# CS577 Group Project Proposal

## 3D character animation from artistic-style photos

**Baoli Hao** **Ivan Gvozdanovic**

## 1 Descriptions and Brief Survey

Object detection and recognition using CNN type networks have been a prominent field of study in the last decade. In every branch of this area of research, the content representation is referred to as the feature responses in higher layers of the network [3]. Furthermore, one of those branches studies ways of how to combine features from a reference image to the source image serving as input. Capturing the texture information and extracting the style of the reference image which in turn is applied to the source image is called "style transfer" [4] (see one example below). This result was generated on the basis of the VGG-Network [6]. The structure of this system consists of two CNN networks. One processes the source image and the other captures the desired style from the *style* image. These two are then combined to get the Figure 1a below.

(a) "Style-transferred" image

(b) Style image

Another interesting study is three-dimension character animation from a single photo input [8]. The team behind the *Photo-wake up* network displayed a novel approach for animating a human subject from a single photo. They developed an application of viewing and animating humans in single photos in 3D firstly, and devised a novel 2D warping technique that enables

the transformation of a poseable template body model to adapt and conform to the intricate contours of a person's silhouette. The whole process is consisted of the following steps. First a character in the input image is detected. Then, his silhouette is isolated which in turn is broken down in different meshes. The meshes with open curvatures are closed and pose bones are added to the different body parts. The original texture and normal map is then applied back to the individual and finally a bone animation is applied in order to move the character. This process results in the generation of a mesh that can be animated.

## 2 Preliminary plan

Our project will be based on these two applications. Our goal is to try and create more interesting images and patterns which will then be animated. In other words, our plan is to divide the process into two parts: style transfer and image animation.

We will firstly transfer the style from the style image to content image including humans, i.e., "Mona Lisa", by using the algorithms in [4] and VGG-Network [6]. Moreover, we will introduce more interesting techniques to enhance the image quality and stability.

Once we get the stylised images, we will pre-process them before we jump into next step, which is animating the human subject in the stylised images. The state-of-the-art algorithms will be applied to perform human detection, segmentation as well as two-dimensional pose estimation. A rigged mesh will be created afterwards [8] so that the animation of these rigged meshes can be handled in a more convenient manner using any 3D motion sequence. More precisely, we are planning to use mask R-CNN [5] for human detection and segmentation (implementation by [1]). We will estimate body pose by using the algorithms in [7], and human segmentation is refined (some algorithms we would like to explore and try more). Once we segment out the person in our stylised image, we will perform PatchMatch [2] to fill in the regions where the person, i.e., Mona Lisa, used to be.

## References


[1] R. Anantharaman, M. Velazquez, and Y. Lee, *Utilizing mask r-cnn for detection and segmentation of oral diseases*, in 2018 IEEE international conference on bioinformatics and biomedicine (BIBM), IEEE, 2018, pp. 2197–2204.

[2] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, *Patchmatch: A randomized correspondence algorithm for structural image editing*, ACM Trans. Graph., 28 (2009), p. 24.

[3] L. Gatys, A. Ecker, and M. Bethge, *Texture synthesis and the controlled generation of natural stimuli using convolutional neural networks (2015)*, CoRR abs/1505.07376.

[4] L. A. Gatys, A. S. Ecker, and M. Bethge, *A neural algorithm of artistic style*, arXiv preprint arXiv:1508.06576, (2015).

[5] K. He, G. Gkioxari, P. Dollár, and R. Girshick, *Mask r-cnn*, in Proceedings of the IEEE international conference on computer vision, 2017, pp. 2961–2969.

[6] K. Simonyan and A. Zisserman, *Very deep convolutional networks for large-scale image recognition*, arXiv preprint arXiv:1409.1556, (2014).

[7] S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, *Convolutional pose machines*, in Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2016, pp. 4724–4732.

[8] C.-Y. Weng, B. Curless, and I. Kemelmacher-Shlizerman, *Photo wake-up: 3d character animation from a single photo*, in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 5908–5917.