COMP9444 Project Summary Report

Kidney (And Kidney Tumour) Segmentation

November 2024

Team: Waiting for GPU
Yu Xie (z5529775)
Aayush Bajaj (z5362216)
Samuel Hodgson (z5416863)
Song Lin (z5362555)
Liam Biddle (z5311885)

1. Introduction

The aim of this report is to evaluate different methods for semantic medical image segmentation. Specifically determining the most accurate model for identifying kidney segmentation from the KITS19 dataset. This work is then extended in the final nnUnet model to identify tumours within the kidney. There are clear applications of this work for medical practitioners who seek to quickly identify the location of any kidney tumours so that they can be removed through partial nephrectomy.
This report only seeks to evaluate pre-existing models and methods but additionally contributes to the conversation by investigating potential extensions of said models. The research was limited by the time and computational resources available to a team of undergraduate students. Thus the more computationally intensive models had to be limited or modified such that they could be feasibly explored.

2. Related Work

In 2015, U-Net was presented. It relied on the strong use of data augmentation to use the available annotated samples more efficiently. The architecture consisted of a contracting path to capture context and a symmetric expanding path that enabled precise localization. Such a network could be trained end-to-end from very few images and outperforms the prior best method (a sliding-window convolutional network) *(Ronneberger et al., 2015)*.
After the great success of U-Net, a lot of medical image segmentation networks were improved based on the U-Net network, and many networks of this type were derived. One of the most famous networks was V-net. It was a volumetric, FCN-based approach to 3D image segmentation which was proposed in 2016 *(Milletari et al., 2016)*. It was trained end-to-end on MRI volumes depicting the prostate, and learned to predict segmentation for the whole volume at once. A new objective function was also introduced in V-Net which was optimised during training based on the Dice coefficient to achieve a state that can handle strong imbalances between foreground and background voxels. When the time came to 2022, to improve the visualisation and quantification of different image regions, the implementation of a 2D version of V-net architecture was proposed. Extensive simulation of the 2D V-net model shows a significant improvement in Dice Coefficient, sensitivity, specificity and accuracy compared with U-Net *(Dangoury et al., 2022)*.
Also proposed in 2022, SAM-Net was a self-supervised learning-based VO framework with Semantic probabilistic and Attention Mechanism. SAM-Net can jointly learn the single view depth, the ego motion of camera and object detection *(Kirillov et al., 2023)*. It provided zero-shot segmentation outputs with or without interactive user inputs, demonstrating notable performance on various objects and image domains without prior training *(Ali et al., 2024).*
Then back in 2019, no-new-Net (nnU-Net) was proposed. It includes a formalism for automatic adaptation to new datasets *(Isensee et al., 2020)*. Based on an automated analysis of the dataset, nnU-Net automatically designs and executes a network training pipeline. Being wrapped around the standard U-Net architecture, the hypothesis was that a systematic and careful choice of all hyperparameters will still yield competitive performance. Indeed, without any manual re-tuning, the method achieves state-of-the-art performance on several well-known medical segmentation benchmarks.

3. Methods

The models we have chosen include Unet, V-net (Volumetric network), SAMnet (Segment Anything Model) by meta and nnUnet.

U-Net

Named for its U-shaped architecture, the original Unet was designed to work in 2D, a fully convolutional neural network (FCNN) made for biomedical segmentation. However, given that most biomedical images are taken in 3D, many adaptations of it have since been made to use 3D images instead *(Içek et al., 2016) (Yao et al., 2022)*. Such examples include DAR-Unet, 3D-Unet, or a hybrid between 3D and 2D implementation *(Ushinksy et al., 2020)*. By far, the 3D variations of Unet are the state of the art in biomedical image segmentation.

V-Net

V-Net is a convolutional neural network which is designed to specially deal with the segmentation of volumetric medical imagery. It operates on the 3D input data from sources such as MRI or CT scans. Using a combination of an encoder which captures the spatial features of the scans, and the decoder which reconstructs the segmentation. The architecture also allows for the usage of connections which skip between the encoder and decoder layers, which allows the presentation of resolution and accuracy. There have been several improvements which can be made on VNet since its conception in 2016, which incorporate modern improvements. For this assignment, several modifications have been made to accommodate the KiTS19 kidney tumour segmentation dataset as well several dependency changes.

SAM-NET

Developed by Meta, the current SAM 2 network allows segmentation of image or video, for our purposes, we have elected to explore its predecessor, which is specialised for images. The difference between SAM and other networks is that it uses panoptic segmentation. The main advantage of SAM is that it is designed to work on almost any device and has been pre trained by Meta on a huge dataset for segmentation. This provides accessibility for lower computational hardware.

SAMNet is however limited as it is unable to naturally classify which mask belongs to the kidney segmentation, not having the necessary domain knowledge to do so. For this reason SAMNet was connected to a secondary fully connected neural network which uses characteristics of the input masks to predict their relative DICE Scores with the target kidney segmentation.

Our findings suggest that combining the masks with the two highest DICE Scores for each slice produces a more accurate result. This is likely a consequence of slices containing two distinct and separated kidneys which SAMNet will naturally treat as two separate objects.

The accuracy of this particular model architecture was greatly limited by false positives. The model is trained to select the two generated masks it believes to most likely represent kidneys, however this system fails when there are no kidneys in the slice. This was the case for many slices in the dataset and thus the overall performance was significantly impacted.

nnU-Net

nnU-Net or known as “no-new-Net” is a framework designed to dynamically adjust itself to any new segmentation task that requires little to no manual intervention *(Isensee & Maier-Hein, 2019)*. The architecture extends that of the original UNet via the integration of preprocessing, flexible architecture design, and a post-processing pipeline. It can adjust by using 2D, 3D and

cascaded segmentation to automatically choose the best approach for the specific dataset. Due to the flexibility and generalisation continually achieves high performance for medical segmentation results.

4. Experimental Setup

Dataset: https://github.com/neheller/kits19

The dataset contains 300 selected CT imaging scans from patients who underwent nephrectomy (kidney surgery) for one or more kidney tumours. The imaging is provided in NIFTI format, which is used primarily for storing and sharing medical imaging data.
Manual segmentation labels were produced to be used as target masks. Background features were labelled 0, with kidneys labelled 1 and the tumours labelled 2.
There are only 210 segmentations in this dataset which we split into training, validation and test sets. Starter code is provided to download 300 images.

Data Analysis

Most 3D images along with segmentations are provided with shape n x 512 x 512 (Unit: voxel), except case_00160, which is n x 512 x 768 (Unit: voxel).
Different images have different voxel spacing, which means they have different physical resolution. But for each image, voxel spacings of the latter two dimensions are the same.
Hounsfield values (HU) of images are between -1000 and 1000. Background occupies about 60% of an image, which has HU values close to -1000. The human body and organs share the left 40% with HU values close to 0.

Data preprocessing

1. Resample voxel spacing to (3.22, 1.62, 1.62) to get Unified resolution.
2.Limit the range of HU to (-79, 304) to enhance image and reduce noise.
3. Use z-score to normalise images.
4. Cut images and segmentations into 2D slices (Aixel, Coronal, Sagittal).
5.Resize images and segmentations to 512 x 512 (Unit: pixel) to fit models.

Evaluation metrics

We use the DICE coefficient to evaluate models. It evaluates the similarity between two samples, defined as twice the area of intersection divided by the sum of the sizes of the two sets. This was also used as the metric for Early Stopping to find the appropriate epoch sizes when training.

5. Results

Model	Kidney Segmentation DICE Score
U-Net	0.6114
V-Net	0.6145
SAM-Net	0.6278
nnU-Net	0.79151 (mean + tumour)
*State of the Art (3D U-Net)	0.974 (only kidney)

Among the evaluated models, nnU-Net achieved the highest Dice score, significantly outperforming other models like UNet, VNet, and SAMNet. Although nnU-Net showed superior performance among 2D networks, it still falls short when compared to the state-of-the-art 3D U-Net (we spent 55 hours training, but did not get the time to train a 3D network).
Compared with other models, nnU-Net's flexible design allows for automated network architecture tuning and hyperparameter optimization, resulting in substantial performance improvements.
With a Dice score close to 0.8, nnUNet demonstrates promise for real-world clinical applications, though a more complicated model (3D), was proved in the Notebook to be the next step for a more accurate Dice score. Because in medical domains where high precision is critical, achieving a Dice coefficient closer to 0.9 or above is often necessary for reliable deployment.

6. Conclusions

The standard implementation of UNet struggles with segmenting complex tasks due to the limited generalising nature of the model's architecture. This implementation segments only the kidney without consideration for the tumour (not strictly part of the spec), which likely would have drastically decreased its performance as 2D networks are known to struggle with detecting small objects in 3D images converted to slices, something already evident in its detection of the kidney on negative samples.
The larger number of parameters and the usage of 3D convolutions in V-Net needs to have high computational requirements, which leads to an increase in training time, limiting the number of epochs needed. On a smaller dataset, and the larger number of parameters of VNet can lead to overfitting.
While the benefits of Meta's pretrained model allow for simple and accurate image segmentation it is unable to naturally identify which of the produced masks belong to the kidney without extensive domain knowledge. And this is the limitation of SAM-Net. Our solution of extracting the mask information and feeding it into a secondary neural network proved to be a relatively simple extension of the model which solved our problem. However, the model architecture continues to be limited by false positives. Since the secondary network has been tasked to identify and predict the combination of the two most kidney-like masks provided to it, it performs poorly when the target segmentation does not contain any kidney. Unfortunately this was the case for many of the image slices and thus the overall performance of the model was limited. Potentially a modification of SAMNet2 which is designed for video segmentation, could be used for a solution to this problem. The 3D volumes could be treated as videos with flat slices being frames and the third dimension being treated as time. Treating the problem in this 3D perspective would resolve the issue of having slices without any kidneys to identify, however it requires extensive computational power and time which were unfortunately unavailable to us for this particular project.
Overall, we believe choosing nnU-Net was a good long term decision, in that many problems will now have the same pipeline, but in the short-term there remains much about the internal workings of this framework that eludes us. There is certainly improvement to be had here with respect to training more sophisticated models, and then on top of that we did not get the chance to leverage data augmentation since we ran out of space on Katana.

References

Ali, L., Alnajjar, F., Swavaf, M., Elharrouss, O., Abd-alrazaq, A., & Damseh, R. (2024, September 17). *Evaluating segment anything model (SAM) on MRI scans of brain tumours*. Scientific Reports. https://doi.org/10.1038/s41598-024-72342-x

Dangoury, S., Sadik, M., Alali, A., & Fail, A. (2022, December 12). *V-net Performances for 2D Ultrasound Image Segmentation*. IEEE Xplore. https://ieeexplore.ieee.org/document/9781973

Içek, O., Abdulkadir, A., Lienkamp, S., Brox, T., & Ronneberger, O. (2016, June 21). *3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation*. arXiv. https://arxiv.org/pdf/1606.06650

Isensee, F., Jaeger, P., Kohl, S., Petersen, J., & Maier-Hein, K. (2020, Dec 7). *nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation*. Nature Methods. https://www.nature.com/articles/s41592-020-01008-z

Isensee, F., & Maier-Hein, K. (2019, October 4). *An attempt at beating the 3D U-Net*. arXiv. https://arxiv.org/pdf/1908.02182

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A., Lo, W.-Y., Dollár, P., & Girshick, R. (2023, April 5). *Segment Anything*. arXiv. https://arxiv.org/pdf/1908.02182

Milletari, F., Navab, N., & Ahmadi, S.-A. (2016, June 15). *V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation*. arXiv. https://arxiv.org/pdf/1606.04797

Ronneberger, O., Fischer, P., & Bronx, T. (2015, May 18). *U-Net: Convolutional Networks for Biomedical Image Segmentation*. arXiv. https://arxiv.org/pdf/1505.04597

Ushinksy, A., Bardis, M., Glavis-Bloom, J., Uchio, E., Chantaduly, C., Nguyentat, M., Chow, D., Chang, P.D., & Houshyar, R. (2020, November 10). *A 3D-2D Hybrid U-Net Convolutional*

Neural Network Approach to Prostate Organ Segmentation of Multiparametric MRI. American Journal of Roentgenology. https://doi.org/10.2214/ajr.19.22168

Yao, K., Su, Z., Huang, K., Yang, X., Sun, J., Hussain, A., & Coenen, F. (2022, March 24). *A Novel 3D Unsupervised Domain Adaptation Framework for Cross-Modality Medical Image Segmentation*. IEEE Journal of Biomedical and Health Informatics. https://ieeexplore.ieee.org/document/9741336