Object Detection on Monocular Images with Two-Dimensional Canonical Correlation Analysis



School of Engineering

USCViterbi

Zifan Yu, Suya You

Task and Intuition

Task Introduction :

We work on object detection on monocular images with twodimensional Canonical Analysis (2D CCA) to fuse the correlated features of the RGB images with the correlated features of the estimated depth data.



Approach keys:

- Inferring scene geometry (depth) directly from monocular RGB imagery without using any real depth sensor.
- Applying two dimensional canonical correlation analysis (2DCCA) to fuse and extract the correlated features between the RGB and depth data.
- Adding a novel 2D-CCA layer to merge two-dimensional CCA into deep neural networks for the object detection task.

$$X_t = l_x^T X r_x , \ Y_t = l_y^T Y r_y$$

$$max corr(X_t, Y_t) = max corr(l_x^T X r_x, l_y^T Y r_y)$$



- The depth information is estimated by a pre-trained encoder-decoder convolutional neural network. The backbone for the RGB stream and estimated depth stream is ResNet. Extracted feature maps by 2D-CCA layer are concatenated for detection.
- The proposed 2D-CCA layer has two pairs of trainable parameters, which correspond to the left transformation and right transformation in 2D-CCA. The parameters are initiated by calculating 2D-CCA transformation, and 2D-CCA is calculated every *x* epochs to update the parameters of the 2D-CCA layer.
- The loss function consists of the category classification loss and the bound box regression loss.

Results and Evaluations

| Results of Kitti: | m | AP | mRecall | | mIol | |
|-------------------|---------|---------|---------|---------|---------|---------|
| | iou=0.7 | iou=0.5 | iou=0.7 | iou=0.5 | iou=0.7 | iou=0.5 |
| Faster-RCNN RGBD | 48.60 | 57.13 | 50.96 | 59.55 | 72.13 | 72.33 |
| Faster-RCNN 2DCCA | 41.72 | 57.94 | 44.74 | 61.47 | 67.88 | 70.04 |

Results of different fusing operation :

| | mAP | | mRecall | | mIoU | | |
|---|---------|---------|---------|---------|---------|---------|--|
| | iou=0.7 | iou=0.5 | iou=0.7 | iou=0.5 | iou=0.7 | iou=0.5 | |
| + | 37.33 | 54.23 | 39.81 | 57.95 | 64.59 | 67.50 | |
| Ð | 41.72 | 57.94 | 44.74 | 61.47 | 67.88 | 70.04 | |

Results of different 2D-CCA calculation frequency:

| | mAP | mRecall | mloU | |
|-------------|-------|---------|-------|--|
| CCA Initial | 46.64 | 48.05 | 62.42 | |
| X = 10 | 57.94 | 61.47 | 70.04 | |
| X = 20 | 42.21 | 44.67 | 59.07 | |

Results of Virtual Kitti:

| | Validation | | | Test | | |
|-------------------|------------|---------|-------|-------|---------|-------|
| | mAP | mRecall | mloU | mAP | mRecall | mloU |
| Faster-RCNN RGBD | 98.01 | 96.69 | 91.31 | 98.29 | 97.18 | 91.80 |
| Faster-RCNN 2DCCA | 97.29 | 94.11 | 88.31 | 98.02 | 95.60 | 88.87 |

