

Benchmarking de méthodes de segmentation vasculaires

Oscar Morand

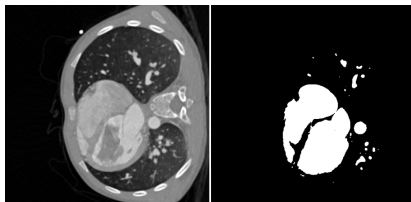
Encadré par: Élodie Puybareau

EPITA

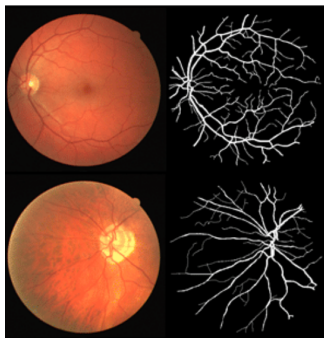
LRE - Équipe IMAGE



Introduction

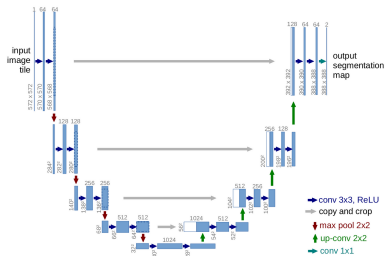


Exemple de segmentation binaire (à gauche) [1], et de segmentation multi-classes (à droite) [2]



Images du jeu de données DRIVE [3]: images originales (à gauche) et segmentations binaires annotées correspondantes (à droite).

Benchmarking, état de l'art



Architecture du modèle U-Net [4]

FCN

Fully Convolutional Network, ex:
U-Net [4], V-Net [5], nnU-Net [6]...

Autres modèles de deep learning

Generative adversarial networks (GAN) [7], Vision Transformers (ViTs) [8]

Autre

- Benchmark: "Blood vessel segmentation algorithms, review of methods, datasets and evaluation metrics" [9]
- Compétition: PARSE Challenge [10]

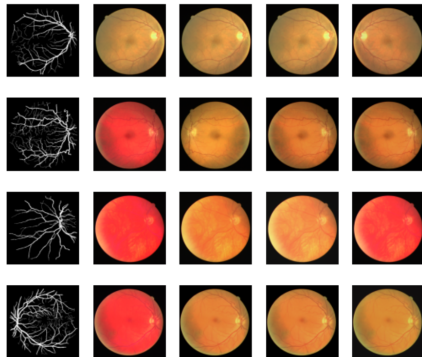
Benchmarking, Travail réalisé (Pré-traitement)

Prétraitement et augmentation de données

- Normalisation des données
- Redimensionner les images
- Data augmentation:
 - ▶ Flip aléatoire
 - ▶ Variations des valeurs HSV

DRIVE

original dataset size: 20 * augmentation factor: 4 = 80



Benchmarking, Travail réalisé (métriques)

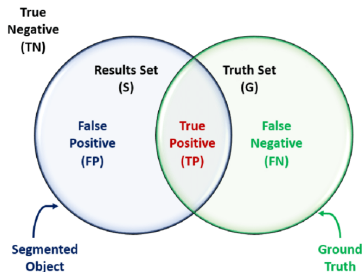


Diagramme [11] illustrant le calcul des valeurs de la matrice de confusion pour la segmentation.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision / Positive predictive value (PPV)} = \frac{TP}{TP + FP}$$

$$\text{Recall / Sensitivity} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$F1 = \frac{2 \times TP}{2 \times TP + FP + FN} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{False Positive Rate / Fallout} = 1 - \text{Specificity} = \frac{FP}{TN + FP}$$

$$\text{Negative Predictive Value (NPV)} = \frac{TN}{TN + FN}$$

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Benchmarking, Travail réalisé (entrainement)

```
# Training parameters
LEARNING_RATE = 1e-4
NUM_EPOCHS = 100
METRICS = [accuracy, sensitivity, specificity, false_positive_rate, positive_predictive_value, negative_predictive_value, MCC, DSC]
BATCH_SIZE = 2
SHUFFLE = False
K = 5

# Model parameters
LOAD_MODEL = True
IMAGE_HEIGHT = 512
IMAGE_WIDTH = 512

# Datasets parameters
datasets_augmentation_factors = {
    "drive": 4,
}

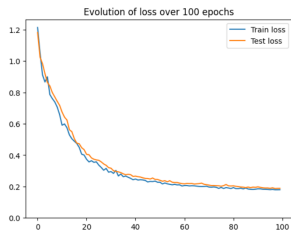
# Other parameters
NUM_WORKERS = 0
PIN_MEMORY = True
```

Paramètres de l'entrainement

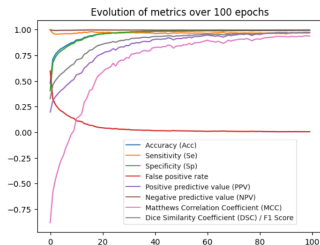
$$BCE = -\frac{1}{N} \sum_{i=0}^N y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i)$$

Fonction de coût (BCE Loss [12] avec pondération) utilisée pour l'entrainement du modèle.

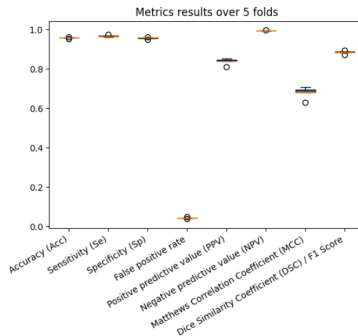
Benchmarking, premiers résultats (1/2)



Évolution de la fonction de coût (Weighted BCE) sur un fold, pour 100 epochs.

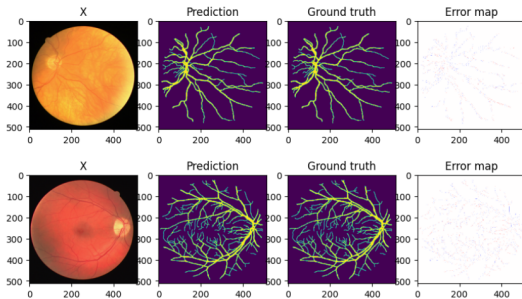


Évolution des métriques sur un fold, pour 100 epochs.

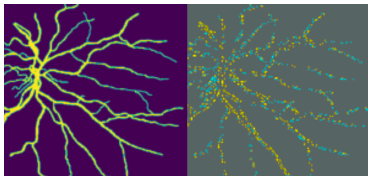


Distribution statistique des résultats des métriques sur 5 folds.

Benchmarking, premiers résultats (2/2)

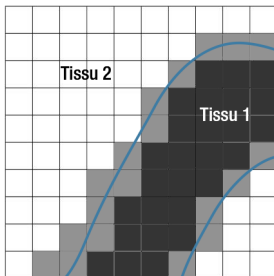


Résultats d'inférence après l'entraînement du modèle sur 100 epochs: image d'origine (à gauche), suivie de la prédiction du modèle, du masque de vérité terrain et de la carte d'erreur (à droite).



Zoom sur la carte d'erreur d'une prédiction, où le jaune indique les faux positifs et le bleu indique les faux négatifs (avec amélioration du contraste pour mettre en évidence les erreurs).

Nouvelle métrique, problématique



Exemple de l'effet de volume partiel. Les pixels gris correspondent à des artéfacts de volume partiel et ont une valeur intermédiaire entre les tissus 1 et 2.

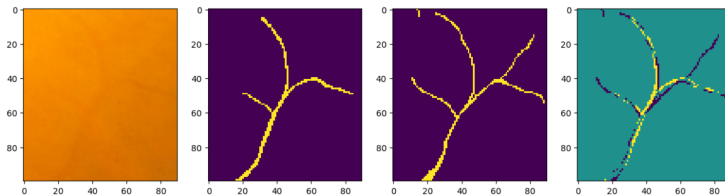
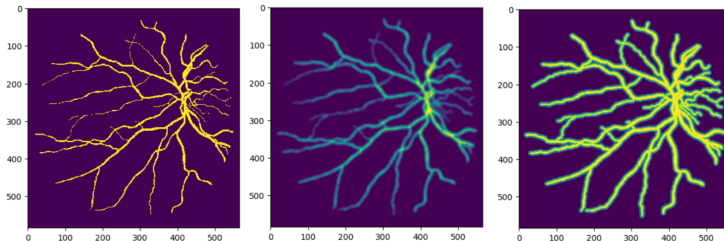
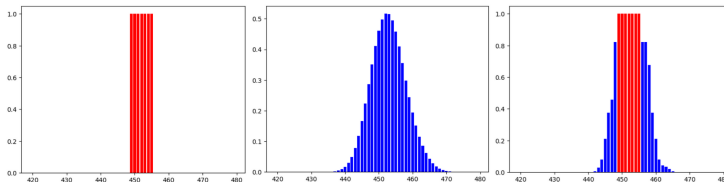


Image du vaisseau sanguin (gauche), vérité terrain faite par la personne numéro 1, vérité terrain numéro 2, carte des différences entre les deux vérités terrain (droite).

Nouvelle métrique, première approche (1/2)



Masque de vérité terrain binaire (à gauche), vérité terrain floutée avec une fonction gaussienne (au centre), vérité terrain lissée avec une fonction de distance et une décroissance exponentielle (à droite).



Vue en coupe transversale du même segment de vaisseau sanguin. Masque de vérité terrain binaire (à gauche), vérité terrain floutée avec une fonction gaussienne (au centre), vérité terrain lissée avec une fonction de distance et une décroissance exponentielle (à droite).

Nouvelle métrique, première approche (2/2)

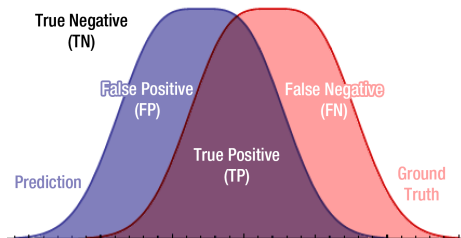
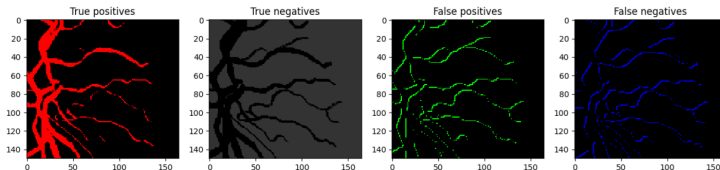
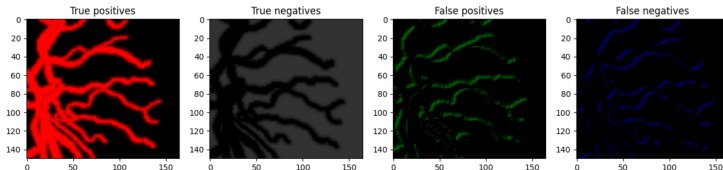


Schéma du nouveau calcul des valeurs de la matrice de confusion, avec vérité terrain et prédiction lissées.

Première approche, analyse des résultats (1/4)



Résultat entre la vérité terrain originale (binaire) et la prédiction décalée. Vrais positifs en rouge (à gauche), vrais négatifs en gris, faux positifs en vert, et faux négatifs en bleu (à droite).



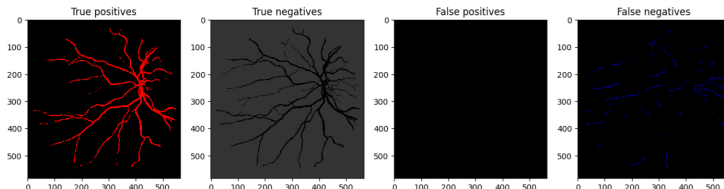
Résultat entre notre vérité terrain lissée et la prédiction décalée.

Première approche, analyse des résultats (2/4)

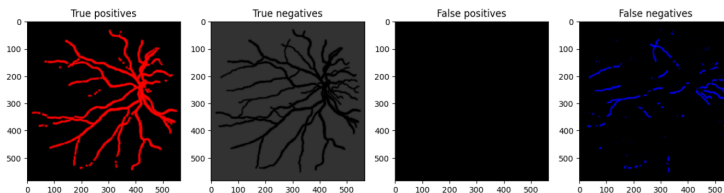
metric_name	binary	smooth	diff
Accuracy (Acc)	0.924485	0.939065	0.014580
Sensitivity (Se)	0.785714	0.909811	0.124097
Specificity (Sp)	0.954922	0.954868	-0.000054
False positive rate	0.045078	0.045132	0.000054
Positive predictive value (PPV)	0.792658	0.915893	0.123235
Negative predictive value (NPV)	0.953090	0.951455	-0.001635
Matthews Correlation Coefficient (MCC)	0.555944	0.790188	0.234245
Dice Similarity Coefficient (DSC) / F1 Score	0.789171	0.912842	0.123671

Tableau comparatif des métriques, appliquant chaque métrique aux deux méthodes (masques binaires et masques lissés) après décalage de la prédiction, et affichant la différence entre eux.

Première approche, analyse des résultats (3/4)



Résultat entre la vérité terrain originale (binaire) et la prédiction après ouverture morphologique. Vrais positifs en rouge (à gauche), vrais négatifs en gris, faux positifs en vert, et faux négatifs en bleu (à droite).



Résultat entre notre vérité terrain lissée et la prédiction après ouverture morphologique.

Première approche, analyse des résultats (4/4)

metric_name	binary	smooth	diff
Accuracy (Acc)	0.990429	0.972226	-0.018203
Sensitivity (Se)	0.854624	0.799359	-0.055265
Specificity (Sp)	1.000000	1.000000	0.000000
False positive rate	0.000000	0.000000	0.000000
Positive predictive value (PPV)	1.000000	1.000000	0.000000
Negative predictive value (NPV)	0.989859	0.968770	-0.021088
Matthews Correlation Coefficient (MCC)	0.919759	0.879998	-0.039762
Dice Similarity Coefficient (DSC) / F1 Score	0.921614	0.888493	-0.033121

Tableau comparatif des métriques, appliquant chaque métrique aux deux méthodes (masques binaires et masques lissés) après ouverture morphologique sur la prédiction, et affichant la différence entre eux.

clDice (1/2)

$$\text{clDice}(V_P, V_L) = 2 \times \frac{\text{Tp}(\text{prec}(S_P, V_L) \times \text{Tsens}(S_L, V_P))}{\text{Tp}(\text{prec}(S_P, V_L) + \text{Tsens}(S_L, V_P))}$$

avec

$$\text{Tp}(\text{prec}(S_P, V_L)) = \frac{|S_P \cap V_L|}{|S_P|} \quad ; \quad \text{Tsens}(S_L, V_P) = \frac{|S_L \cap V_P|}{|S_L|}$$

VL étant le masque de vérité terrain, VP la prédiction de segmentation, SL le squelette de la vérité terrain, SP celui de la prédiction.

clDice (2/2)

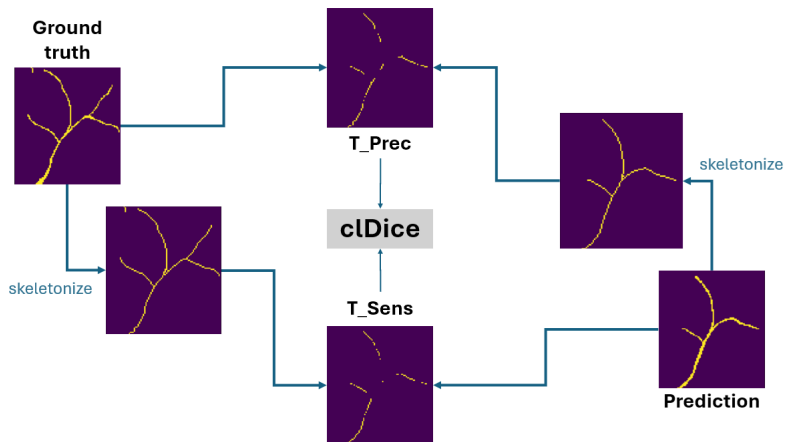


Schéma de l'algorithme du clDice [13].

Smooth cLDice

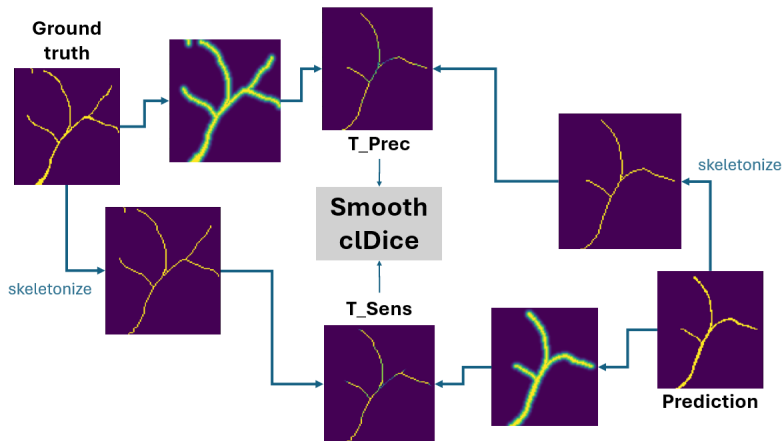
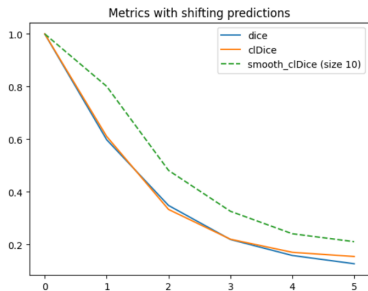
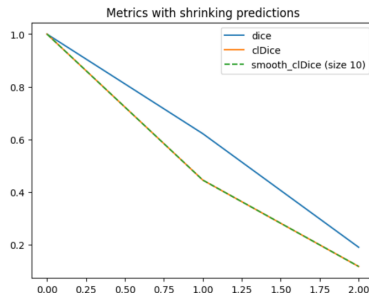


Schéma de l'algorithme du "smooth cLDice", on rajoute une étape de lissage avant le calcul de T_{Prec} et T_{Sens} .

Smooth cIDice, analyse des résultats



Évolution des métriques (dice, cIDice, smooth_cIDice) par rapport à la force du décalage appliqué à la prédiction (en pixels).

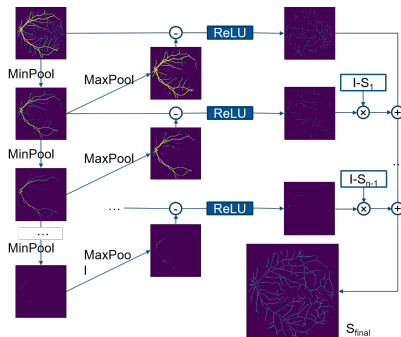


Évolution des métriques (dice, cIDice, smooth_cIDice) par rapport au rayon du kernel de l'ouverture morphologique appliqué à la prédiction (en pixels).

Travail futur

Tâches à réaliser (dans l'ordre)

- 1 Intégrer cLDice et smooth_cLDice aux métriques du benchmark
- 2 Développer une fonction de coût à partir du cLDice
- 3 Réaliser un entraînement avec cette fonction de coût
- 4 Continuer et finir le benchmarking



Algorithme de soft-skeleton [13], pour calculer le squelette morphologique d'un objet de façon différentiable.

Conclusion

Travail réalisé

- Base d'un programme de benchmarking prêt à accueillir plus de modèles / jeux de données
- Premier entraînement concluant d'un modèle
- Nouvelle(s) métrique(s) prometteuse(s)

Bibliographie I

- [1] O. J. Barrowclough, G. Muntingh, V. Nainamalai, and I. Stangeby, "Binary segmentation of medical images using implicit spline representations and deep learning," *Computer Aided Geometric Design*, vol. 85, p. 101972, Feb. 2021. [Online]. Available: <http://dx.doi.org/10.1016/j.cagd.2021.101972>
- [2] M. Hamza, "Multi-class semantic segmentation with u-net (pytorch)," *Medium*, 2023, accessed: 2024-06-15. [Online]. Available: <https://medium.com/@mhamdaan/multi-class-semantic-segmentation-with-u-net-pytorch-ee81a66bba89>
- [3] J. Staal, M. Abramoff, M. Niemeijer, M. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Transactions on Medical Imaging*, vol. 23, no. 4, pp. 501–509, 2004.
- [4] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

Bibliographie II

- [5] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-net: Fully convolutional neural networks for volumetric medical image segmentation,” in *2016 fourth international conference on 3D vision (3DV)*. Ieee, 2016, pp. 565–571.
- [6] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, “nnu-net: a self-configuring method for deep learning-based biomedical image segmentation,” *Nature methods*, vol. 18, no. 2, pp. 203–211, 2021.
- [7] P. Luc, C. Couprie, S. Chintala, and J. Verbeek, “Semantic segmentation using adversarial networks,” *arXiv preprint arXiv:1611.08408*, 2016.
- [8] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, “An image is worth 16x16 words: Transformers for image recognition at scale,” *arXiv preprint arXiv:2010.11929*, 2020.
- [9] S. Moccia, E. De Momi, S. El Hadji, and L. S. Mattos, “Blood vessel segmentation algorithms—review of methods, datasets and evaluation metrics,” *Computer methods and programs in biomedicine*, vol. 158, pp. 71–91, 2018.

Bibliographie III

- [10] G. Luo, K. Wang, J. Liu, S. Li, X. Liang, X. Li, S. Gan, W. Wang, S. Dong, W. Wang *et al.*, “Efficient automatic segmentation for multi-level pulmonary arteries: The parse challenge,” *arXiv preprint arXiv:2304.03708*, 2023.
- [11] M. Mostapha, “A novel diffusion tensor imaging-based computer-aided diagnostic system for early diagnosis of autism.” 2014.
- [12] “Understanding binary cross-entropy / log loss: A visual explanation.” [Online]. Available: <https://towardsdatascience.com/understanding-binary-cross-entropy-log-loss-a-visual-explanation-a3ac6025181a>
- [13] S. Shit, J. C. Paetzold, A. Sekuboyina, I. Ezhov, A. Unger, A. Zhyhka, J. P. Pluim, U. Bauer, and B. H. Menze, “cldice-a novel topology-preserving loss function for tubular structure segmentation,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 16 560–16 569.

Bibliographie IV

- [14] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, “Unet++: A nested u-net architecture for medical image segmentation,” in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*. Springer, 2018, pp. 3–11.