Unsupervised Learning for Global and Local Visual Perception Using Navon Figures

Kayato Nishitsunoi and **Yoshiyuki Ohmura** ({**nishitsunoi, ohmura**}**@isi.imi.i.u-tokyo.ac.jp**)
Department of Mechano-Informatics, Graduate School of Information Science and Technology,
The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan

Yasuo Kuniyoshi (kuniyosh@isi.imi.i.u-tokyo.ac.jp)
Next Generation Artificial Intelligence Research Center (AI Center) and School of Information Science and Technology,
The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan

Abstract

In human visual cognition, there are two types of cognition: holistic cognition, in which the whole is perceived as it is, and featural cognition, in which attention is directed to the components of an object. Navon figures are images that are commonly used for the study of holistic and featural processing in vision. In this paper, we propose a machine learning model that performs unsupervised learning to separate the global and local shapes of Navon figures. In the experiments, by introducing a model that learns image features by exploiting algebraic independence, the global and local shapes of Navon figures were successfully separated and the latent space representing each feature was learned. It was also shown that the feature separation ability was improved by making the structure of the neural network asymmetric. However, the components of the Navon figures used in this study were identical; the proposed model cannot direct attention to each component of Navon figures. Therefore, a model that can direct attention to each component and learn its feature is required in the future.

Keywords: unsupervised learning; global perception; local perception; algebraic independence; Navon figure

Introduction

One of the differences between human and robot perception is the diversity of perception. Humans are capable of multiple visual perceptions of a single image, whereas robots often have only a single interpretation. Two examples of such types of visual perception are holistic and featural perception. Holistic cognition in human vision is a type of cognition in which the whole of an object is taken as it is. Featural cognition is a type of cognition that focuses on the components of an object.

Navon figures are often used in cognitive science to investigate the properties of global and local processing of vision (Martin, 1979; Navon, 1977; Paquet & Merikle, 1984). Figure 1 shows an example of a typical Navon figure, which is a hierarchical visual stimulus that consists of several smaller components. When the local elements of such a Navon figure are relatively small in comparison with the global shape, the local elements are treated as texture and do not affect the overall perception of the figure, even if they are different sub-elements (Kimchi, 1992); only the positions of the local elements are important for perceiving the overall shape (Pomerantz, 2017). In such Navon figures, when two figures are compared, it is possible that the overall structure is "the same" even if the local elements are different, as shown in Figure 2a. Conversely, it is possible that the local elements that compose the Navon figure are "the same" even if the overall image is different, as shown in Figure 2b. Garner (2014) described a separable-dimension stimulus as one in which one attribute does not affect the other. The Navon figure is a visual stimulus that has such a separable-dimension property.

Figure 1: Navon figure.

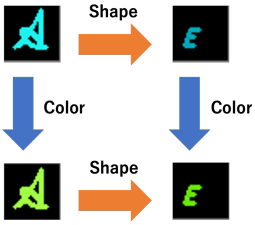
(a) Same global cognition (b) Same local cognition

Figure 2: Various perceptions of Navon figures.

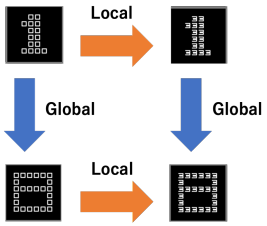
Related to these types of visual perception, there are some machine learning models that deal with such global and local processing, using Navon figures as an evaluation target. One example of such a model is the dual skipping network (Cheng et al., 2018). This model is a neural network based on the idea that the right hemisphere of the brain processes low-spatial-frequency stimuli and the left hemisphere processes high-spatial-frequency stimuli (Kauffmann et al., 2014). Cheng et al. showed that dual skipping networks can recognize global and local shapes of Navon figures with high accuracy. Navon figures were also used as an evaluation target in a study that investigated whether convolutional neural networks (CNNs) have a texture bias, and showed that CNNs can learn both shape and texture elements (Hermann et al., 2020). In the above studies, neural networks were constructed for global and local recognition of Navon figures. However, these models perform supervised learning in which the global and local shapes of Navon figures are provided as teacher data for training.

In the case of humans, even infants as young as three to four months old are capable of both global and local cognition, to some extent (Ghim & Eimas, 1988), and some studies have shown the existence of grouping of local elements

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*



(a) Color and shape independence



(b) Global and local independence

Figure 3: Independent features.

by newborns (Cassia et al., 2002; Farroni et al., 2000). These findings indicate that humans are able to perform such visual cognition to some extent without being taught. Therefore, machine learning models that can learn such global and local processing through unsupervised learning are expected. In this regard, Hsiao et al. (2013) developed Autoencoder models where the distribution of connections between the encodings and the input units differs between the global perception model and the local perception model. They demonstrated the global and local cognition bias using Navon stimuli. This model represents a prior unsupervised machine learning model for global and local perception.

In contrast, this study proposes an unsuperved machine learning model for global and local recognition of Navon figures using a novel machine learning approach recently introduced by Ohmura et al. (2023). We develop a model that uses the algebraic independence (Simpson, 2018) of the global and local features of Navon figures and simple asymmetric structures to separate them into different latent spaces.

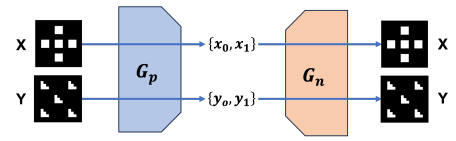
Machine Learning Model Using Algebraic Independence

We selected a recently conceived model that uses algebraic independence to learn image features by unsupervised learning (Ohmura et al., 2023) and used it as the basis of a model to separate the global and local recognition structures of Navon figures. The model of Ohmura et al. (2023) differs from conventional unsupervised representation learning methods using statistical independence (Higgins et al., 2016; Kingma & Welling, 2013) in that it does not assume a prior distribution of latent variables. In addition, it can learn image features in different spaces, rather than on different axes of latent variables.

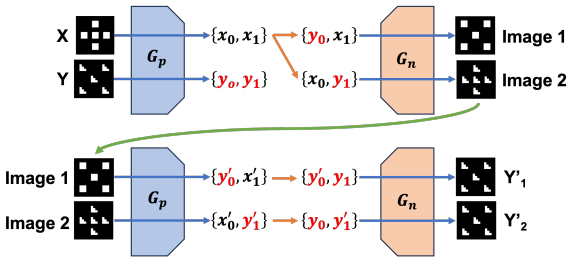
Algebraic Independence

This section describes the algebraic independence introduced in the method of Ohmura et al. (2023). The method focuses on transformations of independent features between image patterns, and learns independent features of images by exploiting the algebraic constraints that these transformations have. Algebraic independence of transformations of independent features requires the following three conditions to hold.

- Identity condition. There exists an identity transformation.



(a) Autoencoder structure



(b) Processing flow in the training step

Figure 4: Schematic of the model using simple Navon figures. Note that Image 1 and Image 2 in Figure 4b are intermediate images generated after proper training, not images during the training process.

- Injectivity condition. Each feature can be transformed separately, and the transformations do not affect each other.
- Commutativity condition. When two features are to be changed, the same transformation occurs regardless of the order in which the two features are changed.

These conditions are derived from the independence structure defined by Simpson (2018). For example, for a transformation that changes color and shape, there exists a transformation that does not change either feature, namely the identity transformation; color and shape can be changed separately; and the transformation of an object has the same result if color is changed before shape or vice versa, as shown in Figure 3a. Therefore, color and shape transformations satisfy the three conditions for algebraic independence. By learning transformations that satisfy such algebraic independence conditions from raw image patterns, this model is able to separate the independent features, color and shape, as different qualia by unsupervised learning.

In this study, we applied the method of Ohmura et al. (2023) to Navon figures to experimentally separate the global shapes and local shapes of the figures into two latent spaces. As shown in Figure 3b, there exist global and local transformations on Navon figures that satisfy the conditions for algebraic independence, such as changes to color and shape. Therefore, it is expected that the latent space for global and local recognition can be learned through unsupervised learning using the method of Ohmura et al. (2023). Using this method to learn features of Navon figures would also provide new insights into the study of global and local cognitive models. This method enables the learning of global and local perception models through the interaction of these models, as detailed in the following section, whereas the prior model (Hsiao et al., 2013) learned each model independently.

Learning Method

We now explain in more detail how the method of Ohmura et al. (2023) separates independent features using simple Navon figure images, as shown in Figure 4. This model consists of an encoder G_p that computes features from an image and a decoder G_n that generates an image from the features, as shown in Figure 4a.

$$(\boldsymbol{x}_0, \boldsymbol{x}_1) = G_p[\mathrm{X}] \tag{1}$$

$$\mathrm{X} = G_n(\boldsymbol{x}_0, \boldsymbol{x}_1) \tag{2}$$

The model learns to make each latent vector represent an independent feature of the input image. In the learning process, the method requires two image patterns as inputs, as shown in Figure 4b. The first image X is the original image pattern before transformation and the second image Y is the image pattern after transformation. The model learns the two independent transformations that transform image X to image Y. It should be noted that the model performs unsupervised learning. Therefore, the two input images are randomly sampled from the same datasets and it is not always possible to transform X to Y in two transformation steps. In such cases, the model learns the identity transformation.

First, from images X and Y, pairs of latent vectors $(\boldsymbol{x}_0, \boldsymbol{x}_1)$ and $(\boldsymbol{y}_0, \boldsymbol{y}_1)$ that represent the features of each image are produced by encoder G_p. Second, pairs of latent vectors $(\boldsymbol{y}_0, \boldsymbol{x}_1)$ and $(\boldsymbol{x}_0, \boldsymbol{y}_1)$ are created by replacing one of the latent vectors in $(\boldsymbol{x}_0, \boldsymbol{x}_1)$ with the corresponding latent vector in $(\boldsymbol{y}_0, \boldsymbol{y}_1)$. Finally, from these pairs of latent vectors, images 1 and 2 are generated by decoder G_n.

$$\mathrm{Image\,1} = G_n(\boldsymbol{y}_0, \boldsymbol{x}_1) \tag{3}$$

$$\mathrm{Image\,2} = G_n(\boldsymbol{x}_0, \boldsymbol{y}_1) \tag{4}$$

The generated images (1 and 2) are the images in which one of the features of image X has been exchanged for the corresponding feature of image Y. Therefore, this step represents one of the two transformations required to obtain image Y from image X. In the second step, pairs of latent vectors are computed from the generated images (1 and 2), again using encoder G_p. This time, the other latent vector, which was not exchanged in the previous step, is exchanged. Images Y'_1 and Y'_2 are generated from these latent vectors by decoder G_n. These two image generation steps result in two transformations from image X to image Y. If these transformations satisfy algebraic independence, images Y'_1 and Y'_2 are the same as image Y.

$$(\boldsymbol{y}'_0, \boldsymbol{x}'_1) = G_p[\mathrm{Image\,1}] \tag{5}$$

$$(\boldsymbol{x}'_0, \boldsymbol{y}'_1) = G_p[\mathrm{Image\,2}] \tag{6}$$

$$\mathrm{Y}'_1 = G_n(\boldsymbol{y}'_0, \boldsymbol{y}_1) \tag{7}$$

$$\mathrm{Y}'_2 = G_n(\boldsymbol{y}_0, \boldsymbol{y}'_1) \tag{8}$$

The final loss $\mathcal{L}$ to be optimized is defined as follows.

$$\mathcal{L} = MSE(\mathrm{Y} - \mathrm{Y}'_1) + MSE(\mathrm{Y} - \mathrm{Y}'_2) \tag{9}$$

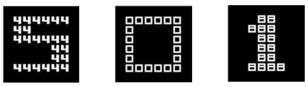


Figure 5: Training dataset.

In this equation, MSE represents the mean squared error between image Y and image Y′. This method learns the encoder G_p and decoder G_n by minimizing this loss to ensure that the transformations between the two image patterns are algebraically independent. That is, the two latent vectors computed by the encoder G_p are learned to be independent features and these vectors form latent spaces that represent each independent feature.

The model can learn to satisfy the identity condition when the two input images are identical, and can satisfy the commutativity condition by minimizing the loss. However, it cannot be trained to explicitly satisfy the injectivity condition; the features to be learned cannot be fully separated but may be mixed into one latent space. Therefore, we propose a model in which the structure of the neural network is asymmetric to promote the differentiation of the features to be learned instead of using the symmetric network used in the study of Ohmura et al. (2023). To confirm the effect of the asymmetric structure, we conducted experiments to compare the separation performance of the asymmetric and symmetric models.

Experiments and Results

This section describes experiments to learn representations of the global and local shapes of Navon figures using the methods described in the previous section.

Training Dataset

The visual stimuli used for training are shown in Figure 5. We created 64 × 64 pixels and 3ch Navon figure datasets in which both global and local shapes represent a digit between 0 and 9 for the training dataset. The images are black-and-white and all components (local shapes) of each Navon figure are the same. In this study, we evaluated how well the model was able to learn the features of the training dataset. It should be noted that, although there are 100 distinct images, 10,000 pairs of input images X and Y are used in the training process.

Experiment 1: Separating the Global and Local Features of Navon Figures

In this experiment, we used the algebraic independence of Navon figures to separate global and local features into different latent spaces by unsupervised learning. We used the method described in the previous section. Figure 6 and Table 1 describe the network architecture of encoder G_p and decoder G_n used in this experiment. The model's architecture was based on the model by Ohmura et al. (2023), originally designed for 32 × 32 images. To accommodate 64 × 64 images, an additional convolutional layer was added. Furthermore, we introduced an asymmetric structure whereby the

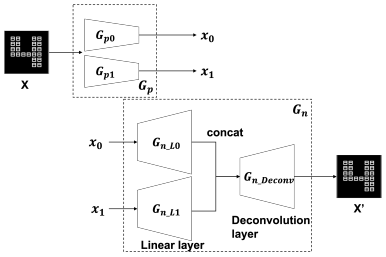


Figure 6: Network structure.

Table 1: Asymmetric network structure.

(a) Encoder G_p

G_{p0}, G_{p1}
Conv2d(3, 32, 4, 2, 1), ReLU
Conv2d(32, 32×2, 4, 2, 1), ReLU
Conv2d(32×2, 32×4, 4, 2, 1), ReLU
Conv2d(32×4, 32×4, 4, 2, 1), ReLU
Flatten
Linear(32×4×4×4, 32×4×4×4), ReLU
Linear(32×4×4×4, 64)

(b) Decoder G_n

G_{n_L0}	G_{n_L1}
Linear(64, 4×64), ReLU	Linear(64, 128×64), ReLU
Linear(4×64, 4×64), ReLU	Linear(128×64, 128×64), ReLU
	Linear(128×64, 128×64), ReLU
Unflatten(4×4, 4, 4)	Unflatten(128×4, 4, 4)
G_{n_Decom}	
Deconv2d((4 + 128)×4, (4 + 128)×4, 4, 2, 1), ReLU	
Deconv2d((4 + 128)×4, (4 + 128)×2, 4, 2, 1), ReLU	
Deconv2d((4 + 128)×2, (4 + 128)×2, 4, 2, 1), ReLU	
Deconv2d((4 + 128)×2, (4 + 128), 4, 2, 1), ReLU	
Conv2d((4 + 128), 3, 1)	

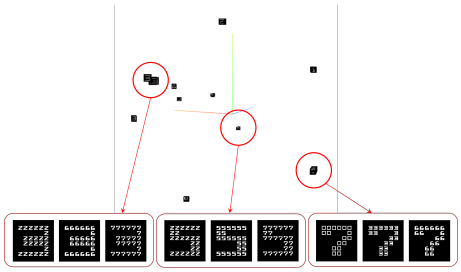
features to be learned are less likely to be mixed in a single latent space. Specifically, we constructed the asymmetric model by making the structure of linear layers G_{n_L0} and G_{n_L1} in decoder G_n asymmetric. G_{n_L1} has more linear layers and parameters than G_{n_L0}. This asymmetric structure facilitates the construction of unbalanced latent spaces for the two latent vectors. The model was trained with a batch size of 100 for 300 training epochs (until the loss fully converged), using RAdam as the optimizer.

Result 1

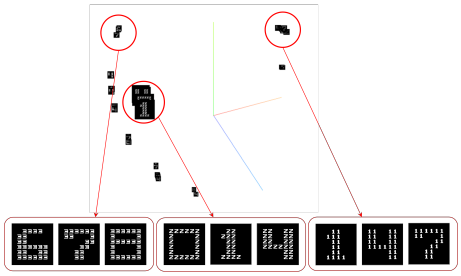
Figure 7 shows the distributions of the two latent vectors of the 100 Navon figures generated in the final training process. Each latent space was analyzed by principal component analysis and plotted in a three-dimensional space. The red circles and arrows in these figures show several Navon figures that are included in the circles in a magnified form. Figure 7a and 7b show that Navon figures with the same global shape are distributed in the same part of the latent space that represents the global shape, and those with the same local shape are distributed in the same part of the latent space that represents the local shape. In summary, by using a machine learning model based on algebraic independence, we were able to construct different recognition structures for global and local shapes of Navon figures.

Experiment 2: Effect of the Asymmetric Structure

In this experiment, we tested the effect of the asymmetric structure by comparing the learning performance of several asymmetric structures with corresponding symmetric structures. We conducted four comparative experiments, creating an asymmetric model and two corresponding symmetric models for each. Table 2 shows the layer depths of the



(a) Global latent space



(b) Local latent space

Figure 7: Latent spaces of Navon figures.

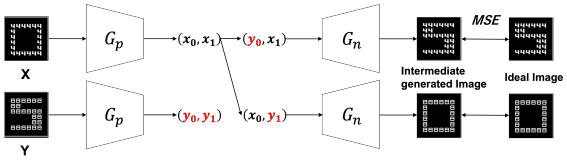


Figure 8: Quantitative evaluation method.

asymmetric and symmetric structures. The layer depths of the asymmetric structures are the same as those used in Experiment 1, with two-depth linear layers for G_{n_L0} and three-depth layers for G_{n_L1}. The asymmetry can be strengthened by adjusting the values of n_0 and n_1 in decoder G_n. We constructed two symmetric models in which the layer depths and parameters of the linear layers of decoder G_n were the same for the two latent vectors. Symmetric model 1 has two-depth linear layers and symmetric model 2 has three-depth linear layers for both G_{n_L0} and G_{n_L1}. For the asymmetric models, we fixed the hyperparameter $n_0 = 4$ and only changed the hyperparameter $n_1 = 4, 66, 128, 190$. The n_0 and n_1 in both symmetric models were set so that the numbers of learnable parameters of decoder G_n were similar to those of the corresponding asymmetric models. The values of hyperparameters and the numbers of learnable parameters for this experiment are shown in Figure 9. The other experimental conditions, such as the datasets used, batch size, number of epochs, and optimizer, were the same as for Experiment 1.

Evaluation Method This section describes the quantitative evaluation method used for Experiment 2. As shown in Figure 8, we calculated the mean squared error (MSE) between the ideally transformed images and the generated images re-

Table 2: Decoder G_n structure. Symmetric models 1 and 2 have the same depth of linear layers for G_{n_L0} and G_{n_L0}, while the asymmetric model has different depths of linear layers. The red parameters n_0 and n_1 can be used to strengthen the asymmetry of the neural network by changing their respective values.

Symmetric model 1		Symmetric model 2		Asymmetric model	
G_{n_L0}	G_{n_L1}	G_{n_L0}	G_{n_L1}	G_{n_L0}	G_{n_L1}
Linear(64, n_0×64) ReLU	Linear(64, n_1×64) ReLU	Linear(64, n_0×64) ReLU	Linear(64, n_1×64) ReLU	Linear(64, n_0×64) ReLU	Linear(64, n_1×64) ReLU
Linear(n_0×64, n_0×64) ReLU	Linear(n_1×64, n_1×64) ReLU	Linear(n_0×64, n_0×64) ReLU	Linear(n_1×64, n_1×64) ReLU	Linear(n_0×64, n_0×64) ReLU	Linear(n_1×64, n_1×64) ReLU
		Linear(n_0×64, n_0×64) ReLU	Linear(n_1×64, n_1×64) ReLU		Linear(n_1×64, n_1×64) ReLU
Unflatten(n_0×4, 4, 4)	Unflatten(n_1×4, 4, 4)	Unflatten(n_0×4, 4, 4)	Unflatten(n_1×4, 4, 4)	Unflatten(n_0×4, 4, 4)	Unflatten(n_1×4, 4, 4)
G_{n_Decomv}					
Deconv2d((n_0+n_1)×4, (n_0+n_1)×4, 4, 2, 1), ReLU					
Deconv2d((n_0+n_1)×4, (n_0+n_1)×2, 4, 2, 1), ReLU					
Deconv2d((n_0+n_1)×2, (n_0+n_1)×2, 4, 2, 1), ReLU					
Deconv2d((n_0+n_1)×2, (n_0+n_1), 4, 2, 1), ReLU					
Conv2d((n_0+n_1), 3, 1)					

sulting from the conversion of the global or local shape of image X to that of image Y. That is, we calculated the MSE between the intermediate images 1 and 2 of Figure 4b and the ideal images. Because the intermediate generated images are not binary images, an appropriate threshold value (0.2 in this experiment) was set and the images were binarized for the calculation. It is difficult to determine which of the two latent spaces learns the global information and which learns the local information. Therefore, there are two ways to calculate the MSE for the two intermediate images. In this study, the MSE was calculated for each learning step; we determined which latent spaces learned global and local information from the combinations with smaller MSE values in the final learning step, and then used these MSE values for evaluation.

Result 2

Figure 9 shows the mean values of MSE between the two intermediate images and the ideal images at each training step after 30 trials for each comparative experiment using an asymmetric model and two corresponding symmetric models. We employed Bonferroni method to analyze the results of each comparative experiment and the differences among the asymmetric structures. In all cases except for comparative experiment 1 using the asymmetric model with $n_0 = 4$ and $n_1 = 4$, the final values of MSE for the asymmetric model were lower than those for symmetric models 1 and 2 ($p < 0.0167$ between the asymmetric model and symmetric model 1 and $p < 0.0167$ between the asymmetric model and symmetric model 2 for each comparative experiment), indicating that, in most cases, the asymmetric model more appropriately transformed the global and local shapes of the Navon figures. However, in the case of comparative experiment 1 using the asymmetric model with $n_0 = 4$ and $n_1 = 4$, the effect of the asymmetric structure is not evident, and the final MSE value was the highest (the average values are 0.0376, 0.0120, 0.0100 and 0.0157 for the asymmetric models of comparative experiment 1, 2, 3 and 4) and significantly different from the other asymmetric models ($p < 0.00833$ between the asymmetric model with $n_0 = 4$ and $n_1 = 4$ and each of the other asymmetric models). It is thought that the asymmetric model has a lower representation ability to learn the features

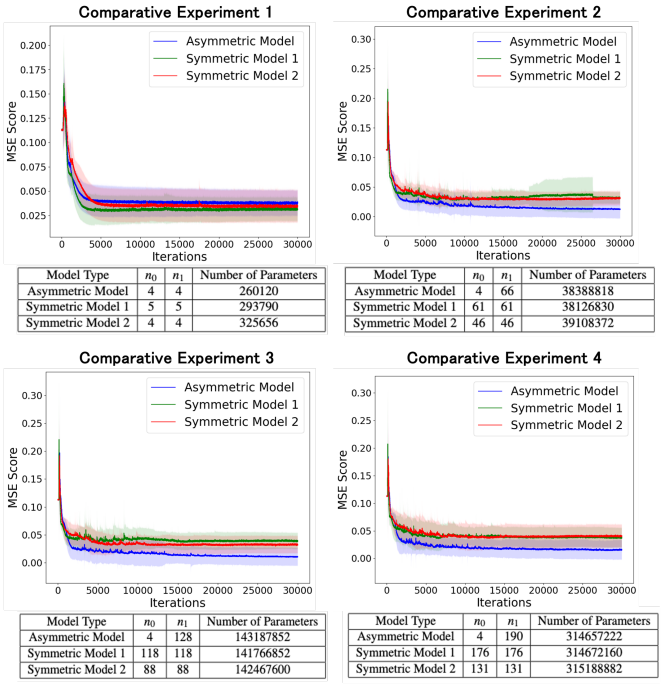


Figure 9: Variation of MSE with the number of learning steps for four comparative experiments with different hyperparameters. The table below each graph shows the hyperparameters and the numbers of learnable parameters of decoder G_n.

of Navon figures due to the low number of learnable parameters in this particular case.

Figure 10 shows the intermediate images generated during the learning process in symmetric model 1 and the asymmetric model of comparative experiment 3. In this figure, we define Image 1 as a globally transformed image and Image 2 as a locally transformed image. This figure shows that, in both models, global learning was completed in the early stages of learning and local learning was achieved in the (approximately) 200 epochs since then. However, in the symmetric model, the local shape was converted along with the global shape and the global feature was not separated from the local

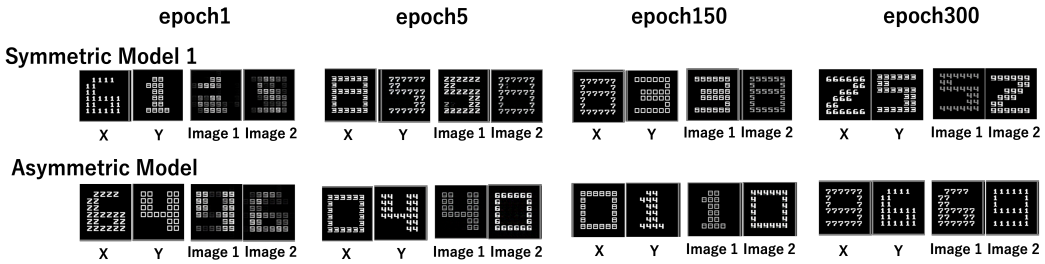


Figure 10: Difference between the symmetric and asymmetric models' learning of the Navon figures.

Figure 11: Navon figure with different local component.

feature. The same phenomenon was also apparent in symmetric model 2. This suggests that the asymmetric structure improves the differentiation of the features to be learned.

In human visual cognition, there is a phenomenon of the lateralization of global and local cognition through hemispheric asymmetry in the processing (Han et al., 2002; Ivry & Robertson, 1998). The differentiation effect of processing features based on asymmetric structure in this study may serve as one hypothesis to explain such a phenomenon. However, since this study only used simple asymmetric structures, further research using models based on actual neuroscientific findings is needed to confirm this connection to lateralization.

Experiment 3: Recognition of Navon Figures With a Different Component

In this experiment, we tested the recognition of Navon figures with a different local element using a model that was already trained. We used Navon figure images with one local element differing from the others, as shown in Figure 11, as evaluation datasets. Specifically, using the model trained in Experiment 1, we generated reconstructed images from the evaluation images and conducted a qualitative evaluation of how the images were recognized.

Result 3

Figure 12 shows the reconstructed image of the Navon figures with one different local element. The left image is the image input to encoder G_p and the right image is the image reconstructed by decoder G_n. This figure shows that, in the reconstructed images, the different local element has been replaced by the surrounding local elements. In this study, we ensured the algebraic independence of global and local transformations of Navon figures by constructing a training dataset that contained only images with unified local features. Therefore, this result is a consequence of the fact that the images in the training dataset that were most similar to the input images

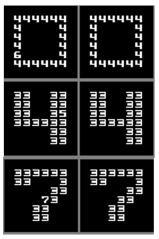


Figure 12: Input image (left) and reconstructed image (right).

contained uniform local components. The proposed model is not one that can direct attention to each local feature; the construction of such a model is a topic for future research.

Conclusion

In this paper, we proposed an unsupervised learning model to form different visual cognitive structures of global and local features. Applying a model that learns image features by optimizing transformations between input patterns to increase their algebraic independence, we proposed a model with an asymmetric structure and conducted experiments to separate global and local features of input images. The experiment used Navon figures, which are hierarchical visual stimuli. We succeeded in generating two latent spaces representing the global and local shapes of the Navon figures by unsupervised learning, exploiting the algebraic independence of the global and local shapes of the Navon figures. We also demonstrated that the asymmetric structure of the model facilitates the differentiation of features to be learned in the learning process. This result suggests a potential correlation with the lateralization of the human brain but further research utilizing models grounded in actual neuroscientific principles is necessary to establish a definitive connection between the observed asymmetric effect and neuroscientific phenomena. And the proposed model was able to learn to satisfy algebraic independence because the local elements of the Navon shapes in the training datasets were all uniform. However, it was unable to recognize hierarchical visual stimuli with different local elements. To learn and recognize such image patterns, a model that can direct attention to each individual element would be required. The construction of such a model is expected in our future research.

Acknowledgments

This work was supported in part by the Grant-in-Aid for Scientific Research (A), under Grant JP22H00528. We thank Takayuki Komatsu, a doctoral student with the Laboratory for Intelligent Systems and Informatics, Graduate School of Information Science and Technology, The University of Tokyo, for his helpful comments on this study.

References

Cassia, V. M., Simion, F., Milani, I., & Umiltà, C. (2002). Dominance of global visual properties at birth. *Journal of Experimental Psychology: General*, *131*(3), 398.

Cheng, C., Fu, Y., Jiang, Y.-G., Liu, W., Lu, W., Feng, J., & Xue, X. (2018). Dual skipping networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4071–4079.

Farroni, T., Valenza, E., Simion, F., & Umilta, C. (2000). Configural processing at birth: Evidence for perceptual organisation. *Perception*, *29*(3), 355–372.

Garner, W. R. (2014). *The processing of information and structure*. Psychology Press.

Ghim, H.-R., & Eimas, P. D. (1988). Global and local processing by 3-and 4-month-old infants. *Perception & Psychophysics*, *43*(2), 165–171.

Han, S., Weaver, J. A., Murray, S. O., Kang, X., Yund, E. W., & Woods, D. L. (2002). Hemispheric asymmetry in global/local processing: Effects of stimulus position and spatial frequency. *Neuroimage*, *17*(3), 1290–1299.

Hermann, K., Chen, T., & Kornblith, S. (2020). The origins and prevalence of texture bias in convolutional neural networks. *Advances in Neural Information Processing Systems*, *33*, 19000–19015.

Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., Mohamed, S., & Lerchner, A. (2016). Beta-vae: Learning basic visual concepts with a constrained variational framework. *International conference on learning representations*.

Hsiao, J. H., Cipollini, B., & Cottrell, G. W. (2013). Hemispheric asymmetry in perception: A differential encoding account. *Journal of Cognitive Neuroscience*, *25*(7), 998–1007.

Ivry, R. B., & Robertson, L. C. (1998). *The two sides of perception*. MIT press.

Kauffmann, L., Ramanoël, S., & Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Frontiers in integrative neuroscience*, *8*, 37.

Kimchi, R. (1992). Primacy of wholistic processing and global/local paradigm: A critical review. *Psychological bulletin*, *112*(1), 24.

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Martin, M. (1979). Local and global processing: The role of sparsity. *Memory & Cognition*, *7*, 476–484.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive psychology*, *9*(3), 353–383.

Ohmura, Y., Shimaya, W., & Kuniyoshi, Y. (2023). An algebraic theory to discriminate qualia in the brain. *arXiv preprint arXiv:2306.00239*.

Paquet, L., & Merikle, P. M. (1984). Global precedence: The effect of exposure duration. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *38*(1), 45.

Pomerantz, J. R. (2017). Perceptual organization in information processing. In *Perceptual organization* (pp. 141–180). Routledge.

Simpson, A. (2018). Category-theoretic structure for independence and conditional independence. *Electronic Notes in Theoretical Computer Science*, *336*, 281–297.