COCO Challenge 2018
Panoptic Segmentation Task

Team name: PKU_360

Team members: Yibo Yang, Xia Li, Hongyang Li, Tiancheng Shen, Zhouchen Lin, Jian Dong, Jiashi Feng, Shuicheng Yan

Peking University

Qihoo Corporation

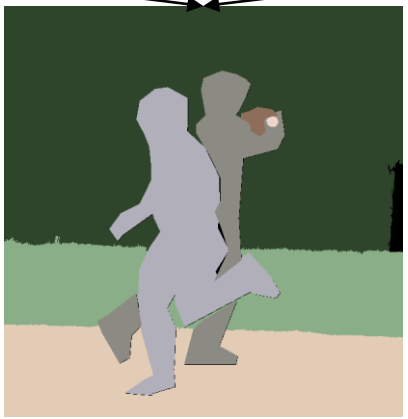
• Task Analysis

Image

Semantic segmentation

Instance segmentation

Panoptic segmentation



Task Analysis

- Occlusion between instances

- Occlusion between instance and semantic pixels

Task Analysis

ECCV 2018
European Conference on Computer Vision
8 – 14 September 2018 | Munich, Germany

- Occlusion between instances
  - Non overlapping detector, such as [1]
  - ⊙ Reasoning to solve occlusion, such as by post processing or learnable NMS.
- Occlusion between instance and semantic pixels

[1] Arnab et al. pixelwise instance segmentation with a dynamically instantiated network, CVPR 2017

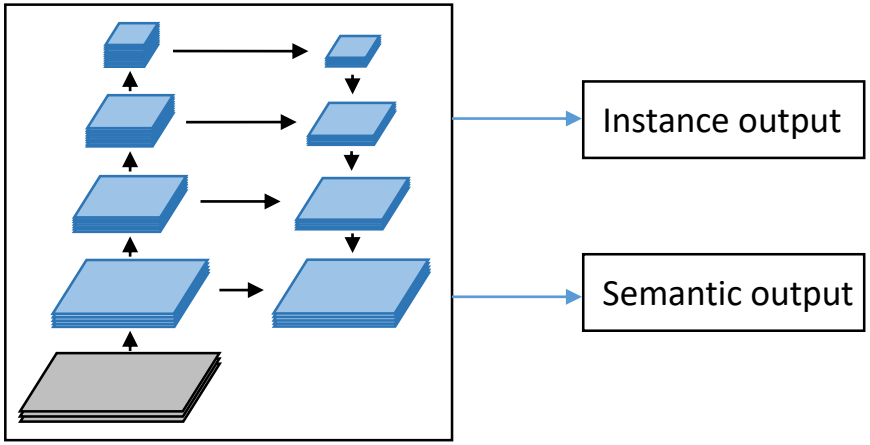
Task Analysis

- Occlusion between instances
  - Non overlapping detector, such as [1]
  - ⊙ Reasoning to solve occlusion, such as by post processing or learnable NMS.
- Occlusion between instance and semantic pixels
  - Comparison between semantic confidence and objectness score.
  - ⊙ Thing segments override stuff segments.

[1] Arnab et al. pixelwise instance segmentation with a dynamically instantiated network, CVPR 2017

Task Analysis

- Training methods
  - Multi-task in an e2e manner



FPN

Instance and semantic segmentation share the same Conv body to extract feature.

- Train instance and semantic segmentation separately

Instance Segmentation

ECCV 2018
European Conference on Computer Vision
8 – 14 September 2018 | Munich, Germany

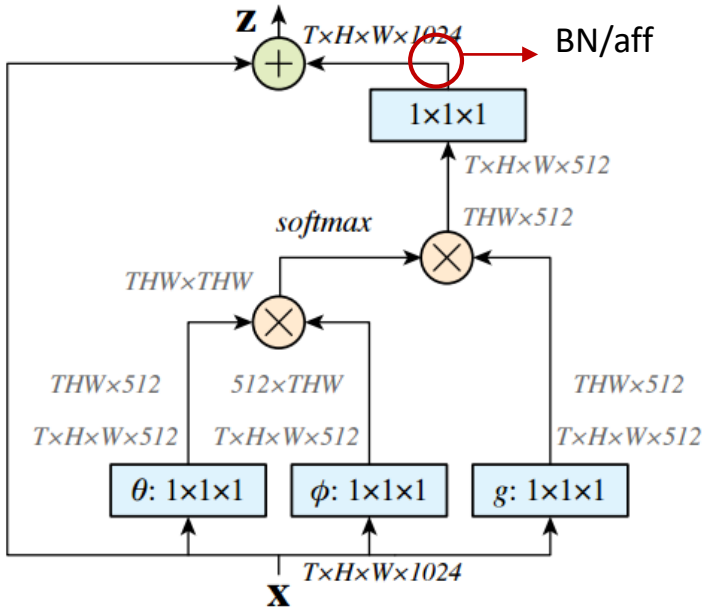
- Based on Mask RCNN
- Backbone
  - ResNeXt-152 trained on ImageNet 5k provided by Facebook.
- Best single model performance
  - 43.5 mask mAP on test-dev (used for our panoptic results)
- Methods
  - Non-local module[1]
  - Squeeze and excitation module[2]
  - Bottom-up path aggregation[3] in an alternate updating manner[4]
  - Synchronized BN, multi-scale training/testing, etc.
- Training details
  - 300k iterations
  - Single image on each GPU
  - Initial lr: 0.01

[1] Wang, et al. Non-local neural networks, CVPR 2018
[2] Hu, et al. Squeeze and excitation networks, CVPR 2018
[3] Liu, et al. Path aggregation network for instance segmentation, CVPR 2018
[4] Yang, et al. Convolutional neural networks with alternately updated clique, CVPR 2018

Instance Segmentation

ECCV 2018
European Conference on Computer Vision
8 – 14 September 2018 | Munich, Germany

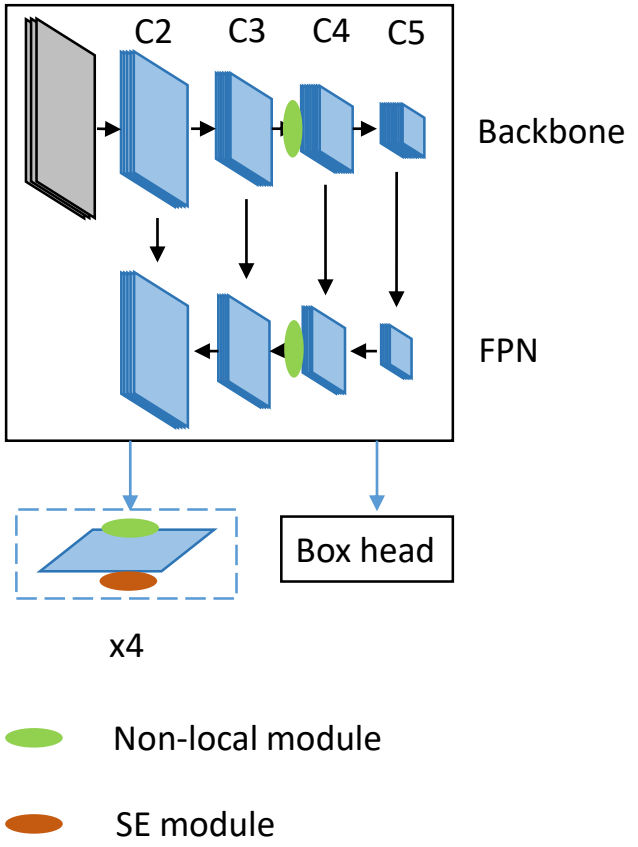
- Non-local module
  - On backbone (Res4)
  - On FPN (the same level with Res4)
  - On mask head (before each conv of the 4-convs head)
  - Synchronized BN or affine operation with scale parameter initialized as 0



Instance Segmentation

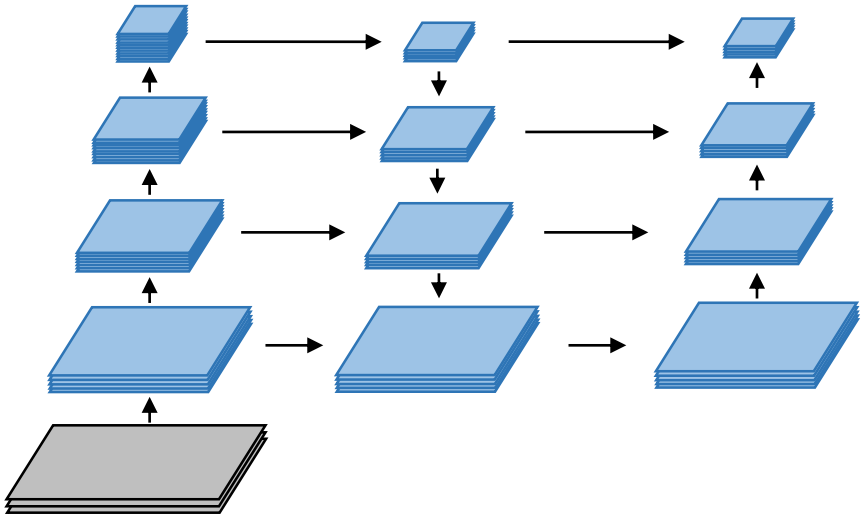
ECCV 2018
European Conference on Computer Vision
8 – 14 September 2018 | Munich, Germany

- Non-local module
  - On backbone (Res4)
  - On FPN (the same level with Res4)
  - On mask head (before each conv of the 4-convs head)
  - Synchronized BN or affine operation with scale parameter initialized as 0
- Squeeze-and-excitation module
  - On mask head (after each conv of the 4-convs head)



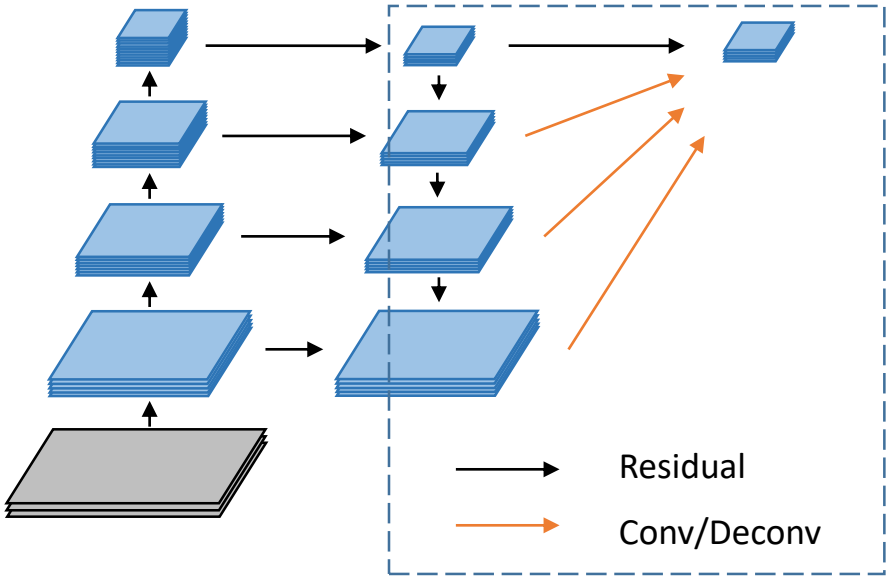
Instance Segmentation

- Bottom-up path aggregation
  - Original



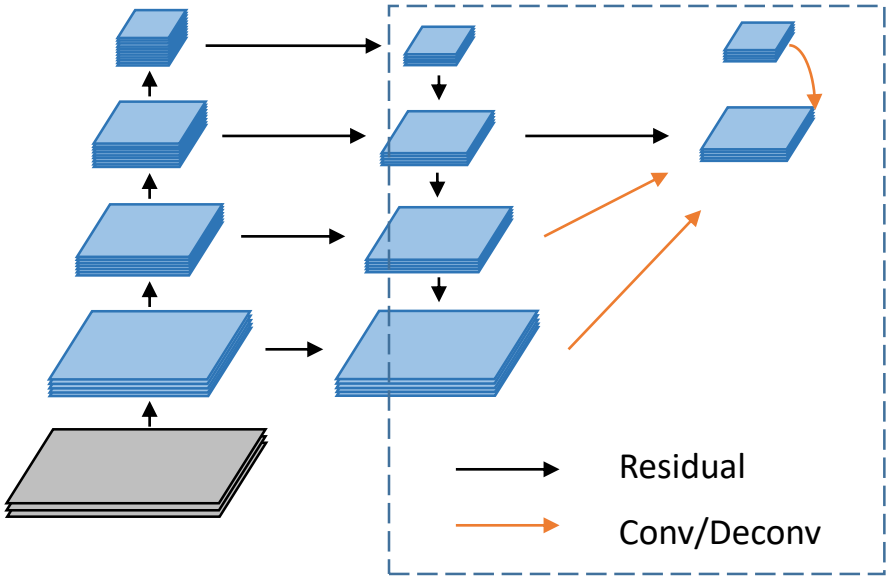
Instance Segmentation

- Bottom-up path aggregation
  - Ours



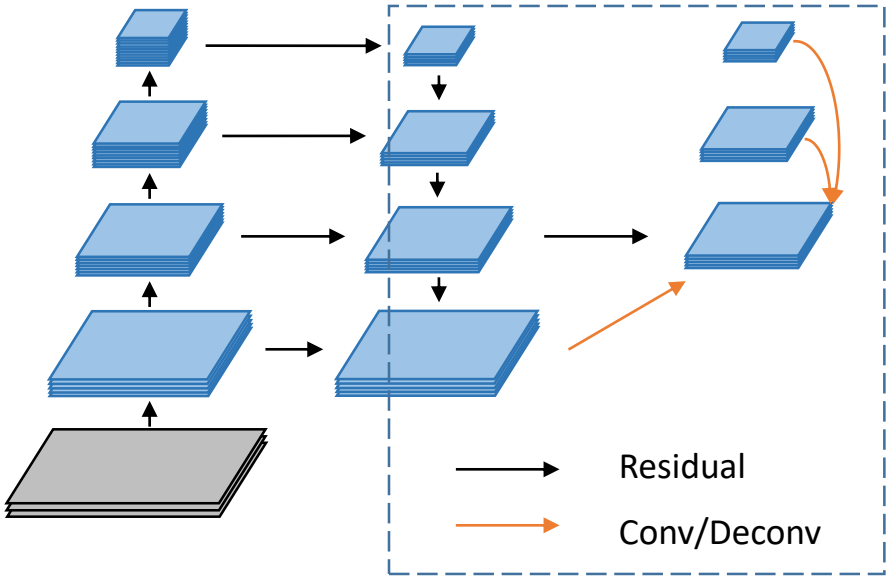
Instance Segmentation

- Bottom-up path aggregation
  - Ours



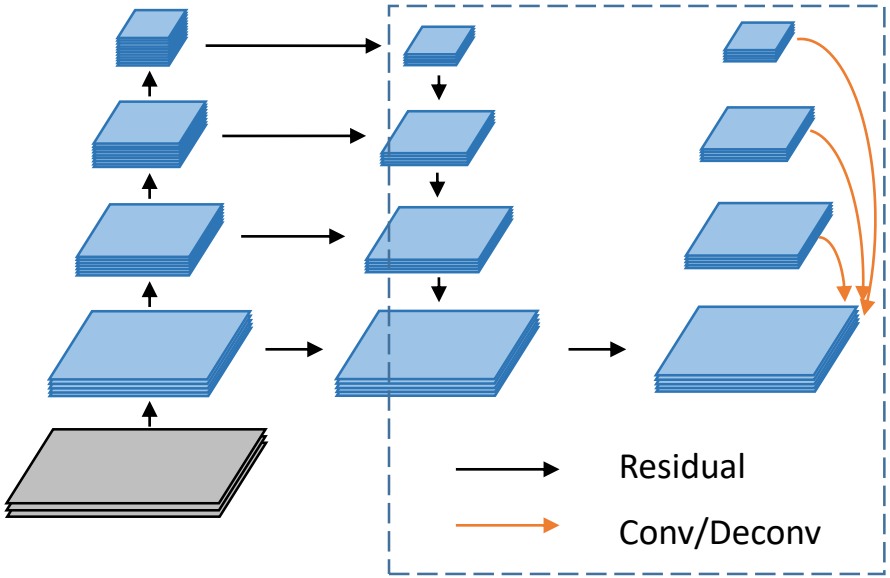
Instance Segmentation

- Bottom-up path aggregation
  - Ours



Instance Segmentation

- Bottom-up path aggregation
  - Ours



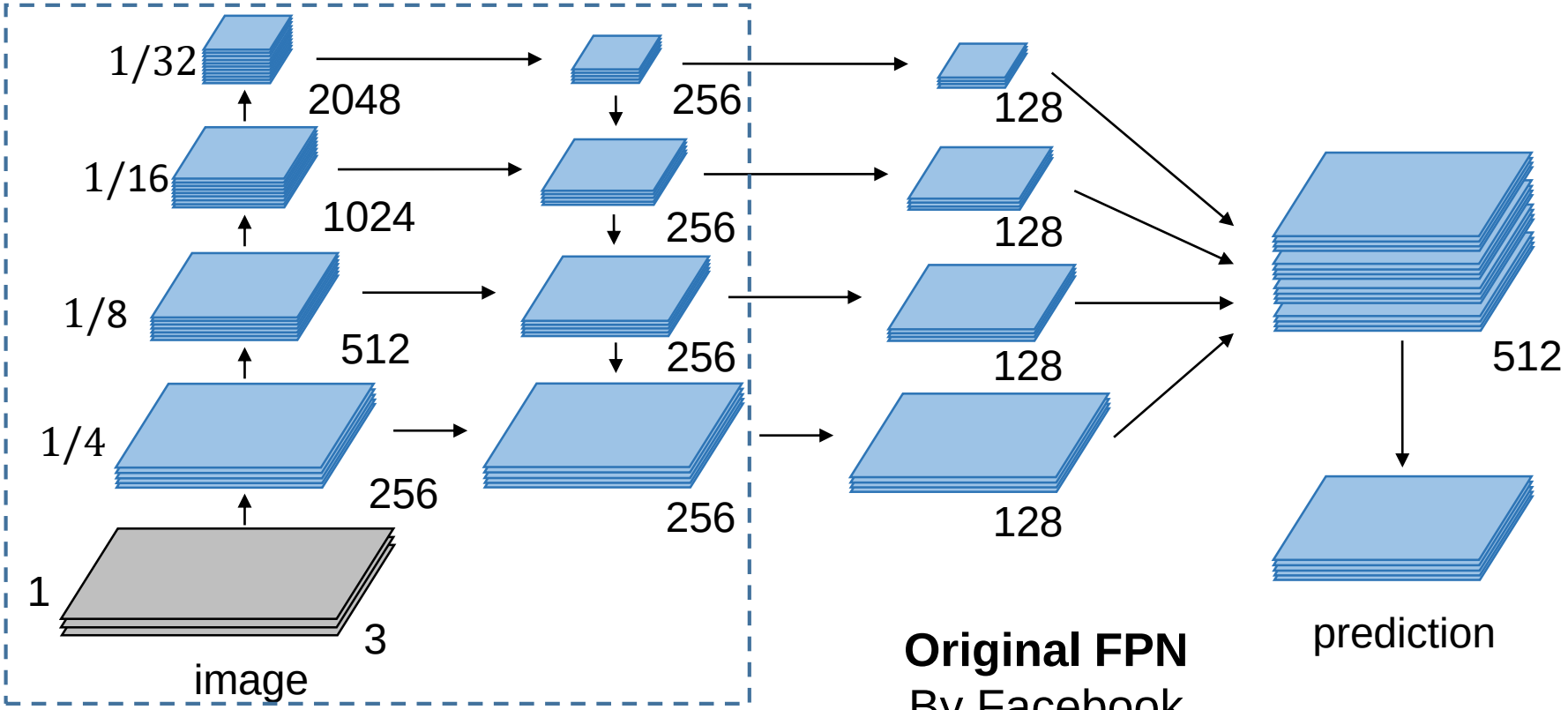
Instance Segmentation

ECCV 2018
European Conference on Computer Vision
8 – 14 September 2018 | Munich, Germany

- Ablation experiments (40000 iterations, no test time augmentation, on val set)

	Box map	Mask map
R-50 baseline	33.66	30.76
+ 4SE mask head	33.83	30.96
+ nonlocal backbone + 4SE mask head	33.83	31.09
+ nonlocal backbone + 4SE mask head + 4nonlocal mask head	33.99	31.15
+ nonlocal backbone + nonlocal FPN	34.02	31.08
+ nonlocal backbone + nonlocal FPN + path aggregation (original)	34.11	31.28
+ nonlocal backbone + nonlocal FPN + path aggregation (ours)	34.60	31.75

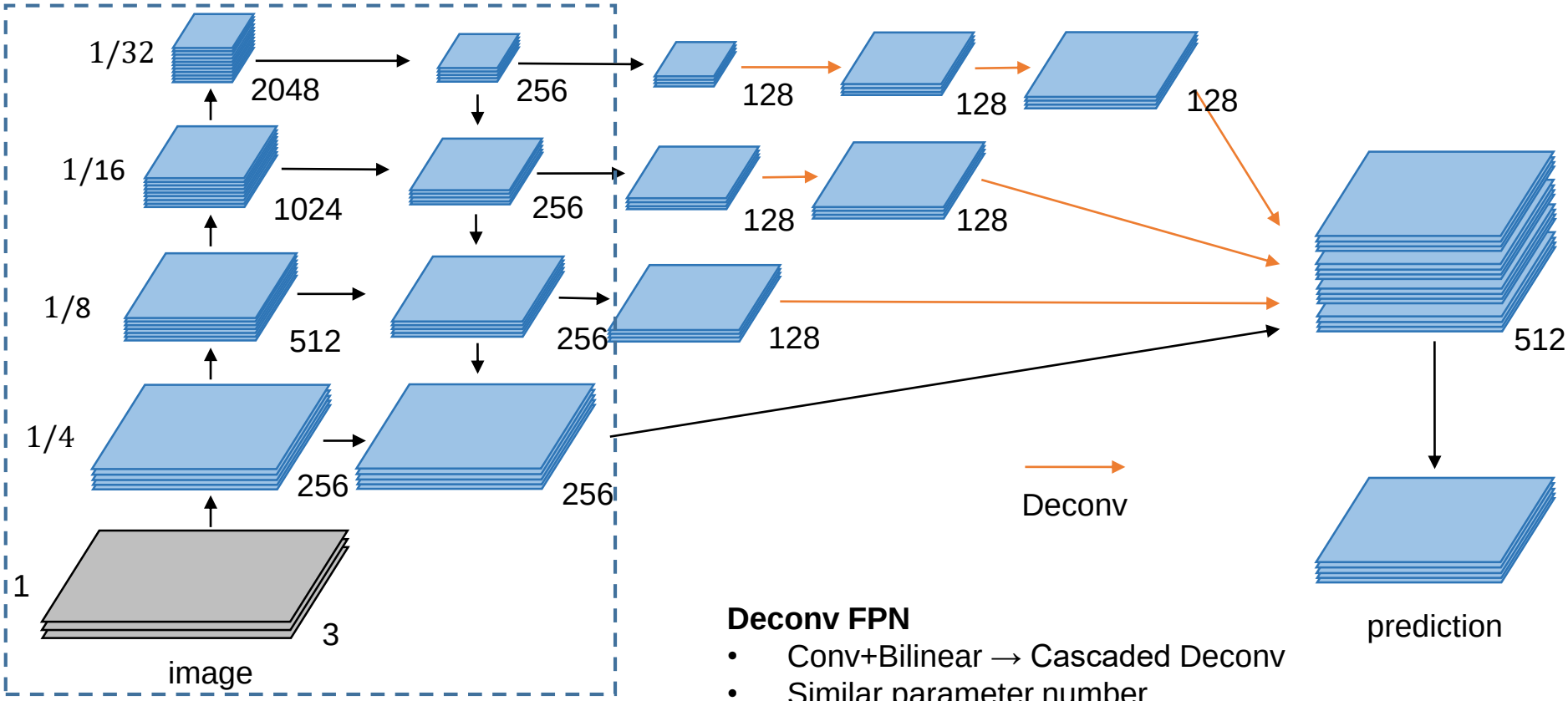
• Semantic Segmentation



Original FPN
By Facebook

• Semantic Segmentation

ECCV 2018
European Conference on Computer Vision
8 – 14 September 2018 | Munich, Germany

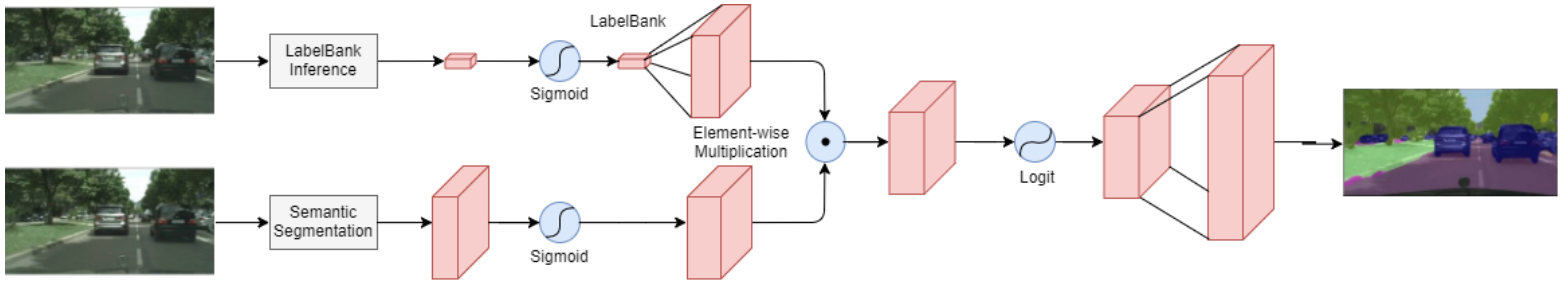


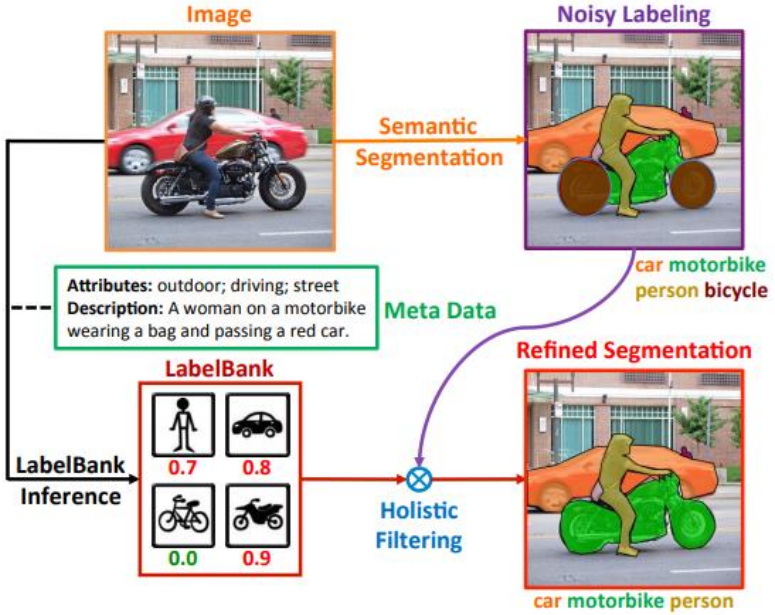
Deconv FPN

- Conv+Bilinear → Cascaded Deconv
- Similar parameter number

Semantic Segmentation

ECCV 2018
European Conference on Computer Vision
8 – 14 September 2018 | Munich, Germany





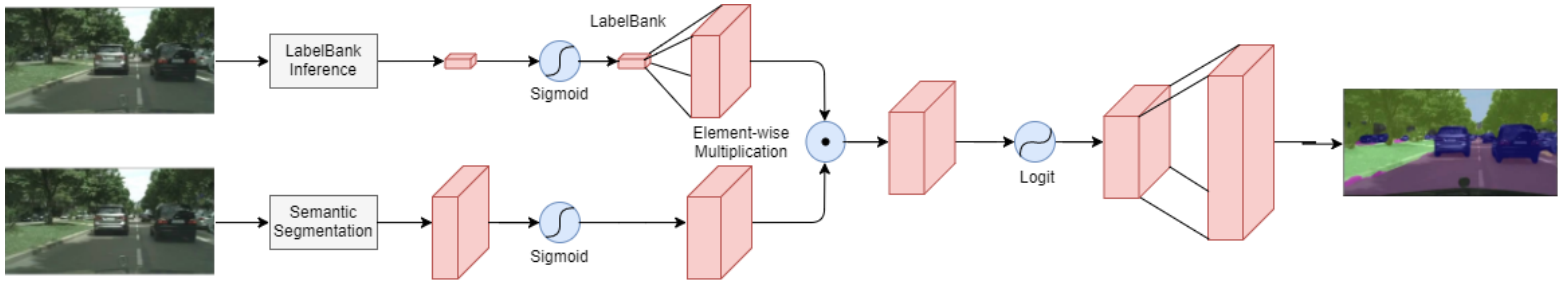
Original Labelbank[1]

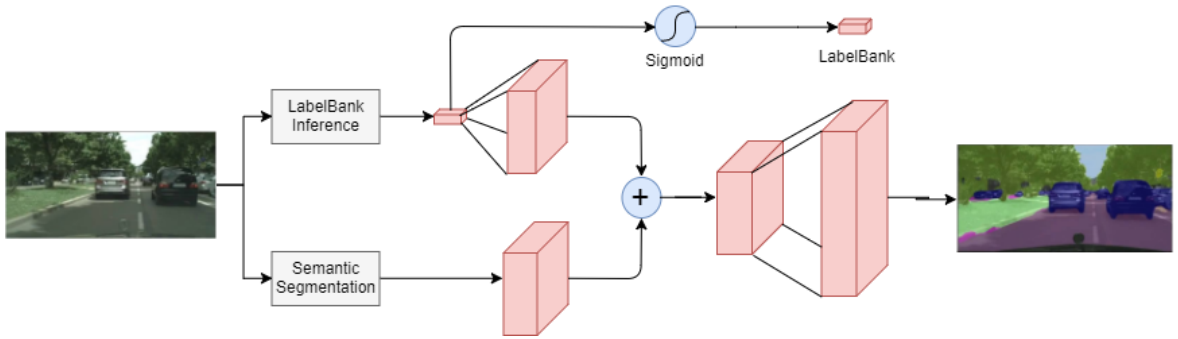
- Auxiliary branch to determine whether a label occurs in an image
- Multiply with seg map to remove non-existing labels in prediction

[1] Hu H, Deng Z, Zhou G T, et al. LabelBank: Revisiting Global Perspectives for Semantic Segmentation, arXiv preprint arXiv:1703.09891, 2017.c

Semantic Segmentation

ECCV 2018
European Conference on Computer Vision
8 – 14 September 2018 | Munich, Germany





Modified Labelbank (LB)

- Share backbone of two branches
- Simplify the 'Merge' operation

Semantic Segmentation

ECCV 2018
European Conference on Computer Vision
8 – 14 September 2018 | Munich, Germany

Comparative Experiment

- Backbone: SE-ResNet50
- Init Learning Rate: 1e-2
- Iteration: 20k
- Optimizer: Adam
- Input size: 512
- Dataset: COCO-stuff 10k

	mIoU	fIoU	mAcc	pAcc
Original FPN	31.19	48.34	42.71	62.54
Deconv FPN	31.52	49.23	42.74	63.64
FPN + LB	**33.12**	50.1	**45.01**	**64.74**
Deeplab[1]	32.37	**50.73**	43.34	65.2
PSPNet[2]	32.58	50.41	43.49	64.93
FPANet[3]	32.14	49.23	43.91	63.69

[1] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation, arXiv preprint arXiv:1706.05587, 2017.
[2] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network, CVPR 2017: 2881-2890.
[3] Li H, Xiong P, An J, et al. Pyramid Attention Network for Semantic Segmentation, arXiv preprint arXiv:1805.10180, 2018.

· Semantic Segmentation

Final Submit

- Backbone: ResNeXt152
- Init Learning Rate:
  - Backbone: 1e-3
  - Seg Head: 1e-2
- Normalization:
  - Backbone: freeze
  - Seg Head: no BN
- Iteration: 60k
- Optimizer: Adam
- Dataset: COCO- Panoptic (Stuff Parts)

Average the two models for panoptic calculation

	Original FPN	Deconv FPN
Input size	800	732
mIoU	49.54	49.39
fIoU	67.53	67.2
mAcc	62.10	62.38
pAcc	79.51	79.29

Panoptic Segmentation

- Baseline method (provided by panoptic cocoapi)
  - Filter out instances (objectness score below a threshold)
  - NMS-like procedure (remove pixels which have been assigned to a segment with higher score, accept the non-overlapping portion if sufficient fraction remains)
  - Filter our semantic segments (area below a threshold)
  - Thing override stuff

Panoptic Segmentation

- Baseline method (provided by panoptic cocoapi)
  - Filter out instances (objectness score below a threshold)
  - NMS-like procedure (remove pixels which have been assigned to a segment with higher score, accept the non-overlapping portion if sufficient fraction remains)
  - Filter our semantic segments (area below a threshold)
  - Thing override stuff

Problem: does not solve occlusion, take object relationships into account

e.g.

Tie -> Person

Spoon -> Bowl -> Dinning table

Panoptic Segmentation

- Our method

  - Filter out instances (objectness score below a threshold);
  - Select the labels that are more likely to be overlapped with other labels according to the frequency;
  - For the selected labels, apply the NMS-like procedure within each label (the procedure is valid only when two segments are of the same label);
  - For the other labels, apply the NMS-like procedure among them;
  - Assign the overlapped pixels according to label prior to solve occlusion;

Panoptic Segmentation

- Our method

  - Filter out instances (objectness score below a threshold);
  - Select the labels that are more likely to be overlapped with other labels according to the frequency;
  - For the selected labels, apply the NMS-like procedure within each label (the procedure is valid only when two segments are of the same label);
  - For the other labels, apply the NMS-like procedure among them;
  - Assign the overlapped pixels according to label prior to solve occlusion;
  - Filter out semantic segments (area below a threshold)
  - Filter out semantic pixels (confidence below a threshold)
  - Assign a semantic pixel to the second highest prediction label when its probability is above a threshold and the highest prediction is void.
  - Thing override stuff

Panoptic Segmentation

- Ablation experiments (on val set)

Method	PQ	SQ	RQ	PQ-t	SQ-t	RQ-t	PQ-s	SQ-s	RQ-s
Baseline	<45.6	-	-	-	-	-	-	-	-
Method 1	45.6	79.9	55.4	57.2	83.5	67.9	28.2	74.4	36.5
Method 2	46.02	79.9	55.9	57.8	83.5	68.7	28.2	74.4	36.5
Method 3	46.06	79.9	55.9	57.9	83.5	68.8	28.2	74.4	36.5

Method 1: Do not apply the procedure on our selected out labels, and apply on the other labels.
Method 2: Apply the procedure within each label for all labels.
Method 3: Apply the procedure within each label for our selected labels, and apply the procedure among the other labels.

Panoptic Segmentation

- Ablation experiments (on test-dev set)

Method	PQ	SQ	RQ
-	44.2	79.5	53.5
+ semantic area threshold	45.6	79.8	55.2
+ semantic area threshold + Method 3	46.3	79.7	56.1

Submitted entry

Panoptic Segmentation

- Some examples (from val set)

Image

Panoptic output

Ground truth

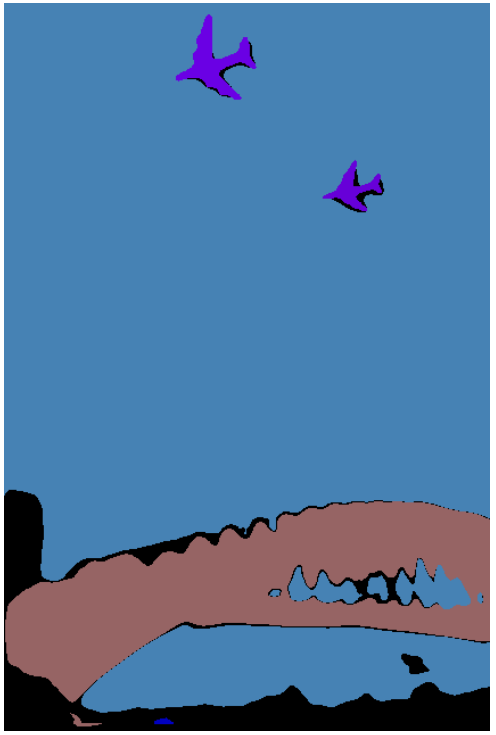
Panoptic Segmentation

ECCV 2018
European Conference on Computer Vision
8 – 14 September 2018 | Munich, Germany

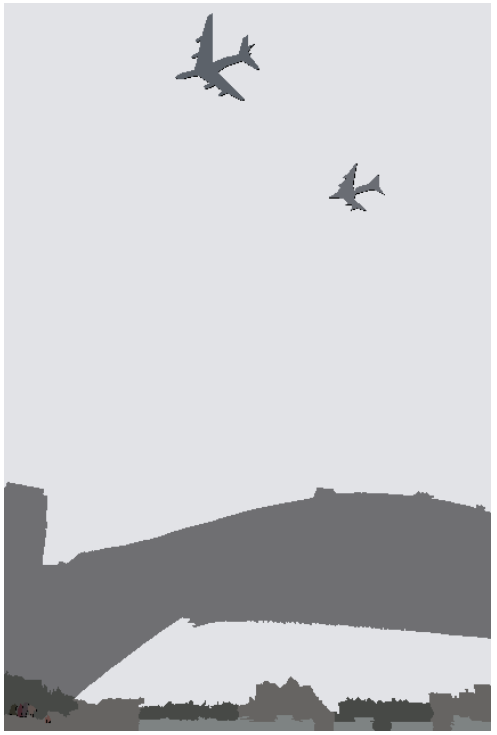
- Some examples (from val set)

Image

Panoptic output



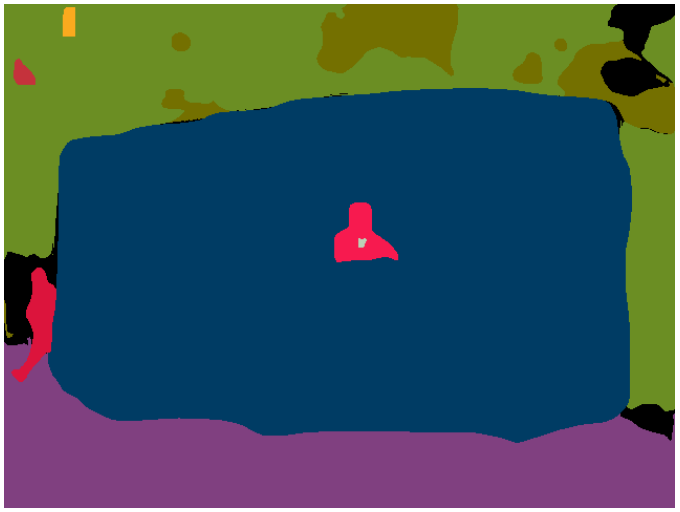
Ground truth



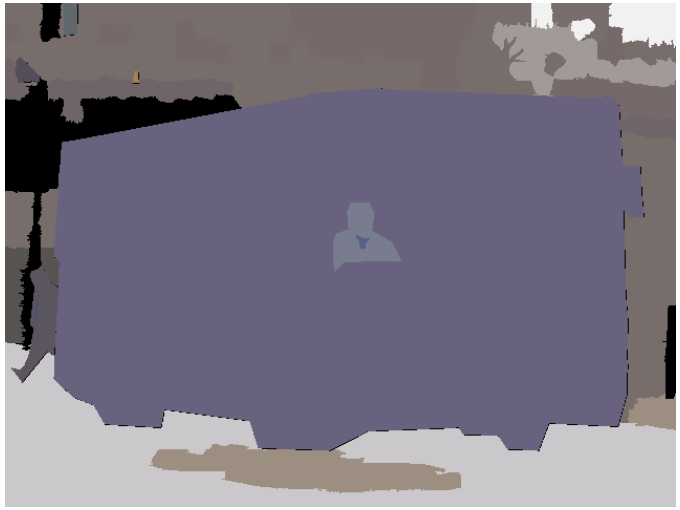
• Panoptic Segmentation

ECCV 2018
European Conference on Computer Vision
8 – 14 September 2018 | Munich, Germany

Image



Panoptic output



Ground truth

Panoptic Segmentation

- Future direction

  - Reasoning object relationships in an e2e manner to resolve the overlap between instances.
  - Semantic and instance segmentation output can be unified into a single framework to resolve the overlap between thing and stuff.

Thank you!

For any question, please contact: ibo@pku.edu.cn