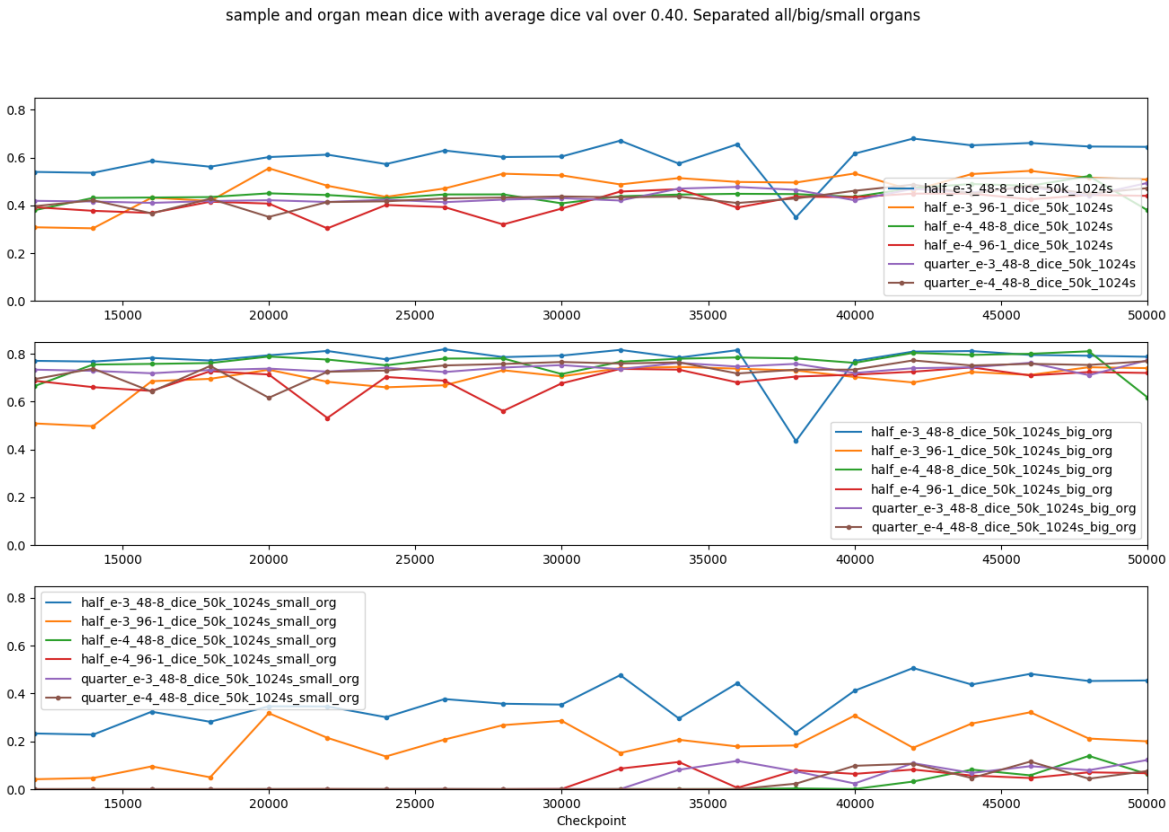
KW47 NiftyNet

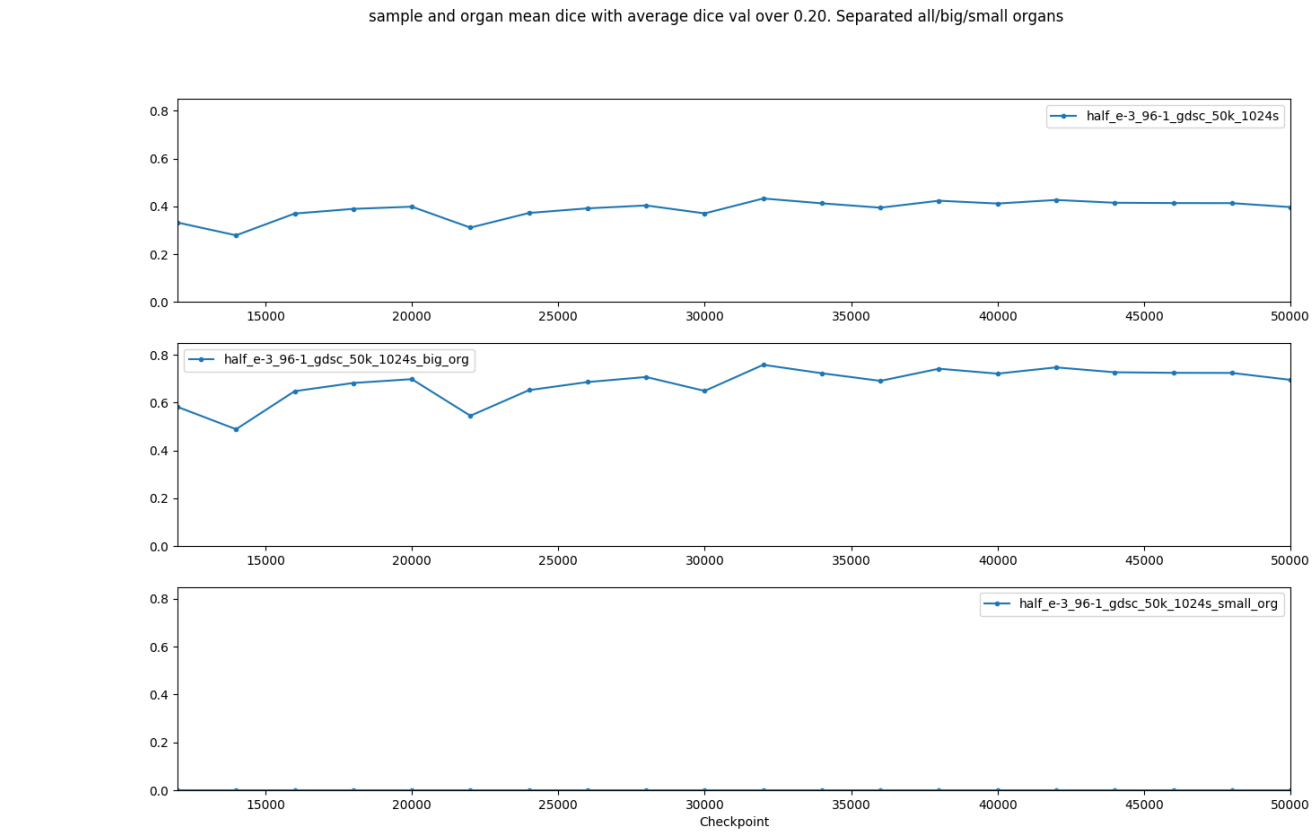
Stage 1 results!

Loss funktion:

Dice, average threshold 0.4



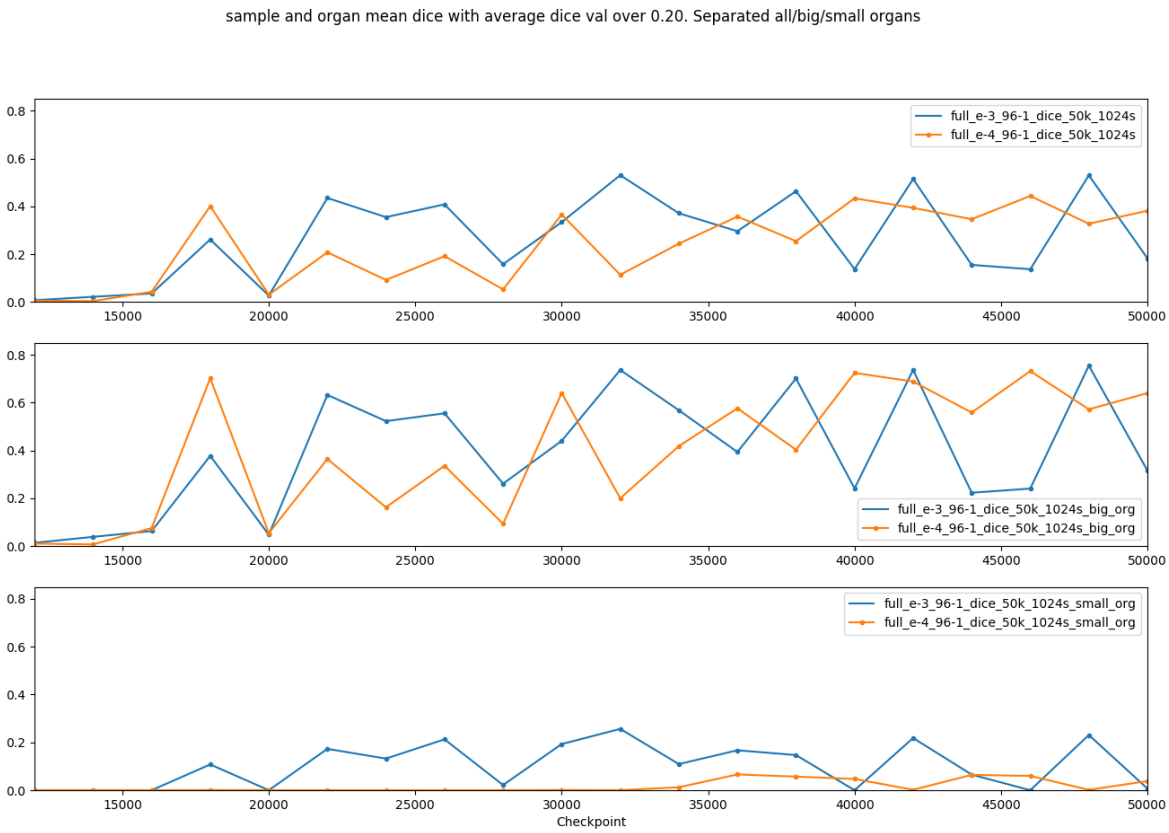
Best results using the dice loss, for half and quarter sized cts and learning rates of 0.001 and 0.0001

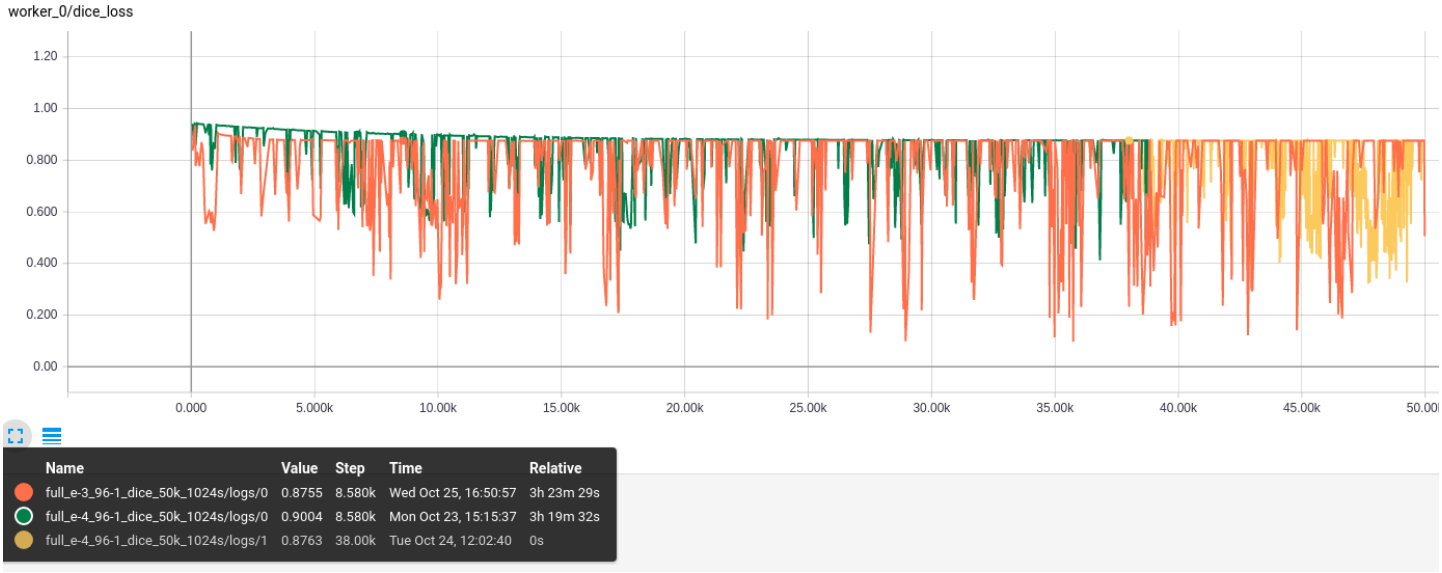


Just one result, all other gdsc based models perform wore than the threshold, for the one working model half_e-3_96-1_gdsc_50k_1024s. Small organs are not detected at all, also the loss does not seem to converge!!

Volume size:

Full size,

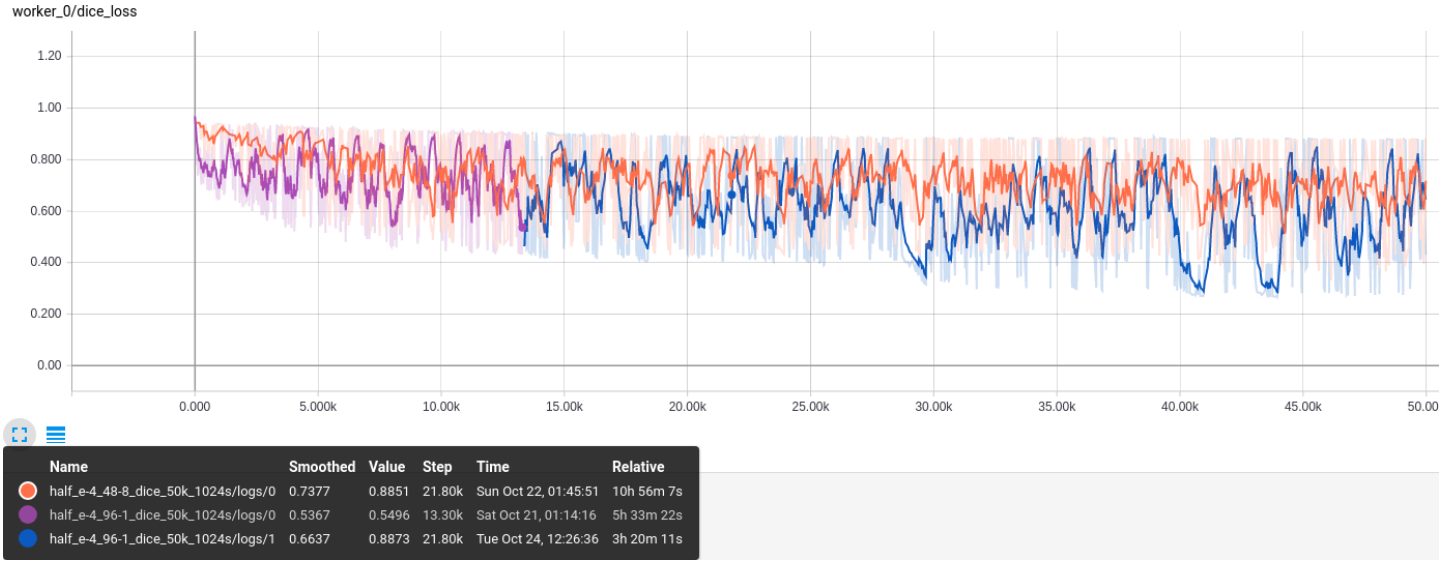




Only two models achive an average test dice performance over 0.2. Results have a lot of noise and small organs are detected bad. The loss is not as noisy as others but does not converge.

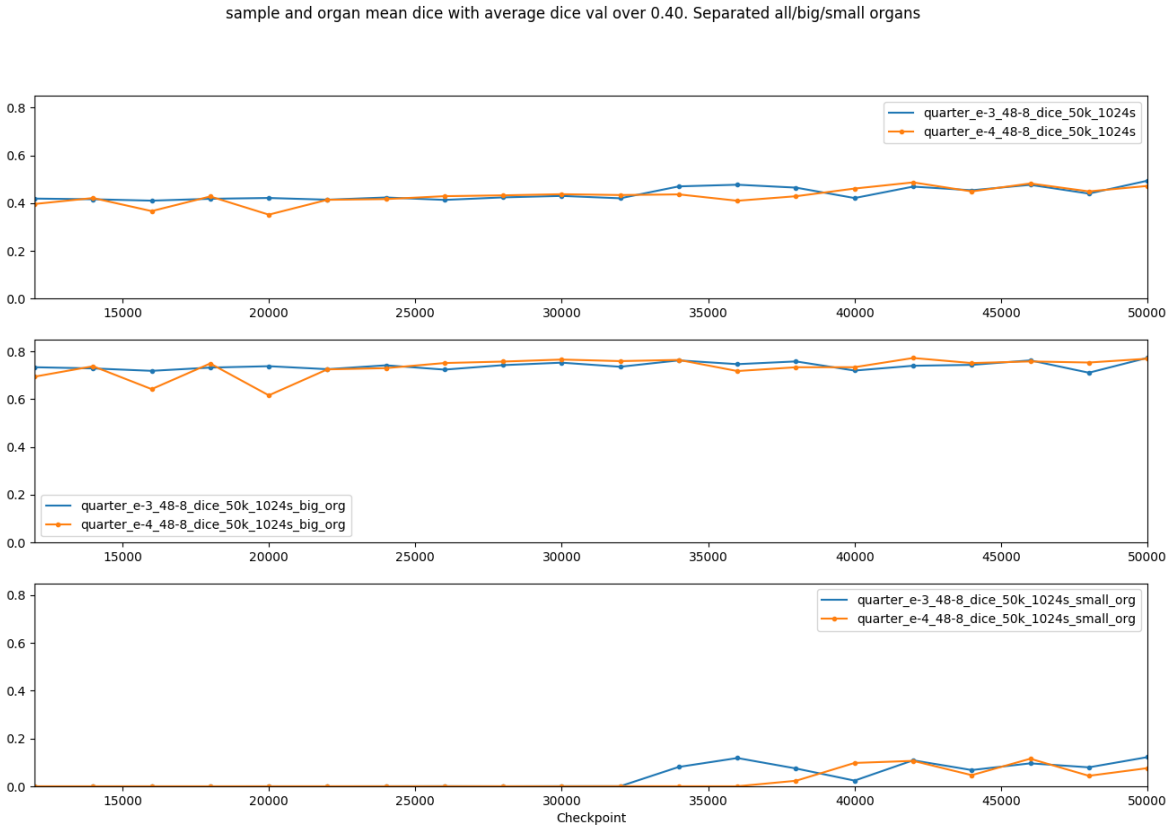

sample and organ mean dice with average dice val over 0.40. Separated all/big/small organs
half_e-3_48-8_dice_50k_1024s
half_e-3_96-1_dice_50k_1024s
half_e-4_48-8_dice_50k_1024s
half_e-4_96-1_dice_50k_1024s
half_e-3_48-8_dice_50k_1024s_big_org
half_e-3_96-1_dice_50k_1024s_big_org
half_e-4_48-8_dice_50k_1024s_big_org
half_e-4_96-1_dice_50k_1024s_big_org
half_e-3_48-8_dice_50k_1024s_small_org
half_e-3_96-1_dice_50k_1024s_small_org
half_e-4_48-8_dice_50k_1024s_small_org
half_e-4_96-1_dice_50k_1024s_small_org
Checkpoint

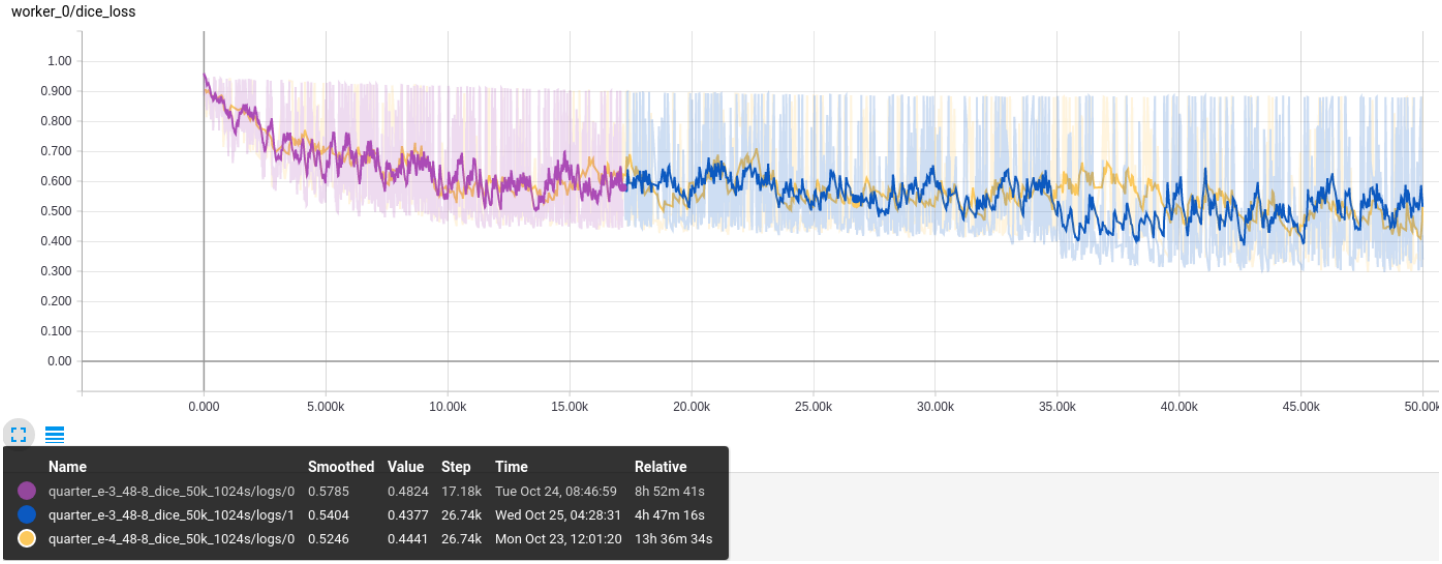

worker_0/dice_loss
Name Smoothed Value Step Time Relative
half_e-3_48-8_dice_50k_1024s/logs/0 0.6212 0.4975 23.06k Sat Oct 21, 01:20:25 11h 31m 59s
half_e-3_96-1_dice_50k_1024s/logs/0 0.7057 0.6855 23.06k Fri Oct 20, 02:47:21 9h 22m 27s



4 models produce good results, all dice, lr 0.001 models are better, especially for small organs, however loss function less noisy for higher learning rates.

Quarter size,



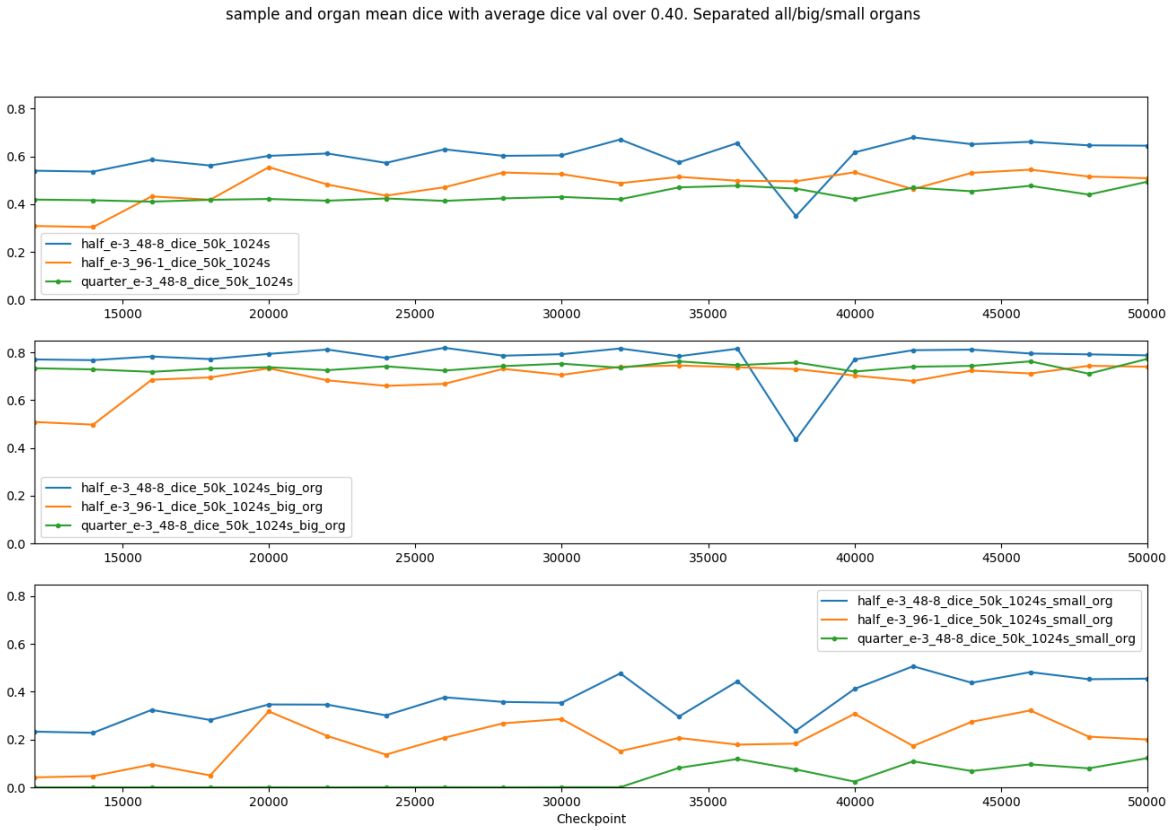


Name	Smoothed	Value	Step	Time	Relative
quarter_e-3_48-8_dice_50k_1024s/logs/0	0.5785	0.4824	17.18k	Tue Oct 24, 08:46:59	8h 52m 41s
quarter_e-3_48-8_dice_50k_1024s/logs/1	0.5404	0.4377	26.74k	Wed Oct 25, 04:28:31	4h 47m 16s
quarter_e-4_48-8_dice_50k_1024s/logs/0	0.5246	0.4441	26.74k	Mon Oct 23, 12:01:20	13h 36m 34s

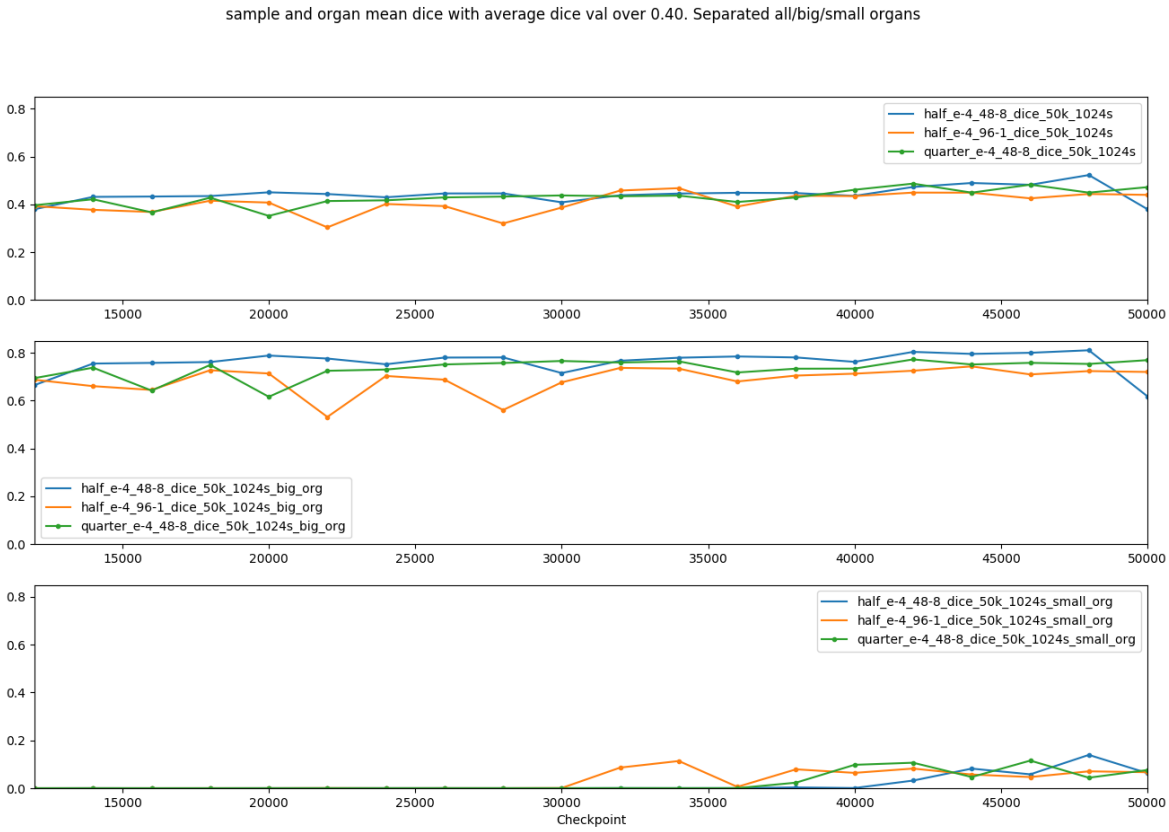
Two models perform well, lr of 0.0001 has less noise and both models seem not to be converged yet. Especially they started to learn the small organs.

Learning rate

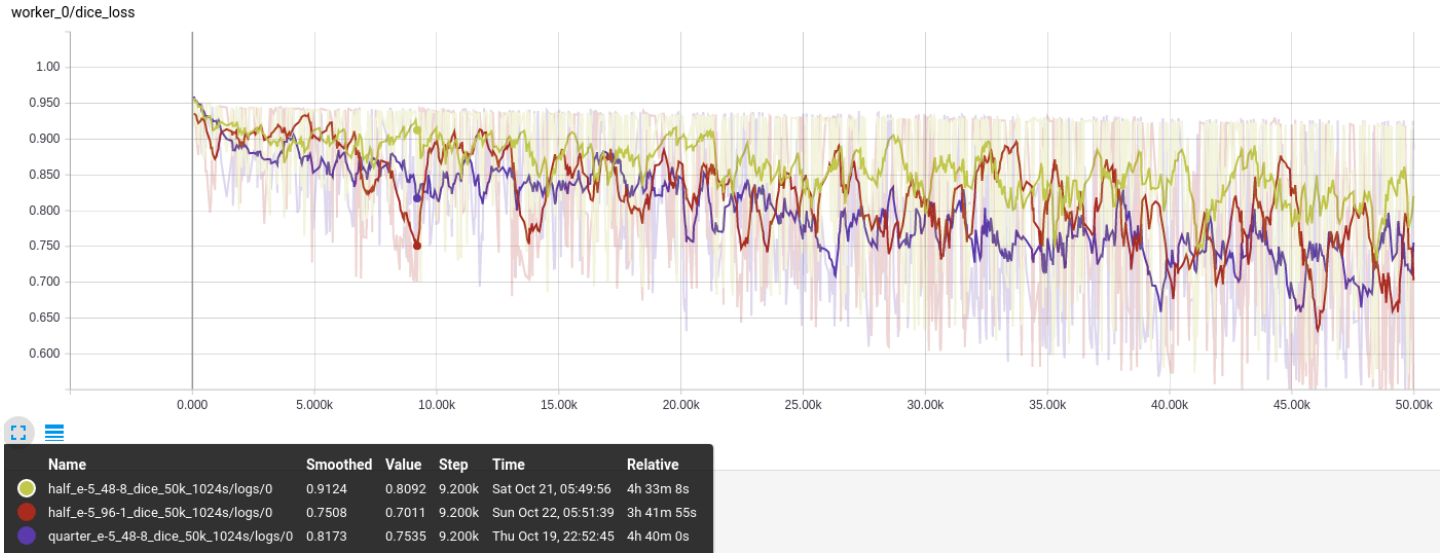
e-3,



Three best models use the dice loss and a lr of e-3, however the loss function is noisy (plotted already previously)



Same three models also work well with a lr of e-4, however performence for small organs are not detected that well, maybe because of the **small lr convergence is slower**. half_e-4_91-1_dice_50k_1024 also looks promising to continue training, but the loss function has way **more noise** than the half_e-4_48-8_dice_50k_1024 model (maybe because of the), thats why the second model is used to continue training.



Three modles achive relative good results. I assume because of the small lr they converge slover. However the lossfunktion is noisy, thus other models are used to continue training.

Overall interpretation:

Gdsc does not work

Lr of 0.001 works best

Few models should be further trained:

half_e-4_48-8_dice_50k_1024 -> loss function not really noisy model started to learn small organs

half_e-3_48-8_dice_50k_1024 -> best model so far

quarter_e-4_48-8_dice_50k_1024 -> big organs already good, loss not noisy started to learn small organs

quarter_e-3_48-8_dice_50k_1024 -> big organs already good, loss not noisier than for the e-4 model but started to learn small organs

Open Questions:

Why does the Gdsc perform so bad?

Why is loss so noisy, best results have loss with more noise then worse results?