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

D. Mukhametshina

April 19, 2019

Notation:

- $\mathcal{X}$ – space of images
- Y – a collection, $(y_1, \ldots, y_n)$, of semantic labels at n spatial locations.
- $\mathcal{C} = \mathcal{C}' \cup \{c^{\text{bg}}\}$ of size k – semantic labels
- $\mathcal{D} = \{(X_i, T_i)\}_{i=1}^{N}$ – training data
- $T_i \subset \mathcal{C}'$ – weak annotation (foreground labels)

Notation:

- $\mathcal{X}$ – space of images
- Y – a collection, $(y_1, \ldots, y_n)$, of semantic labels at n spatial locations.
- $\mathcal{C} = \mathcal{C}' \cup \{c^{\text{bg}}\}$ of size k – semantic labels
- $\mathcal{D} = \{(X_i, T_i)\}_{i=1}^{N}$ – training data
- $T_i \subset \mathcal{C}'$ – weak annotation (foreground labels)

Goal:

Train deep CNN $f(X; \theta)$, where θ is parameter:
$f_{u,c}(X; \theta) = p(y_u = c | X)$, $c \in \mathcal{C}$, $u \in \{1, 2, \ldots, n\}$.

The SEC loss for weakly supervised image segmentation

Notations:

- L_{seed} – provides localization hints to the network
- L_{expand} – penalizes the network for predicting segmentation masks with too small or wrong objects
- $L_{\text{constrain}}$ – encourages segmentations that respect the spatial and color structure of the images

Optimization problem:

$$\min_{\theta} \sum_{(X,T)\in\mathcal{D}} \left[L_{\text{seed}(f(X;\theta),T)} + L_{\text{expand}(f(X;\theta),T)} + L_{\text{constrain}(X,f(X;\theta))} \right]$$

The SEC loss for weakly supervised image segmentation

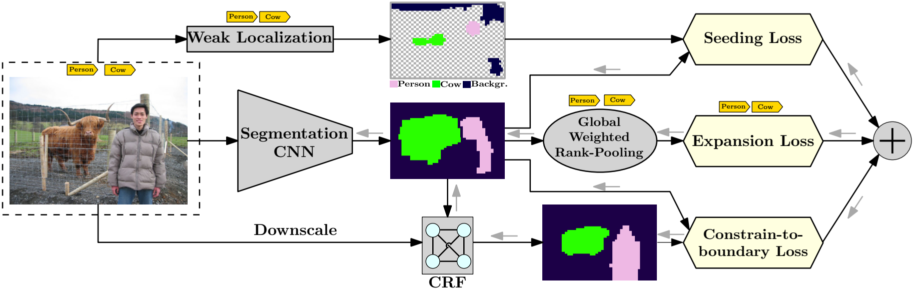


Figure: A schematic illustration of SEC that is based on minimizing a composite loss function consisting of three terms: *seeding loss*, *expansion loss* and *constrain-to-boundary loss*.

Seeding loss with localization cues.

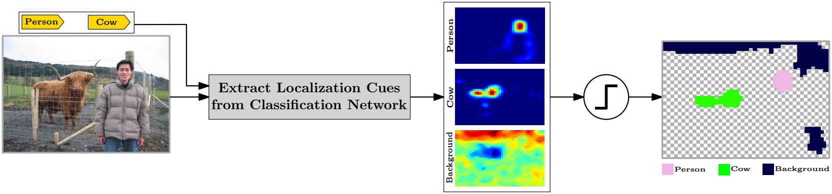


Figure: The schematic illustration of the weak localization procedure.

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

S_c – a set of locations, with class c (from weak localization)

Expansion loss with global weighted rank pooling.

Existing scorings for (c, X):

- *global max-poling* (GMP) – $\max_{u \in \{1,\ldots,n\}} f_{u,c}(X)$
- *global average-pooling* (GAP) – $\frac{1}{n}\sum_{u=1}^{n} f_{u,c}(X)$

Problems:

- GMP underestimates the sizes of objects
- GAP overestimatesthe sizes of objects

Expansion loss with global weighted rank pooling.

$I^c = \{i_1, ..., i_n\}$

$f_{i_1,c}(x) \geq f_{i_2,c}(x) \geq ... \geq f_{i_n,c}(x)$, $c \in \mathcal{C}$

$0 < d_c <= 1$ – decay parameter for class c

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

Decay parameters:

- d_+ – classes that occur in an image, ($d_+ = 0.996$)
- d_- – classes that do not occur in an image, ($d_- = 0$)
- d_{bg} – background ($d_{\text{bg}} = 0.999$)

Problems:

$$L_{\text{expand}}(f(X), T) = -\frac{1}{|T|} \sum_{c \in T} \log G_c(f(X); d_+) -$$

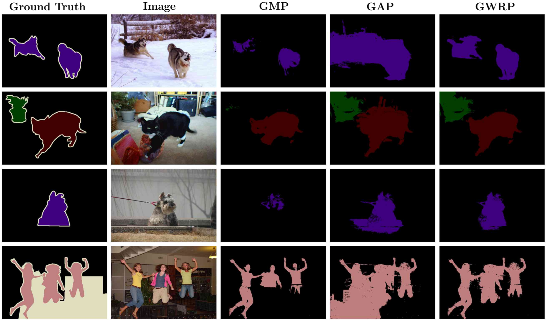
$$-\frac{1}{|\mathcal{C}' \backslash T|} \sum_{c \in \mathcal{C}' \backslash T} \log(1 - G_c(f(X); d_-)) - \log G_{c^{\text{bg}}}(f(X); d_{\text{bg}})$$

Constrain-to-boundary loss.

- Fully-connected CRF, $Q(X, f(X))$, with unary potentials given by the logarithm of the probability scores predicted by the segmentation network, and pairwise potentials of fixed parametric form that depend only on the image pixels.
- X downscaled to the resolution of the segmentation mask.
- $L_{\text{constrain}}(X, f(X)) = \frac{1}{n} \sum_{u=1}^{n} \sum_{c \in \mathcal{C}} Q_{u,c}(X, f(X)) \log \frac{Q_{u,c}(X, f(X))}{f_{u,c}(X)}$

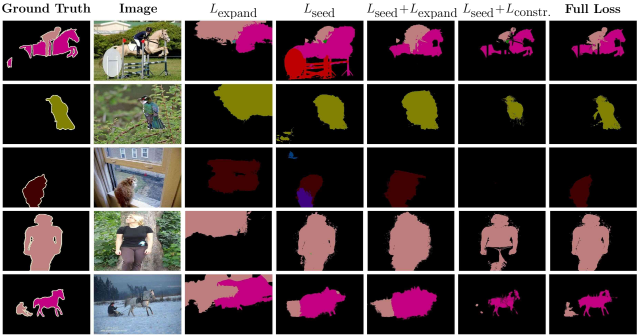
Optimization.

For training the network was used the batched stochastic gradient descent (SGD).
8000 iterations, the batch size is 15, the dropout rate is 0.5 and the weight decay parameter is 0.0005.
The initial learning rate is 0.001 and it is decreased by a factor of 10 every 2000 iterations.



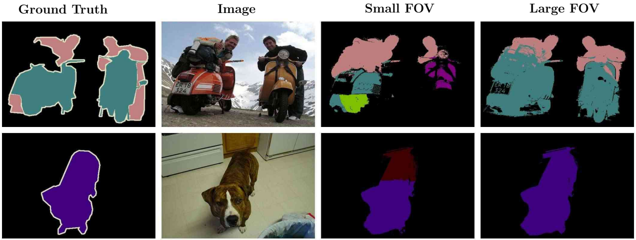
pooling method	fg fraction	mIoU (*val*)
GMP	20.4	46.5
GAP	35.6	45.7
GWRP	25.8	50.7
ground truth	26.7	–

Figure: Results on the *val* set and examples of segmentation masks for models trained with different pooling strategies.



loss function	mIoU *(val)*
L_{expand}	20.6
L_{seed}	45.4
L_{seed} + L_{expand}	44.3
L_{seed} + $L_{\text{constrain}}$	50.4
all terms	50.7

Figure: Results on the *val* set and examples of segmentation masks for models trained with different loss functions.



field of view	mIoU (*val*)
211x211	38.1
378x378	50.7

Figure: Results on the *val* set and examples of segmentation masks for models with small or large field-of-views.

THANK YOU