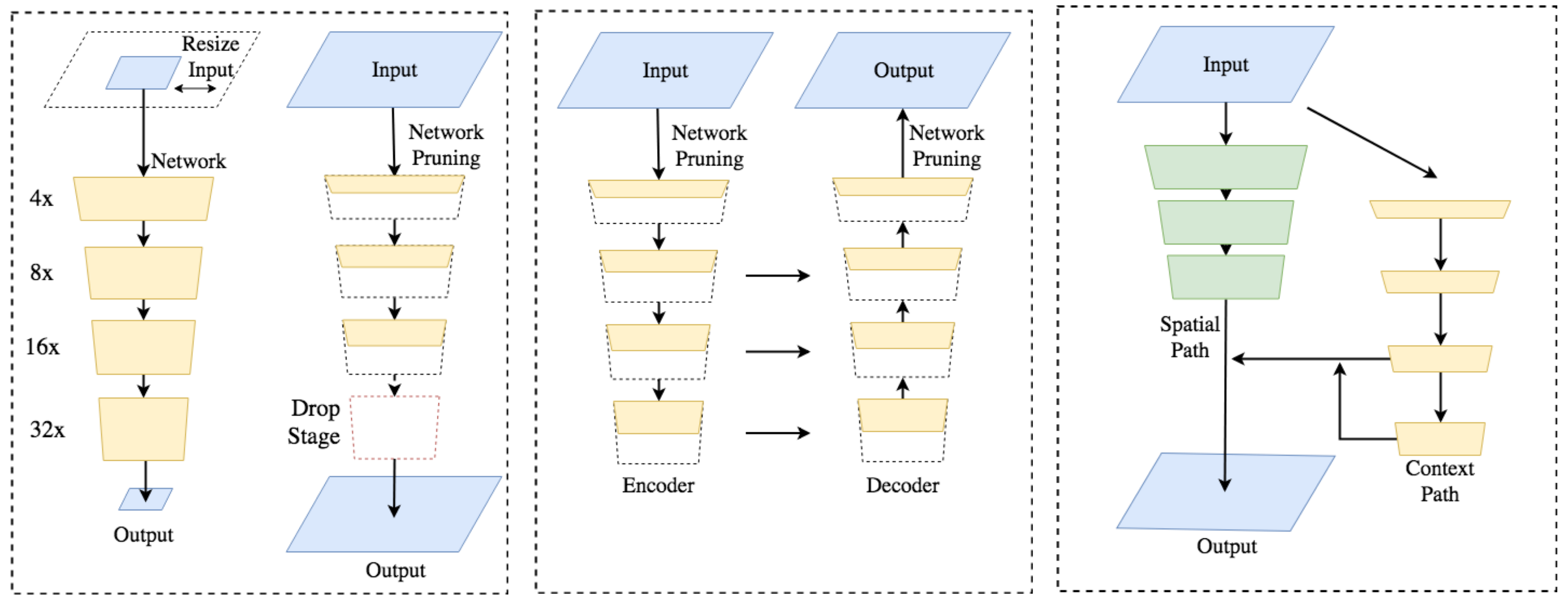
BiSeNet: Bilateral Segmentation Network for Real-time Semantic Segmentation

Changqian Yu*[1], Jingbo Wang*[2], Chao Peng[3], Changxin Gao[1], Gang Yu[3], and Nong Sang[1]

[1]Huazhong University of Science and Technology [2]Peking University [3]Megvii Inc. (Face++)

ECCV 2018
European Conference on Computer Vision

Motivation



- Restrict the input size
- Prune the model
- Drop the last stage

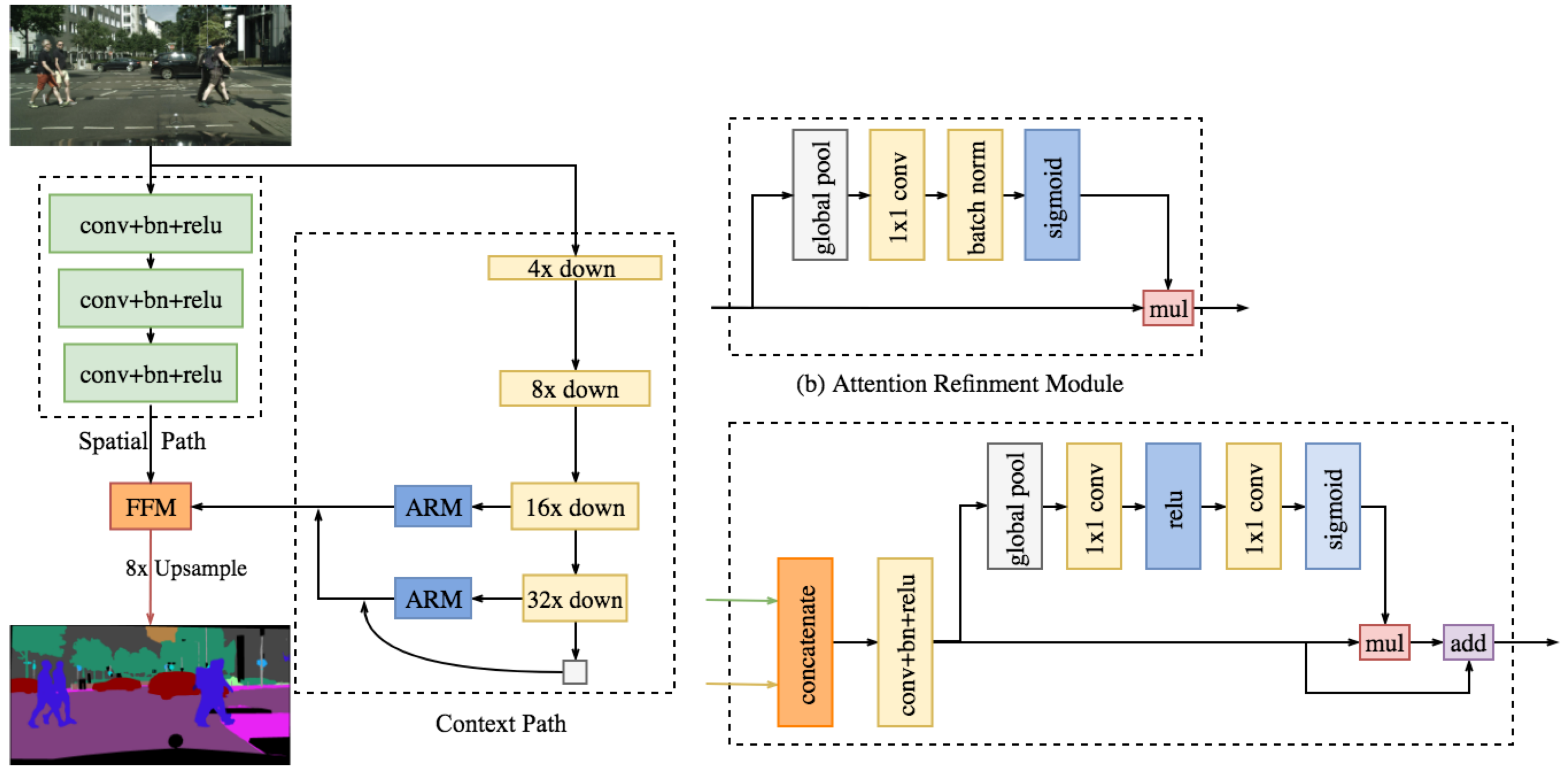
} Loss of spatial details

- Relief: U-shape structure → reduce speed + not completely recovered

Contributions

- We propose a novel approach to decouple the function of spatial information preservation and receptive field offering into two paths. Specifically, we propose a Bilateral Segmentation Network (BiSeNet) with a Spatial Path (SP) and a Context Path (CP).
- We design two specific modules, Feature Fusion Module (FFM) and Attention Refinement Module (ARM), to further improve the accuracy with acceptable cost.
- We achieve impressive results on the benchmarks of Cityscapes, CamVid, and COCO-Stuff. More specifically, we obtain the results of 68.4% on the Cityscapes test dataset with the speed of 105 FPS.

Bilateral Segmentation Network



(a) Network Architecture

(b) Attention Refinement Module

(c) Feature Fusion Module

Experimental Results

Ablation Study

Method	Mean IOU(%)
CP	66.01
CP+SP(Sum)	66.82
CP+SP(FFM)	67.42
CP+SP(FFM)+GP	68.42
CP+SP(FFM)+ARM	68.72
CP+SP(FFM)+GP+ARM	71.40

FLOPS & Parameters

Method	BaseModel	GFLOPS	Parameters
SegNet [1]	VGG16 [29]	286.0	29.5M
ENet [25]	From scratch	3.8	0.4M
Ours	Xception39	2.9	5.8M
Ours	Res18	10.8	49.0M

Speed Comparison on different benchmarks

Method	NVIDIA Titan X						NVIDIA Titan XP					
	640×360		1280×720		1920×1080		640×360		1280×720		1920×1080	
	ms	fps	ms	fps	ms	fps	ms	fps	ms	fps	ms	fps
SegNet [1]	69	14.6	289	3.5	637	1.6	-	-	-	-	-	-
ENet [25]	7	135.4	21	46.8	46	21.6	-	-	-	-	-	-
Ours[1]	**5**	**203.5**	**12**	**82.3**	**24**	**41.4**	**4**	**285.2**	**8**	**124.1**	**18**	**57.3**
Ours[2]	8	129.4	21	47.9	43	23	5	205.7	13	78.8	29	34.4

Speed Comparison on Cityscapes

Method	BaseModel	Mean IOU(%) val	Mean IOU(%) test	FPS
SegNet [1]	VGG16	-	56.1	-
ENet [25]	From scratch	-	58.3	-
SQ [30]	SqueezeNet [14]	-	59.8	-
ICNet [39]	PSPNet50 [40]	67.7	69.5	30.3
DLC [17]	Inception-ResNet-v2	-	71.1	-
Two-column Net [34]	Res50	74.6	72.9	14.7
Ours	Xception39	69.0	68.4	**105.8**
Ours	Res18	**74.8**	**74.7**	65.5

Accuracy Comparison on Cityscapes

Method	BaseModel	Mean IOU(%) val	Mean IOU(%) test
DeepLab [4]	VGG16 [29]	-	63.1
FCN-8s [22]	VGG16	-	65.3
Adelaide [19]	VGG16	-	66.4
Dilation10 [37]	VGG16	68.7	67.1
LRR [10]	VGG16	70.0	69.7
DeepLab-v2+CRF [5]	Res101	71.4	70.4
RefineNet [18]	Res101	-	73.6
DUC [32]	Res152	76.7	76.1
PSPNet [40]	Res101	-	78.4
Ours	Xception39	72.0	71.4
Ours	Res18	78.6	77.7
Ours	Res101	**80.3**	**78.9**

Results on COCO-Stuff (164k/91class)

Method	BaseModel	Mean IOU(%)	Pixel Accuracy(%)
Deeplab-v2	VGG-16	24.0	58.2
Ours	Xception39	22.8	59.0
Ours	Res18	28.1	63.2
Ours	Res101	**31.3**	**65.5**