CycleGAN with auxiliary classifier for chinese calligraphy style transfer and font classification

Rui Feng[1,*]

[1]Department of Electronics and Electrical Engineering, Keio University, Tokyo, 223-8522, Japan

*fengruirua@keio.jp

Abstract. As a minor field that remains traditional, few have tried to integrate machine learning with Chinese calligraphy. This ancient form of visual art, on the contrary, is a considerable stage for computer vision. The numerous preservations of Chinese calligraphy make the rich context for machine learning studies that potentially make the field more valuable both aesthetically and pedagogically. This study transfers between stone inscriptions and ink marks, two unique forms of Chinese calligraphy, and classifies the fonts of the pieces. The model of this study is based on the outline of CycleGAN, a powerful generative algorithm for image style transfer. An auxiliary classifier, whose loss is trained together with the GAN loss and the cycle consistency loss, is deployed on the discriminator of CycleGAN, enabling the network to function as a classifier simultaneously. The model succeeds to convert the forms of stone inscriptions and ink marks, with the properties of each style vividly presented. It also successfully optimizes both the discriminator loss and the classification loss, showing the practicability of an auxiliary classifier on CycleGAN. This study points to a decent potential for further combinations of machine learning techniques with Chinese calligraphy studies, to make the job more versatile and detailed.

Keywords: Style Transfer, CycleGAN, Auxiliary Classifier, Chinese Calligraphy.

1. Introduction

The stone inscription (i.e., Bei) and the ink mark (i.e., Tie) are the two fundamental forms of Chinese calligraphy. Stone inscriptions, serving as memorials, are those carved on steles, and ink marks refer to those directly written on paper. Stone inscriptions are relatively less valuable not only because they are easy to preserve, but due to the various distortions they sustain — they can be misinterpreted by the craftsmen who do the carving; they are open to either natural or artificial damage; they lose mostly the rich contrast in inking, the touch of brush pen with paper, and the tenuous strokes. Arguably, stone inscriptions sacrifice some portion of the value of art at the altar of duplicability. Learning stone inscriptions also demands imagination to recover the missed details, which can be challenging for beginners, particularly kids. Based on the above, a stone-inscription-to-ink-mark translator that brings those vivid details back to life should become a contribution to both appreciation and education.

The field of image style transfer, first implemented via Convolutional Neural Network (CNN) [1], has been rapidly improved since the application of Generative Adversarial Networks (GAN) [2]. Variations of GAN such as pix2pix [3], CycleGAN [4], StyleGAN (PGGAN) [5] [6], etc. have successfully elevated the accuracy and the quality of style transfer to a mature level. Image style transfer has been applied to thousands of kinds of tasks, from painting-to-photograph transfer to facial properties switch. However, as the community of Chinese calligraphy, mostly limited in China and Japan, remains a rarefied minority, communication and education have been held out in the very traditional way for decades. There have been efforts in the issue of text recognition, while no attempt has ever been made to figure out what those stone inscriptions would look like if they were written on paper. Additionally, it can be noticed that CycleGAN is mostly used to solve single tasks that only do the transfer at a time. Given that modifications of normal GAN (or even CycleGAN [7]) have tried adding an additional label input [8], or an auxiliary classifier to implement the classification task simultaneously [9], deploying an auxiliary classifier in CycleGAN could be a potential way to expand its function, which few have yet done. One of the current attempts tried to implement an auxiliary classifier in CycleGAN, to reverse and classify the gender of facial photographs at the same time [10]. But it only distinguishes between the input and the output, not any extra information.

In this research, the dataset consists of square images cropped from the scans of celebrated stone inscriptions and ink marks, each containing 1-2 Chinese characters. The fonts include Regular Script (Kai), Clerical Script (Li), Running Script (Xing), and Cursive Script (Cao). This study uses CycleGAN to transfer from stone inscriptions to ink marks. An auxiliary classifier is attached to the discriminator so that the discriminator simultaneously judges and labels the output of the generator. Note that the classifier does not distinguish between stone inscription and ink mark, which are originally included in the input and output; it tells which kinds of fonts the given images are in. Consequently, the GAN loss, the cycle consistency loss, and the classification loss are integrated as the overall loss function to be optimized. As a result, the CycleGAN properly transferred the color, realistically recovered the contrast in the strength of inking, and to some extent successfully reflected the quality of the paper. The classifier successfully accomplished the auxiliary classification task.

2. Methods

2.1. Dataset description and preprocessing

The dataset of this research is collected from the online sources of the scans of celebrated calligraphy masterpieces (**Figure 1**). Both the datasets of stone inscriptions and ink marks consist of parts from 10 different pieces dating from Eastern Han (25 A.D. – 220 A.D.) to Qing Dynasty (1636 A.D. – 1912 A.D.). The dataset of stone inscriptions includes 1,074 full-color square pictures cropped from the scans, which are later regularized in 256×256-pixel size. 1,000 of them are demarcated as the training set and the rest as the testing set. The dataset of ink marks includes 1, 200 pictures of the same size, with 1,000 as the training set and 200 as the testing set.

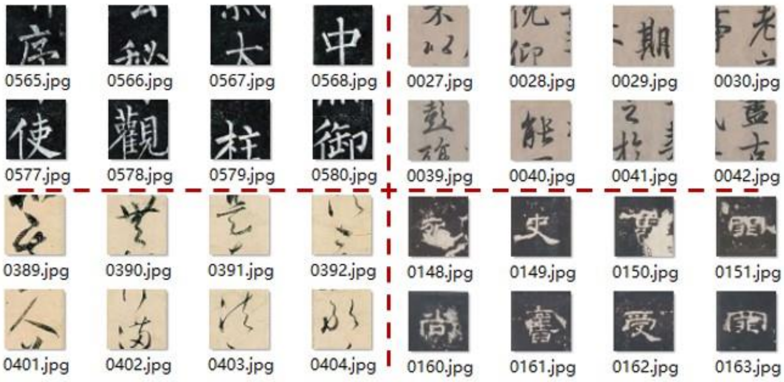

Figure 1. Samples from the datasets of stone inscriptions (upper left and lower right) and ink marks (upper right and lower left) that consist of four types of fonts: Kai (upper left), Xing (upper right), Cao (lower left), and Li (power right).

Each square picture contains around 1-2 Chinese characters, some being full while others being incomplete. The quality of the data varies, for that (i) the dataset contains preservations from ancient ages that have suffered great damage, and that (ii) some scans with relatively low resolution would undergo distortion when regularized into 256*256-pixel size. The whole dataset can be generally divided into four types of fonts. In the training sets, 245 of ink marks and 649 of stone inscriptions are in Kai; 534 of ink marks and 0 of stone inscriptions are in Xing; 221 of ink marks and 53 of stone inscriptions are in Cao; 0 of ink marks and 298 of stone inscriptions are in Li. Note that sub-categories such as Xiao-Kai vs Da-Kai, or Xiao-Cao vs Da-Cao are included but not considered this time.

2.2. Proposed approach

This study chose CycleGAN as the model to achieve the goal based on the fact that Chinese calligraphy pieces are mostly unpaired. Writing on stone and paper are considered different media for expression, serving different purposes. Therefore, the same text is rarely written both on stone and on paper.

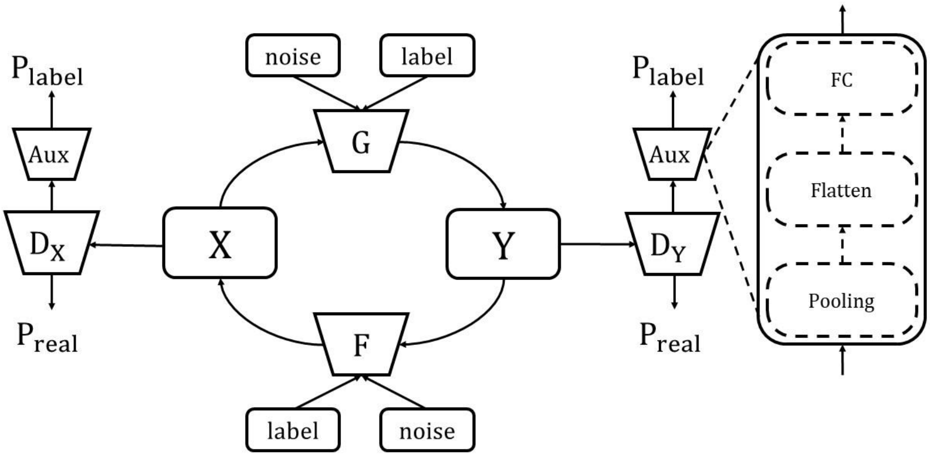

Figure 2. The architecture of CycleGAN with auxiliary classifier. An additional input "label" is given, and the discriminator is connected to the auxiliary classifier.

CycleGAN consists of two sets of GANs that are linked into a cycle. Two domains X and Y are mapped into each other by the two generators G and F, respectively. The two discriminators D_X and D_Y, respectively, judge whether the output is real or fake. Accordingly, one part of the loss of CycleGAN is the regular GAN loss, describing how well the generator tries to mimic the real dataset, and how well the discriminator distinguishes between the real dataset and what the generator fakes:

$$L_{\text{GAN}}(G, D_Y, X, Y) = \mathrm{E}_{y\sim p_{data}(y)}[\log D_Y(y)] + \mathrm{E}_{x\sim p_{data}(x)}\big[\log(1 - D_Y\big(G(x)\big)\big] \tag{1}$$

$$L_{\text{GAN}}(F, D_X, Y, X) = \mathrm{E}_{x\sim p_{data}(x)}[\log D_X(x)] + \mathrm{E}_{y\sim p_{data}(y)}\big[\log(1 - D_X\big(F(y)\big)\big] \tag{2}$$

With the regular GAN loss, the generators can be trained to transfer the style of one domain to something very similar to the other. However, since wildly generated results are not desired, in another word, the main body of domain X needs to be to the greatest extent preserved, the mapping into domain Y should be mapped into domain X again (X') to check if it corresponds to the original. The difference between X and X' is described as the cycle consistency loss:

$$L_{\text{cyc}}(G, F) = \mathrm{E}_{x\sim p_{data}(x)}\big[\big\|F\big(G(x)\big) - x\big\|\big] + \mathrm{E}_{y\sim p_{data}(y)}\big[\big\|G\big(F(y)\big) - y\big\|\big] \tag{3}$$

Inspired by Conditional GAN (cGAN) and GAN with auxiliary classifier (ACGAN), this research seeks to implement the auxiliary classifier on CycleGAN, optimizing the classification loss together with the GAN loss and the cycle consistency loss, instead of training a new classification network. The **Figure**

2 presents the architecture of the proposed method. The font type label is given as an additional input. The discriminator is connected to an auxiliary classifier which includes a pooling layer, a flatten layer, and a fully connected layer. This way, the discriminator is able to classify the font of the calligraphy simultaneously.

The total loss (full objective) is expressed as:

$$L(G,F,D_X,D_Y,cls)=L_{\mathrm{GAN}}(G,D_Y,X,Y)+L_{\mathrm{GAN}}(F,D_X,Y,X)+L_{\mathrm{cyc}}(G,F)+L_{\mathrm{cls}}(cls) \tag{4}$$

2.3. Implementation details

In this research, the following setting of hyperparameters shown in Table 1 is adopted:

Table 1. Hyperparameters of the model.

Hyperparameter	Value
batch size	1
initial gain	0.015
identity loss weight	0.01
learning rate	0.0002
number of epochs with a constant learning rate	100
number of epochs with learning rate linearly decaying to 0	100
number of generator filters in the last conv layer	64
number of discriminator filters in the last conv layer	64
generator architecture	resnet 9 blocks
number of layers in the discriminator	3

As the GPU used has a VRAM of 16GB, the max affordable batch size is 4 based on the 256*256 data size. To insure more fine details transferred, the batch size is set to 1. The identity loss is designed to ensure that when domain Y itself goes through generator G, nothing would be changed. Without identity loss, the generator can change the color arbitrarily. Thus, in term of Chinese calligraphy, which almost covers merely 2 colors, the identity loss weight should be dealt carefully to prevent the color from being abnormally changed.

To record the training process, the 9 types of losses (generator loss A, B; discriminator loss A, B; cycle consistency loss A, B; identity loss A, B; classification loss) of each epoch are plotted, and the transferring results of each epoch are downloaded. After finishing the 200 epochs of training, the generator is used to transfer the testing set.

3. Result and discussion

3.1. Training loss over time

The 9 types of losses are monitored over 200 epochs and exported individually. A part of the results is shown on Figure 3. The classification loss and the discriminator loss decreased rapidly in 20 epochs and converged at around epoch 40. The overall loss of the model also converged at around epoch 40, and the performance kept stable after epoch 120.

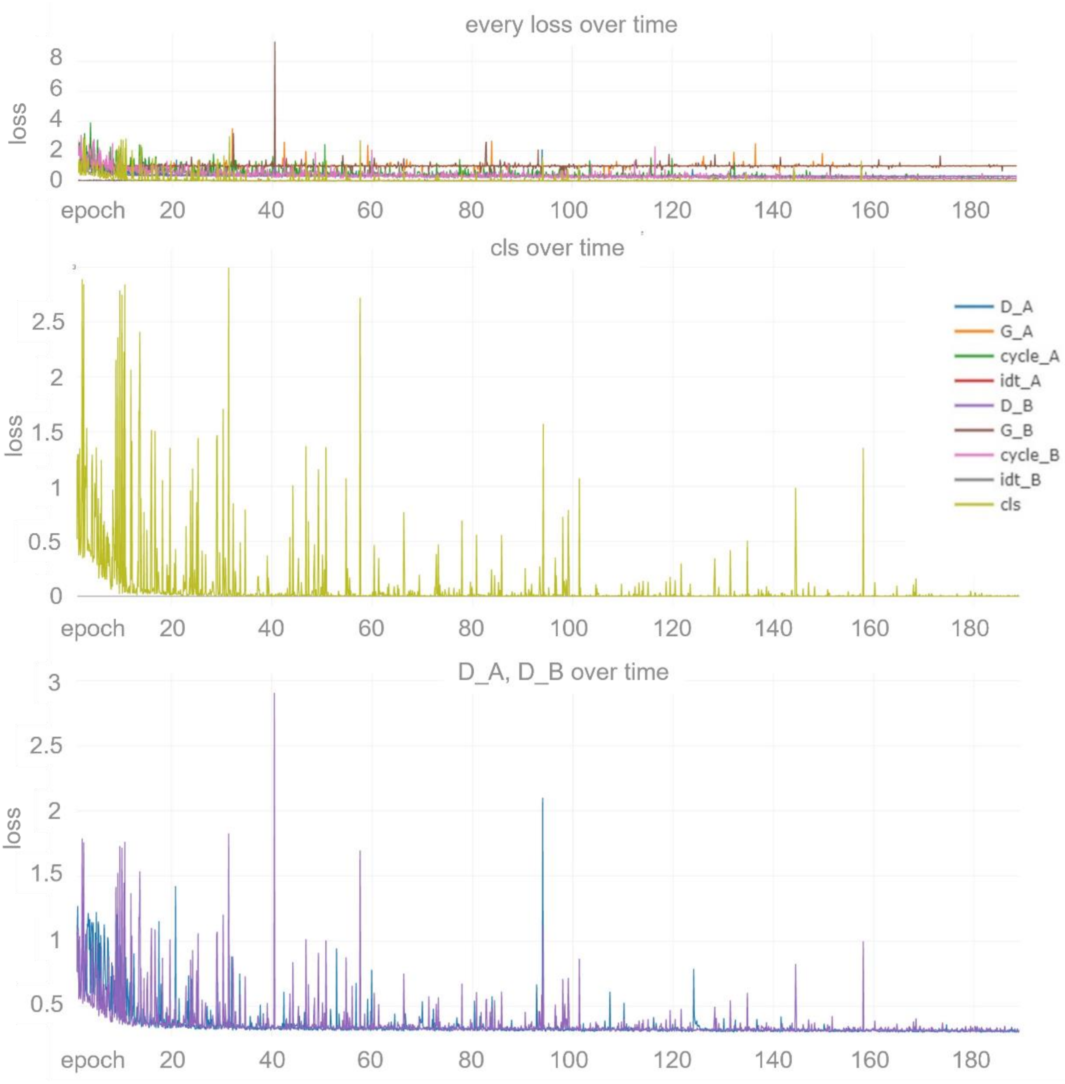

Figure 3. The training losses over time. The aggregate plot of 9 losses (above), the classification loss (middle), and the discriminator loss (below).

The auxiliary classifier performed synchronously with the discriminators, with the loss decreased and converged as expected, showing that the featurization in the discriminator coordinated with that in the auxiliary classifier so that the classification loss could be successfully optimized. The number of training epochs should be largely reduced, as the loss converged, and the model reached its best performance at a very early stage. Redundant training can lead to over-fitting, bringing negative effects to the final performance of the model. The regular metrics of classification tasks such as accuracy are not calculated this time due to the small batch size.

3.2. Transferring performance

Corresponding to the loss over time, the transferring performance was improved in around 40 epochs and began to decline after that point (**Figure 4**). At the early stage of the training, the random color noise still existed. At around epoch 40, the model reached its best performance, not only properly transferred the color, but also vividly expressed the contrast in inking strength and the texture of paper. After epoch 120, apparent over-emphasis on the inking contrast occurred, and when the training finished, the performance became severely poor.

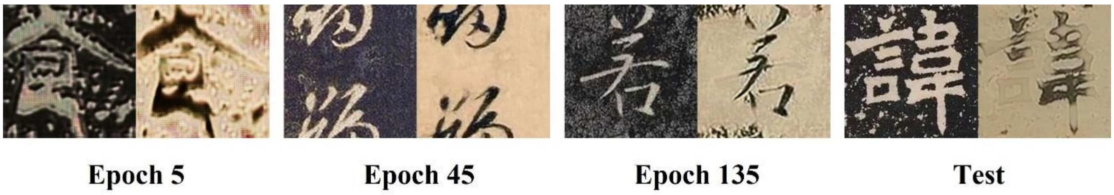

Figure 4. The transferring performance over time, at epoch 5, epoch 45, epoch 135, and test, from left to right. The left are real copies of stone inscriptions, and the right are fake ink marks.

Although with the hyperparameters carefully fine-tuned, the inversed-color problem still occurred in many runs (**Figure 5**), indicating that CycleGAN is sensitive to images with few colors.

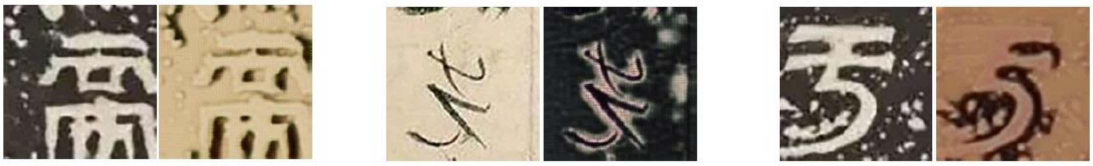

Figure 5. The color gets inversed in many runs. When the left were transferred to the right, the characters show the contrary color of what they were supposed to be in.

It is also noticed that the model also did a decent job transferring backwards — from ink marks to stone inscriptions (**Figure 6**). The model managed to learn that the copies of stone inscriptions usually have bolder strokes than the ink marks, and there are sometimes noises on the picture when the stone inscriptions are made into copies.

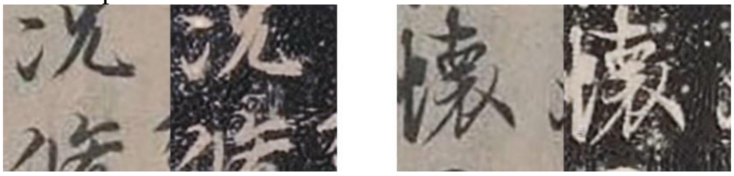

Figure 6. The transfer from ink marks to stone inscriptions in the training progress. The left are real ink marks, and the right are fake copies of stone inscriptions.

3.3. Limitation

The following deficiency can be improved in the future: (i) The classification task is rudimentary. Diversity in the auxiliary classification function such as Chinese character recognition can be achieved if given a more varied set of labels. (ii) Missed characters cannot be predicted. This can be improved if combined with Nature Language Processing. (iii) The styles of different calligraphers are not reflected. (iv) No high-resolution databases are currently available so that it cannot be adapted to StyleGAN to achieve better details.

4. Conclusion

This research aims to transfer the expression form of Chinese calligraphy — from the stone inscriptions that are easy to preserve, to the more precious ink marks, and simultaneously figure out which kind of font are the pieces in. An auxiliary classifier is attached to the discriminator in the CycleGAN model, which inputs both the target image and its label, and is trained to optimize the classification loss together with the traditional GAN loss. Within the early training epochs, the model smoothly transferred the images and embodied the details and properties of ink marks. The auxiliary classifier performed synchronously with the discriminator, with the classification loss successfully converging. In the future, a more thorough classifier, a combination with other artificial intelligence fields, and a better model with a better database should be included.

References

[1] Gatys L A Ecker A S Bethge M A 2015 Neural algorithm of artistic style arXiv:1508.0657

[2] Goodfellow I Pouget-Abadie J Mirza M et al. 2014 Generative adversarial nets Advances in neural information processing systems 2672-2680.

[3] Isola P Zhu J Y Zhou T et al. 2018 Image-to-Image Translation with Conditional Adversarial Networks arXiv:1611.07004v

[4] Zhu J Y Park T Isola P et al. 2017 Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Network

[5] Karras T Laine S Aila T 2019 A Style-Based Generator Architecture for Generative Adversarial Networks arXiv:1812.04948v3

[6] Karras T Aila T Laine S et al. 2018 Progressive Growing of GANs for Improved Quality, Stability, and Variatio

[7] Li Y Tai Y W Tang C K 2018 Attribute-Guided Face Generation Using Conditional CycleGA arXiv:1705.09966v2

[8] Mirza M Osindero S 2014 Conditional Generative Adversarial Nets arXiv:1411.1784v1

[9] Odena A Olah C Shlens J 2017 Conditional Image Synthesis with Auxiliary Classifier GANs arXiv:1610.09585v4

[10] Eveningglow 2017 Multitask-CycleGAN URL: https://github.com/eveningglow/multitask CycleGAN