Fast R-CNN by Ross Girshick

Minsung Cho

Short summary of the paper

1. The problem that is being addressed
   1. R-CNN and SPPnet’s problems
      1. Training is a multi-stage pipeline.
      2. Training is expensive in space and time.
      3. Object detection is slow.
2. Its significance
   1. Training is a multi-stage pipeline:
      1. Stage 1 is object proposal using log loss
      2. Stage 2 is SVM acting as object detectors
      3. Stage 3 is the bounding box regression
   2. Training is expensive in space and time:
      1. It takes 2.5 GPU days for the 5k images and requires hundreds of gigabytes of storage.
   3. Object detection is slow:
      1. On GPU, it takes 47s / image
3. The approach used to solve the problem
   1. The RoI pooling layer
   2. Initializing from pre-trained networks
   3. Fine-tuning for detection:
      1. Multi-task loss
      2. Mini-batch sampling
      3. Back propagation through RoI pooling layers
      4. SGD Hyperparameters
   4. Scale invariant

Assumptions that are made

Three pre-trained ImageNet models: CaffeNet, VGG CNN M 1024, and the very deep VGG16 model.

An assessment of the strengths

1. Why it was better than the existing techniques at the time
   1. Higher detection quality (mAP) than R-CNN, SPPnet
   2. Training is single-stage, using a multi-task loss
   3. Training can update all network layers
   4. No disk storage is required for feature caching

An assessment of the weaknesses

1. When the algorithm might fail
   1. Running the implementation on CPU is slower than GPU
   2. Does not share convolutional layers with detection networks at test time
   3. Spends too much time on region proposals

List of 2-3 questions that I have regarding the work