

## Article

# Deep Neural Network for Cardiac Magnetic Resonance Image Segmentation

David Chen <sup>1</sup>, Huzefa Bhopalwala <sup>2</sup>, Nakeya Dewaswala <sup>2</sup>, Shivaram P. Arunachalam <sup>3,4</sup>, Moein Enayati <sup>5</sup>, Nasibeh Zanjirani Farahani <sup>2</sup>, Kalyan Pasupathy <sup>6</sup>, Sravani Lokineni <sup>2</sup>, J. Martijn Bos <sup>2</sup>, Peter A. Noseworthy <sup>2</sup>, Reza Arsanjani <sup>7</sup>, Bradley J. Erickson <sup>4</sup>, Jeffrey B. Geske <sup>2</sup>, Michael J. Ackerman <sup>2,8,9</sup>, Philip A. Araoz <sup>4</sup> and Adelaide M. Arruda-Olson <sup>2,\*</sup>

- <sup>1</sup> Department of Cardiovascular Surgery, Cleveland Clinic, Cleveland, OH 44195, USA; dchen46@gmail.com  
<sup>2</sup> Department of Cardiovascular Medicine, Mayo Clinic, Rochester, MN 55902, USA; bhopalwalahuzuzaifa@gmail.com (H.B.); nakeya52@gmail.com (N.D.); zanjiranifarahani.nasibeh@mayo.edu (N.Z.F.); sravani.lokineni@gmail.com (S.L.); bos.martijn@mayo.edu (J.M.B.); noseworthy.peter@mayo.edu (P.A.N.); geske.jeffrey@mayo.edu (J.B.G.); ackerman.michael@mayo.edu (M.J.A.)  
<sup>3</sup> Department of Internal Medicine, Mayo Clinic, Rochester, MN 55902, USA; poigaiarunachalam.shivaram@mayo.edu  
<sup>4</sup> Department of Radiology, Mayo Clinic, Rochester, MN 55902, USA; bje@mayo.edu (B.J.E.); paraoz@mayo.edu (P.A.A.)  
<sup>5</sup> Center for the Science of Health Care Delivery, Mayo Clinic, Rochester, MN 55902, USA; enayati.moein@mayo.edu  
<sup>6</sup> Biomedical and Health Information Sciences Department, University of Illinois, Chicago, IL 60612, USA; kap@uic.edu  
<sup>7</sup> Department of Cardiovascular Medicine, Mayo Clinic, Phoenix, AZ 85054, USA; arsanjani.reza@mayo.edu  
<sup>8</sup> Department of Pediatric and Adolescent Medicine, Mayo Clinic, Rochester, MN 55902, USA  
<sup>9</sup> Department of Molecular Pharmacology & Experimental Therapeutics, Mayo Clinic, Rochester, MN 55902, USA  
\* Correspondence: arrudaolson.adelaide@mayo.edu



**Citation:** Chen, D.; Bhopalwala, H.; Dewaswala, N.; Arunachalam, S.P.; Enayati, M.; Farahani, N.Z.; Pasupathy, K.; Lokineni, S.; Bos, J.M.; Noseworthy, P.A.; et al. Deep Neural Network for Cardiac Magnetic Resonance Image Segmentation. *J. Imaging* **2022**, *8*, 149. <https://doi.org/10.3390/jimaging8050149>

Academic Editor: Marcin Balcerzyk

Received: 10 March 2022

Accepted: 6 May 2022

Published: 23 May 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** The analysis and interpretation of cardiac magnetic resonance (CMR) images are often time-consuming. The automated segmentation of cardiac structures can reduce the time required for image analysis. Spatial similarities between different CMR image types were leveraged to jointly segment multiple sequences using a segmentation model termed a multi-image type UNet (MI-UNet). This model was developed from 72 exams (46% female, mean age  $63 \pm 11$  years) performed on patients with hypertrophic cardiomyopathy. The MI-UNet for steady-state free precession (SSFP) images achieved a superior Dice similarity coefficient (DSC) of  $0.92 \pm 0.06$  compared to  $0.87 \pm 0.08$  for a single-image type UNet ( $p < 0.001$ ). The MI-UNet for late gadolinium enhancement (LGE) images also had a superior DSC of  $0.86 \pm 0.11$  compared to  $0.78 \pm 0.11$  for a single-image type UNet ( $p = 0.001$ ). The difference across image types was most evident for the left ventricular myocardium in SSFP images and for both the left ventricular cavity and the left ventricular myocardium in LGE images. For the right ventricle, there were no differences in DCS when comparing the MI-UNet with single-image type UNets. The joint segmentation of multiple image types increases segmentation accuracy for CMR images of the left ventricle compared to single-image models. In clinical practice, the MI-UNet model may expedite the analysis and interpretation of CMR images of multiple types.

**Keywords:** cardiac magnetic resonance imaging; deep learning; image segmentation; hypertrophic cardiomyopathy

## 1. Introduction

Cardiac magnetic resonance (CMR) imaging is well established for the diagnosis and prognostication of many cardiac diseases [1,2]. A valuable diagnostic aspect of CMR imaging is the ability to acquire image types by the variation of sequence parameters

sensitive to different pathologic changes in tissue microstructure [2]. Steady-state free precession (SSFP) sequences produce images with high contrast between the myocardium and the blood pool; this is useful for the characterization of functional and morphological findings, whereas late gadolinium enhancement (LGE) highlights tissue characteristics useful for the identification of myocardial pathologies. In clinical settings, both SSFP and LGE are acquired and utilized for the interpretation of CMR imaging of patients with hypertrophic cardiomyopathy (HCM) [3].

The analysis and interpretation of CMR images are often time-consuming. The automated segmentation of cardiac structures, such as the left ventricular wall, can greatly reduce the amount of time necessary for image analysis. A variety of methods have been proposed to automate this process, and these can be considered either non-deep-learning or deep-learning methods. Non-deep-learning methods leverage prior knowledge of the characteristics of images to determine the boundaries of anatomic structures [4–8]. However, most of these strictly feature-engineered methods have fallen out of favor due to the flexibility and robustness of deep learning [9].

Although deep-learning methods are expensive to train, the execution of a model can be performed without significant computational resources compared to iterative segmentation methods that must be performed for each new image. The most recent advances in CMR segmentation utilize combinations of prior knowledge or models and deep learning [10–12]. Most of these methods have focused on a single CMR image type, primarily cine SSFP, that provides a clear delineation between the blood pool and the myocardium, a high spatial resolution, and a high signal-to-noise ratio. Other methods, such as multi-sensor fusion, yield only a single mask that may be inappropriate for multi-image type segmentation where misregistration and sequence differences may require significantly different masks [13]. The multi-input fusion network (MIFNet) model is based on multi-scale input and feature fusion and showed reasonable performance with limited parameters and less training time; however, it used only SSFP sequences for segmenting CMR images [14]. The focal loss constrained residual network (FR-Net) model demonstrated improved performance in the CMR segmentation of two-dimensional (2D) short-axis CMR images; it trains the model with focal loss and dice loss alternatively [15]. A model based on an ensemble of neural networks was reported that selects the most optimal segmentation based on the predicted segmentation accuracy on the fly; it can be adapted to CMR, echocardiography, and cardiac computed tomography segmentation [16]. The BLU-Net model, which is a nested U-shape network with compressed dense blocks (CDBlocks) and dilated convolution, shows good performance on CMR segmentation based on a larger receptive field by preserving spatial resolution and feature information [17]. Although these novel methods report reasonable performance, they use one specific CMR imaging sequence dataset for model training. The motivation for multi-image type segmentation is to leverage complementary information from SSFP and LGE to provide a deep-learning solution for improving CMR imaging segmentation that saves time spent on image analysis.

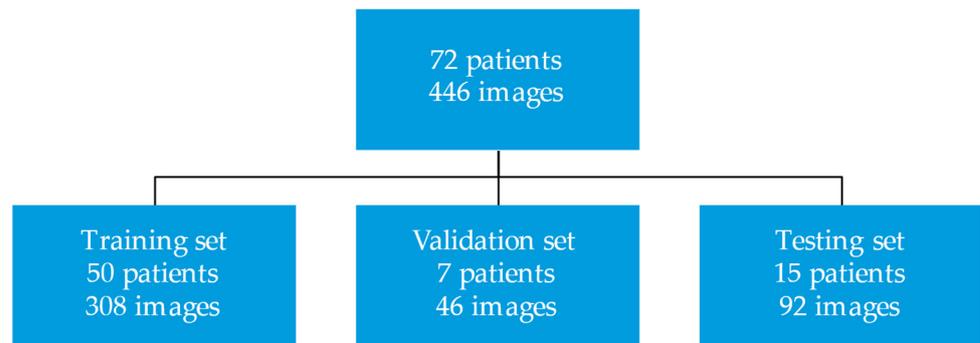
Accordingly, the goal of this study was to create a deep-learning solution that learns to jointly segment CMR images of different types to minimize the image analysis workload of radiologists. The hypothesis tested was that the joint segmentation of SSFP and LGE can provide a deep neural network architecture. We show a proof of concept of this novel model in a cohort of patients with HCM, for which CMR is routinely used both for diagnosis and for sudden cardiac death risk stratification [18–20].

## 2. Materials and Methods

### 2.1. Study Design

This study was approved by the Mayo Clinic institutional review board. All patients provided informed consent for the research use of their health data. This study created a deep learning model to segment CMR SSFP and LGE images. We randomly selected 72 patients who underwent contrast CMR exams for the evaluation of HCM between 2009 and 2016 at Mayo Clinic Rochester, MN. The CMR images of these patients with HCM were

divided into training ( $n = 50$  patients; 308 images), validation ( $n = 7$  patients; 46 images), and test sets ( $n = 15$  patients; 92 images). Figure 1 displays the study design.



**Figure 1.** Study design.

Standard clinical cardiac exams were conducted with short- and long-axis cine SSFP and short- and long-axis LGE in a 1.5T General Electric Optima MR450w (General Electric, Boston, MA, USA). For the segmentation, selected ventricular short-axis slices between the first appearance of the aortic valve and the disappearance of the right ventricle from the end-systolic phase of the cine SSFP and the LGE images were acquired in systole for each study. All images were extracted from the Mayo Clinic image archiving and communication system in Digital Imaging and Communications in Medicine (DICOM) format. A physician manually segmented the images for the left ventricular cavity (LVC), the left ventricular myocardium (LVM), and the right ventricle (RV), including both the myocardial wall and cavity using Radiology Informatics Laboratory Contour (RIL-Contours), which is a freely available annotation tool developed at Mayo Clinic [21]. The physician was instructed to segment ventricular slices that included all 3 structures and in which the left ventricular myocardium was not bisected by the aortic valve. The papillary muscle was included in the LVC segment. The cine SSFP images were segmented to match the cardiac frame during which the LGE images were acquired.

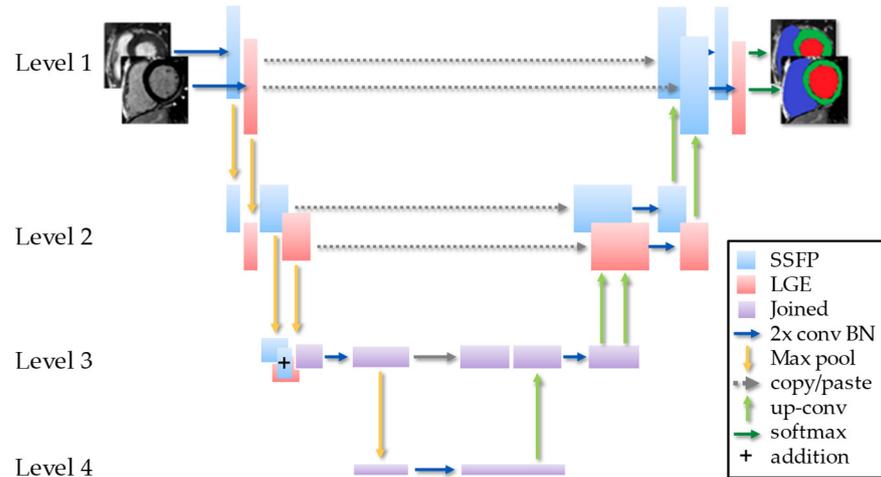
## 2.2. Preprocessing

DICOM images and annotations were converted into numpy arrays using pyDicom (version 1.2), a python package used for manipulating DICOM files [22]. All images were zero-padded to a square matrix, cropped by 50% to focus on a region of interest containing the heart, and under sampled following filtering with a low-pass filter to a  $128 \times 128$  matrix. Data augmentation with rotation between  $-60$  and  $60$  degrees around the center of the image was performed to increase the training sample size and increase the likelihood of training a generalizable model.

## 2.3. Model

A multi-image type bidimensional UNet (MI-UNet) was developed to jointly learn features applicable to the segmentation of two types of CMR images: cine SSFP and LGE. This was based on the observation that multi-task learning problems have often shown improved results compared to learning single tasks individually. The network was derived from the conventional bidimensional UNet developed by Ronneberger et al. with back-to-back cascading encoder-decoder networks with skip connections [23]. In the standard UNet, each "level" consisted of two convolutional, batch normalization, and rectified linear unit activation layers in series. The encoding network used max pooling between levels to minimize feature map size, while the decoding network used deconvolution to recover the dimensions of the original input. Skip connections were also used between equivalent levels in the encoding and decoding networks.

In this work, rather than two independent encoder-decoder networks for each image type, the hidden layers between the two networks were additively joined and then shared at the bottom of the “U”, much like a Siamese network, as shown in Figure 2.



**Figure 2.** Schematic of the MI-UNet with 2 convolutions and batch normalization (2x conv BN).

The shared hidden layer that produced the optimal number of shared layers was identified. When the hidden layers were shared prior to the deepest layer, the subsequent layers acted as a single-input UNet. The final results of the two streams were not fused. Rather, the two streams shared a common deep embedding with separate encoders and decoders. This was to promote shared embedding between two spatially co-located imaging sources.

The conventional UNet utilizes a combination pixel-wise binary cross entropy (BCE) and Dice similarity coefficient (DSC) loss with empirically determined weighting for each [23,24] by Equation (1):

$$L_{seg} = \lambda_1 L_{DSC} + (1 - \lambda_1) L_{BCE} \tag{1}$$

here  $\lambda_1$  is the weight balancing between DSC loss and BCE loss. The DSC loss is determined by Equation (2):

$$L_{DSC}(x, \hat{x}) = 1 - \frac{2 \sum_i x_i \hat{x}_i}{\sum_i x_i^2 + \sum_i \hat{x}_i^2} \tag{2}$$

whereas the BCE loss is determined by Equation (3):

$$L_{BCE}(x, \hat{x}) = -\frac{1}{N} \sum_{i=1}^N \hat{x}_i \cdot \log(p(x_i)) + (1 - \hat{x}_i) \cdot \log(1 - p(x_i)) \tag{3}$$

where  $x$  is the predicted mask,  $\hat{x}$  is the label,  $i$  is index for pixel, and  $p(x)$  is the probability of the label. In the jointly trained model, the loss function needed to be applied to both the SSFP and LGE segments, and this yields Equation (4):

$$L_{seg} = L_{Cine_{seg}} + L_{DE_{seg}} \tag{4}$$

where  $L_{Cine_{seg}}$  and  $L_{DE_{seg}}$  represent the loss returned by each respective branch.

Given that we were learning to segment two anatomically similar images, we used this knowledge to further constrain the segments. First, a total variation (TV) penalty term was added for the gradient of each anatomic segment. This was accomplished a priori such that each segment was locally continuous to minimize the length of the contours around

each segment. The total variation (TV) term is well approximated by an L1 norm, as follows in Equation (5) [25]:

$$TV(x) = \frac{1}{N} \sum_{i=1}^N |\Delta x_i| \quad (5)$$

Second, we assumed that there should be minimum spatial differences between the segments of the two images. This contrast was instituted using a BCE loss between the segments of the two images. The overall loss function is as follows in Equation (6):

$$L_{seg} = \lambda_1 L_{DSC} + (1 - \lambda_1) L_{BCE} + \lambda_{TV} (TV(x_{Cine}) + TV(x_{DE})) + \lambda_{Sp} L_{BCE}(x_{Cine}, x_{DE}) \quad (6)$$

where  $\lambda_{TV}$  and  $\lambda_{Sp}$  are the weights for TV and spatial BCE loss, respectively.

#### 2.4. Training

All models were trained using the training dataset, while hyperparameters were tuned using the validation dataset, and the final metrics were computed based on the test dataset. Model parameters were randomly initialized from a normal distribution, and Adam optimizer was used [26]. Models were trained using data augmentation with rotation between  $-60$  and  $60$  degrees around the center of the image. For each model, the optimized hyperparameters (batch size: 2–24, learning rate:  $10^{-6}$ – $10^{-1}$ , and batch normalization) were found using a grid search maximizing DSC on the validation dataset. Specifically, for the MI-UNet model, the appropriateness of various a priori constraints (smoothness and similarity) was studied by varying the weights ( $\lambda_{TV}$  and  $\lambda_{Sp}$ ) assigned to each respective constraint. Models were implemented in python 3.6 using pyTorch on 3 Nvidia V100 32GB GPUs [27].

#### 2.5. Evaluation

To evaluate performance, the proposed MI-UNet model was compared to three UNet-based architectures, namely the conventional single-image UNet (one model for each image type) and the transfer-learned UNet pretrained on one image type. The single-image UNet was trained on both cine SSFP and LGE images, and the transfer-learned UNet consisted of a conventional UNet that was first pre-trained using the corresponding imaging type. The performance of each segmentation model was evaluated using the standard Dice score (DSC) [24,28]. We used paired two-tailed t-tests at a 5% significance level to evaluate the statistical significance of the DSC of each model compared to a baseline single-contrast UNet. For the MI-UNet, the frame of the SSFP images matched the acquisition window of the LGE images.

### 3. Results

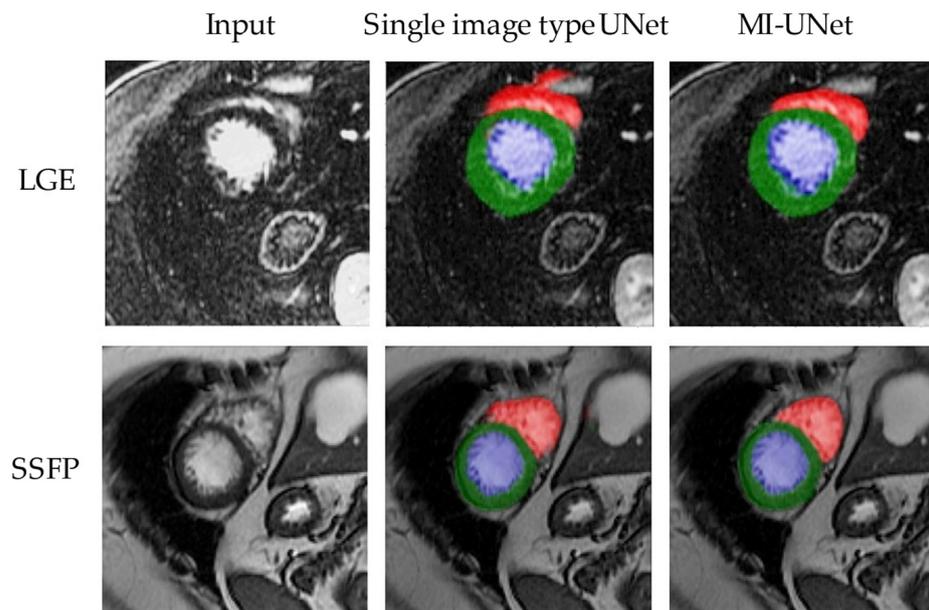
CMR exams performed on 72 patients with a mean age of  $63 \pm 11$  years (46% women) were included in this study. All MI-UNet models that joined encoding hidden layers provided similar DSC scores compared to the single-image type UNet for SSFP or LGE (Table 1).

**Table 1.** Dice Similarity Coefficient for the MI-UNet compared to a single-image type UNet.

	SSFP	<i>p</i> Value*	LGE	<i>p</i> Value *
Single-image type UNet	$0.87 \pm 0.08$		$0.78 \pm 0.12$	
MI-UNet	$0.842 \pm 0.132$	0.126	$0.788 \pm 0.141$	0.6393

SSFP—steady-state free precession; LGE—late gadolinium enhancement. \* *p*-values refer to comparison with single-image type UNets.

Figure 3 shows an example of segments obtained from the MI-UNet model compared with single-image type UNets.



**Figure 3.** Example outputs of the segmentation models. Segments obtained from the proposed model compared with single-image type. In this figure, blue represents the left ventricular cavity, green the left ventricular wall, and red the right ventricle.

The MI-UNet without the use of a priori constraints ( $\lambda_{TV} = 0$  and  $\lambda_{Sp} = 0$ ) had a DSC of  $0.87 \pm 0.11$  for SSFP ( $p = 0.73$ ) and a DSC of  $0.82 \pm 0.16$  for LGE ( $p = 0.049$ ) when compared with a single-image type UNet. The best weights for  $\lambda_{TV}$  and  $\lambda_{Sp}$  were 0.3 and 0.6, respectively. Including the TV constraint and the similarity constraint individually improved the DSC scores by 0.8% and 0.3%, respectively. The two constraints used together improved the DSC scores by 2.0% with weights of 0.15 and 0.5. The MI-UNet with these constraints ( $\lambda_{TV} = 0.15$  and  $\lambda_{Sp} = 0.5$ ) had a superior DSC for both SSFP ( $p < 0.001$ ) and LGE ( $p < 0.001$ ), compared with a single-image type UNet (Table 2).

**Table 2.** Dice Similarity Coefficients of the MI-UNet with Different Constraints.

	SSFP	<i>p</i> Value *	LGE	<i>p</i> Value *
Single-image type UNet	$0.87 \pm 0.08$		$0.78 \pm 0.12$	
MI-UNet ( $\lambda_{TV} = 0, \lambda_{Sp} = 0$ )	$0.87 \pm 0.11$	0.73	$0.82 \pm 0.16$	0.06
MI-UNet ( $\lambda_{TV} = 0.3, \lambda_{Sp} = 0$ )	$0.89 \pm 0.08$	0.03	$0.81 \pm 0.20$	0.21
MI-UNet ( $\lambda_{TV} = 0, \lambda_{Sp} = 0.6$ )	$0.85 \pm 0.11$	0.22	$0.84 \pm 0.16$	0.002
MI-UNet ( $\lambda_{TV} = 0.15, \lambda_{Sp} = 0.5$ )	$0.92 \pm 0.06$	<0.001	$0.86 \pm 0.11$	<0.001

SSFP—steady-state free precession; LGE—late gadolinium enhancement. MI-UNet—multi-image type UNet.  $\lambda_{TV}$ —is the weight for the total variation constraint.  $\lambda_{Sp}$ —is the weight for multi-contrast similarity constraint. \* *p*-values refer to comparison with single-image type UNets.

The MI-UNet achieved a superior mean DSC compared to individually trained UNets for both SSFP and LGE images. The difference across image types was most evident at the left ventricular myocardium in SSFP images and at both the left ventricular cavity and the left ventricular myocardium in LGE images. For the right ventricle, there were no differences in DCS when comparing the MI-UNet with single-image type UNets. The MI-UNet also achieved a superior mean DSC compared to transfer-learned UNets for

SSFP images, with the difference across image types most evident at the left ventricular myocardium. The DSC metrics for the tested models in each segment are summarized in Table 3. The MI-UNet has 59M parameters and is 762MB in size, which is twice the size of the single-image type UNet.

**Table 3.** Comparison of Dice Similarity Coefficients for Tested Models in Each Segment.

MI-UNet.	Comparison with Single-Image Type UNet			Comparison with Transfer-Learned UNet	
			<i>p</i> Value		<i>p</i> Value
<b>SSFP</b>					
LVC	0.92 ± 0.06	0.91 ± 0.04	0.07	0.88 ± 0.14	0.01
LVM	0.90 ± 0.05	0.86 ± 0.07	<0.0001	0.86 ± 0.11	0.001
RV	0.88 ± 0.15	0.84 ± 0.21	0.14	0.83 ± 0.19	0.03
Mean	0.90 ± 0.07	0.87 ± 0.08	0.005	0.86 ± 0.13	0.005
<b>LGE</b>					
LVC	0.86 ± 0.12	0.78 ± 0.12	<0.0001	0.83 ± 0.14	0.11
LVM	0.89 ± 0.07	0.828 ± 0.08	<0.0001	0.86 ± 0.09	0.01
RV	0.75 ± 0.21	0.733 ± 0.21	0.56	0.69 ± 0.29	0.11
Mean	0.83 ± 0.11	0.780 ± 0.10	0.001	0.79 ± 0.15	0.04

SSFP—steady-state free precession; LGE—late gadolinium enhancement. LVC—left ventricular cavity. RV—right ventricle.

SSFP—steady-state free precession; LGE—late gadolinium enhancement; LVC—left ventricular cavity; LVM—left ventricular myocardium; RV—right ventricle.

#### 4. Discussion

Novel findings from this study were the demonstration that a jointly segmented MI-UNet model improved the segmentation accuracy for both SSFP and LGE images of the left ventricle compared to conventional single-image type UNets. This technique has the potential to save time on the interpretation of CMRs by eliminating the need to manually segment different image types for the assessment of the left ventricular cavity and walls. Changes in imaging characteristics and differences in the visualization of different tissues, such as fat or flowing blood, can greatly impact the ability of conventional segmentation methods to generalize to each contrast. However, individualized segments are critically important for the accurate quantification of pathologies, such as the extent of myocardial ischemia or the distribution of fibrosis [29]. Herein, we developed a model to jointly segment cine and LGE CMR images to allow similarities implicit in the underlying cardiac anatomy to inform the segments of each image type.

It has been widely observed that learning multiple tasks using a single network can often improve the generalizability of features, thereby improving the results for each task compared to individually learned networks [30]. There was an improvement in the cine SSFP segmentation results for the left ventricular myocardium, and we also found statistically significant improvements in the DSC scores for LGE segments both of the left ventricular cavity and the left ventricular myocardium. It appears that combining layers in the network allows the network to learn more generalizable features. We also believe that the higher signal-to-noise ratio of the cine SSFP images may provide a strong constraint for segmenting the LGE images; this is further evidenced by the efficacy of the spatial constraint offered.

The spatial constraint offered by joint segmentation provides advantages over pre-defined spatial priors. Most importantly, spatial priors may not be well-suited for patients with hypertrophic or dilated cardiomyopathies, both of which can be associated with

extensive cardiac remodeling, potentially adversely constraining segmentation. In our proposed method, rather than using a pre-conditioned shape prior, we effectively use high signal-to-noise ratio SSFP images as the prior. This constrains the LGE segments with a patient-specific prior.

The significant improvement in LGE images is clinically important because automated quantification is utilized for the diagnosis and management of many diseases, including coronary artery disease and HCM [28,30,31], since both the extent and pattern of LGE are clinically relevant [31]. Automatic quantification of these patterns is difficult because LGE images have poor tissue delineation due to the low signal of the myocardium and the bright signal of the slow-flowing adjacent blood in the cavities [32]. Even the characteristic of myocardial enhancement varies depending on pathology; bright focal lesions are characteristic of coronary artery disease, whereas diffuse, patchy regions are indicative of fibrosis or myocardial remodeling [31]. Despite the significant difference in image characteristics, incorporating features learned from the higher signal-to-noise ratio SSFP images improves the DSC metrics of the LGE images.

A recent literature review on deep learning for cardiac imaging has shown that UNets are the most well-known and most frequently used variant of fully convolutional networks for biomedical image segmentation [33]. Several state-of-the-art cardiac image segmentation methods have adopted the UNet architecture and achieved promising segmentation accuracy [33]. Most approaches used 2D networks rather than 3D networks for segmentation due to the low through-plane resolution and motion artifacts of most cardiac magnetic resonance scanners [34].

Prior cardiac segmentation studies focused on one image type (LGE or SSFP). In contrast, the present study developed a multi-image type segmentation using both LGE and SSFP because these image types provide complementary information necessary for the diagnosis and risk stratification of hypertrophic cardiomyopathy. Additionally, the majority of the prior cardiac ventricle segmentation models for CMR were trained and tested on publicly available datasets. The present study used a clinical dataset of patients with HCM with a broad spectrum of the cardiac phenotypic characteristics of this condition [20]. A previous cardiac segmentation model was trained and tested in a dataset with only eight patients with HCM. In contrast, the present study had a larger sample size of patients with HCM, including cardiac phenotypes from real-world clinical practice. It is also important to underscore that the use of clinical datasets for the development of cardiac segmentation enables the translation of these models to clinical practice.

The motion attentive transition for zero-shot video object segmentation is a promising two-stream network structure [35]. Future experiments will compare the performance of the MI-UNet with motion attentive transition for zero-shot video object segmentation. Cross-image pixel contrast for semantic segmentation is a segmentation method that addresses intra-class compactness and inter-class dispersion [36] and will be incorporated into this segmentation framework in future experiments. Group-wise learning for weakly supervised semantic segmentation is another segmentation method that enables the discovery of relationships among groups of images [37]. This methodology will also be used in future experiments.

There are limitations to this work. First, due to the lack of an external test set, it was not possible to evaluate model portability. Second, only two image types were evaluated, whereas the typical CMR study includes additional image types such as T1- or T2-weighted images. In the future, we plan to apply this technique to other image types, including T2- and T1-mapping. Third, we have only considered variants of the UNet for the comparison of results. In future studies, MI-Net results will be compared with SegNet [38] and ResNet [39] segmentation models. Future experiments will include additional metrics. Fourth, the dataset used for this project is unique, as it is from a cohort of patients with HCM. These patients often have variable patterns of myocardial thickening (e.g., isolated thickening of the septum) that can potentially confound atlas- and shape-based segmentation methods [40,41]. Moreover, the pattern of histopathologic fibrosis burden

varies widely in the population of patients with HCM [42,43]. The inconsistencies in the presentation of HCM between the two imaging types make the segmentation task more challenging compared to a healthy population. In mitigation, the comparable performance of the segmentation model in this cohort suggests the potential for generalizability across sequences and diseases.

## 5. Conclusions

The joint segmentation of multiple image types provides a deep neural network architecture that supports the automated segmentation of CMR images in patients with HCM. Unlike other methods that yield a single mask using multiple sensors, this method does not require a separate mask registration, which is time-consuming and prone to error. Furthermore, the method described yields superior results for left ventricular segmentation with fewer data compared to individually training separate models on different imaging types.

**Author Contributions:** D.C., S.P.A., M.E. and N.Z.F. developed and trained the models and performed statistical analysis. H.B., S.L. and P.A.A. performed the image segmentation. D.C., H.B., N.D., S.P.A., M.E., N.Z.F., K.P., S.L., J.M.B., J.B.G., M.J.A., R.A. and A.M.A.-O. developed the cohort, recruited patients, and acquired the imaging data. D.C., P.A.A., J.M.B. and A.M.A.-O. conceived and designed the experiments. D.C., P.A.N., S.P.A., M.E., N.Z.F. and B.J.E. provided AI architecture innovations. A.M.A.-O. conceived the overall study and handled funding and supervision. D.C., H.B., N.D., S.P.A., M.E., N.Z.F. and A.M.A.-O. drafted the manuscript. All authors read, provided criticism, and approved the final manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** The research reported in this publication was supported by the Paul and Ruby Tsai Family Hypertrophic Cardiomyopathy Research Fund; the National Heart, Lung, and Blood Institute of National Institutes of Health (K01HL124045); and by the Mayo Clinic K2R award. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of the Mayo Clinic (protocol code 17-003392; date of approval 5 April 2017).

**Informed Consent Statement:** Patient consent was waived due to the fact that research involved no more than minimal risk to subjects and because research could not practicably be carried out without the waiver.

**Data Availability Statement:** The data used in this study can be made available to researchers collaborating with Mayo Clinic under a research agreement. However, the data are not publicly available due to the need to preserve the privacy of patient health information.

**Acknowledgments:** The authors thank Tessa L. Flies and Katie M. Nagel for secretarial support.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Captur, G.; Manisty, C.; Moon, J.C. Cardiac MRI evaluation of myocardial disease. *Heart* **2016**, *102*, 1429–1435. [[CrossRef](#)] [[PubMed](#)]
2. Bogaert, J.D.S.; Taylor, A.M.; Muthurangu, V. *Clinical Cardiac MRI*; Springer: Berlin/Heidelberg, Germany; New York, NY, USA; Dordrecht, The Netherlands; London, UK, 2012.
3. Brenes, J.C.; Doltra, A.; Prat, S. Cardiac magnetic resonance imaging in the evaluation of patients with hypertrophic cardiomyopathy. *Glob. Cardiol. Sci. Pract.* **2018**, *2018*, 22. [[CrossRef](#)] [[PubMed](#)]
4. Liu, H.; Hu, H.; Xu, X.; Song, E. Automatic left ventricle segmentation in cardiac MRI using topological stable-state thresholding and region restricted dynamic programming. *Acad. Radiol.* **2012**, *19*, 723–731. [[CrossRef](#)] [[PubMed](#)]
5. Tsai, A.; Yezzi, A.; Wells, W.; Tempny, C.; Tucker, D.; Fan, A.; Grimson, W.E.; Willsky, A. A shape-based approach to the segmentation of medical imagery using level sets. *IEEE Trans. Med. Imaging* **2003**, *22*, 137–154. [[CrossRef](#)]

6. Grosgeorge, D.; Petitjean, C.; Dacher, J.-N.; Ruan, S. Graph cut segmentation with a statistical shape model in cardiac MRI. *Comput. Vis. Image Underst.* **2013**, *117*, 1027–1035. [[CrossRef](#)]
7. Zotti, C.; Luo, Z.; Lalande, A.; Humbert, O.; Jodoin, P.-M. Novel Deep Convolution Neural Network Applied to MRI Cardiac Segmentation. *arXiv* **2017**, arXiv:1705.08943.
8. Zhuang, X.; Hawkes, D.; Crum, W.; Boubertakh, R.; Uribe, S.; Atkinson, D.; Batchelor, P.; Schaeffter, T.; Razavi, R.; Hill, D. Robust registration between cardiac MRI images and atlas for segmentation propagation. In Proceedings of the Medical Imaging 2008: Image Processing, San Diego, CA, USA, 16–21 February 2008; Reinhardt, J.M., Pluim, J., Eds.; SPIE Proceedings: San Diego, CA, USA, 2008; Volume 6914.
9. Shen, D.; Wu, G.; Suk, H.I. Deep Learning in Medical Image Analysis. *Annu. Rev. Biomed. Eng.* **2017**, *19*, 221–248. [[CrossRef](#)]
10. Avendi, M.R.; Kheradvar, A.; Jafarkhani, H. A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. *Med. Image Anal.* **2016**, *30*, 108–119. [[CrossRef](#)]
11. Avendi, M.R.; Kheradvar, A.; Jafarkhani, H. Automatic segmentation of the right ventricle from cardiac MRI using a learning-based approach. *Magn. Reson. Med.* **2017**, *78*, 2439–2448. [[CrossRef](#)]
12. Oktay, O.; Ferrante, E.; Kamnitsas, K.; Heinrich, M.; Bai, W.; Caballero, J.; Cook, S.A.; de Marvao, A.; Dawes, T.; O’Regan, D.P.; et al. Anatomically Constrained Neural Networks (ACNNs): Application to Cardiac Image Enhancement and Segmentation. *IEEE Trans. Med. Imaging* **2018**, *37*, 384–395. [[CrossRef](#)]
13. Liu, Y.; Wang, W.; Wang, K.; Ye, C.; Luo, G. An Automatic Cardiac Segmentation Framework based on Multi-sequence MR Image. *arXiv* **2019**, arXiv:1909.05488.
14. Shi, J.; Ye, Y.; Zhu, D.; Su, L.; Huang, Y.; Huang, J. Automatic segmentation of cardiac magnetic resonance images based on multi-input fusion network. *Comput. Methods Programs Biomed.* **2021**, *209*, 106323. [[CrossRef](#)] [[PubMed](#)]
15. Li, C.; Chen, M.; Zhang, J.; Liu, H. Cardiac MRI segmentation with focal loss constrained deep residual networks. *Phys. Med. Biol.* **2021**, *66*. [[CrossRef](#)]
16. Hann, E.; Popescu, I.A.; Zhang, Q.; Gonzales, R.A.; Barutçu, A.; Neubauer, S.; Ferreira, V.M.; Piechnik, S.K. Deep neural network ensemble for on-the-fly quality control-driven segmentation of cardiac MRI T1 mapping. *Med. Image Anal.* **2021**, *71*, 102029. [[CrossRef](#)] [[PubMed](#)]
17. Zhang, H.; Zhang, W.; Shen, W.; Li, N.; Chen, Y.; Li, S.; Chen, B.; Guo, S.; Wang, Y. Automatic segmentation of the cardiac MR images based on nested fully convolutional dense network with dilated convolution. *Biomed. Signal Processing Control.* **2021**, *68*. [[CrossRef](#)]
18. Chan, R.H.; Maron, B.J.; Olivotto, I.; Pencina, M.J.; Assenza, G.E.; Haas, T.; Lesser, J.R.; Gruner, C.; Crean, A.M.; Rakowski, H.; et al. Prognostic value of quantitative contrast-enhanced cardiovascular magnetic resonance for the evaluation of sudden death risk in patients with hypertrophic cardiomyopathy. *Circulation* **2014**, *130*, 484–495. [[CrossRef](#)] [[PubMed](#)]
19. Maron, M.S.; Rowin, E.J.; Wessler, B.S.; Mooney, P.J.; Fatima, A.; Patel, P.; Koethe, B.C.; Romashko, M.; Link, M.S.; Maron, B.J. Enhanced American College of Cardiology/American Heart Association Strategy for Prevention of Sudden Cardiac Death in High-Risk Patients With Hypertrophic Cardiomyopathy. *JAMA Cardiol.* **2019**. [[CrossRef](#)]
20. Ommen, S.R.; Mital, S.; Burke, M.A.; Day, S.M.; Deswal, A.; Elliott, P.; Evanovich, L.L.; Hung, J.; Joglar, J.A.; Kantor, P.; et al. AHA/ACC Guideline for the Diagnosis and Treatment of Patients With Hypertrophic Cardiomyopathy: Executive Summary: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *Circulation* **2020**, *142*, e533–e557. [[CrossRef](#)]
21. Philbrick, K.A.; Weston, A.D.; Akkus, Z.; Kline, T.L.; Korfiatis, P.; Sakinis, T.; Kostandy, P.; Boonrod, A.; Zeinoddini, A.; Takahashi, N.; et al. RIL-Contour: A Medical Imaging Dataset Annotation Tool for and with Deep Learning. *J. Digit. Imaging* **2019**, *32*, 571–581. [[CrossRef](#)]
22. Mason, D. SU-E-T-33: Pydicom: An Open Source DICOM Library. *Med. Phys.* **2011**, *38*, 3493. [[CrossRef](#)]
23. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
24. Sorensen, T.J. *A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content and Its Application to Analyses of the Vegetation on Danish Commons*; I kommission hos E. Munksgaard: Kobenhavn, Denmark, 1948.
25. Rudin, L.I.; Osher, S.; Fatemi, E. Nonlinear total variation based noise removal algorithms. *Phys. D Nonlinear Phenom.* **1992**, *60*, 259–268. [[CrossRef](#)]
26. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
27. Paszke, A.; Gross, S.; Chintala, S.; Chanan, G.; Yang, E.; Devito, Z.; Lin, Z.; Desmaison, A.; Antiga, L.; Lerer, A. Automatic differentiation in PyTorch. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
28. Zou, K.H.; Warfield, S.K.; Bharatha, A.; Tempany, C.M.; Kaus, M.R.; Haker, S.J.; Wells, W.M., 3rd; Jolesz, F.A.; Kikinis, R. Statistical validation of image segmentation quality based on a spatial overlap index. *Acad. Radiol.* **2004**, *11*, 178–189. [[CrossRef](#)]
29. Vogel-Claussen, J.; Rochitte, C.E.; Wu, K.C.; Kamel, I.R.; Foo, T.K.; Lima, J.A.; Bluemke, D.A. Delayed enhancement MR imaging: Utility in myocardial assessment. *Radiographics* **2006**, *26*, 795–810. [[CrossRef](#)]
30. Ruder, S. An overview of multi-task learning in deep neural networks. *arXiv* **2017**, arXiv:1706.05098.

31. Bohl, S.; Wassmuth, R.; Abdel-Aty, H.; Rudolph, A.; Messroghli, D.; Dietz, R.; Schulz-Menger, J. Delayed enhancement cardiac magnetic resonance imaging reveals typical patterns of myocardial injury in patients with various forms of non-ischemic heart disease. *Int. J. Cardiovasc. Imaging* **2008**, *24*, 597–607. [[CrossRef](#)]
32. Kim, H.W.; Rehwald, W.G.; Jenista, E.R.; Wendell, D.C.; Filev, P.; van Assche, L.; Jensen, C.J.; Parker, M.A.; Chen, E.-I.; Crowley, A.L.C. Dark-blood delayed enhancement cardiac magnetic resonance of myocardial infarction. *JACC Cardiovasc. Imaging* **2018**, *11*, 1758–1769. [[CrossRef](#)]
33. Chen, C.; Qin, C.; Qiu, H.; Tarroni, G.; Duan, J.; Bai, W.; Rueckert, D. Deep Learning for Cardiac Image Segmentation: A Review. *Front. Cardiovasc. Med.* **2020**, *7*. [[CrossRef](#)]
34. Baumgartner, C.; Koch, L.; Pollefeys, M.; Konukoglu, E. An exploration of 2D and 3D deep learning techniques for cardiac MR image segmentation. arXiv e-prints. 2017. *Cornell Univ. (arXiv)* **2017**, *1*. [[CrossRef](#)]
35. Zhou, T.; Wang, S.; Zhou, Y.; Yao, Y.; Li, J.; Shao, L. Motion-Attentive Transition for Zero-Shot Video Object Segmentation. *arXiv* **2020**, *2003*, 04253. [[CrossRef](#)]
36. Wang, W.; Zhou, T.; Yu, F.; Dai, J.; Konukoglu, E.; Van Gool, L. Exploring Cross-Image Pixel Contrast for Semantic Segmentation. *arXiv* **2021**, arXiv:2101.11939.
37. Zhou, T.; Li, X.; Feng, C.-M.; Li, J.; Shao, L. Group-Wise Learning for Weakly Supervised Semantic Segmentation. *IEEE Trans. Image Processing* **2022**, *31*, 799–811. [[CrossRef](#)] [[PubMed](#)]
38. Badrinarayanan Vijay, K.A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *arXiv* **2015**, arXiv:1511.00561.
39. Long, W.; Li, X.; Gao, L. A transfer convolutional neural network for fault diagnosis based on ResNet-50. *Neural. Comput. Applic.* **2020**, *32*, 6111–6124. [[CrossRef](#)]
40. Neubauer, S.; Kolm, P.; Ho, C.Y.; Kwong, R.Y.; Desai, M.Y.; Dolman, S.F.; Appelbaum, E.; Desvigne-Nickens, P.; DiMarco, J.P.; Friedrich, M.G. Distinct subgroups in hypertrophic cardiomyopathy in the NHLBI HCM registry. *J. Am. Coll. Cardiol.* **2019**, *74*, 2333–2345. [[CrossRef](#)]
41. Avendi, M.; Kheradvar, A.; Jafarkhani, H. Fully automatic segmentation of heart chambers in cardiac MRI using deep learning. *J. Cardiovasc. Magn. Reson.* **2016**, *18*, P351. [[CrossRef](#)]
42. Tanaka, M.; Fujiwara, H.; Onodera, T.; Wu, D.; Hamashima, Y.; Kawai, C. Quantitative analysis of myocardial fibrosis in normals, hypertensive hearts, and hypertrophic cardiomyopathy. *Heart* **1986**, *55*, 575–581. [[CrossRef](#)]
43. Aquaro, G.D.; Positano, V.; Pingitore, A.; Strata, E.; Di Bella, G.; Formisano, F.; Spirito, P.; Lombardi, M. Quantitative analysis of late gadolinium enhancement in hypertrophic cardiomyopathy. *J. Cardiovasc. Magn. Reson.* **2010**, *12*, 21. [[CrossRef](#)]