F-CAM: Full Resolution Class Activation Maps via Guided Parametric Upscaling (#1177)

Soufiane Belharbi[1*] & Aydin Sarraf[3] & Marco Pedersoli[1] & Ismail Ben Ayed[1] & Luke McCaffrey[2] & Eric Granger[1]

[1]LIVIA, Dept. of Systems Engineering, École de technologie supérieure, Montreal, Canada
[2]Rosalind and Morris Goodman Cancer Research Centre, Department of Oncology, McGill University
[3]Ericsson, Global AI Accelerator, Montreal, Canada

Context

State-of-the-art methods for weakly supervised object localization (WSOL) rely on CAMs because they are:

- easy to obtain, interpretable, require only global class-image labels

However, they have several drawbacks:

- low resolution (downscale factor up to 32), inaccurate boundaries, cover minimal discriminant regions of object, sensitive to thresholding
- the low resolution of CAM limits their localization potential (`see paper for simulations`). Interpolation is required for full resolution, leading to bloby localization, and inaccurate boundaries

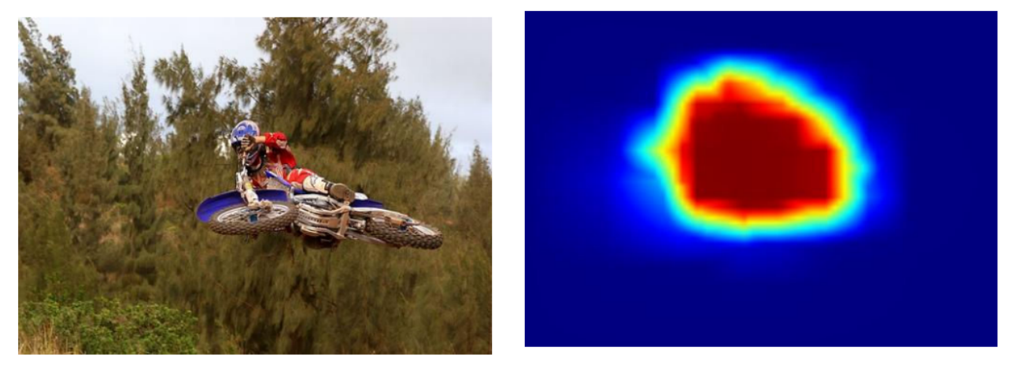

Figure 1. Example of a CAM localization with bloby, inaccurate boundaries.

This work aims to improve the resolution of CAMs using *parametric upscaling with priors*, allowing to improve localization accuracy

F-CAM: Full Resolution CAMs

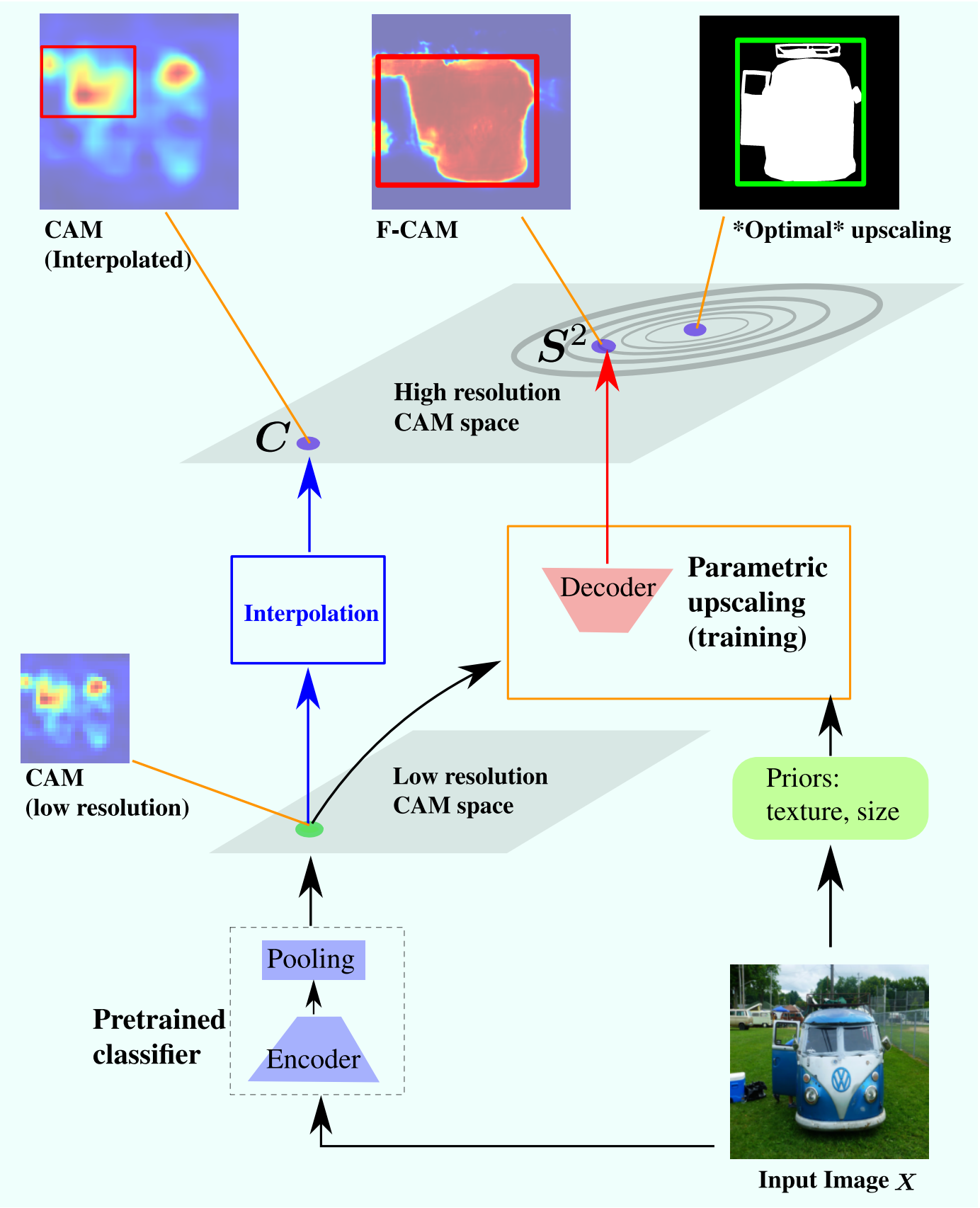

Figure 2. An illustration of the differences between interpolation and our trainable parametric upscaling with priors. $\boldsymbol{C}$ is the interpolated CAM, and $\boldsymbol{S}^2$ is the F-CAM produced using our proposed trainable decoder architecture.

- Avoid interpolation. Use parametric upscaling
- Guide upscale learning using priors: seeds, texture, size

Proposed Approach: Architecture

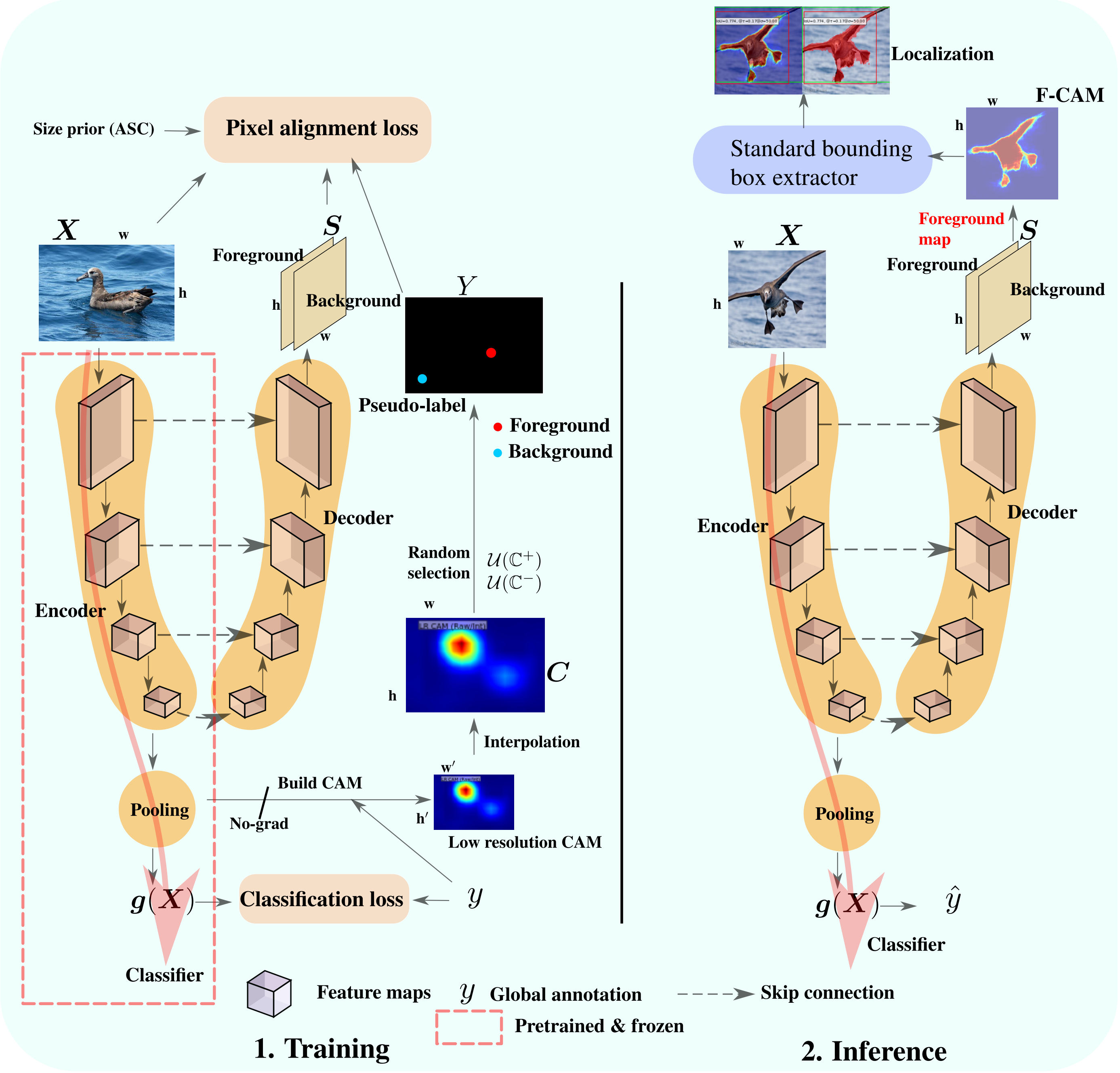

Figure 3. Our proposal: training (left), inference (right).

Proposed Approach: Training

Exploit a pre-trained classifier.

Priors:

- Align pixels via sampled pseudo-labels from the low resolution CAM.
- Use image color/texture to yield consistent boundaries.
- Use unsupervised size prior to yield complete object.

Total training loss:

$$\underbrace{\text{Seed alignment}}_{\text{pixel partial cross-entropy}} + \underbrace{\text{color/texture consistency}}_{\text{CRF}} + \text{Size constraint} \tag{1}$$

$$\begin{aligned} \min_{\boldsymbol{\theta}} \quad & -\log(\boldsymbol{g}(\boldsymbol{X})[y]) + \alpha \sum_{p \in \Omega'} \boldsymbol{H}(Y_p, \boldsymbol{S}_p) + \lambda\, \mathcal{R}(\boldsymbol{S}, \boldsymbol{X}) \;, \\ \text{s.t.} \quad & \sum \boldsymbol{S}^r \geq 0 \;, \quad r \in \{1, 2\} \;, \end{aligned} \tag{2}$$

CRF: $\mathcal{R}$. partial cross-entropy: $\boldsymbol{H}$.

Results

Datasets: CUB, OpenImages.

Methods	CUB (`MaxBoxAcc`) VGG	Inception	ResNet	Mean	OpenImages (`PxAP`) VGG	Inception	ResNet	Mean
CAM *(cvpr,2016)*	71.1	62.1	73.2	68.8	58.1	61.4	58.0	59.1
HaS *(iccv,2017)*	76.3	57.7	78.1	70.7	56.9	59.5	58.2	57.8
ACoL *(cvpr,2018)*	72.3	59.6	72.7	68.2	54.7	63.0	57.8	58.4
SPG *(eccv,2018)*	63.7	62.8	71.4	66.0	55.9	62.4	57.7	58.6
ADL *(cvpr,2019)*	75.7	63.4	73.5	70.8	58.3	62.1	54.3	58.2
CutMix *(eccv,2019)*	71.9	65.5	67.8	68.4	58.2	61.7	58.7	59.5
Best WSOL	76.3	65.5	78.1	70.8	58.3	63.0	58.7	59.5
FSL baseline	86.3	94.0	95.8	92.0	61.5	70.3	74.4	68.7
Center baseline	59.7	59.7	59.7	59.7	45.8	45.8	45.8	45.8
CSTN *(icpr,2020)*	Resnet101: 76.0				–	–	–	–
TS-CAM *(corr,2021)*	Deit-S: 83.8				–	–	–	–
MEIL *(cvpr,2020)*	73.8	–	–	–	–	–	–	–
DANet *(iccv,2019)*	67.7	67.03	–	–	–	–	–	–
SPOL *(cvpr,2021)*	–	–	96.4	–	–	–	–	–
CAM* *(cvpr,2016)*	61.6	58.8	71.5	63.9	53.0	62.7	56.8	57.5
GradCAM *(iccv,2017)*	69.3	62.3	73.1	68.2	59.6	63.9	60.1	61.2
GradCAM++ *(wacv,2018)*	84.1	63.3	81.9	76.4	60.5	64.0	60.2	61.5
Smooth-GradCAM++ *(corr,2019)*	69.7	66.9	76.3	70.9	52.2	61.7	54.3	56.0
XGradCAM *(bmvc,2020)*	69.3	60.9	72.7	67.6	59.0	63.9	60.2	61.0
LayerCAM *(ieee,2021)*	84.3	66.5	85.2	78.6	59.5	63.5	61.1	61.3
CAM* + ours	87.3	82.0	90.3	86.5	67.8	71.9	72.1	70.6
GradCAM + ours	87.5	84.4	90.5	87.4	68.6	70.0	70.9	69.8
GradCAM++ + ours	91.5	84.6	91.0	89.0	64.8	67.1	66.3	66.0
Smooth-GradCAM++ + ours	89.1	86.8	90.7	88.8	60.3	65.4	64.4	63.3
XGradCAM + ours	86.8	84.4	90.4	88.8	68.7	71.3	70.4	70.1
LayerCAM + ours	91.0	85.3	92.4	89.7	64.3	64.9	65.3	64.8
Best WSOL + ours	91.5	86.8	92.4	89.7	68.7	71.9	72.1	70.6

Table 1. Performance on `MaxBoxAcc` and `PxAP` metrics.

Qualitative results: First row: WSOL baseline. Second row: WSOL baseline + ours. First column: CAM. Next column: localization. Colors: predicted boxes in red, and true box in green. Thresholded mask is in red. $\sigma = 50\%$.

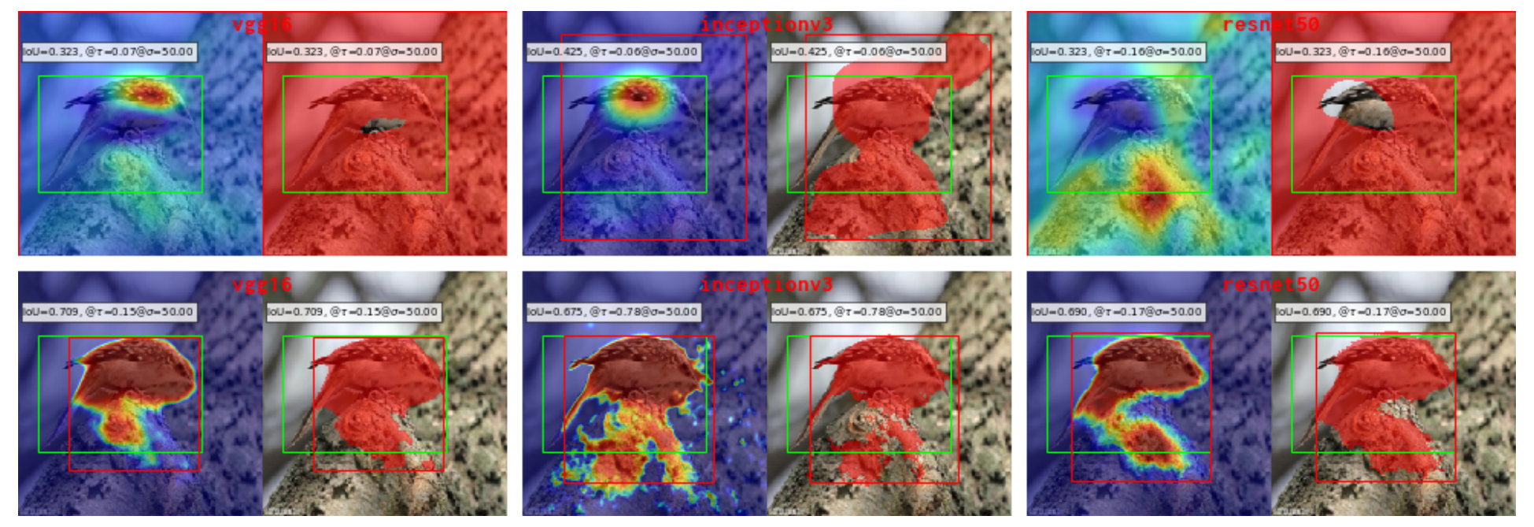

Figure 4. Test samples from CUB (CAM* method).

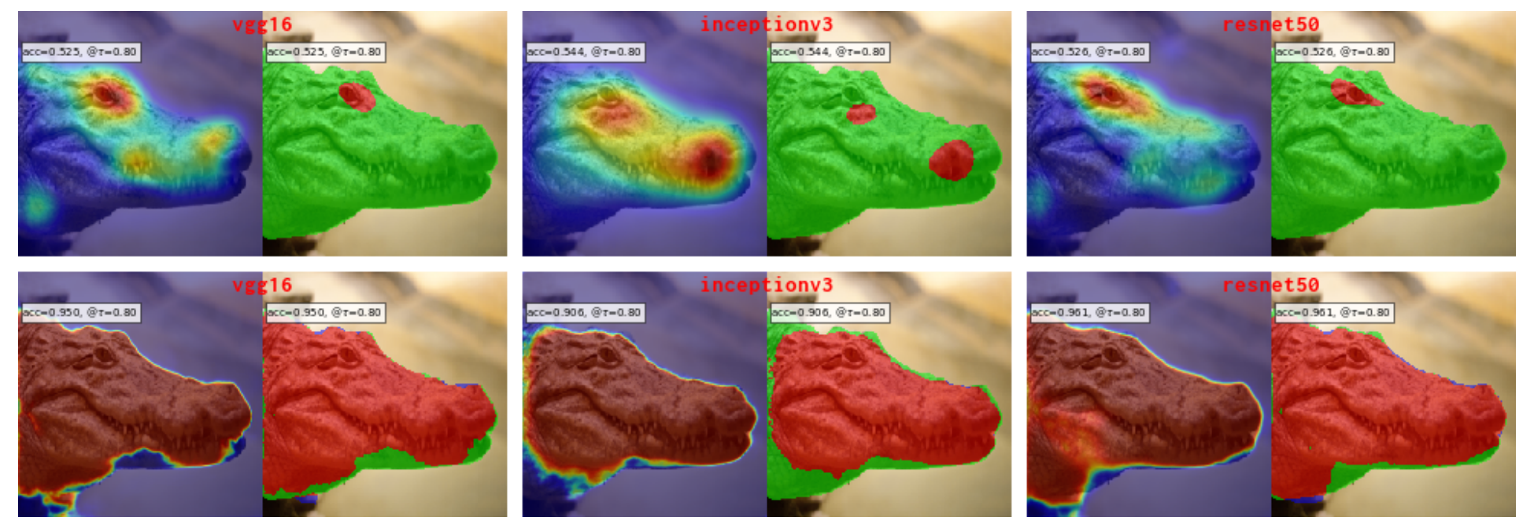

Figure 5. Test samples from OpenImages (LayerCAM method).

Code: https://github.com/sbelharbi/fcam-wsol

WACV 2022

soufiane.belharbi.1@ens.etsmtl.ca