Supplementary Document:
Generalizing Interactive Backpropagating Refinement for Dense Prediction Networks

1. Optimal learning rates for experiments

Learning rate is one of the most important hyperparameter for backpropagation. Therefore, we search for the best learning rate settings for all experiments by using a subset of the test set from each dataset for evaluation using various learning rates. Specifically, we test for 10 learning rates ranging from 0.1×0.5^0 to 0.1×0.5^9. Tables 1-5 shows the obtained best learning rates used for our experiments. The number of layers included in the architecture is denoted as L1, L2 and L3. Architectures for semantic segmentation and image matting can contain up to 3 G-BRS layers, while architectures for interactive segmentation and depth estimation contain 1 G-BRS layer.

2. Quantitative results for all experiments

2.1. Interactive and semantic segmentation

We only report top scores achieved by each G-BRS layer type in the main paper due to the space limit. Tables 6-8 show the AUC as well as maximum mIoU achieved in total number of clicks using all backpropagating refinement settings. We see that the proposed consistency loss $\mathcal{L}_c$ enables very consistent improvement for all G-BRS as well as RGB-BRS settings. The G-BRS-bmconv layer also consistently achieves the best AUC and maximum mIoU for segmentation tasks.

2.2. Image matting

We compute the following standard metrics for the task of image matting: Sum of Absolute Differences (SAD), Mean Squared Error (MSE), Gradient (Grad) and Connectivity (Conn) error. We report results on MSE in the main paper due to the space limit. For simplicity, here we report the metrics obtained for each G-BRS layer type with the layout (#layers) that achieves the best scores. Figure 1 shows that the G-BRS-sb layer that uses the same layer architecture as f-BRS has limited ability for refinement comparing to other G-BRS layers. This is due to G-BRS-sb's lack of ability for localized modification of the features. We also observe that settings that utilize the consistency loss $\mathcal{L}_c$ generally achieves more stable and accurate results.

Interestingly, our results indicate that RGB-BRS achieves best overall quantitative scores for image matting with or without $\mathcal{L}_c$. This finding is helpful for the community as backpropagating refinement has not been implemented on the task of image matting before. However, RGB-BRS does introduce additional inference time and memory consumption. In a later section, we show a detailed report on Seconds Per Click (SPC) for all backpropagating refinement settings so the trade-off between accuracy and efficiency can be decided by the user.

2.3. Depth estimation

We compute the following standard metrics for the task of depth estimation: δ_{1-3}, Abs Rel, Sq Rel, RMSE and RMSElog. We only report results on δ_1 in the main paper due to the space limit. Here we report the AUC computed on each metric over the total number of clicks as well as the best score achieved in the total number of clicks. Table 9 shows that the G-BRS-bmconv layer achieves the best metrics in all settings and $\mathcal{L}_c$ provides consistent improvement.

3. Running time analysis

Inference speed is an important factor for interactive applications. In this section, we provide the Seconds Per Click (SPC) measured for all backpropagating refinement settings using the proposed $\mathcal{L}_c$. We select 100 test instances from each dataset and perform 10 clicks on each instance. A PC with an AMD Ryzen Threadripper 1920X CPU and a RTX 2080 Ti GPU is used in this experiment. Tables 10-14 show the obtained SPC for all backpropagating refinement settings. First, We see that the overall difference between the inference time of different G-BRS layers is very small. In addition, the additional inference time for inserting multiple G-BRS layers is very small, which makes the usage of multiple G-BRS layers more desirable if it can lead to improvement in performance. As expected, RGB-BRS has considerably higher inference time and previous experiments show that it only achieves better performance in the task of image matting. For the task of semantic segmentation, we recognize that the obtained inference time is not ideal for real-time interactive responses. This is due to our selection of a state-of-the-art architecture that prioritizes accuracy over efficiency. Our experiments show that our approach can effectively improve results of top-performing models. As a result, users have the flexibility to apply our approach to other architectures and design the G-BRS configuration to accommodate the specific needs of their applications.

4. Qualitative Comparisons

To compare the performance of backpropagating refinement using our proposed G-BRS layers with the previously proposed channel-wise scale and bias as auxiliary variables by f-BRS, we select the G-BRS-bmconv layer and the G-BRS-sb layer (channel-wise s and b) for qualitative comparison.

For interactive segmentation, we compare with the prior approach f-BRS directly as the G-BRS-sb layer for this task is equivalent to the solution proposed by f-BRS. Figure 2 shows that the G-BRS-bmconv layer can make more detailed refinement. Since the original f-BRS was not implemented for the other applications, we make the best comparison possible by comparing the proposed G-BRS-bmconv layer with the G-BRS-sb layer. We emphasize that the specific G-BRS configurations for various architectures as well as the utilized consistency loss are part of our contribution.

For semantic segmentation, we use 3 G-BRS layers for both G-BRS-sb and G-BRS-bmconv for the best performance as shown in Tables 7-8. Figure 3 shows that both G-BRS-sb and G-BRS-bmconv achieve high accuracy on Cityscapes while the G-BRS-bmconv layer is capable of making more detailed refinement. Figure 4 shows that in a more challenging semantic segmentation dataset (Mapillary Vista) with 65 classes, the G-BRS-bmconv layer outperforms the G-BRS-sb layer with a larger margin. For image matting, we use 3 G-BRS layers as well for the best performance (Figure 1). Figure 5 shows that the G-BRS-bmconv layer is capable of refining alpha matte at a more detailed level. For depth estimation, Figure 6 shows that the G-BRS-sb layer is more susceptible to undesired global error while the G-BRS-bmconv layer produces better depth maps with both global and localized accuracy.

#Layers	w $\mathcal{L}_c$					w/o $\mathcal{L}_c$				
	RGB-BRS	sb	bmsb	bmsb-m	bmconv	RGB-BRS	sb	bmsb	bmsb-m	bmconv
L0	$0.1 \cdot \frac{1}{2}^{9}$	-	-	-	-	$0.1 \cdot \frac{1}{2}^{9}$	-	-	-	-
L1	-	$0.1 \cdot \frac{1}{2}^{1}$	$0.1 \cdot \frac{1}{2}^{4}$	$0.1 \cdot \frac{1}{2}^{2}$	$0.1 \cdot \frac{1}{2}^{9}$	-	$0.1 \cdot \frac{1}{2}^{6}$	$0.1 \cdot \frac{1}{2}^{5}$	$0.1 \cdot \frac{1}{2}^{4}$	$0.1 \cdot \frac{1}{2}^{9}$

Table 1. Learning rate settings for the backpropagating refinement layouts on **SBD**.

#Layers	w $\mathcal{L}_c$					w/o $\mathcal{L}_c$				
	RGB-BRS	sb	bmsb	bmsb-m	bmconv	RGB-BRS	sb	bmsb	bmsb-m	bmconv
L0	$0.1 \cdot \frac{1}{2}^{8}$	-	-	-	-	$0.1 \cdot \frac{1}{2}^{9}$	-	-	-	-
L1	-	$0.1 \cdot \frac{1}{2}^{3}$	$0.1 \cdot \frac{1}{2}^{4}$	$0.1 \cdot \frac{1}{2}^{3}$	$0.1 \cdot \frac{1}{2}^{7}$	-	$0.1 \cdot \frac{1}{2}^{4}$	$0.1 \cdot \frac{1}{2}^{5}$	$0.1 \cdot \frac{1}{2}^{4}$	$0.1 \cdot \frac{1}{2}^{9}$
L2	-	$0.1 \cdot \frac{1}{2}^{4}$	$0.1 \cdot \frac{1}{2}^{4}$	$0.1 \cdot \frac{1}{2}^{2}$	$0.1 \cdot \frac{1}{2}^{7}$	-	$0.1 \cdot \frac{1}{2}^{6}$	$0.1 \cdot \frac{1}{2}^{6}$	$0.1 \cdot \frac{1}{2}^{5}$	$0.1 \cdot \frac{1}{2}^{9}$
L3	-	$0.1 \cdot \frac{1}{2}^{4}$	$0.1 \cdot \frac{1}{2}^{4}$	$0.1 \cdot \frac{1}{2}^{3}$	$0.1 \cdot \frac{1}{2}^{8}$	-	$0.1 \cdot \frac{1}{2}^{6}$	$0.1 \cdot \frac{1}{2}^{7}$	$0.1 \cdot \frac{1}{2}^{5}$	$0.1 \cdot \frac{1}{2}^{9}$

Table 2. Learning rate settings for the backpropagating refinement layouts on **Cityscapes**.

#Layers	w $\mathcal{L}_c$					w/o $\mathcal{L}_c$				
	RGB-BRS	sb	bmsb	bmsb-m	bmconv	RGB-BRS	sb	bmsb	bmsb-m	bmconv
L0	$0.1 \cdot \frac{1}{2}^{4}$	-	-	-	-	$0.1 \cdot \frac{1}{2}^{9}$	-	-	-	-
L1	-	$0.1 \cdot \frac{1}{2}^{2}$	$0.1 \cdot \frac{1}{2}^{3}$	$0.1 \cdot \frac{1}{2}^{3}$	$0.1 \cdot \frac{1}{2}^{7}$	-	$0.1 \cdot \frac{1}{2}^{3}$	$0.1 \cdot \frac{1}{2}^{5}$	$0.1 \cdot \frac{1}{2}^{3}$	$0.1 \cdot \frac{1}{2}^{9}$
L2	-	$0.1 \cdot \frac{1}{2}^{2}$	$0.1 \cdot \frac{1}{2}^{3}$	$0.1 \cdot \frac{1}{2}^{2}$	$0.1 \cdot \frac{1}{2}^{7}$	-	$0.1 \cdot \frac{1}{2}^{4}$	$0.1 \cdot \frac{1}{2}^{6}$	$0.1 \cdot \frac{1}{2}^{5}$	$0.1 \cdot \frac{1}{2}^{9}$
L3	-	$0.1 \cdot \frac{1}{2}^{3}$	$0.1 \cdot \frac{1}{2}^{4}$	$0.1 \cdot \frac{1}{2}^{3}$	$0.1 \cdot \frac{1}{2}^{8}$	-	$0.1 \cdot \frac{1}{2}^{5}$	$0.1 \cdot \frac{1}{2}^{6}$	$0.1 \cdot \frac{1}{2}^{5}$	$0.1 \cdot \frac{1}{2}^{9}$

Table 3. Learning rate settings for the backpropagating refinement layouts on **Mapillary Vista**.

#Layers	w $\mathcal{L}_c$					w/o $\mathcal{L}_c$				
	RGB-BRS	sb	bmsb	bmsb-m	bmconv	RGB-BRS	sb	bmsb	bmsb-m	bmconv
L0	$0.1 \cdot \frac{1}{2}^{9}$	-	-	-	-	$0.1 \cdot \frac{1}{2}^{9}$	-	-	-	-
L1	-	$0.1 \cdot \frac{1}{2}^{1}$	$0.1 \cdot \frac{1}{2}^{4}$	$0.1 \cdot \frac{1}{2}^{4}$	$0.1 \cdot \frac{1}{2}^{7}$	-	$0.1 \cdot \frac{1}{2}^{7}$	$0.1 \cdot \frac{1}{2}^{5}$	$0.1 \cdot \frac{1}{2}^{5}$	$0.1 \cdot \frac{1}{2}^{9}$
L2	-	$0.1 \cdot \frac{1}{2}^{2}$	$0.1 \cdot \frac{1}{2}^{5}$	$0.1 \cdot \frac{1}{2}^{4}$	$0.1 \cdot \frac{1}{2}^{8}$	-	$0.1 \cdot \frac{1}{2}^{8}$	$0.1 \cdot \frac{1}{2}^{6}$	$0.1 \cdot \frac{1}{2}^{6}$	$0.1 \cdot \frac{1}{2}^{8}$
L3	-	$0.1 \cdot \frac{1}{2}^{3}$	$0.1 \cdot \frac{1}{2}^{6}$	$0.1 \cdot \frac{1}{2}^{5}$	$0.1 \cdot \frac{1}{2}^{9}$	-	$0.1 \cdot \frac{1}{2}^{9}$	$0.1 \cdot \frac{1}{2}^{6}$	$0.1 \cdot \frac{1}{2}^{6}$	$0.1 \cdot \frac{1}{2}^{9}$

Table 4. Learning rate settings for the backpropagating refinement layouts on **Composition-1k**.

#Layers	w $\mathcal{L}_c$					w/o $\mathcal{L}_c$				
	RGB-BRS	sb	bmsb	bmsb-m	bmconv	RGB-BRS	sb	bmsb	bmsb-m	bmconv
L0	$0.1 \cdot \frac{1}{2}^{7}$	-	-	-	-	$0.1 \cdot \frac{1}{2}^{8}$	-	-	-	-
L1	-	$0.1 \cdot \frac{1}{2}^{2}$	$0.1 \cdot \frac{1}{2}^{3}$	$0.1 \cdot \frac{1}{2}^{2}$	$0.1 \cdot \frac{1}{2}^{7}$	-	$0.1 \cdot \frac{1}{2}^{2}$	$0.1 \cdot \frac{1}{2}^{3}$	$0.1 \cdot \frac{1}{2}^{2}$	$0.1 \cdot \frac{1}{2}^{7}$

Table 5. Learning rate settings for the backpropagating refinement layouts on **NYU-Depth-V2**.

#Layers	w $\mathcal{L}_c$					w/o $\mathcal{L}_c$				
	RGB-BRS	sb	bmsb	bmsb-m	bmconv	RGB-BRS	sb	bmsb	bmsb-m	bmconv
L0 (AUC)	0.8506	-	-	-	-	0.8251	-	-	-	-
L1 (AUC)	-	0.8521	0.8591	0.8589	0.8594	-	0.8427	0.8529	0.8457	0.8322
L0 (mIoU*)	0.9051	-	-	-	-	0.8605	-	-	-	-
L1 (mIoU*)	-	0.9083	0.9162	0.9169	0.9181	-	0.8931	0.9117	0.9052	0.8867

Table 6. Area Under Curve (AUC) computed using mIoU and maximum mIoU achieved in total number of clicks (mIoU*) on **SBD** for all backpropagating refinement settings.

#Layers	w $\mathcal{L}_c$					w/o $\mathcal{L}_c$				
	RGB-BRS	sb	bmsb	bmsb-m	bmconv	RGB-BRS	sb	bmsb	bmsb-m	bmconv
L0 (AUC)	0.8820	-	-	-	-	0.8678	-	-	-	-
L1 (AUC)	-	0.8847	0.8909	0.8893	0.8964	-	0.8787	0.8849	0.8800	0.8889
L2 (AUC)	-	0.8870	0.8924	0.8912	0.8962	-	0.8806	0.8859	0.8826	0.8856
L3 (AUC)	-	0.8872	0.8932	0.8944	0.8966	-	0.8795	0.8862	0.8826	0.8733
L0 (mIoU*)	0.8965	-	-	-	-	0.8873	-	-	-	-
L1 (mIoU*)	-	0.8943	0.9019	0.9003	0.9080	-	0.8917	0.8966	0.8936	0.9028
L2 (mIoU*)	-	0.8986	0.9049	0.9005	0.9070	-	0.8924	0.8996	0.9006	0.9055
L3 (mIoU*)	-	0.9002	0.9049	0.9055	0.9083	-	0.8973	0.9006	0.9011	0.9000

Table 7. Area Under Curve (AUC) computed using mIoU and maximum mIoU achieved in total number of clicks (mIoU*) on **Cityscapes** for all backpropagating refinement settings.

#Layers	w $\mathcal{L}_c$					w/o $\mathcal{L}_c$				
	RGB-BRS	sb	bmsb	bmsb-m	bmconv	RGB-BRS	sb	bmsb	bmsb-m	bmconv
L0 (AUC)	0.6752	-	-	-	-	0.6543	-	-	-	-
L1 (AUC)	-	0.7517	0.7521	0.7481	0.7680	-	0.7242	0.7154	0.7116	0.7270
L2 (AUC)	-	0.7624	0.7659	0.7640	0.7774	-	0.7360	0.7301	0.7308	0.7420
L3 (AUC)	-	0.7676	0.7714	0.7708	0.7791	-	0.7369	0.7391	0.7380	0.7109
L0 (mIoU*)	0.7266	-	-	-	-	0.7169	-	-	-	-
L1 (mIoU*)	-	0.7896	0.7918	0.7909	0.8148	-	0.7802	0.7620	0.7563	0.7795
L2 (mIoU*)	-	0.8034	0.8092	0.8048	0.8213	-	0.7978	0.7859	0.7851	0.8019
L3 (mIoU*)	-	0.8101	0.8153	0.8154	0.8215	-	0.7967	0.7919	0.7936	0.7868

Table 8. Area Under Curve (AUC) computed using mIoU and maximum mIoU achieved in total number of clicks (mIoU*) on **Mapillary Vista** for all backpropagating refinement settings.

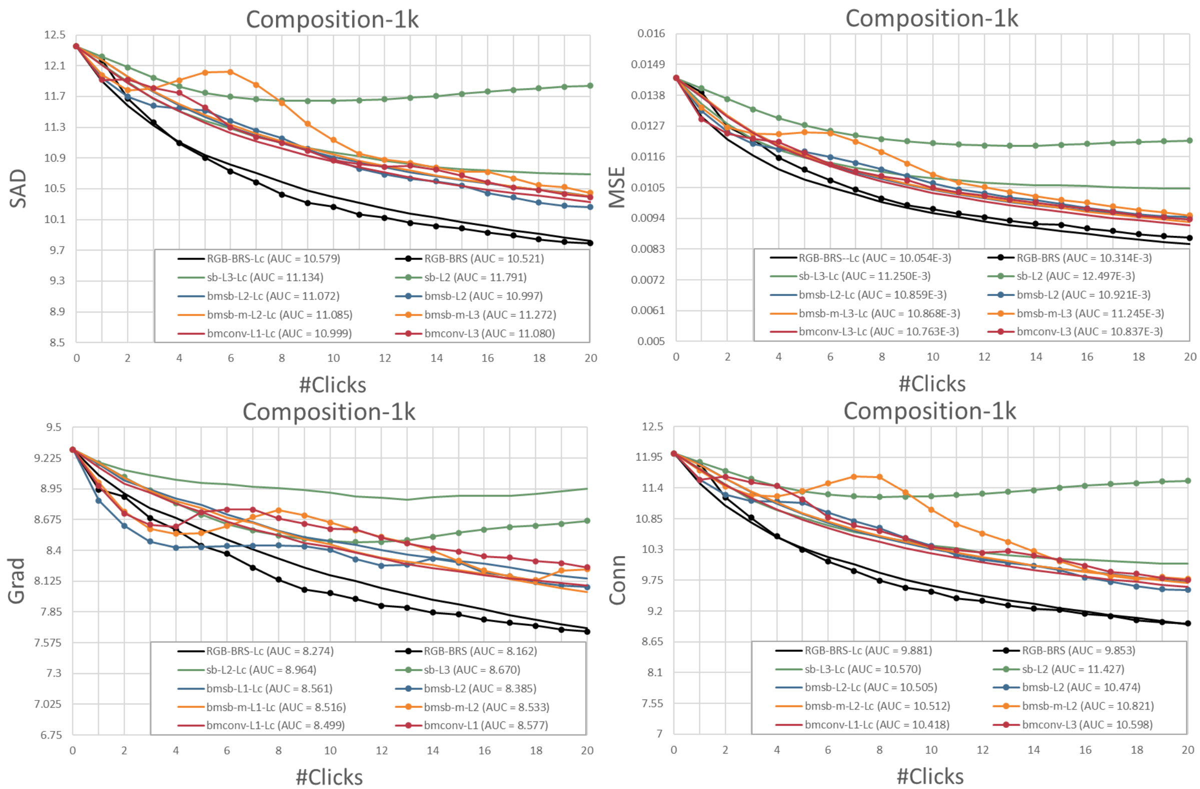


Figure 1. SAD, MSE, Grad and Conn on **Composition-1k** for image matting. Results denoted with $\mathcal{L}_c$ are obtained using the consistency loss.

Metrics	w $\mathcal{L}_c$					w/o $\mathcal{L}_c$				
	RGB-BRS	sb	bmsb	bmsb-m	bmconv	RGB-BRS	sb	bmsb	bmsb-m	bmconv
$\text{AUC}_{\delta 1} \uparrow$	0.9610	0.9554	0.9560	0.9555	0.9628	0.9592	0.9549	0.9556	0.9552	0.9623
$\text{AUC}_{\delta 2} \uparrow$	0.9914	0.9904	0.9906	0.9902	0.9917	0.9908	0.9900	0.9904	0.9901	0.9912
$\text{AUC}_{\delta 3} \uparrow$	0.9974	0.9972	0.9974	0.9972	0.9976	0.9974	0.9971	0.9974	0.9972	0.9975
$\text{AUC}_{\text{AbsRel}} \downarrow$	0.0565	0.0636	0.0628	0.0633	0.0555	0.0578	0.0640	0.0630	0.0636	0.0559
$\text{AUC}_{\text{SqRel}} \downarrow$	0.0289	0.0326	0.0321	0.0332	0.0281	0.0299	0.0335	0.0325	0.0333	0.0289
$\text{AUC}_{\text{RMSE}} \downarrow$	0.2512	0.2722	0.2688	0.2726	0.2487	0.2574	0.2741	0.2702	0.2733	0.2512
$\text{AUC}_{\text{RMSE}\ log} \downarrow$	0.0861	0.0940	0.0927	0.0938	0.0849	0.0880	0.0946	0.0931	0.0941	0.0857
$\delta_1{}^* \uparrow$	0.9830	0.9752	0.9765	0.9754	0.9833	0.9821	0.9745	0.9761	0.9752	0.9827
$\delta_2{}^* \uparrow$	0.9960	0.9942	0.9943	0.9940	0.9959	0.9957	0.9941	0.9957	0.9942	0.9941
$\delta_3{}^* \uparrow$	0.9987	0.9982	0.9983	0.9982	0.9986	0.9988	0.9982	0.9986	0.9983	0.9982
$\text{AbsRel}^* \downarrow$	0.0375	0.0472	0.0450	0.0467	0.0372	0.0384	0.0475	0.0453	0.0467	0.0378
$\text{SqRel}^* \downarrow$	0.0151	0.0210	0.0203	0.0221	0.0151	0.0157	0.0213	0.0208	0.0214	0.0158
$\text{RMSE}^* \downarrow$	0.1858	0.2191	0.2131	0.2181	0.1876	0.1894	0.2200	0.2134	0.2182	0.1898
$\text{RMSE}\ log^* \downarrow$	0.0627	0.0749	0.0726	0.0744	0.0627	0.0639	0.0753	0.0729	0.0744	0.0635

Table 9. Area Under Curve (AUC) computed using various depth estimation metrics and maximum scores achieved in total number of clicks (denoted with "*") on **NYU-Depth-V2** for all backpropagating refinement settings. Since BTSNet has a G-BRS layout that only contains 1 insertion, we omit the #layers in this table. For clarification, RGB-BRS does not utilize G-BRS layers and other G-BRS layout contains 1 G-BRS layer.

#Layers	RGB-BRS	G-BRS-sb	G-BRS-bmsb	G-BRS-bmsb-m	G-BRS-bmconv
L0	1.542	-	-	-	-
L1	-	0.687	0.601	0.450	0.583

Table 10. SPC for all backpropagating refinement layouts with $\mathcal{L}_c$ on **SBD**.

#Layers	RGB-BRS	G-BRS-sb	G-BRS-bmsb	G-BRS-bmsb-m	G-BRS-bmconv
L0	8.362	-	-	-	-
L1	-	5.627	5.641	5.633	5.633
L2	-	5.649	5.652	5.665	5.661
L3	-	5.691	5.716	5.718	5.727

Table 11. SPC for all backpropagating refinement layouts with $\mathcal{L}_c$ on **Cityscapes**.

#Layers	RGB-BRS	G-BRS-sb	G-BRS-bmsb	G-BRS-bmsb-m	G-BRS-bmconv
L0	8.125	-	-	-	-
L1	-	5.153	5.133	5.145	5.134
L2	-	5.159	5.153	5.165	5.159
L3	-	5.222	5.224	5.233	5.231

Table 12. SPC for all backpropagating refinement layouts with $\mathcal{L}_c$ on **Mapillary Vista**.

#Layers	RGB-BRS	G-BRS-sb	G-BRS-bmsb	G-BRS-bmsb-m	G-BRS-bmconv
L0	2.473	-	-	-	-
L1	-	1.350	1.338	1.347	1.344
L2	-	1.348	1.346	1.359	1.355
L3	-	1.360	1.367	1.388	1.375

Table 13. SPC for all backpropagating refinement layouts with $\mathcal{L}_c$ on **Composition-1k**.

#Layers	RGB-BRS	G-BRS-sb	G-BRS-bmsb	G-BRS-bmsb-m	G-BRS-bmconv
L0	3.205	-	-	-	-
L1	-	2.040	2.040	2.045	2.055

Table 14. SPC for all backpropagating refinement layouts with $\mathcal{L}_c$ on **NYU-Depth-V2**.

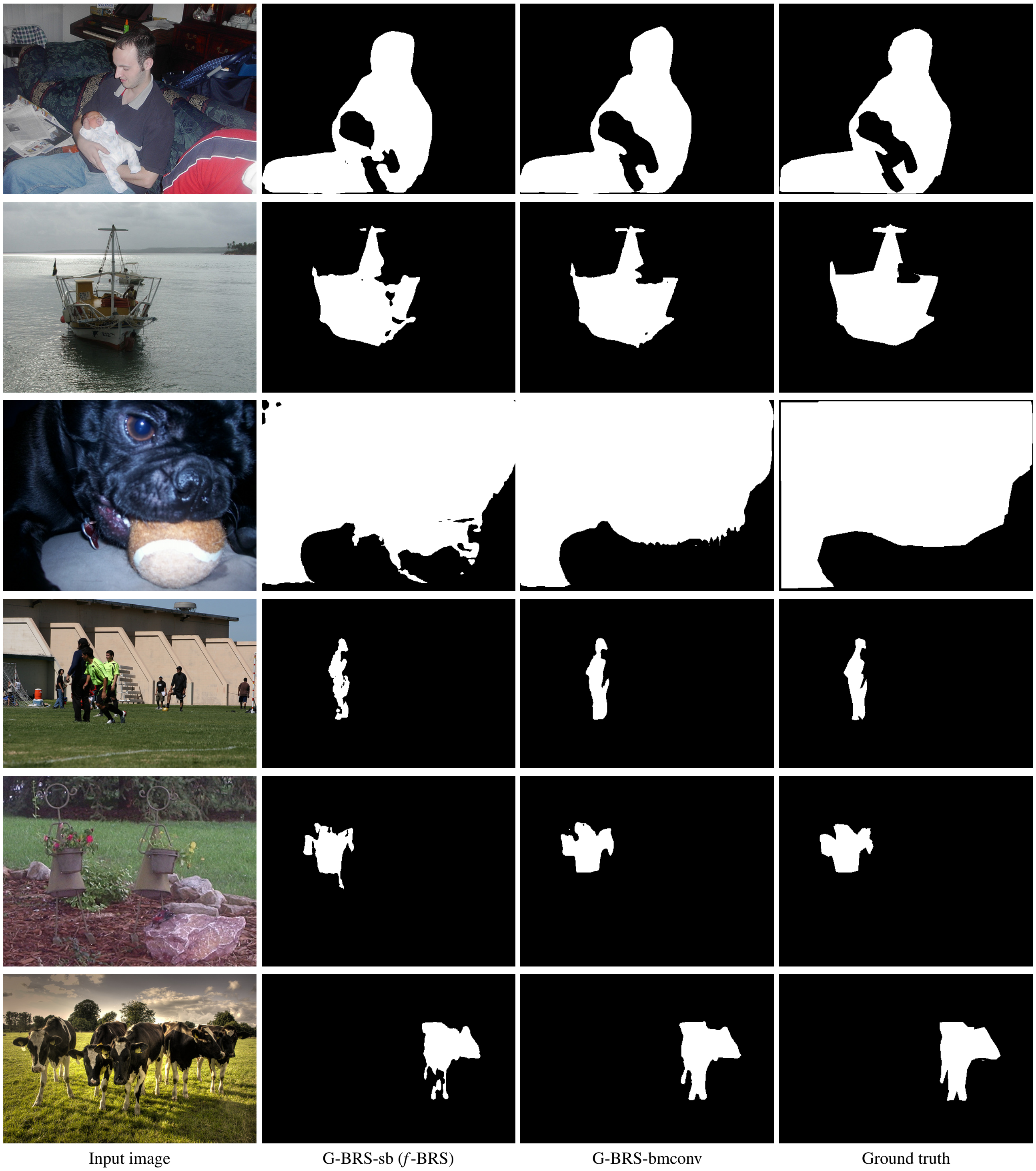
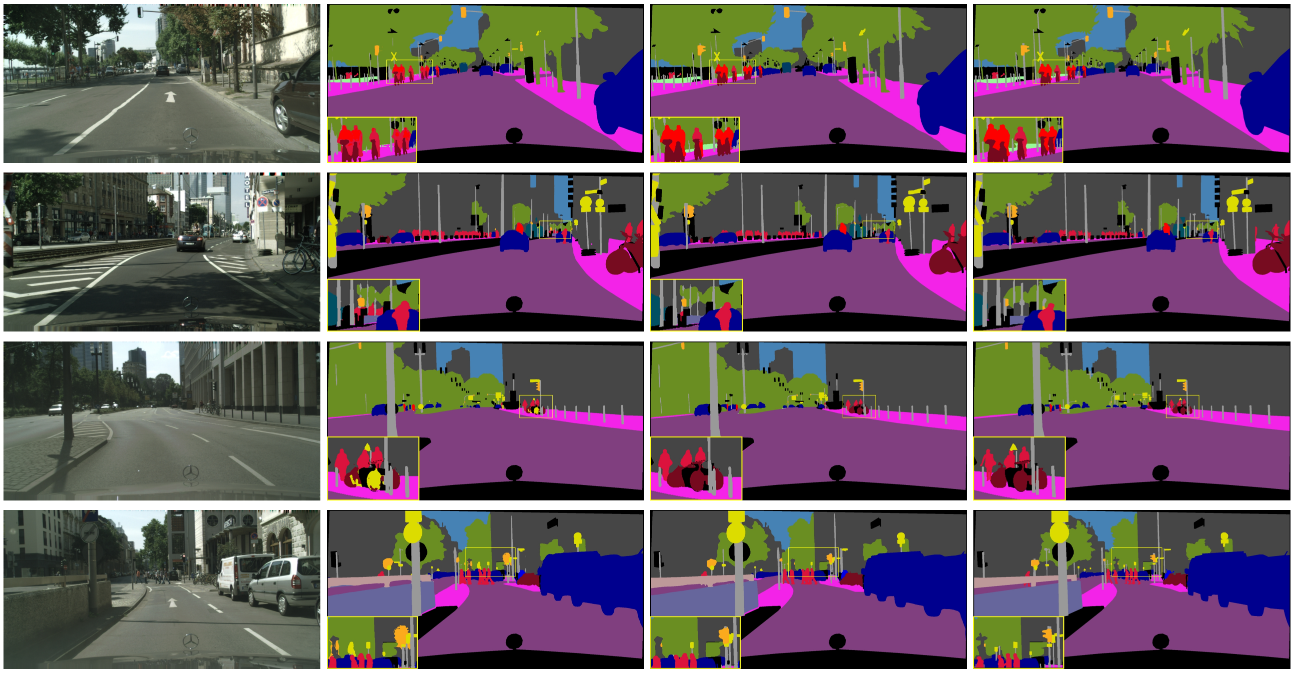
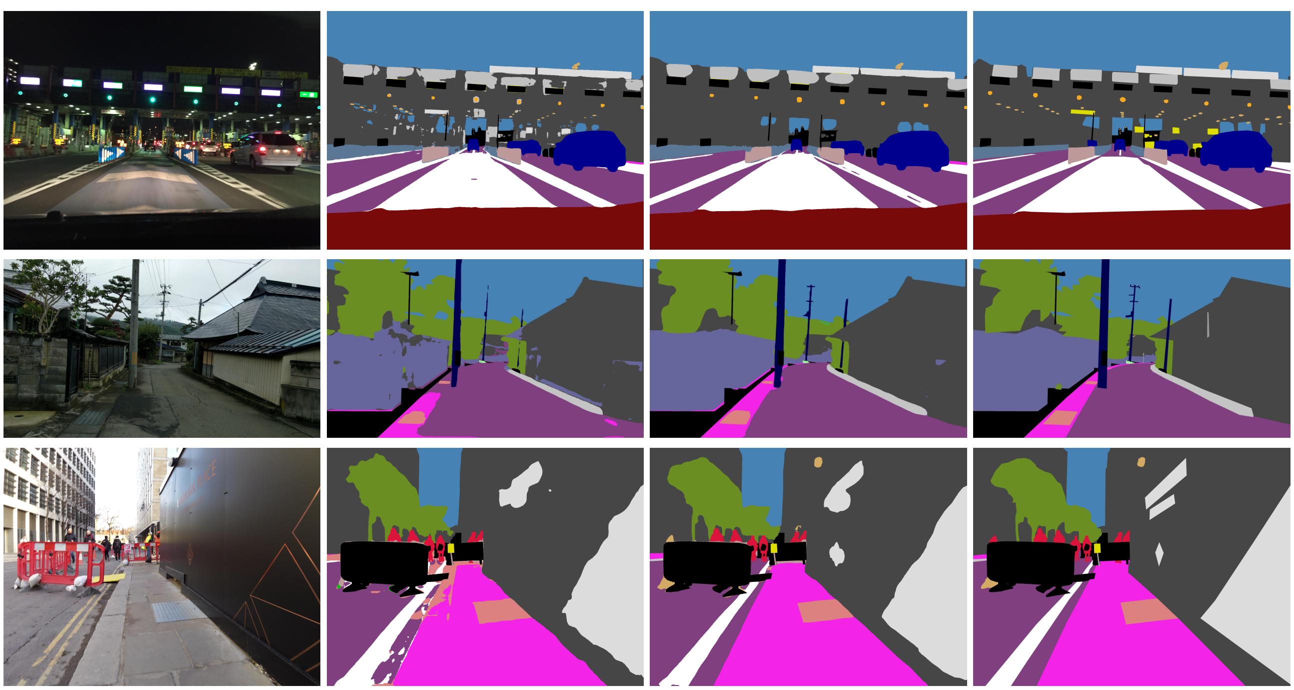


Figure 2. Qualitative comparison between performance of the G-BRS-sb (f-BRS) layer and our G-BRS-bmconv layer for the task of interactive segmentation on **SBD**.



Input image G-BRS-sb G-BRS-bmconv Ground truth

Figure 3. Qualitative comparison between performance of the G-BRS-sb layer and our G-BRS-bmconv layer for the task of semantic segmentation on **Cityscapes**. Since both methods are highly accurate with G-BRS-bmconv providing more refined details, we highlight the differences in the yellow zoomed window. Invalid labels are shown in black. Best viewed in magnification.



Input image G-BRS-sb G-BRS-bmconv Ground truth

Figure 4. Qualitative comparison between performance of the G-BRS-sb layer and our G-BRS-bmconv layer for the task of semantic segmentation on **Mapillary Vista**. Invalid labels are shown in black.

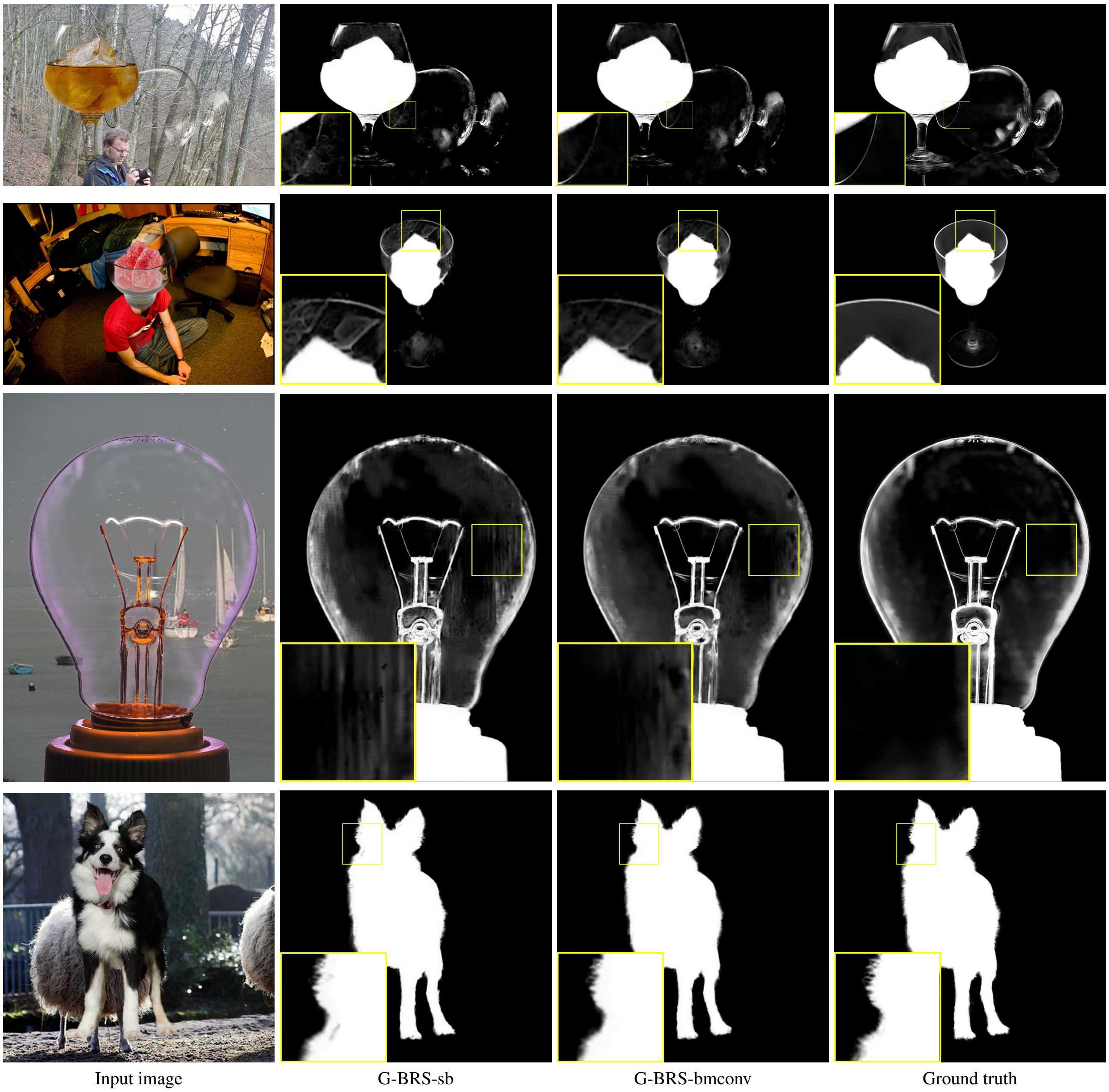


Figure 5. Qualitative comparison between performance of the G-BRS-sb layer and our G-BRS-bmconv layer for the task of image matting on **Composition-1k**.

Figure 6. Qualitative comparison between performance of the G-BRS-sb layer and our G-BRS-bmconv layer for the task of depth estimation on **NYU-Depth-V2**. Invalid region is shown in black in the ground truth.