Tel Aviv University

Deep Learning Course

Project Report

# Universal Style Transfer via Feature Transforms

*Authors*
Eyal Waserman
Carmi Shimon

February 16, 2019

**Abstract**

Style transfer is the technique of recomposing images in the style of other images. Universal style transfer aims to transfer arbitrary visual styles to content images. Existing feed-forward based techniques would need to be trained on pre-defined styles and then fine tuned for new styles. Whereas, this paper presents new methods which are completely independent of the style during train phase making it a "learning-free" approach. In this paper, we present an encoder-decoder architecture where the encoder serves as feature-extractor and the decoder is trained for image reconstruction. The Goal is to build a new Image which has the contents of the Content Image and style of the Style Image. Our results provide new insights handling arbitrary styles in an efficient-Learning-free methods.

# 1 Introduction

Style transfer aims to synthesize an image that preserves some notion of the content but carries characteristics of the style. The key challenge is how to extract effective representations of the style and then match it in the content image. Transferring the style from one image onto another can be considered a problem of texture transfer. In texture transfer the goal is to synthesize a texture from a source image while constraining the texture synthesis in order to preserve the semantic content of a target image. For texture synthesis there exist a large range of powerful non-parametric algorithms that can synthesise photo realistic natural textures by re-sampling the pixels of a given source texture [1, 2, 3, 4]. Most previous texture transfer algorithms rely on these non-parametric methods for texture synthesis while using different ways to preserve the structure of the target image. For instance, Efros and Freeman introduce a correspondence map that includes features of the target image such as image intensity to constrain the texture synthesis procedure [3]. Lee et al. improve this algorithm by additionally informing the texture transfer with edge orientation information [5]. Although these algorithms achieve remarkable results, they all suffer from the same fundamental limitation: they use only low-level image features of the target image to inform the texture transfer. Ideally, however, a style transfer algorithm should be able to extract the semantic image content (e.g. the objects and the general scenery) and then inform a texture transfer procedure to render the semantic content of the target image (content and style image) in the style of the style image. Therefore, a fundamental prerequisite is to find image representations that independently model variations in the semantic image content and the style in which it is presented. To generally separate content from style in natural images is still an extremely difficult problem.
However, the recent advance of Deep Convolutional Neural Networks (CNNs) [6] has produced powerful computer vision systems that learn to extract high-level semantic information from natural images. It was shown that CNNs trained with sufficient labeled data on specific tasks such as object recognition learn to extract high-level image content in generic feature representations that generalize across data sets [7] and even to other visual information processing tasks, including texture recognition [8] and artistic style classification [15].

The main issue is how to properly and effectively apply the extracted style characteristics (feature correlations) to content images in a style-agnostic manner.

In this work we show how the generic feature representations learned by high-performing CNNs (Encoder) followed by efficient feature Whitening-Coloring transforms (WCTs) and a compatible reconstruction (Decoder) can be used to manipulate the content and the style of natural images. We introduce a novel methods which boosts style transfer by taking advantage of the existence of feature representations from state-of-the-art CNNs. We also show new and efficient methods of combining different style images into the target image (content image) by using WCT algorithm efficiently. Our goal was to invent new and efficient ways of UST based on the work by Li et al. [11]. Our method consists of a stylization step and a smoothing step. Both have a closed-form solution1and can be computed efficiently. The stylization step is based on the (WCT) [10], which stylizes images via feature projections. The WCT was designed for artistic stylization. Our results show similar results as presented in [10] while showing efficiency in computation.

## 2 Related Work

Existing stylization methods can be classified into two categories: global and local. Global methods [12, 13] achieve stylization through matching the means and variances of pixel colors [12]. Local methods [14] stylize images through finding dense correspondences between the content and style photos based on either low-level or high-level features. These approaches are slow in practice. Also, they are often developed for specific scenarios. Therefore these methods do not scale to the setting of arbitrary style images well.

Gatys et al. [7, 8] showed remarkable results by using the VGG-19 deep neural network for style transfer. The major step in the algorithm is to solve an optimization problem of matching the Gram matrices of deep features extracted from the content and style photos. A number of methods have been developed [15, 16, 17] to further improve its stylization performance and speed. However, these methods do not aim for preserving photorealism.

Their approach was taken up by various follow-up papers that, among other things, proposed different ways to represent the style within the neural network. Li et al. [15] suggested an approach to preserve local patterns of the style image. Instead of using a global representation of the style, computed as Gram matrix.

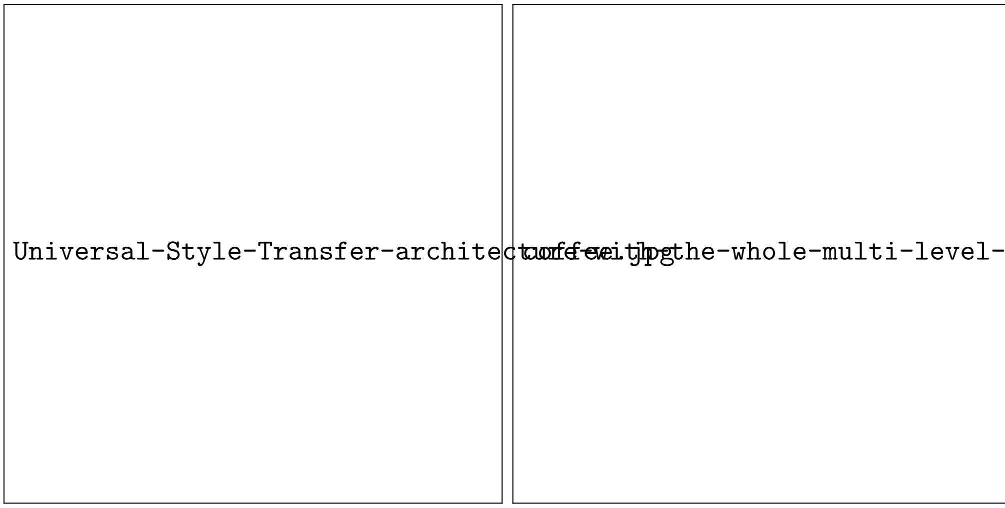
Nikulin et al. [18] tried the style transfer algorithm by Gatys et al. on other nets than VGG and proposed several variations in the way the style of the image is represented to archive different goals like illumination or season transfer. However, this method is developed for specific scenarios which cannot be scaled to the setting of arbitrary style images.

This work is an extension of [11], which is closest to a related work [19], directly adjusts the content feature to match the mean and variance of the style feature. However, the generalization ability of the learned models on unseen styles is still limited.

Different from the existing methods, our approach performs style transfer efficiently in a feed-forward manner while achieving generalization and visual quality on arbitrary styles. Our approach is closely related to [15], where content feature in a particular (higher) layer is adaptively instance normalized by the mean and variance of style feature. This step can be viewed as a sub-optimal approximation of the WCT operation, thereby leading to less effective results on both training and unseen styles. Moreover, our encoder-decoder network is trained solely based on image reconstruction, while [15] requires learning such a module particularly for stylization task. We evaluate the proposed algorithm with existing approaches extensively on both style transfer and texture synthesis tasks and present in-depth analysis.

## 3 Methods

Our proposed algorithm is first to implement [11] which formulates style transfer as an image reconstruction process coupled with feature transformation, i.e., whitening and coloring. The reconstruction part is responsible for inverting features back to the RGB space and the feature transformation matches the statistics of a content image to a style image. We used a pre-trained weights which was trained by VGG-19 [20] encoder using ImageNet dataset (Deng et al.) [21]. Second, we show an improved algorithm which merges two style images bu using WCT algorithm which based on singular value decomposition (SVD). Here, we both implement the original merge algorithm as proposed in [11] as well as introduce three additional efficient methods based on the use WCT.



(a) Encoder WCT Decoder Architecture.

(b) More coffee.

Figure 1: The same cup of coffee. Two times.

### 3.1 Stylization

**WCT**. The WCT [11] formulates stylization as an image reconstruction problem with feature projections. To utilize WCT, an auto-encoder for general image reconstruction is first trained. We used the VGG-19 model [20] as the encoder $\varepsilon$(weights are kept fixed) and trains a decoder D for reconstructing the input image. The decoder is symmetrical to the encoder and uses up-sampling layers to enlarge the spatial resolutions of the feature maps, (see figure [1]). Once the auto-encoder is trained, a pair of projection functions are inserted at the network bottleneck to perform stylization through the whitening ($P_C$) and coloring ($P_S$) transforms. The key idea behind the WCT is to directly match feature correlations of the content image to those of the style image via the two projections. Specifically, given a pair of content image $I_C$ and style image $I_S$, the WCT first extracts their vectorised VGG features $C_f = \varepsilon(I_C)$ and $S_f = \varepsilon(I_S)$, and then transform the content feature $C_f$ via

$$CS_f = P_S P_C H_C \tag{1}$$

Where $P_C = E_C \Lambda_C^{-\frac{1}{2}}$, and $P_S = E_S \Lambda_S^{\frac{1}{2}}$. Here $\Lambda_C$ and $\Lambda_S$ are the diagonal matrices with the eigenvalues of the covariance matrix $C_f C_f^T$ and $S_f S_f^T$ respectively. The matrices $E_C$ and $E_S$ are the corresponding orthonormal matrices of the eigenvalues, respectively. After the transformation, the correlations of transformed features match those of the style features, i.e., $CS_f CS_f^T = S_f S_f^T$. Finally, the stylized image is obtained by directly feeding the transformed feature map into the decoder: $Y = D(CS_f)$. For better stylization performance, Li et al. [11] use a multi-level stylization strategy, which performs the WCT on the VGG features at different layers. The WCT performs well for artistic image stylization. However it generates structural artifacts (e.g., distortions on object boundaries)

# 4 Experiments

# 5 Conclusions

# 6 References

## References


[1] A. Efros and T. K. Leung. Texture synthesis by non-parametric sampling. In Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on,volume 2, pages 1033–1038. IEEE, 1999.

[2] L. Wei and M. Levoy. Fast texture synthesis using tree-structured vector quantization. In Proceedings of the 27th annual conference on Computer graphics and interactive techniques, pages 479–488. ACM Press/Addison-Wesley Publishing Co., 2000.

[3] A. A. Efros and W. T. Freeman. Image quilting for texture synthesis and transfer. In Proceedings of the 28th annual conference on Computer graphics and interactive techniques, pages 341–346. ACM, 2001.

[4] V. Kwatra, A. Schodl, I. Essa, G. Turk, and A. Bobick. Graph cut textures: image and video synthesis using graph cuts. In ACM Transactions on Graphics (ToG), volume 22,pages 277–286. ACM, 2003.

[5] H. Lee, S. Seo, S. Ryoo, and K. Yoon. Directional Texture Transfer. In Proceedings of the 8th International Symposium on Non-Photo realistic Animation and Rendering, NPAR '10,pages 43–48, New York, NY, USA, 2010. ACM.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Image net classification with deep convolutional neural networks. In Advances in neural information processing systems, pages1097–1105, 2012.

[7] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang,E. Tzeng, and T. Darrell. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition.arXiv:1310.1531 [cs], Oct. 2013. arXiv: 1310.1531

[8] M. Cimpoi, S. Maji, and A. Vedaldi. Deep filter banks for texture recognition and segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3828–3836, 2015.

[9] S. Karayev, M. Trentacoste, H. Han, A. Agarwala, T. Darrell,A. Hertzmann, and H. Winnemoeller. Recognizing imagestyle.arXiv preprint arXiv:1311.3715, 2013.

[10] Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollar, P.,Zitnick, C.L.: Microsoft COCO: Common objects in context. In: ECCV, 2014.

[11] [11] Li, Y., Fang, C., Yang, J., Wang, Z., Lu, X., Yang, M.H.: Universal style transfer via feature transforms. In: NIPS, 2017.

[12] [12] Reinhard, E., Ashikhmin, M., Gooch, B., Shirley, P.: Color transfer between images.IEEE Computer graphics and applications21(5) (2001) 34–41.

[13] [13] Freedman, D., Kisilev, P.: Object-to-object color transfer: Optimal flows and smsp transformations. In: CVPR, 2010.

[14] [14] Shih, Y., Paris, S., Durand, F., Freeman, W.T.: Data-driven hallucination of different times of day from a single outdoor photo. In: SIGGRAPH, 2013.

[15] [15] Li, C., Wand, M.: Combining markov random fields and convolutional neural networks for image synthesis. In: CVPR, 2016.

[16] [16] Ulyanov, D., Lebedev, V., Vedaldi, A., Lempitsky, V.: Texture networks: Feed-forward synthesis of textures and stylized images. In: ICML, 2016.

[17] [17] Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: ECCV, 2016.

[18] Nikulin, Y., Novak, R.: Exploring the neural algorithm of artistic style. CoRRabs/1602.07188, 2016.

[19] X. Huang and S. Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In ICCV, 2017.

[20] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scaleimage recognition. In: ICLR, 2015.

[21] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. Imagenet: A large-scale hierarchical imagedatabase. InProc. CVPR, 2009