Deep Spiking Neural Networks with High Representation Similarity Model Visual Pathways of Macaque and Mouse

Liwei Huang[1,2], **Zhengyu Ma**[2*], **Liutao Yu**[2], **Huihui Zhou**[2], **Yonghong Tian**[1,2*]

[1]National Engineering Research Center of Visual Technology, School of Computer Science, Peking University, China
[2]Department of Networked Intelligence, Peng Cheng Laboratory, China
huanglw20@stu.pku.edu.cn, {mazhy, yult, zhouhh}@pcl.ac.cn, yhtian@pku.edu.cn

Abstract

Deep artificial neural networks (ANNs) play a major role in modeling the visual pathways of primate and rodent. However, they highly simplify the computational properties of neurons compared to their biological counterparts. Instead, Spiking Neural Networks (SNNs) are more biologically plausible models since spiking neurons encode information with time sequences of spikes, just like biological neurons do. However, there is a lack of studies on visual pathways with deep SNNs models. In this study, we model the visual cortex with deep SNNs for the first time, and also with a wide range of state-of-the-art deep CNNs and ViTs for comparison. Using three similarity metrics, we conduct neural representation similarity experiments on three neural datasets collected from two species under three types of stimuli. Based on extensive similarity analyses, we further investigate the functional hierarchy and mechanisms across species. Almost all similarity scores of SNNs are higher than their counterparts of CNNs with an average of 6.6%. Depths of the layers with the highest similarity scores exhibit little differences across mouse cortical regions, but vary significantly across macaque regions, suggesting that the visual processing structure of mice is more regionally homogeneous than that of macaques. Besides, the multi-branch structures observed in some top mouse brain-like neural networks provide computational evidence of parallel processing streams in mice, and the different performance in fitting macaque neural representations under different stimuli exhibits the functional specialization of information processing in macaques. Taken together, our study demonstrates that SNNs could serve as promising candidates to better model and explain the functional hierarchy and mechanisms of the visual system.

Introduction

Originally, the prototype of deep neural networks is inspired by the biological vision system (Hubel and Wiesel 1959, 1962). To date, deep neural networks not only occupy an unassailable position in the field of computer vision (LeCun, Bengio, and Hinton 2015), but also become better models of the biological visual cortex compared to traditional models in the neuroscience community (Khaligh-Razavi and Kriegeskorte 2014; Yamins et al. 2014; Yamins and DiCarlo 2016). They have been successful at predicting the neural responses in primate visual cortex, matching the hierarchy of ventral visual stream (Güçlü and van Gerven 2015; Kubilius et al. 2019; Nayebi et al. 2018; Kietzmann et al. 2019), and even controlling neural activity (Bashivan, Kar, and DiCarlo 2019; Ponce et al. 2019). Moreover, as training paradigms of mice (Zoccolan et al. 2009) and techniques for collecting neural activity (de Vries et al. 2020) have been greatly improved, there is a strong interest in exploring mouse visual cortex. Deep neural networks also play an important role in revealing the functional mechanisms and structures of mouse visual cortex (Shi, Shea-Brown, and Buice 2019; Cadena et al. 2019; Nayebi et al. 2022; Bakhtiari et al. 2021; Conwell et al. 2021).

Compared to biological networks, Artificial Neural Networks discard the complexity of neurons (Pham, Packianather, and Charles 2008). Spiking Neural Networks, incorporating the concept of time and spikes, are more biologically plausible models (Maass 1997). To be more specific, because of their capabilities of encoding information with spikes, capturing the dynamics of biological neurons, and extracting spatio-temporal features, deep SNNs are highly possible to yield brain-like representations (Hodgkin and Huxley 1952; Gerstner and Kistler 2002; Izhikevich 2004; Brette et al. 2007; Kasabov et al. 2013). However, deep SNNs have not been employed to model visual cortex due to the immaturity of training algorithms. Recently, a state-of-the-art directly trained deep SNN (Fang et al. 2021), makes it possible to use deep SNNs as visual cortex models.

Contributions. In this work, we conduct large-scale neural representation similarity experiments on SNNs and other high-performing deep neural networks to study the brain's visual processing mechanisms, with three datasets and three similarity metrics (Figure 1). Specifically, to the best of our knowledge, we are the first to use deep SNNs to fit complex biological neural representations and explore the biological visual cortex. We summarize our main contributions in four points as follows.

- We find that SNNs outperform their counterparts of CNNs with the same depth and almost the same architectures in almost all experiments. In addition, even with very different depths and architectures, SNNs can achieve top performance in most conditions.
- By making a more direct comparison between macaques

*Corresponding author.

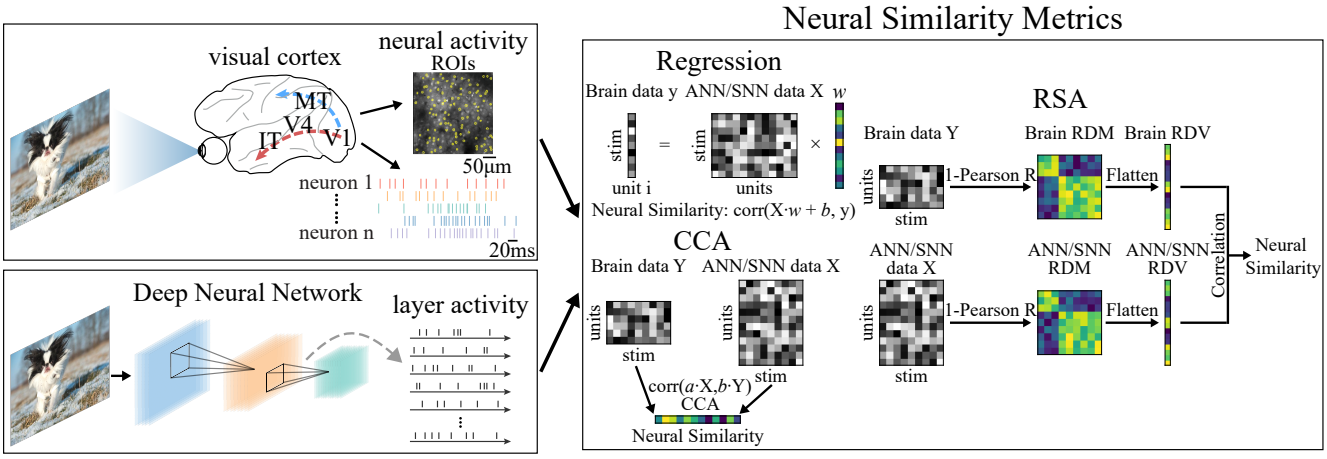


Figure 1: To conduct neural representation similarity experiments, we apply three similarity metrics to a layer-by-layer comparison between the responses of models and the neural activities of visual cortex.

and mice for the first time, we reveal the differences in the visual pathways across the two species in terms of the homogeneity of visual regions and the increases of receptive field sizes across cortical visual pathways, which is consistent with previous physiological work.

- The multi-branch structures in neural networks benefit neural representation similarity to mouse visual cortex, providing computational evidence that parallel information processing streams are widespread between cortical regions in the mouse visual system.
- Comparing the results of two macaque neural datasets under different stimuli, we reveal that the macaque vision system may have functional specialization for processing human faces and other natural scenes.

Altogether, as the first work to apply deep SNNs to fit neural representations, we shed light on visual processing mechanisms in both macaques and mice, demonstrating the potential of SNNs as a novel and powerful tool for research on the visual system. Our codes and appendix are available at *https://github.com/Grasshlw/SNN-Neural-Similarity*.

Related Work

There are plenty of computational models of macaque and mouse visual systems for exploring the visual processing mechanisms recently. We summarize some of the outstanding work in the following.

The network models of macaque visual system. In the early days, studies basically used simple feedforward neural networks as the models of the macaque visual system (Khaligh-Razavi and Kriegeskorte 2014; Yamins et al. 2014; Yamins and DiCarlo 2016). Recently, some bio-inspired or more complex models achieved better performance in fitting the neural representations of macaque visual cortex (Kubilius et al. 2019; Dapello et al. 2020; Zhuang et al. 2021; Higgins et al. 2021). (Kubilius et al. 2019) proposed a brain-like shallow CNN with recurrent connections to better match the macaque ventral visual stream. By mimicking the primary stage of the primate visual system, VOneNets (Dapello et al. 2020) performed more robustly in image recognition while better simulating macaque V1. Moreover, the representations learned by unsupervised neural networks (Zhuang et al. 2021; Higgins et al. 2021) also effectively matched the neural activity of macaque ventral visual stream. Although the above work developed many bio-inspired structures, the networks are still traditional ANNs in nature. Our work introduces deep SNNs for the first time to explore the visual processing mechanisms of macaque visual system.

The network models of mouse visual system. Large-scale mouse neural dataset provided an experimental basis for model studies of mouse visual system (de Vries et al. 2020; Siegle et al. 2021). (Shi, Shea-Brown, and Buice 2019) conducted comparisons between the representations of mouse visual cortex and the VGG16 trained on the ImageNet dataset. In (Bakhtiari et al. 2021), they developed a single neural network to model both the dorsal and ventral pathways with showing the functional specializations. What's more, a large survey of advanced deep networks (Conwell et al. 2021) revealed some hierarchy and functional properties of mice. Similar to the studies of macaque visual system, deep SNNs have never been used to model the mouse visual system. In this work, we not only use SNNs as one of the candidates to fit the representations of mouse visual cortex, but also conduct direct comparisons between macaques and mice to further investigate the functional hierarchy and mechanisms of the two species.

Methods

Neural Datasets

Our work is conducted with three neural datasets. These datasets are recorded from two species under three types of stimuli. More specifically, there are neural responses of mouse visual cortex to natural scene stimuli, and responses of macaque visual cortex to face image and synthetic image stimuli.

Allen Brain mouse dataset. It is part of the Allen Brain Observatory Visual Coding dataset (Siegle et al. 2021) col-

lected using Neuropixel probes from 6 regions simultaneously in mouse visual cortex. Compared to two-photon calcium imaging, Neuropixel probes simultaneously record the spikes across many cortical regions with high temporal resolution. In these experiments, mice are presented with 118 250-ms natural scene stimuli in random orders for 50 times. Hundreds to thousands of neurons are recorded for each brain region. To get the stable neurons, we first concatenate the neural responses (average number of spikes in 10-ms bins across time) under 118 images for each neuron, and then preserve the neurons whose split-half reliability across 50 trials reaches at least 0.8.

Macaque-Face dataset. This dataset (Chang et al. 2021) is composed of neural responses of 159 neurons in the macaque anterior medial (AM) face patch under 2,100 real face stimuli, recorded with Tungsten electrodes. For this dataset, we compute the average number of spikes in a time window of 50-350ms after stimulus onset and exclude eleven neurons with noisy responses by assessing the neurons' noise ceiling. The details of the preprocessing procedure are the same as (Chang et al. 2021).

Macaque-Synthetic dataset. This dataset (Majaj et al. 2015) is also about macaque neural responses which are recorded by electrodes under 3,200 synthetic image stimuli, and used for neural prediction in the initial version of Brain-Score (Schrimpf et al. 2020a). The image stimuli are generated by adding a 2D projection of a 3D object model to a natural background. The objects consist of eight categories, each with eight subclasses. The position, pose, and size of each object are randomly selected. 88 neurons of V4 and 168 neurons of IT are recorded. The neural responses are preprocessed to the form of average firing rate and can be downloaded from Brain-Score.

Models

Since the core visual function of macaque and mouse visual cortex is to recognize objects, the basic premise of model selection is that the model has good performance on object recognition tasks (e.g. classification on ImageNet). Based on this premise, we employ 12 SNNs, 43 CNNs, and 26 vision transformers, all of which are pretrained on the ImageNet dataset and perform well in the classification task. As for SNNs, we use SEW ResNet as the base model, which is the deepest and SOTA directly trained SNN (Fang et al. 2021). Furthermore, by combining the residual block used in SEW ResNet and the hierarchy of the visual cortex, we build several new SNNs and train them on the ImageNet using SpikingJelly (Fang et al. 2020) (see Appendix A for model structures and the details of model training). As for CNNs and vision transformers, we use 44 models from the Torchvision model zoo (Paszke et al. 2019), 22 models from the Timm model zoo (Wightman 2019) and 3 models from the brain-like CNNs, CORnet family (Kubilius et al. 2019). In the feature extraction procedures of all models, we feed the same set of images used in biological experiments to the pretrained models and obtain features from all chosen layers. Different from CNNs and vision transformers, the features of SNNs are spikes in multiple time steps.

Similarity Metrics

To obtain the representation similarity between biological visual cortex and computational models, we apply three similarity metrics to computing similarity scores: representational similarity analysis (RSA) (Kriegeskorte et al. 2008; Kriegeskorte, Mur, and Bandettini 2008), regression-based encoding method (Carandini et al. 2005; Yamins et al. 2014; Schrimpf et al. 2020a,b) and singular vector canonical correlation analysis (SVCCA) (Raghu et al. 2017; Morcos, Raghu, and Bengio 2018). RSA has already been widely used to analyze neural representations of a model and a brain to different stimuli at the population level, while the regression-based encoding method directly fits the model features to neural activity data. SVCCA is originally proposed to compare features of deep neural networks, and then (Shi, Shea-Brown, and Buice 2019) used it to compare representation matrices from mouse visual cortex and DNNs, which demonstrated its effectiveness.

With the same model and same cortical region, we use these metrics for a layer-by-layer comparison to compute the similarity scores. The maximum similarity score across layers for a given cortical region is considered to be the level of representation similarity between the model and the cortical region. Finally, in a given dataset, we take the average score of all cortical regions as the final similarity score for each model, which gives the overall model rankings. The implementation of each similarity metric is as follows.

RSA. For two response matrices $R \in \mathbb{R}^{n \times m}$ from each layer of models and each cortical region, where n is the number of units/neurons and m is the number of stimuli, we calculate the representational similarity between the responses to each pair of image stimuli using the Pearson correlation coefficient r, yielding two representational dissimilarity matrices ($RDM \in \mathbb{R}^{m \times m}$, where each element is the correlation distance $1 - r$). Then, the Spearman rank correlation coefficient between the flattened upper triangles of these two matrices is the metric score.

Regression-Based Encoding Method. Firstly, we run truncated singular value decomposition (TSVD) to reduce the feature dimension of model layers to 40. Secondly, the features after dimensionality reduction are fitted to the representations of each neuron by ridge regression. Finally, we compute the Pearson correlation coefficient between the predicted and ground-truth representations of each neuron and take the mean of all correlation coefficients as the metric score. More specifically, we apply leave-one-out cross-validation to obtain predicted representations of each neuron. For simplicity, we name this method 'TSVD-Reg'.

SVCCA. For both the responses of model layers and cortical regions, we use TSVD to reduce the dimension of unit/neuron to 40, yielding two reduced representation matrices. Then we apply canonical correlation analysis (CCA) to these two matrices to obtain a vector of correlation coefficients (the length of the vector is 40). The metric score is the mean of the vector. Because of the invariance of CCA to affine transformations (Raghu et al. 2017), in this procedure, we only need to ensure that the stimulus dimension is consistent and aligned, even if the unit/neuron dimension is different. Dimensionality reduction plays an important role

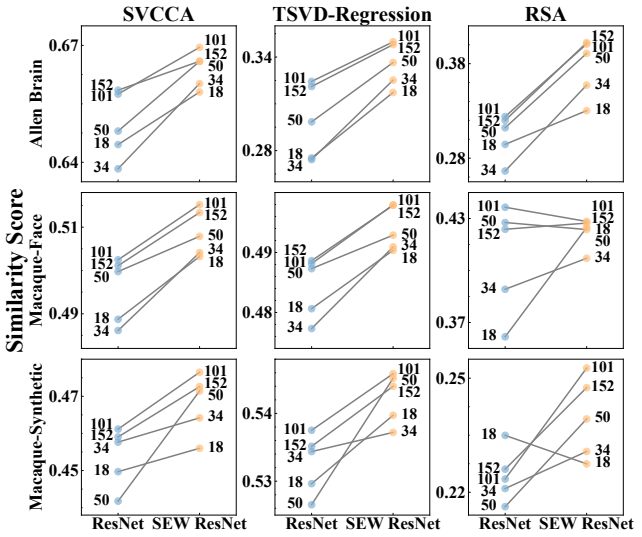


Figure 2: For three datasets and three similarity metrics, each point indicates the final representation similarity score of a model. Each pair of SEW ResNet and ResNet with the same depth are linked by a gray solid line. In almost all conditions, SEW ResNet outperforms ResNet by a large margin.

in this method to make the number of model features comparable to the number of neurons in cortical regions, since the former usually far exceeds the latter. In addition, dimensionality reduction helps to determine which features are important to the original data, while CCA suffers in important feature detection. Using just CCA performs badly, which has been proven by (Raghu et al. 2017).

Results

Comparisons of Representation Similarity Scores between SNNs and Other Types of Models

To check how similar the models are to the visual cortex's mechanisms in visual processing, we rank the final similarity scores of all models and conduct comparisons among three types of models (CNNs, SNNs, and vision transformers). Specially, we focus on comparing SNN (SEW ResNet) and CNN (ResNet) with the same depth and almost the same architectures (Figure 2). The final similarity score of a model is the average similarity score across all cortical regions. (The overall rankings can be found in Appendix B and the comparisons among three types of models are shown in Appendix C.)

Allen brain mouse dataset. No single model achieves the highest final similarity scores with all three metrics. For a fair comparison, we apply the paired t-test to SEW ResNet and ResNet with the same depth. For all three metrics, SEW ResNet performs better than ResNet by a large margin ($t = 5.857$, $p = 0.004$; $t = 7.666$, $p = 0.002$; $t = 7.592$, $p = 0.002$)[1].

[1]The results of the three similarity metrics are separated by semicolons, in the order of SVCCA, TSVD-Reg, and RSA. Other results that appear below also correspond to the three metrics in this order, unless the correspondence is stated in the text.

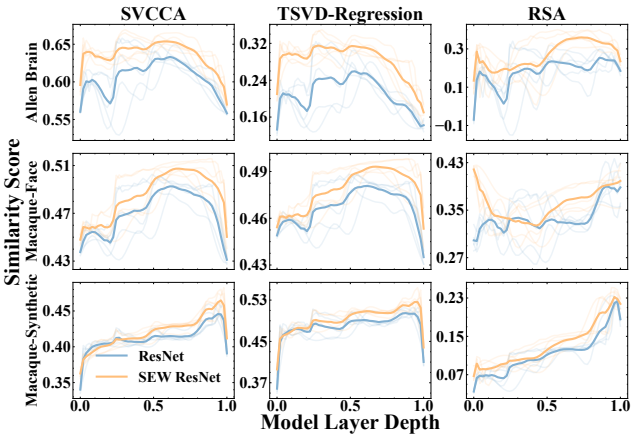


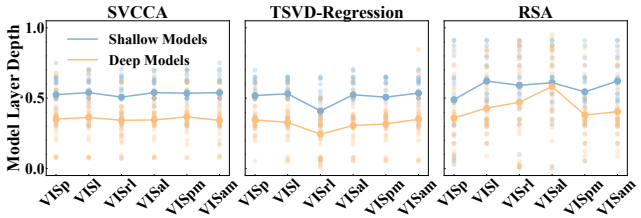
Figure 3: For three datasets and three similarity metrics, we plot the trajectories of similarity score with model layer depth. The models are divided into two groups: ResNet and SEW ResNet. The normalized layer depth ranges from 0 (the first layer) to 1 (the last layer). Because the depths of models are not the same, we first discretize the normalized depth into 50 bins, and then apply the cubic spline interpolation to the scores of each model, yielding the smooth trajectories shown in the plot. The fine, semitransparent lines are the trajectories of each model. The thick lines are the average trajectories among each group.

Macaque-Face dataset. For both SVCCA and TSVD-Reg, Wide-SEW-ResNet14 and Wide-SEW-ResNet8 achieve the first and second highest final similarity scores respectively. But for RSA, TNT-S and Inception-ResNet-V2 take their place and outperform other models by a large margin. As for SEW ResNet and ResNet, the former performs significantly better than the latter for both SVCCA and TSVD-Reg ($t = 8.195$, $p = 0.001$; $t = 7.528$, $p = 0.002$). However, the difference is not significant for RSA ($t = 1.117$, $p = 0.327$). Specifically, the similarity score of SEW ResNet152 is only slightly higher than that of ResNet152, and at the depth of 50 and 101, SEW ResNet's scores are lower than ResNet's.

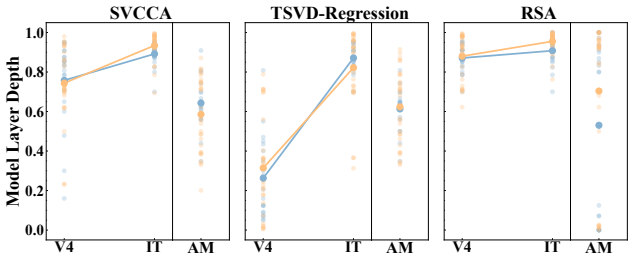
Macaque-Synthetic dataset. Similar to the results of Allen Brain dataset, no model performs best for all three metrics. SEW ResNet performs moderately better than ResNet ($t = 3.354$, $p = 0.028$; $t = 3.824$, $p = 0.019$; $t = 2.343$, $p = 0.079$). The only contrary is that SEW ResNet18 performs worse than ResNet18 for RSA.

Further, to check the details of comparison between the SNNs and their CNN counterparts, we analyze the trajectories of similarity score across model layers (Figure 3). As for ResNet and SEW ResNet with the same depth, the trends of their similarities across model layers are almost the same, but the former's trajectory is generally below the latter's. In other words, the similarity scores of SEW ResNet are higher than those of ResNet at almost all layers.

Taken together, the results suggest that when the overall



(a) Allen Brain mouse dataset



(b) Macaque-Face dataset and Macaque-Synthetic dataset

Figure 4: For three datasets, we plot the normalized depth of the layer that achieves the top similarity score in each cortical region and each metric. Based on model depth, neural networks are divided into two groups: shallow models with less than 50 layers and deep models with more than 50 layers. The normalized layer depth ranges from 0 (the first layer) to 1 (the last layer). Each small point indicates an individual model. The large point indicates the average depth across a group.

architectures and depth are the same, SNNs with spiking neurons perform consistently better than their counterparts of CNNs with an average increase of 6.6%. Besides, SEW ResNet14 also outperforms the brain-like recurrent CNN, CORnet-S, with the same number of layers (see more details in Appendix B). Two properties of SNNs might contribute to the higher similarity scores. On the one hand, IF neurons are the basic neurons of spiking neural networks. The IF neuron uses several differential equations to roughly approximate the membrane potential dynamics of biological neurons, which provides a more biologically plausible spike mechanism for the network. On the other hand, the spiking neural network is able to capture the temporal features by incorporating both time and binary signals, just like the biological visual system during information processing.

Best Layers across Cortical Regions Reveal Functional Hierarchy in the Visual Cortex of Macaques and Mice

To figure out the distinctions in the functional hierarchy between macaques and mice, for each cortical region, we obtain the normalized depth of the layer that achieves the highest similarity score in each model. Then, we divide models (excluding vision transformers) into two groups based on their depths and conduct investigations on these two groups separately. A nonparametric ANOVA is applied to each group for testing whether layer depths change significantly across cortical regions.

For mouse visual cortex (Figure 4 (a)), taking the deep model group as an example, ANOVA shows overall significant changes in depth across cortical regions for TSVD-Reg and RSA (Friedman's $\chi^2 = 49.169$, $p = 2.0 \times 10^{-9}$; $\chi^2 = 19.455$, $p = 0.002$). But there is no significant change for SVCCA ($\chi^2 = 8.689$, $p = 0.122$). According to these results, the differences in depth across regions are indeterminacy and irregular. Meanwhile, the trends of layer depth between some regions contradict the hierarchy observed in physiological experiments of mice (those between VISp and VISrl for TSVD-Reg and between VISal and VISpm for RSA). However, for macaque visual cortex (Figure 4 (b)), there are significant differences ($t = -5.451$, $p = 6.5 \times 10^{-6}$; $t = -8.312$, $p = 2.8 \times 10^{-9}$; $t = -3.782$, $p = 6.9 \times 10^{-4}$, also taking the deep model group as an example) between V4 and IT, and the trend is consistent with the information processing hierarchy in primate visual cortex.

The comparative analyses of the best layer depths of the shallow and deep model groups also exhibit the differences between macaques and mice. For mouse visual cortex, the best layer depths of shallow models are significantly higher than those of deep models. Compared to deep models, most shallow models achieve the top similarity scores in intermediate and even later layers. Differently, for macaque visual cortex, the depth of models has little effect on the depth of the most similar layer. What's more, we find that the most similar layer of mouse visual cortex always occurs after the 28×28 feature map is downsampled to 14×14, which leads to the layer depths' difference between shallow and deep models. Nevertheless, the best layer of macaque IT appears in the last part of networks, where the feature map has been downsampled more times.

In summary, our results might reveal two distinctions in the functional hierarchy between macaques and mice. First, there is a distinct functional hierarchical structure of macaque ventral visual pathway, while there might be no clear sequential functional hierarchy in mouse visual cortex. One explanation is that the mouse visual cortex is organized into a parallel structure and the function of mouse cortical regions are more generalized and homogeneous than those of macaques. Another possibility would be that even though the sequential relations exist among mouse cortical regions as proposed in anatomical and physiological work, they are too weak for the current deep neural networks to capture. Additionally, mice perform more complex visual tasks than expected with a limited brain capacity (Djurdjevic et al. 2018). Consequently, the neural responses of mouse visual cortex may contain more information not related to object recognition that neural networks focus on. Secondly, it is well known that the units in the neural networks get larger receptive fields after downsampling, and through the analyses of differences between two groups of models based on depth, we find the feature map of the best layer for mouse is downsampled fewer times than that for macaque. Based on these results, we provide computational evidence that the increased ratio of the receptive field size in cortical regions across the mouse visual pathway is smaller than those across the macaque visual pathways, which echoes some physio-

Dataset \ Metric	SVCCA	TSVD-Regression	RSA
Allen Brain mouse dataset	$r = -0.654$, $p = 2.0 \times 10^{-6}$	$r = -0.596$, $p = 2.4 \times 10^{-5}$	$r = -0.548$, $p = 1.4 \times 10^{-4}$
Macaque-Face dataset	—	—	—
Macaque-Synthetic dataset	—	—	—

Table 1: The correlation between the similarity scores and the number of parameters. r is Spearman's rank correlation coefficient. "—" indicates that there is no significant correlation.

Dataset \ Metric	SVCCA	TSVD-Regression	RSA
Allen Brain mouse dataset	—	—	—
Macaque-Face dataset	$r = 0.657$, $p = 4.2 \times 10^{-6}$	$r = 0.634$, $p = 1.1 \times 10^{-5}$	$r = 0.527$, $p = 4.7 \times 10^{-4}$
Macaque-Synthetic dataset	—	$r = -0.408$, $p = 0.009$	$r = -0.575$, $p = 1.1 \times 10^{-4}$

Table 2: The correlation between the similarity scores and the model depth. r is Spearman's rank correlation coefficient. "—" indicates that there is no significant correlation.

logical work (Siegle et al. 2021; Zhu and Yang 2013).

Structures and Mechanisms of Models Reveal Processing Mechanisms in the Visual Cortex of Macaques and Mice

To explore the processing mechanisms in the visual cortex of macaques and mice, we investigate the model properties from the whole to the details. As shown in Table 1 and 2, we first measure the correlation between the similarity scores and the sizes (i.e. the number of trainable parameters and the depth) of network models. For Allen Brain mouse dataset, there are significant negative correlations between the similarity scores and the number of parameters for three metrics while there is no correlation with the depth. Conversely, for the two macaque neural datasets, the similarity scores are highly correlated with the depth of networks, but not with the number of parameters. Specifically, there is a positive correlation for Macaque-Face dataset while a negative correlation for Macaque-Synthetic dataset. (We also apply the linear regression to analyze the correlation between the similarity scores and the model size. The results are consistent with Spearman's rank correlation and are shown in Appendix E). Based on these results, we further investigate more detailed properties of neural networks to explain the processing mechanisms in the visual cortex.

For the mouse dataset, on the one hand, the best layer depths show non-significant changes across the mouse cortical regions as mentioned in the previous section. On the other hand, the similarity scores of the mouse dataset are only correlated with the number of model parameters but not with the depth of models. It calls into the question whether any detailed structures in the neural networks help to reduce the number of parameters and improve its similarity to mouse visual cortex. Therefore, we explore the commonalities between models that have the top 20% representation similarities (see Appendix D) for Allen Brain dataset. As expected, the top models contain similar structures, such as fire module, inception module, and depthwise separable convolution. All these structures essentially process information through multiple branches/channels and then integrate the features from each branch. The models with this type of structure outperform other models ($t = 2.411$, $p = 0.024$; $t = 3.030$, $p = 0.007$; $t = 1.174$, $p = 0.247$). Moreover, we apply the depthwise separable convolution to SNNs, which yields a positive effect. The representation similarity of Spiking-MobileNet is higher than SEW-ResNet50 with a similar depth (+0.8%; +3.9%; +12.1%). In fact, some studies using multiple pathways simulate the functions of mouse visual cortex to some extent (Shi et al. 2022; Nayebi et al. 2022). Our results further suggest that not only the mouse visual cortex might be an organization of parallel structures, but also there are extensive parallel information processing streams between each pair of cortical regions (Wang, Sporns, and Burkhalter 2012; Siegle et al. 2021).

For the two macaque datasets with different stimuli, not only are the model rankings significantly different, but also the correlations between the similarity scores and the model depth are totally opposite. These results corroborate the following two processing mechanisms in macaques: the ventral visual stream of primate visual cortex possesses canonical coding principles at different stages; the brain exhibits a high degree of functional specialization, such as the visual recognition of faces and other objects, which is reflected in the different neural responses of the corresponding region (although the face patch AM is a sub-network of IT, they differ in the neural representations). Besides, as shown in Figure 5,

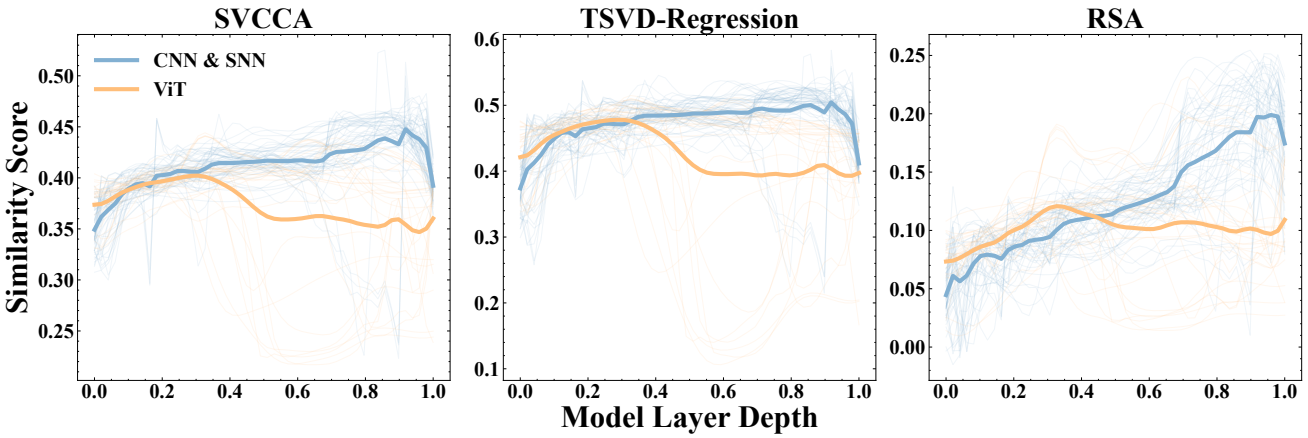


Figure 5: For Macaque-Synthetic dataset, trajectories of similarity score with model layer depth are plotted. The models are divided into two groups: ViT and CNN&SNN. The normalized layer depth ranges from 0 (the first layer) to 1 (the last layer). The calculation and plotting of the trajectories are the same as Figure 3.

the similarity scores of vision transformers reach the maximum in the early layers and then decrease. Differently, the scores of CNNs and SNNs keep trending upwards, reaching the maximum in almost the last layer. On the other hand, Appendix C shows that vision transformers perform well in Macaque-Face dataset but poorly in Macaque-Synthetic dataset. Considering the features extraction mechanism of vision transformers, it divides the image into several patches and encodes each patch as well as their internal relation by self-attention. This mechanism is effective for face images that are full of useful information. However, the synthetic image consists of a central target object and a naturalistic background. When vision transformers are fed with this type of stimuli, premature integration of global information can lead to model representations containing noise from the unrelated background. What's more, when we take all models with the top 20% representation similarities as a whole for analyses, as described in the above paragraph, the properties that enable networks to achieve higher neural similarity are not yet clear. Taken together, the computational mechanism of the better models may reveal core processing divergence to different types of stimuli in the visual cortex.

Discussion

In this work, we take large-scale neural representation similarity experiments as a basis, aided by analyses of the similarities across models and the visual cortical regions. Compared to other work, we introduce SNNs in the similarity analyses with biological neural responses for the first time, showing that SNNs achieve higher similarity scores than CNNs that have the same depth and almost the same architectures. As analyzed in Section 3.1, two properties of SNNs might serve as the explanations for their high similarity scores.

The subsequent analyses of the models' simulation performance and structures indicate significant differences in functional hierarchies between macaque and mouse visual cortex. As for macaques, we observed a clear sequential hierarchy. However, as for mouse visual cortex, some work (Conwell et al. 2021) exhibits that the trend of the model feature complexity roughly matches the processing hierarchy, but other work suggests that the cortex (Shi, Shea-Brown, and Buice 2019; Nayebi et al. 2022) is organized into a parallel structure. Our results are more supportive of the latter. Furthermore, we provide computational evidence not only that the increased ratio of the receptive field size in cortical regions across the mouse visual pathway is smaller than those across the macaque visual pathway, but also that there may be multiple pathways with parallel processing streams between mouse cortical regions. Our results also clearly reveal that the processing mechanisms of macaque visual cortex differ to various stimuli. These findings provide us with new insights into the visual processing mechanisms of macaque and mouse, which are the two species that dominate the research of biological vision systems and differ considerably from each other.

Compared to CNNs, the study of task-driven deep SNNs is just in its initial state. Although we demonstrate that SNNs outperform their counterparts of CNNs, SNNs exhibit similar properties as CNNs in the further analyses. In this work, we only build several new SNNs by taking the hints from the biological visual hierarchy, while many well-established structures and learning algorithms in CNNs have not been applied to SNNs yet. In addition, the neural datasets used in our experiments are all collected under static image stimuli, lacking rich dynamic information to some certain, which may not fully exploit the properties of SNNs. Given that SNNs perform well in the current experiments, we hope to explore more potential of SNNs in future work.

In conclusion, as more biologically plausible neural networks, SNNs may serve as a shortcut to explore the biological visual cortex. With studies on various aspects of SNNs, such as model architectures, learning algorithms, processing mechanisms, and neural coding methods, it's highly promising to better explain the sophisticated, complex, and diverse vision systems in the future.

Ethics Statement

The biological neural datasets used in our experiments are obtained from public datasets or from published papers with the authors' consent.

Acknowledgements

We thank L. Chang for providing Macaque-Face dataset. This work is supported by the National Natural Science Foundation of China (No.61825101, No.62027804, and No.62088102).

References

Bakhtiari, S.; Mineault, P.; Lillicrap, T.; Pack, C.; and Richards, B. 2021. The functional specialization of visual cortex emerges from training parallel pathways with self-supervised predictive learning. In *Advances in Neural Information Processing Systems 34*, 25164–25178.

Bashivan, P.; Kar, K.; and DiCarlo, J. J. 2019. Neural population control via deep image synthesis. *Science*, 364(6439): eaav9436.

Brette, R.; Rudolph, M.; Carnevale, T.; Hines, M.; Beeman, D.; Bower, J. M.; Diesmann, M.; Morrison, A.; Goodman, P. H.; Harris, F. C.; et al. 2007. Simulation of networks of spiking neurons: a review of tools and strategies. *Journal of computational neuroscience*, 23(3): 349–398.

Cadena, S. A.; Sinz, F. H.; Muhammad, T.; Froudarakis, E.; Cobos, E.; Walker, E. Y.; Reimer, J.; Bethge, M.; Tolias, A.; and Ecker, A. S. 2019. How well do deep neural networks trained on object recognition characterize the mouse visual system? In *NeurIPS Neuro AI Workshop*.

Carandini, M.; Demb, J. B.; Mante, V.; Tolhurst, D. J.; Dan, Y.; Olshausen, B. A.; Gallant, J. L.; and Rust, N. C. 2005. Do we know what the early visual system does? *Journal of Neuroscience*, 25(46): 10577–10597.

Chang, L.; Egger, B.; Vetter, T.; and Tsao, D. Y. 2021. Explaining face representation in the primate brain using different computational models. *Current Biology*, 31(13): 2785–2795.

Conwell, C.; Mayo, D.; Barbu, A.; Buice, M.; Alvarez, G.; and Katz, B. 2021. Neural regression, representational similarity, model zoology & neural taskonomy at scale in rodent visual cortex. In *Advances in Neural Information Processing Systems 34*, 5590–5607.

Dapello, J.; Marques, T.; Schrimpf, M.; Geiger, F.; Cox, D.; and DiCarlo, J. J. 2020. Simulating a primary visual cortex at the front of CNNs improves robustness to image perturbations. In *Advances in Neural Information Processing Systems 33*, 13073–13087.

de Vries, S. E.; Lecoq, J. A.; Buice, M. A.; Groblewski, P. A.; Ocker, G. K.; Oliver, M.; Feng, D.; Cain, N.; Ledochowitsch, P.; Millman, D.; et al. 2020. A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex. *Nature neuroscience*, 23(1): 138–151.

Djurdjevic, V.; Ansuini, A.; Bertolini, D.; Macke, J. H.; and Zoccolan, D. 2018. Accuracy of rats in discriminating visual objects is explained by the complexity of their perceptual strategy. *Current biology*, 28(7): 1005–1015.

Fang, W.; Chen, Y.; Ding, J.; Chen, D.; Yu, Z.; Zhou, H.; Masquelier, T.; Tian, Y.; and other contributors. 2020. SpikingJelly. https://github.com/fangwei123456/spikingjelly. Accessed: 2022-07-06.

Fang, W.; Yu, Z.; Chen, Y.; Huang, T.; Masquelier, T.; and Tian, Y. 2021. Deep residual learning in spiking neural networks. In *Advances in Neural Information Processing Systems 34*, 21056–21069.

Gerstner, W.; and Kistler, W. M. 2002. *Spiking neuron models: Single neurons, populations, plasticity*. Cambridge university press.

Güçlü, U.; and 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.

Higgins, I.; Chang, L.; Langston, V.; Hassabis, D.; Summerfield, C.; Tsao, D.; and Botvinick, M. 2021. Unsupervised deep learning identifies semantic disentanglement in single inferotemporal face patch neurons. *Nature communications*, 12(1): 1–14.

Hodgkin, A. L.; and Huxley, A. F. 1952. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, 117(4): 500.

Hubel, D. H.; and Wiesel, T. N. 1959. Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology*, 148(3): 574.

Hubel, D. H.; and Wiesel, T. N. 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160(1): 106.

Izhikevich, E. M. 2004. Which model to use for cortical spiking neurons? *IEEE Transactions on Neural Networks*, 15(5): 1063–1070.

Kasabov, N.; Dhoble, K.; Nuntalid, N.; and Indiveri, G. 2013. Dynamic evolving spiking neural networks for online spatio-and spectro-temporal pattern recognition. *Neural Networks*, 41: 188–201.

Khaligh-Razavi, S.-M.; and Kriegeskorte, N. 2014. Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS computational biology*, 10(11): e1003915.

Kietzmann, T. C.; Spoerer, C. J.; Sörensen, L. K.; Cichy, R. M.; Hauk, O.; and Kriegeskorte, N. 2019. Recurrence is required to capture the representational dynamics of the human visual system. *Proceedings of the National Academy of Sciences*, 116(43): 21854–21863.

Kriegeskorte, N.; Mur, M.; and Bandettini, P. A. 2008. Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, 2: 4.

Kriegeskorte, N.; Mur, M.; Ruff, D. A.; Kiani, R.; Bodurka, J.; Esteky, H.; Tanaka, K.; and Bandettini, P. A. 2008. Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, 60(6): 1126–1141.

Kubilius, J.; Schrimpf, M.; Kar, K.; Rajalingham, R.; Hong, H.; Majaj, N.; Issa, E.; Bashivan, P.; Prescott-Roy, J.; Schmidt, K.; et al. 2019. Brain-like object recognition with high-performing shallow recurrent ANNs. In *Advances in Neural Information Processing Systems 32*, 12785–12796.

LeCun, Y.; Bengio, Y.; and Hinton, G. 2015. Deep learning. *nature*, 521(7553): 436–444.

Maass, W. 1997. Networks of spiking neurons: the third generation of neural network models. *Neural networks*, 10(9): 1659–1671.

Majaj, N. J.; Hong, H.; Solomon, E. A.; and 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.

Morcos, A.; Raghu, M.; and Bengio, S. 2018. Insights on representational similarity in neural networks with canonical correlation. In *Advances in Neural Information Processing Systems 31*, 5732–5741.

Nayebi, A.; Bear, D.; Kubilius, J.; Kar, K.; Ganguli, S.; Sussillo, D.; DiCarlo, J. J.; and Yamins, D. L. 2018. Task-driven convolutional recurrent models of the visual system. In *Advances in Neural Information Processing Systems 31*, 5295–5306.

Nayebi, A.; Kong, N. C.; Zhuang, C.; Gardner, J. L.; Norcia, A. M.; and Yamins, D. L. K. 2022. Mouse visual cortex as a limited resource system that self-learns an ecologically-general representation. *bioRxiv*.

Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. 2019. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, 8024—-8035.

Pham, D. T.; Packianather, M. S.; and Charles, E. 2008. Control chart pattern clustering using a new self-organizing spiking neural network. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 222(10): 1201–1211.

Ponce, C. R.; Xiao, W.; Schade, P. F.; Hartmann, T. S.; Kreiman, G.; and Livingstone, M. S. 2019. Evolving images for visual neurons using a deep generative network reveals coding principles and neuronal preferences. *Cell*, 177(4): 999–1009.

Raghu, M.; Gilmer, J.; Yosinski, J.; and Sohl-Dickstein, J. 2017. SVCCA: Singular vector canonical correlation analysis for deep learning dynamics and interpretability. In *Advances in Neural Information Processing Systems 30*, 6076–6085.

Schrimpf, M.; Kubilius, J.; Hong, H.; Majaj, N. J.; Rajalingham, R.; Issa, E. B.; Kar, K.; Bashivan, P.; Prescott-Roy, J.; Geiger, F.; et al. 2020a. Brain-score: Which artificial neural network for object recognition is most brain-like? *bioRxiv*.

Schrimpf, M.; Kubilius, J.; Lee, M. J.; Murty, N. A. R.; Ajemian, R.; and DiCarlo, J. J. 2020b. Integrative benchmarking to advance neurally mechanistic models of human intelligence. *Neuron*, 108(3): 413–423.

Shi, J.; Shea-Brown, E.; and Buice, M. 2019. Comparison against task driven artificial neural networks reveals functional properties in mouse visual cortex. In *Advances in Neural Information Processing Systems 32*, 5765–5775.

Shi, J.; Tripp, B.; Shea-Brown, E.; Mihalas, S.; and A. Buice, M. 2022. MouseNet: A biologically constrained convolutional neural network model for the mouse visual cortex. *PLOS Computational Biology*, 18(9): e1010427.

Siegle, J. H.; Jia, X.; Durand, S.; Gale, S.; Bennett, C.; Graddis, N.; Heller, G.; Ramirez, T. K.; Choi, H.; Luviano, J. A.; et al. 2021. Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature*, 592(7852): 86–92.

Wang, Q.; Sporns, O.; and Burkhalter, A. 2012. Network analysis of corticocortical connections reveals ventral and dorsal processing streams in mouse visual cortex. *Journal of Neuroscience*, 32(13): 4386–4399.

Wightman, R. 2019. PyTorch Image Models. https://github.com/rwightman/pytorch-image-models. Accessed: 2022-01-17.

Yamins, D. L.; and DiCarlo, J. J. 2016. Using goal-driven deep learning models to understand sensory cortex. *Nature neuroscience*, 19(3): 356–365.

Yamins, D. L.; Hong, H.; Cadieu, C. F.; Solomon, E. A.; Seibert, D.; and DiCarlo, J. J. 2014. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the national academy of sciences*, 111(23): 8619–8624.

Zhu, X.; and Yang, Z. 2013. Multi-scale spatial concatenations of local features in natural scenes and scene classification. *Plos one*, 8(9): e76393.

Zhuang, C.; Yan, S.; Nayebi, A.; Schrimpf, M.; Frank, M. C.; DiCarlo, J. J.; and Yamins, D. L. 2021. Unsupervised neural network models of the ventral visual stream. *Proceedings of the National Academy of Sciences*, 118(3): e2014196118.

Zoccolan, D.; Oertelt, N.; DiCarlo, J. J.; and Cox, D. D. 2009. A rodent model for the study of invariant visual object recognition. *Proceedings of the National Academy of Sciences*, 106(21): 8748–8753.