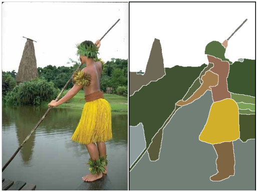
CNN-aware Binary Map for General Semantic Segmentation

Mahdyar Ravanbakhsh, Hossein Mousavi, Moin Nabi, Mohammad Rastegari, Carlo Regazzoni

September 2016

Low-level Segmentation VS. Semantic Segmentation



Low-level Segmentatio

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

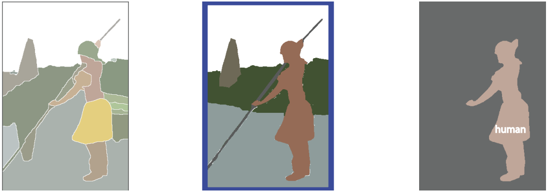
Partitioning a scene into semantic regions and a unique object label is assigned to each region:

- Supervised Fully Convolutional Neural Network[3].

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

[1] Felzenszwalb, P. F. and Huttenlocher, D. P, "Efficient graph-based image segmentation", IJCV 2004.
[2] Achanta, R. et.al. , "SLIC superpixels compared to state-of-the-art superpixel methods", PAMI 2012.
[3] Long, J., Shelhamer, E. and Darrell, T., "Fully convolutional networks for semantic segmentation". CVPR 2015.

General Semantic Segmentation



General Semantic Segmentation

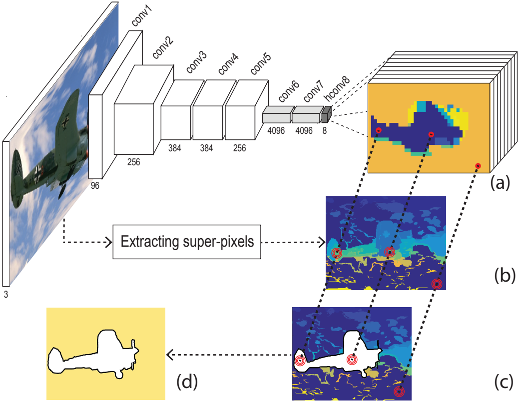
Goal: Narrow down the semantic gap between low-level segmentation and semantic segmentation

By: Inject semantics into general segmentation, while maintaining the method complexity in a manageable level.

Overview

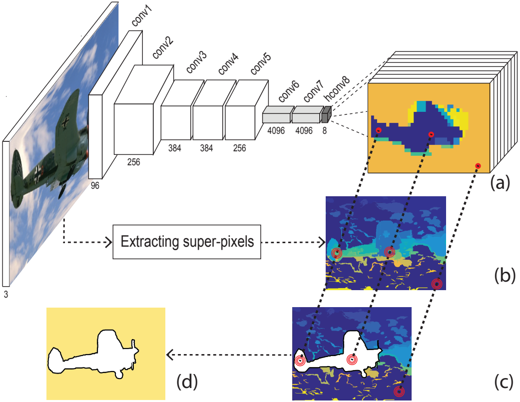
General Idea

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.



Overview

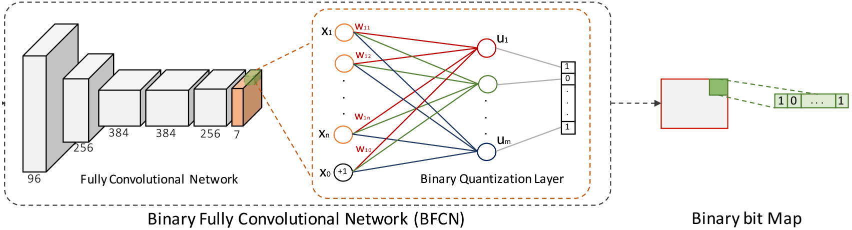
- Given input image to a Fully Convolutional Neural Network (FCN) obtaining a CNN feature map.
- Generate compact binary representation of the CNN features maps through the Binary Encoding Layer.
- Refine the binary bit maps by averaging over superpixels.
- Partition the image by merging the superpixels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



Binary Quantization Layer

Strength: Compact binary representation instead of high-dimensional CNN features, Efficient partitioning with hashing layer for training the network in an end-to-end fashion.

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

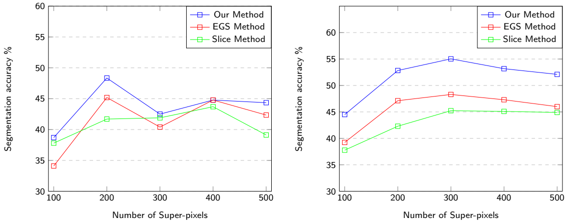


Results

Quantitative results on MSRC and Berkeley datasets

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%**

Segmentation-IoU over superpixel variation

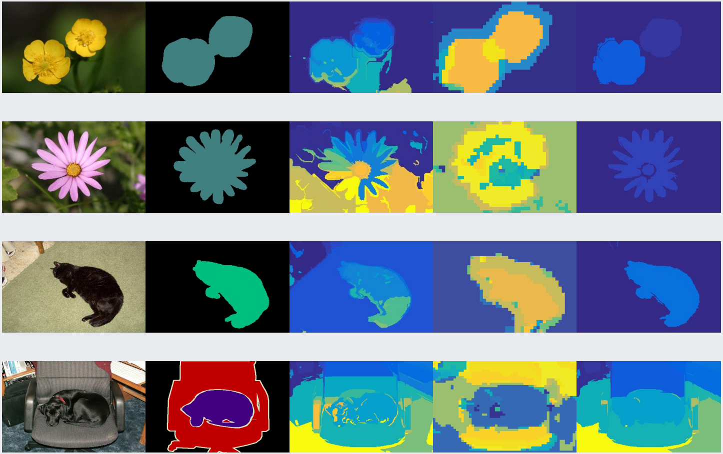


(a) Berkeley (b) MSRC

[1] Felzenszwalb, P. F. and Huttenlocher, D. P, "Efficient graph-based image segmentation", IJCV 2004.
[2] Achanta, R. et.al. , "SLIC superpixels compared to state-of-the-art superpixel methods", PAMI 2012.

Results

Image Samples | Ground truth | EGS [1] | Binary maps visualization | Our method



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