection. Colorization is an ill-posed and highly ambiguous problem as different colors can be assigned to a single object (e.g. a shirt can have different colors, plant leaves can be green, yellow, or brown, etc.).

Classical colorization methods require additional reference color images. However, the process of retrieving reference color images requires manual effort. Recently, deep learning based colorization methods have become more popular because it removes the user interventions and colorizes the images in a fully automatic manner. In this paper, we propose a fully automatic image colorization approach named ColorGAN based on an adversarial learning approach. We use encoder-generator architecture with random noise added to latent space after being processed from several convolution layers to get vivid and diverse results. This manipulation in the latent space helps to get vivid and diverse colors.

The feature map of the noise, after the series of convolution layers, is compared with the feature map of the ground truth color image, which is passed through a pre-trained VGG network, using the L1 distance loss.

2. Methodology

The colorization technique presented in this article uses encoder-generator architecture as shown in Fig 1. In this section, we first describe the overall architecture of the proposed approach. Then we provide the details of the proposed objective function.

2.1 Detailed Model Architecture

The proposed GAN architecture consists of three parts: encoder, generator, and discriminator. The encoder is responsible for getting a feature map of the input image and the generator reconstructs the image with color information.

Given a grayscale image L, the encoder learns the mapping of the grayscale image to a feature map. Instead of the feature vector, we use a feature map to improve the reconstruction process. A learnable feature map from a normal distribution is added to this feature map so that the output image can have more vividness. The *Lab* color space is used instead of RGB where *L* is the luminance channel and *(a, b)* are the chrominance channels. The latent space feature map is fed to the generator for reconstructing the image with *ab* channels which are then combined with the *L* channel to get the output color image. The *ab* channel image is fed to the discriminator along with the *L* channel to calculate

감사의 글 : 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 ICT 명품인재양성 사업의 연구결과로 수행되었음 (IITP-2022-0-01821)

the score based on the generated image. The discriminator structure is based on the Markovian discriminator architecture (PatchGAN)[1], which focuses on capturing high-frequency components using local patches. Unlike other discriminators, it classifies each patch instead of classifying the whole image.

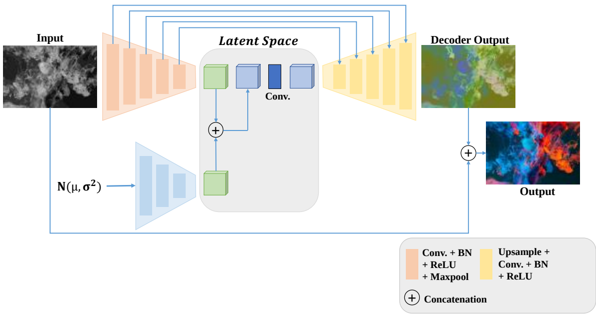

Fig 1. Generator Architecture of proposed Network

2.2 Objective Function

The proposed objective function is defined by

$$\mathcal{L} = \lambda_g \mathcal{L}_g + \lambda_p \mathcal{L}_p + \lambda_{L1} \mathcal{L}_1 + \lambda_s \mathcal{L}_s \quad (1)$$

The first term $\mathcal{L}_g$ denotes adversarial Wasserstein (WGAN) loss [2], to avoid the vanishing gradient problem and achieves stable training for GAN. The term $\mathcal{L}_p$ denotes the perceptual loss which is the L1 distance of features extracted by the pre-trained VGG network. $\mathcal{L}_1$ is the conventional L1 loss for the output colorized image. $\mathcal{L}_s$ is the comparison of the random normal distribution feature map and ground truth image feature map and is defined as

$$\mathcal{L}_s = E\left|\left|G_f\left(N(\mu,\sigma^2)\right) - VGG(y)\right|\right|_1 \quad (2)$$

where, $N(\mu,\sigma^2)$ is the random normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 0.1$. The terms $\lambda_g, \lambda_p, \lambda_{L1,}$ and λ_s are fixed empirically and set to $\{\lambda_g, \lambda_p, \lambda_{L1,} \lambda_s\} = \{0.1, 100, 10, 1\}$.

3. Experiment and Results

3.1 Implementation details

For training, the PASCAL VOC dataset is used, which has 17125 images. 15413 images are used for training while 1712 images are used for testing. The images are rescaled to 256x256 and trained for 300 epochs. The learning rate was set to 1×10^{-4} and 2×10^{-4} for the generator and discriminator respectively. We use a batch size of 16 and an Adam optimizer with $\beta_1 = 0.5$ and $\beta_2 = 0.999$.

3.2 Quantitative metrics

We used PSNR (peak signal-to-noise ratio) and SSIM (structural similarity index) metrics for evaluating the performance of our model.

3.3 Results

Colorization results of the proposed ColorGAN are shown in Fig 2. We compare our results with ChromaGAN [3] and Coltran [4]. Fig 2 shows that the proposed method gives more natural colors and outperforms other methods. ChromaGAN[3] and Coltan[4] suffer from desaturated and unnatural colorization. We further evaluate our method with PSNR and SSIM. Our method outperforms other methods in quantitative metrics as shown in Table 1.

Input | Chroma GAN[3] | ColTran [4] | Proposed | Groundtruth

Fig 2. Qualitative comparison with other colorization methods.

4. Conclusion

We proposed ColorGAN, a fully automatic colorization method using an encoder-generator architecture. We encode the input image to the feature map and add a learnable feature map from the normal distribution in the encoded feature map for diverse and vivid colorization. We evaluate our method and demonstrate that ColorGAN outperforms existing state-of-the-art methods.

Table 1. Quantitative comparison of the proposed method with other colorization methods

	Chroma GAN [3]	ColTran [4]	ColorGAN
PSNR	24.154	23.839	**24.333**
SSIM	0.926	0.868	**0.938**

References

[1] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. “Image-to-image translation with conditional adversarial networks.” In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1125–1134, 2017.

[2] M. Arjovsky, S. Chintala, and L. Bottou. “Wasserstein generative adversarial networks.” In International Conference on Machine Learning, pages 214–223, 2017.

[3] P. Vitoria, L. Raad, C. Ballester Chromagan: “Adversarial picture colorization with thesemantic class distribution.” Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 2445–2454 (2020)

[4] M. Kumar, D. Weissenborn, Kalchbrenner, “Colorization transformer.” International Conference on Learning Representations (2021)