

Appendix for “Hierarchical Image Peeling: A Flexible Scale-space Filtering Framework”

A. Application on saliency detection

	ECSSD	PASCAL-S	HKU-IS	SOD	DUTS-TE
HS [6]	0.226	0.260	0.213	0.280	0.241
HS [6] at $t = 1$	0.214	0.258	0.195	0.269	0.228
HS [6] at $t = 2$	0.203	0.257	0.183	0.257	0.217
HS [6] at $t = 3$	0.223	0.260	0.188	0.267	0.218
HS [6] at $t = 4$	0.230	0.272	0.202	0.281	0.227
PoolNet [2]	0.039	0.074	0.032	0.100	0.039
PoolNet [2] at $t = 1$	0.053	0.100	0.044	0.126	0.056
PoolNet [2] at $t = 2$	0.072	0.118	0.059	0.147	0.073
PoolNet [2] at $t = 3$	0.097	0.141	0.079	0.164	0.091
PoolNet [2] at $t = 4$	0.105	0.147	0.086	0.173	0.097
CSF [1]	0.033	0.068	0.030	0.098	0.037
CSF [1] at $t = 1$	0.050	0.097	0.041	0.121	0.056
CSF [1] at $t = 2$	0.071	0.117	0.057	0.147	0.073
CSF [1] at $t = 3$	0.099	0.141	0.077	0.171	0.091
CSF [1] at $t = 4$	0.108	0.150	0.087	0.182	0.097
EGNet [8]	0.040	0.074	0.031	0.097	0.039
EGNet [8] at $t = 1$	0.053	0.099	0.042	0.124	0.055
EGNet [8] at $t = 2$	0.072	0.117	0.057	0.139	0.071
EGNet [8] at $t = 3$	0.098	0.137	0.076	0.161	0.089
EGNet [8] at $t = 4$	0.107	0.143	0.084	0.171	0.096

Table 1: Quantitative comparison in MAE. The lower the MAE, the better. *Name* at t refers to the saliency map of a filtering/peeling result at t -th recurrent step.

We further conduct the experiments that directly performing salient object detection on the filtering/smoothing results of our framework at 4 recurrent steps. Table 1 shows the quantitative results in terms of MAE on public datasets: ECSSD [6], PASCAL-S [4], HKU-IS [3], SOD [5], and DUTS-TE [7]. It can be seen that for all the results by PoolNet [2], CSF [1] and EGNet [8], the MAE increases with the smoothing level. Though most of the results by HS [6] when performing saliency detection on the filtered images are better, there still exist some worse cases. Table 1 and the quantitative results in our original paper verify that it is our scale-space filtering framework instead of image smoothing works on saliency detection task. The reason may be that different scales make different features prominent, and single scale smoothing cannot cover all the useful features for salient object detection (the textures with high contrast are mistakenly considered as salient regions in small-scales, while some salient structures are overly smoothed in large-scales). Please see visual results in Figure 1 - 4 for evidence.

B. Effect of hyper-parameters

We train our network under different configurations of hyper-parameters in an unsupervised manner. As can be seen from Figure 6, altering $\lambda_{pre}^{\mathcal{P}}$ influences the image edges. Specifically, large $\lambda_{pre}^{\mathcal{P}}$ leads to thick, messy and unnatural edges. From the Figure 7, tuning the weight $\lambda_{con}^{\mathcal{P}}$ for the term $\mathcal{L}_{con}^{\mathcal{P}}$ controls the smoothness of the results. We observe that small $\lambda_{con}^{\mathcal{P}}$ is unable to produce satisfactory results, *e.g.* some regions are not completely smoothed out, such as the first row in Figure 7, the neck of the bear. When enlarging the weight $\lambda_{con}^{\mathcal{P}}$, the structures/edges in filtered results indicated by corresponding guidance maps are gradually becoming blurry. In addition, it can be observed that both large $\lambda_{pre}^{\mathcal{P}}$ and $\lambda_{con}^{\mathcal{P}}$ cause obvious color shift.

C. Additional experimental results

We show more experimental results in this section. The edge confidence maps we use for plotting the precision-recall curve in our main draft are shown in Figure 5, while the scale-space filtering results are shown in Figure 8 - 9.

D. Effectiveness of edge guidance

Edge guidance is critical for our framework, which controls the appearance of smoothing/filtering results. An intriguing question arises as to what if there is a large structural difference between the edge and the input image. To figure out this, we take the edge guidance predicted from one image as the guidance of another image. The results are shown in Figure 10. We can see that even the edges are greatly different from the input image, our HIPe can still produce smoothing/filtering results strictly adhering to the edge guidance.

References

- [1] Shang-Hua Gao, Yong-Qiang Tan, Ming-Ming Cheng, Chengze Lu, Yunpeng Chen, and Shuicheng Yan. Highly efficient salient object detection with 100k parameters. In *ECCV*, 2020.
- [2] Qibin Hou, Ming-Ming Cheng, Xiaowei Hu, Ali Borji, Zhuowen Tu, and Philip Torr. Deeply supervised salient ob-

- ject detection with short connections. *TPAMI*, 41(4):815–828, 2019.
- [3] Guanbin Li and Y. Yu. Visual saliency based on multiscale deep features. In *CVPR*, pages 5455–5463, 2015.
 - [4] Yin Li, Xiaodi Hou, Christof Koch, James M. Rehg, and Alan L. Yuille. The secrets of salient object segmentation. In *CVPR*, pages 280–287, 2014.
 - [5] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *ICCV*, 2001.
 - [6] Qiong Yan, Li Xu, Jianping Shi, and Jiaya Jia. Hierarchical saliency detection. In *CVPR*, pages 1155–1162, 2013.
 - [7] Chuan Yang, Lihe Zhang, Huchuan Lu, Xiang Ruan, and Ming-Hsuan Yang. Saliency detection via graph-based manifold ranking. In *CVPR*, pages 3166–3173, 2013.
 - [8] Jia-Xing Zhao, Jiang-Jiang Liu, Deng-Ping Fan, Yang Cao, Jufeng Yang, and Ming-Ming Cheng. Egnet:edge guidance network for salient object detection. In *ICCV*, 2019.

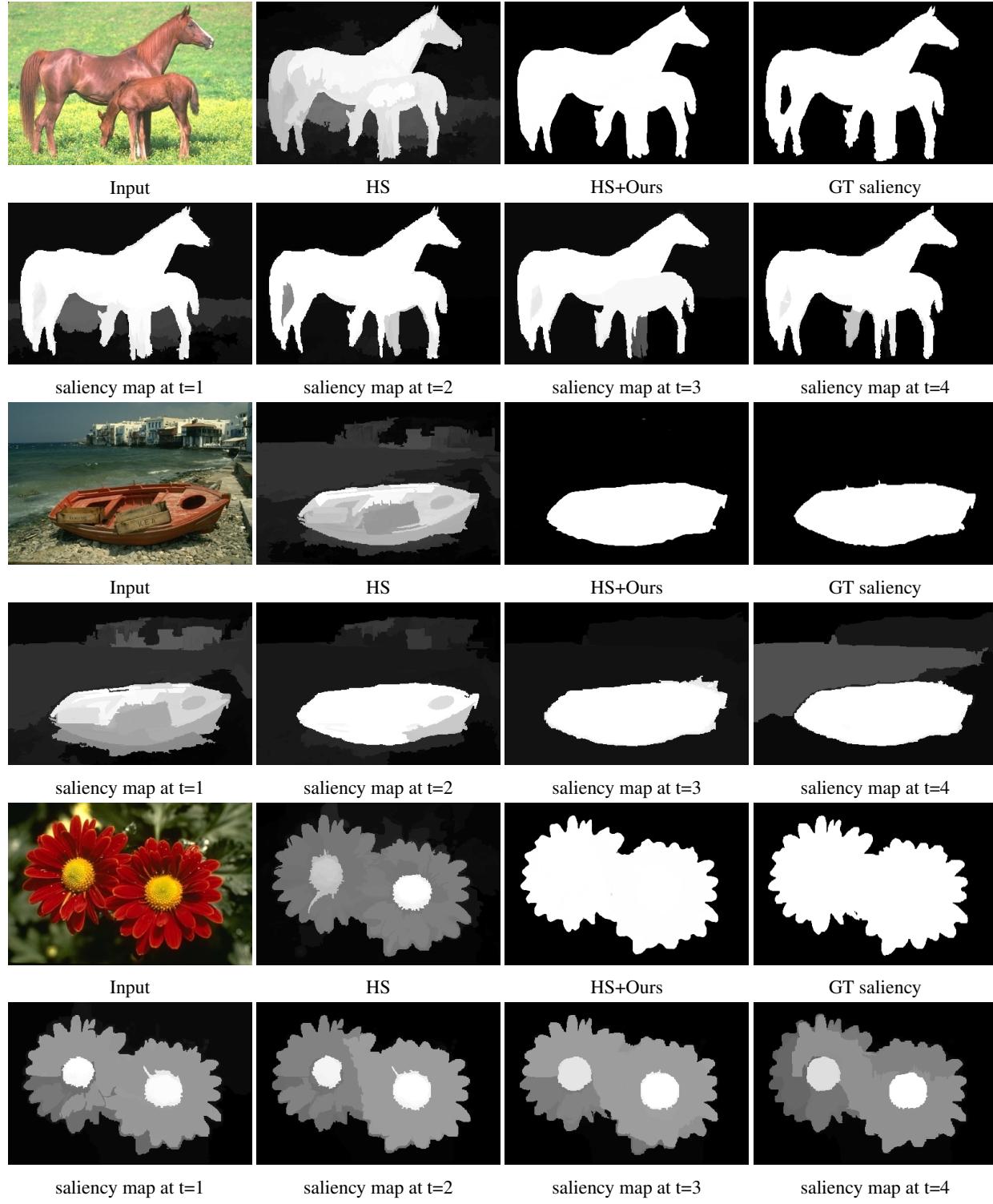


Figure 1: Visual comparison between HS and HS + Ours, and the saliency maps by HS of our four filtering results.

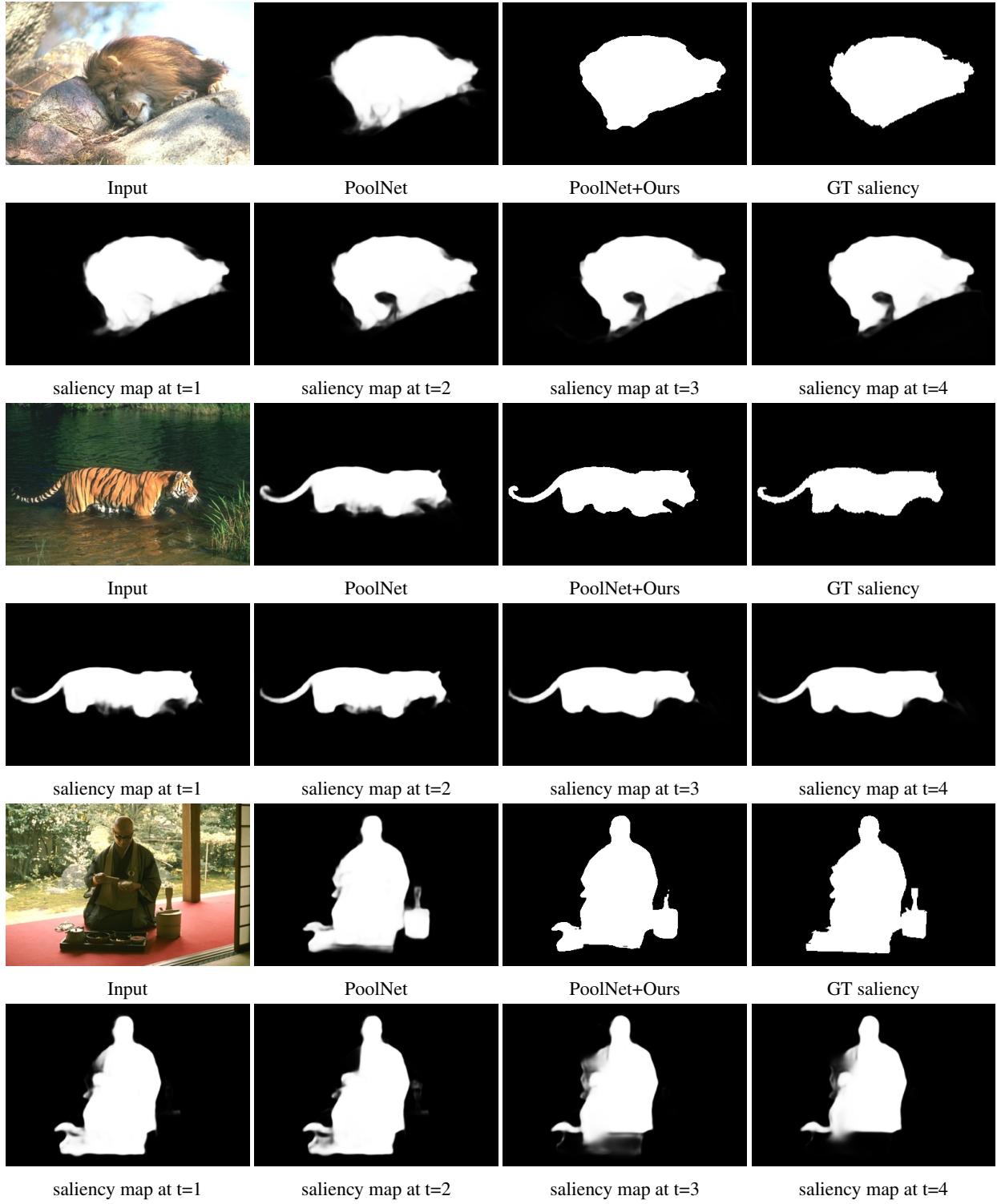


Figure 2: Visual comparison between PoolNet and PoolNet + Ours, and the saliency maps by PoolNet of our four filtering results.

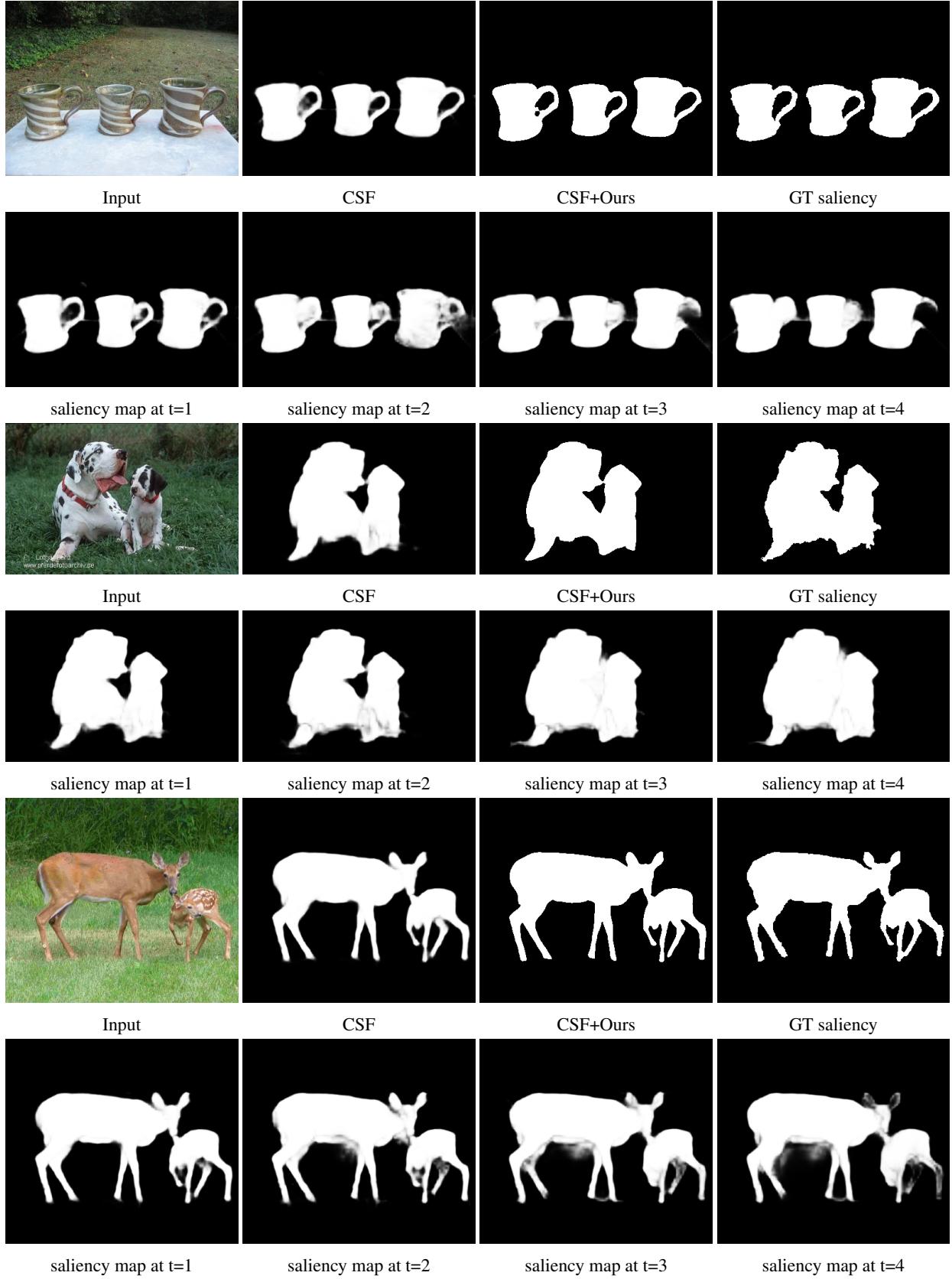


Figure 3: Visual comparison between CSF and CSF + Ours, and the saliency maps by CSF of our four filtering results.

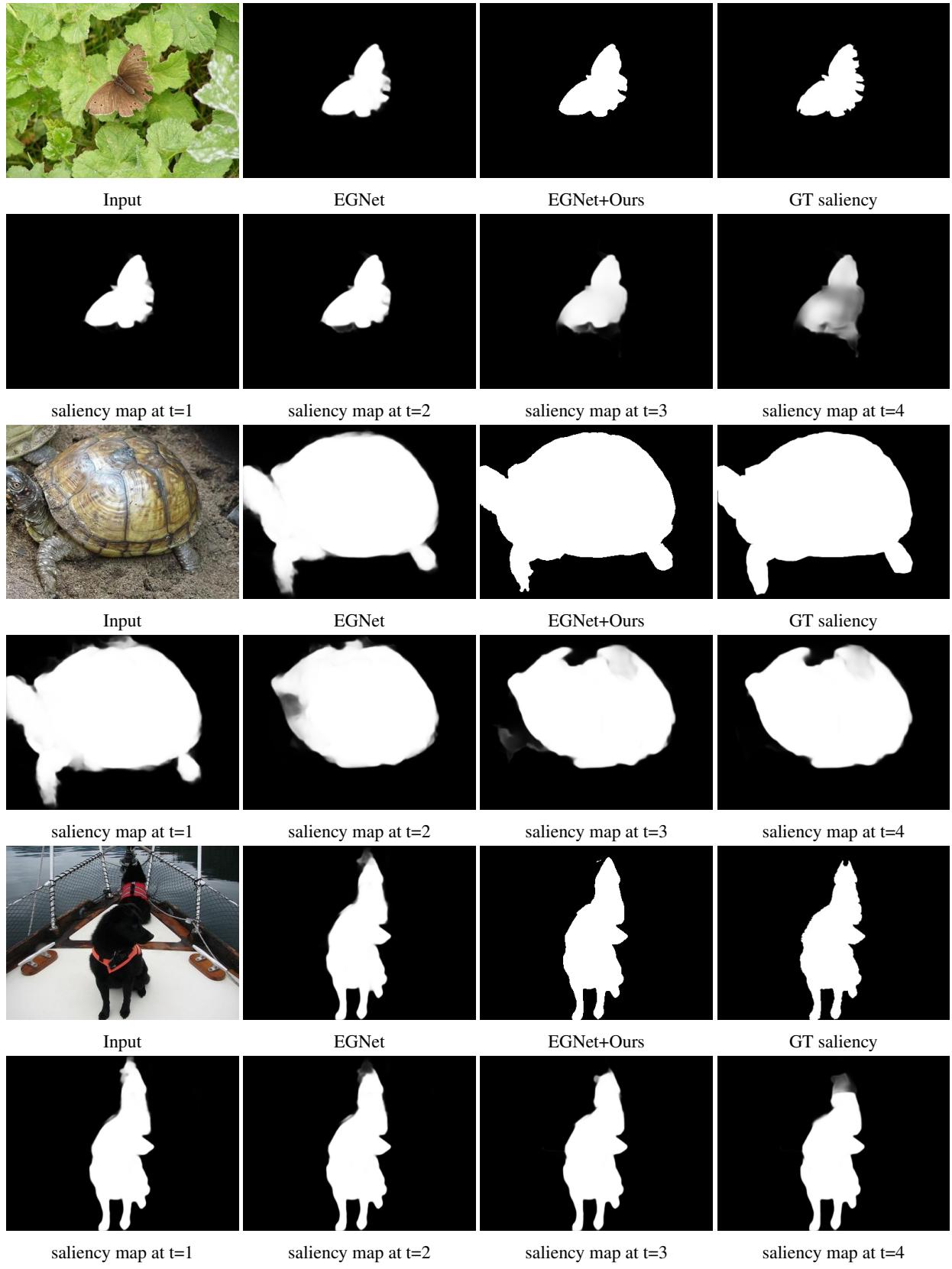


Figure 4: Visual comparison between EGNet and EGNet + Ours, and the saliency maps by EGNet of our four filtering results.

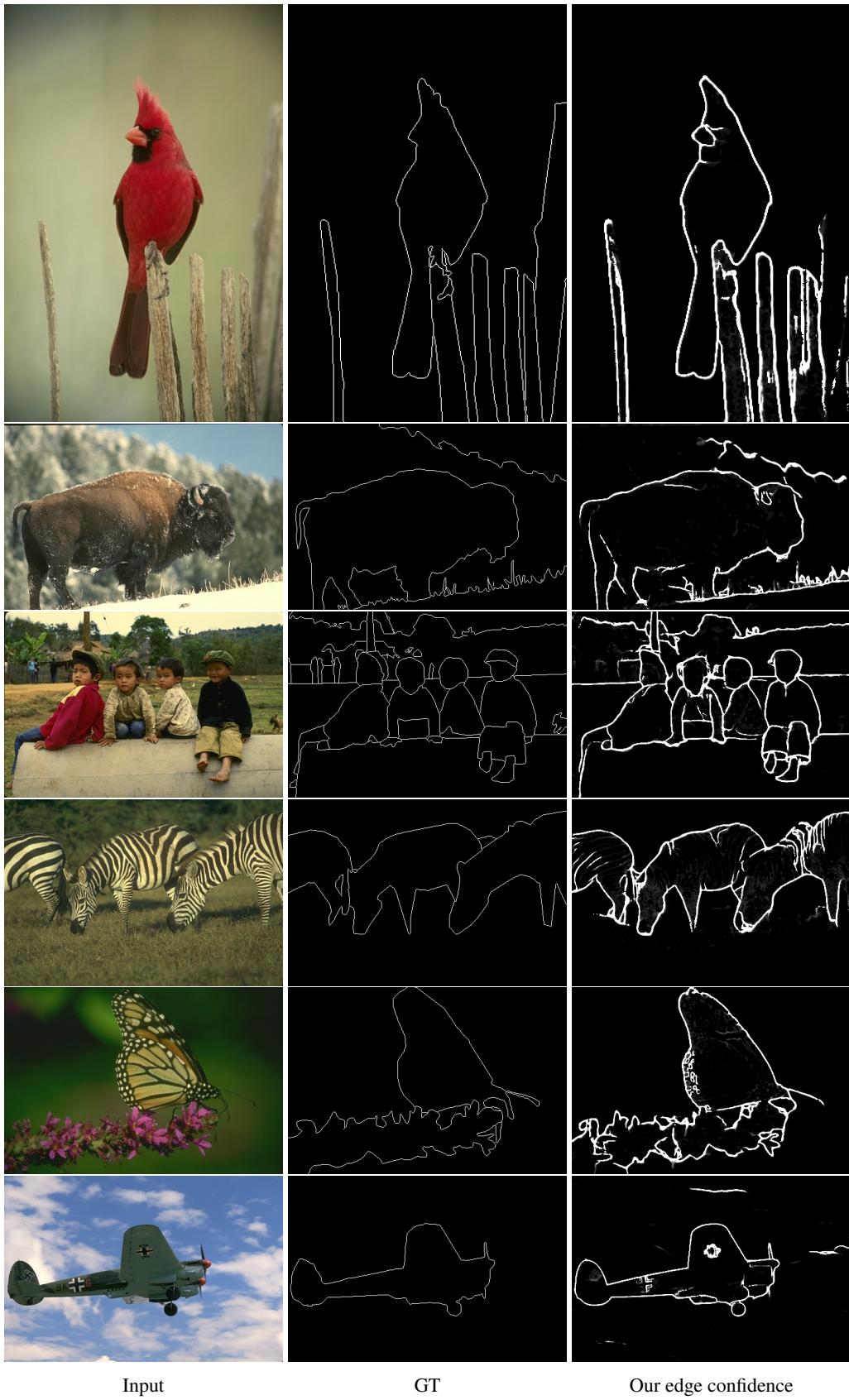


Figure 5: More examples of our edge confidence map.

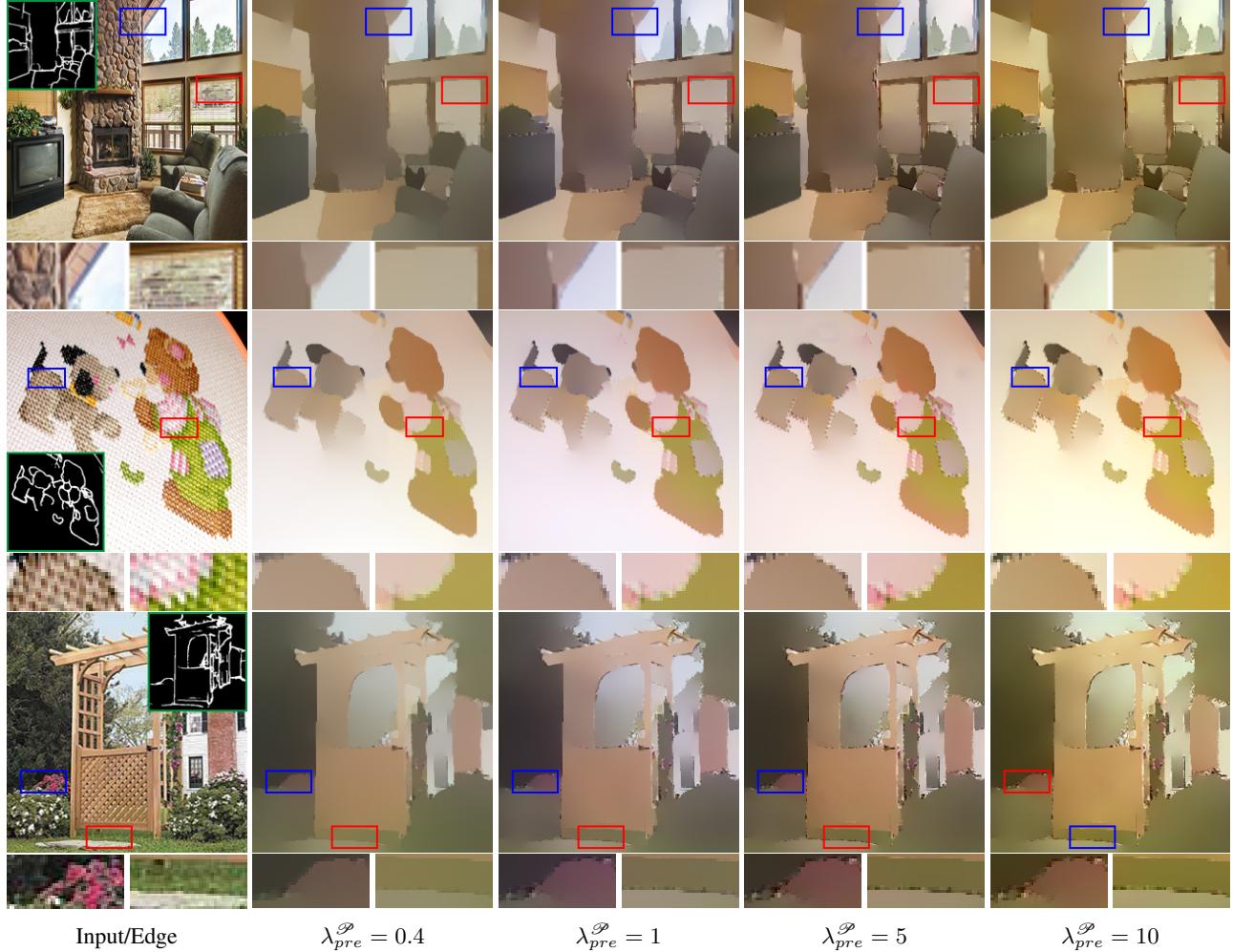


Figure 6: Effect of tuning the weight $\lambda_{pre}^{\mathcal{P}}$ of loss term $\mathcal{L}_{pre}^{\mathcal{P}}$. $\lambda_{con}^{\mathcal{P}}$ is fixed to be 4. The left-most column shows the input images and corresponding edge guidance. The regions of red and blue boxes are highlighted at the bottom of each image.



Figure 7: Effect of tuning the weight λ_{con}^P of loss term L_{con}^P . λ_{pre}^P is fixed to be 0.4. The left-most column shows the input images and corresponding edge guidance. The regions of red and blue boxes are highlighted at the bottom of each image.

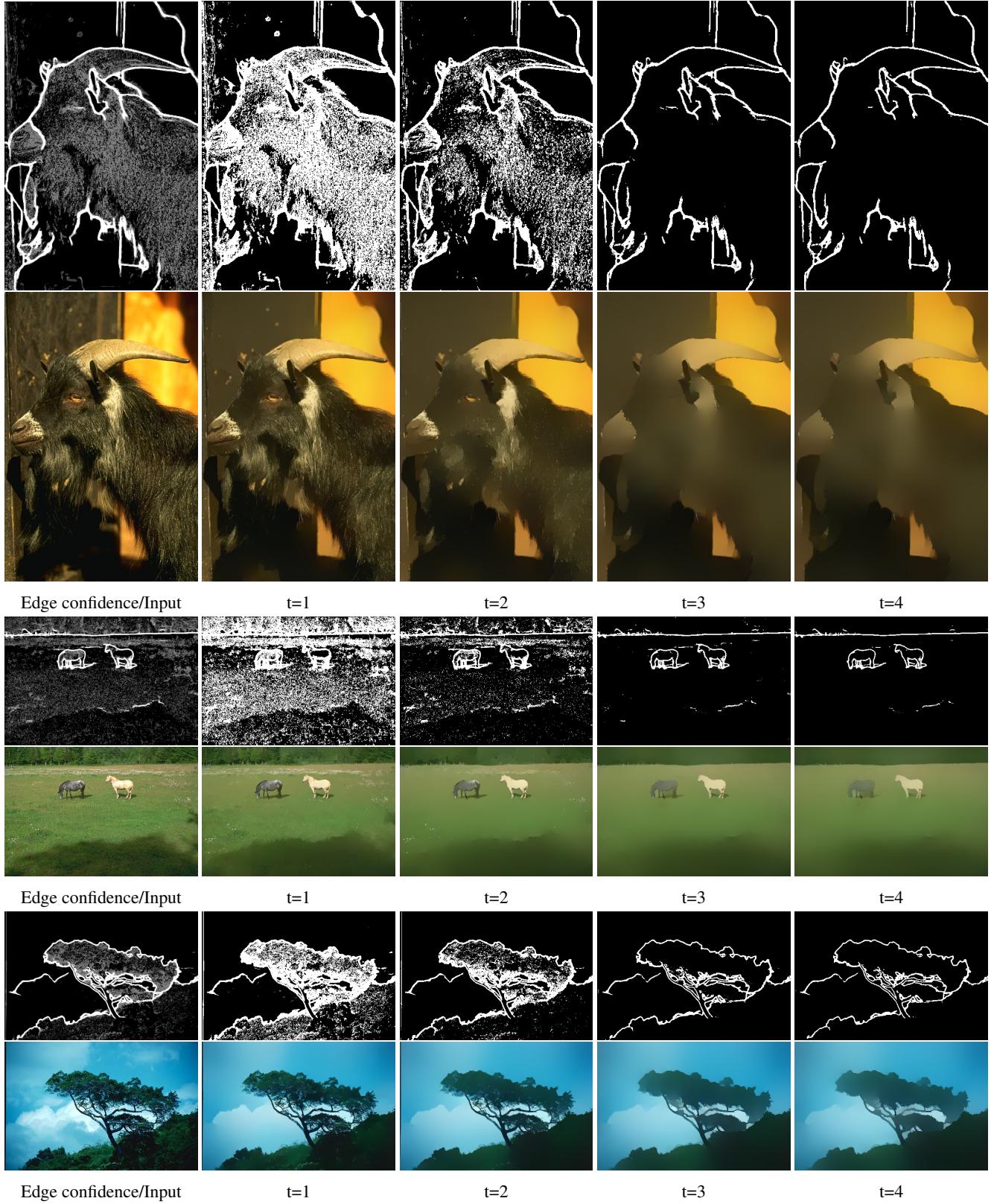


Figure 8: The additional scale-space filtering results. The edge confidence maps in this Figure are obtained by averaging the edges of four recurrent steps.

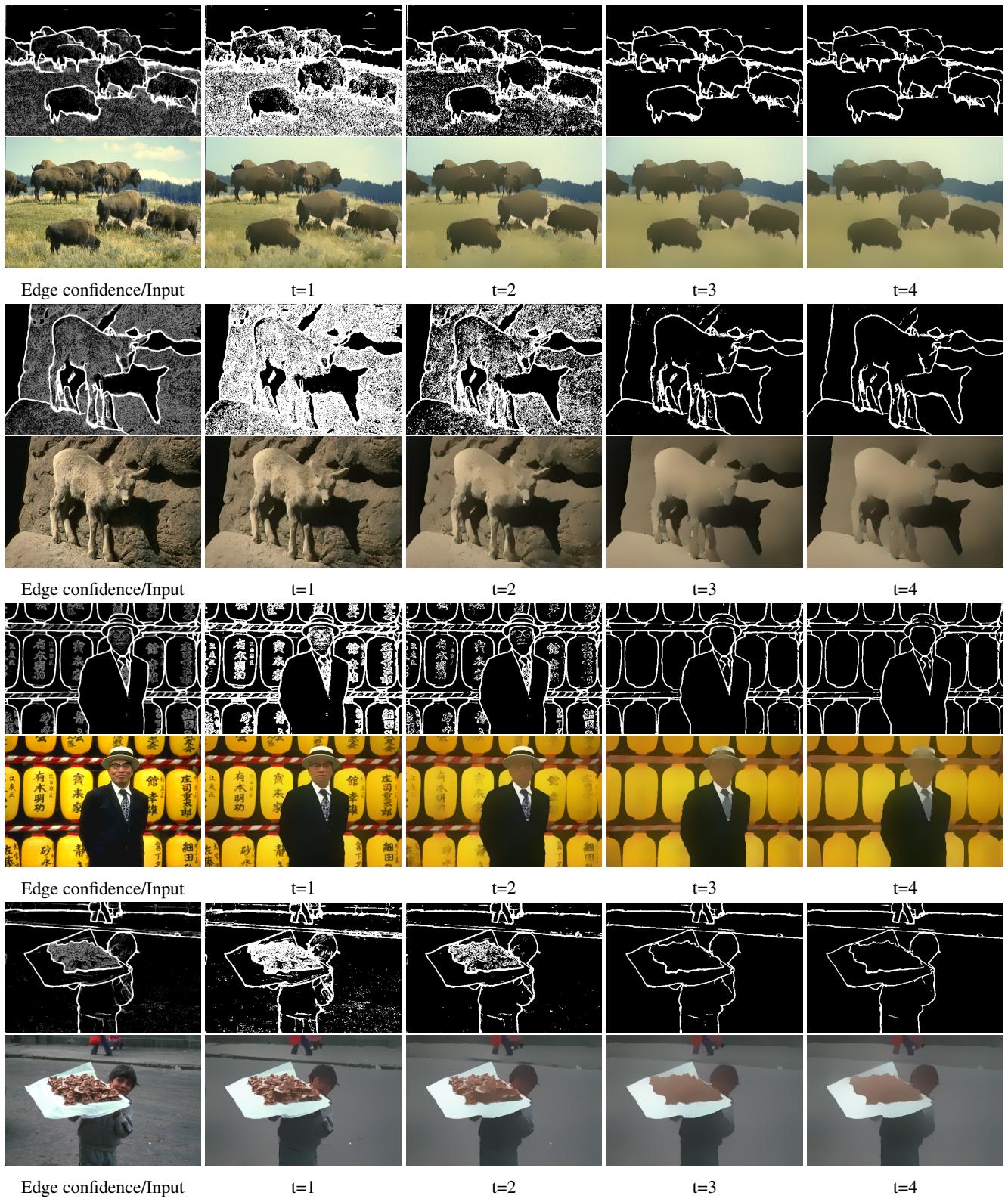


Figure 9: The additional scale-space filtering results. The edge confidence maps in this Figure are obtained by averaging the edges of four recurrent steps.

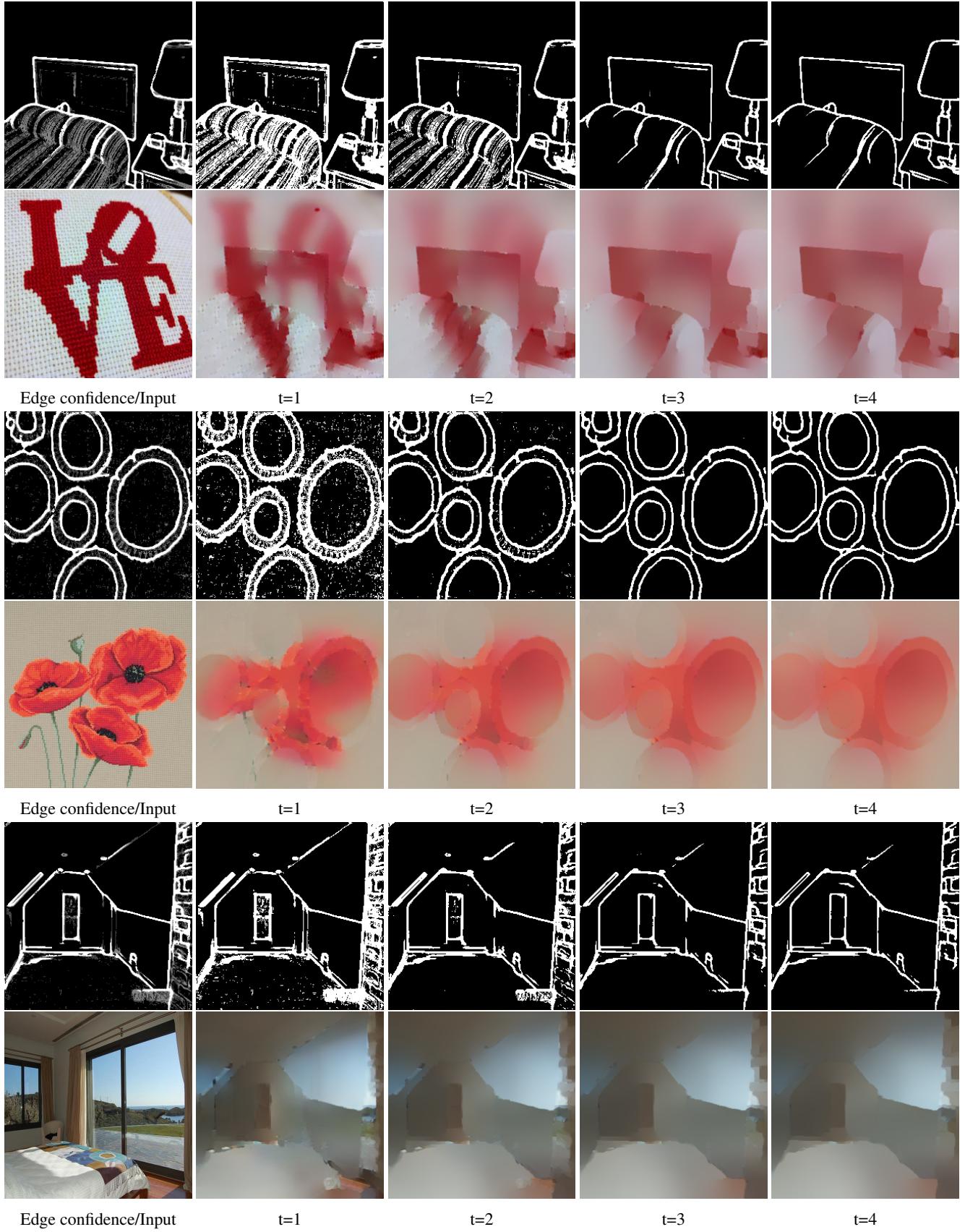


Figure 10: The impact of edge guidance. The edge confidence maps in this Figure are obtained by averaging the edges of four recurrent steps.