Explainable Image Segmentation with Prototypes

class- and scale-independence

Ábel Portik

Faculty of Mathematics and Computer Science,
Babeș–Bolyai University, Cluj-Napoca

29 November 2025

The problem of semantic image segmentation

- Safety-critical application areas:
  - self-driving cars,
  - aviation,
  - medical imaging.
- Deep learning models:
  - achieve high accuracy,
  - lack explanations ("black-box" nature).
- Need for interpretable models:
  - reliability and transparency,
  - understanding the model's behavior.

The problem of semantic image segmentation

CityScapes dataset [3]

- standard benchmark dataset for evaluating semantic segmentation
- contains images of urban environments with pixel-level annotations
- 30 semantic classes, including:
  - road
  - sidewalk
  - building etc.
- high-resolution images (1024×2048 pixels)

Self-explaining deep learning models

- built-in interpretability
  - through explanatory components
  - concept, prototype
- explanations understandable by humans
- combination of:
  - **intuitive reasoning** of classical ML models
  - **accuracy** of deep learning models

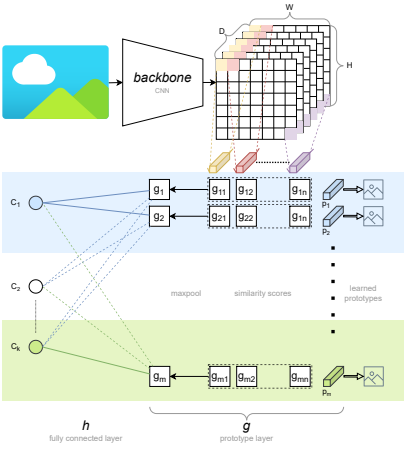
ProtoPNet [1]

strengths:

- patch-level prototypes
- "This looks like..." style reasoning

limitations:

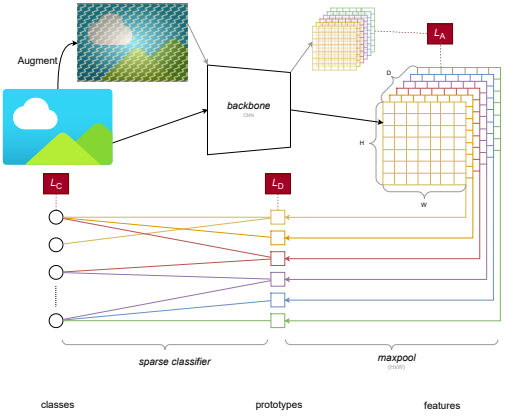
- prototypes have learnable weights
  - prototype – patch representation
- class-specific prototypes



PIP-Net [5]

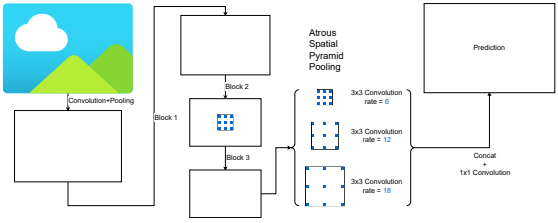
strengths:

- concept activations
  - no learned prototypes
  - unsupervised learning
- "sparse classifier"
- cross-class prototypes



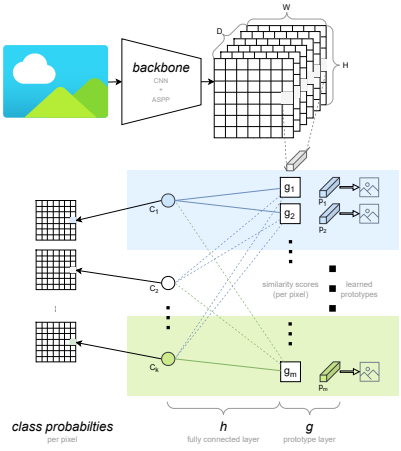
DeepLab v3 [2]

- based on the ResNet [4] architecture
- foundation for multiple prototype-based models
- Atrous Spatial Pyramid Pooling (ASPP) = multi-scale representation



ProtoSeg [7]

- similar to ProtoPNet
- based on the DeepLab v3 architecture
- prototype diversity ensured using Kullback–Leibler divergence

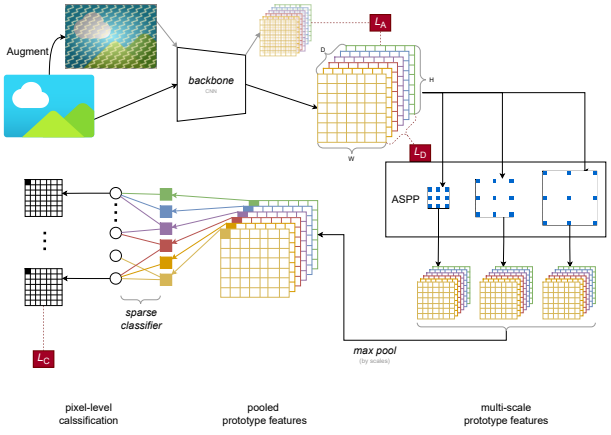


ScaleProtoSeg [6]

- prototypes at multiple scales
- leveraging the capabilities of ASPP

Proposed architecture

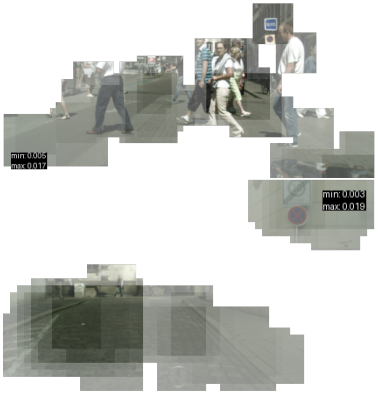
- spatially aligned latent representations
- scale-independent prototypes
  - ASPP: shared weights
  - activations examined at multiple scales
- class independence
  - "sparse classifier"



Several prototypes belonging to one concept

concept contribution to semantic classes:

- road (12.29)
- sidewalk (3.68)
- wall (1.29)
- ground (0.71)



Bibliography I


Chaofan Chen, Oscar Li, Chaofan Tao, Alina Jade Barnett, Jonathan Su, and Cynthia Rudin.
This looks like that: Deep learning for interpretable image recognition, 2019.


Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam.
Rethinking atrous convolution for semantic image segmentation, 2017.


Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele.
The cityscapes dataset for semantic urban scene understanding.
In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.


Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun.
Deep residual learning for image recognition, 2015.

Bibliography II


Meike Nauta, Jörg Schlötterer, Maurice van Keulen, and Christin Seifert.
Pip-net: Patch-based intuitive prototypes for interpretable image classification.
In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2744–2753, June 2023.


Hugo Porta, Emanuele Dalsasso, Diego Marcos, and Devis Tuia.
Multi-scale grouped prototypes for interpretable semantic segmentation.
In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 2869–2880. IEEE, 2025.


Mikołaj Sacha, Dawid Rymarczyk, Łukasz Struski, Jacek Tabor, and Bartosz Zieliński.
Protoseg: Interpretable semantic segmentation with prototypical parts, 2023.

Thank you for your attention!

Questions?

This research was supported by the **"Romanian Hub for Artificial Intelligence – HRIA"**, Smart Growth, Digitization and Financial Instruments Program, 2021–2027, MySMIS no. 351416 project,
and by the **Domus Scholarship of the Hungarian Academy of Sciences**, ID 62/22/2025/HTMT.