Exploring ensemble applications for multi-sequence myocardial pathology segmentation

Markus J. Ankenbrand, David Lohr, Laura M. Schreiber

We tested different loss functions and hyper-parameters using a 2D U-Net architecture (resnet34 backbone) with five-fold cross-validation on the training data.
Pathology specific sequence data (e.g. LGE for scar and T_2 for edema) was used as a sole input for training and in combination with all sequences. We wanted to address the question whether for limited training data it is beneficial to incorporate prior knowledge by predicting classes with their appropriate sequence or if a neural network is able to infer these relationships from a multi-sequence dataset. In addition, we aimed to create a model zoo, combining predictions from models with high performance on individual classes.
Images were cropped to the central 256x256 region as this contained the region of interest in all cases. To improve robustness and learn more general features extensive data augmentation was used, including both MR artifacts (motion, noise) and standard image transformations (zoom, rotation, brightness, contrast). Variations of training data, loss functions and hyper-parameters led to 21 models trained.
The multi-sequence model was trained using all image sequences input via color channels producing pixel-level segmentation for all six classes (background, left ventricle, right ventricle, myocardium, edema, and scar). Cross-entropy as a loss function performed best (metric: dice) for non-pathologic tissue, while pathology weighted focal-loss (0.35 for both scar and edema) had best mean performance on scar and edema.
The mean Dice scores over all cross validation folds using the multi-sequence model were: Dice(LV)=0.853, Dice(RV)=0.787, Dice(myocardium)=0.695, Dice(scar)=0.438, Dice(edema)=0.227. The specific models reached dice scores of: Dice(scar)=0.479, Dice(edema)=0.276
These results indicate that the employed neural network is capable of learning multi-sequence segmentation end-to-end. Combining different outputs from a model zoo based on objective criteria proofed difficult.