Which deep learning model can best explain object representations of within-category exemplars?

Dongha Lee

Cognitive Science Research Group, Korea Brain Research Institute, Daegu, Republic of Korea

Deep neural network (DNN) models realize human-equivalent performance in tasks such as object recognition. Recent developments in the field have enabled testing the hierarchical similarity of object representation between the human brain and DNNs. However, the representational geometry of object exemplars within a single category using DNNs is unclear. In this study, we investigate which DNN model has the greatest ability to explain invariant within-category object representations by computing the similarity between representational geometries of visual features extracted at the high-level layers of different DNN models. We also test for the invariability of within-category object representations of these models by identifying object exemplars. Our results show that transfer learning models based on ResNet50 best explained both within-category object representation and object identification. These results suggest that the invariability of object representations in deep learning depends not on deepening the neural network but on building a better transfer learning model.

Introduction

Visual object recognition refers to the human ability of accurately identifying objects with substantial variation in appearance. Human object representations in the high-level visual cortex are invariant to event-specific idiosyncratic properties of objects, such as different viewing conditions (Booth & Rolls, 1998; Grill-Spector et al., 1999), lighting (DiCarlo & Cox, 2007), mirror reversal (Baylis & Driver, 2001; Rollenhagen & Olson, 2000), retinal location (DiCarlo & Maunsell, 2003), object size (Andrews & Ewbank, 2004; Konen & Kastner, 2008), and distance (Andrews & Ewbank, 2004; Hung, Kreiman, Poggio, & DiCarlo, 2005). These representations allow humans to identify within-category object exemplars' proficiency. Most likely, when we recognize an object, we tend to focus more on semantic feature representations (e.g., tool and face) than visual feature representations (e.g., orientation and color).

A growing number of studies have investigated the neural mechanisms of object representation through neurobiologically inspired feedforward processing (DiCarlo, Zoccolan, & Rust, 2012; Felleman & Van Essen, 1991) or deep neural network (DNN) modeling (Devereux, Clarke, & Tyler, 2018; Khaligh-Razavi & Kriegeskorte, 2014). It has been demonstrated in these studies that object representations emerge hierarchically from lower-level visual areas to higher-level semantic areas. Specifically, low-level visual features (e.g., orientation and edge) are encoded in the early visual cortex, whereas high-level visual features (e.g., a hammer and a knife) are encoded in the ventral temporal cortex (Cadieu et al., 2014; Guclu & van Gerven, 2015; Khaligh-Razavi & Kriegeskorte, 2014). In this respect, the manifestation of object representations from a DNN is becoming increasingly important in emulating human object recognition.

DNN models based on convolutional neural networks, which are useful for identifying patterns in images for object recognition (Krizhevsky, Sutskever, & Hinton, 2012), have been used in various cognitive and behavioral neuroscience applications, yielding remarkable results. Some DNN models are representative of the structure of the human visual system. Investigating the hierarchical similarity of object representations between the brain and DNNs (Khaligh-Razavi & Kriegeskorte, 2014) and extracting hierarchical visual features from images (Horikawa & Kamitani, 2017a, 2017b; Wen et al., 2018) via transfer learning from a pretrained network of DNN models is a feasible approach. Importantly, the DNN models can extract visual features by applying convolution and pooling repeatedly, even though images may be different within individual categories.

With the advance of multivariate analyses based on machine learning algorithms, multivariate approaches such as multivariate pattern analysis (MVPA) classification and representational similarity analysis (RSA) can be leveraged to investigate how information is represented. MVPA classification estimates a binary contrast (correct or incorrect) for new patterns, whereas

Citation: Lee, D. (2021). Which deep learning model can best explain object representations of within-category exemplars?. *Journal of Vision*, *21*(10):12, 1–10, https://doi.org/10.1167/jov.21.10.12.

https://doi.org/10.1167/jov.21.10.12 Received February 17, 2021; published September 14, 2021 ISSN 1534-7362 Copyright 2021 The Authors

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

RSA estimates all pairwise similarities (Freund, Etzel, & Braver, 2021; Lewis-Peacock & Norman, 2013). The RSA approach is feasible to test the representational content (or geometry) by comparing a matrix of pairwise similarities between brains and computational models, different species, and brain and behavior (Kriegeskorte & Kievit, 2013; Kriegeskorte, Mur, & Bandettini, 2008). The underlying representational content of different visual tasks has been demonstrated by combining RSA and DNN transfer learning (Dwivedi & Roig, 2019). However, it has recently been clarified that convolutional neural network–based DNN models process visual stimuli in a different manner than that in humans, owing to noise sensitivity (Zhang, Liu, & Suen, 2020) or the use of local information only (Baker, Lu, Erlikhman, & Kellman, 2018; Geirhos et al., 2018). Additionally, despite the high performance of DNN models in visual object recognition, the accuracy of encoding object representations by DNN models for different exemplars of the same basic object is unclear.

Object representations for manipulating tools are one of the important features to discriminate between humans from animals (Ambrose, 2001). Tool-specific information (e.g., object manipulation/function knowledge) is distributed in the tool-preferring brain regions such as the medial fusiform gyrus and posterior middle temporal gyrus (Almeida, Fintzi, & Mahon, 2013; Chao & Martin, 2000; Garcea & Mahon, 2014; Lee, Mahon, & Almeida, 2019; Mahon, Kumar, & Almeida, 2013; Mahon et al., 2007). In a previous study, we have demonstrated that the representational content of within-category tools is stable across different functional MRI scanning days (Lee & Almeida, 2021). In this regard, we have tried to understand how the human brain and DNNs perform invariant within-category representations of tools. We focused on whether DNNs exhibit distinct representational content of within-category tools. Our approach agrees well with the approach of using both object representation similarity and object identification accuracy to evaluate DNNs (Majaj, Hong, Solomon, & DiCarlo, 2015).

Thus, we examine different DNN models to evaluate which model best explains invariant representations across exemplars of within-category objects using transfer learning of public deep learning models and RSA. Specifically, we focus on the similarity between representational geometries of within-category object exemplars and identification between the exemplars. Our results show that deep neural models with high representation similarity can discriminate between exemplars of within-category objects, indicating that deep learning models may demonstrate the invariability of object representations.

Materials and methods

Data acquisition

Grayscale images of 80 different tools with 10 exemplars per tool (a total of 800 images) were adopted from our previous study (Lee & Almeida, 2021). All images were 400 × 400 pixels in size (~10° of visual angle). The data are available at https://osf.io/yx7rn/.

Transfer learning of DNN models

Transfer learning, a popular time-saving deep learning approach, uses pretrained models for a task to train new models for other similar tasks (Rawat & Wang, 2017). In this study, we used nine DNN models pretrained on the ImageNet data set: AlexNet (Krizhevsky et al., 2012), GoogLeNet (Szegedy et al., 2015), ResNet (ResNet50, ResNet101; He, Zhang, Ren, & Sun, 2016), VGGNet (VGG16, VGG19; Simonyan & Zisserman, 2014), InceptionV3 (Szegedy, Vanhoucke, Ioffe, Shlens, & Wojna, 2016), and MobileNetV2 (Sandler, Howard, Zhu, Zhmoginov, & Chen, 2018). Transfer learning was conducted to extract higher-level (last fully connected layer herein) visual features from tool images using the Deep Learning Toolbox in MATLAB 2019b (The MathWorks, Natick, MA) and Graphics Processing Unit (GPU). To transfer layers from pretrained models to new models, the following four steps were conducted:

1. Dividing the tool images into training ($n = 720$, e.g., 1st–9th exemplars of each tool) and validation ($n = 80$, e.g., 10th exemplar of each tool) data sets
2. Replacing the last fully connected layer of each model with a new fully connected layer with 80 output nodes
3. Training a deep learning model from training images with predefined labels
4. Passing a validation image through the trained models to extract features to the new image

This process was conducted using a 10-fold cross-validation.

Representational similarity analysis

RSA can be used to test computational models of visual object recognition (Khaligh-Razavi & Kriegeskorte, 2014; Kriegeskorte et al., 2008; Nili et al., 2014). To characterize the representational content of within-category objects, we conducted RSA

on the higher-level features of tools extracted from the last fully connected layer in each DNN model. Representational dissimilarity matrices (RDMs) were constructed by calculating the correlation distances (i.e., 1 – Pearson's r) between all pairs of higher-level features for 80 tools. The RDMs were separately organized across exemplars ($n = 10$) for each DNN model. As this was the first analysis, a correlation analysis was also performed to measure representational similarity between the two RDMs using 45 pairs of the 10 exemplars.

Object identification analysis

Object identification analysis was performed on the visual features extracted from the last fully connected layers of the DNN models. The correlation coefficients between the visual features of objects for the 45 pairs of 10 exemplars were calculated. The correlation coefficients were vectorized to label the n objects. The class label L for a given correlation coefficient vector r was determined by the class c of the maximal correlation coefficient as follows:

$$L_n = \underset{c}{\operatorname{argmax}}\,(r) \quad (1)$$

Comparison of DNN and brain representations

To investigate how the brain and DNNs perform invariant within-category tool representations, we compared nine DNN representations with neural representations in the left fusiform gyrus (A37lv, lateroventral area 37) and posterior middle temporal gyrus (V5/MT+) in the human Brainnetome Atlas (Fan et al., 2016) adopted from our previous study (Lee & Almeida, 2021). Briefly, neural similarity patterns were elicited by neural responses to 80 tools in tool-preferring regions (A37lv, V5/MT+). For these tool-preferring regions, neural RDMs (80 × 80) were constructed by calculating the correlation distance (1 − Pearson's r) between two neural patterns. Then, we conducted a correlation analysis to calculate the representational similarity between DNN RDMs and neural RDMs for 80 tools. A detailed description of the neural RDMs can be found in Lee and Almeida (2021).

Results

Visual features in the higher-level layers (i.e., last fully connected layers) for the tools were extracted via transfer learning of DNN models. With these visual features, we performed RSA to first construct exemplar-specific RDMs. Then, we characterized DNN representations of within-category object exemplars (Figure 1A). To test how a DNN can specifically discriminate an object among different objects, we conducted object identification equivalent to N-class classification, where the correlations between a pair of objects were compared, and the object with a higher correlation was chosen for the object label (Lee, Yun, Jang, & Park, 2017; Venkatesh, Jaja, & Pessoa, 2020) (Figure 1B). These analyses were repeated for 45 pairwise comparisons of the 10 exemplars. To further test which DNN can best explain representational geometries of within-category objects across exemplars, we calculated the representational similarity between exemplar-specific RDMs in a pairwise manner using Pearson's correlation (Figure 1C).

Object identification accuracy of the DNN models

Object identification involves identifying the specific visual features of within-category object exemplars. Here, we investigated to what degree the DNN models could correctly discriminate an object among several similar objects using high-level visual features that are represented differently in each object. To address this issue, the correlation coefficients between visual features of object exemplars were calculated, and each object was labeled with the highest correlation coefficient. The identification accuracy of each DNN model is displayed in Figure 2. A significantly higher identification accuracy was observed in VGG19, ResNet50, and ResNet101, and there were no differences between the models with high accuracy. All statistical inferences of identification accuracy were based on paired t tests ($p < 0.05$ with Bonferroni correction for the 36 pairwise comparisons of the nine models).

Relationships between object identification and properties of DNN models

Figure 3 shows the correlation results of the object identification accuracy analysis with respect to the validation accuracy and number of layers that are summarized in Table 1. The identification accuracy showed a significant positive correlation with the validation accuracy (Figure 4A, $r = 0.9706$, $p = 1.4 \times 10^{-5}$), but no significant correlation with the number of layers (Figure 4B, $r = 0.1249$, $p = 0.7488$) was observed.

Object representation similarity of DNN models

To test which DNN model best explained the invariant representations across the different exemplars of similar objects, we investigated the similarity between object representations by computing

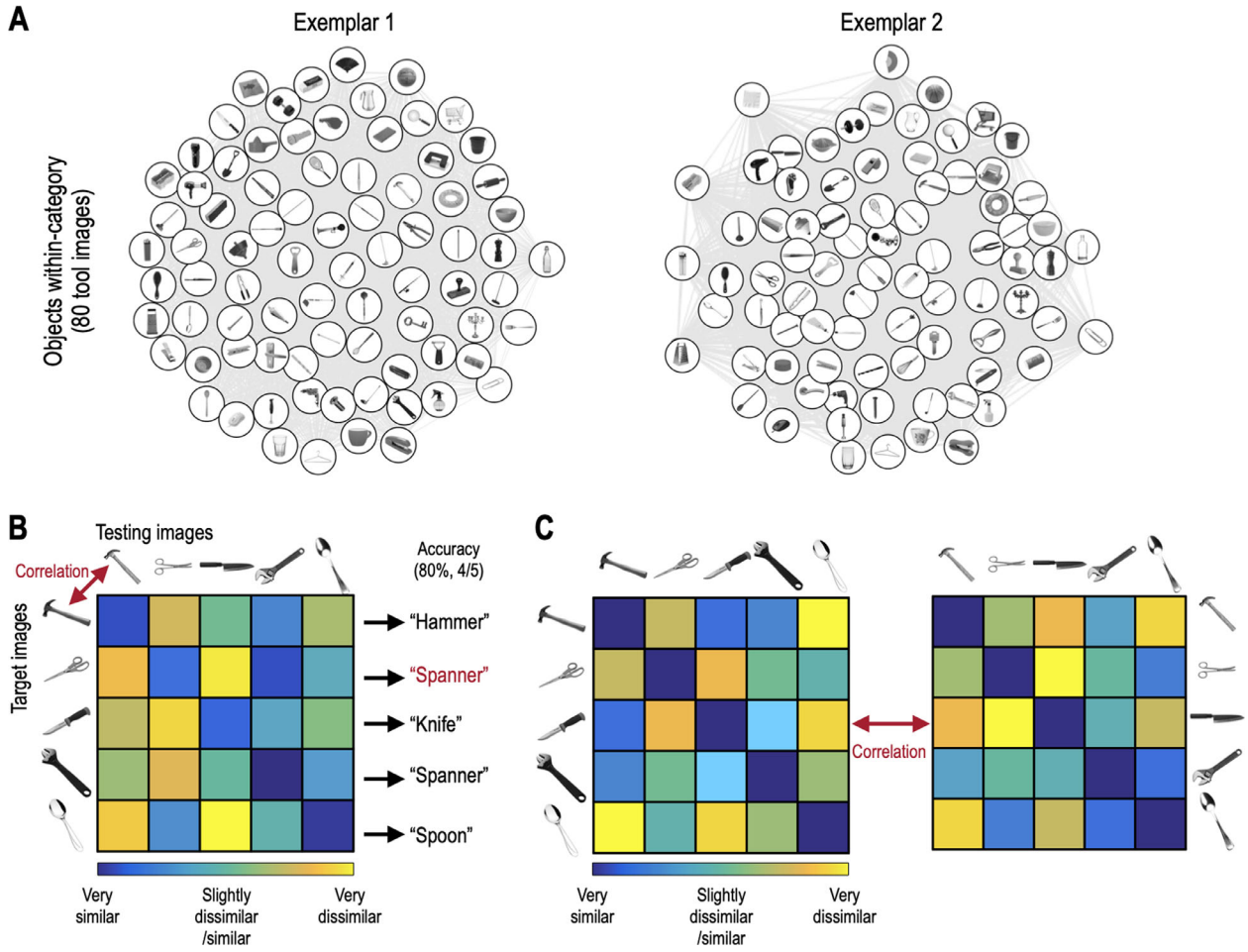

Figure 1. Overview of experimental design. (A) Higher-level visual features for 80 tools were extracted from the last fully connected layers through transfer learning. Multidimensional scaling was used to visualize DNN representations of within-category object exemplars. (B) The identification accuracy was computed by labeling an object in the target images as one of the 80 objects in the testing images. (C) Representational similarity between DNN representations of object exemplars was calculated using the correlation distance between visual features.

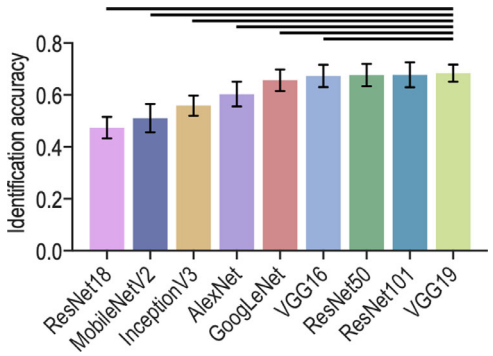

Figure 2. Object identification accuracy for the nine DNN models. The identification accuracies of ResNet50, ResNet101, and VGG19 were significantly higher than those of other DNN models. The lines on the bars indicate the standard error of the mean. The horizontal lines indicate that the aforementioned models performed significantly different from the other models.

the Pearson's correlations between pairs of RDMs for 10 exemplars in each DNN model. As shown in Figure 4, ResNet50 showed a significantly higher representation similarity than that of other DNN models. All statistical inferences of the representation similarity are based on paired t tests, and Bonferroni correction was applied for multiple comparisons of the nine models (adjusted significance level: $p < 0.0056$). Transfer learning using ResNet50 is summarized in Figure 5.

Relationships between DNN and brain representations for within-category object exemplars

The principal finding was that ResNet50 best explained within-category object-specific representations. We employed representational

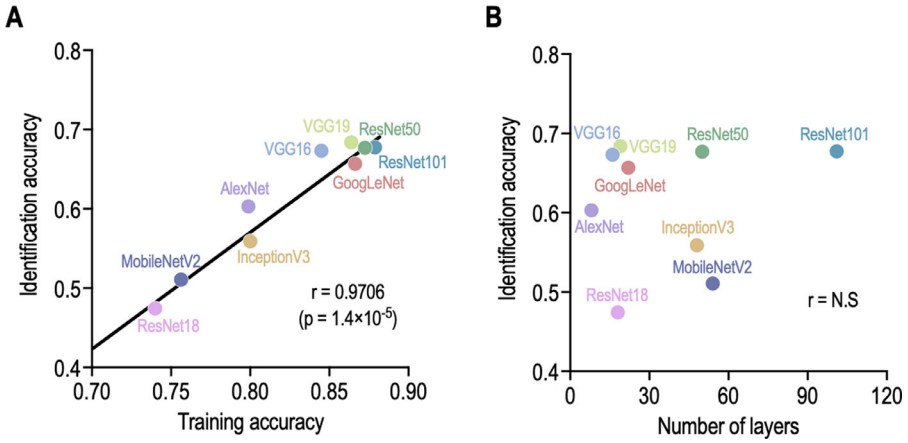

Figure 3. Correlation analysis between object identification accuracy and properties of DNN models. The identification accuracy showed a strong positive correlation with the validation accuracy, whereas no significant correlation was observed between the identification accuracy and number of DNN layers.

DNN models	Validation accuracy (%)	Training time (GPU, s)	Number of layers	Parameters (millions)
ResNet18	74.0 ± 4.1	185.9 ± 9.9	18	11.7
MobileNetV2	75.6 ± 4.4	422.2 ± 6.7	54	3.5
AlexNet	79.9 ± 4.1	131.6 ± 6.7	8	61.0
InceptionV3	80.0 ± 4.1	700.8 ± 21.1	48	23.9
VGG16	84.5 ± 3.6	291.0 ± 3.6	16	138.0
VGG19	86.4 ± 3.7	316.1 ± 4.8	19	144.0
GoogLeNet	86.6 ± 4.0	297.3 ± 8.6	22	7.0
ResNet50	87.3 ± 2.3	428.3 ± 13.3	50	25.6
ResNet101	87.9 ± 4.2	786.4 ± 16.5	101	44.6

Table 1. Training performance of DNN models. *M* ± *SD*.

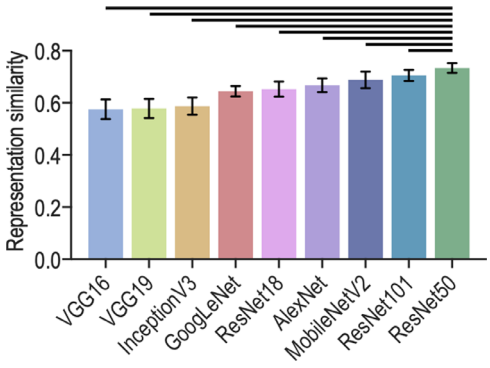

Figure 4. Object representation similarity for nine DNN models. Representation similarity of ResNet50 was significantly higher than that of other DNN models. The lines of the bars indicate the standard error of the mean. The horizontal lines indicate that the performance of ResNet50 was significantly different from the other DNN models.

similarity analysis to further assess how the brain and DNNs perform invariant object representations. As shown in Figure 6, a significantly greater representation similarity was observed for ResNet50, ResNet100, AlexNet, MobileNetV2, InceptionV3, and ResNet10. The ResNet50 model showed the highest representation similarity among the DNN models. This finding may extend our understanding of how humans perform object recognition within a single category.

Discussion

In this study, we report the capabilities of DNN models in discerning the invariance to idiosyncratic properties of within-category object representation. Our results may aid in reducing the gap between human ability and deep learning with respect to object representation. The most important aspect of human identification of objects is the encoding of visual information in terms of the representational geometry for objects. In this respect, it is necessary to note how computational neural networks are functionally similar to the human brain in their handling of object representations (Kriegeskorte et al., 2008). Several studies have investigated the hierarchical similarity of

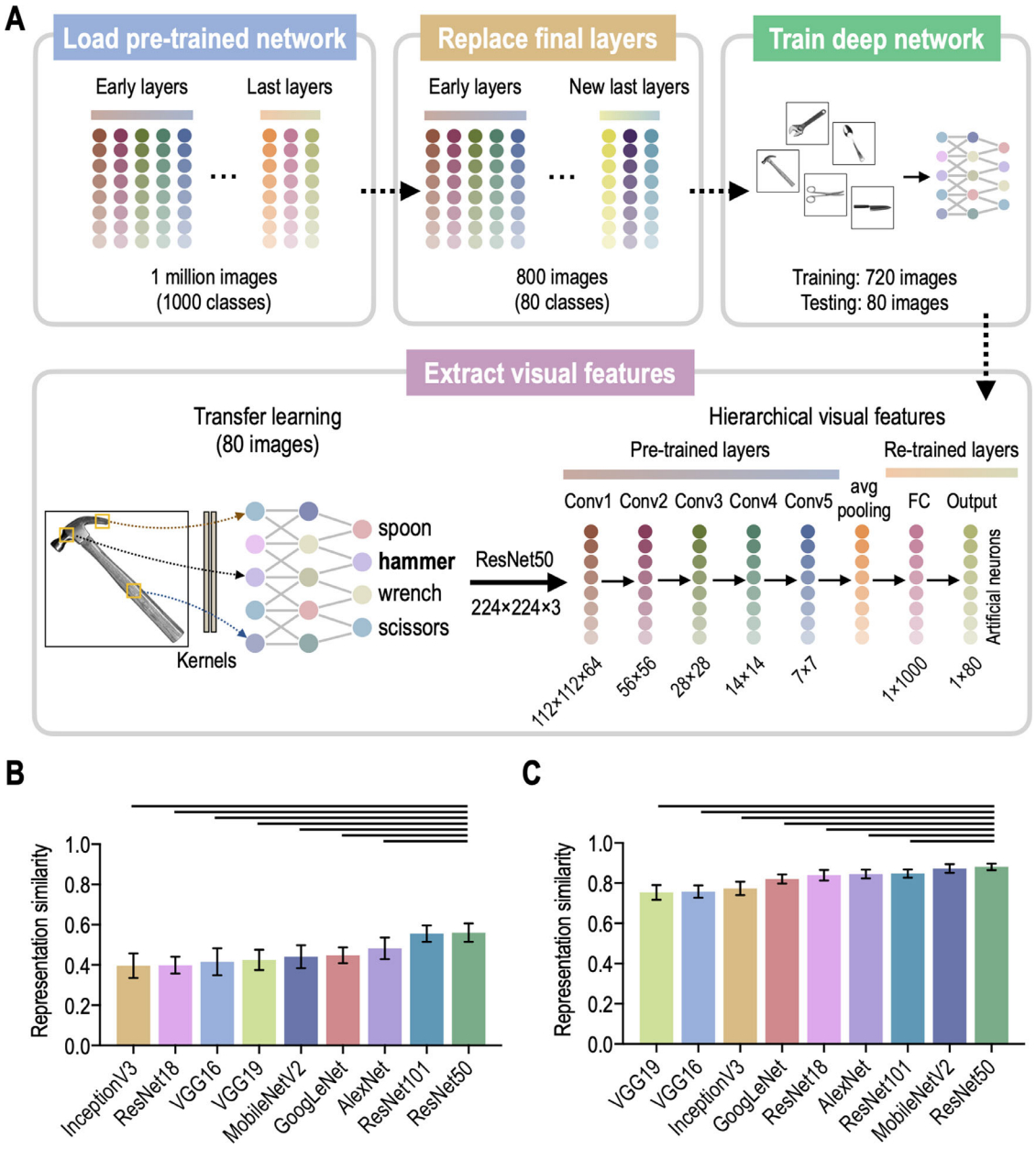

Figure 5. (A) Schematic of transfer learning using the ResNet50 architecture. (B) Object representation similarity using the visual features with low identification accuracy. (C) Object representation similarity using visual features with high identification accuracy.

object representations between the brain and DNNs (Cichy, Khosla, Pantazis, Torralba, & Oliva, 2016; Grill-Spector et al., 1999; Horikawa & Kamitani, 2017a, 2017b; Khaligh-Razavi & Kriegeskorte, 2014; Lee & Almeida, 2021; Wen et al., 2018). Specifically, the human brain has been shown to replicate object representations in the high-level association cortex (Charest, Kievit, Schmitz, Deca, & Kriegeskorte, 2014). This ability is related to human object–similarity judgments that are based on higher-level visual and semantic representations (Mur et al., 2013). In consideration of such findings, we investigated which DNN model best explains invariant object representations across within-category exemplars, as within-category representations are a more a generalizable situation in everyday life than categorical representations.

Our results show that the VGG19, ResNet50, and ResNet101 models performed better object identification than other DNN models. These results are in line with previous findings that indicate that visual features extracted from the high-level DNN layers (e.g., last fully connected layer) are more related to perceived object processing than veridical (i.e., pixel-to-pixel) visual processing (Cichy et al., 2016; Horikawa & Kamitani, 2017b; Lee & Almeida, 2021). These results

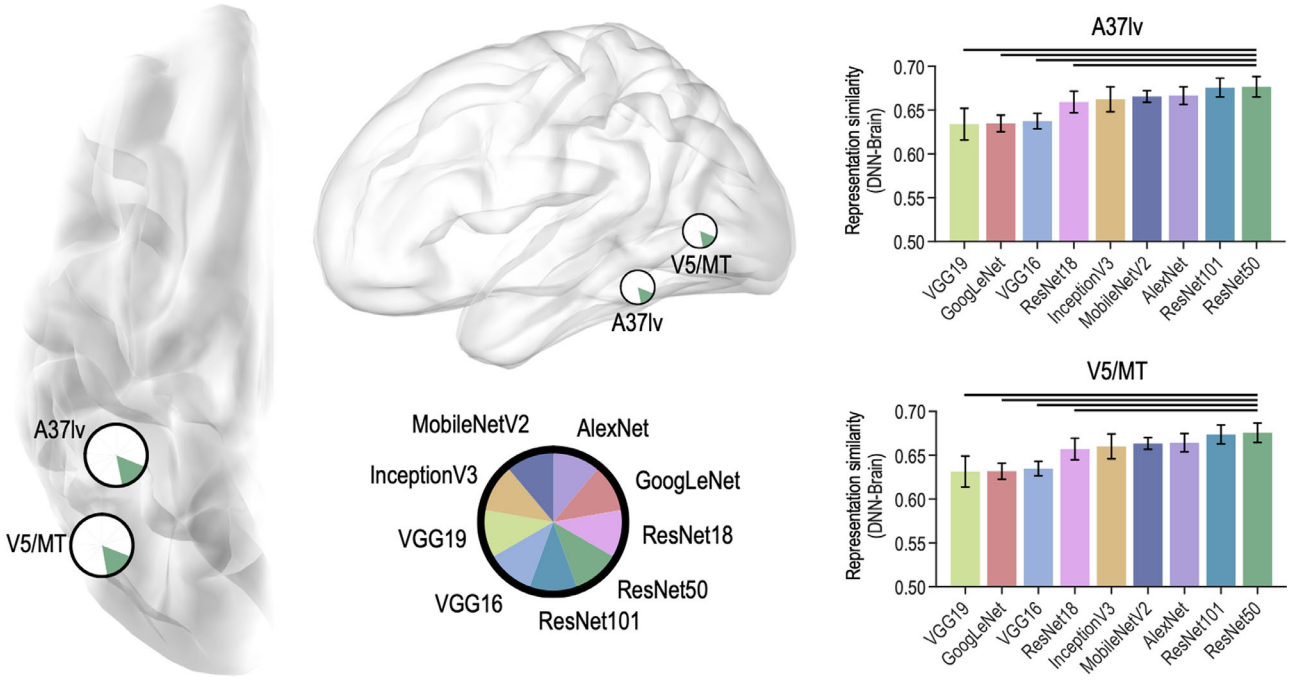

Figure 6. Comparisons of DNN representations in tool-preferring regions. The horizontal lines indicate significant differences in DNN–brain representation similarity between the DNN models.

also agree with the systematic differences observed in the representation of visual images between DNNs and the human brain (Horikawa, Aoki, Tsukamoto, & Kamitani, 2018). Regarding deep learning, it is often considered that deeper networks yield better performance. To test this, we conducted a correlation analysis of object identification accuracy with respect to each DNN's validation accuracy and number of layers. The object identification accuracy strongly correlated with the validation accuracy ($r = 0.97$) but not with number of layers. Thus, building better models may be more important for transfer learning than constructing deeper layers. Importantly, object identification of within-category exemplars is dependent on the tuning of training parameters rather than on the deepening of DNN layers.

By comparing representational geometries for different exemplars of within-category objects, we showed that ResNet50 is superior to other DNN models in within-category object representations across exemplars. Although several DNNs were evaluated on a small sample of objects (10 exemplars of 80 tools), the current results are in agreement with studies that have reported invariant object representations of event-specific idiosyncratic properties of objects (Andrews & Ewbank, 2004; DiCarlo & Cox, 2007; Hung et al., 2005; Konen & Kastner, 2008). These findings suggest that DNN models can explain the invariability of object representations for different exemplars. This invariability seems to be DNN model specific and is used as a measure of the DNN model's goodness of fit for better object representations (Figure 4).

The principal finding was that ResNet50 had both higher object identification accuracy and object representation similarity than other DNN models. As can be seen in Figure 4, within-category object representations were specific to DNN models. Some models exhibited impressive object identification accuracy but low object representation similarity. For example, VGG19 and ResNet50 had indistinguishable accuracies in object identification, but their object similarities were quite different (i.e., on either end of the range measured for this class of objects). To clarify how object representations can differ when object identification accuracies are similar, we investigated the representation similarities of objects showing low identification accuracy and those showing high identification accuracy (Figure 5). In the representation of objects showing low accuracy, VGG19 showed low similarity, whereas ResNet50 showed high similarity (Figure 5B). Similarly, ResNet50 showed high similarity, but VGG19 showed low similarity in the representation of objects showing high accuracy (Figure 5C). These data indicate that even though ResNet50 is much deeper than VGG19, the ResNet50 model was optimized for identification and representation.

One may ask whether DNNs are compared with the human ability for object-specific representations. To confirm DNN capacity and human ability on within-category object representations, we investigated relationships between within-category

object representations using DNN and neural models. When comparing DNN representations and neural representations, ResNet50 showed the best representation similarity among deep learning models (Figure 6). These findings are in agreement with the invariant object recognition between the brain and DNNs reported previously (Majaj et al., 2015). This may be important in understanding how computational neural networks perform within-category object representations similar to the brain's biological neural network.

The present study has a limitation that should be addressed. The object identification accuracies were lower than those of DNN models reported previously in studies on object recognition using ImageNet. This may be owing to the small sample size, which may have led to the poorer performance of the trained DNN models. Thus, transfer learning with a larger sample size should be explored in future studies. Another potential reason for the discrepancy in object identification accuracies may be the complexity of classification. DNN object classification used outputs from the last fully connected layer that were transformed into probabilities by a softmax function, and the probabilities were then used to classify the object. This process is similar to that of object categorization, in which an object is recognized as a member of a single category (i.e., hammer/type level) when satisfying a set of object-general features at the superordinate level (Serre, Oliva, & Poggio, 2007; Warrington & McCarthy, 1987). However, object identification involved recognizing multiple exemplars (i.e., hammer 1, hammer 2/exemplar level) of the same type of object in this study.

Conclusions

We showed that object representation and object identification are most stable across exemplars of within-category objects when using the ResNet50 model. Furthermore, we demonstrated that object identification is dependent on the goodness of fit of the trained DNN. The current results represent a step forward in understanding invariant within-category object representations to develop DNN models with human-equivalent capabilities.

Keywords: invariant object representations, deep neural networks, object exemplars, representation similarity, identification accuracy

Acknowledgments

We thank Jorge Almeida for the stimuli and data and two anonymous reviewers for helpful suggestions.

Supported by the basic research program through the Korea Brain Research Institute, funded by the Ministry of Science and Information Communications Technology (ICT) (21-BR-05-01).

Commercial relationships: none.
Corresponding author: Dongha Lee.
Email: donghalee@kbri.re.kr.
Address: Cognitive Science Research Group, Korea Brain Research Institute, Daegu, Republic of Korea.

References

Almeida, J., Fintzi, A. R., & Mahon, B. Z. (2013). Tool manipulation knowledge is retrieved by way of the ventral visual object processing pathway. *Cortex, 49*(9), 2334–2344.

Ambrose, S. H. (2001). Paleolithic technology and human evolution. *Science, 291*(5509), 1748–1753.

Andrews, T. J., & Ewbank, M. P. (2004). Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. *Neuroimage, 23*(3), 905–913.

Baker, N., Lu, H., Erlikhman, G., & Kellman, P. J. (2018). Deep convolutional networks do not classify based on global object shape. *PLoS Computational Biology, 14*(12), e1006613.

Baylis, G. C., & Driver, J. (2001). Shape-coding in IT cells generalizes over contrast and mirror reversal, but not figure-ground reversal. *Nature Neuroscience, 4*(9), 937–942.

Booth, M. C., & Rolls, E. T. (1998). View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cerebral Cortex, 8*(6), 510–523.

Cadieu, C. F., Hong, H., Yamins, D. L., Pinto, N., Ardila, D., & Solomon, E. A., . . . DiCarlo, J. J. (2014). Deep neural networks rival the representation of primate IT cortex for core visual object recognition. *PLoS Computational Biology, 10*(12), e1003963.

Chao, L. L., & Martin, A. (2000). Representation of manipulable man-made objects in the dorsal stream. *Neuroimage, 12*(4), 478–484.

Charest, I., Kievit, R. A., Schmitz, T. W., Deca, D., & Kriegeskorte, N. (2014). Unique semantic space in the brain of each beholder predicts perceived similarity. *Proceedings of the National Academy of Sciences United States of America, 111*(40), 14565–14570.

Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A., & Oliva, A. (2016). Comparison of deep neural

networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Scientific Reports, 6*, 27755.

Devereux, B. J., Clarke, A., & Tyler, L. K. (2018). Integrated deep visual and semantic attractor neural networks predict fMRI pattern-information along the ventral object processing pathway. *Scientific Reports, 8*, 10636.

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences, 11*(8), 333–341.

DiCarlo, J. J., & Maunsell, J. H. (2003). Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. *Journal of Neurophysiology, 89*(6), 3264–3278.

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron, 73*(3), 415–434.

Dwivedi, K., & Roig, G. (2019). Representation similarity analysis for efficient task taxonomy & transfer learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12387–12396.

Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., & Chen, L., . . . Jiang, T. (2016). The Human Brainnetome Atlas: A new brain atlas based on connectional architecture. *Cerebral Cortex, 26*(8), 3508–3526.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex, 1*(1), 1–47.

Freund, M. C., Etzel, J. A., & Braver, T. S. (2021). Neural coding of cognitive control: The representational similarity analysis approach. *Trends in Cognitive Sciences, 25*(7), 622–638.

Garcea, F. E., & Mahon, B. Z. (2014). Parcellation of left parietal tool representations by functional connectivity. *Neuropsychologia, 60*, 131–143.

Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A., & Brendel, W. (2018). *ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. International Conference on Learning Representations (ICLR)*.

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron, 24*(1), 187–203.

Guclu, U., & van Gerven, M. A. (2015). Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *Journal of Neuroscience, 35*(27), 10005–10014.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778.

Horikawa, T., Aoki, C. S., Tsukamoto, M., & Kamitani, Y. (2019). *Characterization of deep neural network features by decodability from human brain activity. Scientific Data, vol. 6*, 190012.

Horikawa, T., & Kamitani, Y. (2017a). Generic decoding of seen and imagined objects using hierarchical visual features. *Nature Communications, 8*(1), 1–15.

Horikawa, T., & Kamitani, Y. (2017b). Hierarchical neural representation of dreamed objects revealed by brain decoding with deep neural network features. *Frontiers in Computational Neuroscience, 11*, 4.

Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science, 310*(5749), 863–866.

Khaligh-Razavi, S. M., & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Computational Biology, 10*(11), e1003915.

Konen, C. S., & Kastner, S. (2008). Two hierarchically organized neural systems for object information in human visual cortex. *Nature Neuroscience, 11*(2), 224–231.

Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: Integrating cognition, computation, and the brain. *Trends in Cognitive Sciences, 17*(8), 401–412.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis—connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience, 2*, 4.

Krizhevsky, A., Sutskever, I., & Hinton, E. G. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25 (NIPS)*.

Lee, D., & Almeida, J. (2021). Within-category representational stability through the lens of manipulable objects. *Cortex, 137*, 282–291.

Lee, D., Mahon, B. Z., & Almeida, J. (2019). Action at a distance on object-related ventral temporal representations. *Cortex, 117*, 157–167.

Lee, D., Yun, S., Jang, C., & Park, H. J. (2017). Multivariate Bayesian decoding of single-trial event-related fMRI responses for memory retrieval of voluntary actions. *PLoS One, 12*(8), e0182657.

Lewis-Peacock, J. A., & Norman, K. A. (2014). Multi-voxel pattern analysis of fMRI data. In M. Gazzaniga, & R. Mangun (Eds.), *Cognitive*

Neurosciences (pp. 911–920). Cambridge, USA: MIT Press.

Mahon, B. Z., Kumar, N., & Almeida, J. (2013). Spatial frequency tuning reveals interactions between the dorsal and ventral visual systems. *Journal of Cognitive Neuroscience, 25*(6), 862–871.

Mahon, B. Z., Milleville, S. C., Negri, G. A., Rumiati, R. I., Caramazza, A., & Martin, A. (2007). Action-related properties shape object representations in the ventral stream. *Neuron, 55*(3), 507–520.

Majaj, N. J., Hong, H., Solomon, E. A., & DiCarlo, J. J. (2015). Simple learned weighted sums of inferior temporal neuronal firing rates accurately predict human core object recognition performance. *Journal of Neuroscience, 35*(39), 13402–13418.

Mur, M., Meys, M., Bodurka, J., Goebel, R., Bandettini, P. A., & Kriegeskorte, N. (2013). Human object-similarity judgments reflect and transcend the primate-IT object representation. *Frontiers in Psychology, 4*, 128.

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A toolbox for representational similarity analysis. *PLoS Computational Biology, 10*(4), e1003553.

Rawat, W., & Wang, Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review. *Neural Computation, 29*(9), 2352–2449.

Rollenhagen, J. E., & Olson, C. R. (2000). Mirror-image confusion in single neurons of the macaque inferotemporal cortex. *Science, 287*(5457), 1506–1508.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4510–4520.

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences United States of America, 104*(15), 6424–6429.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations (ICLR)*.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., & Anguelov, D., . . . Rabinovich, A. (2015). Going deeper with convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2818–2826.

Venkatesh, M., Jaja, J., & Pessoa, L. (2020). Comparing functional connectivity matrices: A geometry-aware approach applied to participant identification. *Neuroimage, 207*, 116398.

Warrington, K. E., & McCarthy, A. R. (1987). Categories of knowledge: Further fractionations and an attempted integration. *Brain, 110*(5), 1273–1296.

Wen, H., Shi, J., Zhang, Y., Lu, K., Cao, J., & Liu, Z. (2018). Neural encoding and decoding with deep learning for dynamic natural vision. *Cerebral Cortex, 28*(12), 4136–4160.

Zhang, X. Y., Liu, C. L., & Suen, C. Y. (2020). Towards robust pattern recognition: A review. *Proceedings of the IEEE, 108*(6), 894–922.