CROSS-PYRAMID CONSISTENCY REGULARIZATION
FOR SEMI-SUPERVISED MEDICAL IMAGE SEGMENTATION

Ing. Matúš Bojko bojko.matus@gmail.com
Supervisor: Ing. Marek Jakab, PhD.
Slovak University of Technology in Bratislava

In this paper from diploma thesis, we propose a hybrid consistency learning approach to effectively exploit unlabeled data for semi-supervised medical image segmentation, by leveraging Cross-pyramid consistency regularization (CPCR) between two decoders. Firstly, we design a hybrid Dual Branch Pyramid Network (DBPNet) consisting of encoder and two slightly different decoders, each producing a pyramid of perturbed auxiliary predictions from multiple resolution scales. Secondly, we present a learning strategy for this network named CPCR, that combines existing consistency learning and uncertainty minimization approaches on main output predictions of decoders with our novel regularization term. More specifically, in this term we extend the soft label supervision setting to pyramid predictions across decoders to support knowledge distillation in deep hierarchical features.

Introduction

Advancements in deep learning, particularly CNNs and recent Vision Transformers (ViTs), have enabled remarkable performance across a wide range of medical image segmentation tasks. However, fully supervised methods require large and well-annotated datasets for training. Acquiring dense and precise pixel or voxel-level annotations for medical images is both expensive and time-consuming for domain experts. As a result, methods that leverage unlabelled data, such as semi-supervised approaches, have gained significant research attention and become increasingly important. Semi-supervised learning (SSL) enables to train powerful models with assumption of limited carefully labelled data and large number of unlabelled data.

Methodology

In this work, we propose a hybrid network-level consistency learning method for efficient semi-supervised medical image segmentation.

Dual Branch Pyramid Network

Firstly, we propose a hybrid Dual Branch Pyramid Network (DBPNet), to not only obtain diverse predictions from main outputs of decoders (branches), but also obtain auxiliary pyramid predictions from different scales of each decoder.

Proposed DBPNet (in Fig.1) is based on standard U-Net structure with skip-connection and separated decoder to 2 branches. We follow strategy of MCNet+ for obtaining output disperacies by using dropout and different upsampling method in each branch:

- by using transpose convolution
- or 1x1 convolution with bilinear upsampling

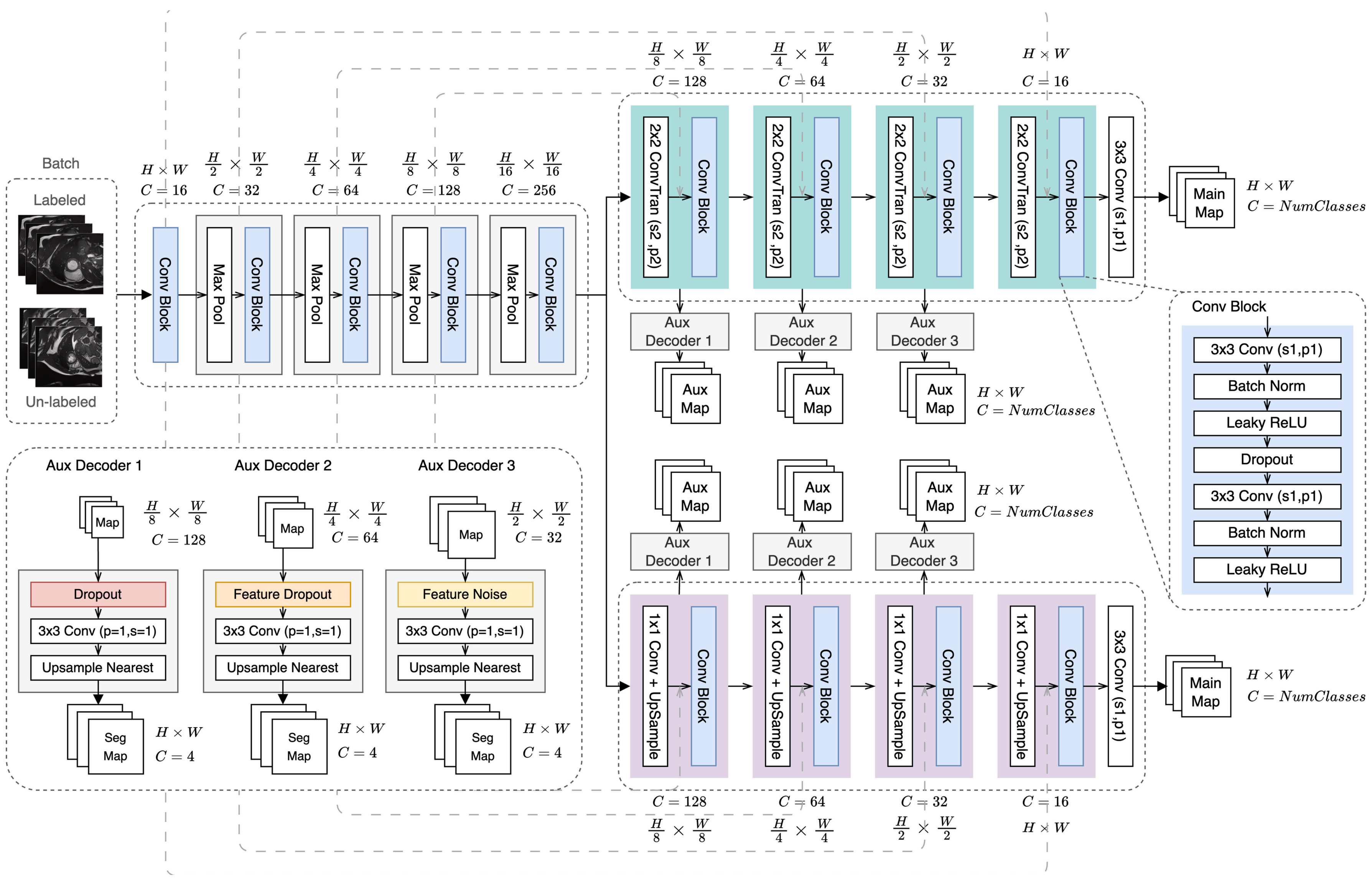

Fig.1 Visualization of architecture details of DBPNet.

To produce the predictions at different scales (pyramid), we added **auxiliary segmentation heads** from URPC. To produce output disperacies at each scale, every head is proceeded by different perturbation approach, either by applying: dropout, feature dropout or feature noise.

Cross pyramid consistency regularization

For this scheme we propose a learning strategy that in single training iteration leverages both labelled and unlabelled images in end-to-end manner for consistency learning and knowledge distillation.

In CPCR, similarly as MTNet, each branch serves as a teacher of the other branch (and vice versa) using soft label supervision between main outputs of decoders. In addition, CPCR applies this supervision on the auxiliary pyramid predictions between pairs of predictions across the adjacent scales of decoders.

This strategy consist of 3 regularization loss terms for consistency learning and uncertainty minimization:

Main consistency loss: enforces consistency between predictions from main outputs of each branch (on scale 4). In this term the main prediction from one branch (TR) serves as pseudo-label for second branch (UP) and vice versa. To address noise present from incorrect predictions, we follow MTNet by using KL divergence as distance metric and T-softmax for generating pseudo-labels as soft probability maps.

$$\mathcal{L}_{\text{con}}^{main} = KL(\tilde{P}_{TR}^{(4)}, \tilde{P}_{UP}^{(4)}) + KL(\tilde{P}_{UP}^{(4)}, \tilde{P}_{TR}^{(4)}) \qquad \tilde{\mathbf{p}}_c = \frac{exp(\mathbf{z}_c/T)}{\sum_c exp(\mathbf{z}_c/T)}$$

Auxiliary consistency loss: The auxiliary predictions upsampled from different lower resolutions still contain different spacial frequencies i.e. they capture the low-frequency component of the segmentation and contain low-level hierarchical features. We introduce this term to enforce consistency between these features across decoders by similarely applying soft label supervision between predictions at each scale of branches to potentially support knowledge distillation across deep layers.

$$\mathcal{L}_{\text{con}}^{aux} = \frac{1}{3}\sum_{s=1}^{3}[KL(\tilde{P}_{TR}^{(s)}, \tilde{P}_{UP}^{(s)}) + KL(\tilde{P}_{UP}^{(s)}, \tilde{P}_{TR}^{(s)})]$$

Overall loss function is weighted sum of proposed terms, supervised loss using dice coefficient, and existing uncertainty minimization term applied on main outputs that leverages entropy minimization to futher encouradge inter-decoder consistency. Regularization terms are applied on all data (labelled and unlabeled).

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{sup}} + 0.1 \cdot (\mathcal{L}_{\text{con}}^{main} + \mathcal{L}_{\text{um}}) + \lambda \cdot \mathcal{L}_{\text{con}}^{aux}$$

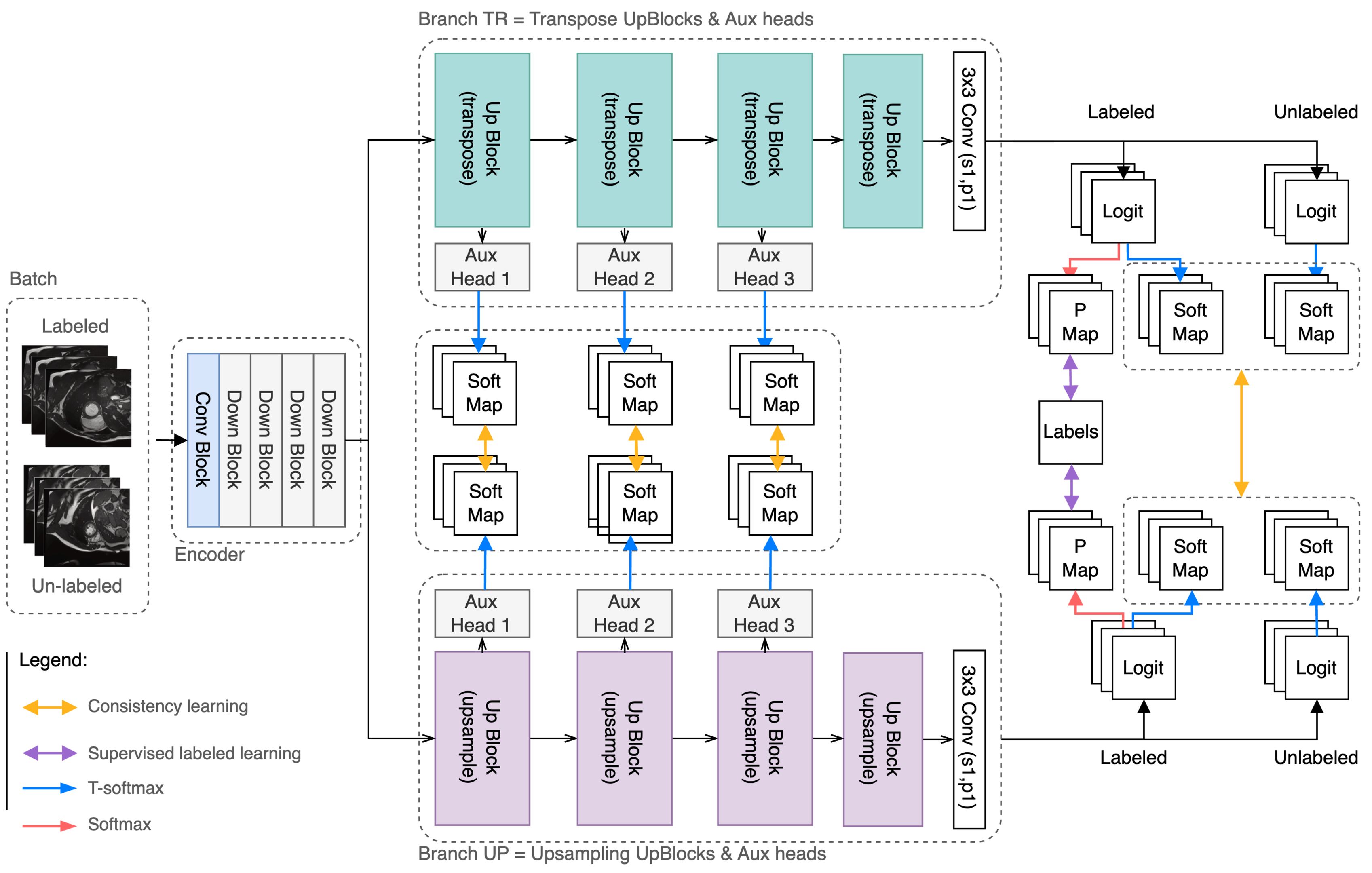

Fig.2 Overview of the proposed learning strategy via CRPC on DBPNet.

Dataset

For training and evaluation we utilized commonly used public benchmark dataset **ACDC (Automated Cardiac Diagnosis Challenge)**, consisting of 200 annotated short-axis cardiac cine-MRI scans from 100 patients. Due to large spacing (5mm) and possible shift between slices (due to respiration), the 3D scans were handled as multi-class segmentation task of the 2D short-axis slices.

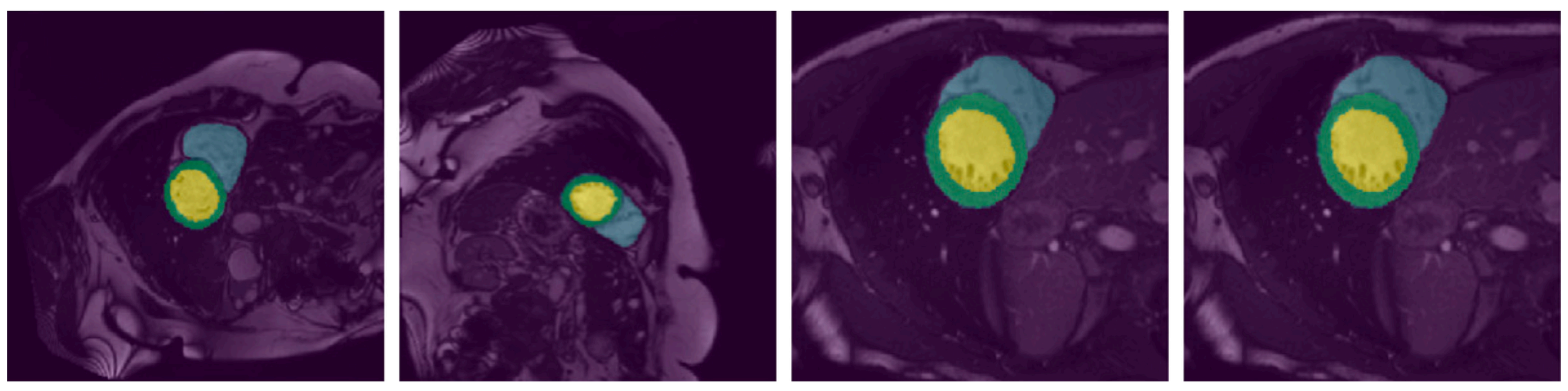

Fig.3 Examples of ACDC dataset slices overlaid with segmentation masks indicating their classes: left ventricular endocardium (yellow), right ventricular endocardium (blue), left ventricular myocardium (green), and background (purple).

Comparison results

We experimented with typical setup of 10% labelled and the rest unlabelled data. Experimental results show that DBPNet with CPCR outperforms five state-of-the-art SSL methods and have comparable performance to the most recent ones.

Method:	Metrics:				Complexity:
	DSC(%)↑	IoU(%)↑	95HD↓	ASD↓	Params.(M)
U-Net (10%)	77.34	66.20	9.18	2.45	1.81
U-Net (All)	91.65	84.93	1.89	0.56	1.81
UA-MT (2019)	81.58	70.48	12.35	3.62	1.81
SASSNet (2020)	84.14	74.09	5.03	1.40	1.81
DTC (2021)	82.71	72.14	11.31	2.99	1.81
URPC (2021)	81.77	70.85	5.04	1.41	1.83
MC-Net (2021)	86.34	76.82	7.08	2.08	2.58
MC-Net+ (2022)	87.10	78.06	6.68	2.00	1.81
DVCPS (2025)	**88.76**	**80.36**	5.03	1.43	–
DBPNet (Ours)	88.11	79.45	**4.12**	**1.11**	1.83

Table.1 Comparison of proposed DBPNet and CPCR with six SSL methods on ACDC test set, using 10% of all annotations for training. Note that all used similar settings and same data split.

This paper is part of diploma thesis: Medical image segmentation using deep learning methods.