# Distinguishing Unsupervised Feature Learning

**Lina Zhu**

**27 May 2018**

## 1   Introduction

The rise of deep neural networks, especially the Convolutional Neural Network (CNN) has led to several breakthroughs in computer vision benchmarking. The most successful models are trained through supervised learning, which requires large data sets to be fully noted as specific tasks. However, getting annotation data is often very expensive or in some cases even not feasible. In recent years, unsupervised learning has gained increasing attention to the community [1]. Our new approach to unsupervised learning stems from some observational object recognition about the results of supervised learning. As shown in Figure 1, the leopard-class image was rated much higher than the class Jaguar rather than the class bookcase. Such observations reveal that a typical discriminatory learning method can automatically discover the apparent similarity category between semantics without a clear guidance to do so. In other words, the obvious similarity is

not a comment learned from semantics but comes from the visual data itself. We put class supervision on extreme case supervision and ask: Can we learn a meaningful measure that reflects the apparent similarity between instances through purely discriminatory learning? An image is unique in its own right, each of which can significantly differ from other images in the same semantic category. If we learn to distinguish the concepts of any semantic category without an individual, we may end up with a representation that captures the apparent similarity between instances, just as classification-supervised learning still maintains a clear similarity in categories. This formulation of unsupervised learning as an instance-level discrimination is also technically attractive because it can benefit from the latest benefits of discriminatory supervisory learning, such as in the new network architecture. However, now we also face a major challenge. The number of "classes" is the size of the entire training set. In order to assess the effectiveness of unsupervised learning, past work has relied on linear classifiers such as support vector machines (SVMs). Functional classes linked to learning are classified during testing. However, it is not yet clear why the learning function through training tasks can be separated linearly for unknown test tasks. We advocate the use of non-parametric methods and tests for both types of training. The functionality of each instance is stored in a separate repository, not the weight of the network. At test time, we performed classification using k-nearest neighbor (kNN) based on learning metrics.
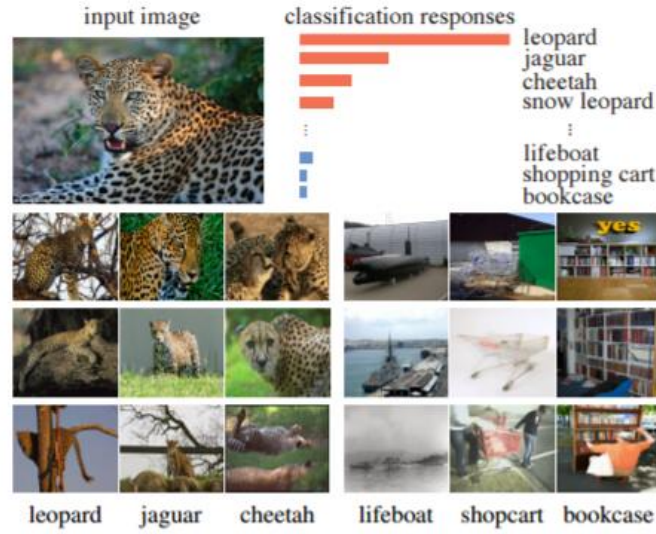
Figure 1: **Supervised learning results that motivate our unsupervised approach. For an image from class leopard, the classes that get highest responses from a trained neural net classifier are all visually correlated, e.g., jaguar and cheetah. It is not the semantic labeling, but the apparent similarity in the data themselves that brings some classes closer than others. Our unsupervised approach takes the class-wise supervision to the extreme and learns a feature representation that discriminates among individual instances.**

# 2  Image classification

We learn feature representations on ImageNet and compare our methods with representative unsupervised learning methods. We select the design parameters through empirical verification, compare our methods with random initialization networks (as lower bounds) and various unsupervised learning methods, including self-supervised learning [2]. Table 1 shows that using AlexNet and the midline linear classification function, our method achieves 35.6% accuracy, surpassing all baselines, including the most advanced technologies.

Table 1:  **Top-1 classification accuracy on ImageNet.**

| method | conv1 | conv2 | conv3 | conv4 | conv5 | kNN | dim |
|--------|-------|-------|-------|-------|-------|-----|-----|
| Random | 11.6 | 17.1 | 16.9 | 16.3 | 14.1 | 3.5 | 10K |
| Data-Init | 17.5 | 23.0 | 24.5 | 23.2 | 20.6 | - | 10K |
| Context | 16.2 | 23.3 | 30.2 | 31.7 | 29.6 | - | 10K |
| Adversarial | 17.7 | 24.5 | 31.0 | 29.9 | 28.0 | - | 10K |
| Color | 13.1 | 24.8 | 31.0 | 32.6 | 31.8 | - | 10K |
| Jigsaw [3] | 19.2 | 30.1 | 34.7 | 33.9 | 28.3 | - | 10K |
| Count | 18.0 | 30.6 | 34.3 | 32.5 | 25.7 | - | 10K |
| SplitBrain | 17.7 | 29.3 | 35.4 | 35.2 | 32.8 | 11.8 | 10K |
| Exemplar | | | 31.5 | | | - | 4.5K |
| Ours Alexnet | 16.8 | 26.5 | 31.8 | 34.1 | 35.6 | 31.3 | 128 |
| Ours VGG16 | 16.5 | 21.4 | 27.6 | 35.1 | 39.2 | 33.9 | 128 |
| Ours Resnet18 | 16.0 | 19.9 | 29.8 | 39.0 | 44.5 | 41.0 | 128 |
| Ours Resnet50 | 15.3 | 18.8 | 24.9 | 40.6 | 54.0 | 46.5 | 128 |

# 3  Summary

We propose an unsupervised feature learning method by using a new nonparametric maximization example to distinguish between softmax formulations. This is apparent from the observation that the result of supervised learning inspired by the observation is similar to the image. Our experimental results show that our method surpasses the latest technology levels ImageNet and Places in image classification, and the compact 128-dimensional representation can be extended with more data and deeper networks. It also provides competitive comprehensive results for semi-supervised learning and object detection tasks.

# References

[1] C. Doersch and A. Gupta. Unsupervised visual representation learning by context prediction. In *ICCV*, 2015.

[2] X. Wang and A. Gupta. Unsupervised learning of visual representations using videos. In *ICCV*, 2015.

[3] M. Noroozi and P. Favaro. Unsupervised learning of visual representations by solving jigsaw puzzles. In *ECCV*, 2016.