Research on Digital Concept Art Illustrations Style Classification based on Deep Learning

Ziyang Li

Harbin Institute of Technology, Nangang District, Harbin, China
120L051020@stu.hit.edu.cn

Abstract. With the development of the Internet, a large number of conceptual art pieces have been created in the form of digital images and uploaded to the internet, giving rise to a diverse range of digital image resources that cater to the needs of contemporary youth subcultures. Therefore, categorizing these digital resources by art style can help in the intelligent management of digital resource platforms. Additionally, it assists internet users who may have limited knowledge in art to understand and learn about various art styles. The deep convolutional neural network model represented by the ResNet network has achieved great success in image classification in the past. In recent years, the Vision Transformer model, which is improved based on the Transformer model that has performed brilliantly in the NLP field, has further improved the accuracy of image classification based on the self-attention mechanism.This article focuses on three commonly seen art styles on the Chinese internet and employs deep learning to examine the accuracy of three neural network models in handling binary and ternary classification problems related to these styles. The test and validation results obtained on the dataset were used to evaluate and compare the three models, and the model with better performance was selected to improve the accuracy of image style classification.

Keywords: Deep Learning, Style Classification, Digital Concept Art

1 Introduction

Concept art is a design approach that expresses ideas in the form of illustrations, commonly used but not limited to films, video games, animations, and comics. The production of concept art is essential in the creation process of many contemporary media artworks. In recent years, a large number of concept art pieces have been produced and uploaded to the internet in the form of digital images, leading to the emergence of a vast array of diverse digital image resources catering to contemporary youth subcultures. Therefore, intelligent identification and classification of common and popular artistic styles in these digital resources can assist in establishing a well-defined digital resource platform and lay the foundation for more refined concept art style classification in the future. In addition, with the development of the AI painting

B. H. Ahmad (ed.), *Proceedings of the 2023 International Conference on Data Science, Advanced Algorithm and Intelligent Computing (DAI 2023)*, Advances in Intelligent Systems Research 180,
https://doi.org/10.2991/978-94-6463-370-2_40

field this year, people have greatly lowered the threshold for artistic creation. By entering keywords related to the content they want to create, they can create the desired artwork through a computer. However, the judgment and input of keywords related to the style of the work still require users to have a certain amount of artistic knowledge. This article hopes to provide some theoretical basis for the development of digital illustration style judgment applications by studying digital image style classification, thereby helping people without artistic knowledge to better complete AI painting art creation.

In the field of image classification, prior deep learning research has made significant progress. For example, researchers such as Òscar Lorente, Ian Riera, and Aditya Rana have studied the performance of various methods, including the use of support vector machines with the visual bag of words classifier, multilayer perceptrons, the existing InceptionV3 architecture, and their own CNN design called "TinyNet." They evaluated the accuracy and loss in each case [1]. Furthermore, in the study of abstract artistic concepts like image style, Quan Wang and Guorui Feng proposed a new two-branch network structure that aggregates graphic style features and global style features. They introduced a graph network to model correlations between artistic image region styles to capture graphic style [2]. Xiaoming Yu and Gan Zhou introduced an image style transfer model called "CrGAN," which encodes images into content and style to achieve continuous image transformations. In this paper [3], The author uses ResNet and Transformer models, which also have corresponding research in the image domain. For instance, Shang-Hua Gao evaluated Res2Net blocks on widely used datasets such as CIFAR-100 and ImageNet, demonstrating their superiority in various computer vision tasks like object detection, class activation mapping, and salient object detection [4]. Nicolas Carion and Francisco Massa proposed a new method based on the Transformer model to effectively simplify the object detection process [5]. Hua-Peng Wei and Ying-Ying Deng focused on comparing and analyzing the shape bias between CNN- and transformer-based models from the view of VST tasks, proposed three kinds of transformer-based visual style transfer (Tr-VST) methods (Tr-NST for optimization-based VST, Tr-WCT for reconstruction-based VST and Tr-AdaIN for perceptual-based VST) [6].

In the previous research on the region of image style classification, most researchers focus on the study of traditional art such as oil painting in Western and ink painting in China. For example, Xu. Z proposed A cross-contrast neural network model which utilizes an information entropy-based similarity measurement method to achieve multi-class classification of art styles and artists for artistic works [7]. This method addresses the issues of low accuracy and limited number of categories in existing neural network approaches for ink style classification.

Jiang. W conducted research on four different neural network model frameworks to effectively extract and fuse features for ink art classification [8]. Wang. H improved the deep learning network models ResNet50 and NTS to enhance the accuracy of painting image style classification recognition and validated the improved models' generalization by using a custom-made dataset of Thangka art [9]. Zhang Y applied various style transfers to the training dataset's images, then added the generated new

images to the training dataset, thereby increasing the number of samples in the training set to improve the model's classification accuracy [10]. Guo K proposed two convolutional neural network models for architectural style image classification, conducted experimental research, and provided corresponding analysis of the experimental results [11].

This paper aims to use current CNN framework and improve it to accomplish high accuracy of some common conceptual art style classification. It can not only help to enhance the efficiency of managing a large number of concept art resources, particularly those represented by game concept art, and to give those who lack of artistic knowledge to assistance to find correct key words to accomplish their own concept art creations using AI drawing tools.

2 Background

2.1 Dataset Description

The cyberpunk style images used in this article were all sourced from Kaggle dataset CyberVerse (Cyberpunk_ImagesDataset). The image datasets for the other two styles used in this article were obtained from the Chinese internet and were digitized image resources for which downloading was authorized.

2.2 Image Style

Given the diversity and complexity of conceptual art illustration styles, it is difficult to completely unify them using a simple hierarchical classification method. Therefore, this article chose several common and widely appreciated artistic styles, using constructed and trained deep learning model to try to classify them. The following will briefly introduce the content, artistic techniques, and other characteristics of several selected digital illustration styles.

Cyberpunk. "Cyberpunk" is a blend of "Cybernetics" and "Punk," often set in a backdrop where there is a juxtaposition of "low life and high tech." These narratives typically feature advanced scientific technology in contrast with a partially deteriorated societal structure. Figure 1 shows a representative scene illustration in the cyberpunk style, The overall picture is dominated by cool tones and has high color contrast.

Fig. 1 Cyberpunk style illustration [12]

From the perspective of artistic style, the overall brightness of the images is relatively low, with high contrast. They often feature a predominant cool color palette, contrasted with localized warm tones mainly represented by neon lights. The overall impression is both vibrant and icy. Numerous neon lights, billboards, towering structures juxtaposed with slums, and augmented individuals with prosthetics and heavy cables are distinctive visual elements that epitomize the cyberpunk style.

Chinese palaeowind. Chinese palaeowind is a new artistic style emerging in modern society, based on elements of traditional Chinese culture. Its content encompasses modern and fantastical elements, reflecting traditional Chinese cultural ideas, distinctive artistic characteristics, and the unique aesthetic taste of Chinese people. Chinese palaeowind illustration often draws inspiration from ancient China as its creative source. It can be based on historically accurate depictions or exist within the realm of fictional ancient settings. Figure 2 shows an illustration in the palaeowind style that imitates the smearing style of Chinese traditional painting.

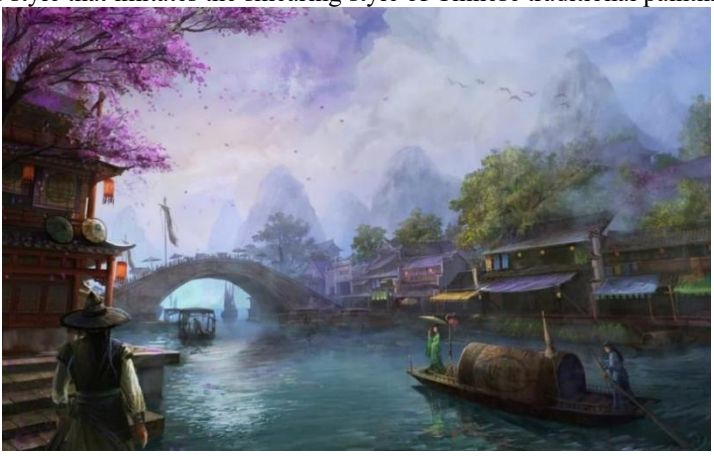

Fig. 2. Chinese palaeowind style illustration [12]

In terms of artistic characteristics, Chinese palaeowind illustrations exhibit distinct external features, primarily demonstrated through the application of traditional lines. These lines are derived from Chinese traditional culture, such as ancient painting and calligraphy, and encompass representative visual elements, including lines of varying thickness, length, and clear brushstrokes. The use of color is often influenced by

traditional colors, giving the illustrations a localized quality. In terms of composition, the integration of Western composition techniques enhances the sense of perspective and dynamism in the artwork, resulting in a certain degree of inclusiveness and diversity in the form of the illustrations.

Manga. Manga are comics or graphic novels originating from Japan. Most manga conform to a style developed in Japan in the late 19th century, and the form has a long history in earlier Japanese art [12]. The term manga is used in Japan to refer to both comics and cartooning. Outside of Japan, the word is typically used to refer to comics originally published in the country. Figure 3 shows a scene illustration with a Manga style.

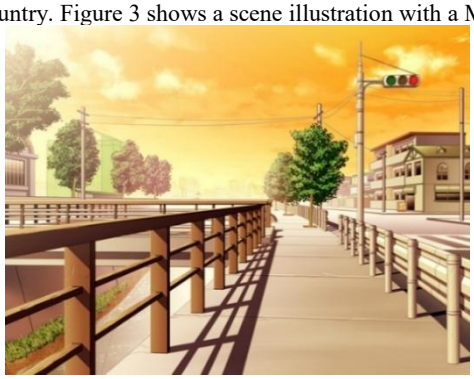

Fig. 3. Manga style illustration [12]

The Japanese artistic style is often characterized by being cute, with themes centered around adorable and clear depictions of beautiful girls and spirited teenagers. The style is concise, soothing, featuring vibrant color tones, and strong lines. Nowadays, many Japanese art styles employ a pseudo-thick painting technique to portray details and enhance the texture of the image. Japanese CG illustrations frequently employ gentle and elegant color palettes, creating a warm and cozy atmosphere in the visuals. This choice of colors aims to evoke a sense of tranquility and harmony.

3 Method

3.1 ResNet CNN model

The ResNet network was proposed by the team led by Kai Ming He at Microsoft Research in 2015. It won the first place in the ImageNet competition's classification task that year, and has since exerted a significant influence on subsequent research and applications. Figure 4 shows the significant solution to the degradation phenomenon of multi-layer neural networks after adopting the ResNet structure.

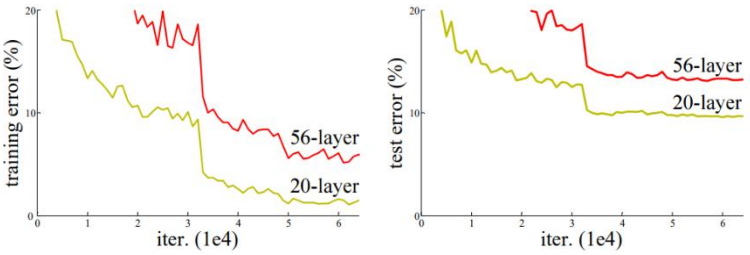

Fig. 4. The change in error rate after using the ResNet network [13]

The ResNet network addressed the common problem of degradation in deep neural networks at that time by introducing a solution based on the concept of residual construction. Figure 5 shows a classic unit structure that adopts the ResNet residual function idea.

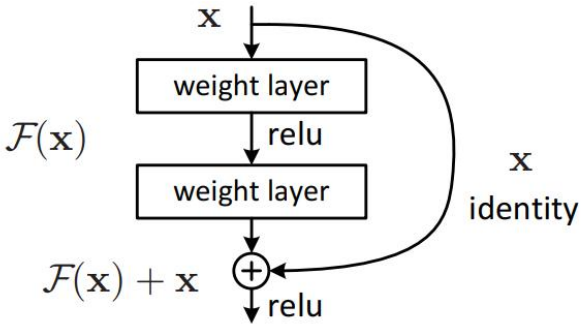

Fig. 5. Classical ResNet block [13]

In ResNet, by stacking shallow networks with identity mappings, the target function F(x) is represented as the difference between the optimal function H(x) and x, which is called the "residual function." As the optimal function shares similarities with a linear function, if the target function is close to the identity function, the training process can be significantly accelerated.

Exactly, that's the key advantage of Residual Networks (ResNets). With the introduction of residual connections, it becomes possible to train deeper networks effectively, which was difficult with traditional deep neural networks due to the vanishing gradient problem. The residual connections allow the network to learn identity mappings, which essentially lets the model preserve earlier layer information without losing it during the training process.

As a result, as the network depth increases, the training error tends to decrease, and the model can extract more meaningful and deeper-level features. This property of ResNets allows them to achieve better performance compared to shallower networks and has been instrumental in pushing the boundaries of deep learning in various tasks such as image classification, object detection, and segmentation.

The CNN model structure used in this paper is displayed in the Figure 6:

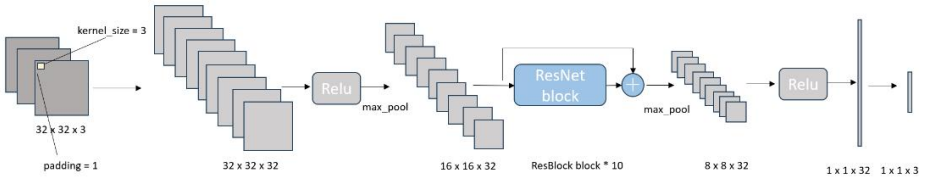

Fig. 6. ResBlock structure used in this experiment [13]

3.2 Vison Transformer

ViT utilizes the Transformer architecture, which has been widely employed in natural language processing tasks. Based on the self-Attention mechanism, ViT first divides the input image into equally-sized patches. Each patch is then transformed into a vector through an embedding layer and augmented with position encodings corresponding to their original positions in the image. Subsequently, using Query, Key, and Value matrices of the same dimension as the input patches, ViT performs image classification through supervised training. In the Encoder structure of the ViT model, a multi-head attention mechanism is adopted, dividing the model into multiple heads, forming multiple subspaces, allowing the model to focus on different aspects of information separately. The final result is determined by the weighted sum of the results of multiple heads, enhancing the robustness and stability of the network. Figure 7 shows the basic structure of the ViT network and the process that the entire model needs to go through to complete.

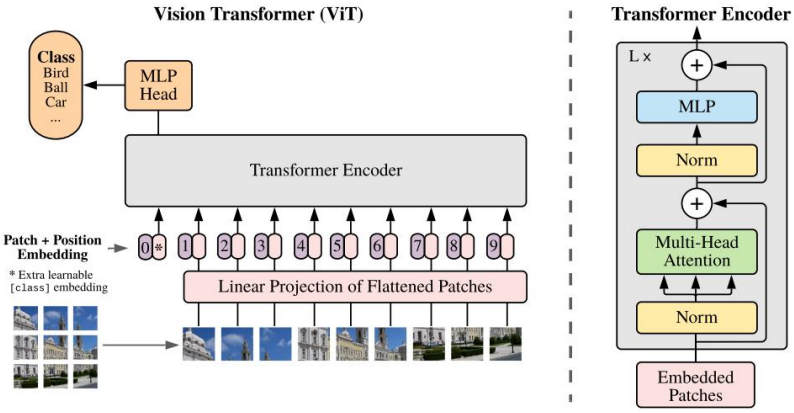

Fig. 7. Classical ViT structure [14]

4 Results

4.1 Outcome of the Model

In this experiment, the final output dimension of the traditional ResNet18 neural network structure is a 512-dimensional vector. In order to make the output results meet the requirements of the problem, this experiment added a fully connected layer at the end of ResNet18, causing some image information to be lost and the model performed poorly on the validation set, with some overfitting problems. Therefore, this experiment made minor adjustments to the traditional ResNet structure, reducing the input dimension of the final fully connected layer from 512 dimensions in ResNet18 to 32 dimensions, effectively improving the accuracy on the validation set. The binary classification problem was improved from 64.1% to 79.5%, and the ternary classification problem was improved from 54.8% to 72.0%.

After replacing the ResNet model with the ViT model, the accuracy on both the training and validation sets was significantly improved, proving that ViT is better than ResNet18 in dealing with image style classification problems. Table 2 and 3 showed the binary classification and three-class classification, respectively.

Table 1. Binary classification

Model	Training Accuracy	ValSet Accuracy
ResNet18	87.2%	64.1%
ReNet CNN	88.2%	79.5%
ViT	89.1%	85.4%

Table 2. Three-class classification

Model	Training Accuracy	ValSet Accuracy
ResNet18	89.0%	54.8%
ResNet CNN	87.0%	72.0%
ViT	97.0%	81.2%

4.2 Visualization of the Results

The training process and validation results of the three models on the training set and validation set are all displayed through visual charts, where Figure 8 and Figure 9 correspond to the training process for binary classification problems and ternary

classification problems, respectively. The blue graph refers to training set, the red one refers to validation set.

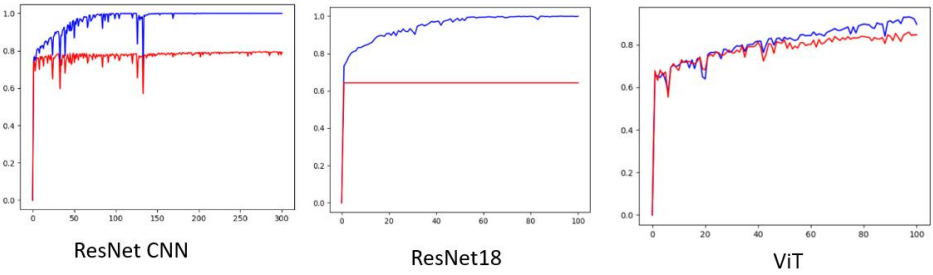

Fig. 8. Binary classification used three models above (Photo/Picture credit: Original)

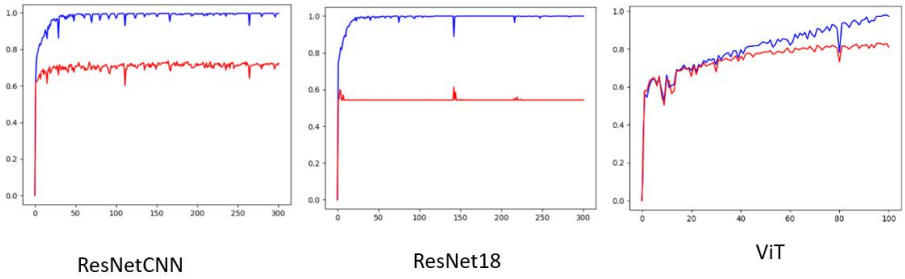

Fig. 9. Three-class classification used three models above (Photo/Picture credit: Original)

5 Conclusion

In the task of digital image style classification, compared to ResNet18, segmenting the input images into smaller units during data preprocessing, applying multiple layers of pooling in the CNN network structure, and using a smaller input dimension in the final fully connected layer effectively improved the model's accuracy and generalization on the validation set. Besides, Using the ViT network architecture for digital image style classification on medium-sized datasets yields higher accuracy compared to traditional CNNs.

This experiment only selected three commonly popular digital image styles for classification, and cannot provide a comprehensive and systematic classification of the diverse styles of a large number of digital images available on the internet. As a result, its practical application value is relatively low. By using a larger dataset with a greater variety of style labels and training with a more suitable model, the aim is to enable the model to accurately classify the style of input digital art images.

References

1. Lorente Ò, Riera I, Rana A. Image classification with classic and deep learning techniques[J]. arXiv preprint arXiv:2105.04895, 202..
2. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S. (2020). End-to-End Object Detection with Transformers. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, JM. (eds) Computer Vision – ECCV 2020. ECCV 2020. Lecture Notes in Computer Science(), vol 12346.
3. Wei, HP., Deng, YY., Tang, F. et al. A Comparative Study of CNN- and Transformer-Based Visual Style Transfer. J. Comput. Sci. Technol. 37, 601–614 (2022)..
4. Gao, S.-H., Cheng, M.-M., Zhao, K., Zhang, X.-Y., Yang, M.-H., & Torr, P. (2019). Res2Net: A New Multi-scale Backbone Architecture. IEEE Transactions on Pattern Analysis and Machine Intelligence.
5. Wang, Q., Feng, G. (2021). Image Style Recognition Using Graph Network and Perception Layer. In: Fang, L., Chen, Y., Zhai, G., Wang, J., Wang, R., Dong, W. (eds) Artificial Intelligence. CICAI 2021. Lecture Notes in Computer Science(), vol 13069.
6. Yu, X., Zhou, G. (2022). CrGAN: Continuous Rendering of Image Style. In: Khanna, S., Cao, J., Bai, Q., Xu, G. (eds) PRICAI 2022: Trends in Artificial Intelligence. PRICAI 2022. Lecture Notes in Computer Science, vol 13631.
7. Xu Ziying., Application and Research of the Classification in Art Style Based on Cross Contrast Neural Network[D]. Nanjing University, 2019.
8. Jiang Wei. Research on Ink-wash Painting Classification Method Based upon Deep Feature Fusion [D].Tianjin University, 2021.
9. Wang Honghui. Research on Painting Image Style Classification and Application Based on Deep Learning [D]. Qinghai University, 2022.
10. Zhang Ye. Research on Image Classification Technology Based on Style Transfer [D].Harbin Institution of Technology, 2019.
11. Guo Kun. Research on Classification of Architectural Style Image Based on Convolution Neural Network [D]. Wuhan University of Technology, 2017.
12. Gravett, Paul. Manga: Sixty Years of Japanese Comics. New York: Harper Design. ISBN 978-1-85669-391-2, 2004.
13. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
14. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.