Sparse Coding on Stereo Video for Object Detection

Sheng Y. Lundquist
Portland State University
New Mexico Consortium
shenglundquist@gmail.com

Melanie Mitchell
Portland State University
Santa Fe Institute

Garrett T. Kenyon
Los Alamos National Laboratory
New Mexico Consortium

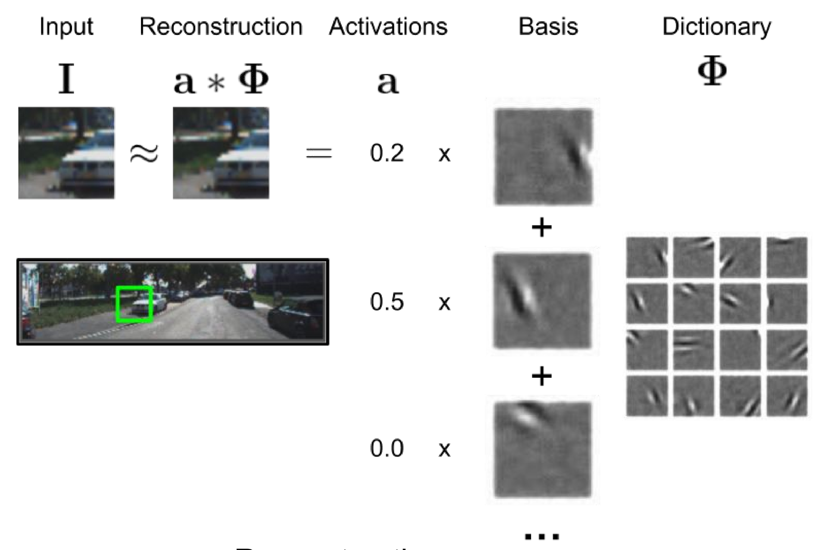
Introduction

Problem: Deep convolutional neural networks (DCNN) perform well at object detection, but require millions of labeled training examples.

Contribution: Given only limited stereo-video data, we show that adding an unsupervised sparse-coding layer to a DCNN improves object-detection performance as compared to fully supervised DCNNs. Additionally, the network that incorporates the sparse-coding achieves more consistent performance than the fully supervised DCNN.

Task: Detect cars using KITTI dataset of 7000 stereo video frames with bounding box labels (Greiger 2012).

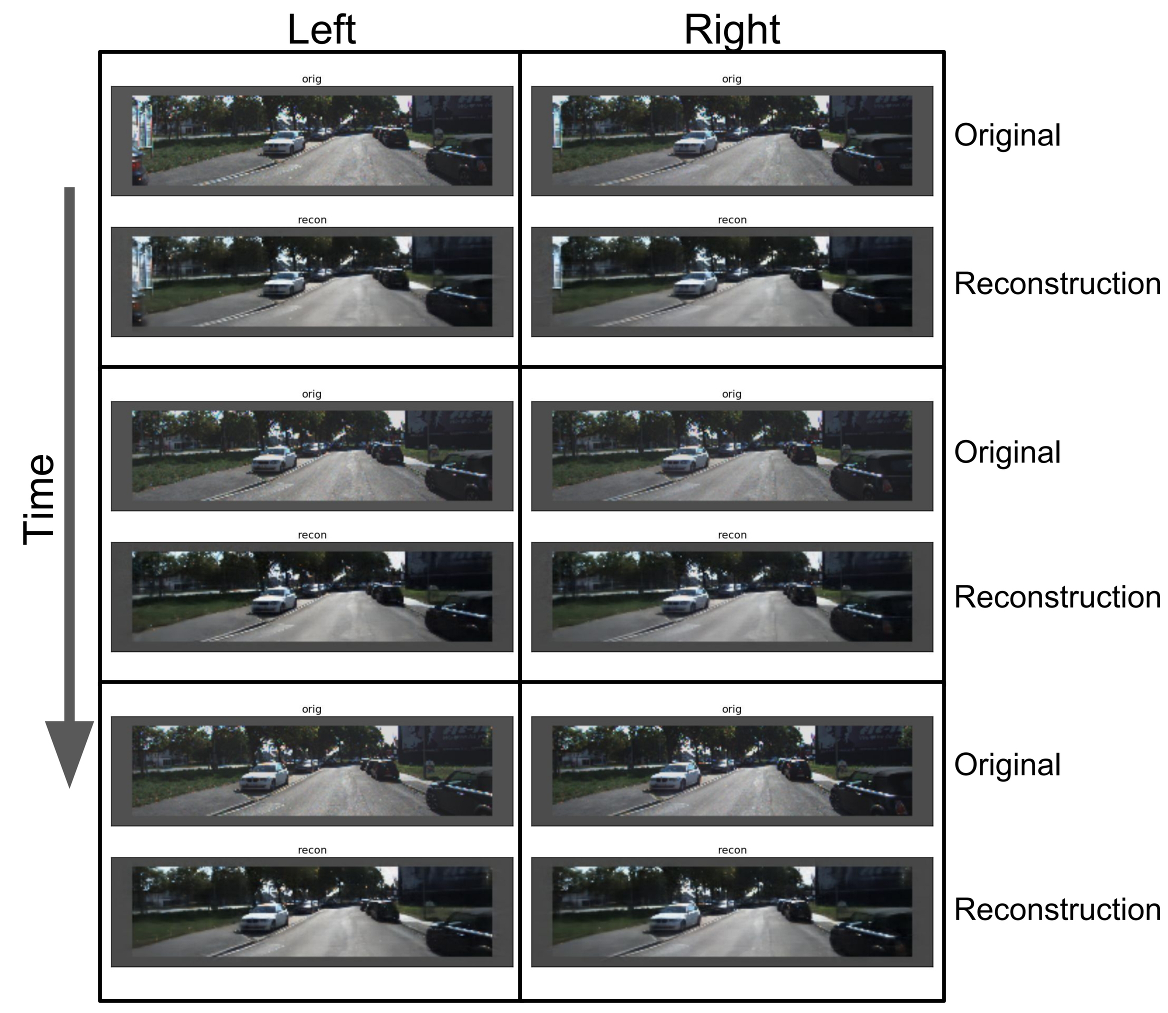
Sparse Coding



$$E = \underbrace{\|\mathbf{I} - \mathbf{a} * \mathbf{\Phi}\|_2^2}_{\text{Reconstruction Error}} + \lambda \underbrace{\|\mathbf{a}\|_1}_{\text{Sparsity}}$$

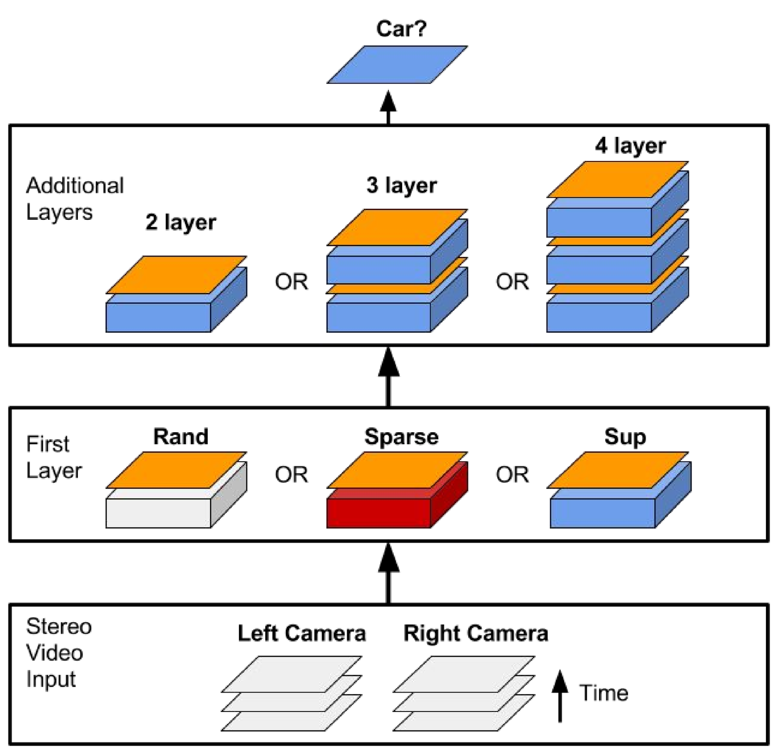
Sparse coding reconstructs a given input with a weighted linear combination of basis functions drawn from an overcomplete dictionary. Weighting coefficients are constrained to be sparse. The reconstruction is calculated via a 3-dimensional deconvolution (Zeiler 2010) on the time, height, and width axes.

Inputs and Reconstructions



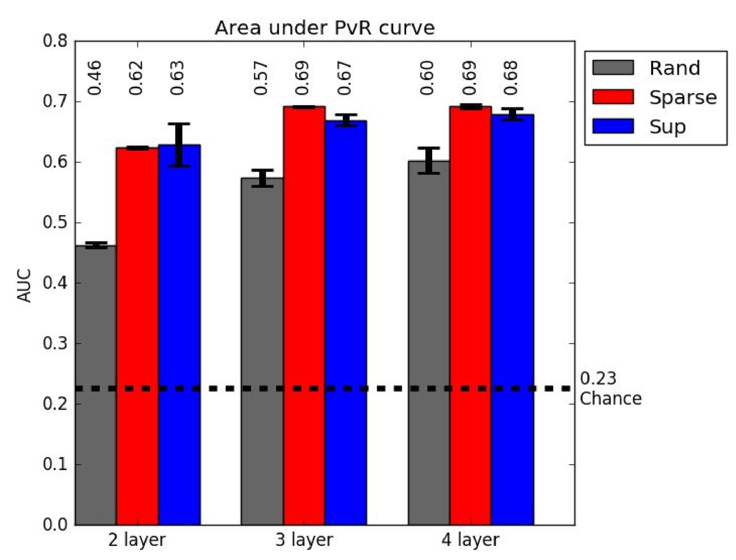
Sample stereo video frames and reconstructions from sparse coding.

Networks



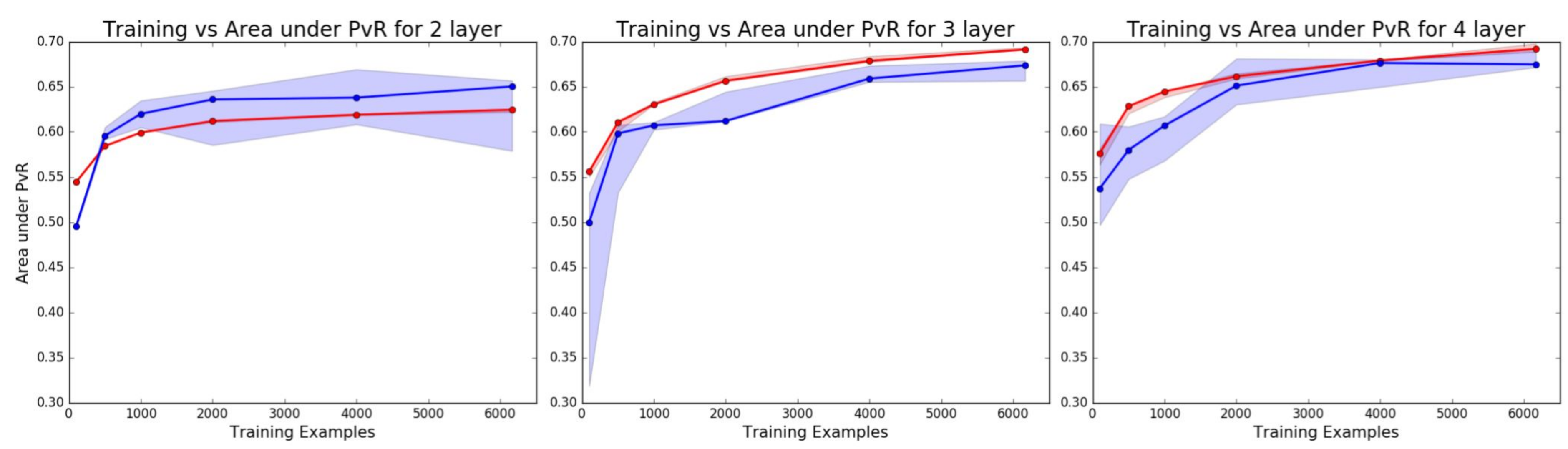
Network schematics for experiments. **Rand**: Convolution, random weights. **Sparse:** Sparse coding, offline unsupervised learning. **Sup**: Convolution, online supervised learning. We vary the total number of layers in the network. Red denotes unsupervised learning; blue denotes supervised learning; grey denotes no learning; orange denotes max pooling.

Sparse Coding Layer Outperforms Supervised Layer



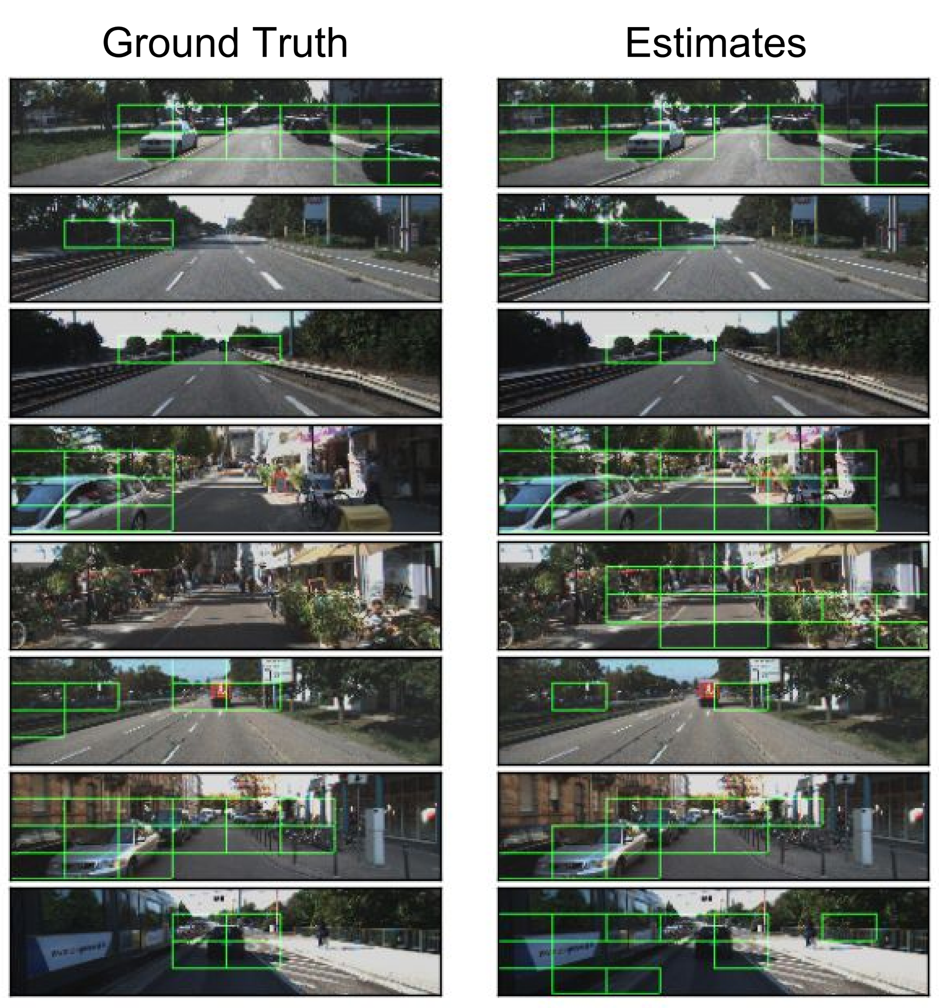
Area under precision vs recall curve for all models trained with all available training data. In models with 3 and 4 layers, a sparse coding layer results in higher performance. Additionally, **Sparse** performance has lower variance than that of the other two models. We find that all models outperform random chance.

Sparse Coding Layer Outperforms Supervised Layer with Less Labeled Training Data



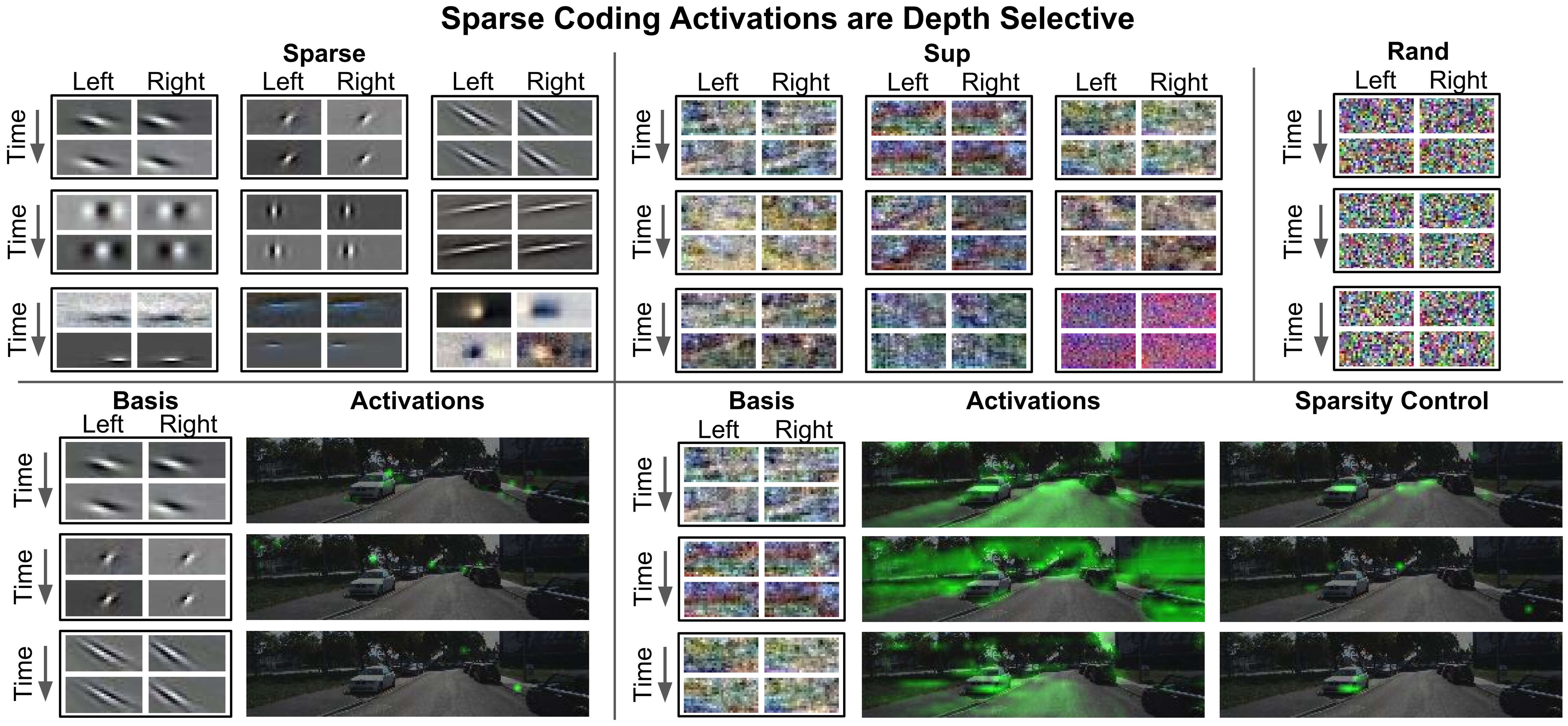
Performance versus number of labeled training examples provided to **Sparse** (red) and **Sup** (blue). The line denotes the median area under precision vs recall curve, with the area between the minimum and maximum performance filled in. We find that **Sparse** consistently outperforms **Sup**, and performs better with less provided training data. Additionally, **Sparse** achieves more consistent performance than **Sup**.

Detection Examples



Left: Ground truth with labeled boxes for cars. **Right:** Boxes are detection from 3-layer network with sparse-coding layer.

Sparse Coding Activations are Depth Selective



Top: Representative basis functions for **Sparse**, **Sup**, and **Rand**. **Bottom left**: Activations for **Sparse** over the input image. **Bottom right**: Activations for **Sup** over the input image. **Sparsity Control:** We threshold **Sup** activations to be, on average, equally sparse as **Sparse** model. This shows that **Sparse** model activations are more depth-selective than **Sup** activations, even when controlling for sparsity.

Research funded by DARPA Cooperative Agreement Award HR0011-13-2-0015