A Appendix

A.1 Non-normalized data corresponding to Figure 1

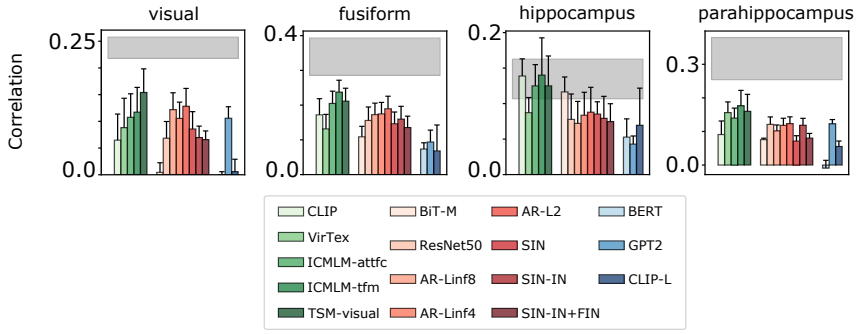

Figure 3: **Non-normalized RSA values between model and brain RDMs.** The brain RDMs are calculated based on selecting 30 voxels from each ROI, as in the main analysis. The gray bands show the upper and lower bounds of the noise-ceilings calculated by adding and subtracting it's s.e.m. respectively.

A.2 Voxel-selection based on a fixed beta-value threshold.

In the main analysis, we selected 30 voxels in each ROI based on the noise-ceiling analysis in the hippocampus. In other words, in each ROI we selected the 30 voxels with the highest beta values.

As a control method, instead of restricting the number of voxels to 30, we used the value of the 30th voxel from hippocampus as a threshold for other ROIs. The number of voxels found in each ROI for each participant is depicted in Table 1 and the RSA values in Figure 4. We observed that this alternate criterion did not affect the overall trend in other regions.

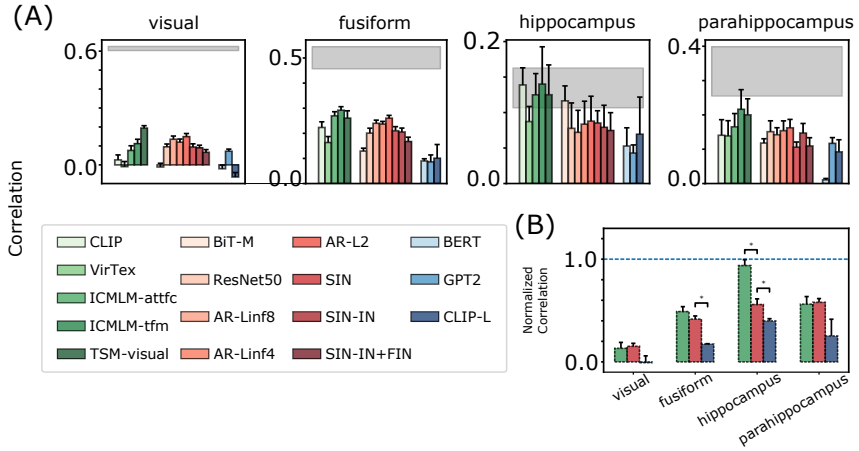

Figure 4: Non-normalized RSA values after using the beta value of the 30th voxel from hippocampus as a threshold for other ROIs for each participant.

Table 1: Number of voxels found in each region after thresholding

	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5
visual	530	831	343	707	592
fusiform	532	376	217	368	508
hippocampus	30	30	30	30	30
parahippocampus	122	67	85	111	167

A.3 RSA computed using different metrics

We verified the robustness of our results by using other metrics to compute the RDMs and RSA.

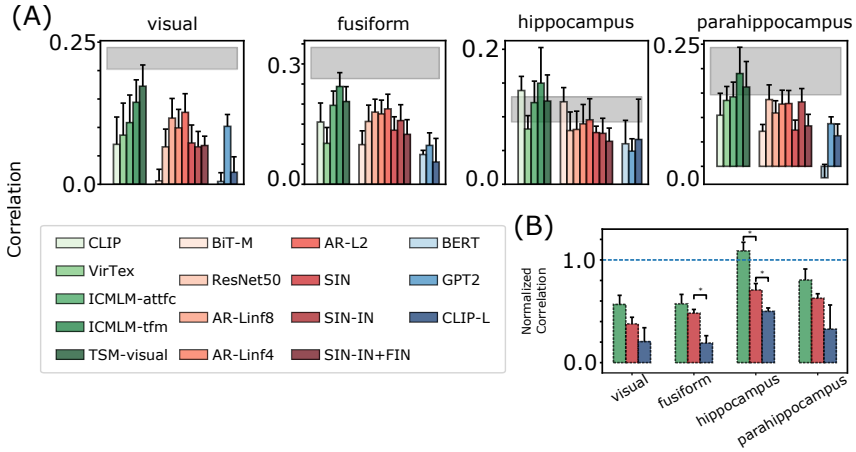

Figure 5: The RDMs were calculated using the Pearson correlation distance, and the Spearman rank correlation was used to compute the RSA.

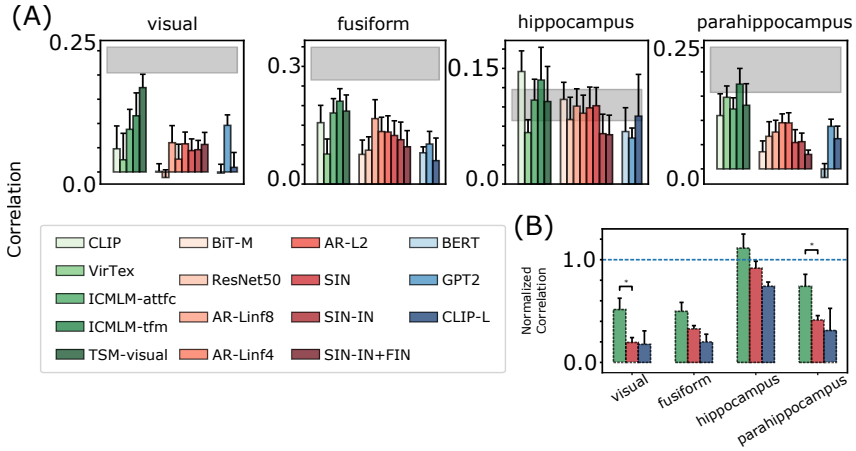

Figure 6: The RDMs were calculated using the Cosine distance, and the Spearman rank correlation was used to compute the RSA.

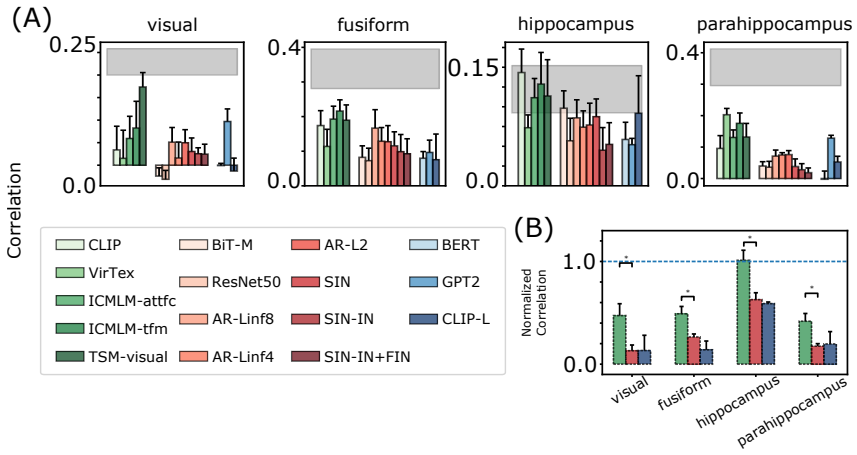

Figure 7: The RDMs were calculated using the Cosine distance, and the Pearson correlation was used to compute the RSA.

A.4 Licenses of the assets used

Asset	License
FreeSurfer	https://surfer.nmr.mgh.harvard.edu/fswiki/FreeSurferWiki
SPM12	GNU GPL
fMRI data	CC0
CLIP	MIT
VirTex	MIT
TSM	Apache-2.0
ICMLM	N/A
BiT-M	Apache-2.0
ResNet	MIT
AR models	MIT
SIN models	https://github.com/rgeirhos/texture-vs-shape/blob/master/DATASET_LICENSE
GPT2	MIT
BERT	Apache-2.0

Table 2: Available Licences of all the assets used in the study. Links to the appropriate webpages are provided for special licenses.

A.5 Broader Impacts

The research discussed above analyses the ability of neural networks to explain the human neural activity, specifically it demonstrates that multimodal neural networks are better than visual or linguistic models in explaining the activity in the hippocampus.

Importantly, this research provides potential insights for designing better bio-plausible networks. Upon diligent use, such networks can elucidate mechanisms in biological brains necessary to help patients. At the same time, we are aware of the possibilities for nefarious use of such systems, and urge all researchers to consider their implications.