

# Understanding Video with Deep Convolutional Neural Networks

Victoria Dean, Carl Vondrick, Antonio Torralba (Electrical Engineering and Computer Science)  
Cisco Undergraduate Research and Innovation Scholar

## Computer Vision

Useful for many tasks, including pedestrian detection for self-driving cars

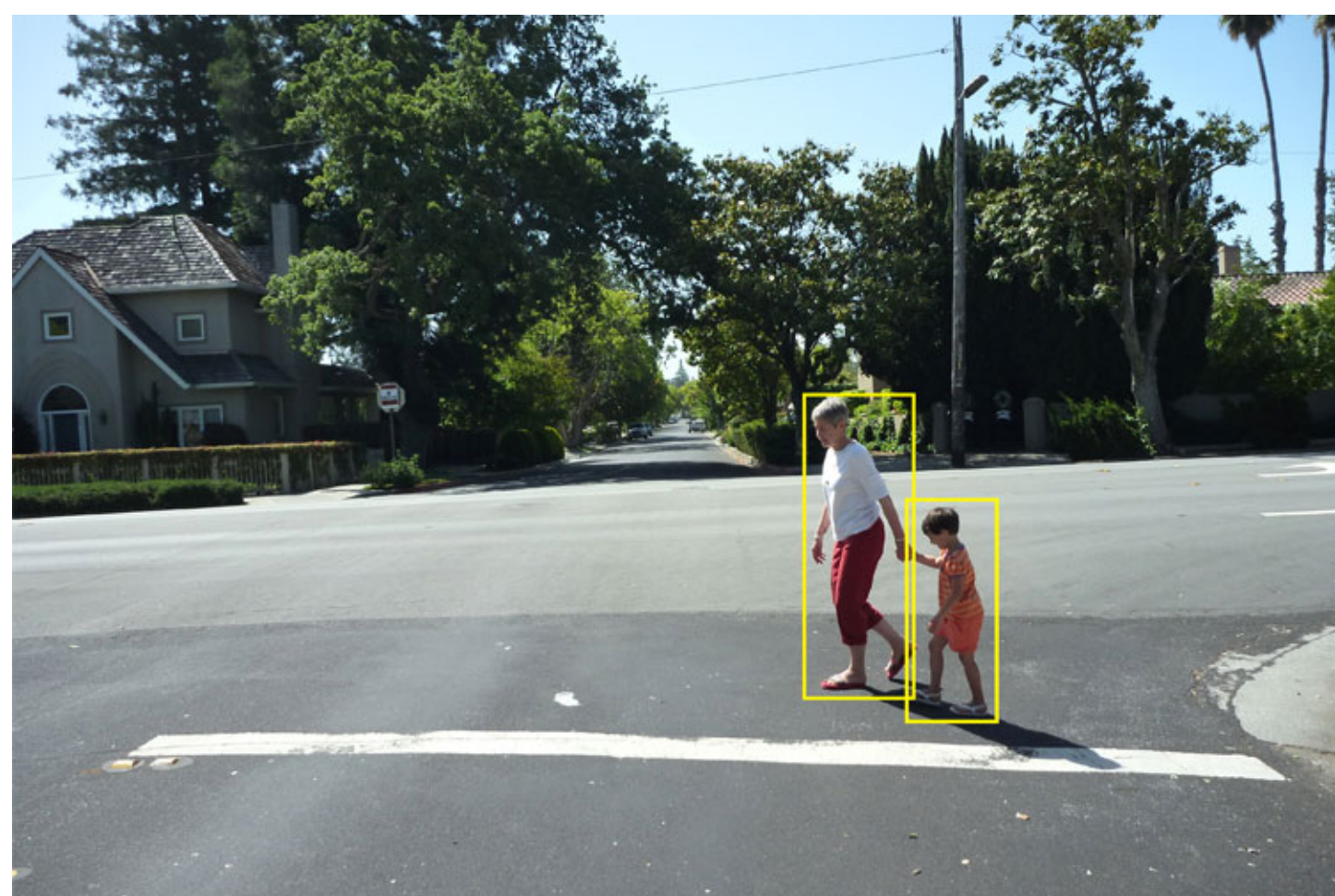
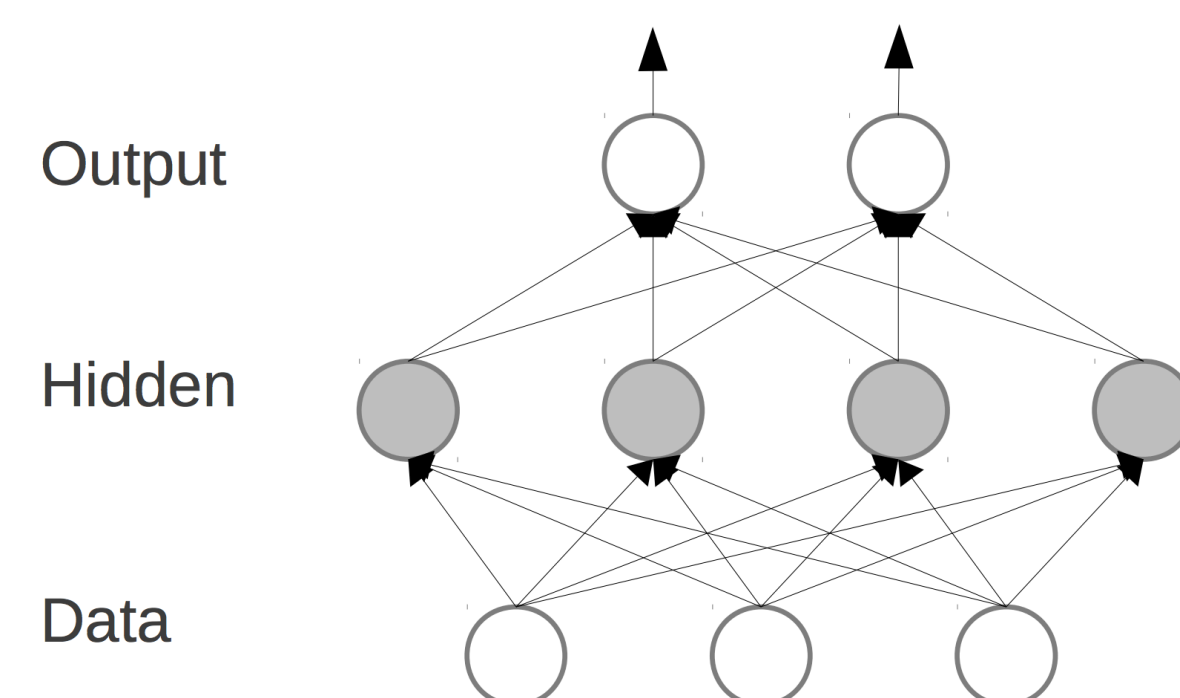


Photo from [1]

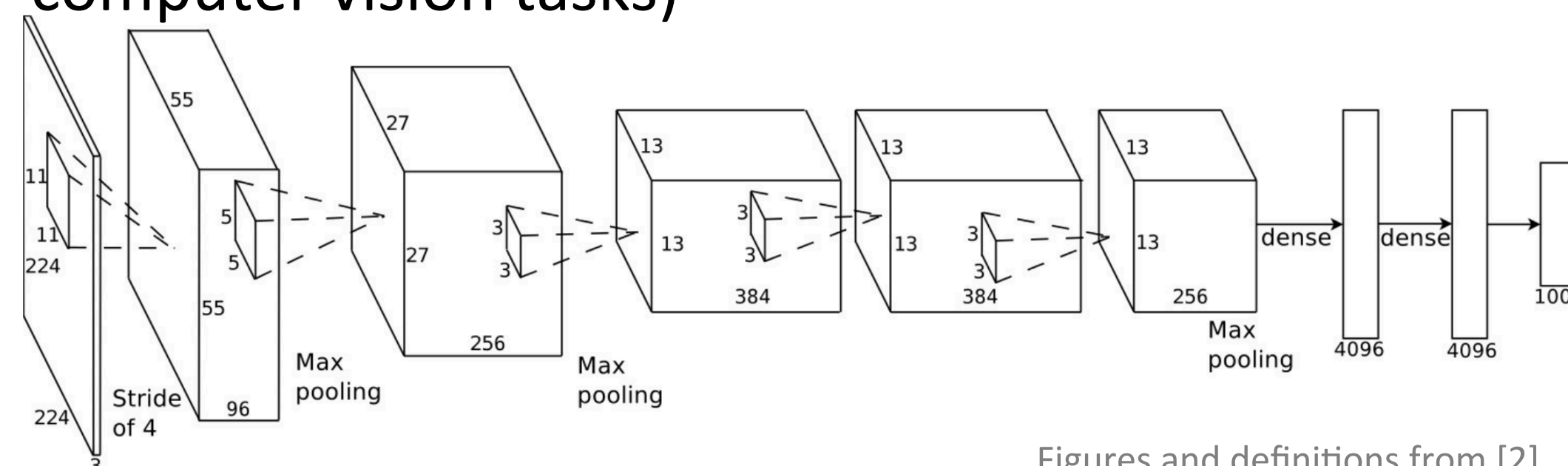
Much successful work on images, but less done on videos (labeling videos is expensive)

## Neural Networks

**Neural network** computes a differentiable function of its input

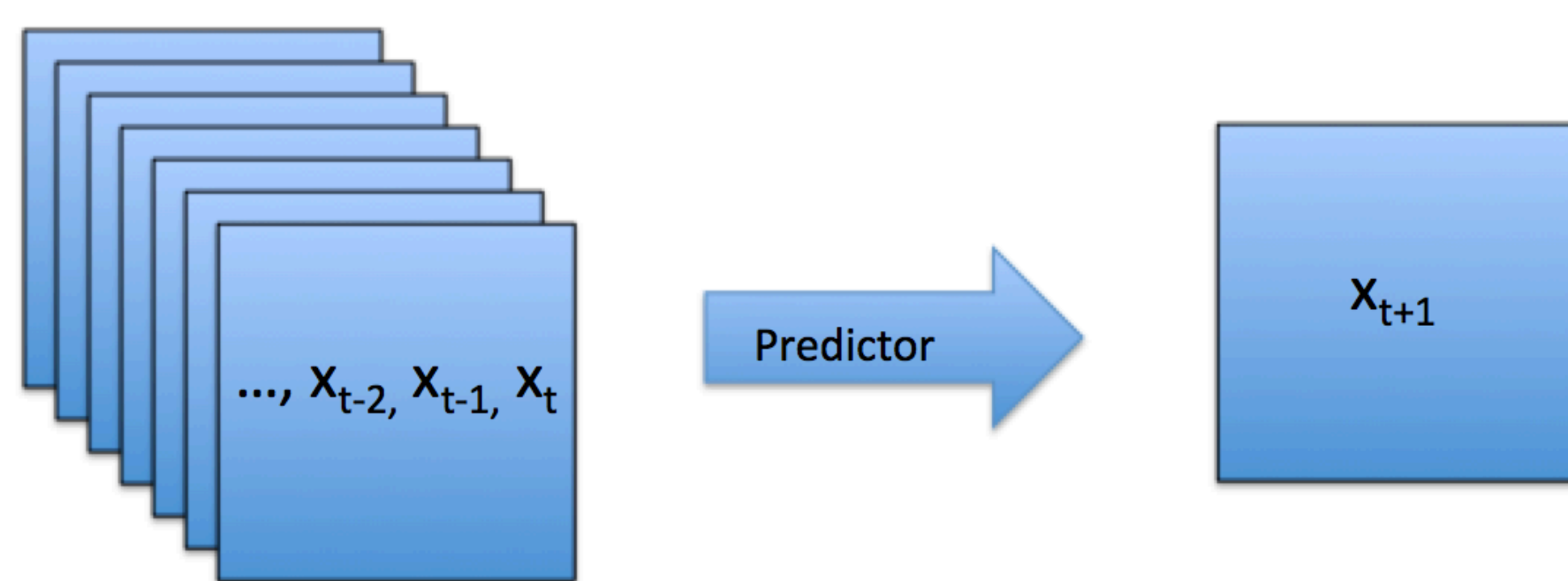


A **convolutional neural network (CNN)** applies the same function to each section of the input (good for computer vision tasks)



Figures and definitions from [2]

## Unsupervised Learning for Frame Prediction



Given past frames, predict future frame

- The previous frames are inputs to the predictor, which attempts to **generate the next frame  $x_{t+1}$**
- Cost function computes how close predicted frame pixels are to actual next frame pixels

## Current Frame Prediction Models

ground truth



our model



ground truth



our model



Figure from [3]

**Problem: Predictor tends to output mean frame**

## Current status

- Implementation of **simple prediction model** in Caffe, a popular open-source library for CNNs [4]
- Current model **generates future frame given one past frame**
- Simplified problem to **prediction of cropped frame** based on object detector bounding boxes [5]
- Considering cost functions that will hopefully improve issue with outputting mean frame

## Next Steps and References

- Try out different **network architectures** (number of layers, dimensions, etc.) and determine which is the most successful for prediction
- Experiment with different **cost functions**: do any yield predictions that are better than just mean?
- Implement **LSTM** (Long Short Term Memory) network for learning greater context in videos

### References

- [1] <http://www.nvidia.com/content/tegra/automotive/images/driver-assistance/pedestrian-detection-large.jpg>
- [2] <http://www.image-net.org/challenges/LSVRC/2012/supervision.pdf>
- [3] <http://arxiv.org/abs/1412.6604>
- [4] <http://caffe.berkeleyvision.org/>
- [5] <http://arxiv.org/abs/1311.2524>