SE125 Machine Learning

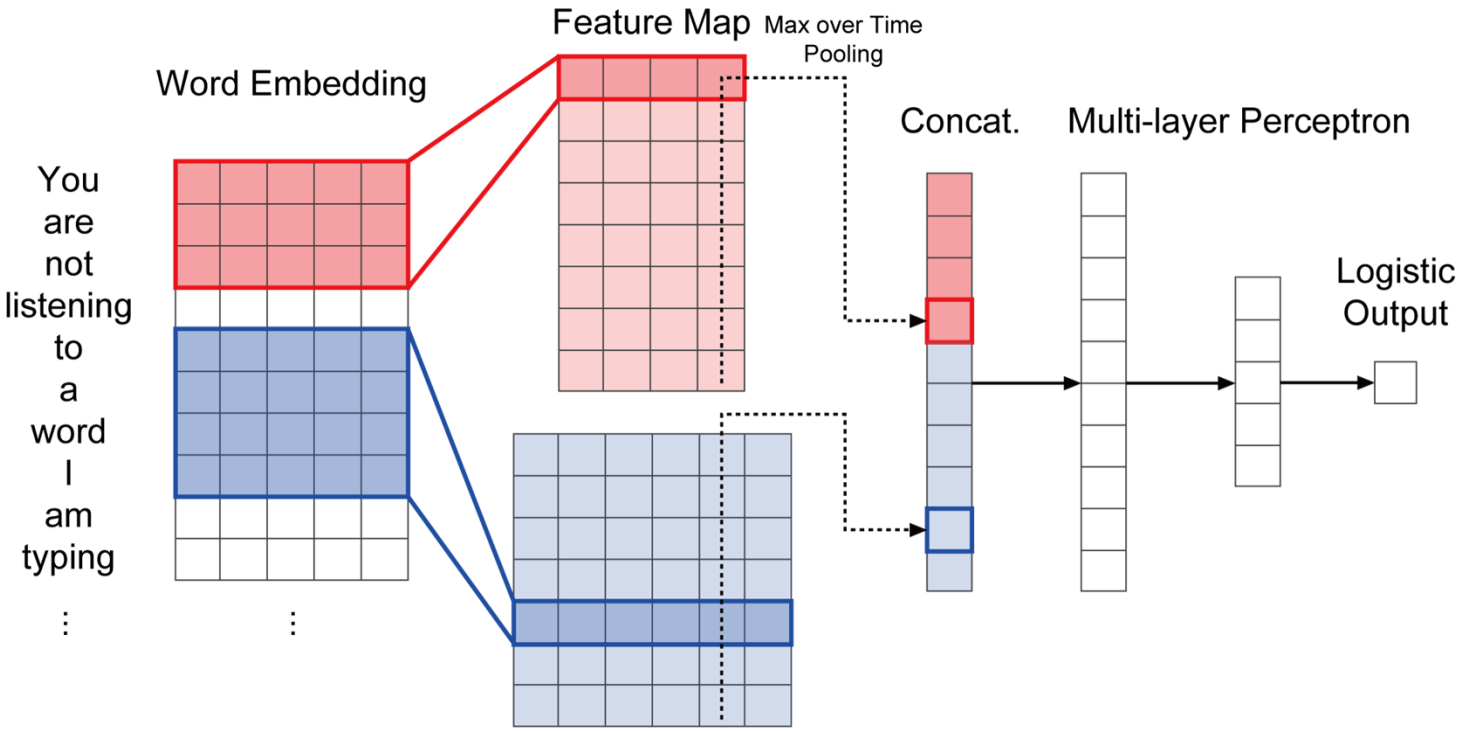
Convolutional Neural Networks Part III

Yue Ding

School of Software, Shanghai Jiao Tong University
dingyue@sjtu.edu.cn

- CNN Applications （扩展内容）

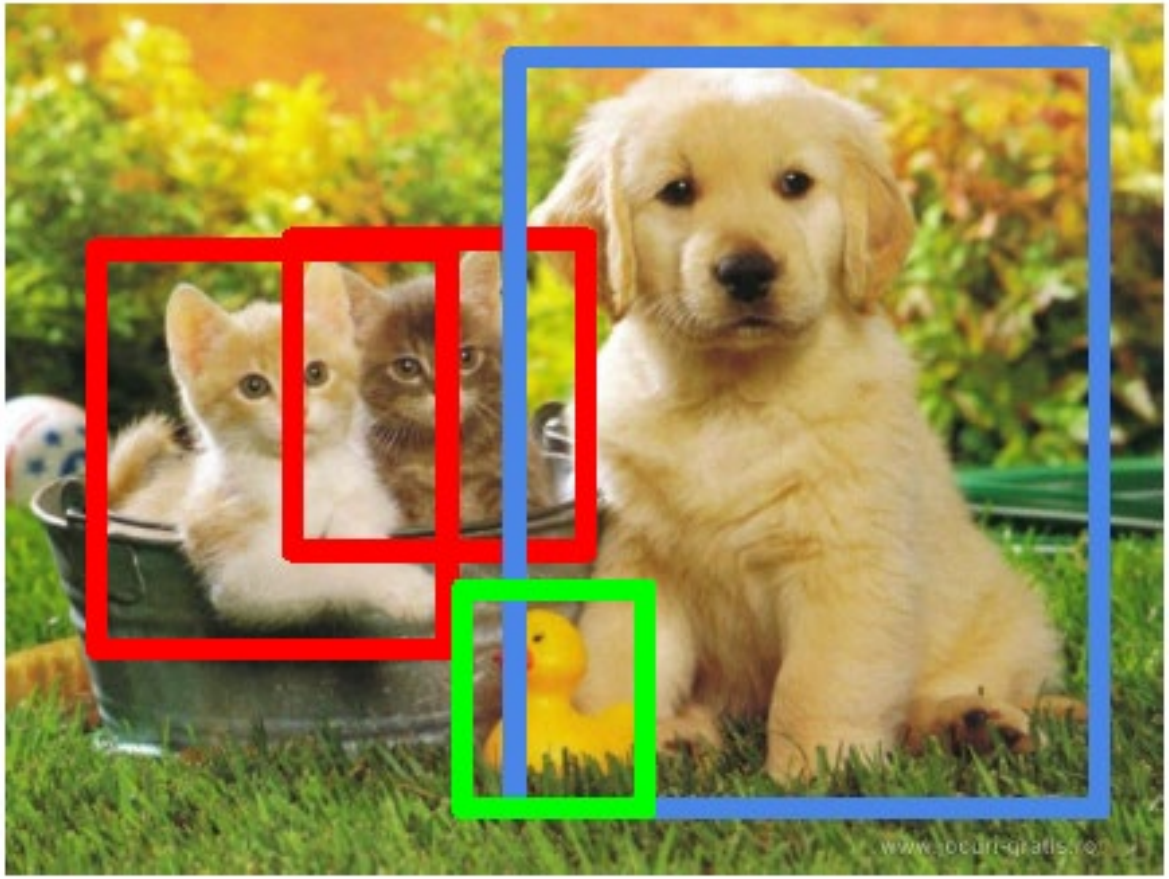
Text Classification



- Word embedding: map each word to a *k*-dimensional dense vector
- CNN kernel: *n* x *k* matrix to explore the neighbor *k* words' patterns
- Max-over-time pooling: find the most salient pattern from the text for each kernel
- MLP: further feature interaction and distill high-level patterns

[Kim, Y. 2014. Convolutional neural networks for sentence classification. EMNLP 2014.]

Object Detection



Object Detection

CAT? NO

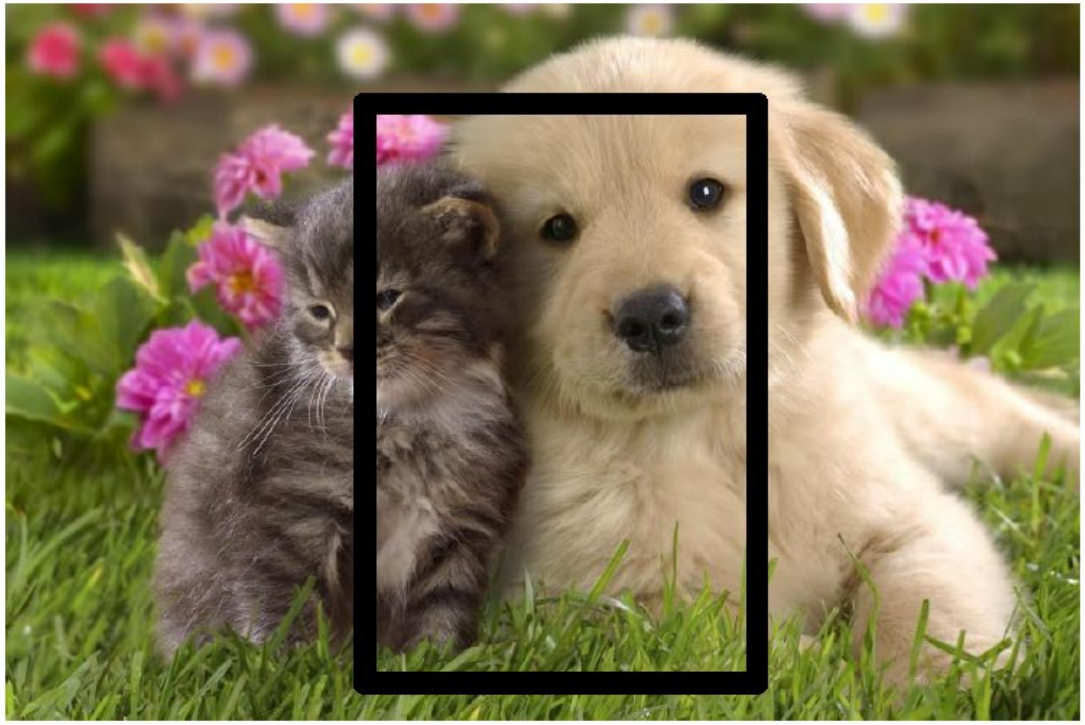
DOG? NO

Object Detection

CAT? YES!

DOG? NO

Object Detection



CAT? NO

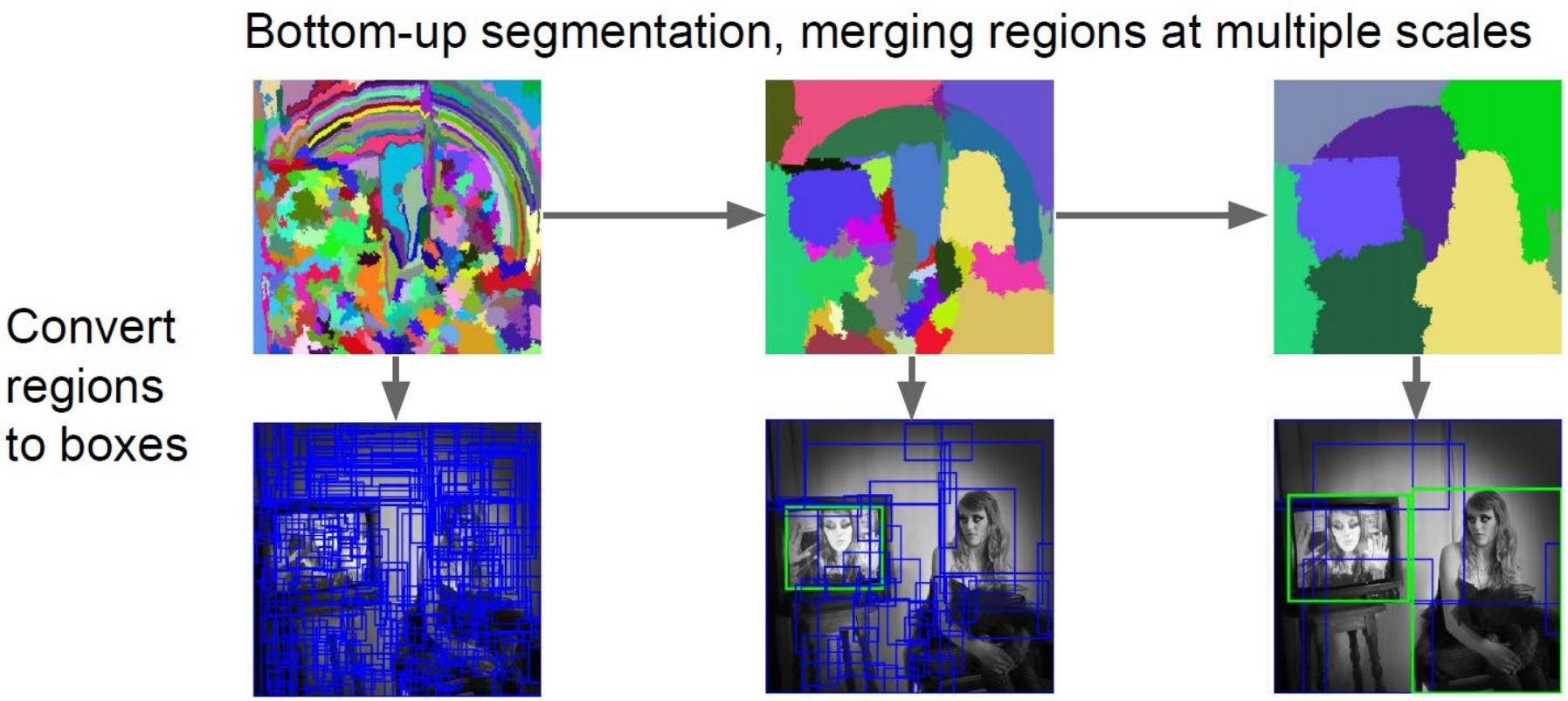
DOG? NO

Object Detection

- Detection as Classification
  - **Problem**: Need to test many positions and scales.
  - **Solution:** If your classifier is fast enough, just do it.

Object Detection

- Region Proposals: Selective Search



Uijlings et al, “Selective Search for Object Recognition”, IJCV 2013

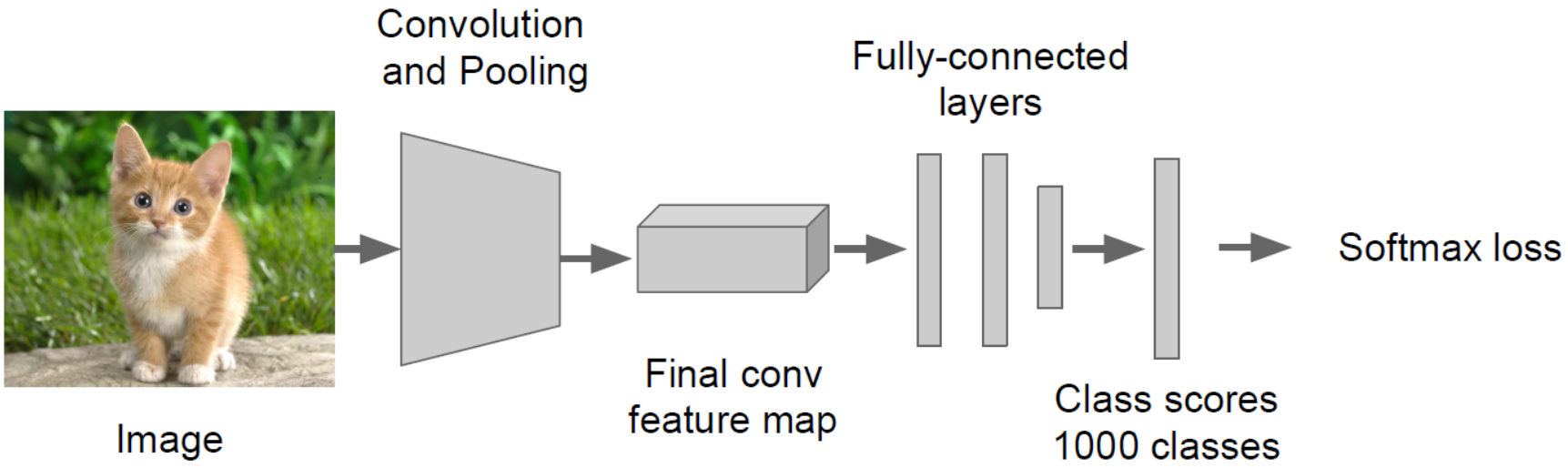
Object Detection

- R-CNN
  - Girschick et al, “Rich feature hierarchies for accurate object detection and semantic segmentation”, CVPR 2014

Object Detection

R-CNN Training

Step 1: Train (or download) a classification model for ImageNet (AlexNet)

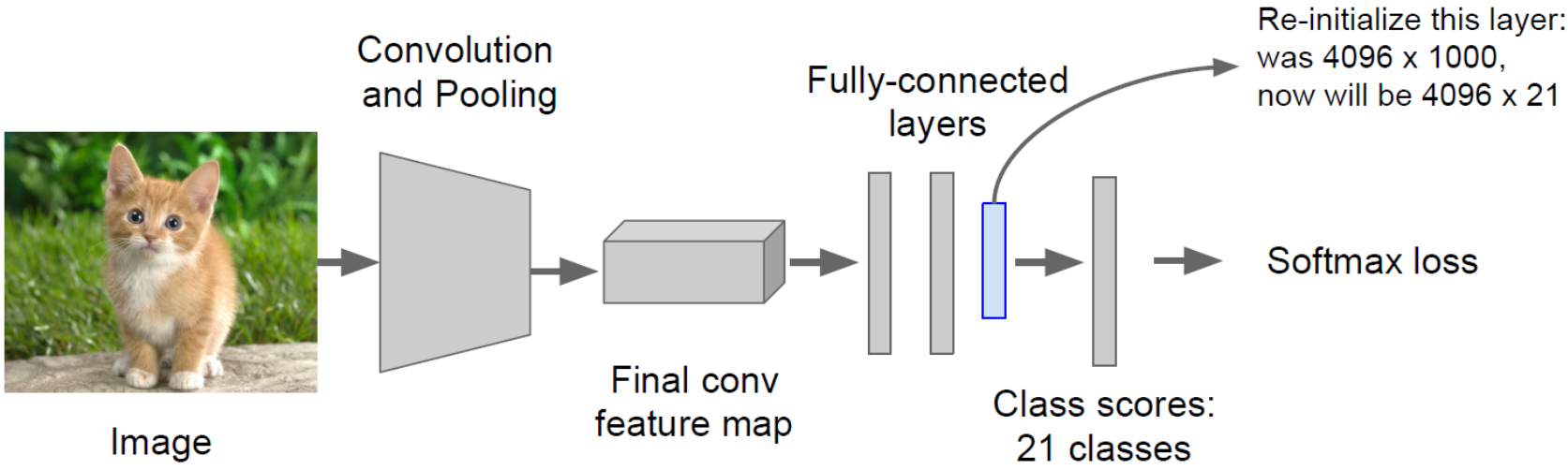


Object Detection

R-CNN Training

Step 2: Fine-tune model for detection

- Instead of 1000 ImageNet classes, want 20 object classes + background
- Throw away final fully-connected layer, reinitialize from scratch
- Keep training model using positive / negative regions from detection images



Object Detection

R-CNN Training

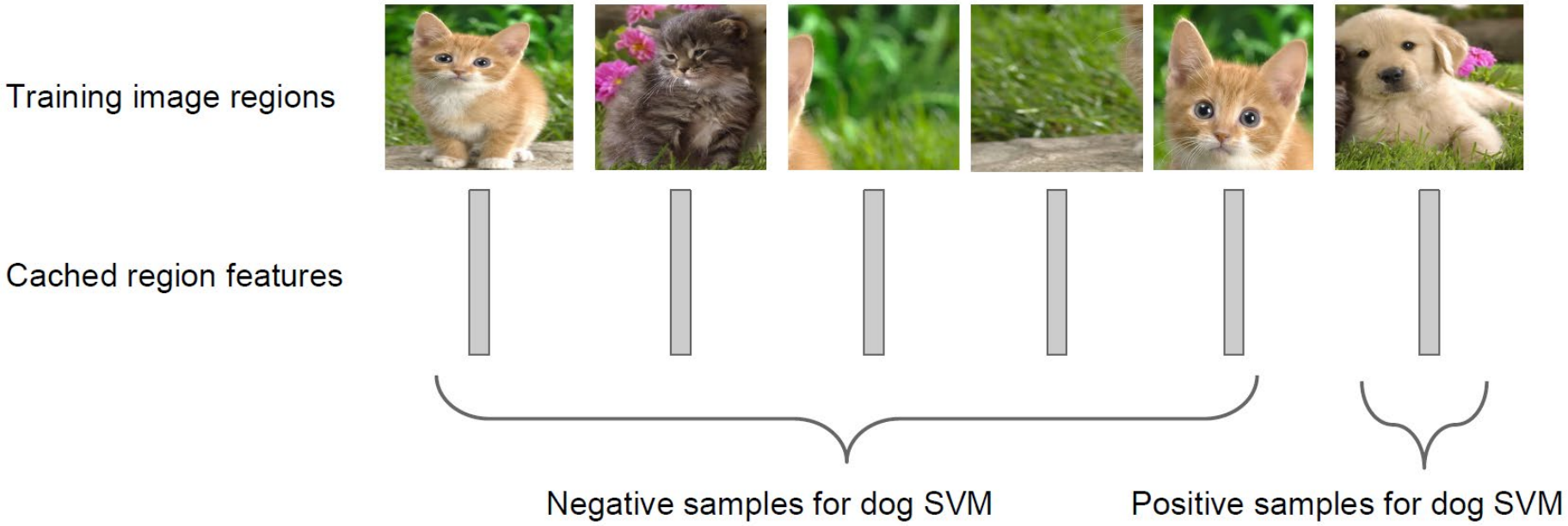
Step 3: Extract features

- Extract region proposals for all images
- For each region: warp to CNN input size, run forward through CNN, save pool5 features to disk
- Have a big hard drive: features are ~200GB for PASCAL dataset!

Object Detection

R-CNN Training

Step 4: Train one binary SVM per class to classify region features



Object Detection

R-CNN Training

Step 5 (bbox regression): For each class, train a linear regression model to map from cached features to offsets to GT boxes to make up for “slightly wrong” proposals

Training image regions

Cached region features

Regression targets
(dx, dy, dw, dh)
Normalized coordinates

(0, 0, 0, 0)
Proposal is good

(.25, 0, 0, 0)
Proposal too
far to left

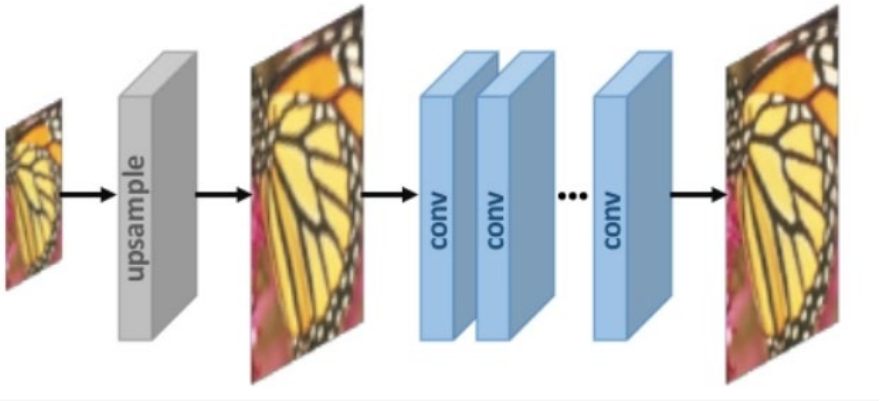
(0, 0, -0.125, 0)
Proposal too
wide

Image Super Resolution

- Super Resolution is the process of recovering a High Resolution (HR) image from a given Low Resolution (LR) image.

Image Super Resolution

- Pre-upsampling



- Post-upsampling