Multi-level Classification: Implications for Human-like Generalization

Joshua Peterson, Paul Soulos, Aida Nematzadeh, and **Tom Griffiths**
Department of Psychology, University of California, Berkeley

Abstract

What labels should we use for object classification to support as many downstream tasks as possible? Immensely popular datasets in computer vision often make use of somewhat arbitrary, flat label sets. By contrast, research on human category learning focuses on the multi-level nature of categorization (e.g., the same object is called both a "Dalmatian" and a "dog"). Such hierarchies of increasing abstraction provide additional constraints on the representations to be learned that support mutual multi-level categorization, and further comprise the structure of the world that humans leverage to make robust generalizations with little data. In the following work, we show that multi-level categorization objectives can learn fully compatible representations, and that labels at different levels of abstractions support both different representations and alternative priors for generalization behavior.

1 Category Learning in Children and AI

One of the challenges for a child learning a language is identifying what level in a hierarchical taxonomy a word – category label – refers to. After hearing the word "dog" upon observing a Dalmatian called Sebastian, a child needs to learn whether the category label "dog" refers to only Sebastian, all the Dalmatians, different breeds of dogs, etc. A classic line of psychological research examined what level of abstraction (*e.g.*, dogs vs. Dalmatians) conveys the most information; this level is called the *basic level* [Rosch et al., 1976]. Another line of research has investigated whether children (and people in general) have an innate or a learned basic-level bias, *i.e.*, a tendency to generalize a new category label to the members of the basic level of the taxonomy [*e.g.*, Markman, 1991, Golinkoff et al., 1994]. In computer vision, this problem is formulated as a classification task where the training data consists of images paired with single labels.

However, this problem formulation is different from what children experience. Multiple labels (*e.g.*, Dalmatian and dog) are commonly used to refer to the same entity in the world. This explicit use of multiple labels can help children learn a better representation of the taxonomic relations between categories – the implicit hierarchical structure underlying the world. Different breeds of dogs form a category not only because of their perceptual similarity but also because they are all referred to by the label "dog". Moreover, forming a hierarchical representation can in turn help children better generalize to new items; for example, a new breed of dog, such as a poodle, will be categorized as a dog because of its similarity to the members of that category.

Here we explore the consequences of multi-level classification by training a deep image classifier on images paired with multiple labels, each corresponding to a different level of the hierarchical taxonomy. In particular, we focus on two sets of labels, basic and subordinate; subordinate labels (such as "Dalmatian") are categories that are below the basic level (such as "dog") in the hierarchical taxonomy. We find that an array of models trained on these two sets achieve good classification accuracy on both, with no apparent competition, yet the representations learned from including basic level labels (or using them exclusively) capture different levels of similarity observed in a hierarchical taxonomy, including high-level groups not explicit in the original training labels. Finally, we show that basic labels result in a better match of basic-level generalization behavior observed in people using a simple model that restricts generalization with the number of consistent examples.

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

2 Training Methodology

We pose the multi-level labeling problem simply as learning a set of independent softmax classifiers that are unconnected to each other but fully connected to the final representation layer of a deep CNN. While other alternatives are possible, we ask here whether a single representation is compatible with both objectives simultaneously. We use the InceptionV3 [Szegedy et al., 2015] as our base model. To speed training and since we were mostly interested in representations learned in late layers, we use the original pre-trained model weights and fully reinitialize the final module to encourage re-learning (approximately 8% of the layers). We use a hyperparameter α to control the weighting of the two crossentropy loss terms: $\alpha\mathcal{L}_{basic} + (1-\alpha)\mathcal{L}_{subordinate}$. We use the 1000 labels from ILSVRC12 as our subordinate classes, and basic level labels from Wang and Cottrell [2015]. Approximately 700,000 images are used from the total set, following the stratification procedure of the original work.

3 Multi-Level Classification Performance

Classification accuracy for the full validation set for all experiments is given in Table 1. All networks were trained for three epochs. Surprisingly, varying the weight between the subordinate and basic losses did not appear to affect the classification accuracy for either label set. This leads us to believe that it is possible to include the basic level human knowledge without sacrificing the ultimate performance of the network. Although the original pretrained network obtains higher accuracy in predicting subordinate classes, our truncated dataset is only approximately 58% the size of the original dataset, and so a moderate decrease in overall performance is expected.

Table 1: Top-1 and Top-5 accuracy for each α.

	Basic Labels		**Subordinate Labels**	
α	Top-1	Top-5	Top-1	Top-5
0	–	–	72.53	91.23
.25	81.88	94.96	72.54	91.25
.50	81.77	94.99	72.42	91.27
.75	82.11	95.15	72.48	91.26
1	81.9	95.04	–	–

4 Exploring Representations

We first visualized the representations using t-SNE in Figure 1, revealing that the model trained only on subordinate labels does not appear to cluster basic level categories (e.g., dog) very well, whereas the model trained on basic labels does a much more effective job of clustering like subordinate classes together. Models trained on both label sets simultaneously exhibited similar clustering (not shown).

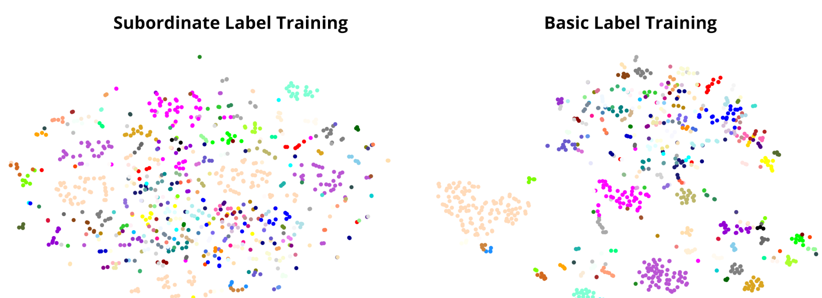

Figure 1: t-SNE visualizations of representations learned for subordinate only training and basic only training, colored by basic label. The basic model has tighter clusters and more distinct boundaries between categories.

We also plot hierarchical clusterings of the the representations in Figure 2. The model trained on basic labels has a clearly defined hierarchy, whereas the model trained on subordinate labels does not. The branch we call "artificial" contains images of artificial objects such as cars, buildings, household objects, sports, and technology. The other main branch includes animals, fish, mushrooms, and other "natural" stimuli. Interestingly, this high-level distinction is not present explicitly in the basic label set, and is one of the defining categorical divisions found in human object representations [Mur et al., 2013].

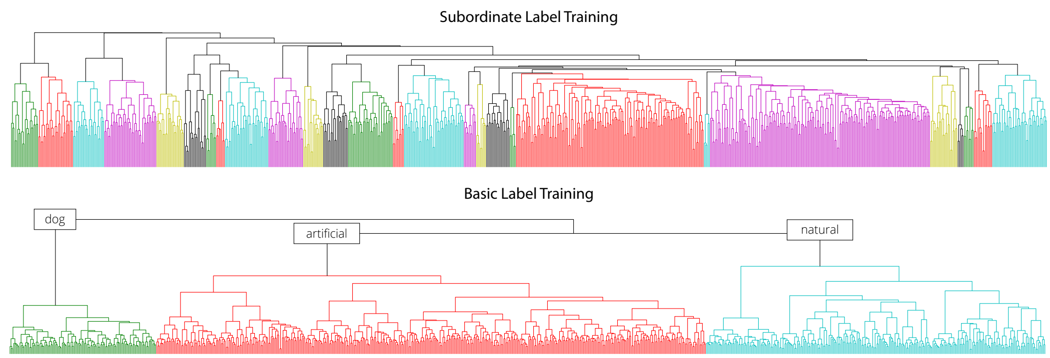

Figure 2: Hierarchical clusterings of the representations. The model trained on subordinate labels (top) has no clear taxonomic structure. The model trained on basic labels (bottom) has a clearly defined structure which divides objects into three high level categories.

5 Generalization Experiments

Xu and Tenenbaum [2007], henceforth X&T, examined how people generalize a novel word label after observing a few examples, and whether the number or the taxonomic level from which the examples were drawn changes the generalization behavior. For example, the participants heard a word label such as “dax” while observing one Dalmatian, three Dalmatians, or three different breeds of dogs. We examine whether our models exhibit the generalization behavior observed in these experiments. We focus on three of their training conditions: “1 sub”, where the learner is given 1 (subordinate) example, such as a Dalmatian, “3 sub”, where the learner is given 3 examples from the same subordinate category, such as 3 Dalmatians, and “3 basic’, where 3 examples are drawn from a basic-level, such as a poodle, a Dalmatian, and a Great Pyrenees.

In the X&T’s experiments, participants were asked after training to pick everything that is a “dax” from a fixed set of examples drawn from different levels of the taxonomy. We focus on the test objects that are relevant to our experimental conditions, *i.e.*, two basic-level and two subordinate matches.[1] For example, after observing one Dalmatian as a “dax”, the participants had to decide whether two other Dalmatians, a poodle, and a golden retriever are also examples of “dax”. The results of their experiments is shown in Figure 3. Two interesting observations can be made from their results: First, people exhibit basic-level bias since they generalize to the basic-level after observing only subordinate examples. Second, their degree of basic-level generalization decreases after observing 3 examples (see “Basic Match” in 1-sub and 3-sub conditions in Figure 3).

Learning from Positive Examples. To mimic generalization experiments with humans, we use a simple k-shot generalization model meant to learn concepts from positive examples only. Specifically, we use exponentiated euclidean distance (in the spirit of Shepard et al. [1987], given that we are interesting in making comparisons to human behavior), normalized over distances to all items in the test set to obtain generalization probabilities. Given that it has been demonstrated that human generalization behavior tightens with increasing examples [Tenenbaum, 2000], we also include the contribution of the number of examples n in our final model:

$$g(q_i, c) = \frac{e^{-nd(q_i,c)}}{\sum_j e^{-nd(q_j,c)}}, \tag{1}$$

where g is the generalization function, c is a concept (a single training example or the mean of n training examples), and q_i is the query (test) image. Since we care only about the relative differences in generalization probabilities across different levels of testing stimuli, we linearly normalize each set of probabilities so that the largest becomes 1. To evaluate this model, we sample test and train stimuli analogous to X&T from 88 of the basic level labels in our set that encompassed at least 3 subordinate classes, a requirement for the final three-example experiment condition, and average the results over 10 iterations.

[1] X&T also included superordinate examples (such as an animal other than a dog). Here we only focus on the subordinate and basic-level matches because our training data only includes those labels.

Results. Results from generalization experiments from Basic and Subordinate Label Training are shown in the left in middle plots of Figure 3. Similarly to humans, both models exhibit basic-level bias after training on 1 example or 3 subordinate examples, and this basic-level generalization decreases as the model receives more examples (compare 1 sub to 3 sub conditions). Both models also exhibit similarity to humans in their subordinate and basic-level generalization behavior after observing 3 examples drawn from a basic-level category. Both of these effects more strongly resemble humans in the case of Basic Label Training.

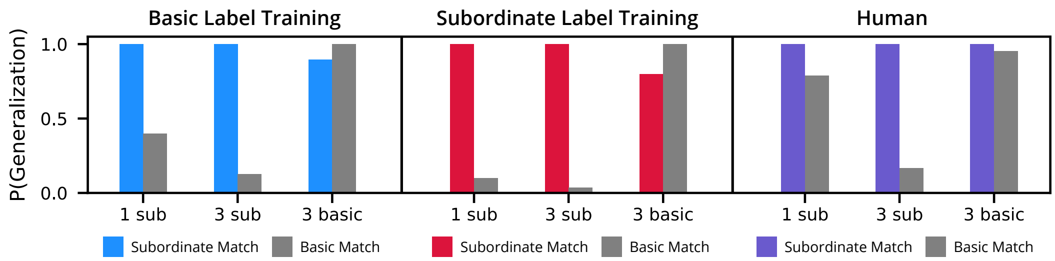

Figure 3: The results of generalization experiments from our trained models and the human data from X&T's experiments.

6 Discussion

Image classification models are often trained on images paired with single labels that correspond to the subordinate level of the hierarchical taxonomy. People, on the other hand, use multiple labels to refer to the same entity in the world (*e.g.*, dog and Dalmatian). This explicit use of multiple labels makes the category learning problem easier as people can use the labels to identify similarities between the concepts (in addition to other features, such as perceptual cues, that signal similarity). This hierarchical nature of the representation is in turn helpful in categorizing and labeling a new entity – a problem that is often formulated as a few or one shot learning task in machine learning.

However, to our knowledge, none of the existing modern image classification models simultaneously train on multiple labels. In this work, we offer initial evidence that training on basic-level and subordinate labels (or just basic-level labels) results in representations that better capture the hierarchical structure of taxonomy of real-world objects. We also show that these representations better match the basic-level bias observed in human generalization behavior. Future work should examine whether such taxonomic biases can be helpful in few-shot learning applications.

References

R. M. Golinkoff, C. B. Mervis, and K. Hirsh-Pasek. Early object labels: The case for a developmental lexical principles framework. *Journal of child language*, 21(01):125–155, 1994.

E. M. Markman. *Categorization and naming in children: Problems of induction*. Mit Press, 1991.

M. Mur, M. Meys, J. Bodurka, R. Goebel, P. A. Bandettini, and N. Kriegeskorte. Human object-similarity judgments reflect and transcend the primate-it object representation. *Frontiers in psychology*, 4, 2013.

E. Rosch, C. B. Mervis, W. D. Gray, D. M, and P. Boyes-Braem. Basic objects in natural categories. *Cognitive Psychology*, 1976.

R. N. Shepard et al. Toward a universal law of generalization for psychological science. *Science*, 237(4820): 1317–1323, 1987.

C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z.-n. Wojna. Rethinking the inception architecture for computer vision. corr abs/1512.00567 (2015), 2015.

J. B. Tenenbaum. Rules and similarity in concept learning. In *Advances in neural information processing systems*, pages 59–65, 2000.

P. Wang and G. W. Cottrell. Basic level categorization facilitates visual object recognition. *arXiv preprint arXiv:1511.04103*, 2015.

F. Xu and J. B. Tenenbaum. Word learning as bayesian inference. *Psychological review*, 114(2):245, 2007.