

# Investigating the Gestalt Principle of Closure in Deep Convolutional Neural Networks

Yuyan Zhang<sup>1</sup>, Derya Soydaner<sup>2,3</sup>, Fatemeh Behrad<sup>2,3</sup>, Lisa Koßmann<sup>2</sup>,  
and Johan Wagemans<sup>2,3</sup> \*

1- KU Leuven, Department of Computer Science, Leuven, Belgium

2- KU Leuven, Department of Brain and Cognition, Leuven, Belgium

3- Leuven.AI, KU Leuven Institute for AI, Leuven, Belgium

**Abstract.** Deep neural networks perform well in object recognition, but do they perceive objects like humans? This study investigates the Gestalt principle of closure in convolutional neural networks. We propose a protocol to identify closure and conduct experiments using simple visual stimuli with progressively removed edge sections. We evaluate well-known networks on their ability to classify incomplete polygons. Our findings reveal a performance degradation as the edge removal percentage increases, indicating that current models heavily rely on complete edge information for accurate classification. The data used in our study is available on Github.

## 1 Introduction

Neural networks were designed inspired by the working mechanism of the human brain and have since achieved remarkable success across various fields. While psychology still aims to better understand the human brain, computer science strives to enhance understanding of neural networks. The primary goal in neural network research is to develop models capable of performing similar tasks as the human brain, rather than to recreate it. Interestingly, neural networks tend to exhibit more human-like behavior than expected, despite not being explicitly designed for this purpose. In particular, recent findings suggest that CNNs may exhibit certain aspects of Gestalt laws of perceptual organization [1], which explain how the human brain interprets complex visual stimuli, albeit possibly subject to certain thresholds and limitations.

The narrow scope of previously explored neural network architectures, coupled with limited datasets and insufficient experiments, necessitates a more comprehensive investigation. We focus on the Gestalt principle of closure, which states that the human brain naturally fills in gaps to perceive figures as complete wholes, when parts are occluded or fragmented. We present a dataset specifically designed to examine closure from various psychology-based perspectives and conduct experiments across a broad range of CNNs to investigate their alignment with this principle. Our work provides a comprehensive analysis of CNNs regarding closure, identifying limits and thresholds that define their applicability in performing closure over gradually manipulated stimulus classes.

---

\*This work was supported by the European Research Council (ERC) for the 'Gestalts Relate Aesthetic Preferences to Perceptual Analysis' (GRAPPA) Project under Grant 101053925.

## 2 Preliminaries

### 2.1 Gestalt laws of perceptual organization

Human visual perception has multiple, hierarchical processing stages. At the low-level processing stages, the visual system encodes features like luminance and color, at the mid-level stages elements are grouped and organized, and at the high-level stages tasks of identifying and understanding are performed. One significant contribution to research on visual perception was the formulation of the principles of perceptual grouping, often also referred to as “Gestalt Laws”. Grouping describes the phenomenon that observers perceive certain elements as belonging to each other compared to others, which appear distinct. The classic laws of grouping are: proximity, similarity, common fate, good continuation, symmetry, parallelism and closure [2].

### 2.2 Law of closure

The law of closure refers to a shared human preference for closed shapes over disjointed individual elements. This leads the visual system to complete contours and to create closed shapes. Depending on the elements and mechanisms at play, this can be referred to as contour integration or contour completion. Contour integration occurs when distinct elements such as dots or lines are integrated into the contour of a shape [2]. Contour completion refers to the integration of smooth contours [2]. If a shape or scene is completed behind an occluder, this is referred to as amodal completion. If the perception of an illusory contour or surface is triggered by stimulus characteristics, this can be referred to as modal completion [3]. The distinction between the two is made by the presence or absence of certain visual qualities in the completion [3].

## 3 Related Work

Amanatiadis *et al.* [4] trained AlexNet [5] and Inception V1 (GoogLeNet) [6] on the MNIST [7] and ImageNet [8] datasets to explore six core Gestalt laws. They measured closure as a function of occlusion percentage and found that the closure principle is effective in the CNNs up to approximately 30% occlusion, beyond which the models’ performance decreases. Another study examined a wider range of models including ResNet-152 [9] and DenseNet-201 [10], and found mixed evidence of perceptual grouping [11]. Ehrensperger *et al.* [12] analyzed how AlexNet and Inception V1 classify Kanizsa triangles (a famous example of modal completion) and modified triangles with sections of the edges removed, reporting closure. Kim *et al.* [13] showed that both Inception and a simple CNN, trained to classify natural images, exhibit closure on synthetic displays of edge fragments. They tested these CNNs on incomplete triangles and demonstrated that they were more likely to be recognized as similar to complete triangles rather than disordered fragments. Baker *et al.* [14] fine-tuned AlexNet for a shape discrimination task, training it to classify wireframes and Kanizsa

squares as ‘fat’ and ‘thin’. In contrast to earlier work, they concluded that neural networks do not perceive illusory contours. With limited stimuli and metrics employed on only a few CNNs, making a definitive statement about the presence of closure is not possible. Our method trains on complete polygons and tests on incomplete ones. Our approach, along with our examination of a broader range of CNN models, provides a more comprehensive assessment of closure performance in CNNs.

## 4 Convolutional neural networks

CNNs have greatly advanced computer vision by effectively learning spatial features from images through a series of convolutional layers. Notable CNN architectures have emerged, each characterized by unique design choices and strengths. Early CNN models like AlexNet [5] and VGG16 [15] achieved impressive accuracy but suffered from high computational complexity due to their reliance on numerous small convolutional filters stacked together. Recent advancements, exemplified by architectures such as ResNet, SqueezeNet [16], and DenseNet, have addressed this challenge by introducing innovative techniques such as residual connections, and efficient filter design.

Moreover, Inception network [6] and ShuffleNet [17] improve accuracy by incorporating filters of varying sizes within a single layer and leveraging channel shuffle operations, respectively. EfficientNet [18] employs a compound scaling method to optimize network depth, width, and resolution according to specific resource constraints. MobileNet [19], with its lightweight design and depth-wise separable convolutions, is particularly suitable for deployment on mobile devices.

In our experiment, we freeze all layers of the mentioned CNNs. The last layer is replaced with a fully-connected layer with 10 nodes corresponding to 10 different polygons in the dataset. The networks are fine-tuned for 25 epochs with a batch size of 32. We use the Adam optimizer [20] with a learning rate of .001 and cross-entropy loss function.

## 5 Experimental results

### 5.1 Dataset

For the experiment, we generated a novel dataset with a training set (Fig. 1 (a)) consisting of 320 complete polygons, which vary in the number of sides (3 to 12),  $\theta_{global}$  (defined as the degree to which the polygon rotates around its centre in the plane;  $0^\circ$ ,  $15^\circ$ ,  $30^\circ$ ,  $45^\circ$ ,  $60^\circ$ ,  $75^\circ$ ,  $90^\circ$ , and  $105^\circ$ ), background colours (white or black), and centre positions (the centre of the canvas or 16 pixels to the left and 16 pixels to the bottom of the canvas’ centre). Assuming the polygon has  $N$  sides, then its  $\theta_{global}$  could have the value of  $180(N - 2) * i / N$  ( $i = 0, 1, 2, 3, 4, 5, 6, 7$ ).

The test set (Fig. 1 (b)) for the classification includes polygons which differ from each other in the removal percentages. This refers to the space between the line fragments that potentially form the contour of the shape. The removal

percentage of a polygon could be 0 (a complete polygon), 10, 20, 30, 40, 50, 60, 70, 80, or 90. A removal percentage means that on each side of a polygon section of the contour has been removed (or become invisible) and the invisible part is always in the middle of each side. This setting will result in 10 test sets (based on the removal percentage) each including 320 images.

## 5.2 Closure Measurement

We employ the percentage of the correctly classified instances in the test set ( $p$ ) as the indicator of the closure effect. A stable classification performance through the different removal percentage stages above 10% (randomly selecting a class for a given instance) would indicate evidence for the existence of the closure effect in the model.

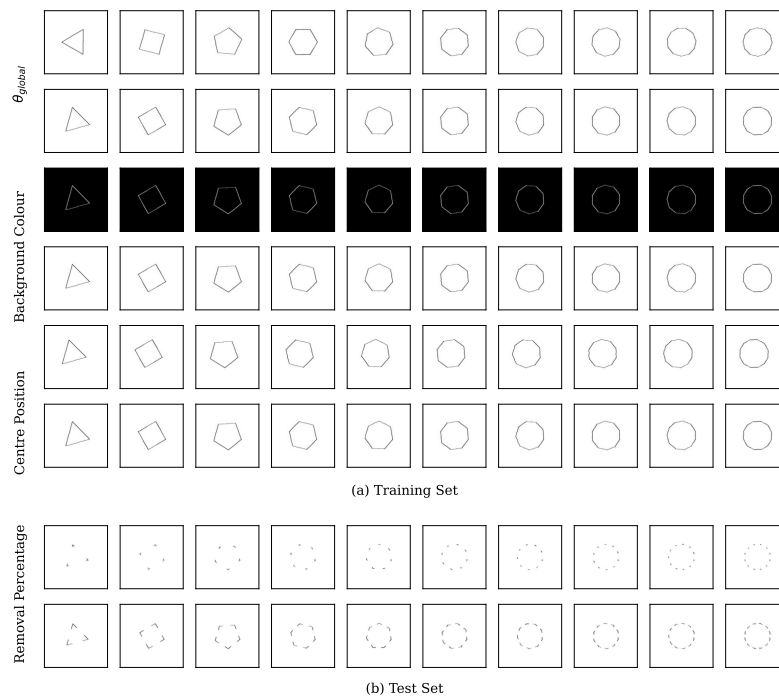


Fig. 1: Image examples used in the (a) training set (b) test set. Each column indicates a different type of polygon (i.e., triangle, square, pentagon, hexagon, and so on). In (a), the top two rows show polygons with different  $\theta_{global}$ , the middle two rows feature different background colours, and the bottom two rows illustrate various positions. In (b), each row shows polygons with varying percentages of removal from each side. The settings for  $\theta_{global}$ , background colours, and centre positions of the polygons in the test set are the same as those used in the training set, although these variables are not depicted in (b).

### 5.3 Results and discussion

The performance of models on the training set (when the removal percentage is 0) is depicted in Fig. 2. VGG16 and SqueezeNet V1.1 perform relatively well with an accuracy of around 90%. AlexNet and ResNet50 also have a good performance and achieve an accuracy of higher than 70%. Inception V3, ShuffleNet V2 and DenseNet121 obtain an accuracy of more than 60%. However, the accuracy of EfficientNet B0 and MobileNetV3 are between 40% and 50%, respectively, due to their lower complexities. It suggests that they cannot do the task reliably.

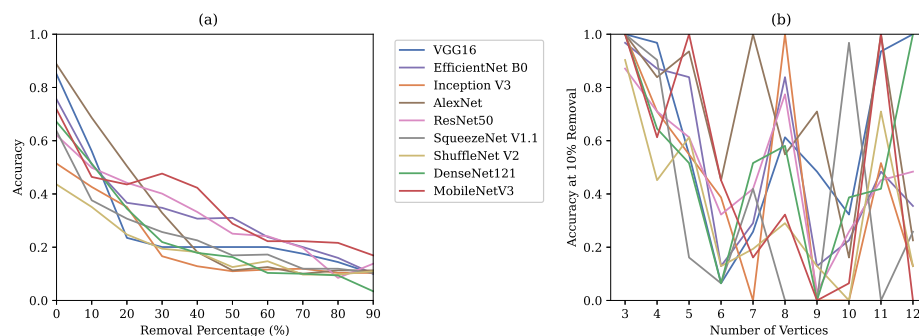


Fig. 2: The percentage of correctly classified examples in the test set versus (a) the removal percentage for each model in each side of a polygon and (b) the number of vertices at 10% removal percentage.

A closer inspection of model performance reveals that accuracy decreases by removal percentage as a general trend (Fig. 2(a)), but the performance of individual CNNs over vertices is inconsistent (Fig. 2(b)). The accuracy of classification remains much higher than chance level (10%) when the removal percentage is not larger than 20% to 30%, although it decreases quickly when the percentage of removal is larger. This aligns with the findings of Amanatiadis *et al.* [4]. However, through the small gradual manipulations in our experiment, we observe a drastic decrease in accuracy already at the first stimulus manipulation. This is not in accordance with contour completion elicited by the closure effect.

## 6 Conclusion

We demonstrate that the classification performance of CNNs deteriorates with partial removal of the contours in incomplete polygons, by using 10 different polygons with gradually manipulated contours. Most models perform the classification task above the chance level for removal percentages under 30%, which could be seen as an indication of performing contour completion. However, classification performance already drastically deteriorates for all networks at only 10% of contour removal, and is inconsistent across number of vertices, which might instead be symptomatic of CNNs high dependence on local features. In

conclusion, CNNs do not display the closure effect with current measurements and datasets.

## References

- [1] M. Wertheimer. Laws of organization in perceptual forms. *Psychologische Forschung*, 4:301–350, 1923.
- [2] J. Wagemans et al. A century of gestalt psychology in visual perception: I. perceptual grouping and figure–ground organization. *Psychological Bulletin*, 138(6):1172, 2012.
- [3] J. Wagemans, R. Van Lier, and B.J. Scholl. Introduction to Michotte’s heritage in perception and cognition research. *Acta Psychologica*, 123(1-2):1–19, 2006.
- [4] A. Amanatiadis, V.G. Kaburlasos, and E.B. Kosmatopoulos. Understanding deep convolutional networks through gestalt theory. In *IEEE International Conference on Imaging Systems and Techniques (IST)*, 2018.
- [5] A. Krizhevsky, I. Sutskever, and G.E. Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.
- [6] C. Szegedy et al. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, 2015.
- [7] Y. LeCun et al. Gradient-based learning applied to document recognition. *IEEE*, 86(11):2278–2324, 1998.
- [8] J. Deng et al. Imagenet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009.
- [9] K. He et al. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.
- [10] G. Huang et al. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4700–4708, 2017.
- [11] V. Biscione and J.S. Bowers. Mixed evidence for gestalt grouping in deep neural networks. *Computational Brain & Behavior*, 6(3):438–456, 2023.
- [12] G. Ehrensperger, S. Stabinger, and A.R. Sánchez. Evaluating cnns on the gestalt principle of closure. In *Artificial Neural Networks and Machine Learning - ICANN 2019: Theoretical Neural Computation*, pages 296–301, 2019.
- [13] B. Kim et al. Neural networks trained on natural scenes exhibit gestalt closure. *Computational Brain & Behavior*, 4(3):251–263, 2021.
- [14] N. Baker et al. Deep convolutional networks do not perceive illusory contours. In *Proceedings of the 40th Annual Conference of the Cognitive Science Society*, Madison, WI, 2018. Cognitive Science Society.
- [15] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [16] F.N. Iandola et al. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and < 0.5 mb model size. *arXiv preprint arXiv:1602.07360*, 2016.
- [17] X. Zhang et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6848–6856, 2018.
- [18] M. Tan and Q. Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*, pages 6105–6114. PMLR, 2019.
- [19] A.G. Howard et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.
- [20] D.P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.