MODELING THE CORTICAL VISUAL PATHWAYS USING ARTIFICIAL NEURAL NETWORKS

by

Zhixian Han

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science

Department of Psychological Sciences

West Lafayette, Indiana

December 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Anne B. Sereno, Chair

Department of Psychological Sciences

Dr. Sebastien Hélie

Department of Psychological Sciences

Dr. Richard Schweickert

Department of Psychological Sciences

Approved by:

Dr. Kimberly P. Kinzig

ACKNOWLEDGMENTS

I would like to thank Professor Anne Sereno for the suggestions on the design of simulations and the comments on the manuscript.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Although in conventional models of visual information processing, object identity and spatial information are processed separately and independently in ventral and dorsal cortical visual pathways respectively, some recent studies have shown that information about both object's identity (of shape) and space are present in both visual pathways. However, it is still unclear whether the presence of identity and spatial information in both pathways have functional roles or not. In a recent study (Han & Sereno, in press), we have tried to answer this question through computational modeling. Our simulation results suggested that two separate cortical visual pathways for identity and space (1) actively retain information about both identity and space; (2) retain information about identity and space differently; (3) that this differently retained information about identity and space in the two pathways may be necessary to accurately and optimally recognize and localize objects. However, in these simulations, there was only one object in each image. In reality, there may be more than one object in an image. In this master's thesis, I have tried to run visual recognition simulations with two objects in each image. My two object simulations suggest that (1) the two separate cortical visual pathways for identity and space (orientation) still retain information about both identity and space (orientation) when there are two objects in each image; (2) the retained information about identity and space (orientation) in the two pathways may be necessary to accurately and optimally recognize objects' identity and orientation. These results agree with our one object simulation results.

INTRODUCTION

The two-stream hypothesis of the human visual system states that there are two separate pathways in the brain that process different aspects of visual information. In general, the ventral pathway mainly processes object identity information and the dorsal pathway mainly processes spatial information. However, the computational properties of these pathways are still unclear.

As reviewed by Han and Sereno (in press), "It is widely documented in neuropsychological, lesion, and anatomical studies that the human visual system has two distinct cortical pathways (Ungerleider & Mishkin, 1982; Mishkin, Ungerleider, & Macko, 1983; Felleman & Essen, 1991). Further, the ventral pathway primarily processes information important for object recognition (Logothetis & Sheinberg, 1996) while the dorsal pathway primarily processes information related to spatial cognition (Colby & Goldberg, 1999). However, some recent studies have challenged this idea (Konen & Kastner, 2008; Freud, Rosenthal, Ganel, & Avidan, 2015; Freud, Plaut, & Behrmann, 2016; Hong, Yamins, Majaj, & DiCarlo, 2016). Some studies have found that representations associated with shape and location processing are present in both visual streams (Konen &Kastner, 2008; Sereno, Lehky, & Sereno, 2020; Hong et al., 2016). However, it remains unclear whether the representations of shape in dorsal stream and the representations of location in ventral stream are non-task-related or whether they might play a functional rule in spatial cognition and object recognition, respectively. Some findings from fMRI and behavioral studies have suggested that spatial processing that operates at the level of the scene, presumably within the dorsal visual pathway, can contribute to shape processing (Zachariou, Klatzky, & Behrmann, 2014). Another study found that correlated activity between ventral and dorsal visual pathways was higher when people were looking at objects with impossible spatial structures compared with when they were looking at objects with possible structures (Freud et al., 2015), which indicated that dorsal pathway processing might help the brain to recognize objects with impossible structures. Furthermore, Hong et al. (2016) found in neural recordings that spatial information increases along the ventral stream, consistent with prior studies demonstrating spatial properties in later stages of the ventral stream (Nowicka & Ringo, 2000; Lehky, Peng, McAdams, & Sereno, 2008). In addition, Hong et al. (2016) suggest that it is likely that the spatial information in the ventral stream does not come from the dorsal stream, in agreement with previous studies arguing that ventral stream spatial representations are distinct and

independent from dorsal stream spatial encodings (Sereno & Lehky, 2011; Sereno, Sereno, & Lehky, 2014).

The studies mentioned above indicate that distinct and independent representations of shape and space may exist in both visual pathways and might have functional roles. Therefore, in a recent study (Han & Sereno, in press), we attempted to tackle these questions through explicit hypothesis testing using computational models. In our study, we found that identity (of shape) and spatial information processing were present in both simulated ventral and dorsal pathways. These simulated ventral and dorsal pathways were trained to do straightforward object recognition and spatial cognition tasks, respectively. Then, we argued that the possible reason for this is that neural networks need to processidentity (shape) and spacial information independently and differently in order to accurately and optimally recognize and localize objects.

One of the limitations in our study is that there was only one object in each image,but there is usually more than one object to be recognized in real life. In this Master'sthesis, I conduct visual pathway simulations with two objects in each image and try to seewhether the findings we obtained with one object simulations are generalizable to the multiple (two) objects simulations.

In this thesis, I use similar methods that we used in our previous study to conduct simulations (see Han and Sereno (in press)). In order to model the cortical visual pathways and study their computational properties, feed-forward multi-layer convolutional neural networks were used to simulate the functions of the two visual pathways in the brainand multi-layer perceptrons were used to simulate the process of decoding information from recorded neural activities in the brain. All networks were trained using supervised learning and backpropagation. When modeling the cortical visual pathways, for simplicity and control, it is assumed that the all pathways have the same computational structure (the numbers of neurons are the same and the structures of the connections between neurons are the same) and receive the same visual input images. The connection weights between the neurons in the pathways are allowed to be modified with training.

There are two artificial neural networks that are used to simulate the ventral pathway (brain identity network) and the dorsal pathway (brain orientation network). The primary goal of the brain identity network is to distinguish different kinds of objects, whereas the primary goal of the brain orientation network is to determine the orientations of objects necessary for interaction (e.g. grasp).

Black and white images consisting of different kinds of tops, pants, and shoes (Xiao, Rasul, & Vollgraf, 2017) were used to randomly construct the images of objects (see Figure 1). Each

image of an object consisted of one top, one pant, and one shoe. Each object image was embedded in a black background and presented at two different locations and four different orientations (all parts - top, pant, shoe - of the same object were presented with the same orientation, and altogether centered at the selected location). One object is always at the bottom right, and the other object is always at the upper left of theimage. The four possible orientations of the objects are up, down, left, and right. The two objects may have different orientations.

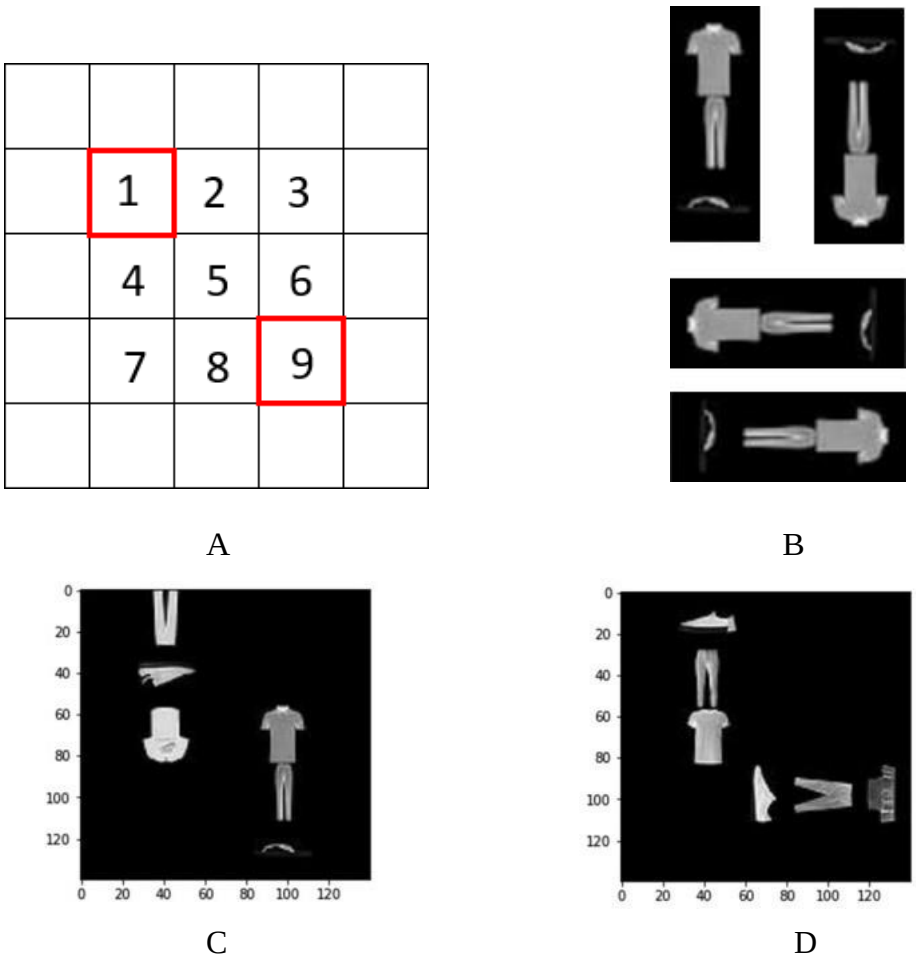

Figure 1. Object image locations and orientations. A. The are nine locations labeled from 1 to 9 in the background. The two objects are always at the upper left (location 1) and bottom right (location 9). B. Four possible orientations of an object image (up, down, left,and right orientations, respectively; going from top to bottom images and, for the first row,left to right images). Note that the alignment of parts within an image are not randomized, are always in the same alignment, and always constrained to the two directions along the long axis. C. An example of a scrambled (S) object with "down" orientation at location 1 and an unscrambled (US) object with "up" orientation at location 9. D. An example of a scrambled (S) object with "up" orientation at location 1 and an unscrambled (US) object with "right" orientation at location 9.

These randomly generated object images with black background were used as visual inputs. For each object in the image, there is 50% probability that the top, the pant, and the shoe were in the "unscrambled" order. That is, the unscrambled order is the normal order of how people are dressed with the pant, but not shoe or top, in the middle. There is 50% probability that the top, the pant, and the shoe were in the "scrambled" order, wherethe order of top, pant, and shoe does not follow the normal order (see Methods for additional details). There is 25% probability for each object to be in each possible orientations (up, down, left, right).

Two artificial neural networks $network_{identity}$ and $network_{orientation}$ were trained to do an identity task (whether the two objects in the image are scrambled or unscrambled), and a spatial task (determine the orientations of the two objects in the image), respectively.The $network_{identity}$ was used to model the ventral pathway, whereas $network_{orientation}$ was used to model the dorsal pathway. These two networks are considered as the brain networks. These brain networks were used to simulate the functions of ventral and dorsal cortical visual pathways in the brain. Various decoder networks were then trained to decode different kinds of information from the later processing stages of these brain networks. These decoder networks were used to simulate the process of decoding information from the recorded neural activities in the brain. It is assumed that if the testing accuracy of the decoder network was higher, then the later processing stages of the brain networks retained more information related to the decoder network's decoding goal.

According to the simulation results, though the $network_{orientation}$ lost some identity information when it was trained to do the orientation task, the later processing stage of the $network_{orientation}$ still retained some of the information that was necessary for object recognition. In addition, though the $network_{identity}$ lost some spatial information when it was trained to do the identity task, it still maintained some information that was necessary for the orientation task. Results suggest that object identity information is retained by a network trained to do a spatial (orientation) task and spatial (orientation) information is retained by a network trained to do object recognition. These results agree with the one object simulation results reported in (Han & Sereno, in press) and suggest that aspects of both object identity and spatial properties might be important for successful object recognition and spatial tasks regardless of the number of objects in the image. However, the information retained was not always sufficient to optimally complete the other goal. Therefore, these results indicate that with two objects in each image, it is still true that multiple pathways are

necessary in order to achieve highest performance on different goals, such as required by the identity task and the orientation task.

METHODS

Object Images

Similar to the methods we used in our previous study (see Han and Sereno, in press), six hundred black and white images were used to train, validate, and test the neural networks. Specifically, 400 images were used for training, 100 were used for validating, and 100 were used for testing. Black and white images of different kinds of tops,pants, and shoes obtained from the tensorflow data set "*Fashion − MNIST*" were used to construct the images of objects (Xiao et al., 2017). Each of these object images consists of three parts: a top (1 of 62 possible), a pant (1 of 66 possible), and a shoe. The shoe could be one of the two following types: sandals (58 possible) and closed shoes (61 possible). In each image, two objects were put in a black background at two locations (bottom left and upper right). The two objects may have different orientations. These two objects images with black background were used as visual inputs. All networks were trained with 200 epochs unless stated otherwise because these networks can reach the highest performance level at the end of training with 200 epochs. Batch size = 256 and the Adam optimization method were used while training.

Object Image Location

Object image locations and object image orientations are shown and explained inFigure 1. In our previous study, the object could be at any of the 9 locations in the background. However, in this two objects simulation study, the two objects can only be at two locations. In each image, the two objects are always positioned at two places (upper left or location 1 and bottom right or location 9) in a 140 × 140 (pixels) black square background (Figure 1A). Note that the two objects never overlap with each other.

Alignment of the Parts Within an Object Image and Orientation of an Object Image

Similar to Han and Sereno (in press), the parts within an object image always had the same alignment (Figure 2). Further, the alignment of the parts within an object and the orientation of the object are always the same (Figure 2). Given an object image, the alignment of the parts within an object image was limited to the two directions along the long axis. For example, if the long axis of

the object image is vertical, then the alignment of the parts could only be up (Figure 2A) or down (Figure 2B). If the long axis of the object image is horizontal, then the alignment of the parts could only be left (Figure 2C) or right (Figure 2D). Hence, the orientations of the object image (as well as the alignment of parts within the object) could have four options: up, down, left, right (Figure 2). There are 4 possible orientations for each object.

Object Image Order: Unscrambled Versus Scrambled

Similar to Han and Sereno (in press), the 6 possible orders for a given object image in the 4 different orientations are illustrated in Figure 2. Despite 6 possible orders, there are only 2 possible classifications by the identity network, unscrambled (US) object or scrambled (S) object. The object image order depends on the orientation of the object. If the orientation of the object is up, then the top part of the object image (order start) is at the top. If the orientation of the object is down, then the top part of the object image (order start) is at the bottom. In 50% of the object images, the top, the pant, and the shoe parts are in the normal order. These images were labeled as unscrambled (images labelled "US" in Figure 2). Just as how people dress themselves and stand up in daily life, the normal order means that the top is at the top, the pant is in the middle and the shoe is at the bottom. If the object image is rotated to another orientation, the normal order stays consistent, just as people sometimes may lie down or do a handstand. In the other 50% of the images, the top, the pant, and the shoe have parts that are in a scrambled order (images labelled "S" in Figure 2). That is, if the order of top, pant, and shoe does not follow the normal order (e.g., shoe, pant, shirt), the object image is labeled as "scrambled" (third image in Figure 2A). In addition, if all the parts are rotated so that the orientation of the object is all upside down and the top is at the top, the pant is in the middle, and the shoe is at the bottom, the object image is also considered "scrambled" (third image in Figure 2B). Thus, with 3 parts in every object image (top, pant, shoe), there are 6 possible spatial orders in total for each orientation and only one of them is in the unscrambled order.

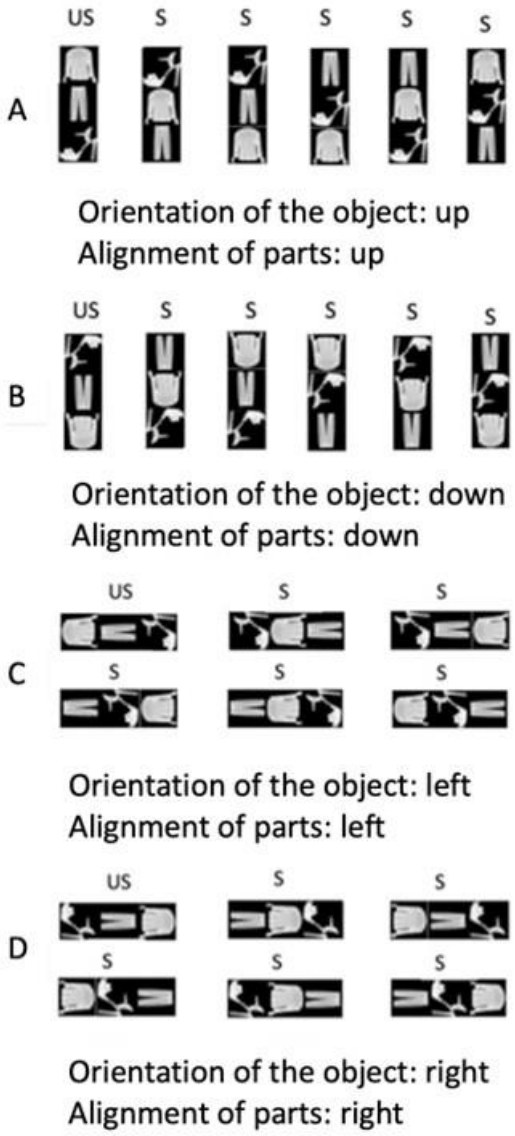

Figure 2. Unscrambled and scrambled object orders. The alignment of the parts within anobject and the orientation of the object are always the same. For each orientation, there are six possible orders of parts. Only the first image for each orientation (first image in each row) is considered as an unscrambled object image (labeled "US"). The other images for a given orientation are scrambled object images (labeled "S"). A. Up orientation. B. Down orientation. C. Left orientation. D. Right orientation. (Adapted from Figure 2 in Han & Sereno (in press)).

Neural Networks

Similar to Han and Sereno (in press), feed-forward multi-layer convolutional artificial neural networks were used to build brain networks to model the visual information processing in the brain. Each brain neural network consists of several hidden layers, including the convolutional layers, the pooling layers, and the fully connected dense layers. ReLu activation function was used at each layer except the final output layer in which a softmax activation function was used. Random dropout was used as a regularization method to improve the performance of the network (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014). These neural networks were implemented using tensorflow and were trained using the supervised learning and the back-propagation method (Rumelhart, Hinton, & Williams, 1986). Simple multi-layer perceptrons were usedto build decoder networks (see additional details below).

Brain Networks: Object Recognition (Identity Task) and Spatial Cognition (Orientation Task) with 2 objects

The structure of the brain networks is shown in Figure 3. The only difference between the different brain networks was in their final output layer. All brain networks take the same set of images as inputs. However, $network_{identity}$ was trained to classify the objects as "scrambled" or "unscrambled" (identity task), whereas $network_{orientation}$ was trained to determine the orientation of the two objects (spatial task). Because there are two objects in each image and each object could have 2 possible identities and 4 possible orientations, there are $2 \times 2 = 4$ possible results for the identities of objects in the image and $4 \times 4 = 16$ possible results for the orientations of objects in the image. Therefore, the chance level testing accuracy for the two tasks are: identity task: 25.0%, orientation task: 6.25%. While training and testing, the activities of the second to last layers of $network_{identity}$ and $network_{orientation}$ were recorded.

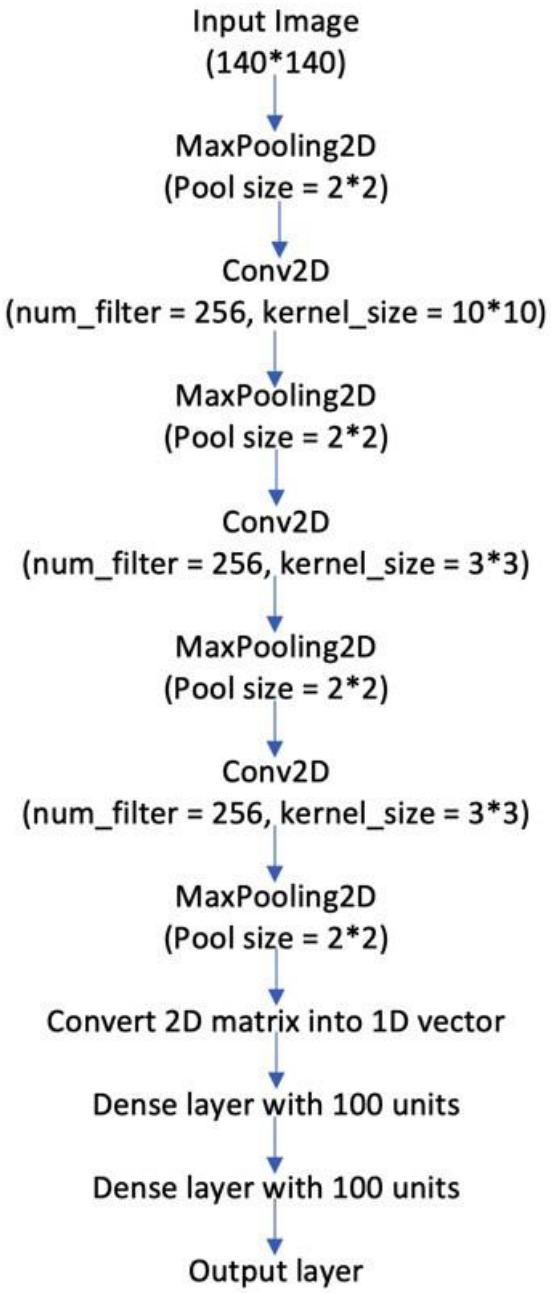

Figure 3. The structure of brain networks. Each neural network consists of several hidden layers, including the convolutional layer, the pooling layer, and the fully connected dense layer. The only difference between different brain networks is the size of their output layer.The size of the output layer depends on the task they were trained to do. (Adapted from Figure 3 in Han & Sereno (in press))

Deoder Networks

In order to analyze the information retained in the later processing stage of the brain networks, simple decoder networks consisting of three fully connected dense layers were used. The structure of the decoder is shown in Figure 4. The decoder network took the artificial neural activities of the second to last layer units of a brain network as inputs and was trained to give different kinds of outputs depending on what kind of information itwas trying to decode. The second to last layer activities of a brain network were different when the input images were different. Therefore, during training and testing of a decoder network, the inputs (second to last layer activities) must be paired with the corresponding true labels of the training and testing images. The reasons for choosing to decode from thesecond to last layer activities of the brain networks are: First, the last layer is the output layer and it only includes information about the final classification decision of the corresponding task, which was different and independent for different networks. Second, the layers before the second to last layer are closer to the input layer and information may not have been fully processed at these layers. As assumed in Han and Sereno (in press), if the decoder is able to use the information from the second to last layer activities to do a task with high accuracy, then that indicates that there is a large amount of task relevant information contained (and/or retained) in the second to last layer activities.

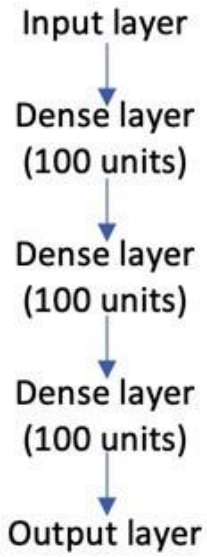

Figure 4. The structure of a decoder network. The input dimension is equal to the number of units in the network layer that it was trained to decode from. The output dimension depends on what kind of information it was trained to decode. (Adapted from Figure 4 in Han & Sereno (in press)).

Comparing Networks

We used a similar method as Han and Sereno (in press) in order to compare networks. Each network (including the decoders) was trained 10 times and testing accuracies were obtained for each of the 10 training sessions. The testing accuracies were obtained by dividing the number of correct classifications by the total number of testing samples (100) during the testing session. The accuracies that are used to compare different networks in this thesis are always referring to the testing accuracies. Unpaired two-samples t-tests were used to compare network accuracies and to determine the significance of the differences.

Baseline Decoder Networks: Getting the Baseline Accuracies

As we noted in Han and Sereno (in press), it is important to first know the decoding accuracy of an untrained network, before trying to decode information from the second to last layer of each brain network. Thus, we used a similar method as Han and Sereno (in press) to determine baseline accuracies from an untrained network. To get the baseline accuracies, an untrained network is used. The untrained network has the same structure as $network_{identity}$ (as the output layer is not important, we could have used the structure of $network_{orientation}$ as well). After all connection weights in the untrained network were randomly initialized, training, validating, and testing images were provided as inputs to the network for 0 epochs and the activities of the second to last layer units were recorded. Because all input data only went through the network once and no training happened during this process (trained for 0 epochs), the connection weights were still random initialized.

Unit activities of the second to last layer of the untrained network served as inputs to the decoder $network_{identity\ two\ objects\ baseline}$. Then the decoder $network_{identity\ two\ objects\ baseline}$ was trained to do the identity task and the accuracy obtained was the baseline accuracy for identity. These unit activities of the second to last layer of an untrained network also served as inputs to the decoder $network_{orientation\ two\ objects\ baseline}$. Then decoder $network_{orientation\ two\ objects\ baseline}$ was trained to determine the orientation information and the accuracy obtained was the baseline accuracy for orientation.

The reason for getting these baseline accuracies is to determine how much information about identity and orientation would still be present in the second to last layer of the network if the network was not trained at all (i.e. all connection weights are random).

Determining the Amount of Information About a Task in the Later Processing Stage of the Brain Network When the Brain Network was Trained to do a Different Task

As we assumed previously (Han & Sereno, in press), it is possible that when the network is trained to do one kind of task, it would extract and keep the task relevant information and ignore any task irrelevant information. Similar to Han and Sereno (in press), we examine here whether the amount of information about a relevant task in the later processing stage of the brain network would still be retained when this network wasfirst trained to do a different irrelevant task.

The inputs and task goals of different decoders are listed in the Table 1. For example, the decoder $network_{(identity,\ orientation)\ two\ objects}$ received intermediate processing information about identity from the brain identity network (i.e., inputs were artificial neural activities from the second to last layer of the brain $network_{identity\ two\ objects}$) but then was trained to decode information about orientations of the two objects from it. Similar explanations can be applied to the other decoder.

Determining Whether Performance on the Identity and Orientation Tasks with 2 objects is Dependent on Whether There is One (Double Sized) Single Pathway or Two Separate Pathways

We tested whether performance was dependent on the number of pathways, similar to what we did in our previous study (Han & Sereno, in press). For $network_{combine\ identity\ and\ orientation\ two\ objects}$, a single pathway takes the images as visual inputs and determines objects' identity and orientation information as 1 of the 64 possible combinations of the two objects' identities (4 possible) and orientations (16 possible). For $network_{separate\ identity\ and\ orientation\ two\ objects}$, two brain network pathways take the images as visual inputs. The brain identity network pathway determines objects' identity and the brain orientation network pathway determines objects' orientation. Later, the results fromthe two network pathways are combined to determine objects' identity and space information as 1 of the 64 possible combinations of the two objects' identities (4 possible) and orientations (16 possible).

Table 1. Inputs and Task Goals of Different Decoders

Decoder Name	Take the Second to Last Layer Activities From	To Do the Task
$network_{(identity,\ orientation)\ two\ objects}$	$network_{identity\ two\ objects}$	orientation
$network_{(orientation,\ identity)\ two\ objects}$	$network_{orientation\ two\ objects}$	identity

The sizes of $network_{combine\ identity\ and\ orientation\ two\ objects}$ and $network_{seperate\ identity\ and\ orientation\ two\ objects}$ are designed to be equal. The only difference is their architectures. In addition, $network_{combine\ identity\ and\ orientation\ two\ objects}$ was trained for 400 epochs and $network_{separate\ identity\ and\ orientation\ two\ objects}$ was trained for 200 epochs because the two brain networks in $network_{separate\ identity\ and\ orientation\ two\ objects}$ had already been trained for 200 epochs in advance. The architectures of these networks are shown Figure 5 ($network_{combine\ identity\ and\ orientation\ two\ objects}$) and Figure 6 ($network_{separate\ identity\ and\ orientation\ two\ objects}$).

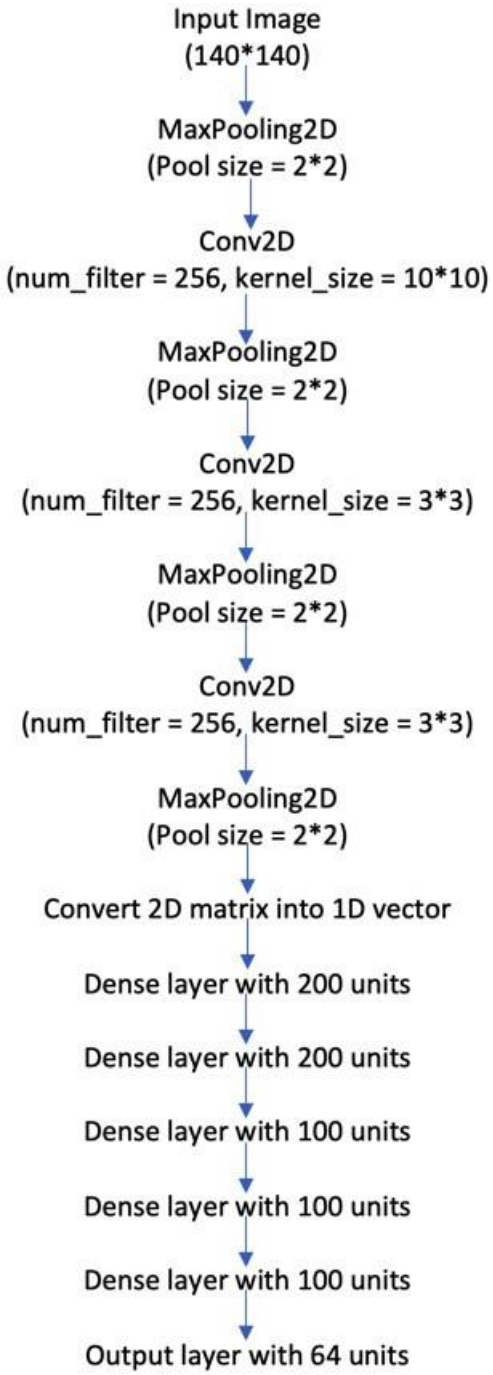

Figure 5. The structure of $network_{combine\ identity\ and\ orientation\ two\ objects}$, the single network that takes the images as visual inputs and determines the two objects' identities and orientations information as 1 of the 64 possible combinations of identities (4 possible) and orientations (16 possible).

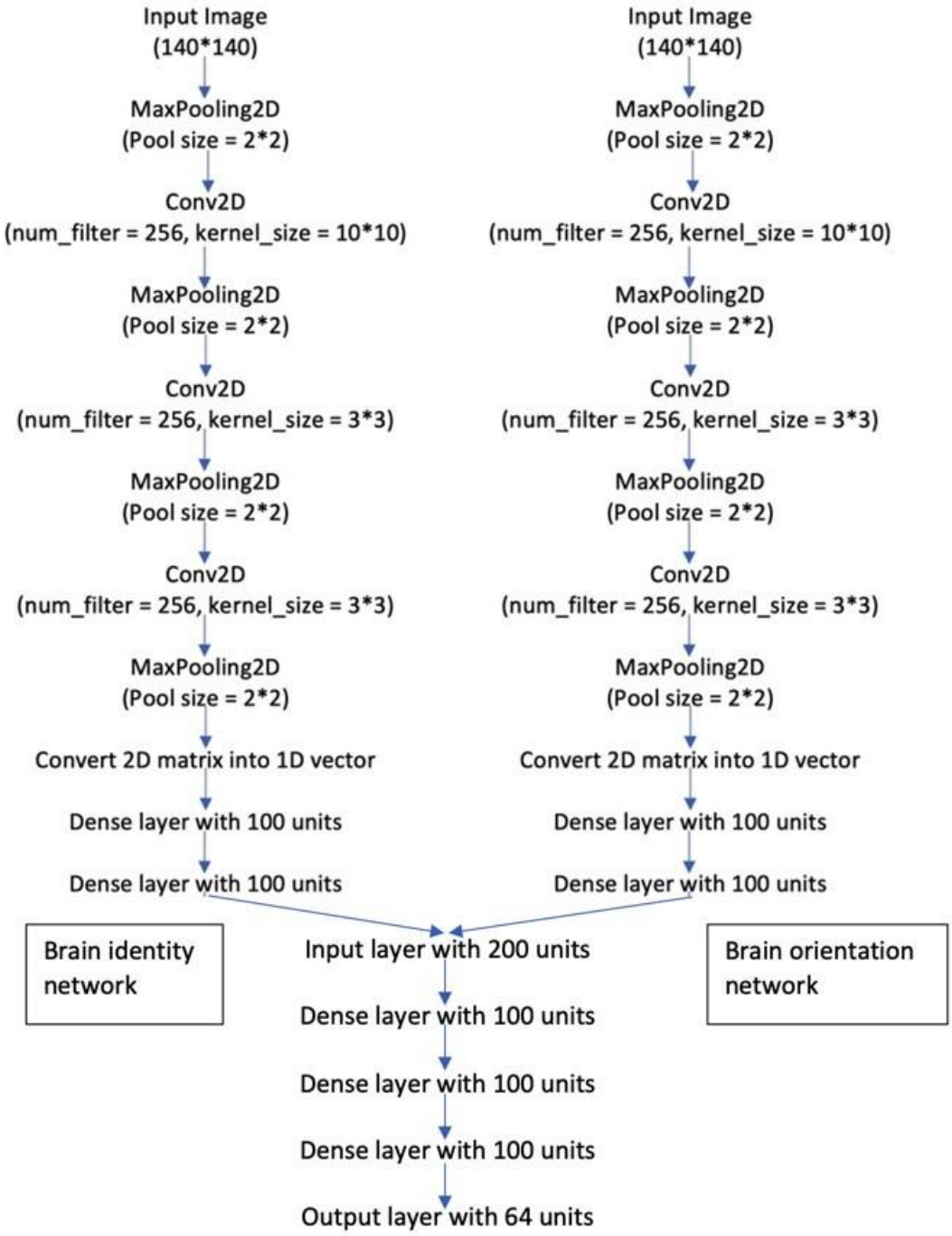

Figure 6. The structure of $network_{separate\ identity\ and\ orientation\ two\ objects}$, the two brain network pathways that take the images as visual inputs. The brain identity network pathway determines objects' identity and the brain space network pathway determines space. Later, the results from the two networks are combined to determine objects' identities and orientation information as 1 of the 64 possible combinations of identities (4possible) and orientations (16 possible).

RESULTS

Similar to methods established in Han and Sereno (in press), it is necessary to perform simulations for multiple times with network weights randomly initialized differently each time to make sure the network does not get stuck at local minimums. The networks were always trained for 10 times and 10 testing accuracies were obtained for each condition after training when obtaining the accuracies in each experimental setting. Unpaired two-samples t-tests were used to compare different testing accuracies and to determine the significance of the differences. The difference is considered to be significant ifthe corresponding p-value < 0.001 (labeled with ***), or p-value < 0.01 (labeled with **), or p-value < 0.05 (labeled with *). The average testing accuracies for different experimental settings are shown in Table 2 and Table 3.

The comparisons of accuracies between different networks are shown in Table 4. Briefly, we found that the second to last layer activities of brain networks that were trained to do a given task had significantly higher decoding accuracies than the baseline when we tried to decode information about a different task. Specifically, we found that a network trained to identify objects' identities actively retained information about objects' orientations, and likewise, a network trained on the orientation task actively retained information about objects' identities.

In addition, simulation results from comparing a single combined pathway versus two segregated pathways in order to accurately identify the identities of the two objects and accurately determine the orientations of the two objects suggest that two separate pathways are advantageous. Separate pathways allow the networks to process the same visual input images in different ways for different tasks or goals. The specific comparisons and findings are discussed in more detail in the Discussion.

These two objects simulation results agree with the one object simulation results previously reported in Han & Sereno (in press). It suggests that these results are independent of the number of objects in each image.

Table 3. Average Testing Accuracies for Various Networks. Definitions of Decoder Networks are Listed in Table 1

Network	Average Accuracy (%)	Chance Level (%)	Standard Deviation (%)
$network_{identity\ two\ objects\ baseline}$	54.1	25.0	1.3
$network_{orientation\ two\ objects\ baseline}$	61.8	12.5	2.8
$network_{identity\ two\ objects}$	93.6	25.0	1.5
$network_{orientation\ two\ objects}$	91.2	12.5	1.4
$network_{(identity,\ orientation)\ two\ objects}$	68.3	25.0	1.8
$network_{(orientation,\ identity)\ two\ objects}$	66.5	12.5	1.6

Table 5. Average Testing Accuracies for $network_{combine\ identity\ and\ orientation\ two\ objects}$ and $network_{separate\ identity\ and\ orientation\ two\ objects}$

Network	Average Accuracy (%)	Chance Level (%)	Standard Deviation (%)
$network_{combine\ identity\ and\ orientation\ two\ objects}$	73.1	1.6	2.0
$network_{separate\ identity\ and\ orientation\ two\ objects}$	87.2	1.6	1.4

Note. The $network_{combine\ identity\ and\ orientation\ two\ objects}$ and the $network_{separate\ identity\ and\ orientation\ two\ objects}$ are used to simulate determining the two objects' identities and orientations with one pathway or two separate pathways.

Table 7. Comparisons of Testing Accuracies Between Different Networks

Network 1	Network 2	Average Difference in Accuracy (%) (*: $p < 0.05$, * *: $p < 0.01$, * * *: $p < 0.001$)	p-value
$network_{(identity,\ orientation)\ two\ objects}$	$network_{orientation\ two\ objects\ baseline}$	6.5 * * *	< 0.001
$network_{(orientation,\ identity)\ two\ objects}$	$network_{identity\ two\ objects\ baseline}$	12.4 * * *	< 0.001
$networks_{separate\ identity\ and\ orientation\ two\ objects}$	$network_{combine\ identity\ and\ orientation\ two\ objects}$	14.1 * * *	< 0.001

Note. The first section examines whether there is information about orientation in the identity network and whether there is information about identity in the orientation network. The final section compares testing accuracies of a network doing the identity and orientation tasks using two separate pathways with a network doing the identity and orientation tasks using a single pathway.

DISCUSSION

In order to better understand whether the presence of identity and spatial properties in cortical areas have a functional role or not and whether the number of objects affect the results, I tried to simulate the ventral identity network and the dorsal orientation network with two objects in each visual input image. Then I compared the two objects simulation results with one object simulation results reported in (Han & Sereno, in press).

I trained networks to do various object and orientation recognition tasks with two objects in each image and the results show that these networks actively retain non-task related information. Both the identity network and the orientation network were independent and trained on a single task and had no cross connections from the other network. Therefore, any non-task related properties that were retained in each of these networks were not coming from the other network.

In sum, these results suggest that (1) the two separate cortical visual pathways for identity and space (orientation) still retain information about both identity and space (orientation) when there are two objects in each image; (2) the retained information about identity and space (orientation) in the two pathways may be necessary to accurately andoptimally recognize objects' identity and orientation. These results agree with our one object simulation results even though the numbers of objects in each image are different.

In our previous study (Han & Sereno, in press), we also had similar findings with one object simulations. In addition, in our previous study, the findings did not depend on the specific parameter settings of the networks. Because very similar methods were used in this thesis, the findings in this thesis should also be independent of the specific parameter settings of the networks. Therefore, these findings should also be valid for the biological brain though their parameter settings may not be the same as our artificial networks.

Is There Information About Orientation in the Brain Identity Network?

The accuracy of the decoder $network_{(identity,orientation)\ two\ objects}$ is significantly higher than the accuracy of the decoder $network_{orientation\ two\ objects\ baseline}$, as shown in Table 4. This result suggests that it is possible to decode information about orientation from the activities of the second to last layer units of $network_{identity\ two\ objects}$. It indicates that even though $network_{identity\ two\ objects}$ was only

trained to identify scrambled/unscrambled objects, its later processing stage still retained information about the orientations of the objects. These findings agree with Han and Sereno (in press)'s one object simulation results.

Is There Information About Object Identity (Scrambled/Unscrambled) in the Brain Orientation Network?

The accuracy of $network_{(orientation,identity)\ two\ objects}$ is significantly higher than the accuracy of $network_{orientation\ two\ objects\ baseline}$. This indicates that even though $network_{orientation\ two\ objects}$ was trained to determine the orientation of the objects, its later processing stage still retained some information that was necessary to do the identity task.

Comparison Between the Performance of a Single Pathway Network and the Performance of a Two Pathways Network

According to Han and Sereno (in press), two separate pathways are advantageous inorder to process the same visual inputs in different ways for different tasks or goals so thatthe network can accurately identify objects and accurately determine the location and orientation of objects. In this thesis, I tried to investigate whether two separate pathways are still advantageous when there are multiple objects in each image.

To address this question, the accuracies of $network_{combine\ identity\ and\ orientation\ two\ objects}$ and $network_{separate\ identity\ and\ orientation\ two\ objects}$ were compared. $network_{combine\ identity\ and\ orientation\ two\ objects}$ was used to simulate the process of doing the identity and orientation tasks using a single pathway and $network_{separate\ identity\ and\ orientation\ two\ objects}$ was used to simulate the process of doing the identity and spatial tasks using two separate pathways. As shown in Table 4, the testing accuracy of $network_{separate\ identity\ and\ orientation\ two\ objects}$ is significantly higher than the accuracy of $network_{combine\ identity\ and\ orientation\ two\ objects}$. It implies that when two pathways are used to determine the two objects' identity and orientation information separately, the neural network has better performance. Our findings suggest there are advantages for the brain touse two separate pathways to determine identity and spatial (orientation) information.

In summary, in order to accurately identify objects and accurately determine the spatial information (orientations) of objects, these findings suggest that two separate pathways are advantageous in order to process the same visual inputs in different ways for different tasks or goals.

Processing information differently using multiple separate pathways may cause a binding problem (Treisman, 2002). In this thesis, the two objects have always been fixed at the same two locations (upper left and bottom right). Therefore, the identities of the two objects can be combined with the orientations of the two objects easily according to their fixed locations. However, the binding problem is still an important problem because the objects may not always be at the same locations. One possible way to solve the binding problem is that the binding problem may be lessened by using the spatial information contained in the identity network and object identity information in the spatial network.

Limitations and Future Directions

According to our one object (Han & Sereno, in press) and two objects simulations,we have suggested that object identity and spatial information is processed independently in both ventral and dorsal pathways. Further, we suggest it is advantageous to have two separate pathways for object recognition and spatial cognition. However, the binding problem caused by processing information in multiple pathways independently remains a problem to be resolved. In our previous study, the binding problem was not a problem with only one object ineach image (Han & Sereno, in press). In this thesis, the binding problem is solved by always fixing the two objects at the same locations. It is obviously not a valid method to solve the binding problem in the general case. In the future, it is important to find a betterway to combine the independently processed information from different pathways when there are multiple objects in each image.

Conclusion

In summary, these simulations imply that with two objects in each image, it is still true that both ventral and dorsal cortical visual pathways contain information about both identity (of shape) and space (orientation), even when trained with a single identity or orientation task. The modeling also suggests that the identity and spatial (orientation) information retained in the two pathways is important for accurately accomplishing the identity and spatial (orientation) tasks. These results agree with our one object simulation results (Han & Sereno, in press). Therefore, it suggests that these results are robust and are not dependent on the number of objects in the visual tasks.

REFERENCES

Colby, C. L., & Goldberg, M. E. (1999). space and attention in parietal cortex. *Annual Review of Neuroscience, 22* , 319-349. https://doi.org/10.1146/annurev.neuro.22.1.319

Felleman, D., & Essen, D. V. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex, 1*(1), 1-47. https://doi.org/10.1093/cercor/1.1.1.

Freud, E., Plaut, D. C., & Behrmann, M. (2016). 'what' is happening in the dorsal visual pathway. *Trends in Cognitive Sciences, 20*(10), 773-784. https://doi.org/10.1016/j.tics.2016.08.003

Freud, E., Rosenthal, G., Ganel, T., & Avidan, G. (2015). Sensitivity to object impossibility in the human visual cortex: Evidence from functional connectivity. *Journal of Cognitive Neuroscience, 27*(5), 1029-1043. https://doi.org/10.1162/jocn_a_00753

Han, Z., & Sereno, A. (in press). Modeling the ventral and dorsal cortical visual pathways using artificial neural networks. *Neural Computation*.

Hong, H., Yamins, D. L. K., Majaj, N. J., & DiCarlo, J. J. (2016). Explicit information for category-orthogonal object properties increases along the ventral stream. *Nature Neuroscience, 19*(4), 613-630. https://doi.org/10.1038/nn.4247

Konen, C. S., & Kastner, S. (2008). Two hierarchically organized neural systems for object information in human visual cortex. *Nature Neuroscience, 11*(2), 224-231. https://doi.org/10.1038/nn2036

Lehky, S. R., Peng, X., McAdams, C. J., & Sereno, A. B. (2008). Spatial modulation of primate inferotemporal responses by eye position. *PLoS ONE, 3*(10), e3492. https://doi.org/10.1371/journal.pone.0003492

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience, 19,* 577-621. https://doi.org/10.1146/annurev.ne.19.030196.003045

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences, 6,* 414-417.https://doi.org/10.1016/0166-2236(83)90190-X

Nowicka, A., & Ringo, J. L. (2000). Eye position-sensitive units in hippocampal formation and in inferotemporal cortex of the macaque monkey. *European Journal of Neuroscience, 12,* 751-759. https://doi.org/10.1046/j.1460-9568.2000.00943.x

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations byback-propagating errors. *Nature, 323,* 533-536. https://doi.org/10.1038/323533a0

Sereno, A. B., & Lehky, S. R. (2011). Population coding of visual space: comparison of spatial representations in dorsal and ventral pathways. *Frontiers in Computational Neuroscience*, *4*(159), 1-16. https://doi.org/10.3389/fncom.2010.00159

Sereno, A. B., Lehky, S. R., & Sereno, M. E. (2020). Representation of shape, space, and attention in monkey cortex. *Cortex*, *122*, 40-60. https://doi.org/10.1016/j.cortex.2019.06.005

Sereno, A. B., Sereno, M. E., & Lehky, S. R. (2014). Recovering stimulus locations using populations of eye-position modulated neurons in dorsal and ventral visual streams of non-human primates. *Frontiers in Integrative Neuroscience*, *8*(28), 1-20. https://doi.org/10.3389/fnint.2014.00028

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014).Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, *15*(56), 19291958. Retrieved from http://jmlr.org/papers/v15/srivastava14a.html

Treisman, A. (2002). The binding problem. *Current Opinion in Neurobiology*, *6*(2), 171-178. https://doi.org/10.1016/S0959-4388(96)80070-5

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In M. Goodale, D. J. Ingle, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549-586).Cambridge, MA: MIT Press.

Xiao, H., Rasul, K., & Vollgraf, R. (2017). Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. https://doi.org/arXiv:1708.07747

Zachariou, V., Klatzky, R., & Behrmann, M. (2014). Ventral and dorsal visual stream contributions to the perception of object shape and object location. *Journal of Cognitive Neuroscience*, *26*(1), 189-209. https://doi.org/10.1162/jocn_a_00475

APPENDIX. MANUSCRIPT

The following manuscript is the authors' final unpublished version of the paper, which has been accepted for publication in *Neural Computation*.

Neural Computation homepage: https://direct.mit.edu/neco

MODELING THE VENTRAL AND DORSAL CORTICAL VISUAL PATHWAYS USING ARTIFICIAL NEURAL NETWORKS

Zhixian Han[1] and Anne Sereno[1,2]

[1]Department of Psychological Sciences, Purdue University, West Lafayette, IN, USA 47907

[2]Weldon School of Biomedical Engineering, Purdue University, West Lafayette, IN, USA 47907

ABSTRACT

Although in conventional models of cortical processing, object recognition and spatial properties are processed separately in ventral and dorsal cortical visual pathways respectively, some recent studies have shown that representations associated with both object's identity (of shape) and space are present in both visual pathways. However, it is still unclear whether the presence of identity and spatial properties in both pathways have functional roles or not. In our study, we have tried to answer this question through computational modeling. Our simulation results show that both a model ventral and dorsalpathway, separately trained to do object and spatial recognition, respectively, each actively retained information about both identity and space. In addition, we also show that these networks retained different amounts and kinds of identity and spatial information. As a result, our modeling suggests that two separate cortical visual pathways for identity and space (1) actively retain information about both identity and space; (2) retain information about identity and space differently; and (3) that this differently retained information about identity and space in the two pathways may be necessary to accurately and optimally recognize and localize objects. Further, modeling results suggests these findings are robust and do not strongly depend on the specific structures of the neural networks.

MODELING THE VENTRAL AND DORSAL CORTICAL VISUAL PATHWAYS USING ARTIFICIAL NEURAL NETWORKS

Introduction

It is widely documented in neuropsychological, lesion, and anatomical studies that the human visual system has two distinct cortical pathways (Ungerleider & Mishkin, 1982; Mishkin, Ungerleider, & Macko, 1983; Felleman & Essen, 1991). Further, the ventral pathway primarily processes information important for object recognition (Logothetis & Sheinberg, 1996) while the dorsal pathway primarily processes information related to spatial cognition (Colby & Goldberg, 1999). However, some recent studies have challenged this idea (Konen & Kastner, 2008; Freud, Rosenthal, Ganel, & Avidan, 2015; Freud, Plaut,& Behrmann, 2016; Hong, Yamins, Majaj, & DiCarlo, 2016). Some studies have found thatrepresentations associated with shape and location processing are present in both visual streams (Konen & Kastner, 2008; A. B. Sereno, Lehky, & Sereno, 2020; Hong et al., 2016).However, it remains unclear whether the representations of shape in dorsal stream and the representations of location in ventral stream are non-task-related or whether they might play a functional rule in spatial cognition and object recognition, respectively. Some findings from fMRI and behavioral studies have suggested that spatial processing that operates at the level of the scene, presumably within the dorsal visual pathway, can contribute to shape processing (Zachariou, Klatzky, & Behrmann, 2014). Another study found that correlated activity between ventral and dorsal visual pathways were higher when people were looking at objects with impossible spatial structures compared with when they were looking at objects with possible structures (Freud et al., 2015), which indicated that dorsal pathway processing might help the brain to recognize objects with impossible structures. Furthermore, Hong et al. (2016) found in neural recordings that spatial information increases along the ventral stream, consistent with prior studies demonstrating spatial properties in later stages of the ventral stream (Nowicka & Ringo, 2000; Lehky, Peng, McAdams, & Sereno, 2008). In addition, Hong et al. (2016) suggest that it is likely that the spatial information in the ventral stream does not come from the dorsal stream, in agreement with previous studies arguing that ventral stream spatialrepresentations are distinct and independent from dorsal stream spatial encodings (A. B. Sereno & Lehky, 2011; A. B. Sereno, Sereno, & Lehky, 2014).

The experimental evidence mentioned above indicates that representations of shape and space exist in both visual pathways and might have functional roles. Therefore, we attempt here to tackle

these questions through explicit hypothesis testing using computational models. In our study, we examine whether identity (of shape) and space processing were found to be present in both simulated ventral and dorsal streams trained to do straightforward object recognition and localization tasks, respectively; we explore possible reasons for why information associated with identity (shape) and space processing were found to be present in both simulated ventral and dorsal streams; and finally, discuss how this information could elucidate our understanding of the computational properties and needs of the two visual streams. Hong et al. (2016) showed with modeling that explicitspatial information is present in the ventral pathway. They did not show whether shape information is present or retained in the dorsal pathway. They also did not show whether different kinds of shape and spatial information are maintained differently in simulations ofthe ventral and dorsal pathways. Their results are not sufficient to suggest why seemingly task-irrelevant information is maintained in a neural network. These are computationally tractable questions that are important and timely.

In order to model the two cortical visual pathways and study their computational properties, feed-forward multi-layer convolutional neural networks were used to simulate the functions of the two visual pathways in the brain and multi-layer perceptrons were used to simulate the process of decoding information from recorded neural activities in the brain. All networks were trained using supervised learning. When modeling the two cortical visual pathways, for simplicity and control, it is assumed that the two pathways use the same computational structure (the numbers of neurons are the same and the structures of the initial connections between neurons are the same) and receive the same visual input images. However, we will allow the connection weights between the neurons inthe two pathways to be modified with training. It is almost certain that the connection weights between the two pathways will be different after training because the training networks have to meet different goals. Specifically, the primary goal of the ventral pathway is to distinguish different kinds of objects by distinguishing different features or different combinations of features, whereas the primary goal of the dorsal pathway is to determine the spatial information (e.g. locations and/or orientations) of objects necessary for interaction (e.g. reach, grasp, and/or avoidance/navigation). We used the back-propagation training method as a tool to capture the computational properties that result from these differing goals of the two visual pathways. Back-propagation is currently the best method for updating connection weights between neurons in artificial neural networks. In general, artificial neural networks trained using the back-propagation methodtend to perform better than models trained using any other weight-updating methods (Lillicrap, Santoro, Marris, Akerman, & Hinton, 2020).

Though biological neural networks are unlikely to be able to perform back-propagation weight updates in the same way as artificial neural networks, some researchers have argued that biological neural networks could compute back-propagation-like effective synaptic updates by using the differences of neural activities induced by feedback connections (Lillicrap et al., 2020; Whittington & Bogacz, 2019). Therefore, the back-propagation training method was used to obtain the results shown in this paper.

Black and white images consisting of different kinds of tops, pants, and shoes (Xiao,Rasul, & Vollgraf, 2017) were used to construct the images of objects (see Figure 1). Each image of an object consisted of one top, one pant, and one shoe. Each object image was put in a black background at one of several possible locations and orientations. These object images with black background were used as visual inputs. In half of the images, the top, the pant, and the shoe were in the "unscrambled" order. In the other half of the images, the top, the pant, and the shoe were in the "scrambled" order (see Methods for additional details).

Five artificial neural networks $network_{identity}$, $network_{shoes}$, $network_{space}$, $network_{location}$, and $network_{orientation}$ were trained to do an identity task (whether the image is scrambled or unscrambled), a shoe task (determine whether the shoe in the imageis a sandal or closed shoe), and three spatial tasks (determine the location and orientation, location alone, or the orientation alone of the image), respectively. Both $network_{identity}$ and $network_{shoes}$ were used to model the ventral pathway, whereas $network_{space}$, $network_{location}$, and $network_{orientation}$ were used to model the dorsal pathway. These five networks are considered the brain networks. These brain networks were used to simulate the functions of ventral and dorsal cortical visual pathways in the brain. Various additional nonlinear and linear decoders were then trained to decode different kinds of information from the later processing stages of these brain networks. These decoders were used to simulate the process of decoding information embedded in the recorded neural activity signals in the brain. It is assumed that if the testing accuracy of the decoder was higher, then the later processing stages of the brain networks retained more information needed bythe decoder's decoding goal.

According to the simulation results, though the $network_{space}$ lost some identity information when it was trained to do the space task, the later processing stage of the $network_{space}$ still retained some of the information that was necessary for distinguishing different kinds of objects (combinations of features). In addition, though the $network_{identity}$lost some spatial information when it was trained to do the identity task, it still maintained some information that was necessary for the

spatial task. Specifically, although $network_{identity}$ maintained both location and orientation information, it maintained more information about orientations of the object images. Results suggest that object information is retained by a network trained to do a spatial task and spatial information isretained by a network trained to do object recognition, suggesting that aspects of both object and spatial properties might be important for successful object recognition and spatial tasks. However, the information retained was not always sufficient to optimally complete the other goal. Therefore, the results indicate that a reason for why there are twovisual pathways in the brain might be that multiple pathways are necessary in order to achieve highest performance on different goals, such as required by the identity, the spatial,and the shoe tasks. More importantly, it also suggests that these multiple pathways retain different aspects and amounts of both object and spatial information to achieve highest performance on spatial and object tasks respectively.

Our main modeling goal is to gain a better understanding of computational issues rather than identifying the specific response features that are similar to the real neural responses of ventral and dorsal cortical areas. That is, a proof of computational concept more than an accurate model of the real human brain. Indeed, known differences in the structure of the two pathways (e.g., different number of areas within each stream, already evident in (Felleman & Essen, 1991)) would complicate direct and controlled comparisonsof such biologically accurate models.

Given that our goal is proof of computational concept, we repeated some of the simulations with slightly different brain network structures (different number of filters, different kernel sizes) to test if our findings are dependent on the specific conditions or structures of the artificial neural networks. Because our findings do not depend on the specific structures we have used or particular parameters chosen, the findings suggest they may reflect more general computational processes. Specifically, our findings may also be valid for the biological brain even though the structures of our artificial neural networks and the structures of the biological brain networks are not the same.

Methods

Object Images

Black and white images consisting of different kinds of tops, pants, and shoes (Xiao et al., 2017) were used to construct the images of objects (see Figure 1). Images of different kinds of tops, pants, and shoes obtained from the tensorflow data set "Fashion-MNIST" were used to construct

the images of objects (Xiao et al., 2017). Each of these object images consists of three parts: a top (1 of 62 possible), a pant (1 of 66 possible), and a shoe. The shoe could be one of the two following types: sandals (58 possible) and closed shoes (61 possible). Each object image was embedded in a black background and presented at different locations and orientations (all parts - top, pant, shoe - of the object were presented with the same orientation, and altogether centered at the selected location).

These object images with black background were used as visual inputs. In half of the images, the top, the pant, and the shoe were in the "unscrambled" order. That is, the unscrambled order is the normal order of how people are dressed with the pant, but not shoe or top, in the middle. In the other half of the images, the top, the pant, and the shoewere in the "scrambled" order, where the order of top, pant, and shoe does not follow the normal order.

Six hundred black and white images were used to train, validate, and test the neuralnetworks. Specifically, 400 images were used for training, 100 were used for validating, and 100 were used for testing. We have used a small dataset in our simulations and did not useimage augmentation because our goal is not to maximize the performance of the artificial neural networks but rather to compare the performance of different neural networks in order to clarify differences in the kinds of information that are retained as well as how much information is retained. Using a very large dataset (60,000 images) caused the testing accuracy of $network_{identity}$ to be above 98% and the accuracy of $network_{space}$ to be 100%. The reason they could reach such high accuracies may be because training images and tasks were simple and the number of possible variations was limited. It is difficult to examine and identify performance differences between different networks if most of the networks have almost 100% accuracies. Nevertheless, to test if dataset size would alter any findings, we repeated some simulations with a larger dataset (1200 images) and found that the size of the dataset did not affect our major findings. All networks were trained with 200 epochs and all of these networks had reached the highest performance level at the end of training with 200 epochs. For some conditions where testing accuracies approached 100%, we added Gaussian noise to the images (including both object and background) to increase task difficulty so that we could better compare performance differences of the different networks. An example of the noisy image is shown in Figure 1D. Batch size = 256and the Adam optimization method were used while training. The initial learning rate of Adam optimization was 0.001.

Object Image Location

Object image locations and object image orientations are shown and explained inFigure 1. The object images were put at different places in a 140 × 140 (pixels) black square background. Specifically, the centers of each object image could have 9 possible locations (Figure 1A).

Alignment of the Parts within an Object Image and Orientation of an Object Image

The parts within an object image always had the same alignment (Figure 2). Further, the alignment of the parts within an object and the orientation of the object are always the same (Figure 2). Given an object image, the alignment of the parts within an object image was limited to the two directions along the long axis. For example, if the longaxis of the object image is vertical, then the alignment of the parts could only be up (Figure 2A) or down (Figure 2B). If the long axis of the object image is horizontal, then the alignment of the parts could only be left (Figure 2C) or right (Figure 2D). Hence, the orientations of the object image (as well as the alignment of parts within the object) could have four options: up, down, left, right (Figure 2). With 9 possible center locations and 4possible orientations, there were 36 spatial combinations of locations and orientations in total.

Object Image Order: Unscrambled versus Scrambled

The six possible orders for a given object image in the four different orientations areillustrated in Figure 2. Despite 6 possible orders, there are only 2 possible classifications bythe identity network, unscrambled (US) object or scrambled (S) object. The object image order is determined by the orientation of the object. If the orientation of the object is up, then the top part of the object image (order start) is at the top. If the orientation of the object is down, then the top part of the object image (order start) is at the bottom. In halfof the object images (300 out of 600), the top, the pant, and the shoe parts are in the normal order. These images were labeled as unscrambled (images labelled "US" in Figure 2). Just as how people dress themselves and stand up in daily life, the normal ordermeans that the top is at the top, the pant is in the middle and the shoe is at the bottom. If the object image is rotated to another orientation, the normal order stays consistent, just as people sometimes may lie down or do a handstand. In the other half of the images (300 out of 600), the top, the pant, and the shoe have parts that are in a scrambled order (images labelled "S" in Figure 2). That is, if the order of top, pant, and shoe does not follow the normal order (e.g., shoe, shirt, pant),

the object image is labeled as "scrambled" (second image in Figure 2A). In addition, if all the parts are rotated so that the orientationof the object is all upside down and the top is at the top, the pant is in the middle, and theshoe is at the bottom, the object image is also considered “scrambled” (third image in Figure 2B). Thus, with 3 parts in every object image (top, pant, shoe), there were 6 possible spatial orders in total for each orientation and only one of them is the unscrambled order.

We chose to use the scrambled-unscrambled or identity task because it is a common task used to identify ventral regions in human fMRI studies (e.g., (Kourtzi & Kanwisher,2000; Grill-Spector, Kourtzi, & Kanwisher, 2001)). It is an object recognition task that includes information about the relations between parts. The shoe identity task is a task that is sensitive to the ability to discriminate shape information of parts. Much work in animals (e.g., with respect to faces, (Perrett, Hietanen, Oram, & Benson, 1992)) as well ashumans (e.g., (Hoffman & Haxby, 2000)) have demonstrated the importance of ventral regions in discriminating lower level visual features (see also, (Bracci, Ritchie, & de Beeck,2017)).

Neural Networks

Feed-forward multi-layer convolutional artificial neural networks were used to buildbrain networks to model the visual information processing in the brain. Each neural network consists of several hidden layers, including the convolutional layers, the pooling layers, and the fully connected dense layers. ReLu activation function was used at each layer except the final output layer in which a softmax activation function was used. Random dropout was used as a regularization method to improve the performance of the network. Random dropout regularization method is a neuroscience-inspired regularizationmethod that is commonly used in the deep learning community (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014). These neural networks were implemented using tensorflow and were trained using the supervised learning and the back-propagation method (Rumelhart, Hinton, & Williams, 1986). Simple multi-layer perceptrons were usedto build decoder networks (see additional details below).

Brain Networks: Global Recognition (Identity Task), Spatial Cognition (Spatial Task, Location Task, Orientation Task) and Feature Recognition Networks (Shoe Task)

The structure of the brain networks is shown in Figure 3. All brain networks share the same structure. The only difference between the different brain networks was in their final output layer. To test whether our findings were dependent on the specific brain network structure (number of filters, kernel sizes), we also repeated some simulations with different brain network structures (different number of filters, different kernel sizes), and show in the results section that our main conclusions did not change when the brain network structures were modified.

All brain networks take the same set of images as inputs. However, $network_{identity}$ was trained to classify the input images as "scrambled" or "unscrambled" (identity task), whereas $network_{space}$ was trained to determine the location and the orientation of the images (spatial task). A third network, $network_{shoes}$ was a variant of $network_{identity}$. It wasidentical to $network_{identity}$ but trained instead to classify the type of shoes in both scrambled and unscrambled images as either a "closed shoe" or "sandal". Two additional networks were variants of $network_{space}$: $network_{location}$ was a variant of $network_{space}$ and trained instead to only determine the locations of the images and $network_{orientation}$ was a variant of $network_{space}$ and trained instead to only determine the orientations of the images(both networks only differing from $network_{space}$ in their final output layer). The chance level testing accuracy for the various tasks are: identity task: 50.0%, spatial task: 2.8%, shoes task: 50.0%, location task: 11.1%, orientation task: 25.0%. While training and testing, the activities of the second to last layers of $network_{identity}$, $network_{space}$, $network_{shoes}$, $network_{location}$, and $network_{orientation}$ were recorded.

Decoders

In order to analyze the information contained in the later processing stage of the convolutional networks, two kinds of decoders were used: nonlinear decoder networks and linear decoders.

A nonlinear decoder network consisting of three fully connected dense layers was used. The structure of the decoder is shown in Figure 4. We used nonlinear decoders because we were trying to simulate the process of decoding information from the brain, where nonlinear decoders have been used (Xu et al., 2019). The decoder network took the artificial neural activities of the second to last

layer units of a brain network as inputs and was trained to give different kinds of outputs depending on what kind of information it was trying to decode. While training the decoder networks, only the parameters in the decodernetwork were adjusted.

The linear decoders we used were linear support vector machines (linear SVMs). The parameters were set as follows. Loss function: squared hinge. Regularization: L2 regularization with regularization strength = 1. The linear decoders also took the artificialneural activities of the second to last layer units of a brain network as inputs and was trained to give different kinds of outputs depending on what kind of information it was trying to decode.

The second to last layer activities of a brain network are different when the input images are different. Therefore, during training and testing of a decoder network, the inputs (second to last layer activities) must be paired with the corresponding true labels ofthe training and testing images. The reasons for choosing to decode from the second to lastlayer activities are: First, the last layer is the output layer and it only includes information about the final classification decision of the corresponding task, which was different for different networks. Second, the layers before the second to last layer are closer to the inputlayer and information may not have been fully processed at these layers. The assumption isthat if the decoder is able to use the second to last layer activities to do a task with high accuracy, then that indicates that there is a large amount of task relevant information contained (and/or retained) in the second to last layer activities.

Comparing Networks

In order to compare networks, each network (including the decoders) was trained 10times and testing accuracies were obtained for each of the 10 training sessions. The testing accuracies were obtained by dividing the number of correct classifications by the total number of testing samples (100) during the testing session. The accuracies that are used tocompare different networks in this paper are always referring to the testing accuracies. Unpaired two-samples t-tests were used to compare network accuracies and to determine the significance of the differences.

Baseline Decoder Networks: Getting the baseline accuracies

Before trying to decode information from the second to last layer of each brain network, it is important to know the accuracy of decoding from an untrained network. To get the baseline

accuracies, an untrained network is used. The untrained network has the same structure as $network_{identity}$ (as the output layer is not important, we could have used the structure of $network_{space}$ as well). After all connection weights in the untrained network were randomly initialized, training, validating, and testing images were provided as inputs to the network for 0 epochs and the activities of the second to last layer units were recorded. Because all input data only went through the network once and no traininghappened during this process (trained for 0 epochs), the connection weights were still random.

Unit activities of the second to last layer of the untrained network served as inputs to the decoder $network_{identity\ baseline}$. Then the decoder $network_{identity\ baseline}$ was trained to do the identity task and the accuracy obtained was the baseline accuracy for identity. These unit activities of the second to last layer of an untrained network also served as inputs to the decoder $network_{space\ baseline}$. Then decoder $network_{space\ baseline}$ was trained to determine the spatial information and the accuracy obtained was the baseline accuracy forspace. When these activities of an untrained network served as inputs to the decoder $network_{shoes\ baseline}$, the decoder $network_{shoes\ baseline}$ was trained to determine the type of shoes and the accuracy obtained was the baseline accuracy for the classification of shoes.

The reason for getting these baseline accuracies is to determine how much information about identity, space and shoes would still be present in the second to last layer of the network if the network was not trained at all (i.e. all connection weights arerandom).

Determining the Amount of Information About a Task in the Later Processing Stage of the Brain Network When the Brain Network was Trained to do a Different Task

It is possible that when the network is trained to do one kind of task, it would extract the task relevant information and throw away task irrelevant information. We examine here whether the amount of information about a relevant task in the later processing stage of the brain network would increase or decrease when this network was first trained to do a different irrelevant task.

The inputs and task goals of different decoders are listed in Table 1. For example, the decoder $network_{(space,identity)}$ received intermediate processing information about spacefrom the brain space network (i.e., inputs were artificial neural activities from the second to last layer of the brain $network_{space}$) but then was trained to decode information about identity from it. Similar arguments can be applied to other decoders.

Determining the Amount of Information About a Task in the Later Processing Stage of the Brain Network When the Brain Network was Trained to do the Same Task

Using decoder networks to decode information from the brain networks is similar to adding more layers to the brain network and then training it for more epochs. The network's testing accuracy could increase or decrease simply because it was trained for more epochs or it has more layers. Training for more epochs or having more layers may increase testing accuracy by extracting more statistical information from the training samples, whereas it could also decrease testing accuracy by over-fitting. Therefore, if we want to determine whether training with a task helps or hurts the network's ability to do another (different) task, we need to determine the accuracy of the decoder network when the brain network was trained again to do the same task.

The inputs and task goals of the relevant decoders are also listed in Table 1. For example, decoder $network_{(identity,identity)}$ received the intermediate processing information of the brain $network_{identity}$ as inputs (i.e., inputs were artificial neural activities from the second to last layer of the brain $network_{identity}$) and then was trained to decode informationabout identity from it. Decoder $network_{(space,space)}$ received the intermediate processing information of $network_{space}$ and then was trained to decode spatial information from it.

Determining Whether Performance on the Identity and Spatial Tasks is Dependent on Whether There is One (Double Sized) Single Network or Two Separate Networks

For $network_{combine\ identity\ and\ space}$, a single network takes the images as visual inputs and determines objects' identity and space information as 1 of the 72 possible combinations of identity (2 possible) and space (36 possible). For $network_{separate\ identity\ and\ space}$, two brainnetworks take the images as visual inputs. The brain identity network determines objects' identity and the brain space network determines space. Later, the results from the two networks are combined to determine objects' identity and space information as 1 of the 72 possible combinations of identity (2 possible) and space (36 possible).

The sizes of $network_{combine\ identity\ and\ space}$ and $network_{separate\ identity\ and\ space}$ are designed to be equal. The only difference is their architectures. In addition, $network_{combine\ identity\ and\ space}$ was trained for 400 epochs and $network_{separate\ identity\ and\ space}$ was trained for 200 epochs because the two brain networks in $network_{separate\ identity\ and\ space}$ had already been trained for 200 epochs in advance.

The architectures of these networks are shown in Figure 5 ($network_{combine\ identity\ and\ space}$) and Figure 6 ($network_{separate\ identity\ and\ space}$).

Results

It is necessary to perform training, validation, and testing for multiple times with network weights randomly initialized differently each time to make sure the network did not get stuck at local minimums. When obtaining the accuracies in each experimental setting, the networks were always trained for 10 times and 10 testing accuracies were obtained for each condition after training. Unpaired two-samples t-tests were used to compare different accuracies and to determine the significance of the differences. The difference is consideredto be significant if the corresponding p-value < 0.05. The level of significance is marked with * (p-value < 0.05), ** (p-value < 0.01), and *** (p-value < 0.001). The average testing accuracies for different experimental settings are shown in Table 2 and Table 3. One possible reason for the baseline accuracies to be higher than the corresponding chancelevels is that although the connection weights were initialized randomly, some information contained within the input images themselves can still be passed on to the second to lastlayer units and this sensory-driven information was decoded by the decoder networks. The comparisons of accuracies between different networks are shown in Table 4.

Briefly, we found that the second to last layer activities of brain networks that were trainedto do a given task had higher decoding accuracies than the baseline when we tried to decode information about a different task. That is, we found that a network trained to identify images actively retained information about space, and likewise, a network trained on a spatial task actively retained information about identity. In addition, the decoding accuracies were lower from the brain networks that were trained to do a different task than from brain networks that were trained to do the same task. Additional modeling to better understand why networks retained seemingly task irrelevant information suggest that this information is retained and preserved uniquely in service of improving the accuracy of the "irrelevant" task. For example, the identity network actively maintained more information about orientation than location because in order to determine whether the object is in the unscrambled or scrambled order, the network needs to determine the object orientation.

Finally, simulation results from comparing a single combined pathway versus two segregated pathways in order to accurately identify objects and accurately determine the location and orientation of objects suggest that two separate pathways are advantageous in order to process the

same input (visual information) in different ways for different tasks or goals. The specific comparisons and findings are discussed in more detail in the Discussion.

One additional comparison, not illustrated in Table 4, was made to compare the difference in the amount of accuracy decrease in percentage from $network_{(location,location)}$ to $network_{(identity,location)}$ versus from $network_{(orientation,orientation)}$ to $network_{(identity,orientation)}$. We found that the accuracy of brain $network_{(location,location)}$ stayed around 100% when this network was trained with more epochs (e.g., 200 epochs) and suggest that the accuracy of this brain network saturates when trained for 200 epochs. Since we are using the accuraciesof the decoders to assess the amount of different kinds of information contained in the second to last layer brain network activities, when the accuracy saturates, it is possible that with additional training the amount of information retained has changed but the accuracy stays the same (around 100%), making it difficult to evaluate whether the amountof information retained has changed or not and difficult to compare these networks' performance with other brain and decoder networks. Therefore, we increased the difficulty for the location and orientation tasks by adding Gaussian white noise to the input images.

With noisy input images, the accuracy of $network_{(location,location)}$ was still very high (97.0%) but did not reach 100%. The range of the accuracy of the location task is from 11.1% (chance level) to 100.0%, or 88.9%. The range of the accuracy of the orientation task is from 25.0% (chance level) to 100.0%, or 75.0%. Given the differences in ranges forthese networks, the change of accuracy in percentage was normalized by these respective ranges. Namely, the normalized change of accuracy was obtained by dividing the amount of change in accuracy by the size of the range of the accuracy of the corresponding task. The accuracy of $network_{(identity,location)}$ is lower than $network_{(location,location)}$ by $(97.3 - 29.9)/88.9 = 75.8\%$ and the accuracy of $network_{(identity,orientation)}$ is lower than $network_{(orientation,orientation)}$ by $(84.1 - 34.5)/75.0 = 66.1\%$. After the network had been trained to do the identity task, the accuracy of determining location decreased more in normalized percentage than did the accuracy of determining orientation. This difference in the amount of accuracy decrease is significant (p-value < 0.001).

We repeated the simulations about whether there is information about space in $network_{identity}$ and whether there is information about identity in $network_{space}$ with some different settings. The comparisons of accuracies between different networks when differentsample sizes, or different network parameter settings were used are shown in Table 5 and Table 6. We repeated the simulations for the following three alternative settings: (1) with 1200 images used as dataset; (2) with the number of

filters in each convolutional layer inbrain networks doubled from the first layer to the last layer (64, 128, 256 filters); (3) with the kernel sizes for the first, second, third convolutional layers in brain networks were reduced to 5 × 5, 2 × 2, 2 × 2 respectively; (4) with decoder networks which have 2 hidden layers; and (5) with decoder networks which have 50 units in each hidden layer. Only one setting (size of dataset, or number of filters, kernel sizes for brain networks, number of hidden layers, or number of units in hidden layers for decoder networks) was changed at a time. The results with these different settings are consistent with the results we obtained with regular settings.

We also repeated the decoding simulations using linear decoders. When obtaining the accuracies in each experimental setting, the linear decoders were always trained for 10 times and 10 testing accuracies were obtained for each condition after training (10 trainingand testing episodes). The input images were permuted each time and different sets of input images were selected from the whole dataset for training and testing during each episode. Unpaired two-samples t-tests were used to compare different accuracies and to determine the significance of the differences.

The average testing accuracies for different experimental settings are shown in Table 7. The comparisons of accuracies between different linear decoders are shown in Table 8. According to these results, unlike the results obtained using nonlinear decoders, the difference between the accuracies of $network_{(location,identity)}$, $network_{(orientation,identity)}$ and $network_{(space,identity)}$ are not significant. Though the accuracy of $network_{(space,shoes)}$ is significantly higher than the baseline (p-value = 0.003) and it is different from the nonlinear decoder result, the accuracy of $network_{(space,shoes)}$ is still significantly lower than$network_{(space,identity)}$ and it is consistent with the nonlinear decoder result. All of the otherresults shown in Table 8 are consistent with the results obtained using nonlinear decoders.

One additional comparison, not illustrated in Table 8, was made to compare the difference in the amount of accuracy decrease in percentage from $network_{(location,location)}$ to$network_{(identity,location)}$ versus from $network_{(orientation,orientation)}$ to $network_{(identity,orientation)}$. Again, we increased the difficulty for the location and orientation tasks by adding Gaussianwhite noise to the input images. With noisy input images, the accuracy of $network_{(identity,location)}$ is lower than $network_{(location,location)}$ by $(98.9 - 15.8)/88.9 = 93.5\%$ and the accuracy of $network_{(identity,orientation)}$ is lower than $network_{(orientation,orientation)}$ by $(94.4 - 28.7)/75.0 = 87.6\%$. After the network had been trained to do the identity task, the accuracy of determining location decreased more in normalized percentage than did the accuracy of determining orientation. This difference in the amount

of accuracy decrease is significant (p-value < 0.05). This result is consistent with the result obtained using nonlinear decoders.

Discussion

Using a computational modeling approach, we aimed to better understand whether the presence of identity and spatial properties in cortical areas important for space and object recognition have a functional role or not. We trained networks to do various object and spatial recognition tasks. We show that these networks actively retain non-task related information. Specifically, these networks retain different amounts of identity and spatial information (as shown for example by the amount of identity information retained by networks identity versus space; or that the space network retains more scrambled/unscramble identity information than type of shoe information) and different kinds of identity and spatial information (as shown for example by the greater retention of orientation than location information by the identity network). Each of these networks wasindependent and trained on a single task and had no cross connections from other networks. Hence, any non-task related properties that were retained in each of these networks were not coming from other networks. We repeated some simulations with different neural network parameters and the results were still consistent with our findings. It implies that our findings are robust and do not depend on specific parameter settings of the neural networks. In sum, based on our results, we (1) suggest that this different retained information about identity and space in the two pathways is functional, (2) demonstrate that this task irrelevant information need not come from another cortical stream or external source, and (3) show that in some cases the task irrelevant information may be necessary to accurately and optimally recognize and localize objects. Because our findings do not depend on specific parameter settings, they should also be valid for the biological brain though their structures may not be the same as our artificial networks.

Is There Information About Space in the Identity Network?

According to both the nonlinear decoder and linear decoder results, the accuracy ofthe decoder $network_{(identity,space)}$ is significantly higher than the accuracy of the decoder $network_{space\ baseline}$ when both of them were trained to decode information about space. This finding suggests it is possible to decode information about space from the activities ofthe second last layer units of $network_{identity}$. It

indicates that even though $network_{identity}$ was only trained to identify scrambled/unscrambled images, its later processing stage still had information about space when it was processing the input images in order to identify scrambled/unscrambled images. Furthermore, the accuracy of $network_{(identity,space)}$ is significantly lower than the accuracy of $network_{(space,space)}$. This may be because as information goes from the input layer to the second to last layer, some information about space may be lost because it is not useful to have very precise information about space for $network_{identity}$'s task (identifying scrambled/unscrambled images).

However, an important question is why the activities of the second to last layer units of $network_{identity}$ still contained spatial information. Was this spatial information actively kept by the network or was it just passively left in the network? While the network was processing input information in order to do a task, it would extract useful information from the inputs and eliminate useless information in the inputs. As a result, some information would be retained, and some information would be lost. "Actively kept by the network" means the network chose to keep spatial information when it was eliminating other useless information. "Passively left in the network" means the network did not choose to keep spatial information, the network just did not actively eliminate all spatial information. If the network did not choose to actively keep the spatial information,then whatever, if any, spatial information was passively left in the network should be equivalent across trained networks and there should be no difference in the spatial information retained in $network_{identity}$ and $network_{shoes}$.

In order to answer this question, the accuracy of $network_{(shoes,space)}$ is compared with the accuracy of $network_{(identity,space)}$. The result is that the accuracy of $network_{(shoes,space)}$ is significantly lower than the accuracy of $network_{(identity,space)}$, which means that there was significantly less space information retained in the activities of the second to last layer units of $network_{shoes}$. This result is the same for both nonlinear and linear decoders. It is likely because when identifying a feature (feature recognition; or thetypes of the shoes) does not need as much spatial information compared to identifying scrambled/unscrambled images (global recognition; or identifying combinations of features). These findings indicate that the space information was actively maintained by $network_{identity}$ even though it was trained to do the identity task. Though most studies assume spatial information in the ventral stream is coming from the dorsal stream, our results indicate the information may be retained/built up within the ventral stream.

These findings agree with Hong et al. (2016)'s computational modeling results. They also used hierarchical convolutional neural networks (HCNN) to model the ventral visual cortical pathway. They trained the HCNN to do category estimation tasks. They took the neural activities of the top hidden layer of their HCNN while training and used these artificial hidden layer neural activities and a decoder to perform category-orthogonalestimation tasks. They found the network performance on category-orthogonal estimation tasks improved as training proceeds. It suggests that the category-orthogonal information was extracted by the HCNN when the HCNN was trained to do category estimation tasks,which is similar to what we found.

What Kind of Spatial Information was Actively Maintained More in the Identity Network?

There are different kinds of spatial information, including the locations of the object images (defined as the location of the center of the object), the orientations of the object images, and the spatial alignments and orders of the parts of the object images and so forth. Two kinds of spatial information were examined in this study: object location and part/object orientation. According to the results presented in the results section, the accuracy of determining object location decreased significantly more than the accuracy of determining part/object orientation after the network had been trained to do the identity task when noisy input images were used. The amounts of accuracy decrease are comparable because they have been normalized according to their different chance level accuracies (see Results section). These findings suggest that the information loss about part/object orientation is smaller than the information loss about object location in the identity network. That is, these findings indicate that the identity network in our study actively maintained more information about part/object orientation than object location.

Why the Identity Network Actively Maintained More Information About Orientation?

To answer why the identity network actively maintained more information aboutpart/object orientation, the accuracies of $network_{(location,identity)}$ and $network_{(orientation,identity)}$ were compared. The assumption is $network_{location}$ would retain more information about object location in its second to last layer while $network_{orientation}$would retain more information about part/object orientation in its second to last layer.

According to the nonlinear decoder results, the accuracy of $network_{(orientation,identity)}$ is significantly higher than the accuracy of the $network_{(location,identity)}$. These findings suggest that part/object orientation information is more important for the identity task. It could be because in order to tell whether the object images are in the unscrambled order ornot, the network needs to determine the part/object orientation. This is because the definition of the order of parts depends on the part/object orientation, as explained in the methods section. However, information about object location is less important for our identity task because in our task, the location of the object image is irrelevant for identifying the scrambled/unscrambled object image. This suggests that spatial information is preserved along the ventral pathway because it is behaviorally useful.

Likewise, given that some object recognition tasks can require an ability to disambiguate the same or similar objects in different locations (Garcia & Buffalo, 2020; Suzuki, Miller, &Desimone, 1997; Byun & Lee, 2010), we would expect that if our identity network was trained with such an object recognition task, spatial location information would be preserved to a greater extent than what we report in our study. Our findings suggest that spatial information is preserved along the ventral pathway when this information is behaviorally useful for the identification task.

When linear decoders are used, the difference between the accuracies of $network_{(orientation,identity)}$ and $network_{(location,identity)}$ are not significant, but in a similar direction as the nonlinear decoders. It is possible that the current methods and experiments are not sensitive enough to detect a small difference using linear decoders. Alternatively, it is possible that the amount of linearly decodable information about identity in $network_{location}$ and $network_{orientation}$ is not significantly different. However, it is important to point out that the question we were trying to answer is whether the question of which of the orientation or location is more important for doing the identity task. The brain is the subject who does the identity task and the brain is nonlinear. Therefore, we feel the results obtained using nonlinear decoders is most relevant to understanding neuraldecoding. Nevertheless, future work is needed to better understand and test the reliabilityof this difference between nonlinear and linear decoders.

Is There Information About Object Identity (Scrambled/Unscrambled or Type of Shoes) in the Space Network?

The accuracy of $network_{(space,identity)}$ is significantly higher than the accuracy of $network_{identity\ baseline}$ when both of them were trained to identify scrambled/unscrambled images (the

identity task). This indicates that even though $network_{space}$ was trained to determine the location and orientation of the parts/images, its later processing stage still had some information that was necessary to do the identity task. It may be because the dorsal pathway processes parts/objects or face representation information in order to better recognize the object/face's configural identity (Freud et al., 2016). For example, though face recognition is believed to be mainly processed by the ventral pathway (Grill-Spector, Weiner, Gomez, Stigliani, & Natu, 2018), in a same–different face detection task, configural but not featural processing of faces was found in the posterior dorsal pathway (Zachariou, Christine V. Nikas, Gotts, & Ungerleider, 2016). TMS centered on the parietal regions impaired performance on configural but not featural face difference detection (Zachariou et al., 2016), which suggested that the dorsal pathway processing is important for the intact perception of object configural information.

Furthermore, the accuracy of $network_{(space,identity)}$ is significantly lower than the accuracy of $network_{(identity,identity)}$. Likely, as information goes from the input layer to thesecond to last layer, some information about scrambled/unscrambled identity may be lost because it is not useful to have very precise information about scrambled/unscrambled identity for $network_{space}$'s task (identifying locations and orientations).

On the other hand, according to the nonlinear decoder results, the accuracy of $network_{(space,shoes)}$ is not significantly higher than the accuracy of $network_{shoes\ baseline}$.According to the linear decoder results, although the accuracy of $network_{(space,shoes)}$ is significantly higher than the accuracy of $network_{shoes\ baseline}$, the accuracy of $network_{(space,shoes)}$ is still significantly lower than $network_{(space,identity)}$. Together, thesefindings indicate that $network_{space}$'s later processing stage retains less information aboutthe identity of shoes (whether it is a sandal or a closed shoe). The reason that the space network contains information about object scrambled/unscrambled identity may be because the global object recognition information (scrambled/unscrambled) retained is relevant to the $network_{space}$'s task, but the specific feature recognition information (theidentity of shoes) is not relevant.

When the space network was determining object orientation or location, it did notneed to know object identity (scrambled/unscrambled). So what might be the benefit of retaining this information? The reason may be relevant to some studies about tool processing. Recent studies found that tool sensitivity undergoes further refinement betweenthe ages of 4 and 8 years, which indicated that sensitivity to objects in the dorsal pathway may require more motor experience and

learning during childhood (Kersey, Clark, Lussier, Mahon, & Cantlon, 2016; Freud et al., 2016). Young children are more likely to categorize objects based on their physical features (e.g., material of the object) rather than their function-related features (Smith, Jones, & Landau, 1996; Landau, Smith, & Jones, 1998). These studies indicate that the dorsal pathway may retain object function-related features (including the spatial relation of parts, important in scrambled/unscrambled) when it is trained to do spatial tasks during motor learning (for tasks that require localization and orientation) and such motor training can help people learn how to categorize different kinds of tools based on these function-related features. The shoe identity may be more similar to an object physical feature (and does not require spatial relation of parts), and thus it is not retained by the dorsal pathway. These arguments would need to be confirmed by future studies.

Independence not Interactions: What These Simulations Imply About the Ventral and Dorsal Cortical Visual Pathways

Some previous studies (A. B. Sereno & Lehky, 2011; A. B. Sereno et al., 2020) havefound that the ventral pathway had representations about space, but these spatial representations were different from the spatial representations in the dorsal pathway. The spatial representations in the ventral pathway were topological ("categorical") whereas thespatial representations in the dorsal pathway were precise and accurate ("coordinate"). They have suggested that it might be that objects' shape and spatial information are differently and independently constructed within each pathway in order to achieve differentfunctions (object recognition or spatial recognition). In addition, Freud et al. (2015) also suggested that object representations in the dorsal pathway can be computed independently from those in the ventral pathway. According to our simulation results using nonlinear decoders, the spatial information retained in $network_{identity}$ had more explicit information about orientation than location. Information about orientation is more useful for identifying scrambled/unscrambled images. Nevertheless, given that $network_{identity}$ is used to model the ventral pathway and $network_{space}$ is used to model the dorsal pathway,these results agree with previous experimental findings as well as the interpretation that objects' identity and spatial information are differently and independently constructed within each pathway in order to achieve different functions as opposed to the idea that these "crossed" signals are coming from the other stream (Zachariou et al., 2014; van Polanen & Davare, 2015). Hong et al. (2016) found that spatial information increased along the ventral stream, and their computational modeling also suggested

that spatial information that is present within the pathway is extracted along the ventral pathway, so itbecomes more explicit at the later processing stages of the ventral pathway.

Some previous studies also found the dorsal pathway had representations about objects' shapes (A. B. Sereno & Maunsell, 1998; M. E. Sereno, Trinath, Augath, & Logothetis, 2002; Konen & Kastner, 2008) and some have argued that these representations of objects' shapes are different from the object representations in the ventral pathway (Lehky & Sereno, 2007; Janssen, Srivastava, Ombelet, & Orban, 2008). According to our simulation results, the identity information retained in $network_{space}$ had more information about the scrambled/unscrambled identity (global recognition) than it did about shoes identity (feature recognition). These findings likely occurred because $network_{space}$ extracted spatial information about the arrangements of features from the inputs and this extracted spatial information was useful for the scrambled/unscrambled (or global) identity task but did not help with the shoe identity (feature recognition) task.

What Might be a Reason for why There are Two Relatively Segregated Visual Pathways in the Brain?

Suppose the ventral pathway in the brain works similar to $network_{identity}$ and the dorsal pathway works similar to $network_{space}$. There are two possible ways to determine anobject's identity and spatial information. One way is to use a single pathway to process visual inputs and determine the object's identity and spatial information (e.g. location andorientation) at the same time. Another way is to segregate these goals (identity and space)and use two separate pathways to process visual inputs. In this dual stream method, one pathway processes the visual inputs and is critical for object identity, whereas the other pathway processes the same visual inputs and is important for spatial information and visuomotor control; with separate cortical regions/streams responsible for the object's identity and spatial information. Experimental evidence has shown that the brain is using the second way to determine object's identity and spatial information (Ungerleider & Mishkin, 1982), but why?

To address this question, the accuracies of $network_{combine\ identity\ and\ space}$ and $network_{separate\ identity\ and\ space}$ were compared. $network_{combine\ identity\ and\ space}$ was used to simulate the process of doing the identity and spatial tasks using a single pathway and $network_{separate\ identity\ and\ space}$ were used to simulate the process of doing the identity andspatial tasks

using two separate pathways. The testing accuracy of $network_{separate\ identity\ and\ space}$ is significantly higher than the accuracy of $network_{combine\ identity\ and\ space}$. It implies that when two pathways are used to determine anobject's identity and spatial information separately, the neural network has better performance. Our findings suggest there are advantages for the brain to use two separate pathways to determine identity and spatial information.

On the other hand, according to the results discussed in previous sections, if there isonly a single combined pathway and this pathway processes space information first, then ata later time point processes identity, there would be less information about object identity. As a result, this single pathway structured brain wouldn't be able to do object recognition accurately. In addition, if a single pathway processed object identity first, then this pathway would lose information about space and wouldn't be able to accurately determine the locations and orientations of the objects.

In summary, in order to accurately identify objects and accurately determine the location and orientation of objects, these findings suggest that two separate pathways are advantageous in order to process the same input (visual information) in different ways fordifferent tasks or goals. However, in some tasks or conditions, the goal may require coordination of the information from these segregated pathways (e.g., reaching for objectsonly if they are edible). In these cases, processing information differently using multiple separate pathways may cause a binding problem (Treisman, 2002). We suggest here, that the binding problem may be lessened by using the spatial information contained in the identity network and object identity information in the spatial network.

We are not aware of any published study examining exactly how much information and what kinds of information are extractable from neural networks solving these different tasks jointly or separately. These are computationally tractable questions that are important and timely. Our simulations using CNNs ignore a lot of details of real world tasks and real biological neural networks (e.g., different cell types and connectivity or the fact that ventral stream has more cortical areas than dorsal). These simplifications are important and necessary to make direct computational comparisons possible. Our intent isnot to claim that the simulation findings emulate physiological conditions of these brain pathways. Our claims concern whether or not there could be a computational need for properties retained in distinct pathways or computational need for separate pathways for recognition and non-recognition tasks.

We repeated the simulations of decoding space information from $network_{identity}$ anddecoding identity information from $network_{space}$ with some different parameter settings in brain networks or

decoder networks. Because our findings do not strongly depend on specific parameter settings of the brain networks or decoder networks, our findings may also be valid for the brain though the structures of biological neural networks may be different. In sum, our computational findings are adequately supported by the results shown above and certainly have relevance to better understanding of the computational constraints of neural computation.

Limitations and Future Directions

First, we constrained the alignments of the three parts in each object image to always be the same, and we defined object orientation according to alignment of parts. In this case, the identity of an object (scrambled/unscrambled) depends on the simple 1D order of parts and the 1D order of parts depend on the alignment of parts. However, in reality, the identity of an object may not be dependent on the 1D order or alignment of parts. In other words, the dependency between object identity and the alignment of parts is just one simple example of the possible dependency between object identity and spatial information of parts. These dependencies in real life could be different and more complex. For example, an object may be unscrambled, even when different parts of an object do not have the same alignment (e.g. the yoga pose of Uttanasana, standing forward bend, wherethe head is upside down, but the feet are right-side up). In this case, the identity of an object (scrambled/unscrambled) no longer depends on the alignment of parts, but it may still depend on other 2D or 3D spatial information of parts (such as the relative distance between parts, relative locations of parts, and other topological information) that affect object identity recognition. A previous study has found that object recognition accuracy can be improved by taking into account the spatial distribution of object parts. They found this via mathematical modeling and did not use artificial neural networks (Morales-González & García-Reyes, 2013). Many objects are different not because they have different physical features like color or texture, but because they have different spatialrelations between parts. Therefore, it is very likely that in general, the identity artificial neural network retains some spatial information of parts because this information increasesobject recognition accuracy. In our current study, we demonstrated that the identity network retains some spatial information when object identity is dependent on the alignment of parts. In addition, we ran simulations where the image order was only dependent on the order of parts and found no differences in the major findings we report. We used relatively simple object images to make sure the variables used in the simulated experiments are well defined and controlled (e.g. all objects consist

of three parts, objects' orientations and locations are clearly defined). If more complex and realistic images were used, then the objects in the images would have much more variations and it would be more difficult to define and control the variables that might increase or affect the computational differences we report. In the future, it is important to examine object recognition accuracy in more general settings where more realistic images are used and the object identity is based on other higher dimensional or topological spatial information.

In addition, motion is also an important property of both the ventral and dorsal cortical visual pathways (M. E. Sereno et al., 2002). As our goal was proof of computational concept, we did not use more complex stimuli and models which could complete tasks using motion. In the future, it would be interesting to use more complex artificial neural network models to test whether our findings still hold when the networks are more elaborate and can respond to more complex moving stimuli, scenes, and tasks. However, given the vast known variety, for example in cell type, receptors, connectivity, or modules, as well as variety of stimuli, tasks and experiences and training that a single brain encounters by the age of 20, even a CNN model that generalized to more complex stimuli, multiple tasks as well as predicted neural responses in visual cortical areas, would still remain disputable as an accurate model of a real human brain.

Finally, we used a supervised learning rule. Many researchers think the brain is mainly using unsupervised learning and reinforcement learning to learn how to accomplish different tasks (Hinton & McClelland, 1988). Although previous work has argued that supervised learning may be biologically plausible (Lillicrap et al., 2020; Whittington & Bogacz, 2019), it would be interesting to examine whether more biologically plausible learning rules affect any of the findings we report. Future studies should also try to localizeobjects more accurately (give more accurate coordinates when localizing objects). In addition, the current study can only localize one object at a time. It would be interestingin future work to use more realistic and biologically plausible networks which can localize and identify multiple objects at the same time.

Conclusion

In summary, our simulations imply that both ventral and dorsal cortical visual pathways contain information about both identity and space, even when trained with a single identity or location task. We have also shown that the ventral pathway does not contain all types of spatial information equally and the dorsal pathway does not contain alltypes of object identity information

equally. In our simulations and tasks, there was more orientation information than location information retained in the ventral pathway.

Likewise, we found that in the dorsal pathway there was more information retained about the whole object (global recognition) than the information about individual features (feature recognition). These modeling findings suggest that the object information retainedin the dorsal pathway and spatial information retained in the ventral pathway are not the same properties respectively retained in the ventral and dorsal pathways themselves. The retained object and spatial information in the dorsal and ventral pathway, respectively, appear to be those aspects of identity and space that are most needed to accomplish spatial and identity tasks, respectively. As a result, the modeling suggests that the identityand spatial information retained in the two pathways needs to be different in order to accurately accomplish different kinds of tasks. Furthermore, we show that two separate pathways are needed in order to process visual information in different ways so that the brain can accomplish different kinds of visual tasks more accurately. Using a computationalapproach, we provide a framework to test the properties and functional consequences of two independent visual pathways (with no cross connections) and show that the findings can provide insight into recent contradictory findings in systems neuroscience.

Acknowledgements

The authors would like to thank Sidney Lehky and Margaret Sereno for commentson the manuscript. This work was partially supported by start-up funds from Purdue University (ABS).

References

Bracci, S., Ritchie, J. B., & de Beeck, H. O. (2017). On the partnership between neural representations of object categories and visual features in the ventral visual pathway. *Neuropsychologia*, *105*, 153-164. https://doi.org/10.1016/j.neuropsychologia.2017.06.010

Byun, J., & Lee, I. (2010). Disambiguation of similar object-place paired associations and the roles of the brain structures in the medial temporal lobe. *Experimental Neurobiology*, *19*(1), 15-22. https://doi.org/10.5607/en.2010.19.1.15

Colby, C. L., & Goldberg, M. E. (1999). space and attention in parietal cortex. *Annual Review of Neuroscience*, *22* , 319-349. https://doi.org/10.1146/annurev.neuro.22.1.319

Felleman, D., & Essen, D. V. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*(1), 1-47. https://doi.org/10.1093/cercor/1.1.1.

Freud, E., Plaut, D. C., & Behrmann, M. (2016). 'what' is happening in the dorsal visualpathway. *Trends in Cognitive Sciences*, *20*(10), 773-784. https://doi.org/10.1016/j.tics.2016.08.003

Freud, E., Rosenthal, G., Ganel, T., & Avidan, G. (2015). Sensitivity to object impossibility in the human visual cortex: Evidence from functional connectivity. *Journal of Cognitive Neuroscience*, *27*(5), 1029-1043. https://doi.org/10.1162/jocn_a_00753

Garcia, A. D., & Buffalo, E. A. (2020). Anatomy and function of the primate entorhinalcortex. *Annual Review of Vision Science*, *6*, 411-432. https://doi.org/10.1146/annurev-vision-030320-041115

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex andits role in object recognition. *Vision Research*, *41* (10-11), 1409-1422. https://doi.org/10.1016/S0042-6989(01)00073-6

Grill-Spector, K., Weiner, K. S., Gomez, J., Stigliani, A., & Natu, V. S. (2018). The functional neuroanatomy of face perception: from brain measurements to deep neural networks. *Interface Focus*. https://doi.org/10.1098/rsfs.2018.0013

Hinton, G. E., & McClelland, J. L. (1988). Learning representations by recirculation. *Neural Information Processing Systems*, 358-366. https://doi.org/10.1016/j.tics.2018.12.005

Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity inthe distributed human neural system for face perception. *Nature Neuroscience*, *3*(1), 80-84. https://doi.org/10.1038/71152

Hong, H., Yamins, D. L. K., Majaj, N. J., & DiCarlo, J. J. (2016). Explicit information for category-orthogonal object properties increases along the ventral stream. *Nature Neuroscience*, *19*(4), 613-630. https://doi.org/10.1038/nn.4247

Janssen, P., Srivastava, S., Ombelet, S., & Orban, G. A. (2008). Coding of shape and position in macaque lateral intraparietal area. *The Journal of Neuroscience*, *28*(26),6679-6690. https://doi.org/10.1523/JNEUROSCI.0499-08.2008

Kersey, A. J., Clark, T. S., Lussier, C. A., Mahon, B. Z., & Cantlon, J. F. (2016). Development of tool representations in the dorsal and ventral visual object processingpathways. *Cerebral Cortex*, *26*, 3135-3145. https://doi.org/10.1093/cercor/bhv140

Konen, C. S., & Kastner, S. (2008). Two hierarchically organized neural systems for object information in human visual cortex. *Nature Neuroscience*, *11*(2), 224-231. https://doi.org/10.1038/nn2036

Kourtzi, Z., & Kanwisher, N. (2000). Activation in human mt/mst by static images withimplied motion. *Journal of Cognitive Neuroscience*, *12*(1), 48-55. https://doi.org/10.1162/08989290051137594

Landau, B., Smith, L., & Jones, S. (1998). Object shape, object function, and objectname. *Journal of Memory and Language*, *38*(1), 1-27. https://doi.org/10.1006/jmla.1997.2533

Lehky, S. R., Peng, X., McAdams, C. J., & Sereno, A. B. (2008). Spatial modulation of primate inferotemporal responses by eye position. *PLoS ONE*, *3*(10), e3492. https://doi.org/10.1371/journal.pone.0003492

Lehky, S. R., & Sereno, A. B. (2007). Comparison of shape encoding in primate dorsal andventral visual pathways. *Journal of Neurophysiology*, *97*, 307-319. https://doi.org/10.1152/jn.00168.2006

Lillicrap, T. P., Santoro, A., Marris, L., Akerman, C. J., & Hinton, G. (2020). Backpropagation and the brain. *Nature Reviews Neuroscience*, *21*, 335-346.https://doi.org/10.1038/s41583-020-0277-3

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience*, *19*, 577-621. https://doi.org/10.1146/annurev.ne.19.030196.003045

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences*, *6*, 414-417.https://doi.org/10.1016/0166-2236(83)90190-X

Morales-González, A., & García-Reyes, E. B. (2013). Simple object recognition based on spatial relations and visual features represented using irregular pyramids. *MultimediaTools and Applications*, *63*, 875-897. https://doi.org/10.1007/s11042-011-0938-3

Nowicka, A., & Ringo, J. L. (2000). Eye position-sensitive units in hippocampal formation and in inferotemporal cortex of the macaque monkey. *European Journal of Neuroscience*, *12*, 751-759. https://doi.org/10.1046/j.1460-9568.2000.00943.x

Perrett, D. I., Hietanen, J. K., Oram, M. W., & Benson, P. J. (1992). Organization andfunctions of cells responsive to faces in the temporal cortex. *Philosophical Transactions of the RoyalSsociety B*, *335*, 23-30. https://doi.org/10.1098/rstb.1992.0003

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations byback-propagating errors. *Nature*, *323*, 533-536. https://doi.org/10.1038/323533a0

Sereno, A. B., & Lehky, S. R. (2011). Population coding of visual space: comparison of spatial representations in dorsal and ventral pathways. *Frontiers in Computational Neuroscience*, *4*(159), 1-16. https://doi.org/10.3389/fncom.2010.00159

Sereno, A. B., Lehky, S. R., & Sereno, M. E. (2020). Representation of shape, space, and attention in monkey cortex. *Cortex, 122,* 40-60. https://doi.org/10.1016/j.cortex.2019.06.005

Sereno, A. B., & Maunsell, J. H. R. (1998). Shape selectivity in primate lateral intraparietal cortex. *Nature*, *395*(6701), 500-503. https://doi.org/10.1038/26752

Sereno, A. B., Sereno, M. E., & Lehky, S. R. (2014). Recovering stimulus locations using populations of eye-position modulated neurons in dorsal and ventral visual streams of non-human primates. *Frontiers in Integrative Neuroscience*, *8*(28), 1-20. https://doi.org/10.3389/fnint.2014.00028

Sereno, M. E., Trinath, T., Augath, M., & Logothetis, N. K. (2002). Three-dimensionalshape representation in monkey cortex. *Neuron*, *33*(4), 635-652. https://doi.org/10.1016/S0896-6273(02)00598-6

Smith, L. B., Jones, S. S., & Landau, B. (1996). Naming in young children: A dumb attentional mechanism? *Cognition*, *60*(2), 143-171. https://doi.org/10.1016/0010-0277(96)00709-3

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). *Journal of Machine Learning Research*, *15*(56), 19291958. Retrieved from http://jmlr.org/papers/v15/srivastava14a.html

Suzuki, W. A., Miller, E. K., & Desimone, R. (1997). Object and place memory in themacaque entorhinal cortex. *Journal of Neurophysiology*, *78*, 1062-1081. https://doi.org/10.1152/jn.1997.78.2.1062

Treisman, A. (2002). The binding problem. *Current Opinion in Neurobiology*, *6*(2), 171-178. https://doi.org/10.1016/S0959-4388(96)80070-5

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In M. Goodale, D. J. Ingle, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549-586).Cambridge, MA: MIT Press.

van Polanen, V., & Davare, M. (2015). Interactions between dorsal and ventral streams for controlling skilled grasp. *Neuropsychologia*, *79*(Pt B), 186-191. https://doi.org/10.1016/j.neuropsychologia.2015.07.010

Whittington, J. C., & Bogacz, R. (2019). Theories of error back-propagation in the brain. *Trends in Cognitive Sciences*, *23*(3), 235-250.https://doi.org/10.1016/j.tics.2018.12.005

Xiao, H., Rasul, K., & Vollgraf, R. (2017). Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. https://doi.org/arXiv:1708.07747

Xu, Z., Wu, W., Winter, S. S., Mehlman, M. L., Butler, W. N., Simmons, C. M., Harvey, R. E., Berkowitz, L. E., Chen, Y., Taube, J. S., Wilber, A. A., & Clark, B. J. (2019). A comparison of neural decoding methods and population coding across thalamo-cortical head direction cells. *Frontiers in Neural Circuits*, *13*, 75. https://doi.org/10.3389/fncir.2019.00075

Zachariou, V., Christine V. Nikas, Z. N. S., Gotts, S. J., & Ungerleider, L. G. (2016). Spatial mechanisms within the dorsal visual pathway contribute to the configuralprocessing of faces. *Cerebral Cortex, 27* , 4124–4138. https://doi.org/10.1093/cercor/bhw224

Zachariou, V., Klatzky, R., & Behrmann, M. (2014). Ventral and dorsal visual stream contributions to the perception of object shape and object location. *Journal of Cognitive Neuroscience*, *26*(1), 189-209. https://doi.org/10.1162/jocn_a_00475

Table 1. *Inputs and task goals of different decoders when the brain network was trained with a different or same task.*

Decoder Name	Take the second to last layer activities from	To do the task
$network_{(space,identity)}$	$network_{space}$	identity
$network_{(location,identity)}$	$network_{location}$	identity
$network_{(orientation,identity)}$	$network_{orientation}$	identity
$network_{(identity,space)}$	$network_{identity}$	space
$network_{(shoes,space)}$	$network_{shoes}$	space
$network_{(identity,location)}$	$network_{identity}$	location
$network_{(identity,orientation)}$	$network_{identity}$	orientation
$network_{(identity,shoes)}$	$network_{identity}$	shoes
$network_{(space,shoes)}$	$network_{space}$	shoes
$network_{(space,identity,shoes)}$	$network_{(space,identity)}$	shoes
$network_{(identity,identity)}$	$network_{identity}$	identity
$network_{(space,space)}$	$network_{space}$	space

Table 2. *Average testing accuracies in percentage (%) ± standard deviations (%) for brain networksand nonlinear decoder networks. The column headers are the names of the brain networks. The row headers are the kinds of information that decoder networks were trying to decode. The data are accuracies obtained by various decoder networks except for the data in the rowlabeled with "Brain". The row header "Brain" means there is no decoder and it is the accuracy obtained by the brain network. Definitions of decoder networks are listed in Table 1 and Table 2. The data for simulations that were not conducted are labeled "NA".*

Brain / Decoders	Identity	Space	Shoes	Orientation	Location
Baseline Accuracy (Decode from the untrained brain network)	60.7 ± 2.0	48.0 ± 3.9	60.7 ± 2.4	38.2 ± 1.1 (noisy inputs)	44.2 ± 4.8 (noisy inputs)
No Decoder (Brain NetworkAccuracy)	80.2 ± 1.8	85.7 ± 3.1	73.4 ± 1.9	82.6 ± 2.3 (noisy inputs)	97.0 ± 1.1 (noisy inputs)
Identity	81.6 ± 0.8	71.2 ± 2.8	NA	72.2 ± 1.2	65.0 ± 2.0
Space	75.8 ± 2.9	86.5 ± 1.2	61.0 ± 1.6	NA	NA
Shoes	NA	57.5 ± 2.1	NA	NA	NA
Orientation	34.5 ± 3.9 (noisy inputs)	NA	NA	84.1 ± 0.7 (noisy inputs)	NA
Location	29.9 ± 4.1 (noisy inputs)	NA	NA	NA	97.3 ± 0.5 (noisy inputs)

Table 3. *Average testing accuracies for* $network_{combine\ identity\ and\ space}$ *and* $network_{separate\ identity\ and\ space}$. *The* $network_{combine\ identity\ and\ space}$ *and the* $network_{separate\ identity\ and\ space}$ *are used to simulate object identification and localization withone pathway or two separate pathways.*

Network	Average Accuracy(%)	Chance Level (%)	Standard Deviation (%)
$network_{combine\ identity\ and\ space}$	72.8	1.4	2.1
$networks_{separate\ identity\ and\ space}$	76.8	1.4	1.5

Table 4. *Comparisons of testing accuracies between different networks. The first two sections examine whether there is information about space in the identity network and why there is information about space in the identity network. The next section examines whether there is information about identity and what kind of identity information is in the space network.The final section compares testing accuracies of a network doing the identity and spatial tasks using two separate pathways with a network doing the identity and spatial tasks usinga single pathway.*

Comparisons of testing accuracies between different networks			
Network 1	Network 2	Average Difference in Accuracy (%) (∗: $p < 0.05$, ∗ ∗: $p < 0.01$, ∗ ∗ ∗: $p < 0.001$)	p-value
$network_{(identity,sace)}$ $network_{(identity,space)}$ $network_{(identity,space)}$	$network_{space\ baseline}$ $network_{(space,space)}$ $network_{(shoes,space)}$	27.8 ∗ ∗ ∗ -10.7 ∗ ∗ ∗ 14.8 ∗ ∗ ∗	< 0.001 < 0.001 < 0.001
$network_{(location,identity)}$ $network_{(orientation,identity)}$ $network_{(location,identity)}$	$network_{(space,identity)}$ $network_{(space,identity)}$ $network_{(orientation,identity)}$	-6.2 ∗ ∗ ∗ 1.0 -7.2 ∗ ∗ ∗	< 0.001 0.324 < 0.001
$network_{(space,identity)}$ $network_{(space,identity)}$ $network_{(space,shoes)}$	$network_{identity\ baseline}$ $network_{(identity,identity)}$ $network_{shoes\ baseline}$	10.5 ∗ ∗ ∗ -10.4 ∗ ∗ ∗ -3.2 ∗ ∗	< 0.001 < 0.001 0.005
$networks_{separate\ identity\ and\ space}$	$network_{combine\ identity\ and\ space}$	4.0 ∗ ∗ ∗	< 0.001

Table 5. *Average testing accuracies for the some networks with different settings (different dataset size, or different number of filters, or different kernel sizes). Only one setting was changed at a time. Definitions of decoder networks are listed in Table 1 and Table 2. If labeled with "1200 samples", then 1200 images were used as dataset. If labeled with "increase filters", then the number of filters in each convolutional layer doubled from the first layer to the lastlayer (64, 128, 256 filters). If labeled with "different kernel sizes", then the kernel sizes for the first, second, third convolutional layers were reduced to* 5×5, 2×2, 2×2 *respectively. If labeled with "2 layer decoder", then the decoder with 2 hidden layers was used. If labeled with "50 units decoder", then the decoder with 50 units in each hidden layer was used. For the other networks, 600 images and regular parameter settings shown in Figure 3 and Figure 4 were used.*

Network	Average Accuracy(%)	Chance Level (%)	Standard Deviation (%)
$network_{identity\ baseline}$	60.7	50.0	2.0
$network_{space\ baseline}$	48.0	2.8	3.9
$network_{identity\ baseline}\ 1200_{samples}$	77.7	50.0	2.1
$network_{identity\ baseline\ increase\ filters}$	69.5	50.0	2.0
$network_{identity\ baseline\ different\ kernels}$	67.3	50.0	2.1
$network_{space\ baseline}\ 1200_{samples}$	57.6	2.8	2.1
$network_{space\ baseline\ increase\ filters}$	62.4	2.8	2.7
$network_{space\ baseline\ different\ kernels}$	50.4	2.8	2.9
$network_{identity\ baseline}\ 2_{layer\ decoder}$	59.4	50.0	2.1
$network_{space\ baseline}\ 2_{layer\ decoder}$	48.9	2.8	3.7
$network_{identity\ baseline}\ 50_{units\ decoder}$	60.0	50.0	2.5
$network_{space\ baseline}\ 50_{units\ decoder}$	41.4	2.8	5.9
$network_{(identity,space)}$	75.8	2.8	2.9
$network_{(identity,space)}\ 1200_{samples}$	79.3	2.8	2.0
$network_{(identity,space)\ increase\ filters}$	67.5	2.8	1.9
$network_{(identity,space)\ different\ kernels}$	69.5	2.8	4.0
$network_{(space,identity)}$	71.2	50.0	2.8
$network_{(space,identity)}\ 1200_{samples}$	87.4	50.0	0.9

Table 5 continued

$network_{(space,identity)\ increase\ \ filters}$	75.5	50.0	2.5
$network_{(space,identity)\ different\ kernels}$	80.8	50.0	1.4
$network_{(identity,space)}\ 2_{layer\ decoder}$	75.4	2.8	1.8
$network_{(space,identity)}\ 2_{layer\ decoder}$	70.4	50.0	1.3
$network_{(identity,space)}\ 50_{units\ decoder}$	73.1	2.8	3.3
$network_{(space,identity)}\ 50_{units\ decoder}$	69.8	50.0	3.4

Table 6. *Comparisons of testing accuracies between different networks. The first section examines whether there is information about space in the identity network and whether there is information about identity in the space network when number of samples = 1200. The second and third sections examine whether there is information about space in the identity network and whether there is information about identity in the space network when different network parameter settings were used. The fourth and fifth sections examine whether the results would change when decoders with different number of hidden layers or different number of units were used. If labeled with "1200 samples", then 1200 images were used as dataset. If labeled with "increase filters", then the number of filters in each convolutional layer doubled from the first layer to the last layer (64, 128, 256 filters). Iflabeled with "different kernel sizes", then the kernel sizes for the first, second, third convolutional layers were reduced to* 5×5, 2×2, 2×2 *respectively. If labeled with "2 layerdecoder", then the decoder with 2 hidden layers was used. If labeled with "decoder 50 units",then the decoder with 50 units in each hidden layer was used. For the other networks, 600 images and regular parameter settings shown in Figure 3 and Figure 4 were used.*

Comparisons of testing accuracies between different networks			
Network 1	Network 2	Average Difference in Accuracy (%) (* * *: $p < 0.001$)	p-value
$network_{(identity,space)}$ $1200_{samples}$	$network_{space\ baseline}$ $1200_{samples}$	21.8 * * *	< 0.001
$network_{(space,identity)}$ $1200_{samples}$	$network_{(identity\ baseline)}$ $1200_{samples}$	9.8 * * *	< 0.001
$network_{(identity,space)}$ *increase filters*	$network_{space\ baseline\ increase\ filters}$	5.1 * * *	< 0.001
$network_{(space,identity)}$ *increase filters*	$network_{(identity\ baseline)}$ *increase filters*	6.0 * * *	< 0.001
$network_{(identity,space)}$ *different kernels*	$network_{(space\ baseline)}$ *different kernels*	19.1 * * *	< 0.001
$network_{(space,identity)}$ *different kernels*	$network_{(identity\ baseline)}$ *different kernels*	13.5 * * *	< 0.001
$network_{(identity,space)}$ $2_{layer\ decoder}$	$network_{(space\ baseline)}$ $2_{layer\ decoder}$	26.5 * * *	< 0.001
$network_{(space,identity)}$ $2_{layer\ decoder}$	$network_{(identity\ baseline)}$ $2_{layer\ decoder}$	11.0 * * *	< 0.001
$network_{(identity,space)}$ $50_{units\ decoder}$	$network_{(space\ baseline)}$ $50_{units\ decoder}$	31.7 * * *	< 0.001
$network_{(space,identity)}$ $50_{units\ decoder}$	$network_{(identity\ baseline)}$ $50_{units\ decoder}$	9.8 * * *	< 0.001

Table 7. *Average testing accuracies in percentage (%) ± standard deviations (%) for linear decoders.The column headers are the names of the brain networks. The row headers are the kinds ofinformation that linear decoder were trying to decode. The data are accuracies obtained by various decoder networks. Definitions of decoders are listed in Table 1 and Table 2. The data for simulations that were not conducted are labeled "NA".*

Brain / Decoders	Identity	Space	Shoes	Orientation	Location
Baseline Accuracy (Decode from the untrained brain network)	52.0 ± 5.1	2.8 ± 1.2	53.3 ± 3.5	24.0 ± 5.3 (noisy inputs)	10.9 ± 5.2 (noisy inputs)
Identity	93.6 ± 1.9	63.2 ± 4.3	NA	63.0 ± 4.2	61.4 ± 6.5
Space	76.1 ± 4.1	96.1 ± 1.7	68.8 ± 6.0	NA	NA
Shoes	NA	58.7 ± 3.7	NA	NA	NA
Orientation	28.7 ± 3.4 (noisy inputs)	NA	NA	94.4 ± 2.1 (noisy inputs)	NA
Location	15.8 ± 5.9 (noisy inputs)	NA	NA	NA	98.9 ± 0.9 (noisy inputs)

Table 8. *Comparisons of testing accuracies between different linear decoders. The first two sections examine whether there is information about space in the identity network and why there is information about space in the identity network. The next section examines whether there is information about identity and what kind of identity information is in the space network.*

Comparisons of testing accuracies between different linear decoders			
Network 1	Network 2	Average Difference in Accuracy (%) ($*$: $p < 0.05$, $**$: $p < 0.01$, $***$: $p < 0.001$)	p-value
$network_{(identity,space)}$	$network_{space\ baseline}$	73.3 $***$	< 0.001
$network_{(identity,space)}$	$network_{(space,space)}$	-20.0 $***$	< 0.001
$network_{(identity,space)}$	$network_{(shoes,space)}$	7.3 $**$	0.006
$network_{(location,identity)}$	$network_{(space,identity)}$	-1.8	0.47
$network_{(orientation,identity)}$	$network_{(space,identity)}$	-0.2	0.92
$network_{(location,identity)}$	$network_{(orientation,identity)}$	-1.6	0.52
$network_{(space,identity)}$	$network_{identity\ baseline}$	11.2 $***$	< 0.001
$network_{(space,identity)}$	$network_{(identity,identity)}$	-30.4 $***$	< 0.001
$network_{(space,shoes)}$	$network_{shoes\ baseline}$	5.4 $**$	0.003
$network_{(space,shoes)}$	$network_{(space,identity)}$	-4.5 $*$	0.02

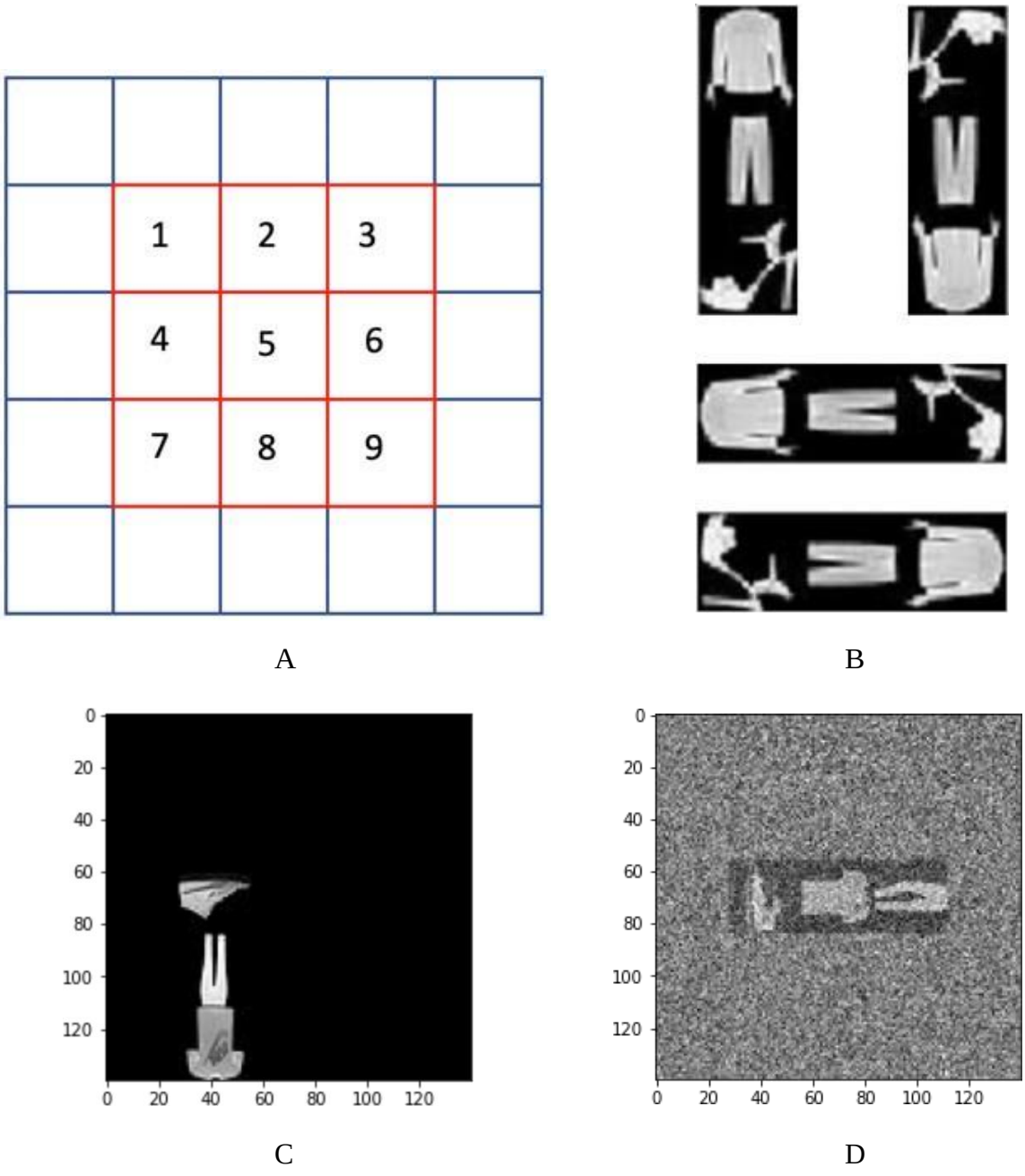

Figure 1. Object image locations and orientations. A. Nine possible locations of the center of an object image. B. Four possible orientations of an object image (up, down, left, and right orientations, respectively; going from top to bottom images and, for the first row, left to right images). Note that the alignment of parts within an image are not randomized, arealways in the same alignment, and always constrained to the two directions along the long axis. C. An example of an unscrambled (US) image with "down" orientation at location 7. D. An example of a noisy scrambled (S) image with "right" orientation at location 5.

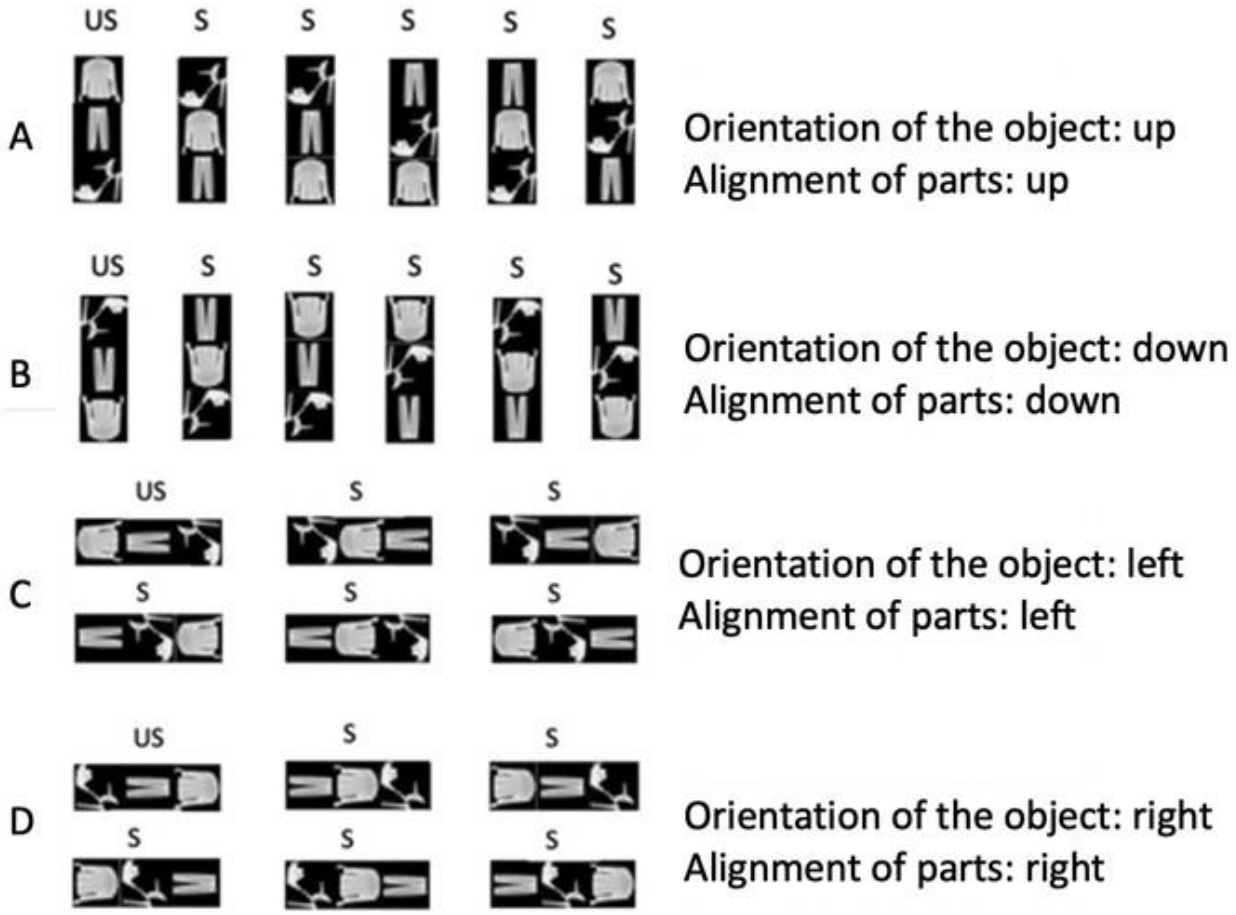

Figure 2. Unscrambled and scrambled object orders. The alignment of the parts within anobject and the orientation of the object are always the same. For each orientation, there are six possible orders of parts. Only the first image for each orientation (first image in each row) is considered as an unscrambled object image (labeled "US"). The other images for a given orientation are scrambled object images (labeled "S"). A. Up orientation. B. Down orientation. C. Left orientation. D. Right orientation.

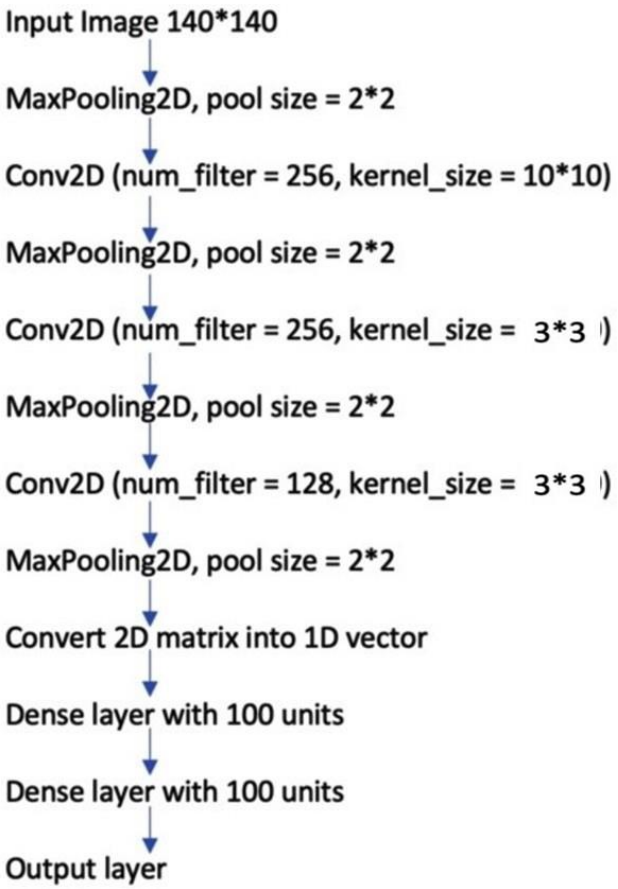

Figure 3. The structure of brain networks. Each neural network consists of several hidden layers, including the convolutional layer, the pooling layer, and the fully connected dense layer. The only difference between different brain networks is the size of their output layer.The size of the output layer depends on the task they were trained to do.

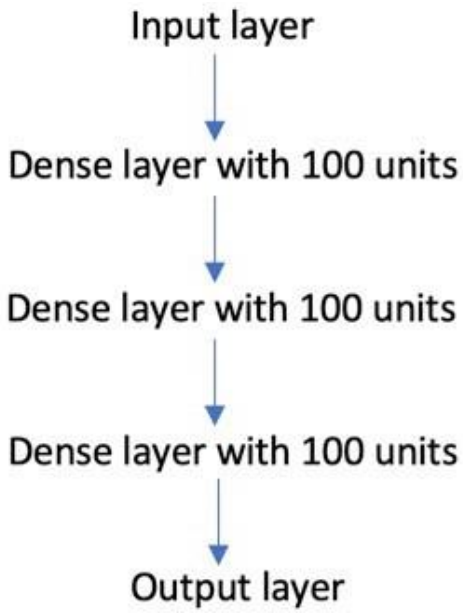

Figure 4. The structure of a decoder network. The input dimension is equal to the numberof units in the network layer that it was trained to decode from. The output dimension depends on what kind of information it was trained to decode.

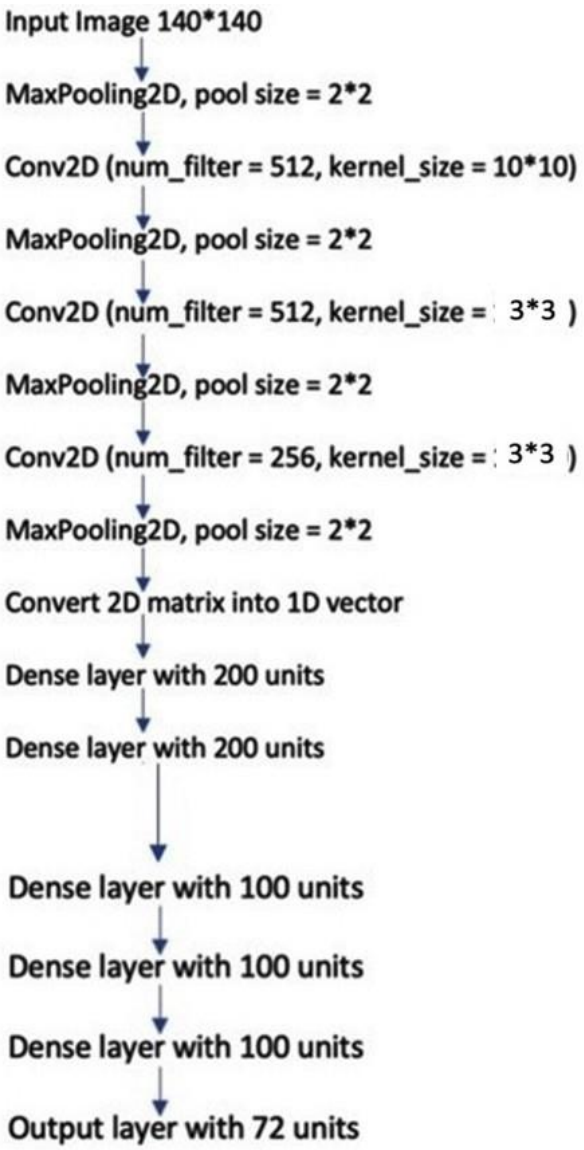

Figure 5. The structure of $network_{combine\ identity\ and\ space}$, the single network that takes the images as visual inputs and determines objects' identity and space information as 1 of the 72 possible combinations of identity (2 possible) and space (36 possible).

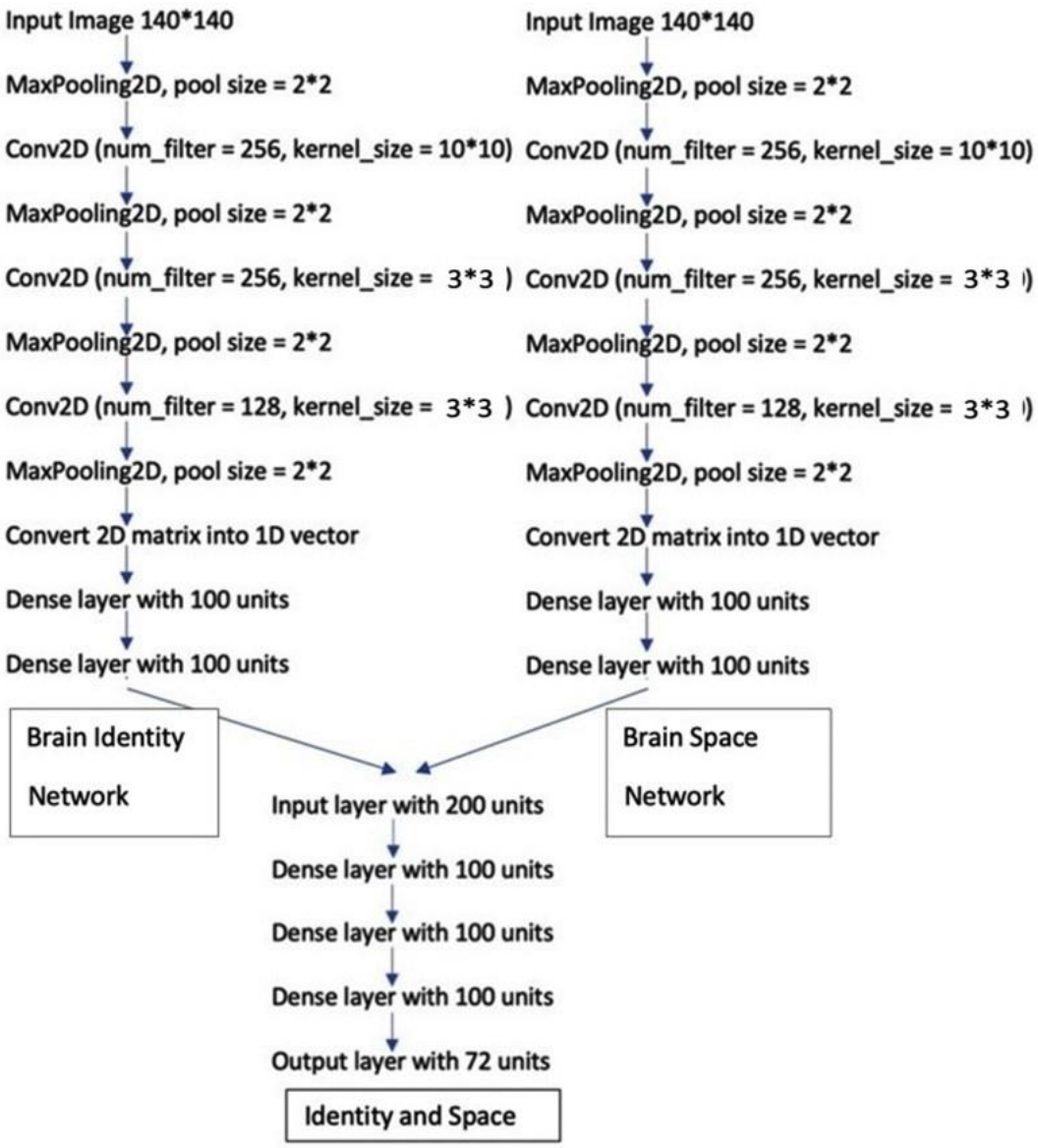

Figure 6. The structure of $network_{separate\ identity\ and\ space}$, the two brain networks that take the images as visual inputs. The brain identity network determines objects' identity and the brain space network determines space. Later, the results from the two networks are combined to determine objects' identity and space information as 1 of the 72 possible combinations of identity (2 possible) and space (36 possible).