- #paper/to-read ~ 2013 CE ~ Object Detection, Semantic Segmentation
  - **Rich feature hierarchies for accurate object detection and semantic segmentation**
  - https://arxiv.org/abs/1311.2524
  - Sequel papers:
    - Fast R-CNN
    - Faster R-CNN
    - Mask R-CNN
    - Context R-CNN
  - Mentioned papers:
    - Regionlets
    - Objectness of Image Windows
    - Recognition Using Regions
    - Category Independent Object Proposals
    - DeCAF
    - Adaptive Deconvolutional Networks
    - Diagnosing Error in Object Detectors
  - Mentioned topics:
    - Deformable Part Models
    - Hard Negative Mining

Summary

- **Architecture**: input image → extracting region proposals (~2k) → feature-extraction CNN → class prediction or bounding box regression SVM (or both).
  - Intersection Over Union, IoU with ground truth labels should be greater than 0.5.
  - At test time, IoU is used to select the dominating prediction for a region.
  - With the enabled bounding box Regression, the model learns to predict shifts of the predicted bounding box relative to the coordinates of the proposed region.

- R-CNN can scale to thousands of classes without resorting to hierarchical tricks or Hashing.
- **Training process** consists of two parts.
  - Supervised pre-training happens on a big Classification dataset to make the CNN part learn without overfitting on (usually) small detection datasets.
  - Domain-specific fine-tuning is used to train SVM heads on region proposals.
    - Those proposals should fit 227×227 pixels so they are usually warped.
- Different CNN architectures yield different mAPs but also require different amount of Compute.
- For region proposals, Selective Search is used but other methods are appropriate too.
- For Semantic Segmentation, Constrained Parametric Min-Cuts, CPMC is used.
