Combinets: Creativity via Recombination of Neural Networks

Matthew Guzdial and Mark Riedl
School of Interactive Computing
Georgia Institute of Technology
Atlanta, GA 30332 USA
mguzdial3@gatech.edu and riedl@cc.gatech.edu

Abstract

One of the defining characteristics of human creativity is the ability to make conceptual leaps, creating something surprising from existing knowledge. In comparison, deep neural networks often struggle to handle cases outside of their training data, which is especially problematic for problems with limited training data. Approaches exist to transfer knowledge from models trained on one problem with sufficient data to new problems with insufficient data, but they tend to require additional training or a domain-specific method of transfer. We present conceptual expansion, a general approach for reusing existing trained models to derive new models without backpropagation. We evaluate our approach on few-shot variations of two tasks: image classification and image generation, and outperform standard transfer learning approaches.

Introduction

Modern deep learning systems perform well with large amounts of training data on known classes but often struggle otherwise. This is a general issue given the invention or discovery of novel classes, rare or illusive classes, or the imagining of fantastical classes. For example, if a new traffic sign were invented tomorrow it would have a severe, negative impact on autonomous driving efforts until enough training examples were collected.

Deep learning success has depended more on the size of datasets than on the strength of algorithms (Pereira, Norvig, and Halevy 2009). A significant amount of training data for many classes exists. But there are also many novel, rare, or fantastical classes with insufficient data that can be understood as derivations or combinations of existing classes. For example, consider a pegasus, a fantastical creature that appears to be a horse with wings, and therefore can be thought of as a combination of a horse and a bird. If we suddenly discovered a pegasus and only had a few pictures, we couldn't train a typical neural network classifier to recognize a pegasus as a new class nor a generative adversarial network to create new pegasus images. However we might be able to approximate both models given existing models trained on horse and bird data.

Various approaches exist that reuse knowledge from models trained on large datasets for a particular problem to try to solve problems with smaller datasets, such as zero-shot and transfer learning. In these approaches, knowledge from a source model trained is applied to a target problem by either retraining the network on the target dataset (Levy and Markovitch 2012) or leveraging sufficiently general or authored features to represent new classes (Xian, Schiele, and Akata 2017). The latter of these two approaches is not guaranteed to perform well depending on source and target problems, and the former of these is limited in terms of what final target models can be learned.

Combinational creativity is the type of creativity humans employ when combining existing knowledge to create something new (Boden 2004). Many algorithms exist that attempt to reflect this process, but they have historically required hand-authored graphical representations of input concepts with combination only occurring across symbolic values (Fauconnier 2001). A neural network is a large, complex graph of numeric values derived from data. If combinational creativity techniques could be applied to recombine trained neural networks, this could allow us to address the novel class problem (e.g. pegasus) without the introduction of outside knowledge or heuristics.

We introduce a novel approach, *conceptual expansion*, that allows for the recombination of an arbitrary number of learned models into a final model without additional training. In the domains of image recognition and image generation we demonstrate how recombination via conceptual expansion outperforms standard transfer learning approaches for fixed neural network architectures. The remainder of the paper is organized as follows: first we discuss related work and differentiate this technique from similar approaches for few-shot problems. Second, we discuss conceptual expansions in detail and the search-based approach we employ to construct them in this paper. Third, we present a variety of experiments to demonstrate the limitations and advantages of the approach.

Related Work

Computational Creativity

Combinational creativity represents both a type of creativity and class of algorithm for knowledge reuse through recombining existing knowledge and concepts for the purposes of inventing novel concepts (Boden 2004). There have many

prior combinational creativity algorithms. Case-based reasoning (CBR) represents a general AI problem solving approach that relies on the storage, retrieval, and adaptation of existing solutions (De Mantaras et al. 2005). The adaptation function has lead to a large class of combinational creativity algorithms (Wilke and Bergmann 1998; Fox and Clarke 2009; Manzano, Ontañón, and Plaza 2011). These tend to be domain-dependent, for example for the problem of text generation or tool creation (Hervás and Gervás 2006; Sizov, Öztürk, and Aamodt 2015).

The area of belief revision, modeling how beliefs change, includes a function to merge prior existing beliefs with new beliefs (Cojan and Lieber 2009; Konieczny and Pérez 2011; Fox and Clarke 2009). Amalgams are an extension of this belief merging process that looks to output the simplest combination (Ontañón and Plaza 2010). The mathematical notion of convolution has been applied to blend weights between two neural nets in work that parallels our desire to combine combinational creativity and machine learning, but with inconclusive results (Thagard and Stewart 2011).

Conceptual blending is perhaps the most popular combinational creativity technique, though it has traditionally been limited to hand-authored input (Fauconnier 2001). Li et al. (2012) introduced goals to conceptual blending, which parallels our usage of training data to optimize the structure of a combination. Conceptual blending has further traditionally relied on symbolic values, which makes it ill-suited to statistical machine-learning. Visual blending (Cunha et al. 2017), combines pieces of images using conceptual blending and parallels our use of combinational creativity with Generative Adversarial Networks, however it requires hand-defined components and combines images instead of models. Guzdial and Riedl (2016) utilized conceptual blending to recombine machine-learned models of video game level design by treating all numbers as ordinal values, but their approach does not generalize to neural networks.

Combinational creativity algorithms tend to have many possible valid outputs. This is typically viewed as undesirable, with general heuristics or constraints designed to pick a single correct combination from this set (Fauconnier 2001; Ontañón and Plaza 2010). This limits the potential output of these approaches, we instead employ a domain-specific heuristic criterion to explore the space of possible combinations for an optimal one.

In Boden's model of three types of creativity (Boden 2009), we can consider our approach to combine elements of combinational and exploratory creativity. Conceptual expansion is a combinational creativity algorithm as, given a set of existing knowledge, it defines a space of possible valid combinations. We then employ a search process, which we call conceptual expansion search, to explore this space for particular combinations that meet some goal or heuristic.

Knowledge Reuse in Neural Networks

A wide range of prior approaches exist for the reuse or transfer of knowledge in neural networks, such as zero-shot, one-shot, and few-shot learning (Xian, Schiele, and Akata 2017; Fei-Fei, Fergus, and Perona 2006), domain adaptation (Daumé III 2009), and transfer learning (Lampert, Nickisch, and Harmeling 2009; Wang and Hebert 2016). These approaches require an additional set of features for transfer, or depend upon backpropagation to refine learned features from some source domain to a target domain. In the former case these additional transfer features can be hand-authored (Lampert, Nickisch, and Harmeling 2009; Kulis, Saenko, and Darrell 2011; Ganin et al. 2016) or learned (Norouzi et al. 2013; Mensink, Gavves, and Snoek 2014; Ba et al. 2015; Elhoseiny et al. 2017). In the case requiring additional training these approaches can freeze all weights of a network aside from a final classification layer or can tune all the weights of the network with standard training approaches (Wong and Gales 2016; Li et al. 2017). As an alternative one can author an explicit model of transfer such as metaphors (Levy and Markovitch 2012) or hypotheses (Kuzborskij and Orabona 2013).

Kuzborskij et al. (2013) investigate the same n to n+1 multiclass transfer learning problem as our image classification experiments, and make use of a combination of existing trained classifiers. However, their approach makes use of Support Vector Machines with a small feature-set and only allows for linear combinations. Rebuffi et al. (2017) extended this work to convolutional neural nets, but still requires retraining via backpropagation. Chao et al. (2016) demonstrated that average visual features can be used for zero-shot learning, which represents a domain independent zero-shot learning measure that does not require human authoring or additional training. We employ this last approach as a baseline.

One alternative to reusing learned knowledge in neural networks, is to extend a dataset to new classes using query expansions on the web (Yao et al. 2017) . However, we are interested primarily in the question of how existing learned features can be applied to problems in which no additional training data exists, even online, due to the class in question being new, fantastical, or rare. Similarly, Neuroevolution is an approach to train neural networks via evolutionary search, which includes an explicit recombination step (Floreano, Dürr, and Mattiussi 2008). However, this approach does not transfer knowledge from one domain to another.

Conceptual Expansion

Imagine tomorrow we discover that a pegasus exists. Initially we lack enough images of this newly discovered flying horse to build a traditional classifier or image generator. However, suppose we have neural network classifiers and generators trained on classes including horses and birds. Conceptual expansion, a combinational creativity algorithm, allows us to reuse the learned features from machine learned model(s) to produce new models without additional training or additional transfer features.

The intuition behind conceptual expansion is that it defines a high-dimensional, parameterized search space from an arbitrary number of pretrained input models, where each point of this search space is a new model that can be understood as a combination of existing models. We can then explore this space to find models that better meet a goal or heuristic. Each point of this space—each combined model—is a valid conceptual expansion. We can consider

the case where a class or concept (c_X) is a combination of other classes ($c1, ...c_n$) and that the learned features of models of classes $c_1, ...c_n$ can be recombined to create the features of a model of c_X. In these cases, we hypothesize that conceptual expansions can represent models one cannot necessarily discover using conventional machine learning techniques with the available data. Furthermore, we hypothesize that these conceptual expansion models may perform better on specific tasks than standard models in cases with small amounts of available data, such as identifying or generating new classes of objects. In prior work (2018), we demonstrated the application of conceptual expansion to non-neural graphs with hundreds of thousands of edges, which suggests the potential for their application to neural networks. We can use a heuristic informed by this small amount of training data to guide the search for our final conceptual expansion. This process is inspired by the human ability to make conceptual leaps, but is not intended as an accurate recreation.

A conceptual expansion of concept X is represented as the following function:

$$\mathrm{CE}^{\mathrm{X}}(F, A) = a_1 * f_1 + a_2 * f_2 ... a_n * f_n \qquad (1)$$

Where $F = \{f_1, ...f_n\}$ is the set of all mapped features and $A = \{a_1, ...a_n\}$ is a set of filters each dictating what of and what amount of mapped feature f_i should be represented in the final conceptual expansion. In the ideal case $X = CE^X$ (e.g. a combined model of birds and horses equals our ideal pegasus model). The exact shape of a_i depends upon the feature representation. If features are symbolic, a_i can have values of either 0 or 1 (including the mapped feature or not), or vary from 0 to 1 if features are numeric or ordinal. Note that for numeric values one may choose a different range (e.g. -1 to 1) dependent on the domain. If features are matrices, as in a neural net, each a_i is also a matrix. In the case of matrices the multiplication is an element-wise multiplication or Hadamard product. As an example, in the case of neural image recognition, $\{f_1, ..., f_n\}$ are the variables in a convolutional neural network learned via backpropagation. Deriving a conceptual expansion is the process of finding an A for known features F such that $\mathrm{CE}^{\mathrm{X}}(\cdot)$ optimizes a given objective or heuristic towards some target concept X.

In this representation, the space of conceptual expansions is a multidimensional, parameterized search space over possible combinations of our input models. There exists an infinite number of possible conceptual expansions for non-symbolic features, which makes naively deriving this representation ill-advised. Instead, as is typical in combinational creativity approaches, we first derive a *mapping*. The mapping determines what particular prior knowledge—in this case the weights and biases of a neural network—will be combined to address the novel case. This will determine the starting point of the later search process we employ to explore the space of possible conceptual expansions.

Given a mapping, we construct an initial conceptual expansion—a set of $F = \{f_1, ..., f_n\}$ and an $A = \{a_1, ..., a_n\}$—that is iterated upon to optimize for domain specific notions of quality (in the example pegasus case image recognition accuracy). In the following sections we discuss the creation of the mapping and then the refinement of the conceptual expansion.

Algorithm 1: Conceptual Expansion Search

input : available data $data$, an initial model $model$, a mapping m, and a score $score$
output: The maximum expansion found according to the heuristic

```
1  maxE ← DefaultExpansion(model)+m;
2  maxScore ← score;
3  v ← [maxE ];
4  improving ← 0;
5  while improving < 10 do
6      n ← maxE.GetNeighbor(v);
7      v ← v + n;
8      s ← Heuristic(n, data);
9      oldMax ← maxScore maxScore, maxE ←
         max([maxScore, maxE ], [s, n ]);
10     improving ←oldMax < maxScore?0:improving ++
11 return maxE;
```

Mapping Construction

Constructing the initial mapping is relatively straightforward for the purposes of this paper. As input we assume we have an existing trained model or models (CifarNet trained on CIFAR-10 for the purposes of this example (Krizhevsky and Hinton 2009)), and data for a novel class (whatever pegasus images we have). We construct a mapping with the novel class data by examining how the model or models in our knowledge base perform on the data for the novel class. The mapping is constructed according to the ratio of the new images classified into each of the old classes. For example, suppose we have a CifarNet trained on CIFAR-10 and we additionally have four pegasus images. Say CifarNet classifies two of the four pegasus images as a horse and two as a bird. We construct a mapping of: f_1 consisting of the weights and biases associated with the horse class, and f_2 consisting of the weights and biases associated with the bird class. We initialize the A values for both variables to all be 0.5—the classification ratio—meaning a floating point a value for the biases and an a matrix for the weights.

Conceptual Expansion Search

The space of potential conceptual expansions grows exponentially with the number of input features, and the mapping construction stage gives us an initial starting point in this space from which to search. We present the pseudocode for the Conceptual Expansion Search in Algorithm 1. Line 1 creates an initial expansion by combining a default expansion with the mapping information. The exact nature of this depends on the final network architecture. For example, the mapping may overwrite the entirety of the network if the input models and final model have the same architecture or just the final classification layer if not (as in the case of adding an additional class). In this case a default expansion is a conceptual expansion equivalent to the original model(s), in that each variable is replaced by an expanded variable with

its original f_i value and an a_i of 1.0 (or matrix of 1.0's). This means that the initial expansion is functionally identical to the original model, beyond any weights impacted by the mapping. This initial conceptual expansion derived at the end of the mapping construction will be a linear combination of the existing knowledge, but the final conceptual expansion need not be a linear combination.

Once we have a mapping we search for a set of F and A for which the conceptual expansion performs well on a domain-specific measure $Heuristic$ (e.g. pegasus classification accuracy). For the purposes of this paper we implement a greedy optimization search that checks a fixed number of neighbors before the search ends. The $GetNeighbor$ function randomly selects between one of the following: altering a single element of a single a_i, replacing all of the values of a single a_i replacing values of x_i with a randomly selected alternative x_j, or adding an additional x_i and corresponding random a_i to an expanded variable. The final output of this process is the maximum scoring conceptual expansion found during the search. For the purposes of clarity we refer to these conceptual expansions of neural networks as *combinets*.

CifarNet Experiments

In this section we present a series of experiments meant to demonstrate the strengths and limitations of conceptual expansions for image classification with deep neural networks. We chose CIFAR-10 and CIFAR-100 (Krizhevsky and Hinton 2009) as the domains for this approach as these represent well-understood datasets. It is not our goal to achieve state of the art on CIFAR-10 or CIFAR-100; we instead use these datasets to construct problems in which a system must identify images of a class not present in some initial training set given limited training data on the novel class. We then apply our approach to these problems, comparing them with appropriate baselines. For the source deep neural network model we chose CifarNet (Krizhevsky and Hinton 2009), again due to existing understanding of its performance on the more traditional applications of these datasets. We chose not to make use of a larger dataset like ImageNet or a larger architecture (Deng et al. 2009), as we aim to compare how our approach constructs new features given a limited set of input features. We do not include a full description of CifarNet but note that it is a two-layer convolutional neural net with three fully-connected layers.

For each experiment, we ran our conceptual expansion search algorithm ten times and took the most successful combinet found across the ten runs in terms of training accuracy. We did this to ensure we had found a near optimal conceptual expansion, but anticipate that future work will explore more sophisticated optimization strategies. We note that this approach was still many times faster than initially training the CifarNet on CIFAR-10 with backpropagation.

Our first experiment expands a CifarNet trained on CIFAR-10 to recognize one additional class selected from CIFAR-100 that is not in CIFAR-10. We vary the amount of training data for the newly introduced class. This allows us to evaluate the performance of recombination via conceptual expansions under a variety of controlled conditions. Our second experiment fully expands a CifarNet model trained on CIFAR-10 to recognize the one-hundred classes of CIFAR-100 with limited training data. Finally, we investigate the running example throughout this paper: expanding a CifarNet model trained on CIFAR-10 to classify pegasus images.

CIFAR-10 + Fox/Plain

For our initial experiment we chose to add fox and plain (as in a grassy field) recognition to the CifarNet, as these classes exist within CIFAR-100, but not within CIFAR-10 (CIFAR-10 is made up of the classes: airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck). We chose foxes and plains for this initial case study because they represented illustrative examples of conceptual expansion performance. There exists a number of classes in CIFAR-10 that can be argued to be similar to foxes, but no classes similar to plains.

For training data we drew from the 50,000 training examples for the ten classes of CIFAR-10, adding a varying number of training instance of fox or plain. For test data we made use of the full 10,000 CIFAR-10 test set and the 100 samples in the CIFAR-100 test set for each class. For each size slice of training data (i.e. 1, 5, 10, 50, and 100) we constructed five unique random slices. We chose five for consistency across all the differently sized slices, given that there was a maximum of 500 training images for fox and plain, and our largest slice size was 100. We present the average test accuracy across all approaches and with all sample sizes in Table 1. This table shows results when we provide five slices of fox or plain images in the quantities of 1, 5, 10, 50, or 100. For each slice, we provide the accuracy on the original CIFAR-10 images and the accuracy of identifying the 11th class (either fox or plains).

We evaluate against three baselines. Our first baseline (standard) trains CifarNet with backpropagation with stratified branches on the 10,000 CIFAR-10 images and newly introduced foxes or plains. This baseline makes the assumption that the new class was part of the same domain as the other classes as in (Daumé III 2009). For our second baseline we took inspiration from transfer learning and student-teacher models (Wong and Gales 2016; Li et al. 2017; Furlanello et al. 2017), and train an initial CifarNet on only the CIFAR-10 data and then retrain the classification layers to predict the eleventh class with the newly available data. We note that transfer learning typically involves training on a larger dataset, such as ImageNet, then retraining the final classification layer. However, we wished to compare how these different approaches alter the same initial features for classifying the new class. For our third baseline we drew on the zero-shot approach outlined in (Chao et al. 2016), using the average activation of the trained CifarNet classification layer on the training data to derive feature classification vectors. In all cases we trained the model until convergence.

There exist many other transfer approaches, but other approaches tend to require additional human authoring of transfer methods or features and/or an additional dataset to draw from. We focus on comparing the behavior of these approaches in terms of altering or leveraging learned features.

Table 1: A table with the average test accuracy for the first experiment. The orig. column displays the accuracy for the 10,000 test images for the original 10 classes of CIFAR-10. The 11th column displays the accuracy for the CIFAR-100 test images.

	100		50		10		5		1	
Fox	11th	orig.	11th	orig.	11th	orig.	11th	orig.	11th	orig.
combinet	**34.0±3.5**	81.8±2.2	**26.0±5.2**	81.59±1.9	**28.3±3.5**	79.1±1.6	**23.0±8.5**	80.6±1.2	**12.0±9.8**	80.7±7.2
standard	7.0±2.7	62.04	0.0±0.0	62.17	0.0±0.0	62.34	0.0±0.0	62.44	0.0±0.0	76.44±3.5
transfer	5.0±4.3	**87.2±0.5**	0.0±0.0	**87.9±0.2**	0.0±0.0	**88.1±0.4**	0.0±0.0	**87.7±0.2**	0.0±0.0	**88.0±1.1**
zero-shot	11.0±0.7	86.2±0.4	11.0±1.0	86.2±0.8	9.6±2.3	86.2±0.2	10.0±4.6	86.0±1.4	6.0±3.3	83.2±2.5
Plain	11th	orig.	11th	orig.	11th	orig.	11th	orig.	11th	orig.
combinet	**53.0±10.0**	84.0±3.6	**45.7±7.6**	84.2±7.8	**31.3±22.0**	83.9±2.4	**28.3±12.6**	82.3±2.2	**23.0±17.4**	84.0±2.4
standard	50.0±7.7	62.54	42.0±3.2	62.18	16.0±12.8	61.67	0.0±0.0	62.27	0.0±0.0	62.27
transfer	4.5±3.0	**86.92**	0.0±0.0	**86.91**	0.0±0.0	**86.96**	0.0±0.0	**87.20**	0.0±0.0	**87.20**
zero-shot	23.0±0.7	86.2±0.5	23.6±1.1	86.2±0.3	22±2.8	86.1±13.9	18.6±3.8	83.7±3.4	15.6±7.3	82.7±2.9

As can be seen in Table 1, the combinet consistently outperforms the baselines at recognizing the newly added eleventh class. We note that the expected CifarNet test accuracy for CIFAR-10 is 85%. Combinets achieve the best accuracy on the newly added class while only losing a small amount of accuracy on average on the 10 original classes. The combinet loss in CIFAR-10 accuracy was almost always due to overgeneralizing. The transfer approach did slightly better than the expected CIFAR-10 accuracy, but this matches previously reported accuracy improvements from retraining (Furlanello et al. 2017).

Foxes clearly confused the baselines, leading to no correctly identified test foxes for the standard or transfer baselines at the lowest values. Compared to plains, foxes had significant overlap in terms of features with cats and dogs. With these smaller size samples transfer and standard were unable to learn or adapt suitable discriminatory features. Comparatively, the conceptual expansion approach was capable of combining existing features into new features that were more successfully able to discriminate between these classes. The zero-shot approach did not require additional training and instead made use of secondary features to make predictions, which was more consistent, but still not as successful as our approach in classifying the new class. In comparison plain was much easier to recognize for our baselines, likely due to the fact that it represented a class that differed significantly from the existing ten. However, our approach was still able to outperform this, creating novel features that could better differentiate the plain class.

Note that combinets do not always outperform these other approaches. For example, the standard approach beats out combinets, getting an average of 83% accuracy with access to all 500 plain training images, while the combinet only achieves an accuracy of roughly 50%. This suggests that combinets are only suited to problems with low training data with this current approach.

Expanding CIFAR-10 to CIFAR-100

For the prior experiments we added a single eleventh class from CIFAR-100 to a CifarNet trained on CIFAR-10. This experiment looks at the problem of expanding a trained CifarNet from classifying the ten classes of the CIFAR-10 dataset to the one-hundred classes of the CIFAR-100 dataset.

For this experiment we limited our training data to ten randomly chosen samples of each CIFAR-100 class. We altered our approach to account for the change in task, constructing an initial mapping for each class individually as if we were expanding a CifarNet to just that eleventh class. We utilized the same three baselines as with the first experiment.

We note that one would not typically utilize CifarNet for this task. Even given access to all 50,000 training samples of CIFAR-100 a CifarNet trained using backpropagation only achieves roughly 30% test accuracy for CIFAR-100. We mean to show the relative scale of accuracy before and after conceptual expansion and not an attempt to achieve state of the art on CIFAR-100 with the full dataset. We tested on the 100,000 test samples available for CIFAR-100.

The average test accuracy across all 100 classes are as follows: the combinet achieves 11.13%, the standard baseline achieves 1.20%, the transfer baseline achieves 6.43%, and the zero-shot baseline achieves 4.10%. We note that our approach is the only one to do better than chance, and significantly outperforms all other baselines. However no approach reaches anywhere near the 30% accuracy that could be achieved with full training data for this architecture.

Pegasus

We return to our running example of an image recognition system that can recognize a pegasus. Unfortunately we lack actual images of a pegasus. To approximate this we collected fifteen photo-realistic, open-use pegasus images from Flickr. Using the same combinet as the above two experiments we ran a 10-5 training/test split and a 5-10 training/test split. For the former we recognized 4 of the 5 pegasus images (80% accuracy), with 80% CIFAR-10 accuracy, and for the latter we recognized 5 of the 10 pegasus images (50% accuracy) with 82% CIFAR-10 accuracy.

DCGAN Experiment

In this section we demonstrate the application of conceptual expansions to generative adversarial networks (GANs). Specifically, we demonstrate the ability to use conceptual expansions to find GANs that can generate images of a class without traditional training on images of that class. We also demonstrate how our approach can take as input an arbitrary number of initial neural networks, instead of the one network for the classification experiments. We make use

Table 2: Summary of results for the GAN experiments.

	combiGAN		combi+N		combi+T		Naive		Transfer	
Samples	I	KL	I	KL	I	KL	I	KL	I	KL
500	3.83±0.32	0.33	**4.61±0.22**	**0.28**	3.05±0.23	0.31	2.98±0.25	0.33	3.38±0.19	1.05
100	4.23±0.15	**0.10**	4.38±0.37	0.29	**4.40±0.19**	0.43	1.76±0.04	0.33	3.26±0.23	0.36
50	**4.05±0.24**	0.22	4.03±0.35	**0.12**	1.69±0.05	2.36	1.06±0.00	10.8	3.97±0.22	0.21
10	4.67±0.44	0.44	**4.79±0.28**	0.13	3.06±0.19	1.20	1.20±0.01	10.8	4.40±0.19	**0.11**

of the DCGAN (Radford, Metz, and Chintala 2015) as the GAN architecture for this experiment, as it has known performance on a number of tasks. We make use of the CIFAR-100 dataset from the prior section and in addition use the Caltech-UCSD Birds-200-2011 (Wah et al. 2011), the CAT (Zhang, Sun, and Tang 2008), the Stanford Dogs (Khosla et al. 2011), FGVC Aircraft (Maji et al. 2013), and the Stanford Cars (Krause et al. 2013) datasets. We make use of these five datasets as they represent five of the ten CIFAR-10 classes, but with significantly more images and images of higher quality. Sharing the majority of classes between experiments allows us to draw comparisons between results.

We trained a DCGAN on each of these datasets till convergence, then used all five of these models as the original knowledge base for our approach. Specifically, we built mappings by testing the proportion of training samples the discriminator of each GAN classified as real. We then built a combinet discriminator for the target class from the discriminators of each GAN. Finally we built a combinet generator from the generators of each GAN, using the combinet discriminators as the heuristic for the conceptual expansion search. We nickname these combinet discriminators and generators *combiGANs*. As above we made use of the fox images of CIFAR-100 for the novel class, varying the number of available images.

We built two baselines: (1) A naive baseline, which involved training the DCGAN on the available fox images in the traditional manner. (2) A transfer baseline, in which we took a DCGAN trained on the Stanford Dogs dataset and retrained it on the fox dataset. We also built two variations of combiGAN: (1) A combiGAN baseline in which we used the discriminator of the naive baseline as the heuristic for the combinet generator (Combi+N). (2) Same as the last, but using the transfer baseline discriminator (Combi+T). We further built a baseline trained on the Stanford Dogs, CAT dataset, and Fox images simultaneously as in (Cheong and Teo 2018), but found that it did not have any improvement over the other baselines. We omit it to save space. We do not include the zero shot baseline from the prior section as it is only suitable for classification tasks.

CombiGAN Results

We made use of two metrics: the inception score (Salimans et al. 2016) and Kullback-Leibler (KL) divergence between generated image classification and true image classification distributions. We acknowledge that inception score was originally designed for ImageNet; since we do not train on ImageNet, we cannot use this as an objective score, but we can use it as a comparative metric of objectness. For the second metric we desired some way to represent how fox-like the generated images were. Thus we made use of the standard classifier trained on 500 foxes, though we could have made use of any classifier in theory. We compare the distribution over classes of real CIFAR-100 fox images and the fake images with the KL divergence. We generated 10,000 images from each GAN to test each metric. We summarize the results of this experiment in Table 2.

We note that in almost all cases our approach or one of its variations (combi+N and combi+T) outperform the two baselines. In the case with 10 training images the transfer baseline beats our approach on our fox-like measure, but this 0.11 differs only slightly from the 0.13 combi+N value. In Figure 1, we include the most fox-like image in terms of classifier confidence from the training samples (real) and each baseline's output. We note that the combiGAN output had a tendency to retain face-like features, while the transfer baseline tended to revert to fuzzy blobs. We also include

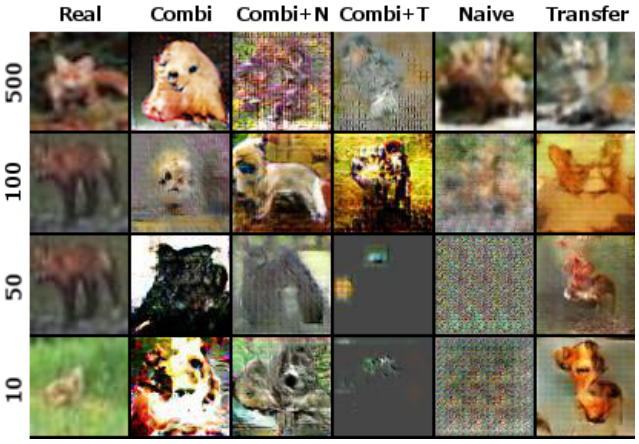

Figure 1: Most fox-like output according to our model for each baseline and sample size.

Figure 2: Four fox-like images hand-picked by the authors from the first 1,000 images output by the combiGAN trained on 500 foxes.

four hand-picked fox-like images from the 500 sample case in Figure 2.

Discussion and Limitations

Conceptual expansions of neural networks—combinets and combiGANs—outperform standard approaches on problems with limited data without additional knowledge engineering. We refer to this approach generally as conceptual expansion, which is inspired by the human ability to make conceptual leaps by combining existing knowledge. Our main contribution in this paper is the initial exploration of conceptual expansion of neural networks; we speculate that more sophisticated optimization search routines may achieve greater improvements.

We anticipate the future performance of conceptual expansions to depend upon the extent to which the existing knowledge base contains relevant information to the new problem and ability for the optimization function to find helpful conceptual expansions. We note that one choice of optimization function could be human intuition, and we have had success hand-designing conceptual expansions for sufficiently small problems.

Conceptual expansions appear less dependent on training data than existing transfer learning approaches as evidenced by the comparative performance of the approach with low training data, This is further evidenced by those instances where conceptual expansion outperformed itself with less training data. We anticipate further exploration of this in future work.

Conclusions

We present conceptual expansion of neural networks: *combinets*, an approach to produce recombined versions of existing machine learned deep neural net models. We ran four experiments of this approach compared to common baselines, and found we were able to achieve greater accuracy with less data. Our technique relies upon a flexible representation of recombination of existing knowledge that allows us to represent new knowledge as a combination of particular knowledge from existing cases. To our knowledge this represents the first attempt at applying a model of combinational creativity to neural networks.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. IIS-1525967. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

Ba, L. J.; Swersky, K.; Fidler, S.; and Salakhutdinov, R. 2015. Predicting deep zero-shot convolutional neural networks using textual descriptions. In *International Conference on Computer Vision*, 4247–4255.

Boden, M. A. 2004. *The creative mind: Myths and mechanisms*. Psychology Press.

Boden, M. A. 2009. Computer models of creativity. *AI Magazine* 30(3):23–23.

Chao, W.-L.; Changpinyo, S.; Gong, B.; and Sha, F. 2016. An empirical study and analysis of generalized zero-shot learning for object recognition in the wild. In *European Conference on Computer Vision*, 52–68. Springer.

Cheong, B., and Teo, H. 2018. Can we train dogs and humans at the same time? gans and hidden distributions. Technical report, Stanford.

Cojan, J., and Lieber, J. 2009. Belief merging-based case combination. In *ICCBR*, 105–119. Springer.

Cunha, J. M.; Gonçalves, J.; Martins, P.; Machado, P.; and Cardoso, A. 2017. A pig, an angel and a cactus walk into a blender: A descriptive approach to visual blending. *arXiv preprint arXiv:1706.09076*.

Daumé III, H. 2009. Frustratingly easy domain adaptation. *arXiv preprint arXiv:0907.1815*.

De Mantaras, R. L.; McSherry, D.; Bridge, D.; Leake, D.; Smyth, B.; Craw, S.; Faltings, B.; Maher, M. L.; T COX, M.; Forbus, K.; et al. 2005. Retrieval, reuse, revision and retention in case-based reasoning. *The Knowledge Engineering Review* 20(3):215–240.

Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Fei-Fei, L. 2009. Imagenet: A large-scale hierarchical image database. In *CVPR09*.

Elhoseiny, M.; Zhu, Y.; Zhang, H.; and Elgammal, A. 2017. Zero shot learning from noisy text description at part precision. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Fauconnier, G. 2001. Conceptual blending and analogy. *The analogical mind: Perspectives from cognitive science* 255–286.

Fei-Fei, L.; Fergus, R.; and Perona, P. 2006. One-shot learning of object categories. *IEEE transactions on pattern analysis and machine intelligence* 28(4):594–611.

Floreano, D.; Dürr, P.; and Mattiussi, C. 2008. Neuroevolution: from architectures to learning. *Evolutionary Intelligence* 1(1):47–62.

Fox, J., and Clarke, S. 2009. Exploring approaches to dynamic adaptation. In *Proceedings of the 3rd International DiscCoTec Workshop on Middleware-Application Interaction*, 19–24. ACM.

Furlanello, T.; Lipton, Z. C.; Amazon, A.; Itti, L.; and Anandkumar, A. 2017. Born again neural networks. In *NIPS Workshop on Meta Learning*.

Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; Marchand, M.; and Lempitsky, V. 2016. Domain-adversarial training of neural networks. *The Journal of Machine Learning Research* 17(1):2096–2030.

Guzdial, M., and Riedl, M. 2016. Learning to blend computer game levels. In *Seventh International Conference on Computational Creativity*.

Guzdial, M., and Riedl, M. 2018. Automated game design

via conceptual expansion. In *Fourteenth Artificial Intelligence and Interactive Digital Entertainment Conference*.

Hervás, R., and Gervás, P. 2006. Case-based reasoning for knowledge-intensive template selection during text generation. In *European Conference on Case-Based Reasoning*, 151–165. Springer.

Khosla, A.; Jayadevaprakash, N.; Yao, B.; and Fei-Fei, L. 2011. Novel dataset for fine-grained image categorization. In *First Workshop on Fine-Grained Visual Categorization, IEEE Conference on Computer Vision and Pattern Recognition*.

Konieczny, S., and Pérez, R. P. 2011. Logic based merging. *Journal of Philosophical Logic* 40(2):239–270.

Krause, J.; Stark, M.; Deng, J.; and Fei-Fei, L. 2013. 3d object representations for fine-grained categorization. In *4th International IEEE Workshop on 3D Representation and Recognition (3dRR-13)*.

Krizhevsky, A., and Hinton, G. 2009. Learning multiple layers of features from tiny images. *Citeseer*.

Kulis, B.; Saenko, K.; and Darrell, T. 2011. What you saw is not what you get: Domain adaptation using asymmetric kernel transforms. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, 1785–1792. IEEE.

Kuzborskij, I., and Orabona, F. 2013. Stability and hypothesis transfer learning. In *International Conference on Machine Learning*, 942–950.

Kuzborskij, I.; Orabona, F.; and Caputo, B. 2013. From n to n+ 1: Multiclass transfer incremental learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3358–3365.

Lampert, C. H.; Nickisch, H.; and Harmeling, S. 2009. Learning to detect unseen object classes by between-class attribute transfer. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 951–958. IEEE.

Levy, O., and Markovitch, S. 2012. *Teaching machines to learn by metaphors*. Technion-Israel Institute of Technology, Faculty of Computer Science.

Li, B.; Zook, A.; Davis, N.; and Riedl, M. O. 2012. Goal-driven conceptual blending: A computational approach for creativity. In *Proceedings of the 2012 International Conference on Computational Creativity, Dublin, Ireland*, 3–16.

Li, J.; Seltzer, M. L.; Wang, X.; Zhao, R.; and Gong, Y. 2017. Large-scale domain adaptation via teacher-student learning. *arXiv preprint arXiv:1708.05466*.

Maji, S.; Kannala, J.; Rahtu, E.; Blaschko, M.; and Vedaldi, A. 2013. Fine-grained visual classification of aircraft. *arXiv preprint arXiv:1306.5151*.

Manzano, S.; Ontañón, S.; and Plaza, E. 2011. Amalgam-based reuse for multiagent case-based reasoning. In *International Conference on Case-Based Reasoning*, 122–136. Springer.

Mensink, T.; Gavves, E.; and Snoek, C. G. 2014. Costa: Co-occurrence statistics for zero-shot classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2441–2448.

Norouzi, M.; Mikolov, T.; Bengio, S.; Singer, Y.; Shlens, J.; Frome, A.; Corrado, G. S.; and Dean, J. 2013. Zero-shot learning by convex combination of semantic embeddings. *arXiv preprint arXiv:1312.5650*.

Ontañón, S., and Plaza, E. 2010. Amalgams: A formal approach for combining multiple case solutions. In *Case-Based Reasoning. Research and Development*. Springer. 257–271.

Pereira, F.; Norvig, P.; and Halevy, A. 2009. The unreasonable effectiveness of data. *IEEE Intelligent Systems* 24(undefined):8–12.

Radford, A.; Metz, L.; and Chintala, S. 2015. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*.

Rebuffi, S.-A.; Kolesnikov, A.; Sperl, G.; and Lampert, C. H. 2017. icarl: Incremental classifier and representation learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Salimans, T.; Goodfellow, I.; Zaremba, W.; Cheung, V.; Radford, A.; and Chen, X. 2016. Improved techniques for training gans. In *Advances in Neural Information Processing Systems*, 2234–2242.

Sizov, G.; Öztürk, P.; and Aamodt, A. 2015. Evidence-driven retrieval in textual cbr: bridging the gap between retrieval and reuse. In *International Conference on Case-Based Reasoning*, 351–365. Springer.

Thagard, P., and Stewart, T. C. 2011. The aha! experience: Creativity through emergent binding in neural networks. *Cognitive science* 35(1):1–33.

Wah, C.; Branson, S.; Welinder, P.; Perona, P.; and Belongie, S. 2011. The caltech-ucsd birds-200-2011 dataset. Technical Report CNS-TR-2011-001, California Institute of Technology.

Wang, Y.-X., and Hebert, M. 2016. Learning to learn: Model regression networks for easy small sample learning. In *European Conference on Computer Vision*, 616–634. Springer.

Wilke, W., and Bergmann, R. 1998. Techniques and knowledge used for adaptation during case-based problem solving. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, 497–506. Springer.

Wong, J. H., and Gales, M. J. 2016. Sequence student-teacher training of deep neural networks. *International Speech Communication Association*.

Xian, Y.; Schiele, B.; and Akata, Z. 2017. Zero-shot learning-the good, the bad and the ugly. *arXiv preprint arXiv:1703.04394*.

Yao, Y.; Zhang, J.; Shen, F.; Hua, X.; Xu, J.; and Tang, Z. 2017. Exploiting web images for dataset construction: A domain robust approach. *IEEE Transactions on Multimedia* 19(8):1771–1784.

Zhang, W.; Sun, J.; and Tang, X. 2008. Cat head detection-how to effectively exploit shape and texture features. In *European Conference on Computer Vision*, 802–816. Springer.