Large-Scale Pre-training for Segmentation of Organs-at-Risk and Gross Tumor Volume of NPC for Radiotherapy Planning

Constantin Ulrich[1,3,4] and Klaus H. Maier-Hein[1,4]

1 Division of Medical Image Computing, German Cancer Research Center (DKFZ), Heidelberg, Germany
2 Pattern Analysis and Learning Group, Department of Radiation Oncology, Heidelberg University Hospital, Heidelberg, Germany
3 National Center for Tumor Diseases (NCT), NCT Heidelberg, A partnership between DKFZ and University Medical Center Heidelberg
4 Medical Faculty Heidelberg, University of Heidelberg, Heidelberg, Germany

GERMAN CANCER RESEARCH CENTER IN THE HELMHOLTZ ASSOCIATION

Research for a Life without Cancer

Solution for SegRap 2023 challenge

- **Training dataset consists of 120 cases including both a CT and a contrast CT scan**
- **2 Subtasks: Segmentation of 45 overlapping Organs at Risk and of two Tumor classes**
- **Our solution employs a large scale supervised pre-training using “MultiTalent”**
- **Task specific nnU-Net modifications for fine-tuning on both subtasks**

Initial experiments

- initial experiments to find right modality for each task
  - CT for OAR (Task 1)
  - contrast CT for tumors (task 2)
- a few task specific modifications
  - tested foreground oversampling strategies
  - increased GPU limit for nnU-Net’s experiment planning
  - extended training schedule to improve convergence

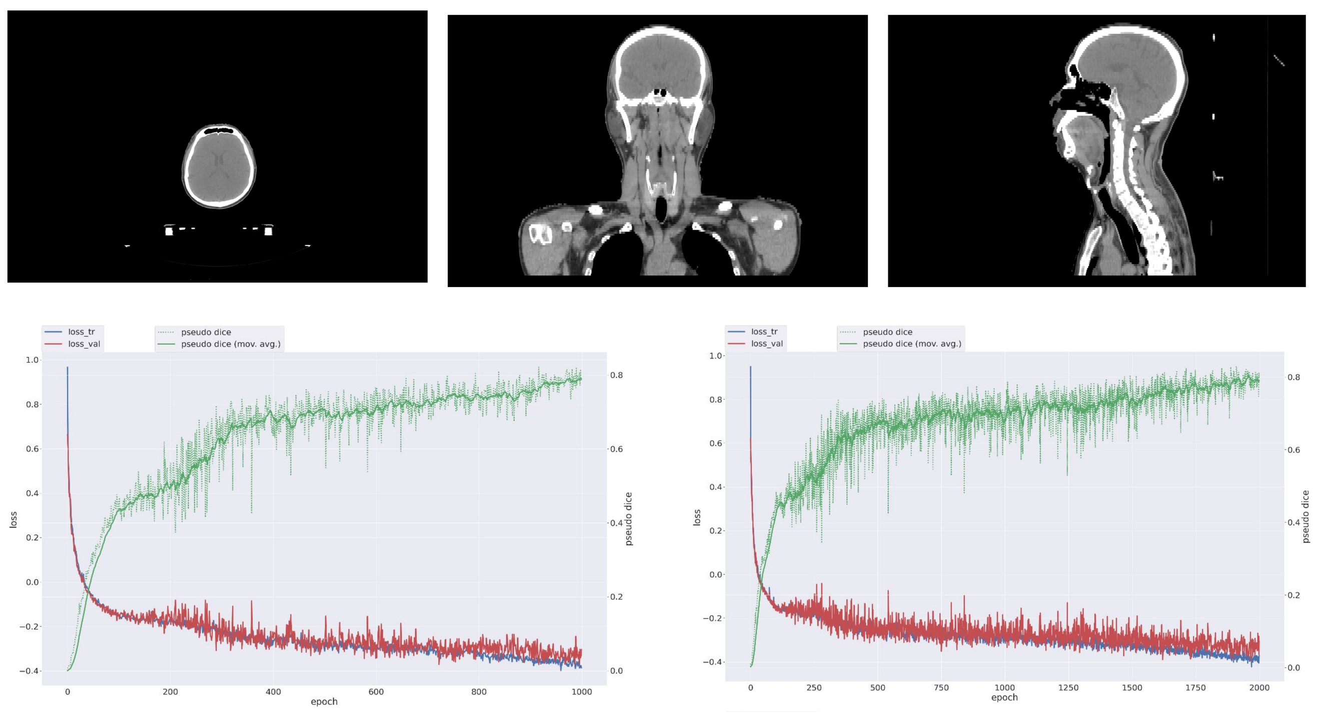

Fig. 4: Motivation for the task specific modifications: Top: The small head region covers only a small part of the CT volume. Bottom: For segmentation of the OARs, the default nnU-Net seams not to be fully converged after 1000 epochs.

MultiTalent
A novel strategy for a combined training using multiple partially labeled datasets for a comprehensive structure segmentation

Pre-training

- extended dataset collection (included totalsegmentator and AMOS dataset) resulting in >2800 images
- Preprocessed as nnU-Net suggested for the SegRap dataset

Fine-tuning

- fine-tuning using default nnU-Net configurations unless stated otherwise
- finetuning schedule:
  1. 10 epoch only heads warm up, lr linearly increased to 0.001
  2. 50 epochs whole network warm up, lr linearly increased to 0.01
  3. default nnU-Net scheme

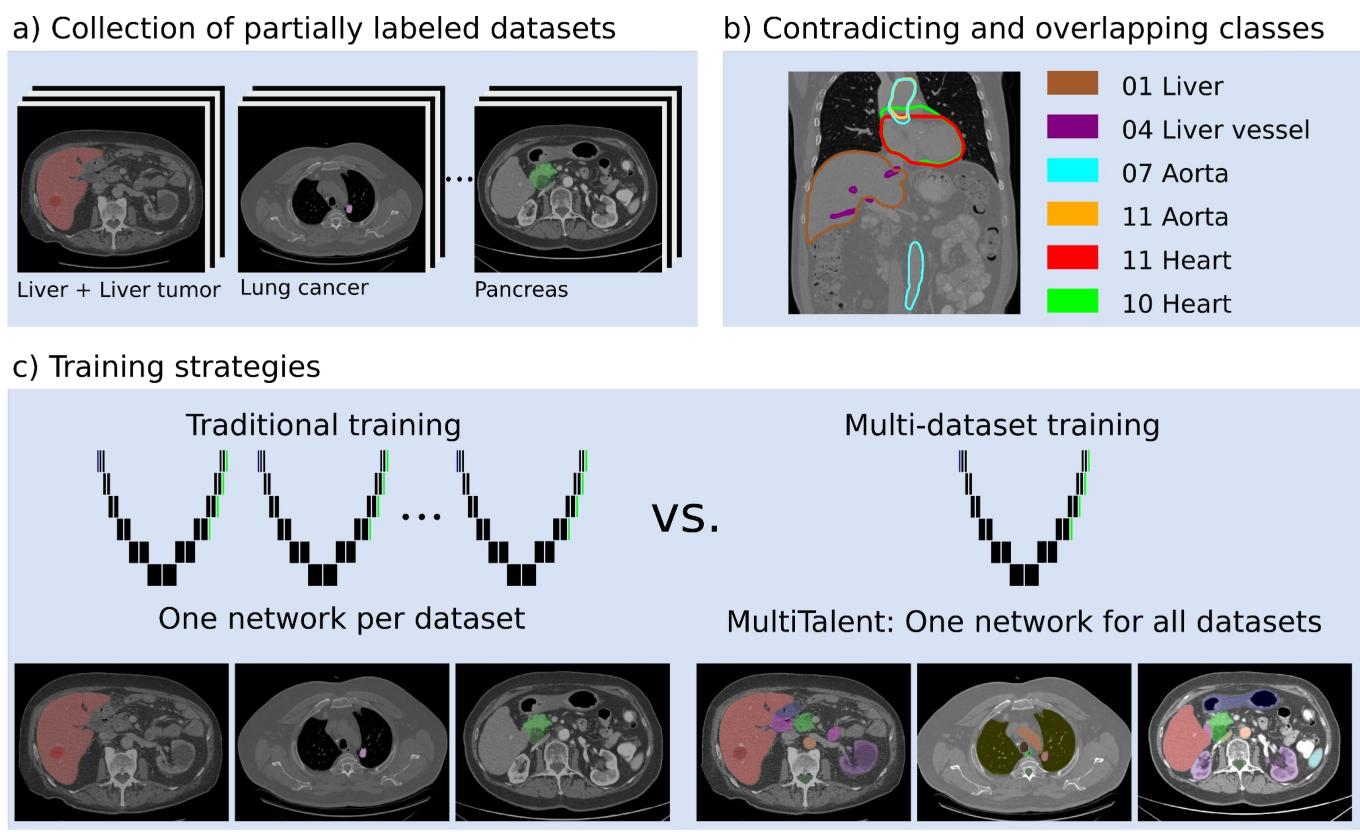

Fig. 1: MultiTalent deals with different annotation protocols for the same target classes

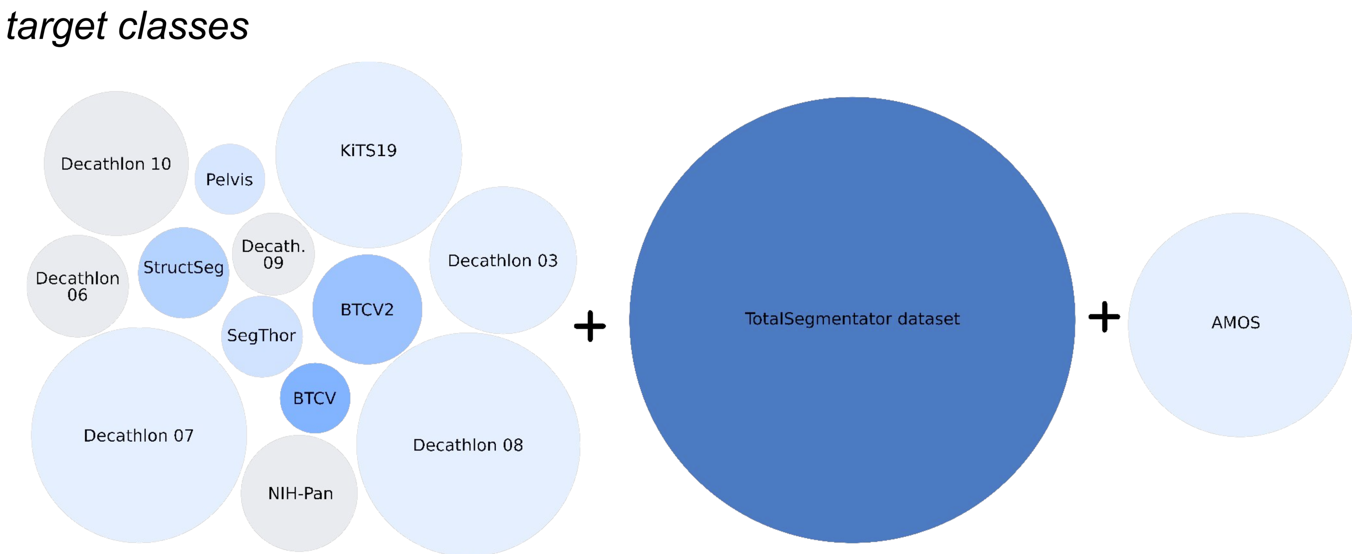

Fig. 2: Overview over all datasets used for pre-training. Totalsegmentator has 1204, Amos 300 and the original dataset collection has 1460 images.

Results - Task 1

Architecture	epochs	patch size	batch size	over-sampling	pre-training	Dice 1-fold	Dice 5-fold
U-Net	1000	[28, 256, 256]	2	default	no	77.88	-
Resenc U-Net	2000	[32, 320, 256]	2	default	no	78.04	-
Resenc U-Net	2000	[32, 320, 256]	2	33%	no	72.63	-
Resenc U-Net	2000	[32, 320, 256]	2	default	yes	82.48	82.02
Resenc U-Net	2000	[32, 320, 256]	2	33%	yes	82.76	82.23
Resenc U-Net	2000	[32, 320, 256]	4	default	yes	82.54	82.34
Resenc U-Net	2000	[32, 320, 256]	4	33%	yes	82.26	**82.53**

Table 1: Dice averaged over 54 non-overlapping classes

Task 2

Architecture	epochs	patch size	batch size	over-sampling	pre-training	Dice 1-fold	Dice 5-fold
U-Net	1000	[28, 256, 256]	2	default	no	75.02	-
U-Net	2000	[28, 256, 256]	2	0.33%	no	75.78	76.44
U-Net	2000	[28, 256, 256]	4	0.33%	no	75.22	76.42
U-Net	2000	[32, 320, 256]	2	0.33%	no	75.35	76.19
Resenc U-Net	2000	[32, 320, 256]	2	default	no	74.39	-
Resenc U-Net	2000	[32, 320, 256]	2	default	yes	74.66	-
Resenc U-Net	2000	[32, 320, 256]	2	0.33%	yes	74.90	-
Resenc U-Net	2000	[32, 320, 256]	4	0.33%	yes	75.70	**76.44**

Table 2: Dice averaged over 2 tumor classes

Out of Competition for task 1: Not valid submission because of issues regarding the RAM and time limit.
Test Set: DSC avg: 84.77, NSD avg: 84.41

Test Set (5 fold ensemble: no test time augmentation):
GTVp: Dice avg.: 77.71, NSD avg.: 35.60
GTVnd: Dice avg.: 69.18, NSD avg.: 64.76

constantin.ulrich@dkfz-heidelberg.de