# Image to Illustration Translation via Generative Adversarial Networks

Mert Canatan

***Abstract*—In this project, I explored illustrations in children's books as a new domain in un-paired image-to-image translation. Current state of the art image-to-image translation models successfully transfer either the style or the content. However, they fail to transfer both at the same time. I addressed to this problem and I achieved better balance between style and content. There are no well-defined or agreed-upon evaluation metrics for unpaired image-to-image translation and so far, the success of the image translation models has been based on subjective, qualitative visual comparison on a limited number of images. To address this issue, I used a new framework [1] for the quantitative evaluation of the image-to-illustration models, where both content and style are taken into account using separate classifiers.**

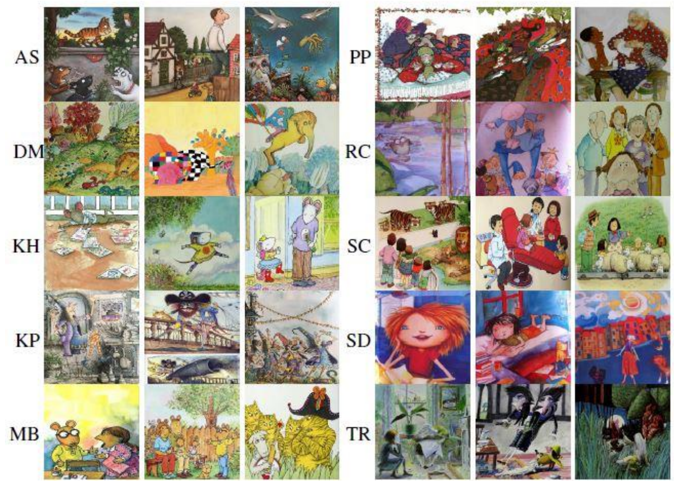
***Index Terms*—Generative Adversarial Networks, Image to Image Translation, Illustrations, Style Transfer**

## INTRODUCTION AND RELATED WORK

Style transfer in image-to-image translation became popular after Gatsy et al.'s precursor work [3]. Various approaches have been derived by researchers for the style transfer problem. These approaches included paired transfer [3, 4], unpaired transfer [5, 6, 7], online methods [3] and offline methods such as Generative Adversarial Networks (GANs) [5, 6, 7, 10] and Convolutional Neural Networks (CNNs) [8]. Style transfer is used in many different applications such as converting natural images to art paintings [5, 9], transferring style of a certain animal to another animal [5], converting weather/season of the scene [10], transferring sketches [11], completing sketches [12], converting face images to sketches [6], and making a cartoon image from an image [7].

This project aims to show a novel style transfer method using illustrations of childen's books (Fig 1.) to create a perfect balance between style and content image-to-image transfer. It is claimed to be called a new domain because illustrations are qualitavely different than cartoon and art paintings. They contain objects like trees, people, etc., and the abstraction level might be higher than cartoons and art paintings. Therefore, existing methods such as CycleGAN [5] might suffer from transferring content or DualGAN [6] might suffer from transferring style.

In the following sections of this report description, evaluation, related work and summary and conclusion will be introduced respectively.




*. Fig 1. Illustration examples in children's books*

## DESCRIPTION

Image-to-illustration tasks that have been done so far showed that there is a problem in transferring the content and the style at the same time [5,6, 7]. In this project, novel generator network GANILLA (Fig. 2) is introduced to handle this problem (Fig 3.). GANILLA structure utilizes low level features to preserve content and at the same time it transfers the style.

There is two stages in this model: the downsampling stage which is a modified ResNet-18. ResNet-18 is modified by concatenating features from previous layers at each downsampling layers and the upsampling stage which is a feeded lower-levedl features using the outputs of each layer in downsampling stage via long, skip connections showed in blue arrows in Figure 3.

The discriminator network in this model is a 70 by 70 PatchGAN [5] which is used for image to image translation models.

While training proposed GANILLA model cycle-consistency idea is followed. First set tries to map source images into a target domain and at the same time the second set takes the input as the target domain and tries to generate sources images and this happens in a cyclic manner.

Loss function consist of two Minimax losses and a cycle consistency loss. Minimax losses are for each Generator and Discriminator pair and cycle consistency loss makes sure that a generated sample could be mapped back to source domain by using $L_1$ norm.

In this project PyTorch is used to implement the models. Models are trained using 200 epoch with Adam solver with a learning rate of 0.0002. All networks are trained from scratch without using any pretrained network weights.

For training, natural images from CycleGAN training dataset [5] is used as source domain and new illustration dataset [1] is used as target domain. In testing, 751 image from CycleGAN test set is used.For the illustration set, the dataset is extended by adding new images. The dataset size is almost doubled after this addition. In this project the used dataset includes approximately 9500 illustrations taken from 363 books and 24 artists [2].

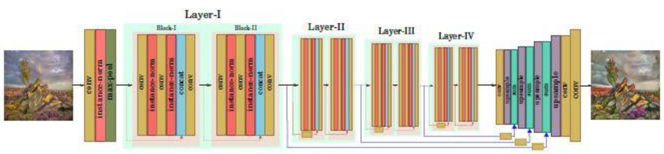



*Fig 2. GANILLA generator network.*


*Fig 3. CycleGAN,CartoonGAN, DualGAN and GANILLA structure in terms of layers. Green color represents convolutional layers, pink color represents residual layers with additive connection, blue color represents deconvolutional layer, purple color represents residual layers with concatenative connection, and the red color represents feature pyramid network.*

## EVALUATION

GANILLA method is compared with the current state of the art GAN methods including CycleGAN [5], DualGAN [6] and CartoonGAN [7]. There is a problem of comparing GAN methods due to lack of quantitative evaluations. In this project a new quantitative evaluation method is introduced [1] to handle this problem. For the qualitative evaluation between CycleGAN, DualGAN and GANILLA Figure 4 can be used as reference. Also the user study which is a survey including 66 people can be found in Hicsonmez et al.'s work [1].

For quantitative evaluation of the quality of style transfer can be done using style classifier. To train a style classifier, training images has to be detached from their visual content while keeping the style. For this basis, illustration images were randomly cropped by small patches and these patches were used to train the style classifier (Style CNN) [1] (Table 1) .

| Style CNN (%) | | | |
|---|---|---|---|
| Cartoon GAN | Cycle GAN | Dual GAN | GAN ILLA |
| 77.8 | **99.2** | 94.8 | 95.3 |
| 64.4 | **88.8** | 0.5 | 81.4 |
| 87.1 | **99.0** | 0.2 | 81.6 |
| 0.7 | **99.2** | 94.1 | 89.6 |
| 0.1 | **94.0** | 1.4 | 67.0 |
| 85.0 | **96.4** | 86.1 | 91.7 |
| 98.9 | **99.8** | 0.1 | 85.6 |
| 92.8 | **99.8** | 1.3 | 59.4 |
| 68.0 | 96.6 | 94.0 | **98.4** |
| 30.9 | **99.3** | 92.0 | 94.7 |
| 60.6 | **97.2** | 46.5 | 84.5 |

*Table 1. Style CNN comparison between CycleGAN, DualGAN and GANILLA*

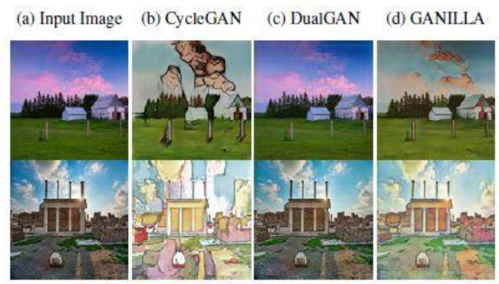



*Fig 4. Comparison of CycleGAN, DualGAN and GANILLA*

## SUMMARY AND CONCLUSIONS

In this project, extensive children's books illustration dataset and novel generator network for image-to-illustration transfer is presented. Due to high level of abstract objects and shapes in illustrations current state of the art GAN structures lack to preserve the content and the style at the same time. This problem is tackled by using a novel GAN structure GANILLA. The main problem while comparing GAN structures lies on lack of the quantitative evaluation of the generator networks. In this project this problem is aimed to tackle by introducing a new quantitative evaluation metric which evaluates both content and style. According to this metric the GANILLA shows the best performance among the existing GAN structures [1].

## REFERENCES

[1] S. Hicsonmez, N. Samet, E. Akbas, and P. Duygulu, "GANILLA: Generative Adversarial Networks for Image to Illustration Translation," arXiv:2002.05638 [cs.CV], Mar. 2020.

[2] S. Hicsonmez, N. Samet, F. Sener, P. Duygulu, Draw: Deep networks for recognizing styles of artists who illustrate children's books, in: ACM International Conference on Multimedia Retrieval, ACM, 2017, pp. 3381–346.

[3] L. A. Gatys, A. S. Ecker, M. Bethge, A neural algorithm of artistic style, arXiv preprint arXiv:1508.06576.

[4] L. A. Gatys, A. S. Ecker, M. Bethge, Image style transfer using convolutional neural networks, IEEE Conference on Computer Vision and Pattern Recognition.

[5] J.-Y. Zhu, T. Park, P. Isola, A. A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, IEEE International Conference on Computer Vision.

[6] Z. Yi, H. Zhang, P. Tan, M. Gong, Dualgan: Unsupervised dual learning for image-to-image translation, IEEE International Conference on Computer Vision.

[7] Y. Chen, Y.-K. Lai, Y.-J. Liu, Cartoongan: Generative adversarial networks for photo cartoonization, in: IEEE Conference on Computer Vision and Pattern Recognition, 2018.

[8] D. Ulyanov, V. Lebedev, A. Vedaldi, V. S. Lempitsky, Texture networks: Feedforward synthesis of textures and stylized images., in: International Conference on Machine Learning, 2016, pp. 1349–1357.

[9] M. Amodio, S. Krishnaswamy, Travelgan: Image-to-image translation by transformation vector learning, in: IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 8983–8992.

[10] L. Karacan, Z. Akata, A. Erdem, E. Erdem, Learning to generate images of outdoorscenes from attributes and semantic layouts (2016). arXiv:1612.00215.

[11] W. Chen, J. Hays, Sketchygan: Towards diverse and realistic sketch to image synthesis, in: IEEE Conference on Computer Vision and Pattern Recognition, 2018.

[12] F. Liu, X. Deng, Y.-K. Lai, Y.-J. Liu, C. Ma, H. Wang, Sketchgan: Joint sketch completion and recognition with generative adversarial network, in: IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 5830–5839.

[13] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Advances in Neural Information Processing Systems, 2014, pp. 2672–2680.