New Machine Learning Program Recognizes Handguns in Even Low-Quality Video

Jingyue Shen

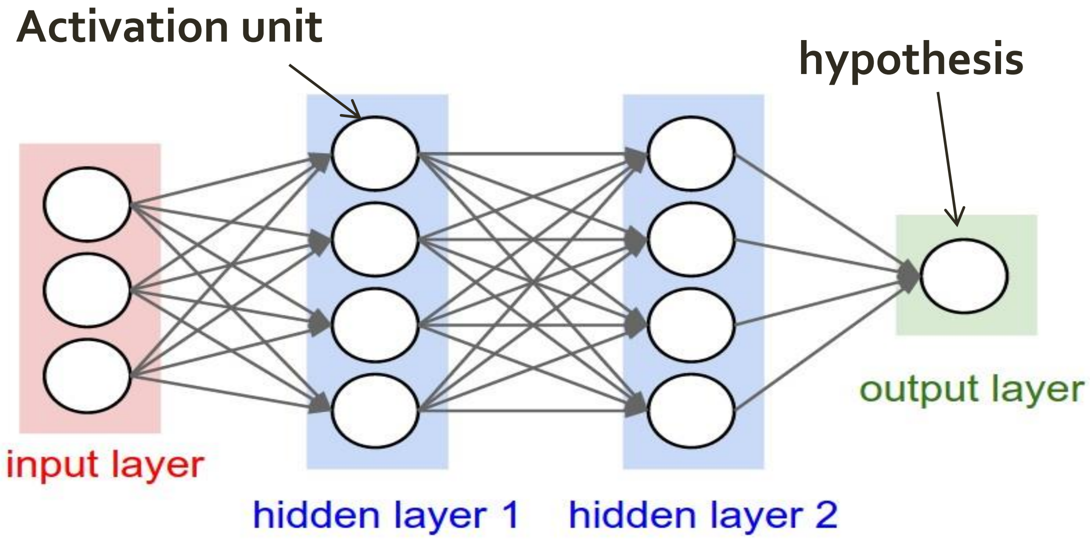
outcome

- Handgun detection alert system that can be used on surveillance or control camera
- capable of catching guns from even low-quality YouTube footage just under a quarter second

Keywords

- Deep Convolutional Neural Networks
- Transfer learning

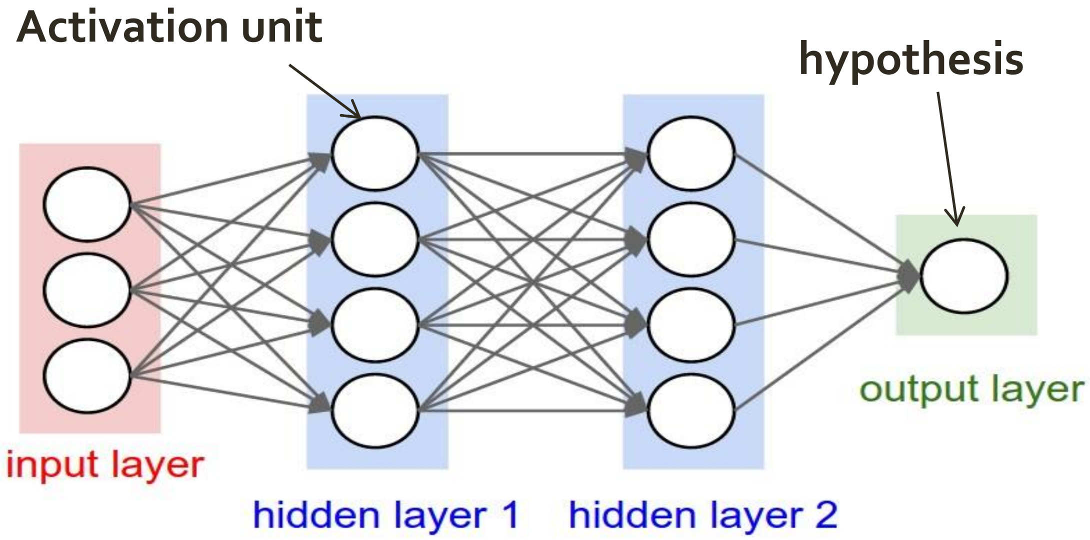
Neural Network



$$\begin{bmatrix} x_0 \\ x_1 \\ x_2 \end{bmatrix} \rightarrow [\quad] \rightarrow h_\theta(x)$$

- From one layer to another
  - Weights(θ)
  - Activation units($a_i^{(j)}$)
  - Sigmoid function($\frac{1}{1+e^{-\theta^T x}}$)
  - Resulting hypothesis function(h_θ (x))

Neural Network



$$\begin{bmatrix}x_0\\x_1\\x_2\end{bmatrix}\rightarrow[\quad]\rightarrow h_\theta(x)$$

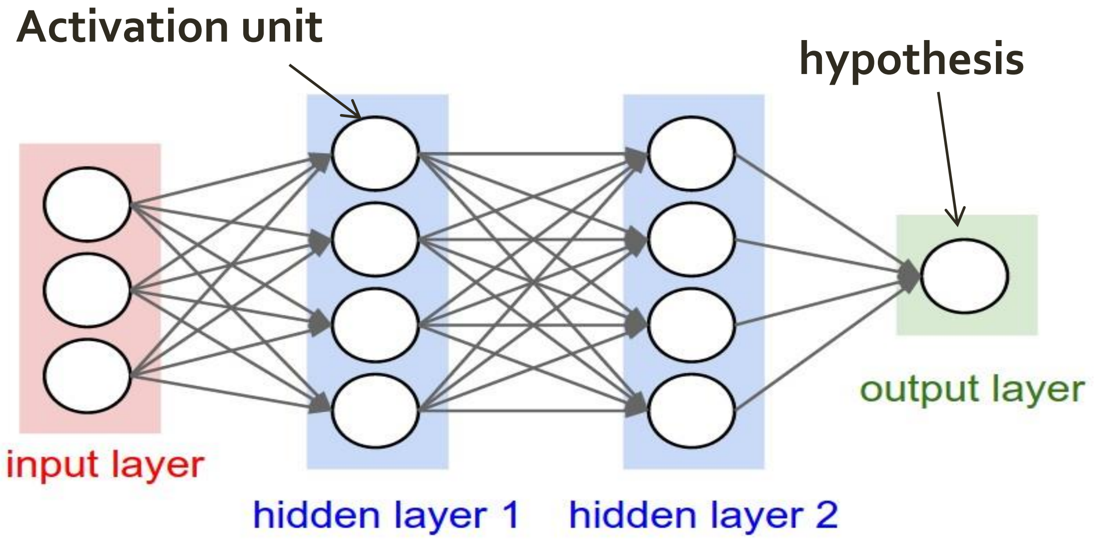
- Cost function

$$J(\Theta) = -\frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{K}\left[y_k^{(i)}\log((h_\Theta(x^{(i)}))_k) + (1-y_k^{(i)})\log(1-(h_\Theta(x^{(i)}))_k)\right] + \frac{\lambda}{2m}\sum_{l=1}^{L-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}(\Theta_{j,i}^{(l)})^2$$

- Calculate gradient
  - Backward propagation algorithm

(calculate "error term" for each node and use them to find gradient of J(θ))

Neural Network

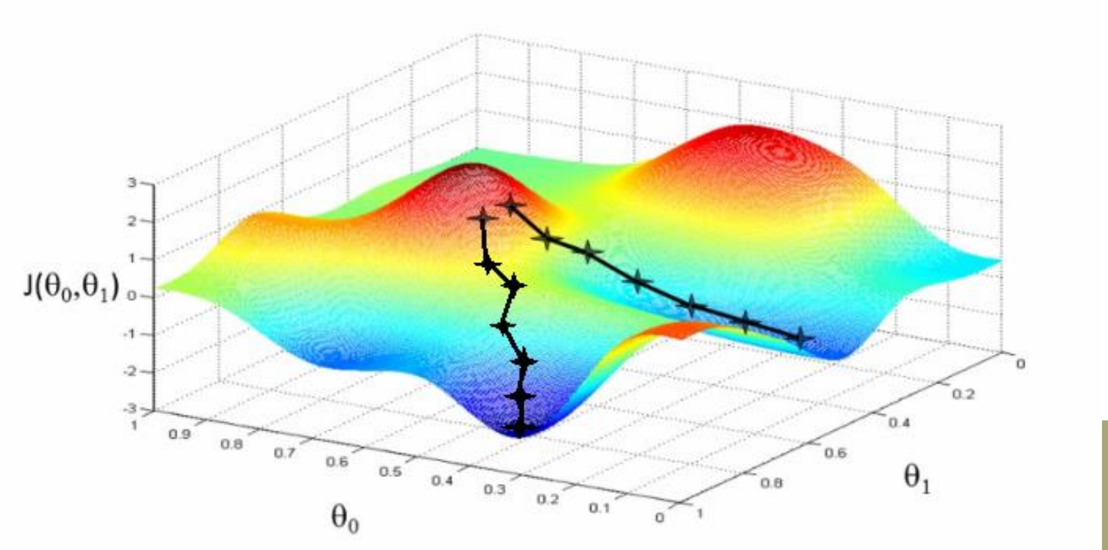


$$\begin{bmatrix} x_0 \\ x_1 \\ x_2 \end{bmatrix} \rightarrow [\quad] \rightarrow h_\theta(x)$$

- Update weights
  - Gradient descent

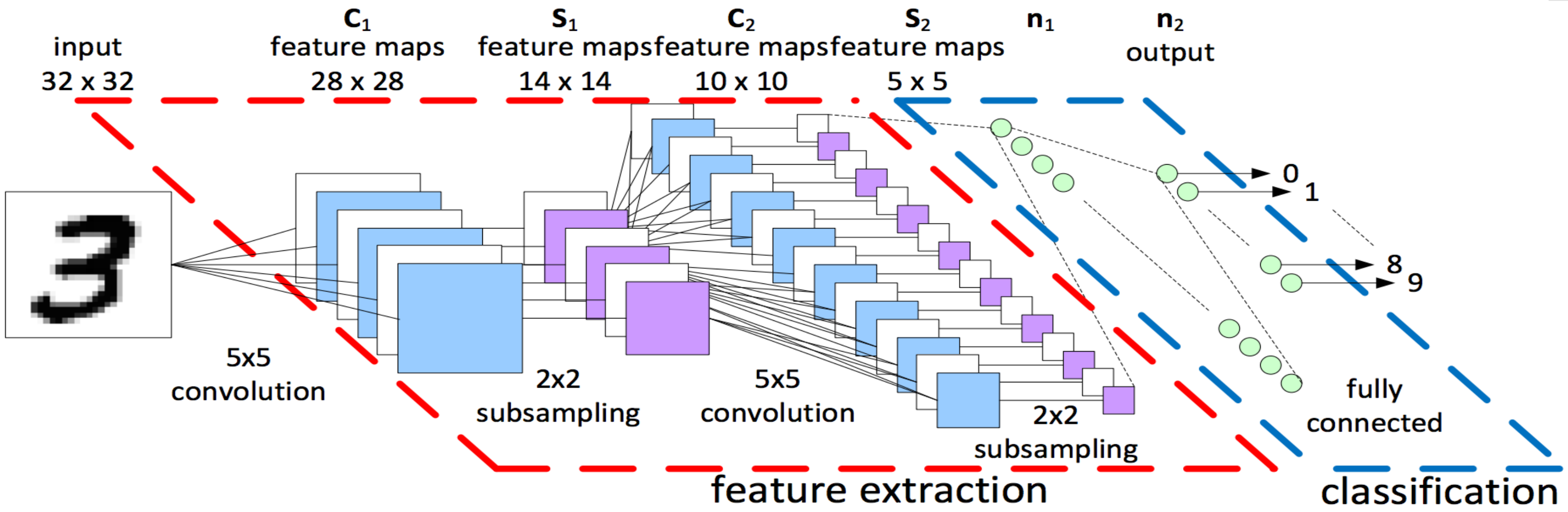
$$\Theta_t = \Theta_t - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

- Repeat all processes until cost is relatively low



Deep Convolutional Neural Network (CNNs)

- Far more complicated
  - Not fully connected
  - 3D volumes of neurons
  - ……
  - 3 main types of layers(convolutional layer, pooling layer, fully-connected layer) stack together
  - ……



Deep Convolutional Neural Network

But……

Deep Convolutional Neural Network

- Have similar general procedures
  - Propagate forward to calculate the hypothesis(score)
  - Propagate backward to calculate gradient
  - Use the gradient to update weights

Typical Solution & Challenges

- Convolutional Neural Networks(CNNs)
- Millions of training images

- Researchers only have 3000 images of pistols

(extracted from online gun catalogues, gun use tutorials and gun advertisement)

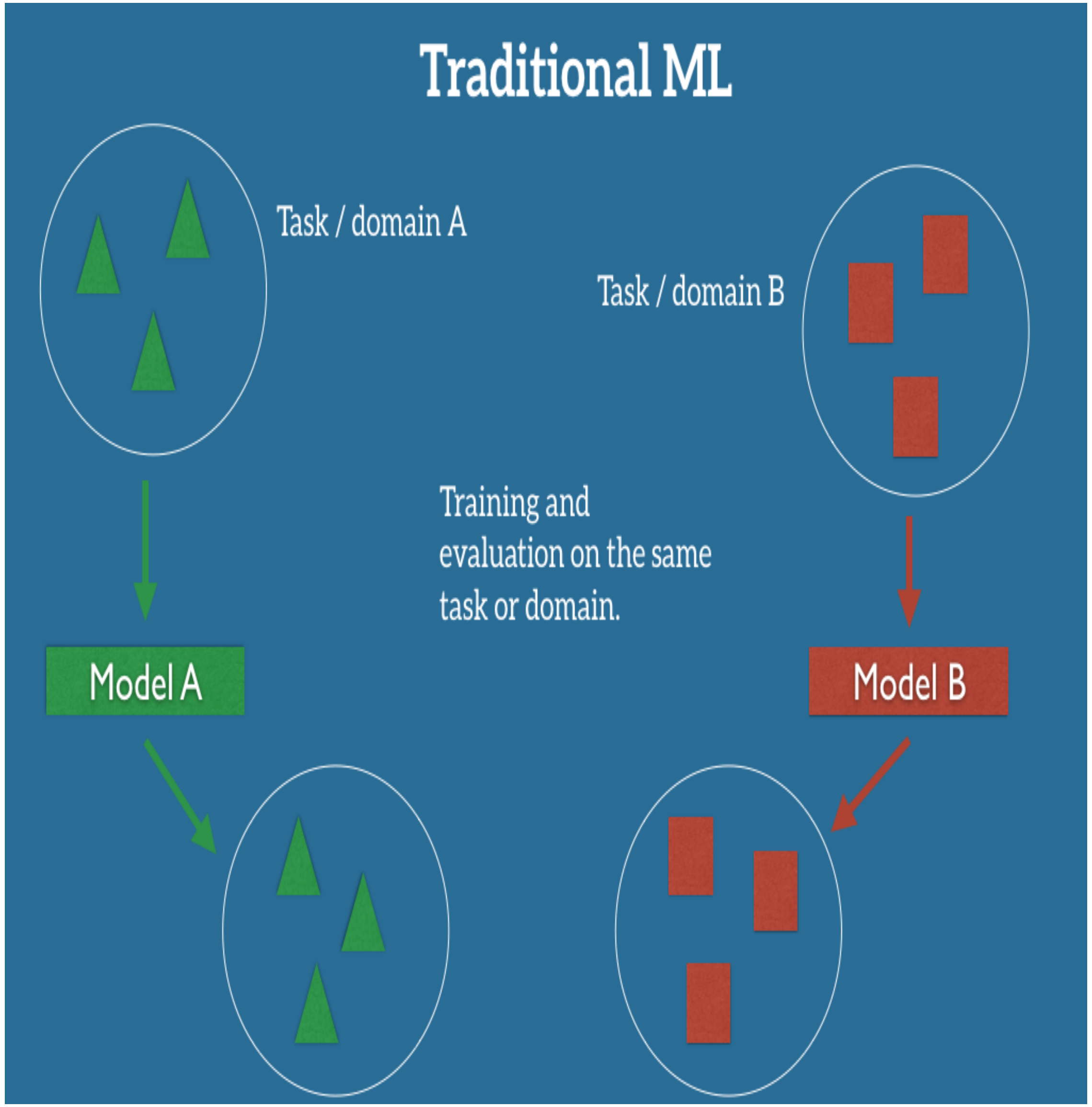
Solution

- Transfer Learning

  - creation of visual recognition models even when data is scarce

  - use knowledge about one category of thing and apply that to other related category.

(“fine tune” an existing related model)



Traditional ML

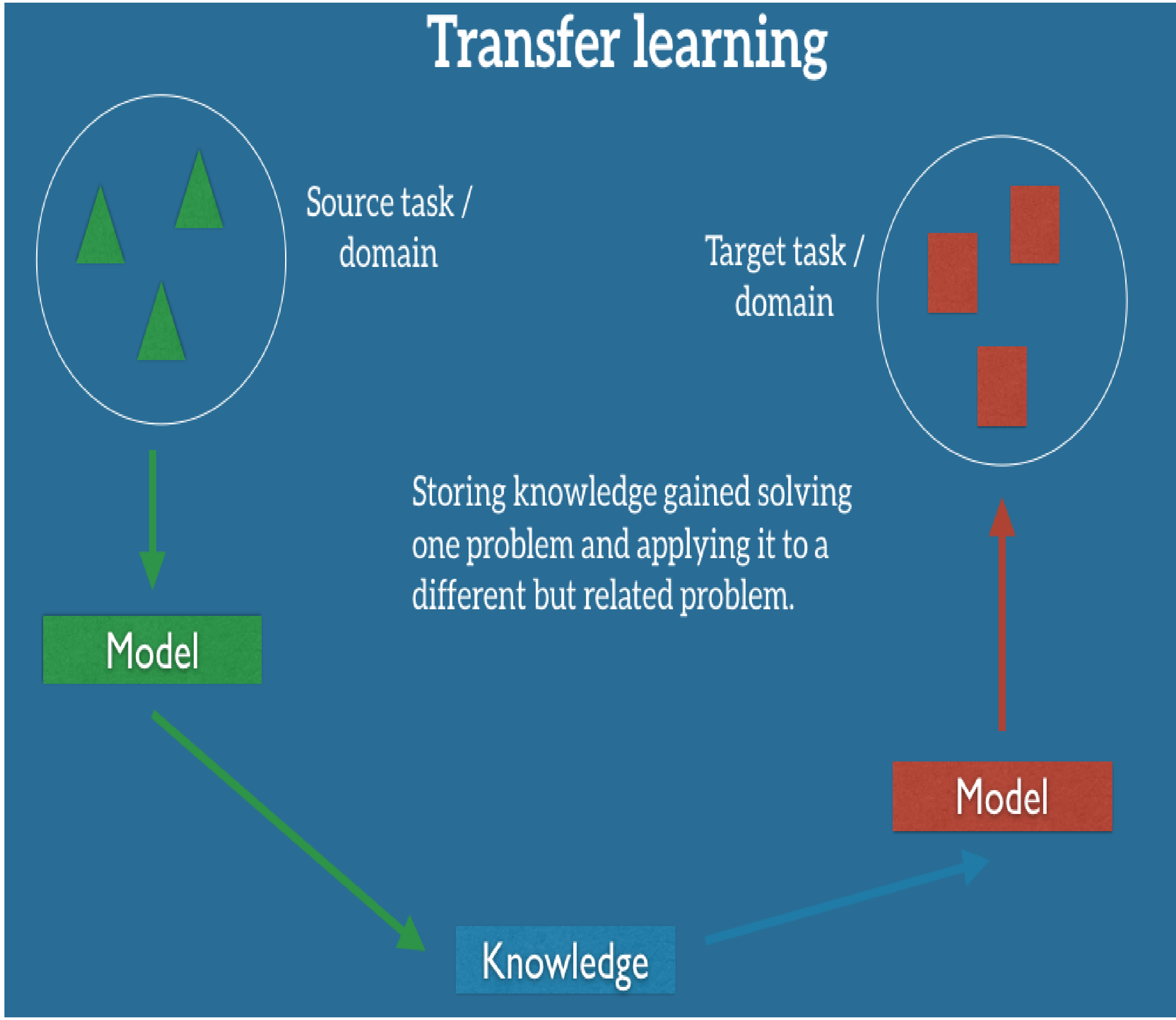
Task / domain A

Task / domain B

Training and evaluation on the same task or domain.

Model A

Model B



Transfer learning

Source task / domain

Target task / domain

Storing knowledge gained solving one problem and applying it to a different but related problem.

Model

Knowledge

Model

Classifier Model Used

- VGG-16
  - 144million parameters, trained based on a 1.28 million image dataset known as ImageNet

  - Can classify input images with respect to 1,000 different object classes

Training Classifier

- Training
  - Prediction loss minimization

Cost function

weight

weight decay

$$J(w) = \frac{1}{N}\sum_{i=1}^{N} L(f(w;x_i),y_i) + \lambda R(w)$$

#data instances

Loss function

Input images

output
class labels

  - Stochastic Gradient Descent

$$w_{t+1} = \mu w_t - \alpha \Delta J(w_t)$$

Whole Process

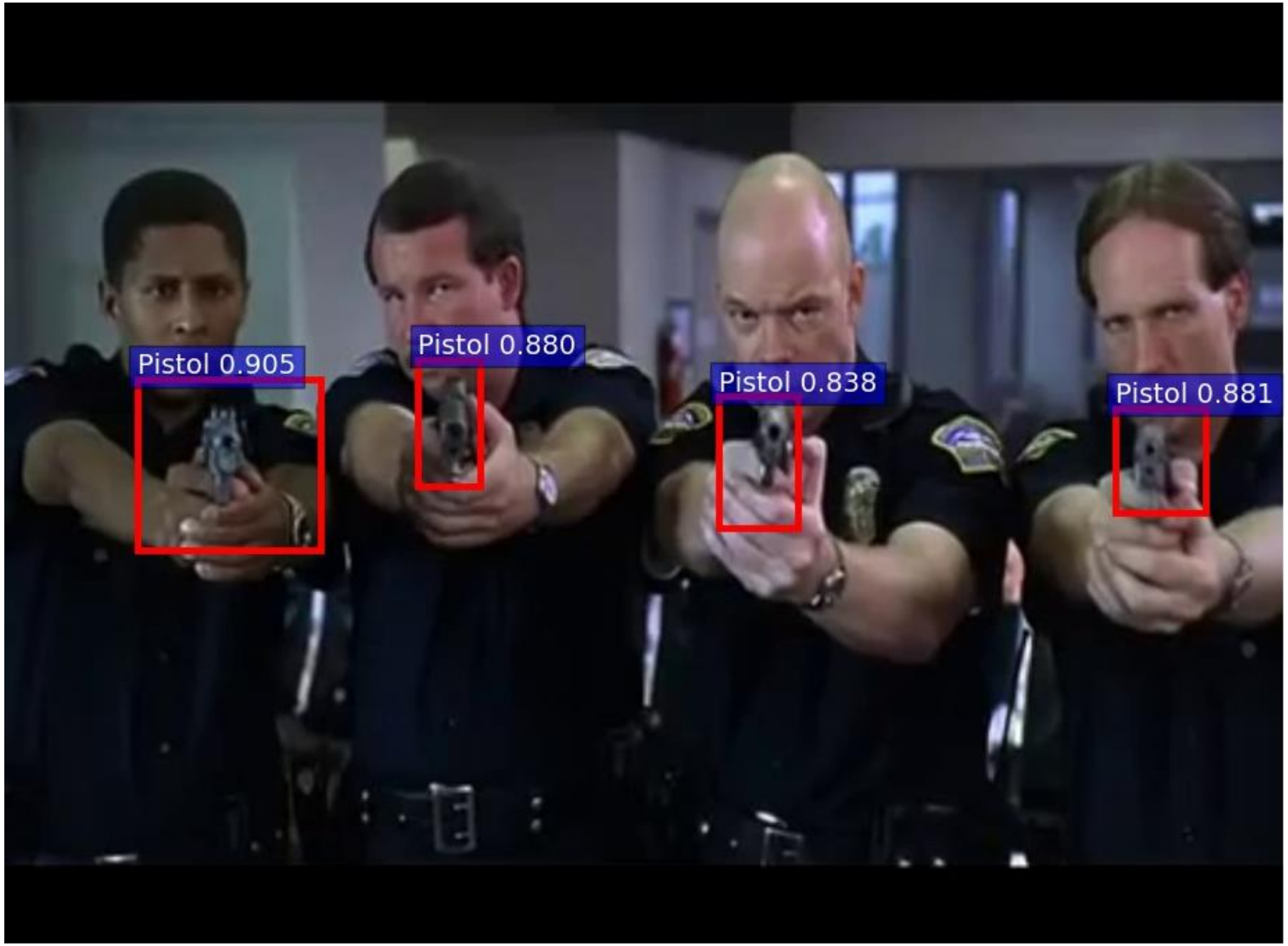
- Train VGG-16 classifiers based on manually-constructed different new datasets
- Object detection process
  - run it on a number of areas of the input image using
    - sliding window approach
    - region proposals approach
- Goal: minimize the number of false positives while reaching a near-real-time detection

Achievement

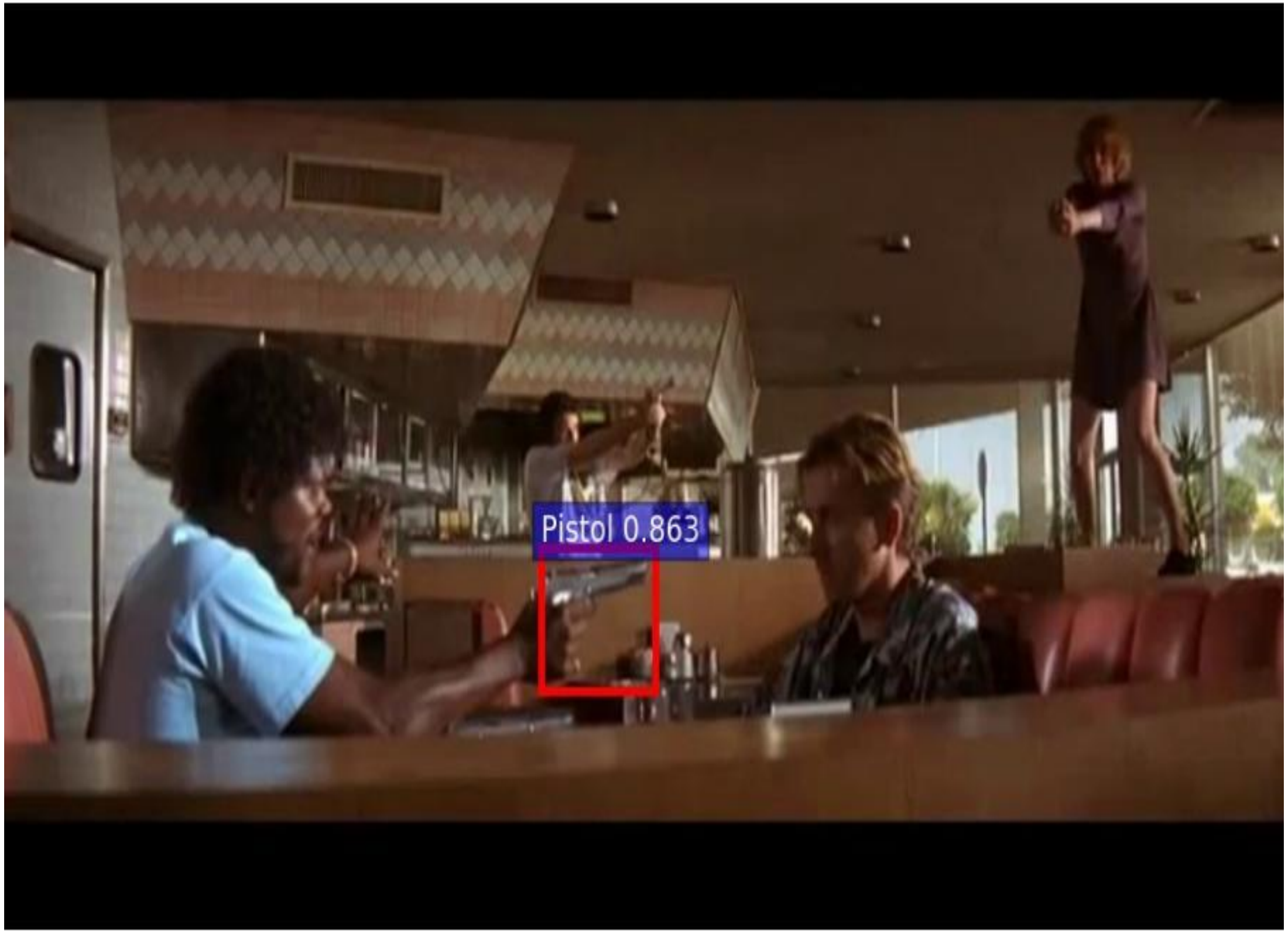
- Accuracy :between 90% and 95%
  - Eliminate almost all false positive
  - Has significant set of false negative due to very low contrast and luminosity

- Reaction time: can recognize pistols even in low-quality video under a quarter second

Example of correct prediction



Example of False Negative



Conclusion

- Fill in the gulf between ultra high-tech and crudely obvious approach.
  - NYPD testing a system that tracks guns based on the radiative signatures of human bodies
  - a firm called Shooter Detection Systems is pushing a system that automatically detects and reports gunfire: "a smoke alarm for gun fire detection."
- Though computationally taxing, doing pistol recognition is really quick

Reference

- http://cs231n.github.io/convolutional-networks/#overview
- https://www.coursera.org/learn/machine-learning/programming/AiHgN/neural-network-learning
- https://arxiv.org/pdf/1702.05147.pdf

Thank you!