Influence of CNN Layer Depth on Spiral Visual Illusions

Kenji Aoki
University of Miyazaki, 1-1 Gakuenkibanadai-nishi, Miyazaki-city, Miyazaki, 889-2192, Japan

Makoto Sakamoto
University of Miyazaki, 1-1 Gakuenkibanadai-nishi, Miyazaki-city, Miyazaki, 889-2192, Japan
Email: aoki@cc.miyazaki-u.ac.jp, saka2000@cc.miyazaki-u.ac.jp

Abstract

Understanding the mechanism of visual illusion generation through Convolutional Neural Networks (CNNs) that mimic the receptive fields of the visual cortex can contribute to elucidating the mechanisms of visual information processing in the brain. In our previous research, we demonstrated the potential for Fraser's spiral illusion to manifest in CNNs. In this study, we focused on the depth of the CNN layer structure and examined the impact of the number of layers on the manifestation of the visual illusion. We provided 14 types of spiral illusion images to three different CNN patterns with varying layer structures and tasked them with distinguishing between concentric circles and spirals. The results indicated that CNNs with fewer layers were more prone to the illusion, whereas CNNs with more layers were less likely to exhibit the illusion. These results suggest that the number of layers in a CNN influences the manifestation of visual illusions.

Keywords: Convolutional Neural Network, CNN, Visual Illusion, Spiral Illusion

1. Introduction

Convolutional Neural Networks (CNNs), which mimic the local receptive fields of the visual cortex, have been extensively studied and utilized in various fields [1], including image recognition [2]. In human vision, under certain conditions, phenomena known as "visual illusions" occur, where shapes, sizes, lengths, colors, and directions are perceived differently from their physical reality. The occurrence of visual illusions suggests that visual information from the retina is not transmitted directly to the brain but undergoes some processing. Therefore, investigating whether CNNs can exhibit illusions and understanding the mechanisms behind them, could contribute to elucidating the mechanisms of visual information processing.

In the study by Watanabe et al., experiments were conducted using deep neural networks to observe motion from "rotating snakes" images, a type of motion illusion, to determine if the same illusions observed in humans could be replicated [3]. Motion illusions are phenomena where stationary images are perceived as moving. In the case of "rotating snakes," disk-shaped images resembling snakes appear to rotate. The results showed that the neural network predicted rotational motion from the "rotating snakes" images, indicating that deep neural networks can exhibit illusions.

Our previous research also demonstrated the potential for Fraser's spiral illusion to occur in CNNs [4]. Fraser's spiral illusion is a phenomenon where concentric circles are perceived as a spiral. This study suggested that the structure of the CNN model influences the occurrence of visual illusions. Horikawa et al. compared CNNs with the human brain and reported homology in information representation between each layer of the CNN and each region of the human visual cortex [5]. This implies that altering the number of convolutional layers in CNNs could affect the occurrence of illusions. Additionally, Simonyan et al. reported that the deeper the layers of a CNN, the higher the accuracy of image classification [6]. Therefore, in our research, we hypothesized that increasing the depth of the layers would enhance image discrimination accuracy, leading to the correct identification of physically concentric images as concentric, thereby reducing the occurrence of visual illusions. In this study, we focused on the depth of the CNN layer structure and examined the influence of the number of layers on the occurrence of visual illusions.

2. Methodology

2.1. *CNN models*

The CNN constructed in this study is composed of several layers, including convolutional layers, normalization layers, and pooling layers. In the input layer, the pixel values (0-255) of the input image are provided. The convolutional layers extract local features by multiplying the input values with the filter values and summing the neighboring output values. The results of the convolution are converted into output signals using an activation function. The ReLU (Rectified Linear Unit) function was used as the activation function. The ReLU function outputs the input value directly if it is greater than 0, and outputs nothing if it is 0 or less.

The pooling layers perform down-sampling to reduce the complexity of the subsequent layers, which is equivalent to reducing the resolution in image processing [7]. This helps to mitigate the effect on the output results when the input image is slightly shifted. In this study, max pooling was used. Max pooling divides the image into small rectangular regions and outputs only the maximum value within each region to the next layer.

Batch Normalization, developed by Serger et al., is a technique to accelerate deep learning [8]. It also reduces dependency on initial values and helps prevent overfitting. Dropout is a technique that randomly deletes neurons during training to prevent overfitting.

These layers were combined to construct the CNN model. The structure of the three models constructed in this study is shown in Fig. 1. The dimension of the input layer is 150 x 150 for all models. Grayscale images of 150 x 150 pixels were provided as input, and the images passed through multiple convolutional layers (Conv2d in Fig. 1), Batch Normalization layers (BatchNormalization in Fig. 1), and pooling layers (MaxPooling in Fig. 1). Finally, the images passed through a global average pooling layer (GlobalAveragePooling in Fig. 1), Dropout (Dropout in Fig. 1), and a fully connected layer (Dense in Fig. 1) to output either "0" or "1". An output value of "0" indicates a concentric circle, while "1" indicates a spiral.

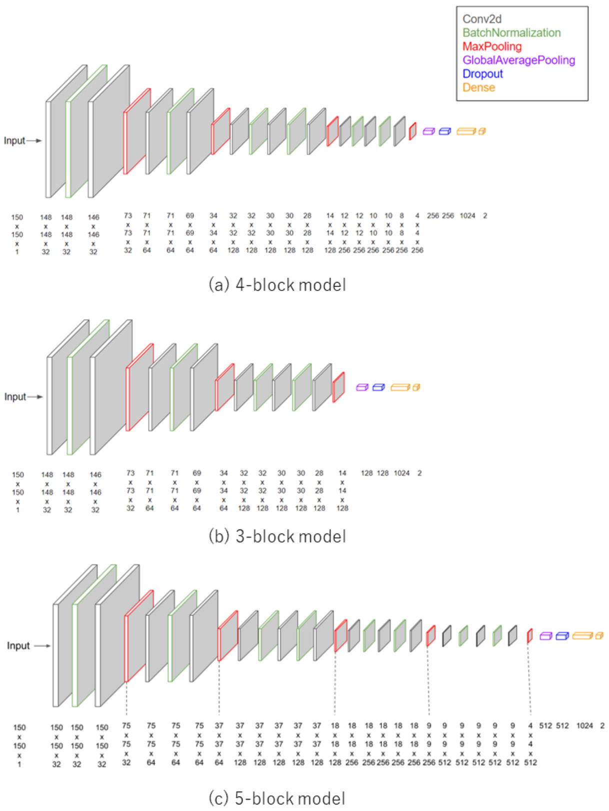

Fig.1 Composition of the 3 CNN models

The convolutional layers, Batch Normalization layers, and Max Pooling layers were combined into a "block," and models with different depths were constructed by combining four, three, and five blocks. The configurations of these models are shown in Fig. 1 (a), (b), and (c), respectively. Each section separated by dotted lines in Fig. 1 (c) represents one block. In our previous research, we demonstrated the potential for spiral illusion occurrence using a model composed of four blocks (4-block model in Fig. 1 (a)) [4]. We investigated whether visual illusions occurred in these three models and whether there were any changes in the occurrence of visual illusions.

2.2. *Computing Environment*

In this study, we utilized Python (ver. 3.9.6), a programming language rich in machine learning libraries, and TensorFlow (ver. 2.5.0), an open-source machine learning software library developed by Google, for machine learning using CNNs. The computer used for this study had the following specifications: OS - Windows 10, CPU - Intel Xeon X3480, GPU - NVIDIA GeForce RTX3060, and RAM - 20.0GB.

2.3. *Training Datasets*

To train the constructed CNN to distinguish between spiral and concentric circle images, we created a training dataset. First, we generated 50 images each of 150 x 150 pixels for both concentric circles and spirals. Additionally, we augmented these images by applying horizontal and vertical shifts, flips, and scaling. Including the original images, we prepared 250 images each for concentric circles and spirals. From these, 200 images of each type were randomly selected as training data, and 50 images of each type were selected as test data. Fig. 2 shows some of the created training data images. Fig. 2 (a) shows concentric circle images, and Fig. 2 (b) shows spiral images. In addition to images of concentric circles and spirals drawn with solid and dashed lines, we created images with features seen in spiral illusions. The left side of Fig. 2 shows images drawn with solid lines, the center with dashed lines, and the right side shows images combining black and white lines and triangular endpoints, similar to those seen in spiral illusion images (e.g., Fig. 3 06). We also prepared patterns with different combinations of background and line colors, such as white and black, black and white, and gray and black. Note that the training dataset does not include spiral illusion images.

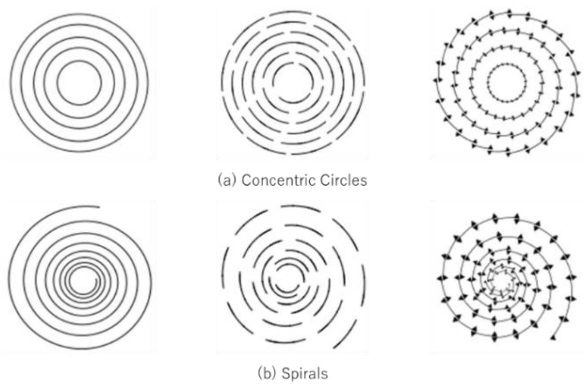

Fig.2 Example of training data set

2.4. *Validation Images*

After training the CNN model, we provided 14 spiral illusion images as input and checked the output. The spiral illusion images used for validation are shown in Fig. 3. These images were prepared with reference to those published on the "Akiyoshi Kitaoka's Illusion Pages" [9]. The images were standardized to a size of 150 pixels by 150 pixels and converted to 256-level grayscale. Images 01 to 14 in Fig. 3 are actually concentric circles, but they exhibit a visual illusion phenomenon where they appear spiral when viewed by humans.

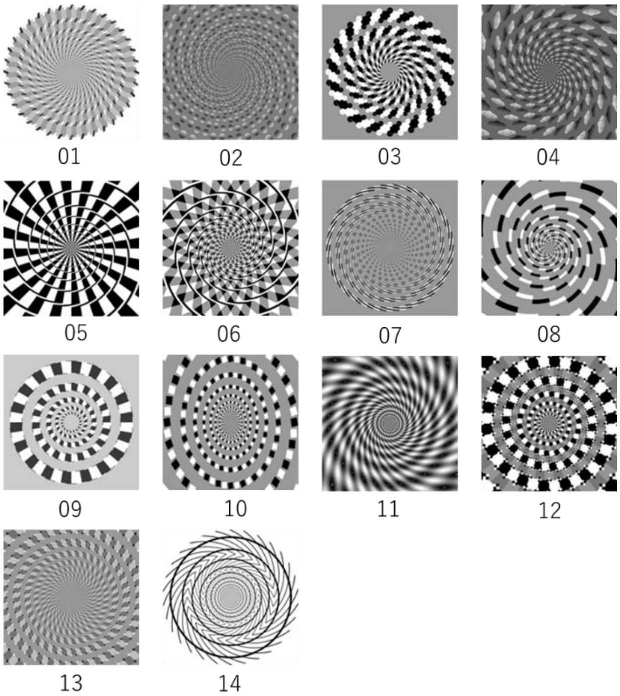

Fig.3 Spiral illusion images for verification

3. Results and Discussion

We checked the changes in the loss function values and accuracy with respect to the number of training iterations during the training process of the constructed CNN models. Cross-entropy error was used as the loss function. For all three models shown in Fig. 1, it was confirmed that the loss function values and accuracy for both the training and test data stabilized after 300 epochs, so the models were trained for up to 500 epochs for validation.

To eliminate the possibility of random classification as spirals, we randomly changed the initial values and repeated the training and validation 10 times. Fig. 4 shows the number of times each validation image was classified as a visual illusion. The vertical axis represents the number of times the image was classified as a spiral out of 10 validations, and the horizontal axis corresponds to the image names shown in Fig. 3. The results for the 4-block model, 3-block model, and 5-block model are shown in blue, red, and yellow, respectively.

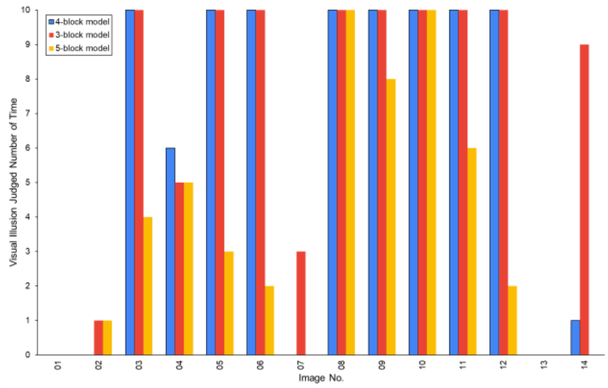

Fig.4 Comparison of the number of spiral illusion occurrences among three models

For the 4-block model (blue in Fig. 4), 8 out of 14 images (57%) were consistently classified as spirals in all 10 trials, indicating the occurrence of spiral illusions. Additionally, 4 out of 14 images (29%) were consistently classified as concentric circles, indicating no illusion. The remaining images had mixed classifications.

For the 3-block model (red in Fig. 4), 8 out of 14 images (57%) were consistently classified as spirals in all 10 trials, indicating the occurrence of spiral illusions. Additionally, 2 out of 14 images (14%) were consistently classified as concentric circles, indicating no illusion.

For the 5-block model (yellow in Fig. 4), 2 out of 14 images (14%) were consistently classified as spirals in all 10 trials, indicating the occurrence of spiral illusions. Additionally, 4 out of 14 images (29%) were consistently classified as concentric circles, indicating no illusion.

The 3-block model showed a similar rate of spiral classification as the 4-block model. However, images "02" and "07", which were not classified as spirals in the 4-block model, were classified as spirals in the 3-block model. Additionally, image 14 was classified as a spiral more frequently in the 3-block model than in the 4-block model. The 5-block model classified images as spirals less frequently than the 4-block model. These results suggest that deeper layers may reduce the occurrence of visual illusions. However, it was not clearly demonstrated whether deeper layers lead to more physically accurate classifications.

4. Conclusion

In this study, we constructed CNNs with varying depths, trained them with concentric circle and spiral images, and used the models to test for the occurrence of spiral illusions. By comparing the results across different models, we found that deeper layers might reduce the occurrence of visual illusions. In the future, we plan to investigate the influence of changing the training images and further modifying the model depth on the occurrence of illusions. Additionally, we aim to examine and compare the features extracted by each layer to understand the differences.

References

1. A. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, K. Kavukcuoglu, WaveNet: A Generative Model for Raw Audio, 2016. doi:10.48550/arXiv.1609.03499
2. Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition. Proceedings of the IEEE, vol. 86 (11), 1998, pp. 2278-2324. doi:10.1109/5.726791
3. E. Watanabe, A. Kitaoka, K. Sakamoto, M. Yasugi, K. Tanaka, Illusory Motion Reproduced by Deep Neural Networks Trained for Prediction. Frontiers in Psychology, vol. 9, 2018. doi:10.3389/fpsyg.2018.00345
4. K. Aoki., T. Togo., M. Sakamoto. Investigating Visual Illusions in Convolutional Neural Networks Using Spiral Illusion Images, Proceedings of the Sixteenth International Conference on Genetic and Evolutionary Computing, Genetic and Evolutionary Computing, Lecture Notes in Electrical Engineering 1322, 2024, (in press).
5. T. Horikawa, Y. Kamitani, Generic decoding of seen and imagined objects using hierarchical visual features. Nature Communications, vol. 8 (1), 2017, p. 15037. doi:10.1038/ncomms15037
6. K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition., 2014. doi:10.48550/arXiv.1409.1556
7. S. Albawi, T. A. Mohammed, S. Al-Zawi, Understanding of a convolutional neural network. 2017 International Conference on Engineering and Technology (ICET), 2017, pp.1-6. doi:10.1109/ICEngTechnol.2017.8308186
8. S. Ioffe, C. Szegedy, Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, 2015. dio:10.48550/arXiv.1502.03167
9. A. Kitaoka, Akiyoshi's illusion pages. http://www.psy.ritsumei.ac.jp/~akitaoka/index-j.html, last accessed 2022/02/07.

Authors Introduction

Dr. Kenji Aoki

He received a Ph.D. in Engineering from Kagoshima University. He is working in the Information Technology Center at University of Miyazaki as Associate Professor. His research interests include visual information processing, artificial intelligence, campus information systems and educational information systems. He is a member of IPSJ and JSET.

Dr. Makoto Sakamoto

He received the Ph.D. degree in computer science and systems engineering from Yamaguchi University. He is a professor in the Faculty of Engineering, University of Miyazaki. He is a theoretical computer scientist, and his current main research interests are computer science and information processing.