Masked Supervised Learning for Semantic Segmentation

Hasib Zunair and A. Ben Hamza

Motivation

- Existing semantic segmentation methods overly focus on attention-based methods to model long-range context.
- Does not explicitly leverage context (self-supervised masked autoencoders, two stage training)
- We ask:
  - Is short range useful?
  - Does it work well with attention-based methods?

Failure Cases

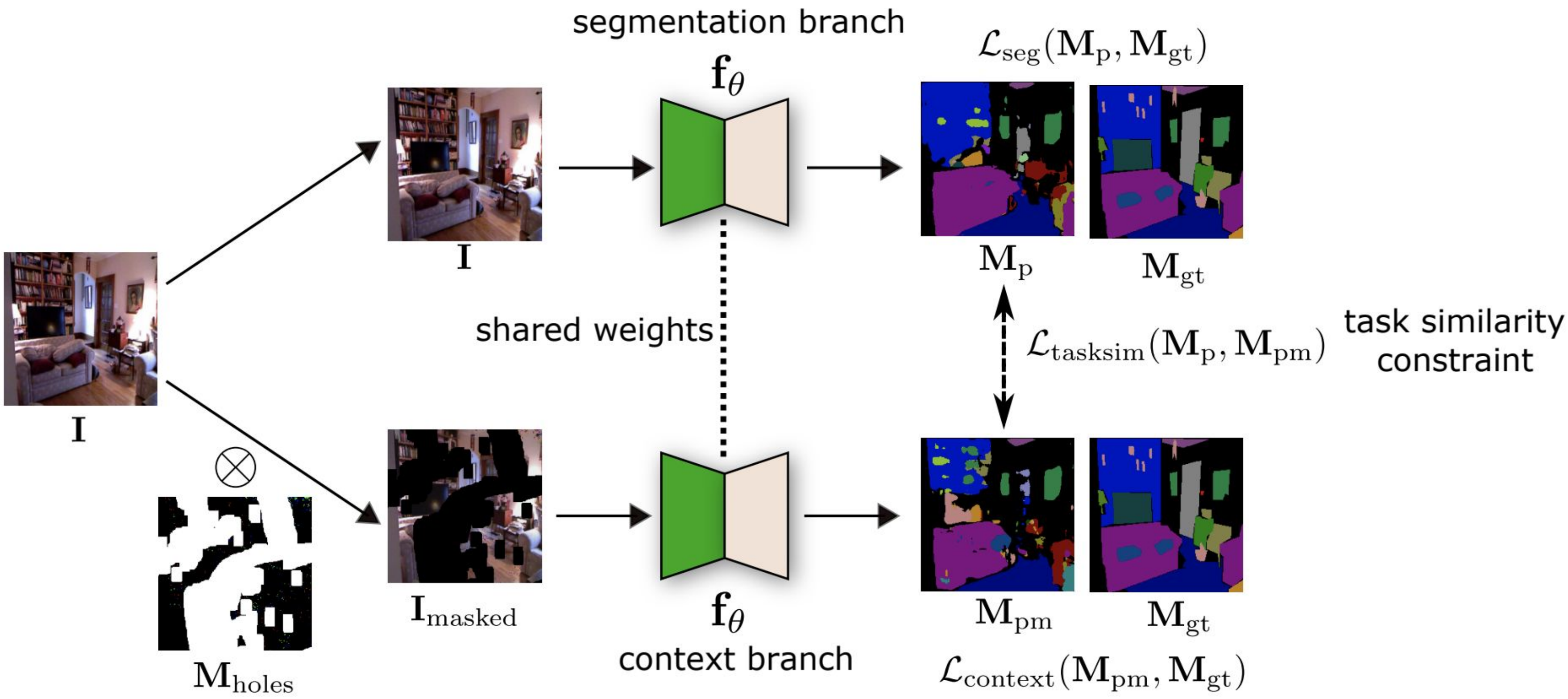
- Over-segment regions of interest (ROIs)
- Noisy and discontinuous predictions
- Fail to predict boundary regions
- Poorly segment minority classes & misclassify in multi-class semantic segmentation

These failure cases lead to poor segmentation performance.

Université Concordia University
BMVC 2022
DECATHLON

Masked Supervised Learning for Semantic Segmentation

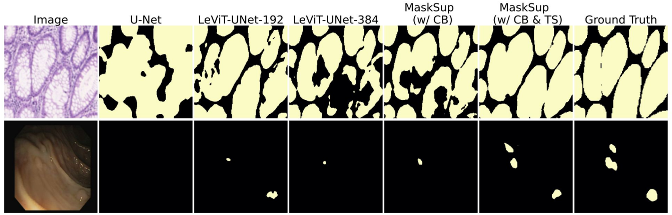
Hasib Zunair and A. Ben Hamza

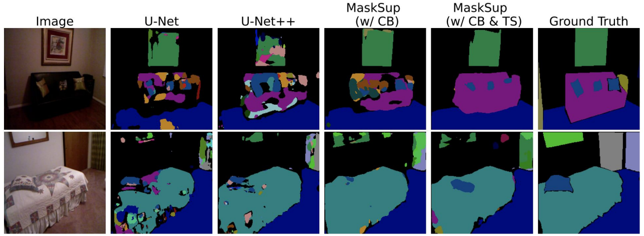


$$\mathcal{L}_{\mathrm{total}} = \alpha_1 \mathcal{L}_{\mathrm{seg}}(\mathbf{M}_{\mathrm{p}}, \mathbf{M}_{\mathrm{gt}}) + \alpha_2 \mathcal{L}_{\mathrm{context}}(\mathbf{M}_{\mathrm{pm}}, \mathbf{M}_{\mathrm{gt}}) + \alpha_3 \mathcal{L}_{\mathrm{tasksim}}(\mathbf{M}_{\mathrm{p}}, \mathbf{M}_{\mathrm{pm}})$$

Masked Supervised Learning for Semantic Segmentation

Hasib Zunair and A. Ben Hamza





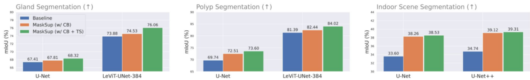
MaskSup can better i) segment both natural and medical images, ii) model shape of small ROI iii) segment minority classes

Method	GLaS, mIoU (↑)	CVC-Clinic-DB, mIoU (↑)	NYUDv2 (↑)
U-Net [18]	67.41	69.74	33.60
FCN [15]	50.84	-	29.20
U-Net++[30]	69.10	72.90	34.74
HRNet-18 [26]	-	-	33.18
ResU-Net [27]	65.95	-	-
ResU-Net++ [9]	-	79.60	-
SFA [7]	-	60.70	-
Attention U-Net [17]	-	82.70	-
Axial Attention U-Net [25]	63.03	-	-
MedT [21]	69.61	-	-
KiU-Net [22]	72.78	-	-
LeViT-UNet-128 [29]	70.45	-	-
LeViT-UNet-192 [29]	71.83	79.16	-
LeViT-UNet-384 [29]	<u>73.88</u>	<u>81.38</u>	-
PAD-Net [28] △	-	-	33.10
HybridNet A2 [13] △	-	-	34.30
MTI-Net [23] △	-	-	<u>37.49</u>
MaskSup (**Ours**)	**76.06**	**84.02**	**39.31**

MaskSup outperforms baselines on three datasets

Method	GLaS, mIoU (↑)	CVC-Clinic-DB, mIoU (↑)	NYUDv2 (↑)
MAE [8]	75.04	82.50	37.42
MaskSup (**Ours**)	**76.06**	**84.02**	**39.31**

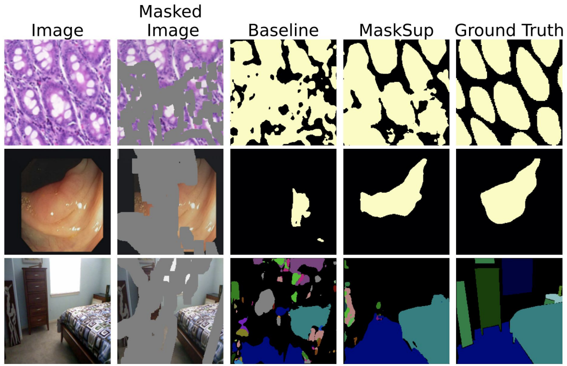
MaskSup outperforms MAE in mIOU and is more efficient



CB and TS both improve performance of different architectures

Masking	GLaS, mIoU (↑)	CVC-Clinic-DB, mIoU (↑)	NYUDv2 (↑)
Low	75.65	81.80	35.33
High	**76.06**	**84.02**	**39.31**

Heavy masking works better in MaskSup!



MaskSup can segment regions even when input is heavily masked (shape aware) 🤯

Segments **small** ROIs (short-range context)

Segments **large** ROIs (long-range context)

Method	Params (M) (↓)	GLaS, mIoU (↑)	CVC-Clinic-DB, mIoU (↑)	NYUDv2 (↑)
LeViT-384 [29]	51	73.88	81.38	-
MaskSup (LeViT-192)	**19**(2.6x)	**74.44**(+0.75)	**82.17**(+0.97)	-
U-Net++ [30]	9	-	-	34.74
MaskSup (U-Net)	**3**(3x)	-	-	**38.54**(+10.91)

MaskSup is computationally efficient and achieves superior performance with fewer parameters.