Image Colorization Based on VAE-GAN and the Mixture Density Model

Jifeng Sun, Kai Wang and Yibin Lin

School of Electronic and Information, South China University of Technology, Guangzhou, China
Email: ecjfsun@scut.edu.cn

Abstract. A diverse image colorization method based on VAE-GAN and mixture density network is proposed by this paper. The VAE-GAN is used to learn the color representation of images, while the mixture density network is to model multimodal distribution of color with learned representation, so that we can draw samples from the distribution and decode them with VAE-GAN to achieve diverse colorization. In comparison with vanilla VAE, the VAE-GAN is trained with adversarial loss and perceptual loss on the discriminator, which ensures better color representation learned by the model and more consistent color result.

Keywords. Diverse colorization; variational auto-encoder; generative adversarial networks; mixture density network; deep learning.

1. Introduction

Color is very important feature of the images, because the vision of human is sensitive to the color which provides very plentiful experience. For Example, in the case of the weather, the color of the scene represents whether it is sunning, of cloudy or rainy. In the case of the organs inside the human body, the color of the organ in the image represents the organ is heart, lungs or blood vessel. On the other hands, the color of in the scene of the countryside or downtown represents some part is wheat field, highway, house, apartment or office building.

The colorization of gray image rise from two kinds of purposes: (a) the colorization of early gray movies, gray documentary films, gray TV dramas or gray photos produced 50 years ago; (b) the colorization of night scene vision and MRI/CT medical image.

The image colorization based on AI, especially deep learning, improves the effectiveness compared with the man-made way.

There are three kinds of image colorization scheme, i.e., the scheme based on the reference images [1], the scheme based on the color label [2, 3] and the scheme without reference image [4-8]. Reinhard et al. proposed the image color transfer algorithm based on statistical characteristics.

On the other hand, to do the image colorization faces two basic problems, i.e., information dimension increase from the gray image to the color image, and the color ambiguity of colorization.

The scheme based on the reference images is an early method which is effective to solve the problem of the color ambiguity to some extent.

The scheme based on the color label [2-3] provides some location information for the colorization, and needs no reference image.
The scheme without reference image [4-8] is a colorization method based on the maximum a posteriori probability estimation of the gray images and the Gauss-Markov random field of the color image.

2. Method Based on VAE-GAN and the Mixture Density Model

In the case of no reference information, multiple different colorization results will be got for a single gray image. But the colorization result of most deep learning algorithms is single, which limits the application of those algorithms. The common colorization schemes based on deep learning, e. g. those based on GAN or auto-regression, have some shortages, such as instability, no convergence in the iteration and slowness.

In order to improve the diversity and uniformity of the colorization, colorization based on VAE-GAN and mixture density model is used as figure 1.

![Figure 1](image)

**Figure 1.** The structure of the colorization method based on VAE-GAN and the mixture density model.

2.1. The Principle of VAE-GAN and the Mixture Density Network

Variational Auto-encoder. Variational Auto-Encoder (VAE) is proposed by Kingma [9] in 2013, in which is an auto-coding of the unsupervised representation. For \( X = \{x_1, x_2, \ldots, x_N\} \), each data \( x_i \) can be obtained as following two-step algorithm:

- **Sampling** \( z_i \) from prior distribution \( p(z) \)
- **Sampling** \( x_i \) from conditional distribution \( p_o(x \mid z) \)

where conditional distribution \( p_o(x \mid z) \) come from parametric distribution family, \( \theta \) is unknown parameter, in the case of generative model, the estimation of \( \theta \) may provide new sample according to above two-step algorithm.

In order to estimate the parameter, the Log likelihood function of the dataset can be defined as follows:

\[
\log p(X) = \sum_{i=1}^{N} \log p(x_i) = \sum_{i=1}^{N} \log \int p(z) p(x_i \mid z) dz
\]

But when the dimension of latent variable \( z \) is too high, the Log likelihood function of the dataset is difficult to be solved, which brings big difficulty for estimating the unknown parameter \( \theta \).

The idea of VAE is to introduce inferential network \( q_o(z \mid x) \) to be approximately equal to posterior conditional probability \( p_o(z \mid x) \), the non-negativity of Kullback-Leibler divergence is helpful to achieve lower bound.
\[ L(\theta, \varphi; x) = \log p_\theta(x) - D_{KL} \left( q_\varphi(z \mid x) \parallel p_\theta(z \mid x) \right) \]
\[ = -D_{KL} \left( q_\varphi(z \mid x) \parallel p_\theta(z) \right) + E_{q_\varphi(z \mid x)} \left[ \log p_\theta(x \mid z) \right] \]
\[ = -L_{VAE} \]

where \( L(\theta, \varphi; x) \) is Evidence Lower Bound, (ELBO). Therefore variational lower bound can be maximized to be approximately equal to the real likelihood function. In the case of optimum, Evidence Lower Bound is equal to Log likelihood, inferential network \( q_\varphi(z \mid x) \) can be used to get posterior conditional probability \( p_\theta(z \mid x) \).

The variational lower bound be divided into binomical, the first is the regular term of an auto-coding, the second is the negative reconstruction of the auto-encoder. Because of inferring Evidence Lower Bound, so this kind is called variational auto-encoder. VAE is shown as figure 2.

**Figure 2.** The structure of VAE model.

In practice, suppose distribution function \( q_\varphi(z \mid x) \) and \( p_\theta(z \mid x) \) all obey Gaussian distribution, and the reconstruction loss \( E_{q_\varphi(z \mid x)} \left[ \log p_\theta(x \mid z) \right] \) can be optimized by the reparametrization tricks.

VAE can not only be used for the production of new sample, but also possesses the learning ability of the intrinsic from the sample.

Because \( E_{q_\varphi(z \mid x)} \left[ \log p_\theta(x \mid z) \right] \) is equivalent to L2 loss, which cause the extremely fuzzy effect, so the colorization based on VAE can not keep the quantity.

### 2.2 VAE-GAN

GAN which works as discriminator can introduced to solve the shortage of VAE. The constraint which caused by GAN discriminator to VAE can solve this problem. Figure 3 shows such VAE-GAN scheme.

**Figure 3.** The structure of VAE-GAN.

### 2.3 Mixture Density Network

Mixture Density Network, (MDN) [10] is one kind of neural networks for learning the multi-peak distribution which can not be properly solved by Gaussian mixture model. It is proved that Mixture Density Network shown figure 4 in can provide better solution for this problem, assume \( x \) is input, \( \pi_k \) is mean value of the Gaussian distribution, \( \sigma_k \) is the variance, \( \mu_k, \sigma_k \) are used as the output, the distribution of Mixture Density Network is given as
\[ p(y \mid x) = \sum_{k=1}^{K} \pi_k N(y \mid \mu_k, \Sigma_k) \] (3)

If their K components in the Mixture Density Network, the mixture coefficient \( \sum_{k=1}^{K} \pi_k = 1 \), which can be implemented by the normalization based softmax. For ordinary colorization, only one color image corresponds to a gray image. For the purpose of diverse colorization, the color distribution of the color modeling of the Mixture Density Network as shown in figure 4 should be used.

2.4. The Colorization Algorithm based on VAE-GAN and the Mixture Density Network
As shown above, the colorization based on VAE-GAN can overcome the problems, such as blurry effect, of that based on VAE.

The training process can be divided to steps: (1). a and b two color domain channels of the dataset is applied as the input and label. The flowchart of training process and testing process of VAE-GAN are respectively given in figures 5a and 5b.

3. Experimental Results
The condition of the experiment is given as follows, Ubuntu 16.04 operating system, Tensorflow1.14 deep learning structure, Nvidia GeForce 1080Ti GPU, Intel i7-7700 CPU, 32GB memory, Taking LSUN-Church as the train set as an example, the train of VAE-GAN needs setting set 32, 100
iteration. Using a second order Adam as the optimizer, Log decay rate is $\beta_1 = 0.9$, $\beta_2 = 0.99$, initial learning rate is $1 \times 10^{-3}$, the learning rate after 30 generations $1 \times 10^{-4}$, the learning rate after 70 generations is, the dataset is used to run in the model, and the intrinsic vector is stored to work as the label data of the train of MDN.

In the training process of MDN, suppose size is 64, 50 iterations, SGD is used for the optimization, the learning rate is $1 \times 10^{-5}$. The training parameters of two other dataset are same except the learning rate.

The datasets LFW, LSUN-Church and ImageNet-Val is used to valuate the colorization of our scheme and Zhang [11], Larsson [12], Iizuka [13], Cao [14], both in single mode and the diverse model.

3.1. Comparison of the Colorization in Single Mode

Figure 6 show the single colorization results of Cao’s method and our method in the colorization in single mode, where the first column is the gray images, the second column is real color image, third to sixth column are the colorization of six methods.

From figure 6, our method based on MDN are better than other five previous methods.
3.2. Comparison of Diverse Colorization Mode
Figures 7 and 8 show the diverse colorization results from Cao’s method [14] and our method, where each method can be used to get different colorization in any one of three sampling. In each figure, the first column correspond to the gray images, the second to the original real color image and the third to the result of the colorization.

From these results, our scheme is better than Cao’s method.

3.3. MOS Score of Subjective Quantity Criteria
A subjective quantity criteria MOS is applied to valuate the effectiveness of the colorization methods, in which 10 volunteers are asked to score the various colorization results on one to five score. Such a results is given in table 1, when $p < 0.01$. 
Figure 8. Diverse colorization from our method.

Table 1. MOS Scores of several colorization methods.

| Method | LFW   | LSUN-Church | ImageNet-Val |
|--------|-------|-------------|--------------|
| Cao    | 2.477 | 2.741       | 2.528        |
| Deshpande | 2.352 | 2.440       | 2.426        |
| Zhang  | 2.43  | 2.70        | 2.51         |
| Larsson| 2.83  | 2.95        | 2.89         |
| Iizuka | 2.79  | 2.92        | 2.93         |

Compared to other five methods, the colorization scheme based VAE-GAN and mixture density network in this paper achieves a high MOS score, which proves the effectiveness of our method.
3.4. FID Score of Objective Quantity Criteria
In order to analyze the quality and diversity of the colorization, Frechet Inception Distance (FID) is given to our scheme and other five methods as table 2.

|       | LFW   | LSUN-Church | ImageNet-Val |
|-------|-------|-------------|-------------|
| Cao   | 82.35 | 78.91       | 85.82       |
| Deshpande | 78.40 | 81.48       | 79.47       |
| Zhang | 75.25 | 76.25       | 78.17       |
| Larsson | 73.71 | 73.01       | 76.73       |
| Iizuka | 73.58 | 73.25       | 77.02       |
| ours  | **72.03** | **71.68**  | **76.52**  |

4. Conclusion
A two-step diverse colorization based on VAE-GAN and mixture density network is proposed, where VAE-GAN is used to learn the colors and features of the images, and the mixture density network is used to model the complete color distribution according to such features, which provides the diverse colorization. The training on the adversarial model and discriminator is applied to constrain the features of the color domain. The future work may be the colorization of 3D human body based on VAE-GAN, mixture density network and geometric generative model.

Acknowledgement
We thank the kind help from Professor Lianwen Jin of South China University of Technology in high performance workstations. The project is supported by NSFC No. 62071183.

References
[1] Gupta R K, Chia A Y S, Rajan D, et al. 2012 Image colorization using similar images Proceedings of the 20th ACM International Conference on Multimedia pp 369-378.
[2] Horiuchi T 2002 Estimation of color for gray-level image by probabilistic relaxation Object recognition supported by user interaction for service robots 16th International Conference on Pattern Recognition pp 867-870.
[3] Horiuchi T and Hirano S 2003 Colorization algorithm for grayscale image by propagating seed pixels Proceedings 2003 International Conference on Image Processing p I-457.
[4] Levin A, Lischinski D and Weiss Y 2004 Colorization using optimization ACM Transactions on Graphics 689-694.
[5] Noda H, Niimi M and Korekuni J 2006 Simple and efficient colorization in YCbCr color space 18th International Conference on Pattern Recognition (ICPR'06) (IEEE) pp 685-688.
[6] Cheng Z, Yang Q and Sheng B 2015 Deep colorization Proceedings of the IEEE International Conference on Computer Vision pp 415-423.
[7] Reinhard E, Adhikhmin M, Gooch B, et al. 2001 Color transfer between images IEEE Computer Graphics and Applications 21 (5) 34-41.
[8] Kingma D P and Welling M 2013 Auto-encoding variational bayes arXiv:1312.6114.
[9] Larsen A B L, Sønderby S K, Larochelle H, et al. 2016 Autoencoding beyond pixels using a learned similarity metric International Conference on Machine Learning PMLR pp 1558-1566.
[10] Bishop C M 1994 Mixture density networks (Neural Computing Research Group, Aston University) Technique Report NCRG/94/004 pp 1-25.
[11] Zhang R, Isola P and Efros A A 2016 Colorful image colorization European Conference on Computer Vision (Cham: Springer) pp 649-666.
[12] Larsson G, Maire M and Shakhnarovich G 2016 Learning representations for automatic colorization *European Conference on Computer Vision* (Cham: Springer) pp 577-593.

[13] Iizuka S, Simo-Serra E and Ishikawa H 2016 Let there be color! Joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification *ACM Transactions on Graphics (ToG)* **35** (4). 1-11.

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