Convolutional Neural Networks

Shrey Gupta

Applied Machine Learning (HOUSECS 59-01), Duke University

November 7, 2018

Problems

- Classification: —.
- Localization: —.
- Object detection: —.
- Landmark detection: —.

Classification

- —

Localization

- —

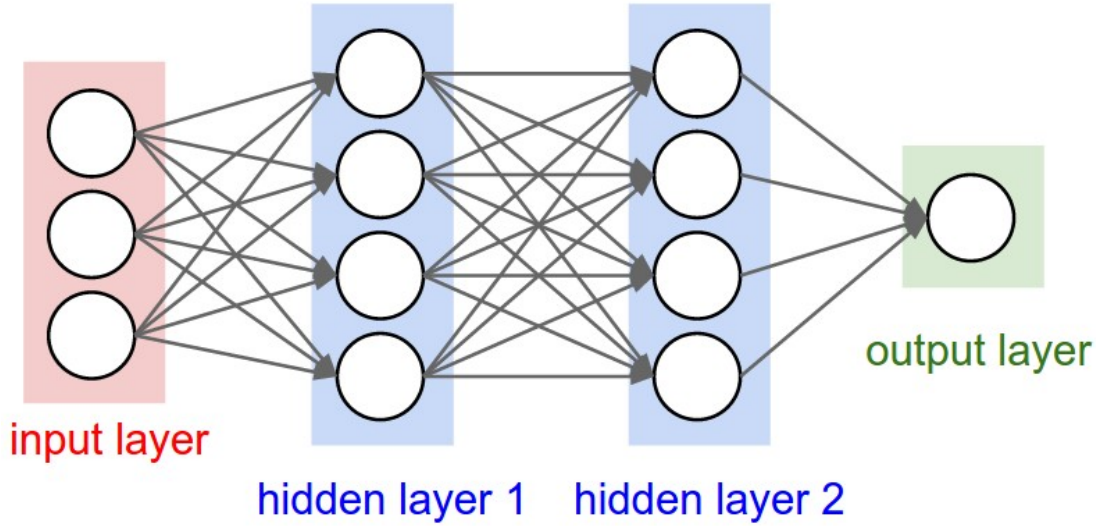
Object Detection

- —

Landmark Detection

- —

Recall



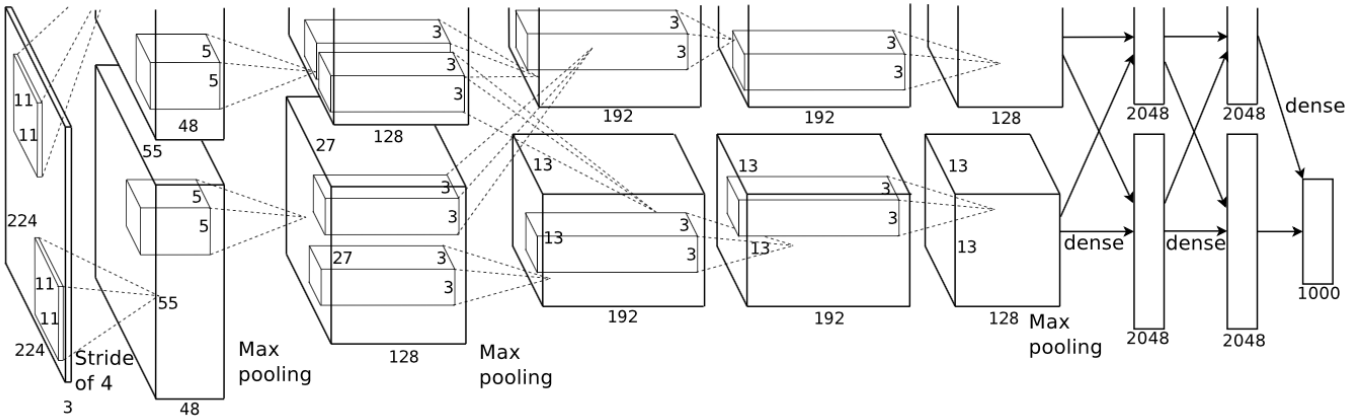
- Input layer: single vector of feature inputs.
- Hidden layer(s): sets of neurons with nonlinear activation, fully connected by weights (with biases) to other layers.
- Output layer: single vector of output scores.

Image source: Stanford University

Issues

- Doesn't scale well to large features, such as images.
- Speech and images are rich in structure (e.g. hierarchy). In the case of images, how can we utilize the structure to our advantage?
- Take advantage of the properties of images to create a new network.

Architecture Summary



- Architecture: convolutional layers, pooling layers, and fully-connected layers.
- Note: consider intensity (for black-and-white images) or RGB values over pixels as inputs.

Image source: Conference on Neural Information Processing Systems

Templates

- —

Convolutional Layers

- Recall: neurons match inputs to patterns.
- Create $w \times h$ templates for $w \times h$ receptive fields centered at (x, y) locations in the image, with stride s (usually small) along each dimension.
  - Each receptive field is hence *local*.
- Match the template to the field, and compute a score via the *dot product*.
- Construct a new image of the activation of the scores.

Stride

- —

Aside: Padding

- —

Volume

- —

Pooling Layers

- Reduces the number of parameters (and hence computations).
- Controls overfitting: intuitively, allows for translational invariance.
- Compute the maximum score for each $w \times h$ subimage across the image (overlap possible).
- Construct a new, smaller image of the maximum scores.
- Common variations of the sizes include 2×2 and 2×3 (with overlapping).

Fully-Connected Layers

- These layers are identical to the hidden layers as before: sets of neurons with nonlinear activation (usually ReLU), fully connected by weights (with biases) to other layers.
- Input from the final convolutional layer, consisting of transformed and smaller features.

Aside: Convolution

- Why are these called convolutional neural networks?
- Recall the convolutional layers in the network, which perform template-matching; with stride 1, this can be viewed as a convolution over the image.

Backpropagation

- Backpropagation is performed in a manner identical to fully-connected neural networks (i.e. as before).

Example

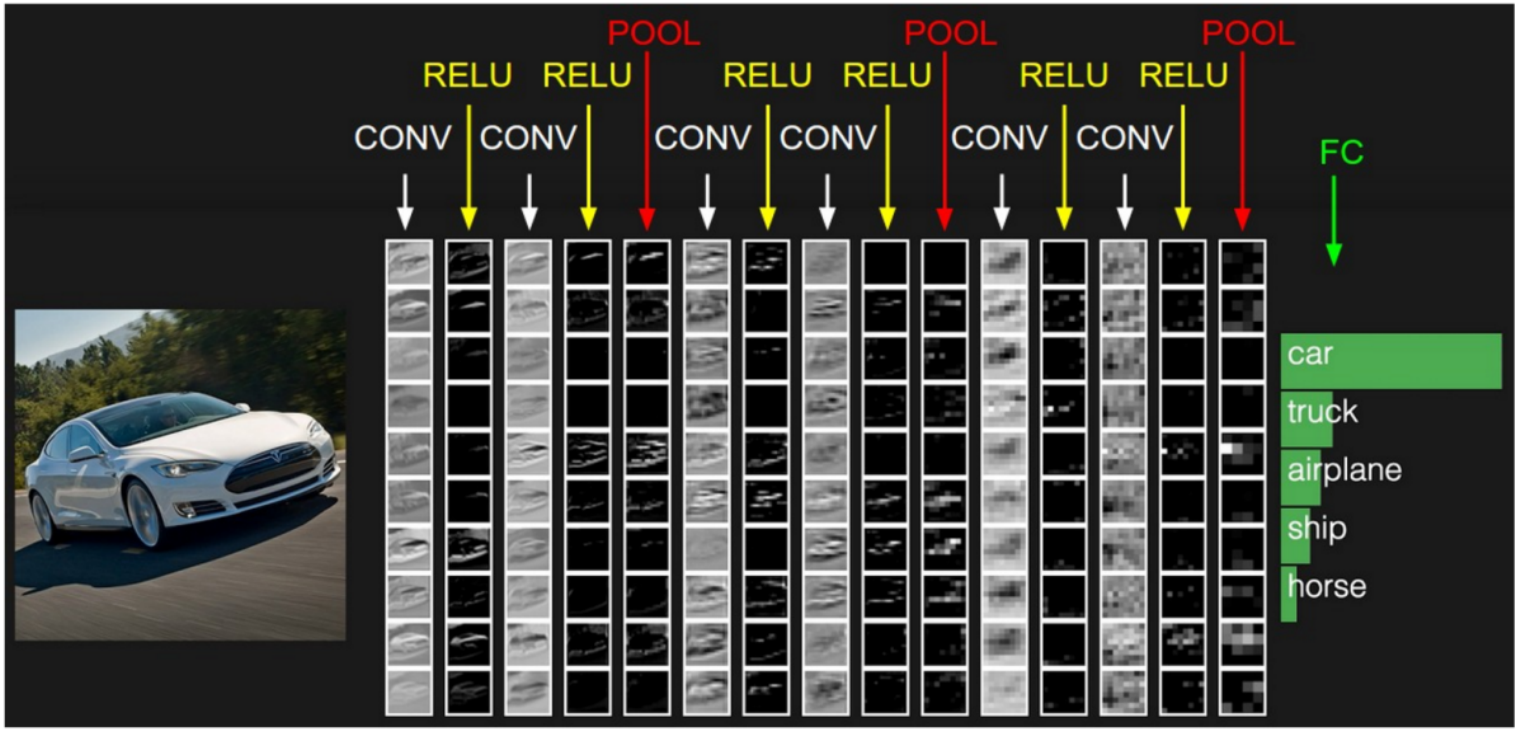


Image source: Stanford University

Layer Representation

- —

LeNet

- —

AlexNet

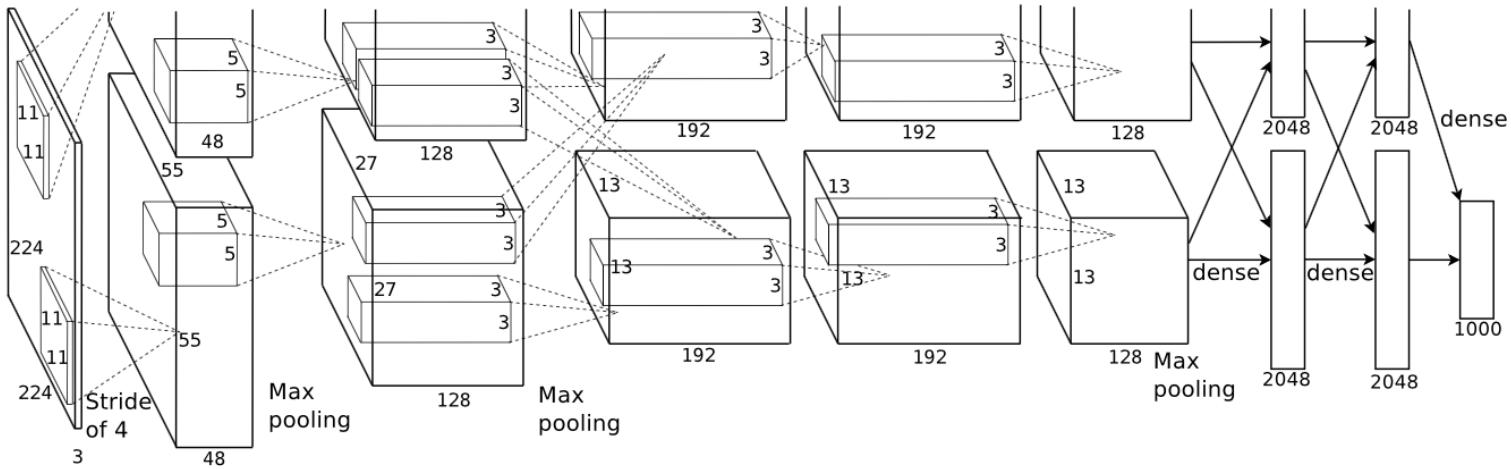


Image source: Conference on Neural Information Processing Systems

ResNet

- —

VGGNet

- —

Inception Network

- —

YOLO

- —

RetinaNet

- —

Non-maximum Supression

- —

Data Augmentation

- Important to include larger sets of training data.
- Ex: horizontal/vertical flips, rotations, resizing, cropping, changes in contrast/brightness, and/or distortions.

Transfer Learning

- —

Notebook

- Today's notebook will work through an example of convolutional neural networks.

References

- *Deep Convolutional Neural Nets*
- *Convolutional Neural Networks for Visual Recognition*