An Efficient Mobile-Edge Collaborative System for Video Photorealistic Style Transfer

Ang Li[1], Chunpeng Wu[1],Yiran Chen[1], Bin Ni[2]
[1]Department of Electrical and Computer Engineering, Duke University
[2]Quantil Inc.
Email: {al380, chunpeng.wu, yiran.chen}@duke.edu, nibin@quantil.com

ACM Reference Format:
Ang Li[1], Chunpeng Wu[1],Yiran Chen[1], Bin Ni[2]. 2019. An Efficient Mobile-Edge Collaborative System for Video Photorealistic Style Transfer. In *SEC '19: ACM/IEEE Symposium on Edge Computing, November 7–9, 2019, Arlington, VA, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3318216.3363332

1 INTRODUCTION

In the past decade, convolutional neural networks (CNNs) have achieved great practical success in image transformation tasks, including style transfer, semantic segmentation, etc. CNN-based style transfer, which denotes transforming an image into a desired output image according to a user-specified style image, is one of the most popular techniques in image transformation. It has led to to many successful industrial applications with significant commercial impacts, such as Prisma and DeepArt. Figure 1 shows the general workflow of the CNN-based style transfer. Given a content image and a user-specified style image, the content features and style features can be extracted using a pre-trained CNN, and then be merged to generate the stylized image. The CNN model is trained for generating a stylized image that has similar content features as the content image's and similar style features as the style image's. In this example, we can see the content image is captured at a lake in the daytime, while the style image is another similar scene captured at dusk. After performing style transfer, the content image is successfully transformed to the dusky scene while keeping the content unchanged as the content image.

However, directly applying existing style transfer techniques on videos is very challenging, since frame-by-frame transformations are very slow even running on powerful GPUs. Although mobile phones have been widely adopted to record our daily, and then edit and share images and videos on social networks or with friends, performing video style transfer on mobile phones is still technically unaffordable due to the limited computing sources on mobile phones. Besides, another critical challenge is "photorealistic" style transfer. Artistic style can tolerate distortions in stylized images, however, loyally preserving the content structure is required in photorealistic style transfer. Therefore, the visual quality of photorealistic stylized images can be evaluated by humans. There is a necessity to design a system that can efficiently realize photorealistic style transfer of videos on mobile phones while keeping high visual quality of the stylized videos.

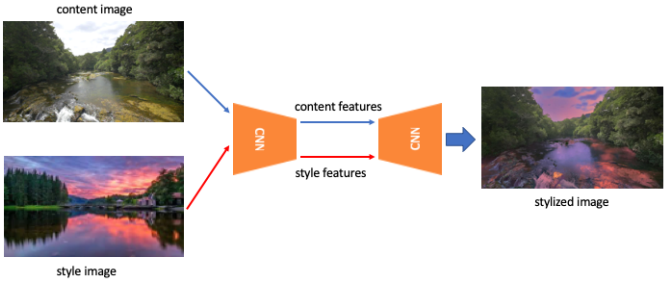

Figure 1: An example of photorealistic style transfer.

To address above challenges, we proposed an efficient mobile-edge collaborative system for performing photorealistic style transfer of videos. Instead of performing stylization frame by frame, only extracted key frames need to be processed using pre-trained CNN models on edge servers, while the rest of intermediate frames are generated on-the-fly using our proposed optical-flow-based frame interpolation algorithm on mobile phones. The interpolation relies on the optical flow information between the intermediate frames and key frames. Since edge servers are closer to users and hence can provide real-time response to mobile phones, we adopt edge servers for assistance rather than cloud servers. We also design a meta-smoothing module to efficiently address two problems in the style transfer: *dynamically upscaling* and *removing distortions*. The meta-smoothing module can be trained in an end-to-end manner. Experimental results demonstrate our proposed system can successfully perform stylizations of videos with even better visual quality compared to the state-of-the-art method while achieving significant speedup with high resolutions.

2 SYSTEM DESIGN

As Figure 2 shows, the system consists of two major modules: *CNN-based stylizer* and *optical-flow-based interpolation*. The work flow is as follows: When a user performs style transfer of a video on the mobile phone, the key frames will be extracted and then sent to the edge server associated with a user-specified style image. After that, the received key frames will be transformed according to the style image using a pre-trained CNN-based stylizer on the edge server. The stylized key frames will then be returned to the mobile client. Finally,

SEC '19, November 7–9, 2019, Arlington, VA, USA

ACM ISBN 978-1-4503-6733-2/19/11.
https://doi.org/10.1145/3318216.3363332