Graduate School of Neural Information Processing
University of Tübingen

Evaluating Learning Objectives in Visual Perceptual Learning Using Deep Neural Networks

Lab Rotation Report

Ali Gholamzadeh

The study was supervised by

Dr. Shahab Bakhtiari

The Systems Neuroscience and AI Lab
Université de Montréal

Duration of the lab rotation: March 11 – Nov 24, 2023
Deadline for submission: 11 March 2024

Abstract

This report presents an analysis of Visual Perceptual Learning (VPL) through the lens of advanced Deep Neural Network (DNN) models, specifically focusing on enhancements made to existing models to achieve a more accurate representation of VPL. By incorporating anatomical skip connections and integrating a Recurrent Neural Network (RNN) for dynamic processing and reaction time measurement, alongside the exploration of unsupervised loss functions, this study aims to bridge the gap between computational modeling and biological plausibility in VPL. Our findings demonstrate the models' ability to learn and differentiate between various task precisions, showing behaviors akin to those observed in human and primate VPL tasks, thus offering new insights into the mechanisms of perceptual learning and its applications in artificial intelligence and vision rehabilitation.

1. Introduction

Visual Perceptual Learning, the process of improving perceptual skills through practice, is a fundamental aspect of sensory psychology and neuroscience, revealing how experiences shape our sensory modalities. VPL, in particular, highlights how exposure to visual stimuli enhances our ability to discriminate between different features, such as orientation, contrast, and motion, contributing significantly to our understanding of sensory processing and learning mechanisms. Despite considerable advancements, the intricate dynamics of VPL continue to elude comprehensive theoretical explanation, necessitating further exploration into the neural correlates and computational models that can accurately replicate these learning processes.

Recent studies leveraging Deep Convolutional Neural Networks (DCNNs), modeled after the hierarchical structure of the visual cortex, have shown promise in mirroring the complexities of human perceptual learning. These models, capable of emulating the functional intricacies of early, intermediate, and late visual areas, present a compelling framework for investigating VPL. However, traditional approaches often fall short in aligning with physiological data, prompting the need for refined models that consider the contributions of multiple visual areas and incorporate biologically plausible learning mechanisms. This paper builds on the foundational work of Wenliang and Seitz [26], extending the model to include skip connections and RNN components, aiming to enhance biological plausibility and measure dynamic processes such as reaction time. By exploring these modifications within the context of VPL, we aim to deepen our understanding of perceptual learning's underlying mechanisms and its implications for artificial intelligence, machine learning, and vision rehabilitation.

2. Background

2.1 Visual Perceptual Learning

Perceptual learning, the improvement in perceptual tasks through practice, has been a subject of interest for sensory psychologists since the early days of experimental psychology [1]. It occurs in all sensory modalities, such as vision, audition, touch, smell, taste, and combinations of these [2]. This type of learning can lead to significant performance enhancements and can persist for extended periods [3,4].

Visual perceptual learning (VPL) involves the enhancement of sensitivity to visual stimuli through training or experience. This phenomenon has been demonstrated in the discrimination of simple features such as orientation, contrast, and motion direction, as well as more complicated patterns [5, 6, 7, 8, 9]. Many studies on visual perceptual learning employ forced-choice tasks. In these tasks, such as the two-interval two-alternative forced choice (2I-AFC) detection task, participants are first presented with a reference stimulus, followed by a target stimulus. They are then required to determine if the target is more clockwise or more counterclockwise compared with the reference. Typically, feedback is provided to inform the participant whether their judgment was accurate. The evaluation of a participant's performance, in terms of accuracy or reaction time, is usually conducted over a series of trials within a block (which includes dozens of trials) or a session (comprising hundreds of trials). The results are graphically represented through a learning curve that maps performance improvement over the number

of blocks or sessions. As the training progresses, participants tend to show improvements in both speed and accuracy, indicating that the task becomes less challenging. (fig. 1)

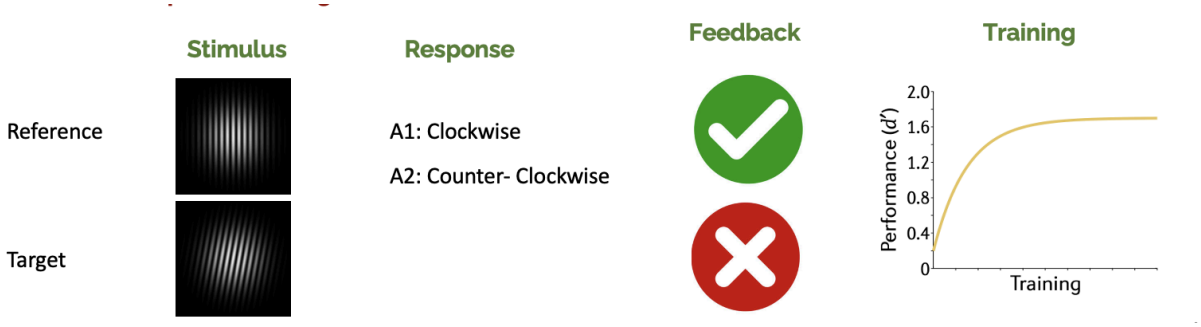

Figure 1. Standard paradigm of Visual Perceptual Learning

The hallmark of perceptual learning, particularly evident in laboratory settings, is its specificity: the enhancement in performance achieved through training on a specific task and stimulus often does not extend to similar tasks and stimuli [2]. Visual perceptual learning is specifically related to factors such as the retinal location, the involved eye, and the stimulus or task being trained. However, this specificity is not absolute; a degree of learning transfer to similar stimuli and tasks is often observed, influenced by a variety of factors. The observation of specificity in behavioral outcomes could be consistent with enhanced weighting or read-out processes in the initial cortical regions, rather than directly indicating plastic changes within the early visual cortex [10]. The degree of specificity and its transfer are measured by the specificity index (SI), which can vary significantly across different tasks].

2.2 Theories of perceptual learning

Perceptual learning in the visual domain is explained by two primary theories: representation enhancement and information reweighting. Representation enhancement suggests that learning modifies the responses or tuning of neurons in early visual areas, while information reweighting involves adjusting the emphasis on relevant versus irrelevant inputs during perceptual decision-making, without altering the underlying neural representations [7,11,12,13]. Although both approaches improve signal-to-noise ratios, reweighting, which can occur at various stages in the visual processing pathway, is considered the more prevalent mechanism in adapting to specific tasks [14,15].

Evidence largely supports the stability of visual representations in early cortical areas, with significant plasticity observed in higher visual areas during active tasks, indicating a transient, task-specific influence rather than permanent changes [16,17]. Neurofeedback and imaging studies further suggest that while early cortical areas can be influenced by targeted interventions, perceptual learning primarily engages higher-level areas through the reweighting of information [16,19,20].

The majority of computational models and empirical data support the reweighting theory, highlighting its role in accounting for the observed dynamics of perceptual learning without necessitating persistent changes in early sensory representations [11,21,22,23]. This suggests that perceptual learning largely reflects a reevaluation of sensory information rather than a fundamental alteration of sensory representations.

Most current models for Visual Perceptual Learning (VPL) use artificial neural networks and are trained with Hebbian-like [22, 23,24] rules or optimal decoding methods [25]. However, they often fall short in aligning with physiological data and considering how multiple visual areas contribute to learning [26]. While some models like the Reverse Hierarchy Theory [13] (RHT) and the Dual Plasticity Model [12] propose theoretical frameworks for VPL, they do not predict specific neuron tuning changes. A recent study by Wenliang & Seitz [26] reveals that a Deep Neural Network serves as an effective computational model for visual perceptual learning, accurately replicating essential behavioral and physiological observations despite not being specifically designed for this purpose.

2.3 Deep convolutional neural networks

Deep convolutional neural networks (DCNNs) are designed for tasks like image recognition, learning from vast datasets of images and their labels to fine-tune the connections between layers, achieving high accuracy in image classification. Although these networks are primarily tools in artificial intelligence and not models of human behavior, they encounter similar challenges as those found in the study of visual perceptual learning, including issues of specificity versus generalizability and balancing plasticity with stability.

DCNNs, modelled after the visual cortex, are composed of hierarchically structured layers, each receiving input from the one before it. They can be comprehensively trained to model intricate input-output relationships and can achieve a human-like precision in categorizing natural images [29]. Upon training, DCNNs' feature detectors mirror the functional properties of the neurons across the visual cortex, extracting simple features in initial layers and complex visual forms in deeper layers [30]. Furthermore, DCNNs have demonstrated significant similarities with human behaviors and neural data from early, intermediate, and late visual areas (such as the inferior temporal cortex (IT)) [28]. This makes DCNNs a natural choice for modeling and understanding VPL [22] in line with the RHT [31]. Although early studies have used simple neural network architectures and shallow networks [32] to mimic varied perceptual training conditions, the capacity of DCNNs to accurately represent VPL's physiological data is still a largely unexplored research avenue.

In the context of modeling perceptual learning, DCNNs and the integrated reweighting theory present distinct approaches [26]. DCNNs learn representations and adjust connections from the ground up, while the integrated reweighting theory starts from established properties of the visual system, acknowledging that humans have some level of proficiency in visual tasks even prior to training. After extensive training, the early layers of DCNNs show similarities to the processing in early visual cortical areas [27,28]. However, DCNNs rely on supervised learning with labeled images, whereas the integrated reweighting theory, mimicking natural learning processes, can operate without explicit feedback and proves more resilient to slight changes in visual inputs.

2.4 Deep Learning Model of Perceptual Learning

The Reverse Hierarchy Theory (RHT) does not explicitly detail how training induces selective neural plasticity based on task precision. However, a study by Wenliang and Seitz [26] demonstrated that the principles of RHT could be observed in a deep neural network (DNN), specifically designed and trained for this purpose. By retraining a DNN known as AlexNet, initially pretrained on image classification tasks, on an orientation discrimination task, Wenliang and Seitz (2018) were able to replicate key findings from VPL research (fig. 2A). This experiment aimed to show that a DNN could not only exhibit the behavioral outcomes seen in VPL but also emulate the specific neurophysiological changes observed in studies with nonhuman primates(fig. 2B and 2C). These changes were expected to reflect a layer-specific plasticity within the network, influenced by the precision of the task at hand, aligning with predictions made by the RHT.

The proposed DNN was engaged in an orientation discrimination task at various precision levels (0.5–10° separation angle), assessing its learning outcomes not only with the trained stimuli but also with untrained spatial frequencies and orientations. This approach mirrored human VPL processes, where the network exhibited diminished sensitivity transfer to new stimuli with higher precision training, akin to findings in prior research [8] (fig. 2D and 2E). Interestingly, the network achieved peak accuracy more swiftly on tasks requiring less precision, highlighting a parallel with human learning challenges in high-precision tasks.

The proposed model, along with earlier theories, suggested that VPL-related behavioral improvements could be attributed to the retuning of sensory neurons [33], although VPL improvements might not always stem from sensory neuron retuning. The DNN model, representing a scenario where sensory information is ideally integrated and interpreted downstream from the visual cortex, proposes that VPL enhances performance by adjusting sensory neurons' tuning to be more task-relevant. This concept aligns with the model but might not fully apply to the human visual system. In cases where the information readout from sensory neurons isn't optimal, VPL improvements primarily involve

fine-tuning this readout process, reflecting a reweighting mechanism [23] rather than altering neuron tuning.

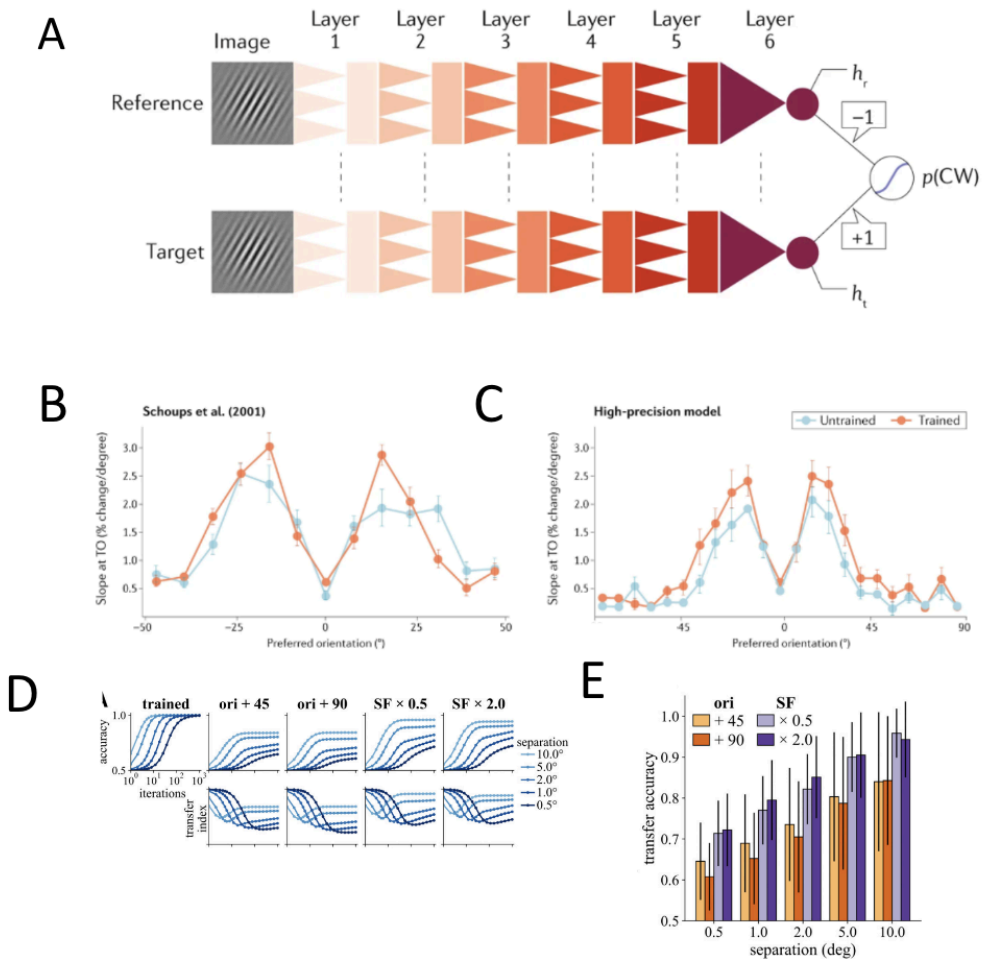

Figure 2. A) A deep network model of perceptual learning. Clockwise-oriented or anticlockwise-oriented visual inputs flow through layers of weights to an output layer that reports the direction of rotation. h_t and hr denote the last hidden unit for the target and reference images, respectively; p(CW) is the probability that the target image is clockwise with respect to the reference[18]. B) Tuning curve slope changes due to learning measured in primate primary visual cortex (V1) [16]. C) Alterations in the slope of tuning curves resulting from the application of gradient descent learning in the model depicted in part A. D) displays the trajectory of accuracy (top) and transfer index (bottom) across training iterations, with darker blue signifying greater precision. E) Transfer of behavioral performance to varying orientations (Ori) and spatial frequencies (SF) following training on tasks with different angular separations[26].

Adjusting the tuning of sensory neurons (representation model) and modifying how information is weighted during the decision-making process (reweighting model) both contribute to the behavioral outcomes observed in Visual Perceptual Learning (VPL), yet they operate through distinct processes. The reweighting model, different from the structure of deep neural networks (DNNs), allows the decision-making neuron to interact directly with neurons across all layers. In tasks requiring high precision, this model emphasizes the importance of initial visual processing stages by assigning greater significance to them, enhancing task specificity without altering the neurons' response characteristics [11]. On the other hand, Wenliang and Seitz (2018) demonstrated that within DNNs' layered hierarchy, learning is influenced by the specific adjustments in the early layers, even without direct communication between these layers and the decision-making component, raising questions about the primary drivers of VPL specificity.

Wenliang and Seitz [26] found that within the strict hierarchy of DNNs, which does not naturally accommodate the reweighting approach, the observed enhancements largely stem from modifications in the response patterns of neurons across different layers. Introducing "skip

connections" that link all layers directly to the decision-making layer could integrate the reweighting and representation theories, allowing for a more comprehensive analysis of their respective roles in improving behavior. This concept of skip connections finds support in the anatomical structure of the visual cortex, as noted in study by Felleman and Van Essen [34], suggesting a potential parallel in how the brain might integrate these mechanisms.

To enhance the accuracy of the existing model initially proposed by Wenliang and Seitz [26], we implemented several modifications. First, to better integrate the concept of anatomical skip connections and align the model more closely with the reweighting approach, we incorporated skip connections into the AlexNet architecture. Second, considering the significance of reaction times in studies of visual perceptual learning, which the original model's static framework could not capture, we introduced a recurrent neural network (RNN) as the decision-making component to reflect the dynamic nature of task processing and decision-making within individual trials. Finally, whereas the original model was trained entirely in a supervised manner, we also explored the impact of introducing unsupervised feedback to the network, aiming to investigate its role further.

3. Material & Methods

3.1 Model Description

In our study, we utilized a deep neural network (DNN) based on the AlexNet architecture, enhanced with skip connections and a recurrent neural network (RNN) component, to simulate visual perceptual learning (VPL), following the approach outlined in [25]. For a detailed explanation of the network architecture, we direct readers to [29]. Originally, AlexNet comprises eight layers, with the first five layers featuring units that connect to small, retinotopically arranged patches in the preceding layer or input image. This arrangement facilitates spatial replication of feature extraction across all locations via weight sharing, a process known as convolution. The final three layers of AlexNet, fully connected, originally map to object labels for classification tasks.

In our adaptation, to mirror early visual processing more accurately and streamline the model, we retained only the convolutional layers of AlexNet, excluding the three fully connected layers. This modification reflects findings suggesting the latter layers' high representational similarity to inferotemporal (IT) cortex regions, thus aligning more with object classification tasks than with the nuances of VPL, particularly for tasks like Gabor orientation discrimination [28, 35]. A fully connected layer has been incorporated, establishing complete integration with all convolutional layers through skip connections. This configuration forms a comprehensive feature representation of the stimulus within the newly constituted sixth layer of our model. (fig 3.)

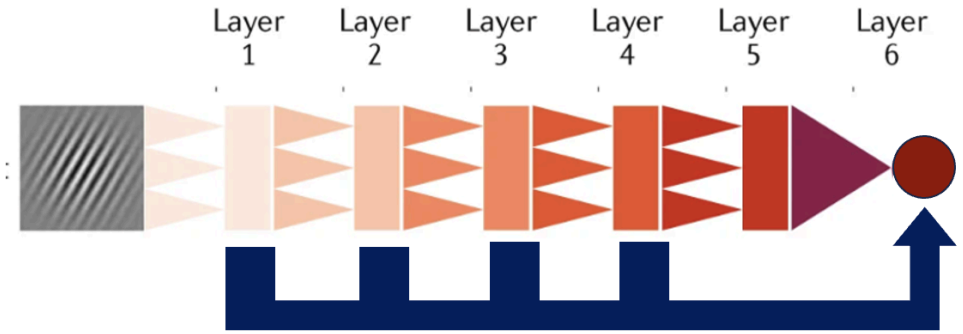

Figure 3. AlexNet with skip connections

Upon feature extraction using AlexNet equipped with skip connections, the resulting features are inputted into a Recurrent Neural Network (RNN) module. RNNs, a subset of neural networks, are distinguished by their ability to use previous inputs to maintain and update their hidden states. The standard architecture of RNNs is depicted (fig. 4).

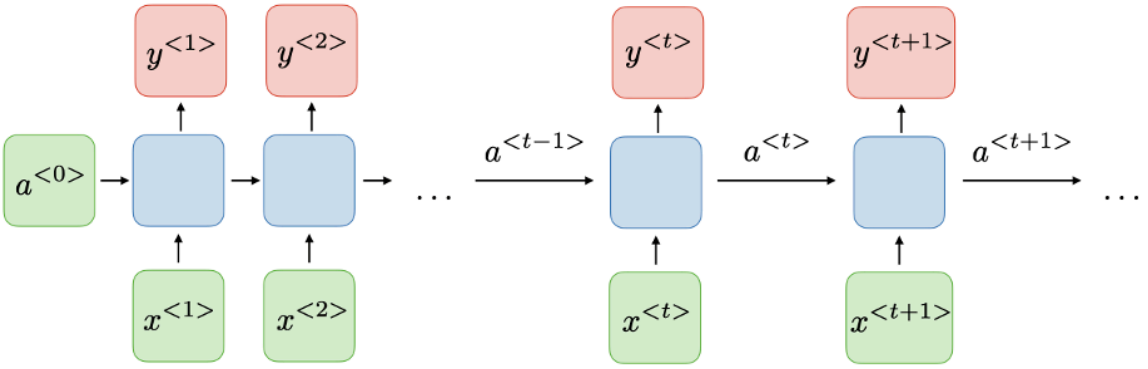

Figure 4. Architecture of a traditional RNN [41]

For each time step t, activation $a^{<t>}$ and output the output $y^{<t>}$ are computed as follows:

$$a^{<t>} = g_1(W_{aa}a^{<t-1>} + W_{ax}x^{<t>} + b_a) \quad (1)$$

$$y^{<t>} = g_2(W_{ya}a^{<t>} + b_y) \quad (2)$$

In these expressions, $x^{<t>}$, represents the input at time t, $a^{<t>}$ denotes the hidden state at time t, encapsulating information from current and past inputs. The matrices W_{ax}, W_{aa} and W_{ya} correspond to the weights from input to hidden layer, hidden layer to itself, and hidden layer to output layer, respectively, while b_a and b_o are bias terms. The function g typically represents a nonlinear activation function, such as the hyperbolic tangent or sigmoid function[36].

For this research, we employed a specific variant of Recurrent Neural Networks known as the Gated Recurrent Unit (GRU). The Gated Recurrent Unit (GRU) is a type of Recurrent Neural Network (RNN) architecture designed to solve the vanishing gradient problem that traditional RNNs face, enabling it to capture dependencies from long sequences of data more effectively. GRUs streamline the architecture of RNNs through the use of two key mechanisms: the update gate and the reset gate. The update gate helps the model decide how much of the past information to carry forward to the next step, while the reset gate determines how much of this information to forget. This allows GRUs to effectively manage information over long durations, making them particularly useful for tasks that involve sequential data processing [37].

The six-layer network, augmented with an RNN module, was adapted to simulate decision-making within the two-interval two-alternative forced choice (2I-2AFC) paradigm, as illustrated in Figure 5. In this setup, a reference stimulus is presented for a brief period, followed by a target stimulus. The model is tasked with determining if the target stimulus is rotated more clockwise or counterclockwise relative to the reference. In our DNN framework, both the reference and target stimuli undergo processing through the identical six-layer network. The DNN architecture processes both stimuli using the same six-layer setup, ending in a scalar output from the RNN. A readout layer positioned atop the RNN then displays the probability of the image being oriented clockwise.

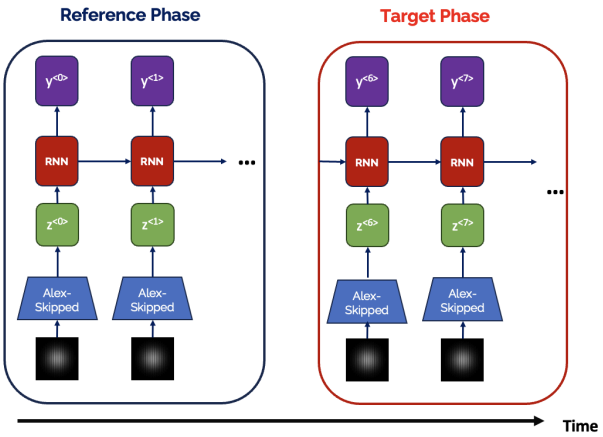

Figure 5. Dynamic VPL Deep Neural Network

3.2 Task and stimuli

In the two experiments outlined below, the stimuli were composed of 5 reference images and 5 target images, all centered on 8-bit, 227 by 227-pixel frames with a gray background.

3.3 Experiments

Experiment 1:

The network underwent training to determine if a target Gabor stimulus was oriented clockwise or counter-clockwise relative to a reference stimulus. The training parameters included a reference Gabor orientation of 0° and a spatial frequency of 0.05 cycles per pixel. To mimic the brief exposure to the stimulus and the presence of sensory noise, we maintained a low contrast level and introduced isotropic Gaussian noise with a standard deviation of 0.001 to each image. The training covered various angle separations between the reference and target stimuli, specifically at 0.5°, 1.0°, 2.0°, 5.0°, and 10.0°. To evaluate the network's ability to generalize and transfer learning to new tasks, we tested the trained model using images with a different spatial frequency of 0.1.

At each time step, the network produces an output. The input includes 5 reference images and 5 target images, with Gaussian noise added to the latter. We utilized a binary cross-entropy loss function to minimize the difference between the network's predictions for the five target images and their actual labels.

$$BCE = -\frac{1}{N}\sum_{1}^{N}(y_i log(\widehat{y_i}) + (1- y_i)log(1 - \widehat{y_i})) \quad (3)$$

- N is is the number of samples,
- y_i is the actual label of the ith sample,
- $\widehat{y_i}$ is the predicted probability that the ith sample belongs to the positive class.

Experiment 2:

Experiment 2 closely mirrored the first experiment, with a key distinction. In training the network, we implemented a compound loss function that combined classification error, derived from a binary cross-entropy function, and prediction error, obtained from a mean squared error function. This required modifying the network to include an additional readout layer on top of the recurrent neural network, tasked with predicting the next image representation. We then calculated the mean distance between the output of this layer and the actual embedded representation. Consequently, the loss function was a weighted sum of both the classification and prediction errors.

$$MSE = mean(z_t - \widehat{z_t})^2 \quad (4)$$

Compound Loss = Classification Loss + λ Prediction Loss (5)

- z_t: Actual feature representation at time
- $\widehat{z_t}$: Predicted feature representation at time
- λ: Prediction error weight

3.4 Training procedure

In both experiments, we initialized the network weights by importing them from an AlexNet pre-trained on the ImageNet Dataset for the convolutional layers, while all other weights were set randomly. We employed stochastic gradient descent (SGD) as the learning algorithm, adjusting the weights to reduce the gap between the network's output and the given stimulus labels. The learning rate was set to 0.001, and the momentum was set to 0.9, both of which remained constant throughout the training process. The loss function, which was cross-entropy loss for the first experiment and compound loss for the second, varied based on the network weights, the input images batch of 20 sequences, and their respective labels. The calculation of gradients for weight adjustment was performed using backpropagation through the network layers [38].

3.5 Specificity Index (SI)

The Specificity Index (SI) is a measure used to quantify the degree to which learning or improvement in one task transfers to other tasks or conditions. In the context of Visual Perceptual Learning (VPL), the SI assesses how training on a specific visual task affects performance on similar but untrained tasks. A high SI indicates that the learning is highly specific to the trained task, showing little to no transfer to other tasks, which suggests that the skills or improvements gained are closely tied to the particular stimuli or conditions of the training. Conversely, a lower SI suggests a broader transfer of learning, indicating that the improvements are not strictly bound to the trained conditions and can enhance performance on different but related tasks. The SI is essential for comprehending the boundaries and capabilities of perceptual learning, offering insights into how learning transfers across different sensory or cognitive areas and the generalizability of training effects. For this task we compute the SI as follows:

$$\text{Specificity Index (SI)} = \frac{test\text{-}loss^{<T>} - train\text{-}loss^{<T>}}{train\text{-}loss^{<0>} - train\text{-}loss^{<T>}} \quad (6)$$

3.6 Estimating learning in layers

Following the training period, the weights in each layer were consolidated into a singular vector, and the extent of learning was quantified by the deviation from the values prior to training. For a given layer that has N total connections to its preceding layer, let's denote the initial N-dimensional weight vector, which was trained for object classification, as w (where N and w are as specified in AlexNet). The alteration in this weight vector attributable to perceptual learning is represented as Δw. Consequently, the change in the layer is determined as follows:

$$d_{rel1} = \frac{\sum_{i}^{N} |\delta w_i|}{\sum_{i}^{N} |w_i|} \quad (7)$$

where i indexes each element in the weight vector

4. Results:

Following the described training methodology, we successfully trained the networks in the first experiment across all separation angles. As anticipated, increasing the task's precision (by decreasing the separation angle) made training the network more challenging, requiring additional epochs to achieve 100 percent accuracy, as illustrated in Figure 5.A. The observed fluctuations in accuracy highlight the model's uncertainty during training, caused by the strict threshold applied for calculating accuracy.

In the second experiment, we were also able to train the network using a compound loss across all separation angles, as shown in Figure 5.B. Incorporating unsupervised loss into the training necessitated more epochs to reach a stable state compared to the first experiment. Notably, in this setup, the classification loss steadily decreased, while the prediction loss initially increased before decreasing. This pattern suggests that the model initially focuses on accurately predicting the labels for the target images before leveraging the prediction error to stabilize, ultimately enabling it to predict both the next image and the correct target label simultaneously.

.

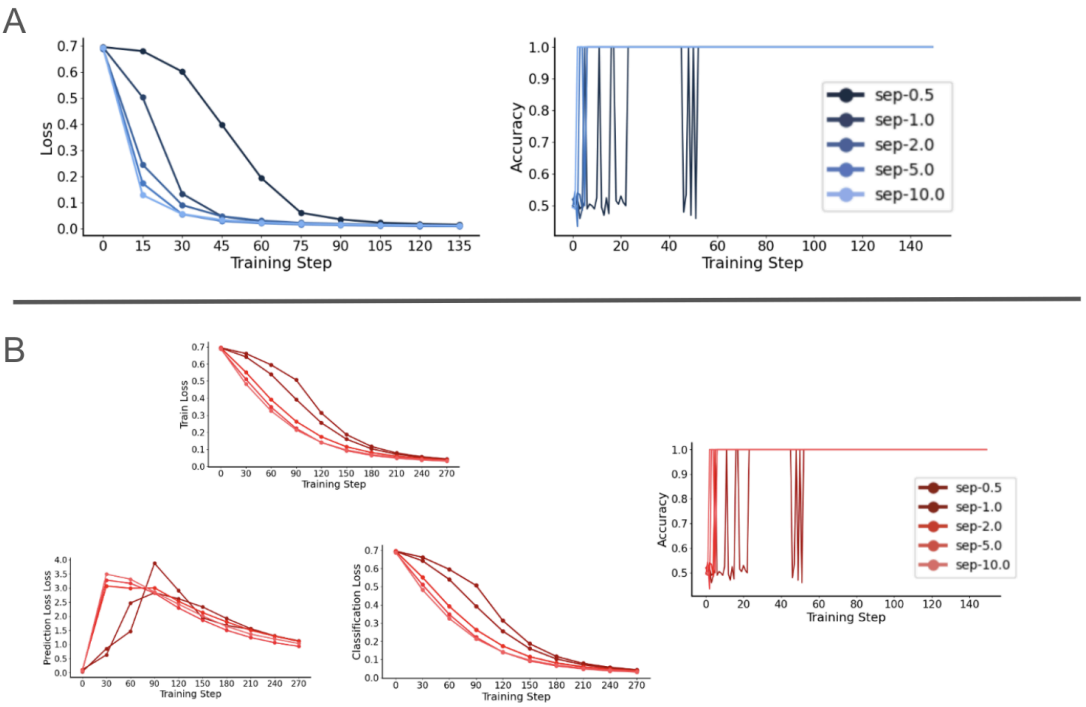

Figure 5. Training Loss Function and Accuracy. A) Experiment 1. B) Experiment 2.

For both experiments, we calculated the specificity index using test images with a different spatial frequency, as depicted in Figure 6. In line with findings from human experiments, the specificity index increased with task precision, indicating lower generalizability at smaller separation angles for both experiments. However, the decline in specificity index exhibited different patterns between the two experiments: the first experiment displayed a sharp decrease when moving from a separation angle of 1.0 to 2.0, whereas the second experiment showed a more uniform decreasing pattern. Overall, the second experiment demonstrated a higher specificity index, suggesting that the inclusion of prediction loss did not enhance the model's ability to generalize effectively in this task.

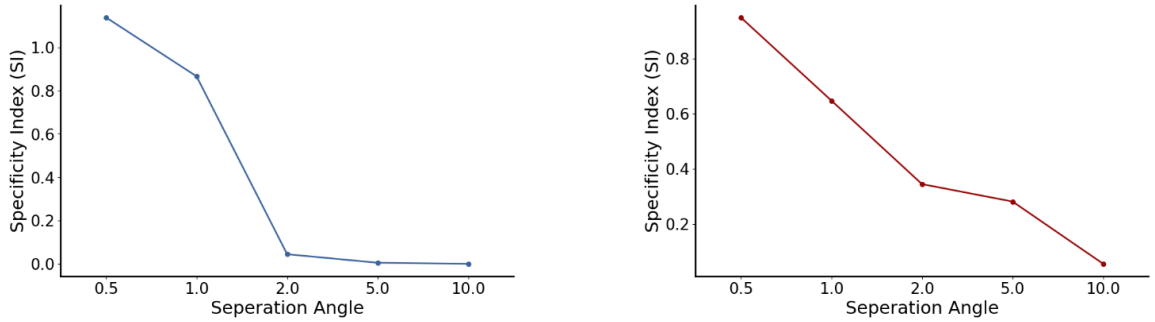

Figure 6. Specificity Index (SI). A) Experiment 1. B) Experiment 2.

Additionally, we explored the model's reaction time. This was assessed by taking the average inverse of the slopes between the output of the last reference image and the first target image, serving as an indicator of reaction time. Our observations revealed that the model's reaction time decreased with training, enabling it to reach the decision threshold more quickly (fig. 7). This observation is consistent with findings from other studies, which have frequently linked a reduction in reaction time to the learning process, highlighting a noticeable decrease as training progresses.

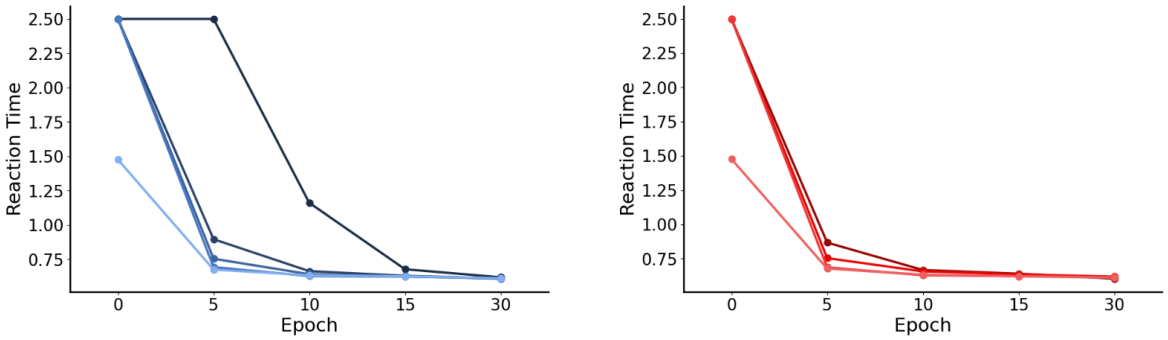

Figure 7. Reaction Times. A) Experiment 1. B) Experiment 2.

Finally, for both experiments, we assessed the extent of learning across different layers by applying Equation 7 to various epochs. It was noted that most of the training occurred within the weights of the dense layer, which is responsible for selecting features to be input into the RNN, as well as within the weights of the RNN itself. Conversely, the weights of the convolutional layers exhibited minimal changes over the course of training, as illustrated in Figure 8. This observation is in line with the reweighting theory, which suggests that during VPL tasks, the representation itself does not alter. Instead, what changes is how information from this representation is read, specifically enhancing the emphasis on task-relevant information while diminishing the focus on task-irrelevant information to improve performance.

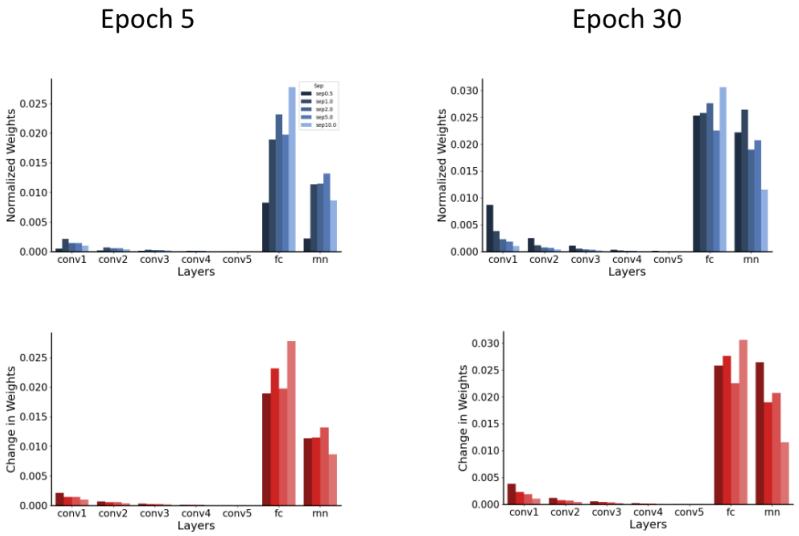

Figure 8.Learning in Different Layers. A) Experiment 1. B) Experiment 2.

5. Discussion

In this study, we implemented several modifications to the model initially proposed by Wenliang and Seitz [25] to develop a more precise representation of Visual Perceptual Learning (VPL). By integrating skip connections, we aimed to enhance the model's biological plausibility. Additionally, the incorporation of a Recurrent Neural Network (RNN) component introduced dynamics to the model, enabling the measurement of reaction time. Furthermore, we explored the impact of incorporating an unsupervised loss function into the fundamentally supervised model, acknowledging that VPL can occur even without external feedback.

The adapted models successfully learned the task, distinguishing between various separation angles throughout the training process. These models demonstrated behaviors akin to those observed in humans and primates during VPL tasks, including similar specificity patterns. Specificity is a crucial aspect of VPL in laboratory experiments and is considered a defining characteristic of VPL behavior. The models also accounted for the reduction in reaction time with continued training. Moreover, the addition of skip connections to AlexNet suggested that this model is more closely aligned with the information reweighting theory, the prevailing explanation for performance enhancements in VPL.

Utilizing deep neural networks (DNNs) to model Visual Perceptual Learning (VPL) provides key benefits for both theoretical knowledge and practical applications. On a theoretical level, it advances our understanding of how VPL functions in the brain by examining different DNN architectures and learning techniques, thereby offering insights into the cognitive processes underlying perceptual learning. This knowledge can improve the design of AI and machine learning models, enhancing their ability to perform perception-related tasks in a manner akin to human learning.

Secondly, this approach has the potential to drive significant advancements in Artificial Intelligence (AI) and machine learning. By modeling VPL using various DNN architectures, we can improve AI systems' capabilities in perception-related tasks, allowing them to process and learn from sensory information in ways that more closely resemble human learning. Additionally, a better model of VPL can have practical applications in vision rehabilitation, offering new strategies for treating visual impairments and optimizing therapies for those recovering from visual system injuries or diseases. Furthermore, investigating the transferability aspect of perceptual learning could revolutionize AI models by enhancing their ability to apply learned knowledge to new, unencountered tasks, thereby boosting computational efficiency and model generalization.

While the DNN model shows a remarkable similarity to neuronal and behavioral data, it's important to acknowledge that several decisions made regarding the training paradigm, noise

injection, learning rule, and learning rate could have significantly influenced the results. Initially, the network training utilized fixed orientation angle separations, which deviates from the standard method commonly employed in numerous VPL studies. Furthermore, the learning rule Stochastic Gradient Descent (SGD) does not align well with more biologically plausible Hebbian-like learning methods [22,23], although more biologically plausible versions have been proposed 40.

In conclusion, while the DNN seems to be promising to provide a theoretical explanation for many aspects of VPL, further investigation of different network architectures and learning rules is required. Moreover, to thoroughly understand the limitations of the DNN in shedding light on this phenomenon, it is crucial to compare the model's performance with empirical data from human and animal studies.

References

1. James, W. The Principles of Psychology (Henry Holt, 1890).
2. Lu, Z.-L. & Dosher, B. A. Current Directions in Visual Perceptual Learning. Nature Reviews Psychology 1, 654–668 (2022).
3. Karni, A. & Sagi, D. Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. Proc. Natl Acad. Sci. USA 88, 4966–4970 (1991).
4. Zhou, Y. et al. Perceptual learning improves contrast sensitivity and visual acuity in adults with anisometropic amblyopia. Vis. Res. 46, 739–750 (2006).
5. Fiorentini, A. & Berardi, N. Perceptual learning specific for orientation and spatial frequency. *Nature* **287,** 43–44 (1980).
6. Ball, K. & Sekuler, R. A specific and enduring improvement in visual motion discrimination. *Science* **218**, 697–698 (1982).
7. Karni, A. & Sagi, D. Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. *Proc. Natl Acad. Sci. USA* **88**, 4966–4970 (1991)
8. Ahissar, M. & Hochstein, S. Task difficulty and the specificity of perceptual learning. *Nature* **387**, 401–406 (1997).
9. Mastropasqua, T., Galliussi, J., Pascucci, D. & Turatto, M. Location Transfer of Perceptual Learning: Passive Stimulation and double training. *Vision Research* **108,** 93–102 (2015).
10. Dosher, B. A. & Lu, Z.-L. Mechanisms of perceptual learning. Vis. Res. 39, 3197–3221 (1999).
11. Dosher, B. A. & Lu, Z.-L. Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. Proc. Natl Acad. Sci. USA 95, 13988–13993 (1998).
12. Watanabe, T. & Sasaki, Y. Perceptual learning: toward a comprehensive theory. Annu. Rev. Psychol. 66, 197–221 (2015).
13. Ahissar, M. & Hochstein, S. The reverse hierarchy theory of visual perceptual learning. Trends Cogn. Sci. 8, 457–464 (2004).
14. Dosher, B. & Lu, Z.-L. Perceptual Learning: How Experience Shapes Visual Perception (MIT Press, 2020).
15. Fahle, M. & Morgan, M. No transfer of perceptual learning between similar stimuli in the same retinal position. Curr. Biol. 6, 292–297 (1996).
16. Schoups, A., Vogels, R., Qian, N. & Orban, G. Practising orientation identification improves orientation coding in V1 neurons. Nature 412, 549–553 (2001).
17. Law, C.-T. & Gold, J. I. Neural correlates of perceptual learning in a sensory- motor, but not a sensory, cortical area. Nat. Neurosci. 11, 505–513 (2008).
18. Saxe, A., Nelli, S. & Summerfield, C. If deep learning is the answer, what is the question? *Nature Reviews Neuroscience* **22,** 55–67 (2020).
19. Law, C.-T. & Gold, J. I. Neural correlates of perceptual learning in a sensory- motor, but not a sensory, cortical area. Nat. Neurosci. 11, 505–513 (2008)
20. Shibata, K., Watanabe, T., Sasaki, Y. & Kawato, M. Perceptual learning incepted by decoded fMRI neurofeedback without stimulus presentation. Science 334, 1413–1415 (2011).
21. Petrov, A. A., Dosher, B. A. & Lu, Z.-L. The dynamics of perceptual learning: an incremental reweighting model. Psychol. Rev. 112, 715–743 (2005).
22. Sotiropoulos, G., Seitz, A. R. & Seriès, P. Perceptual learning in visual hyperacuity: a reweighting model. Vis. Res. 51, 585–599 (2011).
23. Dosher, B. A., Jeter, P., Liu, J. & Lu, Z.-L. An integrated reweighting theory of perceptual learning. Proc. Natl Acad. Sci. USA 110, 13678–13683 (2013).
24. Herzog, M. H., Aberg, K. C., Frémaux, N., Gerstner, W. & Sprekeler, H. Perceptual Learning, roving and the unsupervised bias. Vision Research 61, 95–99 (2012).
25. Zhaoping, L., Herzog, M. H. & Dayan, P. Nonlinear ideal observation and recurrent preprocessing in perceptual learning. *Netw. Comput. Neural Syst.* **14**, 233–247 (2003).
26. Wenliang, L. K. & Seitz, A. R. Deep neural networks for modeling visual perceptual learning. *J. Neurosci.* **38**, 6028–6044 (2018).
27. Yamins, D. L. & DiCarlo, J. J. Eight open questions inthe computational modeling of higher sensory cortex. Curr. Opin. Neurobiol. 37, 114–120 (2016).
28. Khaligh-Razavi SM, Kriegeskorte N) Deep supervised, but not unsupervised, models may explain IT cortical representation. PLoS Comput Biol 10:e1003915 (2014).

29. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Paper presented at Advances in Neural Information Processing Systems 25 (NIPS 2012), Lake Tahoe, NV, December. (2012).
30. Zeiler, M. D. & Fergus, R. Visualizing and understanding Convolutional Networks. *Computer Vision – ECCV 2014* 818–833 (2014).
31. Bakhtiari, S. Can deep learning model perceptual learning? *The Journal of Neuroscience* **39,** 194–196 (2019).
32. Cohen, G. & Weinshall, D. Hidden layers in perceptual learning. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017).
33.Goldstone, R. L. Perceptual learning. *Annual Review of Psychology* **49,** 585–612 (1998).
34. Felleman, D. J. & Van Essen, D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex* **1,** 1–47 (1991).
35. Guclu, U. & van Gerven, M. A. Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *Journal of Neuroscience* **35,** 10005–10014 (2015).
36. Sherstinsky, A. Fundamentals of Recurrent Neural Network (RNN) and long short-term memory (LSTM) network. Physica D: Nonlinear Phenomena 404, 132306 (2020).
37. Cho, K. *et al.* Learning phrase representations using RNN encoder–decoder for statistical machine translation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (2014). doi:10.3115/v1/d14-1179
38. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. Nature 323, 533–536 (1986).
39. Green, C. S., Pouget, A., & Bavelier, D. (2010). Improved probabilistic inference as a general learning mechanism with action video games. Current Biology, 20(17), 1573–1579.
40. Lillicrap, T. P., Cownden, D., Tweed, D. B. & Akerman, C. J. Random synaptic feedback weights support error backpropagation for deep learning. Nature Communications 7, (2016).
41. .Recurrent neural networks cheatsheet star. *CS 230 - Recurrent Neural Networks Cheatsheet* Available at: https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-recurrent-neural-networks.