

An Improved Style Transfer Approach for Videos

Rong-Jie Chang
Institute of Multimedia Engineering
National Chiao Tung University
Hsinchu 300, Taiwan
Email: cps9257@gmail.com

Der-Lor Way
Department of NewMedia Art
Taipei National University of Arts
Taipei 112, Taiwan

Chin-Chen Chang
Department of Computer Science and
Information Engineering
National United University
Miaoli 360, Taiwan
Email: ccchang@nuu.edu.tw

Zen-Chung Shih
Department of Computer Science
National Chiao Tung University
Hsinchu 300, Taiwan

Abstract—In this paper, we present an improved approach to transfer style for videos based on semantic segmentation. We segment foreground objects and background, and then apply different styles respectively. A fully convolutional neural network is used to perform semantic segmentation. We increase the reliability of the segmentation, and use the information of segmentation and the relationship between foreground objects and background to improve segmentation iteratively. We also use segmentation to improve optical flow, and apply different motion estimation methods between foreground objects and background. This improves the motion boundaries of optical flow, and solves the problems of incorrect and discontinuous segmentation caused by occlusion and shape deformation.

Keywords—*Semantic segmentation; Motion estimation; Neural network; Style transfer*

I. INTRODUCTION

Recently, style transfer using deep neural networks just outperforms other applications in both quantity and quality. By extracting high-level features from a style image, the objective is not restricted to any specific style. Numerous applications using this technique only process a single image, such as using an artist painting as a style image and transferring this style to a photo-realistic image. Even when this technique is applied to videos, it still uses only one style image to do the whole style transfer. This design loses the degrees of freedom in creativity. In addition, it makes no difference applying style to objects and environment.

Temporal coherence in a stylized video is crucial. The stylized video will result in terribly flickering without temporal coherence. This problem can be solved by motion estimation. Optical flow, one of the motion estimation methods, is the most widely used motion estimation method. Semantic segmentation provides information of object location and boundaries, which can solve the problem of inaccurate motion boundaries.

In this paper, we propose an improved approach that combines semantic segmentation and style transfer for videos. User can input a video and one or two style images. Then our approach automatically produces coarse foreground object

segments and background segments frame by frame. We use flood fill algorithm to separate unconnected segments, treat each of them as an individual cluster, and improve them with fully connected conditional random field (CRF) and grab cut algorithm. We take these improved segments as constraints to improve optical flow, and then we apply style transfer with these improved results.

The main contribution of our system is that we use semantic segmentation in style transfer to increase the degrees of freedom in creativity, and strengthen the reliability of the inference from the fully convolutional neural network. Besides, the temporal coherence can be enhanced by the usage of semantic segmentation.

II. METHOD

The aim of our approach is to enhance the quality of artistic style videos by using semantic segmentation. Given a video sequence and two style images, our approach extracts foreground and background objects frame by frame and transfers the style accordingly. We improve the problems of inaccurate motion boundaries, discontinuous segmentations and low-level noise in stylized images. Our system diagram is shown in Fig. 1.

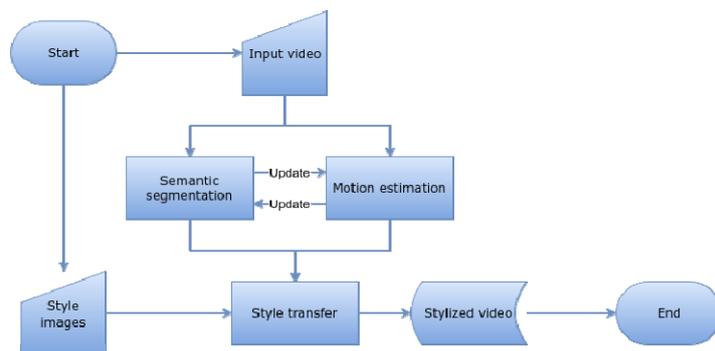


Fig 1. Our system diagram.

III. RESULTS

Fig. 2 shows the results of our approach. The results show that we can obtain more precise motion boundaries.

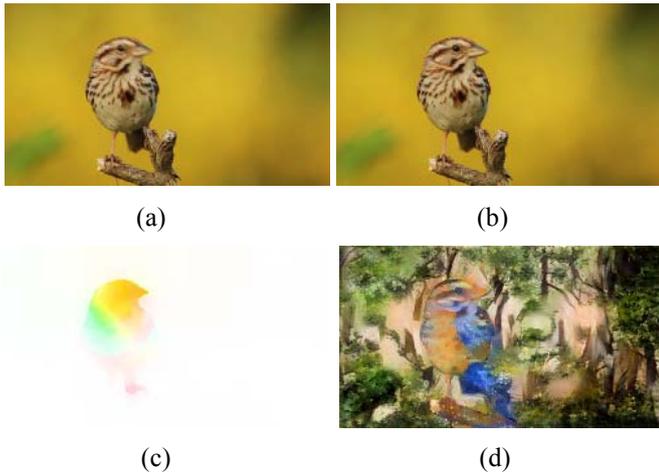


Fig. 2. Results of our approach. (a) Frame t , (b) Frame $t+1$, (c) Our improved flow, and (d) Our stylized result.

IV. CONCLUSION

We have proposed a method to improve semantic segmentation and motion estimation, and transfer style to videos. Our method solves the unstable segmentation from FCN and classifies objects into foreground and background. We use the foreground and background segments as inputs for grab cut, and then use dynamic bounding boxes to iteratively expand and refine the segmentation. We choose different optical flow methods depending on the characteristic of the segmentation. Our approach outperforms spatial homogeneous optical flow method and has more precise motion boundaries if the movement of the foreground object is greatly different from the background. With these improved data we can apply different styles in the foreground and background and create artistic style videos with coherence. In addition, we can segment a style image to specify the region for feature extraction.

REFERENCES

- [1] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," *Computer Vision and Pattern Recognition*, 2016.
- [2] M. Ruder, A. Dosovitskiy, and T. Brox, "Artistic style transfer for videos," *German Conference on Pattern Recognition*, 2016.
- [3] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *Computer Vision and Pattern Recognition*, 2015.