

Analyse de l'encodeur pour la segmentation d'une base de données hétérogène de photographies de plaies chroniques avec peu d'annotations

Guillaume PICAUD^{1,3}, Marc CHAUMONT^{1,2}, Gérard SUBSOL¹, Luc TEOT³
{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



1. Contexte : les plaies chroniques

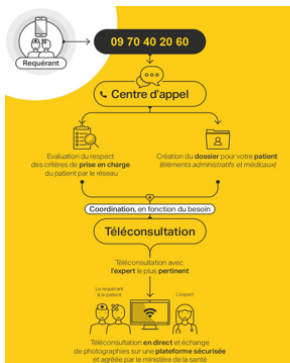
- 4 à 6 semaines d'évolution sans cicatrisation
- ulcère 66%, escarre 23%, pied diabétique 11%
- 700 000 porteurs, 8 000 amputations, 3 500 décès/an en France



3 plaies chroniques : pied diabétique, ulcère, escarre

1. Contexte : le Cicat-Occitanie et la télémédecine

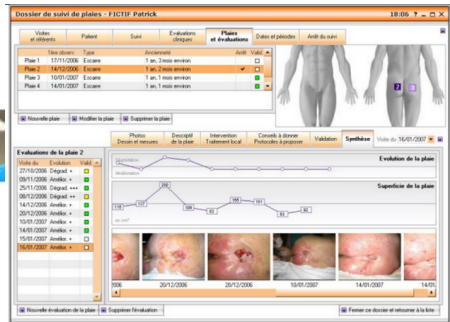
- La téléconsultation : une aide aux équipes paramédicales¹



¹Téot et al. 2020. "Complex wound healing outcomes for outpatients receiving care via telemedicine, home health, or wound clinic: a randomized controlled trial" *The International Journal of Lower Extremity Wounds*

1. Contexte : base de données Cicat-Occitanie

- Dossiers médicaux de 19 000 patients
- 133 000 images
- Acquisition non contrôlée : appareils, observateurs, environnements, points de vue, ...



1. Contexte : comparaison de bases de données



- Pas d'annotation
- Images "into the wild"
- Tout type de plaies

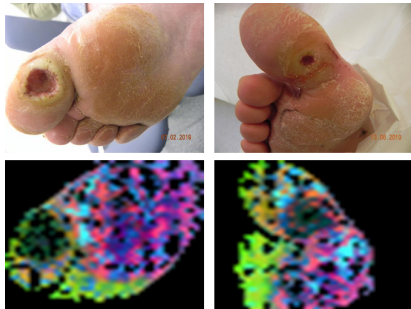
- 2 000 images annotées²
- Acquisitions contrôlées
- Plaies du pied diabétique



²Kendrick Connah et al. 2022. "Translating Clinical Delineation of Diabetic Foot Ulcers into Machine Interpretable Segmentation" *arXiv*

1. Contexte : différentes approches pour la segmentation

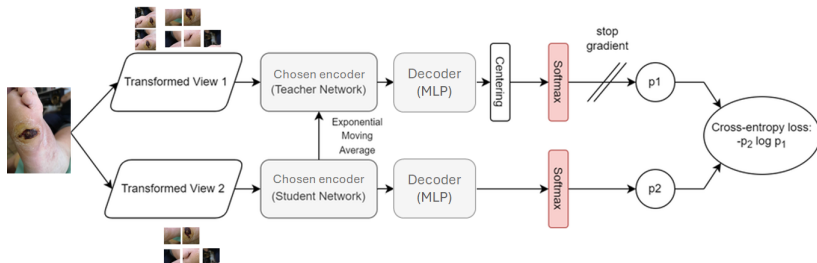
- Approche supervisée compromise :
 - Peu de données annotées
 - Changement de domaine
- Approche autosupervisée (SSL) :
 - Base de données non annotées
 - Pretraining encodeur tâche prétexte
 - Fine-tuning avec peu de données
 - DINO : Distillation with no label^{3,4}



³Caron et al. 2021. "Emerging properties in self-supervised vision transformers" ICCV

⁴Oquab et al. 2023. "Dinov2: Learning robust visual features without supervision" arXiv

2.1 SSL DINO : Distillation with no labels

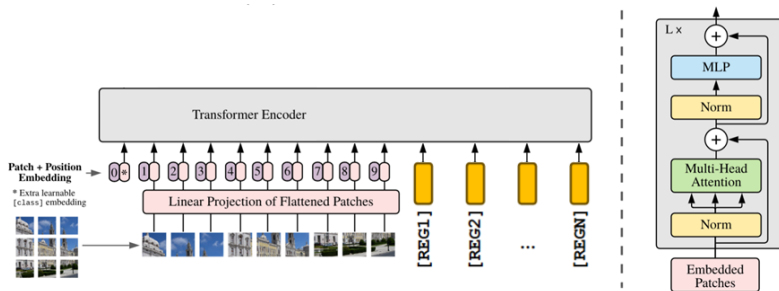


- Base de données LVD-142M pour l'apprentissage des encodeurs^{5,6}

³Caron et al. 2021. "Emerging properties in self-supervised vision transformers" *ICCV*

⁴Oquab et al. 2023. "Dinov2: Learning robust visual features without supervision" *arXiv*

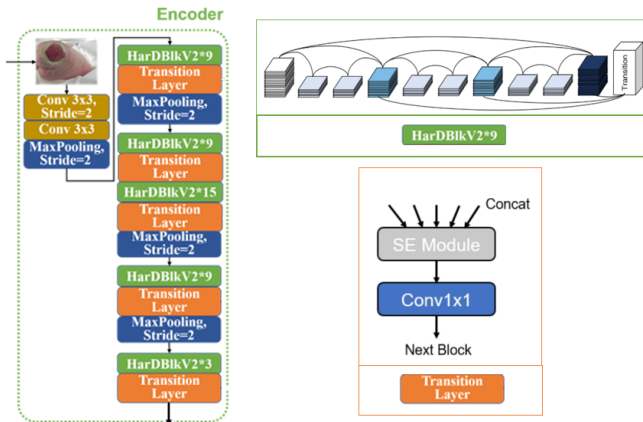
2.2 Encodeurs génériques : les Vision Transformers



- ViTs14_reg : 21M params
- ViTl14_regs : 307M params

⁶Timothée Darcet et al. 2023. "Vision transformers need registers" *arXiv preprint arXiv:2309.16588*

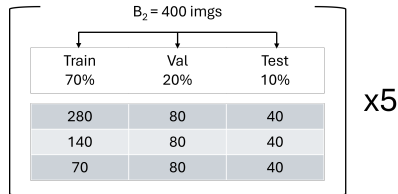
2.3 Encodeur spécialisé : HardNet-DFUS + Lawin Transformer



⁷Ting-Yu et al. 2022. *HardNet-DFUS: An Enhanced Harmonically-Connected Network for Diabetic Foot Ulcer Image Segmentation and Colonoscopy Polyp Segmentation* arXiv

3.1 Protocole : Base de données

- B_1 : 88 727 images (1 plaie/image) du Cicat-Occitanie
- B_2 : 400 images annotées manuellement par un expert
- Pre-training SSL sur B_1
- Fine-tuning supervisé segmentation B_2 train $\Rightarrow$ B_2 test (5 fois)



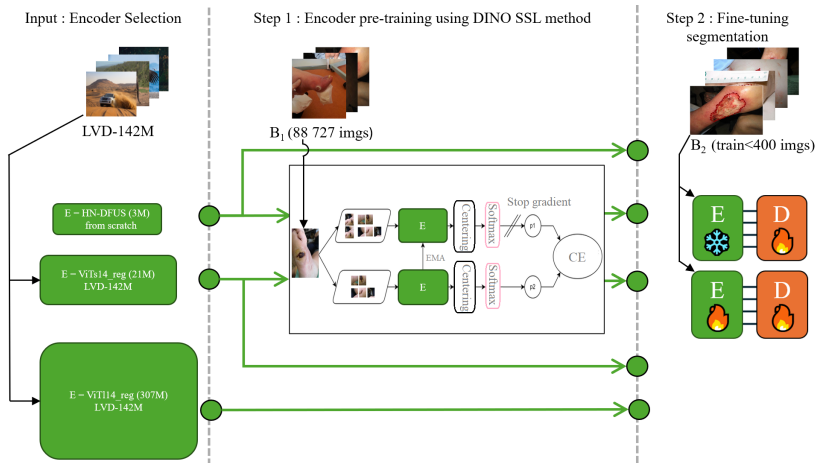
$B_1 = 88\,727$ imgs



$B_2 = 400$ imgs



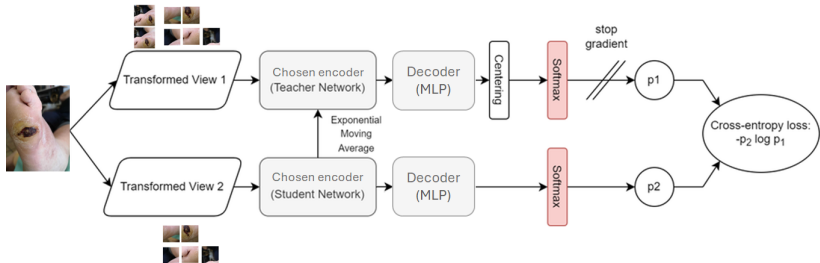
3.2 Expériences : Scénarios d'entraînement



- Initialisation : Choix de l'encodeur
- Etape 1 : Pre-training "à la DINO" sur B_1
- Etape 2 : Fine-tuning supervisé en segmentation sur B_2

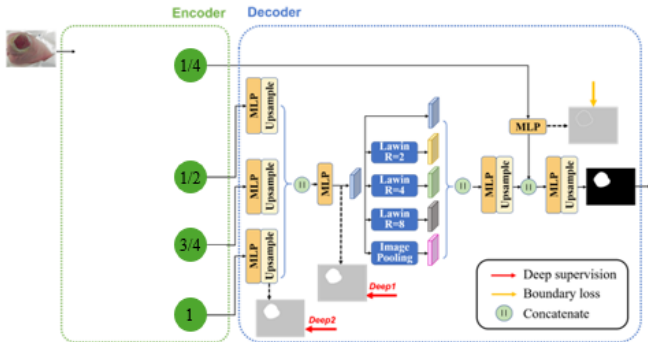
3.3 Initialisation & Etape 1 : Pretraining "à la DINO" (SSL) sur B_1

- Encodeurs retenus :
 - HardNet-DFUS (3 M) poids aléatoire
 - ViTs14_reg (21 M) poids LVD-142M
 - ViTl14_reg (307 M) poids LVD-142M
- Fonction de coût et optimisation : cross-entropy



3.3 Etape 2 : Fine-tuning supervisé en segmentation sur B_2

- Décodeur retenu : Lawin Transformer



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

- G vérité terrain ; G_B contour de vérité terrain
- P prédiction finale ; P_B contour prédiction ; P_i prédiction bloc i

3.4 Expériences : résultats

Encodeur	SSL(B ₁)	Optimisation (B ₂)	Train=70	Train=140	Train=280
HardNet-MSEG _{rdm}	✓	❄	0.724±0.032	0.737±0.011	0.755±0.017
		🔥	0.756±0.034	0.784±0.009	0.798±0.013
	✗	🔥	0.694±0.038	0.736±0.033	0.771±0.023
ViTs14_reg	✓	❄	0.594±0.055	0.644±0.035	0.665±0.034
		🔥	0.674±0.017	0.709±0.023	0.720±0.025
	✗	❄	0.571±0.043	0.646±0.028	0.653±0.020
		🔥	0.685±0.025	0.721±0.021	0.729±0.006
ViT14_reg	✗	❄	0.637±0.042	0.639±0.024	0.701±0.032

Pre-training sur LVD-142M : paramètres ↑ = *performances* ↑

3.4 Expériences : résultats

Encodeur	SSL(B ₁)	Optimisation (B ₂)	Train=70	Train=140	Train=280
HardNet-MSEG _{rdm}	✓	❄	0.724±0.032	0.737±0.011	0.755±0.017
		💧	0.756±0.034	0.784±0.009	0.798±0.013
	✗	💧	0.694±0.038	0.736±0.033	0.771±0.023
ViTs14_reg	✓	❄	0.594±0.055	0.644±0.035	0.665±0.034
		💧	0.674±0.017	0.709±0.023	0.720±0.025
	✗	❄	0.571±0.043	0.646±0.028	0.653±0.020
		💧	0.685±0.025	0.721±0.021	0.729±0.006
ViT14_reg	✗	❄	0.637±0.042	0.639±0.024	0.701±0.032

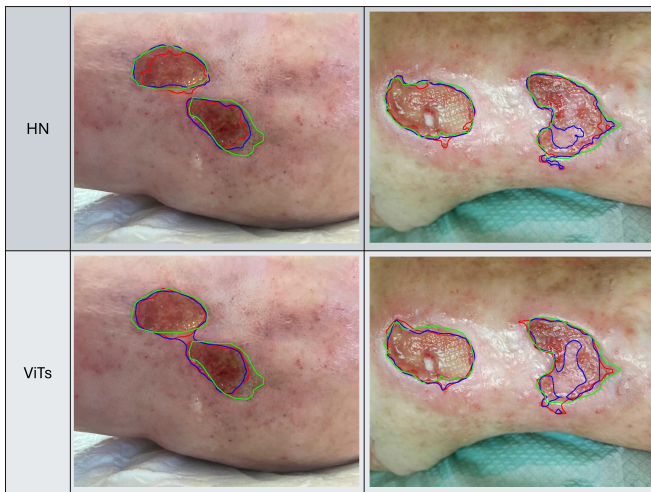
Encodeur léger spécialisé > encodeurs larges initialisés sur LVD-142M

3.4 Expériences : résultats

Encodeur	SSL(B_1)	Optimisation (B_2)	Train=70	Train=140	Train=280
HardNet-MSEG _{rdm}	✓	❄	0.724±0.032	0.737±0.011	0.755±0.017
		💧	0.756±0.034	0.784±0.009	0.798±0.013
	✗	💧	0.694±0.038	0.736±0.033	0.771±0.023
ViTs14_reg	✓	❄	0.594±0.055	0.644±0.035	0.665±0.034
		💧	0.674±0.017	0.709±0.023	0.720±0.025
	✗	❄	0.571±0.043	0.646±0.028	0.653±0.020
		💧	0.685±0.025	0.721±0.021	0.729±0.006
ViT114_reg	✗	❄	0.637±0.042	0.639±0.024	0.701±0.032

Pre-training SSL sur B_1 bonnes performances avec peu d'annotations

3.5 Expériences : Visualisation des résultats



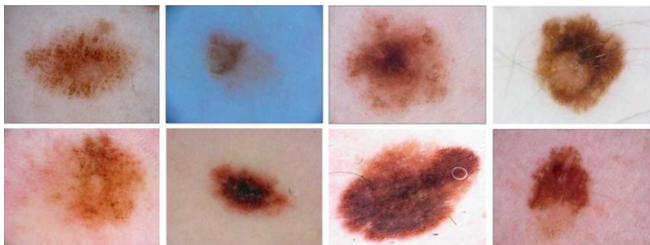
Visualisation pour Train = 280 images avec 🔥

vert : vérité terrain ; rouge avec SSL B_1 ; bleu sans SSL B_1

4. Conclusion

Dans le contexte d'une petite base de données annotées :

- encodeur spécialisé léger > Encodeurs génériques larges
- Pre-training SSL "à la DINO" > approche supervisée
- Perspectives : extension de B_1 avec d'autres bases :



⁶<https://www.isic-archive.com/>

Analyse de l'encodeur pour la segmentation d'une base de données hétérogène de photographies de plaies chroniques avec peu d'annotations

Guillaume PICAUD^{1,3}, Marc CHAUMONT^{1,2}, Gérard SUBSOL¹, Luc TEOT³
{guillaume.picaud, marc.chaumont, gerard.subsol}@lirmm.fr , l-teot@chu-montpellier.fr

¹LIRMM, équipe ICAR, Univ. Montpellier, CNRS, France

²Univ. Nîmes Place Gabriel Péri, France

³Cicat-Occitanie, Montpellier, France

