Deep Weakly-Supervised Learning Methods for Classification and Localization in Histology Images: A Survey

(#M004)

Jérôme Rony[1] & Soufiane Belharbi[1*] & Jose Dolz[2] & Ismail Ben Ayed[1] & Luke McCaffrey[3] & Eric Granger[1]

[1]LIVIA, Dept. of Systems Engineering, ETS Montreal, Canada
[2]LIVIA, Dept. of Software and IT Engineering, ETS, Montreal, Canada
[3]Rosalind and Morris Goodman Cancer Research Centre, Department of Oncology, McGill University

Context

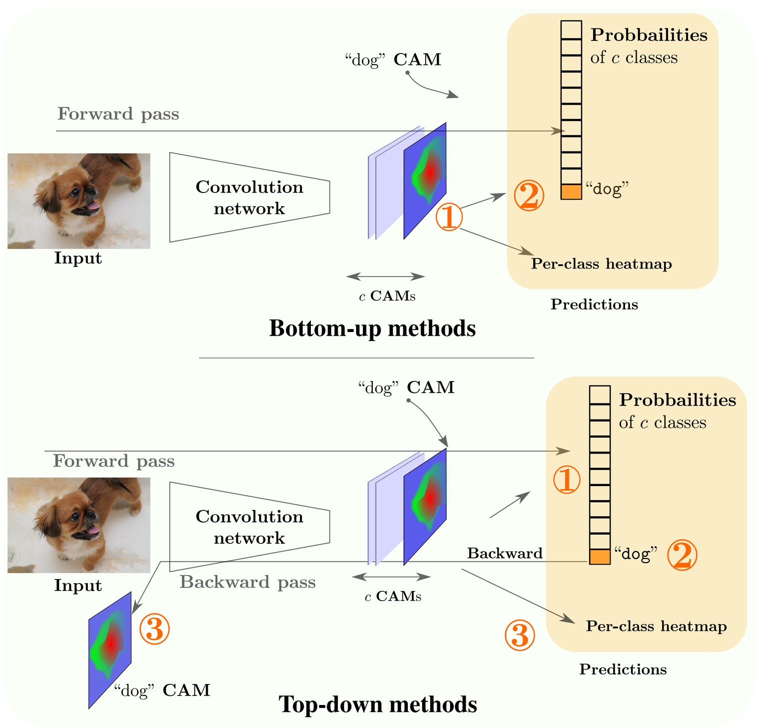
We explore weakly-supervised object localization methods in histology images and to what extent they are able to localize regions of interest (ROIs), i.e., cancerous regions, using only image label supervision.

Weak supervision: only global image class is available.

Challenges:

- Large images
- Label ambiguity
- Large stain variability
- Unstructured images
- Similar foreground/background (no salient patterns)

Taxonomy



A taxonomy of WSOL methods

- **Bottom-up WSOL**
  - Spatial pooling
    - GAP ('13)
    - MAX-Pool ('15)
    - LSE ('16)
    - CAM ('16)
    - WILDCAT ('17)
    - PRM ('18)
    - Deep MIL ('18)
  - CAM refinement
    - Data augmentation
      - HaS ('17)
      - SPN ('17)
      - AE ('17)
      - Two-Phase ('17)
      - ACoL ('18)
      - GAIN ('18)
      - CutMix ('19)
      - ADL ('19)
      - RecMin ('19)
      - PuzzleMix ('20)
      - MEIL ('20)
      - GC-Net ('20)
      - SaliencyMix ('21)
      - ScoreMix ('22)
      - MAXMIN ('22)
    - Features enhancement
      - MDC ('18)
      - FickleNet ('19)
      - DANet ('19)
      - NL-CCAM ('20)
      - I2C ('20)
      - ICL ('20)
      - CSTN ('20)
      - TS-CAM ('21)
      - Pseudo-annotation
        - SPG ('18)
        - Pair-Sim ('20)
        - PSOL ('20)
        - SPOL ('21)
        - F-CAM ('22)
        - NEGEV ('22)
- **Top-down WSOL**
  - Biologically-inspired
    - Feedback layer ('15)
    - Excitation-backprop ('18)
  - Grad-based aggregation
    - Grad-CAM ('17)
    - Grad-CAM++ ('18)
    - Smooth-Grad-CAM++ ('19)
    - XGrad-CAM ('20)
    - LayerCAM ('21)
  - Confidence-based aggregation
    - Score-CAM ('20)
    - SS-CAM ('20)
    - IS-CAM ('20)
    - Ablation-CAM ('20)

Empirical Results

Datasets: `GlaS` for colon cancer, and `CAMELYON16` for breast cancer.

Localization performance of different WSOL Methods

Methods / Metric	GlaS VGG	GlaS Inception	GlaS ResNet	GlaS Mean	CAMELYON16 VGG	CAMELYON16 Inception	CAMELYON16 ResNet	CAMELYON16 Mean
	PxAP (B-LOC)							
Bottom-up WSOL								
GAP *(corr,2013)*	58.5	57.5	56.2	57.4	37.5	24.6	43.7	35.2
MAX-Pool *(cvpr,2015)*	58.5	57.1	46.2	53.9	42.1	40.9	20.2	34.4
LSE *(cvpr,2016)*	63.9	62.8	59.1	61.9	63.1	29.0	42.1	44.7
CAM *(cvpr,2016)*	68.5	50.5	64.4	61.1	25.4	48.7	27.5	33.8
HaS *(iccv,2017)*	65.5	65.4	63.5	64.8	25.4	47.1	29.7	34.0
WILDCAT *(cvpr,2017)*	56.1	54.9	60.1	57.0	44.4	31.4	31.0	35.6
ACoL *(cvpr,2018)*	63.7	58.2	54.2	58.7	31.3	39.3	31.3	33.9
SPG *(eccv,2018)*	63.6	58.3	51.4	57.7	45.4	24.5	22.6	30.8
Deep MIL *(icml,2018)*	66.6	61.8	64.7	64.3	53.8	51.1	57.9	54.2
PRM *(cvpr,2018)*	59.8	53.1	62.3	58.4	46.0	41.7	23.2	36.9
ADL *(cvpr,2019)*	65.0	60.6	54.1	59.9	19.0	46.0	46.0	37.0
CutMix *(eccv,2019)*	59.9	50.4	56.7	55.6	56.4	44.9	20.7	40.6
TS-CAM *(corr,2021)*	t:54.5	b:57.8	s:55.1	52.8	t:46.3	b:21.6	s:42.2	36.7
MAXMIN *(tmi,2022)*	75.0	49.1	81.2	68.4	50.4	**80.8**	**77.7**	**69.6**
NEGEV *(midl,2022)*	**81.3**	**70.1**	**82.0**	**77.8**	**70.3**	53.8	52.6	58.9
Top-down WSOL								
GradCAM *(iccv,2017)*	75.7	56.9	70.0	67.5	40.2	34.4	29.1	34.5
GradCAM++ *(wacv,2018)*	76.1	65.7	70.7	70.8	41.3	43.9	25.8	37.0
Smooth-GradCAM++ *(corr,2019)*	71.3	67.6	75.5	71.4	35.1	31.6	25.1	30.6
XGradCAM *(bmvc,2020)*	73.7	66.4	62.6	67.5	40.2	33.0	24.4	32.5
LayerCAM *(ieee,2021)*	67.8	66.1	70.9	68.2	34.1	25.0	29.1	29.4
Fully supervised								
U-Net*(miccai,2015)*	96.8	95.4	96.4	96.2	83.0	82.2	83.6	82.9

Results

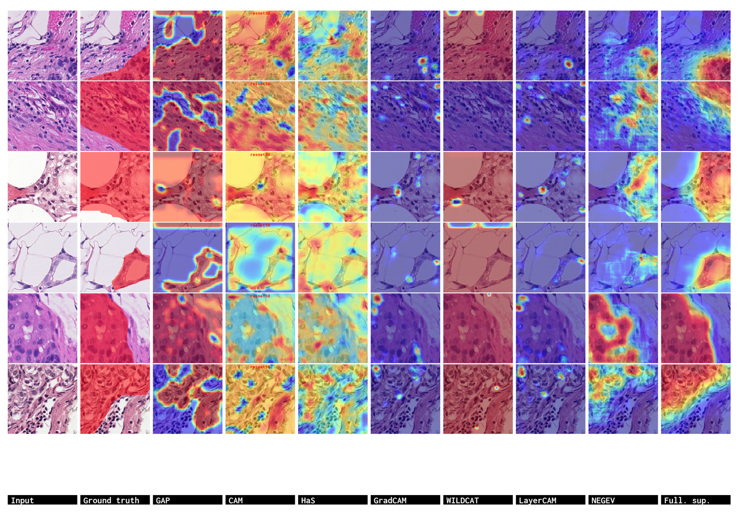


Figure 1. Predictions over **metastatic** test samples for `CAMELYON16`.

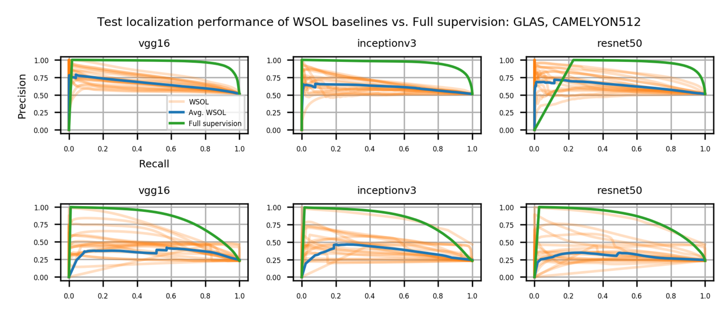


Figure 3. Localization sensitivity to thresholding: WSOL methods (orange), average WSOL methods (blue), fully supervised method (green). Top: `GlaS`. Bottom: `CAMELYON16`.

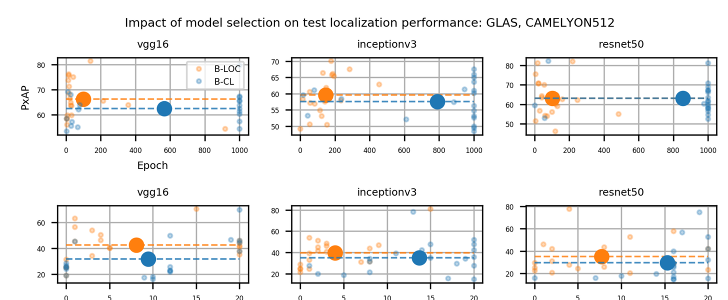


Figure 2. **Localization**: Impact of model selection (`B-LOC`: orange. vs. `B-CL`: blue) over test **localization** (`PxAP`) performance. Each point indicates the epoch (x-axis) at which the best model is selected and its corresponding localization performance (y-axis). Large circles indicate the average over all WSOL methods. Top: `GlaS`. Bottom: `CAMELYON16`.

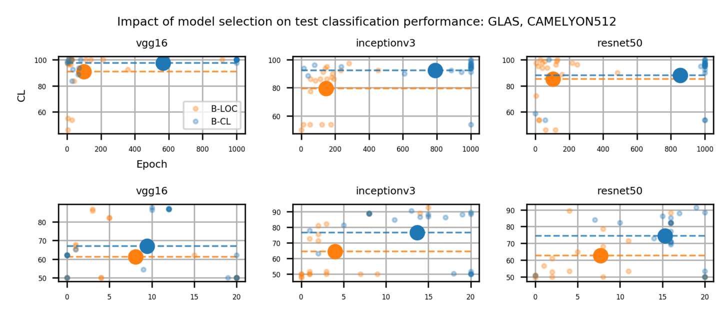


Figure 4. **Classification**: Impact of model selection (`B-LOC`: orange. vs. `B-CL`: blue) over test **classification** (`CL`) performance. Each point indicates the epoch (x-axis) at which the best model is selected and its corresponding classification performance (y-axis). Large circles indicate the average over all WSOL methods. Top: `GlaS`. Bottom: `CAMELYON16`.

Ongoing Challenges for WSOL in Histology Data

- Under activation (high false negative), Over activation (high false positive)
- Sensitivity to thresholding
- Model selection

Directions:

- Unsupervised size constraints
- Pseudo-labels
- Validation free

Code: https://github.com/jeromerony/survey_wsl_histology

soufiane.belharbi.1@ens.etsmtl.ca