Comparing CNN to K-Means for Medical Image Segmentation

Ian Esselburn, Maxim Mamotlivi

Computer Science Dept., Kent State University
1300 University Esplanade, Kent, OH 44240, United States
`iesselbu.edu`
`mmamotli@kent.edu`

ABSTRACT— **The subject of this project regards the comparison of two different methods in finding and outlining spleens from CT scan images; U-Net Convolutional Neural Network (CNN) and a restricted K-means clustering algorithm. For this, a dataset of NIfTI formatted CT scan images was converted into PNG format along with corresponding spleen-masks. A U-Net CNN was then trained on this data to produce a Keras model, and the parameters of K-means were minimally adjusted to better fit the data format. The results of each model/algorithm's operations on a separate test set of images were then compared via a Dice similarity coefficient. We find that the U-Net trained Keras Model produced significantly more accurate spleen image masks, especially in instances where images weren't very clear. K-means operations required less time in aggregate, however the resulting image masks were substantially less accurate. This project assists in expanding the understanding of the tradeoffs involved between these approaches for medical image analysis.**

I. INTRODUCTION

The practice of medical imaging holds great importance in the course of diagnosing and treating patients. The process of locating and outlining organs in these images, called segmentation, may be performed manually by doctors, however this can become a time intensive task for increasingly large sets of image data. Computer programs utilizing machine learning can greatly help expedite such processes. Thus, it is necessary to measure the effectiveness of various image segmentation approaches to identify which are most suitable towards the specific conditions of each task.

Because this project pursues a direct-comparative approach, the spleen was selected as the target organ for segmentation. The primary motivating factors in this selection were: the relatively simple shape of the spleen, which facilitates segmentation while still posing challenges in distinguishing it from surrounding organs, and the availability of comprehensive spleen imaging data. Our first approach utilized a U-Net specialized convolutional neural network trained for precise medical image segmentation. The second approach involved restricted K-means clustering: a more or less traditional unsupervised machine learning algorithm, calibrated to account for certain regularities in the CT scan images. By testing both methods on the same dataset and test images we aimed to see which method would be superior for the given task of spleen segmentation.

The comparison between U-Net CNN and restricted K-means clustering aimed to assess their trade-offs both in accuracy, computational efficiency, and adaptability throughout. Generally, the U-Net deep learning approach excels in capturing complex image patterns, offering high precision but demanding substantial computational resources and labeled data. Conversely, K-means, as an unsupervised method, does not require training or label-data but struggles with variability in images and noise. Practical constraints like processing speed and data needs were considered along with accuracy throughout the course of evaluation.

II. RELATED WORK

Automated organ segmentation in medical imaging has been a key research area for years,

aiming to improve diagnostic precision and efficiency. The U-Net CNN architecture, introduced in 2015 by Ronneberger et al. [2], demonstrated exceptional performance in segmenting medical images, even with limited training data, due to its ability to capture detailed spatial features through a contracting and symmetric-expanding path structure [2]. Thus much research centers on the expansion, modification and comparison of U-Net for specific use-cases as U-Net is well known for its general effectiveness [8, 9].

Meanwhile, K-means clustering is an established unsupervised method, grouping similar pixels without requiring labeled data, providing a fast method for segmentation. Because K-means functions in images to isolate pixels in areas of interest from the background, it depends on a base level of image regularity [6]. If image quality varies within a dataset along multiple degrees (brightness, shifts in the area of interest, noise, etc.) it becomes necessary to either restrict parts of the image from consideration or subtractively filter data points throughout the course of interaction (accounting for the specific form of variance) [6, 7]

Recent studies reflect the continuing progress in deep learning-based segmentation. The Medical Segmentation Decathlon analyzed the versatility of architectures like U-Net among other approaches, across various medical imaging tasks, providing also a good repository of open source data [1]. A 2021 study by Meddeb et al. [3] compared customized neural network algorithms for spleen segmentation in CT scans, finding high accuracy for normal spleens but challenges with abnormal ones, suggesting case-specific limitations. Other works by Guo et al. [4] and Kayalibay et al. [5] further confirmed the effectiveness of convolutional neural networks in organ segmentation for multimodal images. Conversely, K-means clustering, while computationally efficient, encounters the aforementioned issues, resulting in both under-segmentation and over-segmentation; as explored by Dhanachandra et al. [6]. However, K-means and U-Net are rarely compared directly because of their paradigmatic differences.

III. METHODOLOGY

The project methodology can be roughly divided into three distinct phases: dataset preparation, the implementation of each approach (CNN & K-means), and testing followed by evaluation.

To this end, the Python programming language and related libraries were used throughout the entire project – in each phase. The libraries utilized and their respective purposes are as follows:

```
os
    -   Operating system interactions
        (file/folder handling).
numpy
    -   Numerical computing and array
        manipulation.
tensorflow
    -   Deep learning framework for model
        training and inference.
cv2 (OpenCV)
    -   Computer vision tasks (image/video
        processing, feature detection, basic
        K-means).
matplotlib
    -   Data visualization and plotting.
nibabel
    -   Neuroimaging file format handling
        (conversion, reading/writing).
imageio
    -   Reading and writing image files
        (format conversion).

*Note that tensorflow is only compatible
within Python versions 3.9 - 3.12
```

A. Dataset Preparation

The original spleen dataset was obtained from the repository made available by the Medical Segmentation Decathlon Project website [1]. Because the data inside this archive had apparently been transferred from a Mac computer, the immediate task required the removal of useless metadata files. Having isolated only the image archives, subdivided into training images, training image masks, and a test set, conversion from NIfTI image format to PNG was initiated via Python program. The resulting dataset of images was then meticulously and manually refined to include only those CT scan slices which displayed the spleen, in original image, image mask, and test image. Only by the completion of preprocessing did it become possible to implement the approaches in question.
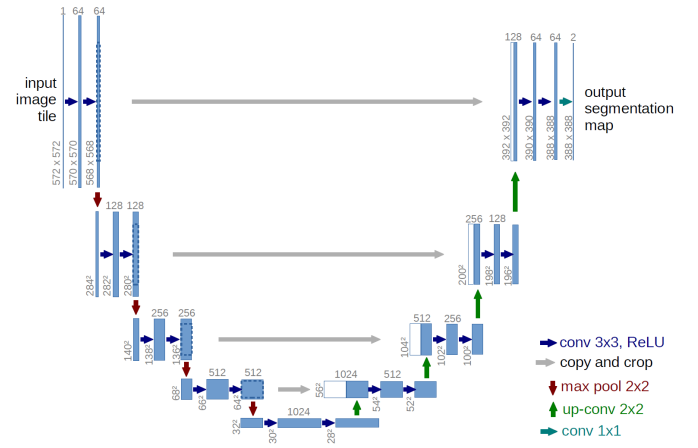
B. U-Net Approach

U-Net derives its name from the U-shaped architecture governing the approach. This U-Net implementation is structured with an encoder-decoder design that operates to segment the spleen accurately from the background.

The first component, the encoder path, compresses the image while extracting features. This implementation uses three encoder blocks. Each block contains a convolutional layer that extracts features, followed by a max pooling layer which reduces the image size by half each time. The feature channels are then increased, starting at 64, then 128, then 256. At the base of the U shape lies the 512-filter convolutional layer, at which the image is most compressed.

The decoder path, in an order mirroring the preceding stages, expands the image back to full size. It is composed of three blocks that each increase the resolution. Each decoder block first up-samples the image to double its size, then uses skip connections to reintegrate details from the corresponding encoder level. Convolutions are applied to refine the features.

The model was trained using binary cross-entropy loss, since the membership of a pixel in the spleen-class is a binary distinction; it is either a part of the spleen shape, or it is not. The Adam optimizer, which adapts the learning rate automatically, was utilized with an initial learning rate of 0.001. Training spanned 20 epochs, meaning the network processed the entire dataset 20 times. Attempts at more epochs led to overfitting. Each batch contained 8 images, balancing GPU constraints. The network learned the likelihood of each pixel being part of the spleen, within a training time of approximately 2 hours.

U-Net diagram from Ronneberger et al. [2] for reference, the general structure of which is quite similar to the U-Net approach employed here..

C. K-Means Approach

As K-means is an unsupervised clustering algorithm, the extent of modification depended on restricting the area of consideration to a region of interest (ROI). Given the manner in which the CT scan slices were taken, ROI was defined as the center of the upper half of the image (middle 1/4 of width, upper 1/4 of height), as this region would always contain the spleen at a level brighter than the background.
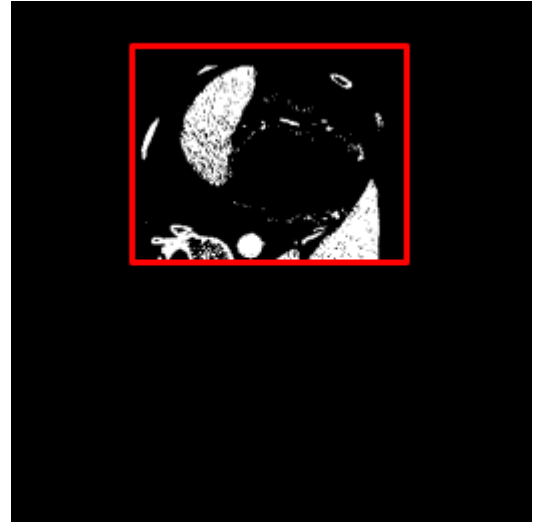

Fig. displaying mask produced by restricted K-means with limited ROI highlighted in red.

The number of clusters, K, was set to 3 after visual inspection of early segmentation results determined that this better separated the spleen, background, and intermediate features. The choice

of K=3 was also influenced by the binary nature of the task (spleen or not), but allowed for an additional cluster to capture some noise or middle regions, because the alternative scenario (K<3) proved to mask in excess. The iterations were tuned to a maximum of 25 in the termination criteria, and the number of attempts for K-means initialization was set to 25, both adjusted after, again, visually inspecting the mask quality to balance computational efficiency and convergence.

Epsilon value as termination criteria remained 1.0 (common default value). A smaller epsilon could have increased computational cost without significantly improving the clustering outcome. Thus, the changes made reasonably exhausted the extent of optimization given the grayscale image's intensity distribution.

RESULTS | EVALUATION

The U-Net and K-Means methods were quantitatively assessed using the Dice coefficient score which quantitatively measures the overlap between predicted segmentations and ground-truth masks. The Dice score ranges from 0 to 1, with 1 indicating perfect overlap. In testing, the U-Net model achieved a Dice score of 0.95 on a test image (slice_075.png), demonstrating high accuracy in delineating spleen boundaries. By contrast, the K-Means algorithm predictably scored a significantly lower Dice score of 0.38, indicating lower segmentation accuracy and inconsistent performance across varied CT slices.

The computational requirements of both methods were also analyzed by accounting. The U-Net Keras model required approximately 2 hours for training due to its deep learning architecture and iterative optimization. Conversely, K-Means generated segmentations practically instantaneously without requiring a training phase, directly clustering pixel intensities [6]. However, these times are also subject to adjustment by changing the number of iterations or modifying the termination threshold.

Visual comparison revealed several differences between the two methods. U-Net produced overall smoother and more accurate boundaries abounding the spleen mask, while K-means produced jagged edges with isolated pixels. K-means frequently included other organs with similar grayscale brightness intensity to the spleen, creating false positives. On slices where the spleen was less visible due to poor contrast, U-Net still performed reasonably well whereas K-means results were quite poor. U-Net output image masks of relatively consistent quality across different images. K-means performance varied depending on the image quality and contrast. Visual differences clearly demonstrated that the U-Net Keras model produced more accurate segmentations across a wide variety of the test images.
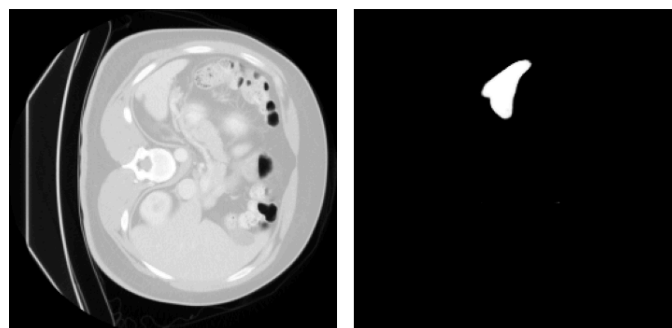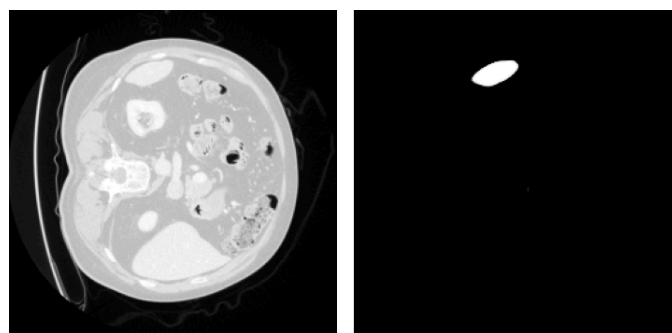

Fig. showing testing image on left and CNN prediction on right

Fig. showing ground truth on right and K-means generated mask on left

IV. CONCLUSIONS

Throughout the project knowledge was acquired regarding various techniques for segmenting medical images. The implementation and evaluation of a Convolutional Neural Network (CNN), specifically the U-Net Keras model architecture, compared to a restricted K-means clustering algorithm for spleen segmentation revealed the distinct advantages of U-Net over K-means.

The accuracy comparison gave predictable insights. U-Net achieved a superior segmentation accuracy with a Dice score of 0.95, in contrast to K-means, which recorded a Dice score of 0.38. This metric is critical in medical applications, where precise localization of the spleen is essential.

U-Net notably demonstrated exceptional performance in handling complex cases with low contrast and ambiguous boundaries, where K-means underperformed, particularly in images where the spleen's intensity resembled that of surrounding organs. U-Net consistently delivered reliable results across diverse imaging conditions [1,2].

For real-world applications where accuracy is critical, our findings establish U-Net as the superior method for medical image segmentation. But more importantly, reveals that the method of restrictive optimization and parameter tuning for K-means is only effective in circumstances where images have a high degree of regularity within a dataset. K-means may still be suitable for rapid preliminary analysis but should not be relied upon for final clinical decisions.

If more time was available to extend the project, a few different directions could be explored. Implementing a 3D U-Net could incorporate inter-slice information from CT scans to further improve accuracy. Enhancing K-means clustering techniques through more in-depth geometric optimizations could also yield better image mask quality. Testing both methods on cases with enlarged spleens or abnormal anatomy would offer valuable insights into their robustness beyond typical healthy spleen cases [3]. This project provided practical experience in applying CNN and K-means for medical image segmentation, allowing a comparison of deep learning and traditional methods.

REFERENCES

[1] Antonelli, M., Reinke, A., Bakas, S. et al. The Medical Segmentation Decathlon. Nat Commun 13, 4128 (2022). https://doi.org/10.1038/s41467-022-30695-9.

[2] Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:1505.04597 (2015). https://arxiv.org/abs/1505.04597.

[3] Meddeb A, Kossen T, Bressem KK, Hamm B, Nagel SN. Evaluation of a Deep Learning Algorithm for Automated Spleen Segmentation in Patients with Conditions Directly or Indirectly Affecting the Spleen. Tomography. 2021 Dec 13;7(4):950-960. doi: 10.3390/tomography7040078.

[4] Z. Guo, X. Li, H. Huang, N. Guo and Q. Li. Deep Learning-Based Image Segmentation on Multimodal Medical Imaging. IEEE Transactions on Radiation and Plasma Medical Sciences, vol. 3, no. 2, pp. 162-169, March 2019, doi: 10.1109/TRPMS.2018.2890359.

[5] Kayalibay, B., Jensen, G. & van der Smagt, P. CNN-based Segmentation of Medical Imaging Data. arXiv:1701.03056 (2017). https://arxiv.org/abs/1701.03056.

[6] Dhanachandra, N., Manglem, K. & Chanu, Y. J. Image Segmentation Using K-means Clustering Algorithm and Subtractive Clustering Algorithm. Procedia Comput. Sci. 54, 764–771 (2015). https://doi.org/10.1016/j.procs.2015.06.090.

[7] M. Luo, Yu-Fei Ma and Hong-Jiang Zhang. A Spatial Constrained K-Means Approach To Image Segmentation. Fourth International Conference on Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint, Singapore, 2003, pp. 738-742 vol.2, doi: 10.1109/ICICS.2003.1292554.

[8] Gupta, D. Image Segmentation Keras: Implementation of Segnet, FCN, UNet, PSPNet and other models in Keras. arXiv:2307.13215 (2023). https://arxiv.org/abs/2307.13215.

[9] Momin, A. Image Segmentation Using Deep Learning: Tensorflow and Keras Implementation in Python. Hildebrand Department of Petroleum and Geosystems Engineering, The University of Texas at Austin (2020). https://ahadmomin.github.io/files/Final%20Report-%20Ahad%20Momin.pdf.