

On the Use of K-Competitive Networks for Writing Style Transfer

Luca Leaser
Cornell University
ll698@cornell.edu

Ryan Butler
Cornell University
rjb392@cornell.edu

Yuji Akimoto
Cornell University
ya242@cornell.edu

Cameron Ibrahim
Cornell University
cai29@cornell.edu

ABSTRACT

Writing style is a key component of the quality of any on-line review - a detailed and informative review is only useful insofar as it is able to retain a reader's attention. In this paper, we explore the use of k -competitive layers in convolutional neural networks to classify the author of a given piece of text. We then apply the techniques used in neural style transfer for images to transfer the writing style of one author onto the content written by another. Given sufficient examples of well-written reviews, our techniques can be used to improve the quality of reviews on online platforms such as Yelp or Amazon.

1. INTRODUCTION

Neural style transfer [5] has proven very successful at transferring the artistic style of an image onto the content of another. Due to the layered structure of convolutional neural networks, it was observed that later layers in the network produce feature representations of the content of an image, while the correlations between these features give a good representation of artistic style. More specifically, given a content image and style image, the algorithm treats the input to the network as a variable and simultaneously minimizes the ℓ_2 norm between the input and content image's feature representations, and the input and style image's style representations.

Figure 1: Cornell's Gates Hall stylized in the form of Vincent Van Gogh's *Starry Night*.



Modifications that optimize computational time [8], preserve the texture of the style image [6], and preserve facial structures in the content image [10], among others, have further advanced the art of neural style transfer for images. However, applying similar techniques to text has proven difficult for a number of reasons. Firstly, textual data lives in

an extremely sparse space, and therefore it is difficult to iteratively produce better stylized text while avoiding phrases that make no sense. Secondly, style and content are much more closely linked in text than they are in images - the same writer may write two passages of text in a completely different tone, yet maintain the same underlying style in a manner that is very subtle.

In this paper, we explore the use of k -competitive convolutional layers, a generalization of k -competitive layers designed for autoencoders in [2], which in turn are based on k -sparse layers as defined in [15]. These layers are integrated into a word-level convolutional neural network [11], and we perform a qualitative analysis of the quality of content and style representations generated by these networks. A key benefit of our method is that it does not require a training set of translations between writing styles, which are very difficult to obtain.

2. RELATED WORK

Early methods at quantitatively determining the writing style of a piece of text focused on statistical analyses of the rates of passive and active sentences, the ratio of pronouns to nouns to adjectives, and the likelihood of adjacent words given their distribution in the English language. However, these methods rely on unreliable tokenizations and are difficult to adapt to style transfer. Xu et. al. [19] provided some of the first algorithmic style transfer methods between Shakespearean and modern English, but relied on a phrase table and did not take advantage of modern machine learning methods.

Recent methods based on neural networks fall into two broad categories: those that treat the problem as a monolingual machine translation problem, and those that attempt to generate style and content representations of text. [1] and [7] fall under this first category, and generate stylized text that perform well according to BLEU score [16]. However, these translation-based methods require a large dataset of tupled phrases that are identical in content and differ only in style. Use of passages in the Bible in [1] and SparkNotes' Shakespearean translations in [7] present interesting but impractical test cases, while generating rap lyric translations with Google Translate via French [7] leaves much to be desired in terms of accuracy.

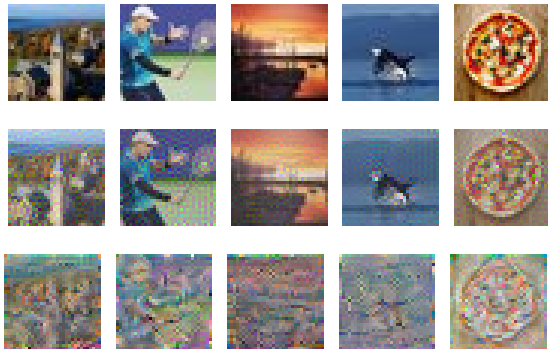
[3] and [9] both consider a similar method to ours in terms of generating style and content representations of text, and

[9] defines three evaluation criteria: soundness (preservation of content), coherence (grammatical correctness & common sense), and effectiveness (matches intended style). However, these lack sufficient experiments with which to evaluate the quality of their results. [4] provides a quantitative assessment of their content and style representations, in addition to a more complex alternative architecture. All of the aforementioned works use a recurrent neural network encoder-decoder structure (otherwise referred to as a “seq2seq” model), with either a long short-term memory (LSTM) or gated recurrent unit (GRU) used as the underlying structure. These methods seek to encode the content of the input text into a low dimensional space, and then use the decoder to generate text in a specified style. Our key contribution is in the use of convolutional neural networks, which have proven very successful at tasks such as sentiment analysis [11].

3. PRELIMINARY ANALYSIS

The majority of implementations of style transfer for images use the VGG-19 [18] architecture to generate content and style representations of an image. This network was trained to classify images into one of the one thousand classes of the ImageNet challenge dataset [17] and therefore is able to generate informative feature representations of a very large variety of objects. Given the resources required to train a similarly robust classifier for text, we first explore the tradeoffs of using an underlying network that is trained to discriminate between a much smaller number of classes by training a VGG network with the CIFAR-10 dataset [12] (here, we use pre-trained weights provided by [13]). Although this modified network is far less generalizable in terms of classification, we found that the content representations provided by later layers of the network were still informative enough to reconstruct the original image, even for images containing objects not included in the 10 classes of CIFAR-10. By reconstruction, we refer to treating the input to the network as a variable, and minimizing ℓ_2 loss between the input image and original image’s feature representations in a certain layer.

Figure 2: Several 32x32 images (top) reconstructed in the conv2.2 layer (middle) and conv4.2 layer (bottom) of VGG-19 trained on CIFAR-10 data; maximum 10000 iterations.



In the reconstruction phase, we use average-pooling in place of max-pooling, as suggested in [5]. Although none of the images in Figure 2 resemble any of the ten classes present

in CIFAR-10 data, the network still produces highly informative feature representations through convolution. This is a key observation in making the transition to text data, as writing style is subjective and therefore it is difficult to collect data from a large number of distinct classes that span the entire space of writing styles.

However, we also observe that this reduction in generalizability of the underlying network architecture is not without tradeoffs. When using the conv4.2 layer at which to match the feature representations of two images (the layer used in Figure 1), the reconstructed images are mostly unrecognizable to the human eye, and takes a much larger number of iterations to produce something even remotely close to the original. While we do not dismiss the possibility that this loss in reconstruction accuracy could be mitigated by hyperparameter optimization, we conclude that high generalizability as a classifier is an important quality of the underlying architecture although not paramount.

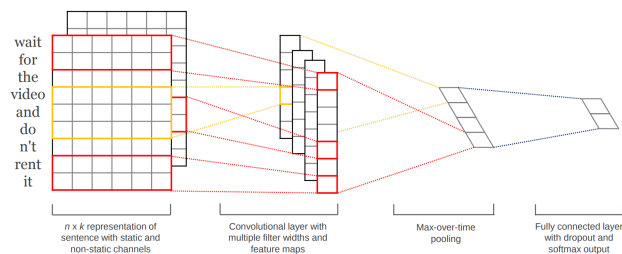
4. MODEL

4.1 Convolutional Neural Networks for Text Classification

Although recurrent neural networks have proven adept at language recognition tasks, evaluating the overall content and style of written text is a task that is less dependent on sequential information, and a hidden state is likely too small to encode enough information for our purposes. We believe that a convolutional neural network allows us to take advantage of the success of style transfer for images, in contrast to aforementioned approaches that treat style transfer as a machine translation problem.

As the baseline for our comparisons we use a word-level convolutional neural network as designed by Yoon Kim in [11]. A 1-d convolutional layer varies from those commonly found in architectures for image classification in that the receptive field of a filter is one-dimensional, and subsequent pooling layers operate along the time axis. The network takes as input a sentence, with each word encoded as a vector. As a “shallow” architecture, the network performs layer of convolution with filter lengths of 3, 4, 5 with 100 feature maps each, before a max-over-time pooling operation and a softmax layer.

Figure 3: Structure of the baseline convolutional neural network model.



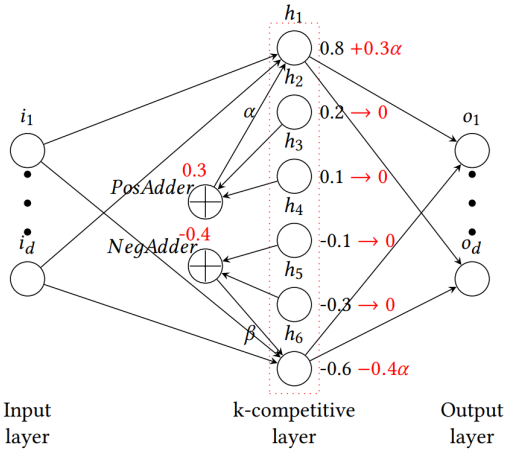
4.2 K-Competitive Convolutional Layers

It has been observed that encouraging sparsity when learning representations acts as a capable regularizer that improves performance on classification tasks. In [15], Makhzani and Frey designed the k -sparse autoencoder and demonstrated its effectiveness as a preprocessing step on image

data. In the feedforward phase, sparsity is used as the sole non-linearity: after computing the activations $z = W^T x + b$, only the largest k entries of z are kept and the rest are set to 0.

Since textual data often has high-dimensionality, sparsity, and power-law distributions [2], autoencoders for text have often learned trivial features such as proper nouns specific to a certain passage. Chen and Zaki made a modification to k -sparse layers by introducing competition in the place of truncation: after computing the activations $z = \tanh(W^T x + b)$, only the k entries with the largest absolute value are kept non-zero. These k most “competitive” neurons are then augmented (with an amplification constant) with the **energy** of the “losers” that were made inactive, where the energy of a subset of neurons H is defined as $\sum_{h_i \in H} |z_i|$.

Figure 4: A fully connected k -competitive layer.



Here, we propose a generalization of k -competitive layers to convolutional layers. Let z_{ijk} be the activation volume produced by a convolutional layer, where $z_{x,y,:}$ denotes the feature vector at spatial location (x, y) . Then, we define the **energy map** of an activation volume, E , as the matrix with entries:

$$e_{ij} = \sum_{k=1}^K |z_{ijk}|$$

We then perform a k -competitive non-linearity, where we select the k most competitive entries, and masking the activations of the “loser” neurons while augmenting the activations of the “winner” neurons. If we denote the set of indices of the k “winner” neurons by C_k , then:

$$kcomp(z)_{i,j,:} = \mathbf{0} \quad \forall (i, j) \notin C_k$$

$$kcomp(z)_{i,j,:} = z_{i,j,:} + \alpha \sum_{(i,j) \notin C_k} z_{i,j,:} \quad \forall (i, j) \in C_k$$

where α , the amplification constant, is a hyperparameter.

5. EXPERIMENTS

5.1 Datasets

We use the most prolific reviewers on Yelp¹ to generate a large amounts of text for various styles of writing. Although

¹<https://www.yelp.com/dataset/challenge>

the reviews for a frequent user of Yelp may span a number of years, we make the assumption that a user’s writing style remains constant over that period of time. This may not strictly be the case as language, particularly of the colloquial variety that is often found online, evolves over time and certain words serve as clear indicators of certain time periods. However, dealing with a time-varying vocabulary set is an extension that we have not yet considered. After identifying the most frequent reviewers, we tokenize all of their reviews using NLTK [14] to generate a corpus.

5.2 Comparison of Architectures

5.3 Hyperparameters and Training

6. RESULTS

7. CONCLUSION

8. FUTURE EXTENSIONS

9. ACKNOWLEDGMENTS

The authors would like to thank Professor Thorsten Joachims at the department of Computer Science at Cornell University for his guidance, and their fellow members of the Cornell Data Science NLP Research Team for their support.

10. REFERENCES

- [1] K. Carlson, A. Riddell, and D. N. Rockmore. Zero-shot style transfer in text using recurrent neural networks. *CoRR*, abs/1711.04731, 2017.
- [2] Y. Chen and M. J. Zaki. Kate: K-competitive autoencoder for text. *KDD*, 2017.
- [3] T. Edirisooriya and M. Tenney. Applying artistic style transfer to natural language.
- [4] Z. Fu, X. Tan, N. Peng, D. Zhao, and R. Yan. Style transfer in text: Exploration and evaluation. 2017.
- [5] L. A. Gatys, A. S. Ecker, and M. Bethge. A neural algorithm of artistic style. *CoRR*, abs/1508.06576, 2015.
- [6] L. A. Gatys, A. S. Ecker, and M. Bethge. Texture synthesis and the controlled generation of natural stimuli using convolutional neural networks. *CoRR*, abs/1505.07376, 2015.
- [7] S. W. Jang, J. Min, and M. Kwon. Writing style conversion using neural machine translation. 2016.
- [8] J. Johnson, A. Alahi, and F. Li. Perceptual losses for real-time style transfer and super-resolution. *CoRR*, abs/1603.08155, 2016.
- [9] J. Kabbara and J. C. K. Cheung. Stylistic transfer in natural language generation systems using recurrent neural networks. 2016.
- [10] P. Kaur, H. Zhang, and K. J. Dana. Photo-realistic facial texture transfer. *CoRR*, abs/1706.04306, 2017.
- [11] Y. Kim. Convolutional neural networks for sentence classification. *CoRR*, abs/1408.5882, 2014.
- [12] A. Krizhevsky. Learning multiple layers of features from tiny images. Technical report, 2009.

- [13] S. Liu and W. Deng. Very deep convolutional neural network based image classification using small training sample size. 2015.
- [14] E. Loper and S. Bird. Nltk: The natural language toolkit. In *In Proceedings of the ACL Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics. Philadelphia: Association for Computational Linguistics*, 2002.
- [15] A. Makhzani and B. J. Frey. k-sparse autoencoders. *CoRR*, abs/1312.5663, 2013.
- [16] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu. Bleu: a method for automatic evaluation of machine translation. 2002.
- [17] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.
- [18] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.
- [19] W. Xu, A. Ritter, W. B. Dolan³, R. Grishman, and C. Cherry. Paraphrasing for style. 2012.

APPENDIX