algorithms

MDPI

Article

Automatic Myocardium Segmentation in Delayed-Enhancement MRI with Pathology-Specific Data Augmentation and Deep Learning Architectures

Gonzalo Mosquera-Rojas [1], Cylia Ouadah [1], Azadeh Hadadi [2,3], Alain Lalande [1,4,*] and Sarah Leclerc [1]

[1] IFTIM, ICMUB Laboratory, CNRS UMR 6302, University of Burgundy, 21000 Dijon, France; gonzalo.mosquerarojas2000@gmail.com (G.M.-R.); cylia.ouadah@gmail.com (C.O.); sarah.leclerc@u-bourgogne.fr (S.L.)
[2] Arts et Metiers Institute of Technology, LISPEN, HESAM Université, UBFC, 71100 Chalon-sur-Saône, France; a.hadadi1363@gmail.com
[3] Institute for Information Management in Engineering, Karlsruhe Institute of Technology, 76133 Karlsruhe, Germany
[4] Department of Medical Imaging, University Hospital of Dijon, 21000 Dijon, France
* Correspondence: alain.lalande@u-bourgogne.fr

Abstract: The extent of myocardial infarction (MI) can be evaluated thanks to delayed enhancement (DE) cardiac MRI. DE MRI is an imaging technique acquired several minutes after the injection of a contrast agent where MI appears with a bright signal. The automatic myocardium segmentation in DE MRI is quite challenging, especially when MI is present, since these areas usually showcase a heterogeneous aspect in terms of shape and intensity, thus obstructing the myocardium visibility. To overcome this issue, we propose an image processing-based data augmentation algorithm where diverse synthetic cases of MI were created in two different ways: fixed and adaptive. In the first one, the training set is enlarged by a specific factor, whereas in the second, the method receives feedback from the segmentation model during training and performs the augmentation exclusively on complex cases. The method performance was evaluated in single and multi-modality settings. In this latter, information from kinetic images (Cine MRI), which are acquired along DE MRI in the same examination, is also used, and the extracted features from both modalities are fused. The results show that applying the data augmentation in a fixed fashion on a multi-modality setting leads to a more consistent segmentation of the myocardium in DE MRI. The segmentation models, which were all UNet-based architectures, can better relate MI areas with the myocardium, thus increasing its overall robustness to pathology-specific local pattern perturbations.

Keywords: data augmentation; cardiac imaging; myocardial infarction; delayed enhancement MRI; image segmentation

Citation: Mosquera-Rojas, G.; Ouadah, C.; Hadadi, A.; Lalande, A.; Leclerc, S. Automatic Myocardium Segmentation in Delayed-Enhancement MRI with Pathology-Specific Data Augmentation and Deep Learning Architectures. *Algorithms* **2023**, *16*, 488. https://doi.org/10.3390/a16100488

Academic Editor: Antonio Lanatà

Received: 4 September 2023
Revised: 6 October 2023
Accepted: 16 October 2023
Published: 20 October 2023

1. Introduction

According to the World Health Organization, cardiovascular diseases (CVD) are among the leading causes of global mortality. In 2005, they accounted for 17 out of 58 million worldwide deaths, of which 7.6 million were due to coronary heart disease (CHD) [1]. Myocardial infarction (MI), also known as heart attack, is one of the ways in which CHD can appear [1]. This condition refers to the damage or death of a particular portion of the heart muscle (myocardium) due to an unexpected interruption or decrease of blood supply, which is caused by coronary artery blocks.

Nowadays, several strategies are used in medical practice for MI diagnosis, namely, analysis of electrocardiogram (ECG), tracking of the levels of biomarkers that could suggest potential myocardial tissue damage or death, and medical imaging. This latter embodies different techniques, which include echocardiography, scintigraphy, magnetic resonance imaging (MRI), and computed tomography (CT) [2].

After the heart has suffered from an event of MI, it is essential to evaluate the viability of the myocardium [3,4]. According to Thomson et al. [5], one of the de facto definitions of viability is "recovery of contractile function following revascularization". In medical practice, there are two main cardiac MRI techniques for evaluating the structure and function of the heart: kinetic (Cine) and delayed enhancement (DE) MRI. As its name suggests, the latter is performed several minutes after the injection of a gadolinium-based contrast agent [3]. Due to its nature, Cine MRI is more suitable for performing local and global heart motion and contraction analysis, whereas DE MRI provides valuable insights to assess myocardial tissue damage since the contrast agent allows these areas to be highlighted. The principle of the presence of hypersignal in diseased areas on DE MRI is based on the delayed washout of the gadolinium-based contrast agent in the myocardium due to an increased proportion of extracellular space. This abnormal proportion of extracellular space arises from the presence of fibrosis or edema.

The viability of the myocardium involves analyzing the MI extent, which is generally performed on short-axis images and requires the segmentation of the myocardium. This segmentation can be performed manually, but this process is time consuming and prone to inter-operator variability. Therefore, automatized approaches appear to be an appropriate alternative to perform such a task, which can be very challenging due to different factors such as artifacts and noise in the image, but notably due to the varying contrast between normal and abnormal myocardial tissue since the MI appearance ranges from being bright and well defined to being subtle and heterogeneous. International challenges such as EMIDEC in 2020 [3,4] validate the need for having reliable and robust automatic methods to perform this task. It is important to notice that DE-MRI is not only performed for the evaluation of MI. Indeed, this technique can be used to evaluate other diseases, where there is the presence of fibrosis or edema, such as myocarditis, and hypertrophic cardiomyopathy, among others. The main difference between these diseases is the localization of the bright signal at the level of the myocardium on DE MRI.

When assessing myocardial viability, both MRI modalities can offer complementary information. Therefore, it is natural to think that leveraging information from Cine MRI can have a positive influence on the segmentation performance in DE MRI, which entails addressing the problem as a multi-modality one. For instance, in the work of Dikici et al. [6], they automatically segmented the corresponding Cine MRI and used it as a segmentation prior that was deformed to maximize the overlap with the correct segmentation on DE MRI. This fit was performed through an affine registration procedure that included translation, shearing, and scaling operations. In the work of Ciofolo et al. [7], they initialized a geometrical template from DE MRI and then deformed it to match the image data, using some prior knowledge regarding shape, contours, and region intensity distribution. Afterwards, they built a 3D mesh from Cine MRI and registered it to DE MRI space. Finally, they deformed this mesh towards the 2D contours obtained in the first step, resulting in a mesh representing the myocardium in 3D. In the work of Huellebrand et al. [8], they performed DE MRI segmentation using pre-trained convolutional neural networks on a Cine MRI dataset.

Although there are several ways to use information from different modalities to segment a target, data fusion is one of the most common approaches nowadays. A review of deep learning segmentation architectures with different fusion strategies was presented by Zhou et al. [9]. They divided them into input-level fusion, layer-level fusion, and decision-level fusion. In the first one, the modalities are combined channel by channel and then fed to the segmentation network. In the second one, each modality is used as input for separate segmentation networks, and the individual features are fused within the layers of the networks. In the third one, each modality is used to train separate segmentation networks, and the output is then combined to obtain the final segmentation result. This schematic can also be thought of as output fusion.

In Ouadah et al. [10], several UNet-based fusion architectures were compared in terms of performance for left ventricle segmentation from multi-modal MRI. They found out that

an intermediate fusion UNet-based architecture (DualUNet) proved to work best when segmenting both the left ventricle cavity and the myocardium. This architecture is inspired by the work of Xue et al. [11], where T1 and T1 flipped brain scan images are fused before the decoding path for stroke lesion segmentation.

Although deep-learning-based segmentation methods have shown promising results in a variety of different applications, they do need big amounts of data in order to generalize well. This is quite problematic in the field of medical imaging, since the amount of available data is generally limited, and the annotation task is complex and costly.

Data augmentation emerged as a technique to solve the above-mentioned problem. As its name suggests, it involves generating new data samples by applying several transformations to the existing one. Some of the most common transformations reported in the literature include rotations, translations, scaling, and deformations. Additionally, there is a group of deep learning methods, commonly known as generative models, which can automatically create synthetic images, and thus have also been used for data augmentation.

There have been some previous works where data augmentation has been used for segmentation in cardiac MRI. For instance, Chen et al. [12] proposed an adversarial data augmentation pipeline that can dynamically optimize its image transformation parameters to produce results that are realistic and plausible in typical medical imaging setups. Lin et al. [13] proposed a shape-based data augmentation algorithm where they learn different representations of the left ventricle and capture structural relationships between shapes. Given a dataset of images with their corresponding left ventricle contours, one image is selected as a reference, and the contours of the rest are deformed to match the reference one. The transformation parameters are stored and taken as shape features. New shapes are generated from an orthonormal feature basis that is computed based on eigenvector analysis on the correlation matrix of shape features. Skandarani et al. [14] proposed the use of variational autoencoders and generative adversarial networks (GANs) [15] for generating realistic synthetic MRI. Their results show that training convolutional neural networks with the data generated by their model achieved competitive performance with respect to other traditional techniques.

The main objective of this work was to propose a data augmentation algorithm to improve the segmentation of the myocardium in DE MRI since deep-learning-based segmentation models struggle a lot to segment this structure when myocardial infarction is present. Through our method, synthetic myocardial infarction samples are created on healthy myocardium to provide the segmentation models with diverse pathological data so that they can further learn from patterns in these cases. This can be applied under two scenarios: fixed and adaptive. In the first one, the training set is enlarged at different rates and fed to the segmentation model. In the second one, feedback from the model is received during training to identify the cases where it is struggling the most, and new synthetic data samples are exclusively created from these particular cases. The work is validated on single-modality and multi-modality settings, using a slightly modified version of the traditional UNet architecture [16] and the DualUNet architecture [10].

2. Materials and Methods

2.1. Datasets

2.1.1. CINEDE Dataset

The CINEDE dataset used in this work consists of 124 patients with Cine and DE MRI, as shown in Figure 1.

The exams were acquired using 1.5 T and 3 T magnets (Siemens Healthineers, Erlangen, Germany) with a phased thoracic coil. DE MRI was acquired 10 min after the injection of a gadolinium-based contrast agent at a concentration of 0.1 or 0.2 mmol/kg. A T1-weighted phase-sensitive inversion recovery (PSIR) sequence was used (TR = 3.5 ms, TE = 1.42 ms, TI = 400 ms, flip angle = 20°). Cine MRI was performed using a breath-hold SSFP sequence (TR = 3.7 ms, TE = 1.6 ms, flip angle = 20°). The number of slices in short-axis orientation per volume varies between 7 and 12, with a total of 984 2D slices in short-axis

orientation. The median voxel spacing in the x, y, and z axes for the Cine and DE modalities are $[1.367, 1.367, 10]$ mm^3 and $[1.875, 1.875, 10]$ mm^3, respectively. All images have their corresponding manual annotations (ground truth) with two labels: left ventricle cavity and myocardium, corresponding to the epicardial border of the muscle.

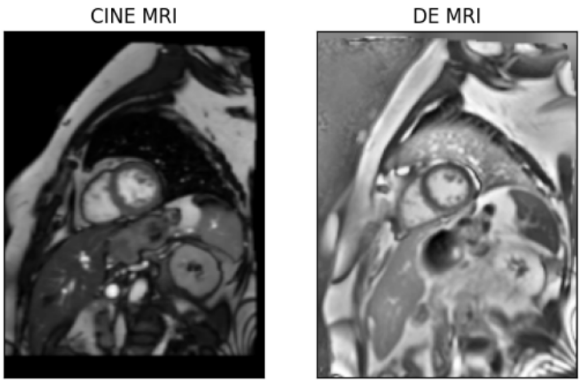

Figure 1. Cine and DE MRI slices in short-axis orientation corresponding to the same patient and localization.

The dataset represents cases from five different conditions, namely, normal examinations, dilated cardiomyopathy, myocarditis, myocardial infarction, hypertrophic cardiomyopathy, and other pathology. This diversity in the dataset translates to a large set of distinct features that the model must learn as they might impact how the anatomical structures appear, thus increasing the difficulty of the segmentation task. The sample distribution among different classes is uneven.

2.1.2. EMIDEC Dataset

EMIDEC dataset is an auxiliary data source that was used for the extraction of prior information that is fed into our data augmentation algorithm. This dataset was proposed by Lalande et al. [3] within the framework of the automatic evaluation of myocardial infarction from the DE MRI challenge. It consists of 150 DE MRI examinations with a text file that contains clinical information from the patient and the corresponding ground-truth image that maps four labels: left ventricle cavity, myocardium, myocardial infarction, and persistent microvascular obstruction (MVO) area. The data sample distribution between normal and pathological cases is one-third and two-thirds, respectively. All abnormal cases correspond to patients with acute myocardial infarction. The pixel spacing of the examinations ranges from 1.25×1.25 mm^2 to 2×2 mm^2, the slices have a thickness of 8 mm, and the distance between slices is 10 mm [3]. The other acquisition parameters were the same as for the CINEDE dataset.

2.2. Method

The problem solution comprises three critical steps, namely, data pre-processing, selection of segmentation architecture and definition of training details, and post-processing.

2.2.1. Data Pre-Processing

For this particular problem, the pre-processing consisted of several stages. First, the image pixel intensity was normalized within the range of 0–255. In the multi-modality approach, the data fusion architecture expects input images of equal size. Therefore, the size of both images in a given Cine–DE MRI pair was made equal by zero-padding the smaller one. As a third step, the spatial correspondence between the two modalities was uniform to have consistent contextual information across modalities. This was achieved by resizing all images to match the median voxel spacing of DE MRI, as it is the target modality in this study. This was performed through bilinear interpolation since it provided image aspect ratio preservation. As a final step, Cine MRI was registered to DE MRI to

make them suitable for fusion-based deep learning architectures under the multi-modality setting. Image size was fixed at 256 × 256 in order to speed up the training process and thus decrease the computational cost. In the work of Rukundo [17], it was also found that a segmentation network trained on 256 × 256 DE MRI obtains lower cross-entropy loss than the one obtained when trained on 128 × 128 images, which further supports the selection of 256 × 256 as the working image size for this particular problem.

2.2.2. Segmentation Architectures

The problem was addressed under two different scenarios: single and multi-modalities. As their name suggests, in the first one, only the information from the target modality (DE MRI) was taken into account, while in the second one, Cine MRI was also considered, so the data become of paired nature. For the single-modality approach, the UNet architecture, proposed by Ronneberger et al. [16], was used. Batch normalization layers were added to the building blocks of the network.

For the multi-modality approach, the "DualUNet" architecture, which was investigated in the work of Ouadah et al. [10], was used. This architecture consists of a modified version of the original UNet, but it allows data fusion between the Cine and DE MRI at an intermediate step. Figure 2 shows that one independent UNet-like encoder path is used for each modality in order to extract separate distinct features. The output of the convolution blocks at level I are combined through a fusion block where several operations occur. First, both feature maps are stacked, then fusion features are calculated through a 3D convolutional layer, and the result is taken back to the original feature map dimension via squeeze operation.

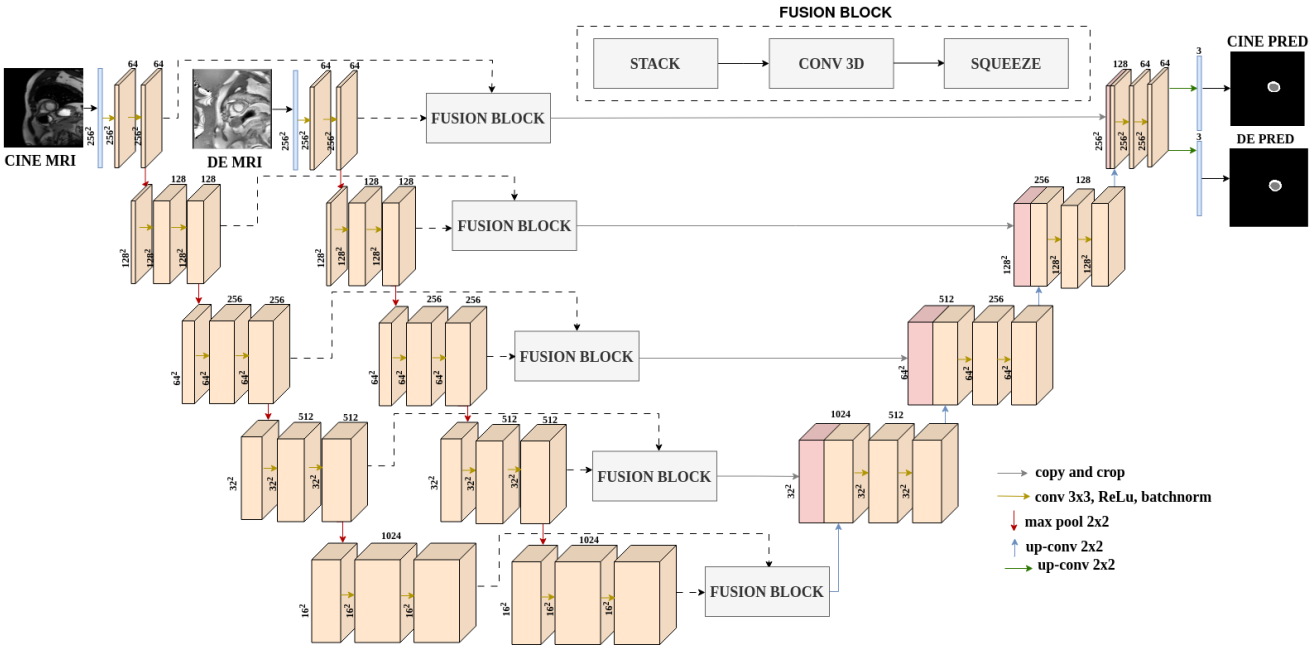

Figure 2. DualUNet architecture for multi-modal cardiac MRI segmentation.

2.2.3. Extraction of Prior Information from EMIDEC Dataset

In this work, a pathology-specific data augmentation algorithm is proposed. Synthetic myocardial infarction cases are generated from images of healthy patients. Afterwards, these newly created images are added to the training set and fed to the segmentation models, with the goal of helping them to learn that the MI scar, if present, is part of the myocardium and should be segmented as such.

The algorithm feeds itself from some prior information in terms of type, location, area, and intensity to generate cases that are as realistic as possible. These priors are extracted from the pathological cases of the EMIDEC dataset such as:

Type: From an algorithmic point of view, the myocardial infarction was constrained to have two types: it either crosses the whole myocardium (transmural MI, called TMI hereinafter) or not (non-transmural MI, called NTMI hereinafter). Figure 3 shows an example of the masks for these two types of MI.

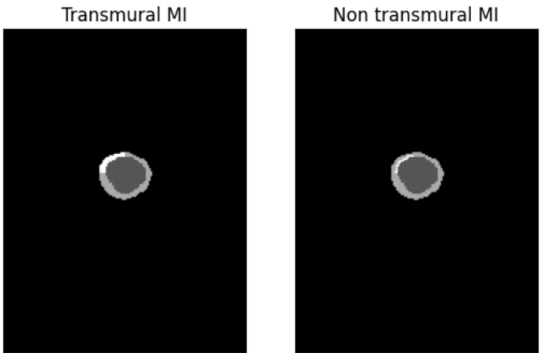

Figure 3. Examples of transmural and non-transmural MI. Dark gray: left ventricle cavity; light gray: myocardium; white: myocardial infarction.

Location: To understand the typical location of MI, all images from the EMIDEC dataset were divided into four equal parts, or "quadrants", corresponding to a modified version of the "Bull's Eye Plot" proposed by Cerqueira et al. [18]. Considering these 4 quadrants for the basal, mid, and apical localizations, although there is a higher prevalence of the MI in the first quadrant at basal and mid slices, we found that the variation is rather subtle, which means that the difference is not big enough to establish a location pattern of the MI structure robustly. Hence, it was decided to keep the location prior to the algorithm as a random variable that can take values between 1 and 4 with equal probability.

Size: The size prior was defined in terms of the percentage of MI scar with respect to the myocardium area, whose mean was 0.208 ± 0.114. In order to account for outliers with bigger scars, the size prior to the MI was defined as a portion of the total myocardium that ranges from 10% to 50%, with an equal probability of selection.

Intensity: Once the myocardium mask is generated using the priors mentioned above, perturbation must be performed on the MRI to generate the synthetic image. The intensity of the MI area was defined as the mean intensity of the left ventricle cavity plus a prior knowledge value. This value is sampled from a generated normal distribution whose mean is the average of the differences between the maximum intensity value of the MI (red contour in Figure 4) minus the mean value of the left ventricle cavity (blue contour in Figure 4) across the entire pathological cases on EMIDEC dataset. The selection of a normal distribution is performed to ensure a consistent MI shape in terms of intensity, and the starting value is selected as the mean intensity of the left ventricle since it was also found that the MI is always brighter.

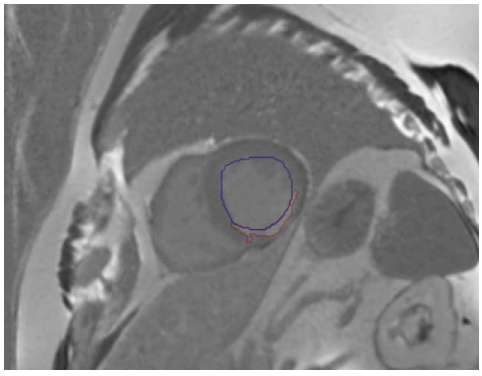

Figure 4. Example of pathological case on EMIDEC dataset [3]. The red and blue contours correspond to the MI and left ventricle cavity, respectively.

2.2.4. Data Augmentation Approach

Once the priors are defined, the data augmentation algorithm consists of the following steps:

Initialization: Given the original image and its corresponding ground truth as inputs, the type and location of MI are defined randomly. The size of the MI corresponds to a percentage sampled randomly from a uniform distribution within the range [0.1,0.5]. The algorithm then calculates the adequate MI size as a multiplication of the total myocardium area by the randomly sampled percentage.

Extraction of contours: As a second step, the contours for the left ventricle cavity and the myocardium are extracted, as shown in image b from Figure 5.

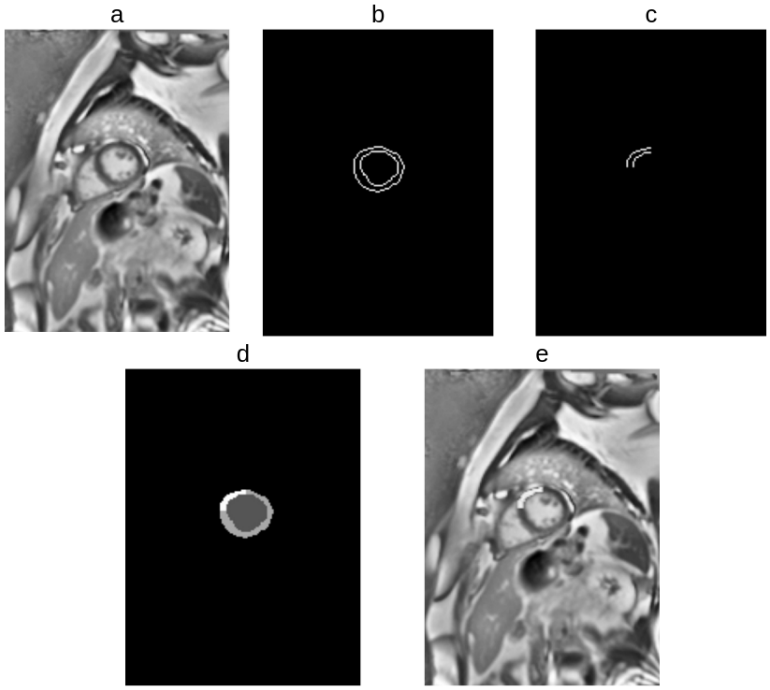

Figure 5. Summary of the data augmentation method for a case of TMI: (**a**) original DE MRI; (**b**) segmentation of myocardial contours; (**c**) selection of working quadrant; (**d**) MI ROI generation; (**e**) MI generation in DE MRI.

Division of contours in quadrants: Once the contours have been extracted successfully, four binary images are generated, which contain the parts of both contours that fall within the range of coordinates of the previously defined quadrants. As a final step, one of the quadrants is selected for the MI generation, and its binary image is kept.

Definition of boundary points: As a fourth step, one random pair of coordinates from the left ventricle cavity contour, located within the selected quadrant, is chosen as the starting point (*start CV*) for generating the MI shape. Afterwards, the left ventricle cavity end point (*end CV*) is chosen as the one whose distance with the *start CV* is the closest to the effective MI size. The starting and end points of the myocardium correspond to the closest points (measured in Euclidean distance) to *start CV* and *end CV*, respectively.

MI contour generation (TMI): An ROI is generated from the boundary points by first painting all pixels belonging to each contour and falling between the extreme points, then closing it through lines using Breseham's line algorithm [19].

MI contour generation (NTMI): The process is quite similar to the one described for TMI, with the difference that the myocardium contour is not painted since the MI falls between the two contours. The intermediate coordinates are determined by a thickness variable that is sampled from a random distribution that ranges between [0.2,0.8]. It cannot be assured that the intermediate points can always be connected through a line since the algorithm selects them randomly in order to generate a variety of MI scar shapes. Therefore, the breadth first search algorithm is used to close the contour and generate

the final ROI [20]. An example of an intermediate-generated path from this algorithm for NTMI is shown in Figure 6.

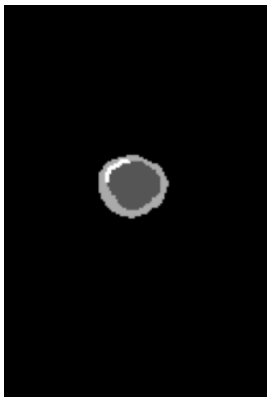

Figure 6. ROI generation of NTMI using breadth first search algorithm.

Creation of MI mask: Once the contour is closed, it is filled, and the final MI is generated, as shown in image d from Figure 5.

Generation of MI perturbation on the MRI: Following the rule described in Section 2.2.4, the MI is generated as the mean value of the left ventricle cavity plus the prior knowledge value. As a final step, some blurring was applied to smooth the intensity variations in the borders of the MI. The blurring was applied on the area delimited by a dilated mask (for one iteration) of the myocardium using a 3×3 kernel. The final result of the algorithm is depicted in image e from Figure 5.

2.3. Experiments

The experimentation protocols for this work were defined as follows:

- **Single modality baseline**, which corresponds to training the modified UNet for the target task.
- **Multi-modality baseline**, which corresponds to training the DualUNet for the target task.
- **Evaluation of the proposed data augmentation algorithm**, which refers to measuring the impact of applying the data augmentation under two different scenarios: random and adaptive, which are explained in detail in the next section.

2.4. Network Training Details

2.4.1. 5-Fold Cross Validation

All the experiments were performed using 5-fold cross validation. Within each fold, the data were split in the following way: 60% (590 images) for training, 20% (197 images) for validation, and 20% (197 images) for testing. The splits were made so that the classes were equally distributed.

2.4.2. Application of the Proposed Augmentation Algorithm

Random augmentation: This strategy refers to randomly enlarging the training dataset at different rates, e.g., 50%, 100%, and 200%, and evaluating its impact on the segmentation performance for both single-modality and multi-modality approaches.

Adaptive augmentation: In this case, the augmentation was applied by obtaining feedback on the model training, e.g., monitoring the cases in which it struggled the most. This criterion of "struggle" was defined as any case for which the Dice score for the segmentation of the myocardium was less than 0.8 during training.

2.4.3. Auxiliary Data Augmentation Techniques

On top of the proposed data augmentation algorithm, some other typical transformations were randomly applied to the data, namely, horizontal and vertical flips, rotations,

random brightness contrast, random gamma (change the image contrast by raising its intensity values to the gamma power), and contrast-limited adaptive histogram equalization (CLAHE) [21] with clip limit equal to 2 and a grid size of (8,8). Structural transformations (flips and rotations) were applied to both the image and its corresponding mask, whereas pixel-intensity-based transformations were only applied to the image.

2.4.4. Optimization

Throughout the experiments, an initial phase of hyperparameter tuning was always performed. The loss function was varied between Dice loss and focal Dice loss (a weighted average between the focal and Dice loss). The selected optimizer was ADAM [22], with a learning rate randomly varying between 10^{-1} and 10^{-5} and a weight decay fixed at 10^{-5}. The batch size varied between 4, 8, and 16. Early stopping was also used, and it was set with a patience of 20 epochs and a maximum number of 200 epochs.

The parameters for which the best results were consistently obtained in the baseline architectures were the following: focal Dice loss, ADAM [22] optimizer with a learning rate of 3^{-4}, weight decay of 10^{-5}, and a batch size of 8. A step scheduler was fixed during all training experiments, which decreased the learning rate every 5 epochs by a multiplicative factor of 0.95.

2.4.5. Post-Processing

Once the segmentation masks were generated, the post-processing for all experiments consisted of two simple steps: extraction of the more significant area component and hole filling.

2.5. Evaluation

The Dice score is one of the most commonly used metrics for evaluating the performance of image segmentation pipelines. It measures the overlap between the predicted segmentation mask and the ground truth. This quantity ranges from 0 (no match) to 1 (perfect match). Some of its main advantages are that it is sensitive to true-positive matches, robust to class imbalance and intuitive to interpret. Considering these benefits, the Dice score is selected as the primary evaluation metric for all the experiments and stages of this work, which includes training optimization and model comparison.

The Hausdorff distance is a standard metric for evaluating image segmentation results. It typically quantifies the dissimilarity between two sets of points or shapes (contours) by measuring the maximum distance between any point in one set and its nearest point in the other. In the context of image segmentation, this means assessing to which extent the predicted mask deviates from the ground truth with respect to shape and spatial arrangement. The smaller the Hausdorff distance, the better the alignment between the prediction and the ground truth. Our work used this metric as an auxiliary local metric to provide a comprehensive overview of the selected pipeline performance in the final stage.

To evaluate the effectiveness of applying our data augmentation algorithm to improve the segmentation of the target structure (myocardium on DE MRI), the Wilcoxon signed-rank test [23] is used. This method was selected since it allows us to test the null hypothesis that two dependent samples, which are not normally distributed (as in our case), come from the same distribution. If rejected, this means there is sufficient statistical evidence to conclude that the samples are different. Therefore, in our case, the algorithm had a positive impact on improving the segmentation performance.

3. Results

3.1. Single Modality

The first set of experiments was performed under a single-modality scenario, i.e., only using image information from DE MRI. The segmentation architecture was the traditional UNet [16], adding batch normalization layers.

The performance of the proposed data augmentation method was assessed in its fixed and adaptive version. Table 1 presents the segmentation results evaluated through the Dice score and its standard deviation.

Table 1. Mean test Dice score for single-modality approach trained with fixed (DA) and adaptive (ADA) data augmentation for the myocardium segmentation. Best result in bold.

UNet	UNet-50% DA	UNet-100% DA	UNet-200% DA	UNet-ADA
0.834 ± 0.056	0.838 ± 0.065	0.844 ± 0.058	$\mathbf{0.845 \pm 0.052}$	0.843 ± 0.051

The results show that the augmentation method increases the Dice score for the myocardium as more training data become available, finding its peak value at 200% (when two MI infarction cases are created from each normal image), and its standard deviation also decreases. For larger augmentation rates (>200% DA), the segmentation performance was either maintained or decreased.

It is worth mentioning that the adaptive data augmentation also achieves good results that are even better than its fixed counterpart (as in the case of 50% DA), and it uses much less data since the number of cases to augment based on the model feedback varies from 20 to 25.

When doing a case-by-case analysis, it was found that most of the time, the algorithm was able to notoriously improve the segmentation on the structure of interest (the myocardium), as shown in the top part of Figure 7. However, the segmentation performance dropped slightly in a few cases, as shown in the bottom part of Figure 7.

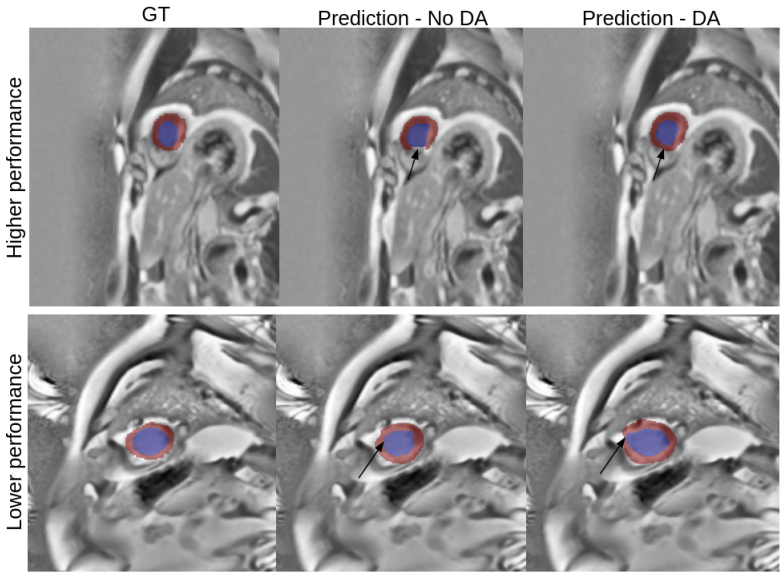

Figure 7. Example of a higher (**top**) and lower (**bottom**) segmentation performance for the single-modality model after applying data augmentation (indicated by the black arrows). From left to right: ground truth, prediction of model without data augmentation, prediction of model with best data augmentation rate (200%). The left ventricle cavity and myocardium are presented in blue and red, respectively.

3.2. Multi Modality

The same group of experiments was also performed under the multi-modality approach, which consisted of using DualUNet as a segmentation architecture, in which data from the Cine and DE modalities are fused at an intermediate step.

Table 2 presents the segmentation results for the test set, assessed through the mean Dice score and its corresponding standard deviation.

Table 2. Mean test Dice score for single-modality approach trained with fixed (DA-%) and adaptive (ADA) data augmentation for the myocardium segmentation. Best result in bold.

DualUNet	DualUNet-50% DA	DualUNet-100% DA	DualUNet-200% DA	DualUNet-ADA
0.845 ± 0.044	0.847 ± 0.046	**0.856 ± 0.040**	0.852 ± 0.042	0.853 ± 0.039

It can be noted that, similarly to the single-modality approach, the segmentation model manages to consistently increase the Dice score and lower its standard deviation as more synthetic data become available for training. However, it reaches a plateau at 100%, after which the Dice score decreases and the standard deviation increases. For larger augmentation rates (>200% DA), the segmentation performance was either maintained or decreased.

In the adaptive augmentation case, the obtained results are also good considering that the amount of synthetic data generated is much less when compared to any of the fixed augmentation scenarios. However, the best performance in the test set is obtained when the DualUNet is trained on a dataset with 100% data augmentation.

Throughout a case-specific analysis, it was found that, in 89 out of 124 cases (71.77%), the Dice score was higher after applying the data augmentation algorithm. One example of this higher performance is presented in Figure 8. When focusing on the areas indicated by the arrows, it can be noted that the data augmentation algorithm indeed makes the model learn from different myocardial infarction shapes, thus making it capable of delineating the myocardium contour much more thoroughly, covering areas corresponding to myocardial infarction that were missed before applying the algorithm.

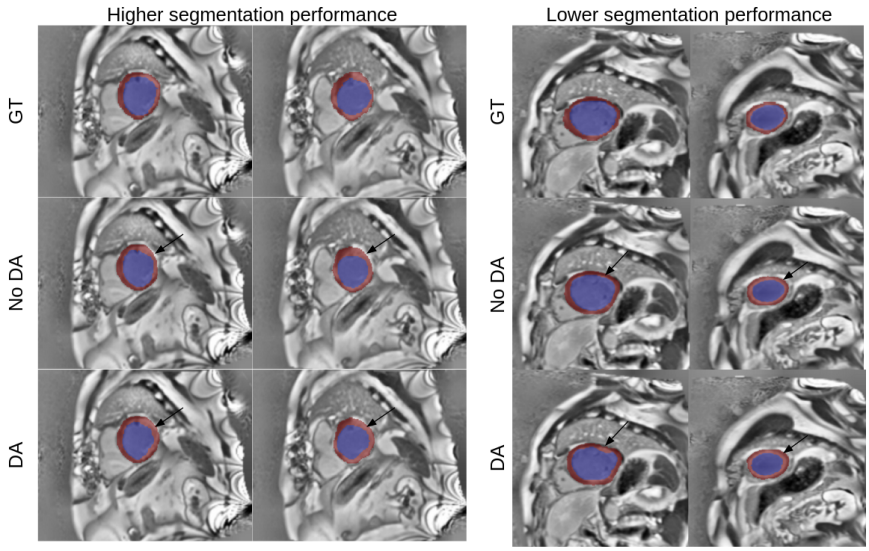

Figure 8. Examples of higher (**left**) and lower (**right**) segmentation performance for the multi-modality model after applying data augmentation (differences shown by the black arrows). From top to bottom: ground truth (GT), prediction of model without data augmentation (No DA), prediction of model with the best data augmentation rate (DA: 100%). The left ventricle cavity and myocardium are presented in blue and red, respectively.

In cases where the Dice was lower, the visual difference was generally difficult to spot. Figure 8 presents examples of such cases, and it can be observed that in the area selected by the arrow, there is a slightly worse segmentation of the myocardium as it becomes thicker (taking some area that should have been delimited as the left ventricle cavity).

3.3. Single vs. Multi-Modality

Based on the information presented in Tables 1 and 2, it can be seen that the performance of the best models for the myocardium segmentation under single and multi-

modality approaches are 0.845 ± 0.052 and 0.856 ± 0.040, respectively. Under the Wilcoxon test, the null hypothesis was rejected ($p_{value} < 10^{-5}$), thus implying that crossing information between Cine and DE modalities had a positive impact on the segmentation of DE myocardium. Figure 9 shows an example of the visual performance under the two scenarios, and it is observed that in the multi-modality approach, a more consistent segmentation of the myocardium is produced. Thus, the proposed pipeline (DualUNet trained at 100% enlarged dataset, hereinafter called "DualUNet100"), is deemed to be the overall best-performing method for segmenting the myocardium in DE MRI.

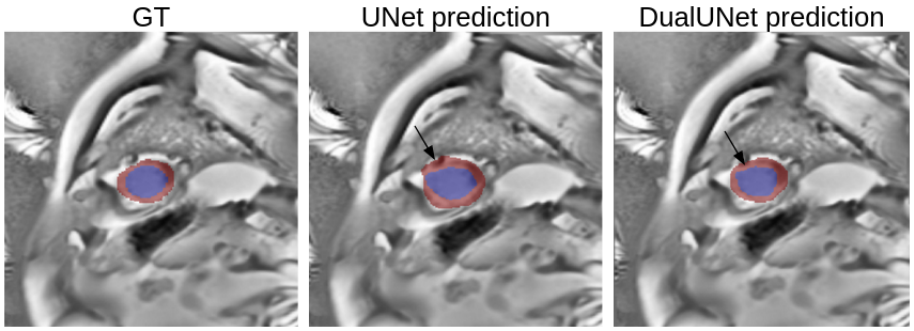

Figure 9. Example of segmentation performance under single and multi-modality approaches (highlighted by the black arrows). The left ventricle cavity and myocardium are in blue and red, respectively.

3.4. Best Model Further Results

The final experiment in this stage consisted of comparing DualUNet vs. DualUNet100. Table 3 shows the segmentation results for these two models regarding Dice score and Hausdorff distance.

Table 3. Mean Dice score and Hausdorff distance of the best model (DualUNet) after applying the best data augmentation strategy (DualUNet100). Best results in bold.

DualUNet DSC	DualUNet (100% DA) DSC	DualUNet HD (mm)		DualUNet (100% DA) HD (mm)	
0.845 ± 0.044	0.856 ± 0.040	Endocardium	**5.73 ± 2.12**	Endocardium	5.77 ± 2.23
		Epicardium	6.08 ± 1.75	Epicardium	**5.90 ± 1.93**

It can be seen that the myocardium segmentation Dice score is consistently higher after applying the data augmentation strategy, and the standard deviation is also less. This difference was proven to be statistically significant under a Wilcoxon test with $p_{value} < 10^{-5}$. Regarding the Hausdorff distance since the region of interest is bordered by two contours, namely, endocardium and epicardium, they both have to be taken into account. It can be observed that the Hausdorff distance is slightly higher for the endocardium and slightly lower for the epicardium. However, the increase is lower than the decrease. This might indicate that the segmentation performance at the contour level is approximately maintained.

In a final stage, the segmentation performance exclusively on MI cases was analyzed. If the proposed data augmentation worked as expected, it should reduce the outliers for the cases classified as myocardial infarction.

Figure 10 shows the distribution of the Dice score for the myocardium segmentation before and after applying the data augmentation algorithm, with the DualUNet architecture trained with 100% data augmentation. As expected, the overall distribution of the metric went up, and two of the three outliers were eliminated. Although the worst one was not removed, its score improved significantly, which gives evidence of an increase in the robustness of the model, better delineating the myocardium despite the presence of MI.

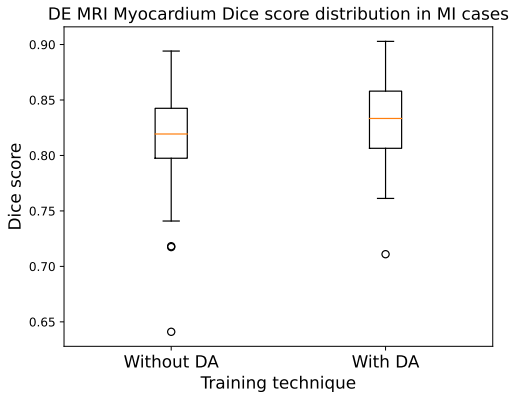

Figure 10. Box-and-whisker plot of the Dice score distribution for the myocardium on myocardial infarction cases before and after applying data augmentation. The white circles correspond to the outliers.

4. Discussion

In this study, we proposed an image processing-based data augmentation algorithm dedicated to the segmentation of the myocardium on DE MRI. The method can create synthetic infarction cases using prior information extracted from an external dataset. The algorithm works under two possible scenarios: fixed and adaptive. In the first one, the training dataset is enlarged at a given amount, whereas in the second, feedback from the model is obtained while training.

The first set of experiments was performed under the single-modality approach. This was performed using a traditional UNet with batch normalization layers as a segmentation model. Afterwards, the data augmentation algorithm was applied at fixed rates, namely, 50%, 100%, and 200%. It could be noticed that the segmentation performance on the myocardium was higher as more artificial data were available. The best result is obtained at a rate of 200%. The adaptive data augmentation strategy also offered an increase in the Dice score of the myocardium, but this was less than the increase generated with the 200% strategy.

Similarly, the results of the multi-modality approach show that the segmentation is consistently improved when increasing the percentage of data augmentation up to 100%. Once again, the results were also improved with the adaptive data augmentation strategy, but to a smaller extent when comparing it to the 100% dataset enlargement.

In both single- and multi-modality scenarios, it was observed that the models have a good generalization capability. Furthermore, it was demonstrated that adding the data augmentation helps to better delineate the myocardium in two ways: it either completes missed contours (top part of Figure 7) or includes MI areas that were previously missed (left part of Figure 8). Although the improvement in Dice score might seem small when seen in the general view, the produced segmentation masks are much more consistent and clinically relevant. It was also shown in Figure 10 that, when focused only on cases related to MI (the target group for our study), the quantitative improvement is more significant, and most of the outliers (corresponding to the poorly segmented cases) are eliminated. These improvements also prove that when models are fed with enough synthetic MI, they can better segment the myocardium.

There were a few cases in which the segmentation performance was lower after applying the data augmentation (bottom and right parts of Figures 7 and 8, respectively). In these cases, the myocardium area becomes thicker than it should, which means that left ventricle cavity pixels are mistakenly classified as part of the myocardium. Further investigation should be performed in order to limit these minority cases.

When comparing the single- vs. multi-modality approach, it was found that the latter consistently outperformed the first. This supports the hypothesis that the information from Cine MRI is indeed helpful in improving the segmentation performance on DE MRI.

One of the limitations of the data augmentation method is the intensity pattern for synthetic MI generation. In particular, our current strategy does not take into account the generation of MI with persistent microvascular obstruction. Therefore, future work should be focused on investigating additional strategies that would allow us to create synthetic images that represent this particular condition in MI cases and therefore make the models more robust. Moreover, our pathology-specific data augmentation could be applied to the automatic segmentation of DE MRI concerning other diseases where the DE MRI is abnormal, such as myocarditis or hypertrophic cardiomyopathy.

5. Conclusions

In this work, we proposed a pathology data augmentation algorithm where we create synthetic myocardial infarction shapes to allow deep learning models to further learn from pathological cases and improve the segmentation of the myocardium on delayed-enhancement MRI. We proved that the method made the overall pipeline more robust in both single- and multi-modality settings, obtaining higher Dice scores with less variability and more plausible segmentation masks, where the myocardium is better delineated even in the presence of myocardial infarction.

Author Contributions: Conceptualization, G.M.-R., C.O., A.H., A.L. and S.L.; methodology, G.M.-R., A.L. and S.L.; software, G.M.-R.; validation, G.M.-R. and A.L.; formal analysis, G.M.-R.; investigation, G.M.-R., C.O. and A.H.; resources, A.L.; data curation, C.O. and A.H.; writing—original draft preparation, G.M.-R.; writing—review and editing, A.L. and S.L.; visualization, G.M.-R., A.L. and S.L.; supervision, A.L. and S.L.; project administration, A.L. and S.L.; funding acquisition, A.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Belonging to the University Hospital of Dijon (France), data cannot be shared publicly without the permission of this institution.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

Cine MRI	kinetic MRI
DE MRI	delayed-enhancement MRI
MI	myocardial infarction
TMI	transmural myocardial infarction
NTMI	non-transmural myocardial infarction
BFS	breadth first search
CLAHE	contrast-limited adaptive histogram equalization
MVO	microvascular obstruction
HD	Hausdorff distance
DSC	Dice score coefficient
CV	left ventricle cavity
MYO	myocardium
CVD	cardiovascular disease
CHD	coronary heart disease

References

1. Mendis, S.; Thygesen, K.; Kuulasmaa, K.; Giampaoli, S.; Mähönen, M.; Ngu Blackett, K.; Lisheng, L.; Writing group on behalf of the participating experts of the WHO consultation for revision of WHO definition of myocardial infarction. World Health Organization definition of myocardial infarction: 2008–09 revision. *Int. J. Epidemiol.* **2011**, *40*, 139–146. [CrossRef] [PubMed]
2. Thygesen, K.; Alpert, J.S.; Jaffe, A.S.; Simoons, M.L.; Chaitman, B.R.; White, H.D. Third universal definition of myocardial infarction. *Circulation* **2012**, *126*, 2020–2035. [CrossRef] [PubMed]
3. Lalande, A.; Chen, Z.; Decourselle, T.; Qayyum, A.; Pommier, T.; Lorgis, L.; de la Rosa, E.; Cochet, A.; Cottin, Y.; Ginhac, D.; et al. Emidec: A database usable for the automatic evaluation of myocardial infarction from delayed-enhancement cardiac MRI. *Data* **2020**, *5*, 89. [CrossRef]

4. Lalande, A.; Chen, Z.; Pommier, T.; Decourselle, T.; Qayyum, A.; Salomon, M.; Ginhac, D.; Skandarani, Y.; Boucher, A.; Brahim, K.; et al. Deep learning methods for automatic evaluation of delayed enhancement-MRI. The results of the EMIDEC challenge. *Med. Image Anal.* **2022**, *79*, 102428. [CrossRef] [PubMed]
5. Thomson, L.E.; Kim, R.J.; Judd, R.M. Magnetic resonance imaging for the assessment of myocardial viability. *J. Magn. Reson. Imaging Off. J. Int. Soc. Magn. Reson. Med.* **2004**, *19*, 771–788. [CrossRef]
6. Dikici, E.; O'Donnell, T.; Setser, R.; White, R.D. Quantification of delayed enhancement MR images. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2004: 7th International Conference, Saint-Malo, France, September 26–29. 2004. Proceedings, Part I 7*; Springer International Publishing: Cham, Switzerland, 2004; pp. 250–257.
7. Ciofolo, C.; Fradkin, M.; Mory, B.; Hautvast, G.; Breeuwer, M. Automatic myocardium segmentation in late-enhancement MRI. In Proceedings of the 2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Paris, France, 14–17 May 2008; pp. 225–228.
8. Huellebrand, M.; Ivantsits, M.; Zhang, H.; Kohlmann, P.; Kuhnigk, J.M.; Kuehne, T.; Schönberg, S.; Hennemuth, A. Comparison of a hybrid mixture model and a cnn for the segmentation of myocardial pathologies in delayed enhancement MRI. In *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges: 11th International Workshop, STACOM 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers 11*; Springer International Publishing: Cham, Switzerland, 2021; pp. 319–327.
9. Zhou, T.; Ruan, S.; Canu, S. A review: Deep learning for medical image segmentation using multi-modality fusion. *Array* **2019**, *3*, 100004. [CrossRef]
10. Ouadah, C.; Hadadi, A.; Lalande, A.; Leclerc, S. Comparison of CNN Fusion Strategies for Left Ventricle Segmentation from Multi-modal MRI. In *International Conference on Functional Imaging and Modeling of the Heart*; Springer Nature: Cham, Switzerland, 2023; pp. 265–273.
11. Xue, Y.; Farhat, F.G.; Boukrina, O.; Barrett, A.; Binder, J.R.; Roshan, U.W.; Graves, W.W. A multi-path 2.5 dimensional convolutional neural network system for segmenting stroke lesions in brain MRI images. *Neuroimage Clin.* **2020**, *25*, 102118. [CrossRef]
12. Chen, C.; Qin, C.; Ouyang, C.; Li, Z.; Wang, S.; Qiu, H.; Chen, L.; Tarroni, G.; Bai, W.; Rueckert, D. Enhancing MR image segmentation with realistic adversarial data augmentation. *Med. Image Anal.* **2022**, *82*, 102597. [CrossRef] [PubMed]
13. Lin, A.; Wu, J.; Yang, X. A data augmentation approach to train fully convolutional networks for left ventricle segmentation. *Magn. Reson. Imaging* **2020**, *66*, 152–164. [CrossRef] [PubMed]
14. Skandarani, Y.; Painchaud, N.; Jodoin, P.M.; Lalande, A. On the effectiveness of GAN generated cardiac MRIs for segmentation. In Proceedings of the Medical Imaging with Deep Learning Conference, Montreal, QC, Canada, 6–8 July 2020.
15. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. ACM* **2020**, *63*, 139–144. [CrossRef]
16. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9. 2015, Proceedings, Part III 18*; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.
17. Rukundo, O. Effects of image size on deep learning. *Electronics* **2023**, *12*, 985. [CrossRef]
18. American Heart Association Writing Group on Myocardial Segmentation and Registration for Cardiac Imaging; Cerqueira, M.D.; Weissman, N.J.; Dilsizian, V.; Jacobs, A.K.; Kaul, S.; Laskey, W.K.; Pennell, D.J.; Rumberger, J.A.; Ryan, T.; et al. Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart: A statement for healthcare professionals from the Cardiac Imaging Committee of the Council on Clinical Cardiology of the American Heart Association. *Circulation* **2002**, *105*, 539–542.
19. Bresenham, J.E. Algorithm for computer control of a digital plotter. *IBM Syst. J.* **1965**, *4*, 25–30. [CrossRef]
20. Cormen, T.H.; Leiserson, C.E.; Rivest, R.L.; Stein, C. *Introduction to Algorithms*; MIT Press: Cambridge, MA, USA, 2022; Chapter 20; pp. 554–562.
21. Pizer, S.M.; Amburn, E.P.; Austin, J.D.; Cromartie, R.; Geselowitz, A.; Greer, T.; ter Haar Romeny, B.; Zimmerman, J.B.; Zuiderveld, K. Adaptive histogram equalization and its variations. *Comput. Vision, Graph. Image Process.* **1987**, *39*, 355–368. [CrossRef]
22. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
23. Wilcoxon, F. *Individual Comparisons by Ranking Methods*; Springer: Berlin/Heidelberg, Germany, 1992; pp. 191–195.