OBJECT LOCALIZATION

DAVID LU

INTRODUCTION

PREPROCESSING AND DATA AUGMENTATION

Determining the location of a phone in images

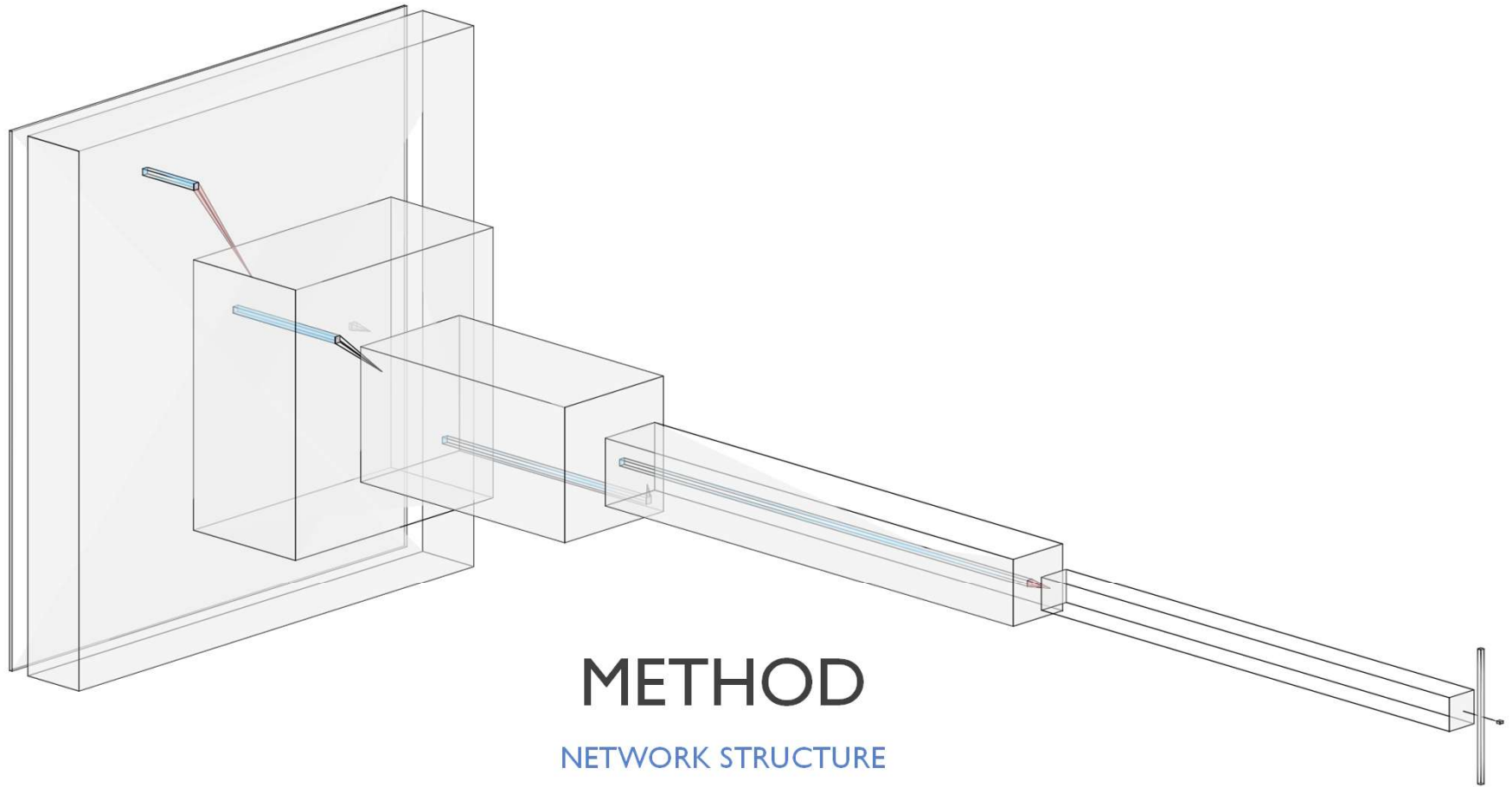Return the object coordinates with respect to the image

RATIONALE

Coordinate Localization is regressed even in complex detection networks

Retraining feature detectors on Scarce datasets is foolish

Use weights in VGG trained on Imagenet to jumpstart custom model

Update convolutional parameters to specialize network to task

METHOD

NETWORK STRUCTURE

METHOD

256x256 input squares

Convolutional and Pooling layers similar to VGG

Fully Connected Layer for regression of two coordinates

AD HOC AUGMENTATION DURING TRAINING

PREPROCESSING AND DATA AUGMENTATION

Random Vertical Flipping
Random Horizontal Flipping
Random Brightness Adjustment
Random Contrast Adjustment
Random Saturation Adjustment
Random Hue Adjustment
Normalization

HYPER PARAMETERS AND TRAINING

100 epochs
16 batch
SGD with 0.001 learning rate 0.9 momentum
0.5x / 20 epoch weight decay
MSE Loss

PARAMETERS AND LAYERS

```
----------------------------------------------------------------
       Layer (type)          Output Shape        Param #
================================================================
         Conv2d-1          [-1, 64, 256, 256]      1,792
          ReLU-2           [-1, 64, 256, 256]        0
        MaxPool2d-3        [-1, 64, 128, 128]        0
         Conv2d-4         [-1, 128, 128, 128]      73,856
          ReLU-5          [-1, 128, 128, 128]        0
        MaxPool2d-6        [-1, 128, 64, 64]         0
         Conv2d-7          [-1, 256, 64, 64]      295,168
          ReLU-8           [-1, 256, 64, 64]         0
         Conv2d-9          [-1, 256, 64, 64]      590,080
         ReLU-10           [-1, 256, 64, 64]         0
       MaxPool2d-11        [-1, 256, 32, 32]         0
        Conv2d-12          [-1, 512, 32, 32]     1,180,160
         ReLU-13           [-1, 512, 32, 32]         0
        Conv2d-14          [-1, 512, 32, 32]     2,359,808
         ReLU-15           [-1, 512, 32, 32]         0
       MaxPool2d-16        [-1, 512, 16, 16]         0
        Conv2d-17          [-1, 512, 16, 16]     2,359,808
         ReLU-18           [-1, 512, 16, 16]         0
        Conv2d-19          [-1, 512, 16, 16]     2,359,808
         ReLU-20           [-1, 512, 16, 16]         0
       MaxPool2d-21         [-1, 512, 8, 8]          0
  AdaptiveAvgPool2d-22      [-1, 512, 7, 7]          0
        Linear-23              [-1, 64]         1,605,696
       Sigmoid-24              [-1, 64]              0
       Dropout-25              [-1, 64]              0
        Linear-26              [-1, 2]              130
================================================================
```
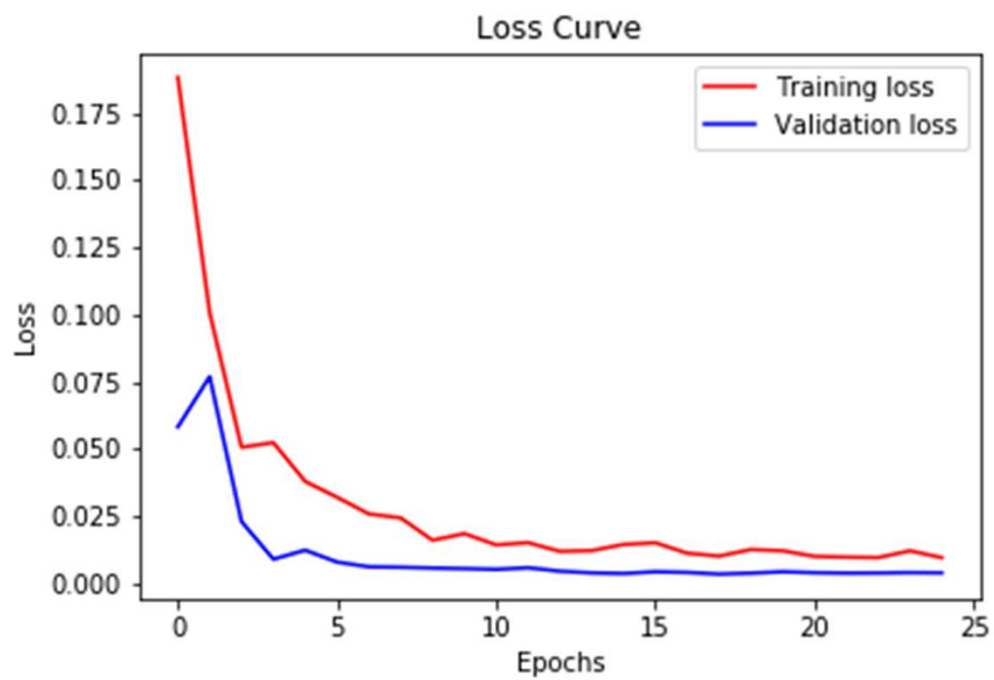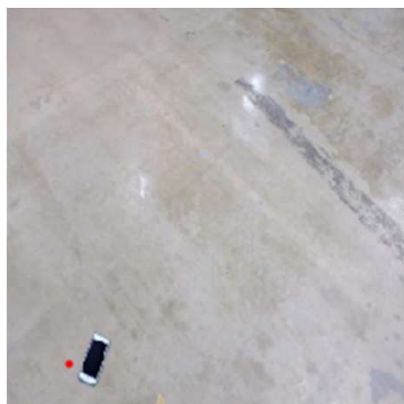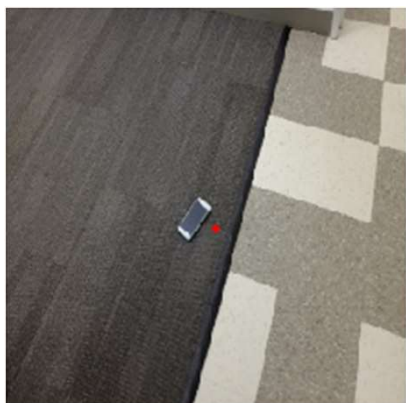
Loss Curve

RESULTS

LOSS CURVE

120.jpg

121.jpg

122.jpg

123.jpg

124.jpg

126.jpg

RESULTS

PREDICTED VISUALIZATIONS

Name	X	Y
120.jpg	0.7976	0.7551
121.jpg	0.1528	0.8774
122.jpg	0.1555	0.8610
123.jpg	0.5137	0.5422
124.pg	0.7865	0.2362
126.jpg	0.4059	0.2117

RESULTS

COORDINATES

CONCLUSION

Generalization to test images
Overfitting with larger structures
Custom dataloader
Sensitive to scale

CONCLUSION

Multi-scale network for different scale phones

Skip connections to avoid overfitting

Usage of UNET style network to remove background before dense layers