The Transfer of Film Style Based on Meta-Learning

Keqin Chen†
Southwestern University of Finance and Economics
Chengdu 611130, China
chenkeqin@nsu.edu.cn

Yixin Meng†
Chengdu Neusoft University
Chengdu 611130, China
1608093132@qq.com

Gongfan Zhang
Chengdu Neusoft University
Chengdu 611130, China
2428549208@qq.com

Kun Zhu
Chengdu Neusoft University
Chengdu 611130, China
1031219242@qq.com

Wenbin Luo†
St. Mary's University
San Antonio, TX78228, U.S.A.
wluo@stmarytx.edu

ABSTRACT
The film style refers to the combination and configuration of colors in the film, which is often dominated by one color, making the picture show a certain tendency. However, the creation of special effects not only requires special professional skills, but also takes a lot of manual labor. If artificial intelligence technology can be transferred to the picture style of the film industry, production costs will be greatly reduced. In this paper, we propose a technique which combines the style transfer and meta-learning to create a new way of thinking. Compared with traditional image style transfer, the transfer of film style based on meta-learning could save the cost of film production significantly and take much less time to perform the transfer process. Toward the end, extensive experimental results were presented to validate our proposed method, which clearly outperforms the traditional image style transfer.

CCS CONCEPTS
• Computing methodologies ~ Artificial intelligence • Computing methodologies ~ Neural networks

KEYWORDS
Transfer Learning; Meta-learning; Film Style; Neural Network

1 Introduction
The traditional non-parametric image style transfer method is mainly based on the rendering of physical model and texture synthesis [14] Efros [1] proposed a simple texture algorithm, which combined and recombined sample textures to synthesize new textures. Hertzmann [5] proposed a method based on analogism to effectively implement image style transfers through those high-level abstract feature representations and obtained images with new textures synthesized through image feature mapping. Gatys [4] proposed the image style migration based on convolutional neural network. It is found that the convolutional neural network can be used to separate the content abstract feature representation and style abstract feature representation of the image, and to effectively realize the image style migration by dealing with those high-level abstract feature representations independently. It achieved very impressive artistic effect. The core idea of the algorithm is used in the pre-training VGG Model [10]. ZhuâĂŹs work has attracted wide attention from both academia and industry, and a large number of follow-up studies have been proposed, including two approaches, one based on image iteration and the other model-based iteration [18]. Among them, according to the different ways of image style acquisition, the method based on image iteration usually uses techniques, such as maximum mean difference [6]. Markov random field (MRF) [17] , and depth image analogy (DIA) [16]. On the other hand, the model-based iterative method commonly employs reconstruction decoder based on the construction model [8] [9] [12] and image [7], respectively. Mor [11] et al. proposed a method for achieving musical genre migration between different instruments, genres and styles. Style migration is used to convert between different styles. Chen [3] et al. used the generated confrontation network to present the semantic content of the picture into different artistic styles. This paper proposes a new meta-learning method for image style transfer and we apply it successively to the transfer of film style.

2 Meta-Learning And Neural Styl
Meta-learning is about changing the learning algorithm itself. This could be an external optimization loop that changes an internal optimization iteration, or a self-referencing algorithm that can change itself. Transfer learning or meta-learning is a learning algorithm that takes the initial parameters as part of itself. One of Chelsea Finn's latest algorithms, MAML [13], can be used for model-based intensive learning, where models can be dynamically changed quickly. In the evolutionary strategy gradient, a gradient descent based algorithm, the loss function uses the stochastic

AICS 2019, July 12–13, 2019, Wuhan, Hubei, China

ACM ISBN 978-1-4503-7150-6/19/07...$15.00
https://doi.org/10.1145/3349341.3349478

gradient descent method to optimize the parameters of the strategy. As a result, the parameters of the loss function are improved. Fig. 1 shows the meta-learning optimization process. However, the loss of the learned optimizer can become difficult to control, as the update steps unfold. Because the loss function of the optimizer parameters become more and more complex, minor changes in parameter values can be linked to the dramatic changes in the final performance. In the case of learning optimization, the variational optimization, a principled explanation of the evolutionary strategy as a way to mitigate the loss situation, is used to solve the aforementioned problem in this paper.

3 Network Structure and Improvement

3.1 Film Style and Color

In the film "Shadow" as shown in Fig. 2, the development of the whole story has been in the black and white gray scene under the ups and downs, showing a dull atmosphere, accurate black and white contrast into the contrast between good and evil. Also, the film's black and white main tone, color weakening, did not dilute the theme. Instead, it highlighted the cruelty of war, human nature of the cold. Unique black and white ash has also made the film more accurately reflect the complexity of human nature, better highlighting the theme of the film with Yin and Yang balance. The formation of color main tone directly affects the formation of the theme style of the film.

3.2 Mod VGG Network

Before implementing style migration, let us take a look at the VGG network as shown in Fig. 3 below and discuss how to use the VGG network for image style migration.

VGG19 is a DCNN structure proposed by Google DeepMind in a paper published in ICLR 2015 [15]. As it is well known, the CNN network performs well in image processing. After VGG19 was proposed, it was also applied to the area of image processing. Generally speaking, the training of deep convolutional neural network is a step-by-step process, which includes extracting data set features ranging from simple features to more complex ones. The trained model learns to extract image features. Therefore, the trained model can be used to extract other image features directly, at least in theory, which is also the basis of transfer learning.

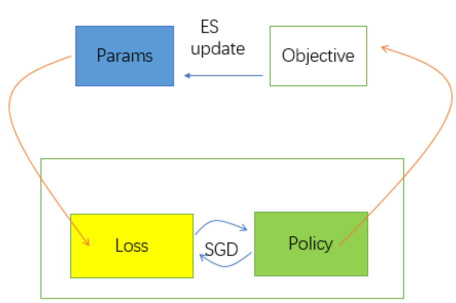

Figure 1: Meta-Learning Optimization Process

Figure 2: A Screenshot of the Film "Shadow"

In real applications, however, the results obtained are often not as good as those obtained by retraining the model on new data. Regardless, it can save a lot of training time, which is very useful under certain circumstances.

3.3 Style Migration

Two things should be made clear about style migration for a piece of video: 1. The generated image shall retain the characteristics of the original image; 2. The generated images need to have texture features of stylized images. According to these, it can be determined that in order to realize style transfer, there are two loss values: one is the loss of the content features between the generated image and the original image, and the other is the loss of the texture features between the generated image and the style image. However, to extract different features of an image (the content features and the texture features), we only need to use different convolutional structures for training. In this paper, we will use two neural networks.

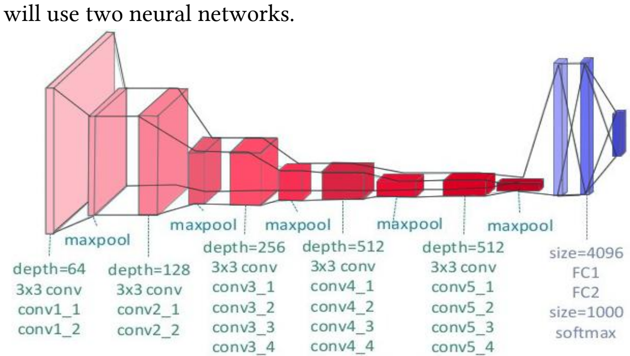

Figure 3: VGG19 Network Structure Diagram

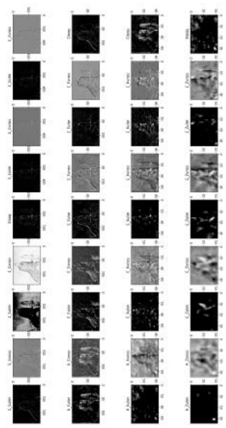

Figure 4: Represents Figure 2 Through VGG19 Features of Network Learning

A VGG network continuously used convolutional layers to extract features, and used those features to classify items. Therefore, the parameters of the extracted content and texture features in the network can be migrated and used.

As shown in Fig. 6, suppose the initial image x (the input image) is a random image, which is passed through fw (Image Transform Net) network to generate image y. At this time, y needs to calculate the features with the style picture ys to get a loss style and calculate the features with the content picture yc to get a loss content. Assuming that loss = loss style + loss content, the network parameters of fw can be trained.

3.4 Performance Measurements

Since the model used in Fig.1 is VGG-19, it is equivalent to calculating the characteristics obtained by the two images with relu3-3 of VGG-19, and the calculated functions are as follows:

$$\ell_{feat}^{\emptyset,J}(\hat{y},y)=\frac{1}{C_jH_JW_J}\parallel \emptyset_j(\hat{y})-\emptyset_j(y)\parallel^2$$

In short, suppose yc, the characteristic matrix obtained is φ(Y), and the characteristic matrix obtained by the image is φ(Yˆ), and c = ϕ.channel, w = ϕ.weight, h = ϕ.height, Then there are:

$$\boldsymbol{loss}_{content}=\frac{\sum_{i=0,f=0}^{w,h}\left(\emptyset_{i,j(\hat{y})}-\emptyset_{i,j(\hat{y})}\right)^2}{c*w*h}$$

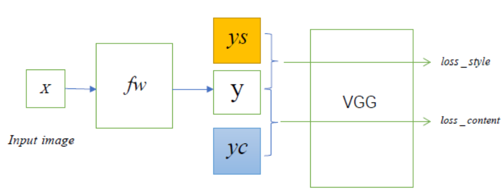

Figure 5: Parameter Training Process

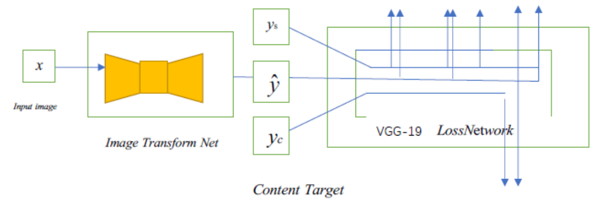

Figure 6: The Loss in VGG Value Procedure.

FPS is the number of frames per second on the screen. It refers to the number of frames in an animation or video. FPS measures the amount of information used to save and display dynamic video. The more frames per second, the smoother the action.

$fps = frameCount\ /\ elapsedTime$

Esab refers to the stability error and measures the loss between t and t-1. Here, we evaluate the stability of the results. The smaller the stability error, the more stable the result. For the entire video, we replace it with the average error.

$e_{stab}(t,t-1)=Mg\odot\parallel t-w^t(t-1)\parallel^2$

4 Experimental Results

4.1 Data Sources

This paper uses two short videos (one animated movie episode and the other one is a real video of our own) as our training dataset. In order to obtain a ground true stream similar to each of the two consecutive frames in these videos, we used Deepflow [17] to calculate the bidirectional light flow and used the reverse flow as the ground true stream. For the ground-truth of the component mask Mg, we use the method exployed in literature [13] to detect occlusion and motion boundaries. We mask two types of pixels, set to 0:2 in Mg, and the moving boundary pixels with large flow gradients tend to be less accurate, which can lead to ghost phenomena in the composition. All other pixels in Mg are set to 1.

Table 1: Comparison of Stability Errors with Different Methods of Runtime

Methods	Stablilty Error esab		Runtime (Fps)
	Selfie video coversion effect	Totoro	
Johnson [6] et al.	0.0199	0.0240	38.17
Chen [2] et al.	0.0159	0.0181	20.60
Our proposed method	0.0128	0.0113	15.07

The proposed method worked with baselines per frame [10]. The video showed a lower stability loss, even though GPUs with only 2.52.8Ghz is relatively slow. Compared with fixed flow subnets, the fine-tuning flow subnet has a better consistency.

4.2 Implementation Details

In this experiment, the training set uses two types of Pretraining Style Network. The image resolution for all videos is 640x360 and batch size of the network is 1 (frame) to carry out K iteration. The Adam optimization method was employed, the initial learning rate was 1e-4, and the attenuation was 0.

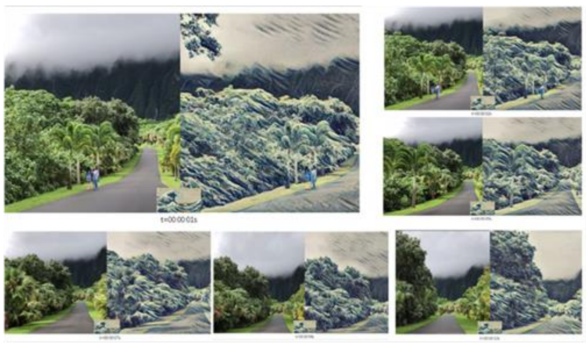

Figure 7: Selfie video conversion effect

Figure 8: The effect of stylistic transfer in the Totoro movie

After 500 rounds of iteration, we produced the video screenshots shown in Figures 7 and 8. We can certainly increase the number of training iterations to obtain more effective style transfer images. In practical applications, we applied image style conversion technology to a real video and the Hayao Miyazaki's Totoro movie, which achieved significant shifting and rendering effects.

Although Gatys et al.'s method [4] can generate good neural style transfer results, its speed is very slow. Instead, you can use a pre-training network (usually pre-training on ImageNet) and define a loss function that achieves the goal of style migration, and then optimize the loss function over time. This includes: content loss, style loss and total variation loss. Each part is calculated separately and then combined in a primitive loss function. Content, style, and total variation loss can be optimized in turn by minimizing the element loss function. The first step is to train a network to produce the desired style. Once the network is trained, it can be applied to any content video.

Johnson [6] et al.'s method, with fast computing speed, can be used to quickly stylize images. However, as the mainstream technology of industrial application software, image quality needs to be further improved, requiring a large amount of training data. Chen [2] et al.'s method has two branches and uses alternate training, so it is necessary to define two separate loss functions. In terms of training strategy, for every T+1 iterations, the stylized branch is trained for the first T iterations, and then the self-coding branch is trained for one time. Because two loss functions need to be defined, it is very troublesome in training. How could you learn good features when the amount of annotated data is small? We need to generalize those rare categories without additional training, since training is not rewarding due to the lack of data, the high cost, and the long duration. The goal of meta-learning is to create a model for various learning tasks, so that some new learning tasks can be solved with very limited samples. The model created by meta-learning combines previous knowledge with a small amount of information from the current new task. As a result, it can avoid overfitting on the new data. Our proposed method clearly outperforms both Johnson [6] et al. and Chen [2] et al. in terms of stability error and runtime, as shown in Table 1.

4.3 Known Issues and Future Research Directions

Our proposed algorithm of image style migration based on deep learning has achieved remarkable results. However, there are still some known issues that need to be solved before its being widely adopted. This section discusses several issues at present and puts forward our suggestions for future research.

Parameter adjustment: In order to obtain satisfactory results, both the image iterative method and the model-based iterative method need to be manually tuned. Especially with the model iteration method, the model should be re-trained after each adjustment of the parameters. Although the method based on image reconstruction encoder can alleviate the problem of parameter adjustment and does not need to train the model separately for different styles, the training process of image refactoring decoder is cumbersome, and the image generation effect is not ideal. Local smoothing can improve the performance of the method based on image refactoring decoder, but it will make the texture of stylized image disappear, and the final effect is almost similar to Image color migration. Therefore, finding a method, which is both simple and controllable and can guarantee the image quality, is an important research direction in the next step. If the model storage capacity is not taken into account, further improving the image generation quality based on image reconstruction coding method is a very worthy research direction, because this method can effectively avoid the problem of parameter adjustment.

Limitations of pre-training model: currently, most of the image style transfer methods based on deep learning adopt the VGG model before training to extract image features. Although VGG is a good convolutional neural network model before training and performs well in feature extraction, it is a heavyweight model with a large volume and a large amount of calculation. It was not designed specifically for image style transfer at the beginning. Therefore, getting rid of the dependence on the pre-exercise VGG model, or designing a more refined and effective feature extractor, is an important way to promote the further development of image style transfer based on deep learning.

The generation of confrontation network can solve the limitation of the pre-training model. Its real image generation effect is conducive to improving the quality of the generated image. The optimization method based on the discrete distribution is similar to the associated method based on image iteration. The method of confrontation training has a good effect in the application scenario of acquiring new features.

Improvement of transfer learning theory: Image style transfer is a special case in transfer learning. At present, the transfer learning method based on in-depth learning is still in its infancy and needs better mathematical methods and theoretical guidance. The perfection of transfer learning theory is of great significance to the further development of image style transfer based on in-depth learning.

In order to improve the migration learning ability of the model, a powerful and universal neural network model is proposed in the related research work of the general model, which plays an important role in guiding the further development of image style migration

Pre-processing and post-processing methods: In order to make the final results more suitable for practical application, some pre-processing and post-processing methods can be used, such as semantic segmentation, fusion, color migration, smoothing and so on. These pre-processing and post-processing methods play an important role in improving the image style migration effect.

5 Application

With the continuous improvement of algorithms and theories in image style migration based on deep learning, the effect of style migration has been greatly improved, and it has broad commercial application prospects. The style migration method proposed in this paper can be used in the following areas.

A) Image processing: At present, most of the images circulating on social networks are processed by software, and image beautification is a popular application technology. The traditional image processing technology can only perform fixed pattern processing on images, and the emergence of image style migration based on neural network brings more imagination space to image style design. It can also be applied to mobile apps that turn a user's video into a high-quality art video in a matter of seconds. Subsequent charges for video style migration applications can generate certain commercial value. With the help of these applications, people can easily create their own style of art without special expertise.

B) Video processing: In the film and television entertainment industry, such as movies, television, animation etc., film and television special effects technology can be seen everywhere. However, the creation of film and television special effects technology requires not only special professional skills, but also a lot of manual labor. If more artificial intelligence technology can be used, the cost-of online videos be greatly reduced. You can also delve into the more advanced parameter spaces in image style migration and impressively styled movie scenes.

C) Auxiliary tools for style design: Image style migration can serve as a useful aid, such as art painting, architectural art design, costume art design, game scene design, and more. This is likely to become a research hotspot in the future. With the current research progress, the image style migration based on deep learning is developing rapidly, and its potential commercial value needs to be further explored.

6 Conclusion

In this paper, the transfer of film style based on meta-learning was proposed, and its application scenario, existing problems, and future research directions were discussed and analyzed in depth. Although successful application cases were presented in this paper with the proposed method, further research and improvement is needed before it can be adopted widely in commercial applications. In general, the transfer of basic style is a challenging, emerging topic, which not only attracted wide attention of the academic community, but also has a great demand in the industry, with important research significance and broad application prospects.

REFERENCES

[1] A EFROS ANDW and T FREEMAN. [n.d.]. Image quilting for texture synthesis and transfer. In Proc. ACM Conf. Comp. Graphics (SIGGRAPH). 341–346.
[2] Dongdong Chen, Yuan Lu, Liao Jing, Nenghai Yu, and Hua Gang. 2017. StyleBank: An Explicit Representation for Neural Image Style Transfer. (2017).
[3] Xinyuan Chen, Chang Xu, Xiaokang Yang, Li Song, and Dacheng Tao. 2018. Gated-GAN: Adversarial Gated Networks for Multi-Collection Style Transfer. IEEE Transactions on Image Processing 28, 2 (2018), 1–1.
[4] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. 2015. A neural algorithm of artistic style. arXiv preprint arXiv:1508.06576 (2015).
[5] Aaron Hertzmann, Charles E Jacobs, Nuria Oliver, Brian Curless, and David H Salesin. 2001. Image analogies. In Proceedings of the 28th annual conference on Computer graphics and interactive techniques. ACM, 327–340.
[6] Justin Johnson, Alexandre Alahi, and Fei Fei Li. 2016. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In European Conference on Computer Vision.
[7] Jan Eric Kyprianidis, John Collomosse, Tinghuai Wang, and Tobias Isenberg. 2013. State of the "Art": A Taxonomy of Artistic Stylization Techniques for Images and Video. (2013).
[8] Yijun Li, Fang Chen, Jimei Yang, Zhaowen Wang, and Ming Hsuan Yang. 2017. Universal Style Transfer via Feature Transforms. (2017).
[9] Yijun Li, Ming Yu Liu, Xueting Li, Ming Hsuan Yang, and Jan Kautz. 2018. A Closed-form Solution to Photorealistic Image Stylization. (2018).
[10] Yanghao Li, Naiyan Wang, Jiaying Liu, and Xiaodi Hou. 2017. Demystifying neural style transfer. arXiv preprint arXiv:1701.01036 (2017).
[11] Noam Mor, Lior Wolf, Adam Polyak, and Yaniv Taigman. 2018. A Universal Music Translation Network. (2018).
[12] Yaroslav Nikulin and Roman Novak. 2016. Exploring the Neural Algorithm of Artistic Style. (2016).
[13] Manuel Ruder, Alexey Dosovitskiy, and Thomas Brox. 2016. Artistic Style Transfer for Videos. (2016).
[14] Manuel Ruder, Alexey Dosovitskiy, and Thomas Brox. 2018. Artistic style transfer for videos and spherical images. International Journal of Computer Vision 126, 11 (2018), 1199–1219.
[15] Karen Simonyan and Andrew Zisserman. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. international conference on learning representations (2015).
[16] Xin Wang, Geoffrey Oxholm, Da Zhang, and Yuan Fang Wang. 2016. Multimodal Transfer: A Hierarchical Deep Convolutional Neural Network for Fast Artistic Style Transfer. (2016).
[17] Philippe Weinzaepfel, Jerome Revaud, Zaid Harchaoui, and Cordelia Schmid. 2014. DeepFlow: Large Displacement Optical Flow with Deep Matching. In IEEE International Conference on Computer Vision.
[18] Xizhou Zhu, Yuwen Xiong, Jifeng Dai, Lu Yuan, and Yichen Wei. 2017. Deep feature flow for video recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2349–2358.