CNN-AWARE BINARY MAP FOR GENERAL SEMANTIC SEGMENTATION

Mahdyar Ravanbakhsh mahdyar.ravan@ginevra.dibe.unige.it — **University of Genova**

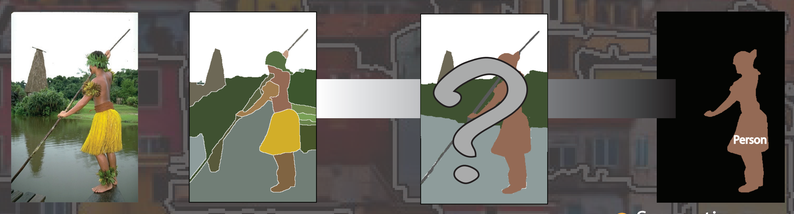
Hossein Mousavi hossein.mousavi@iit.it — **Italian Institute of Technology (IIT)**

Moin Nabi moin.nabi@unitn.it — **University of Trento**

Mohammad Rastegari mohammadr@allenai.org — **Allen Institute for Artificial Intelligence (AI2)**

Carlo Regazzoni carlo@dibe.unige.it

Motivation



- Generic segments
- Low-level features
- Unsupervised

General Semantic Segmentation

- Semantic segments
- High-level consepts
- Strongly supervised

Backgrounds

Low-level Segmentation

Partitioning an image based on the low-level image features:

- Graph-based approaches (e.g, EGS[1])
- Gradient-ascent-based approaches (e.g., SLIC[2])

Problems: Lack of semantic, not invariant to illumination and occlusion.

Semantic Segmentation

Partitioning a scene into semantic regions and a unique object label is assigned to each region:

- Supervised Fully Convolutional Neural Network [3].

Problems: Supervision is biased, non-comprehensive, and not scalable

General Idea

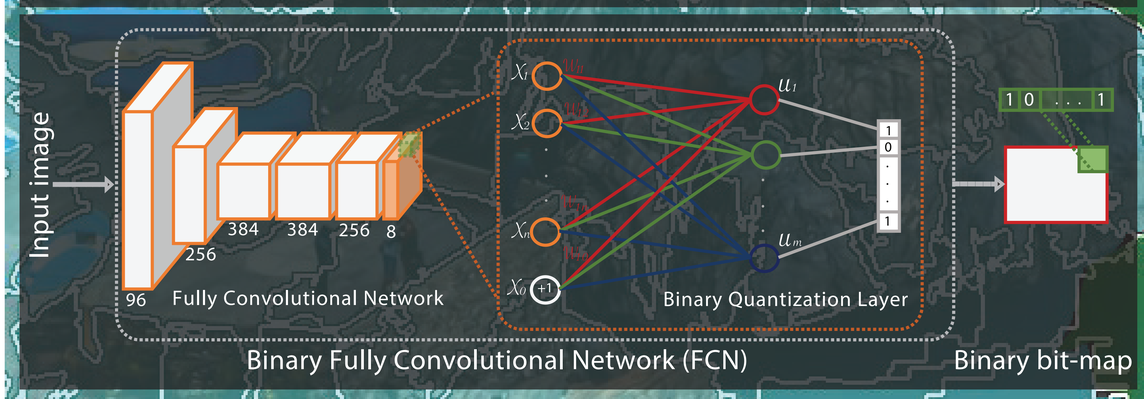
Goal: Narrow down the semantic gap between low-level segmentation and semantic segmentation:

By: Inject semantics (inherited from generic CNN representations trained on smaller set of categories) into general segmentation, while maintaining the method complexity in a manageable level.

Binary Quantization Layer

Strength: Compact binary representation instead of high-dimensional CNN features, Efficient partitioning with hashing techniques embedded as layer for end-to-end training of the net.

Binary Encoding: Binarizing CNN feature maps by a linear transformation (implemented simply as convolution) where the weights are initialized with Locality Sensitive Hashing (LSH).

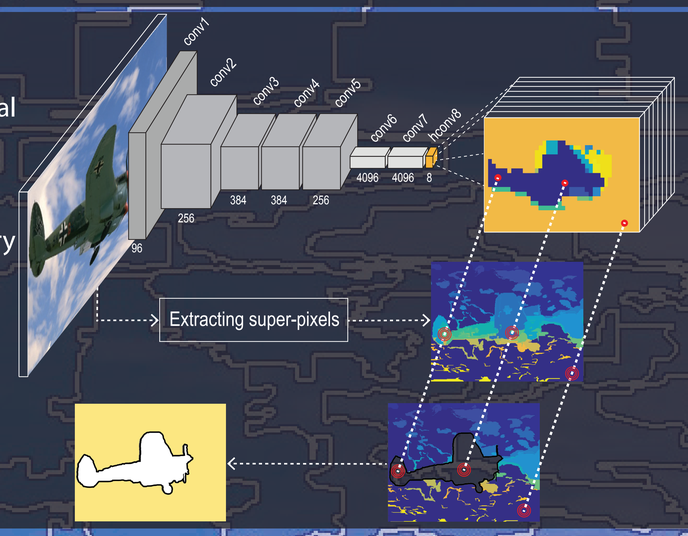


Proposed Framework

Intuition: The nearby pixels should have similar visual attributes unless they undergo a large semantic change. We proposed a Binary Convolutional Neural Network which provide the means to represent the visual attributes as the binary patterns.

Algorithm

- Given input image to a Fully Convolutional Neural Network (FCN) obtaining a CNN feature map.
- Generate compact binary representation of the CNN features maps through the Binary Encoding Layer.
- Refine the binary bit maps by averaging over superpixels.
- Partition the image by merging the superpixels with the similar binary pattern.



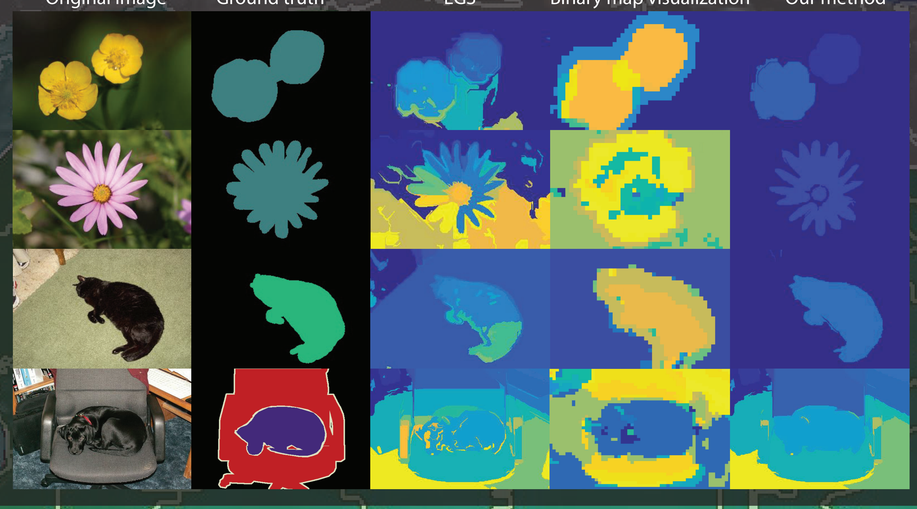
Results

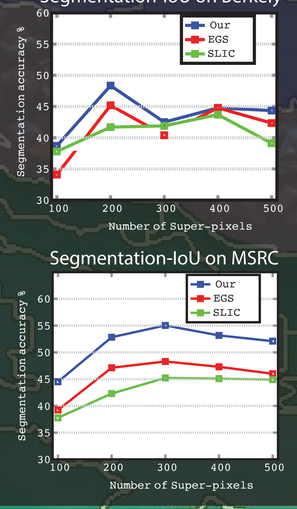
Our method is compared with EGS [1], and SLIC[2] in terms of Intersection over Union (IoU) measure.

Datasets:

Berkeley Segmentation Dataset (BSDS500), Microsoft Research Cambridge database (MSRC).

MSRC		Berkeley	
Method	IoU	Method	IoU
EGS [1]	50.3%	EGS [1]	45.19%
SLIC [2]	48.7%	SLIC [2]	43.70%
Our method	55.03 %	Our method	48.35%





References

[1] Pedro F Felzenszwalb and Daniel P Huttenlocher, "Efficient graph-based image segmentation," in IJCV 2004.

[2] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Susstrunk, " SLIC superpixels compared to state-of-the-art superpixel methods," in PAMI 2012.

[3] Jonathan Long, Evan Shelhamer, and Trevor Darrell, "Fully convolutional networks for semantic segmentation," in CVPR 2015.