Paper

1. Panoptic Segmentation
2. UPSNet
3. Panoptic Feature Pyramid Networks
4. Panoptic-DeepLab

01 Panoptic Segmentation

- CVPR2019
- Facebook AI Research (FAIR)、 HCI (海德堡图像处理中心)

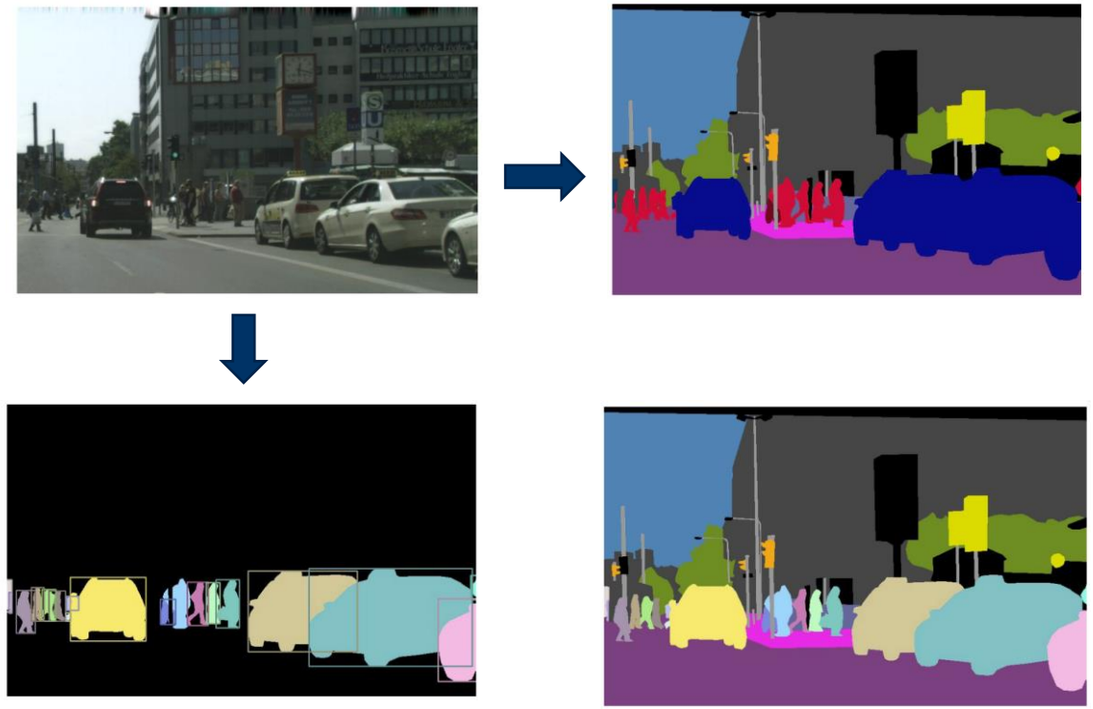
- **Contribution**
  - New task: Panoptic Segmentation
  - New metrics: panoptic quality(PQ)
  - Basic algorithm： combine PSPNet and Mask R-CNN

01 Panoptic Segmentation

- **Panoptic, Instance, Semantic Segmentation**

Things: people, animals, tools – received the dominant share of attention

Stuff : grass, sky, road – amorphous regions of similar texture



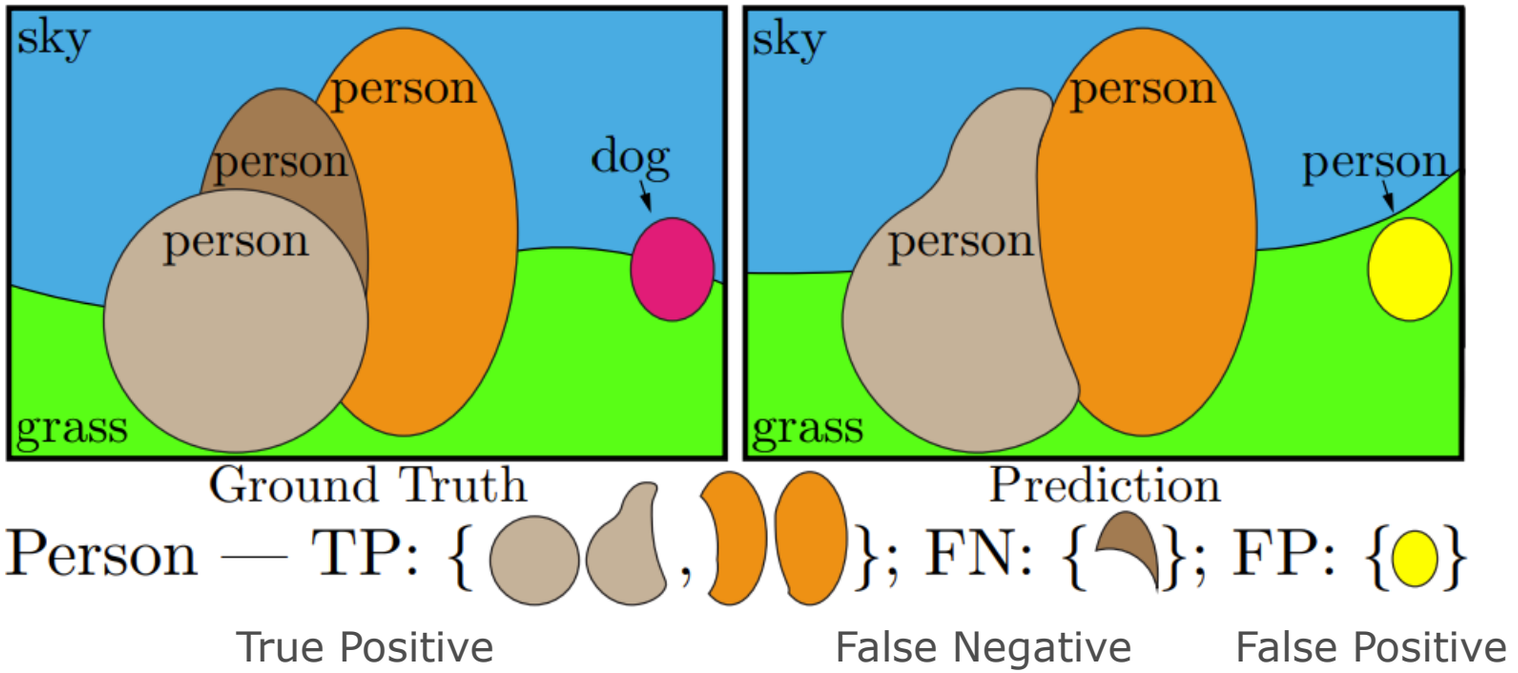
Task format:

$$(l_i, z_i) \in \mathcal{L} \times \mathbb{N}$$

01 Panoptic Segmentation

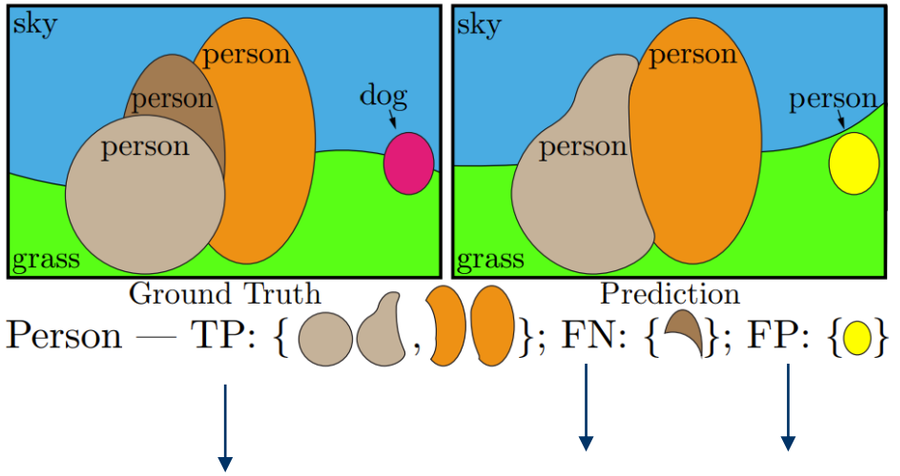
- **Panoptic Quality**
  - Segment Matching

IoU is greater than 0.5



01 Panoptic Segmentation

- **Panoptic Quality**
  - PQ Computation



4 2 1

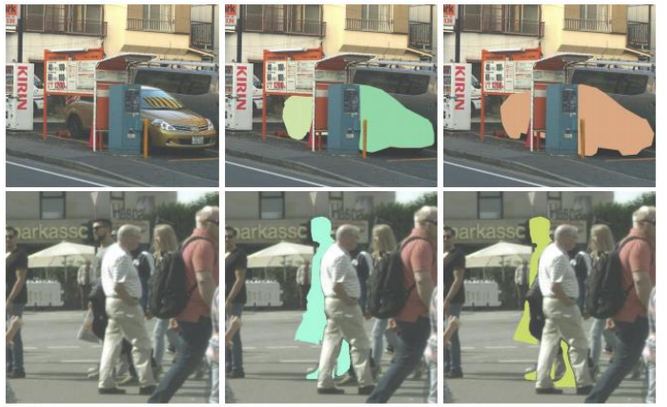
匹配得分：

$$\mathrm{PQ} = \frac{\sum_{(p,g)\in TP} \mathrm{IoU}(p,g)}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}$$

$$RQ = \frac{4}{4+1+0.5}$$

$$\mathrm{PQ} = \underbrace{\frac{\sum_{(p,g)\in TP} \mathrm{IoU}(p,g)}{|TP|}}_{\text{segmentation quality (SQ)}} \times \underbrace{\frac{|TP|}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}}_{\text{recognition quality (RQ)}}.$$

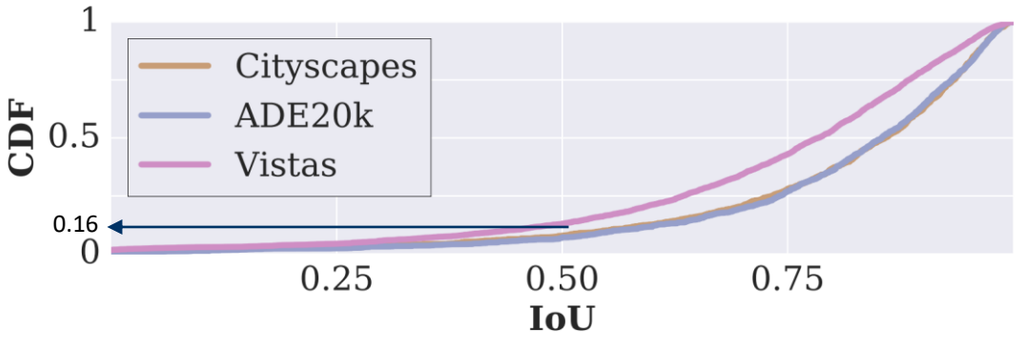
01 Panoptic Segmentation



- **Human Consistency Study**

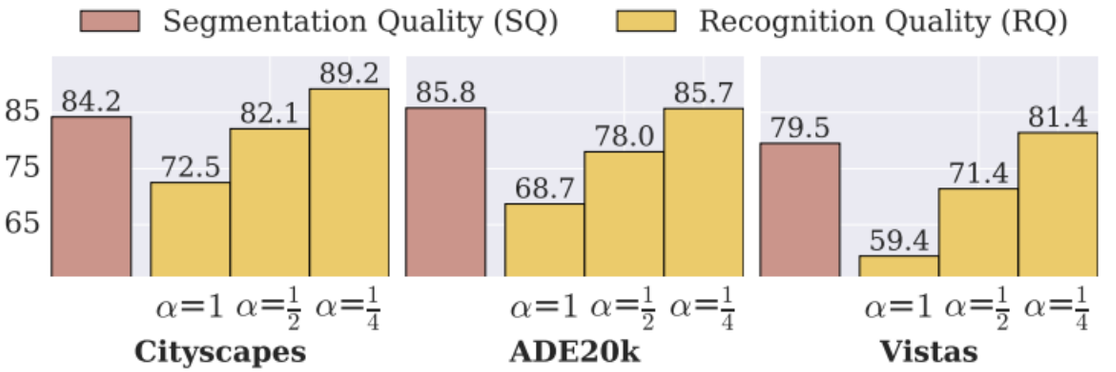
（1）IOU threshold

二分图最大权匹配得到的结果:



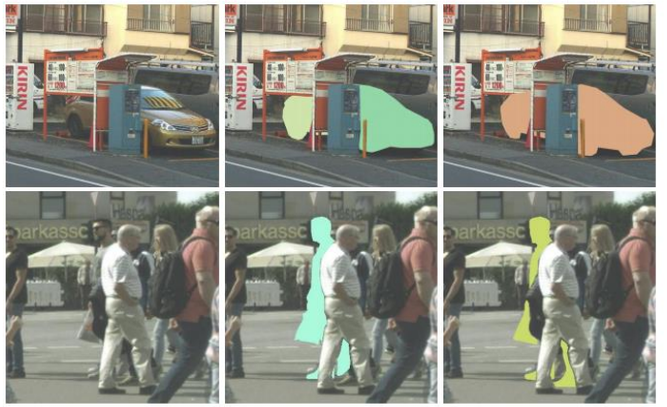
（2）SQ vs. RQ balance

$$\mathrm{RQ}^{\alpha} = \frac{|TP|}{|TP| + \alpha|FP| + \alpha|FN|}.$$



01 Panoptic Segmentation

- **Human Consistency Study**



(3) Stuff vs. things

	PQ	PQ^{St}	PQ^{Th}	SQ	SQ^{St}	SQ^{Th}	RQ	RQ^{St}	RQ^{Th}
Cityscapes	69.7	71.3	67.4	84.2	84.4	83.9	82.1	83.4	80.2
ADE20k	67.1	70.3	65.9	85.8	85.5	85.9	78.0	82.4	76.4
Vistas	57.5	62.6	53.4	79.5	81.6	77.9	71.4	76.0	67.7

(4) Small vs. large objects

	PQ^{S}	PQ^{M}	PQ^{L}	SQ^{S}	SQ^{M}	SQ^{L}	RQ^{S}	RQ^{M}	RQ^{L}
Cityscapes	35.1	62.3	84.8	67.8	81.0	89.9	51.5	76.5	94.1
ADE20k	49.9	69.4	79.0	78.0	84.0	87.8	64.2	82.5	89.8
Vistas	35.6	47.7	69.4	70.1	76.6	83.1	51.5	62.3	82.6

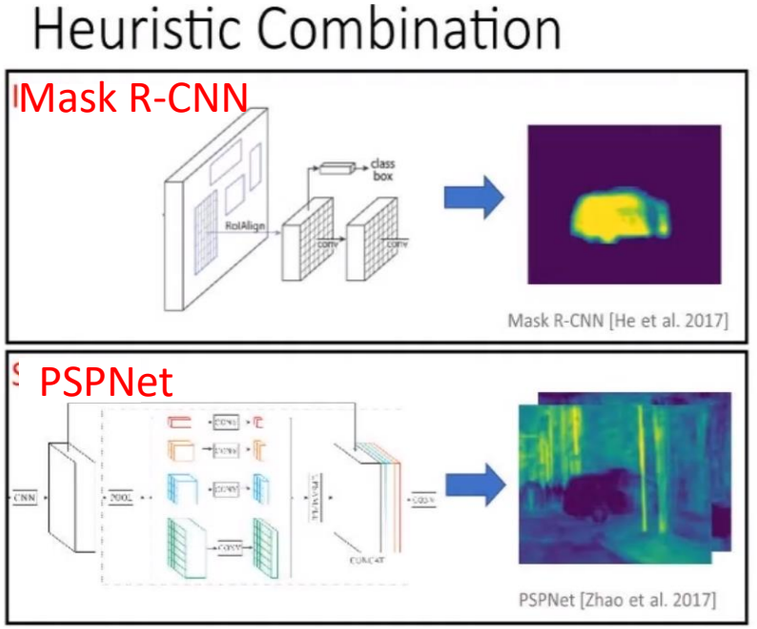
01 Panoptic Segmentation

- **Basic algorithm： Combine PSPNet and Mask R-CNN**

（1）we combine those segments with semantic segmentation results by resolving any overlap between thing and stuff classes in favor of the thing class

（2） Starting from the most confident instance.

（3）For each instance, remove pixels which have been assigned to previous segments, if a sufficient fraction of the segment remains, accept the non-overlapping portion

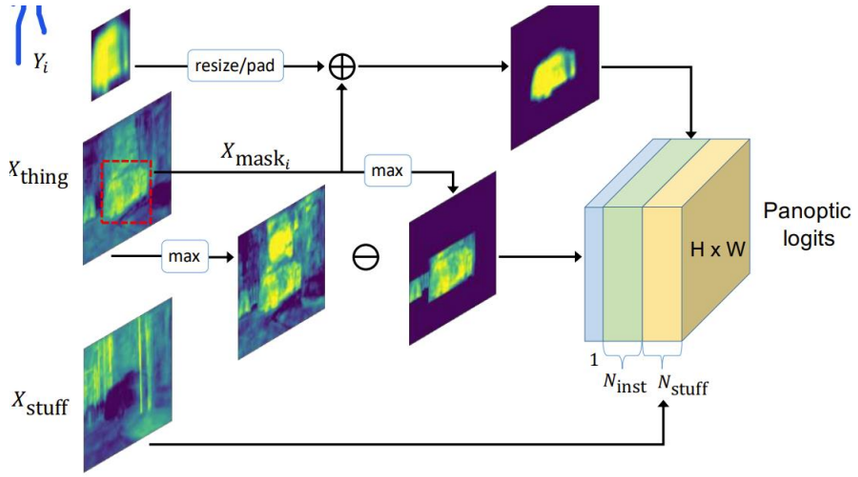


02 UPSNet—A Unified Panoptic Segmentation Network

- CVPR2019
- Uber ATG

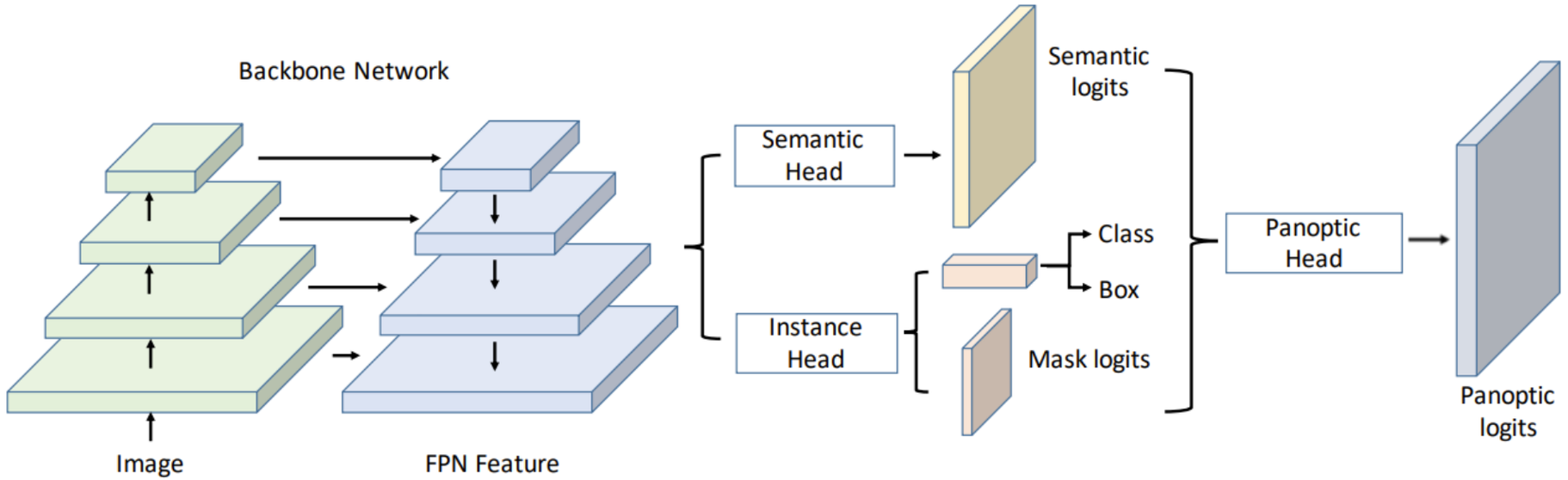
  University of Toronto

  The Chinese University of Hong Kong



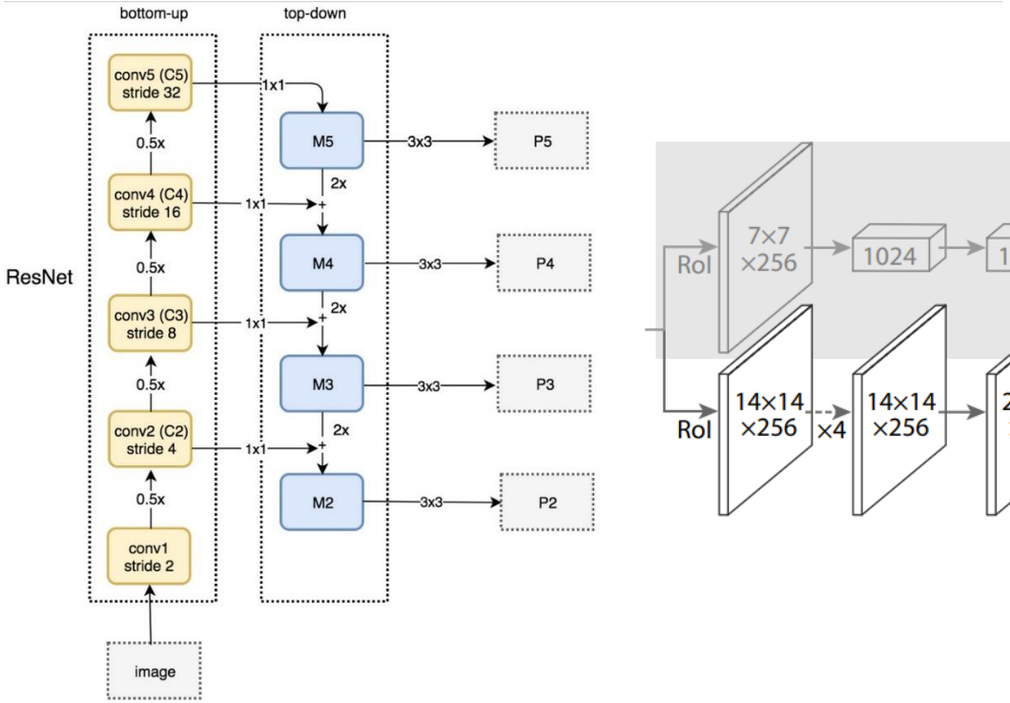
02 UPSNet—A Unified Panoptic Segmentation Network

- **Contribution** 单个backbone+两个轻量网络头部可以获得很好的语义和实例分割
  - A end-to-end network: Unified Panoptic Segmentation Network



02 UPSNet—A Unified Panoptic Segmentation Network

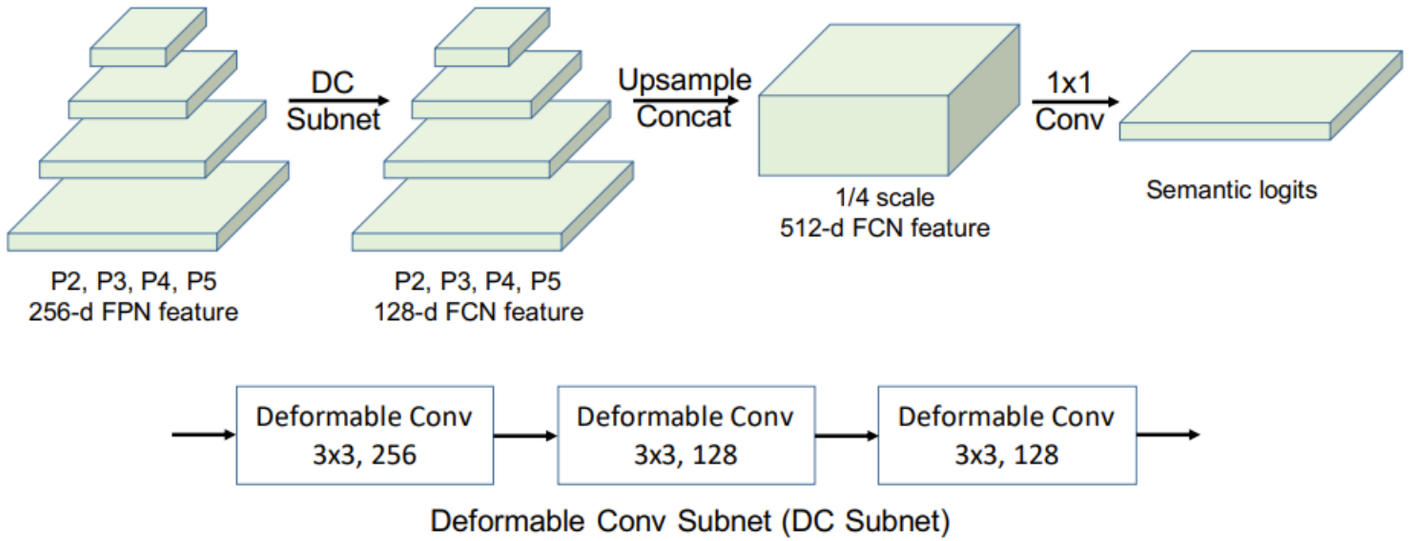
- **Instance Head (Mask R-CNN)**



$$k = \lfloor k_0 + \log_2(\sqrt{wh}/224) \rfloor.$$

02 UPSNet—A Unified Panoptic Segmentation Network

- **Semantic Head**



Deformable convolution

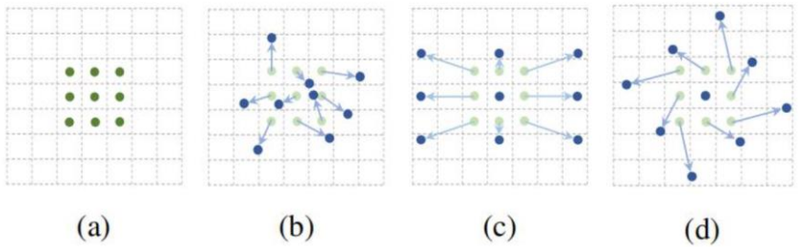
+

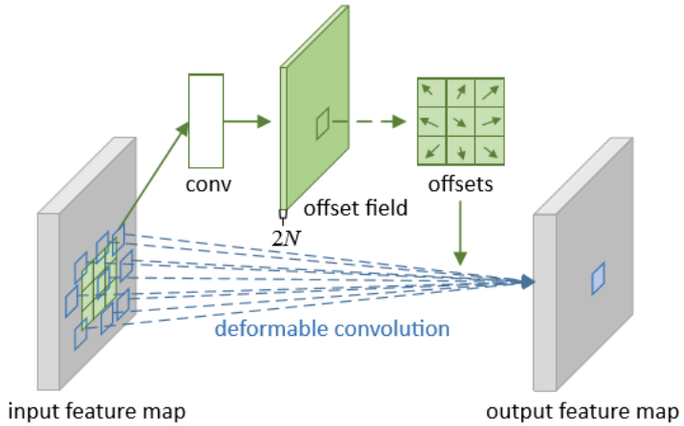
Multi scale feature cascade

→ **PSPNet**

02 UPSNet—A Unified Panoptic Segmentation Network

- **Semantic Head**
  - Deformable convolution





deformation modules	DeepLab mIoU@V / @C	class-aware RPN mAP@0.5 / @0.7	Faster R-CNN mAP@0.5 / @0.7	R-FCN mAP@0.5 / @0.7
atrous convolution (2,2,2) (default)	69.7 / 70.4	68.0 / 44.9	78.1 / 62.1	80.0 / 61.8
atrous convolution (4,4,4)	73.1 / 71.9	72.8 / 53.1	78.6 / 63.1	80.5 / 63.0
atrous convolution (6,6,6)	73.6 / 72.7	73.6 / 55.2	78.5 / 62.3	80.2 / 63.5
atrous convolution (8,8,8)	73.2 / 72.4	73.2 / 55.1	77.8 / 61.8	80.3 / 63.2
deformable convolution	**75.3 / 75.2**	**74.5 / 57.2**	78.6 / 63.3	81.4 / 64.7

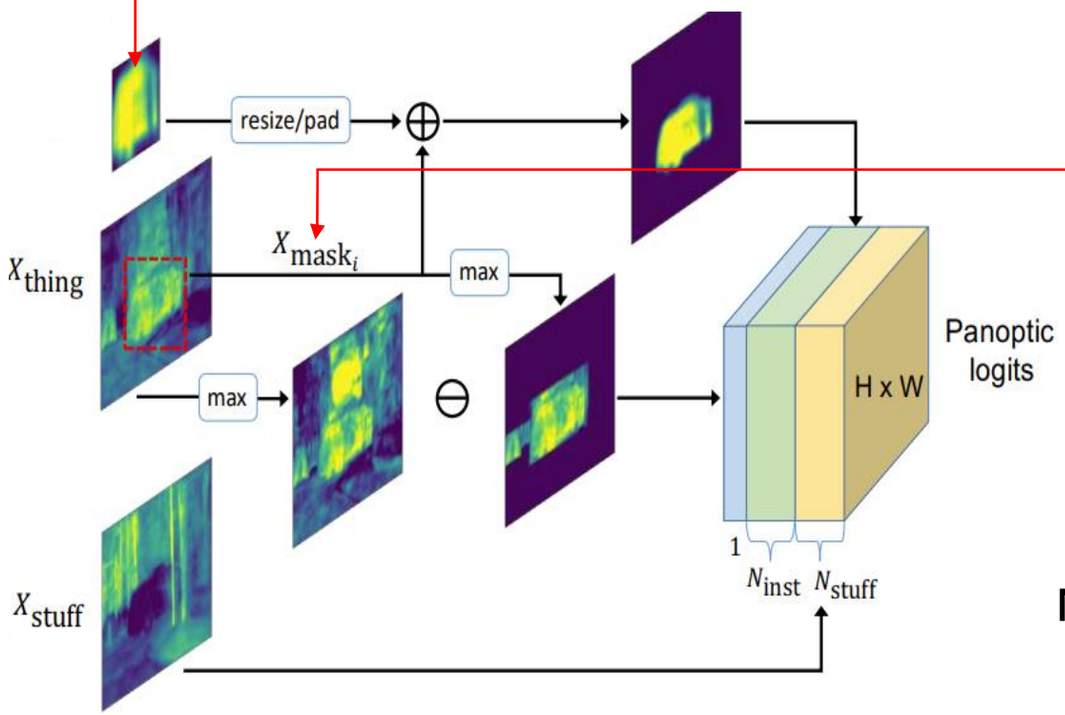
02 UPSNet—A Unified Panoptic Segmentation Network

- **Panoptic Head**

mask大小为28*28，将其reshape成对应b-box

Instance head:

Semantic head：



X_{maski}：由Semantic结果里面对应通道上对应b-box的点组成

N_{inst}

train: 由ground-truth里面实例个数给定

predict: 由mask的个数确定

02 UPSNet—A Unified Panoptic Segmentation Network

- **8 Loss function**

  - **RPN**
    box classification : cross entropy loss
    box regression : smoth L1

  - **Instance Head**
    box classification : cross entropy loss
    box regression : smoth L1
    mask segmentation : cross entropy loss

  - **Semantic Head**
    pixel-wise classification loss: cross entropy loss
    ROL Loss: cross entropy loss

  - Panoptic **Head**
    pixel-wise classification loss: cross entropy loss

ROL Loss:

To put more emphasis on the foreground objects

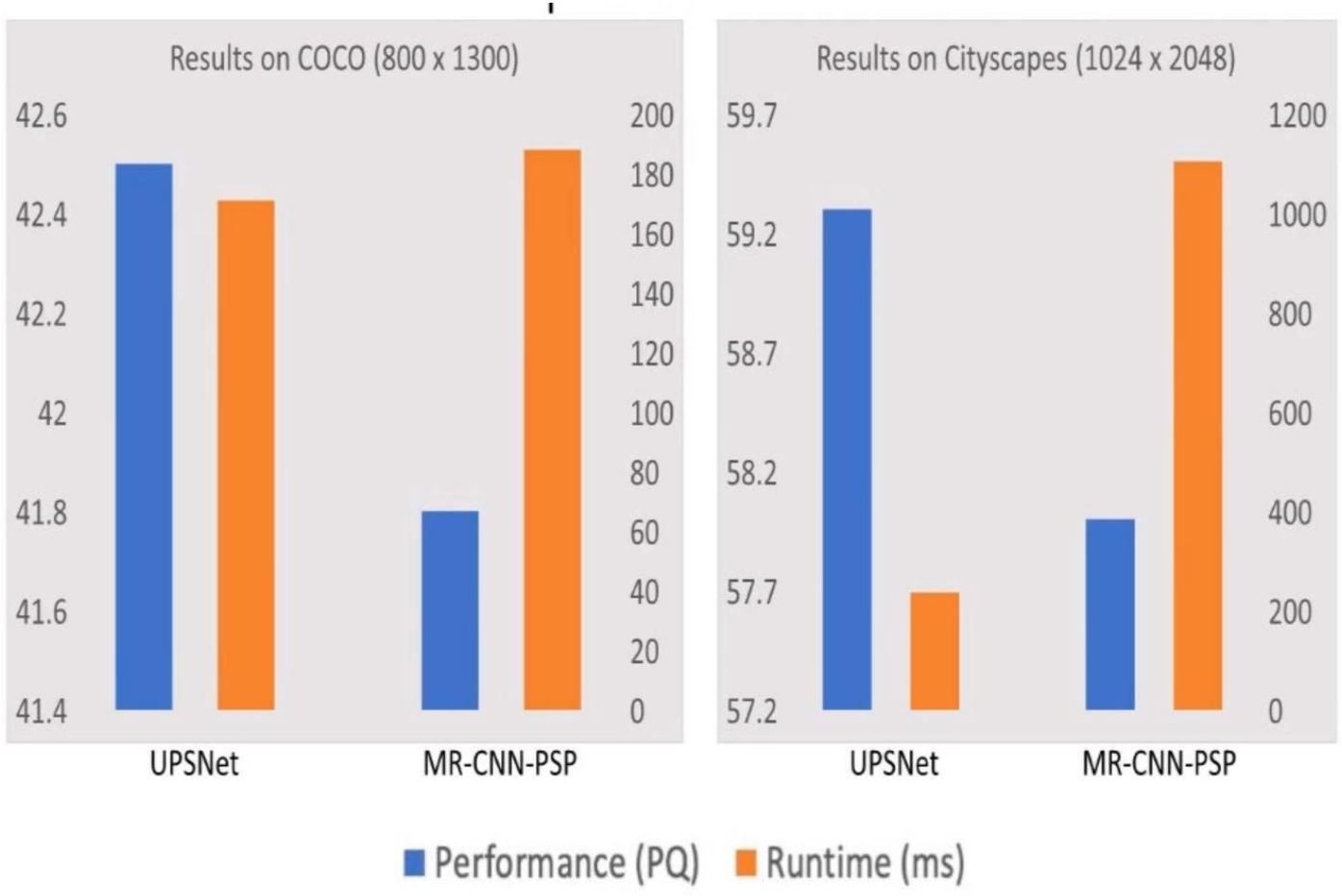
1、use the ground truth bounding box of the instance to crop the sementic logits map

2、 resize it to 28 × 28

3、 cross entropy loss computed over 28 × 28 patch

02 UPSNet—A Unified Panoptic Segmentation Network

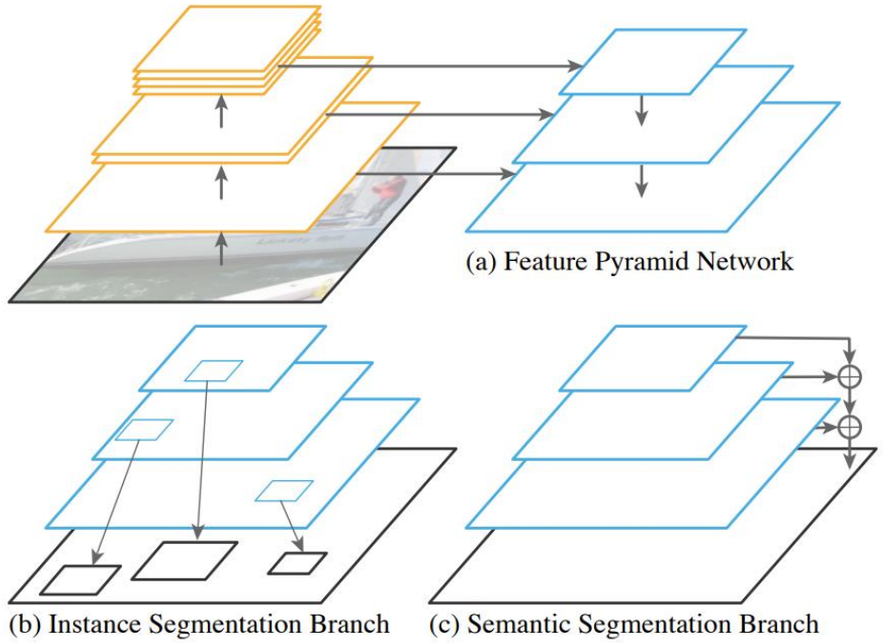
- **Results on COCO and Cityscapes**



03 Panoptic Feature Pyramid Networks

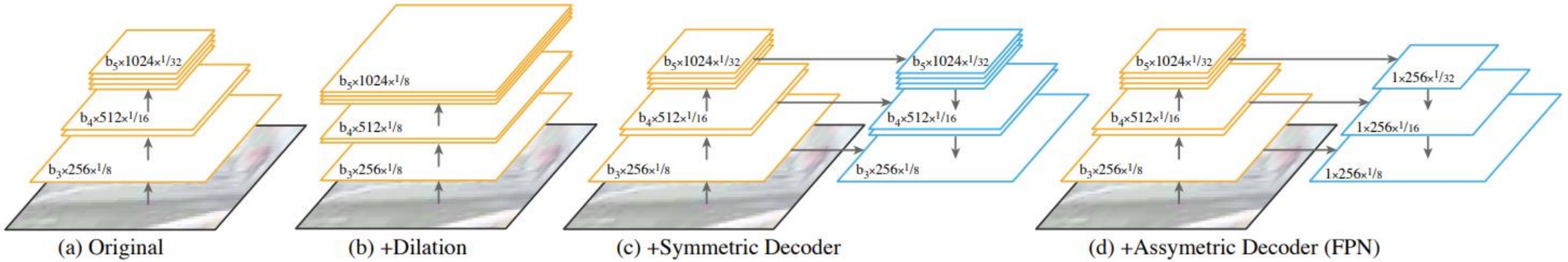
- CVPR2019
- Facebook AI Research (FAIR)

- **Contribution**
  - a single network for instance and semantic segmentation tasks



03 Panoptic Feature Pyramid Networks

- **backbone**



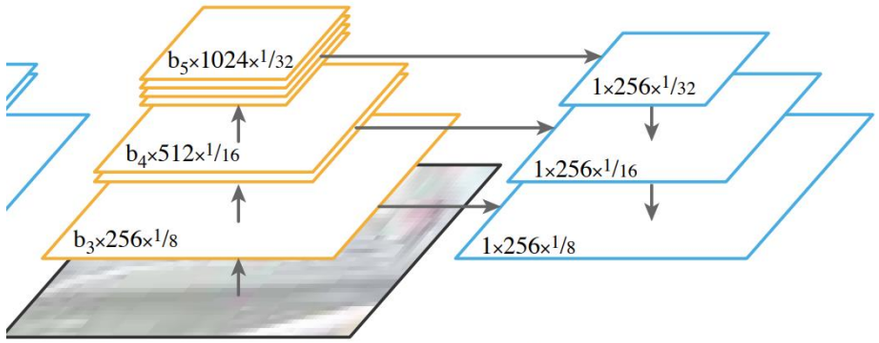
symmetric decoder

asymmetric, lightweight decoder

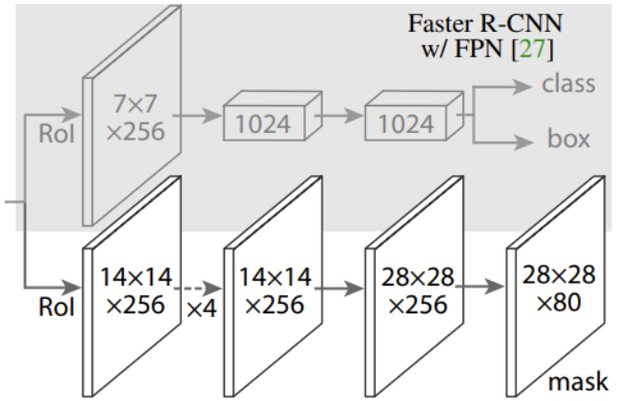
top-down pathway has only one block

03 Panoptic Feature Pyramid Networks

- **Instance segmentation branch**
  - The design of FPN same channel applies a shared network branch to predict a refined box and class label for each region.



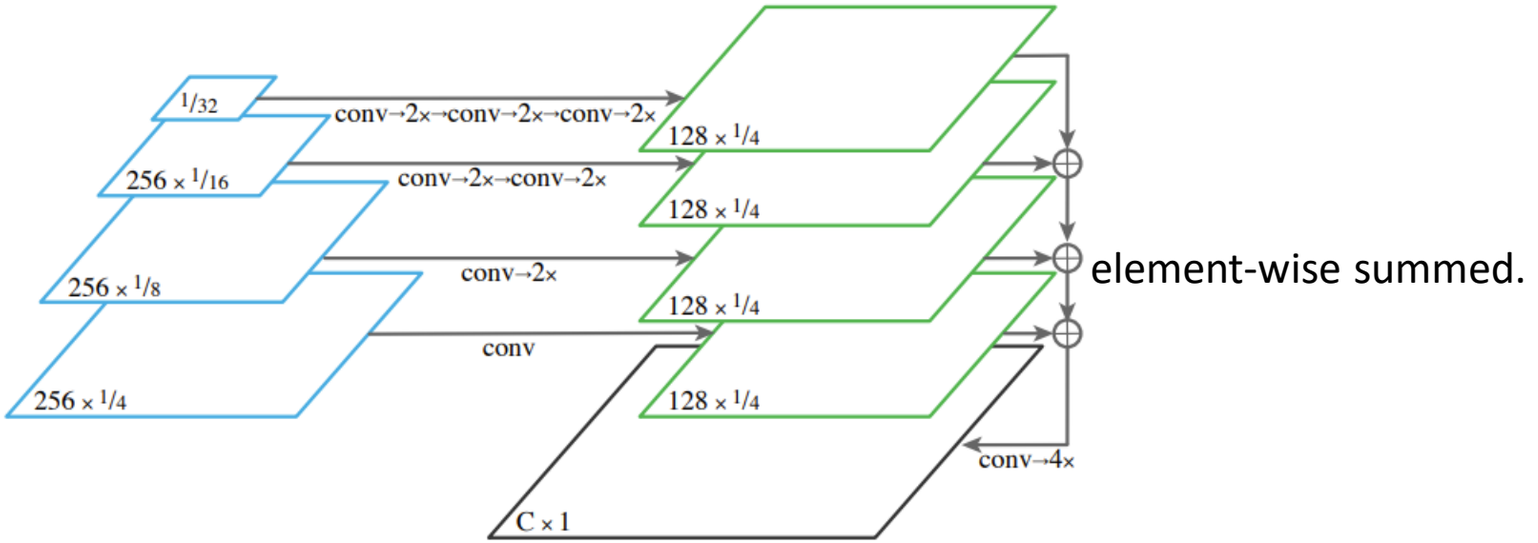
(d) +Assymetric Decoder (FPN)



'×4' denotes a stack of four consecutive convs.

03 Panoptic Feature Pyramid Networks

- **Semantic segmentation branch**



03 Panoptic Feature Pyramid Networks

- **Loss**

$$L = \lambda_{\mathrm{i}} \left(L_{\mathrm{c}} + L_{\mathrm{b}} + L_{\mathrm{m}} \right) + \lambda_{\mathrm{s}} L_{\mathrm{s}}$$

Instance：

λ_s	mIoU	AP	AP_{50}	AP_{75}	PQ^{Th}
0.0	-	33.9	55.6	35.9	46.6
0.1	37.2	**34.0**	55.6	**36.0**	**46.8**
0.25	39.6	33.7	55.3	35.5	46.1
0.5	41.0	33.3	54.9	35.2	45.9
0.75	41.1	32.6	53.9	34.6	45.0
1.0	41.5	32.1	53.2	33.6	44.6
		+0.1	*+0.0*	*+0.1*	*+0.2*

(a) Panoptic FPN on COCO for **instance** segmentation ($\lambda_i = 1$).

λ_s	mIoU	AP	AP_{50}	PQ^{Th}
0.0	-	32.2	58.7	51.3
0.1	68.3	32.5	59.2	52.9
0.25	71.8	32.8	59.6	52.7
0.5	72.0	32.7	59.5	52.9
0.75	73.4	32.8	58.8	52.3
1.0	74.2	**33.2**	**59.7**	**52.4**
		+1.0	*+1.0*	*+1.1*

(b) Panoptic FPN on Cityscapes for **instance** segmentation ($\lambda_i = 1$).

Semantic：

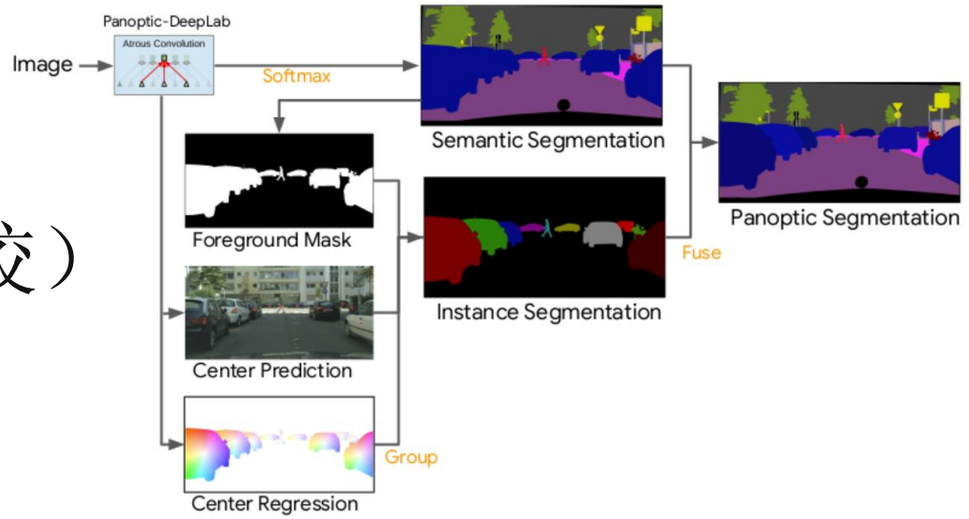
λ_i	AP	mIoU	fIoU	PQ^{St}
0.0	-	40.2	67.2	27.9
0.1	20.1	40.6	67.5	28.4
0.25	25.5	41.0	67.8	28.6
0.5	29.2	41.3	68.0	28.9
0.75	30.8	41.1	68.2	28.9
1.0	32.1	**41.5**	**68.2**	**29.0**
		+1.2	*+1.0*	*+1.1*

(c) Panoptic FPN on COCO for **semantic** segmentation ($\lambda_s = 1$).

λ_i	AP	mIoU	iIoU	PQ^{St}
0.0	-	74.5	55.8	62.4
0.1	27.4	75.3	57.6	62.5
0.25	30.5	**75.5**	**58.3**	**62.5**
0.5	32.0	75.0	58.2	62.2
0.75	32.6	74.3	58.2	61.7
1.0	33.2	74.2	57.4	61.4
		+1.0	*+2.5*	*+0.1*

(d) Panoptic FPN on Cityscapes for **semantic** segmentation ($\lambda_s = 1$).

04 Panoptic-DeepLab



- CVPR2020
- UIUC（伊利诺伊大学厄巴纳-香槟分校）
- Google Research

- **Contribution**
  - a bottom-up 、 one-stage approach could deliver state-of-the-art results on panoptic segmentation

04 Panoptic-DeepLab

- **Difference between the top-down and the bottom-up**
  - the top-down methods(UPSNet, Panoptic FPN)
    - Two –stage
    - attaching another semantic segmentation branch to Mask R-CNN
  - the bottom-up methods(Panoptic-DeepLab)
    - One –stage
    - semantic segmentation + class-agnostic offset regression

04 Panoptic-DeepLab

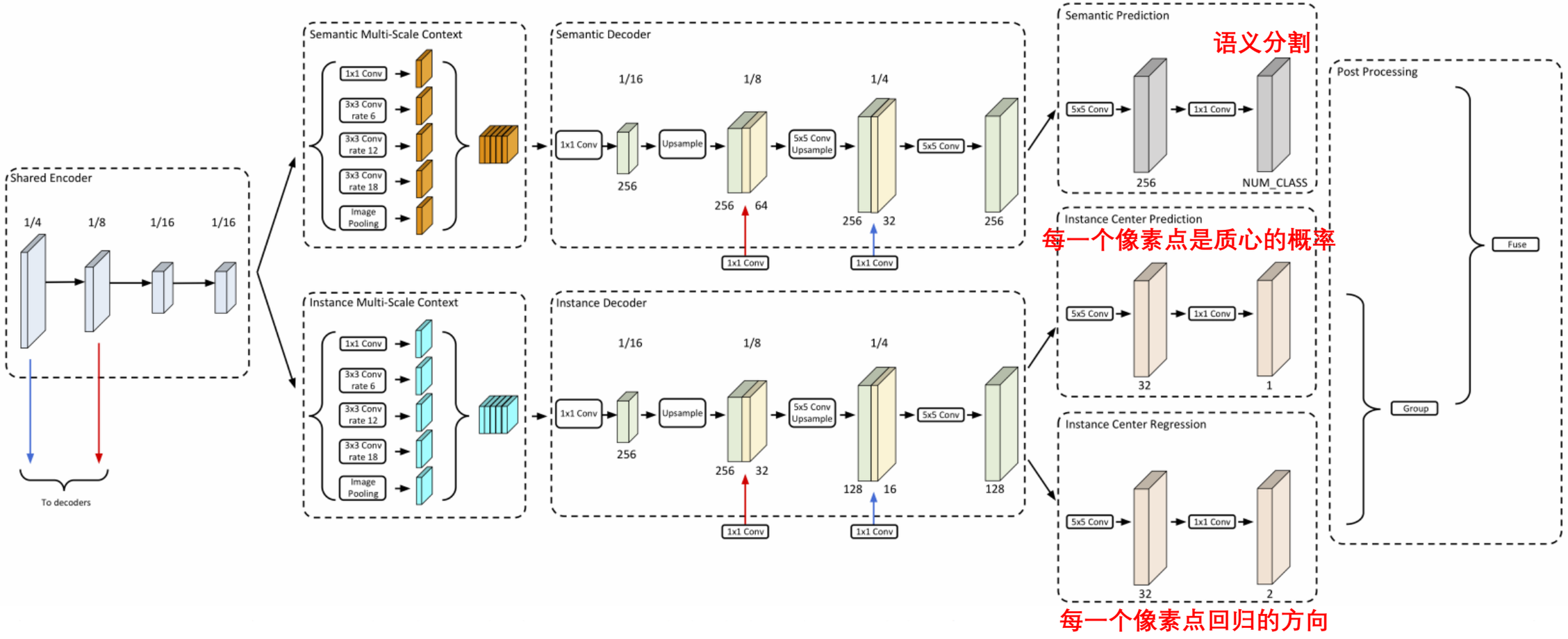
- **Network**

an encoder backbone

decoupled ASPP modules

decoupled decoder modules

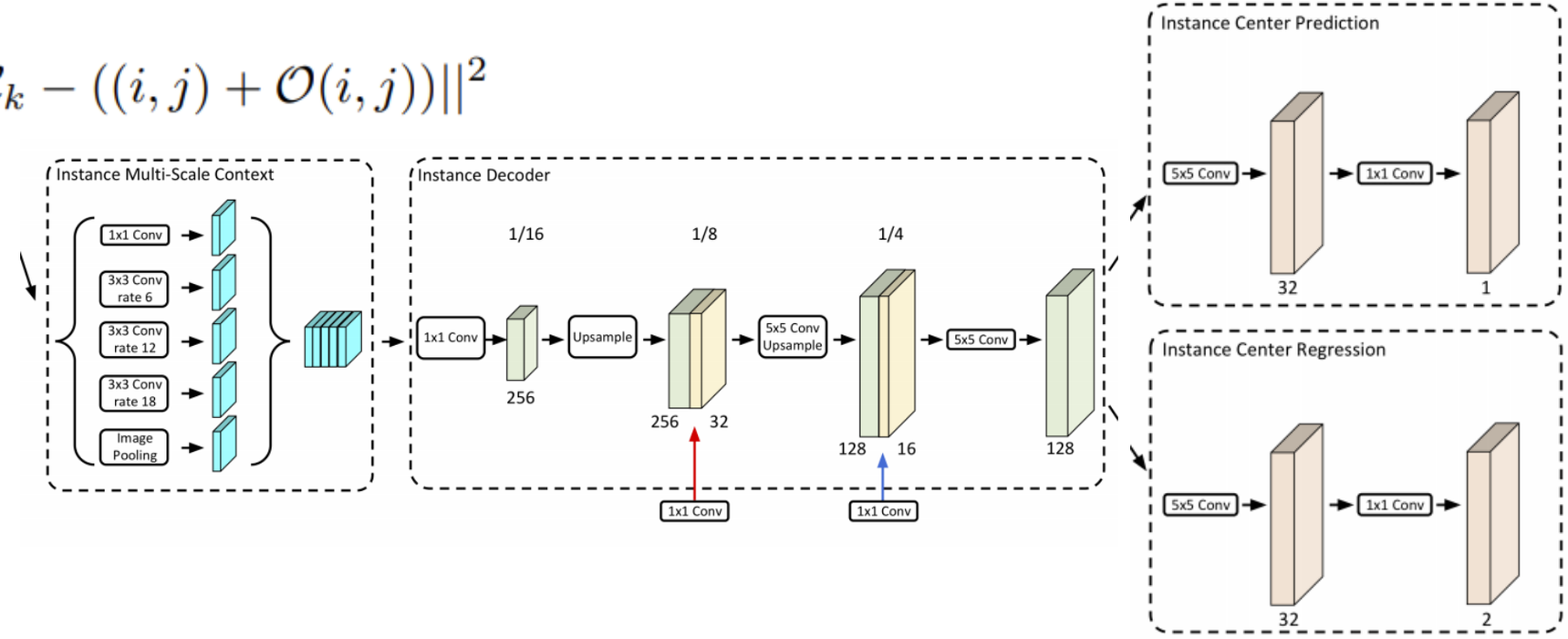
task-specific prediction heads



04 Panoptic-DeepLab

- **Simple instance representation**
  - max-pooling with kernel size 7　保留Pool前后未改变的坐标，作为实例的中心
  - top-k highest confidence scores are kept (k = 200)
- **Simple instance grouping**

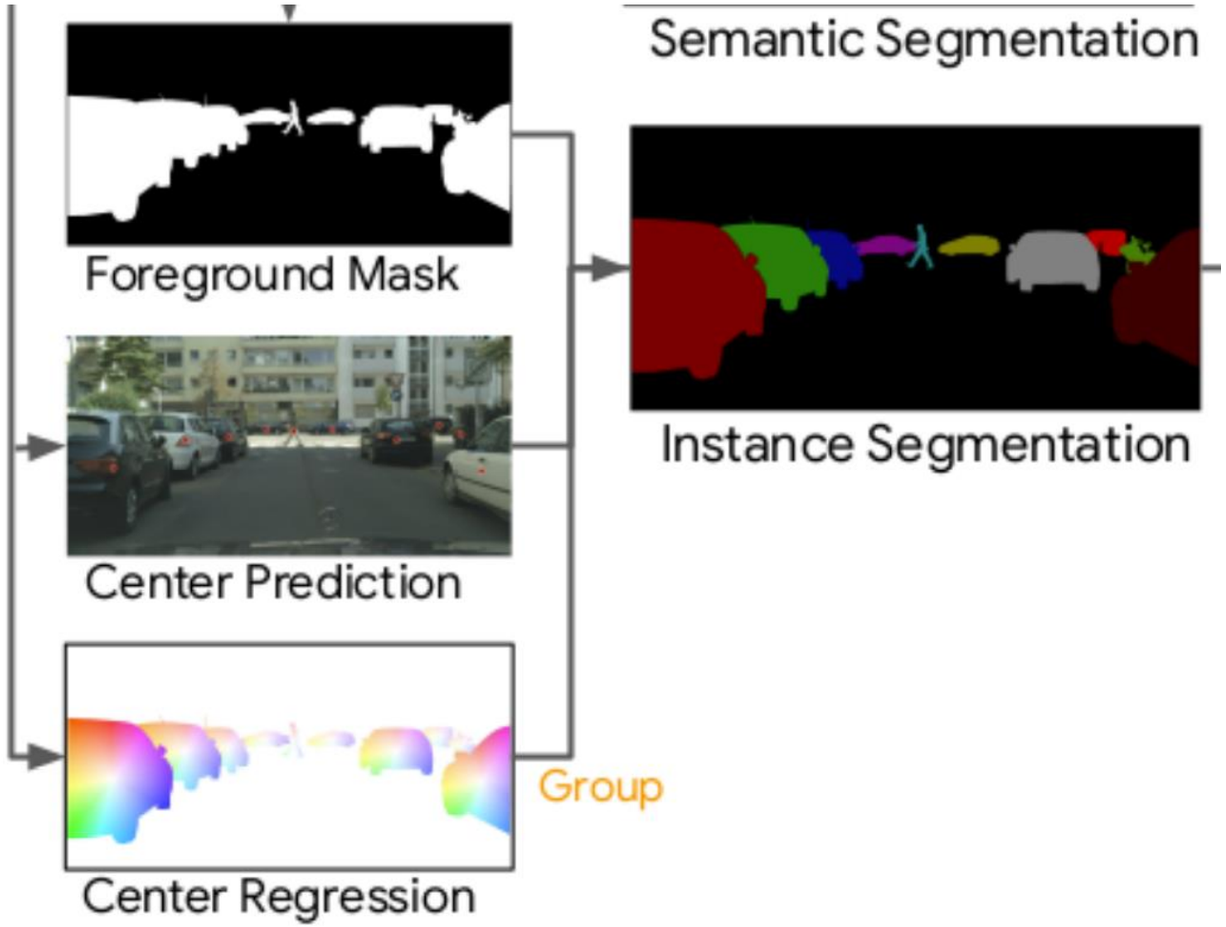
$$\hat{k}_{i,j} = \underset{k}{\operatorname{argmin}} ||\mathcal{C}_k - ((i,j) + \mathcal{O}(i,j))||^2$$



04 Panoptic-DeepLab

- **Post processing**

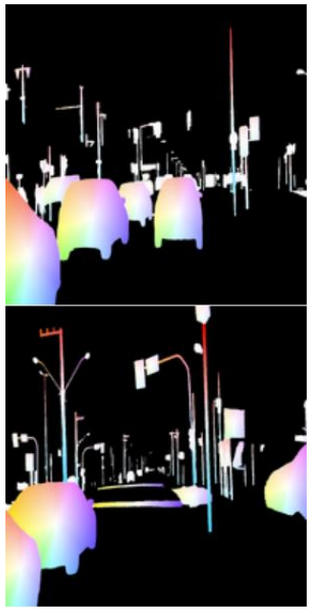
  - generate the final panoptic segmentation result by the ”majority vote” proposed（多数投票算法）



04 Panoptic-DeepLab

- **Loss function**

$$\mathcal{L} = \lambda_{sem}\mathcal{L}_{sem} + \lambda_{heatmap}\mathcal{L}_{heatmap} + \lambda_{offset}\mathcal{L}_{offset}$$



$\mathcal{L}_{sem}$

cross entropy loss

$\mathcal{L}_{heatmap}$

cross entropy loss

$\mathcal{L}_{offset}$

Mean Squared Error (MSE) :to minimize the distance between the predicted heatmap and mask