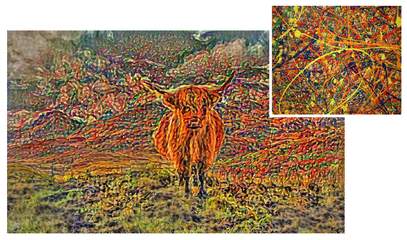
Localized Style Transfer With Semantic Segmentation

Alex Wells, Jeremy Wood, Minna Xiao

INTRODUCTION

With the seminal 2015 publication of “A Neural Algorithm of Artistic Style,” Gatys, et al introduced the application of CNNs to the problem of image style transfer. Style transfer involves the transfer of the style of an image to another, while preserving the content of the target image. Research using Gatys, et al’s algorithm, and subsequent improvements, have produced impressive results of images in the rendering of various artistic styles, from Van Gogh to Kandinsky. Now, with the popularity of Facebook’s mobile app Prisma, which offers style transfer for photos - and real-time transfer of live video, style transfer has been brought to the masses.

Original image of a cow.



Gatys, et al style transfer of Pollock’s *Jump In*.

PROBLEM

Existing research and applications on style transfer thus far have primarily focused on style transfer onto the entire image. We would like to focus instead on style transfer onto specific parts of an image, using semantic segmentation. This approach allows for the selection of specific regions in the original image to be altered, while the rest of the original image maintains its appearance.

DATA

To generate the masks needed for our localized style transfer, we use a CRF-RNN network (condition random fields as recurrent neural networks) for semantic image segmentation. We trained this network on the PASCAL-VOC dataset, which consists of 20 classes. The style transfer implementation uses a pretrained VGG-19 network trained on ImageNet.

METHODS AND RESULTS

Blending with Markov Random Fields

Markov Random Fields describe a markov relationship between nodes in a random field. These fields can be used in image processing to blend two images together (i.e. a style transfered image and the original) such that they minimize a unary constraint on pixels and a binary relationship between pixels. In our case, we first determined a border region on the edge of the segmentation mask and then used a unary cost that encouraged pixels along the border to be assigned to foreground or background based on proximity and then applied a binary cost that minimized the resulting contrast within the border region after the masked style transfer has been applied to the original image (anti-aliases it). Thus pixels in the border region adapt a foreground or background label based on what creates the smoothest blend between stylized image and original image.

Naively, this approach can be used on a fully stylized image (no segmentation), the segmentation mask, and the original image to blend them all together. However we also applied this approach to our masked style loss approach to further revert the non-masked portion of the image to the original content.

$$U(p,l) = \|p - c^l\|. \qquad B(p_1,l_1,p_2,l_2) = |I_{l_1}(p_1) - I_{l_2}(p_1)|^2 + |I_{l_2}(p_2) - I_{l_1}(p_2)|^2.$$

Naive masked style transfer.

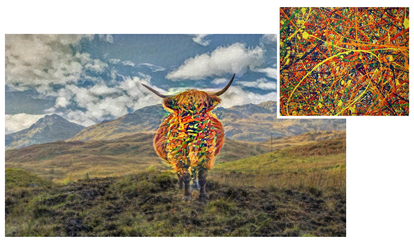
MRF blending with only unary costs.

Full MRF blending.

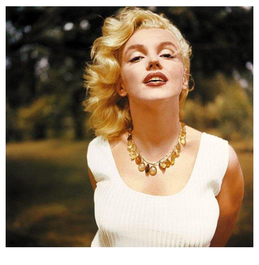
Masked Style Loss

The algorithm introduced by Gatys et al minimizes a custom loss function that sums over a style loss, content loss and total-variation loss: $L_{total} = \alpha L_{style} + \beta L_{content} + \gamma L_{tv}$. In order to achieve style transfer on a portion of the image rather than the entire image, we implement a modification to the style loss calculation, where the original style loss calculation is the sum of the style losses for a set of layers $\mathcal{L}$ (conv1_1, conv2_1, conv3_1, conv4_1, conv5_1): $L_{style} = w \sum_{\ell \in \mathcal{L}} (G^{\ell} - A^{\ell})^2$. At each layer, for the current image *x* and the source style image *s*, we have feature maps F^{ℓ}_x and $F^{\ell}_s \in \mathbb{R}^{WxHxD}$ that represent the activations of the *D* filters for each of the spatial positions in *x* and *s*. Prior to computing the Gram matrices representing the feature correlations for *x* and *s*, we create a mask volume $mask \in \mathbb{R}^{WxHxD}$ from the binary mask generated by our segmentation system that we apply to the feature maps in order to mask over the spatial extent of each filter in the volume.

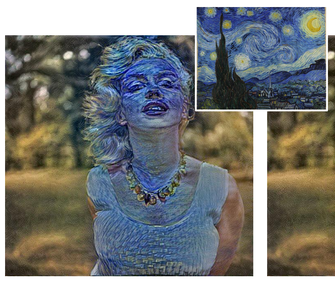
Segmentation mask generated by CRF-RNN.



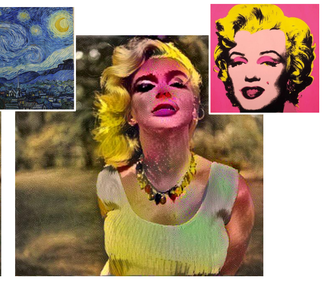
Localized style transfer of Pollock’s *Jump In*, using masked style loss.



Original image of Marilyn Monroe.



Localized style transfer of Van Gogh’s *Starry Night*.



Localized style transfer of Warhol’s *Marilyn Monroe 31*.

FUTURE WORK

In order to improve upon our segmentation pipeline, we would like to implement the strategy outlined in “Mask R-CNN” by Kaiming He et al. This framework has achieved state-of-the-art results on the COCO 2016 challenge, and would bolster our ability to extract regions of interest in our input images, particularly from more challenging images. We also plan on performing a more thorough fine-tuning of the content and style loss hyperparameters in order to improve our current style transfer results.