CS6501: Deep Learning for Visual Recognition

Object Detection: RCNN, Fast-RCNN, Faster-RCNN
Today’s Class

- Object Detection
- The RCNN Object Detector (2014)
- The Fast RCNN Object Detector (2015)
- The Faster RCNN Object Detector (2016)
- YOLO (CVPR 2016)
- SSD (ECCV 2016)
Object Detection

deer
Object Detection

Class Scores
- Deer: 0.9
- Cat: 0.05
- Umbrella: 0.01
...

Box Coordinates
(x, y, w, h)

Fully Connected:
4096 to k

Fully Connected:
4096 to 4
Object Detection

Deer: (x, y, w, h)
Cat: (x, y, w, h)
Object Detection

Penguin: (x, y, w, h)
Penguin: (x, y, w, h)
Penguin: (x, y, w, h)
Penguin: (x, y, w, h)
...

4096
Object Detection as Classification

---

deer?
cat?
background?
Object Detection as Classification

deer?
cat?
background?
Object Detection as Classification

CNN

deer?
cat?
background?
Object Detection as Classification with Sliding Window

deer?
cat?
background?
Object Detection as Classification with Box Proposals

RCNN

First stage: generate category-independent region proposals.

- 2000 Region proposals for every image

Selective Search: combine the strength of both an exhaustive search and segmentation. Uijlings et al. IJCV 2013.

ref
RCNN

First stage: generate category-independent region proposals.
- 2000 Region proposals for every image

Second stage: extracts a fixed-length feature vector from each region.
- A 4096-dimensional feature vector from each region proposal
RCNN

First stage: generate category-independent region proposals.
- 2000 Region proposals for every image

Second stage: extracts a fixed-length feature vector from each region.
- A 4096-dimensional feature vector from each region proposal

Third stage: a set of class-specific linear SVMs.
- Object category and location
RCNN

• Simple and scalable.
• improves mAP.

• A multistage pipeline.
• Training is expensive in space and time (features are extracted from each region proposal in each image and written into disk).
• Object detection is slow.

Fast-RCNN

?
Fast-RCNN

https://arxiv.org/abs/1504.08083
Fast R-CNN. Girshick. ICCV 2015.

Idea: No need to recompute features for every box independently
Fast-RCNN

Process the whole image with several convolutional (conv) and max pooling layers to produce a conv feature map.

A region of interest (RoI) pooling layer extracts a fixed-length feature vector from the region feature map.

FC+ softmax

K + 1 categories

FC+ regressor

four real-valued numbers for each of the K object classes.
RCNN vs Fast-RCNN

Figure adapted from: http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture11.pdf
RCNN

- Simple and scalable.
- Improves mAP.
- A multistage pipeline.
- Training is expensive in space and time (features are extracted from each region proposal in each image and written into disk).
- Object detection is slow.

Fast-RCNN

- Higher mAP.
- Single stage, end-to-end training.
- No disk storage is required for feature caching.
- Proposals are the computational bottleneck in detection systems.

Faster-RCNN

?
Faster-RCNN

Idea: Integrate the Bounding Box Proposals as part of the CNN predictions

https://arxiv.org/abs/1506.01497
Ren et al. NIPS 2015.
Faster-RCNN

Region Proposal Networks:

- Feature map
  - Sliding window, nxn
  - nxn conv layer
  - 1x1 conv layer
- k anchors boxes
- 2k scores
- Object or not object
- 4k coordinates
- Bounding box proposal
- cls layer
- reg layer

RPN

Shared conv layers

Fast-RCNN

Object is a cat
- Classification loss
- Bounding-box regression loss

Refine BB position

Rol pooling

proposals

Region Proposal Network
RCNN vs Fast-RCNN

Figure adapted from: http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture11.pdf
RCNN

- Simple and scalable.
- Improves mAP.
- A multistage pipeline.
- Training is expensive in space and time (features are extracted from each region proposal in each image and written into disk).
- Object detection is slow.

Fast-RCNN

- Higher mAP.
- Single stage, end-to-end training.
- No disk storage is required for feature caching.
- Proposals are the computational bottleneck in detection systems.

Faster-RCNN

- Compute proposals with a deep convolutional neural network -- Region Proposal Network (RPN)
- Merge RPN and Fast R-CNN into a single network, enabling nearly cost-free region proposals.
YOLO - You Only Look Once

Idea: No bounding box proposal. A single regression problem, straight from image pixels to bounding box coordinates and class probabilities.

- extremely fast
- reason globally
- learn generalizable representations

https://arxiv.org/abs/1506.02640
Redmon et al. CVPR 2016.
YOLO - You Only Look Once

Divide the image into 7x7 cells.
Each cell trains a detector.
The detector needs to predict the object’s class distributions.
The detector has 2 bounding-box predictors to predict bounding-boxes and confidence scores.
SSD: Single Shot Detector

Idea: Similar to YOLO, but denser grid map, multiscale grid maps. + Data augmentation + Hard negative mining + Other design choices in the network.

Liu et al. ECCV 2016.
Questions?