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 these 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.

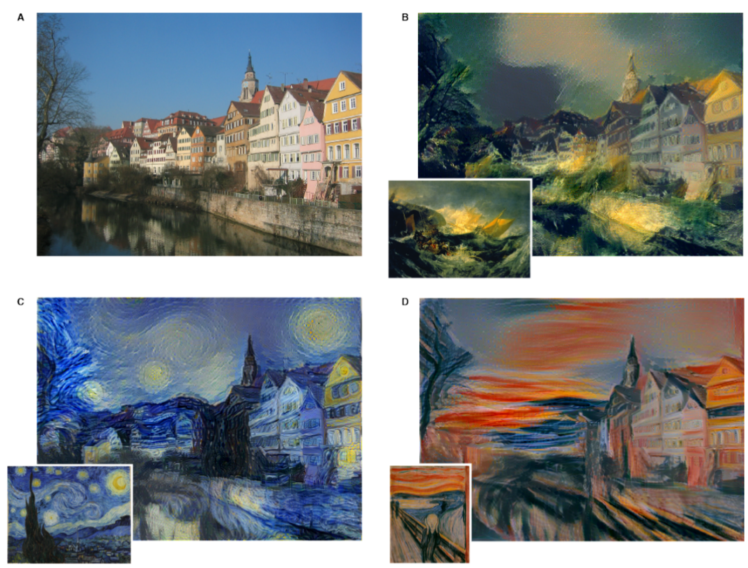

Figure 1: Examples of the algorithm by Gatys et. al. Image A is the content image; B,C,D are stylized images with the style image in the lower left corner.

As shown above, this has proven very successful for images, with variations that provide improvements on speed and preserving facial structures. 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. For our model to be useful, it must conduct the style transfer quickly enough that it is faster than a human re-writing the text manually, so our final implementation should be a generative model as opposed to a descriptive model (details below).

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. We will need to validate the following:

- Whether our designed architecture encodes meaning: we will do this by starting from white noise and iterating until the feature representation in a specific 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.
- Whether our designed architecture encodes style (independent of meaning): again, we will do this by starting from white noise and iterating until we have a style representation that matches the original. We will then qualitatively determine if writing style is preserved in a manner independent of the meaning.
- Whether our algorithm transfers writing style: our final evaluation will be mostly qualitative, and will rely on human judgment to determine whether the meaning of the original review was preserved, and the style was changed.

3.2 Research Timeline

3.2.1 Descriptive Model

The initial goalposts of our research is to create a descriptive model in the style of that created by Gatys et. al. for the use on text style transfer. However, there are a number of tasks that must be completed in order to first determine the feasibility of our goal, let alone implementation of our final model.

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 building such a model, selecting a single one to prepare for presentation on **29 Oct.**

Of these candidates, those with the strongest ability to encode meaning will then be used to assess the viability of text reconstruction using neural networks. 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 The next significant hurdle will be to find an accurate way to represent text style, likely using gradient descent to reconstruct original text and checking layer by layer to determine which layers best code for style in the piece of text. 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 as been preserved. 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 by at latest **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

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.