



**POLITECNICO**  
**MILANO 1863**

SCUOLA DI INGEGNERIA INDUSTRIALE  
E DELL'INFORMAZIONE

PROJECT REPORT

## Healthy Abdominal Organ Segmentation

MSC IN COMPUTER SCIENCE AND ENGINEERING - ARTIFICIAL INTELLIGENCE

Author: PIETRO CAFORIO

Academic year: 2023-2024

### 1. Introduction

This is a project for the "Biomedical Computer Vision" course of the Passion in Action catalogue of extracurricular courses proposed by Politecnico di Milano. The goal of this project is to predict segmentations of abdominal organs on dicom scans through deep learning models.

### 2. The dataset

The dataset proposed is the one from the CHAOS (Combined Healthy Abdominal Organ Segmentation) Challenge. The CHAOS dataset [3] is a collection of data used for research in medical imaging, particularly for abdominal organ segmentation (liver, kidneys, and spleen) from computed tomography (CT) and magnetic resonance imaging (MRI) data with their corresponding segmentation masks. The CHAOS dataset has been widely used in the development of automatic segmentation algorithms for abdominal organs. These algorithms aim to automatically identify and delineate the liver, kidneys, and spleen from CT or MRI scans, replicating the work of radiologists but with the potential for faster and more consistent results. In this specific case, only the MRI in T1DUAL configuration samples were used.

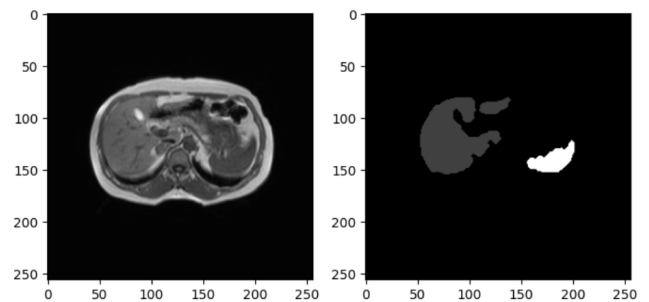


Figure 1: Sample slice with its corresponding label mask identifying different organs through different gray scales

### 3. Data Preprocessing

Using Pydicom, we load in-phase DICOM images, standardize voxel dimensions, and reshape volumes to a consistent format. Normalizing pixel distributions enhances data suitability for machine learning. The preprocessed dataset is saved as a NumPy file. Ground truth masks undergo similar reshaping and align seamlessly with volumes. A visual check confirms dataset integrity. The result is a uniformly processed dataset ready for machine learning model integration.

### 4. Models

In this section I'm describing the models and thought process used during the development to obtain the final performance.

#### 4.1. UNet model

The first baseline model I implemented is the UNet model. This model consists of an encoder to capture features and a decoder to generate segmentation maps. Key features include skip connections between corresponding encoder and decoder layers, enabling the retention of fine details. U-Net is widely used in biomedical image analysis, especially for tasks like identifying organs or tumors in medical images, hence it's one of the best choices for this specific project.

To try to obtain better performance I used image augmentation (section 5) which didn't improve the performance of the model. This made me realize that probably the model was not "complex" enough to encode the information from augmented images and that's why I decided to implement the next model.

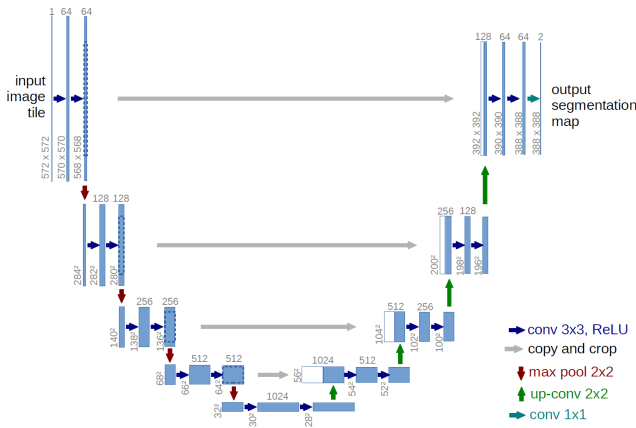


Figure 2: UNet architecture schema showing encoding (on the left side) and decoding (on the right side) part of the architecture and the skip connections and concatenation operations between the upsampling steps

#### 4.2. VGG16-Unet architecture attempt

Motivated by the intriguing concept of applying transfer learning and fine-tuning, I delved into the exploration of cutting-edge Convolutional Neural Networks (CNNs) to seamlessly integrate them into the UNet architecture. During my research, I found an implementation of VGG16 tailored to process single-channel images [1]. Eager to elevate the performance of UNet, I strategically fused this VGG16 variant into the encoding phase, leveraging its capabilities, and harnessed the outputs of selected layers for the

network's forwarding segment.

However, the anticipated breakthrough remained elusive. This implementation failed to yield any discernible improvement, leaving the model unable to effectively learn the segmentation task.

#### 4.3. MultiResUnet model

Researching for some sophisticated simple to implement model I bumped into the MultiResUnet model [2]. This architecture consists in enhancing the U-Net architecture for biomedical image segmentation by introducing a MultiRes block inspired by the Inception family networks. This block replaces the original U-Net's convolutional layer pairs and incorporates multiresolution analysis by factorizing 5x5 and 7x7 convolution operations into 3x3 ones. A residual path is added to this configuration. Additionally, to address semantic distance between Encoder and Decoder networks, Res Paths are introduced, including extra convolutions along the shortcut path. The combination of MultiRes blocks and Res paths results in the proposed architecture called MultiResUNet

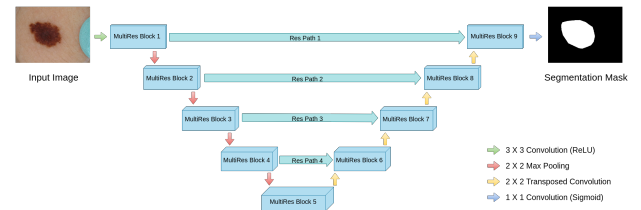


Figure 3: MultiResUnet architecture schema

This model allowed me to implement more emphasized data augmentation.

#### 4.4. Final Ensemble model

Once I got some nice performing models, to improve the generalization and consistency of the predictions I used a weighted ensemble of the predictions of the models.

### 5. Image augmentation

Data Augmentation has been implemented using the Albumentations library, a powerful tool for image augmentation. The augmentation pipeline is encapsulated within the A.Compose class, allowing for a sequential application of

transformations. Specifically, two augmentations have been employed:

- **Random Crop:** Images in the training set undergo a random cropping operation with a fixed width and height of 256 pixels. This operation introduces spatial variability to the training data, enhancing the model’s ability to handle diverse image compositions.
- **Blur:** A blurring operation is applied with a high probability of 80%. This introduces a degree of smoothness to the images, potentially aiding the model in learning robust features and reducing sensitivity to fine details. Upon further consideration, I have opted to incorporate an additional augmentation pipeline that includes the introduction of a single layer of Gaussian Noise to the images. This is a scenario that could feasibly occur in real-world situations.

## 6. Models training

For every model I used the AdamW optimizer, a variant of the Adam optimizer that incorporates weight decay directly into its update step. Weight decay is a regularization technique that penalizes large parameter values to prevent overfitting. In addition to that I used some callbacks, in particular:

- **early stopping:** to prevent overfitting (hence improve generalization capabilities), monitoring the mean validation IoU during training and stopping the training process when the performance on this metric stops improving for a certain number of steps (patience parameter).
- **reduce learning rate on plateau:** to fine-tune the training process for better convergence by adjusting the learning rate when the mean validation IoU stops improving for a certain number of steps (less than the patience parameter). I slightly tweaked the hyperparameters for each specific case to find the most suitable configuration.

## 7. Model evaluation

In order to evaluate the models the IoU (Intersection Over Union) metric has been used. It’s a commonly used metric for evaluating the performance of segmentation models. It measures the overlap between the predicted segmentation

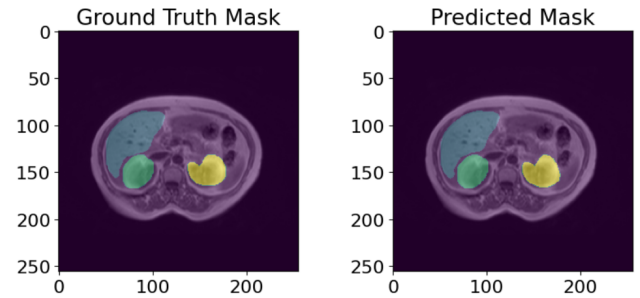
and the ground truth segmentation. The IoU is calculated as the ratio of the area of intersection between the predicted and ground truth regions to the area of their union. It is expressed as a value between 0 and 1, where 0 indicates no overlap, and 1 indicates perfect overlap.

### 7.1. Final test set performances

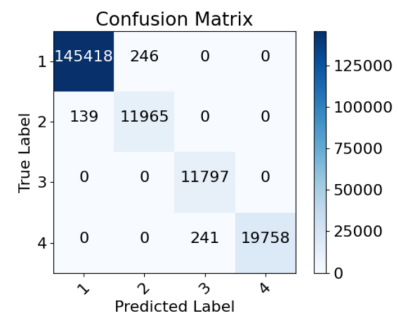
**Test set mean IoU of the implemented models**

	Mean IoU	Accuracy
<b>UNet</b>	0.776	0.989
<b>MRUNet</b>	0.7703	0.9886
<b>Ensemble</b>	0.777	0.9891

**Table 1:** Test-set performance of the final implemented models



**Figure 4:** Prediction example of the model



**Figure 5:** Confusion Matrix of the final ensemble model.

Class 1 : Liver

Class 2 : Right Kidney

Class 3 : Left Kidney

Class 4 : Spleen

## 8. Conclusions

Ultimately, the combination of ensemble techniques and augmentation did not yield significant enhancements in the performance of the UNet and MultiResUnet models, only contributing marginal percentage increases in test set accuracy and mean IoU. This outcome underscores the consistent and state-of-the-art performance that the UNet architecture offers in the biomedical domain. Overall, this project proved valuable as it allowed me to practically delve into techniques and architectures that were previously only within the realm of theoretical understanding.

## References

- [1] Saif Gazali. Cell nuclei segmentation using vgg16-unet and double-unet, 2021.
- [2] Nabil Ibtehaz and M Sohel Rahman. Multiresunet: Rethinking the u-net architecture for multimodal biomedical image segmentation. *Neural Networks*, 121:74–87, 2020.
- [3] Ali Emre Kavur, M. Alper Selver, Oğuz Dicle, Mustafa Barış, and N. Sinem Gezer. CHAOS - Combined (CT-MR) Healthy Abdominal Organ Segmentation Challenge Data, April 2019.