BIOINF 593, , pp. 1–5

doi:

Organ Segmentation from MRI images using U-Net

Farzad Siraj, Ashwin Saxena, Clark Wakeland and Yuyang Xia

EECS, University of Michigan

Abstract

Motivation: Cross sectional body images are one of the primary tools employed by medical professionals when diagnosing and assessing a patient's condition. The accurate identification and segmentation of organs in medical images, particularly computed topography (CT) scans, is a crucial step in modern healthcare. Precise organ segmentation enables clinicians to diagnose and treat diseases more effectively, but manual segmentation is time-consuming and prone to errors. Our project aims to address this challenge by developing a machine learning solution using a U-Net architecture for organ segmentation in MRI images.
Results: The trained U-Net model produced training and validations accuracies of 0.954 and 0.918 respectively compared to the given ground truth data in terms of measured overlapping surface area. Additionally, the relative absolute volume difference between the segmented and reference sets was 4.485% for the training set and 7.01% for the validation set.
Availability: All results, tools, and models are available on request. The python notebook used for training is available here.
Contact: fsiraj@umich.edu, ashwinsa@umich.edu, cwakelnd@umich.edu, yyxia@umich.edu

Key words: Organ Segmentation, U-Net, CHAOS Challenge

Introduction

Within the past 50 years, the development of cross sectional body imaging has proven to be a fundamental diagnostic tool for medical professionals. However, the images are only as valuable as the information that can be interpreted from them, leading to a need for in depth analysis and extensive training for the professionals utilizing these images.

Manual segmentation has been the conventional method for identifying organs within body scans. This process demands considerable time and effort from radiologists and professionals while being inherently subjective and prone to human errors. As the volume of medical imaging data continues to grow, the need for efficient, consistent, and error-free segmentation methods becomes increasingly evident.

In response to the increasing demands of the medical industry for analysis of body images, the CHAOS challenge was created. The CHAOS (Combined (CT-MR) Healthy Abdominal Organ Segmentation) Challenge is a scientific competition and dataset in the field of medical image analysis. It was created to encourage the development of advanced algorithms and models for the precise segmentation of abdominal organs in medical images, focusing on the liver, spleen, and kidneys [2] .

In pursuit of solutions to the CHAOS challenge, we employed a state-of-the-art U-Net model and trained it using given ground truth values from CT and MRI images provided in the CHAOS dataset. The U-Net architecture offers a framework for addressing the organ segmentation through the use of an encoder capturing feature information and a decoder upscaling the information to generate a segmentation mask.We employed a combination of Dice and Binary Cross Entropy loss functions when training the model to obtain appropriate levels of feature identification for the organs of interest.

This model has the ability to reduce the time spent on analysis of CT and MRI images, which, in turn, would lead to expedited and more efficient diagnoses. Instead of relaying on manual processing by medical professionals, images could be quickly and accurately processed, allowing for immediate distribution to relevant stakeholders. This streamlining of analysis would enable doctors to allocate more of their time and expertise toward enhancing the overall well being and care of their patients.

Background

There are multiple clinical reasons to accurately measure the volume, size, and shape of healthy abdominal organs. The liver for example can be influenced by various diseases such as congestive heart failure, cancer, cirrhosis, infections, metabolic disorders, and congenital diseases. Analyzing the liver's dimensions can offer insights into the disease's severity. Additionally, monitoring the liver's growth pattern and changes during treatment can provide valuable information about the disease's significance and prognosis. Therefore, it is crucial to assess whether the liver is enlarged, calculate its volume, and identify associated effects accurately. Precise segmentation is also essential for planning liver transplant surgeries. Determining whether a portion of the liver to be resected is sufficient for the recipient patient and whether the remaining liver will be sufficient for the donor is an important part of treatment decisions. Taking into account other organs,

the spleen enlarges in cases of portal hypertension, infections, and various other diseases, making a precise organ segmentation and volumetric modelling of it clinically valuable. Additionally, segmentation of the kidneys can be used for observation of tissue thickness and monitoring the development of potential tumors on the organs.

Unfortunately, the amount of training data for organ segmentation is limited due to the high expense of gathering and annotating medical volumetric datasets [3]. Deep models trained on relatively small datasets generally result in overfitting and lack of satisfactory general performance. As a result, historical attempts to create segmentation models have either had unsatisfactory performance or are unfeasible to train. In response to this, leading medical researchers in the field developed the CHAOS challenge to encourage new development and new training methods of these models.

CHAOS Challenge

Understanding prerequisites of complex medical procedures plays an important role in the success of the operations. Physicians have to use advanced tools such as three-dimensional visualization and printing, which require extraction of the object(s) of interest from DICOM images. Accordingly, the precise segmentation of abdominal organs (i.e. liver, kidney(s) and spleen) has critical importance for several clinical procedures including like pre-evaluation of liver for living donor-based transplantation surgery. This motivates ongoing research to achieve better segmentation results and overcoming countless challenges originating from both highly flexible anatomical properties of abdomen and limitations imposed by the characteristics of the image modalities. The CHAOS challenge was held to motivate research in this field [4].

Specifically, the CHAOS challenge encompassed two primary objectives:

1. Segmentation of the liver from CT data sets, which are typically acquired at the portal phase following the injection of contrast agents. This segmentation is essential for the pre-evaluation of living donors for liver transplantation.
2. Segmentation of four essential abdominal organs—namely, the liver, spleen, and right and left kidneys—from MRI data sets. These data sets are acquired using two different MRI sequences: T1-DUAL and T2-SPIR.

Notably, the CHAOS challenge presented tasks that involved the combination of segmenting multiple organs, recognizing the intricate nature of medical imaging and the holistic approach required for comprehensive abdominal organ segmentation. Through this challenge, the medical research community was inspired to innovate, collaborate, and ultimately contribute to advancements in the field of abdominal organ segmentation, thereby enhancing the success and precision of vital clinical procedures.

Methodology

Data Preparation

The CHAOS dataset consists of both MRI and CT scans, and for the purposes of this project we focused solely on MRI, of which we have both T1-DUAL and T2-SPIR images. Our dataset consisted of scans from 20 patients of both MRI modalities for a total of 1270 grayscale 500×500 images in DICOM format. Each has a corresponding label image where the presence of organs is indicated by specific float values, for example, the liver is 0.24705882.

For the images, we resized each to 128×128 and then took a 96×96 center crop as the bordering pixels were just background. Then we normalized the pixel values to the range $[0, 1]$ and 0-centered them. For the label images, we applied the same resizing and center-cropping and converted them from a 1-channel image to a 4-channel one-hot encoding where each channel corresponded to 1 of the 4 organs present in the dataset.

This dataset was split into training and validation sets with a $80 : 20$ split. The test set was separate with no labels provided.

Model Architecture

For our model, we chose the U-Net architecture [6] which was developed specifically for the purpose of Biomedical Image Segmentation where there is a limited quantity of data. This is a convolutional neural network that follows a very similar architecture to auto-encoders where there is a contraction path (analogous to an encoder) and an expansion path (analogous to decoders). One major difference is that there are residual connections between each layer in the contraction path with the corresponding layer in the expansion path.

Our implementation mimics the original design presented in Figure 1, however, we also added Batch Normalization layers following each Convolution layer, and used a Leaky ReLu activation function. The expansion path took as input a $1 \times 96 \times 96$ image and halved the spatial dimensions in each step using MaxPooling layers taking the 96×96 input down to 6×6 and doubled the channel dimensions from 64 to 1024. The expansion path reversed this until we got a final output of $4 \times 96 \times 96$. Simple Upsampling layers were used to increase the spatial dimensions.

This gave us a model with a total of $34,526,084$ trainable parameters across 74 layers.

Hyperparameter Tuning

We tried learning rates ranging from 1e1 to 1e$-$5 and found that 1e$-$3 gave us good convergence. We also tried using various weight decay values but found that they restricted our model too much and finally settled with a value of 0 which surprisingly did not cause overfitting. These parameters were used in conjunction with an Adam optimizer with betas $= (0.9, 0.999)$. For batch size, we tried various powers of 2 and went with a typical 32 as we found that the only difference this made was the speed of training and convergence.

Loss Metrics

We initially used exclusively binary cross entropy loss as this is a multi-class classification problem, however, we found that this caused our model to predict background values for all pixels as that is overwhelmingly the most prevalent class. To remedy this, we also used Dice score and made our final loss function a weighted combination of Dice and Binary Cross Entropy loss. After iteration of various weight combinations, we settled on 50$-$50 as that gave satisfactory results.

Training

We trained our model for 100 epochs as we did not see any significant improvement beyond that. Training was done on a Tesla T4 GPU in a Google Colab environment and took

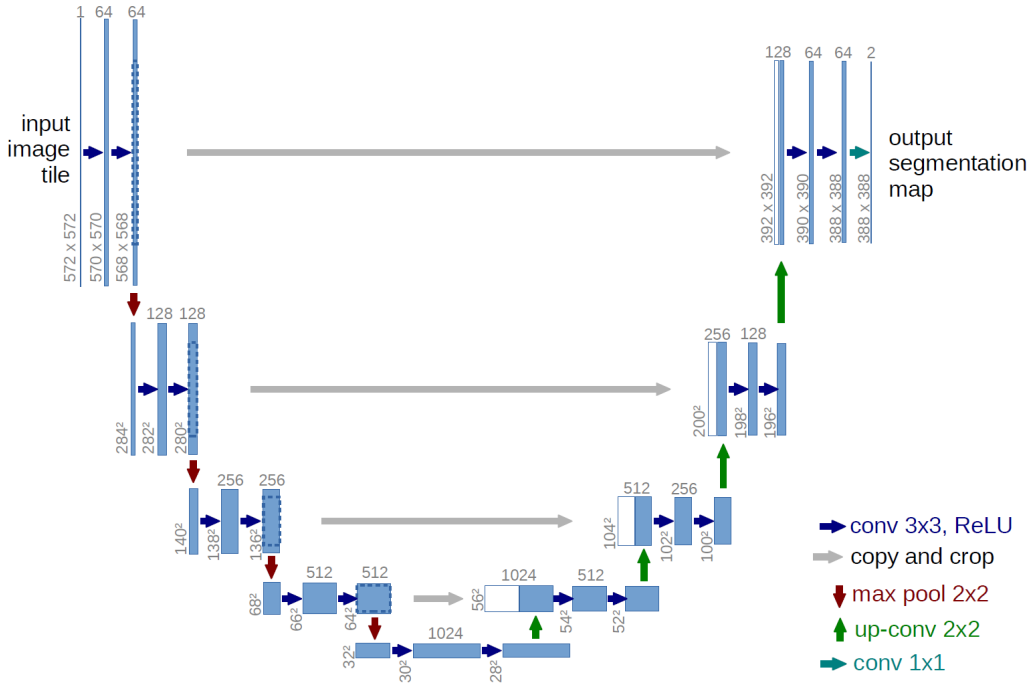

Fig. 1. Base U-Net architecture

approximately 1 hour. We discuss our results in the sections that follow.

Evaluation of Methodology

Training Losses and Validation Losses

During training, we employed a combination of loss functions to guide the model. The Dice loss aimed to optimize the overlap between predicted and ground truth masks, while Binary Cross-Entropy (BCE) loss addressed pixel-wise classification. To handle class imbalance, we introduced a weighted loss, assigning higher importance to minority class pixels.

Figure 2 shows the training losses and validation losses using Weighted loss, Binary Cross-Entropy (BCE) loss and Dice loss. We found all of them rapidly decreased in the beginning, and reached plateau around epoch 50. The BCE loss has the least loss values among them, suggesting the model is effectively fitting the training data with respect to binary classification, and with BCE loss the model is learning more quickly and effectively.

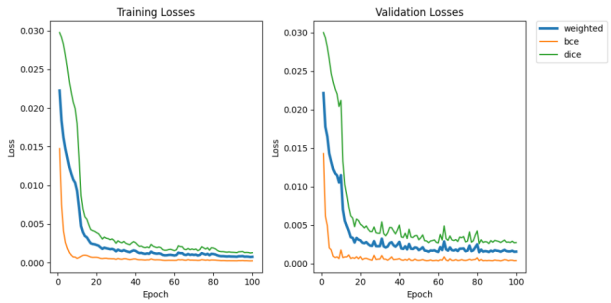

Fig. 2. Training Losses and Validation Losses

Notice that there are several subtle increases in validation loss occurs in later epochs for Dice Loss and Weighted loss, implying some possible approaches for future exploration and improvement, such as refining the model architecture, optimizing hyperparameters, etc.

The utilization of the U-Net network in our model was strategic, leveraging its training approach that heavily incorporates data augmentation. This methodology efficiently maximizes the utility of our limited annotated samples, enabling effective training even with a relatively small dataset. With only scans from 20 patients at our disposal, the choice of U-Net aligns seamlessly with our data constraints, proving to be an apt and resource-efficient solution for our specific dataset.

Other Possible Approaches

There are also other possible network models that we can use for our project.

One strategy is to implement a Recurrent Convolutional Neural Network (RCNN) [1], incorporating U-Net and Residual Network. It would be helpful for training deeper networks and improve feature representation for segmentation tasks. Another option is to utilize the Attention U-Net [5], which removes extraneous image features, reducing computational overhead and enhancing model sensitivity and prediction accuracy. Alternatively, the Nested U-Net approach connects networks through nested, dense skip pathways, which reduces the semantic gap between the feature maps of the encoder and decoder sub-networks [7].

It is crucial to acknowledge that the availability of training datasets for medical segmentation tasks is typically constrained. The considerable effort involved in collecting and annotating medical data often limits the dataset size. In this case, designing an excessively complex model might not lead to improved performance, since such models have high chances leading to model overfitting due to the limited dataset size.

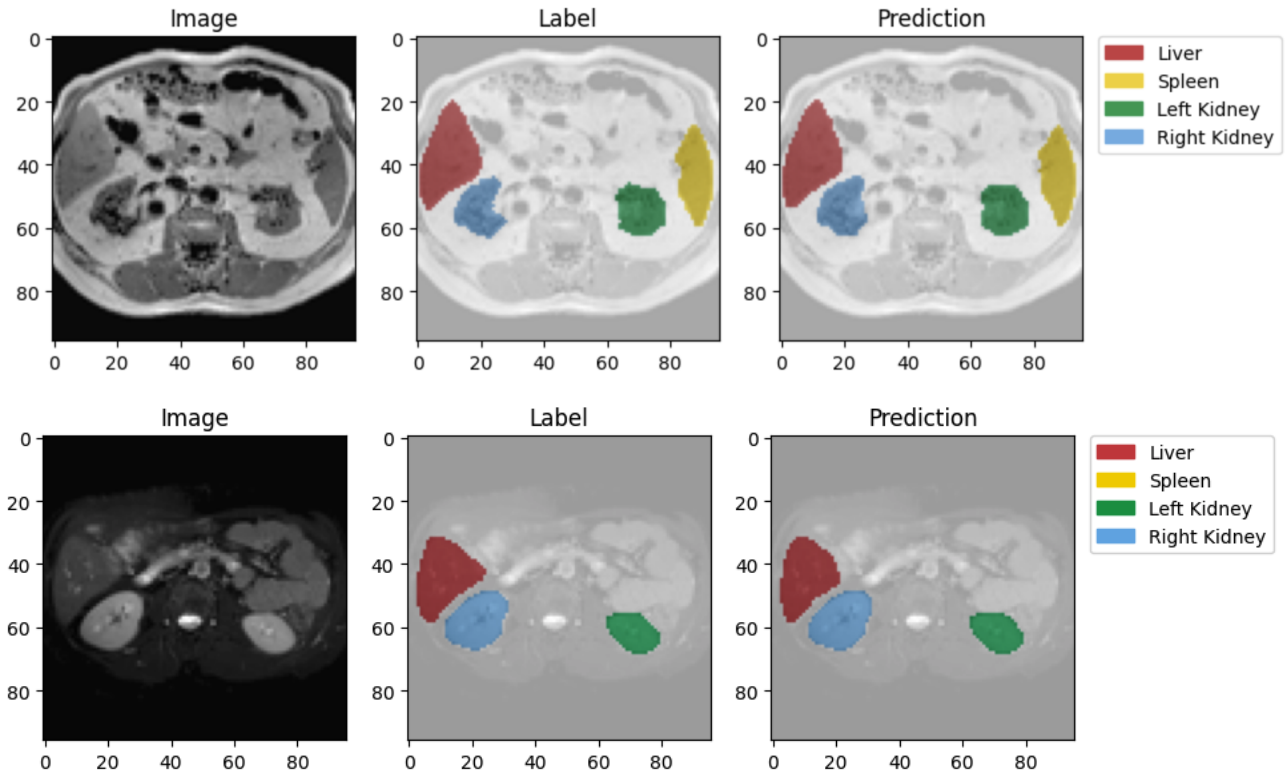

Fig. 3. Example segmentation results from our U-Net model from the validation set. Our model is able to segment all four organs that were labeled in the CHAOS dataset. Note: Not all organs are visible or labelled in some of the data.

Results

Quantitative Results

According to previous studies in literature, it is not possible to define a single evaluation metric for the organ segmentation problem. That is why, multiple evaluation metrics are used similar to the challenges before (i.e. SLIVER07). The four evaluation metrics used for evaluating CHAOS challenge are the following:

1. **Sørensen–Dice coefficient (DICE):** This value provides information about the overlapping parts of segmented and reference volumes in mm3. This value is 1 for a perfect segmentation, 0 for the worst case (no overlap).
2. **Relative absolute volume difference (RAVD):** This provides information about the differences between volumes between segmented and reference organs, but values the differences more than overlap. This value is 0% for a perfect segmentation and 100% for the worst case.
3. **Average symmetric surface distance (ASSD):** This value represents the average difference between the surface of the segmented object and the reference in 3D. After the border voxels of segmentation and reference are determined, those voxels that have at least one neighbor from a predefined neighborhood that does not belong to the object are collected. For each collected voxel, the closest voxel in the other set is determined and the average of all these distances gives the ASSD. This value is 0 mm for a perfect segmentation and the max distance of image for the worst case.
4. **Maximum symmetric surface distance (MSSD):** This value is similar to ASSD but particularly important for surgical operations as it determines the maximum margin of error by selecting the biggest of all calculated distances (0 mm for a perfect segmentation, max distance of image for the worst case).

	DICE	RAVD	ASSD	MSSD
Train Set	0.954	4.485	0.018	5.776
Validation Set	0.918	7.010	0.044	7.811

Table 1. Evaluation Metrics of our methodology on train and validation datasets

We calculated the aforementioned metrics for our model on the training and validation sets. The metrics are shown in Table 1. Our Dice Sørensen–Dice coefficient (DICE) scores are reasonably close to 1 conveying that we have a well trained segmentation model. The RAVD values are also in the single digits. While both ASSD and MSSD values are close to 0.

Qualitative Results

The CHAOS dataset contained labels for training and validation set, however, due to the challenge aspect of the CHAOS dataset, there are no publicly available labels for the test dataset. We ran our model on the dataset and evaluated the results. As shown in Figure 3 our model performs well at segmenting the organs correctly. The area and positions of the predicted segmented regions are very similar to the label areas.

Figure 4 shows our model running on one of the test dataset. From the original image, we can see the regions the model

predicts correspond fairly well to the rough shapes of the organs.

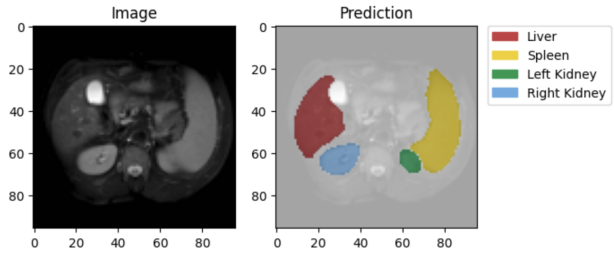

Fig. 4. Example result of our model running on Test Data set.

Overall, both the quantitative and qualitative show that our modifications to the UNet model were quite successful. The addition of a BatchNorm layer and the transformations applied to the images were a vital innovation that aided in the success of our model in segmenting abdominal organs.

Future Work

In our project, the primary emphasis centered on the segmentation of multi-modal MRI images (liver, spleen, and kidneys) sourced from the CHAOS challenge, leveraging the accompanying ground truth data. Our model is expected to work on both T1-DUAL (in-phase and oppose-phase) and T2-SPIR sequences. Looking ahead, our aspirations extend to including liver segmentation within the context of CT images that also contained in the CHAOS challenge, and broadening our scope to encompass the segmentation of various organs using mixed datasets that amalgamate both CT and MRI modalities since the input data for CT and MRI are the same.

Furthermore, our conviction extends beyond the immediate focus on liver, spleen, and kidneys. We envisage the adaptability of our segmentation model to transcend organ-specific tasks, demonstrating efficacy in diverse datasets. This could extend to applications beyond our primary targets, encompassing tasks such as brain segmentation, lung segmentation, bone segmentation, whole-body segmentation, and more. We believe that the versatility of our model positions it as a valuable asset across a spectrum of medical imaging datasets.

We could also explore the potential of enhancing segmentation accuracy by finding alternative models beyond the U-Net architecture. This includes investigating the efficacy of models like RCNN, Attention U-Net, and Nested U-Net. The diversification in model selection aims to optimize segmentation performance and explore the strengths and capabilities offered by these different architectures.

Conclusion

The overall performance of the model obtained is sufficient enough to warrant further experimentation and implementation. Our model achieved promising quantitative results, with Dice scores close to 1, relative absolute volume differences in single digits, and average symmetric surface distances and maximum symmetric surface distances close to 0, demonstrating its effectiveness in organ segmentation. Qualitatively, our model successfully segmented organs in MRI images from the validation set. Future architecture may involve exploring alternative network architectures, such as RCNN, Attention U-Net, and Nested U-Net, to further enhance segmentation accuracy and efficiency.

Indeed, applications beyond the training set of abdominal imaging immediately become clear. Similar model training techniques could be used in other areas of the body that are frequently imaged such as the upper torso and brain. Upper torso imaging could benefit from lung segmentation, heart delineation, breast tumor identification, and spinal cord analysis, while brain imaging could benefit from the segmentation of tumors, brain structures, and stroke lesions. The multitude of structures within these regions suggests diagnostic applications for our model in these areas, potentially altering the conventional method of processing medical images and ultimately enhancing patient care. Additionally, the versatility of our approach opens doors for broader research into automating complex image analysis tasks across various medical specialties.

We hope our work inspires other researchers to pursue the development of new architectures and training techniques for image segmentation and volumetric organ modelling.

References

1. Md Zahangir Alom, Mahmudul Hasan, Chris Yakopcic, Tarek M. Taha, and Vijayan K. Asari. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation, 2018.
2. A. Emre Kavur, N. Sinem Gezer, Mustafa Barış, Sinem Aslan, Pierre-Henri Conze, Vladimir Groza, Duc Duy Pham, Soumick Chatterjee, Philipp Ernst, Savaş Özkan, Bora Baydar, Dmitry Lachinov, Shuo Han, Josef Pauli, Fabian Isensee, Matthias Perkonigg, Rachana Sathish, Ronnie Rajan, Debdoot Sheet, Gurbandurdy Dovletov, Oliver Speck, Andreas Nürnberger, Klaus H. Maier-Hein, Gözde Bozdağı Akar, Gözde Ünal, Oğuz Dicle, and M. Alper Selver. CHAOS Challenge - combined (CT-MR) healthy abdominal organ segmentation. *Medical Image Analysis*, 69:101950, April 2021.
3. A. Emre Kavur, Naciye Sinem Gezer, Mustafa Barış, Yusuf Şahin, Savaş Özkan, Bora Baydar, Ulaş Yüksel, Çağlar Kılıkçıer, Şahin Olut, Gözde Bozdağı Akar, Gözde Ünal, Oğuz Dicle, and M. Alper Selver. Comparison of semi-automatic and deep learning based automatic methods for liver segmentation in living liver transplant donors. *Diagnostic and Interventional Radiology*, 26:11–21, January 2020.
4. Ali Emre Kavur, M. Alper Selver, Oğuz Dicle, Mustafa Barış, and N. Sinem Gezer. CHAOS - Combined (CT-MR) Healthy Abdominal Organ Segmentation Challenge Data, April 2019.
5. Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention u-net: Learning where to look for the pancreas, 2018.
6. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.
7. Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation, 2018.