

Challenge : Réseau de neurones pour la segmentation

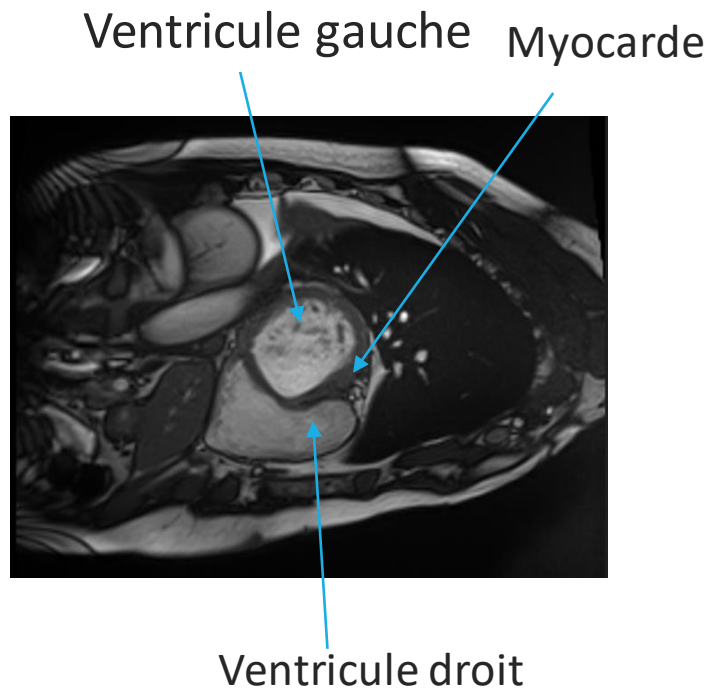
MTI881 – HIVER 2023

- EMMA OLLIVIER
- BAPTISTE VIERA
- STEVE LEVESQUE
- PAUL WALIGORA



Jeu de données

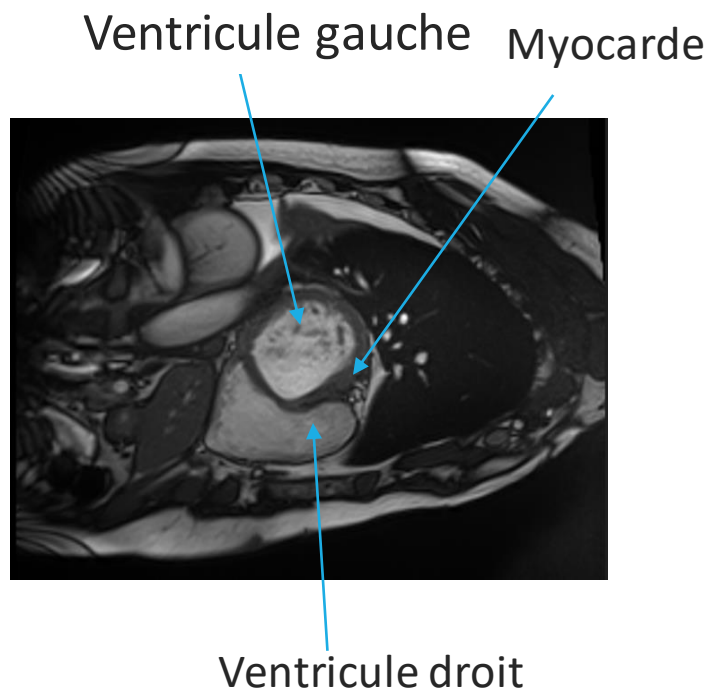
3 classes à segmenter :



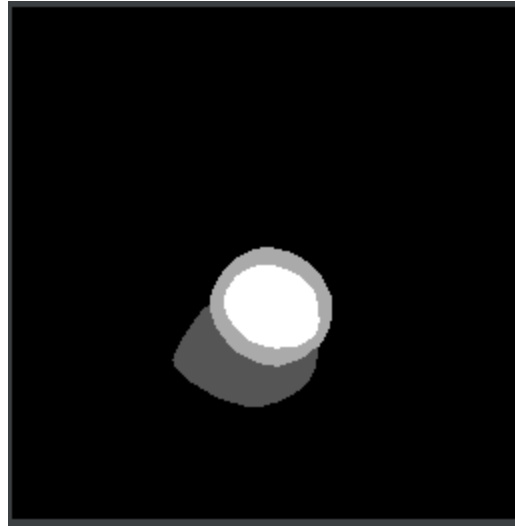
- Image de 256*256 pixels
- 204 images étiquetées pour l'entraînement
- 1004 images non étiquetées pour l'entraînement
- 74 images étiquetées pour la validation

Jeu de données

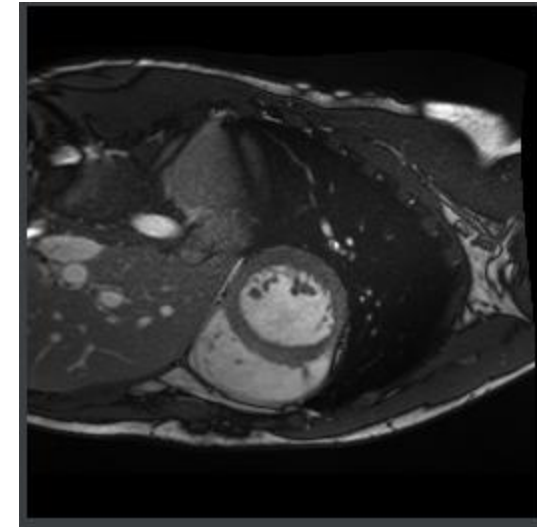
3 classes à segmenter :



GT



Img



patient078_01_3

Architecture : Transformers

Article :

- "Attention is All You Need" – 2017
- "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale" – 2019

Auto-attention :

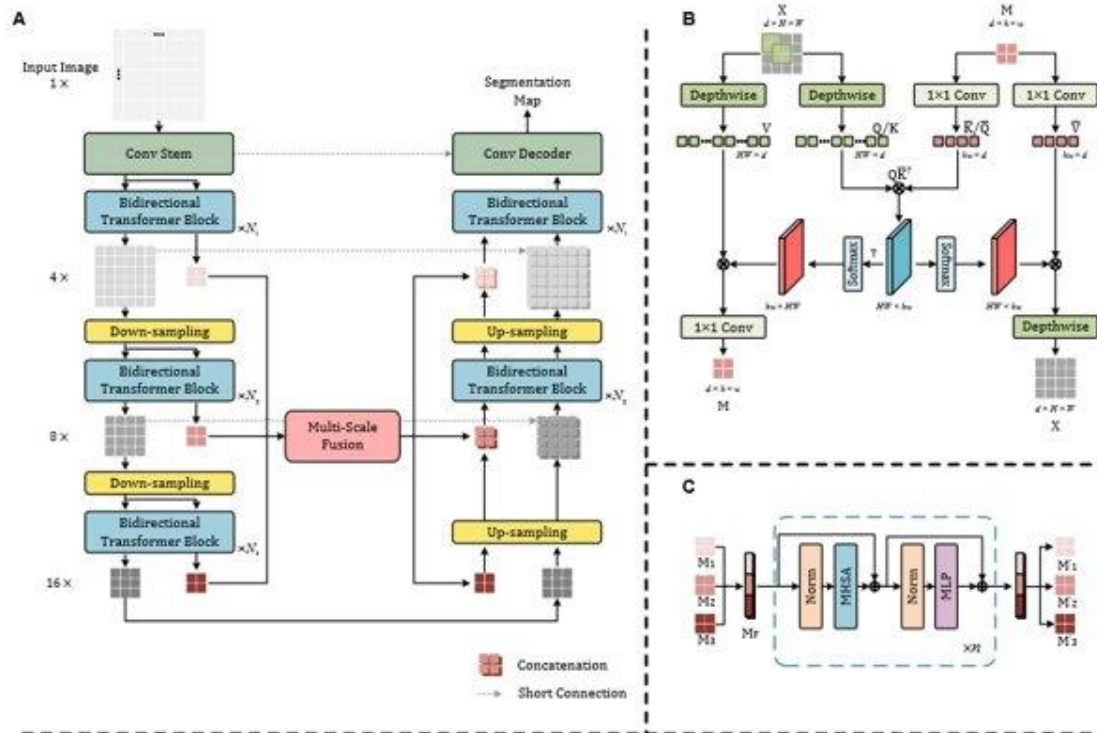
- Calcul des poids d'attention en mesurant la similarité des pixels de l'image deux à deux avec prise en compte des relations globales entre les pixels
- Complexité quadratique par rapport à la longueur de la sequence d'entrée

Limites :

- Performances moindres sur des petits jeux de données
- Equilibrage difficile entre le calcul et les details fins pour les entrées à hautes resolutions
- Moyenne capacité de généralisation sur un large éventail de tâches médicales

Architecture : MedFormer

Architecture hybride



Zhang, Y., Tian, Y., Kong, X., Liu, B., & Fu, Y. "A Data-scalable Transformer for Medical Image Segmentation: Architecture, Model Efficiency, and Benchmark." *IEEE Transactions on Medical Imaging*, vol. 40, no. 3, pp. 1223-1232, 2021.

Architecture : MedFormer

Architecture hybride :

- **Convolution séparable en profondeur dans le réseau de projection :**
 - Capture les textures locales (caractéristiques locales)
 - Favorable à l'analyse de données à petite échelle avec le biais inductif
 - Couches peu profondes
- **Feed-Forward dans les blocs transformers :**
 - Modélisation des structures anatomiques complexes de manière hiérarchique
 - Amélioration des relations locales et à longue portée
 - Couches intermédiaires et profondes

Architecture : MedFormer

Attention multi-têtes bidirectionnelle (B- MHA)

- Elimination des tokens redondants
- Conservation des éléments les plus informatifs
- Réduction de la complexité quadratique à un niveau linéaire
- Robuste face au changement de domaine
- Compromis favorable entre les performances, le nombre de paramètres et le coût de calcul

Architecture : MedFormer

Tableau 1 : Résultats DSC sur le grand ensemble de données d'IRM cardiaque avec différents ratios de données d'entraînement

Arch.	Modèles	5%	10%	40%	70%	100%
CNN	UNet ⁸	86.51	87.17	87.96	88.37	88.59
	Attn UNet ³²	86.74	87.46	88.20	88.38	88.63
	UNet++ ⁹	86.54	87.33	88.21	88.52	88.59
	ResUNet	86.65	87.48	88.31	88.54	88.49
	R50-UNet ³³	86.42	87.48	88.39	88.59	88.72
TFM	TransUNet ²⁵	86.53	87.56	88.33	88.55	88.56
	TransUNet †	86.30	87.22	88.01	88.57	88.55
	SwinUNet ²⁶	75.25	82.19	85.71	86.20	86.83
	SwinUNet†	86.71	86.91	87.61	87.70	88.01
	UTNet ³⁴	86.70	87.50	88.41	88.58	88.69
	MedFormer	87.72	87.99	88.80	88.92	89.05

† indique que le modèle est initialisé avec des poids pré-entraînés sur ImageNet.

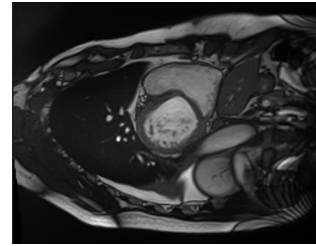
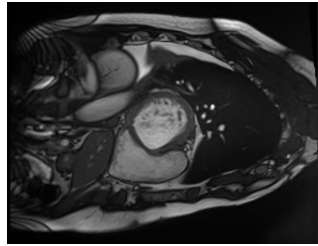
Zhang, Y., Tian, Y., Kong, X., Liu, B., & Fu, Y. "A Data-scalable Transformer for Medical Image Segmentation: Architecture, Model Efficiency, and Benchmark." *IEEE Transactions on Medical Imaging*, vol. 40, no. 3, pp. 1223-1232, 2021.

Limites

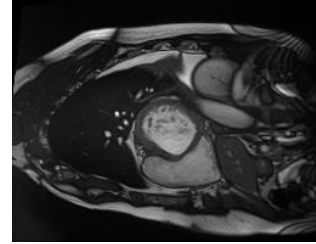
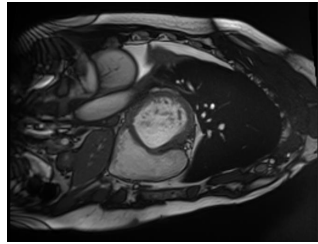
- Medformer, dans l'article, n'a été uniquement analysé dans le cadre d'un apprentissage supervisé

Augmentation des données étiquetées

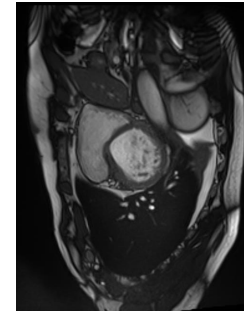
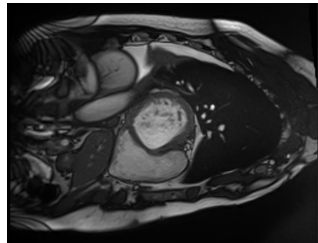
Flip



Mirror



Rotate



Adaptive Moment Estimation Optimizer (Adam)

Définition de l'optimisateur ADAM:

Algorithme d'optimisation couramment utilisé pour entraîner des réseaux de neurones profonds. Combine le Stochastic Gradient Descent (SGD) et le Momentum.

Fonctionnement:

- 1 - L'estimation adaptative du moment du premier ordre (moyenne mobile exponentielle des gradients)
- 2 - L'estimation du moment du deuxième ordre (moyenne mobile exponentielle des carrés des gradients)

Utilisé pour **ajuster** la taille de la mise à jour des **poids** lors de la descente de gradient stochastique

Adaptive Moment Estimation Optimizer (Adam)

Explication de la moyenne mobile (Exemple en analyse de marché)



Adaptive Moment Estimation Optimizer (Adam)

L'algorithme d'Adam:

- => Permet de trouver un minimum global de manière très efficace
- => Populaire pour les réseaux de neurones profonds en raison de sa capacité à ajuster efficacement la taille de pas de mise à jour des poids
- => Permet à l'algorithme de converger plus rapidement et plus efficacement vers un minimum global

Adaptive Moment Estimation (Adam)

$$m_t = \gamma_1 m_{t-1} + (1 - \gamma_1) \frac{\partial \mathcal{L}}{\partial w}$$

Premier moment

$$v_t = \gamma_2 v_{t-1} + (1 - \gamma_2) \frac{\partial \mathcal{L}^2}{\partial w}$$

Second

$$\hat{m}_t = \frac{m_t}{1 - \gamma_1^t} \quad \hat{v}_t = \frac{v_t}{1 - \gamma_2^t}$$

Stabilisation

$$\theta^{t+1} = \theta^t - \frac{\eta}{\sqrt{v^t + \epsilon}} m^t$$

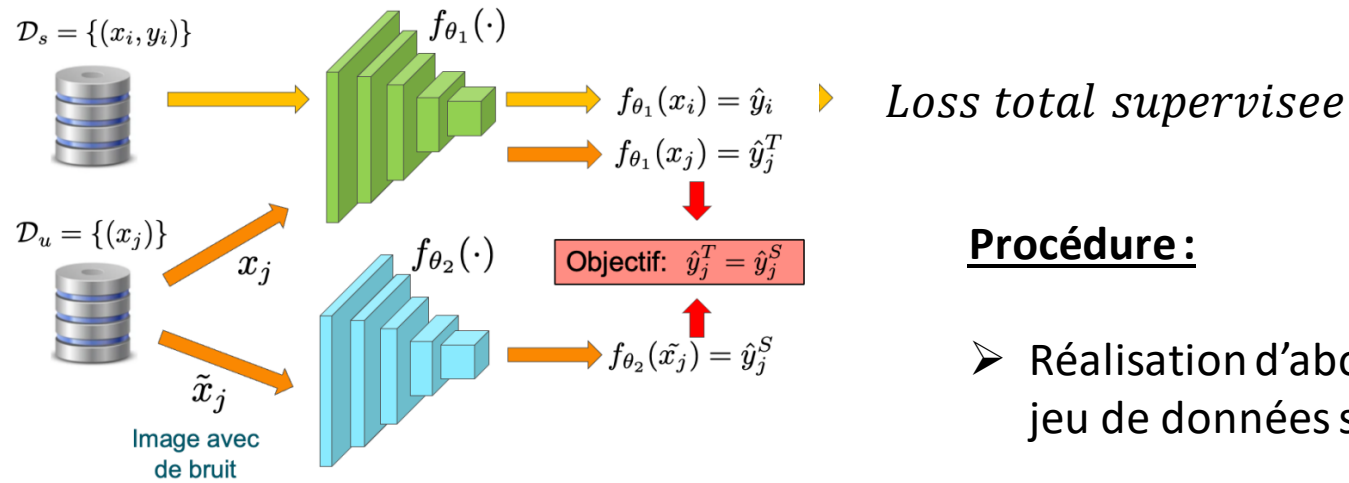
Résultat (itération t)

Fonction de coût

$$Loss\ Total\ Supervisee = \frac{3}{4} \times Dice\ loss + \frac{1}{4} \times Cross\ entropy\ loss$$

- Choix des hyperparamètres de régularisation pour donner trois fois plus d'importance au dice qu'à l'entropie croisée
- Utilisation d'une fonction dice multi classe permettant de faire abstraction du fond [1]

Utilisation des données non supervisées

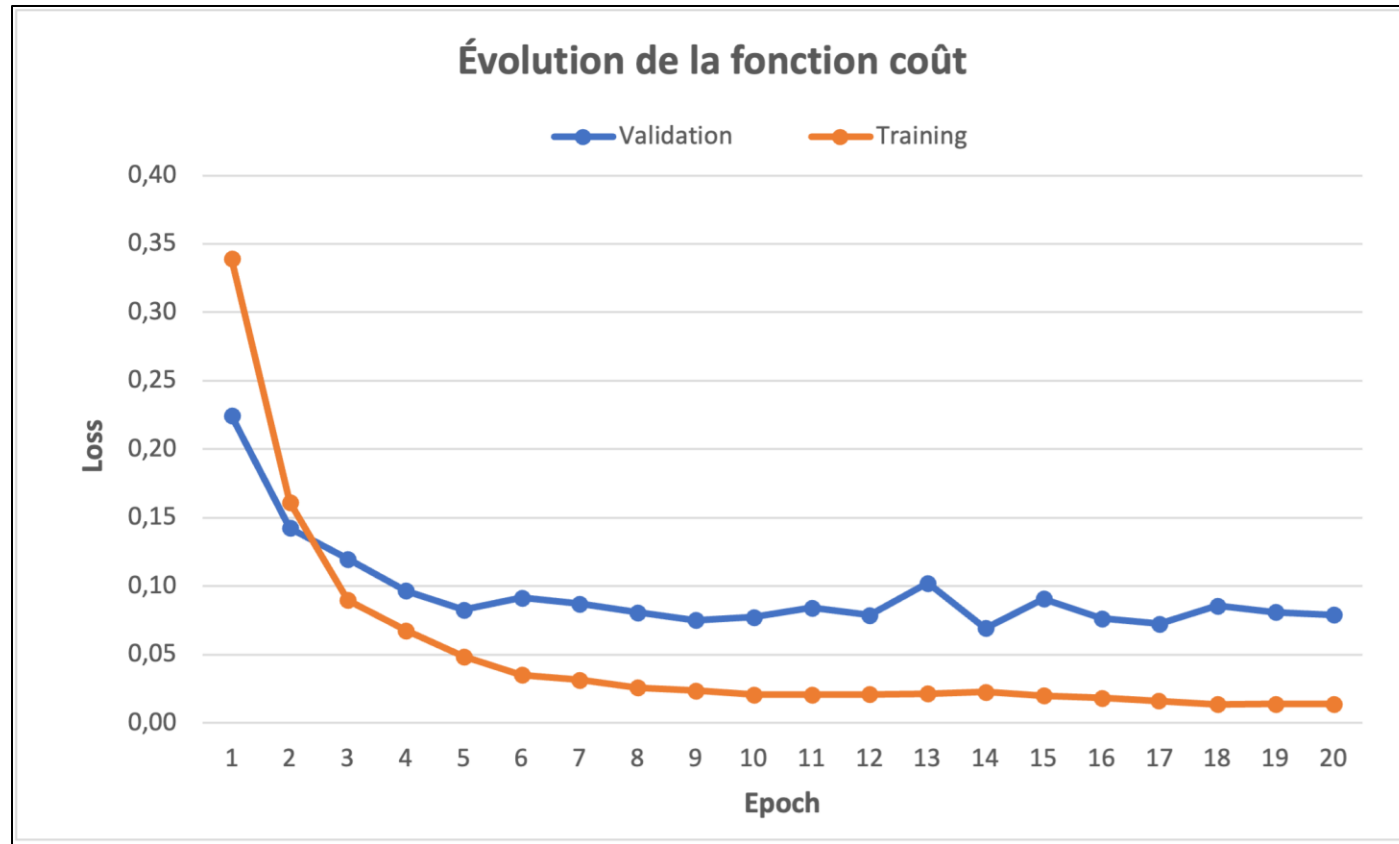


Procédure :

- Réalisation d'abord de plusieurs epochs avec uniquement le jeu de données supervisées
- Utilisation ensuite des deux jeux de données simultanément

$$\text{Loss de consistance} = \left\| \hat{y}_j^T - \hat{y}_j^S \right\|^2$$

$$\text{Loss Total} = \frac{4}{7} \times \text{Loss total supervisee} + \frac{3}{7} \times \text{Loss de consistance}$$



Epoch	validation	Training
1	0,2244	0,3391
2	0,1425	0,1612
3	0,1197	0,0898
4	0,0966	0,0676
5	0,0828	0,0486
6	0,0914	0,0351
7	0,0871	0,0314
8	0,0806	0,0257
9	0,0752	0,0235
10	0,0774	0,0206
11	0,0841	0,0206
12	0,0787	0,0209
13	0,1023	0,0214
14	0,0693	0,0226
15	0,0907	0,0199
16	0,0762	0,0182
17	0,0723	0,0159
18	0,0856	0,0135
19	0,0809	0,0140
20	0,0790	0,0140

Exemple d'un entraînement de notre modèle

Learning rate: 0.001

Epoch : 20

Batch size (training) : 15

Batch size (validation): 1

Résultats

Accuracy Total : 98%

Dice : 93%

	Précision	Recall	Accuracy
Classe 1	93%	95%	99%
Classe 2	59%	93%	99%
Classe 3	41%	97%	99%

Merci pour votre « attention » !

Bibliographie

- HILAB-GIT. SSL4MIS: Self-supervised learning for medical image segmentation. Github Repository, 2021. <https://github.com/HILAB-GIT/SSL4MIS>.
- Qubvel. Segmentation Models Pytorch. Github Repository, 2021. https://github.com/qubvel/segmentation_models.pytorch
- Dolz, J (2023), cours de MTI881, ETS Montreal
- Zhang, Y., Tian, Y., Kong, X., Liu, B., & Fu, Y. "A Data-scalable Transformer for Medical Image Segmentation: Architecture, Model Efficiency, and Benchmark." IEEE Transactions on Medical Imaging, vol. 40, no. 3, pp. 1223-1232, 2021.
- Gao, Y. CBIM-Medical-Image-Segmentation: Convolutional Block Attention Module for Medical Image Segmentation. Github Repository, 2021. <https://github.com/yhygao/CBIM-Medical-Image-Segmentation>.