Patterns

A critical period for developing face recognition

Highlights

- Deep artificial neural networks exhibit critical periods
- Face learning can only be restored within but not outside of the critical period
- A computational account by learning rate explains the properties of critical period
- Knowledge distillation and attention transfer partially recover face learning

Authors

Jinge Wang, Runnan Cao, Puneeth N. Chakravarthula, Xin Li, Shuo Wang

Correspondence

xli48@albany.edu (X.L.), shuowang@wustl.edu (S.W.)

In brief

Face learning has important critical periods during development, but the underlying computational mechanisms remain unknown. Wang et al. provide a full computational account for face-learning behaviors by using deep artificial neural networks and show that face learning can only be restored when providing information within the critical period.

Wang et al., 2024, Patterns *5*, 100895
February 9, 2024
https://doi.org/10.1016/j.patter.2023.100895

Patterns

CellPress
OPEN ACCESS

Article

A critical period for developing face recognition

Jinge Wang,[1] Runnan Cao,[1,2] Puneeth N. Chakravarthula,[2] Xin Li,[1,3,*] and Shuo Wang[1,2,4,*]
[1]Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV 26506, USA
[2]Department of Radiology, Washington University in St. Louis, St. Louis, MO 63110, USA
[3]Department of Computer Science, University at Albany, Albany, NY 12222, USA
[4]Lead contact
*Correspondence: xli48@albany.edu (X.L.), shuowang@wustl.edu (S.W.)
https://doi.org/10.1016/j.patter.2023.100895

THE BIGGER PICTURE Just as humans have critical development phases for facial recognition, deep artificial neural networks exhibit similar periods. These critical phases determine the network's ability to acquire and process facial information, and like in humans, impaired face learning can occur when information is lacking during these critical periods. Fortunately, we found that restoration is possible if the necessary input is provided within this critical window. Beyond this time frame, the model's capacity to absorb new information wanes. Our work not only uncovers the computational foundations of face learning but also offers insights into its behavior and strategies for recovering impaired face learning.

Proof-of-Concept: Data science output has been formulated, implemented, and tested for one domain/problem

SUMMARY

Face learning has important critical periods during development. However, the computational mechanisms of critical periods remain unknown. Here, we conducted a series of *in silico* experiments and showed that, similar to humans, deep artificial neural networks exhibited critical periods during which a stimulus deficit could impair the development of face learning. Face learning could only be restored when providing information *within* the critical period, whereas, *outside of* the critical period, the model could not incorporate new information anymore. We further provided a full computational account by learning rate and demonstrated an alternative approach by knowledge distillation and attention transfer to partially recover the model outside of the critical period. We finally showed that model performance and recovery were associated with identity-selective units and the correspondence with the primate visual systems. Our present study not only reveals computational mechanisms underlying face learning but also points to strategies to restore impaired face learning.

INTRODUCTION

A critical period is a time window during development when some particular experience must be undergone for the complete development of language and sensory systems to occur.[1] The critical period hypothesis was originally proposed for the acquisition of a second language[2] and visual perception.[3] In children born with opacity or deviation of the eyes, the deprived eye will suffer from lack of a cortical response despite a healthy retina. The consequences of such sensory deprivation can lead to lifelong amblyopia (due to ocular dominance plasticity).[4,5] Similarly, it has been hypothesized that there is a critical period for the development of the fusiform face area (FFA), which has an intriguing connection (e.g., atypical fixation patterns in autism) with the difference in face processing by individuals with autism spectrum disorder (ASD).[6–13] People with ASD have an increased tendency to saccade away from the eye region of faces when information is present in those regions[14] and instead have an increased preference to fixate on the location of the mouth.[15] However, it has remained an open question whether there are critical periods in the development of face processing,[10] what the computational mechanisms of critical periods are, and what the developmental trajectory of facial feature selection is.

It has been argued that the neural coding of visual stimuli can change over development.[16] A study using functional MRI to examine the development of several functionally defined regions, including object, face, and place-selective cortices in different age groups (children, adolescents, and adults), has shown that development that occurred through the expansion

Please cite this article in press as: Wang et al., A critical period for developing face recognition, Patterns (2023), https://doi.org/10.1016/j.patter.2023.100895

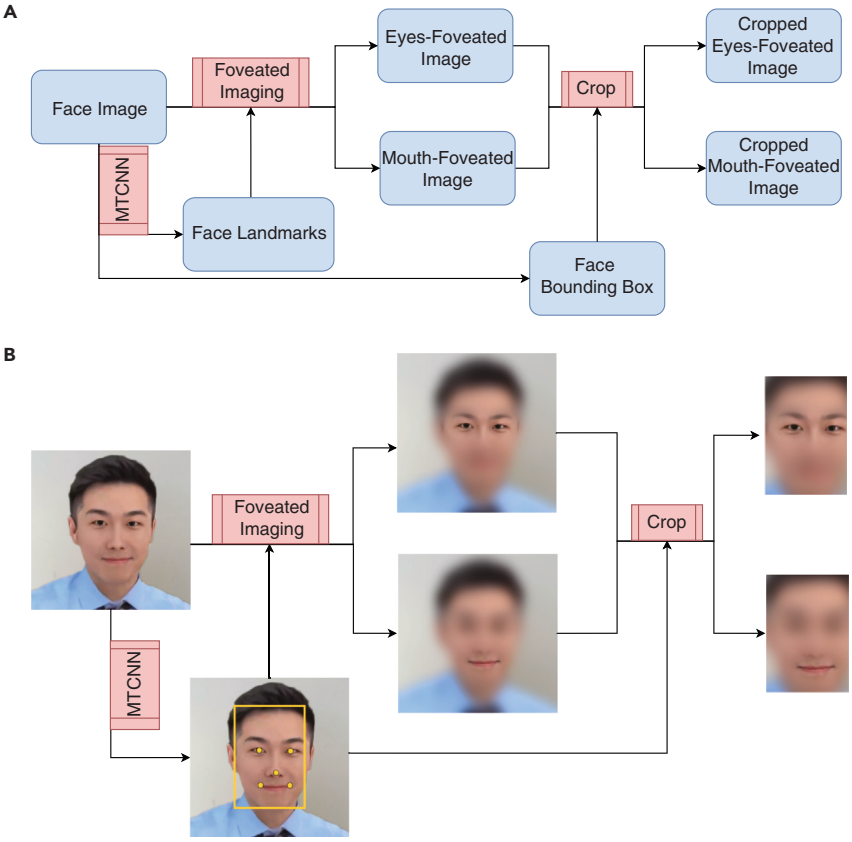

Figure 1. Image processing pipeline

(A) Image processing pipeline. MTCNN was applied to the original face images to detect a tight bounding box outlining the face area and facial landmarks (centers of the eyes, nose tip, and corners of the mouth). Foveation imaging was applied to the original face images to derive two sets of foveated images (eye-foveated and mouth-foveated). We finally cropped the images based on the bounding box derived using the MTCNN.

(B) An example showing the image processing pipeline. For copyright reasons, we replaced the original stimulus image in this and subsequent figures with a similar picture.

of the FFA and parahippocampal place area (PPA) into the surrounding cortex is correlated with improved recognition memory for faces and places, respectively.[17] Furthermore, microstructural proliferation in the human cortex is coupled with the development of face processing.[18] In addition to the developmental trajectory, critical periods play an important role in the computational mechanisms of face learning. Our recent study has provided a neuronal mechanism for face learning: neuronal distance between face identities increases as a function of exposure, suggesting that faces become more neurally distinct after learning.[19] Notably, the core components of face processing and their neuromaturational time course in typical development (TD) may facilitate our understanding of face processing deficits in autism as well as development of clinical tools for early diagnosis and remediation.[20]

Given the challenges of creating a critical period in the physical world, developing computational surrogate models[21] has become an appealing alternative. Deep neural networks (DNN) such as VGG-Face[22] and FaceNet[23] have achieved comparable or even superior face recognition performance compared with human observers. These DNN-based surrogate models have made it convenient to conduct experiments with deprived stimuli or perturbation of network architectures.[24,25] For example, the existence of a "critical period" (usually the first few epochs) in the DNN has been shown experimentally,[21] suggesting that critical periods are not restricted to biological systems but can emerge naturally in learning systems, whether biological or artificial, due to fundamental constraints arising from learning dynamics and information processing. It has been shown that the use of unsupervised learning in DNNs can provide a quantitative model of the ventral visual processing system and serve as a biologically plausible computational model of primate sensory learning.[26] More broadly, DNNs provide an important approach to testing the computational benefits of fundamental organizational features of the visual system.[27] DNNs create a highly organized face similarity structure where natural image variation is organized hierarchically, offering an important theoretical framework to understand identity coding.[28] Furthermore, it has been shown that brain-like functional specialization emerges spontaneously in DNNs and reflects a computational optimization for face recognition.[29] In sum, *in silico* experiments with DNNs have provided unprecedented opportunities to understand face coding and learning, especially when artificial models show correspondence with brain models.[25,30–33]

In this study, we hypothesize that, similar to humans and animals, deep artificial neural networks exhibit critical periods during which a stimulus deficit can impair the development of face learning. We further hypothesize that face learning can only be restored when providing information *within* the critical period but not *outside of* the critical period. We seek a computational account for critical periods and explore possible ways to restore face learning. We hypothesize that the learning rate is a key factor for critical periods. We finally explore the correspondence with primate visual systems, which, in turn, may explain the recovery mechanism from the critical periods.

RESULTS

A surrogate model for developing face recognition

We first used full-face images to train a DNN based on the ResNet50 architecture (Figures 1 and 2A). We observed a rapid increase in performance in early training epochs (Figure 2B), which reached a plateau after 30 epochs. We next quantified information utilization in the images using gradient-weighted class activation mapping (Grad-CAM; methods). The heatmaps reflected the regions in the face that contributed to the correct classification of face identities (Figure 2E). In the full-face model, we found that the network utilized information from both the

Please cite this article in press as: Wang et al., A critical period for developing face recognition, Patterns (2023), https://doi.org/10.1016/j.patter.2023.100895

CellPress
OPEN ACCESS

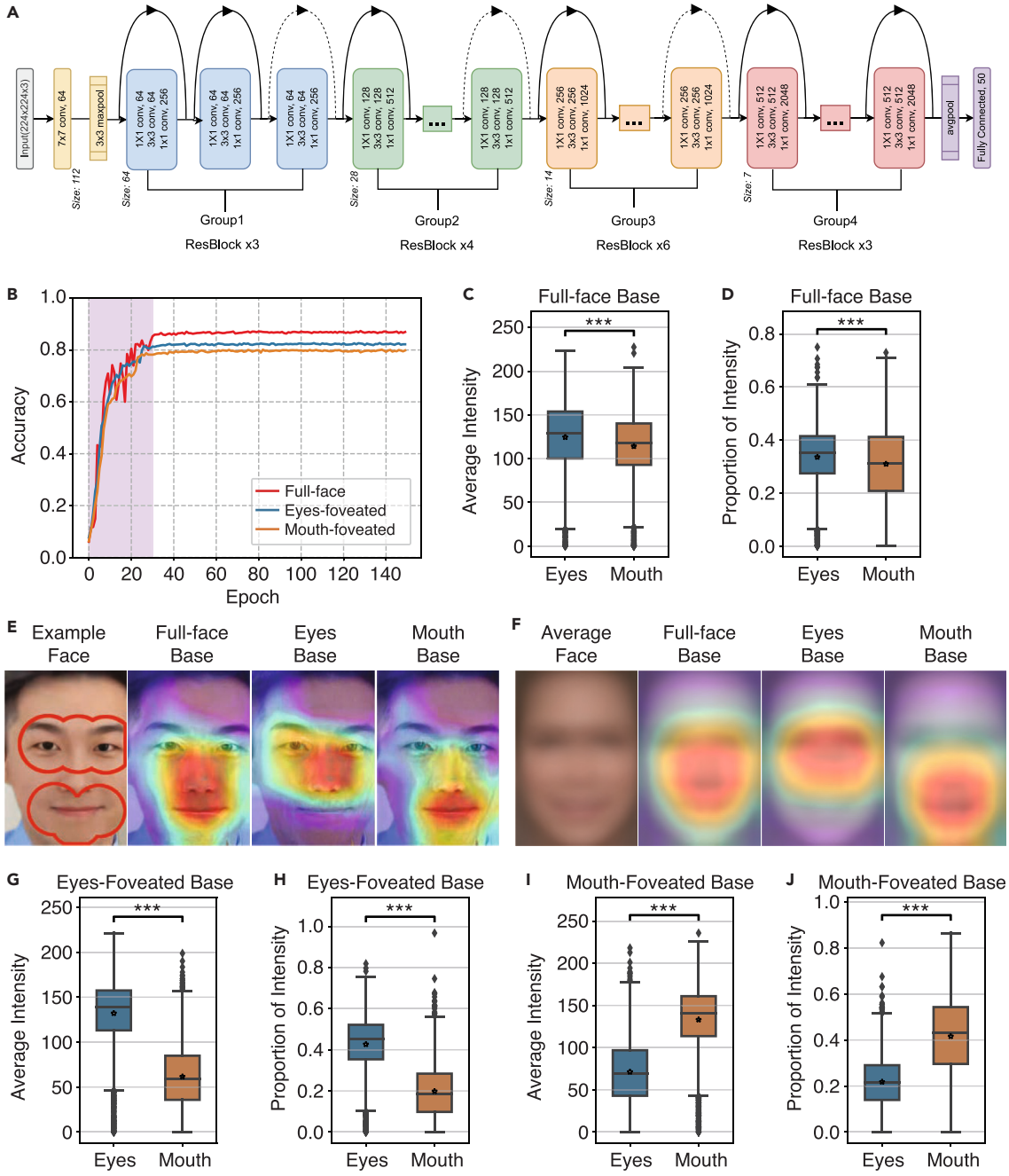

Figure 2. Face recognition models with different training stimuli

(A) ResNet50 architecture. ResNet50 has 4 stages. It performs the initial convolution and max-pooling using 7 × 7 and 3 × 3 kernels, respectively. Subsequently, the inputs go through the 4 stages. All stages contain the basic residual blocks. For ResNet50, there are 3, 4, 6, and 3 residual blocks in stages 1, 2, 3, and 4, respectively. In each residual block, 3 convolution layers (1 × 1, 3 × 3, and 1 × 1) are stacked. The 1 × 1 convolution layers are responsible for reducing and then restoring the dimensions. The 3 × 3 layer is left as a bottleneck with smaller input/output dimensions. The curved arrows are skip connections or "shortcuts." Solid connections refer to the identity connection. The dashed connection denotes that the convolution operation in the residual block is performed with stride 2. As feature maps progress from one stage to another, the channel width is doubled, and the size of the input is reduced to half. Finally, the network has an average pooling layer followed by a fully connected layer having 50 neurons (number of different identities).

(B) Network learning curve. The validation accuracy of face identity recognition is plotted as a function of the model training epoch. The shaded area denotes the critical period.

(C, G, and I) Average Grad-CAM intensity for each region of interest (ROI).

(D, H, and J) The proportion of Grad-CAM intensity for each ROI. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, and the whiskers extend to the most extreme data points the algorithm considers to not be outliers. Asterisks indicate a significant difference using a two-tailed paired t test. ***$p < 0.001$.

(legend continued on next page)

Please cite this article in press as: Wang et al., A critical period for developing face recognition, Patterns (2023), https://doi.org/10.1016/j.patter.2023.100895

eyes and mouth (see Figure 2E for an example; see Figure 2F for group average), although the network utilized more information from the eye region than the mouth region (Figure 2C; average Grad-CAM intensity: eyes: 124.47 ± 39.97, mouth: 114.10 ± 37.51; two-tailed two-sample t test: $t(7536) = 11.62$, $p < 10^{-30}$; Figure 2D; proportion of Grad-CAM intensity: eyes: 0.34 ± 0.11, mouth: 0.31 ± 0.14; $t(7536) = 8.79$, $p < 10^{-17}$).

To study the impact of facial information, we employed foveated imaging and created eye-foveated and mouth-foveated images (Figure 1; methods). As expected, with reduced information, the models reached a lower performance (Figure 2B). Interestingly, the eye-foveated model had a better performance compared with the mouth-foveated model, indicating that the eyes contained more information than the mouth for face recognition. Notably, we observed a similar "critical period" (the first 30 epochs; see formal definitions below) compared with the full-face model, suggesting that the foveated models had a similar learning process.

We also quantified information utilization in the images in these models (Figures 2E and 2F). Indeed, the eye-foveated model utilized more information in the eyes than the mouth (Figure 2G; average Grad-CAM intensity: eyes: 132.08 ± 37.71, mouth: 61.59 ± 35.62; $t(7536) = 83.42$, $p < 10^{-30}$; Figure 2H; proportion of Grad-CAM intensity: eyes: 0.43 ± 0.14, mouth: 0.20 ± 0.13; $t(7536) = 73.61$, $p < 10^{-30}$), whereas the mouth-foveated model utilized more information in the mouth than the eyes (Figure 2I; average Grad-CAM intensity: eyes: 71.25 ± 37.65, mouth: 133.14 ± 40.25; $t(7536) = 68.95$, $p < 10^{-30}$; Figure 2J; proportion of Grad-CAM intensity: eyes: 0.22 ± 0.11, mouth: 0.42 ± 0.18; $t(7536) = 58.55$, $p < 10^{-30}$). This result confirmed that reducing certain visual inputs into training would lead to reduced utilization of the corresponding visual information. On the other hand, the eye-foveated model had a higher average Grad-CAM intensity in the eyes than the full-face model (Figure 2G vs. Figure 2C; two-tailed paired t test: $t(3768) = 12.03$, $p < 10^{-30}$), and the mouth-foveated model had a higher average Grad-CAM intensity in the mouth than the full-face model (Figure 2I vs. Figure 2C; $t(3768) = 30.39$, $p < 10^{-30}$), suggesting that the network could adjust to focus on available information.

Last, we showed that another popular DNN model for face recognition (i.e., VGG-Face) had a similar learning curve and critical period and that full-face models outperformed eye-foveated models and mouth-foveated models (Figures S2A and S2B). Therefore, we confirmed that our results were not idiosyncratic to the DNN model used in the present study.

Recovery with full-face images within vs. outside of the critical period

Above, we have revealed a critical period during DNN training (i.e., learning face identities) and illustrated the information utilization during this process. We next investigated whether training with restricted stimuli (eye-foveated faces or mouth-foveated faces) could be recovered with additional visual information.

We first used full-face images to recover impaired models. We found that providing full-face information to the network *within* the critical period led to better performance, and this was the case for both the eye-foveated model (Figure 3A) and mouth-foveated model (Figure 3B). However, providing full-face information to the network *outside of* the critical period did not improve the performance (Figures 3A and 3B), and it could even deteriorate the accuracy for the mouth-foveated model (Figure 3B). This result was confirmed with different starting points of recovery within or outside of the critical period (Figures S1B and S1C).

Importantly, the change in performance with recovery was associated with different utilization of facial information. For the eye-foveated model, recovering within the critical period led to an increased utilization of mouth information compared with recovering outside the critical period (see Figure 3C for an example and Figure 3D for group average; Figure 3E; average Grad-CAM intensity: within: 90.03 ± 37.21, outside: 67.27 ± 32.00; $t(3768) = 53.68$, $p < 10^{-30}$; Figure 3F; proportion of Grad-CAM intensity: within: 0.25 ± 0.13, outside: 0.21 ± 0.12; $t(3768) = 45.65$, $p < 10^{-30}$). For both recovery conditions, the eyes still contributed more information than the mouth (Figure 3E; average Grad-CAM intensity: within: $t(7536) = 53.82$, $p < 10^{-30}$, outside: $t(7536) = 89.85$, $p < 10^{-30}$; Figure 3F; proportion of Grad-CAM intensity: within: $t(7536) = 43.84$, $p < 10^{-30}$, outside: $t(7536) = 72.29$, $p < 10^{-30}$).

Similarly, for the mouth-foveated model, recovering within the critical period led to increased utilization of eye information compared with recovering outside of the critical period (see Figure 3C for an example and Figure 3D for group average; Figure 3G; average Grad-CAM intensity: within: 109.18 ± 40.77, outside: 76.03 ± 35.50; $t(3768) = 63.39$, $p < 10^{-30}$; Figure 3H; proportion of Grad-CAM intensity: within: 0.29 ± 0.11, outside: 0.23 ± 0.10; $t(3768) = 56.29$, $p < 10^{-30}$). For both recovery conditions, the mouth still contributed more information than the eyes (Figure 3G; average Grad-CAM intensity: within: $t(7536) = 19.15$, $p < 10^{-30}$, outside: $t(7536) = 67.13$, $p < 10^{-30}$; Figure 3H; proportion of Grad-CAM intensity: within: $t(7536) = 16.34$, $p < 10^{-30}$, outside: $t(7536) = 54.58$, $p < 10^{-30}$).

We next investigated the extent to which the foveated models recovered by comparing them with the full-face model. Although the eye-foveated model recovered with full-face images within the critical period improved performance, it did not fully reach the level of the full-face model in model performance (85.67% vs. 87.32%; Figure 3A vs. Figure 2B) and utilization of facial information (Figure 3E vs. Figure 2C; average Grad-CAM intensity: eyes: $t(3768) = 22.72$, $p < 10^{-30}$, mouth: $t(3768) = 42.73$, $p < 10^{-30}$; Figure 3F vs. Figure 2D; proportion of Grad-CAM intensity: eyes: $t(3768) = 42.60$, $p < 10^{-30}$, mouth: $t(3768) = 45.06$, $p < 10^{-30}$), suggesting that the impaired

(E) The Grad-CAM intensity maps for an example face.
(F) The Grad-CAM intensity maps for group average across faces. The intensity values indicate the contribution/importance of pixels for face recognition. The red contours in the example face delineate the eyes and mouth ROIs for this face.
(C and D) Full-face model.
(G and H) Eye-foveated model.
(I and J) Mouth-foveated model.

Please cite this article in press as: Wang et al., A critical period for developing face recognition, Patterns (2023), https://doi.org/10.1016/j.patter.2023.100895

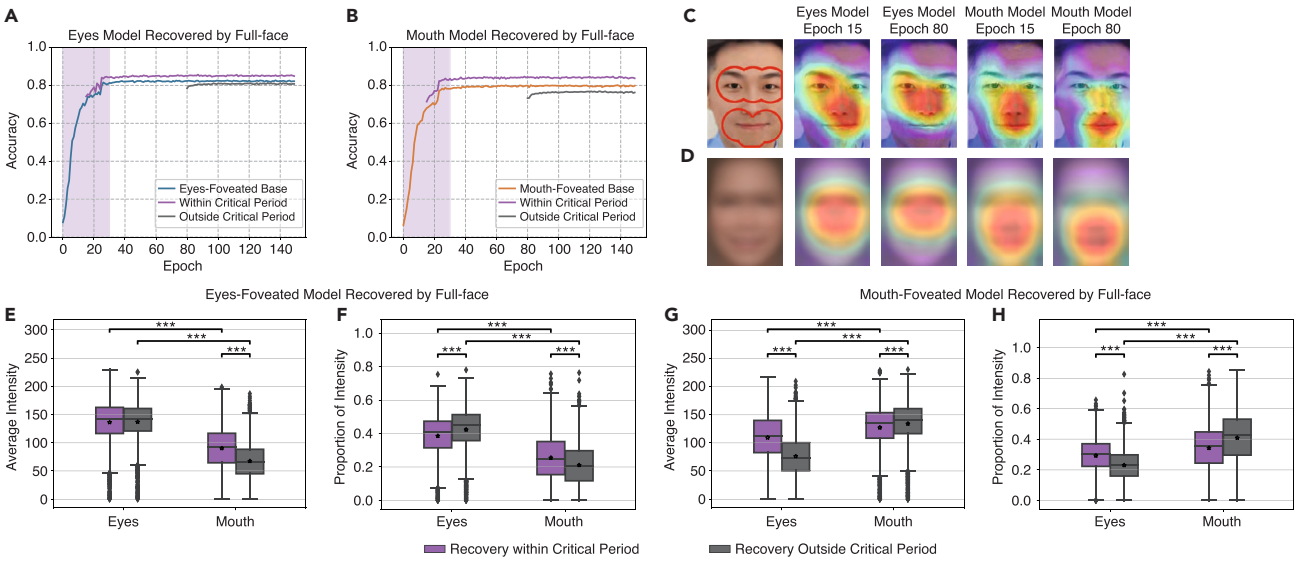

Figure 3. Recovery with full-face images

(A) The learning curve for the eye-foveated model.
(B) The learning curve for the mouth-foveated model.
(C) The Grad-CAM intensity maps for an example face.
(D) The Grad-CAM intensity maps for group average across faces.
(E and G) Average Grad-CAM intensity for each ROI.
(F and H) The proportion of Grad-CAM intensity for each ROI.
(E and F) Eye-foveated model.
(G and H) Mouth-foveated model.
Legend conventions are as in Figure 2.

eye-foveated model could be partially recovered within the critical period. Similarly for the mouth-foveated model, when recovery with full-face images happened within the critical period, the resultant model improved model performance (84.66% vs. 87.32%; Figure 3B vs. Figure 2B) and utilization of facial information (Figure 3G vs. Figure 2C; average Grad-CAM intensity: eyes: $t(3768) = 26.30$, $p < 10^{-30}$, mouth: $t(3768) = 22.56$, $p < 10^{-30}$; Figure 3H vs. Figure 2D; proportion of Grad-CAM intensity: eyes: $t(3768) = 34.66$, $p < 10^{-30}$, mouth: $t(3768) = 34.40$, $p < 10^{-30}$) toward the full-face model, although the impaired mouth-foveated model was partially recovered.

Together, our results suggest that providing information in the critical period can recover model performance and information usage in the impaired models. However, providing information outside of the critical period cannot recover impaired models anymore, which justifies the importance of timing for the recovery of impaired models.

Recovery with complementary information within vs. outside of the critical period

Above, we have shown recovery with full-face images, which contain full information of the faces. Can we recover impaired models with complementary information (i.e., providing an eye-foveated model with mouth-foveated images or a mouth-foveated model with eye-foveated images)?

To answer these questions, we input mouth-foveated images into the eye-foveated model within and outside of the critical period. Recovering within the critical period using complementary stimuli led to a similar model performance (Figure 4A), but interestingly, recovering outside of the critical period even deteriorated the model performance (Figure 4A). As expected, recovering outside of the critical period did not change information utilization; the eyes still contributed more information than the mouth (see Figure 4C for an example and Figure 4D for group summary; Figure 4E; average Grad-CAM intensity: eyes: 128.15 ± 38.63, mouth: 77.75 ± 37.45; $t(7536) = 57.51$, $p < 10^{-30}$; Figure 4F; proportion of Grad-CAM intensity: eyes: 0.38 ± 0.13, mouth: 0.24 ± 0.14; $t(7536) = 49.23$, $p < 10^{-57}$). However, notably, recovering within the critical period led to an opposite pattern of information utilization: the mouth contributed more information than the eyes (see Figure 4C for an example and Figure 4D for group summary; Figure 4E; average Grad-CAM intensity: eyes: 97.56 ± 41.72, mouth: 120.46 ± 39.65; $t(7536) = 24.42$, $p < 10^{-30}$; Figure 4F; proportion of Grad-CAM intensity: eyes: 0.28 ± 0.12, mouth: 0.36 ± 0.16; $t(7536) = 21.95$, $p < 10^{-30}$), a pattern of results that were more similar to the mouth-foveated model. This result suggests that new complementary information provided during the critical period overrode the original information utilization. In other words, it indicates that the network mainly takes information provided later. It is worth noting that the model performance seemed to switch as well (Figures 4A and 4B): the eye-foveated model turned into the mouth-foveated model.

Similarly, when we input eye-foveated images into the mouth-foveated model, we found that recovering outside of the critical period even deteriorated the model performance (Figure 4B), and the critical period did not change information utilization: the mouth still contributed more information than the eyes (see

Please cite this article in press as: Wang et al., A critical period for developing face recognition, Patterns (2023), https://doi.org/10.1016/j.patter.2023.100895

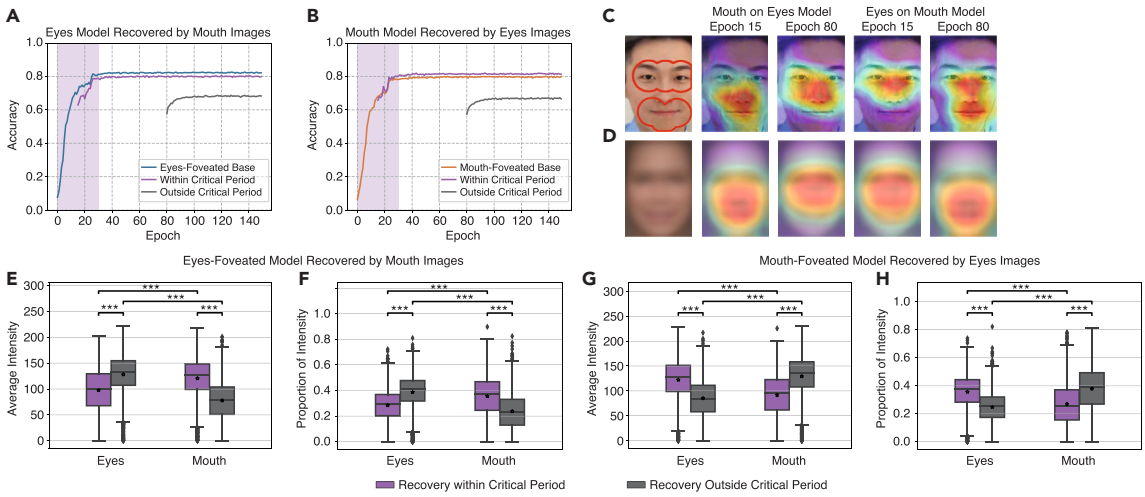

Figure 4. Recovery with images with complementary information

(A) The learning curve for the eye-foveated model.

(B) The learning curve for the mouth-foveated model.

(C) The Grad-CAM intensity maps for an example face.

(D) The Grad-CAM intensity maps for group average across faces.

(E and G) Average Grad-CAM intensity for each ROI.

(F and H) The proportion of Grad-CAM intensity for each ROI.

(E and F) Eye-foveated model.

(G and H) Mouth-foveated model.

Legend conventions are as in Figure 2.

Figure 4C for an example and Figure 4D for group summary; Figure 4G; average Grad-CAM intensity: eyes: 84.84 ± 38.36, mouth: 128.98 ± 41.65; $t(7536) = 47.86$, $p < 10^{-30}$; Figure 4H; proportion of Grad-CAM intensity: eyes: 0.24 ± 0.11, mouth: 0.38 ± 0.17; $t(7536) = 40.42$, $p < 10^{-30}$). However, again, recovering within the critical period led to an opposite pattern of information utilization: the eyes contributed more information than the mouth (see Figure 4C for an example; Figure 4D for group summary; Figure 4G; average Grad-CAM intensity: eyes: 122.19 ± 10.42, mouth: 91.48 ± 40.97; $t(7536) = 32.76$, $p < 10^{-30}$; Figure 4H; proportion of Grad-CAM intensity: eyes: 0.36 ± 0.13, mouth: 0.27 ± 0.15; $t(7536) = 27.48$, $p < 10^{-30}$), a pattern of results that was more similar to the eye-foveated model. This result again suggests that new complementary information provided during the critical period overrode the original information utilization, and the mouth-foveated model turned into the eye-foveated model. In addition, the model performance seemed to switch as well (Figures 4A and 4B; i.e., the mouth-foveated model turned into the eye-foveated model), and this result could be further replicated by a different DNN (Figures S2C and S2D).

Using complementary information for recovery, we not only confirm that providing new information outside of the critical period cannot alter the model anymore but also show that providing new information within the critical period will override the original model.

Computational mechanism underlying recovery from the critical period

We next investigated why there was a critical period and why the information provided in the critical period could override previous information. We hypothesize that the decrease in learning rate could explain the above results. This hypothesis is in line with the developmental trajectory of the primate visual system, where learning is decreased as a function of age.[34] It is worth noting that, during model training, the learning rate evolved based on an adaptive rule (Methods) that is consistent with neurodevelopment, and we did not preset the epoch-by-epoch learning rates (see also Figure S1A for validation with different initial learning rates).

Indeed, we showed that the learning rate monotonically dropped as a function of the training epoch (Figure 5A), suggesting that the learning became more local in the later stage, which is consistent with the idea of a critical period during development. Specifically, at epoch 80, the learning rate dropped below 0.001 (Figure 5A), which prevented the model from learning using full-face images (Figure 5B, illustrated using the mouth-foveated model; note that there was also an initial drop in performance due to a change in training stimuli from mouth-foveated images to full-face images; cf. Figure 3B). However, notably, when we restored a larger learning rate in later epochs, the learning process was recovered (Figure 5B). Interestingly, a smaller change in the learning rate (0.001) initially resulted in a smaller drop in performance compared with a larger change in the learning rate (0.01), although over time a larger change in the learning rate ultimately led to higher performance. Therefore, a larger learning rate could lead to a better recovery, and the same learning rate as the initial training phase (0.01, the learning rate in the critical period) could best recover the model (Figure 5B). In sum, the inability to recover outside of the critical period could be explained by the reduced learning rate: the network could not get out of the local minima to restore the learning for new information. This also explained why the network overrode the

Please cite this article in press as: Wang et al., A critical period for developing face recognition, Patterns (2023), https://doi.org/10.1016/j.patter.2023.100895

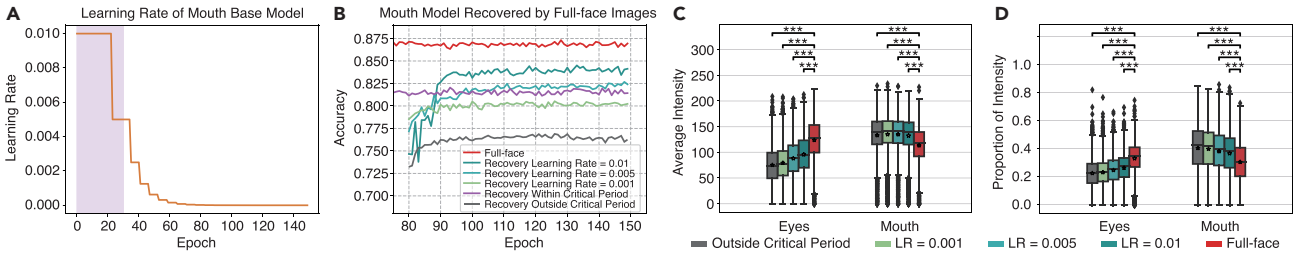

Figure 5. Improving model recovery by adjusting the learning rate

(A) Learning rate as a function of training epoch in the mouth-foveated model.

(B) The learning curve for each learning rate. Model recovery ability varied as a function of learning rate. The learning curves for the full-face model and the mouth-foveated model recovered by full-face images within and outside of the critical period are shown as a reference.

(C) Average Grad-CAM intensity for each learning rate.

(D) The proportion of Grad-CAM intensity for each learning rate (LR).

Legend conventions are as in Figure 2.

previously learned information for recovery within the critical period: the network presumably converged to another local minimum with the new information for learning.

Furthermore, we showed that, consistent with improved model performance, utilization of eye information increased as a function of learning rate toward the full-face model (Figure 5C; average Grad-CAM intensity: recovery outside of the critical period: 76.03 ± 35.50, learning rate = 0.001: 80.11 ± 35.13, learning rate = 0.005: 89.02 ± 36.87, learning rate = 0.01: 96.50 ± 37.92, full-face model: 124.47 ± 39.97; one-way repeated-measure ANOVA across learning rates: $F(2,7536) = 1{,}220$, $p < 10^{-20}$; Figure 5D; proportion of Grad-CAM intensity: recovery outside the critical period: 0.23 ± 0.10, learning rate = 0.001: 0.24 ± 0.10, learning rate = 0.005: 0.25 ± 0.10, learning rate = 0.01: 0.27 ± 0.11, full-face model: 0.34 ± 0.11; $F(2,7536) = 1{,}086$, $p < 10^{-20}$). Moreover, we derived similar results with the eye-foveated model.

Together, our results suggest that the learning rate is a determining factor for the critical period and can explain the network performance and information utilization concerning model recovery.

Knowledge distillation and attention transfer for model recovery

Can we achieve the same improvement in learning without modifying the learning rate but by applying knowledge distillation and attention transfer? Specifically, we used the full-face model as the teacher model and applied attention transfer to improve the mouth-foveated model outside of the critical period (i.e., the student model, which is the same as the recovery outside of the critical period; Figure 6A; methods). Indeed, attention transfer improved model performance outside of the critical period (Figure 6B) and increased information utilization in the eyes (see Figure 6C for an example and Figure 6D for group average; Figure 6E; average Grad-CAM intensity: no attention transfer [i.e., recovery outside of the critical period]: 76.03 ± 35.50, attention transferred: 89.49 ± 37.27; $t(3768) = 45.52$, $p < 10^{-30}$; Figure 6F; proportion of Grad-CAM intensity: no attention transfer: 0.23 ± 0.10, attention transferred: 0.25 ± 0.10; $t(3768) = 33.82$, $p < 10^{-30}$). However, the model after attention transferred still did not reach the same level of performance (Figure 6B; 80.76% vs. 87.32%; the performance of the attention transferred model was similar to recovery with a learning rate of 0.001 [80.53%]; Figure 5B) and utilization of eye information (Figure 6E; average Grad-CAM intensity: $t(3768) = 47.91$, $p < 10^{-30}$; Figure 6F; proportion of Grad-CAM intensity: $t(3768) = 39.01$, $p < 10^{-30}$) as the full-face model (teacher model). Furthermore, similar results were found for the eye-foveated model.

Together, our results suggest that knowledge distillation and attention transfer can partially recover an impaired model outside of the critical period (i.e., when the learning rate is low), although the extent of recovery is limited compared with directly adjusting the learning rate.

Identity selectivity

We have shown before that those identity-selective units (methods) are key building blocks of the DNN for face recognition.[25] We next investigated the change of identity selectivity during the development of face recognition by summarizing the percentage of identity-selective units in each model. We focused on the top DNN layer (Conv4), where identity selectivity is most established and relevant.[25,33] Indeed, we found that the full-face model (Figure 7; 87.5%) had a higher percentage of identity-selective units than the eye-foveated model (78.9%; χ^2 test: $p < 10^{-10}$) and a control forehead-foveated model (72.8%; $p < 10^{-10}$; note that the eye-foveated model also had a higher percentage of identity-selective units than the control forehead-foveated model: $p < 10^{-10}$). Importantly, for the eye-foveated model, recovery with full-face images within the critical period increased the percentage of identity-selective units (Figure 7; 84.2%; $p < 10^{-10}$), but recovery outside of the critical period did not increase the percentage of identity-selective units (79.1%; $p = 0.33$; within vs. outside: $p < 10^{-10}$). Similarly, for the forehead-foveated model, recovery with full-face images within the critical period increased the percentage of identity-selective units (Figure 7; 81.9%; $p < 10^{-10}$), although recovery outside of the critical period also increased the percentage of identity-selective units (76.3%; $p < 10^{-10}$), albeit to a lesser extent (within vs. outside: $p < 10^{-10}$). Together, our results further suggested a recovery mechanism using identity-selective units: restricted visual information impaired the formation of identity-selective units, and recovering within the critical period could increase and recover identity-selective units.

Please cite this article in press as: Wang et al., A critical period for developing face recognition, Patterns (2023), https://doi.org/10.1016/j.patter.2023.100895

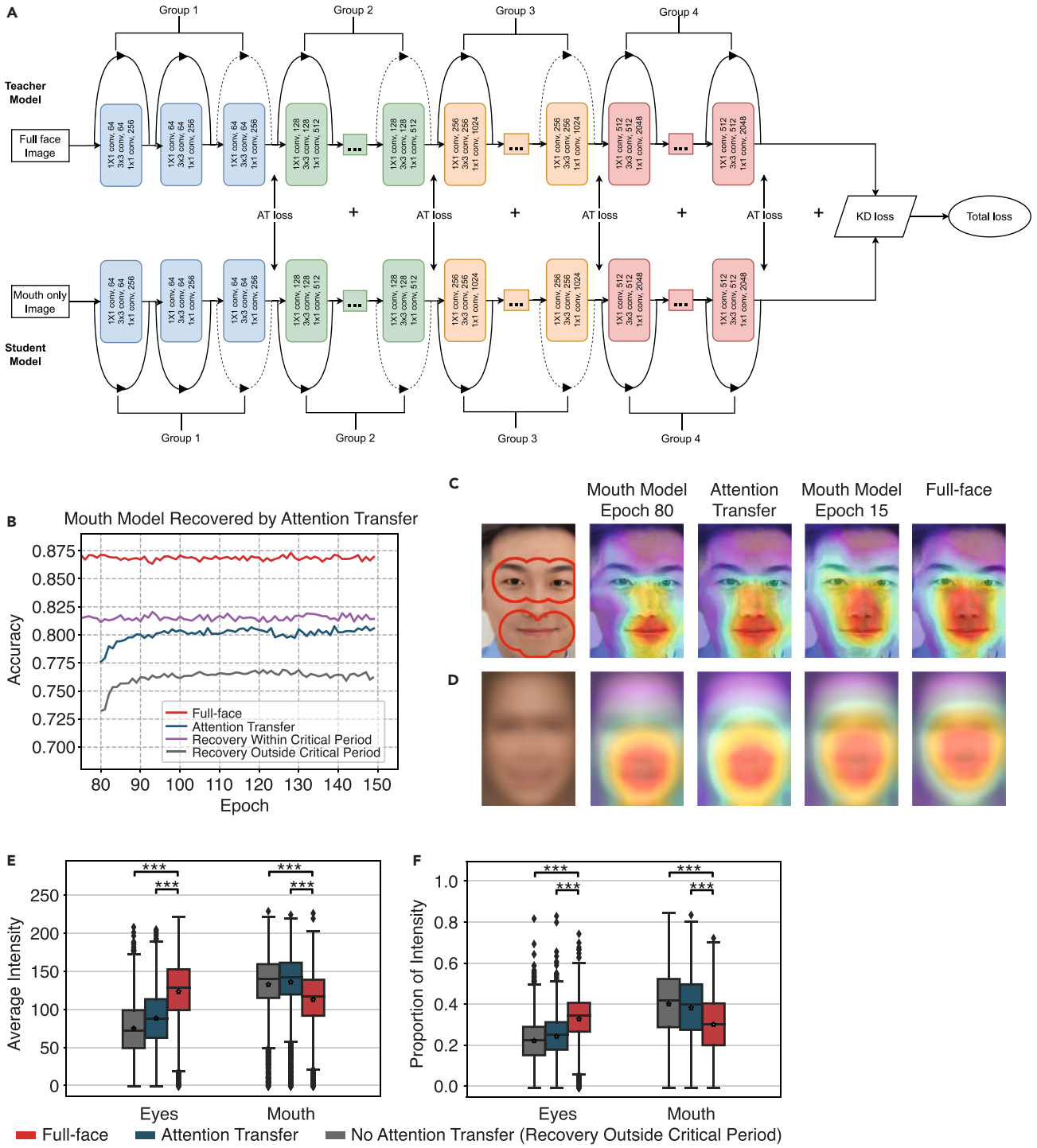

Figure 6. Improving model recovery by applying KD and AT

(A) A computational framework for KD and AT. We used the model trained by full-face images as the teacher model to improve the performance of the mouth-foveated model (student model) outside of the critical period. Both the teacher model and the student model had the same ResNet50 architecture.

(B) The learning curves for the mouth-foveated model with vs. without AT. The learning curves for the full-face model and the mouth-foveated model recovered by full-face images within and outside the critical period are shown as a reference.

(C) The Grad-CAM intensity maps for an example face.

(D) The Grad-CAM intensity maps for group average across faces.

(E) Average Grad-CAM intensity for each ROI.

(F) The proportion of Grad-CAM intensity for each ROI.

Legend conventions are as in Figure 2.

Please cite this article in press as: Wang et al., A critical period for developing face recognition, Patterns (2023), https://doi.org/10.1016/j.patter.2023.100895

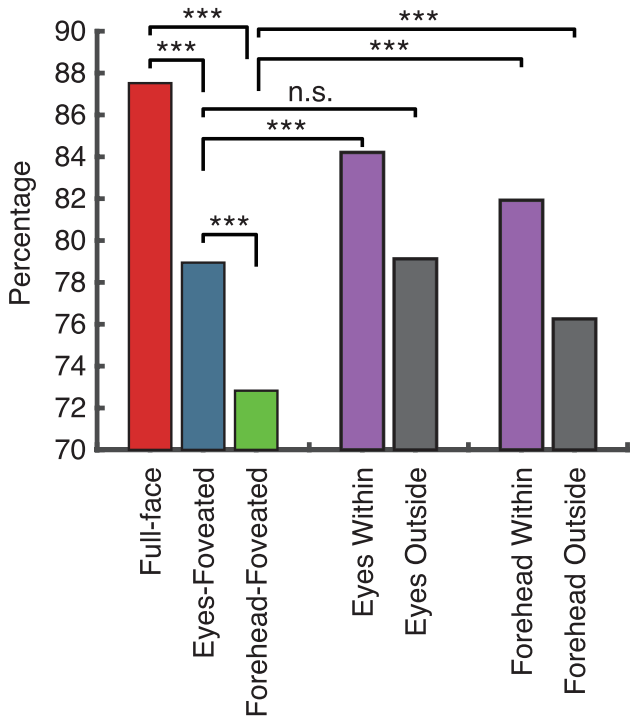

Figure 7. Percentage of identity-selective units for each model

Eyes within, eye-foveated model recovered by full-face images within the critical period; eyes outside, eye-foveated model recovered by full-face images outside of the critical period; forehead within, forehead-foveated model recovered by full-face images within the critical period; forehead outside, forehead-foveated model recovered by full-face images outside of the critical period. Asterisks indicate a significant difference between models using χ^2 test. ***$p < 0.001$; n.s., not significant.

Establishing the relationship between artificial DNN units and real primate neurons

The DNN performs face recognition tasks similarly to humans, and it has been suggested that DNNs share similarities with the primate visual system and can therefore help us better understand the neural mechanisms of face recognition[26,31,35] (see more under introduction). We finally investigated whether the development of face recognition in DNNs had a similar biological basis.

First, we analyzed whether the ensemble of DNN units shared representational similarity with the ensemble of monkey inferotemporal (IT) neurons (Figures 8A–8D). We used an independent set of stimuli from the CelebA dataset (500 natural face images of 50 celebrities)[36] to compare artificial DNN units and real IT neurons. We recorded neuronal activity using two Utah arrays in the anterior and central IT cortex (methods) while the monkey performed a passive viewing task (Figure 8A). We identified 53 multi-unit activity (MUA) channels that showed sufficient internal consistency, and we focused on these channels for further analysis. We found that, for all models, the pairwise distance from the DNN (methods) significantly correlated with the neuronal pairwise distance from the monkey IT cortex (Figure 8C; see Figure 8D for temporal dynamics of each model), and for the full-face model, there was an increase of correlation toward the top DNN layer (Figure 8C). Notably, we found that the full-face model had a better correspondence with IT neurons than the eye-foveated model and mouth-foveated model (Figure 8C). We also found that the eye-foveated model recovered with full-face images (Figure 3A) within the critical period increased correspondence with IT neurons compared with the recovery outside of the critical period (Conv4: permutation $p < 0.001$). Therefore, the correspondence between DNN units and real IT neurons could reflect model performance and recovery (cf. Figure 3A).

Second, we analyzed whether the ensemble of DNN units shares representational similarity with the ensemble of the human amygdala and hippocampal neurons (Figures 8B, 8E, and 8F). We used the same stimuli (500 natural face images of 50 celebrities) as for monkey recordings and recorded from 667 neurons in the human amygdala and hippocampus (340 neurons from the amygdala, 222 neurons from the anterior hippocampus, and 105 neurons from the posterior hippocampus; firing rate > 0.15 Hz) of 8 neurosurgical patients (23 sessions in total).[33] Patients performed a one-back task (Figure 8B), and they could well recognize the faces.[33] The responses of 76 of 667 neurons (11.39%) differed between different face identities in a window of 250–1,250 ms following stimulus onset, and these neurons were the real human identity-selective neurons. We grouped amygdala and hippocampal neurons as a single neuronal population (i.e., medial temporal lobe [MTL] neurons) for further analysis because they show very similar identity selectivity responses.[33,38] We found that the pairwise distance from the top DNN layer significantly correlated with the neuronal pairwise distance from the human MTL, consistent with the processing stage along the ventral visual pathway (Figure 8E; see Figure 8F for temporal dynamics of the full-face model). We also found that the full-face model had a better correspondence with MTL neurons than the eye-foveated model and mouth-foveated model (Figures 8E and 8F), and the mouth-foveated model recovered with full-face images within the critical period increased correspondence with MTL neurons compared with the recovery outside of the critical period (Conv2: permutation $p = 0.001$).

By comparing artificial units and real primate neurons, we not only revealed a systematic correspondence between the two face recognition systems but also showed that such correspondence was associated with DNN model performance and recovery.

DISCUSSION

In this study, we systematically investigated face learning and facial information utilization during a critical period. Specifically, we revealed a critical period during development that has the following properties. (1) Under the baseline condition, reduced facial information resulted in reduced model performance and subsequent inability to use information from the corresponding facial parts. (2) When full-face information was provided *within* the critical period, full recovery could be achieved, but recovery did not happen when full-face information was provided *outside of* the critical period. (3) When complementary information was provided within the critical period, it could even override the original model and become a model like that trained with new information alone. We further provided a computational account with a learning rate that could explain the properties of critical

Please cite this article in press as: Wang et al., A critical period for developing face recognition, Patterns (2023), https://doi.org/10.1016/j.patter.2023.100895

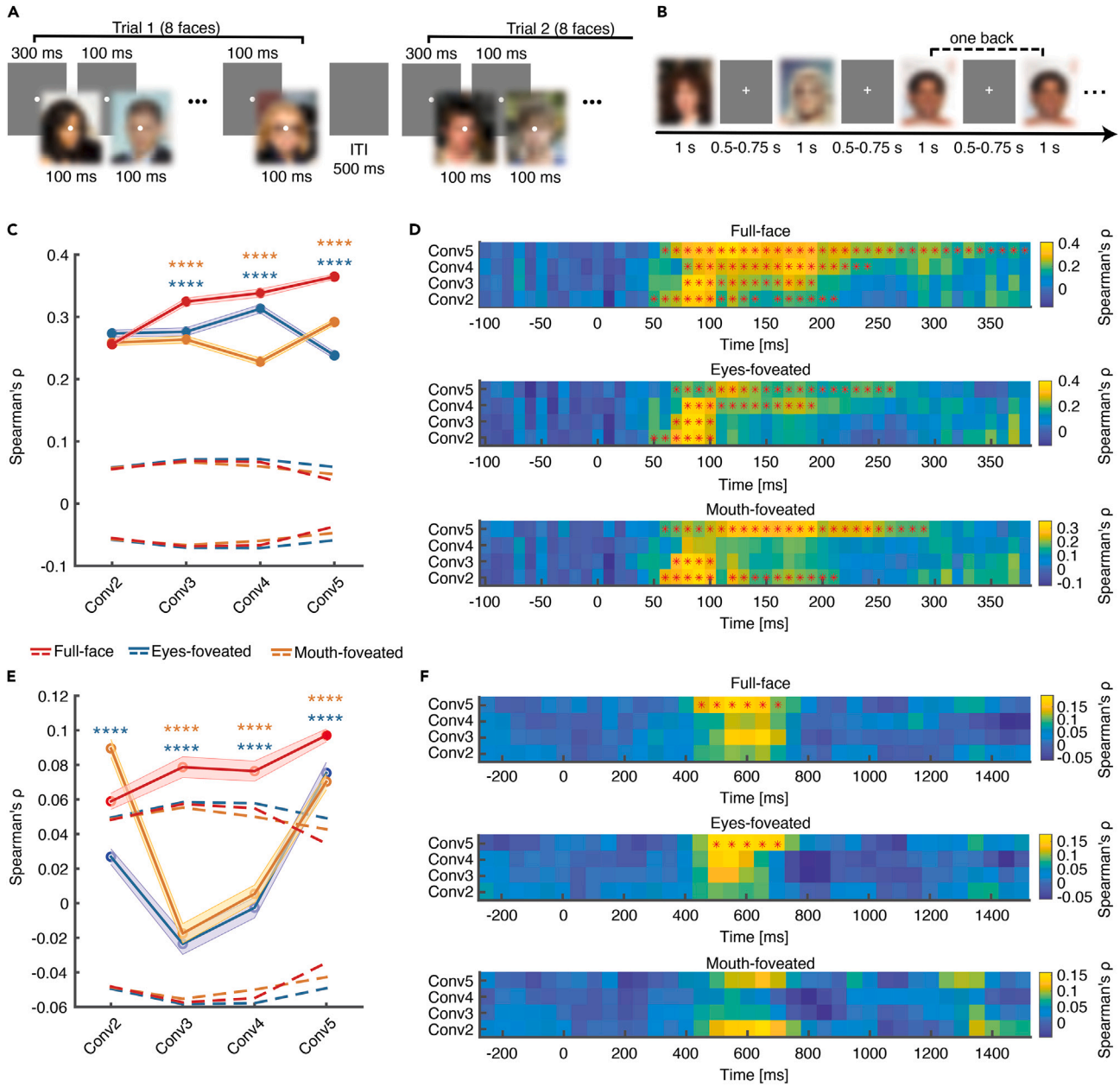

Figure 8. Match between the deep neural network (DNN) units and (real) primate neurons

(A, C, and D) Monkey inferotemporal (IT) cortical neurons.

(B, E, and F) Human amygdala and hippocampal neurons.

(A) Task used to acquire neural responses from a monkey. In each trial, 8 faces were presented for 100 ms each, followed by a fixed inter-stimulus-interval (ISI) of 100 ms. There was a central fixation point of 300 ms at the beginning of each trial, and there was an inter-trial-interval (ITI) of at least 500 ms following each trial. The central fixation point persisted through the trial.

(B) Task used to acquire single-neuron responses from humans. We employed a one-back task in which patients responded whenever an identical famous face was repeated. Each face was presented for 1 s, followed by a jittered ISI of 0.5–0.75 s. Face images are blurred for illustration purposes only.

(C and E) Correlation between pairwise distance in the primate neuronal face space and pairwise distance in the DNN face space.

(C) Here we used the mean firing rate in a time window of 70 ms–180 ms after stimulus onset as the response to each face, and we averaged the responses to 10 faces for each face identity.

(E) Here we used the mean firing rate in a time window of 250 ms–1000 ms after stimulus onset as the response to each face, and we averaged the responses to 10 faces for each face identity. Dashed lines denote ±SD across permutation runs (n = 1,000), and solid circles represent a significant correlation (permutation test: $p < 0.05$, Bonferroni correction across layers). The shaded area denotes ±SEM across bootstrap runs (n = 1,000; each resample contained 35 identities), and asterisks indicate a significant difference between models using one-tailed two-sample t test. ****$p < 0.001$.

(legend continued on next page)

Please cite this article in press as: Wang et al., A critical period for developing face recognition, Patterns (2023), https://doi.org/10.1016/j.patter.2023.100895

periods, and we showed that, by altering the learning rate, learning could be recovered. We also demonstrated an alternative approach (i.e., knowledge distillation and attention transfer) that can partially recover the model outside of the critical period. Finally, we showed that model performance and recovery were associated with identity-selective units as well as the correspondence with the primate visual systems. Together, our results not only highlight the importance of a critical period in face learning but also elucidate its underlying computational mechanism and restoration strategies.

Our present findings are consistent with the neurodevelopment concepts. First, brains have developmental critical periods. Classical studies have documented critical periods affecting a range of species and systems, from visual acuity in kittens[39,40] to song learning in birds.[41] Uncorrected eye defects (e.g., strabismus, cataracts) during the critical period for visual development lead to amblyopia.[4,5] Second, our results show that, outside of the critical period, the learning process could not be restored, which is likely accounted for by the restricted learning rate. In analogy, brain rewiring is significantly reduced after the critical period,[3] which may cause a reduced learning rate. On the other hand, pathways have been discovered in animal models through which critical periods may be re-opened in adults, making it possible to re-awaken a brain's youth-like plasticity and, thus, repair brain injury, recover sensory deficits, and treat neurodevelopmental disorders.[42] Third, network attention transfer could partially improve model performance, which is analogous to learning after development. Our recent study has revealed a neural mechanism underlying face learning and shown that the neuronal population geometry in the human amygdala and hippocampus, quantified by the representational distance, encodes face familiarity, similarity, and learning. Specifically, the neuronal separation between different face identities expands with increased exposure, suggesting that faces become more distinctly represented in the neural network following the learning process.[19] Together, through a series of experiments in an artificial neural network, our present results implicate the importance of the critical period during model training/learning, which is consistent with the neurobiology of animal development.

Developmental prosopagnosia (DP) is an impairment in recognizing faces despite normal vision, intelligence, and socio-cognitive abilities and no history of brain damage.[43] The impaired development of face processing during the critical period may lead to prosopagnosia.[44] In this study, we demonstrated that information provided during the critical period determines subsequent information utilization, which is also consistent with the development of visual attention.[45] Recent arguments suggest that neural coding strategies during development may exhibit high levels of dynamism.[16] Additionally, research has shown that early emotional processing in young children differs from that observed in adolescents, who are more adultlike.[46] One hypothesis is that, in TD, the brain undergoes a process of specialization for face processing where specific regions such as the FFA become dedicated to face recognition. It is believed that, during the critical period, particular experiences and interactions with faces shape and refine the neural circuits involved in face processing. In the case of DP, it is proposed that disruptions or abnormalities may occur during this process of face specialization in the critical period. Consequently, this can lead to impaired development of the neural circuits involved in face recognition, resulting in persistent difficulties with recognizing faces throughout an individual's life.

Our present results are relevant to neurodevelopmental disorders such as ASD. Many studies have documented abnormal face processing in people with ASD,[14,15,47–52] and such a deficit has both a developmental[53] and genetic[54] root. In particular, people with ASD demonstrate impaired utilization of facial information. During viewing naturalistic social videos, people with autism demonstrate abnormal patterns of social visual pursuit that are consistent with reduced saliency of eyes and increased saliency of mouths, bodies, and objects.[49] When viewing static faces, people with autism view non-feature areas of the faces significantly more often but core feature areas of the faces (e.g., eyes and mouth) significantly less often than controls,[50] and they have piecemeal rather than configural strategies.[55] Similarly, some research suggests that people with ASD demonstrate active avoidance of fixating the eyes in faces, which, in turn, influences the recognition performance of emotions,[48] whereas other research suggests that children with ASD demonstrate gaze indifference and passive insensitivity to social signals in others' eyes at the time of initial diagnosis.[56] The atypical facial fixations are complemented by neuronal evidence of abnormal processing of information from the eye region of faces in blood-oxygen-level-dependent (BOLD) fMRI[57] and single-neuron responses in the amygdala.[58]

Our present study provided a possible computational account for such a deficit in ASD: reduced access to eye information during the critical period resulted in impaired utilization of eye information and, thus, gaze to the eyes after development. Therefore, our results point to a potential way to recover from such face processing deficits by early training with guided fixation onto the eyes. On the other hand, although our results suggest that recovery outside of the critical period could not restore normal function, the network attention transfer has provided an important alternative to recover learning, which is also consistent with the behavioral training strategy currently being applied in ASD. It is worth noting that, although our results highlight the importance of critical periods, a future study is needed to understand whether such deficits in ASD are the cause or consequence of a critical period. In addition, using human single-neuron recordings, it has been shown that neurons in the human amygdala and hippocampus encode facial features (e.g., the eyes and mouth) and eye movement to these facial features,[59] which, in turn, may be related to abnormal facial feature representation in ASD.[58,60] A future study is needed to directly investigate the neuronal mechanisms for face learning concerning critical periods.

(D and F) Temporal dynamics of correlation of pairwise distance between primate neurons and DNN units.
(D) Monkey IT neurons (bin size = 40 ms, step size = 10 ms).
(F) Human MTL neurons (bin size = 500 ms, step size = 50 ms). Color coding indicates the Spearman's correlation coefficient. Asterisks indicate a significant correlation in that bin (permutation test: *$p < 0.05$, false discovery rate [FDR][37] corrected across time bins for each layer).

Please cite this article in press as: Wang et al., A critical period for developing face recognition, Patterns (2023), https://doi.org/10.1016/j.patter.2023.100895

DNNs currently provide the most compelling quantitative models of the response patterns of neurons throughout the primate ventral visual stream based on their predictive power, biological plausibility, generalization ability, and performance comparisons. In this study, we used DNNs as surrogate models that can serve as accurate representations of many aspects of face learning. It is worth noting that our current simulation approach with surrogate computational models can be generalized to other sensorimotor domains (e.g., auditory), and our findings were not restricted to the model or visualization method used but could be replicated with other models and visualization methods. It has also been shown that DNNs trained with unsupervised contrastive embedding can well simulate infant learning during development.[26] Interestingly, such unsupervised learning produces brain-like representations even when trained solely with real human child developmental data collected from head-mounted cameras. In addition, our recent study has developed a computational model that illustrates an increase in the representational distance of artificial units with training, aligning with neuronal findings.[19]

On the other hand, artificial neural network models can have a biological correspondence with both human and non-human primate neurons, which has been illustrated in face recognition.[25,33] In this study, we compared typically developed primate neurons with artificial units at various stages of face learning. The results demonstrated that the full-face model showed a stronger correspondence with IT and MTL neurons compared with the foveated models. Additionally, we observed that foveated models, when trained with full-face images during the critical period, displayed increased correspondence with primate neurons compared with models trained outside of the critical period. Consequently, these results imply that our artificial DNN unit modeling holds strong biological relevance and is well suited for understanding real brain processes. The systematic correspondence between the two face recognition systems allows us to gain a deeper understanding of the computational mechanisms involved in the development of face perception.

Conclusions

This study used DNNs as surrogate models to explore the critical period in face processing development. DNNs, like humans and animals, exhibit critical periods where temporary stimulus deficits impair learning. Comparisons were made between DNN computations and monkey/human single-neuron recordings. Key findings include the following: (1) revealing the critical period and its properties, such as reduced performance with limited facial information and the importance of timing for recovery; (2) providing a computational account with a learning rate explaining critical period properties and demonstrating the role of identity-selective DNN units in recovery; (3) illustrating learning restoration approaches, including adjusting the learning rate and employing knowledge distillation and attention transfer; and (4) establishing correspondence between artificial and human/monkey neuron responses. This systematic investigation highlights the critical period's importance, clarifies its computational mechanism and restoration strategies, and sheds light on brain development. It contributes to computational modeling of the critical period in face processing, with implications for understanding ASD etiology.

EXPERIMENTAL PROCEDURES

Resource availability

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Shuo Wang (shuowang@wustl.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

All data and statistical analysis code are available on Zenodo (https://doi.org/10.5281/zenodo.10014797).[61]

Methods

Training and testing data

We used a subset of images from the CASIA_WebFace dataset as the training and testing data.[62] The CASIA-WebFace dataset has been used for various face verification and identification tasks. The dataset contains 494,414 face images of 10,575 real identities collected from the internet. In this study, we selected images from 50 identities that have more than 400 images in the dataset. The identities were diverse in race, sex, and age. Because the image quality varies, we manually removed the images that have a low resolution, major facial occlusion, or extreme facial angles. As a result, our training and testing dataset contained 37,000 images from 50 different identities, with each identity having more than 300 image samples.

Image processing

We applied the face and facial landmark detection tool Multi-Task Cascaded Convolutional Neural Networks (MTCNN)[63] to crop the faces from the images and label the facial landmarks. MTCNN is a framework developed as a solution for both face detection and face alignment. It is one of the most popular and accurate face-detection tools. The process consists of three stages of convolutional neural networks (CNNs). It uses a shallow CNN as the first step to produce candidate windows quickly. Through a more intricate CNN, it improves the suggested candidate windows in the second step. To further refine the outcome and output face landmark positions, a third CNN that is more complicated than the others is used in the third step. After applying the MTCNN, we derived a tight bounding box outlining the face area as well as the coordinates for the centers of the eyes, nose tip, and two corners of the mouth.

We next applied foveated imaging to generate images that mimic human foveation/fixation (i.e., the spatial resolution is highest at the point of the fovea and drops rapidly away from that point as a function of eccentricity, and thus the region around the point of fixation [or foveation point] is sampled with the highest intensity and perceived with the highest sensitivity; Figure 1). Foveated imaging is a method of digital image processing where the level of detail or resolution varies across the image according to one or more fixation points. We utilized the open-source Python implementation of image retina transformation for foveated imaging (https://github.com/ouyangzhibo/Image_Foveation_Python) and produced two groups of foveated images (eye-foveated and mouth-foveated) based on the facial landmarks identified by the MTCNN (Figure 1). Specifically, in eye-foveated images, only the eye region was clear, and the rest of the image was blurry, whereas in mouth-foveated images, only the mouth region was clear, and the rest of the image was blurry.

The locations of the eyes and mouth were detected by the MTCNN, which, in turn, determined the size of the eye and mouth regions. It is worth noting that we did not set a region of interest for the eyes or mouth, and foveated imaging was based on the center of the facial landmarks of the eyes or mouth. Because different images had slightly different locations of the detected facial landmarks, and we used the same parameters for foveated imaging across images, there might be slight differences in the content of the eye or mouth region under foveation. We used the default parameters from the foveated imaging toolbox, which are suitable for the faces detected by the MTCNN. Additionally, we manually verified each foveated image to ensure its quality.

We finally cropped the images based on the bounding box derived using the MTCNN (Figure 1). All subsequent analyses were based on the cropped images.

Model training and testing

We used the well-known DNN implementation based on the ResNet50[64] CNN architecture (see details in Figure 2A). Because the goal of the present study is

to understand model performance during training, we trained the network from scratch.

We first trained three base models, using full-face images, eye-foveated images, and mouth-foveated images. For each model, 80% of the images were used as the training set, and the remaining 20% of the images were used as the testing/validation set. We used the stochastic gradient descent (SGD) optimizer with an initial learning rate of 10^{-2}, and all models were trained for 150 epochs. An adaptive learning rate scheduler was applied, which halved the current learning rate when the loss of validation did not drop for 5 epochs. To update the weights, we computed the cross-entropy loss on random batches of 32 images (scaled to 224 × 224 pixels) for back-propagation. We derived similar results using different initial learning rates (Figure S1A).

We next trained recovery models based on base models by providing base models with different information (i.e., same images with different foveation) during their training. We implemented the recovery at different stages of model training, and the model training only continued with the new set of images. It is worth noting that, to facilitate a direct comparison between base models and recovery models, we used the same parameters for the recovery models as the corresponding base models, including the epoch-by-epoch learning rate. The recovery models were trained with the new set of images until they reached 150 epochs. For instance, if the base model was recovered at epoch 15, then the recovery model would continue the training using the learning rate at epoch 15 and inherit all the subsequent learning rates from the base model. The recovery model would continue the training for another 135 epochs to have a total of 150 training epochs.

To compare different models, we always used the same set of original images (not foveated) to test all models.

Critical period

The critical period is a time window of early post-natal development during which sensory deficits can lead to permanent skill impairment.[1] Similar to humans and other animals, deep artificial neural networks exhibit critical periods during which a temporary stimulus deficit can impair the model's performance.[21] In this study, we defined the critical period of the DNN as the epochs of the early fast learning phase (following the same definition as in Achille et al.[21]). Specifically, based on the learning curves of the base models, the first 30 epochs were defined as the critical period. We thus chose epoch 15 and epoch 80 to compare recovery models starting within vs. outside of the critical period, and we derived similar results using other epochs (e.g., epoch 10 vs. epoch 90) to compare recovery models (Figures S1B and S1C).

Model visualization and quantification

In our experiments, we adopted the Grad-CAM[65] as our visualization tool. Grad-CAM is a popular technique for visualizing which regions in the original image contribute to the final output. It uses the gradients of the target category flowing into a certain convolution layer, usually the last one, to produce a coarse localization map highlighting the important pixels/regions in the image for predicting the category. This approach reveals the implicit attention of the model to make the real contributor of features in the input image distinguishable. Grad-CAM is an improvement from the previous approach, CAM,[66] for both versatility and accuracy.

We further quantified Grad-CAM intensity in the eye and mouth regions of interest (ROIs) (Figure 2B). We defined the eye and mouth ROIs in the image based on the facial landmarks for each image. It is worth noting that the eye ROI and the mouth ROI were of similar size across images (eyes: 4,894.76 ± 1,668.83 pixels, mouth: 4,843.66 ± 2,110.94; two-tailed two-sample t test: $t(7536) = 1.17$, $p = 0.24$). Because most of the nose region was covered by both the eyes and mouth ROIs, we did not separately analyze the Grad-CAM intensity for the nose region. In addition to the average Grad-CAM intensity in each ROI, we also calculated the proportion of total intensity in each ROI by dividing the total Grad-CAM intensity of an image.

Knowledge distillation and attention transfer

The basic idea behind knowledge distillation (KD) is to train a small, lightweight model using supervised information from a larger model with superior performance to improve its performance. It was first proposed by Hinton et al.[67] in 2015. The large model is known as the teacher model, while the small model is characterized as the student model. The supervised information from the output of the teacher model is called "knowledge," and the process of student learning to migrate the supervised information from the teacher is called "distillation." Our recovery experiment, in contrast to the original KD concept, is built upon two identical architectures (Figure 6A). The gap in model performance was mainly reflected in the different stimuli. One model was trained by the full-face images, which were regarded as the teacher model (Figure 6A, top), whereas the other model was trained by the mouth-foveated images, which were regarded as the student model (Figure 6A, bottom). Our purpose here was to guide the mouth-foveated model to learn new features with information from the full-face model when the mouth-foveated model had already missed the critical period (i.e., recovery started at epoch 80, which is outside of the critical period; note that this is the same model for recovery outside of the critical period).

DNN models can barely learn new features when they have passed the critical period, especially when the learning rate becomes extremely low. To reinforce the recovery effect, we used another technique, attention transfer (AT),[68] which can work together with KD. We used the average feature map of each group of convolutional layers as the attention and transferred the attention from the teacher model to the student model. It is worth noting that only the student model was updated during the process, while the teacher model acted as a supervisor, enabling the student model to learn from its information, and, as a result, all weights in the teacher model were frozen.

We made the learning rate of the student model identical to that of the foveated model to determine whether the KD-AT method could effectively aid in the recovery of the original foveated model under extremely low learning rates. We computed the AT loss after each convolution group using the following loss function:

$$L_{AT} = L(W_S, x) + \sum_{j \in I} \| \frac{Q_S^j}{\| Q_S^j \|_2} - \frac{Q_T^j}{\| Q_T^j \|_2} \|_2$$

where $L(W_S, x)$ denotes the standard cross-entropy loss, and I denotes the indices of all teacher-student activation layer pairs. $Q_S^j = vec(Avg(A_S^j))$ and $Q_T^j = vec(Avg(A_T^j))$ are the j-th pair of student and teacher attention maps in vectorized form, respectively, and the attention map is the cross-channel average of the activation tensor A.

Finally, we added the KD loss between the output of the teacher model (y_t) and the student model (y_s) to the previous loss function. As a result, the final total loss was obtained as follows:

$$L_{AT} = L(W_S, x) + \sum_{j \in I} \| \frac{Q_S^j}{\| Q_S^j \|_2} - \frac{Q_T^j}{\| Q_T^j \|_2} \|_2 + L_{KD}(y_t, y_s)$$

With this total loss function, we aimed to enable the student model to not only make correct predictions but also to learn similar feature representations as the teacher model.

Selection of identity-selective DNN units and primate neurons

To select identity-selective units,[25] we used a one-way ANOVA to identify identity-selective units that had a significantly unequal response to different identities ($p < 0.01$). We further imposed an additional criterion to identify a subset of identity-selective units with selective identities; the response of identity was 2 standard deviations (SDs) above the mean of responses from all identities. These identified identities whose response stood out from the global mean were the encoded identities.

We followed the same selection procedure as for primate neurons.[25,33] We used the mean firing rate in a time window of 250–1,000 ms after stimulus onset as the response to each face for primate neurons. Note that we also used this response to study the correlation between DNN units and primate neurons.

Neural recordings from a monkey

The detailed procedure has been described in our previous study.[25] Briefly, we recorded from the anterior and central IT cortex in one male rhesus macaque (*Macaca mulatta*) using two Utah arrays (Blackrock Microsystems) (see Kar et al.[32] and Kar and DiCarlo[32,69] for details). We detected the multiunit spikes after the raw data were zero-phase band-pass filtered between 300 and 6,000 Hz (MATLAB ellip function, fourth order with 0.1-dB pass-band ripple and 40-dB stop-band attenuation), and we used MUA for analyses. To test with an independent dataset, the monkey passively viewed 500 images from

the CelebA dataset.[36] In each trial, the monkey first viewed a white central fixation point (0.2 degrees of visual angle [DVAs]) on a gray background for 300 ms to initiate a trial. Then, 8 faces were presented for 100 ms each, each followed by a blank (gray) screen for an inter-stimulus interval (ISI) of 100 ms. The central fixation point persisted throughout the trial, and a fluid reward was given when the monkey successfully fixated through the entire trial. The inter-trial interval (ITI) of the blank gray screen was at least 500 ms. We recorded 4,155 trials in total, and we rejected 666 trials where the monkey broke the fixation (±2 DVAs). For each round of presentation, we generated a random sequence for the 500 faces, and we used different sequences for different rounds of presentation. All procedures conformed to local and US National Institutes of Health guidelines, including the US National Institutes of Health Guide for the Care and Use of Laboratory Animals. All experiments were performed with the approval of the MIT Institutional Animal Care and Use Committee (IACUC).

Neural recordings from human neurosurgical patients

The detailed procedure has been described in our previous study.[25,70] Briefly, we recorded from implanted depth electrodes in the amygdala and hippocampus from 8 neurosurgical patients (23 sessions in total) with pharmacologically intractable epilepsy. Bipolar wide-band recordings (0.1–9,000 Hz), using one of the eight microwires as a reference, were sampled at 32 kHz and stored continuously for offline analysis with a Neuralynx system. The raw signal was filtered with a zero-phase-lag 300- to 3,000-Hz band-pass filter, and spikes were sorted using a semi-automatic template-matching algorithm as described previously.[71] Units were carefully isolated, and recording and spike sorting quality were assessed quantitatively. Only units with an average firing rate of at least 0.15 Hz (entire task) were considered. Only single units were considered. Trials were aligned to stimulus onset, and we used the mean firing rate in a time window of 250 ms–1000 ms after stimulus onset as the response to each face. We employed a one-back task with the same 500 CelebA images as for monkey recordings. In each trial, a single face was presented at the center of the screen for a fixed duration of 1 s, with uniformly jittered ITI of 0.5–0.75 s. Patients pressed a button when the present face image was identical to the immediately previous image. All participants provided written informed consent using procedures approved by the Institutional Review Board of West Virginia University (WVU).

Match between DNN units and primate neurons

We employed a pairwise distance metric[31] to compare the neural coding of face identities between primate neurons and DNN units. For each pair of identities, we used the dissimilarity value (1 – Pearson's r)[72] as a distance metric. The primate neuronal distance metric was calculated between firing rates of all recorded neurons, and the DNN distance metric was calculated between activation of all DNN units. We then correlated the primate neuronal distance metric and the DNN distance metric. To determine statistical significance, we used a non-parametric permutation test with 1,000 runs. In each run, we randomly shuffled the face labels and calculated the correlation between the primate neuronal distance metric and the DNN distance metric. The distribution of correlation coefficients computed *with* shuffling (i.e., null distribution) was eventually compared with the one *without* shuffling (i.e., observed response). If the correlation coefficient of the observed response was greater than 95% of the correlation coefficients from the null distribution, then it was considered significant. A significant correlation indicated that the DNN face space corresponded to the primate neuronal face space.[31] We computed the correlation for each DNN layer so that we could determine the specific layer that the neuronal population encoded. For each face identity, we averaged the response of all faces of that identity to get a single mean firing rate.

To get temporal dynamics, for human neurons, we used a moving window with a bin size of 500 ms and a step size of 50 ms (given the sparseness of human MTL neurons, this time window is commonly used[33,60,73]). The first bin started at −300 ms relative to trial onset (the bin center was thus 50 ms before trial onset), and we tested 19 consecutive bins (the last bin was thus from 600–1,100 ms after trial onset). For monkey neurons, we used a moving window with a bin size of 40 ms and a step size of 10 ms. The first bin started at −70 ms relative to stimulus onset (the bin center was thus 50 ms before stimulus onset), and we tested 26 consecutive bins (the last bin was thus from 180–220 ms after stimulus onset). We used Bonferroni correction to correct for multiple comparisons across DNN layers and false discovery rate (FDR)[37] to correct for multiple comparisons across time bins.

We used a bootstrap with 1,000 runs to compare between models (full-face vs. eye-foveated and full-face vs. mouth-foveated). In each run, data from 70% of the identities (i.e., 35 identities) were randomly selected to calculate the correspondence between DNN units and primate neurons. We thus created a distribution of correspondence for each model.

We further used a permutation test with 1,000 runs to statistically compare the correspondence for recovery within vs. outside of the critical period. In each run, we shuffled the recovery labels (within vs. outside) and calculated the difference in correspondence between recoveries. We then compared the observed difference in correspondence between recoveries with the permuted null distribution to derive statistical significance.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.patter.2023.100895.

ACKNOWLEDGMENTS

We thank all patients for their participation, staff from WVU Ruby Memorial Hospital for support with patient testing, and Alina Peter and James DiCarlo for providing the monkey physiology data. This research was supported by the McDonnell Center for Systems Neuroscience, AFOSR (FA9550-21-1-0088), NSF (BCS-1945230, IIS-2114644), and NIH (R01MH129426). The funders had no role in study design, data collection, and analysis; decision to publish; or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

J.W., P.N.C., X.L., and S.W. designed research. J.W. and R.C. performed experiments. J.W., R.C., and P.N.C. analyzed data. J.W., P.N.C., X.L., and S.W. wrote the paper. All authors discussed the results and contributed to the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 17, 2023
Revised: November 9, 2023
Accepted: November 14, 2023
Published: December 26, 2023

REFERENCES

1. Berardi, N., Pizzorusso, T., and Maffei, L. (2000). Critical periods during sensory development. Curr. Opin. Neurobiol. *10*, 138–145.
2. Birdsong, D. (1999). Second Language Acquisition and the Critical Period Hypothesis (Routledge).
3. Hensch, T.K. (2005). Critical period plasticity in local cortical circuits. Nat. Rev. Neurosci. *6*, 877–888.
4. Daw, N.W. (1998). Critical Periods and Amblyopia. Arch. Ophthalmol. *116*, 502–505.
5. Hensch, T.K., and Quinlan, E.M. (2018). Critical periods in amblyopia. Vis. Neurosci. *35*, E014.
6. Geldart, S., Mondloch, C.J., Maurer, D., De Schonen, S., and Brent, H.P. (2002). The effect of early visual deprivation on the development of face processing. Dev. Sci. *5*, 490–501.
7. Hensch, T.K. (2004). Critical period regulation. Annu. Rev. Neurosci. *27*, 549–579.
8. McKone, E., Wan, L., Pidcock, M., Crookes, K., Reynolds, K., Dawel, A., Kidd, E., and Fiorentini, C. (2019). A critical period for faces: Other-race face recognition is improved by childhood but not adult social contact. Sci. Rep. *9*, 12820.

9. Parr, L.A. (2011). The evolution of face processing in primates. Philos. Trans. R. Soc. Lond. B Biol. Sci. *366*, 1764–1777.
10. Pascalis, O., Fort, M., and Quinn, P.C. (2020). Development of face processing: are there critical or sensitive periods? Current Opinion in Behavioral Sciences *36*, 7–12.
11. Pierce, K., Müller, R.A., Ambrose, J., Allen, G., and Courchesne, E. (2001). Face processing occurs outside the fusiform 'face area' in autism: evidence from functional MRI. Brain *124*, 2059–2073.
12. Röder, B., Ley, P., Shenoy, B.H., Kekunnaya, R., and Bottari, D. (2013). Sensitive periods for the functional specialization of the neural system for human face processing. Proc. Natl. Acad. Sci. USA *110*, 16760–16765.
13. Sugita, Y. (2009). Innate face processing. Curr. Opin. Neurobiol. *19*, 39–44.
14. Spezio, M.L., Adolphs, R., Hurley, R.S.E., and Piven, J. (2007). Analysis of face gaze in autism using "Bubbles. Neuropsychologia *45*, 144–151.
15. Neumann, D., Spezio, M.L., Piven, J., and Adolphs, R. (2006). Looking you in the mouth: abnormal gaze in autism resulting from impaired top-down modulation of visual attention. Soc. Cognit. Affect Neurosci. *1*, 194–202.
16. Avitan, L., and Goodhill, G.J. (2018). Code Under Construction: Neural Coding Over Development. Trends Neurosci. *41*, 599–609.
17. Golarai, G., Ghahremani, D.G., Whitfield-Gabrieli, S., Reiss, A., Eberhardt, J.L., Gabrieli, J.D.E., and Grill-Spector, K. (2007). Differential development of high-level visual cortex correlates with category-specific recognition memory. Nat. Neurosci. *10*, 512–522.
18. Gomez, J., Barnett, M.A., Natu, V., Mezer, A., Palomero-Gallagher, N., Weiner, K.S., Amunts, K., Zilles, K., and Grill-Spector, K. (2017). Microstructural proliferation in human cortex is coupled with the development of face processing. Science *355*, 68–71.
19. Cao, R., Wang, J., Brunner, P., Willie, J., Li, X., Rutishauser, U., Brandmeir, N., and Wang, S. (2023). Neural mechanisms of face familiarity and learning in the human amygdala and hippocampus. Cell Reports *42*, 113520.
20. Golarai, G., Grill-Spector, K., and Reiss, A.L. (2006). Autism and the development of face processing. Clin. Neurosci. Res. *6*, 145–160.
21. Achille, A., Rovere, M., and Soatto, S. (2019). Critical Learning Periods in Deep Neural Networks (ICLR).
22. Parkhi, O., Vedaldi, A., and Zisserman, A. (2019). Deep Face Recognition (British Machine Vision Association).
23. Schroff, F., Kalenichenko, D., and Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. Preprint at bioRxiv, 815–823.
24. O'Toole, A.J., Castillo, C.D., Parde, C.J., Hill, M.Q., and Chellappa, R. (2018). Face Space Representations in Deep Convolutional Neural Networks. Trends Cognit. Sci. *22*, 794–809.
25. Wang, J., Cao, R., Brandmeir, N.J., Li, X., and Wang, S. (2022). Face identity coding in the deep neural network and primate brain. Commun. Biol. *5*, 611.
26. Zhuang, C., Yan, S., Nayebi, A., Schrimpf, M., Frank, M.C., DiCarlo, J.J., and Yamins, D.L.K. (2021). Unsupervised neural network models of the ventral visual stream. Proc. Natl. Acad. Sci. USA *118*, e2014196118.
27. Grill-Spector, K., Weiner, K.S., Gomez, J., Stigliani, A., and Natu, V.S. (2018). The functional neuroanatomy of face perception: from brain measurements to deep neural networks. Interface Focus *8*, 20180013.
28. Hill, M.Q., Parde, C.J., Castillo, C.D., Colón, Y.I., Ranjan, R., Chen, J.-C., Blanz, V., and O'Toole, A.J. (2019). Deep convolutional neural networks in the face of caricature. Nat. Mach. Intell. *1*, 522–529.
29. Dobs, K., Martinez, J., Kell, A.J.E., and Kanwisher, N. (2022). Brain-like functional specialization emerges spontaneously in deep neural networks. Sci. Adv. *8*, eabl8913.
30. Yamins, D.L.K., 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. Proc. Natl. Acad. Sci. USA *111*, 8619–8624.
31. Grossman, S., Gaziv, G., Yeagle, E.M., Harel, M., Mégevand, P., Groppe, D.M., Khuvis, S., Herrero, J.L., Irani, M., Mehta, A.D., and Malach, R. (2019). Convergent evolution of face spaces across human face-selective neuronal groups and deep convolutional networks. Nat. Commun. *10*, 4934.
32. Kar, K., Kubilius, J., Schmidt, K., Issa, E.B., and DiCarlo, J.J. (2019). Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. Nat. Neurosci. *22*, 974–983.
33. Cao, R., Wang, J., Lin, C., Rutishauser, U., Todorov, A., Li, X., Brandmeir, N., and Wang, S. (2020). Feature-based encoding of face identity by single neurons in the human medial temporal lobe. Preprint at bioRxiv.
34. Deng, W., Aimone, J.B., and Gage, F.H. (2010). New neurons and new memories: how does adult hippocampal neurogenesis affect learning and memory? Nat. Rev. Neurosci. *11*, 339–350.
35. Yamins, D.L.K., and DiCarlo, J.J. (2016). Using goal-driven deep learning models to understand sensory cortex. Nat. Neurosci. *19*, 356–365.
36. Liu, Z., Luo, P., Wang, X., and Tang, X. (2015). Deep Learning Face Attributes in the Wild. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 3730–3738.
37. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. Roy. Stat. Soc. B *57*, 289–300.
38. Quiroga, R.Q., Reddy, L., Kreiman, G., Koch, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. Nature *435*, 1102–1107.
39. Wiesel, T.N., and Hubel, D.H. (1963). Effects of visual deprivation on morphology and physiology of cells in the cat's lateral geniculate body. J. Neurophysiol. *26*, 978–993.
40. Wiesel, T.N. (1982). Postnatal development of the visual cortex and the influence of environment. Nature *299*, 583–591.
41. Konishi, M. (1985). Birdsong: from behavior to neuron. Annu. Rev. Neurosci. *8*, 125–170.
42. Hensch, T.K., and Bilimoria, P.M. (2012). Re-opening Windows: Manipulating Critical Periods for Brain Development (Dana Foundation).
43. Grill-Spector, K., Weiner, K.S., Kay, K., and Gomez, J. (2017). The Functional Neuroanatomy of Human Face Perception. Annu. Rev. Vis. Sci. *3*, 167–196.
44. Rivolta, D. (2014). Cognitive and Neural Aspects of Face Processing. In Prosopagnosia: When all faces look the same (Springer Berlin Heidelberg)), pp. 19–40.
45. Amso, D., and Scerif, G. (2015). The attentive brain: insights from developmental cognitive neuroscience. Nat. Rev. Neurosci. *16*, 606–619.
46. Batty, M., and Taylor, M.J. (2006). The development of emotional face processing during childhood. Dev. Sci. *9*, 207–220.
47. Adolphs, R., Sears, L., and Piven, J. (2001). Abnormal Processing of Social Information from Faces in Autism. J. Cognit. Neurosci. *13*, 232–240.
48. Kliemann, D., Dziobek, I., Hatri, A., Steimke, R., and Heekeren, H.R. (2010). Atypical Reflexive Gaze Patterns on Emotional Faces in Autism Spectrum Disorders. J. Neurosci. *30*, 12281–12287.
49. Klin, A., Jones, W., Schultz, R., Volkmar, F., and Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. Arch. Gen. Psychiatr. *59*, 809–816.
50. Pelphrey, K.A., Sasson, N.J., Reznick, J.S., Paul, G., Goldman, B.D., and Piven, J. (2002). Visual Scanning of Faces in Autism. J. Autism Dev. Disord. *32*, 249–261.
51. Spezio, M.L., Adolphs, R., Hurley, R.S.E., and Piven, J. (2007). Abnormal Use of Facial Information in High-Functioning Autism. J. Autism Dev. Disord. *37*, 929–939.
52. Wang, S., and Adolphs, R. (2017). Reduced specificity in emotion judgment in people with autism spectrum disorder. Neuropsychologia *99*, 286–295.
53. Jones, W., and Klin, A. (2013). Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism. Nature *504*, 427–431.

54. Constantino, J.N., Kennon-McGill, S., Weichselbaum, C., Marrus, N., Haider, A., Glowinski, A.L., Gillespie, S., Klaiman, C., Klin, A., and Jones, W. (2017). Infant viewing of social scenes is under genetic control and is atypical in autism. Nature *547*, 340–344.

55. Dawson, G., Webb, S.J., and McPartland, J. (2005). Understanding the Nature of Face Processing Impairment in Autism: Insights From Behavioral and Electrophysiological Studies. Dev. Neuropsychol. *27*, 403–424.

56. Moriuchi, J.M., Klin, A., and Jones, W. (2017). Mechanisms of Diminished Attention to Eyes in Autism. Am. J. Psychiatr. *174*, 26–35.

57. Kliemann, D., Dziobek, I., Hatri, A., Baudewig, J., and Heekeren, H.R. (2012). The Role of the Amygdala in Atypical Gaze on Emotional Faces in Autism Spectrum Disorders. J. Neurosci. *32*, 9469–9476.

58. Rutishauser, U., Tudusciuc, O., Wang, S., Mamelak, A.N., Ross, I.B., and Adolphs, R. (2013). Single-Neuron Correlates of Atypical Face Processing in Autism. Neuron *80*, 887–899.

59. Cao, R., Li, X., Brandmeir, N.J., and Wang, S. (2021). Encoding of facial features by single neurons in the human amygdala and hippocampus. Commun. Biol. *4*, 1394.

60. Cao, R., Lin, C., Hodge, J., Li, X., Todorov, A., Brandmeir, N.J., and Wang, S. (2022). A neuronal social trait space for first impressions in the human amygdala and hippocampus. Mol. Psychiatr. *27*, 3501–3509.

61. Wang, J., Cao, R., Chakravarthula, P.N., Li, X., and Wang, S. (2023). Data release for: A critical period for developing face recognition. Zenodo.

62. Yi, D., Lei, Z., Liao, S., and Li, S.Z. (2014). Learning face representation from scratch. Preprint at arXiv.

63. Zhang, K., Zhang, Z., Li, Z., and Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Process. Lett. *23*, 1499–1503.

64. He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778.

65. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 618–626.

66. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., and Torralba, A. (2016). Learning deep features for discriminative localization. Preprint at arXiv, 2921–2929.

67. Hinton, G., Vinyals, O., and Dean, J. (2015). Distilling the knowledge in a neural network. Preprint at arXiv.

68. Zagoruyko, S., and Komodakis, N. (2017). Paying More Attention to Attention: Improving the Performance of Convolutional Neural Networks via Attention Transfer (ICLR).

69. Kar, K., and DiCarlo, J.J. (2021). Fast Recurrent Processing via Ventrolateral Prefrontal Cortex Is Needed by the Primate Ventral Stream for Robust Core Visual Object Recognition. Neuron *109*, 164–176.e5.

70. Cao, R., Lin, C., Brandmeir, N.J., and Wang, S. (2022). A human single-neuron dataset for face perception. Sci. Data *9*, 365.

71. Rutishauser, U., Schuman, E.M., and Mamelak, A.N. (2006). Online detection and sorting of extracellularly recorded action potentials in human medial temporal lobe recordings, in vivo. J. Neurosci. Methods *154*, 204–224.

72. Kriegeskorte, N., Mur, M., and Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. Front. Syst. Neurosci. *2*, 4.

73. Wang, S., Yu, R., Tyszka, J.M., Zhen, S., Kovach, C., Sun, S., Huang, Y., Hurlemann, R., Ross, I.B., Chung, J.M., et al. (2017). The human amygdala parametrically encodes the intensity of specific facial emotions and their categorical ambiguity. Nat. Commun. *8*, 14821.