DICS: Deep Image-to-Comic Synthesis

A Deep Learning Project Proposal

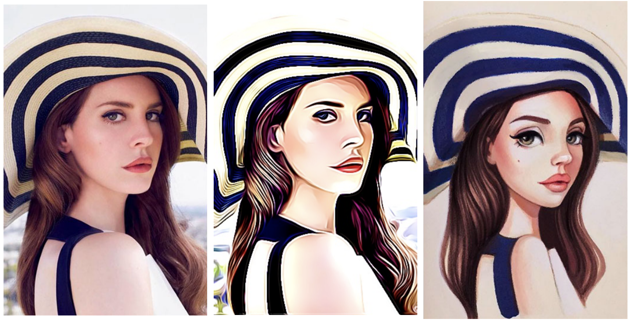
COUVERCHEL, Vincent 2018280584 HEDIA, Mohamed-Laid 2018280581 ZHANG, Naifu 2018280351

Abstract

We investigate the application of variants of Generative Adversarial Network or/and Convolutional Neural Network to perform image-to-image style transfer. Specifically, we attempt to translate real life images of humans to comic or manga pictures.

1. Background

Commercially available apps (BeFunky) exploit the idea of neural style transfer to perform color, texture and painter style transformation to images, à la *Figure 1(b)*. However, these image analogy models do not truly produce comic-style images, à la *Figure 1(c)* (Kiryakova), when applied to human image-to-comic translation.



(a) Input image (b) Existing output (c) Desired output

Figure 1. Comparison of existing and desired human image-to-comic synthesis

We aim to produce output closer to *Figure 1(c)* (Kiryakova). These have not only artistic value but also wide commercial use case such as Instagram filters etc.

2. Literature Review & Theory

Recent researches focus on **Deep Neural Style Transfer** (Gatys et al., 2015) (Dumoulin et al., 2016). Whilst they promise to combine the style of one image with the content of another, these networks only capture low-level texture information and fail to accommodate global shape deformations (Bau et al., 2017).

A promising model is **Generative Adversarial Networks** (GANs) (Goodfellow et al., 2014) that achieves good results in complex domains mapping. A major challenge of such models is that they need input-and-output pairs during training phase, and we do not have such datasets.

3. Plan

Since we do not have paired training data, we will explore **Unsupervised Image Translation** that can generate cycle mapping between inputs and outputs, such mapping can be represented as input→output→input and output→input→output connections. This mapping can be used as training data for GANs - (Gokaslan et al., 2018) describes this model, but provides only a partial implementation.

The model uses a cyclic loss to learn an as much bijective as possible mapping between the inputs and outputs by preserving the most important information.

As the aim is to have convicted results for human sight, a perceptual loss is used to emphasizes shape and appearance similarity.

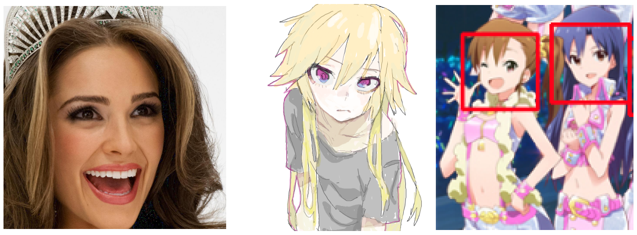
Our main goal is to run the model on different comic styles to achieve the desired outputs. That said, we will also explore improving the model by augmenting it with other techniques, such as **Partial CNN** (Liu et al., 2018) or **Semantic Image Synthesis** (Park et al., 2019) etc.

4. Data

For our work, we need a dataset for human faces and comic faces each.

For human faces, the CelebA-HQ dataset contains more than 200k face image with attributes (Liu et al.).

For anime faces, we will try to build our data set using different mangas (One Piece, DragonBall etc.). Our output will depend on the anime used during the training stage. To build the dataset, we can perform segmentation to extract faces from mangas (Gatys et al., 2017). Should this not work out, we can rely on Danbooru dataset which contains more than 3.3M tagged anime images (Gwern Branwen, 2019)



(a) CelebA-HQ data (b) Danbooru data (c) Our custom data

Figure 2. Example training data

5. Evaluation

We shall first try to **quantitatively** evaluate the model using **multi-scale structure similarity loss (MS-SSIM)** (Wang et al., 2004), which is a oft-cited loss function:

$$\mathcal{L}_{total} = \lambda_{GAN} SLN(\mathcal{L}_{GAN}) + \lambda_{FM} SLN(\mathcal{L}_{FM}) + \lambda_{CYC} SLN(\lambda_{SS}\mathcal{L}_{SS} + \lambda_{L1}\mathcal{L}_{L1})$$

It's ultimately the human viewer's subjective **qualitative** assessment that truly matters.

References

Bau, D., Zhou, B., Khosla, A., Oliva, A., and Torralba, A. Network dissection: Quantifying interpretability of deep visual representations. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul 2017. doi: 10.1109/cvpr.2017.354. URL http://dx.doi.org/10.1109/CVPR.2017.354.

BeFunky. Photo-to-cartoon app. URL https://www.befunky.com/create/photo-to-cartoon/.

Dumoulin, V., Shlens, J., and Kudlur, M. A learned representation for artistic style, 2016.

Gatys, L. A., Ecker, A. S., and Bethge, M. A neural algorithm of artistic style, 2015.

Gatys, L. A., Ecker, A. S., and Bethge, M. Local binary pattern cascadeanime face, 2017. URL https://github.com/nagadomi/lbpcascade_animeface.

Gokaslan, A., Ramanujan, V., Ritchie, D., Kim, K. I., and Tompkin, J. Improving shape deformation in unsupervised image-to-image translation. *Lecture Notes in Computer Science*, pp. 662678, 2018. ISSN 1611-3349. doi: 10.1007/978-3-030-01258-8_40. URL http://dx.doi.org/10.1007/978-3-030-01258-8_40.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative adversarial nets. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'14, pp. 2672–2680, Cambridge, MA, USA, 2014. MIT Press. URL http://dl.acm.org/citation.cfm?id=2969033.2969125.

Gwern Branwen, Aaron Gokaslan, A. t. D. c. Danbooru2018: A large-scale crowdsourced and tagged anime illustration dataset. https://www.gwern.net/Danbooru2018, January 2019. URL https://www.gwern.net/Danbooru2018. Accessed: DATE.

Kiryakova, L. Image to comic drawings. URL https://www.artstation.com/lerika.

Liu, G., Reda, F. A., Shih, K. J., Wang, T.-C., Tao, A., and Catanzaro, B. Image inpainting for irregular holes using partial convolutions. *CoRR*, abs/1804.07723, 2018.

Liu, Z., Luo, P., Wang, X., and Tang, X. Celeba-hq dataset. URL http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html.

Park, T., Liu, M.-Y., Wang, T.-C., and Zhu, J.-Y. Semantic image synthesis with spatially-adaptive normalization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

Wang, Z., Bovik, A., Rahim Sheikh, H., and Simoncelli, E. Image quality assessment: From error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13:600 – 612, 05 2004. doi: 10.1109/TIP.2003.819861.