DPF: Learning Dense Prediction Fields with Weak Supervision
Supplementary Material

Xiaoxue Chen[1], Yuhang Zheng[2], Yupeng Zheng[3]
Qiang Zhou[1], Hao Zhao[1], Guyue Zhou[1], Ya-Qin Zhang[1]
[1]AIR, Tsinghua University [2]BUAA [3]CASIA
{chenxiaoxue, zhaohao}@air.tsinghua.edu.cn, zyh_021@buaa.edu.cn

In this supplementary material, we provide more experiment results to show the effectiveness of DPFs. In the following sections, we first provide our training details, then we present per-category evaluation results on scene parsing. Afterward, we provide more visualization results for both scene parsing and intrinsic decomposition. Finally, we give more analysis experiments to investigate DPFs.

1. Training Details

We use the "poly" learning rate policy in which the current learning rate equals the base one multiplying $(1 - \frac{epoch_{\text{current}}}{epoch_{\max}})^{0.9}$. On PascalContext, we set the base learning rate to 0.028, while on ADE20k, we set the base learning rate to 0.035. For the IIW dataset, the base learning rate is set to 0.007. We set power to 0.9 on all datasets. Momentum and weight decay are set to 0.9 and 0.0001 respectively. The number of training epochs is set to 70 for PascalContext, 60 for ADE20k, and 30 for IIW dataset. For data augmentation, we exploit random rotation between -10 and 10 degrees and a random scale between 0.5 and 2 for all datasets. In addition, we add a random crop with the size of 512×512 and a random horizontal flip. The comprehensive data augmentation scheme makes the network resist overfitting. Last, we normalize the RGB channels of images from a 0∼255 range to a -1∼1 range which can speed up the convergence of the model and improve the performance.

For PascalContext and ADE20k, we use *DistributedDataParallel* and train DPFs with four GPUs with batch size 2 per GPU. For IIW, we train the model on a single GPU while setting the batch size to 2.

2. Per-Category Evaluation

Fig. 1, 2 show per-category semantic IoU comparison between DPF and baseline on PASCALContext [3] and ADE20k respectively [4]. Tab. 5, 4 provide specific IoU. It is manifested that our DPFs achieve better performance in most categories. Notably, the baseline results are 0 in some categories, (e.g., mouse, truck, and bus on PascalContext) while DPFs achieve significant growth of IoU in these categories. We believe this is credited to the guidance feature, which provides detailed semantic information and benefits the fine-grained segmentation.

Weight	PASCAL	ADE20k	IIW
Distance	44.2	32.1	12.9
Learned	**45.3**	**33.8**	**11.9**

Table 1. Quantitative results on the options of weights.

3. More Qualitative Results

Scene parsing. Fig.3, 4 shows some qualitative results for scene parsing on PascalContext and ADE20k datasets respectively. Compared with the baseline prediction, DPF produces more precise results. Specifically, the predictions of DPF have fewer noise patches, and the segmented shapes of objects are also more reasonable and clearer, which is credited to the smoothness of DPFs.

Intrinsic decomposition. Fig. 5 presents some representative qualitative results on IIW. As shown in the area marked by the red box in the figure, DPF can distinguish the shadows of objects and successfully decompose the wall or ground into consistent reflectance. Also, it is manifested from the yellow box that our model can also predict the clothes with wrinkles as the same reflectance. These results illustrate the capability of DPF on intrinsic decomposition.

4. More Ablation Study

Interpolation weights. We conduct an experiment to investigate the importance of using an MLP to predict the interpolation weights. We use bilinear interpolation with relative distances as weights, and the results are provided in Tab. 1. Obviously, the results of using the predicted weights are better, because the predicted weights not only leverage the spatial affinity between the query point and its neighbor

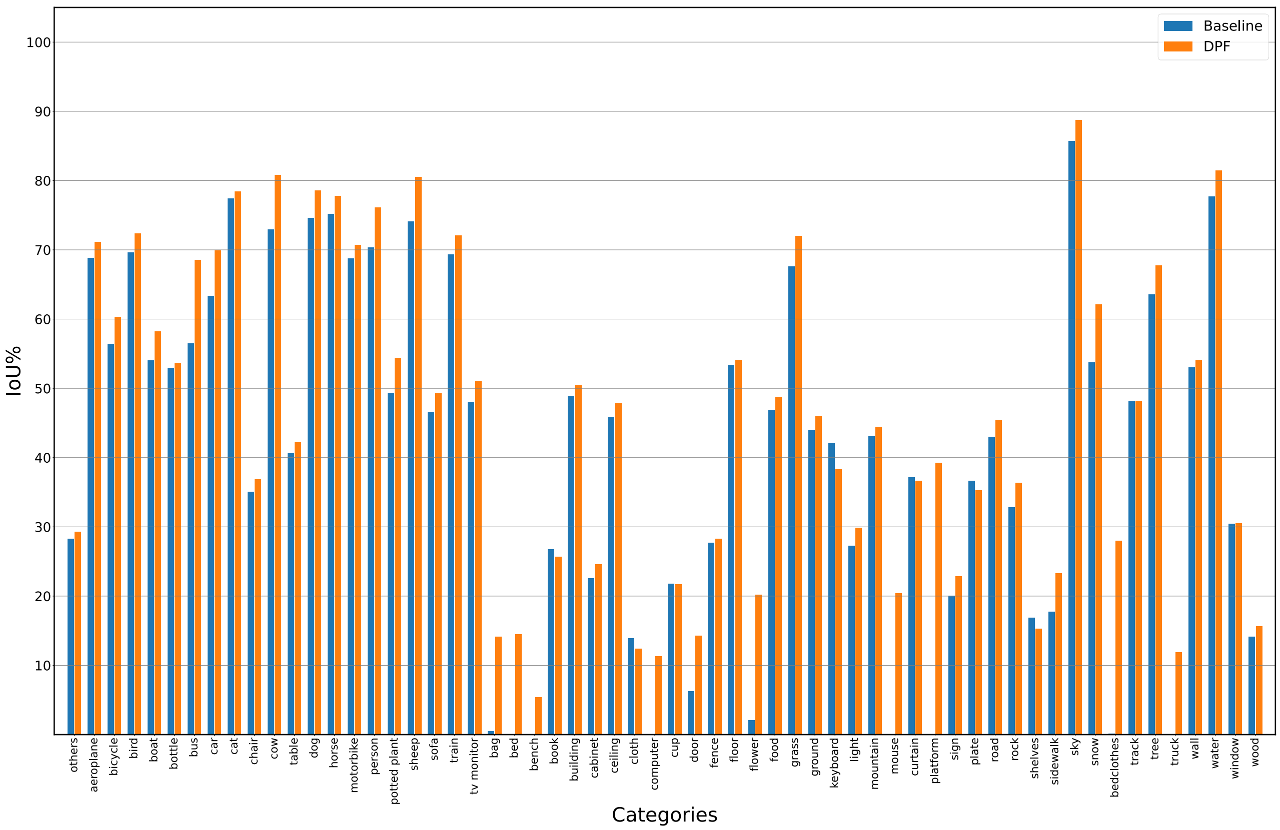

Figure 1. Per-category evaluation on PASCALContext dataset.

Method	PASCAL	ADE20k
DPT	40.4	30.4
DPT+denseCRF	41.3	31.1
DPT+DPF	**45.3**	**33.8**

Table 2. Quantitative results on the effects of DPF.

pixels but also consider the semantic correlation provided by the latent code.

Comparsions with CRF method. CRFs are conventional techniques that smooth the dense prediction results on top of dense prediction baselines using low-level features, which work in a similar spirit to DPF. To compare with CRF, we additionally post-process the outputs of DPT baselines using denseCRF [2], and the results are shown in Tab. 2. Since the prediction of intrinsic decomposition is a continuous value between 0-1, and the denseCRF method requires a discrete label set, we only provide the results on ADE20k and PASCAL. As shown in the table, DPF outperforms the denseCRF method with clear margins, demonstrating its superiority with CRF.

5. Analysis Experiment

DPF is effective in the intrinsic decomposition of different materials. Tab. 6 presents the reflectance precision on different materials. We use a pre-trained model of [1] to pre-process the IIW dataset and get attribute maps for each test image. For a pair of comparison points, we can easily get their attribute from the coordinates on the attribute maps. If the predicted relative reflectance of the comparison pair is correct, we consider the reflectance predictions of the two points to be correct. With the above preliminaries, we calculate the per-attribute precision of both baseline and DPF. The results show that DPF has superior performance in different attributes compared with baseline. Moreover, DPF gets higher precision on attributes such as wall and painted, while glass and metal have lower precision, which demonstrates that the model works better on diffuse materials, while intrinsic decomposition with specular reflection is relatively difficult.

6. Transductive Learning

DPF achieves superior performance on both inductive learning and transductive learning. Inductive learn-

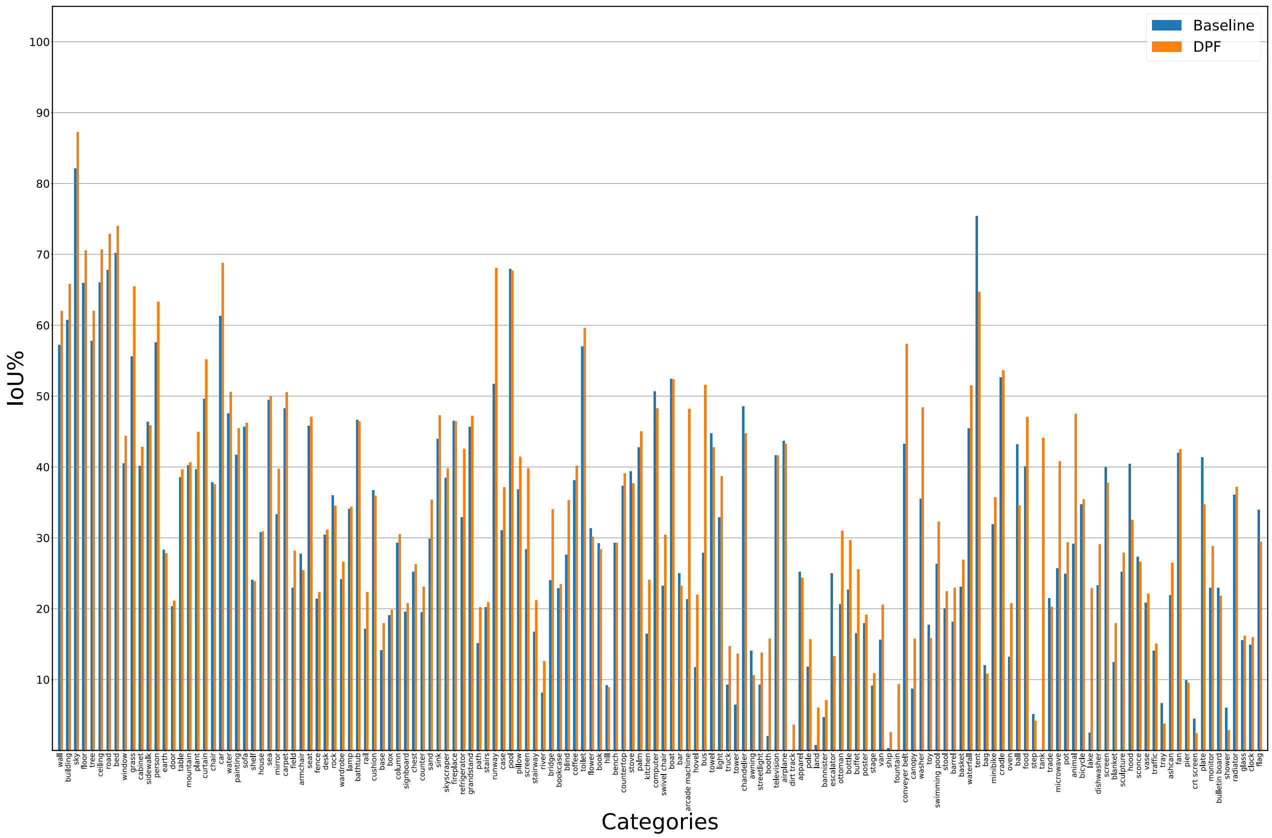

Figure 2. Per-category evaluation on ADE20k dataset.

ing, which represents the generalization of models, means learning from one sample and testing on another unseen sample. While transductive learning refers to testing on the unlabeled sample where we have obtained features during training. Commonly, weakly supervised learning may refer to both inductive learning and transductive learning because of the number of unlabeled pixels in the image. In Table. 3 , it is obvious that our DPF is also effective in transductive learning on both datasets. Especially on PascalContext, DPF has a strong lead of 14.7%. This demonstrates that our proposed DPF is a general framework that is not specific to the inductive or transductive solution.

Dataset	Method	mIoU
PASCAL	Baseline(%)	46.5
	w/DPF (%)	**61.2** (+14.7)
ADE20k	Baseline(%)	38.0
	w/DPF (%)	**42.4** (+4.4)

Table 3. Transductive learning on PASCALContext and ADE20k unlabeled training sets, compared with DPT(baseline).

References

[1] Xiaoxue Chen, Tianyu Liu, Hao Zhao, Guyue Zhou, and Ya-Qin Zhang. Cerberus transformer: Joint semantic, affordance and attribute parsing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19649–19658, 2022. 2

[2] Philipp Krähenbühl and Vladlen Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. *Advances in neural information processing systems*, 24, 2011. 2

[3] Roozbeh Mottaghi, Xianjie Chen, Xiaobai Liu, Nam-Gyu Cho, Seong-Whan Lee, Sanja Fidler, Raquel Urtasun, and Alan Yuille. The role of context for object detection and semantic segmentation in the wild. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014. 1

[4] Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Scene parsing through ade20k dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 633–641, 2017. 1

Method	wall	building	sky	floor	tree	ceiling	road	bed	window	grass
Baseline(%)	57.2	60.8	82.1	66.0	57.8	66.0	67.8	70.2	40.5	55.6
DPF(%)	**62.0**	**65.8**	**87.2**	**70.5**	**62.0**	**70.7**	**72.9**	**74.0**	**44.4**	**65.5**
Method	cabinet	sidewalk	person	earth	door	table	mountain	plant	curtain	chair
Baseline(%)	40.2	**46.4**	57.6	**28.3**	20.3	38.6	40.2	39.7	49.6	**37.8**
DPF(%)	**42.9**	45.9	**63.3**	27.8	**21.1**	**39.7**	**40.7**	**44.9**	**55.2**	37.5
Method	car	water	painting	sofa	shelf	house	sea	mirror	carpet	field
Baseline(%)	61.3	47.6	41.7	45.6	**24.1**	30.8	49.4	33.3	48.2	23.0
DPF(%)	**68.8**	**50.6**	**45.4**	**46.2**	23.9	**30.9**	**50.0**	**39.8**	**50.5**	**28.1**
Method	armchair	seat	fence	desk	rock	wardrobe	lamp	bathtub	rail	cushion
Baseline(%)	**27.7**	45.8	21.4	30.4	**36.0**	24.2	34.1	**45.6**	17.2	**36.7**
DPF(%)	25.4	**47.1**	**22.3**	**31.2**	34.5	**26.6**	**34.4**	46.5	**22.3**	35.9
Method	base	box	column	signboard	chest	counter	sand	sink	skyscraper	fireplace
Baseline(%)	14.1	19.1	29.3	19.6	25.2	19.5	29.9	43.9	38.4	**46.5**
DPF(%)	**17.9**	**19.8**	**30.5**	**20.8**	**26.3**	**23.1**	**35.4**	**47.3**	**39.8**	46.4
Method	refrigerator	grandstand	path	stairs	runway	case	pool	pillow	screen	stairway
Baseline(%)	32.9	45.7	15.2	20.2	51.7	31.0	**68.0**	36.9	28.4	16.7
DPF(%)	**42.6**	**47.2**	**20.2**	**20.9**	**68.1**	**37.1**	67.7	**41.4**	**39.8**	**21.2**
Method	river	bridge	bookcase	blind	coffee	toilet	flower	book	hill	bench
Baseline(%)	8.2	24.0	22.9	27.6	38.1	57.0	**31.3**	**29.2**	**9.2**	29.3
DPF(%)	**12.6**	**34.0**	**23.4**	**35.3**	**40.1**	**59.6**	30.1	28.4	8.9	**29.3**
Method	countertop	stove	palm	kitchen	computer	swivel chair	boat	bar	arcade machine	hovel
Baseline(%)	37.3	**39.4**	42.8	16.5	**50.6**	23.2	**52.5**	**25.0**	21.3	11.8
DPF(%)	**39.1**	37.7	**45.0**	**24.1**	48.2	**30.4**	52.4	23.2	**48.2**	**21.9**
Method	bus	towel	light	truck	tower	chandelier	awning	streetlight	booth	television
Baseline(%)	27.9	**44.8**	32.9	9.3	6.4	**48.5**	**14.1**	9.3	2.0	41.6
DPF(%)	**51.6**	42.7	**38.7**	**14.7**	**13.6**	44.7	10.6	**13.8**	**15.7**	**41.6**
Method	airplane	dirt track	apparel	pole	land	bannister	escalator	ottoman	bottle	buffet
Baseline(%)	**43.7**	0.0	**25.2**	11.8	0.7	4.7	**25.0**	20.6	22.6	16.5
DPF(%)	43.3	**3.7**	24.3	**15.7**	**6.0**	**7.1**	13.3	**31.0**	**29.6**	**25.6**
Method	poster	stage	van	ship	fountain	conveyer belt	canopy	washer	toy	swimming pool
Baseline(%)	17.9	9.1	15.6	0.3	0.0	43.2	8.7	35.5	**17.7**	26.3
DPF(%)	**19.1**	**10.9**	**20.6**	**2.5**	**9.3**	**57.4**	**15.8**	**48.4**	15.8	**32.3**
Method	stool	barrel	basket	waterfall	tent	bag	minibike	cradle	oven	ball
Baseline(%)	20.1	18.1	23.1	45.4	**75.4**	**12.0**	31.9	52.7	13.2	**43.2**
DPF(%)	**22.5**	**23.0**	**26.9**	**51.5**	64.7	10.8	**35.7**	**53.6**	**20.8**	34.6
Method	food	step	tank	trade	microwave	pot	animal	bicycle	lake	dishwasher
Baseline(%)	40.1	**5.1**	0.0	**21.5**	25.7	24.9	29.2	34.8	2.5	23.3
DPF(%)	**47.1**	4.2	**44.1**	20.2	**40.8**	**29.4**	**47.5**	**35.4**	**22.9**	**29.1**
Method	screen	blanket	sculpture	hood	sconce	vase	traffic	tray	ashcan	fan
Baseline(%)	**40.1**	12.5	25.2	**40.5**	**27.3**	20.8	14.1	**6.7**	21.9	42.0
DPF(%)	37.7	**17.9**	**27.9**	32.6	26.6	**22.1**	**15.0**	3.8	**26.5**	**42.5**
Method	pier	crt screen	plate	monitor	bulletin board	shower	radiator	glass	clock	flag
Baseline(%)	**9.9**	**4.5**	**41.4**	22.9	**23.0**	**6.0**	36.1	15.5	14.9	**33.9**
DPF(%)	9.5	2.4	34.7	**28.8**	21.8	2.9	**37.2**	**16.2**	**16.0**	29.4

Table 4. Per-category semantic parsing results on ADE20k.

Method	others	aeroplane	bicycle	bird	boat	bottle	bus	car	cut	chair
Baseline(%)	28.3	68.8	56.4	69.6	54.0	52.9	56.5	63.3	77.4	35.1
DPF(%)	**29.3**	**71.1**	**60.3**	**72.4**	**58.2**	**53.7**	**68.5**	**69.9**	**78.4**	**36.8**
Method	cow	table	dog	horse	motorbike	person	potted plant	sheep	sofa	train
Baseline(%)	72.9	40.6	74.6	75.2	68.7	70.3	49.3	74.1	46.5	69.3
DPF(%)	**80.8**	**42.2**	**78.5**	**77.8**	**70.7**	**76.1**	**54.4**	**80.5**	**49.2**	**72.1**
Method	TV monitor	bag	bed	bench	book	building	cabinet	ceiling	cloth	computer
Baseline(%)	48.1	0.4	0.0	0.0	**26.7**	48.9	22.6	45.8	**13.9**	0.0
DPF(%)	**51.1**	**14.1**	**14.5**	**5.4**	25.7	**50.4**	**24.6**	**47.8**	12.4	**11.3**
Method	cup	door	fence	floor	flower	food	grass	ground	keyboard	light
Baseline(%)	**21.8**	6.2	27.7	53.3	2.1	46.9	67.6	43.9	**42.0**	27.3
DPF(%)	21.7	**14.3**	**28.3**	**54.1**	**20.1**	**48.7**	**72.0**	**45.9**	38.3	**29.8**
Method	mountain	mouse	curtain	platform	sign	plate	road	rock	shelves	sidewalk
Baseline(%)	43.1	0.0	**37.1**	0.0	20.0	**36.6**	43.0	32.8	**16.8**	17.7
DPF(%)	**44.4**	**20.4**	36.6	**39.2**	**22.8**	35.3	**45.4**	**36.3**	15.3	**23.3**
Method	sky	snow	bedclothes	track	tree	truck	wall	water	window	wood
Baseline(%)	85.7	53.7	0.1	48.1	63.6	0.0	53.0	77.7	30.4	14.1
DPF(%)	**88.7**	**62.1**	**27.9**	**48.2**	**67.7**	**11.9**	**54.1**	**81.4**	**30.5**	**15.7**

Table 5. Per-category semantic parsing results on PascalContext.

Method	**Wood**	**Painted**	**Paper**	**Glass**	**Brick**	**Metal**	**Flat**	**Plastic**	**Textured**	**Glossy**	**Shiny**
Baseline(%)	71.0	82.3	30.6	29.5	79.7	7.4	3.1	18.5	57.6	21.8	55.1
w/DPF (%)	72.5(+1.5)	84.0(+1.7)	31.4(+0.8)	30.5(+1.0)	81.3(+0.6)	7.9(+0.5)	3.1(+0)	18.9(+0.4)	58.4(+0.8)	22.2(+0.4)	56.3(+1.2)

Table 6. Precision of reflectance prediction on different attributes.

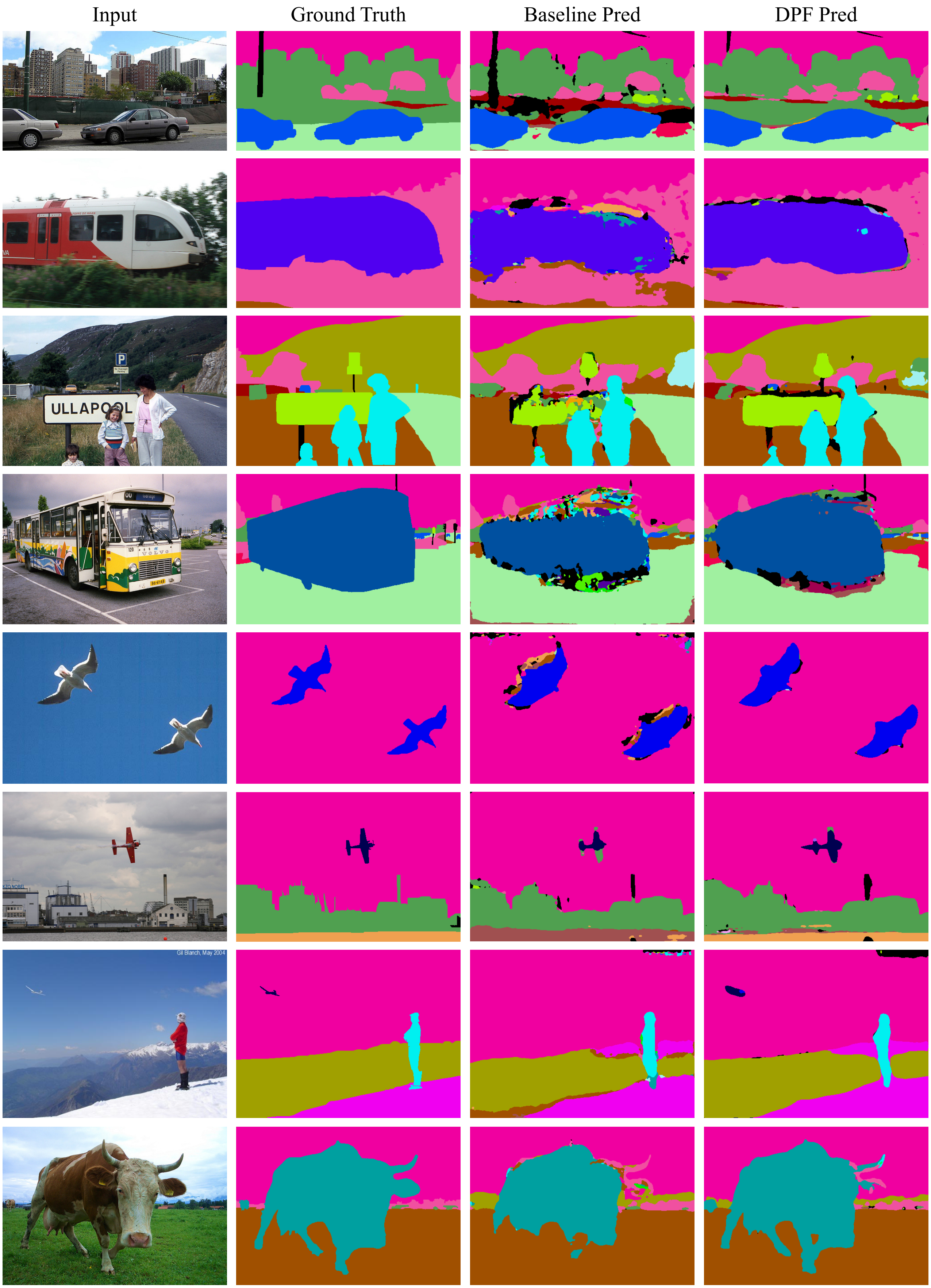

Figure 3. More qualitative results on PascalContext.

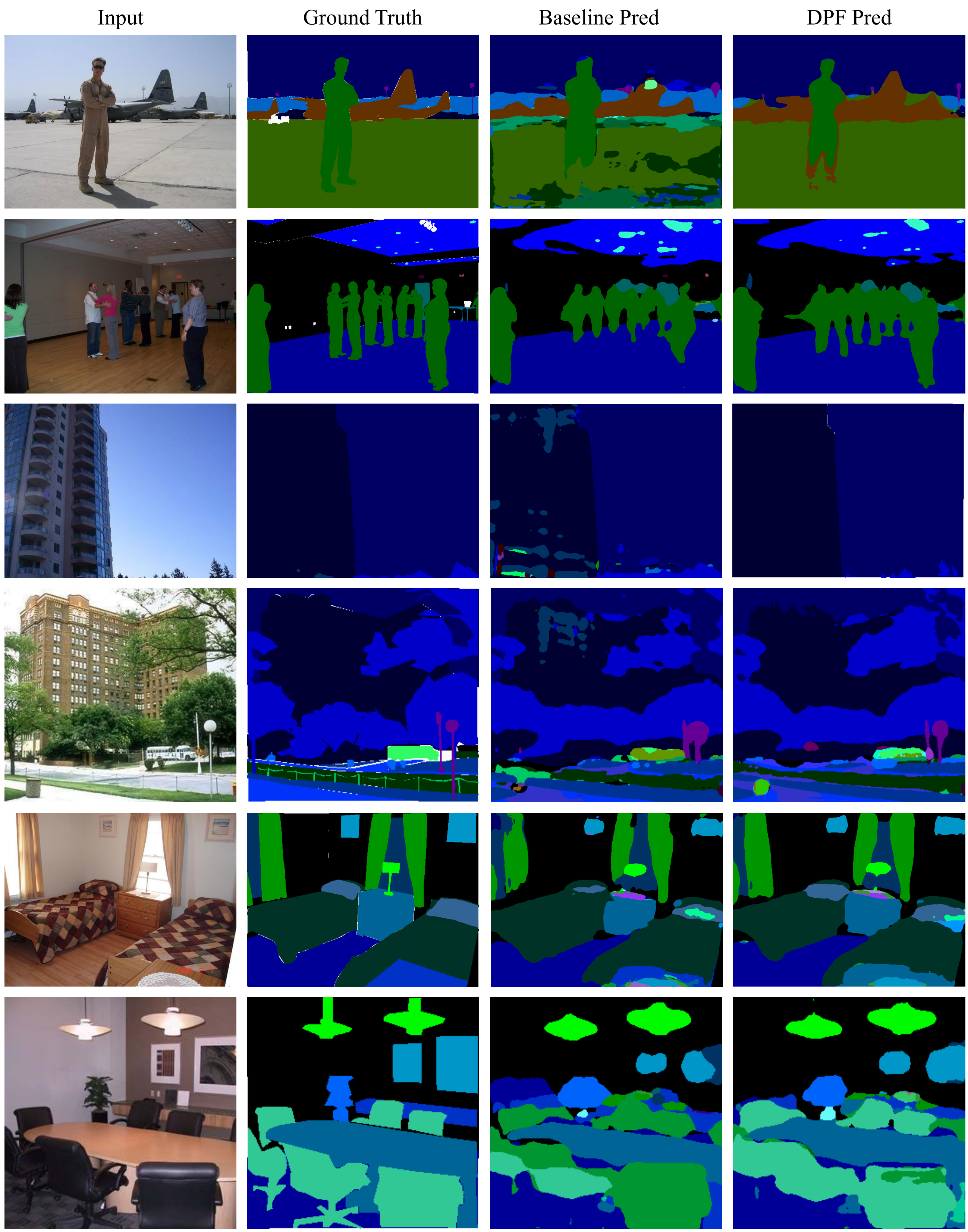

Figure 4. More qualitative results on ADE20k. White stands for mask.

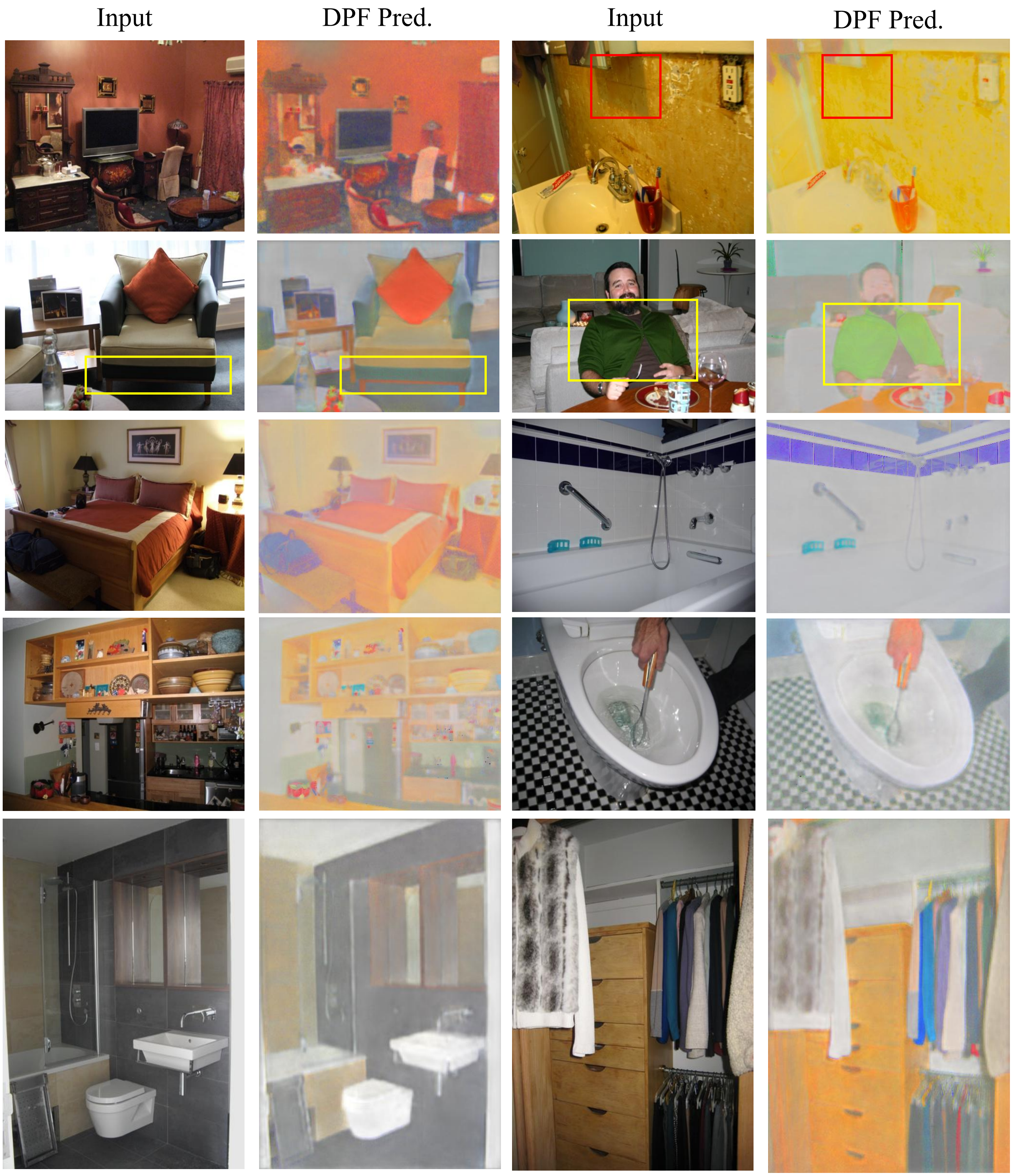

Figure 5. More qualitative results on IIW.