IMAGENET

ImageNet is an image database organized according to the WordNet hierarchy (currently only the nouns), in which each node of the hierarchy is depicted by hundreds and thousands of images.

Each meaningful concept in WordNet, possibly described by multiple words or word phrases, is called a "synonym set" or "synset". There are more than 100,000 synsets in WordNet; Hypernyms are synsets that are more general e.g., "organism" is a hypernym of "plant". Hyponyms are synsets that are more specific e.g., "aquatic" is a hyponym of "plant". The majority of them are nouns (80,000+). This hierarchy makes it useful for computer vision tasks. If the model is not sure about a subcategory, it can simply classify the image higher up the hierarchy where the error probability is less.

In ImageNet, the aim is to provide on average 1000 images to illustrate each synset. Images of each concept are quality-controlled and human-annotated.

This hierarchy makes it useful for computer vision tasks. If the model is not sure about a subcategory, it can simply classify the image higher up the hierarchy where the error probability is less.

In its completion, it is expected that ImageNet will offer tens of millions of cleanly labelled and sorted images for most of the concepts in the WordNet hierarchy.

The project has been instrumental in advancing computer vision and deep learning research. The data is available for free to researchers for non-commercial use.

Why ImageNet?

The ImageNet project was inspired by two important needs in computer vision research.

1) The need to establish a clear North Star problem in computer vision. While the field enjoyed an abundance of important tasks to work on, from stereo vision to image retrieval, from 3D reconstruction to image segmentation, object categorization was recognized to be one of the most fundamental capabilities of both human and machine vision. Hence there was a growing demand for a high-quality object categorization benchmark with clearly established evaluation metrics.
2) There was a critical need for more data to enable more generalizable machine learning methods. Ever since the birth of the digital era and the availability of web-scale data exchanges, researchers in these fields have been working hard to design more and more sophisticated algorithms to index, retrieve, organize and annotate multimedia data. But good research requires good resources. To tackle this problem at scale (think of your growing personal collection of digital images, or videos, or a commercial web

search engine's database), it was critical to provide researchers with a large-scale image database for both training and testing.

Does ImageNet own the images? Can I download the images?

No, ImageNet does not own the copyright of the images. ImageNet only compiles an accurate list of web images for each synset of WordNet. For researchers and educators who wish to use the images for non-commercial research and/or educational purposes, they can provide access through their site under certain conditions and terms.

What is the ImageNet Challenge and what's its connection with the dataset?

- ImageNet Large Scale Visual Recognition Challenge (ILSVRC) was an annual computer vision contest held between 2010 and 2017. It's also called ImageNet Challenge.

- For this challenge, the training data is a subset of ImageNet: 1000 synsets, 1.2 million images. Images for validation and test are not part of ImageNet and are taken from Flickr and via image search engines. There are 50K images for validation and 150K images for testing. These are hand-labeled with the presence or absence of 1000 synsets.

- The Challenge included three tasks: image classification, single-object localization (since ILSVRC 2011), and object detection (since ILSVRC 2013). More difficult tasks are based upon these tasks. In particular, image classification is the common denominator for many other computer vision tasks. Tasks related to video processing, but not part of the main competition, were added in ILSVRC 2015. These were object detection in video and scene classification.

How were the images labelled in ImageNet?

- In the early stages of the ImageNet project, a quick calculation showed that by employing a few people, they would need 19 years to label the images collected for ImageNet. But in the summer of 2008, researchers came to know about an Amazon service called Mechanical Turk. This meant that image labelling can be crowdsourced via this service. Humans all over the world would label the images for a small fee.

- Humans make mistakes and therefore we must have checks in place to overcome them. Each human is given a task of 100 images. In each task, 6 "gold standard" images are placed with known labels. At most 2 errors are allowed on these standard images, otherwise the task has to be restarted.

- In addition, the same image is labelled by three different humans. When there's disagreement, such ambiguous images are resubmitted to another human with tighter quality threshold (only one allowed error on the standard images).

What is meant by a pretrained ImageNet model?

- A model trained on ImageNet has essentially learned to identify both low-level and high-level features in images. However, in a real-world application such as medical image analysis or handwriting recognition, models have to be trained from data drawn from those application domains. This is time consuming and sometimes impossible due to lack of sufficient annotated training data.

- One solution is that a model trained on ImageNet can use its weights as a starting point for other computer vision task. This reduces the burden of training from scratch. A much smaller annotated domain-specific training may be sufficient. By 2018, this approach was proven in a number of tasks including object detection, semantic segmentation, human pose estimation, and video recognition.

What are the criticisms or shortcomings of ImageNet?

- Though ImageNet has a large number of classes, *most of the classes don't represent everyday entities.* One researcher, Samy Bengio, commented that the WordNet categories don't reflect the interests of common people. He added, "Most people are more interested in Lady Gaga or the iPod Mini than in this rare kind of diplodocus".

- *Images are not uniformly distributed across subcategories.* One research team found that by considering 200 subcategories, they found that the top 11 had 50% of the images, followed by a long tail.

- *When classifying people, ImageNet uses labels that are racist, misogynist and offensive.* People are treated as objects. Their photos have been used without their knowledge. About 5.8% labels are wrong.

- *ImageNet lacks geodiversity.* Most of the data represents North America and Europe. China and India are represented in only 1% and 2.1% of the images respectively. This implies that models trained on ImageNet will not work well when applied for the developing world.

- Another study from 2016 found that 30% of *ImageNet's image URLs are broken.* This is about 4.4 million annotations lost. Copyright laws prevent caching and redistribution of these images by ImageNet itself.

How is Tiny ImageNet related to ImageNet?

- Tiny ImageNet and its associated competition are part of Stanford University's CS231N course. It was created for students to practise their skills in creating models for image classification.

- The Tiny ImageNet dataset has 100,000 images across 200 classes. Each class has 500 training images, 50 validation images, and 50 test images. Thus, the dataset has 10,000 test images. The entire dataset can be downloaded from a Stanford server.

What are some technical details of ImageNet?

- ImageNet consists of 14,197,122 images organized into 21,841 subcategories. These subcategories can be considered as sub-trees of 27 high-level categories. Thus, ImageNet is a well-organized hierarchy that makes it useful for supervised machine learning tasks.

- On average, there are over 500 images per subcategory. The category "animal" is most widely covered with 3822 subcategories and 2799K images. The "appliance" category has on average 1164 images per subcategory, which is the most for any category. Among the categories with least number of images are "amphibian", "appliance", and "utensil".

- As many as 1,034,908 images have been annotated with *bounding boxes*. For example, if an image contains a cat as its main subject, the coordinates of a rectangle that bounds the cat are also published on ImageNet. This makes it useful for computer vision tasks such as object localization and detection.

- Then there's *Scale-Invariant Feature Transform (SIFT)* used in computer vision. SIFT helps in detecting local features in an image. ImageNet gives researchers 1000 subcategories with SIFT features covering about 1.2 million images.

- Images vary in resolution but it's common practice to train deep learning models on sub-sampled images of 256x256 pixels.

References:

1) https://www.image-net.org/
2) Devopedia. 2021. "ImageNet." Version 16, April 7. Accessed 2022-06-15. https://devopedia.org/imagenet