Paper:

Müller-Lyer illusion is Replicated by Higher Layer of Pre-trained Deep Neural Network for Object Recognition

Hongtao Zhang, Zhen Li, Shinichi Yoshida

*School of Information, Kochi University of Technology, Kami, Kochi 782-8502, Japan
E-mail: 258004n@gs.kochi-tech.ac.jp
**Department of Psychology, The University of Hong Kong, Hong Kong, China
E-mail: jczhenli@hku.hk
***School of Information, Kochi University of Technology, Kami, Kochi 782-8502, Japan
E-mail: yoshida.shinichi@kochi-tech.ac.jp

Abstract. Deep neural networks (DNNs) are commonly used in neuroscience research, however comparing neural networks to human perception remains challenging. With the development of the disciplinary intersection of deep learning and neuroscience, researchers have begun to use deep neural networks (DNNs) and human behavioral perception to assess the relevance of DNNs to brain neural networks. Here, we explore the similarity between DNNs and human visual perception through the well-known Müller-Lyer illusion. Specifically, by presenting the classical visual illusory stimuli, we compared representation by VGG-19 and ResNet-101 with human perception. Results indicated that similar to human visual perception, Muller-Lyer illusion is also shown in DNNs. Also, we further find that DNNs exhibit low physical-perceptual bias but present similar judgments in analogy with the layers of human perceptual hierarchy.

Keywords: deep neural networks; illusions; Müller- Lyer illusion; visual perception; deep learning

1. Introduction

Cognitive psychology is an essential discipline that studies the brain's neural networks. Cognitive psychology emphasizes understanding various brain functions, such as vision, language, reasoning, and consciousness. Among them, illusions of the senses, especially visual illusions, are investigated by scholars in brain neurology [1]. The concept of illusion itself is not uniform and is mainly divided into perceptual, logical, and geometric levels. Figure 1 shows a well-known geometric illusion, the Müller-Lyer illusion [2]. With the development of machine learning, scientists have started to analyze the mechanisms of visual illusions through deep neural networks (DNN), which are accurate and effective as a learning template for brain science models [3, 4]. They even outperform humans in object recognition.

In recent years, even with the continuous improvement and advancement of DNNs, it is still highly controversial

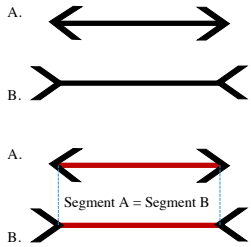


Fig. 1. Müller-Lyer Illusion. Straight-line A and straight-line B visually exhibit sensory judgments of different lengths but remain the same in length.

to discuss whether DNNs can be used as an explanation for neuroscience research. At present, the interpretation of human vision remains highly problematic and complex. DNNs, which as one of the explanations of the brain neural network, can promote and support the study of the brain that are currently inconclusive and unexplained [4].

DNNs can be good models for human brain in many aspects [4, 5]. At the same time, DNNs hold great promise for simulating human behaviors and are somehow able to exhibit performance close to that of humans [5–8]. DNN models trained for image classification seem to explain some aspects of biological vision well. Although DNNs fail to explain some aspects of the visual system and the similarity between DNNs and the visual system is inconclusive, it is still reasonable for the current neuroscience research. In addition, several studies have shown that hierarchical visual areas and DNN layers correspond in visual feature representation and that deeper stages of DNN exist at a level of perception similar to that of the brain [8–10]. In this study, we are biased to go for DNNs to seek and compare human differences and synergistic relationships at the visual level and to explore the possibility of their mechanisms(figure 2).

This paper explores the similarity between DNN and human perception in a visual illusion. We organize experiments through the Muller illusion to further verify that DNNs also have brain-like performance and feedback in other visual illusion tasks. In the study, we de-

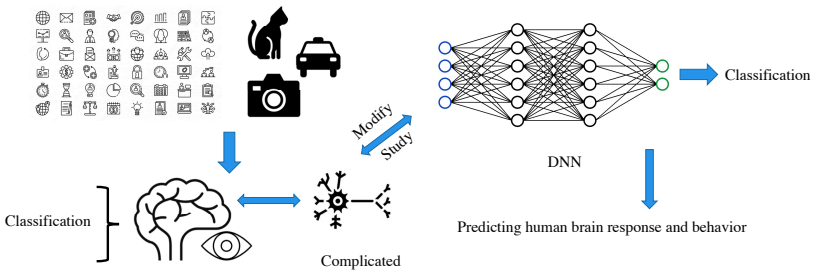


Fig. 2. DNNs provide an explanation of brain mechanisms by simulating brain neural networks. By the same token, as the explanation within the brain continues to improve, it further drives the development of DNNs.

sign human subject experiments to obtain human behavioral data through the Muller-Lyer illusion and compare the same behavioral data by judging them through DNNs [11–14] (Figure 3). The results show that DNNs have similar illusory judgments to humans. Based on different stages of DNNs, we further find that DNNs reflect human perceptual judgments of optical illusions in some aspects, and the representation of their performance resembles the characterization of human perceptual representations [12]. Nevertheless, there is still a black-box problem with the specific generation mechanism, and there is still a need to improve the DNN model further while enhancing the reference comparison of multiple cognitive science issues [12, 13, 15].

2. Method

2.1. Behavioral experiment

2.1.1. Participants

A total of twelve volunteers were recruited for the behavioral experiment. All twelve volunteers were male (mean age: 24.58, sd = ±2.72), and the participants were not aware of the purpose of the experiment. None of the participants had problems with color blindness or visual defects. In addition, all participants had normal or corrected to the normal range of visual acuity. This study was approved by the Human Research Ethics Committee of Kochi University of Science and Technology and followed the relevant guidelines and regulations. Written informed consent was obtained from all participants.

2.1.2. Apparatus

We demonstrated the stimuli with a Pixio PX248 Prime monitor (refresh rate: 144Hz, resolution: 1920 × 1080) in a dark room. The experiment was written in a Python environment with Psychopy (ADD reference). The distance between the participant and the display was approximately 64 cm. Subjects were required to adjust the stimuli via a keyboard (Magic Keyboard, Apple)[1].

2.1.3. Stimuli

Participants were required to perform ten trials of training before experiment. During the experiment, stimuli with a standard straight line and a straight line with arrows were randomly displayed on the screen. The length of the standard straight lines ranged from 200 to 380 pixels with 20-pixel intervals and total of 10 categories. The straight lines with arrows ranged from 100 to 500 pixels and were randomized in length. The left and right positions of the straight line with arrows are presented in a random order, and there are two random cases of the outward and inward-facing arrows. The arrows had an angle of 45 degrees and a length of 25 pixels.

1. https://www.apple.com

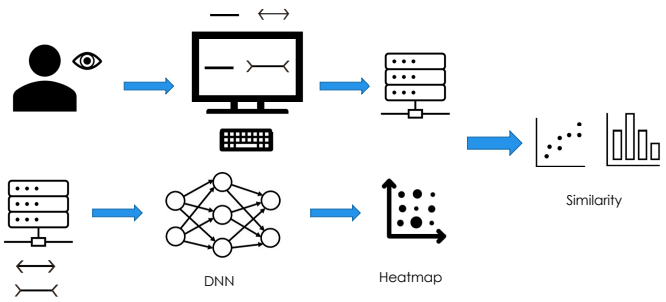


Fig. 3. The main processes of study. The behavioral and psychological data of the subjects were collected through a designed optical illusion experiment, based on the subjects' data, to perform the same test on the DNN and compare the sensory performance of both.

As shown in Figure 4, The A, S and K, L keys adjusted the straight line with arrows until it felt like the length of the straight line with arrows was equal to that of the standard straight line. Then the participants pressed the space bar to complete the adjustment for the current trial and then moved on to the next trial, with the A, S key being the "fine" adjustment of ±1 pixel and the K, L key being the "coarse" adjustment of ±10 pixels. The experiment was presented in four groups, each with group comprised of 40 trials. The behavioral data acquired were subjects' perceptual data in ten length categories (spaced at 20 pixels) from 200 to 380. Each length category includes eight perceived lengths with arrows facing outward in a straight line and eight perceived lengths with arrows facing inward in a straight line.

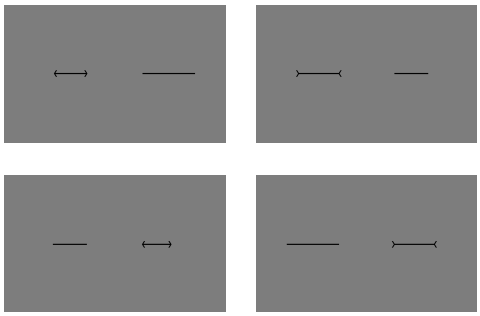


Fig. 4. The stimulis in experiment.

2.2. DNN models

2.2.1. Nerual networks

We perform DNN model selection from Brain-score (https://www.brain-score.org/) [10, 13, 16, 17]. The experiments use the conventional VGG-19 [18] and ResNet-101 [19] as the initial models. According to relevant visual studies and perceptual comparison experiments, Resnet-101 is a more brain-like artificial neural network. In contrast, VGG-19 is more consistent with the interpretation of human visual perception, although its performance is lower than the others [5, 10]. We perform feature extraction in DNN using subjects' perceptual data, mainly in FC7 of VGG-19 and FC of ResNet-101, for perceptual data in opposite arrow directions, respectively, to perform similarity analysis. In addition, the experiments also set up a control group for feature extraction to further verify the performance of DNN. The straight line with opposite arrow directions but the same actual length is used as the control group. In addition, we also compare the analysis at different depths of the network to further investigate the representation of these lines at different layers of DNN.

2.2.2. Equipment

The experiment is based on Linux[2] system as a whole. We build DNN networks in the Pytorch[3] framework. GPU (Tesla V100-SXM2 (32GB), DGX)[4].

2.3. Data Analysis

2.3.1. Perceived length group

The subjects' adjusted length data are available in the subjects' experiments, corresponding to each length label. Therefore, we use the subjects' adjusted lengths as the perceived length data. Each length label corresponds to one outward arrow and one inward arrow for the perceived length data. For example, a length label of 200 pix has data with an arrow facing outward at 210.34 pix and an arrow facing inward at 196.45 pix. We present these data as images and assemble them into a perceptual length dataset.

2.3.2. Real length group

Similarly, corresponding to each length label, we used the same length data as the length label in each length label in order to verify the effect of the presence or absence of arrows on the length judgment. For example, a 200pix length label has 200pix of data with arrows facing outward and inward. We also assemble the actual length dataset in the form of images.

2.3.3. Main analysis

The experiments are analyzed and compared using a characterized similarity approach. We use representative similarity analysis (RSA) to obtain the representational difference matrix (RDM) [4, 20] for the perceptual data in the opposite arrow direction and the control group. The RDM consists of the extracted feature tensor distances. The Euclidean distance is used as the distance metric for the experiments. We perform the distance calculation on the feature tensor in the penultimate or third layer of the network and further normalize. In addition, we performed the difference comparison in each module of the DNN network. Specifically, we compute the linear images and images with arrows of the perceptual length group and the actual length group separately in the DNN and then obtain their feature tensor $t_{mn}^{\text{perceived}}$ and t_{mn}^{real}. m and n represent the rows and columns of the tensor.

The specific formula for the Euclidean distance is as follows:

$$r_{\text{L2}} = \sqrt{\sum_{m=1}^{M}\sum_{n=1}^{N}\left(t_{mn}^{\text{perceived}} - t_{mn}^{\text{real}}\right)^2} \quad \ldots\ldots \quad (1)$$

Then, we normalize R_{L2} by different length(u represents different label of length):

$$R_{\text{L2}} = \frac{r_{\text{L2(u)}}}{\overline{r_{\text{L2}}}} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

$t_{mn}^{\text{perceived}}$ is the tensor of the perceived distance, and t_{mn}^{real} represents the tensor of the actual distance. After calculating the L2 distance by tensor, we normalized the L2 distance for different length labels.

2. https://www.linux.org/
3. https://pytorch.org/
4. https://www.nvidia.com/ja-jp/

3. Results

3.1. Behavioral results

We set up multiple sets of visual illusion stimuli to allow subjects to make adjustments in their perceptual judgments. In Figure 5, the blue dashed line represents the standard length, i.e., the actual length. In general, the two straight lines are the same length and remain the same with the addition of the arrow, i.e., the blue dashed line. The solid red line reflects the actual perceived length of the subject. From the figure, the straight line with the arrow facing outward needs to be adjusted longer to match the actual length, and the subject's perceived distance to the straight line with the arrow facing outward is reduced. Similarly, the perceived performance of a straight line with the arrow facing outward is smaller than the actual length and requires minor adjustment to match the perceived actual length.

3.2. ResNet-101

By obtaining perceptual data from the participants, we compare the feature tensor representation difference matrix [20]. Figure 6 shows the heat map form of the representation difference matrix for opposite arrow directions under different network modules. Darker colors represent low correlations, and brighter colors represent high correlations.

As the depth of the neural network increases, both the perceived length group and the actual length group (control group) maintain a constant trend. In the perceptual length group, the diagonal similarity is high. In the actual length group, straight lines with different arrow directions also showed a positive correlation, but the high correlation is shifted upward compared to the perceptual group. As the depth of the network deepened, the strength of the relationship between the different length variables became weaker, and the correlation remained essentially constant under the diagonal.

3.3. VGG-19

We perform the same analysis in the VGG-19 network. As seen in Figure 7, the similarity trend between the perceived length group and the actual length group (control group) in equal length units is the same as the depth of the neural network increases. The high similarity in the perceptual length group gradually fits the diagonal region. In the control group, the high similarity initially approached to fit the diagonal region and finally started to shift upward, and the final result is consistent with the case of ResNet-101.

3.4. Dissimilarity on different layers

We further put the line pictures of the perceived length group and the actual length group into each layer of the DNN structure to calculate them separately, extract the feature vectors, and calculate the Euclidean distance by Eq. (1) to obtain the dissimilarity of each layer of the network. From Figure 8, two groups exhibit a very low dissimilarity in the input layer of both networks and then gradually increase. There is no particular correlation between the physical and perceptual properties of DNNs.

4. Discussion

In the subject experiments, we first obtained data on the psychological aspects of the subject's behavior. Subjects produced perceptual biases from lines with different arrows. The lines with arrows caused subjects to show a visual effect of longer or shorter lengths (Figure 5). For example, in a straight line with a length of 260 pix, subjects perceive the straight line with arrows facing outward to have a matching length of 260 pix in the 275.1135 ± 12.7522 pix interval. Moreover, the straight line with the arrow facing inward was matched on the 251.6981 ± 10.2164 pix interval. Overall, the sensory performance of the straight line with the outward arrows is perceived shorter than actual length; on the contrary, the straight line

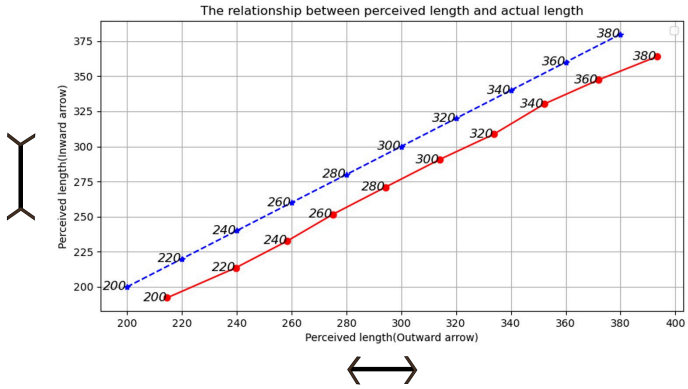


Fig. 5. The perceived illusion from Human subject. The blue dashed lines represent straight lines of the same actual length, with arrows in opposite directions. The solid red line represents the subject's perceived length. The graph corresponds to each actual length, and the red aspect shows an offset, showing that the straight lines with arrows facing inward are longer and those with arrows facing outward are smaller.

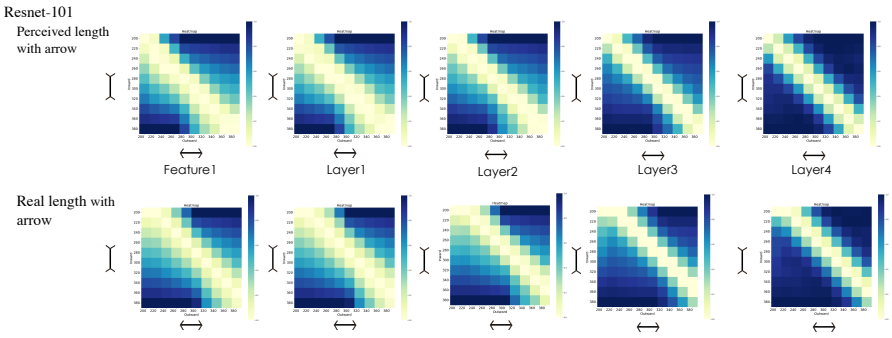


Fig. 6. Perceptual judgments of the perceptual dataset and the actual length set as expressed on ResNet-101. Darker colors represent lower similarity, and brighter colors represent higher similarity. Both groups exhibit high similarity performance that tends to be diagonal, with a more upward bias in the actual length group. The deepening of the network depth does not change the high similarity performance but instead deepens the dissimilarity between the labels of different lengths.

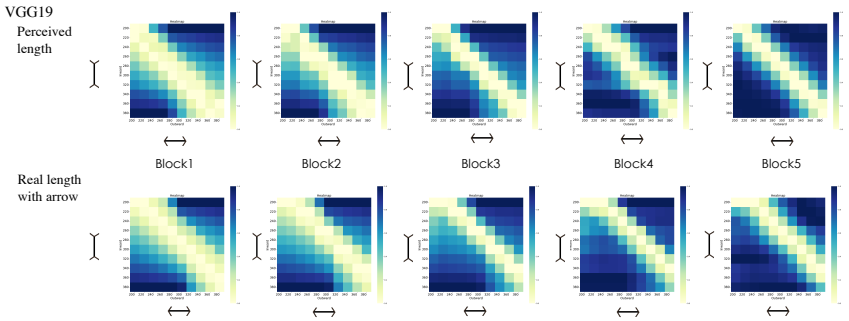


Fig. 7. The heat map shows the representation similarity between lines of different types and lengths on VGG-19. Darker colors represent lower similarity, and brighter colors represent higher similarity. Both groups also show high similarity performance tending to the diagonal, with a more upward bias in the actual length group. Moreover, the high similarity shifts upward as the depth increases.

with the inward arrows is perceived longer than the actual length.

In the DNN comparison experiments, both DNN models exhibit similar perceptual biases to humans. Figure 6 shows that ResNet-101 exhibits an illusional response from the beginning, and the dissimilarity only rises with increasing depth for different length variables. The high similarity on the diagonal remains unchanged. The correlation between the straight line with the arrow facing outward and the straight line with the arrow facing inward at the same perceptual length is high. The DNN exhibits the illusional response similarly as the subjects [9, 16, 21]. In the actual length group, straight lines with different arrow directions also showed a positive correlation, but the high correlation is shifted upward compared to the perceptual group.

In figure 7, VGG-19 shows an upward shift of high similarity, eventually reaching the same performance as that of ResNet-101. We believe this is related to VGG19 is closer to human visual perception. However, both remain consistent in the final results.

The high similarity in the diagonal from the perceived group indicates that the DNN makes an illusory performance that straight lines with the same actual length but different arrow directions are not equal in length. The length of the arrow in outward looked shorter and in inward looked longer, matching the subjects' perceptual performance.

In further experimental comparisons, we compared the two groups' data in every block of the network before the classification layer. We believe that it may show different similarities at the initial stage and then multiply the opposite trend, which would have some relevance to the mechanism of the brain to process the pictures of illusions. For example, focusing on the physical property aspect in the early stage, a high similarity would exist for straight lines of the same length in different arrow directions. The opposite would be true for perceived straight lines of different lengths. At the perceptual level down the line, it also happens to be the other way around. Combined with the response of DNN, the perceived length group has unequal straight-line lengths but sensory length consistency. In contrast, the control group is equal in length and inconsistent in sensory length. Therefore, the opposite trend of similarity will be shown in DNN. the DNN may be biased to physical properties first and then to the perceptual domain. The straight lines of perceptual length groups show higher dissimilarity in DNN and then gradually decrease

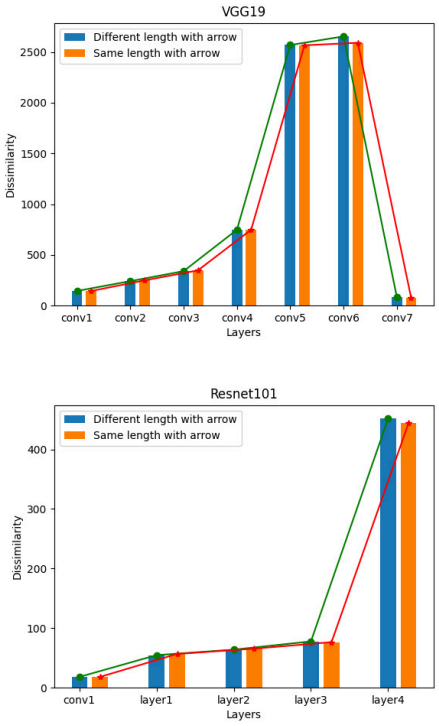


Fig. 8. The dissimilarity between VGG-19 and ResNet-101. As the number of network layers increases, both groups in VGG-19 and ResNet-101 show a gradual upward trend and remain the same in terms of similarity, with VGG-19 showing a significant drop after the conv6 layer but remaining consistent in terms of similarity.

to reach the minimum at the perceptual layer.

A similar pattern is not shown in Figure 8 but rather the same trend. It implies that the part of the DNN that produces perceptual bias may be at the perceptual level, i.e., the part where humans produce hallucinatory processing may also be in the perceptual domain further back, rather than at the physical processing field.

In this study, human brain-like cognitive performance is presented in DNNs that were not trained on a similar task. Our findings explain whether DNNs can be used in brain cognition studies.

5. Conclusion

In general, through the comparisons between DNNs and human behavior, we found that DNNs can produce judgments similar to human perception, and there may be similar mechanisms between deep neural networks and brain nerves at the perceptual level. Based on DNN, an influential class of theoretical analysis methods, it has helped further the understanding of neural networks in the brain. A single optic illusion may only show mechanistic similarity in one aspect. In the future, we will combine multiple visual illusions and brain fMRI data to compare and make inferences with advanced DNN models.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Numbers JP22K12786, JP21H03553, JP20H00267.

References:

[1] Todorović, Dejan. ”What Are Visual Illusions?.” Perception 49, no. 11 (2020): 1128-1199.

[2] Oyama, Tadasu. ”Japanese studies on the so-called geometrical-optical illusions.” Psychologia 3, no. 1 (1960): 7-20.

[3] Villa, Alexander Gómez, Adrián Martín, Javier Vazquez-Corral, and Marcelo Bertalmío. ”Convolutional Neural Networks Deceived by Visual Illusions.” CoRR (2018). Phil. Trans. Roy. Soc. London, Vol.A247, pp. 529-551, April 1955.

[4] Sun, Eric D., and Ron Dekel. ”ImageNet-trained deep neural networks exhibit illusion-like response to the Scintillating grid.” Journal of Vision 21, no. 11 (2021): 15-15.

[5] Lindsay, Grace W. ”Convolutional neural networks as a model of the visual system: Past, present, and future.” Journal of cognitive neuroscience 33, no. 10 (2021): 2017-2031.

[6] Kubilius, Jonas, Martin Schrimpf, Aran Nayebi, Daniel Bear, Daniel LK Yamins, and James J. DiCarlo. ”Cornet: Modeling the neural mechanisms of core object recognition.” BioRxiv (2018): 408385.

[7] Geirhos, Robert, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A. Wichmann, and Wieland Brendel. ”ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness.” arXiv preprint arXiv:1811.12231 (2018).

[8] Geirhos, Robert, Kristof Meding, and Felix A. Wichmann. ”Beyond accuracy: Quantifying trial-by-trial behaviour of CNNs and humans by measuring error consistency.” arXiv preprint arXiv:2006.16736 (2020).

[9] Geirhos, Robert, Carlos R. Medina Temme, Jonas Rauber, Heiko H. Schütt, Matthias Bethge, and Felix A. Wichmann. ”Generalisation in humans and deep neural networks.” arXiv preprint arXiv:1808.08750 (2018).

[10] Schrimpf, Martin, Jonas Kubilius, Ha Hong, Najib J. Majaj, Rishi Rajalingham, Elias B. Issa, Kohitij Kar et al. ”Brain-score: Which artificial neural network for object recognition is most brain-like?.” BioRxiv (2020): 407007.

[11] Gomez-Villa, Alexander, Adrian Martin, Javier Vazquez-Corral, and Marcelo Bertalmío. ”Convolutional neural networks can be deceived by visual illusions.” In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 12309-12317. 2019.

[12] Cichy, Radoslaw M., and Daniel Kaiser. ”Deep neural networks as scientific models.” Trends in cognitive sciences 23, no. 4 (2019): 305-317.

[13] Mehrer, Johannes, Courtney J. Spoerer, Nikolaus Kriegeskorte, and Tim C. Kietzmann. ”Individual differences among deep neural network models.” Nature communications 11, no. 1 (2020): 1-12.

[14] Mehrer, Johannes, Courtney J. Spoerer, Nikolaus Kriegeskorte, and Tim C. Kietzmann. ”Individual differences among deep neural network models.” Nature communications 11, no. 1 (2020): 1-12.

[15] Golan, Tal, Prashant C. Raju, and Nikolaus Kriegeskorte. ”Controversial stimuli: Pitting neural networks against each other as models of human cognition.” Proceedings of the National Academy of Sciences 117, no. 47 (2020): 29330-29337.

[16] Majaj, Najib J., and Denis G. Pelli. ”Deep learning—Using machine learning to study biological vision.” Journal of vision 18, no. 13 (2018): 2-2.

[17] Storrs, Katherine R., Tim C. Kietzmann, Alexander Walther, Johannes Mehrer, and Nikolaus Kriegeskorte. ”Diverse deep neural networks all predict human IT well, after training and fitting.” BioRxiv (2020).

[18] Simonyan, Karen, and Andrew Zisserman. ”Very deep convolutional networks for large-scale image recognition.” arXiv preprint arXiv:1409.1556 (2014).

[19] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. ”Deep residual learning for image recognition.” In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778. 2016.

[20] Kriegeskorte, Nikolaus, Marieke Mur, and Peter A. Bandettini. ”Representational similarity analysis-connecting the branches of systems neuroscience.” Frontiers in systems neuroscience (2008): 4.

[21] Dekel, Ron. ”Human perception in computer vision.” arXiv preprint arXiv:1701.04674 (2017).