Title
Generalization and Transfer Learning in Neural Networks Performing Shape, Size, and Color Classification

Permalink
https://escholarship.org/uc/item/1q2683zm

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44)

Authors
Huang, Aishi
Huebner, Philip A
Willits, Jon

Publication Date
2022

Peer reviewed

Generalization and Transfer Learning in Neural Networks Performing Shape, Size, and Color Classification

Aishi Huang (aishih2@illinois.edu),
Department of Psychology, University of Illinois
Champaign, IL 61820

Philip A. Huebner (huebner3@illinois.edu)
Department of Psychology, University of Illinois
Champaign, IL 61820

Jon A. Willits (jwillits@illinois.edu)
Department of Psychology, University of Illinois
Champaign, IL 61820

Abstract

We investigated neural networks' ability to generalize during visual object recognition. In three experiments, we show that while basic multilayer neural networks easily learn to classify the objects on which they are trained, they show serious difficulties transferring that knowledge to novel items. However, our experiments also show that when the previously trained networks are then trained on the novel items, they learn to respond correctly to the novel items much faster than untrained networks. This shows that these networks are learning abstract representations that go beyond the simple items on which they were trained. We argue that this demonstrates that regarding abstract rule learning, the problem with neural networks is not their inability to learn abstractions, but their ability to apply that knowledge when classifying new objects.

Keywords: neural networks, transfer learning, object recognition, knowledge representation

Introduction

Excitement about artificial neural networks, both as a theory of cognition, and as a form of artificial intelligence, has waxed and waned over time. Recently, new architectures, training algorithms, and increased computational power have led to neural networks that are increasingly successful at a wide range of tasks (Krizhevsky et al., 2012, Devlin et al., 2018). However, behavioral research has shown that human behavior is sometimes qualitatively different than the performance of state-of-the-art deep learning models (Geirhos et al., 2018, Linzen et al., 2016), and questions still exist about the fundamental representational capabilities of these models (Marcus, 1998; Martin & Doumas, 2020).

For example, Marcus argued that neural networks are incapable of generalizing "outside of their training set", in a way that seems so effortless for humans. Marcus cites many examples, but perhaps the most demonstrative is a model based on 7-month-old infants in a rule-learning experiment (Marcus et al., 1999). In the experiment, infants heard sequences of phonemes that followed one of three rule-based structures: ABA sequences (where the first and third syllable were the same), ABB sequences (where the second and third were the same), and AAB sequences (where the first and second were the same). In this experiment, infants who heard sequences following one rule for two minutes showed subsequent discrimination between sequences that did and did not follow the rule, even when the novel sequences used a completely novel set of phonemes. Marcus et al. concluded that infants learn generalizable rules independent of the items that they use as the basis of induction for those rules.

Marcus used a simple recurrent network (SRN, Elman, 1990) to simulate this experiment, demonstrating that the network did not show similar generalization. In Marcus's model, each phoneme was represented as both an input and output unit, and the model was trained to predict the sequences of phonemes, where success was defined as learning which of the three rules (ABB/AAB/ABA) the sequences followed. Marcus showed that while SRNs learned to predict the sequences for phonemes occurring in the training sequence, they failed to show generalization to phonemes that didn't occur during training. Marcus argued that the only thing the neural network could do was learn associations between specific syllables. Abstract rules exist at the level of relations between variables, and neural networks cannot represent variables in this manner.

In contrast, Willits (2013) argued that ABA-style transition rules were encoded in the SRN's recurrent weights but could not be used during testing because the network was inhibited against ever predicting the novel sounds – a sensible outcome given that these sounds never occurred during training. Analyses of the network's representations demonstrated that alternation and/or repetition rules were encoded in the network's recurrent weights, but that strong inhibitory weights to the novel output units prevented this from being expressed. Willits further showed that if the network was given just a little bit of training on the new test items, it learned rule-consistent test items much more quickly than rule-inconsistent items. Willits argued that the network had learned that there were items belonging to an 'A' category, and items belonging to a 'B' category, and had learned the sequential relationships of A and B items. In the transfer learning scenario, if the new sequence was rule consistent, all the network needed to learn was to categorize

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society*.

the novel items as either an A or B, and the sequences would be predicted correctly. In contrast, to learn rule-inconsistent items the network needed to re-learn input, output, and new recurrent weights reflecting the new sequential rule.

Willits's demonstration is an example of what has since become a very popular neural network training technique: transfer learning (Pan & Yang, 2009). In recent years, deep learning systems are rarely trained "from scratch". Instead, an existing network that already performs well at one task is used as a base model and trained to do a second task. If the source and target task share underlying structure, this speeds the model's learning of the target task. Willits (2013) showed this principle applies to abstract relational rules (like ABA sequence rules) applying to entirely novel items participating in those sequences. This demonstration suggests that the criticism of a neural network's ability to show rule-based transfer is not a problem with learning or representing abstract rules, but instead with applying a rule to novel input.

The goal of this paper is to show that this principle can be used to better understand the representational capabilities and limitations of neural networks in a much wider range of situations, particularly visual object recognition. We used a simple artificial visual world (Polyomino World) to carefully control the structure and statistics of visual images. Using this framework, we then investigated the basic capabilities of simple artificial neural networks regarding their ability to successfully classify the objects in these visual images. Next, we employed transfer learning (providing additional training for new items to a previously trained network), investigating the ways in which pre-trained networks showed facilitated learning for novel items. If pre-trained networks show faster learning than untrained networks, then this shows the network did have useful representations that could have been applied to the novel object, and that immediate failure to show transfer was a problem with associating that representation with the novel item, not with learning the representation in the first place.

Polyomino World

In the current study, we wanted a perfectly controlled dataset to simplify analysis and understanding of the neural network's behavior. To this end, we created the artificial world "Polyomino World". Polyomino World is a world consisting of scenes of objects, objects defined in terms the set of pixels the objects occupy. In this paper, we created a very simple set of scenes consisting of an 8x8 grid of grey pixels (the background), with each scene containing a single object. In the following experiments, we defined all our objects to be simple polyominoes (plane geometric figures formed by joining one or more equal squares edge to edge) of size 1 (monomino), size 2 (domino), size 3 (tromino), and size 4 (tetromino), with examples shown in Figure 1.

There are certain properties of the polyominoes worth noting. Considering all polyominoes of size 1-4, there are 9 distinct orientation-independent shape types, shown in Figure 2. Some of the shapes can be placed in different orientations.

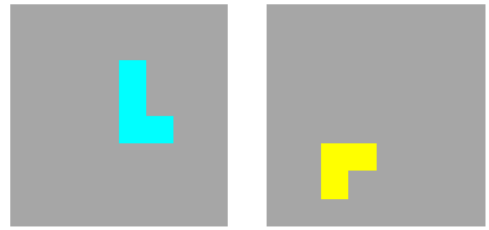

Figure 1. Example images from Polyomino World, with one colored shape on an 8x8 grid with a grey background.

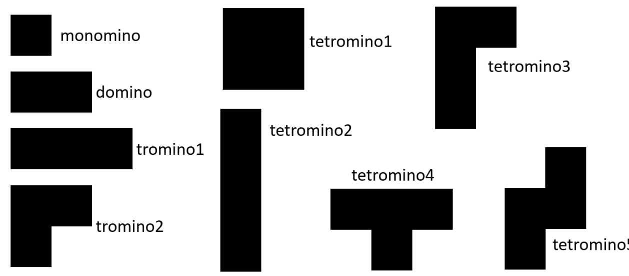

Figure 2. The full set of nine polyomino shapes.

A domino and tetromino2 can be either horizontal or vertical. A tromino2 and tetromino4 can be placed in four different orientations, and a tetromino3 and tetromino5 can be placed in eight different orientations. This allows us to test a neural network's ability to learning representations of the objects independent of their orientation.

In Polyomino World we can specify the color of each pixel of each object. In this paper, all the pixels of each object were the same color, and came from the set of 8 colors making up the corners of the RGB cube: red (+1,-1,-1), blue (-1,+1,-1), green (-1,+1,-1), cyan (-1,+1,+1), magenta (+1,-1,+1), yellow (+1,+1,-1), white (+1,+1,+1), black (-1,-1,-1). Each shape occurred against a grey background, the center of the RGB cube (0,0,0). Thus, each input scene was represented as a 192-element vector (8x8x3) containing 0's where each cell was empty/grey, and the color-appropriate RGB-values of +1 and -1 where a colored shape was present.

Network Architecture and Training

The main goal of this paper was to obtain a better understanding of how neural networks work, and so the architecture used was a simple, fully interconnected network with a single hidden layer, as shown in Figure 3. All models were initialized with weights chosen from a random uniform distribution ranging from -0.01 to 0.01.

Each network was trained in the following manner. First, it was presented with a set of input vectors representing each image in the training set. For each input, activation was propagated along the weighted connections to the hidden layer, put through a sigmoid activation function, and then propagated along the weighted connections to the output layer, also put through a sigmoid activation function.

These output activations were compared to the correct output activation, which constituted three 1's (one each for the correct color, shape, and size labels), and a zero for all other units. The binary cross entropy was used as an error

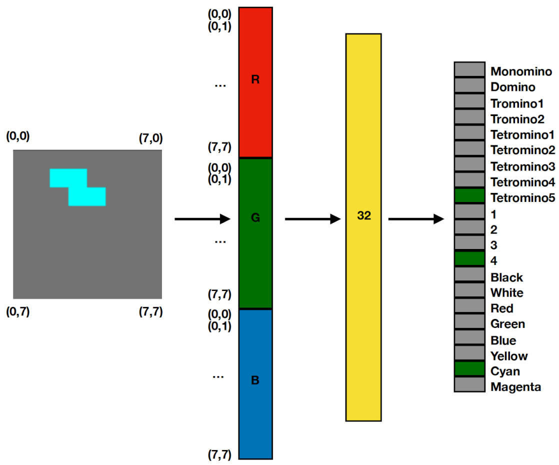

Figure 3. Network architecture for all experiments: an input layer with 192 units (8x8 units for each R, B, and G color pixel); a 32-unit hidden layer; and an output layer with 22 units, one unit for each of the 9 shape labels, 4 sizes, and 8 colors. The model was trained to simultaneously classify all three features (label, color, and size).

function that was used to train the weights using stochastic gradient descent with a learning rate of 0.40.

Experiment 1: Novel Positions

Experiment 1 had two goals. First, to test the neural network's ability to show immediate transfer to objects occurring in previously unexperienced portions of the visual array. We tested this by training a model to classify the shape, size, and color of objects that were presented in only half of the visual array (either top half or bottom half) and tested the model on the same shapes presented in the omitted half. Second, we tested how this pre-training affected the learning of objects appearing in the previously omitted half.

This experiment acts as a visual replication of the previously described models of Marcus et al.'s ABA experiment. To succeed, the model needs to learn to represent each shape (a set of pixels that are the same color at the same time) and transfer that knowledge to pixels that have never been any color but grey (and thus always had values of zero in the input). As in the ABA task, the network should fail to show immediate transfer. The model will never see any input but zero in the omitted space. Thus, weights connected to those inputs will never contribute to prediction error and thus will never be adjusted by the error-driven learning algorithm. When the model is tested with input to this region, the result should be a hidden layer activation that is effectively a random vector, with poor transfer performance. However, we also predict benefits from this prior training in the transfer learning phase. If the model has learned representations for the particular shapes, sizes, and colors that appeared in half of the visual array, then when it is trained on the other half, all it should need to do is adjust the input weights so those inputs instantiate the same representations, and learning to do so should be much faster than learning to correctly classify the same inputs without having already learned to classify those shapes in the other part of the visual array.

Method

Stimuli. We created three datasets: "Full", "Omit top", and "Omit bottom". For the "Full" dataset, we created a training set of images each containing only a single object. An image was created for all possible orientations of each shape, in each color, in all possible positions in the 8x8 grid. The total number of scenes was 11,496 (the number of colors, times the number of rotational and flipping variant for each shape, times the number of legal positions of each shape). For the other two datasets, we restricted the legal positions to those where the shape was entirely in the bottom ("Omit top") or entirely in the top ("Omit bottom") of the 8x8 grid.

Procedure. Using the general training procedure described above, to test the model's immediate transfer we trained five models each on the "Omit top" and "Omit bottom" datasets. These models were trained for 2,000,000 trials, and then tested on the items from the omitted half of the visual array. To test transfer learning, we took these 10 pre-trained models and trained them for an additional 2,000,000 trials on the "Full" dataset. We then compared the performance of these models to 10 Control models that were trained for 2,000,000 trials on "Full" dataset but without pre-training. Critically, both the Pre-trained and Control models saw the critical items (the items from either the top or the bottom of the visual array) the same number of times.

Results and Discussion

Each network was evaluated at every 50,000 training steps (i.e., after seeing 50,000 objects and performing 50,000 weight updates). The networks were evaluated for their ability to correctly classify each object's shape label, size, and color. For each of these features, the network's guess at the correct choice was made by choosing the item from each set with the highest activation level, and if that was the correct feature, the guess was evaluated as correct.

On the trained items (items appearing only in half of the visual array), the models achieved perfect classification accuracy on all three features: for color after 50k training instances, for size after 200k training instances, and for shape label after 500k training instances.

In the test of immediate transfer, as predicted the models failed to show transfer. They learned to correctly classify items in the trained half of the visual array, but utterly failed to show any kind of transfer to the same objects presented in the omitted half of the visual array. Classification accuracy was 100% for objects in the trained part of the visual array but was at chance for all three features (shape, size, and color) for objects in the untrained part of the visual array.

Also as predicted, in the test of transfer learning facilitation the pre-trained models had significantly faster learning on objects presented in the previously omitted half of the visual field, compared to models without pretraining. This facilitation effect was present for learning to correctly classify the object's shape label, as shown in Figure 4. Some

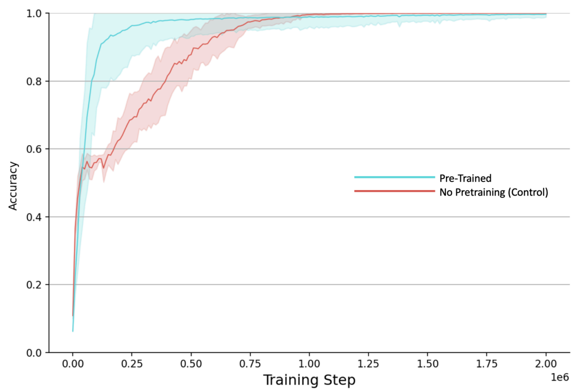

Figure 4. Accuracy at correctly activating the correct shape labels as a function of amount of training. The "Pre-Trained" curve shows the learning trajectory for items presented in the portion of the visual array omitted during pre-training. The curve for "No Pretraining" curve shows the learning trajectory for the same set of items in a model starting with randomly initialized weights.

pre-trained models (as evidenced by the error bars shown as the shaded portion of the figure) reached perfect shape label classification performance within 100,000 trials (seeing each color of each shape in each position about 10 times). In contrast, models without pre-training took at least 600,000 trials to reach the same level of performance. There was also significant facilitation in reaching perfect performance on size classification (40,000 trials for pre-trained models vs. 60,000 trials for non-pretrained models). There was no significant difference for learning to classify color, as both models achieved perfect performance within 20,000 trials.

The failure of immediate transfer demonstrates what has been argued by Marcus: that neural networks do have a serious problem when it comes to generalizing knowledge. Marcus and others have argued that failures of this sort demonstrate that neural networks are not representing information in a symbolic fashion, as humans do. However, the transfer learning results suggest the problem is a different one. Critics argue that neural networks learn simple input-output mappings and do not represent abstract structure that binds together items of the same shape. But if this were true, then the models that were pre-trained on half of the visual array, and then further trained on the other half, would have seen no learning advantage compared to a model starting from scratch. But this was not the case. This facilitated learning strongly suggests that there was, for example, a learned representation for "domino", and that new instances of dominoes could, with further training, be quickly associated with that representation. The problem with the neural network is not a learning or representation problem; the problem is providing a mechanism to automatically bind new instances to those representations.

Experiment 2: Novel Shapes

One potential criticism of Experiment 1 is that it is an unrealistic model of the situations that humans encounter. There is no point in development where half of the visual field is not exposed to stimuli. Experiment 2 explores a situation much more naturalistic, experiencing novel objects.

In Experiment 2, the network should "learn to fail" to correctly classify the label of the unseen object. In the pre-training phase, the network will be trained to classify eight of the objects, and then tested on the untrained object. Prior to any training, the network's randomly initialized weights will lead to it predicting that all nine shape objects are equally likely to be the label for the object. However, as the model continues to experience all objects but one, it will adjust its weights to never predict the unseen item. Thus, failure to show immediate transfer to the unseen object's label is a perfectly sensible thing for the network to learn to do.

Of more interest are two things. First, as the network is learning about the other eight objects, will it continue to be able to correctly classify the unseen object's color and size? One criticism of neural networks is that, due to their distributed representations, they are not able to learn separable features of objects. However, if the network shows immediate transfer for color and size of the unseen object, this will provide evidence that the network can learn these properties and transfer them to new, unseen objects. Second, how will the network respond when, as in Experiment 1, it is given a second round of "transfer learning" and given the opportunity to learn about the object that had previously been omitted. If the pre-trained model that had learned about the other eight objects can learn to label the ninth object more quickly than a model without the pre-training, this will show that the network learned features that help it to classify, not just things it has seen, but as-of-yet unseen objects, evidence the network is learning some kind of abstract representation of shape that can serve as a basis of generalization.

Method

Stimuli. Experiment 2 had two classes of datasets: 1) "Full", which like in Experiment 1, contained each object in all possible colors, orientations, and legal positions in the 8x8 grid, and 2) "Omit a Shape", where all examples of one of the nine different polyomino shapes was removed from the dataset. There were nine different versions of this dataset; in each, one of the different shapes was omitted.

Procedure. The procedure in this experiment was much like in Experiment 1. Using the general training procedure described above, in the "Immediate Transfer" condition we trained five randomly initialized models on each of the nine different "Omit a Shape" datasets. These models were trained for 2,000,000 trials, and then tested on their classification accuracy when presented with the shape that was omitted during training. The models were evaluated in the same way as in Experiment 1. To test transfer learning, we took these pre-trained models and trained them for an additional 2,000,000 trials on the "Full" dataset. We then compared the performance of these models to 10 control models that were trained for 2,000,000 trials on "Full" dataset without pre-training. Critically, both the pre-trained and control models saw the critical items (the shape omitted during pre-Training for the pre-trained models) the same number of times.

Results and Discussion

On trained items, the models had perfect performance for color (after 50k training steps), for size (after 200k training steps), and for shape label (after 300k training steps). In the test of immediate transfer, as predicted the models "learned to fail", reaching 0% accuracy for choosing the correct label for the untrained shape within 50k trials. In contrast, they showed perfect accuracy classifying the color and size of the untrained object, with accuracy levels for these features that were not significantly different than accuracy for these features on trained objects. In the test of transfer learning facilitation, the pre-trained models once again outperformed the models without pre-training, as shown in Figure 5.

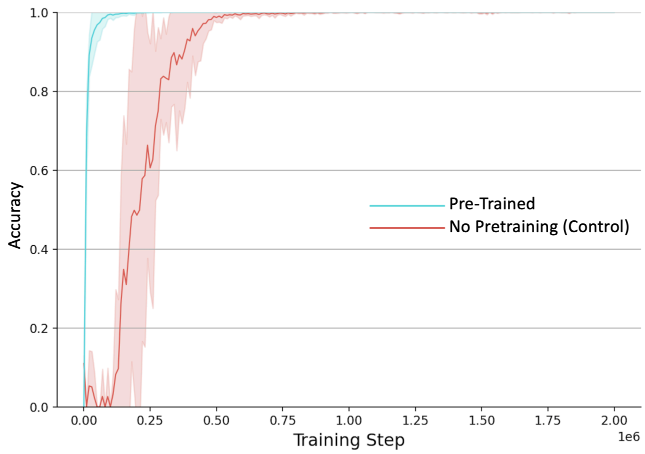

Figure 5. Shape label learning trajectory for a previously untrained shape (tromino1), on a model that was pre-trained on all other shapes, compared to a model with no pre-training.

Prior experience learning to classify the other eight shapes significantly speeded the learning of the new shape. The model without pretraining took, on average, 200k training steps to reach perfect shape label classification accuracy on a shape, whereas the pretrained models took, on average, 500k training steps. In addition to being significantly faster overall, this difference was significant for each individual shape except monominoes (which were at ceiling). The average time for the models (pre-trained and not) to reach 100% accuracy on each of the nine shapes is shown in Table 1.

Table 1. Mean training trials required to reach 100% shape label classification, as a function of whether the model was pre-trained on the other 8 shapes. All numbers are thousands of training steps.

Shape	M	D	Tr1	Tr2	Te1	Te2	Te3	Te4	Te5
Pre-trained	50	50	150	100	800	200	350	300	100
Not Pre-trained	50	100	500	400	1350	750	550	650	300

The results from Experiment 2 replicate Experiment 1. The faster learning of the previously unseen shape by the pre-trained models demonstrates that the weights being learned by the model help it instantiate distributed representations composed of features that are useful when they need to be applied to a new object. When the model is given additional training on the new object, on many objects all it needs to do is learn to adjust its output weights to correctly assign an output label to that new shape's hidden representation. This is much faster than learning an entirely new representation from scratch. In addition, the demonstration that the model continues to correctly classify unseen objects' size and color, while not knowing how to classify the object's name, is evidence the model is learning representations of the different features (shape, size, and color) that are at least somewhat independent, a necessary pre-requisite for more advanced reasoning using these features.

Experiment 3: Novel Orientations

One could argue that the models in Experiment 2 suffer from the same problem as the models in Experiment 1. Both models, like Marcus's ABA model, have either input units (in Experiment 1) or output units (in Experiment 2) that are never used during training. This property, it has been argued, makes them bad models of human learning, since humans don't come to learning situations with inputs or outputs that have never occurred in previous training, and thus have been explicitly trained expect null input (Seidenberg & Elman, 1999). Experiment 3 models a third situation of transfer and transfer learning that completely avoids this problem: the classification of previously experienced shapes that are occur in unique orientations (either rotated or flipped). In this experiment, all shapes are experienced during training, in all possible positions in the space. However, during training the shapes will only have occurred in half of their possible orientations. For example, during training the model might see dominos placed only horizontally, or tetromino3's (the one that looks like an L), but only in four of the eight possible orientations that shape can occur (counting all four 90-degree rotations, placed both forwards and backwards. The model is then tested on the untrained orientations to see if it can make correct color, size, and shape label classifications.

Method

Stimuli. The items in Experiment 3 were limited to those that could appear in multiple orientations (i.e., monominoes and tetromino1s were not included). There were two classes of datasets: 1) "Full", which contained each of the 7 objects in all possible colors, orientations, and legal positions in the 8x8 grid, and 2) "Omit Half of Orientations", where half of the orientation variants of each shape were omitted. There were two different versions of this dataset; in each one a different half of the variants were omitted.

Procedure. In the "Immediate Transfer" condition we trained five models on each of the "Omit Half" datasets. These models were trained for 2,000,000 trials and then tested on the omitted half of the variants. For the "Pre-trained" condition, we took these 10 pre-trained models and trained them an additional 2,000,000 trials on the Control dataset (containing all the variants). We compared the performance of these models to 10 "Control" models, trained for 2,000,000 trials on Control dataset without pre-training. Critically, both the Pre-trained and Control models saw the critical items the same number of times.

Results and Discussion

On trained items, the models achieved perfect accuracy for color, size, and shape. In the test of immediate transfer (classification accuracy for shape, size, and color on the untrained variants of each shape), the models reached perfect accuracy for size and color, transferring knowledge of these features as quickly as they were learned on the training set. In contrast, shape accuracy plateaued at 73% after 700,000 training steps. This was considerably above chance (14.3%, or one out of seven possible shape labels that could be guessed), but still failing to show perfect transfer to unobserved items. This is the kind of failure to show transfer discussed by Marcus, and without the issue of relying on input or output units trained on null data. The models were learning to map each observed variant to a response, but whatever representation they were using to do so did not lead to correct classification of new, unseen versions of those shapes. These results are shown in Figure 6.

However, these results are also evidence against the notion the networks were only learning simple mappings between each independent variant of each shape and its output label. If that were the case, immediate transfer would have been at chance (17%), not 73%, as each shape variant's representation would have provided no help with any of the other unseen variants. This argument is strongly supported by the results of the transfer learning (shown in Figure 7), which show that with additional training on the previously unseen items, the pre-trained models achieved perfect accuracy much more quickly (after about 350,000 training steps) than the untrained models (1,000,000 training steps).

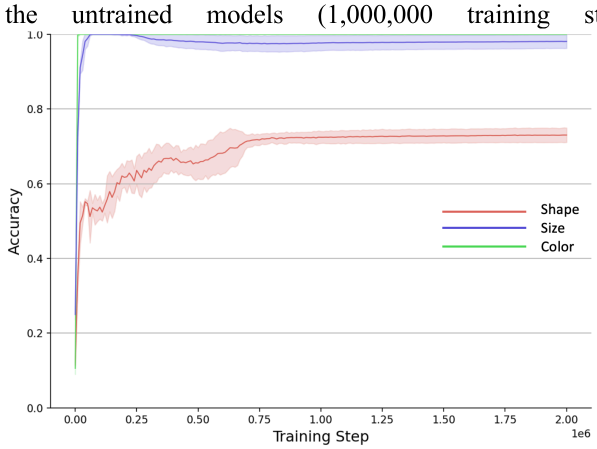

Figure 6. Immediate transfer for shape, size, and color for untrained variants (rotations and flips) of trained shapes.

General Discussion:

These results demonstrate that neural networks have a problem transferring knowledge to new situations, even in situations that don't rely on untrained input or output units. However, the fact that in all three experiments, the pre-trained models learned the new items faster that untrained models shows that the models are learning representations

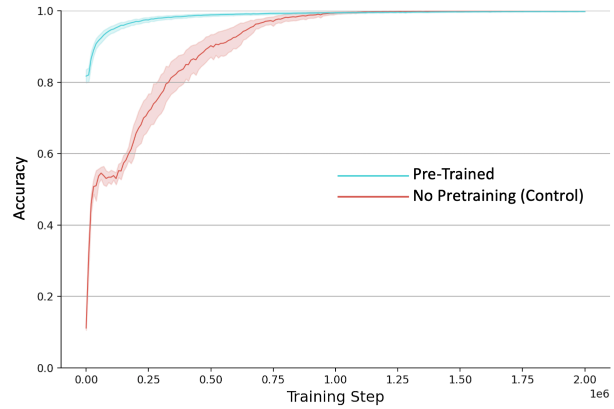

Figure 7. Shape label learning for a previously untrained variants of shapes (flips and rotations), on a model that was pre-trained on all other variants, compared to a model without pre-training.

that provide a basis for transfer to the new items. That transfer is just not being immediately applied to new items. This strongly suggests that the problem with the neural networks regarding learning abstract relational representations is not a problem with learning or representation, but a problem with application of those representations to new items.

Neural network enthusiasts should not, however, get too excited about these results, as they demonstrate several very serious difficulties that would need to be overcome for the models to show human like performance. First, even though the networks do show transfer benefits from previous knowledge, the networks still require considerable additional training even with this prior learning, for example of 300,000 training steps in Experiment 3. That is equivalent to having experienced each new variant over 10,000 times (collapsing over color and position). Human learners take less than that. What conclusion should we draw from this? We argue that successful transfer learning is evidence that there is representation that is useful as a basis of transfer, but that transfer learning (or anything like it) is an unlikely candidate mechanism for that transfer to occur. What other options are there, within a neural network framework?

Of course, before we can understand whether the representations being learned by neural networks can truly support generalization, we need to understand more about exactly how the neural networks are representing what they are learning. Truly representing a shape in a way that leads to successful transfer can be done, and maybe must be done by, representing a relational rule. Can such a rule be represented in a neural network? In principle, yes. But is that how these neural networks are learning to represent those shapes? If not, can they be convinced to do so? Before deciding whether neural networks are fundamentally incapable of learning, representing, and using information like humans, we have a lot to learn about how they work.

References

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179-211.

Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A., & Brendel, W. (2018). ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. arXiv preprint arXiv:1811.12231.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*, 1097-1105.

Linzen, T., Dupoux, E., & Goldberg, Y. (2016). Assessing the ability of LSTMs to learn syntax-sensitive dependencies. *Transactions of the Association for Computational Linguistics*, *4*, 521-535.

Marcus, G. F. (1998). Rethinking eliminative connectionism. *Cognitive Psychology*, *37*, 243-282.

Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, *283*, 77-80.

Martin, A. E., & Doumas, L. A. (2020). Tensors and compositionality in neural systems. *Philosophical Transactions of the Royal Society B*, *375*, 20190306.

Seidenberg, M. S., & Elman, J. L. (1999). Do infants learn grammar with algebra or statistics?. *Science*, 284, 433-433.

Willits, J. (2013). Learning nonadjacent dependencies in thought, language, and action: Not so hard after all. *Proceedings of the Annual Meeting of the Cognitive Science Society.*