



Automatic Segmentation of Healthy Abdominal Organs using CT and MR images from the CHAOS Challenge Dataset

Prepared by: Juan Pablo Ramirez (10990150)

1. Introduction

Medical imaging has been a crucial tool for modern medicine by providing detailed representations of internal body structures. Advanced technologies like Computed Tomography (CT) (see Table 6) and Magnetic Resonance Imaging (MRI) (see Table 6) have transformed medical diagnostic techniques by permitting non-invasive and highly accurate examinations of organs and tissues. Therefore, the precise interpretation of these images is critical for diagnosing health issues, tracking progress of a treatment, and assisting in surgical procedures.¹ Nevertheless, manual analysis and segmentation (i.e. demarcation of structure boundaries) of these images is labor-intensive, time-consuming, and prone to human error. As a result, automated organ segmentation has become decisive in medical image analysis, with the goal of improving accuracy and efficiency while reducing the workload of healthcare professionals.²

Hence, this paper aims to explore and evaluate different computer vision techniques and Deep Learning models for the automated segmentation of organs in CT and MR images. Specifically, the primary objectives include segmenting the liver from CT scans and delineating solid abdominal organs (liver, spleen, and kidneys) from MR sequences using the CHAOS (Combined Healthy Abdominal Organ Segmentation) Challenge dataset.³

2. About the Dataset

The CHAOS Challenge dataset is a renowned data collection that has served as a benchmark for evaluating methods in medical image segmentation due to its diverse set of CT and MR imaging of healthy abdominal organs (i.e. without any pathological abnormalities) including various imaging conditions and anatomical variations. The dataset was sourced from 80 routine abdominal screening procedures conducted at the Dokuz Eylul University Hospital (Izmir, Turkey), with scan images stored in DICOM format (see Table 6) and corresponding annotations as PNG files. It is important to remark that CT data only includes annotations for the liver, while the MR sequences encompass T1-DUAL in-phase (see Table 6), out-phase (see Table 6) (both registered), and T2-SPIR (see Table 6) (not registered) images, alongside annotations for the liver, right kidney, left kidney, and spleen.³

Furthermore, this study exclusively uses the training subset of the dataset, containing 20 patients for CT scans and an equivalent number for MR sequences. The exploration of the dataset is encapsulated in Table 1 and Figure 1, revealing distinct statistical characteristics between CT and MR images due to their inherent differences in data acquisition and resolution. Comparing the two modalities, CT scans exhibit a higher 2D spatial resolution of 512x512 pixels with a broader range of axial slices per patient (ranging from 81 to 266), but each pixel representing a smaller voxel (see Table 6) in the real-world. On the other hand, MRI scans typically have a lower 2D spatial resolution of 256x256 pixels containing fewer slices (around 32 per patient), but each pixel representing a bigger voxel (see Table 6) in the real-world.

Attribute	CT	MR
# patients	20	20
# sets	20	20(x3)
2D Spatial resolution	512x512	256x256/288x288/320x320
# axial slices per patient [min-max]	[81-266]	[26-50]
Average axial slices	144	32(x3)
Total axial slices	2874	647(x3)
Voxel X spacing (mm) [min-max]	[0.568-0.791]	[1.441-2.031]
Voxel Y spacing (mm) [min-max]	[0.568-0.791]	[1.441-2.031]
Voxel Z spacing (mm) [min-max]	?	[5.5-9.0]
Sampling overlap (mm) [min-max]	?	[0.5-1.0]
Resolution (hz/pixel)	?	523 (T1) / [2.71-2.84] (T2)

Table 1. CHAOS training subset CT and MR Imaging Attributes

3. Methodology

To address the organ segmentation problem, a structured pipeline with five phases was implemented: data exploration, data pre-processing, model experimentation, model selection, and final model assessment.

The first phase involved loading the original DICOM files from the CHAOS Challenge dataset and extracting useful metadata and raw images in order to conduct exploratory data analysis (EDA) and gain a complete understanding of the dataset. The second phase aimed to apply distinct pre-processing techniques to the CT and MR images in order to reduce intrinsic disturbances, extract pertinent features for segmentation, and prepare the datasets for subsequent model training. The third phase involved an iterative experimentation with various deep learning architectures and hyperparameter tuning, separately for CT and MR datasets. Finally, the fourth and fifth phases focused on selecting the models that demonstrated the best performance on the validation set and assessing them on a test dataset, ensuring its generalizability and robustness.

All phases of the pipeline were implemented through Python programming and Python libraries executed within the Google Colaboratory environment, taking advantage of GPU acceleration and cloud-based resources for model training and computation-intensive tasks.

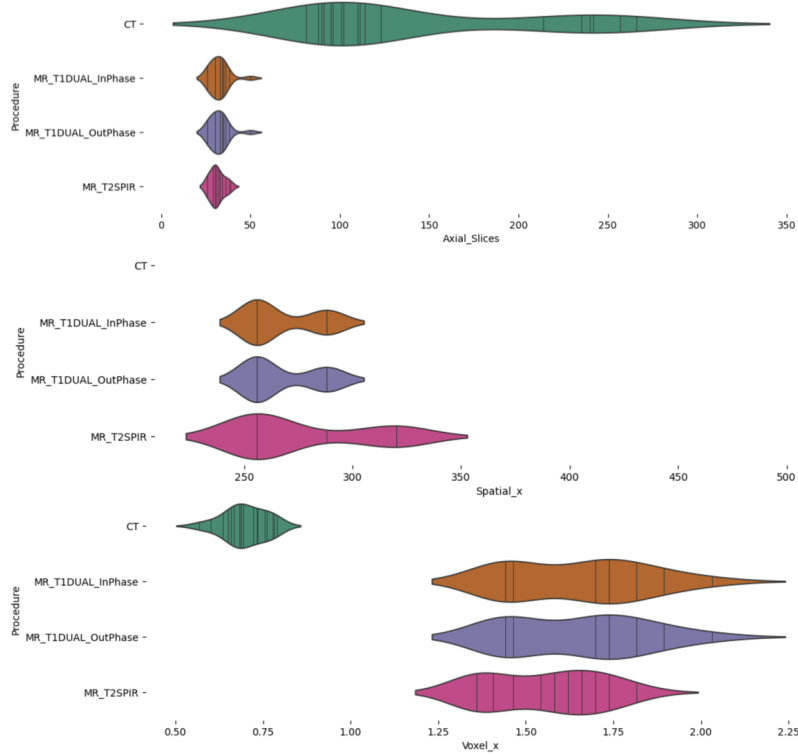


Fig. 1. Distribution of axial slices, spatial and voxel dimensions of the dataset medical images.

3.1. Exploratory Data Analysis (EDA)—This phase involved an examination of the CHAOS Challenge dataset to reveal its key characteristics and distributions.

3.1.1. Data acquisition—The original CHAOS Challenge dataset was obtained from Zenodo.⁴ Specifically, the 890.8MB train set with DICOM files and ground-truth annotations was used. This portion of the dataset comprised CT, MR T1-weighted inphase, MR T1-weighted outphase, and MR T2-weighted SPIR imaging modalities of 20 different patients’ abdomen.

3.1.2. Data loading and metadata extraction—DICOM files were loaded into memory, and key metadata (axial slices number, spatial dimensions, voxel dimensions, sample overlap, resolution, and aspect ratio) were extracted to facilitate posterior image pre-processing. Subsequently, a metadata dictionary was constructed to organize and store the extracted metadata efficiently. It was serialized and saved using the `pickle` module, enabling fast retrieval and utilization in subsequent processing steps. Finally, to speed up future data loading procedures, both the raw images and their annotations were serialized and saved as numpy files to avoid the inherent delay associated with extracting DICOM images directly.

3.1.3. Analysis and visualization—A complete analysis of the dataset was performed using the efficient dictionary data format that included the dataset metadata. Violin plots on Figure 1 were used to depict the distributions of essential features such as axial slices, spatial dimensions, and voxel dimensions across various imaging modalities. These charts provided insights into

the variability and distribution of dataset characteristics to guide the pre-processing and model selection strategies.

3.2. Dataset pre-processing—This phase aimed to improve the quality and utility of both CT and MR images for subsequent organ segmentation tasks. Additionally, this step is crucial to produce better performance on Deep learning models.⁵

3.2.1. Computed Tomography (CT)—

- **[Resampling]** In segmentation tasks, it is generally preferred to have input images with uniform physical spacing and consistent resolution to prevent findings that are specific to the center or dependent on reconstruction techniques.⁵ However, given that all CT images had same resolution (512x512 pixels) and the complete voxel data was unavailable, resampling was **not performed** to avoid potential loss of information.
- **[Denoising]** CT images can be affected by various disturbances, primarily including beam hardening, patient movements, scanner malfunctions, low resolution, intrinsic low-dose radiation, and the presence of metal implants.⁵ Therefore, various denoising techniques were explored, including **averaging, Gaussian, median, and bilateral filters**.
- **[Sharpening]** To accentuate organ boundaries,⁵ multiple sharpening filters were evaluated, including **Laplacian, sharpen filter, emboss filter, and unsharp mask**.
- **[Normalization]** Intensity values were normalized at slice-level to the range **[0, 255]** to standardize pixel intensities across all CT images. This makes the dataset fit efficiently into 8 bit images monotonically.⁵
- **[Background cleaning]** An **Otsu threshold filter** was employed to remove non-relevant background elements such as the CT bed, ensuring focus on the target organs.
- **[Histogram equalization]** Standard **histogram equalization, contrast limited adaptive histogram equalization (CLAHE), and gamma correction** were tested to enhance the contrast of the CT images by redistributing the intensity values across the entire range.⁶
- **[Label one-hot encoding]** Given the segmentation task, CT label annotations were one-hot encoded to assign a **value of 1 for the liver** pixels and 0 for the rest of the image.

3.2.2. Magnetic Resonance Imaging (MRI)—

- **[Resampling]** Given the different resolutions and physical spacing of each MR, **resampling was performed** (using dataset statistics) to ensure uniform image size and isotropic voxel dimensions, enhancing comparability and facilitating subsequent processing steps.
- **[Denoising]** Denoising is typically integrated into modern MR scanners, but in practice, additive Gaussian, Rician, and Speckle noise, along with respiratory, body movements

and aliasing, can still introduce significant noise.⁵ Therefore, similar to CT images, **averaging, Gaussian, median, and bilateral filters** was tested on the MR images.

- **[Bias Field correction]** "Bias field" distortion is a low-frequency non-uniformity in MR data that causes MR intensity values to vary across images taken from the same scanner, the same patient, and even the same tissue.⁵ Hence, **N4 expectation-maximization algorithm** was applied for bias field correction of all MR images.
- **[Standardization]** Intensity values were **standardized** to maintain consistency across MR images, ensuring uniformity in feature representation within the same body region across all different slices and patients.⁵
- **[Registration]** MR T1-weighted inphase and outphase images were already registered so they were just **aligned and joined** spatially.
- **[Label one-hot encoding]** Given the multi-class segmentation task, MR label annotations were one-hot encoded to represent five slices: background, liver, right kidney, left kidney, and spleen.

3.3. Model training and experimentation—This phase employed a systematic methodology to develop and evaluate deep learning models for organ segmentation using the CHAOS Challenge dataset. For the CT models, the inputs consisted of 512x512 2D images, and the outputs were 512x512 segmentation maps where liver pixels were labeled as 1 and all other pixels as 0. For the MR T1-weighted models, the inputs comprised 256x256x2 2D images, which included aligned and combined inphase and outphase samples to enhance the complementary information. The outputs for these models were 256x256x5 segmentation maps, with each slice representing the map for a specific organ (liver, right kidney, left kidney, spleen, or background). Finally, for the MR T2-weighted SPIR models, the inputs were 256x256 2D images, and the outputs were similarly 256x256x5 segmentation maps, with each slice corresponding to the map for each organ.

3.3.1. Naïve thresholding on CT pre-processed images—As an initial step, a naïve intensity value thresholding method was applied to the pre-processed CT images to segment the liver. This simple approach served as a baseline to compare against the performance of more sophisticated deep learning models.

3.3.2. Dataset preparation—Four distinct datasets were prepared to facilitate comprehensive model training and evaluation.

- (1) Raw CT images with liver annotations.
- (2) Pre-processed CT images with liver annotations.
- (3) Pre-processed MR T1-weighted inphase and outphase (joined) images with annotations for all organs (liver, right kidney, left kidney, spleen).
- (4) Pre-processed MR T2-weighted SPIR images with annotations for all organs (liver, right kidney, left kidney, spleen).

Moreover, data was split at the patient level to ensure that the training, validation, and test sets were mutually exclusive. Hence, models were evaluated on unseen data, providing a realistic assessment of their generalization capabilities.

- **[Training Set]** 16 patients (80%)
- **[Validation Set]** 2 patients (10%)
- **[Test Set]** 2 patients (10%)

3.3.3. *Model parameter setup*—

- **[Batch size]** 16
- **[Epochs]** 12-20
- **[Optimizer]** Adam with an initial learning rate of 1×10^{-4} (decrease on plateau)
- **[Loss function]** DICE loss (used in original CHAOS challenge)³
 - $\text{DICE loss} = 1 - 2 \frac{|S \cap G|}{|S| + |G|}$
- **[Metrics]** DICE coefficient, Relative Absolute Volume Difference (RAVD), mean Intersection over Union (IoU), and accuracy (used in original CHAOS challenge)³
 - $\text{DICE} = 2 \frac{|S \cap G|}{|S| + |G|}$
 - $\text{RAVD} = \frac{\text{abs}(|S| - |G|)}{|G|} \times 100$
 - $\text{IoU} = \frac{S \cap G}{S \cup G}$

3.3.4. *Model architectures*—Chosen for their proven efficacy in medical image segmentation tasks and computational limitations on training.

- **[Autoencoder]**
- **[Small U-net]**
- **[Large U-net]**
- **[Trans U-net]**

3.3.5. *Experiments*—

- (1) Liver segmentation on CT raw images
- (2) Liver segmentation on CT pre-processed
- (3) Multi-class segmentation of liver, kidneys and spleen on pre-processed MR T1-weighted inphase and outphase images
- (4) Separate independent segmentation of liver, kidneys and spleen on pre-processed MR T1-weighted inphase and outphase images
- (5) Multi-class segmentation of liver, kidneys and spleen on pre-processed MR T2-weighted images
- (6) Separate independent segmentation of liver, kidneys and spleen on pre-processed MR T2-weighted images

3.4. *Model selection and evaluation*—In this phase, models were chosen based on their performance metrics (DICE, RAVD, mena IoU, and accuracy) on the validation set. Subsequently, the models demonstrating superior performance across these metrics were assessed on the independent test data to validate their generalization capabilities.

4. Results

4.1. CT pre-processing—After exploring various computer vision techniques as outlined in the methodology, the final pre-processing pipeline for CT images was selected to optimize clarity and contrast. The pre-processing result is illustrated in Figure 2.

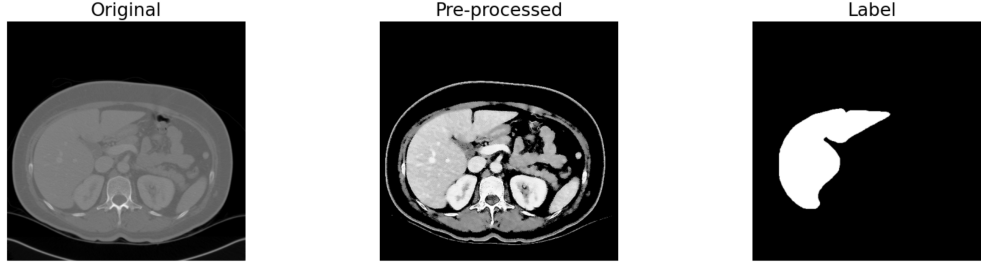


Fig. 2. Pre-processing result of CT images.

- (1) **[Bilateral Filter]** 9x9 filter with spatial sigma and color sigma both set to 75 was able to reduce noise while preserving edges.
- (2) **[Normalization]** Intensity values were normalized to an 8-bit range [0-255].
- (3) **[Sharpen Filter]** 3x3 sharpening filter with the kernel $\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$ was applied to enhance the edges within the CT images.
- (4) **[Background cleaning]** Otsu threshold filter was employed to remove non-relevant background elements, such as the CT bed.
- (5) **[Standard Histogram Equalization]** Applied to enhance the contrast of the CT images and improve visibility of different tissue types.

4.2. CT liver intensity threshold segmentation—Initially, intensity threshold segmentation was used as a baseline strategy to detect the liver in pre-processed CT images. This approach used pixel intensity values to separate the liver from the surrounding tissues.

As depicted in Figure 3, organ tissues were successfully segmented. However, each image required its own specific intensity value for correct thresholding. Additionally, the liver was not segmented in isolation, but all the tissues within the abdominal section were included with this segmentation approach. This serves as a baseline, highlighting that deep learning models should not only segment all organ tissues but also discriminate and isolate the target organ accurately.

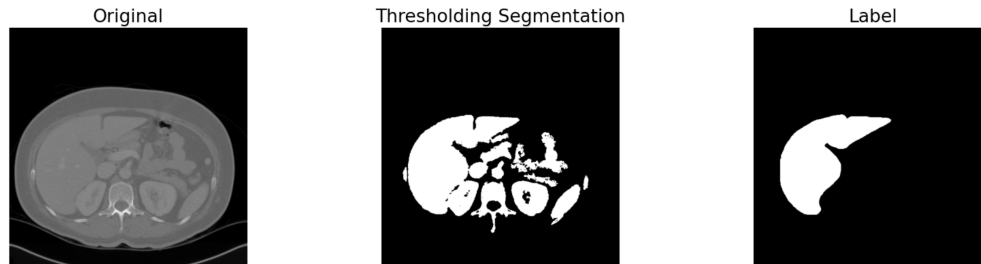


Fig. 3. CT liver intensity threshold segmentation.

4.3. CT liver segmentation models—As illustrated in Table 2, the experiments highlight the effectiveness of advanced deep learning architectures, particularly the Trans U-Net architecture, for CT liver segmentation. Moreover, the use of pre-processed CT images provided additional benefits, however the most significant gains were due to the model’s complexity and architecture.

Architecture	CT Dataset	Parameters	DICE coef	RAVD (%)	Mean IoU	Accuracy (%)
Autoencoder	Raw	332,801	0.6544	？**	0.7118	96.39
Small U-net	Raw	573,249	0.7526	？**	0.8118	97.01
Large U-net	Raw	31,030,593	0.9306	？**	0.9328	99.27
Large U-net	Pre-processed	31,030,593	0.9306	4.16	0.9313	99.30
Trans U-net	Pre-processed	37,332,289	0.9470	2.04	0.9466	99.46

Table 2. Validation CT liver segmentation performance on different architectures.

** value was not generated due to half-precision floating point numbers for more efficient model training

Specifically, the validation metrics indicate a clear performance hierarchy among the tested architectures, with more complex models generally outperforming simpler ones. The autoencoder, with the fewest parameters, showed the lowest performance across all metrics, with a DICE coefficient of 0.6544 and mean IoU of 0.7118. This suggests that the model was able to capture general features, but lack the detailed spatial information needed for precise liver segmentation. In contrast, the small U-Net demonstrated an improvement over the autoencoder, achieving a DICE coefficient of 0.7526 and a mean IoU of 0.8118.

Furthermore, the large U-Net, when applied to the raw dataset, significantly outperformed both the autoencoder and small U-Net, with a DICE coefficient of 0.9306 and a mean IoU of 0.9328. This improvement in performance reveals the effectiveness of deeper and more complex architectures for detailed segmentation tasks. When the large U-Net was trained on pre-processed CT data, the DICE coefficient remained unchanged, but accuracy and mean IoU showed slight improvements. These results suggest that pre-processing can enhance model performance by providing cleaner and more uniform input data, even if small improvement in this case.

Finally, the Trans U-Net showed the best performance across all metrics, with a DICE coefficient of 0.9470, the lowest RAVD at 2.04%, and a mean IoU of 0.9466. This model with the transformer mechanism enabled to better capture global context and detailed spatial relationships, leading to a better segmentation performance.

4.4. MR pre-processing—Following some testing on computer vision techniques outlined in the methodology, the selected pre-processing pipeline is depicted in Figure 4.

- (1) **[Bilateral Filter]** 9x9 filter with spatial sigma and color sigma both set to 75 was able to reduce noise while preserving edges.
- (2) **[Bias Field Correction]** N4 expectation-maximization bias field correction was utilized to address intensity non-uniformity caused by the magnetic field.

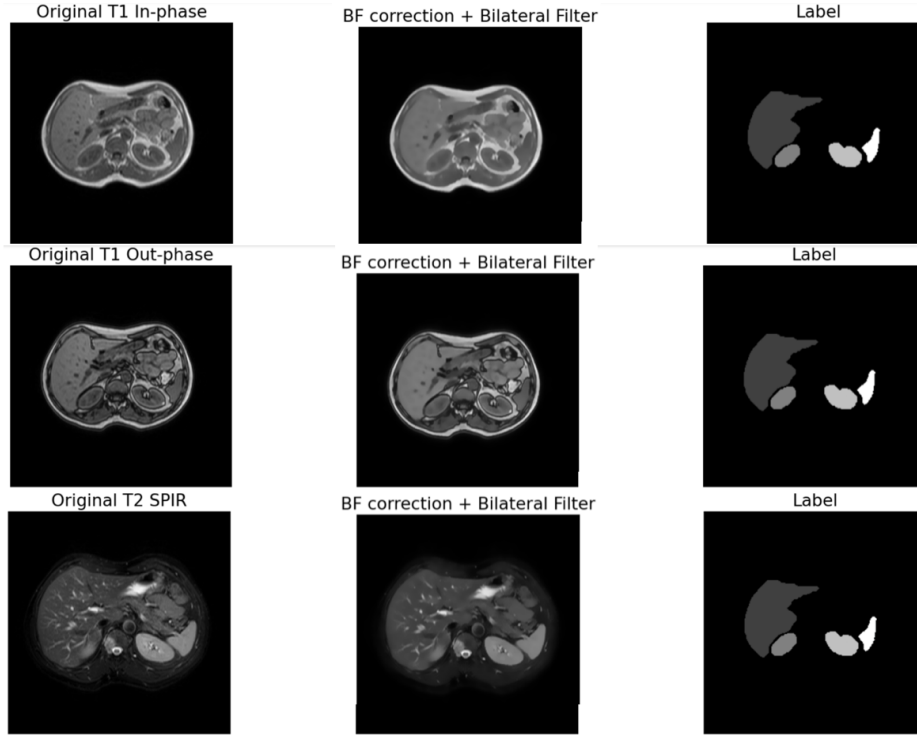


Fig. 4. Pre-processing result of MR images.

- (3) **[Resampling]** MR volumes were resampled to a target shape of $32 \times 256 \times 256$, representing the mean axial slice count and the most repeated resolution among the MR images. Additionally, resampling was performed to achieve isotropic voxel dimensions of $1.5 \times 1.5 \times 1.5$ mm, ensuring uniformity in spatial resolution across all axes.
- (4) **[Normalization]** Intensity values were normalized to an range $[0-1]$.
- (5) **[Standardization]** Scaling intensity values to have zero mean and unit variance across all MR images of same type to ensure consistent intensity distributions among tissues.

4.5. MR organ segmentation models—The results presented in Table 3 and Table 4 reveal the diverse performance across different architectures and MR datasets for organ segmentation tasks. The overall segmentation appears to be sub-optimal, potentially influenced by the chosen architecture, pre-processing pipeline, training strategies, computational constraints, or distinctive features and tissue contrasts inherent in the T1-weighted and T2-weighted datasets. However, certain models demonstrated better performance in segmenting individual organs, compared to the joint prediction task of simultaneous segmentation of multiple organs.

4.5.1. Joint segmentation of organs—In the case of T1-weighted inphase and outphase images, the autoencoder exhibited moderate performance, with DICE coefficients ranging from 0.2525 for the right kidney to 0.5682 for the liver. However, the small U-net and large U-net models failed to produce meaningful segmentations, producing DICE coefficients close to zero for most organs. Similarly, for T2-weighted SPIR images, while the autoencoder achieved relatively higher DICE coefficients, particularly for the liver and spleen, the small U-net and large U-net models showed a low performance, with DICE coefficients varying across organs. Interestingly, the Trans U-net yielded zero DICE coefficients across all organs and datasets, indi-

cating a complete failure to segment any structures, demonstrating limitations in its architecture or training strategy.

Architecture	MR Dataset	Parameters	Liver DICE	R kidney DICE	L kidney DICE	Spleen DICE
Autoencoder	T1 in+out	5'496,741	0.5682	0.2525	0.3532	0.2752
Small U-net	T1 in+out	573,461	0.0	0.0116	0.0074	0.0073
Large U-net	T1 in+out	31'031,429	0.0	0.5	0.25	0.0075
Trans U-net	T1 in+out	37'333,125	0.0	0.5	0.25	0.0
Autoencoder	T2 SPIR	5'496,453	0.4974	0.6849	0.5056	0.5167
Small U-net	T2 SPIR	573,317	0.3591	0.1024	0.0897	0.0986
Large U-net	T2 SPIR	31'030,853	0.3786	0.7951	0.0074	0.0094
Trans U-net	T2 SPIR	37'332,549	0.0	0.0076	0.0072	0.0

Table 3. Validation MRI organ segmentation performance of joint prediction on different architectures

4.5.2. Single independent segmentation of organs—In the context of single independent prediction using the large U-net architecture, T1-weighted images achieved DICE coefficients of 0.5396 for the liver, 0.0156 for the right kidney, 0.0698 for the left kidney, and 0.0185 for the spleen. Similarly, T2-weighted SPIR images demonstrated improved segmentation performance, particularly for the liver and right kidney, with DICE coefficients of 0.5760 and 0.8301, respectively. These results suggest that, despite improvement from joint segmentation, there remains room for improvement, especially in the segmentation of smaller and less represented structures in the dataset such as the left kidney and spleen.

Architecture	MR Dataset	Parameters	Liver DICE	R kidney DICE	L kidney DICE	Spleen DICE
Large U-net	T1 in+out	31'030,593	0.5396	0.0156	0.0698	0.0185
Large U-net	T2 SPIR	31'030,593	0.5760	0.8301	0.2396	0.3867

Table 4. Validation MRI single independent organ segmentation performance on large U-net

4.6. Model selection and evaluation—Models and dataset for each organ were selected based on their performance on the validation set, prioritizing the DICE coefficient metric (the higher the better). These top-performing models were evaluated on the test set to verify generalization capabilities, as illustrated in Table 5.

Organ	Architecture	Dataset	Parameters	Test DICE coef
Liver	Trans U-net	CT pre-processed	37,332.289	0.963
Right kidney	Large U-net	MR T2 SPIR	31,030.593	0.8293
Left kidney	Autoencoder	MR T2 SPIR	5,496.453	0.6239
Spleen	Autoencoder	MR T2 SPIR	5,496.453	0.5167

Table 5. Evaluation of top-performing model on test set for each organ

For liver segmentation, the Trans U-net architecture trained on pre-processed CT images was the best model, achieving a DICE coefficient of 0.9470 in validation and 0.963 in testing (see

Figure 5), approaching the state-of-the-art with a DICE coefficient of 0.979.³ This outstanding result highlights the efficacy of the Trans U-net model in accurately delineating liver boundaries from properly pre-processed CT scans

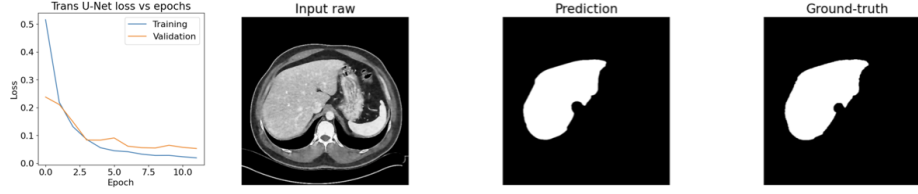


Fig. 5. Liver segmentation from CT images with Trans U-net model.

Regarding the right kidney segmentation, the Large U-net architecture trained on T2-weighted SPIR MR images for single-organ prediction, demonstrated great performance with a DICE coefficient of 0.8301 in validation and 0.8293 in testing (see Figure 6).

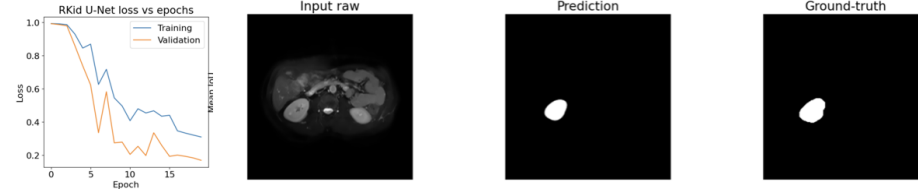


Fig. 6. Right kidney segmentation from MR T2SPIR images with large U-net model.

Furthermore, the autoencoder architecture trained on MR T2-weighted SPIR images for joint prediction of multiple organs achieved a moderate performance in segmenting the left kidney and spleen (see Figure 7). For left kidney segmentation, the model achieved a DICE coefficient of 0.5056 in validation and 0.6239 in testing. For spleen segmentation, it achieved a DICE coefficient of 0.5167 in both validation and testing.

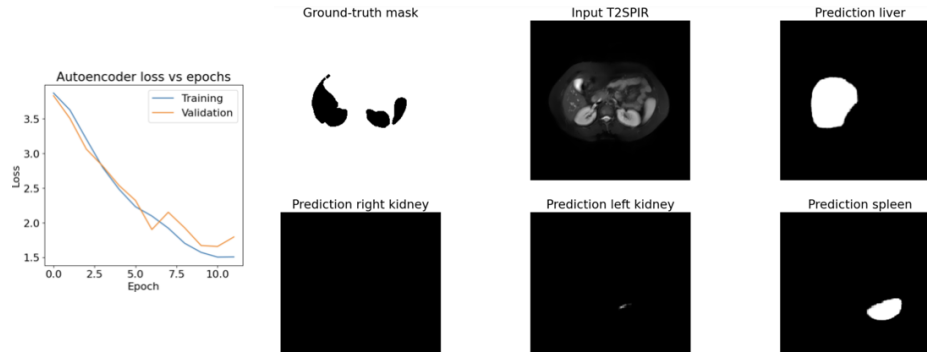


Fig. 7. Left kidney and spleen segmentation from MR T2SPIR images with autoencoder model.

Overall, all models outperformed the intensity thresholding baseline as expected. These results demonstrated the importance of factors such as imaging modality, pre-processing pipelines,

anatomical complexity, and model architecture for achieving accurate and reliable automatic organ segmentation for clinical practice.

5. Conclusions

This project demonstrated the potential of advanced computer vision techniques and deep learning models in the automated segmentation of organs from CT and MR images from the CHAOS Challenge dataset. Through careful data pre-processing and experimentation with various deep learning architectures, a significant performance of organ segmentation have been achieved. Specifically, the Trans U-net model, combined with pre-processed CT images, exhibited superior performance, demonstrating the importance of both sophisticated model architectures and robust pre-processing pipelines in medical image analysis.

Furthermore, the results indicate that while simple thresholding methods can serve as a baseline, they do not accurately isolate target organs. Therefore, deep learning models are capable of capturing intricate spatial relationships and providing precise segmentation maps.

Finally, future research may involve integrating CT and MR imaging, exploring transfer learning, fine-tuning and larger deep learning architectures with more computational resources, incorporating other medical imaging modalities, tracking different metrics, follow different pre-processing strategies, and using newer methods for automatic medical image segmentation.

References

- ¹ Vision XRAY group “Importance of Medical Imaging.” (2013) (accessed 2024) <https://xray.com.au/importance-of-medical-imaging/>.
- ² Pinto-Coelho, L. “How Artificial Intelligence Is Shaping Medical Imaging Technology: A Survey of Innovations and Applications.” *Bioengineering (Basel)* **10.12** <https://doi.org/10.3390/2Fbioengineering10121435>.
- ³ Kavur, E., *et al.* “CHAOS Challenge - combined (CT-MR) healthy abdominal organ segmentation.” **69** <https://doi.org/10.1016/j.media.2020.101950>.
- ⁴ Kavur, A. E., Selver, M. A., Dicle, O., Barış, M., Gezer, N. S. “CHAOS - Combined (CT-MR) Healthy Abdominal Organ Segmentation Challenge Data.” (2019) doi:10.5281/zenodo.3362844 URL <https://doi.org/10.5281/zenodo.3362844>.
- ⁵ Masoudi, S., *et al.* “Quick guide on radiology image pre-processing for deep learning applications in prostate cancer research.” *J Med Imaging (Bellingham)* **8.1** <https://doi.org/10.1117/1.JMI.8.1.010901>.
- ⁶ Sadia, R., Chen, J., Zhang, J. “CT image denoising methods for image quality improvement and radiation dose reduction.” *Journal of Applied Clinical Medical Physics* <https://doi.org/10.1002/acm2.14270>.
- ⁷ Preston, D. “Magnetic Resonance Imaging (MRI) of the Brain and Spine: Basics.” (2006) (accessed 2024) <https://case.edu/med/neurology/NR/MRI%20Basics.htm>.
- ⁸ Merkle, E., Nelson, R. “Dual Gradient-Echo In-Phase and OpposedPhase Hepatic MR Imaging: A Useful Tool for Evaluating More Than Fatty Infiltration or Fatty Sparing.” *RadioGraphics* **16.5** <https://doi.org/10.1148/rg.265055711>.
- ⁹ Zee, C., *et al.* “SPIR MRI in spinal diseases.” *Journal of Computer Assisted Tomography* **16.3** 356–360 (1992) <https://doi.org/10.1097/00004728-199205000-00004>.
- ¹⁰ Medical Imaging Technology Association (MITA) “About DICOM: Overview.” (accessed 2024) <https://www.dicomstandard.org/about>.

¹¹ Gerber, A., Peterson, B. “What is an Image?” *J Am Acad Child Adolesc Psychiatry* **47.3** 245–248 (2008)
<https://doi.org/10.1097%2FCHI.0b013e318161e509>.

Appendix A: Medical Vocabulary

Term	Description
Computed Tomography (CT)	Generated by measuring x-ray attenuation through a rotating energy source, resulting in detailed cross-sectional images of the body. CT imaging offers valuable anatomical insights by leveraging tissue density variations. However, its use may be limited in certain clinical contexts due to concerns about ionizing radiation exposure and its relatively constrained soft-tissue contrast. ⁵
Magnetic Resonance Imaging (MRI)	Depict the radio-frequency energy emitted during proton realignment in a robust magnetic field. Renowned for their superior resolution and soft-tissue contrast, MR images often present challenges in display and analysis due to the absence of a standardized intensity scale. Moreover, even with consistent scanner usage, significant variations in MR images persist, complicating interpretation. ⁵
MRI T1-weighted	MRI sequences capitalize on the longitudinal relaxation time (T1) to accentuate tissue differences. Short Time to Echo (TE) and Repetition Time (TR) are employed to produce images where protons in fat and water are either in-phase or out-of-phase. ⁷ In-phase alignment yields constructive interference, while out-of-phase alignment leads to signal cancellation, especially at fat-water interfaces. This technique is particularly adept at detecting lesions containing both fat and water components, such as hepatic lesions in the liver. ⁸
MRI T2-weighted	MRI sequences employ the transverse relaxation time (T2) to measure the duration for spinning protons to lose phase coherence perpendicular to the main magnetic field. These sequences emphasize variations in tissue T2 relaxation times, achieved through longer Time to Echo (TE) and Repetition Time (TR). ⁷ Additionally, the SPIR (Spectral Presaturation with Inversion Recovery) technique selectively saturates signals from specific tissues while nullifying signals from others (e.g. fat) to enhance image contrast and provide exceptional soft tissue contrast. ⁹
Digital Imaging and Communications in Medicine (DICOM)	International standard for medical images format with patient, study, series, and instance information. ¹⁰
Volume Element or Voxel	In medical imaging, each digital pixel corresponds to a 3D square or rectangular chunk of a body structure. ¹¹

Table 6. Vocabulary terms

