

Applying nnU-Net to the KiTS19 Grand Challenge ^{*}

J. B. Graham-Knight¹, A. Djavadifar², Dr. P. Lasserre¹, and Dr. H. Najjaran²

¹ Department of Computer Science, University of British Columbia, Kelowna BC, CA

² School of Engineering, University of British Columbia, Kelowna BC, CA

Abstract. U-Net, conceived in 2015, is making a resurgence in medical semantic segmentation tasks. This comeback is largely thanks to the excellent performance of nnU-Net in recent competitions. nnU-Net generalizes well, as proven by its first-place finish in the Medical Segmentation Decathlon. Notably, nnU-Net focuses on the training process rather than algorithmic improvements, and can often beat more complex algorithms. This paper shows the results of applying nnU-Net to the KiTS19 Kidney Segmentation Grand Challenge. Each of the 5 cross-validation training folds achieves good results, with scores nearing or exceeding 0.9 after approximately 500 epochs per fold.

1 Introduction

Renal Cell Carcinoma is one of the most prevalent diseases affection the kidney. Treatment of RCC consists of three general options: partial nephrectomy (PN), radical nephrectomy (RN), or ablation. Nephrectomy refers to removal of the kidney, where PN refers to removing only the part of the kidney affected by the tumor and RN refers to removal of the entire kidney. Ablation does not entail removing any part of the kidney, but rather the killing of cancerous cells with no incision; this is accomplished, for example, through the application of radio waves, heat, or cold. This is a less invasive procedure then nephrectomy, but is used sparingly because it has a generally lower success rate; as such, it is generally reserved for patients which would be unlikely to survive surgery. Partial nephrectomy is preferred over radical nephrectomy, and is generally accomplished through laproscopic surgery (LPN); sometimes, partial nephrectomy is assisted by robot (RAPN). LPN generally achieves a decrease in operative blood loss and shorter hospital stays [4].

KiTS19 ([5]) is a Grand Challenge in Biomedical Image Analysis [1] where the goal is to correctly identify the kidney and cancerous tissue. Images and segmentations are supplied in the NIFTI format. Data is available from 300 patients. 210 contain both imaging and segmentation labels and are used for training; 90 contain only imaging and are used for evaluation. Patients had undergone a partional or radical nephrectomy for one or more kidney tumors between 2010 and 2018 at the University of Minnesota Medical Center.

^{*} In Conjunction With The Stone Centre at VGH and University of British Columbia, Vancouver Campus

2 Segmentation

Segmentation is the task of removing unwanted features of an image while retaining desired features. A CT scan contains much information which is not always useful to treatment. KiTS19 ([5]) is a 3D semantic segmentation challenge where labels correspond to: 0 - unwanted information; 1 kidney; and, 2, kidney cancer.

Object detection is a simpler task than segmentation, but the two are related. In an object detection task, rather than producing pixel-wise output, a bounding box is defined which contains some region of interest. If an object detection task was applied to Kidney Cancer detection, three bounding boxes might be produced: one containing the right kidney, one containing the left kidney, and one containing the cancerous tissue. Object detection, being a simpler task, is computationally cheaper to perform. It also produces a less useful result, however, as it is not granular.

The 2017 Kaggle Data Science Bowl (DSB) was won by a team employing a 3D object detector. CT scans were combined into normalized 3D models using traditional CV techniques. Blocks of images were then fed through a neural net to identify probability that such block was part of a nodule. The output of this neural network is a sequence of $32 \times 32 \times 32 \times 3$ features, where each feature is a 5-dimensional vector $(\hat{o}, \hat{d}_x, \hat{d}_y, \hat{d}_z, \hat{d}_r)$. \hat{o} represents confidence that this block contains a nodule, and the remaining values represent the centroid and radius of the nodule. The blocks were then processed, without using a neural network, to combine overlapping regions. This provided a tensor of 3-dimensional bounding spheres and probabilities as its output [8]. The same general approach used to win the DSB 2017 was featured recently in a paper on lung cancer screening [2]. They achieved an AUC of 95.9%, shown in Figure 1.

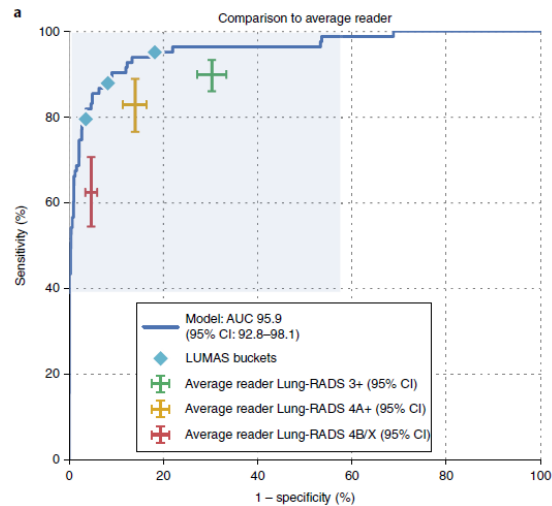


Fig. 1. ROC curve for part 1 results from [2]

Many more competitions have been run with the goal of evaluating medical images. In February 2019, ten of the previous datasets were combined into a single challenge called the Medical Segmentation Decathlon [11, 10]. In contrast to previous competitions, where task-specific optimisations could provide an edge, the decathlon seeks to provide a more generalisable algorithm. Evaluation metrics include the Dice Score (Equation 1) and Normalised Surface Distance, which measures the difference between the calculated and ground truth surfaces.

$$DSC = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (1)$$

TP = true positives
 FP = false positives
 FN = false negatives

The winner of the decathlon was nnU-Net [7]. In contrast with other approaches, which sought to improve neural network architecture, Isensee et. al sought to use a well known and relatively simple architecture (U-Net [9], specifically 3D U-Net [3]), but to train it very well.

3 Method

Training is done using 5-fold cross validation; to accomplish this, 1/5 of the training set is withheld for evaluation. If the training loss and validation loss diverge, it is an indication that the model is overfitting. Training is also stopped if there is no improvement in model performance for some time. This is repeated five times, known as folds, with the validation data set changed in each fold. This allows the algorithm to converge to five different critical points, effectively producing five different models. These models are then combined into an ensemble, which is the final result.

It takes many epochs, or iterations over the entire data set, to produce good results when training a fold. In this case, it took approximately 500 epochs for each fold; one fold was near 700 epochs. The code for nnU-Net [6] was used on the training data from the 2019 Kidney Tumor Segmentation Challenge [5]. 3-dimensional image segmentation is a computationally heavy task; training of nnU-Net on the KiTS19 data took over two weeks with a single Titan GPU. There have been some attempts to optimize this time. One such optimization is the employment of an object detector before segmentation, which reduces the area of interest. Another optimization is to use U-Net in cascade, which is possible within the nnU-Net framework; the first phase does a coarse segmentation, and the second phase does a finer segmentation. Depending on the features present in the images, cascaded U-Net may not be wise as resolution is lost in the first stage. In this application, only the full-resolution 3D U-Net algorithm was tested. The 2D and 3D-cascaded algorithms were not used.

4 Results

Results are available in Figures 2 through 6. The blue line indicates training loss and the red line indicates validation loss, which is calculated from the held-back port of the data set for the fold. For loss, lower is better, and the scale is on the left axis. The green line represents the evaluation metric, and in this case it is training to reasonably good results with all folds being above 0.8 and some fold exceeding 0.9; the scale is on the right axis. The model is converging well, so reasonably good results are expected in predictions. Each graph represents a single fold, and the folds are combined into a final ensemble for final predictions.

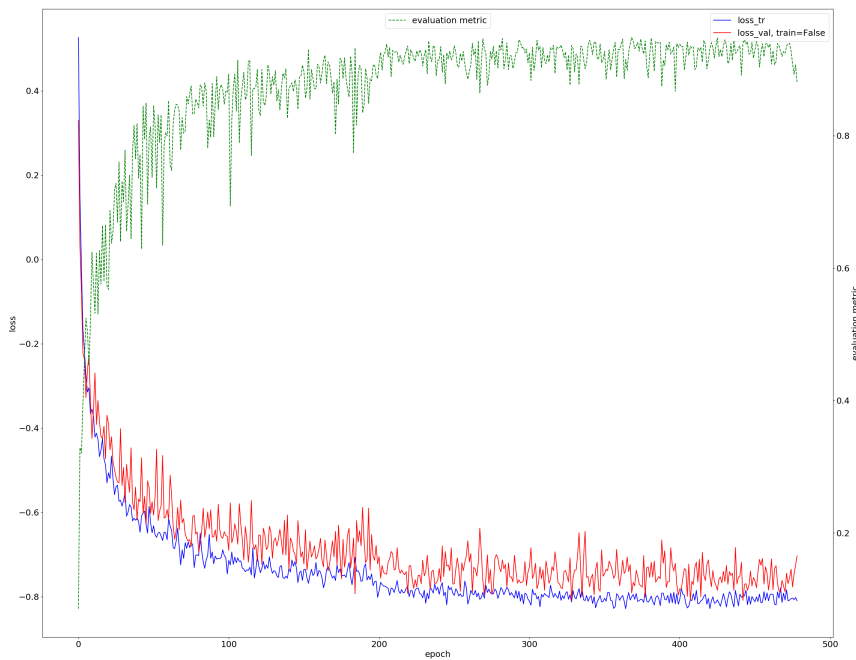


Fig. 2. Results of Fold 1

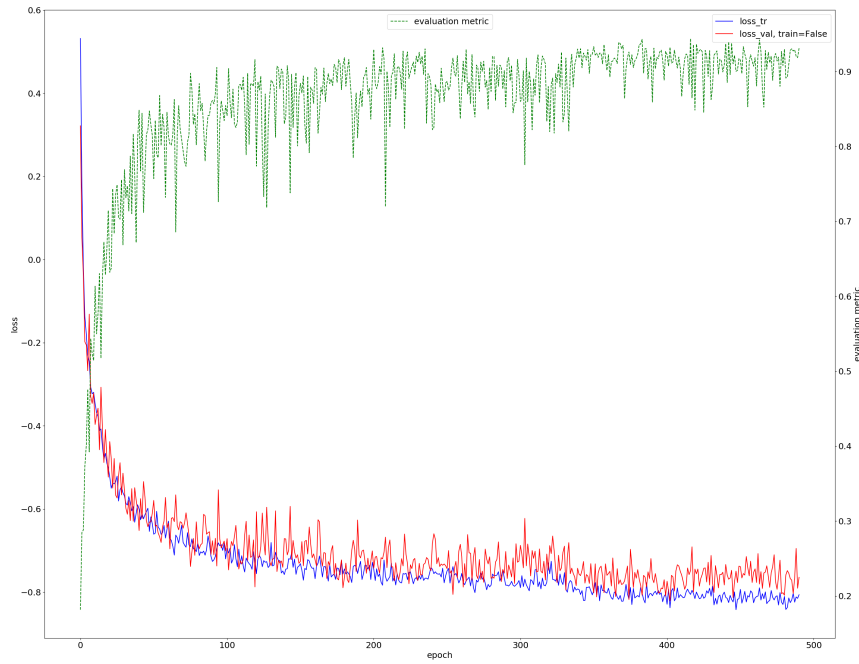


Fig. 3. Results of Fold 2

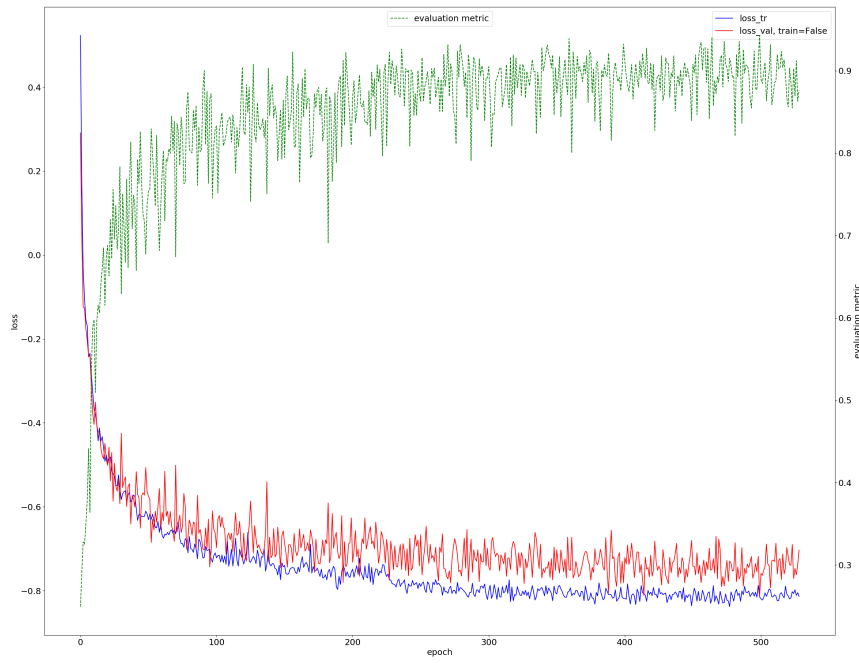


Fig. 4. Results of Fold 3

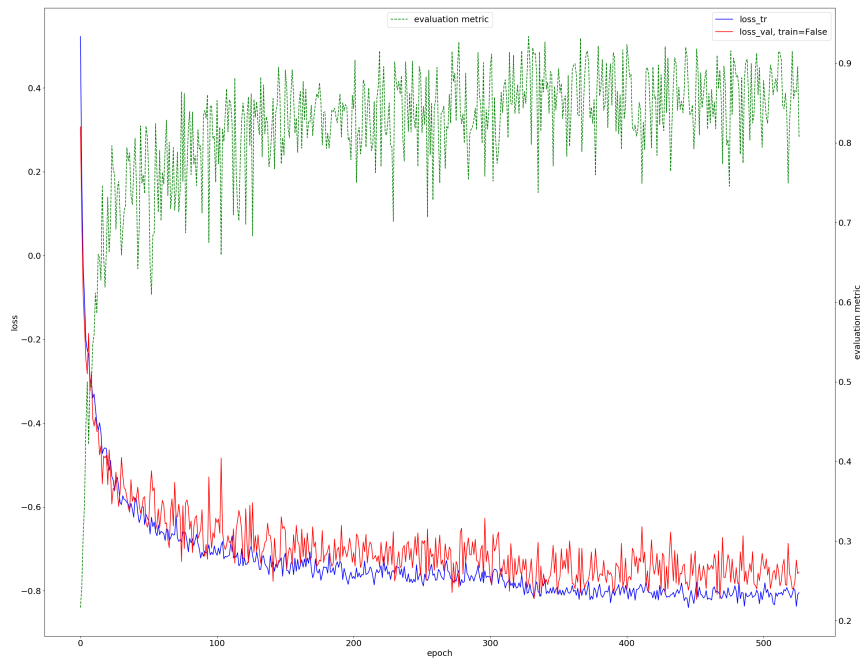


Fig. 5. Results of Fold 4

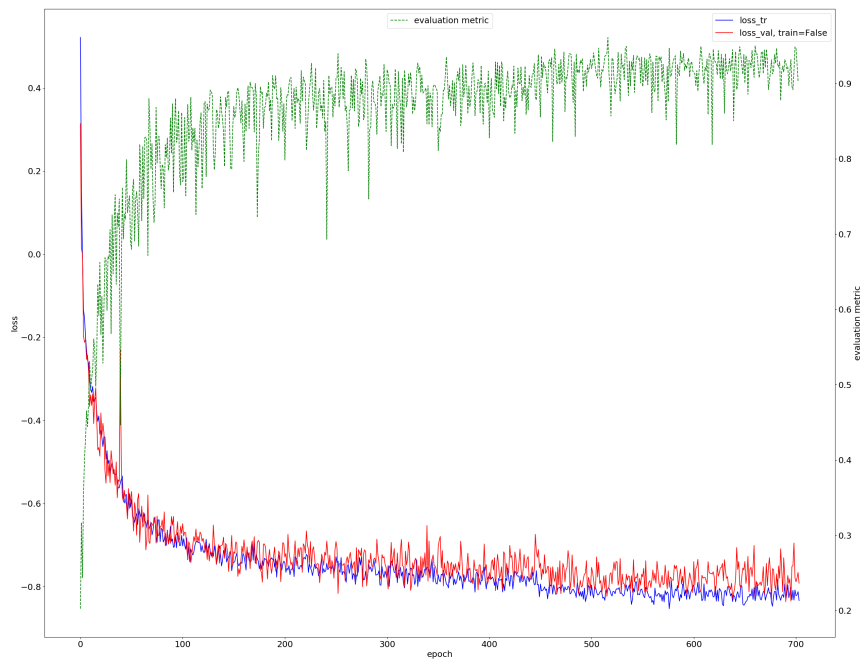


Fig. 6. Results of Fold 5

References

1. Grand challenges in biomedical image analysis. <https://grand-challenge.org/> (2019), [Online; accessed 29-July-2019]
2. Ardila, D., Kiraly, A.P., Bharadwaj, S., Choi, B., Reicher, J.J., Peng, L., Tse, D., Etemadi, M., Ye, W., Corrado, G., Naidich, D.P., Shetty, S.: Author correction: End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature Medicine* (2019). <https://doi.org/10.1038/s41591-019-0536-x>, <https://doi.org/10.1038/s41591-019-0536-x>
3. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: *International conference on medical image computing and computer-assisted intervention*. pp. 424–432. Springer (2016)
4. Detmer, F.J., Hettig, J., Schindele, D., Schostak, M., Hansen, C.: Virtual and augmented reality systems for renal interventions: A systematic review. *IEEE Reviews in Biomedical Engineering* **10**, 78–94 (2017). <https://doi.org/10.1109/RBME.2017.2749527>
5. Heller, N., Sathianathan, N., Kalapara, A., Walczak, E., Moore, K., Kaluzniak, H., Rosenberg, J., Blake, P., Rengel, Z., Oestreich, M., et al.: The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes. *arXiv preprint arXiv:1904.00445* (2019)
6. Isensee, F.: nn-unet source code. <https://github.com/MIC-DKFZ/nnUNet> (2019), [Online; accessed 4-June-2019]
7. Isensee, F., Petersen, J., Kohl, S.A., Jäger, P.F., Maier-Hein, K.H.: nnu-net: Breaking the spell on successful medical image segmentation. *arXiv preprint arXiv:1904.08128* (2019)
8. Liao, F., Liang, M., Li, Z., Hu, X., Song, S.: Evaluate the malignancy of pulmonary nodules using the 3-d deep leaky noisy-or network. *IEEE transactions on neural networks and learning systems* (2019)
9. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*. pp. 234–241. Springer (2015)
10. Simpson, A.L., Antonelli, M., Bakas, S., Bilello, M., Farahani, K., van Ginneken, B., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., et al.: A large annotated medical image dataset for the development and evaluation of segmentation algorithms. *arXiv preprint arXiv:1902.09063* (2019)
11. Simpson, A.L., et al.: Medical segmentation decathlon. <http://medicaldecathlon.com/> (2019), [Online; accessed 12-June-2019]