Adversarial Scene Editing: Automatic Object Removal from Weak Supervision

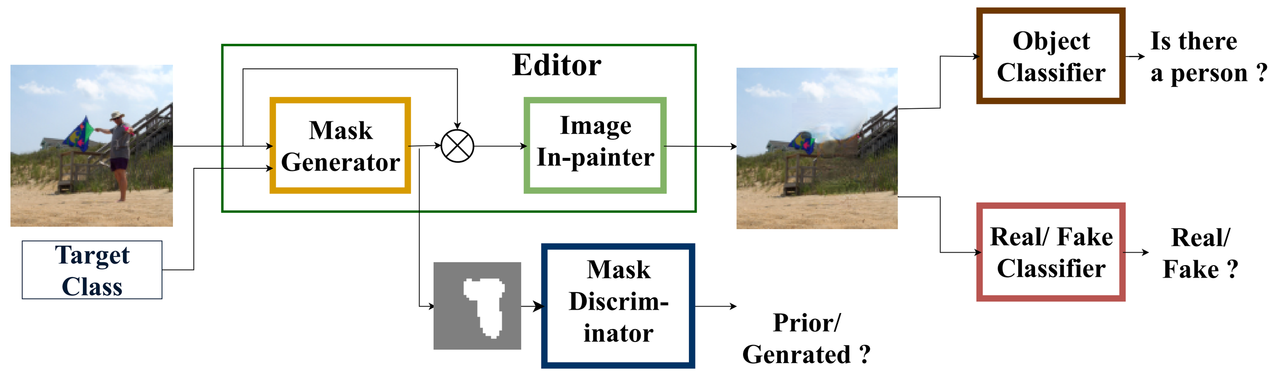
Rakshith Shetty[1,3] Mario Fritz[2,3] Bernt Schiele[1,3]

[1]Max Planck Institute for Informatics [2]CISPA Helmholtz Center i.G. [3]Saarland Informatics Campus

Contact: rshetty@mpi-inf.mpg.de bit.ly/asecode

Summary

Task: Learning to remove desired object class from input images with only weak image-level labels
Why: Test-case generation, data augmentation and privacy filters



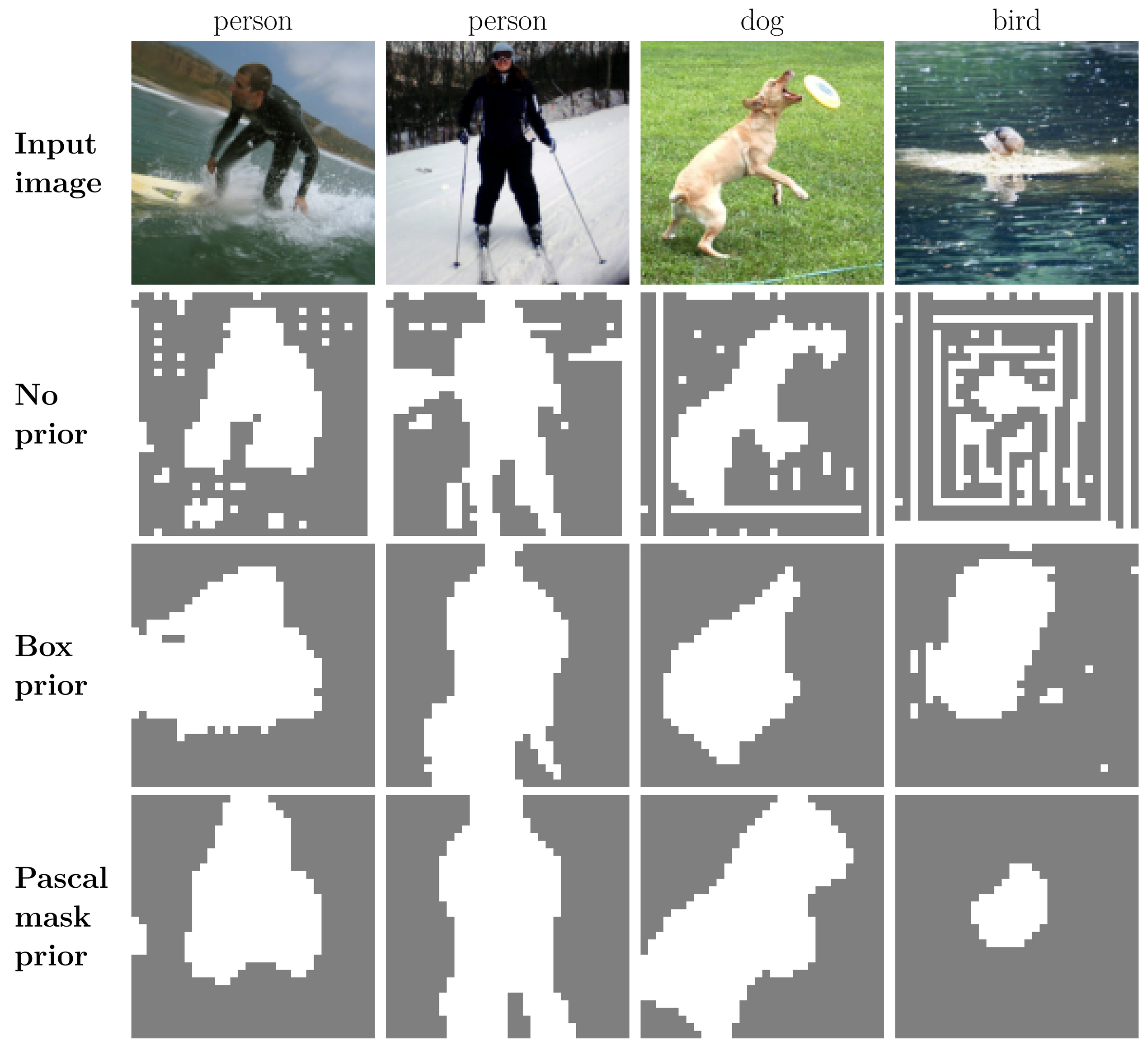
Two-stage editor network to remove objects from input image

- The two-stage architecture with jointly trained **mask generator** and **in-painter** helps avoid adversarial perturbations.
- **Mask generator** learns by fooling the **object classifier** and the **mask discriminator**.
- **In-painter** learns by fooling the **real-fake classifier**.
- Adversarial shape prior to incorporate the knowledge of object shapes.

Learns to remove objects using supervision from image labels and unpaired shape priors. Achieves removal success rate on-par with fully supervised Mask-RCNN.

Qualitative results of removal

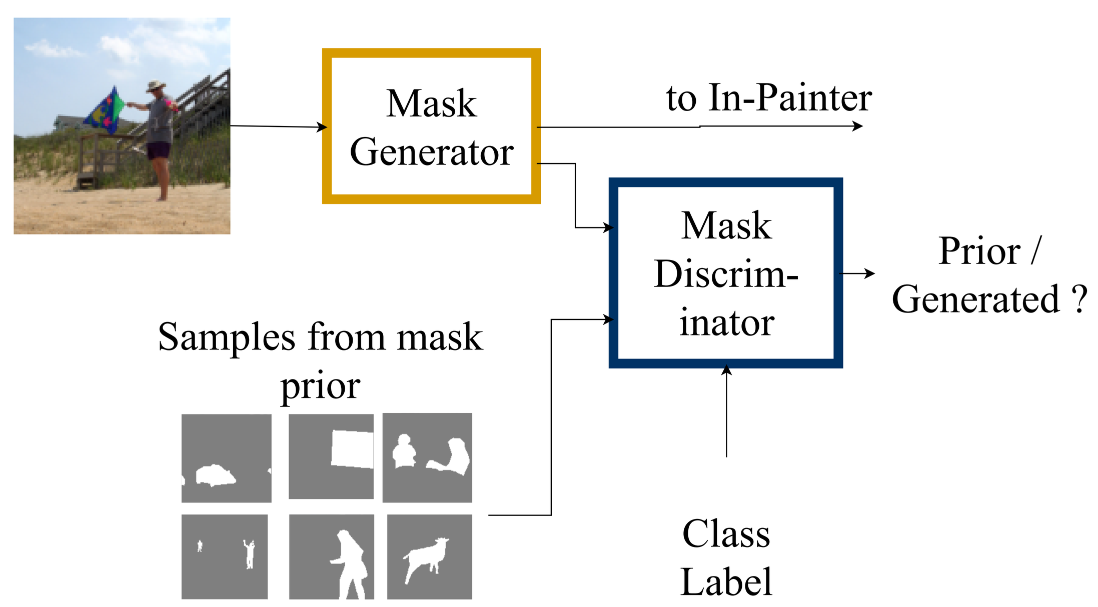
Effect of adversarial shape prior



Using more accurate priors (top to bottom) leads to improved mask generation which better adheres to the object shape.

Model

Training the mask generator



- **Mask generator** learns with combination of three losses, $L(G_M)$

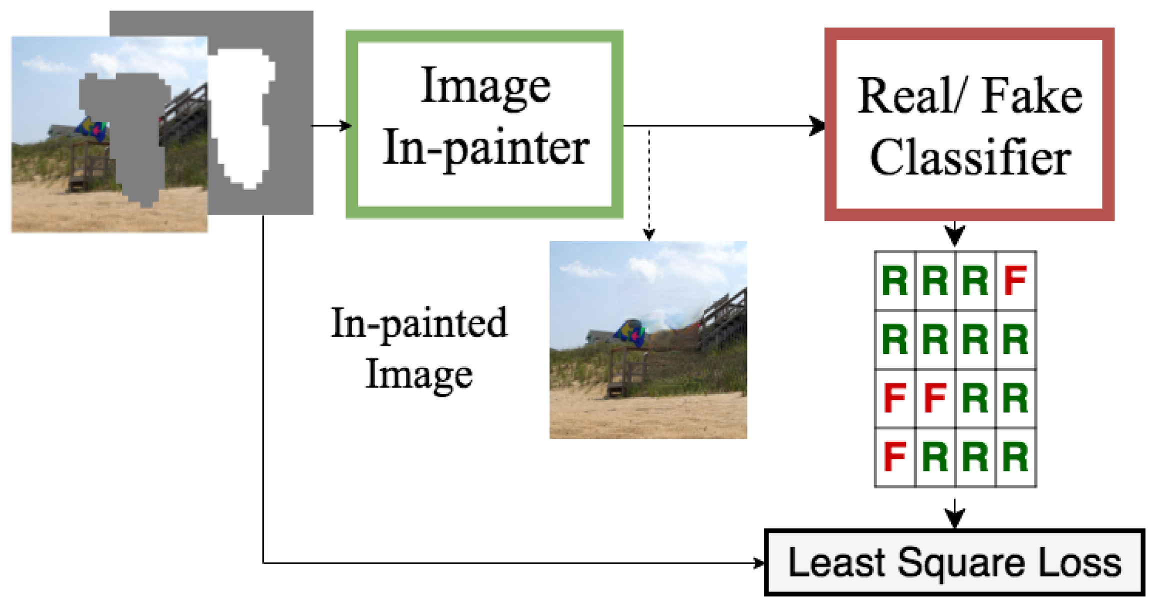
$$m = G_M(x, c_t)$$

$$y = (1 - m) \cdot x + m \cdot G_I\left((1 - m) \cdot x\right)$$

$$L(G_M) = \underbrace{-\mathbb{E}_x\left[\log(1 - D_{cls}(y, c_t))\right]}_{\text{Fool object classifier}} \underbrace{- \mathbb{E}_x\left[D_M(m, c_t)\right]}_{\text{Adversarial prior}} + \underbrace{\exp(\Sigma\, m_{ij})}_{\text{Size regularization}}$$

- Adversarial shape prior offers flexibility. We experiment with randomly generated rectangles and unpaired masks.

Training the in-painter



- **In-painter** learns with a combination of reconstruction loss, adversarial real-fake loss and style loss [1].
- Local real-fake classifier is trained with mask as the label.

$$L(D_{\text{rf}}) = \underbrace{\frac{\Sigma_{ij}(1 - m_{ij})(D_{\text{rf}}(y)_{ij} - 1)^2}{\Sigma_{ij}(1 - m_{ij})}}_{\text{Real pixels}} + \underbrace{\frac{\Sigma_{ij} m_{ij} \cdot (D_{\text{rf}}(y)_{ij} + 1)^2}{\Sigma_{ij} m_{ij}}}_{\text{Edited pixels}}$$

- Local real-fake loss focuses discriminator to the edited region.

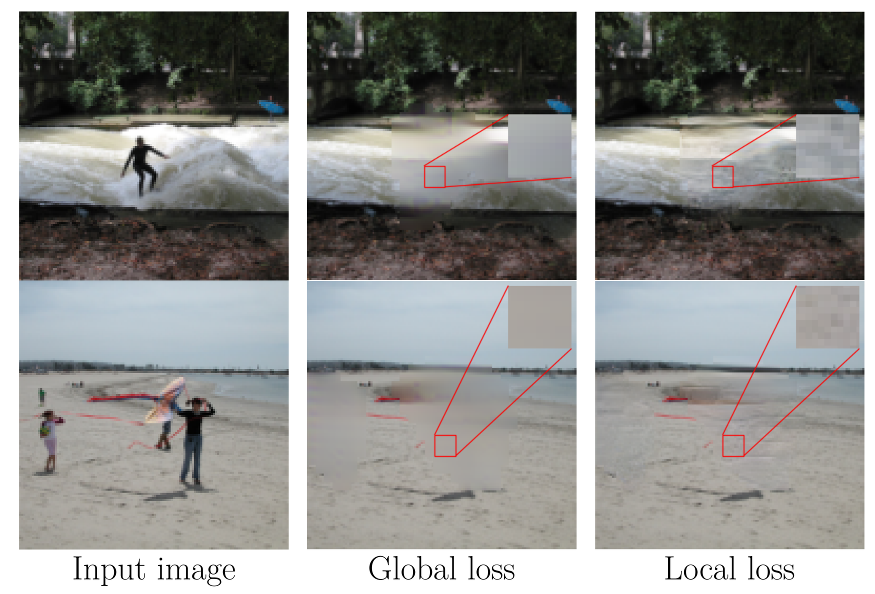
Quantitative results and ablations

Model	Supervision	Removal Performance			Image quality metrics		
		removal success ↑		false ↓	perceptual loss ↓	pSNR ↑	ssim ↑
		all	person	removal			
GT masks	-	66	72	5	**0.04**	**27.43**	**0.930**
Mask RCNN	Segmentation masks & bounding boxes	68	73	6	0.05	25.59	0.900
Mask RCNN (dil. 7x7)		75	77	10	0.07	24.13	0.882
Ours-pascal	image labels & unpaired masks	73	**81**	16	0.08	22.64	0.803

Quantitative comparison to supervised segmentation based removal Our weakly supervised model trained with pascal mask prior achieves on-par removal success rate as the fully supervised Mask-RCNN based object removal.

Prior	Removal Performance			Perceptual loss ↓	Mask accuracy (mIoU ↑)
	removal success ↑		false ↓		
	all	person	removal		
None	**94**	**96**	36	0.13	0.15
boxes	83	88	23	0.11	0.18
pascal (10)	67	59	17	**0.07**	**0.23**
pascal (100)	70	75	**16**	**0.07**	0.22
pascal (all)	73	81	**16**	0.08	0.22

Better shape priors improve the mask accuracy and reduce false removal. Moving down the table from the no prior case to the box priors and then to the class specific shape priors from the Pascal dataset masks the masks smaller, improves the mIoU and also reduces the false removal rate.



Qualitative comparison of global vs local loss. Local real-fake loss improves the in-painting results producing sharper, texture-rich

References

[1] G. Liu, F. A. Reda, K. J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro, "Image inpainting for irregular holes using partial convolutions," *arXiv*, 2018.

[2] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *ICCV*, 2017.