Deep Learning Models for Weakly-Supervised Object Localization and Segmentation

Soufiane Belharbi, Eric Granger and Ismail Ben Ayed

ICPR 2022 Tutorial
August 21, 2022
Montreal, Canada

ÉTS
Le génie pour l'industrie

Overview

1) **Introduction**
2) **Review of WSOL Methods**
   a) bottom-up and top-down methods
   b) case studies

Coffee Break (pause at 10h for 30 mins)

3) **Review of WSSS Methods**
4) **Applications of WSOL / WSSS**
5) **Key Challenges and Future Directions**

Le génie pour l'industrie

LIVIA

LABORATOIRE D'IMAGERIE, DE VISION ET D'INTELLIGENCE ARTFICIELLE

Belharbi, Soufiane: Post-doc. LIVIA, Dept. of Systems Engineering, ETS Montreal, Canada.

Areas: Machine Learning, Deep Learning, Computer Vision.

Granger, Eric: Professor, LIVIA, Dept. of Systems Engineering, ETS Montreal, Canada.

Areas: Machine Learning, Pattern Recognition, Computer Vision, Information Fusion, Affective Computing, Biometrics, Video Surveillance

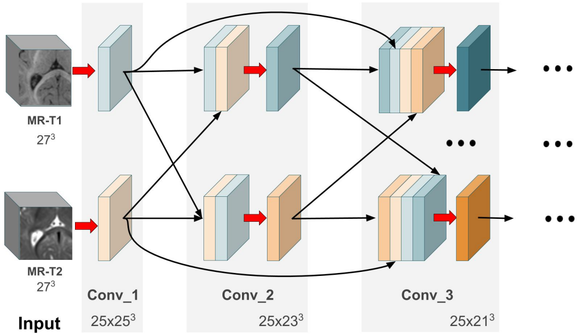
Ben Ayed, Ismail: Professor, LIVIA, Dept. of Systems Engineering, ETS Montreal, Canada.

Areas: Computer Vision, Machine Learning, Pattern Recognition, Optimization, Medical Image Analysis, Information Theory

http://liviamtl.ca/

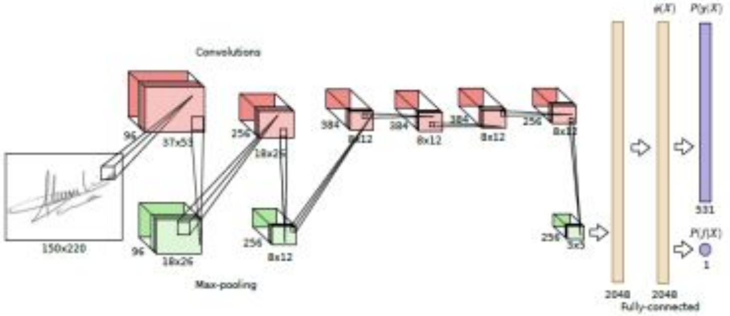
ÉTS Le génie pour l'industrie

LABORATOIRE D'IMAGERIE, DE VISION ET D'INTELLIGENCE ARTFICIELLE



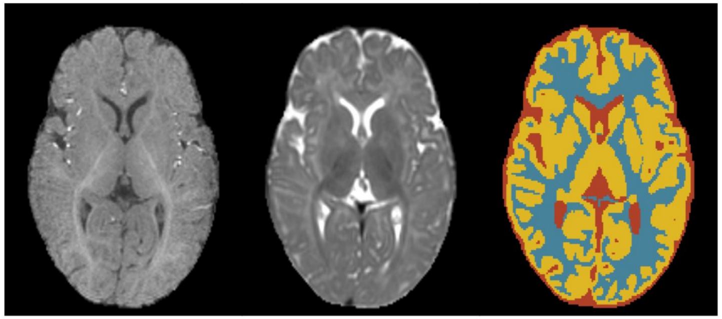
Research Axes:

- machine learning
- computer vision: perception in 2D and 3D scenes
- pattern recognition in static and dynamically-changing environments
- information fusion
- optimization of complex systems



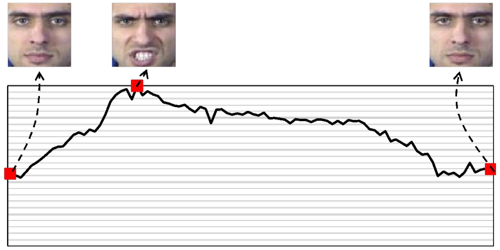
Le génie pour l'industrie

Application Areas:



- analysis of medical, aerial images
- video analytics and surveillance

- biometrics (face, voice and signature)
- document analysis
- affective computing



Part 1

Introduction

- Weakly-Supervised Learning
- Focus of this Tutorial

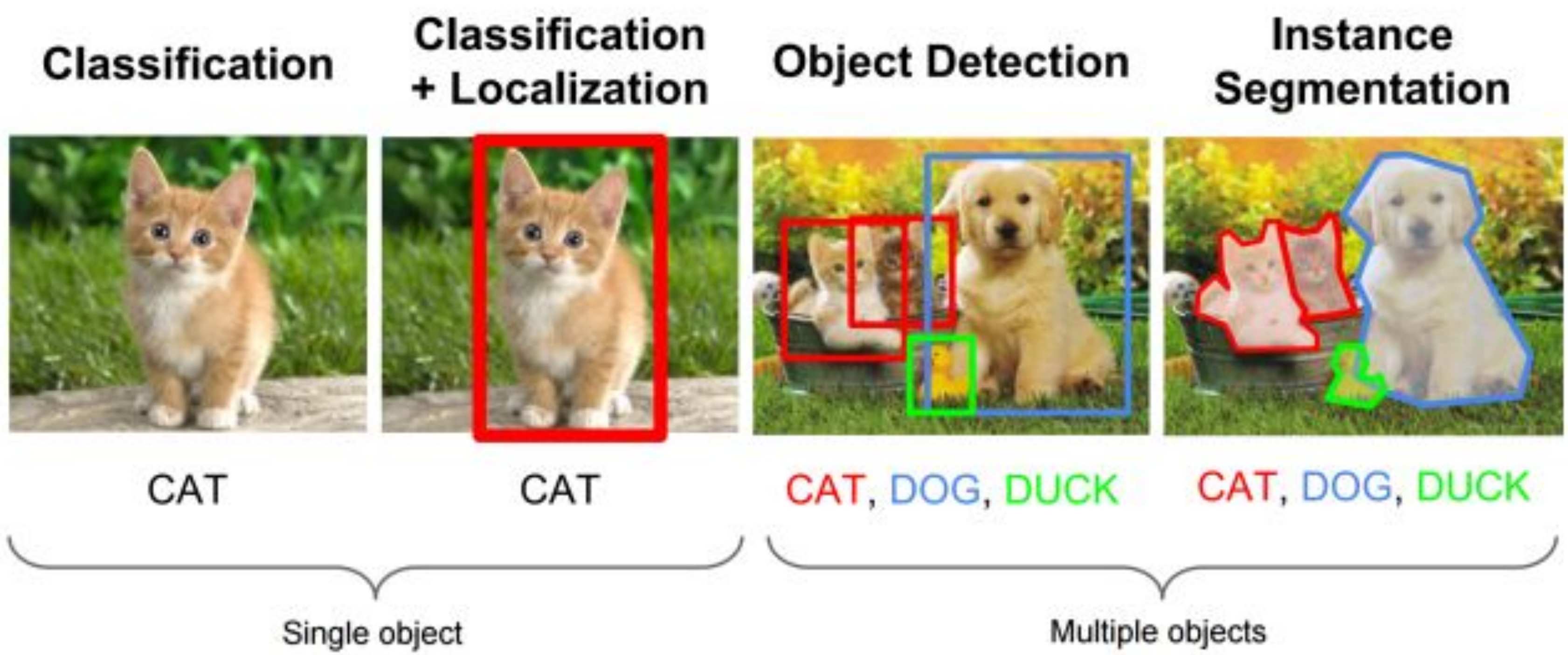
ÉTS
Le génie pour l'industrie

Visual Recognition Tasks

Differences between localization, detection and segmentation

- OL: aims to locate the main (or most visible) object in an image
- OD/IS: amis to find all the instances of objects and their boundaries

Semantic segmentation associates every pixel of an image with a class label

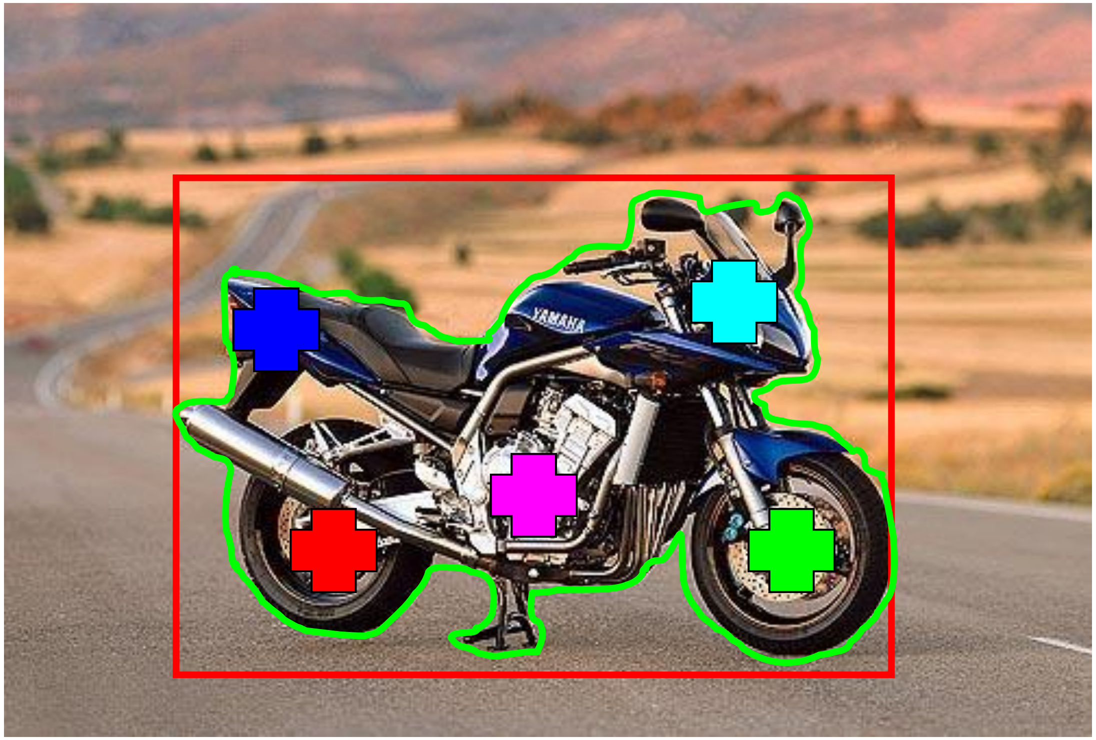


https://towardsdatascience.com/object-localization-in-overfeat-5bb2f7328b62

Visual Recognition Tasks

- **Supervised learning**: images in the training set are annotated with the “correct answer” that the model is expected to produce

Contains a motorbike



Levels of annotations:

- Image labels: for a classification task
- Patches or bounding boxes: for a detection or localization tasks
- Points: for a point-wise localization task
- Pixel regions: for a segmentation task

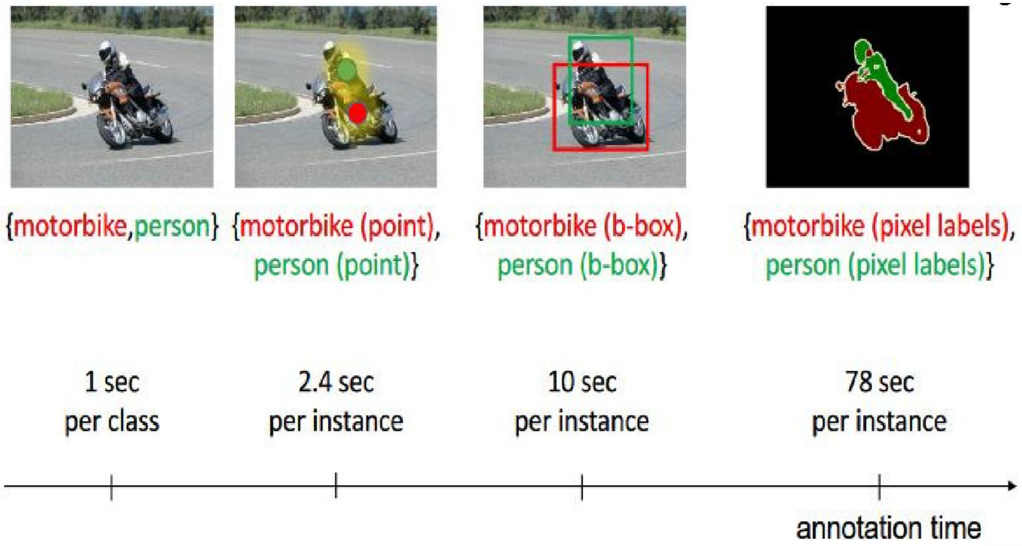
Visual Recognition Tasks

- **Why learn from data with *weak annotations?***

To optimize large DL models end-to-end, and this requires much data to determine all the model parameters

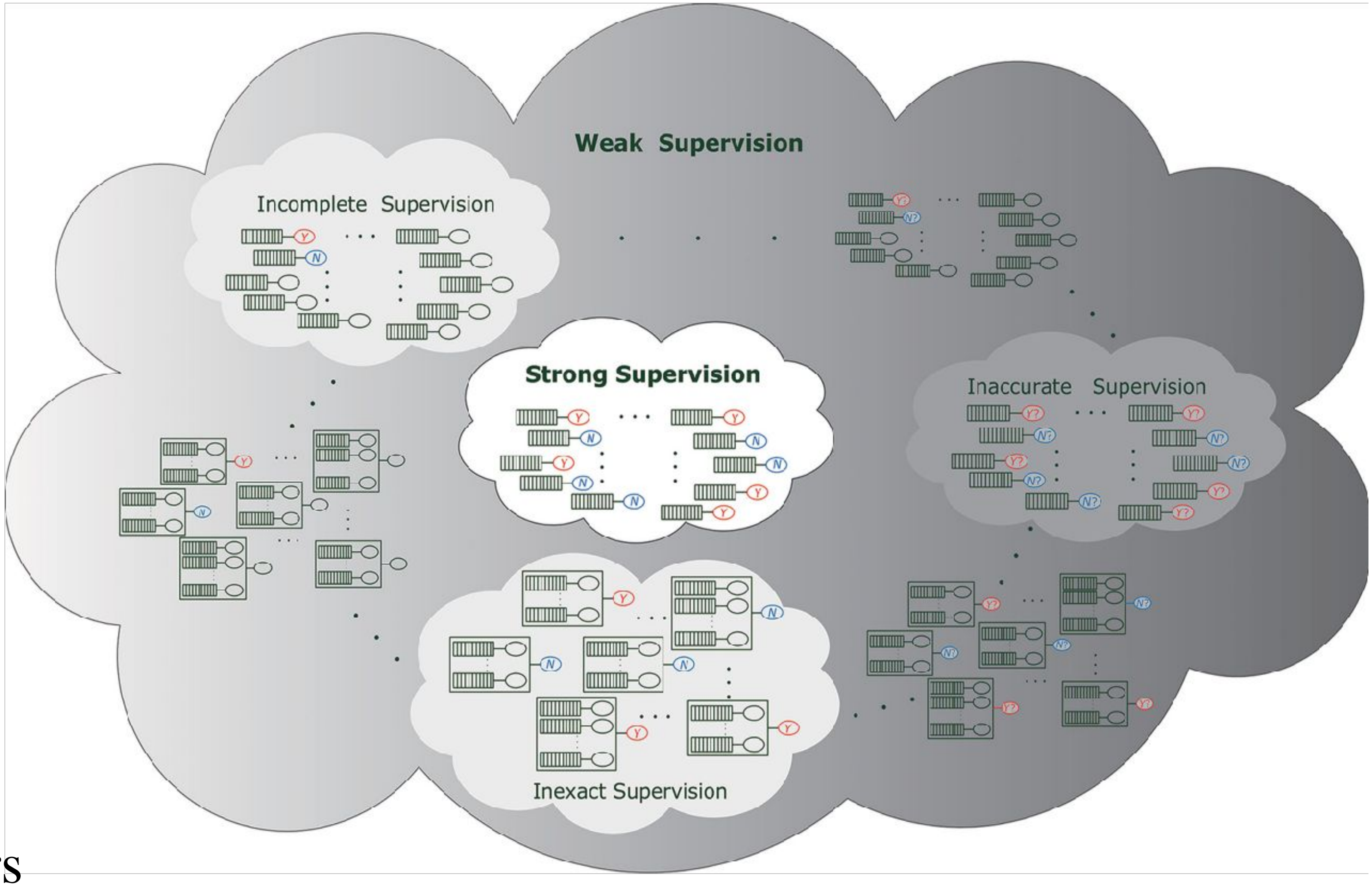
Collection and annotation of labeled training data generally costly

- full supervision may involve outlining objects, patches or marking points, pixels with category labels, etc.
- labels are ambiguous in some applications



Source: Bearman, Amy, et al. "What's the point: Semantic segmentation with point supervision." ECCV 2016.

Weak Supervised Learning Scenarios



- bars = vectors
- red/blue ovals = labels
- "?" = inaccurate labels

Weak Supervised Learning Scenarios

1) ***Incomplete supervision***: when only a small subset of training data has labels, although unlabelled data is abundant

WSL techniques:

- AL (active learning)
- SSL (semi-supervised learning)

2) ***Inexact supervision***: when training on labelled data with coarse labels

WSL technique:

- MIL (multiple instance learning)

3) ***Inaccurate supervision*****:** when labels may suffer from errors or noise

WSL techniques:

- data-editing methods
- crowdsourcing with majority vote

Weak Supervised Learning Techniques

1) ***Incomplete supervision***:

- **AL (active learning):** query an expert to label most relevant samples
- **SSL (semi-supervised learning):** involves training a model using both fully labeled and unlabeled examples

2) ***Inexact supervision***:

- **MIL (multiple-instance learning):** uses training examples grouped into sets (*bags*). Supervision is provided only for an entire set of instances.

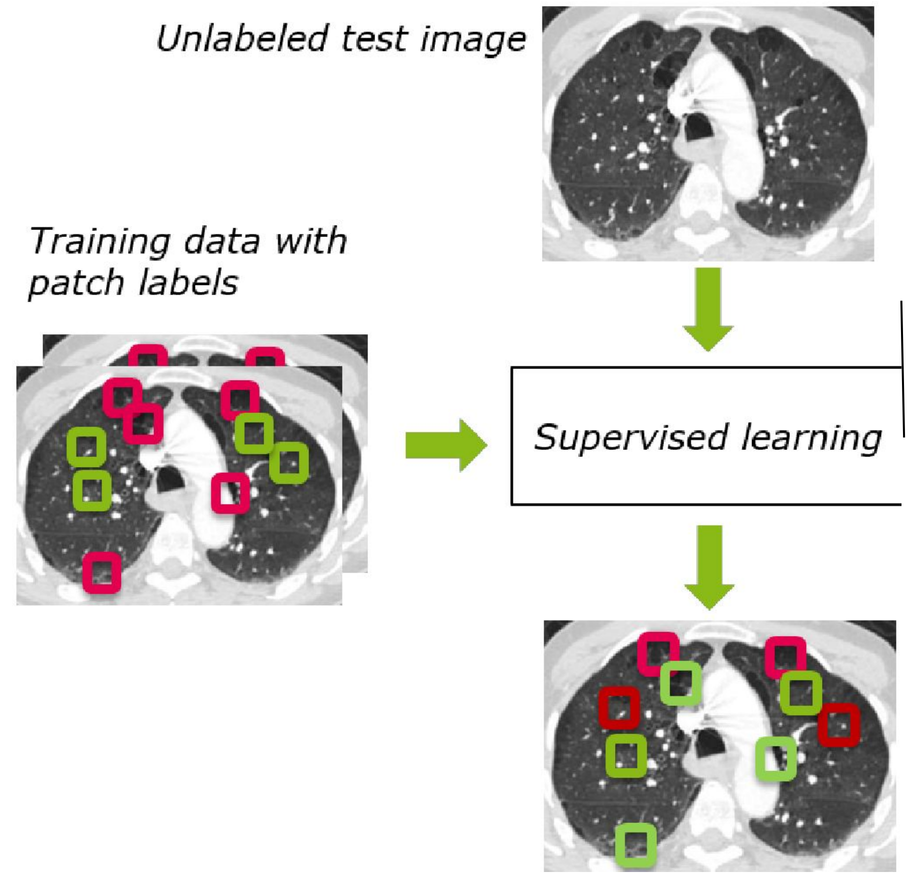
3) ***Inaccurate supervision***:

- **Data-editing methods:** determine outlier annotations
- **Crowdsourcing with majority vote:** synthesis of responses from a large population of annotators

Weak Supervised Learning Techniques

Example: classifying healthy (green) vs emphysema (red) patches of tissue in chest CT images

1) *Supervised learning:* labeled healthy and abnormal patches available

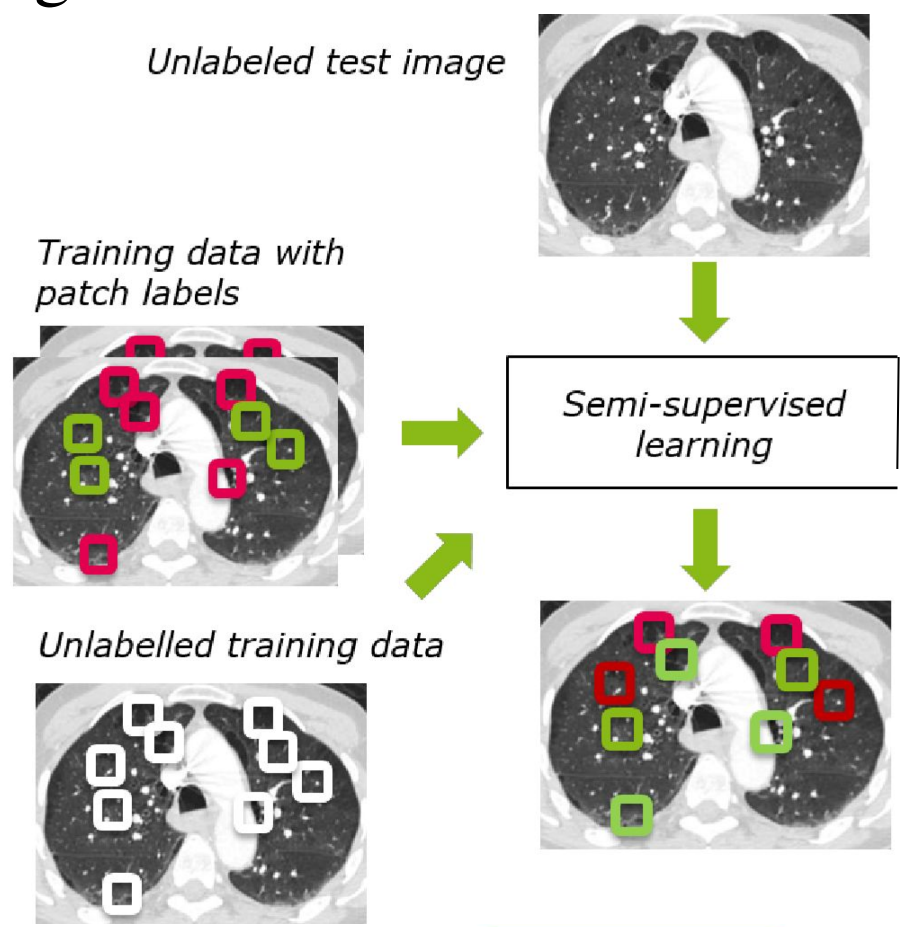


Source: V. Cheplygina et al., 'Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis,' *Medical Image Analysis*, 2019.

Weak Supervised Learning Techniques

Example: classifying healthy (green) vs emphysema (red) patches of tissue in chest CT images

2) Semi-supervised learning: in addition to healthy and abnormal patches, unlabeled patches are available.

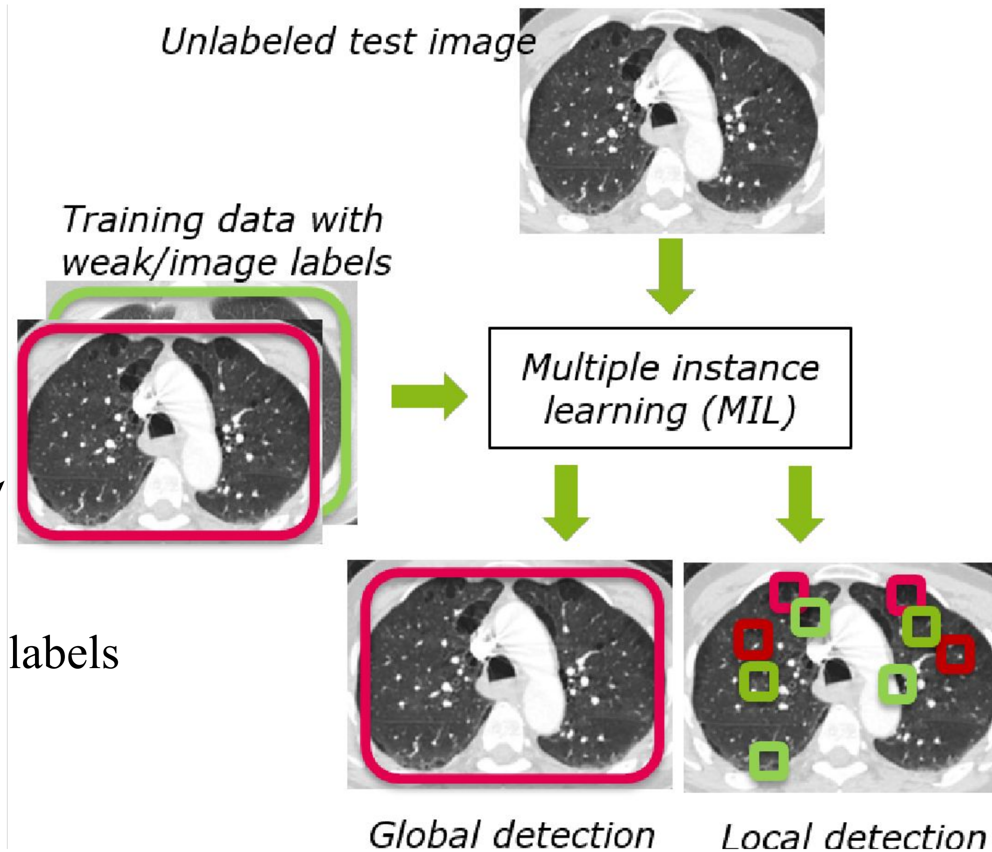


Source: V. Cheplygina et al., 'Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis,' *Medical Image Analysis*, 2019.

Weak Supervised Learning Techniques

Example: classifying healthy (green) vs emphysema (red) patches of tissue in chest CT images

3) Multiple instance learning: labeled patches are not available, but Image-level labels are.



global Image class labels

Source: V. Cheplygina et al., ‘Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis,’ *Medical Image Analysis*, 2019.

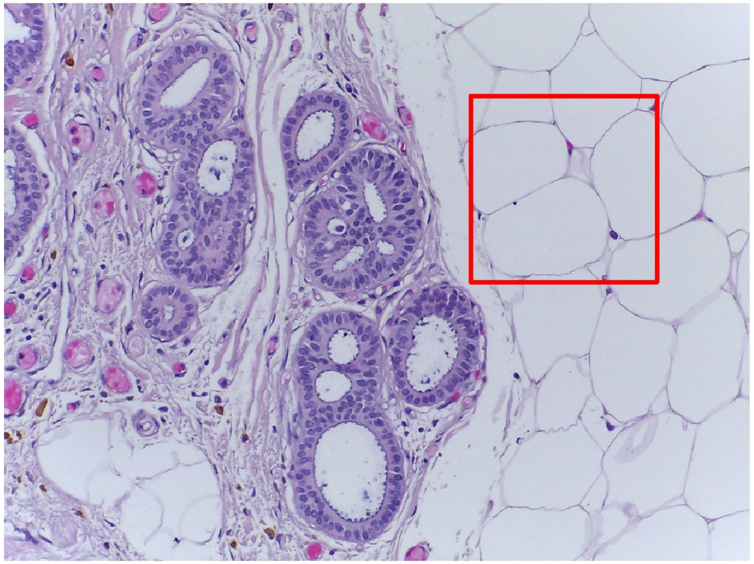
Computer Aided Diagnosis

- **Example:** MIL in computer aided diagnosis

  given a large histological image, predict if a subject is diseased or healthy and locate regions of interest

- **Image (bag) =** set of segments or patches (instances) with global annotation

- **Database -** example of an image from the ICIAR BACH Challenge 2018



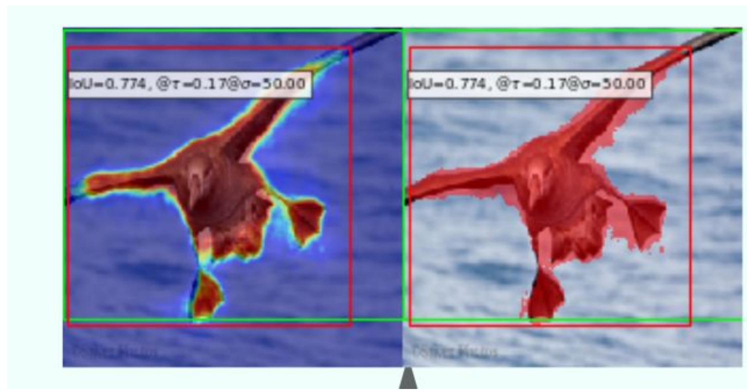
Source: V. Cheplygina et al., 'Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis,' *Medical Image Analysis*, 2019.

Focus of this Tutorial

Section 2:

Weakly-Supervised Object Localization (WSOL):

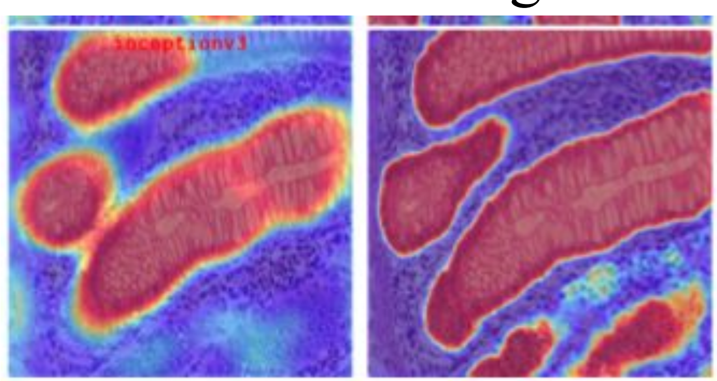
Localizing objects based on training data with global image-class labels



Section 3:

Weakly-Supervised Semantic Segmentation (WSSS)

segmenting objects based on training data with global image-class labels



Le génie pour l'industrie

Part 2

Review of WSOL Methods

- **WSOL literature: bottom-up and top-down methods**
- **Case studies:**
  a) F-CAM for improved interpolation
  b) Transformer-based models

Le génie pour l'industrie

Part 2. Review of WSOL methods: Setup

Supervised object localization

- Regression task
- One object
- Supervision: bounding box

Part 2. Review of WSOL methods: Setup

Supervised object localization

- Regression task
- One object
- Supervision: bounding box [high cost, prevents scaling up]

Is there other type of CHEAP SUPERVISION ?

Le génie pour l'industrie

Part 2. Review of WSOL methods: Setup

Supervised object localization

- Regression task
- One object
- Supervision: bounding box [high cost, prevents scaling up]

Is there other type of CHEAP SUPERVISION ?

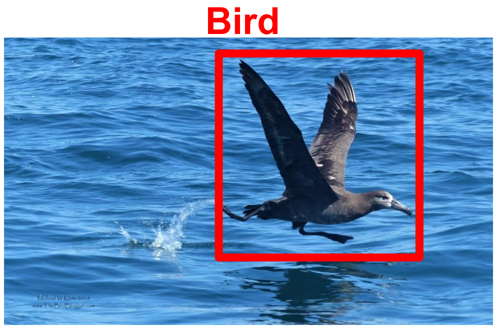
Yes, global image class !!! – *Weak supervision*

Part 2. Review of WSOL methods: Setup

Weakly Supervised object localization: WSOL

Input

WSOL model



Bounding box + image class

- One object
- Supervision: Image class
- Output: Bounding box + image class

Part 2. Review of WSOL methods: Setup

Weakly Supervised object localization: WSOL

Image
+
Class label

Train set

Image
+
Class label
+
Bounding box

Test set

Part 2. Review of WSOL methods: Setup

Weakly Supervised object localization: WSOL

Image
+
Class label

Train set

Image
+
Class label
+
Bounding box

Test set

Evaluate: classification, localization, classification + localization

Datasets:

- CUB: birds species (200 classes)
- Imagenet-1k: common objects (1k classes)

Le génie pour l'industrie

Part 2. Review of WSOL methods: Setup

Weakly Supervised object localization: WSOL

How a bounding box is produced in WSOL?

Le génie pour l'industrie

Part 2. Review of WSOL methods: Setup

How a bounding box is produced in WSOL?

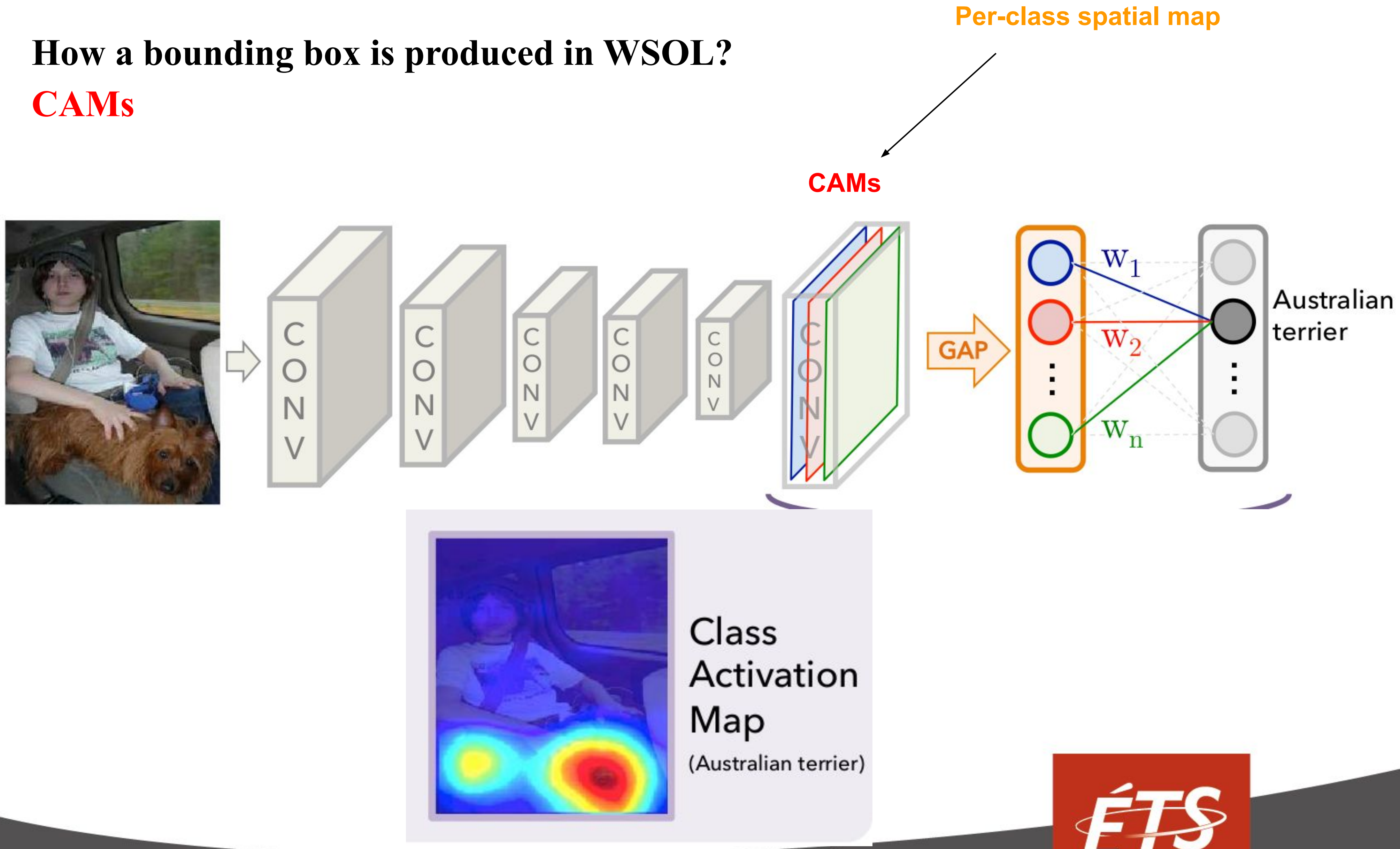
- Deep learning methods
- Class Activation Maps (CAMs)

Le génie pour l'industrie

Part 2. Review of WSOL methods: Setup

How a bounding box is produced in WSOL?

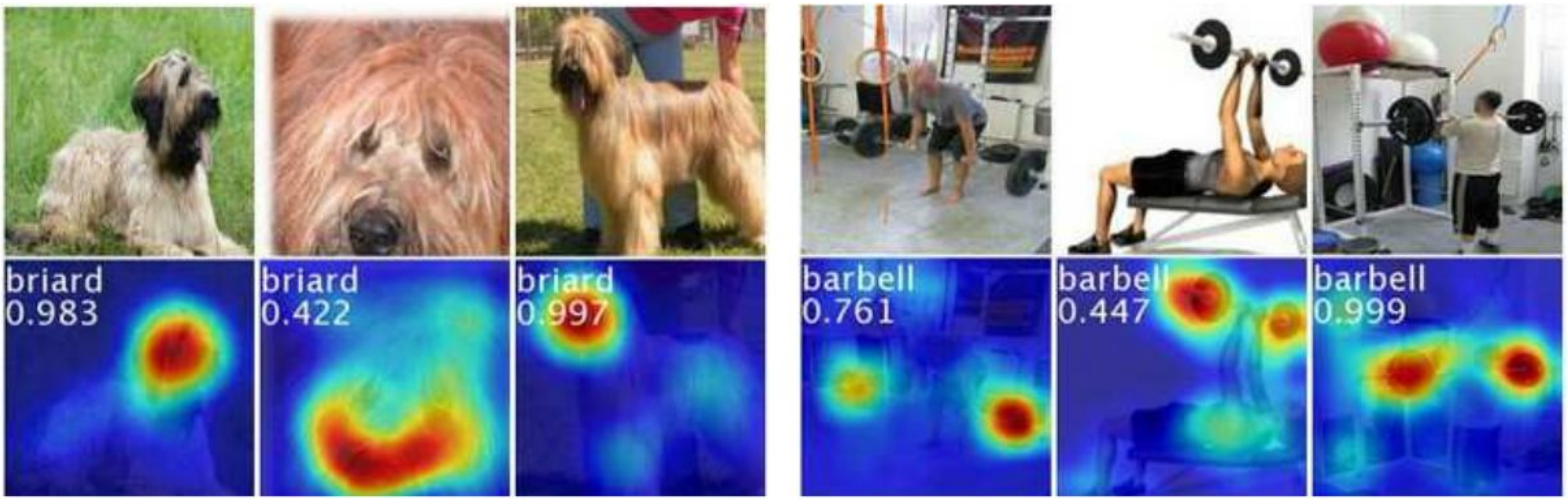
CAMs



Part 2. Review of WSOL methods: Setup

How a bounding box is produced in WSOL?

CAMs



Part 2. Review of WSOL methods: Setup

How a bounding box is produced in WSOL?

CAMs

Input

Original CAM

Interpolated CAM

Le génie pour l'industrie

Part 2. Review of WSOL methods: Setup

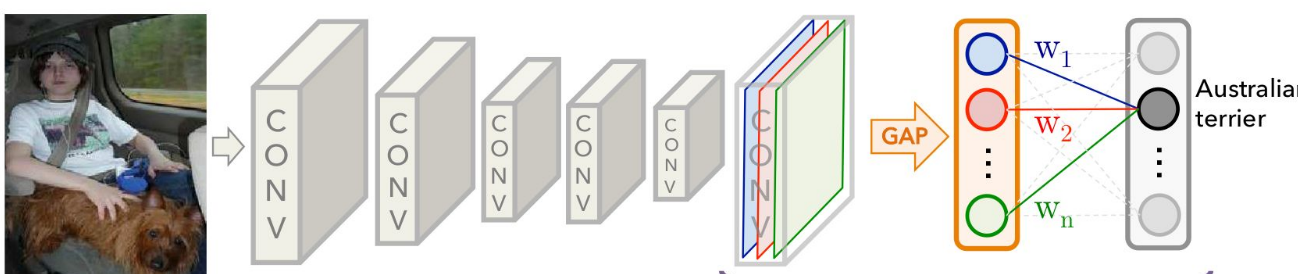
How a bounding box is produced in WSOL?

CAMs

Input

Original CAM

Interpolated CAM



Smaller map size

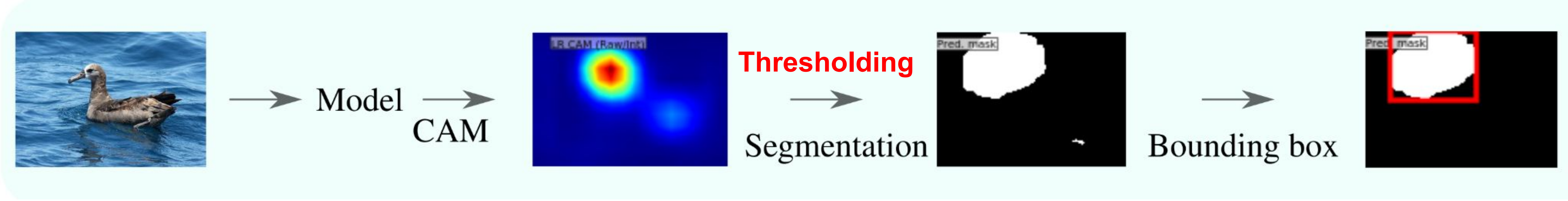
Learning Deep Features for Discriminative Localization. Zhou et al. CVPR, 2016

Part 2. Review of WSOL methods: Setup

How a bounding box is produced in WSOL?

CAMs

Image processing, contours, tightest bbx

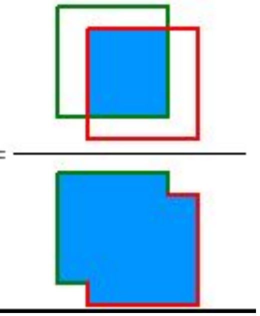


Part 2. Review of WSOL methods: Setup

WSOL: Evaluation/metrics

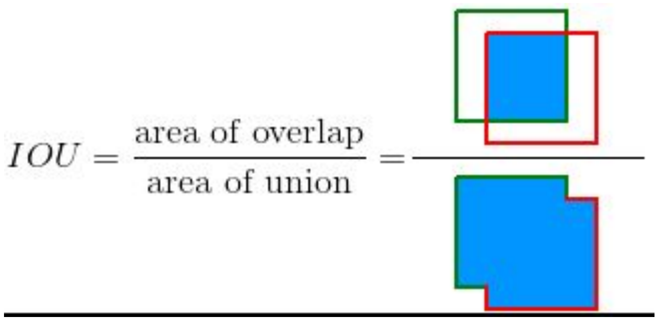
Intersection Over Union

$$IOU = \frac{\text{area of overlap}}{\text{area of union}} =$$



Part 2. Review of WSOL methods: Setup

WSOL: Evaluation/metrics



Standard **localization** metric:

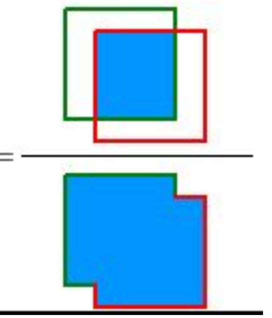
- **CorLoc**: Correct localization
  =
- **GT-known** localization accuracy

1 if IOU >= δ = 0.5 else 0

Part 2. Review of WSOL methods: Setup

WSOL: Evaluation/metrics

$$IOU = \frac{\text{area of overlap}}{\text{area of union}} =$$



Standard **localization** metric:

- **CorLoc**: Correct localization
  =
- **GT-known** localization accuracy

1 if IOU >= δ = 0.5 else 0

In CVPR 2020, 2 new **localization** metrics:

- MaxBoxAcc
- MaxBoxAccV2

Take in consideration CAM thresholding

(localization via CAMs is threshold-dependent)

Evaluating Weakly Supervised Object Localization Methods Right. Choe et al. CVPR 2020.

Evaluating Weakly Supervised Object Localization Methods Right. Choe et al. CVPR 2020.

Part 2. Review of WSOL methods: Setup

WSOL: Evaluation/metrics

WSOL **localization** metric: **MaxBoxAcc**

Normalized CAM

IoU threshold

G-truth bbox

$$\text{BoxAcc}(\tau, \delta) = \frac{1}{N} \sum_{n} 1_{\text{IoU}\left(\text{box}(s(\mathbf{X}^{(n)}), \tau), B^{(n)}\right) \geq \delta}$$

CAM threshold

Tightest bbox around largest connected component

Evaluating Weakly Supervised Object Localization Methods Right. Choe et al. CVPR 2020.

Part 2. Review of WSOL methods: Setup

WSOL: Evaluation/metrics

WSOL **localization** metric: **MaxBoxAcc**

$$\texttt{BoxAcc}(\tau, \delta) = \frac{1}{N} \sum_n 1_{\mathrm{IoU}\left(\mathrm{box}(s(\mathbf{X}^{(n)}), \tau), B^{(n)}\right) \geq \delta}$$

$$\texttt{MaxBoxAcc}(\delta) := \max_\tau \texttt{BoxAcc}(\tau, \delta)$$

τ in [0, 1], step: 0.001.
δ = 0.5

Evaluating Weakly Supervised Object Localization Methods Right. Choe et al. CVPR 2020.

Le génie pour l'industrie

Part 2. Review of WSOL methods: Setup

WSOL: Evaluation/metrics

WSOL **localization** metric: **MaxBoxAccV2**

Account for variable object sizes

$$1/3 \sum_{\delta} MaxBoxAcc(\delta), \quad \delta \in \{0.3, 0.5, 0.7\}$$

MaxBoxAccV2 more difficult than MaxBoxAcc

Part 2. Review of WSOL methods: Setup

WSOL: Evaluation/metrics

WSOL **localization and classification** metric: **top1-localization, top5-localization**

Top1-localization

MaxBoxAcc(0.5) = 1 and true class = top1 prediction

Top5-localization

MaxBoxAcc(0.5) = 1 and true class in **top5** predictions

Evaluating Weakly Supervised Object Localization Methods Right. Choe et al. CVPR 2020.

Part 2. Review of WSOL methods: Setup

WSOL: Model selection (validation / early stopping)

Image
+
Class label

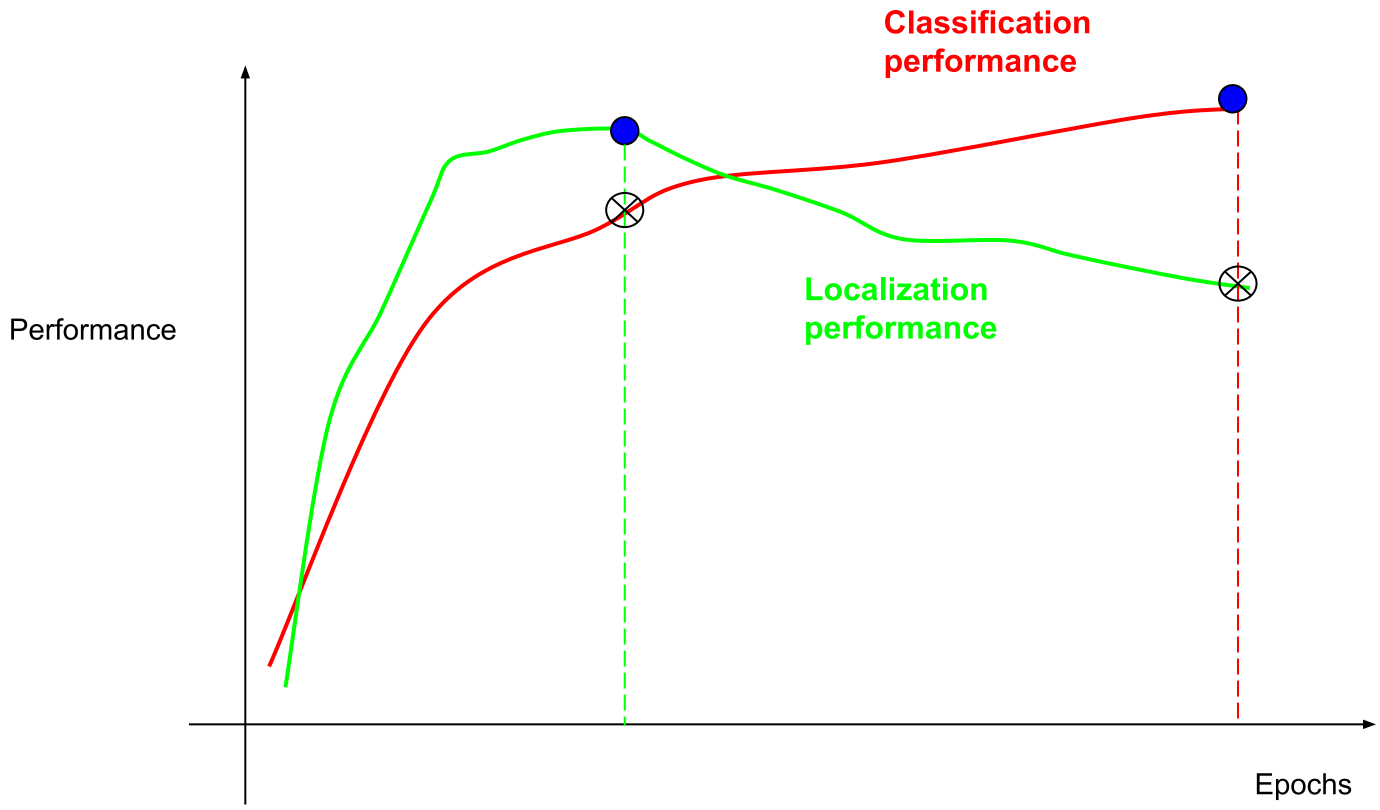
Train set

Image
+
Class label
+
Bounding box

Test set

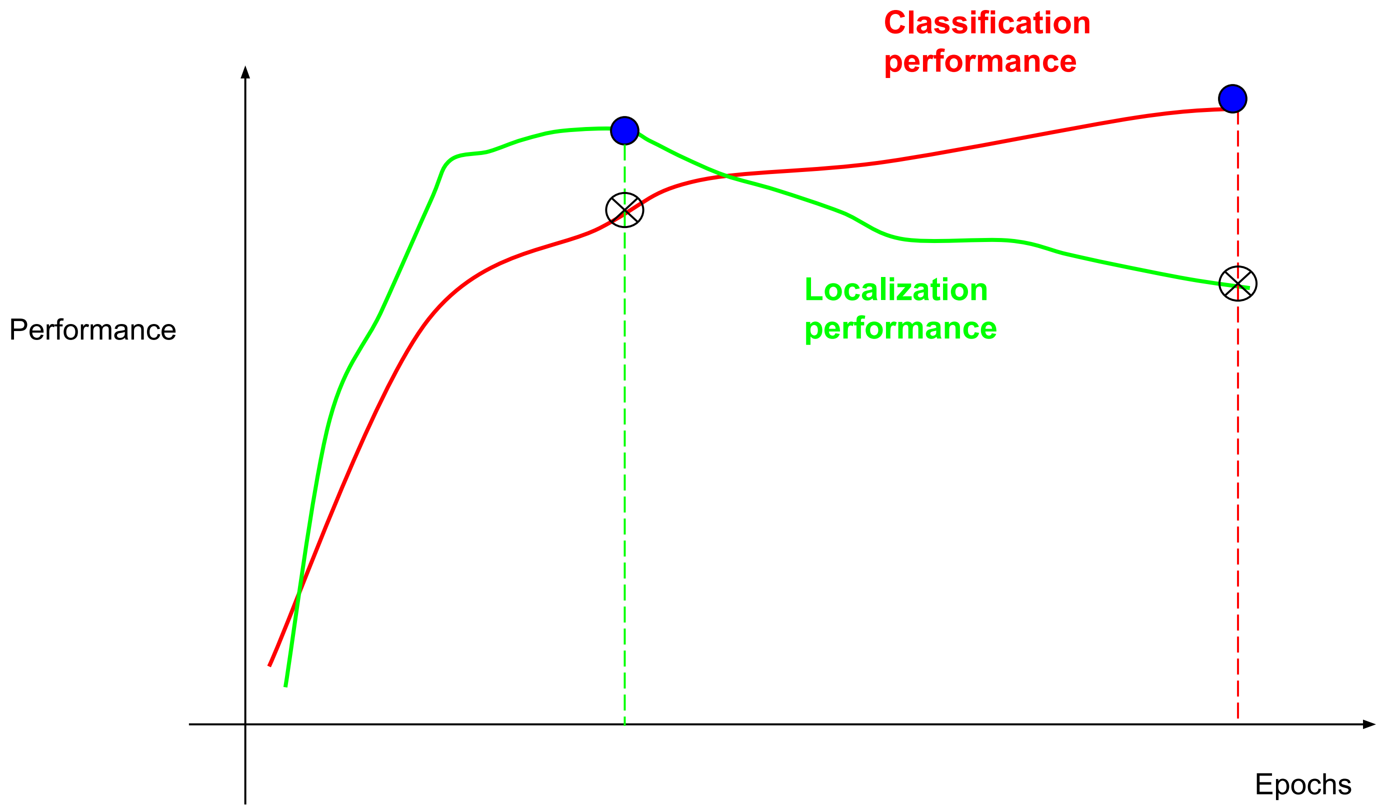
Part 2. Review of WSOL methods: Setup

WSOL: Model selection (validation / early stopping)



Part 2. Review of WSOL methods: Setup

WSOL: Model selection (validation / early stopping)



Classification, localization: antagonist tasks (WSOL)

Part 2. Review of WSOL methods: Setup

WSOL: Model selection (validation / early stopping)

Image
+
Class label

Train set

Image
+
Class label
+
Bounding box

Test set

User bias over test set !!!

Part 2. Review of WSOL methods: Setup

WSOL: Model selection (validation / early stopping)

Train set	Valid set	Test set
Image + Class label	Image + Class label + Bounding box	Image + Class label + Bounding box

Part 2. Review of WSOL methods: Setup

WSOL: Model selection (validation / early stopping)

Image
+
Class label

Train set

Image
+
Class label
+
Bounding box

Valid set

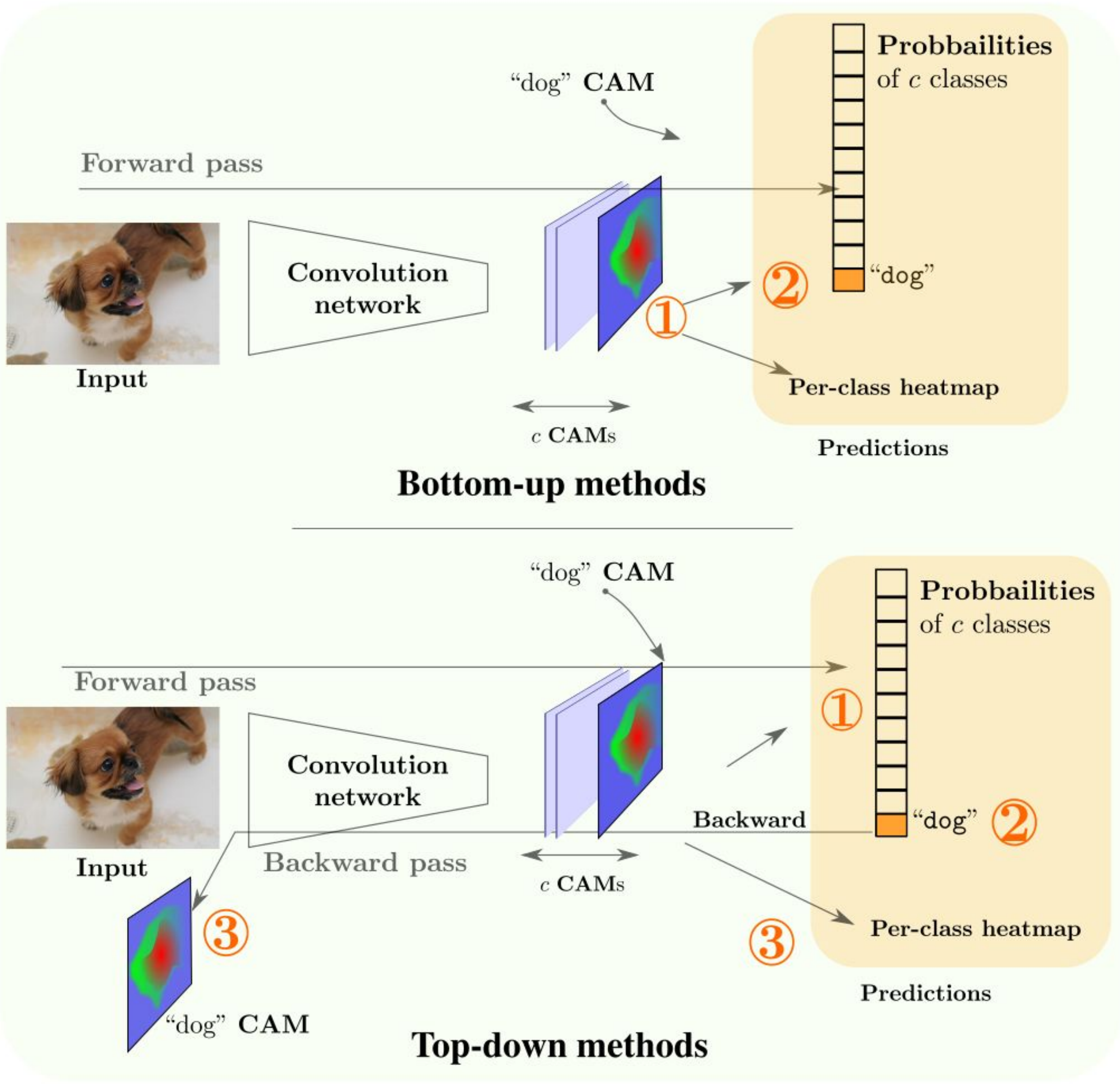
Image
+
Class label
+
Bounding box

Test set

- Allow more fair comparison between methods
- Realistic?

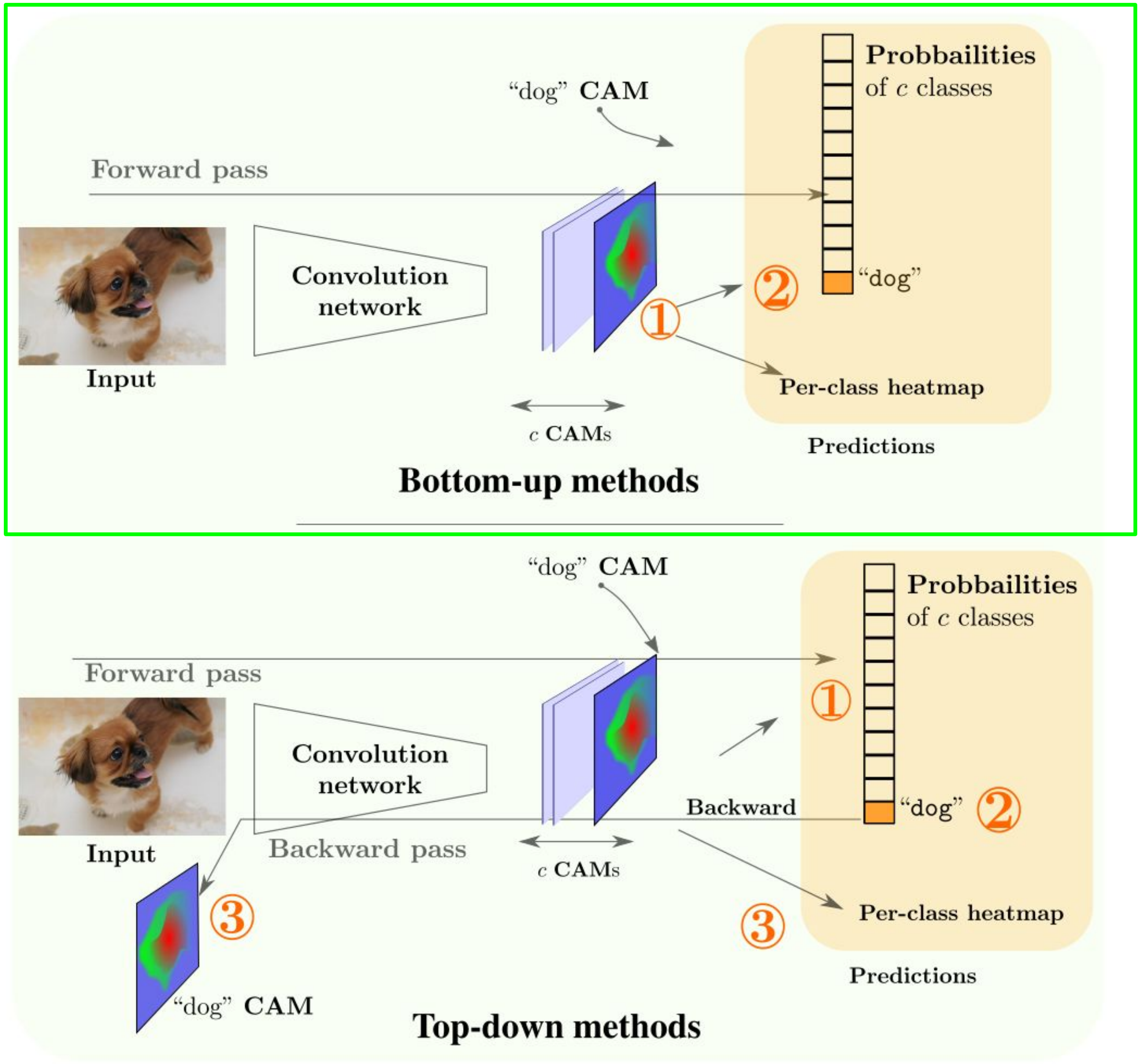
Part 2. Review of WSOL methods: Literature

Taxonomy



Part 2. Review of WSOL methods: Literature

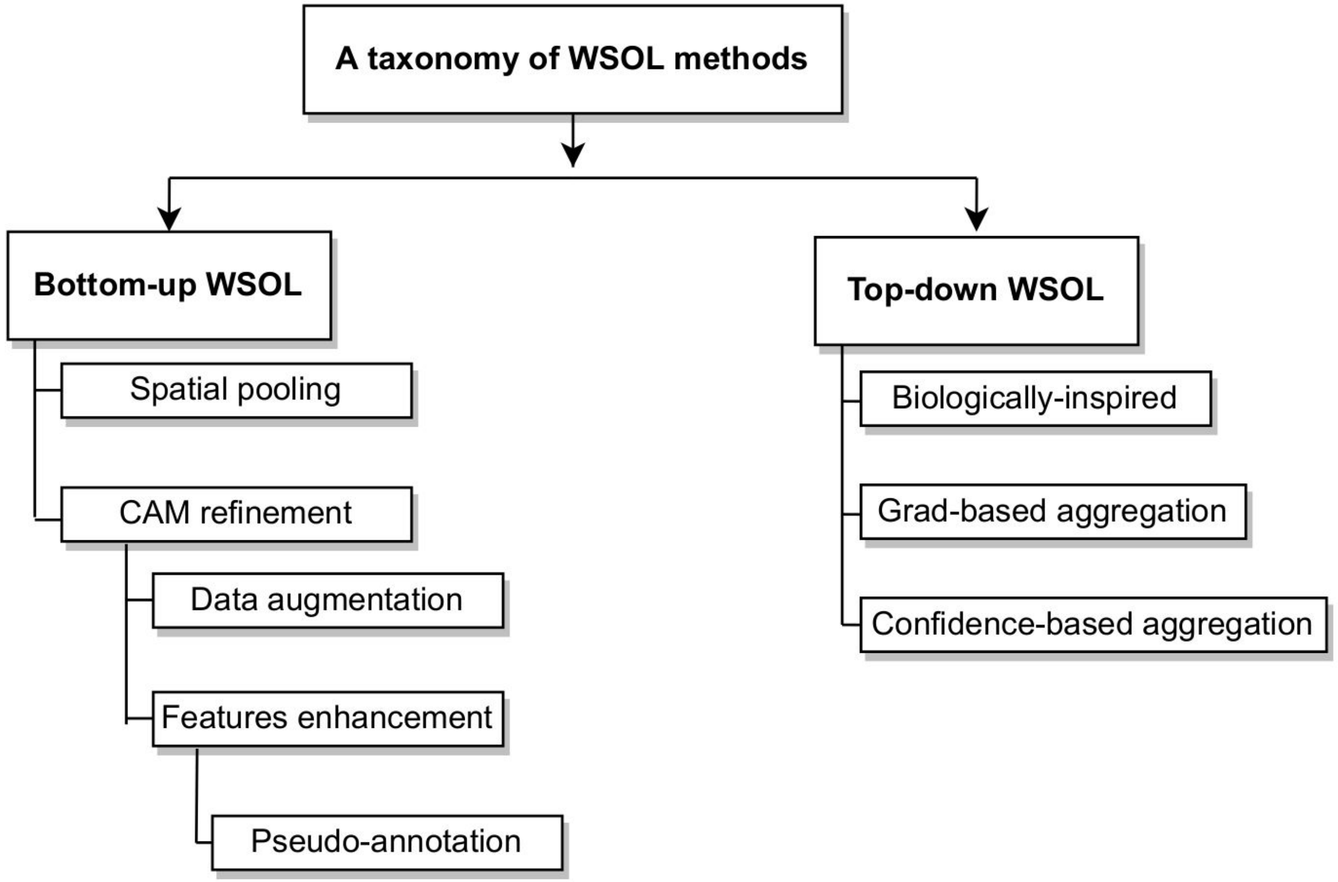
Taxonomy



More common
in WSOL

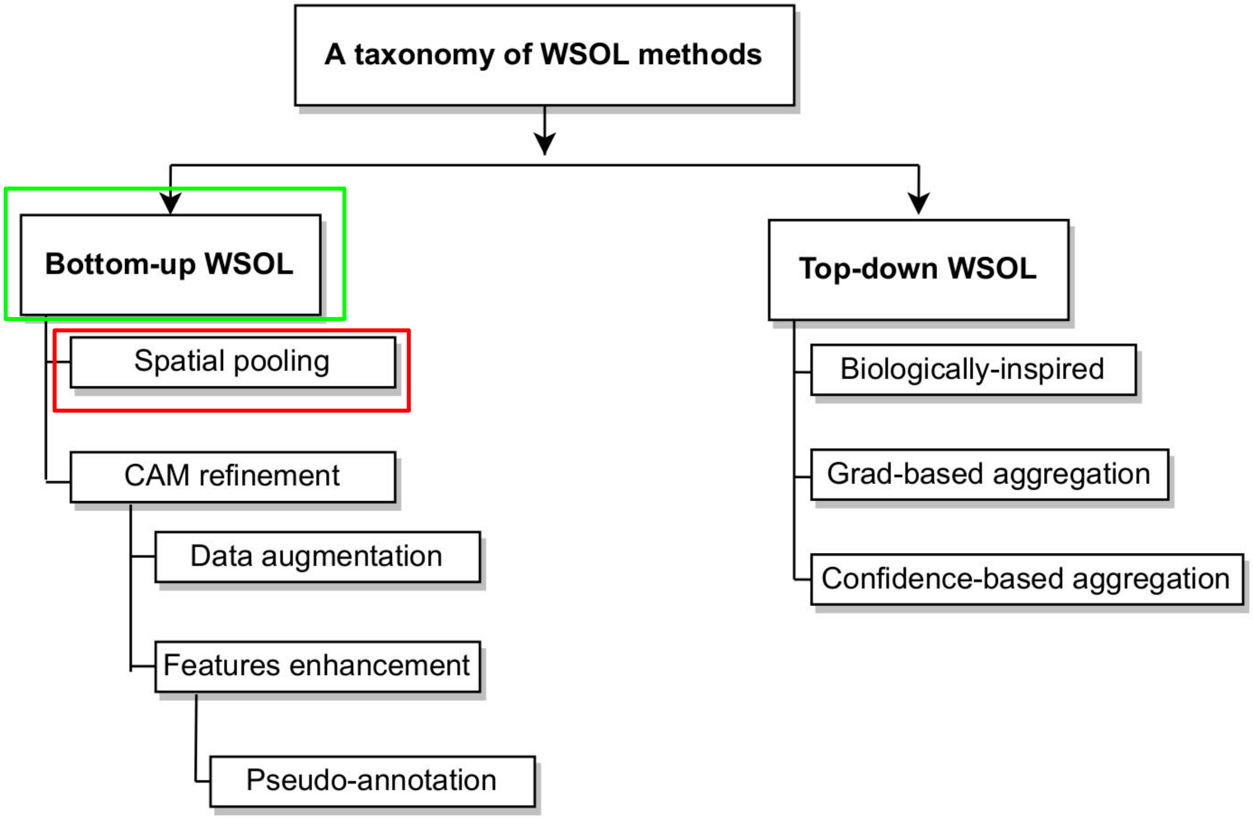
Part 2. Review of WSOL methods: Literature

Taxonomy



Part 2. Review of WSOL methods: Literature

Taxonomy



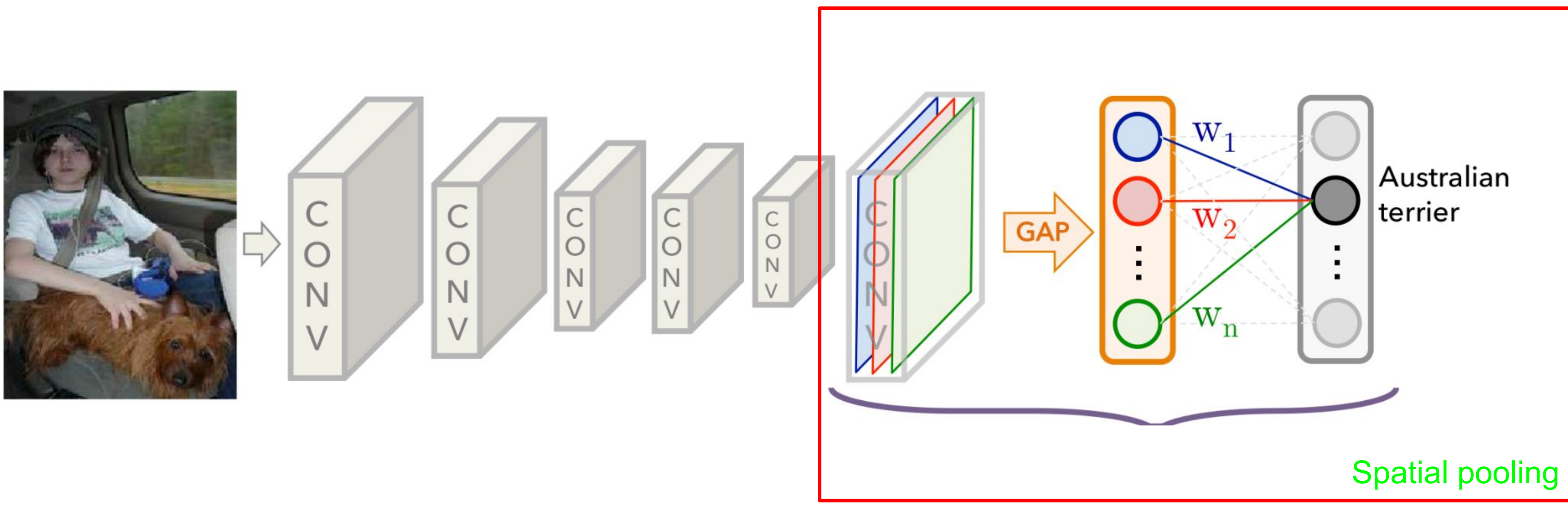
Part 2. Review of WSOL methods: Literature

Bottom-up WSOL

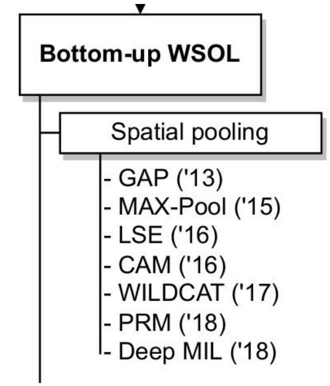
- Spatial pooling
  - GAP ('13)
  - MAX-Pool ('15)
  - LSE ('16)
  - CAM ('16)
  - WILDCAT ('17)
  - PRM ('18)
  - Deep MIL ('18)

Taxonomy: Bottom-up
Spatial pooling

Spatial pooling function = function to compute posterior per-class probability from spatial maps



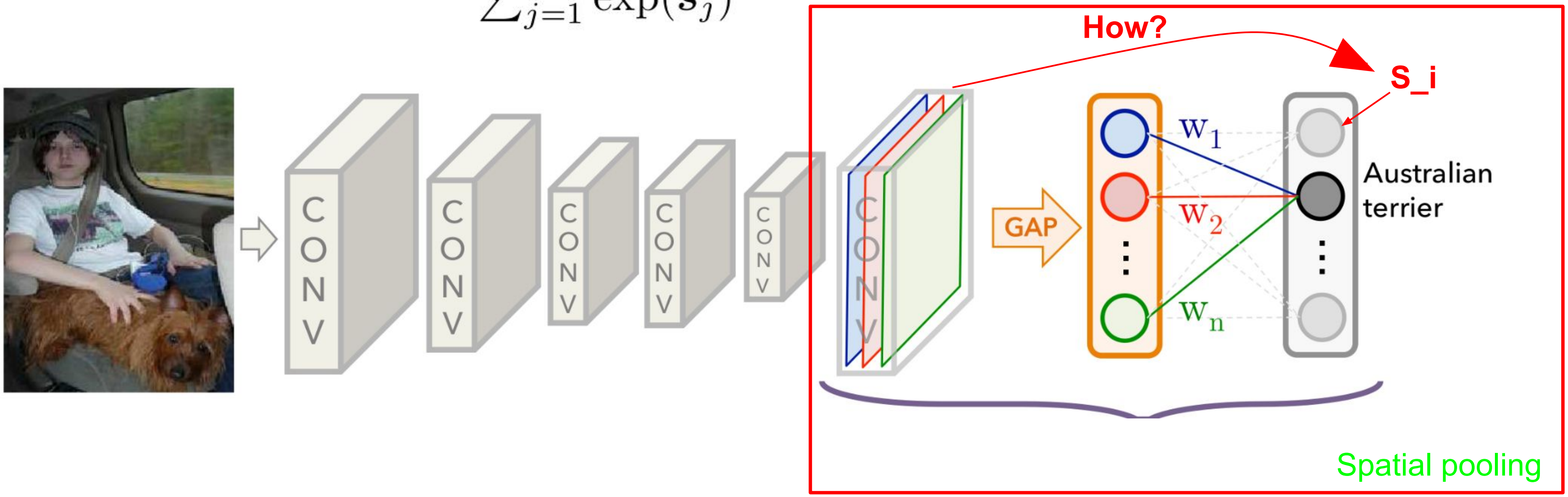
Part 2. Review of WSOL methods: Literature



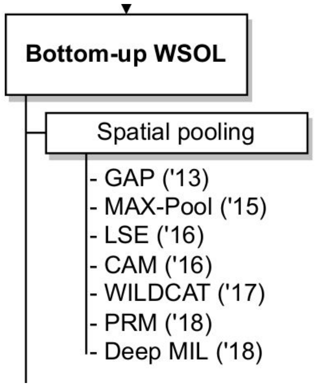
Taxonomy: Bottom-up
Spatial pooling

Spatial pooling function = function to compute posterior per-class probability from spatial maps

$$\Pr(y = i|\mathbf{x}) = \text{softmax}(\boldsymbol{s})_i = \frac{\exp(\boldsymbol{s}_i)}{\sum_{j=1}^{C} \exp(\boldsymbol{s}_j)} .$$



Part 2. Review of WSOL methods: Literature

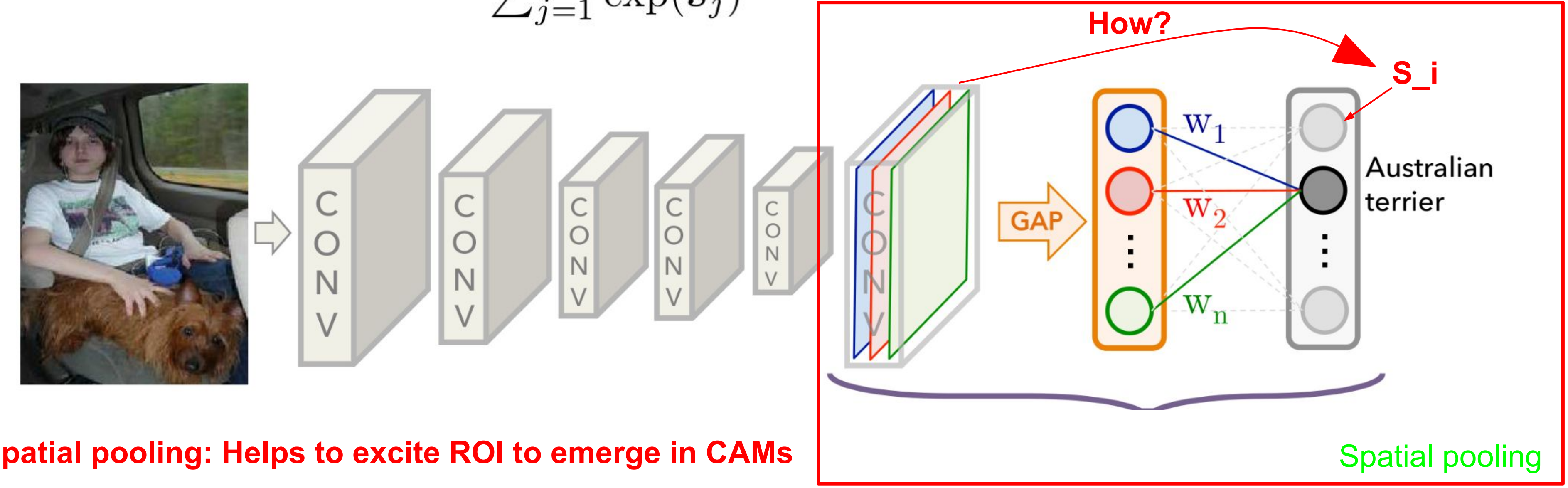


Taxonomy: Bottom-up

Spatial pooling

Spatial pooling function = function to compute posterior per-class probability from spatial maps

$$\Pr(y=i|\mathbf{x}) = \mathrm{softmax}(\boldsymbol{s})_i = \frac{\exp(\boldsymbol{s}_i)}{\sum_{j=1}^{C}\exp(\boldsymbol{s}_j)} .$$



Spatial pooling: Helps to excite ROI to emerge in CAMs

ÉTS Le génie pour l'industrie

Part 2. Review of WSOL methods: Literature

Bottom-up WSOL

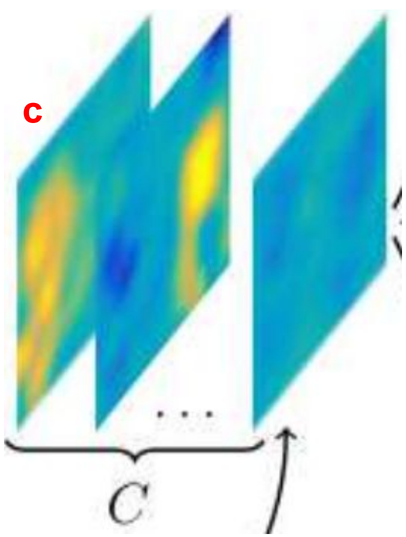
- Spatial pooling
  - GAP ('13)
  - MAX-Pool ('15)
  - LSE ('16)
  - CAM ('16)
  - WILDCAT ('17)
  - PRM ('18)
  - Deep MIL ('18)

Taxonomy: Bottom-up
Spatial pooling

Spatial pooling function = function to compute posterior per-class probability from spatial maps

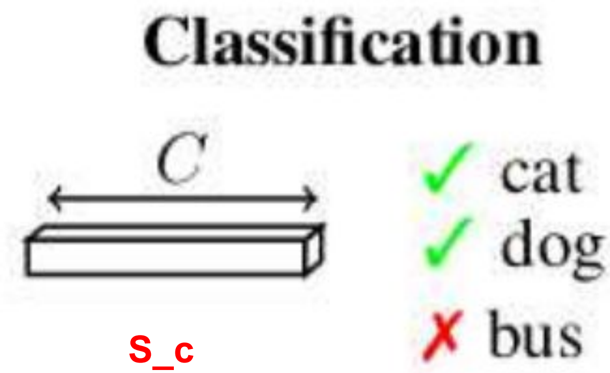
M^c

GAP: **G**lobal **A**verage **P**ooling



All locations take part in scoring!!!!

$$s_c = \frac{1}{h \times w} \sum_{i,j} M^c_{i,j}$$



Lin, M., Chen, Q., and Yan, S. (2013). Network in network. coRR, abs/1312.4400.

Part 2. Review of WSOL methods: Literature

- **Bottom-up WSOL**
  - Spatial pooling
    - GAP ('13)
    - MAX-Pool ('15)
    - LSE ('16)
    - CAM ('16)
    - WILDCAT ('17)
    - PRM ('18)
    - Deep MIL ('18)

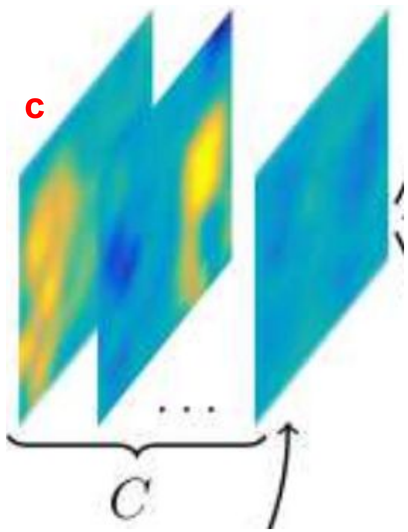
Taxonomy: Bottom-up
Spatial pooling

Spatial pooling function = function to compute posterior per-class probability from spatial maps

GAP: **G**lobal **A**verage **P**ooling

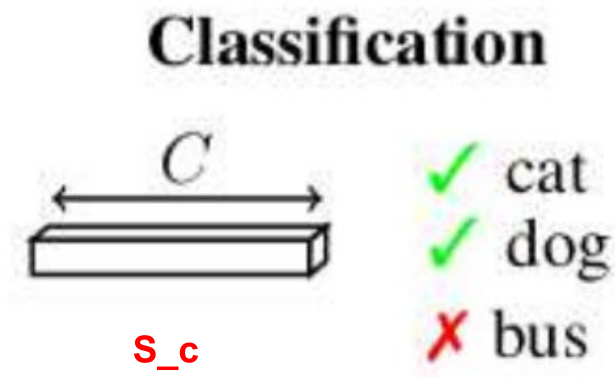
M^c

c



All locations take part in scoring!!!!

$$s_c = \frac{1}{h \times w} \sum_{i,j} M^c_{i,j}$$



S_c

Only small part of the object activates!!!

Lin, M., Chen, Q., and Yan, S. (2013). Network in network. coRR, abs/1312.4400.

Part 2. Review of WSOL methods: Literature

Bottom-up WSOL

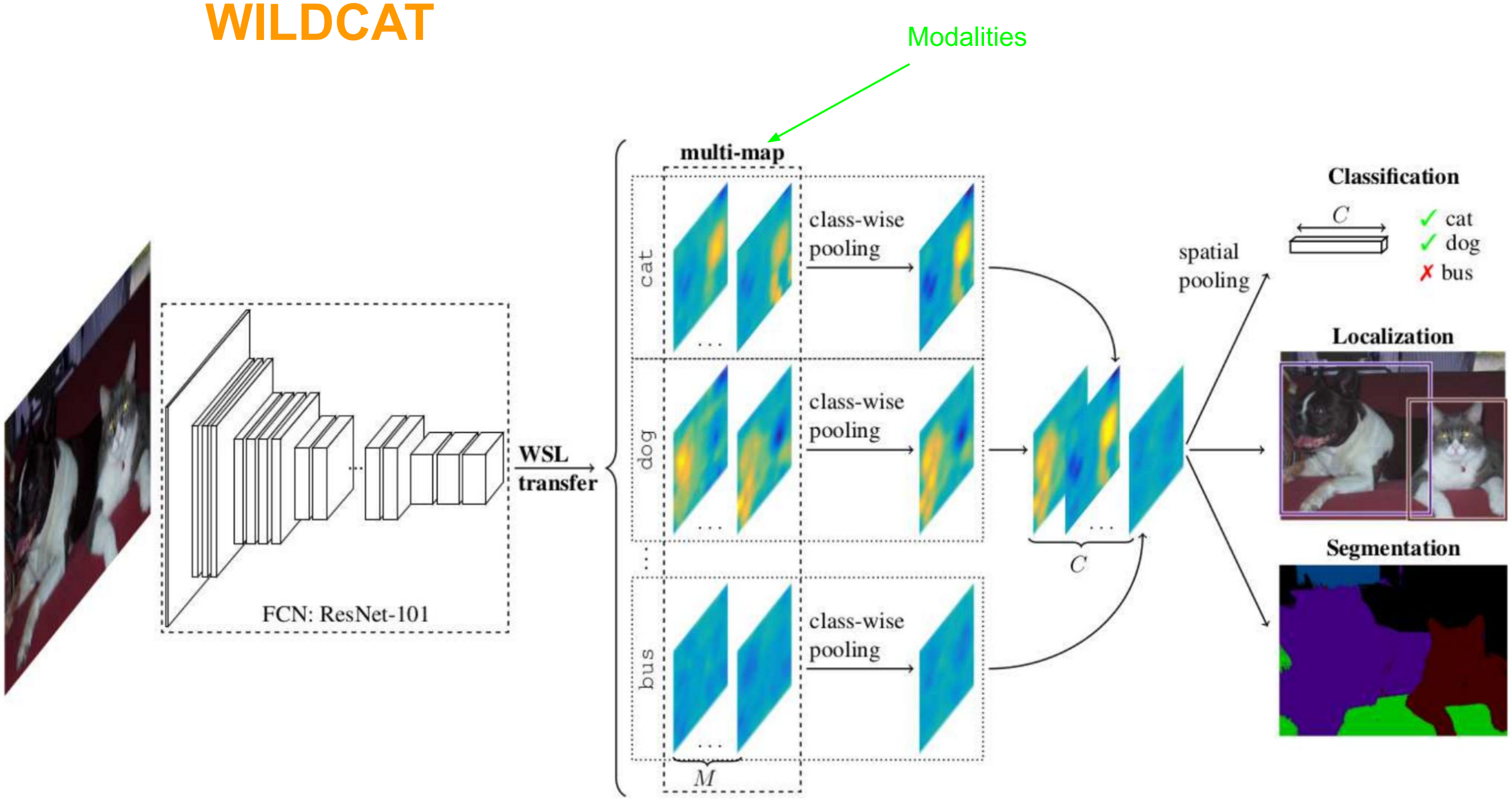
Spatial pooling

- GAP ('13)
- MAX-Pool ('15)
- LSE ('16)
- CAM ('16)
- WILDCAT ('17)
- PRM ('18)
- Deep MIL ('18)

Taxonomy: Bottom-up
Spatial pooling

Spatial pooling function = function to compute posterior per-class probability from spatial maps

WILDCAT



Durand, T., Mordan, T., Thome, N., et al. (2017). Wildcat: Weakly supervised learning of deep convnets for image classification, pointwise localization and segmentation. In CVPR.

Le génie pour l'industrie

Part 2. Review of WSOL methods: Literature

Bottom-up WSOL

Spatial pooling

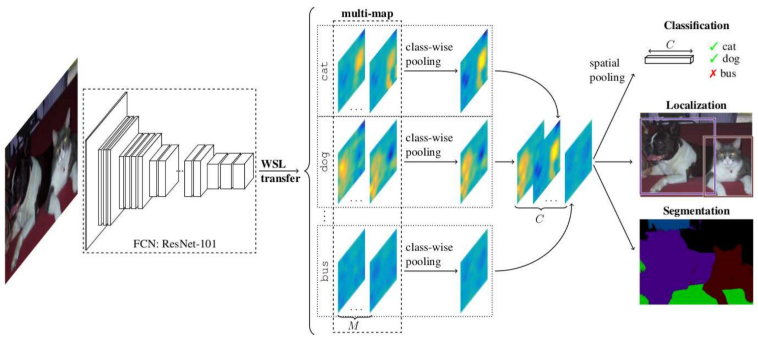
- GAP ('13)
- MAX-Pool ('15)
- LSE ('16)
- CAM ('16)
- WILDCAT ('17)
- PRM ('18)
- Deep MIL ('18)

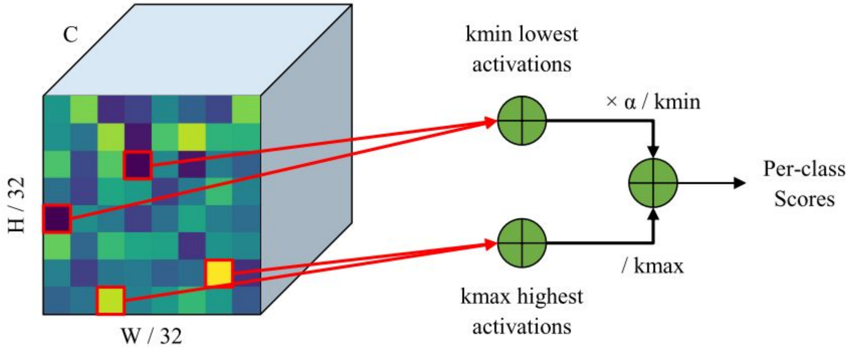
Taxonomy: Bottom-up
Spatial pooling

Spatial pooling function = function to compute posterior per-class probability from spatial maps

WILDCAT

$$s^c = \max_{\mathbf{h} \in \mathcal{H}_{k^+}} \frac{1}{k^+} \sum_{i,j} h_{i,j} \bar{z}^c_{i,j} + \alpha \left(\min_{\mathbf{h} \in \mathcal{H}_{k^-}} \frac{1}{k^-} \sum_{i,j} h_{i,j} \bar{z}^c_{i,j} \right)$$





Durand, T., Mordan, T., Thome, N., et al. (2017). Wildcat: Weakly supervised learning of deep convnets for image classification, pointwise localization and segmentation. In CVPR.

Le génie pour l'industrie

Part 2. Review of WSOL methods: Literature

Bottom-up WSOL

Spatial pooling

- GAP ('13)
- MAX-Pool ('15)
- LSE ('16)
- CAM ('16)
- WILDCAT ('17)
- PRM ('18)
- Deep MIL ('18)

Taxonomy: Bottom-up

Spatial pooling

Main issue:
Minimal coverage of objects
(most discriminative parts only)

Part 2. Review of WSOL methods: Literature

Bottom-up WSOL

Spatial pooling

- GAP ('13)
- MAX-Pool ('15)
- LSE ('16)
- CAM ('16)
- WILDCAT ('17)
- PRM ('18)
- Deep MIL ('18)

Taxonomy: Bottom-up

Spatial pooling

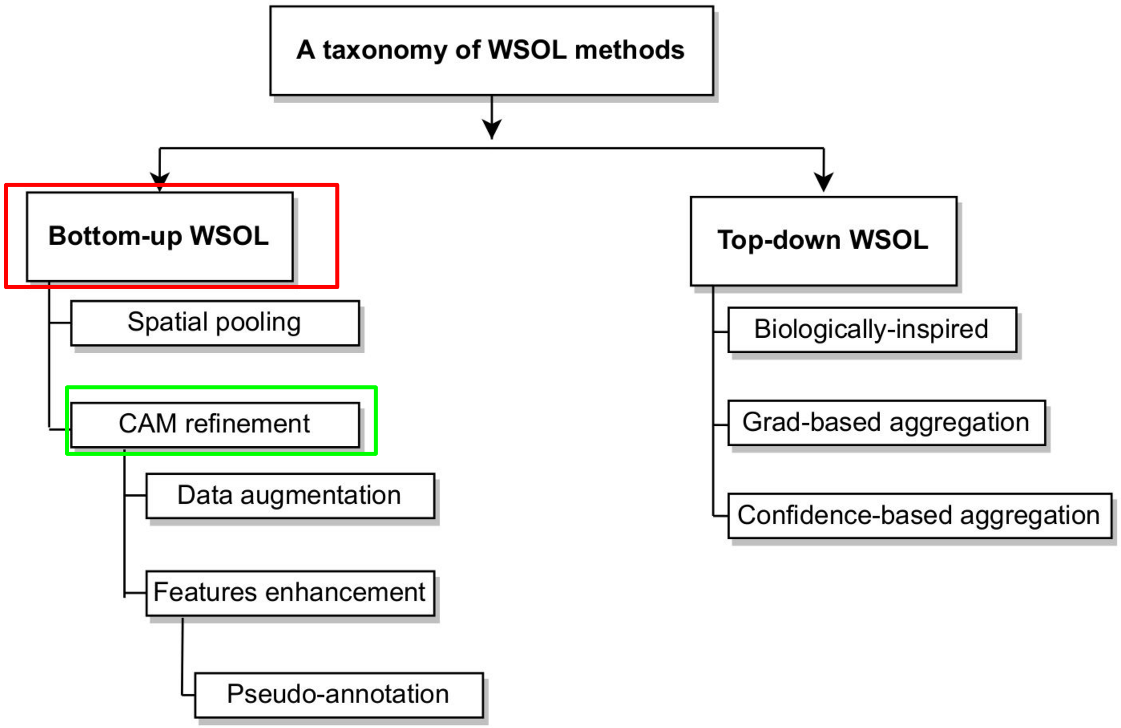
Main issue:
Minimal coverage of objects
(most discriminative parts only)

Solution: Refine the CAM!!!

Part 2. Review of WSOL methods: Literature

Taxonomy: Bottom-up CAM refinement

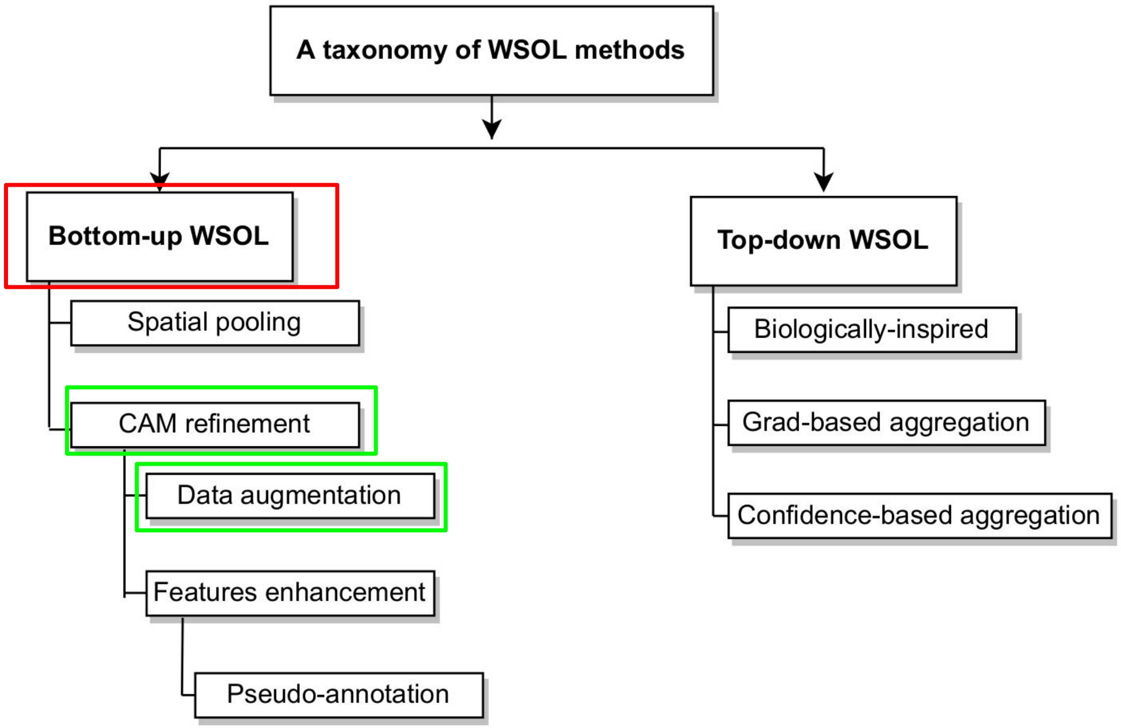
CAM refinement:
- Data augmentation
- Features enhancement
- Pseudo-annotation



Goal: how to recover full object?

Part 2. Review of WSOL methods: Literature

Taxonomy: Bottom-up
CAM refinement



CAM refinement:

- **Data augmentation**
- Features enhancement
- Pseudo-annotation

Goal: how to recover full object?

Part 2. Review of WSOL methods: Literature

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)

Taxonomy: Bottom-up

CAM refinement: Data augmentation

- Uses simple pooling functions
- Data augmentation: **prevent model from overfitting over single small part of the object**
- **Mine** complet objects by **perturbing image / features: Information suppression (erasing)**

Le génie pour l'industrie

Part 2. Review of WSOL methods: Literature

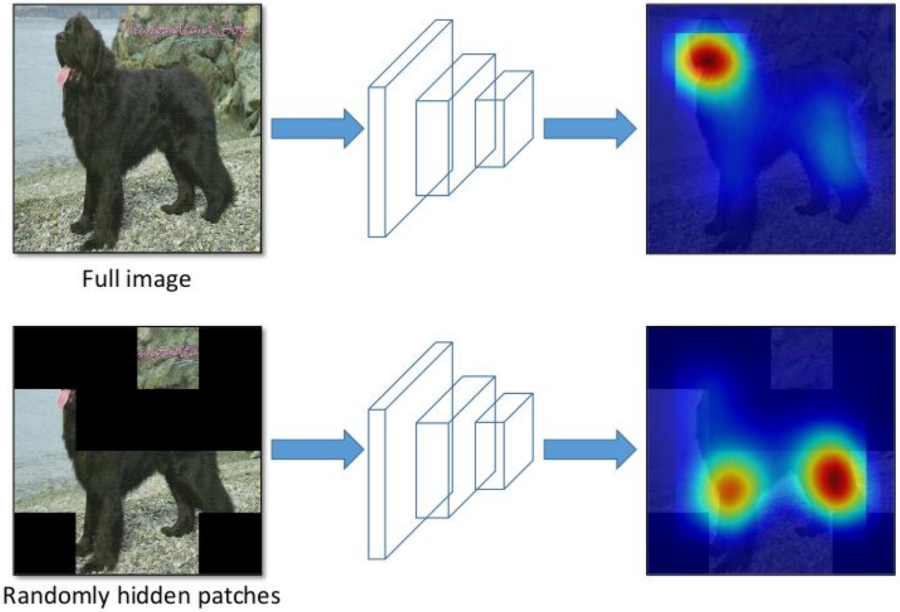
Data augmentation

- 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)

Taxonomy: Bottom-up

CAM refinement: Data augmentation

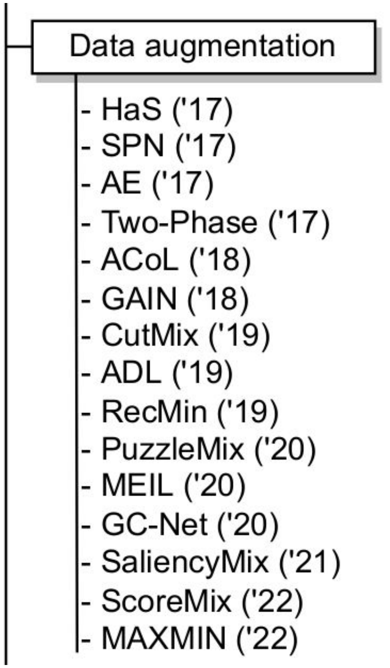
HaS: **H**ide **a**nd **S**eek



Le génie pour l'industrie

Singh, K. and Lee, Y. (2017). Hide-and-seek: Forcing a network to be meticulous for weakly-supervised object and action localization. In ICCV.

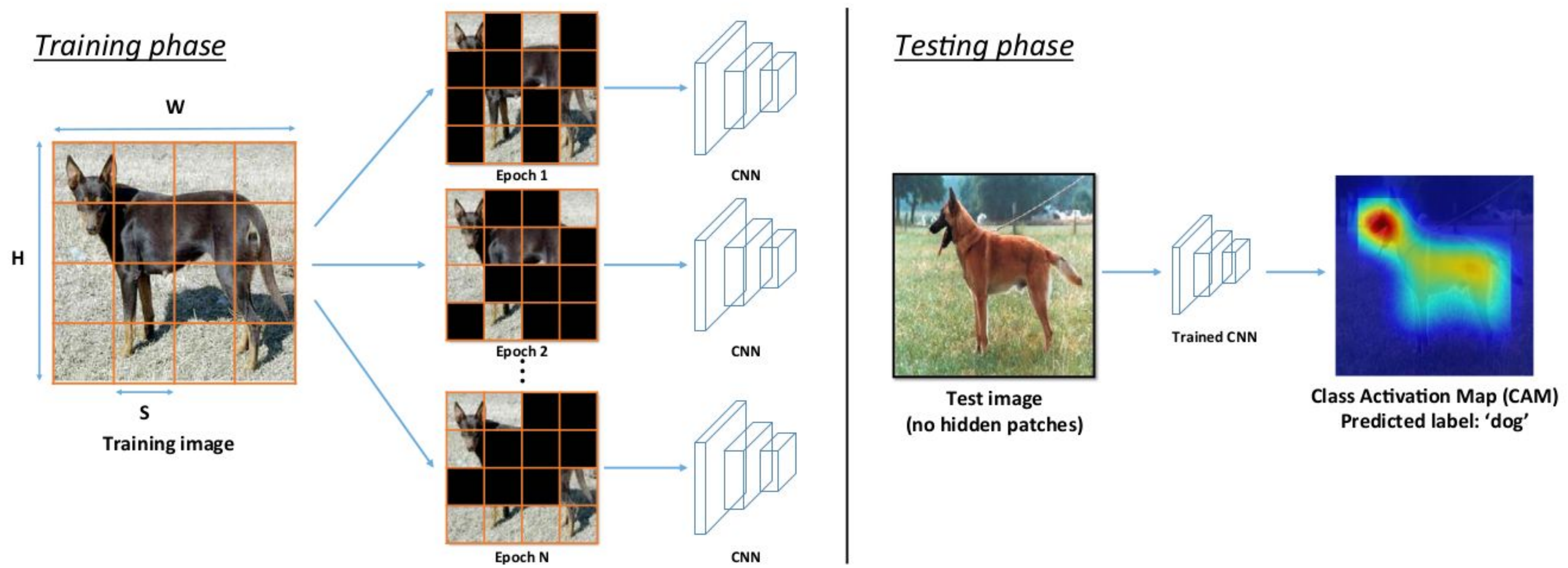
Part 2. Review of WSOL methods: Literature



Taxonomy: Bottom-up

CAM refinement: Data augmentation

HaS: **H**ide **a**nd **S**eek



Drop patch with prob. p

Singh, K. and Lee, Y. (2017). Hide-and-seek: Forcing a network to be meticulous for weakly-supervised object and action localization. In ICCV.

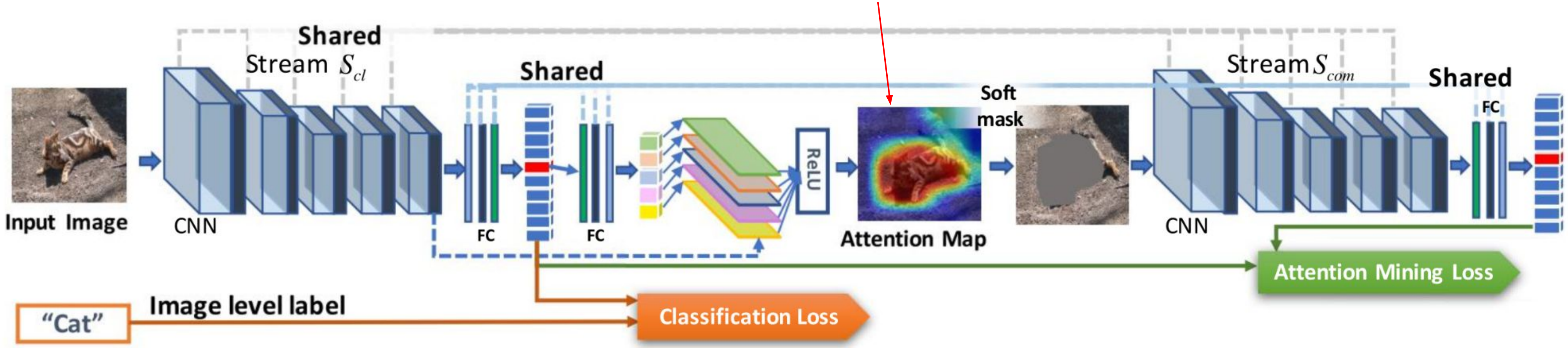
Part 2. Review of WSOL methods: Literature

- 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)

Taxonomy: Bottom-up
CAM refinement: Data augmentation

GAIN: Erasing (image)

Heated CAM



Part 2. Review of WSOL methods: Literature

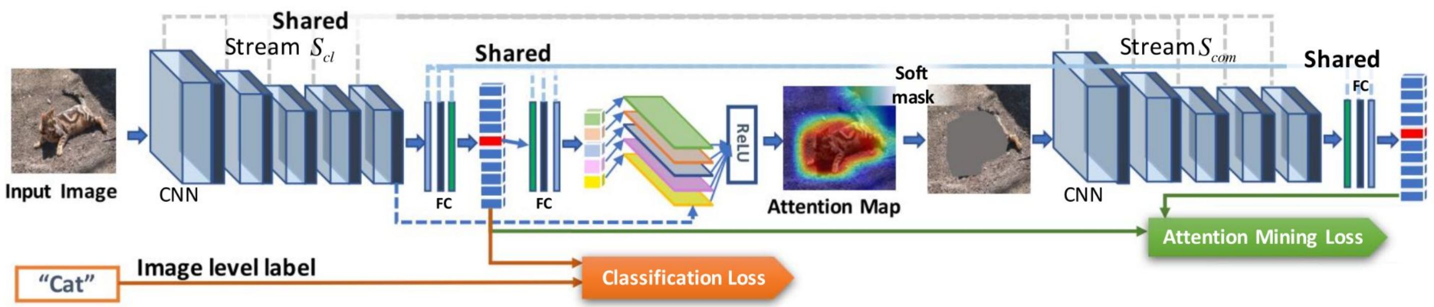
Data augmentation

- 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)

Taxonomy: Bottom-up

CAM refinement: Data augmentation

GAIN: Erasing (image)



Train loss

$$L_{self} = L_{cl} + \alpha L_{am},$$

cross-entropy

$$L_{am} = \frac{1}{n} \sum_{c} s^c(I^{*c}),$$

Le génie pour l'industrie

Li, K., Wu, Z., Peng, K., et al. (2018). Tell me where to look: Guided attention inference network. In CVPR.

Part 2. Review of WSOL methods: Literature

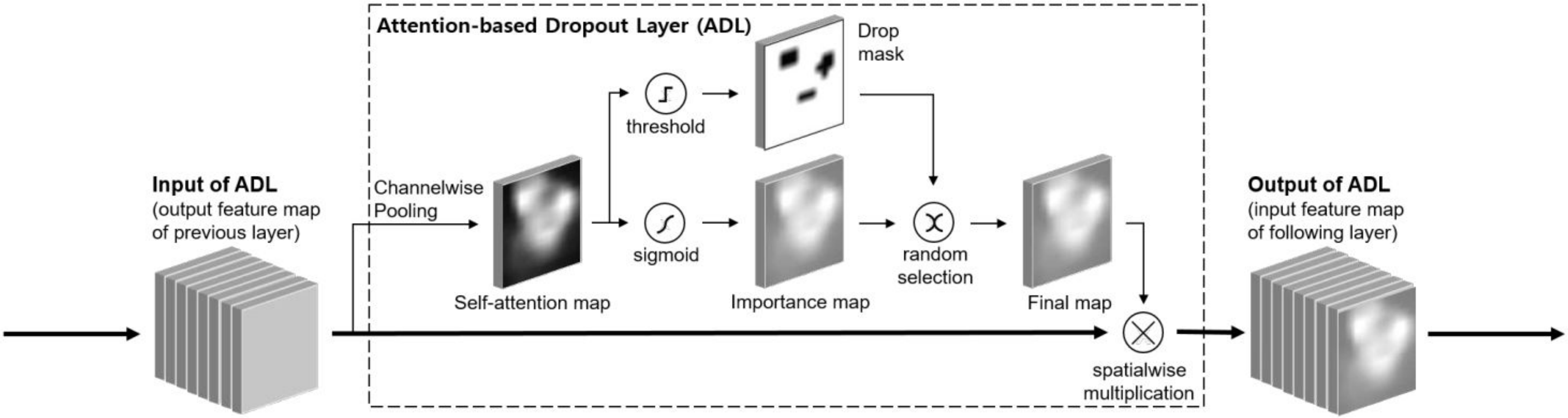
Data augmentation

- 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)

Taxonomy: Bottom-up
CAM refinement: Data augmentation

ADL: Erasing spatial features

Mining ROI via dropout!



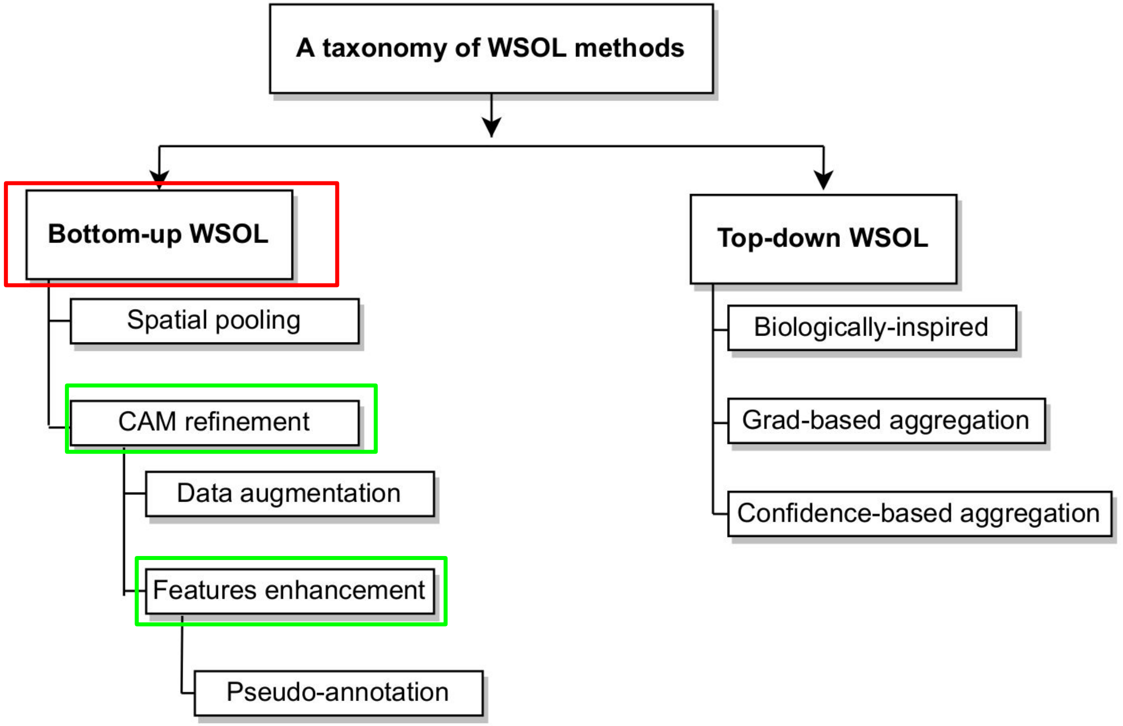
Choe, J. and Shim, H. (2019). Attention-based dropout layer for weakly supervised object localization. In CVPR.

Part 2. Review of WSOL methods: Literature

Taxonomy: Bottom-up
CAM refinement: Features enhancement

CAM refinement:
- Data augmentation
- **Features enhancement**
- Pseudo-annotation

Change architecture
+
Use low features



Part 2. Review of WSOL methods: Literature

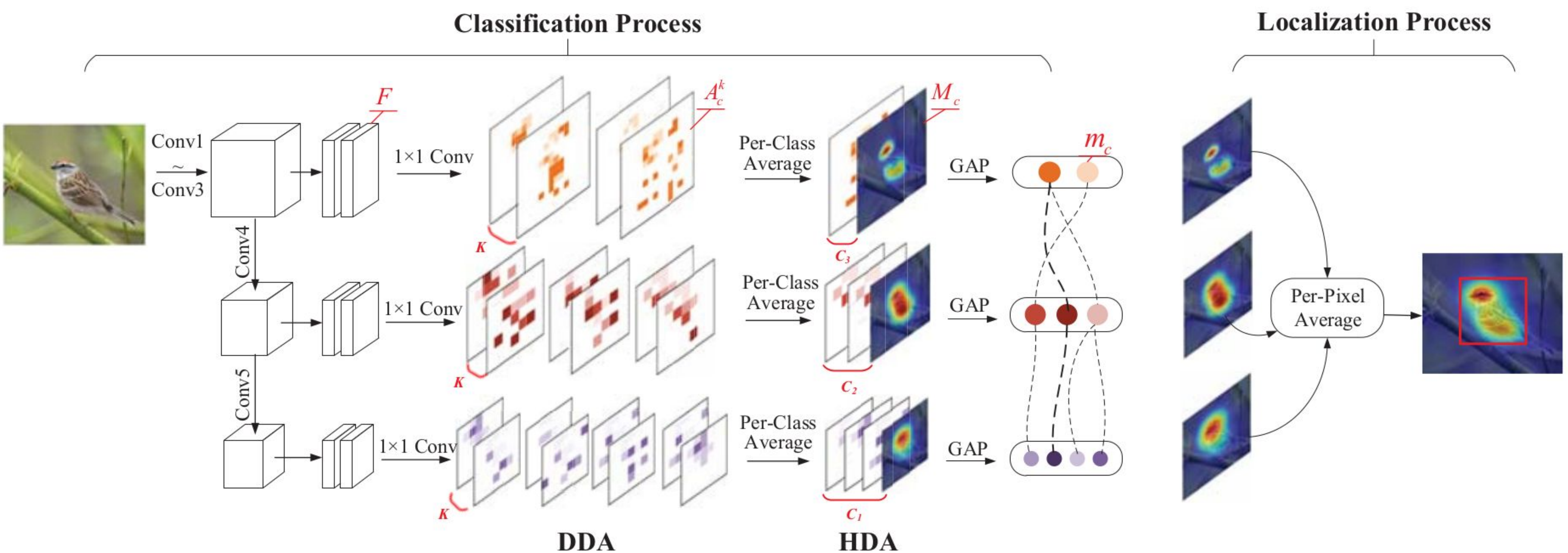
Features enhancement

- MDC ('18)
- FickleNet ('19)
- DANet ('19)
- NL-CCAM ('20)
- I2C ('20)
- ICL ('20)
- CSTN ('20)
- TS-CAM ('21)

Taxonomy: Bottom-up

CAM refinement: Features enhancement

DANet



Xue, H., Liu, C., Wan, F., Jiao, J., Ji, X., and Ye, Q. (2019). Danet: Divergent activation for weakly supervised object localization. In ICCV.

Part 2. Review of WSOL methods: Literature

Features enhancement

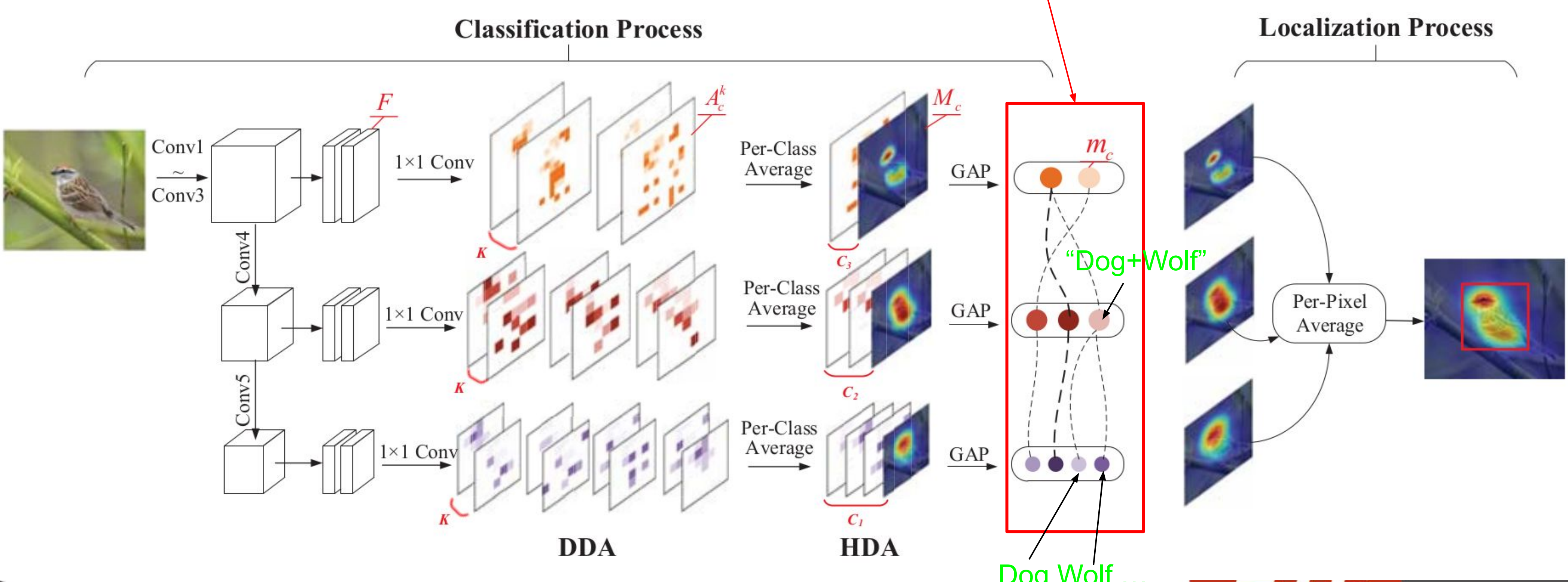
- MDC ('18)
- FickleNet ('19)
- DANet ('19)
- NL-CCAM ('20)
- I2C ('20)
- ICL ('20)
- CSTN ('20)
- TS-CAM ('21)

Taxonomy: Bottom-up

CAM refinement: Features enhancement

Define:
a hierarchy of classes, and
new parent-classes

DANet



Xue, H., Liu, C., Wan, F., Jiao, J., Ji, X., and Ye, Q. (2019). Danet: Divergent activation for weakly supervised object localization. In ICCV.

Part 2. Review of WSOL methods: Literature

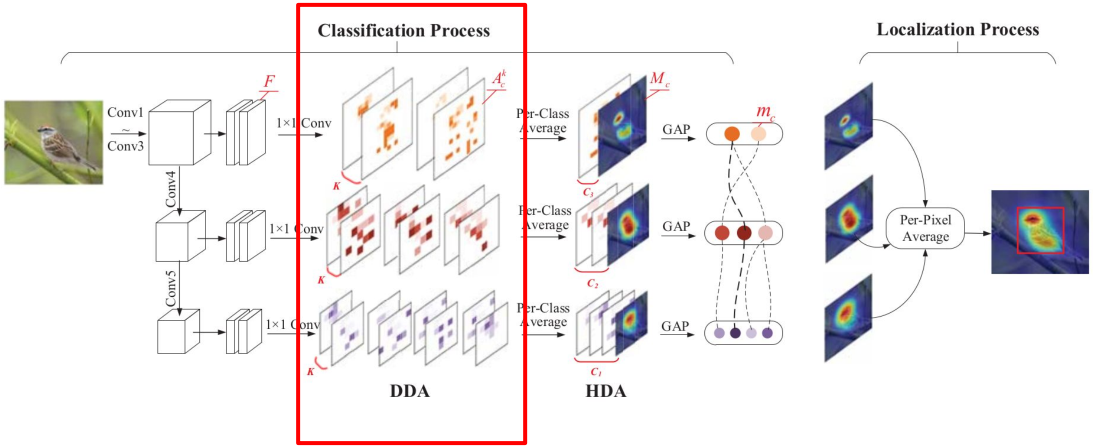
Features enhancement

- MDC ('18)
- FickleNet ('19)
- DANet ('19)
- NL-CCAM ('20)
- I2C ('20)
- ICL ('20)
- CSTN ('20)
- TS-CAM ('21)

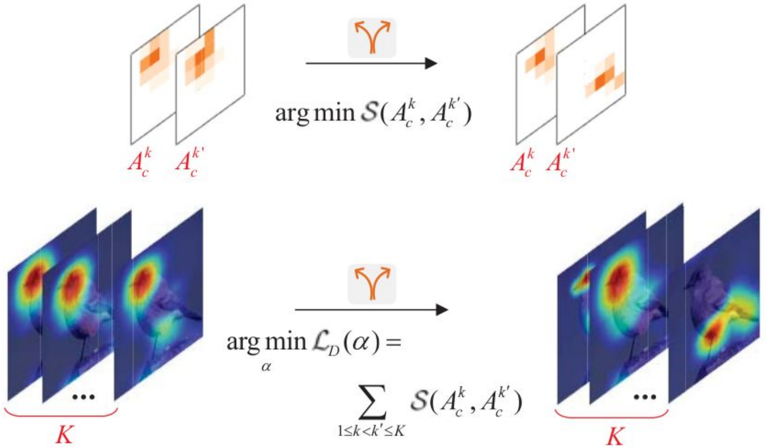
Taxonomy: Bottom-up

CAM refinement: Features enhancement

DANet



DDA loss: **D**iscrepant **D**ivergent **A**ctivation



$$\arg\min_{\alpha} \mathcal{L}_D(\alpha) = \sum_{1\leq k<k'\leq K} \mathcal{S}(A_c^k, A_c^{k'}),$$

$$\mathcal{S}(A_c^k, A_c^{k'}) = \frac{A_c^k \cdot A_c^{k'}}{\|A_c^k\| \cdot \|A_c^{k'}\|}$$

Xue, H., Liu, C., Wan, F., Jiao, J., Ji, X., and Ye, Q. (2019). Danet: Divergent activation for weakly supervised object localization. In ICCV.

Part 2. Review of WSOL methods: Literature

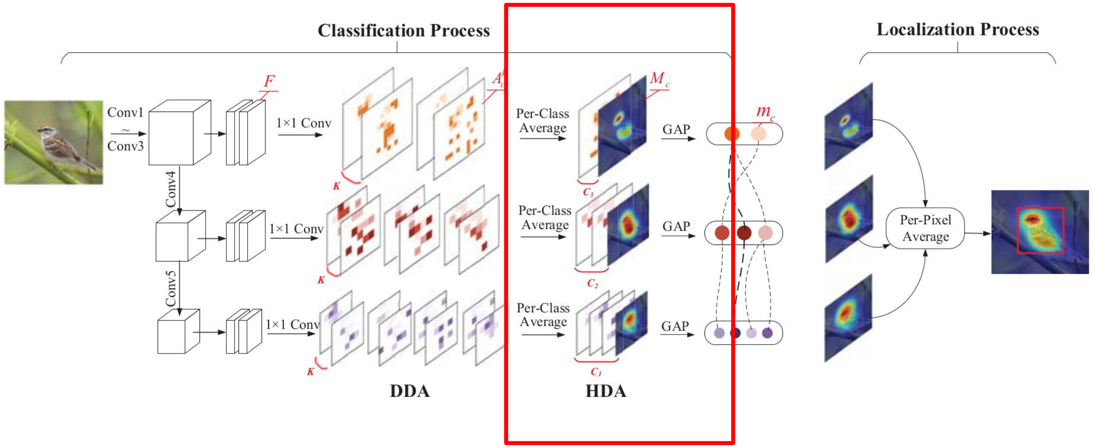
Features enhancement

- MDC ('18)
- FickleNet ('19)
- DANet ('19)
- NL-CCAM ('20)
- I2C ('20)
- ICL ('20)
- CSTN ('20)
- TS-CAM ('21)

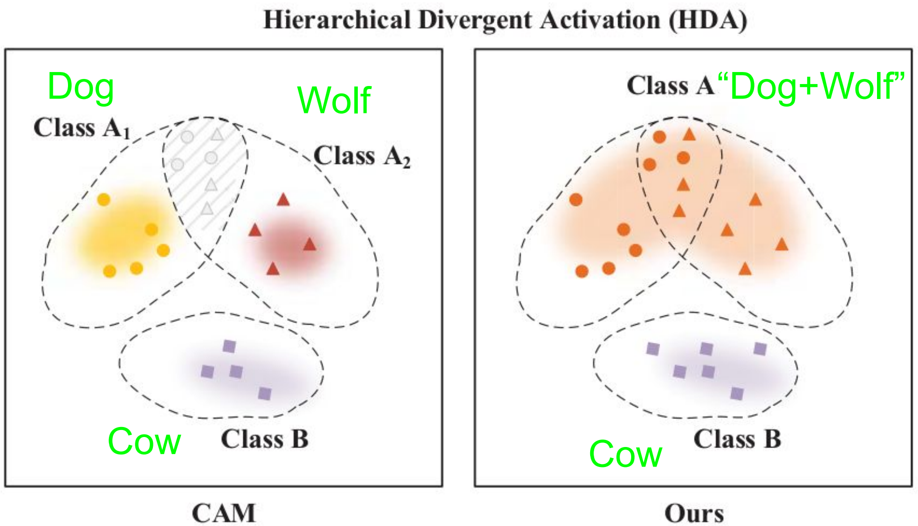
Taxonomy: Bottom-up

CAM refinement: Features enhancement

DANet



HDA loss: **H**ierarchical **D**ivergent **A**ctivation



$$\arg\min_{\alpha} \mathcal{L}_H(\alpha) = \sum_h \mathcal{L}_h(\alpha) = -\sum_h \frac{1}{C^h} \sum_c y_c^h log(p_c^h),$$

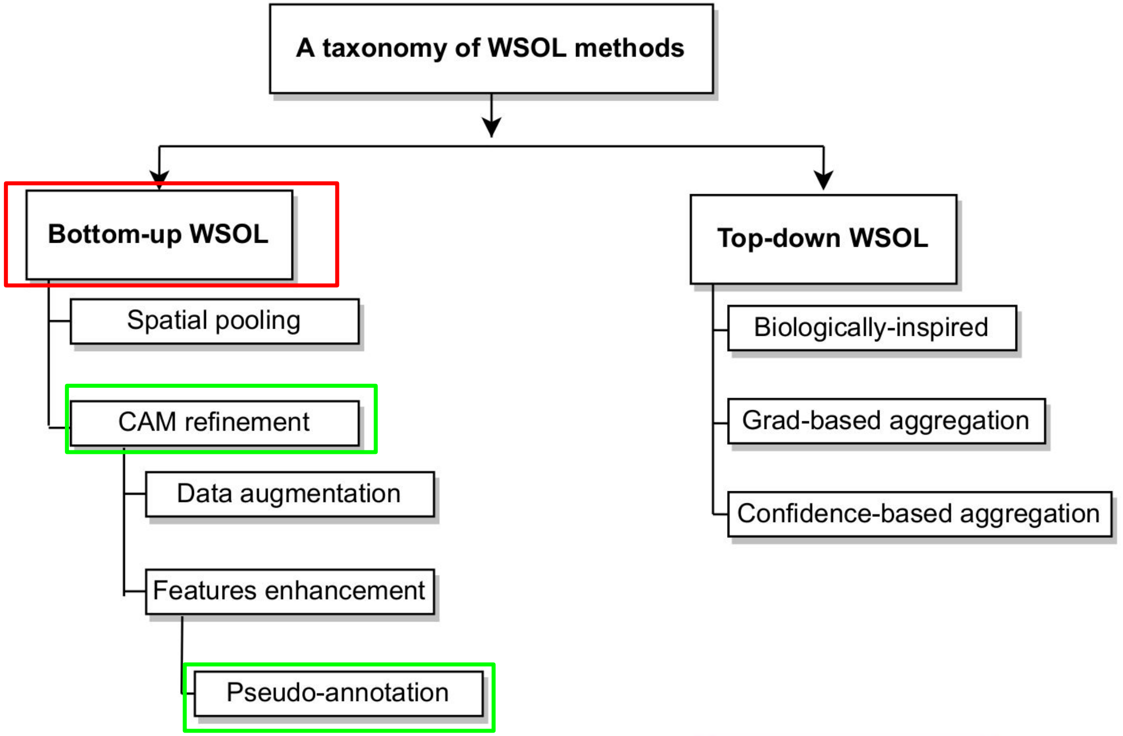
Standard classification loss

Xue, H., Liu, C., Wan, F., Jiao, J., Ji, X., and Ye, Q. (2019). Danet: Divergent activation for weakly supervised object localization. In ICCV.

Part 2. Review of WSOL methods: Literature

Taxonomy: Bottom-up
CAM refinement: Features enhancement

CAM refinement:
- Data augmentation
- Features enhancement
- **Pseudo-annotation**

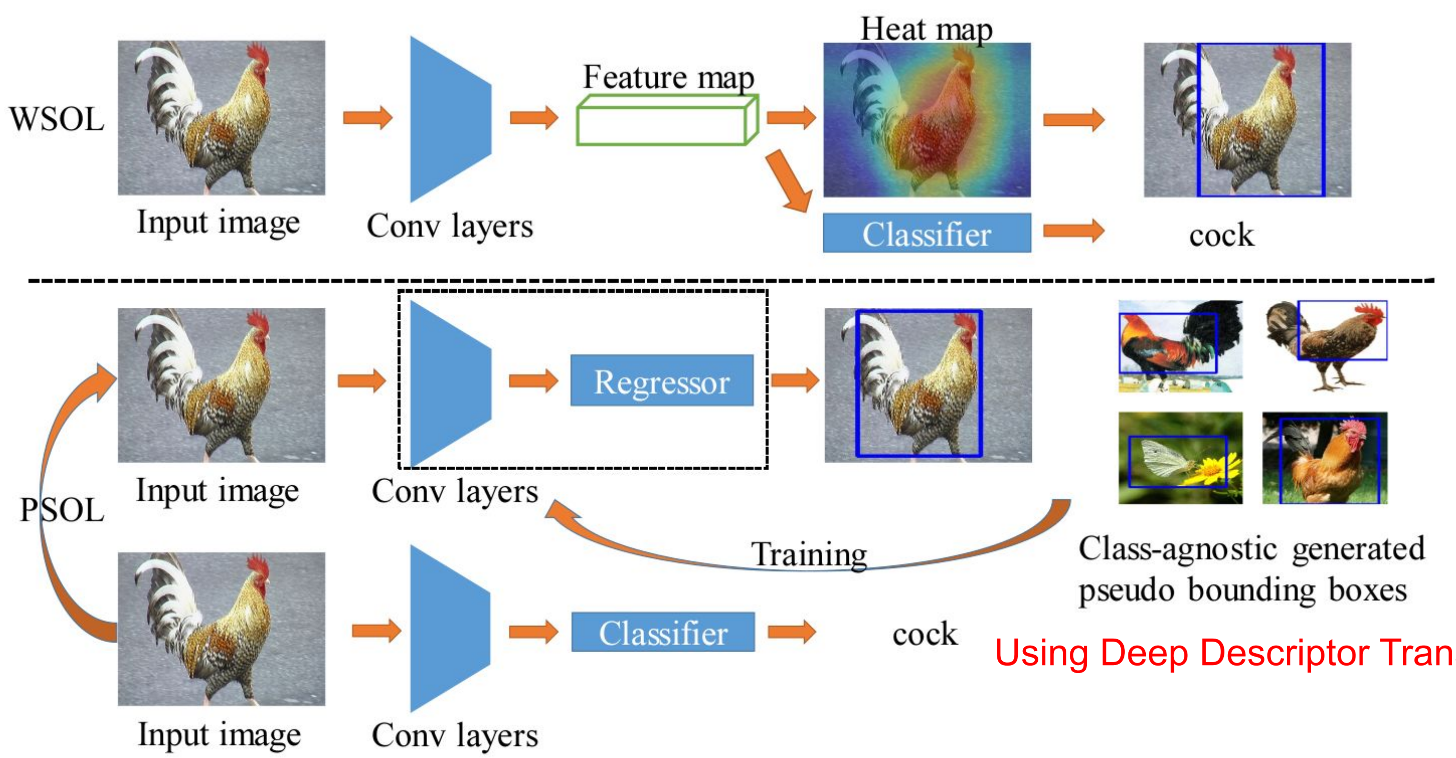


Part 2. Review of WSOL methods: Literature

Pseudo-annotation
- SPG ('18)
- Pair-Sim ('20)
- PSOL ('20)
- SPOL ('21)
- F-CAM ('22)
- NEGEV ('22)

Taxonomy: Bottom-up
CAM refinement: Features enhancement

PSOL: Pseudo-Supervised Object Localization



In WSOL:
Classification task
And
Localization task

Are antagonist
→ need to be separated

Using Deep Descriptor Transformation (DDT) method [2]

[1] Zhang, C., Cao, Y., and Wu, J. (2020a). Rethinking the route towards weakly supervised object localization. In CVPR.
[2] X.-S. Wei, C.-L. Zhang, J. Wu, C. Shen, and Z.-H. Zhou. Unsupervised object discovery and co-localization by deep descriptor transformation. PR, 88:113–126, 2019.

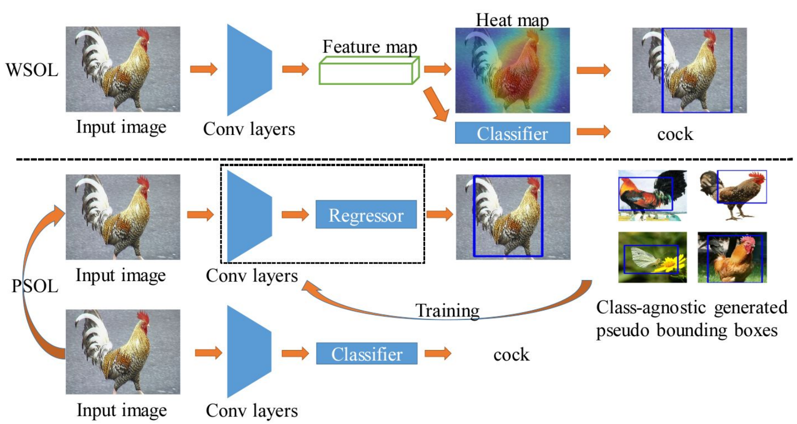
Part 2. Review of WSOL methods: Literature

Pseudo-annotation
- SPG ('18)
- Pair-Sim ('20)
- PSOL ('20)
- SPOL ('21)
- F-CAM ('22)
- NEGEV ('22)

Taxonomy: Bottom-up

CAM refinement: Features enhancement

PSOL: Pseudo-Supervised Object Localization



Algorithm 1 Pseudo Supervised Object Localization

Input: Training images I_{tr} with class label L_{tr}
Output: Predicted bounding boxes b_{te} and class labels L_{te} on testing images I_{te}

1: Generate pseudo bounding boxes $\tilde{b}_{tr}$ on I_{tr}
2: Train a localization CNN F_{loc} on I_{tr} with $\tilde{b}_{tr}$
3: Train a classification CNN F_{cls} on I_{tr} with L_{tr}
4: Use F_{loc} to predict b_{te} on I_{te}
5: Use F_{cls} to predict L_{te} on I_{te}
6: **Return:** b_{te}, L_{te}

In WSOL:
Classification task
And
Localization task

Are antagonist
→ need to be separated

Zhang, C., Cao, Y., and Wu, J. (2020a). Rethinking the route towards weakly supervised object localization. In CVPR.

Le génie pour l'industrie

Part 2. Review of WSOL methods: Literature

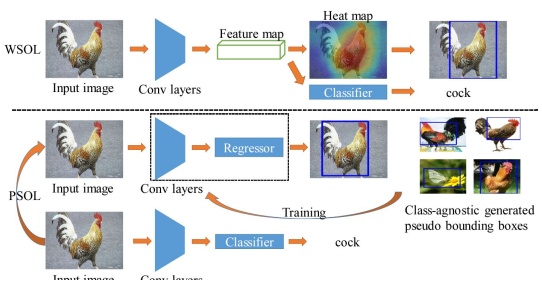
Pseudo-annotation

- SPG ('18)
- Pair-Sim ('20)
- PSOL ('20)
- SPOL ('21)
- F-CAM ('22)
- NEGEV ('22)

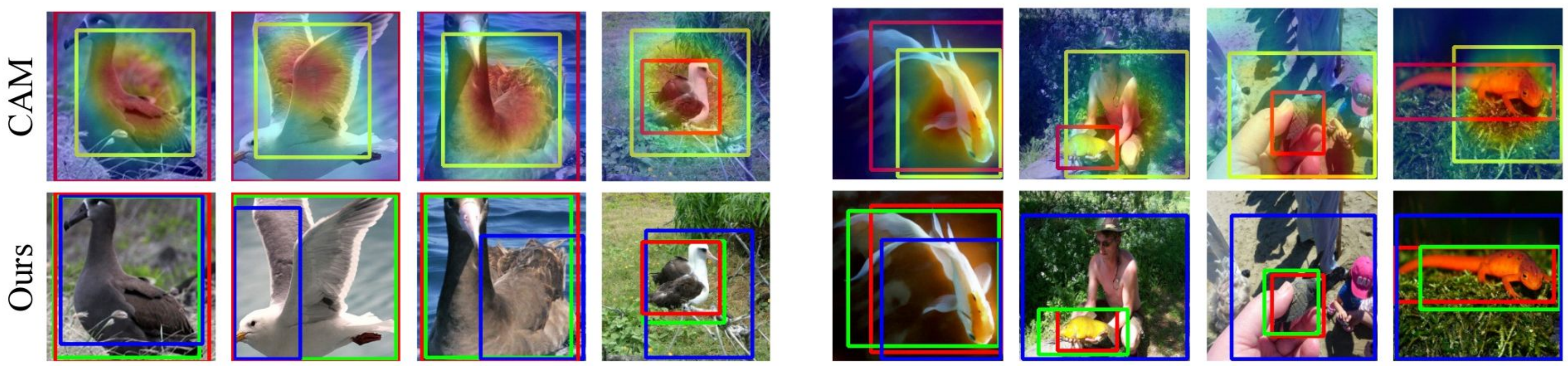
Taxonomy: Bottom-up

CAM refinement: Features enhancement

PSOL: Pseudo-Supervised Object Localization



Red: ground truth
Yellow: CAM predicted bbox
Green: PSOL predicted bbox



(a) CUB-200-2011 (b) ImageNet-1k

Zhang, C., Cao, Y., and Wu, J. (2020a). Rethinking the route towards weakly supervised object localization. In CVPR.

Part 2. Review of WSOL methods: Literature

Pseudo-annotation

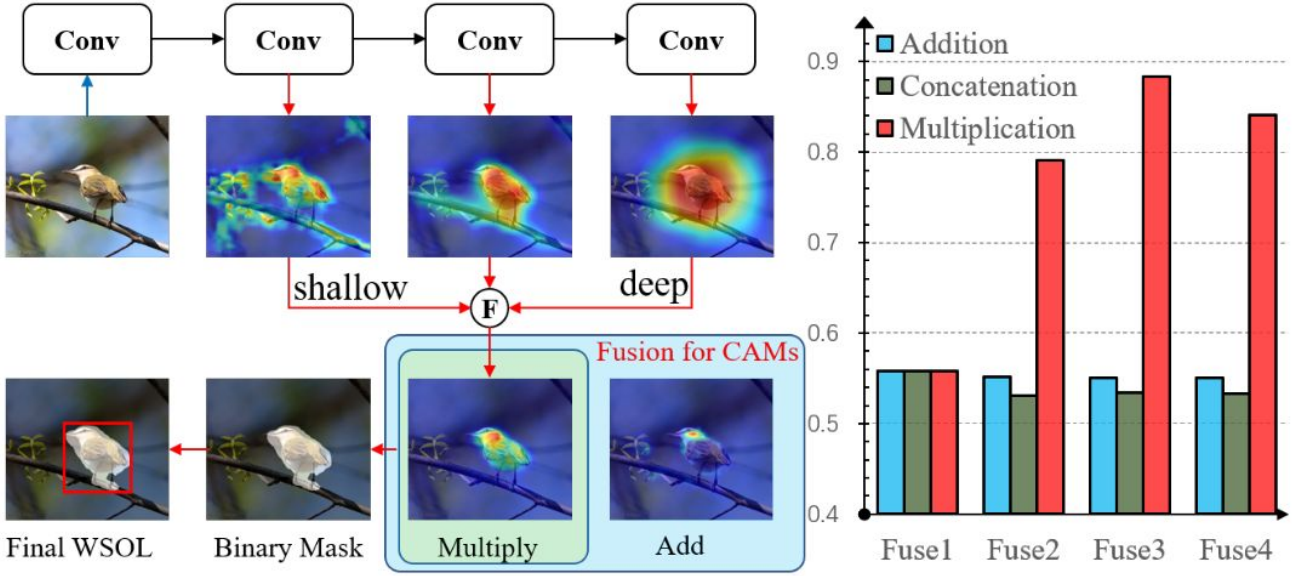
- SPG ('18)
- Pair-Sim ('20)
- PSOL ('20)
- SPOL ('21)
- F-CAM ('22)
- NEGEV ('22)

Taxonomy: Bottom-up

CAM refinement: Features enhancement

SPOL: Shallow **P**seudo-supervised **O**bject **L**ocalization

Shallow features are useful for localization!



(a) Improved CAMs from multiplicative fusion using shallow features for WSOL

(b) Different fusion strategies from different layers

Fusex: x last layers

Wei, J., Wang, Q., Li, Z., Wang, S., Zhou, S. K., and Cui, S. (2021). Shallow feature matters for weakly supervised object localization. In CVPR.

Part 2. Review of WSOL methods: Literature

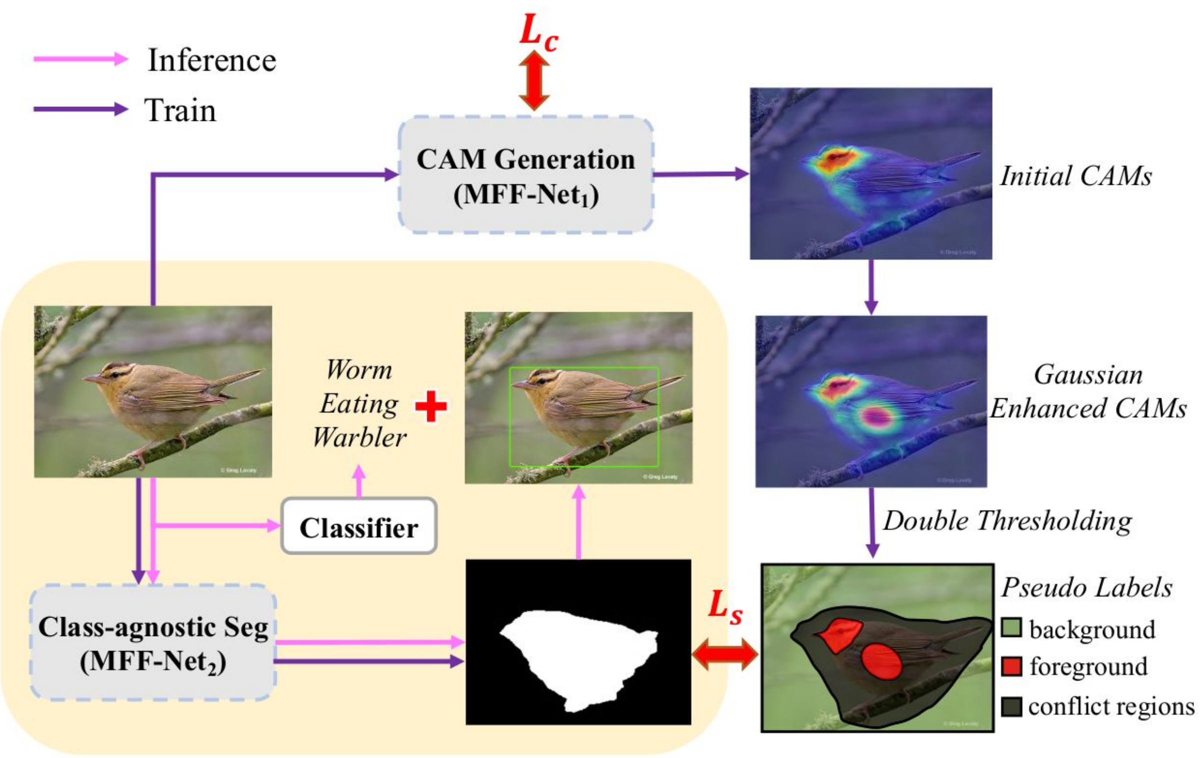
Pseudo-annotation

- SPG ('18)
- Pair-Sim ('20)
- PSOL ('20)
- SPOL ('21)
- F-CAM ('22)
- NEGEV ('22)

Taxonomy: Bottom-up

CAM refinement: Features enhancement

SPOL: Shallow **P**seudo-supervised **O**bject **L**ocalization



(a) SPOL

Wei, J., Wang, Q., Li, Z., Wang, S., Zhou, S. K., and Cui, S. (2021). Shallow feature matters for weakly supervised object localization. In CVPR.

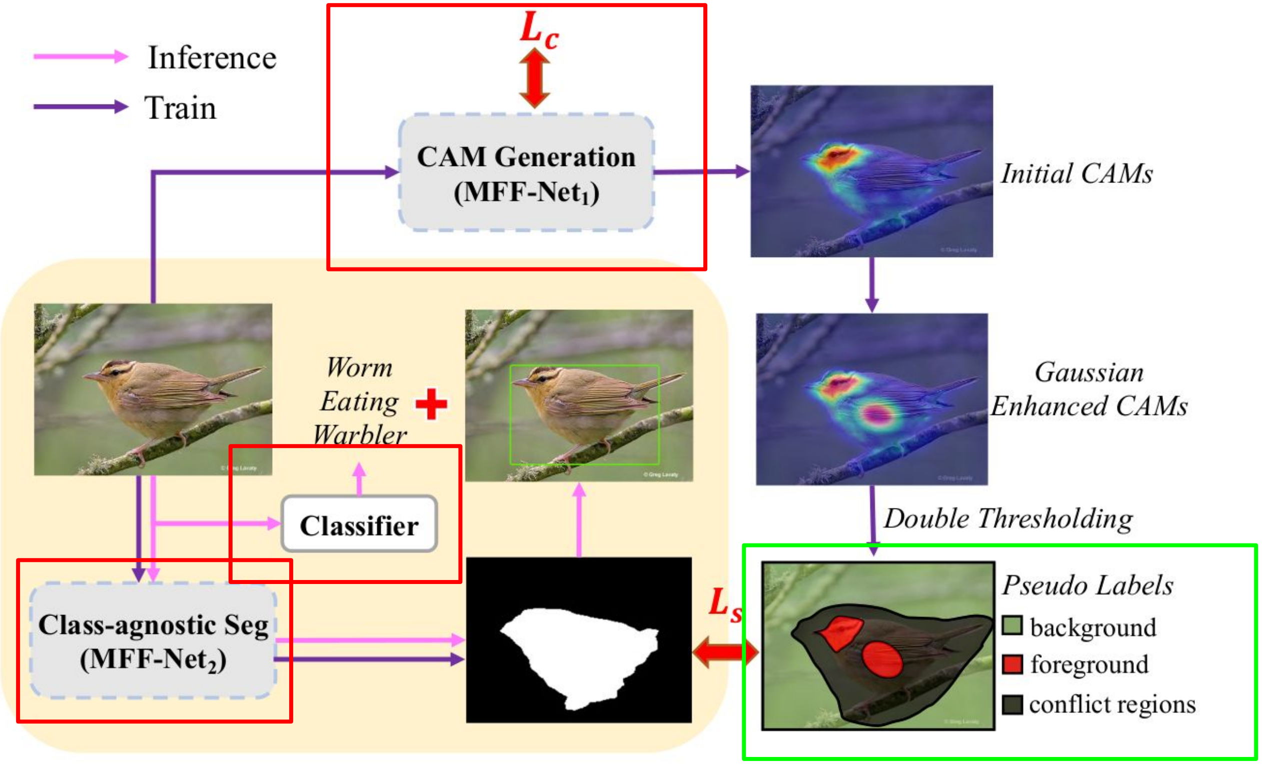
Part 2. Review of WSOL methods: Literature

Pseudo-annotation
- SPG ('18)
- Pair-Sim ('20)
- PSOL ('20)
- SPOL ('21)
- F-CAM ('22)
- NEGEV ('22)

Taxonomy: Bottom-up

CAM refinement: Features enhancement

SPOL: Shallow **P**seudo-supervised **O**bject **L**ocalization



(a) SPOL

Train loss:

Lc: standard cross-entropy

Ls: partial cross-entropy over only foreground and background

Wei, J., Wang, Q., Li, Z., Wang, S., Zhou, S. K., and Cui, S. (2021). Shallow feature matters for weakly supervised object localization. In CVPR.

ÉTS Le génie pour l'industrie

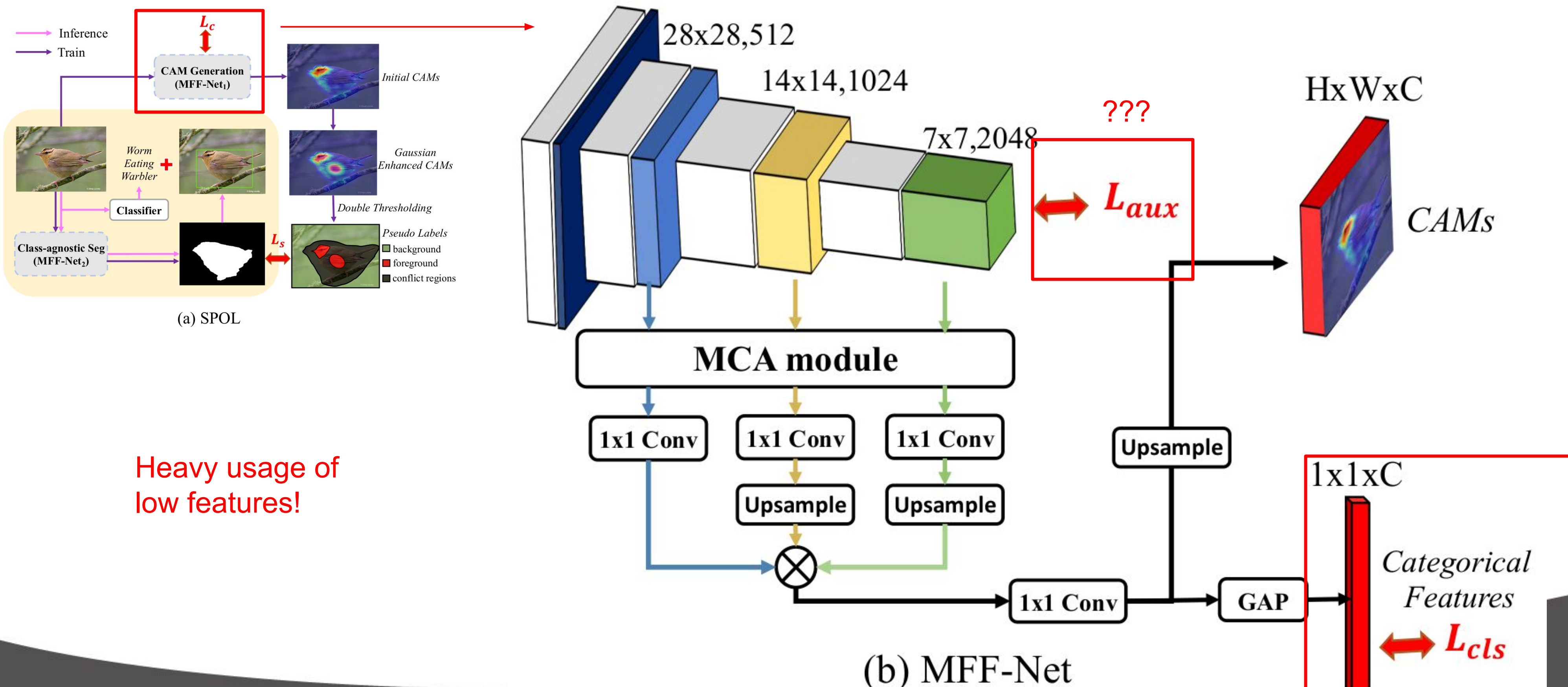
Part 2. Review of WSOL methods: Literature

Pseudo-annotation
- SPG ('18)
- Pair-Sim ('20)
- PSOL ('20)
- SPOL ('21)
- F-CAM ('22)
- NEGEV ('22)

Taxonomy: Bottom-up

CAM refinement: Features enhancement

SPOL: Shallow Pseudo-supervised Object Localization



Heavy usage of low features!

Wei, J., Wang, Q., Li, Z., Wang, S., Zhou, S. K., and Cui, S. (2021). Shallow feature matters for weakly supervised object localization. In CVPR.

Part 2. Review of WSOL methods: Literature

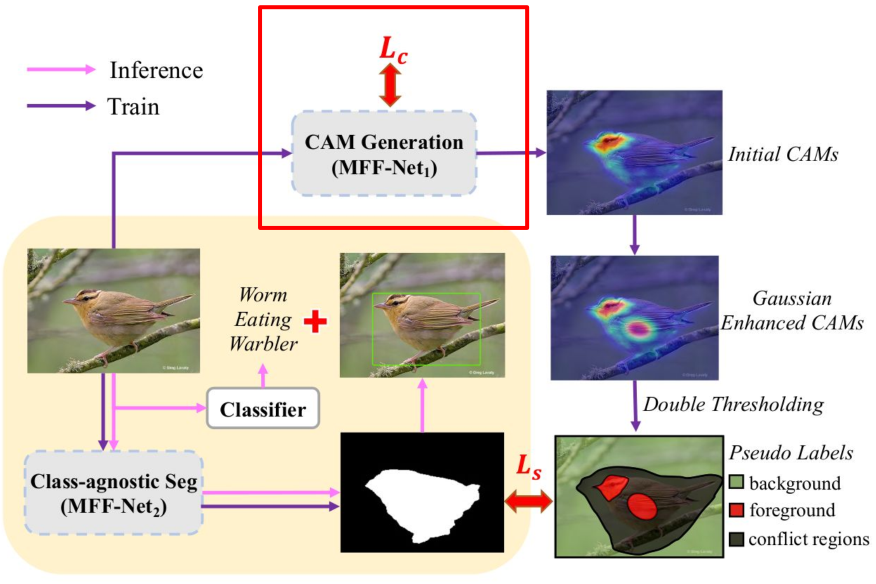
Pseudo-annotation

- SPG ('18)
- Pair-Sim ('20)
- PSOL ('20)
- SPOL ('21)
- F-CAM ('22)
- NEGEV ('22)

Taxonomy: Bottom-up

CAM refinement: Features enhancement

SPOL: Shallow **P**seudo-supervised **O**bject **L**ocalization



(a) SPOL

Algorithm 1: SPOL

Input: Training images I_{tr} with class label L_{tr}
Output: Predicted bounding boxes B_{te} and class labels L_{te} on testing images I_{te}

1 // Training Phase
2 Train MFF-Net$_1$ F_w on I_{tr} with L_{tr}
3 Use F_w to generate pseudo label M_{tr} on I_{tr}
4 Train MFF-Net$_2$ F_s on I_{tr} for Seg. with M_{tr}
5 Train a classifier F_c on I_{tr} with L_{tr}
6 // Inference Phase
7 Use F_s to predict M_{te} on I_{te}
8 Extract object bounding box B_{te} from M_{te}
9 Use F_c to predict L_{te} on I_{te}
10 **Return:** B_{te}, L_{te}

Wei, J., Wang, Q., Li, Z., Wang, S., Zhou, S. K., and Cui, S. (2021). Shallow feature matters for weakly supervised object localization. In CVPR.

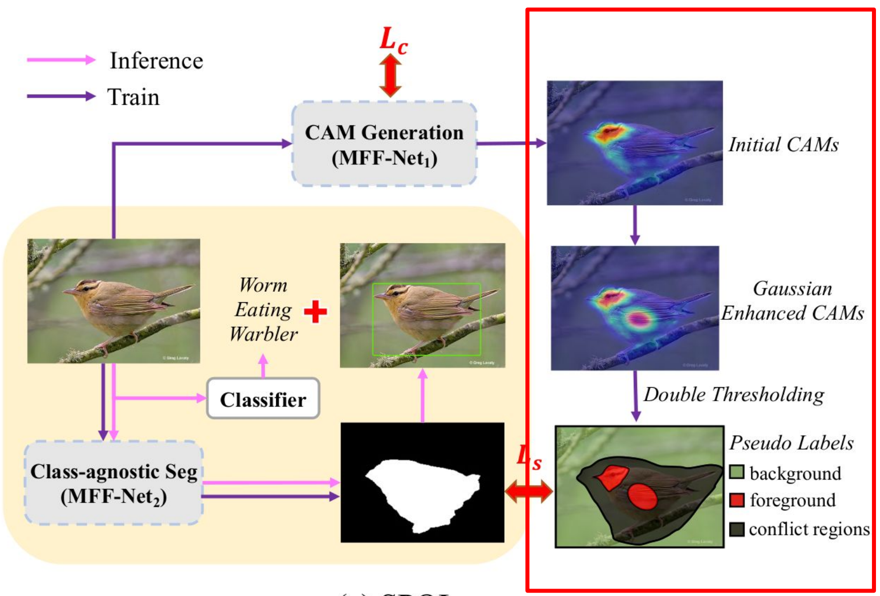
Part 2. Review of WSOL methods: Literature

Pseudo-annotation
- SPG ('18)
- Pair-Sim ('20)
- PSOL ('20)
- SPOL ('21)
- F-CAM ('22)
- NEGEV ('22)

Taxonomy: Bottom-up

CAM refinement: Features enhancement

SPOL: Shallow **P**seudo-supervised **O**bject **L**ocalization



(a) SPOL

Algorithm 1: SPOL

Input: Training images I_{tr} with class label L_{tr}
Output: Predicted bounding boxes B_{te} and class labels L_{te} on testing images I_{te}

// Training Phase
Train MFF-Net$_1$ F_w on I_{tr} with L_{tr}
Use F_w to generate pseudo label M_{tr} on I_{tr}
Train MFF-Net$_2$ F_s on I_{tr} for Seg. with M_{tr}
Train a classifier F_c on I_{tr} with L_{tr}

// Inference Phase
Use F_s to predict M_{te} on I_{te}
Extract object bounding box B_{te} from M_{te}
Use F_c to predict L_{te} on I_{te}
Return: B_{te}, L_{te}

Wei, J., Wang, Q., Li, Z., Wang, S., Zhou, S. K., and Cui, S. (2021). Shallow feature matters for weakly supervised object localization. In CVPR.

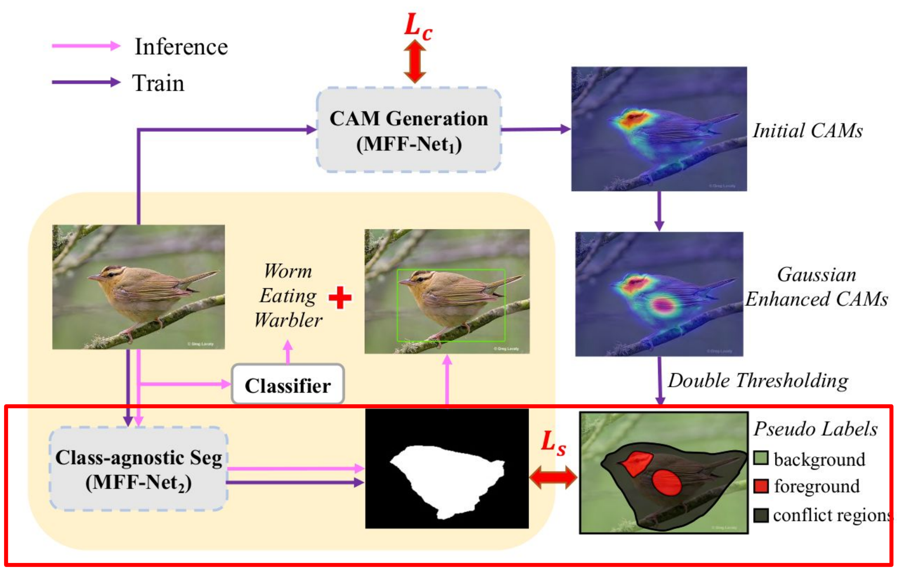
Part 2. Review of WSOL methods: Literature

Pseudo-annotation
- SPG ('18)
- Pair-Sim ('20)
- PSOL ('20)
- SPOL ('21)
- F-CAM ('22)
- NEGEV ('22)

Taxonomy: Bottom-up

CAM refinement: Features enhancement

SPOL: Shallow Pseudo-supervised Object Localization



(a) SPOL

Algorithm 1: SPOL

Input: Training images I_{tr} with class label L_{tr}

Output: Predicted bounding boxes B_{te} and class labels L_{te} on testing images I_{te}

// Training Phase
Train MFF-Net$_1$ F_w on I_{tr} with L_{tr}
Use F_w to generate pseudo label M_{tr} on I_{tr}
Train MFF-Net$_2$ F_s on I_{tr} for Seg. with M_{tr}
Train a classifier F_c on I_{tr} with L_{tr}

// Inference Phase
Use F_s to predict M_{te} on I_{te}
Extract object bounding box B_{te} from M_{te}
Use F_c to predict L_{te} on I_{te}
Return: B_{te}, L_{te}

Wei, J., Wang, Q., Li, Z., Wang, S., Zhou, S. K., and Cui, S. (2021). Shallow feature matters for weakly supervised object localization. In CVPR.

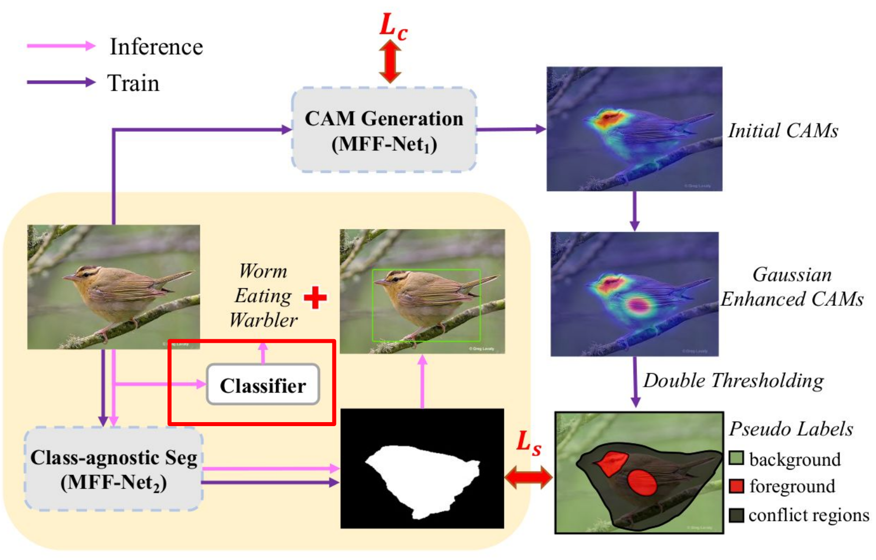
Part 2. Review of WSOL methods: Literature

Pseudo-annotation
- SPG ('18)
- Pair-Sim ('20)
- PSOL ('20)
- SPOL ('21)
- F-CAM ('22)
- NEGEV ('22)

Taxonomy: Bottom-up

CAM refinement: Features enhancement

SPOL: Shallow Pseudo-supervised Object Localization



(a) SPOL

Algorithm 1: SPOL

Input: Training images I_{tr} with class label L_{tr}

Output: Predicted bounding boxes B_{te} and class labels L_{te} on testing images I_{te}

// Training Phase
Train MFF-Net$_1$ F_w on I_{tr} with L_{tr}
Use F_w to generate pseudo label M_{tr} on I_{tr}
Train MFF-Net$_2$ F_s on I_{tr} for Seg. with M_{tr}
Train a classifier F_c on I_{tr} with L_{tr}

// Inference Phase
Use F_s to predict M_{te} on I_{te}
Extract object bounding box B_{te} from M_{te}
Use F_c to predict L_{te} on I_{te}
Return: B_{te}, L_{te}

Wei, J., Wang, Q., Li, Z., Wang, S., Zhou, S. K., and Cui, S. (2021). Shallow feature matters for weakly supervised object localization. In CVPR.

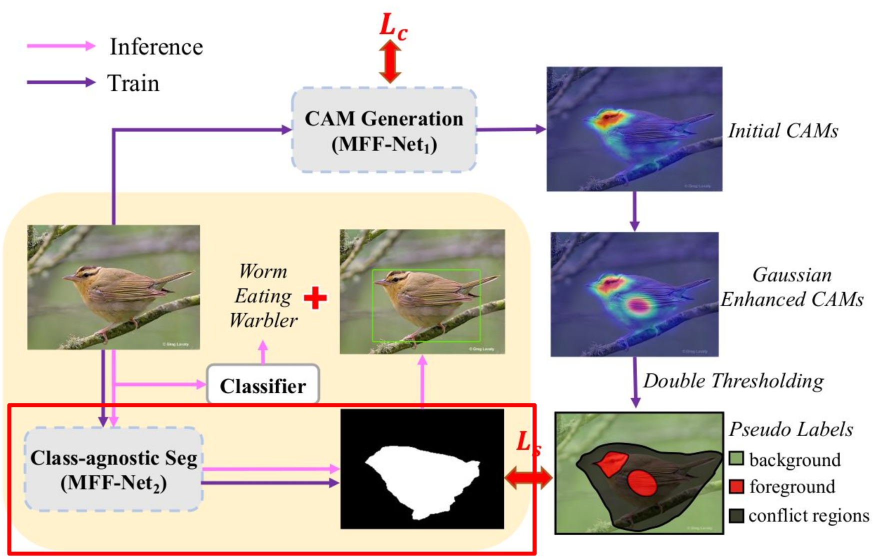
Part 2. Review of WSOL methods: Literature

Pseudo-annotation
- SPG ('18)
- Pair-Sim ('20)
- PSOL ('20)
- SPOL ('21)
- F-CAM ('22)
- NEGEV ('22)

Taxonomy: Bottom-up

CAM refinement: Features enhancement

SPOL: **S**hallow **P**seudo-supervised **O**bject **L**ocalization



(a) SPOL

Algorithm 1: SPOL

Input: Training images I_{tr} with class label L_{tr}

Output: Predicted bounding boxes B_{te} and class labels L_{te} on testing images I_{te}

1 // Training Phase
2 Train MFF-Net$_1$ F_w on I_{tr} with L_{tr}
3 Use F_w to generate pseudo label M_{tr} on I_{tr}
4 Train MFF-Net$_2$ F_s on I_{tr} for Seg. with M_{tr}
5 Train a classifier F_c on I_{tr} with L_{tr}
6 // Inference Phase
7 Use F_s to predict M_{te} on I_{te}
8 Extract object bounding box B_{te} from M_{te}
9 Use F_c to predict L_{te} on I_{te}
10 **Return:** B_{te}, L_{te}

Wei, J., Wang, Q., Li, Z., Wang, S., Zhou, S. K., and Cui, S. (2021). Shallow feature matters for weakly supervised object localization. In CVPR.

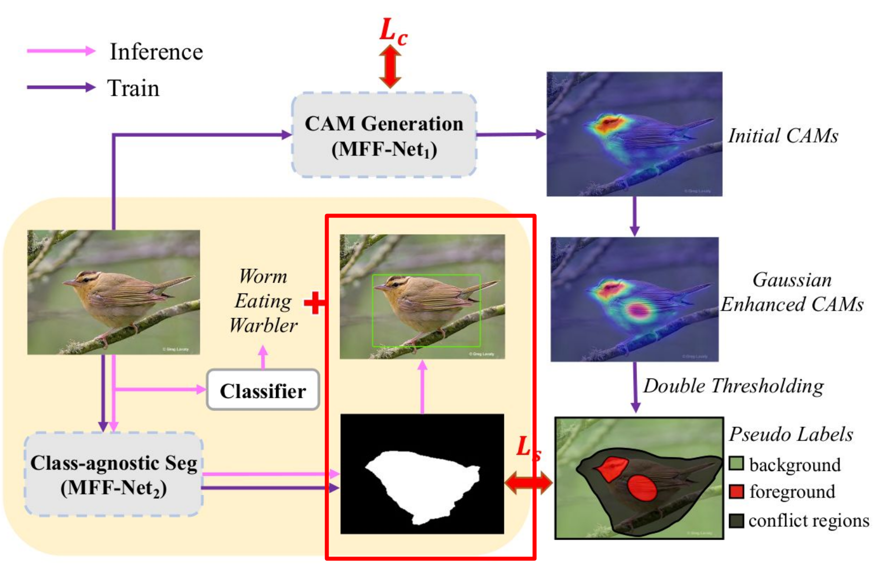
Part 2. Review of WSOL methods: Literature

Pseudo-annotation
- SPG ('18)
- Pair-Sim ('20)
- PSOL ('20)
- SPOL ('21)
- F-CAM ('22)
- NEGEV ('22)

Taxonomy: Bottom-up

CAM refinement: Features enhancement

SPOL: Shallow **P**seudo-supervised **O**bject **L**ocalization



(a) SPOL

Algorithm 1: SPOL

Input: Training images I_{tr} with class label L_{tr}
Output: Predicted bounding boxes B_{te} and class labels L_{te} on testing images I_{te}

// Training Phase
Train MFF-Net$_1$ F_w on I_{tr} with L_{tr}
Use F_w to generate pseudo label M_{tr} on I_{tr}
Train MFF-Net$_2$ F_s on I_{tr} for Seg. with M_{tr}
Train a classifier F_c on I_{tr} with L_{tr}
// Inference Phase
Use F_s to predict M_{te} on I_{te}
Extract object bounding box B_{te} from M_{te}
Use F_c to predict L_{te} on I_{te}
Return: B_{te}, L_{te}

Part 2. Review of WSOL methods: Literature

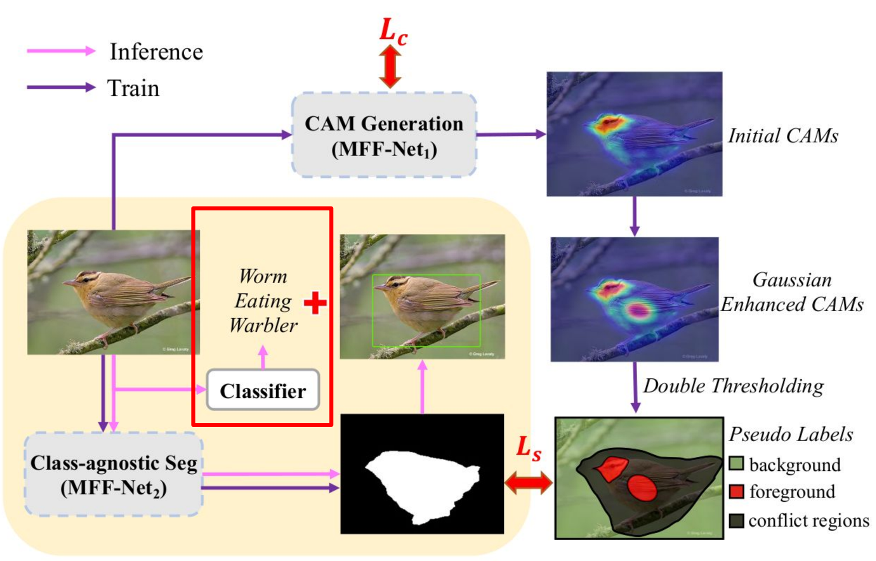
Pseudo-annotation

- SPG ('18)
- Pair-Sim ('20)
- PSOL ('20)
- SPOL ('21)
- F-CAM ('22)
- NEGEV ('22)

Taxonomy: Bottom-up

CAM refinement: Features enhancement

SPOL: Shallow **P**seudo-supervised **O**bject **L**ocalization



(a) SPOL

Algorithm 1: SPOL

Input: Training images I_{tr} with class label L_{tr}
Output: Predicted bounding boxes B_{te} and class labels L_{te} on testing images I_{te}

// Training Phase
Train MFF-Net$_1$ F_w on I_{tr} with L_{tr}
Use F_w to generate pseudo label M_{tr} on I_{tr}
Train MFF-Net$_2$ F_s on I_{tr} for Seg. with M_{tr}
Train a classifier F_c on I_{tr} with L_{tr}

// Inference Phase
Use F_s to predict M_{te} on I_{te}
Extract object bounding box B_{te} from M_{te}
Use F_c to predict L_{te} on I_{te}
Return: B_{te}, L_{te}

Wei, J., Wang, Q., Li, Z., Wang, S., Zhou, S. K., and Cui, S. (2021). Shallow feature matters for weakly supervised object localization. In CVPR.

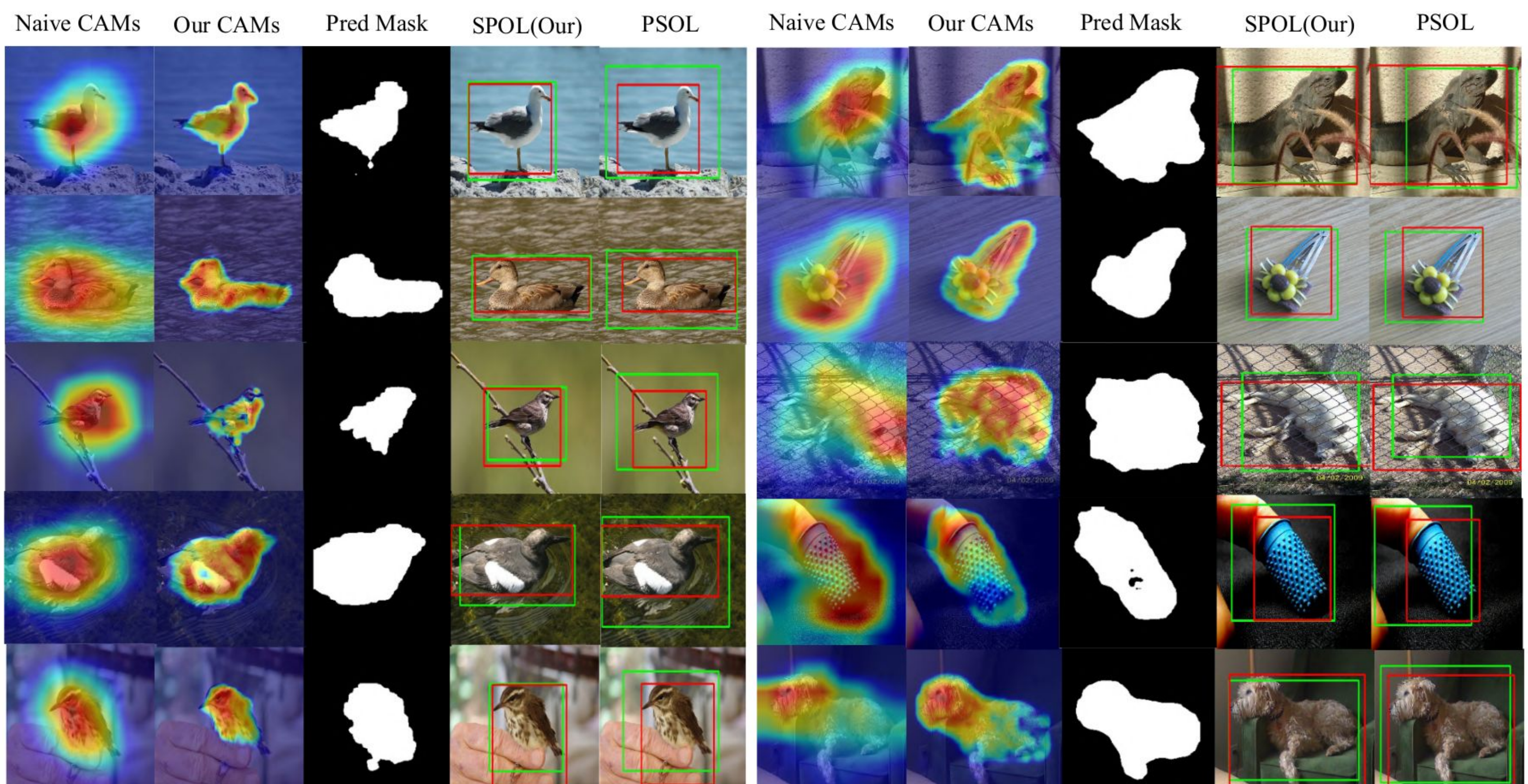
Part 2. Review of WSOL methods: Literature

Pseudo-annotation
- SPG ('18)
- Pair-Sim ('20)
- PSOL ('20)
- SPOL ('21)
- F-CAM ('22)
- NEGEV ('22)

Taxonomy: Bottom-up

CAM refinement: Features enhancement

SPOL: Shallow Pseudo-supervised Object Localization

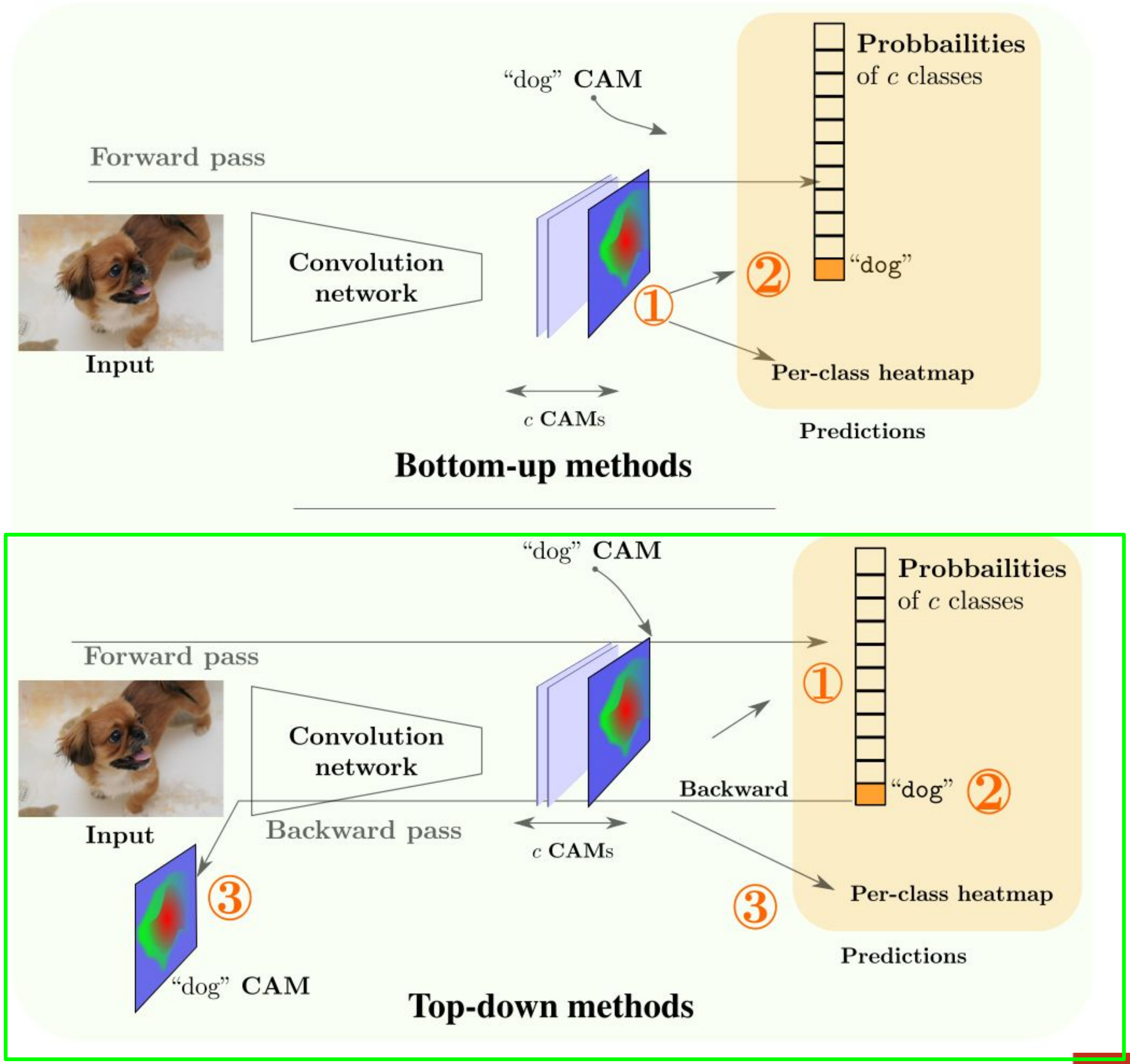


(a) CUB200 dataset (b) ImageNet-1K dataset

Wei, J., Wang, Q., Li, Z., Wang, S., Zhou, S. K., and Cui, S. (2021). Shallow feature matters for weakly supervised object localization. In CVPR.

Part 2. Review of WSOL methods: Literature

Taxonomy

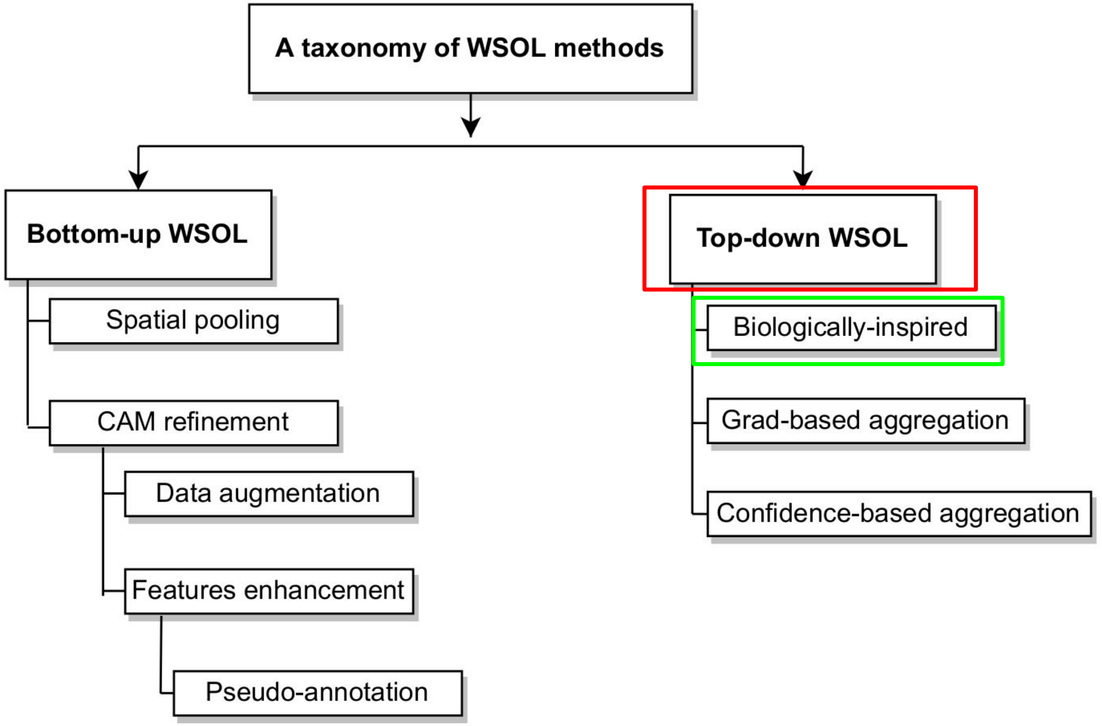


Less common in WSOL

Part 2. Review of WSOL methods: Literature

Taxonomy

- Cognitive science
- Human visual attention (top-down mechanism)



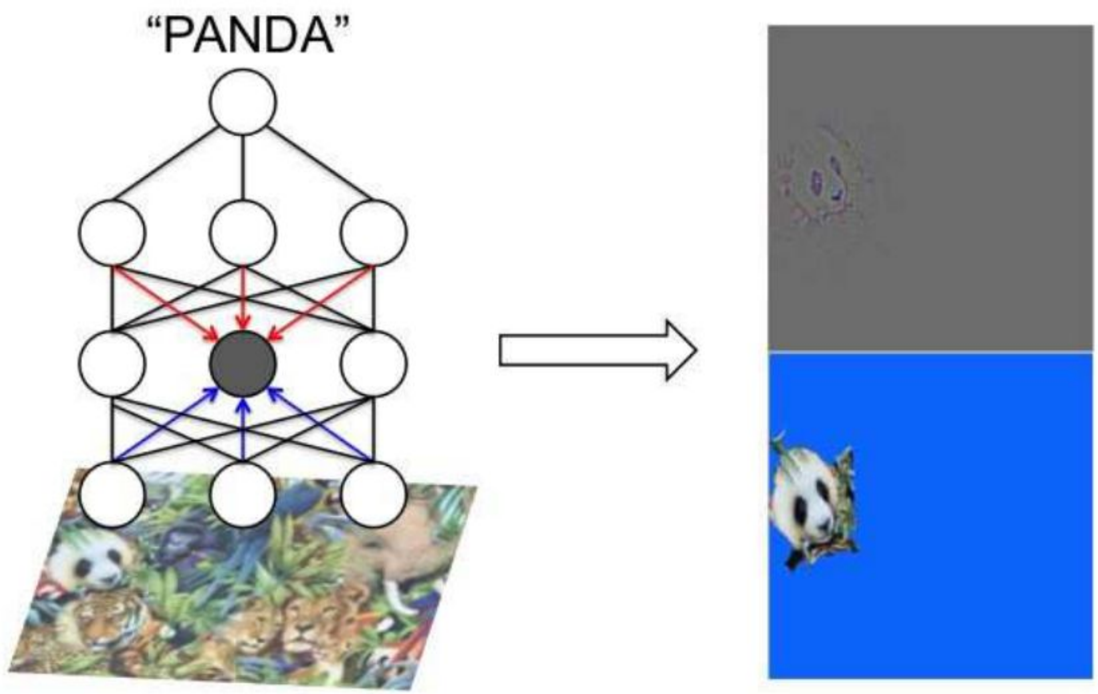
Part 2. Review of WSOL methods: Literature

Biologically-inspired
- Feedback layer ('15)
- Excitation-backprop ('18)

Taxonomy: Top-Down
Biologically-inspired

Feedback layer

Use bottom-up
input image
top-down semantic
label to infer
hidden neuron
activation



Part 2. Review of WSOL methods: Literature

- Biologically-inspired
  - Feedback layer ('15)
  - Excitation-backprop ('18)

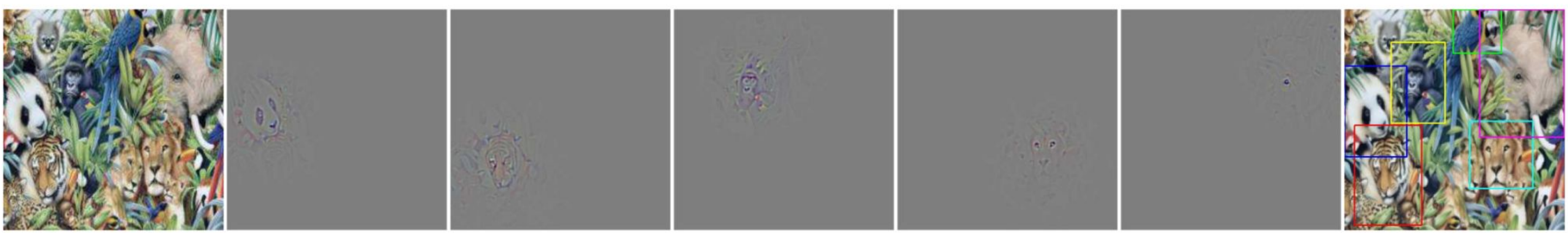
Taxonomy: Top-Down

Biologically-inspired

Feedback layer

Localization over input neurons

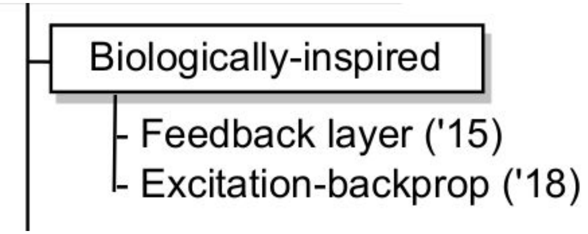
Bounding box localization



(a) Input Image (b) Panda (c) Tiger (d) Gorilla (e) Lion (f) Elephant (g) Localization

Cao, C., Liu, X., Yang, Y., et al. (2015). Look and think twice: Capturing top-down visual attention with feedback convolutional neural networks. In ICCV.

Part 2. Review of WSOL methods: Literature

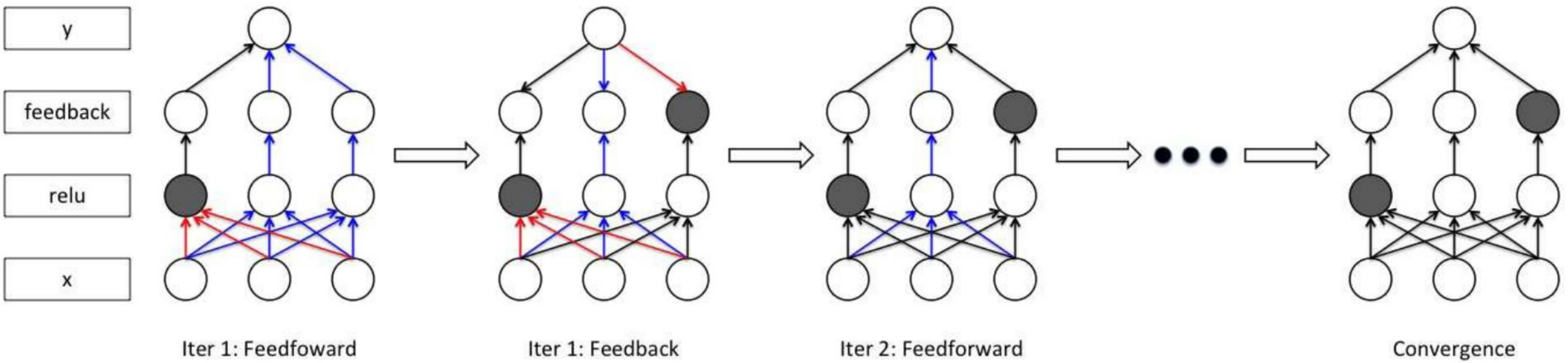


Taxonomy: Top-Down
Biologically-inspired

Feedback layer

Feedback layer: network of binary control gates **z**

Iterative optimization process to find **z**



Cao, C., Liu, X., Yang, Y., et al. (2015). Look and think twice: Capturing top-down visual attention with feedback convolutional neural networks. In ICCV.

Part 2. Review of WSOL methods: Literature

Biologically-inspired
- Feedback layer ('15)
- Excitation-backprop ('18)

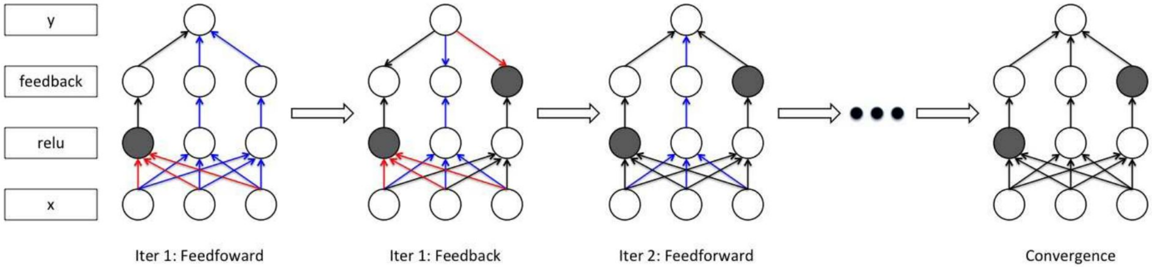
Taxonomy: Top-Down
Biologically-inspired

Feedback layer

Feedback layer: network of binary control gates **z**

Promote selectivity

Iterative optimization process to find **z** **(relaxed)**



$$\max_{\mathbf{z}} \quad s_k(I, \mathbf{z}) - \lambda ||\mathbf{z}||_1$$
$$s.t. \quad 0 \leq z^l_{i,j,c} \leq 1, \ \forall\, l, i, j, c$$

Gradient ascent:

$$\mathbf{z}_{t+1} = \mathbf{z}_t + \alpha \cdot \left(\frac{\partial s_k}{\partial \mathbf{z}} |_{\mathbf{z}_t} - \lambda \right)$$

Le génie pour l'industrie

Cao, C., Liu, X., Yang, Y., et al. (2015). Look and think twice: Capturing top-down visual attention with feedback convolutional neural networks. In ICCV.

- Biologically-inspired
  - Feedback layer ('15)
  - Excitation-backprop ('18)

Part 2. Review of WSOL methods: Literature

Taxonomy: Top-Down
Biologically-inspired

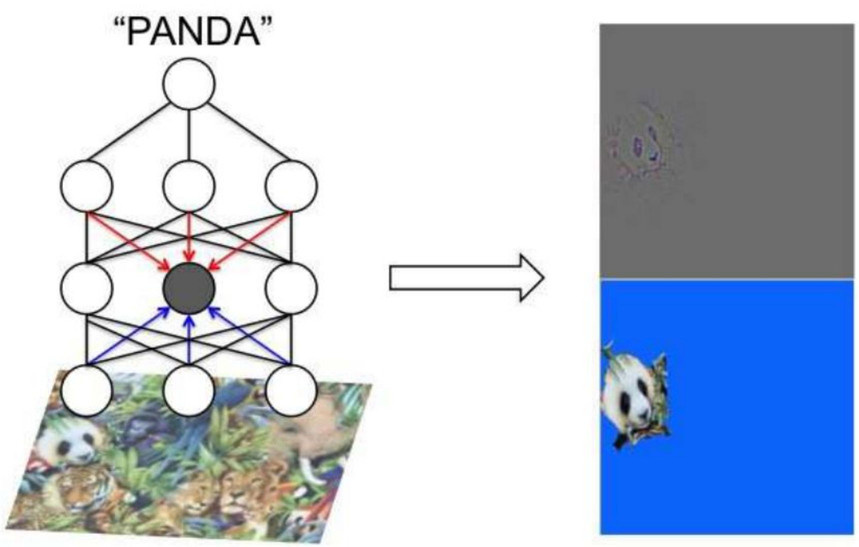
Feedback layer

Feedback layer: network of binary control gates **z**

Promote selectivity

Iterative optimization process to find **z** **(relaxed)**

$$\max_{\mathbf{z}} \quad s_k(I, \mathbf{z}) - \lambda ||\mathbf{z}||_1$$
$$s.t. \quad 0 \leq z^l_{i,j,c} \leq 1, \; \forall \, l, i, j, c$$



Gradient ascent:

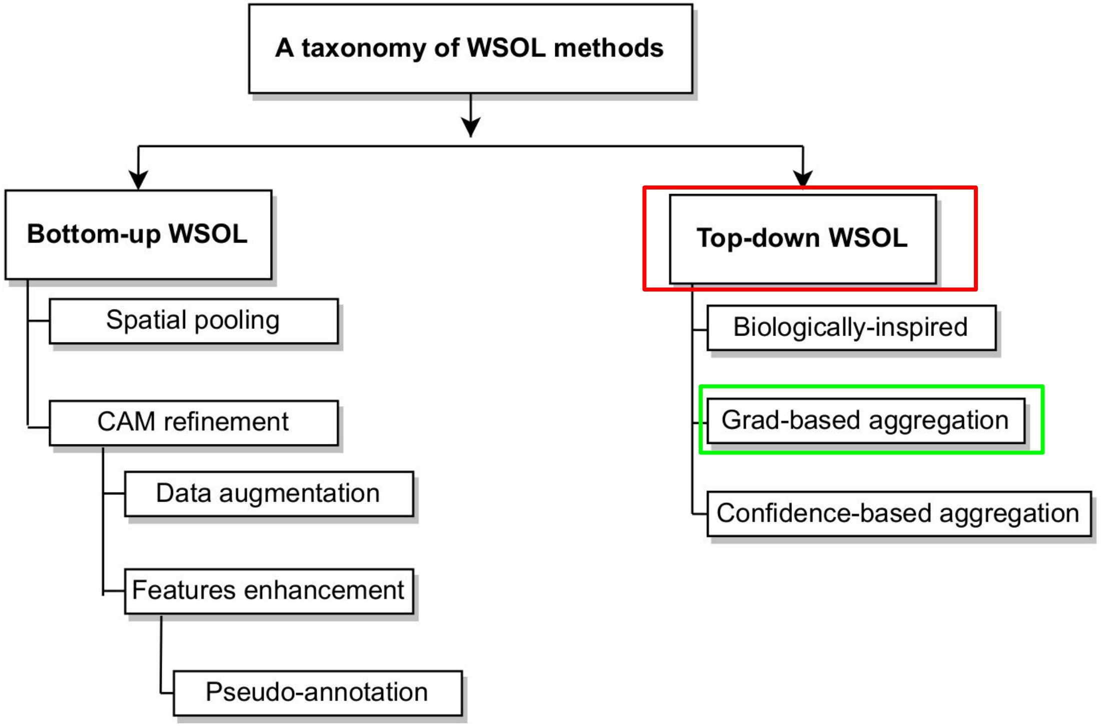
$$\mathbf{z}_{t+1} = \mathbf{z}_t + \alpha \cdot \left(\frac{\partial s_k}{\partial \mathbf{z}}\Big|_{\mathbf{z}_t} - \lambda\right)$$

Cao, C., Liu, X., Yang, Y., et al. (2015). Look and think twice: Capturing top-down visual attention with feedback convolutional neural networks. In ICCV.

Part 2. Review of WSOL methods: Literature

Taxonomy

- Aggregation of feature maps using gradient



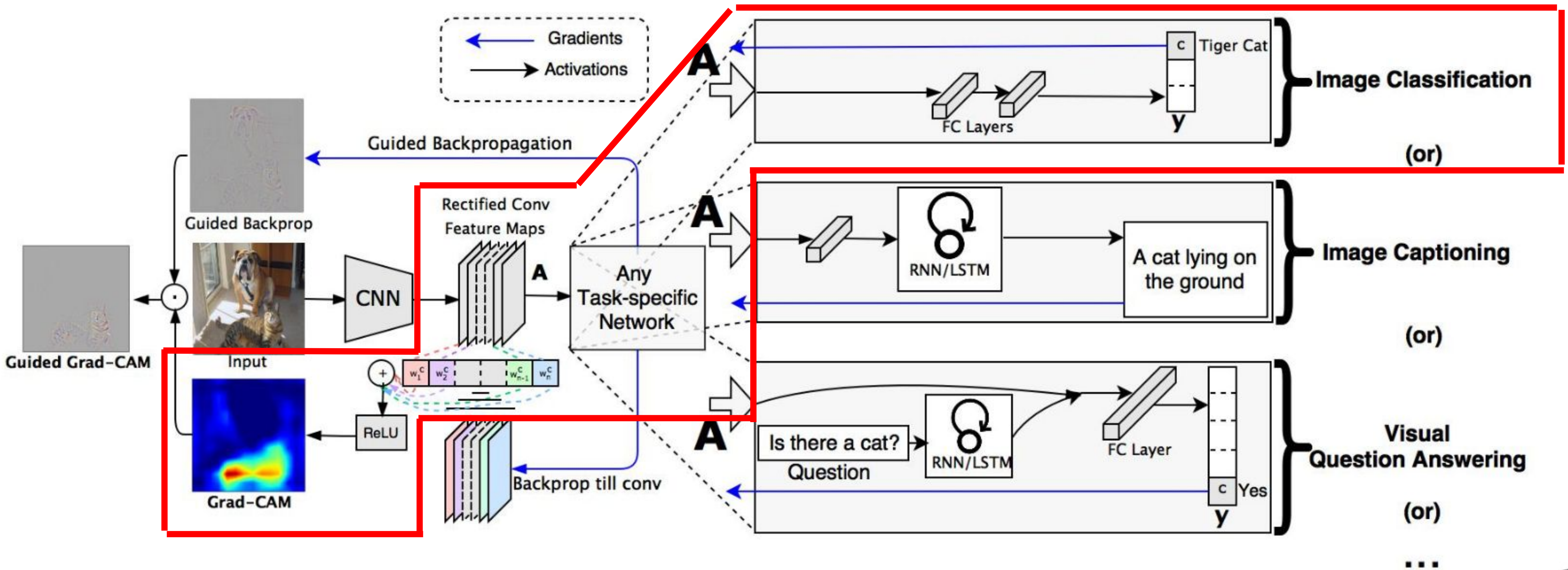
Part 2. Review of WSOL methods: Literature

Grad-based aggregation
- Grad-CAM ('17)
- Grad-CAM++ ('18)
- Smooth-Grad-CAM++ ('19)
- XGrad-CAM ('20)
- LayerCAM ('21)

Taxonomy: Top-Down

Grad-based aggregation

Grad-CAM



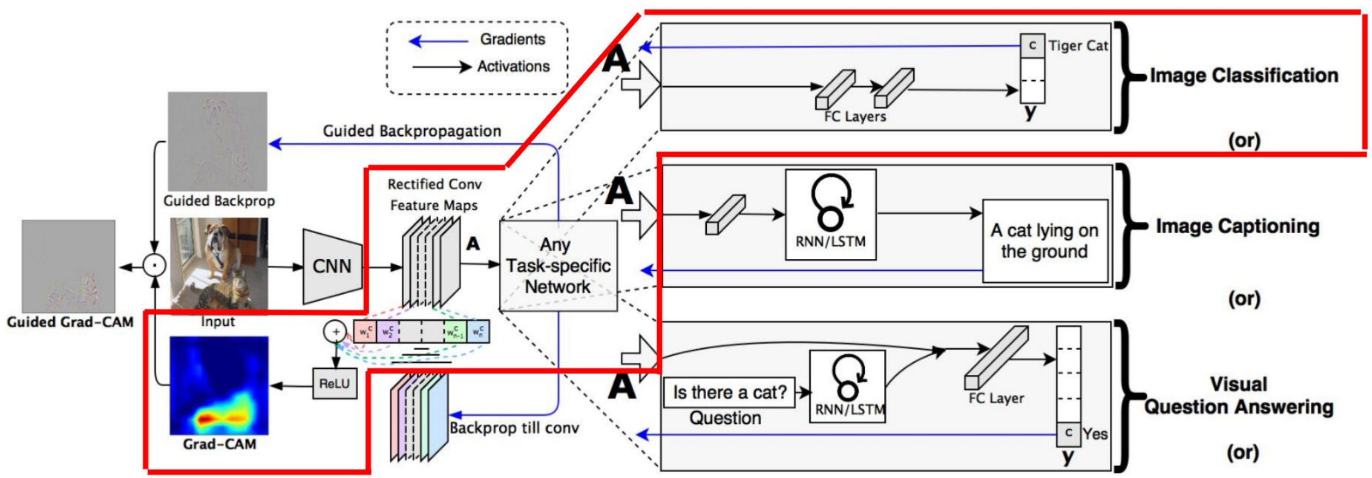
Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In ICCV.

Part 2. Review of WSOL methods: Literature

Grad-based aggregation
- Grad-CAM ('17)
- Grad-CAM++ ('18)
- Smooth-Grad-CAM++ ('19)
- XGrad-CAM ('20)
- LayerCAM ('21)

Taxonomy: Top-Down
Grad-based aggregation

Grad-CAM



$$L^c_{\text{Grad-CAM}} = ReLU \underbrace{\left(\sum_k \alpha^c_k A^k \right)}_{\text{linear combination}}$$

$$\alpha^c_k = \overbrace{\frac{1}{Z} \sum_i \sum_j}^{\text{global average pooling}} \underbrace{\frac{\partial y^c}{\partial A^k_{ij}}}_{\text{gradients via backprop}}$$

Le génie pour l'industrie

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In ICCV.

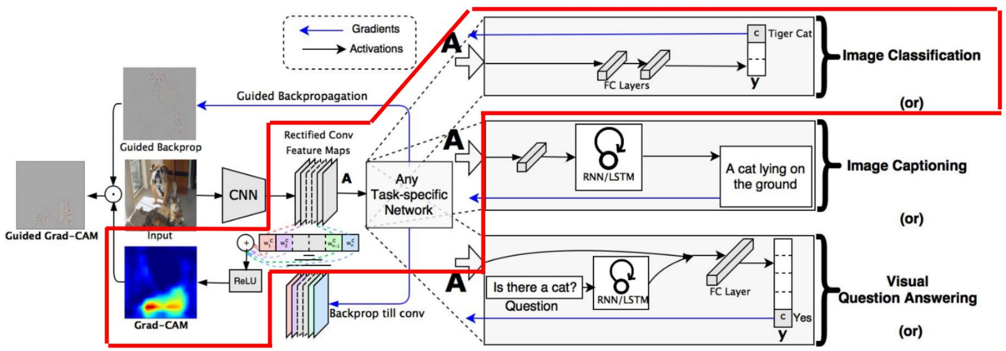
Part 2. Review of WSOL methods: Literature

Grad-based aggregation

- Grad-CAM ('17)
- Grad-CAM++ ('18)
- Smooth-Grad-CAM++ ('19)
- XGrad-CAM ('20)
- LayerCAM ('21)

Taxonomy: Top-Down
Grad-based aggregation

Grad-CAM

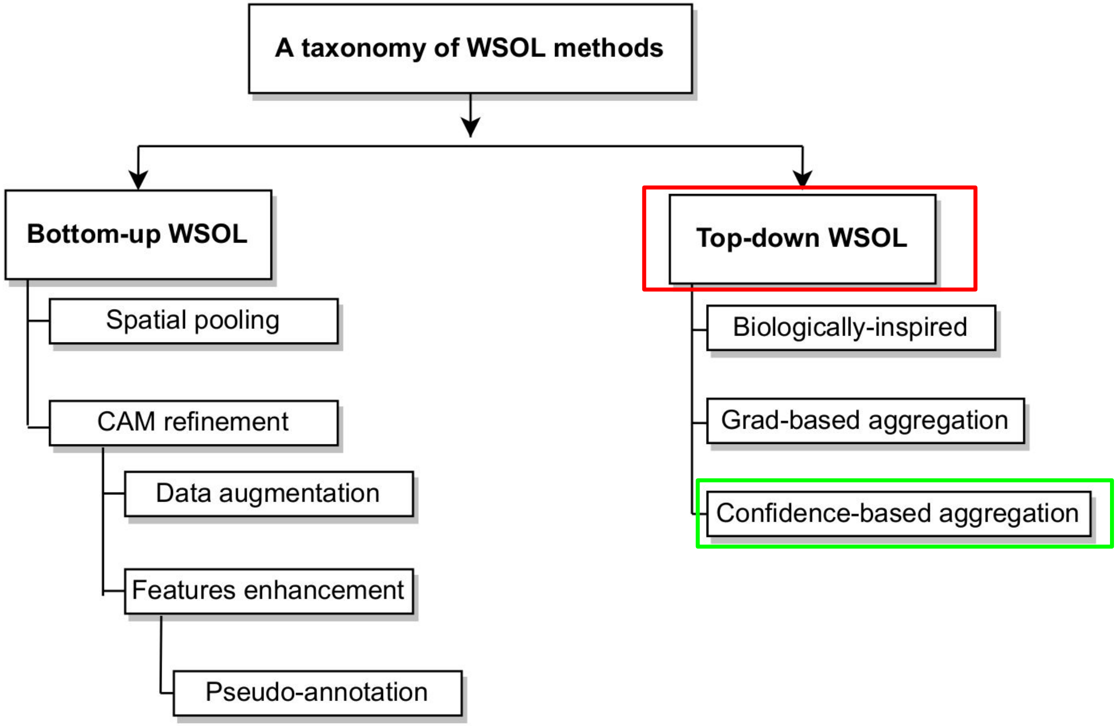


Dog Cat

Part 2. Review of WSOL methods: Literature

Taxonomy

- Aggregation of feature maps using classification confidence



Part 2. Review of WSOL methods: Literature

Confidence-based aggregation

- Score-CAM ('20)
- SS-CAM ('20)
- IS-CAM ('20)
- Ablation-CAM ('20)

Taxonomy: Top-Down

Confidence-based aggregation

Score-CAM

$$\mathbf{M}^{c}_{Score-CAM} = ReLU\left(\sum_{k} \alpha^{c}_{k} \mathbf{F}^{l}_{k}\right),$$

In function of classifier response

Le génie pour l'industrie

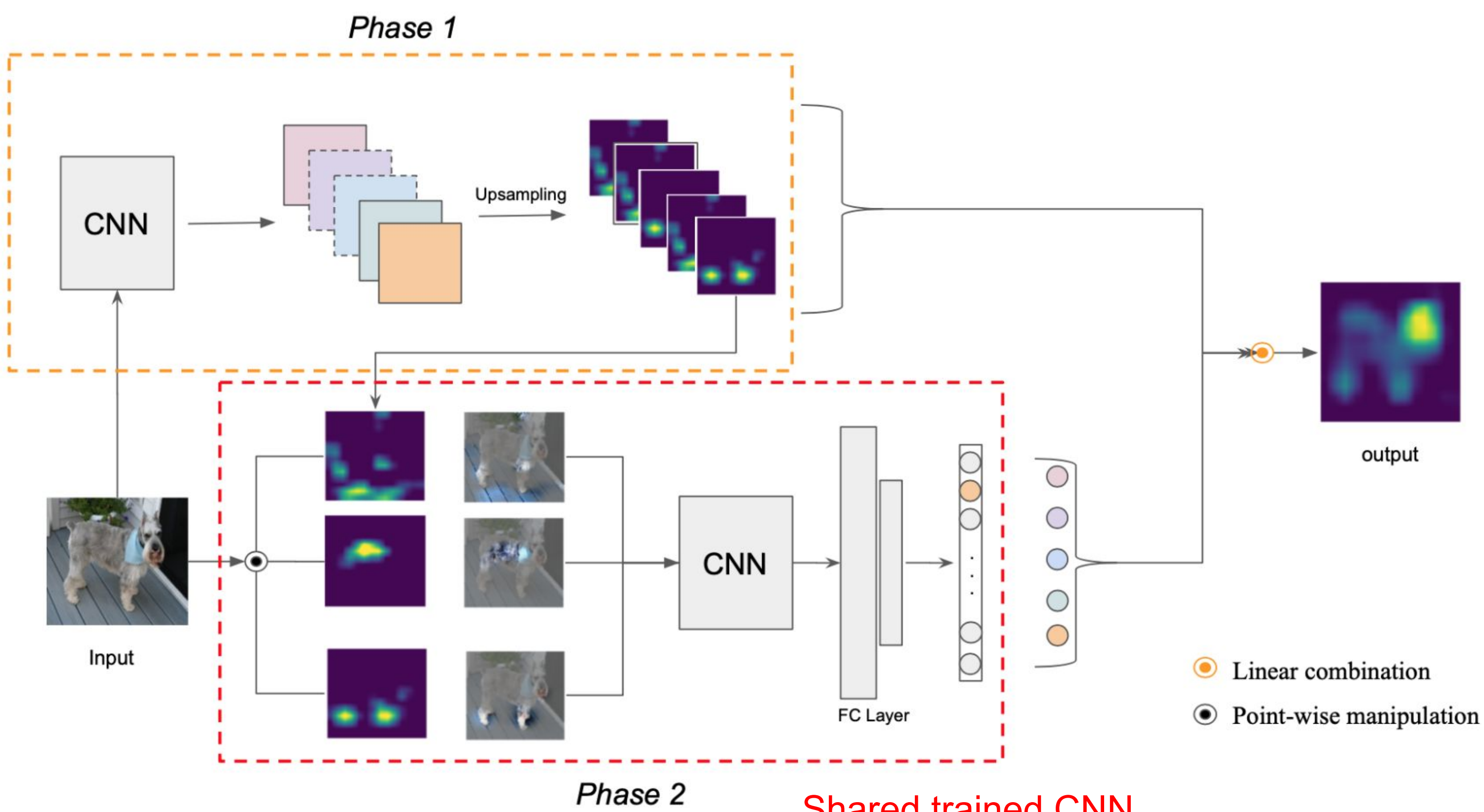
Wang, H., Wang, Z., Du, M., Yang, F., Zhang, Z., Ding, S., Mardziel, P., and Hu, X. (2020). Score-cam: Score-weighted visual explanations for convolutional neural networks. In CVPR workshop.

Part 2. Review of WSOL methods: Literatur

Confidence-based aggregation
- Score-CAM ('20)
- SS-CAM ('20)
- IS-CAM ('20)
- Ablation-CAM ('20)

Taxonomy: Top-Down
Confidence-based aggregation

Score-CAM



Shared trained CNN

Wang, H., Wang, Z., Du, M., Yang, F., Zhang, Z., Ding, S., Mardziel, P., and Hu, X. (2020). Score-cam: Score-weighted visual explanations for convolutional neural networks. In CVPR workshop.

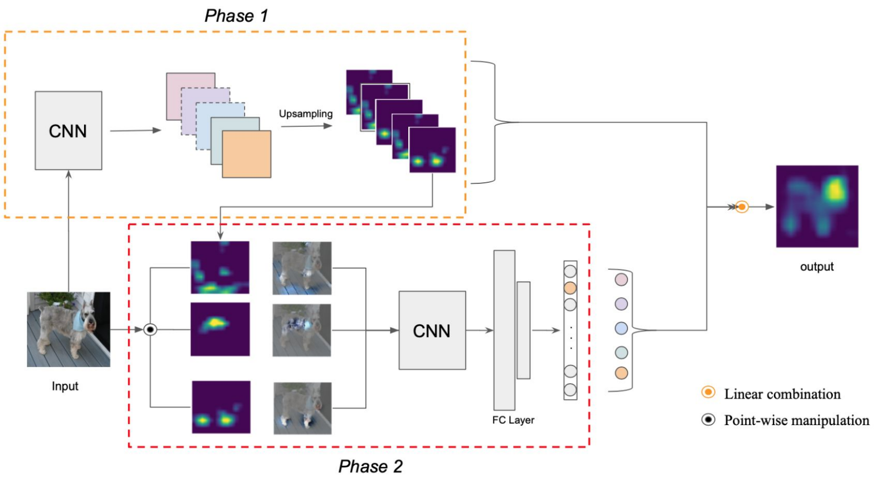
Part 2. Review of WSOL methods: Literatur

Confidence-based aggregation
- Score-CAM ('20)
- SS-CAM ('20)
- IS-CAM ('20)
- Ablation-CAM ('20)

Taxonomy: Top-Down

Confidence-based aggregation

Score-CAM



$$\mathbf{M}^{c}_{Score-CAM} = ReLU\left(\sum_{k} \alpha^{c}_{k} \mathbf{F}^{l}_{k}\right) \ ,$$

$$\alpha^{c}_{k} = C(A^{k}_{l})$$

Channel-wise Increase of Confidence (CIC)

$$C(A^{k}_{l}) = f(\underline{X \circ H^{k}_{l}}) - f(X_{b})$$

Perturbed input image

Reference image

Wang, H., Wang, Z., Du, M., Yang, F., Zhang, Z., Ding, S., Mardziel, P., and Hu, X. (2020). Score-cam: Score-weighted visual explanations for convolutional neural networks. In CVPR workshop.

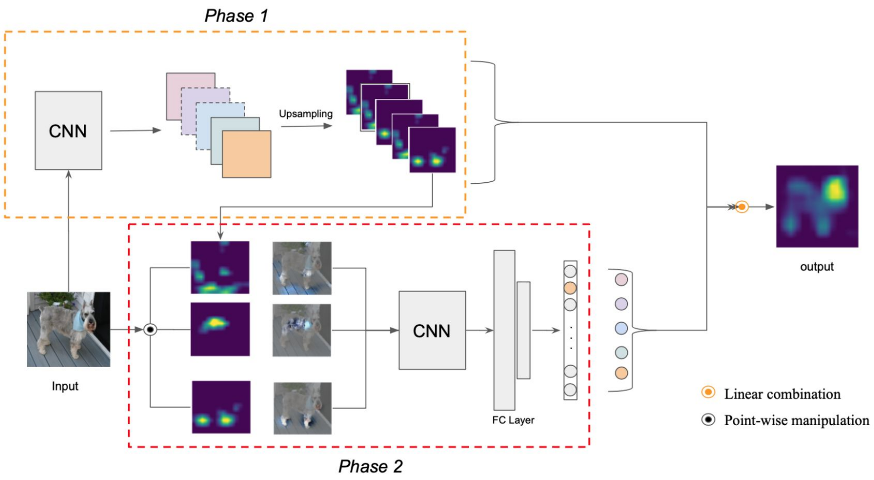
Le génie pour l'industrie

Part 2. Review of WSOL methods: Literatur

Confidence-based aggregation
- Score-CAM ('20)
- SS-CAM ('20)
- IS-CAM ('20)
- Ablation-CAM ('20)

Taxonomy: Top-Down
Confidence-based aggregation

Score-CAM



$$\mathbf{M}^c_{Score-CAM} = ReLU\left(\sum_k \alpha^c_k \mathbf{F}^l_k\right),$$

$$\alpha^c_k = C(A^k_l)$$

Channel-wise Increase of Confidence

$$C(A^k_l) = f(\underline{X \circ H^k_l}) - f(X_b)$$

Perturbed input image

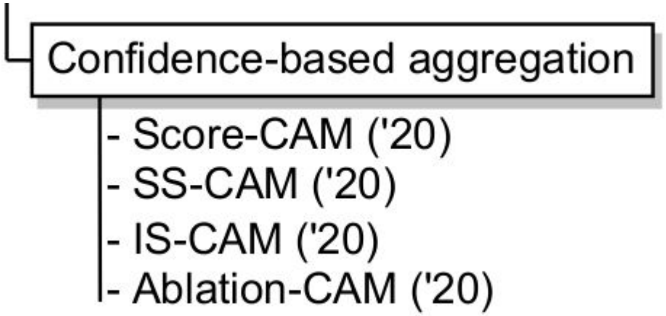
Reference image

Normalization:

$$\alpha^c_k = \frac{\exp\left(C(A^k_l)\right)}{\sum_k \exp\left(C(A^k_l)\right)}$$

Wang, H., Wang, Z., Du, M., Yang, F., Zhang, Z., Ding, S., Mardziel, P., and Hu, X. (2020). Score-cam: Score-weighted visual explanations for convolutional neural networks. In CVPR workshop.

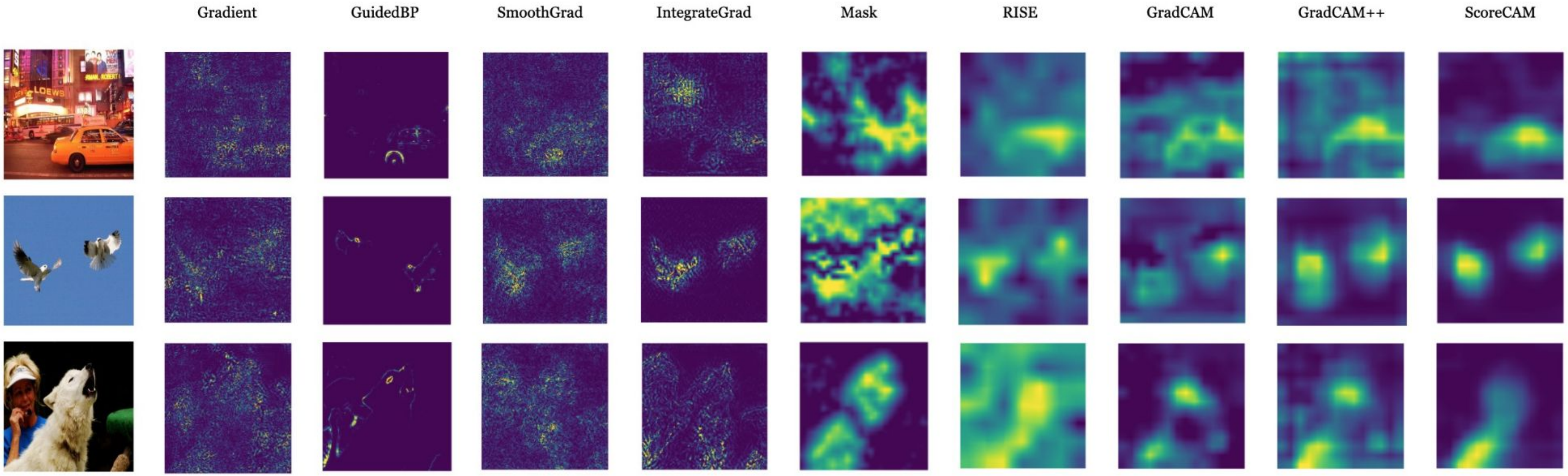
Part 2. Review of WSOL methods: Literatur



Taxonomy: Top-Down

Confidence-based aggregation

Score-CAM



Wang, H., Wang, Z., Du, M., Yang, F., Zhang, Z., Ding, S., Mardziel, P., and Hu, X. (2020). Score-cam: Score-weighted visual explanations for convolutional neural networks. In CVPR workshop.

Part 2. Review of WSOL methods: Literatur

Confidence-based aggregation
- Score-CAM ('20)
- SS-CAM ('20)
- IS-CAM ('20)
- Ablation-CAM ('20)

Taxonomy: Top-Down
Confidence-based aggregation

Are these methods practical?

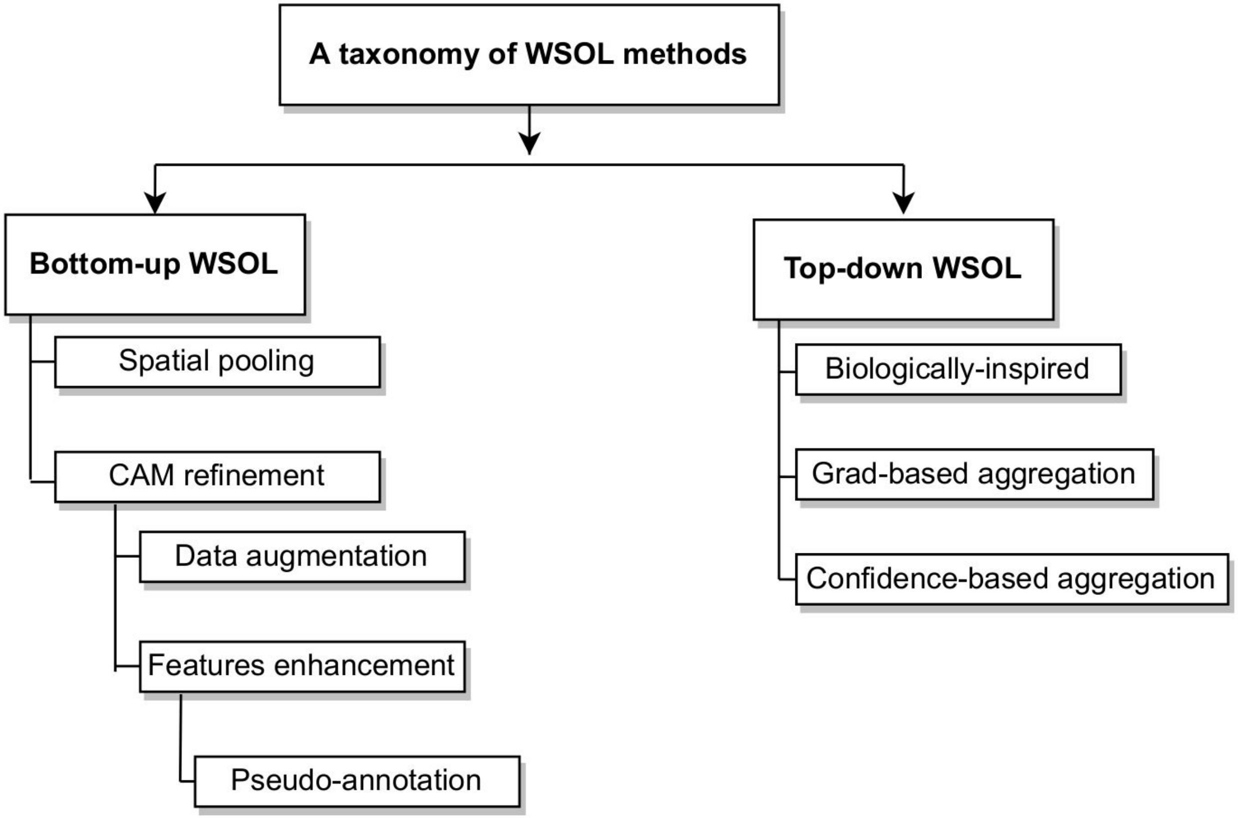
Backbones (encoders) / Methods	VGG16 #PCL	VGG16 #NFM	VGG16 SFM	VGG16 #PDEC	Inception #PCL	Inception #NFM	Inception SFM	Inception #PDEC	ResNet50 #PCL	ResNet50 #NFM	ResNet50 SFM	ResNet50 #PDEC
Details	≈19.6	1024	28x28	≈23.1	≈25.6	1024	28x28	≈5.7	≈23.9	2048	28x28	≈9
CAM* [58]	.2ms				.2ms				.3ms			
GradCAM [32]	7.7ms				21.1ms				27.8ms			
GradCAM++ [7]	23.5ms				23.7ms				28.0ms			
Smooth-GradCAM [25]	62.0ms				150.7ms				136.2ms			
XGradCAM [12]	2.9ms				19.2ms				14.2ms			
LayerCAM [15]	3.2ms				18.2ms				17.9ms			
Mean	16.6ms				38.8ms				37.4ms			
ours + STDCL	6.2ms				25.5ms				18.5ms			
ACoL [55]	12.0ms				19.2ms				24.9ms			
SPG [56]	11.0ms				18ms				23.9ms			
ADL [9]	6.4ms				16.0				14.4ms			
ScoreCAM [44]	1.9sec				3.4sec				9.3sec			
SSCAM [24]	1min45sec				2min16sec				5min49sec			
IS-CAM [23]	30.1sec				39.0sec				1min39sec			

Time necessary to build a full size CAM over an idle **Tesla P100 GPU** for one random RGB image of size **224 × 224** with 200 classes. Methods SSCAM [24] (N = 35, σ = 2), IS-CAM [23] (N = 10), IS-CAM [23] (N = 10) are evaluated with batch size 32 with their original hyper-parameters (N, and σ).

Belharbi, S., Sarraf, A., Pedersoli, M., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022c). F-cam: Full resolution class activation maps via guided parametric upscaling. In WACV.

Part 2. Review of WSOL methods: Literature

Taxonomy



Completed.

Part 2. Review of WSOL methods: Literature

Taxonomy: Bottom-up
NON-CAM methods

CSTN: Convolutional **STN**

Are there NON-CAM WSOL methods?

Meethal, A., Pedersoli, M., Belharbi, S., and Granger, E. (2020). Convolutional stn for weakly supervised object localization and beyond. In ICPR.

Le génie pour l'industrie

Part 2. Review of WSOL methods: Literature

Taxonomy: Bottom-up
NON-CAM methods

CSTN: Convolutional **STN**

Are there NON-CAM WSOL methods?

Yes, but very limited.
why?

Meethal, A., Pedersoli, M., Belharbi, S., and Granger, E. (2020). Convolutional stn for weakly supervised object localization and beyond. In ICPR.

ÉTS
Le génie pour l'industrie

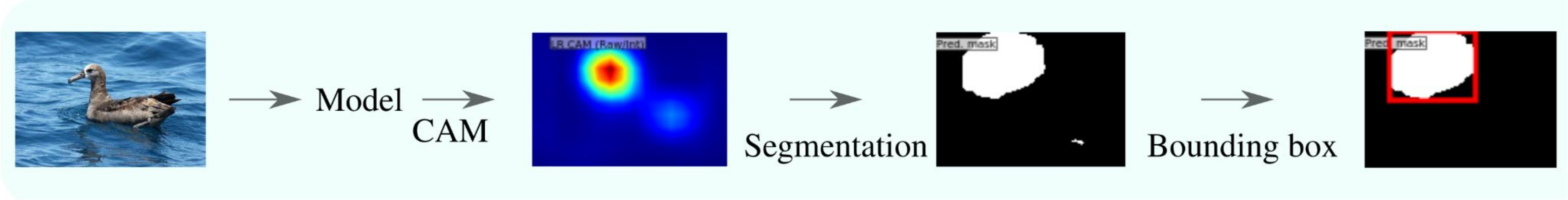
Part 2. Review of WSOL methods: Literature

Taxonomy: Bottom-up
NON-CAM methods

CSTN: Convolutional **STN**

Are there NON-CAM WSOL methods?

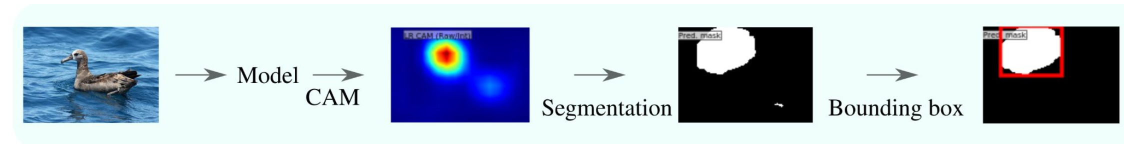
Yes, but very limited.
why?



Meethal, A., Pedersoli, M., Belharbi, S., and Granger, E. (2020). Convolutional stn for weakly supervised object localization and beyond. In ICPR.

Part 2. Review of WSOL methods: Literature

Taxonomy: Bottom-up NON-CAM methods



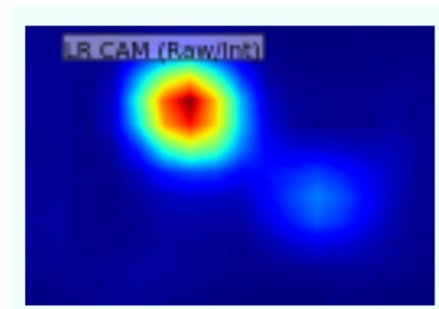
CSTN: Convolutional **STN**

Are there NON-CAM WSOL methods?

Yes, but very limited.
why?

1: Model

Input, image class



CAM
Soft-segmentation
(**Dense** localization)

2: Image post-processing (non-differentiable)

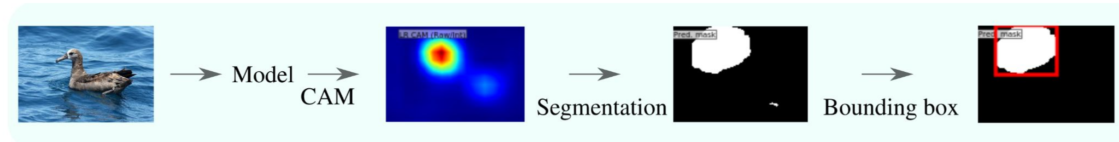
Bounding box localization
(**Coarse** localization)

Le génie pour l'industrie

Meethal, A., Pedersoli, M., Belharbi, S., and Granger, E. (2020). Convolutional stn for weakly supervised object localization and beyond. In ICPR.

Part 2. Review of WSOL methods: Literature

Taxonomy: Bottom-up
NON-CAM methods

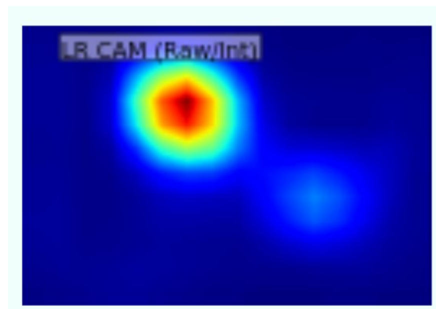


CSTN: Convolutional STN

Are there NON-CAM WSOL methods?

Yes, but very limited.
why?

1: Model



CAM
Soft-segmentation
(**Dense** localization)

2: Image post-processing (non-differentiable)

Input, image class

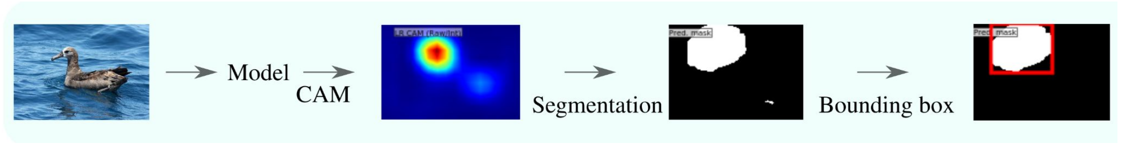
Why not?

Bounding box localization
(**Coarse** localization)

Meethal, A., Pedersoli, M., Belharbi, S., and Granger, E. (2020). Convolutional stn for weakly supervised object localization and beyond. In ICPR.

Part 2. Review of WSOL methods: Literature

Taxonomy: Bottom-up NON-CAM methods

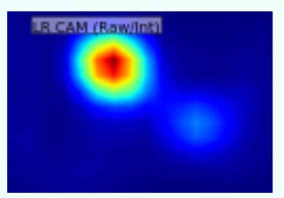


CSTN: Convolutional **STN**

Are there NON-CAM WSOL methods?

**Yes, but very limited.
why?**

1: Model



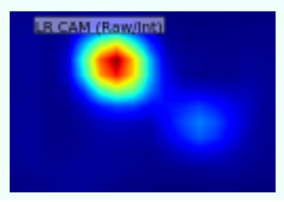
2: Image post-processing (non-differentiable)

CAM
Soft-segmentation
(**Dense** localization)

Input, image class

Why not?

Bounding box localization
(**Coarse** localization)

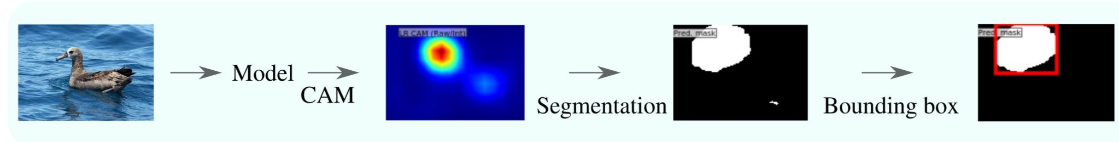


- **CAM**: emerging property from convolution over visible pixels in image [--> easy]

Meethal, A., Pedersoli, M., Belharbi, S., and Granger, E. (2020). Convolutional stn for weakly supervised object localization and beyond. In ICPR.

Part 2. Review of WSOL methods: Literature

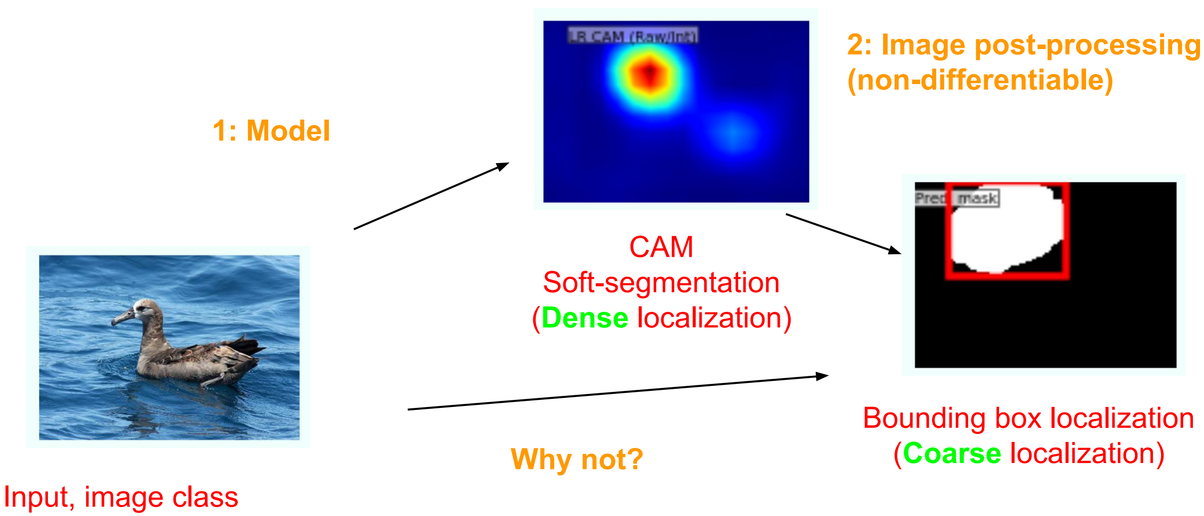
Taxonomy: Bottom-up NON-CAM methods

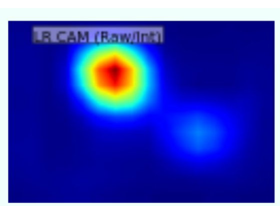


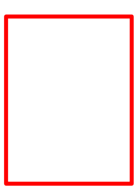
CSTN: Convolutional STN

Are there NON-CAM WSOL methods?

Yes, but very limited.
why?







- **CAM**: emerging property from convolution over visible pixels in image [--> easy]
- **Bounding box**: abstract concept (invisible in image) [--> more difficult]

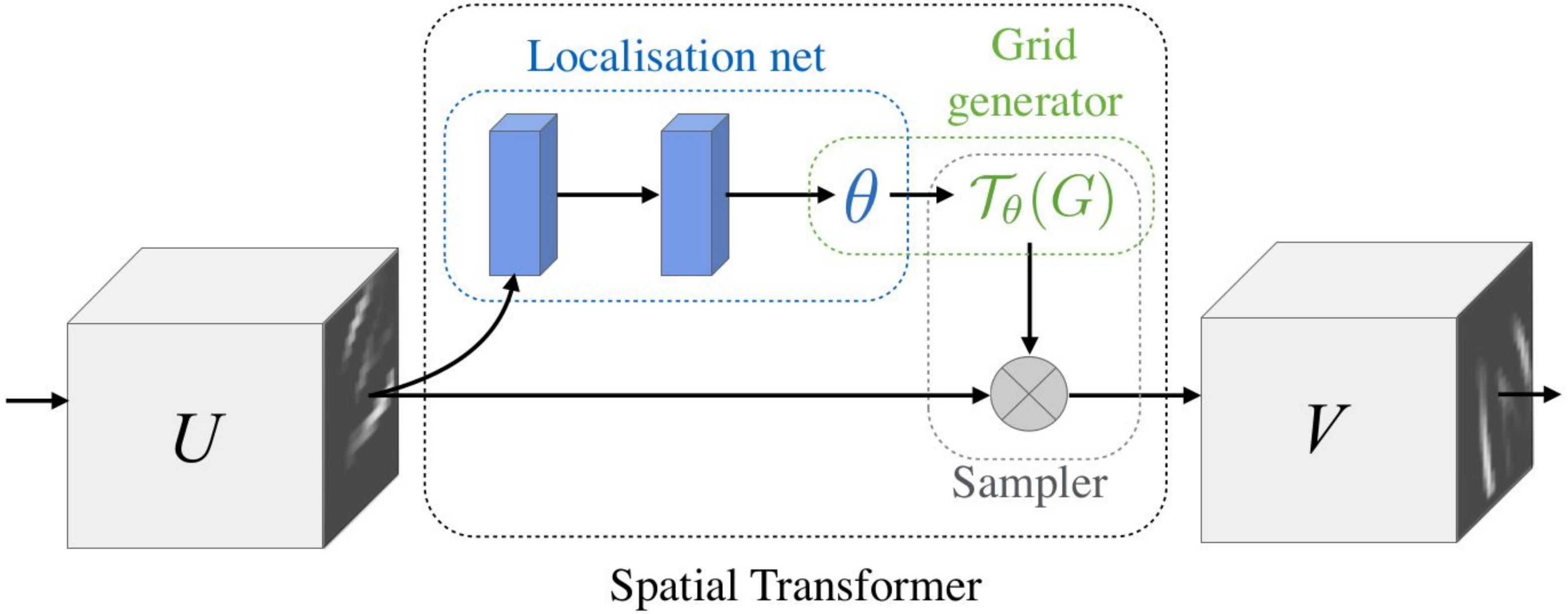
Meethal, A., Pedersoli, M., Belharbi, S., and Granger, E. (2020). Convolutional stn for weakly supervised object localization and beyond. In ICPR.

Part 2. Review of WSOL methods: Literature

Taxonomy: Bottom-up
NON-CAM methods

Spatial **T**ransformer **N**etworks (**STN**)
M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, “Spatial transformer networks,” in eurIPS, 2015.

CSTN: Convolutional **STN**



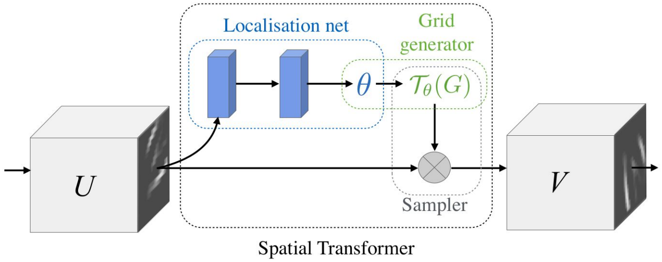
Meethal, A., Pedersoli, M., Belharbi, S., and Granger, E. (2020). Convolutional stn for weakly supervised object localization and beyond. In ICPR.

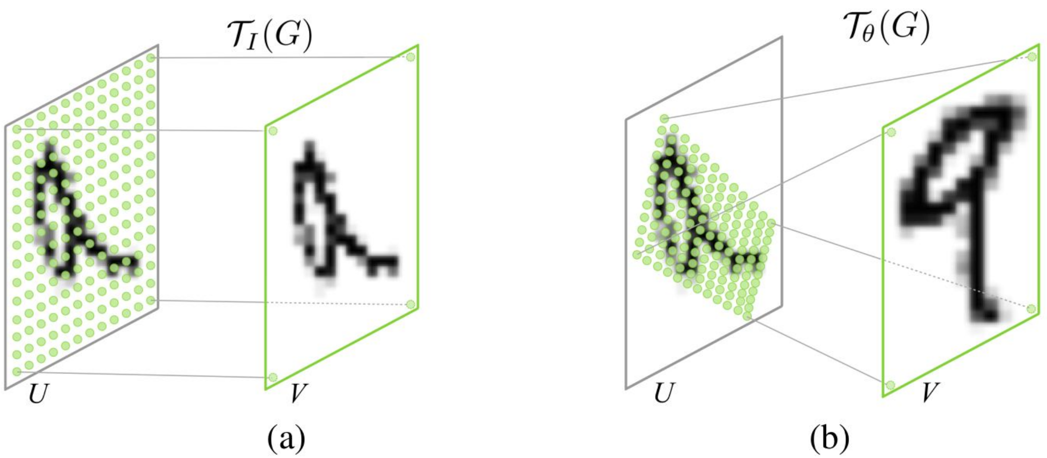
Part 2. Review of WSOL methods: Literature

Taxonomy: Bottom-up NON-CAM methods

Spatial **T**ransformer **N**etworks (**STN**)
M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, “Spatial transformer networks,” in eurIPS, 2015.

CSTN: Convolutional **STN**





Differentiable ‘cropping’

Model/Invariant to affine transformations: translation, scale, rotation, …

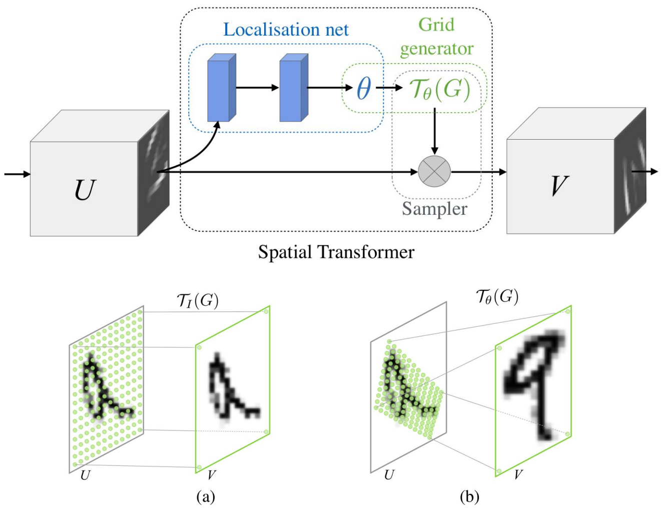
Part 2. Review of WSOL methods: Literature

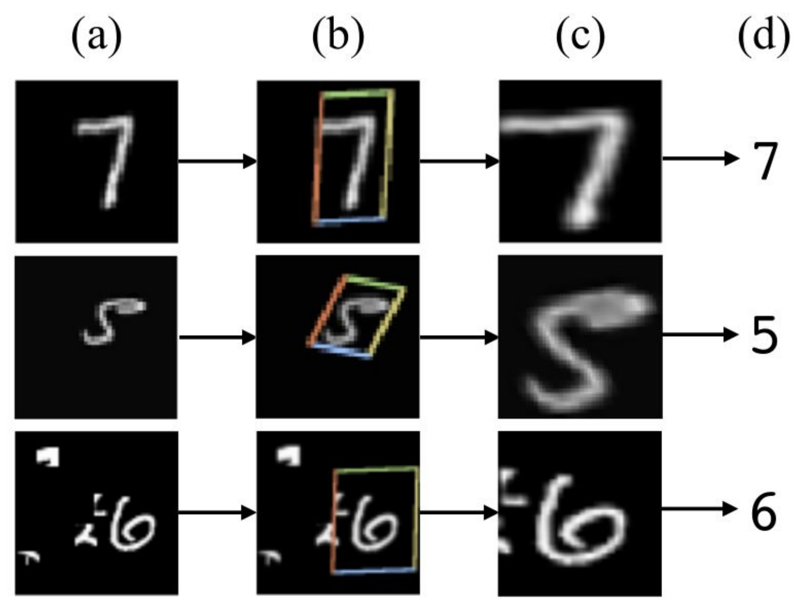
Taxonomy: Bottom-up NON-CAM methods

Spatial **T**ransformer **N**etworks (**STN**)

M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transformer networks," in eurIPS, 2015.

CSTN: Convolutional **STN**





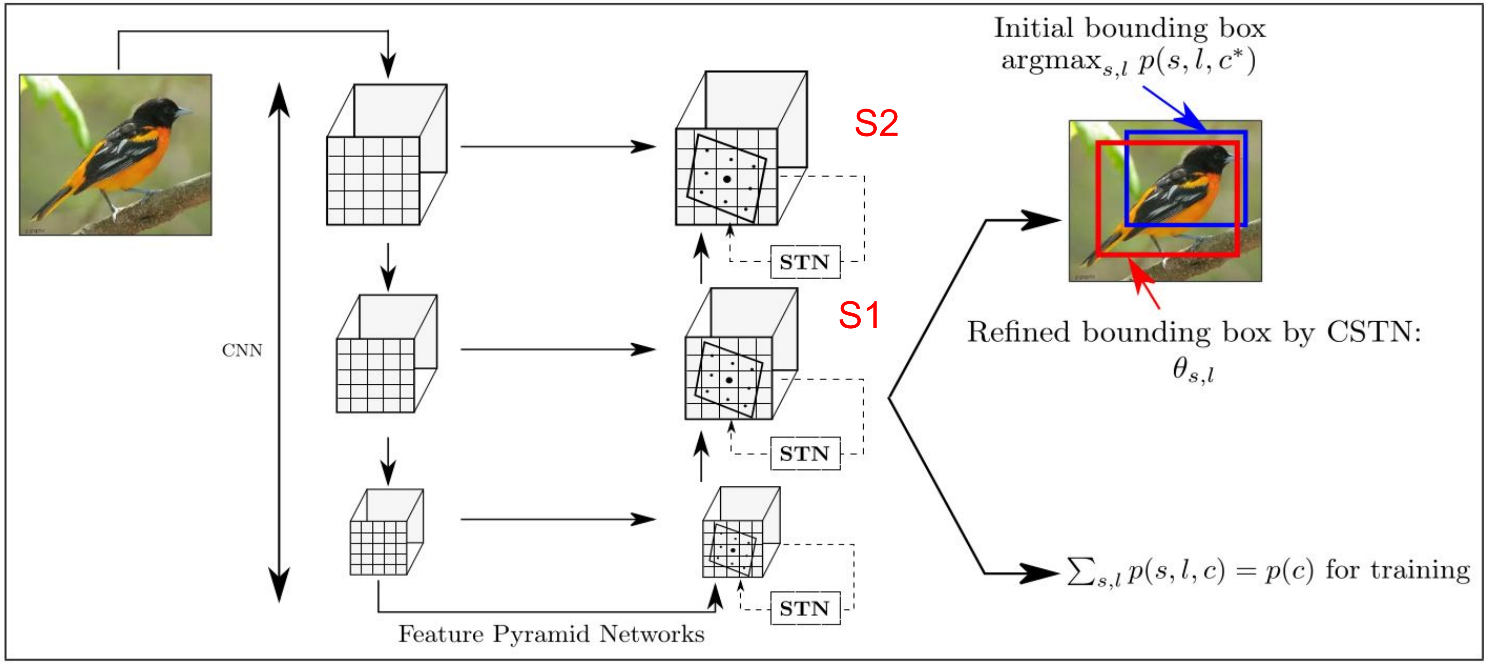
Localization

Meethal, A., Pedersoli, M., Belharbi, S., and Granger, E. (2020). Convolutional stn for weakly supervised object localization and beyond. In ICPR.

Part 2. Review of WSOL methods: Literature

Taxonomy: Bottom-up
NON-CAM methods

CSTN: Convolutional **STN**



Meethal, A., Pedersoli, M., Belharbi, S., and Granger, E. (2020). Convolutional stn for weakly supervised object localization and beyond. In ICPR.

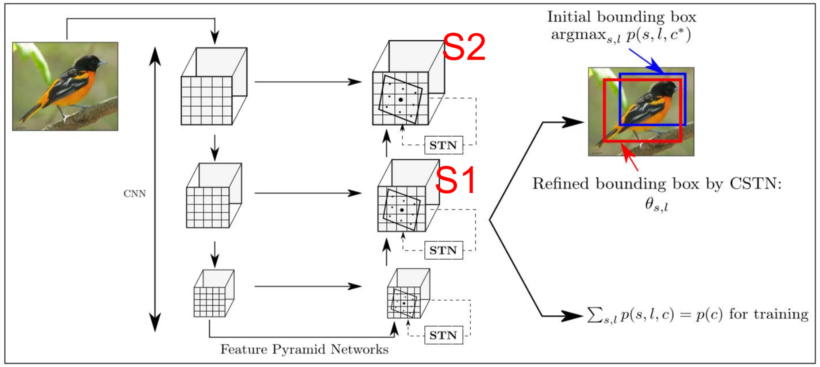
Part 2. Review of WSOL methods: Literature

Taxonomy: Bottom-up
NON-CAM methods

Train loss

CSTN: Convolutional STN

$$L(x,y) = L_{cls}(x,y) + \lambda L_\theta + \alpha L_{scale}(x)$$



Meethal, A., Pedersoli, M., Belharbi, S., and Granger, E. (2020). Convolutional stn for weakly supervised object localization and beyond. In ICPR.

Le génie pour l'industrie

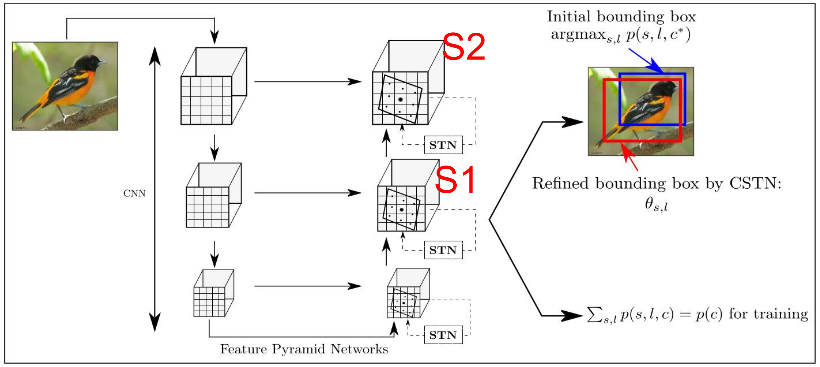
Part 2. Review of WSOL methods: Literature

Taxonomy: Bottom-up
NON-CAM methods

Train loss

CSTN: Convolutional **STN**

$$L(x,y) = \underbrace{L_{cls}(x,y)}_{\text{Standard classification loss}} + \lambda L_\theta + \alpha L_{scale}(x)$$

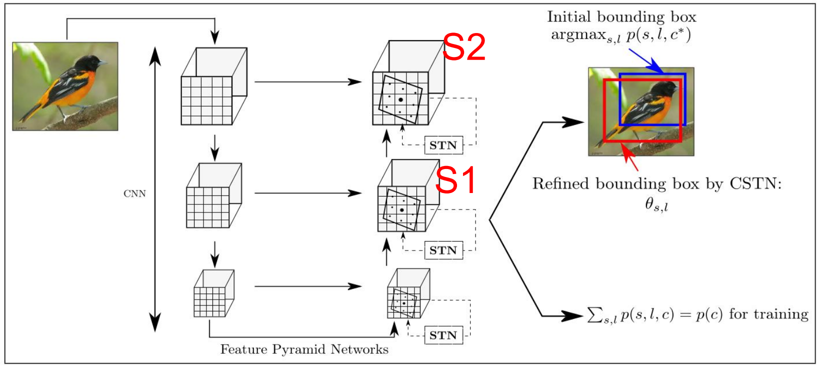


Meethal, A., Pedersoli, M., Belharbi, S., and Granger, E. (2020). Convolutional stn for weakly supervised object localization and beyond. In ICPR.

Part 2. Review of WSOL methods: Literature

Taxonomy: Bottom-up
NON-CAM methods

CSTN: Convolutional STN



Train loss

$$L(x,y) = L_{cls}(x,y) + \lambda L_{\theta} + \alpha L_{scale}(x)$$

Deal with degenerate transformations

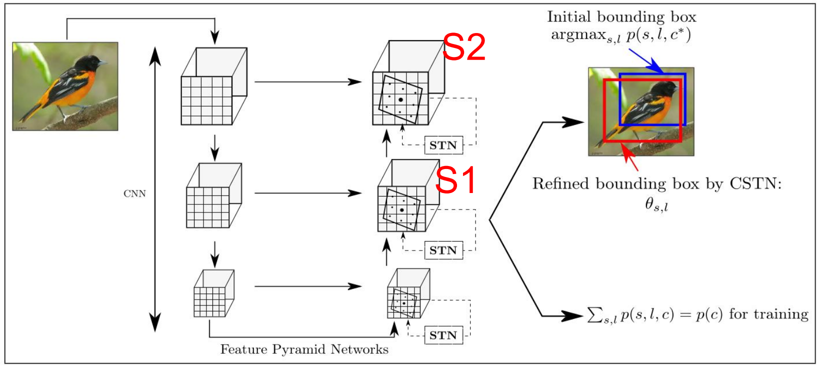
$$L_{\theta} = \sum_{s \in S} \sum_{i=1}^{h_s \times w_s} ||\theta_{ref} - \theta_i||^2 \ .$$

Reference transformation (identity)

Part 2. Review of WSOL methods: Literature

Taxonomy: Bottom-up
NON-CAM methods

CSTN: Convolutional STN



Train loss

$$L(x,y) = L_{cls}(x,y) + \lambda L_\theta + \alpha L_{scale}(x)$$

Deal with scale issue:
Large objects are selected at low level layers (small part only) → allow top layer to select full object

$$L_{scale}(x) = \max\left(0, \max_l p(s = s_1, l, c = c^*|x) - \max_l (p(s = s_2, l, c = c^*|x)\right)$$

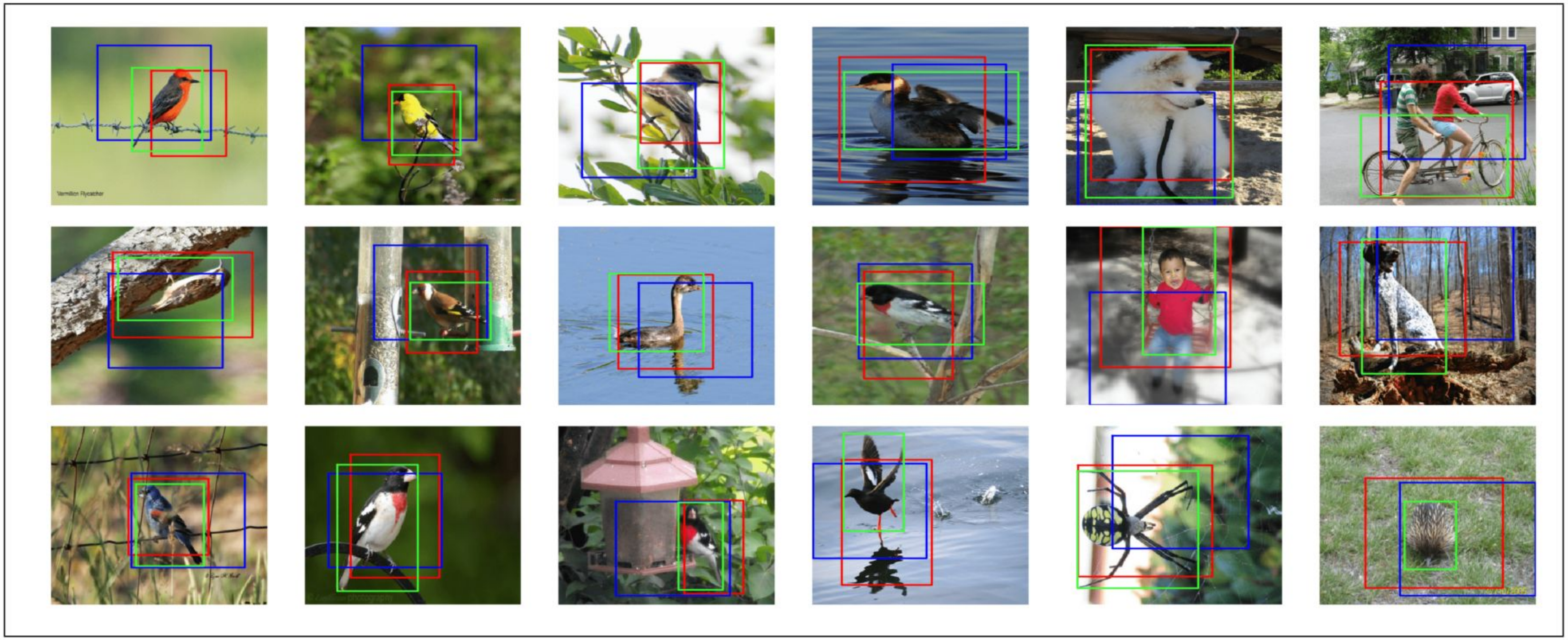
Meethal, A., Pedersoli, M., Belharbi, S., and Granger, E. (2020). Convolutional stn for weakly supervised object localization and beyond. In ICPR.

Part 2. Review of WSOL methods: Literature

Taxonomy: Bottom-up
NON-CAM methods

Green: ground truth
Blue: without STN
Red: with STN

CSTN: Convolutional **STN**



Le génie pour l'industrie

Meethal, A., Pedersoli, M., Belharbi, S., and Granger, E. (2020). Convolutional stn for weakly supervised object localization and beyond. In ICPR.

Part 2

Review of WSOL Methods

- **WSOL literature: bottom-up and top-down methods**
- **Case studies:**
  a) F-CAM for improved interpolation
  b) Transformer-based models

Case Study (a): F-CAM for Improved Interpolation

F-CAM: Full Resolution Class Activation Maps via Guided Parametric Upscaling

Soufiane Belharbi, Aydin Sarraf, Marco Pedersoli, Ismail Ben Ayed, Luke McCaffrey, Eric Granger

WACV 2022: Winter Conf. on Applications of Computer Vision

EMPOWER YOURSELF ML / AI ERICSSON Montreal

McGill

LIVIA LABORATOIRE D'IMAGERIE, DE VISION ET D'INTELLIGENCE ARTIFICIELLE

ÉTS Le génie pour l'industrie

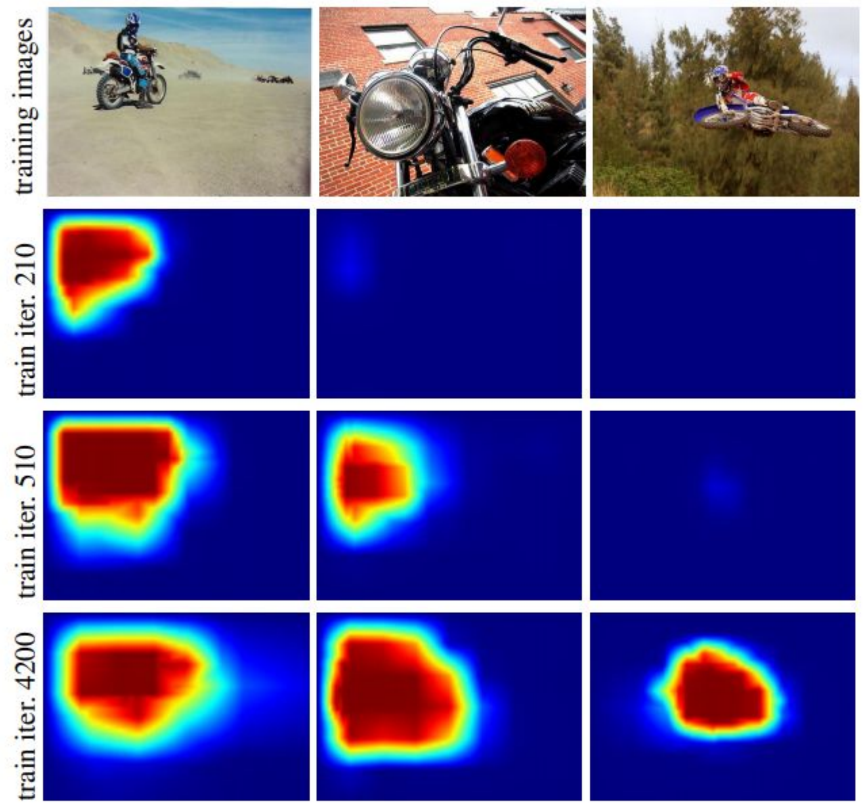
Case Study (a): F-CAM for Improved Interpolation

- **A Challenge with CAMs:** low resolution (due to convolution and pooling) has negative impact on localization performance

Standard interpolated from CAM of 8x8 resolution (downscale factor of 32)

(a) Input (b) ResNet-18

Standard interpolated CAMs

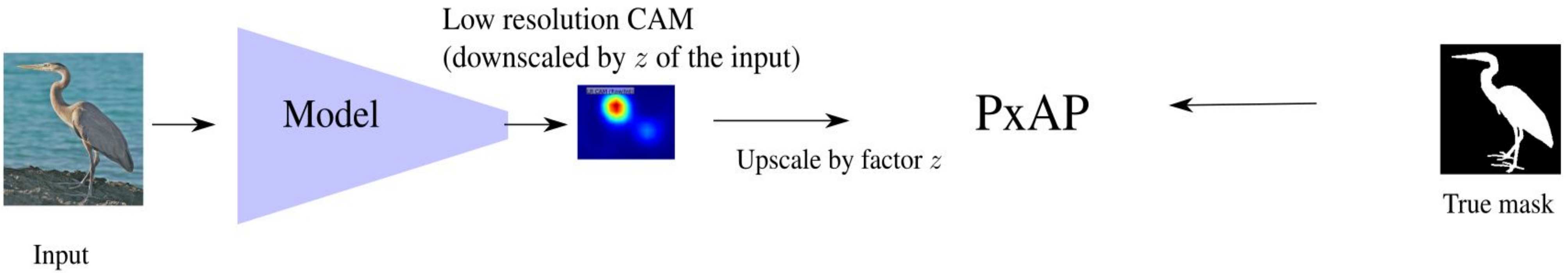


Source: F. Yu, V. Koltun, and T. Funkhouser, Dilated residual networks, CVPR 2017

Source: Oquab, M., et al., Is object localization for free?-weakly-supervised learning with CNNs. In CVPR 2015

Case Study (a): F-CAM for Improved Interpolation

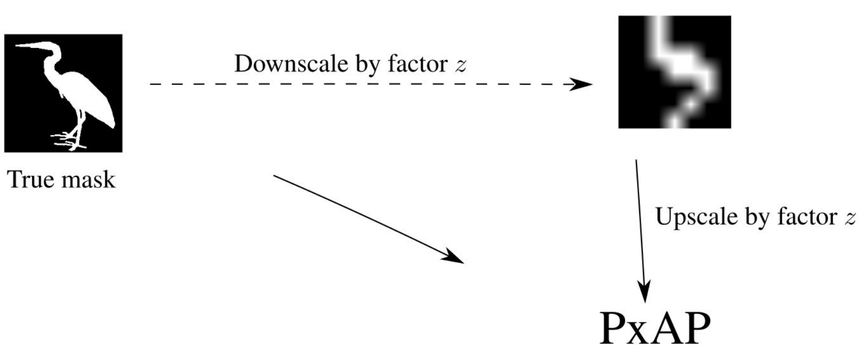
- **Challenges:** Evaluating the impact on localization performance of CAM size (CUB dataset)
  - the CNN produces low-resolution CAMs that are interpolated (by a factor z) to return to the input image size
  - PxAP measures localization accuracy: it evaluates interpolated images against the true segmentation masks



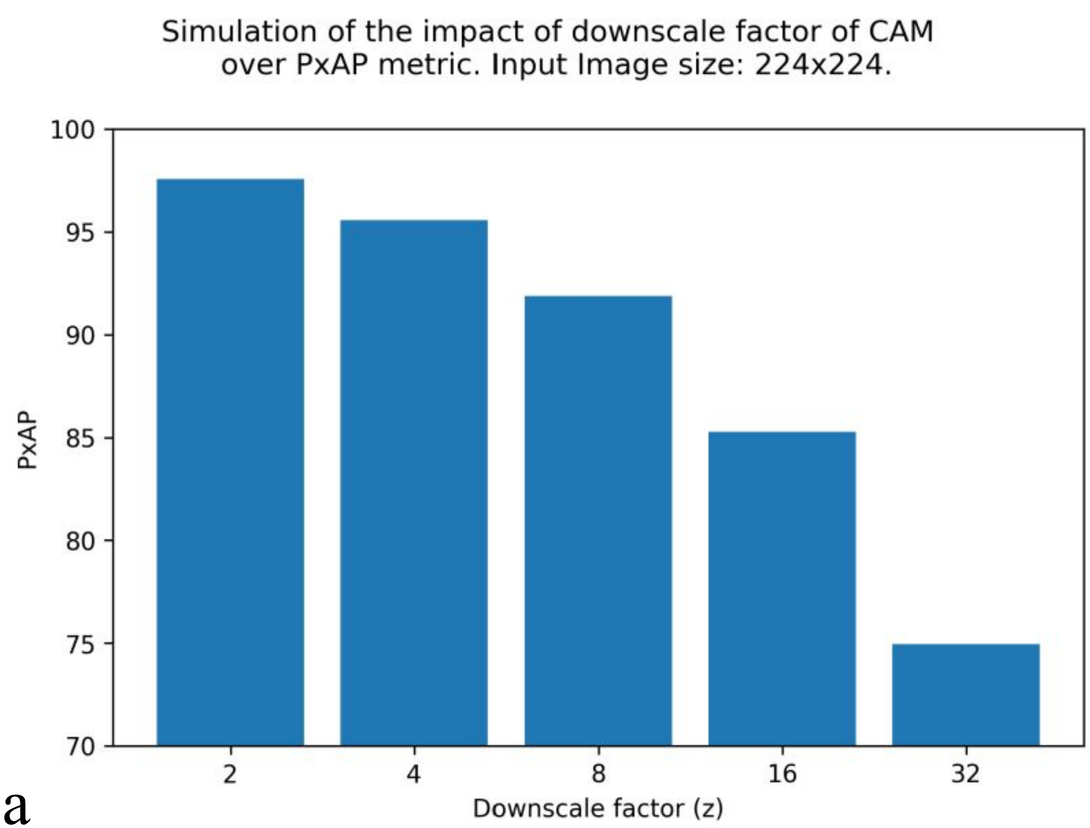
Standard procedure

Case Study (a): F-CAM for Improved Interpolation

- **Challenges:** Impact of CAMs size on localization performance



Our simulation



Results: increasing the downscaling factor (z) leads to a considerable decline in localization accuracy

ÉTS Le génie pour l'industrie

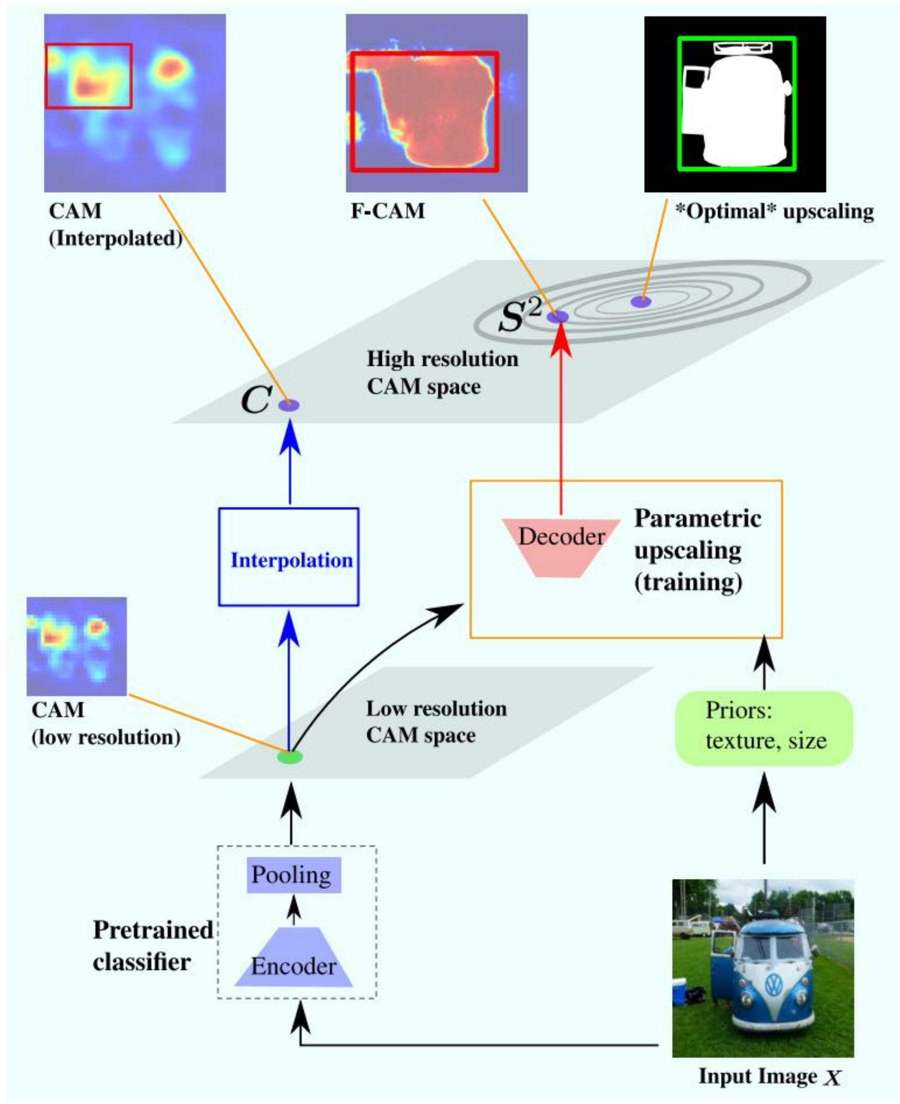
Case Study (a): F-CAM for Improved Interpolation

- **Interpolation:** common, but it does not consider statistical properties of an object, such as color and texture or its shape
- **Litterature:** alternatives to avoid bi-cubic interpolation for producing higher resolution CAMs - *learnable upscaling*
  - residual dilation networks [Yu, CVPR 2017]
  - an end-to-end weakly supervised semantic segmentation approach that upscaled the feature maps [Zhang, AAAI 2020]
  - U-Net architecture to reconstruct the image [Tagaris. ICIP 219]

These methods either produce smaller CAMs, are difficult to scale to large number of classes, or require costly post-processing

Case Study (a): F-CAM for Improved Interpolation

- **Proposed F-CAM with Guided Parametric Upscaling**



Encoder: any pre-trained CNN classifier,

L_c = classification loss (supervised)

Decoder: trained to perform parametric upscaling

L_D = pixel alignment loss (unsupervised)

= SR (CAM) + CRF (image) + ASC (size)

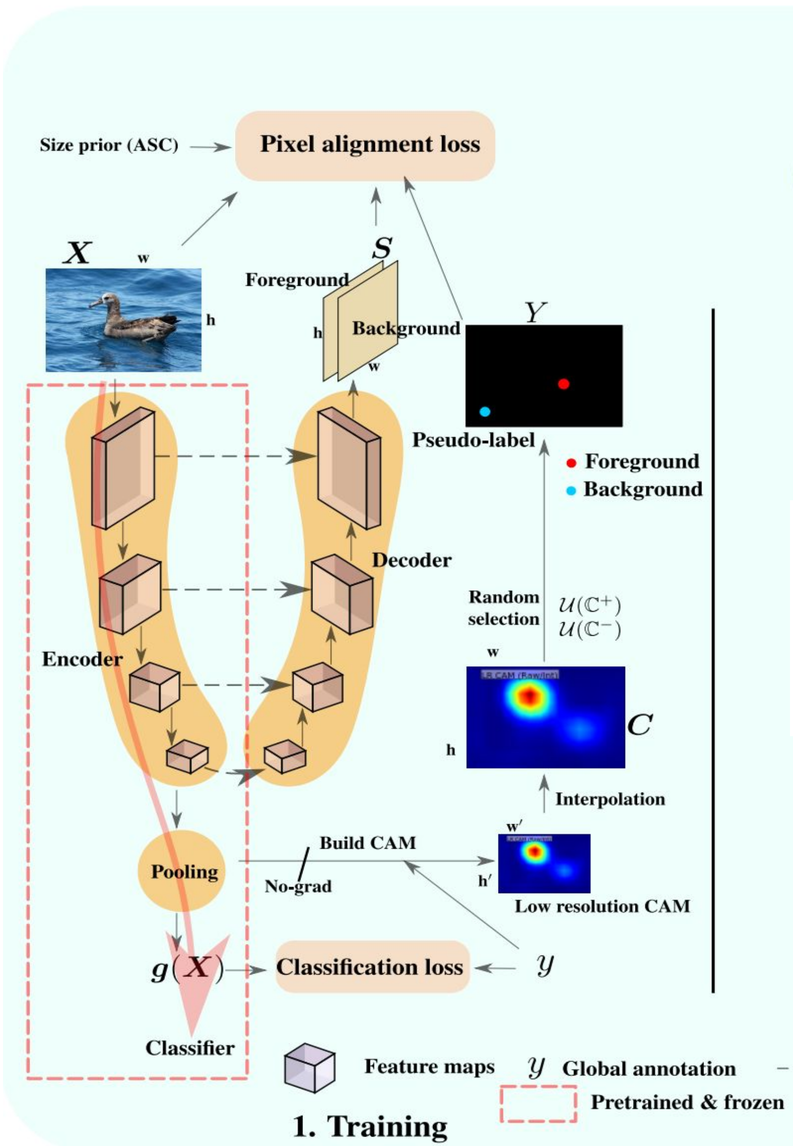
where

- SR: seeds (positive/negative evidence at pixel level)
- CRF: image properties
- ASC: unsupervised size constraint

Le génie pour l'industrie

Case Study (a): F-CAM for Improved Interpolation

- **Proposed F-CAM:** training models the foreground and background



Overall loss for end-to-end training

$$L_c = L_{CE} \qquad L_D = L_{SR} + L_{CRF}$$

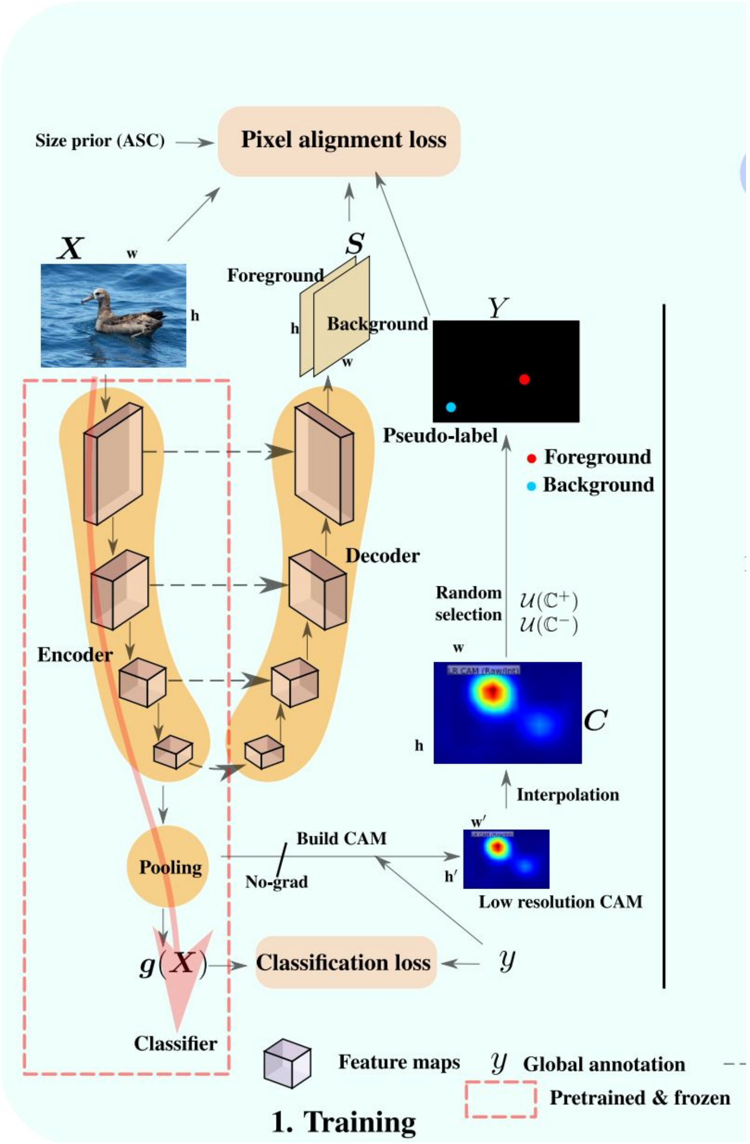
$$\min_{\boldsymbol{\theta}} \quad -\log(\boldsymbol{g}(\boldsymbol{X})[y]) + \alpha \sum_{p \in \Omega'} \boldsymbol{H}(Y_p, \boldsymbol{S}_p) + \lambda\, \mathcal{R}(\boldsymbol{S}, \boldsymbol{X}) \, ,$$

$$\text{s.t.} \quad \sum \boldsymbol{S}^r \geq 0 \, , \quad r \in \{1, 2\} \, ,$$

ASC: area size constraint

Case Study (a): F-CAM for Improved Interpolation

- **Proposed F-CAM:** training models the foreground and background



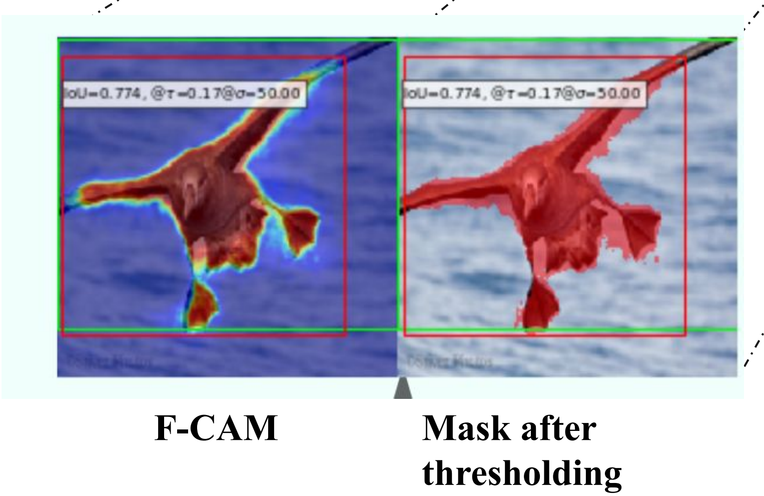
L_{SR} **: Random sampling from *noisy evidence***

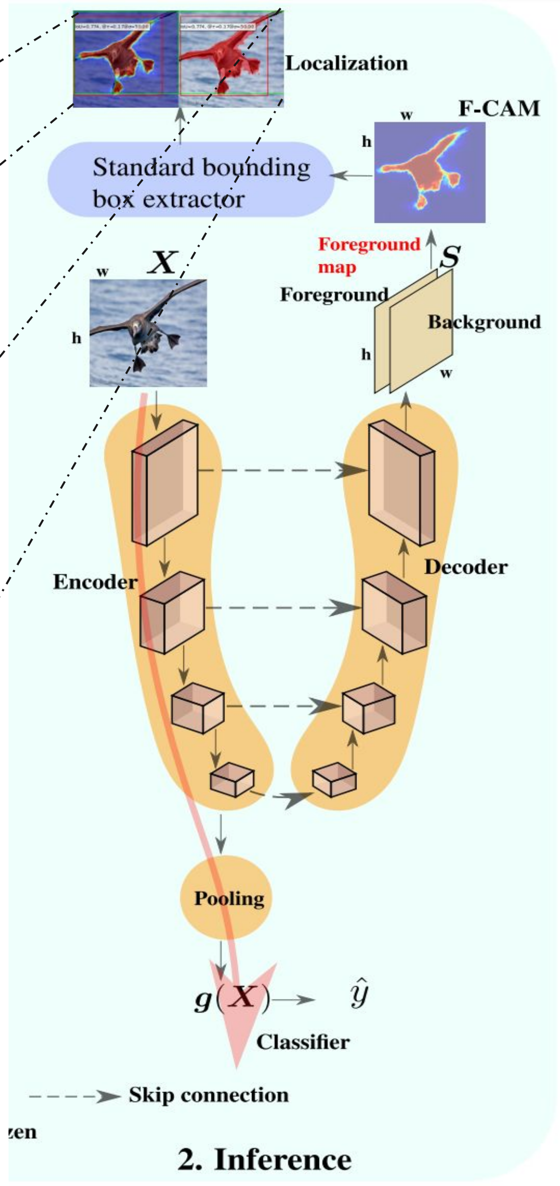
- use stochastic sampling (extreme dropout) to avoid overfitting wrong labels,
- give the model enough time to allow consistent ROI to emerge
- at each SGD step, select 2 new random pixels (foreground and background)

Case Study (a): F-CAM for Improved Interpolation

- **Proposed F-CAM:**

  At inference time





Case Study (a): F-CAM for Improved Interpolation

- **Experiments:**

  Visual results on images from the CUB dataset

VGG16 | Inception v3 | ResNet50

CAM

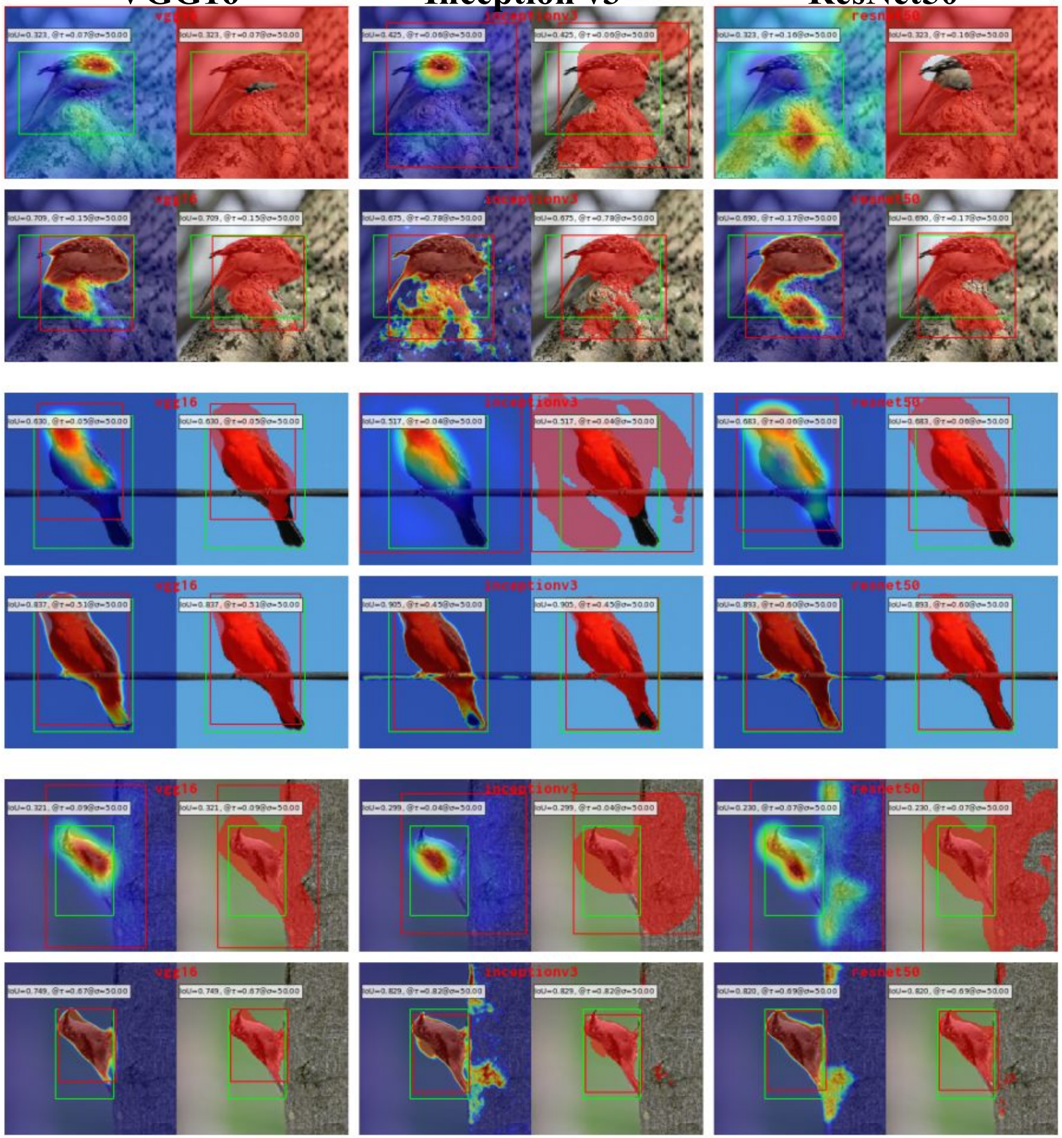
+ F-CAM

Grad CAM

+ F-CAM

Grad CAM++

+ F-CAM



Le génie pour l'industrie

Case Study (a): F-CAM for Improved Interpolation

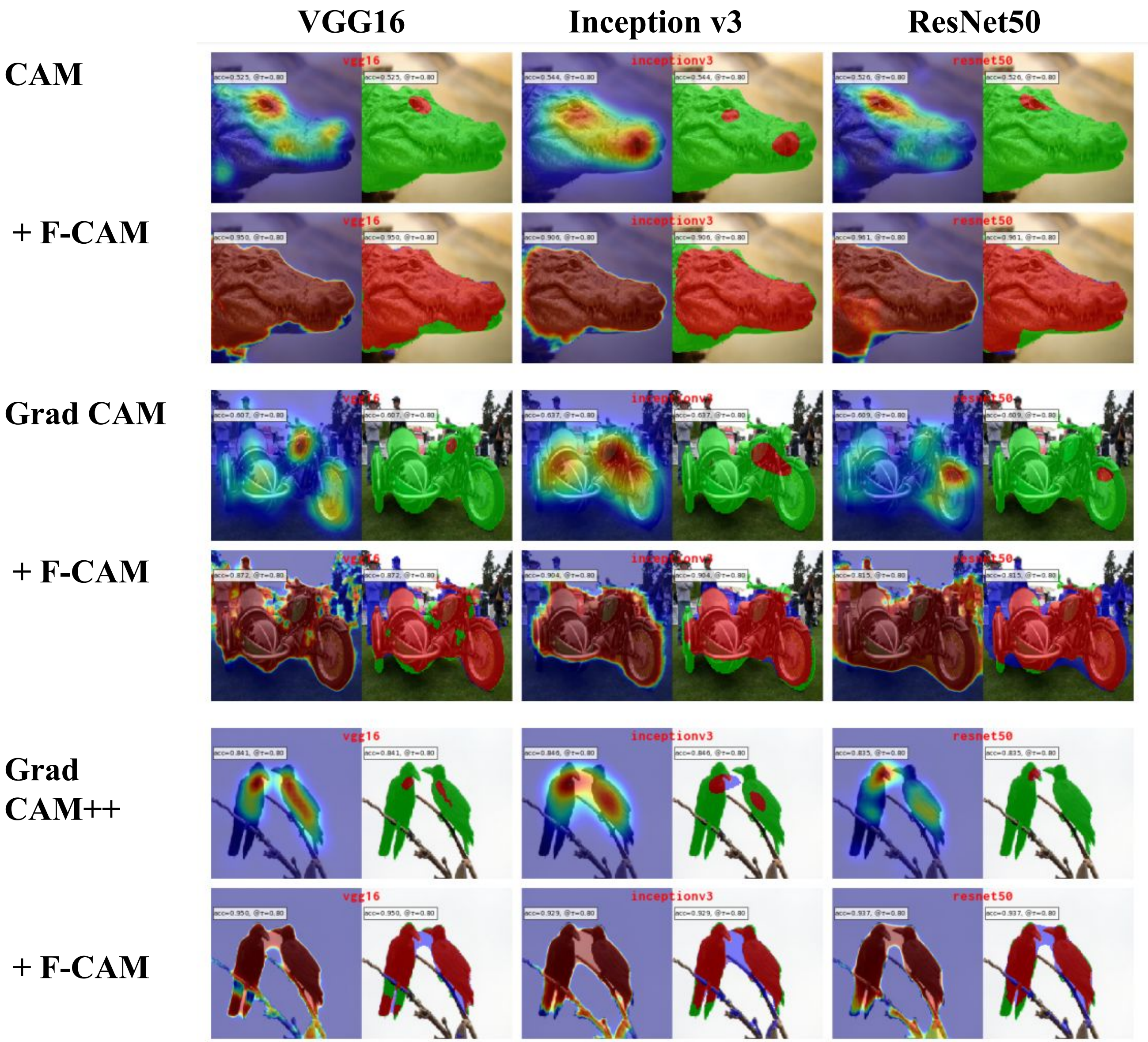
- **Experiments:**

Visual results on images from the OpenImages dataset

Compare CAMs to segmentation masks

Pixel classification

- Red: true positive pixels
- Blue: false positive pixels
- Green: false negative pixels



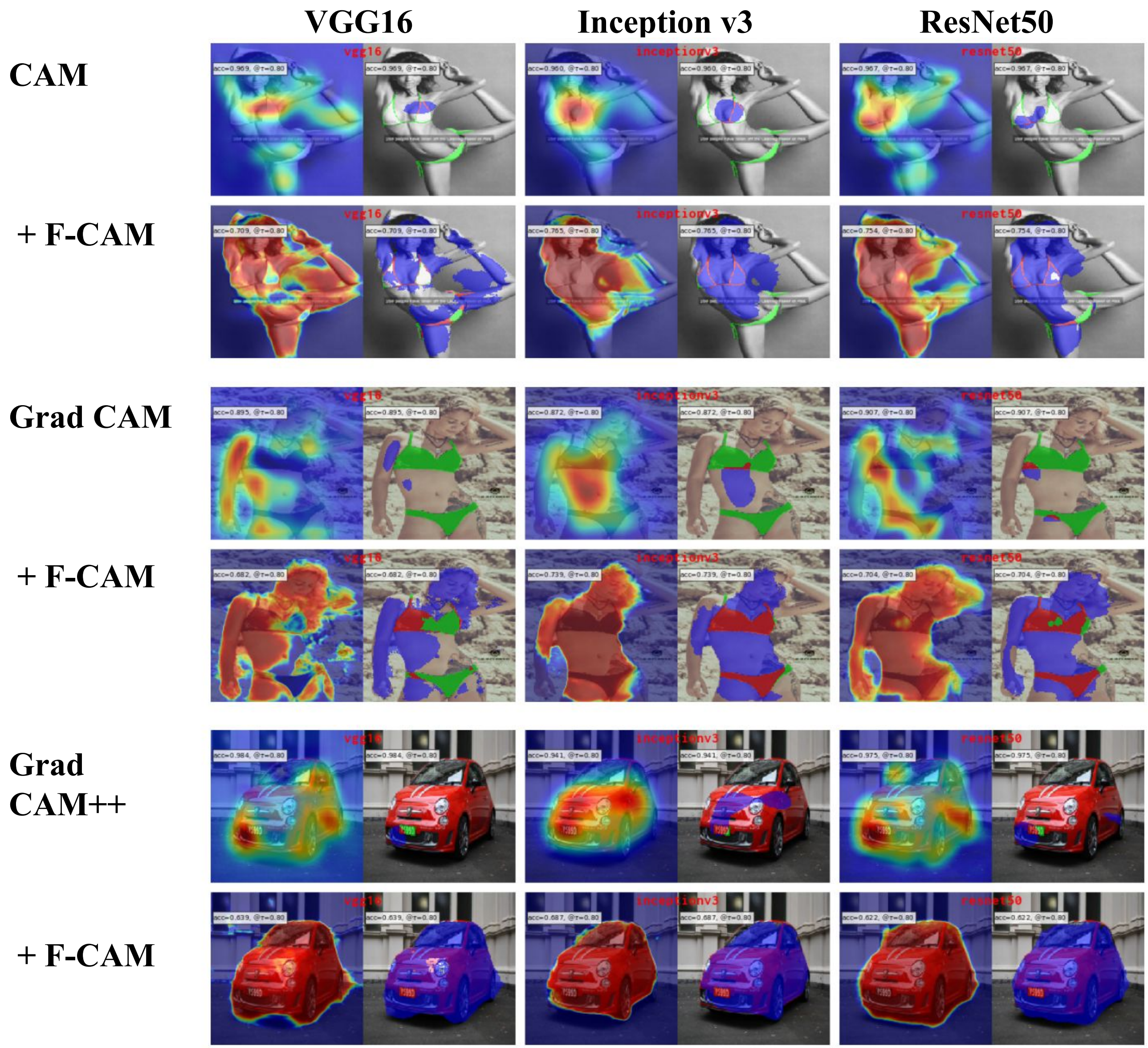
Case Study (a): F-CAM for Improved Interpolation

- **Experiments:**

  Visual results on images from the OpenImages dataset

  Failure cases

  - due to sampling errors
  - performance is tied to quality of CAM from pre-trained CNN classifier



Case Study (a): F-CAM for Improved Interpolation

- **Experiments:**

Localization accuracy on the CUB and OpenImages datasets

3 CNNs

Performance measures:

- MaxBoxAcc (CUB)
- PxAP (OpenImages)

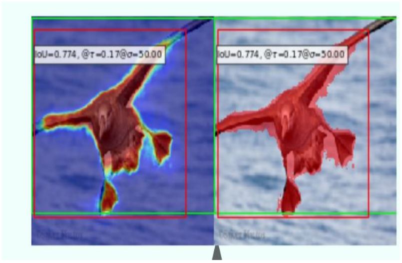
Methods	CUB (MaxBoxAcc)				OpenImages (PxAP)			
	VGG	Inception	ResNet	Mean	VGG	Inception	ResNet	Mean
CAM [57] *(cvpr,2016)*	71.1	62.1	73.2	68.8	58.1	61.4	58.0	59.1
HaS [34] *(iccv,2017)*	76.3	57.7	78.1	70.7	56.9	59.5	58.2	57.8
ACoL [53] *(cvpr,2018)*	72.3	59.6	72.7	68.2	54.7	63.0	57.8	58.4
SPG [54] *(eccv,2018)*	63.7	62.8	71.4	66.0	55.9	62.4	57.7	58.6
ADL [9] *(cvpr,2019)*	75.7	63.4	73.5	70.8	58.3	62.1	54.3	58.2
CutMix [51] *(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 [22] *(icpr,2020)*	Resnet101 [14]: 76.0				–	–	–	–
TS-CAM [13] *(corr,2021)*	Deit-S [39]: 83.8				–	–	–	–
MEIL [21] *(cvpr,2020)*	73.8	–	–	–	–	–	–	–
DANet [47] *(iccv,2019)*	67.7	67.03	–	–	–	–	–	–
SPOL [44] *(cvpr,2021)*	–	–	96.4	–	–	–	–	–
CAM* [57] *(cvpr,2016)*	61.6	58.8	71.5	63.9	53.0	62.7	56.8	57.5
GradCAM [32] *(iccv,2017)*	69.3	62.3	73.1	68.2	59.6	63.9	60.1	61.2
GradCAM++ [7] *(wacv,2018)*	84.1	63.3	81.9	76.4	60.5	64.0	60.2	61.5
Smooth-GradCAM++ [25] *(corr,2019)*	69.7	66.9	76.3	70.9	52.2	61.7	54.3	56.0
XGradCAM [12] *(bmvc,2020)*	69.3	60.9	72.7	67.6	59.0	63.9	60.2	61.0
LayerCAM [15] *(ieee,2021)*	84.3	66.5	85.2	78.6	59.5	63.5	61.1	61.3
CAM* [57] + ours	87.3	82.0	90.3	86.5	67.8	71.9	72.1	70.6
GradCAM [32] + ours	87.5	84.4	90.5	87.4	68.6	70.0	70.9	69.8
GradCAM++ [57] + ours	91.5	84.6	91.0	89.0	64.8	67.1	66.3	66.0
Smooth-GradCAM++ [57] + ours	89.1	86.8	90.7	88.8	60.3	65.4	64.4	63.3
XGradCAM [57] + ours	86.8	84.4	90.4	88.8	68.7	71.3	70.4	70.1
LayerCAM [57] + 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.

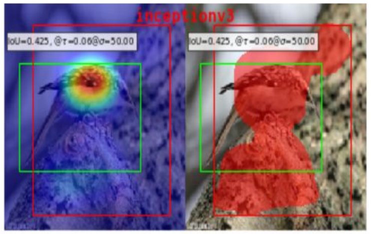
Case Study (a): F-CAM for Improved Interpolation

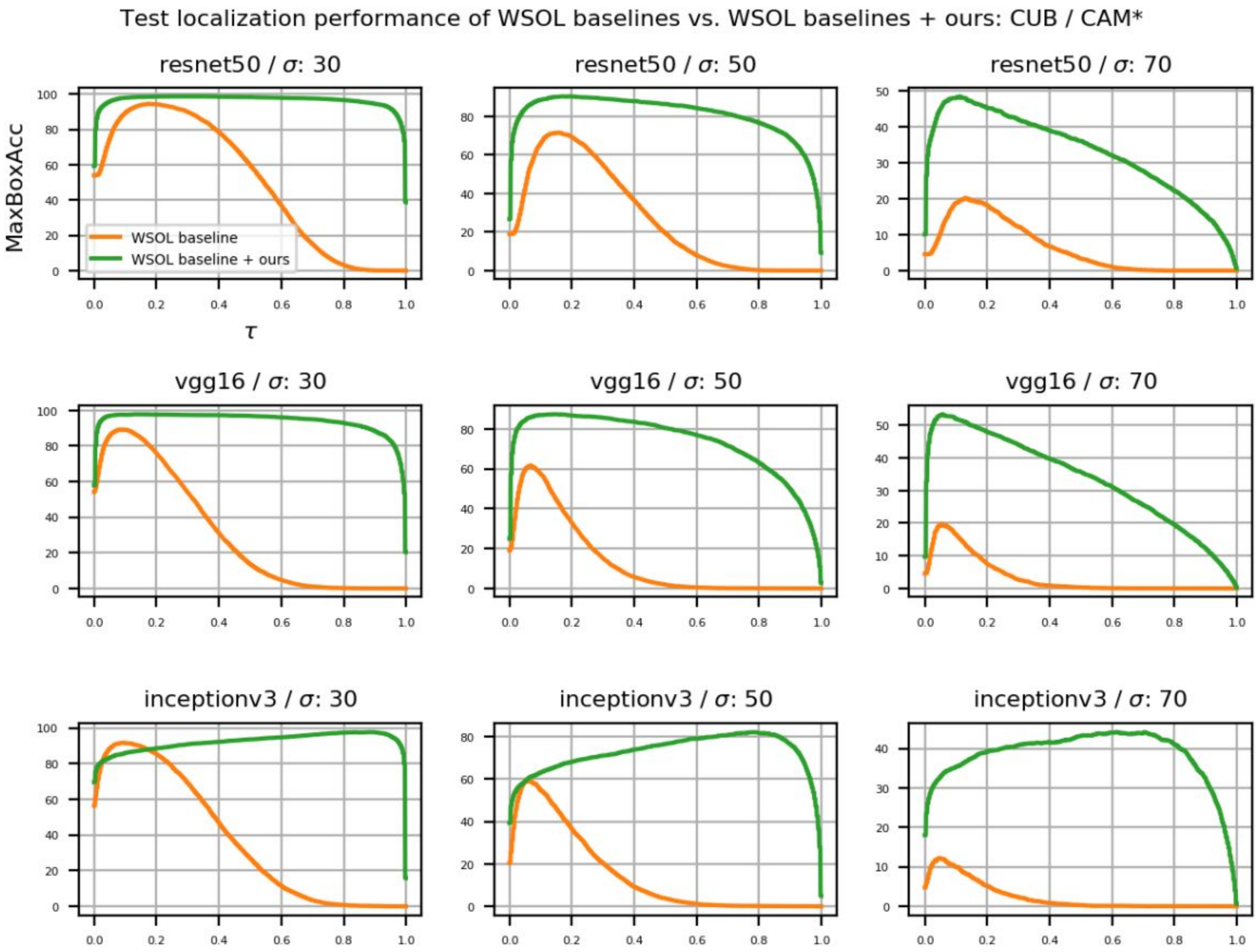
- **Experiments:** sensitivity to threshold values on the CUB dataset

CAM + ours



CAM

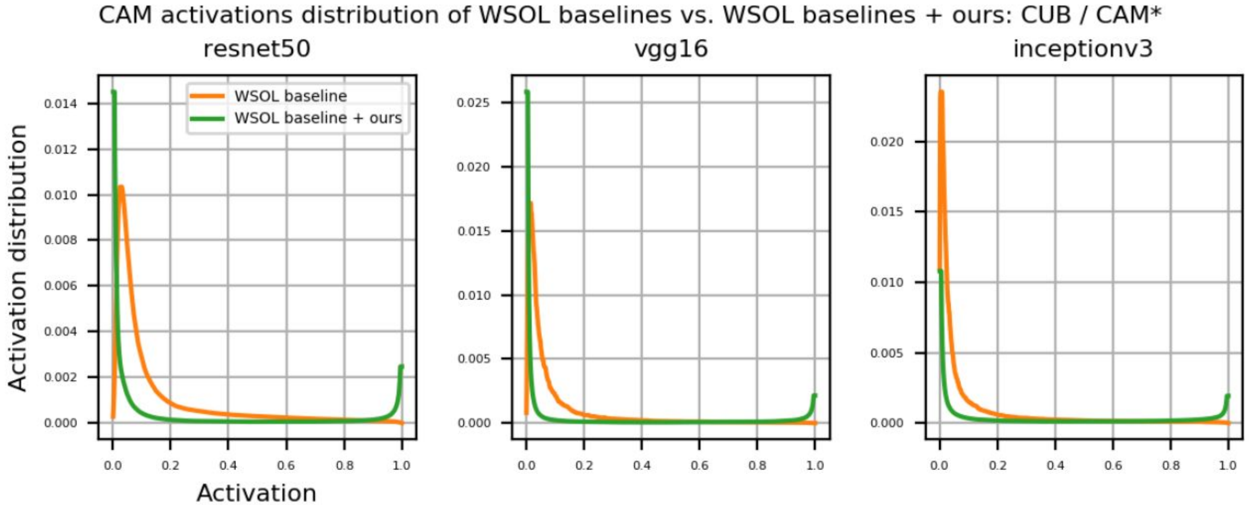


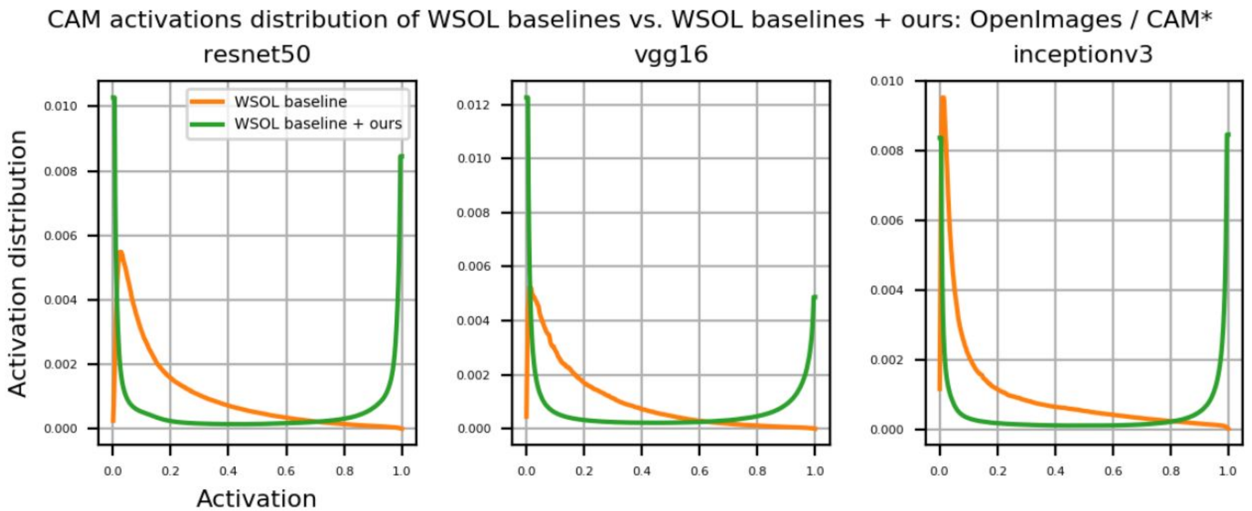


(a) MaxBoxAcc metric.

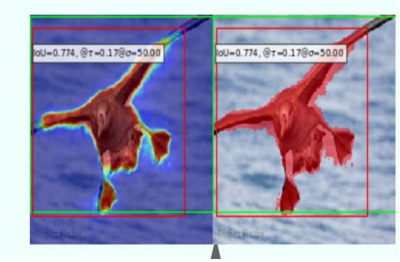
Case Study (a): F-CAM for Improved Interpolation

- **Experiments:** sensitivity to threshold values on the CUB dataset

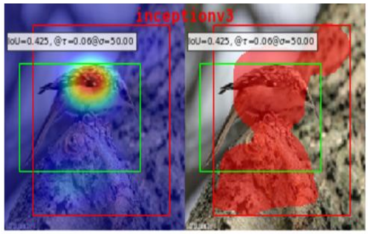




CAM + ours



CAM



Case Study (a): F-CAM for Improved Interpolation

- **Experiments:** Ablation study on the impact on loss components

Methods	CUB (MaxBoxAcc) VGG	Inception	ResNet	Mean	OpenImages (PxAP) VGG	Inception	ResNet	Mean
CAM* [58]	61.6	58.8	71.5	63.9	53.0	62.7	56.8	57.5
CAM* [58] + SR	84.2	73.0	82.2	79.8	64.5	64.1	63.8	64.1
CAM* [58] + SR + ASC	82.9	74.1	83.2	80.0	63.9	63.4	62.0	63.1
CAM* [58] + SR + CRF	84.6	78.9	86.1	83.2	66.3	68.3	67.5	67.3
CAM* [58] + SR + CRF + ASC	87.3	82.0	90.3	86.5	67.8	71.9	72.1	70.6
Improvement	+25.7	+23.2	+18.8	+22.5	+14.8	+9.2	+15.3	+12.8

Le génie pour l'industrie

Case Study (a): F-CAM for Improved Interpolation

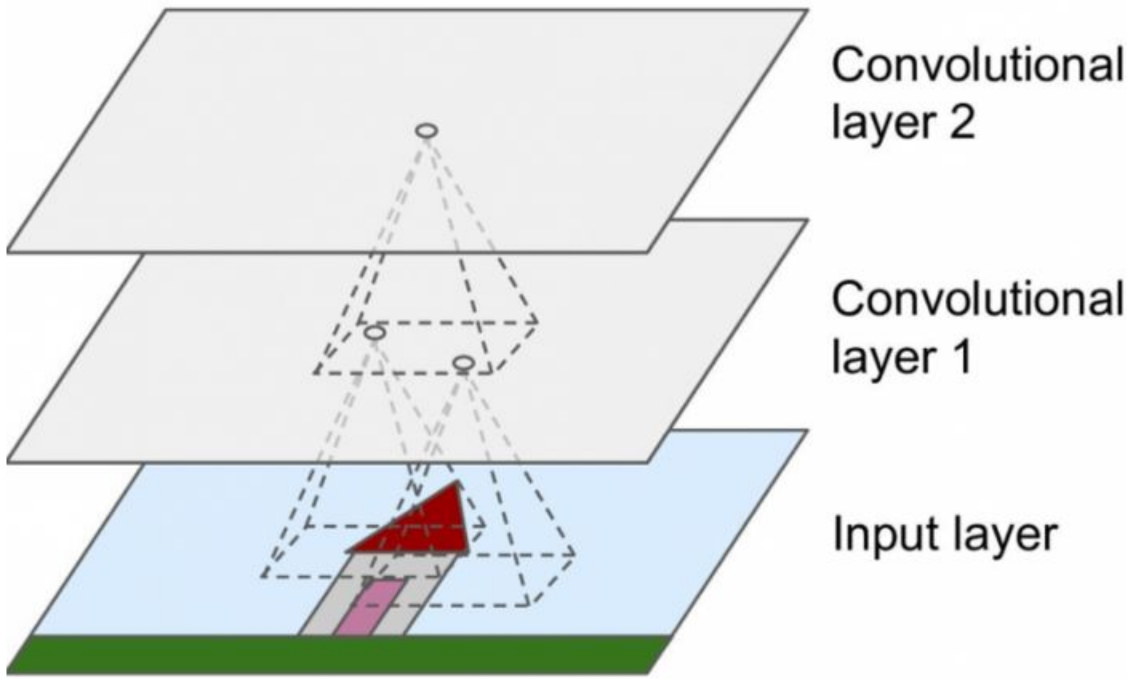
- **Experiments:** Complexity

adding the decoder for upscaling was a competitive runtime during inference

Backbones (encoders)	VGG16				Inception				ResNet50			
Methods	#PCL	#NFM	SFM	#PDEC	#PCL	#NFM	SFM	#PDEC	#PCL	#NFM	SFM	#PDEC
Details	≈19.6	1024	28x28	≈23.1	≈25.6	1024	28x28	≈5.7	≈23.9	2048	28x28	≈9
CAM* [58]	.2ms				.2ms				.3ms			
GradCAM [32]	7.7ms				21.1ms				27.8ms			
GradCAM++ [7]	23.5ms				23.7ms				28.0ms			
Smooth-GradCAM [25]	62.0ms				150.7ms				136.2ms			
XGradCAM [12]	2.9ms				19.2ms				14.2ms			
LayerCAM [15]	3.2ms				18.2ms				17.9ms			
Mean	16.6ms				38.8ms				37.4ms			
ours + STDCL	6.2ms				25.5ms				18.5ms			
ACoL [55]	12.0ms				19.2ms				24.9ms			
SPG [56]	11.0ms				18ms				23.9ms			
ADL [9]	6.4ms				16.0				14.4ms			
ScoreCAM [44]	1.9sec				3.4sec				9.3sec			
SSCAM [24]	1min45sec				2min16sec				5min49sec			
IS-CAM [23]	30.1sec				39.0sec				1min39sec			

Case Study (b): Transformer-based models for WSOL (Ongoing work)

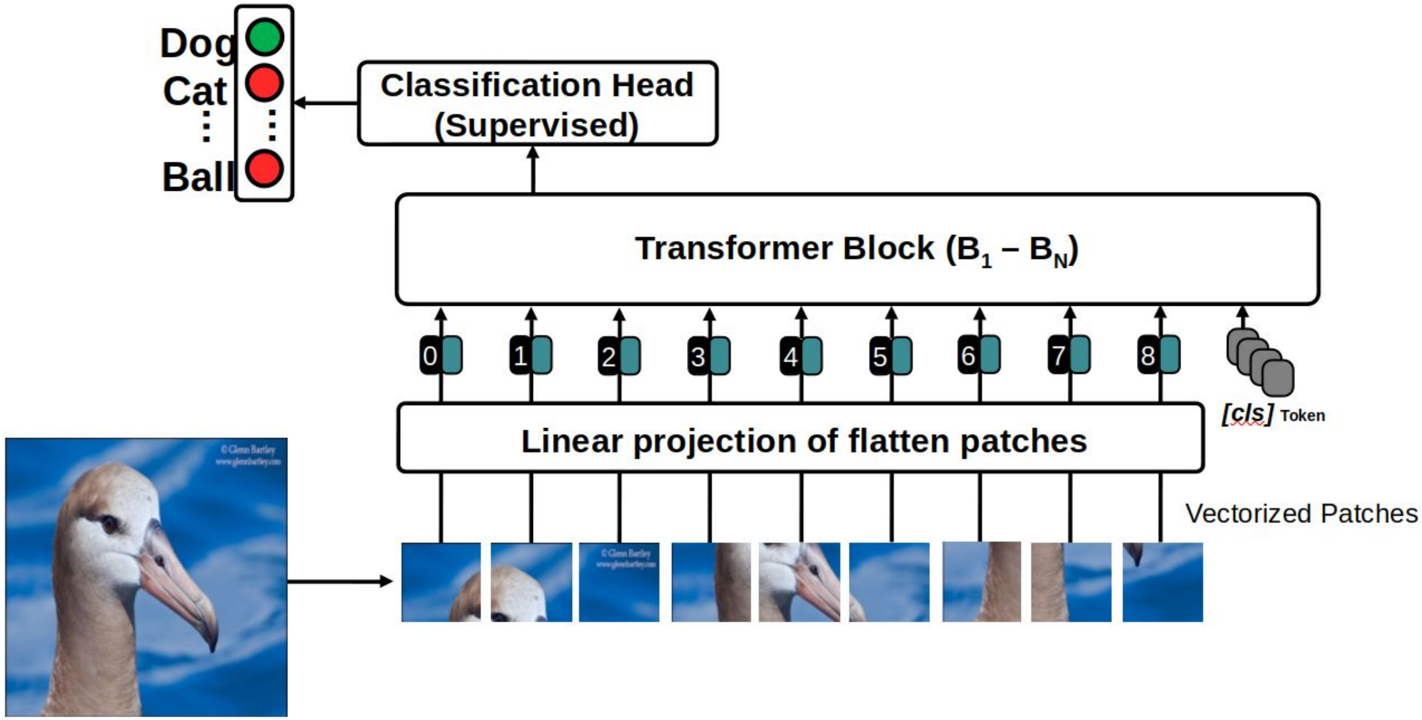
- Small local receptive field of CNN



https://miro.medium.com/max/700/1*j7eTJrId7YIdBYCKNpnWQA.png

Case Study (b): Transformer-based models for WSOL (Ongoing work)

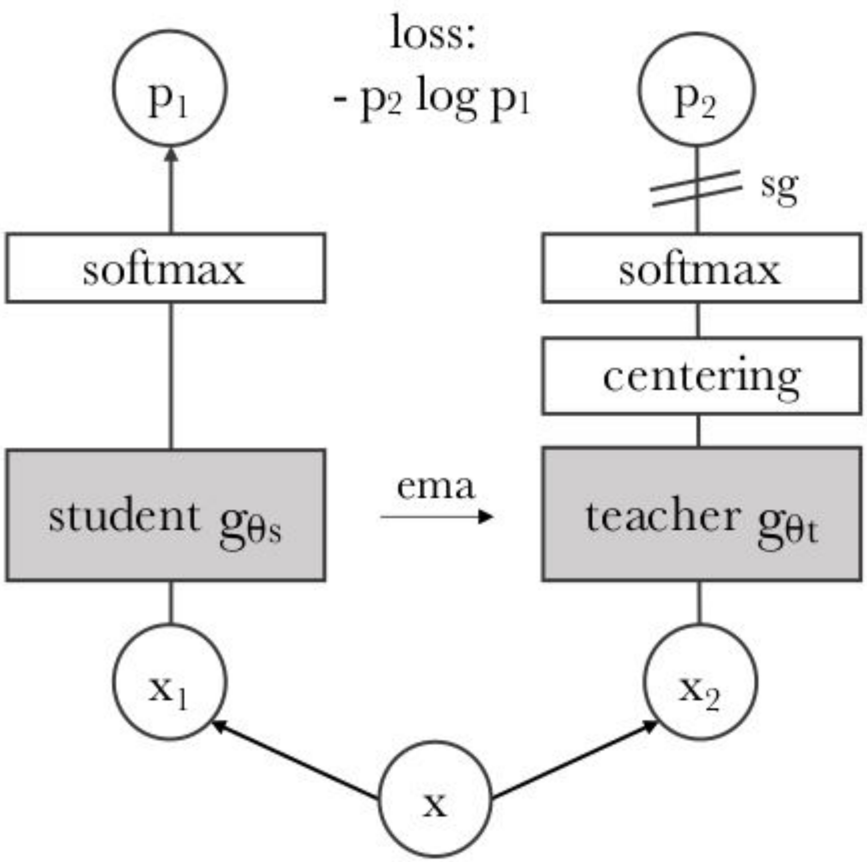
- Long receptive field of transformers: Long range



Case Study (b): Transformer-based models for WSOL (Ongoing work)

- Dino: self-distillation

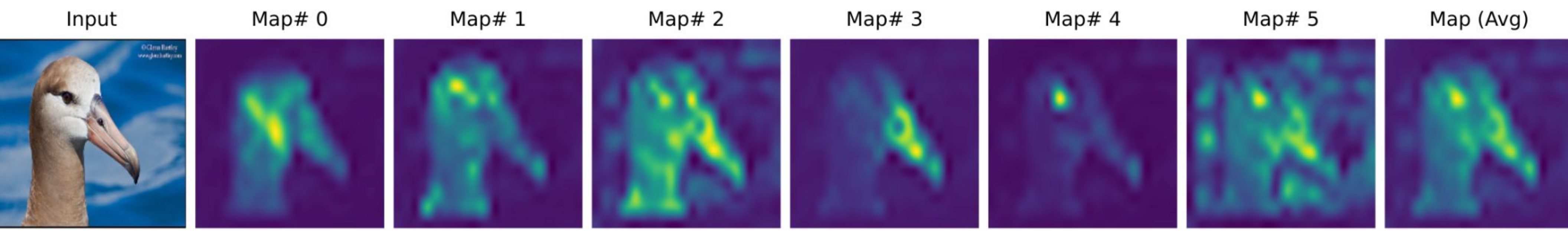
ViT: Vision Transformers



Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., & Joulin, A. (2021). Emerging properties in self-supervised vision transformers. In ICCV.

Case Study (b): Transformer-based models for WSOL (Ongoing work)

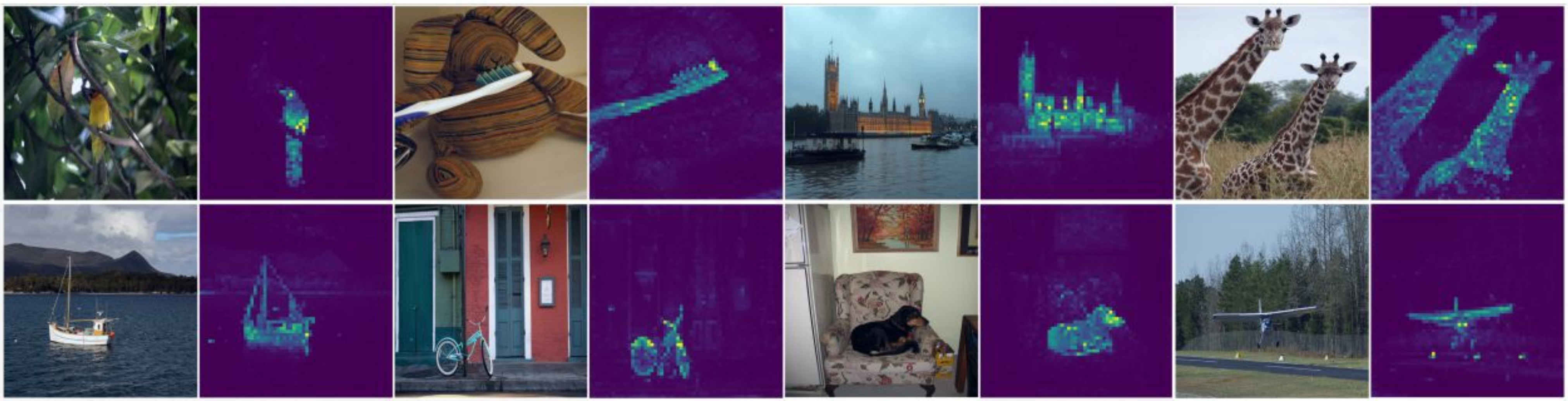
- Dino: self-distillation



Localization at **output 6 heads.**
Which head to select?

Case Study (b): Transformer-based models for WSOL
(Ongoing work)

- Dino: self-distillation



Localization at Dino's **output heads:**
Best head selected using ground truth

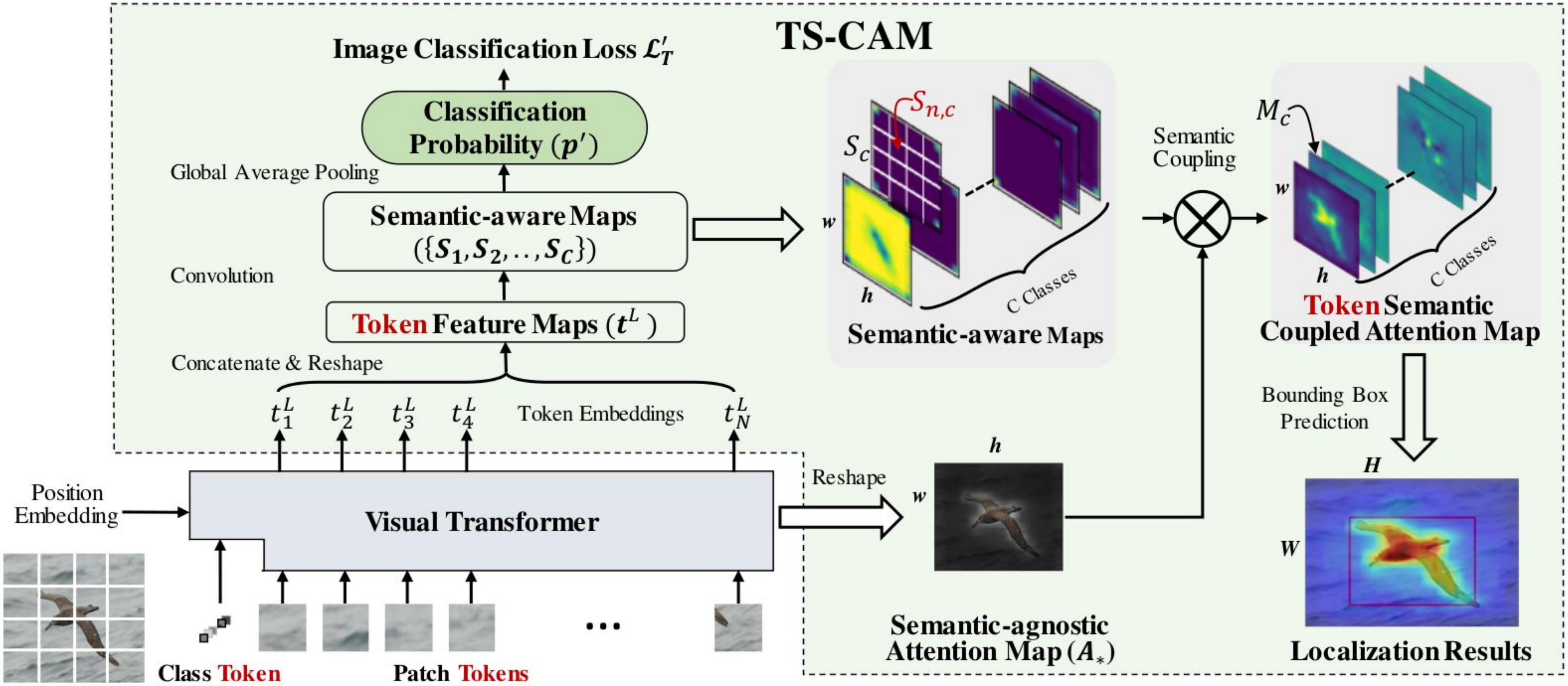
Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., & Joulin, A. (2021). Emerging properties in self-supervised vision transformers. In ICCV.

Le génie pour l'industrie

Case Study (b): Transformer-based models for WSOL (Ongoing work)

- TS-CAM: Token Semantic CAM

How to build CAM?



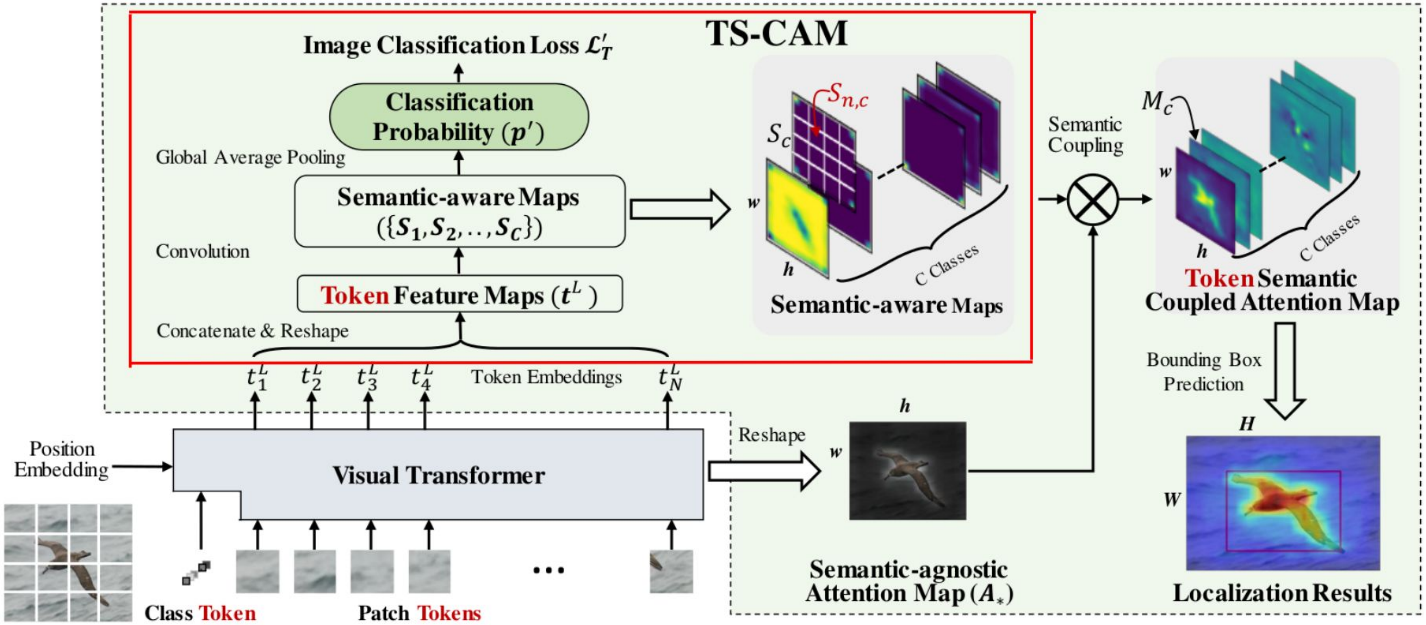
Gao, W., Wan, F., Pan, X., Peng, Z., Tian, Q., Han, Z., ... & Ye, Q. (2021). Ts-cam: Token semantic coupled attention map for weakly supervised object localization. In ICCV..

Case Study (b): Transformer-based models for WSOL (Ongoing work)

- TS-CAM: Token Semantic CAM

How to build CAM?

CAMs



$$S_c = \sum_d \mathbf{t}_d^L * k_{c,d},$$

3x3 conv

$$\mathcal{L}'_T = -\log p'_y$$
$$= -\log \frac{\exp\left(\sum_n S_{n,y}/N\right)}{\sum_c \exp\left(\sum_n S_{n,c}/N\right)},$$

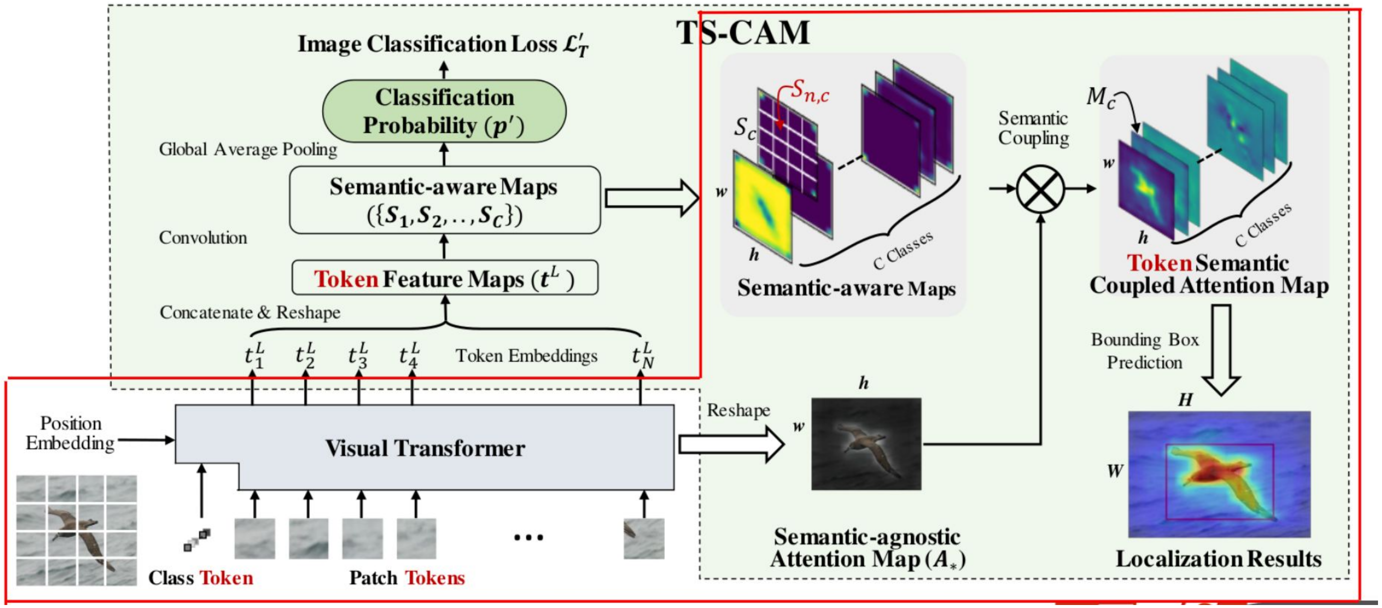
Le génie pour l'industrie

Gao, W., Wan, F., Pan, X., Peng, Z., Tian, Q., Han, Z., ... & Ye, Q. (2021). Ts-cam: Token semantic coupled attention map for weakly supervised object localization. In ICCV..

Case Study (b): Transformer-based models for WSOL (Ongoing work)

- TS-CAM: Token Semantic CAM

How to build CAM?



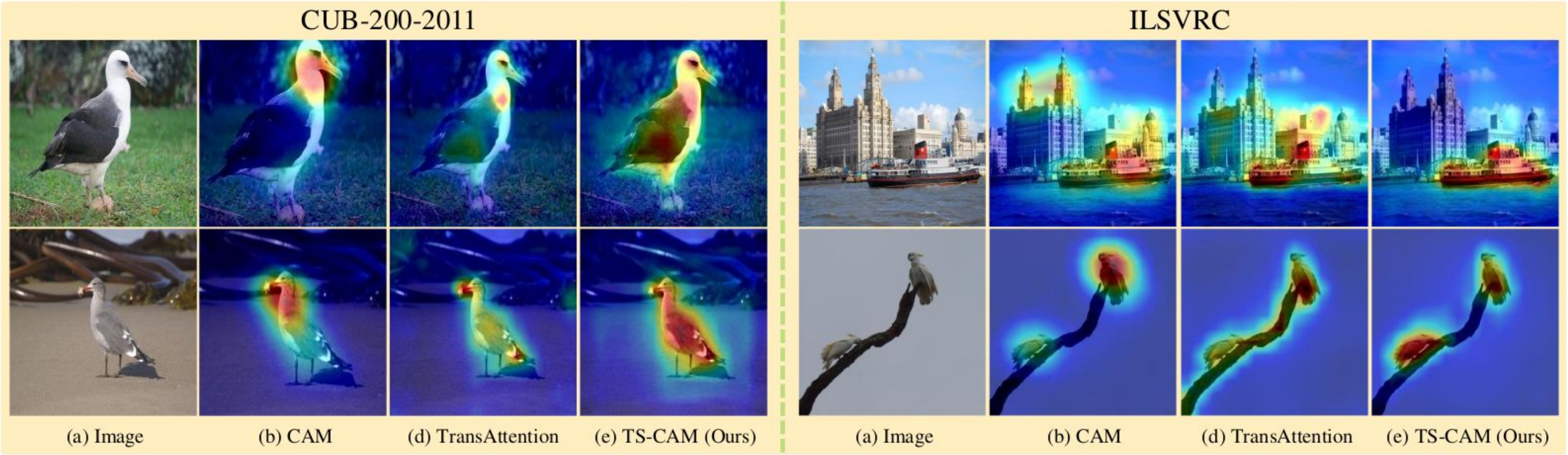
Cross-layers attention:

$$A_* = \frac{1}{L} \sum_l A_*^l,$$

Case Study (b): Transformer-based models for WSOL (Ongoing work)

- TS-CAM: Token Semantic CAM

How to build CAM?



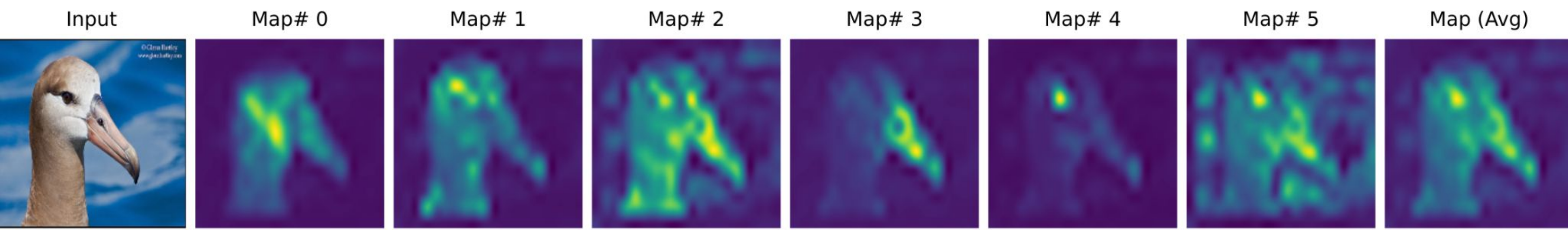
Gao, W., Wan, F., Pan, X., Peng, Z., Tian, Q., Han, Z., ... & Ye, Q. (2021). Ts-cam: Token semantic coupled attention map for weakly supervised object localization. In ICCV..

ÉTS Le génie pour l'industrie

Case Study (b): Transformer-based models for WSOL (Ongoing work)

- Ongoing work

How to build CAM?



Localization at **output 6 heads.**
Which head to select?

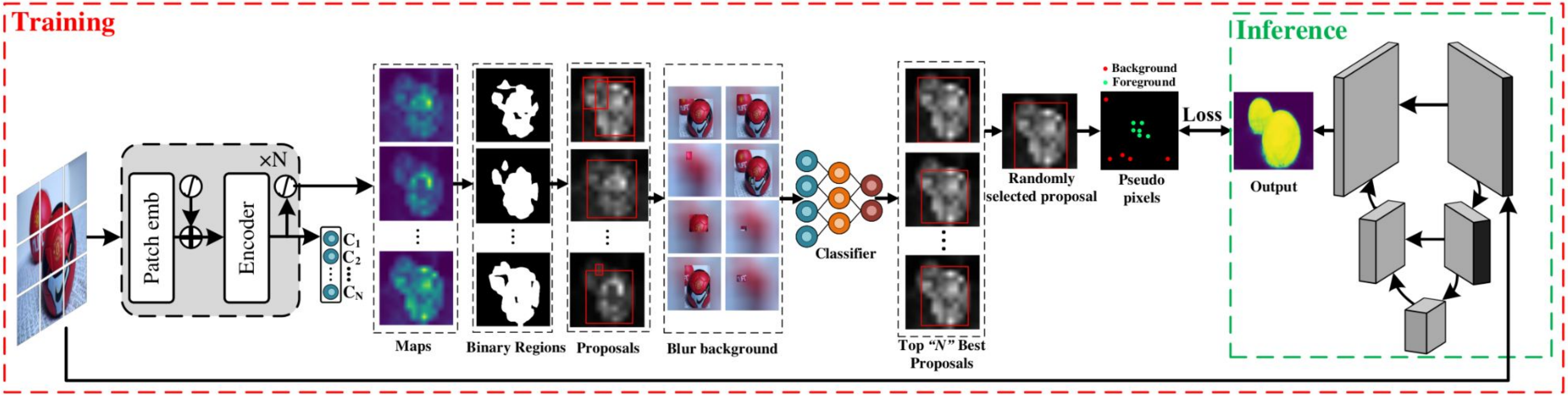
–> Fuse heads!

Le génie pour l'industrie

Case Study (b): Transformer-based models for WSOL (Ongoing work)

- Ongoing work: fusion of heads

How to build CAM?

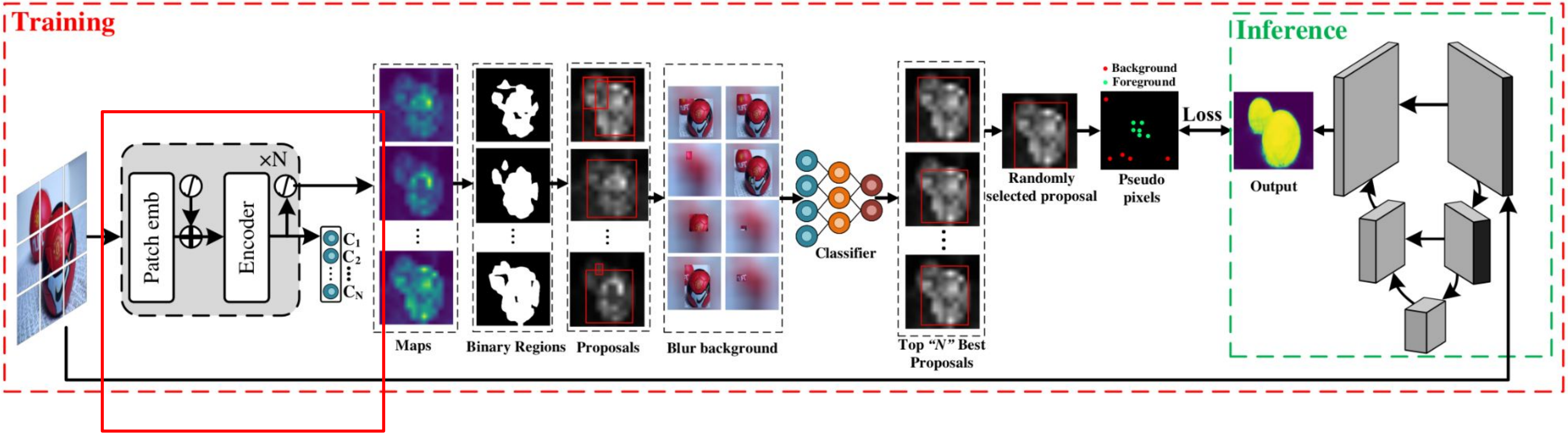


Case Study (b): Transformer-based models for WSOL (Ongoing work)

- Ongoing work: fusion of heads

How to build CAM?

Training



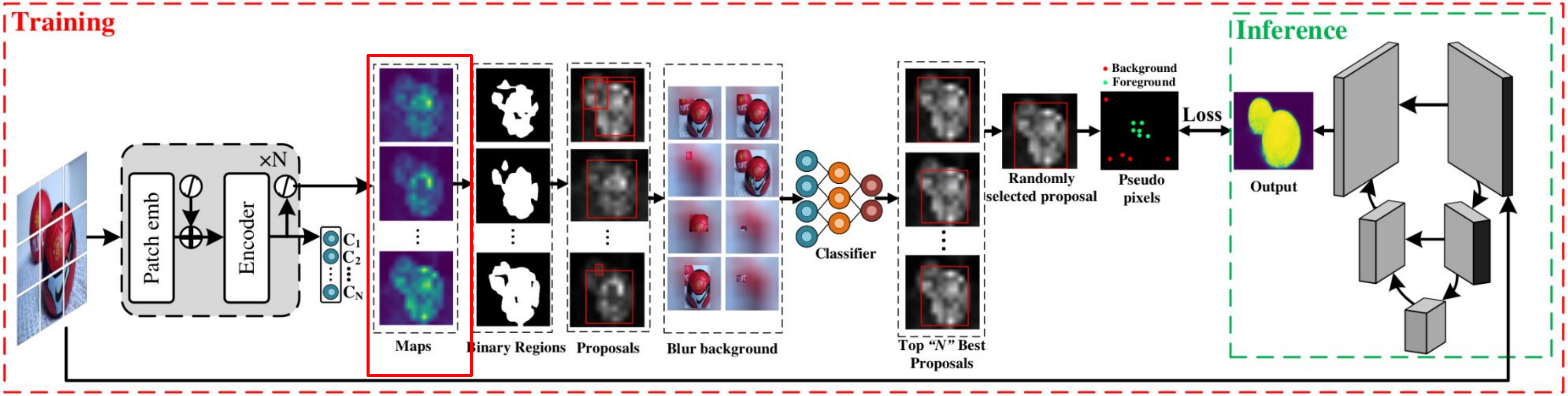
Dino

Case Study (b): Transformer-based models for WSOL (Ongoing work)

- Ongoing work: fusion of heads

How to build CAM?

Training



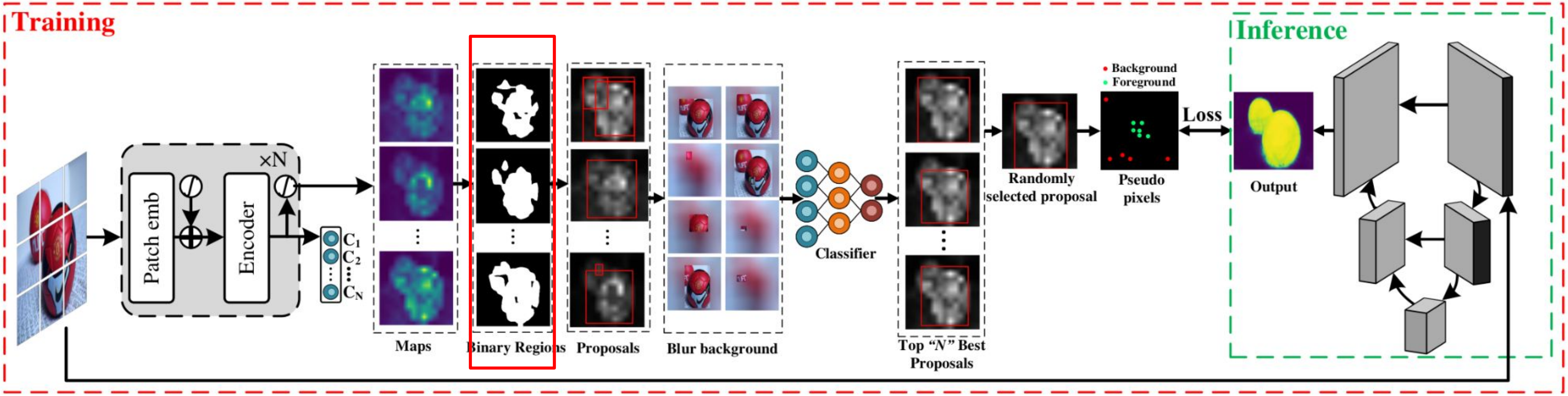
K heads

Case Study (b): Transformer-based models for WSOL (Ongoing work)

- Ongoing work: fusion of heads

How to build CAM?

Training



Proposals generation

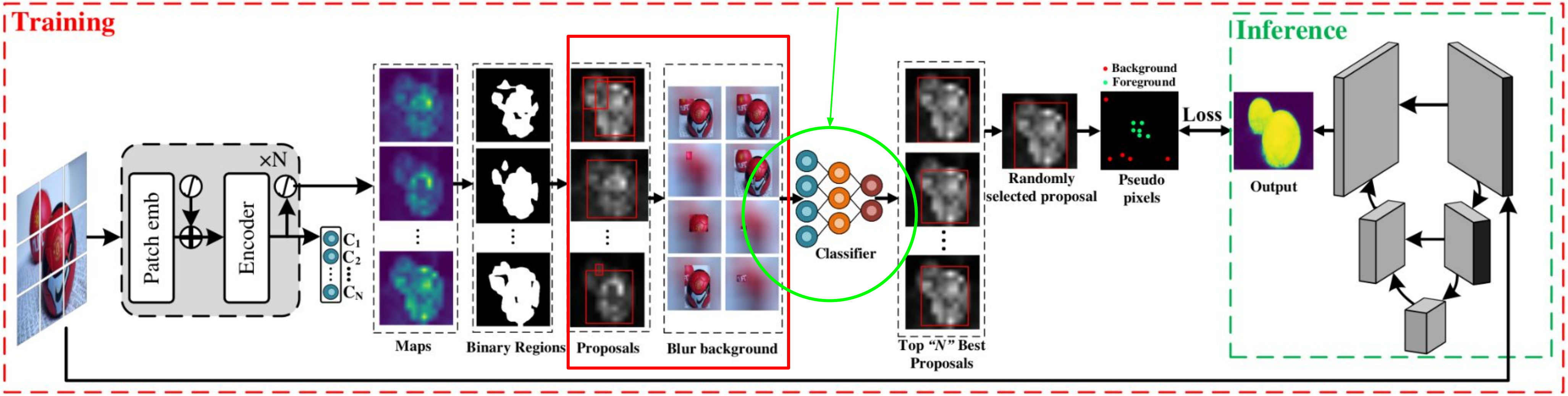
Case Study (b): Transformer-based models for WSOL (Ongoing work)

- Ongoing work: fusion of heads

How to build CAM?

Training

Pretrained classifier



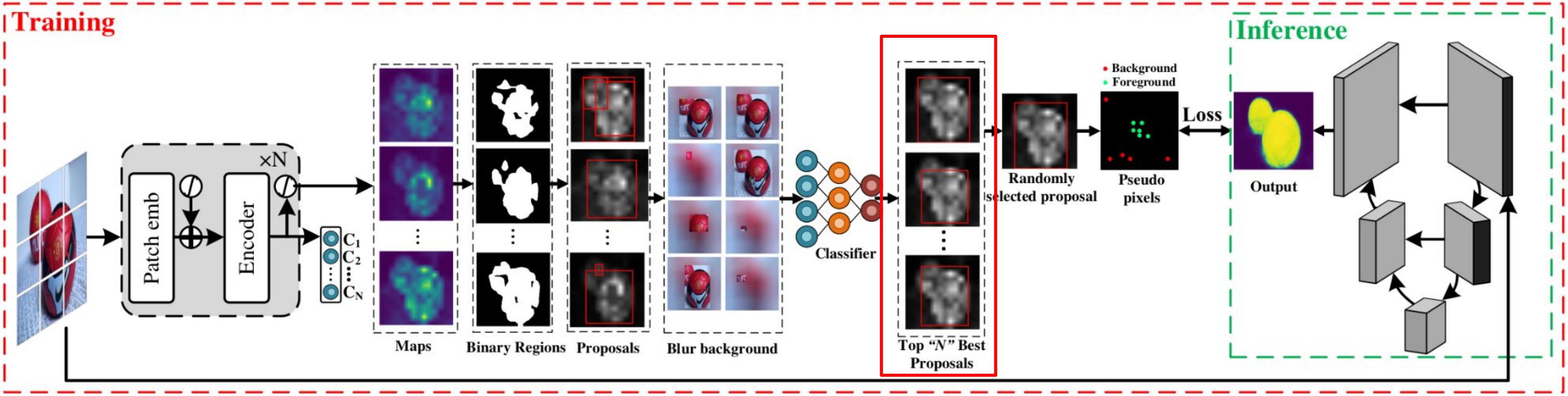
Discriminative
Proposals selection

Case Study (b): Transformer-based models for WSOL (Ongoing work)

- Ongoing work: fusion of heads

How to build CAM?

Training



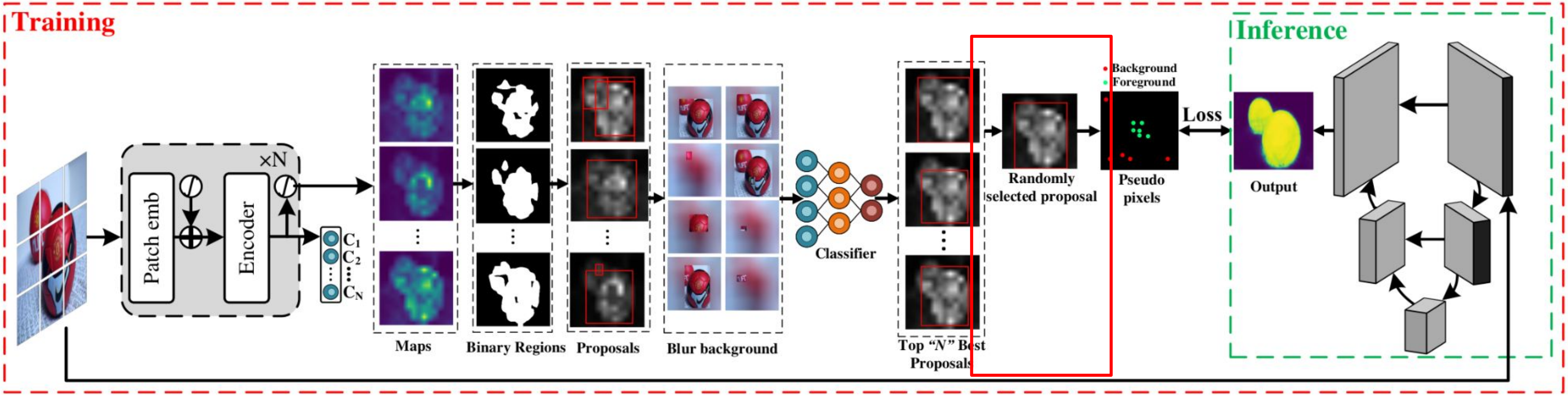
Top-n proposals

Case Study (b): Transformer-based models for WSOL (Ongoing work)

- Ongoing work: fusion of heads

How to build CAM?

Training



Random selection of a proposal from the top-n proposals

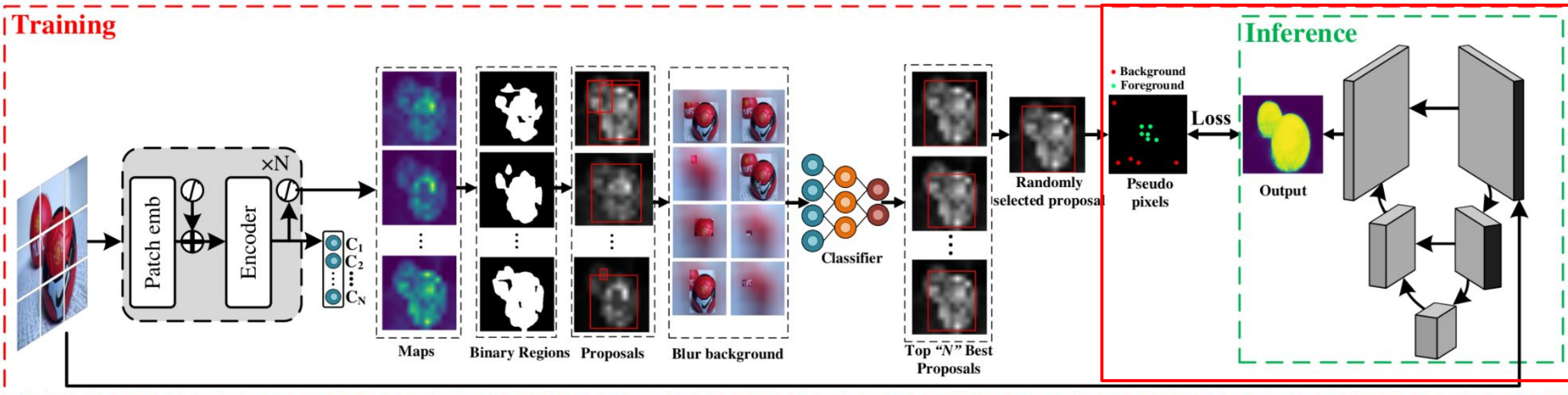
Le génie pour l'industrie

Case Study (b): Transformer-based models for WSOL (Ongoing work)

- Ongoing work: fusion of heads

How to build CAM?

Training



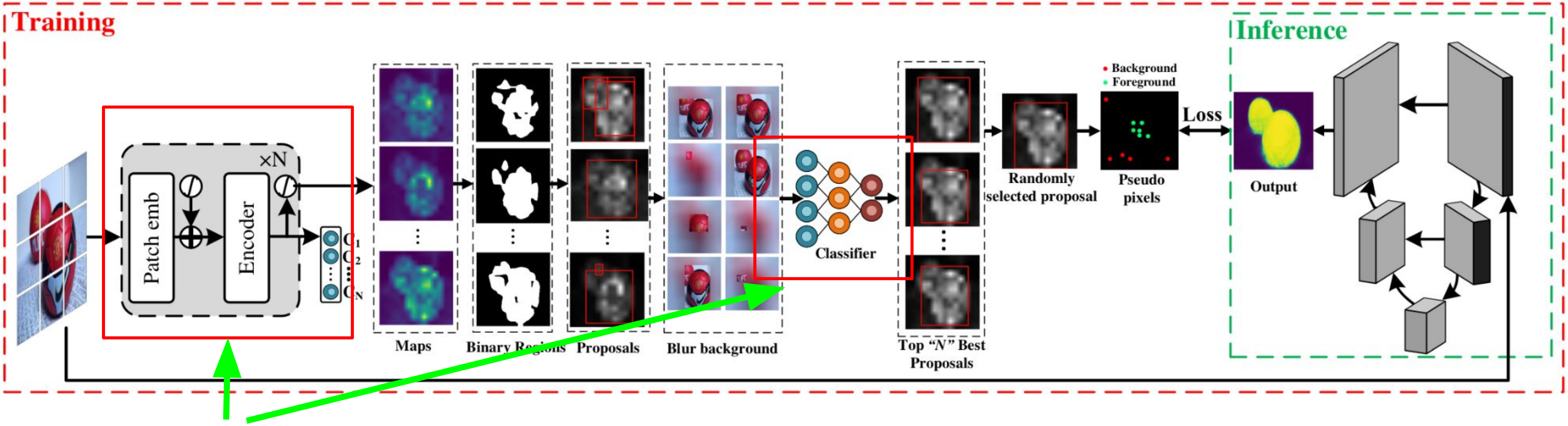
Training U-Net like net using sampled seeds (FG/BG) from proposal

Case Study (b): Transformer-based models for WSOL (Ongoing work)

- Ongoing work: fusion of heads

How to build CAM?

Training



Frozen

Train loss: U-Net params

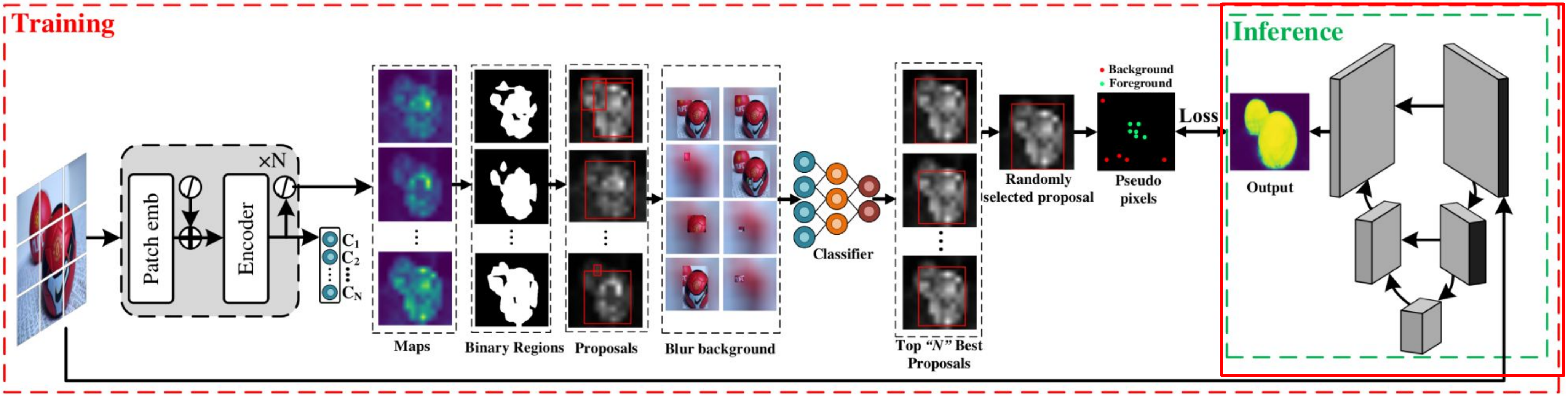
$$\mathcal{L}_{Total} = \min_{\theta} \lambda_{CLS} + \lambda_{CPA}\mathcal{L}_{CPA} + \lambda_{CRF}\mathcal{L}_{CRF}$$

Case Study (b): Transformer-based models for WSOL (Ongoing work)

- Ongoing work: fusion of heads

How to build CAM?

Inference

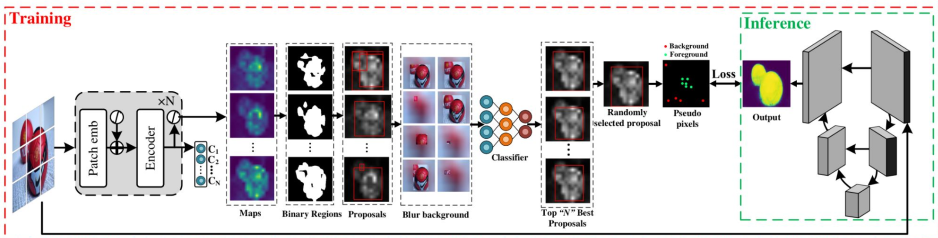


Localization: Using only input image and U-Net.

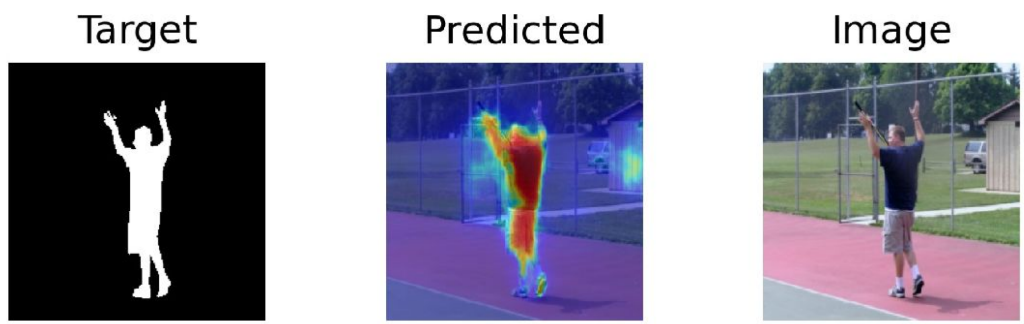
Case Study (b): Transformer-based models for WSOL (Ongoing work)

- Ongoing work: fusion of heads

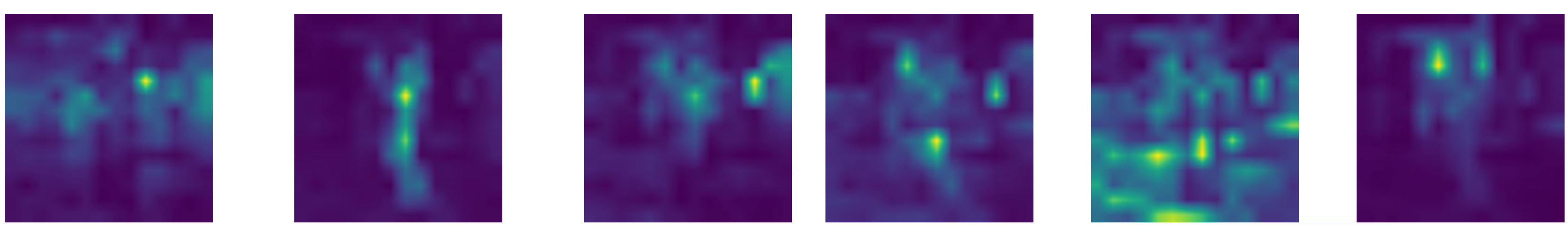
How to build CAM?



Model prediction



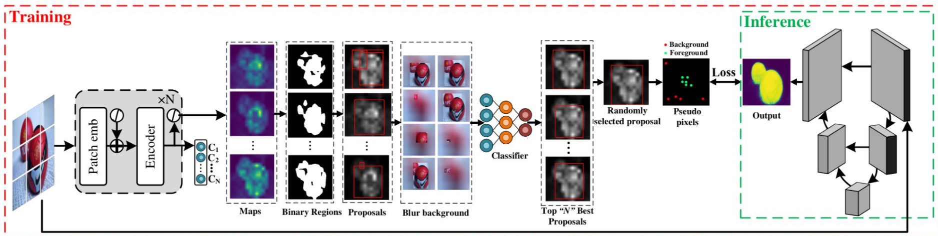
Dino heads



Case Study (b): Transformer-based models for WSOL (Ongoing work)

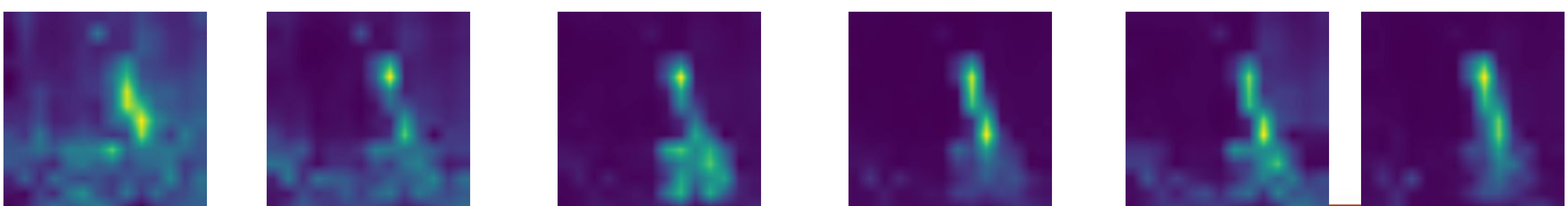
- Ongoing work: fusion of heads

How to build CAM?



Model prediction

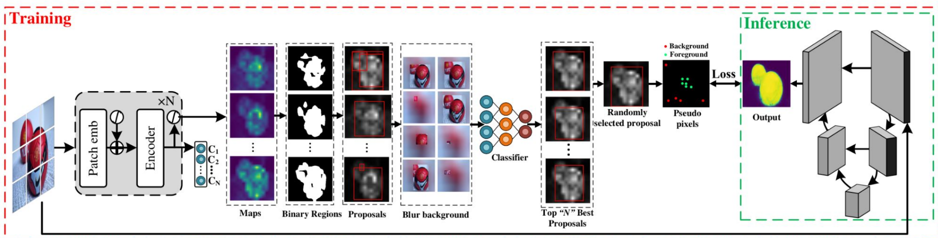
Dino heads



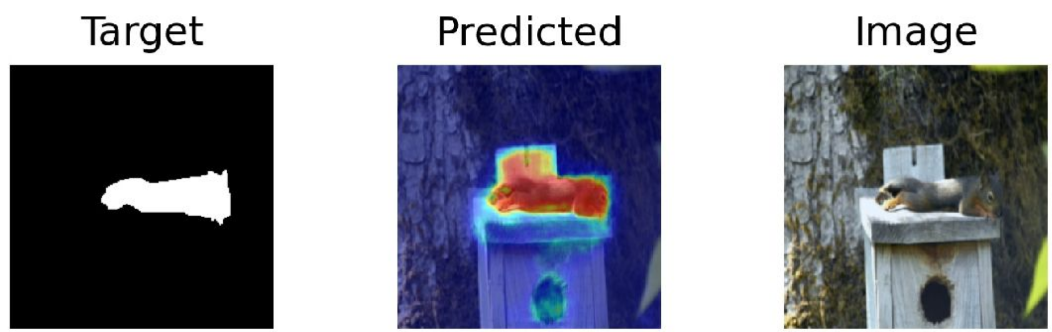
Case Study (b): Transformer-based models for WSOL (Ongoing work)

- Ongoing work: fusion of heads

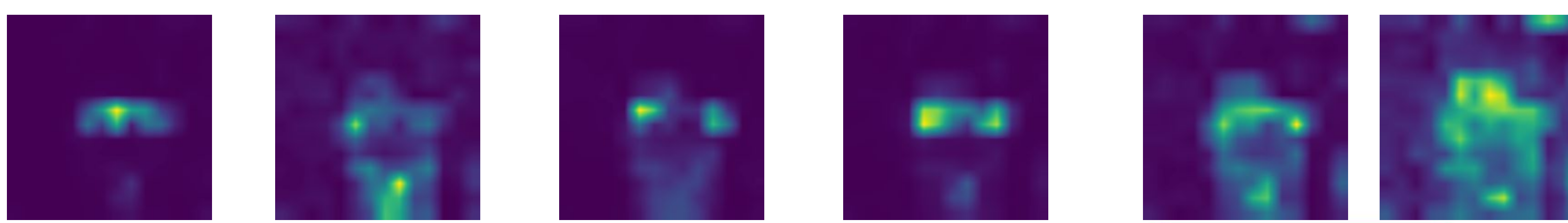
How to build CAM?



Model prediction

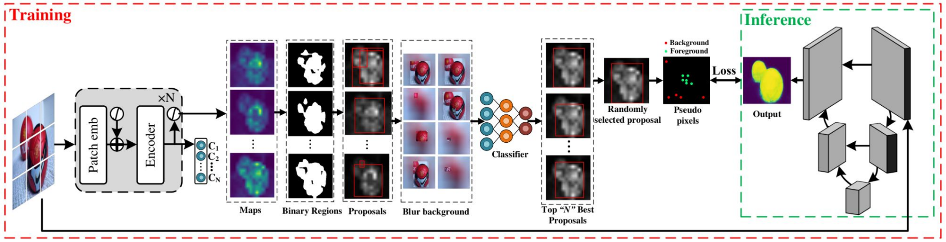


Dino heads



Case Study (b): Transformer-based models for WSOL
(Ongoing work)

- Ongoing work: fusion of heads



Paper + code: will be available early september.

Online Resources

F-CAM (WACV 2022):

- paper: https://arxiv.org/abs/2109.07069
- code: https://github.com/sbelharbi/fcam-wsol

F-CAM extension to histology (MIDL 2022):

- paper: https://arxiv.org/abs/2201.02445
- code: https://github.com/sbelharbi/negev

Part 3:

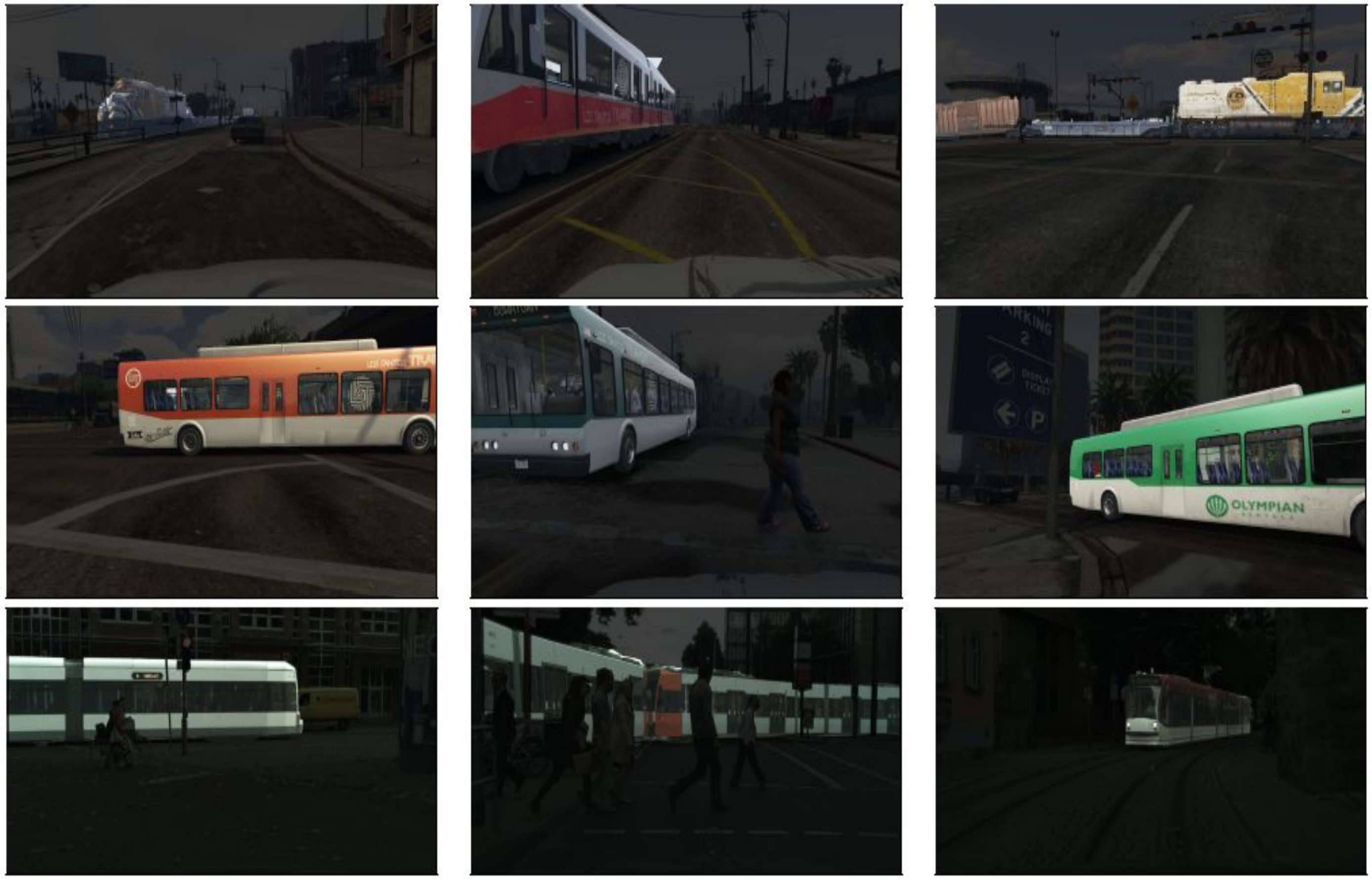
Review of Loss Functions and Settings for WSSS methods

Le génie pour l'industrie

Dense pixel-wise labels are very expensive

Le génie pour l'industrie

Domain shifts make things worse



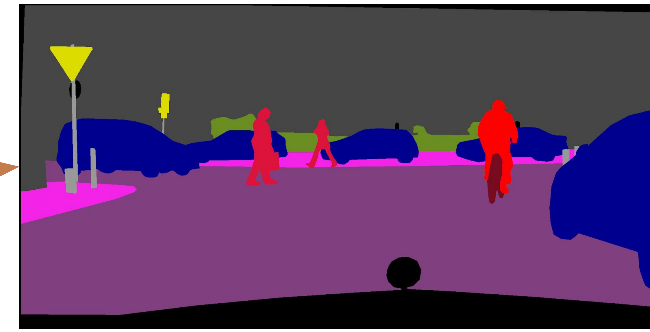
"train"
GTA

"bus"
GTA

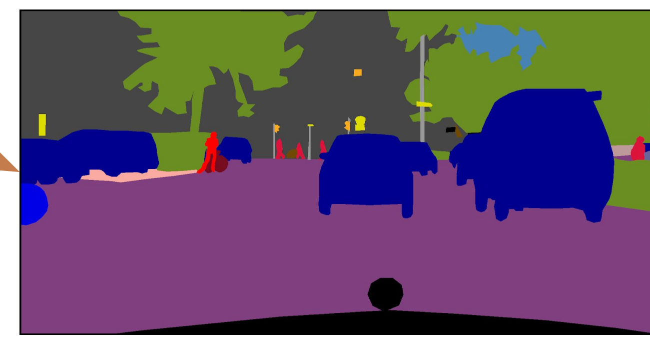
"train"
Cityscapes

Figures from [Zhang et al., A Curriculum Domain Adaptation Approach to the Semantic Segmentation of Urban Scenes TPAMI 2019]

Domain shifts make things worse



Frankfurt



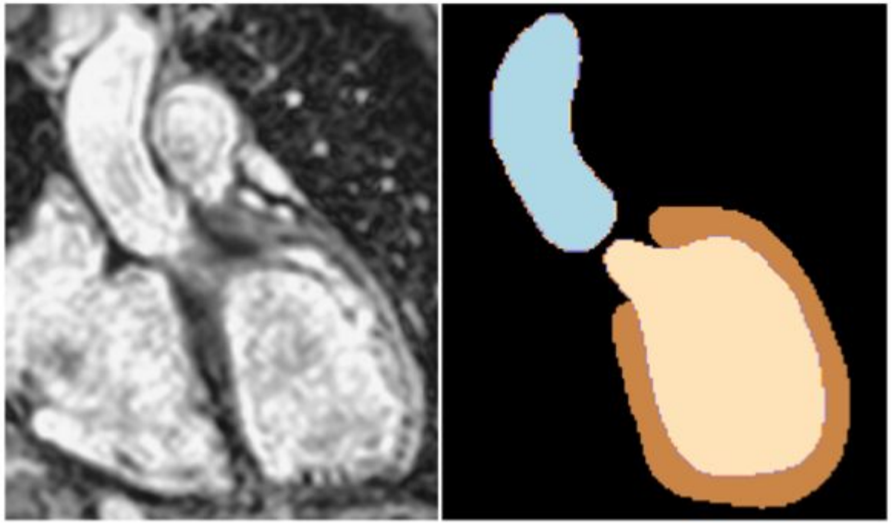
Zurich

road	sidewalk	building	wall	fence	pole	traffic light	traffic sign	vegetation	terrain
sky	person	rider	car	truck	bus	train	motorcycle	bicycle	unlabeled

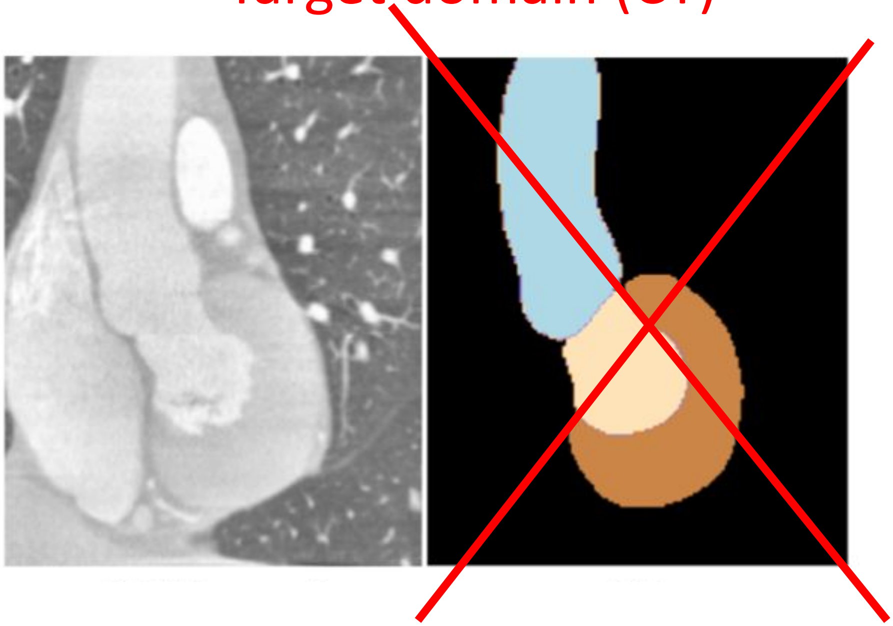
Cityscapes (5000 images): labeling of 1 image takes 90 min at average
[Cordt et al., CVPR 2016]

Domain shifts make things worse

Source domain (MRI)



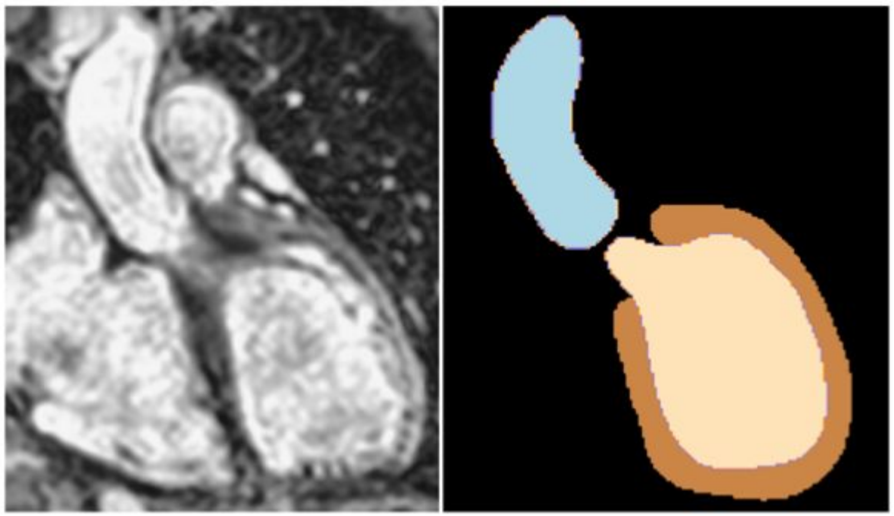
Target domain (CT)



[Images from Bateson et al., Constrained Domain Adaptation for Image Segmentation, TMI'21]

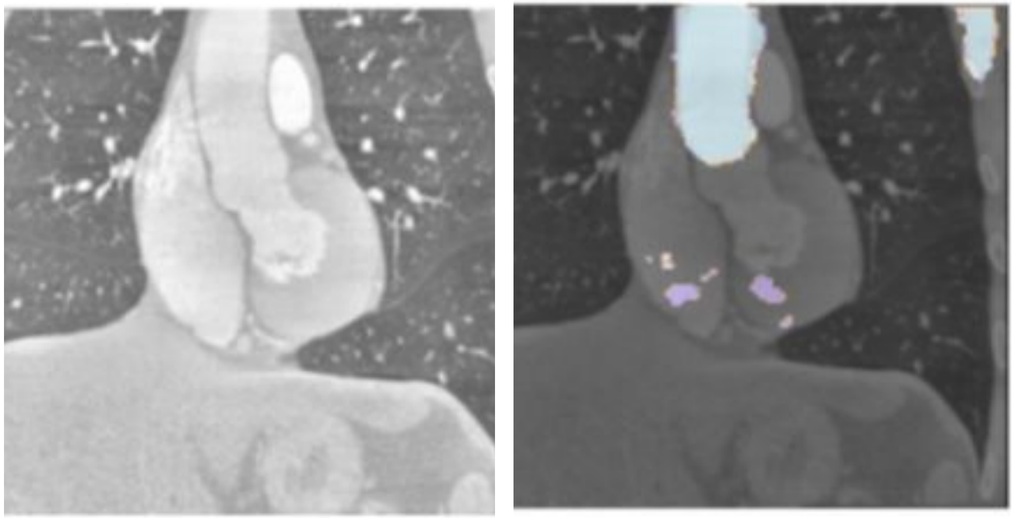
Domain shifts make things worse

Source domain (MRI)



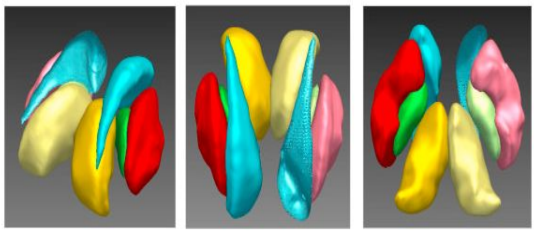
No adaptation

(bad generalization to the target)

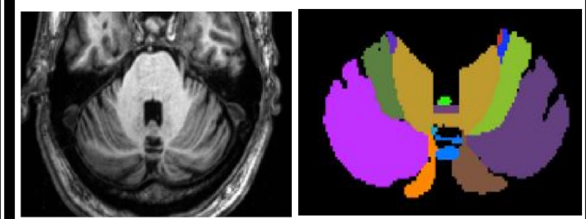


[Images from Bateson et al., Constrained Domain Adaptation for Image Segmentation, TMI'21]

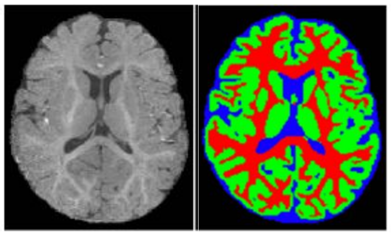
…and more complex is some applications (e.g medical image analysis)



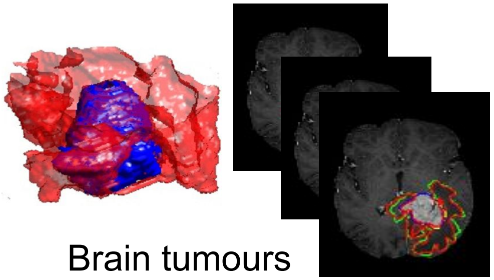
Subcortical structures
(Dolz et al., Neuroimage 2018)



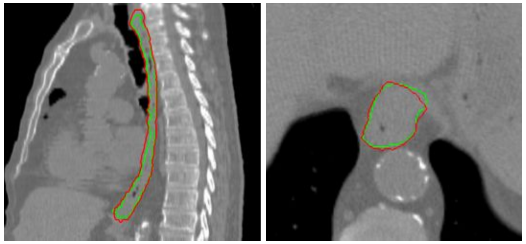
Cerebellum parcellation
(Carass et al., Neuroimage 2018)



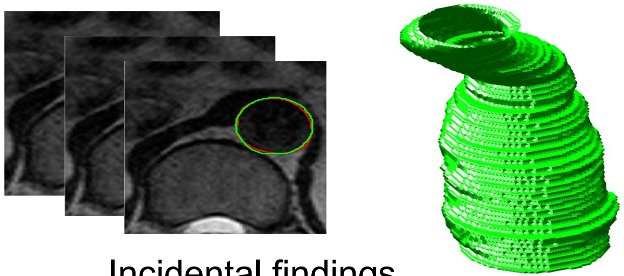
Brain tissues (6-month infant)
(Li et al., TMI 2019)



Brain tumours
(Njeh et al., CMIG 2015)



Organs at risk
(Dolz et al., Med. Phys. 2017)



Incidental findings
(Ben Ayed et al., MICCAI 2014)

Dense labels are more complicated in medical imaging:

Not anywhere close to the 10k images of Pascal VOC and the 5k of Cityscapes

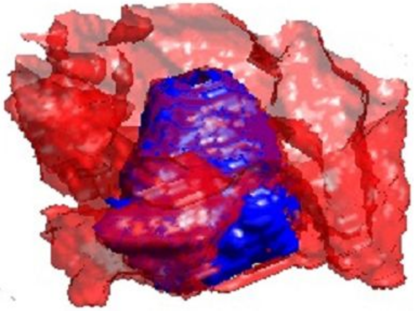
Crowdsourcing?

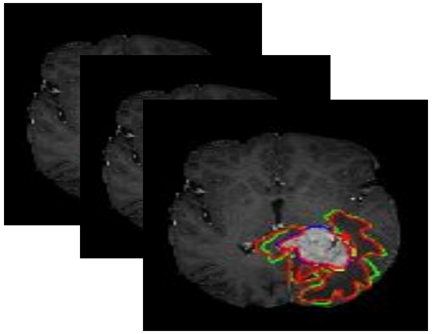
Dense labels are more complicated in medical imaging:

Not anywhere close to the 10k images of Pascal VOC and the 5k of Cityscapes

Crowdsourcing?

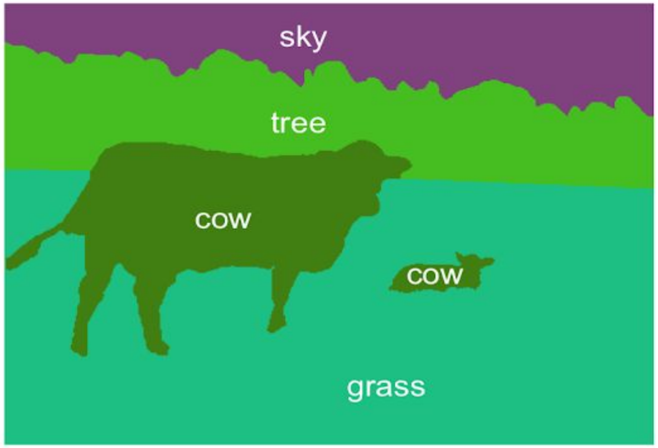
Dense 3D annotations: several hours (of radiologist time)



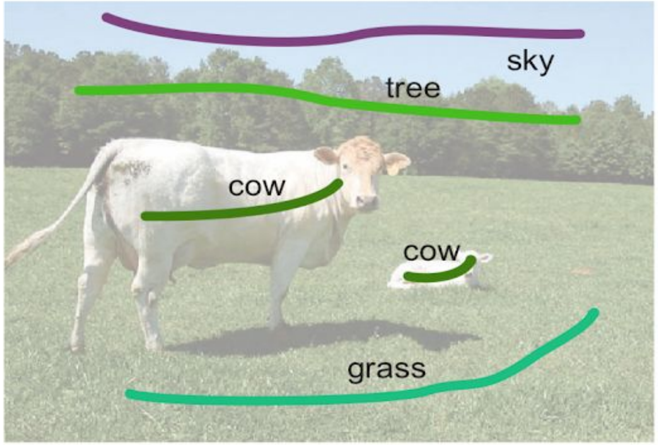


Semi-supervised learning (SSL)

A lot of unlabeled data, and only a fraction of points are labeled



Full annotations



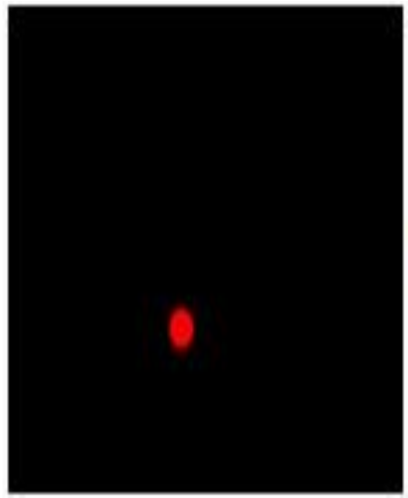
Semi-supervised

Figures from Lin et al. Scribblesup: Scribble-supervised convolutional networks for semantic segmentation, CVPR 2016

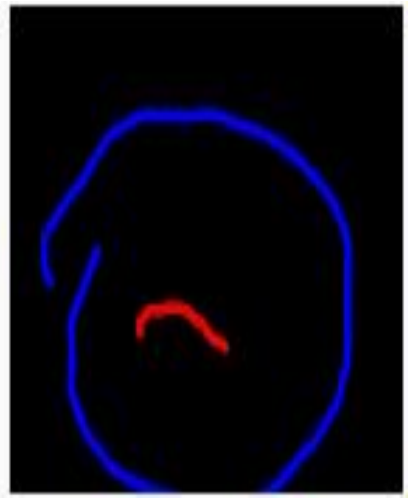
Forms of weak supervision in segmentation

Car
Parking
Sky
No person

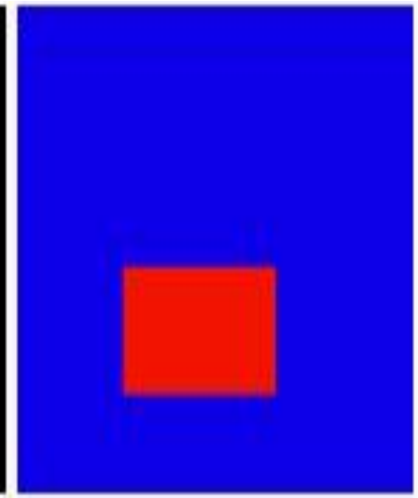
Image tags
(MIL)



Points
(SSL)



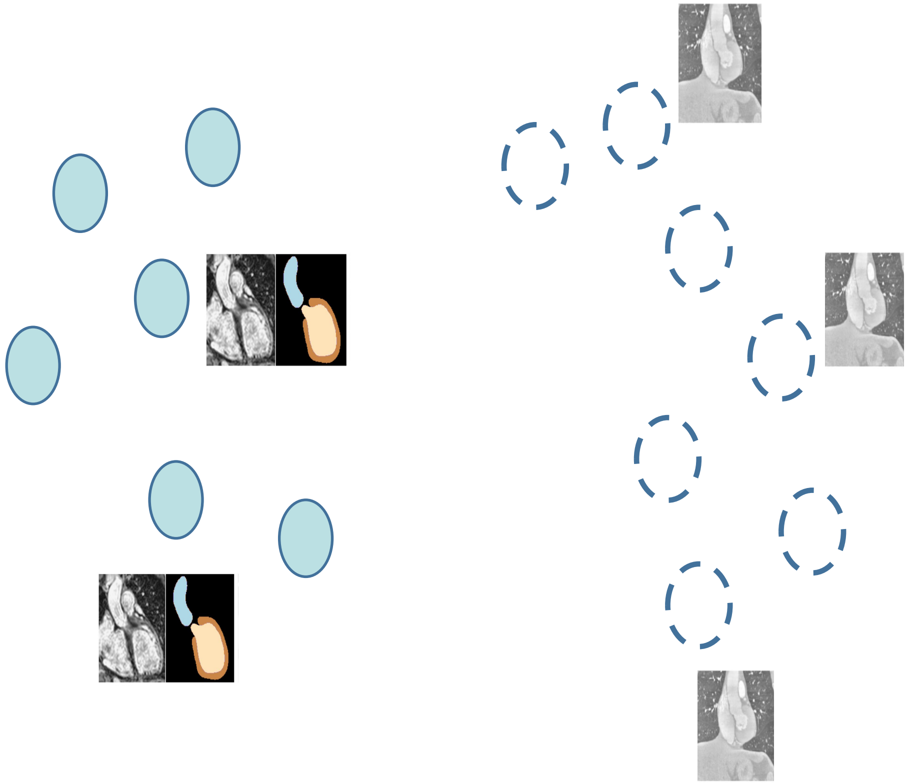
Scribbles
(SSL)



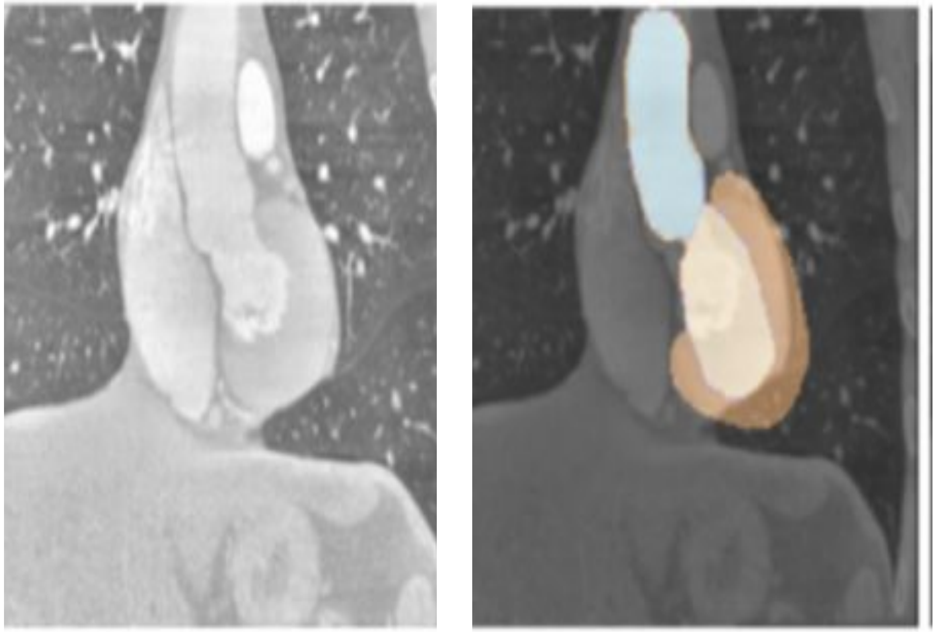
Boxes
(MIL)

Closely related problem: Unsupervised Domain Adaptation (UDA)

Training on both
labeled and unlabeled data

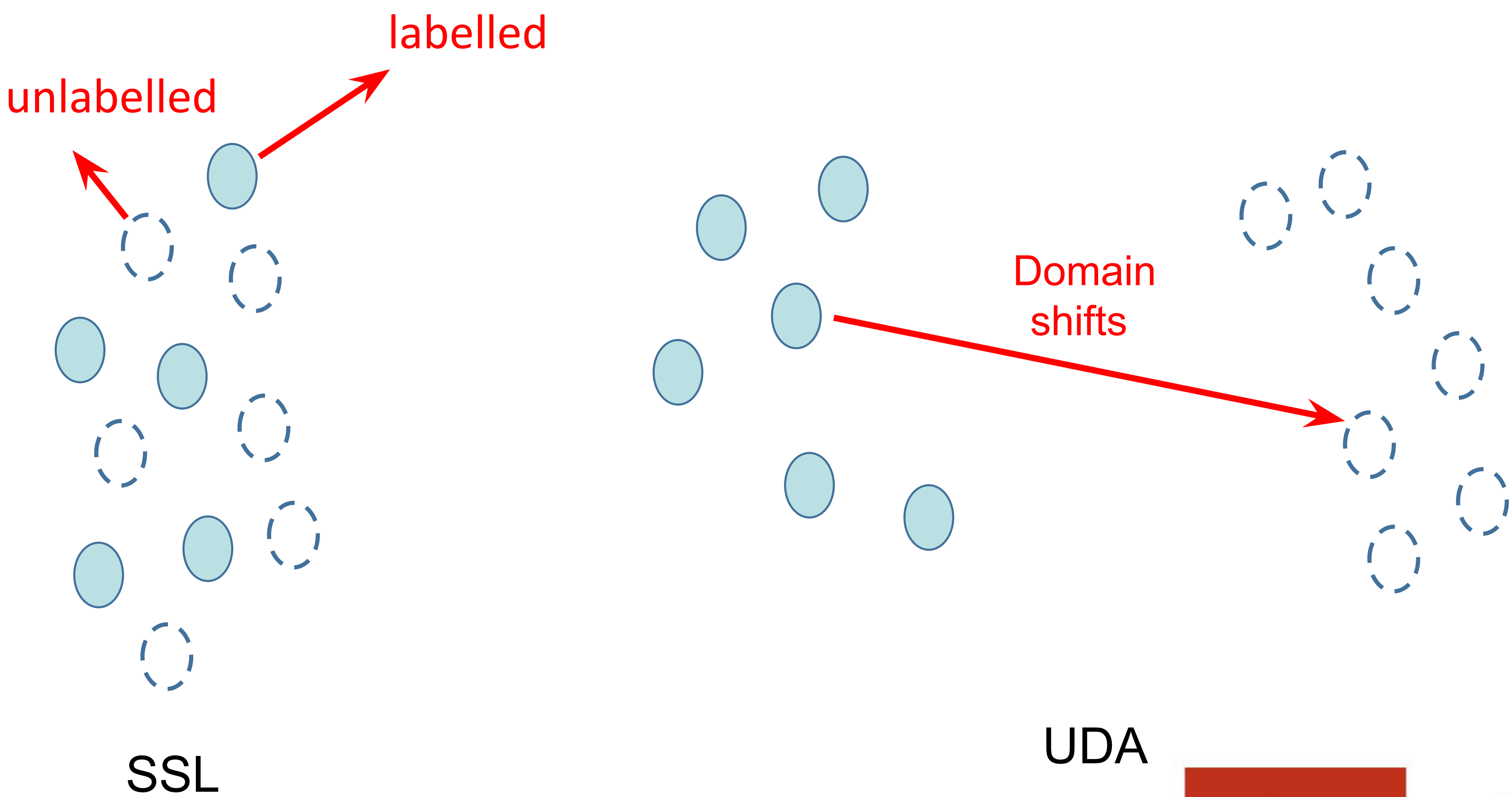


Corrected result with UDA



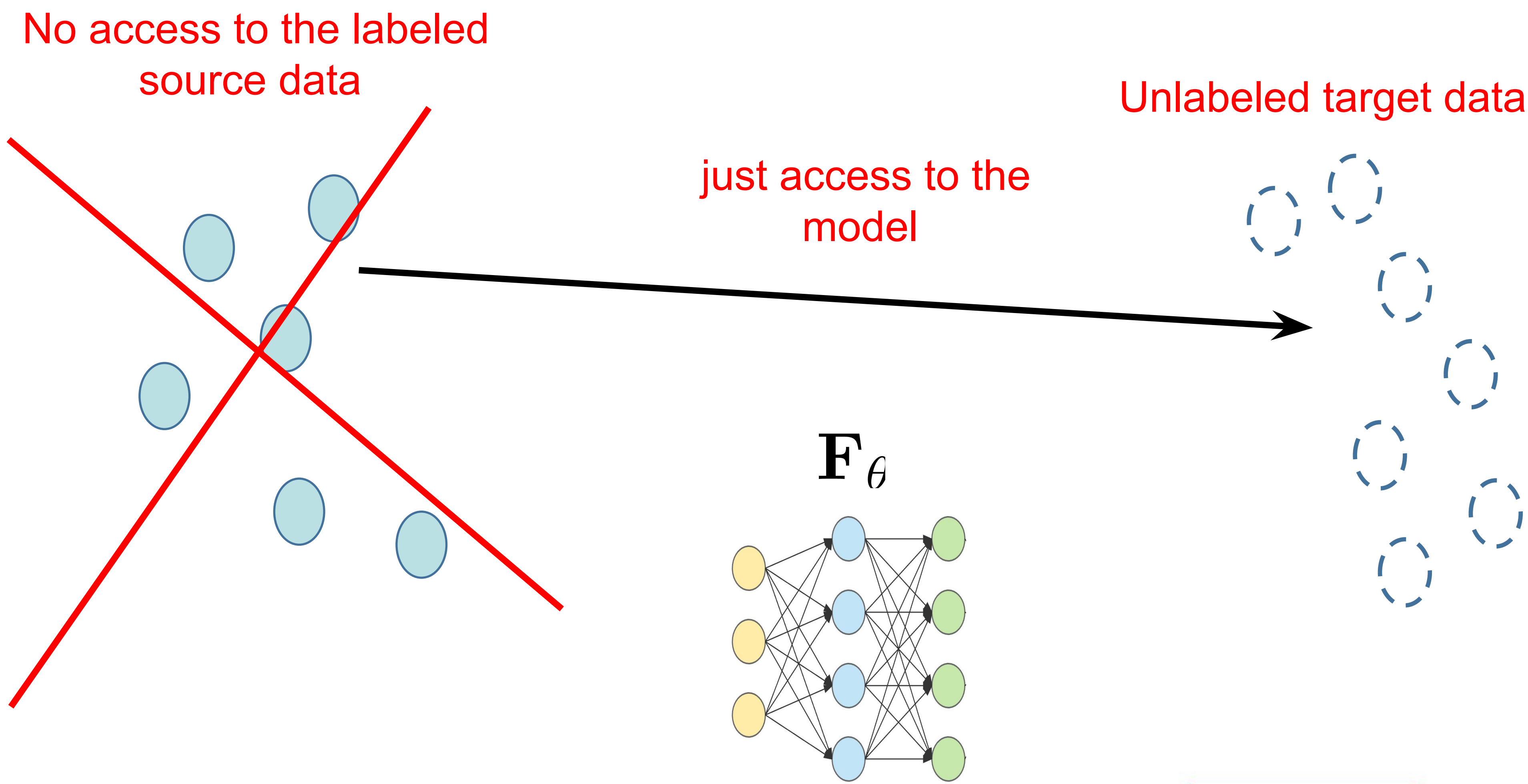
[Images from Bateson et al., Constrained Domain Adaptation for Image Segmentation, TMI’21]

UDA = SSL +Domain Shifts



Test-time adaptation (TTA)

UDA without access to the source data



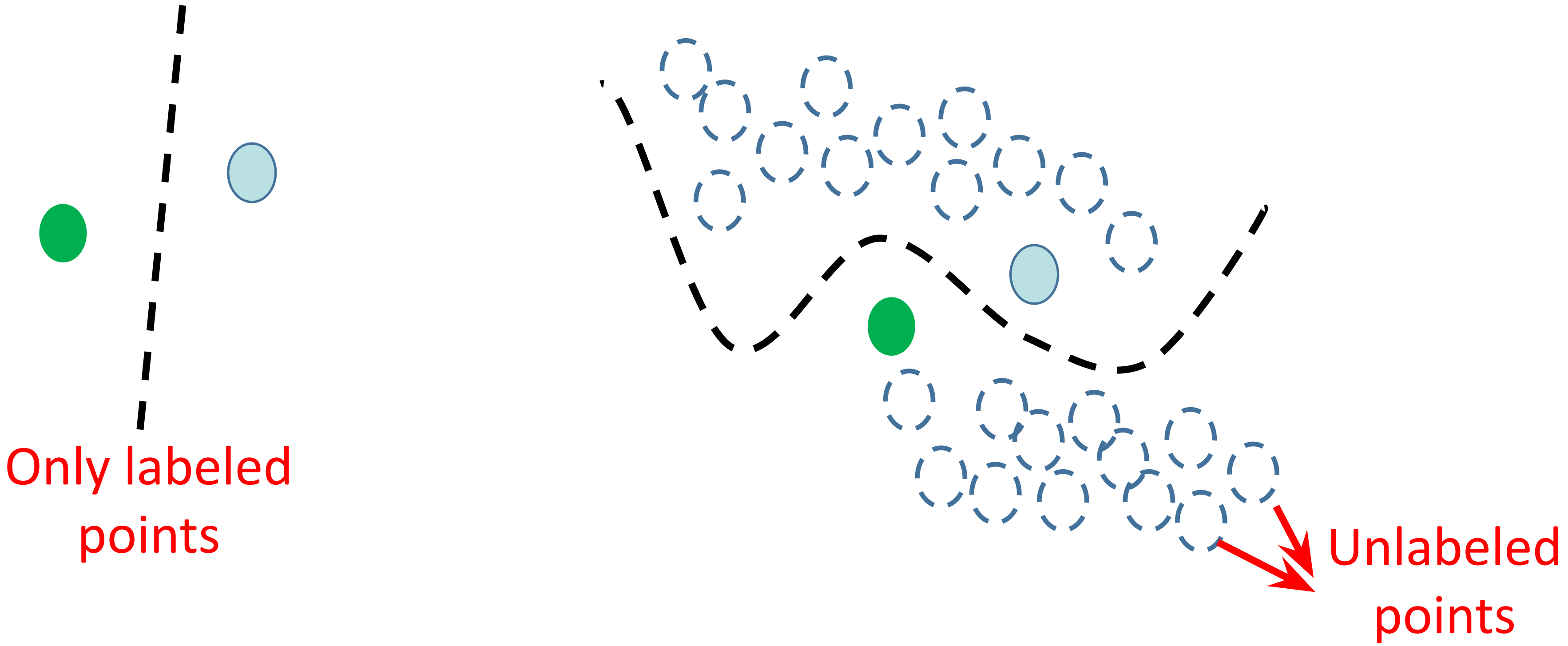
SSL/MIL/UDA/TTA unsupervised loss functions in a nutshell

Leveraging unlabeled data with priors

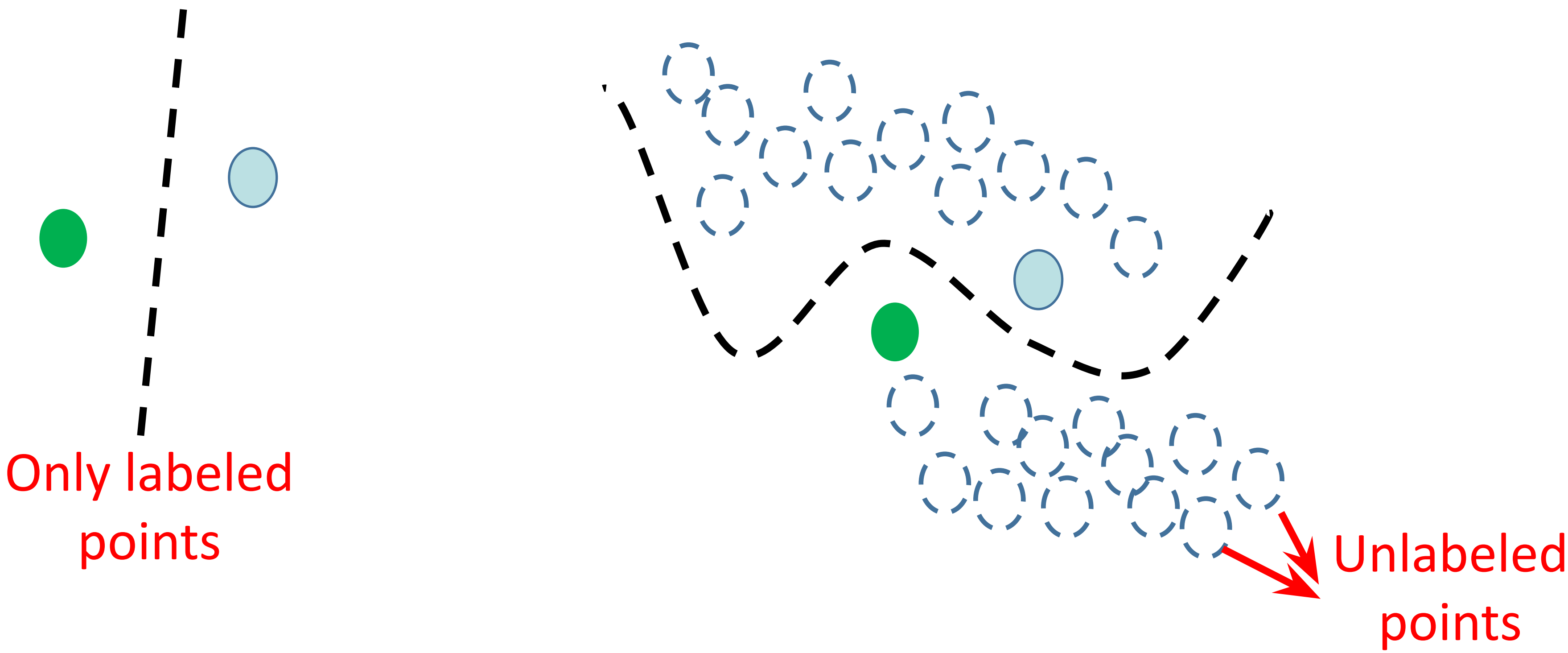
- Structure-driven priors: Regularization
- Knowledge-driven priors (e.g., anatomical constraints)
- Invariance priors (e.g., contrastive learning)
- Multi-modal priors (e.g., text info associated with the images)

Unsupervised Manifold Regularization

In SSL: Learning from both unlabeled and labeled data

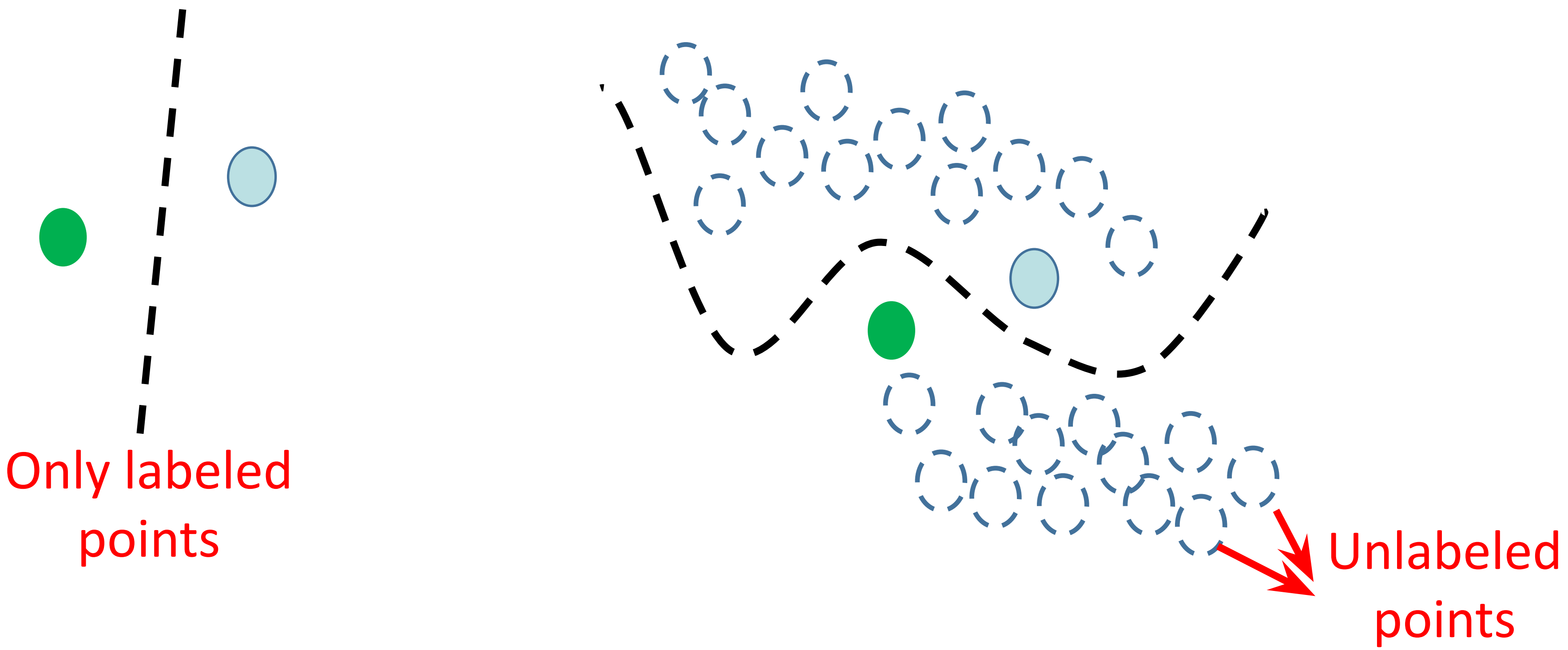


In SSL: Learning from both unlabeled and labeled data



$$\sum_{p \in \mathcal{L}} l(\mathbf{f}_\theta^p, \mathbf{y}^p) + \sum_{p,q \in \mathcal{L} \cup \mathcal{U}} g(\mathbf{f}_\theta^p, \mathbf{f}_\theta^q, w_{p,q})$$

In SSL: Learning from both unlabeled and labeled data



$$\sum_{p\in\mathcal{L}} l(\mathbf{f}_\theta^p, \mathbf{y}^p) + \sum_{p,q\in\mathcal{L}\cup\mathcal{U}} g(\mathbf{f}_\theta^p, \mathbf{f}_\theta^q, w_{p,q})$$

Labeled points

Unlabeled points

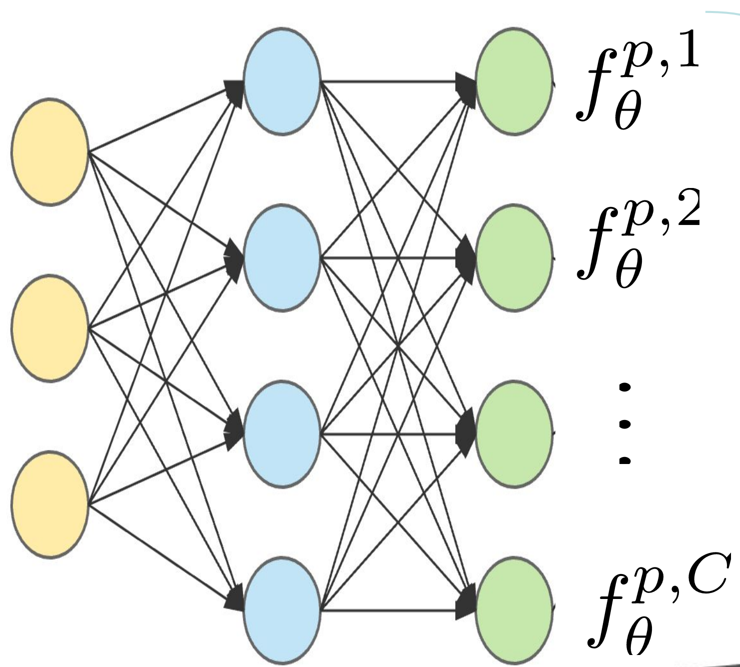
In SSL: Learning from both unlabeled and labeled data

$$\sum_{p \in \mathcal{L}} l(\mathbf{f}_\theta^p, \mathbf{y}^p) + \sum_{p,q \in \mathcal{L} \cup \mathcal{U}} g(\mathbf{f}_\theta^p, \mathbf{f}_\theta^q, w_{p,q})$$

Labels

e.g.:
cross-entropy

e.g.: *softmax outputs*



$$\mathbf{f}_\theta^p = \mathbf{s}_\theta^p \in [0, 1]^K$$

Laplacian regularization: Standard in classical SSL

$$\sum_{p \in \mathcal{L}} l(\mathbf{f}_\theta^p, \mathbf{y}^p) + \sum_{p,q \in \mathcal{L} \cup \mathcal{U}} g(\mathbf{f}_\theta^p, \mathbf{f}_\theta^q, w_{p,q})$$

Labels

e.g.: **Laplacian regularization**

$$w_{p,q} \|\mathbf{f}_\theta^p - \mathbf{f}_\theta^q\|^2$$

e.g.:
cross-entropy

e.g.: *softmax outputs*

Laplacian regularization: Standard in classical SSL

$$\sum_{p \in \mathcal{L}} l(\mathbf{f}_\theta^p, \mathbf{y}^p) + \sum_{p,q \in \mathcal{L} \cup \mathcal{U}} g(\mathbf{f}_\theta^p, \mathbf{f}_\theta^q, w_{p,q})$$

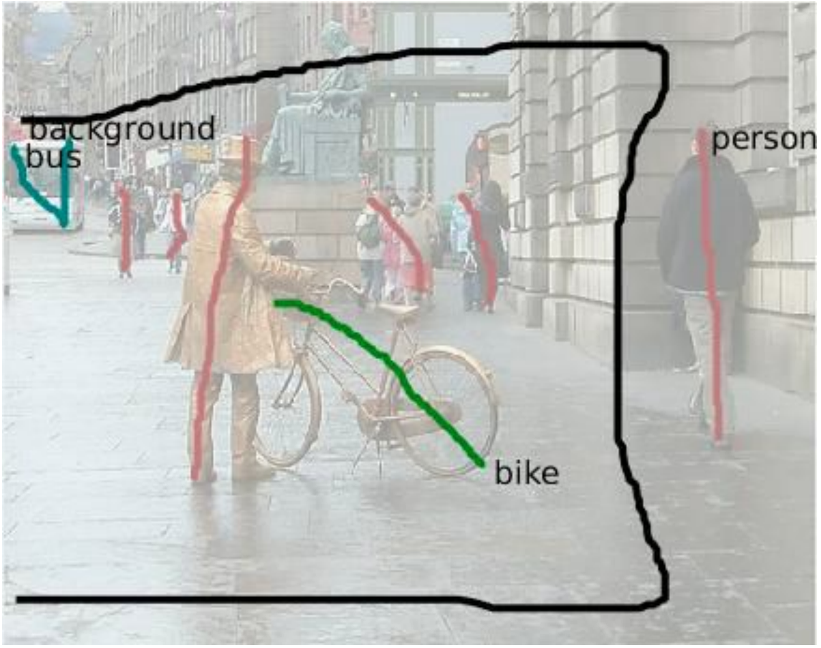
$$w_{p,q} \|\mathbf{f}_\theta^p - \mathbf{f}_\theta^q\|^2$$

- [Weston et al., Deep Learning via semi-supervised embedding, ICML 2008]
- [Belkin et al., Manifold regularization: a geometric framework for learning from Labeled and Unlabeled Examples, JMLR 2006]
- [Zhu et al., Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions, ICML 2003]

Le génie pour l'industrie

Semi-supervision loss in segmentation

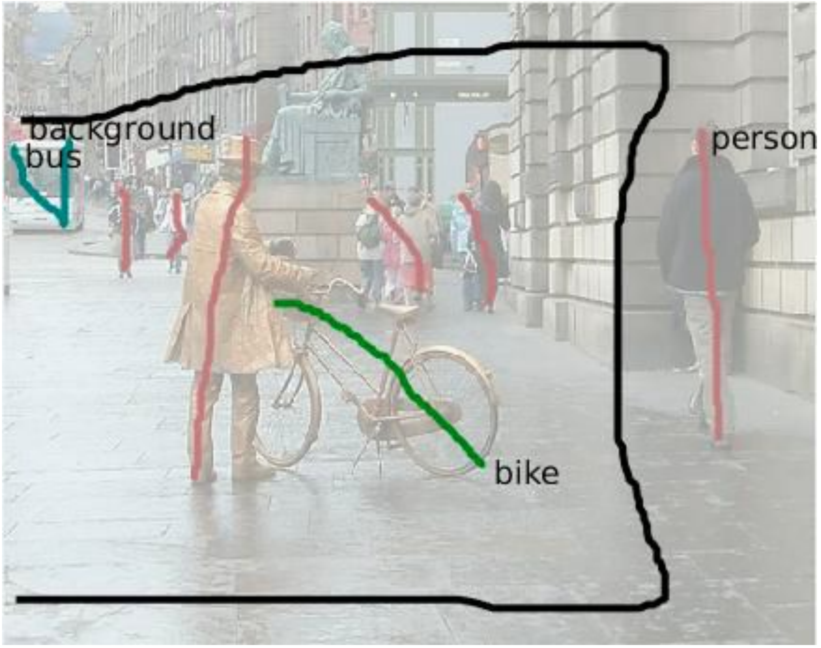
$$\min_{\theta} \sum_{p \in \mathcal{L}} l(\mathbf{y}^p, \mathbf{s}_\theta^p) + \sum_{p,q \in \mathcal{L} \cup \mathcal{U}} w_{p,q} \|\mathbf{s}_\theta^p - \mathbf{s}_\theta^q\|$$



[Tang et al., On regularized losses for weakly supervised segmentation, ECCV 2018]

Semi-supervision loss in segmentation

$$\min_{\theta} \sum_{p \in \mathcal{L}} l(\mathbf{y}^p, \mathbf{s}_\theta^p) + \sum_{p,q \in \mathcal{L} \cup \mathcal{U}} w_{p,q} \|\mathbf{s}_\theta^p - \mathbf{s}_\theta^q\|$$

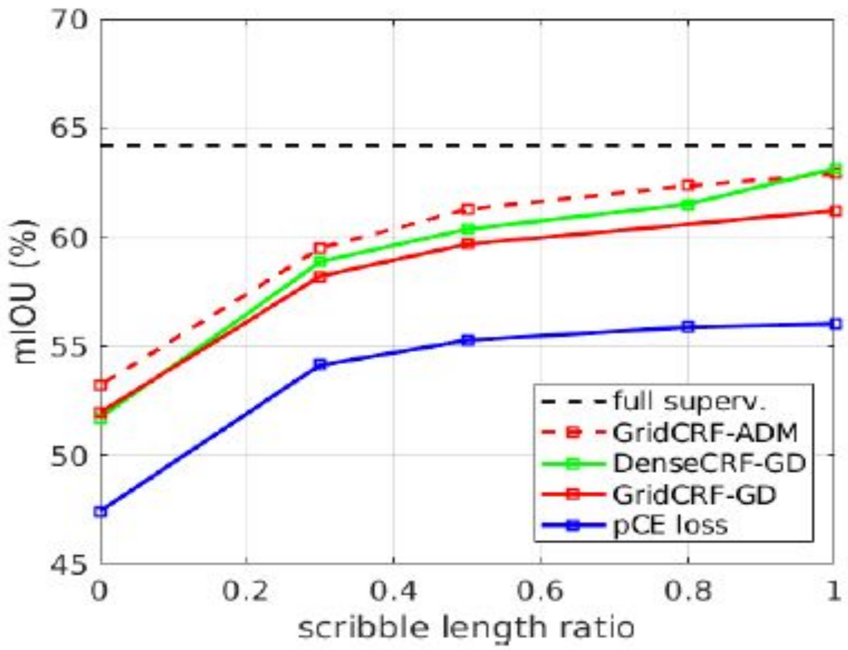


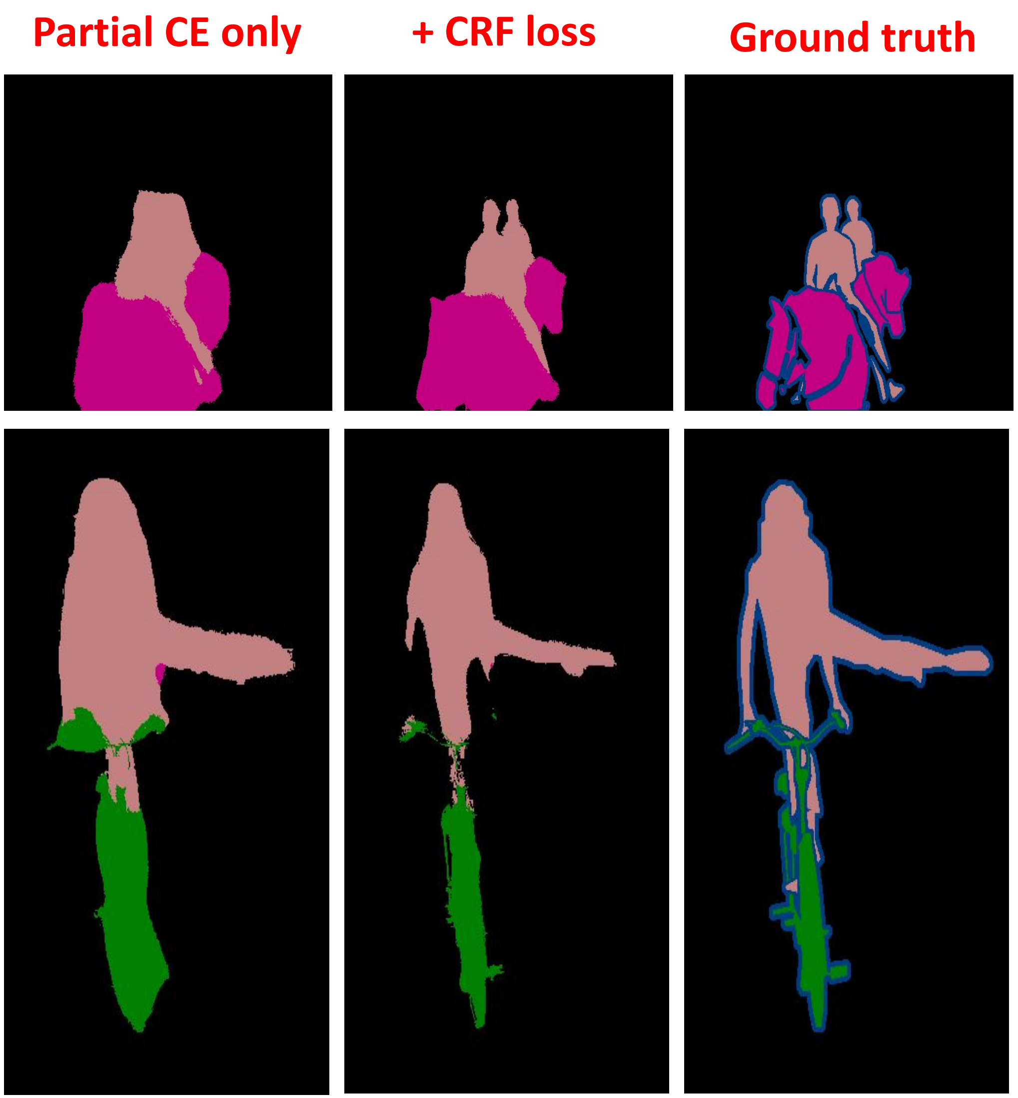
On the vertices of the simplex (binary variables), this is exactly the popular **Potts model in CRFs**

[Tang et al., On regularized losses for weakly supervised segmentation, ECCV 2018]

ÉTS
Le génie pour l'industrie

Semi-supervision loss in segmentation

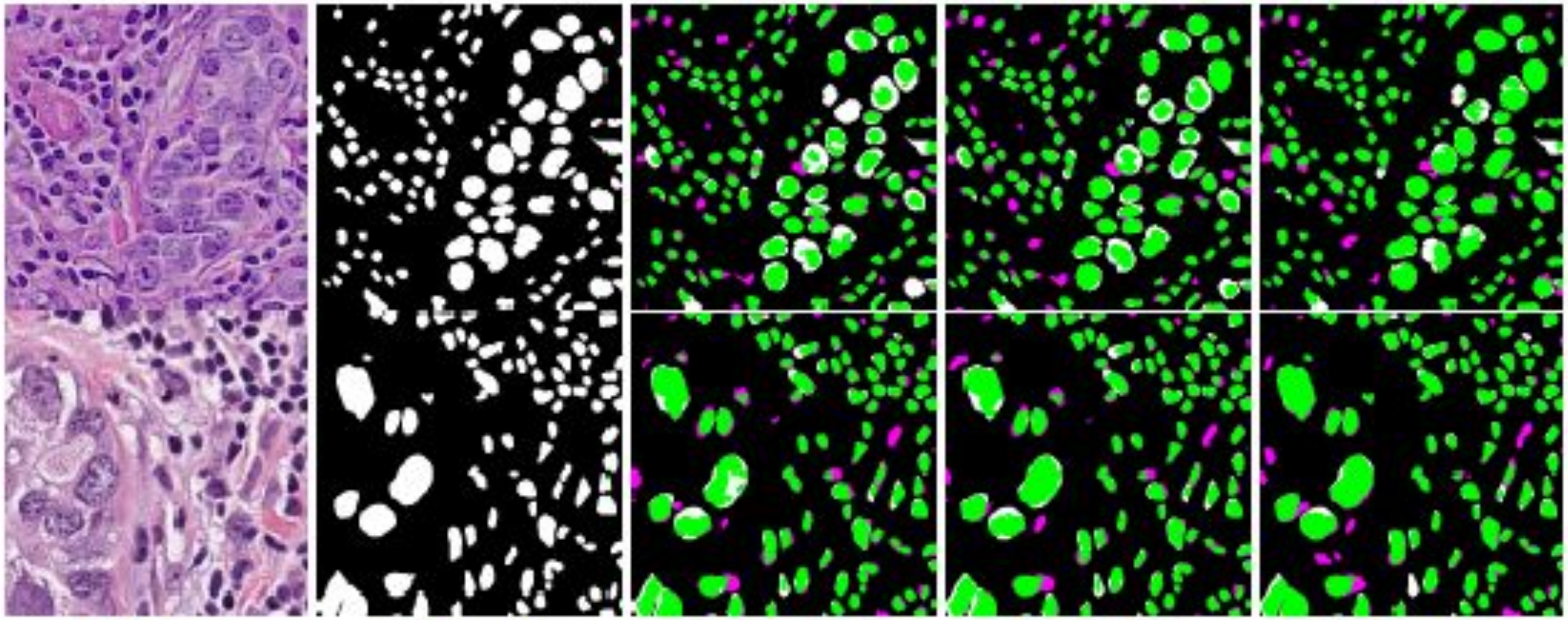




[Tang et al., On regularized losses for weakly supervised segmentation, ECCV 2018]

Quite used in medical imaging

White (FN); Magenta (FP); Green (TP)



(a) images (b) gt masks (c) without CRF (d) with CRF (e) full annotation

- Figures from Qu et al., Weakly Supervised Deep Nuclei Segmentation using Points Annotation in Histopathology Images, MIDL 2019 [Histology, point annotation]

- Ji et al., Scribble-Based Hierarchical Weakly Supervised Learning for Brain Tumor Segmentation, MICCAI 2019
[Brain tumor images, scribble annotations]

Unsupervised Entropy Regularization

Entropy minimization in SSL

$$\min_{\theta} - \sum_{p \in \mathcal{L}} \sum_{c=1}^{C} y^{p,c} \log s_{\theta}^{p,c} - \sum_{p \in \mathcal{U}} \sum_{c=1}^{C} s_{\theta}^{p,c} \log s_{\theta}^{p,c}$$

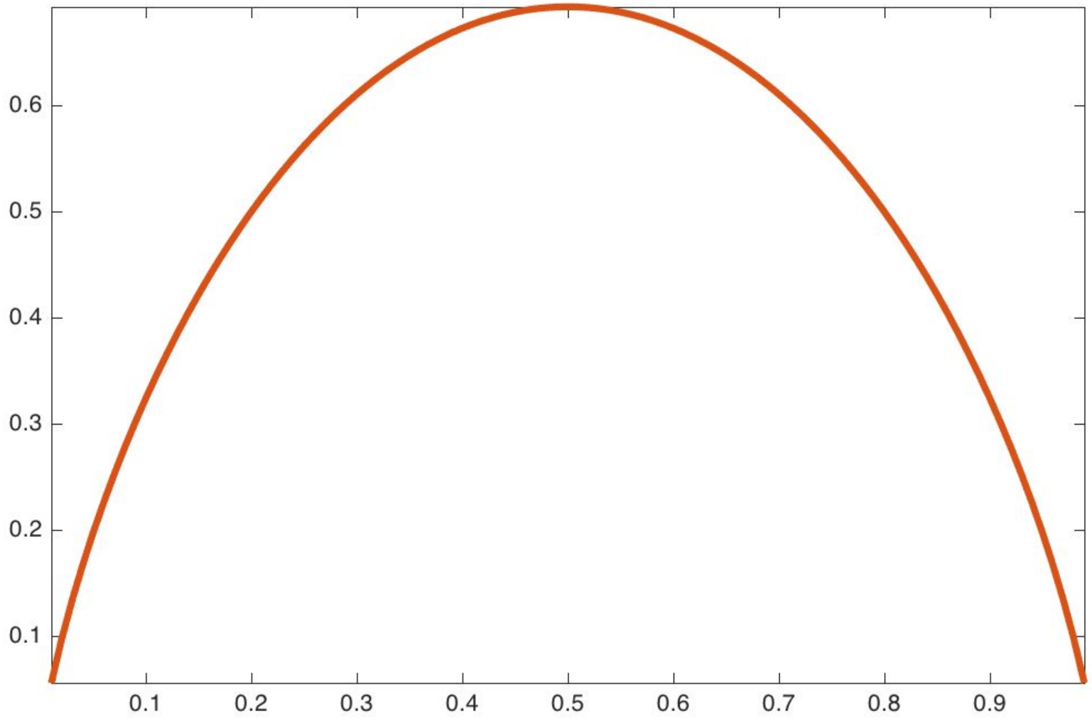
Shannon Entropies: “unsupervised cross-entropies (with unknown labels)”

- Grandvalet & Bengio, Semi-supervised learning by entropy minimization, NIPS 2005
- Gomes et al., Discriminative clustering by regularized information maximization, NIPS 2010

Effect of the Unsupervised Entropy (Why Is It Good for SSL)

It makes predictions confident (just like the cross-entropy)

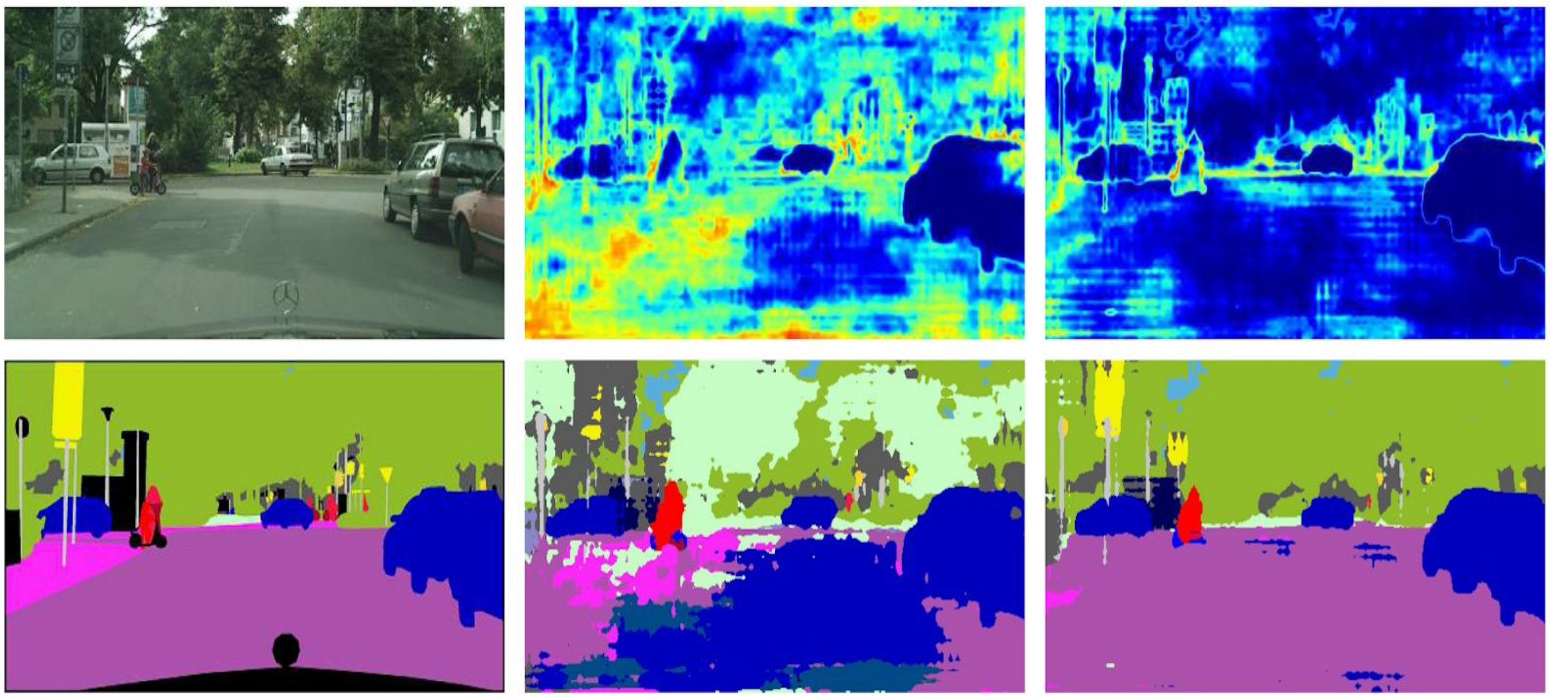
$$-s_\theta^p \log s_\theta^p - (1 - s_\theta^p) \log(1 - s_\theta^p)$$



Entropy Minimization in UDA

It makes predictions confident (just like the cross-entropy)

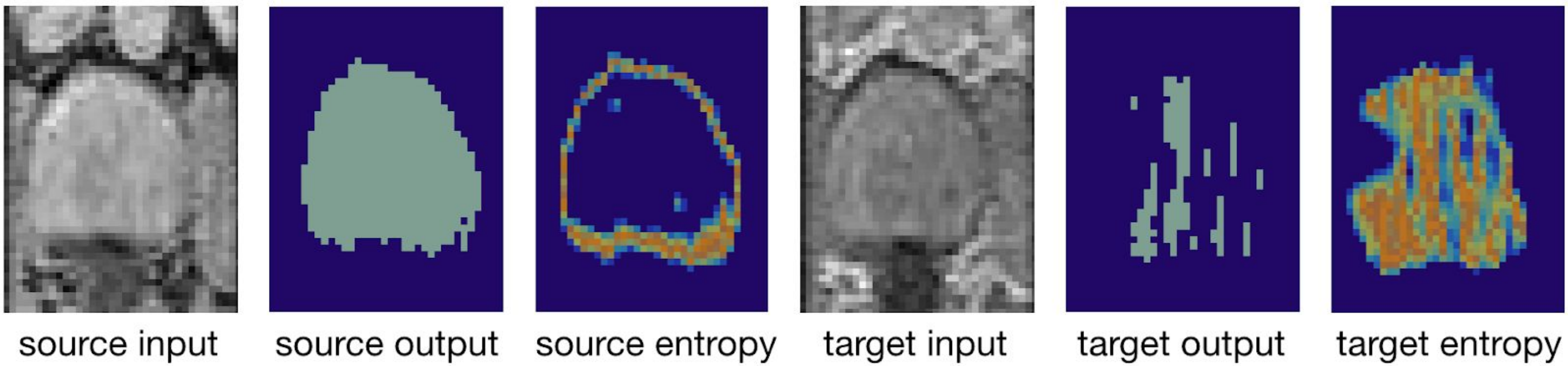
Input image + GT | Without adaptation | Entropy minimization



Images from Vu et al., ADVENT: Adversarial Entropy Minimization for Domain Adaptation in Semantic Segmentation, CVPR 2019

Entropy Minimization in UDA

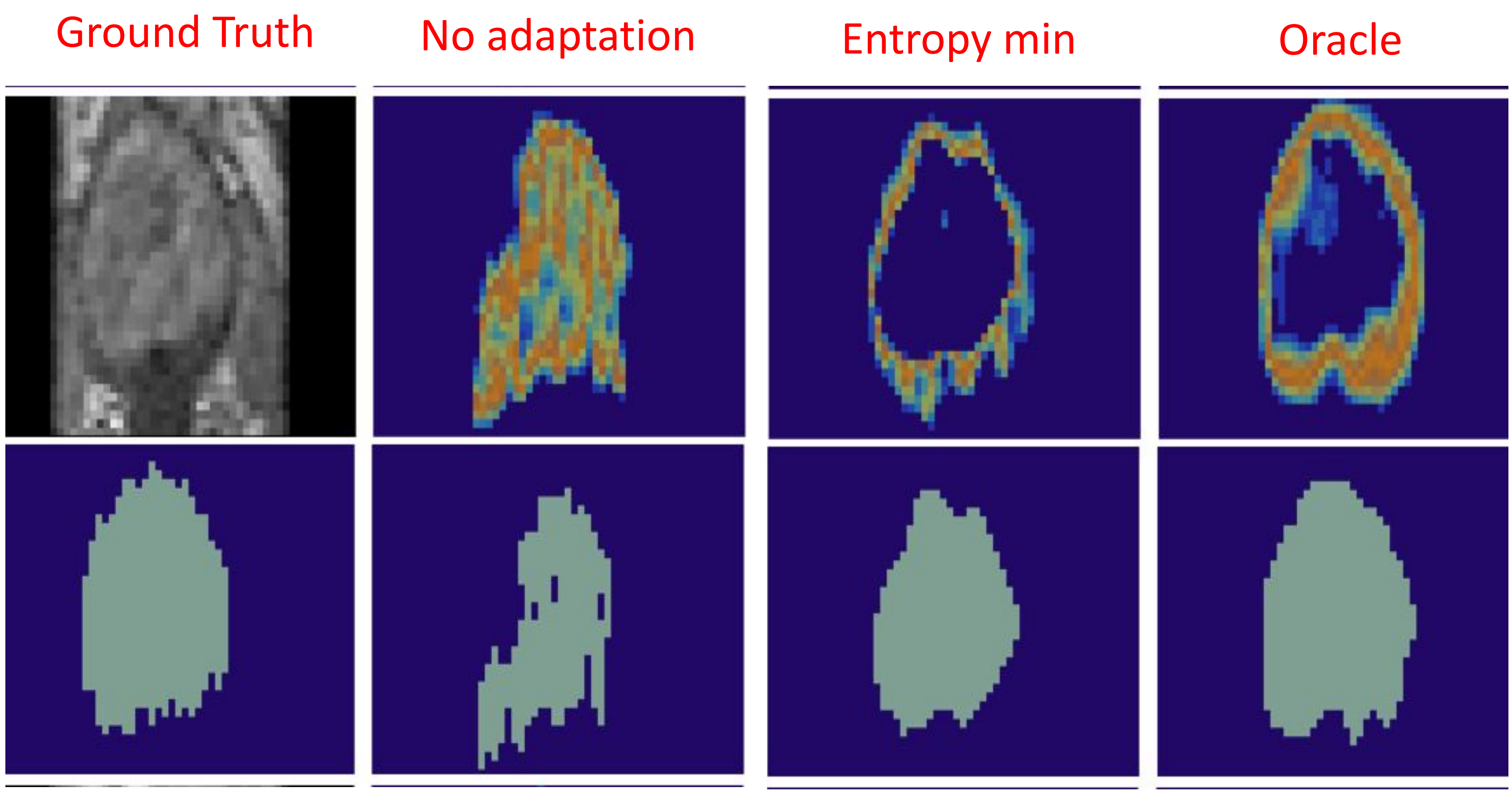
It makes predictions confident (just like the cross-entropy)



Images from Bateson et al., Source-relaxed domain adaptation for segmentation, MICCAI 2020

Entropy Minimization in UDA

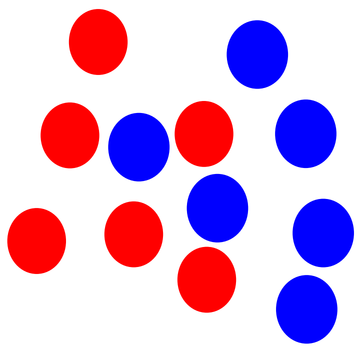
It makes predictions confident (just like the cross-entropy)



Images from Bateson et al., Source-relaxed domain adaptation for segmentation, MICCAI 2020

Why Entropy Minimization is good

It increases the margin between the classes



High entropy
(low confidence)

Low entropy
(high confidence)

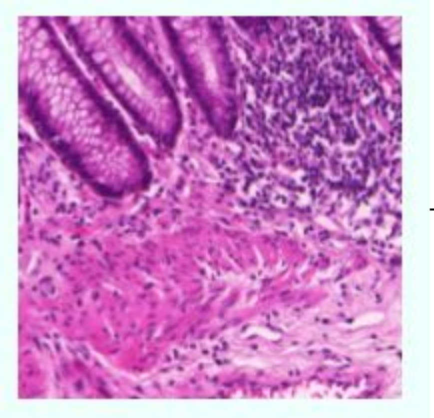
PAUSE (30 mins)

Part 4:
Applications of WSOL / WSSS

(a) **Medical Cancer Grading and ROI Localization in Histology**
(b) Weakly-Supervised Video Object Localization
(c) Person ReID: Embedding Networks
(d) Medical Semantic Segmentation

(a) Cancer Grading and Localization in Histology:

- **Task**



Model

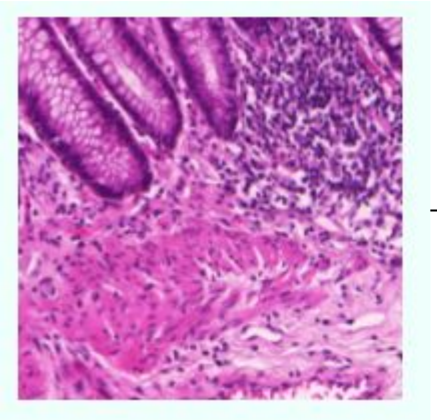
- Classification (grading)
- Localization of ROI at pixel-level (segmentation)

Input: histology image

Microscopic data for cancer diagnostic

(a) Cancer Grading and Localization in Histology:

- **Task**



Model

- Classification (grading)
- Localization of ROI at pixel-level (segmentation)

Input: histology image

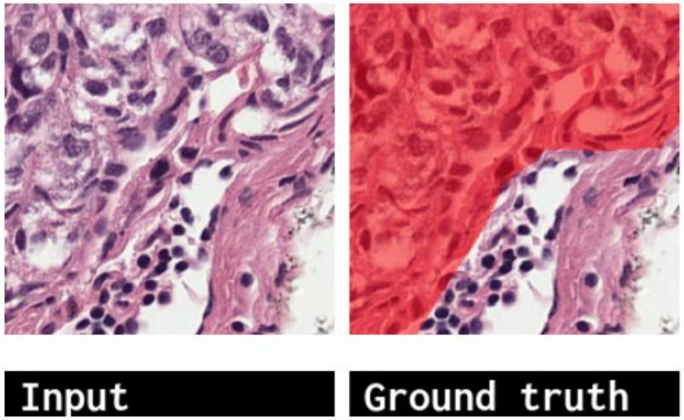
Microscopic data for cancer diagnostic

Available supervision: global image grad (class)

(a) Cancer Grading and Localization in Histology:

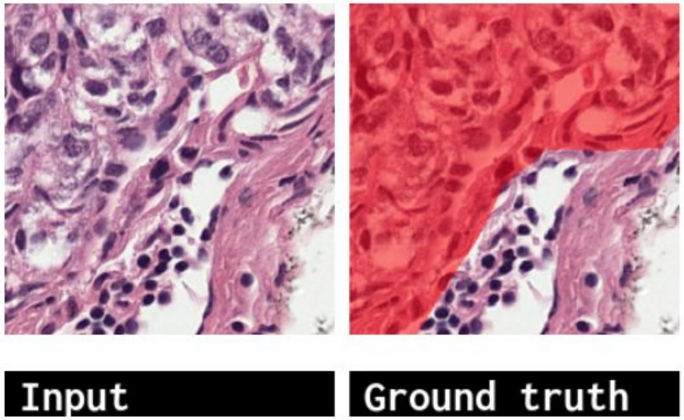
- **Histology data challenges**

  - Object non-salient (foreground is similar to background)



(a) Cancer Grading and Localization in Histology:

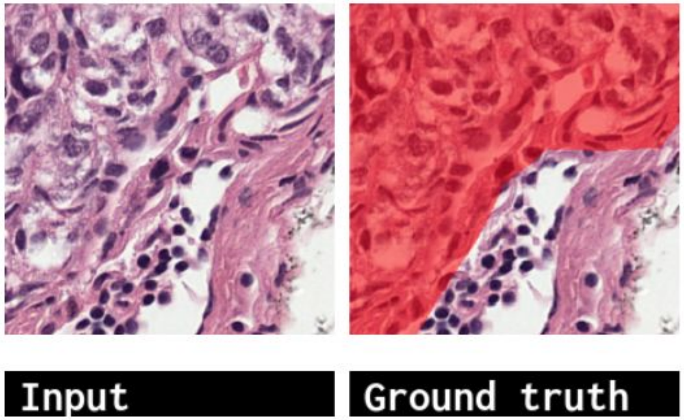
- **Histology data challenges**



  - Object non-salient (foreground is similar to background)
  - ROI with arbitrary shapes – no common global structure – opposite to natural scene images.

(a) Cancer Grading and Localization in Histology:

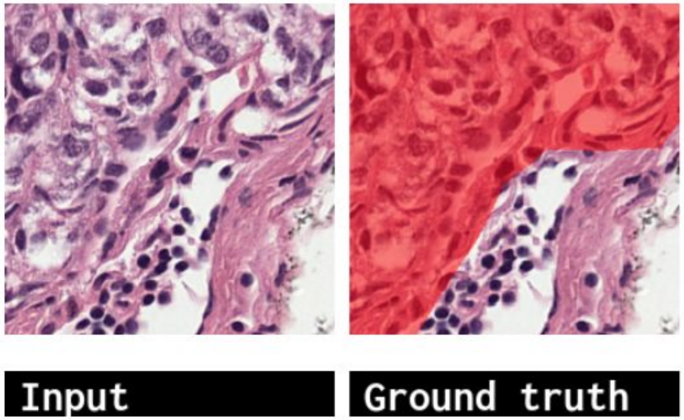
- **Histology data challenges**



- Object non-salient (foreground is similar to background)
- ROI with arbitrary shapes – no common global structure – opposite to natural scene images.
- Stain variation – Hematoxylin and Eosin (H&E)

(a) Cancer Grading and Localization in Histology:

- **Presented work**

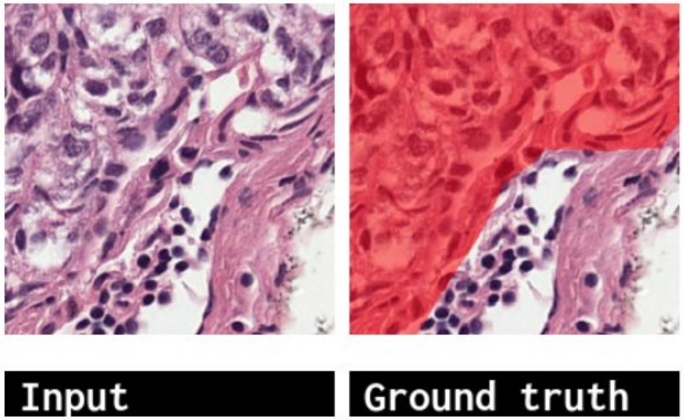


- Belharbi, S., Rony, J., Dolz, J., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022). **Deep interpretable classification and weakly-supervised segmentation of histology images via max-min uncertainty.** IEEE Transactions on Medical Imaging, 41:702–714.

- Belharbi, S., Pedersoli, M., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022). **Negative evidence matters in interpretable histology image classification**. In Medical Imaging with Deep Learning (MIDL).

(a) Cancer Grading and Localization in Histology:

- **Presented work**



- Belharbi, S., Rony, J., Dolz, J., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022). **Deep interpretable classification and weakly-supervised segmentation of histology images via max-min uncertainty.** IEEE Transactions on Medical Imaging, 41:702–714.

- Belharbi, S., Pedersoli, M., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022). **Negative evidence matters in interpretable histology image classification**. In Medical Imaging with Deep Learning (MIDL).

(a) Cancer Grading and Localization in Histology:

Deep interpretable classification and weakly-supervised segmentation of histology images via max-min uncertainty

Transactions on Medical Imaging, 2022

Code:

https://github.com/sbelharbi/deep-wsl-histo-min-max-uncertainty

Belharbi, S., Rony, J., Dolz, J., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022b). Deep interpretable classification and weakly-supervised segmentation of histology images via max-min uncertainty. IEEE Transactions on Medical Imaging, 41:702–714.

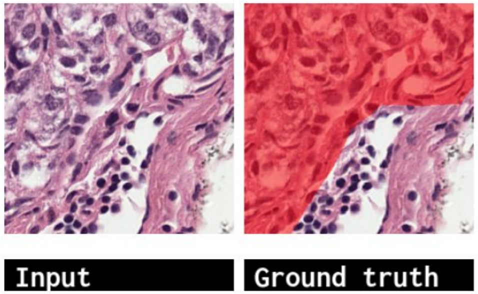
Le génie pour l'industrie

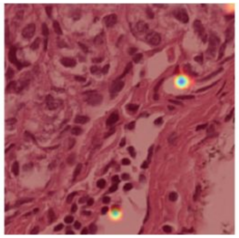
(a) Cancer Grading and Localization in Histology:

- **Issue**

  - Non-salient object → visual similarity between foreground/background

    High false positives/negatives





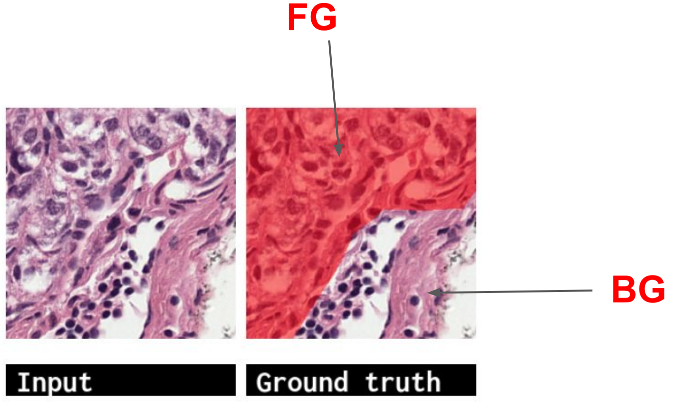
Le génie pour l'industrie

Belharbi, S., Rony, J., Dolz, J., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022b). Deep interpretable classification and weakly-supervised segmentation of histology images via max-min uncertainty. IEEE Transactions on Medical Imaging, 41:702–714.

(a) Cancer Grading and Localization in Histology:

- **Constrain CAMs**

  - Explicit modeling foreground/background map



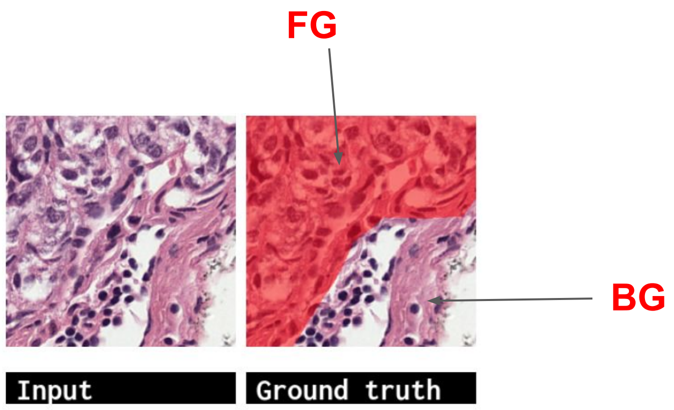
Belharbi, S., Rony, J., Dolz, J., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022b). Deep interpretable classification and weakly-supervised segmentation of histology images via max-min uncertainty. IEEE Transactions on Medical Imaging, 41:702–714.

ÉTS
Le génie pour l'industrie

(a) Cancer Grading and Localization in Histology:

- **Constrain CAMs**

  - Explicit modeling foreground/background map
  - Constrain the presence of both FG / BG using size constraints

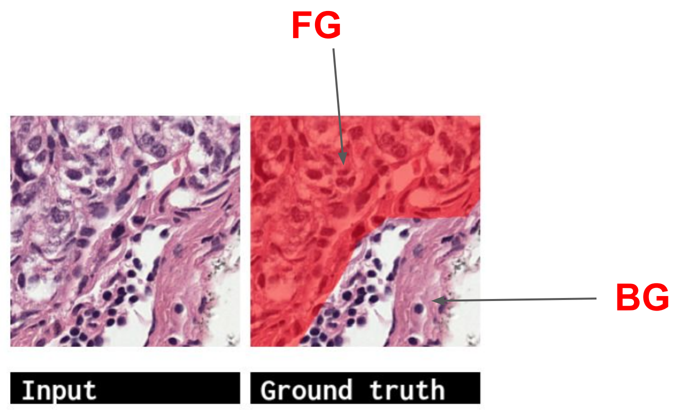


Belharbi, S., Rony, J., Dolz, J., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022b). Deep interpretable classification and weakly-supervised segmentation of histology images via max-min uncertainty. IEEE Transactions on Medical Imaging, 41:702–714.

(a) Cancer Grading and Localization in Histology:

- **Constrain CAMs**

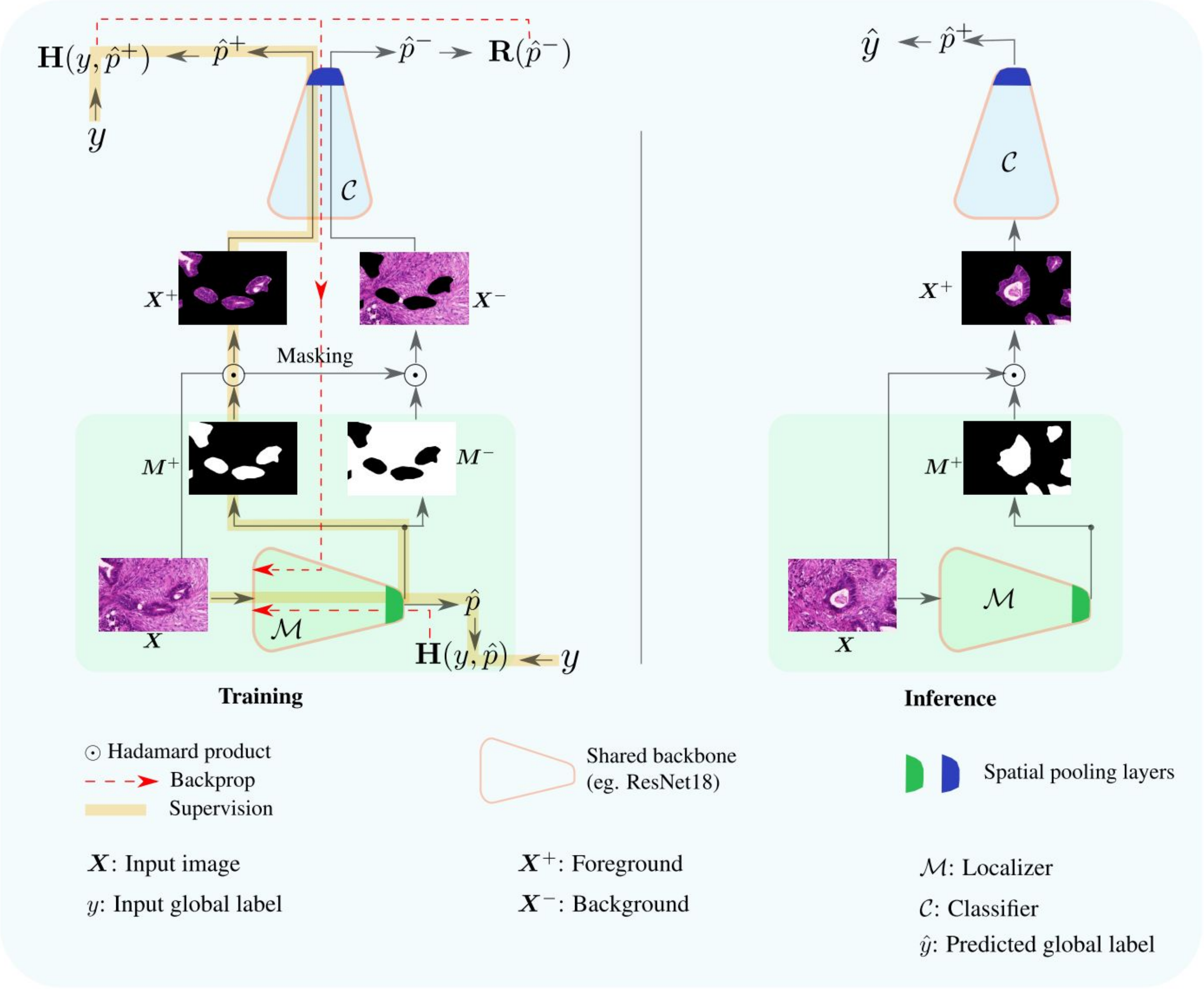
  - Explicit modeling foreground/background map
  - Constrain the presence of both FG / BG using size constraints
  - Ensure that each map is consistent using classifier response.



Belharbi, S., Rony, J., Dolz, J., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022b). Deep interpretable classification and weakly-supervised segmentation of histology images via max-min uncertainty. IEEE Transactions on Medical Imaging, 41:702–714.

(a) Cancer Grading and Localization in Histology:

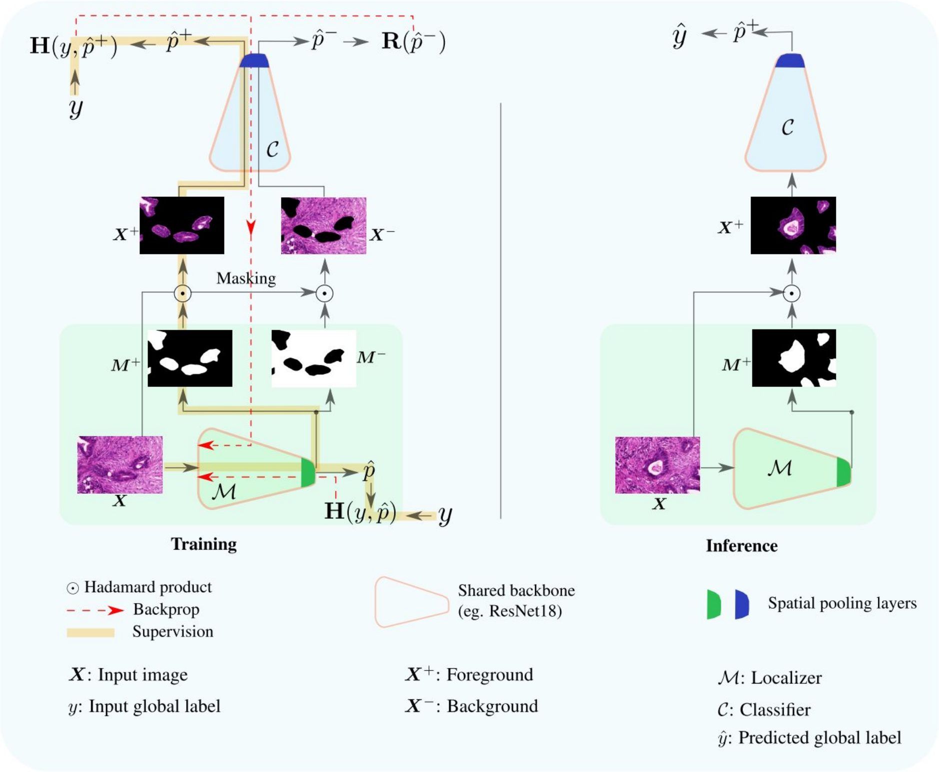
- **Our architecture**



Belharbi, S., Rony, J., Dolz, J., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022b). Deep interpretable classification and weakly-supervised segmentation of histology images via max-min uncertainty. IEEE Transactions on Medical Imaging, 41:702–714.

(a) Cancer Grading and Localization in Histology:

- **Training loss**



$$\min_{\boldsymbol{\theta}_\mathcal{C}} \quad \mathbf{H}(p, \hat{p}^+) + \lambda\, \mathbf{R}(\hat{p}^-) - \frac{1}{t}\left[\log \boldsymbol{s}^+ + \log \boldsymbol{s}^-\right],$$

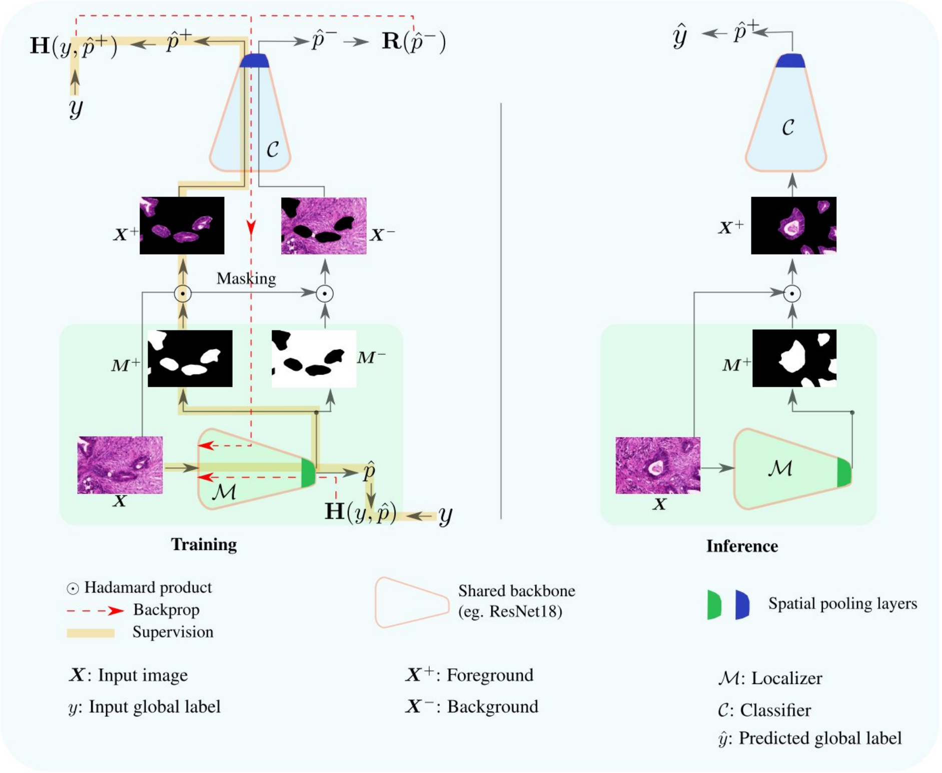
Maximize classifier
Response over FG

Belharbi, S., Rony, J., Dolz, J., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022b). Deep interpretable classification and weakly-supervised segmentation of histology images via max-min uncertainty. IEEE Transactions on Medical Imaging, 41:702–714.

ÉTS
Le génie pour l'industrie

(a) Cancer Grading and Localization in Histology:

- **Training loss**



$$\min_{\boldsymbol{\theta}_{\mathcal{C}}} \quad \mathbf{H}(p, \hat{p}^{+}) + \lambda\, \mathbf{R}(\hat{p}^{-}) - \frac{1}{t}\left[\log \boldsymbol{s}^{+} + \log \boldsymbol{s}^{-}\right],$$

The BG has no discriminative regions left.
Max uncertainty

$$\mathbf{R}(\hat{p}^{-}) = -\mathbf{H}(\hat{p}^{-}); \quad \text{or} \quad \mathbf{R}(\hat{p}^{-}) = \mathbf{H}(q, \hat{p}^{-}),$$

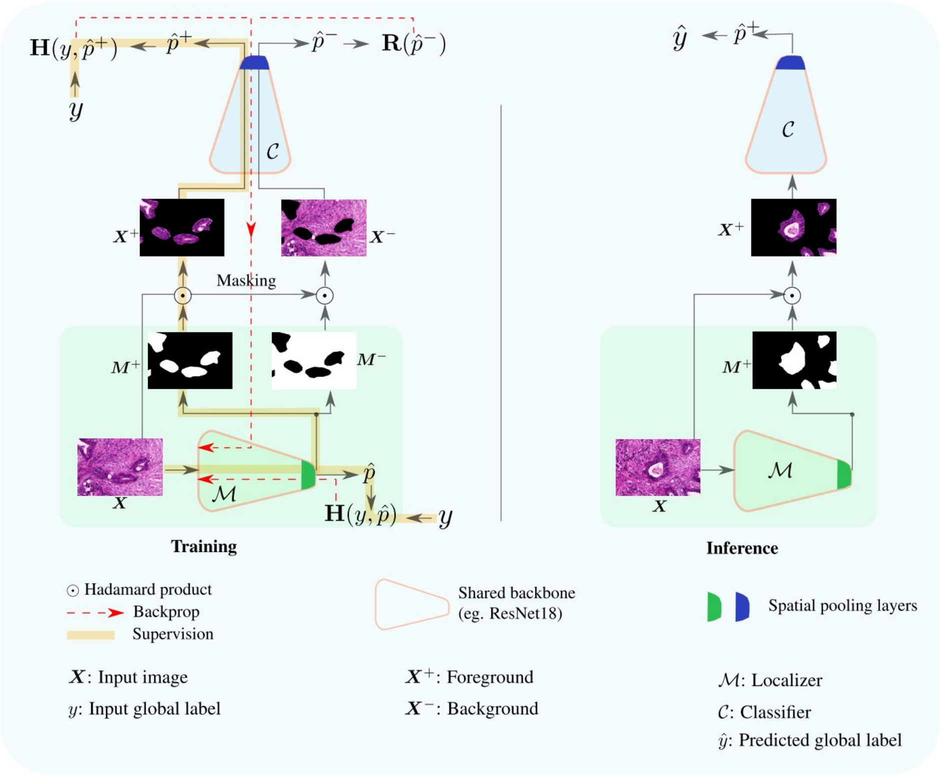
Explicit Entropy Maximization (**EEM**)

Surrogate for explicit Entropy Maximization (**SEM**). q: uniform dist.

Belharbi, S., Rony, J., Dolz, J., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022b). Deep interpretable classification and weakly-supervised segmentation of histology images via max-min uncertainty. IEEE Transactions on Medical Imaging, 41:702–714.

(a) Cancer Grading and Localization in Histology:

- **Training loss**



ASC: Absolute Size Constraint

$$\min_{\boldsymbol{\theta}_{\mathcal{C}}} \quad \mathbf{H}(p, \hat{p}^+) + \lambda\, \mathbf{R}(\hat{p}^-) - \frac{1}{t}\left[\log \boldsymbol{s}^+ + \log \boldsymbol{s}^-\right],$$

Ensure both FG/BG are present: max size.

$$\boldsymbol{s}^+ = \textstyle\sum_{z\in\Omega} \boldsymbol{M}^+(z)\,, \quad \boldsymbol{s}^- = \textstyle\sum_{z\in\Omega} \boldsymbol{M}^-(z)$$

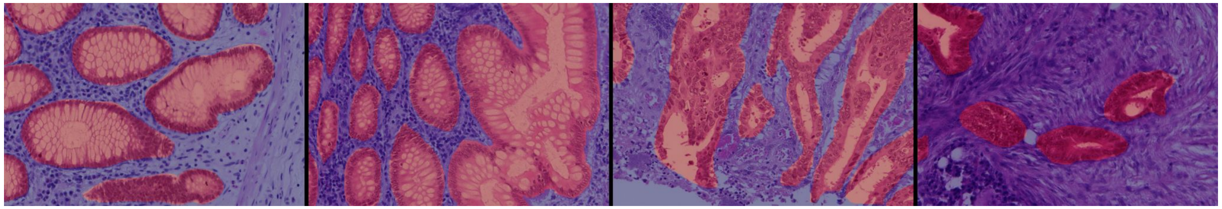
Sizes

Belharbi, S., Rony, J., Dolz, J., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022b). Deep interpretable classification and weakly-supervised segmentation of histology images via max-min uncertainty. IEEE Transactions on Medical Imaging, 41:702–714.

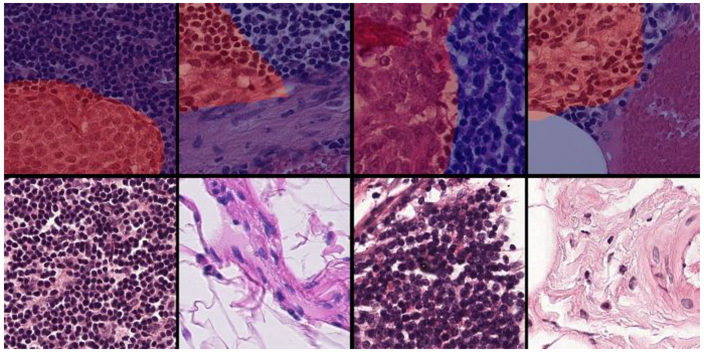
(a) Cancer Grading and Localization in Histology:

- **Experiments**

  - Task: classify and localize ROI
  - 2 public datsets: GlaS, Camelyon16 patches.



GlaS: colon cancer diagnosis



Camelyon16 patches: breast cancer

Belharbi, S., Rony, J., Dolz, J., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022b). Deep interpretable classification and weakly-supervised segmentation of histology images via max-min uncertainty. IEEE Transactions on Medical Imaging, 41:702–714.

ÉTS Le génie pour l'industrie

(a) Cancer Grading and Localization in Histology:

- **Results**

Method	Image level Cl. error (%)	Pixel level F1$^+$ (%)	F1$^-$ (%)
All-ones (Lower-bound)	−−	66.01	00.00
PN [39]	−−	65.52	24.08
ERASE [80]	7.50	65.60	25.01
CAM-Max [52]	1.25	66.00	26.32
CAM-LSE [57, 74]	1.25	66.05	27.93
Grad-CAM [64]	**0.00**	66.30	21.30
CAM-Avg [88]	**0.00**	66.90	17.88
Wildcat [20]	1.25	67.21	22.96
Deep MIL [33]	**2.50**	68.52	41.34
Ours (EEM)	**0.00**	**72.11**	69.07
Ours (SEM)	**0.00**	71.94	**69.23**
U-Net [60] (Upper-bound)	−−	90.19	88.52

GlaS

Method	Image level Cl. error (%)	Pixel level F1$^+$ (%)	F1$^-$ (%)
All-ones (Lower-bound)	−−	59.44	00.00
PN [39]	−−	31.15	37.36
ERASE [80]	8.61	31.30	42.48
CAM-Max [52]	10.06	48.28	81.92
CAM-LSE [57, 74]	1.51	64.31	63.78
Grad-CAM [64]	2.40	62.78	79.05
CAM-Avg [88]	2.40	62.75	79.05
Wildcat [20]	**1.48**	62.73	72.59
Deep MIL [33]	1.93	59.01	36.94
Ours (EEM)	6.26	67.98	**88.80**
Ours (SEM)	6.95	**68.26**	88.55
U-Net [60] (Upper-bound)	−−	71.11	89.68

Camelyon16

Belharbi, S., Rony, J., Dolz, J., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022b). Deep interpretable classification and weakly-supervised segmentation of histology images via max-min uncertainty. IEEE Transactions on Medical Imaging, 41:702–714.

(a) Cancer Grading and Localization in Histology:

- **Visual results**

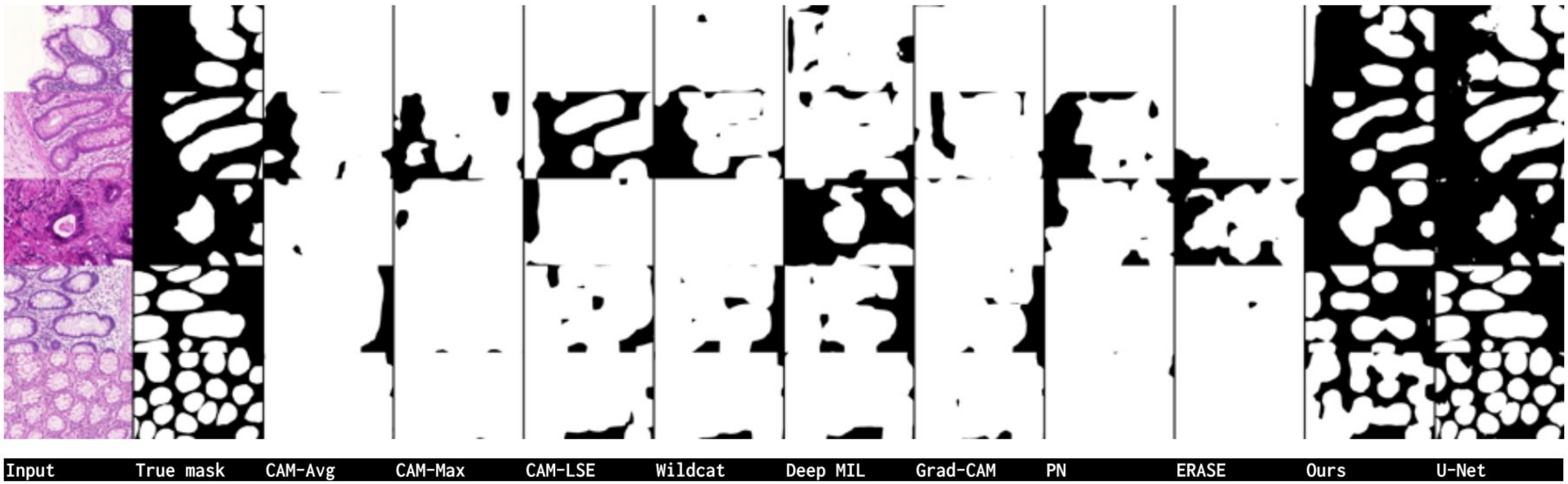


Figure 2: **GlaS dataset**: Qualitative results of the predicted binary mask for each method on several GlaS test images. Our method, referred to as *Ours*, is the SEM version with the ASC regularization term. (Best visualized in color.)

Belharbi, S., Rony, J., Dolz, J., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022b). Deep interpretable classification and weakly-supervised segmentation of histology images via max-min uncertainty. IEEE Transactions on Medical Imaging, 41:702–714.

Le génie pour l'industrie

(a) Cancer Grading and Localization in Histology:

- **Visual results**

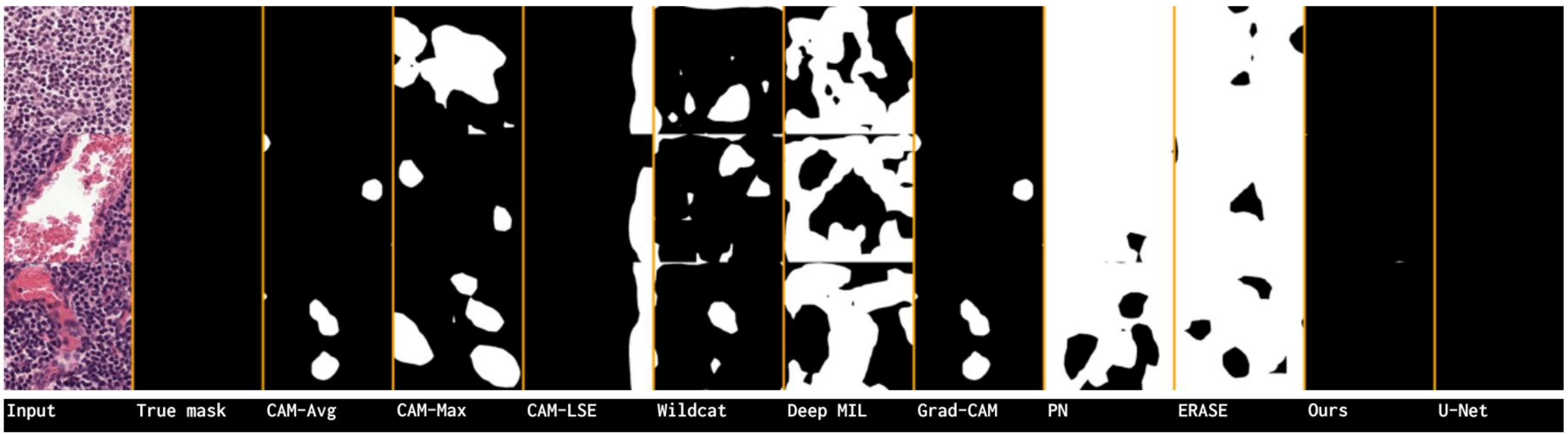


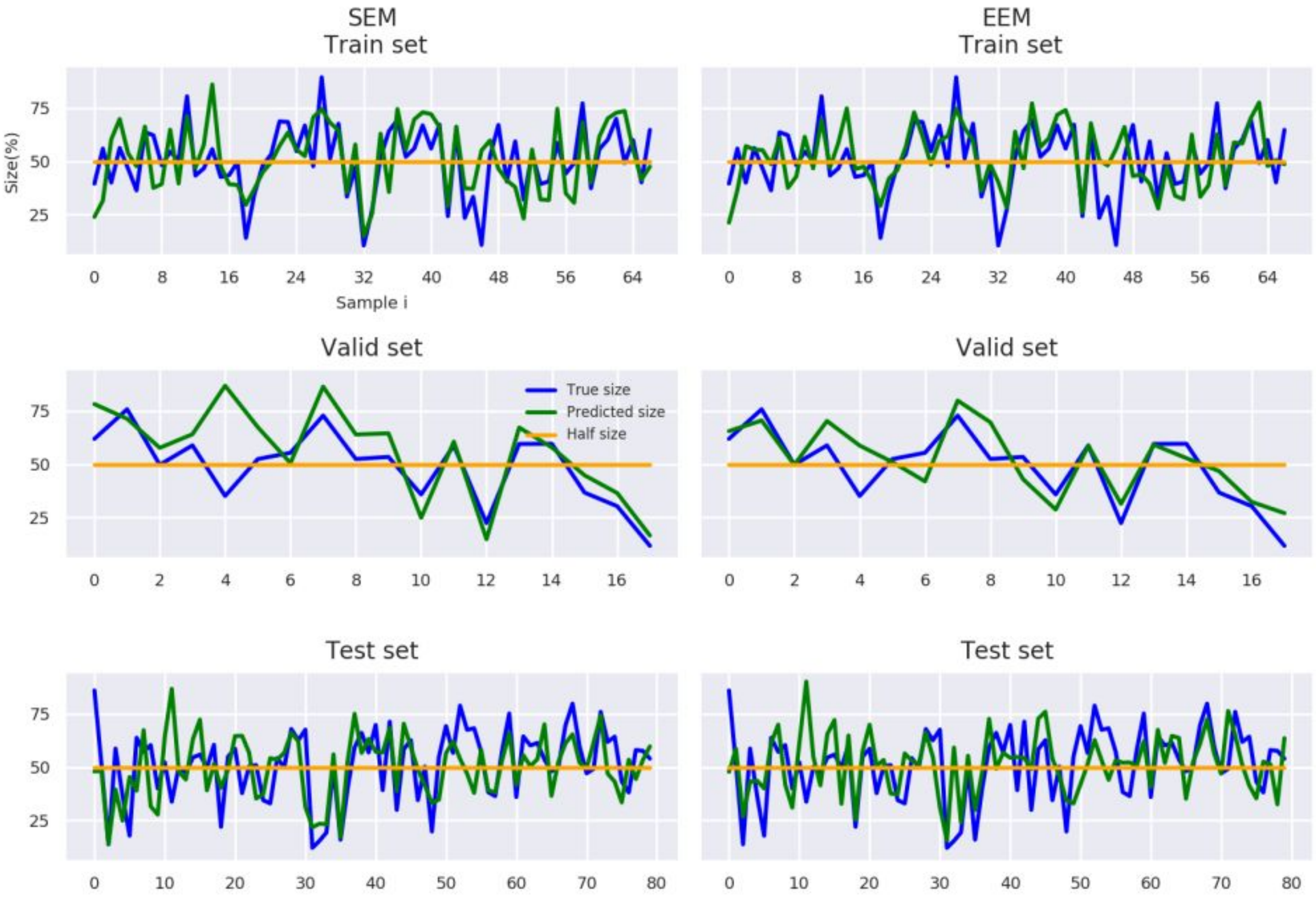
Figure 3: **Camelyon16-P512 benchmark**: Examples of mask predictions over **normal** samples from the testing set. White pixels indicate metastatic regions, while black pixels indicate normal tissue. This illustrates false positives. Note that normal samples do not contain any metastatic regions. Ours is SEM version with the ASC regularization. (Best visualized in color.)

Belharbi, S., Rony, J., Dolz, J., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022b). Deep interpretable classification and weakly-supervised segmentation of histology images via max-min uncertainty. IEEE Transactions on Medical Imaging, 41:702–714.

ÉTS Le génie pour l'industrie

(a) Cancer Grading and Localization in Histology:

- **Size constraint**



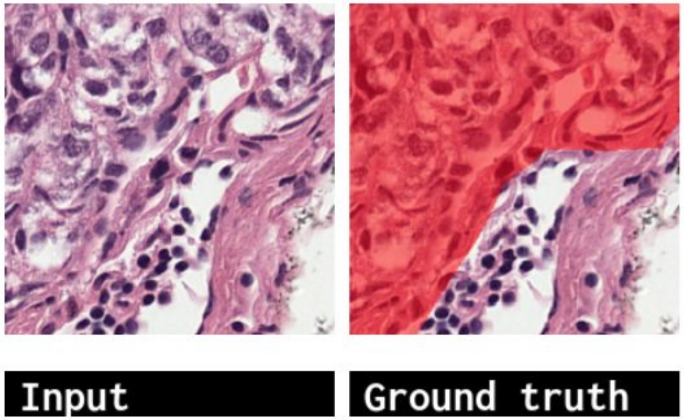
$$\min_{\boldsymbol{\theta}_C} \quad \mathbf{H}(p, \hat{p}^{+}) + \lambda\, \mathbf{R}(\hat{p}^{-}) - \frac{1}{t}\left[\log \boldsymbol{s}^{+} + \log \boldsymbol{s}^{-}\right],$$

Belharbi, S., Rony, J., Dolz, J., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022b). Deep interpretable classification and weakly-supervised segmentation of histology images via max-min uncertainty. IEEE Transactions on Medical Imaging, 41:702–714.

ÉTS Le génie pour l'industrie

(a) Cancer Grading and Localization in Histology:

- **Presented work**



- Belharbi, S., Rony, J., Dolz, J., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022). **Deep interpretable classification and weakly-supervised segmentation of histology images via max-min uncertainty.** IEEE Transactions on Medical Imaging, 41:702–714.

- Belharbi, S., Pedersoli, M., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022). **Negative evidence matters in interpretable histology image classification**. In Medical Imaging with Deep Learning (MIDL).

(a) Cancer Grading and Localization in Histology:

Negative evidence matters in interpretable histology image classification

Medical Imaging with Deep Learning (MIDL), 2022

Code:

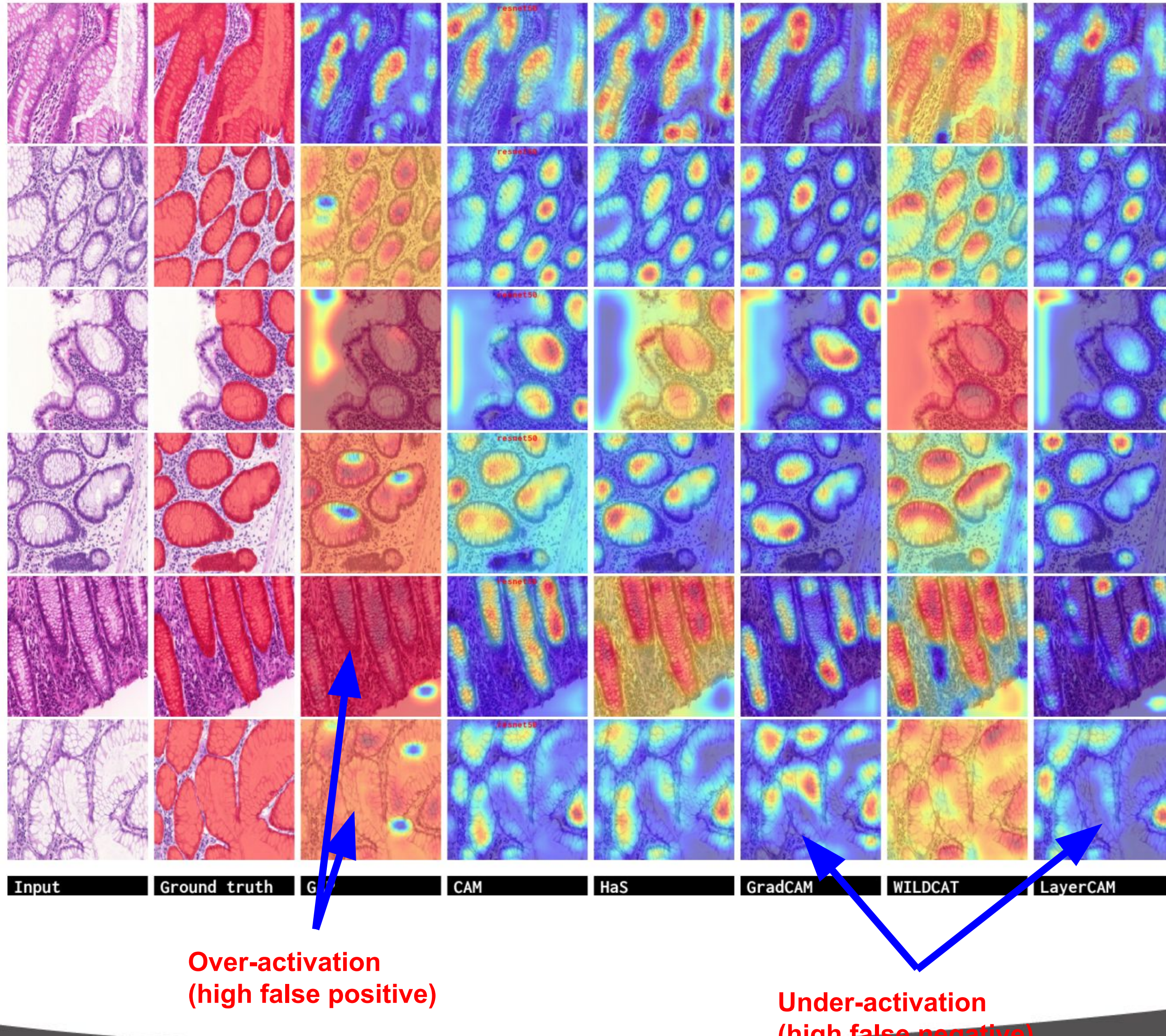
https://github.com/sbelharbi/negev

Belharbi, S., Pedersoli, M., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022). Negative evidence matters in interpretable histology image classification. In Medical Imaging with Deep Learning (MIDL).

Le génie pour l'industrie

(a) Cancer Grading and Localization in Histology:

- **CAMs' challenges in histology images**



Belharbi, S., Pedersoli, M., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022). Negative evidence matters in interpretable histology image classification. In Medical Imaging with Deep Learning (MIDL).

Le génie pour l'industrie

(a) Cancer Grading and Localization in Histology:

- **Using negative knowledge**

  - To reduce mis-predictions, guide the CAM learning with available Negative knowledge.

  Negative knowledge = all what is not ROI.

Belharbi, S., Pedersoli, M., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022). Negative evidence matters in interpretable histology image classification. In Medical Imaging with Deep Learning (MIDL).

(a) Cancer Grading and Localization in Histology:

- **Using negative knowledge**

  - 2 sources of negative knowledge

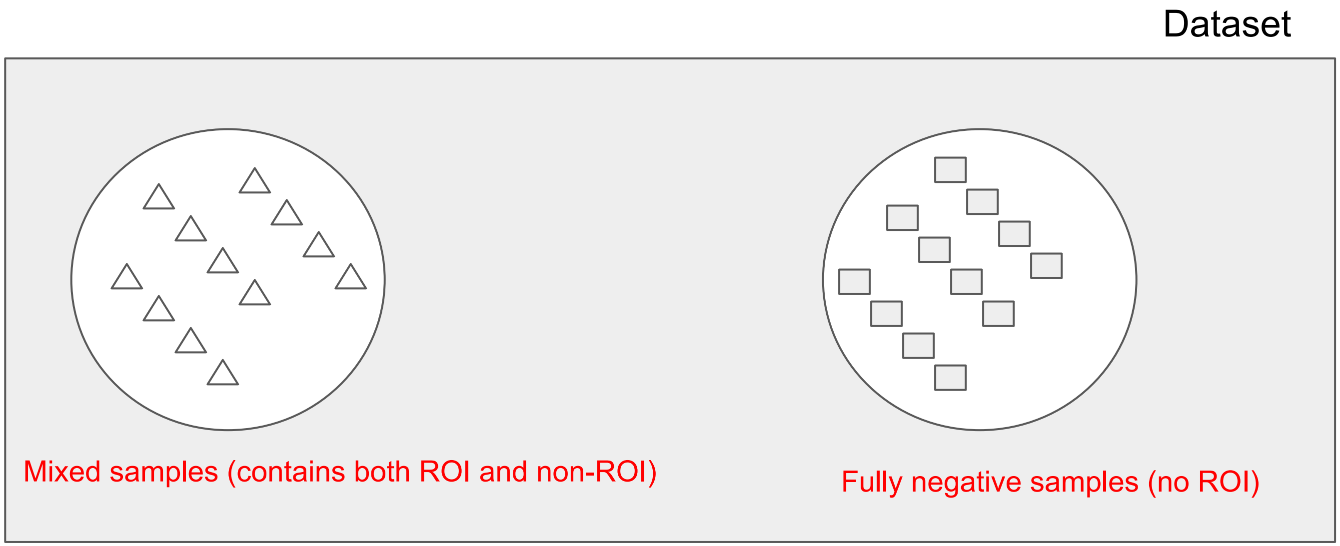
Belharbi, S., Pedersoli, M., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022). Negative evidence matters in interpretable histology image classification. In Medical Imaging with Deep Learning (MIDL).

(a) Cancer Grading and Localization in Histology:

- **Using negative knowledge**

  - 2 sources of negative knowledge

1 - Naturally occurring in dataset



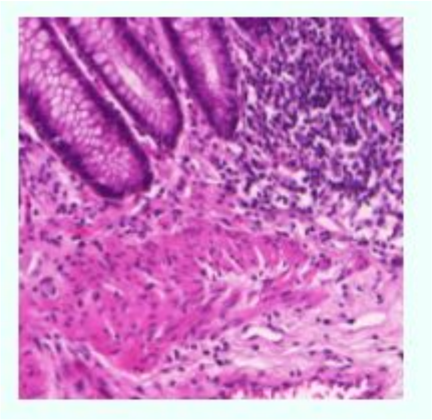
Belharbi, S., Pedersoli, M., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022). Negative evidence matters in interpretable histology image classification. In Medical Imaging with Deep Learning (MIDL).

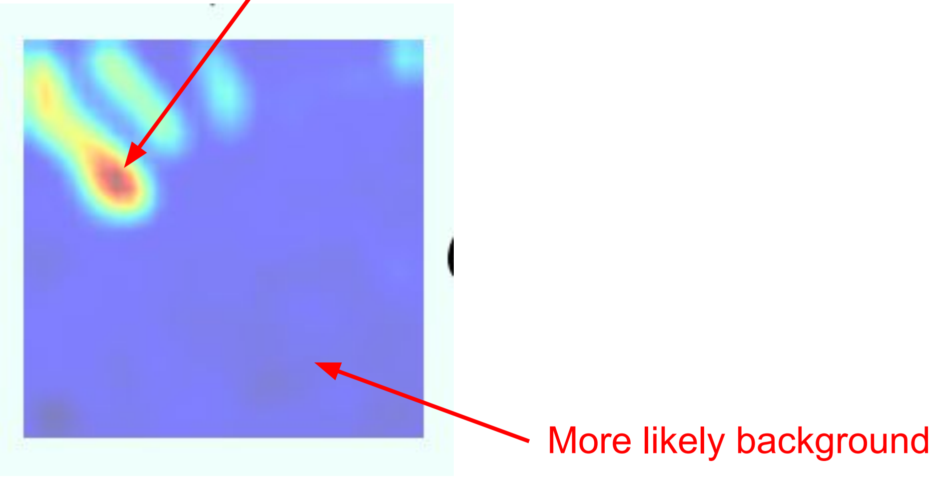
(a) Cancer Grading and Localization in Histology:

- **Using negative knowledge**

  - 2 sources of negative knowledge

2 - Low activation in CAMs



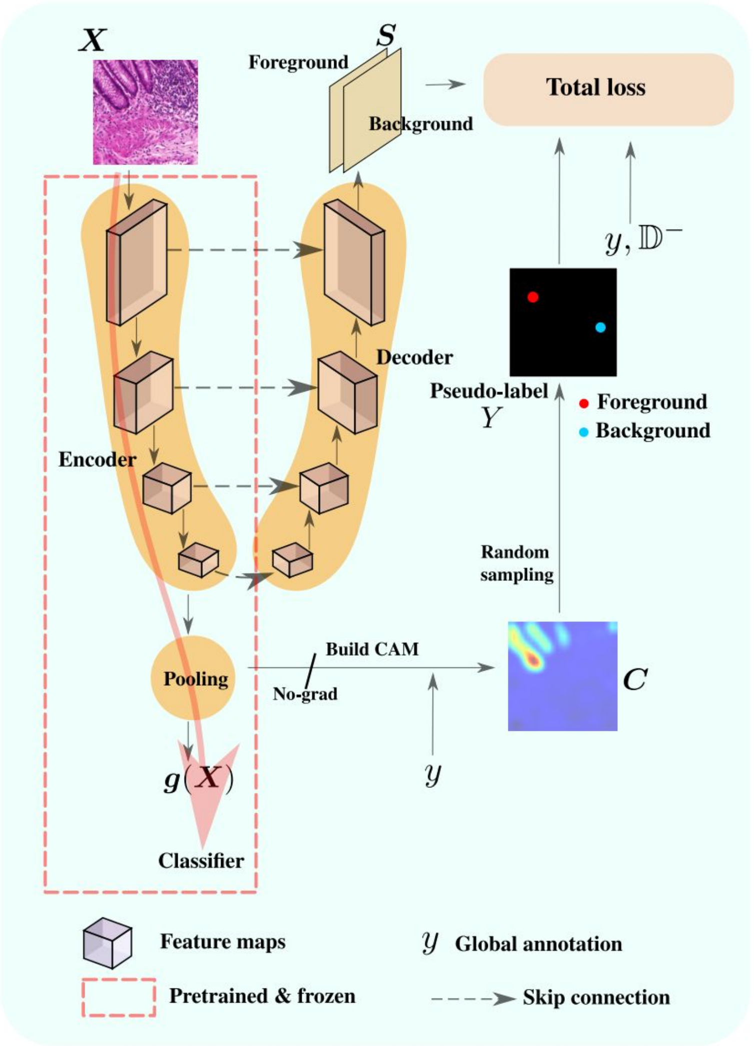


Belharbi, S., Pedersoli, M., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022). Negative evidence matters in interpretable histology image classification. In Medical Imaging with Deep Learning (MIDL).

(a) Cancer Grading and Localization in Histology:

- **Architecture**

- Requires only image class for training



Le génie pour l'industrie

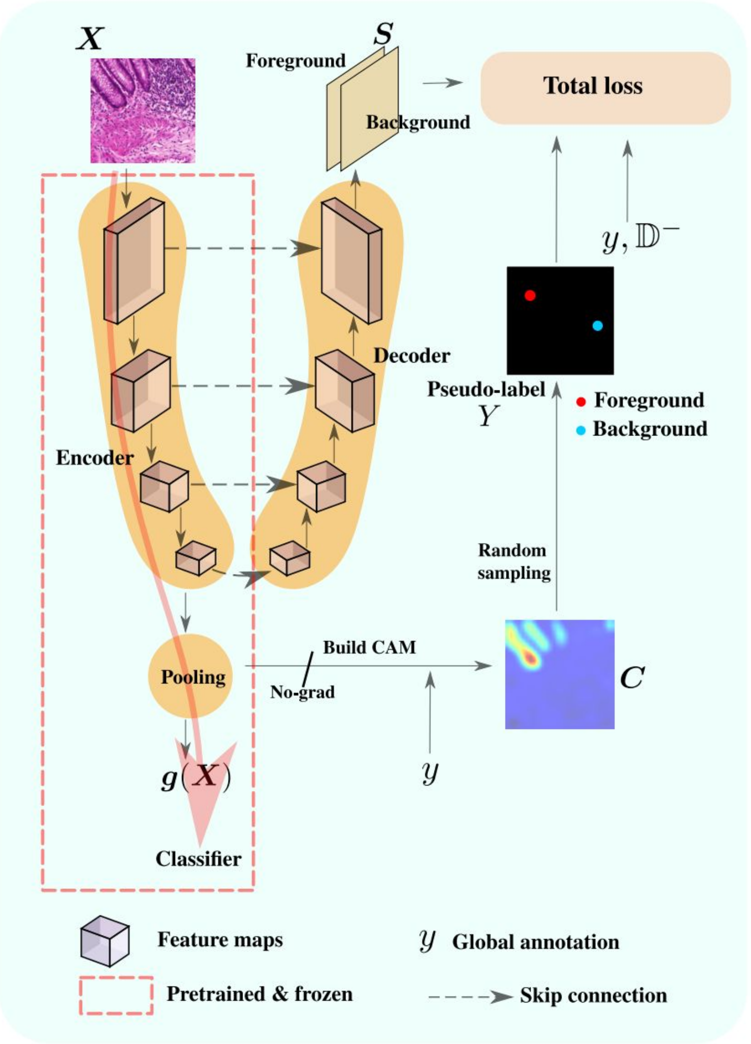
Belharbi, S., Pedersoli, M., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022). Negative evidence matters in interpretable histology image classification. In Medical Imaging with Deep Learning (MIDL).

(a) Cancer Grading and Localization in Histology:

- **Training**

1- Exploit CAM positive/negative information

$$\min_{\boldsymbol{\theta}} \sum_{p \in \{\mathbb{C}^+ \cup \mathbb{C}^-\}} \boldsymbol{H}(Y_p, \boldsymbol{S}_p) \; .$$



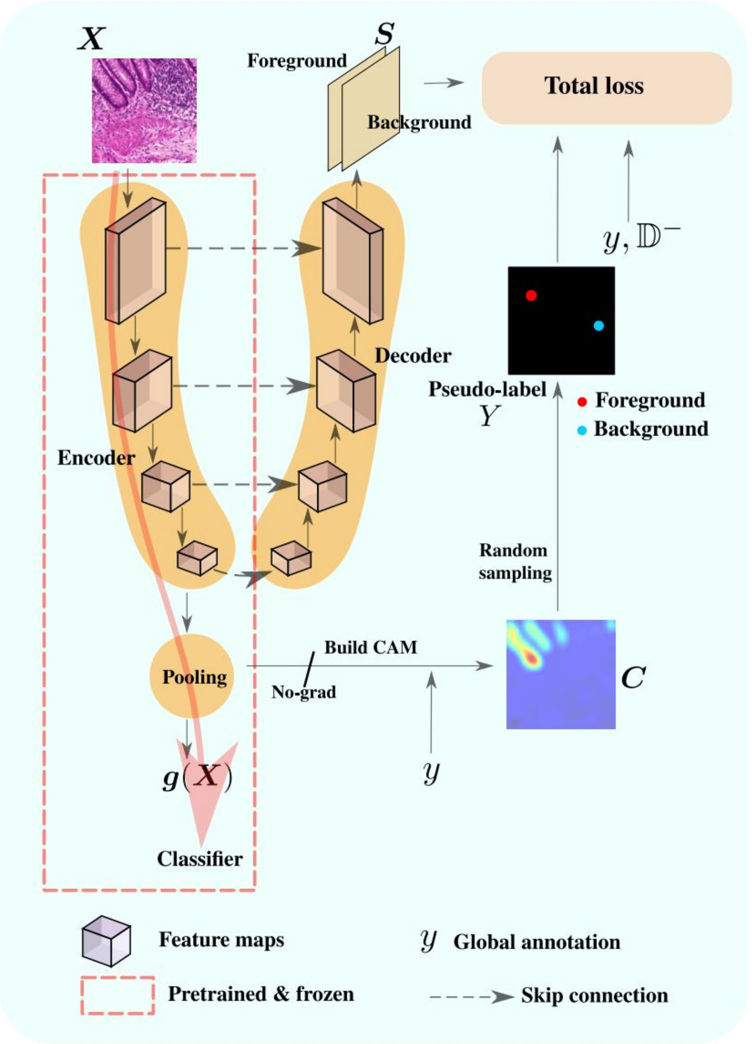
Belharbi, S., Pedersoli, M., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022). Negative evidence matters in interpretable histology image classification. In Medical Imaging with Deep Learning (MIDL).

(a) Cancer Grading and Localization in Histology:

- **Training**

2- Fully negative samples

$$\min_{\boldsymbol{\theta}} \quad \sum_{p \in \Omega} -\log(1 - \boldsymbol{S}^0_p) \ , \forall \boldsymbol{X} \in \mathbb{D}^- \ .$$



Belharbi, S., Pedersoli, M., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022). Negative evidence matters in interpretable histology image classification. In Medical Imaging with Deep Learning (MIDL).

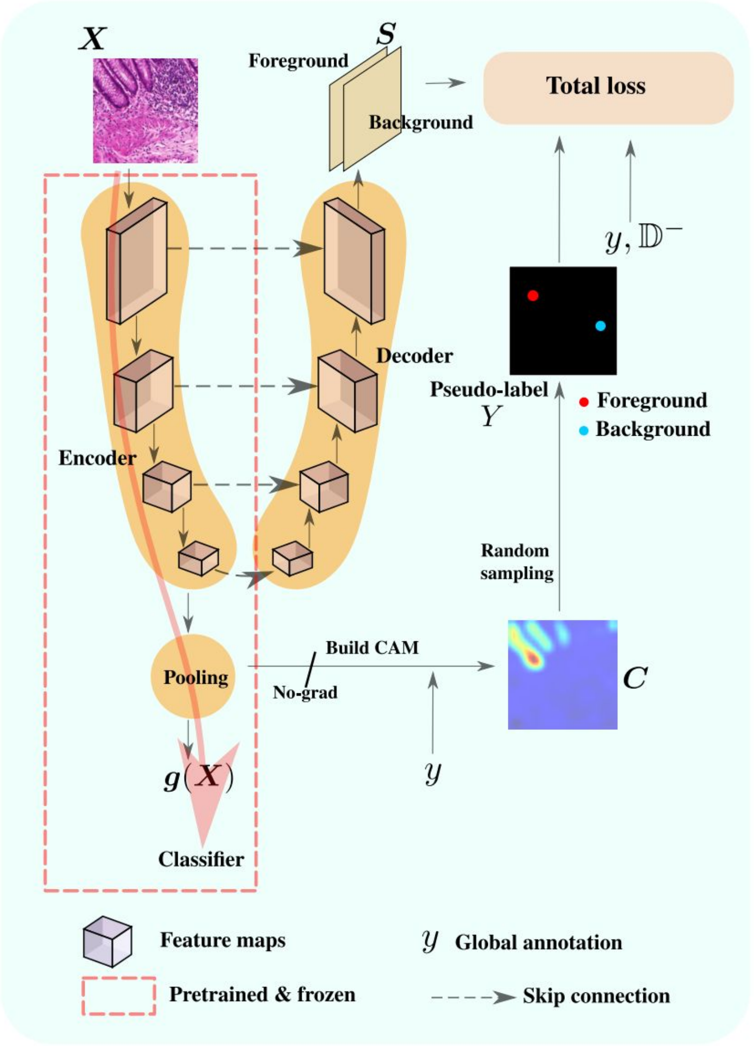
Le génie pour l'industrie

(a) Cancer Grading and Localization in Histology:

- **Training**

Total adaptive loss

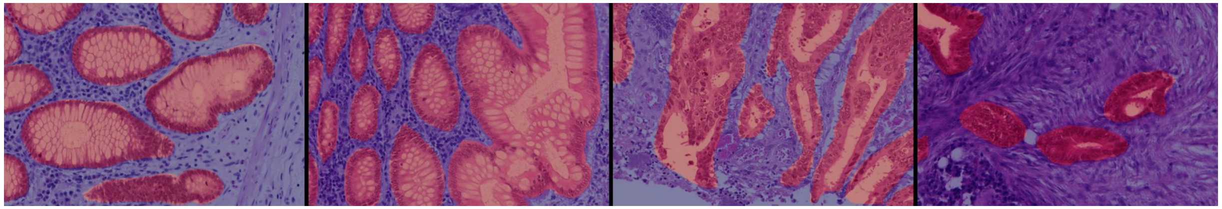
$$\min_{\boldsymbol{\theta}} \quad \mathbb{1}_{\boldsymbol{X} \in \mathbb{D}^-} \Big(\sum_{p \in \Omega} - \log(1 - \boldsymbol{S}_p^0) \Big)$$
$$+ (1 - \mathbb{1}_{\boldsymbol{X} \in \mathbb{D}^-}) \Big(\lambda \sum_{p \in \{\mathbb{C}^+ \cup \mathbb{C}^-\}} \boldsymbol{H}(Y_p, \boldsymbol{S}_p) \Big) ,$$



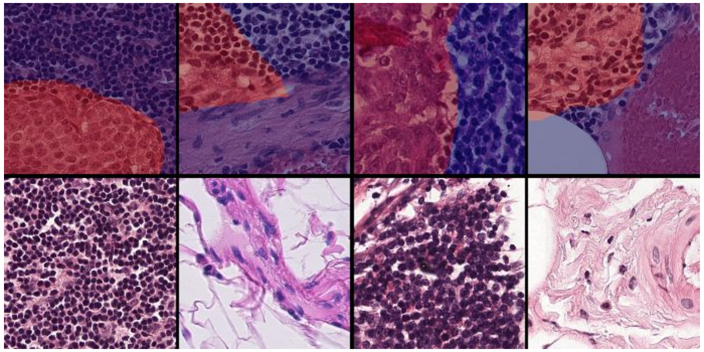
(a) Cancer Grading and Localization in Histology:

- **Experiments**

- Task: classify and localize ROI
- 2 public datsets: GlaS, Camelyon16 patches.



GlaS: colon cancer diagnosis



Camelyon16 patches: breast cancer

Belharbi, S., Pedersoli, M., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022). Negative evidence matters in interpretable histology image classification. In Medical Imaging with Deep Learning (MIDL).

(a) Cancer Grading and Localization in Histology:

- **Results**

	GlaS				CAMELYON16			
	VGG	Inception	ResNet	Mean	VGG	Inception	ResNet	Mean
Metric	PxAP							
WSL								
GAP (Lin et al., 2013) *(corr,2013)*	58.5	57.5	56.2	57.4	37.5	24.6	43.7	35.2
MAX-POOL (Oquab et al., 2015) *(cvpr,2015)*	58.5	57.1	46.2	53.9	42.1	40.9	20.2	34.4
LSE (Sun et al., 2016) *(cvpr,2016)*	63.9	62.8	59.1	61.9	**63.1**	29.0	42.1	44.7
CAM (Zhou et al., 2016) *(cvpr,2016)*	68.5	50.5	64.4	61.1	25.4	48.7	27.5	33.8
HaS (Singh and Lee, 2017) *(iccv,2017)*	65.5	65.4	63.5	64.8	25.4	47.1	29.7	34.0
GradCAM (Selvaraju et al., 2017) *(iccv,2017)*	75.7	56.9	70.0	67.5	40.2	34.4	29.1	34.5
WILDCAT (Durand et al., 2017) *(cvpr,2017)*	56.1	54.9	60.1	57.0	44.4	31.4	31.0	35.6
ACoL (Zhang et al., 2018a) *(cvpr,2018)*	63.7	58.2	54.2	58.7	31.3	39.3	31.3	33.9
SPG (Zhang et al., 2018b) *(eccv,2018)*	63.6	58.3	51.4	57.7	45.4	24.5	22.6	30.8
GradCAM++ (Chattopadhyay et al., 2018) *(wacv,2018)*	**76.1**	65.7	70.7	70.8	41.3	43.9	25.8	37.0
Deep MIL (Ilse et al., 2018) *(icml,2018)*	66.6	61.8	64.7	64.3	53.8	**51.1**	**57.9**	54.2
PRM (Zhou et al., 2018) *(cvpr,2018)*	59.8	53.1	62.3	58.4	46.0	41.7	23.2	36.9
ADL (Choe and Shim, 2019) *(cvpr,2019)*	65.0	60.6	54.1	59.9	19.0	46.0	46.0	37.0
CutMix (Yun et al., 2019) *(eccv,2019)*	59.9	50.4	56.7	55.6	56.4	44.9	20.7	40.6
Smooth-GradCAM (Omeiza et al., 2019) *(corr,2019)*	71.3	**67.6**	75.5	71.4	35.1	31.6	25.1	30.6
XGradCAM (Fu et al., 2020) *(bmvc,2020)*	73.7	66.4	62.6	67.5	40.2	33.0	24.4	32.5
LayerCAM (Jiang et al., 2021) *(ieee,2021)*	67.8	66.1	**70.9**	68.2	34.1	25.0	29.1	29.4
NEGEV (ours)	**81.3**	**70.1**	**82.0**	**77.8**	**70.3**	**53.8**	52.6	**58.9**
Fully supervised								
U-Net (Ronneberger et al., 2015)*(miccai,2015)*	96.8	95.4	96.4	96.2	83.0	82.2	83.6	82.9

Table 1: PxAP performance over GlaS and CAMELYON16 test sets.

Belharbi, S., Pedersoli, M., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022). Negative evidence matters in interpretable histology image classification. In Medical Imaging with Deep Learning (MIDL).

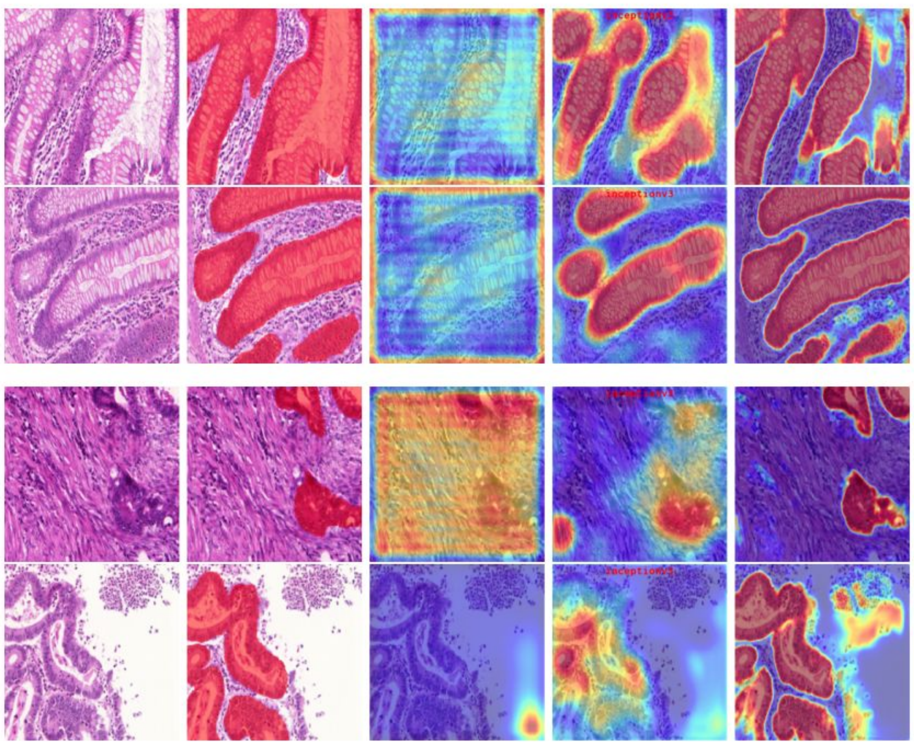
ÉTS
Le génie pour l'industrie

(a) Cancer Grading and Localization in Histology:

- **Results**

Full-sup

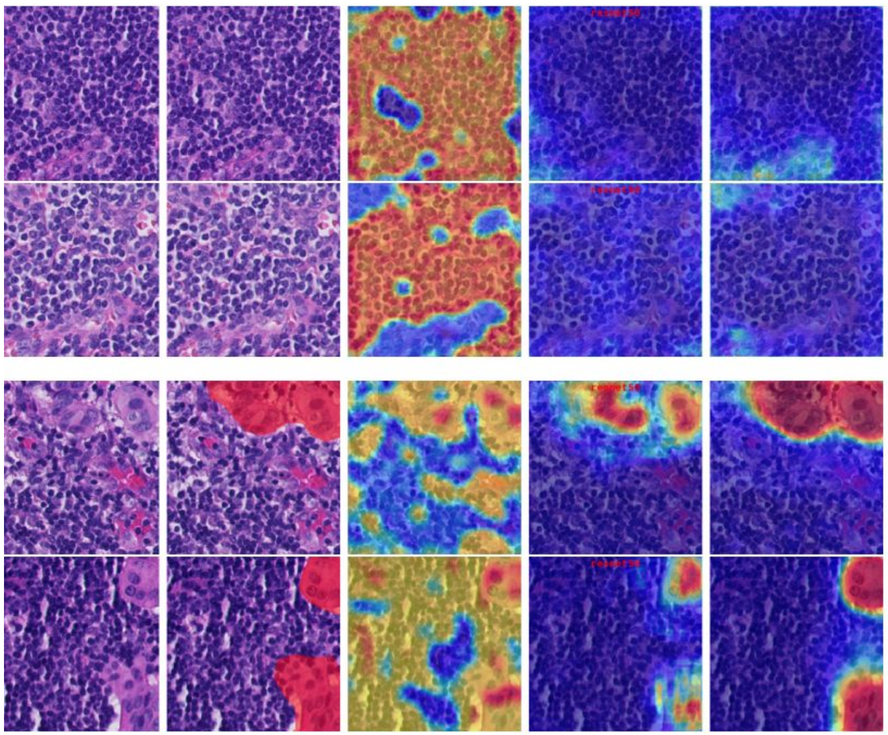
Img gt CAM Ours Unet



GlaS

Full-sup

Img gt CAM Ours Unet



Camelyon16

Le génie pour l'industrie

Belharbi, S., Pedersoli, M., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022). Negative evidence matters in interpretable histology image classification. In Medical Imaging with Deep Learning (MIDL).

(a) Cancer Grading and Localization in Histology:

- **Ablations**

Methods	GlaS VGG	GlaS Inception	GlaS ResNet	GlaS Mean	CAMELYON16 VGG	CAMELYON16 Inception	CAMELYON16 ResNet	CAMELYON16 Mean
CAM (Zhou et al., 2016)	68.5	50.5	64.4	61.1	25.4	48.7	27.5	20.3
Ours + $\mathbb{C}^+$	81.3	53.3	81.3	71.9	38.1	36.5	30.8	35.1
Ours + $\mathbb{C}^+$ + $\mathbb{C}^-$	81.3	70.1	82.0	77.8	38.1	35.3	30.2	34.5
Ours + $\mathbb{C}^+$ + $\mathbb{C}^-$ + $\mathbb{D}^-$	–	–	–	–	70.3	53.8	52.6	58.9
Improvement	+12.8	+19.6	+17.6	+16.6	+44.9	+5.0	+25.1	+25.0

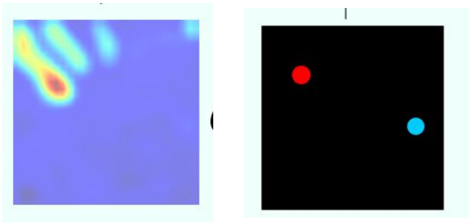
Impact of different terms

$$\min_{\boldsymbol{\theta}} \quad \mathbb{1}_{\boldsymbol{X}\in\mathbb{D}^-}\Big(\sum_{p\in\Omega} -\log(1-\boldsymbol{S}_p^0)\Big) + (1-\mathbb{1}_{\boldsymbol{X}\in\mathbb{D}^-})\Big(\lambda \sum_{p\in\{\mathbb{C}^+\cup\mathbb{C}^-\}} \boldsymbol{H}(Y_p, \boldsymbol{S}_p)\Big) \;,$$

Belharbi, S., Pedersoli, M., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022). Negative evidence matters in interpretable histology image classification. In Medical Imaging with Deep Learning (MIDL).

(a) Cancer Grading and Localization in Histology:

- **Ablations**



Methods	GlaS VGG	GlaS Inception	GlaS ResNet	GlaS Mean	CAMELYON16 VGG	CAMELYON16 Inception	CAMELYON16 ResNet	CAMELYON16 Mean
CAM (Zhou et al., 2016)	68.5	50.5	64.4	61.1	25.4	48.7	27.5	33.8
Ours ($n = 1$, random selection)	81.3	70.1	82.0	77.8	70.3	53.8	52.6	58.9
Ours ($n = 1$, static selection)	77.7	60.3	76.5	71.5	57.5	47.4	42.8	49.2
Performance drop	-3.6	-9.8	-5.5	-6.3	-12.8	-6.4	-9.8	-9.6

Fixed vs random seeds selection

Le génie pour l'industrie

Belharbi, S., Pedersoli, M., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022). Negative evidence matters in interpretable histology image classification. In Medical Imaging with Deep Learning (MIDL).

(a) Cancer Grading and Localization in Histology:

- **Ablations**

	GlaS				CAMELYON16			
n	VGG	Inception	ResNet	Mean	VGG	Inception	ResNet	Mean
1	81.3	70.1	82.0	77.8	70.3	53.8	52.6	58.9
2	81.3	52.9	81.3	71.8	69.7	51.1	47.2	56.0
3	81.3	52.9	81.3	71.8	69.7	51.9	47.2	56.2
4	81.3	55.0	81.3	72.5	69.7	50.0	47.2	56.6
5	81.3	52.9	81.3	71.8	69.7	53.4	47.2	56.7
10	81.3	53.7	81.3	72.1	69.7	52.6	47.2	56.5
20	81.3	52.9	82.2	72.1	69.7	51.3	47.2	56.0
50	81.3	52.0	81.3	71.5	69.7	53.8	50.3	57.9
100	81.3	53.4	81.3	72.0	69.7	50.5	47.2	55.8
500	81.3	52.9	81.3	71.8	69.7	51.5	47.6	56.2
1k	81.3	53.7	81.3	72.1	69.7	51.2	48.5	56.4
2k	81.3	53.0	81.3	71.8	69.7	51.5	47.2	56.1
3k	81.3	54.2	81.3	72.2	69.7	50.4	48.5	56.2
4k	81.3	52.9	81.3	71.8	69.7	52.9	47.2	56.6
5k	81.3	53.2	82.7	72.4	69.7	51.4	47.7	56.2
10k	81.3	52.9	81.3	71.8	69.7	52.1	47.2	56.3
CAM (Zhou et al., 2016)	68.5	50.5	64.4	61.1	25.4	48.7	27.5	33.8

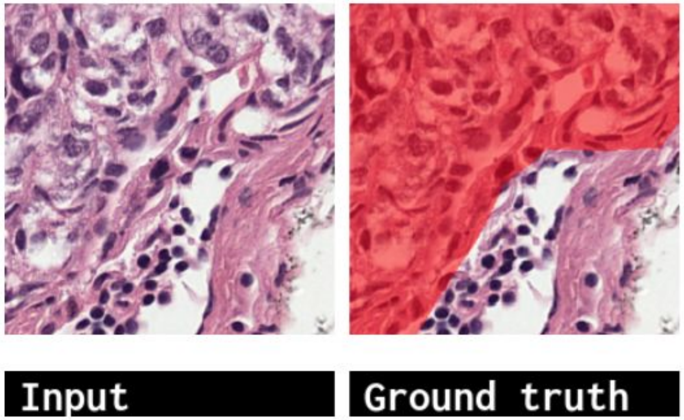
How many pixels to sample?

Le génie pour l'industrie

Belharbi, S., Pedersoli, M., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022). Negative evidence matters in interpretable histology image classification. In Medical Imaging with Deep Learning (MIDL).

(a) Cancer Grading and Localization in Histology:

- **Presented work**



- Belharbi, S., Rony, J., Dolz, J., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022). **Deep interpretable classification and weakly-supervised segmentation of histology images via max-min uncertainty.** IEEE Transactions on Medical Imaging, 41:702–714.

- Belharbi, S., Pedersoli, M., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022). **Negative evidence matters in interpretable histology image classification**. In Medical Imaging with Deep Learning (MIDL).

Completed.

ÉTS Le génie pour l'industrie

Part 4: Applications of WSOL / WSSS

(a) Medical Cancer Grading and ROI Localization in Histology

(b) Weakly-Supervised Video Object Localization

(c) Person ReID: Embedding Networks

(d) Medical Semantic Segmentation

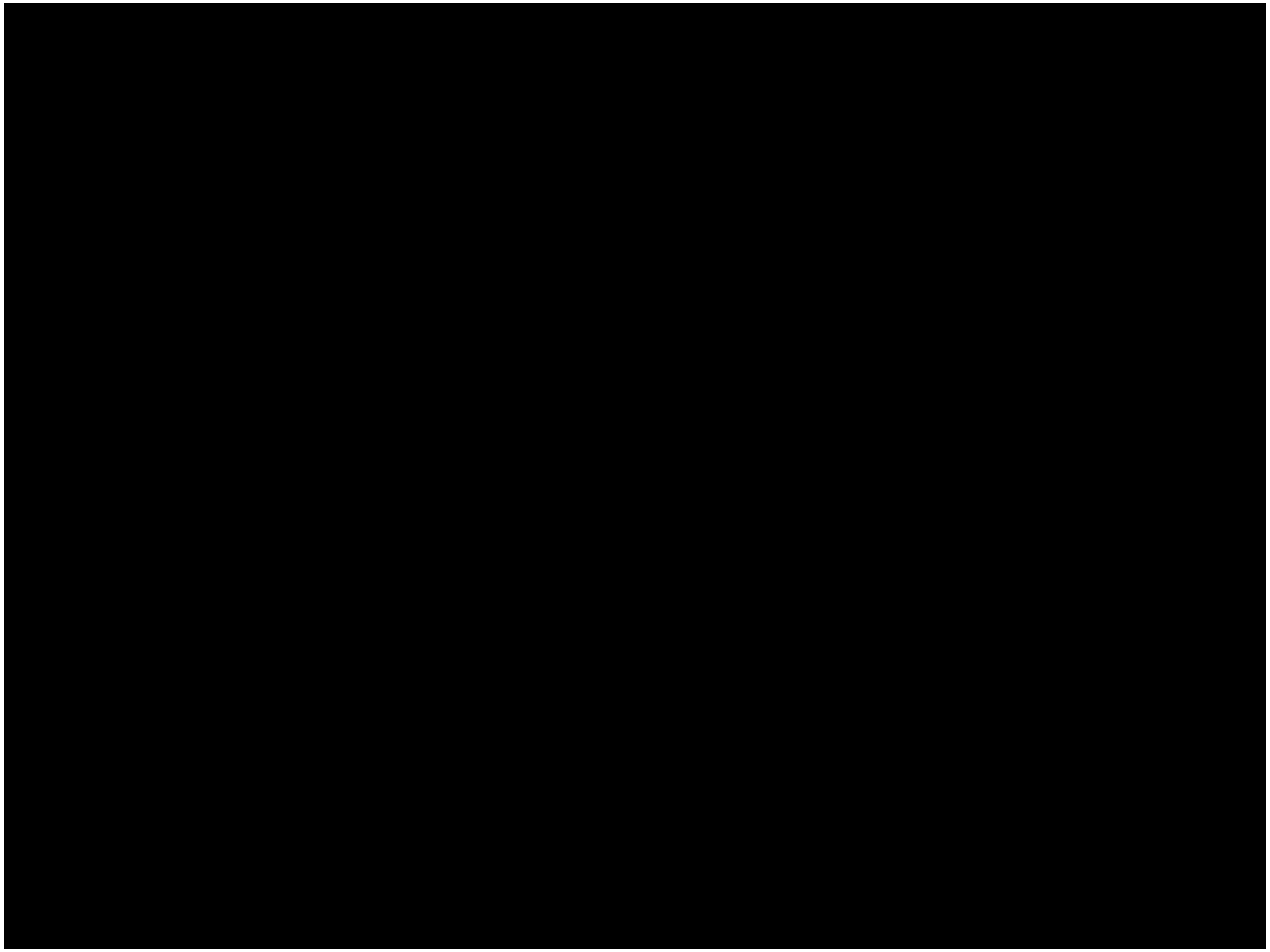
(b) Weakly-Supervised Video Object Localization: (Ongoing work)

TCAM: Temporal Class Activation Maps for Object Localization in Weakly-Labeled Unconstrained Videos

Will be available early Sept. (paper + code)

(b) Weakly-Supervised Video Object Localization: (Ongoing work)

- **Task**



Moving objects

Camera motions

View-point changes

Decoding artifacts

Editing effects

Unconstrained videos
(Dataset: YouTube-Objects v1.0)

Video:
https://docs.google.com/file/d/1sF9cynslvdcS_pAtDispmyxUPGa2Y7dg/preview

TCAM: Temporal Class Activation Maps for Object Localization in Weakly-Labeled Unconstrained Videos. 2022

(b) Weakly-Supervised Video Object Localization: (Ongoing work)

- **Task: Supervision**

  - Labeling all frames via bbox is expensive

  - Use **weak supervision: global video tag (cheap!)**

  Global video tag: main object class in the video (not necessarily present in all frames)

(b) Weakly-Supervised Video Object Localization: (Ongoing work)

- **Task**

Video → Model → Localize object in each frame (bbox)

Model is trained using weak labels (global video tag)

(b) Weakly-Supervised Video Object Localization: (Ongoing work)

- **Task**

Video object localization helps:

- Localizing object of interest in video
- Video content understanding
- Improve subsequent tasks: video summarization, event detection, video object detection, video tracking, …

(b) Weakly-Supervised Video Object Localization: (Ongoing work)

- **Current state-of-the-art**

  - Stagnated (<= 2020)

(b) Weakly-Supervised Video Object Localization: (Ongoing work)

- **Current state-of-the-art**

- Stagnated (<= 2020)

- Multiple sequential, independent stages:
    1. Generate spatio-temporal segments/proposals (visual and motion cues)
    2. Identify prominent object
    3. Refine

(b) Weakly-Supervised Video Object Localization: (Ongoing work)

- **Current state-of-the-art**

- Stagnated (<= 2020)

- Multiple sequential, independent stages:
  1. Generate spatio-temporal segments/proposals (visual and motion cues)
  2. Identify prominent object
  3. Refine

- Video tags are used only to cluster video

(b) Weakly-Supervised Video Object Localization: (Ongoing work)

- **Current state-of-the-art**

  - Stagnated (<= 2020)
  - Multiple sequential, independent stages:
    1. Generate spatio-temporal segments/proposals (visual and motion cues)
    2. Identify prominent object
    3. Refine
  - Video tags are used only to cluster video
  - ROI are not necessarily discriminative

(b) Weakly-Supervised Video Object Localization: (Ongoing work)

- **Current state-of-the-art**

  - Stagnated (<= 2020)
  - Multiple sequential, independent stages:
    1. Generate spatio-temporal segments/proposals (visual and motion cues)
    2. Identify prominent object
    3. Refine
  - Video tags are used only to cluster video
  - ROI are not necessarily discriminative
  - Motion cues (optical flow) is not necessarily discriminative, is noisy and requires further post-processing

(b) Weakly-Supervised Video Object Localization: (Ongoing work)

- **Current state-of-the-art**

- Stagnated (<= 2020)
- Multiple sequential, independent stages:
  1. Generate spatio-temporal segments/proposals (visual and motion cues)
  2. Identify prominent object
  3. Refine
- Video tags are used only to cluster video
- ROI are not necessarily discriminative
- Motion cues (optical flow) is not necessarily discriminative, is noisy and requires further post-processing
- Localization is done by solving an optimization problem over a cluster of videos or single video (slow inference time, build model per-class/video)

(b) Weakly-Supervised Video Object Localization: (Ongoing work)

- **Proposal**

Leverage CAMs for weakly supervised video object localization

(b) Weakly-Supervised Video Object Localization: (Ongoing work)

- **Proposal**

WSOL CAM methods trained on still images yield descent performance

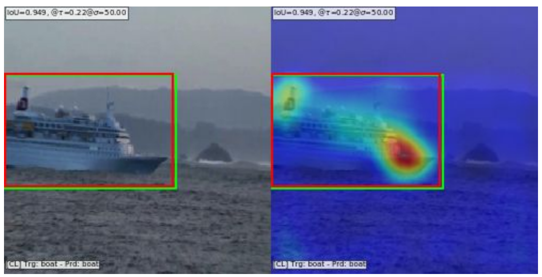
TCAM: Temporal Class Activation Maps for Object Localization in Weakly-Labeled Unconstrained Videos. 2022

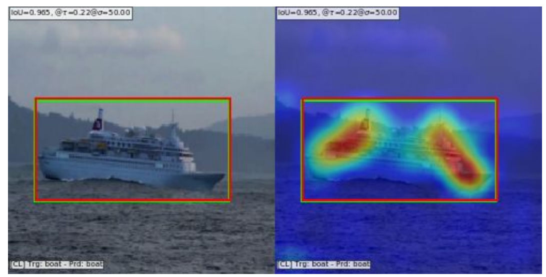
Le génie pour l'industrie

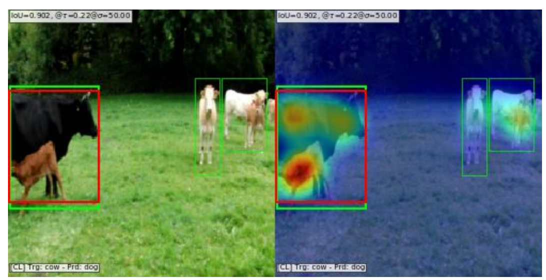
(b) Weakly-Supervised Video Object Localization:
(Ongoing work)

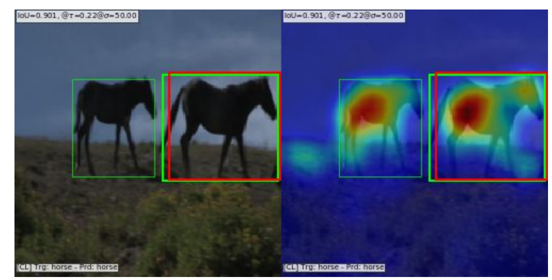
- **Proposal**

WSOL CAM methods trained on still images yield descent performance









Do not account for spatio-temporal dependency!

Le génie pour l'industrie

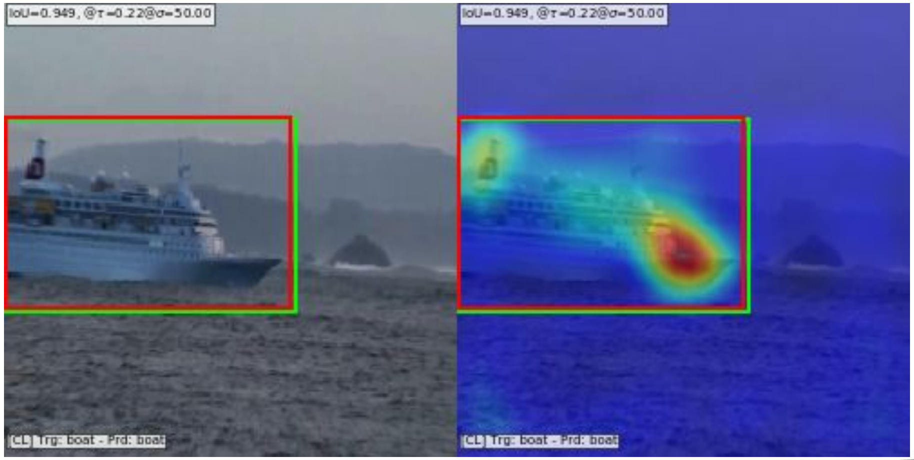
TCAM: Temporal Class Activation Maps for Object Localization in Weakly-Labeled Unconstrained Videos. 2022

(b) Weakly-Supervised Video Object Localization: (Ongoing work)

- **Proposal**

Observation:

Slight variation in consecutive frames leads to variation in CAMs



(b) Weakly-Supervised Video Object Localization: (Ongoing work)

- **Proposal**

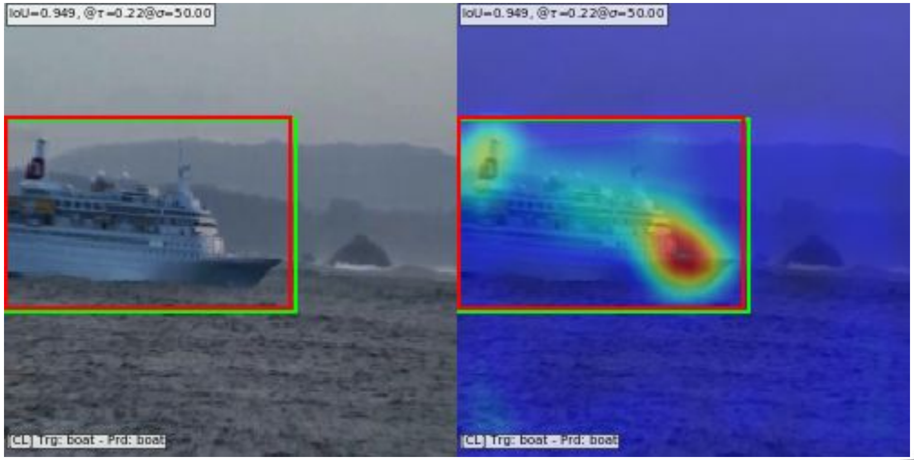
Observation:

Slight variation in consecutive frames leads to variation in CAMs

—>

Aggregate consecutive CAMs to build complete CAM

And use it to sample pseudo-labels

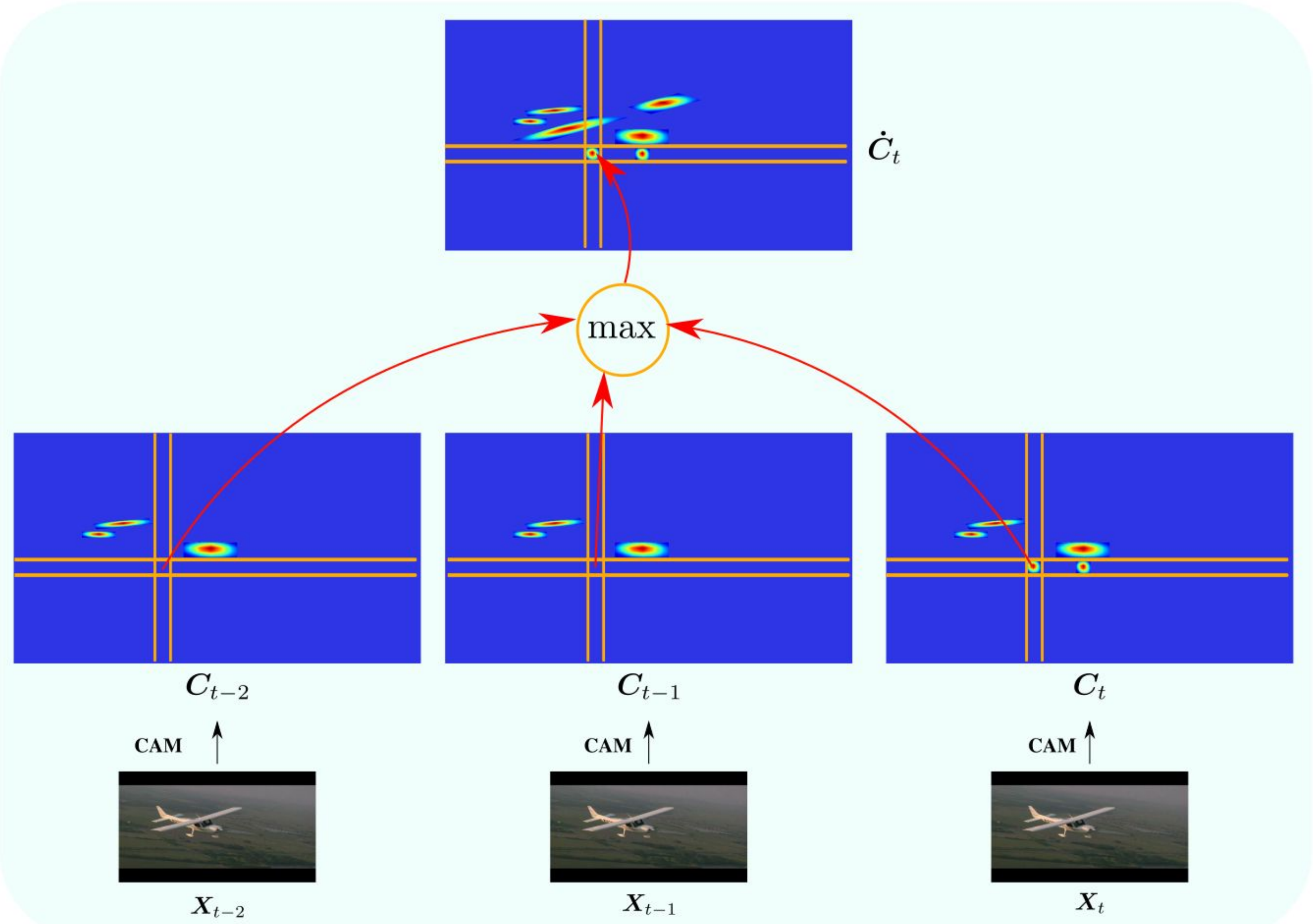


(b) Weakly-Supervised Video Object Localization: (Ongoing work)

- **Proposal**

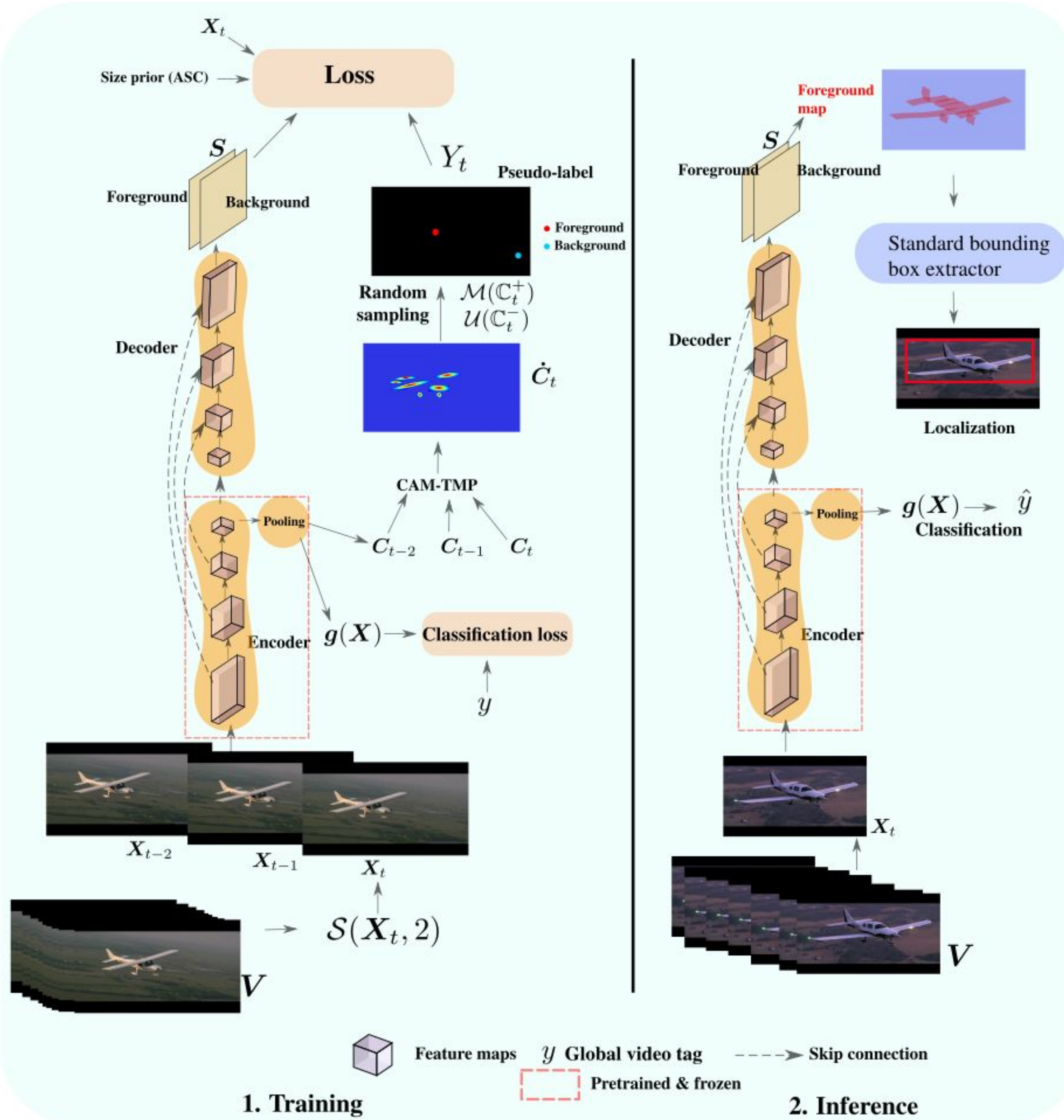
CAMs aggregation

CAM-**T**emporal **M**ax **P**ooling (**CAM-TMP**)



(b) Weakly-Supervised Video Object Localization: (Ongoing work)

- **Proposal**

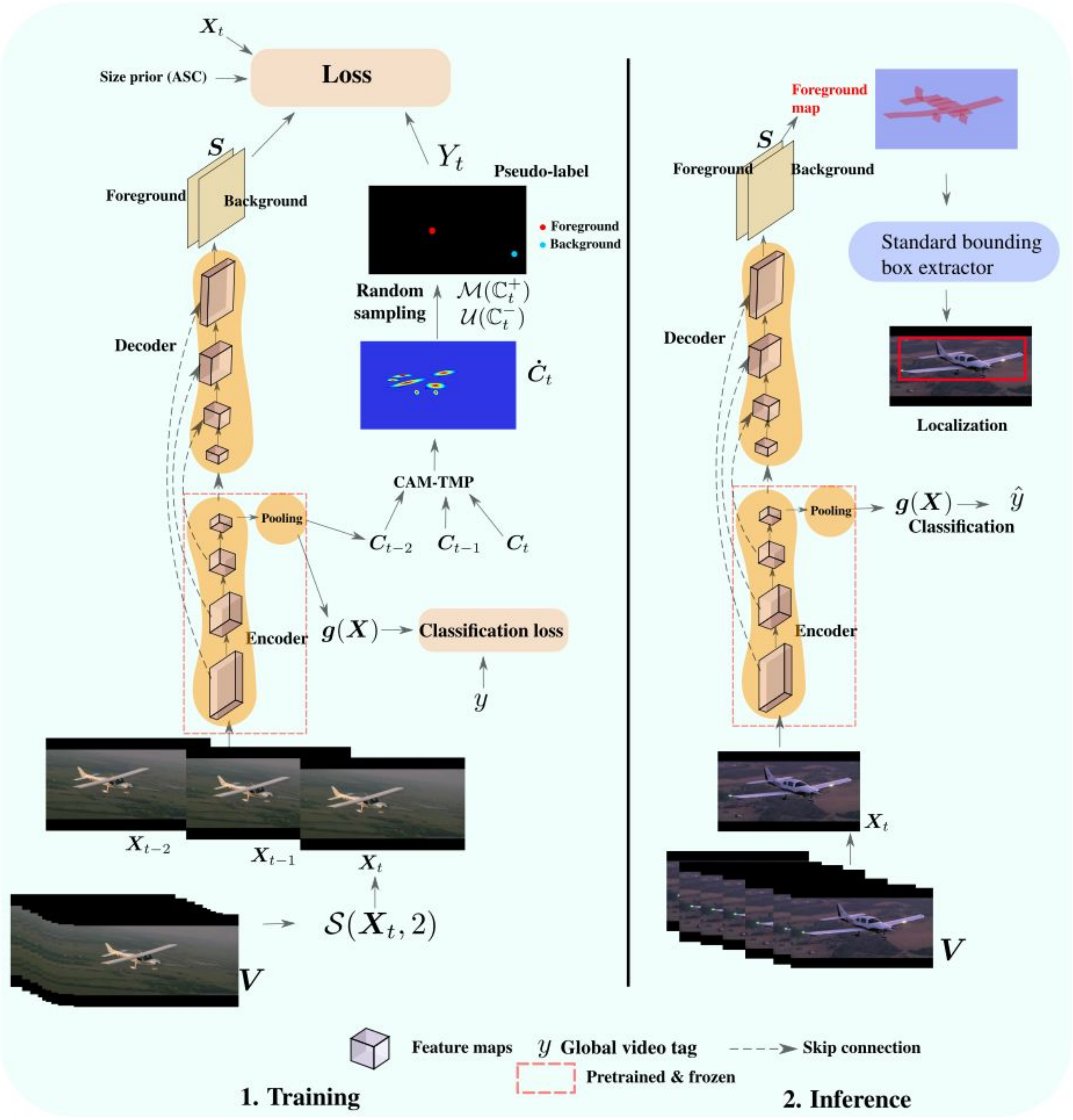


(b) Weakly-Supervised Video Object Localization: (Ongoing work)

- **Proposal**

Training:
accounts for spatio-temporal dependency

$$\min_{\boldsymbol{\theta}} \quad \sum_{p \in \Omega'_t} \boldsymbol{H}_p(Y_t, \boldsymbol{S}_t) + \lambda\, \mathcal{R}(\boldsymbol{S}_t, \boldsymbol{X}_t)\,,$$
$$\text{s.t.} \quad \sum \boldsymbol{S}^r(t) \geq 0\,, \quad r \in \{0,1\}\,,$$



(b) Weakly-Supervised Video Object Localization: (Ongoing work)

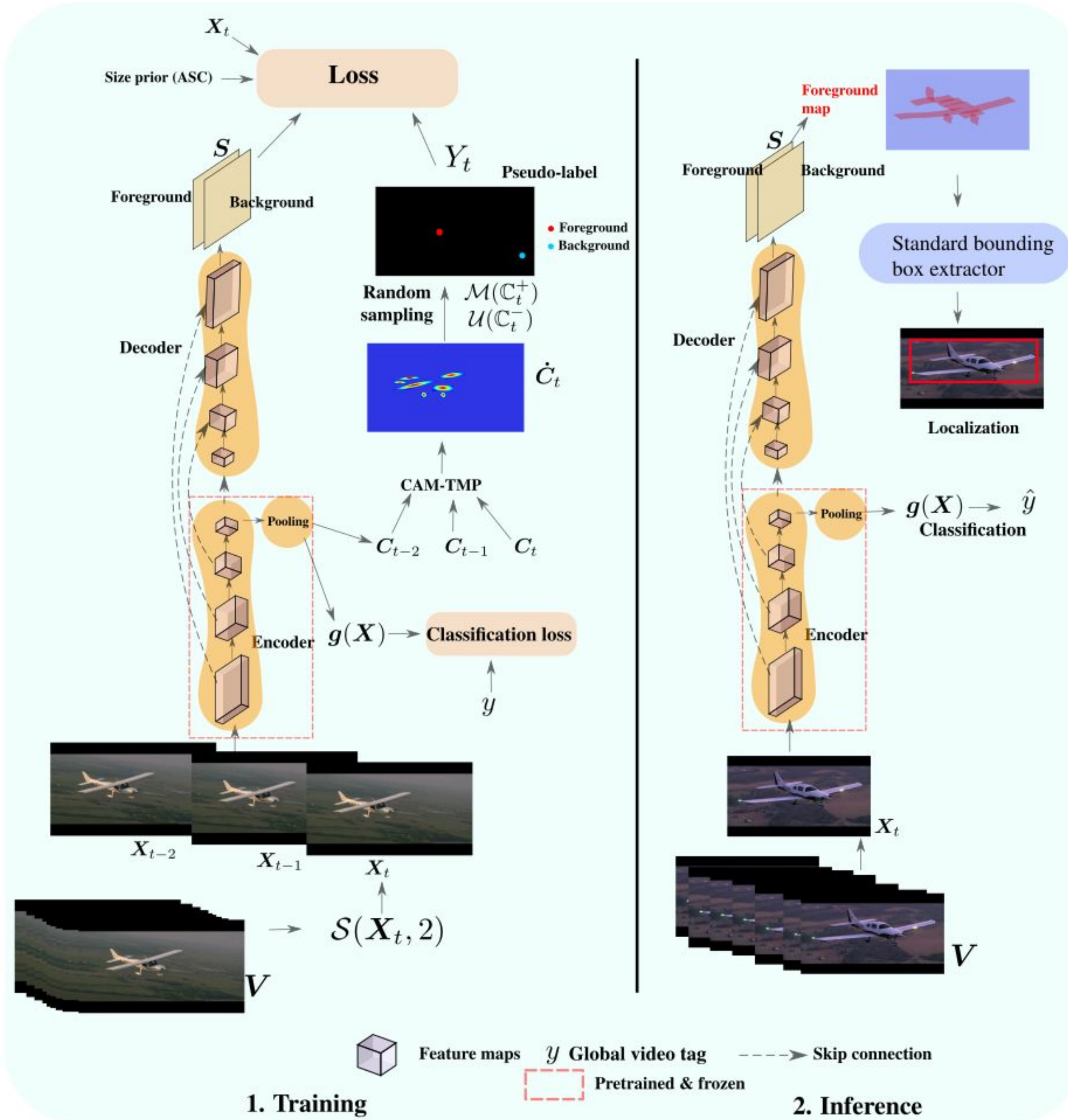
- **Proposal**

Training:
accounts for spatio-temporal dependency

$$\min_{\boldsymbol{\theta}} \quad \sum_{p \in \Omega'_t} \boldsymbol{H}_p(Y_t, \boldsymbol{S}_t) + \lambda\, \mathcal{R}(\boldsymbol{S}_t, \boldsymbol{X}_t)\;,$$
$$\text{s.t.} \quad \sum \boldsymbol{S}^r(t) \geq 0\;, \quad r \in \{0, 1\}\;,$$

Methods	CorLoc
Layer-CAM [22] *(ieee,2021)*	63.0
Ours + $\mathbb{C}^+$ + $\mathbb{C}^-$	68.5
Ours + $\mathbb{C}^+$ + $\mathbb{C}^-$ + CRF	69.6
Ours + $\mathbb{C}^+$ + $\mathbb{C}^-$ + ASC	66.2
Ours + $\mathbb{C}^+$ + $\mathbb{C}^-$ + CRF + ASC	70.5
Ours + $\mathbb{C}^+$ + $\mathbb{C}^-$ + CRF + ASC + CAM-TMP	72.8
Improvement	+9.8

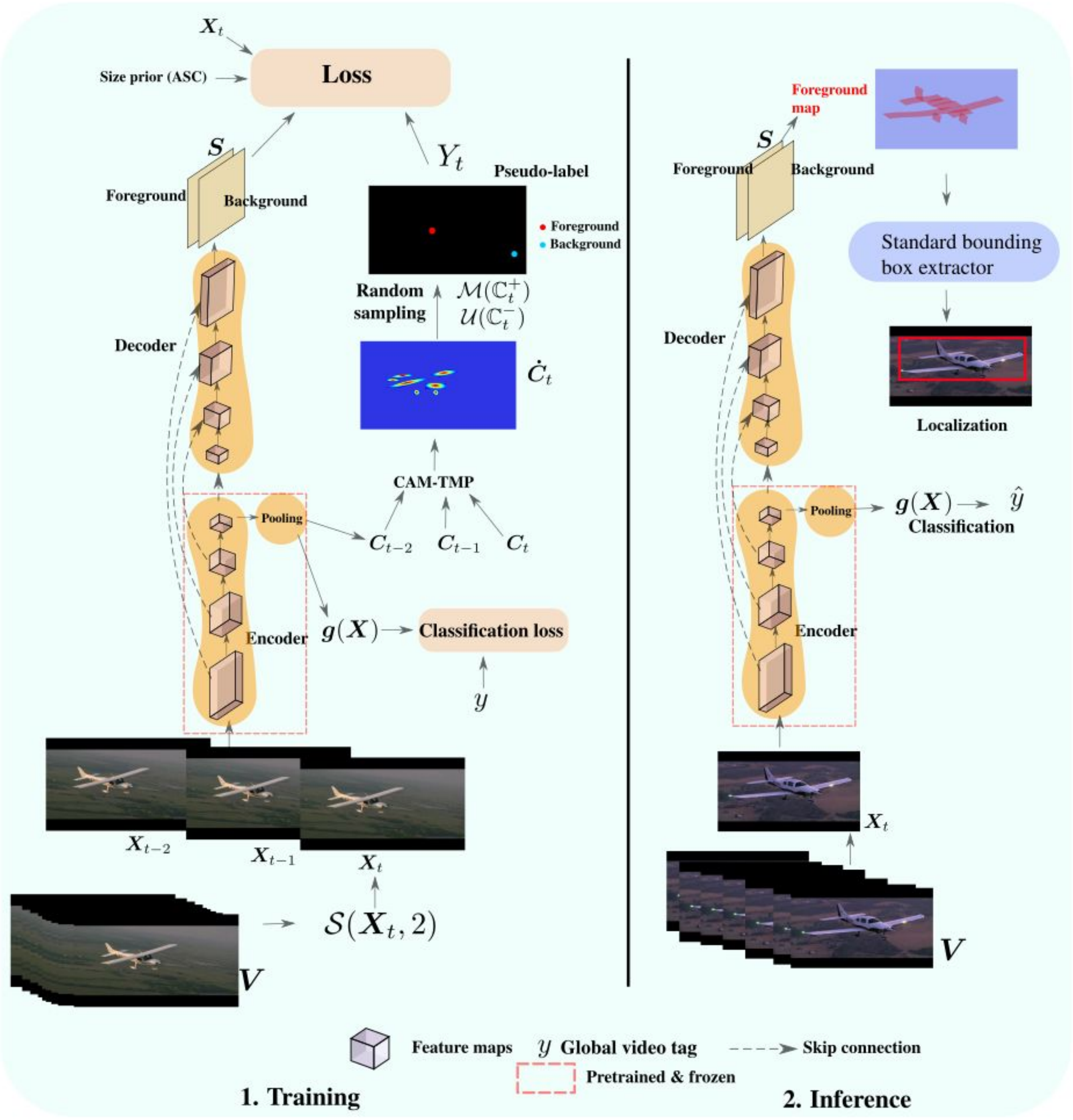
Ablation



(b) Weakly-Supervised Video Object Localization: (Ongoing work)

- **Proposal**

Inference:
Independent frames → fast



Le génie pour l'industrie

(b) Weakly-Supervised Video Object Localization: (Ongoing work)

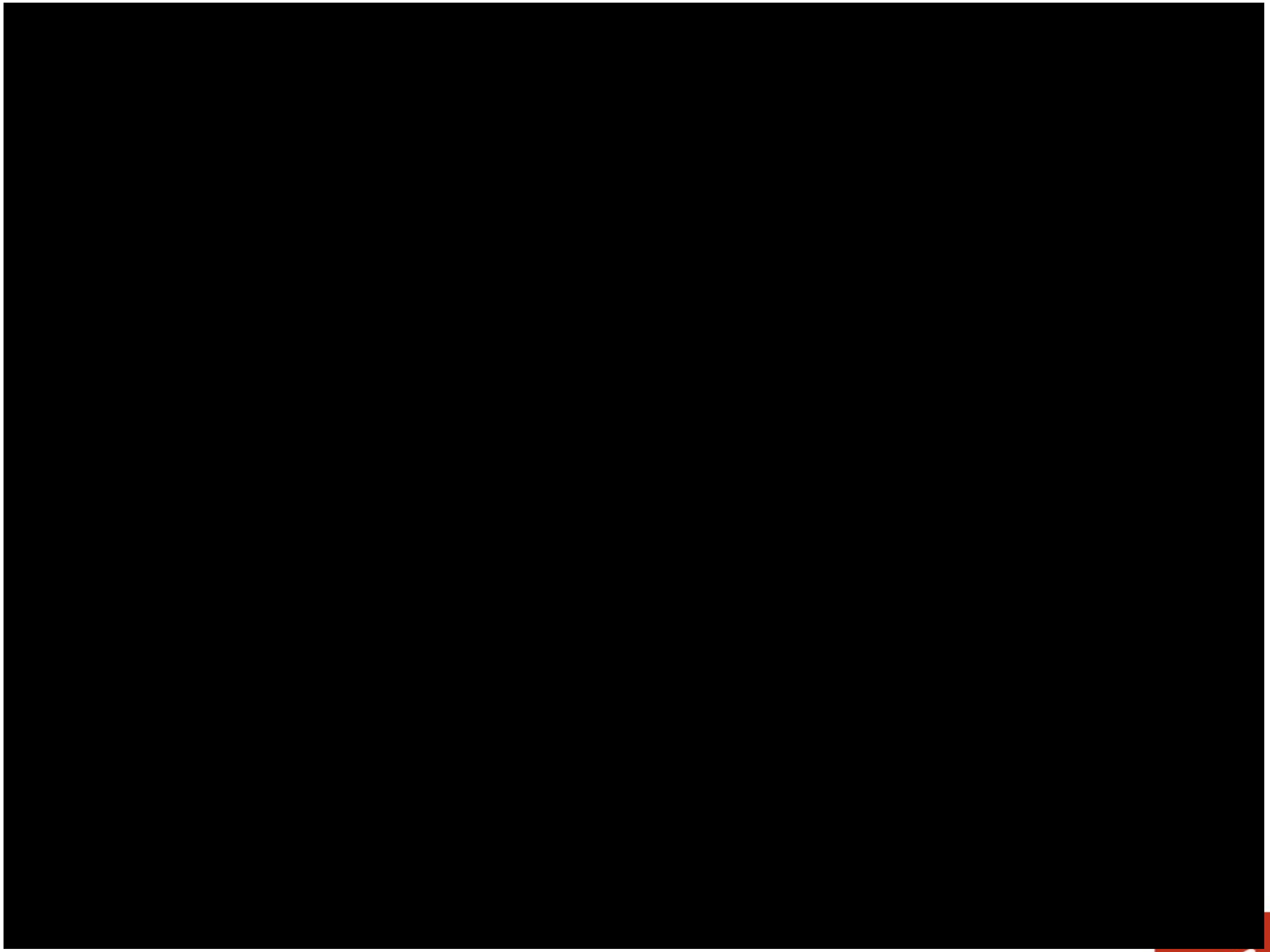
- **Results**

Video:
https://docs.google.com/file/d/1KbrQu35oX2NpoH8NiKDt0cY5Qy1kqbwo/preview

"Plane"

(b) Weakly-Supervised Video Object Localization: (Ongoing work)

- **Results**



Video:
https://docs.google.com/file/d/1_UP0Dwdp7TI4gs84BPyK2_rN7ENPysmK/preview

"Horse"

Le génie pour l'industrie

(b) Weakly-Supervised Video Object Localization: (Ongoing work)

- **Results**

Video:
https://docs.google.com/file/d/1wCxm1votCm_1M-cENBq414ZA0tnZXCC5/preview

"Car"

ÉTS
Le génie pour l'industrie

(b) Weakly-Supervised Video Object Localization:
(Ongoing work)

- **Results**

Video:
https://docs.google.com/file/d/1dckaUlkaaqPSyeQyIcbdNFyTTguVG-vQ/preview

"Car"

(b) Weakly-Supervised Video Object Localization: (Ongoing work)

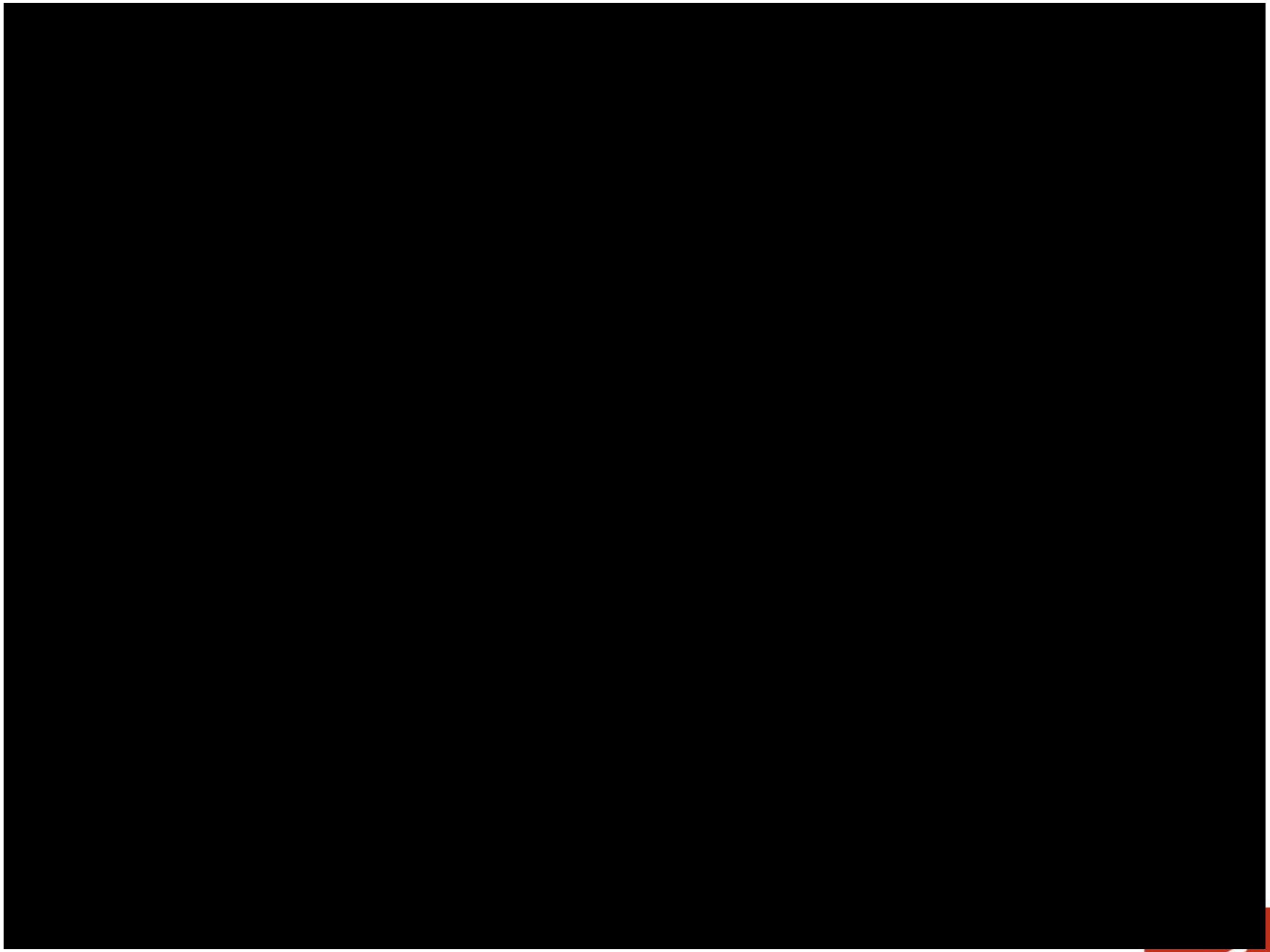
- **Results**

Video:
https://docs.google.com/file/d/1XkZBEjtb-I-bu5nK2gSSsKH-FfASqOQ-/preview

"Car"

(b) Weakly-Supervised Video Object Localization: (Ongoing work)

- **Results**



Video: https://docs.google.com/file/d/1Gf8y2SUNEKhqWkkxARIo0xY65kJPorXq/preview

“Car”

TCAM: Temporal Class Activation Maps for Object Localization in Weakly-Labeled Unconstrained Videos. 2022

ÉTS Le génie pour l’industrie

(b) Weakly-Supervised Video Object Localization: (Ongoing work)

TCAM: Temporal Class Activation Maps for Object Localization in Weakly-Labeled Unconstrained Videos

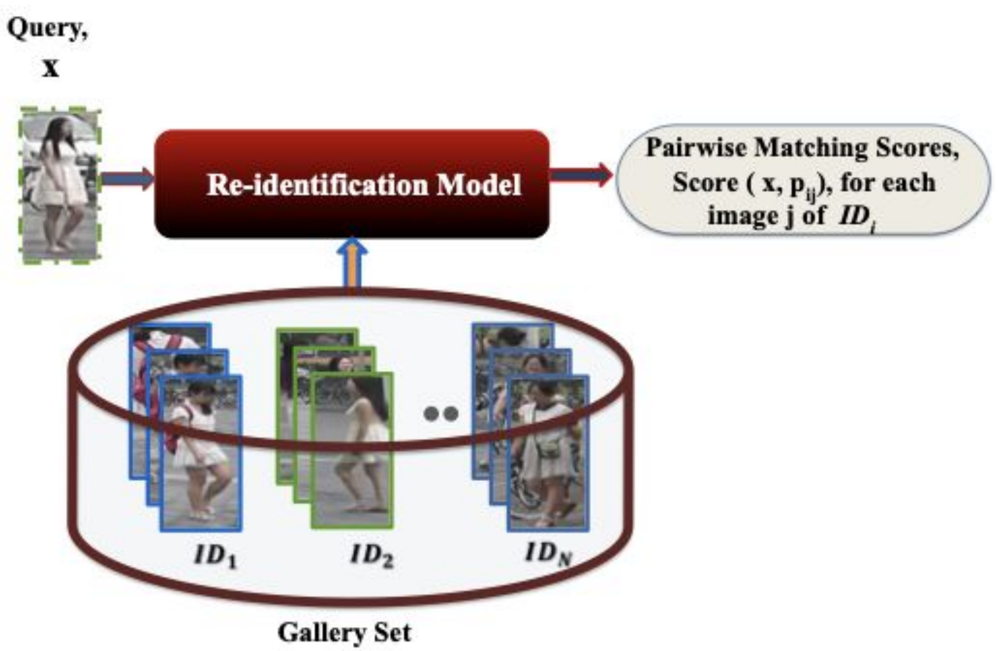
Will be available early Sept. (paper + code)

Completed.

(c) Person ReID: Embedding Networks

Person Re-Identification:

recognize individuals captured over a distributed set of non-overlapping video camera views



Challenges in real-world scenarios:

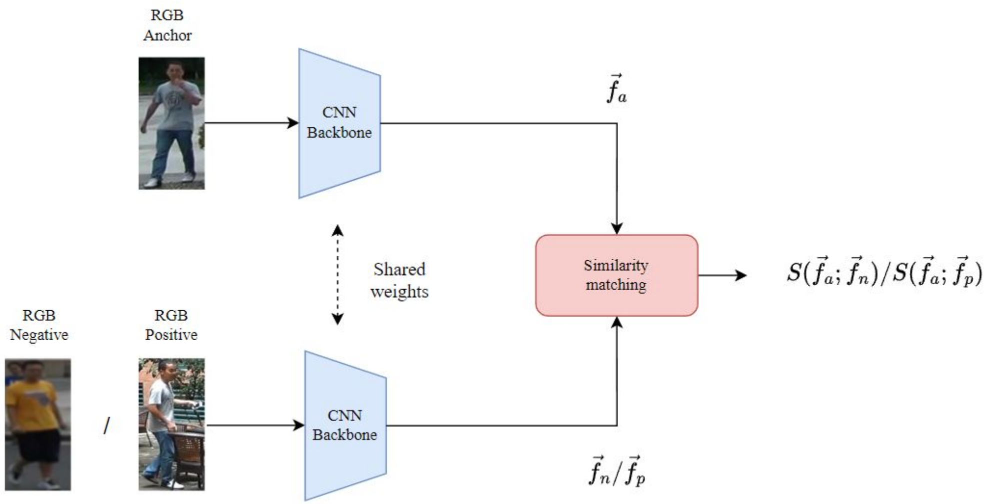
- low resolution, scales,
- motion blur
- occlusions, poor viewpoint
- variations in pose and viewpoint
- variation in illumination
- open set scenarios
- pedestrians with similar clothes
- misalignment over different views

(c) Person ReID: Embedding Networks

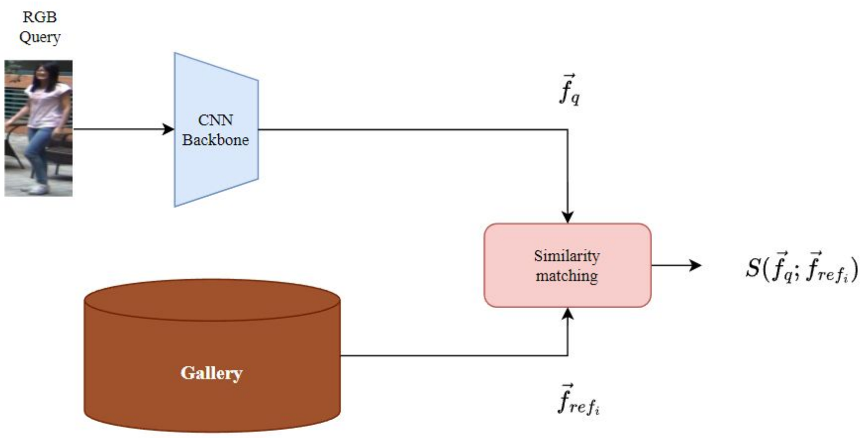
Deep Siamese CNNs for pairwise similarity matching:

- **matching**: given query and reference images, they assess similarity, using Euclidean, cosine distance, etc., between feature representations
- **metric learning**: specific losses, like triplet, contrastive, and magnet losses are used to learn coherent embeddings

Training model



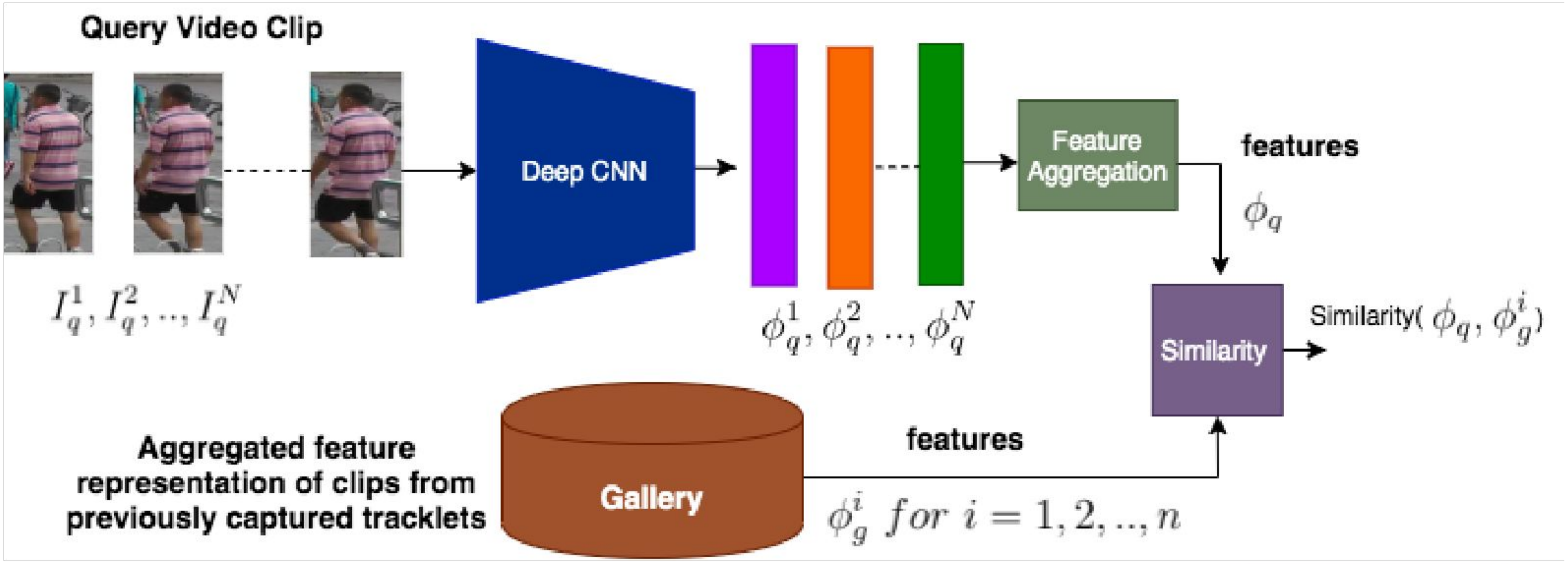
Test model



(c) Person ReID: Embedding Networks

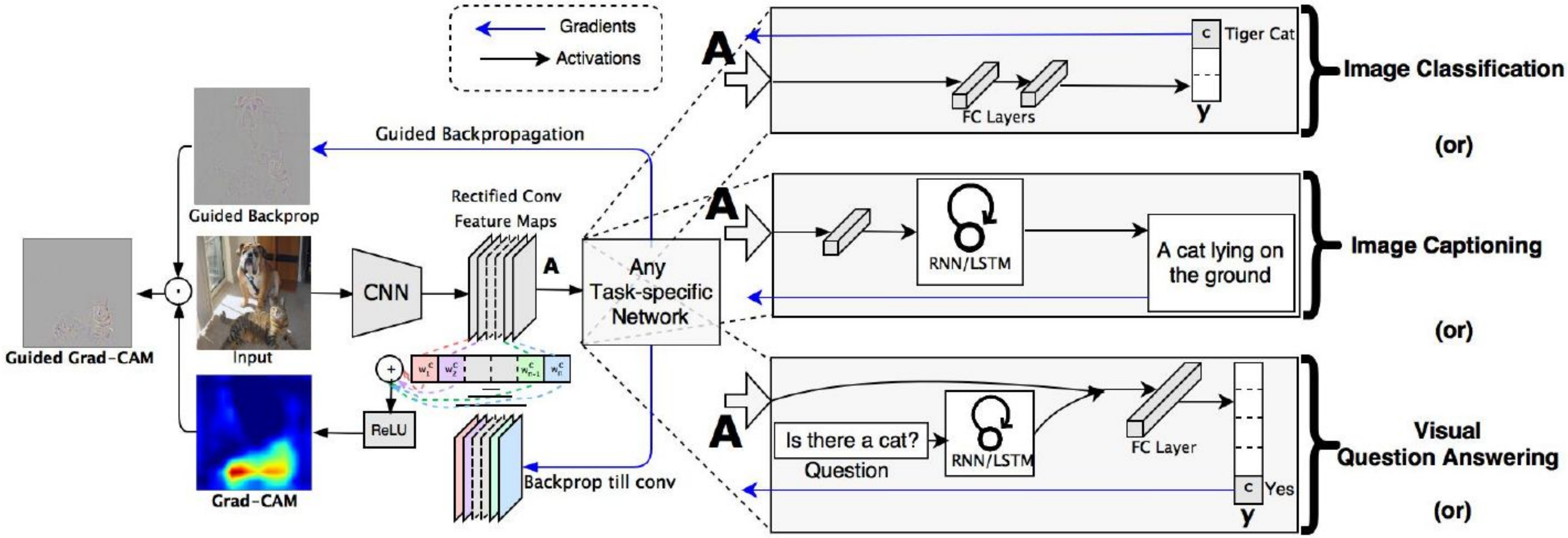
Deep Siamese CNNs for pairwise similarity matching:

- well adapted for person recognition is video surveillance
  - metric learning without training on reference data of the persons being sought
- can also be used for video person ReID, based on video clips



(c) Person ReID: Embedding Networks

Grad-CAM: First extension of CAMs to diverse CNN models

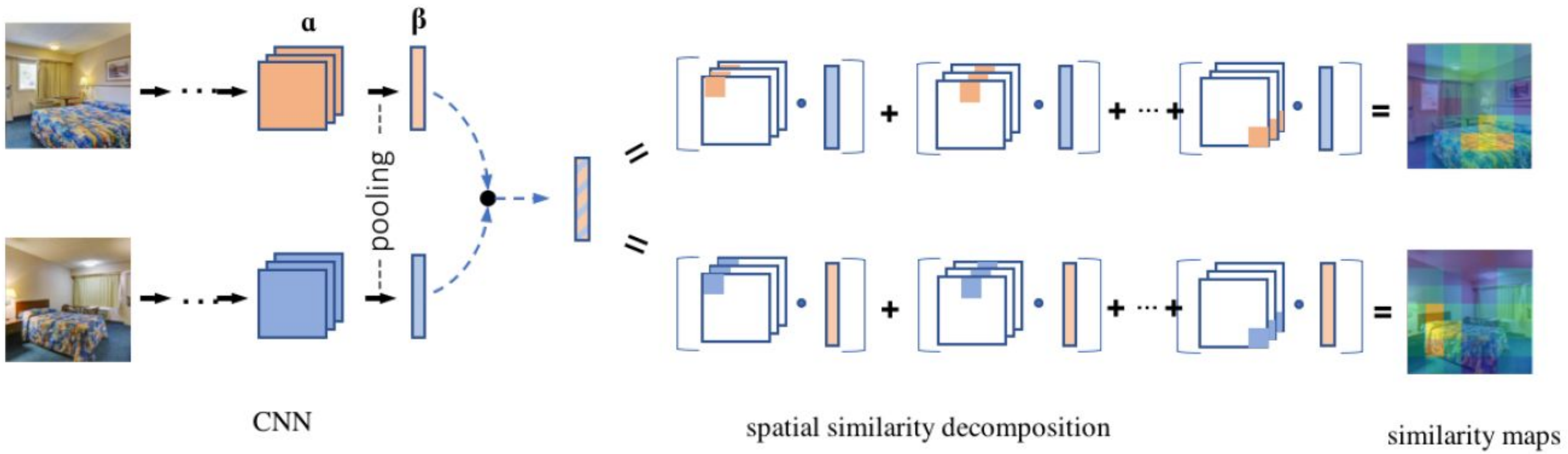


- **Principle**: the gradient for a given class is used after the last conv layer to produce weights for the localization map
- **Not adapted** to embedding networks since it does not produce class-related gradients at inference time - classes not encountered by the model

Source: R. Selvaraju, et al., Grad-CAM: Visual explanations from deep networks via gradient-based localization. CVPR 2017.

(c) Person ReID: Embedding Networks

Visualizing Similarity Networks:



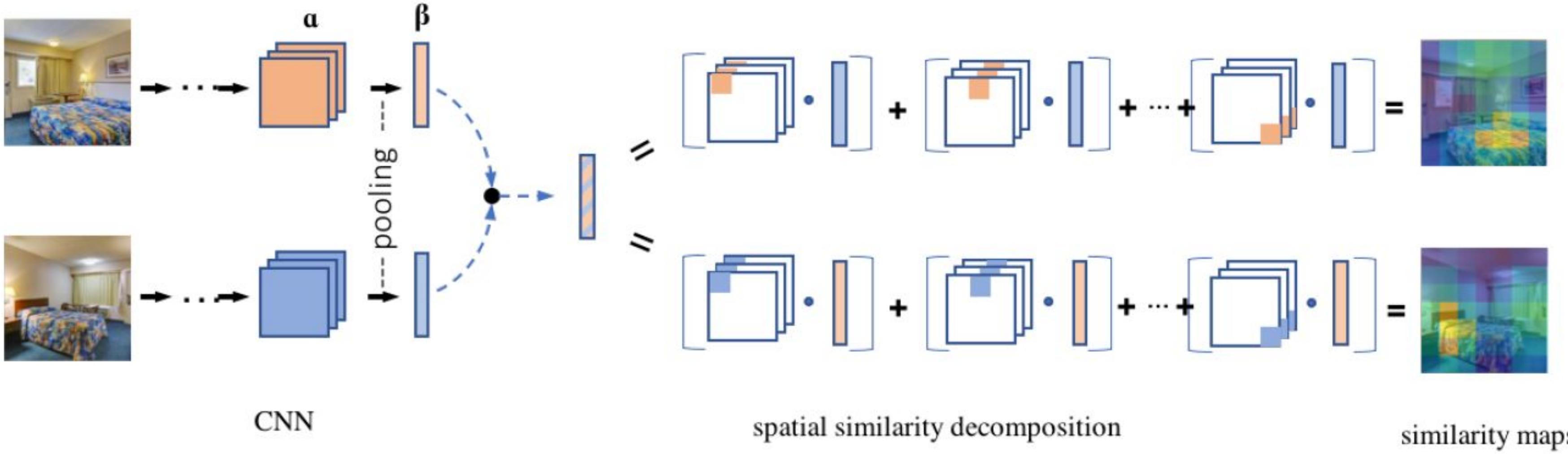
- Produces spatial similarity maps for a pair of images $I^{(i)}$ and $I^{(j)}$
- In these maps, the cosine similarity between two image feature vectors $\beta^{(i)}$ and $\beta^{(j)}$ is spatially decomposed to highlight the contribution of image regions to the overall pairwise similarity:

$$s(\boldsymbol{\beta}^{(i)}, \boldsymbol{\beta}^{(j)}) = \frac{\boldsymbol{\beta}^{(i)} \cdot \boldsymbol{\beta}^{(j)}}{\left\|\boldsymbol{\beta}^{(i)}\right\| \left\|\boldsymbol{\beta}^{(j)}\right\|}$$

Source : A Stylianou, R Souvenir, and R Pless, Visualizing deep similarity networks, WACV 2019.

(c) Person ReID: Embedding Networks

Visualizing Similarity Networks:



- **Approach** – the similarity measure $s(\beta^{(i)}, \beta^{(j)})$ is spatially decomposed to observe the influence of each part of the image in this measure
- Computing the similarity maps depends on the pool operation.
- **Example:** using global average pooling $\boldsymbol{\beta} = \frac{1}{K^2} \sum_{x,y} \boldsymbol{\alpha}_{(x,y)}$

$$s(\boldsymbol{\beta}^{(i)}, \boldsymbol{\beta}^{(j)}) = \frac{\boxed{\boldsymbol{\beta}^{(i)}} \cdot \boldsymbol{\beta}^{(j)}}{\left\|\boldsymbol{\beta}^{(i)}\right\| \left\|\boldsymbol{\beta}^{(j)}\right\|} = \frac{\boldsymbol{\alpha}^{(i)}_{(1,1)} \cdot \boldsymbol{\beta}^{(j)} + \ldots + \boldsymbol{\alpha}^{(i)}_{(K,K)} \cdot \boldsymbol{\beta}^{(j)}}{Z} \longrightarrow$$

arrange terms spatially and visualize as a similarity map

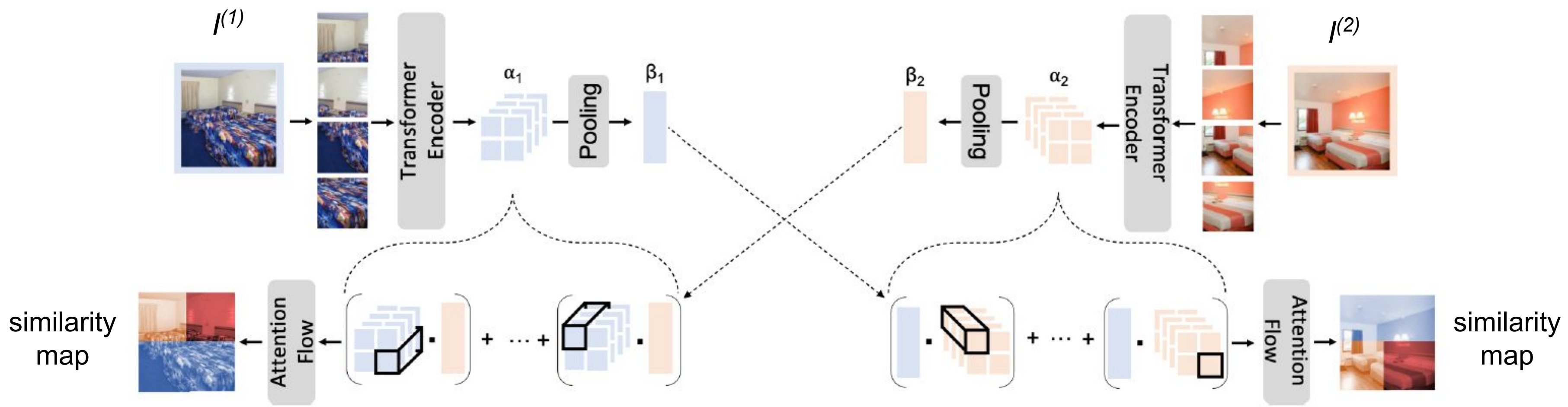
Z: normalizing factor $K^2 \left\|\boldsymbol{\beta}^{(i)}\right\| \left\|\boldsymbol{\beta}^{(j)}\right\|$

Source : A Stylianou, R Souvenir, and R Pless, Visualizing deep similarity networks, WACV 2019.

Le génie pour l'industrie

(c) Person ReID: Embedding Networks

Visualizing Similarity Networks: extend the same idea for interpretation of transformer embedding networks



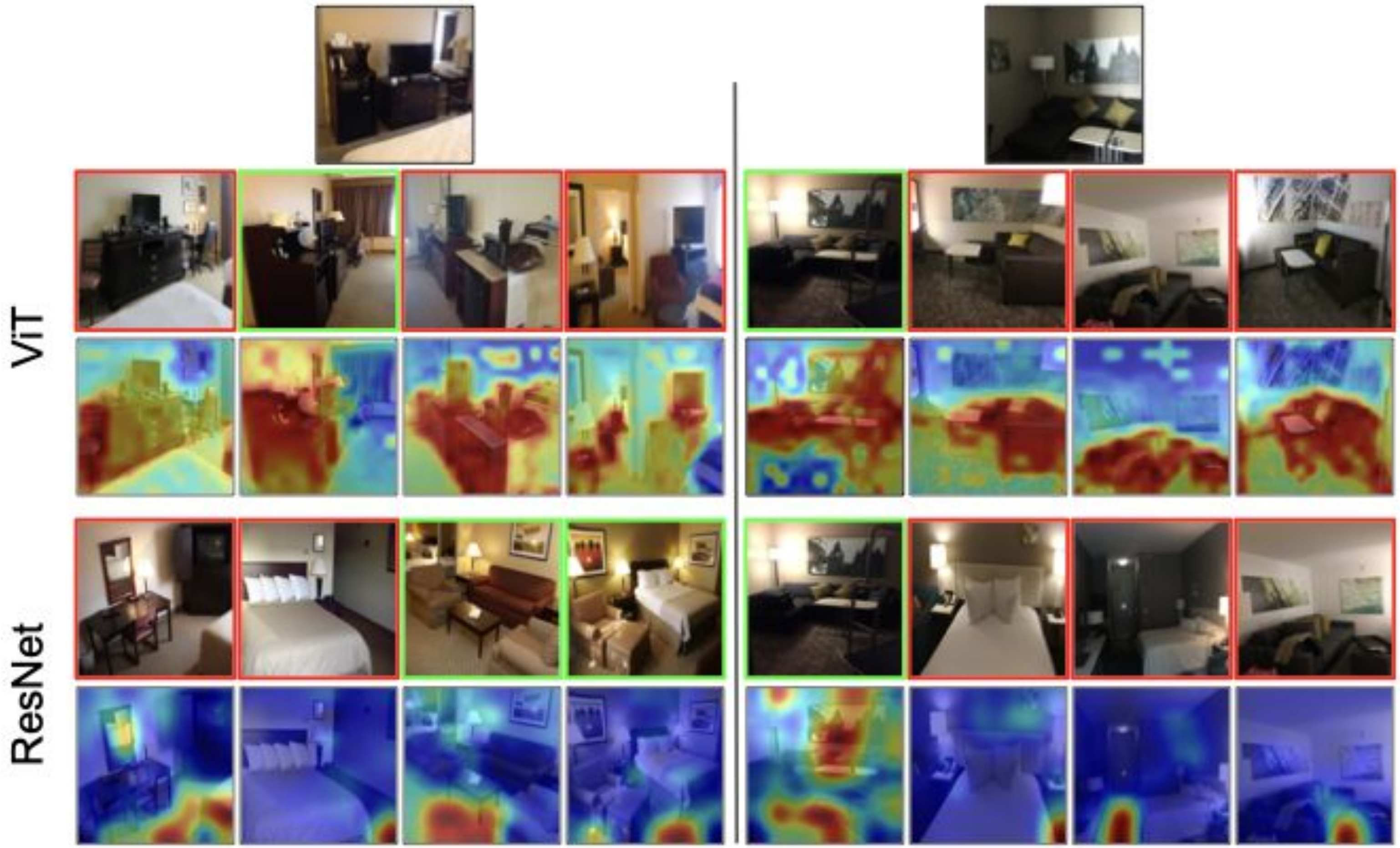
- Instead of using the feature map of the last conv. layer to decompose the similarity $s(\beta^{(i)}, \beta^{(j)})$, it relies the last output tokens
- **Idea** – Use rollout algorithm to approximately match token activation in the similarity measure with the correct spatial image location

$$s(\boldsymbol{\beta}^{(i)}, \boldsymbol{\beta}^{(j)}) = \frac{\boldsymbol{\beta}^{(i)} \cdot \boldsymbol{\beta}^{(j)}}{\left\|\boldsymbol{\beta}^{(i)}\right\| \left\|\boldsymbol{\beta}^{(j)}\right\|}$$

Source: S Black, et al., "Visualizing Paired Image Similarity in Transformer Networks." WACV 2022

(c) Person ReID: Embedding Networks

Visualizing Similarity Networks: extend the same idea for interpretation of transformer embedding networks

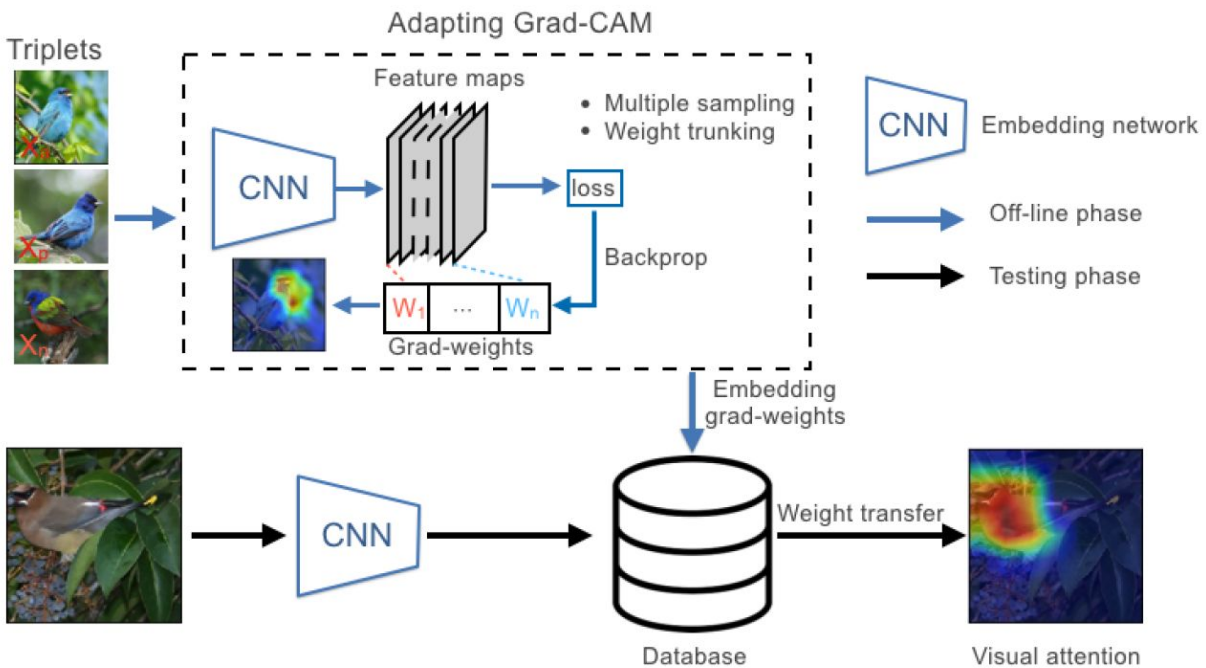


Source: S Black, et al., "Visualizing Paired Image Similarity in Transformer Networks." WACV 2022

(c) Person ReID: Embedding Networks

Adapted Grad-CAM for Embedding Nets:

Replace the per-class gradient by the triplet loss gradient $g(A^k) = \frac{\partial \mathcal{L}_{tri}}{\partial A^k}$.



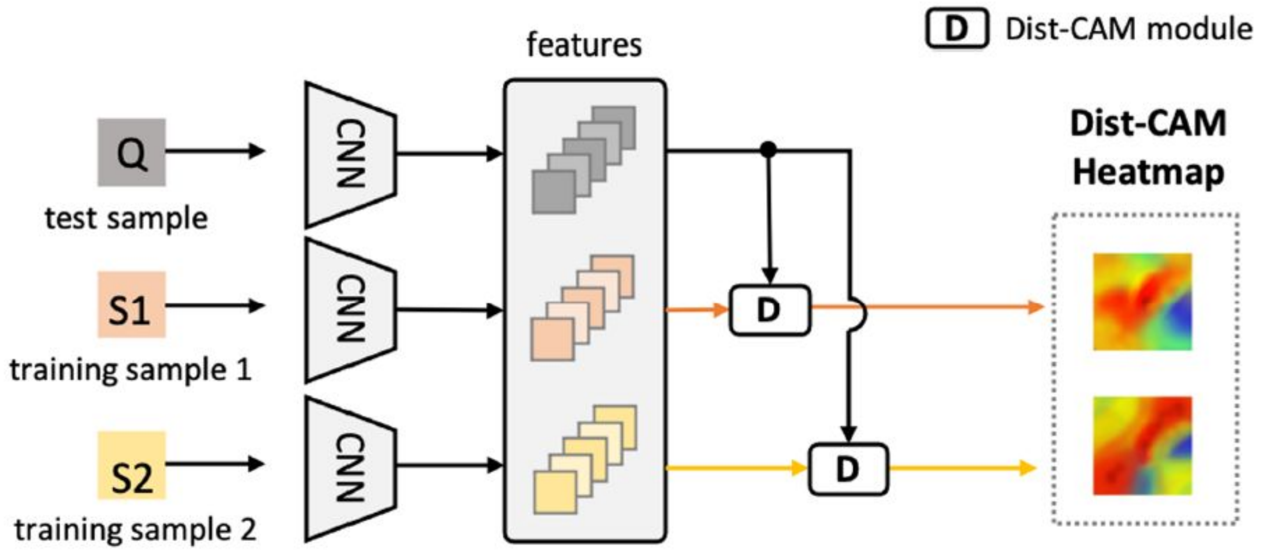
- query the nearest neighbor embedding in the training set for a given test set embedding
- apply the grad-weights of the nearest neighbor to produce the CAM visualization

Source : L Chen, et al., Adapting grad-cam for embedding networks. WCCV 2020.

ÉTS Le génie pour l'industrie

(c) Person ReID: Embedding Networks

Distance-based CAM: Produce a map from the closest training sample weights



- query the closest embedding in the training set for a given test set embedding
- adapt the FC layer weights of the closest training sample regarding each channel distance from the test to the training feature map.

Source: Shen, Yeqing, et al. Distance-Based Class Activation Map for Metric Learning. PRCV 2021.

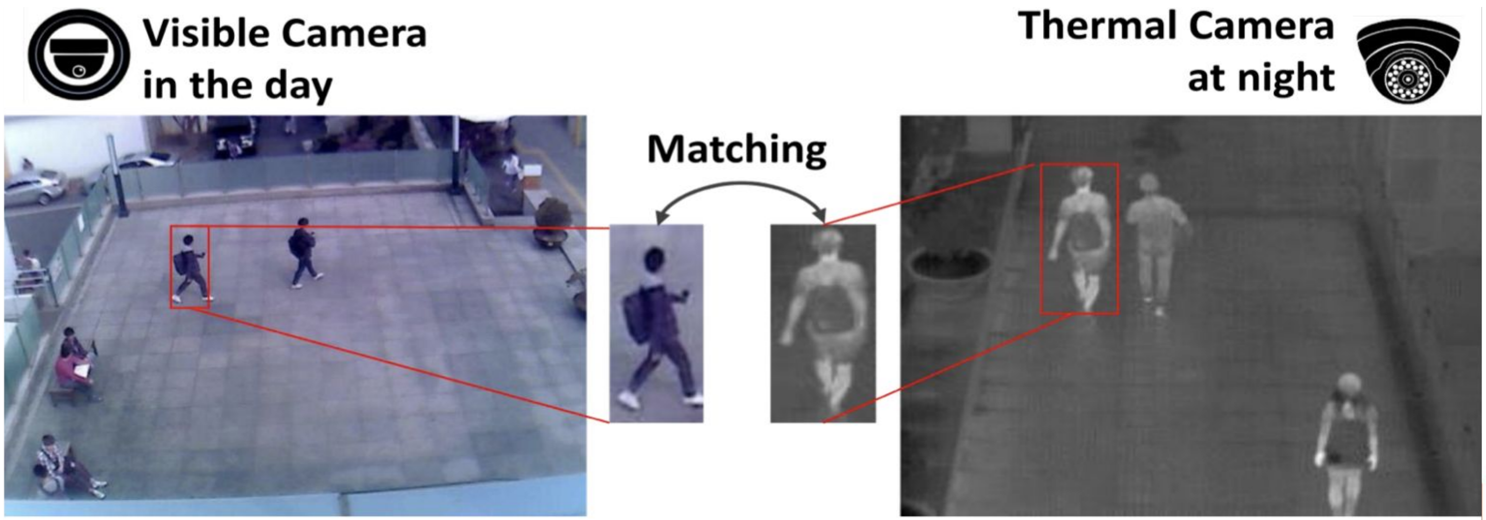
ÉTS Le génie pour l'industrie

(c) Person ReID: Embedding Networks

Visible-Infrared (cross-modality) Person ReID:

- RGB can provide high quality color information (dependant on illumination)
- IR works well at night, or under poor lighting conditions

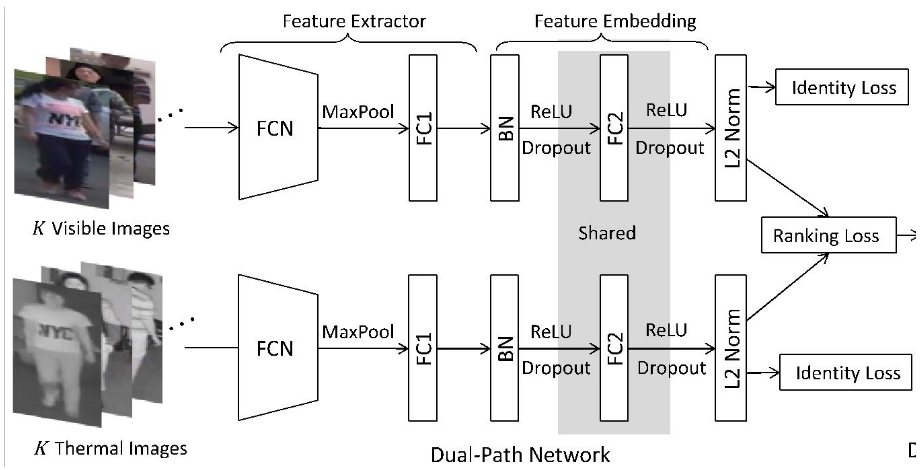
Objectives: match a same person across a network of different spectrum cameras (RGB and IR) - *large domain shift*

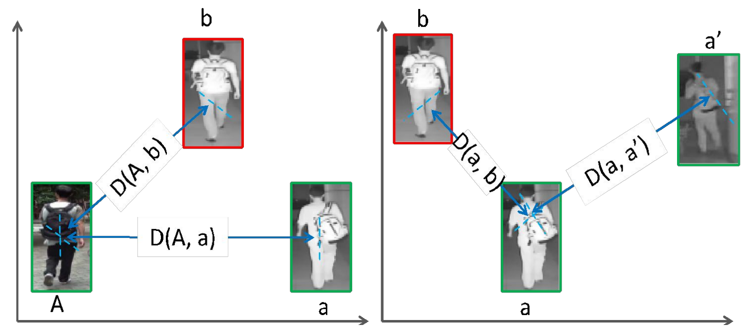


(c) Person ReID: Embedding Networks

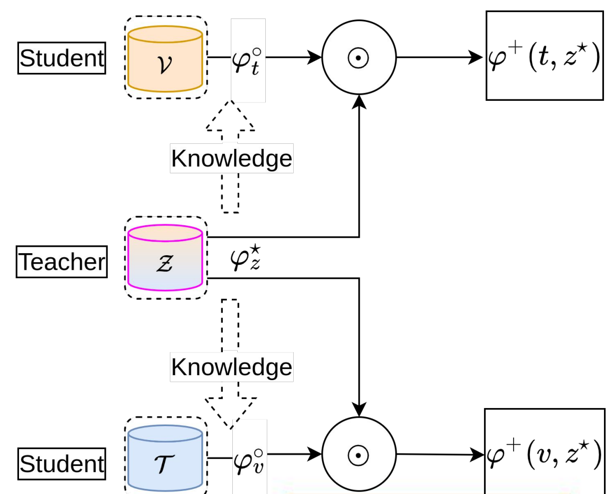
Visible-Infrared (cross-modality) Person ReID:

- **BDTR framework:** dual-path network for feature extraction and bi-directional dual-constrained top-ranking loss





- Network trained using privileged information from intermediate virtual domains (teacher, *Z*), between infrared and visual (students, *V* and *T*)

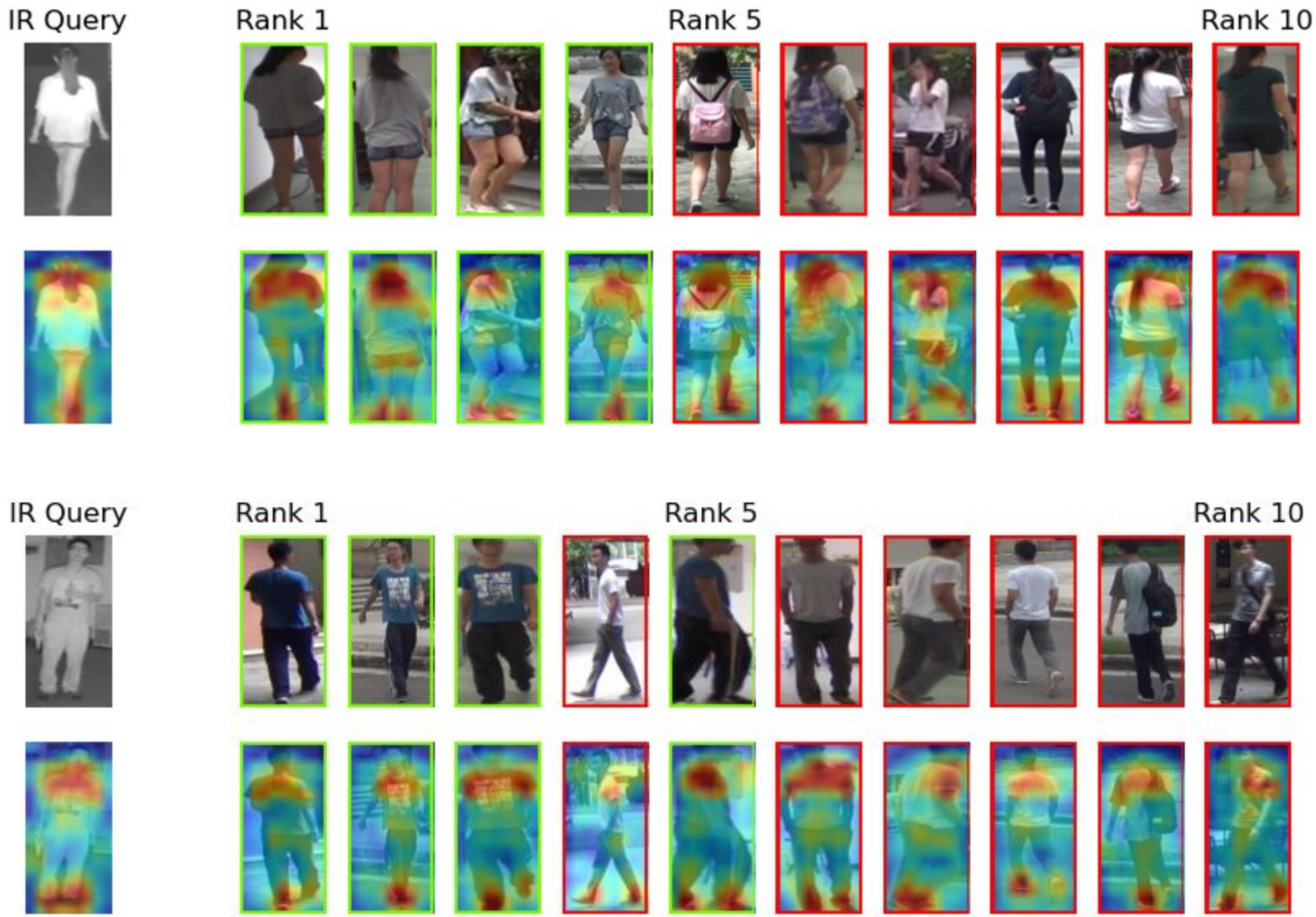


Source: M Ye, et al., Bi-directional center-constrained top-ranking for visible thermal person re-identification. *IEEE TIFS* 2019
Source: M Alehdaghi, et al., Visible-Infrared Person Re-Identification Using Privileged Intermediate Information. ECCVw 2022.

(c) Person ReID: Embedding Networks

Visible-infrared Person ReID:

Visual results on SYSU dataset using "similarity CAM" (Stylianou, 2019)



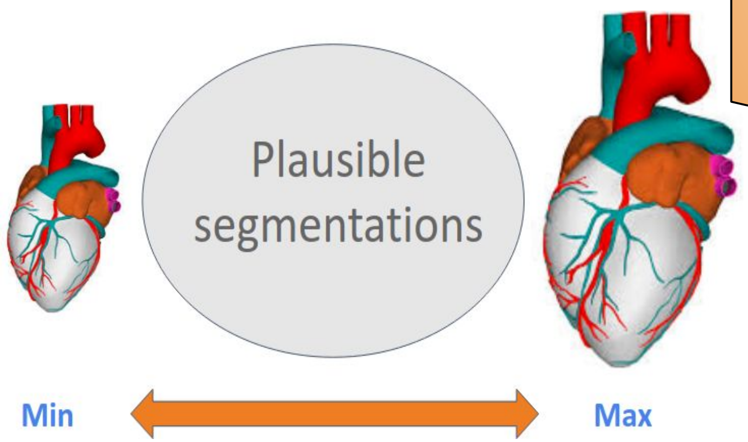
Source: M Alehdaghi, et al., Visible-Infrared Person Re-Identification Using Privileged Intermediate Information. ECCVw 2022.

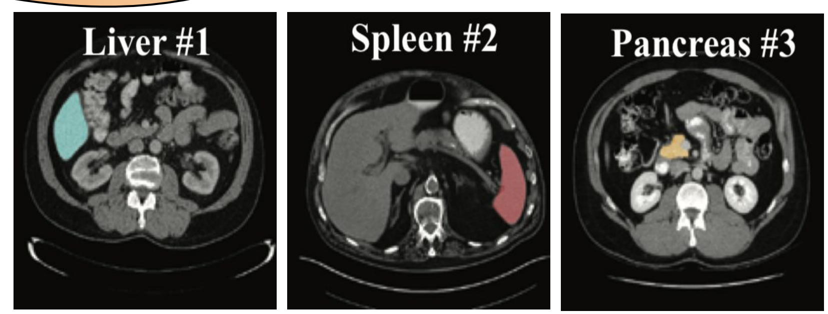
Medical Semantic Segmentation

Anatomical Constraints

Data meets domain knowledge

A lot of priors in medical imaging

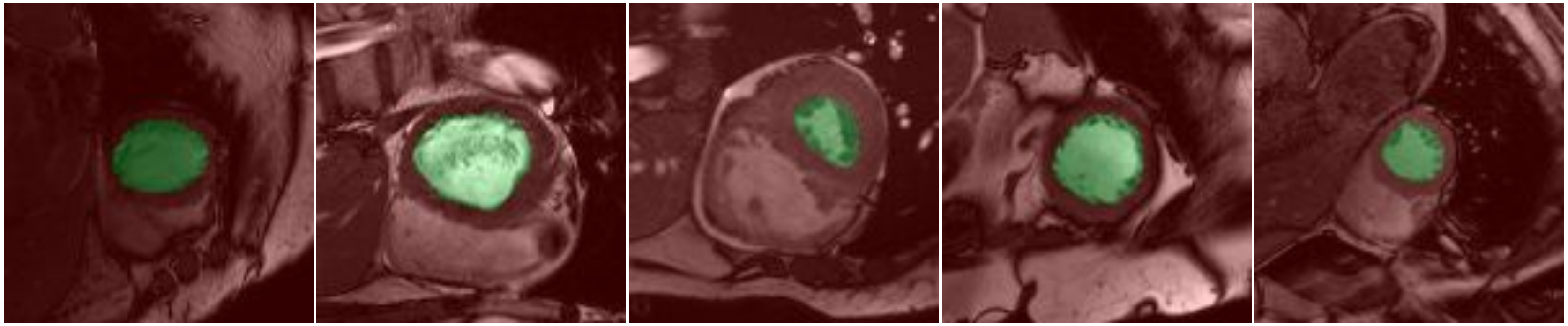




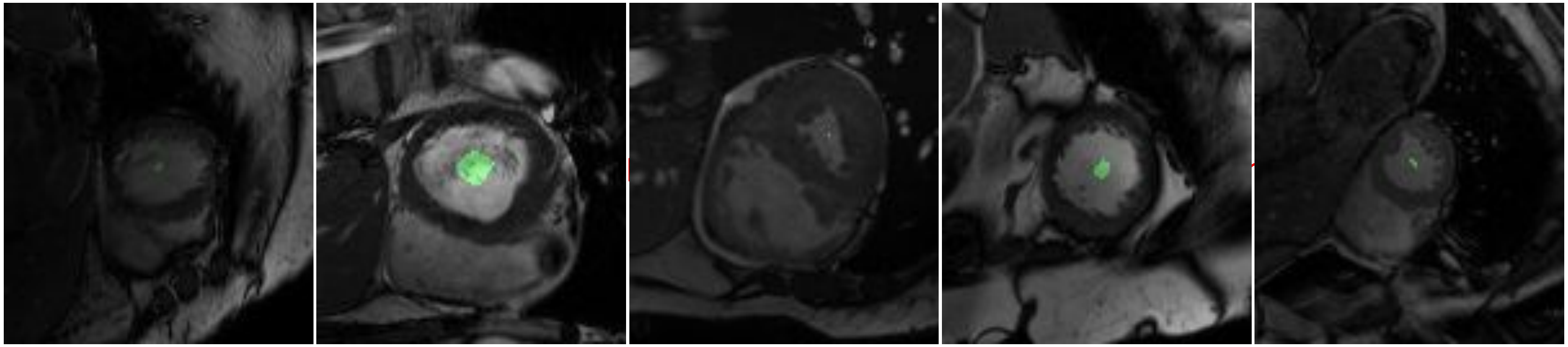
Anatomical priors (e.g., shapes)

Partially labeled data
(e.g., exploiting organ relationships)

Example: Left Ventricle Segmentation in Cardiac MRI with Volumetric Constraints



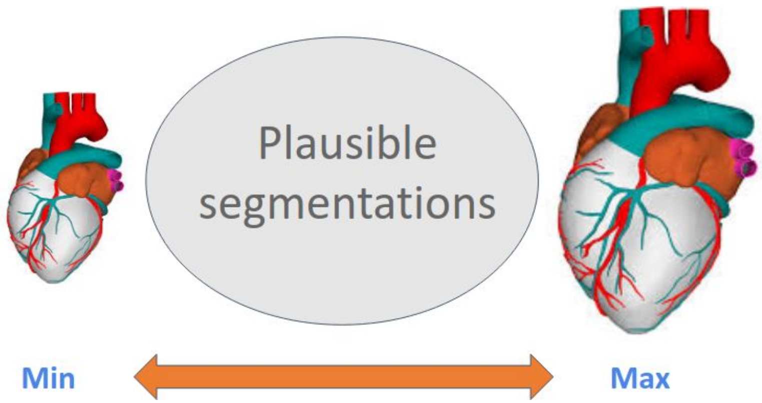
Full annotations



Partial annotations for cross-entropy

Example: Left Ventricle Segmentation in Cardiac MRI with Volumetric Constraints

$$\min \leq \sum_{p \in \Omega} s_\theta^{p,c} \leq \max$$

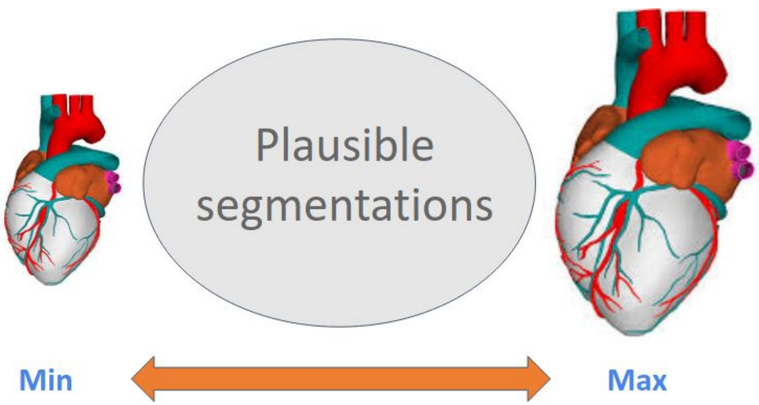


Size information

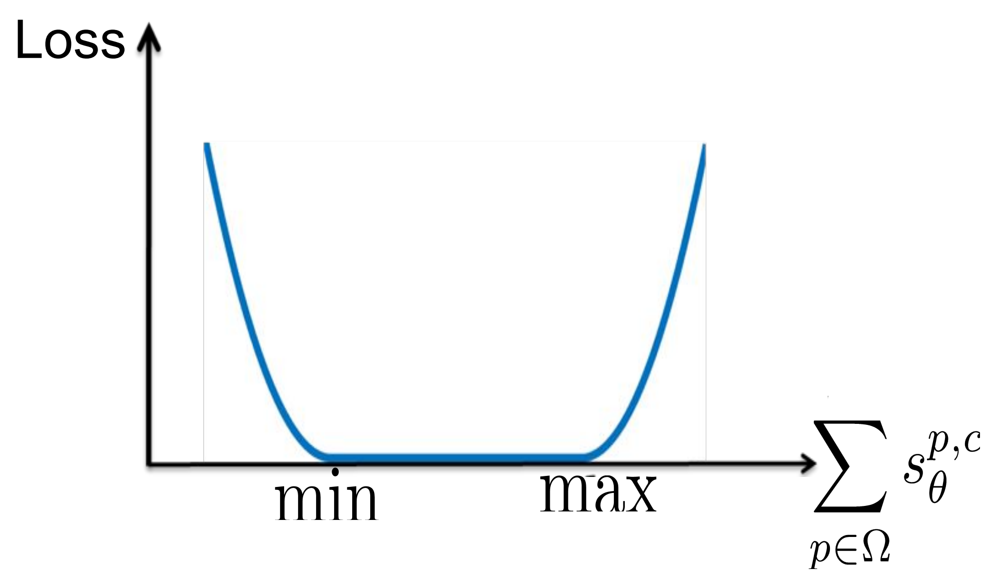
Kervadec et et al., Constrained-CNN Losses for Weakly Supervised Segmentation, MedIA'19

Example: Left Ventricle Segmentation in Cardiac MRI with Volumetric Constraints

$$\min \leq \sum_{p \in \Omega} s_\theta^{p,c} \leq \max$$

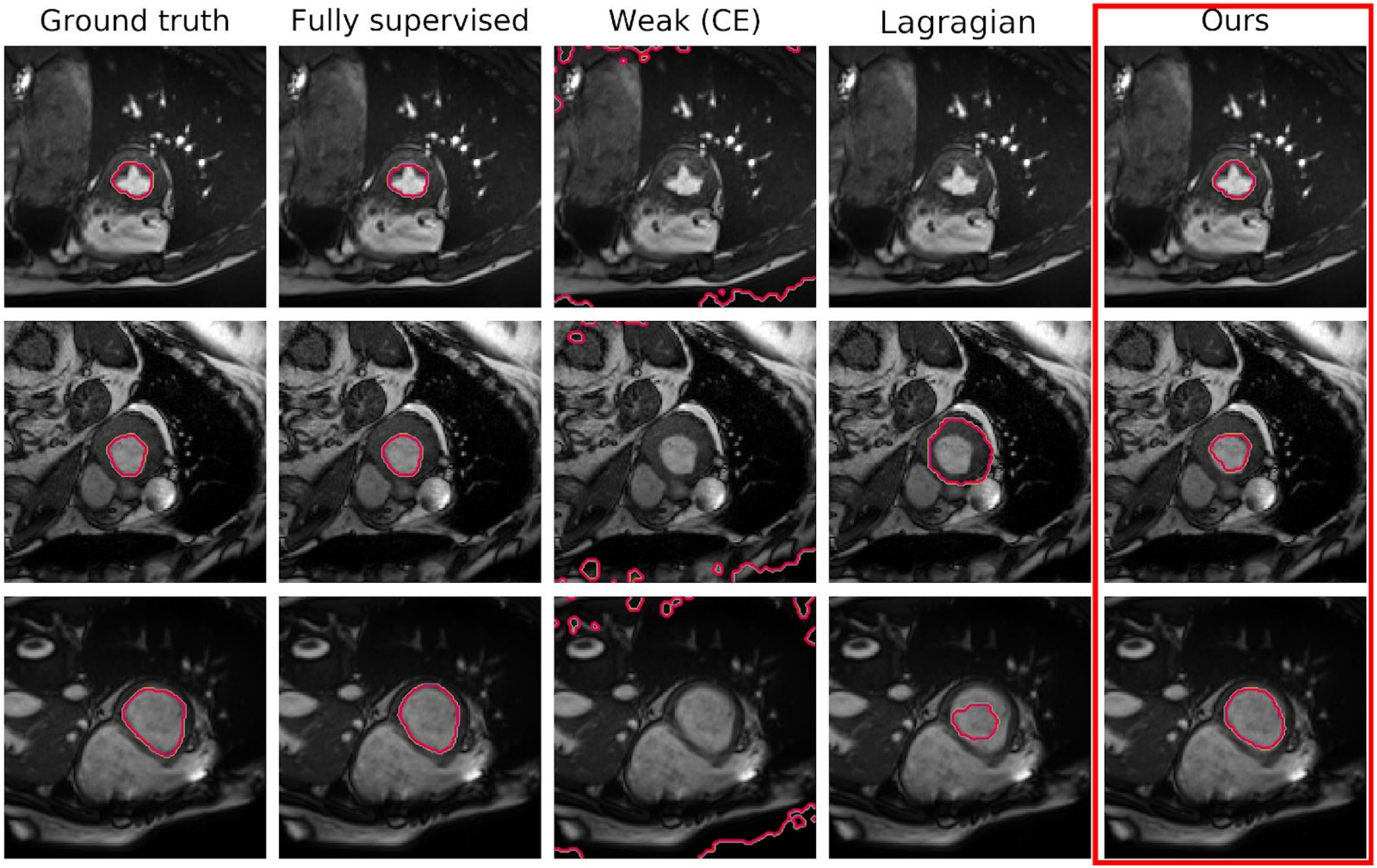


Size information



Kervadec et et al., Constrained-CNN Losses for Weakly Supervised Segmentation, MedIA'19

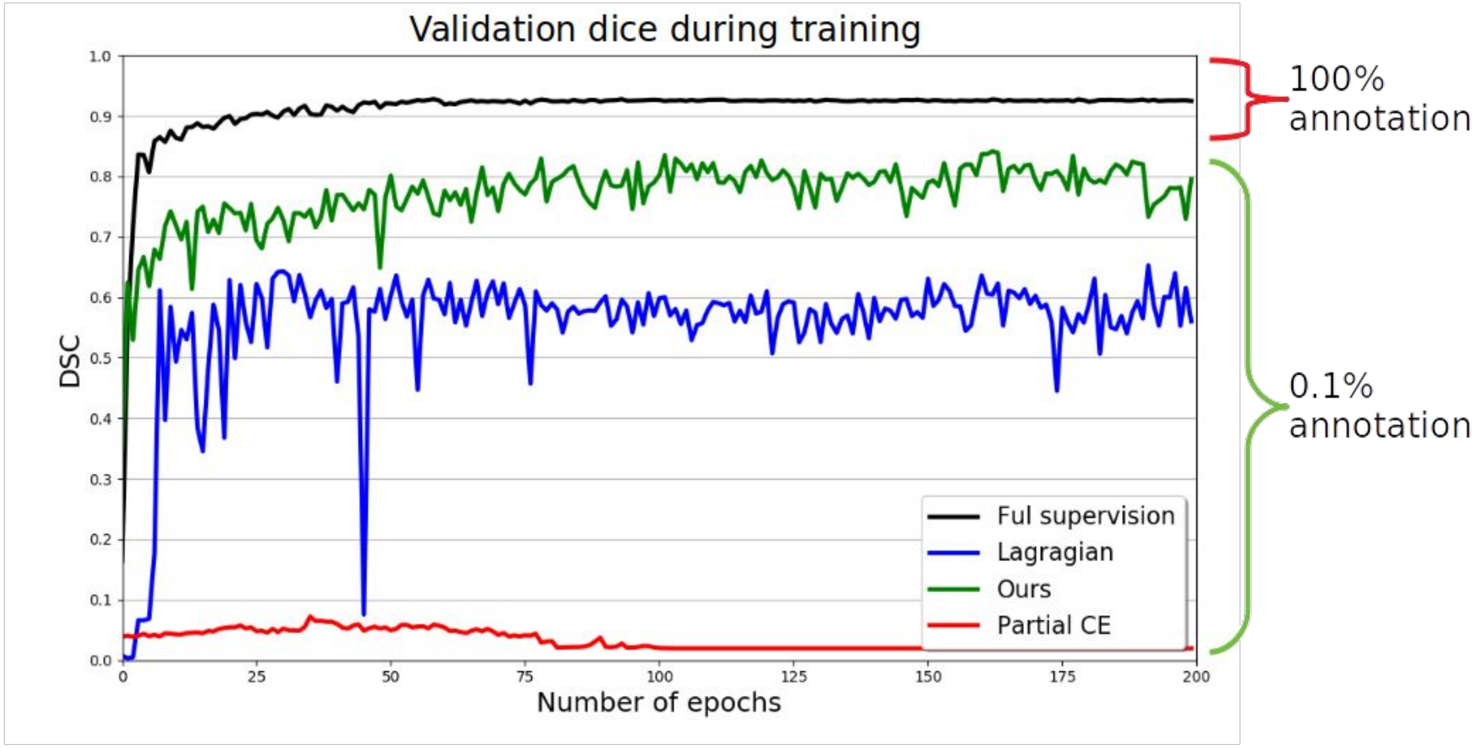
Example: Left Ventricle Segmentation in Cardiac MRI with Volumetric Constraints



Kervadec et et al., Constrained-CNN Losses for Weakly Supervised Segmentation, MedIA’19

Le génie pour l’industrie

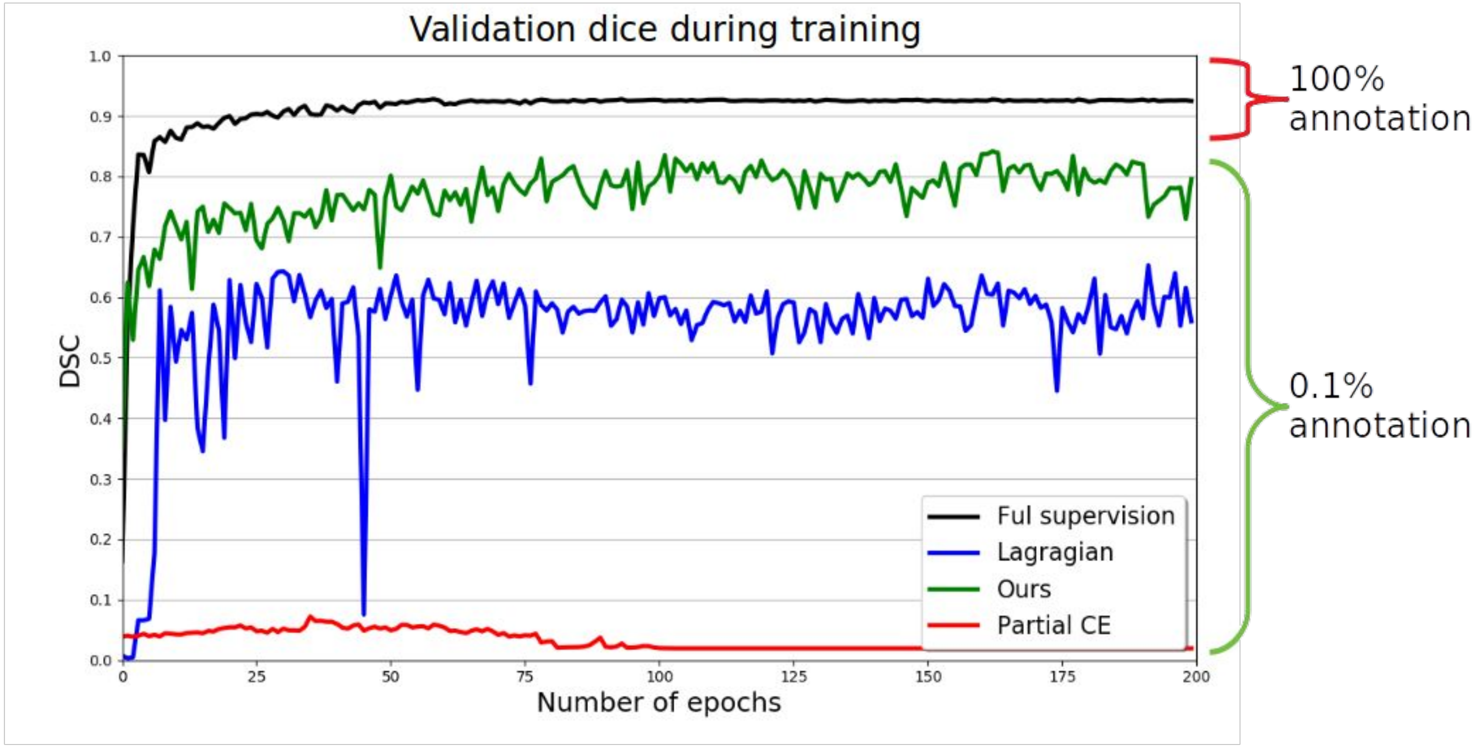
Example: Left Ventricle Segmentation in Cardiac MRI with Volumetric Constraints



Kervadec et et al., Constrained-CNN Losses for Weakly Supervised Segmentation, MedIA’19

Example: Left Ventricle Segmentation in Cardiac MRI with Volumetric Constraints

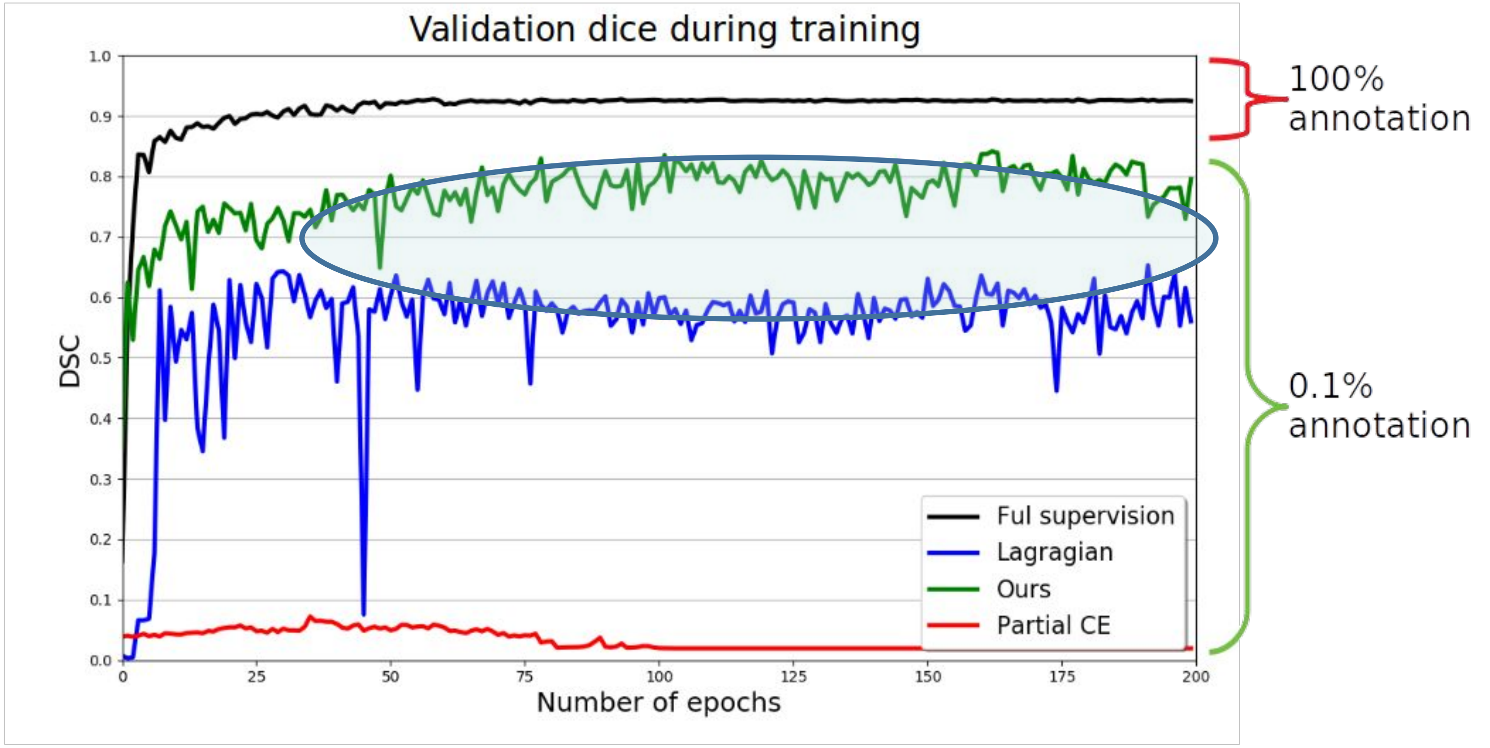
The exciting part: 90% of full supervision Dice with 0.1% of labels



Kervadec et et al., Constrained-CNN Losses for Weakly Supervised Segmentation, MedIA'19

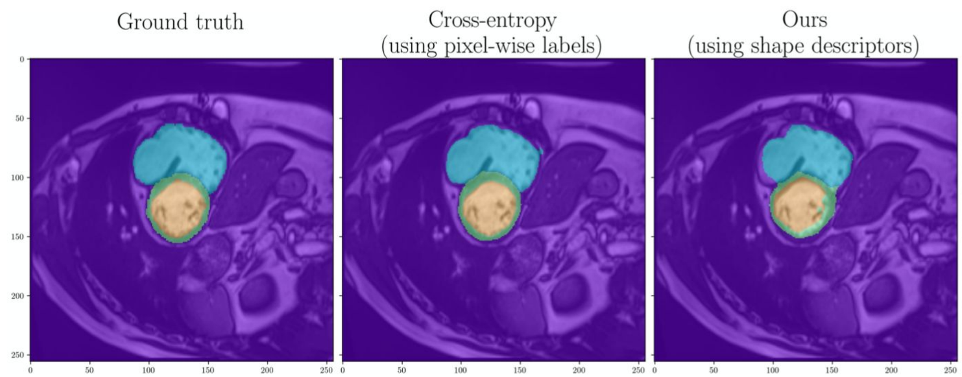
Example: Left Ventricle Segmentation in Cardiac MRI with Volumetric Constraints

The surprising part: Lagrangian optimization is much worse than a simple penalty



Kervadec et et al., Constrained-CNN Losses for Weakly Supervised Segmentation, MedIA'19

Beyond size: Exploring shape priors



(a) A visual comparison of the different supervision methods on the ACDC dataset.

Pixel	Label
0	RV
1	BACKGROUND
2	LV
⋮	
65536	BACKGROUND

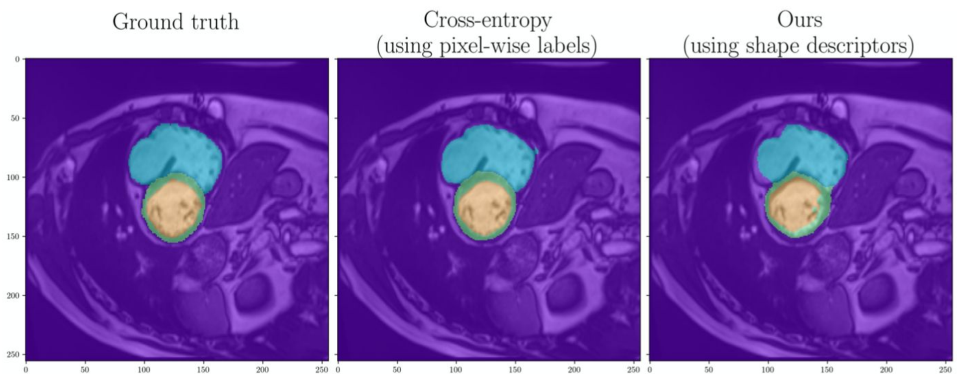
(b) Pixel-wise labels (**65k** discrete values)

Shape descriptor (in pixels)	Class RV	Class MYO	Class LV
Object volume $\mathfrak{V}$	3100	800	1600
Centroid location $\mathfrak{C}$	(125, 80)	(125, 125)	
Avg. dist. to centroid $\mathfrak{D}$	(20, 15)	(15, 20)	(10, 10)
Object length $\mathfrak{L}$	750	1000	500

(c) Shape descriptors (**16** continuous values)

Kervadec et et al., Beyond pixelwise supervision: A few shape descriptors might be surprisingly good! MIDL'21

Beyond size: Exploring shape priors



(a) A visual comparison of the different supervision methods on the ACDC dataset.

$$\mu_{i,j}^{c} = \sum_{p} s_{\theta}^{p,c} x_{p}^{i} y_{p}^{j}$$

Spatial coordinates

Pixel	Label
0	RV
1	BACKGROUND
2	LV
⋮	
65536	BACKGROUND

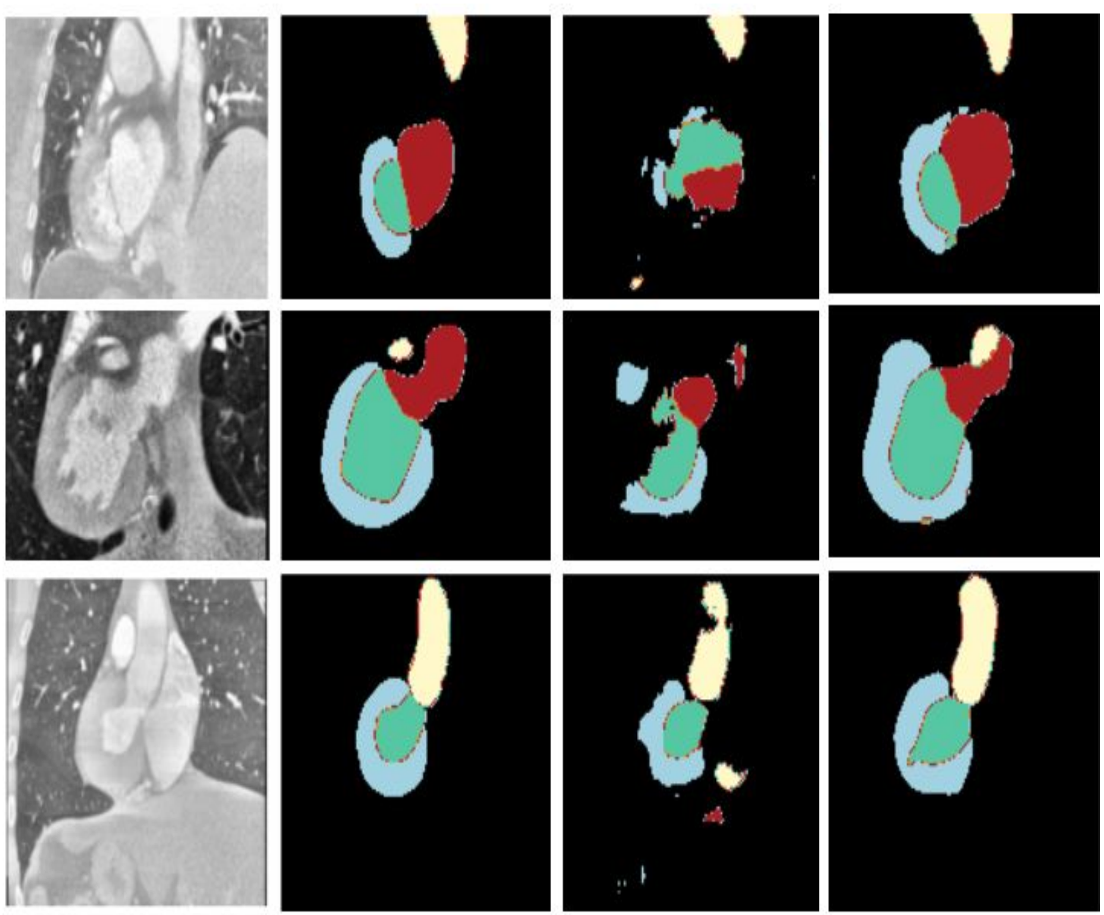
(b) Pixel-wise labels (**65k** discrete values)

Shape descriptor	Class		
(in pixels)	RV	MYO	LV
Object volume $\mathfrak{V}$	3100	800	1600
Centroid location $\mathfrak{C}$	(125, 80)	(125, 125)	
Avg. dist. to centroid $\mathfrak{D}$	(20, 15)	(15, 20)	(10, 10)
Object length $\mathfrak{L}$	750	1000	500

(c) Shape descriptors (**16** continuous values)

Kervadec et et al., Beyond pixelwise supervision: A few shape descriptors might be surprisingly good! MIDL'21

Shape moments: Also powerful in test-time adaptation



Loss optim. w.r.t scale and bias param. of batch norm layers

Bateson et et al., Test-Time Adaptation with Shape Moments for Image Segmentation, MICCAI 2022

Part 5:

Key challenges and future directions

Part 5. Key challenges and future directions

Take away message

- WSOL is mostly done via CAMs

Le génie pour l'industrie

Part 5. Key challenges and future directions

Take away message

- WSOL is mostly done via CAMs

- Bottom-up methods are dominant

Part 2. Review of WSOL methods: Literature

Take away message

- WSOL is mostly done via CAMs
- Bottom-up methods are dominant
- What currently works better:
    - Leveraging low level features
    - Pseudo-labels

Part 5. Key challenges and future directions

Take away message

Ongoing issues of CAMs

Part 5. Key challenges and future directions

Take away message

Ongoing issues of CAMs

- Cover full discriminative objects

Part 5. Key challenges and future directions

Take away message

Ongoing issues of CAMs

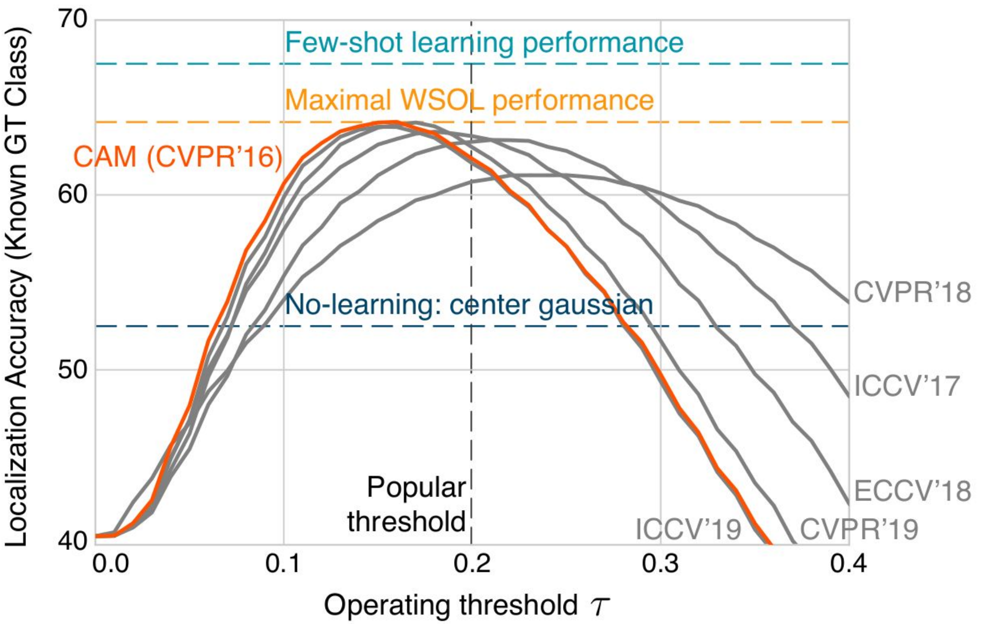
- Cover full discriminative objects
- Deal with background (complex scene, non-salient objects)

Part 5. Key challenges and future directions

Take away message

Ongoing issues of CAMs

- Cover full discriminative objects
- Deal with background (complex scene, non-salient objects)
- Threshold dependence



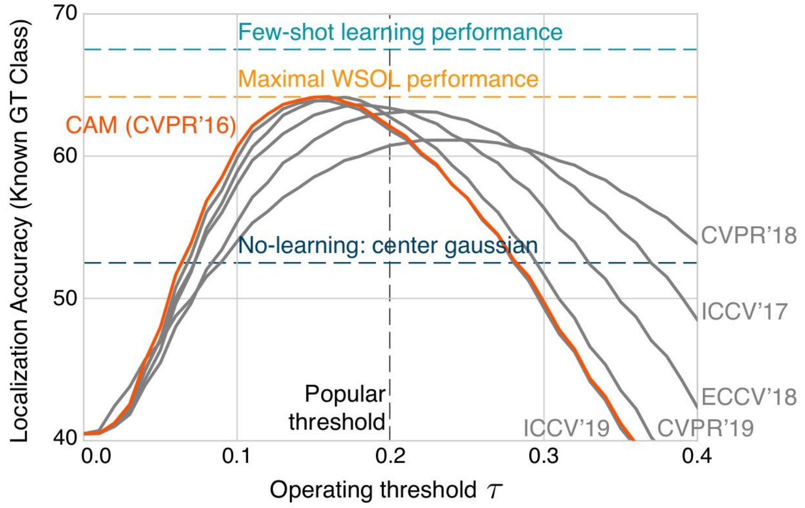
Choe, J., Oh, S. J., Lee, S., Chun, S., Akata, Z., and Shim, H. (2020). Evaluating weakly supervised object localization methods right. In CVPR.

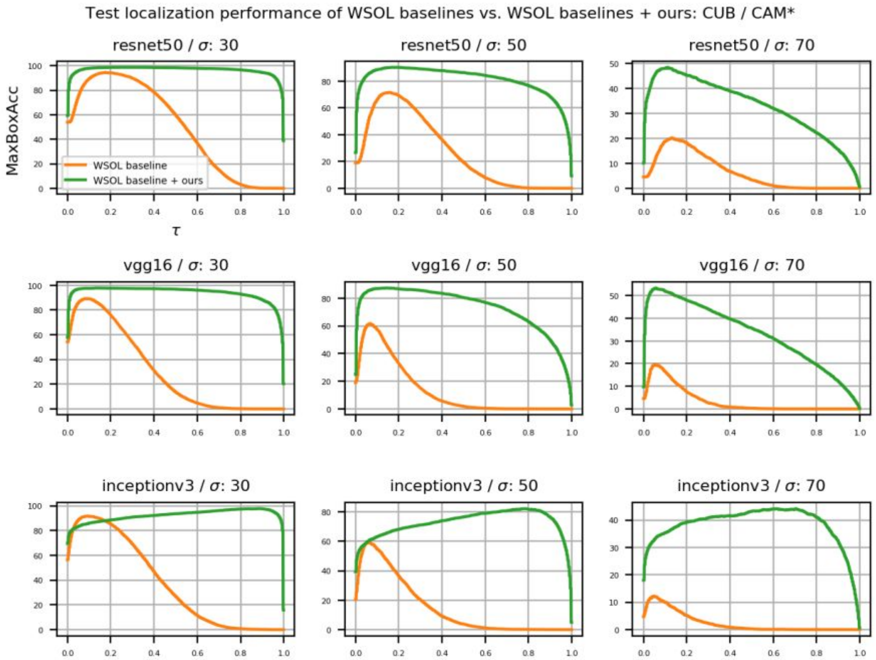
Part 5. Key challenges and future directions

Take away message

Ongoing issues of CAMs

- Cover full discriminative objects
- Deal with background (complex scene, non-salient objects)
- Threshold dependence





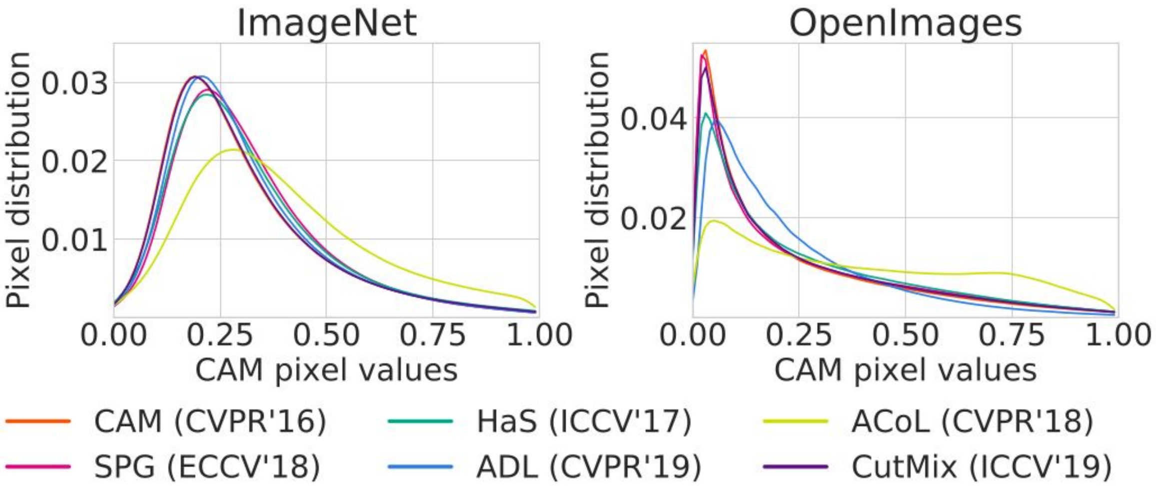
Choe, J., Oh, S. J., Lee, S., Chun, S., Akata, Z., and Shim, H. (2020). Evaluating weakly supervised object localization methods right. In CVPR.
Belharbi, S., Sarraf, A., Pedersoli, M., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022c). F-cam: Full resolution class activation maps via guided parametric upscaling. In WACV.

Part 5. Key challenges and future directions

Take away message

Ongoing issues of CAMs

- Cover full discriminative objects
- Deal with background (complex scene, non-salient objects)
- Threshold dependence
- Uncertainty



Choe, J., Oh, S. J., Lee, S., Chun, S., Akata, Z., and Shim, H. (2020). Evaluating weakly supervised object localization methods right. In CVPR.

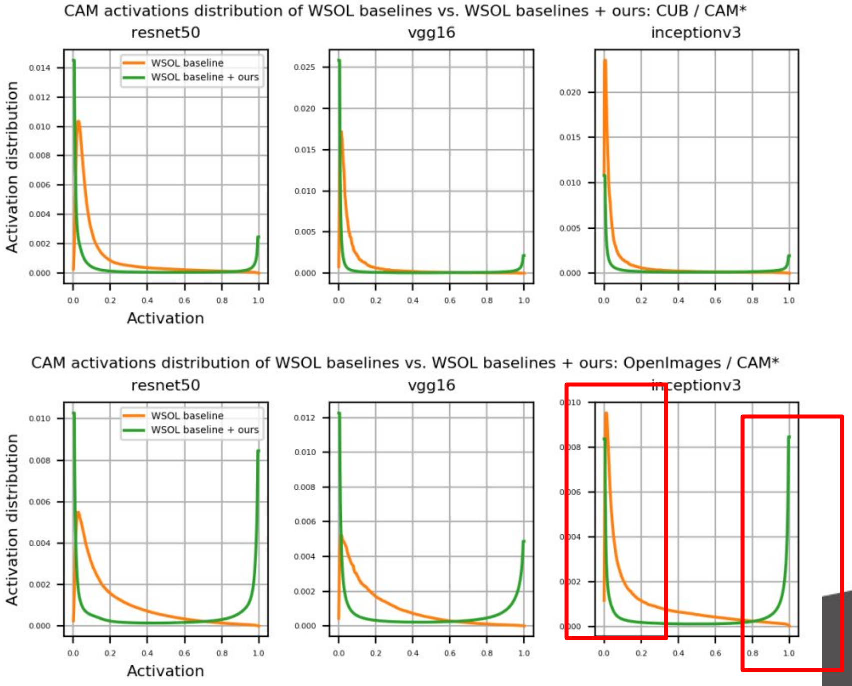
Part 5. Key challenges and future directions

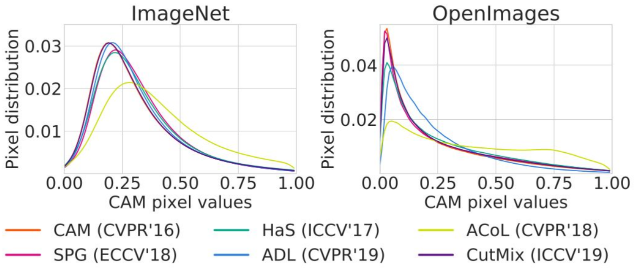
Take away message

Ongoing issues of CAMs

- Cover full discriminative objects
- Deal with background (complex scene, non-salient objects)
- Threshold dependence
- Uncertainty

2 modes





Choe, J., Oh, S. J., Lee, S., Chun, S., Akata, Z., and Shim, H. (2020). Evaluating weakly supervised object localization methods right. In CVPR.

Part 5. Key challenges and future directions

Take away message

Ongoing issues of CAMs

- Cover full discriminative objects
- Deal with background (complex scene, non-salient objects)

- Threshold dependence
- Uncertainty
- Tiny objects

https://library.sportingnews.com/styles/crop_style_16_9_desktop/s3/2022-04/john%20tavares%20david%20savard%20040822.jpg?itok=QNvrM1DT

Part 2. Review of WSOL methods: Literature

Take away message

Ongoing issues of CAMs

- Cover full discriminative objects
- Deal with background (complex scene, non-salient objects)
- Threshold dependence
- Uncertainty
- Tiny objects

Class: ‘Puck’

Part 5. Key challenges and future directions

Take away message

Ongoing issues of CAMs

- Cover full discriminative objects
- Deal with background (complex scene, non-salient objects)
- Threshold dependence
- Uncertainty
- Tiny objects
- Objects co-occurrence

Class: ‘Puck’

Le génie pour l'industrie

https://library.sportingnews.com/styles/crop_style_16_9_desktop/s3/2022-04/john%20tavares%20david%20savard%20040822.jpg?itok=QNvrM1DT

Part 5. Key challenges and future directions

Take away message

WSOL datasets: Saturation

- CUB dataset: ~97% MaxBoxAcc
- Imagenet-1k dataset: ~70% MaxBoxAcc

Code

- **Deep Weakly-Supervised Learning Methods for Classification and Localization in Histology Images: A Survey**. Rony et al. 2022. **Code**: https://github.com/jeromerony/survey_wsl_histology
- **F-cam: Full resolution class activation maps via guided parametric upscaling**. Belharbi, S., Sarraf, A., Pedersoli, M., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022). In WACV. **Code**: https://github.com/sbelharbi/fcam-wsol
- **Negative evidence matters in interpretable histology image classification**. Belharbi, S., Pedersoli, M., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022). In Medical Imaging with Deep Learning (MIDL). **Code**: https://github.com/sbelharbi/negev
- **Deep interpretable classification and weakly-supervised segmentation of histology images via max-min uncertainty**. Belharbi, S., Rony, J., Dolz, J., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022). IEEE Transactions on Medical Imaging, 41:702–714. **Code**: https://github.com/sbelharbi/deep-wsl-histo-min-max-uncertainty
- **Convolutional stn for weakly supervised object localization and beyond**. Meethal, A., Pedersoli, M., Belharbi, S., and Granger, E. (2020). In ICPR. **Code**: https://github.com/akhilpm/ConvSTN
- **TCAM: Temporal Class Activation Maps for Object Localization in Weakly-Labeled Unconstrained Videos**. Belharbi, S., Ben Ayed, I., McCaffrey, L., and Granger, E.. 2022. **Code (soon!)**: https://github.com/sbelharbi/tcam-wsol-video
- **Holistic Guidance for Occluded Person Re-Identification**. M. Kiran, R G. Praveen, L. T. Nguyen-Meidine, S. Belharbi, L.-A. Blais-Morin, E. Granger. BMVC 2021. **Code**: https://github.com/madhukiranets/HolisitcGuidanceOccReID2
- **Deep Active Learning for Joint Classification & Segmentation with Weak Annotator**. S. Belharbi, I. Ben Ayed, L. McCaffrey, E. Granger. WACV 2021. **Code**: https://github.com/sbelharbi/deep-active-learning-for-joint-classification-and-segmentation-with-weak-annotator
- **Constrained-CNN Losses for Weakly Supervised Segmentation**. Kervadec et et al, MedIA'19. **Code**: https://github.com/LIVIAETS/SizeLoss_WSS
- **Beyond pixelwise supervision: A few shape descriptors might be surprisingly good!**. Kervadec et et al.,MIDL'21. **Code**: https://github.com/hkervadec/shape_descriptors
- **Evaluating Weakly Supervised Object Localization Methods Right.** J. Choe, S. Joon Oh, S. Lee, S. Chun, Z. Akata, H. Shim. CVPR 2020. **Code**: https://github.com/clovaai/wsolevaluation

References

- Bearman, Amy, et al. "**What's the point: Semantic segmentation with point supervision.**" ECCV 2016.
- Z. Zhou. '**A brief introduction to weakly supervised learning**.' National Science Review, 5(1):44–53, 2018.
- V. Cheplygina et al., '**Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis**,' Medical Image Analysis, 2019.
- **Learning Deep Features for Discriminative Localization**. Zhou et al. CVPR, 2016
- **Evaluating Weakly Supervised Object Localization Methods Right**. Choe et al. CVPR 2020.
- **Deep Weakly-Supervised Learning Methods for Classification and Localization in Histology Images: A Survey**. Rony et al. 2022.
- Lin, M., Chen, Q., and Yan, S. (2013). **Network in network**. coRR, abs/1312.4400.
- Durand, T., Mordan, T., Thome, N., et al. (2017). **Wildcat: Weakly supervised learning of deep convnets for image classification, pointwise localization and segmentation.** In CVPR.
- Singh, K. and Lee, Y. (2017). **Hide-and-seek: Forcing a network to be meticulous for weakly-supervised object and action localization**. In ICCV.
- Li, K., Wu, Z., Peng, K., et al. (2018). **Tell me where to look: Guided attention inference network**. In CVPR.
- Choe, J. and Shim, H. (2019). **Attention-based dropout layer for weakly supervised object localization**. In CVPR.
- Xue, H., Liu, C., Wan, F., Jiao, J., Ji, X., and Ye, Q. (2019). **Danet: Divergent activation for weakly supervised object localization**. In ICCV.
- Zhang, C., Cao, Y., and Wu, J. (2020a). **Rethinking the route towards weakly supervised object localization**. In CVPR.
- X.-S. Wei, C.-L. Zhang, J. Wu, C. Shen, and Z.-H. Zhou. **Unsupervised object discovery and co-localization by deep descriptor transformation**. PR, 88:113–126, 2019.
- Wei, J., Wang, Q., Li, Z., Wang, S., Zhou, S. K., and Cui, S. (2021). **Shallow feature matters for weakly supervised object localization**. In CVPR.
- Cao, C., Liu, X., Yang, Y., et al. (2015). **Look and think twice: Capturing top-down visual attention with feedback convolutional neural networks**. In ICCV.
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). **Grad-cam: Visual explanations from deep networks via gradient-based localization**. In ICCV.
- Wang, H., Wang, Z., Du, M., Yang, F., Zhang, Z., Ding, S., Mardziel, P., and Hu, X. (2020). **Score-cam: Score-weighted visual explanations for convolutional neural networks**. In CVPR workshop.
- **Spatial Transformer Networks (STN)**, M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transformer networks," in eurIPS, 2015.

References

- Meethal, A., Pedersoli, M., Belharbi, S., and Granger, E. (2020). **Convolutional stn for weakly supervised object localization and beyond**. In ICPR.
- Belharbi, S., Sarraf, A., Pedersoli, M., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022). **F-cam: Full resolution class activation maps via guided parametric upscaling**. In WACV.
- F. Yu, V. Koltun, and T. Funkhouser, **Dilated residual networks**, CVPR 2017
- Oquab, M., et al., **Is object localization for free?-weakly-supervised learning with CNNs**. In CVPR 2015
- Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., & Joulin, A. (2021). **Emerging properties in self-supervised vision transformers**. In ICCV.
- Gao, W., Wan, F., Pan, X., Peng, Z., Tian, Q., Han, Z., ... & Ye, Q. (2021). T**s-cam: Token semantic coupled attention map for weakly supervised object localization**. In ICCV..
- Figures from [Zhang et al., **A Curriculum Domain Adaptation Approach to the Semantic Segmentation of Urban Scenes,** TPAMI 2019
- Bateson et al., **Constrained Domain Adaptation for Image Segmentation**, TMI'21
- Lin et al. **Scribblesup: Scribble-supervised convolutional networks for semantic segmentation**, CVPR 2016
- Tang et al., **On regularized losses for weakly supervised segmentation**, ECCV 2018
- Qu et al., **Weakly Supervised Deep Nuclei Segmentation using Points Annotation in Histopathology Images**, MIDL 2019
- Ji et al., **Scribble-Based Hierarchical Weakly Supervised Learning for Brain Tumor Segmentation**, MICCAI 2019
- Grandvalet & Bengio, **Semi-supervised learning by entropy minimization**, NIPS 2005
- Gomes et al., **Discriminative clustering by regularized information maximization**, NIPS 2010
- **ADVENT: Adversarial Entropy Minimization for Domain Adaptation in Semantic Segmentation**, CVPR 2019
- **Source-relaxed domain adaptation for segmentation**, MICCAI 2020
- Belharbi, S., Rony, J., Dolz, J., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022). **Deep interpretable classification and weakly-supervised segmentation of histology images via max-min uncertainty**. IEEE Transactions on Medical Imaging, 41:702–714.
- Belharbi, S., Pedersoli, M., Ben Ayed, I., McCaffrey, L., and Granger, E. (2022). **Negative evidence matters in interpretable histology image classification**. In Medical Imaging with Deep Learning (MIDL).
- A Bhuiyan, et al., **Pose Guided Gated Fusion for Person Re-identification**, WACV 2020.
- R. Selvaraju, et al., **Grad-CAM: Visual explanations from deep networks via gradient-based localization**. CVPR 2017.
- A Stylianou, R Souvenir, and R Pless, **Visualizing deep similarity networks**, WACV 2019.

ÉTS Le génie pour l'industrie

References

- S Black, et al., "**Visualizing Paired Image Similarity in Transformer Networks**." WACV 2022
- L Chen, et al., **Adapting grad-cam for embedding networks.** WCCV 2020.
- Shen, Yeqing, et al. **Distance-Based Class Activation Map for Metric Learning**. PRCV 2021.
- M Ye, et al., **Bi-directional center-constrained top-ranking for visible thermal person re-identification**. IEEE TIFS 2019
- M Alehdaghi, et al., **Visible-Infrared Person Re-Identification Using Privileged Intermediate Information**. ECCVw 2022
- Kervadec et et al., **Constrained-CNN Losses for Weakly Supervised Segmentation**, MedIA’19
- Kervadec et et al., **Beyond pixelwise supervision: A few shape descriptors might be surprisingly good!** MIDL’21
- Bateson et et al., **Test-Time Adaptation with Shape Moments for Image Segmentation**, MICCAI 2022
- **TCAM: Temporal Class Activation Maps for Object Localization in Weakly-Labeled Unconstrained Videos**. 2022 [soon]

Le génie pour l’industrie

Discussion

Le génie pour l'industrie

Slides

These slides will be available at:
https://sbelharbi.github.io/publications/icpr-tutorial-wsl-2022/slides.pdf