HM: Hybrid Masking for Few-Shot Segmentation (Supplementary Material)

Seonghyeon Moon[1], Samuel S. Sohn[1], Honglu Zhou[1], Sejong Yoon[2], Vladimir Pavlovic[1], Muhammad Haris Khan[3], and Mubbasir Kapadia[1]

[1] Rutgers University, New Jersey, USA
{sm2062,samuel.sohn,honglu.zhou,vladimir,mubbasir.kapadia}@rutgers.edu
[2] The College of New Jersey, New Jersey, USA yoons@tcnj.edu
[3] Mohamed Bin Zayed University of Artificial Intelligence, Abu Dhabi, UAE
muhammad.haris@mbzuai.ac.ae

Our method (HM) reveals improvement in mIoU on three standard benchmarks compared to the existing state-of-the-art HSNet [5] by generating features indicative of objects' contours, edges, and complex patterns through proposed hybrid masking. In this supplementary material, we further include the following materials:

1. Further analysis of training efficiency.
2. Additional qualitative results on three bench marks.
3. Further analysis of failure cases.

1 Training efficiency

Compared to the existing HSNet [5] (baesline), out method (HSNet-HM) shows improved training efficiency not only on the COCO-20^i dataset [4] but also on the other two benchmark datasets, PASCAL-5^i [1] and FSS-1000 [3]. Fig. 1 and Fig. 2 display that our method tends to converge faster than HSNet [5]. Table 1 shows the number of epochs it takes to get the best performing model on the validation set. In particular, the results of COCO-20^i with ResNet101 [2] show that our method (HM) has a better training efficiency.

Table 1: **Number of epochs to reach the best model.**

Backbone feature	Masking methods	PASCAL-5^i 1-shot 20^0	20^1	20^2	20^3	mEpoch	COCO-20^i 1-shot 20^0	20^1	20^2	20^3	mEpoch	FSS-1000^i 1-shot Epoch
ResNet50 [2]	HSNet [5]	345	433	204	244	306.5	262	249	160	295	241.5	530
	HSNet-HM	**188**	**117**	**45**	**56**	**101.5**	**41**	**32**	**32.8**	**23**	**35**	**177**
ResNet101 [2]	HSNet [5]	177	185	136	199	174.3	235	251	345	355	296.5	886
	HSNet-HM	**73**	**95**	**30**	**72**	**67.5**	**52**	**27**	**14**	**14**	**26.8**	**298**

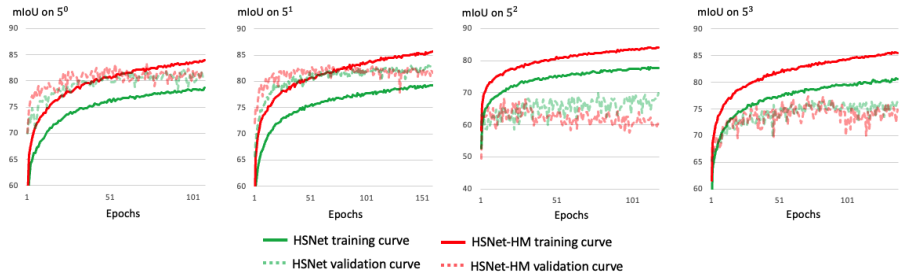

Fig. 1: Training profiles of HSNet[5] and HSNet-HM on PASCAL-5^i [1]

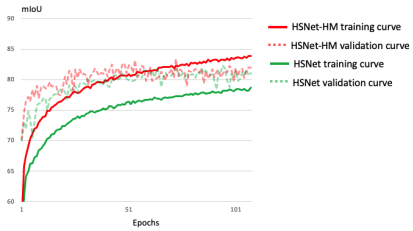

Fig. 2: Training profiles of HSNet[5] and HSNet-HM on FSS-1000.

2 Qualitative results

HM reveal better performance in mIoU over other methods on three standard benchmarks. Fig. 3,4,5 shows the qualitative results of each benchmark. HM is able to segment the target object in the query image precisely. Also, even if only a part of the object exists in the support set or query image because of occlusion, HM can segment the target object more accurately.

3 Failure Cases

Although HM provides significant improvements over existing best methods, it can be further improved after analyzing the failure cases shown in Fig. 6. We discuss these failure cases below:

- When the target object in the support set or query image is very small, HM was unable to segment accurately. We believe that the reason for this is that feature maps could not contain enough information about the target object.
- A problem arose when the shape of the target object of the support set and the target image of the query image were different. HM was able to detect where the target object was, but the details failed to segment.
- If there is an object with a similar texture to the target object in the query image, there is a difficulty in segmentation. For example, the textures of cows and horses were quite similar, and because of this, the trained model could not distinguish between horses and cows.

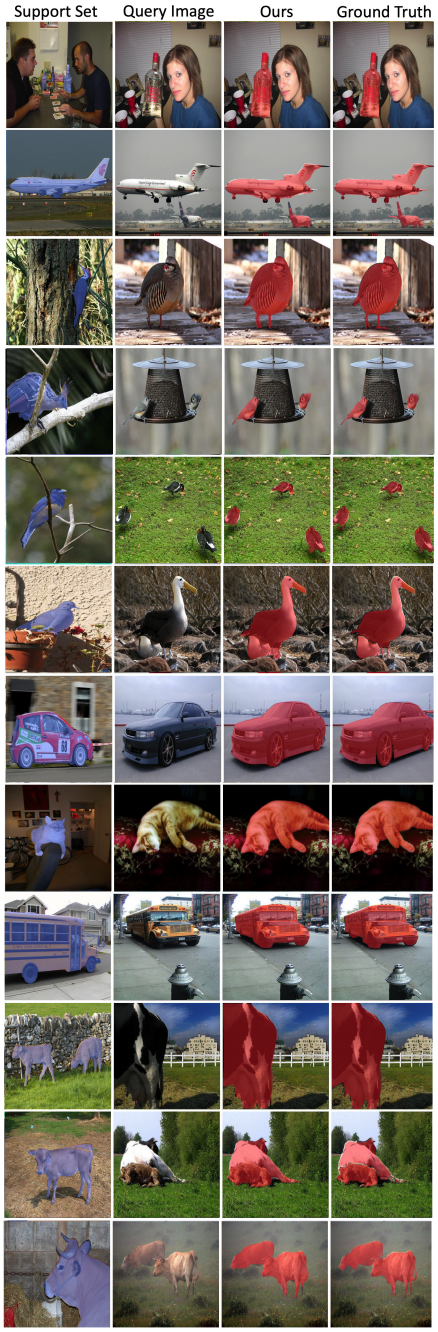

Fig. 3: HSNet-HM Qualitative result on PASCAL-5^i [1]

Fig. 4: HSNet-HM Qualitative result on FSS-1000 [3]

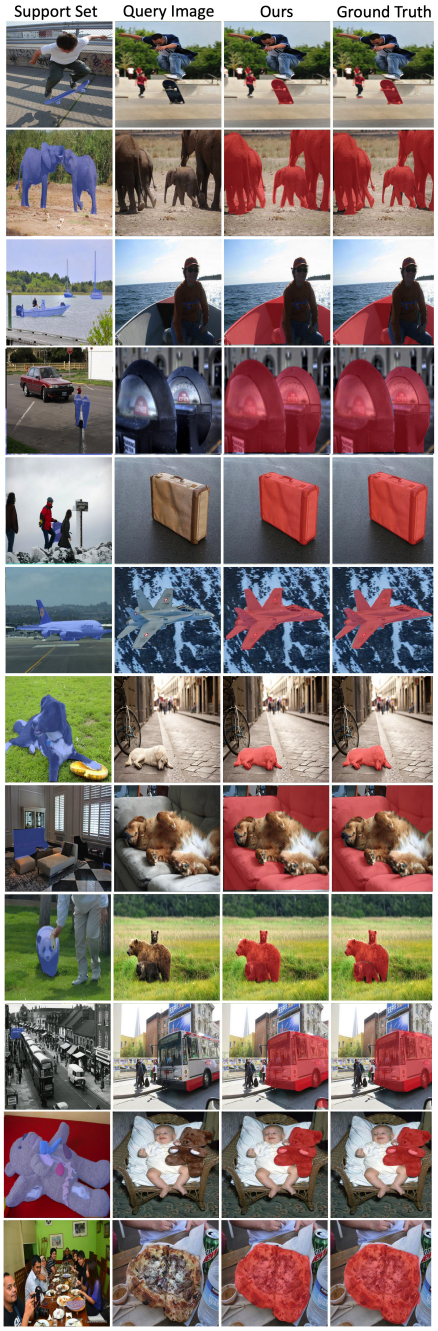

Fig. 5: HSNet-HM Qualitative result on COCO-20^i [4]

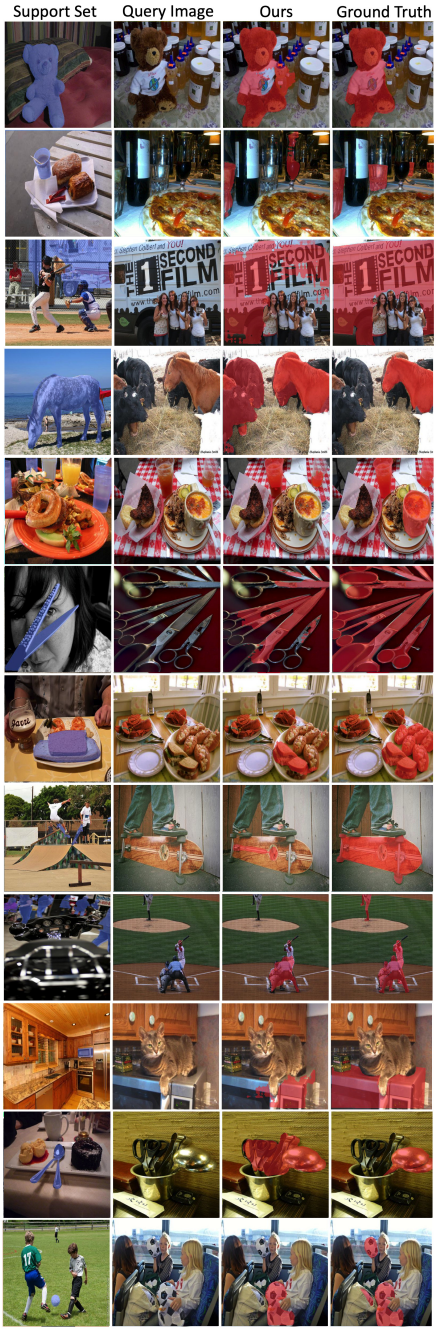

Fig. 6: HSNet-HM Failure cases on COCO-20^i [4]

References

1. Everingham, M., Van Gool, L., Williams, C., Winn, J., Zisserman, A.: The pascal visual object classes (voc) challenge. International Journal of Computer Vision **88**, 303–338 (06 2010). https://doi.org/10.1007/s11263-009-0275-4
2. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778. IEEE Computer Society (2016). https://doi.org/10.1109/CVPR.2016.90, https://doi.org/10.1109/CVPR.2016.90
3. Li, X., Wei, T., Chen, Y.P., Tai, Y.W., Tang, C.K.: Fss-1000: A 1000-class dataset for few-shot segmentation. CVPR (2020)
4. Lin, T.Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C.L., Dollár, P.: Microsoft coco: Common objects in context (2015)
5. Min, J., Kang, D., Cho, M.: Hypercorrelation squeeze for few-shot segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 6941–6952 (October 2021)