Implicit Representations for Constrained Image Segmentation

Jan Philipp Schneider[1] Mishal Fatima[1] Jovita Lukasik[1]
Andreas Kolb[1] Margret Keuper[1,2,3] Michael Moeller[1]

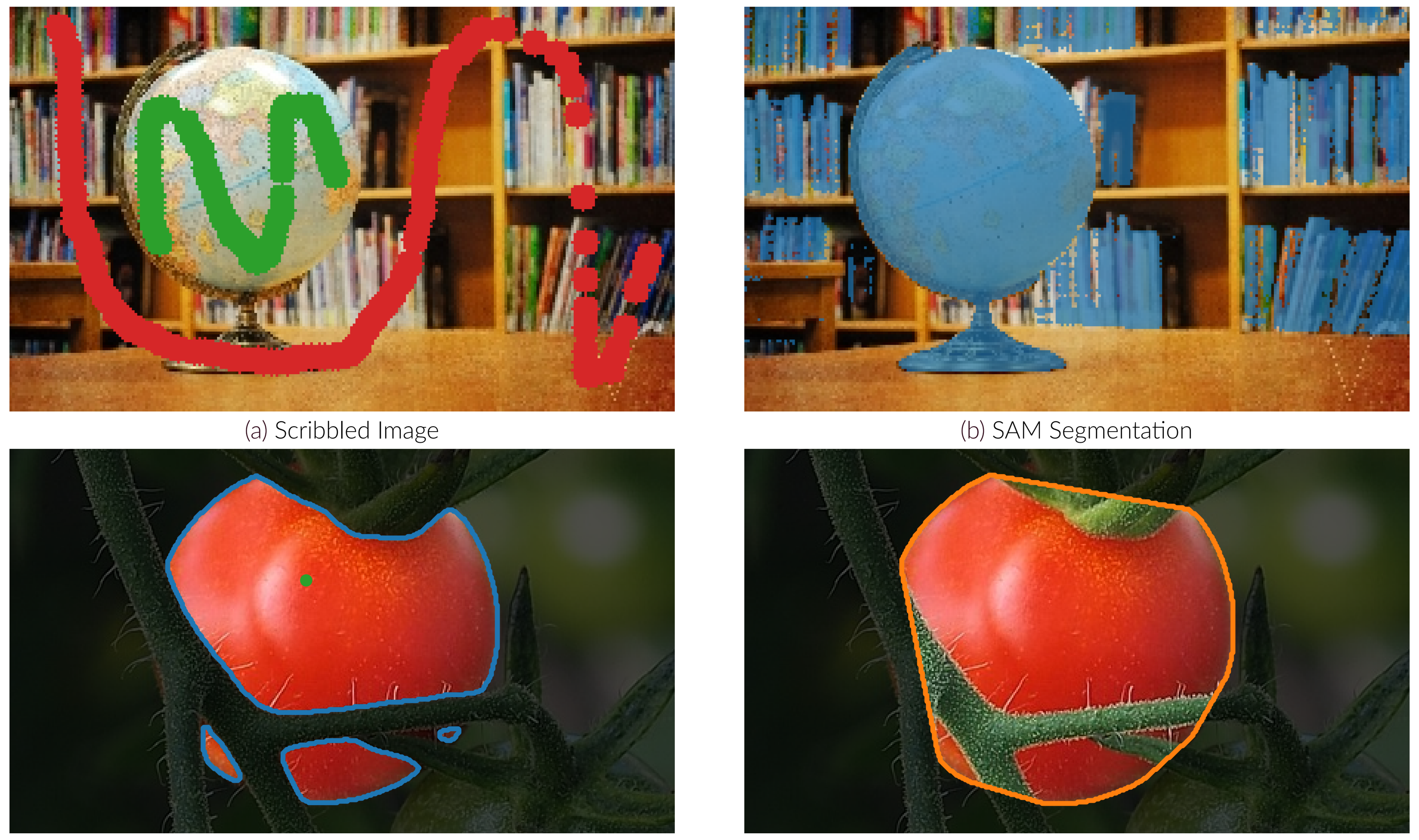
[1] University of Siegen [2] University of Mannheim [3] Max-Planck-Institute for Informatics, Saarland Informatics Campus

Motivation

Segmentation **Foundation Models**, like Segment Anything (SAM) [2], can still **fail if**:

- Data is scarce
- Objects are occluded
- Examples are out-of-distribution

Also: The prior knowledge of geometric properties is rarely considered in image segmentation
But: **Provably** ensuring constraints is hard and costly



(a) Scribbled Image (b) SAM Segmentation
(c) Occluded Tomato (d) Convex Hull of Tomato

Figure 1. Segmenting a scribbled image (a). with SAM (b) leads to a scattered result. The partial occlusion of a tomato (c) leads to a disjointed segmentation in SAM, while for volume estimation, a volume-preserving segmentation (d) would be advantageous.

Can provable geometric constraints improve segmentation quality?

Proposal

Implicit Segmentation Representations (ISR)

- Mapping spatial coordinates to fore- or background
- Foreground is **provably** assured to have one of the properties:
  Convex, Star-Shaped, Mirror & Rotational Symmetric or Path-Connected

Regularization for Shapes

- Using a input convex neural network (ICNN) [3], different input coordinate systems or diffeomorphism + ICNN
- Goes with any prediction architecture: Constrained ISR is used as a regularizer
- Capable of joint optimization

Figure 2. Naïve segmentations (blue) and learned convex, star-shaped, mirror-symmetric, and path-connected ISRs (orange).

Method

Constraint ISR

- Represent segmentations as a *function* $\mathcal{G}_\nu(x) : \mathbb{R}^2 \to \mathbb{R}$ *implicitly* via a neural network
  $\Rightarrow \mathcal{G}_\nu(x)$ maps every pixel to its foreground likeliness
- Operating in the (image) coordinate system $x = c(f), x \in \mathbb{R}^2$ for an image $f \in \mathbb{R}^{n_y \times n_x \times 3}$
- Optional: coordinate transform function $c : \Omega \subset \mathbb{R}^2 \to \mathbb{R}^2$ mapping spatial image domain Ω
- For complete definitions and proofs of all ISRs we refer to our paper [4].

Project or Regularize

Given:

- Segmentation predictor $\mathcal{N}_\theta(f) \in [0,1]^{n_y \times n_x}$; can be any neural network, or simple unaries
- One of our proposed constrained ISRs $\mathcal{G}_\nu$

The output of $\mathcal{N}_\theta$ can be *projected* on $\mathcal{G}_\nu$ by:

$$\text{dist}(\mathcal{N}_\theta(f), S) = \min_\nu \|\sigma(\mathcal{N}_\theta(f)) - \sigma(\mathcal{G}_\nu(c(f)))\| \tag{1}$$

for S as the set of functions represented by $\mathcal{G}_\nu$, and σ beeing the sigmoid function.
Further: (1) can be used as a **regularizer** during training of $\mathcal{N}_\theta \Rightarrow$ optimizing θ and ν *jointly*

Numerical Experiments

- Convex ISR is investigated using a scribble-based convexity dataset and two simple architectures for $\mathcal{N}_\theta$ [1]:
  1. Convolutional Neural Network (CNN)
  2. Fully Connected Network (FCN)
- Using RGB, spatial, and/or semantic input features per pixel

Table 1. Intersection over union (IoU) of foreground objects with predictors $\mathcal{N}_\theta$ and our proposed convex ISR

	RGB+semantic		RGB+spatial+semantic	
	CNN / convex	FCN / convex	CNN / convex	FCN / convex
seq.	0.726 / **0.843**	0.714 / **0.851**	**0.778** / 0.766	0.736 / **0.746**
joint	0.818 / **0.899**	0.635 / **0.894**	0.805 / **0.809**	0.768 / **0.769**

(a) Scribbled Image (b) Segmentation (c) Seq. Projection (d) Joint Projection

Figure 3. Results of the FCN architecture, trained on scribbles with RGB + semantic features.

- Path-connected ISR was evaluated on the FBMS-59 dataset
- Primary predictor $\mathcal{N}_\theta$ is a UNet with training scheme proposed by [5]
- Sequential fit yields 1 %, joint projection 2%, IoU increase over baseline across all image sequences
- Path-Connected ISR is particularly useful when $\mathcal{N}_\theta$ segments multiple objects. In joint training, these disappear completely, see Fig. 4.

Figure 4. Left: Sequential prediction (blue) and projection (orange), Right: Both, after joint training. ISR enforces $\mathcal{N}_\theta$ to focus on the right car and accuracy increases significantly.

Numerical Experiments Cont.

- Constraint ISRs are not limited to spatial image dimensions
- The input x of $\mathcal{G}_\nu$ can be configured to include further features like depth or time

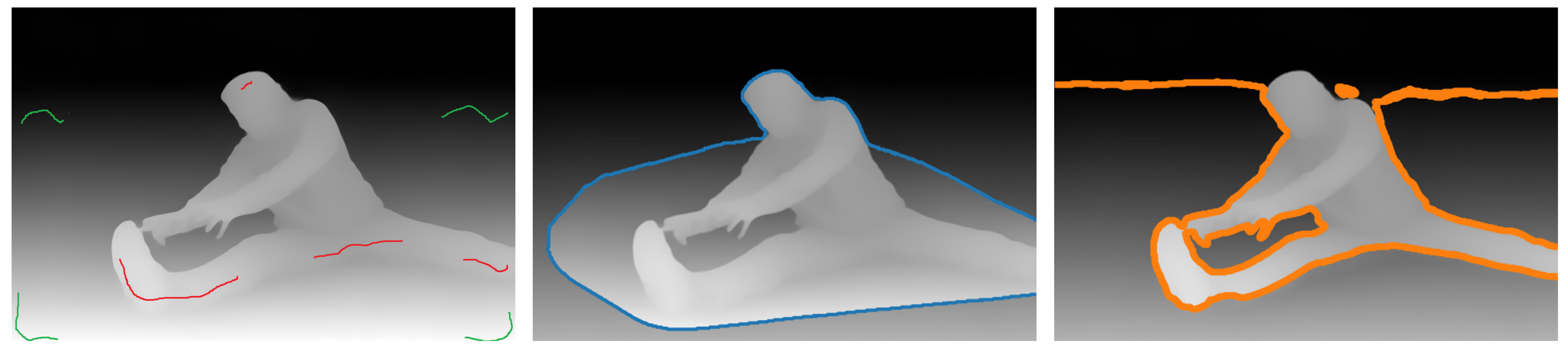


Figure 5. Illustrating of a naïve segmentation (blue) and convex ISR (orange) in $(x,\ y,$ depth)-space. For the ISR, we segment the floor "plane", which is convex in (x, y, depth).

- Combining spatial image dimensions, time, and the path-connected ISR yields a provable spatio-temporal path-connected ISR
- Allows for accurate super-resolution in space and time (Fig. 6)
- Particularly advantageous if the unaries are noisy or the objects are occluded or covered (Fig. 7)
- Could be in-depth explored in segmentation object-trackers in the future, to increase object awareness

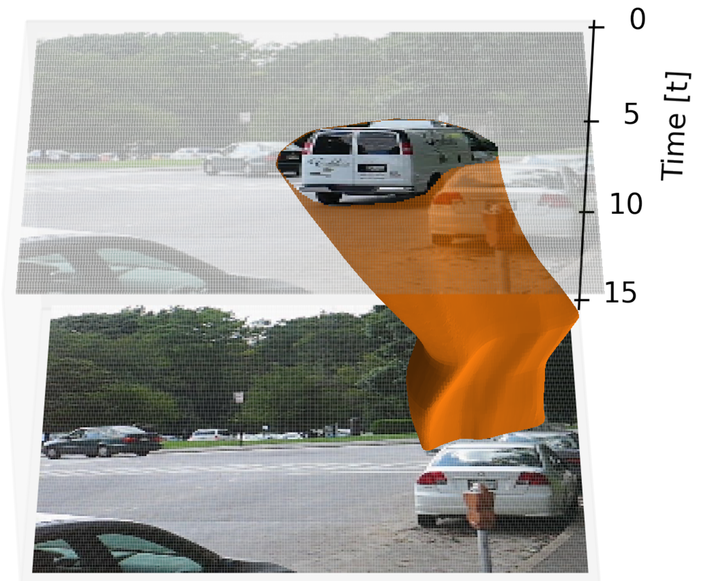


Figure 6. Spatio-temporal path-connected ISR along an image sequence. ISR was 6x super-resolved in time.

Figure 7. Spatio-temporal path-connected ISR, when trained on unaries which were partially (20 %) replaced by pure noise. The frames shown are index 10 to 14 of the *cars3* sequence.

Conclusion

- Proposed a variety of **provable** geometric constraints using *Implicit Representations* for image segmentation
- The ISRs with **implicitly enforced constraints can improve the results**:
  Especially if the missing condition is the main point of failure
- The **joint fit is superior** to a sequential one
- ISRs can be used in a variety of ways and further extended beyond segmentations in the spatial image domain

References

[1] Hannah Dröge and Michael Moeller. Learning or modelling? an analysis of single image segmentation based on scribble information. In *International Conference on Image Processing (ICIP)*, 2021.

[2] Alexander Kirillov et al. Segment anything. *CoRR*, 2023.

[3] Brandon Amos et al. Input convex neural networks. In *International Conference on Machine Learning*, 2017.

[4] Jan Philipp Schneider et al. Implicit Representations for Constrained Image Segmentation. In *International Conference on Machine Learning*, 2024.

[5] Amirhossein Kardoost and Margret Keuper. Uncertainty in minimum cost multicuts for image and motion segmentation. In *Uncertainty in Artificial Intelligence*, 2021.

The authors acknowledge support from the DFG research unit 5336 - Learning to Sense. Experiments were conducted using GPU resources of the OMNI cluster at the University of Siegen, which we were kindly given access to by the HPC team.

For the code and link to our paper, scan the QR-Code or visit: `https://github.com/jp-schneider/awesome`