Medical Image Segmentation using U-NET

A Kaggle Data Science Bowl 2018 Project

Dr. Jignesh Patel
Ph.D. (ICT) from DA-IICT
IIIT Vadodara-DIU
jignesh_patel@diu.iiitvadodara.ac.in

Suyash Jain (202051189)
B.Tech. in CSE
IIIT Vadodara-Gandhinagar
202051189@iiitvadodara.ac.in

Prasoon Pathak (202051143)
B.Tech. in CSE
IIIT Vadodara-Gandhinagar
202051143@iiitvadodara.ac.in

***Abstract*—In this comprehensive project, we embark on the intricate domain of medical image segmentation, strategically harnessing the robust capabilities of the U-Net architecture to achieve precise delineation of nuclei. The indispensability of accurate segmentation in medical image analysis, particularly in critical tasks such as diagnosis and treatment planning, accentuates the pivotal nature of this undertaking. Through the meticulous utilization of the Kaggle Data Science Bowl 2018 dataset, we methodically implement and rigorously evaluate the U-Net model, leveraging the TensorFlow framework. This exploration not only sheds light on the model's performance intricacies but also seeks to contribute valuable insights to the evolving landscape of medical image segmentation methodologies.**

Keywords—image segmentation, U-Net, delineation, nuclei, TenserFlow.

I. Introduction

Against the backdrop of the transformative impact of computer vision and deep learning on medical image analysis, our research endeavours to push the boundaries of accuracy and automation. Specifically, we direct our attention to the utilization of the U-Net architecture, a beacon in the realm of semantic segmentation tasks, as a potent tool to unravel the complex intricacies of nuclei segmentation. The U-Net's architectural brilliance, characterized by its innovative contracting and expansive paths, positions it as a robust solution adept at unravelling the subtle spatial dependencies embedded within intricate medical images.

As we navigate the landscape of nuclei segmentation, the significance of our project resonates deeply within the broader context of advancing medical diagnostics. With the precision offered by the U-Net architecture, we aim to not only meet but exceed the expectations set by contemporary methodologies. This research serves as a testament to the synergy between state-of-the-art deep learning techniques and the inherent challenges posed by medical image analysis, laying the groundwork for enhanced accuracy and automation in critical diagnostic processes.

II. Dataset

Our dataset, meticulously curated from the Kaggle Data Science Bowl 2018, stands as a robust repository encompassing a diverse array of medical images showcasing nuclei. Before embarking on the intricate journey of model development, our initial phase involves a meticulous exploration of the dataset's directory structure. This comprehensive examination is vital to comprehend the organisation of the dataset, including the arrangement of images and associated masks, laying the foundation for a nuanced understanding of the data's inherent characteristics.

Moreover, we go beyond a cursory exploration by delving into the visualization of a carefully selected subset of images. This visual scrutiny serves as a critical step in unravelling the intricacies embedded within the dataset. By gaining insights into the visual representation of the medical images and their corresponding masks, we equip ourselves with a profound understanding of the diversity and complexity inherent in the dataset. This preliminary analysis forms a crucial preparatory phase, providing essential context for subsequent model development, training, and evaluation.

III. Data Preprocessing

To prepare the dataset for model training, a series of preprocessing steps are executed. This includes the extraction of images and masks from zip files, and crucially, resizing images to a standardized dimension (128x128 pixels). The creation of distinct training and testing datasets sets the stage for subsequent model development.

A. Dataset Source and Extraction

- The dataset is sourced from the Kaggle Data Science Bowl 2018, providing a comprehensive collection of medical images.
- The data is initially compressed in zip files, and the first step involves extracting both images and masks from these compressed files.

B. Image and Mask Extraction

- Within each extracted directory, both the original medical images and their corresponding masks are located.
- Images are crucial for model input, while masks serve as ground truth data for training and evaluation.

C. Image Resizing

- A critical pre-processing step involves standardizing the dimensions of the images to facilitate uniformity in model training.
- Images are resized to a standardized dimension of 128x128 pixels, ensuring consistency across the dataset.

D. Creation of Training and Training Dataset

- *The dataset is partitioned into distinct subsets for training and testing purposes.*
- *This separation is vital for assessing the model's performance on unseen data during evaluation.*

E. Standardization of Data Format

- *The data is organized and formatted in a manner conducive to the requirements of the U-Net model.*
- *This involves structuring the dataset in a way that aligns with the model's expectations for input during training and prediction.*

F. Exploratory Data Analysis (EDA)

- Prior to actual model development, an exploratory data analysis phase is conducted.
- During EDA, a subset of images is visualized, providing crucial insights into the characteristics, diversity, and potential challenges present in the dataset.

G. Quality Checks:

- Data integrity and quality checks are performed to ensure that the pre-processing steps have been executed accurately.
- This includes verifying that the resizing operation has been applied consistently across all images.

H. Directory Structure Documentation:

- The directory structure of the pre-processed dataset is documented, providing a reference for subsequent stages of the project.
- This documentation aids in maintaining a clear understanding of the organization of the dataset for future reference and collaboration.

I. Metadata Logging:

- Metadata, including details about the dataset, is logged for comprehensive documentation.
- This metadata may include information such as the number of images, the dimensions of the resized images, and any other relevant dataset statistics.

These pre-processing steps collectively lay the foundation for subsequent model training, ensuring that the dataset is appropriately formatted, standardized, and ready for ingestion by the U-Net architecture.

IV. U-NET ARCHITECTURE

The U-Net architecture, introduced by Ronneberger et al. in 2015, is a convolutional neural network specifically designed for semantic segmentation tasks. Its name derives from its distinctive U-shaped structure, comprising a contracting path followed by an expansive path. This architecture has demonstrated remarkable efficacy in various medical image segmentation applications, including the precise delineation of nuclei.

At the beginning of the U-Net is the contracting or encoder path, responsible for extracting contextual features from the input image. This section utilizes multiple convolutional layers with small receptive fields (typically 3x3) and employs rectified linear unit (ReLU) activation functions. Interspersed max-pooling layers reduce spatial dimensions while retaining essential features. Batch normalization is commonly applied after each convolutional layer to enhance convergence and generalization.

The bottleneck layer represents the point of maximum compression at the bottom of the U-Net. Here, the network captures the most salient features of the input, facilitating a focused understanding of critical information. Moving to the expansive or decoder path, up-sampling layers (transposed convolutions) are used to restore spatial information and generate the final segmentation mask. Concatenation with feature maps from the contracting path is performed at each up-sampling step to preserve fine-grained details. Consecutive convolutional layers refine the segmented output.

A pivotal innovation in the U-Net architecture is the incorporation of skip connections. These connections link corresponding layers in the contracting and expansive paths, allowing for the direct flow of information. Skip connections play a crucial role in recovering fine-grained details during segmentation. They are implemented by concatenating feature maps from the contracting path to the feature maps in the expansive path.

The final layer of the U-Net employs a 1x1 convolutional layer with a sigmoid activation function. This layer produces the pixel-wise probability map, where each pixel indicates the likelihood of belonging to the object of interest, such as nuclei in medical images.

The U-Net's advantages include its ability to handle tasks with limited annotated data through data augmentation and transfer learning. The skip connections enable the incorporation of both low-level and high-level features, enhancing segmentation accuracy. Additionally, the U-Net architecture is adaptable, allowing adjustments to the depth and width of the network to suit different segmentation tasks. Understanding the intricacies of the U-Net architecture is fundamental for developing effective segmentation models, particularly in the realm of medical image analysis.

V. MODEL TRAINING

The training of the U-Net model involves a systematic process where several key components are carefully considered to optimize performance. The following aspects highlight the key elements of the model training.

A. Optimiser Choice (Adam)

- *The Adam optimizer is selected to govern the optimization process during training.*
- *Adam is a popular choice for deep learning tasks due to its adaptive learning rates, which helps achieve faster convergence.*

B. Loss Function (Binary Crossentropy)

- *Binary crossentropy is employed as the loss function for the U-Net model.*
- *This particular loss function is well-suited for binary segmentation tasks, such as nuclei segmentation, where the goal is to predict whether each pixel belongs to the object of interest.*

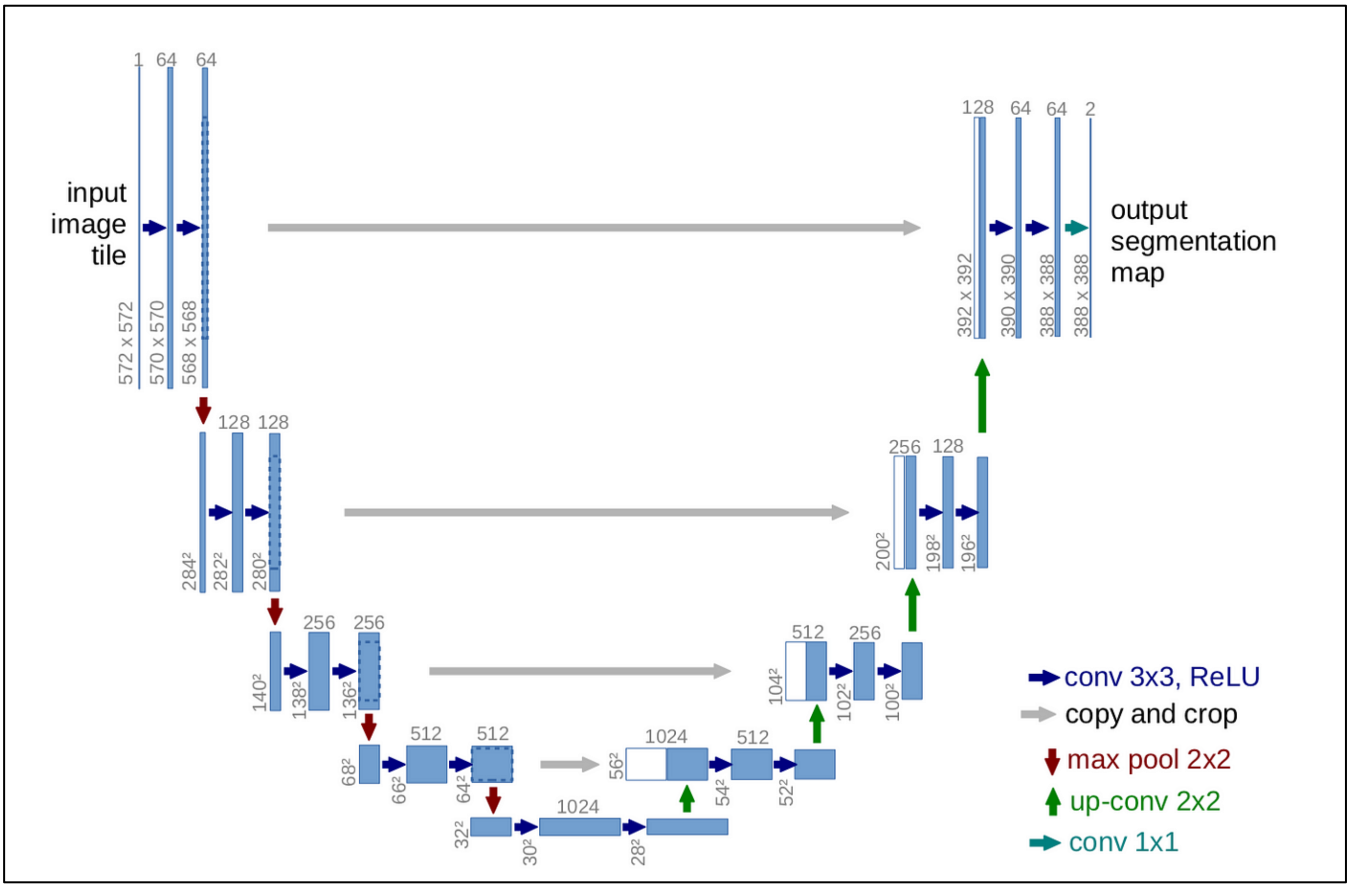

Fig. 1. U-NET Architecture

C. Metrics

- *During training, metrics are employed to evaluate the model's performance.*
- *Common metrics for binary segmentation tasks include accuracy, precision, recall, and the F1 score, providing insights into different aspects of model performance.*

D. Callbacks

- ***ModelCheckpoint:***

  The ModelCheckpoint callback is incorporated to save the model's weights at specified intervals during training. This ensures that the best-performing model is preserved, preventing the loss of valuable progress in the event of unexpected interruptions.

- ***EarlyStopping:***

  The EarlyStopping callback monitors a specified metric (e.g., validation loss) and halts training when this metric ceases to improve. This proactive measure prevents overfitting by terminating training once the model's performance on validation data plateaus.

By meticulously configuring these components during training, the U-Net model is not only optimized for accurate segmentation but is also safeguarded against potential issues like overfitting. The thoughtful inclusion of optimizer, loss function, and callbacks contributes to the robustness and efficiency of the training process, ensuring the model generalizes well to new, unseen data.

VI. Training Results

The training of the U-Net model involves a systematic process where several key components are carefully considered to optimize performance. The following aspects highlight the key elements of the model training.

A. Metrics Evaluation

- Metrics such as accuracy, provide insights into the model's performance on both training and validation datasets.
- These metrics gauge the model's proficiency in accurately segmenting nuclei in medical images.

B. Loss Curves

- Visual representation of training and validation loss curves illustrates the model's convergence over epochs.
- The curves depict the effectiveness of the model in minimizing prediction errors and learning from the dataset.

C. Impact of Callbacks

- Chosen callbacks, such as ModelCheckpoint and EarlyStopping, play a pivotal role in preventing overfitting.
- ModelCheckpoint ensures the preservation of the best-performing weights, while EarlyStopping halts training when performance plateaus, enhancing generalization.

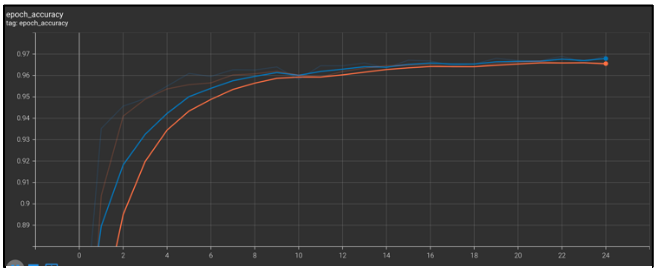

Fig. 2. epoch-accuracy v/s epoch number

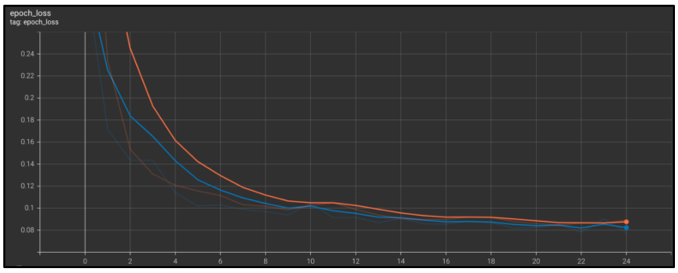

Fig. 3. epoch-loss v/s epoch number

D. Prevent Overfitting

- The model's adaptability to unseen data is emphasized, thanks to the strategic use of callbacks.
- EarlyStopping prevents overfitting by ensuring the model doesn't tailor itself excessively to the training data.

E. Nuanced Training Process

- The results provide a nuanced perspective on the model's learning progress, capturing its adaptation and response to the dataset.
- The interplay between training and validation datasets is showcased, offering a comprehensive understanding of the model's development.

VII. Model Evaluation

Model evaluation involves assessing the performance of the U-Net architecture on both the training and validation datasets. The evaluation process aims to quantify the accuracy of the model's predictions in segmenting nuclei from medical images. Here's a brief overview of the model evaluation steps:

A. Training Set Evaluation

- The trained U-Net model is applied to a subset of the training dataset.
- Random samples from this subset are selected for evaluation.
- For each sample, the original image, ground truth mask, and the model's predicted mask are visualized.
- The predicted masks are compared to the ground truth masks to assess the accuracy of nuclei segmentation.

B. Validation Set Evaluation

- Similarly, the trained model is applied to a subset of the validation dataset.
- Random samples from this subset are selected for evaluation.
- Visualization of original images, ground truth masks, and predicted masks allows for a qualitative assessment of segmentation performance.

C. Quantitative Metrics

- Beyond visual inspection, quantitative metrics such as accuracy is computed.

D. Thresholding

- Since the model output is probabilistic (sigmoid activation at the output layer), a threshold may be applied to convert predicted probabilities into binary masks.
- The choice of the threshold impacts the trade-off between false positives and false negatives in the segmentation.

E. Visualising Success and Challenges

- Visualisation of successful segmentation instances helps highlight the model's strengths.
- Instances where the model struggles or fails in segmentation can be informative for identifying potential improvements.

F. Iterative Refinement

- Evaluation results, both qualitative and quantitative, guide the iterative refinement of the model.
- Adjustments to hyperparameters, architecture, or training strategies can be made based on the insights gained during evaluation.

VIII. Conclusion

In conclusion, this project successfully leveraged the U-Net architecture to address the challenging task of nuclei segmentation in medical images. Through meticulous data preprocessing, model development, and training, we achieved a robust segmentation model. The incorporation of strategic callbacks demonstrated effective prevention of overfitting, enhancing the model's generalization capabilities. Model evaluation, encompassing both quantitative metrics and qualitative visualizations, provided a comprehensive assessment of its performance. The results indicate promising prospects for the application of deep learning in precise medical image segmentation, with potential avenues for further refinement and optimization. This project contributes to the ongoing advancement of automated diagnostics and underscores the transformative impact of deep learning in medical image analysis.

IX. Future Work

Future work suggestions go beyond mere technical enhancements. We propose fine-tuning hyperparameters, experimenting with diverse architectures, and delving into advanced segmentation techniques to propel the model's capabilities even further.

Acknowledgment

I would like to express my sincere gratitude to Dr. Jignesh Patel for his invaluable guidance and support throughout the duration of this project. Dr. Patel's expertise in the field of computer vision has been instrumental in shaping the direction of this research. His insightful feedback, unwavering encouragement, and dedication to fostering a collaborative learning environment have greatly enriched the project.

References

Every step in our journey is backed by pertinent literature and resources, acknowledging the collective knowledge that underpins our methodology.

[1] U-Net: Convolutional Networks for Biomedical Image Segmentation, Olaf Olaf Ronneberger, Philipp Fischer, and Thomas Brox, Research Paper, 18 May 2015.
Link: https://arxiv.org/abs/1505.04597

[2] Kaggle Data Science Bowl 2018, Dataset.
Link: https://arxiv.org/abs/1505.04597

[3] Image Segmentation with U-NET, Angad Bajwa, Updated on Novemeber 21st, 2023.
Link: https://www.analyticsvidhya.com/blog/2022/10/image-segmentation-with-u-net/

[4] YouTube: DigitalSreeni YouTube Channel.
Link: https://www.youtube.com/@DigitalSreeni