8-2023

Towards Reliable Hepatocytic Anatomy Segmentation in Laparoscopic Cholecystectomy Using U-Net with Auto-Encoder

Koloud Najem Alkhamaiseh
Western Michigan University

TOWARDS RELIABLE HEPATOCYTIC ANATOMY SEGMENTATION IN LAPAROSCOPIC CHOLECYSTECTOMY USING U-NET WITH AUTO-ENCODER

Koloud Najem Alkhamaiseh, Ph.D.

Western Michigan University, 2023

Despite the advantages of minimally invasive surgeries that depend heavily on vision, the indirect access and lack of the 3D field of view of the area of interest introduce some complications in the desired procedures. Fortunately, the recorded videos from these procedures offer the opportunity for intra-operative and post-operative analyses, to improve future performance and safety.

Deep learning models for surgical video analysis could therefore support visual tasks such as identifying the critical view of safety (CVS) in laparoscopic cholecystectomy (LC), potentially contributing to the reduction of the current rates of bile duct injuries in LC. Most bile duct injuries during LC occur due to visual misperception leading to the misinterpretation of anatomy.

In this study, a deep neural network is proposed comprising a segmentation model to highlight hepatocytic anatomy to predict CVS criteria achievement during LC. This network was trained and tested using 200 LC videos, which include the cholec80, m2cai16-tool, and m2cai16-workflow challenges datasets, the World Laparoscopy Hospital videos, and the Society of American Gastrointestinal and Endoscopic Surgeons (SAGES) videos. Expert surgeons preprocessed and reviewed these videos to train the proposed model. The model takes advantage of the effectiveness of Auto-Encoder weights as starting weights of the U-Net encoder to explore accurate medical image segmentation. Semantic segmentation is more appropriate than the bounding boxes method, which may produce incorrect or overlapping annotations of structures.

U-Net has been chosen because its network architecture is very simple, and straightforward, and it can be trained fast. In addition, U-Net creates highly detailed segmentation maps using very few samples, and that is a very important aspect for the medical imaging community because the number of available labeled images is often quite low. These properties make U-Net a suitable choice.

Five experiments were developed to prove the efficiency of the proposed model, where each experiment was trained with different weights and data preprocessing approaches. Each was then tested on a part of the dataset. Five cross-validation techniques were deployed for each experiment to estimate the performance of unseen data, and to ensure that the model can be generalized. A hybrid loss function was used to calculate the loss of the output results from different levels, and other evaluation metrics were calculated to evaluate the model's performance as a segmentation model.

The efficiency of the proposed approach outperforms some of the state-of-the-art studies reported in the literature for automatic hepatocytic landmarks identification with an accuracy of 92%, 93.9% for precision, and 74.7% for mean IoU, despite the limited number of videos available for this study. In addition, the proposed model has proven its efficiency in segmenting challenging cases.

Moreover, the model can segment dynamic LC videos and generate two views of the segmented video. This work can analyze dynamic videos assuming that it would facilitate CVS identification.

Finally, 1,550 laparoscopic cholecystectomy image frames from 200 video clips were selected and annotated at the pixel level for 5 classes confirmed by expert surgeons, which are identified as CVS in laparoscopic cholecystectomy surgery.

TOWARDS RELIABLE HEPATOCYTIC ANATOMY SEGMENTATION IN LAPAROSCOPIC CHOLECYSTECTOMY USING U-NET WITH AUTO-ENCODER

by

Koloud Najem Alkhamaiseh

A dissertation submitted to the Graduate College
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
Electrical and Computer Engineering
Western Michigan University
August 2023

Doctoral Committee:

Janos L. Grantner, Ph.D., Chair
Ikhlas Abdel-Qader, Ph.D.
Saad Shebrain, M.D.

ACKNOWLEDGMENTS

I am extremely grateful to my husband and my kids for their support and patience during my academic journey.

I would like to express my sincere gratitude to Dr. Saad Shebrain for his guidance, and constant encouragement and support. His mentorship was principal in providing me a varied experience that helped me to shape my long-term career goals in research, and for his surgical knowledge, skills and procedural consultation of the work in this dissertation

Also, it is a pleasure to thank to my dissertation advisor Prof. Janos Grantner and Prof. Ikhlas Abdel-Qader for serving on my dissertation committee.

This dissertation and all research publications have not been possible without their instructive comments and valuable suggestions.

Koloud Najem Alkhamaiseh

TABLE OF CONTENTS

Table of Contents—continued

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Surgeries are a pivotal part of medical procedures and especially minimally invasive surgeries that are inserted in different and major parts of surgical treatments [1]. In this technique, the surgeon moves and controls the surgical instruments inserted in the abdominal cavity through hollow tubes called Trocars by monitoring a video stream captured by an endoscopic camera inserted through the navel.

The main goal of Minimally Invasive Surgeries (MIS) is to reduce the size and the number of incisions compared to traditional open surgery, this will result in less pain and blood loss, shorter time to recovery, and reduced risks of infection [2]. As a potential example, laparoscopic cholecystectomy video analysis can be used for improving procedure performance and/or safety. Despite these significant advantages, it remains quite complex from the surgical standpoint due to limited surgical vision and indirect manipulation i.e., lack of depth perception and tactile sensation [3].

Laparoscopic Cholecystectomy (LC) started in 1985 [4], like other laparoscopic procedures, patients preferred it over open surgery [5], as it is a safe and often outpatient procedure performed by almost a million US citizens annually [6].

Laparoscopic cholecystectomy (LC) is the standard procedure for symptomatic cholelithiasis and cholecystitis and remains one of the most commonly performed procedures in the United States and across the world, averaging 750,000 cases per year in the US alone [7]. It is also one of the most common procedures performed during general surgery residency. Unfortunately, LC is associated with a rare but very dreadful complication, iatrogenic bile duct injuries (BDI) with an incidence of 0.3% (3 per 1000 procedures) [8], which translates into approximately 2,500-3,000 bile ducts injury a year in the USA. This complication is associated with a threefold increase in mortality at one-year reported [9], and costs about one billion dollars in the United States alone, in one year [10].

Bile duct injury usually occurs because of visual perceptual illusion (i.e., the surgeon believes to be clipping and cutting the cystic duct while actually dividing the common bile duct) [11,12]. To prevent the visual illusion leading to BDI, Strasberg et al. [11] introduced the Critical View of Safety (CVS) as a technique for the conclusive identification of the cystic duct. To achieve CVS, a surgeon should clean the hepatocytic triangle from fatty and fibrous tissue and anatomize the gall bladder lower part from the liver bed to get a clear view of only 2 tubular structures, i.e., the cystic duct and artery, entering the gallbladder [13], as shown in Figure 1.1.

However, overconfidence with LC and deficient implementation of CVS explain the non-decreasing rate of BDI [14]. Additionally, there is a lack of CVS visual records analysis, as the CVS assessment is qualitative and subject to observer interpretation and accomplished manually by either recorded videos or doublet photography [14, 15]. Thus, an automatic method for identifying CVS and highlighting landmarks anatomy in LC videos is desirable for surgical assessment and quality improvement.

Surgical data science (SDS), a novel multidisciplinary field based on the large amount of digital data gained from surgery to improve surgical care could be exploited to address this concern [16]. Recently, the advances in computer vision (CV) coupled with the rich availability of surgical videos has resulted in Artificial Intelligence (AI) applications as deep learning algorithms that can train machines to analyze and extract meaningful information from intraoperative videos [17] and could provide valuable feedback on critical steps, which should improve surgical assessment and surgical care.

Nowadays, deep learning models are used in minimally invasive surgeries for both medical imaging and hardware-based solutions, by processing information through multiple neuron layers networks [18], where convolutional neural networks (CNNs) are the common neuronal arrangement, that has a structure similar to human's visual cortex, allowing it to learn visual object recognition complex tasks [19].

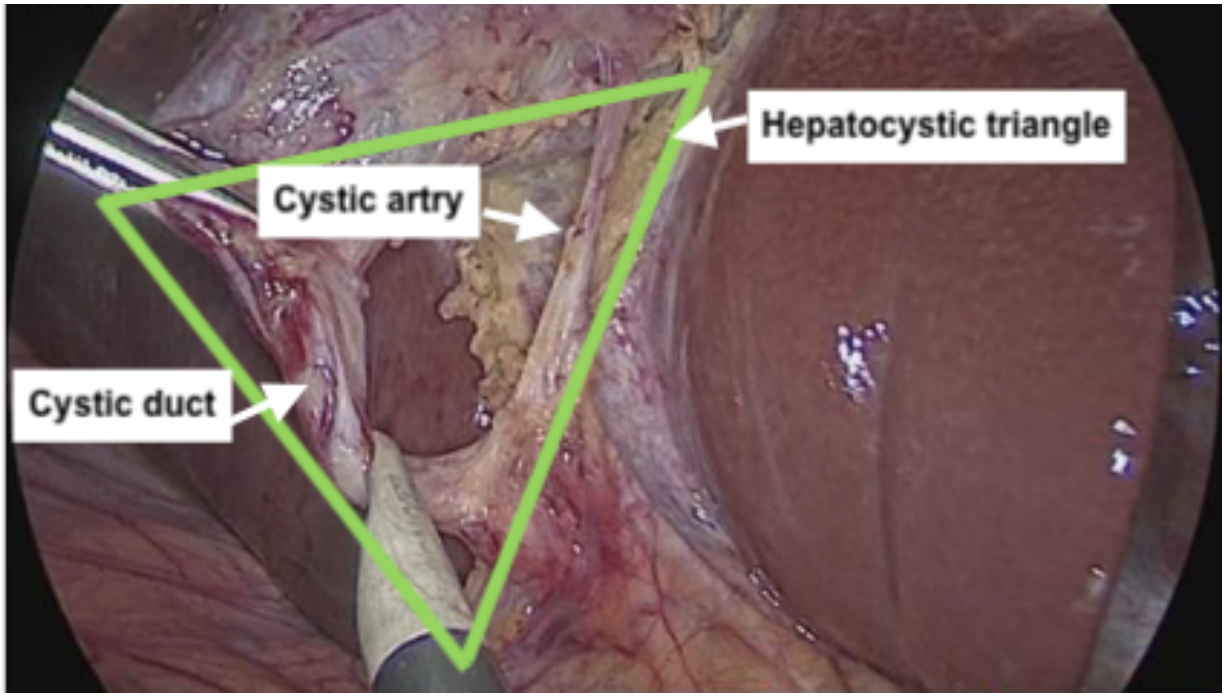


Figure 1.1:The critical view of safety, where the cystic duct and artery, and the hepatocystic triangle entering the gallbladder [41].

However, analyzing surgical videos is challenging, as surgery is filled with visual disturbances i.e. surgical smoke, blood, blurry cameras, no clear view of anatomical structures and surgical planes, and lying hidden within other tissues [20]. Further, artery/vein classification is more challenging as there is a shortage of a gold standard that allows objective comparison of approaches for this problem [21]. In order to that, a robust and accurate image analysis model is needed, that can be applied to intraoperative video analysis.

With the rapid and significant improvement in artificial intelligence, especially deep learning (DL) [22], image segmentation models based on deep learning achieved great performance, compared with traditional machine learning and computer vision models, deep learning gain better segmentation accuracy and speed.

However, such a potential of image segmentation models is only limited by the need for a large amount of annotated data i.e., a fully supervised model. Furthermore, the process of generating huge, annotated data needs experts with extensive manual efforts. This process is a very tedious, expensive task in terms of time, and subjective to human error. To that, different strategies are used to train the model with limited annotated data such as self-supervised learning, transfer learning, and data augmentation.

Medical image segmentation is the common approach used for highlighting tissue abnormalities, key organs, or blood vascular systems [23]. The encoder-decoder U-Net model [24] is widely used in medical image segmentation, in which multi-scale features are combined to produce an accurate segmentation. On the other hand, an autoencoder is a type of artificial neural network that learns data representations in an unsupervised manner to capture the most important parts of the input image by ignoring the signal noise. Practically, autoencoder is used as a data preparation technique to perform feature extraction on raw data that can be used to train a different machine learning model. Additionally, an autoencoder is a type of artificial neural network that learns data representations in an unsupervised manner to capture the most important parts of the input image by ignoring the signal noise.

U-Nets and Autoencoders solve two completely different problems. While U-Net works on image-to-image translation, autoencoder encodes and decodes a high-dimensional data space into a lower dimensional representation. In general, the architectures of the two models look quite similar, potentially they are used for different tasks, U-net is used to predict masks in image segmentation, and autoencoders try to represent data in lower dimensional space.

The classical architecture of autoencoder works on reducing the input as it goes through the layers of its encoder blocks, where at the final layer of the encoder the input is represented as a linear feature. Then, the decoder works on upsampling the gained linear feature representation. So that the output of the autoencoder unit is of the same dimension as the input it received. Practically, this architecture deploys a linear compression of the input that leads to a bottleneck that does not transmit all features, while it is ideal in preserving the dimensionality of input/output.

The U-Net architecture also has an encoder and decoder, but it uses deconvy blocks and adds skip connections to overcome the bottleneck limitation, which allows feature representations to pass through the bottleneck.

The similarity between U-Net and autoencoder architecture inspirits us to utilize it to build a segmentation model that can transfer gained features from unsupervised learning of autoencoder on part of the prepared dataset, which will be used as starting weights of U-Net segmentation model to identify the intended landmarks in LC procedures, the general

architecture of the proposed model is illustrated in figure 1.2 below, where training U-Net with initial encoder pre-trained weights would dramatically speed up the training process.

Practically, autoencoder is used as a data preparation technique to perform feature extraction on raw data that can be used to train a different machine learning model.

The main contribution of this work is to build a deep neural network that combines Auto-Encoder with a U-Net segmentation model to automatically highlight cystic duct and cystic artery anatomy which is very important in the prediction of CVS criteria achievement.

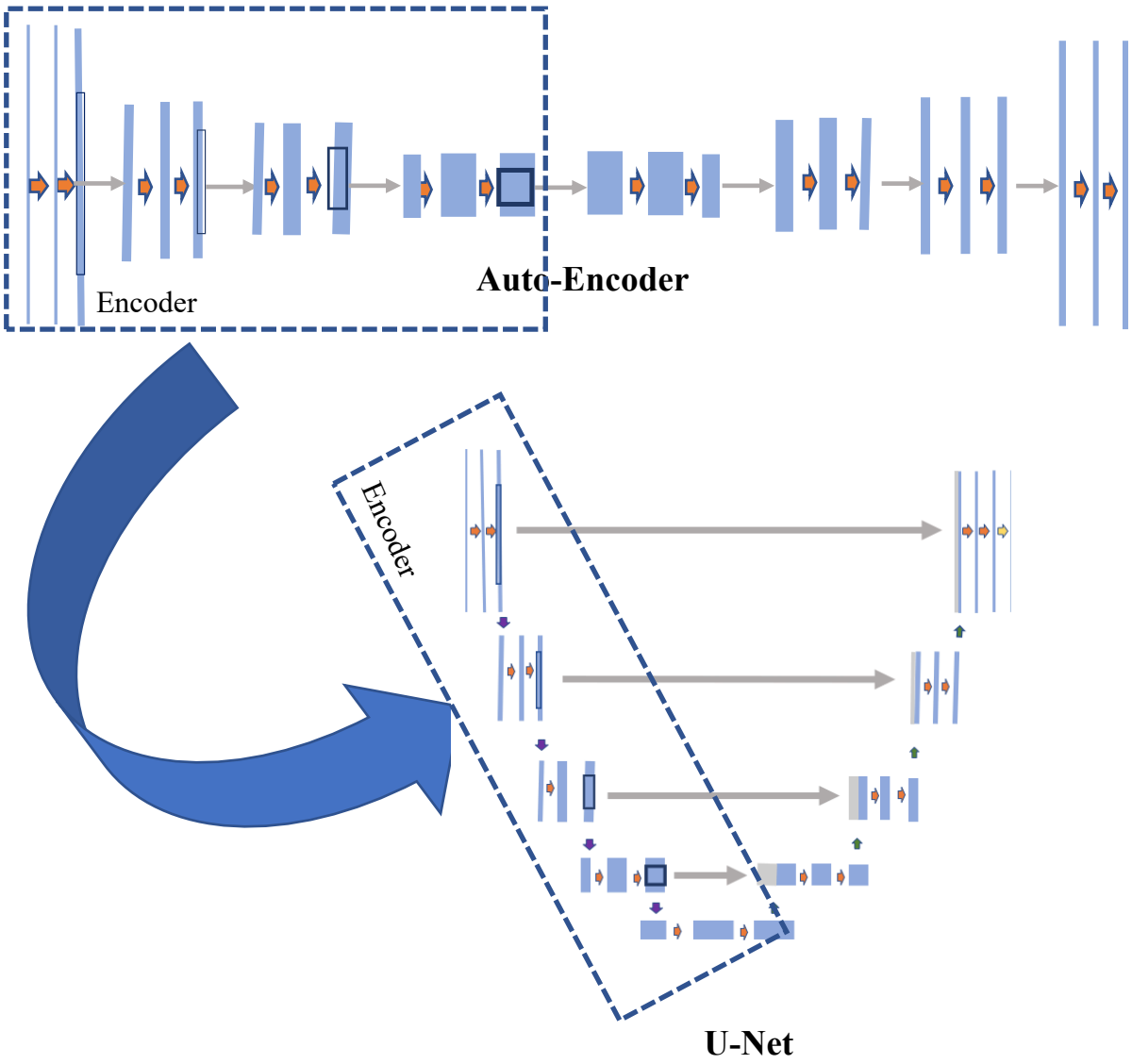


Figure 1.2: Transferring Auto-Encoder weights to U-Net encoder layers.

Moreover, A critical issue in medical image segmentation is the lack of large and annotated datasets, in addition to per-pixel labeling that is desired instead of the image level label [25].

Therefore, preparing a robust and precise dataset for Laparoscopic Cholecystectomy can participate in the translation of deep learning models to enhance the safety and the efficiency of surgical care, and can be published for future research.

The proposed model combines the effectiveness of Auto-Encoder pre-trained weights as starting weights of U-Net encoder layers to segment the intended landmarks, i.e., cystic duct, cystic artery, Gallbladder, and liver.

1.2 Statement of the Problem

Different incidents occur during minimally invasive surgery mainly due to visual perception errors and overconfidence that lead to anatomy misinterpretation. The problem of stable rates of bile duct injuries (BDI) in laparoscopic cholecystectomy (LC) due to poor implementation and subjective interpretation of the critical view of safety (CVS) is a critical medical issue.

Therefore, deep learning as a subfield of AI can potentially provide an intraoperative model for surgical video analysis and can be trained to guide the surgeons toward reliable hepatocytic anatomy segmentation and produce selective video documentation of this safety step of LC. In the proposed model, Auto-Encoder with U-Net segmentation was deployed to identify hepatocytic anatomical landmarks in laparoscopic cholecystectomy videos.

1.3 Research Questions

- What are the techniques/approaches that could be applied to automatically segment the intended anatomical landmarks accurately?
- Is it possible to utilize the effectiveness of combining Auto-Encoder and U-Net for identifying the hepatocytic anatomy?
- How to efficiently utilize the benefits of Auto-Encoder with U-Net to meet workload requirements, such as execution time, and memory requirements?

- How to address the weaknesses/vulnerabilities of combining Auto-Encoder and U-Net to produce a more efficient model for the intended task?

1.4 Dissertation Organization

The rest of this dissertation is organized as follows: Chapter Two present a literature review of convolutional neural networks (CNNS) and variants in computer vision, medical image segmentation models, and deep learning neural networks for identifying surgical anatomy. Chapter Three presents a brief background in the fundamental knowledge regarding deep learning techniques used in this work. Chapter Four explains the design and the implementation of Auto-Encoder with U-Net segmentation model. Chapter Five presents the experimental work and the results of the proposed model. In Chapter Six, a detailed discussion of the gained results is presented. Chapter Seven proves the efficiency of the model in segmenting different challenging cases during LC procedures. Finally, Chapter Eight presents some drawn conclusions and suggests future work.

CHAPTER 2
LITERATURE REVIEW

2.1 Introduction

Recently more and more researchers have published articles on machine learning applications in laparoscopic cholecystectomy (LC). Since LC is a high-volume surgery, producing a large data set, is used in education and benchmarking, keyframe extraction, and predicting the remaining time of surgery studies.

In benchmarking or skill assessment of surgeons, the performance of surgeons has been improved by analyzing surgery phases, instrument usage and its path length, time of use of each instrument, the number of hand motions, how smoothly movements are, and measuring the applied force [26, 27].

Some studies have focused on keyframe extraction [28], where 81% of the ground truth keyframes were extracted using their trained network. This can be used in education, automatic generation of surgery summary reports, and in building deep learning models for the surgery phase and task recognition. For example, authors in [29] built an automatic phase detection model that segments and classifies the frames into an instrument and non-instrument pixels in cholecystectomy. The Support Vector Machine (SVMs) and Oriented FAST and Rotated BRIEF (ORB) features are used, the instrument ones are categorized, and the cholecystectomy phases are detected based on a predefined set of rules.

Moreover, based on surgery phase information, if accurate calculation of surgery remaining time is possible, more efficient preparations for the next surgery can be processed automatically. In addition to that, optimal planning can be achieved by using the available surgery rooms and medical staff with more patients and less waiting time [30, 31].

Further, medical image segmentation based on artificial intelligence has played an important role in automatic diagnosis and treatment and it is the preference of modern medicine and analysis obtained extensively [32, 33].

2.2 Convolutional Neural Networks (CNNs) and Variants in Computer Vision

CNNs and its variants serve as the primary backbone architectures for computer vision applications, variants with deeper and more effective convolutional neural architectures, e.g., VGG [34], GoogleNet [35], ResNet [36], DenseNet [37], HRNet [38], and EfficientNet [39]. In addition to these architectural advances, there is a strong potential of Transformer-like architectures for unified modeling between vision and language such as self-attention layers and Transformer architectures in the natural language processing (NLP) field. Researchers replace some or all of the spatial convolution layers in the popular ResNet with these self-attention layers [40]. However, while they achieve better accuracy/FLOPs trade-offs than that of the ResNet architecture, it's very expensive in terms of memory usage with more latency than that of the convolutional networks.

On the other hand, other works deployed Vision Transformer (ViT) [41] on non-overlapping medium-sized image patches for image classification. The results are encouraging in image classification, when ViT models are applied to object detection and semantic segmentation by direct up-sampling or deconvolution, they show relatively lower performance [42, 43].

To address these limitations, authors in [44] proposed a hierarchical Transformer, represented by shifted windows between consecutive layers, which limits self-attention computation to non-overlapping local windows and allows cross-window connection at different scales, and has linear computational complexity relative to the image size. These features make it more efficient and compatible with various vision tasks, where in image classification it gives 87.3 top-1 accuracy on ImageNet-1K, while in object detection it gives 58.7 box AP and 51.1 mask AP on COCO testdev, and 53.5 mIoU on ADE20Kval in semantic segmentation.

2.3 Medical Image Segmentation Models

Medical image semantic segmentation provides an important cornerstone in image analysis and understanding. As there are rapid advancements in deep learning methods, conventional U-Net segmentation networks have been applied in many fields [45]. It is built based on a contracting path to capture context and a very symmetrical encoder-decoder structure expanding path that enables precise localization. The model is very fast as it takes

less than a second for the segmentation of a 512 x 512 image on a recent GPU. Consequently, recent works have been developed to benefit from its effectiveness in different applications and approaches, where many researchers have improved the U-Net architecture and proposed various U-Net variants.

Bhakti et al. [46] presented the Eff-UNet hybrid model, where EfficientNets were deployed as an encoder for feature extraction, combined with U-Net decoder for reconstructing the fine-grained segmentation map. The proposed model achieved 0.7376 and 0.6276 mean Intersection over Union (mIoU) on the validation and the test datasets respectively.

Hang et al. [47] proposed a Memory U-Net model that combined U-Net and memory networks to separate and segment individual lesions in spatially connected lesions, where U-Net was used for feature extraction and the memory network as the alternative code book for generalized Hough voting. A density map of lesion locations was obtained by aggregating votes from all lesion voxels, to generate the intended instance segmentation results.

Reza et al. [25] presented a bi-directional Co-nvLSTM (BConvLSTM) U-Net, and densely connected convolutions were built for medical image segmentation. By this combination, the feature maps extracted from the corresponding encoding path and the previous decoding up-convolutional layer in a non-linear way gave good results.

In [48], Spider U-Net was designed for 3D Blood Vessel Segmentation (BVS), where a long short-term memory (LSTM) for the capture of the context of the consecutive data, was deployed into the baseline model. The results justified the consideration of inter-slice connectivity with LSTM to improve the model's performance in the 3D BVS task.

A hybrid model that combined semantic and instance segmentation models in a sequential way was presented in [49]. Here, the U-Net encoder-decoder architecture created a segmentation probability map that was fed as the input to the Mask R-CNN network to produce a final instance segmentation result with a mean Average Precision (mAP) of 0.724 and a Dice coefficient of 0.9284, on the author's proposed dataset.

A two-phase approach for high-accuracy automatic pancreas segmentation in computed tomography (CT) imaging was proposed in [50]. A pancreas localization phase was

built using Mask R-CNN model on the 2D CT slice, and then pancreas segmentation was produced by refining the candidate pancreas region with 3D U-Net on the 2D sub-CT slices coming from the first phase. The qualitative and quantitative results of the proposed approach showed better results than other existing approaches.

A deep learning segmentation method called CMM-Net was developed in [23]. It combined the dilated convolution and pyramid pooling models in U-Net encoder network with an inversion recovery model, that retrieved all the segmented images from the augmented testing data into a single output using “OR” and “AND” logical operators. The proposed approach achieved better performances on three different biomedical imaging tasks compared to the recent deep learning approaches.

H-DenseUNet [51] combined dense connections of 2D and 3D to retain the effective features to the maximum extent. Alom et al. proposed R2U-Net [52] that upgraded the original U-Net structure by combining the residual unit and RCNN. On these grounds, authors in ResUNet-a [53] added residual connections, but also combined pyramid scene parsing pool and multi-task inference in the new network. These additions improve the overall performance of the segmentation task with more processing in terms of memory and speed needed.

UNet + + [54] extended U-Net by integrating multiple U-Nets of different depths into a single network and by redesigning the skip connections between the encoder and the decoder and fusing the features of different semantic scales. The approach delivered an average IoU gain of 3.9 points over U-Net. Afterward, an attention mechanism was added to it to propose ANU-Net [55]. As an extension to UNet++, UNet 3+ [56] added a dense interconnection between the encoder and the decoder, and a dense interconnection among decoder levels instead of the crossed convolution blocks at the encoder-decoder path at UNet++.

Authors in [22] proposed a scale-attention deep learning network (SA-Net), where different scales of features were extracted in a residual model and an attention model was deployed to carry out the scale-attention capability. Experiment results showed good performance results in retinal vessel detection, lung segmentation, artery/vein classification, and blastocyst segmentation tasks.

Authors in [57] proposed a model called Segmentation Auto-Encoder (SAE), which leverages all available unlabeled scans and requires segmentation prior as a single unpaired segmentation image. From experiments that applied SAE to brain MRI scans, results show good quality segmentations, particularly when the prior is good.

2.4 Deep Learning Neural Networks for Identifying Surgical Anatomy

A recent and interesting approach in laparoscopic surgery is the ability to identify surgical anatomy using deep learning neural networks. Namazi et al. [58, 59] proposed a deep learning model trained to identify the CVS directly using the cholec80 dataset, where the extracted frames (20 frames per second) from LC videos are manually labeled as either 1 (CVS) or 0 (no CVS). In this research, authors addressed the problem of high imbalance in the training set by over-sampling and under-sampling. The accuracy gained was 94.6%, the precision of 70%, and 65.5% for recall.

Authors in [60] focused on segment IV of the liver and its diagonal line (D-line) as a clear landmark for carrying out difficult LC. The dataset compromise sixty-two consecutive difficult LC were managed by the segment IV approach. Gallbladder was extracted along the D-line was achieved in 44 (71%) cases.

Altieri et al. [61] developed an artificial intelligence (AI) model that can identify anatomical landmarks and safe and dangerous zones of dISSection using 264 extracted frames from 63 videos annotated by two expert surgeons using a freehand annotator and reviewed by a hepatobiliary surgeon. The results were 95% (±2), 63% (±20), 98% (±1), 78% (±9), and 96% (±2) for mean accuracy, sensitivity, specificity, Positive Predictive Value (PPV), and Negative Predictive Value (NPV), respectively, for identifying safe zones. These metrics were 97% (±2), 84% (±13), 99% (±1), 93% (±6) and 97% (±2), respectively for dangerous zones.

A study was built in [62] to detect bile leakage in laparoscopic cholecystectomy video frames based on a deep learning model. They used 62380, both bile and no bile leakage images, that were extracted from the cholec80 dataset and surgeries performed in the Meander Medical Centre. These images were used to train and evaluate two convolution neural networks with different parameter settings to achieve an optimal bile leakage detection algorithm. The trained

model delivered 83% sensitivity, 80% specificity, and 0.91 of an AUC score for the testing dataset.

Tokuyasu et al. [63] proposed a landmark indication system based on a deep learning object detection algorithm i.e., YOLOv3, where 2000 endoscopic images of the region of Calot's triangle in the gallbladder neck were extracted from 76 LC videos. These images were used to train and evaluate the model to identify four landmarks: the cystic duct, common bile duct, lower edge of the left medial liver segment, and Rouviere's sulcus. The average precision values for each landmark were as follows 0.320, 0.074, 0.314, and 0.101, respectively. In addition, a prototype was constructed and used in the operation for a patient with cholelithiasis as a verification experiment.

Scheikl et al. [64] compares different neural networks, loss functions, and training strategies that deployed as a semantic segmentation model on different organs and tissue types in laparoscopic images. TernausNet-11 was trained with Soft-Jaccard loss using a pretrained, trainable encoder. The model gains a segmentation quality of (78.31% mean Intersection over Union [IoU]) and inference time (28.07ms) on a single GTX 1070 GPU.

Other studies targeted hepatocytic anatomy segmentation to define the critical view of safety during cholecystectomy before ligation [16, 17, 65, and 66]. For example, Pietro Mascagni started his work by formalizing a reproducible method for objective video reporting of CVS in 100 LC procedures performed at the Digestive Surgery Department of the Nouvel Hôpital Civil (Strasbourg, France). The video segments, i.e., which start 60s before the division of cystic structures, were reviewed by two independent surgeons. The assessment of CVS was based on an adaptation of the doublet view 6-point scale and a new binary method that checked if the criterion either was achieved or not [17]. Then he developed a deep learning model to automatically segment hepatocystic anatomy and a classification model to predict CVS criteria achievement using 5-fold cross-validation. 2854 images were extracted from 201 LC videos for training and testing. The results showed 66.6% for mean intersection over union for segmentation, and 71.9%, 71.4% for mean average precision and balanced accuracy, respectively [65]. Recently, Mascagni in [16] proposed a protocol, checklists, and visual examples to set a guideline to annotate surgical images, to get a consistent hepatocystic anatomy segmentation and CVS criteria assessment.

Madani et al. [20] developed a deep learning model to identify a safe zone that covers the hepatocytic triangle, and an unsafe zone that covers the hepatic duodenal ligament, liver hilum, and inferior structures, deep to the hepatocytic triangle. In this work, senior surgeons annotated extracted frames from 290 videos to identify the location of the two zones. The proposed model was able to recognize the safe and unsafe zones with an accuracy of 0.94 and 0.95, respectively. In addition to that, a dynamic overlay of the zones was carried out on the videos.

Researchers in [66] developed a deep learning model consisting of a Convolutional Neural Networks (CNN) and a Recurrent Neural Network (RNN) which were trained and validated on 120 LC videos combining the cholec80 and EndoVis workflow challenge datasets to detect the presence of the three CVS criteria. That is (1) seeing only two structures entering the gallbladder, (2) exposing the inferior 1/3 of the cystic plate, and (3) clearing the hepatocytic triangle of fat and fibrous tissue. Results achieved were 0.98 for the accuracy, 0.81 for F1-score, 0.96 and 0.70 for precision and sensitivity, respectively.

As it can be concluded from these recent studies, deep learning algorithms can be trained to automatically segment hepatocytic anatomy and assess CVS criteria in still laparoscopic images and videos reliably in LC videos. However, only limited number of works in this field can prove the potential of such methods in the assessment of the safety of a cholecystectomy procedure. There is a room to get better results, which could accelerate the translation of deep learning models to enhance the safety and efficiency of cholecystectomy.

CHAPTER 3

BACKGROUND

The widespread application of deep learning, which leverages multiple layers of neural networks for extracting multiple levels of representations through a variety of methods, is to be used for detection or classification tasks. Recently, these methods have promoted the state-of-the-art in most challenging problems in different domains of computer vision such as data dimensionality reduction, handwritten number recognition, pattern recognition, image recognition, image repair, image segmentation, object tracking, scene analysis, etc., showing very high effectiveness [67].

This chapter provides fundamental knowledge regarding deep learning techniques used in this proposal. Such as Artificial Neural Networks (ANNs) and Convolutional Neural Networks (CNNs), followed by deep learning applications: Image Segmentation, and ultimately the state-of-the-art models that deployed in this work.

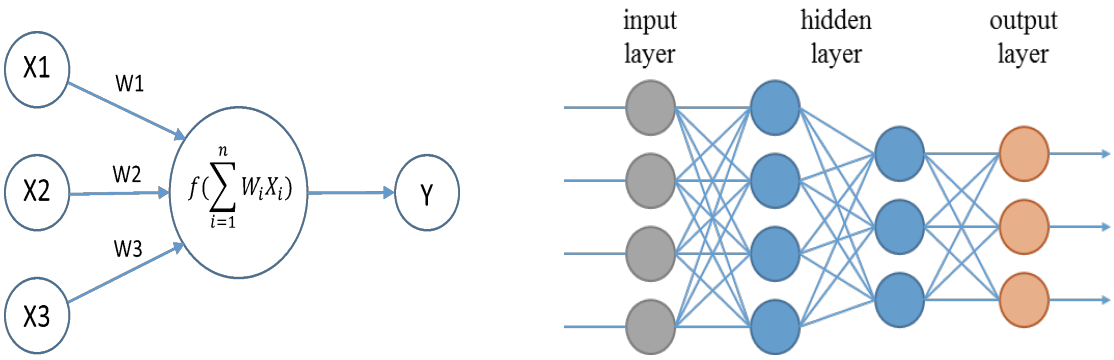
3.1 Deep Learning Components

3.1.1 Artificial Neural Networks

Artificial Neural Networks (ANNs) are computing systems that simulate the neural system in the human brain, where each ANN combines multiple neurons (units) that process their inputs. Figure 3.1(a) shows an example of an artificial neuron, where *f* is the activation function, and the output of a neuron is the weighted sum of the inputs plus a bias, that will be activated by *f*. In Figure 3.1(b) an ANN is presented, in which the layers following the input layer are called hidden layers and the depth of the network is identified by its number. In general, the output for an *L*-hidden-layer ANN is calculated as in equation 3.1[68]:

$$F_W(x) = W_0^T f\left[W_L^T f\left(W_{L-1}^T \cdots f(W_1^T)\right)\right] \quad \textit{Equation 0.1}$$

where *ml* is the number of units in the *l*th layer, *Wl = (wl,0, ...wl, ml)* is the weight matrix for layer $0 < l \leq$ L, and *Wo* is the output weight matrix. The summation of all weights *W* is the total weights of an ANN.



(a) A single neuron. (b) A deep fully connected neural network with two hidden layers.

Figure 3.1: Artificial Neural Networks.

The activation function *f* is used to add non-linearity to the weighted output that can constrain the output and form a general mapping function. The most commonly used activation functions are: sigmoid, hyperbolic tangent (tanh), and Rectified Linear unit (RELU), as shown in Figure 3.2.

3.1.1.1 Optimization

Supervised learning has proven to be the most promising algorithm, where its aim is to find the best mapping between the given inputs to their corresponding outputs, with the availability of ground truth. In ANNs, the optimization process finds the best weights *W* that minimizes experimental cost (loss) function through training. Given *n* training samples $(x1, y1)\ldots(xn,\ yn)$, where *yi* is the ground-truth for the *i*th sample *xi*, the loss function minimization formula is as [69]:

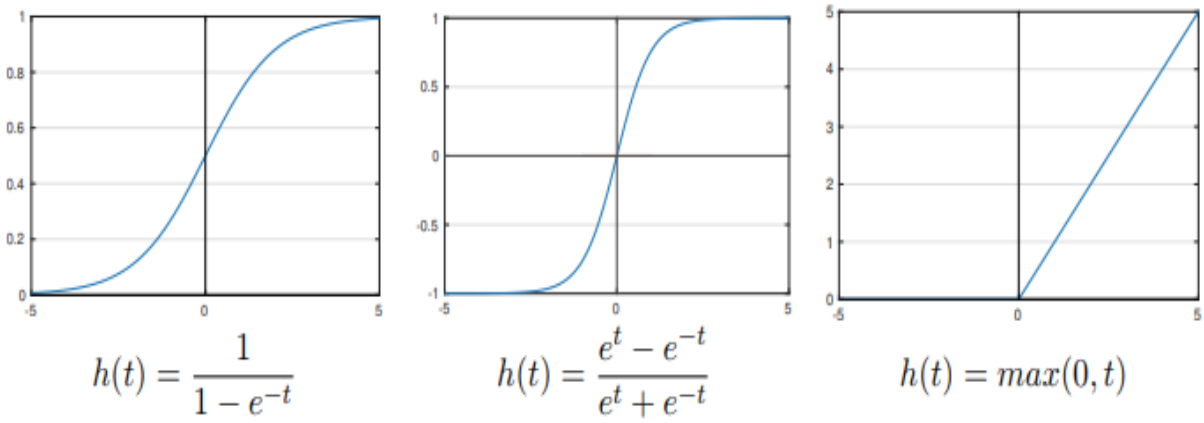


Figure 3.2: Common activation functions: sigmoid (left), tanh (middle), RELU (right).

$$min_W \mathcal{L}(W) = \frac{1}{n}\sum_{i=1}^{n} \ell\big(y_i, F_W(x_i)\big)$$ *Equation 0.2*

Gradient Descent (GD) method is used to minimize the loss function, by computing the gradient and going through steps analogous to gradient negative values at each iteration through training. Given the initial weights matrices $W(0)$, the update process at step k for each layer l can be written as [68]:

$$W_l^{(k)} = W_l^{(k-1)} - \eta \nabla_{W_1} \mathcal{L}\Big(W_l^{(k-1)}\Big),\ l = 1, \dots, L$$ *Equation 0.3*

Where $\eta > 0$ is the step size i.e. (learning rate) and is the gradient operation. In GD-based algorithms, the gradients are calculated based on the backpropagation method [70], in which the gradients for each layer l are computed using the chain rule, starting recursively from the last layer until converging to a local or a global minimum. The cost function as in equation 3.3 need to be computed for all of the samples in the training set, which might be expensive in terms of memory usage, especially when the training set size is large. While in the Stochastic Gradient Descent (SGD) method, the gradients are computed for each sample or a mini batch of size nb that is selected randomly from the total batch n, which makes this method faster and more efficient than GD for larger datasets. On the other hand, SGD algorithm has slower convergence coming from the noisy nature of the gradients as the batch size is small. To solve this problem, the exponential weighted moving average of the update term of the current and previous step is taken [71], where the hyper-parameter is the momentum, as in the following formula:

$$\begin{aligned} W^{(k)} &= W^{(k-1)} - \eta \Delta W^{(k)} \\ \Delta W^{(k)} &= \beta \Delta W^{(k-1)} + (1-\beta)\nabla_W \mathcal{L}\big(W^{(k)}\big) \end{aligned}$$ *Equation 0.4*

Several approaches have emerged to improve the convergence of SGD, based on adaptive learning rate and momentum, such as AdaGrad [72], AdaDelta [73], ADAM [74], NADAM [75], etc. The formulas for ADAM update rules are as follows:

$$W^{(k)} = W^{(k-1)} - \frac{\eta}{\sqrt{\hat{V}(k-1)} - \epsilon}\hat{M}^{(k-1)}$$ *Equation 0.5*

Where $\hat{V}$ and $\hat{M}$ are calculated as:

$$\begin{aligned} \hat{M}^{(k-1)} &= \frac{M^{(k-1)}}{1-\beta_1} \\ \hat{V}^{(k-1)} &= \frac{V^{(k-1)}}{1-\beta_2} \end{aligned}$$ *Equation 0.6*

The hyper-parameters β_1 and β_2 are selected to be close to 1 when calculating the mean and the variance estimation as follows:

$$M^{(k-1)} = \beta_1 M^{(k-2)} + (1-\beta_1)\nabla_w \mathcal{L}\left(W^{(k-1)}\right)$$
$$V^{(k-1)} = \beta_2 V^{(k-2)} + (1-\beta_2)\left[\nabla_w \mathcal{L}\left(W^{(k-1)}\right)\right]^2 \quad \textit{Equation 0.7}$$

3.1.1.2 Regularization

Deep learning models are vulnerable to overfitting when the model does well on the training set but can't generalize for the invisible samples. Therefore, different regularization techniques are proposed to reduce the generalization error by reducing the capacity of the network or the optimization process, as in equation 3.8 [68]:

$$M^{(k-1)} = \beta_1 M^{(k-2)} + (1-\beta_1)\nabla_w \mathcal{L}\left(W^{(k-1)}\right)$$
$$V^{(k-1)} = \beta_2 V^{(k-2)} + (1-\beta_2)\left[\nabla_w \mathcal{L}\left(W^{(k-1)}\right)\right]^2 \quad \textit{Equation 0.8}$$

Other regularization techniques use data augmentation, where random transformations and noise are applied to the input to get more diversity and size of the training set.

3.1.1.3 Normalization

In deep neural networks, normalization techniques are used to minimize the layers' input range or the activation maps. Batch normalization [76] is the simplest form of normalization method, where normalization processes each mini batch in the activation maps. It starts by calculating the mean and the variance of that feature, then subtracting the mean from the features and dividing the result by the standard deviation of the mini batch. However, this technique has some limitations [77]. To that, different normalization approaches are proposed, i.e. weight normalization [78], layer normalization [79], instance normalization [80], group normalization [81], etc. Normalization methods have proven to be a primary element in improving optimizer stability and convergence [82].

3.1.2 Convolutional Neural Network Architectures

A Convolutional Neural Network (CNN) is a classic artificial neural network structure that has gradually emerged in recent years, based on the combination of deep learning and

image-processing technology. Four basic building blocks create such a network, which are: convolutional layers, max-pooling layer, flatten, and fully connected layers as in Figure 3.3.

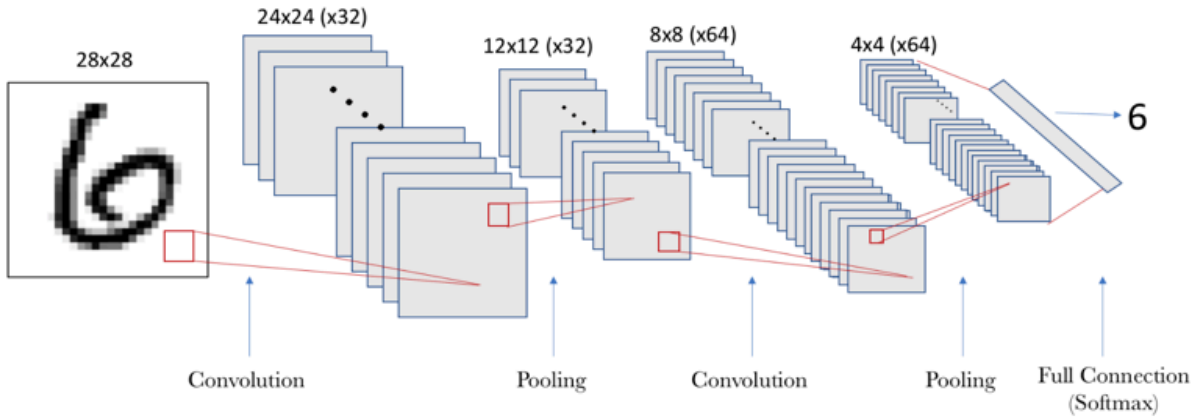


Figure 3.3: A convolutional neural network [83].

Thus, in CNN, each neuron is connected to a small number of neurons from the previous layer, which can reduce many parameters. Weight sharing is another technique used to reduce the parameters, where a set of connections share the same weights.

Additionally, downsampling is used to simplify the network by removing the odd columns and even rows, and this will reduce the number of samples per layer, further reducing the number of parameters, and improving the robustness of the model. To avoid the shrinking of the image size during the convolution, phase paddings are used, wherein each convolution layer, kernel size, kernel stride, and padding are set.

Moreover, the pooling layer usually is used to downsample the features spatially and can filter useful information and analyze them in the next convolution layer. Both the convolution layers and the pooling layers repeat many times to extract more advanced features. Further, the three-dimensional feature is flattened to a one-dimensional feature that will be processed by a fully connected network to classify these features into a particular class. As in color images, each pixel of these images has three color channels (red, green, blue).

Recently, rapid advances in hardware coupled with parallel processing, as in Graphical Processing Units (GPUs) had a significant impact on deep learning development. In

addition to the availability of online datasets and competitions such as Imagenet (Large Scale Vision Recognition Challenge) [84] which encourages researchers to deploy their ideas using large manually annotated datasets, they made many achievements including image feature extraction and classification, pattern recognition, etc.

The Alexnet [85] architecture was one of the biggest breakthroughs in pattern recognition and machine learning domain, as it was the most successful at image classification of the training set of ImageNet, making convolutional neural networks become the key research object in computer vision, and this research continues to deepen. The main architecture consists of five convolutional layers, followed by three fully connected layers as in Figure 3.4.

Even after, different architectures emerged as the VGGnet [34] that adds more convolutional layers for better training techniques. One more evolution stepped out in 2015, with skip (residual) connections in the Resnet architecture [36], these connections provide a shortcut path for the gradients to pass back to the beginning. This approach limits the vanishing gradient phenomenon, which allows deeper architectures of up to 1000 layers to be possible with residual connections.

3.2 Image Segmentation

Image segmentation is a very important and difficult branch in computer vision, aiming at dividing the image into different disjointed meaningful regions [87] according to features such as color, grayscale, spatial texture, and geometric shapes. Based on these features consistency or similarity can be recognized. There are three major image segmentation categories, semantic segmentation, instance segmentation, and panoramic segmentation. Currently, more and more research branches of image segmentation, i.e., medical image segmentation, satellite image segmentation, autonomous driving [88], etc.

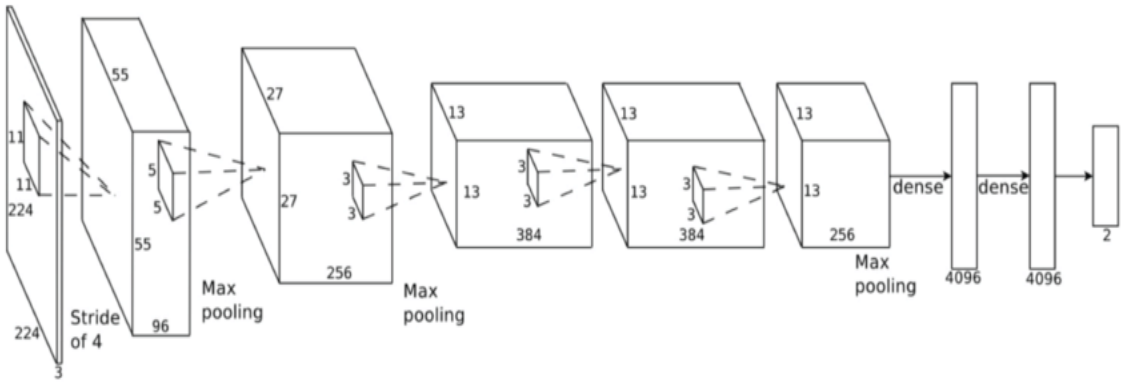


Figure 3.4: The Alexnet architecture [86].

Deep learning segmentation methods can no longer be compared with traditional image segmentation methods, such as the threshold-based segmentation method [89], edge detection-based segmentation method [90], and region-based image segmentation method [91], in which digital image processing is combined with mathematics to segment an image. In these methods, the calculation is simple and the segmentation speed is fast, but details segmentations accuracy cannot be guaranteed. While deep learning segmentation methods have made a remarkable breakthrough, with high accuracy compared to traditional segmentation methods.

A fully convolutional network is the first deep learning network that successfully did image semantic segmentation. Then other outstanding segmentation networks were proposed, such as U-Net [24], Mask R-CNN [44], RefineNet [92], and DeconvNet [93] that segment fine edges successfully. However, for different segmentation applications, there is no universal segmentation model that is suitable for all images.

According to segmentation coarse and fine granularity properties. Semantic segmentation is the primary task in computer vision, where the visual input is divided into different semantically interpretable categories i.e., the classification categories are meaningful in the real world. For example, if we need to identify all pixels in an image that belongs to a bus and paint them yellow.

However, clustering as an unsupervised model can be used for segmentation, where the classes trained for detecting regional boundaries can't be subdivided, so the results are not necessarily semantic. While in semantic segmentation more detailed understanding of images is performed than image classification or target detection, which makes it an attractive platform in many areas, such as autonomous driving, robotics, and medical images.

Further, the instance segmentation approach performs objects detection and contours drawing on all the categories in an image, but as the number of instances is not known in advance, this makes the task more difficult than the semantic segmentation task, and most instance segmentation architectures are not end-to-end and extremely complex compared to regional-based convolutional neural networks [94].

3.2.1 Medical Image Segmentation

Medical image segmentation is the process of recognizing the location and contours of organs through medical images that can help doctors in the automatic diagnosis and treatment and follow-up of many diseases. Even though the automation of medical image segmentation has been extensively studied in the past, manual annotations practically are still used in clinical practice, which is expensive in terms of time, and it's a subjective process. Thus, there is a high demand for accurate and reliable automatic segmentation methods that can improve workflow proficiency in clinical scenarios and mitigate radiologists and other medical experts' workloads [95].

The human body has multiple organs and tissues, and each part has its specificities, for example, diagnosing brain tumors and lung nodules needs a relatively large segmentation area, while in retinal blood images, blood vessels are segmented with high accuracy. Based on these peculiarities, researchers design more efficient and accurate segmentation algorithms for different organs [95].

Medical image segmentation is a challenging task: due to many issues in medical data. Authors in [96] summarize these challenges in eight categories, i.e. extrinsic variability, intrinsic variability, spatial complexity, moving or deforming targets, extremely small targets, and similar adjacent structures. Out of these challenges, medical image data limitations, and the difficulty in collecting data and labeling in many other segmentation tasks, as it is time-consuming and labor-intensive. However, the common challenges for all deep learning applications are high computation costs and a lack of interpretability [96].

3.3 The State-of-the-Art Models

In this section, the state-of-the-art models that are utilized in this proposed work will be introduced. The proposed approach starts with Auto-Encoder, followed by U-Net for segmenting hepatocytic anatomical landmarks.

3.3.1 Auto-Encoder

An Auto-Encoder is an unsupervised artificial neural network that is built from two components: an encoder and a decoder. The encoder part learns how to efficiently compress and encode data, followed by the decoder which learns how to reconstruct the data back from the reduced encoded representation to a representation as close to the original input as possible. The Auto-Encoder design reduces the data dimensions by learning how to ignore data noise [97].

The network structure of Auto-Encoders can vary between a simple Feed Forward network, LSTM network, or Convolutional Neural Network depending on the application. Figure 3.5 shows the general structure of Auto-Encoder, which consists of an input layer and an output layer that are connected through one or more hidden layers. Using this architecture, features can be extracted by reducing the dimension of its hidden layer [125], which allows the autoencoder to focus on capturing the main features that best represent the data.

As one can see, this network reconstructs the input by transforming inputs into outputs as closely as possible without simply distorting the input. This type of neural network is used mainly in unsupervised learning problems as well as transfer learning [97]. In addition to that, Autoencoders can be used for image denoising, image data generation, and image compression.

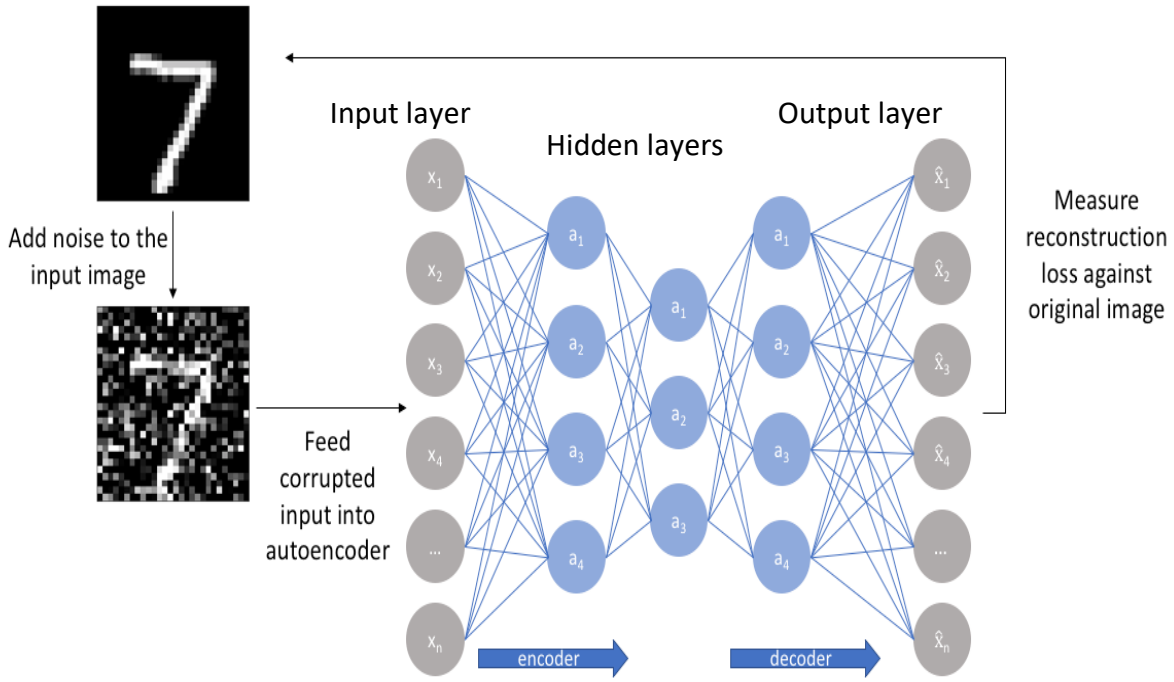


Figure 3.5: The structure of an Auto-Encoder network.

3.3.2 U-Net

The result obtained from FCN segmentation is not fine enough, i.e., relatively blurry and smooth, and not sensitive to image details. Ronneberger et al. [24] proposed U-Net, which has a similar architecture to FCN except that U-Net uses the concatenation of the feature map in the convolution layer and transpose convolution layer and can be trained effectively with a small dataset i.e., (around 30 images)

The entire U-Net network architecture is similar to the uppercase U letter, with a very symmetrical encoder-decoder structure, where the encoder and the decoder are on the left and right sides of the network, respectively, that has a basic convolutional layer and activation layer ReLU on both sides. The encoder has a 2 2 max-pooling layer followed by a convolutional layer to increase the robustness to disturbance. While in the decoder, upsampling restores the features extracted by the convolutional layer from the low-resolution feature map. This scenario is repeated until it is decoded to resolve the input. Using skip connections between the encoder and the decoder can reduce the semantic information loss caused by downsampling, the full architecture is shown in Figure 3.6 [24].

The novelty of U-Net is in the expansive path that at each stage the feature map is upsampled using 2x2 up-convolution. Followed by the feature map from the corresponding layer in the contracting path that is cropped and concatenated onto the upsampled feature map. Then two successive 3x3 convolutions and ReLU activation are placed. Finally, an additional 1x1 convolution is applied that can limit the feature map to the required number of channels and build the segmented image. Cropping is necessary as pixel features in the edges have the least amount of contextual information and can be discarded.

As the U-Net network does not contain fully connected layers, the resulting segmentation map contains only pixels of the input image that allows large images to be segmented by an overlap-tile strategy to predict the pixels in the image border region. This approach can be utilized to be applied to large images and can save GPU memory [24].

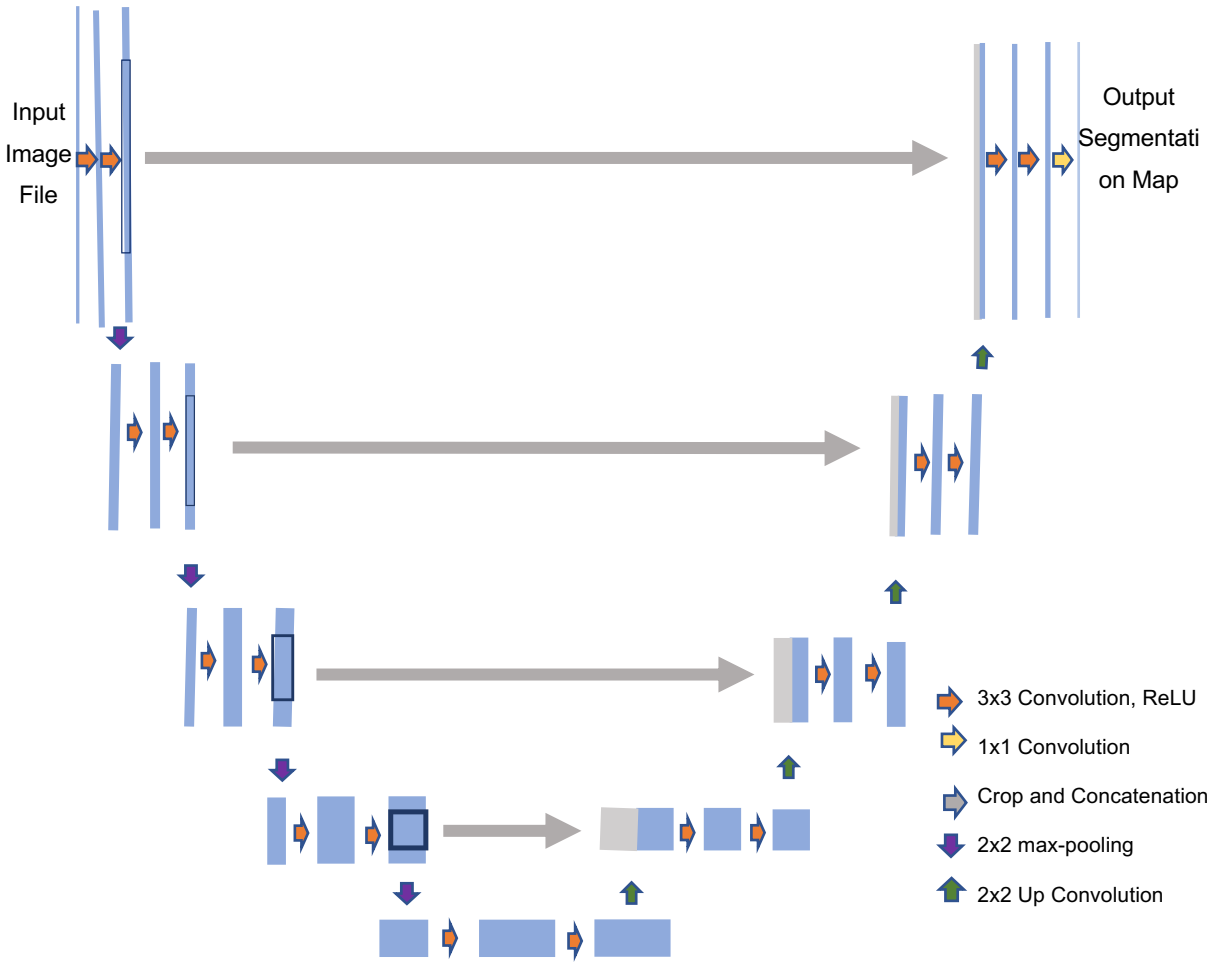


Figure 3.6: U-net architecture (with 32x32 pixels in the lowest resolution as an example). Each blue box relates to a multi-channel feature map. The gray boxes represent copied feature maps, and the arrows denote different operations [24].

The energy function for the entire network is as in following equation:

$$E = w(x)log\sum\left(p_{k(x)}(x)\right) \quad \textit{Equation. 3.9}$$

Where p_k is the pixel-wise SoftMax function applied after the final feature map.

$$p_k(x) = \frac{e^{ak(x)}}{\sum_k^E = a^{a\alpha(x)r}} \quad \textit{Equation. 3.10}$$

And a$k(x)$ denotes the activation in channel k.

3.3.3 TensorFlow

TensorFlow is an open-source software library for numerical calculations based mainly on data flow graphs [98] that were developed by researchers and engineers at the Google Brain Team. Its main objective is to accelerate deep neural networks and machine intelligence research. It is very useful for fast-performing graphics-based calculations. As TensorFlow API is flexible, it can deploy models between multiple devices through its GPU-enabled architecture, and now it has a large active community with rich low-level and high-level API. However, the main drawback of TensorFlow is the need to redefine all the APIs as it is not Python-friendly, which requires developers to spend more time learning.

Since the computation is based on the static data flow graph, creating a data flow graph is needed first, and then feeding data into the data flow graph, where nodes in the diagram represent the mathematical operations and can represent the data flow end or start. The multidimensional data arrays are represented by the edge that connects two nodes which are called a tensor. During training, the tensor flows from one node to another node in the data flow graph continuously. Once all the prepared tensors are ready, nodes will be assigned to different computing devices to work asynchronously in parallel.

CHAPTER 4

METHODOLOGY

In this chapter, the effectiveness of using Auto-Encoder with the U-Net segmentation model will be investigated, in particular, to automatically highlight cystic duct and cystic artery anatomy and to predict CVS criteria achievement in laparoscopic cholecystectomy video.

Figure 4.1 illustrates the main steps and methods the proposed work goes through ~~it~~ to identify the intended landmarks of laparoscopic cholecystectomy. In the first step, datasets are collected and processed. It is followed by building the Auto-Encoder to prepare the pretrained weights for the next step, where U-Net will be trained using the prepared dataset and the pretrained weights. Finally, the model will be used to predict landmarks in a laparoscopic cholecystectomy video, and the evaluation criteria will be calculated.

4.1 Data Collection

The primary source of data that is used for hepatocytic anatomy segmentation and CVS assessment is the LC procedures endoscopic videos. There are two extensive datasets available i.e., Cholec80 [99] and EndoVis [100], that contain LC videos with annotation of laparoscopic instruments and surgery phases [101, 102].

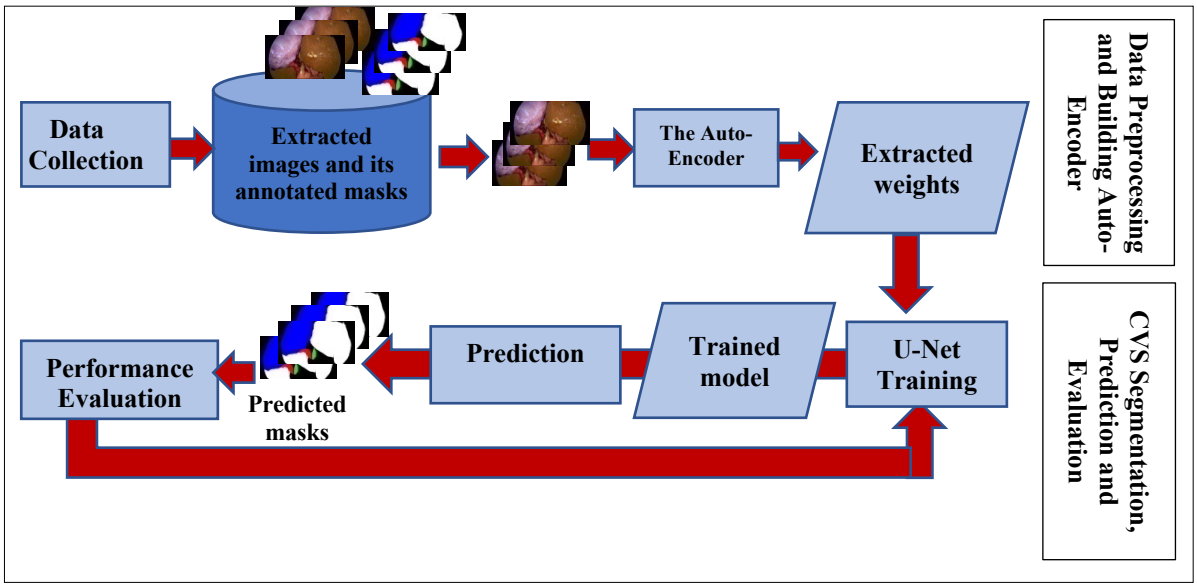


Figure 4.1: Schematic diagram of the proposed model main steps and methods.

Image labeling or image annotation is the process of identifying or recognizing different objects in an image; to make images readable for computer vision. There are different types of image annotations. Some of the most common types of image annotation for computer vision are bounding boxes, polygonal segmentation, line annotation, spatial annotation, landmark annotation, 3D cuboids, semantic segmentation, etc.

In the case of the CVS anatomical landmarks identification, semantic segmentation may be more appropriate than bounding boxes, where bounding boxes may produce incorrect or overlapping annotation of structures, and some organs may have fixed ‘starts’ and ‘ends’ (i.e., the edge of the liver). Other organs may be more flexible, such as a cystic artery, where some connective tissue surrounding them makes the labeling more difficult as the border between the artery and the gallbladder is somehow ambiguous [103].

4.2 Building the Auto-Encoder

The main idea in building an Auto-Encoder is that the encoder block has the same structure for both the Auto-Encoder and the U-Net models as described in the standard U-Net [24]. While the decoder block is built without skip connections, the hepatocytic anatomy features were extracted from the input images and represented in a low dimensional space. The weights gained from training the Auto-Encoder on a part of the prepared dataset were used as starting weights for the U-Net encoder block in the model, as illustrated in Figure 1.2. Training the U-Net with initial encoder pretrained weights would dramatically speed up the training process.

4.3 CVS Segmentation, Evaluation, and Prediction

The flowchart shown in Figure 4.3 illustrates all the steps needed for building the segmentation model. Since the extracted images and the corresponding masks have different dimensions, each image and mask were resized to 256 x 256.

As there is a high correlation between the neighboring frames in surgical videos, under-sampling/over-sampling was used to balance the classes and avoid over-fitting. Based on the number of samples for each targeted object, over-sampling/up-sampling can be applied

to the minority classes and under-sampling/down-sampling for the majority classes to get a uniform distribution of the combination of the labels.

Furthermore, the augmentation process of the training datasets is recommended to improve the deep learning performance by teaching it the desired invariance and robustness properties [104], as the appearance of the abdominal organs differs between patients, and there are individual differences in the skill of each endoscopic operator. Keras [105] "ImageDataGenerator class" was used to augment the training datasets. Only horizontal, vertical, and reflective, augmentations were used. Additionally, each mask will be converted to categorical using Keras utilities.

Recently, CNNs deliver the best, state-of-the-art performance in different visual recognition tasks, because it is powerful and has nonlinear feature extraction capabilities. In medical image segmentation, FCN was one of the first deep networks applied to image segmentation, followed by U-Net, which achieved good segmentation results leveraging the need for a large amount of training data. Despite these useful results, the processing step in these networks is performed separately for the two sets of feature maps that will be then simply concatenated [95].

U-Net particularly creates highly detailed segmentation maps using very small samples. That is very important especially in the medical imaging community as the available labeled images are often limited. This property is achieved by using random elastic deformation that lets the network learn these variations without the need for new labeled data [24]. In addition, U-Net utilizes a weighted loss function to separate touching objects of the same class, which penalizes the model if it fails in separating two objects. Finally, U-Net can be considered the much faster segmentation model to train than most other segmentation models based on its context-based learning.

Due to the above settings, U-Net has become the preferred choice when segmenting medical images, i.e., where the output result can locate the target category, and the input training data are patches, that are similar to data augmentation and help when the number of biomedical images is small.

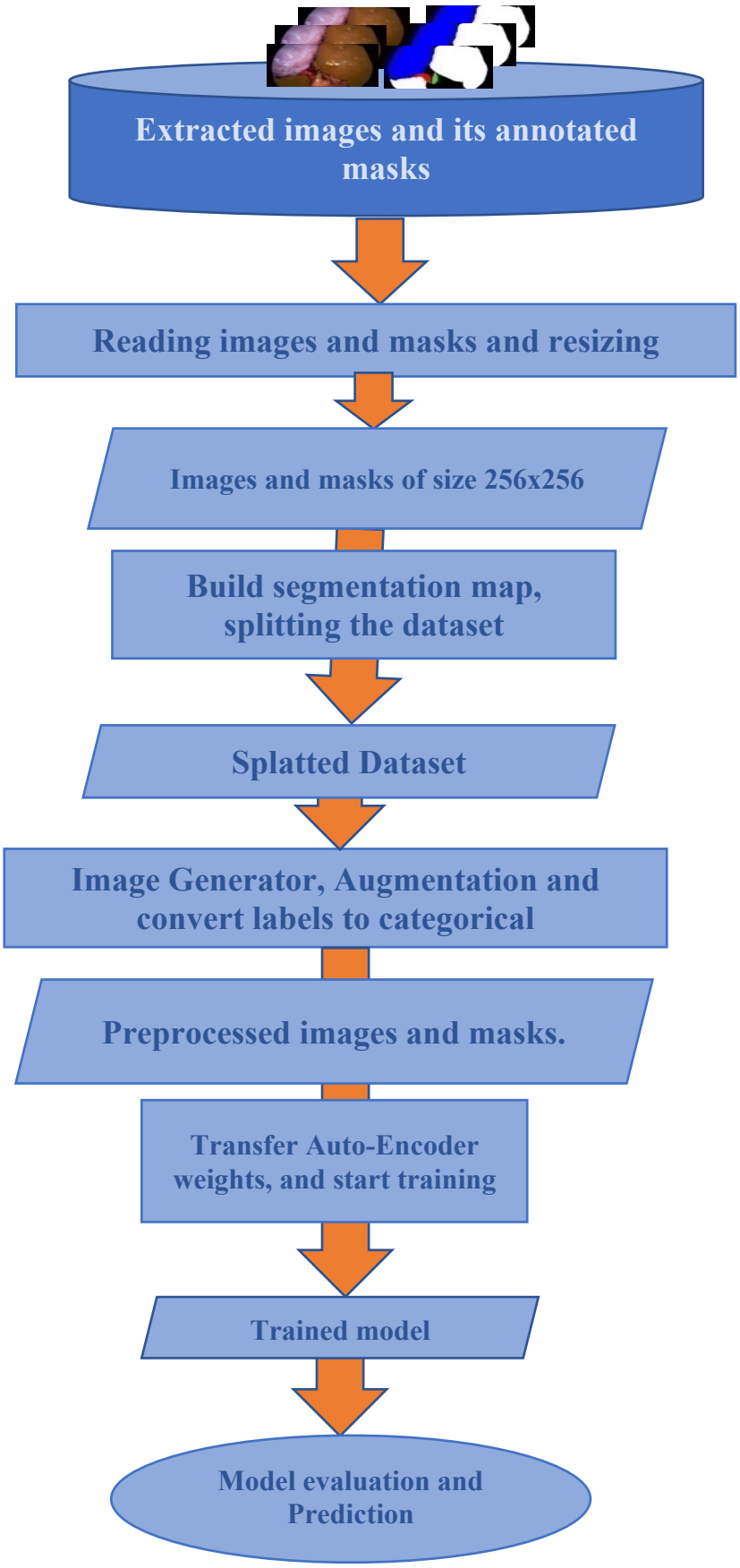


Figure 4.2: Segmentation, evaluation, and prediction flowchart.

Typically, U-Net is trained from scratch starting with randomly initialized weights. It is well known that to avoid over-fitting through training, the data set needs to be large, with millions of images. Training networks on the Imagenet [120] data set are widely used as initialization weights.

In the self-supervised learning [124] approach, a pre-training step with an unsupervised strategy is deployed for learning useful and better representations of the data samples. This strategy is an emerging technology that is effectively closing the gap with fully supervised methods on large computer vision benchmarks [126].

To do that, U-Net [24] was deployed as the segmentation network in the proposed model, in which the gained weights from Auto-Encoder training were used as the U-Net's encoder starting weights.

Thus, a hybrid loss function [88] was used to calculate the loss of the output results from different levels making the loss function more balanced, it combined three loss functions: dice coefficient, cross-entropy, and focal loss, as illustrated in Equation 4.1. The dice coefficient is a metric that is widely used in computer vision models to calculate the similarity between two images. Cross-entropy is a measure of the difference between two probability distributions for a random variable, and it works well for classification and segmentation models. Focal loss is a variation of Binary Cross-Entropy, where it focuses more on learning hard examples and works well for highly imbalanced class scenarios [106].

$$Hybrid\ Loss = \sum_{i=1}^{4} \left(1 - \left[\frac{b \times \bar{Y} \times log\ Y_i}{/Y_i - 0.5/^r} + \frac{2 \times Y_i \times \bar{Y} + s}{Y_i^2 + \bar{Y}^2 + s}\right]\right) \quad \textit{Equation 0.1}$$

where $\bar{Y}$ stands for the exact segmentation result gained from manual labeling, Y_i is the segmentation prediction produced in block X_{0_i} in the network, *s* is the smooth factor, *b* is the balance factor in Focal loss, and *r* is the hyperparameter used for difficult samples.

4.4 Optimization Techniques and Performance Evaluation

During training, the hybrid loss function (Equation 4.1) was used for backpropagation and network parameters update. The Adam optimizer was used with a learning rate of 0.01.

In addition, to obtain more reliable results, 5-fold cross-validation was ~~deployed~~ employed on the dataset, where it's a method for estimating the performance of a model on

unseen data. It is a resampling technique without replacement. The dataset is organized into 5 folds, making sure the statistical properties were preserved across folds, and the validation set will be used to pick the model hyperparameters and to determine when to stop training. This technique helps to avoid overfitting, which can occur when a model is trained using all of the data, and the model will be tested on k different data sets, which helps to ensure that the model is generalizable [107].

In medical image segmentation tasks, the evaluation metrics are divided into pixel-based and overlap-based methods [88]. The following popular indicators are used to evaluate the segmentation quality of the proposed model: Intersection over Union (IoU), precision, recall, sensitivity, specificity, and the Hausdorff distance. IoU measures the percentage of overlap between the segmentation output and the ground truth annotation. The better the segmentation, the larger the overlapping area and IoU is closer to 1. Precision measures how accurate is the trained system predictions. i.e., the percentage of predictions' correctness, while recall measures how well the system finds all the positives, for example, finding 80% of the possible positive cases in the top predictions. Sensitivity and specificity are the two widely used statistical measures for binary classification test performance in the medical field, where sensitivity measures the percentage of actual positive values that are identified, whereas specificity measures the percentage of negative values that are correctly identified.

The Hausdorff distance is added to evaluate the segmentation on boundaries, where it measures the degree of similarity between two sets of points, i.e., the distance between the two boundaries of ground truth and the segmentation result input to the network [88]. The mathematical expressions of these metrics are as follows:

$$IoU = \frac{TP}{TP+FP+FN} \quad \textit{Equation 0.2}$$

$$Precision = \frac{TP}{TP+FP} \quad \textit{Equation 0.3}$$

$$Recall = \frac{TP}{TP+FN} \quad \textit{Equation 0.4}$$

$$Specificity = \frac{TN}{TN+FP} \quad \textit{Equation 0.5}$$

$$Sensitivity = \frac{TP}{TP+FN} \quad \textit{Equation 0.6}$$

$$\textit{Hausdorff Distance}\ (Y, \bar{Y}) = m\ \{d_{Y\bar{Y}}, d_{\bar{Y}Y}\} = m\left\{\max_{\substack{y_1 \in Y \\ Y_2 \epsilon Y}} (y_1, y_2), \max_{y_2 \epsilon Y, y_1 \in Y} d(y_1, y_2)\right\} \quad \textit{Equation 0.7}$$

where TP stands for the True Positive, FP is the False Positive, FN is the False Negative, and TN is the True Negative. $\overline{\Upsilon}$ is the segmentation result, and Υ is the annotated result. The $d()$ means the Manhattan distance calculated between Υ_1 and Υ_2, and ϵ denotes the boundary of the object.

CHAPTER 5
EXPERIMENTAL WORK AND RESULTS

Laparoscopic Cholecystectomy (LC) is vastly accepted worldwide, and it is currently the standard procedure for cholelithiasis and/or cholecystitis [63]. The reported incidence of bile duct injury (BDI) during LC is much higher by two to five times than during abdominal surgery [108].

Table 1.1 describes the standard operative procedural steps involved in LC [109], where it's obvious that identifying the cystic duct, cystic artery, and common bile duct is very important to retrieve the gallbladder safely.

The Society of American Gastrointestinal and Endoscopic Surgeons Safe Chole Task Force defines the Critical View of Safety (CVS) to avoid misidentifying the common bile duct or an aberrant duct as the cystic duct that can prevent bile duct injuries. In the CVS procedure, the hepatocystic triangle is firstly cleared from fat and fibrous tissue, then the lower part of the gallbladder is separated from the liver to reveal the cystic plate, where only two structures should be seen entering the gallbladder, i.e., cystic duct and the cystic artery [110]. Achieving CVS is confirmed to be the safest technique to detect Calot's triangle landmarks and it significantly prevents intraoperative complications [111].

Recently, video reporting has become a very important and valuable tool in documenting operative quality and safety in laparoscopic surgery. Endoscopic videos are the richest source of intraoperative information, providing active and efficient information to improve surgical care and safety [13].

In fact, CNN-based deep learning models can capture and objectively analyze surgical videos, and this could be the basis for real-time, intraoperative artificial intelligence (AI) guidance and additional context-specific information to be presented to the surgeon during endoscopic surgeries. They may include detecting safe areas of dissection, where hepatocytic anatomy reliably segmented and unequivocally assess CVS, and producing a video database of these safety steps of LC. Intraoperatively, these models could be used to improve surgeons' awareness, and potentially contribute towards surgical safety [44].

Table 1.1: Standard procedural steps involved in laparoscopic cholecystectomy.

Step	Procedure
1	Retracting the gallbladder (GB) to get the field of view.
2	Calot's triangle Confirmation.
3	Developing Calot's triangle plane in the area and identifying its boundaries based on effective retraction of the GB.
4	Revealing the cystic artery and right hepatic artery.
5	Running direction of common bile duct Confirmation.
6	Dissection around the cystic duct, and performance of the clipping method.
7	Height of the cut-line of the cystic duct.
8	Dissection of the GB from the GB bed with an adequate layer.
9	Bleeding control from the GB bed.
10	Retrieve the GB.

5.1 Experiment Setup

Google Colab notebooks [112] were used as the basic environment, whereas a Jupyter notebook runs in the cloud and is integrated with Google Drive, which makes them easy to set up, share, and access. Collaboratory ("Colab" for short) is a data analysis and machine learning tool that can allow users to combine executable Python code with rich text, charts, HTML, LaTeX, images, and more into one document saved on Google Drive. It connects to the very powerful Google Cloud Platform runtimes, which enables users to easily share and collaborate on their work.

In this work, all the experiments were running on the Tesla P100-PCIE-16 GB GPU with up to 56 multiprocessors and 16,280 MB of memory that were obtained by subscribing to google colabpro+. Several Python libraries were used to implement the process and validate the performance of the proposed deep learning model scripts (e.g., Keras, TensorFlow, and segmentation models).

5.2 Dataset Collection and Preprocessing

A dataset containing 200 videos of cholecystectomy surgeries from the Computational Analysis and Modeling of Medical Activities research group [113], free videos from the World Laparoscopy Hospital [120], and free YouTube videos from the Society of American

Gastrointestinal and Endoscopic Surgeons (SAGES) [121]. The dataset from the Computational Analysis and Modeling of Medical Activities research group contains three groups:

1) The m2cai16-tool dataset contains 15 videos of cholecystectomy procedures from the University Hospital of Strasbourg/IRCAD (Strasbourg, France) [123].
2) The m2cai16-workflow dataset contains 41 videos of cholecystectomy procedures from the University Hospital of Strasbourg/IRCAD (Strasbourg, France) and Hospital Klinikum Rechts der Isar (Munich, Germany).
3) The Cholec80 dataset contains 80 videos of cholecystectomy procedures recorded at the University Hospital of Strasbourg/IRCAD (Strasbourg, France).

These videos are made in 12 different HD resolutions which are as follows: (569 x 334), (578 x 324), (584 x 328), (612 x 344), (612 x 344), (716 x 402), (762 x 428), (774 x 434), (780 x 438), (810x 454), (846 x 476), (854 x 480), and (1920 x1080), with a constant bit rate of 35 MBit/sec. The frame rates of the videos are 30fps, and the video codec is MPEG-2 [116].

Given the anatomical variations in biliary anatomy and the varying degrees of inflammation in cholecystitis, and to maintain a high degree of accuracy of the datasets, we collaborated with expert surgeons from the Department of General Surgery at Western Michigan University Homer Stryker M.D. School of Medicine [117]. Labelling was confirmed by three expert surgeons with 10-20 years of experience. Three chief surgical residents were also involved in this process. The annotation protocol includes initial input from one of the expert surgeons who was trained in minimally invasive surgery with more than a thousand laparoscopic and open cholecystectomies. CVS aims to identify two structures attached to the gallbladder in Calot's triangle (i.e., cystic duct and cystic artery). The primary surgeons and the senior residents discussed the project's focus and testing of the system's ability to detect the intended structures of CVS, and they independently assessed each video/image annotation to allow the deep learning model to learn the most held interpretation of CVS, rather than a single assessment [118]. Based on their recommendations, 14 videos were excluded from initial modelling because of poor image quality due to either bleeding or less visible or obscured landmarks, anatomical differences or misses, or procedures that do not follow the general standard of cholecystectomy.

Each video in this study was trimmed, so the segment of interest starts when CVS was identified until the cystic duct and cystic artery were clipped or ligated. This is contrary to the previous studies which focused on the 60 or 120 seconds before the clipping. The total time of all clipped videos is 65,672 minutes, with an average time of 8 minutes for each video (it ranges from 2:03 minutes to 42:39 minutes). A total of 1,550 images with identified landmarks were extracted from these videos and selected based on feedback from the surgeons.

A Fiji plugin, i.e., annotator [119], was used to efficiently annotate the extracted images with CVS criteria to facilitate qualitative and manual annotation. Figure 5.1 illustrates some examples of annotations for three selected frames extracted from the prepared dataset, where five classes were identified with their related artificial colors that were confirmed by surgeons.

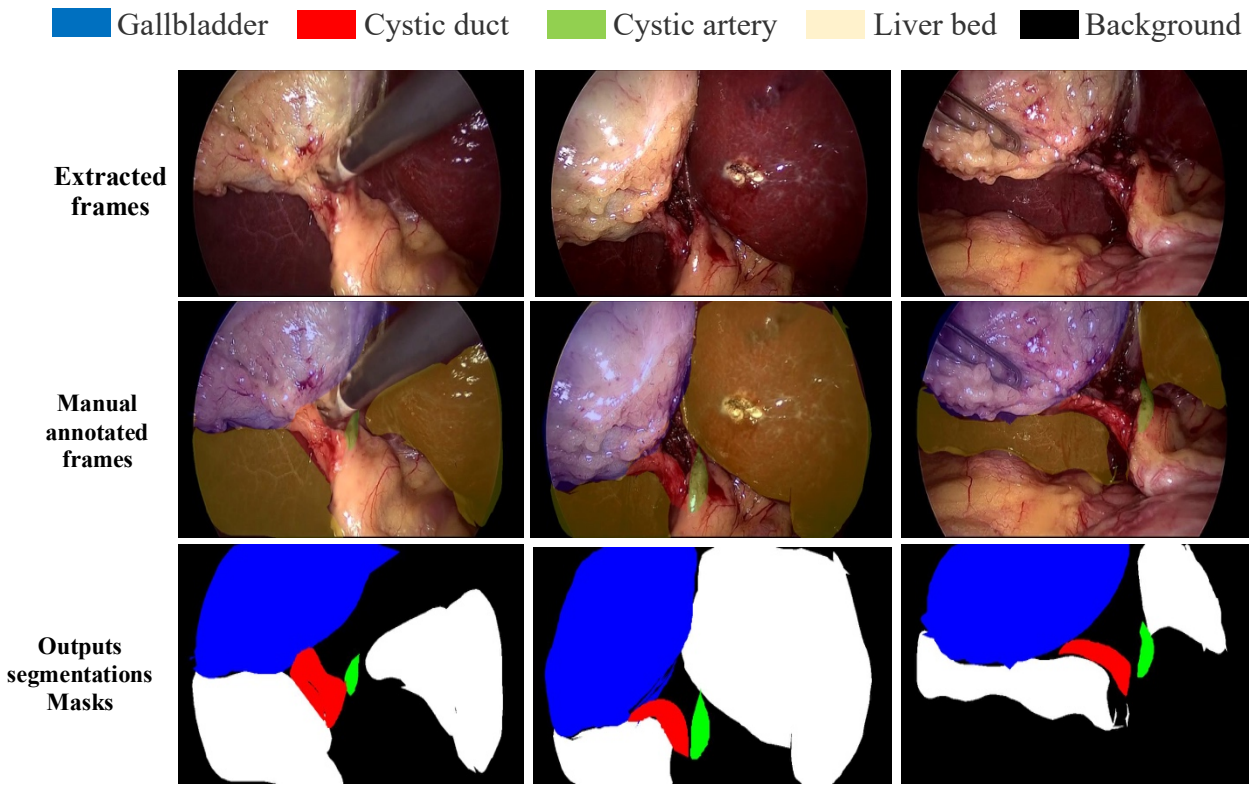


Figure 5.1: Examples of extracted frames annotations from the dataset (first row), their corresponding manual annotated frames (second row), and the output segmentation masks (third row).

5.3 Building the Auto-Encoder

The Auto-Encoder was trained on a part of the dataset to gain pretrained weights that will be transferred to the U-Net encoder layers, where 430 images taken from the prepared dataset were trained for 100 epochs. The training loss was 0.00023727 and accuracy was 91.01%. These weights were transferred to the first 40 layers of the U-Net encoder. Figure 5.2 shows an example of the trained Auto-Encoder input and output images, where the constructed image is very close to the input image with 91% accuracy.

5.4 Experiment Design

We performed five experiments with different weights, models, and techniques to evaluate the proposed model. Each experiment starts by reading the extracted images and masks and resizes them to 256 x 256 pixels. After that the segmentation map was built for the intended labels. The next step was splitting the dataset into training, validation, and testing sets with ratios of 80:10:10, respectively. This was followed by creating an image generator and augmenting the dataset with horizontal and vertical flips and with reflection fill mode.

However, this step was not used in all experiments. The main differences among the experiment's procedures were the ways to select the pretrained weights and the augmentation step decision, as demonstrated in the following flowchart in Figure 5.3. Masks were converted to categories, where all these preprocessing steps are based on the Keras pre-processing image model [100].

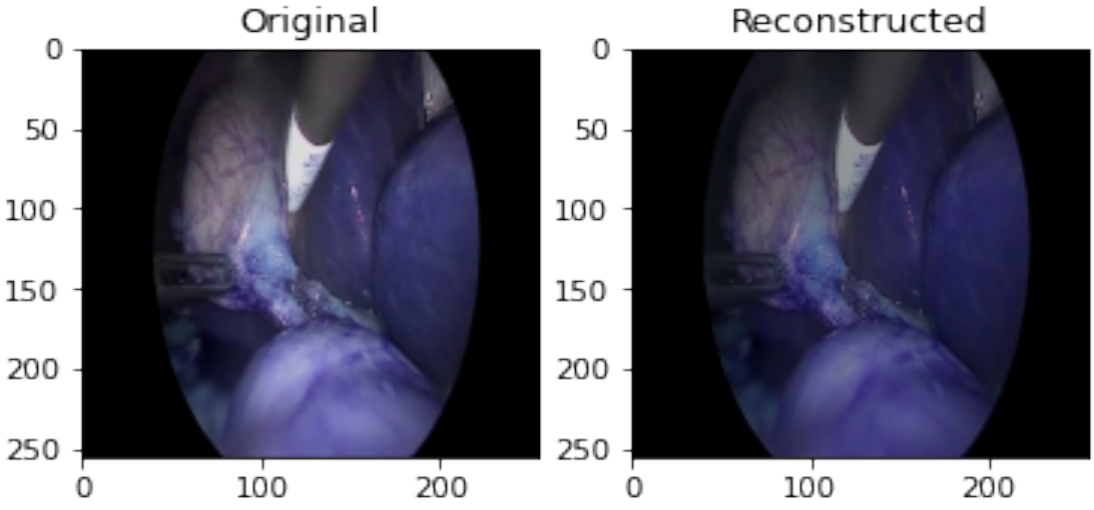


Figure 5.2: Auto-Encoder input and output images.

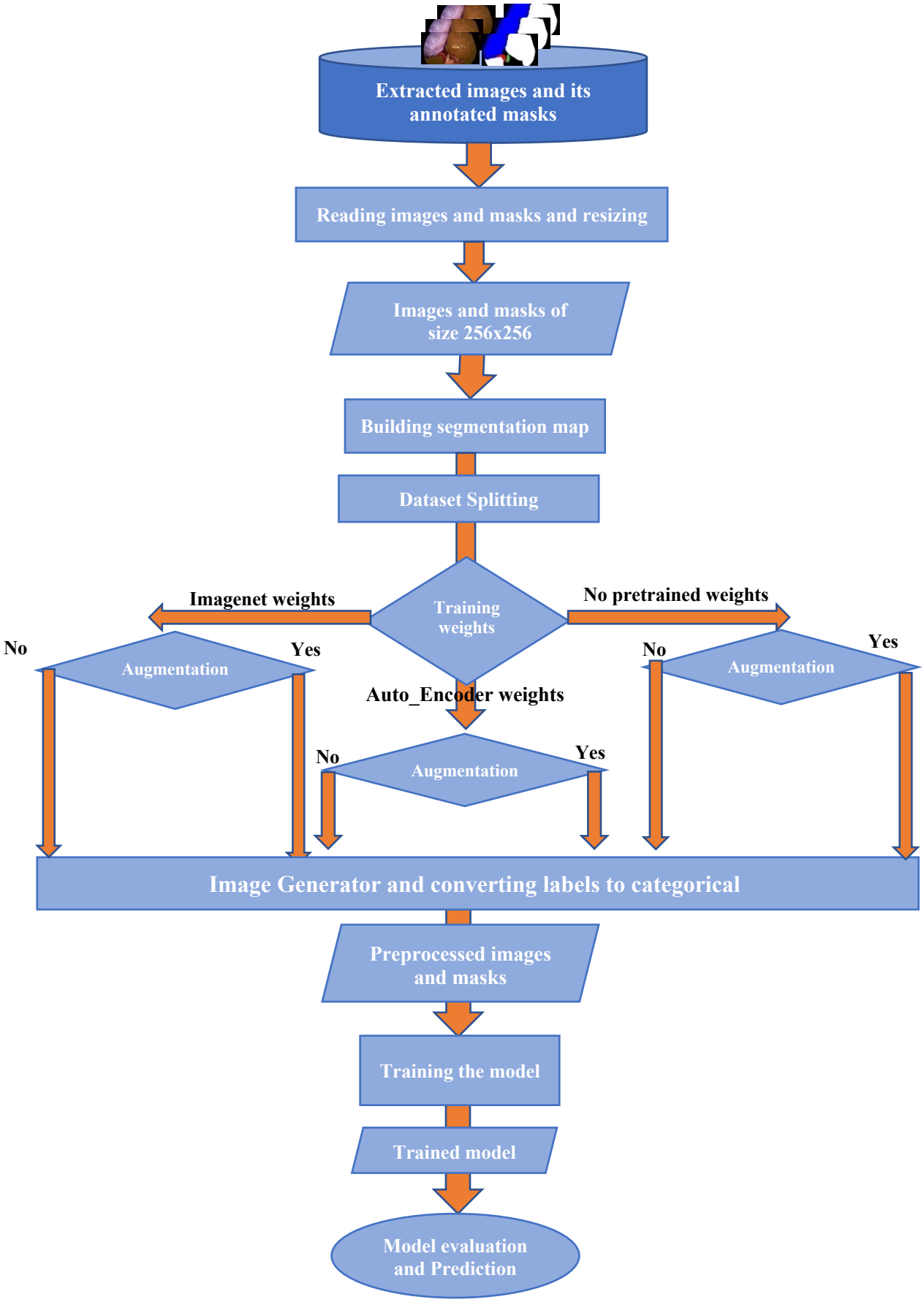


Figure 5.3: Experiment design flowchart.

5.4.1 Experiment 1 (U-Net with Auto-Encoder Weights and without Data Augmentations)

In this experiment, the Autoencoder weights were transferred to U-Net encoder layers and the dataset was not augmented. Figure 5.4 illustrates the training and validation metrics curves, i.e., Accuracy, Loss, IoU, Precision, Recall, Sensitivity, and Specificity, where the training values go smoothly and reach 0.92, 0.83, 0.87, 0.939, 0.87, 0.99, and 0.99 respectively. The training values go smoothly, but the validation values stay in the same range after epoch 20 or 30, except for Sensitivity validation values that dropped after 30 epochs. The mean IoU is 0.747, and the Hausdorff distance is 28.

5.4.2 Experiment 2 (U-Net with Auto-Encoder Weights and Data Augmentations)

In this experiment, the Autoencoder weights were transferred to U-Net encoder layers and the dataset was augmented. Figure 5.5 illustrates the training and validation metrics curves, i.e., Accuracy, Loss, IoU, Precision, Recall, Sensitivity, and Specificity, where the training values go smoothly and reach 0.91, 0.78, 0.87, 0.92, 0.91, 0.99, and 0.99 respectively. While validation values stay in the same range after epoch 20 or 30, except Sensitivity validation values dropped slightly after 40 epochs. The mean IoU is 0.74, and the Hausdorff distance equals 29.27.

5.4.3 Experiment 3 (U-Net backbone with Data Augmentations)

In this experiment, the U-Net was trained without pre-trained weights, and the dataset was augmented. In Figure 5.6, the training and validation metrics curves go smoothly with some small variations in validation values. The training values are summarized in Table 1.2, where the mean IoU is 0.737, and the Hausdorff distance is 27.

5.4.4 Experiment 4 (U-Net Backbone without Data Augmentations)

In this experiment, the U-Net was trained without pre-trained weights and the dataset was not augmented. In Figure 5.7, the training and validation metrics curves are shown, where the training curve goes smoothly, while the validation values fluctuate through the training time, especially in Sensitivity values that fluctuate in the same range, and the Precision values dropped at the beginning and start increasing with fluctuation after epoch 10. The training values are summarized in Table 1.2, where the mean IoU is 0.71, and the Hausdorff distance equals 31.8.

5.4.5 Experiment 5 (Resnet backbone with imagenet Pretrained Weights)

In this experiment, the dataset was trained using Resnet backbone with imagenet [120] pretrained weights. In Figure 5.8, the training and validation metrics curves fluctuate strongly through the training time, especially the validation values, where they fluctuate in the same region and decrease in Sensitivity and Specificity values. The mean IoU is 0.809, and the Hausdorff distance equals 15.8.

Following the standard cross-validation strategy, we performed fivefold cross-validation on a random split of the prepared dataset. The training iterations were stopped at 100 epochs for each fold in the designed experiments detailed above. Table 1.3 includes the quantitative results of the performance assessment procedures for the proposed approach for all folds, in each experiment.

A qualitative comparison for all experiments is illustrated in Figure 5.9, each row shows the original tested image of size 256x256, combined with its ground truth annotated image, each experiment segmentation prediction result, and its calculated Hausdorff Distance. Each color corresponds to a different label, where yellow color labels background, light green is used for the liver, presidential blue for the cystic artery, cadet blue for the gallbladder, and royal blue for the cystic duct.

Finally, the produced model was tested using new videos to tune it based on the gained results. More details on this step will be discussed in the next chapter.

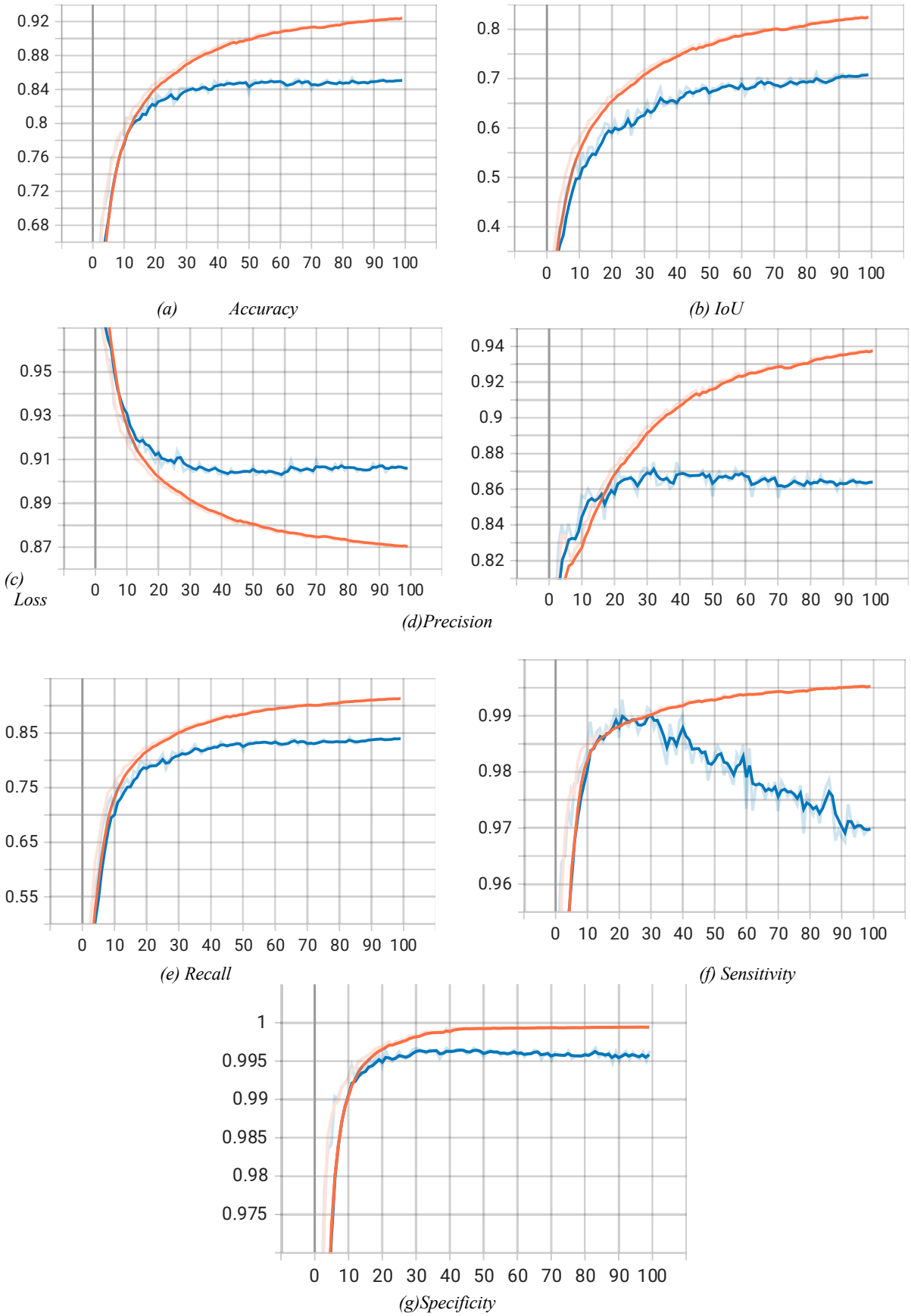


Figure 5.4 : U-Net with Auto-Encoder weights and without data augmentations experiment evaluation metrics.

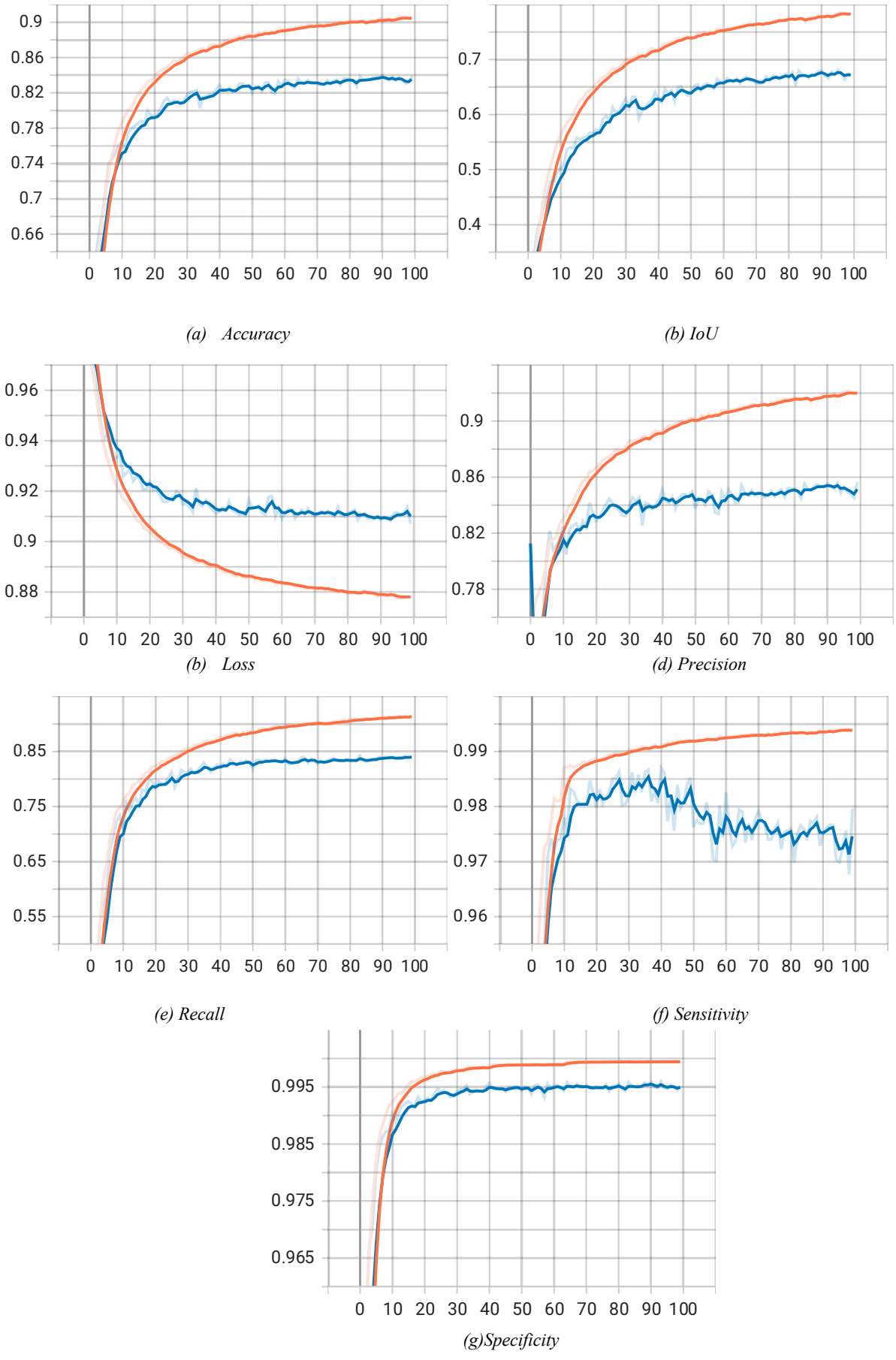
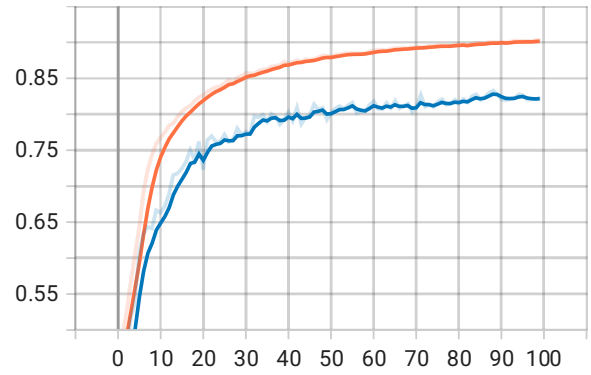
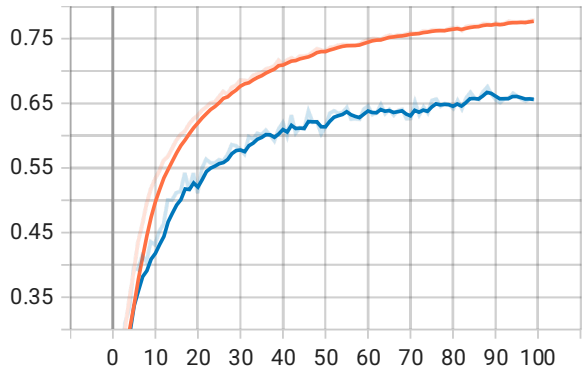


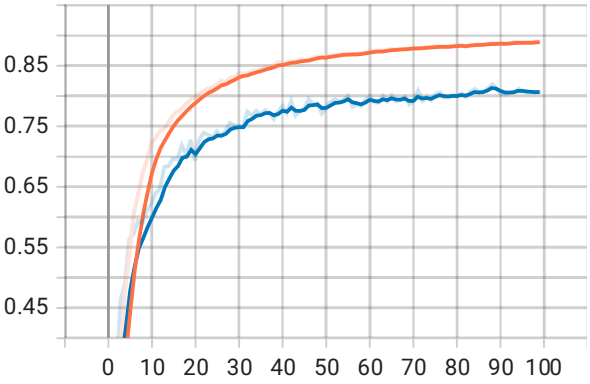
Figure 5.5: U-Net with Auto-Encoder weights and data augmentations experiment evaluation metrics.



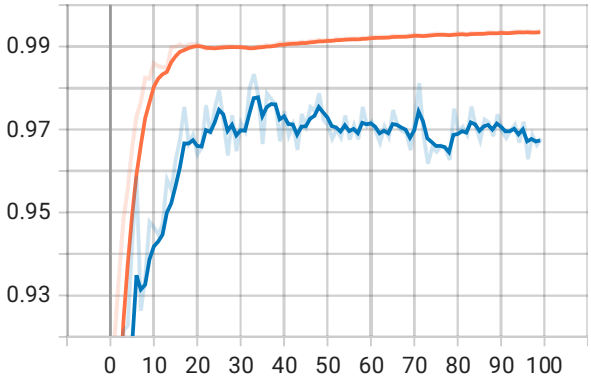
(a) Accuracy



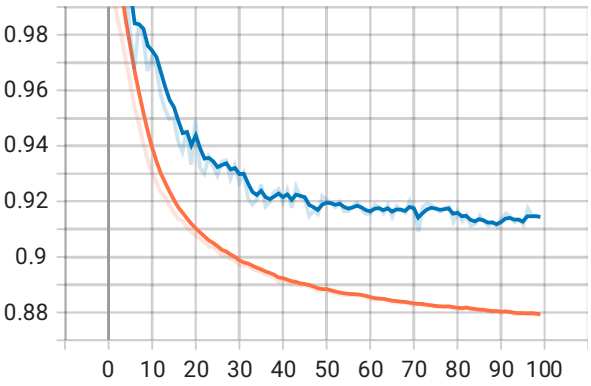
(b) IoU



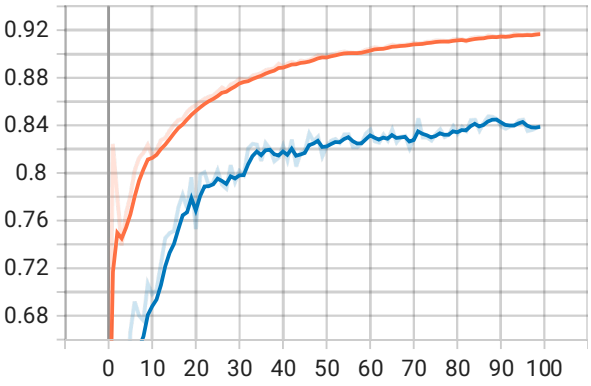
(c) Loss



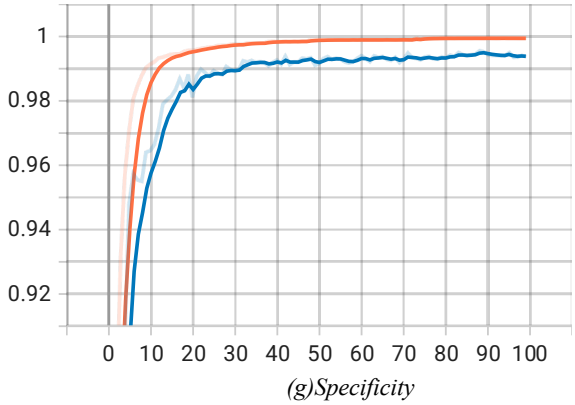
(d) Precision



(e) Recall



(f) Sensitivity



(g)Specificity

Figure 5.6: U-Net backbone with data augmentations experiment evaluation metrics.

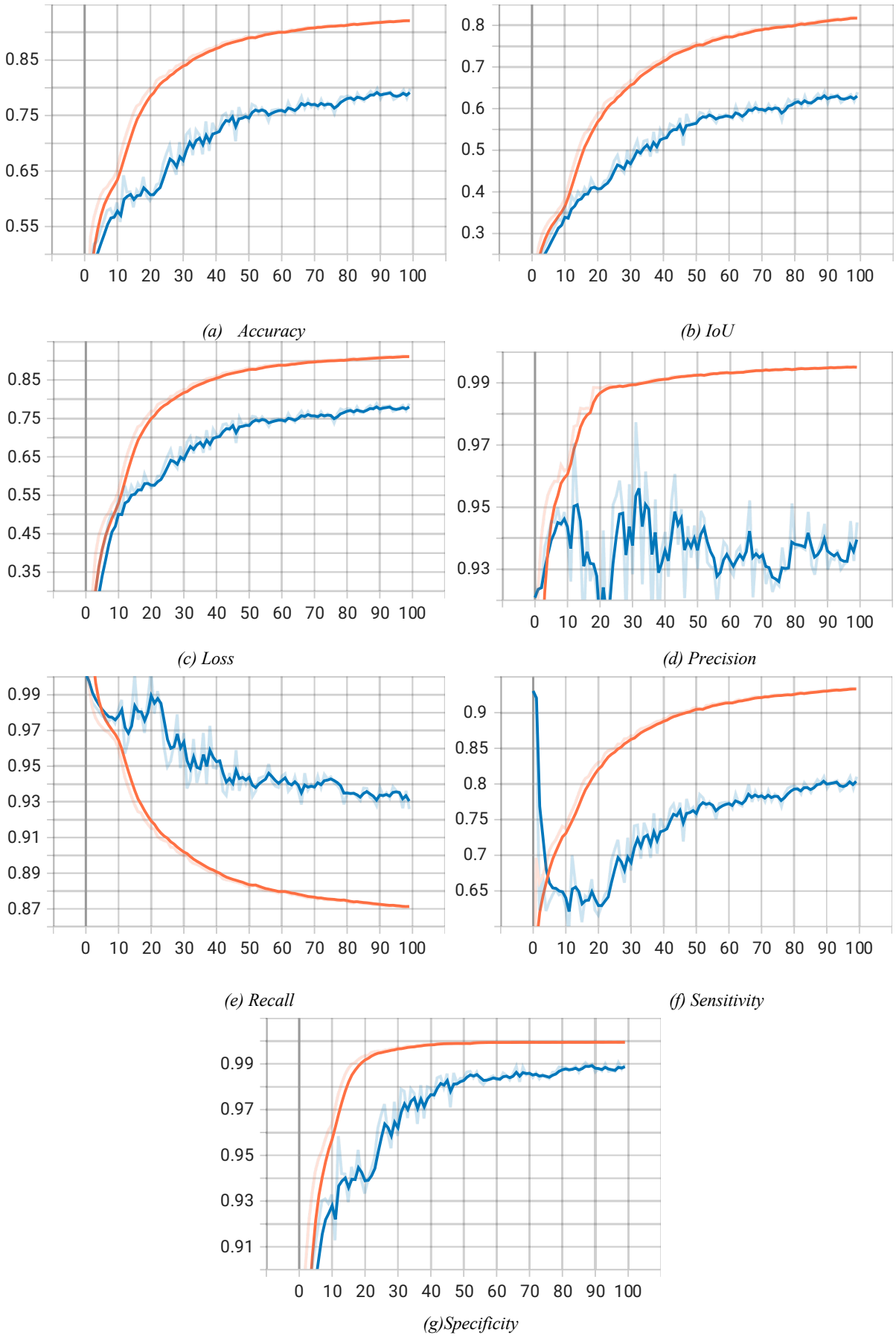


Figure 5.7: U-Net backbone without data augmentations experiment evaluation metrics.

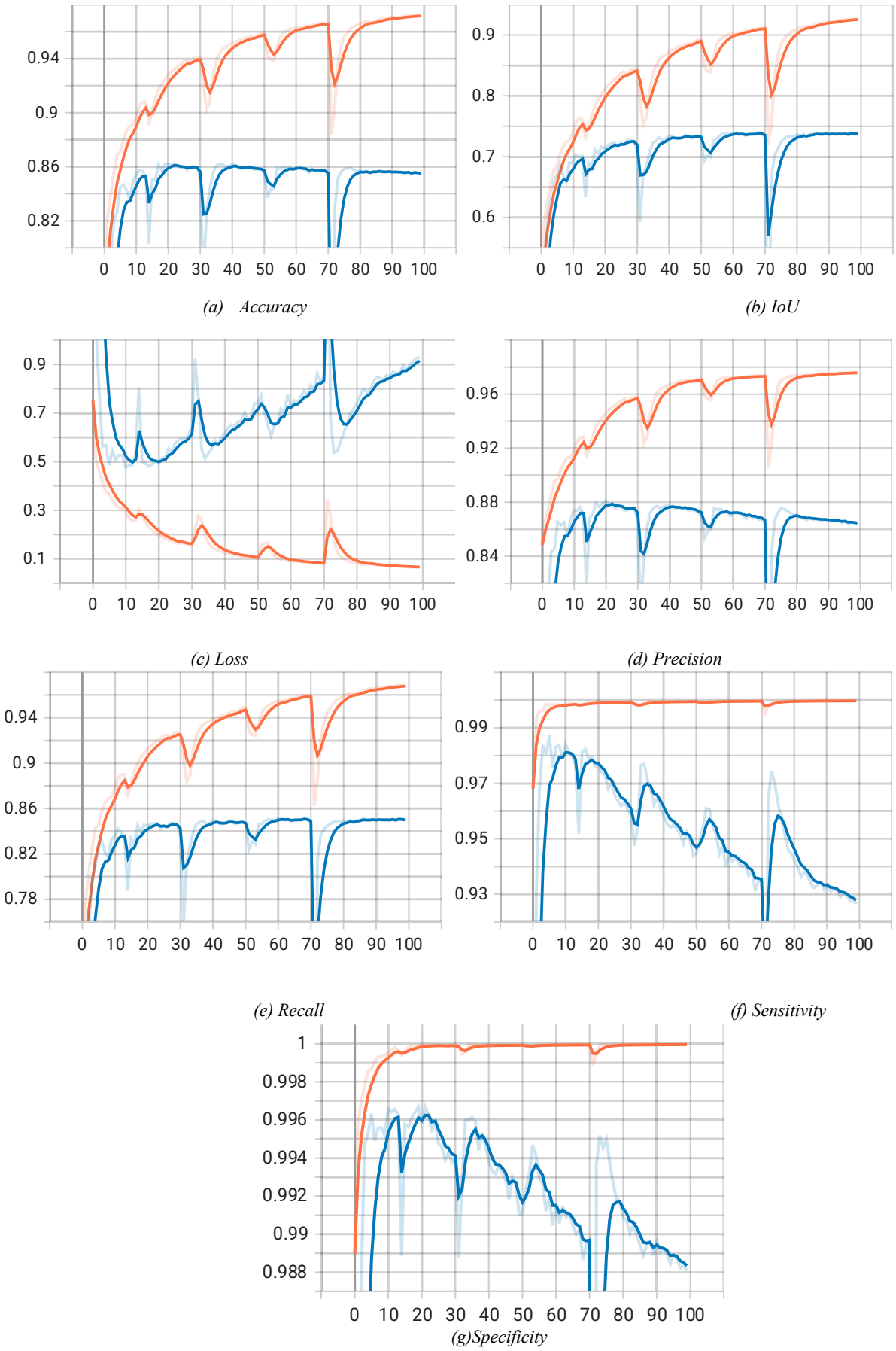


Figure 5.8: Resnet backbone with imagenet pretrained weights experiment evaluation metrics.

Table 1.2: A summary of all experiments Quantitative Results.

Experiment	accuracy %	IoU %	loss %	precision %	recall %	sensitivity %	specificity %	mean IoU%	hausdorff distance
U-Net with Auto-Encoder weights and without Data Augmentations	**92**	**83**	87	**93.9**	87	99.95	99.99	**74.65**	28
U-Net with Auto-Encoder weights and Data Augmentations	**91**	78	92	87	91	99.5	99.95	**74.23**	29.274563
U-Net backbone with Data Augmentations	90	77	88	91	84	99.5	99.99	73.71	27
U-Net Backbone without Data Augmentations	90.5	81	87	93	85	99.5	99.99	71.27	31.764761
Resnet backbone with imagenet pretrained weights	**95**	93	9	**97**	96.5	99.99	99.99	**80.87**	15.779734

››

Table 1.3: Quantitative results of performance assessment procedures for the proposed approach for all folds.

Fold Number	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Experiment 1 (U-Net with Auto-Encoder weights and without Data Augmentations)					
Accuracy%	84.40	88.67	91.20	91.49	**91.99**
Loss%	90.90	88.88	87.65	87.48	**87.32**
Experiment 2(U-Net with Auto-Encoder weights Data Augmentations)					
Accuracy%	82.73	87.51	87.62	89.17	89.10
Loss%	91.51	89.22	89.31	88.62	88.53
Experiment 3 (U-Net backbone with Data Augmentations)					
Accuracy%	83.48	86.46	87.68	88.57	89.34
Loss%	90.95	89.84	89.45	88.84	88.53
Experiment 4 (U-Net Backbone without Data Augmentations)					
Accuracy%	83.41	88.66	89.39	91.86	91.69
Loss%	91.05	88.75	88.47	87.28	87.35
Experiment 5 (Resnet backbone with imagenet pretrained weights)					
Accuracy%	85.78	92.39	94.42	96.79	**97.67**
Loss%	89.97	38.88	26.77	12.97	**9.52**

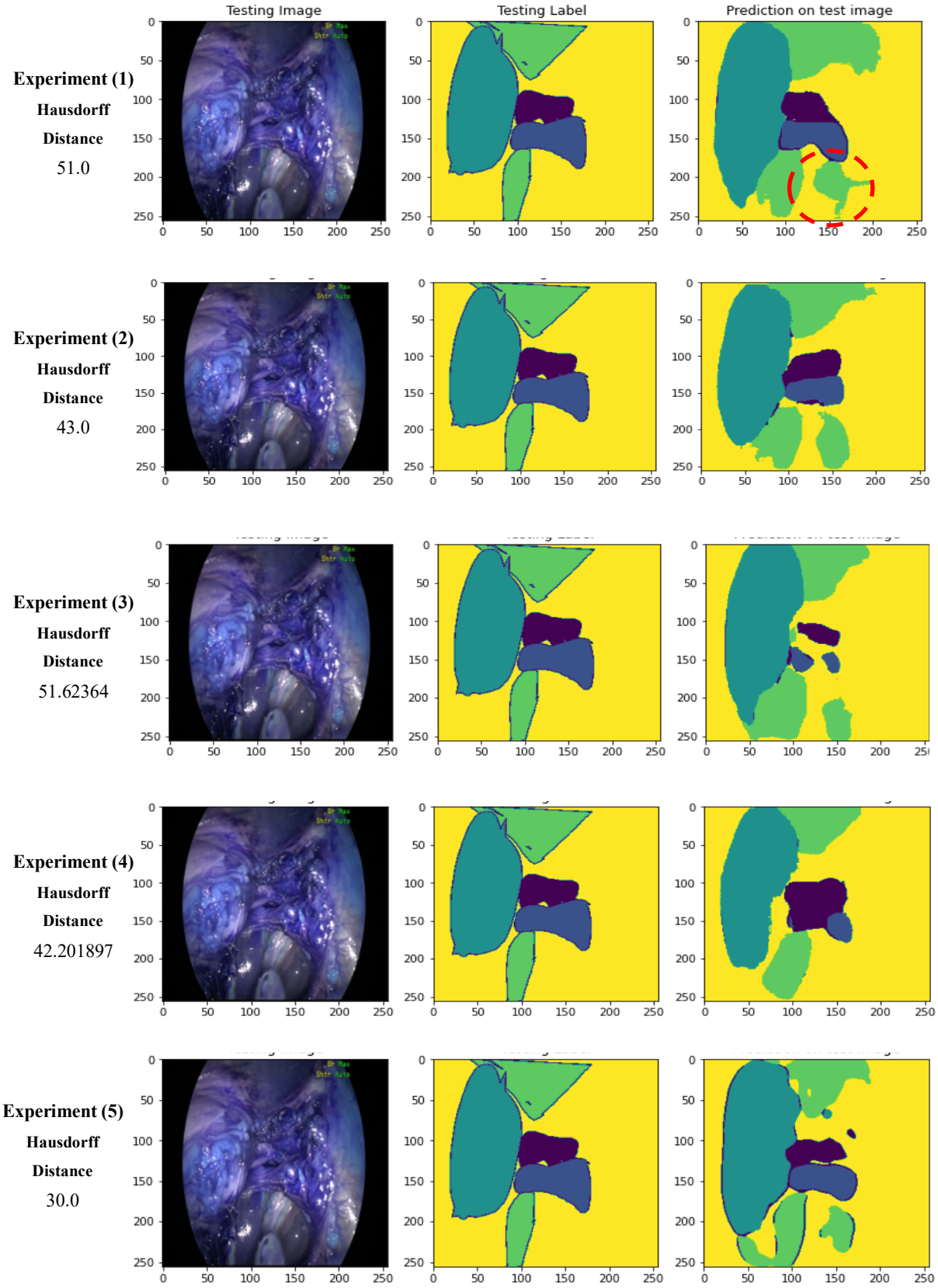


Figure 5.9: A qualitative comparison for all experiments, each row displays the original tested image of size of 256x256, combined with its ground truth annotated image, and each experiment segmentation prediction result, and its calculated Hausdorff Distance.

CHAPTER 6

RESULTS AND DISCUSSION

This chapter will discuss the proposed model results illustrated earlier in the previous chapter. These results are accepted for publication in the Journal Surgical Endoscopy [127]. This work uses deep learning neural networks that combine Auto-Encoder with U-Net to segment cystic duct and cystic artery anatomy, which are very important in the prediction of CVS criteria achievement in Laparoscopic Cholecystectomy.

6.1 Quantitative and Qualitative Results

Based on the results listed in Tables 1.2-1.3, the proposed model that uses Auto-Encoder weights as pretrained weights for U-Net encoder layers without augmentation gave the highest accuracy, mean IoU, and loss, compared to all other scenarios, except Resnet backbone with imagenet [120] pretrained weights, which is expected, as imagenet weights are very large compared to the limited prepared Auto-Encoder weights used in the model. It should be noted that imagenet contains more than 14 million images that belong to more than 20,000 classes. This leads us to work on preparing more data that can help create larger weights and produce better results. This was also approved through 5-fold cross validation results bolded in Table 1.3.

Figures 5.4 through 5.8 demonstrate the smoothness and the continuity of the model evaluation metrics curves through the training and validation compared to the other scenarios, except for the validation sensitivity curve that has some fluctuations and decreases due to classes imbalance as shown in Figure 5.4 (f), where it gave better results when the data was augmented as in Figure 5.5 (f). In case of the Resnet backbone with imagenet pretrained weights, the training and validation curves have strong fluctuations, which indicates random classification.

From the qualitative comparison illustrated in Figure 5.9, the proposed model produces a smooth, accurate, and sometimes even better prediction of the tested image compared to the ground truth annotated image. This is obvious in the first, second, and last rows, where some liver parts were not labeled in the ground truth image and are still predicted correctly in the results. In addition to that, the third and fourth experiments failed to predict

cystic duct as illustrated in the third and fourth rows. This approves the important role of using Auto-Encoder weights in the proposed model compared to training the model without any weights, where pretrained weights help in better feature extraction and fast convergence time.

6.2 Video Segmentation

Based on the results we had gained from testing the proposed model on different images, we deployed our model to segmented LC videos. The processing time for each segmented frame was around 24 milliseconds, the segmented video view was clear, and it had a flexible frame per second (fps) rate.

In addition to that, the model can generate two views of the segmented videos, with the first view showing only the segmented parts, and the second view adding the original image as the background which is converted to grayscale. Figure 6.1 presents three selected frames from each generated video view.

This work can analyze dynamic videos with the assumption that it would facilitate CVS identification. These dynamic views can provide richer information than flat still images, and can be particularly important in improving the performance of deep learning models when assessing the clearance of the intended landmarks, i.e., (cystic duct and cystic artery).

6.3 Comparison with Related Approaches in Literature

Five approaches reported in the literature to try to identify hepatocytic anatomy reliably and assess the availability of CVS in laparoscopic cholecystectomy videos using deep learning neural networks. Table 1.4 includes a comparison summary between these approaches and the proposed model. Included in the summary are the dataset volume, the number of extracted images, the surgeon's confirmation, the annotation tools used, the model built, and the results.

Compared to previously reported approaches and models, the efficiency of the proposed approach in our study performed well for autonomous hepatocytic landmarks identification with an accuracy of 92%, and 74.65% for mean IoU, despite the limited number

of videos. The difference could be explained by the way the models were constructed in these studies.

Furthermore, not all approaches can segment the intended landmarks, only Mascagni et al. [57], Madani et al. [19] and Namazi et al. [64] did that with no details of the segmentation model structure they used, and with limited results compared to the proposed model's results.

However, the authors in [60] only defined safe or non-safe zones without segmenting the hepatocytic landmarks. The authors in [62] used object detection with bounding boxes to detect the anatomical landmarks with low precision results. This is expected, as we mentioned earlier, that semantic segmentation is more appropriate than bounding boxes, where structures have more complex geometric configurations and blend with their background.

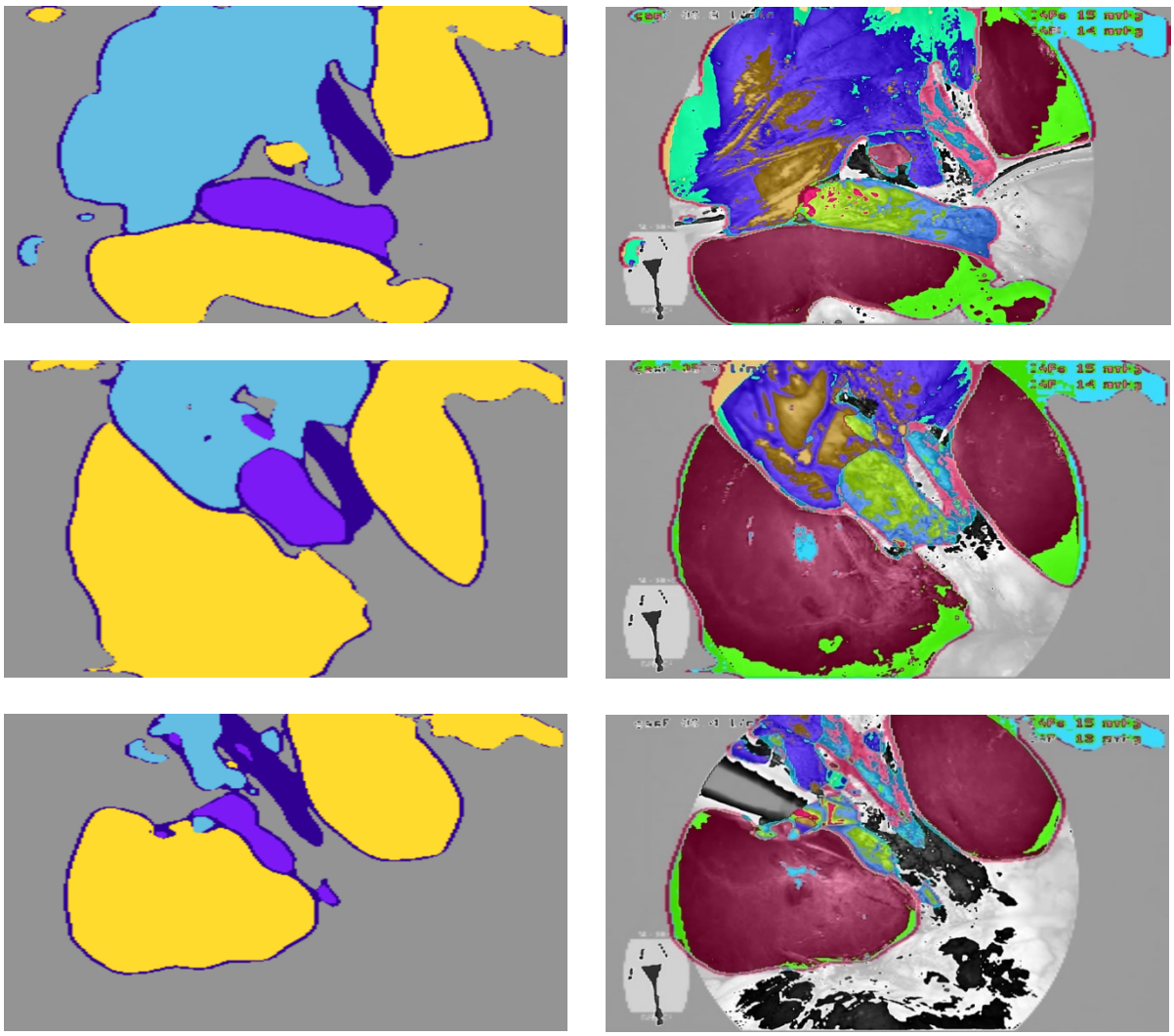


Figure 6.1: Video segmentation views, the first column represents the first view, and the second column represents the second view of the selected frames.

Table 1.4: A comparison summary of the proposed model with the related reported approaches in the literature.

Criteria	**Dataset**				**Model**	**Results**	
Approach	**Size**	**Extracted Images**	**Surgeons Confirmation**	**Annotation Tool**		**Accuracy%**	**Mean IoU%**
The proposed model	200 LC videos	1550 images were extracted from the moment of a clear view of cystic structures until just prior to clipping of it	Two expert surgeons and three 5th- year residents	A Fiji plugin (annotator) [119]	U-Net with Auto-Encoder	**92**	**74.65414**
Madani et al. [20]	308 LC videos	2627 frames were randomly-selected from the moment the gallbladder was grasped until just before clipping of the cystic structures	Three acute care and minimally-invasive surgeons, and one high-volume hepatobiliary surgeon	Think Like A Surgeon software [121]	ResNet50 with multi-scale pyramid pooling.	90.0	50.0
Mascagni et al. [58]	201 LC videos	402 images were selected from 2854 extracted frames starting just 60 seconds precedes the first clipping of cystic structures (ie, 1 frame per second)	Two Surgeons	Pixel Annotation Tool [122]	DeepCVS combines a Segmentation model with multi-label classification network	71.9	66.6
Namazi et al. [65]	120 LC videos	Frames were extracted from two minutes snippet just before clipping of cystic structures (one per second)	Trained medical students	Not reported	A Convolutional Neural Network (CNN) with a Recurrent Neural Network (RNN)	81	Not reported
Altieri et al. [61]	63 LC videos	264 frames	Two high-volume surgeons and reviewed for accuracy by a hepatobiliary surgeon	Free-hand annotations	Not reported	89	Not reported
Tokuyasu et al. [63]	99 LC videos	2339 frames were extracted, Where images with a certain degree of similarity were selected for landmark labeling	Two expert surgeons	Not reported	YOLOv3 Object detection with bounding boxes	Average precision = 20.225	Not reported

In this study, we showed that the proposed approach allows when and how to select the extracted frames from the LC videos. The proposed model dataset's images were not selected randomly or within a fixed window of time, unlike previous studies; they were carefully selected from the beginning of the video to allow the model to learn more features, and to handle different situations and times.

This work illustrates how U-Net performance can be enhanced using a fine-tuning technique to initialize encoder weights in a segmentation network. Potentially, fine-tuning in image segmentation is very important, as collecting a large volume of the training dataset (in particular for medical images) and qualitatively labeling it is very difficult, tedious, expensive, and subjective to human error. Moreover, pre-trained networks reduce the training time and also prevent over-fitting, as was demonstrated by the model validation curve results.

Additionally, the overall performance improved when using Resnet backbone with imagenet [120] pretrained weights.

On these grounds, we believe that the proposed model of deep learning will have a potential role in surgical education and performance assessment. The model’s output (the segmented images and videos) may provide feedback on performance and facilitate deliberate practices (such as identifying hepatocytic landmarks).

CHAPTER 7

CHALLENGING CASES SEGMENTATION

As surgical videos are filled with visual disturbances, that makes anatomical structure identification very difficult. The following sections will discuss how the proposed model can successfully deal with different challenging cases during LC procedures.

7.1 Blurry View

Fogging and polluting the lens are the main reasons for the blurry view during surgeries. The proposed model can predict the intended landmarks despite a blurry view with the Hausdorff Distance equals ~~to~~ 17.03 as illustrated in Figure 7.1.

7.2 Blood Bleeding Challenge

One of the biggest challenges during surgery is blood bleeding, which covers the anatomical landmarks and makes it difficult to identify their structures. The proposed model can predict the intended landmarks successfully with blood bleeding as demonstrated in Figure 7.2, with the Hausdorff Distance equals 41.79 for the first example and 26.25 for the second one.

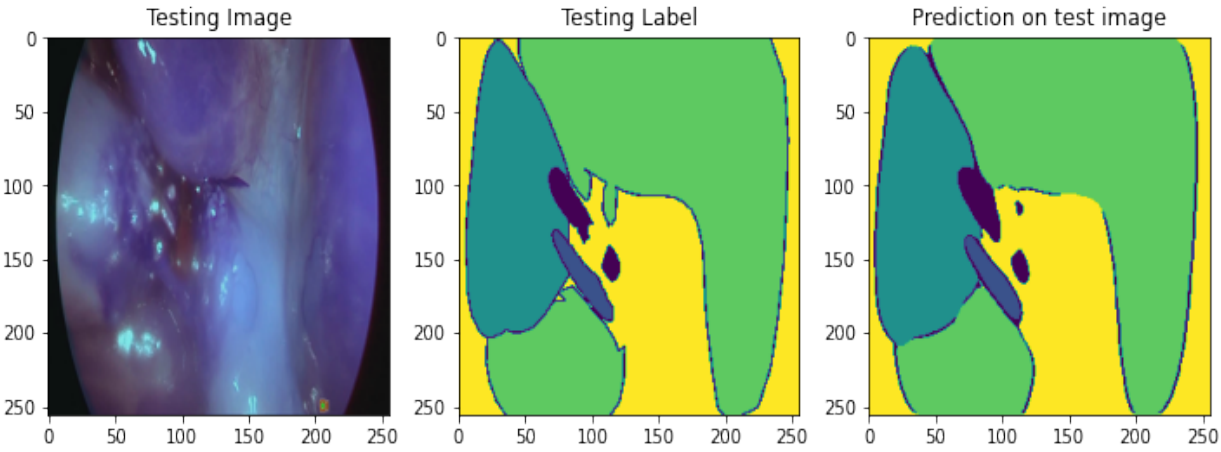


Figure 7.1: Blurry view challenge prediction.

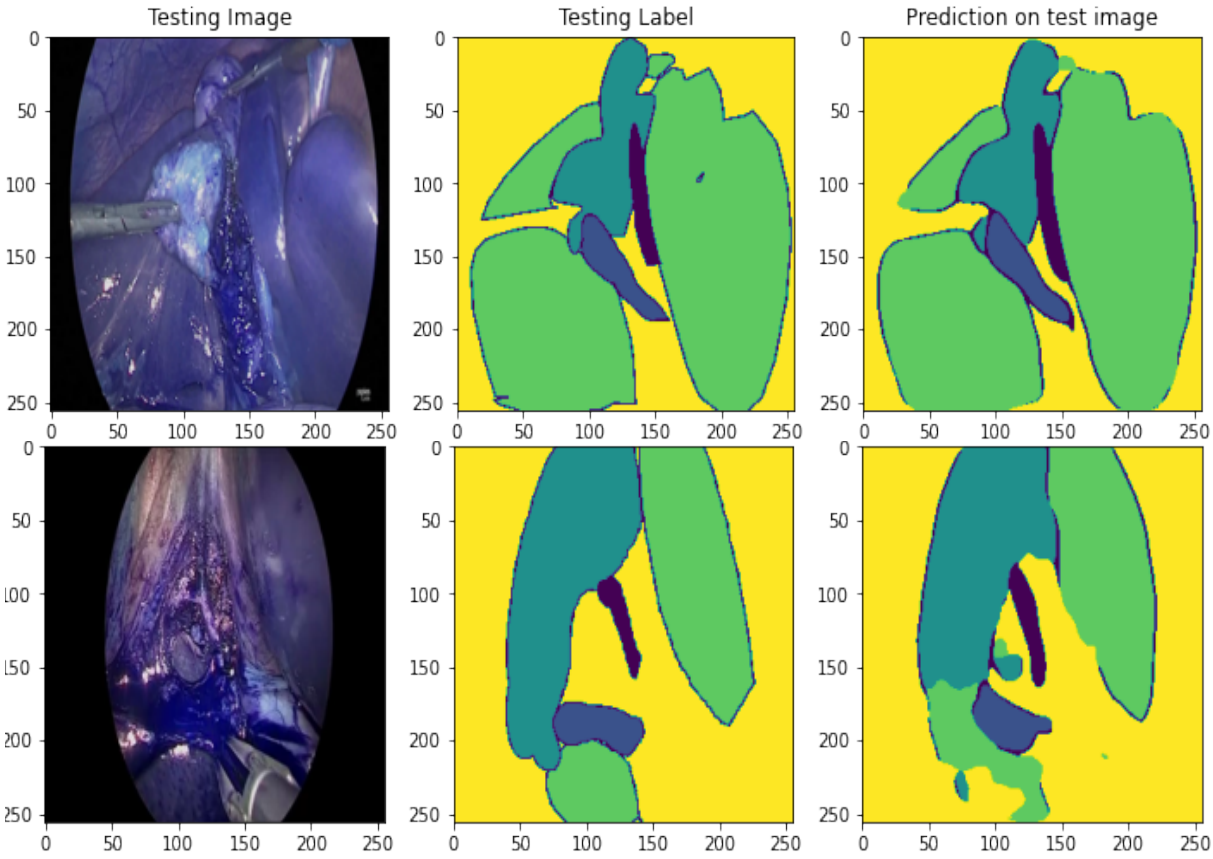


Figure 7.2: Structures with blood bleeding challenge prediction.

7.3 Diverse Structures' Shapes, Sizes, and Being Covered by Other Tissues

As abdominal organs differ between patients, and it's not an easy task to identify the intended landmarks during surgery, where these structures have different shapes, sizes, colors, and sometimes lie hidden within other tissues such as fat or another structure. The proposed model confirms its efficiency in detecting different hidden landmarks with high accuracy as shown in Figure 6.3 below with Hausdorff Distance equals 22.63, 12.37, 22.36, and 12, respectively, for each row. In the first-row example, it's very difficult to detect the cystic artery as it's not clear and it's also hidden by other tissues. In the second row, there is a lot of fat surrounding the structures. The third and fourth row examples demonstrate different shapes and sizes of the intended landmarks. The proposed model can succeed in most cases by segmenting the cystic duct and the cystic artery correctly.

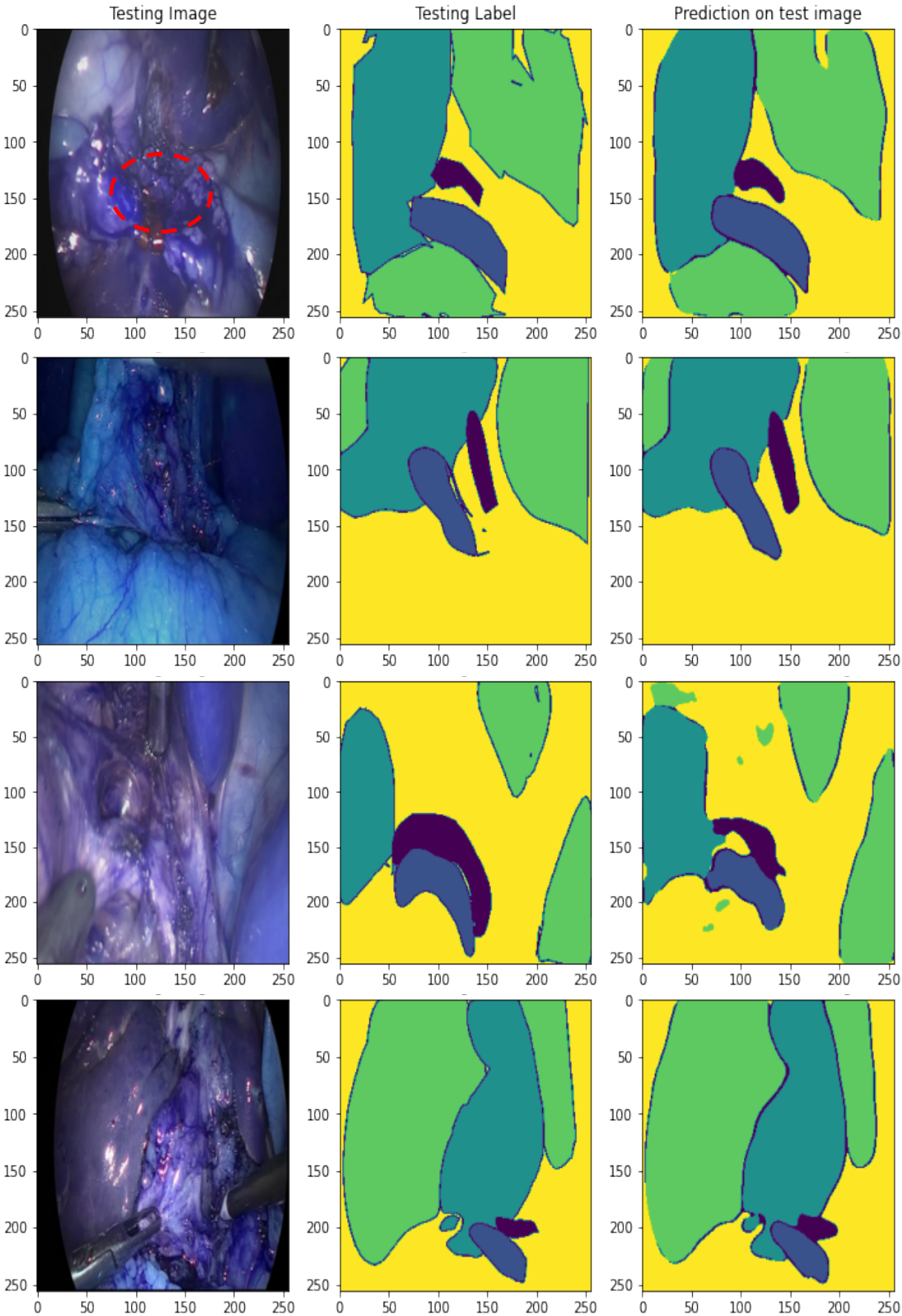


Figure 7.3: Diverse structures shapes, sizes and lie hidden within other tissues prediction.

7.4 Cystic Artery and Cystic Duct Cutting Identification

Detecting Cystic Artery and Cystic Duct Cutting during surgery is very important, and can help surgeons, especially, when the view is not clear to identify that the structure is already cut. Figure 7.4 demonstrates how the proposed model can predict the cutting successfully, the first example shows the prediction of cystic duct cutting with the Hausdorff Distance of 17.72. The second example shows the prediction of the cystic artery even better than the ground truth mask, where it detects the two cut parts with the Hausdorff Distance equals 20.88.

7.5 Accurate Extraction of the Surgical Tools

Segmenting any unrelated objects from the intended landmarks as the background is very important, making the prediction clearer and more accurate as illustrated in Figure 7.5, where two surgical tools are extracted accurately in the proposed model with the Hausdorff Distance of 15.

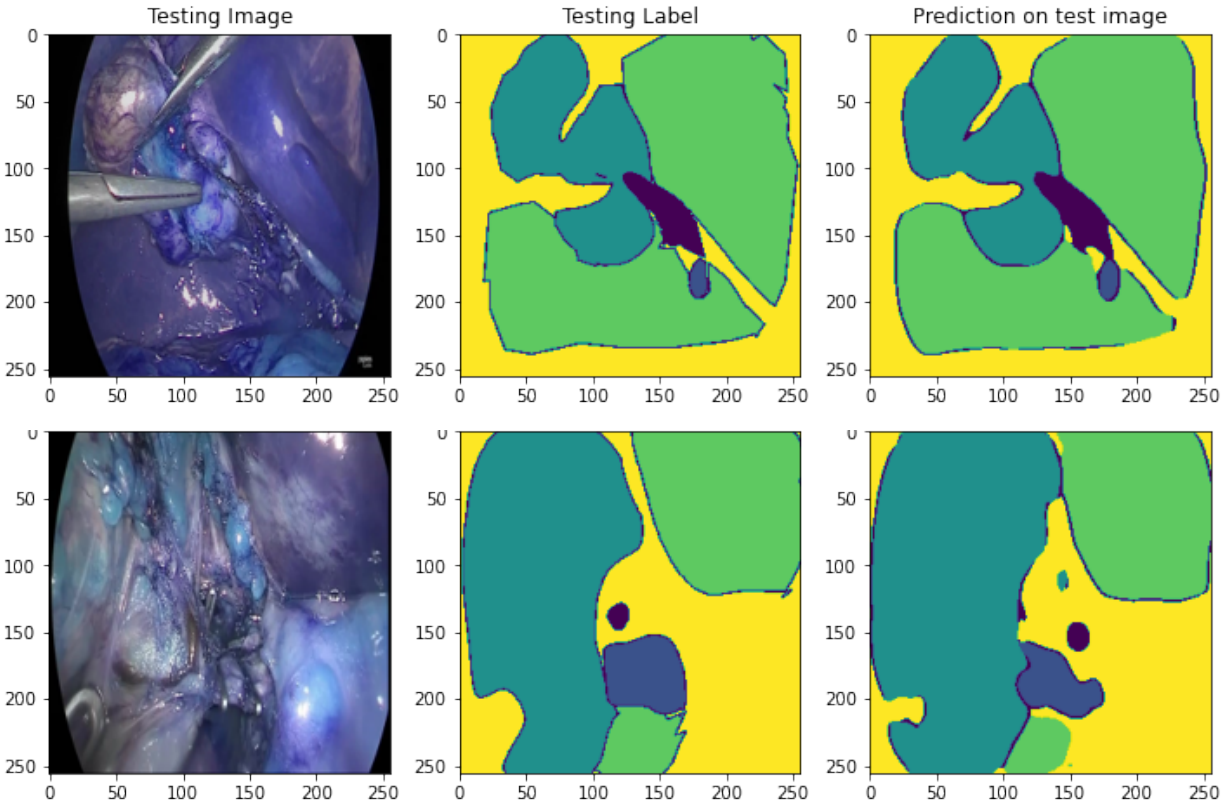


Figure 7.4: Cystic artery and cystic duct cutting prediction.

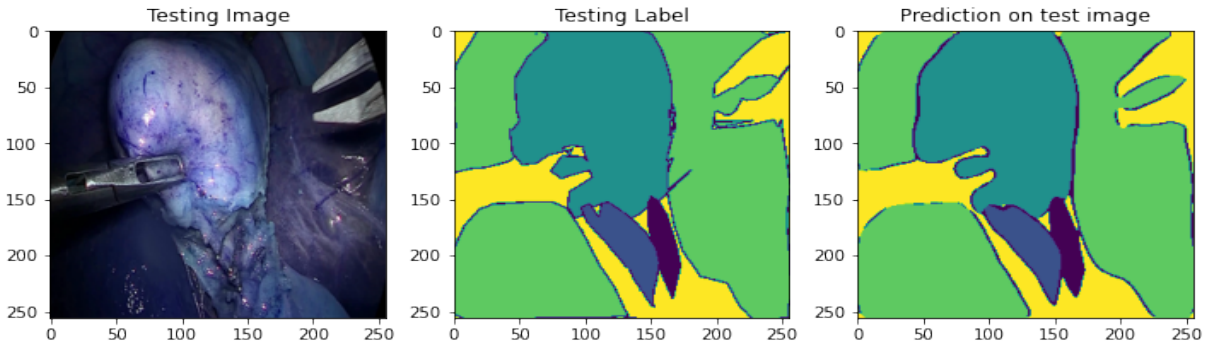


Figure 7.5: Segmenting unrelated objects as background.

7.6 Segmenting Special Landmarks Structures

As each patient has its unique abdominal with different organs shape and sometimes numbers, in a few cases they have duplicated organs, i.e., two gallbladders, two cystic arteries, two cystic ducts, etc. Figure 7.6 below shows different cases; the first case has two gallbladders with two cystic ducts and two cystic arteries. The second case has two cystic ducts illustrated in the second row, while the third row shows a case containing two cystic arteries segmentations. It is important to note that the trained dataset doesn't include any cases with duplicate organs. Based on these results, the proposed model proves its efficiency in segmenting ~~new~~ cases with duplicate organs and other challenges.

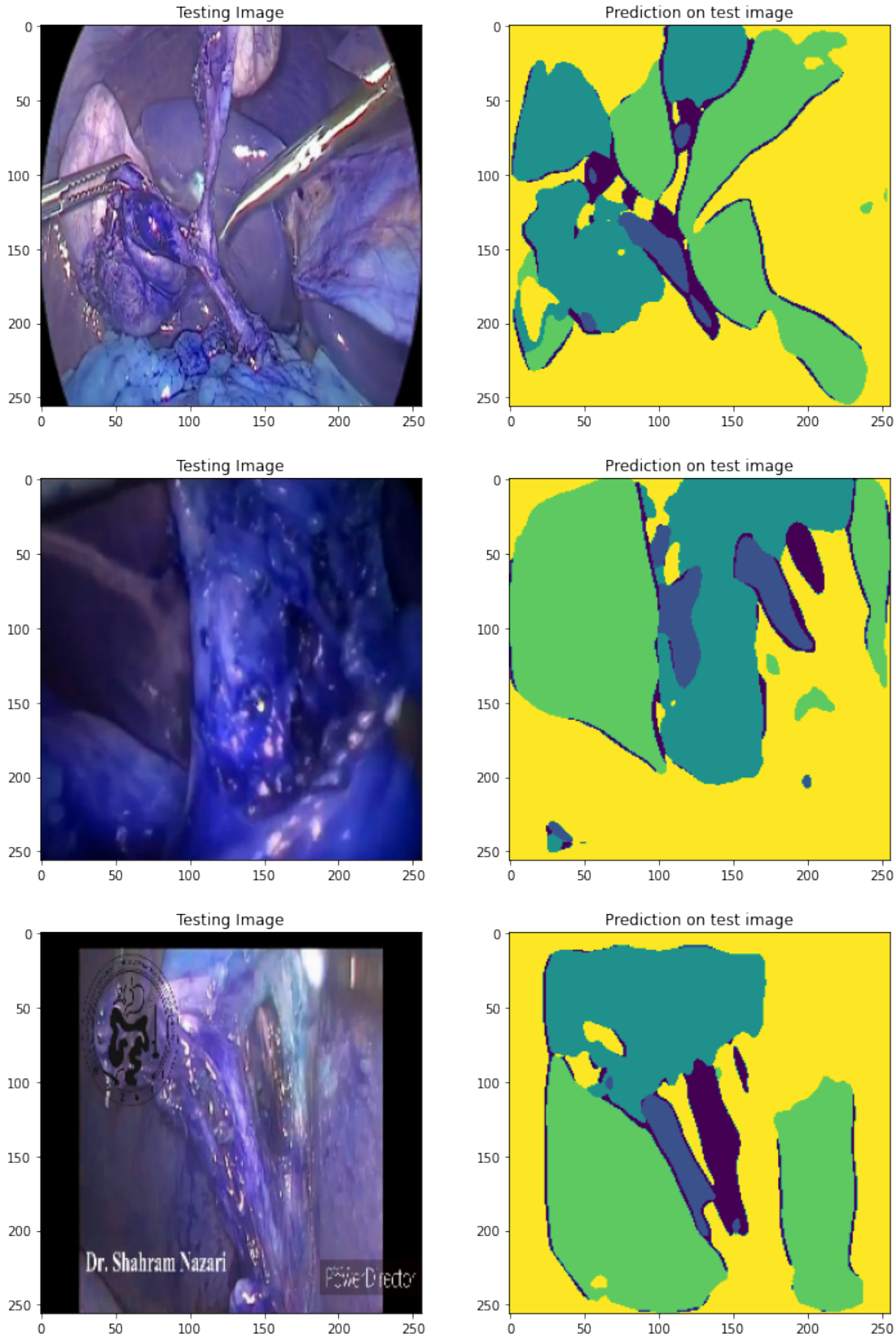


Figure 7.6: Segmenting special landmarks structures with two gallbladders, two cystic ducts and two cystic arteries.

CHAPTER 8

CONCLUSION AND FUTURE WORK

8.1 Summary of Work

In this dissertation, a deep artificial neural network model was proposed for the CVS criteria. It combines the effectiveness of the Auto-Encoder weights with U-Net, where these weights are used as starting weights to automatically segment the hepatocytic anatomy in laparoscopic images. The model can provide reliable guidance and minimize the risk of adverse events and offline for objective performance assessment in LC.

A dataset that combines 1550 images was prepared by extracting selected frames from Laparoscopic Cholecystectomy videos and was annotated by the Fiji annotator to feed the proposed segmentation model. An Auto-Encoder model was built and trained on part of the prepared dataset to produce the desired weights that were used as initial weights for the U-Net encoder layers.

Five experiments were developed to prove the efficiency of the proposed model, where each experiment was trained with different weights and data preprocessing approaches and was tested on part of the dataset. Five cross validation techniques were deployed for each experiment to estimate the performance of unseen data, and to ensure that the model can be generalized. A hybrid loss function was used to calculate the loss of the output results from different levels, and different evaluation metrics were calculated to evaluate the model performance as a segmentation model.

A very powerful Google Cloud Platform runtime was used to build the model, where a Tesla P100-PCIE-16GB GPU with up to 56 multiprocessors and 16280MB of memory was deployed. The trained model was tested using part of the dataset kept for the prediction, in addition to new images that contain challenging cases to check the model efficiency in segmenting the intended hepatocytic landmarks. The proposed model was used to segment LC videos, where the processing time for each frame equals 24ms with a flexible frame per second rate and it can generate two views of the segmented video.

8.2 Contributions

- Contributions of this dissertation include building a deep neural network that combines Auto-Encoder as weight initialization with a U-Net segmentation model to automatically highlight cystic duct and cystic artery anatomy that are very important in the prediction of CVS criteria achievement. Autoencoders of U-Net can efficiently predict hepatocytic landmarks when applied to LC images. The proposed model was tested and showed promising results with smooth and accurate predictions on tested images and videos.

- Another contribution of this work is preparing a precisely confirmed dataset for Laparoscopic Cholecystectomy. The dataset compromising 1550 images with its related labeled masks was prepared and confirmed by expert surgeons.

8.3 Limitations

- Memory usage, where image segmentation task needs large memory. The system fails when the images were cropped into patches instead of resizing, which limits extracting more features.
- The sample size of videos (i.e., 200 videos) could be viewed as a relatively small sample size, although this study has the second largest sample size among the reported studies.
- The proposed model's timing performance applies on prerecorded videos only.

8.4 Future Work

Recommendations for future research projects are as follows:

- Obtaining more Laparoscopic Cholecystectomy videos from different sources can improve and generalize the proposed model results.
- Working with a more powerful machine with more RAM can help in utilizing the image's resolution without the need of resizing, by cropping each image into patches, which will help in extracting more features.
- Deploying different segmentation models with different modifications.
- Working towards real time segmentation during LC procedures.

BIBLIOGRAPHY

[1] ELLIS R. (2005). FROM SCANS TO SUTURES: COMPUTER-ASSISTED ORTHOPEDIC SURGERY IN THE TWENTY-FIRST CENTURY. CONFERENCE PROCEEDINGS: ... ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY. IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY. ANNUAL CONFERENCE, 2005, 7234–7237. HTTPS://DOI.ORG/10.1109/IEMBS.2005.1616180

[2] SHABANZADEH, D. M., & SØRENSEN, L. T. (2012). LAPAROSCOPIC SURGERY COMPARED WITH OPEN SURGERY DECREASES SURGICAL SITE INFECTION IN OBESE PATIENTS: A SYSTEMATIC REVIEW AND META-ANALYSIS. ANNALS OF SURGERY, 256(6), 934–945. HTTPS://DOI.ORG/10.1097/SLA.0B013E318269A46B.

[3] BALLANTYNE G. H. (2002). THE PITFALLS OF LAPAROSCOPIC SURGERY: CHALLENGES FOR ROBOTICS AND TELEROBOTIC SURGERY. SURGICAL LAPAROSCOPY, ENDOSCOPY & PERCUTANEOUS TECHNIQUES, 12(1), 1–5. HTTPS://DOI.ORG/10.1097/00129689-200202000-00001

[4] LITYNSKI G. S. (1998). ERICH MÜHE AND THE REJECTION OF LAPAROSCOPIC CHOLECYSTECTOMY (1985): A SURGEON AHEAD OF HIS TIME. JSLS : JOURNAL OF THE SOCIETY OF LAPAROENDOSCOPIC SURGEONS, 2(4), 341–346.

[5] KEUS, F., DE JONG, J. A., GOOSZEN, H. G., & VAN LAARHOVEN, C. J. (2006). LAPAROSCOPIC VERSUS OPEN CHOLECYSTECTOMY FOR PATIENTS WITH SYMPTOMATIC CHOLECYSTOLITHIASIS. THE COCHRANE DATABASE OF SYSTEMATIC REVIEWS, (4), CD006231. HTTPS://DOI.ORG/10.1002/14651858.CD006231

[6] PUCHER, P. H., BRUNT, L. M., DAVIES, N., LINSK, A., MUNSHI, A., RODRIGUEZ, H. A., FINGERHUT, A., FANELLI, R. D., ASBUN, H., AGGARWAL, R., & SAGES SAFE CHOLECYSTECTOMY TASK FORCE (2018). OUTCOME TRENDS AND SAFETY MEASURES AFTER 30 YEARS OF LAPAROSCOPIC CHOLECYSTECTOMY: A SYSTEMATIC REVIEW AND POOLED DATA ANALYSIS. SURGICAL ENDOSCOPY, 32(5), 2175–2183. HTTPS://DOI.ORG/10.1007/S00464-017-5974-2.

[7] SHAFFER E. A. EPIDEMIOLOGY AND RISK FACTORS FOR GALLSTONE DISEASE: HAS THE PARADIGM CHANGED IN THE 21ST CENTURY? CURR GASTROENTEROL REP. 2005;7(2):132-140. DOI:10.1007/S11894-005-0051-8

[8] BUDDINGH, K. T., WEERSMA, R. K., SAVENIJE, R. A., VAN DAM, G. M., & NIEUWENHUIJS, V. B. (2011). LOWER RATE OF MAJOR BILE DUCT INJURY AND INCREASED INTRAOPERATIVE MANAGEMENT OF COMMON BILE DUCT STONES AFTER IMPLEMENTATION OF ROUTINE INTRAOPERATIVE CHOLANGIOGRAPHY. JOURNAL OF THE AMERICAN COLLEGE OF SURGEONS, 213(2), 267–274. HTTPS://DOI.ORG/10.1016/J.JAMCOLLSURG.2011.03.004

[9] TÖRNQVIST, B., STRÖMBERG, C., PERSSON, G., & NILSSON, M. (2012). EFFECT OF INTENDED INTRAOPERATIVE CHOLANGIOGRAPHY AND EARLY DETECTION OF BILE DUCT INJURY ON SURVIVAL AFTER CHOLECYSTECTOMY: POPULATION BASED COHORT STUDY. BMJ (CLINICAL RESEARCH ED.), 345, E6457. HTTPS://DOI.ORG/10.1136/BMJ.E6457

[10] BERCI, G., HUNTER, J., MORGENSTERN, L., ARREGUI, M., BRUNT, M., CARROLL, B., EDYE, M., FERMELIA, D., FERZLI, G., GREENE, F., PETELIN, J., PHILLIPS, E., PONSKY, J., SAX, H.,

SCHWAITZBERG, S., SOPER, N., SWANSTROM, L., & TRAVERSO, W. (2013). LAPAROSCOPIC CHOLECYSTECTOMY: FIRST, DO NO HARM; SECOND, TAKE CARE OF BILE DUCT STONES. SURGICAL ENDOSCOPY, 27(4), 1051–1054. HTTPS://DOI.ORG/10.1007/S00464-012-2767-5

[11] STRASBERG, S. M., HERTL, M., & SOPER, N. J. (1995). AN ANALYSIS OF THE PROBLEM OF BILIARY INJURY DURING LAPAROSCOPIC CHOLECYSTECTOMY. JOURNAL OF THE AMERICAN COLLEGE OF SURGEONS, 180(1), 101–125.

[12] WAY, L. W., STEWART, L., GANTERT, W., LIU, K., LEE, C. M., WHANG, K., & HUNTER, J. G. (2003). CAUSES AND PREVENTION OF LAPAROSCOPIC BILE DUCT INJURIES: ANALYSIS OF 252 CASES FROM A HUMAN FACTORS AND COGNITIVE PSYCHOLOGY PERSPECTIVE. ANNALS OF SURGERY, 237(4), 460–469. HTTPS://DOI.ORG/10.1097/01.SLA.0000060680.92690.E9

[13] MASCAGNI, P., FIORILLO, C., URADE, T., EMRE, T., YU, T., WAKABAYASHI, T., FELLI, E., PERRETTA, S., SWANSTROM, L., MUTTER, D., MARESCAUX, J., PESSAUX, P., COSTAMAGNA, G., PADOY, N., & DALLEMAGNE, B. (2020). FORMALIZING VIDEO DOCUMENTATION OF THE CRITICAL VIEW OF SAFETY IN LAPAROSCOPIC CHOLECYSTECTOMY: A STEP TOWARDS ARTIFICIAL INTELLIGENCE ASSISTANCE TO IMPROVE SURGICAL SAFETY. SURGICAL ENDOSCOPY, 34(6), 2709–2714. HTTPS://DOI.ORG/10.1007/S00464-019-07149-3

[14] WAY, L. W., STEWART, L., GANTERT, W., LIU, K., LEE, C. M., WHANG, K., & HUNTER, J. G. (2003). CAUSES AND PREVENTION OF LAPAROSCOPIC BILE DUCT INJURIES: ANALYSIS OF 252 CASES FROM A HUMAN FACTORS AND COGNITIVE PSYCHOLOGY PERSPECTIVE. ANNALS OF SURGERY, 237(4), 460.

[15] NIJSSEN, M. A., SCHREINEMAKERS, J. M., MEYER, Z., VAN DER SCHELLING, G. P., CROLLA, R. M., & RIJKEN, A. M. (2015). COMPLICATIONS AFTER LAPAROSCOPIC CHOLECYSTECTOMY: A VIDEO EVALUATION STUDY OF WHETHER THE CRITICAL VIEW OF SAFETY WAS REACHED. WORLD JOURNAL OF SURGERY, 39(7), 1798–1803. HTTPS://DOI.ORG/10.1007/S00268-015-2993-9

[16] MAIER-HEIN, L., VEDULA, S. S., SPEIDEL, S., NAVAB, N., KIKINIS, R., PARK, A., EISENMANN, M., FEUSSNER, H., FORESTIER, G., GIANNAROU, S., HASHIZUME, M., KATIC, D., KENNGOTT, H., KRANZFELDER, M., MALPANI, A., MÄRZ, K., NEUMUTH, T., PADOY, N., PUGH, C., SCHOCH, N., … JANNIN, P. (2017). SURGICAL DATA SCIENCE FOR NEXT-GENERATION INTERVENTIONS. NATURE BIOMEDICAL ENGINEERING, 1(9), 691–696. HTTPS://DOI.ORG/10.1038/S41551-017-0132-7

[17] MASCAGNI, P., FIORILLO, C., URADE, T., EMRE, T., YU, T., WAKABAYASHI, T., FELLI, E., PERRETTA, S., SWANSTROM, L., MUTTER, D., MARESCAUX, J., PESSAUX, P., COSTAMAGNA, G., PADOY, N., & DALLEMAGNE, B. (2020). FORMALIZING VIDEO DOCUMENTATION OF THE CRITICAL VIEW OF SAFETY IN LAPAROSCOPIC CHOLECYSTECTOMY: A STEP TOWARDS ARTIFICIAL INTELLIGENCE ASSISTANCE TO IMPROVE SURGICAL SAFETY. SURGICAL ENDOSCOPY, 34(6), 2709–2714. HTTPS://DOI.ORG/10.1007/S00464-019-07149-3

[18] HASHIMOTO, D. A., ROSMAN, G., RUS, D., & MEIRELES, O. R. (2018). ARTIFICIAL INTELLIGENCE IN SURGERY: PROMISES AND PERILS. ANNALS OF SURGERY, 268(1), 70–76. HTTPS://DOI.ORG/10.1097/SLA.0000000000002693

[19] LECUN, Y., BENGIO, Y., & HINTON, G. (2015). DEEP LEARNING. NATURE, 521(7553), 436–444. HTTPS://DOI.ORG/10.1038/NATURE14539

[20] MADANI, A., NAMAZI, B., ALTIERI, M. S., HASHIMOTO, D. A., RIVERA, A. M., PUCHER, P. H., NAVARRETE-WELTON, A., SANKARANARAYANAN, G., BRUNT, L. M., OKRAINEC, A., & ALSEIDI, A. (2022). ARTIFICIAL INTELLIGENCE FOR INTRAOPERATIVE GUIDANCE: USING SEMANTIC SEGMENTATION TO IDENTIFY SURGICAL ANATOMY DURING LAPAROSCOPIC CHOLECYSTECTOMY. ANNALS OF SURGERY, 276(2), 363–369. HTTPS://DOI.ORG/10.1097/SLA.0000000000004594

[21] HU, J., WANG, H., WANG, J., WANG, Y., HE, F., & ZHANG, J. (2021). SA-NET: A SCALE-ATTENTION NETWORK FOR MEDICAL IMAGE SEGMENTATION. PLOS ONE, 16(4), E0247388. HTTPS://DOI.ORG/10.1371/JOURNAL.PONE.0247388

[22] GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A.; BENGIO, Y. DEEP LEARNING; MIT PRESS: CAMBRIDGE, UK, 2016.

[23] AL-MASNI, M. A., & KIM, D. H. (2021). CMM-NET: CONTEXTUAL MULTI-SCALE MULTI-LEVEL NETWORK FOR EFFICIENT BIOMEDICAL IMAGE SEGMENTATION. SCIENTIFIC REPORTS, 11(1), 10191. HTTPS://DOI.ORG/10.1038/S41598-021-89686-3

[24] RONNEBERGER, O., FISCHER, P. AND BROX, T. (2015) U-NET: CONVOLUTIONAL NETWORKS FOR BIOMEDICAL IMAGE SEGMENTATION. IN: INTERNATIONAL CONFERENCE ON MEDICAL IMAGE COMPUTING AND COMPUTER-ASSISTED INTERVENTION, SPRINGER, CHAM, 234-241.

HTTPS://DOI.ORG/10.1007/978-3-319-24574-4_28

[25] AZAD, R.; ASADI-AGHBOLAGHI, M.; FATHY, M.; ESCALERA, S. BI-DIRECTIONAL CONVLSTM U-NET WITH DENSLEY CONNECTED CONVOLUTIONS. IN PROCEEDINGS OF THE IEEE/CVF INTERNATIONAL CONFERENCE ON COMPUTER VISION WORKSHOP (ICCVW), SEOUL, KOREA, 27–28 OCTOBER 2019.

[26] LOUKAS C. (2018). VIDEO CONTENT ANALYSIS OF SURGICAL PROCEDURES. SURGICAL ENDOSCOPY, 32(2), 553–568. HTTPS://DOI.ORG/10.1007/S00464-017-5878-1

[27] WANG, Z., & MAJEWICZ FEY, A. (2018). DEEP LEARNING WITH CONVOLUTIONAL NEURAL NETWORK FOR OBJECTIVE SKILL EVALUATION IN ROBOT-ASSISTED SURGERY. INTERNATIONAL JOURNAL OF COMPUTER ASSISTED RADIOLOGY AND SURGERY, 13(12), 1959–1970. HTTPS://DOI.ORG/10.1007/S11548-018-1860-1

[28] LOUKAS, C., VARYTIMIDIS, C., RAPANTZIKOS, K., & KANAKIS, M. A. (2018). KEYFRAME EXTRACTION FROM LAPAROSCOPIC VIDEOS BASED ON VISUAL SALIENCY DETECTION. COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE, 165, 13–23. HTTPS://DOI.ORG/10.1016/J.CMPB.2018.07.004

[29] PRIMUS, M. J., SCHOEFFMANN, K., & BÖSZÖRMENYI, L. (2016, JUNE). TEMPORAL SEGMENTATION OF LAPAROSCOPIC VIDEOS INTO SURGICAL PHASES. IN 2016 14TH INTERNATIONAL WORKSHOP ON CONTENT-BASED MULTIMEDIA INDEXING (CBMI) (PP. 1-6). IEEE.

[30] N. PADOY, \MACHINE AND DEEP LEARNING FOR WORK OW RECOGNITION DURING SURGERY," MINIMALLY INVASIVE THERAPY AND ALLIED TECHNOLOGIES, VOL. 28, NO. 2, PP. 82{90, 2019.

[31] PADOY N. (2019). MACHINE AND DEEP LEARNING FOR WORKFLOW RECOGNITION DURING SURGERY. MINIMALLY INVASIVE THERAPY & ALLIED TECHNOLOGIES : MITAT : OFFICIAL

JOURNAL OF THE SOCIETY FOR MINIMALLY INVASIVE THERAPY, 28(2), 82–90. HTTPS://DOI.ORG/10.1080/13645706.2019.1584116

[32] KUANAR, S., ATHITSOS, V., MAHAPATRA, D., RAO, K. R., AKHTAR, Z., & DASGUPTA, D. (2019, SEPTEMBER). LOW DOSE ABDOMINAL CT IMAGE RECONSTRUCTION: AN UNSUPERVISED LEARNING BASED APPROACH. IN 2019 IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING (ICIP) (PP. 1351-1355). IEEE.

[33] KUANAR, S., ATHITSOS, V., PRADHAN, N., MISHRA, A., & RAO, K. R. (2018, APRIL). COGNITIVE ANALYSIS OF WORKING MEMORY LOAD FROM EEG, BY A DEEP RECURRENT NEURAL NETWORK. IN 2018 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING (ICASSP) (PP. 2576-2580). IEEE.

[34] SIMONYAN, K., & ZISSERMAN, A. (2014). VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION. ARXIV PREPRINT ARXIV:1409.1556

[35] SZEGEDY, C., LIU, W., JIA, Y., SERMANET, P., REED, S., ANGUELOV, D., ... & RABINOVICH, A. (2015). GOING DEEPER WITH CONVOLUTIONS. IN PROCEEDINGS OF THE IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (PP. 1-9).

[36] HE, K., ZHANG, X., REN, S., & SUN, J. (2016). DEEP RESIDUAL LEARNING FOR IMAGE RECOGNITION. IN PROCEEDINGS OF THE IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (PP. 770-778)

[37] HUANG, G., LIU, Z., VAN DER MAATEN, L., & WEINBERGER, K. Q. (2017). DENSELY CONNECTED CONVOLUTIONAL NETWORKS. IN PROCEEDINGS OF THE IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (PP. 4700-4708).

[38] WANG, J., SUN, K., CHENG, T., JIANG, B., DENG, C., ZHAO, Y., ... & XIAO, B. (2020). DEEP HIGH-RESOLUTION REPRESENTATION LEARNING FOR VISUAL RECOGNITION. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, 43(10), 3349-3364.

[39] TAN, M., & LE, Q. (2019, MAY). EFFICIENTNET: RETHINKING MODEL SCALING FOR CONVOLUTIONAL NEURAL NETWORKS. IN INTERNATIONAL CONFERENCE ON MACHINE LEARNING (PP. 6105-6114). PMLR.

[40] HU, H., ZHANG, Z., XIE, Z., & LIN, S. (2019). LOCAL RELATION NETWORKS FOR IMAGE RECOGNITION. IN PROCEEDINGS OF THE IEEE/CVF INTERNATIONAL CONFERENCE ON COMPUTER VISION (PP. 3464-3473).

[41] DOSOVITSKIY, A., BEYER, L., KOLESNIKOV, A., WEISSENBORN, D., ZHAI, X., UNTERTHINER, T., ... & HOULSBY, N. (2020). AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE. ARXIV PREPRINT ARXIV:2010.11929.

[42] BEAL, J., KIM, E., TZENG, E., PARK, D. H., ZHAI, A., & KISLYUK, D. (2020). TOWARD TRANSFORMER-BASED OBJECT DETECTION. ARXIV PREPRINT ARXIV:2012.09958.

[43] ZHENG, S., LU, J., ZHAO, H., ZHU, X., LUO, Z., WANG, Y., ... & ZHANG, L. (2021). RETHINKING SEMANTIC SEGMENTATION FROM A SEQUENCE-TO-SEQUENCE PERSPECTIVE WITH

TRANSFORMERS. IN PROCEEDINGS OF THE IEEE/CVF CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (PP. 6881-6890).

[44] LIU, Z., LIN, Y., CAO, Y., HU, H., WEI, Y., ZHANG, Z., ... & GUO, B. (2021). SWIN TRANSFORMER: HIERARCHICAL VISION TRANSFORMER USING SHIFTED WINDOWS. ARXIV PREPRINT ARXIV:2103.14030.

[45] HE, K., GKIOXARI, G., DOLLÁR, P., & GIRSHICK, R. (2017). MASK R-CNN. IN PROCEEDINGS OF THE IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION (PP. 2961-2969).

[46] BAHETI, B., INNANI, S., GAJRE, S., & TALBAR, S. (2020). EFF-UNET: A NOVEL ARCHITECTURE FOR SEMANTIC SEGMENTATION IN UNSTRUCTURED ENVIRONMENT. IN PROCEEDINGS OF THE IEEE/CVF CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION WORKSHOPS (PP. 358-359).

[47] ZHANG, H., ZHANG, J., YANG, G., SPINCEMAILLE, P., NGUYEN, T. D., & WANG, Y. (2021). MEMORY U-NET: MEMORIZING WHERE TO VOTE FOR LESION INSTANCE SEGMENTATION.

[48] LEE, K.; SUNWOO, L.; KIM, T.; LEE, K.J. SPIDER U-NET: INCORPORATING INTER-SLICE CONNECTIVITY USING LSTM FOR 3D BLOOD VESSEL SEGMENTATION. APPL. SCI. 2021, 11, 2014. HTTPS://DOI.ORG/10.3390/ APP11052014.

[49] KONOPCZYŃSKI, T., HEIMAN, R., WOŹNICKI, P., GNIEWEK, P., DUVERNOY, M. C., HALLATSCHEK, O., & HESSER, J. (2020, OCTOBER). INSTANCE SEGMENTATION OF DENSELY PACKED CELLS USING A HYBRID MODEL OF U-NET AND MASK R-CNN. IN INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND SOFT COMPUTING (PP. 626-635). SPRINGER, CHAM.

[50] DOGAN, R. O., DOGAN, H., BAYRAK, C., & KAYIKCIOGLU, T. (2021). A TWO-PHASE APPROACH USING MASK R-CNN AND 3D U-NET FOR HIGH-ACCURACY AUTOMATIC SEGMENTATION OF PANCREAS IN CT IMAGING. COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE, 207, 106141.

[51] LI, X., CHEN, H., QI, X., DOU, Q., FU, C. W., & HENG, P. A. (2018). H-DENSEUNET: HYBRID DENSELY CONNECTED UNET FOR LIVER AND TUMOR SEGMENTATION FROM CT VOLUMES. IEEE TRANSACTIONS ON MEDICAL IMAGING, 37(12), 2663-2674.

[52] ALOM, M. Z., HASAN, M., YAKOPCIC, C., TAHA, T. M., & ASARI, V. K. (2018). RECURRENT RESIDUAL CONVOLUTIONAL NEURAL NETWORK BASED ON U-NET (R2U-NET) FOR MEDICAL IMAGE SEGMENTATION. ARXIV PREPRINT ARXIV:1802.06955.

[53] DIAKOGIANNIS, F.I.; WALDNER, F.; CACCETTA, P.; WU, C. RESUNET-A: A DEEP LEARNING FRAMEWORK FOR SEMANTIC SEGMENTATION OF REMOTELY SENSED DATA. ISPRS J. PHOTOGRAMM. REMOTE SENS. 2020, 162, 94–114. [CROSSREF]

[54] ZHOU, Z., SIDDIQUEE, M. M. R., TAJBAKHSH, N., & LIANG, J. (2019). UNET++: REDESIGNING SKIP CONNECTIONS TO EXPLOIT MULTISCALE FEATURES IN IMAGE SEGMENTATION. IEEE TRANSACTIONS ON MEDICAL IMAGING, 39(6), 1856-1867.

[55] LI, C., TAN, Y., CHEN, W., LUO, X., HE, Y., GAO, Y., & LI, F. (2020). ANU-NET: ATTENTION-BASED NESTED U-NET TO EXPLOIT FULL RESOLUTION FEATURES FOR MEDICAL IMAGE SEGMENTATION. COMPUTERS & GRAPHICS, 90, 11-20.

[56] HUANG, H., LIN, L., TONG, R., HU, H., ZHANG, Q., IWAMOTO, Y., ... & WU, J. (2020, MAY). UNET 3+: A FULL-SCALE CONNECTED UNET FOR MEDICAL IMAGE SEGMENTATION. IN ICASSP 2020-2020 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING (ICASSP) (PP. 1055-1059). IEEE.

[57] YU, E. M., IGLESIAS, J. E., DALCA, A. V., & SABUNCU, M. R. (2020). AN AUTO-ENCODER STRATEGY FOR ADAPTIVE IMAGE SEGMENTATION. ARXIV PREPRINT ARXIV:2004.13903.

[58] NAMAZI, B., SANKARANARAYANAN, G., DEVARAJAN, V. & FLESHMAN, J. A DEEP LEARNING SYSTEM FOR AUTOMATICALLY IDENTIFYING CRITICAL VIEW OF SAFETY IN LAPAROSCOPIC CHOLECYSTECTOMY VIDEOS FOR ASSESSMENT. IN SAGES 2017 ANNUAL MEETING (SAGES, HOUSTON, TX, 2017).

[59] NAMAZI, B., PHD. EE .(2019) TOWARDS AUTOMATED UNDERSTANDING OF LAPAROSCOPIC VIDEOS, HTTP://HDL.HANDLE.NET/10106/28603.

[60] KITAMURA, H., FUJIOKA, S., HATA, T., MISAWA, T., & YANAGA, K. (2019). SEGMENT IV APPROACH FOR DIFFICULT LAPAROSCOPIC CHOLECYSTECTOMY. ANNALS OF GASTROENTEROLOGICAL SURGERY, 4(2), 170–174. HTTPS://DOI.ORG/10.1002/AGS3.12297

[61] ALTIERI, M., HASHIMOTO, D., RIVERA, A. M., NAMAZI, B., ALSEIDI, A., OKRAINEC, A., ... & MADANI, A. (2020). USING ARTIFICIAL INTELLIGENCE TO IDENTIFY SURGICAL ANATOMY, SAFE ZONES OF DISSECTION, AND DANGEROUS ZONES OF DISSECTION DURING LAPAROSCOPIC CHOLECYSTECTOMY. JOURNAL OF THE AMERICAN COLLEGE OF SURGEONS, 231(4), E21-E22.

[62] GERKEMA, M. H. (2020). DEEP LEARNING FOR IDENTIFICATION OF GALLBLADDER LEAKAGE DURING LAPAROSCOPIC CHOLECYSTECTOMY (MASTER'S THESIS, UNIVERSITY OF TWENTE).

[63] STEINHILBER, B., HOFFMANN, S., KARLOVIC, K., PFEFFER, S., MAIER, T., HALLASHEH, O., ... & SIEVERT, K. D. (2015). DEVELOPMENT OF AN ARM SUPPORT SYSTEM TO IMPROVE ERGONOMICS IN LAPAROSCOPIC SURGERY: STUDY DESIGN AND PROVISIONAL RESULTS. SURGICAL ENDOSCOPY, 29(9), 2851-2858.

[64] SCHEIKL, P., LASCHEWSKI, S., KISILENKO, A., DAVITASHVILI, T., MÜLLER, B., CAPEK, M., MÜLLER-STICH, B., WAGNER, M. & MATHIS-ULLRICH, F. (2020). DEEP LEARNING FOR SEMANTIC SEGMENTATION OF ORGANS AND TISSUES IN LAPAROSCOPIC SURGERY. CURRENT DIRECTIONS IN BIOMEDICAL ENGINEERING, 6(1), 20200016. HTTPS://DOI.ORG/10.1515/CDBME-2020-0016

[65] MASCAGNI, P., VARDAZARYAN, A., ALAPATT, D., URADE, T., EMRE, T., FIORILLO, C., ... & PADOY, N. (2022). ARTIFICIAL INTELLIGENCE FOR SURGICAL SAFETY: AUTOMATIC ASSESSMENT OF THE CRITICAL VIEW OF SAFETY IN LAPAROSCOPIC CHOLECYSTECTOMY USING DEEP LEARNING. ANNALS OF SURGERY, 275(5), 955-961.

[66] NAMAZI, B., IYENGAR, N., HASAN, S., BALACHANDRA, S., MADANI, A., HASHIMOTO, D., ... & SANKARANARAYANAN, G. (2020). AI FOR AUTOMATED DETECTION OF THE ESTABLISHMENT OF

CRITICAL VIEW OF SAFETY IN LAPAROSCOPIC CHOLECYSTECTOMY VIDEOS. JOURNAL OF THE AMERICAN COLLEGE OF SURGEONS, 231(4), E48.

[67] ZEILER, M. D., & FERGUS, R. (2014, SEPTEMBER). VISUALIZING AND UNDERSTANDING CONVOLUTIONAL NETWORKS. IN EUROPEAN CONFERENCE ON COMPUTER VISION (PP. 818-833). SPRINGER, CHAM.

[68] BISHOP, C. M., & NASRABADI, N. M. (2006). PATTERN RECOGNITION AND MACHINE LEARNING (VOL. 4, NO. 4, P. 738). NEW YORK: SPRINGER.

[69] ELLACOTT, S. W. (1994). ASPECTS OF THE NUMERICAL ANALYSIS OF NEURAL NETWORKS. ACTA NUMERICA, 3, 145-202.

[70] RUMELHART, D. E., HINTON, G. E., & WILLIAMS, R. J. (1986). LEARNING REPRESENTATIONS BY BACK-PROPAGATING ERRORS. NATURE, 323(6088), 533-536.

[71] QIAN, N. (1999). ON THE MOMENTUM TERM IN GRADIENT DESCENT LEARNING ALGORITHMS. NEURAL NETWORKS, 12(1), 145-151.

[72] DUCHI, J., HAZAN, E., & SINGER, Y. (2011). ADAPTIVE SUBGRADIENT METHODS FOR ONLINE LEARNING AND STOCHASTIC OPTIMIZATION. JOURNAL OF MACHINE LEARNING RESEARCH, 12(7).

[73] ZEILER, M. D. (2012). ADADELTA: AN ADAPTIVE LEARNING RATE METHOD. ARXIV PREPRINT ARXIV:1212.5701.

[74] KINGMA, D. P., & BA, J. (2014). ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION. ARXIV PREPRINT ARXIV:1412.6980

[75] DOZAT, T. (2016). INCORPORATING NESTEROV MOMENTUM INTO ADAM.

[76] OFFE, S., & SZEGEDY, C. (2015, JUNE). BATCH NORMALIZATION: ACCELERATING DEEP NETWORK TRAINING BY REDUCING INTERNAL COVARIATE SHIFT. IN INTERNATIONAL CONFERENCE ON MACHINE LEARNING (PP. 448-456). PMLR.

[77] SKOROKHODOV, I., & BURTSEV, M. (2019). LOSS LANDSCAPE SIGHTSEEING WITH MULTI-POINT OPTIMIZATION. ARXIV PREPRINT ARXIV:1910.03867.

[78] SALIMANS, T., & KINGMA, D. P. (2016). WEIGHT NORMALIZATION: A SIMPLE REPARAMETERIZATION TO ACCELERATE TRAINING OF DEEP NEURAL NETWORKS. ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS, 29.

[79] BA, J. L., KIROS, J. R., & HINTON, G. E. (2016). LAYER NORMALIZATION. ARXIV PREPRINT ARXIV:1607.06450.

[80] ULYANOV, D., VEDALDI, A., & LEMPITSKY, V. (2016). INSTANCE NORMALIZATION: THE MISSING INGREDIENT FOR FAST STYLIZATION. ARXIV PREPRINT ARXIV:1607.08022.

[81] WU, Y., & HE, K. (2018). GROUP NORMALIZATION. IN PROCEEDINGS OF THE EUROPEAN CONFERENCE ON COMPUTER VISION (ECCV) (PP. 3-19).

[82] HEATON, J. (2018). IAN GOODFELLOW, YOSHUA BENGIO, AND AARON COURVILLE: DEEP LEARNING.

[83] TORRES, T. V. (2018). FIRST CONTACT WITH DEEP LEARNING: PRACTICAL INTRODUCTION WITH KERAS. KINDLE DIRECT PUBLISHING.

[84] RUSSAKOVSKY, O., DENG, J., SU, H., KRAUSE, J., SATHEESH, S., MA, S., HUANG, Z., KARPATHY, A., KHOSLA, A., BERNSTEIN, M.S., BERG, A.C., & FEI-FEI, L. (2015). IMAGENET LARGE SCALE VISUAL RECOGNITION CHALLENGE. INTERNATIONAL JOURNAL OF COMPUTER VISION, 115, 211-252.

[85] KRIZHEVSKY, A., SUTSKEVER, I., & HINTON, G. E. (2017). IMAGENET CLASSIFICATION WITH DEEP CONVOLUTIONAL NEURAL NETWORKS. COMMUNICATIONS OF THE ACM, 60(6), 84-90.

[86] GALLEGO, J., PEDRAZA, A., LOPEZ, S., STEINER, G., GONZALEZ, L., LAURINAVICIUS, A., & BUENO, G. (2018). GLOMERULUS CLASSIFICATION AND DETECTION BASED ON CONVOLUTIONAL NEURAL NETWORKS. JOURNAL OF IMAGING, 4(1), 20. HTTPS://DOI.ORG/10.3390/JIMAGING4010020

[87] LIU, D., XIONG, Y., PULLI, K., & SHAPIRO, L. (2011, AUGUST). ESTIMATING IMAGE SEGMENTATION DIFFICULTY. IN INTERNATIONAL WORKSHOP ON MACHINE LEARNING AND DATA MINING IN PATTERN RECOGNITION (PP. 484-495). SPRINGER, BERLIN, HEIDELBERG.

[88] LIU, X., SONG, L., LIU, S., & ZHANG, Y. (2021). A REVIEW OF DEEP-LEARNING-BASED MEDICAL IMAGE SEGMENTATION METHODS. SUSTAINABILITY, 13(3), 1224.

[89] XU, A., WANG, L., FENG, S., & QU, Y. (2010, NOVEMBER). THRESHOLD-BASED LEVEL SET METHOD OF IMAGE SEGMENTATION. IN 2010 THIRD INTERNATIONAL CONFERENCE ON INTELLIGENT NETWORKS AND INTELLIGENT SYSTEMS (PP. 703-706). IEEE.

[90] YU-QIAN, Z., WEI-HUA, G., ZHEN-CHENG, C., JING-TIAN, T., & LING-YUN, L. (2006, JANUARY). MEDICAL IMAGES EDGE DETECTION BASED ON MATHEMATICAL MORPHOLOGY. IN 2005 IEEE ENGINEERING IN MEDICINE AND BIOLOGY 27TH ANNUAL CONFERENCE (PP. 6492-6495). IEEE.

[91] CIGLA, C., & ALATAN, A. A. (2008, OCTOBER). REGION-BASED IMAGE SEGMENTATION VIA GRAPH CUTS. IN 2008 15TH IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING (PP. 2272-2275). IEEE.

[92] LIN, G., MILAN, A., SHEN, C., & REID, I. (2017). REFINENET: MULTI-PATH REFINEMENT NETWORKS FOR HIGH-RESOLUTION SEMANTIC SEGMENTATION. IN PROCEEDINGS OF THE IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (PP. 1925-1934).

[93] NOH, H., HONG, S., & HAN, B. (2015). LEARNING DECONVOLUTION NETWORK FOR SEMANTIC SEGMENTATION. IN PROCEEDINGS OF THE IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION (PP. 1520-1528).

[94] WU, L. (2019). BIOMEDICAL IMAGE SEGMENTATION AND OBJECT DETECTION USING DEEP CONVOLUTIONAL NEURAL NETWORKS.

[95] SINHA, A., & DOLZ, J. MULTI-SCALE GUIDED ATTENTION FOR MEDICAL IMAGE SEGMENTATION. ARXIV 2019. ARXIV PREPRINT ARXIV:1906.02849.

[96] LIU, L., WOLTERINK, J. M., BRUNE, C., & VELDHUIS, R. N. (2021). ANATOMY-AIDED DEEP LEARNING FOR MEDICAL IMAGE SEGMENTATION: A REVIEW. PHYSICS IN MEDICINE & BIOLOGY, 66(11), 11TR01.

[97] HTTPS://TOWARDSDATASCIENCE.COM/AUTO-ENCODER-WHAT-IS-IT-AND-WHAT-IS-IT-USED-FOR-PART-1-3E5C6F017726.

[98] GOOGLERESEARCH, "TENSORFLOW: LARGE-SCALE MACHINE LEARNING ON HETEROGENEOUS SYSTEMS," GOOGLE RES., 2015.

[99] HTTP://CAMMA.U-STRASBG.FR/DATASETS.

[100] HTTPS://ENDOVIS.GRAND-CHALLENGE.ORG.

[101] TWINANDA, A. P., SHEHATA, S., MUTTER, D., MARESCAUX, J., DE MATHELIN, M., & PADOY, N. (2017). ENDONET: A DEEP ARCHITECTURE FOR RECOGNITION TASKS ON LAPAROSCOPIC VIDEOS. IEEE TRANSACTIONS ON MEDICAL IMAGING, 36(1), 86–97. HTTPS://DOI.ORG/10.1109/TMI.2016.2593957

[102] R. STAUDER, D. OSTLER, M. KRANZFELDER, S. KOLLER, H. FEUSSNER, AND N. NAVAB, \THE TUM LAPCHOLE DATASET FOR THE M2CAI 2016 WORK FOW CHALLENGE," ARXIV, VOL. ABS/1610.09278, 10 2016.

[103] WARD, T. M., FER, D. M., BAN, Y., ROSMAN, G., MEIRELES, O. R., & HASHIMOTO, D. A. (2021). CHALLENGES IN SURGICAL VIDEO ANNOTATION. COMPUTER ASSISTED SURGERY, 26(1), 58-68.

[104] PEREZ, L., & WANG, J. (2017). THE EFFECTIVENESS OF DATA AUGMENTATION IN IMAGE CLASSIFICATION USING DEEP LEARNING. ARXIV PREPRINT ARXIV:1712.04621.

[105] KERAS IMAGE PREPROCESSING. AVAILABLE AT: MARCH 2015; ACCESSED 3 NOVEMBER 2021.

[106] JADON, S. (2020, OCTOBER). A SURVEY OF LOSS FUNCTIONS FOR SEMANTIC SEGMENTATION. IN 2020 IEEE CONFERENCE ON COMPUTATIONAL INTELLIGENCE IN BIOINFORMATICS AND COMPUTATIONAL BIOLOGY (CIBCB) (PP. 1-7). IEEE.

[107] SONG, Q. C., TANG, C., & WEE, S. (2021). MAKING SENSE OF MODEL GENERALIZABILITY: A TUTORIAL ON CROSS-VALIDATION IN R AND SHINY. ADVANCES IN METHODS AND PRACTICES IN PSYCHOLOGICAL SCIENCE, 4(1), 2515245920947067.

[108] NAVEZ, B., UNGUREANU, F., MICHIELS, M., CLAEYS, D., MUYSOMS, F., HUBERT, C., ... & GIGOT, J. F. (2012). SURGICAL MANAGEMENT OF ACUTE CHOLECYSTITIS: RESULTS OF A 2-YEAR PROSPECTIVE MULTICENTER SURVEY IN BELGIUM. SURGICAL ENDOSCOPY, 26(9), 2436-2445.

[109] SOPER, N. J., STOCKMANN, P. T., DUNNEGAN, D. L., & ASHLEY, S. W. (1992). LAPAROSCOPIC CHOLECYSTECTOMY THE NEW'GOLD STANDARD'?. ARCHIVES OF SURGERY, 127(8), 917-923.

[110] SAGES. THE SAGES SAFE CHOLECYSTECTOMY PROGRAM. AVAILABLE AT: HTTPS ://WWW.SAGES .ORG/SAFE-CHOLE CYSTECTOMY-PROGRAM/. SEPTEMBER 2014;

[111] SGARAMELLA, L. I., GURRADO, A., PASCULLI, A., DE ANGELIS, N., MEMEO, R., PRETE, F. P., ... & TESTINI, M. (2021). THE CRITICAL VIEW OF SAFETY DURING LAPAROSCOPIC CHOLECYSTECTOMY: STRASBERG YES OR NO? AN ITALIAN MULTICENTRE STUDY. SURGICAL ENDOSCOPY, 35(7), 3698-3708.

[112] HTTPS://COLAB.RESEARCH.GOOGLE.COM/SIGNUP\

[113] HTTP://CAMMA.U-STRASBG.FR/DATASETS

[114] HTTPS://WWW.LAPAROSCOPYHOSPITAL.COM/ONLINE_LAPAROSCOPIC_VIDEOS.HTML

[115] HTTPS://WWW.YOUTUBE.COM/C/SAGESVIDEO/FEATURED

[1156] PARIHAR, A. S., GUPTA, Y. K., SINGODIA, Y., SINGH, V., & SINGH, K. (2020, JUNE). A COMPARATIVE STUDY OF IMAGE DEHAZING ALGORITHMS. IN 2020 5TH INTERNATIONAL CONFERENCE ON COMMUNICATION AND ELECTRONICS SYSTEMS (ICCES) (PP. 766-771). IEEE.

[117] HTTPS://MED.WMICH.EDU/NODE/66

[118] MASCAGNI, P., ALAPATT, D., GARCIA, A., OKAMOTO, N., VARDAZARYAN, A., COSTAMAGNA, G., ... & PADOY, N. (2021). SURGICAL DATA SCIENCE FOR SAFE CHOLECYSTECTOMY: A PROTOCOL FOR SEGMENTATION OF HEPATOCYSTIC ANATOMY AND ASSESSMENT OF THE CRITICAL VIEW OF SAFETY. ARXIV PREPRINT ARXIV:2106.10916.

[119] THOMAS, L., SCHAEFER, F., & GEHRIG, J. (2021). FIJI PLUGINS FOR QUALITATIVE IMAGE ANNOTATIONS: ROUTINE ANALYSIS AND APPLICATION TO IMAGE CLASSIFICATION. F1000RESEARCH, 9, 1248. HTTPS://DOI.ORG/10.12688/F1000RESEARCH.26872.2

[120] HTTPS://IMAGE-NET.ORG/UPDATE-MAR-11-2021.PHP.

[121] HTTPS://THINKLIKEASURGEON.CA/

[122] BREHERET A. PIXEL ANNOTATION TOOL 2017. AVAILABLE AT: HTTPS://GITHUB.COM/

ABREHERET/PIXELANNOTATIONTOOL.

[123] HTTPS://WWW.IRCAD.FR/

[124] JING, L., & TIAN, Y. (2020). SELF-SUPERVISED VISUAL FEATURE LEARNING WITH DEEP NEURAL NETWORKS: A SURVEY. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, 43(11), 4037-4058.

[125] HINTON, G. E., MCCLELLAND, J. L., & RUMELHART, D. E. (1986). PARALLEL DISTRIBUTED PROCESSING: EXPLORATIONS IN THE MICROSTRUCTURE OF COGNITION, VOL. 1. DISTRIBUTED REPRESENTATIONS, 77-109.

[126] FERREIRA, M.F., CAMACHO, R. & TEIXEIRA, L.F. USING AUTOENCODERS AS A WEIGHT INITIALIZATION METHOD ON DEEP NEURAL NETWORKS FOR DISEASE DETECTION. BMC MED INFORM DECIS MAK 20 (SUPPL 5), 141 (2020). HTTPS://DOI.ORG/10.1186/S12911-020-01150-W

[127] KOLOUD. N. A, JANOSE. L.G, SAAD. S. AND IKHLAS. A., " TOWARDS RELIABLE HEPATOCYTIC ANATOMY SEGMENTATION IN LAPAROSCOPIC CHOLECYSTECTOMY USING U-NET WITH AUTO-ENCODER ", SURG ENDOSC (2023). HTTPS://DOI.ORG/10.1007/S00464-023-10306-4.