Knowledge-Driven Hallucination for Low-Shot Classification

The field of visual recognition has made many strides in the last few years. However, high performance for this task relies on a fully supervised learning paradigm, which is often impractical, as gathering richly annotated data is expensive and inefficient. Moreover, reliance on these methods is not ideal, as we humans are readily able to generalize our understanding of previously-learned classes to the unknown. Thus, we require only a small number of examples to learn how to recognize something new. Inspired by this, our project aims to enable computers to learn classifiers for new object categories with only a few labelled examples after establishing a solid knowledge base. This paradigm of learning is called low-shot learning.

In this project, we build on the work of Hariharan and Girshick[1]. In [1], the set of training examples for low-shot categories is artificially expanded through the process of "hallucination." Within any category, there are visual "transformations" that we can observe between any two samples (e.g., sitting dog to standing dog). The method in [1] takes the transformations found in known classes (i.e., classes with a large number of training samples) and applies them randomly to samples in low-shot classes to create new low-shot data-points. In this work, we extend this method by incorporating explicit knowledge of inter-category relationships.

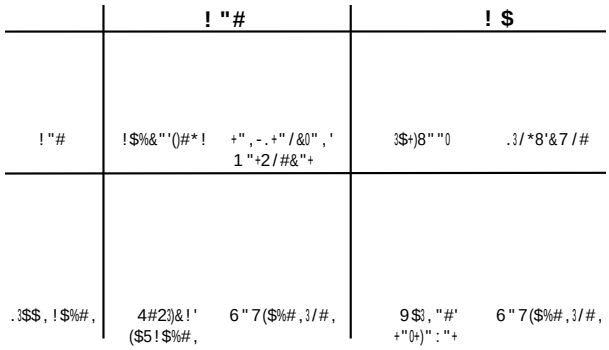

Figure 1: Illustration of how we use categorical relationships in our framework; the first column is a low-shot class, the second and third column show examples of related classes according to Word2Vec and WordNet respectively.

The goal of this project is to leverage knowledge of how different categories are related to each other for low-shot recognition. In preliminary experiments, we have explored two ways of incorporating inter-category relationships in the hallucination method. First, we only hallucinate new samples in low-shot classes by taking transformations from a semantically similar known class ("RG" in Table 1). Second, we consider a higher-order analysis for inter-category relationships. Here, we not only borrow from one related class, but from many, giving a more robust basis of knowledge for hallucination. To that end, we average across multiple generated data-points using transformations from various related known classes, and use that average as the new sample in the low-shot class ("GG" in Table 1). For both these setups, we use the semantic hierarchy from WordNet and a Word2Vec model as knowledge sources ("KB" in Table 1) to ensure that the given low-shot class and any borrowed known classes are related.

For the low-shot benchmark [1], our method of introducing semantic similarity to sample generation improves over the baseline in very low-shot settings (Table 1, $n = 1, 2$). Our method improves the accuracy of recognition for low-shot classes alone, as well as for all (known and low-shot) classes. We also show that our averaging method improves on the baseline for these same scenarios.

Table 1: ***Preliminary quantitative results****: Top-1 and top-5 accuracy metrics from ImageNet are reported below for low-shot and all classes (see [1] for benchmark details). "n" refers to n-shot classification (e.g., n=2 means 2 labeled ground-truth samples per class). We use a total of 5 samples (generated+ground truth) per low-shot class. All methods use ResNet10.* ***Legend****: KB = Knowledge base, RG = Related Generation, GG = Grouped Generation, W2V = Word2Vec, WN = WordNet*

				Low-Shot Classes				All Classes			
				Top-1		Top-5		Top-1		Top-5	
	KB	RG	GG	n=1	n=2	n=1	n=2	n=1	n=2	n=1	n=2
Baseline [1]				8.7	16.5	31.0	45.6	34.0	38.0	53.6	61.8
Ours	W2V	X		10.5	17.0	**34.4**	47.0	**34.5**	**38.3**	**55.1**	**62.5**
		X	X	10.1	16.8	34.0	47.0	34.3	38.0	54.9	62.4
	WN	X		**10.6**	16.9	34.2	**47.1**	**34.5**	38.0	**55.1**	62.4
		X	X	9.6	**17.2**	32.7	47.0	34.2	38.2	54.3	**62.5**

Given these promising preliminary results for knowledge-driven hallucination, we plan to extend this work in higher-order analysis. We propose to encode explicit category relationships into a knowledge graph, using image features or semantic category names as input. Then, we propose to use this knowledge graph in a graph convolutional network (GCN) framework to learn to generate samples of low-shot classes with a more sophisticated understanding of the known classes. In addition, we propose to do diagnostics on this hallucination method. In particular, we want to know which canonical low-shot examples are the most beneficial for seeding hallucinations through learned transformations, so as to discourage the generation of visually outlying samples.

[1] [1] B. Hariharan and R. Girshik. Low-shot Visual Recognition by Shrinking and Hallucinating Features. *ICCV*, 2017