Few-shots learning of anatomic and oncologic structures in radiology

RAIDIUM Data Challenge Report

Julien AJDENBAUM

Project for Stéphane MALLAT's class *"Apprentissage et génération par échantillonnage aléatoire"* at *Collège de France*

MVA - Télécom Paris

March 19, 2024

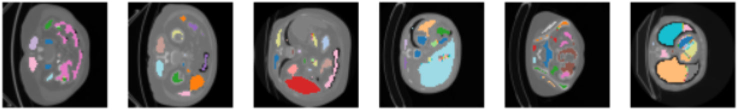
Table des matières

Problem Presentation

- 2D CT-Scan images
- Anatomical and oncological structures segmentation
- Class-Agnostic
- 2k images, 20% Annotated

How to measure a segmentation's accuracy?

$\implies$ Average Rand Index

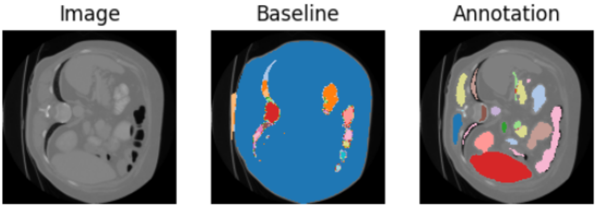


Early Tests: Baseline

Baseline:

- Sobel Filter
- Blurring Convolution
- Gradient of denoised image
- Watershed algorithm

Not too bad: 14% RI Accuracy

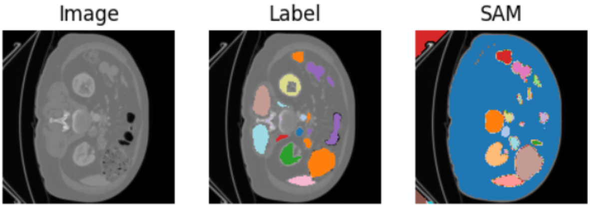


Early Tests: SAM

Segment Anything [1]:

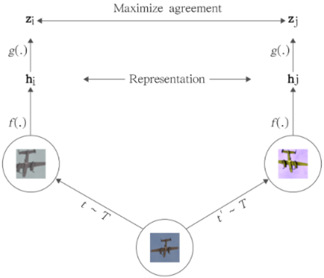
- Foundation model for class-agnostic segmentation
- No prompting

14 % RI Accuracy
$\Longrightarrow$ Much longer and more complex than baseline for the same results



SSL: SimCLR

Similarity Contrastive Learning [2] SSL pret-training only the encoder: in our case ResNet-50



$\Longrightarrow$ the decoder cannot be pre-trained

SSL: SimCLR

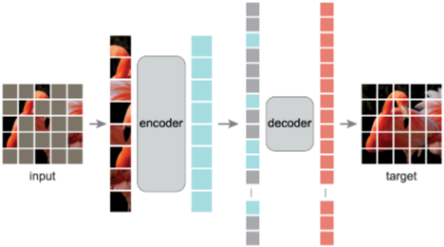
NT-Xent :

$$\ell(i,j) = -\log \frac{\exp\left(s_{i,j}/\tau\right)}{\sum_{k=1}^{2N} \not\Vdash_{[k\neq i]} \exp\left(s_{i,k}/\tau\right)} \tag{1}$$

Conclusion: Not enough images

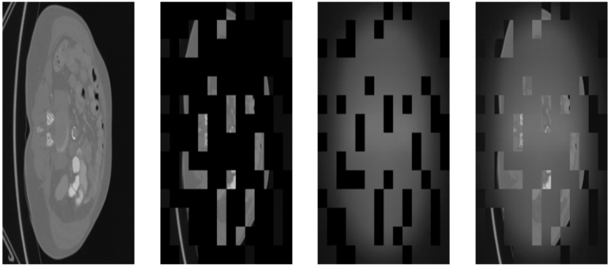
Masked Auto-Encoders

SSL Pre-training of an entire AE, with masked Auto-Encoders [3]



Masked Auto-Encoders

Due to the low number of images, we use Swin [4], an MAE specifically developped for the segementation of images with few examples



Dice metrics related works

$\Longrightarrow$ After unsuccessful attempts at Self Supervised Learning, I wanted something simple yet efficient

$\Longrightarrow$ The difficulty was not really segmenting the areas themselves, but more about understanding which segments went together and which did not

$\Longrightarrow$ This led me to a paper called Adaptive t-vMF Dice Metric [5]

Dice metrics related works

We take the Dice Loss:

$$\text{Dice loss } = \frac{1}{C}\sum_{i=1}^{C}\left(1 - \frac{2\sum_n A_{in}B_{in} + \gamma}{\sum_n A_{in}^2 + \sum_n B_{in}^2 + \gamma}\right) \tag{2}$$

We normalize the regions :

$$\text{Dice loss}_{\text{norm}} = \frac{1}{C}\sum_{i=1}^{C}(1 - \cos\theta_i) \tag{3}$$

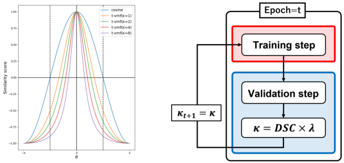
Finally, we extend the cosine similarity by using the t-vMF similarity :

$$\phi_t(\cos\theta;\kappa) = \frac{1+\cos\theta}{1+\kappa(1-\cos\theta)} - 1 \tag{4}$$

Including this t-vMF, we obtain the following final t-vMF Dice Loss :

$$t\text{-vMF Dice loss } = \frac{1}{C}\sum_{i=1}^{C}\left(1 - \phi_t\left(\cos\theta_i;\kappa\right)\right)^2 \tag{5}$$

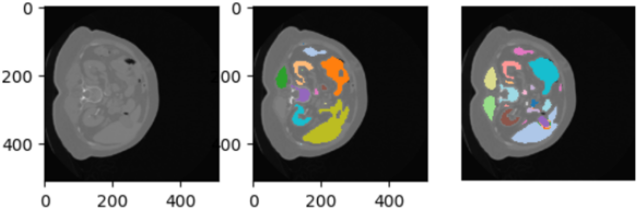
Dice metrics related works



(a) (b)

Fig. 1: (a) Comparison of cosine similarity and t-vMF-based similarities. (b) Overview of the training and validation flows using Adaptive t-vMF Dice loss.

In order to get classes, we just label disconnected regions.
$\Longrightarrow$Acceptable results : 86% Rand Index Score on the validation set
$\Longrightarrow$Due to zero-shot cases, 74% Rand Index Score on the test set
Main problem: labeling the regions



Conclusion : What would I have do differently ?

Simple Auto-Encoder, trained on the whole dataset
Find a loss adapted to class-agnostic segmentation
$\Longrightarrow$ Not try SSL, not enough images

Conclusion : What would I have do differently ?

- A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick, "Segment anything," 2023.
- T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," 2020.
- K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," 2021.
- Z. Xu, Y. Dai, F. Liu, W. Chen, Y. Liu, L. Shi, S. Liu, and Y. Zhou, "Swin mae: Masked autoencoders for small datasets," 2023.
- S. Kato and K. Hotta, "Adaptive t-vmf dice loss for multi-class medical image segmentation," 2022.