Facial representation comparisons between human brain and deep convolutional neural network reveal a fatigue repetition suppression mechanism

Zitong Lu[1], & Yixuan Ku[2,3]

[1] Department of Psychology, The Ohio State University, Columbus, OH, USA
[2] Center for Brain and Mental Well-being, Department of Psychology, Sun Yat-sen University, Guangzhou, China
[3] Peng Cheng Laboratory, Shenzhen, China

*Correspondence should be addressed to: kuyixuan@mail.sysu.edu.cn

Abstract

Repetition suppression for faces, a phenomenon that neural responses are reduced to repeated faces in the visual cortex, have long been studied. However, the underlying primary neural mechanism of repetition suppression remains debated. In recent years, artificial neural networks can achieve the performance of face recognition at human level. In our current study, we combined human electroencephalogram (EEG) and the deep convolutional neural network (DCNN) and applied reverse engineering to provide a novel way to investigate the neural mechanisms of facial repetition suppression. First, we used brain decoding approach to explore the representations of faces and demonstrates its repetition suppression effect in human brains. Then we constructed two repetition suppression models, Fatigue and Sharpening models, to modify the activation of DCNNs and conducted cross-modal representational similarity analysis (RSA) comparisons between human EEG signals and activations in two modified DCNNs, respectively. We found that representations of human brains were more similar to representations of Fatigue-modified DCNN instead of Sharpening modified DCNN. Our results suggests that the facial repetition suppression effect in face perception is more likely caused by the fatigue mechanism suggesting that the activation of neurons with stronger responses to face stimulus would be attenuated more. Therefore, the current study supports the fatigue mechanism as a more plausible neural mechanism of facial repetition suppression. The comparison between representations in the human brain and DCNN provides a promising tool to simulate and infer the brain mechanism underlying human behaviors.

Key words: repetition suppression, reverse engineering, EEG, DCNN, RSA

Introduction

In our daily life, in addition to perceiving new things, we often see the same type of stimulus, or even exactly the same stimulus again. In this process of repeatedly receiving the same information input, neural activity in our brains will always be weaker than the first time you saw the input which is called repetition suppression. Many electrophysiological studies have observed that the response of neurons sensitive to visual information in interior temporal cortex would decreased under the condition of repeated stimuli (Baylis & Rolls, 1987; Kaliukhovich & Vogels, 2011, 2012; Miller et al., 1991; Ringo, 1996; Sawamura et al., 2006; Sobotka & Ringo, 1994). Also, researchers found that repeated stimulus could reduce the blood oxygenation level-dependent (BOLD) in functional magnetic resonance imaging (fMRI) (Henson & Rugg, 2003). In face perception, many electroencephalogram (EEG) and Magnetoencephalography (MEG) studies have found various event-related potential (ERP) components were related to facial repetition suppression, like N170 (Kloth et al., 2010; Kloth & Schweinberger, 2010; Kovács et al., 2006; Maurer et al., 2008; Mercure et al., 2011; Schweinberger et al., 2007; Walther et al., 2013), P200 (Burkhardt et al., 2010; Kaufmann & Schweinberger, 2012; Latinus & Taylor, 2006; Schulz et al., 2012; Zheng et al., 2012), N250r (Dörr et al., 2011; Herzmann et al., 2004; Pfütze et al., 2002; Schweinberger et al., 1995; Schweinberger & Burton, 2003; Wiese et al., 2013), and N400 (Barrett & Rugg, 1989; Bentin et al., 1985; Rugg, 1985; Schweinberger, 1996; Stevenage et al., 2014). However, the underlying neuronal mechanism of repetition suppression remains debated.

Previous studies related to facial repetition suppression focused on the use of univariate analysis, ignoring the dynamics and differences in patters of neural representations. In the past decade, more and more researchers in cognitive neuroscience have used multivariate analysis methods, such correlation or classification-based MVPA (Cox & Savoy, 2003; Golomb & Kanwisher, 2012; Haxby et al., 2001; Kamitani & Tong, 2005; Norman et al., 2006) and representational similarity analysis (RSA) (Kriegeskorte, Mur, & Bandettini, 2008; Kriegeskorte, Mur, Ruff, et al., 2008), to explore the neural mechanism of complex cognitive processes more deeply. These multivariate analysis methods can help us understand the representational patterns of how our brains encode information. In addition, RSA can be applied to conduct representational comparisons between different modes, like brain activity and activations in models (Cichy et al., 2016; Dobs et al., 2019; Güçlü & van Gerven, 2015; Kuzovkin et al., 2018; Urgen et al., 2019; Xie et al., 2020; Xu & Vaziri-Pashkam, 2021; Yamins et al., 2014).

From the perspective of neural coding pattern, Grill-Spector et al., 2006 proposed three possible repetition suppression models (Fatigue, Sharpening, and Facilitation models) based on previous studies on single cell recordings, fMRI and EEG/MEG. For Fatigue model, they assumed that the number of repetition suppression neurons in

the neurons that had stronger response to the stimulus would be more than that in other neurons. For Sharpening model, they assumed that neurons encoding features unrelated to the stimulus would show repetition suppression, resulting in sparse representation of stimulus encoding. For Facilitation model, they assumed that repetition would lead to faster stimulus processing, that was, shorter waiting time or duration. However, can multivariable analysis be applied to explore which is the more possible neuronal mechanism of facial repetition suppression in human brains?

In computer vision domain, some face recognition models based on deep convolutional neural network (DCNN) structure (Parkhi et al., 2015; Ranjan et al., 2019; Schroff et al., 2015; Taigman et al., 2014) have reached the human level (Phillips et al., 2018). Increasingly cognitive neuroscientists and computer scientists have begun to explore the similarities and differences between human brains and artificial intelligence (AI) models in information processing. Especially in the field of vision, research combining brain activity and DCNNs has found that the hierarchical structure of the ventral visual pathway and DCNNs had similar processing representations of visual information (Cichy et al., 2016; Güçlü & van Gerven, 2015; Kietzmann et al., 2019; Yamins et al., 2014). Reverse engineering provides a possible solution to explore the brain mechanism by combining AI models and human brains. We could modify the representation of the model based on different theoretical hypotheses to find the mechanism closest to the representation of human brains. Thus, can we use the idea of reverse engineering to investigate the neural mechanism of repetition suppression?

Our current study aimed to understand the neural mechanism of facial repetition suppression using novel computational methods. We first explore dynamic representations of face information and confirmed facial repetition suppression effect in human brains using classification-based MVPA method based on human EEG. Then we applied reverse engineering to construct two possible repetition suppression models, Fatigue and Sharpening models, to modify the activation of DCNN and conducted cross-modal RSA between human brain activity and activations in modified DCNNs. Our results suggested that fatigue mechanism was a more plausible neural mechanism of facial repetition suppression.

Material and methods

Data and experimental information

The data used in this study were based on an EEG dataset of face perception from OpenNeuro (https://openneuro.org/datasets/ds002718/), which belonged to a multi-subject, multi-modal human neuroimaging dataset (Wakeman & Henson, 2015). In this study, we only used EEG data with 70 valid channels of 18 subjects in that dataset for analysis. All subjects conducted a face perception task (Figure 1A). 450 grey face

stimuli consisting of 150 familiar faces, 150 unfamiliar faces and 150 scrambled faces were used in the task. Each face stimulus was presented twice. Subject would see the same face again immediately or after 5 to 15 trials with 50% probability, respectively. Refer to Wakeman & Henson, 2015 for more details about experimental design.

Here, we labeled first time the subject saw a certain face 'new', second time the subject saw that face immediately 'immediate (repetition)', and second time the subject saw that face after some trials 'delayed (repetition)'. Thus, there were, 3 face conditions (F: familiar, U: unfamiliar, S: scrambled) by 3 presentation conditions (N: new, I: immediate, D: delayed), 9 conditions (FN, FI, FD, UN, UI, UD, SN, SI, SD) in total (Figure 1B).

EEG data has gone through partly preprocessing and been resampled to 250 Hz in the dataset. We band-filtered the data between 0.1 and 30 Hz. Independent components analysis (ICA) was applied to identify and remove the blinks and eye movements (Drisdelle et al., 2017; Jung et al., 2000) using EEGLAB (Delorme & Makeig, 2004). Epochs of -500 to 1500ms from stimulus onset were created. No baseline correction was applied.

Classification-based EEG decoding

***Time-by-time EEG decoding*:** Each trial of EEG data corresponded to two kinds of labels: one face condition label (F, U, or S) and one presentation conditions (N or I or D). 9 separate classification-based decoding analyses were conducted to distinguish neural representations between different presentation conditions: FN vs. FI, UN vs. UI, SN vs. SI, FN vs. FD, UN vs. UD, SN vs. SD, FI vs. FD, UI vs. UD, SI vs. SD.

We applied linear-SVM to do the classification-based decoding between each pair of two conditions for each time-point. First, we binarized the trial labels of the two categories corresponding to two conditions. We downsampled EEG data by averaging each 5 time-points. The original 500 time-points from -500 to 1500ms were compressed into 100 time-points. Therefore, each subject corresponded a label vector, including labels of all trials, and a three-dimension matrix, including time, trial, and channels dimensions, for time-by-time classification. Next, we shuffled the order of trials, and the data was averaged every five trials. The classifier was trained and tested at each time-point separately. Here, we trained on 2/3 trials randomly selected and tested on remaining 1/3 trials to evaluate the performance for each iteration. The entire process of random shuffling, averaging, classification training and testing were iterated 100 times. These above steps were conducted for each time. Then we averaged classification accuracies across all iterations to obtain more reliable time-by-time decoding accuracies. We repeated the time-by-time decoding for 18 subjects and 9 classification condition pairs.

***Cross-temporal EEG decoding*:** Further, we conducted cross-temporal EEG decoding to build time generalization matrices, which was an extension of time-by-time decoding. The essence of cross-temporal decoding was to train the classifier on the data at a time-point and test it on the data at other time-points to determine whether

the encoding mode of information we were interested in at different times was consistent. Therefore, based on the above implementation of time-by-time decoding, we tested each classifier on all 100 time-point data (including the time-point used for training) (Figure 1C). Similarly, we got the final decoding accuracies across 100 iterations. Thus, each subject would get 9 temporal generalization matrices of cross-temporal decoding corresponding to 9 classification condition pairs. All the above EEG decoding processes were implemented based on NeuroRA toolbox (Lu & Ku, 2020).

DCNN models

In this study, we selected a DCNN model used in the field of face recognition, VGG-Face (Parkhi et al., 2015) (Supplementary Figure 1A). This model was pretrained on a dataset of 2622 identities with 1000 face images per person. Its test accuracies on IFW dataset and YouTube Faces dataset could reach 97.27% and 92.8% respectively. VGG-Face was essentially a VGG-16 model structure, which included 13 convolutional layers and 3 full connection layers. Here, VGG-Face model was used as a DCNN model for acquiring face features. As a contrast, we also used an additional VGG-16 model without training. This untrained VGG model with random weights was taken as a DCNN model of learning no face features.

RSA

***EEG RDMs*:** Since there were 3 face conditions by 3 presentation conditions, we used EEG classification-based decoding accuracy as the dissimilarity index to construct 9× 9 neural RDMs. In addition, we constructed cross-temporal RDMs (CTRDMs) instead of traditional RDMs to conduct cross-temporal RSA (CTRSA) (Lu, 2020). Compared with traditional RSA, we applied cross-temporal decoding accuracy into CTRSA. As shown in Figure 1C, a SVM classifier was trained on EEG data of FD vs. SD at t_A and tested on data of FD vs. SD at t_B. This test accuracy was taken as the dissimilarity index between t_A FD and t_B SD conditions in t_A -> t_B CTRDM of this subject. Due to the direction from training time to test time, the CTRDM of t_A -> t_B was different from the CTRDM of t_B -> t_A. The former was obtained by using the data of t_A to train and t_B to test, while the latter was obtained by using the data of t_B to train and t_A to test. Since we couldn't do classification between two same conditions, the diagonal values in decoding based CTRDMs were all 0. Therefore, we constructed a CTRDM for any pair of directed two time-points (from one time-point to another time-point) according to the above method. We could get 100 (time-points) × 100 (time-points) CTRDMs based on cross-temporal EEG decoding. All the EEG RDMs part was implemented based on NeuroRA toolbox (Lu & Ku, 2020).

***DCNN RDMs*:** Since there were a large number of nodes in each layer of the VGG-16

model. We first used principal component analysis (PCA) to reduce the feature dimension. For example, the second layer contained 64 112×112 feature maps corresponding to 802,816 nodes. For each image input into the VGG-16 model. The activation of layer 2 could be output as a 1×802,816 vector. We calculated principal components which were arranged according to their contribution rates from the largest to the smallest using PCA. Principal components with total contribution rates greater than 95% were selected as valid features after dimension reduction. As shown in Figure Supplementary Figure 1B, the number of feature dimension of layer 2 after PCA was reduced to 307. In this way, we reduced the feature dimension of each layer in both VGG-face and untrained VGG models. This dimension reduction part was implemented based on Scikit-learn toolkit (Pedregosa et al., 2011).

We input each image into VGG-Face or untrained VGG and obtained the activation vector of each layer of feature after dimension reduction. 450 images corresponded to 450 activation vectors of each layer. Due to the huge amount of calculation, all calculations were based on even layers (layer 2, 4, …, and layer 16). For each layer, we calculated the Pearson correlation coefficient r between the activation vectors corresponding to any two images. And 1-r was used as the dissimilarity index. A 450×450 RDM for a certain layer was constructed in the order of 150 familiar faces, 150 unfamiliar faces, and 150 scrambled faces. We obtained DCNN RDMs of all even layers in VGG-Face and untrained VGG models. The DCNN RDMs calculation part was implemented based on NeuroRA toolbox (Lu & Ku, 2020).

***Repetition suppression simulations in DCNN*:** In order to use 'reverse engineering' to investigate the neural mechanism of facial repetition suppression, we proposed two possible neuronal level models, Fatigue model and Sharpening model, to simulate the neural presentation of the human brain under facial repetition suppression on DCNNs. We set the activation vector of a face image p at layer i of a DCNN (before PCA) as $A = (a_1, a_2, \dots\dots, a_n)$, where the activation values were ordered in descending order ($a_1 > a_2 > a_3 > \dots > a_{m-1} > a_m = a_{m+1} = \cdots = a_n = 0$), which meant that there were *n-m* nodes with nonzero activation value and m nodes with activation value of zero).

For Fatigue model, we assumed that the activation of the node with higher response to face stimulus was weaken under repetition suppression condition. The activation of the node with low response remained unchanged, but the node with higher activation had more attenuations. Thus, if the face image p was viewed repeatedly, the new activation vector obtained based on Fatigue model would be $A_F = (a_{F1}, a_{F2}, \dots\dots, a_{Fn})$, and its internal activations were:

$$a_{Fi} = \begin{cases} \left(1 - \alpha + \dfrac{(i-1)\alpha}{\beta n}\right) \cdot a_i & (1 \le i \le \beta n) \\ a_i & (\beta n < i \le n) \end{cases}$$

where α was the maximum fatigue coefficient, and β was the proportion of the nodes

that would be attenuated. Thus, the first βn nodes would be attenuated when the same face image was viewed repeatedly, and the nodes from the strongest one to the βnth one would be weakened in proportion from α to $\alpha/(\beta n)$.

For Sharpening model, we assumed that the nodes with lower response to face stimulus were no longer activated under repetition suppression condition, and the nodes with higher response kept same activations. Thus, the new activation vector obtained based on Sharpening model would be $A_S = (a_{S1}, a_{S2}, \dots\dots, a_{Sn})$, and its internal activations were:

$$a_{Si} = \begin{cases} a_i & (1 \le i < (1-\theta)n) \\ 0 & ((1-\theta)n \le i \le n) \end{cases}$$

where θ was the proportion of the nodes which would be not activated. Thus, the last θn nodes' activations would become zero when the same face image was viewed repeatedly.

Figure 1E is a schematic diagram of how these two repetition suppression models simulated in DCNNs. Here, Figure 1E shows the original activations of 30 active nodes and the activations under repetition suppression condition based on Fatigue model (α=0.5, β=0.5) and Sharpening model (θ=0.5) respectively. For all 450 images, we input them into both VGG-Face and untrained VGG respectively, and then we calculated activation vectors of different layers based on different simulation models of repetition suppression.

***Modified DCNN RDMs*:** Based on the above two facial repetition suppression models, we set model parameters corresponding to three types of face stimuli here: For Fatigue model, α was set to 0.9, which meant that the first 90% nodes with nonzero activation value were set as the nodes with high activation. And we set the maximum fatigue coefficient β to 0.5 for immediate repetition condition and 0.05 for delayed repetition condition. For Sharpening model, θ was set to 0.5 for immediate repetition condition and 0.05 for delayed repetition condition. Therefore, three activation vectors corresponding to new, immediate, and delayed conditions were calculated from the activation in each layer in DCNNs of each image based on each repetition suppression model. Then, we input these vectors into PCA to get feature vectors after dimension reduction. For each even layer, face image and repetition suppression model, we calculated the dissimilarity (1-Pearson correlation coefficient) between each pair of two feature vectors and got two 1350×1350 RDMs for VGG-Face and untrained VGG.

***RSA between EEG and DCNNs*:** For VGG-Face and untrained VGG, 8 1350×1350 RDMs corresponding to 8 even layers were obtained. For EEG, each subject corresponded 100 time-by-time 9×9 RDMs and 100×100 cross-temporal 9×9 CTRDMs. To establish the connection between DCNNs and human brains, we averaged the cells under 9 conditions respectively (FN, FI, FD, UN, UI, UD, SN, SI, SD) in 1350×1350 RDMs to get compressed 9×9 DCNN RDMs. Then, we calculated the partial Spearman correlation coefficient as the similarity between neural (CT)RDMs and DCNN RDMs for each time-point. To obtain the similarity between the representation of face information that a DCNN learned for face recognition and

human brains, we calculated the valid representational similarity S_{valid} below:

$$S_{valid} = S_{VGG-Face} - S_{Untrained\ VGG}$$

where $S_{VGG-Face}$ was the representational similarity between VGG-Face and neural activity, and $S_{Untrained\ VGG}$ was the representational similarity between untrained VGG and neural activity. Here, we applied untrained VGG's representations as a baseline that there was not enough face-specific information and calculated valid similarities for two repetition suppression models, respectively. To get face-specific repetition suppression mechanism, the pre-train VGG-Face trained on face recognition task would have stronger similarity with human brains than untrained VGG.

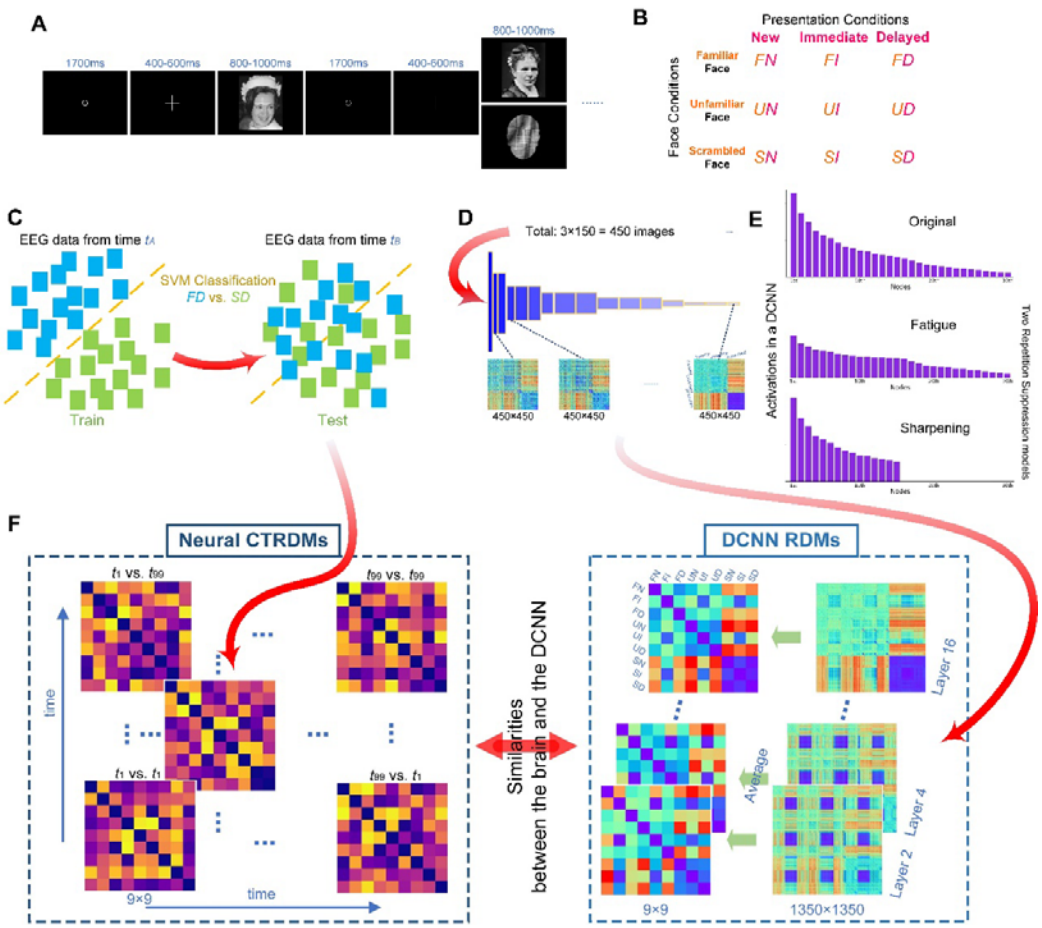


Figure 1. Experimental procedure and key analysis flow chart. (A) Experimental sequences. (B) 9 experimental conditions (3 face conditions by 3 presentation conditions). (C) Schematic diagram of cross-temporal EEG decoding. (D) Schematic diagram of calculating DCNN RDMs. (E) Schematic diagram of Fatigue and Sharpening repetition suppression models. (F) Procedure of cross-modal RSA comparisons between EEG and DCNNs. Two face images shown are in public domain and available at https://commons.wikimedia.org for illustrative purposes only. The actual images used during the experiment were described in Wakeman & Henson, 2015.

Statistical analysis

For classification-based decoding results, if neural representations of brains did encode the information at a certain time-point, we assumed that neural patterns between two conditions would be linearly classified. Then decoding accuracy would be greater than chance, 50%. For RSA results, we tested whether valid similarity was significantly greater than zero. If valid similarity was zeros or less than zero, it suggested that the corresponding repetition suppression mechanism was not specific to face information. To compare decoding accuracy to chance and valid similarity to 0 at each time-point while controlling for multiple comparisons, we used cluster-based permutation tests to assess statistical significance. First, we calculated t-values for each time-point and extracted significant clusters. And we calculated the clustering statistic as the sum of t-values in each cluster. Then we conducted 5000 permutations to calculate the maximum permutation cluster statistic. Finally, we assigned p-values to each cluster of the actual decoding accuracies or similarities by comparing its cluster statistic with the permutation distribution.

Code availability

All analysis scripts are available online at osf.io/unhzm.

Results

Facial repetition suppression in human brains

Classification-based EEG decoding results of three presentation conditions are shown in Figure 2. Figure 2A shows time-by-time decoding results. For New vs. Immediate decoding, decoding accuracies of both familiar and unfamiliar faces were significantly greater than chance from 200ms to 1500ms, and decoding accuracies of scrambled faces were significantly greater than chance from 360ms to 1500ms. Decoding accuracies of familiar faces were significantly greater than decoding accuracies of unfamiliar faces from 620ms to 800ms, 880ms to 1100ms, and 1120ms to 1220ms. Decoding accuracies of familiar faces were significantly greater than decoding accuracies of scrambled faces from 240ms to 1500ms. Decoding accuracies of unfamiliar faces were significantly greater than decoding accuracies of scrambled faces from 240ms to 960ms.

For New vs. Delayed decoding, we only found significant decoding accuracies of familiar faces from 660ms to 780ms, 860ms to 920ms, and 1100ms to 1160ms. Decoding accuracies of familiar faces were significantly greater than decoding accuracies of scrambled faces from 580ms to 800ms.

For Immediate vs. Delayed decoding, decoding accuracies of familiar faces were significantly greater than chance from 280ms to 1180ms. Decoding accuracies of unfamiliar faces were significantly greater than chance from 240ms to 980ms. Decoding accuracies of scrambled faces were significantly greater than chance from 560ms to 900ms. Decoding accuracies of familiar faces were significantly greater than decoding accuracies of scrambled faces from 440ms to 840ms, and decoding accuracies of unfamiliar faces were significantly greater than decoding accuracies of scrambled faces from 440ms to 780ms.

For cross-temporal results (Figure 2B), differences of neural patterns between New and Immediate conditions were strongest. There was nearly no significant difference between New and Delayed conditions. Also, the differences of neural patterns of familiar faces were stronger than unfamiliar faces than scrambled faces.

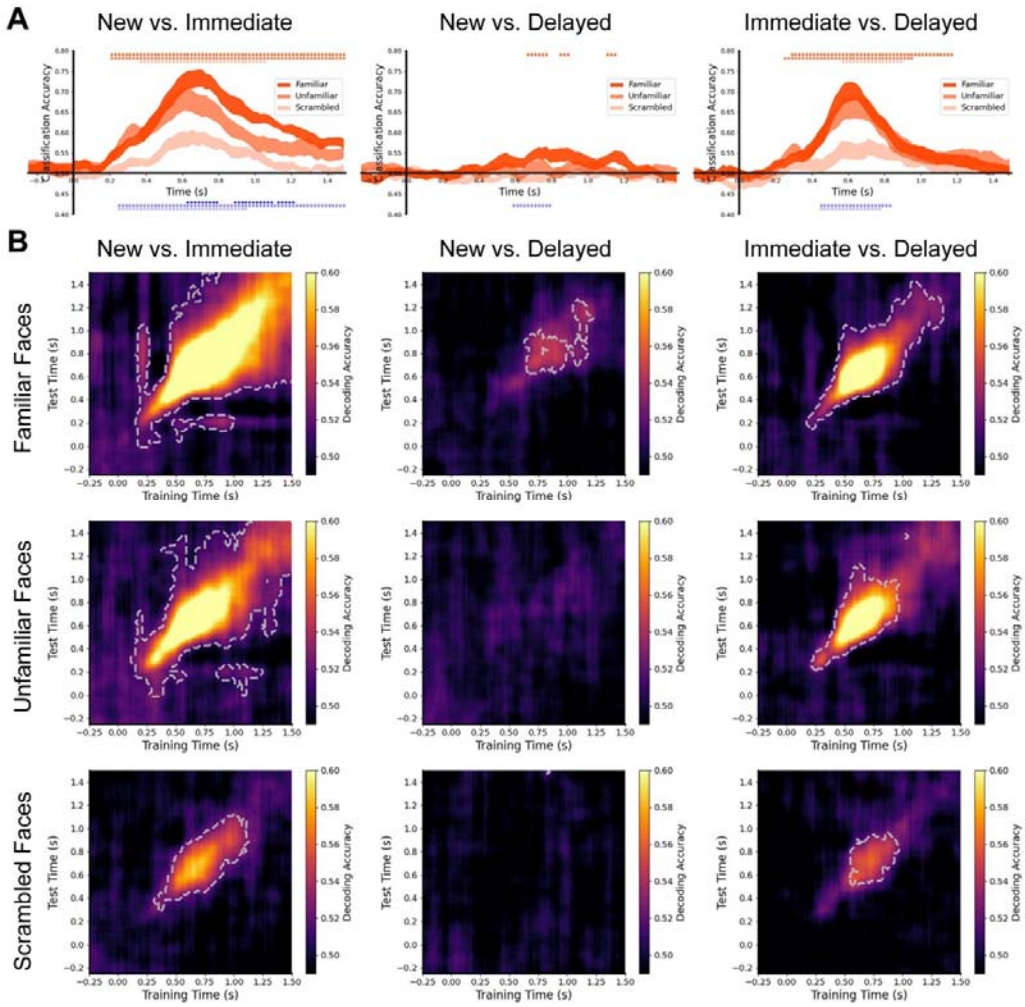


Figure 2. Temporal facial repetition suppression in human brains. (A) Time-by-time decoding results. Color-coded small squares at the top of each plot indicate p<0.01 (cluster-based permutation test) of decoding accuracy significantly greater than chance (from dark to light orange: familiar, unfamiliar, and scrambled faces). Color-

coded small squares at the bottom of each plot indicate $p<0.01$ (cluster-based permutation test) of significant difference of decoding accuracy between two face conditions (from dark to light blue: familiar vs. unfamiliar faces, familiar vs. scrambled faces, and unfamiliar vs. scrambled faces). Line width reflects ±SEM. (B) Cross-temporal decoding results. The baseline of classification-based decoding accuracy is 50%. Light grey outline indicates that average accuracy is significantly greater than chance (cluster-based permutation test, $p<0.01$).

These results revealed the neural mechanism of facial repetition suppression in human brains. Although the input face images were same, the neural representations were significantly different between new and immediate repetition conditions. This suppression effect decreased as the interval between twice became longer. Thus, if the same face image showed repeatedly after some trials, the neural representation would become more similar to the neural representation of that certain image presented at the first time. In addition, familiar faces induced a stronger repetition suppression effect than unfamiliar faces than scrambled faces.

Modified representations based on different repetition suppression models in DCNNs

For DCNNs, we firstly extracted features corresponding to 450 face images from all even layers and calculated 450×450 RDMs (Supplementary Figure 1B and 2). To simulate facial repetition suppression effect in DCNNs, we modified neural representations in both VGG-Face and untrained VGG based on Fatigue and Sharpening models, respectively. Figure 2 shows DCNN RDMs of layer 16 from DCNN's internal activations modified by two repetition suppression models. Based on Fatigue model, only the activation of the node with higher response to face stimulus had repetition suppression effect, and the node with higher activation had more suppressions. Based on Sharpening model, only the activation of the node with lower response to face stimulus had repetition suppression effect, and the nodes with low response would not be activated under repetition condition. Modified 1350×1350 RDMs for all even layers are shown in Supplementary Figure 3.

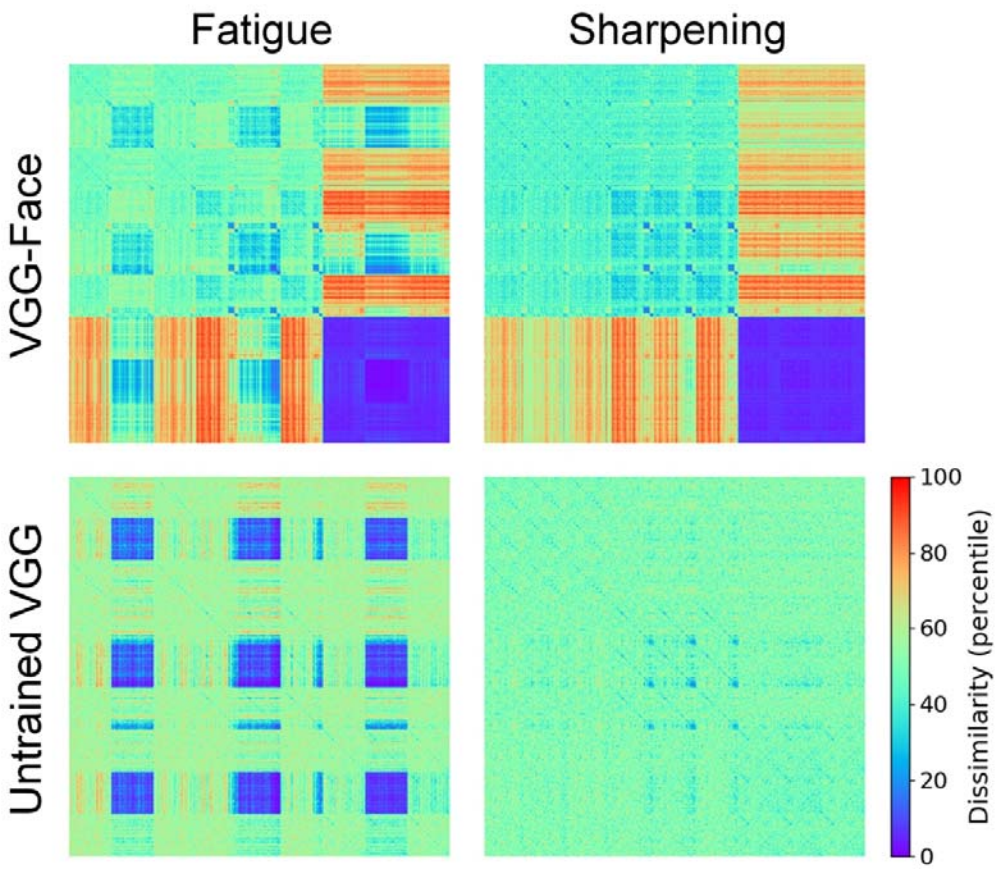


Figure 3. Modified DCNN RDMs of layer 16 based on two different repetition suppression mechanisms.

Comparisons between brains and DCNNs revealing a fatigue mechanism

After compressing the DCNN RDMs, we calculated the similarity between EEG RDMs and modified DCNN models then calculated the difference of cross-modal similarity between VGG-Face and untrained VGG as valid similarity. Figure 4A shows valid representational similarity between activations modified by Fatigue model of all even layers in DCNNs and EEG signals. DCNN representations modified by Fatigue model had significant representational similarity with human brains. We found significant valid similarities on many layers (layer 2: 120ms to 640ms, 700ms to 1420ms; layer 4: 120ms to 460ms, 1040ms to 1220ms; layer 10: 480ms to 560ms, 1380ms to 1500ms; layer 12: 260ms to 380ms, 420ms to 560ms, 1380ms to 1500ms; layer 14: 160ms to 1000ms; layer 16: 200ms to 1320ms). However, DCNN representations modified by Sharpening model almost had no significance with human brains. We only found significant valid similarities from 640ms to 680ms on layer 8.

In detailed, Figure 4B shows cross-temporal valid similarity of layer 16 which was the last layer in VGG structure and contain most relevant information of face

recognition. Based on Fatigue model's modification, the DCNN's representations showed strong and extensive valid similarity with human brains. However, representations of the DCNN modified by Sharpening model only had weak similarities with brain activity.

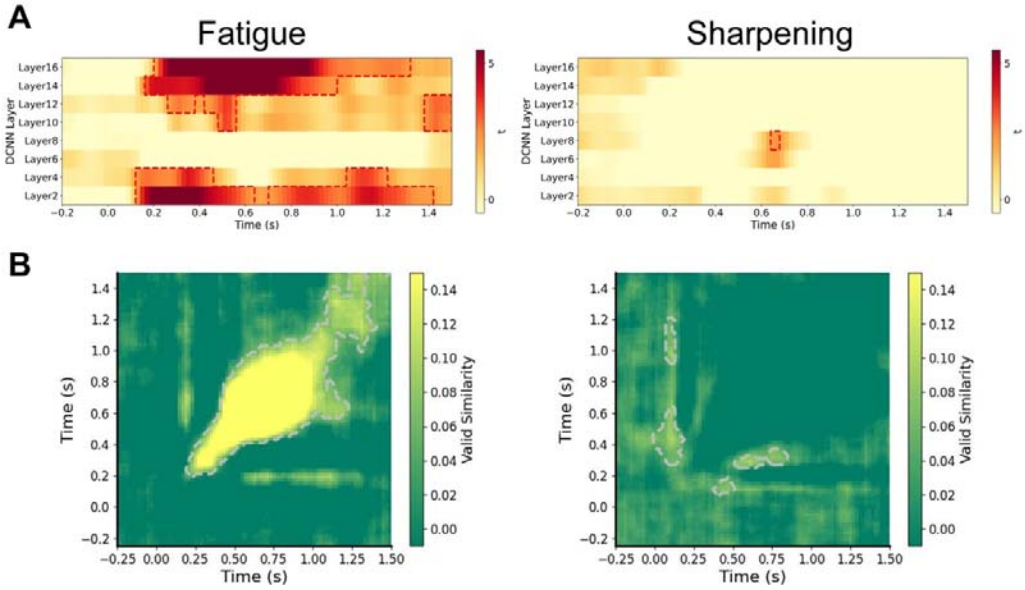


Figure 4. Representational comparison between brain activity and modified DCNN activations. (A) Layer-by-layer temporal valid similarity between EEG and modified DCNNs. (B) Cross-temporal valid similarity between EEG and modified DCNNs on layer 16. The baseline of valid similarity is zero. Outline indicates the significant cluster (cluster-based permutation test, $p<0.05$).

Cross-modal comparisons suggested that simulating the activation in DCNNs based on a fatigue instead of sharpening mechanism could induce more similar representations with human brains' activities. Thus, facial repetition suppression effect in face perception is more likely caused by the fatigue mechanism.

Discussion

In this study, we combined human EEG and DCNNs to explore the neural mechanism of repetition suppression in face perception using classification-based EEG decoding and cross-modal RSA. First, we decoded neural representations between different face presentation conditions based on EEG signals. The decoding results gave evidence of facial repetition suppression effect. It was stronger for familiar faces and weaker for scrambled faces, and it decreased if the repeated interval became longer. Then using reverse engineering, we explored the mechanism of facial repetition suppression in human brains by modifying the activation in AI models which could achieve the performance of face recognition at human level. Since VGG-Face was trained based on a large amount of face images for the purpose of face

identification, we focused on repetition suppression mechanism for face-specific information. In order to obtain the face-relevant information in the DCNN learned through the face training set, the result obtained by subtracting untrained VGG from the results obtained by VGG-Face was used as the valid representational similarity between trained DCNN with face-specific information and the human brain. Based on the results of valid representational similarity, the representation of DCNN model modified by Fatigue model was significantly similar to the representation of human brains. Cross-temporal results also showed that it was similar to brains' representations in a wider range. Meanwhile, these similarities appeared not only in the late layers of the DCNN, but also in the early layers. Previous studies have found that the hierarchical structure of DCNN is similar to that of human brain visual pathway (Cichy et al., 2016; Güçlü & van Gerven, 2015). These results suggested that the activation of neurons with stronger activation was more attenuated in the process of repetition suppression, and this kind of fatigue-based repetition suppression happened not only on neurons processing high-level face features, but also on neurons processing low-level face information.

Although multiple human fMRI studies (Andics et al., 2013; Ewbank et al., 2016; Grotheer et al., 2014; Grotheer & Kovács, 2014; Kovács et al., 2012, 2013; Larsson & Smith, 2012; Mayrhauser et al., 2014) supported that repetition suppression reflected a reduction in prediction error from predictive coding framework (Friston, 2005; Summerfield et al., 2008), several electrophysiological studies in recent years have challenged this that they supported fatigue mechanism with bottom-up or local adaptation instead of sharpening or sparseness representations and predictive coding hypothesis. However, due to the hardness of collecting human electrophysiological data, there were few human neuroimaging studies (Stam et al., 2021) to confirm these findings. Our current study is the first time to provide novel computational ways to explore the neural mechanism of facial repetition suppression using noninvasive human EEG and DCNNs and give strong evidence to support fatigue mechanism of repetition suppression.

Also, there still remains some deficiencies and problems to be solved. Firstly, are there other mechanisms of facial repetition suppression? Because pure DCNN models does not have temporal processing capability, our present study only evaluated two kinds of repetition suppression models, Fatigue and Sharpening models. Some recent studies found recurrent ANN models showed higher similarity with brain activity than models without recurrent componence in object recognition (Kietzmann et al., 2019; Spoerer et al., 2017). The further study can try to add timing process components to DCNN models, such as recurrent structures, to explore whether there may be a Facilitation mechanism (Grill-Spector et al., 2006) of repetition suppression. Also, the mechanism of repetition suppression might be different for different facial information or task types. Thus, it is worth further separating different conditions and finding the differences under different condition. Secondly, is DCNN the best brain-like model? Although DCNN is the most commonly used model in neuroscience, with

the rapid development of computer vision, some new ANN models also show very high task performance or effective feature space, such as generative adversarial network (GAN) (Goodfellow et al., 2014), vision transformer (ViT) (Dosovitskiy et al., 2020), unsupervised models (He et al., 2019), contrastive language-image pretraining (CLIP) (Radford et al., 2021), and spiking neural network (SNN) (Ghosh-Dastidar & Adeli, 2011). It is hard to evaluate which model is more similar to the human brain, which is also a cutting-edge research area both computer scientists and neuroscientists care about.

In addition, limited by the original experimental design of the open dataset we used, there are still many detailed face information worth exploring that have not been carefully and deeply analyzed, such as hair style, skin color, viewpoint, upright or inverted, gender, race, and facial expression. Therefore, we need to add more types of experimental stimuli to further explore the richer and more complex process of face perception and face repetition suppression. At the same time, the research in this part is all based on EEG data, and its purpose is to explore the dynamic neural representation in time series. In the future, we can further explore the spatial information processing based on fMRI and MEG data.

This study is a preliminary exploration of neuroscience problems using novel methods combined AI. By modifying the representation of AI model, we found the closest condition to the representation of human brains to try to answer what the neural mechanism of facial repetition suppression was. This comparison of cross modal representation can not only help us to uncover some neuroscience problems that are difficult to find evidence through noninvasive fMRI or EEG experiments, but also help give AI models more inspirations on the neuroscience level. Such research will provide important enlightenment for the development of brain like intelligence in the future.

Reference

Andics, A., Gál, V., Vicsi, K., Rudas, G., & Vidnyánszky, Z. (2013). FMRI repetition suppression for voices is modulated by stimulus expectations. *NeuroImage*, *69*, 277–283. https://doi.org/10.1016/J.NEUROIMAGE.2012.12.033

Barrett, S. E., & Rugg, M. D. (1989). Event-related potentials and the semantic matching of faces. *Neuropsychologia*, *27*(7), 913–922. https://doi.org/10.1016/0028-3932(89)90067-5

Baylis, G. C., & Rolls, E. T. (1987). Responses of neurons in the inferior temporal cortex in short term and serial recognition memory tasks. *Experimental Brain Research*, *65*(3), 614–622. https://doi.org/10.1007/BF00235984/METRICS

Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and Clinical Neurophysiology*, *60*(4), 343–355. https://doi.org/10.1016/0013-4694(85)90008-2

Burkhardt, A., Blaha, L. M., Jurs, B. S., Rhodes, G., Jeffery, L., Wyatte, D., Delong, J., & Busey, T. (2010). Adaptation modulates the electrophysiological substrates of perceived facial distortion:

Support for opponent coding. *Neuropsychologia*, *48*(13), 3743–3756. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2010.08.016

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 2016 6:1*, *6*(1), 1–13. https://doi.org/10.1038/srep27755

Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage*, *19*(2), 261–270. https://doi.org/10.1016/S1053-8119(03)00049-1

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/J.JNEUMETH.2003.10.009

Dobs, K., Isik, L., Pantazis, D., & Kanwisher, N. (2019). How face perception unfolds over time. *Nature Communications 2019 10:1*, *10*(1), 1–10. https://doi.org/10.1038/s41467-019-09239-1

Dörr, P., Herzmann, G., & Sommer, W. (2011). Multiple contributions to priming effects for familiar faces: Analyses with backward masking and event-related potentials. *British Journal of Psychology*, *102*(4), 765–782. https://doi.org/10.1111/J.2044-8295.2011.02028.X

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020). *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. https://doi.org/10.48550/arxiv.2010.11929

Drisdelle, B. L., Aubin, S., & Jolicoeur, P. (2017). Dealing with ocular artifacts on lateralized ERPs in studies of visual-spatial attention and memory: ICA correction versus epoch rejection. *Psychophysiology*, *54*(1), 83–99. https://doi.org/10.1111/PSYP.12675

Ewbank, M. P., von dem Hagen, E. A. H., Powell, T. E., Henson, R. N., & Calder, A. J. (2016). The effect of perceptual expectation on repetition suppression to faces is not modulated by variation in autistic traits. *Cortex*, *80*, 51–60. https://doi.org/10.1016/J.CORTEX.2015.10.011

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/RSTB.2005.1622

Ghosh-Dastidar, S., & Adeli, H. (2011). Spiking neural networks. *International Journal of Neural Systems*, *19*(4), 295–308. https://doi.org/10.1142/S0129065709002002

Golomb, J. D., & Kanwisher, N. (2012). Higher Level Visual Cortex Represents Retinotopic, Not Spatiotopic, Object Location. *Cerebral Cortex*, *22*(12), 2794–2810. https://doi.org/10.1093/CERCOR/BHR357

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Networks. *Science Robotics*, *3*(January), 2672–2680. https://doi.org/10.48550/arxiv.1406.2661

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, *10*(1), 14–23. https://doi.org/10.1016/J.TICS.2005.11.006

Grotheer, M., Hermann, P., Vidnyánszky, Z., & Kovács, G. (2014). Repetition probability effects for inverted faces. *NeuroImage*, *102*(P2), 416–423. https://doi.org/10.1016/J.NEUROIMAGE.2014.08.006

Grotheer, M., & Kovács, G. (2014). Repetition Probability Effects Depend on Prior Experiences.

Journal of Neuroscience, *34*(19), 6640–6646. https://doi.org/10.1523/JNEUROSCI.5326-13.2014

Güçlü, U., & van Gerven, M. A. J. (2015). Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. *Journal of Neuroscience*, *35*(27), 10005–10014. https://doi.org/10.1523/JNEUROSCI.5023-14.2015

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*(5539), 2425–2430. https://doi.org/10.1126/SCIENCE.1063736/SUPPL_FILE/HAXBYWEB.PDF

He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2019). Momentum Contrast for Unsupervised Visual Representation Learning. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 9726–9735. https://doi.org/10.48550/arxiv.1911.05722

Henson, R. N. A., & Rugg, M. D. (2003). Neural response suppression, haemodynamic repetition effects, and behavioural priming. *Neuropsychologia*, *41*(3), 263–270. https://doi.org/10.1016/S0028-3932(02)00159-8

Herzmann, G., Schweinberger, S. R., Sommer, W., & Jentzsch, I. (2004). What's special about personally familiar faces? A multimodal approach. *Psychophysiology*, *41*(5), 688–701. https://doi.org/10.1111/J.1469-8986.2004.00196.X

Jung, T. P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology*, *111*(10), 1745–1758. https://doi.org/10.1016/S1388-2457(00)00386-2

Kaliukhovich, D. A., & Vogels, R. (2011). Stimulus Repetition Probability Does Not Affect Repetition Suppression in Macaque Inferior Temporal Cortex. *Cerebral Cortex*, *21*(7), 1547–1558. https://doi.org/10.1093/CERCOR/BHQ207

Kaliukhovich, D. A., & Vogels, R. (2012). Stimulus repetition affects both strength and synchrony of macaque inferior temporal cortical activity. *Journal of Neurophysiology*, *107*(12), 3509–3527. https://doi.org/10.1152/JN.00059.2012/ASSET/IMAGES/LARGE/Z9K0131214220014.JPEG

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience 2005 8:5*, *8*(5), 679–685. https://doi.org/10.1038/nn1444

Kaufmann, J. M., & Schweinberger, S. R. (2012). The faces you remember: Caricaturing shape facilitates brain processes reflecting the acquisition of new face representations. *Biological Psychology*, *89*(1), 21–33. https://doi.org/10.1016/J.BIOPSYCHO.2011.08.011

Kietzmann, T. C., Spoerer, C. J., Sörensen, L. K. A., Cichy, R. M., Hauk, O., & Kriegeskorte, N. (2019). Recurrence is required to capture the representational dynamics of the human visual system. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(43), 21854–21863. https://doi.org/10.1073/PNAS.1905544116/SUPPL_FILE/PNAS.1905544116.SM06.MP4

Kloth, N., & Schweinberger, S. R. (2010). Electrophysiological correlates of eye gaze adaptation. *Journal of Vision*, *10*(12), 17–17. https://doi.org/10.1167/10.12.17

Kloth, N., Schweinberger, S. R., & Kovács, G. (2010). Neural Correlates of Generic versus Gender-specific Face Adaptation. *Journal of Cognitive Neuroscience*, *22*(10), 2345–2356. https://doi.org/10.1162/JOCN.2009.21329

Kovács, G., Iffland, L., Vidnyánszky, Z., & Greenlee, M. W. (2012). Stimulus repetition probability effects on repetition suppression are position invariant for faces. *NeuroImage*, *60*(4), 2128–2135.

https://doi.org/10.1016/J.NEUROIMAGE.2012.02.038

Kovács, G., Kaiser, D., Kaliukhovich, D. A., Zoltán, V., & Vogels, R. (2013). Repetition Probability Does Not Affect fMRI Repetition Suppression for Objects. *Journal of Neuroscience*, *33*(23), 9805–9812. https://doi.org/10.1523/JNEUROSCI.3423-12.2013

Kovács, G., Zimmer, M., Bankó, É., Harza, I., Antal, A., & Vidnyánszky, Z. (2006). Electrophysiological Correlates of Visual Adaptation to Faces and Body Parts in Humans. *Cerebral Cortex*, *16*(5), 742–753. https://doi.org/10.1093/CERCOR/BHJ020

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*(NOV), 4. https://doi.org/10.3389/NEURO.06.004.2008/BIBTEX

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., & Bandettini, P. A. (2008). Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. *Neuron*, *60*(6), 1126–1141. https://doi.org/10.1016/J.NEURON.2008.10.043

Kuzovkin, I., Vicente, R., Petton, M., Lachaux, J. P., Baciu, M., Kahane, P., Rheims, S., Vidal, J. R., & Aru, J. (2018). Activations of deep convolutional neural networks are aligned with gamma band activity of human visual cortex. *Communications Biology 2018 1:1*, *1*(1), 1–12. https://doi.org/10.1038/s42003-018-0110-y

Larsson, J., & Smith, A. T. (2012). fMRI Repetition Suppression: Neuronal Adaptation or Stimulus Expectation? *Cerebral Cortex*, *22*(3), 567–576. https://doi.org/10.1093/CERCOR/BHR119

Latinus, M., & Taylor, M. J. (2006). Face processing stages: Impact of difficulty and the separation of effects. *Brain Research*, *1123*(1), 179–187. https://doi.org/10.1016/J.BRAINRES.2006.09.031

Lu, Z. (2020). *PyCTRSA: A Python package for cross-temporal representational similarity analysis-based E/MEG decoding*. https://doi.org/10.5281/ZENODO.4273674

Lu, Z., & Ku, Y. (2020). NeuroRA: A Python Toolbox of Representational Analysis From Multi-Modal Neural Data. *Frontiers in Neuroinformatics*, *14*, 61. https://doi.org/10.3389/FNINF.2020.563669/BIBTEX

Maurer, U., Rossion, B., & McCandliss, B. D. (2008). Category specificity in early perception: Face and word N170 responses differ in both lateralization and habituation properties. *Frontiers in Human Neuroscience*, *2*(DEC), 18. https://doi.org/10.3389/NEURO.09.018.2008/BIBTEX

Mayrhauser, L., Bergmann, J., Crone, J., & Kronbichler, M. (2014). Neural repetition suppression: Evidence for perceptual expectation in object-selective regions. *Frontiers in Human Neuroscience*, *8*(1 APR), 225. https://doi.org/10.3389/FNHUM.2014.00225/ABSTRACT

Mercure, E., Kadosh, K. C., & Johnson, M. H. (2011). The N170 shows differential repetition effects for faces, objects, and orthographic stimuli. *Frontiers in Human Neuroscience*, *5*(JANUARY), 1–10. https://doi.org/10.3389/FNHUM.2011.00006/BIBTEX

Miller, E. K., Li, L., & Desimone, R. (1991). A Neural Mechanism for Working and Recognition Memory in Inferior Temporal Cortex. *Science*, *254*(5036), 1377–1379. https://doi.org/10.1126/SCIENCE.1962197

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), 424–430. https://doi.org/10.1016/J.TICS.2006.07.005

Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition. *BMVC 2015 - Proceedings of the British Machine Vision Conference 2015*.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M.,

Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *The Journal of Machine Learning Research*, *12*, 2825–2830. https://doi.org/10.5555/1953048.2078195

Pfütze, E. M., Sommer, W., & Schweinberger, S. R. (2002). Age-related slowing in face and name recognition: Evidence from event-related brain potentials. *Psychology and Aging*, *17*(1), 140–160. https://doi.org/10.1037/0882-7974.17.1.140

Phillips, P. J., Yates, A. N., Hu, Y., Hahn, C. A., Noyes, E., Jackson, K., Cavazos, J. G., Jeckeln, G., Ranjan, R., Sankaranarayanan, S., Chen, J. C., Castillo, C. D., Chellappa, R., White, D., & O'Toole, A. J. (2018). Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(24), 6171–6176. https://doi.org/10.1073/PNAS.1721355115/SUPPL_FILE/PNAS.1721355115.SD02.CSV

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). *Learning Transferable Visual Models From Natural Language Supervision*. https://doi.org/10.48550/arxiv.2103.00020

Ranjan, R., Patel, V. M., & Chellappa, R. (2019). HyperFace: A Deep Multi-Task Learning Framework for Face Detection, Landmark Localization, Pose Estimation, and Gender Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(1), 121–135. https://doi.org/10.1109/TPAMI.2017.2781233

Ringo, J. L. (1996). Stimulus specific adaptation in inferior temporal and medial temporal cortex of the monkey. *Behavioural Brain Research*, *76*(1–2), 191–197. https://doi.org/10.1016/0166-4328(95)00197-2

Rugg, M. D. (1985). The Effects of Semantic Priming and Word Repetition on Event-Related Potentials. *Psychophysiology*, *22*(6), 642–647. https://doi.org/10.1111/J.1469-8986.1985.TB01661.X

Sawamura, H., Orban, G. A., & Vogels, R. (2006). Selectivity of Neuronal Adaptation Does Not Match Response Selectivity: A Single-Cell Study of the fMRI Adaptation Paradigm. *Neuron*, *49*(2), 307–318. https://doi.org/10.1016/J.NEURON.2005.11.028

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A Unified Embedding for Face Recognition and Clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 815–823).

Schulz, C., Kaufmann, J. M., Kurt, A., & Schweinberger, S. R. (2012). Faces forming traces: Neurophysiological correlates of learning naturally distinctive and caricatured faces. *NeuroImage*, *63*(1), 491–500. https://doi.org/10.1016/J.NEUROIMAGE.2012.06.080

Schweinberger, S. R. (1996). How gorbachev primed yeltsin: Analyses of associative priming in person recognition by means of reaction times and event-related brain potentials. *Journal of Experimental Psychology: Learning Memory and Cognition*, *22*(6), 1383–1407. https://doi.org/10.1037/0278-7393.22.6.1383

Schweinberger, S. R., & Burton, A. M. (2003). Covert Recognition and the Neural System for Face Processing. *Cortex*, *39*(1), 9–30. https://doi.org/10.1016/S0010-9452(08)70071-6

Schweinberger, S. R., Kaufmann, J. M., Moratti, S., Keil, A., & Burton, A. M. (2007). Brain responses to repetitions of human and animal faces, inverted faces, and objects — An MEG study. *Brain Research*, *1184*(1), 226–233. https://doi.org/10.1016/J.BRAINRES.2007.09.079

Schweinberger, S. R., Pfütze, E. M., & Sommer, W. (1995). Repetition Priming and Associative Priming of Face Recognition: Evidence From Event-Related Potentials. *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, *21*(3), 722–736. https://doi.org/10.1037/0278-7393.21.3.722

Sobotka, S., & Ringo, J. L. (1994). Stimulus specific adaptation in excited but not in inhibited cells in inferotemporal cortex of Macaque. *Brain Research*, *646*(1), 95–99. https://doi.org/10.1016/0006-8993(94)90061-2

Spoerer, C. J., McClure, P., & Kriegeskorte, N. (2017). Recurrent convolutional neural networks: A better model of biological object recognition. *Frontiers in Psychology*, *8*(SEP), 1551. https://doi.org/10.3389/FPSYG.2017.01551/BIBTEX

Stam, D., Huang, Y. A., Vansteelandt, K., Sunaert, S., Peeters, R., Sleurs, C., Vrancken, L., Emsell, L., Vogels, R., Vandenbulcke, M., & Van den Stock, J. (2021). Long term fMRI adaptation depends on adapter response in face-selective cortex. *Communications Biology 2021 4:1*, *4*(1), 1–9. https://doi.org/10.1038/s42003-021-02235-6

Stevenage, S. V., Hale, S., Morgan, Y., & Neil, G. J. (2014). Recognition by association: Within- and cross-modality associative priming with faces and voices. *British Journal of Psychology*, *105*(1), 1–16. https://doi.org/10.1111/BJOP.12011

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience 2008 11:9*, *11*(9), 1004–1006. https://doi.org/10.1038/nn.2163

Taigman, Y., Marc', M. Y., Ranzato, A., & Wolf, L. (2014). DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1701–1708).

Urgen, B. A., Pehlivan, S., & Saygin, A. P. (2019). Distinct representations in occipito-temporal, parietal, and premotor cortex during action perception revealed by fMRI and computational modeling. *Neuropsychologia*, *127*, 35–47. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2019.02.006

Wakeman, D. G., & Henson, R. N. (2015). A multi-subject, multi-modal human neuroimaging dataset. *Scientific Data 2015 2:1*, *2*(1), 1–10. https://doi.org/10.1038/sdata.2015.1

Walther, C., Schweinberger, S. R., Kaiser, D., & Kovács, G. (2013). Neural correlates of priming and adaptation in familiar face perception. *Cortex*, *49*(7), 1963–1977. https://doi.org/10.1016/J.CORTEX.2012.08.012

Wiese, H., Kachel, U., & Schweinberger, S. R. (2013). Holistic face processing of own- and other-age faces in young and older adults: ERP evidence from the composite face task. *NeuroImage*, *74*, 306–317. https://doi.org/10.1016/J.NEUROIMAGE.2013.02.051

Xie, S., Kaiser, D., & Cichy, R. M. (2020). Visual Imagery and Perception Share Neural Representations in the Alpha Frequency Band. *Current Biology*, *30*(13), 2621-2627.e5. https://doi.org/10.1016/J.CUB.2020.04.074

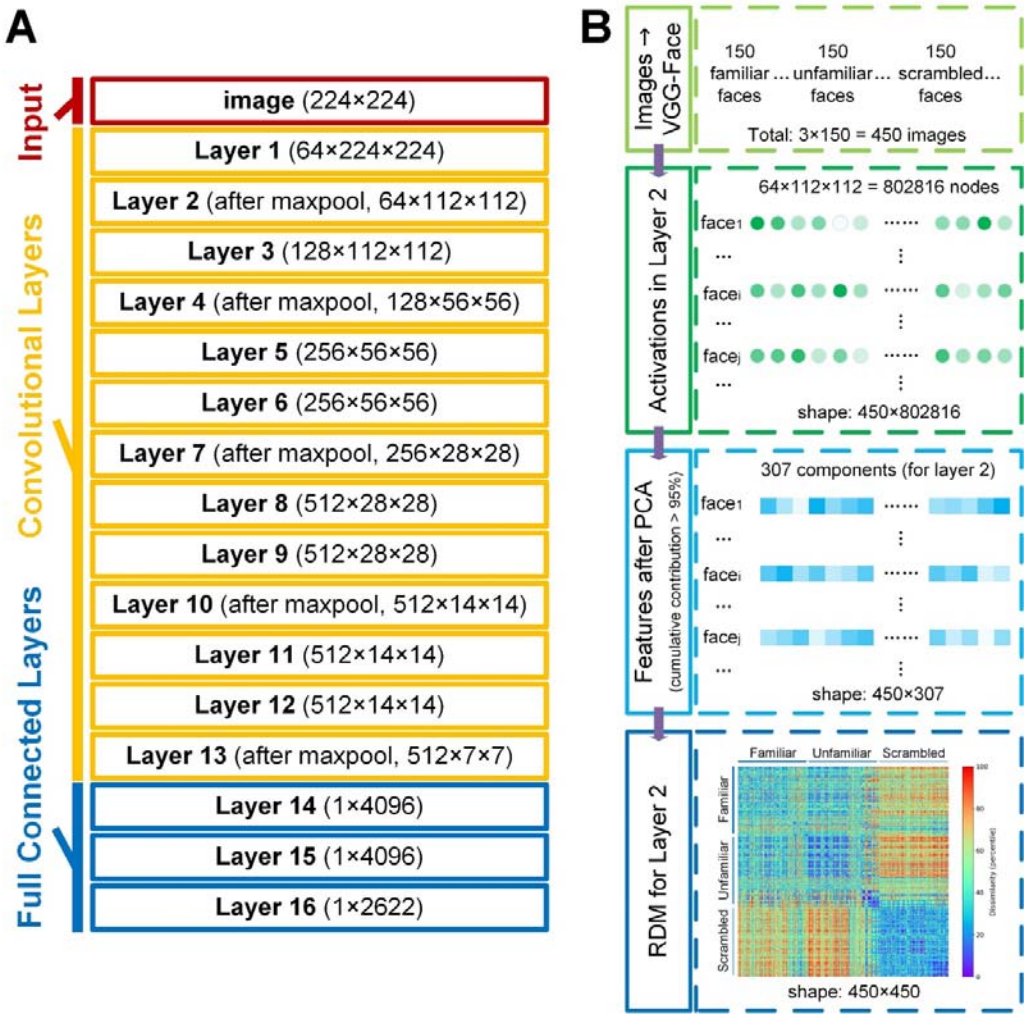
Xu, Y., & Vaziri-Pashkam, M. (2021). Examining the Coding Strength of Object Identity and Nonidentity Features in Human Occipito-Temporal Cortex and Convolutional Neural Networks. *Journal of Neuroscience*, *41*(19), 4234–4252. https://doi.org/10.1523/JNEUROSCI.1993-20.2021

Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(23), 8619–8624. https://doi.org/10.1073/PNAS.1403112111/SUPPL_FILE/PNAS.201403112SI.PDF

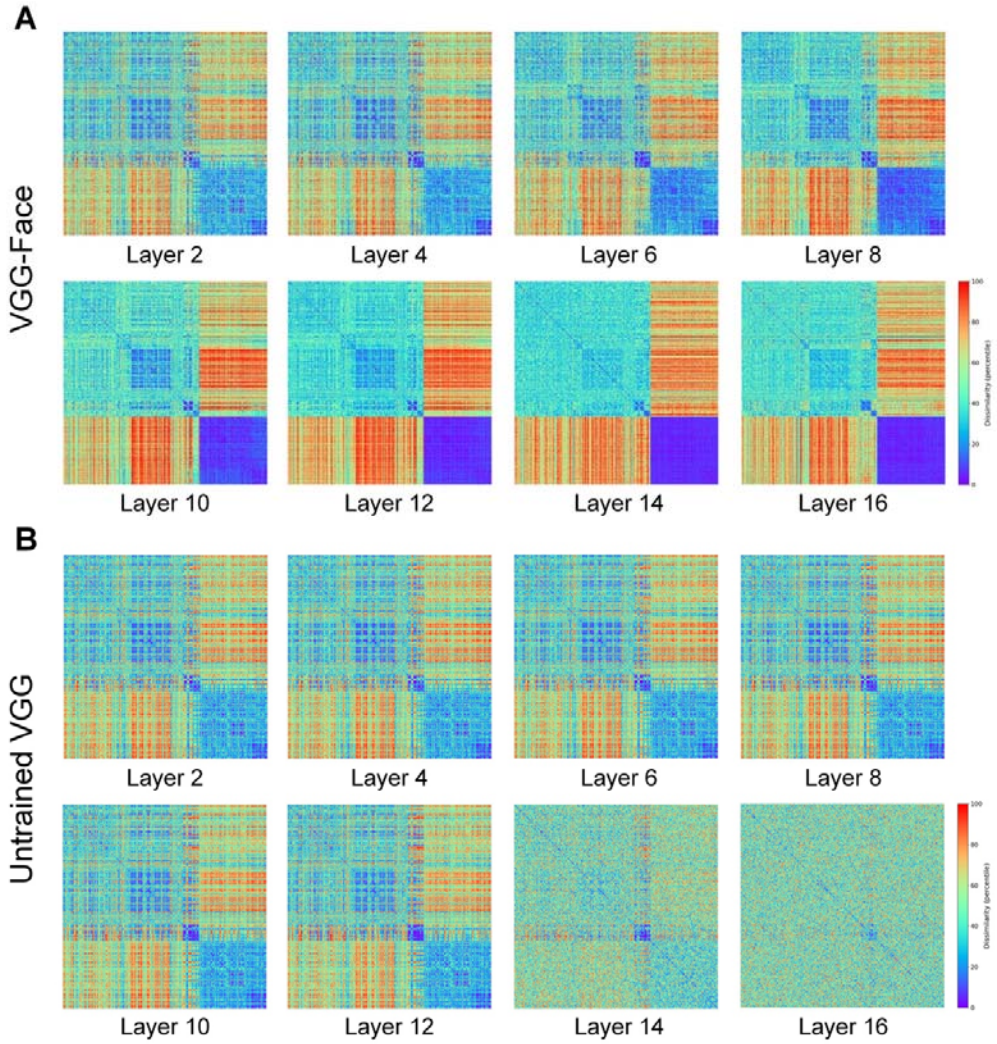
Zheng, X., Mondloch, C. J., & Segalowitz, S. J. (2012). The timing of individual face recognition in

the brain. *Neuropsychologia*, *50*(7), 1451–1461. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2012.02.030

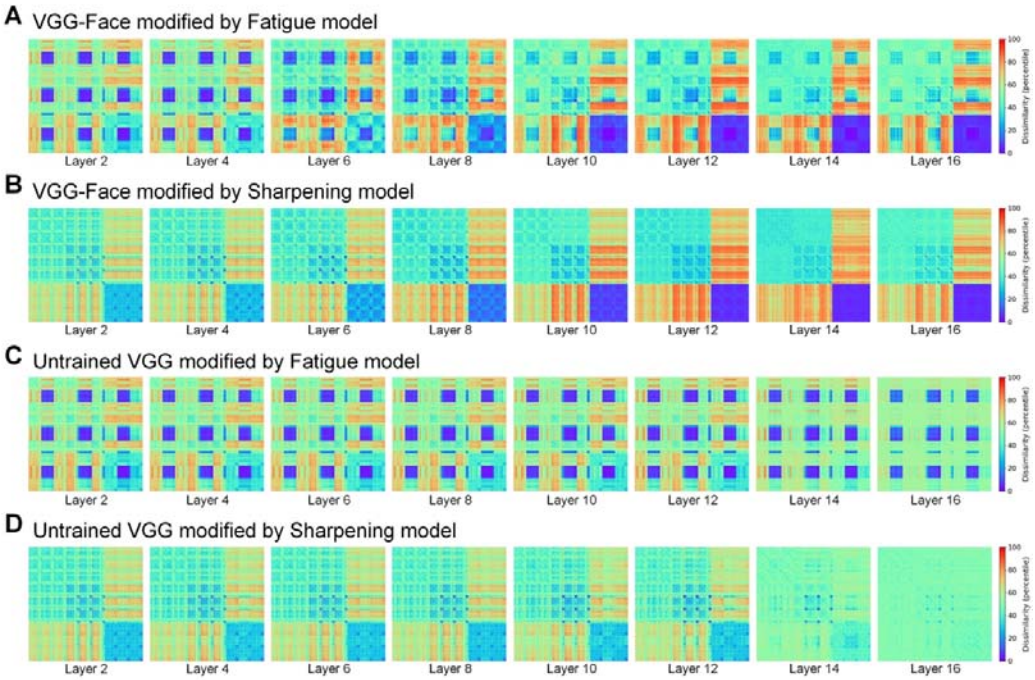
Supplementary



Supplementary Figure 1. (A) VGG-Face structure and (B) calculation steps of DCNN RDMs.



Supplementary Figure 2. DCNN 450×450 RDMs based on (A) VGG-Face and (B) Untrained VGG.



Supplementary Figure 3. 1350×1350 RDMs based on (A) VGG-Face modified by Fatigue model, (B) VGG-Face modified by Sharpening model, (C) untrained VGG modified by Fatigue model, and (D) untrained VGG modified by Sharpening model.