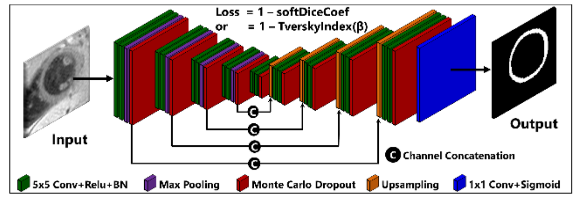
Deep Convolutional Neural Network for Segmentation of Myocardial ASL Short-Axis Data: Accuracy, Uncertainty, and Adaptability

Hung P. Do[1], Yi Guo[2], Andrew J. Yoon[3], and Krishna S. Nayak[4]

[1]Canon Medical Systems USA, Inc.; [2]Snap Inc.; [3]Long Beach Memorial Medical Center, University of California, Irvine; [4]University of Southern California

PURPOSE: Myocardial Arterial Spin Labeling (**ASL**) is a non-contrast quantitative perfusion technique that can assess coronary vascular disease [**1**]. Manual segmentation of left ventricular (**LV**) myocardium is a required step in the post-processing pipeline, and is a major bottleneck due to the low SNR and low and inconsistent blood-myocardium contrast in the source images. This study is motivated by the major success of deep convolutional neural networks (**DCNN**) in automatic medical image segmentation [**2**]. In this work, we (**i**) optimize UNET for myocardial ASL data and evaluate the accuracy of automatic segmentation, (**ii**) quantify the uncertainty of the automatic segmentation model using Monte Carlo (**MC**) dropout as a Bayesian approximation [**3**], and (**iii**) adapt/adjust the predicted masks using the Tversky's similarity measure [**4**] as a loss function.

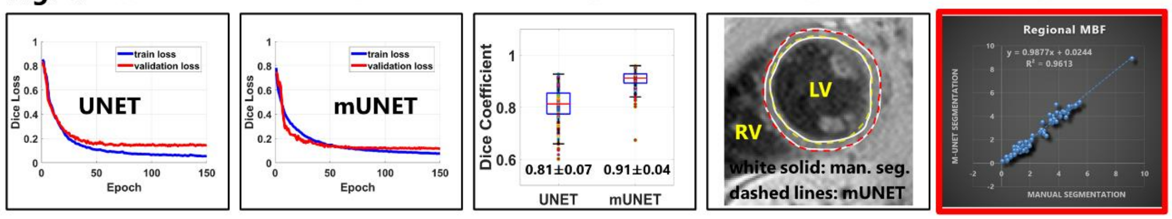


METHODS: A total of 478 ASL images (control and labeled images) from 22 subjects were randomly divided into training and validation sets of 437 and 40 images, respectively. Algorithms were tested on 144 "un-seen" ASL images at rest and Adenosine stress from 6 heart transplant recipients [**5**]. Manual segmentation (ground-truth label) and quantitative myocardial blood flow (**MBF**) were available to evaluate accuracy of the DCNN model.

Original UNET [**2**] and the modified UNET were implemented in Keras with TensorFlow backend. The modified UNET (**mUNET**) was tailored to our specific application, in which the modifications were (**i**) increased filter size to 5x5 to reduce false positive pixels outside the myocardium (white arrow in **Fig2B**), (**ii**) added batch-normalization (**BN**) after every conv layer, and (**iii**) added dropout (**dropout rate = 0.5**) after every two conv layers. Training took 50 min (Adam optimizer, learning rate = 1e-4, batch size = 12, loss function = soft Dice Coefficient) and inferencing took 18ms/image on a 13-Gb memory NIVIDIA K80 GPU. Modified UNET reduced overfitting (**Fig1A and Fig1B**) and significantly increased (**P<0.001**) accuracy compared to the original UNET (**Fig1C**).

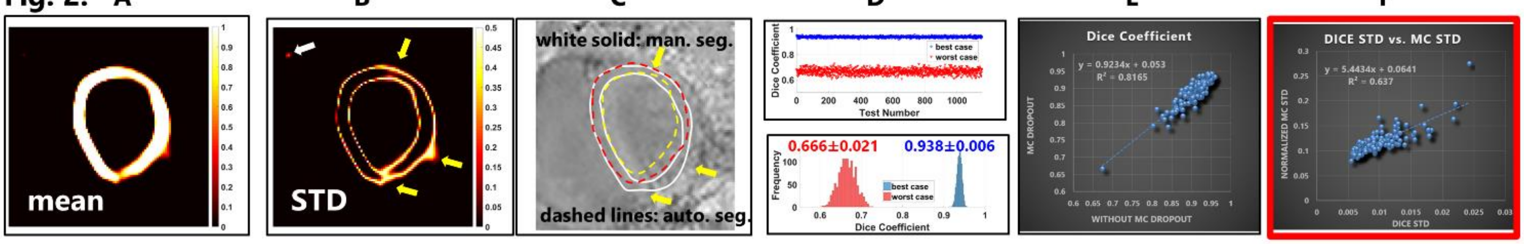
RESULTS:

Fig. 1: A B C D E



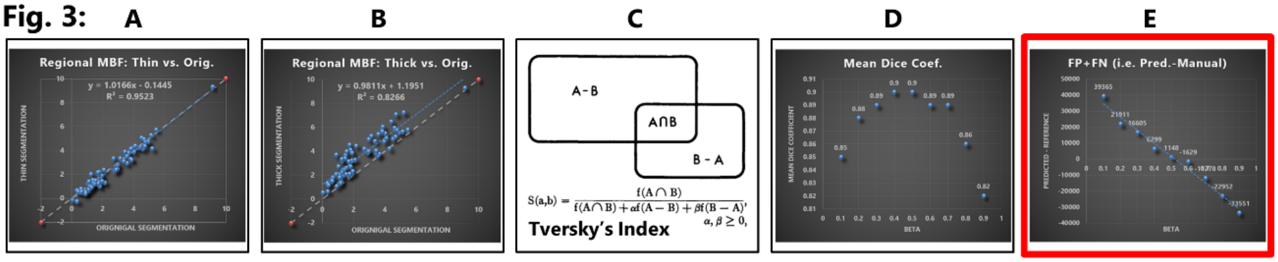
<u>**Accuracy:**</u> Unlike CINE images, labeled ASL images have very low contrast to noise ratio between myocardium and blood pools (see **Fig1D and Fig2C**), a major challenge. mUNET achieved **0.91 ± 0.04** Dice Coefficient on the test set. More importantly, the quantitative MBF measured using the modified UNET predicted masks vs. manual masks were well correlated (**R^2=0.96**) as seen in **Fig1E** (**the key result**).

Fig. 2: A B C D E F



<u>**Uncertainty:**</u> The ground-truth label (manual mask) is often not available therefore the knowledge on model uncertainty is crucial to inform decision. We demonstrated the use of MC dropout [**3**] at inference time as a Bayesian Approximation to estimate model uncertainty and compare it against the error estimated as if the ground-truth was available (**Fig2F**). Mean and uncertainty maps (**Fig2A and Fig2B**) showed the areas with high uncertainty that were edges of the mask and where the prediction and ground-truth label are mis-matched (**yellow arrows**). **Fig1D** showed the output of the mUNET with MC dropout of the best and worst case in the test set. **Fig1E** showed comparable Dice Coefficient between mUNET without vs. mUNET with MC dropout. **Fig2F** (**the key result**) showed good correlation (**R^2=0.64**) between Dice STD (with the knowledge of the ground-truth) vs. MC STD (w/o the knowledge of the ground-truth).

Fig. 3: A B C D E



$$S(a,b) = \frac{f(A \cap B)}{f(A \cap B) + \alpha f(A - B) + \beta f(B - A)}, \quad \alpha, \beta \geq 0,$$

<u>**Adaptability:**</u> Partial volume effect are an issue for most MRI myocardial tissue characterization techniques, including ASL. In ASL, the ventricular blood pools or epicardial fat can have >50x higher signal compared to the ASL signal itself. Careful and conservative segmentation are often required. Soft Dice Coefficient is commonly used as a loss function however it weights false positive (**FP**) and false negative (**FN**) equally that maybe not desired for the task. We demonstrated the use of Tversky's Index [**4**] (**Fig3C**) as a loss function that can control any range of desired FP and FN as seen in **Fig3E**. **Fig3B** showed overestimation of MBF when the masks were thickening (due to partial volume effects) while negligible effects in thinned masks (**Fig3A**). **Fig3D** and **Fig3E** (**the key result**) showed Dice Coefficient and the level of FP and FN of the prediction that can be adjusted by changing the free-parameter beta in the Tversky's Loss.

DISCUSSIONS: This study demonstrates the use of a DCNN model for a specialized cardiac MRI segmentation task and demonstrates modification of the network architecture to improve accuracy, measure uncertainty, and adapt to the specific false positive vs. false negative needs of the application.

<u>**References:**</u> [**1**]. Kober F et al, J Cardiovasc Magn Reson 2016;18:22. [**2**]. Ronneberger O et al, arXiv:1505.04597. [**3**]. Gal Y et al, arXiv:1506.02142. [**4**]. Tversky A et al, Psychol. Rev. 1977;84:327–52. [**5**]. Do HP et al, Magn Reson Med 2017 May;77(5):1975-1980.