A Unified Architecture for Instance and Semantic Segmentation


Alexander Kirillov


Kaiming He


Ross Girshick


Piotr Dollár

UNIVERSITÄT HEIDELBERG

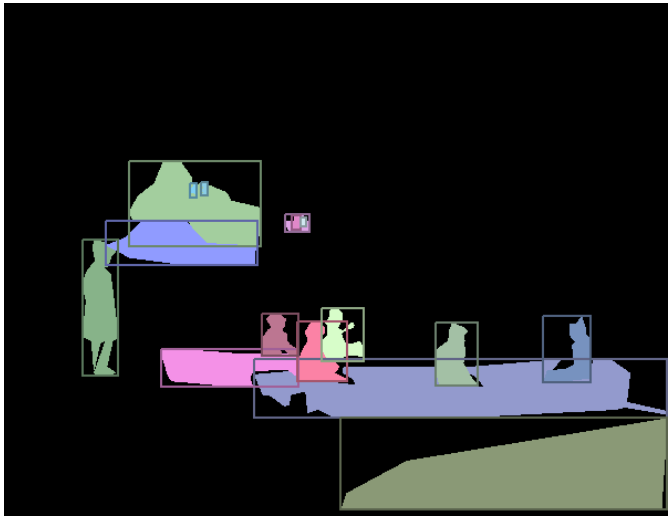
FACEBOOK AI RESEARCH

Object Detection vs Semantic Segmentation

Object Detection

Semantic Segmentation

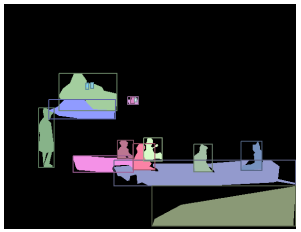
Object Detection vs Semantic Segmentation



Object Detection/Seg

Semantic Segmentation

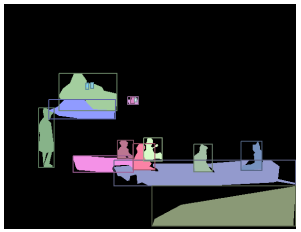
Deep Networks in Object Recognition



Object Detection/Seg

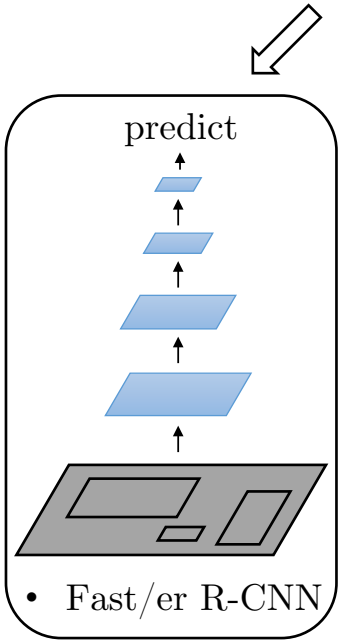
Semantic Segmentation

Deep Networks in Object Recognition



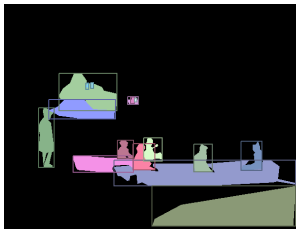
Object Detection/Seg

Semantic Segmentation



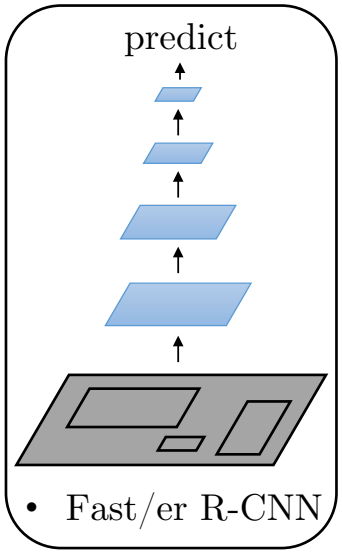
classification net

Deep Networks in Object Recognition

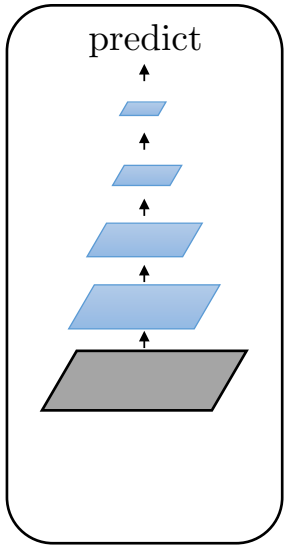


Object Detection/Seg

Semantic Segmentation

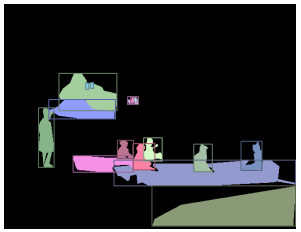


classification net



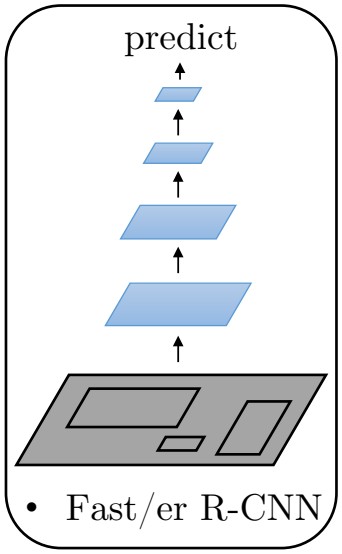
dilated net

Deep Networks in Object Recognition

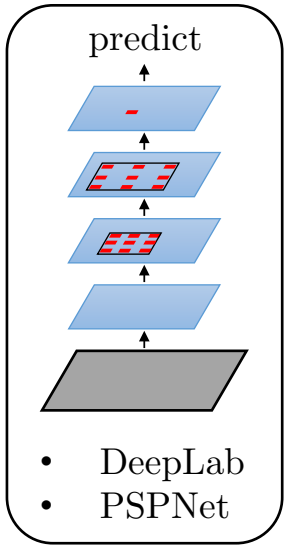


Object Detection/Seg

Semantic Segmentation

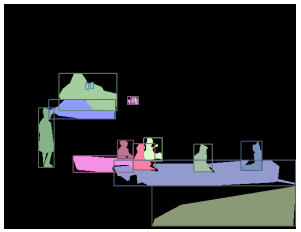


classification net



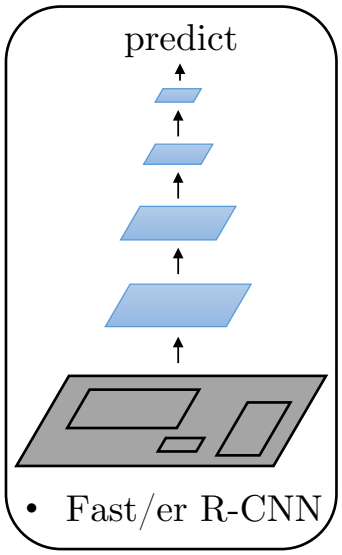
dilated net

Deep Networks in Object Recognition

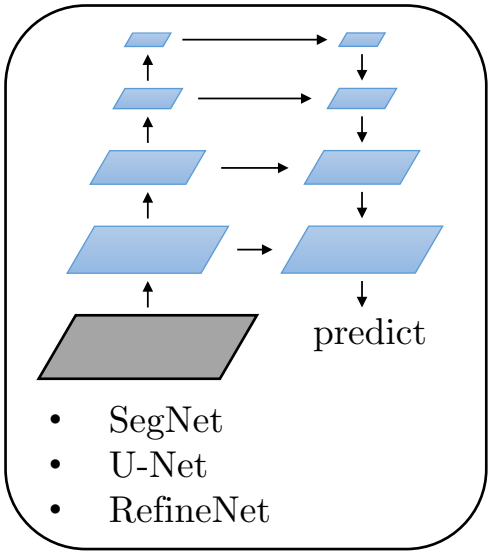


Object Detection/Seg

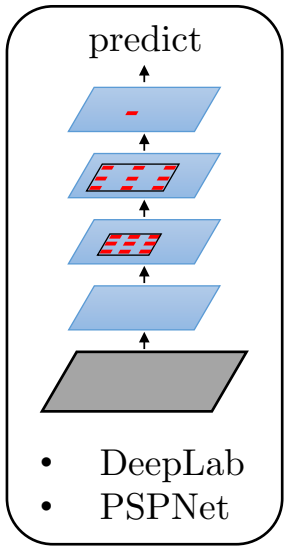
Semantic Segmentation



classification net

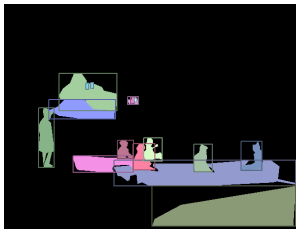


decoder-encoder net



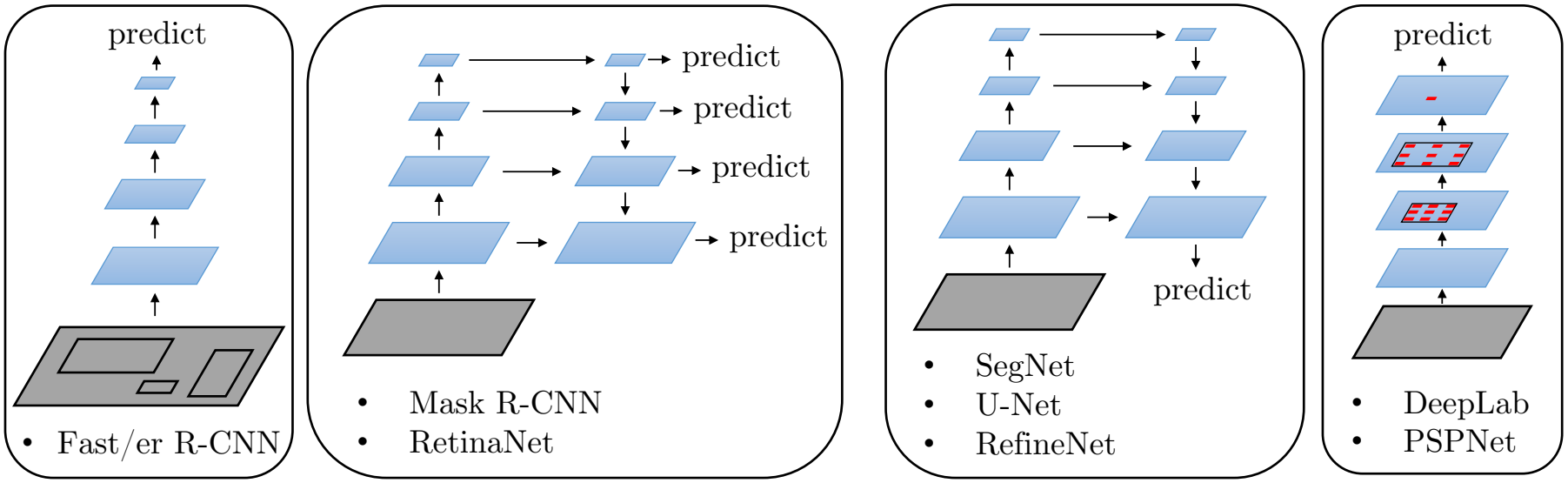
dilated net

Deep Networks in Object Recognition



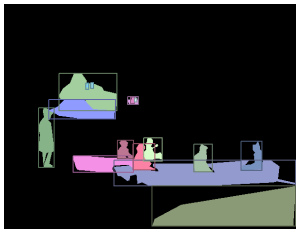
Object Detection/Seg

Semantic Segmentation



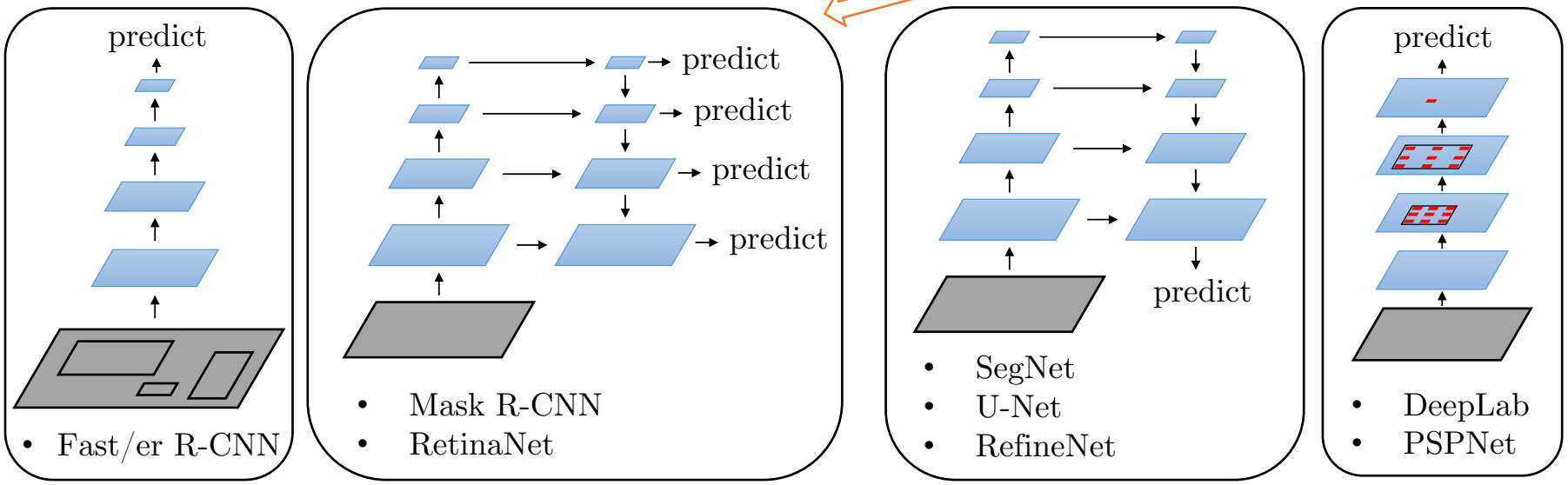
classification net | FPN net | decoder-encoder net | dilated net

Deep Networks in Object Recognition



Object Detection/Seg

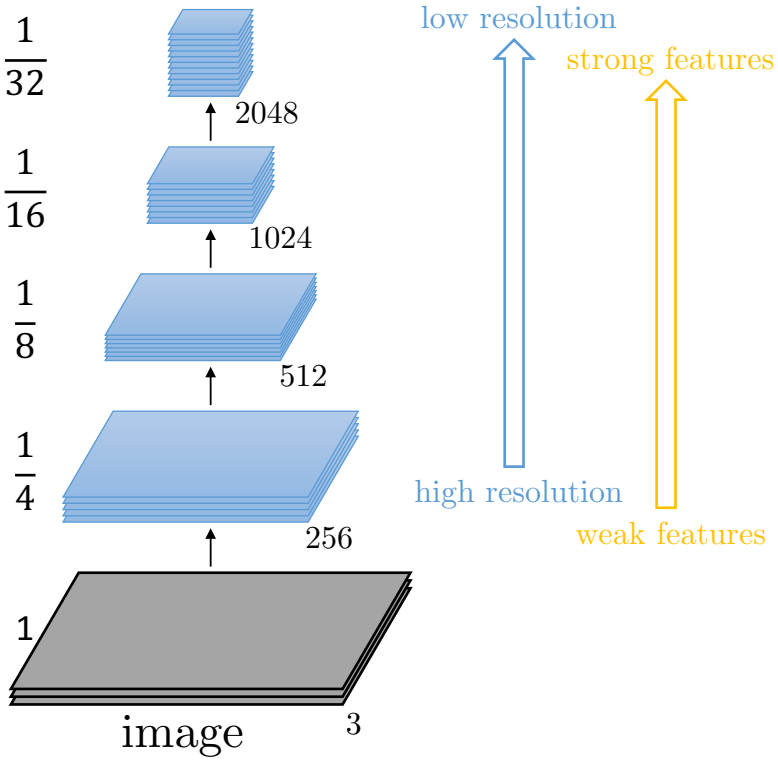
Semantic Segmentation



classification net | FPN net | decoder-encoder net | dilated net

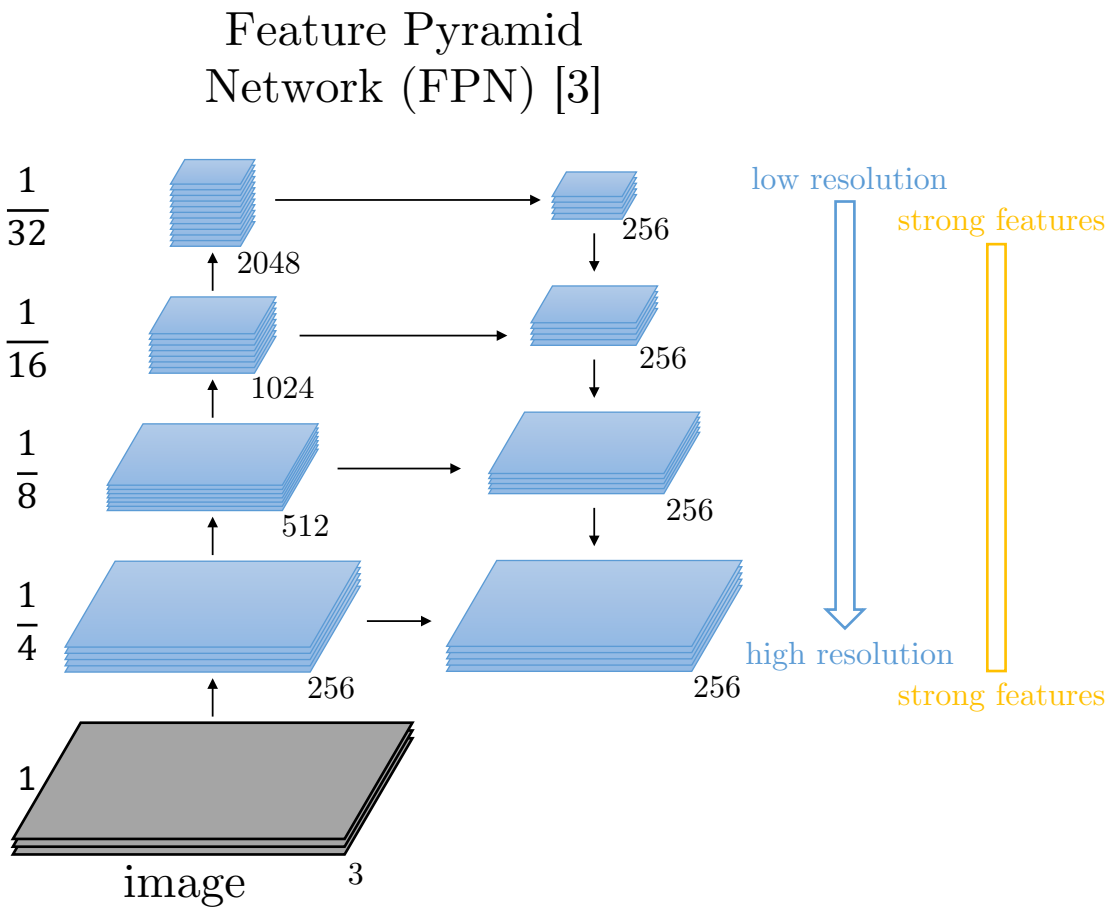
FPN Architecture

ResNet152 [1] /
ResNeXt152 [2]



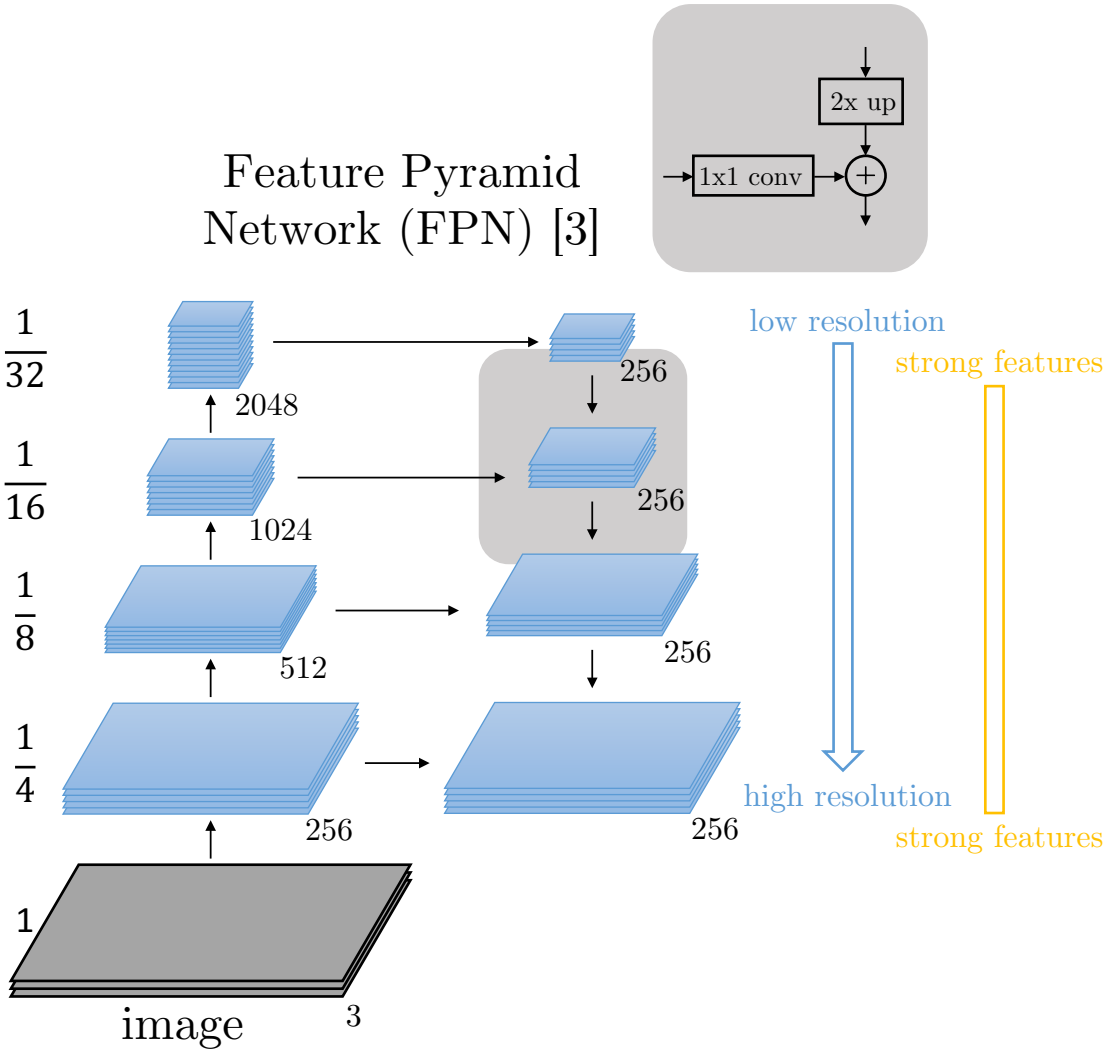
[1] He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. CVPR 2016.
[2] Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. Aggregated residual transformations for deep neural networks. CVPR 2017.

FPN Architecture



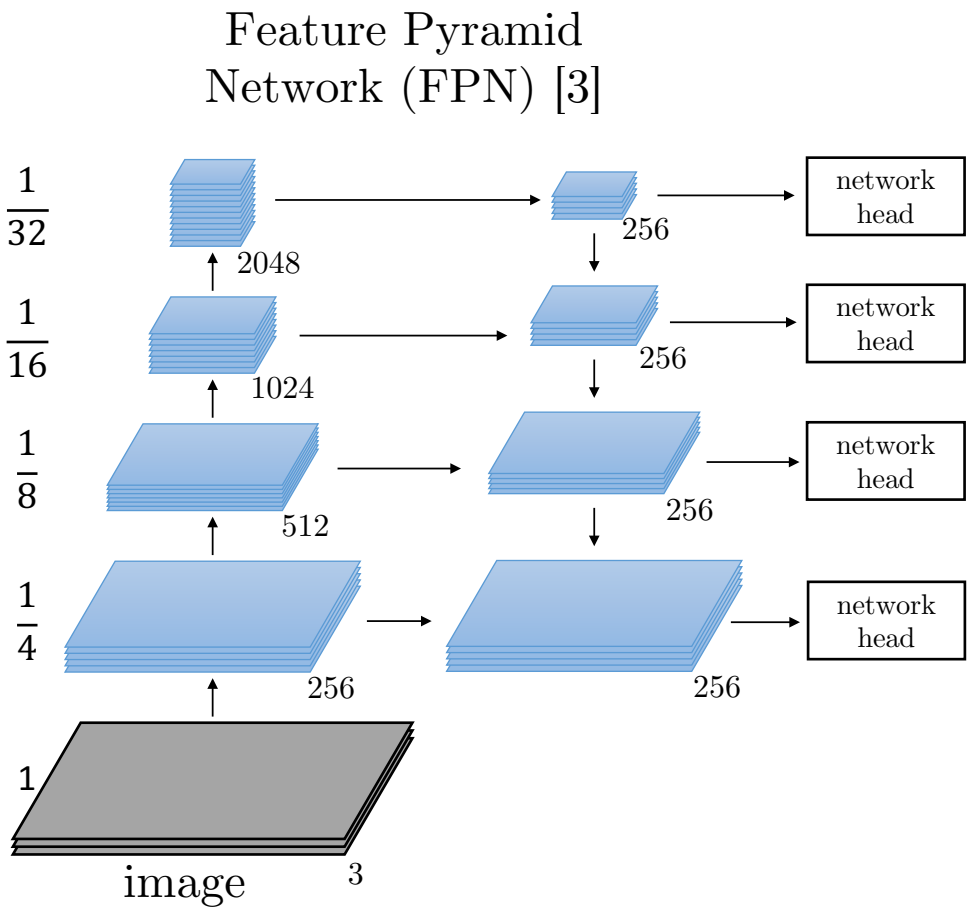
[1] He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. CVPR 2016.
[2] Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. Aggregated residual transformations for deep neural networks. CVPR 2017.
[3] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. Feature pyramid networks for object detection. CVPR 2017.

FPN Architecture



[1] He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. CVPR 2016.
[2] Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. Aggregated residual transformations for deep neural networks. CVPR 2017.
[3] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. Feature pyramid networks for object detection. CVPR 2017.

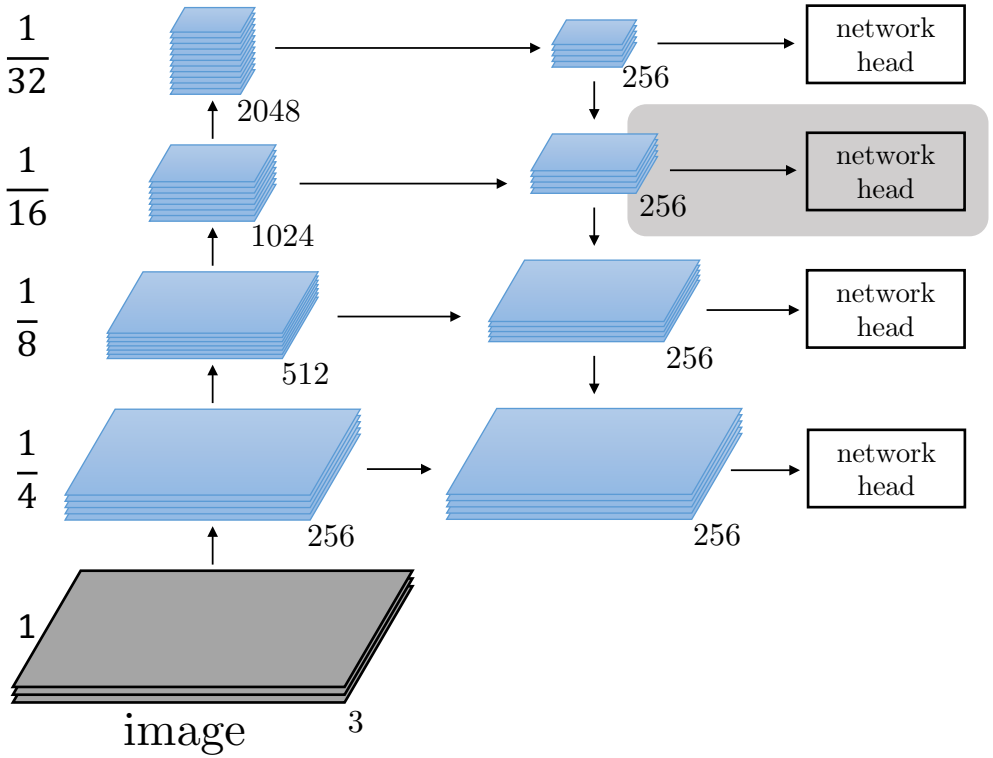
FPN Architecture



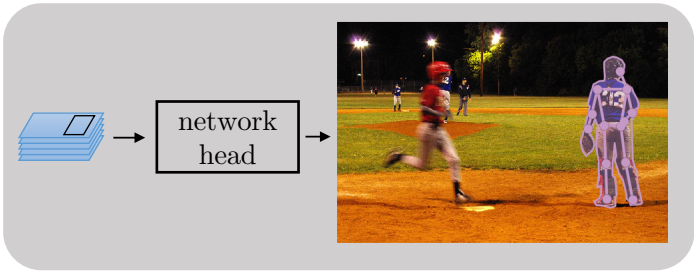
[1] He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. CVPR 2016.
[2] Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. Aggregated residual transformations for deep neural networks. CVPR 2017.
[3] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. Feature pyramid networks for object detection. CVPR 2017.

FPN Architecture

Feature Pyramid Network (FPN) [3]



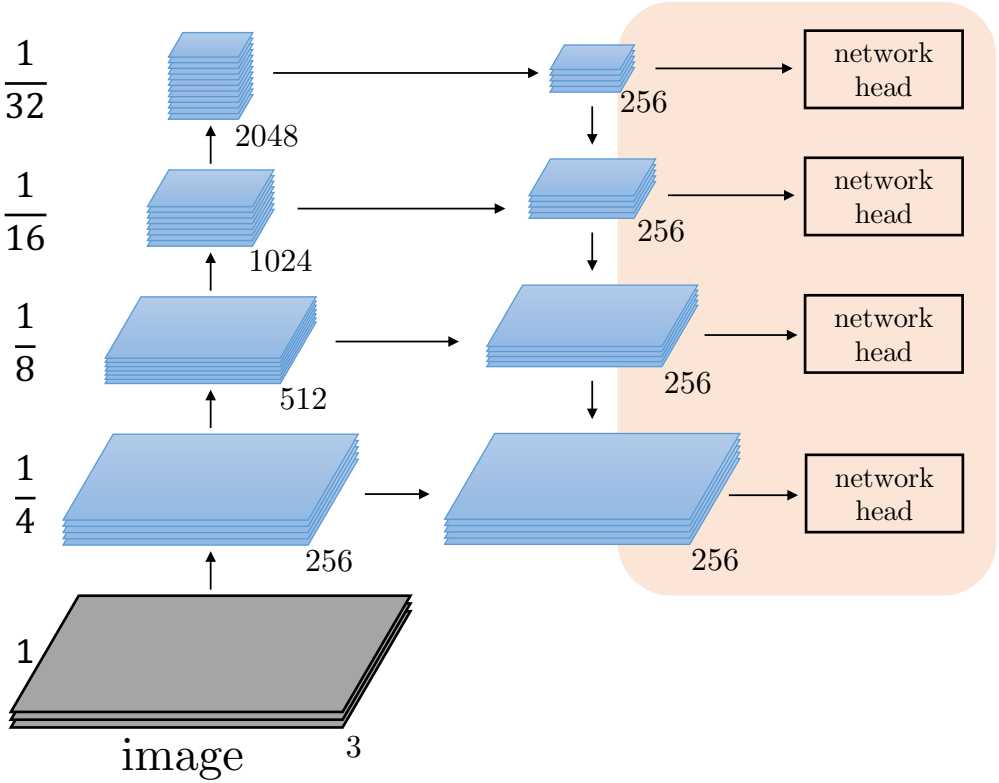
Mask R-CNN[4]



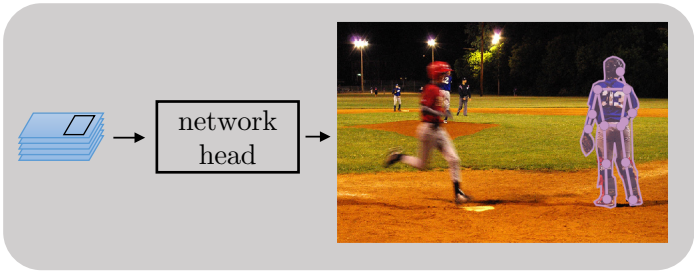
[1] He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. CVPR 2016.
[2] Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. Aggregated residual transformations for deep neural networks. CVPR 2017.
[3] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. Feature pyramid networks for object detection. CVPR 2017.
[4] He, K., Gkioxari, G., Dollár, P., & Girshick, R. Mask R-CNN. ICCV 2017.

FPN Architecture

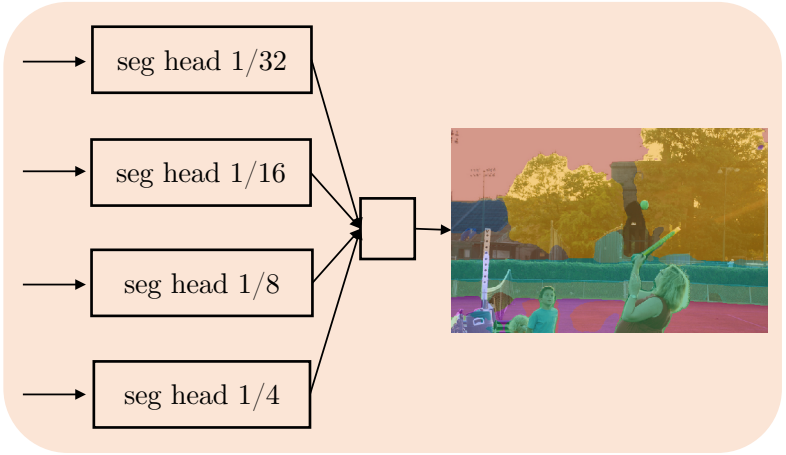
Feature Pyramid Network (FPN) [3]



Mask R-CNN[4]

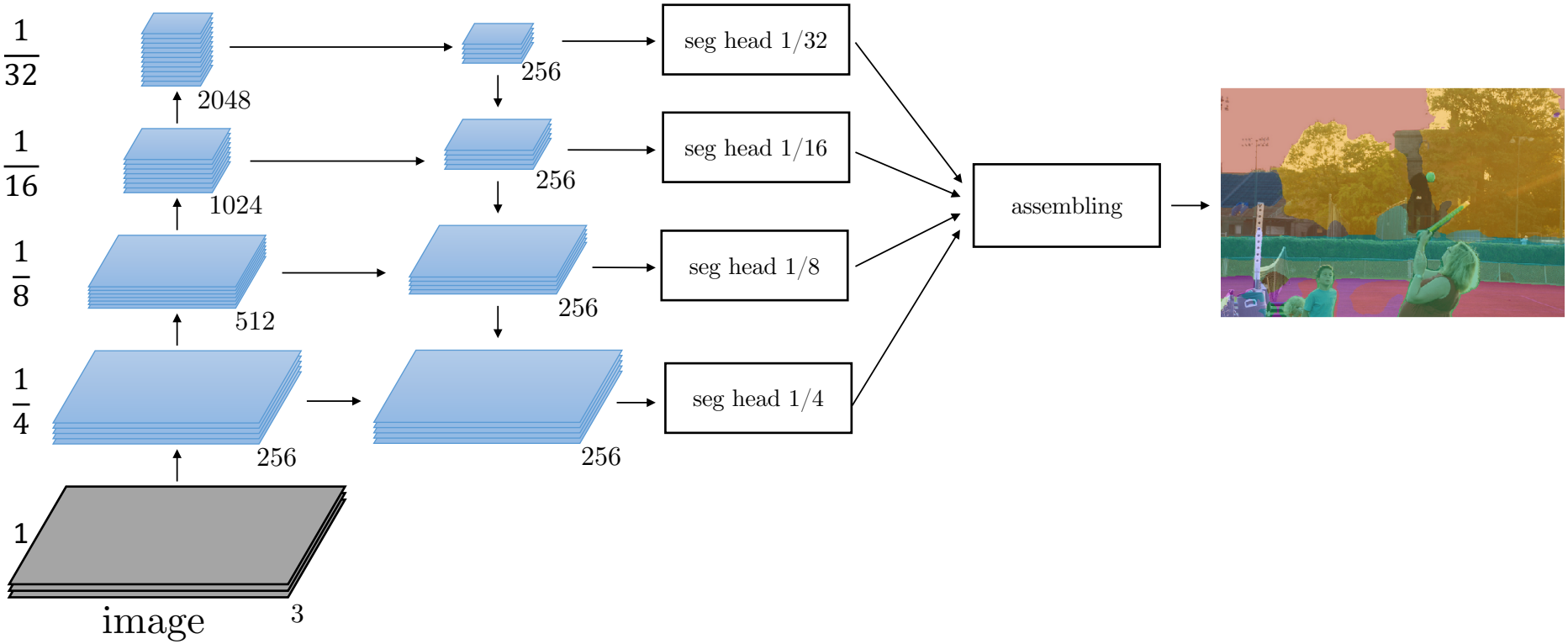


our work



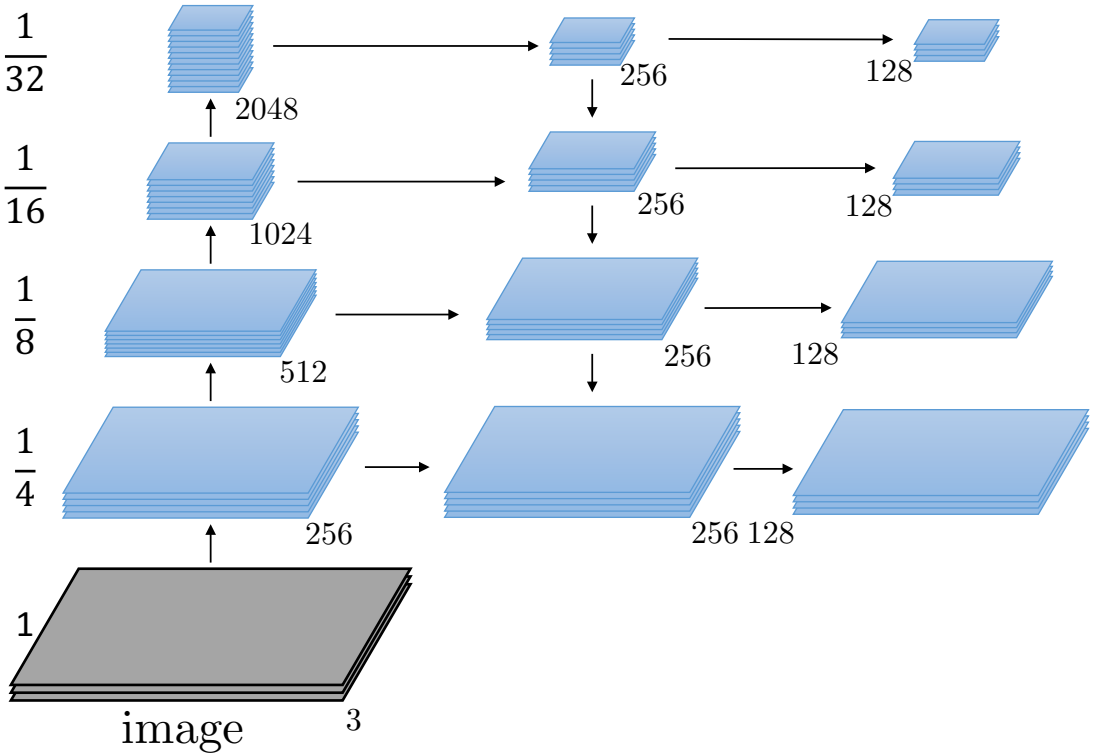
[1] He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. CVPR 2016.
[2] Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. Aggregated residual transformations for deep neural networks. CVPR 2017.
[3] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. Feature pyramid networks for object detection. CVPR 2017.
[4] He, K., Gkioxari, G., Dollár, P., & Girshick, R. Mask R-CNN. ICCV 2017.

FPN for Semantic Segmentation



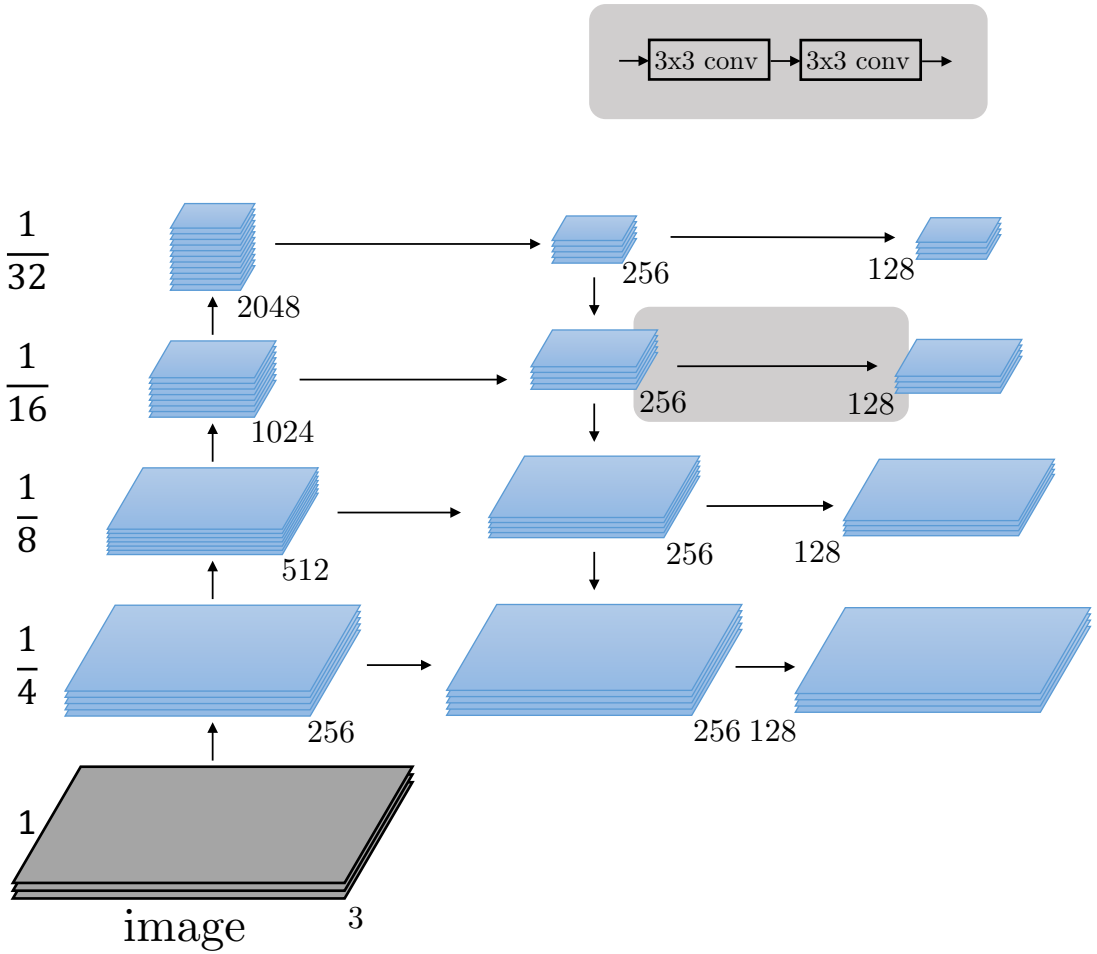
[1] He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. CVPR 2016.
[2] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. Feature pyramid networks for object detection. CVPR 2017.
[3] He, K., Gkioxari, G., Dollár, P., & Girshick, R. Mask R-CNN. ICCV 2017.

FPN for Semantic Segmentation



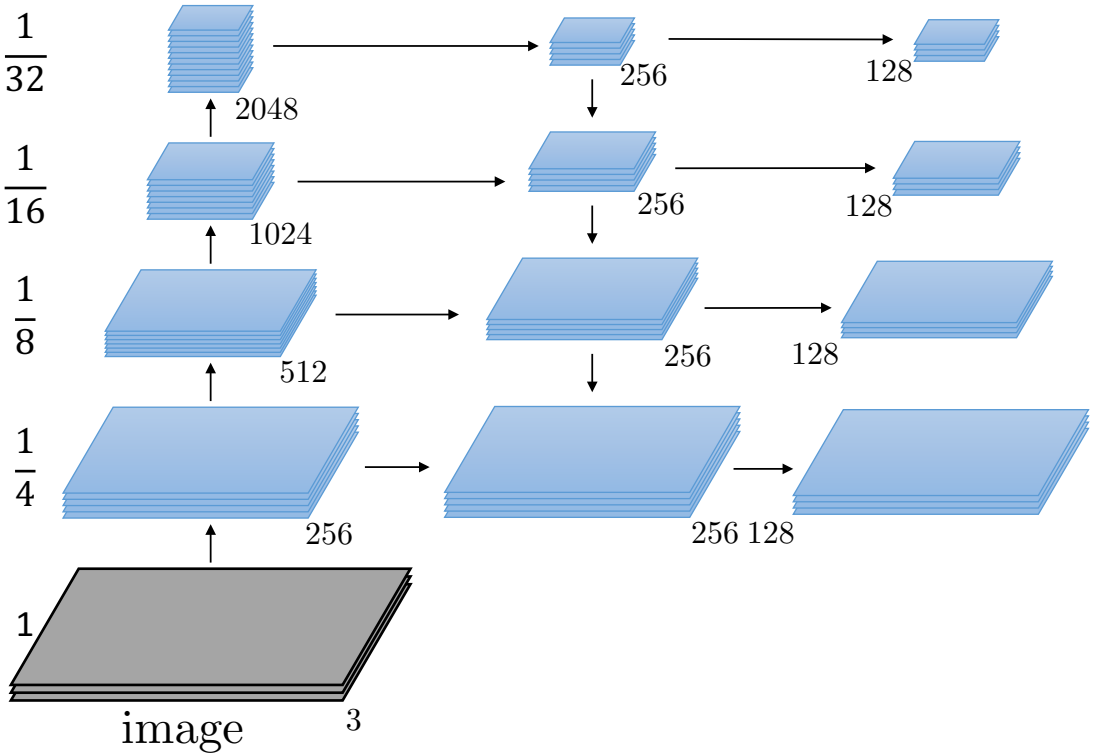
[1] He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. CVPR 2016.
[2] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. Feature pyramid networks for object detection. CVPR 2017.
[3] He, K., Gkioxari, G., Dollár, P., & Girshick, R. Mask R-CNN. ICCV 2017.

FPN for Semantic Segmentation



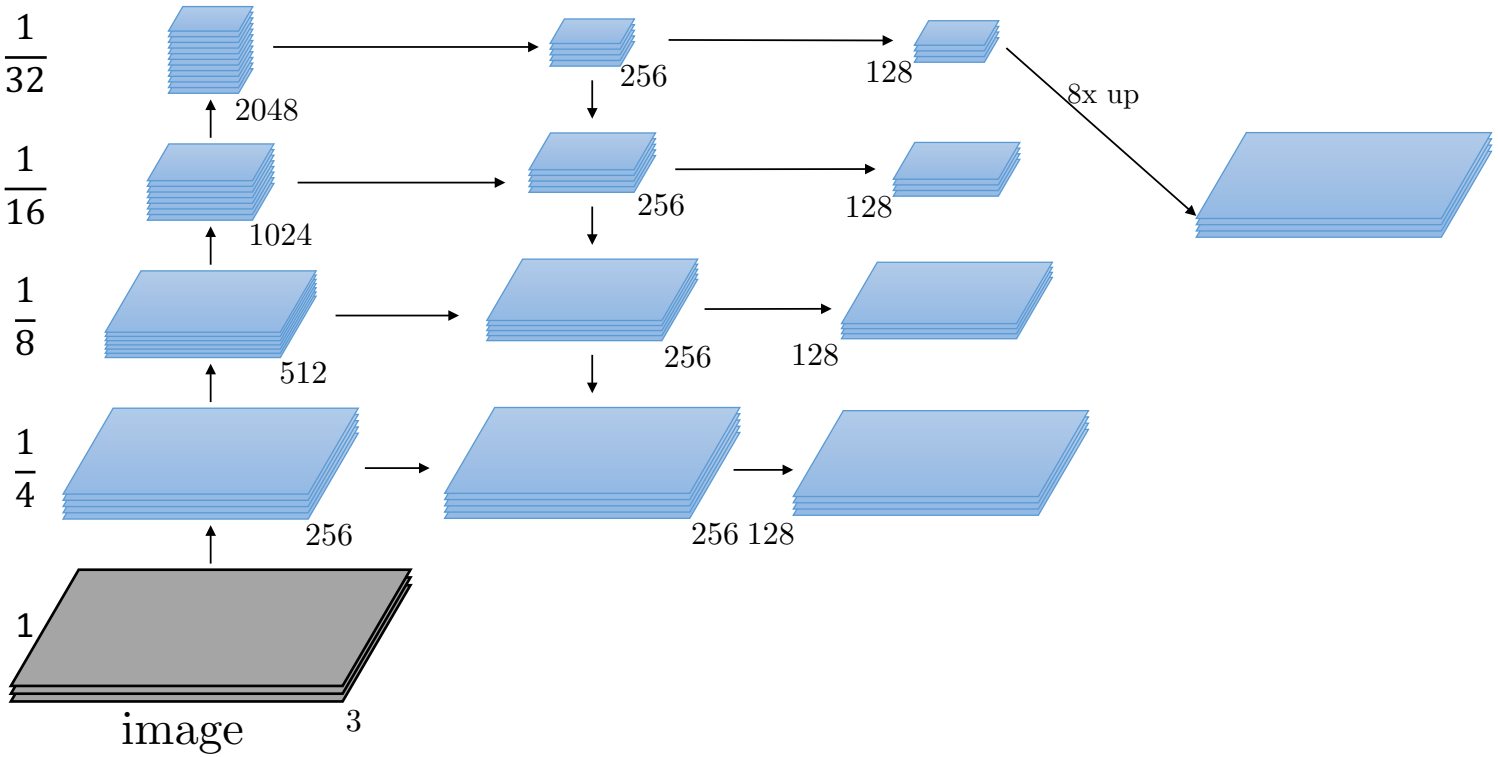
[1] He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. CVPR 2016.
[2] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. Feature pyramid networks for object detection. CVPR 2017.
[3] He, K., Gkioxari, G., Dollár, P., & Girshick, R. Mask R-CNN. ICCV 2017.

FPN for Semantic Segmentation



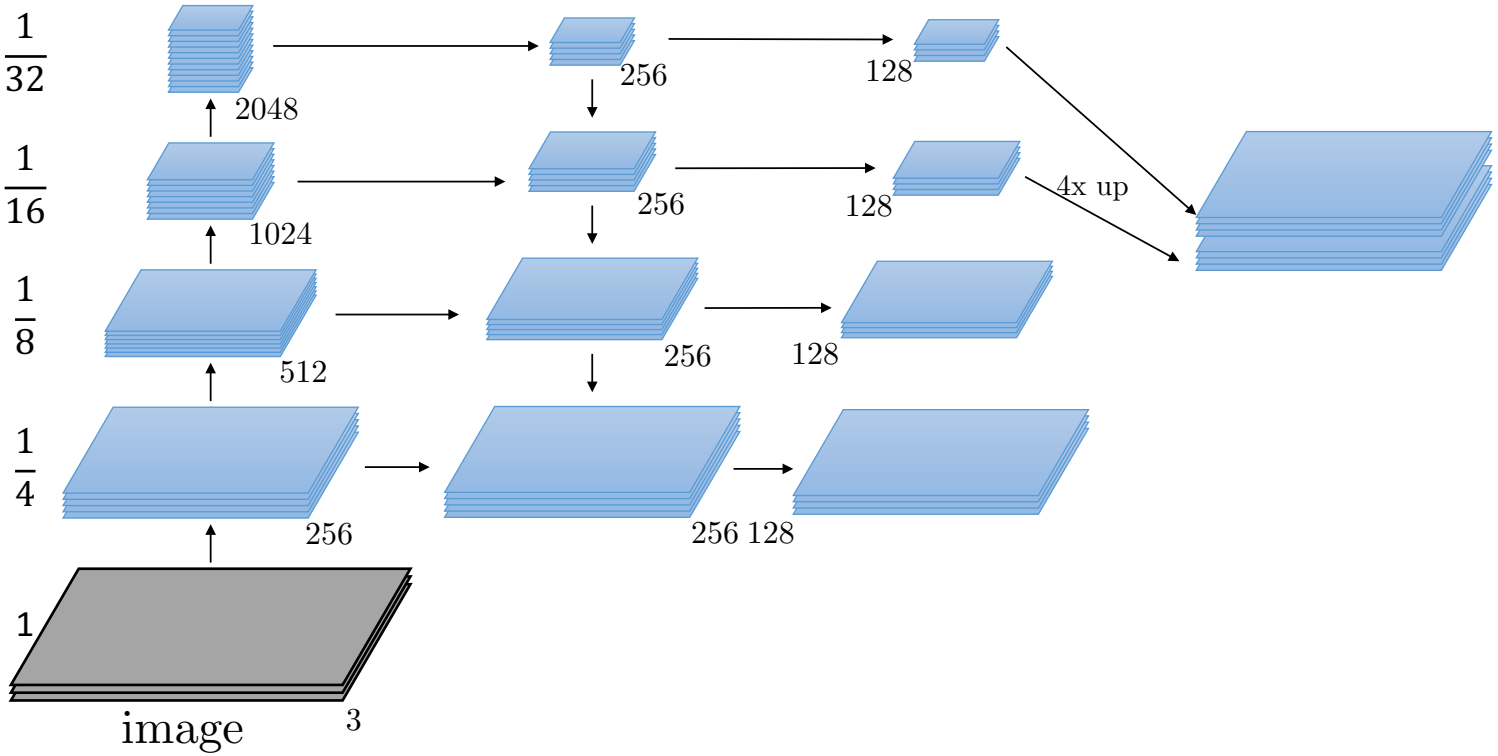
[1] He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. CVPR 2016.
[2] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. Feature pyramid networks for object detection. CVPR 2017.
[3] He, K., Gkioxari, G., Dollár, P., & Girshick, R. Mask R-CNN. ICCV 2017.

FPN for Semantic Segmentation



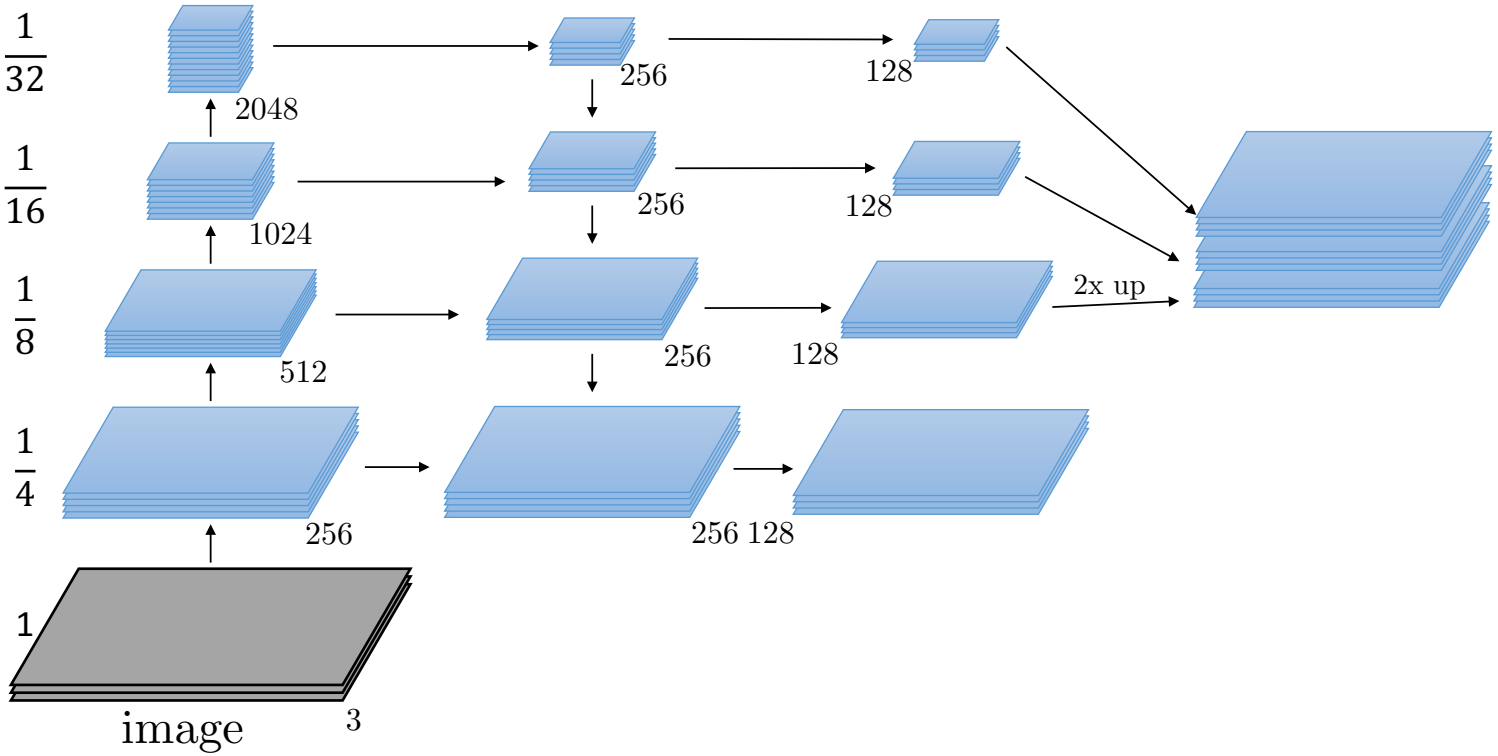
[1] He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. CVPR 2016.
[2] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. Feature pyramid networks for object detection. CVPR 2017.
[3] He, K., Gkioxari, G., Dollár, P., & Girshick, R. Mask R-CNN. ICCV 2017.

FPN for Semantic Segmentation



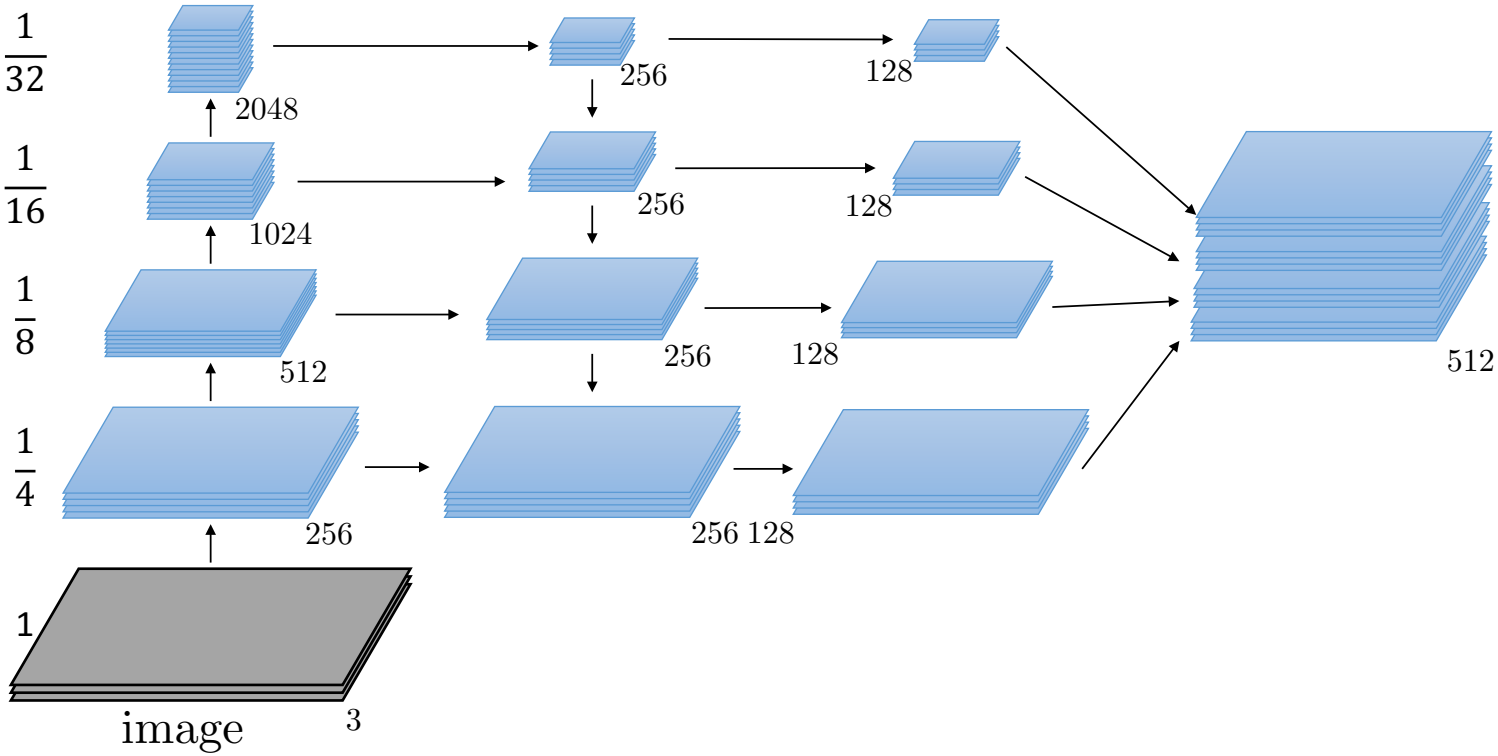
[1] He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. CVPR 2016.
[2] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. Feature pyramid networks for object detection. CVPR 2017.
[3] He, K., Gkioxari, G., Dollár, P., & Girshick, R. Mask R-CNN. ICCV 2017.

FPN for Semantic Segmentation



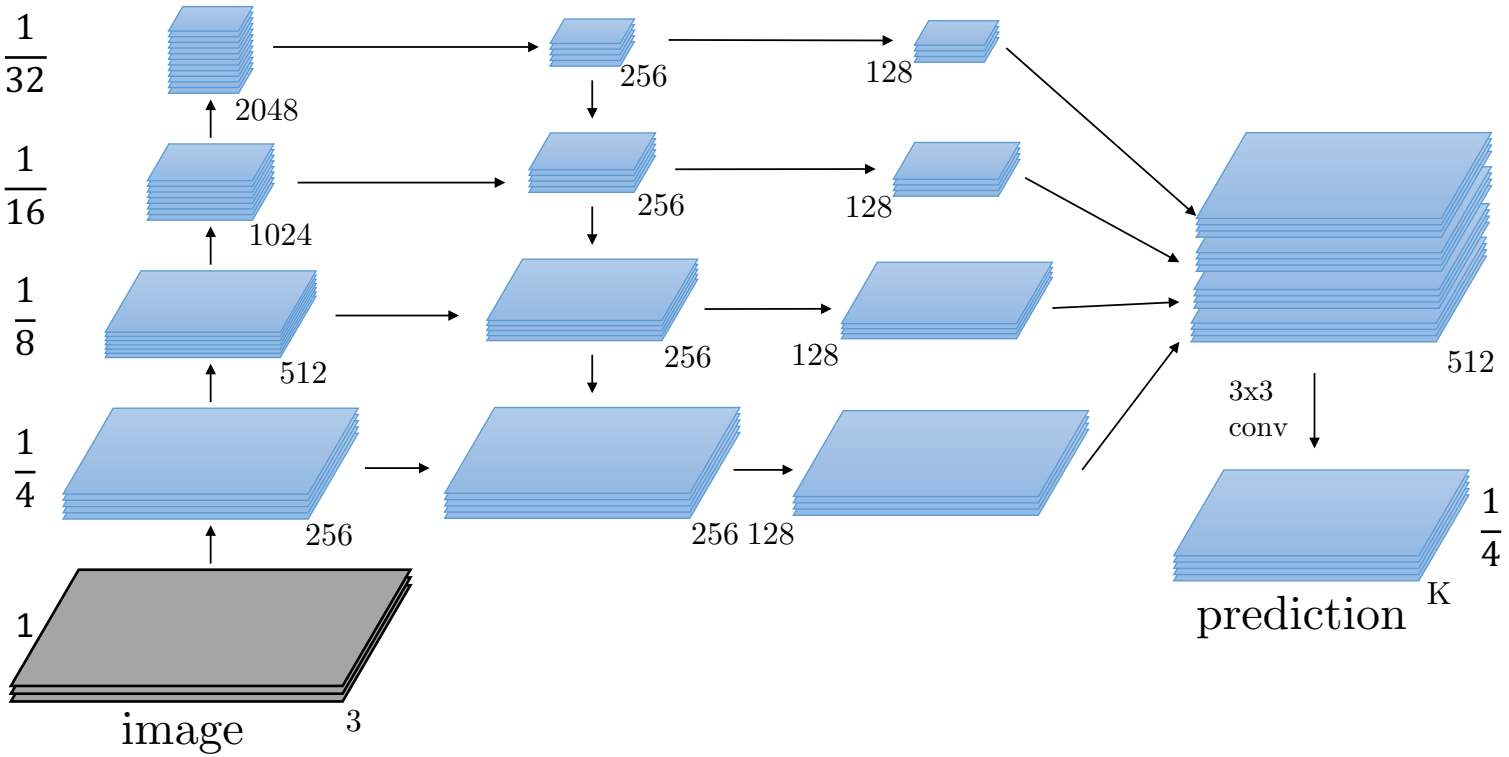
[1] He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. CVPR 2016.
[2] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. Feature pyramid networks for object detection. CVPR 2017.
[3] He, K., Gkioxari, G., Dollár, P., & Girshick, R. Mask R-CNN. ICCV 2017.

FPN for Semantic Segmentation



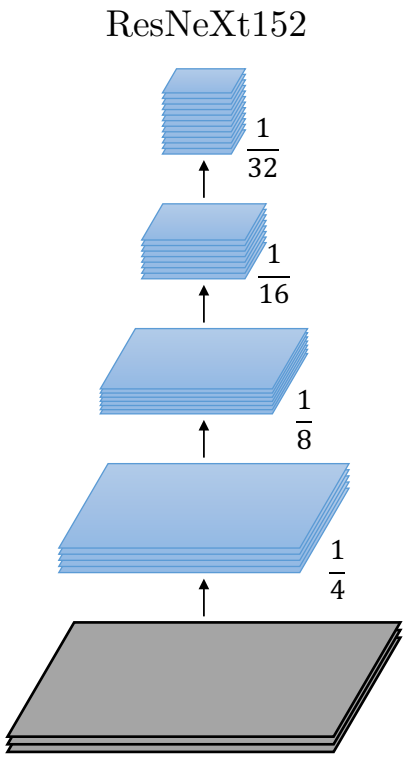
[1] He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. CVPR 2016.
[2] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. Feature pyramid networks for object detection. CVPR 2017.
[3] He, K., Gkioxari, G., Dollár, P., & Girshick, R. Mask R-CNN. ICCV 2017.

FPN for Semantic Segmentation

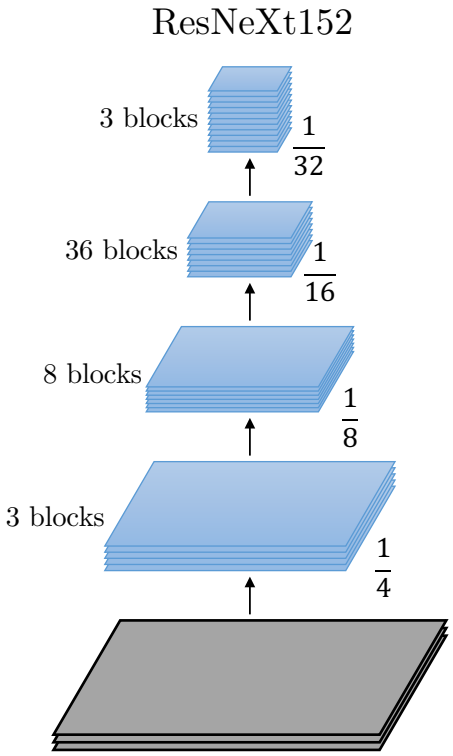


[1] He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. CVPR 2016.
[2] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. Feature pyramid networks for object detection. CVPR 2017.
[3] He, K., Gkioxari, G., Dollár, P., & Girshick, R. Mask R-CNN. ICCV 2017.

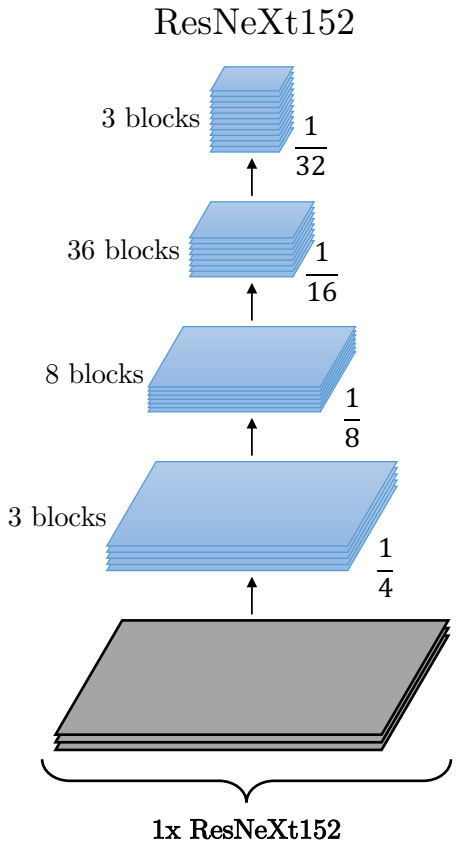
ResNeXt-FPN Efficiency

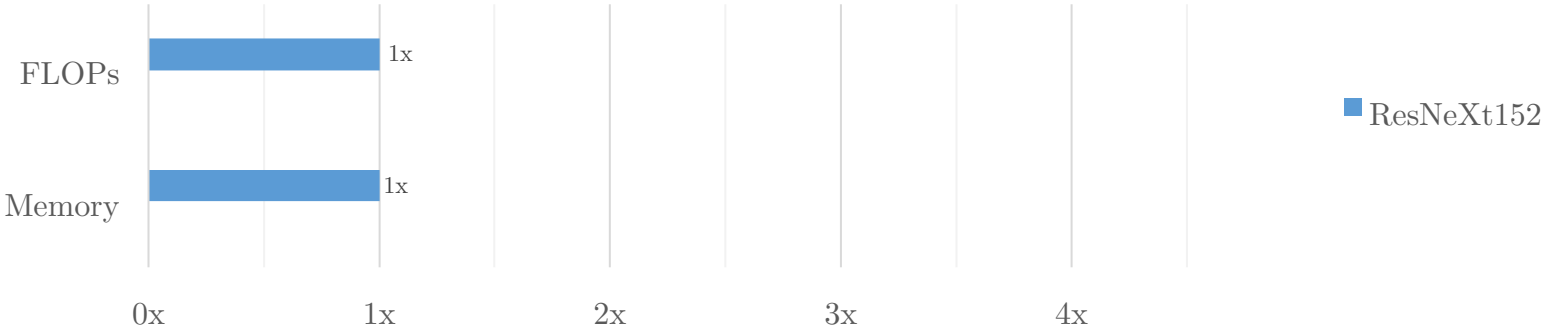


ResNeXt-FPN Efficiency



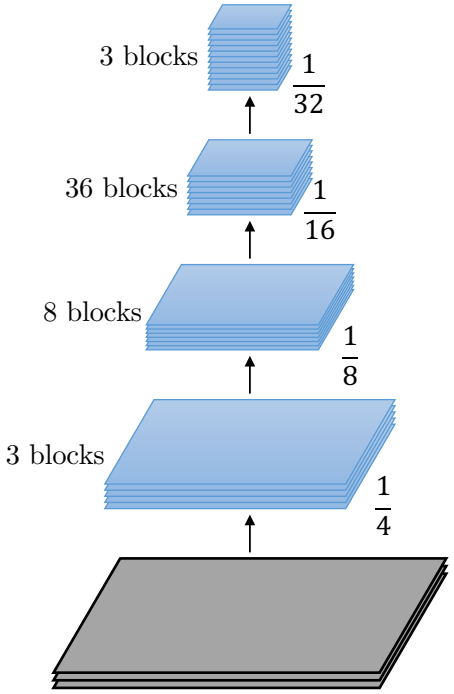
ResNeXt-FPN Efficiency

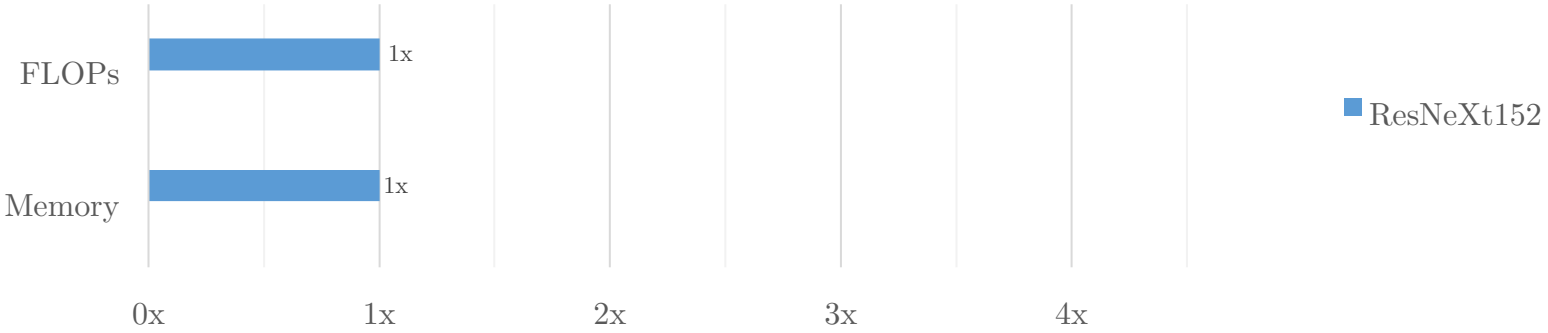




ResNeXt-FPN Efficiency

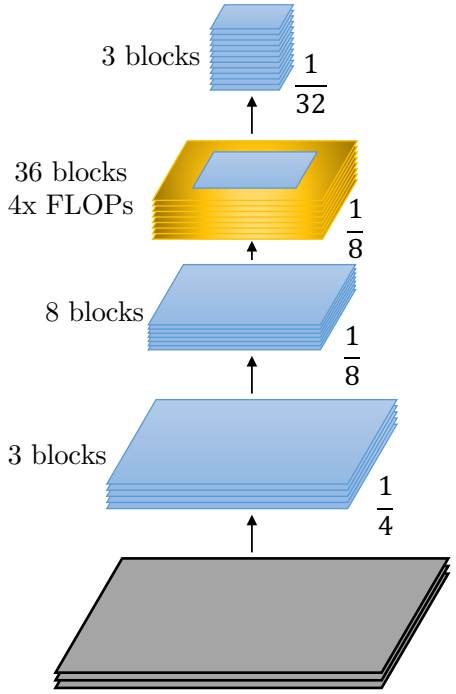
ResNeXt152-dilation (stride 8)

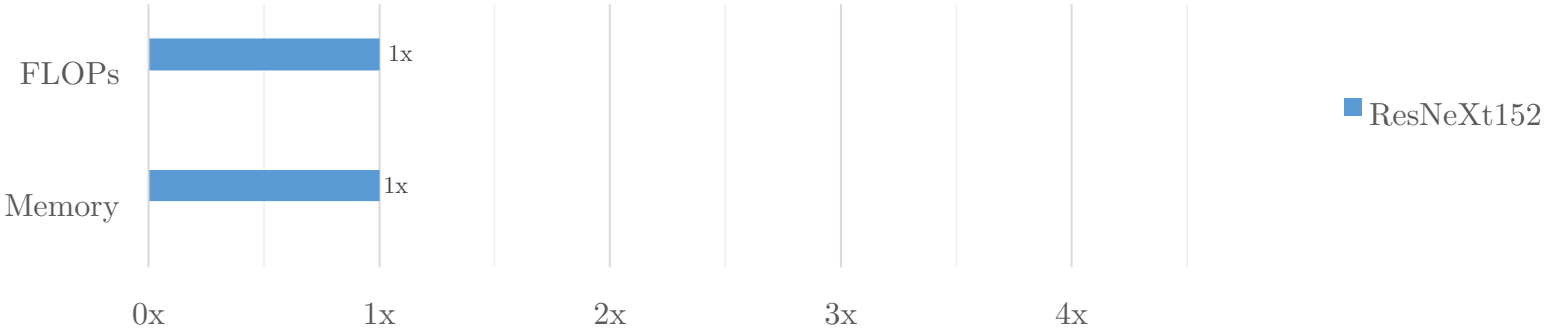




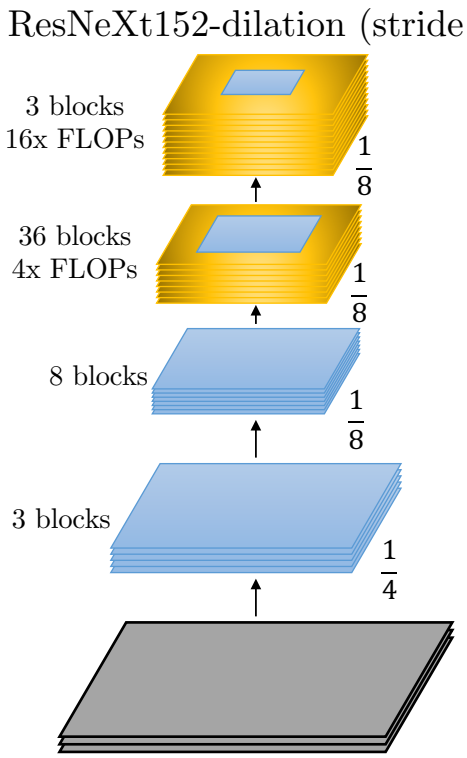
ResNeXt-FPN Efficiency

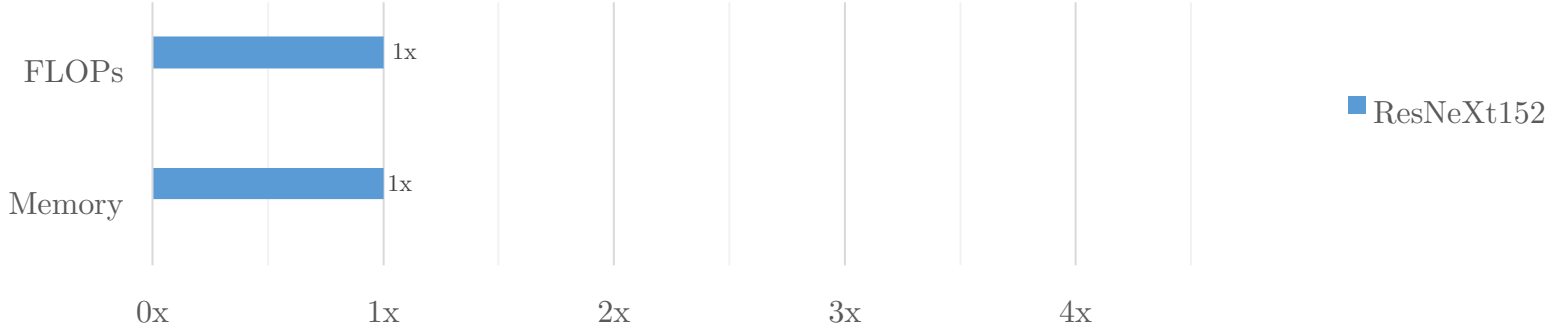
ResNeXt152-dilation (stride 8)



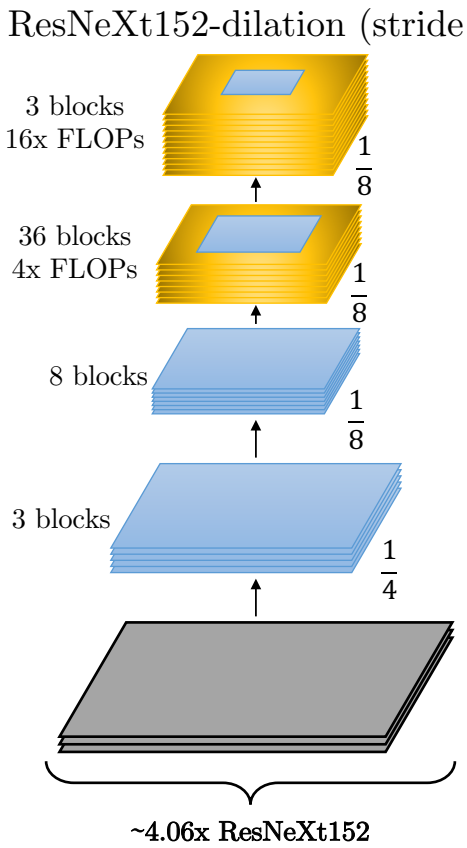


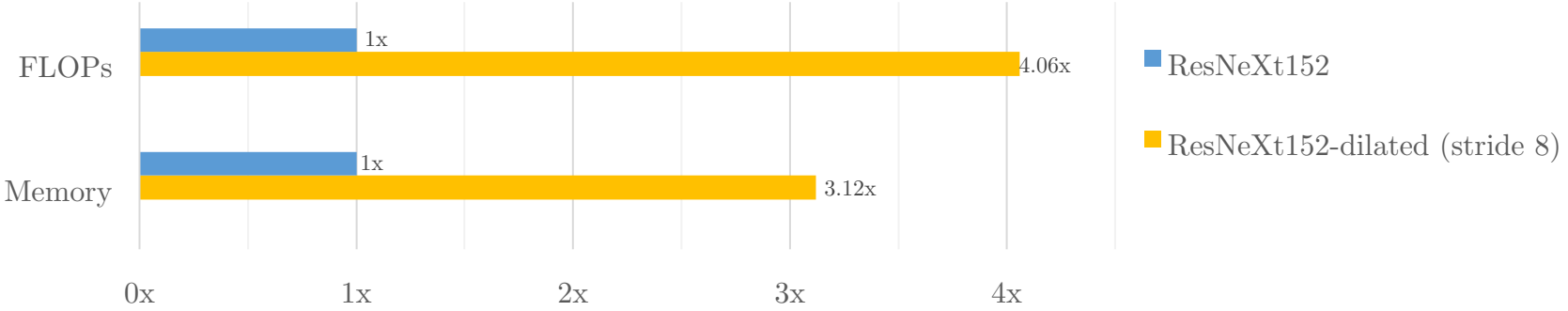
ResNeXt-FPN Efficiency



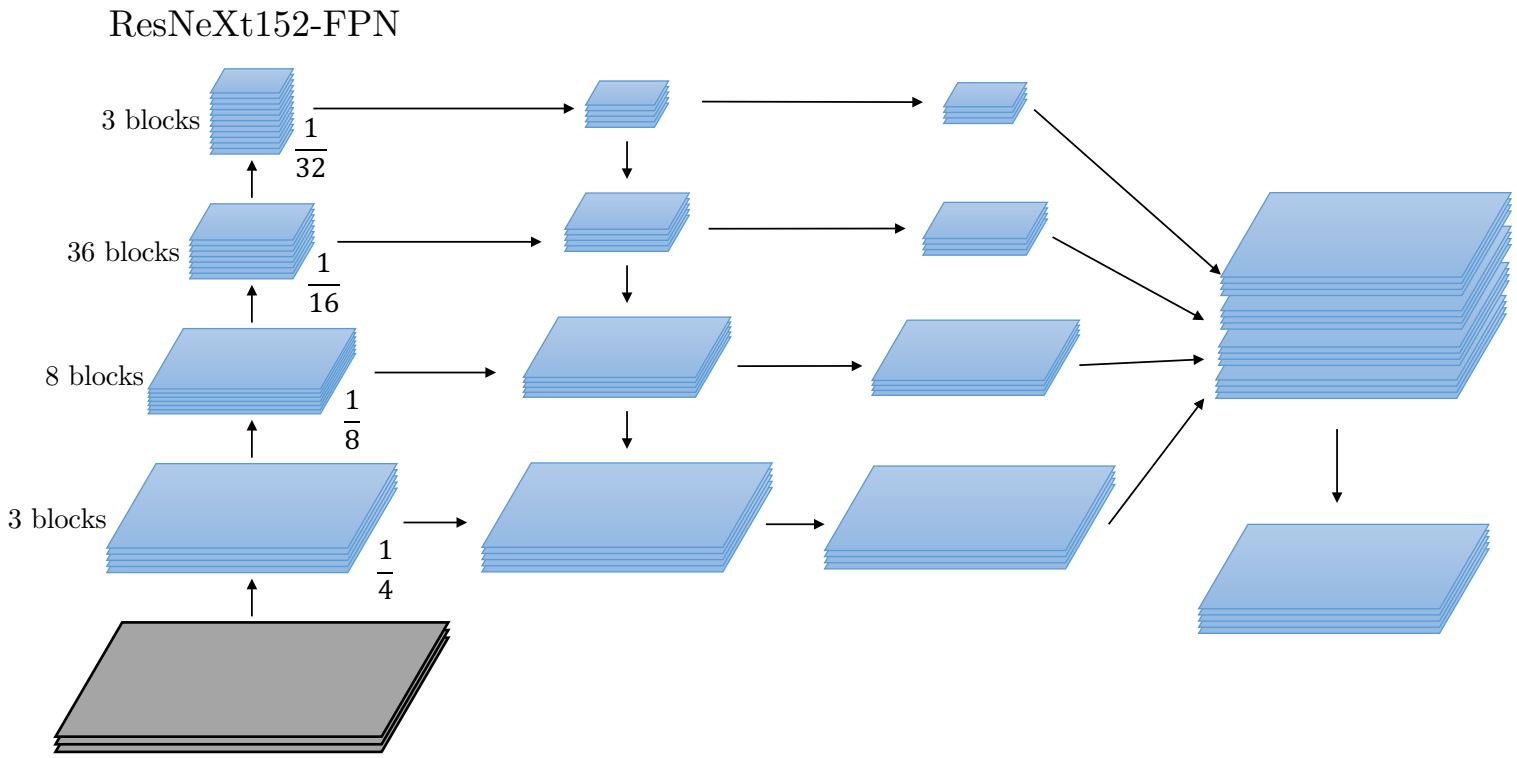


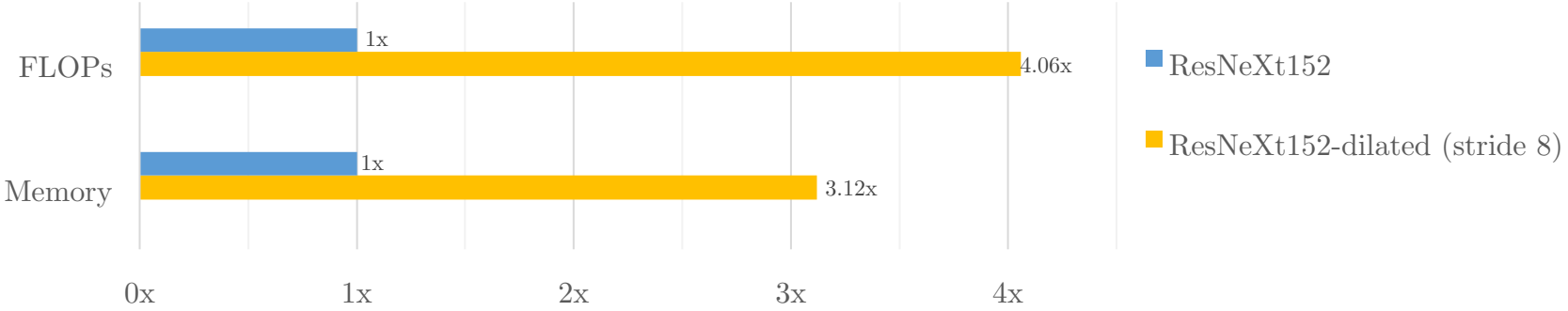
ResNeXt-FPN Efficiency



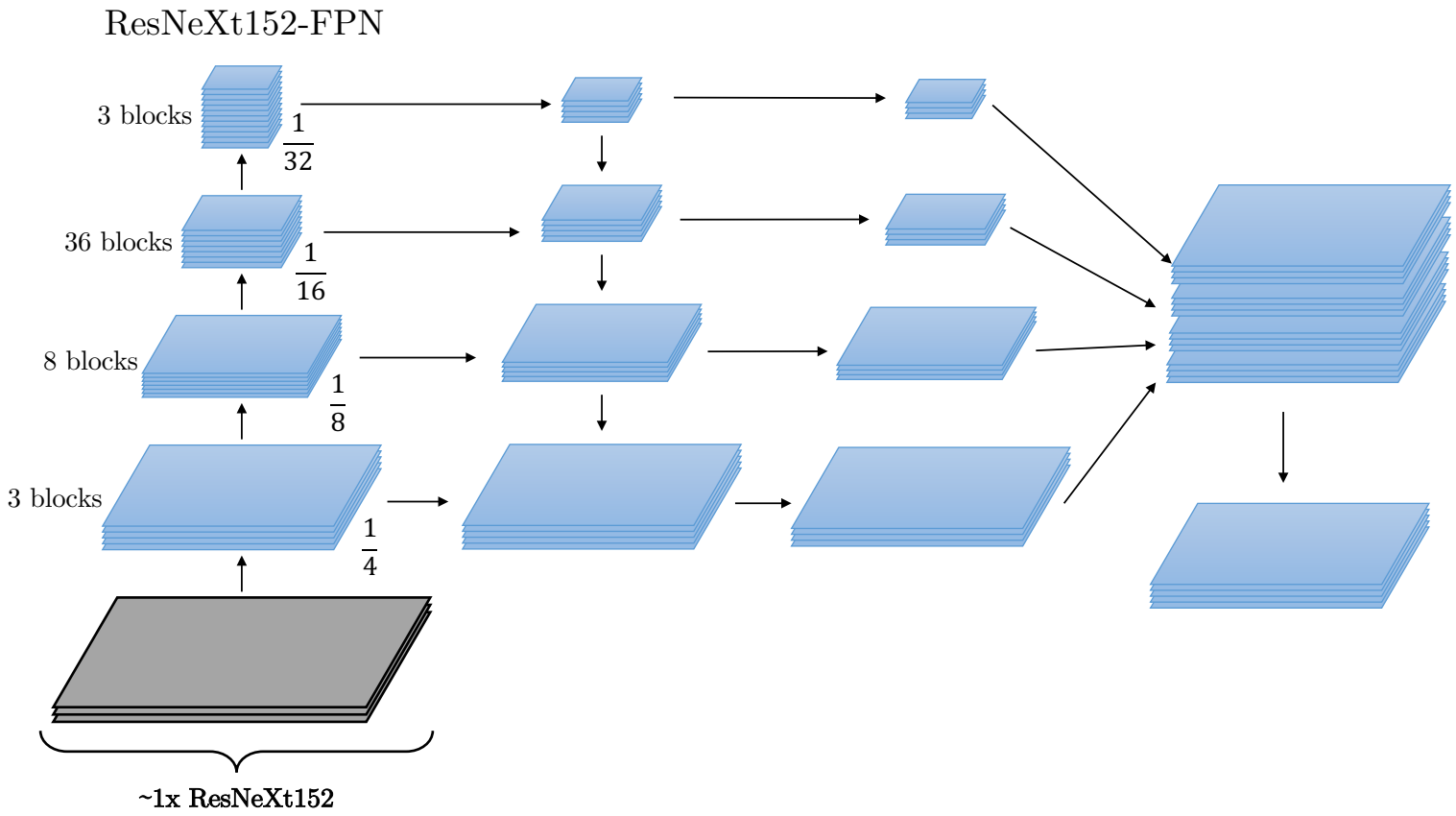


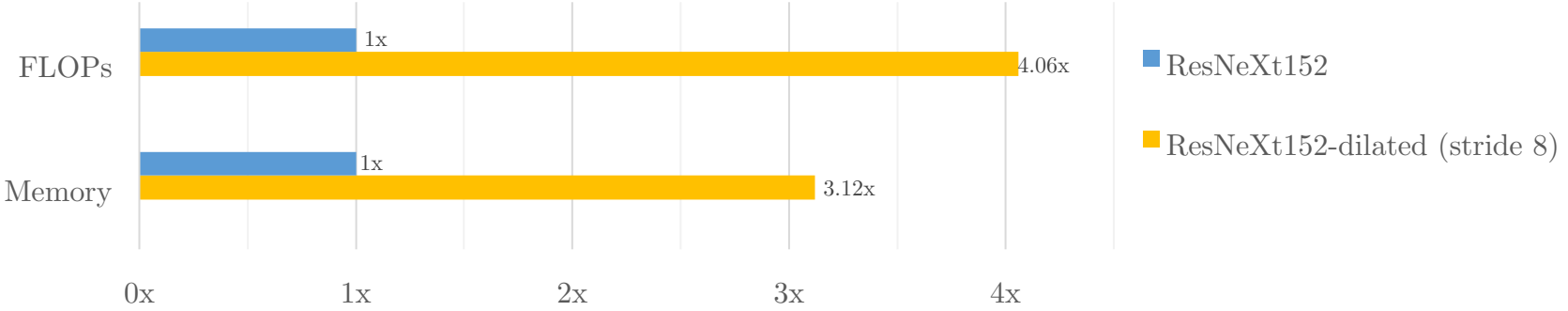
ResNeXt-FPN Efficiency



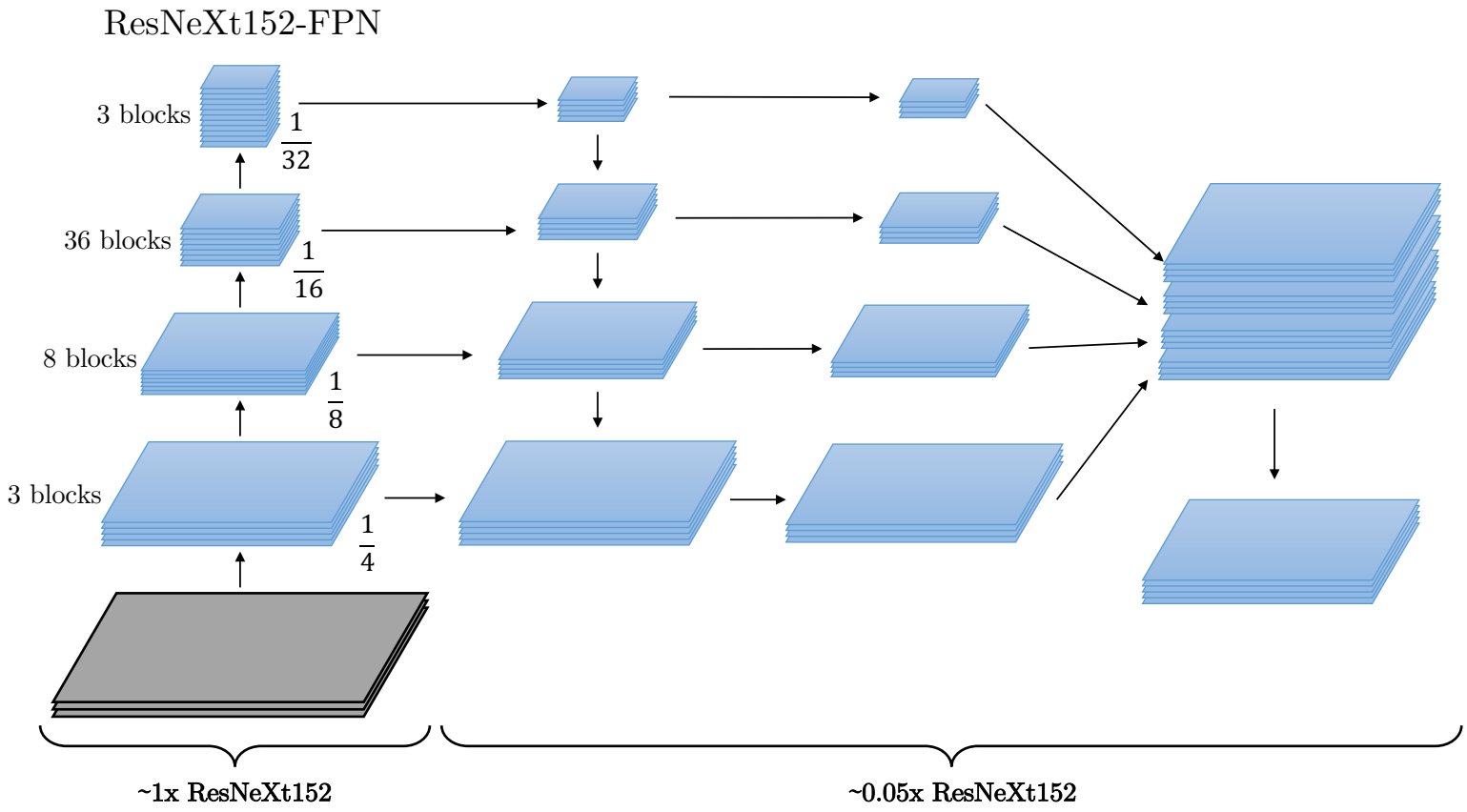


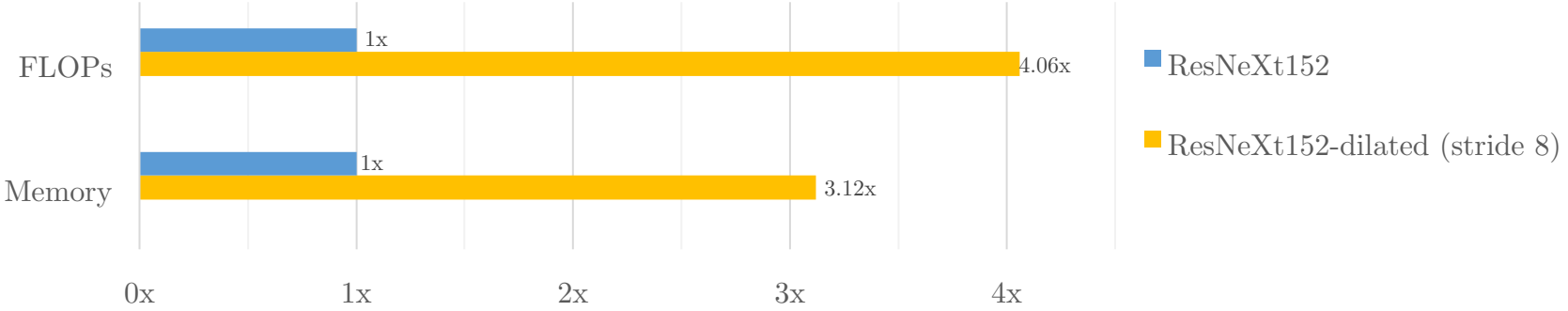
ResNeXt-FPN Efficiency



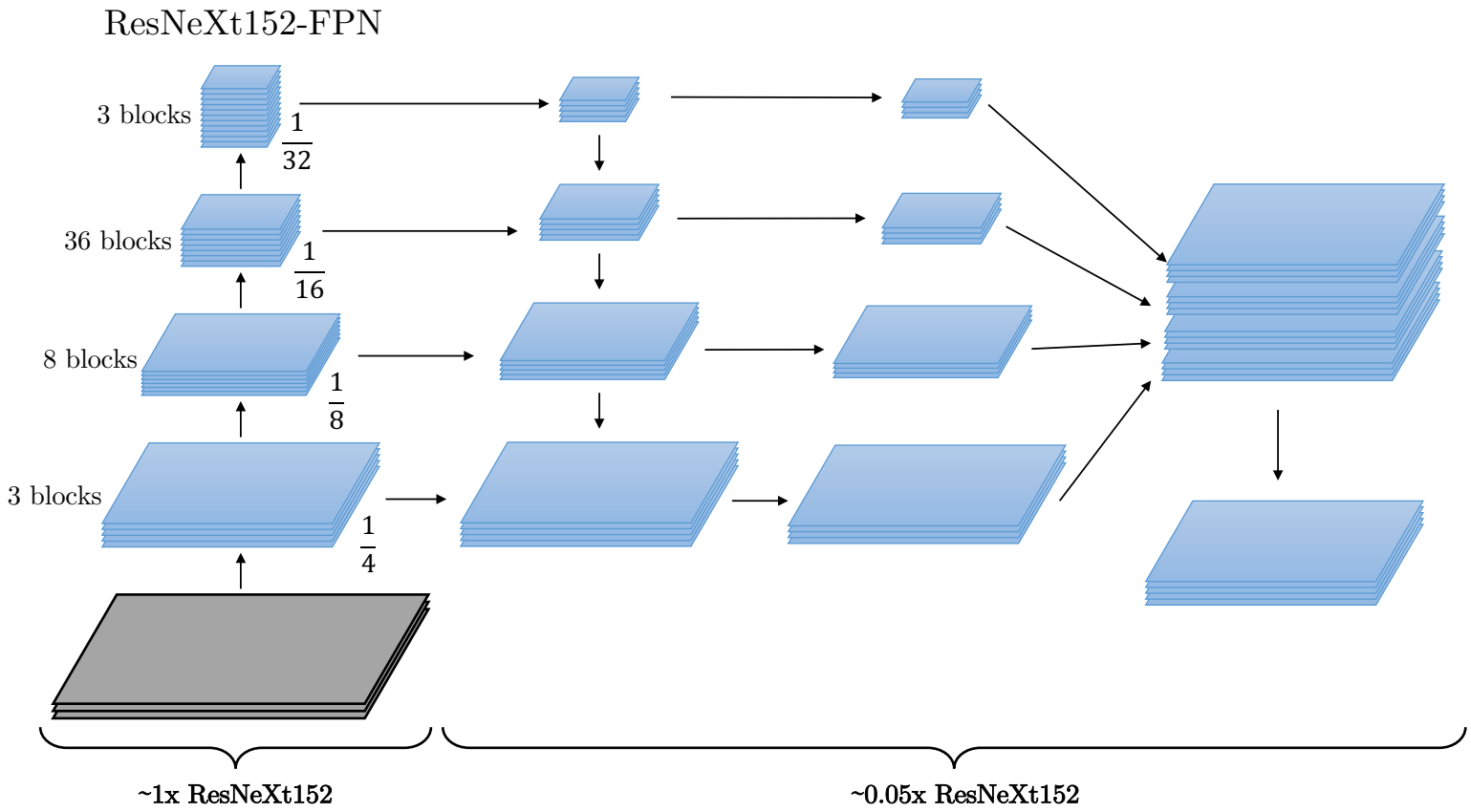


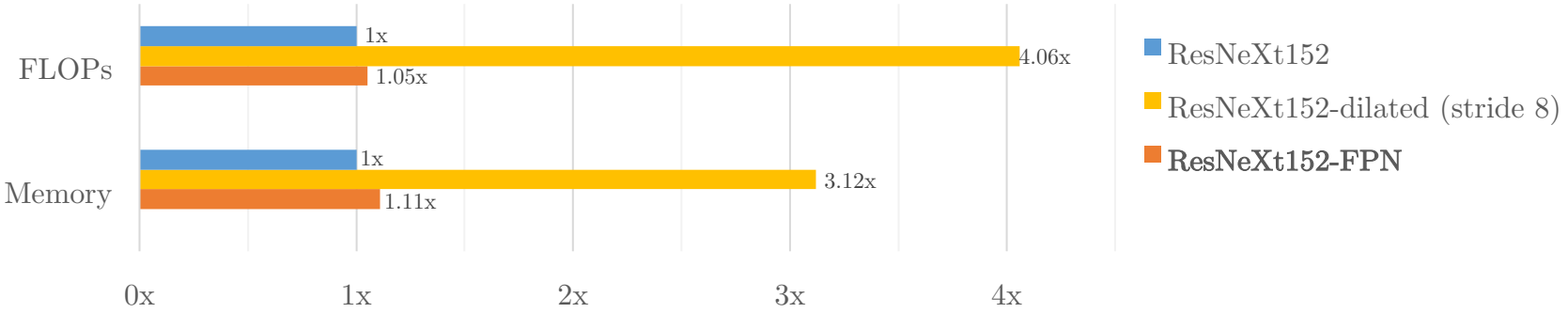
ResNeXt-FPN Efficiency





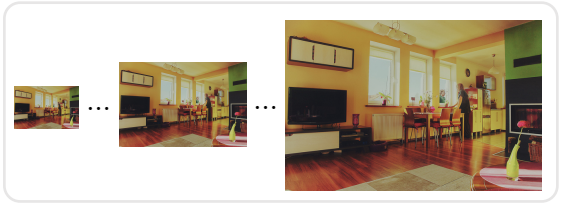
ResNeXt-FPN Efficiency



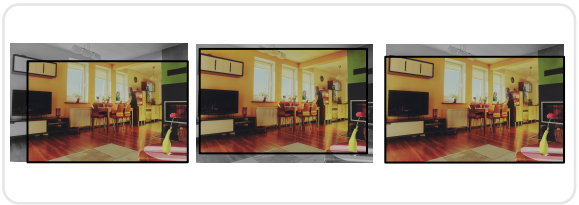


ResNeXt-FPN Training Details

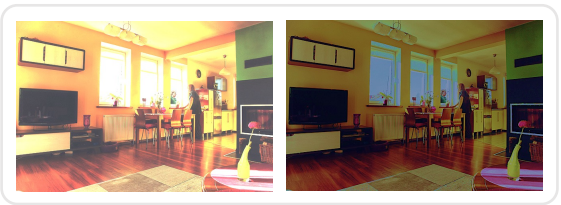
Data augmentation:



Scaling: 0.5x – 2.0x



Crop: 0.9x – 1.0x (max 800x800)

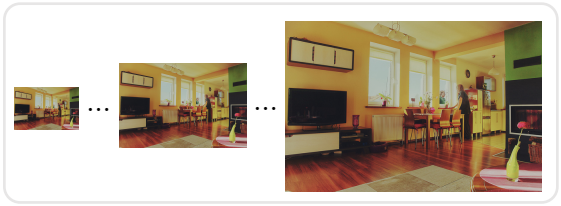


Color augmentation [1]

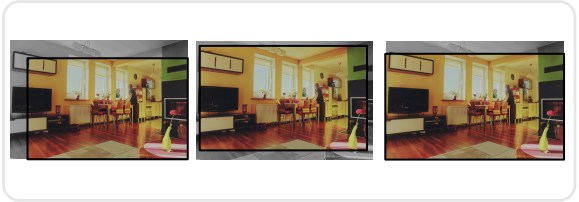
[1] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. SSD: Single shot multibox detector. ECCV 2016.
[2] Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. Aggregated residual transformations for deep neural networks. CVPR 2017.

ResNeXt-FPN Training Details

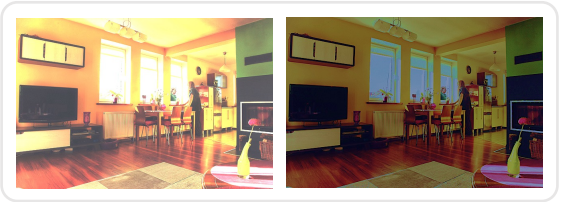
Data augmentation:



Scaling: 0.5x – 2.0x



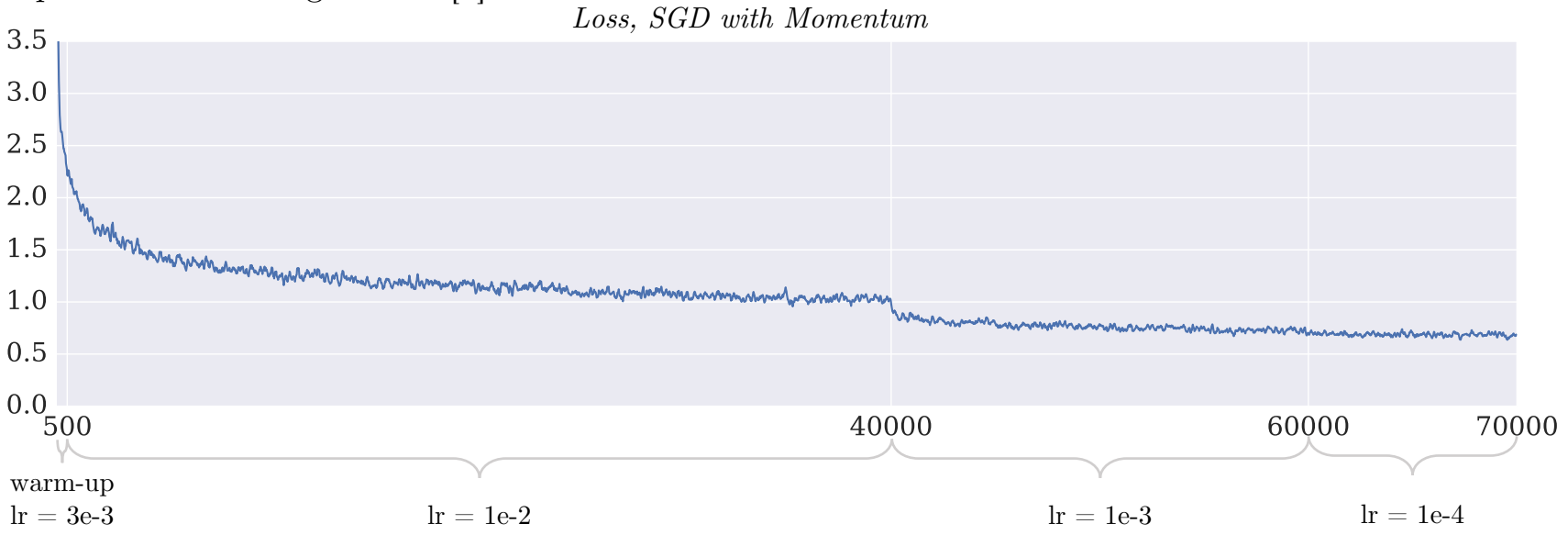
Crop: 0.9x – 1.0x (max 800x800)



Color augmentation [1]

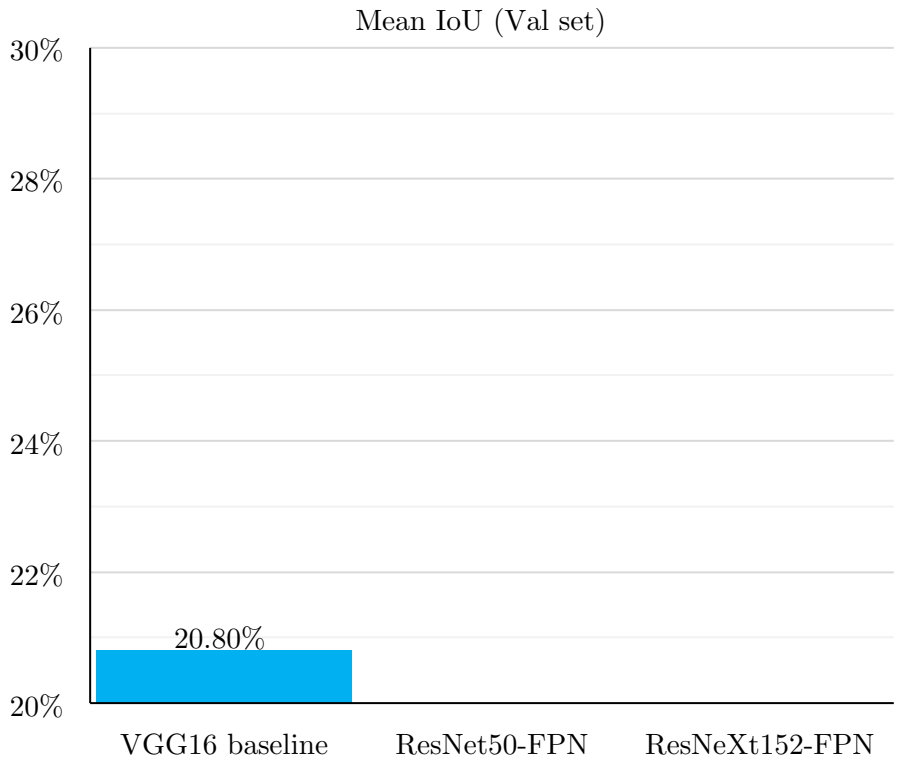
Train stats:

- training time: 27 hours (8 P100 GPUs)
- per GPU memory usage: 14.5GB
- batch size: 16 (2 x 8 GPU)
- pretrained on ImageNet-5k [2]



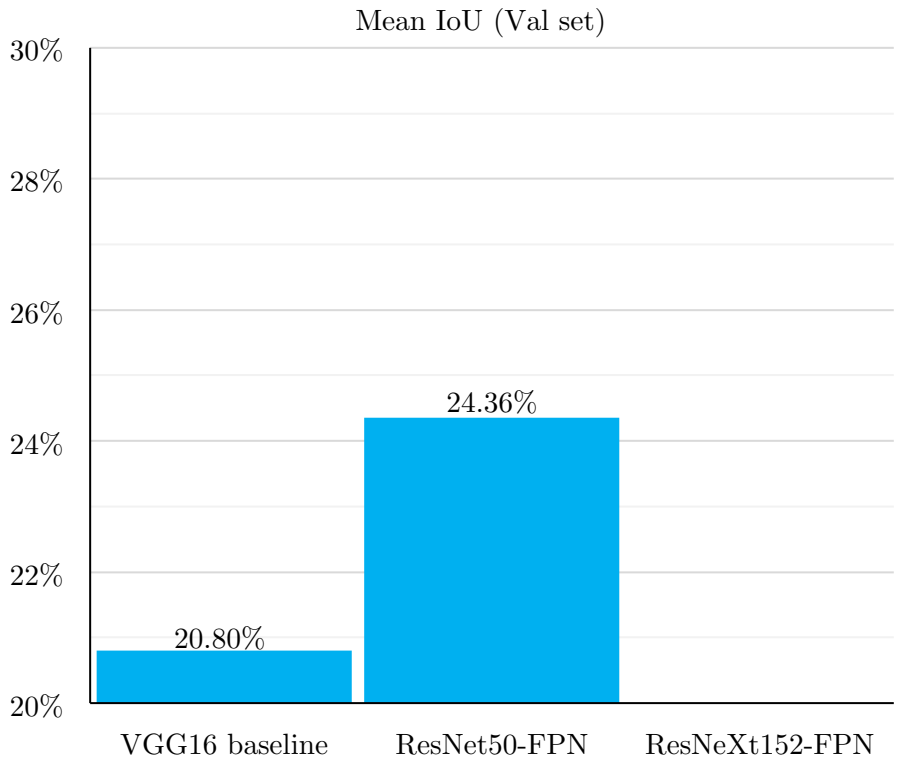
[1] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. SSD: Single shot multibox detector. ECCV 2016.
[2] Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. Aggregated residual transformations for deep neural networks. CVPR 2017.

ResNeXt-FPN Inference



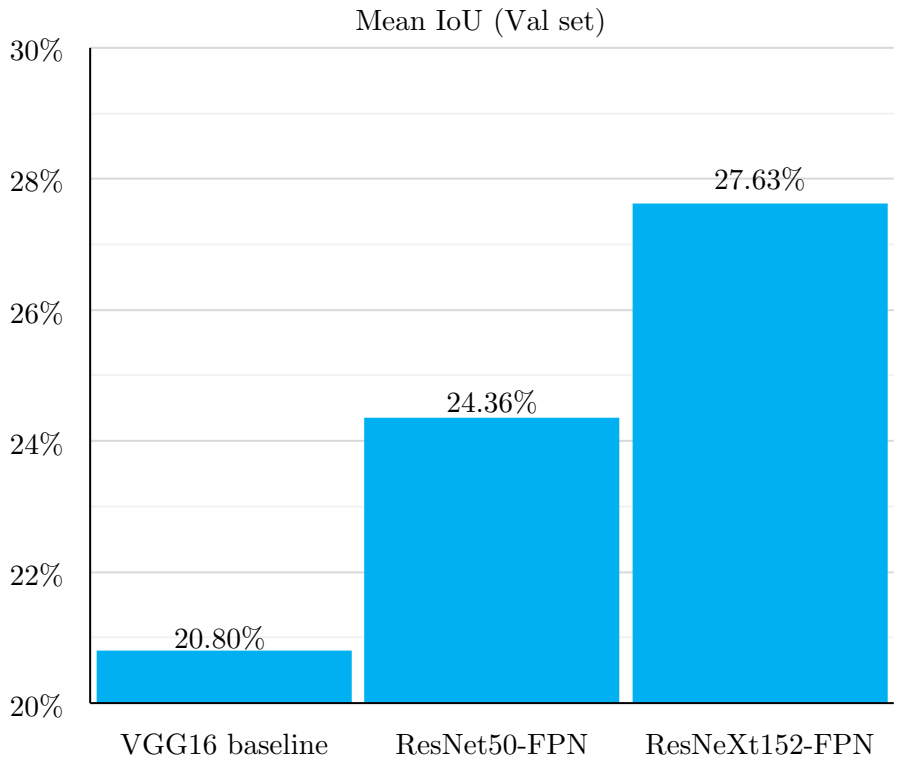
- no test time augmentation

ResNeXt-FPN Inference



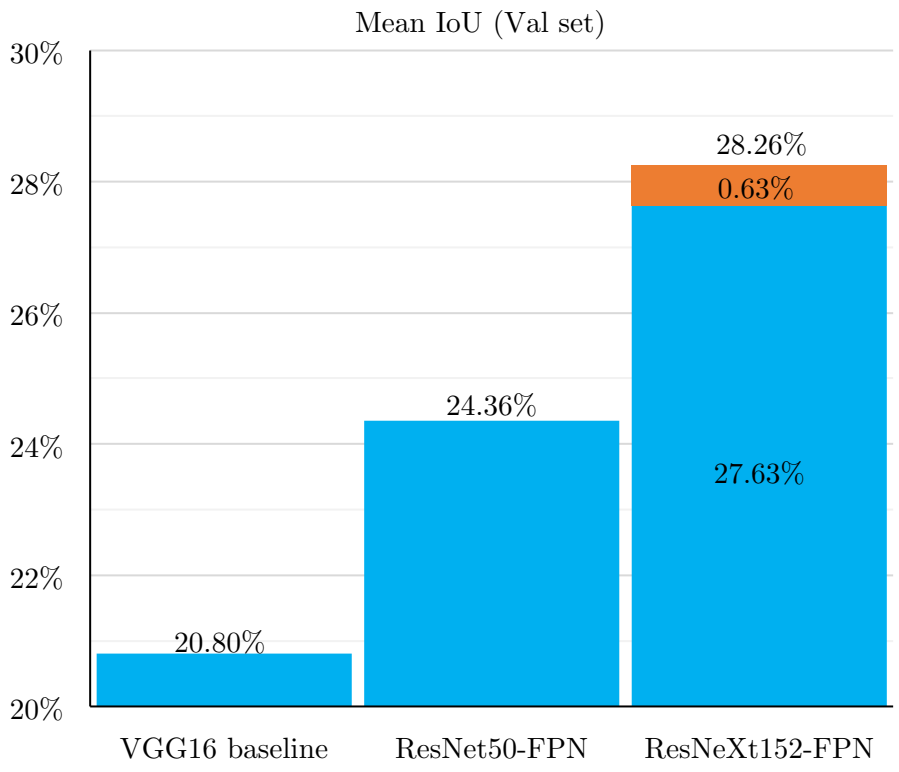
- no test time augmentation

ResNeXt-FPN Inference



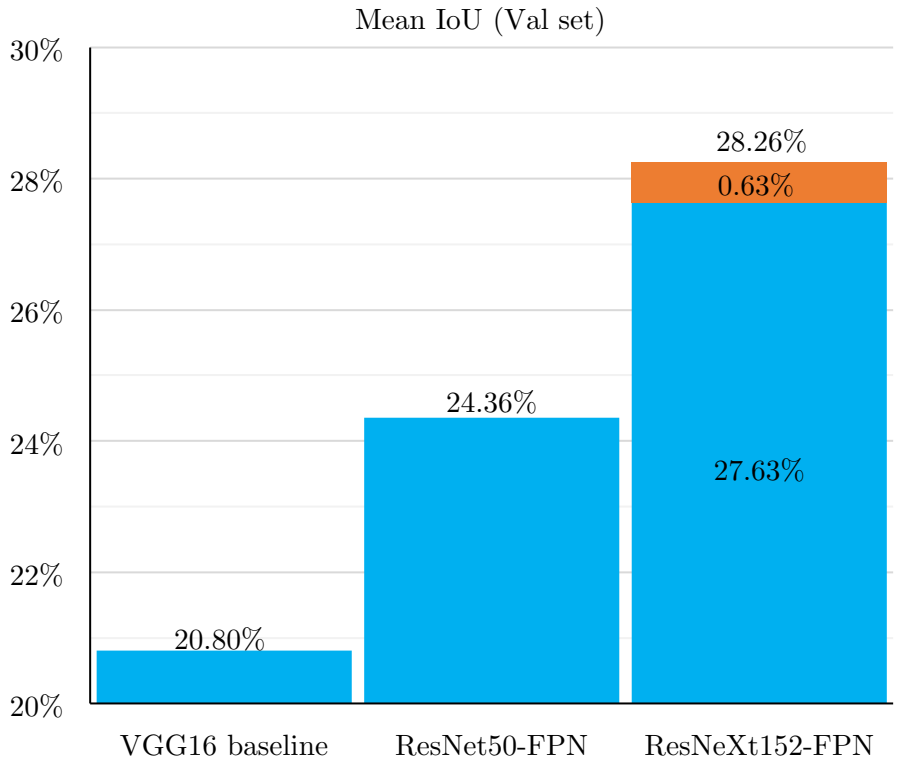
- no test time augmentation

ResNeXt-FPN Inference

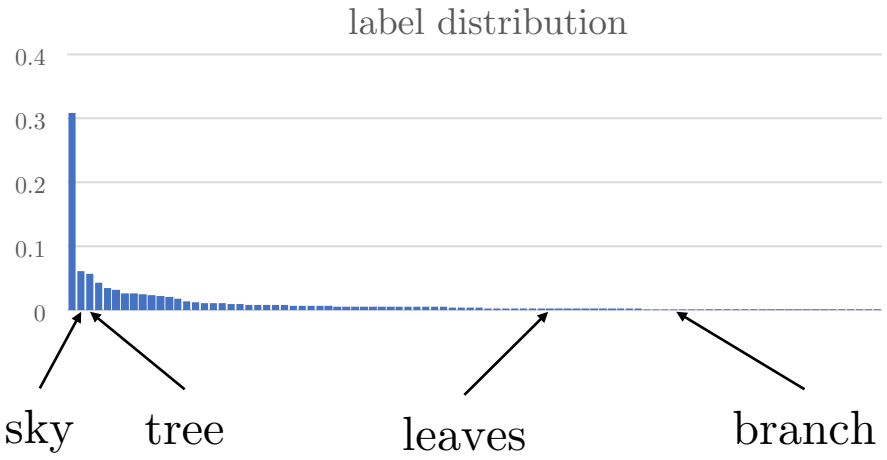


- no test time augmentation
- test time augmentation:
  - flip, multi-scale

ResNeXt-FPN Inference



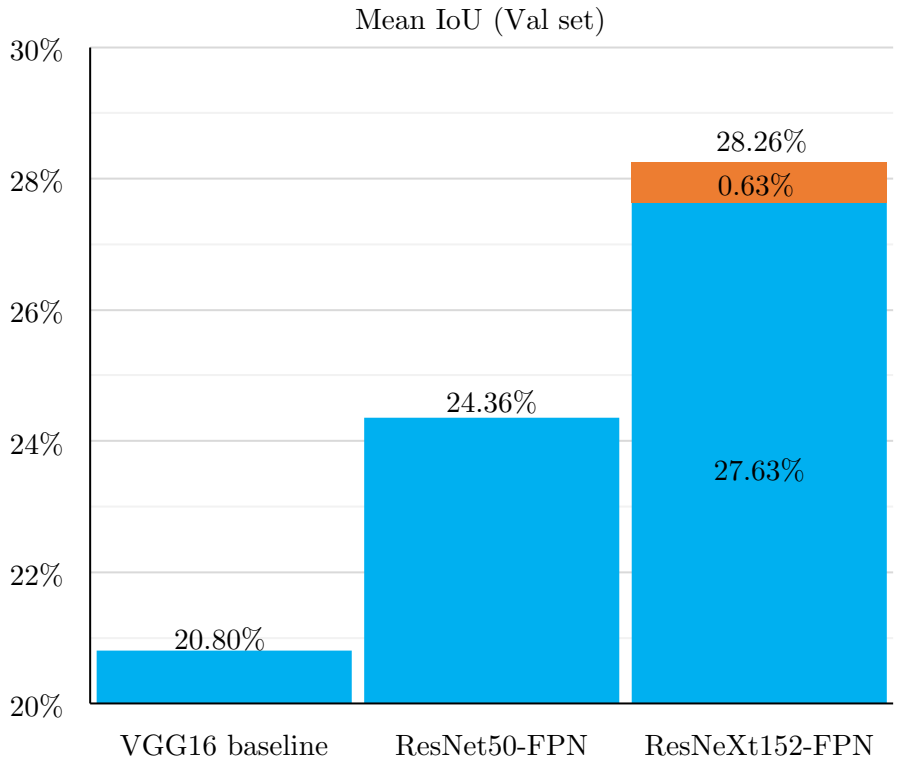
- no test time augmentation
- test time augmentation:
  - flip, multi-scale



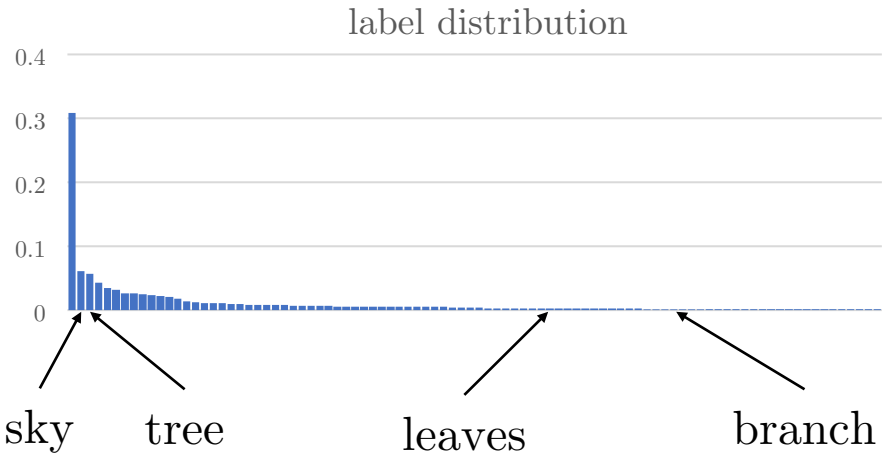
	sky	tree	leaves	branch
$\log p(label\|image)$	**69.8**	71.3	20.3	15.2

IoU per Category

ResNeXt-FPN Inference



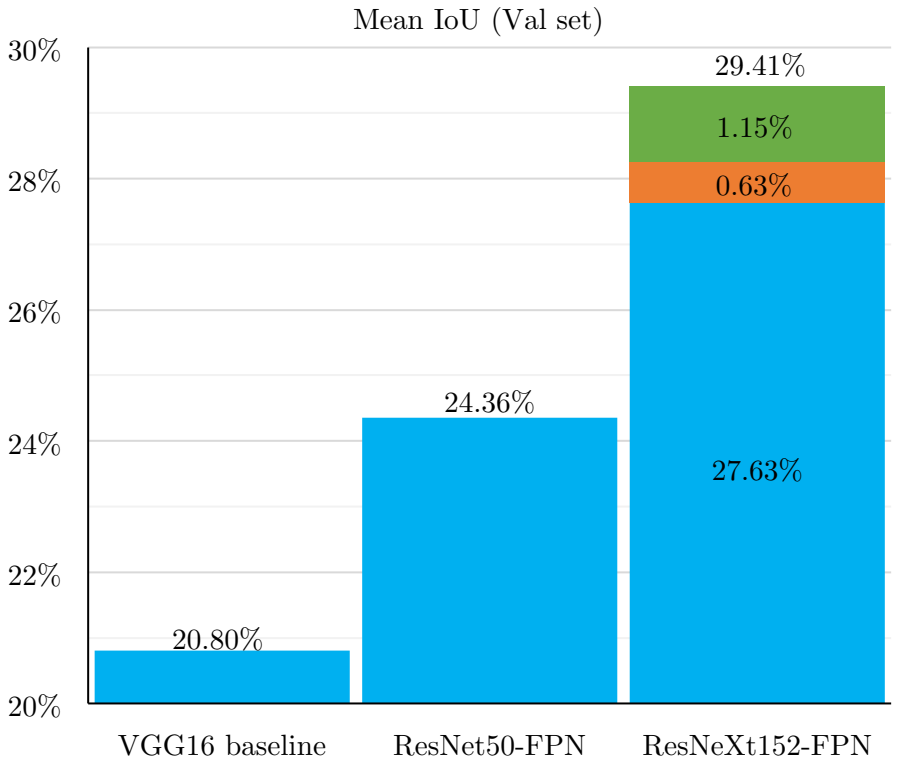
- no test time augmentation
- test time augmentation:
  - flip, multi-scale



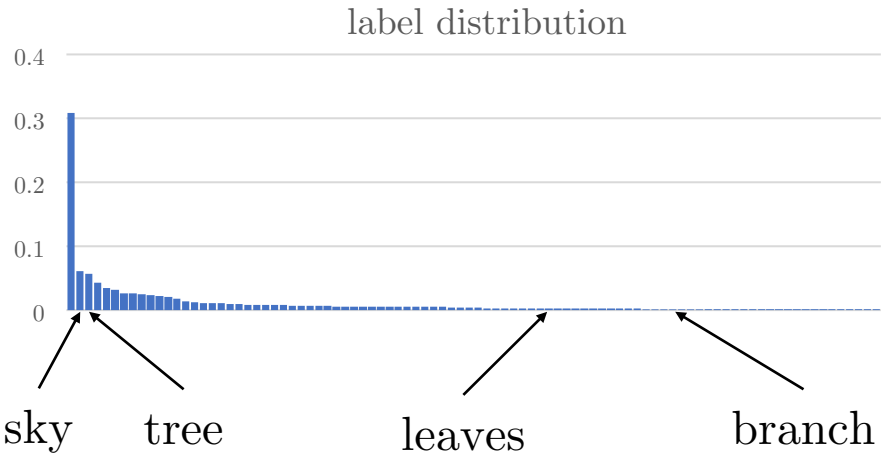
	sky	tree	leaves	branch
$\log p(label\|image)$	**69.8**	71.3	20.3	15.2
$\log p(label\|image) - \alpha \cdot \log p(label)$	69.7	**71.8**	**24.6**	**20.9**

IoU per Category

ResNeXt-FPN Inference



- no test time augmentation
- test time augmentation:
  - flip, multi-scale
- boost the score of rare labels



	sky	tree	leaves	branch
$\log p(label\|image)$	**69.8**	71.3	20.3	15.2
$\log p(label\|image) - \alpha \cdot \log p(label)$	69.7	**71.8**	**24.6**	**20.9**

IoU per Category

ResNeXt-FPN Take-Away

- Winning entry in COCO stuff 2017 competition

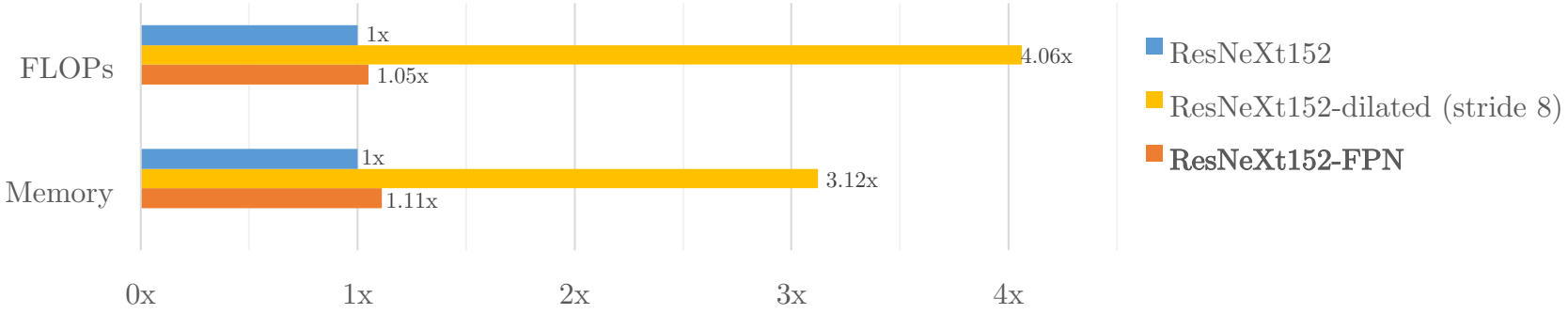
Team name	mIoU	fIoU	mACC	pACC	mIoUS	fIoUS	mACCS	pACCS
ResNeXt152-FPN	**28.80%**	**55.70%**	**42.30%**	**69.20%**	**56.20%**	**68.70%**	**70.30%**	**80.50%**
G-RMI	26.60%	51.90%	40.40%	65.40%	52.40%	64.70%	67.80%	77.40%
Oxford Active Vision Lab	24.20%	50.60%	34.80%	66.00%	50.30%	63.60%	62.20%	77.40%
Baseline Deeplab VGG-16	20.20%	47.60%	28.20%	64.70%	45.90%	60.10%	57.00%	75.10%
Vllab	12.40%	38.90%	17.50%	57.70%	34.90%	50.70%	44.20%	67.90%

ResNeXt-FPN Take-Away

- Winning entry in COCO stuff 2017 competition

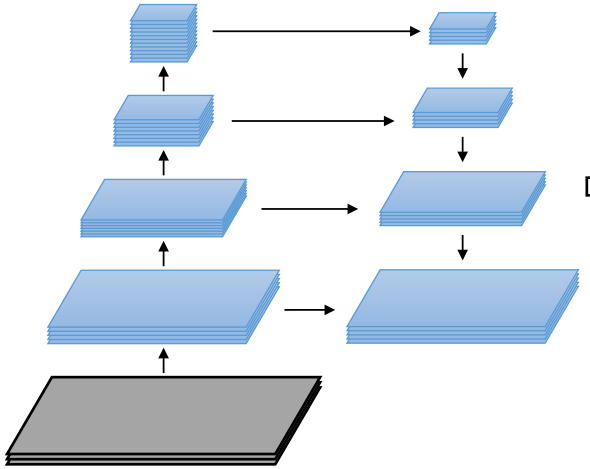
Team name	mIoU	fIoU	mACC	pACC	mIoUS	fIoUS	mACCS	pACCS
ResNeXt152-FPN	**28.80%**	**55.70%**	**42.30%**	**69.20%**	**56.20%**	**68.70%**	**70.30%**	**80.50%**
G-RMI	26.60%	51.90%	40.40%	65.40%	52.40%	64.70%	67.80%	77.40%
Oxford Active Vision Lab	24.20%	50.60%	34.80%	66.00%	50.30%	63.60%	62.20%	77.40%
Baseline Deeplab VGG-16	20.20%	47.60%	28.20%	64.70%	45.90%	60.10%	57.00%	75.10%
Vllab	12.40%	38.90%	17.50%	57.70%	34.90%	50.70%	44.20%	67.90%

- FPN backbone has good [accuracy]/[memory and speed] trade-off

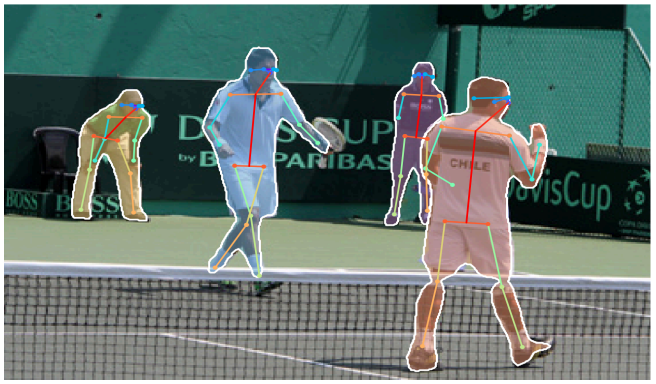


ResNeXt-FPN Take-Away

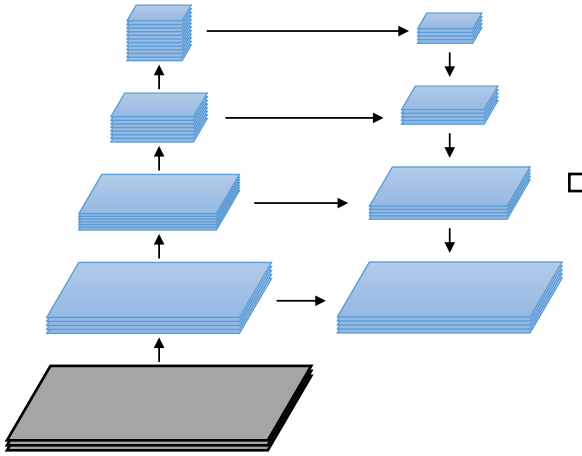
- FPN – unified backbone architecture for object recognition



FPN



instance segmentation
and key points



FPN

semantic segmentation