ImageNet [4]: A Large-Scale Hierarchical Image Database [5]

Qi Zhao

June 6, 2018

1. Introduction

The explosion of image data on the Internet has the potential to foster more sophisticated and robust models and algorithms to index, retrieve, organize and interact with images and multimedia data. The article introduces here a new database called ImageNet, a large-scale ontology of images built upon the backbone of the WordNet [1] structure. ImageNet aims to populate the majority of the 80,000 synsets of WordNet with an average of 500-1000 clean and full resolution images. ImageNet is much larger in scale and diversity and much more accurate than the current image datasets. It believes that a large-scale ontology of images is a critical resource for developing advanced, large-scale content-based image search and image understanding algorithms, as well as for providing critical training and benchmarking data for such algorithms. ImageNet uses the hierarchical structure of WordNet. Each meaningful concept in WordNet, possibly described by multiple words or word phrases, is called a synonym set [3] or synset. There are around 80, 000 noun synsets in Word Net. In ImageNet, we aim to provide on average 500-1000 images to illustrate each synset. Images of each concept are quality-controlled and human-annotated. Figure 1 shows a snapshot of two branches of the mammal and vehicle subtrees.

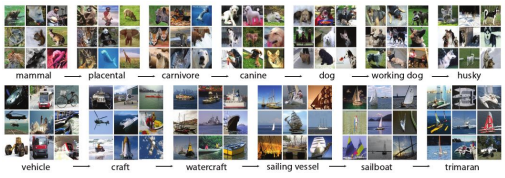

Figure 1. A snapshot of two root-to-leaf branches of ImageNet: the top row is from the mammal subtree; the bottom row is from the vehicle subtree. For each synset, 9 randomly sampled images are presented.

2. Constructing ImageNet

The first stage of the construction of ImageNet involves collecting candidate images for each synset. It therefore collects a large set of candidate images. After intra-synset duplicate removal, each synset has over 10K images on average. We collect candidate images from the Internet by querying several image search engines. For each synset, the queries are the set of WordNet synonyms. Search engines typically limit the number of images retrievable (in the order of a few hundred to a thousand). To obtain as many images as possible, we expand the query set by appending the queries with the word from parent synsets, if the same word appears in the gloss of the target synset.

The second stage of the construction of ImageNet is cleaning candidate image. In each of our labeling tasks, it presents the users with a set of candidate images and the definition of the target synset (including a link to Wikipedia). For each synset, we first randomly sample an initial subset of images. At least 10 users are asked to vote on each of these images. We then obtain a confidence score table, indicating the probability of an image being a good image given the user votes. For each of remaining candidate images in this synset, we proceed with the Amazon Mechanical Turk(AMT) [2] user labeling until a predetermined confidence score threshold is reached. The algorithm successfully filters the candidate images, resulting in a high percentage of clean images per synset , which is shown in the Figure 2.

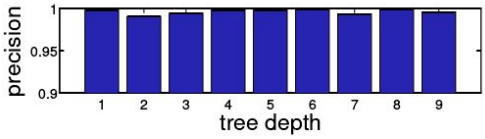

Figure 2. Percent of clean images at different tree depth levels in ImageNet. A total of 80 synsets are randomly sampled at every tree depth of the mammal and vehicle subtrees. An independent group of subjects verified the correctness of each of the images. An average of 99.7% precision is achieved for each synset.

References

[1] C. Fellbaum and G. Miller. *WordNet: An electronic lexical database*. MIT Press, 1998. 1

[2] L. Irani. *Amazon mechanical turk*. John Wiley & Sons, Ltd, 2017. 1

[3] J. Mccrae and N. Collier. Synonym set extraction from the biomedical literature by lexical pattern discovery. *BMC Bioinformatics*, 9(1):1–13, 2008. 1

[4] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, and M. Bernstein. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2014. 1

[5] H. Wu, L. Wang, F. Zhang, and Z. Wen. Automatic leaf recognition from a big hierarchical image database. *International Journal of Intelligent Systems*, 30(8):871–886, 2015. 1