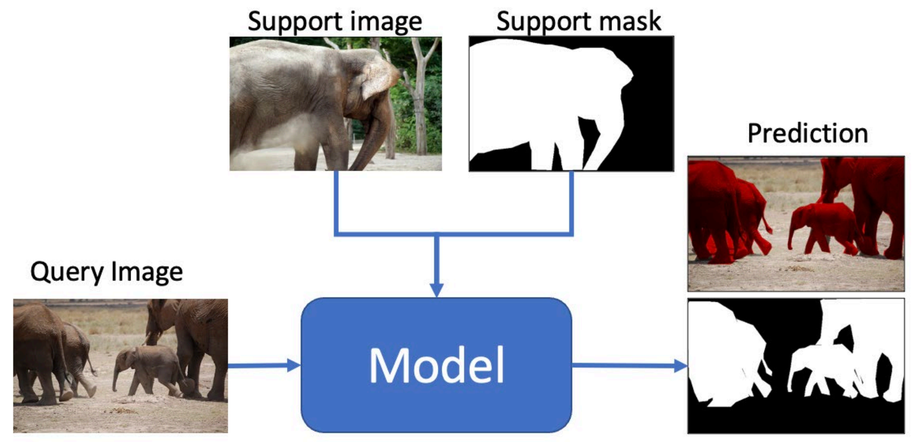
HM : Hybrid Masking for Few-Shot Segmentation

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

1: sm2062@cs.rutgers.edu

Problem

- The goal of few-shot segmentation is to train a model that can identify the target object in a query image with only few annotated samples (Support set).



Contribution

- Propose a simple, effective, and efficient way to enhance the prevalent feature masking technique(FM [1]) with input masking(IM [2]).
- HM with HSNet delivers up to 5.3% gain in mIoU and speeds up its training convergence by around 11x times on average on COCO-20i.

Motivation

- Feature Masking (FM [1]) was widely adopted to remove background from features.

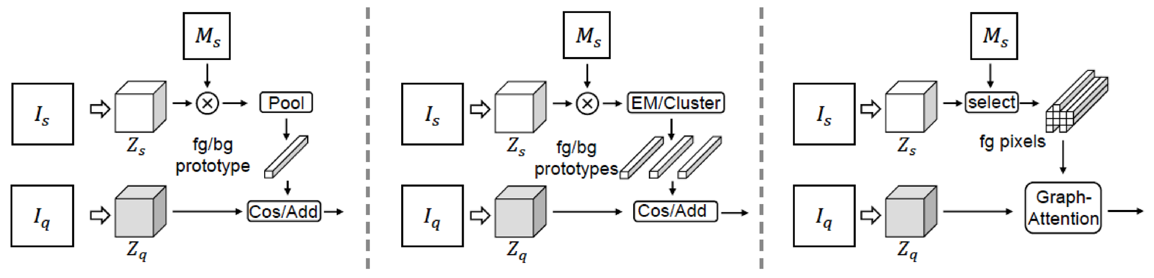
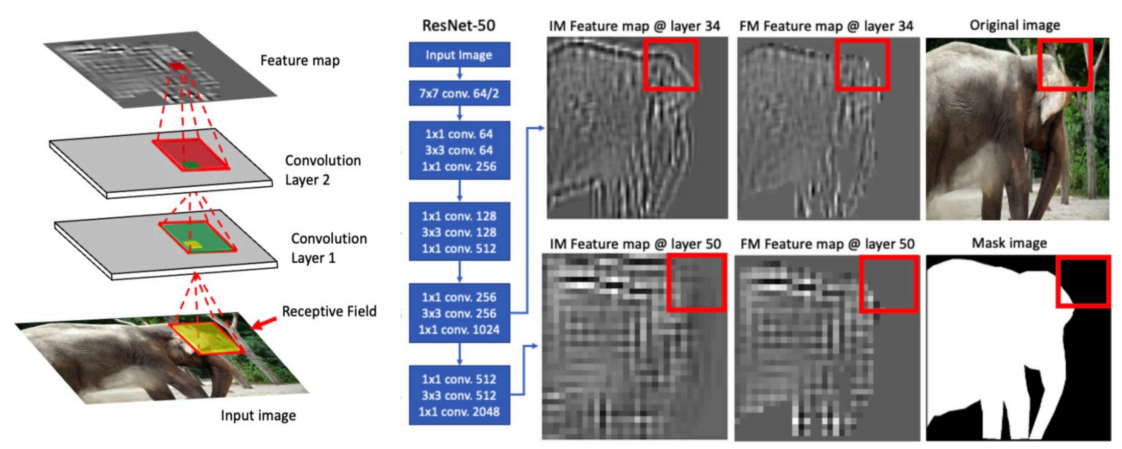
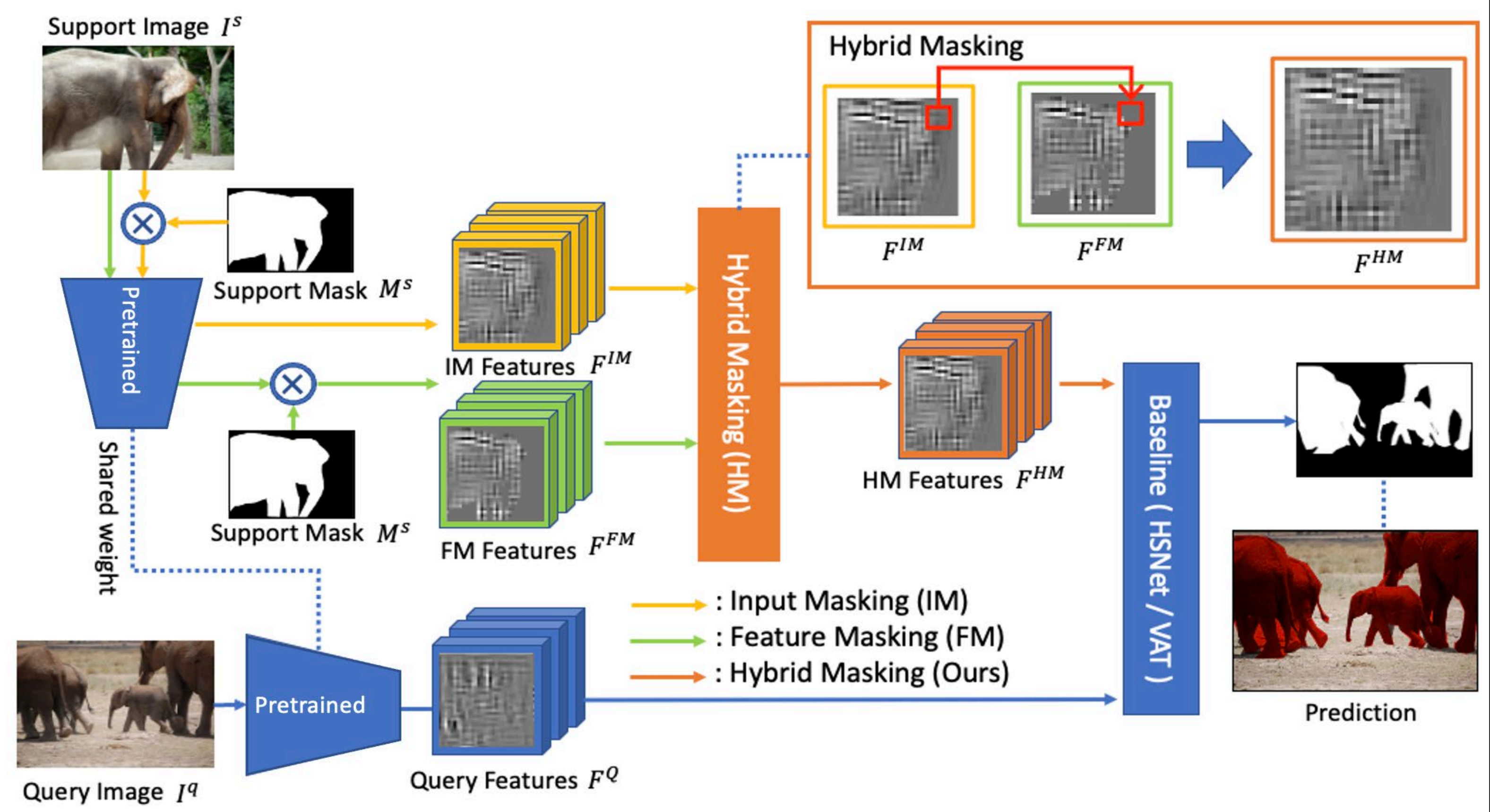


Image from CyCTR [4]

- FM loses useful information through its masking and progressively worsens with deeper layers.



Proposed Method



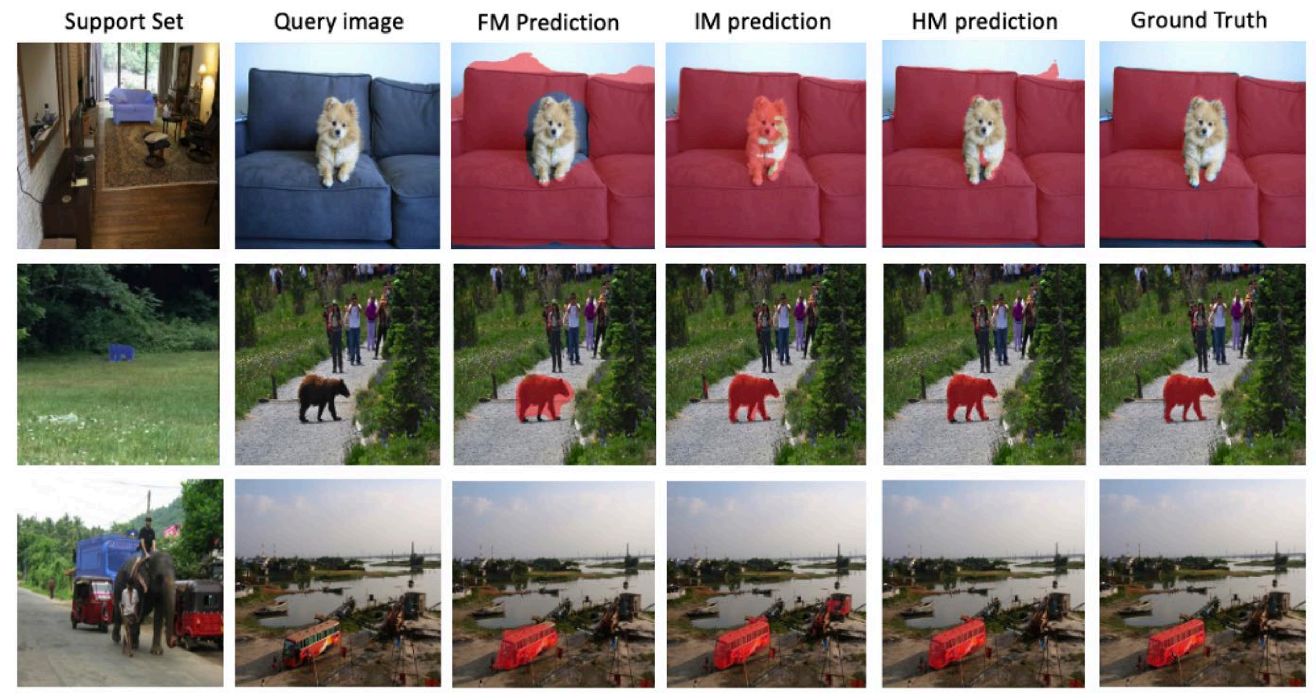
Algorithm : Hybrid Masking

Input : IM feature maps F^{IM} and FM features maps F^{FM}
Each channel i, $f_i^{IM} \in F^{IM}$ and $f_i^{FM} \in F^{FM}$
for i =1, . . . , c **do**
  Set $f_i^{HM} = f_i^{FM}$
  for *Entire pixels* $\in f_i^{HM}$ **do**
    Find an inactive pixel, $p \in f_i^{HM}$
    if $p \leq 0$ **then**
      Replace the pixel, p, with corresponding pixel $\in f_i^{IM}$
    end
  end
end
Output: HM feature maps F^{HM}

1. FM and IM features are computed according to the existing methods.
2. The inactivated values in the FM features are then replaced with IM features.

Analysis

Comparison on three masking techniques on COCO-20i



- FM [1] fails to precisely recover target details, such as target boundaries.
- IM [2] struggles in distinguishing objects from the background.
- HM clearly distinguishes between the target objects and the background and recovers precise details such as, target boundaries.

Results

Performance comparison on PASCAL-5i

Backbone feature	Methods	1-shot 5^0	5^1	5^2	5^3	mIoU	FB-IoU	5-shot 5^0	5^1	5^2	5^3	mIoU	FB-IoU
ResNet50 [8]	RePRI [1]	59.8	68.3	62.1	48.5	59.7	-	64.6	71.4	**71.1**	59.3	66.6	-
	CyCTR [41]	67.8	**72.8**	58.0	58.0	64.2	-	71.1	73.2	60.5	57.5	65.6	-
	HSNet [24]	64.3	70.7	60.3	60.5	64.0	76.7	70.3	73.2	67.4	**67.1**	69.5	80.6
	HSNet*	63.5	70.9	61.2	60.6	64.3	**78.2**	70.9	73.1	68.4	65.9	69.6	80.6
	VAT [9]	67.6	71.2	**62.3**	60.1	65.3	77.4	72.4	**73.6**	68.6	65.7	**70.0**	**80.9**
	HSNet*-HM	**69.0**	70.9	59.3	61.0	65.0	76.5	69.9	72.0	63.4	63.3	67.1	77.7
	VAT-HM	68.9	70.7	61.0	**62.5**	**65.8**	77.1	**71.1**	72.5	62.6	66.5	68.2	78.5
ResNet101 [8]	RePRI [1]	59.6	68.6	62.2	47.2	59.4	-	66.2	71.4	67.0	57.7	65.6	-
	CyCTR [41]	69.3	**72.7**	56.5	58.6	64.3	72.9	73.5	74.0	58.6	60.2	66.6	75.0
	HSNet [24]	67.3	72.3	62.0	63.1	66.2	77.6	71.8	74.4	67.0	68.3	70.4	80.6
	HSNet*	67.5	**72.7**	63.5	63.2	66.7	77.7	71.7	74.8	68.2	68.7	70.8	80.9
	VAT [9]	68.4	72.5	**64.8**	64.2	67.5	78.8	73.3	75.2	68.4	**69.5**	**71.6**	**82.0**
	HSNet*-HM	69.8	72.1	60.4	64.3	66.7	77.8	72.2	73.3	64.0	67.9	69.3	79.7
	VAT-HM	**71.2**	**72.7**	62.7	**64.5**	**67.8**	**79.4**	**74.0**	**75.5**	65.4	68.6	70.9	81.5

Performance comparison on COCO-20i

Backbone feature	Methods	1-shot 20^0	20^1	20^2	20^3	mIoU	FB-IoU	5-shot 20^0	20^1	20^2	20^3	mIoU	FB-IoU
ResNet50 [8]	RePRI [1]	32.0	38.7	32.7	33.1	34.1	-	39.3	45.4	39.7	41.8	41.6	-
	HSNet [24]	36.3	43.1	38.7	38.7	39.2	68.2	43.3	51.3	48.2	45.0	46.9	70.7
	CyCTR [41]	38.9	43.0	39.6	39.8	40.3	-	41.1	48.9	45.2	47.0	45.6	-
	VAT [9]	39.0	43.8	42.6	39.7	41.3	68.8	44.1	51.1	50.2	46.1	47.9	72.4
	ASNet [11]	41.5	44.1	42.8	40.6	42.2	69.4	**48.0**	52.1	49.7	48.2	**49.5**	**72.7**
	HSNet-HM	41.0	45.7	**46.9**	43.7	44.3	**70.8**	45.3	**53.1**	**52.1**	47.0	49.4	72.2
	VAT-HM	42.2	43.3	45.0	42.2	43.2	70.0	45.2	51.0	50.7	46.4	48.3	71.8
	ASNet-HM	**42.8**	**46.0**	44.8	**45.0**	**44.7**	70.4	46.3	50.2	48.4	**48.6**	48.4	72.2
ResNet101 [8]	FWB [25]	17.0	18.0	21.0	28.9	21.2	-	19.1	21.5	23.9	30.1	23.7	-
	DAN [35]	-	-	-	-	24.4	62.3	-	-	-	-	29.6	63.9
	PFENet [33]	36.8	41.8	38.7	36.7	38.5	63.0	40.4	46.8	43.2	40.5	42.7	65.8
	HSNet [24]	37.2	44.1	42.4	41.3	41.2	69.1	45.9	53.0	51.8	47.1	49.5	72.4
	ASNet [11]	41.8	45.4	43.2	41.9	43.1	69.4	**48.0**	52.1	49.7	48.2	49.5	72.7
	HSNet-HM	41.2	**50.0**	**48.8**	45.9	**46.5**	**71.5**	46.5	**55.2**	51.8	48.9	**50.6**	72.9
	ASNet-HM	**43.5**	46.4	47.2	**46.4**	45.9	71.1	47.7	51.6	**52.1**	**50.8**	**50.6**	**73.3**

Performance comparison on FSS-1000

Backbone feature	Methods	mIoU 1-shot	5-shot	Backbone feature	Methods	mIoU 1-shot	5-shot
ResNet50 [8]	FSOT [18]	82.5	83.8	ResNet101 [8]	DAN [35]	85.2	88.1
	HSNet [24]	85.5	87.8		HSNet [24]	86.5	88.5
	VAT [9]	**89.5**	**90.3**		VAT [9]	90.0	**90.6**
	HSNet-HM	87.1	88.0		HSNet-HM	87.8	88.5
	VAT-HM	89.4	89.9		VAT-HM	**90.2**	90.5

Number of best 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**

Visual Comparison with HSNet [3]

Conclusion

- We proposed a new effective masking approach, termed as hybrid masking. It aims to enhance the feature masking (FM [1]) technique, that is commonly used in existing SOTA methods.
- We instantiate HM in strong baselines and the results reveal that utilizing HM surpasses HSNet [3] by visible margins in mIoU (on average 0.4% on PASCAL and 5% on COCO) and reduces training time by a factor of 11x on average.

Reference

[1] Zhang, X., Wei, Y., Yang, Y., Huang, T.: Sg-one: Similarity guidance network for one-shot semantic segmentation. IEEE Transactions on Cybernetics 50, 3855–3865 (2020)
[2] Shaban, A., Bansal, S., Liu, Z., Essa, I., Boots, B.: One-shot learning for semantic segmentation. Proceedings of the British Machine Vision Conference (BMVC 2018).
[3] 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)
[4] Zhang, G., Kang, G., Yang, Y., Wei, Y.: Few-shot segmentation via cycle-consistent transformer (NIPS 2021)

ECCV EUROPEAN CONFERENCE ON COMPUTER VISION TEL AVIV 2022
October 23-27, 2022, Tel Aviv

RUTGERS
TCNJ THE COLLEGE OF NEW JERSEY
MBZUAI