



HHS Public Access

Author manuscript

Annu Int Conf IEEE Eng Med Biol Soc. Author manuscript; available in PMC 2023 February 23.

Published in final edited form as:

Annu Int Conf IEEE Eng Med Biol Soc. 2021 November ; 2021: 4072–4078. doi:10.1109/EMBC46164.2021.9629581.

Deep Learning-Based Segmentation and Uncertainty Assessment for Automated Analysis of Myocardial Perfusion MRI Datasets Using Patch-Level Training and Advanced Data Augmentation

Dilek Mirgun Yalcinkaya,

Cedars-Sinai Medical Center, UCLA Dept. of Bioengineering, and the Laboratory for Translational Imaging of Microcirculation, Krannert Cardiovascular Research Center, Indiana University School of Medicine, Indianapolis

Khalid Youssef [Member, IEEE],

Cedars-Sinai Medical Center, UCLA Dept. of Bioengineering, and the Laboratory for Translational Imaging of Microcirculation, Krannert Cardiovascular Research Center, Indiana University School of Medicine, Indianapolis

Bobby Heydari,

Cumming School of Medicine, University of Calgary, Alberta, Canada.

Luis Zamudio,

Cedars-Sinai Medical Center, UCLA Dept. of Bioengineering, and the Laboratory for Translational Imaging of Microcirculation, Krannert Cardiovascular Research Center, Indiana University School of Medicine, Indianapolis

Rohan Dharmakumar,

Krannert Cardiovascular Research Center, Dept. of Medicine, and IU Health/IUSM Cardiovascular Institute, Indiana University School of Medicine, Indianapolis.

Behzad Sharif [Member, IEEE]

Cedars-Sinai Medical Center, UCLA Dept. of Bioengineering, and the Laboratory for Translational Imaging of Microcirculation, Krannert Cardiovascular Research Center, Indiana University School of Medicine, Indianapolis

Abstract

In this work, we develop a patch-level training approach and a task-driven intensity-based augmentation method for deep-learning-based segmentation of motion-corrected perfusion cardiac magnetic resonance imaging (MRI) datasets. Further, the proposed method generates an image-based uncertainty map thanks to a novel spatial sliding-window approach used during patch-level training, hence allowing for uncertainty quantification. Using the quantified uncertainty, we detect the out-of-distribution test data instances so that the end-user can be alerted that the test data is not suitable for the trained network. This feature has the potential to enable a more reliable integration of the proposed deep learning-based framework into clinical practice. We test our

(corresponding author: Behzad Sharif, PhD; phone: 310-423-7758; behzad.sharif@cshs.org, bzsharif@gmail.com).

approach on external MRI data acquired using a different acquisition protocol to demonstrate the robustness of our performance to variations in pulse-sequence parameters. The presented results further demonstrate that our deep-learning image segmentation approach trained with the proposed data-augmentation technique incorporating spatiotemporal (2D+time) patches is superior to the state-of-the-art 2D approach in terms of generalization performance.

I. INTRODUCTION

Automated image-analysis approaches based on deep learning hold significant promise to enable rapid and objective quantification of cardiac magnetic resonance (CMR) imaging datasets by eliminating the need for any manual steps specifically in the segmentation of CMR images [1]. In the context of stress/rest perfusion CMR imaging for diagnosis of ischemic heart disease, such approaches have shown promise in removing the cumbersome step of contouring myocardial borders given the dynamic (2D+time) changes in myocardial contrast during the first pass of the contrast agent [2,3,4].

Modern neural networks may suffer from miscalibration where model confidence and accuracy deviate from each other. Thus, caution is needed on the potential overconfidence that may be incorrectly implied in segmentations generated by deep learning models. An accompanying uncertainty assessment of the deep learning-based pipeline may help add an interpretable measure of confidence which may, in turn, help prevent the end-user from interpreting the results without any nuance in this regard. Furthermore, uncertainty assessment of deep learning-based applications in medical imaging is essential in understanding the performance bounds and limitations of the underlying models. Uncertainty visualization of deep neural networks has been studied in multiple modalities [5] including CMR [6]. In this work, we introduce a method to extract image-based (aleatoric) uncertainty based on patch-level training to detect out-of-distribution (OOD) test CMR data.

OOD test data detection can play an important role in clinical imaging applications where clinicians (the end-user) interact with deep learning-based tools. For instance, an automated segmentation model can alert the end-user based on a human-in-the-loop framework wherein the end-user is asked to check the segmentation quality only for the OOD test data on which the network has shown a low level of confidence. Based on this framework, the human-in-the-loop avoids the burden of having to verify the machine-generated segmentation (e.g., myocardial contours) for every single time frame, and instead focus on quality-control of a small subset of the imaging data that is detected to be OOD.

Our contributions in this work can be summarized as follows:

- We propose a spatiotemporal (2D+time) deep learning method with patch-level training using a spatial sliding-window to segment motion-corrected CMR perfusion images.
- We train the patch-level network using a task-driven augmentation technique, including an intensity modulation method, for improved generalization (robustness) of the network.

- To evaluate the robustness of the proposed patch-level training method against variations in MRI pulse sequence parameters, we trained/validated our model on a T1-weighted dataset (acquired internally at our center) and tested it on T2/T1-weighted CMR images obtained at an external center.
- Using the fact that a pixel belongs to multiple patches during patch-level testing, we obtain the network's image-based uncertainty in segmenting the myocardium with a 2D uncertainty map UM and locate the uncertainty in form of a 6-sector segmentation model according to the established American Heart Association (AHA) segmentation model.
- Using the uncertainty map UM generated by our proposed method, we quantify the overall uncertainty of segmentation (denoted by U) which quantifies the patch-level network's confidence in the segmentation output. Next, we establish a classifier based on U to identify OOD data in the test dataset. The threshold involved in designing this classifier is determined based on the absolute quantification of myocardial blood flow (MBF).

To the best of our knowledge, this work is the first to apply patch-level training for image-based uncertainty assessment and to use intensity-modulated data augmentation for deep learning-based segmentation of myocardial perfusion CMR datasets.

II. Methods

A. Dataset

We used a two-center short-axis respiratory motion-corrected perfusion CMR dataset acquired at 3T with different pulse sequences. For all of the volunteer and patient imaging studies, local Institutional Review Board (IRB) approval and written informed consent was obtained before each imaging exam/study. Training and validation of the deep neural network were done on 96 stress/rest perfusion studies with suspected ischemia using an SR-prepared FLASH (RF-spoiled gradient recalled echo) pulse sequence with T1-weighted image contrast acquired at 3T. Testing was performed on 40 independent adenosine-stress CMR perfusion studies obtained at an external site using an SR-prepared balanced steady-state free precession (bSSFP) pulse sequence with T2/T1-weighted contrast (also acquired at 3T). There was no leak from the training/validation set to the test set.

Original raw perfusion series was truncated to 30-time frames using previously proposed method. Then, we spatially upsampled the time-series and cropped the perfusion series at the center of the left ventricular (LV), resulting in a size of $128 \times 128 \times 30$ for each slice. Each image volume was normalized to an intensity range of $[0,1]$.

For training/validation images, ground truth segmentation labels for myocardium and right ventricular (RV) were available. LV masks were obtained from the myocardium masks. Hence, the CNNs were trained to perform the segmentation task in the heart for the following classes: myocardium, LV, RV, and background.

B. Patch-level model and training

We used a vanilla UNet as the convolutional neural network (CNN) [7]. The input to the UNet is 3D patches extracted from the motion-corrected perfusion time series using a spatial sliding window. We call this network that we propose *patch3-UNet* and perform the segmentation from 3D patch to 2D patch with the preserved spatial size. The output of *patch3-UNet* is 2D segmentation of the 3D patches as a result of working with motion-corrected images. Each pixel at the output of *patch3-UNet* CNN has a probability of belonging to one of the four classes described above, also known as softmax probability. In patch-level training, patches of the desired size are extracted from the input images with chosen stride and fed to the network, making each 3D patch an independent data sample for the network. Likewise, during testing, the network segments the 3D (2D+time) patches and outputs the 2D segmentation result for each patch in the same size. Then, patches are combined back together to yield the final segmentation result. Generally, overlapping patches are preferred over non-overlapping to minimize the edge effects during testing. For the pixels that are in multiple patches, we take the mean of the myocardium softmax probabilities at each pixel. Fig 1 summarizes a test image's segmentation process using the patch-level method.

Patch-level training makes the network more robust to overfitting by “increasing” the dataset's size. For example, it leads to a 9-fold increase in the training set size when utilized with a patch- and stride-size of 64 in each spatial dimension when each spatial dimension of the input time-series is 128 (as in Fig 1). Also, with an optimal patch-size choice, patch-level training enables the network to simultaneously focus on the image's local and global details. With the spatial sliding window approach, more patches are extracted from the center of the image series than the edges, which is helpful because we localize our images such that the heart is in the middle. Besides, there is no redundancy in the extracted patches, such as an outside-of-the-ROI patch, since our experiments showed that using a relatively larger patch size is favorable for the segmentation problem we aim to solve. We optimized cross-entropy loss with Adam optimizer for all models we trained. Model selection was made using early stopping according to the segmentation performance on the validation set.

C. Advanced Data Augmentation

We first apply commonly used augmentation techniques, including various affine transformations (rotation, scaling, vertical and horizontal flips, image scaling, and shear) and additive white noise to enhance our training set. In addition, we introduce the following data-driven techniques to augment the perfusion images: 1) Perlin noise: Perlin noise is used in various computer vision applications and has a natural-seeming and smooth texture. 2) Intensity modulation: we proposed a task-driven augmentation technique with the motivation that intensity-based augmentation techniques have been shown to improve the shape-based representation capabilities of CNNs in computer vision applications [8,9]. Further, different myocardial intensity levels can be induced by modulating the images with the intensity gradient maps shown in Fig 2, including patterns similar to lateral-wall signal drop-off often seen in raw perfusion images. Finally, we employed contrast augmentation with a probability of 0.50 for each training time series in the advanced augmentation setting. These

techniques vary the texture and contrast levels of the training images and improve the generalizability of *patch3-UNet* trained on FLASH and tested on the bSSFP dataset.

D. Evaluation

We evaluate the trained models on the held-out test set and report the myocardium Dice similarity coefficient between the manual expert labels and the automatic labels found by the proposed pipeline. Further, we report myocardial blood flow (MBF) correlation between manual and automatic segmentation based on American Heart Association (AHA) 48-segment model using Fermi-constrained deconvolution. We average the 48-segment MBF results to obtain the slice-averaged MBF numbers.

As a baseline method, we trained another patch-level UNet using the exact same architecture but with 2D patches, which we name *patch2-UNet* to compare it with the proposed *patch3-UNet*. We aim to assess how using 3D patches affects the generalization of the patch-level method when tested on the external test set obtained with a different pulse sequence. Also, we hypothesize that the proposed *patch3-UNet* method improves the network's performance by learning the temporal connections inherent in the perfusion image series. Following previous studies, *patch2-UNet* was trained with 2D patches extracted from the peak-enhancement LV frame of the perfusion time-series [4]. The peak-enhancement LV frame of each time series is detected using earlier methods [10]. Both *patch2-UNet* and *patch3-UNet* were trained with and without the proposed augmentation techniques to evaluate the effect of the proposed augmentation approach. In total, we trained 4 networks corresponding to 4 different input dimension and augmentation settings while keeping the same setup with no changes made to the vanilla UNet architecture.

E. Uncertainty Estimation & Out-of-Distribution (OOD) Test Data Detection

When utilized with a small stride during testing, patch-level prediction enables the same pixel to be segmented multiple times by the network. Thus, the number of softmax probabilities obtained for each pixel depends on the number of patches that pixel is in. We define the set $\Gamma(x)$ which is the set of all the patches that have pixel x in them thanks to the spatial sliding window approach. The bar plot in Fig 3b illustrates how the cardinality of $\Gamma(x)$ changes depending on the pixel location. Different voting schemes to decide the label of pixel x exist, such as 1) thresholding softmax probabilities first and then getting the majority vote, 2) getting mean of the mean of the softmax probabilities and then threshold (experiments with these methods are described in the Results section). Regardless of the voting fashion, let us define the segmentation output of the network as $b(P_i(x))$, where $P_i(x)$ is the 2D matrix of myocardium probabilities (softmax output) for slice i and $b(\cdot)$ is the binarization operation applied at each pixel x .

In addition to the thresholded myocardial segmentation $b(P_i(x))$, the network outputs an uncertainty map thanks to patch-level training. We use the multiple softmax probabilities of each pixel to create an uncertainty map. $UM_i(x)$ is the 2D segmentation uncertainty map of image slice i and standard deviation of the softmax outputs of distinct patches at each spatial location x . Uncertainty map UM_i shows the segmentation agreement and reveals the areas where the network is challenged the most when segmenting perfusion time series slice i . If

a pixel x is segmented as myocardium most of the time across several patch segmentations, then the network is confident at that pixel. When the network is indecisive, however, the same region collects varying probabilities from different patches. Therefore, this region has higher intensities in the uncertainty map and consequently highlighted as unsure. Fig 3a shows how the same area is segmented several times by being in many patches.

To quantify uncertainty, we sum up the values in each 6 AHA subsegment of the uncertainty map UM_i and normalize by myocardium area of the subsegment to make the uncertainty metric invariant to the heart's size. RV segmentation ability of the network assists in RV insertion point (RVIP) detection and subsegment division. Binary mask of each subsegment k is denoted by M_k (see colorful subsegments in Fig 5). Quantified uncertainty U_i of the acquisition slice i is the mean uncertainty of the 6 AHA subsegments:

$$U_i = \frac{1}{6} \sum_{k=1}^6 \frac{1}{A_k} \sum_x UM_i(x) \odot M_k(x) \quad (1)$$

$$A_k = \sum_x b(P_i(x)) \odot M_k(x) \quad (2)$$

where A_k is the area of the subsegment k —the number of the pixels in the segmentation result $b(P_i(x))$ after masked with $M_k(x)$ — and \odot is the Hadamard product.

Because the underlying assumption on the training images is respiratory motion correction, we are encouraged to show where it fails in the test set and, therefore, detect unreliable segmentation, which would be OOD data. To define the ground truth OOD labels accurately, we use the masks generated by the baseline state-of-the-art 2D network *patch2-UNet* and label the cases having > 5% discrepancy between the 2D automatic and ground truth slice-averaged myocardial blood flows (MBFs) as OOD. The intuitive explanation is simple, if the mask generated by the 2D network using the information from a single time frame has a certain MBF error, then that frame is not informative enough about the flow through the myocardium, implying that the motion correction possibly fails.

Lastly, we establish a classifier between the detected OOD labels and $1 - U$ (to get positive correlation) values from the *patch3-UNet* with advanced augmentation. Then, we set a threshold t through receiver operating characteristic (ROC) curve such that slice i having $1 - U_i < t$ is labeled as “uncertain”. Further, we locate the source of uncertainty in the OOD cases according to AHA's 6-segment model before referring them to the end-user for segmentation revision. So, the human-in-the-loop spends time correcting only for one AHA subsegment. The most challenging subsegment region is identified as having the uncertainty map subsegment having the highest total variation (TV).

In short, UM_i provides a visual map for how confident the network is in its segmentation result, whereas U_i indicates – with a single number– where the segmentation of slice i stands among the others in terms of network confidence in the test set.

III. Results

A. Effect of patch and stride size & voting method

We experimented with different patches and stride sizes. Patch size of 64 turned out to result in the best from the set {16,32,64} according to the validation set performance. The networks trained with smaller patch sizes had difficulty focusing on global details and had errors.

We used different stride sizes for segmentation and uncertainty map results. Stride size of 50% of the patch size (32-pixels) was utilized in either spatial direction for segmentation during training or testing. We used 3.125% of the patch size (2-pixels) as step size for uncertainty map creation. Using a tiny stride minimized edge effects and allowed a single pixel to be segmented many times. Therefore, the uncertainty maps were created with the knowledge from multiple segmentation and reflected the challenging regions in the images. Center pixels being in more patches than the edge pixels were favorable given the images were cropped at the center of the LV cavity. Still, the pixels at the edge of the ROI were in an adequate number of patches to detect uncertainty accurately since we utilized a very small stride.

We experimented with 1) majority vote, 2) mean of the softmax probabilities voting methods. In majority vote, softmax probabilities are rounded at each pixel, resulting in 0's (not myo) and 1's (myo). For segmentation, if 1's are more than 0's that pixel is segmented as myo and vice versa. In case of a tie, we tossed a coin. We calculated the standard deviation of the decisions after binarization operation for uncertainty maps. As a second method, we used the mean of softmax probabilities of each patch that have pixel x in them and then applied the binarization for segmentation result. For uncertainty maps, we obtained the standard deviation of softmax probabilities in the set $\Gamma(x)$. The results did not exhibit a significant difference, and we proceeded with the mean of the softmax probabilities in the rest of the findings.

B. Myocardium Segmentation

We trained four networks in total: *patch2-UNet* and *patch3-UNet* for both proposed advanced augmentation and conventional augmentation. Mean Dice score for *patch3-UNet* with advanced data augmentation was 0.86 on 120 test set slices (3 acquisition slices of 40 stress perfusion studies). We also summarize sector-wise Dice scores in Table 1 for slice-averaged, 6-sector, and 12-sector (2 radial, 6 angular division) cases. The subsegment division was done with the help of RVIP detection ability of the networks. The gap between the proposed method Dice and the compared methods widens as we go from slice-averaged to the 12-sector case. Also, segmentation results of two patients (3 acquisition slices each) are shown in Fig 4.

C. Perfusion Quantification

We quantified myocardial blood flow (MBF) with Fermi-constrained deconvolution using American Heart Association (AHA) 48-sector (24 segment and 2 radial divisions) model and averaged 48-sector MBF numbers to obtain transmural MBF numbers for each slice.

Comparison of MBF numbers obtained from automated analysis using *patch3-UNet* and *patch2-UNet* (trained with the advanced augmentation) vs. the ground truth masks showed very strong correlation ($R^2 > 0.95$ for both). However, Bland-Altman analysis between *patch3-UNet* and ground truth masks exhibited a mean value of 0.007, and limit of agreements (at + and -1.96 standard deviation away from the mean) of 0.174 and -0.161 . In contrast, Bland-Altman analysis between *patch2-UNet* and the same ground truth masks exhibited a mean value of -0.073 , and limit of agreements of 0.200 and -0.347 . In other words, *patch3-UNet* exhibited tighter limits of agreement vs. ground truth.

D. Uncertainty Map Visualization & OOD Detection

Fig 5 shows two examples of *patch3-UNet*'s segmentation and uncertainty maps outputs on two challenging test acquisition slices. The process of localizing the uncertainty in challenging cases using RVIP and uncertainty maps is also summarized.

Fig 6 shows the correlation between Dice and $1 - U$ values, red dots corresponding to out-of-distribution (OOD) acquisition slices labels. Test data labels that was not suitable for the trained model (i.e., OOD labels) were obtained where 2D patch-level network has $>5\%$ myocardial blood flow (MBF) error with respect to ground truth segmentation, hence implying possible failing of motion correction. ROC analysis between OOD data labels obtained using *patch2-UNet* with advanced augmentation and $1 - U$ values resulted in an AUC of 0.936, specificity of 0.918, and sensitivity of 0.826. The optimal operating point of the ROC curve was $t = 0.80$, indicating that slices having $1 - U < t$ are classified as OOD by the proposed method. The mean Dice score for the test cases having $1 - U < t$ was 0.81 which would be the "unreliable" segmentations generated by the model. Likewise, the mean Dice score for the "reliable" test slices was 0.87, i.e., test cases that have $1 - U > t$. As shown in Table 1, the proposed method achieved a mean Dice score of 0.86 across all 40 stress perfusion CMR test data.

In Fig 6, we also demonstrate two outlier slices that do not have a 5% MBF error with *patch2-UNet* but having very high uncertainty (small $1 - U$). These slices are also OOD data but with a different definition, e.g., small blood pool and subtle fat to distinguish in the lateral wall. Other definitions of OOD can be used for these slices. Still, the proposed method is uncertain in these slices.

IV. Discussion

In this work, we have demonstrated that by employing the proposed patch-based 2D+time training approach, the performance of UNet-based deep-learning techniques for segmentation of motion-corrected perfusion CMR datasets can be improved considerably. To the best of our knowledge, this work is the first to apply patch-based network training and intensity-modulated data augmentation for CNN-based segmentation of myocardial perfusion CMR datasets.

A limitation of the proposed method is that it is trained with the complete first-pass time series images. Thus, further experiments are needed to assess its robustness in datasets that are partially truncated in time. Also, we used the raw signal intensities in the deconvolution-

based quantification of MBF; therefore the derived blood flow numbers (and hence the thresholds) may be confounded by saturation effects in the arterial input function. This issue can be remedied by conversion of signal intensities to contrast-agent concentration prior to Fermi-constrained deconvolution. Finally, we did not evaluate the performance of the proposed method in test data acquired at a field strength other than 3T. Therefore, our future work involves testing of *patch3-UNet* on data acquired at 1.5T.

Acknowledgments

Research supported by National Heart, Lung and Blood Institute through National Institutes of Health (NIH) grants R00-HL124323 and R01-HL153430 (PI: B. Sharif) and R01-HL136578 and R01-HL148788 (PI: R. Dharmakumar).

References

- [1]. Leiner T, Rueckert D, Suinesiaputra A et al. Machine learning in cardiovascular magnetic resonance: basic concepts and applications. *J. Cardiovasc. Magn. Reson* 2019;21:61. doi:10.1186/s12968-019-0575-y [PubMed: 31590664]
- [2]. Sandoval Z, Van Dyke J, Dharmakumar R, Sharif B. Rapid Automatic Quantification of Myocardial Blood Flow in Free-breathing Myocardial Perfusion MRI without the Need for Motion Correction: A Novel Spatio-temporal Deep Learning Approach. *Proceedings of ISMRM* 2019;27:1230
- [3]. Xue H, Tseng E, Knott KD, Kotecha T, Brown L, Plein S et al. Automated detection of left ventricle in arterial input function images for inline perfusion mapping using deep learning: A study of 15,000 patients. *Magn Reson Med*. 2020;84(5):2788–2800. doi:10.1002/mrm.28291 [PubMed: 32378776]
- [4]. Scannell CM, Veta M, Villa ADM et al. Deep- Learning-Based Preprocessing for Quantitative Myocardial Perfusion MRI. *J Magn Reson Imaging*. 2020;51(6):1689–1696. doi:10.1002/jmri.26983 [PubMed: 31710769]
- [5]. Roy AG, Conjeti S, Navab N, Wachinger C, Alzheimer’s Disease Neuroimaging Initiative. Bayesian quicknat: model uncertainty in deep whole-brain segmentation for structure-wise quality control. *NeuroImage*. 2019 Jul 15;195:11–22. [PubMed: 30926511]
- [6]. Do HP, Guo Y, Yoon AJ, Nayak KS. Accuracy, uncertainty, and adaptability of automatic myocardial ASL segmentation using deep CNN. *Magn Reson Med*. 2020 May;83(5):1863–74. [PubMed: 31729078]
- [7]. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Proceedings of MICCAI*. 2015;9351:234–241. doi:10.1007/978-3-319-24574-4_28
- [8]. Geirhos R, Rubisch P, Michaelis C, Bethge M, Wichmann FA, Brendel W, ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. *Proceedings of ICLR* 2019. arXiv:1811.12231
- [9]. Coupé P, Manjón JV, Fonov V, Pruessner J, Robles M, Collins DL. Patch-based segmentation using expert priors: application to hippocampus and ventricle segmentation. *Neuroimage*. 2011;54(2):940–954. doi:10.1016/j.neuroimage.2010.09.018 [PubMed: 20851199]
- [10]. Jacobs M, Benovoy M, Chang LC, Arai AE, Hsu LY. Evaluation of an automated method for arterial input function detection for first-pass myocardial perfusion cardiovascular magnetic resonance. *Journal of Cardiovascular Magnetic Resonance*. 2016;18:17. [PubMed: 27055445]

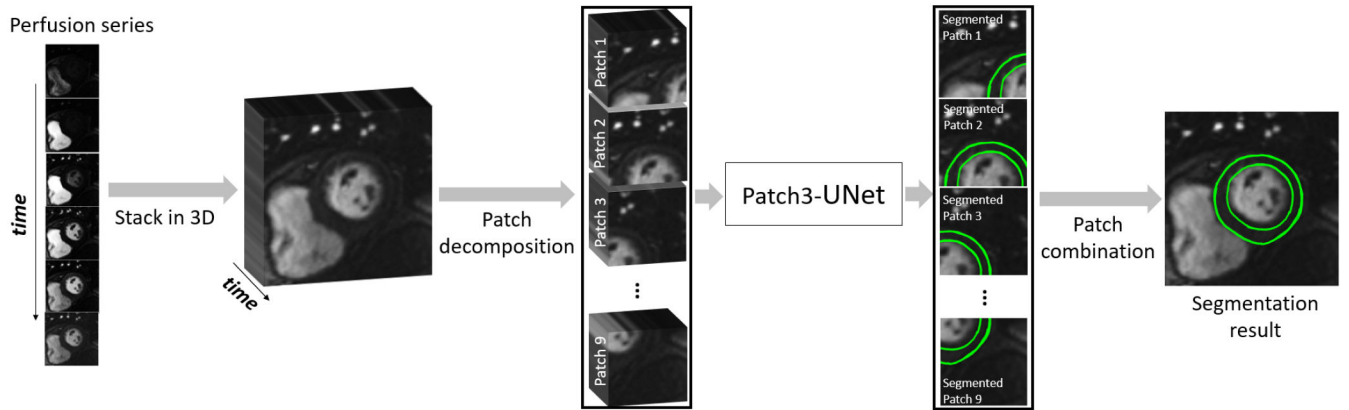


Figure 1. The data processing pipeline for the proposed patch-level approach aimed at segmentation of dynamic CMR perfusion image series.

First, the motion-corrected 2D+time perfusion image series is decomposed into spatiotemporal patches by applying a spatial sliding window (further explained below). The decomposed patches are then fed to a multi-channel vanilla U-Net architecture, which jointly processes dynamic time frames for each patch and detects the myocardial pixels within each 2D patch. The segmented individual patches (output of the network) are combined to yield the segmentation result.

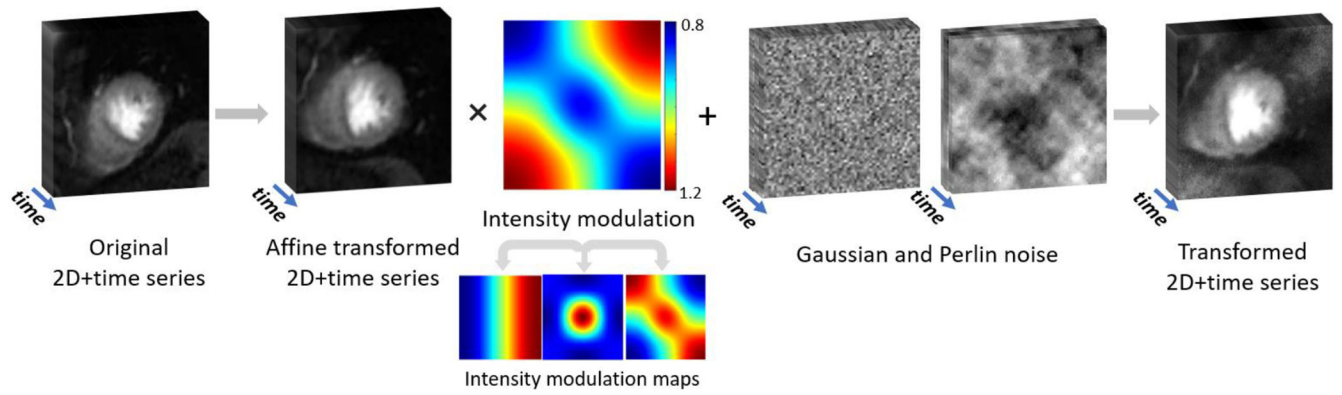


Figure 2. The proposed advanced data-augmentation strategy.

The proposed augmentation technique is performed in three steps. First, various affine transforms (rotation, scaling, vertical and horizontal flips, image scaling, and shear) are applied to the 2D+time series CMR perfusion images. The resulting image series is then modulated with elaborately designed intensity maps. Finally, both temporally-varying Gaussian noise and Perlin noise are added to the transformed time series images. Perlin noise and intensity modulation vary the texture and contrast levels in the time-series images.

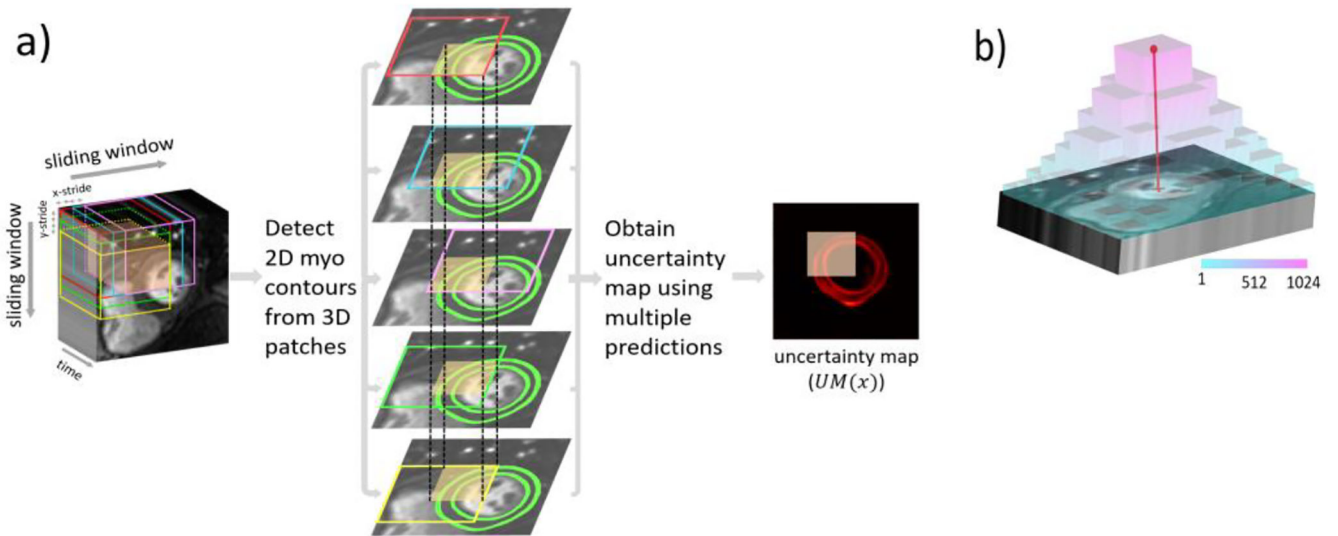


Figure 3. Uncertainty visualization using the proposed patch-level method employing a novel spatial sliding-window approach.

(a) Patch-level testing enables the same region (see the orange volume) to be segmented multiple times. Therefore, we get multiple solutions for the same myocardium (myo) segmentation, and this information is used to create an uncertainty map to assess how uncertain the model is. (b) bar plot illustrates the number of patches that each pixel is in during the spatially sliding-window analysis. Projection in red shows that the center pixel is in 1024 patches.

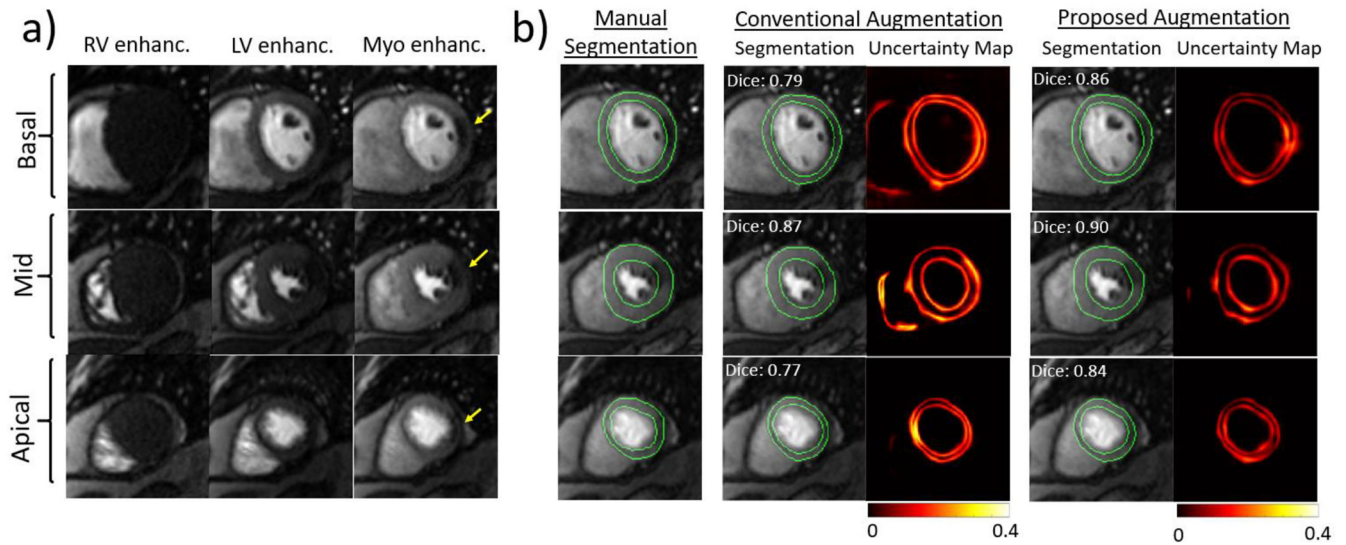


Figure 4. Segmentation results.

a) Stress perfusion CMR time frames acquired in a patient with ischemic heart disease (stress-induced perfusion deficit) are shown at three stages of the first pass of the contrast agent (RV enhancement, LV enhancement, and myocardial enhancement) for three slices (Basal, Mid, Apical). The yellow arrows point to the perfusion deficit. **b)** Segmentation results in the same dataset for two augmentation settings: proposed and conventional, together with manual-expert segmentation. Both conventional and proposed augmentation results employ the proposed patch-level training by extracting patches from the full-ROI images. Segmentation results are accompanied with the uncertainty maps which is a byproduct of the proposed patch-level network. A higher value in the generated uncertainty maps (yellow/white colors) demonstrates high uncertainty at a pixel level.

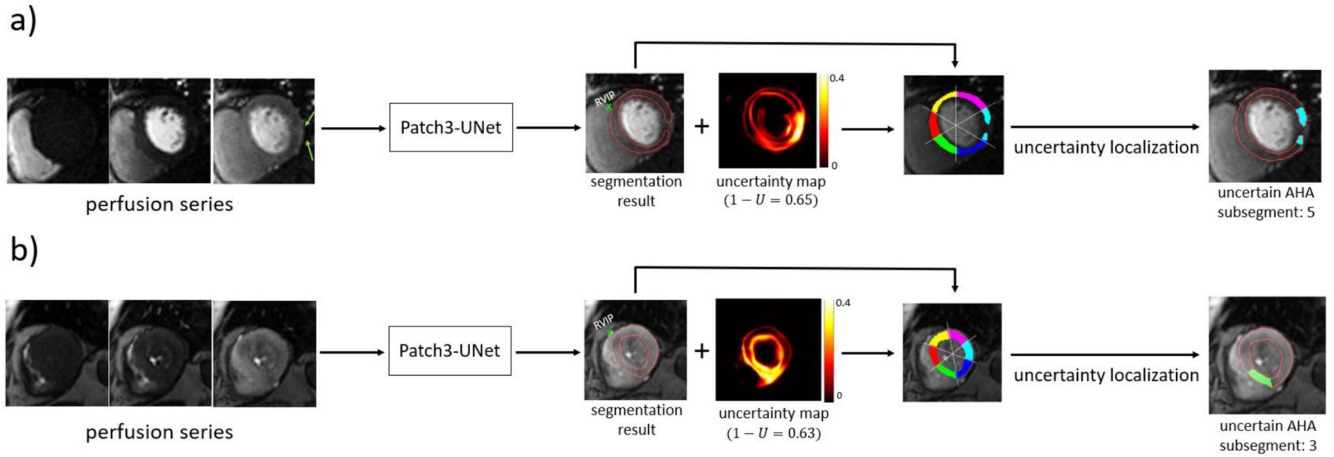
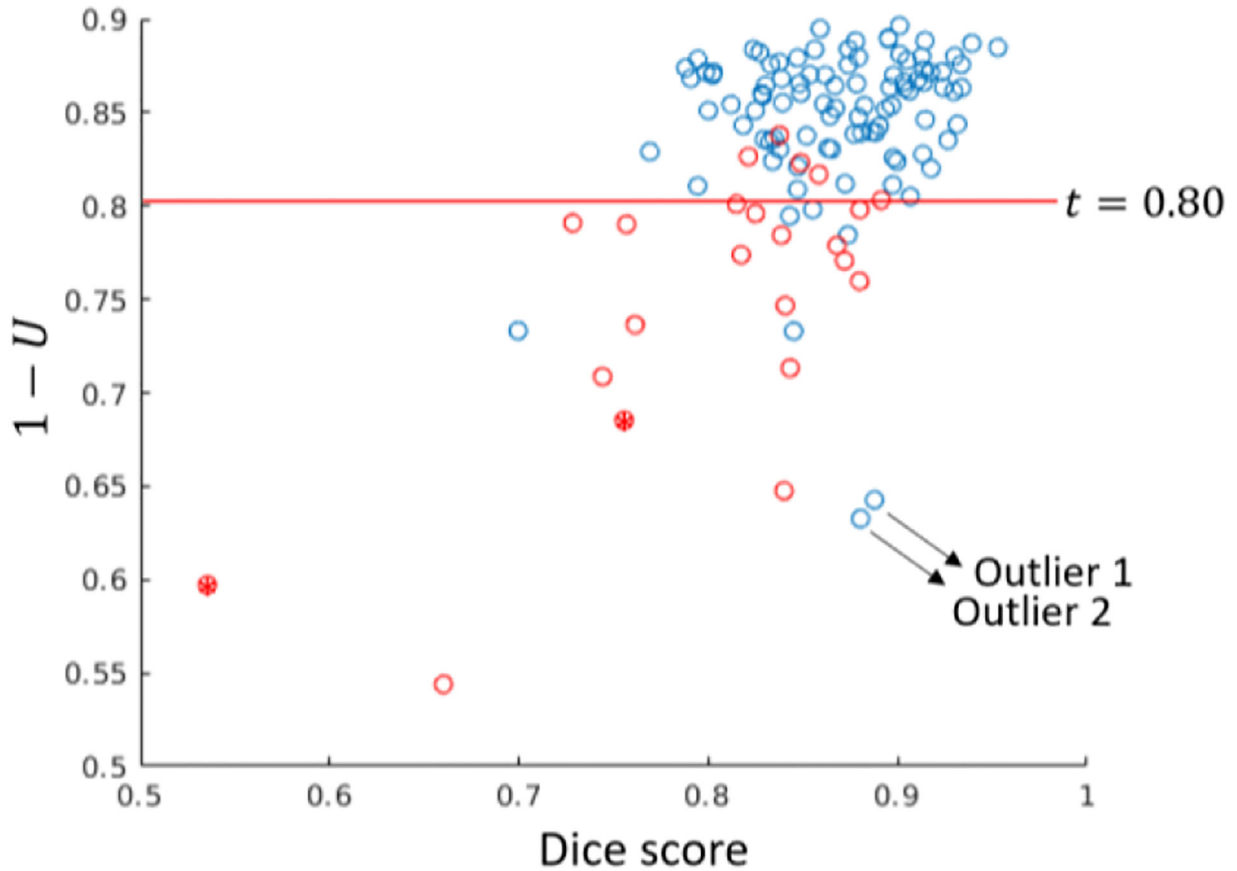


Figure 5. Uncertainty visualization and localization on two challenging slices.

Patch3-UNet takes the 2D+time perfusion series decomposed into patches as input and generates the segmentation (myocardial contours) as well as the uncertainty map. The network is also capable of segmenting the right ventricle (RV) and detection of the RV insertion point (RVIP). Subsequently, the segmentation result is divided into 6 sectors according to the established American Heart Association (AHA) model. Total variation (TV) of each subsegment in the uncertainty map is calculated, and “uncertain segmentation” is localized as the subsegment that has the highest TV. **(a)** shows a challenging Basal slice with a thin layer of epicardial fat in the lateral wall (shown by green arrows on the perfusion series images). Although the segmentation result is not contiguous, patchUNet is mostly uncertain in subsegment 5, which is the most challenging sector to segment. **(b)** shows a Mid slice with small LV blood pool (end systole) from a different patient. As shown on the uncertainty map, the network struggles mostly in the septal and inferior sectors which are very difficult to differentiate from the RV blood pool even for an expert reader. Nevertheless, the proposed *Patch3-UNet* approach segments this challenging test-case successfully.



- : MBF errors > 5% of *patch2-UNet* trained w/ proposed advanced augmentation
- * : MBF errors > 5% of *patch3-UNet* trained w/ proposed advanced augmentation

Figure 6. Out-of-distribution (OOD) test data detection.

The plot shows the Dice score vs. $1-U$ metric. U is defined to quantify the uncertainty from the uncertainty maps with a myocardium area-normalized approach against variations in the heart size. Red circles show the points where 2D patch-level network has >5% myocardial blood flow (MBF) error with respect to ground truth segmentation. Red asterisks are the points where 3D patch-level network has >5% MBF error with respect to expert contouring. Outliers 1 and 2 highlighted on the plot are the two challenging test-cases shown in Fig 5. They result as false positives in the classifier which detects poor motion correction OOD since they do not have any motion correction issues. Indeed, they are also OOD data, but they have their own challenging properties as explained in Fig 5 (small blood pool and thin epicardial fat in the lateral wall) where they would be detected as OOD when described with another definition not based on 2D MBF flow.

Table 1.
Summary of Dice scores for the test dataset.

This table summarizes the mean Dice scores across 40 stress perfusion CMR exams (40 patients; 120 myocardial slices in total) for different augmentation and input-patch dimension settings.

		Whole slice	6-sector	12-sector
Conventional agumentation	patch2-UNet	0.80	0.77	0.62
	patch3-UNet	0.83	0.77	0.66
Proposed agumentation	patch2-UNet	0.82	0.78	0.65
	patch3-UNet	0.86	0.83	0.71

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript