Hidden Layers in Perceptual Learning

By

GAD COHEN LEV-ARI

Under the supervision of

PROF. DAPHNA WEINSHALL

THE HEBREW
UNIVERSITY
OF JERUSALEM

Faculty of Computer Science and Engineering
THE HEBREW UNIVERSITY OF JERUSALEM

A thesis submitted to the Hebrew University of Jerusalem as a partial fulfillment of the requirements of the degree of MASTER OF SCIENCE in the Faculty of Computer Science and Engineering.

APRIL 2020

Abstract

Studies in visual perceptual learning investigate the way human performance improves with practice, in the context of relatively simple (and therefore more manageable) visual tasks. Building on the powerful tools currently available for the training of Convolution Neural Networks (CNN), networks whose original architecture was inspired by the visual system, we revisited some of the open computational questions in perceptual learning. We first replicated two representative sets of perceptual learning experiments by training a shallow CNN to perform the relevant tasks. These networks qualitatively showed most of the characteristic behavior observed in perceptual learning, including the hallmark phenomena of specificity and its various manifestations in the forms of transfer or partial transfer, and learning enabling. We next analyzed the dynamics of weight modifications in the networks, identifying patterns which appeared to be instrumental for the transfer (or generalization) of learned skills from one task to another in the simulated networks. These patterns may identify ways by which the domain of search in the parameter space during network retraining can be significantly reduced, thereby accomplishing knowledge transfer.

Dedication and acknowledgements

I was fortunate to follow my heart and research a subject that fascinated and intrigued me.
To my supervisor, Professor Daphna Weinshall, for her inspiration, turning this journey into an endless thirst in pursuit of knowledge and questions. Thank you for your guidance and support. It was a great privilege to be your student.
To my wife Inbal and my daughter Gaia, you add so much joy and light to my life. Thank you for your support, encouragement, and patience throughout this project.
To caffeine, for helping me go through many a long night in the lab.

TABLE OF CONTENTS

List of Figures

Figure **Page**

CHAPTER 1

INTRODUCTION

Performance in relatively simple perceptual tasks is known to be affected by practice. The study of this process revealed some surprising results [16], providing a window into human learning mechanisms. One of the most striking results has been the repeated observation that many of the acquired skills are specific to low level properties of the stimuli (*e.g.*, orientation), and do not *transfer* (*e.g.*, do not generalize to other orientations). These and other results were used to constrain computational modeling of human perceptual learning, as briefly reviewed Section 2.1.

In this paper we revisit these computational studies in the context of recent advances in deep learning, and specifically model the learner by a Convolution Neural Network (CNN). This modeling choice is justified by the resemblance between the CNN architecture and the organization of low level visual areas in the brain. In Section 3 we describe the simulations of representative perceptual learning experiments, where the learner is a shallow CNN, and investigate its emerging properties. These properties are directly compared with the actual perceptual learning results in a qualitative manner. In Section 4 we analyze the learning process. We track the dynamics of weight modifications as learning proceeds, and identify patterns of change which facilitate subsequent learning sessions, *i.e.*enable learning transfer. This facilitation is likely achieved by reducing the search space when re-training the network.

The model we investigate here is a relatively shallow Convolution Neural Network (CNN) with two hidden conv-pool layers (effectively the learned features), and one output layer that integrates the responses of the features using modifiable weights. The main difference with respect to previous modeling attempts, reviewed below, is that this is a generic model, which resembles the visual system in its pipeline hierarchical structure (although it is not a physiologically accurate model of the visual processing areas). It is a general learning machine that learns visual features

from scratch, as well as decision classifiers. Thus it can be used to investigate the relative contribution of features and classification weights to the learning process.

CHAPTER 2

BACKGROUND

Extraordinary cognitive learning abilities, arguably superior to any other living creature, have allowed human beings to achieve incredible things. Here we focus on the less glamorous but related ability of perceptual learning, where learners improve their perceptual skills (and specifically visual skills) as a result of practice and training.

2.1 Generalization in perceptual learning

Perceptual learning studies usually measure the effect of practice on the performance of simple visual tasks, such as motion direction discrimination or line orientation detection. Typically the learner is given feedback, but perceptual learning is known to take place also without direct feedback.

Early studies showed improvement in sensitivity of basic (low level) visual tasks, as basic as hyper-acuity. It was soon shown that most of these improvements were rather specific, selective to stimulus orientation, spatial-frequency, and retinal location [3, 6]. This seemed to imply that learning-related modulations were taking place in early areas of visual processing, indicating somewhat unexpected plasticity in the adult brain. Thus, the issue of learning *Specificity* became central to the study of perceptual learning, with evidence accumulating for the lack of *Transfer*, namely, perceptual learning typically would not lead to improved performance in the same task when slightly modified (*e.g.*, by shifting the stimulus to a different retinal position). A central question emerged [16]: "does learning involve rewiring of neurons in early visual areas, or can it all be explained by improved efficiency in the readout of unchanged early neuronal representations?". This question provided a central motivation for the current study.

The question of learning *Specificity* continued to inspire additional studies, further complicating the picture. Thus learning *Specificity* was shown to be correlated with difficult perceptual

tasks. In easier tasks, *e.g.*, in discrimination tasks involving stimuli with high SNR, generalization was sometimes seen by way of immediate improvement in the novel task, or a shorter learning period [9, 12]. Moreover, a new phenomenon called learning *Enabling*, or Eureka, was reported. This time, when the difficulty of the perceptual task was manipulated, a new form of transfer was observed [2]: after training with an easy perceptual task (*e.g.*, using high SNR stimulus), observers were suddenly able to learn the corresponding difficult condition (low SNR) they had previously been unable to learn. This is reminiscent of similar phenomena in cognitive learning, and the related concept of *curriculum learning* [4].

Two prominent computational models of perceptual learning were developed to explain this pattern of results. The reverse hierarchy theory [1] postulates a hierarchical architecture, where learning is governed by a top-down (rather than the customary bottom-up) information flow. Specifically, learning is first achieved at some rather abstract high level layer, which is task specific; only later, if and when necessary, further learning is achieved at lower level layers which correspond more directly to stimuli processing and the computation of feature maps. The reweighting model [9] assumes a shallow network, and seeks to explain all observed perceptual learning phenomena based on the reweighting of an unchanged set of features. Thus, while not postulating a reverse learning order, this model also looks for the primary loci of modulations (or learning) at some integration level - where the the task-relevant decision (or classification) is taking place, rather than at the feature level. However, since there seems to be neurophysiological evidence that perceptual learning corresponds to changes in low-level visual processing areas as well [16], this theory is not fully satisfying.

2.2 Convolution Neural Networks

Convolution Neural Network (CNN) is a model whose architecture is based on the neocognitron model [7, 8], whose architecture in turn was inspired to a large extent by our understanding of the early visual system as it took shape in the 1960's and 70's. Each layer in the network computes a number of feature maps (or channels), where each channel corresponds to a certain filter which convolves with patches in the image, in a manner similar to passing a sliding window. In addition to a convolution sub-layer, each layer includes additional operations (sub-layers). Some correspond with known operations in the visual system, like max-pooling which computes the maximal response over a small window defined over the convolution sub-layer, or Rectified Linear Units (ReLU) which correspond to a non-linear operation on the responses, nulling out all negative responses. Additional operations which are not necessarily biologically motivated have been added along the way to artificial CNNs to increase their power.

Current CNN models are based on models developed in the 1980's [15]. Unlike the neocognitron model, these were learning machines which modified the weights of the network based on propagating the error signal from the output layer all the way to the input layer. The back-

propagation model was not biologically motivated, in itself being equivalent to gradient descent of a natural loss function in order to update the network's weights. However, the error signal derived from the loss function approximates the Hebbian rule to some extent, and therefore this model was used to investigate biological learning mechanisms (*e.g.*, [13, 14]).

In recent years we have seen an ever increasing use of CNN models for practical applications, which now dominate the state of the art in many computer vision sub-fields like object recognition (*e.g.*, [10, 11, 17, 18]). This was made possible by the availability of *big* data - a large number of images collected into publicly available databases, which were used to train deep CNN models more effectively than ever before. In this paper we take advantage of this opportunity - the widespread and the availability of very effective tools to train CNNs, in order to revisit questions in perceptual learning, and investigate the learning of CNN models in the context of perceptual learning tasks.

CHAPTER 3

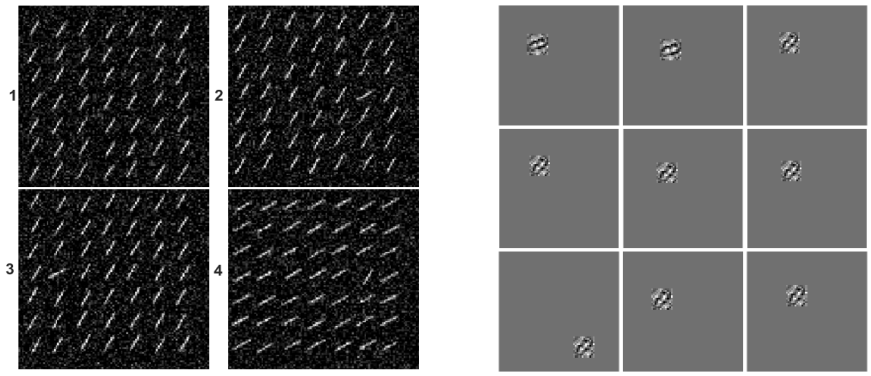
THE EMERGENCE OF *SPECIFICITY* AND *ENABLING* IN NETWORK TRAINING

We focused our investigation on two hallmark characteristics of perceptual learning (see discussion in the previous section): *Specificity* and *Enabling*.

3.1 Methods

CNN network We trained a two layer CNN using vanilla SGD, with a fixed learning rate and batch size (50). The network was initialized using a weights vector out of a fixed randomly generated set. The first layer included 6 channels, each with 5x5 conv, stride 1, ReLU and pooling. The second layer included 16 channels, each with 5x5 conv, stride 1 ReLU and pooing. This layer was connected to the output of the network, 2 output neurons. Overall, the network had 8846 parameters including weights and biases. Each experiment was repeated 32 times, training a new network each time. The plots in this section show average results across all repetitions, with the appropriate standard deviation (*std*). In all conditions the network was trained long enough with the initial learning task to achieve convergence.

Visual stimuli All images were grayscale of size 108×108, with added Gaussian noise. For each experimental setup (*e.g.*, pop-out task with $\theta = 8^o$), a dataset comprising of 3000 train and 1500 test images was created. In the pop-out experiment (Section 3.2), each image showed a 7×7 array of parallel oriented line segments. The experiment involved 4 conditions illustrated in Fig. 3.1a: no pop-out (a1), pop-out in either the right or left location (a2 and a3 respectively), and the same two conditions with swapped orientations (a4). Each image was randomly translated



a) b)

Figure 3.1: a) Pop-out discrimination task with differently oriented line segments in two locations. b) Orientation discrimination task.

by up to 4 pixels in each direction along the x and y axes. In the orientation discrimination experiment (Section 3.3), images were grayscale of size 68×68. A tilted Gabor patch was shown near one of the 4 corners of the image (see Fig. 3.1b), after which it was rotated by a fixed angle in either clockwise or counter clockwise direction.

3.2 Specificity and Enabling in pop-out tasks

We first replicated the task described in [2], training the Convolution Neural Network described above to perform the discrimination task described next. We specifically looked for the emerging properties of learning as they are related to the results reported in [2].

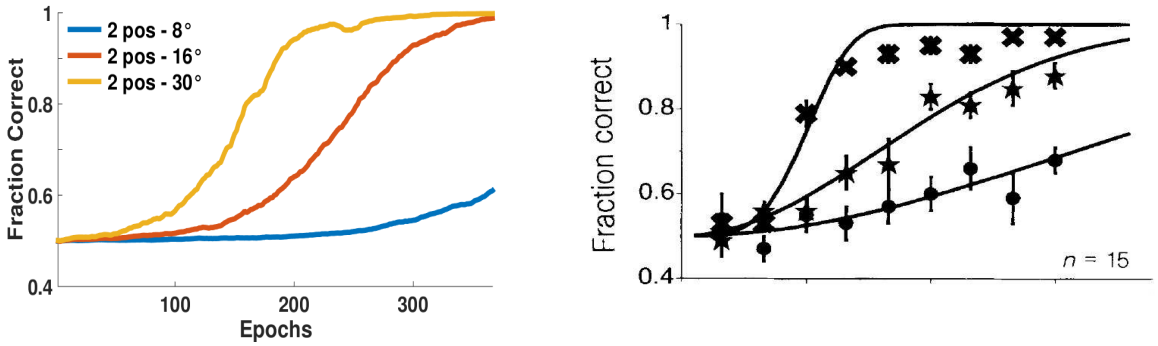
3.2.1 Experimental setup:

The relevant perceptual task was similar to the task described in Fig. 3.1a. The task required to determine whether the displayed array contained an odd line segment (as in Fig. 3.1-a2) or not (as in Fig. 3.1-a1). The angular difference between the odd segment and the remaining segments θ controlled the level of difficulty (or SNR) of the task. In [2] there was an additional parameter which controlled task difficulty - the SOA (Stimulus Onset Asynchrony); this physiological parameter was not replicated in our simulations, as it required physiological modeling beyond the scope of this work.

3.2.2 Learning as a function of task difficulty:

Our CNN model was trained to perform the discrimination of an odd element as illustrated in Fig. 3.1a. The results (percent correct) are shown in Fig. 3.2. We note that it took about 300

epochs to achieve 0.3% test error with the easy task (30^o), and about 1500 epochs for 3.1% test error with the hard task (8^o).



a) b)

Figure 3.2: Results of training to perform the relevant discrimination task with a few levels of difficulty. a) CNN network: the difficulty level is controlled by the angle difference between the odd line segment and the background segments: $8^o, 16^o, 30^o$. b) [2] Fig. 1b. The two plots show qualitatively similar behavior - longer training time for more difficult discrimination tasks. We note, however, that both difficulty level and improvement were measured quite differently in a) and b).

We repeated the manipulation of task difficulty as reported in [2] by changing the angle difference between the odd line segment and the background segments. We used 3 conditions: 8^o (hardest), 16^o (intermediate), and 30^o (easiest). As described in Section 2.1, perceptual learning experiments revealed different characteristics when the task difficulty was manipulated. In the extreme cases, no learning was seen with very difficult tasks, while learning was fast or even immediate with easy tasks. This pattern emerged in our experiments as well, as can be seen in Fig. 3.2a.

3.2.3 Learning generalization (*Transfer*):

To check the *Specificity* of learning, the trained network was tested on the same discrimination task as described above. As in [2], we first tested for transfer in the same image position. Without changing the difficulty of the task, this was originally accomplished by using the swapped task (see Fig. 3.1-a4). Results are shown in Fig. 3.3, showing similar qualitative behavior of larger transfer for easier task, and vice versa.

Next, we investigated the transfer of learning to similar stimuli shown in different image locations. Now the odd line segment was shown in a different image location, in the adjacent grid position to the left or right of the trained location. The results (percent correct) are shown in Fig. 3.4. Once again, we see similar qualitative behavior in the simulated and biological learning outcomes, namely, training with the easy task (30^o angle difference) transferred (or generalized)

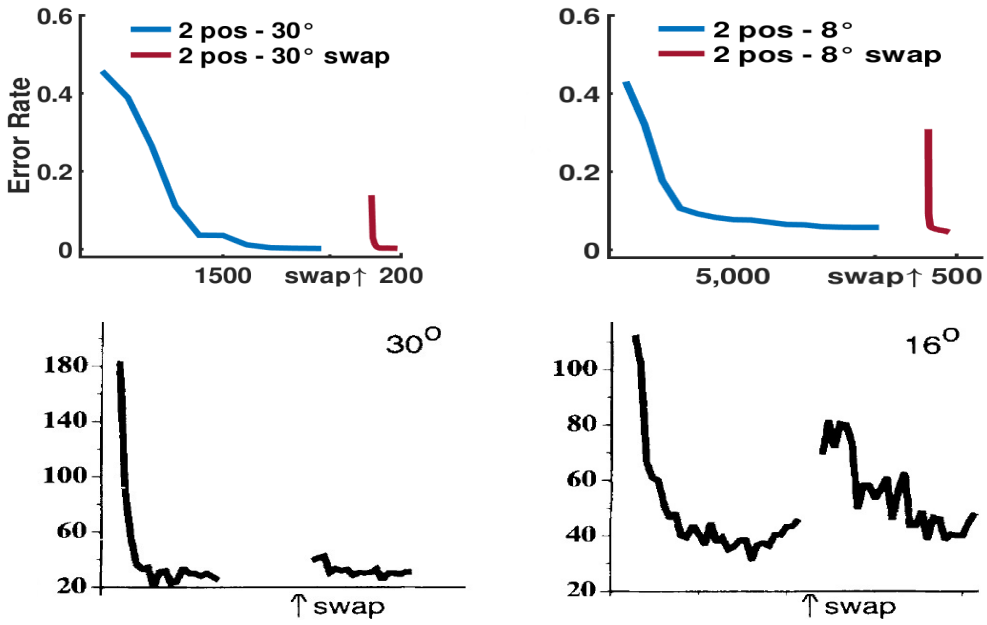


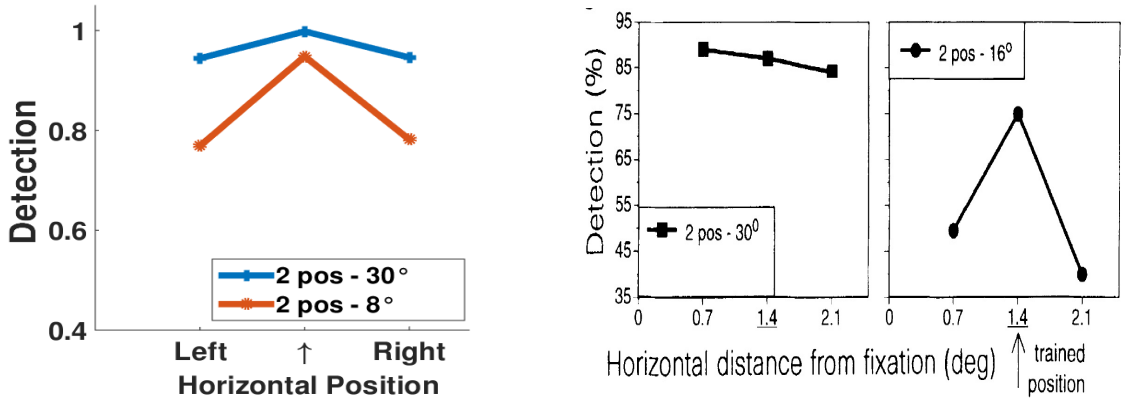
Figure 3.3: Transfer in the same image position between two similar tasks of the same difficulty level. Top: CNN network, transfer in the easier task with $\theta = 30^o$ (left), and the more difficult task with $\theta = 8^o$ (right). Bottom: a qualitatively similar phenomenon was reported in [2] Fig. 2a.

quite significantly to new locations in the visual field, while training with the more difficult task (16^o angle difference) hardly affected performance in the new locations. We see one minor difference: humans transferred more readily to the position between the two trained locations, while our network did not display this preference.

3.2.4 Learning *enabling* (Eureka):

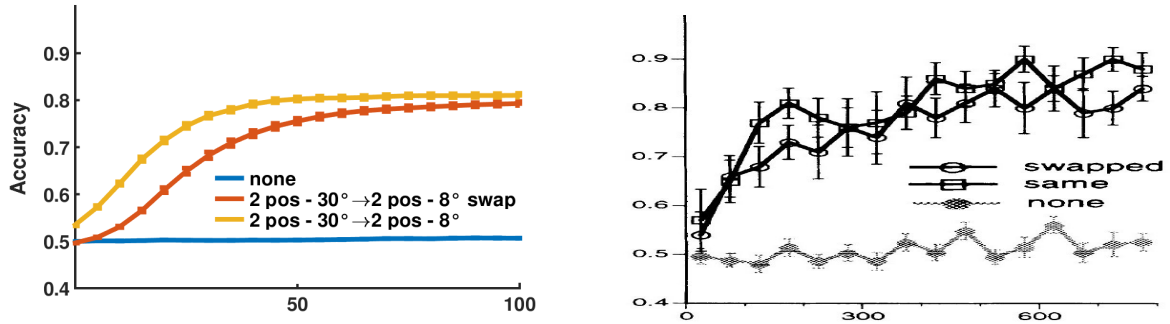
Finally, we investigated the *enabling* of learning. In very difficult tasks, it may be the case that participants do not improve with practice, either because they cannot learn the task, or because they learn it too slowly for the outcome to be measurable. In [2] it was shown that after a short training session with the easy task (or a single long exposure), there was a sudden change and observers began to improve very fast while learning the difficult task. This phenomenon was termed in [2] the *Enabling* of learning; when the change was instantaneous, it was called Eureka.

In our simulations a similar phenomenon emerged, where training with the hard task took significantly longer than training with the easy task. If, however, the network was first trained with the easy task, subsequent training with the hard task became very fast with some instantaneous improvement. Improvement was evident in both accuracy and the speed of convergence, achieving a test error smaller by 1.5% in less than half the number of accumulated iterations, in agreement with the results reported in [4]. These results are shown in Fig. 3.5.



a) b)

Figure 3.4: Results of testing the trained network with the same discrimination task, but where the odd element was shifted to a different location from the trained one. a) CNN network. b) A qualitatively similar phenomenon was reported in [2] Fig. 2c.



a) b)

Figure 3.5: a) CNN: training to detect an odd segment with $\theta = 8^o$ (a difficult task) takes a very long time (bottom line). However, when the network has been first trained with a similar easy task using $\theta = 30^o$, instantaneous improvement is seen, followed by speedy learning. The same happens even if the orientations of the background and odd elements are swapped between the odd element and the background. b) A qualitatively similar phenomenon was reported in [2] Fig. 5b.

3.3 Specificity in orientation discrimination

Our second representative perceptual learning task is inspired by the one used in [9].

3.3.1 Experimental setup:

The relevant perceptual task was similar to the task described in Fig. 3.1b. The task required to determine whether the stimulus rotated clockwise or counter-clockwise. The rotation angle θ controlled the level of difficulty (or SNR) of the task. In each experimental session, only two corners along one image diagonal were used for training, while the other two corners were used to probe for location transfer. In order to test for orientation transfer, another oriented stimulus was presented, corresponding to a rotation by 90^o of the stimulus presented during training.

3.3.2 Learning precision:

In this task, as in the previous task described in Section 3.2, learning characteristics strongly depended on task difficulty. Like before, learning was fast for easier tasks (with relatively large orientation difference θ), and much slower for difficult tasks (small θ). However, [9] observed that what mattered was the difficulty of the target (test) task, a characteristic which they called *task precision*, and not the difficulty of the initial training task, see Fig. 3.6-bottom[1].

In accordance, we repeated this experiments and simulated the 4 relevant conditions. In the first 2 conditions shown in Fig. 3.6-left, we trained the network with either a difficult training task ($\theta = 16^o$) or a relatively easy task ($\theta = 30^o$), and tested transfer to the easy condition of $\theta = 30^o$. In the last 2 conditions shown in Fig. 3.6-right, we used the same training but tested instead transfer to the difficult condition of $\theta = 16^o$. The simulation results show similar qualitative behavior an in the perceptual learning task, although in our simulations the difficulty of the training task also played a role in the efficacy of transfer, and transfer was not instantaneous even with the easy task.

3.3.3 Learning generalization (*Transfer*):

[5] investigated the question of learning interference. Specifically, they compared the transfer to a new task in 3 conditions: when changing only stimulus absolute orientation, changing only the image location where the stimulus was present, or changing both. Based on the results described above, we expect that if the task is not too difficult there should be some transfer in all 3 conditions. Interestingly, [5] observed that when the stimulus absolute orientation was changed, transfer was stronger when it was presented in a new location, as compared to presenting the modified stimulus in the same location. This may indicate the existence of some destructive interference between the learning of different basic features in the same image location.

We investigated whether the same can be seen in our model. Like before, we trained the CNN using the original stimulus, and then tested performance using the same 3 manipulations: change of stimuli absolute orientations while keeping the orientation difference fixed, change

[1]The difference between the initial learning curves (denoted 'Training') in the two conditions, shown on the left side of each panel of Fig. 3.6-bottom, are due to random differences between subjects, since different subjects participated in the two experiments. The learning task, however, was identical.

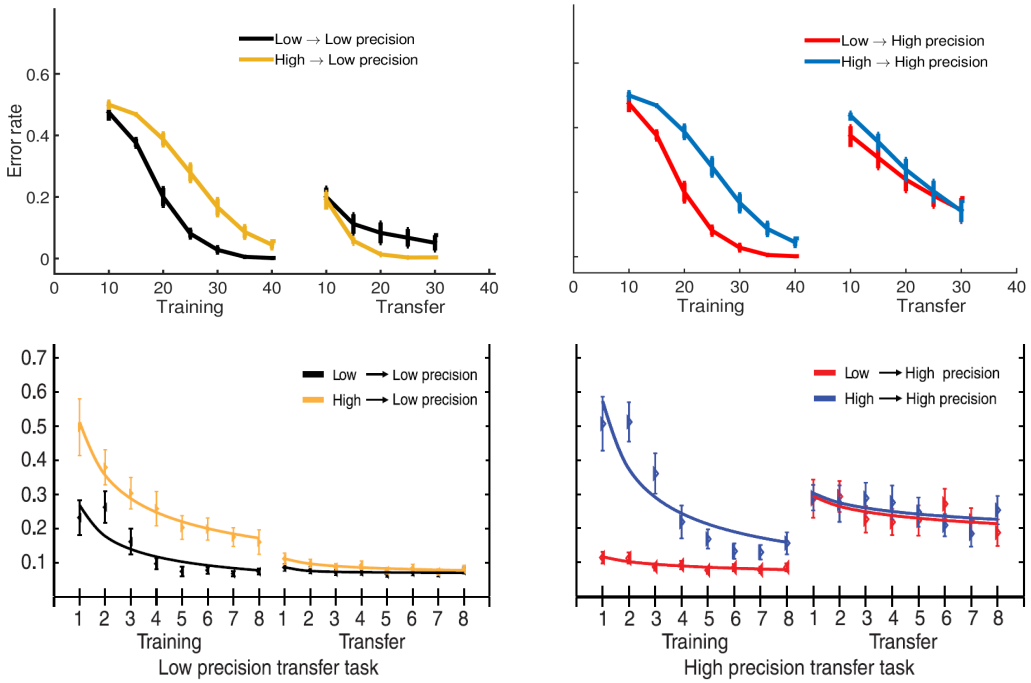


Figure 3.6: The dependence of transfer on the difficulty of the training task vs. the difficulty of the target (test) task. Transfer results are shown separately depending on the difficulty of the transfer (test) task. Left: easy test task; right: difficult test task. Top: CNN simulations (see explanation in text). Bottom: a qualitatively similar phenomenon was reported in [9] Fig. 2.

of location, and change of both. The results are shown in Fig. 3.7 for $\theta = 30^o$, the easy task, in qualitative agreement with the perceptual learning results. (The relevant results reported in [5] only mention the easy task of $\theta = 30^o$. In our simulations this phenomenon was also evident with $\theta = 16^o$, a harder task.)

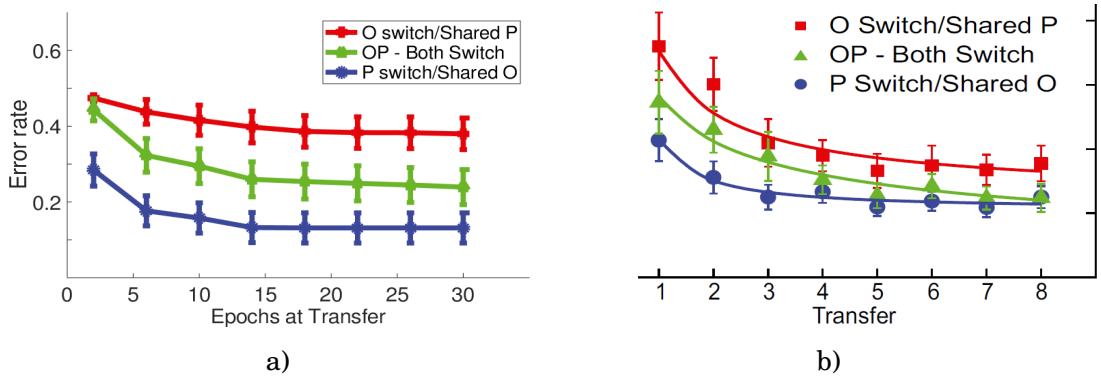


Figure 3.7: Transfer to a new location and new stimulus orientation. a) CNN, using $\theta = 30^o$. b) A qualitatively similar phenomenon was reported in [5] Fig. 2.

3.4 Discussion

Simulating the task described in [2], the learning outcome showed qualitatively similar characteristics to the results described in human perceptual learning. Specifically, as the difficulty of the perceptual task decreased, learning time decreased. Learning transferred (or generalized) to similar tasks much more readily when the task was easy. Finally, training with an easy task enabled subsequent training with difficult tasks by significantly shortening the time needed for learning. These qualitative results capture the essence of almost all the observations reported in [2].

Simulating the tasks described in [5, 9], we saw parallels to additional effects: transfer depended more strongly on the difficulty of the target task rather than the training task; transfer was more effective when image location was changed as compared to when stimulus orientation was changed; and finally, we observed some learning interference when teaching similar tasks with different stimuli in the same image location. The latter observation did not depend on the number of channels in each layer.

CHAPTER 4

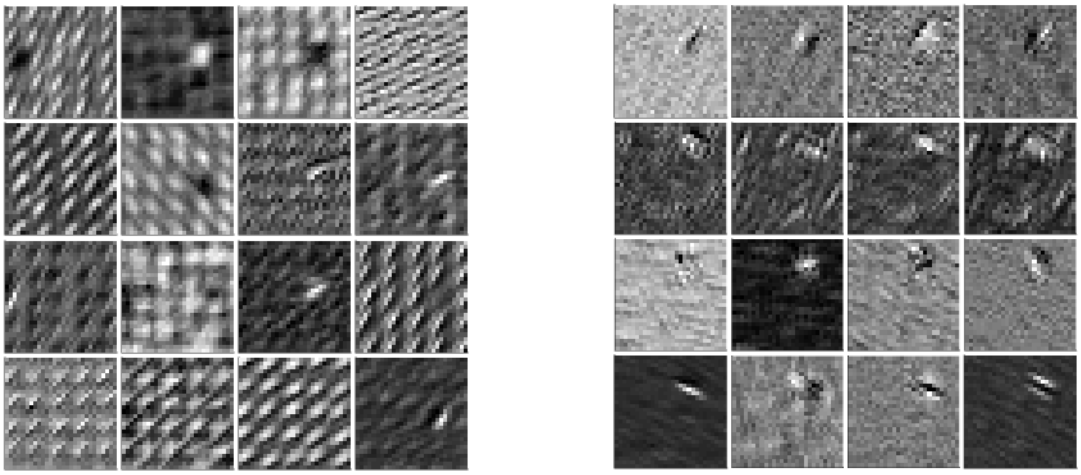
DYNAMICS OF WEIGHT MODIFICATIONS

In the previous section we showed how our simulations, when training a shallow CNN to perform visual discrimination tasks, were able to replicate many of the phenomena observed repeatedly in perceptual learning experiments. In this section we go a step further, and investigate the network's weights modification patterns which underlie these phenomena.

4.1 Feature maps and filters

We first note that, in both experiments, the CNN learned simple edge-like features matching the displayed stimuli, as can be readily seen in Fig. 4.1. More specifically, Fig. 4.1a shows typical patterns of activation in the channels of the second CNN layer in a pop-out experiment. In many of these channels, the location of the odd element appears as a highlighted region or a gap in the background pattern.

In Fig. 4.1b we see detectors of line segments matching the locations and orientations of the training segments (first and last row). Interestingly, features learned for one task with orientation o_1 were able to partially detect the target segment in another task with orientation o_2 (second row). Slight weight modifications, limited to the bias elements only, improved the detection somewhat (compare the third row with the fourth row). This may explain the orientation transfer we see in this condition, where re-training with a new orientation is much faster than training from scratch in the new orientation.



a) b)

Figure 4.1: Representative feature maps. a) Pop-out detection tasks (Section 3.2). The first 3 columns from left to right correspond to a different angle difference: 8^o, 16^o and 30^o. The 4^{th} column corresponds to a task with swapped orientations. In each column, 4 arbitrary channels are shown. b) Orientation discrimination tasks (Section 3.3). Each column corresponds to a different channel. The first row shows 4 feature maps from 4 channels when using orientation o_1 during both train and test. The last row shows the same channels when using o_2, the angle obtained when rotating o_1 by 90^o, during both train and test. The 2^{nd} row shows the same channels when o_1 is used for training the network, and o_2 is used for the computation of the feature map. The 3^{rd} row shows the same channels when o_1 is used for training the network, after which the biases are copied from the network of the last row, and finally o_2 is used for the computation of the feature map.

4.2 Generalization to new locations

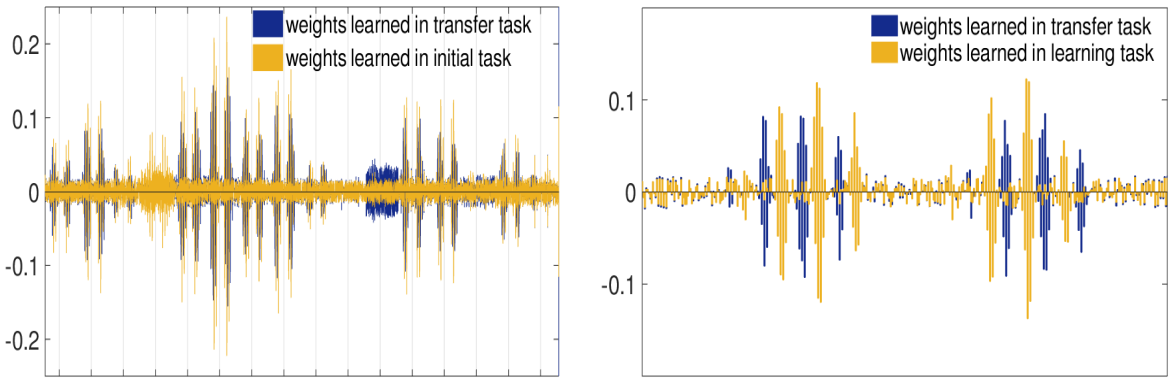
Learning *Transfer* (*i.e.*, generalization) is the term used in the perceptual learning literature to describe the phenomenon where some initial training with a visual task in a certain image location improves performance in a different image location. We replicated two such results, as shown in Fig. 3.4a and Fig. 3.7a.

To investigate *Transfer*, we analyzed the network modifications in the experiment described in Fig. 3.7. Recall that in this setup, a network was first trained with the discrimination of oriented edges in one part of the image (intermediate weight values), and subsequently the network was trained with the same discrimination task in a different image location (final weight values). We Analyzed the changes in the network weights between the intermediate and final states.

First, we noticed that all the significant (normalized) changes occurred in the connections between the last conv-pool layer and the output neurons. These modifications are shown[1] in

[1]In order to visualize the changes in weights, we reshaped the network weights into a $1D$ vector. Thus, for each 2-dimensional filter, the vectorization procedure transformed the matrix of weights to a vector by concatenating the

Fig. 4.2. We see a range of changes in Fig. 4.2a, including inhibition and excitation. However, many of the channels exhibited a specific pattern of changes, as shown for one typical channel in Fig. 4.2b. Here, the pattern of learned weights displays alternating peaks corresponding to the location of the stimuli in the image. This pattern is an artifact of the vectorization procedure.



a) b)

Figure 4.2: a) The weights of the connections between the last conv-pool layer and the output neurons. Faint vertical line separate the weights into 16 groups, each corresponding to weights originating from one channel in the last conv-pool layer. The intermediate weights (yellow) are super-imposed on the final weights (blue). Since the intermediate weights hardly changed in the second phase of training, the final weights are in fact described by the union of the yellow and blue bars. b) Close-up on a typical channel, showing only weights originating from one of the channels in the last conv-pool layer.

Fig. 4.2b should be interpreted as follows: The left pattern corresponds to the left half of the image; here, blue bars correspond to pixels in the top of each column, while yellow bars correspond to pixels in the bottom. The intermediate weights were trained when the edge appeared at the bottom-left part of the image, thus the connections to these locations were amplified (yellow bars). In the second learning phase (the transfer task), the edge appeared at the top-left corner of the image, leading to the amplification of the connections to these pixels (blue bars).[2]

Thus, what we see is that the network maintained the features it had learned in the intermediate phase of training, as seen by the stability of the weights from the first conv-pool layer to the second one. When the same stimulus appeared in a new location, all the network had to do was to modify the weights of the connections between the last conv-pool layer and the output neurons, corresponding to the new image locations. This restriction of the search space, during the second training phase, allowed the network to converge much faster to a good solution for stimuli presented in new locations. In other words, in our network location transfer was the result of re-using features learned previously, having to re-learn only the weights between features and

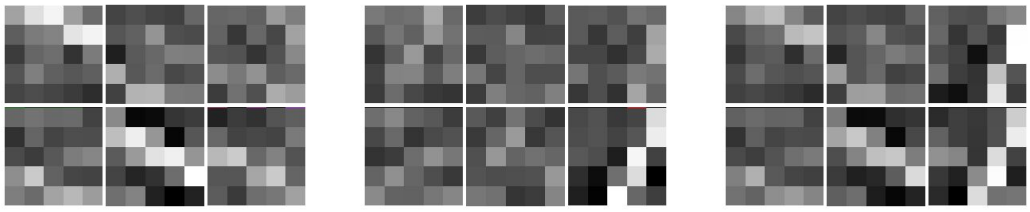
columns of the matrix scanned from left to right.

[2]The right pattern, corresponding to the right half of the image, should be interpreted in a similar manner.

output layer (the readout weights).

4.3 Generalization to new orientations

In the experiments described in Fig. 3.7a, we investigated transfer to new stimulus orientations (as well as new locations). When we examined the dynamics of weight modifications in our network, changing the stimulus orientation while keeping the location constant, we saw a different picture as compared to the one described above. This time all the weights in the network changed, and in particular, new features were being learned, giving rise to weight changes in the first and second conv-pool layers. In the decision layer, connecting the last conv-pool layer to the output neurons, weights were adjusted to match the orientation of the second (transfer) stimulus.



a) b) c)

Figure 4.3: The 6 convolution filters learned in the first conv-pool layer of the CNN when: a) the network was trained to discriminate orientation o_1 only; b) the network was trained to discriminate orientation o_2 only; c) the network was trained to discriminate orientation o_2 after being trained with orientation o_1.

The 6 convolution filters (for each of the 6 channels) learned in the first conv-pool layer of the CNN in one simulation are shown in Fig. 4.3. When training with a single orientation, either o_1 or o_2, some of the emerging filters captured the displayed orientation. Interestingly, when training with one orientation (o_1) and then the other (o_2), new filters appropriate for o_2 appeared in the second phase of learning, but the best filters for the discrimination of o_1 had not been significantly modified. This interesting property, where training with a new task modified primarily channels which had been less significant for the previous task, characterized the emerging filters of our network in all 32 repetitions. We note that in a few repetitions, the CNN failed to learn new appropriate filters, and it also failed to learn the new discrimination task with orientation o_2.

Swapping the line segments: In the pop-out experiment, transfer to a new orientation was investigated by swapping the orientations of the odd segment and the background segments. We inspected the dynamics of weight modifications in the corresponding simulation, as illustrated in Fig. 4.4. In this example, the strong positive weights correspond to a channel (Fig. 4.4a) activated

by the odd segment in the initial learning task. After orientation swap this channel is activated by background segments, and the output neuron indicates whether an input is "non-pop-out". In order for the output neuron to properly classify an input as such, the network suppressed the negative weights emanating from this channel and corresponding to background segments (higher weights in the yellow as compared to the blue curve). Otherwise, their effect, which is summed over all background segments locations, could multiply and decrease the total sum, causing the output neuron to misclassify.

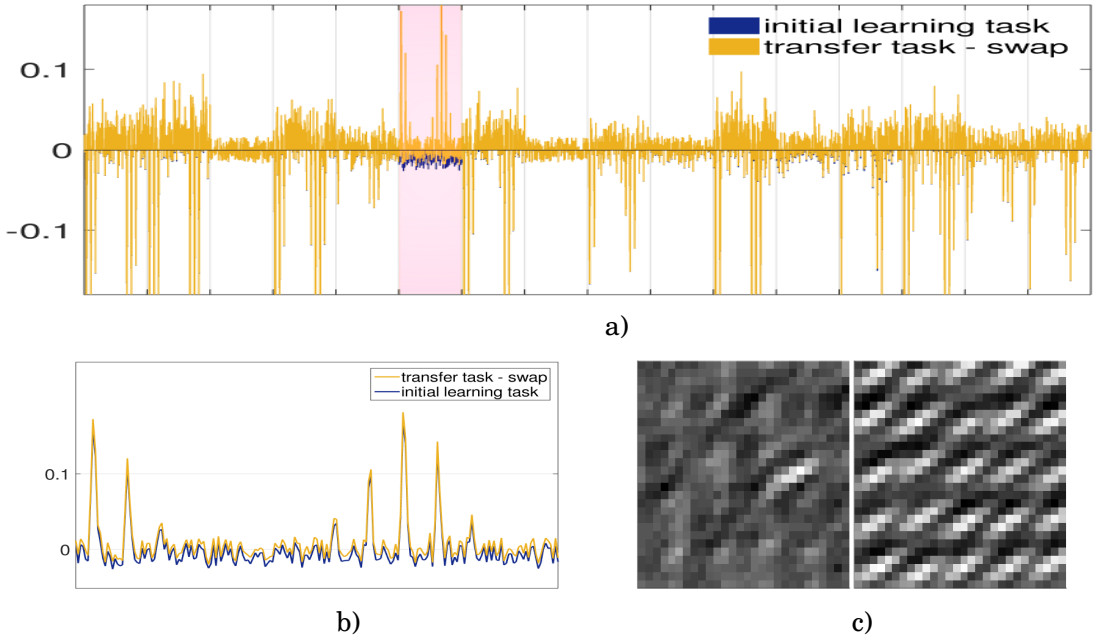


Figure 4.4: a) The weights of the connections between the last conv-pool layer and the output neurons. Showing 16 channels each fenced by two faint vertical bars. The significant changes occurred primarily in the channel marked in pink background. b) Zoom-in on the channel of interest from above. The final weights at the end of re-training with orientation swap (yellow), are super-imposed on the intermediate weights before re-training (blue). c) The feature map of the channel activated by the pop-out edge during the initial task (left). The same channel feature map, at the end of the second learning session after orientation swap (right).

4.4 Enabling

Learning *Enabling* is the term used to describe the phenomenon where initial training with a certain task in an easy condition (high SNR) enables subsequent learning of the same task in a difficult condition (low SNR). We described two such cases in Fig. 3.5a and Fig. 3.6-left. To investigate learning *Enabling*, we analyzed the network modifications in the experiment described in Fig. 3.6. Note that in Fig. 3.6 we used the somewhat confusing terminology introduced in [9]

of *low precision* and *high precision*, which in the following discussion is replaced by the more accurate terminology of high SNR and low SNR respectively.

Once again, we checked the modifications which occurred in the weights of the trained network during the two learning phases. In Fig. 4.5 we show the weights of the network at the end of the first learning phase (end of training with the high SNR task) superimposed on the weights at the end of the second learning phase (end of training with the low SNR task). We first observe that in general, the absolute values of the weights increased substantially, in accordance with our empirical observation that independent training with high SNR and low SNR tasks leads in general to higher absolute weights in the first case as compared to the second. Interestingly, we don't see weights redistribution after the second phase of learning, but what appears like an amplification of the weights: for the most part the weights just got stronger in absolute value while keeping the same sign (see Fig. 4.5-right). We see weight amplification all over the network, including the first and second convolution layers, and the last decision layer. We note that had we trained the network from scratch on the low SNR difficult task, without previously training on the high SNR task, the overall effect would have been the re-distribution of weights rather than weight amplification.

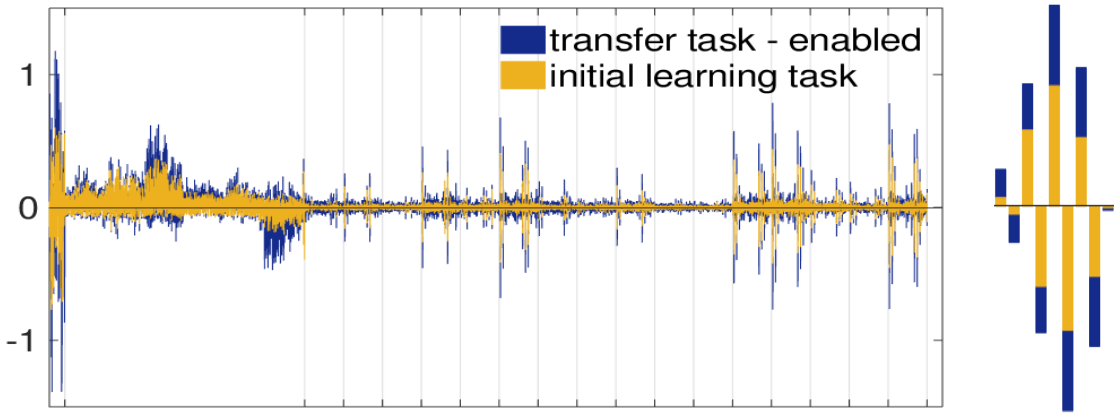


Figure 4.5: Left. Visualization of the weights of the trained network: blue bars show the weights at the end of the process, and super-imposed yellow bars show the intermediate weights after training with the initial easy (high SNR) task. Almost all weight changes correspond to amplification (more positive or more negative values); very few changes correspond to decrease of absolute value or change of sign. Right. To demonstrate this point, we show a close-up on a part of the plot, focusing on the changes in a few of the final weights going from channels to decision neurons. This is typical of all the other weight changes in the network.

Interestingly, it appears like learning *Enabling* in our network was the result of its achieving a certain state, after the first learning phase, which proved to be a powerful initial condition for the second learning phase; from this starting point, weight amplification alone allowed, for the most part, convergence to a good solution in the second task. This reduced the search space

Task	Accuracy	#bits
2 pos - 30^o	99.79%	5.094
2 pos - 16^o	99.40%	5.656
2 pos - 8^o	96.90%	5.7188
8^o enabled	98.43%	5.7188

Table 4.1: Number of bits needed to maintain original accuracy.

significantly, thus leading to much faster convergence, or the enabling of learning.

Finally, we evaluated the networks trained for tasks with different SNR, calculating the minimal number of bits required to store the network's weights without reducing performance by more than 1%, see Table 4.1. Clearly, more bits (or higher precision) were required for harder discrimination tasks. Moreover, we see that the enabled network (row 4 in the Table) reached higher accuracy while requiring the same precision (the same number of bits) as compared with the network trained only with the hard task (row 3).

CHAPTER 5

SUMMARY AND CONCLUSIONS

We described in this paper two sets of results and observations, based on the simulation of perceptual learning experiments. We first trained a shallow Convolution Neural Network to perform these tasks, to be compared with human learners. We were able to show many parallels between the emerging properties in both learning scenarios, especially those concerning learning transfer and enabling.

We then analyzed patterns of weight modifications in the network, identifying characteristic patterns which may have been instrumental for the observed transfer of learning. When the new task occurred in a different image location, the network typically re-used the features learned for the first task, changing only the readout weights in the final classification layer. This agrees with the model proposed in [9]. However, some transfer tasks (such as change in orientation) required the changing of weights across the board, in disagreement with the model proposed in [9]. It may, however, be more consistent with evidence for plasticity in early visual areas during perceptual learning. We further postulate, that increasing the constraints on the architecture, *i.e.*decreasing the number of channels, might increase the interference effect seen in Fig. 3.7, as the filters might need to be unlearned.

Learning enabling, when learning a difficult task becomes possible only after being trained with an easier task, also emerged in our simulations. Specifically, training the network with an easy task left the network in a state where subsequent training with the difficult task converged to excellent performance. This was accomplished by selectively amplifying the weights of the network. This significant reduction of the search space during training could account for the increased learning rate observed as part of learning *Enabling* as suggested in [12]. This explanation of the *Enabling* phenomenon is very different from the explanation proposed in the reverse hierarchy theory [2], which postulated a top-down order of learning. In our model all the

weights of the network were changed simultaneously as a result of error propagation, and still learning *Enabling* emerged.

In conclusion, the results and analysis provided in this paper can be used to rethink the mechanism underlying perceptual learning, and give practical considerations on how to facilitate knowledge transfer in CNN.

BIBLIOGRAPHY

[1] M. AHISSAR AND S. HOCHSTEIN, *The reverse hierarchy theory of visual perceptual learning*, Trends in cognitive sciences, 8 (2004), pp. 457–464.

[2] M. AHISSAR, S. HOCHSTEIN, ET AL., *Task difficulty and the specificity of perceptual learning*, Nature, 387 (1997), pp. 401–406.

[3] K. BALL AND R. SEKULER, *A specific and enduring improvement in visual motion discrimination*, Science, 218 (1982), pp. 697–698.

[4] Y. BENGIO, J. LOURADOUR, R. COLLOBERT, AND J. WESTON, *Curriculum learning*, in Proceedings of the 26th annual international conference on machine learning, ACM, 2009, pp. 41–48.

[5] B. A. DOSHER, P. JETER, J. LIU, AND Z.-L. LU, *An integrated reweighting theory of perceptual learning*, Proceedings of the National Academy of Sciences, 110 (2013), pp. 13678–13683.

[6] A. FIORENTINI AND N. BERARDI, *Perceptual learning specific for orientation and spatial frequency.*, Nature, (1980).

[7] K. FUKUSHIMA, *Neocognitron: A hierarchical neural network capable of visual pattern recognition*, Neural networks, 1 (1988), pp. 119–130.

[8] ——, *Artificial vision by multi-layered neural networks: Neocognitron and its advances*, Neural networks, 37 (2013), pp. 103–119.

[9] P. E. JETER, B. A. DOSHER, A. PETROV, AND Z.-L. LU, *Task precision at transfer determines specificity of perceptual learning*, Journal of Vision, 9 (2009), pp. 1–1.

[10] H. KATAOKA, Y. MIYASHITA, M. HAYASHI, K. IWATA, AND Y. SATOH, *Recognition of transitional action for short-term action prediction using discriminative temporal cnn feature*, in British Machine Vision Conference (BMVC), 2016.

[11] A. KRIZHEVSKY, I. SUTSKEVER, AND G. E. HINTON, *Imagenet classification with deep convolutional neural networks*, in Advances in Neural Information Processing Systems

25, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds., Curran Associates, Inc., 2012, pp. 1097–1105.

[12] Z. LIU AND D. WEINSHALL, *Mechanisms of generalization in perceptual learning*, Vision research, 40 (2000), pp. 97–109.

[13] A. NAVEEN AND M. JOSEPHINE, *Analysis of influences of memory on cognitive load using neural network back propagation algorithm*, International Journal of Data Mining Techniques and Applications, 3 (2014), pp. 336–341.

[14] D. C. PLAUT, J. L. MCCLELLAND, M. S. SEIDENBERG, AND K. PATTERSON, *Understanding normal and impaired word reading: computational principles in quasi-regular domains*, Psychological review, 103 (1996), p. 56.

[15] D. E. RUMELHART, G. E. HINTON, AND R. J. WILLIAMS, *Learning internal representations by error propagation*, tech. rep., DTIC Document, 1985.

[16] D. SAGI, *Perceptual learning in vision research*, Vision research, 51 (2011), pp. 1552–1566.

[17] A. SHARIF RAZAVIAN, H. AZIZPOUR, J. SULLIVAN, AND S. CARLSSON, *Cnn features off-the-shelf: an astounding baseline for recognition*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2014, pp. 806–813.

[18] C. SZEGEDY, W. LIU, Y. JIA, P. SERMANET, S. REED, D. ANGUELOV, D. ERHAN, V. VANHOUCKE, AND A. RABINOVICH, *Going deeper with convolutions*, in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2015.