What makes ImageNet good for transfer learning?

Presentation for Stanford CS331B
Nov. 2 2016
Trevor Standley

Paper authors:

(On arXiv Aug 2016)

Minyoung Huh
Pulkit Agrawal
Alexei A. Efros

Berkeley Artificial Intelligence Research (BAIR) Laboratory

Representations Pre-trained on ImageNet

- When you have enough data to train a vision system end-to-end, do that.
- For every other situation people use ImageNet pre-training.

IMAGENET

Transfer Learning

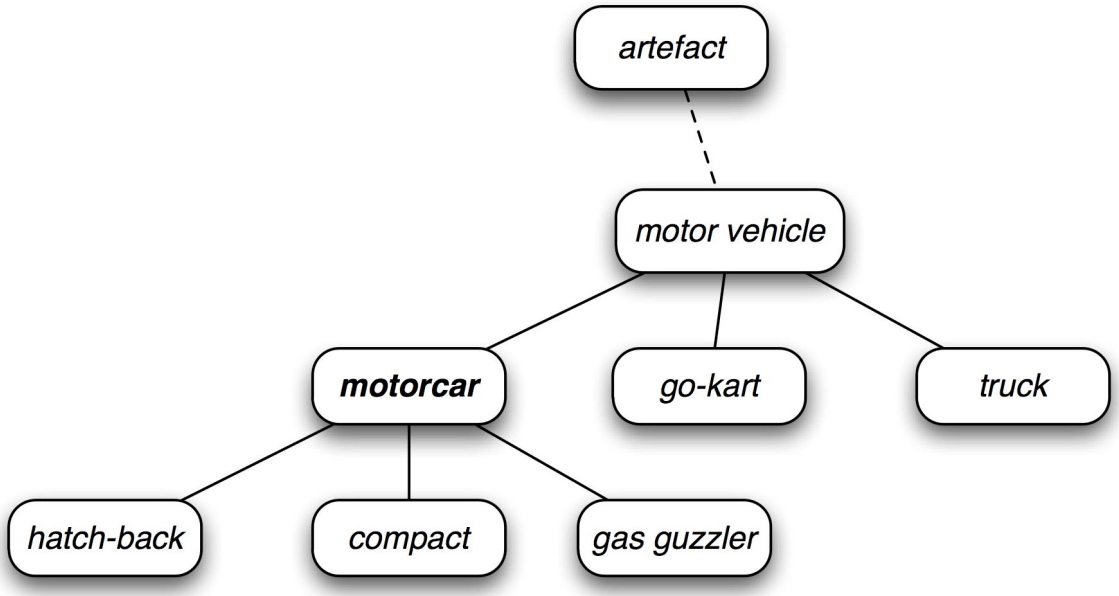
- Storing knowledge gained while solving one problem and applying it to a different but related problem.
- This paper investigates pre-training and fine-tuning.
  - Train a neural network for some task (ImageNet classification in this case).
  - Remove the bottom layer(s), add the appropriate layers for the transfer task.
  - Train on transfer task data but start with the pre-trained weights.

Related Work

- Pretraining hyperparameters and what layers should be used for transfer learning has been studied. In contrast, this work focuses on the effects of different pre-training data.
- Many unsupervised methods for initial training have been tried. ImageNet pretraining has always been found superior.

ImageNet

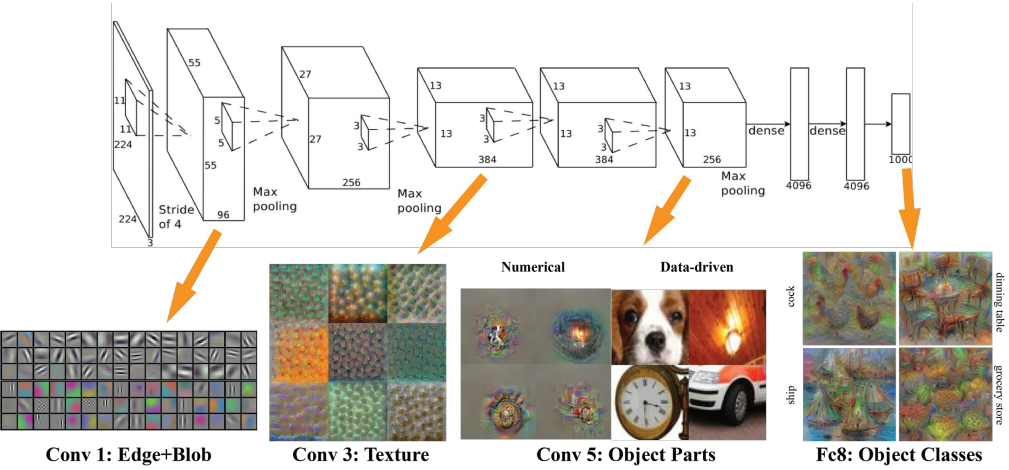
- Systems typically use 1,000 classes with >1,000 images each
- Classes belong to the WordNet hierarchy of Nouns.



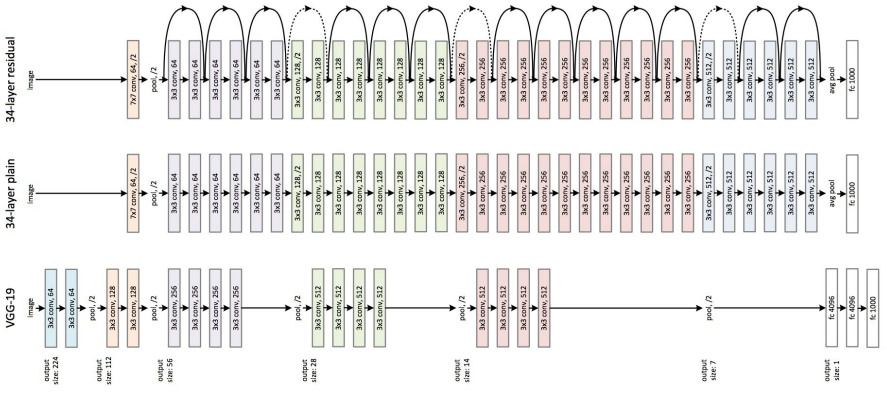
The Gold Standard in Representations

- It is extremely popular and effective to download a recent model pretrained with ImageNet features and fine tune the weights for a given domain. **Why**?

AlexNet



ResNet



Hypotheses

- It is the sheer size of the dataset (1.2 million labelled images) that forces the representation to be general.
- It is the large number (1000) of distinct object classes, which forces the network to learn a hierarchy of generalizable features.
- It is not just the large number of classes, but the fact that many of these classes are visually similar (e.g. many different breeds of dogs), turning this into a fine-grained recognition task, and therefore pushing the representation to work harder.

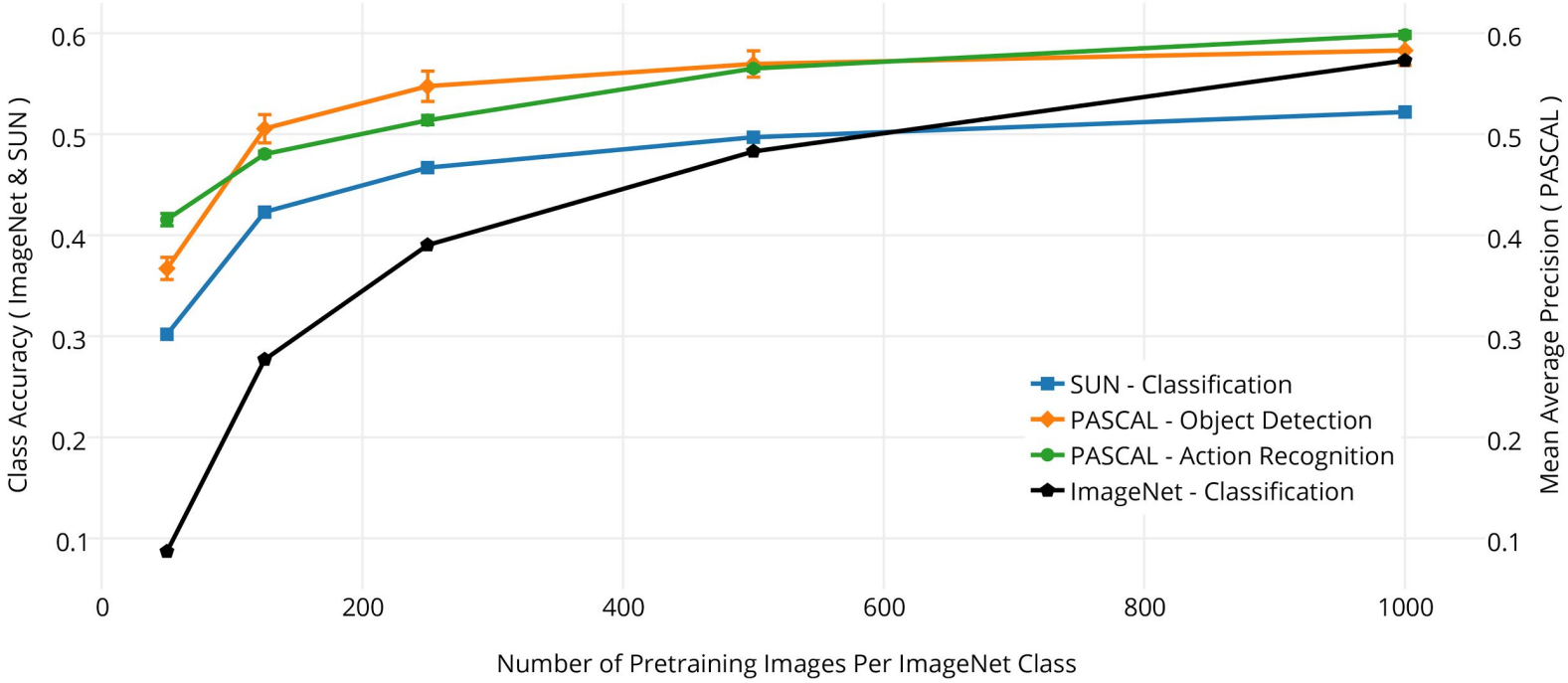
The transfer learning tasks

- PASCAL VOC 2007 object detection (PASCAL-DET)
- Scene classification on SUN dataset (SUN-CLS)
- PASCAL VOC 2012 action recognition (PASCAL-ACT-CLS)

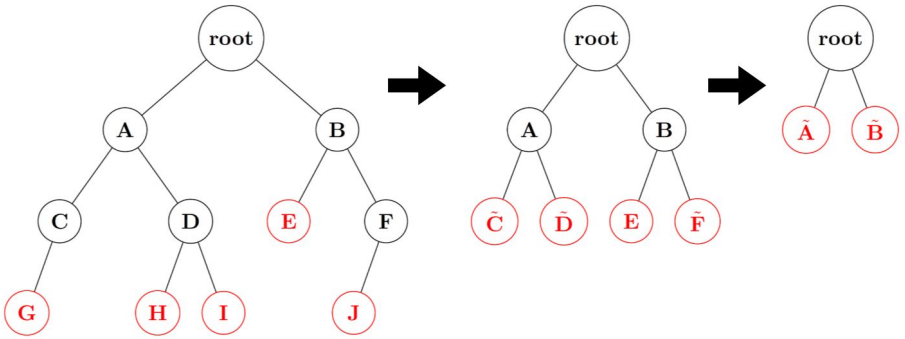
Investigations

1. How many pre-training ImageNet examples are sufficient for transfer learning?
2. How many pre-training ImageNet classes are sufficient for transfer learning?
3. How important is fine-grained recognition for learning good features for transfer learning?
4. Given the same budget of pre-training images, should we have more classes or more images per class?
5. Is more pretraining data always helpful?

1. Pre-training with fewer instances per class



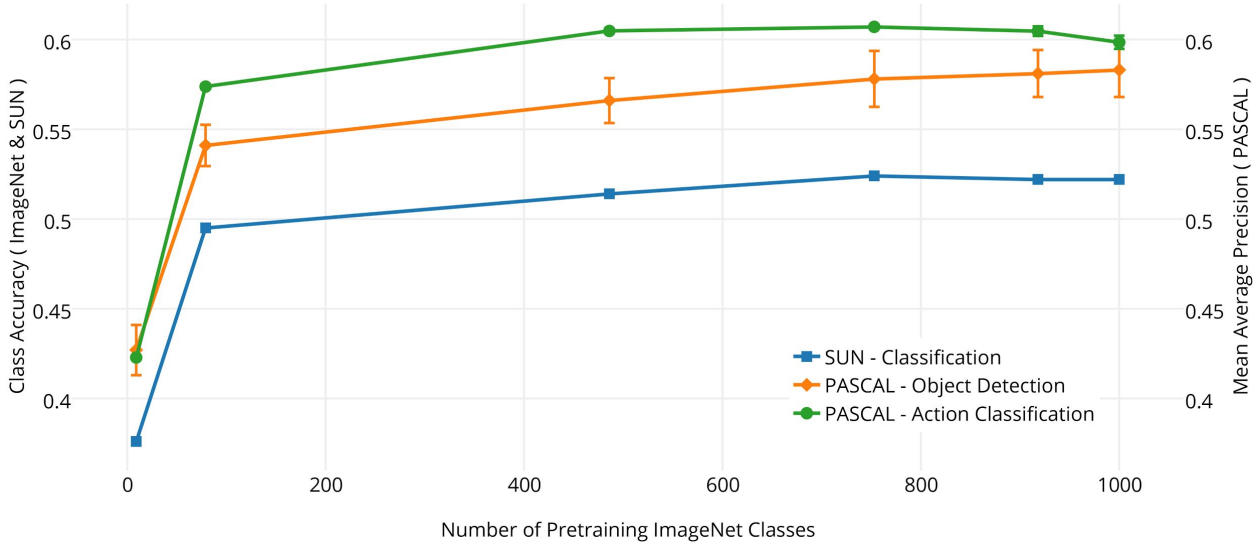
2. Pretraining with fewer classes



Original label set

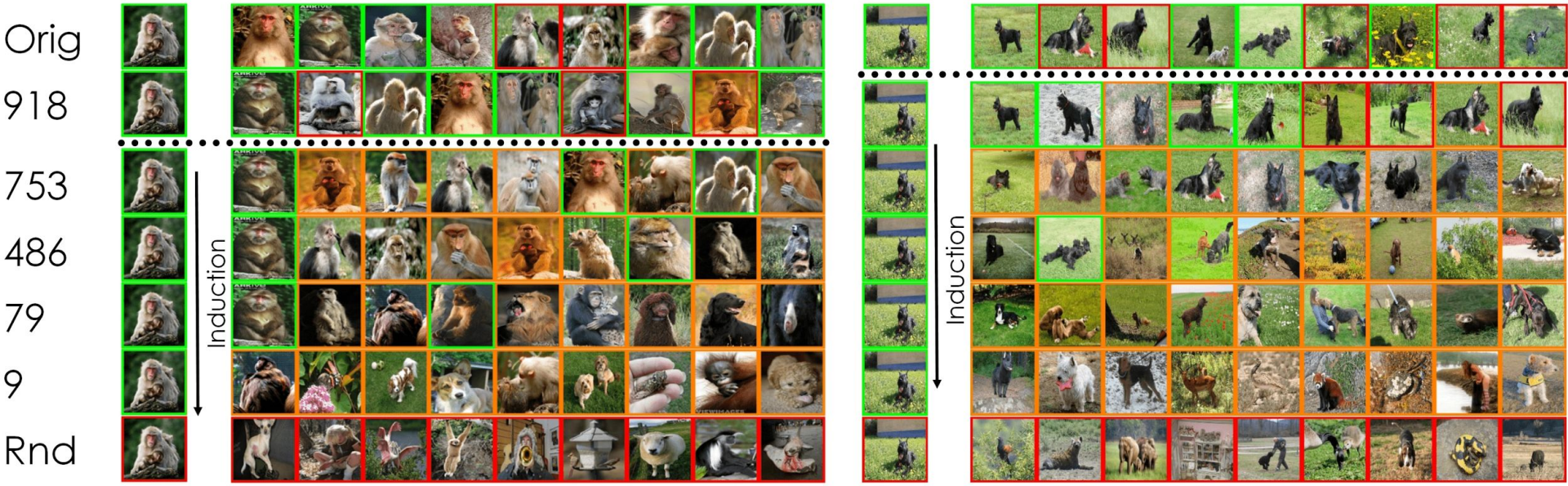
Label set 1

Label set 2



3. Induction Using Coarsely Trained Embeddings

- The previous experiment shows that fine grained classes aren't drastically superior.
- How good are the embeddings at nearest neighbor tasks?

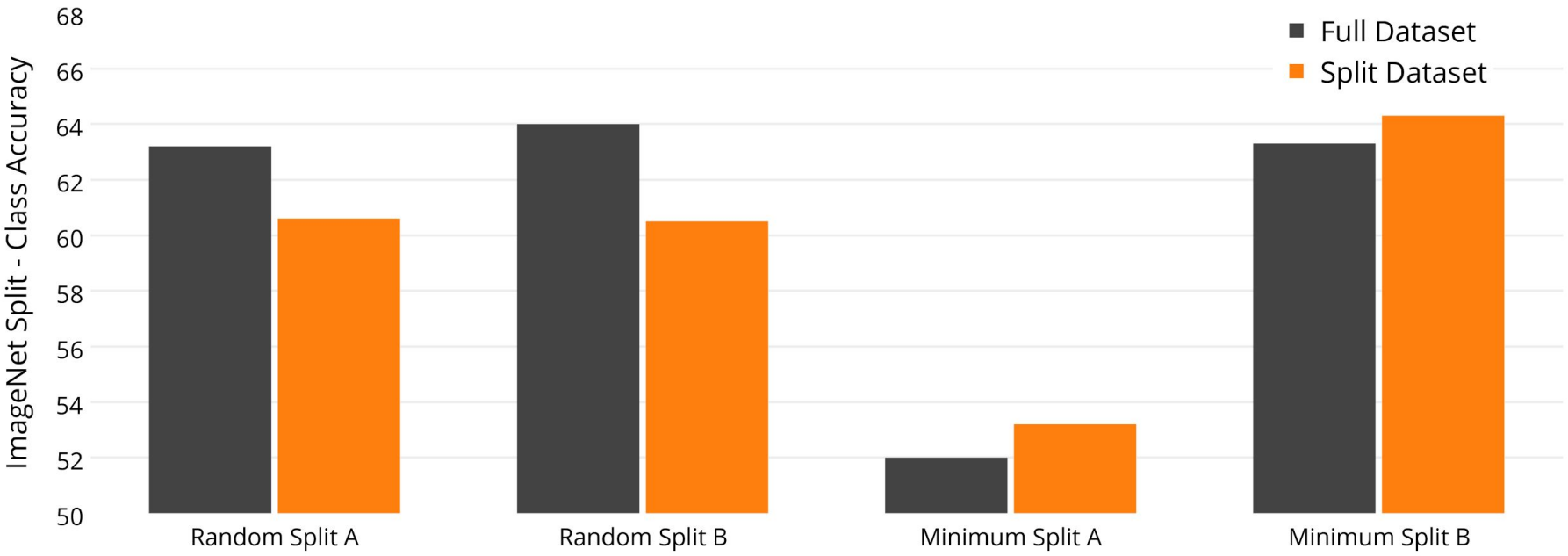


More Classes vs More Examples Per Class

Dataset	PASCAL			SUN		
Data size	500K	250K	125K	500K	250K	125K
More examples/class	57.1	54.8	50.6	50.6	45.7	42.2
More classes	57.0	52.5	49.8	49.7	46.7	42.3

Add pretraining data from non-target classes

If we want to detect types of dogs does it help to add something unrelated like fire trucks?



Hypotheses (revisited)

- It is the sheer size of the dataset (1.2 million labelled images) that forces the representation to be general.
  - No, pretraining with half the data does almost as well on the transfer task.
- It is the large number (1000) of distinct object classes, which forces the network to learn a hierarchy of generalizable features.
  - No, pretraining with data from half the classes does almost as well on the transfer task.
- It is not just the large number of classes, but the fact that many of these classes are visually similar (e.g. many different breeds of dogs), turning this into a fine-grained recognition task, and therefore pushing the representation to work harder.
  - No, transferable features are learned even when classes are very visually distinct.

Take home

- We know that we should use at least 500k images and at least 127 classes
- It will probably work well to skip unrelated classes.
- We also know that labeled pretraining seems to outperform other methods.

Issues (my commentary)

- AlexNet is old and not competitive. It's not clear that lessons learned will generalize to newer architectures.
- The transfer learning tasks are all quite similar to object classification. I would have liked to see results for a language task or a 3d task.
- They don't test against any other pre-training technique or dataset (labeled or unlabeled).
- No one variable seems to matter that much, they each contribute a very small amount, but baselines are always good.