Semi-supervised Semantic Segmentation in Earth Observation

Javiera Castillo-Navarro[1,2]
Bertrand Le Saux[3], Alexandre Boulch[4],
Nicolas Audebert[5] & Sébastien Lefèvre[2]

[1]ONERA, Université Paris Saclay
[2]Université Bretagne-Sud/ IRISA
[3]European Space Agency
[4]valeo.ai
[5]CNAM

ONERA cnes UMR IRISA

Motivation

Figure: Automated cartography.

- Our goal: semantic segmentation $\rightsquigarrow$ pixel-wise classification.

Motivation

Figure: Automated cartography.

- Our goal: semantic segmentation $\rightsquigarrow$ pixel-wise classification.
- Deep learning is the state of the art.
- **General knowledge:** We need a **big amount of annotated data!**

Motivation: Semi-supervised learning

- Labeled data is scarce and hard to obtain.
- Raw data is abundant, especially in Earth Observation.

$\rightsquigarrow$ **Semi-supervised learning to leverage unlabeled data!**
And improve the generalization capacities of the models.

Motivation: Semi-supervised learning

- Labeled data is scarce and hard to obtain.
- Raw data is abundant, especially in Earth Observation.

$\rightsquigarrow$ **Semi-supervised learning to leverage unlabeled data!**
And improve the generalization capacities of the models.

Problem: Limitations of existing EO datasets

- Modest surface coverage
- Lack of variety of scenes
- Only one class or land cover classes

Motivation: Semi-supervised learning

- Labeled data is scarce and hard to obtain.
- Raw data is abundant, especially in Earth Observation.

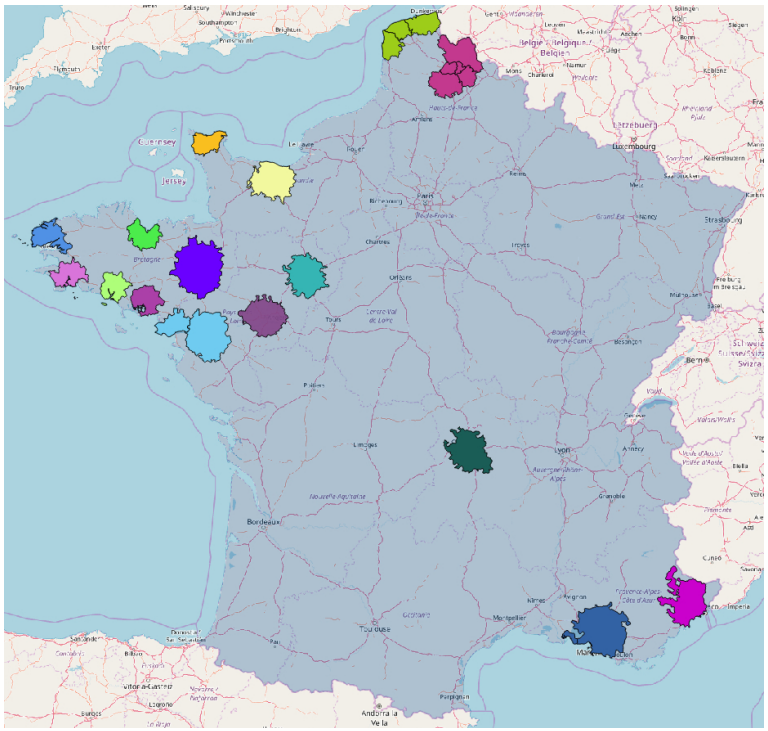
$\rightsquigarrow$ **Semi-supervised learning to leverage unlabeled data!**
And improve the generalization capacities of the models.

Problem: Limitations of existing EO datasets

- Modest surface coverage
- Lack of variety of scenes
- Only one class or land cover classes
- **All designed for fully supervised methods**

The MiniFrance Suite: A dataset for Semi-Supervised Semantic Segmentation.

The MiniFrance dataset



- Very large dataset for semantic segmentation.
- > 53000 km^2 of surface coverage and ~ 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 UrbanAtlas.

ONERA THE FRENCH AEROSPACE LAB

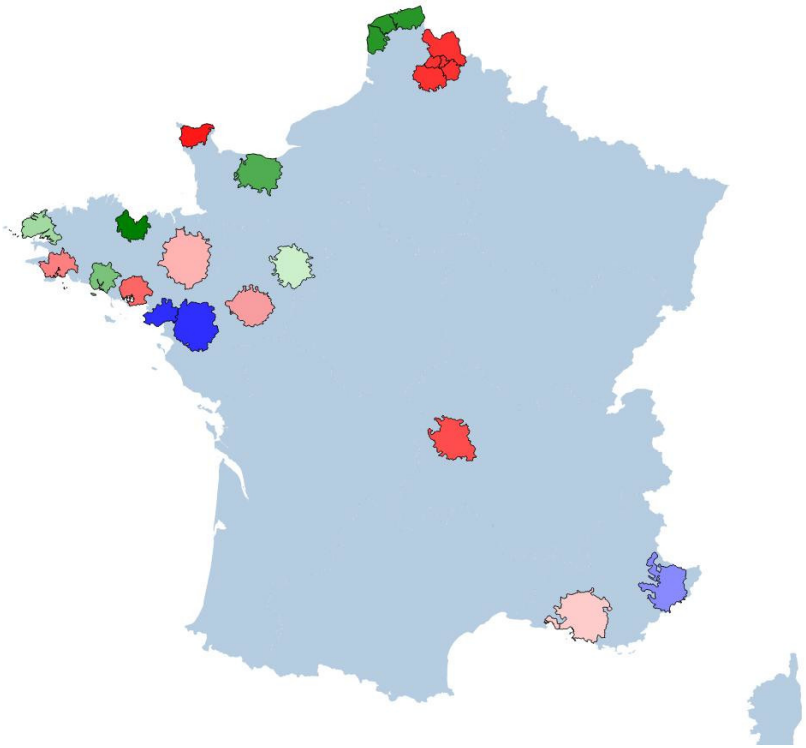
MiniFrance properties

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

ONERA THE FRENCH AEROSPACE LAB

MiniFrance: The Semi-Supervised Partition

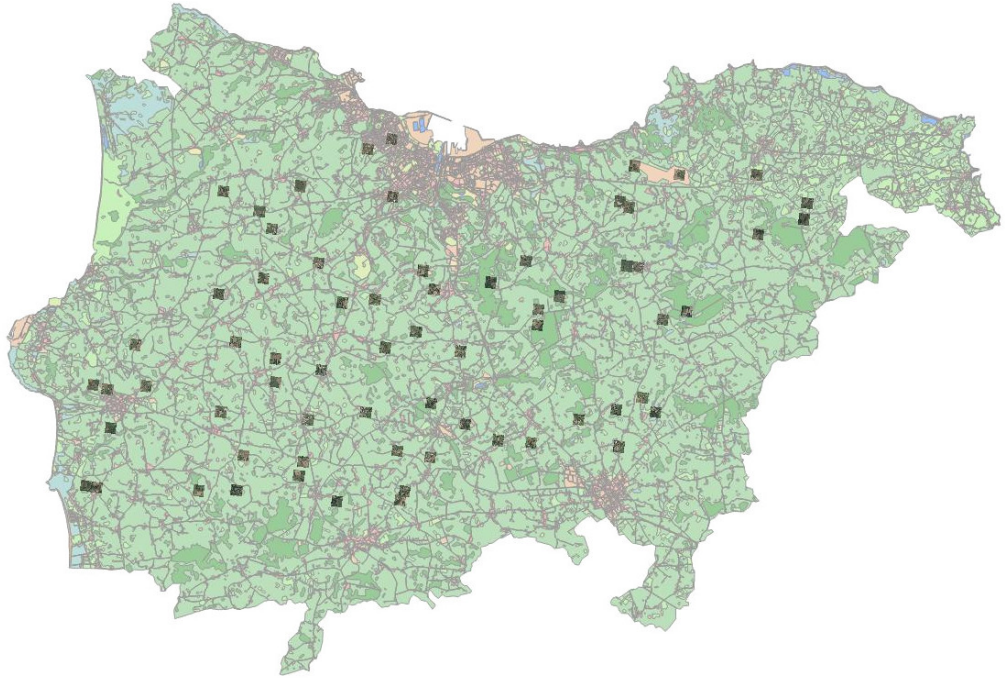
- First remote sensing dataset designed for **Semi-supervised Semantic Segmentation**.



		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	

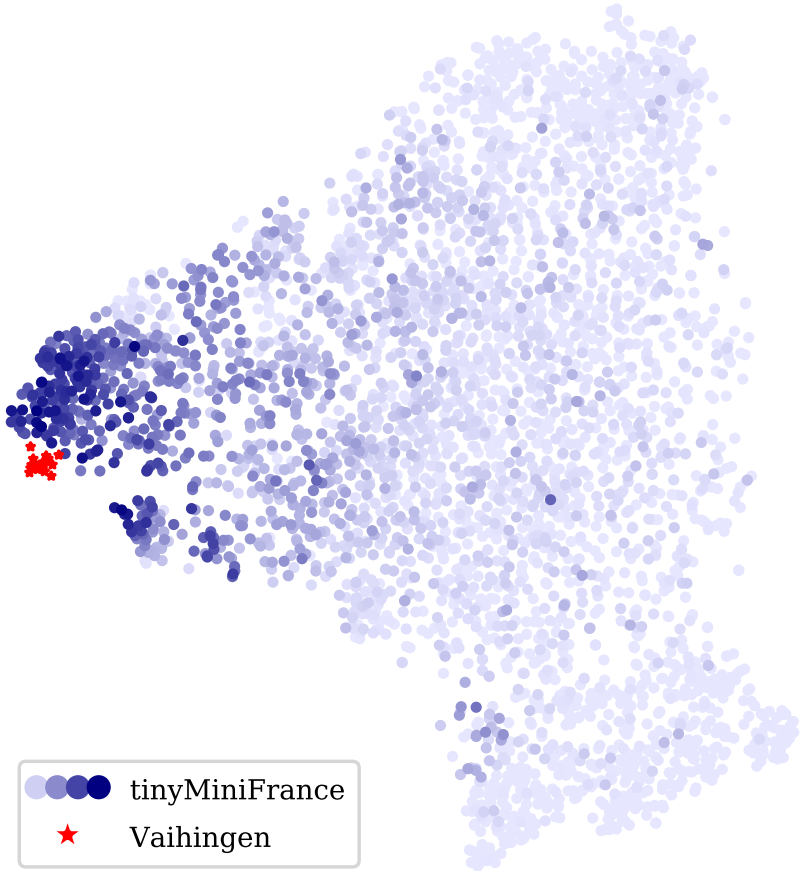
Tiny-MiniFrance

- **Tiny-MiniFrance version of the dataset:**
  → $\sim$ 1.5% randomly sampled from the original data.
  → Fast development and training, without losing generality.



ONERA THE FRENCH AEROSPACE LAB

Diversity compared to classic EO datasets



tinyMiniFrance colored according to the proportion occupied by the class "urban fabric"

⇝ Vaihingen is close to urban images from tinyMiniFrance.

⇝ **tinyMiniFrance presents a much larger variety of scenes.**

Figure: t-SNE visualization of tinyMiniFrance and ISPRS Vaihingen.

Dataset Analysis: Is MiniFrance Suitable for Semi-Supervision?

Tools for multi-location dataset analysis

Two conditions for good semi-supervised learning:

→ Class representativeness

→ Appearance similarity

ONERA THE FRENCH AEROSPACE LAB

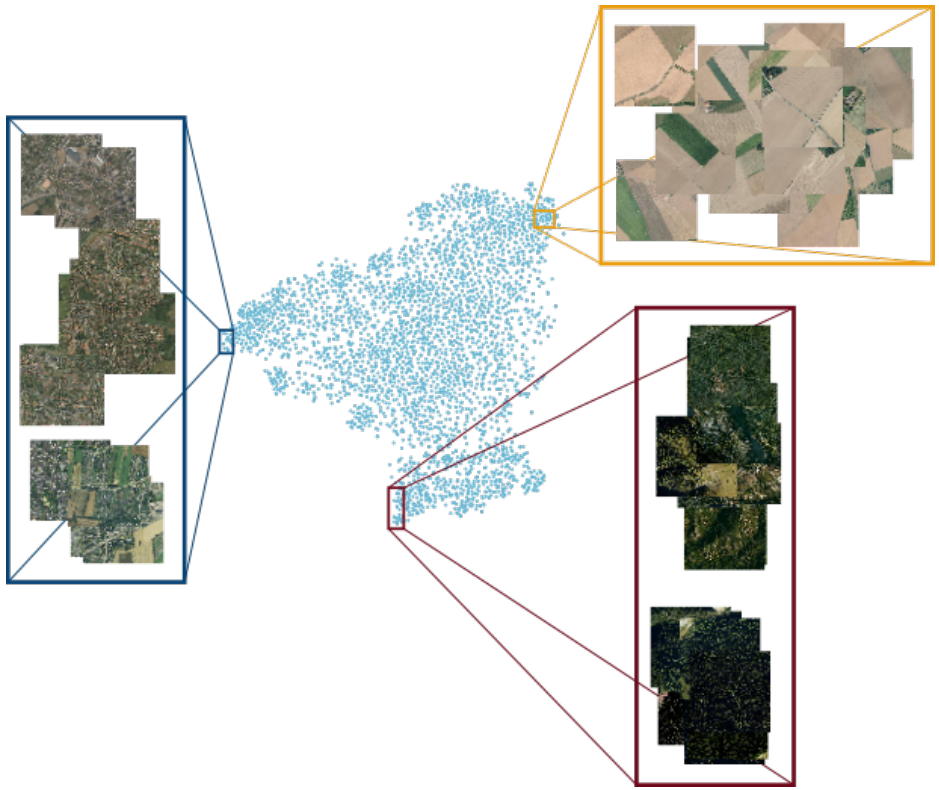
Tools for multi-location dataset analysis

Two conditions for good semi-supervised learning:

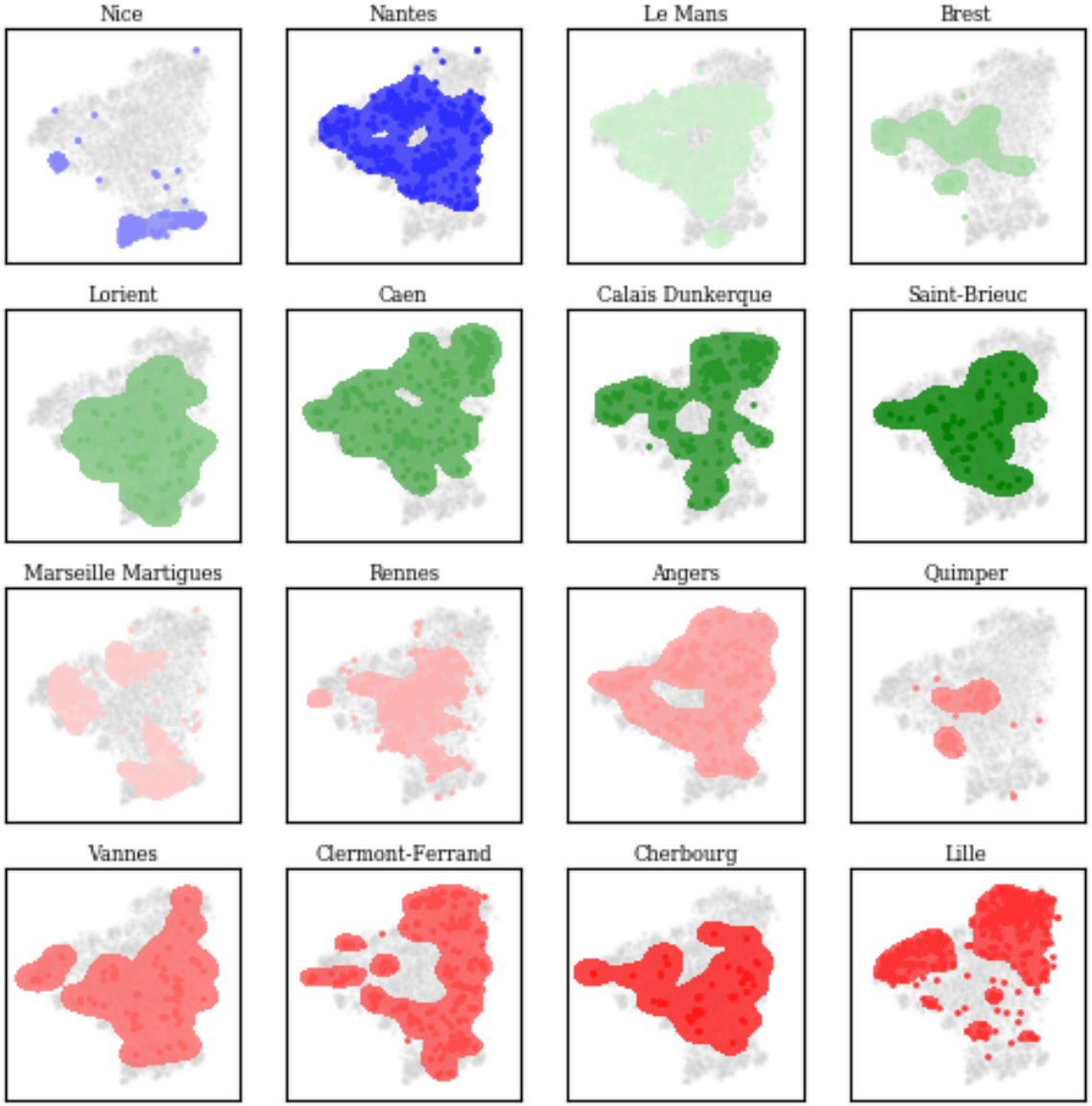
→ Class representativeness
→ Appearance similarity

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.

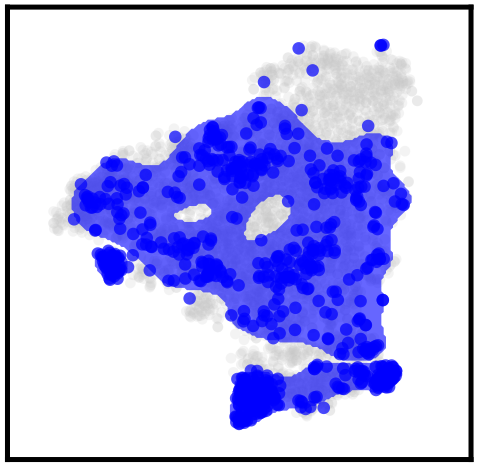


Tools for multi-location dataset analysis

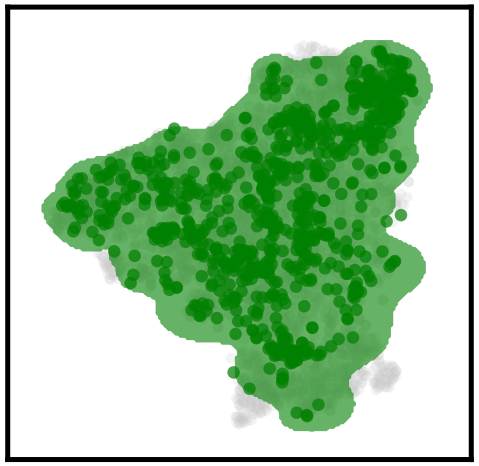


Assessing appearance similarity

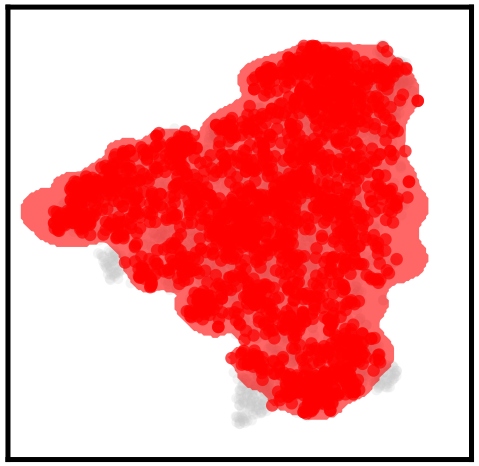
Labeled split



Unlabeled split



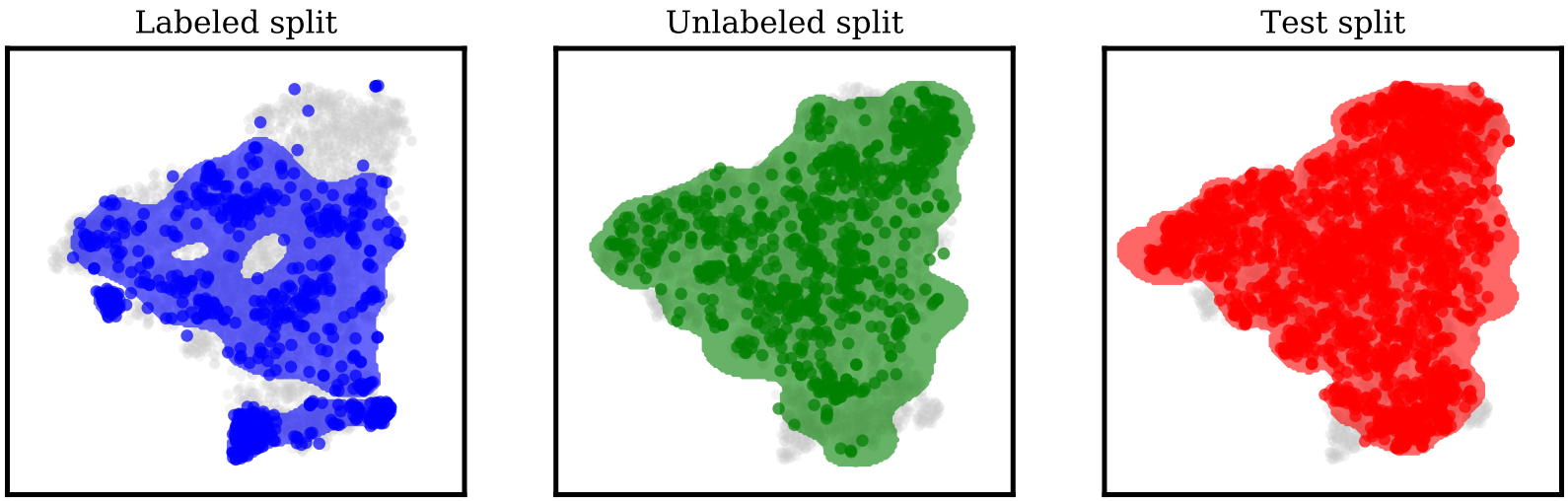
Test split



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

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

Assessing appearance similarity



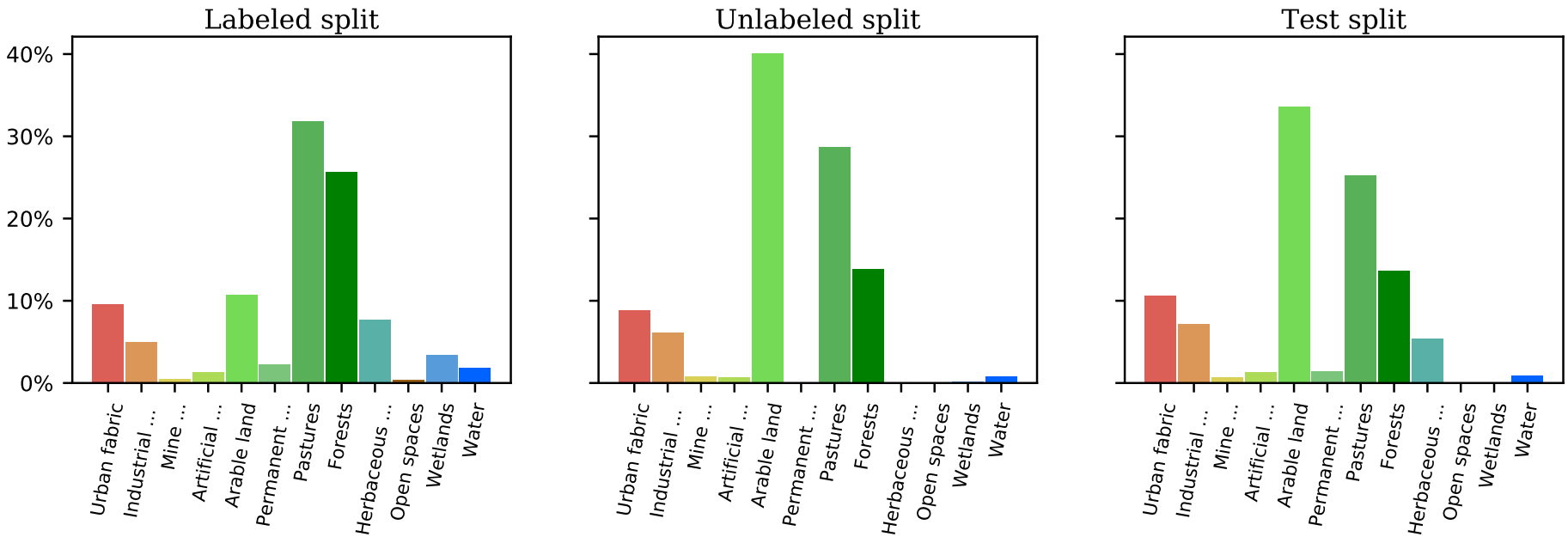
$$IoU(S_1, S_2) = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|}$$

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

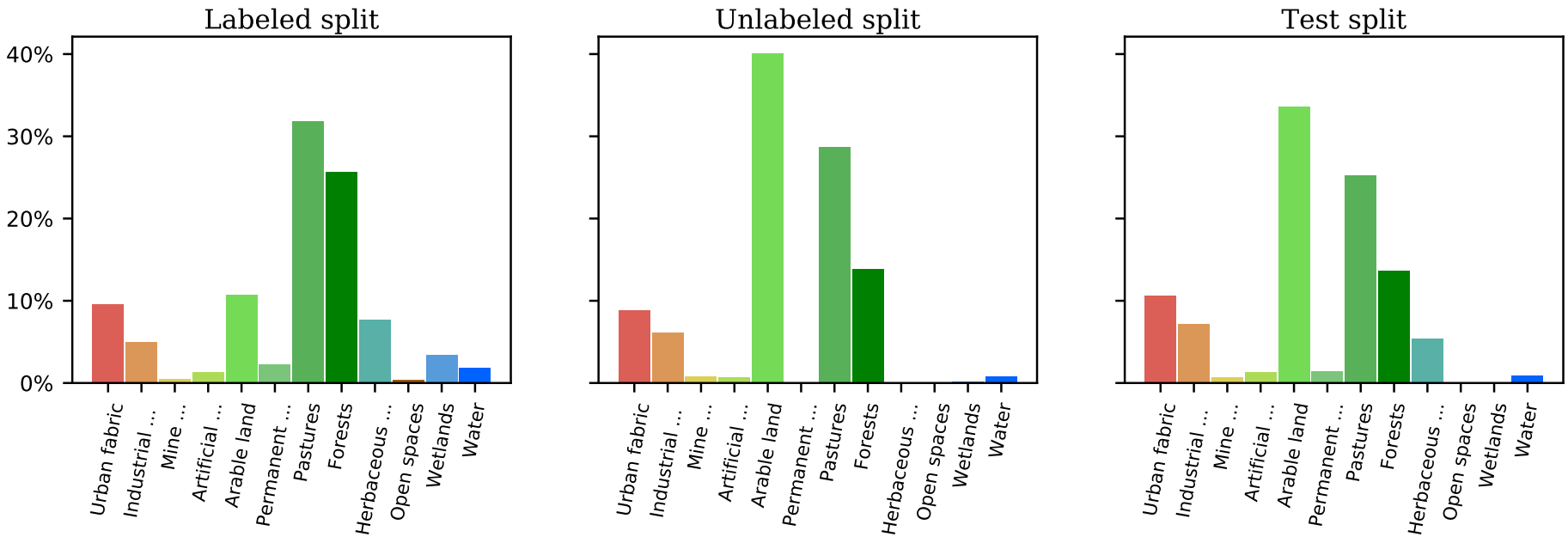
Table: IoU and IoT scores between training data – labeled and unlabeled – and test data.

S_1 - S_2	$IoU(S_1, S_2)$	$IoT(S_1, S_2)$
Labeled - Test	0.63 %	0.64 %
Unlabeled - Test	0.87 %	0.93 %

Assessing class representativeness



Assessing class representativeness

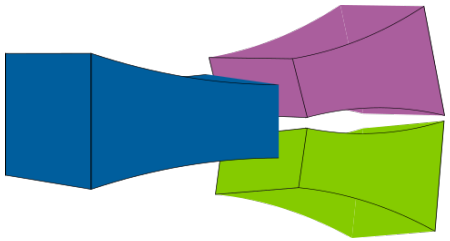


→ All classes in the test split have training examples in the labeled split.

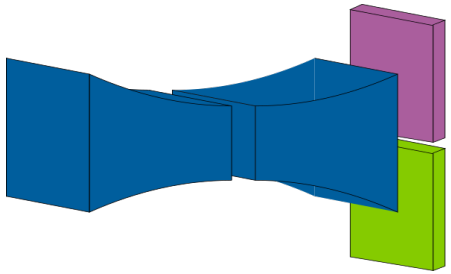
Semi-Supervised Semantic Segmentation with Deep Neural Networks

Semi-supervised approaches

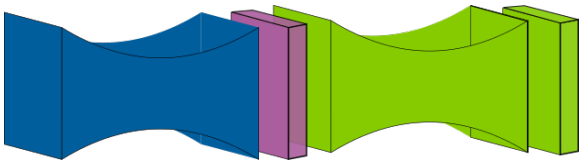
- Study of different neural network architectures and shared parameters configuration to perform semi-supervised learning.



BerundaNet-early



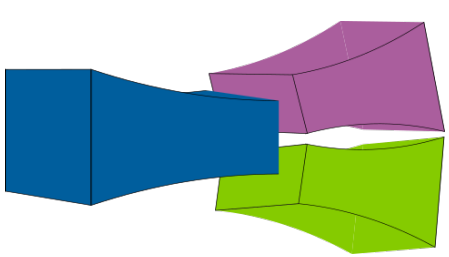
BerundaNet-late



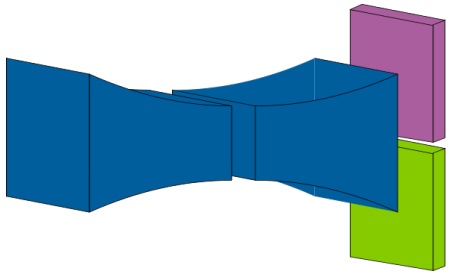
W-Net

Semi-supervised approaches

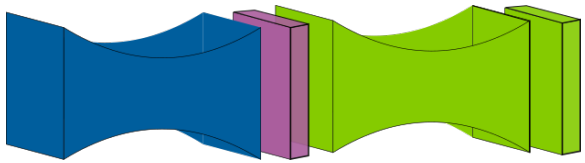
- Study of different neural network architectures and shared parameters configuration to perform semi-supervised learning.



BerundaNet-early



BerundaNet-late



W-Net

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

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

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

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

Unsupervised tasks for semantic segmentation

- The choice of $\mathcal{L}_u$ depends on the task to perform along with semantic segmentation.
- Study of different unsupervised tasks and losses:
  $\rightsquigarrow$ Reconstruction: $\mathcal{L}_1$, $\mathcal{L}_2$ losses.
  $\rightsquigarrow$ Image segmentation: Mumford-Shah loss, invariant information clustering, dense CRF, **relaxed k-means loss (ours)**.

Results: Can unlabeled data help?

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

Results: Can unlabeled data help?

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

- MiniFrance is challenging.

Results: Can unlabeled data help?

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

- MiniFrance is challenging.
- The amount of labeled data influences a lot the performance of supervised settings.

Results: Can unlabeled data help?

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

- MiniFrance is challenging.
- The amount of labeled data influences a lot the performance of supervised settings.
- **Semi-supervised strategies exhibit promising results**, whatever the architecture used as backbone.

And visually

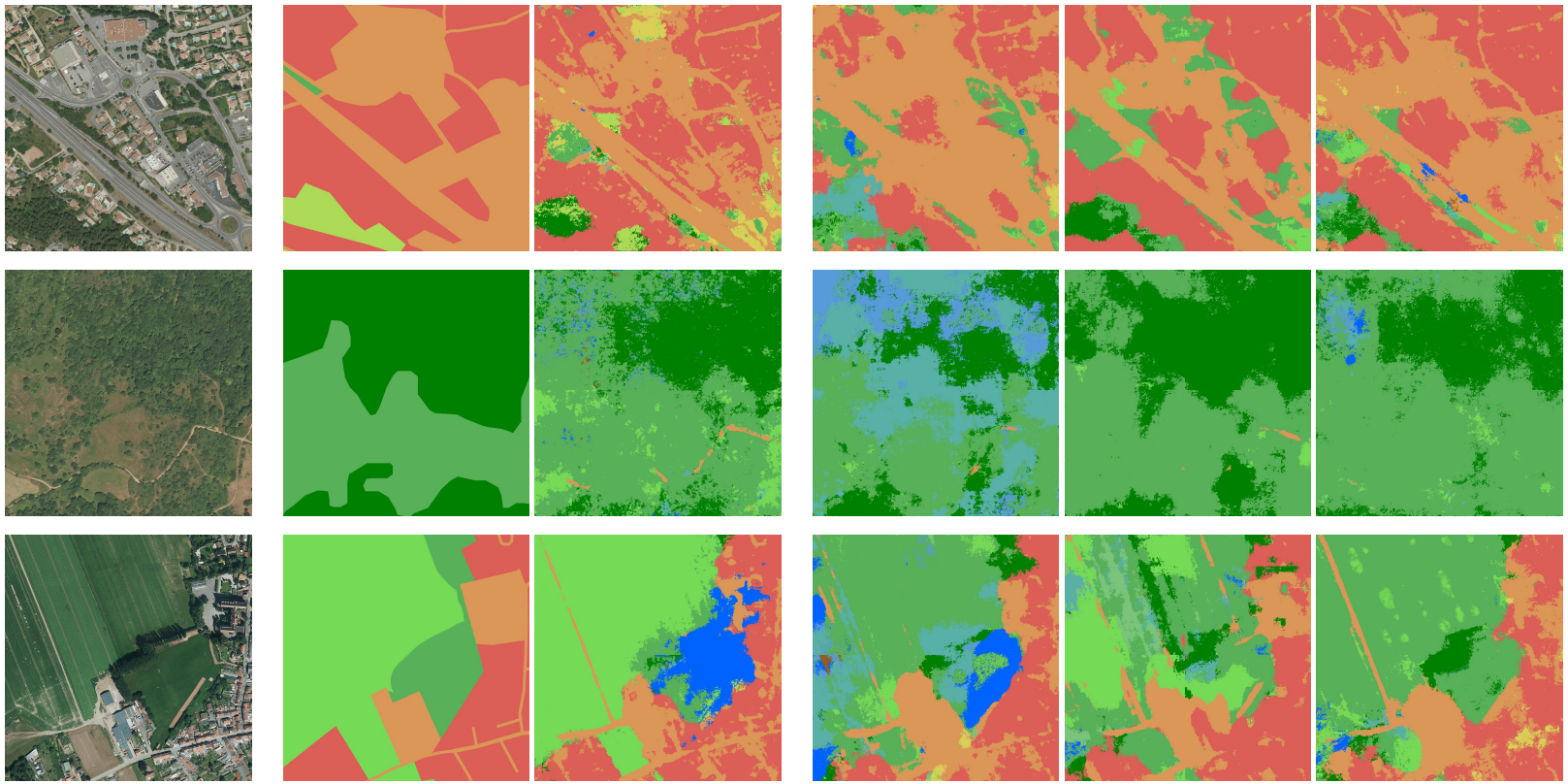


Image GT Oracle Sup Semi ($\mathcal{L}_1$) Semi ($\mathcal{L}_{km}$)

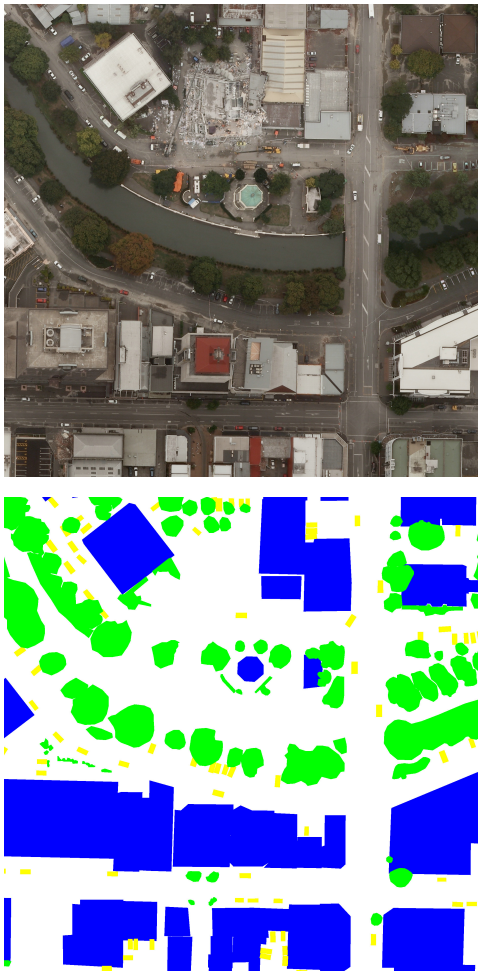
Undisclosed Results

ONERA THE FRENCH AEROSPACE LAB

Semi-supervision over the CASD dataset

- Aerial imagery over Christchurch, New Zealand.
- 4 classes to identify.
- 2 labeled tiles for training + 20 unlabeled tiles for training.

- We train BerundaNet-late with different tasks and losses:
  → Reconstruction: $\mathcal{L}_1$, $\mathcal{L}_2$.
  → Segmentation: $\mathcal{L}_{km}$, $\mathcal{L}_{MS}$.
  → Self-supervision: Jigsaw puzzle ($\mathcal{L}_{js}$), inpainting ($\mathcal{L}_{ca}$).

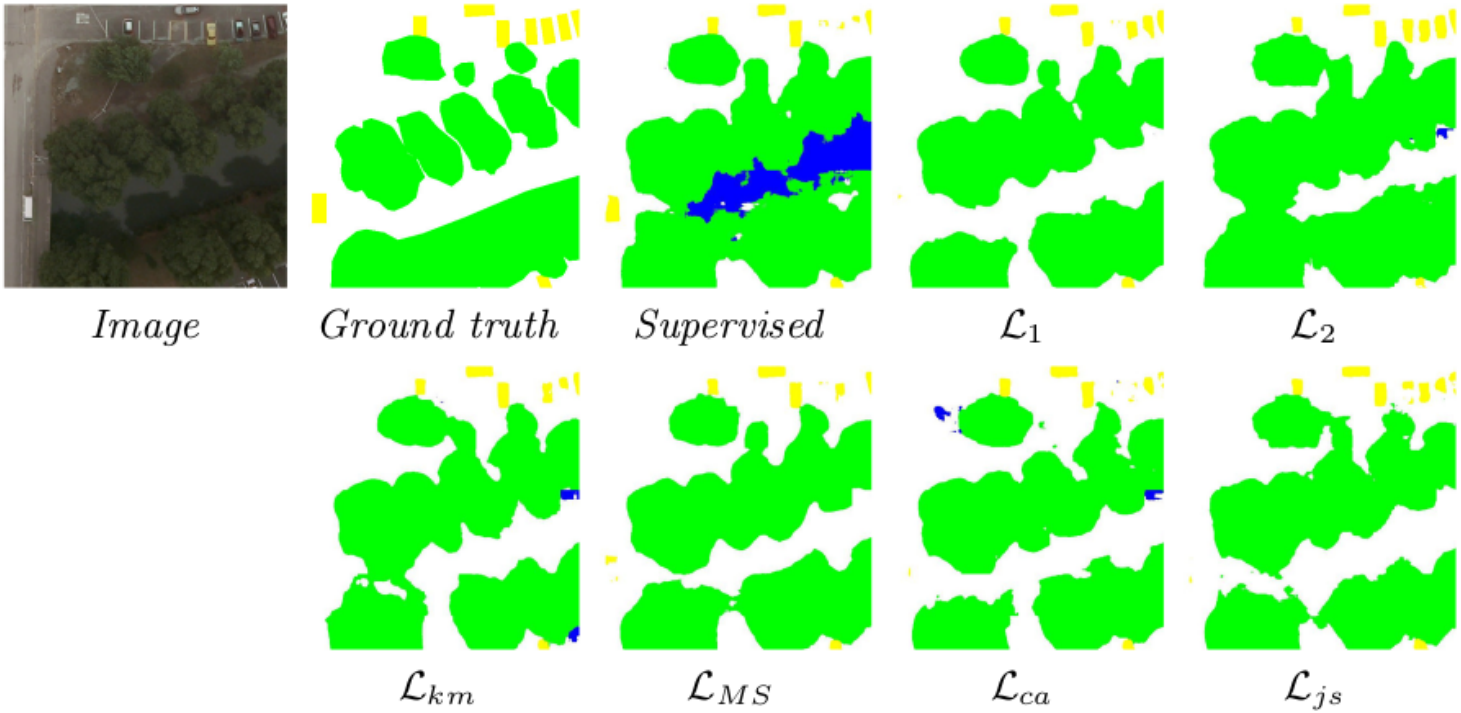


Results over Christchurch

Mode	*Aux. Task*	*Aux. Loss*	λ	*OA (%)*	*mIoU (%)*
Sup	-	-	-	81.06 ± 0.46	67.43 ± 0.49
Semi-sup	Rec	$\mathcal{L}_1$	0.5	82.28 ± 0.55	68.78 ± 1.27
		$\mathcal{L}_2$	5	82.36 ± 0.42	68.99 ± 0.85
	Seg	$\mathcal{L}_{km}$	1	$\mathbf{83.03 \pm 0.42}$	$\mathbf{70.82 \pm 0.35}$
		$\mathcal{L}_{MS}$	1	82.94 ± 0.26	70.24 ± 0.84
	Self	$\mathcal{L}_{ca}$	5	82.57 ± 0.59	69.47 ± 0.7
		$\mathcal{L}_{js}$	0.5	82.88 ± 0.95	70.17 ± 1.12

$\rightsquigarrow$ Semi-supervised methods outperform supervised approach in all cases!

And visually...



Conclusions

- MiniFrance challenges the potential of deep networks and provides lifelike use-cases:
  - → Images are more diverse.
  - → Classes have a higher semantic level.
  - → First dataset designed for semi-supervised learning.

Conclusions

- MiniFrance challenges the potential of deep networks and provides lifelike use-cases:
  - → Images are more diverse.
  - → Classes have a higher semantic level.
  - → First dataset designed for semi-supervised learning.
- We have proposed semi-supervised networks, based on multi-task learning, capable to handle labeled and unlabeled data during training.
- We have validated our methods over a MiniFrance and the CASD datasets, with promising results.

Questions?

Thank you!

Acknowledgements

- Javiera Castillo-Navarro's work is partially funded by a cnes research grant.