Superpixel Anything: A general object-based framework for accurate yet regular superpixel segmentation

Julien Walther, Rémi Giraud, Michaël Clément

IMS Laboratory, University of Bordeaux, CNRS UMR 5218, Bordeaux INP, 33405 Talence, France

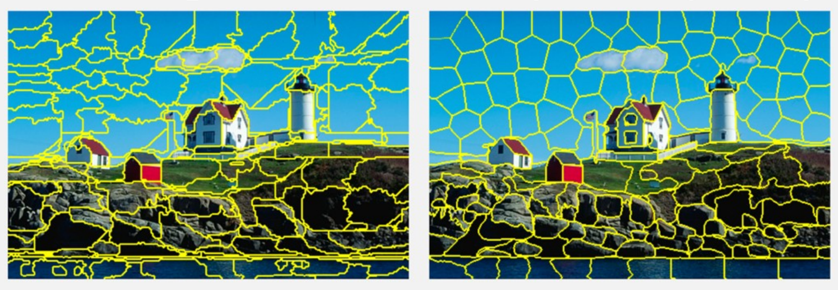
Julien.walther@u-bordeaux.fr

BMVC 2025

Introduction

- **Context and relevance of superpixels**
  - Superpixels help to reduce computational complexity by providing an irregular image under-segmentation
  - Historical trade-off between accuracy and regularity: Recent deep learning-based approaches only focus on segmentation accuracy ⇨ Lack of interpretability

How to improve both accuracy and regularity?

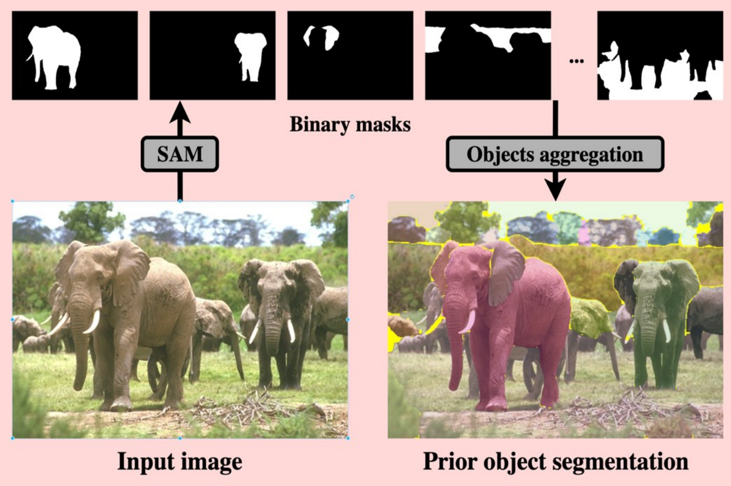


AINet [1] | **SPAM**

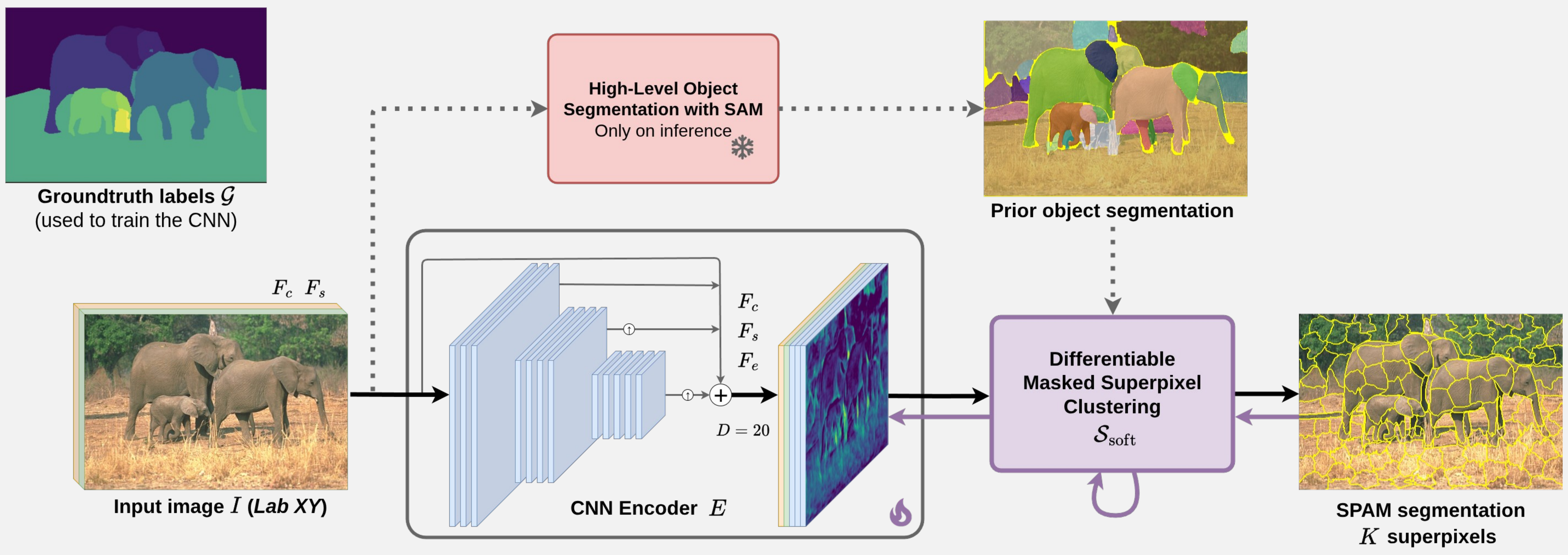
- **Contributions**
  - SPAM combines low-level and trainable high-level features for accurate superpixel clustering
  - At inference, clustering may be **guided by prior object-level maps** allowing to produce **accurate yet regular** superpixels within objects
  - **New adaptive modes** to adjust superpixel density within objects
  - SPAM delivers the most interpretable superpixels and **outperforms state-of-the-art in accuracy and regularity**

High-Level Object Segmentation with SAM

- **Objects Proposals**
  - Segment Anything Model (SAM) [2] generates independent **high-level segmentation masks**
- **Aggregation**
  - Possible mask overlaps between object proposals, leaving **unlabeled pixels**
  - **Remove overlaps** by subtracting smaller from larger objects



Method: Superpixel Anything Model (SPAM)



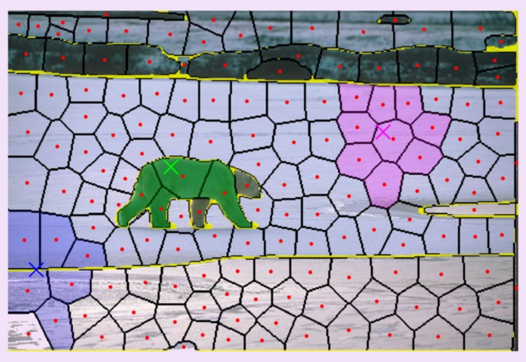
- **Training strategy**
  - Supervised on the BSD dataset $\{I, \mathcal{G}\}$
  - Loss: segmentation and regularity terms

$$\mathcal{L} = \mathcal{L}_{\text{seg}}(\mathcal{G}, \mathcal{S}_{\text{soft}}) + \lambda\,\mathcal{L}_{\text{compact}}(F_s, \mathcal{S}_{\text{soft}})$$

- **Model architecture**
  - Lightweight CNN encoder
  - Inference time < 200ms
- **Inference**
  - Prior object map with uncertainty regions can be used
  - Connexity of superpixels within objects is ensured
  - Superpixels are regular and visually interpretable

Differentiable Masked Superpixel Clustering

- **Initialisation of the grid**
  - Seeds are proportionally placed inside each object of the prior segmentation
- **Constrained K-means clustering**
  - Local candidates are contained in the same object
  - Uncertainty pixels (yellow) are not constrained

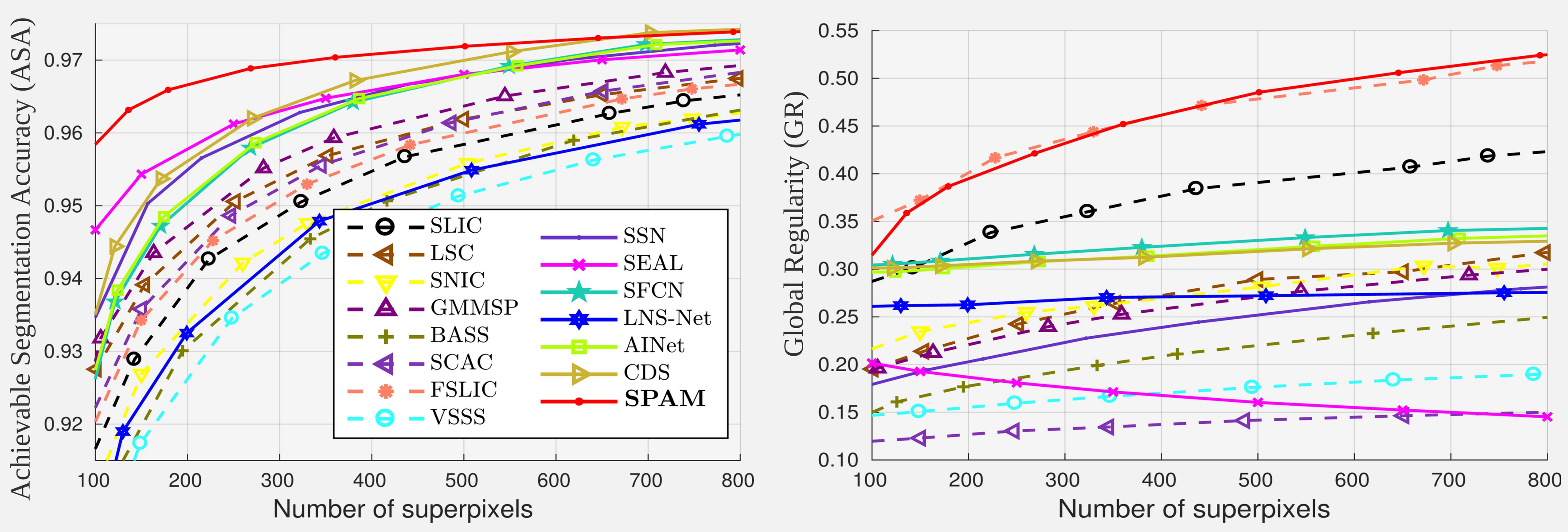


Adaptive Clustering based on Visual Attention (VA)

- The number of superpixels is increased or decreased by a **factor r in regions of interest**

Automatic visual attention using saliency map [3] | User-driven attention mode using clicks

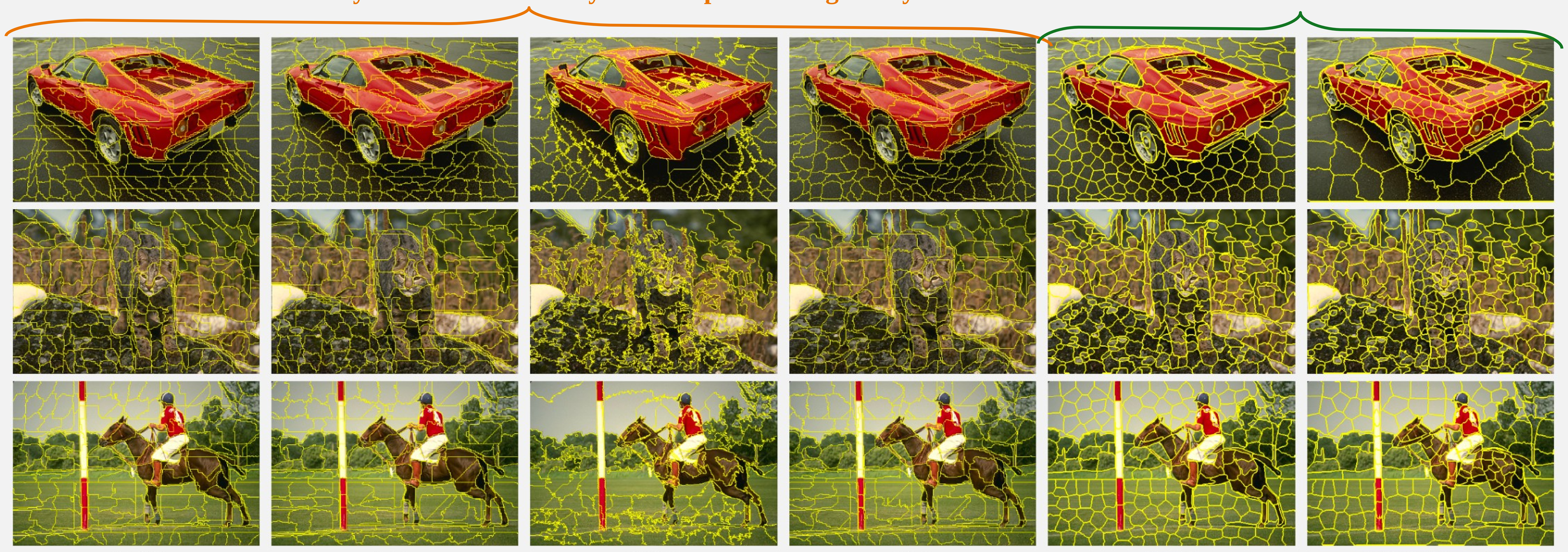
Results



- **Quantitative results**
  - SPAM outperforms SOTA in **both** segmentation accuracy and regularity
  - Same results on BSD, NYUV2 and SBD datasets
- **Qualitative results**
  - Superpixels are much more regular
  - Each one is easily interpretable

⇨ SPAM produces most accurate yet most regular and interpretable superpixels

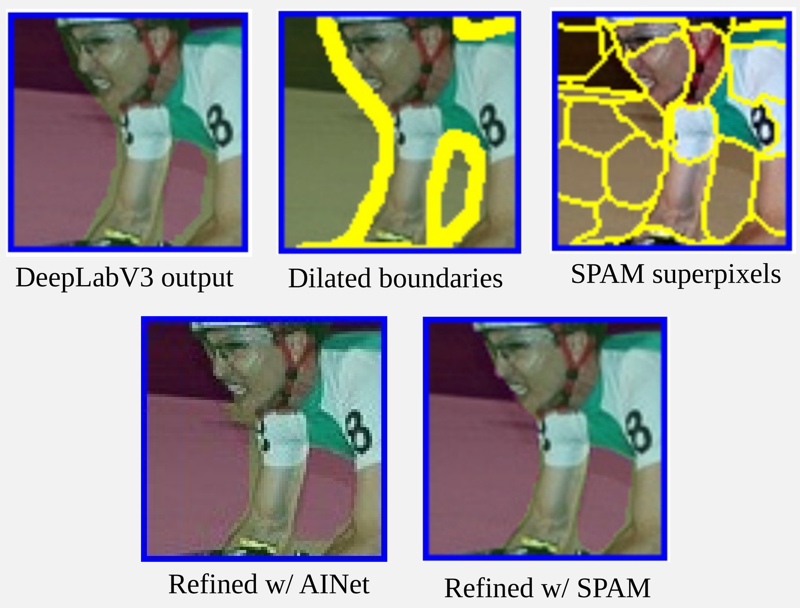
Recent methods try to increase accuracy at the expense of regularity



SFCN [4] | AINet [1] | VSSS [5] | CDS [6] | SPAM | SPAM (VA r=2)

Segmentation Refinement

- **SPAM can be used to refine semantic segmentation** around object borders.
- On **DeepLabV3 [7] outputs**, using a 5×5 dilation to define uncertain areas
- Achieves better **mIoU** than baseline and other superpixel methods on PASCAL VOC2012



DeepLabV3 output | Dilated boundaries | SPAM superpixels

Refined w/ AINet | Refined w/ SPAM

Conclusion and next steps

- **Most accurate yet regular superpixels**, outperforming state-of-the-art methods
- Supports **any prior segmentation** and handles **uncertainty regions** effectively
- Offers **adaptive and interactive modes**
- **Next Steps**
  - Hierarchical decomposition
  - Extension to video

References

[1] Yaxiong Wang, Yunchao Wei, Xueming Qian, Li Zhu, and Yi Yang. AINet: Association implantation for superpixel segmentation. In ICCV, 2021
[2] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloé Rolland, Laura Gustafson, Tete Xiao, head, Alexander C. Berg, Wan-Yen Lo, Piotr Dollàr, and Ross B. Girshick. Segment anything. In ICCV, 2023
[3] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In CVPR, 2021
[4] Fengting Yang, Qian Sun, Hailin Jin, and Zihan Zhou. Superpixel segmentation with fully convolutional networks. In CVPR, 2020
[5] Pei Zhou, Xuejing Kang, and Anlong Ming. Vine spread for superpixel segmentation. IEEE TIP, 32:878–891, 2023
[6] Sen Xu, Shikui Wei, Tao Ruan, and Lixin Liao. Learning invariant inter-pixel correlations for superpixel generation. In AAAI, 2024
[7] Liang-Chieh Chen. Rethinking atrous convolution for semantic image segmentation. arXiv:1706.05587, 2017