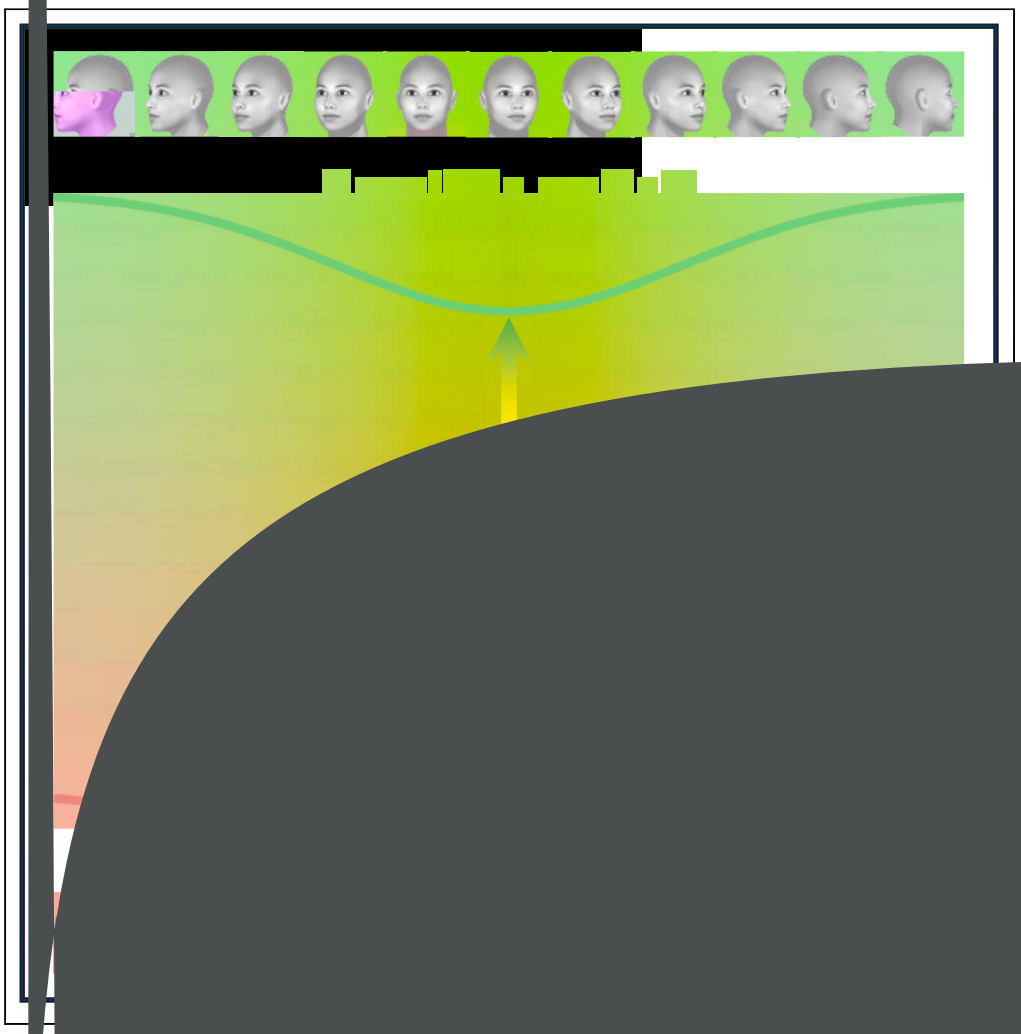
iScience

CellPress
OPEN ACCESS

Article

Profiles of visual perceptual learning in feature space



Shiqi Shen, Yueling Sun, Jiachen Lu, ..., Ce Mo, Fang Fang, Xilin Zhang

xlzhang@m.scnu.edu.cn

Highlights

Profiles of VPL in feature space depend on the visual feature hierarchy

iScience

Article

Profiles of visual perceptual learning in feature space

Shiqi Shen,[1,2,7] Yueling Sun,[1,2,7] Jiachen Lu,[1,2,7] Chu Li,[1,2,7] Qinglin Chen,[1,2] Ce Mo,[3] Fang Fang,[4,5,6] and Xilin Zhang[1,2,8,*]

SUMMARY

Visual perceptual learning (VPL), experience-induced gains in discriminating visual features, has been studied extensively and intensively for many years, its profile in feature space, however, remains unclear. Here, human subjects were trained to perform either a simple low-level feature (grating orientation) or a complex high-level object (face view) discrimination task over a long-time course. During, immediately after, and one month after training, all results showed that in feature space VPL in grating orientation discrimination was a center-surround profile; VPL in face view discrimination, however, was a monotonic gradient profile. Importantly, these two profiles can be emerged by a deep convolutional neural network with a modified AlexNet consisted of 7 and 12 layers, respectively. Altogether, our study reveals for the first time a feature hierarchy-dependent profile of VPL in feature space, placing a necessary constraint on our understanding of the neural computation of VPL.

INTRODUCTION

Visual perceptual learning (VPL), a long-term improvement in visual performance through practices or trainings, has been demonstrated in the detection or discrimination of various stimuli, ranging from simple low-level features to complex high-level objects.[1–12] One of the central questions in VPL is its specificity and generalization (transfer), which have profound implications for the underlying neural mechanisms.[13–15] Indeed, the specificity and generalization has inspired various models and theories that interpret VPL as a result from training not only induced tuning curve plasticity of neurons in the task-relevant sensory areas[16–22] but also improved readout of sensory signals through response reweighting within either visual cortex[23–27] or higher decision areas.[28–31] It is likely, therefore, that VPL reflects plasticity in a complex set of brain networks and may occur at multiple levels (for reviews[2,7,10]).

The specificity of what is learned is a fundamental and prominent property of VPL, in which learned improvements are confined to the particular trained visual attributes, such as the orientation of the trained stimulus.[18,19,27,32–45] However, a number of previous studies have also indicated that VPL can significantly, and almost completely, generalize to the untrained visual attributes and this generalization depends on several factors,[2,46] such as the difficulty[32,47–49] and the processing level[50] of the task, the duration[51] and the state of induced adaptation[52] of the training, the precision demand[53] and the exact procedure (i.e., the double-training paradigm[48,54,55]) of the transfer task, the categorization between the trained and untrained stimuli,[56] and the feature hierarchy (simple low-level versus complex high-level) of the trained stimulus.[1,42,57] Although for several decades VPL has been regarded as a distinct format of learning as its specificity, the generalization of VPL is more important in practical applications.

Previous literature on visual attention have indicated a structured manner regarding how attention demarcates the target of interest from various distractors, either a center-surround profile[58,59] or a monotonic gradient profile.[60,61] Intriguingly, VPL faces the same situation that demarcates the trained visual attribute (specificity) from various untrained visual attributes (generalization), and therefore, an important question in this regard is whether and how these profiles are at play within the VPL. This issue is particularly important since such learning profile in feature space could offer us a unique opportunity to give insight into the whole picture of VPL, thereby furthering our understanding of the neural mechanism underlying VPL and how the visual system adapts to its changing environment. However, the specificity and generalization of VPL are usually assessed by comparing the trained condition versus another or a few untrained conditions, we therefore still know little about the profile of VPL in feature space.

[1]Key Laboratory of Brain, Cognition and Education Sciences, Ministry of Education, South China Normal University, Guangzhou, Guangdong 510631, China
[2]School of Psychology, Center for Studies of Psychological Application, and Guangdong Provincial Key Laboratory of Mental Health and Cognitive Science, South China Normal University, Guangzhou, Guangdong 510631, China
[3]Department of Psychology, Sun-YatSen University, Guangzhou, Guangdong 510275, China
[4]School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing 100871, China
[5]IDG/McGovern Institute for Brain Research, Peking University, Beijing 100871, China
[6]Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China
[7]These authors contributed equally
[8]Lead contact
*Correspondence: xlzhang@m.scnu.edu.cn
https://doi.org/10.1016/j.isci.2024.109128

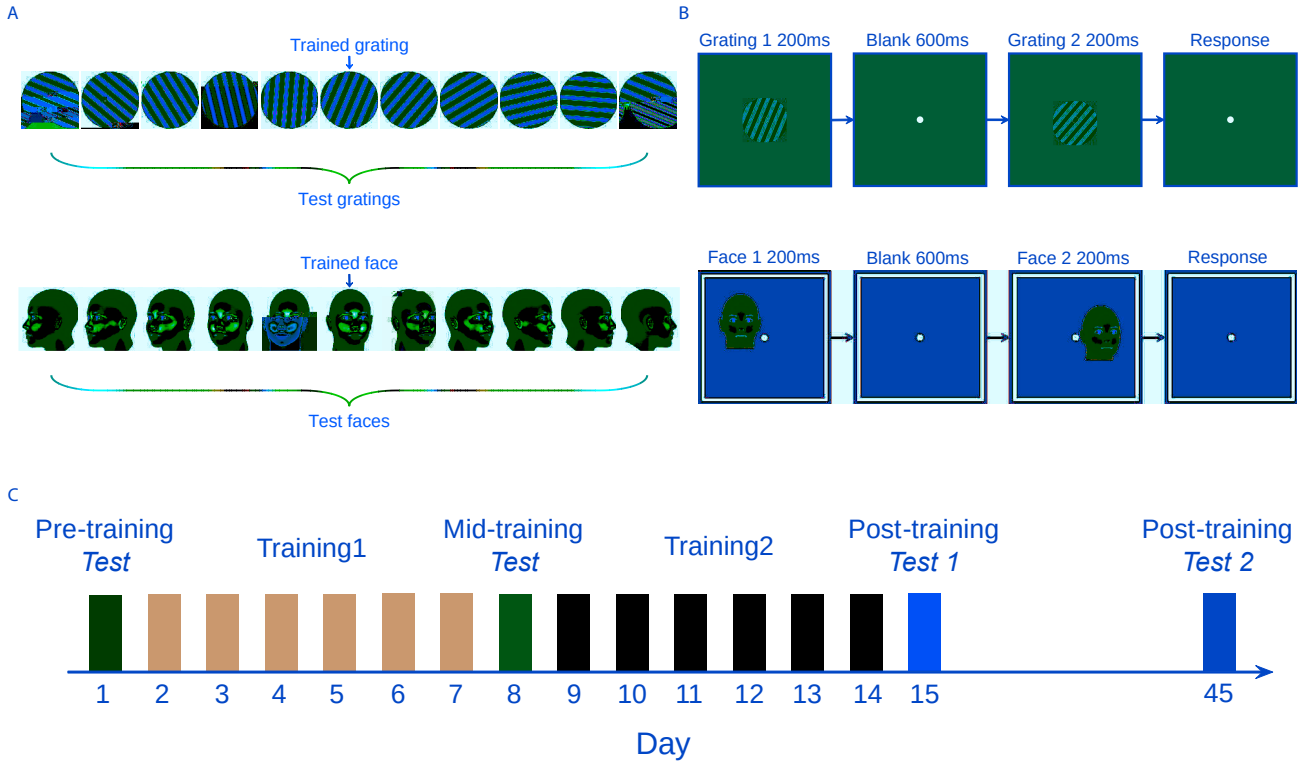


Figure 1. Stimuli and Psychophysical Protocol

(A) Exemplar gratings (up) and faces (bottom) in the grating orientation discrimination (GOD) and face view discrimination (FVD) task, respectively. For both tasks, there were six possible distances in feature space between the trained and test stimuli, ranging from Δ0° through Δ90°, with a step size of 18°.

(B) Schematic description of a two-alternative forced-choice (2-AFC) trial in a QUEST staircase for measuring grating orientation (up) or face view (bottom) discrimination thresholds.

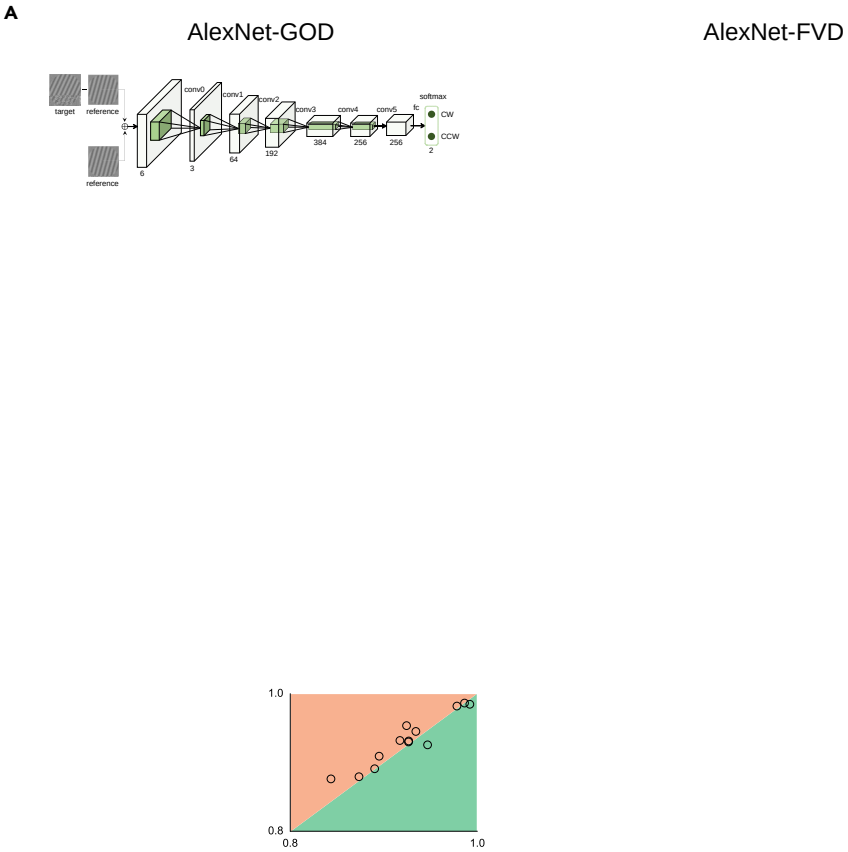
(C) Experimental protocol. Both GOD and FVD tasks consisted of six phases – pre-training test (Pre), discrimination-training 1 (Training1), mid-training test (Mid), discrimination-training 2 (Training2), post-training test 1 (Post1), and post-training test 2 (Post2). Pre, Mid, Post1, and Post2 took place on the days before, during, immediately after and one month after training, respectively. During the two training phases (Training1 and Training2), each subject underwent six daily training sessions.

Besides, deep convolutional neural networks (DCNN) have shown impressive correspondences to various behaviors and neural responses from early to higher visual areas.[62,63] This brain-like hierarchical system provides new ways of studying VPL from behavior to physiology.[64] Indeed, using various artificial neural networks, previous studies have reproduced both experimental and theoretical analyses that resembled predictions of the reverse hierarchy theory[65] of VPL,[66] replicated relative performances of training conditions within a wide range of behavioral data,[67] and emerged both the specificity[68] and generalization[69] of VPL. To date, whether and how the DCNNs can appropriately model the underlying profile of VPL remain unexplored.

To address these issues, here human subjects were trained to perform either a simple low-level feature (grating orientation) or a complex high-level object (face view) discrimination task over a long-time course. For both tasks, we manipulated the distance in feature space between the trained and test stimuli, ranging from Δ0° through Δ90° with a step size of 18°, to measure the profile of VPL (Figure 1A). Unexpectedly, during, immediately after and one month after training, all results confirmed that in feature space, VPL in grating orientation discrimination was a center-surround profile (Figure 2); VPL in face view discrimination, however, was a monotonic gradient profile (Figure 3). More importantly, both profiles can be reproduced by DCNNs qualitatively (Figure 4). Our results reveal for the first time a visual feature hierarchy-dependent profile of VPL in feature space, thereby placing a necessary constraint on our understanding of the neural computation underlying VPL.

RESULTS

Subjects in our study were trained to perform either the grating orientation discrimination (GOD) or face view discrimination (FVD) task. Each task consisted of six possible distances in feature space between the trained and test stimuli, ranging from Δ0° through Δ90°, with a step size of 18° (Figure 1A). For both tasks, there were six phases – pre-training test (Pre), discrimination-training 1 (Training1), mid-training test (Mid), discrimination-training 2 (Training2), post-training test 1 (Post1), and post-training test 2 (Post2). Pre, Mid, Post1, and Post2 took place on the days before, during, immediately after and one month after training, respectively (Figure 1C). During the two training phases (Training1 and



$t_{14} > 3.608$, $p < 0.043$, Cohen's $d > 1.863$; Post2: all $t_{14} > 3.522$, $p < 0.051$, Cohen's $d > 1.819$) and FVD (Mid: all $t_{14} > 4.387$, $p < 0.009$, Cohen's $d > 2.265$; Post1: all $t_{14} > 4.409$, $p < 0.009$, Cohen's $d > 2.277$; Post2: all $t_{14} > 4.029$, $p < 0.019$, Cohen's $d > 2.081$, except for $\Delta 0^\circ$ vs. $\Delta 54^\circ$: $t_{14} = 2.972$, $p = 0.152$, Cohen's $d = 1.545$) tasks. These results supported both the specificity and generalization of VPL, whereby the learning effect was the greatest for the trained stimulus and significantly transferred to the other untrained stimuli, respectively.

Gaussian and Mexican-hat models fitting and comparison

To further assess the shape of these learning effects, we fitted a monotonic model and a nonmonotonic model to the average learning effect across distances in both GOD and FVD tasks. The monotonic and nonmonotonic models were implemented as the Gaussian and Mexican-hat functions, respectively.[61] To compare these two models to our data, we first computed the Akaike information criterion (AIC)[71] and Bayesian information criterion (BIC)[72] with the assumption of a normal error distribution. Then, we calculated the Likelihood ratio (LR) and Bayes factor (BF) of the Gaussian model over the Mexican-hat model based on AIC[73] and BIC[74] approximation, respectively. Results showed that, at Mid, Post1, and Post2, all the LR/BFs were smaller than 1 (Table 1, left) and therefore favored the Mexican-hat model over the Gaussian model for the GOD task (Figure 2C). Notably, we also conducted similar model comparisons for each subject's data and found that the Mexican-hat model was favored over the Gaussian model in 13, 10, and 12 of 15 subjects, at Mid, Post1, and Post2, respectively (Figure 2D). For the FVD task, however, all the LR/BFs were larger than 1 (Table 1, right) and therefore favored the Gaussian model over the Mexican-hat model

Table 1. LR/BF of model comparisons

	GOD			FVD		
	Mid	Post1	Post2	Mid	Post1	Post2
G over M	8.83×10^{-4}	8.05×10^{-2}	3.80×10^{-2}	2.83×10^{3}	3.11×10^{9}	1.18×10^{4}

G, Gaussian model; M, Mexican-hat model; GOD, grating orientation discrimination; FVD, face view discrimination.

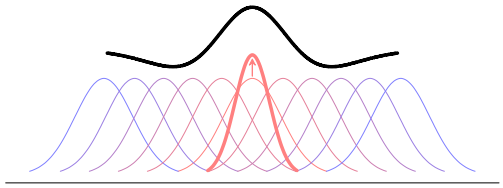
(Figure 3C). The model comparison based on fitting individual data also demonstrated that the Gaussian model was favored over the Mexican-hat model in 10, 9, and 9 of 15 subjects, at Mid, Post1, and Post2, respectively (Figure 3D). Together, these results constituted strong evidence for the center-surround and monotonic gradient profiles of VPL in simple low-level feature and complex high-level object discriminations, respectively. However, it could be argued that the feature hierarchy-dependent profile of VPL might be explained by pre-existing differences that were equalized through learning between GOD and FVD tasks. To address this issue, for the Pre, using a mixed ANOVA with task (GOD and FVD) as the between-subjects factor and distance (Δ0°, Δ18°, Δ36°, Δ54°, Δ72°, and Δ90°) as the within-subjects factor, we compared the discrimination threshold of each distance between two tasks. Results argue against this explanation by showing that the main effect of distance ($F_{5, 140} = 2.198$, $p = 0.069$, $\eta_p^2 = 0.073$) and the interaction ($F_{5, 140} = 2.406$, $p = 0.050$, $\eta_p^2 = 0.079$) between these two factors was (marginally) significant, but the main effect of task ($F_{1, 28} = 0.479$, $p = 0.495$, $\eta_p^2 = 0.017$) was extremely insignificant. Post hoc paired *t* tests (Bonferroni-corrected) further showed that the discrimination threshold of GOD task was significantly higher than that of FVD task for Δ0° ($t_{28} = 3.141$, $p = 0.004$, Cohen's $d = 1.187$), but not for Δ18°, Δ36°, Δ54°, Δ72°, or Δ90° (all $t_{28} < 0.735$, $p > 0.469$, Cohen's $d < 0.278$). Note that this difference in Δ0° could have an influence on the peak (trained stimulus) difference of VPL between two tasks, but not on their profile differences.

Deep convolutional neural network for the profile of VPL

Our results demonstrate the center-surround and monotonic gradient profiles of VPL in simple low-level feature and complex high-level object discriminations, respectively, yet it remains unclear whether DCNN models could emerge these profiles. Here we trained two DCNN models: AlexNet-GOD and AlexNet-FVD, modified from AlexNet[68,75] to perform our GOD and FVD tasks, respectively. AlexNet-GOD consisted of 6 convolutional (conv) layers and 1 fully connected (fc) layer, whereas AlexNet-FVD consisted of 9 conv layers and 3 fc layers (Figure 4). Note that these architects were built to mimic our hypothesis of the visual pathways involved in these two tasks. During the training, layers 2–6 of both the AlexNet-GOD and AlexNet-FVD are initialized with the weights of the first 5 layers of the pre-trained AlexNet, and the other weights are initialized randomly. The last layer of each network was trained to capture the difference between the target and reference and finally obtain the classification by softmax, to model decision making in our 2AFC paradigm (Figure 1B), in which subjects were asked to make a 2AFC judgment of the orientation in GOD task or the view in FVD task of the second stimulus (target) relative to the first one (reference). Parallel to our psychophysical experiments, both the AlexNet-GOD and AlexNet-FVD were independently trained 15 times, and for each distance (Δ0°, Δ18°, Δ36°, Δ54°, Δ72°, and Δ90°), the training effect was defined as the accuracy difference between the pre- and post-training. Similar to our psychophysical results (Table 1), the LR/BF of the AlexNet-GOD training (Gaussian model over Mexican-hat model: 5.729×10^{-1}) was smaller than 1, and therefore favored the Mexican-hat model over the Gaussian model (Figure 4C, left). The model comparison based on fitting individual data advocated that the Mexican-hat model was favored over the Gaussian model in 13 of 15 training data (Figure 4C, right). Besides, across subjects, a non-parametric Wilcoxon signed-rank test was conducted to compare the R^2 of two models, and results significantly advocated the Mexican-hat model over the Gaussian model ($z = 2.215$, $p = 0.013$, $r = 0.572$). Conversely, the LR/BF of the AlexNet-FVD training (Gaussian model over Mexican-hat model: 1.250×10^{8}) was larger than 1, and therefore favored the Gaussian model over the Mexican-hat model (Figure 4D, left). The model comparison based on fitting individual data demonstrated that the Gaussian model was favored over the Mexican-hat model in 14 of 15 training data (Figure 4D, right). Similarly, the non-parametric Wilcoxon signed-rank test again advocated that the Gaussian model was significantly favored over the Mexican-hat model ($z = 2.385$, $p = 0.008$, $r = 0.616$). More importantly, to further conform the effectivity of our constructed DCNN models (AlexNet-GOD and AlexNet-FVD), we performed cross-validation across tasks, i.e., AlexNet-GOD and AlexNet-FVD for face views and grating orientations, respectively. Results showed that, across subjects, for AlexNet-GOD of face views, there was no significant difference in the R^2 between Gaussian and Mexican-hat models (non-parametric Wilcoxon signed-rank test: $z = 0.284$, $p = 0.402$, $r = 0.073$); for AlexNet-FVD of grating orientations, the Gaussian model was significantly favored over the Mexican-hat model (non-parametric Wilcoxon signed-rank test: $z = 2.783$, $p = 0.002$, $r = 0.719$), further supporting our DCNN models' potential to perform brain-like representation. Together, these results further conform the center-surround and monotonic gradient profiles of VPL in simple low-level feature and complex high-level object discriminations, respectively.

DISCUSSION

The present results reveal a feature hierarchy-dependent (simple low-level versus complex high-level) profile of VPL in feature space. Specifically, we found that VPL in grating orientation (simple low-level) discrimination was a center-surround profile, with the maximum learning effect of the trained-orientation and suppressed learning effects of orientations similar to the trained orientation relative to orientations more distinct from it (Figure 2C). VPL in face view (complex high-level) discrimination, however, was a monotonic gradient profile with the



learning effect falling off gradually as the rotated similarity from the trained view (Figure 3C). For both the grating orientation and face view tasks, given the consistent results across during, immediately after and one month after training, their distinct profiles thus cannot be explained by either the undertraining or overtraining of VPL, and showed a classical persistency. However, it might be argued that their distinct profiles could be derived from an attentional selection mechanism by which aspects of information is prioritized over others thereby guiding VPL and behavior. That is to say, subjects may naturally pay different attention to various test stimuli based on their feature similarities away from the trained stimuli, thus yielding various profiles of VPL in feature space. Indeed, previous studies have suggested that attention plays a critical role in both the specificity and generalization of VPL[4,10,12,65,76–88] and could display as either a center-surround profile[58,59] or a monotonic gradient profile.[60,61] However, it is important to note that, in our study, for each test stimulus, subjects performed the same discrimination task at threshold, measured by the QUEST staircase procedure (75% correct),[70] which could maximally (although not completely) control the difference in task difficulty or, presumably, attention, among the distances. In addition, in our study, the gratings were varied in 2D space, whereas the faces were generated by projecting 3D models rotated in depth into a 2D plane (Figure 1A). Thus, one could argue that the distinct profile of VPL in grating orientation and face view discriminations was derived from the difference between 2D and 3D rather than that between simple low-level and complex high-level features. In other words, our study may not demonstrate a feature hierarchy-dependent profile of VPL, but instead, it reveals the distinct profile of VPL between 2D and 3D. To address this argument, we carried out a supplemental experiment (n =*10*), which were identical to the FVD experiment except for using the wire-like objects[89] (Figure S1). The wire-like objects were constructed using very simple bars and generated by projecting 3D models rotated in depth into a 2D plane also, thus offering an excellent opportunity to separate the influence on the profile of VPL between feature hierarchy and feature dimensionality. If the profile of VPL is modulated by feature hierarchies, then the profile of wire-like object learning would similar with that of the grating orientation learning (i.e., the center-surround profile); however, if the profile of VPL is modulated by feature dimensionalities, then the profile of wire-like object learning would similar with that of the face view learning (i.e., the monotonic gradient profile). Our results argue against the feature dimensionality explanation by showing a center-surround profile of VPL for the wire-like object.

The center-surround profile of VPL evident here provides the first behavioral evidence supporting that VPL of grating orientation could involve the simultaneous operation of neural enhancement and neural suppression in feature space that may optimize internal noise reduction during VPL.[2,24,25] That is, the center-surround profile represents an activity distribution in orientation space that is optimal to demarcate the trained from untrained orientations, specifically attenuating inputs from nearby distractors that would be at the largest risk to confuse trained orientation discrimination processes. How does VPL induce an inhibitory zone surrounding the orientation of the trained grating? We proposed that this center-surround profile could be derived from simultaneously increasing the gain and sharpening the tuning of neurons toward the trained orientation (Gain & tuning hypothesis, Figure 5, left). Although speculative, our hypothesis is consistent with a large number of previous studies and theories that interpret VPL as a result of training induced tuning curve plasticity of neurons in the task-relevant sensory areas.[16–22,90–92] More importantly, these hypothesized changes – amplification and sharpening of tuning curves – are also consistent with previous neurophysiological[20,79,90,93,94] studies, which have reported that the tuning curves in early and midlevel visual areas, such as V1 and V4, are sharpened and amplified during the VPL of orientation. Given those studies have identified amplification and sharpening of tuning curves as the neural basis of VPL,[17,21,92] we thus believed that the center-surround profile of VPL in orientation space could also be accounted by this same mechanism.

For the face view discrimination, however, we proposed that its monotonic gradient profile could be accounted by the gain enhancement of neurons toward the trained view only, i.e., the amplification of tuning curves (Gain-only hypothesis, Figure 5, right). Notably, although local features in faces could provide more or less information about face view, VPL might be more reliable to extract the view from the configural information, especially when the face views were randomly presented in a small area.[14,35,95] We thus believed that the face view training in our study improved the ability of computing face view from the configural information of face views, rather than the configural information itself or face parts. Indeed, previous studies have demonstrated that face view learning takes place at cortical areas containing neurons sensitive to face view but tolerant to face size, local information, position, and identity changes.[95] Such neurons have been indicated in monkey inferior temporal[96] and superior temporal sulcus (STS)[97] areas, as well as human fusiform face area, occipital face area, and STS.[98] Intriguingly, several

neurophysiological[99] and brain imaging[35,100,101] studies indeed found that face learning significantly enhanced neuronal responses in these areas. More importantly, the monotonic tuning functions of face views in these areas have also been demonstrated in previous neurophysiological[97,102] and brain imaging[103,104] studies, as well as the artificial neural network.[105] Compared to VPL of grating orientation, which involves the simultaneous operation of neural enhancement and neural suppression in feature space, face view training is a simple monotonic gradient that contains an excitatory peak but without a narrow inhibitory zone in view space. We speculate that this manner may be optimal to transfer the learning effect from the trained view to untrained views, specifically for inputs from nearby of the trained view. Crucially, our speculation has been supported by previous studies. Parallel to low-level vision, the effect of face training was also specific to the trained set of faces but showed a higher degree of generalization than low-level vision, which is in accordance with the low-level feature-invariant face representation in high-level visual cortex.[1,95,101,106–108]

Although we proposed that the feature hierarchy-dependent profile of VPL could be accounted by distinct tuning curve plasticity of sensory neurons (Figure 5), on the one hand, lacking directly evidence with neurophysiological techniques or ultrahigh field fMRI; on the other hand, we cannot deny a potential contribution from other cognitive processes, such as decision, action selection, top-down task relevance, and processing of feedback. Indeed, many studies and models have suggested that VPL, despite the feature hierarchy (simple low-level versus complex high-level) of the trained stimuli, is a complex process that occurs within a complex set of brain networks and might be the result of plasticity at multiple processing levels.[2,5,7,10,65] In particular, previous studies have supported that the training-induced improvement of the readout for sensory signals through response reweighting within either sensory cortex[23–27] or higher decision areas[28–31,109] also plays a key role in VPL. Further work is thus needed to examine how the profile of VPL in feature space is constrained by these response reweighting changes, as well as to parse the relative contributions between tuning curve plasticity of sensory neurons and response reweighting changes to various potential profiles of VPL.

Additionally, the emerged distinct profile of VPL for grating orientation and face view discriminations in the pretrained AlexNet with six (high representational-similarity to early visual areas) and nine (high representational-similarity to object/face areas) layers, respectively, is not only in line with previous studies and theories that interpret VPL as a result of training induced tuning curve plasticity of neurons in the task-relevant sensory areas,[16–22,90,91] but also adds strong evidence supporting DCNNs' potential to perform human-like representation, such as the specificity[68] and generalization[69] of VPL, visual hierarchical coding,[110] and face processing.[111] Although these similarities between DCNNs and humans were mostly qualitative, the DCNN can provide new ways of studying VPL from behavior to physiology, serving as a test bed for various theories and assisting in generating predictions for physiological studies.

Ultimately, the present study opens several questions: First, VPL has been documented in virtually all kinds of tasks at different levels of visual analysis. It is thus worthwhile to address whether our conclusion can be generalized to other visual stimuli, such as contrast, spatial frequency, phase, acuity, color, and motion direction for the simple low-level features, and biological motion, natural images, shapes, and objects for complex high-level stimuli. In addition, perceptual learning is known to occur in not only the vision domain but also other sensory modalities, including audition, touch, smell, taste, and multimodal combinations.[2,5,7,10,12] Revealing their profiles could pave the way for a better practical application of perceptual learning in the education, rehabilitation of patients, and training of expertise. Second, although numerous studies and models have demonstrated that VPL is also specific to the trained retinal location,[2,5,7,13,18,19,112] previous studies have employed the double training technique[55,113] or increasing the variability of task-irrelevant features[69] to enable VPL transfer to new retinal locations, and further studies should therefore, reveal this spatial profile of VPL and how its spatial and feature-spatial profiles interactively shape our learning of the world. Third, stimuli we see in real life are often in a crowded environment or embedded in external noise, while subjects in our study were presented with isolated and noiseless stimuli. Dosher and Lu[2,24,25] have suggested that VPL mechanism reflects a combination of external noise exclusion and internal noise reduction. Previous studies have indicated that they are two independent processes[114] and may have distinct neural mechanisms for plasticity in the brain.[115] Future studies comparing the profile of VPL with and without external noise would help to improve our understanding of the VPL as a whole. Finally, the reverse hierarchy theory[65] and corresponding studies[32,52] have previously put forward that the relative degree of specificity and transfer of VPL depends on task precision. Easier tasks may be learned on the basis of neurons in the higher order visual cortex, resulting in a strong generalization, whereas difficult tasks require high-precision information to be available in early visual areas and thus show a strong specificity. Further work is needed to address how task precision shapes the profile of VPL for both the simple low-level and complex high-level stimuli.

In sum, our study provides, to the best of our knowledge, the first evidence for a feature hierarchy-dependent (simple low-level versus complex high-level) profile of VPL in feature space, thereby furthering our understanding of the relationship between its specificity and generalization, and how they mutually inspire various models and theories for VPL in the literature.

Limitations of the study

First, there were several differences in experimental design between GOD and FVD tasks, i.e., phase randomization across gratings rather than faces; retinotopic location and size randomization across faces rather than gratings (Figure 1B). Although these designs closely followed previous VPL studies using gratings[45,55,56,83,93,113] and faces,[1,35,95,100,101] respectively, we cannot deny a potential contamination by discrepancies between two tasks in their distinct profiles of VPL. Second, the profile of VPL, either center-surround or monotonic gradient, in our study was proposed by the symmetric shape. For each distance, the data of test stimuli was deviated from the trained stimulus either clockwise or counterclockwise and either left or right rotations of GOD and FVD tasks, respectively, was the same, lacking separate measurements to examine the asymmetrical profile of VPL. Third, the DCNN models in our study were modified from AlexNet[75] but did not know whether these models can be constructed from other classic baselines, such as ResNet,[116] GoogleNet,[117] and VGG16.[118] Finally, the feature hierarchy-dependent profile of VPL evident

here derive mainly from psychophysics and artificial neural networks, lacking directly evidence with neurophysiological or brain imaging techniques.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Apparatus
  - Stimuli
  - Procedure
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Gaussian and Mexican-hat models fitting and comparison
  - Deep convolutional neural network for the profile of VPL
  - Statistical analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.109128.

ACKNOWLEDGMENTS

We acknowledge the subjects for their contribution to this study. This work was supported by National Outstanding Youth Science Fund Project of National Natural Science Foundation of China (32022032), and National Natural Science Foundation of China General Program (32271099).

AUTHOR CONTRIBUTIONS

Conceptualization, X.Z.; Methodology, X.Z., C.M., and F.F.; Formal Analysis, S.S., Y.S., J.L., and C.L.; Investigation, S.S., Y.S., J.L., C.L., and Q.C.; Writing – Original Draft, X.Z.; Writing – Review and Editing, X.Z., S.S., Y.S., J.L., and C.L.; Supervision, X.Z.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 17, 2023
Revised: January 22, 2024
Accepted: February 1, 2024
Published: February 6, 2024

REFERENCES

1. Bi, T., and Fang, F. (2013). Neural plasticity in high-level visual cortex underlying object perceptual learning. Front. Biol. *8*, 434–443. https://doi.org/10.1007/s11515-013-1262-2.
2. Dosher, B., and Lu, Z.L. (2017). Visual perceptual learning and models. Annu. Rev. Vis. Sci. *3*, 343–363. https://doi.org/10.1146/annurev-vision-102016-061249.
3. Fahle, M., Poggio, T., and Poggio, T.A. (2002). Perceptual Learning (MIT Press).
4. Goldstone, R.L. (1998). Perceptual learning. Annu. Rev. Psychol. *49*, 585–612. https://doi.org/10.1146/annurev.psych.49.1.585.
5. Huxlin, K.R. (2008). Perceptual plasticity in damaged adult visual systems. Vis. Res. *48*, 2154–2166. https://doi.org/10.1016/j.visres.2008.05.022.
6. Lu, Z.L., and Dosher, B.A. (2022). Current directions in visual perceptual learning. Nat. Rev. Psychol. *1*, 654–668. https://doi.org/10.1038/s44159-022-00107-2.
7. Maniglia, M., and Seitz, A.R. (2018). Towards a whole brain model of Perceptual Learning. Curr. Opin. Behav. Sci. *20*, 47–55. https://doi.org/10.1016/j.cobeha.2017.10.004.
8. Sagi, D. (2011). Perceptual learning in vision research. Vis. Res. *51*, 1552–1566. https://doi.org/10.1016/j.visres.2010.10.019.
9. Seitz, A.R. (2017). Perceptual learning. Curr. Biol. *27*, R631–R636. https://doi.org/10.1016/j.cub.2017.05.053.
10. Seitz, A., and Watanabe, T. (2005). A unified model for perceptual learning. Trends Cognit. Sci. *9*, 329–334. https://doi.org/10.1016/j.tics.2005.05.010.
11. Watanabe, T., and Sasaki, Y. (2015). Perceptual learning: toward a comprehensive theory. Annu. Rev. Psychol. *66*, 197–221. https://doi.org/10.1146/annurev-psych-010814-015214.
12. Yang, J., Yan, F.F., Chen, L., Xi, J., Fan, S., Zhang, P., Lu, Z.L., and Huang, C.B. (2020). General learning ability in perceptual learning. Proc. Natl. Acad. Sci. USA *117*, 19092–19100. https://doi.org/10.1073/pnas.200290311.
13. Fahle, M. (2005). Perceptual learning: specificity versus generalization. Curr. Opin. Neurobiol. *15*, 154–160. https://doi.org/10.1016/j.conb.2005.03.010.
14. Gilbert, C.D., Sigman, M., and Crist, R.E. (2001). The neural basis of perceptual

learning. Neuron *31*, 681–697. https://doi.org/10.1016/S0896-6273(01)00424-X.
15. Li, W. (2016). Perceptual learning: Use-dependent cortical plasticity. Annu. Rev. Vis. Sci. *2*, 109–130. https://doi.org/10.1146/annurev-vision-111815-114351.
16. Adini, Y., Sagi, D., and Tsodyks, M. (2002). Context-enabled learning in the human visual system. Nature *415*, 790–793. https://doi.org/10.1038/415790a.
17. Bejjanki, V.R., Beck, J.M., Lu, Z.L., and Pouget, A. (2011). Perceptual learning as improved probabilistic inference in early sensory areas. Nat. Neurosci. *14*, 642–648. https://doi.org/10.1038/nn.2796.
18. Karni, A., and Sagi, D. (1991). Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. Proc. Natl. Acad. Sci. USA *88*, 4966–4970. https://doi.org/10.1073/pnas.88.11.4966.
19. Schoups, A.A., Vogels, R., and Orban, G.A. (1995). Human perceptual learning in identifying the oblique orientation: retinotopy, orientation specificity and monocularity. J. Physiol. *483*, 797–810. https://doi.org/10.1113/jphysiol.1995.sp020623.
20. Schoups, A., Vogels, R., Qian, N., and Orban, G. (2001). Practising orientation identification improves orientation coding in V1 neurons. Nature *412*, 549–553. https://doi.org/10.1038/35087601.
21. Teich, A.F., and Qian, N. (2003). Learning and adaptation in a recurrent model of V1 orientation selectivity. J. Neurophysiol. *89*, 2086–2100. https://doi.org/10.1152/jn.00970.2002.
22. Zhaoping, L., Herzog, M.H., and Dayan, P. (2003). Nonlinear ideal observation and recurrent preprocessing in perceptual learning. Network *14*, 233–247. https://doi.org/10.1088/0954-898X_14_2_304.
23. Dosher, B.A., Jeter, P., Liu, J., and Lu, Z.L. (2013). An integrated reweighting theory of perceptual learning. Proc. Natl. Acad. Sci. USA *110*, 13678–13683. https://doi.org/10.1073/pnas.1312552110.
24. Dosher, B.A., and Lu, Z.L. (1998). Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. Proc. Natl. Acad. Sci. USA *95*, 13988–13993. https://doi.org/10.1073/pnas.95.23.13988.
25. Dosher, B.A., and Lu, Z.L. (1999). Mechanisms of perceptual learning. Vis. Res. *39*, 3197–3221. https://doi.org/10.1016/S0042-6989(99)00059-0.
26. Petrov, A.A., Dosher, B.A., and Lu, Z.L. (2005). The dynamics of perceptual learning: an incremental reweighting model. Psychol. Rev. *112*, 715–743. https://doi.org/10.1037/0033-295X.112.4.715.
27. Poggio, T., Fahle, M., and Edelman, S. (1992). Fast perceptual learning in visual hyperacuity. Science *256*, 1018–1021. https://doi.org/10.1126/science.1589770.
28. Kahnt, T., Grueschow, M., Speck, O., and Haynes, J.D. (2011). Perceptual learning and decision-making in human medial frontal cortex. Neuron *70*, 549–559. https://doi.org/10.1016/j.neuron.2011.02.054.
29. Law, C.T., and Gold, J.I. (2008). Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. Nat. Neurosci. *11*, 505–513. https://doi.org/10.1038/nn2070.
30. Law, C.T., and Gold, J.I. (2009). Reinforcement learning canount for associative and perceptual learning on a visual-decision task. Nat. Neurosci. *12*, 655–663. https://doi.org/10.1038/nn.2304.
31. Mollon, J.D., and Danilova, M.V. (1996). Three remarks on perceptual learning. Spatial Vis. *10*, 51–58. https://doi.org/10.1163/156856896x00051.
32. Ahissar, M., and Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. Nature *387*, 401–406. https://doi.org/10.1038/387401a0.
33. Ball, and Sekuler, R. (1987). Direction-specific improvement in motion discrimination. Vis. Res. *27*, 953–965. https://doi.org/10.1016/0042-6989(87)90011-3.
34. Berardi, N., and Fiorentini, A. (1987). Interhemispheric transfer of visual information in humans: spatial characteristics. J. Physiol. *384*, 633–647. https://doi.org/10.1113/jphysiol.1987.sp016474.
35. Bi, T., Chen, J., Zhou, T., He, Y., and Fang, F. (2014). Function and structure of human left fusiformcortex are closely associated with perceptual learning of faces. Curr. Biol. *24*, 222–227. https://doi.org/10.1016/j.cub.2013.12.028.
36. Chen, N., Cai, P., Zhou, T., Thompson, B., and Fang, F. (2016). Perceptual learning modifies the functional specializations of visualcortical areas. Proc. Natl. Acad. Sci. USA *113*, 5724–5729. https://doi.org/10.1073/pnas.1524160113.

62. Güçlü, U., and van Gerven, M.A.J. (2015). Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. J. Neurosci. *35*, 10005–10014. https://doi.org/10.1523/jneurosci.5023-14.2015.
63. Khaligh-Razavi, S.M., and Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. PLoS Comput. Biol. *10*, e1003915. https://doi.org/10.1371/journal.pcbi.1003915.
64. Kriegeskorte, N. (2015). Deep neural networks: a new framework for modeling biological vision and brain information processing. Annu. Rev. Vis. Sci. *1*, 417–446. https://doi.org/10.1146/annurev-vision-082114-035447.
65. Ahissar, M., and Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. Trends Cognit. Sci. *8*, 457–464. https://doi.org/10.1016/j.tics.2004.08.011.
66. Lee, R., and Saxe, A. (2014). Modeling perceptual learning with deep networks. In Annual Meeting of the Cognitive Science Society, *36*Annual Meeting of the Cognitive Science Society.
67. Cohen, G., and Weinshall, D. (2017). Hidden layers in perceptual learning. In Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 4554–4562. https://doi.org/10.1109/cvpr.2017.568.
68. Wenliang, L.K., and Seitz, A.R. (2018). Deep neural networks for modeling visual perceptual learning. J. Neurosci. *38*, 6028–6044. https://doi.org/10.1523/jneurosci.1620-17.2018.
69. Manenti, G.L., Dizaji, A.S., and Schwiedrzik, C.M. (2023). Variability in training unlocks generalization in visual perceptual learning through invariant representations. Curr. Biol. *33*, 817–826.e3. https://doi.org/10.1016/j.cub.2023.01.011.
70. Watson, A.B., and Pelli, D.G. (1983). QUEST: A Bayesian adaptive psychometric method. Percept. Psychophys. *33*, 113–120. https://doi.org/10.3758/bf03202828.
71. Akaike, H. (1973). Information theory as an extension the maximumlikelihood principle. In Second International Symposium on Information Teory, B.N. Petrov and F. Csaki, eds. (Akademiai Kiado).
72. Schwarz, G. (1978). Estimating the dimension a model. Ann. Stat. *6*, 461–464.
73. Burnham, K.P., and Anderson, D.R. (2002). A practical information-theoretic approach. In Model selection and multimodel inference, 2nd ed. (Springer), p. 2.
74. Wagenmakers, E.J. (2007). A practical solution to the pervasive problems of p values. Psychon. Bull. Rev. *14*, 779–804. https://doi.org/10.3758/bf03194105.
75. Krizhevsky, A., Sutskever, I., and Hinton, G.E. (2012). ImageNet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. *25*, 1–9.
76. Donovan, I., Shen, A., Tortarolo, C., Barbot, A., and Carrasco, M. (2020). Exogenous attention facilitates perceptual learning in visual acuity to untrained stimulus locations and features. J. Vis. *20*, 18. https://doi.org/10.1167/jov.20.4.18.
77. Hung, S.C., and Carrasco, M. (2021). Feature-based attention enables robust, long-lasting location transfer in human perceptual learning. Sci. Rep. *11*, 13914–14013. https://doi.org/10.1038/s41598-021-93016-y.
78. Ito, M., Westheimer, G., and Gilbert, C.D. (1998). Attention and perceptual learning modulate contextual influences on visual perception. Neuron *20*, 1191–1197. https://doi.org/10.1016/s0896-6273(00)80499-7.
79. Li, W., Piëch, V., and Gilbert, C.D. (2004). Perceptual learning and top-down influences in primary visual cortex. Nat. Neurosci. *7*, 651–657. https://doi.org/10.1038/nn125-281.4(percep)h.659810TD(.)Tj-74.3896-1.1625TD809. ug,ZG.L., J.o, M mechanismes perceptual learning withed ng. Vis.

STAR★METHODS

KEY RESOURCES TABLE

RESOURCE AVAILABILITY

Lead coTA(i3sa/F3F21T7.727.7257.5775.7331T2rg(Further)-32.1(information)-321(and)-327(reuests)-321.2(for)-333.(resources)-31.4(and)

(hereafter referred to as distance in orientation space Δ0°, Δ18°, Δ36°, Δ54°, Δ72°, and Δ90°). All gratings were set at 2.5° diameter, 4 cycles/°, and 50% contrast, with the phase randomized for every presentation (Figure 1A). During the face view discrimination (FVD) task, all three-dimensional (3D) face models were generated by FaceGen Modeler 3.4 (http://www.facegen.com/). No hair was rendered and the value of texture gamma correction was set to 2.0. Face stimuli (extended 3 ° × 3 ° of visual angle) were generated by projecting a 3D face model with variant in-depth rotation angles onto the monitor plane with the front view (0°) as the initial position. Both left and right rotations were executed with a step size of 0.2°, which was used to generate a total of 1,501 face views from -150° (left tilted) to +150° (right tilted). For each subject, one of the face views ($\theta°$) was randomly selected for training, and other 6 face views for testing, which were 0°, 18°, 36°, 54°, 72°, and 90° deviated from the trained face view ($\theta°$), either left or right rotations (hereafter referred to as distance in view space Δ0°, Δ18°, Δ36°, Δ54°, Δ72°, and Δ90°, parallel to GOD task, Figure 1A) to ensure them were within the range from -150° (left tilted) to +150° (right tilted).

Procedure

Both GOD and FVD tasks consisted of six phases – pre-training test (Pre), discrimination-training 1 (Training1), mid-training test (Mid), discrimination-training 2 (Training2), post-training test 1 (Post1), and post-training test 2 (Post2). Pre, Mid, Post1, and Post2 took place on the days before, during, immediately after and one month after training, respectively (Figure 1C).

For both GOD and FVD tasks, during the two training phases (Training1 and Training2), each subject underwent six daily training sessions and a daily session (about 1 hour) consisted of 30 QUEST staircases[70] of 40 trials. In a trial, two targets ($\theta°$ and $\theta° \pm \Delta\theta°$) were each presented for 200-ms and separated by a 600-ms blank interval (Figure 1D), and their temporal order was randomized. Subjects were asked to make a two-alternative forced-choice (2AFC) judgment of the orientation in GOD task or the view in FVD task of the second target relative to the first one (left or right), and received auditory feedback if their response was incorrect. The $\Delta\theta°$ varied trial by trial and was controlled by the QUEST staircase to estimate subjects' discrimination thresholds (75% correct). To measure the time course of the training effect (learning curve), discrimination thresholds from 25 staircases in a daily training session were averaged, and then plotted as a function of training day. During the four test phases (Pre, Mid, Post1, and Post2), we measured discrimination thresholds in the each task for each distance (Δ0°, Δ18°, Δ36°, Δ54°, Δ72°, and Δ90°) and each subject. Each test phase consisted of 48 QUEST staircases of 40 trials: 8 QUEST staircases (same as above) were completed for each distance and the order of six distances was counterbalanced within individual subjects. Discrimination thresholds from the 8 staircases for each distance were averaged as a measure of subjects' discrimination performance. Subjects' performance improvement (i.e., the learning effect) for each distance was calculated as follows:

$$\text{Learning effect} = \frac{Threshold_{pre} - Threshold_{post}}{Threshold_{pre}} * 100\%$$

where $Threshold_{pre}$ is the measured discrimination thresholds at Pre; $Threshold_{post}$ could be the measured discrimination thresholds at Mid, Post1, or Post2. Differently, for the GOD task, the two sequentially presented gratings were always in the fovea; whereas for the FVD task, the spatial positions of two sequentially presented faces were randomly distributed within a 6.5 ° × 6.5 ° area whose center was coincident with the fixation point, with a constraint that these two faces were separated by at least 1.2° of visual angle.

QUANTIFICATION AND STATISTICAL ANALYSIS

Gaussian and Mexican-hat models fitting and comparison

During both GOD and FVD tasks, we fitted a monotonic model and a non-monotonic model to the learning effect for each subject. The monotonic and non-monotonic models were implemented as the Gaussian and Mexican-hat functions, respectively,[61] as follows:

$$Gaussian\ function: y = y0 + \frac{2A}{w\sqrt{2\pi}} e^{-2\left(\frac{x}{w}\right)^2}$$

$$Mexican - hat\ function: y = \frac{2H}{\frac{1}{\sqrt{3m}\pi\sqrt[4]{4}}} e^{-\frac{x^2}{2m^2}} \left(1 - \frac{x^2}{m^2}\right) + y1$$

where y is the learning effect, x is the distance in feature space between the trained and test stimuli (Δ0°, Δ18°, Δ36°, Δ54°, Δ72°, and Δ90°); w, A, and $y0$ are the three parameters controlling the shape of the Gaussian function; m, H, and $y1$ are the three free parameters controlling the shape of the Mexican-hat function. To compare these two models to our data, we first computed the Akaike information criterion (AIC)[71] and Bayesian information criterion (BIC),[72] with the assumption of a normal error distribution as follows:

$$AIC = N \ln\left(\frac{RSS}{N}\right) + 2K + \frac{2K(K+1)}{N - K - 1}$$

$$BIC = N \ln\left(\frac{RSS}{N}\right) + K \ln(N)$$

where N is the number of observations, K is the number of free parameters, and RSS is residual sum of squares. Then, we further calculated the Likelihood ratio (LR) and Bayes factor (BF) of the Gaussian model over the Mexican-hat model based on AIC[73] and BIC[74] approximation, respectively, as follows:

$$LR = e^{\left(\frac{AIC_M - AIC_G}{2}\right)}$$

$$BF = e^{\left(\frac{BIC_M - BIC_G}{2}\right)}$$

where AIC_G and BIC_G are for the Gaussian model, AIC_M and BIC_M are for the Mexican-hat model.

Deep convolutional neural network for the profile of VPL

We trained a deep convolutional neural network (DCNN) model modified from AlexNet[75] to perform our GOD and FVD tasks. The original AlexNet is a classical CNN model consisting of five convolutional (conv) layers and three fully connected (fc) layers, where the fc layers are placed after all the conv layers and the last layer is classified by softmax. In the first five layers, the model extracts features from the input image by convolution, from simple low-level features to complex high-level features with gradually increasing receptive fields along five conv layers. In the last three fc layers, the model integrates and classifies the extracted features. Here we adjusted the number of layers of original AlexNet and constructed two networks: AlexNet-GOD and AlexNet-FVD for our GOD and FVD tasks, respectively. Similar to the deep learning model from Wenliang and Seitz,[68] we took original five conv layers and discarded the last two fc layers to reduce model complexity for the AlexNet-GOD. However, we added three conv layers and took original 3 fc layers for the AlexNet-FVD since these late layers of network exhibited low representational-similarity to early visual areas but high similarity to object/face areas, such as inferior temporal[96] and superior temporal sulcus[97] areas, fusiform face and occipital face areas[98] and thus may be more relevant to face-view classification here.[62,63] Notably, AlexNet-GOD and AlexNet-FVD here were trained to classify whether the target was tilted clockwise or counterclockwise and rotated leftward or rightward compared with the reference, respectively. We thus first obtained the pixel difference between the target and reference images, which was then superimposed on the channels of the reference image. Finally, we used a conv layer (i.e., conv0) to match the number of the channels between superimposed feature maps and the pre-trained AlexNet. Therefore, the AlexNet-GOD in our study consisted of 6 conv layers and 1 fc layer, whereas AlexNet-FVD consisted of 9 conv layers and 3 fc layers (Figure 4). During the training, layers 2-6 of both the AlexNet-GOD and AlexNet-FVD were initialized with the weights of the first 5 layers of the pre-trained AlexNet, and the other weights were initialized randomly. The last layer (seven- and twelve-layers of AlexNet-GOD and AlexNet-FVD, respectively) was trained to capture the difference between the target and reference and finally obtain the classification by softmax, to model decision making in our 2AFC paradigm (Figure 1B), in which subjects were asked to make a 2AFC judgment of the orientation in GOD task or the view in FVD task of the second stimulus (target) relative to the first one (reference). Moreover, the angle separation (grating orientation and face view differences in the AlexNet-GOD and AlexNet-FVD, respectively) between the target and reference in the network was set to 5°.

Parallel to our psychophysical experiments, both the AlexNet-GOD and AlexNet-FVD were independently trained 15 times. For each time, the trained orientation (θ°) of AlexNet-GOD was chosen randomly from 0° to 180°; the 11 test gratings were 0°, ± 18°, ± 36°, ± 54°, ± 72°, and ± 90° deviated (clockwise and counterclockwise) from the trained orientation. All grating stimuli (phase: random) were centered on 227 × 227-pixels images with gray background. The trained view (θ°) of AlexNet-FVD was chosen randomly from -150° to 150°; the 11 test faces were 0°, ± 18°, ± 36°, ± 54°, ± 72°, and ± 90° deviated (left and right rotations) from the trained view. All face stimuli were presented randomly on 227 × 227-pixels images with black background. To improve the robustness of our model, we trained the network on all combinations of several parameters: contrast (0.1, 0.15, 0.2. 0.25, 0.3, 0.4, 0.5, 0.6, 0.7, and 0.8) and SD of the Gaussian additive noise (5, 10, 15, 20, 25, 30, 35, 40, 45, and 50) for both the AlexNet-GOD and AlexNet-FVD; spatial wavelength (5, 10, 15, 20, 25, 30, 40, 50, 60, and 80 pixels) and SD of the Gaussian additive blur (0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, and 5.0) for AlexNet-GOD and AlexNet-FVD, respectively. For each training, there were thus a total of 2,000 images; 1,600 images were the training set and the other 400 images were the test set. For both the AlexNet-GOD and AlexNet-FVD, there were six different distances (Δ0°, Δ18°, Δ36°, Δ54°, Δ72°, and Δ90°), and for each distance, the training effect was defined as the accuracy difference between the pre- and post-training.

Statistical analysis

Analyses were performed using paired t test to compare two conditions and repeated measures ANOVA with both post-hoc analyses and Bonferroni correction for multiple comparisons. The assumption of homogeneity of variance was used to determine whether the data met assumptions of the statistical approach. Sample size and statistical tests are also reported in the figure notes. No subject was excluded from any analyses and all results presented here are from all subjects.