Exploiting Saliency for Segmenting Objects from Image Level Labels

Seong Joon Oh[1], Rodrigo Benenson[1], Anna Khoreva[1], Zeynep Akata[1,2], Mario Fritz[1], Bernt Schiele[1]

[1]Max-Planck Institute for Informatics, [2]University of Amsterdam

max planck institut informatik

goo.gl/KygSeb

Task & Motivation

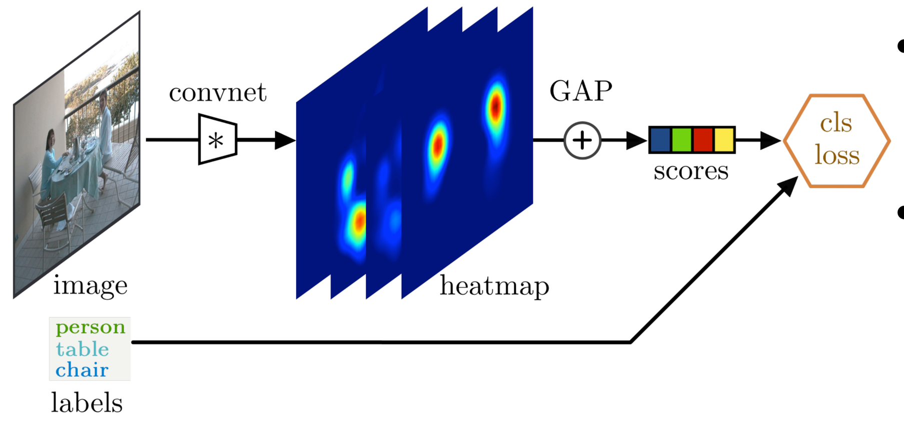
Learning to segment objects from image label annotations.

- Cheaper than full supervision.
- Humans can do.

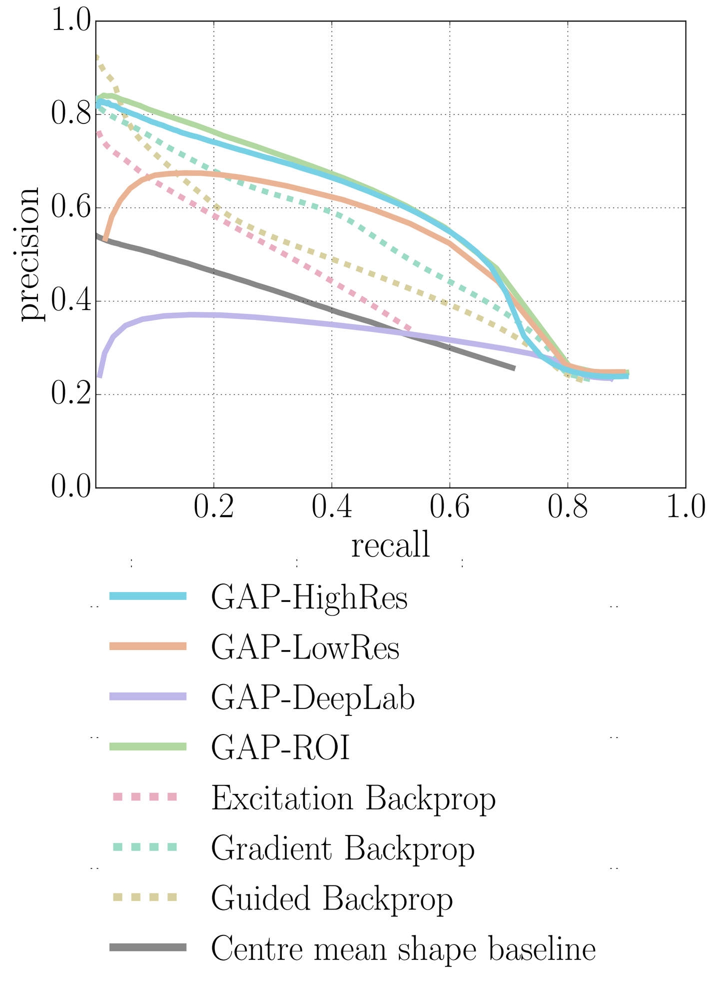
Training: Image Labels → Final task: Semantic Labelling

1. Seed : Encode Image Labels

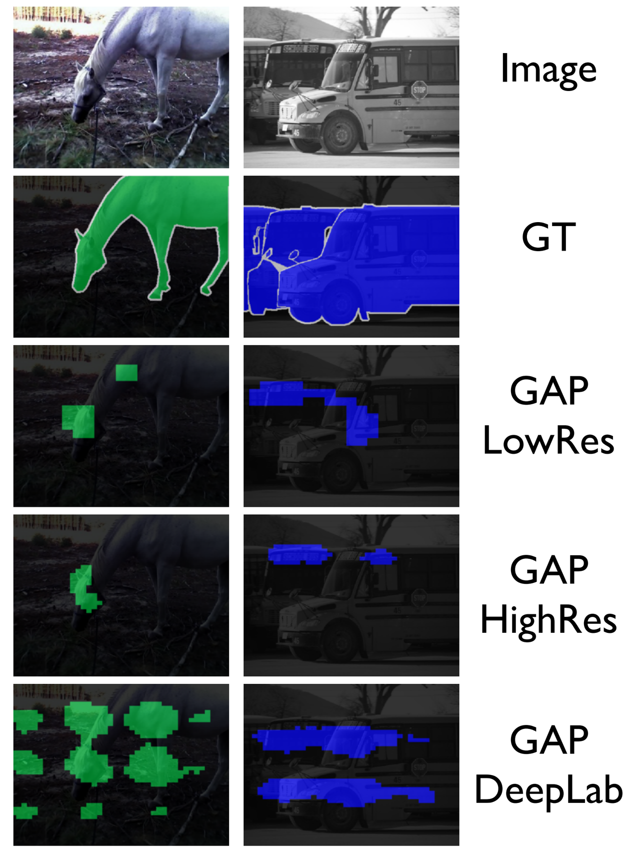
Discriminative object locations from image-level classifiers.



- **Data**: Pascal images + image labels.
- **Model**: fully convolutional network + global average pooling (GAP) [1,2].

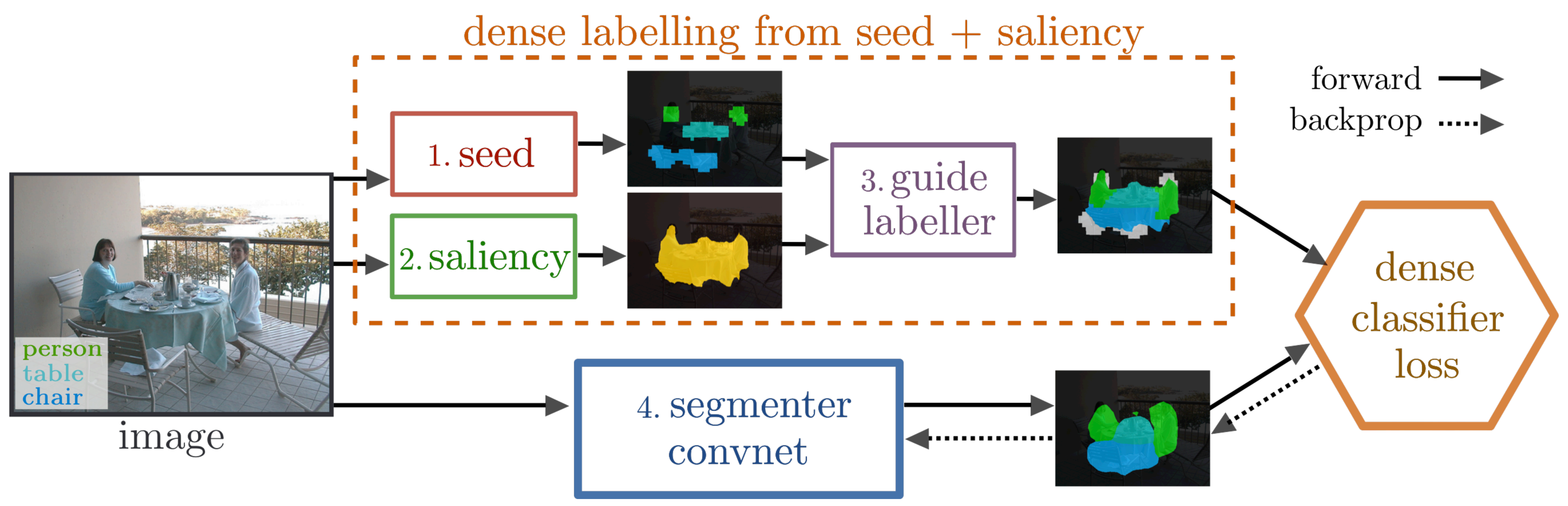


FG class-averaged precision-recall curves for GAP and Gradient based seeds.



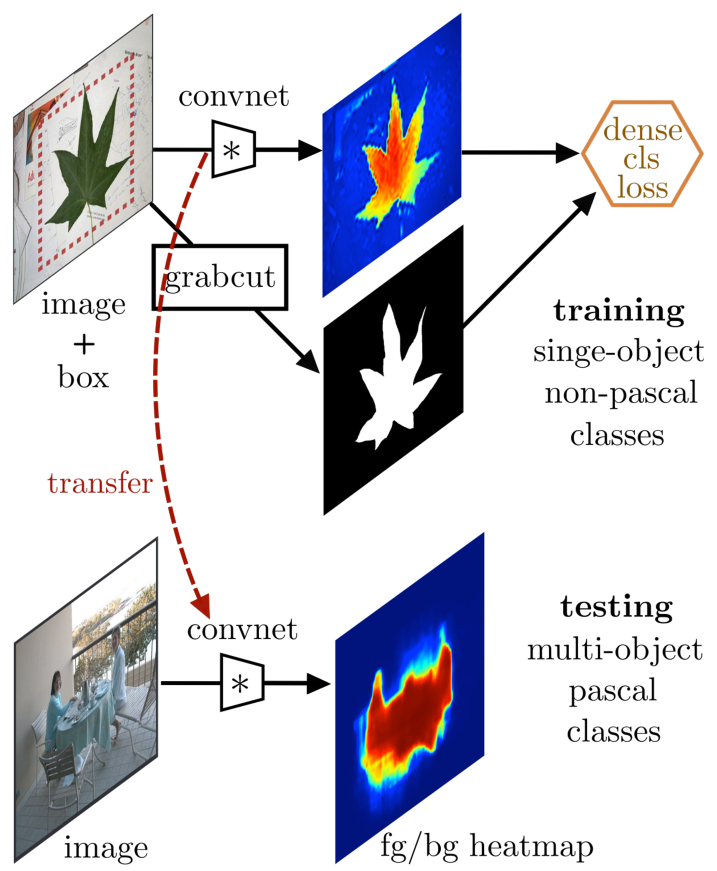
Qualitative results for different GAP types.

Approach : Guided Segmentation

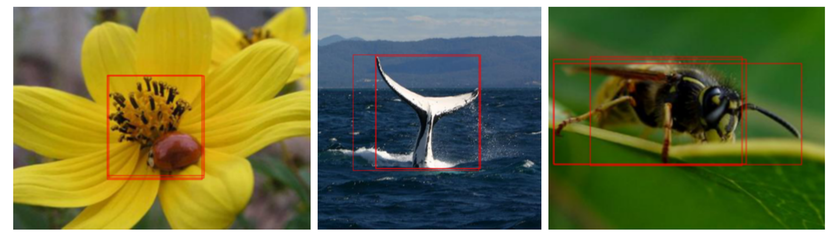


1. Get discriminative object locations from an image-level classifier [1,2] (**seed**).
2. Image labels alone do not give full object extent information (e.g. train and rail); we propose to exploit class-agnostic image-level saliency (**saliency**).
3. Combine the two sources of information (**guide labels**).
4. Refine the labelling by training a **segmenter** (e.g. DeepLab [4]) with the guide labels.

2. Saliency : Encode "Objectness" Prior

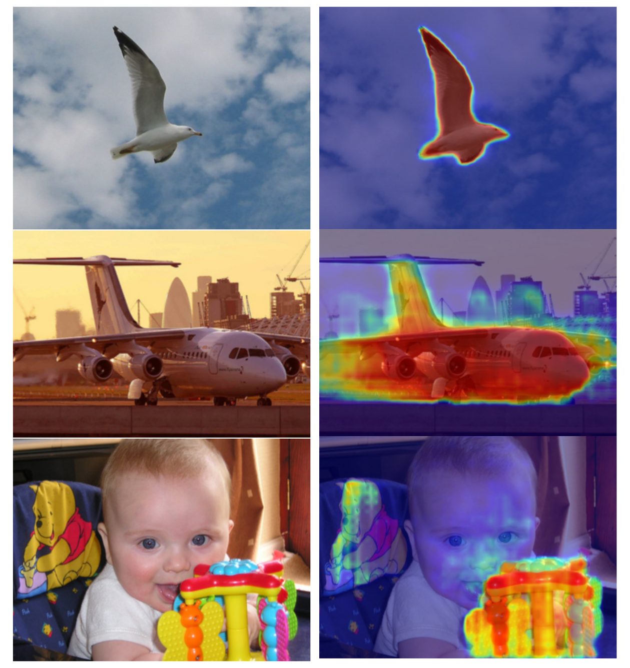


Foreground mask of generic object class.



MSRA non-Pascal training data.

- **Data**: 11k MSRA single-object images with boxes [3]. Only non-Pascal classes are used for the class-genericity of the mask.
- **Model**: DeepLab [4].

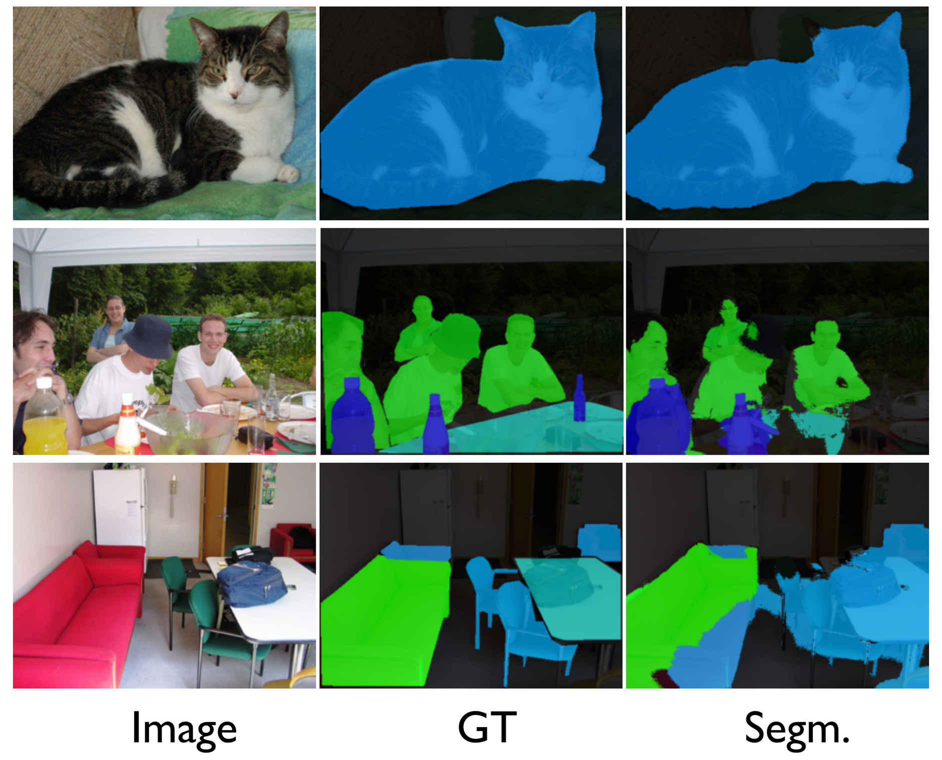


Predicted saliency on Pascal.

4. Segmentation Result & Comparison

Method	Data	Val mIoU	Test mIoU	FS%
MIL-FCN ICLRW'15	I+P	25.0	25.6	36.5
DCSM ECCV'16	I+P	44.1	45.1	64.2
SEC ECCV'16	I+P	50.7	51.7	73.5
STC arXiv'15	I+P+S+E_{40k}	49.8	51.2	72.8
CheckMask ECCV'16	I+P+ μ	51.5	-	-
MicroAnno BMVC'16	I+P+ μ	51.9	53.2	75.7
GuideLabel	**I+P+S**	**55.7**	**56.7**	**80.6**
DeepLabv1 [4]	I+P_{full}	67.6	70.3	100

I	ImageNet pretrain	E_n	*n* images with labels
P	Pascal image labels	μ	Human in the loop
S	Saliency	P_{full}	Pascal full supervision



- Reach **80% of the fully supervised** performance.
- Better saliency model will further improve the result; oracle saliency gives 61.8 mIoU.

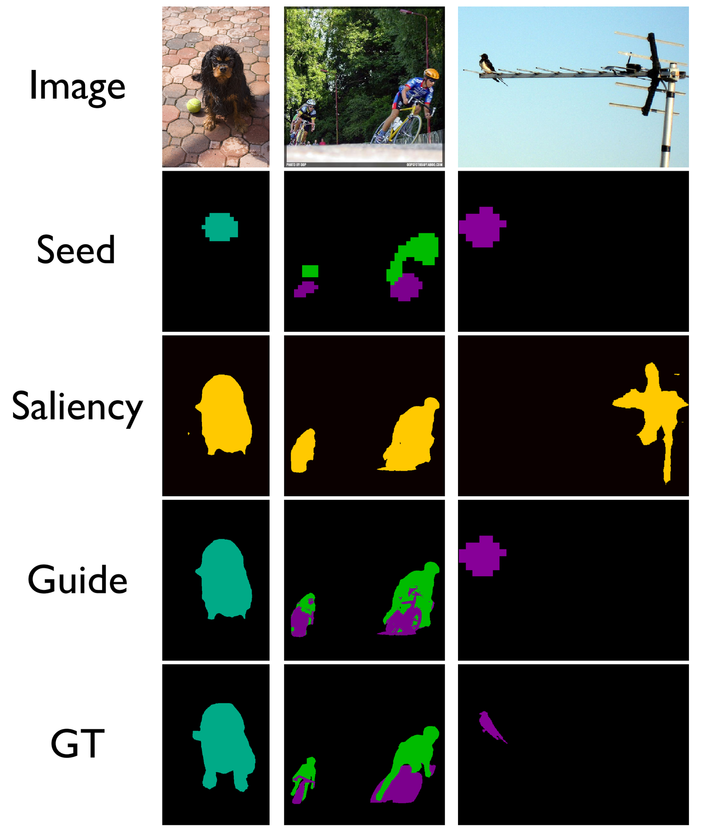
3. Guide Label : Seed + Saliency

Combination algorithm

i. Break seed and saliency into connected components.

ii. If seeds touch saliency: diffuse seeds inside saliency with dense CRF.

iii. If seed is alone, label as FG; If saliency is alone, label as BG.



References

[1] Zhou et al. Learning Deep Features for Discriminative Localization. CVPR'16.
[2] Kolesnikov et al. Seed, Expand, Constrain: Three Principles for Weakly-Supervised Image Segmentation. ECCV'16.
[3] Cheng et al. Global Contrast Based Salient Region Detection. TPAMI'15.
[4] Chen et al. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. ICLR'15.

Acknowledgement: Supported by German Research Foundation (DFG CRF 1223).