上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

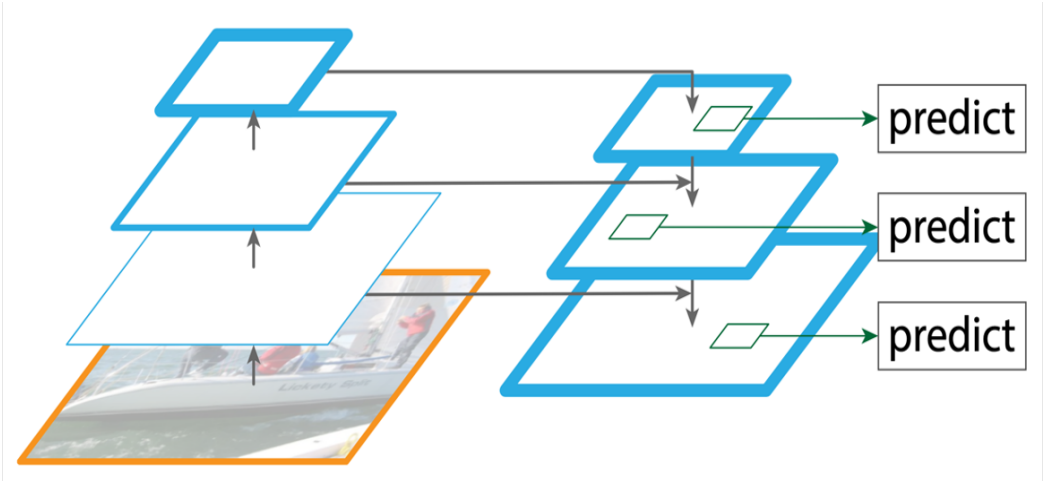
SFPN: Semantic Feature Pyramid Network for Object Detection

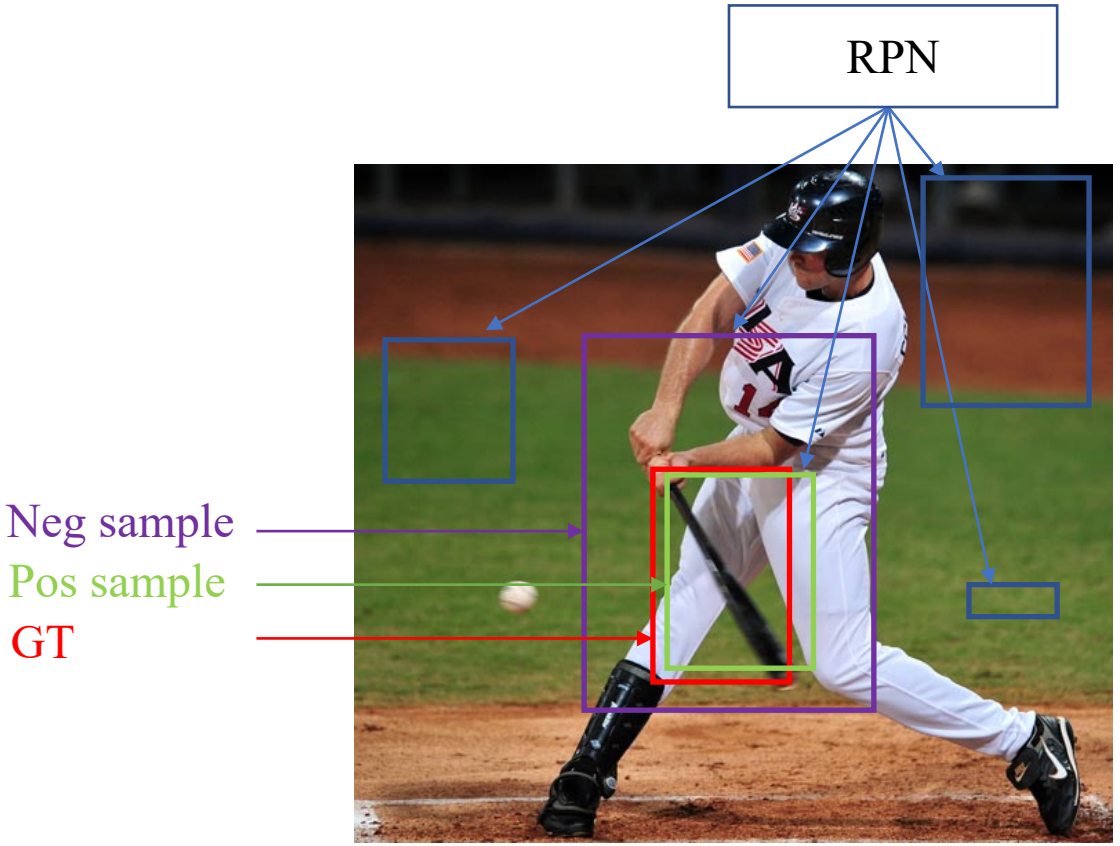
Gan, Yi; Xu, Wei; Su, Jianbo
Shanghai Jiao Tong University

Content

- problem
- motivation
- architecture
- result
- conclusion

01 problem



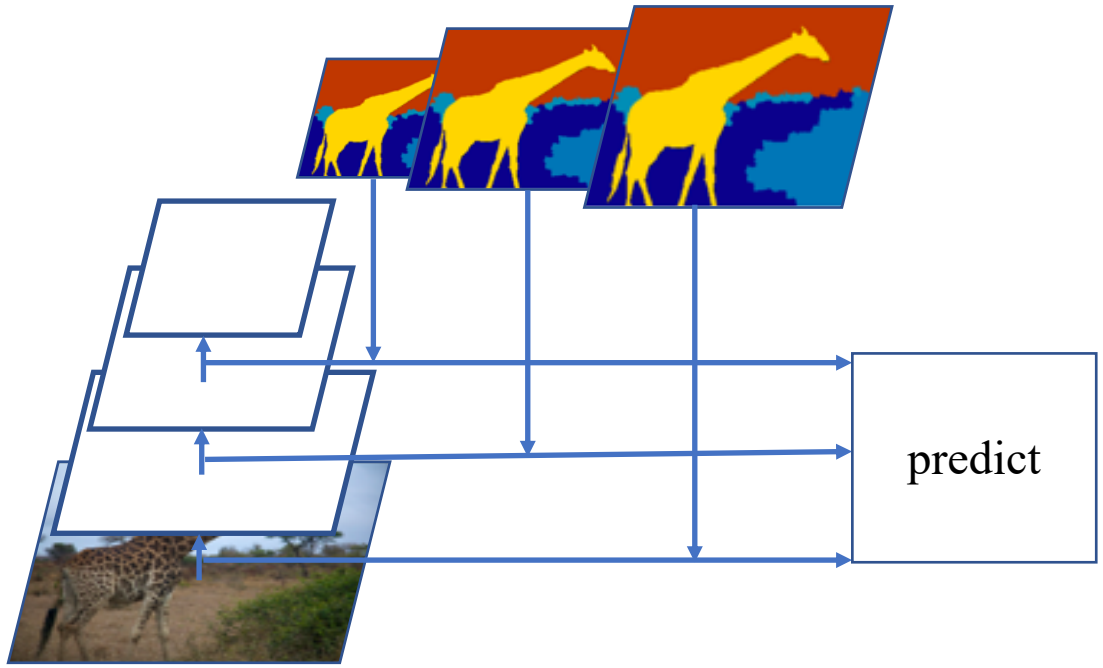


Problem of FPN:
Dilution issue -> low level lack of semantic information
Stacked convs and poolings -> high level lack of localization information

02 Motivation

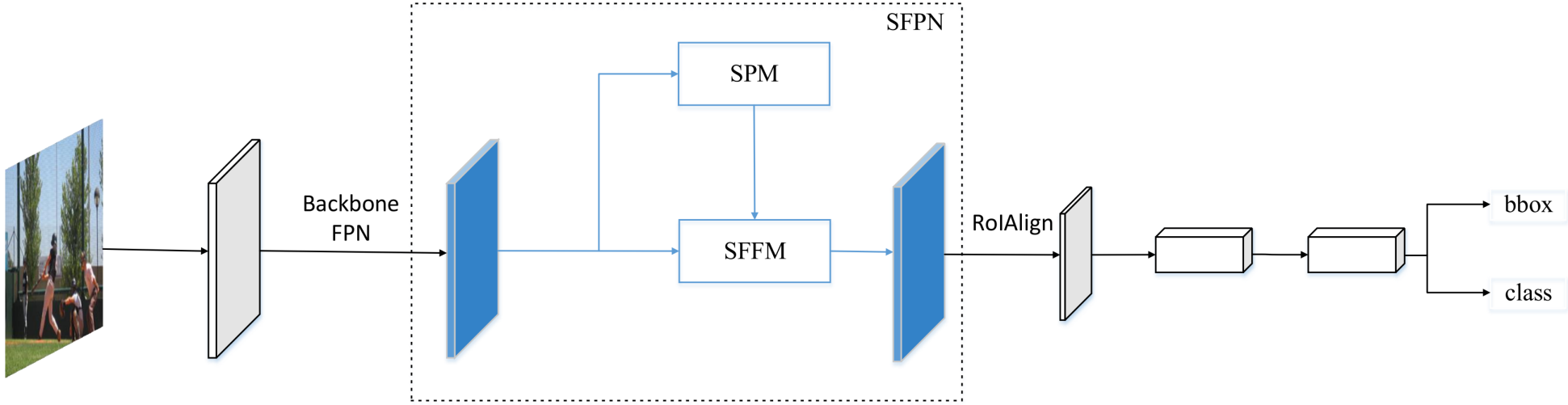
advantage of semantic segmentation:

- It consists of both high level semantic information and low level localization details.
- More scale-invariant than detection



03 Architecture

overall structure



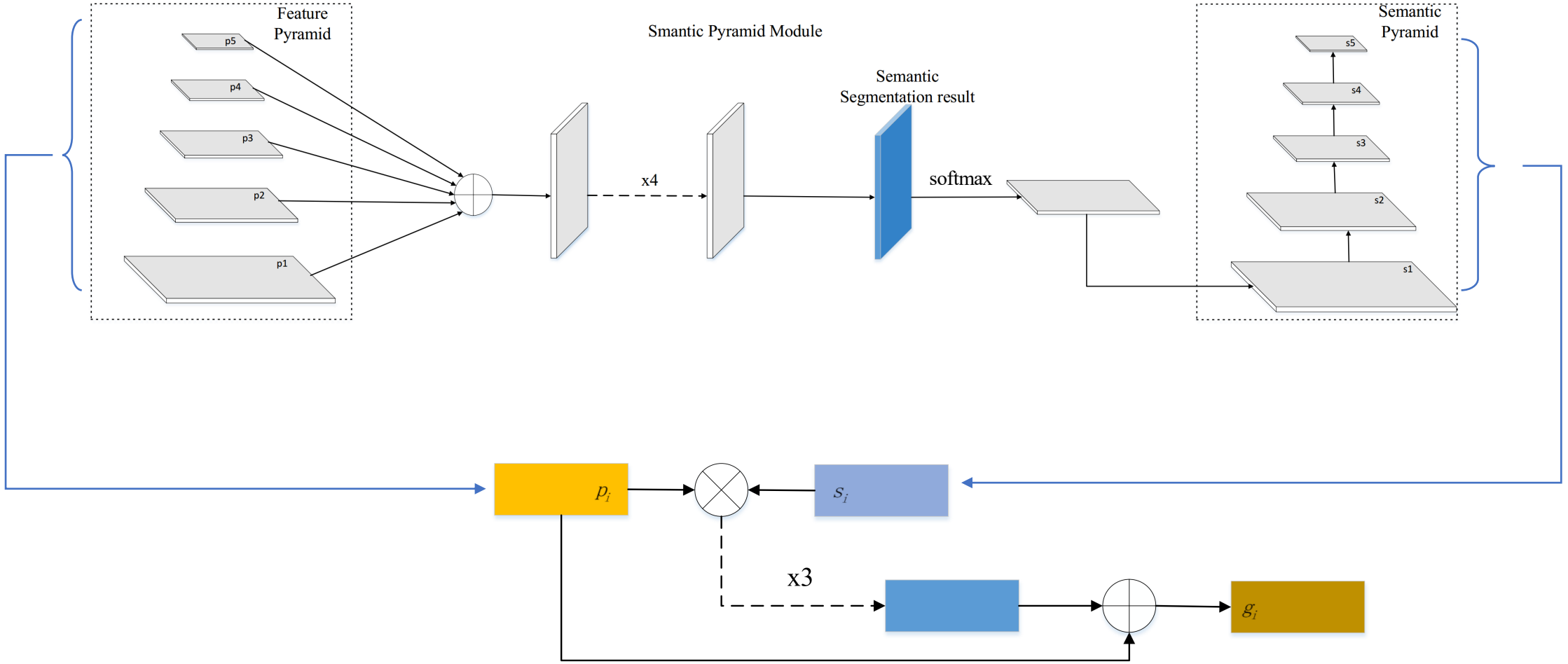
loss function

$$L = L_{bbox} + \alpha L_{segm},$$
$$L_{bbox} = L_{cls}(c_t, \hat{c}_t) + L_{loc}(r_t, \hat{r}_t),$$
$$L_{segm} = CE(e_t, \hat{e}_t).$$

$$L_{loc}(r_t, \hat{r}_t) = smooth\_L1(r_t, \hat{r}_t),$$
$$L_{cls}(c_t, \hat{c}_t) = CE(c_t, \hat{c}_t),$$

03 Architecture

structure details



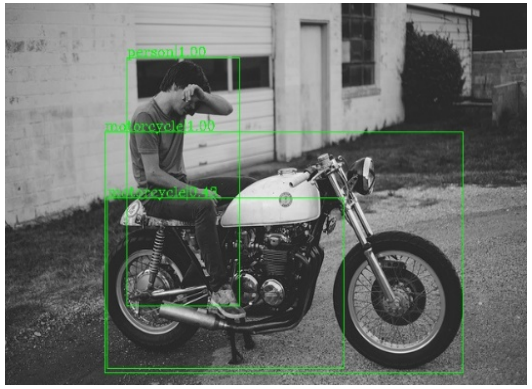
04 Result

TABLE I
COMPARISONS WITH STATE-OF-THE-ART METHODS ON COCO TEST-DEV.

Method	Backbone	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
One-Stage							
YOLOv2 [21]	DarkNet-19	21.6	44.0	19.2	5.0	22.4	35.5
SSD [22]	VGG-16	28.8	48.5	30.3	10.9	31.8	43.5
DSSD [37]	ResNet-101	31.2	50.4	33.3	10.2	34.5	49.8
RetinaNet [23]	ResNet-101	39.1	59.1	42.3	21.8	42.7	50.2
RefineDet [26]	ResNet-101	36.4	57.7	39.5	16.6	39.9	51.4
AlignDet [24]	ResNet-101	42.0	62.4	46.5	24.6	44.8	53.3
Anchor-Free							
RPDet [29]	ResNet-101	41.0	62.9	44.3	23.6	44.1	51.7
CornerNet [27]	Hourglass-104	40.5	56.5	43.1	19.4	42.7	53.9
ExtremeNet [38]	Hourglass-104	40.1	55.3	43.2	20.3	43.2	53.1
FCOS [39]	ResNeXt-101	42.1	62.1	45.2	25.6	44.9	52.0
FSAF [30]	ResNeXt-101	42.9	63.8	46.3	26.6	46.2	52.7
FoveaBox [40]	ResNeXt-101	42.1	61.9	45.2	24.9	46.8	55.6
Two-Stage							
Faster R-CNN [13]	ResNet-101	36.2	59.1	39.0	18.2	39.0	48.2
Mask R-CNN [1]	ResNet-101	38.2	60.3	41.7	20.1	41.1	50.2
Deformable R-FCN [41]	Aligned-inception-ResNet	37.5	58.0	40.8	19.4	40.1	52.5
Libra R-CNN [35]	ResNeXt-101	43.0	64.0	47.0	25.3	45.6	54.6
Cascade R-CNN [14]	ResNet-101	42.8	62.1	46.3	23.7	45.5	55.2
Grid R-CNN [18]	ResNeXt-101	43.2	63.0	46.6	25.1	46.5	55.2
TridentNet [42]	ResNet-101	42.7	63.6	46.5	23.9	46.6	56.6
Faster R-CNN∗	ResNet-50	36.5	58.7	39.1	27.5	39.7	44.6
Faster R-CNN∗	ResNet-101	38.9	60.9	42.3	22.4	42.4	48.3
Mask R-CNN∗	ResNet-50	37.5(34.4)	59.4(56.3)	40.6(36.6)	22.1(18.6)	40.6(37.2)	46.2(44.5)
Mask R-CNN∗	ResNet-101	39.8(36.3)	61.6(58.5)	43.3(38.7)	22.9(19.2)	43.2(39.3)	49.7(47.4)
Cascade R-CNN∗	ResNet-101	42.4	61.1	46.1	23.6	45.4	54.1
Faster RCNN w SFPN	ResNet-50	38.3[**+1.8**]	60.3	41.6	21.8	40.5	49.2
Faster RCNN w SFPN	ResNet-101	40.3[**+1.4**]	62.5	43.9	22.7	43.3	51.9
Mask RCNN w SFPN	ResNet-50	39.3[**+1.8**](35.8[**+1.4**])	61.1(57.9)	42.7(38.2)	22.9(16.8)	42.0(37.6)	49.4(50.8)
Mask RCNN w SFPN	ResNet-101	41.1[**+1.3**](37.3[**+1.0**])	62.9(59.6)	44.9(40.0)	23.0(16.8)	44.0(39.3)	52.6(53.8)
Cascade RCNN w SFPN	ResNet-101	43.5[**+1.1**]	62.1	47.2	23.8	46.0	56.0

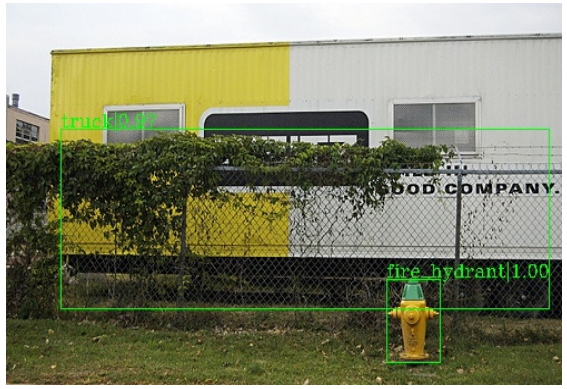
04 Result


remote|0.99
sink|0.52
remote|0.41

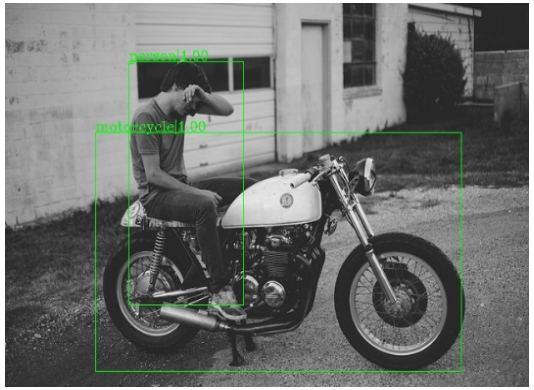

person|1.00
motorcycle|1.00

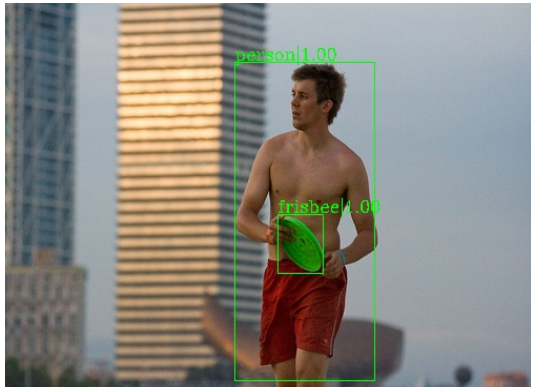

person|1.00
frisbee|1.00
chair|0.40


bird|0.36
boat|0.37
boat|0.40


truck|0.47
fire hydrant|1.00


remote|0.94


person|1.00
motorcycle|1.00


person|1.00
frisbee|1.00


truck|0.33
fire hydrant|1.00

05 Conclusion

SFPN: Semantic Feature Pyramid Network for Object Detection

Pros：

- Imporves the overall mAP
- Can be applied to multiple tasks

Cons：

- Semantic branch brings additional computation cost
- Semantic segmentation labels are needed

Future works：

- Reduce the dependency of semantic segmentation label
- Multi-task network