Object detection

Monday, March 11, 2019 1:58 AM

A. Sliding windows detectors
   1. Sliding windows
      a. Scale & size
      b. Aspect of ratio
      c. Multi response - choose the local maximum
   2. HOG detector
      a. Histogram of Gradients
      b. Input image ->
         normalized gamma & color ->
         compute gradients ->
            // for a cell includes 8x8 pixels
            // for each pixel
            // use 3x1 1x3 or 3x3 Sobel detector
            // calculate direction and overall gradient
         weighted vote into spatial orientationally cells ->
            // set bins e.g. 20 * 9 = 180
            // set weighted sum of each direction
         Contrast normalize over overlapping spatial blocks ->
            // set a block e.g. 2x2 cells
            // use L1-norm or L2-norm to normalize histogram of gradient directions vector magnitude
         Collect HOG's over detection window ->
            // the windows = cell i.e. 8x8 pixels
         SVM ->
         output
   3. Viola-Jones face detector
      a. Haar feature - Rectangular features
            // black - white
      b. Integral images
            // for haar feature calculation - for each window at least 2(B & W)*3(+ & -) = 6 operations
      c. Feature selection - boosting

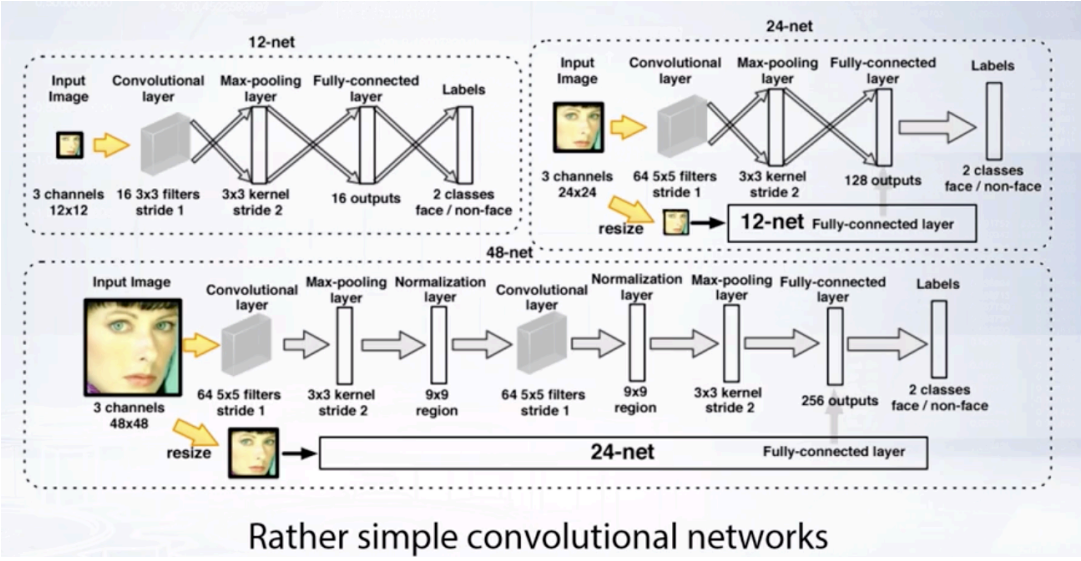
// reweight sample

d. Attentional cascade

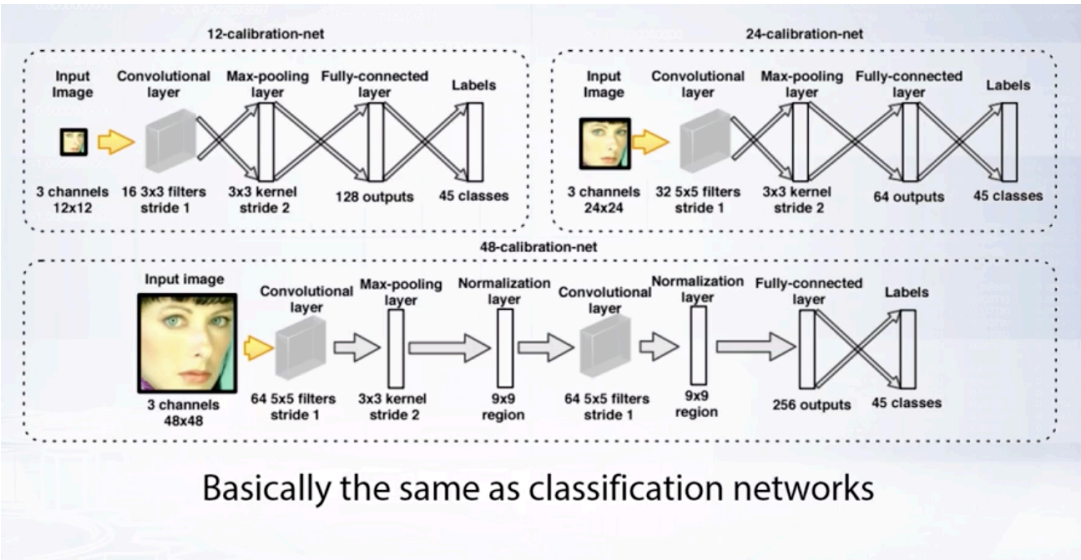
// simple clf(stage #1) to reject big windows, complex clf(stage #2, #3…) to refine the sub-windows

4. Attentional cascade and Neural network

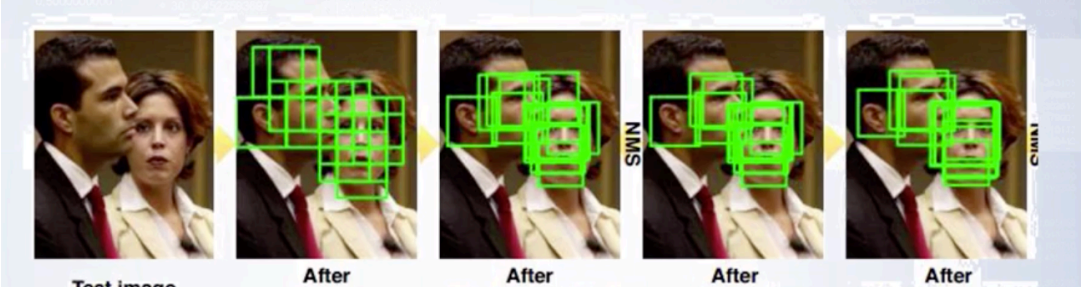
a. Classification network - Whether is a face in image

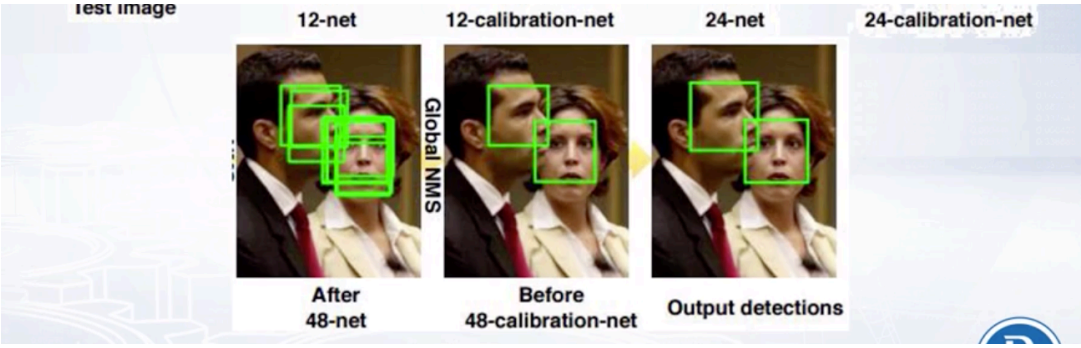


b. Calibration network - refine the position of bounding box



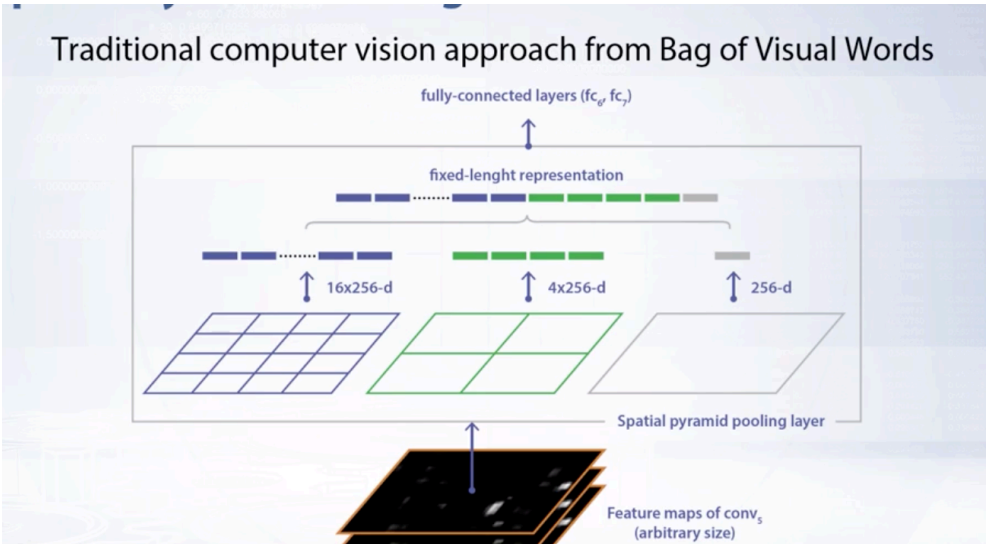
c. Combine of above two network





B. Modern detector architectures

1. Region-based Convolutional network (R-CNN)
   Two-stage cascade
   a. Select object proposal(object area/region)
      Use selective search to select proposals
   b. Apply strong CNN clf to those regions
      // pre-train CNN for feature extraction (image classification - large dataset)
      // fine-tune CNN for object detection (small target dataset)
      // linear clf and bounding box regressor are train on top of CNN features extracted from object proposals
2. Fast R-CNN
   a. Spatial pyramid pooling
      Input image ->
      Conv layers ->
      Spatial pyramid pooling -> 1+4+16+….
      Fc layers ->
      Output



b. ROI-pooling
Special SPP with only one pyramid level
c. Fast R-CNN : efficient SGD

3. Faster R-CNN

Faster R-CNN = Fast R-CNN + Region Proposal Network (RPN)

a. RPN
- Small network to classifying whether is an object
- Regressing bounding box locations

b. Four loss function
RPN classification (good/bad anchor)
RPN regression (anchor -> proposal)
Fast R-CNN classification (over classes)
Fast R-CNN regression (proposal -> box)

c. Multiple scale
Difficult for RPN to handle multiple scale
Then we can train a set of RPNs for various scales

4. Region-based Fully-Convolutional Network (R-FCN)
5. Single shot detector (SSD)

E.g. you only look once (YOLO)

Image is split into grid
Each cell predict boxes and confidence

Output parameterization:

Each cell predicts:

For each bounding bod:

4 coordinates(x,y,width,height)
confidence value

Some number of class probabilities

For each cell prediction : Number of boxes * (4 + 1) + number of classes