Neural Style Transfer for Text

Cameron Ibrahim, Luca Leeser, Ryan Butler, Yuji Akimoto

October 22, 2017

1 Research Context

Neural Style Transfer is a technique conventionally used to stylize images, introduced by Gatys et. al. in 2015. It was observed that due to the layered structure of Convolutional Neural Networks, later layers in the network produce feature representations of the content of the image, while correlations between features give a good representation of artistic style. Using a loss function with a content term and a style term, this algorithm is able to iteratively produce an image that combines the content of one image with the style of another.

Figure 1: Examples of the algorithm by Gatys et. al. Image A is the content image; B,C,D are stylized images with the style image in the lower left corner.

As shown above, this has proven very successful for images, with variations that provide improvements on speed and quality. However, attempts at style transfer for text have been limited and have not been as successful. We plan to use the concepts from neural style transfer for images to expand on existing research in writing style conversion.

If successful, our algorithm will be able to convert a Yelp review from one writing style to another. This can be used to convert reviews into a style that is more likely to be useful to other users. This is beneficial to Yelp as more people are likely to use their service if reviews are useful, and to individual users as they do not have to put as much effort into writing eloquent reviews. This would also ensure that reviews are judged on the basis of their content instead of preferences over writing style.

2 Research Objective

Our objective is to design an algorithm that conducts style transfer for text and implement it in TensorFlow. We will develop a discriminative model to do this, first, and if time permits we will then develop a generative model to improve the speed of the style transfer.

3 Research Road Map

3.1 Experiment Design

Implementations of neural style transfer for images use the VGG-19 convolutional neural network architecture to determine both the content and style of an image. We will need to design and train a classifier or encoder-decoder architecture to replace this when conducting style transfer for text. VGG-19 has several important properties that are critical for style transfer:

- It can classify thousands of different objects, which means it has some way to extract useful features from a large variety of images. This is important because it means useful features can be found on images with a variety of styles and content.
- The earlier layers of the network identify lower level features like edges and colors. This makes them relevant for detecting style, as similarly styled images have similar relationships between the feature maps that these layers create.
- The latter layers of the network identify higher level features that actually describe the content of the image. This makes them relevant for identifying the content of the image.

These properties of VGG-19 are essential for a style transfer task, as the features in its layers are used in the style transfer process. Hence, a convolutional network with similar properties must be identified for text if our approach is to be successful. To identify such a network architecture, we will do the following:

- We will determine if our designed architecture encodes meaning by reconstructing the original text from the feature representations of various layers, especially the later layers. This is performed by generating a white nose image and then performing gradient descent on the image values so that the feature representation in the selected layer matches the original feature representation. We will qualitatively assess whether meaning has been preserved, and can also use quantitative measures such as BLEU score.
- We will determine if our designed architecture encodes style (independent of meanining) in a similar manner, except we use the Gram matrix of the selected layer to do our comparisons between the original image and the style reconstruction instead. We will then qualitatively determine if writing style is preserved in a manner independent of the meaning.

Once such an architecture is identified, developing a descriptive model for style transfer in text should follow using the same techniques as those used in neural style transfer for images.

3.2 Research Timeline

3.2.1 Descriptive Model

The initial goal of our research is to create a descriptive model in the style of that created by Gatys et. al. to be used for text style transfer. We have already outlined our approach to make such a model, so the following is a timeline to perform those tasks.

Encoding Meaning The first hurdle is to identify, implement, and test potential architectures of the use of encoding pieces of text while paying respect to meaning. Significant emphasis will be placed on autoencoders and classifiers during this portion of the research. Success will be determined qualitatively by checking that words deemed to have similar meanings will have similar embeddings. Each of the four members of this group will be responsible for selecting such a model to prepare for presentation on **29 Oct.**

Of these candidate models, those with the strongest theoretical ability to encode meaning will then be used to assess the viability of text reconstruction using neural networks. We will implement the model most likely to succeed first, and then evaluate it using the criteria and methodology outlined previously. If we find a suitable model, we can continue on to the next step, otherwise we will build and evaluate the next candidate model. If we cannot find a viable model, we will at a minimum have the ability to write about what the different layers in the architectures that we looked at are doing to properly function as classifiers/autoencoders. This completion of this objective is vital to the feasibility of creating a descriptive model for text; as such, significant time will be given to this task, with a deadline set for **12 Nov.**.

Style Representation Assuming we find a viable network architecture that we can transfer features from, the next significant hurdle will be to find an accurate way to represent text style. We will build a descriptive model for neural style transfer in a similar fashion to the research done for images. Further research will then need to be made on the use of Graham matrices to represent style in text, as they are likely to be the most common method for doing so with images. With this, we have what is needed to create a descriptive model. We will then use highly stylized text (e.g. Shakespeare, Lewis Carol, Dr. Seuss) to subjectively determine whether content and style has successfully been transferred. The deadline for the completion of this objective is **19 Nov.**

3.2.2 Generative Model

Should we succeed in the creation of a descriptive model for style transfer, the final stage is to take the concepts studied (i.e. text embedding, style representation with text) and apply them in the creation of a generative neural style transfer model. As the majority of the components should have been firmly established during the descriptive model stage, this should be completed if time permits, ideally by **26 Nov.**

3.3 Resources

Our primary resource is the Yelp review dataset. As writing style may not differ significantly across the Yelp dataset, we will also make use of Sparknotes' modern translations of Shakespeare's works and other examples of highly stylized text. We will also need several candidate neural architectures to evaluate for style representation and content representation as outlined previously. We will need significant computational resources, such as GPUs, to be able to train and evaluate enough models within the timeframe given.

References

[1] Leon A. Gatys, Alexander S. Ecker, Matthias Bethge. *A Neural Algorithm of Artistic Style*. CoRR, 2015.

[2] Yongcheng Jing, Yezhou Yang, Zunlei Feng, Jingwen Ye, Mingli Song. *Neural Style Transfer: A Review*. CoRR, 2017.

[3] Justin Johnson, Alexandre Alahi, Fei-Fei Li. *Perceptual Losses for Real-Time Style Transfer and Super-Resolution*. Stanford University.

[4] Chuan Li, Michael Wand. *Precomputed Real-Time Texture Synthesis with Markovian Generative Adversarial Networks*. CoRR, 2016.

[5] Se Wong Jang, Jesik Min, Mark Kwon. *Writing Style Conversion using Neural Machine Translation*. Stanford University.

[6] Tianxiao Shen, Tao Lei, Regina Barzilay, Tommi Jaakkola. *Style Transfer from Non-Parallel Text by Cross-Alignment*. CoRR, 2017.

[7] Barak Oshri, Nishith Khandwala. *There and Back Again: Autoencoders for Textual Reconstruction*. Stanford University.