# Report Project Work Medical Imaging

Salvatore Bruno, Mario Apicella, Emanuela Fiorillo, Domenico Schettini, Alexandre Engel
Group 2

***Abstract*—Liver segmentation is a fundamental task in medical imaging, playing a crucial role in diagnosis, treatment planning, and surgical navigation. However, deep learning models often struggle with domain shifts between imaging modalities, such as CT and MRI, leading to a loss in generalization and performance degradation when applied across different datasets.**
**This work aims to evaluate the effectiveness of domain augmentation techniques, particularly AugMix and structured augmentations, in improving cross-modality liver segmentation. Additionally, a CycleGAN-based approach was explored to generate domain-transformed images (CT to MRI and vice versa) and was used exclusively as a baseline for comparison against our augmentation-based approach.**
**The study utilizes the CHAOS dataset, consisting of 20 CT and 20 MRI abdominal scans, to train and test U-Net segmentation models. The AugMix-based model was compared to the CycleGAN-based model, assessing segmentation accuracy through Dice Score. Results indicate that augmentation strategies significantly mitigate domain shift, with the AugMix-trained model achieving better generalization than the CycleGAN-based approach, which introduced artifacts affecting segmentation performance.**
**These findings demonstrate that structured augmentations provide a robust alternative to unsupervised domain adaptation methods, ensuring greater anatomical preservation in cross-modality segmentation tasks. Future work may focus on hybrid approaches combining generative models and augmentation techniques for further performance improvements.**

## I. Introduction

LIVER segmentation is a crucial task in abdominal medical imaging, essential for diagnosis, surgical planning, and treatment monitoring. However, deep learning models struggle with **domain shifts** between Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) due to differences in acquisition protocols, contrast, and scanner characteristics, leading to reduced performance on unseen modalities. To address this issue, this work explores domain augmentation techniques to enhance segmentation robustness across modalities. Before any augmentation or training phase, all images undergo a preprocessing step to standardize intensity distributions, normalize pixel values, and binarize segmentation masks. This ensures consistency across CT and MRI images, reducing inter-scanner variability and improving model generalization.

The primary augmentation approach leverages **AugMix** and structured augmentations, introducing controlled photometric and geometric transformations to enhance robustness. Additionally, **CycleGAN** was employed to generate domain-transformed images (CT to MRI and vice versa) for comparison purposes. However, CycleGAN was **not used in the final model** but rather as a baseline to evaluate its effectiveness against augmentation-based strategies.

A U-Net-based model is trained on the CHAOS dataset following a **Leave-One-Domain-Out** scheme to rigorously evaluate cross-modality generalization. The training pipeline consists of:

1) **Preprocessing of images**
2) **Training the model** on one domain (**CT** or **MRI**) without domain augmentation.
3) **Testing the model** on the opposite domain (**MRI** or **CT**) without domain augmentation.
4) **Retraining the model** on the same initial domain with domain augmentation (**AugMix** and structured augmentations).
5) **Testing the augmented model** on the second domain to assess the effectiveness of the domain adaptation strategy.

Performance is evaluated using the **Dice Score (DS)**, a widely used metric in medical imaging that measures the overlap between the predicted segmentation and the ground truth. It is defined as:

$$DS = \frac{2 \times |S_{\text{Algorithm}} \cap S_{\text{GroundTruth}}|}{|S_{\text{Algorithm}}| + |S_{\text{GroundTruth}}|} \tag{1}$$

where:

- $S_{\text{Algorithm}}$ represents the predicted segmentation mask.
- $S_{\text{GroundTruth}}$ represents the ground truth segmentation mask.
- The numerator captures the intersection between the two sets, while the denominator normalizes the score.

The Dice Score ranges from 0 to 1, where 1 indicates perfect segmentation (complete overlap) and 0 indicates no overlap. Additionally, different loss functions—**Dice Loss, Binary Cross-Entropy (BCE), and Jaccard Loss**—are analyzed with assigned relative weights to optimize segmentation performance and model convergence.

## II. Preprocessing

Medical image preprocessing is a crucial step in the pipeline of automated analysis, ensuring that input data is clean, standardized, and suitable for downstream tasks such as segmentation or classification. Raw medical images often suffer from noise, intensity variations, and inconsistencies due to different acquisition devices and protocols. Preprocessing techniques aim to enhance image quality, normalize data distributions, and remove irrelevant artifacts, ultimately improving model robustness and generalization. In this work, preprocessing is particularly important to mitigate domain shifts between **(CT)** and **(MRI)**, facilitating cross-modality segmentation of the liver.

### A. DICOM Image Preprocessing

Medical imaging datasets often contain raw **DICOM** images with varying intensity distributions, scanner-dependent artifacts, and modality-specific characteristics. To ensure consistency and improve segmentation performance, a preprocessing pipeline is applied to the dataset. The following steps are performed:

*1) Rescale Slope and Intercept Correction:* DICOM images store pixel values that require a transformation to **Hounsfield Units (HU)** or other intensity scales based on scanner calibration. The values are adjusted using the formula:

$$I_{\text{corrected}} = I_{\text{raw}} \times \text{Slope} + \text{Intercept} \tag{2}$$

where **Slope** and **Intercept** are metadata parameters stored in the DICOM headers. This ensures that the intensity values are correctly mapped to meaningful ranges for segmentation.

*2) Windowing (Contrast Enhancement):* Different imaging modalities (CT, MRI) exhibit varying intensity distributions. To enhance contrast and highlight relevant structures, **Window Center (WC)** and **Window Width (WW)** are applied. The windowing transformation is defined as:

$$I_{\text{windowed}} = \frac{\max(\min(I, WC + WW/2), WC - WW/2) - (WC - WW/2)}{WW} \tag{3}$$

This operation standardizes the intensity range while preserving important anatomical details.

*3) Normalization:* To ensure uniform intensity distribution across different samples, pixel values are normalized to the range $[0, 1]$:

$$I_{\text{normalized}} = \frac{I - \min(I)}{\max(I) - \min(I)} \tag{4}$$

This step is crucial for stabilizing deep learning model training, as it reduces the impact of intensity variations across different scans.

*4) Handling Segmentation Masks:* Segmentation masks in the dataset require specific preprocessing steps to ensure compatibility with the segmentation model. The approach differs for **CT** and **MRI**.

*a) CT Mask Processing:* CT segmentation masks are **binary**, where:

$$M_{\text{binary}}(x, y) = \begin{cases} 1, & \text{if } M(x, y) > 127 \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where $M(x, y)$ represents the pixel intensity in the original mask. This ensures a clear distinction between the **liver region** and the background.

*b) MRI Mask Processing:* MRI segmentation masks are **multiclass**, containing annotations for multiple organs such as the liver, kidneys, spleen, and blood vessels. Since our goal is to segment only the **liver**, it was necessary to extract its corresponding mask. To achieve this, the grayscale intensity distribution of MRI masks was analyzed. The **liver** was found to correspond to a **dark gray intensity range**, approximately between **15 and 100**. The binarization process was performed using the following thresholding approach:

$$M_{\text{binary}}(x, y) = \begin{cases} 1, & \text{if } T_{\text{low}} \leq M(x, y) \leq T_{\text{high}} \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

where $T_{\text{low}} = 15$ and $T_{\text{high}} = 100$ define the valid liver intensity range. To further refine the selection, the threshold was set to the value closest to **50**, ensuring consistency across different MRI scans.

*c) Ensuring Consistency Between Modalities:* Since MRI masks originally contained multiple labels, all **non-liver structures were removed**, setting their pixel values to zero. If no valid liver region was detected, the mask was replaced with a fully black image to avoid incorrect annotations. This **harmonization step** ensures that both **CT and MRI masks are processed in a consistent binary format**, facilitating cross-domain training and improving the segmentation model's generalization capabilities.

## III. Training Without Augmentation

To establish a baseline, the U-Net model was trained and tested separately on CT and MRI data without any augmentation. The model's performance was evaluated using Dice Score, Binary Cross-Entropy (BCE) Loss, Dice Loss, and Jaccard Loss . The results show high segmentation accuracy when trained and tested on the same domain.

TABLE I
PERFORMANCE METRICS FOR MRI → MRI

| Metric | Value |
|---|---|
| Dice Score | 0.8929 |
| BCE Loss | 0.0173 |
| Dice Loss | 0.1071 |
| Jaccard Loss | 0.1927 |
| Total Loss | 0.0979 |

TABLE II
PERFORMANCE METRICS FOR CT → CT

| Metric | Value |
|---|---|
| Dice Score | 0.9242 |
| BCE Loss | 0.0845 |
| Dice Loss | 0.0827 |
| Jaccard Loss | 0.1508 |
| Total Loss | 0.0889 |

The higher Dice Scores (0.89 for MRI, 0.92 for CT) confirm that the model performs well when trained and tested on the same domain.

### A. Cross-Domain Generalization

To evaluate the impact of domain shift, the trained models were tested on the opposite modality (CT model on MRI and vice versa). The results indicate a severe degradation in segmentation performance, with Dice Scores close to zero.

TABLE III
PERFORMANCE METRICS FOR CT → MRI

| Metric | Value |
|---|---|
| Dice Score | 0.0097 |
| BCE Loss | 0.6166 |
| Dice Loss | 0.9897 |
| Jaccard Loss | 0.9948 |
| Total Loss | 0.8675 |

TABLE IV
PERFORMANCE METRICS FOR MRI → CT

| Metric | Value |
|---|---|
| Dice Score | 0.00015 |
| BCE Loss | 0.6863 |
| Dice Loss | 0.9998 |
| Jaccard Loss | 0.9999 |
| Total Loss | 0.9067 |

### B. Observations

- When tested on the same domain, the model achieves high segmentation accuracy(Dice Score: 0.89-0.92).
- When tested on a different domain, performance drops drastically (Dice Score: 0.00015-0.0097), confirming the presence of a severe domain shift between CT and MRI.

- The results highlight the need for domain adaptation techniques, motivating the use of augmentation strategies to improve generalization. (se avanza spazio mettere i grafici)

## IV. Data Augmentation

To enhance segmentation robustness across different imaging modalities, **data augmentation** techniques were applied in two stages:

1) **Initial Data Augmentation:** Applied directly to raw CT and MRI images to simulate cross-modality characteristics.
2) **AugMix Augmentation:** Performed on the raw images again, to generate additional diverse variants of the original scans.

This two-step approach ensures both **domain adaptation** (CT $\rightarrow$ MRI and vice versa) and **intra-domain variability**, enhancing the model's robustness to intensity variations and structural noise.

### A. Stage 1: Cross-Modality Domain Transformation

The first augmentation step transforms images between CT and MRI domains by modifying intensity distributions while preserving anatomical structures.

*a) CT to MRI Transformation:* CT images were converted into MRI-like images through a multi-step, tissue-aware transformation:

- **Edge Detection**: Extracts tissue boundaries using the Canny edge detector.
- **Distance-Based Intensity Adjustment**: Modifies pixel intensities according to their distance from tissue edges.
- **Tissue Classification**: Segments pixels into air, cerebrospinal fluid (CSF), fat, muscle, bone marrow, organs, and bone using predefined intensity thresholds.
- **MRI Weighting Simulation**:
  - **T1-weighted MRI**: CSF appears dark, fat is very bright, muscle has moderate intensity, bone marrow is bright, and bones and air remain dark.
  - **T2-weighted MRI**: CSF is bright, fat appears darker than in T1, muscle is dark, bone marrow has intermediate intensity, and bones remain dark.
- **Bilateral Filtering**: Reduces noise while preserving structural details.
- **Adaptive Histogram Equalization**: Enhances local contrast to resemble MRI characteristics.
- **Gaussian Blurring**: Smoothens intensity transitions while preserving details.

This process generates both **T1-weighted** and **T2-weighted** MRI representations from CT images.

*b) MRI to CT Transformation:* MRI images were adapted to resemble CT scans using intensity and contrast modifications:

- **For T1DUAL (InPhase and OutPhase):**
  - **Histogram Equalization and Contrast Inversion**: Simulates X-ray attenuation characteristics by adjusting intensity distribution.
  - **Gamma Correction**: to enhance CT-like contrast.
  - **Laplacian Filtering**: Highlights tissue boundaries.
  - **Sobel Edge Detection**: Enhances anatomical structures by refining contours.
  - **Adaptive Background Suppression**: Maintains dark air regions, improving contrast.
  - **Final Gaussian Blur**.
- **For T2SPIR:**
  - **Histogram Equalization and Contrast Adjustment**: Improves contrast for a more CT-like intensity range.
  - **Gaussian Smoothing**: Reduces noise and enhances homogeneity.
  - **Edge Enhancement (Laplacian Filtering)**: to improve visualization.
  - **Sharpening and Gaussian Noise**: to mimic CT grain.

### B. Stage 2: AugMix-Based Augmentation

After cross-modality transformations, a second augmentation step was applied to the **raw images** (before conversion) using **AugMix**, a technique designed to enhance robustness by generating diverse augmented versions of the same image.

*a) AugMix Pipeline:* AugMix constructs new images by:

- Applying multiple **photometric transformations**:
  - **Noise Addition**: Salt-and-pepper noise, Gaussian noise.
  - **Blur Filters**: Gaussian blur, median blur, bilateral filtering.
  - **Contrast, Brightness and Sharpness Adjustments**.
  - **Power-Law (Gamma) Transformation**.
  - **Histogram Equalization (CLAHE)**.
- Applying **geometric transformations**:
  - **Rotation**: Random rotations within $[-80°, +80°]$.
  - **Translation**: Random shifts along X and Y axes within $[-10°, +10°]$ pixels.
  - **Flipping**: Horizontal or both horizontal and vertical flipping.
- Constructing multiple augmented versions and blending them using a **Dirichlet mixture model**.
- Enforcing similarity constraints with the **Structural Similarity Index (SSIM)** to ensure realistic variations.

*b) Consistency in Segmentation Masks:* To ensure label consistency, the same **geometric transformations** (rotation, flipping, translation) were applied to the segmentation masks, while photometric augmentations were **excluded** to avoid altering the class labels.

### C. Impact of Augmentation

The dual-step augmentation approach resulted in:

- Improved **domain adaptation** by generating realistic CT-to-MRI and MRI-to-CT transformations.
- Enhanced **intra-domain variability** via AugMix, reducing sensitivity to noise and contrast changes.
- Better **generalization** across imaging conditions, improving segmentation robustness to unseen modalities.

The combination of tissue-aware transformations and statistical augmentation techniques ensures that the dataset remains anatomically valid while introducing substantial variations in contrast and texture, ultimately improving model performance.

## V. Testing and Generalization of the Augmented Model

The augmented model was evaluated to assess its ability to generalize across imaging modalities. As previously observed, when trained and tested on the same domain (**CT → CT** and **MRI → MRI**), the model maintained **high segmentation accuracy**. However, the primary focus of this evaluation is to analyze the model's performance on **cross-domain testing**, where segmentation is performed on an unseen modality.

### A. Cross-Domain Generalization Results

The performance of the model trained on one modality and tested on the other (**CT → MRI** and **MRI → CT**) is summarized in Table V and VI.

TABLE V
Performance Metrics for CT → MRI

| Metric | Value |
|---|---|
| Dice Score | 0.3419 |
| BCE Loss | 0.5672 |
| Dice Loss | 0.6581 |
| Jaccard Loss | 0.7946 |
| Total Loss | 0.6581 |

TABLE VI
Performance Metrics for MRI → CT

| Metric | Value |
|---|---|
| Dice Score | 0.7502 |
| BCE Loss | 0.1565 |
| Dice Loss | 0.2498 |
| Jaccard Loss | 0.3988 |
| Total Loss | 0.2514 |

### B. Analysis of Generalization Performance

The results indicate a significant improvement in **cross-domain generalization** compared to the non-augmented model:

- **CT → MRI**: The Dice Score increased from **0.0097** (non-augmented) to **0.3419**, demonstrating a substantial improvement.
- **MRI → CT**: The Dice Score improved from **0.00015** (non-augmented) to **0.7502**, indicating strong generalization.

While the model performs better on **MRI → CT** than **CT → MRI**, this discrepancy is likely due to the higher structural consistency between MRI sequences and CT images following augmentation. The visualization of segmentation outputs, saved as prediction plots, further highlights these differences.

### C. Impact of Augmentation on Domain Shift

The results confirm that augmentation strategies significantly mitigate domain shift, enabling the model to extract features that generalize across modalities. Although segmentation performance on **CT → MRI** remains lower than **MRI → CT**, the improvement over the non-augmented model demonstrates the effectiveness of applying structured augmentation techniques.

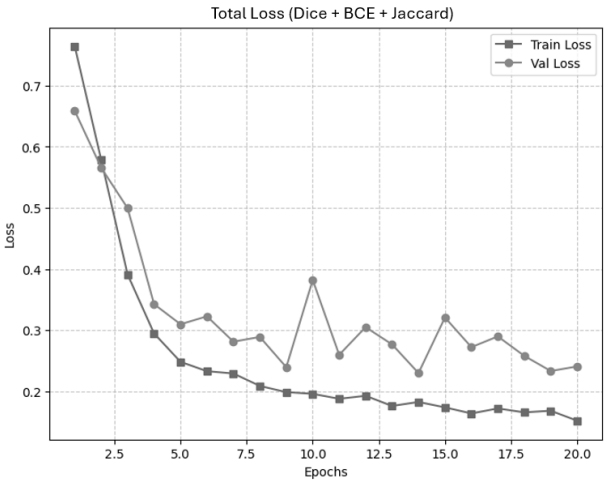


Fig. 1. Training and validation loss over epochs. The combination of Dice, BCE, and Jaccard loss ensures stable training while addressing class imbalance. The training loss steadily decreases, indicating proper learning, whereas the validation loss fluctuates due to domain variability.

The plot in Fig. 1 shows the evolution of the total loss function during training and validation. The training loss follows a smooth decreasing trend, demonstrating that the model is effectively learning. In contrast, the validation loss exhibits fluctuations, which are expected due to the differences between CT and MRI domains. These oscillations suggest that the model is sensitive to domain shifts, highlighting the importance of data augmentation strategies. Despite this, the general downward trend in validation loss indicates an improvement in generalization performance across imaging modalities.

## VI. CycleGAN: Cross-Modality Domain Adaptation

As an **additional experiment**, a **CycleGAN** model was trained to perform unsupervised image-to-image translation between **CT and MRI** scans. The goal was to **generate synthetic images** that resemble the target modality, reducing domain shift effects in liver segmentation. This step was not used for the final model but served as a **comparison benchmark** against our augmentation-based approach.

### A. CycleGAN Training

The CycleGAN model consists of:

- **Two Generators** ($G$, $F$): Convert CT to MRI and MRI to CT.
- **Two Discriminators** ($D_{CT}$, $D_{MR}$): Distinguish real from synthetic images.
- **Cycle Consistency Loss**: Ensures that translating an image to the opposite domain and back preserves its structure.
- **Identity Loss**: Enforces stability when translating already domain-consistent images.

The training process follows:

1) **Real CT images** are translated into synthetic MRI scans, and vice versa.
2) **Discriminators** learn to distinguish real from translated images.
3) **Cycle consistency loss** ensures anatomical structures remain intact.
4) **Identity loss** prevents unnecessary modifications in domain-consistent images.

The CycleGAN was trained on CT and MRI datasets, processing images in pairs. The dataset was structured with:

- **CT images from DICOM scans** (preprocessed).
- **MRI images from T1DUAL and T2SPIR sequences**.

### *B. U-Net Training with CycleGAN Transformed Images*

After training CycleGAN, the synthetic MRI and CT images were used to retrain the U-Net segmentation model, following the same pipeline:

- Train on **CycleGAN-generated CT images**, test on real MRI.
- Train on **CycleGAN-generated MRI images**, test on real CT.

### *C. Results and Comparison*

The performance of the U-Net trained on CycleGAN-generated images was compared to our previous augmentation-based approach.

TABLE VII
CYCLEGAN-BASED U-NET PERFORMANCE FOR CT → MRI

| Metric | Value |
|---|---|
| Dice Score | 0.1904 |
| BCE Loss | 0.7478 |
| Dice Loss | 0.8096 |
| Jaccard Loss | 0.8946 |
| Total Loss | 0.8077 |

TABLE VIII
CYCLEGAN-BASED U-NET PERFORMANCE FOR MRI → CT

| Metric | Value |
|---|---|
| Dice Score | 0.1378 |
| BCE Loss | 1.2501 |
| Dice Loss | 0.8622 |
| Jaccard Loss | 0.9243 |
| Total Loss | 0.9911 |

The results indicate that while CycleGAN-generated images improve domain adaptation, they do not match the effectiveness of the augmentation-based approach. The **Dice Score** remains significantly lower compared to models trained with AugMix and structured augmentations.

## VII. CONCLUSION

In this work, we explored the challenges of **liver segmentation across different imaging modalities** (CT and MRI) and proposed an approach to improve cross-domain generalization using **data augmentation** and **domain adaptation techniques**.

The results demonstrate that:

- When trained and tested on the same domain (**CT → CT** and **MRI → MRI**), the model achieves **high segmentation accuracy**, with Dice Scores above **0.89**.
- Without augmentation, the model fails to generalize across domains (**CT → MRI**: Dice Score **0.0097**, **MRI → CT**: Dice Score **0.00015**).
- **AugMix and image augmentations** significantly improve generalization, increasing Dice Scores to **0.3419** (CT → MRI) and **0.7502** (MRI → CT).
- **CycleGAN-generated images**, used as an alternative domain adaptation approach, led to lower performance (CT → MRI: **0.1904**, MRI → CT: **0.1378**), suggesting that direct augmentation remains more effective.

These findings highlight the importance of structured augmentations in mitigating **domain shift**, demonstrating that **hybrid approaches combining augmentation and domain translation could further improve segmentation robustness**. Future work could explore:

- Combining **CycleGAN and AugMix** to enhance domain generalization.
- Incorporating **self-supervised learning** to leverage unlabeled data.
- Applying **multi-scale feature fusion** to improve anatomical consistency across modalities.

The results reinforce the need for domain adaptation in medical imaging and suggest that data augmentation strategies remain a powerful tool for improving segmentation performance in cross-modality tasks.

## REFERENCES

[1] ResearchGate. The loss and the Jaccard index curves of U-Net model training using our dataset. Available at: The-loss-and-the-Jaccard-index-curves-of-U-Net-model-training.
[2] Kaggle. CycleGAN Tutorial from Scratch: Monet to Photo. Available at: CycleGAN-Tutorial-from-Scratch.
[3] G. Percannella, U. Petruzzello, F. Tortorella, and M. Vento. *Addressing Domain Shift in Cross-Modality Medical Image Segmentation*. Published in ScienceDirect. Available at: ScienceDirect-Domain-Shift-in-Medical-Imaging.
[4] GeeksforGeeks. Cycle Generative Adversarial Network (CycleGAN). Available at: GeeksforGeeks-CycleGAN.
[5] Kaggle. CHAOS Combined CT-MR Healthy Abdominal Organ Dataset. Available at: CHAOS-Combined-CT-MR-Dataset.
[6] GitHub. PyTorch U-Net: Semantic Segmentation Model. Available at: PyTorch-UNet.
[7] D. Hendrycks, N. Mu, E. D. Cubuk, B. Zoph, J. Gilmer, and B. Lakshminarayanan. *AugMix: A Simple Data Processing Method to Improve Robustness and Uncertainty*. Available at: AugMix-ArXiv.