

Motivation



Low Classification Score

Ground Truth

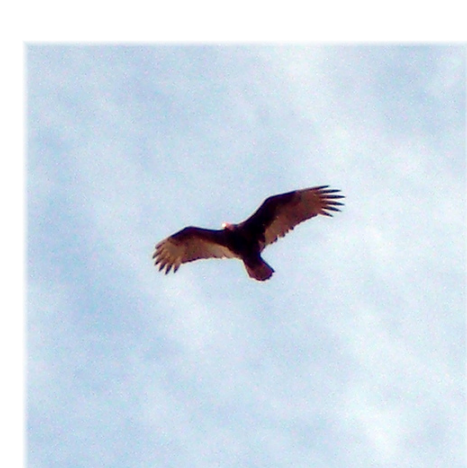
High Classification Score

- ✓ Can classification score of a specific class tell us location information?
 - ✓ Cropped image will have low classification score if there's little overlap between the target object and cropped area
 - ✓ Cropped image will have high classification score if there's large overlap between the target object and cropped area
 - ✓ By evaluating multiple crops of various size and locations, we can infer object location information using classification scores only

Method

Classification Network

- ✓ 3 GoogLeNets with 10 crops (8.53% Top-5 acc on validation set)



Input image



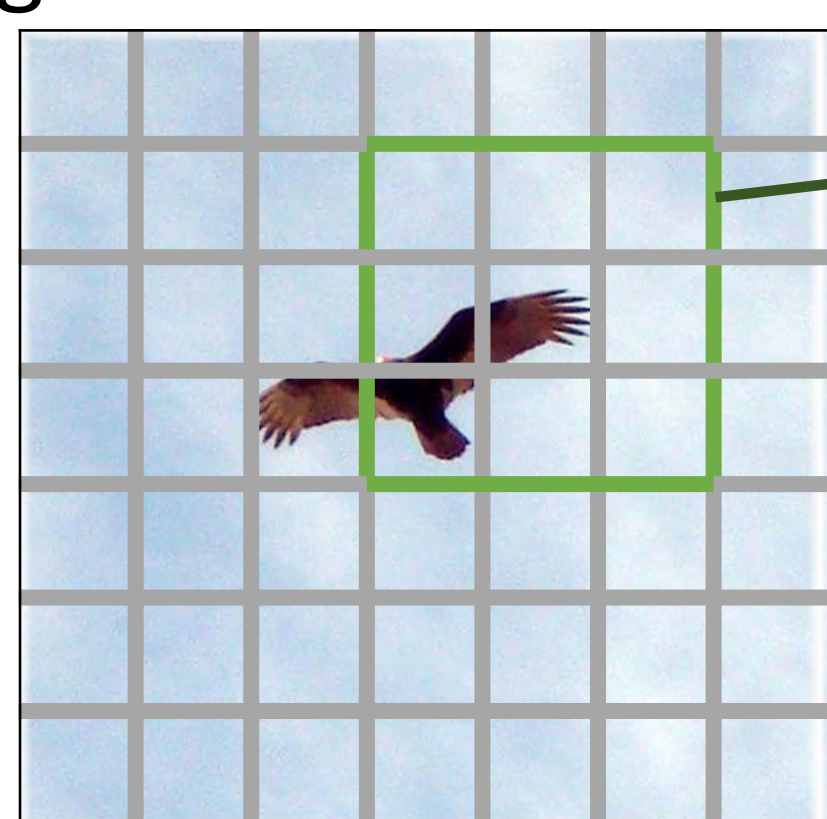
GoogLeNet trained in image-level

Classification

Class prediction

Object Scoring Network

- ✓ Finetune GoogLeNet with images cropped in object-level bounding boxes
- ✓ We exploit the classification scores (before softmax) of predicted class as a feature
- ✓ Dividing the image into 7x7 grid, 140 possible square crops are evaluated to get the classification scores and generate a 140-dim feature vector



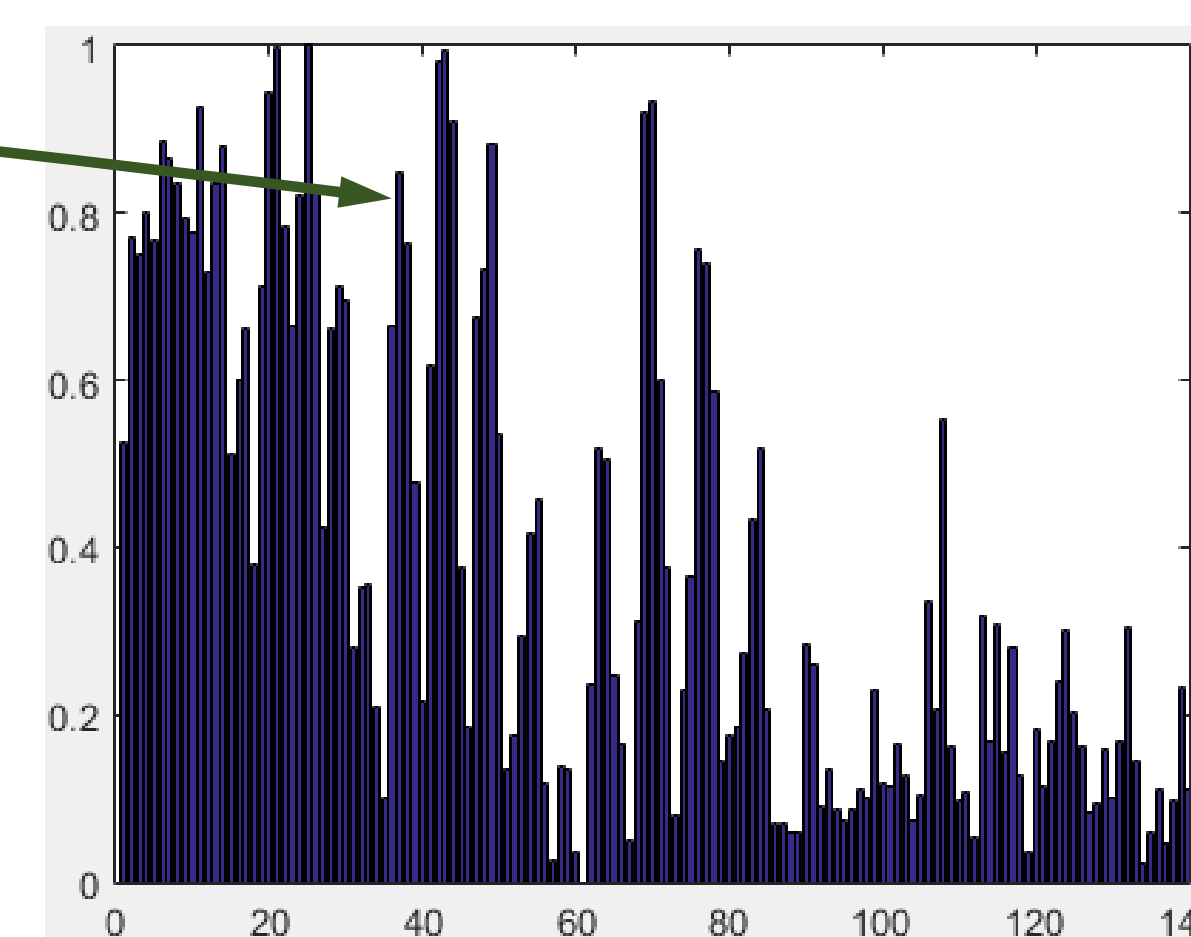
Divide image into 7x7 grid

Each square crop corresponds to a specific dimension of a feature vector



GoogLeNet finetuned in object-level

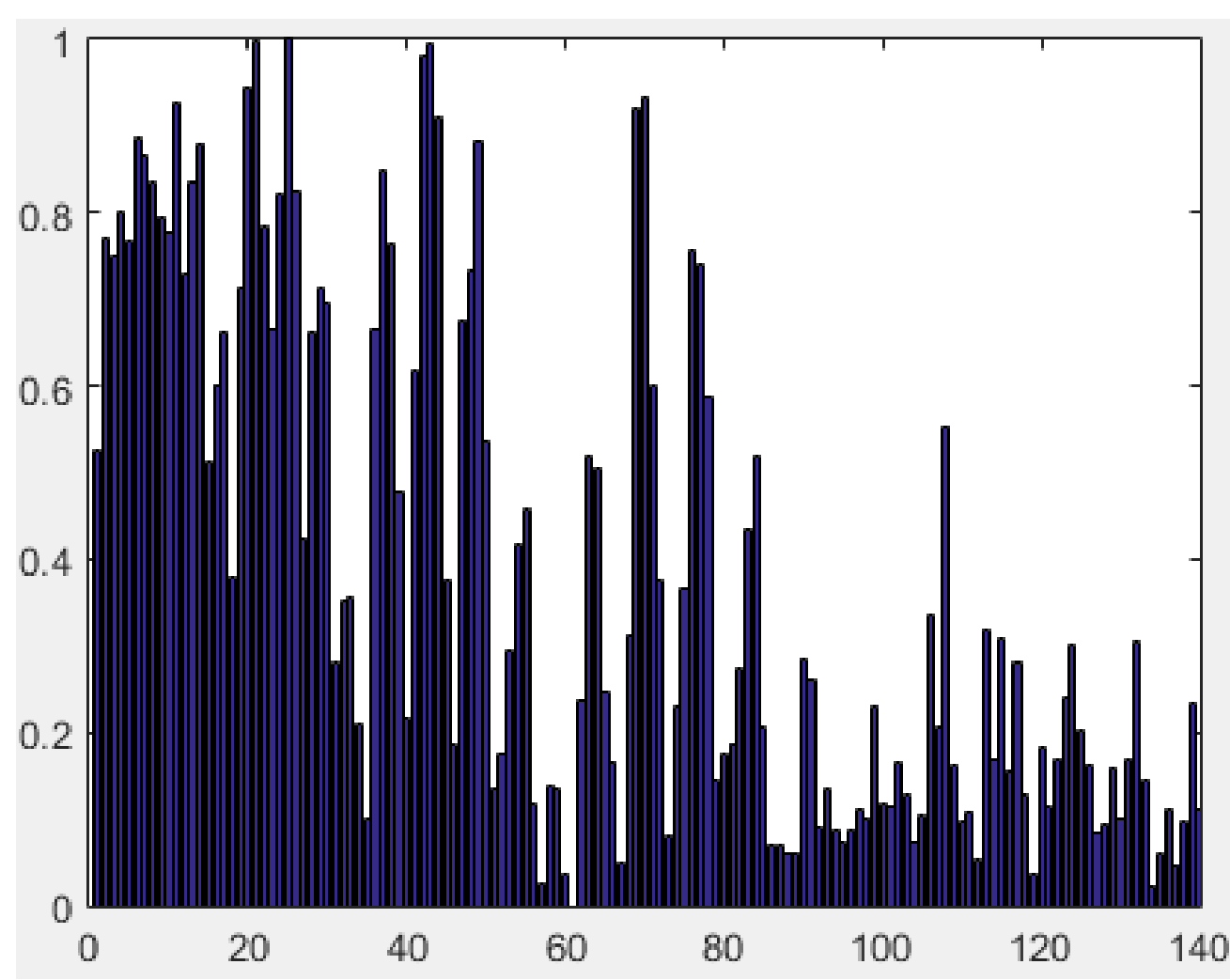
Classification score with 140 square crops



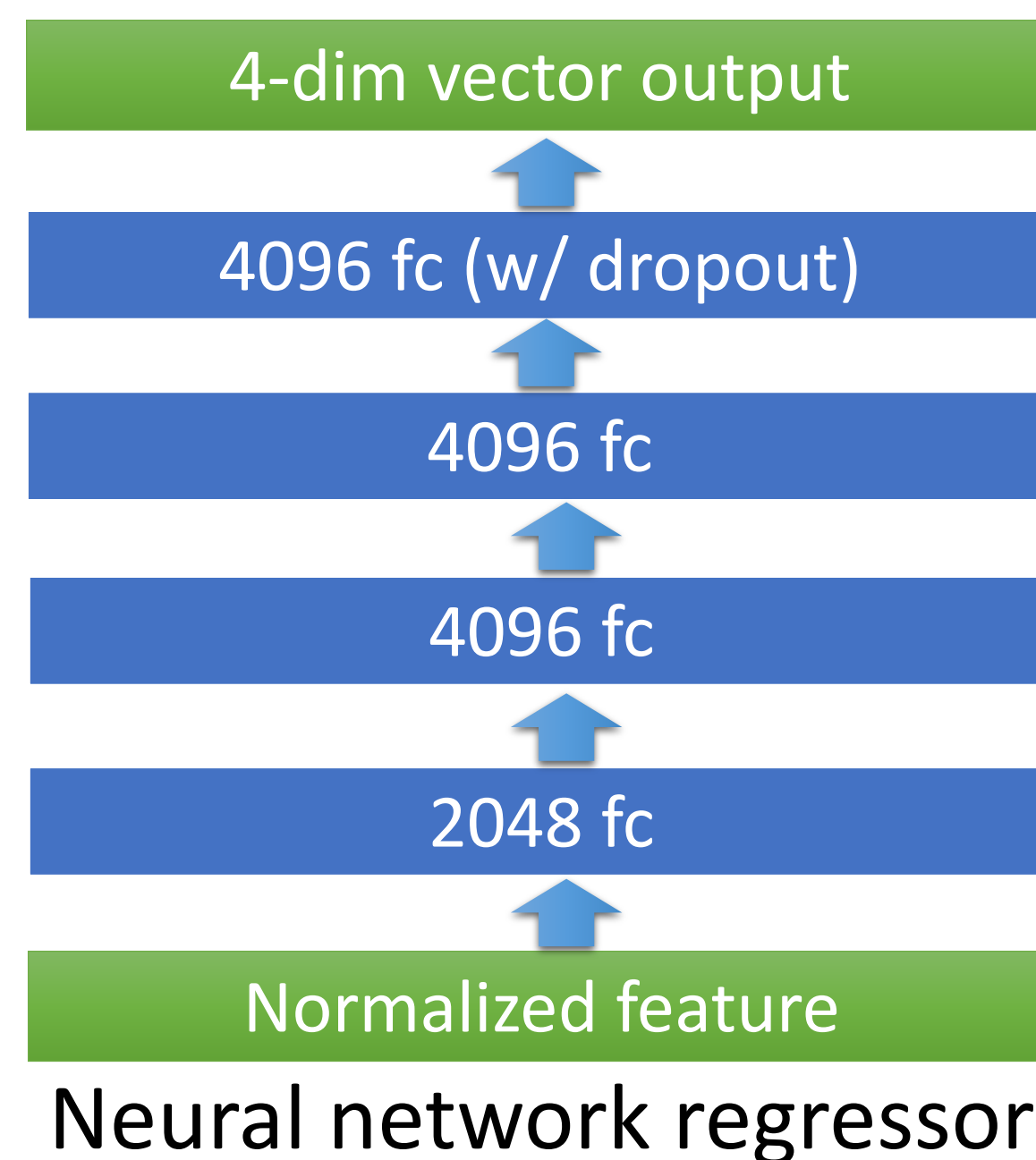
140-dim normalized feature

Localization Network

- ✓ Training neural net regressor which minimizes Euclidean loss
- ✓ Perform class-wise finetuning



140-dim normalized feature



Neural network regressor



Bounding box prediction

Performance

- ✓ Localization performance on validation set

Method	GT class + box regression
Localization Error	32.3%

- ✓ Future works

- ✓ Dealing with images that have multiple objects
- ✓ Various cropping strategies