Do not Treat Boundaries and Regions Differently: An Example on Heart Left Atrial Segmentation

Zhou Zhao Élodie Puybareau Nicolas Boutry Thierry Géraud

EPITA Research and Development Laboratory (LRDE), France

zhou.zhao@lrde.epita.fr

January 15, 2021

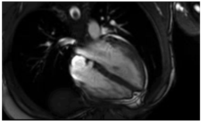
Introduction

Cardiovascular diseases are the leading cause of death globally.
— — — according to the World Health Organization (WHO).

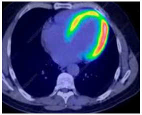
Medical imaging

- Computed tomography (CT)
- Magnetic resonance imaging (MRI)
- Positron emission tomography (PET)
- Single photon emission computed tomography (SPECT)
- Ultrasound (US)

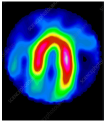
CT



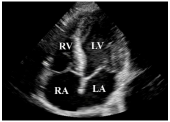
MRI



PET



SPECT



US

MRI: Great contrast between soft tissues and relatively high spatial resolutions.

The difficulties of cardiac segmentation on the MR Images are shown as follow:

- **Poor contrast**

  —between myocardium and surrounding structures
- **Brightness**

  —in left ventricular/right ventricular cavities due to blood flow
- **Non-homogeneous partial volume**

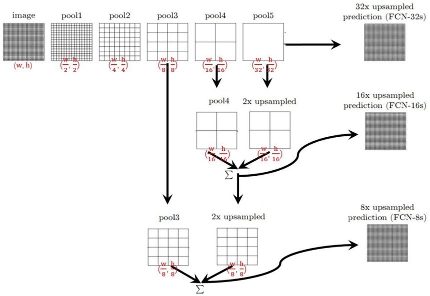
  —limited CMR resolution(1.5T, 3.0T...) along the long-axis
- **Noise**

  —motion artifacts and heart dynamics
- **Shape and intensity variability**

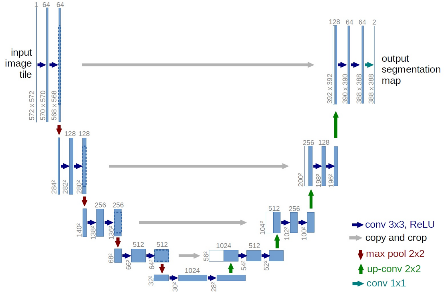
  —different patients and pathologies

Fully Convolutional Networks (FCN) and UNet

– using upsampling layers and combining the feature maps from lower to higher resolutions.



FCN



UNet

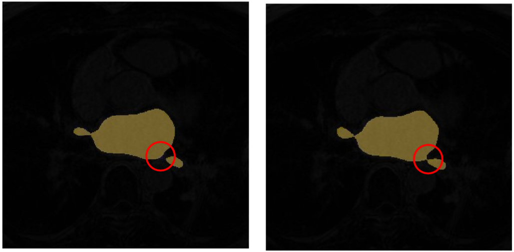
Main Issues

– the lack of precision regarding the boundaries of the segmented objects.

– the loss of small objects and small parts of objects.

Loss Function: Dice and Cross-entropy (CE)

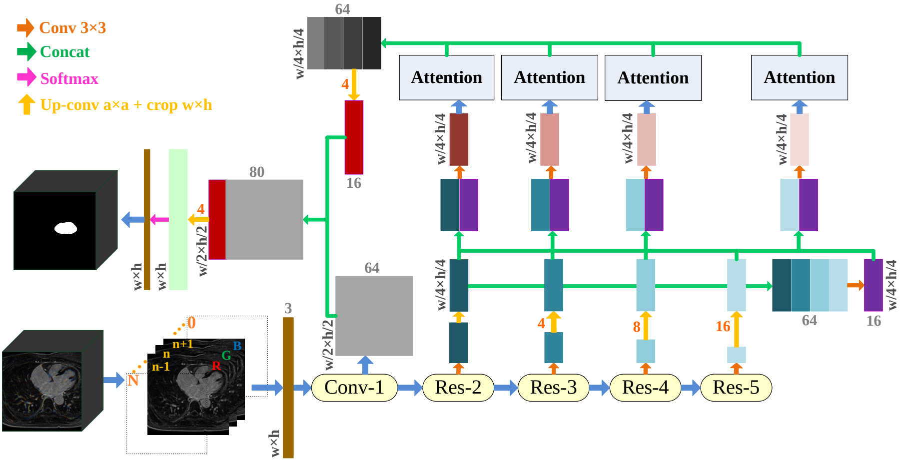
- CE: highly unbalanced problems.
- dice: difficulty dealing with very small structures.
- region-based.



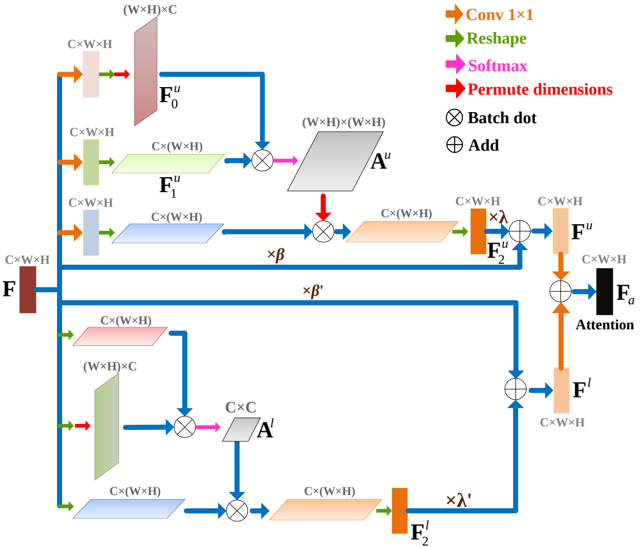
There are three challenging problems applied on cardiac imaging:

- enlarging the receptive field of a CNN.
- improving the segmentation accuracy on small parts of objects.
- balancing the importance of the regions and the boundaries of objects.

Methodology



Architecture of our networks.



Attention Module. λ, λ', β and β' as hyperparameters, which is trained like the convolutional kernel. They decrease the weight of the unimportant feature maps.

Hybrid Loss

The hybrid loss consists of two parts: region loss and boundary loss.

Region Loss: $\ell_{\mathrm{R}} = \ell_{\mathrm{CCE}} + \ell_{\mathrm{SSIM}} + \ell_{\mathrm{DC}}$

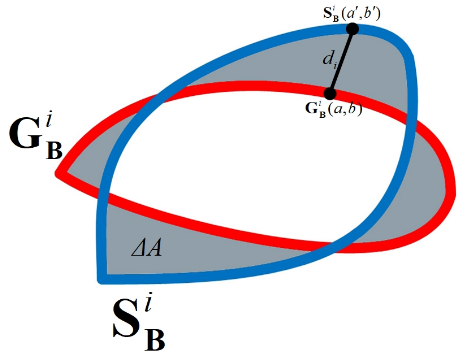
- **Categorical Cross Entropy (CCE)**-pixel-level:
  $\ell_{\mathrm{CCE}} = -\sum_{i=1}^{C} \sum_{a=1}^{H} \sum_{b=1}^{W} y_{(a,b)}^{i} \ln y_{*}{}_{(a,b)}^{i}$,
  − − −multi-class classification and segmentation.

- **Structural Similarity (SSIM)**-patch-level:
  $\ell_{\mathrm{SSIM}} = 1 - ((2\mu_{\mathrm{S}}\mu_{\mathrm{G}} + \varepsilon_1)(2\sigma_{\mathrm{SG}} + \varepsilon_2)) \,/\, ((\mu_{\mathrm{S}}^2 + \mu_{\mathrm{G}}^2 + \varepsilon_1)(\sigma_{\mathrm{S}}^2 + \sigma_{\mathrm{G}}^2 + \varepsilon_2))$,
  − − −assessing image quality and is used to capture the structural information.

- **Dice Coefficient (DC)**-map-level:
  $dice_i = (\epsilon + 2\sum_{n=1}^{N_i} y_n^i \, y_{*}{}_n^i) \,/\, (\epsilon + \sum_{n=1}^{N_i} (y_n^i + y_{*}{}_n^i))$;
  $\ell_{\mathrm{DC}} = 1 - \sum_{i=1}^{C} dice_i / (N_i + \epsilon)$
  − − −measuring the similarity between two sets.

The region loss guides the network to learn the transformation in a three-level hierarchy.

Boundary Loss: $\ell_{\mathrm{B}} = \sum_i^C \int_{\mathbf{G}_{\mathrm{B}}^i} \left\| \mathbf{S}_{\mathrm{B}}^i(a', b') - \mathbf{G}_{\mathrm{B}}^i(a, b) \right\|^2 \mathrm{d}(a, b)$



The boundary loss function is to optimize the segmentation result.

Experiments and Results

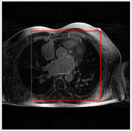
Dataset description

- 100 annotated 3D MRIs

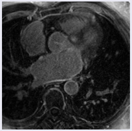
  − − −80 cases for training and 20 cases for validating.
- Pixel spacing: $0.625 \times 0.625 \times 0.625$ mm/pixel.
- Image sizes: $88 \times 576 \times 576$ and $88 \times 640 \times 640$ pixels.

Pre-processing

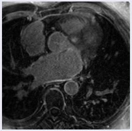
- Cropping each slice to 346×346 pixels.
- Gaussian normalization.
- Stacking 3 successive 2D frames.



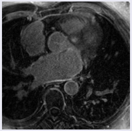
Initial image



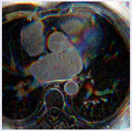
$n-1$



n



$n+1$



“3D-Like”

Post-processing

- Keeping the greatest connected component of the segmentation.
- Padding with zeros to get back a initial width and height of a slice.

Implementation and Experimental Setup

- Keras/TensorFlow using a Nvidia Quadro P6000 GPU.
- Hybrid loss function

  −softmax to get a probability distribution over classes.
- Adam optimizer

  −batchsize=3, $\beta 1$=0.9, $\beta 2$=0.999, epsilon=0.001, lr = 0.01.
- Epoch=30.
- Evaluation method:

  −Dice coefficient;

  −95% Hausdorff distance (95HD);

  −Average Hausdorff distance (AHD).

Results on training datasets

Comparison of our method and other state-of-the-art architectures using a 5 fold cross-validation.

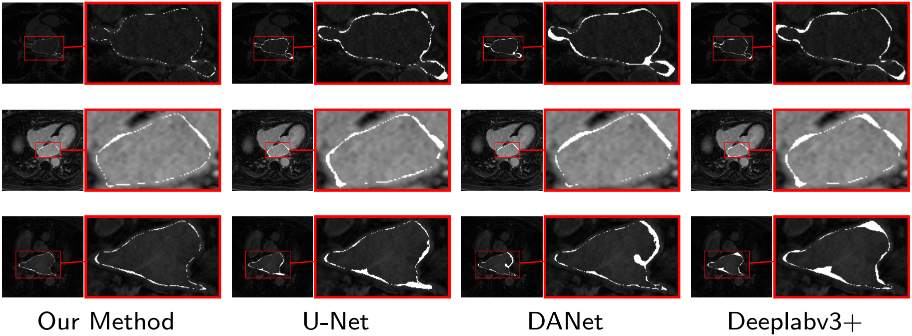
Method	Att. Module	Hyb. Loss	DC/%	95HD/mm	AHD/mm
U-Net [3]			88.556(±2.586)	4.447(±0.996)	0.212(±0.077)
		✔	89.613(±2.257)	4.169(±0.960)	0.210(±0.118)
DANet [2]			84.229(±3.774)	6.145(±2.341)	0.514(±0.477)
		✔	87.584(±2.765)	4.903(±1.448)	0.280(±0.179)
Deeplabv3+ [1]			85.444(±3.079)	5.872(±2.345)	0.504(±0.614)
		✔	87.556(±1.155)	5.210(±1.087)	0.273(±0.074)
Our Method			90.774(±1.568)	3.312(±1.277)	0.158(±0.092)
	✔		91.326(±1.174)	3.097(±0.810)	0.143(±0.055)
	✔	✔	**91.792(±1.065)**	**2.868(±0.667)**	**0.130(±0.042)**

Chen, L.C., et al.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proc. of ECCV. pp. 801–818 (2018)

Fu, J., et al.: Dual attention network for scene segmentation. In: Proc. of CVPR. pp. 3146–3154 (2019)

Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: Proc. of MICCAI. LNCS, vol. 9351, pp. 234–241. Springer (2015)

Comparison of the proposed method and other state-of-the-art architectures. The white pixels are the differences between the prediction and the GT.



Conclusion

- **Treat boundaries and regions fairly**
- **Hybrid loss**

  −considering regions and boundaries of objects equally by combining region loss with boundary loss.
- **Attention module**

  −preventing the interferences between the surrounding similar tissues and to capture large-scale and thiner structures
- **Temporal-like method**

  −taking advantage of the temporal information by stacking 3 successive 2D frames.

Thank you!