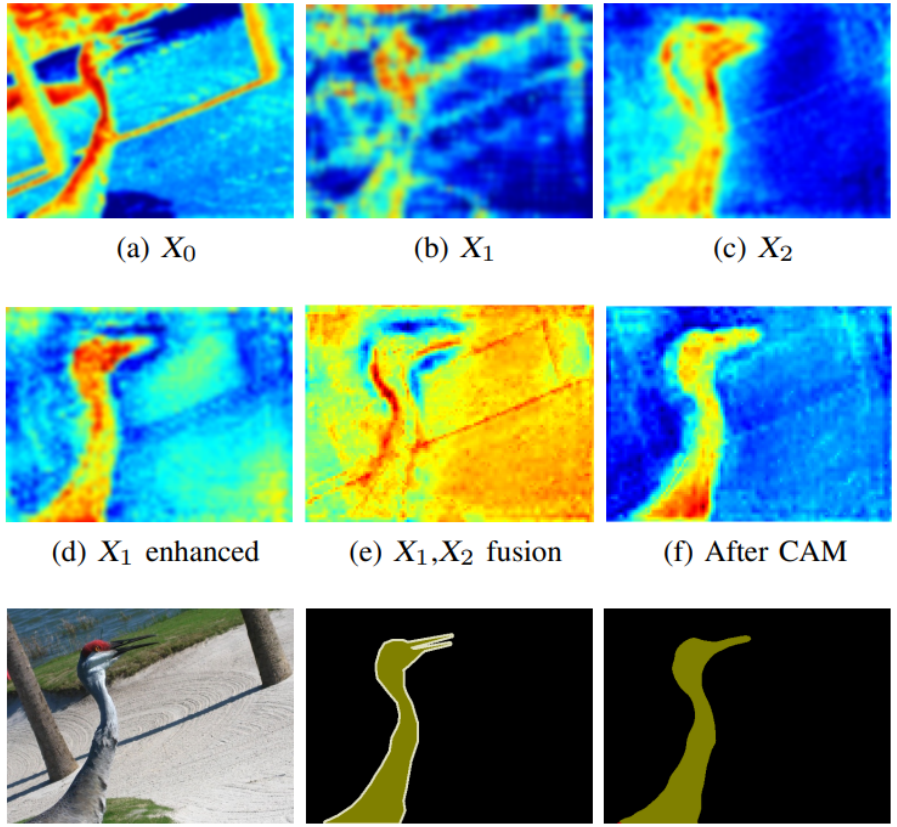
Enhanced Feature Pyramid Network for Semantic Segmentation

Mucong Ye, Jingpeng Ouyang, Ge Chen, Jing Zhang*, Xiaogang Yu
College of Software, Beihang University, Beijing, China
Email: {mucongye, zhang_jing}@buaa.edu.cn

Motivation

Skip connection is frequently used to aggregate multi—scale context information. The features from shallow layers contain low—level appearance information (e.g., edges, lines, and corners), while the features from deep layers contain more semantic information to distinguish different classes.
Simply combining the shallow features with the deep features will bring some background "noise", which will affect the robustness of the features and may lead to semantic consistency. Therefore, the direct fusion of the shallow features and deep features is not optimal for the task of semantic segmentation.



(a) X_0 (b) X_1 (c) X_2

(d) X_1 enhanced (e) X_1,X_2 fusion (f) After CAM

(g) Image (h) Ground truth (i) EFPN results

Overall

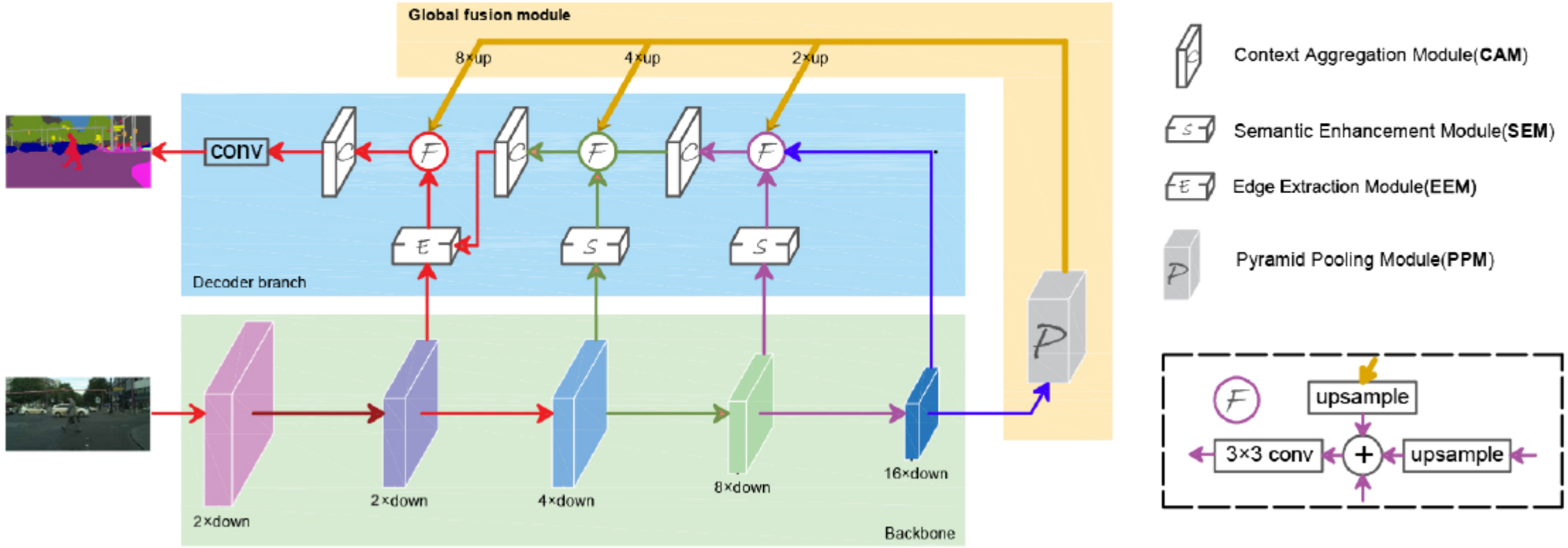

Global fusion module
8×up
4×up
2×up
conv
Decoder branch
2×down
2×down
4×down
8×down
16×down
Backbone
Context Aggregation Module(CAM)
Semantic Enhancement Module(SEM)
Edge Extraction Module(EEM)
Pyramid Pooling Module(PPM)
upsample
3×3 conv
upsample

TABLE II

PER-CLASS RESULTS ON THE PASCAL VOC 2012 TESTING SET. THE PROPOSED EFPN OUTPERFORMS MOST STATE-OF-THE-ART METHODS.

Method	air	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	tarin	tv	mIoU(%)
FCN [9]	76.8	34.2	68.9	49.4	60.3	75.3	74.7	77.6	21.4	62.5	46.8	71.8	63.9	76.5	73.9	45.2	72.4	37.4	70.9	55.1	62.2
DeepLabv2 [2]	84.4	54.5	81.5	63.6	65.9	85.1	79.1	83.4	30.7	74.1	59.8	79.0	76.1	83.2	80.8	59.7	82.2	50.4	73.1	63.7	71.6
CRF-RNN [24]	87.5	39.0	79.7	64.2	68.3	87.6	80.8	84.4	30.4	78.2	60.4	80.5	77.8	83.1	80.6	59.5	82.8	47.8	78.3	67.1	72.0
DeconvNet [25]	89.9	39.3	79.7	63.9	68.2	87.4	81.2	86.1	28.5	77.0	62.0	79.0	80.3	83.6	80.2	58.8	83.4	54.3	80.7	65.0	72.5
PSPNet [3]	91.8	71.9	94.7	71.2	75.8	95.2	89.9	95.9	39.3	90.7	71.7	90.5	**94.5**	88.8	89.6	**72.8**	89.6	64.0	85.1	**76.3**	82.6
AAF [26]	91.3	**72.9**	90.7	68.2	77.7	95.6	**90.7**	**94.7**	**40.9**	89.5	72.6	**91.6**	94.1	88.3	88.8	67.3	**92.9**	62.6	85.2	74.0	82.2
EncNet [4]	94.1	69.2	**96.3**	76.7	**86.2**	**96.3**	**90.7**	94.2	38.8	90.7	**73.3**	90.0	92.5	88.8	**97.9**	68.7	92.6	59.0	**86.4**	73.4	82.9
EFPN(Ours)	**94.9**	68.9	94.9	**79.8**	85.3	95.5	90.6	93.9	39.1	**91.4**	73.0	91.2	93.7	**90.9**	96.1	72.6	92.0	**65.9**	86.2	76.0	**83.6**
With COCO Pre-training																					
Dilation8 [11]	91.7	39.6	87.8	63.1	71.8	89.7	82.9	89.8	37.2	84.0	63.0	83.3	89.0	83.8	85.1	56.8	87.6	56.0	80.2	64.7	75.3
DeepLabv2 [2]	92.9	60.4	91.6	63.4	76.3	95.0	88.4	92.6	32.7	88.5	67.6	89.6	92.1	87.0	87.4	63.3	88.3	60.0	86.8	74.5	79.7
RefineNet [27]	95.0	73.2	93.5	78.1	84.8	95.6	89.8	94.1	43.7	92.0	77.2	90.8	93.4	88.6	88.1	70.1	92.9	64.3	87.7	78.8	84.2
ResNet38 [28]	96.2	75.2	95.4	74.4	81.7	93.7	89.9	92.5	48.2	92.0	79.9	90.1	95.5	91.8	91.2	73.0	90.5	65.4	88.7	80.6	84.9
PSPNet [3]	95.8	72.7	**95.0**	78.9	84.4	94.7	92.0	95.7	43.1	91.0	**80.3**	91.3	96.3	**92.3**	90.1	71.5	94.4	66.9	88.8	**82.0**	85.4
DeepLabv3 [13]	96.4	76.6	92.7	77.8	**87.6**	96.7	**90.2**	95.4	47.5	93.4	76.3	91.4	97.2	91.0	92.1	71.3	90.9	**68.9**	**90.8**	79.3	85.7
EncNet [4]	95.3	76.9	94.2	**80.2**	85.2	96.5	90.8	96.3	47.9	93.9	80.0	92.4	96.6	90.5	91.5	70.8	93.6	66.5	87.7	80.8	85.9
EFPN(Ours)	**97.3**	**81.0**	94.9	79.5	87.2	**96.8**	**92.2**	**96.9**	**48.6**	**94.1**	79.7	**93.3**	**97.6**	90.9	**92.4**	**72.8**	**93.9**	67.8	90.1	81.3	**86.4**

TABLE V

PER-CLASS RESULTS ON THE CITYSCAPES TESTING SET. THE PROPOSED EFPN OBTAINS AN MIOU OF 82.3% WITH TRAINING ON FINE-LABELED DATA.

Methods	road	side	building	wall	fence	pole	T-ligth	T-sign	vege	terrain	sky	person	rider	car	truck	bus	train	motor	bicycle	mIoU(%)
RefineNet [27]	98.2	83.3	91.3	47.8	50.4	56.1	66.9	71.3	92.3	70.3	94.8	80.9	63.3	94.5	64.6	76.1	64.3	62.2	70.0	73.6
DUC-HDC [12]	98.5	85.5	92.8	58.6	55.5	65.0	73.5	77.9	93.3	72.0	95.2	84.8	68.5	95.4	70.9	78.8	68.7	65.9	73.8	77.6
ResNe38 [28]	98.5	85.7	93.1	55.5	59.1	67.1	74.8	78.7	93.7	72.6	95.5	86.6	69.2	95.7	64.5	78.8	74.1	69.0	76.7	78.4
DepthSeg [29]	98.5	85.4	92.5	54.4	60.9	60.2	72.3	76.8	93.1	71.6	94.8	85.2	68.9	95.7	70.1	86.5	75.5	68.3	75.5	78.2
AAF [26]	98.5	85.6	93.0	53.8	58.9	65.9	75.0	78.4	93.7	72.4	95.6	86.4	70.5	95.9	73.9	82.7	76.9	68.7	76.4	79.1
DenseASPP [30]	98.7	87.1	93.4	60.7	62.7	65.6	74.6	78.5	93.6	72.5	95.4	86.2	71.9	96.0	**78.0**	90.3	80.7	69.7	76.8	80.6
DANet [5]	98.6	86.1	93.5	56.1	63.3	69.7	77.3	81.3	93.9	72.9	95.7	87.3	72.9	96.2	76.8	89.4	86.5	72.2	78.2	81.5
EFPN(Ours)	**98.7**	**87.1**	**93.9**	**61.5**	**63.4**	**70.8**	**78.2**	**82.0**	**94.0**	**73.5**	**95.9**	**87.5**	**72.7**	**96.4**	75.8	**93.2**	**88.2**	**73.4**	**78.3**	**82.3**