MICCAI 2024

Marrakesh

MOROCCO

27TH INTERNATIONAL CONFERENCE ON MEDICAL IMAGE COMPUTING AND COMPUTER ASSISTED INTERVENTION

6-10 OCTOBER 2024

PALMERAIE ROTANA RESORT

MARRAKESH / MOROCCO

SSL based encoder pre-training for segmenting a heterogeneous chronic wound image database with few annotations

Guillaume PICAUD[1,3], Marc CHAUMONT[1,2], Gérard SUBSOL[1], Luc TEOT[3]

{guillaume.picaud, marc.chaumont, gerard.subsol}@lirmm.fr, l-teot@chu-montpellier.fr

[1]LIRMM, équipe ICAR, Univ. Montpellier, CNRS, France
[2]Univ. Nîmes Place Gabriel Péri, France
[3]Cicat-Occitanie, Montpellier, France

A large wound image database to segment

Teleconsultation center

Cicat Occitanie

- 130,000 images to segment
- Images ”into the wild”
- All chronic wound types
- A very heterogeneous database

A large wound image database to segment

Teleconsultation center

Cicat Occitanie

→ But only 400 images manually labeled by expert

What approach to use?

- Direct supervised approaches (see DFUC2022 *[Yap et al. 2024]*) compromised

→ Try a Self Supervised Learning (SSL) approach :

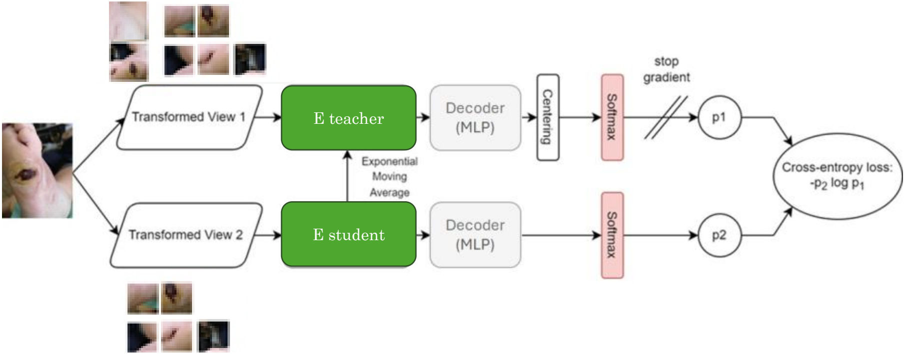
- Based on a large unlabeled database and a small labelled database
- Step 1: Pre-train an encoder with a ‘’pretext task’’ on the large unlabeled database
- Step 2: Fine-tune with few labelled data with the small labelled database
- A SOTA method : Distillation with no label DINO [Caron et al. 2021 ; Oquab et al.2023]

Yap et al. 2024. “Diabetic foot ulcers segmentation challenge report : Benchmark and analysis” Medical Image Analysis

Caron et al. 2021. “Emerging properties in self-supervised vision transformers” *ICCV*

Oquab et al. 2023. “Dinov2: Learning robust visual features without supervision”

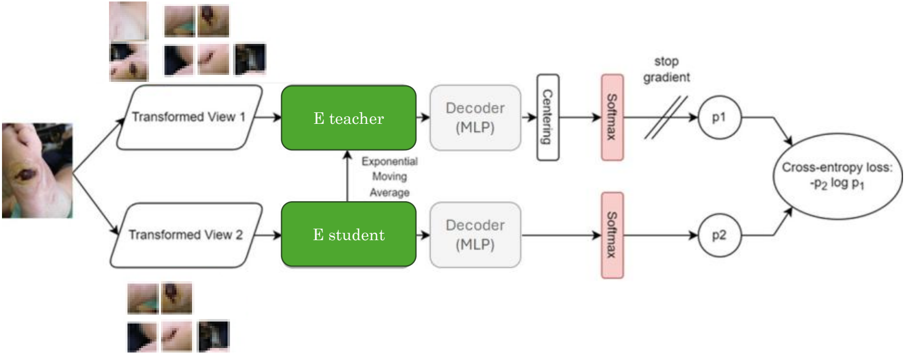
Step 1: Pre-train encoder with a "pretext task"



- Here, pretext task = *Build a robust model whatever the considered image part*

  - Random Local and global crops
  - Teacher-Student architecture based on a same encoder
  - Updating Student weights using cross-entropy loss
  - Updating Teacher weights applying EMA on Student weights

Step 1: Pre-train encoder with a "pretext task"

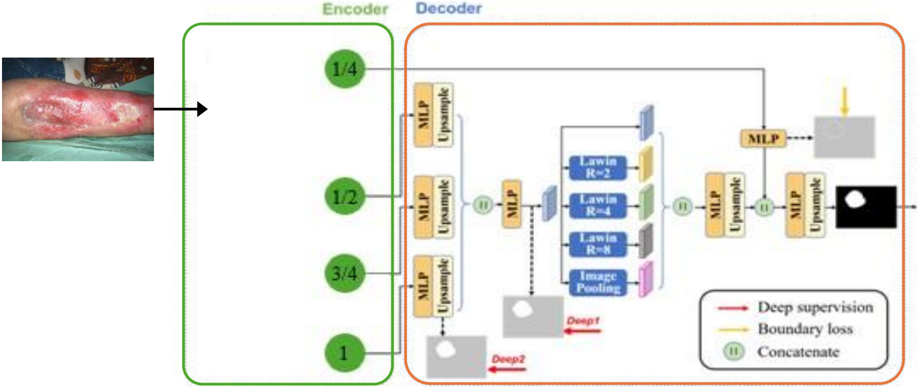


- Selected Encoders :
  - A small Transformer ViTs14_reg *[Darcet et al. 2023]* (21 M) weights from LVD-142M
  - A large Transformer ViTl14_reg *[Darcet et al. 2023]* (307 M) weights from LVD-142M
  - A dedicated lightweight encoder HardNet-DFUS *[Ting-Yu et al. 2022]* (3 M) random weights

Darcet et al. 2023. "Vision transformers need registers" *arXiv*
Ting-Yu et al. 2022. *HarDNet-DFUS: An Enhanced Harmonically-Connected Network for Diabetic Foot Ulcer Image Segmentation and Colonoscopy Polyp Segmentation arXiv*

Step 2: Fine-tune with few labelled data

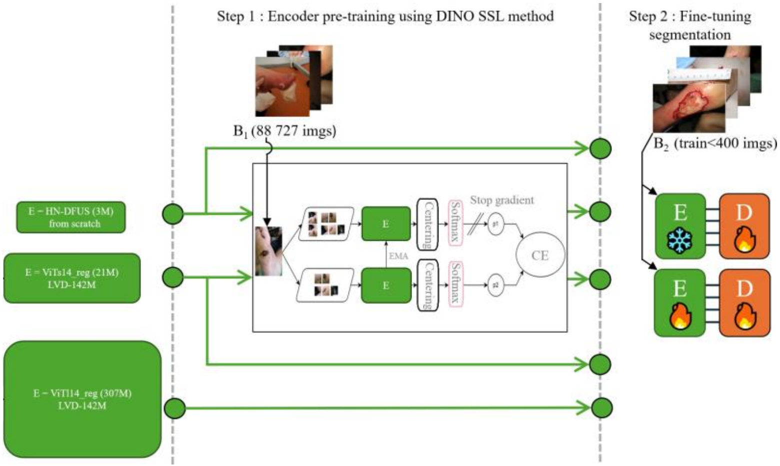
- Selected decoder: Lawin Transformer from HardNet-DFUS *[Ting-Yu et al. 2022]*



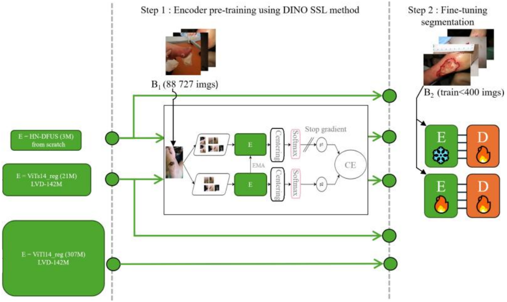
$$Loss = l^{w}_{BCE}(G, P) + l^{w}_{IoU}(G, P) + l_{BCE}(G_B, P_B) + \sum_{i=1}^{n} [l^{w}_{BCE}(G, P_i) + l^{w}_{IoU}(G, P_i)]$$

- G ground truth ; G_B boundary ground truth
- P final prediction ; P_B boundary prediction ; P_i prediction block i

Experimental test bench



Experimental test bench



B_2 = 400 imgs

Train 70%	Val 20%	Test 10%
280 (i.e B_2^{280})	80	40
140 (i.e B_2^{140})	80	40
70 (i.e B_2^{70})	80	40

x5

Performances evaluation with Dice metrics

Experiences : Results

Performances evaluation with Dice metrics

Encodeur	SSL(B_1)	Optimisation (B_2)	B_2^{70}	B_2^{140}	B_2^{280}
HardNet-MSEG$_{rdm}$	✔	❄	0.72 ± 0.03	0.74 ± 0.01	0.76 ± 0.02
		🔥	0.76 ± 0.03	0.78 ± 0.01	0.80 ± 0.01
	✖	🔥	0.70 ± 0.04	0.74 ± 0.03	0.77 ± 0.02
ViTs14_reg	✔	❄	0.59 ± 0.06	0.64 ± 0.04	0.67 ± 0.03
		🔥	0.67 ± 0.02	0.71 ± 0.02	0.72 ± 0.03
	✖	❄	0.57 ± 0.04	0.65 ± 0.03	0.65 ± 0.02
		🔥	0.69 ± 0.03	0.72 ± 0.02	0.73 ± 0.01
ViTl14_reg	✖	❄	0.64 ± 0.04	0.64 ± 0.02	0.70 ± 0.03

- No SSL ✖ → parameters ↑= *performances* ↑

Experiences : Results

Performances evaluation with Dice metrics

Encodeur	SSL(B_1)	Optimisation (B_2)	B_2^{70}	B_2^{140}	B_2^{280}
HardNet-MSEG$_{rdm}$	✓	❄	0.72 ± 0.03	0.74 ± 0.01	0.76 ± 0.02
		🔥	0.76 ± 0.03	0.78 ± 0.01	0.80 ± 0.01
	✗	🔥	0.70 ± 0.04	0.74 ± 0.03	0.77 ± 0.02
ViTs14_reg	✓	❄	0.59 ± 0.06	0.64 ± 0.04	0.67 ± 0.03
		🔥	0.67 ± 0.02	0.71 ± 0.02	0.72 ± 0.03
	✗	❄	0.57 ± 0.04	0.65 ± 0.03	0.65 ± 0.02
		🔥	0.69 ± 0.03	0.72 ± 0.02	0.73 ± 0.01
ViTl14_reg	✗	❄	0.64 ± 0.04	0.64 ± 0.02	0.70 ± 0.03

- No SSL ✗ → parameters ↑= *performances* ↑
- No SSL ✗ → dedicated lightweight encoder better

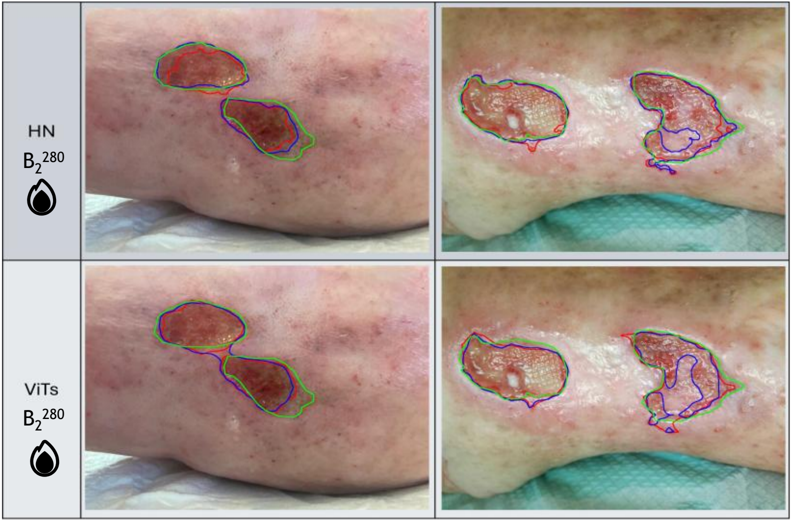
Experiences : Results

Performances evaluation with Dice metrics

Encodeur	SSL(B_1)	Optimisation (B_2)	B_2^{70}	B_2^{140}	B_2^{280}
HardNet-MSEG$_{rdm}$	✔	❄	0.72 ± 0.03	0.74 ± 0.01	0.76 ± 0.02
		🔥	0.76 ± 0.03	0.78 ± 0.01	0.80 ± 0.01
	✖	🔥	0.70 ± 0.04	0.74 ± 0.03	0.77 ± 0.02
ViTs14_reg	✔	❄	0.59 ± 0.06	0.64 ± 0.04	0.67 ± 0.03
		🔥	0.67 ± 0.02	0.71 ± 0.02	0.72 ± 0.03
	✖	❄	0.57 ± 0.04	0.65 ± 0.03	0.65 ± 0.02
		🔥	0.69 ± 0.03	0.72 ± 0.02	0.73 ± 0.01
ViTl14_reg	✖	❄	0.64 ± 0.04	0.64 ± 0.02	0.70 ± 0.03

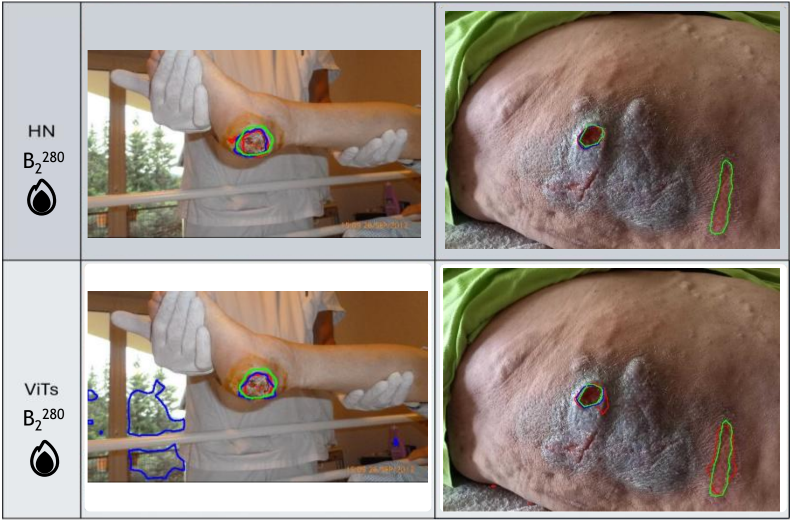
- No SSL ✖ → parameters ↑= *performances* ↑
- No SSL ✖ → dedicated lightweight encoder better
- With SSL ✔ → **much better performances** on dedicated lightweight encoder
- With SSL ✔ → **gain performances ↑ when training database size ↓**

Inference Visualization



Green: GT
blue No SSL
red with SSL

Inference Visualization



Green: GT
blue No SSL

red with SSL

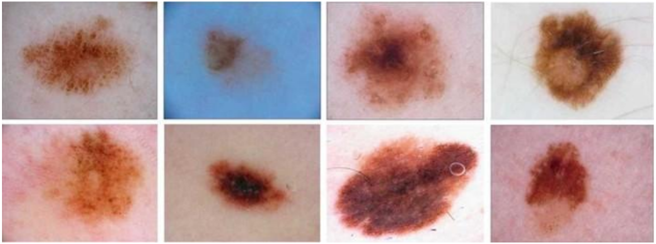
Conclusion

In the context of a large unlabelled database and small annotated database :

SSL > Direct Supervised approach

Perspectives : Other pretext task

Perspectives : Expanding B_1 with databases from other domains[6] :



[6]https://www.isic-archive.com/

Thank you for your attention!