THE EXPERIMENT OF NEURAL NETWORK ON THE COGNITION OF STYLE

WEI HU
[1]*College of Architecture and Urban Planning, Tongji University*
[1]*1990huwei@sina.com*

Abstract. This paper introduces a method to obtain quantified style description vector which is for computer analysis input by using image style classification task. In the experiment, 3331 architectural photos of three styles obtained by crawling and filtering were used as training data. A deep convolutional neural network was trained to map architectural images to high-dimensional feature space, and then the high-dimensional style description vector was used to output the measurement results of style cognition with fully connected neural network. Tested by test data-set of 371 architectural pictures, the accuracy rate of style cognition reached more than 80%. The neural network using architectural data training was applied to the style cognition of non-architectural objects, high accuracy rate was also achieved, it proved that this quantified style description vector did include the information about style cognition to some extent instead of simply classification. Finally, the similarities and differences between the cognitive characteristics of style of neural network and human beings are investigated.

Keywords. Deep neural network; style cognition experiment; eye tracker.

1. Introduction

In fine arts and performing arts, any art object or performance has its creation process and follows some unique generation ways and methods. If such methods are repeated over time to reproduce similar forms or actions, a characteristic pattern emerges and a style is formed(Chan, 2000). Most studies of style have focused on describing important forms of a style or reviewing the background to their development, and further exploring the background and relevance of their interactions with other forms. However, our definition of a style is almost always descriptive and qualitative, rather than numerical and quantitative, which leads to a question: Can styles be numerical described to facilitate input into the computer for computation or generation?

It is easy to map a picture of an architecture into a quantitative vector, but how to get a style vector with the right dimensions and retain the original style information is the key challenge. Maybe we can learn from the word embedding method in NPL, that is, can use the training of word prediction model in text to get

PROJECTIONS, Proceedings of the 26th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2021, Volume 2, 61-70.

the word vector in the intermediate step (Bengio, 2000). In this study, we use the task of architectural style recognition to obtain the style vector in the intermediate step.

This kind of high-dimensional spatial mapping transformation problem of images is actually the field of artificial neural network algorithms. The processing of the data related to the image or block is the advantage of Convolutional Neural Networks (CNN). In the 1990s, LeCun (LeCun et al.1989) introduced Backpropagation (BP) algorithm and simplified the Fukushima's (1975) network structure, marking the birth of modern convolutional neural network structure. Subsequently, complex convolutional neural networks composed of standard CNN structures were proposed, such as Lenet-5 (LeCun and Bottou 1998) proposed in 1998. With the progress of hardware technology, especially the successful application of GPU to neural network computing in 2006, the computing capacity of neural network training has been greatly improved, which makes the training of large and deeper neural network possible. A variety of complex neural networks with convolutional structure layer have been successively proposed, such as AlexNet (Krizhevsky et al. 2012), GoogLeNet (Szegedy et al. 2015), ResNet (He et al.2016), VGG (Simonyan and Zisserman 2015), etc., and have been applied to the image classification challenge competition, which has greatly improved the accuracy of image classification and surpassed human performance in this task in 2015.

Many artists and researchers have practiced the application of convolutional neural networks in style-related problems. Many of them are researches on style transfer. Gatys et al. (2016) and Shahriari et al. (2014) used convolutional neural network to transfer famous painting styles to architectural photos. Özel et al. (2019) used the convolutional neural network to transfer the two-dimensional painting style to the three-dimensional building model. In the process of style transformation, the content and style of the target building and the input image were separated and combined at multiple levels using the neural network, and the specific features were extracted and deployed. Zhang et al. (2020) studied the style transfer of 3D model images by using two GAN networks containing the convolutional neural layer. Some researchers try to use convolutional neural network to classify architectural styles. Obeso et al. (2017) used convolutional neural network combined with sparse features and primary color pixel values to classify three Mexican building images. Zhao et al. (2019) studied building classification using the convolutional neural network based on GoogLeNet. Yi et al. (2020) used convolutional neural network to conduct classification research on 8 types of American houses, but the classification accuracy was only about 40%.

This paper introduces a method to obtain quantified style description vector by using image style classification task, with the following main contents:

First, through the experiment on the cognition of architectural style of convolutional neural network, the results obtained the accuracy rate of more than 80%, which preliminarily verified the feasibility of artificial intelligence on the cognition of design style. In the intermediate step of neural network cognition, architectural style is mapped to high-dimensional space to obtain high-dimensional quantified style description vector, that's the style vector we're looking for.

Secondly, the style cognition ability of the neural network is further verified. Applying the neural network trained with architectural picture data to the style cognition of other kinds of objects, such as jewelry, clothing and furniture, has also achieved a high accuracy rate, which indicates that the method presented in this paper has a certain ability to recognize styles across different kinds, rather than simply classifying architectural pictures. This also proves that the style vector includes the style information of the non-building object, not the building category information.

Thirdly, on the one hand, the internal feature images of the convolutional neural network are visualized to observe the focus of attention of the neural network; on the other hand, the eye tracker was used to test the attention focus of human beings when they recognized the image style of buildings. In this way, the similarities and differences between the cognitive characteristics of style of neural network and human beings are investigated.

2. Method

In this study, an artificial neural network was used to recognize the style of Baroque, Byzantine and Gothic architectural pictures. The use of deep neural network is divided into four steps: First, collect and preprocess the data used in training and testing neural network; Second, select the appropriate neural network structure; Thirdly, use applicable method to train neural network; Finally, the trained network will be tested and applied.

2.1. DATA ACQUISITION

The image data used in the training of this paper is crawled by search websites, which are sourced from Google, Baidu and Naver. About 200,000 images of Baroque architecture, Byzantine architecture and Gothic architecture were downloaded by the crawler. However, most of these crawled pictures are duplicate pictures and inappropriate pictures such as images of books and text, etc. After deduplication and screening, there are about 3,700 pictures for this research, and the number of pictures of buildings of the three styles is roughly equal. The picture samples used in this research are shown in Figure 1.

Figure 1. Sample data images used in this experiment.

2.2. CONVOLUTIONAL NEURAL NETWORK STRUCTURE

The neural network used in this study is expanded based on VGG network(Simonyan and Zisserman 2015). The structure is shown in Figure

2. The main structure of VGG is divided into 5 groups of convolutional layers structure, each group includes 2-3 convolutional layers and 1 pooling layer. The convolutional layer is responsible for extracting the image features, while the pooling layer is responsible for reducing the size of the feature image. After each pooling operation, the size of feature image decreases, and the extracted features become more and more "abstract" and "macro". In this paper, the feature images will be visualized after each group of convolution operations, and their positions are indicated by the numbers 1-5 in Figure 2. After five groups of convolution operations with VGG, the feature images are mapped into a 256-dimensional feature vector by full-connected neural network, which will contain the neural network's "cognition" of the input image. Finally, this cognitive feature vector is output three numbers between 0 and 1 by using full-connected neural network, respectively representing the probability of Baroque, Byzantine and Gothic styles of neural network cognitive of the input image.

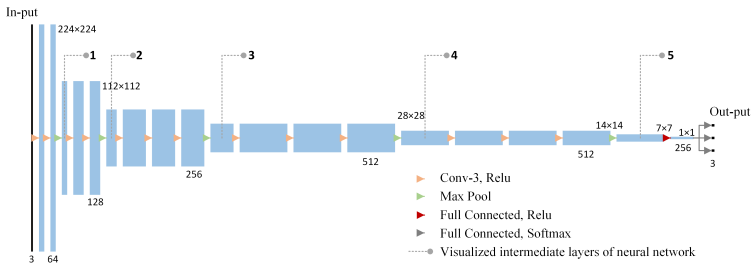

Figure 2. Network structure diagram for this study.

In fact, in order to achieve the research goal of this paper, that is, to find a quantitative vector to describe the style characteristics of the picture, so that it can be easily input into the computer, we carry out the training of style classification is only the means. In the method presented in this paper, the 256-dimensional vector before entering the classified fully connected neural network is the quantitative description vector we seek. We have made some attempts. On the one hand, there are enough vector dimensions of 256-dimension to preserve enough information, which has been proved by the high enough accuracy of classification in the following paper. On the other hand, the vector dimensions of 256-dimension are not too large, and good training can be obtained on relatively small data sets.

2.3. TRAINING

Human's ability to understand and generalize knowledge is far beyond the current neural network algorithm. All the textbooks that we use from elementary school to university can be digitized into a USB stick, and with such a small amount of data, humans can acquire quite a lot of knowledge. However, even teaching artificial intelligence algorithms to recognize buildings in satellite images, a very simple task in human eyes, requires more than 10GB training data (Sun and Hu 2020). Human can use very little information to learn a lot of knowledge, because humans have a special ability, analogy. We learn highly summarized the general theory

of knowledge, deduce different details, would be generally useful in different application scenarios. With the expansion of the scale of neural network, the training needs more and more data and computing power, so it becomes more and more difficult to train a new network from scratch. At this time, it is necessary for artificial intelligence to imitate human's ability of analogy, and the method of transfer learning (Sinno and Qiang 2009) emerges at the right moment. To put it simply, transfer learning is to help train a new network or a new data-set by using some network parameters that have been trained with old data-set, which can greatly reduce the requirement of training data and training time for the new training. VGG is a large scale deep neural network, so this paper uses the method of transfer learning to speed up training. The parameters we transferred were part of the VGG convolutional neural network trained with the ImageNet data-set, which was applied to the classification task of 1000 categories of images. We have reason to believe that there is a great deal of commonality in the extraction of some underlying features, such as edges, textures and simple geometric shapes, whether it is in the cognition of picture design style, or in the classification of flowers, dogs, cats and vehicles. After experiments, the transfer learning method in this study can greatly improve the training efficiency, and the model is convergent in a short time.

3. Result

After the neural network model was trained to converge with the training data-set, the accuracy test was carried out with the test data-set containing 371 pictures which were not included in the training set. After calculation, the model's style cognition accuracy on the test data-set is 80.86%.

Figure 3 is examples of the test images, No. 1-4 are Gothic (Notre Dame cathedral), No. 5-7 are Baroque (Church of Saint Carlo), and No. 8-11 are Byzantine (Hagia Sophia). The cognitive results are shown in Table 1. The accuracy of cognition is relatively high. Several of them deserve special mention. For example, in picture No. 10, the data used in this experimental neural network training are architectural photos, while picture No. 10 is the engineering drawing of the section of Hagia Sophia cathedral. The probability that the neural network considered it Byzantine was as high as 96.57%. This indicates that the neural network has a certain ability of "analogy" in terms of the type of image input for the style cognition task. Picture No. 6, 7, 8 and 9 show the interior roofs of the Baroque and Byzantine buildings, respectively, with a greater similarity in appearance. The neural network correctly recognized three of the pictures, and the wrong No. 6, it can be seen that the neural network is not very sure about the wrong judgment. Picture No. 4, the top of the gate and sculpture of Notre Dame cathedral is another result of cognitive error. Similarly, the neural network is not very sure about the wrong cognitive result, but it also shows that the neural network is not very strong in "analogy" from the training of relatively overall buildings to the local features like relief or sculpture.

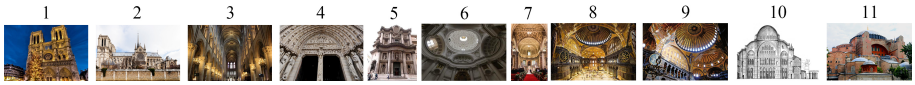

Figure 3. Architectural images for further testing.

Table 1. Probability results of neural network cognition.

	1	2	3	4	5	6	7	8	9	10	11
Baroque	5.66%	4.01%	30.34%	66.37%	97.56%	27.82%	87.81%	1.62%	1.98%	2.45%	0.70%
Byzantine	26.45%	4.99%	4.04%	5.62%	1.11%	70.55%	5.21%	98.35%	97.42%	96.57%	99.20%
Gothic	67.89%	91.00%	65.62%	28.01%	1.33%	1.64%	6.98	0.03%	0.60%	0.98%	0.10%
Correct/Wrong	Correct	Correct	Correct	Wrong	Correct	Wrong	Correct	Correct	Correct	Correct	Correct

4. Further verification

In order to test whether the neural network actually recognized the perceived style of human beings or merely performed a simple image classification task, we designed an experiment for further verification.

When we talk about the Baroque, or the Gothic, we don't just mean the style of buildings, we can also mean furniture, jewelry, clothing or even music or literature. Although it is difficult to quantify, we can at least sum up some qualitative knowledge of architectural style. However, in terms of styles of objects across different categories, we can hardly even make qualitative generalization. If the neural network can recognize the styles of other kinds of objects across categories while only learning architectural styles, then it can be verified that the neural network has indeed learned the styles recognized by human beings.

We used architectural photos of three styles to train the neural network. In the above test, we have preliminarily verified that the correct rate of architectural style cognition is relatively high, and it even have a certain cross-category style cognitive ability, such as the cognitive ability of section drawing of picture No.10. In order to further test the style cognition ability of the neural network, we will test the style cognition of the neural network trained by architectural photos with non-architectural pictures. As shown in Figure 4, picture No. 1 is modern art in Baroque style; picture No. 9 and 11 are baroque style furniture; No. 5 is an oil painting whose interior furnishings and costumes are in the late Baroque style. Picture No. 2 is part of Chanel's "Paris-Byzantium" collection; Picture No. 4 is a Byzantine jewelry; Decorative paintings in picture No. 6 and 7 are in the Byzantine style. Picture No. 3 is gothic furniture; Picture No. 8 is a gothic fascinator for a modern doll; Picture No. 10 is a gothic style photo with modern Gothic makeup and clothing.

Figure 4. Non-architectural pictures for further testing.

As shown in Table 2, the test results show that the neural network trained by architectural photos also has a high accuracy rate in style recognition of non-architectural cross-category items. But there are some issues worth mentioning. Picture No. 2 Chanel "Paris-Byzantium" series of jewelry is recognized as Baroque style; Although the cognition of picture No. 5 oil painting is correct, the certainty is not very high. It seems that the neural network also thinks that it belongs to Byzantine style with certain possibility; According to the prediction results, the neural network seems to be lost in the style cognition of picture No.10, the probability of these three styles is relatively close and none of them is more than 50%.

Table 2. Probability results of further tests.

	1	2	3	4	5	6	7	8	9	10	11
Baroque	96.15%	86.36%	2.01%	3.43%	51.49%	0.72%	0.36%	3.71%	69.87%	42.38%	76.47%
yzantine	2.61%	11.63%	1.16%	94.76%	44.86%	97.63%	99.35%	0.10%	27.60%	32.08%	14.77%
Gothic	1.24%	2.00%	96.83%	1.80%	3.66%	1.65%	0.28%	96.19%	2.53%	25.55%	8.76%
rect/Wrong	Correct	Wrong	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Wrong	Correct

5. Experiment of attention focus in style cognition

Figure 5 shows the visualization process of the middle layer of the neural network from the shallow to the deep 5 positions, and finally superposing all the layers to obtain the thermal diagram of the neural network attention focus. The visualization method is to highlight the sensory field corresponding to highly activated neurons in the convolutional neural network, that is, the features "seen" by neurons in this layer.

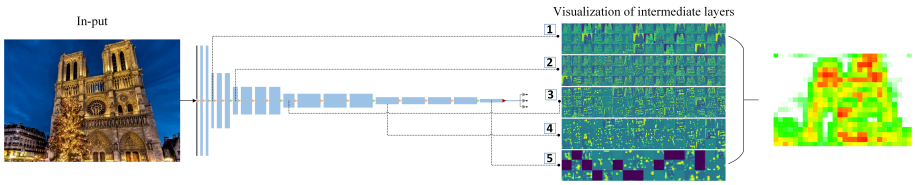

Figure 5. Neural network internal feature image visualization.

It can be seen that the features extracted from the feature images with serial numbers from 1 to 5 change from micro to macro and from details to the abstractions. The feature images extracted from position 1 are the most basic and detailed edge lines, angles and colors, etc. At position 3, they become larger

parts, such as small windows, columns and eaves, etc., while at position 5, they become more integral large parts, such as large rose windows, the portal and even the whole front facade. This hierarchical perception of detail is to some extent similar to human perception of art or the process of creation- We may have similar experience when painting and looking at art exhibitions. one moment, we observe microscopic features such as details and brush strokes closely, and the other moment, we stand back to observe macroscopic features such as overall picture proportions, light and shade contrast and color collocation.

By further stacking all the feature images, the heat map of "attention focus" of all activated neurons in the neural network can be obtained. From the heat map of attention, we can analyze the areas that the neural network focuses on when cognizing the style of a picture, infer the possible judgment basis that the neural network thinks that a picture is a certain style, and try to peer into the corner of the black box of the neural network.

At the same time, we conducted an experiment with 15 architecture students, asking them to look at architectural pictures and judge their style. Eye tracker was used to record their focus of attention during the experiment, as shown in Figure 6.

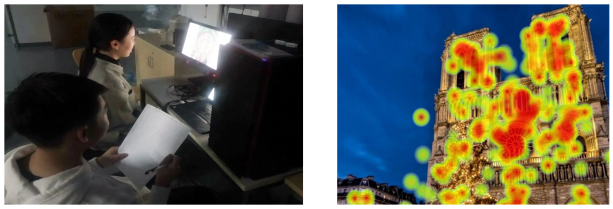

Figure 6. Eye tracker recording of architectural style identification by architecture students.

The results of neural network visual thermal diagram and eye tracker recording are shown in Figure 7. Above is the original picture of behavioral architecture, in the middle is the behavioral eye tracker recording the attention focus of architectural students, and below is the behavioral neural network attention-focus thermal diagram.

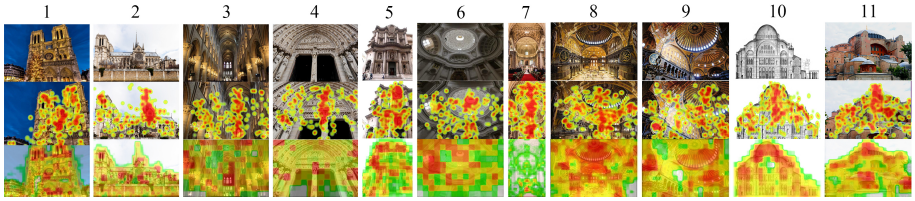

Figure 7. Comparison of the attention focus diagram of the neural network and the architecture students.

After comparison, the following conclusions can be drawn: First, the focus of attention of neural network is more dispersed and average than the visual focus of human beings, which may be because human beings have a better ability of

generalization when observing, and they do not need to observe some regular patterns separately, but only need to focus on one of them. Second, there are many common points of attention between neural networks and human beings, such as rose windows, long lancet windows and spires in Gothic architecture; The junction of eaves and columns, decorative relief in Baroque architecture; the dome in Byzantine architecture. Thirdly, the attention focus of neural network is also different from that of human beings, such as attention to the junction of structures, attention to the vicinity of the edges of architectural components, and attention to sculptures or figures in architecture.

6. Conclusions

In this study, the cognitive ability of neural network to design style was investigated experimentally. In the intermidia step of the neural network, we extracted the quantified style description vector that we want. Based on this style vector, the probability that the input images are of Baroque, Byzantine and Gothic styles is recognized.

Architectural photos of three different styles were crawled through the network to train a neural network training. For this large neural network, the transfer learning method was used to speed up the training. After the training, the performance of the three styles cognition was tested in the test data-set, and the results showed that the accuracy rate of the three styles was 80.86%.

Furthermore, the neural network which was trained of architectural photo data-set is used to cognize the style of non-architectural objects, and the result still shows a high accuracy rate. To some extent, this indicates that the neural network has "discovered" some cross-category common style characteristics in the high-dimensional feature space, so that it can recognize the style of objects across categories without prior training of corresponding categories, that is to say, it has preliminarily acquired certain ability of "analogy" in terms of style cognition.

In order to further explore how the neural network cognizes architectural style, the feature images extracted from the middle layer of the neural network are visualized, showing that as the number of layers deepens, the architectural features extracted from the neural network also change from the subtle geometric features to the macroscopic architectural components. This cognitive pattern has some similarities with human beings. Then, all feature images are superimposed to explore the distribution of neural network's attention when "observing" specific images, so as to analyze the basis of its cognitive judgment. In addition, the eye tracker experiment of architecture students was carried out to record the focus of human attention when observing architectural style and compare it with the neural network. It is found that the focus of neural network is partly similar to that of human beings.

7. Discussion and Futures studies

It is difficult and controversial to describe a style qualitatively in a way that human beings can understand. However, the method presented in this paper can be regarded as a preliminary attempt to obtain a vector that can input a quantitative

description of style into a computer without needing to be in a dimension that humans can understand. Although we cannot directly understand the specific meaning of this style vector, we can input the image style information into the computer in this way, which provides an input interface for us to deal with style-related tasks. Much further work is needed, such as trying to parse these style vectors on a human-understandable dimension; To further improve the accuracy of description; Study how to use style vector for generation and so on.

References

Bengio, Y., Ducharme, R. and Vincent, P.: 2000, A Neural Probabilistic Language Model, *Advances in Neural Information Processing Systems 13*.

Chan, C.: 2000, Can style be measured?, *DESIGN STUDIES*, **21(3)**, 277-291.

Fukushima, K.: 1975, Cognitron: A self-organizing multilayered neural network, *Biological Cybernetics*, **20(3-4)**, 121-136.

Gatys, L.A. and Ecker, A.S.: 2016, Image Style Transfer Using Convolutional Neural Networks, *Computer Vision & Pattern Recognition. IEEE*.

He, K., Zhang, X. and Ren, S.: 2016, Deep Residual Learning for Image Recognition, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE*.

Krizhevsky, A., Sutskever, I. and Hinton, G.E.: 2012, ImageNet classification with deep convolutional neural networks, *NIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems*, 1097–1105.

LeCun, Y., Boser, B. and Denker, J.S.: 1989, Backpropagation Applied to Handwritten Zip Code, *Neural Computation*, **V1N4**, 541-551.

LeCun, Y. and Bottou, L.: 1998, Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, **86(11)**, 2278-2324.

Obeso, A.M., Benois-Pineau, J. and Acosta, A.A.R.: 2017, Architectural style classification of Mexican historical buildings using deep convolutional neural networks and sparse features, *Journal of Electronic Imaging*, **26(1)**, 011-016.

Shahriari, K. and Shahriari, M.: 2017, IEEE standard review — Ethically aligned design: A vision for prioritizing human wellbeing with artificial intelligence and autonomous systems, *Humanitarian Technology Conference. IEEE*, 197-201.

Simonyan, K. and Zisserman, A.: 2015, Very Deep Convolutional Networks for Large-Scale Image Recognition, *ICLR, 2015*.

Sinno, J.P. and Qiang, Y.A.: 2009, A Survey on Transfer Learning, *IEEE Transactions on Knowledge and Data Engineering*, **2009**, 1345-1359.

Sun, C. and Hu, W.: 2020, A Rapid Building Density Survey Method Based on Improved Unet, *the 25th CAADRIA Conference*, Bangkok, Thailand, 649-658.

Szegedy, C., Liu, W. and Jia, Y.: 2015, Going Deeper with Convolutions, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, **07**, 1-9.

Yi, Y.K., Zhang, Y. and Myung, J.: 2020, House style recognition using deep convolutional neural network, *AUTOMATION IN CONSTRUCTION*, **118**, 103307.

Zhang, H. and Blasetti, E.: 2020, 3D Architectural Form Style Transfer through Machine Learning, *the 25th CAADRIA Conference*, Bangkok, Thailand, 659-668.

Zhao, P., Miao, Q. and Liu, R.: 2019, *Architectural Style Classification Based on DNN Model*, Springer, Cham.

Özel, G. and Ennemoser, B.: 2019, Interdisciplinary AI, *the 39th Annual Conference of the Association for Computer Aided Design in Architecture (ACADIA)*, Austin, Texas, 380-391.