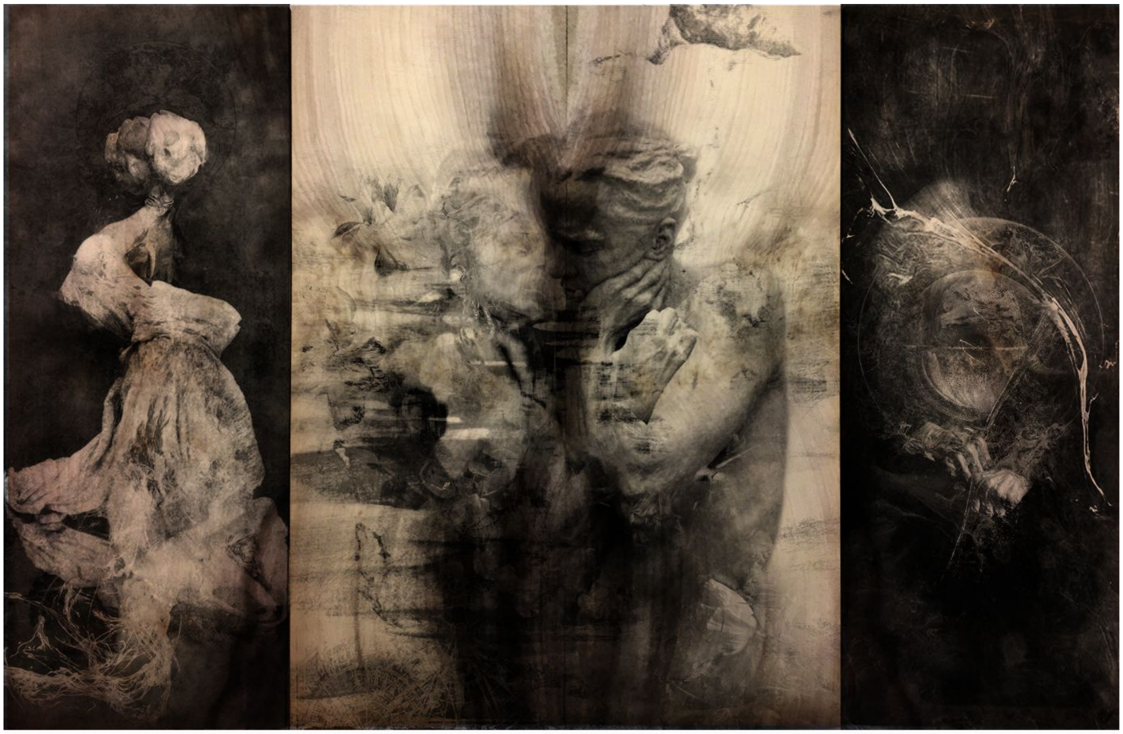
ART AND MACHINE LEARNING
CMU 2019 SPRING
FINAL PROJECT

Poe's Fable, Witkin's Tableau: An ML-aided Synergy



Yixiao Fu, Yi-Chin Lee, Jiyuan Li, Zixin Yao

DESCRIPTION

1 Concept

Author of The Creative Cure, Carrie Barron once remarked, "Novelists or painters are largely solo operators. When it is pure art or self-expression or a deeply original idea that needs to be developed, solitude serves." Yet our project aims to instigate an impossible collaboration across a spatial-temporal spectrum between two prominent artists, Edgar Allan Poe with his dark romanticism literature and Joel-Peter Witkin with his transgressive photographies. A predilection for the macabre in one's work of art would only intensify the aversion toward other ghoulish minds in the sense that the artistic ego loathes similitude. However perfect art as a self defense mechanism, it's the most searingly vivid evidence for one's operational boundary in regards to the material world. ML tools have empowered us to exploit these evidence and catalyze the confluence, regardless of the circuitry of creative currency.

Our process foregrounds not an omnipotent AI artist but a sly yet savvy art persona that knows when to lay back and boss their digital aids around. The most prominent ML techniques employed in this pipeline dealt predominantly with the issue of information overload in the context of art production, TL;DR, FOMO and the sorts. They alleviate the imposed undertakings of artist/archivist, artist/wrangler, artist/arbitrageur, artist/Renaissance Man, artist/scrutinizer as well as reinforce and perpetuate the artist-centered modus operandi, namely, to judge and to fantasize. Our art persona reduces themselves to their coupling of connoisseurship and OCDs, engrossed in an idiosyncratic repetition of reappropriation and internalization.

2 Technique

2.1 VAE + Transformer

The model is a variation of autoencoder (VAE) with an LSTM as the encoder [1] and Transformer [2] as the decoder. In one word, this architecture learns a global representation of each sentence (line of input) so that it models better the style, topic and high-level syntactic structure for the sentence, compared with RNN language model that learns a latent representation for each time step. The encoder outputs a posterior distribution over the latent space $p(z|x)$ so now for each sentence, the z is no longer deterministic. The decoder is a Transformer. A Transformer is described in [2]. It contains Scaled Dot-Product Attention module and Multi-Head Attention module so that in the decoding phase, it queries the previous decoding steps and the encoder output. Tokens are learned embeddings.

2.2 K-Means

Kmeans is an unsupervised clustering method. Given a set of data points, it first randomly initialize the centroids (cluster centers), and then for each data point, it calculates the distance between it and all centroids and it chooses the nearest centroid as its centroid. Then, for each cluster, it recalculates the centroids using the previously assigned data points. It repeats until it converges (e.g. when the clusters do not change any more).

2.3 NN & Distributed NN using efficient data structures

NN (Nearest Neighbor) search aims to find the nearest data point given a data point. The nearest is defined by Euclidean distance in our model. However, if the data size is very large, it is very inefficient to perform linear search (i.e. iterate over all the data points). Therefore, we need some data structures to enable us query the data set in a shorter time than linear but with a high probability to get the right answer. In our implementation, we use K-D tree (K dimensional tree). Each node splits the hyperplane, and left subtree corresponds to left half of the hyperplane, and right subtree corresponds to right half of the hyperplane. Also, we can distribute the data set to different worker nodes to search in parallel. We use celery and rabbitMQ to set up master and workers nodes and let them communicate and do the job in our code (thank you, Prof. Barnabas from 10-605!).

3 Process

3.1 Generating texts

Our data is The Complete Poetical Works of Edgar Allan Poe from Gutenberg (book).

3.1.1 Data Cleaning

We have three steps to clean up the data.

1. We manually delete the content not written by Poe, e.g. some preface, notes, and footnotes by editors and copyright from Gutenberg.
2. We first transform the data into the format of paragraph as a line, and then into sentence as a line, and then we tokenize each sentence to output poe-sent.txt.
3. We limit the vocabulary size by 15000 and 18000, and for each vocabulary size, we try sentence as a piece of data, and 100 and 140 tokens in a line as a piece of data.

3.1.2 Training

We then feed our data into our model and collect the output sentences.

3.1.3 Story composing

We used a VAE to Learn and Model Poe's Anthology and Compose in Poe's Literary Style and then highlight quality snippets in the generated results to make up a tripartite fable and trigger the visual process. The process resembles the practice of master studios nowadays where the masters regulate creative currencies rather than populate the buckets.

3.2 Finding source images for collage

Our next step is to find related source images for collage. Our first attempt is to use Witkin's photography as the sources, but we find those photos already too complicated to be the source so we use images online.

3.2.1 Use K-means to understand Witkin's photography

We use K-means to cluster Witkin's photos hoping to let the model reveal some semantic meanings in the photos not seen by us. We use normalized pixel values as input feature vectors and set the number of clusters from 3 to 10. However, it turns out that the Kmeans can only see some obvious properties of images such as intensity and structure, so we do not get what we want in this algorithm. Also, Witkin's photos are already full of elements so it may be too messy if we use them as source images to collage. Therefore, we look up the Internet to find images.

3.2.2 Use NN to find structurally related source images on Pinterest

We know that we are interested in objects eye, tiger, boats, gay, skull, etc so we start searching them in Pinterest. We tried eye objects. Before we can run NN, we need to have pictures. We crawled Pinterest and using dozens of eye related keywords (e.g. art history eye, bizarre eye and black eyes; the complete list available in github). We finally have ~20k images. We give the model an eye from Witkin as input and hope to retrieve semantically related and aesthetically satisfying pictures from our crawled pictures. We use pretrained VGG16 as the feature extractor. Unfortunately, the retrieved results are not ideal.

3.2.3 Use Pinterest's recommendation

Finally, we resort to using Pinterest's recommendation pictures.

3.3 Collage & Wood Color Printing

We printed the final Collage Images and Leitmotif on a Tripartite Wood Panels Set.

4 Result(next page)

II
Crime Passionnel

As the contour of the second morning arose with stars , their chamber in the fire, like a groan, even more unendurable. The swarming track of the morning had been found by the room. A slight bruise swung his manhood. Twice a lady strangled, her ankles an inch pale. Blackness of madame joyeuse.

Human Interpretation:

The young crony who was promiscuous by nature had an affair with a maritime maiden on board. They committed adultery right in the capitan's chamber bed when he was on a graveyard shift. When he made it back to his chamber by dawn, he found his lover's golden hair coiled up a maiden's fillet of linen. Flabbergasted, the distraught pirate brutally slaughtered the couple and cured the corpses on deck for days.

I
Chemistry

The spirit had taught me that the water is a pirate of guilt, however gay and soft. The gentleman who was merely more than twenty one with golden intense arms, humming imaginative love of this time about their grace.

Human Interpretation:

A pirate in charge of a frigate engaged in a matelotage with one of his young cronies whose physique and youthful charm he immensely adored.

III
Cannibal

His nerves had swallowed his eyes. The yellow haired young lady is not the sequel. He had his majesty, he lavished back. A voyage from the skull into new term to our folly. The parchment gave him a black horizon. Gently vigor of the coffin eyes.

Human Interpretation:

After wreaking blood-soaked vengeance on the couple who wronged him, he embarked on a new journey. He forged his young lover's eyes into a talisman so that he could experience the same nautical scenes with him, who would no longer betray him again.

5 Reflection

5.1 Corpus size for text generator

We thought that the corpus size might be too small compared with the corpus we used in Project 3. However, the results turn out to be good. This is probably due to the consistency in word choices and writing styles by Poe and the fact that we limit the vocabulary size. Remember Zipf' Law tells us that the infrequent types only occupy a very little fraction of our total number of tokens and the top several frequent types dominate the corpus.

5.2 Effectiveness of Kmeans

We only get pixel-level information from Kmeans because we only provide pixel-level information to it. However, if we tailor the labels for the images and train a model on that and then feed the trained features into Kmeans, the clusters will be likely to try to cluster labels, which violates our intention to let the model learn things by itself since it is an unsupervised method.

5.3 Effectiveness of NN

We are sad to see NN do not work either. We have three possible explanations. First, the pretrained model do not have related labels for our eyes images so some high level features in the CNN are not accurate enough. Second, we do not retrieve enough number of images. Third, Pinterest does better than us because they have other features that can help identify similar images such as click-through rate (i.e. if a user clicks two images within a session, then these two images are likely to be similar in a sense), but we do not have that kind of features.

5.4 Future work: Text2Image alternative methods

Text2image is a very difficult task because it is hard to get labeled data and even if we get data, things can be very subjective. That is to say, for the image classification task, we are clear that given a bird image, we output bird label. By contrast, given a piece of text, different people can have different interpretations so it is not even reasonable to give a ground truth image label. On the other hand, image2text is kind of easier because we can have an objective description of an image. However, if we want to be creative then it is another story… We believe in the future works, we need to set an prior for image style. A very rough pipeline is that we can first crawl a lot of contents on the Internet and get (text, image) pairs (which is already difficult enough), and then crawl many artistic images and label them by the artist, and then we learn a mapping from text to image. The models would include feature extractors and generators. We suspect that the dataset size can be extremely ultra large as the key to ensuring the quality of output is the dataset size.

5.5 Future work: Learn similarity between text & image representations

Now, we manually deduce that Poe x Witkin would be an inspiring pair of (writer, painter) / (text, image). An more interesting way is to let ML models find those kind of mappings, which reminds us of Visual Question Answering task because for that, we need both image and text inputs. So attention mechanism might be the key - we can find related regions in images and snippets in texts. The current image models can help us identify the objects in the image, so a more difficult part is to identify some subjective things in the image using texts (e.g. a sad story often yields a very dark or dim picture), and perhaps we need some rule-based feature extractors (like knowledge from art side) also. Actually, this is a very interesting topic!

6 Reference

[1] Bowman, S. R., Vilnis, L., Vinyals, O., Dai, A. M., Jozefowicz, R., and Bengio, S. (2015). Generating sentences from a continuous space. arXiv preprint arXiv:1511.06349.

[2] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. In Neural Information Processing Systems (NIPS), 2017.

CODE: https://github.com/zxyao/late-night-cafe

RESULT: https://github.com/zxyao/late-night-cafe