PNAS
www.pnas.org

Supplementary Information for

Recurrence is required to capture the representational dynamics of the human visual system

Tim C Kietzmann, Courtney J Spoerer, Lynn K A Sörensen, Radoslaw M Cichy, Olaf Hauk, and Nikolaus Kriegeskorte

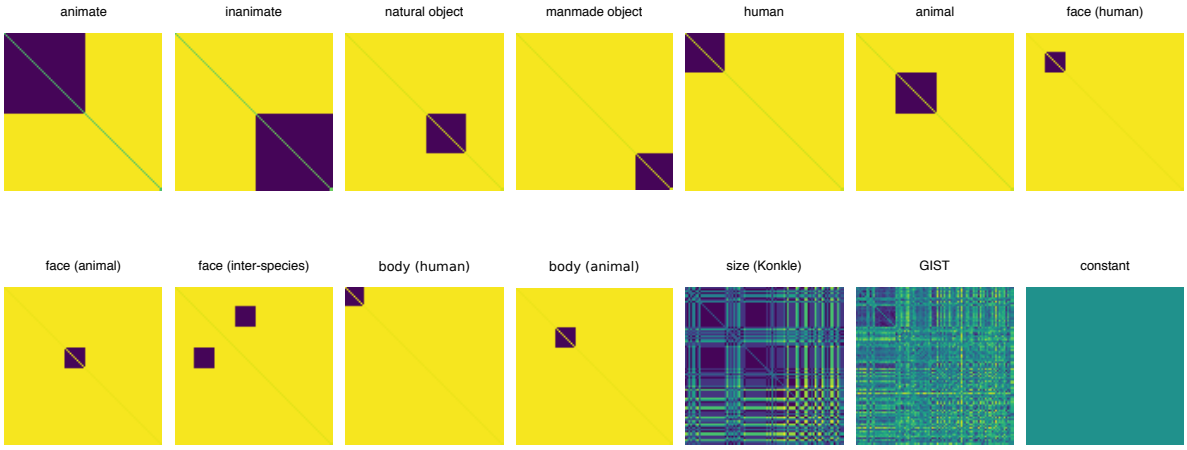
Corresponding author: Tim C Kietzmann
Email: tim.kietzmann@mrc-cbu.cam.ac.uk

This PDF file includes:

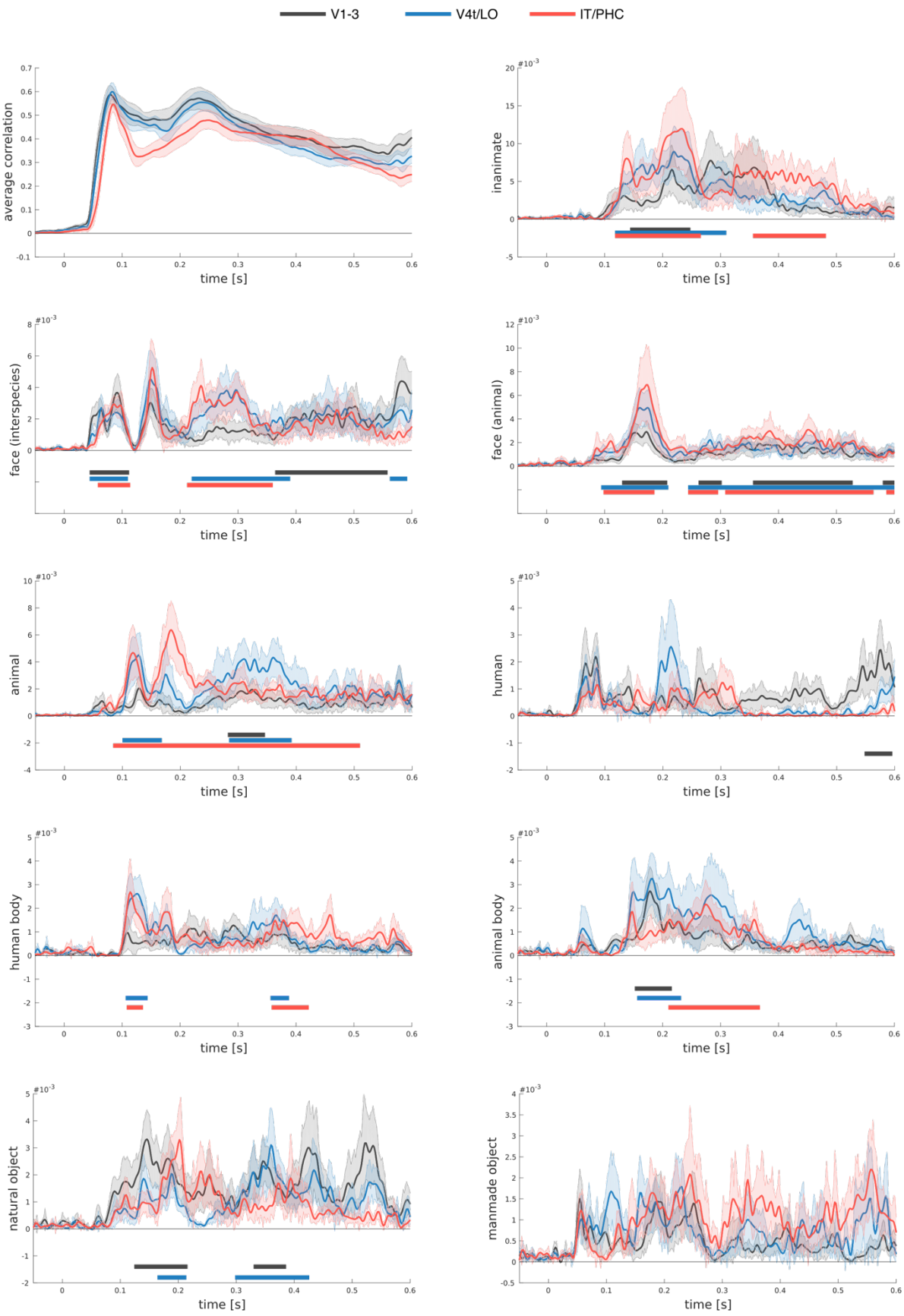
Figures	S1-S9
Tables	S1

Other supplementary materials for this manuscript include the following:

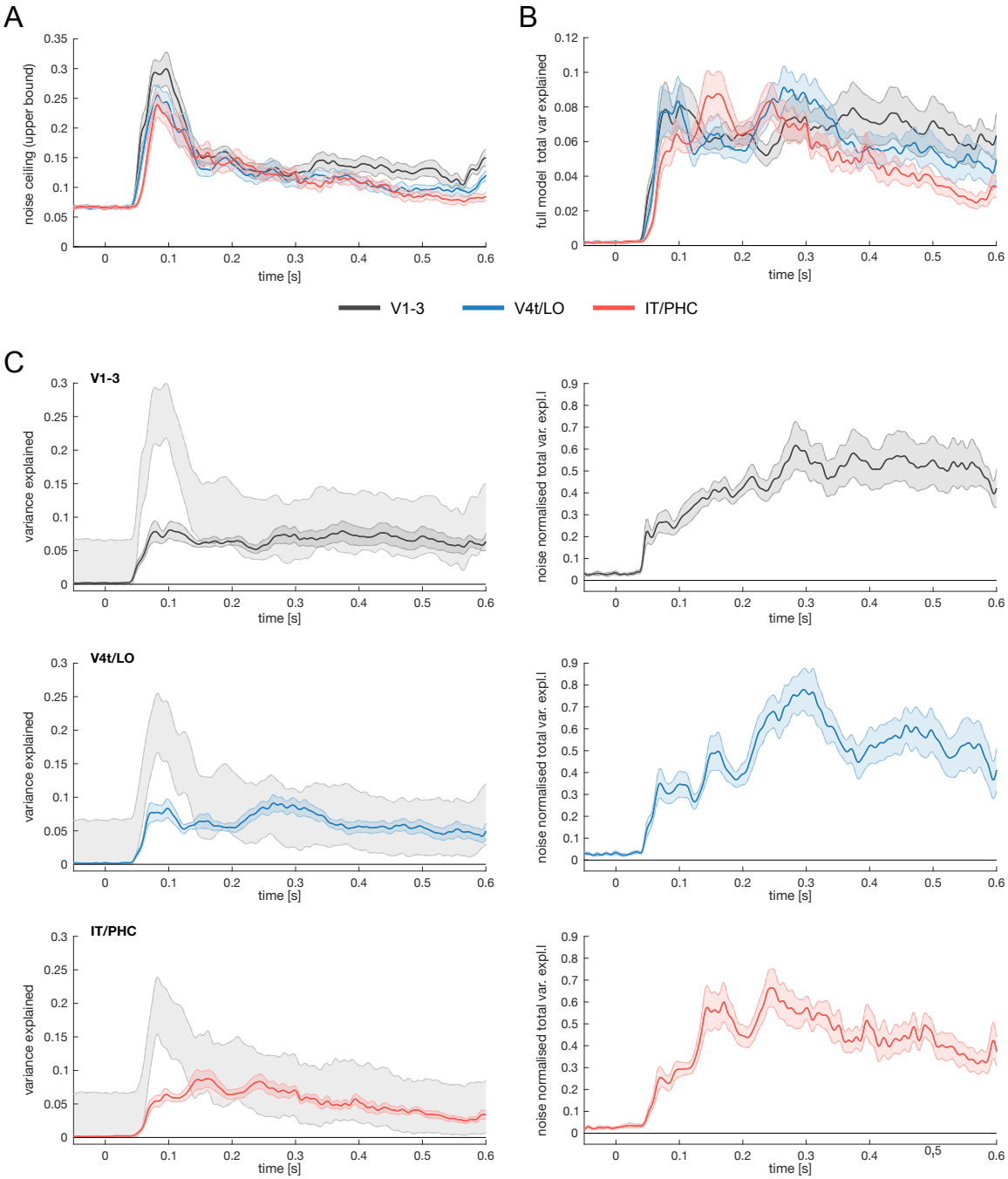
Movies S1-S6



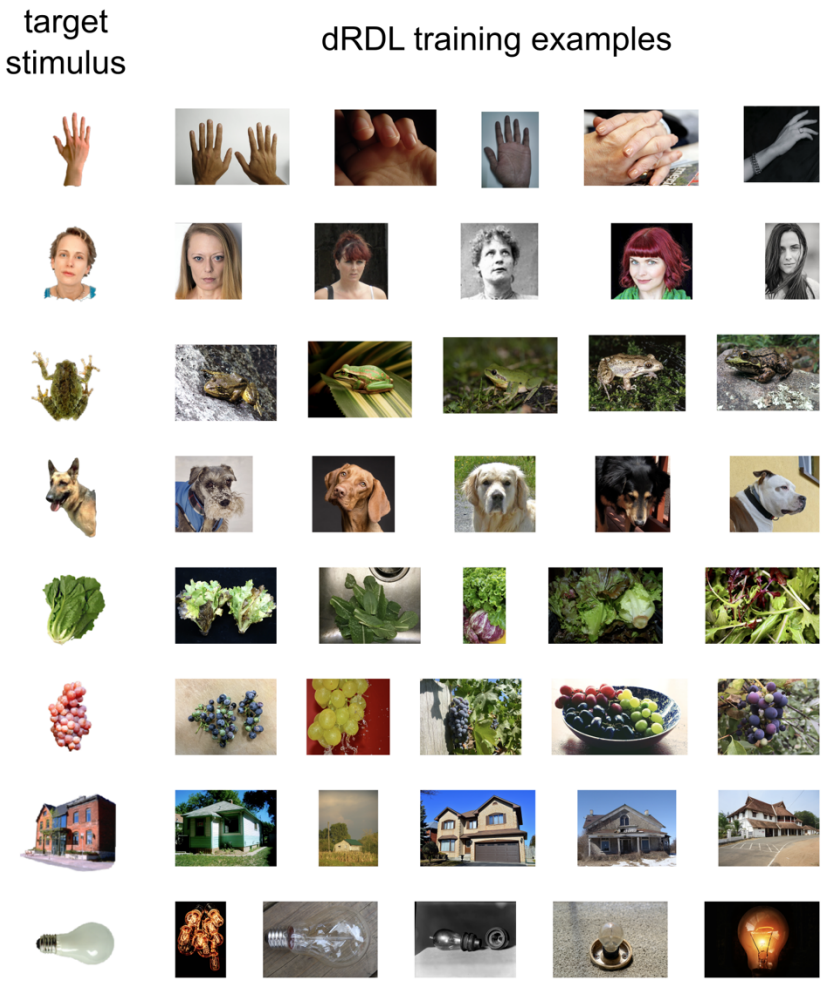
Supplemental Figure 1. RDM components used as predictors during hierarchical GLM modelling. The overall GLM included 11 categorical predictors, as well as a predictor derived from low-level Gist features, real-world size, and a constant.



Supplemental Figure 2. Supplemental results of the hierarchical GLM model fitting procedure. Top left: average pattern distance across time for the three ventral stream ROIs. All others: unique variance explained by all control model predictors.

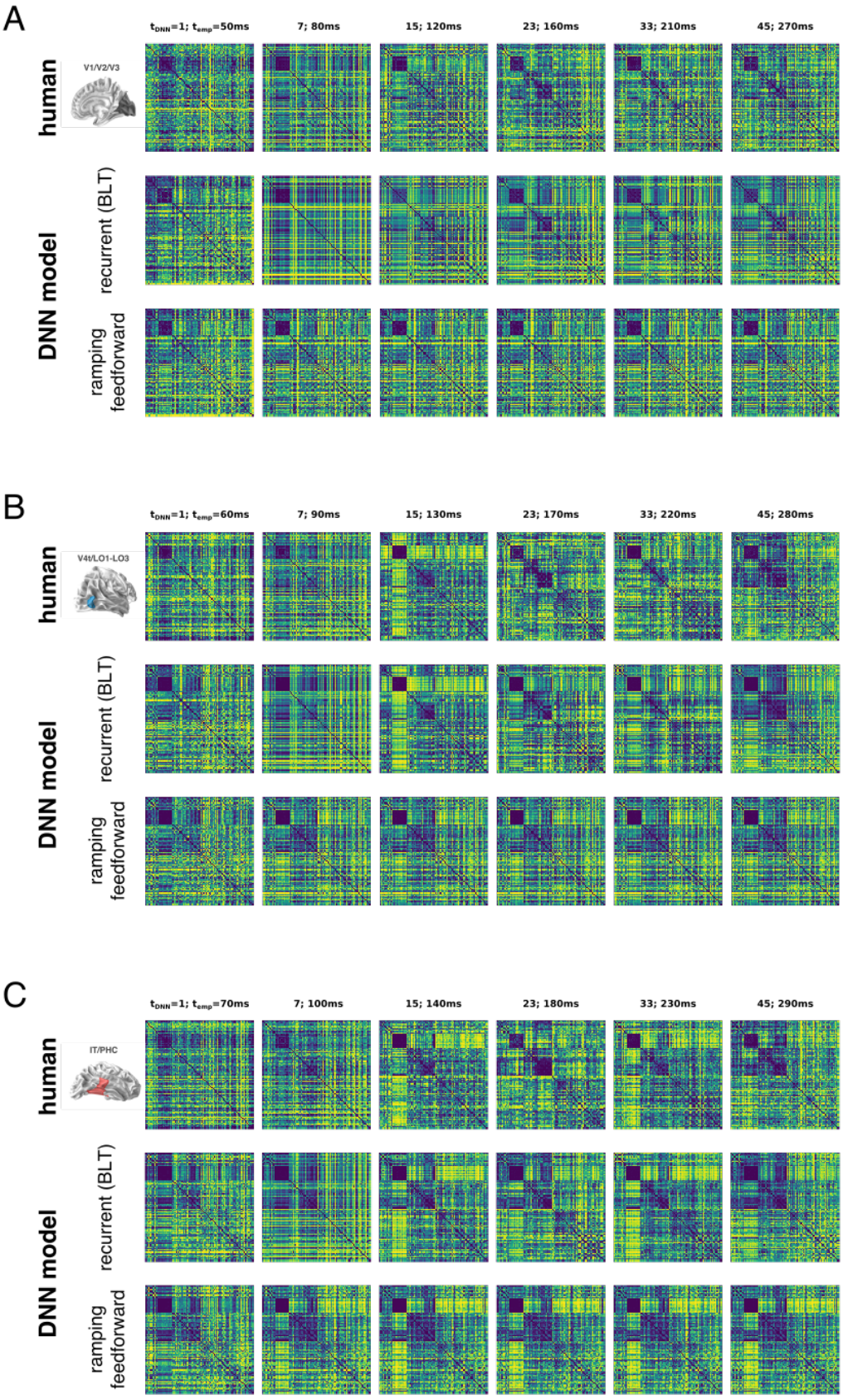


Supplemental Figure 3. Noise ceiling estimates and total variance explained. (A) Upper bound of the noise ceiling computed for each ventral stream ROI. (B) Total variance explained by the linear model used to dissect the ventral stream RDMs. (C) Upper and lower bound of the noise ceiling shown together with the total variance explained by the linear model (left column). The total variance explained was divided by the upper bound of the noise ceiling to express the percentage variance explained of the total explainable variance (right column).

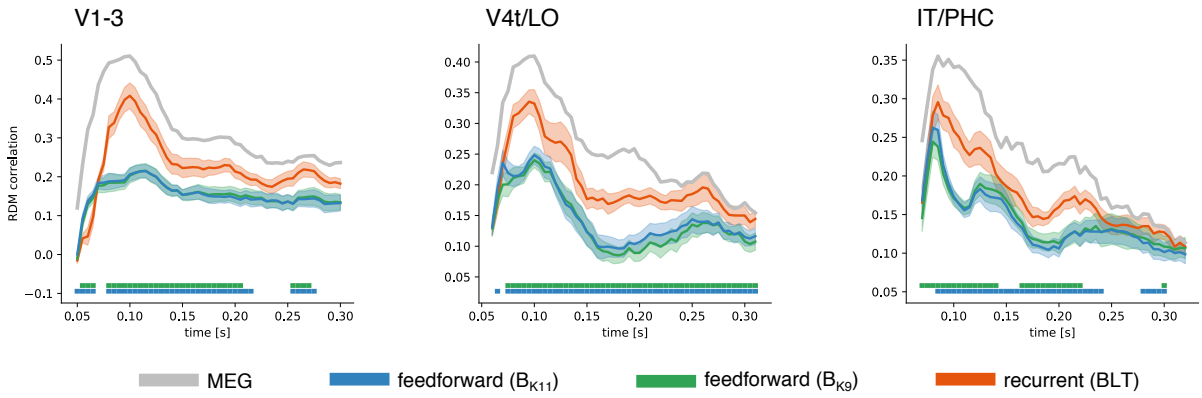


Supplemental Figure 4. Example images used for RDL training. The training set consisted of a total of 141k images, matching the categorical structure of the 92 experimental stimuli, while not including the original experimental stimuli. 61 categories were included for network training, directly mapping to 89 out of 92 experimental stimuli (3 stimuli were excluded from RDL training due to an insufficient number of images for training). Above images serve as illustrative examples from the public domain and are not necessarily contained in the training set.

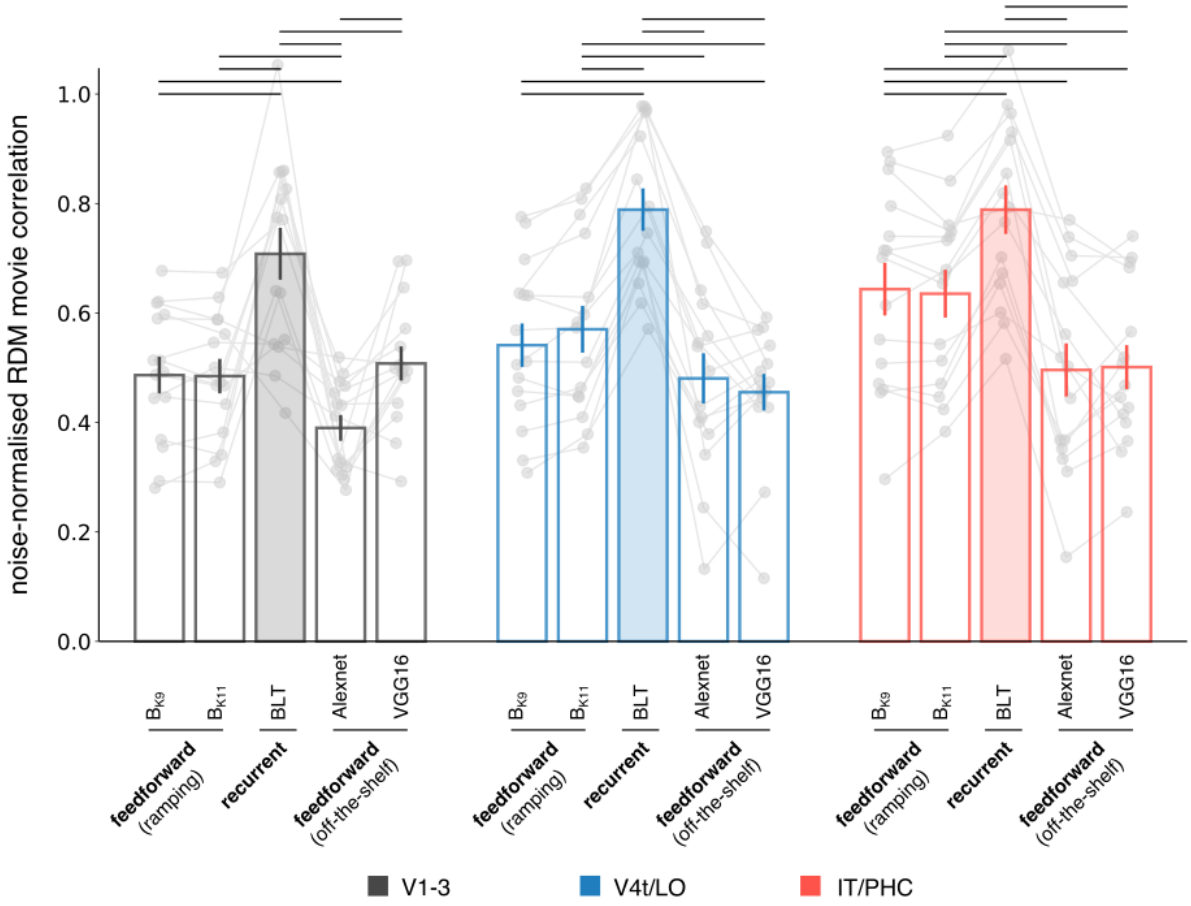
- Hand images courtesy of Flickr/Nate Steiner, anika, krihsna sir, Internet Archive Book Images, zhrefch, and apfelauge.
- Frog images courtesy of Flickr/Public.Resource.Org, USFWS Midwest Region, Bernard Spragg. NZ, Pacific Southwest Region USFWS, and bastiendaubie.
- Face images courtesy of Flickr/Mon Mer, kevin ryder, Bibliothèque de Toulouse, and Noval Goya.
- Lettuce images courtesy of Flickr/Angelica Perduta, Terri Bateman, Shawn Rosvold, Scot Nelson, and Alan Levine.
- Grape images courtesy of Flickr/Mike Linksvayer, Picdrome Public Domain Pictures, Alan Levine, and Suzy Hazelwood.
- Dog images courtesy of Flickr/Soeradjoen, Magda K, Frank-3, Alan Levine, and Radosław Magiera.
- House images courtesy of Flickr/Kim Siever, Hillbraith, ezhuth pura, Maxim Gangalo, and Open Grid Scheduler / Grid Engine.
- Lightbulb images courtesy of Flickr/Darren Cowley, Marcu Ioachim, George Eastman Museum, Alexjo .png, and Josbert Lonnee.



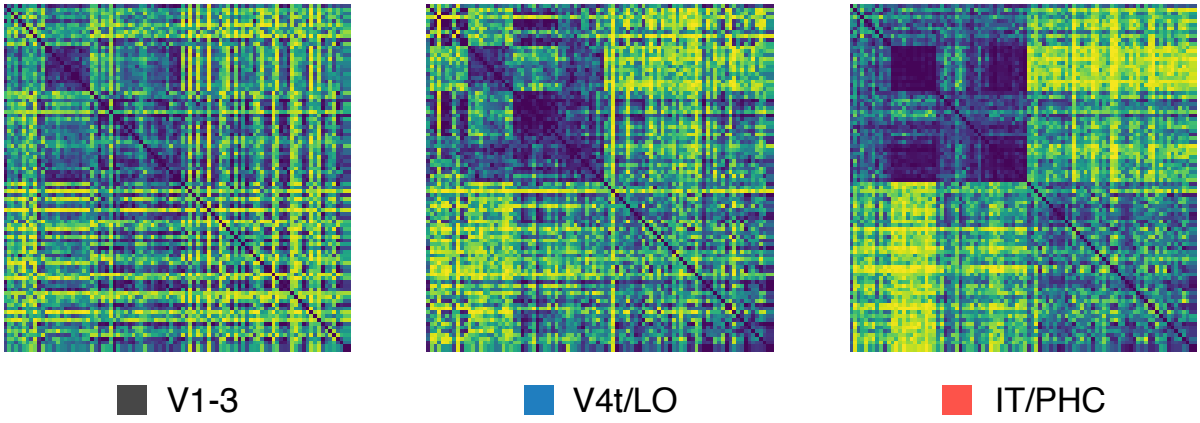
Supplemental Figure 5. RDM movie frames for the representational trajectories of the human ventral stream ROIs and model architectures (ramping feedforward and recurrent (BLT) DNNs). A-C show V1-3, V4t/LO, and IT/PHC, respectively. Numbers on top relate to the respective network and empirical brain times (t_{DNN} and t_{emp}).



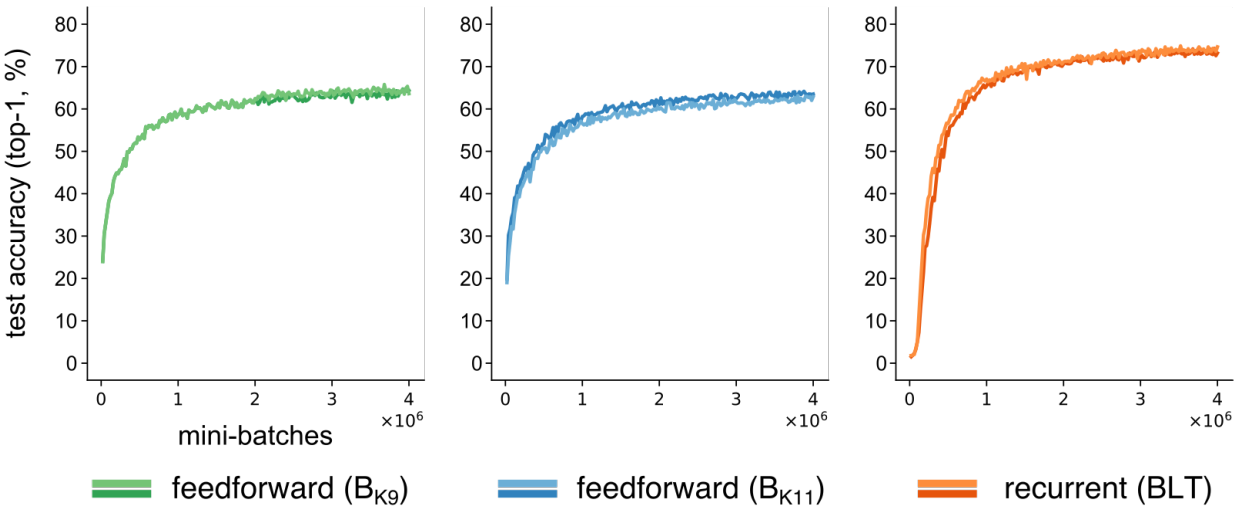
Supplemental Figure 6. Frame by frame traces of RDM correlations between DNNs and brain data. Split-half correlation for the MEG data, as used for training, shown in grey.



Supplemental Figure 7. Comparison of our CNN models (ramping feedforward and recurrent) with off-the-shelf computer vision models Alexnet and VGG16. Average frame-by-frame RDM correlation between deep network models and brain. Alexnet and VGG16 layers were selected via cross-validation. Contrary to the recurrent and ramping feedforward models, their RDM predictions are not time-varying, as classic feedforward models do not exhibit within-layer dynamics.



Supplemental Figure 8. fMRI analyses. RDMs extracted from fMRI data obtained from the same participants and the same three ventral stream ROIs.



Supplemental Figure 9. Image classification test performance across training for the different model types. Ramping feed-forward (B_{K9}, B_{K11}; shown in green and blue, respectively) and recurrent (BLT; shown in orange) models.

Table 1 Network architectural parameters

Layer	Feature maps	Image size	B_{K9} – kernel size (effective)	B_{K11} – kernel size (effective)	BLT – kernel size (effective)
1	64	96 × 96	9 (9)	11 (11)	5 (5)
2	64	48 × 48	9 (9)	11 (11)	5 (5)
3	96	24 × 24	9 (9)	11 (11)	5 (5)
4	96	12 × 12	9 (9)	11 (11)	5 (5)
5	128	6 × 6	9 (9)	11 (11)	5 (5)
6	128	3 × 3	9 (5)	11 (5)	5 (5)

Supplemental Table 1. Details on the network architectures.

Supplemental Movies

1. **MEG RDM movies for three ventral stream ROIs.**
2. **Recurrent DNN RDM movies showing the reconstruction of the representational dynamics observed across three ventral stream ROIs.**
3. **Feedforward DNN RDM movies showing the reconstruction of the representational dynamics observed across three ventral stream ROIs.**
4. **MEG, feedforward and recurrent network RDM movies for area V1-3.**
5. **MEG, feedforward and recurrent network RDM movies for area V4t-LO**
6. **MEG, feedforward and recurrent network RDM movies for area IT/PHC**