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.

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 (**??**). 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 (**??**).

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, 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

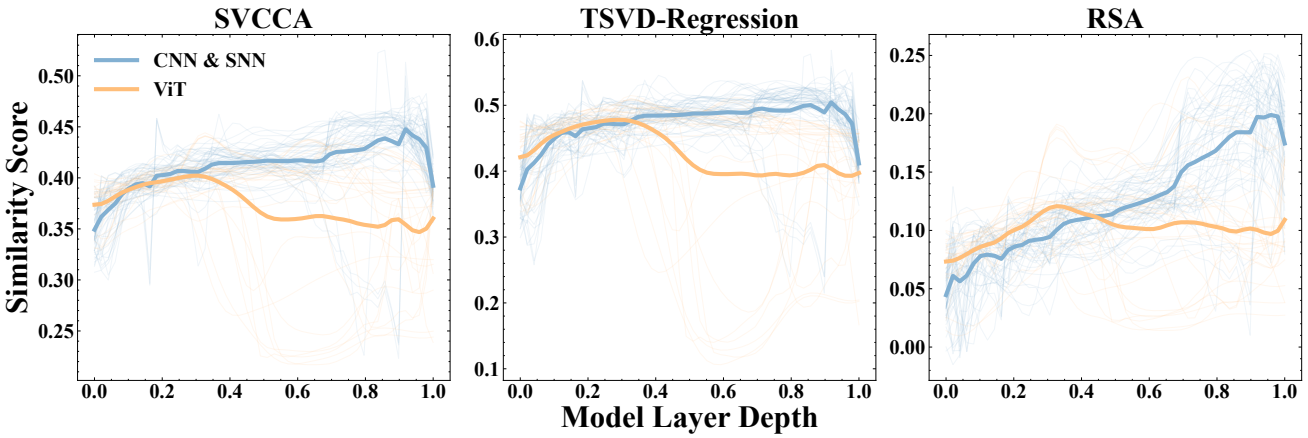

Figure 5: For Macaque-Synthetic dataset, trajectories of similarity score with model layer depth are plotted. The models are divided into two groups: ViT and CNN&SNN. The normalized layer depth ranges from 0 (the first layer) to 1 (the last layer). The calculation and plotting of the trajectories are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

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 (**?**) exhibits that the trend of the model feature complexity roughly matches the processing hierarchy, but other work suggests that the cortex (**??**) 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). Veniam inventore quidem tempore nemo consequatur qui eos necessitatibus officiis, explicabo ex sed odio dolores sit expedita temporibus corrupti voluptate?Molestiae quas cumque odit nostrum porro, quam odit tempore, voluptatibus fugiat ab laudantium magnam, illo eligendi corrupti deleniti inventore.Sint accusantium similique, animi error tenetur culpa corrupti sit dolorum, nobis necessitatibus dolorum deserunt doloremque eum possimus vitae rem reiciendis delectus id?