SMILE, a deep learning method for training ConvNets with partially annotated data for the task of semantic segmentation.

GdR ISIS

Olivier Petit

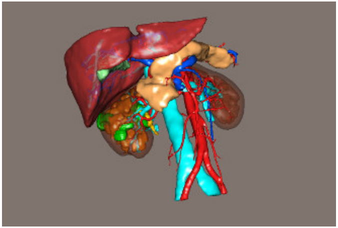
Visible Patient and le CNAM

10/05/2019

Table of contents

Introduction and Motivations

This work focuses on **organ segmentation** in abdominal CT-scans with **Deep Convolutional neural Networks** (ConvNets)



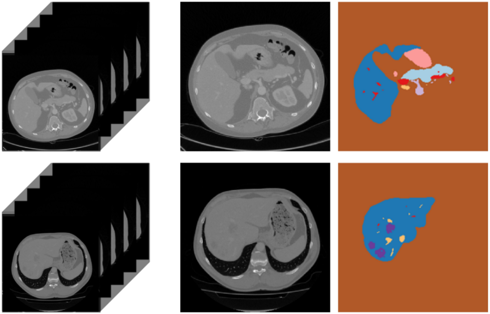
Problem

- training deep ConvNets requires large amount of data
- the annotation process is extremely time consuming and requires high qualified professionals
- clinical experts focus on specific organs or anatomical structures

Introduction and Motivations

Example

Our dataset from VP/IRCAD is partially annotated



Problem

How can we train a ConvNet ?

SMILE method

SMILE: Semantic segmentation with MIssing Labels and ConvNEts [Petit et al., 2018]

The main ideas:

- Learning only with good annotations, and ignore uncertain ones
- SMILEr $\rightarrow$ *semi-supervised* method with reannotation

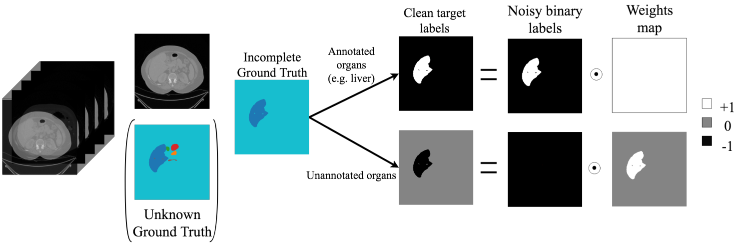
Hypothesis

- If there is an annotation for an organ, it's complete in the entire volume.
- All organs are visible in the 3D image.

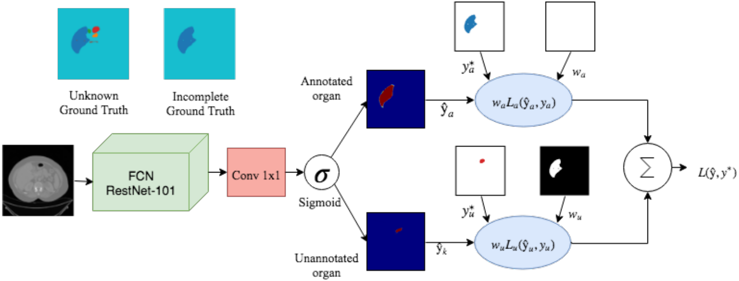
SMILE method

Handling missing annotations

- The first step is to consider K binary classifiers instead of 1 multiclass (K+1) classifier (replacing the softmax activation by a sigmoid)
- We introduce an ambiguity map $\mathbf{W}$; $w_c \in \{0, 1\}$ to ignore ambiguous annotations



SMILE method



Loss function

Each binary classifier has a binary cross-entropy loss :

$$L_k(\hat{y}_k, y_k^*) = -(y_k^* \ log(\hat{y}_k) + (1 - y_k^*) \ log(1 - \hat{y}_k))$$

The final loss is the aggregation of the K losses :

$$L(\hat{y}, y^*) = \sum_{k=1}^{K} w_k \ L_k(\hat{y}_k, y_k^*)$$

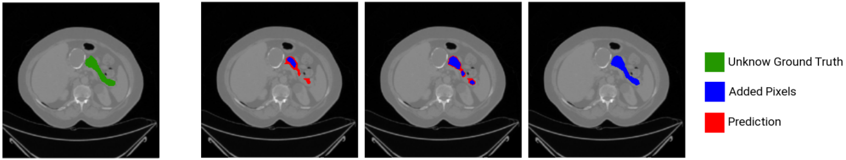
SMILEr incremental self-supervision and relabeling

Reannotation of the missing labels through self-supervision : **curriculum learning strategy** [Bengio et al., 2009].

Procedure

Initialization with SMILE (easy examples);
for $t \leftarrow 1$ *to* T **do**
  Select $\gamma^t = \frac{t}{T}\gamma_{max}$ top scoring pixels among $\hat{y}_i^+$;
  Train with the new labels (hard examples);
end

Interative reannotation $(T = 3,\ \gamma_{max} = 1.0)$



Unknown Ground Truth | $t = 1,\ \gamma_1 = 0.33$ | $t = 2,\ \gamma_2 = 0.66$ | $t = 3,\ \gamma_3 = 1.0$

Experimentations

Experimental setup

Model

FCN based on a ResNet-101 [He et al., 2016].
We train:

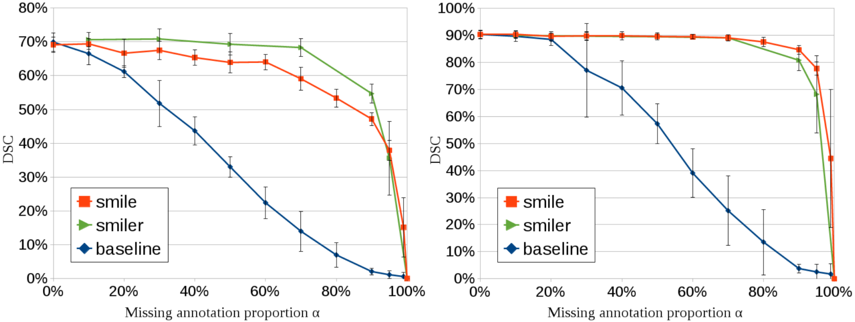
- A baseline model on the raw data
- The same model with the SMILE and SMILEr method

Data

Initial data: 72 CT-scans (complete liver, pancreas and stomach)
Training data: we remove α% of each annotation $\alpha = \{0 \rightarrow 100\%\}$
Split: 80% training; 20% testing

Experimentations

Results

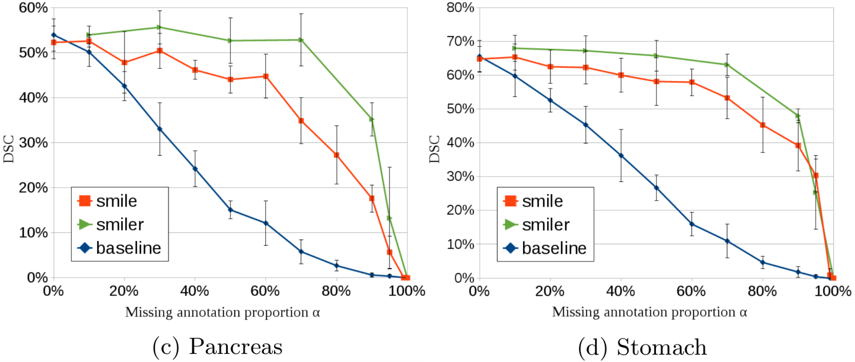


(a) Mean (b) Liver

SMILEr $\alpha = 70\% \sim$ baseline $\alpha = 0\%$

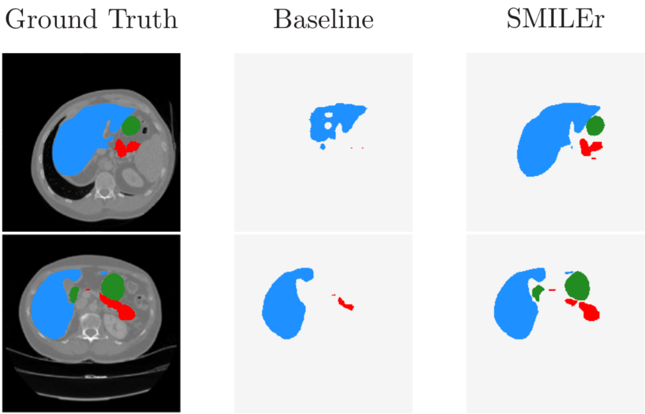
Experimentations

Results



SMILEr: improvements are more pronounced for small organs like the pancreas and the stomach

Qualitative results



Conclusion

We proposed a method for training deep ConvNets on partially annotated data.

- We showed that SMILEr can achieve comparable performances to a model trained with complete annotations with only 30% of the labels

Perspectives

- Training better FCN architectures (e.g. U-Net [Ronneberger et al., 2015])
- Improving the pixel selection for SMILEr (e.g. Bayesian uncertainty criterion [Gal and Ghahramani, 2016])

References I

Bengio, Y., Louradour, J., Collobert, R., and Weston, J. (2009).
Curriculum learning.
In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, pages 41–48.

Gal, Y. and Ghahramani, Z. (2016).
Dropout as a Bayesian approximation: Representing model uncertainty in deep learning.
In *Proceedings of the 33rd International Conference on Machine Learning (ICML-16)*.

He, K., Zhang, X., Ren, S., and Sun, J. (2016).
Deep residual learning for image recognition.
In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 770–778.

References II

Petit, O., Thome, N., Charnoz, A., Hostettler, A., and Soler, L. (2018).
Handling missing annotations for semantic segmentation with deep convnets.
In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 20–28, Cham. Springer International Publishing.

Ronneberger, O., Fischer, P., and Brox, T. (2015).
U-Net: Convolutional Networks for Biomedical Image Segmentation.
In *MICCAI (3)*, volume 9351 of *Lecture Notes in Computer Science*, pages 234–241. Springer.

Appendix

Analyzing the TP/FP ratio

- β_k = voxel ratio for organ k
- α = unannotated organ ratio

Baseline FCN

GT \ Used	Pos	Neg
Pos	$(1-\alpha)\cdot\beta_k$	$\alpha\cdot\beta_k$
Neg	0	$1-\beta_k$

SMILE

GT \ Used	Pos	Neg
Pos	$(1-\alpha)\cdot\beta_k$	0
Neg	0	$(1-\alpha)\cdot(1-\beta_k)+\epsilon$

- Baseline : we learn with all the TN but also all the FN
- SMILE : we learn with no FN but we have removed some TN

Because the number of background labels is high, removing some of the TN have no incidence on the training.