EPFL

Semi-supervised medical image segmentation

Speaker

Emery Sébastien

Supervisor

Behzad Bozorgtabar

January 2020

École polytechnique fédérale de Lausanne

Overview

Generalities

Cross-consistency

Multi-site Prostate Segmentation of MRI data

Generalities

Semantic segmentation

- Pixelwise annotation
- Large amount of labeled data

Costly and time consuming to build a dataset

COCO
Common Objects in Context

CITYSCAPES DATASET

PASCAL2
Pattern Analysis, Statistical Modelling and Computational Learning

Semi-supervised learning

- ❑ Goal : Use the large proportion of unlabeled data combined with the small proportion of labeled example

Labeled example

Current SSL methods

- ❑ Consistency training
- ❑ Pseudo labeling
- ❑ Entropy minimization
- ❑ Bootstrapping

Unlabeled example

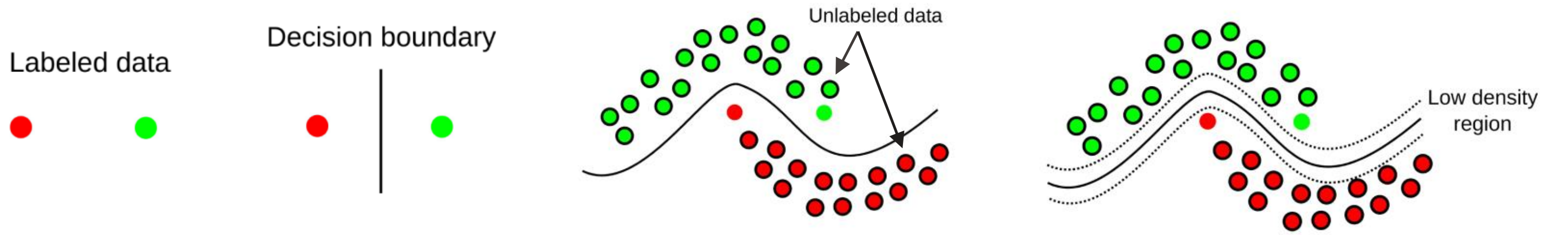
Overview of SSL methods

- Pseudo-labelling[1]
  - Train a network in a supervised manner
  - Predict the unlabeled data points
  - Use the prediction as ground truth and retrained the network with the newly annotated data
- Entropy minimization[2]
  - Add an Entropy penalty to the loss
  - The entropy is computed on the output probability density over classes
  - Prioritize the more confident outputs
- Bootstrapping[3]
  - Train multiples identical network on sub-sampled data and aggregate
  - Train different network with the same data and aggregate

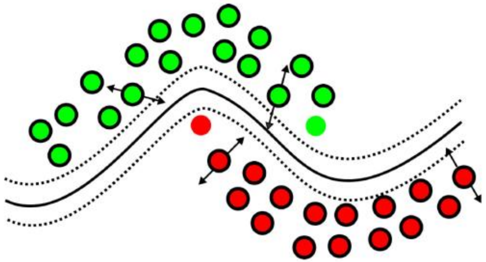
1 : Dong-Hyun Lee [4] 2: Yves Grandvalet and Yoshua Bengio [5] 3: Siyuan Qiao,Wei Shen, Zhishuai Zhang, BoWang, and Alan Yuille [6]

Consistency

Cluster assumption



Consistency training



$$f(x_u) = f(x_u + \epsilon)$$

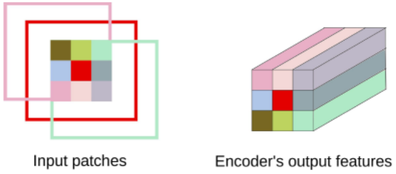
A small perturbation applied to the unlabeled example should not change the prediction

Cross-consistency

Cluster assumption

Reference Paper : “Semi-supervised semantic segmentation with cross-consistency training”
Authors : Yassine Ouali, Céline Hudelot, Myriam Tami

Average Euclidian distance over patches



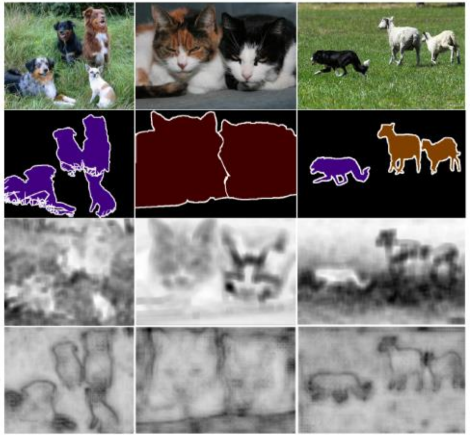
Cluster assumption

- ❑ Violated at the input level
- ❑ Valid at the encoder’s output

Perturbation applied at the encoder’s output

Input smoothness

Feature smoothness

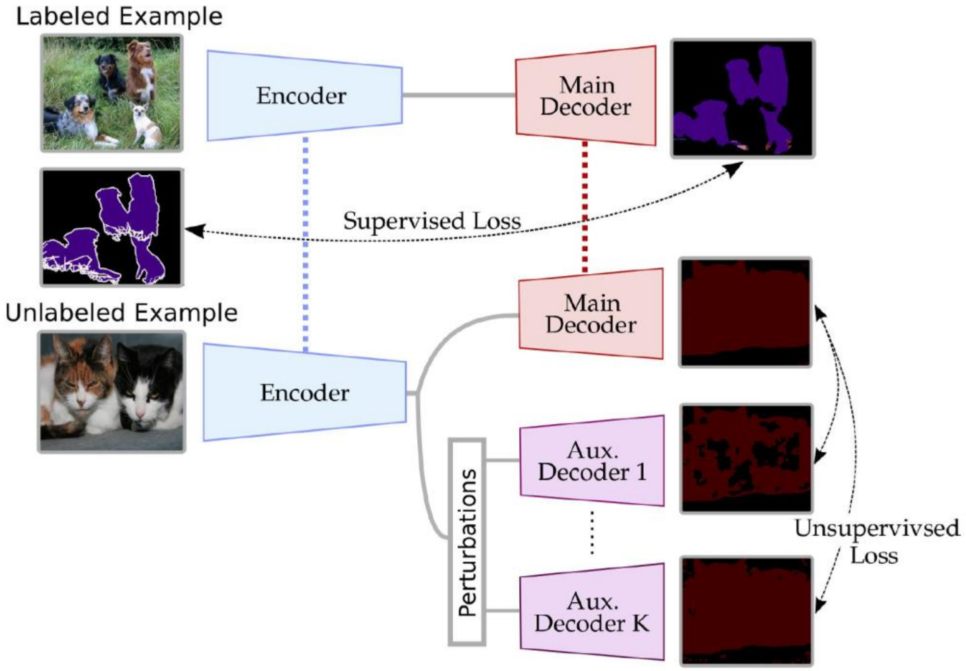


Cross-Consistency training

- Train the Encoder/Decoder in a supervised manner using labeled data
- Enforce consistency of predictions of the unlabeled data over the features

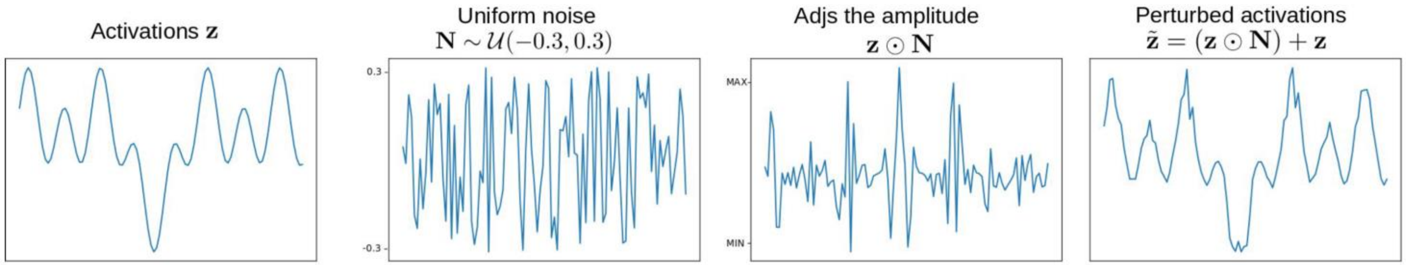
$$\mathcal{L} = \mathcal{L}_S + \omega_U \mathcal{L}_U$$

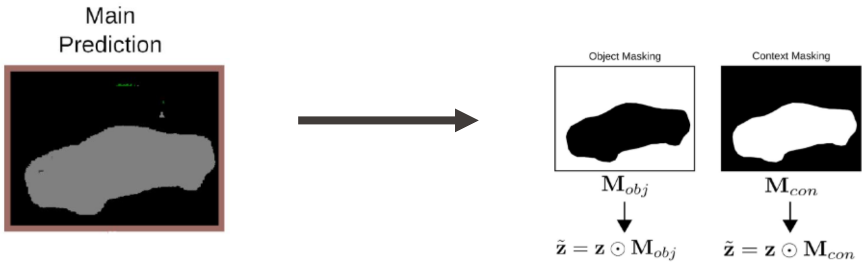
- Ramps up starting from zero the unsupervised loss to avoid initial noise



Perturbations

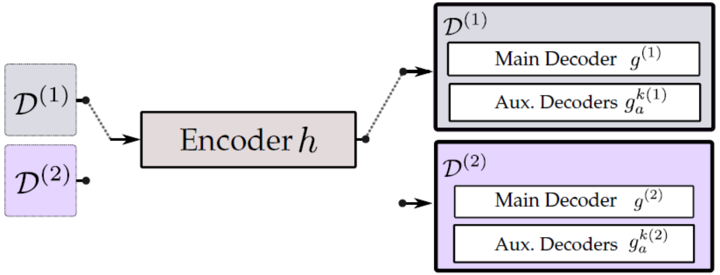
Feature based	Prediction based	Random perturbation
F-noise	Guided-Masking	Spatial dropout
F-Drop	Guided-cutout	-
-	Intermediate VAT	-





Cross-modality

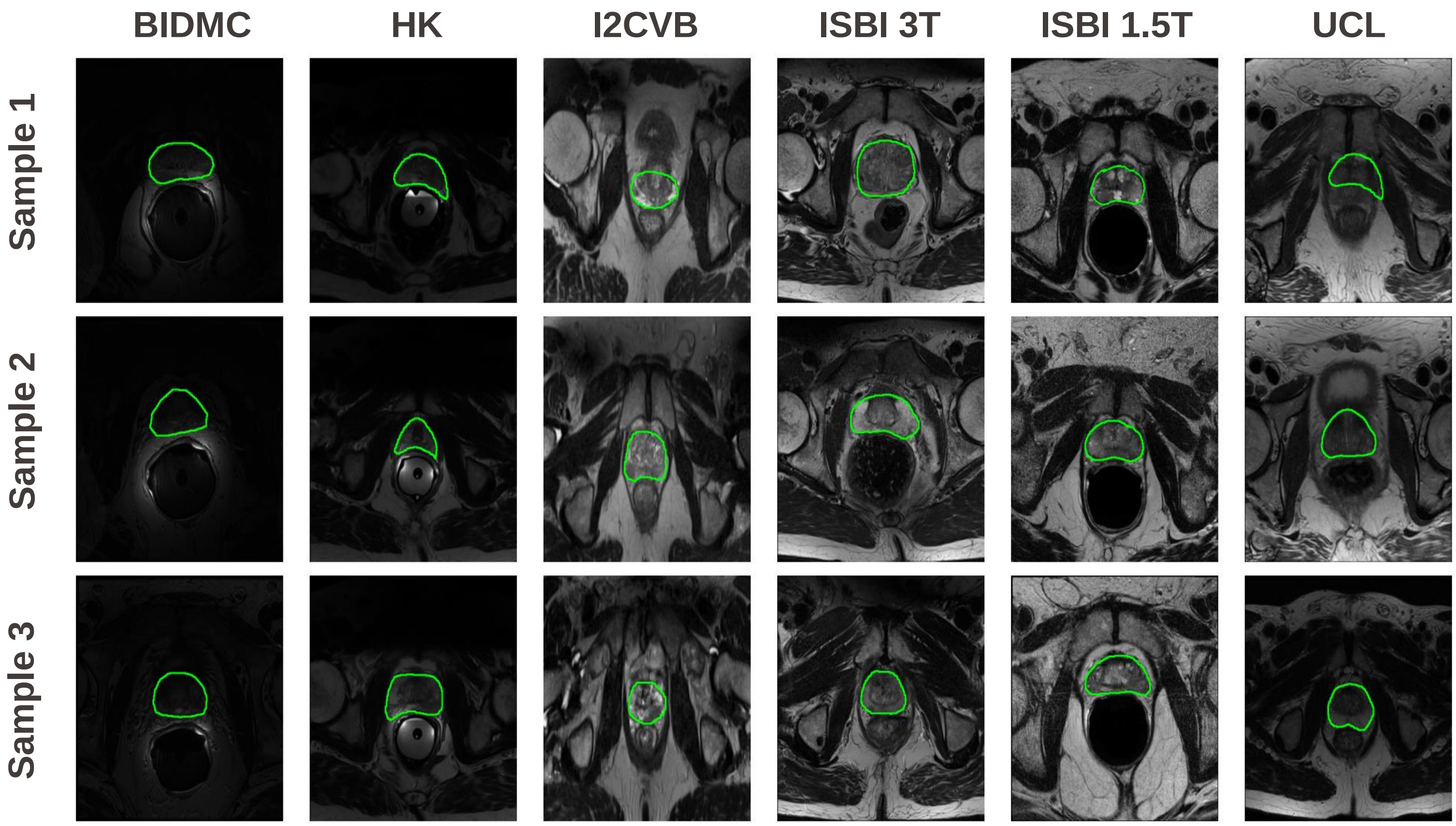
- Two datasets $\{\mathcal{D}^1, \mathcal{D}^2\}$
- Partially or fully non-overlapping space
- Alternate between each domain at each epoch



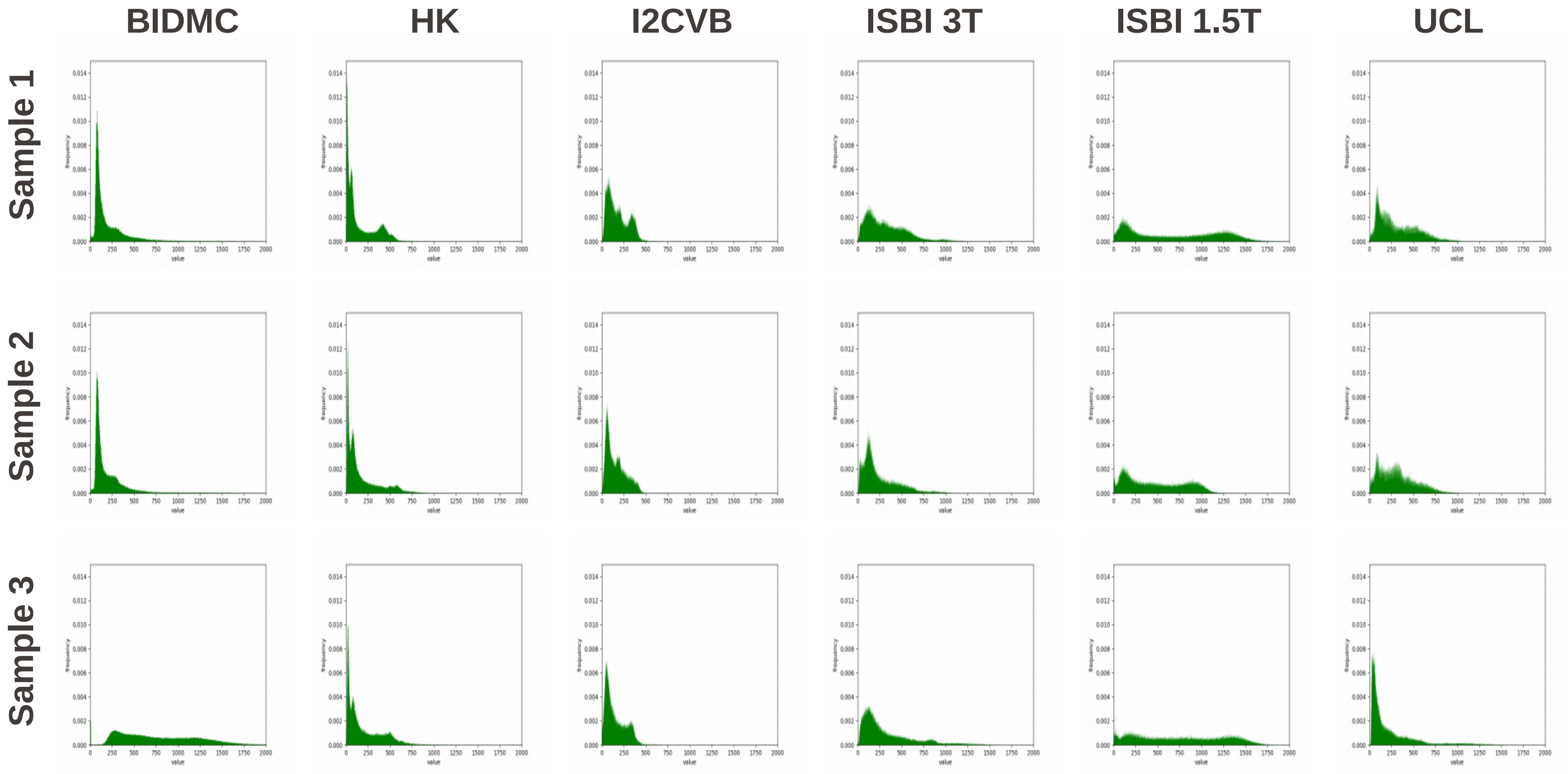
Enforcing consistency over both unlabeled dataset $\mathcal{D}_U{}^{(1)}$ and $\mathcal{D}_U{}^{(2)}$ might impose invariance of the encoder's representation over the two domains

Multi-site Prostate Segmentation of MRI data

Images



Histograms



Dataset

- T_2-weighted MRI from six different sites
- Different field strength
- MRI and segmentation volume provided as .nii files
- State of the art performance for multi-site learning : MS-Net (fully supervised)

Dataset	Institution	Case number	Field strength (T)	Resolution (in/through plane) (mm)	Endorectal coil	Manufactor
ISBI	RUNMC	30	3	0.6-0.625/3.6-4	Surface	Siemens
ISBI 1.5	BMC	30	1.5	0.4/3	Endorectal	Philips
I2CVB	HCRUDB	19	3	0.67-0.79/1.25	NO	Siemens
UCL	UCL	13	1.5 and 3	0.325-0.625/3-3.6	NO	Siemens
BIDMC	BIDMC	12	3	0.25/2.2-3	Endorectal	GE
HK	HK	12	1.5	0.625/3.6	Endorectal	Siemens

Apply the cross-consistency training on multiple domains to generalize the encoder's output to the six modalities

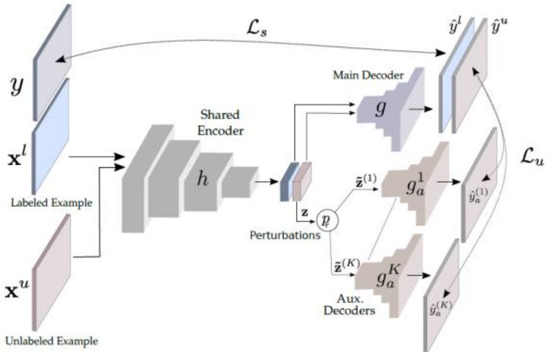
Training

Shared sub-modules

- Main encoder

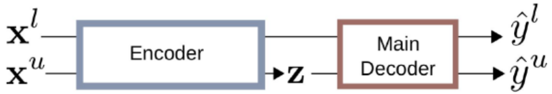
Site specific sub-modules

- Main decoder
- K auxiliary decoders

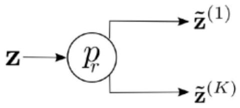


1-Iterate over each site A-B-C-D-E-F

2- Forward both labeled and unlabeled images through the encoder & main decoder:



3- Apply K perturbations to the encoder's output:



4- Compute the aux. predictions:

5- Compute the supervised and unsupervised losses:

$$\mathcal{L}_s = \frac{1}{|\mathcal{D}_l|} \sum_{\mathbf{x}_i^l, y_i \in \mathcal{D}_l} \mathbf{H}(y_i, f(\mathbf{x}_i^l))$$

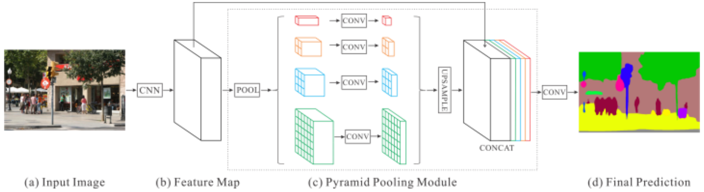
$$\mathcal{L}_u = \frac{1}{|\mathcal{D}_u|} \frac{1}{K} \sum_{\mathbf{x}_i^u \in \mathcal{D}_u} \sum_{k=1}^{K} \mathbf{d}(g(\mathbf{z}_i), g_a^k(\mathbf{z}_i))$$

$$\boxed{\mathcal{L} = \mathcal{L}_s + \omega_u \mathcal{L}_u}$$

Modules

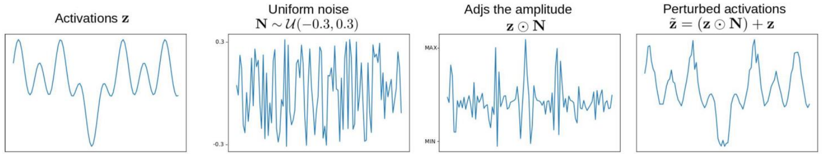
Space	Dimension
Input	Nx1x384x384
ResNet50	Nx2048x48x48
PSP	Nx512x48x48
Decoder	Nxclassesx384x384

Encoder

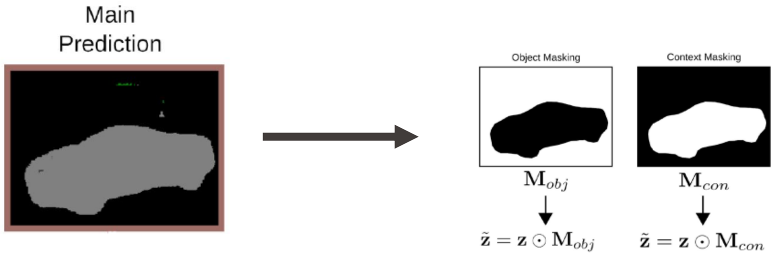


Sub-module	Quantity
Encoder	1
Main decoder	1
DropOut	6
F-Drop	6
F-noise	6
I-VAT	2
Obj-Msk	2
Con-Msk	2
G-Cutout	6

Overall



Per site



Experimental setup

Data splitting

Dataset	Supervised training	Unsupervised training	Validation	Test
ISBI	198	83	53	87
ISBI 1.5	180	70	47	87
I2CVB	219	102	73	74
UCL	94	29	27	25
BIDMC	125	42	46	38
HK	74	23	28	33

- Cases splitted across the set
- Volume converted to slices
- Only slices with prostate kept

Data Normalization

Whitening scaled imagenet (1)	Whitening scaled slices (2)	Whitening slices (3)	Case whitening (4)	Case volume whitening (5)
Slices map to [0,1]	Slices map to [0,1]	Normalization with the slices	Normalization with the volume	Volume Normalization
Normalization with imagenet	Normalization with the slices	-	-	-

Normalization experiment

Goal : Determine the best normalization strategy on single site training

Table : Dice and IoU baseline mean score with variance for different combination of upsampling and normalization on I2CVB test set

Metrics	(1) and relu	(1) and tanh	(2) and relu	(2) and tanh	(3) and relu	(3) and tanh	(4) and relu	(4) and tanh	(5) and relu	(5) and tanh
Dice background	0.768 ± 0.33	0.983 ± 0.01	0.759 ± 0.3	0.986 ± 0.01	0.995 ± 0.001	0.987 ± 0.01	0.994 ± 0.001	0.985 ± 0.01	0.806 ± 0.23	0.983± 0.01
Dice foreground	0.358 ± 0.22	0.664 ± 0.14	0.386± 0.25	0.687 ± 0.11	0.80 ± 0.03	0.696 ± 0.1	0.80 ± 0.01	0.679 ± 0.12	0.381 ± 0.24	0.662 ± 0.13
Mean Dice	0.563± 0.24	0.824± 0.07	0.573 ± 0.27	0.836 ± 0.06	0.90±0.01	0.841 ± 0.05	0.895 ± 0.01	0.832± 0.07	0.593 ± 0.23	0.823 ± 0.07
IoU background	0.701± 0.33	0.967 ±0.02	0.697 ± 0.35	0.972 ± 0.02	0.989 ± 0.001	0.974 ± 0.01	0.989 ± 0.001	0.970 ± 0.02	0.733 ± 0.29	0.967 ± 0.02
IoU foreground	0.243 ±0.16	0.52 ± 0.15	0.277 ± 0.21	0.542 ± 0.13	0.697 ± 0.03	0.548± 0.11	0.691± 0.02	0.970± 0.02	0.267 ± 0.19	0.516 ± 0.15
Mean IoU	0.472 ± 0.24	0.744±0.08	0.487± 0.27	0.758± 0.07	0.843 ± 0.01	0.761 ± 0.06	0.840 ±0.01	0.751 ± 0.08	0.50 ± 0.24	0.741 ± 0.02

Table : Dice and IoU semi mean score with variance for different combination of upsampling and normalization on I2CVB test set

Metrics	(1) and relu	(1) and tanh	(2) and relu	(2) and tanh	(3) and relu	(3) and tanh	(4) and relu	(4) and tanh	(5) and relu	(5) and tanh
Dice background	0.994 ± 0.0.001	0.995 ± 0.0	0.994 ± 0.001	0.995 ± 0.0	0.983 ± 0.0.007	0.978 ± 0.02	0.994 ± 0.001	0.981 ± 0.02	0.995 ± 0.001	0.995± 0.0
Dice foreground	0.807 ± 0.02	0.824 ± 0.003	0.79± 0.01	0.818 ± 0.0.007	0.651 ± 0.08	0.608 ± 0.17	0.795 ± 0.009	0.621 ± 0.17	0.81 ± 0.02	0.822 ± 0.004
Mean Dice	0.901± 0.001	0.909± 0.001	0.892 ± 0.007	0.907 ± 0.003	0.817±0.04	0.793 ± 0.09	0.895 ± 0.005	0.80± 0.09	0.902 ± 0.01	0.908 ± 0.0002
IoU background	0.988± 0.001	0.99 ±0.0	0.988 ± 0.001	0.99 ± 0.001	0.968 ± 0.01	0.957 ± 0.03	0.988 ± 0.001	0.963 ± 0.03	0.989 ± 0.001	0.989 ± 0.001
IoU foreground	0.70 ±0.02	0.712 ± 0.004	0.687 ± 0.02	0.705 ± 0.01	0.497 ± 0.09	0.458± 0.18	0.687± 0.02	0.481± 0.18	0.706 ± 0.02	0.711 ± 0.004
Mean IoU	0.845 ± 0.01	0.851±0.002	0.837± 0.01	0.847± 0.006	0.733 ± 0.05	0.708 ± 0.11	0.838 ±0.01	0.721 ± 0.11	0.848 ± 0.01	0.85 ± 0.002

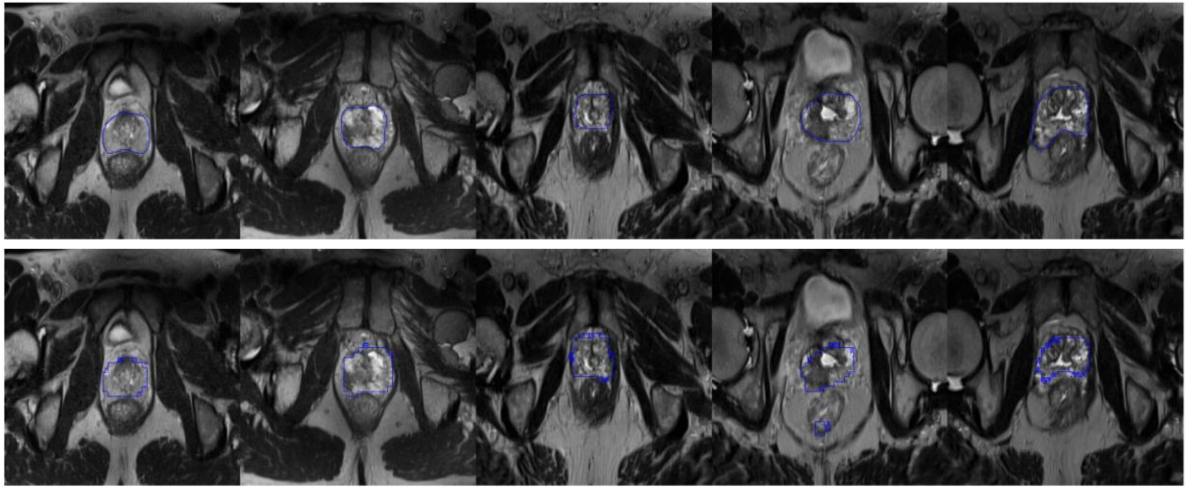
Mean and standard deviation of three seed initialization

Normalization strategy : (1) – (5)
Up-sampling strategy :

- Direct up-sampling by a factor 8 with relu activation function
- Three consecutive up-samling by a factor 2 with tanh activation function

Normalization experiment

In most of the case semi-supervised training improved the performance
We set the last combination (5) and tanh as the standard



Ground truth

Predictions

Baseline, semi and multi-semi training

Goal : Train each site individually (baseline and semi) to compare with the multi site training

Table : Dice and IoU baseline mean score with variance for each sites

Metrics	BIDMC	HK	I2CVB	ISBI	ISBI15	UCL
Dice background	0.983 ± 0.01	0.991 ± 0.002	0.983± 0.01	0.992 ± 0.001	0.993 ± 0.001	0.993 ± 0.0
Dice foreground (1)	0.669 ± 0.11	0.6 ± 0.02	0.662 ± 0.13	0.141 ± 0.1	0.356 ± 0.2	0.568 ± 0.05
Dice foreground (2)	-	-	-	0.683 ± 0.006	0.726 ± 0.02	-
Mean Dice	0.826 ± 0.06	0.796 ± 0.01	0.823 ± 0.07	0.606 ± 0.03	0.692 ± 0.07	0.781 ± 0.02
IoU background	0.967 ± 0.02	0.983 ± 0.004	0.967 ± 0.02	0.984 ± 0.001	0.985 ± 0.001	0.986 ± 0.001
IoU foreground (1)	0.546 ± 0.147	0.535 ± 0.03	0.711 ± 0.004	0.181 ± 0.13	0.314 ± 0.19	0.511 ± 0.05
IoU foreground (1)	-	-	-	0.693 ± 0.01	0.647 ± 0.02	-
Mean IoU	0.756 ± 0.09	0.759 ± 0.01	0.741 ± 0.02	0.619 ± 0.04	0.649 ± 0.07	0.748 ± 0.03

Use the (5) normalization with tanh upscaling
Mean and standard deviation of three seed initialization

Table : Dice and IoU semi-supervised mean score with variance for each sites

Metrics	BIDMC	HK	I2CVB	ISBI	ISBI15	UCL
Dice background	0.994 ± 0.001	0.989 ± 0.004	0.995± 0.0	0.992 ± 0.001	0.992 ± 0.001	0.992 ± 0.001
Dice foreground (1)	0.772 ± 0.003	0.412 ± 0.29	0.822 ± 0.004	0.129 ± 0.1	0.336 ± 0.21	0.445 ± 0.12
Dice foreground (2)	-	-	-	0.682 ± 0.02	0.722 ± 0.032	-
Mean Dice	0.882 ± 0.002	0.701 ± 0.15	0.908 ± 0.0002	0.601 ± 0.04	0.683 ± 0.08	0.719 ± 0.06
IoU background	0.987 ± 0.001	0.979 ± 0.007	0.989 ± 0.001	0.984 ± 0.001	0.984 ± 0.002	0.984 ± 0.002
IoU foreground (1)	0.694 ± 0.01	0.376 ± 0.27	0.711 ± 0.004	0.17 ± 0.13	0.305 ± 0.2	0.388 ± 0.12
IoU foreground (2)	-	-	-	0.687 ± 0.23	0.645 ± 0.04	-
Mean IoU	0.841 ± 0.005	0.678 ± 0.14	0.85 ± 0.002	0.613 ± 0.05	0.645 ± 0.08	0.686 ± 0.06

There is no clear trend of improvement from the baseline to the joint semi-supervised training

Table : Dice and IoU joint semi-supervised mean score with variance for each sites

Metrics	BIDMC	HK	I2CVB	ISBI	ISBI15	UCL
Dice background	0.994 ± 0.001	0.993 ± 0.001	0.977± 0.02	0.990 ± 0.001	0.989 ± 0.001	0.992 ± 0.001
Dice foreground (1)	0.777 ± 0.04	0.615 ± 0.1	0.598 ± 0.22	0.07 ± 0.05	0.356 ± 0.05	0.424 ± 0.1
Dice foreground (2)	-	-	-	0.633 ± 0.008	0.17 ± 0.12	-
Mean Dice	0.886 ± 0.02	0.804 ± 0.05	0.787 ± 0.12	0.564 ± 0.015	0.504 ± 0.049	0.708 ± 0.05
IoU background	0.989 ± 0.001	0.986 ± 0.003	0.956 ± 0.04	0.981 ± 0.001	0.979 ± 0.01	0.983 ± 0.003
IoU foreground (1)	0.733 ± 0.04	0.581 ± 0.09	0.477 ± 0.23	0.085 ± 0.06	0.283 ± 0.09	0.342 ± 0.11
IoU foreground (2)	-	-	-	0.641 ± 0.024	0.104 ± 0.08	-
Mean IoU	0.861 ± 0.0204	0.784 ± 0.05	0.716 ± 0.14	0.569 ± 0.02	0.455 ± 0.04	0.663 ± 0.06

Single-site and joint-sites training (1)

Goal : Train each site individually (semi) to compare with the multi site training

(s) : single site training
(m) : joint training with the six sites

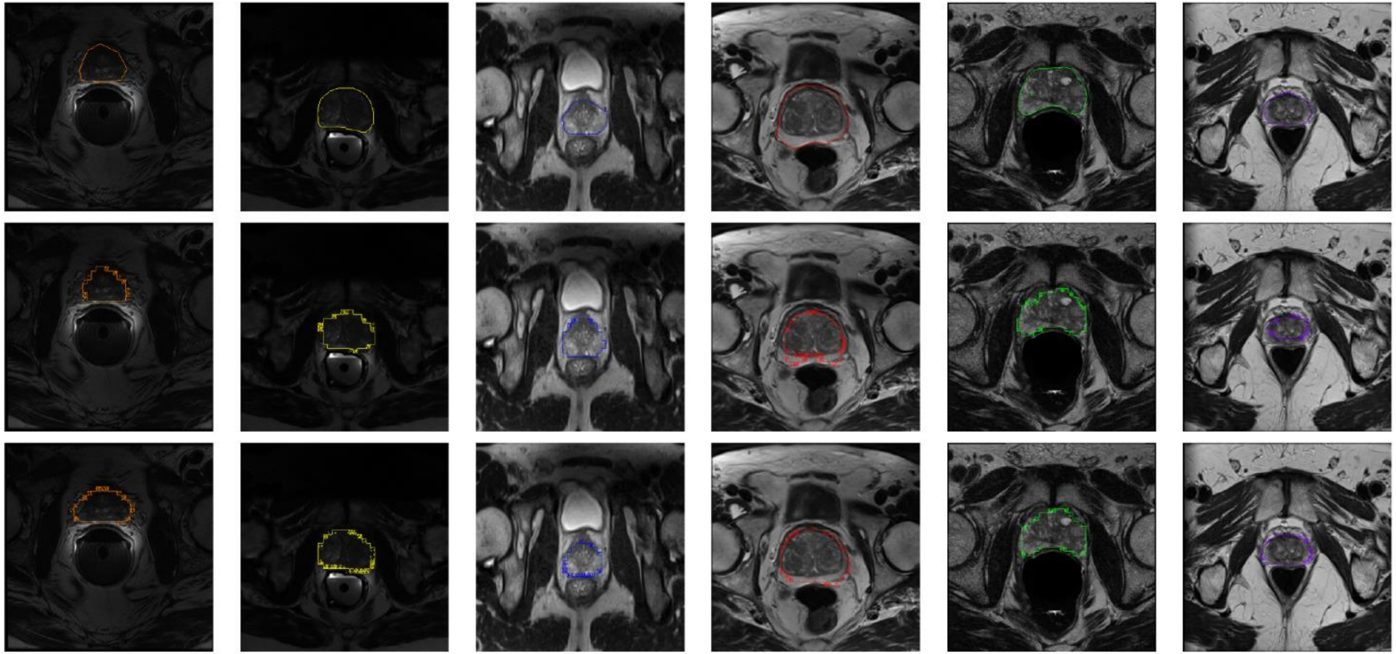
Use the (1) normalization with relu upscaling
Training from seed initialization

Table : Dice and IoU score for single and joint training unsing (1) normalization with relu on test set

Metrics	BIDMC (s)	BIDMC (m)	HK (s)	HK (m)	I2CVB (s)	I2CVB (m)	ISBI (s)	ISBI (m)	ISBI15 (s)	ISBI15 (m)	UCL (s)	UCL (m)
Dice background	0.995	0.995	0.994	0.995	0.995	0.995	0.991	0.99	0.992	0.993	0.991	0.996
Dice foreground	0.76	0.781	0.671	0.707	0.799	0.823	0.191/0.691	0.218/0.688	0.527/0.718	0.551/0.688	0.495	0.748
Mean Dice	0.878	0.888	0.832	0.851	0.897	0.909	0.624	0.632	0.745	0.773	0.743	0.872
IoU background	0.99	0.991	0.988	0.989	0.99	0.991	0.982	0.98	0.984	0.987	0.983	0.992
IoU foreground	0.735	0.767	0.615	0.664	0.698	0.728	0.319/0.667	0.359/0.597	0.5/0.628	0.507/0.695	0.422	0.736
Mean IoU	0.862	0.879	0.801	0.827	0.844	0.86	0.656	0.645	0.704	0.73	0.702	0.864

It seems that the joint semi supervised training increases the performance compared to the single semi-supervised training

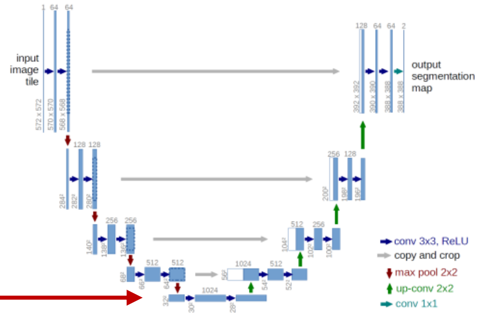
Single-site and joint-sites training (2)



Unet

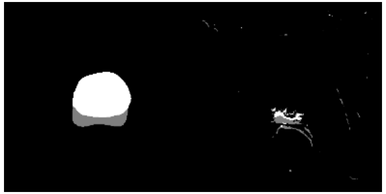
Goal : Use the cross-consistency framework with the unet architecture as encoder/decoder pair

Perturbations applied at the last level of the down sampling

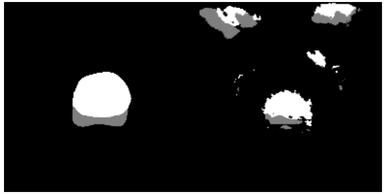


Semi-supervised training

- Auxiliary decoders set to a maximum of 2
- Trained only on ISBI 1,5 for one seed
- 50 and 100 epochs



50



100

Conclusion

Drawbacks

- Should perform a normalization study combining every sites
- Should compute results for more than one seed in the second experiment
- Should perform a complete study on the unet architecture

Discussion

- The resnet50-PSP/decoder module seems not to be able to perform well on this dataset
- The unet architecture could be able to perform well

Future directions

- Self-supervised learning

EPFL

Thank you

Emery Sébastien

École
polytechnique
fédérale
de Lausanne

Reference papers

[1] Yassine Ouali, Celine Hudelot, and Myriam Tami. Semi-supervised semantic segmentation with cross-consistency training. In The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2020.

[2] Quande Liu, Qi Dou, Lequan Yu, and Pheng Ann Heng. Ms-net: Multi-site network for improving prostate segmentation with heterogeneous mri data. IEEE Transactions on Medical Imaging, 2020.

[3] Quande Liu, Qi Dou, and Pheng Ann Heng. Shape-aware meta-learning for generalizing prostate mri segmentation to unseen domains. In International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), 2020.

[4] Dong-Hyun Lee. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In Workshop on Challenges in Representation Learning, ICML, volume 3, page 2, 2013.

[5] Yves Grandvalet and Yoshua Bengio. Semi-supervised learning by entropy minimization. In Advances in neural information processing systems, pages 529–536, 2005.

[6] Siyuan Qiao,Wei Shen, Zhishuai Zhang, BoWang, and Alan Yuille. Deep co-training for semi-supervised image recognition. In Proceedings of the European Conference on Computer Vision (ECCV), pages 135–152, 2018.