# 2nd Place Solution to Instance Segmentation of IJCAI 3D AI Challenge 2020

Kai Jiang[1]*, Xiangyue Liu[2]*, Zheng Ju[3]*, Xiang Luo[1]
[1]LinkDoc Technology, [2]Beihang University, [3]Huaxin consulting Co.,Ltd

## Introduce

- Instance Segmentation is a hot-pot topic in Computer Vision in recent year.
- There are no large-scale well organized benchmarks which providing realistic synthetic indoor images. 3D-FUTURE fill the blank.

## Data Analysis

According to the official split, we adopt 12,144 images for training, 2,024 images for validation, and 6,072 images for the test. We analyze the datasets from three main aspects: (1) **category distribution**, (2) **aspect ratio**, and (3) **area of instance**.

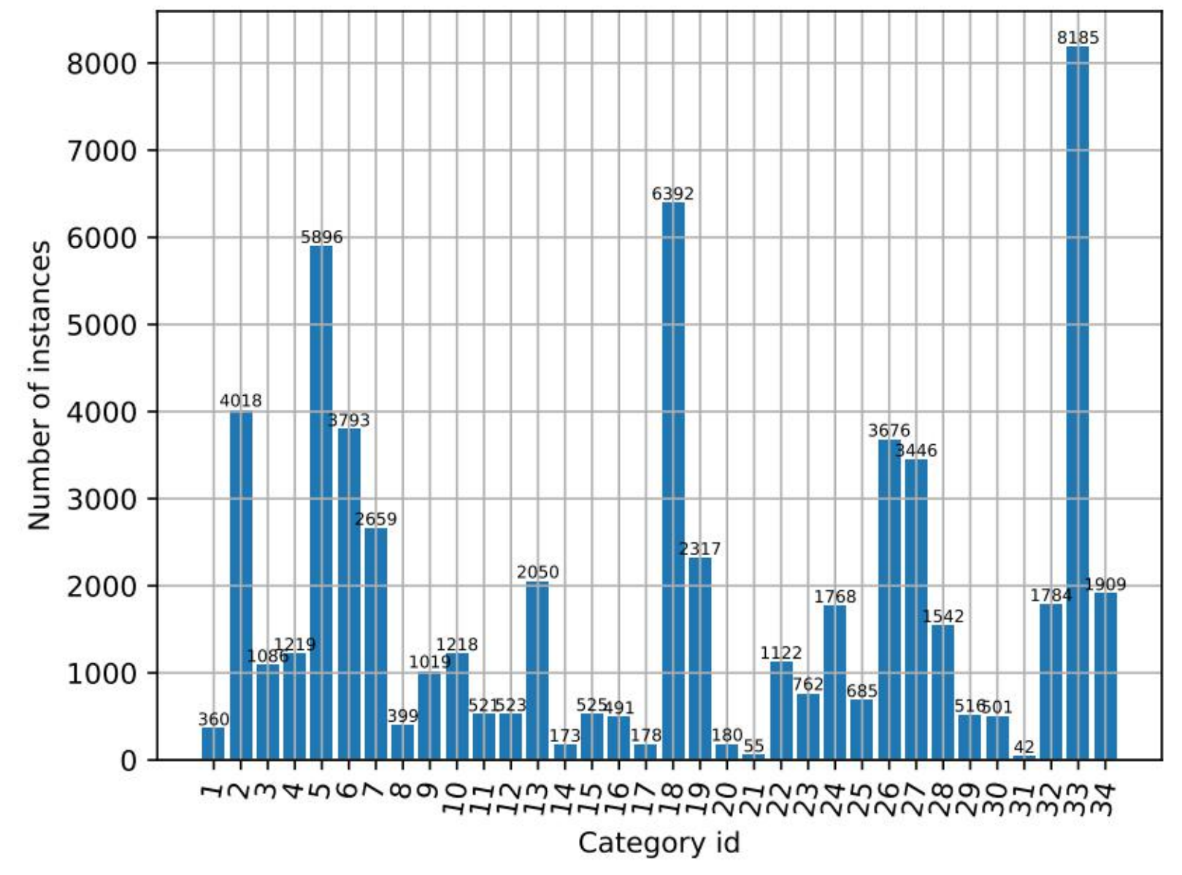

Figure 1: **The number of instances for 34 categories.**

| | H | W | H/W |
|---|---|---|---|
| count | 61010 | 61010 | 61010 |
| 20% | 169 | 157 | 0.555 |
| 40% | 230 | 222 | 0.826 |
| 50% | 258 | 259 | 0.966 |
| 60% | 288 | 303 | 1.103 |
| 80% | 370 | 477 | 1.465 |

Table 1: **Aspect ratio(height/width) of all instances in the training set.**

Compared with COCO with 24% large objects (area>96x96 pixels), 3D-FUTURE has 81.78% large objects.

## Data Process

- **Conservative Image Augmentation**

RandomHorizontalFlip and one of RandomSaturation, RandomContrast, RandomBrightness.

- **Mask Correct**

Some unexplainable pixel labels in mask are observed during the data visualization. We think these mislabeled pixels will cause two problems: (1) Training will be unstable due to wrong labels, (2) The mask fitting will be affected by the noises. There_x0002_fore, masks are corrected before inputting the net.

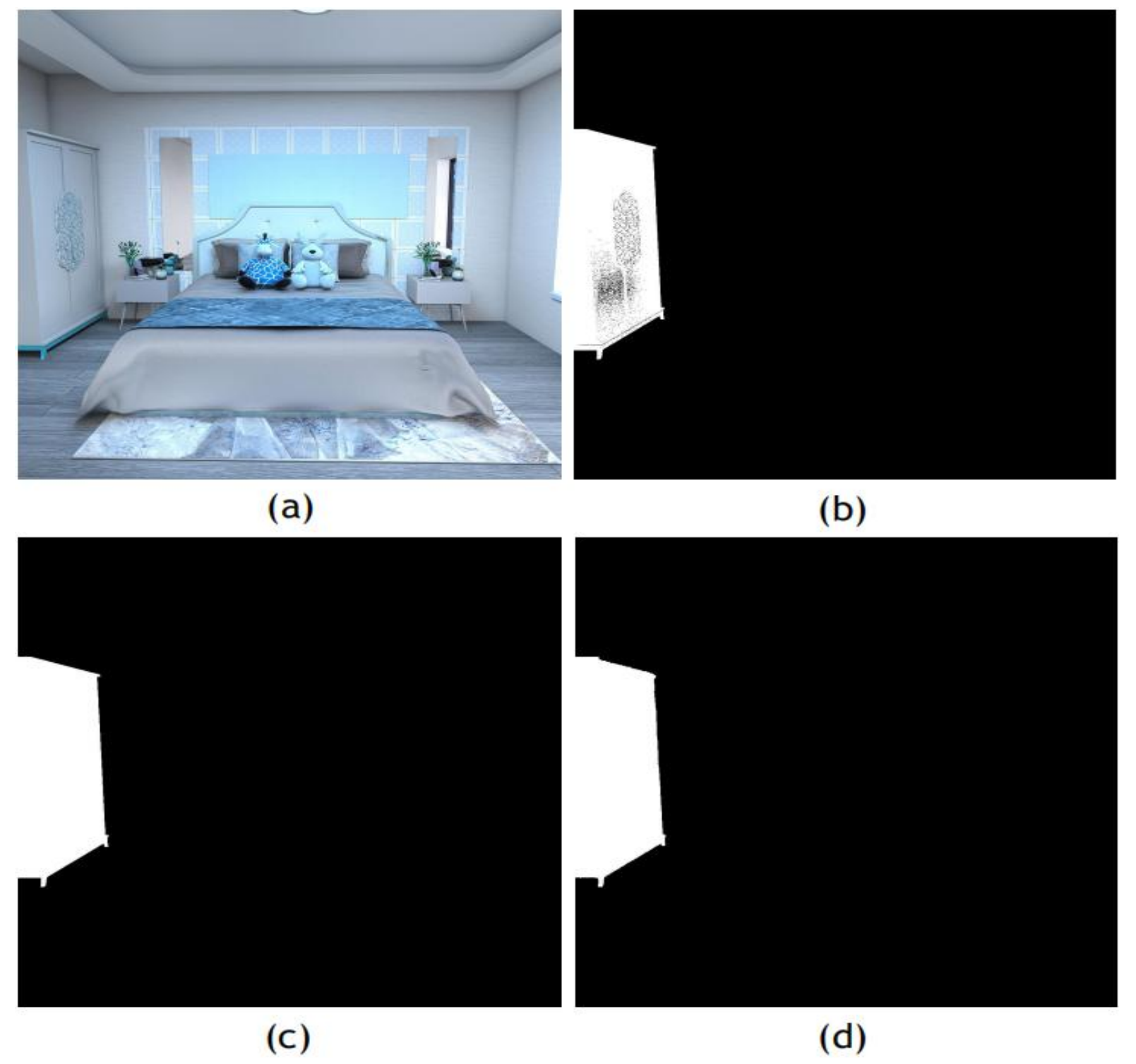

Figure 2: **Example of mask correct.** (a) Original image. (b) Original mask. (c) Corrected mask. (d) mask prediction.

## Method

### Network

- **Overview**

An enhanced PointRend is trained to segment furniture from images. Its backbone is a fusion of ResNeSt, FPN and DCNv2 and this strong engine is responsible for complex features. A cascade technique is used to capture better proposal features. This segmentor is called SPR-Net for short.

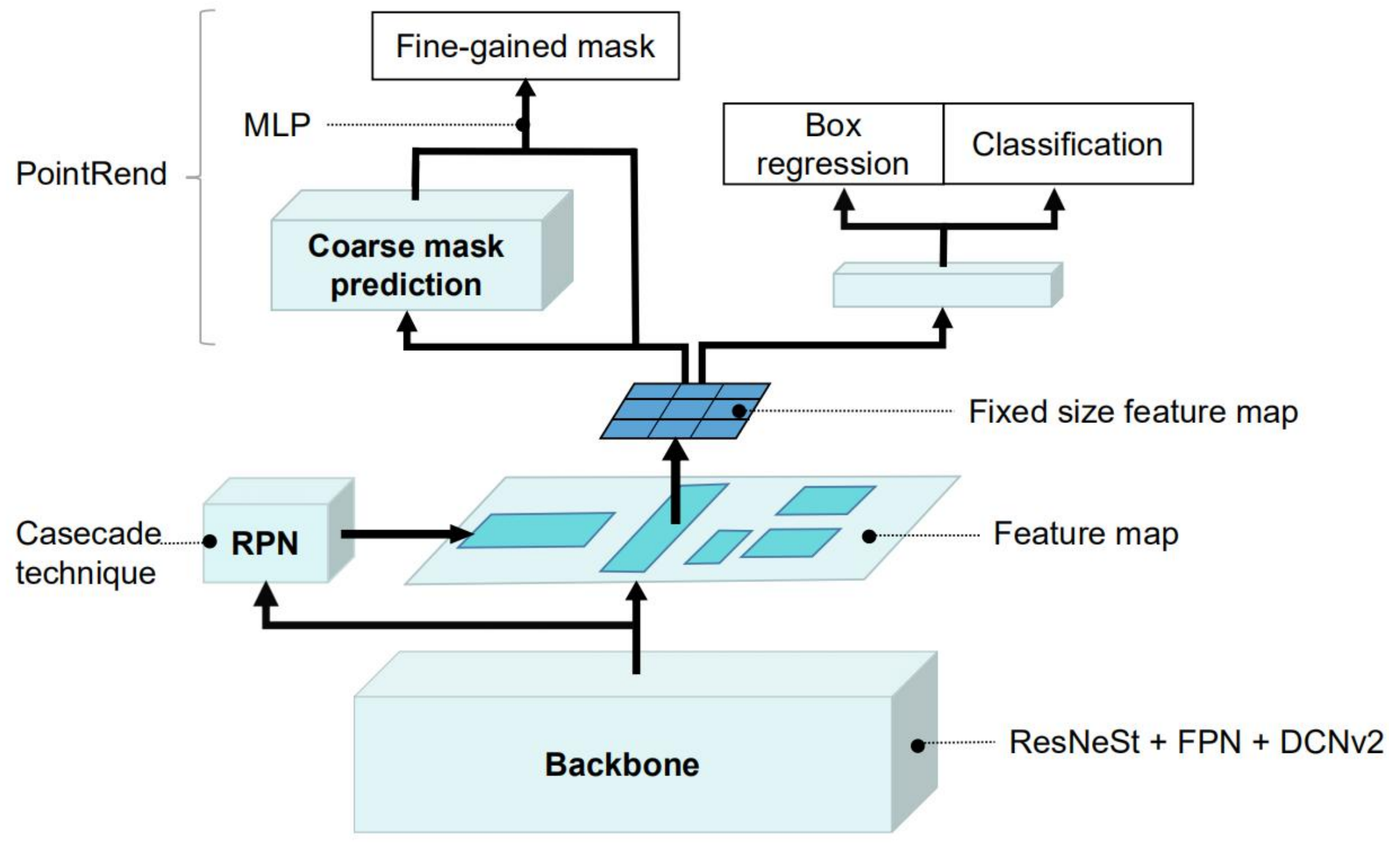

Figure 3: **The structure of SPR-Net.**

- **PointRend**

Due to the detection and segmentation for large objects are relatively easy, the key to segmenting large instance lies in the boundary. Therefore, we use PointRend to refine the boundaries of instance.

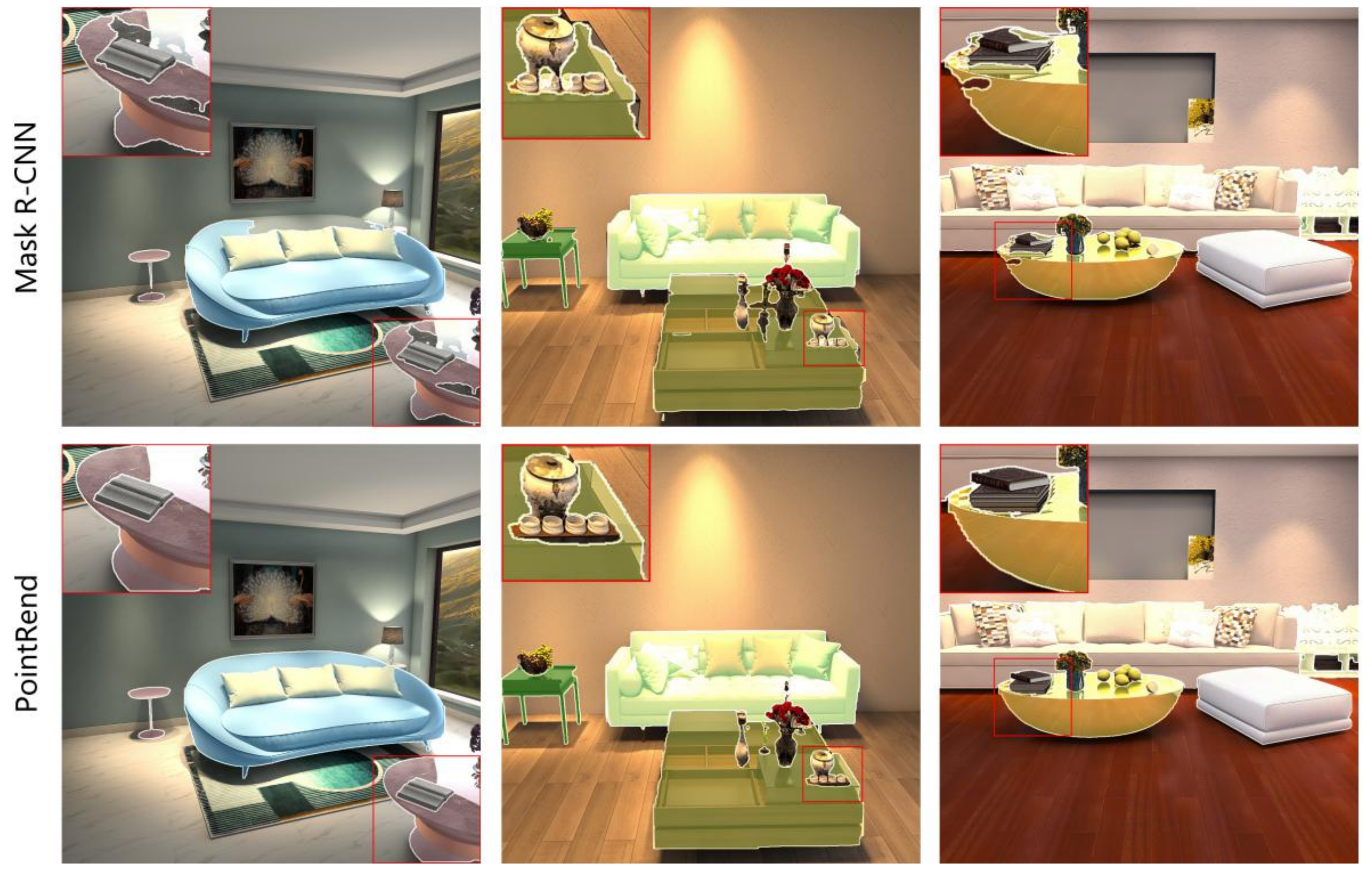

Figure 4: **Examples of segmentation result from Mask R-CNN with default mask head vs. with PointRend, using ResNet101 with FPN.**

### Loss Function

- **Focal Loss**

multi-class focal loss & mask focal loss.

As mentioned above, the dataset has a large class imbalance. In SPR-Net, the cross-entropy loss for classification is replaced by a multi-class focal loss. The categories with a poor mAP and a small number are up-weighted. Finally, SPR-Net equipped with focal loss gives a gain of 0.66 mAP (Table 2).

## Trick

- **Multi-Scale Training**
- **Test Time Augmentation**
- **Model ensemble**

| | X101 | S101 | DCNv2 | Focal | MC | MST | TTA | mAP | AP50 | AP75 | APs | APm | APl |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PointRend+<br>Cascade+<br>FPN | √ | | | | | | | 70.32 | 87.49 | 75.30 | 43.71 | 64.17 | 78.30 |
| | | √ | | | | | | 71.82 | 88.09 | 77.04 | 49.01 | 67.27 | 79.28 |
| | | √ | √ | | | | | 72.76 | 88.39 | 78.07 | 52.03 | 66.61 | 79.55 |
| | | √ | √ | √ | | | | 73.40 | 88.71 | 78.52 | 52.63 | 66.76 | 79.65 |
| | | √ | √ | √ | √ | | | 74.38 | 89.00 | 79.62 | 54.46 | 66.54 | 79.69 |
| | | √ | √ | √ | √ | √ | | 75.53 | 89.92 | 80.55 | 54.68 | 67.36 | 81.63 |
| | | √ | √ | √ | √ | √ | √ | **77.12** | **90.46** | **82.70** | **57.42** | **71.60** | **81.80** |

Table 2: **PointRend's gradual performance improvement on validation set.** ResNetXt101 denoted "X101", ResNeSt101 denoted "S101", Deformable Convolution Network v2 denoted "DCNv2", Mask Correct denoted "MC", Focal Loss denoted "Focal", Multi-Scale Training denoted "MST", Test Time Augmentation denoted "TTA".

## Conclusions

In this report, we present the main details of our scheme that utilizes reasonable data processing, effective models, model ensemble, and other strategies, which gradually increase the leaderboard score step by step, allowing us to achieve the 2nd place in the Instance Segmentation of IJCAI 3D AI Challenge 2020.

## Reference

[1] Zhang, H.,Wu, C.,Zhang, Z.,Zhu, Y. (2020). ResNeSt: Split-Attention Networks.

[2] Kirillov, A., Wu, Y., He, K., & Girshick, R. (2019). Pointrend: image segmentation as rendering.

[3] Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, Piotr. (2017). Focal loss for dense object detection. IEEE Transactions on Pattern Analysis & Machine Intelligence, PP(99), 2999-3007.