# Beyond Labels!

Weakly-supervised, Continual, and Semi-supervised Learning for Earth Observation

Bertrand **Le Saux**

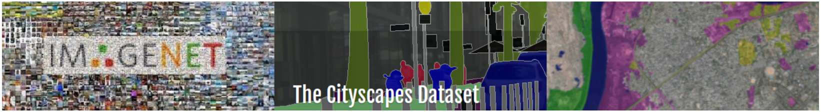
January 10, 2021

ESA/ESRIN, Φ-Lab, I-00044 Frascati (RM), Italy

# Introduction

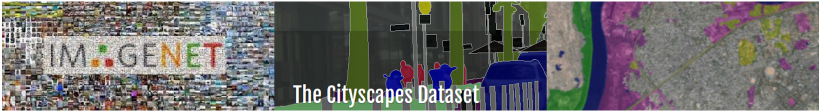
# Context

- The **availability of large, public datasets** has been key for the progress in computer vision and image processing.



# Context

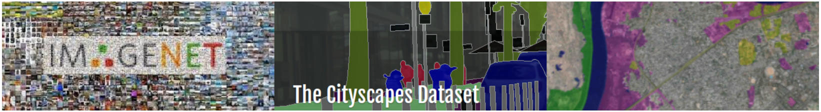
- The **availability of large, public datasets** has been key for the progress in computer vision and image processing.




- The **remote sensing** community has also developed public datasets: **land cover mapping, change detection, building detection**, etc.

# Context

- The **availability of large, public datasets** has been key for the progress in computer vision and image processing.




- The **remote sensing** community has also developed public datasets: **land cover mapping, change detection, building detection**, etc.

⚠ Main issues:

→ Limited surface covered w.r.t. the planet.

→ Classes (mostly land-cover) limited w.r.t. ImageNet.

→ Everyday, new data capture a *changing* world.

→ **Almost all designed for fully supervised methods**

# Motivation and current status

- A generic goal: semantic segmentation (good old pixel-wise classification) $\rightsquigarrow$ automatic cartography.
- Deep learning is the state of the art:
  - 92% ISPRS Vaihingen or Potsdam;
  - 80+% Houston DFC2018;
  - 75+% DeepGlobe Buildings;
  - SEN2MS/DFC2020 55-60%,
- **General knowledge:** With **enough annotated data**, one can **train and predict everywhere!**

# In real life...

**Deep Learning Models to learn *Beyond Labels***

Because the world is not fully labelled...

- The good, the bad and the ugly *label* †:
  limited data with inadequate labels.
- For a few *labels* more †:
  limited data with labels, and a few labels on new data.
- For a fistful of *labels* †:
  limited data with labels, and lots of unlabelled data.

† **Sergio Leone, "Dollar trilogy", 1964-1966.**

# In real life... new solutions are needed

**Deep Learning Models to learn *Beyond Labels***

Because the world is not fully labelled...

- The good, the bad and the ugly *label* †:
  limited data with inadequate labels.
  → Weakly-supervised learning
- For a few *labels* more †:
  limited data with labels, and a few labels on new data.
  → Continual learning
- For a fistful of *labels* †:
  limited data with labels, and lots of unlabelled data.
  → Semi-supervised learning

† **Sergio Leone, "Dollar trilogy", 1964-1966.**

# The good, the bad and the ugly *label*: Weakly-supervised Learning

# High Resolution Semantic Change Detection

- **Automatically generated** from open databases
  - Images: IGN's BD ORTHO
  - Labels: Parcel-based Copernicus Urban Atlas Change 2006-2012
- 291 10000x10000 image pairs
- **High resolution** (50 cm/pixel), **7275 km$^2$** of total imaged area
- **Multitask**: change detection and land cover mapping.

  ⇝ Understand the types of changes that the images contain.


Image 1

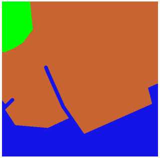
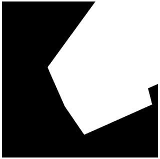

Land-Cover 1


Image 2


Land-Cover 2


Change map

---

**Daudt, Le Saux, Boulch & Gousseau, *Multitask Learning for Large-scale Semantic Change Detection* CVIU 2018.**
**Dataset available from:** `https://rcdaudt.github.io/hrscd/`

# Label Noise

## Data

HRSCD has label noise due to: *automatic vector annotations* and *temporal misalignment* between images and labels.

## Aim

Improve the accuracy of the predictions with respect to the imaged objects.

(a)

(b)

(c)

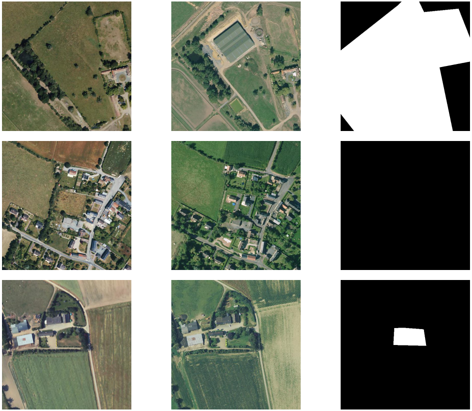


**Figure 1:** HRSCD examples of: (a) too large change markings, (b) false negatives, and (c) false positives.

# Supervision with Noisy Labels

- Direct supervision on HRSCD labels leads the network to predict blobs around detected changes to compensate for ground truth inaccuracies.
- Structure of label noise leads network to make biased predictions.
- Real and perceived class imbalance are different, which makes class weight calculations less accurate.


**Figure 2:** Result of training network with noisy labels.

# Guided Anisotropic Diffusion [1+2]

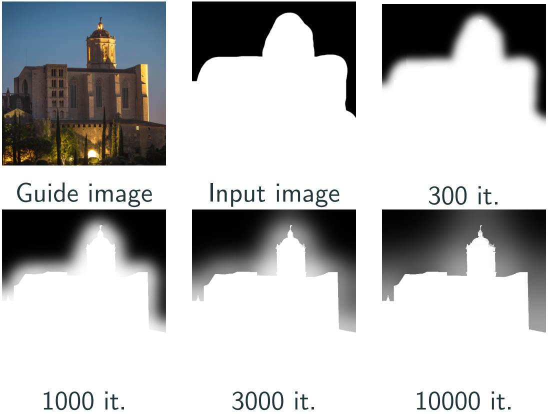



**Figure 3:** Results of Guided Anisotropic Diffusion. Edges in the guide image are preserved in the image to filter by various GAD iterations.

[1] Perona & Malik, *Scale-space and edge detection using anisotropic diffusion* TPAMI 1990.
[2] He, Sun & Tang, *Guided Image Filtering*, ECCV 2010.

# Guided Anisotropic Diffusion as Post-Processing

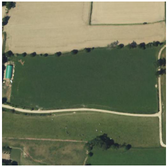


Image 1

Image 2

First prediction

2000 it.

5000 it.

20000 it.

**Figure 4:** Guided anisotropic diffusion allows edges from the guide images to be transferred to the target image, improving the results.

# Iterative Training and Label Cleaning

## Main Idea

Fix the incorrect reference labels by using network predictions! (with caution)

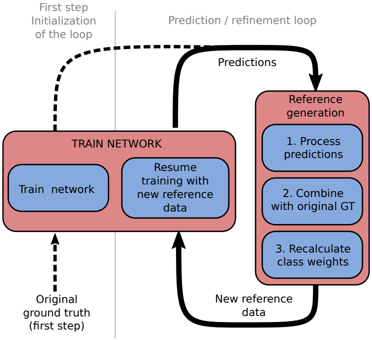



**Figure 5:** Alternate optimisation of segmentation network / label cleaning.

---

**Daudt**, **Le Saux**, **Boulch & Gousseau**, ***Guided Anisotropic Diffusion and Iterative Learning for Weakly Supervised Change Detection***, **CVPR/EarthVision 2019**.

# Results: Iterative training and label cleaning

Referring back to the reference data at each iteration is essential to avoid performance degradation.

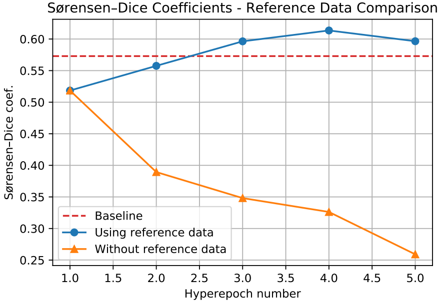



**Figure 6:** Ablation study: referring back to reference data at each iteration is essential to avoid performance degradation.

# Results: Examples

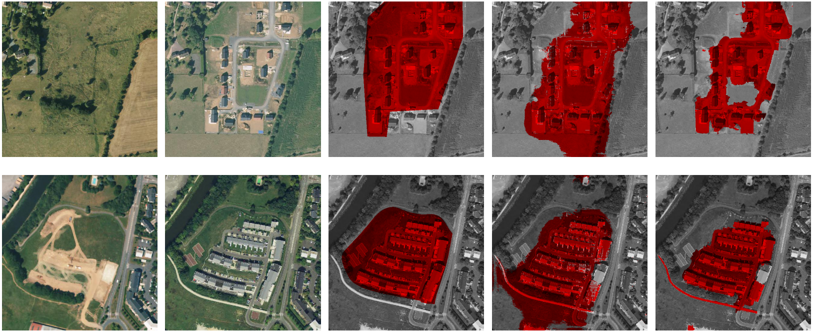


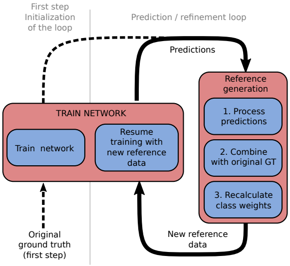
Image 1 | Image 2 | Noisy ref. data | Naive pred. | GAD

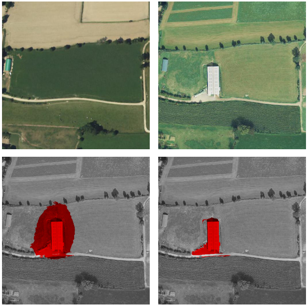
**Figure 7:** Results using the iterative network optimisation with GAD data cleaning with complete inference pipeline.

# Weak-supervision conclusion

## Noisy Labels and Weakly Supervised Learning

- Reduced the effect of label noise through iterative training
- Guided anisotropic diffusion algorithm for post-processing results



**Code**: https://github.com/rcdaudt/guided_anisotropic_diffusion.

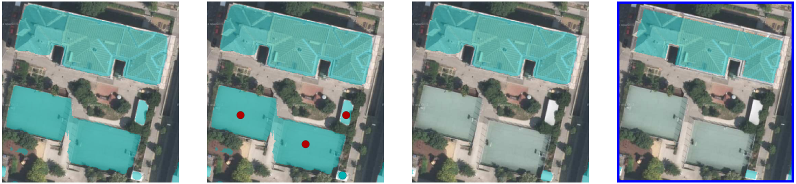
# For a few *labels* more:
# Continual Learning

## Automatic cartography by semantic segmentation

Dense classification of an image now done by Deep Neural Networks.

- EO use cases: Land cover classification, building detection,...
- DNNs are powerful *but* may fail when:
  - they face constraints such as domain shifts
  - training data is limited or labels are flawed.

➥ **Our solution:** Add a human in the loop to **interactively** refine the segmentation maps.



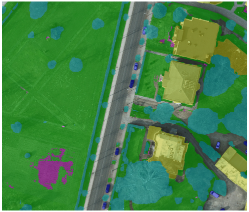
**1 - Initial segmentation** | **2 - Annotation phase** | **3 - Refined segmentation** | **Ground-truth**

Annotations lead to an easy false positive buildings removal (source: INRIA dataset)

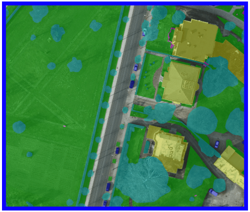
# State-of-the-art example

## Baseline: LinkNet

An efficient neural network architecture designed for semantic segmentation with a encoder/decoder architecture relying on ResNet.



**Initial prediction**



**Ground-truth**

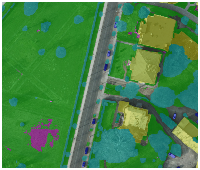
Segmentation map initially proposed by the neural network
(source: ISPRS Potsdam dataset)

Chaurasia & Culurciello, *LinkNet: Exploiting encoder representations for efficient semantic segmentation* VCIP 2017.

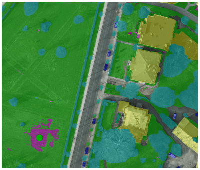
# DISIR: Interactive learning with no retraining

**DISIR: Deep Image Segmentation with Interactive Refinements**

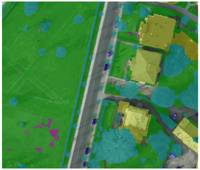
A framework for semantic segmentation with a human-in-the-loop to interactively guide a neural network to enhance its performances using user annotations as guidance



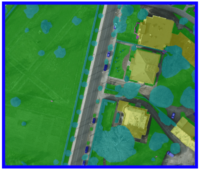
**Initial prediction**



**Initial prediction with one annotation**



**Prediction with the annotation as input**

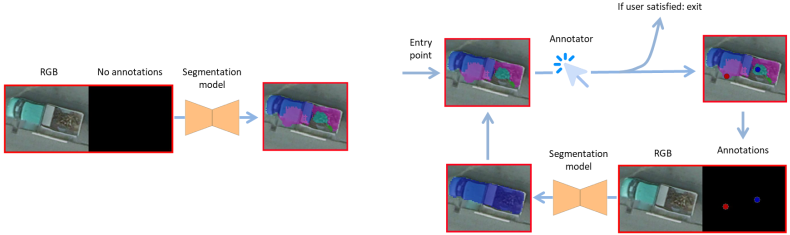


**Ground-truth**

The annotation *almost* leads to a correction of the segmentation map (source: ISPRS Potsdam dataset).

**Lenczner, Le Saux, Luminari, Chan-Hon-Tong & Le Besnerais**
***DISIR**: Deep Image Segmentation with Interactive Refinements* **ISPRS Annals 2020**.
**Code**: `https://github.com/delair-ai/DISIR`

# DISIR: Inference and Interactive Refinement



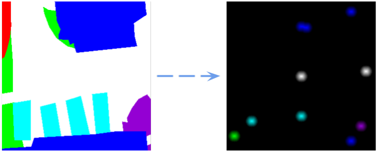

**Initialization phase**

**Interactive loop**

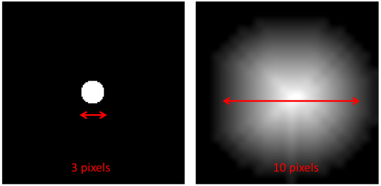
- A human in the loop interactively improves segmentation maps given by a neural network
- **Annotations:** Points representing the label of the clicked pixel
- **Key idea:** Concatenation of annotations and RGB image at input
- **No retraining**: guarantees the swiftness of the process

# DISIR: Trick[1] for Training

- At inference: annotations are clicked by the user for refinement...
- At training: annotations are simulated from the ground-truth
- Extended to multi-class labelling
- Extended representation:
  - Positioning: **Inside clicks** or border clicks
  - Encoding: Binary disks or **euclidean distance transform**



Annotations sampled from the ground-truth



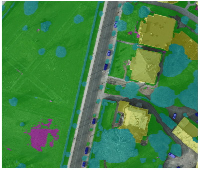

Binary (left) vs distance transform (right)

[1] Xu et al., *Deep Interactive Object Selection*, CVPR 2016.

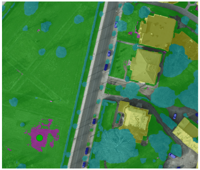
# DISCA: Continual Learning

## DISCA: Deep Image Segmentation with Continual Adaptation

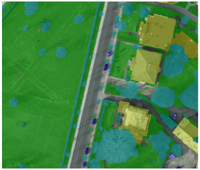
A framework for semantic segmentation with a human-in-the-loop to interactively retrain a neural network to enhance its performances using user annotations as a sparse ground truth



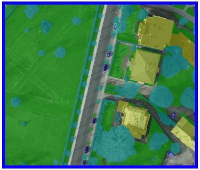
**Initial prediction**



**Initial prediction with one annotation**



**Prediction after retraining**

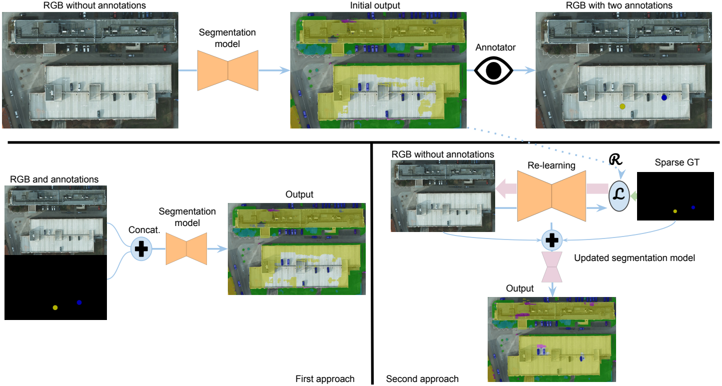


**Ground-truth**

The annotation leads to a correction of the segmentation map
(source: ISPRS Potsdam dataset)

**Lenczner, Chan-Hon-Tong, Luminari, Le Saux & Le Besnerais**
***Interactive Learning for Semantic Segmentation in Earth Observation*** **ECML-PKDD/MACLEAN 2020**.
**Code**: `https://github.com/delair-ai/DISCA`

# Dvelving into DISCA




Top: Initial segmentation phase.
Left: Interactive guidance. Right: Interactive learning.

- **Learn** from the annotations used as a sparse reference.
- Avoid forgetting by using the initial prediction as **regularization**.

| | DISIR | DISCA |
|---|---|---|
| Before | 70.6 | |
| After | 71.3 | **72.2** |

(a) ISPRS Potsdam

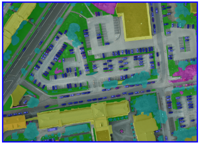
| | DISIR | DISCA |
|---|---|---|
| Before | 85.4 | |
| After | 86.4 | **86.5** |

(b) INRIA buildings

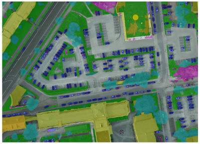
| | DISIR | DISCA |
|---|---|---|
| Before | 85.9 | |
| After | 89.5 | **90.6** |

(c) AIRS

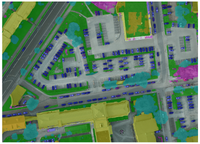
Mean IoU obtained before and after the two interactive processes of only 10 clicks (without or with modified weights).



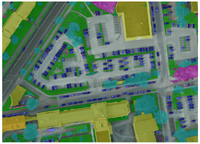
**Ground-truth**



**Initial prediction with one annotation**



**DISIR**



**DISCA**

# Results

| | DISIR | DISCA |
|---|---|---|
| Before | 70.6 | |
| After | 71.3 | **72.2** |

(a) ISPRS Potsdam

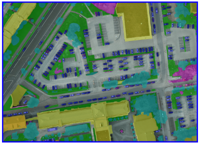
| | DISIR | DISCA |
|---|---|---|
| Before | 85.4 | |
| After | 86.4 | **86.5** |

(b) INRIA buildings

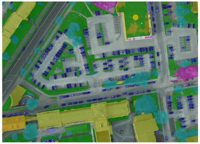
| | DISIR | DISCA |
|---|---|---|
| Before | 85.9 | |
| After | 89.5 | **90.6** |

(c) AIRS

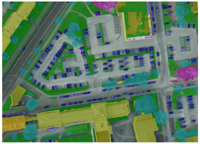
Mean IoU obtained before and after the two interactive processes of only 10 clicks (without or with modified weights).



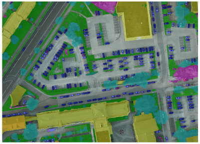
**Ground-truth**



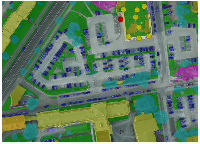
**Initial prediction with one annotation**



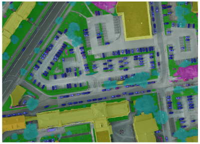
**DISIR**



**DISCA**



**More annotations**

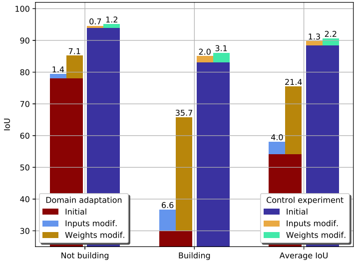


**DISIR after more annotations**

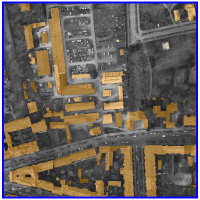
# Domain adaptation: Transfer your model on new locations!

- Train to segment buildings on AIRS; apply model on ISPRS Potsdam
- Interactive training on 10 annotations
- Compared to a network trained to segment buildings *directly* on ISPRS Potsdam

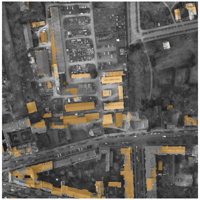
➥ The network is **able to adapt quickly!**.



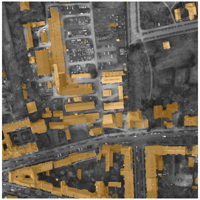

IoU evolution in a domain adaptation setup



**Ground-truth**



**Initial prediction**



**Prediction after learning on 10 annotations**

**Control prediction**

Building segmentation from the ISPRS validation dataset with a network pre-trained on AIRS.

# In a nutshell...

## Take away message

Two complementary approaches to *interactively enhance* segmentation maps proposed by a neural network with *user annotations*.

1. Modify the inputs of the network: Fast and local
   - ➥ *DISIR: Deep Image Segmentation with Interactive Refinement.* Annotations as an add. input, simulated from ground-truth at training.
2. Modify the weights of the networks: Slower and global
   - ➥ *DISCA: Deep Image Segmentation with Continual Adaptation.* Annotations' loss is back-propagated trough the model, using initial prediction as a regularisation.

## What's next

Reinforcement policies in order to better leverage information provided by the user.

# For a few labels more: Crowdsourcing

## Context

Help flood mapping from satellite imagery with in-situ information.

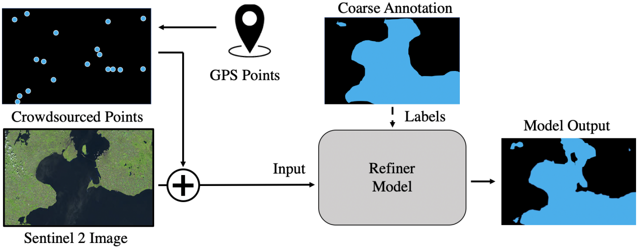



**Figure 8:** Improving automatic cartography with geo-located information.

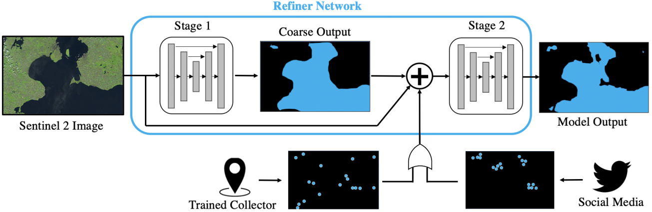
**Data: Cloud2Street's SEN1Floods11** https://github.com/cloudtostreet/Sen1Floods11

# For a few labels more: Crowdsourcing

## Continual Learning as a refiner network

Two strategies to collect street information:

- Social media scraping (low dispersion)
- Trained data collector on site (high dispersion)

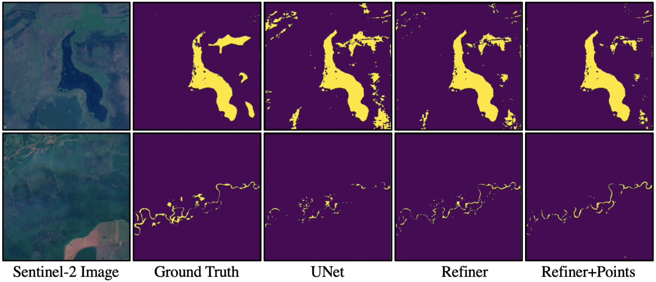



Sunkara, Purri, Le Saux & Adams, *Improving Flood Maps With Crowdsourcing and Semantic Segmentation*, NeurIPS/CCAI 2020.

# For a few labels more: Crowdsourcing

## Results

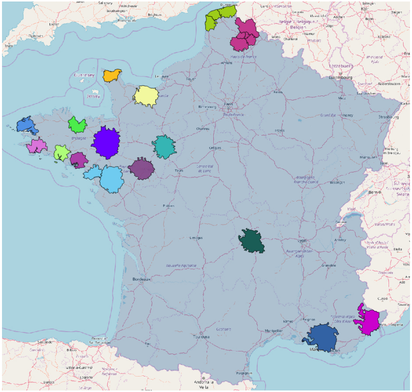
- Geo-localised information helps!
- High dispersion (dedicated info collectors) leads to better improvement




**Sunkara, Purri, Le Saux & Adams, *Improving Flood Maps With Crowdsourcing and Semantic Segmentation*, NeurIPS/CCAI 2020.**

# For a fistful of *labels*: Semi-supervised Learning

# MiniFrance: An EO Benchmark for Semi-supervised Learning



## MiniFrance in numbers

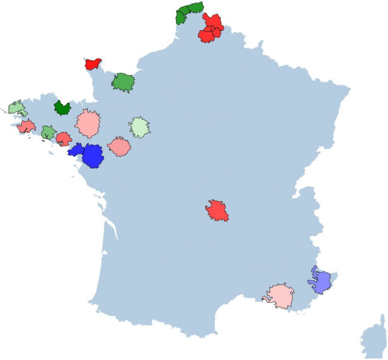
- Very large dataset for semantic segmentation.
- $> 53000$ km$^2$ of surface coverage and $\sim 150$ GB of data.
- 16 conurbations all over France.
- Aerial images from BD ORTHO (IGN) at 50cm/pixel resolution and RGB encoding.
- 15 land-use classes from Copernicus UrbanAtlas.

# MiniFrance in images

- Quantity and variability of data.
- Higher semantics, not visual classes.
- Different class appearances.
- Urban and countryside scenes.

# MiniFrance: The Semi-Supervised Partition

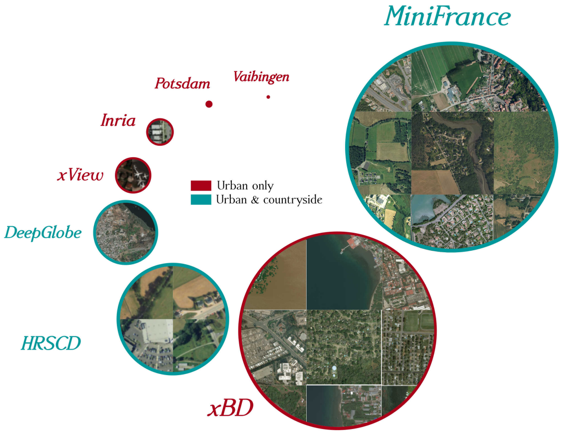
- First remote sensing dataset designed for **Semi-supervised Semantic Segmentation**.



| | | *Conurbation* | *Color* |
|---|---|---|---|
| *Training* | *Labeled* | Nice | |
| | | Nantes | |
| | *Unlabeled* | Le Mans | |
| | | Brest | |
| | | Lorient | |
| | | Caen | |
| | | Dunkerque | |
| | | Saint-Brieuc | |
| *Test* | | Marseille | |
| | | Rennes | |
| | | Angers | |
| | | Quimper | |
| | | Vannes | |
| | | Clermont-Ferrand | |
| | | Lille | |
| | | Cherbourg | |

# MiniFrance: An EO Semi-supervised Learning Benchmark

⇝ MiniFrance w.r.t EO datasets at sub-meter resolution (circle area proportional to the surface covered)



# Tools for multi-location dataset analysis

**Two conditions for good semi-supervised learning:**

→ Appearance similarity

→ Class representativeness

# Tools for multi-location dataset analysis

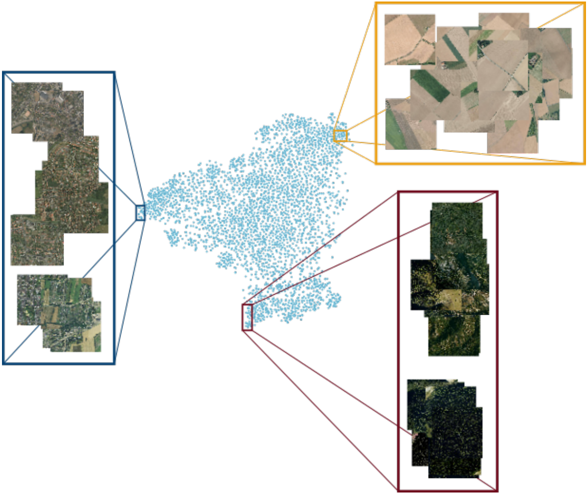
**Two conditions for good semi-supervised learning:**

→ Appearance similarity

→ Class representativeness

**How can we assess it?**

- Encode images with pre-trained CNN.
- Use t-SNE for 2D visualization.



# Tools for multi-location dataset analysis

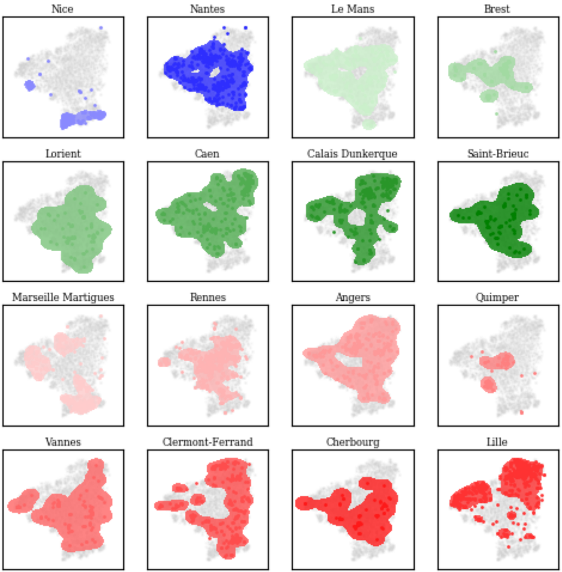
**Two conditions for good semi-supervised learning:**

→ Appearance similarity

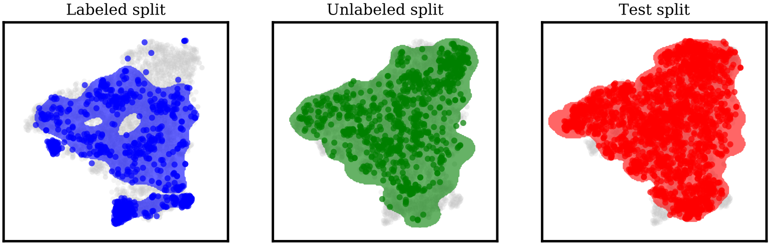
→ Class representativeness

**How can we assess it?**

- Encode images with pre-trained CNN.
- Use t-SNE for 2D visualization.
- Use one-class SVM to estimate city distributions on the 2D space.
- Evaluate appearance similarity.



# Assessing appearance similarity



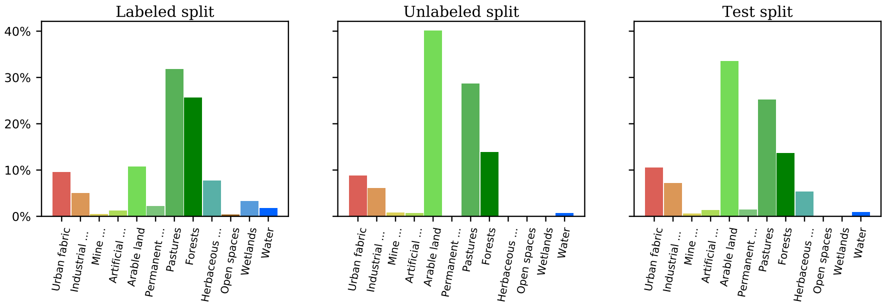

$$IoU(S_1, S_2) = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|}$$

$$IoT(S_1, S_2) = \frac{|S_1 \cap S_2|}{|S_2|}$$

**Table 1:** IoU and IoT scores between training data – labelled and unlabelled – and test data.

| $S_1$ - $S_2$ | $IoU(S_1, S_2)$ | $IoT(S_1, S_2)$ |
|---|---|---|
| *Labelled - Test* | 63 % | 64 % |
| *Unlabelled - Test* | 87 % | 93 % |

# Assessing class representativeness




Assumption: to learn a class, one should see at least one example of it.

→ All classes in the test split have training examples in the labelled split.

# MiniFrance in brief

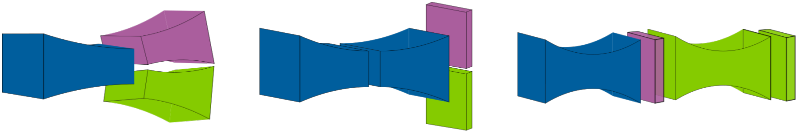
- **MiniFrance is a challenging dataset that provides lifelike use-cases:**
  → Diversity of images.
  → Land use/land cover classes with high semantic level.
  → First dataset designed for **semi-supervised learning** in EO.

- The MiniFrance suite is **publicly available for download!** at: `https://ieee-dataport.org/open-access/minifrance`

Castillo-Navarro, Audebert, Le Saux, Boulch & Lefèvre,
*Semi-Supervised Semantic Segmentation in Earth Observation: The MiniFrance Suite, Dataset Analysis, and Multi-task Network Study*, Machine Learning 2020.

# Semi-supervised learning cast as multi-task

- Study of different neural network architectures and shared parameters configuration to perform semi-supervised learning.



*BerundaNet-early* *BerundaNet-late* *W-Net*

- In this context the loss to optimize is expressed as:

$$\mathscr{L}(x) = \mathscr{L}_s(\phi_s(x), y) + \lambda \mathscr{L}_u(\phi_u(x), x)$$

$x$: input image, $y$: target, $\phi_s(x)$ and $\phi_u(x)$: supervised and unsupervised output of the network, respectively.

- $\mathscr{L}_s$ is a supervised loss for semantic segmentation (usually cross entropy) and $\mathscr{L}_u$ an unsupervised loss term.

# Semi-supervised learning results

- The choice of $\mathscr{L}_u$ depends on the task to perform along with semantic segmentation, e.g.: Reconstruction ( $\mathscr{L}_1$, etc.), Image segmentation (relaxed k-means loss, etc.).

| Backbone | Oracle $\mathcal{L}_{ce}$ | | Supervised $\mathcal{L}_{ce}$ | | Semi-supervised (BerundaNet-late) | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | $\mathcal{L}_{ce} + \lambda\mathcal{L}_1$ | | $\mathcal{L}_{ce} + \lambda\mathcal{L}_{km}$ | |
| | OA | mIoU | OA | mIoU | OA | mIoU | OA | mIoU |
| SegNet | 59.06 | 23.95 | 36.76 | 14.03 | **45.52** | 14.43 | 42.26 | **15.75** |
| U-Net | 57.71 | 25.25 | 46.30 | 18.18 | **47.90** | **18.70** | 46.92 | 18.26 |

- Supervised settings vary a lot depending on quantity of labelled data.
- **Semi-supervised strategies exhibit promising results**, whatever the architecture used as backbone.

# Semi-supervised segmentation maps

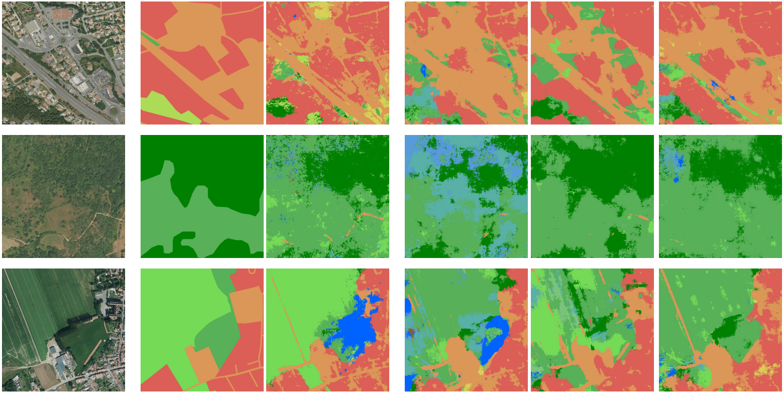


Image | GT | Oracle | Sup | Semi ($\mathscr{L}_1$) | Semi ($\mathscr{L}_{km}$)

Undisclosed | Results

# Dvelving into auxiliary tasks

1. **Reconstruction**

   → Generate an output as close as possible to the original input, using standard **p-norms** e.g. $\mathscr{L}_1$, $\mathscr{L}_2$ losses.

2. **Unsupervised Segmentation**

   → Partition an image into multiple segments, where pixels in a segment share some properties, like color, intensity, or texture, e.g. **Mumford-Shah** functional $\mathscr{L}_{MS}$, **Relaxed** K-**means** $\mathscr{L}_{km}$.

3. **Self-supervision**

   → Build a supervised task from completely unlabelled data by producing labels from the data itself e.g. **Inpainting:**, **Jigsaw puzzle** $\mathscr{L}_{js}$.

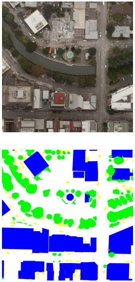
---

Castillo-Navarro, Le Saux, Boulch & Lefèvre, *On Auxiliary Losses for Semi-Supervised Semantic Segmentation*, ECML-PKDD/MACLEAN 2020.

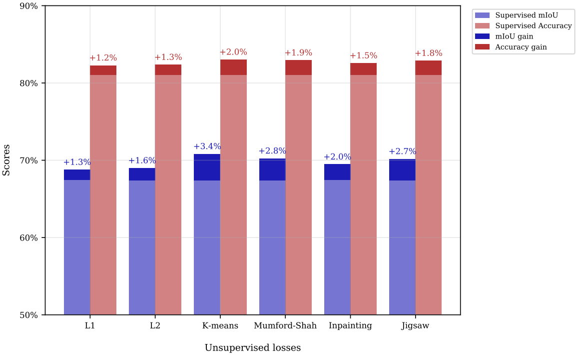
# Results

## Christchurch (NZ) Aerial Semantic Dataset [1]

VHR images at 10cm/pix.; 4 classes, 2 labelled / 20 unlabelled / 2 valid. tiles.



CASD ex.



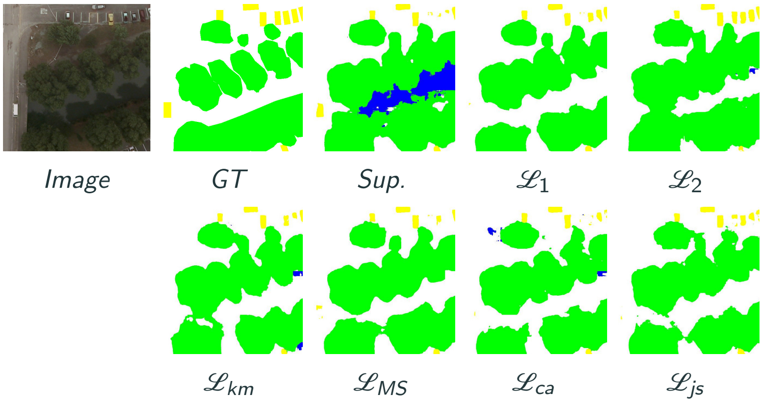

→ **Semi-supervised approaches outperform the supervised setting!**

→ Best scores are obtained with segmentation losses ($\mathscr{L}_{km}$ and $\mathscr{L}_{MS}$.)

---

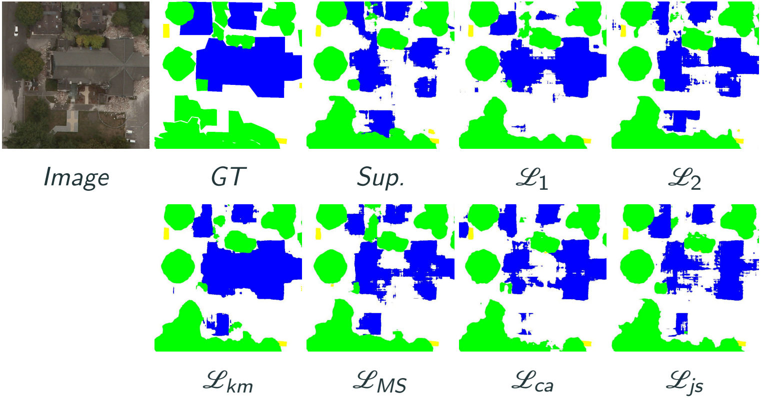
1 **Randrianarivo, Le Saux, & Ferecatu,** ***Man-made structure detection with deformable part-based models*** **IGARSS 2013.**
**CASD available from: Zenodo /** `https://blesaux.github.io/data/`

# And Visually...




→ The supervised approach is the only one that mistakes the shadow of trees over the river as a building.

# And Visually...




→ The $\mathscr{L}_{km}$ loss is the only one that correctly segments the central building.

# Semi-supervised learning conclusions

**Unlabelled data and semi-supervised learning (SSL)**

- A **new benchmark** for SSL: MiniFrance challenges the potential of deep networks and provides lifelike use-cases.
- Various **semi-supervised networks based on multi-task learning** (BerundaNet), to handle labelled and unlabelled data at training.
- Semi-supervision **improves** classification results on MiniFrance and CASD datasets.
- **Segmentation losses** for the auxiliary task seem to be the more appropriate, quite intuitively w.r.t. to the primary task.

# Concluding remarks

# What's next

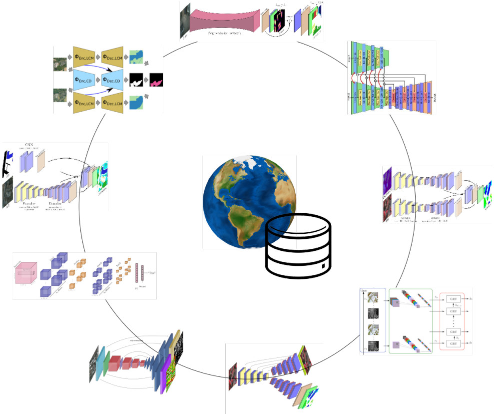
## Space Today

- 1200+ satellites are now evolving around Earth
- Constellations will be tomorrow's standard, with unprecedented high acquisition frequency and data volume

## What it implies

➥ A major change is coming in the way we process EO data

# What's next

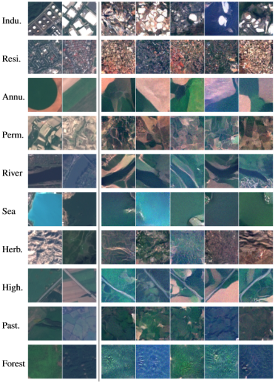


## Dealing with unlabelled data

- Reinforcement, continual, active learning
- Unsupervised, self-supervised, semi-supervised learning
- Few or zero-shot learning, transfer learning

In this PRRS workshop:
Kölle et al, *Remembering Both the Machine and the Crowd when Sampling Points: Active Learning for Semantic Segmentation of ALS Point Clouds*.
Leenstra et al., *Self-supervised pre-training enhances change detection in Sentinel-2 imagery*

# What's next




## Dealing with unlabelled data

- Continual learning over time, upgrading the models place after place $\rightsquigarrow$ go beyond the "fixed dataset" paradigm and move towards life-long learning;
- Unsupervised statistics, with generative models to estimate the underlying distribution of EO data $\rightsquigarrow$ allow both more efficient downstream tasks and simulation.

# The End

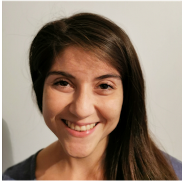
## Thank you for your attention !

Primary contributors:

Rodrigo Caye Daudt

Gaston Lenczner

Veda Sunkara



Javiera Castillo-Navarro

And: Alexandre Boulch, Yann Gousseau, Nicola Luminari, Adrien Chan-Hon-Tong, Guy Le Besnerais, Matthew Purri, Jennifer Adams, Nicolas Audebert, Sébastien Lefèvre.

**Mail:** bertrand.le.saux@esa.int **Web:** `https://blesaux.github.io`