Seed, Expand and Constrain: Three Principles for Weakly-Supervised Image Segmentation

Alexander Kolesnikov
(joint work with Christoph Lampert)
IST Austria

Outline

- Short introduction in semantic image segmentation
  - fully-supervised learning
  - weakly-supervised learning
- SEC: Seed, Expand and Constrain
  - Localization loss (seed)
  - Classification loss (expand)
  - Boundary-aware loss (constrain)
- Evaluation and discussion

Outline

- Short introduction in semantic image segmentation
  - **fully-supervised learning**
  - weakly-supervised learning
- SEC: Seed, Expand and Constrain
  - Localization loss (seed)
  - Classification loss (expand)
  - Boundary-aware loss (constrain)
- Evaluation and discussion

Deep network for Image Classification

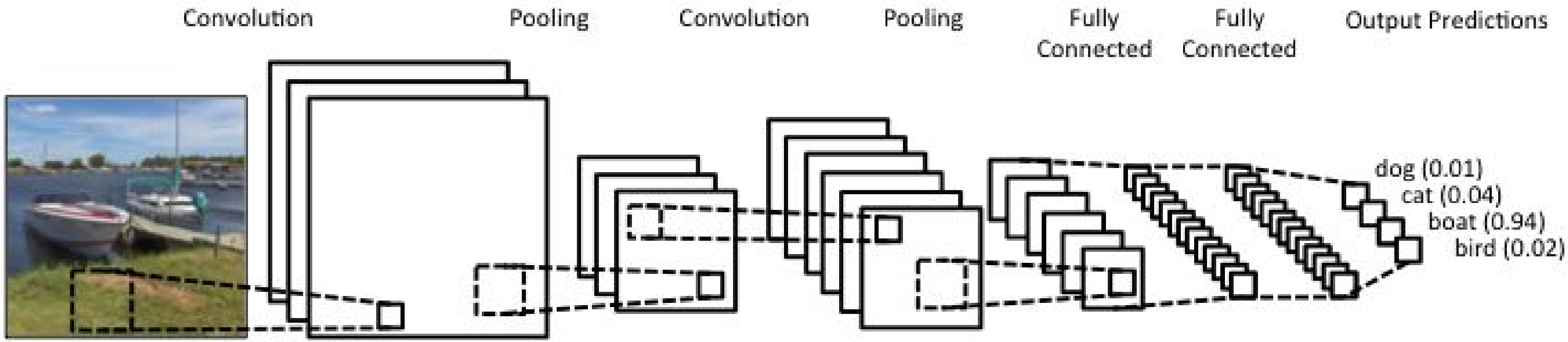


Figure credit: http://www.clarifai.com/technology

Deep network for Image Segmentation

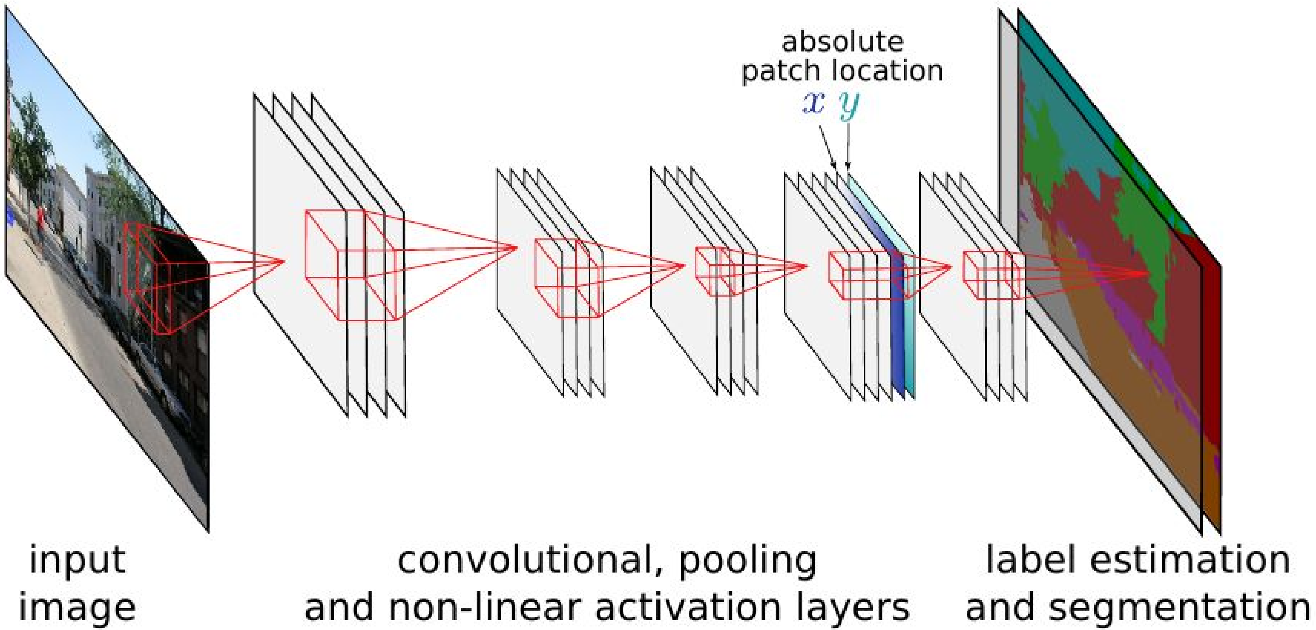


Figure credit: Simon, CVPR-WS 2015

DeepLab-CRF

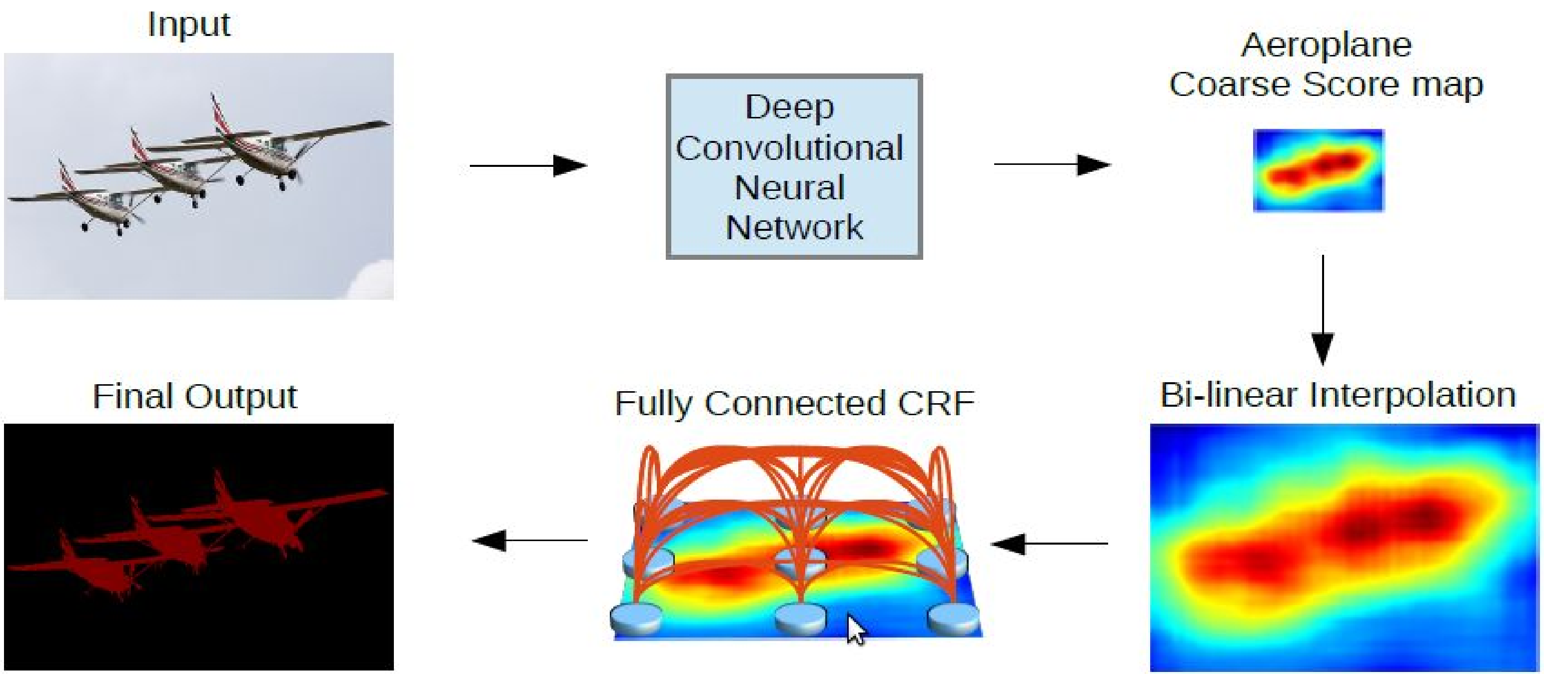


Figure credit: Chen, ICLR 2015

DeepLab-CRF results

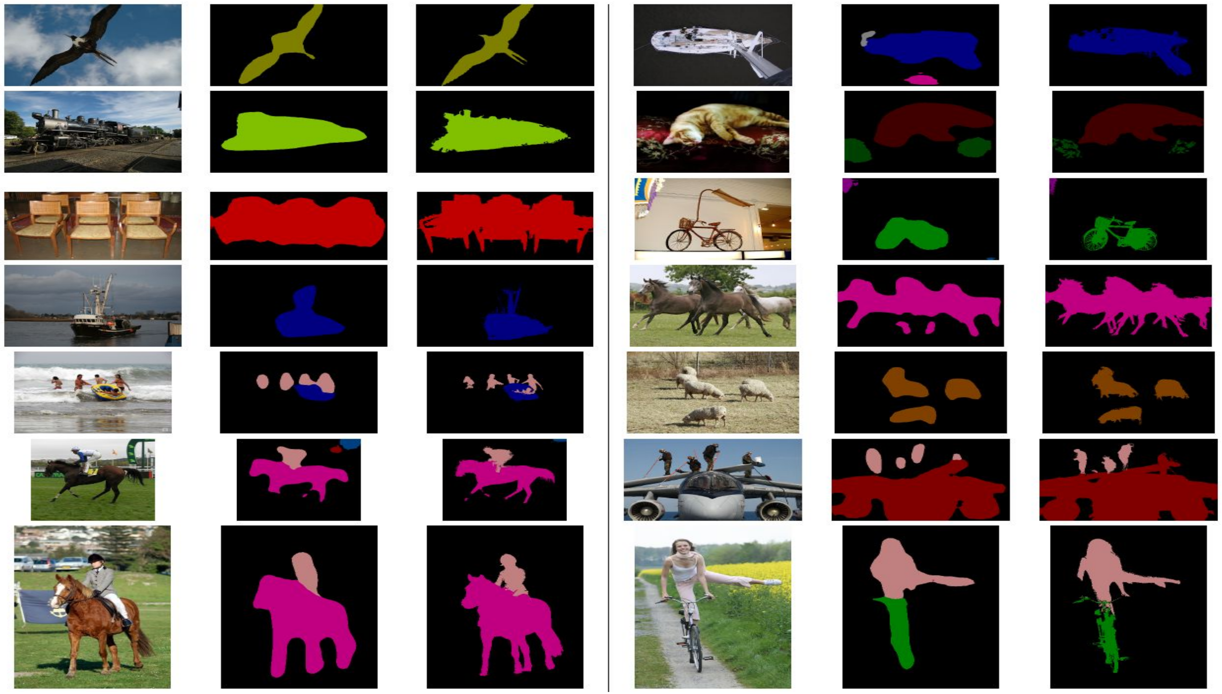


Figure credit: Chen, ICLR 2015

Outline

- Short introduction in semantic image segmentation
  - fully-supervised learning
  - **weakly-supervised learning**
- SEC: Seed, Expand and Constrain
  - Localization loss (seed)
  - Classification loss (expand)
  - Boundary-aware loss (constrain)
- Evaluation and discussion

Weakly-Supervised Segmentation

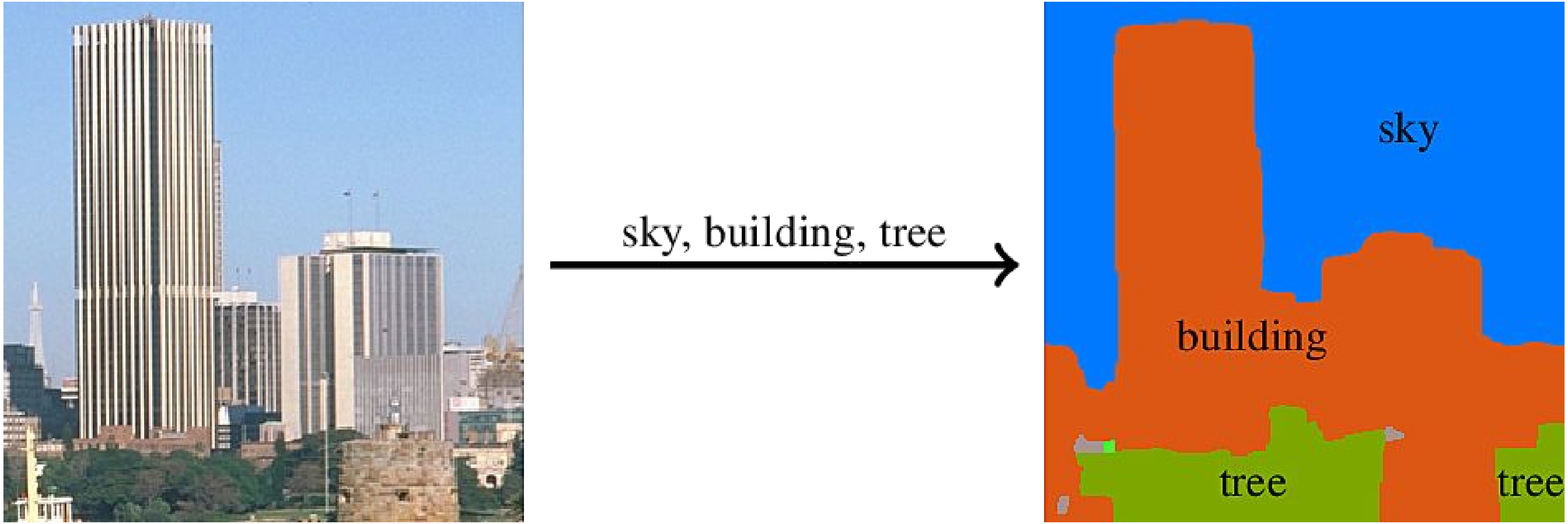


Figure credit: Xu, CVPR 2014

Multiple Instance Learning

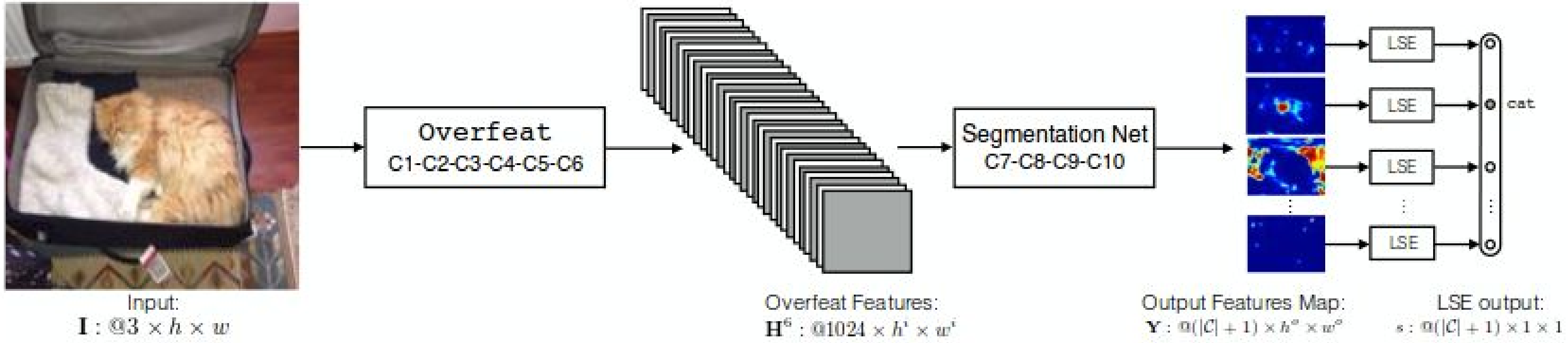


Figure credit: Pinheiro, CVPR15

Self-training (Deeplab)

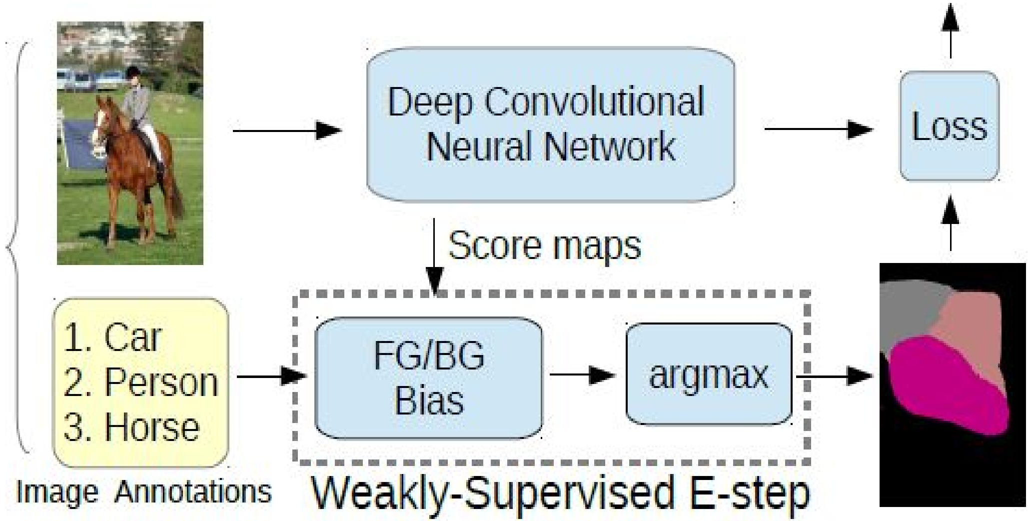
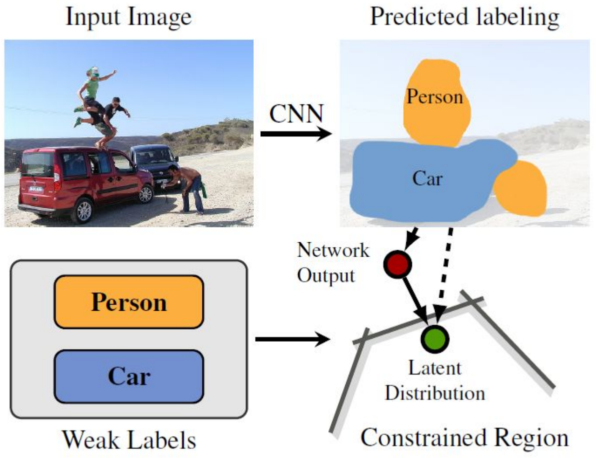
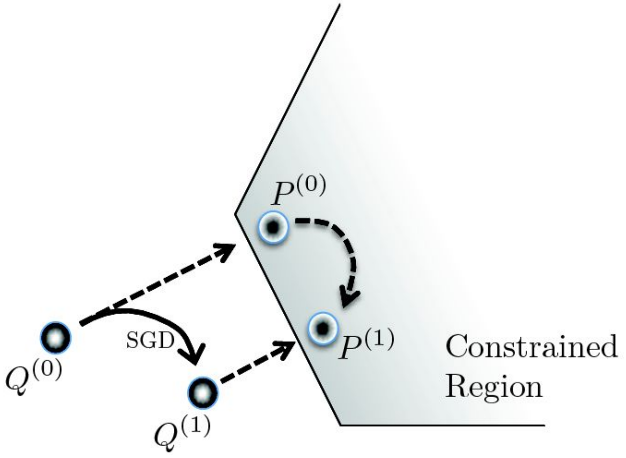


Figure credit: Papandreou , ICCV 2015

Self-training (Constrained CNN)





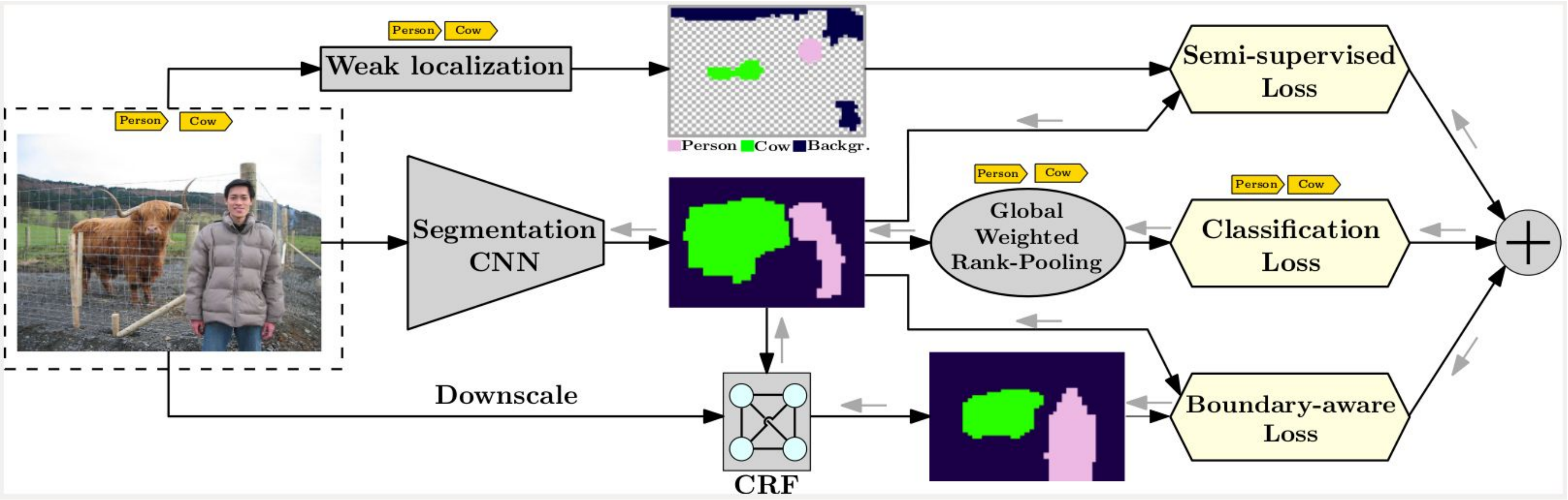
$$\begin{array}{lll} \text{find} & \theta & \\ \text{subject to} & A_I \vec{Q}_I \geq \vec{b}_I & \forall I \end{array} \longrightarrow \begin{array}{ll} \underset{\theta, P}{\text{minimize}} & D\left(P(X) \| Q(X|\theta)\right) \\ \text{subject to} & A\vec{P} \geq \vec{b}, \quad \sum_X P(X) = 1 \end{array}$$

Figure credit: Pathak, ICCV 2015

Outline

- Short introduction in semantic image segmentation
  - fully-supervised learning
  - weakly-supervised learning
- **SEC: Seed, Expand and Constrain**
  - Localization loss (seed)
  - Classification loss (expand)
  - Boundary-aware loss (constrain)
- Evaluation and discussion

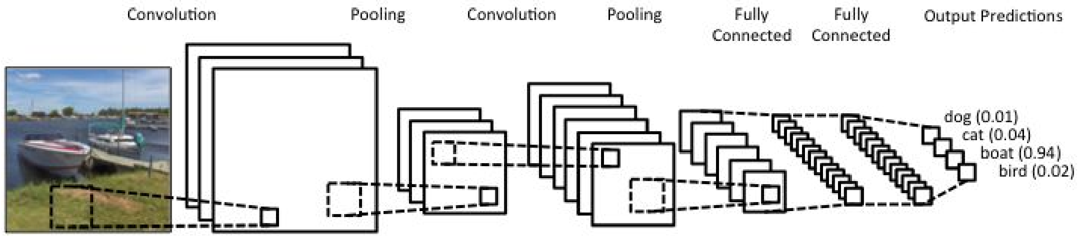
SEC: overview



Outline

- Short introduction in semantic image segmentation
  - fully-supervised learning
  - weakly-supervised learning
- SEC: Seed, Expand and Constrain
  - **Localization loss (seed)**
  - Classification loss (expand)
  - Boundary-aware loss (constrain)
- Evaluation and discussion

Classification



GT: letter opener
1: drumstick
2: candle
3: wooden spoon
4: spatula
5: ladle

GT: letter opener
1: Band Aid
2: ruler
3: rubber eraser
4: pencil box
5: wallet

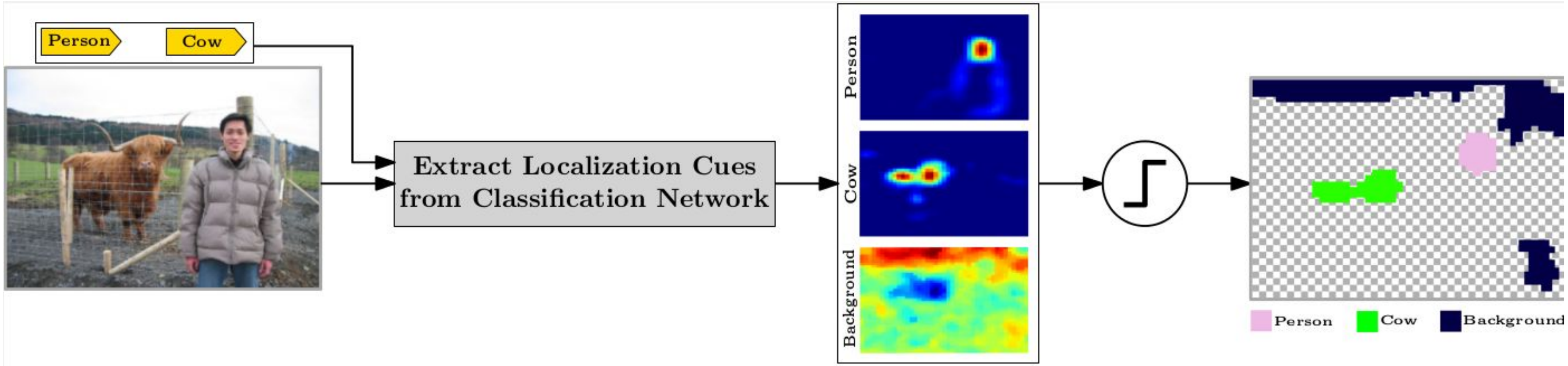
GT: letter opener
1: fountain pen
2: ballpoint
3: hammer
4: can opener
5: ruler

GT: spotlight
1: grand piano
2: folding chair
3: rocking chair
4: dining table
5: upright piano

GT: spotlight
1: acoustic guitar
2: stage
3: microphone
4: electric guitar
5: banjo

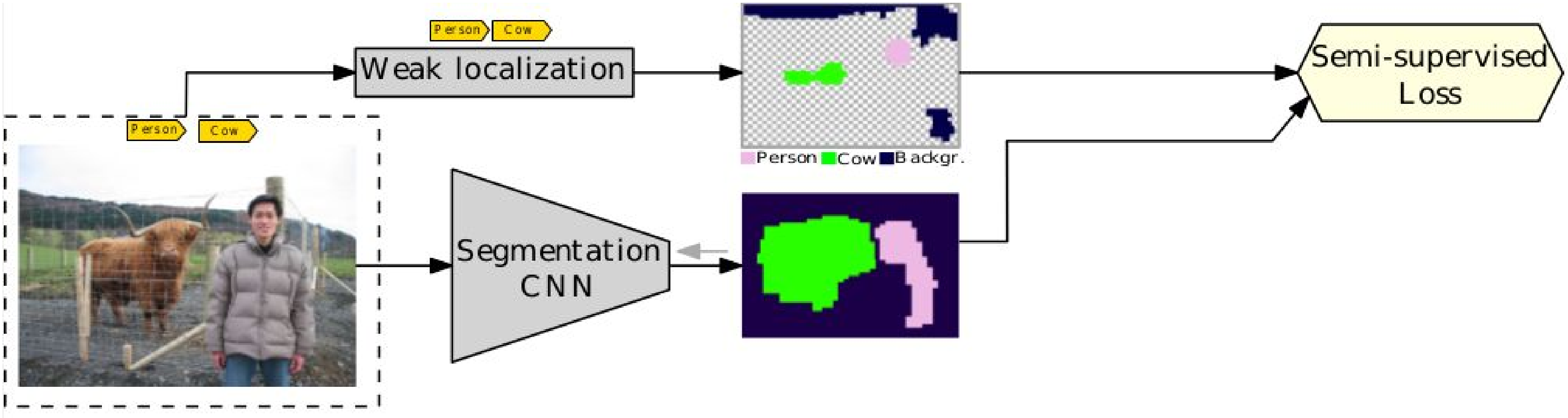
GT: spotlight
1: altar
2: candle
3: perfume
4: restaurant
5: confectionery

Weak localization



Oquab, CVPR15
Zhou, CVPR16
Simonyan, ICLR14
Bazzani, WACV16
Zhou, ICLR15

Localization loss

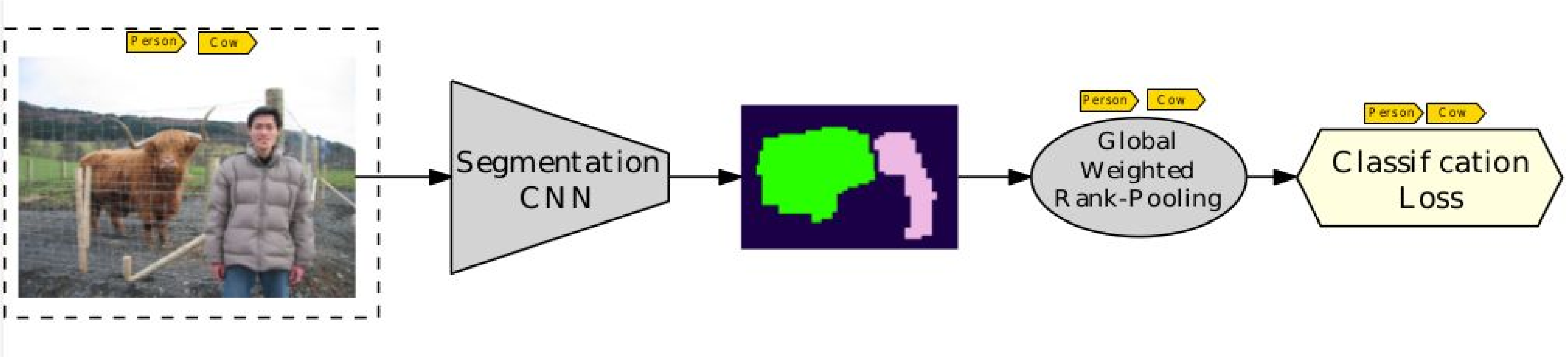


$$L_{\text{semi}}(f(X),T) = \frac{1}{\sum_{c\in T}|S_c|}\sum_{c\in T}\sum_{u\in S_c}\log f_{u,c}(X)$$

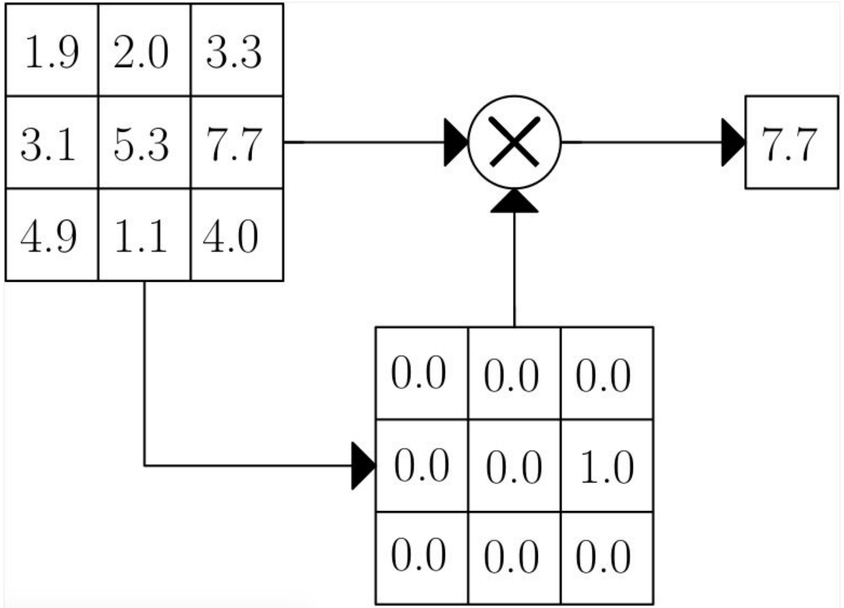
Outline

- Short introduction in semantic image segmentation
  - fully-supervised learning
  - weakly-supervised learning
- SEC: Seed, Expand and Constrain
  - Localization loss (seed)
  - **Classification loss (expand)**
  - Boundary-aware loss (constrain)
- Evaluation and discussion

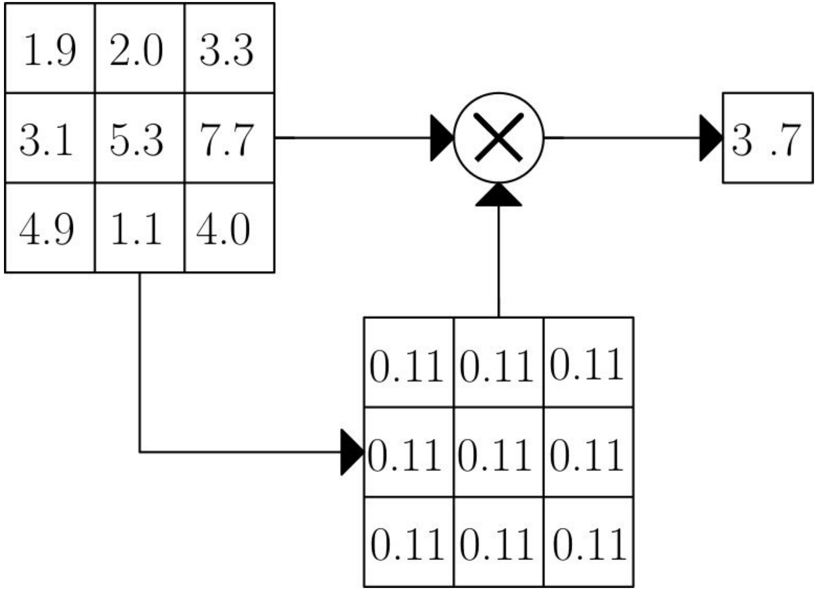
Classification loss



Global pooling strategies

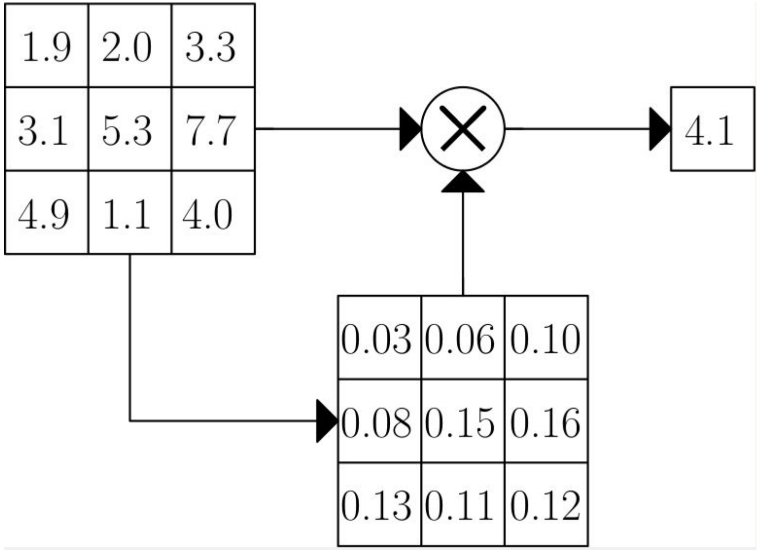


Global max-pooling



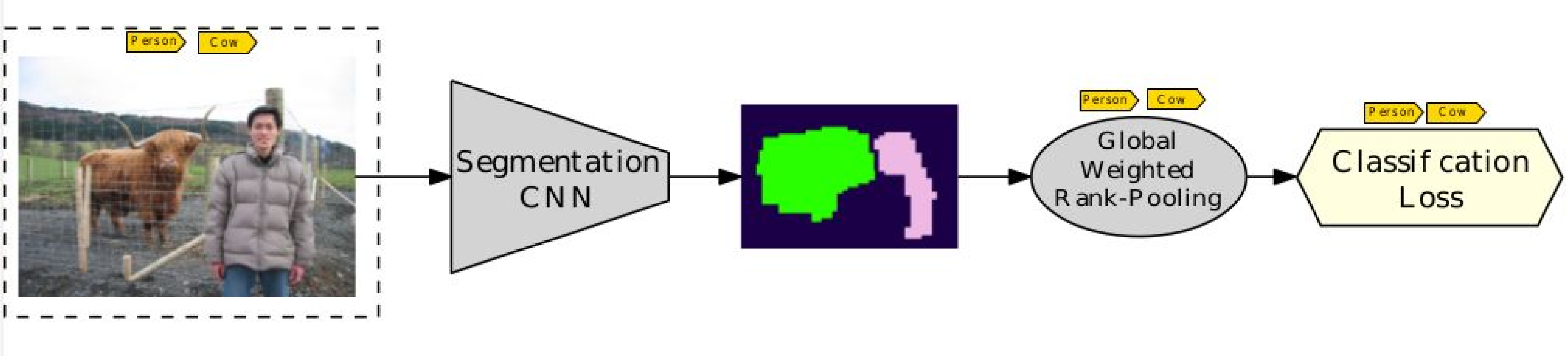
Global avg-pooling

Global weighted rank-pooling



$$G_c(X; d_c) = \frac{1}{Z(d_c)} \sum_{j=1}^{n} (d_c)^{j-1} f_{i_j,c}(X), \text{ where } Z(d_c) = \sum_{j=1}^{n} (d_c)^{j-1}$$

Classification loss

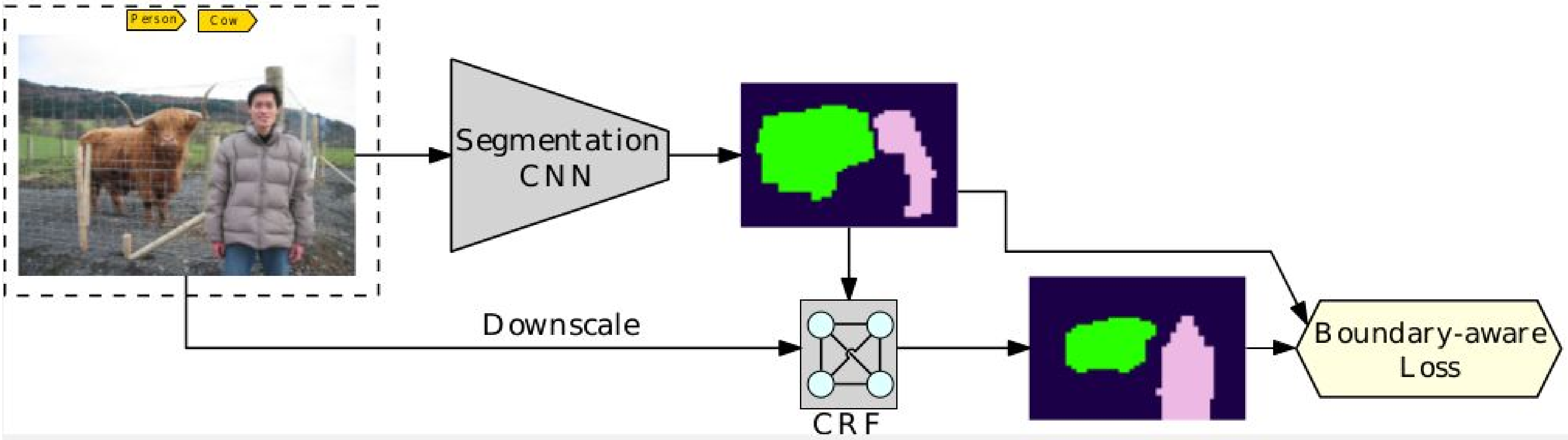


$$L_{\text{class}}(f(X),T) = \sum_{c\in T}\frac{\log G_c(X;d_+)}{|T|} + \sum_{c\in\mathcal{C}'\setminus T}\frac{\log(1-G_c(X;d_-))}{|\mathcal{C}'\setminus T|} + \log G_{c^{\text{bg}}}(X;d_{\text{bg}})$$

Outline

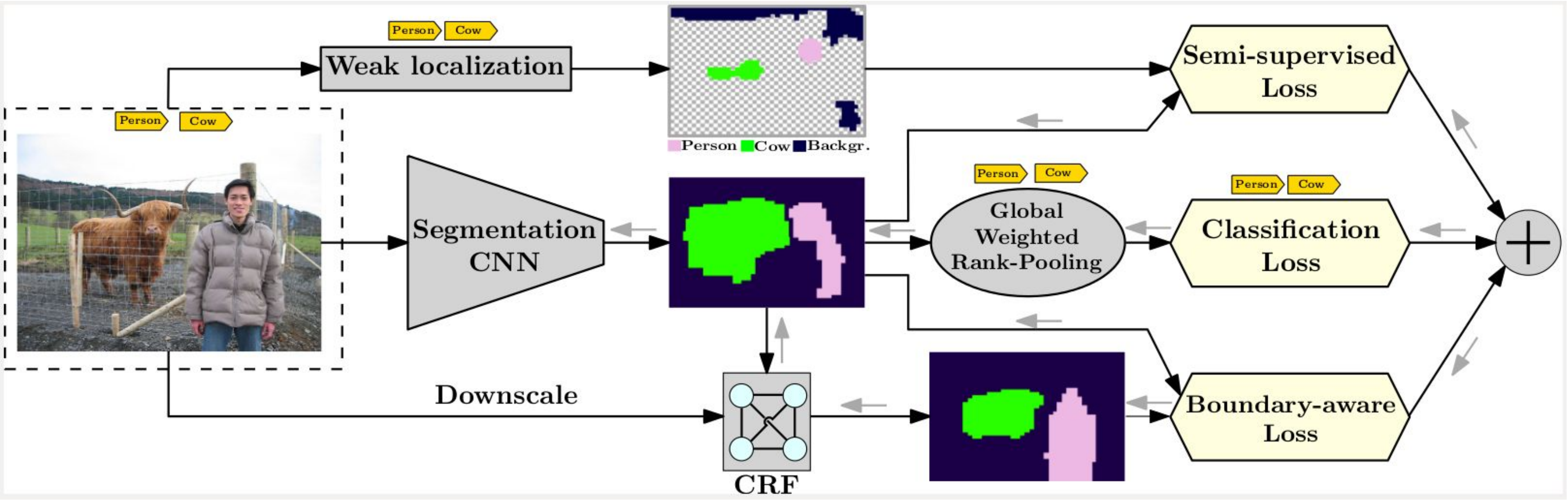
- Short introduction in semantic image segmentation
  - fully-supervised learning
  - weakly-supervised learning
- SEC: Seed, Expand and Constrain
  - Localization loss (seed)
  - Classification loss (expand)
  - **Boundary-aware loss (constrain)**
- Evaluation and discussion

Boundary-aware loss



$$L_{\text{bound}}(X, f(X)) = \frac{1}{n} \sum_{u=1}^{n} \sum_{c \in \mathcal{C}} Q_{u,c}(X) \log \frac{Q_{u,c}(X)}{f_{u,c}(X)}$$

SEC: overview



Outline

- Short introduction in semantic image segmentation
  - fully-supervised learning
  - weakly-supervised learning

- SEC: Seed, Expand and Constrain
  - Localization loss (seed)
  - Classification loss (expand)
  - Boundary-aware loss (constrain)

- **Evaluation and discussion**

Experimental setup

Dataset: PASCAL VOC 2012

- Three parts: train (10582), val (1449)and test (1456)
- 20 semantic classes

Metric: Mean IoU

Deep CNN: from Chen, ICLR2015

CRF: from Krähenbühl, NIPS2011

Software: *Caffe* + *Python* with *Theano*

Hardwrare: GeForce TITAN-X

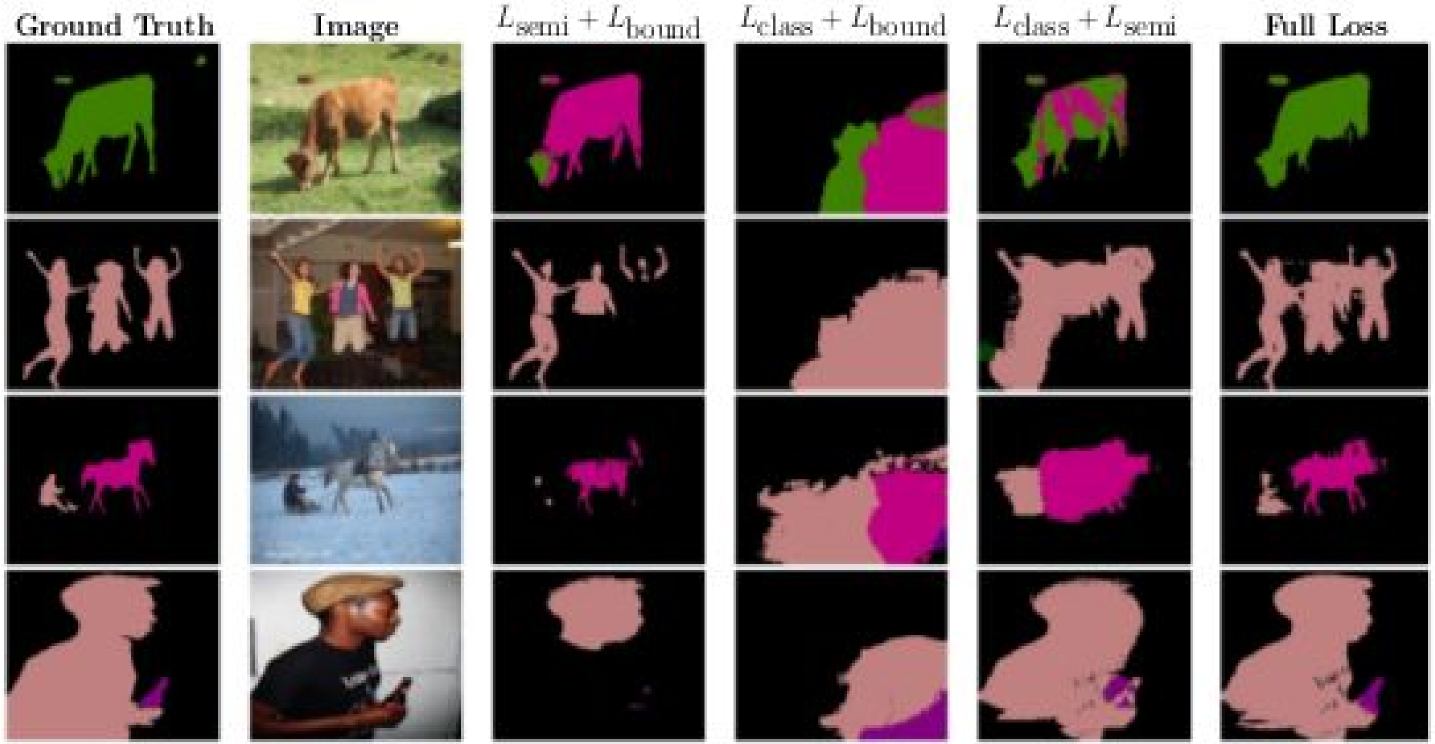
Optimization: SGD (8000 iterations, batch size 15)

PASCAL VOC 2012 *val* set	[2] (Img+Obj)	[44] (stage1)	EM-Adapt (re-impl. of [33])	CCNN [33]	MIL+ILP +SP-sppxl† [5]	SEC (proposed)
background		71.7*	67.2	68.5	77.2	**82.2**
aeroplane		30.7*	29.2	25.5	37.3	**61.7**
bike		30.5*	17.6	18.0	18.4	**26.0**
bird		26.3*	28.6	25.4	25.4	**60.4**
boat		20.0*	22.2	20.2	**28.2**	25.6
bottle		24.2*	29.6	36.3	31.9	**45.6**
bus		39.2*	47.0	46.8	41.6	**70.9**
car		33.7*	44.0	47.1	48.1	**63.2**
cat		50.2*	44.2	48.0	50.7	**72.2**
chair		17.1*	14.6	15.8	12.7	**20.9**
cow		29.7*	35.1	37.9	45.7	**52.9**
diningtable		22.5*	24.9	21.0	14.6	**30.6**
dog		41.3*	41.0	44.5	50.9	**62.8**
horse		35.7*	34.8	34.5	44.1	**56.8**
motorbike		43.0*	41.6	46.2	39.2	**63.5**
person		36.0*	32.1	40.7	37.9	**57.1**
plant		29.0*	24.8	30.4	28.3	**32.2**
sheep		34.9*	37.4	36.3	44.0	**60.6**
sofa		23.1*	24.0	22.2	19.6	**32.3**
train		33.2*	38.1	38.8	37.6	**44.8**
tv/monitor		33.2*	31.6	36.9	35.0	**42.3**
average	32.2*	33.6*	33.8	35.3	36.6	**50.7**

PASCAL VOC 2012 *test* set	MIL-FCN [31]	CCNN [33]	MIL+ILP +SP-sppxl† [5]	SEC (proposed)
background		≈71‡	74.7	**83.0**
aeroplane		24.2	38.8	**55.6**
bike		19.9	19.8	**27.4**
bird		26.3	27.5	**61.1**
boat		18.6	21.7	**22.9**
bottle		38.1	32.8	**52.4**
bus		51.7	40.0	**70.2**
car		42.9	50.1	**58.8**
cat		48.2	47.1	**70.0**
chair		15.6	7.2	**22.1**
cow		37.2	44.8	**54.3**
diningtable		18.3	15.8	**27.9**
dog		43.0	49.4	**67.4**
horse		38.2	47.3	**59.4**
motorbike		52.2	36.6	**70.7**
person		40.0	36.4	**59.0**
plant		33.8	24.3	**38.7**
sheep		36.0	44.5	**58.6**
sofa		21.6	21.0	**38.1**
train		33.4	31.5	**37.6**
tv/monitor		38.3	41.3	**45.2**
average	25.7	35.6	35.8	**51.5**

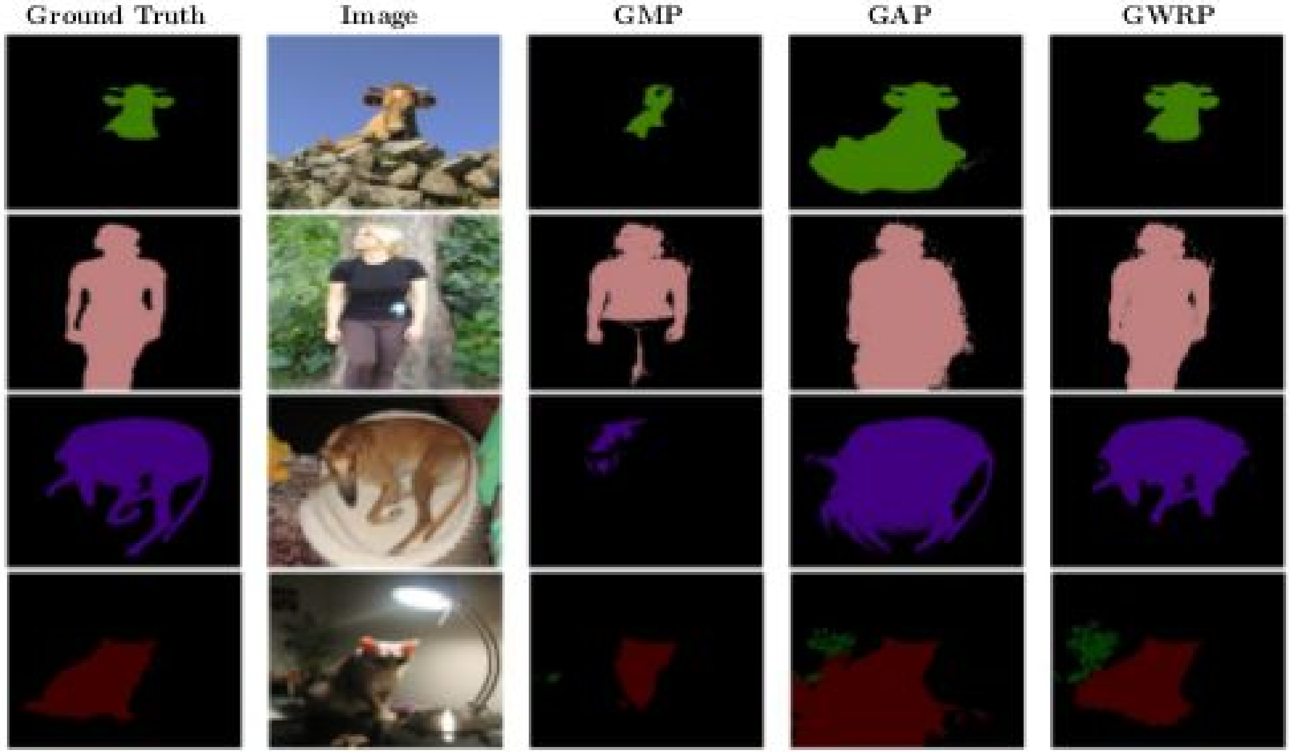
(* results from unpublished/not peer-reviewed manuscripts, † trained on ImageNet, ‡ value inferred from average)

Ablation study



loss function	mIoU (*val*)
L_{class}	27.8
L_{semi}	49.2
$L_{semi} + L_{bound}$	49.4
$L_{class} + L_{bound}$	17.2
$L_{semi} + L_{class}$	45.7
all terms	50.7

Pooling strategies



pooling method	fg fraction	mIoU *(val)*
GMP	21.0	47.3
GAP	37.5	45.1
GWRP	26.7	50.7
ground truth	27.1	–