Motivation

The two-step detection framework.

Task 1: Proposal generation
- Ideal:
  - Output only object proposals.
- Reality:
  - A majority of samples are background (num_classes becomes N+1);
  - With cross-entropy objective, CNN learns biased representation, and it is hard to capture fine-grained variance of each category.

Wrong detections of Fast R-CNN:
- Reality:
  - Background regions
- Ideal:
  - Object categories of interest.

Task 2: Object classification
- Ideal:
  - Classify proposals into N object categories of interest.
- Reality:
  - A majority of proposals overlap multiple sources;
  - 1% absolute mAP gain.

Wrong detections of Fast R-CNN: 'tv_monitor', 0.8 'boat', 0.7 'dog', 0.5

Motivation

Using 'divide and conquer' philosophy to further decompose and better solve each of the two tasks;
- Each task is solved with a carefully designed neural network cascade.

Approach

CRAFT (Cascade RPN And Fast-R-CNN)
- Each task is solved with a carefully designed neural network cascade.
- A two-step detection framework?
- Why don’t we take more baby steps?

Solution

CRAFT enjoys other advances in object detection like ION, ResNet; for fast-r-cnn, and provides an alternative solution.

Results

Object proposal on VOC07 test:
- Recall analysis on difficult categories

Object detection on VOC07/12 test and ILSVRC val2:
- Recall analysis at various IoUs and the detection mAP

ImageNet 2015 Object Detection from Video (VID) Competition
- Detection on VOC07 and ILSVRC val2

Discussion

- CRAFT enjoys other advances in object detection like ION, ResNet;
- The cascade structure used in proposal task plays the role of hard example mining for the following detection task;
- The cascade structure used in detection task points out a potential drawback of current loss function choice for fast-r-cnn, and provides an alternative solution.