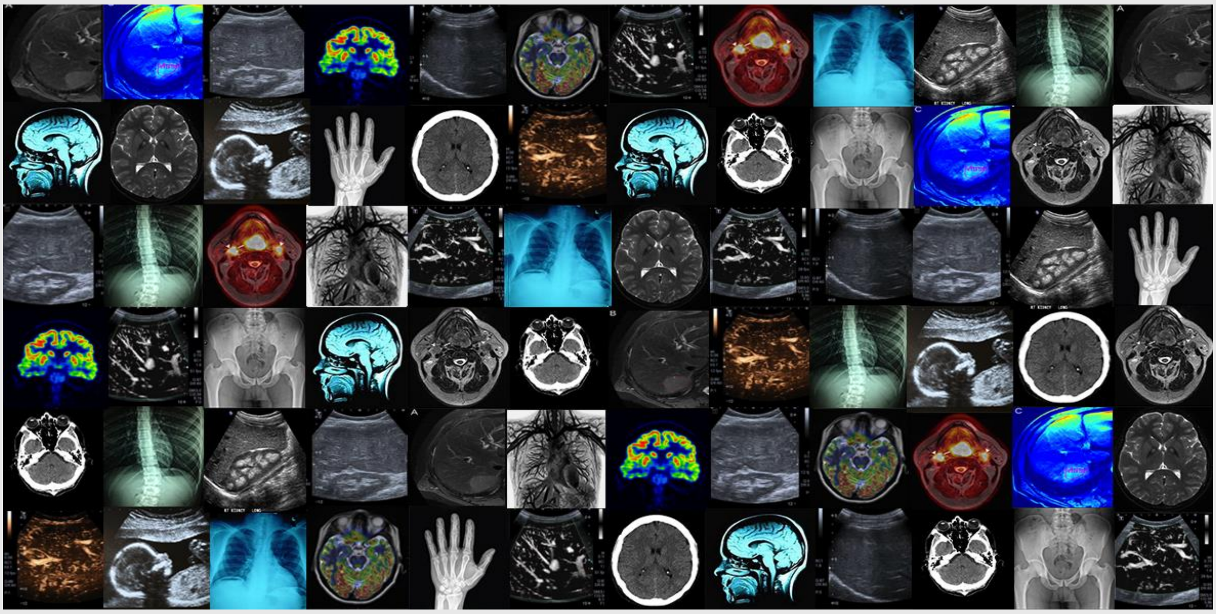
PROJECT 10

A NOVEL APPROACH TO MEDICAL IMAGE SEGMENTATION USING HYBRID ARCHITECTURES



Project sponsor

- Dr Feryal **WINDAL**

Client

- Dr Fares **BOUGOURZI**

Written by:

- Adrien **TCHUEM TCHUENTE** – IA
- Abir **TALHA** – Big Data
- Bassit **NKOUONJOM** – Big Data
- Chaimae **MAAROUF** – Big Data
- Sacha **BINANZER** – IA

From 19th September 2023 to 29th May 2024

M2 Students - ISEN 2023 -2024

SUMMARY

INTRODUCTION

In the field of health, medical imaging technology represents a major advance in the identification of diseases dangerous to humans. Technologies such as magnetic resonance imaging[1] (MRI) or computed tomography[2] (CT scans) have proven to be crucial in the daily work of doctors. Despite their undeniable contribution, the interpretation of these images is subject to human error due to the complexity and variability of the anomalies that can arise from the human body.

Since the early 2010s, artificial intelligence (AI) has emerged as a major ally in solving this problem · Researchers have developed numerous segmentation and classification tools and techniques to better identify anomalies in medical images reference, such as the Synapse Dataset.

To date, there exist effective models such as TransUNet[3] and UNet[4]. This ambitious project focuses on organ segmentation using convolutional neural networks (CNN) integrated in a hybrid architecture. Our work focuses on Hybrid CNN-Transformer architectures which combines the advantages of Transformer and convolutional neural networks to improve the accuracy and efficiency of medical image segmentation.

The ultimate goal of the project is to provide doctors with AI-based tools to better segment organs from medical images such as MRI and CT scans, making disease identification easier. Throughout the project, we conducted rigorous scientific research, systematically testing and optimizing each model to identify and refine the best performing models.

This approach has allowed doctors to better segment parts of diseased organs in the field of medical imaging, offering broad perspectives in scientific research. This new AI-assisted medical segmentation technology such as the UNet Model represents a step forward towards a considerable improvement in the medical diagnosis of diseases such as cancers and the detection of tumors.

---

[1] *https://www.nibib.nih.gov/science-education/science-topics/magnetic-resonance-imaging-mri*

[2] *https://en.wikipedia.org/wiki/CT_scan*

[3] *https://arxiv.org/abs/2102.04306*

[4] https://arxiv.org/abs/1505.04597

PART 1: METHODOLOGY

To carry out this project, we followed an approach structured in several key stages, adopting project management based on Agile methodology. Our project, which took place over a period of 17 weeks, was organized into sprints of 3 weeks each. At the end of each sprint, we had to deliver to our client a part of the project previously defined during our weekly meetings.

APPROACH

1. **Planning and Distribution of Tasks**:
    - Identification of Objectives: We started by defining the objectives of the project, in consultation with our client.
    - Team Meetings: The team met to discuss the different strengths and weaknesses of each member and thus distribute tasks in the most optimal way.
    - Distribution of Tasks: The tasks have been distributed according to the skills of each member of the team.
2. **State of the art and bibliography**: It was a question for us to carry out intensive and rigorous research on the subject. It helped us to identify, understand and compare the documents and techniques relevant to the project.
3. **Development and Integration :**
    - Development Environment: We set up a development environment including tools such as Python, Visual Studio Code and GitHub for code management.
    - Development: Development was done cyclically, with deliveries at the end of each sprint.
4. **Validation and Customer feedback** : At the end of each sprint, we presented the results to the client for validation and to collect feedback.
5. **Optimization and Adjustments** :
    - Model Optimization: Models have been continuously optimized to improve segmentation accuracy and efficiency.

- Adjustments Based on Feedback: Customer feedback and testing led to regular adjustments to the project.

6. **Flexibility and Adaptability:** The project, being research-oriented, was subject to several modifications over the weeks. This flexibility has been essential in adapting to new discoveries and changing requirements. The Agile approach allowed us to respond quickly to challenges and opportunities, ensuring we remained aligned with the client's objectives while integrating the latest technological advancements.

By adopting an Agile methodology and following a structured approach, we were able to carry out this project effectively and efficiently. The use of 3-week sprints allowed fine management of time and resources, while ensuring constant communication with the client. This approach has allowed us to make significant advances in the segmentation of medical images, paving the way for innovative and promising clinical applications.

RESSOURCES

TECH STACK

The technical aspect of our project was mainly carried out in Python. *Python is an interpreted, multi-paradigm, and cross-platform programming language. It promotes structured, functional and object-oriented imperative programming.*[5]

After installing Python, we created a virtual environment[6] to manage and store all the modules and dependencies needed for the project. Microsoft Visual Studio Code has been our primary code editor, providing robust features for editing and debugging code.

[5] *https://fr.wikipedia.org/wiki/Python_(langage)*

[6] *https://www.geeksforgeeks.org/python-virtual-environment/*

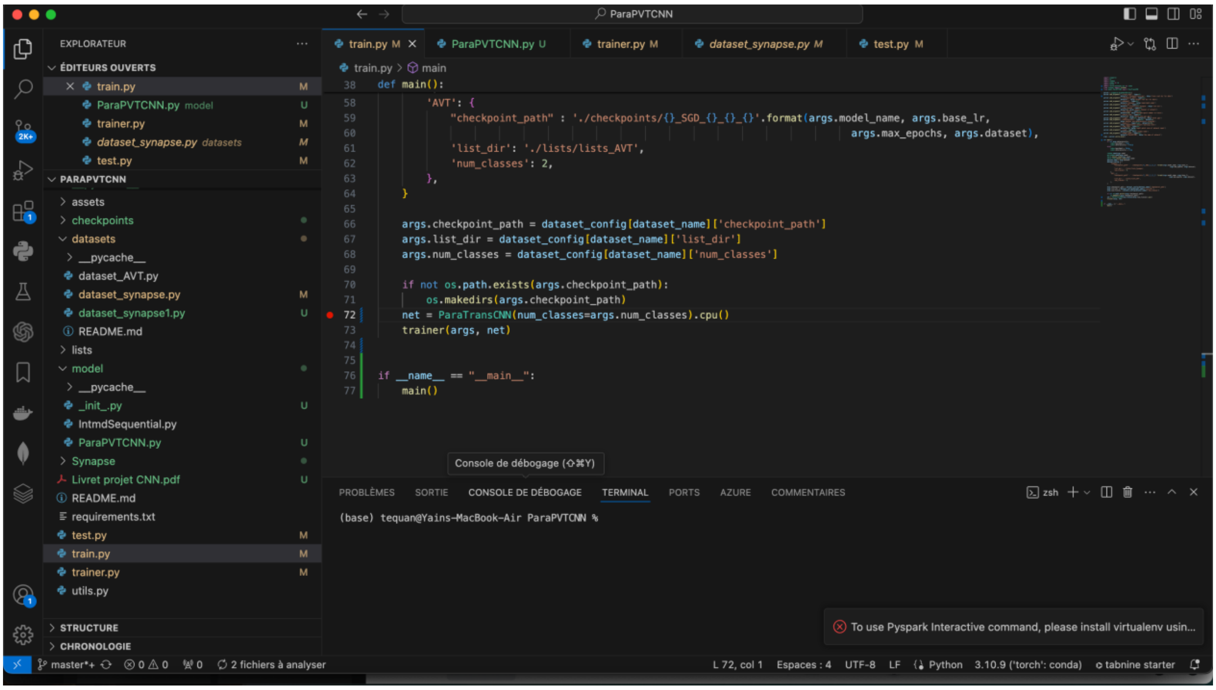


Fig 1: Visual Studio Code interface

At the same time, we set up a GitHub repository to facilitate collaboration and versioning of source code. GitHub was essential for tracking changes, integrating team contributions, and maintaining a clear history of project developments.

PROJECT MANAGEMENT TOOLS

As mentioned above, we adopted an Agile approach when it came to project management. Here are some of the tools we used :

1. Trello

Trello is a visual project management tool that helps organize project steps into boards, lists, and tasks. We have created four lists for each possible stage of the assigned tasks:

- **'À faire'** : Grouping the tasks to be carried out.
- **'En cours'** : For tasks that were currently in progress
- **'En attente/Bloqué'**: For task that needed more information or encountered obstacles
- **'Terminé'** : For terminated tasks

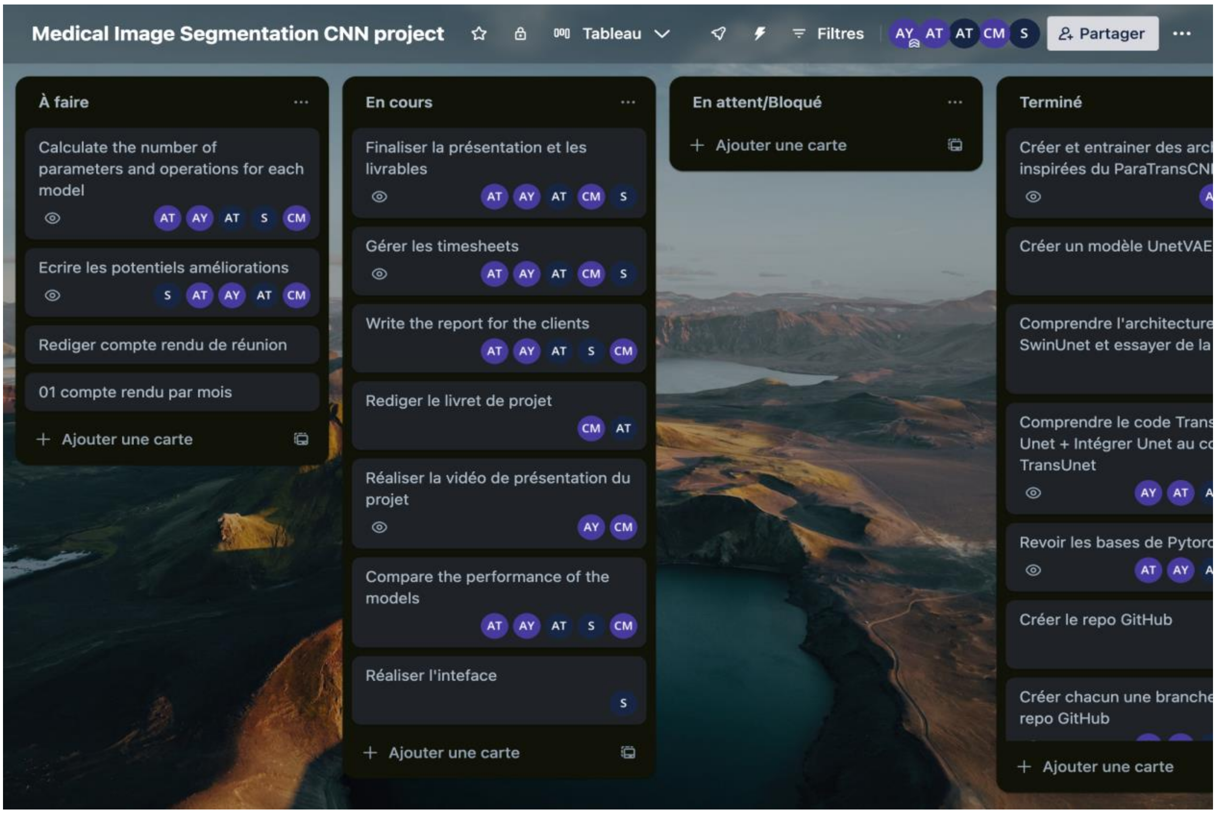


Fig 2: Atlassian Trello project management board

The use of Trello allowed clear and shared visibility into the progress of the project, facilitating collaboration and coordination within the team.

2. Notion

Notion is a powerful, no-code productivity tool used for knowledge management and to make teamwork more user-friendly and efficient. As part of this project, Notion was used to manage timesheets, track hours worked and document progress. This centralization of information made it possible to maintain rigorous organization and increased transparency within the team.

The integration of these tools was crucial to the success of our project, allowing agile, collaborative and structured management.

PART 2: MODELS STUDIED AND RESULTS

During this project, we researched several medical image segmentation models to find the most efficient model in terms of segmentation, here is an overview of the models sought:

UNET MODEL

U-Net is a widely used learning architecture that was first introduced in the article “U-Net: Convolutional Networks for Biomedical Image Segmentation”[7]. The U-Net architecture consists of a network of encoders which is also called a contracting network, a decoder and skip connections.

This encoder network is made up of 4 encoder blocks. Each block contains two convolutional layers with a 3*3 conv and valid padding, followed by a Relu activation function. This goes into a max pooling layer with a kernel size of 2*2. With the max pooling layer, the spatial dimensions of the image are reduced by half, thereby reducing the computational cost of training the model. [2]

Between the encoder and decoder network, there is a bottleneck layer. This is the lowest layer, as we can see in the model above. It consists of 2 convolutional layers followed by Relu. Now, what makes U-Net so effective in image segmentation is the absence of decoder connections and networks. [2]

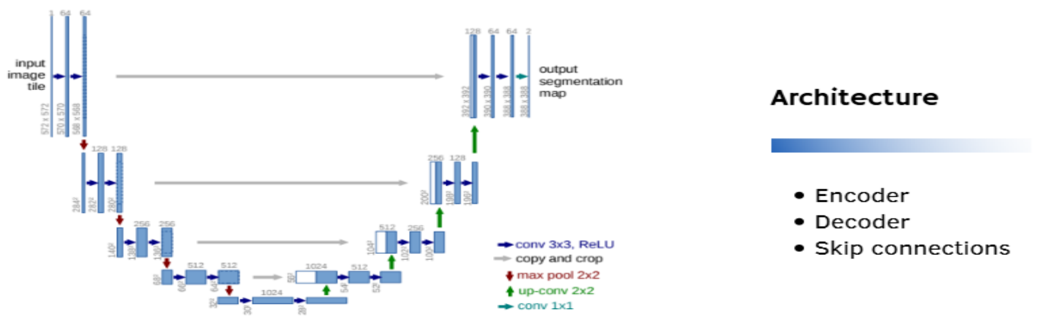


Fig 3 : Unet Architecture

[7] https://arxiv.org/abs/1505.04597

TRANS-UNET

The TransUnet model is like the predecessor Unet model, the TransUNet also includes an encoder and a decoder to encode and decode image information to produce segmentation.

However, unlike traditional U-Nets, the TransUNet instead uses a hybrid CNN-Transformer architecture as an encoder to learn both the high-resolution spatial information from the CNNs and the global contextual information from the Transformers.

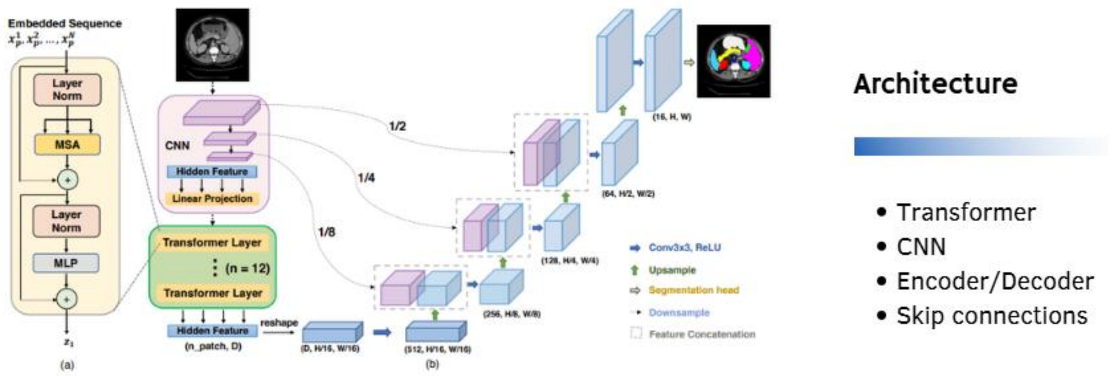


Fig 4 : Trans-Unet Architecture

SWIN-UNET

Swin-UNet is a version of the widely used U-Net architecture that combines the windowed self-attentions mechanism and the U-Net framework.

The Swin Transformer has demonstrated exceptional performance in various vision applications, especially when processing large images.

This architecture is recently used for image segmentation, the Swin-Transformer builds on the Vision-Transformer by computing attention limited to a local window and using shifted windows to provide connections between windows which significantly improve the power of architectural modeling. [3]

Restricting attention to a local window allows it to have a linear computational complexity to capture the image size relative to the quadratic of Vision Transformer.[3]

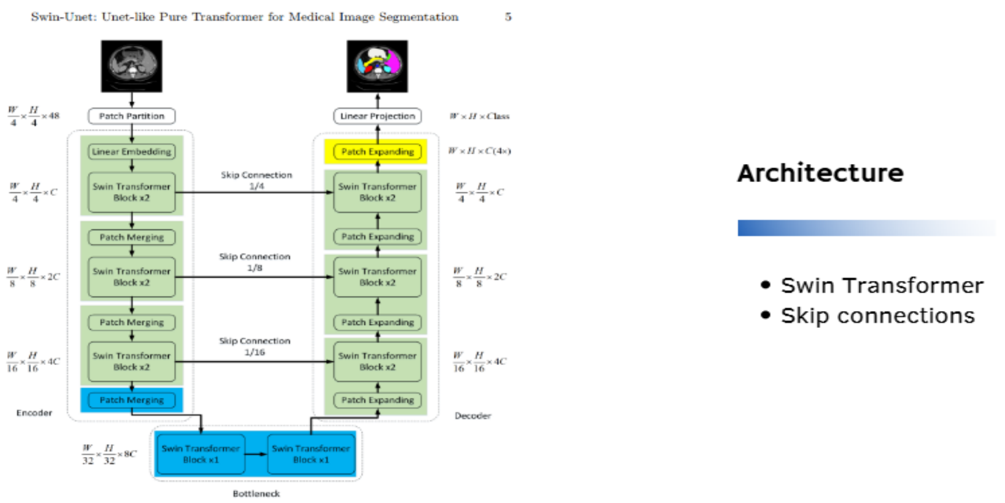


Fig 5 : Swin-Unet Architecture

PARATRANS-CNN

The latter model has a parallelized encoder structure, in which one branch uses ResNet to extract local information from images, while the other branch uses Transformer to extract global information.

Additionally, it integrates pyramid structures into the Transformer to extract global information at different resolutions, especially in intensive prediction tasks.

To efficiently use the different information from the parallelized encoder at the decoder stage, this model uses a channel attention module to merge the encoder functionalities and propagate them through skip connections and bottlenecks. [4]

Intensive numerical experiments were performed on aortic, cardiac, and multi-organ vessel tree datasets. Comparing with state-of-the-art medical image segmentation methods such as Unet and TransUnet.

This method is considered to have better segmentation accuracy, especially on small organs.

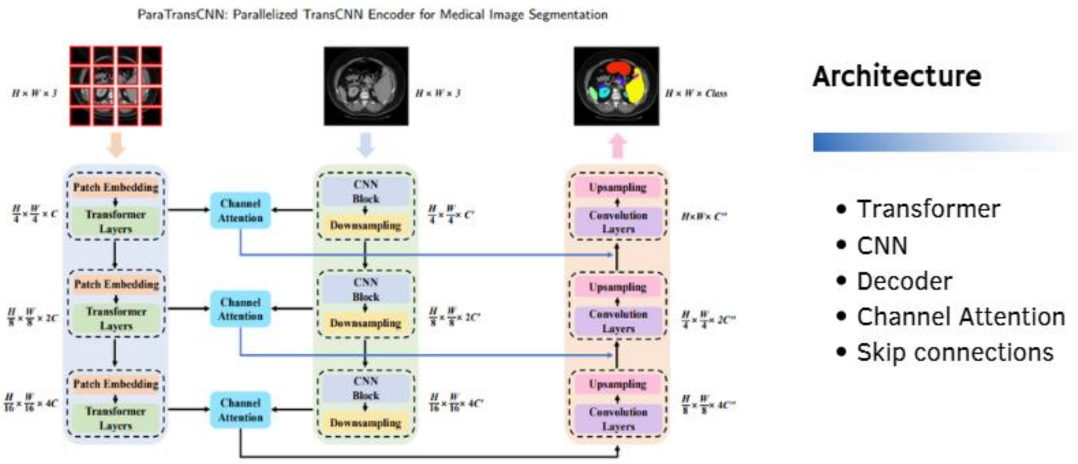


Fig 6 : Para-Trans CNN Architecture

From the ParaTransCNN architecture, we have implemented **03 new hybrid architectures:**

- **ParaPVTCNN**: we replace each Transformers Layers with a Pyramid Vision Transformers (more precisely the PVT v2).

- **ParaPVTCNN with Spatial Attention**: here, in addition to using Pyramid Vision Transformers, we replace Channel Attention blocks with Spatial Attention blocks.

- **ParaPVTCNN with Channel and Spatial Attention combined**: here, in addition to using Pyramid Vision Transformers, we replace the Channel Attention blocks with blocks combining both Channel Attention and Spatial Attention.

COMPARISON OF MODELS PERFORMANCE

Now, after this brief presentation of the models sought and their architectures, we proceed to a detailed comparison of the performances of the latter after training carried out by our team.

Models	Dataset	Batchsize	Max epochs	Max iterations	Mean Dice	Mean hd95
ParaPVTCNN	Synapse	24	150	20000	0.81	21.48
ParaPVTCNN_CA_SA	Synapse	24	150	20000	0.80	21.10
ParaPVTCNN_SA	Synapse	24	150	20000	0.73	24.03
TransUnet	Synapse	24	150	20000	0.75	37.28
Unet	Synapse	24	60	20000	0.75	35.67
SwinUnet	Synapse	24	150	20000	0.78	21.55

Table 1: Comparison of metrics of trained models

GRAPHIC INTERFACE

To ensure an easy way to test and visualise the results of our research, we developed a light interface using the framework Tkinter on Python 3.9. This interface, as demanded in the original specifications, can be used by researchers or doctors to see the before and after pipeline of our trained models.

As you can see, you can select a model and upload a file to be loaded in the application. You can choose already trained models to compute in real time the mask output of your image or models that have output files that were previously computed (resource efficient).

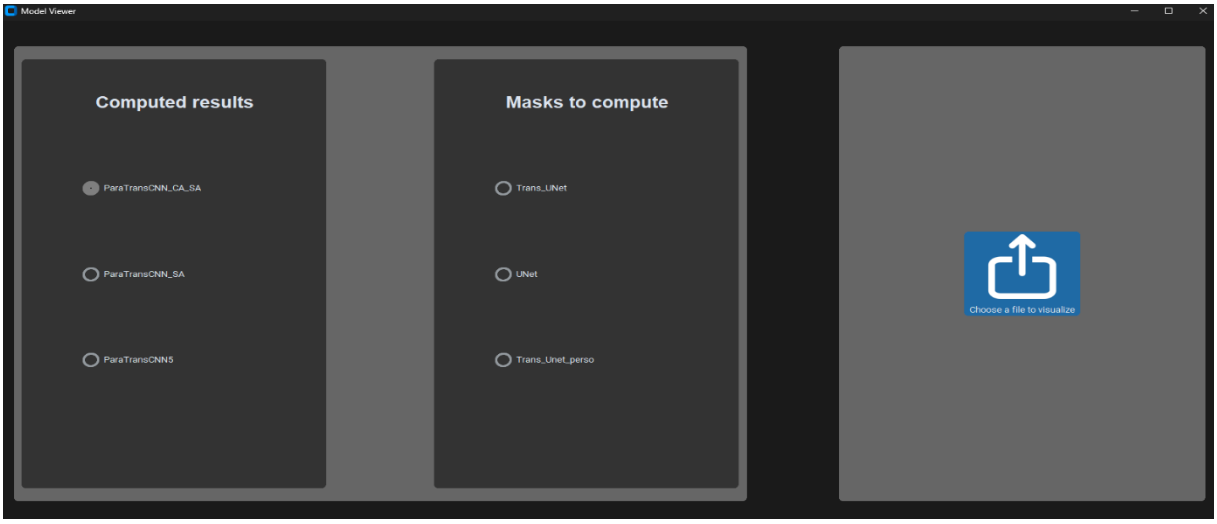


Fig 7 : Home page of the interface

Then, a new page will open displaying your selected image ("*raw image*") and the computed mask ("*mask image*"). A slider with an open area is displayed to select which slice of your image you want to see and its meta-data when available.

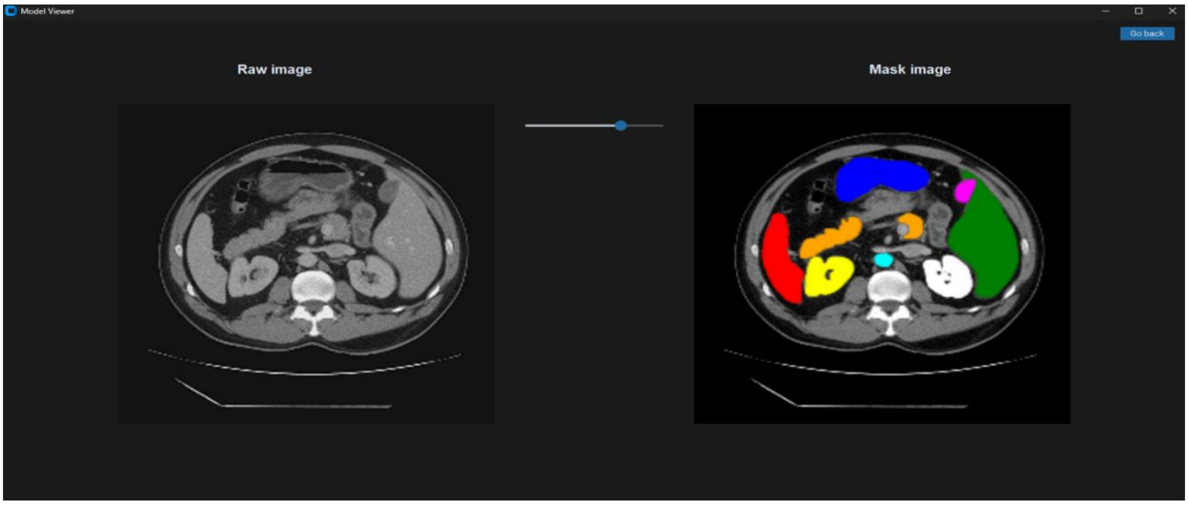


Fig 8 : displayed results on the interface

PART 3 : EXPLANATION OF OUR STRATEGY

The models to be trained were chosen based on the state of the art that we had carried out. This literature review enabled us to obtain a ranking of the models already trained on a dataset like ours, and thus to retain the models with the best performance.

Depending on the feasibility, our time constraints and our allocated computing power, we opted for architectures with pre-trained weights on a dataset like ours, as this optimizes model training times. In addition, we chose architectures that we could understand from our current Deep Learning knowledge.

In general, our decisions were taken in line with the customer's needs and suggestions. For example, the ParaPVTCNN model is the result of a customer's suggestion that we modify the ParaTransCNN model.

Finally, to choose the final model, we compared the performance of the different models we trained. More concretely, we chose the ParaPVTCNN model because it presented the best "dice_score" and "95% Hausdorff distance" metrics on the test set.

PART 4: CHALLENGES AND SOLUTIONS

During our research project, we encountered several major difficulties. First, we invested a great deal of time and effort in scientific research, without always coming up with concrete solutions. The exploratory nature of the project often led us to dead ends, forcing us to test new approaches and pursue our research. However, this allowed us to develop our expertise in Deep Learning.

At the outset, we found it difficult to pinpoint the project's purpose, as the initial objectives were a bit complex, leading to confusion and a lack of clear direction for the team. In addition, as a research-led project, we didn't always have the knowledge or resources to carry out the required tasks. To manage this challenge, we focus on communication with the customer to be sure that we understood and fulfilled his needs at each step of the project. Concretely, we had weekly meetings with the customer to show him our progress and results.

Another challenge was the fact that the hardware available to test and train each model was not directly available to us, so it slowed our progress and consumed a lot of time. Our models were trained and tested by the client on his servers.

The diversity of specializations (AI Engineer, Data Engineer, Data Scientist) within our group sometimes led to divergent approaches, with everyone having different expectations of the concrete results of the project. Moreover, all the group members had different timetables, so it was a bit hard for us to handle the project time management. To manage this challenge, we made up an Agile team management as explained before. This had a real boost effect on our productivity.

Furthermore, learning the PyTorch library, which we had never used before, represented a significant challenge. It required considerable adaptation on the part of the whole team. To manage this, we followed some online tutorials in order to handle and better understand this library.

These challenges tested our ability to manage uncertainties and obstacles, but also strengthened our determination to see this project through despite the difficulties encountered.

CONCLUSION

When it comes to medical image segmentation, our project introduced a novel approach by combining Parallel Pyramid Vision Transformers (ParaPVT) and Convolutional Neural Networks (CNN) to create the ParaPVTCNN model. This innovative model significantly outperformed our initial expectations, surpassing the performance of the models it was inspired by.

This project provided invaluable learning experiences for both the group and individual members. We gained deep insights into overcoming complex challenges and thriving in a collaborative research environment. Additionally, we enhanced our skill sets in understanding hybrid architectures and the intricacies of image segmentation.

Overall, our work not only contributed to the field of medical image analysis but also demonstrated the potential of advanced AI techniques in improving diagnostic accuracy and efficiency.

BIBLIOGRAPHY

[1] TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation (https://arxiv.org/abs/2102.04306)

[2] U-Net: Convolutional Networks for Biomedical Image Segmentation (https://arxiv.org/abs/1505.04597)

[3] Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation (https://arxiv.org/abs/2105.05537)

[4] ParaTransCNN: Parallelized TransCNN Encoder for Medical Image Segmentation (https://arxiv.org/abs/2401.15307)

[5] PVT v2: Improved Baselines with Pyramid Vision Transformer (https://arxiv.org/abs/2106.13797)

[6] CBAM: Convolutional Block Attention Module (https://arxiv.org/abs/1807.06521v2)

[7] Medical Image Segmentation on Synapse multi-organ CT (https://paperswithcode.com/sota/medical-image-segmentation-on-synapse-multi)