CHAPTER 5

ENDOSCOPIC ARTEFACT SEGMENTATION

5.1 INTRODUCTION

Endoscopic artefacts are irregular in shape. The bounding box overestimates the majority of them because they are non-rectangular. Thus, an efficient segmentation algorithm could accurately segment the boundaries of the artefact. Effectively outlining all possible artefacts will be a preliminary step for restoration. Therefore, if an artefact is detected, the edges of the artefact must be segmented and the segmented region alone can be considered for restoration. To achieve this, an efficient algorithm is required. In practice, lots of traditional and AI-based algorithms are available. Identifying the suitable algorithm for segmenting various artefacts is only a trial-and-error process. Considering standard algorithms, a few works best when the difference in pixel values between the target and background is large. For example, threshold-based segmentation algorithms work fine for artefacts such as, specular reflection and saturation. Few regions affected by artefacts such as, bubbles, hold almost the same value as the background pixel so traditional algorithm cannot solve the purpose.

Similarly, the imaging artefact has vast variations and few images are blurred with streak of light. Few are affected by chromatic aberration. Since the artefact are different the same algorithm cannot be used to segment every artefact. Again, selecting parameters for traditional algorithms like threshold value and seed pixel is laborious. Therefore, traditional algorithms are not much preferred.

AI-based algorithms take lead roles in such cases. When AI-based algorithms, especially DL algorithms are used, a huge dataset with images constituting all artefacts could eventually help for better training and segmentation (S. Ahuja,2019). Thus, various traditional and DL -based algorithms for segmentation of multiple artefacts are experimented.

5.2 SIMULATION SETUP

5.2.1. Dataset

EAD 2020 dataset is a public dataset for endoscopic artefact segmentation. It holds annotations for detection and generalization. It contains 643 images for segmentation with a

binary mask for five artefacts: saturation, specular reflections, bubbles, instruments and miscellaneous artefacts. The dataset covers organs such as, the esophagus, stomach, duodenum, liver and colon. The organs are imaged using standard endoscopes in various imaging modalities like NBI, white light and AFI. The dataset embraces images of the patient from countries like the United Kingdom, Russia and France. The dataset is available for public use.

5.2.2 Data Augmentation

An image in the segmentation dataset may contain an instance of an artefact or many instances of various artefacts. Therefore, a separate binary mask is given for each artefact present in the image. Thus, each image in the dataset holds five binary masks. If the image is affected by saturation and specular reflections, the dataset has a mask only for those two artefacts, and the remaining masks are left blank. The dataset is divided into 80% for training and 20% for testing. After partitioning, 515 images are available for training, which is considered too low for a DL -based algorithm to train itself; hence data augmentation techniques are deployed. Augmentation techniques such as, random crops (https://blog.roboflow.com), image blurring, image sharpening, flipping, Gaussian noise and colour manipulation (https://neptune.ai) are deployed to boost the size of the training dataset.

- ***Random Crops*:** Training images from the dataset are randomly cropped to the size of 320 * 320.
- ***Flipping**:* Horizontal flip with 0.5 as probability is used to flip.
- ***Random Gaussian noise***: Gaussian noise is added to the image with probability as 0.2.
- ***Colour manipulation***: This function will initialize change in contrast or Hue Saturation Value (HSV). The probability is set to 1. Thus, every image will undergo any one of the augmentations mentioned.
- ***Image blurring and sharpening***: This function either blurs or sharpen the image. The probability is equal to 1. So, either one of the operations will happen at random.

The augmentation techniques are inherited from the albumentations library (https://albumentations.ai/). The albumentations library consists of many other augmentation

techniques such as, compose, Contrast Limited Adaptive Histogram Equalization (CLAHE), grid distortion, elastic transform, center crop, IAAadditive Gaussian noise, and optical distortion. Selective augmentation techniques are chosen for expanding the dataset.

5.2.3 System Configuration and Programming

The traditional segmentation techniques such as, Otsu, gray threshold, multi threshold, adaptive threshold, active contour, lazy snapping and super pixel are coded using python. The algorithms are executed and tested using system powered by 1.3GHz dual-core intel i5 processor with 4 GB DDR3 RAM and Intel HD Graphics 5000 series card. The code for DL based networks are partially extracted from segmentation models Application Programming Interface (API) (https://segmentation-models.readthedocs.io/en/latest/api.html). The data augmentation techniques are chosen from albumentations library. The complete program is coded using python in Jupiter notebook. The network training and testing is done using google co-lab.

5.3 TRADITIONAL ALGORITHMS FOR ARTEFACT SEGMENTATION

Image segmentation algorithms range from simple threshold-based to DL-based segmentation algorithms. There are vast variety of traditional algorithms, namely Otsu, adaptive thresholding, multi-threshold, gray threshold, active contour, lazy snapping, super pixel and watershed algorithm. Few of these algorithms are experimented with in this study.

In the conventional thresholding technique, a fixed threshold value is selected. The pixel values above the threshold will be made one, and those less than the threshold will be made zero. But identifying this threshold value is only a trial-and-error process. Hence it cannot be an effective technique.

Otsu Binarization: It considers the histogram of the input image. Two peaks are identified: one belongs to the background and the other to the foreground. From the histogram, a middle value can be approximately taken as the threshold. This technique suits well when the image is bimodal, i.e., it has two peaks in its histogram. When there are more than two, the technique may not be accurate for implementation.

Gray Threshold: This threshold technique converts grayscale to binary images. The algorithm chooses a specific gray value to act as a threshold.

Multi Threshold: This technique segments a gray image into several classes by selecting multiple threshold values. This technique would work fine if the image is complex and has more objects to separate from the background.

Adaptive Threshold: Adaptive threshold technique considers a gray scale or an RGB image. The output is a binary image. The adaptive thresholding technique follows either the chow and Kaneko approach or local thresholding. The common idea behind both techniques is that when a smaller region is considered, the pixel values in the smaller region will be almost uniform. Such methods are said to be computationally expensive.

Active Contour: It is a technique to obtain deformable modules present in an image using constraints and forces. It defines the object borders to generate a contour. The desired contour shape is obtained by defining the right energy function.

Lazy Snapping: It is an interactive image segmentation system. In this technique, a foreground and background mask are designated to perform the lazy snapping algorithm. It is an interactive algorithm based on graph cut. Based on the lines drawn, the algorithm determines the foreground and the background. The line drawn to define the foreground and the background need not be precise.

Super-Pixel: It divides the image into multiple segments. The algorithm works based on Simple Linear Iterative Clustering (SLIC). The algorithm divides images and groups them based on regions with similar values.

5.4 CNN ARCHITECTURE

CNN has been utilized to address a number of CV related challenges. Recently, semantic segmentation has developed to the point that it is used as an effective segmentation technique for medical images. The three-dimensional kernels in the CNN architecture enable viewing the three-dimensional structure of medical images. The first two dimensions of a picture are used to gauge its resolution, and the third dimension represented by the RGB channels is used to gauge its colour intensity. Neural networks often shrink the dimensionality of input images to speed up computation and prevent underfitting.

The network takes tensors of shape, image height x image width x 3. It is laborious, time-consuming, and computationally expensive to train a network from scratch. A sizable number of images in the dataset assisted with corresponding binary masks are required to achieve the desired accuracy. The availability of annotated datasets for specific applications like artefact segmentation is sparse. It is challenging to train a network from scratch with random initialized weights since the arithmetic involved is challenging and a lot of data is needed. Transfer learning is therefore preferred. A common CV technique used to better initialize the network weights is transfer-learning. The usage of pre-trained models is the most typical transfer learning technique. Figure 5.1 portrays the simple CNN architecture. The network has three important layers: the convolution layer, the pooling layer and the fully connected layer. Each layer is detailed in the following section:

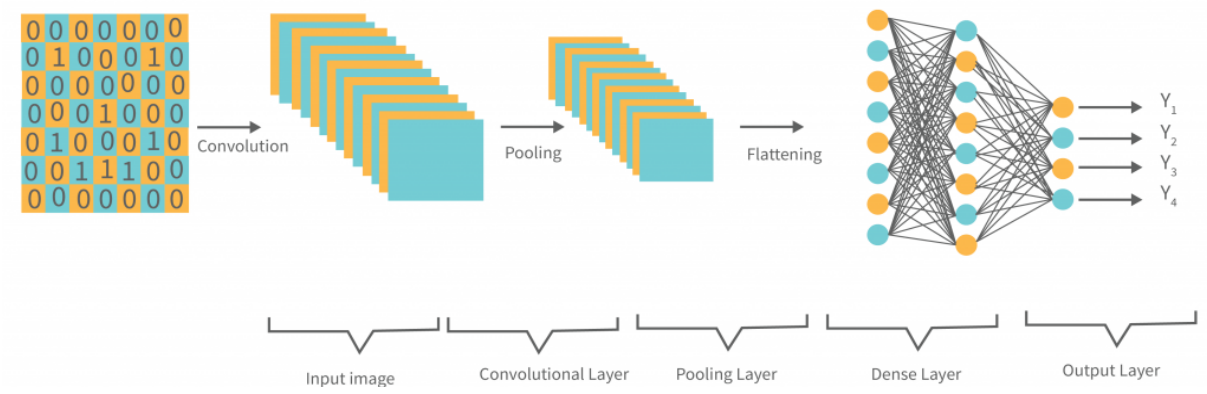

Figure 5.1 CNN Architecture (https://www.interviewbit.com)

5.4.1 Convolutional Layer

The network's convolutional operations are carried out by the convolutional kernel. It is the first and most important fundamental layer of CNN. In this layer, the convolution operation is processed by a kernel. Depending on the stride rate, the kernel adjusts both horizontally and vertically as it scans the image. Kernels are smaller than actual images yet have more depth. In this situation, the kernel's height and width will be fairly small, but its depth will encompass the image's three RGB channels. Figure 5.2 shows the convolution layer operation. Nonlinear activation functions are an important element of convolutional layers. It is customary to add an

activation layer immediately following a convolutional layer because convolution is a linear process and does not naturally include non-linearity into the activation map. The Sigmoid, Hyperbolic tangent (tanh) and ReLU are examples of frequently used non-linear procedures. ReLU has a six-fold faster convergence rate than sigmoid and tanh and it is also very dependable.

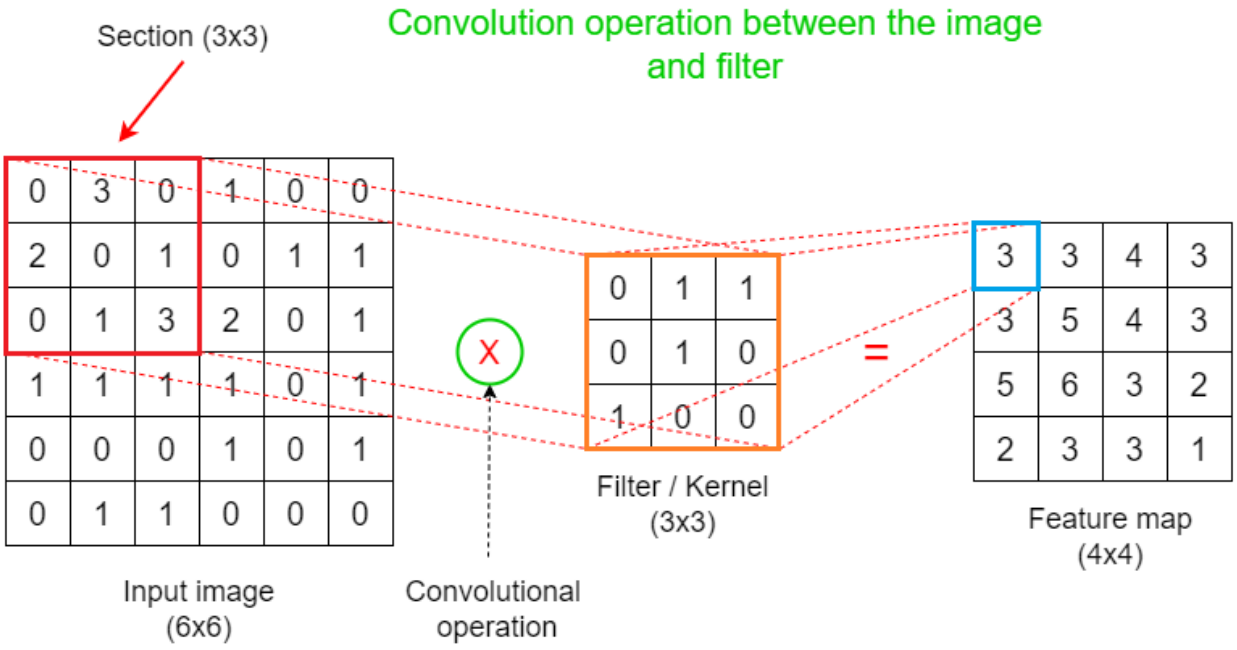

Figure 5.2 Convolutional Layers (https://towardsdatascience.com)

5.4.2 Pooling Layer

This layer minimises the number of dimensions in the image, making it simpler and requiring less memory to process the input. Additionally, pooling decreases the number of parameters, which speeds up the training. The region of the image that the kernel has covered delivers the highest value when the maximum pooling is used. The average pooling, on the other hand, determines the average each values in the area under consideration. The functioning of maximum and average pooling is depicted in Figure 5.3.

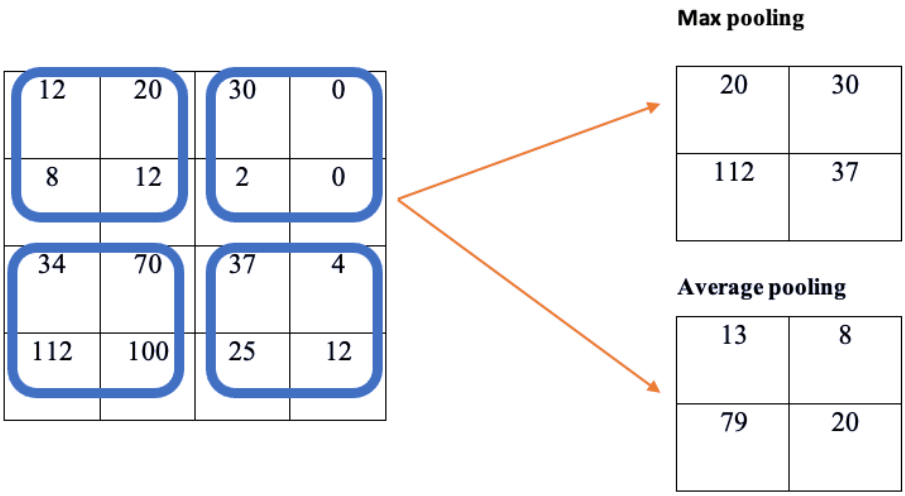

Figure 5.3 Max Pooling and Average Pooling Operation

5.4.3 Fully Connected (FC) Layer

In an FC layer, each neuron is directly connected to every other neuron in the layer beneath it. In a CNN architecture, the FC layer is frequently the last one. By increasing the number of hidden units, the network's learning capacity can be increased, but accuracy reaches a saturation point after a certain level. There is no formula for calculating the amount of concealed units because it is typically a trial and error process.

5.5 CNN BASED IMAGE SEGMENTATION ALGORITHM UNDER STUDY

To experiment the performance of segmentation algorithm in segmenting endoscopic artefact, three different algorithms are chosen as follows:

- U-Net
- FPN
- Link-Net

All the three models are adopted from segmentation model API. The API holds models along with pre-trained weights. A significant benchmark dataset is used for pre-training in order to give a helpful foundation. In order to benefit from the well-known deep CNN model, a backbone trained on the ImageNet dataset is used in place of simple

convolution blocks for feature extraction in the encoding route. The backbone is the underlying network architecture that takes the image as input and extracts the feature maps. All three models use a common EfficientNetB3 backbone pretrained on ImageNet dataset. All the networks are designed to segment a single artefact, as the comparison is to evaluate the performance of DL based algorithm against traditional algorithms. The structure of U-Net and Link-net is explained in section 5.7.1. The architecture of FPN is shown in the Figure 5.4.

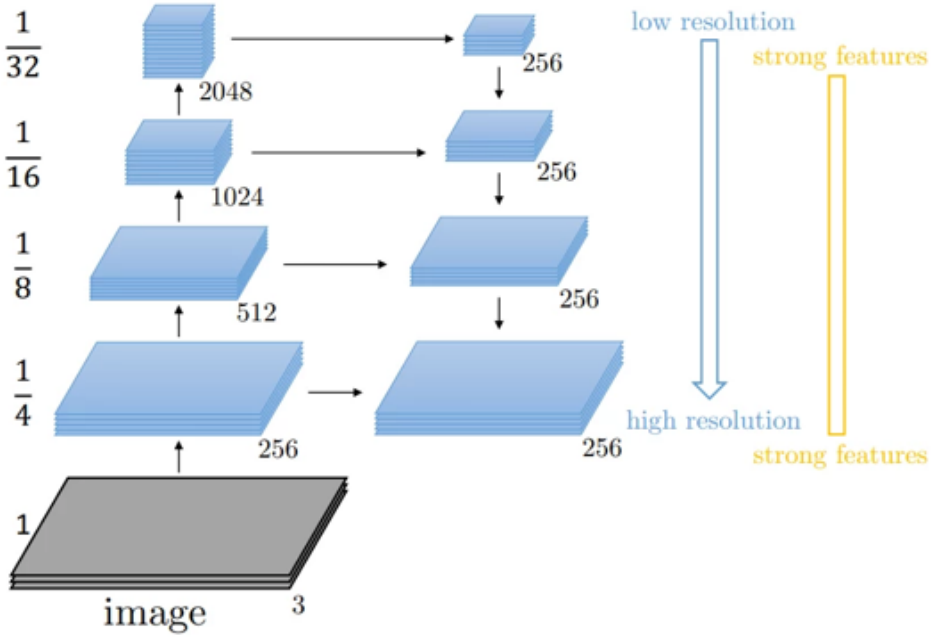

Figure 5.4 Structure of FPN (https://hasty.ai/)

Image segmentation at various scales especially at smaller scale is very tedious. In this research, artefacts such as, specular reflections and bubbles vary in scales form small to large. FPN is composed of top down and bottom up pathway. The bottom up pathway is a replica of a Convolutional Neural Network. As the stage progress in bottom up pathway the spatial resolution decreases. On the other hand in the top down approach, the features are up sampled. The feature maps extracted from the top-down pathway are up sampled and added element by element. After element wise addition the feature map again goes through a 3x3 convolution. This step essentially reduces the aliasing effect. This step is followed by another 3x3 convolution, batch

normalization and activation function. The feature map has then to be reduced based on number of classes then the map is up sampled to match the image resolution required through bilinear interpolation.

5.6 PERFORMANCE METRICS AND SIMULATION RESULTS

5.6.1 Test Set for Artefact Segmentation

Three images are chosen at random for the test set. The images are given as the input to the traditional algorithm. For DL algorithms, images from EAD segmentation dataset is chosen for testing. 20% of the dataset images are carefully handpicked for testing. This collection is pertaining to DL-based segmentation models.

5.6.2 Performance Metrics to Evaluate Traditional and DL Algorithms Under Study

- **Specificity**

  This metric is used to assess how well the trained model predicted the true negative as true negative. Equation 5.1 represents the metric Specificity.

$$Specificity = \frac{TN}{(TN+FP)} \tag{5.1}$$

- **Accuracy**

  It is the measure to check the functionality of the model across all the given classes. This metric gives the ratio of correct predictions out of total input samples. Equation 5.2 represents the metric accuracy.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{5.2}$$

- **Sensitivity**

It is the probability that how well the trained model can predict the positive samples as positive. Equation 5.3 represents the metric sensitivity.

$$Sensitivity = \frac{TP}{TP+FN} \tag{5.3}$$

- **Fβ Score**

A weighted average of precision and recall is the Fβ Score. The formula to calculate Fβ Score is given in the Equation 5.4.

$$F_\beta = (1+\beta^2)\frac{precision*recall}{\beta^2*precision+recall} \quad (5.4)$$

To calculate Fβ Score, precision and recall values are important, which can be obtained using the Equation 5.5 & 5.6. when β =1, we get the F1 score and if β=2, we get F2 score. F1 score is otherwise called as Dice score.

- **Precision**

Precision is a measure that conveys how many predicted positive pixels are actually positive. The formula is given in the Equation 5.5.

$$Precision = \frac{TP}{TP+FP} \quad (5.5)$$

- **Recall**

The recall is a measure to know how many predictions are positive out of total positives. Equation 5.6 is used to calculate recall.

$$Recall = \frac{TP}{TP+FN} \quad (5.6)$$

- **Jaccard Score**

This metric is otherwise called as IoU. IoU is calculated by considering the overlap between the predicted and the target image divided by the union of the predicted and the target image. The formula to calculate IoU is given in the Equation 5.7.

$$IoU = \frac{Area\ of\ overlap\ between\ predicted\ and\ target\ object}{Area\ of\ union\ between\ predicted\ and\ target\ object} \quad (5.7)$$

5.6.3 Simulation Results of Traditional and DL based Algorithms Under Study

To assess the need of DL based algorithm for multiple artefact segmentation, few traditional and DL algorithms are trained and tested with images from the EAD dataset. Figures 5.9 to 5.13 present the binary output of various traditional and DL-based image segmentations algorithm and their performance measures are shown in Table 5.1 to Table 5.6 for different artefacts. The binary segmented result of every model is evaluated against the annotated mask of each artefact. The performance metrics are measured and tabulated.

- **Segmentation of saturation artefact**

Artefact saturation has a property of having bright white pixels. The artefact affected region completely turns white. Thus, there will be a large variation in pixel intensities between the saturated region and the background. Hence even basic threshold-based algorithm can perform well due to its distinct features. Figure 5.5 (a) and (b) shows the original image and the corresponding mask given for saturation. Figure 5.5 (c) to (l) portrays the segmented output obtained from algorithms such as, Otsu, Gray threshold, multi threshold, adaptive threshold, active contour, lazy snapping, super pixel, U-Net, FPN and Link-Net. It is clear from the output presented in Figure 5.5 (f) that; other than adaptive threshold algorithm all other algorithms segment the artefact region with good performance scores. The performance measures are tabulated in the Table 5.1.

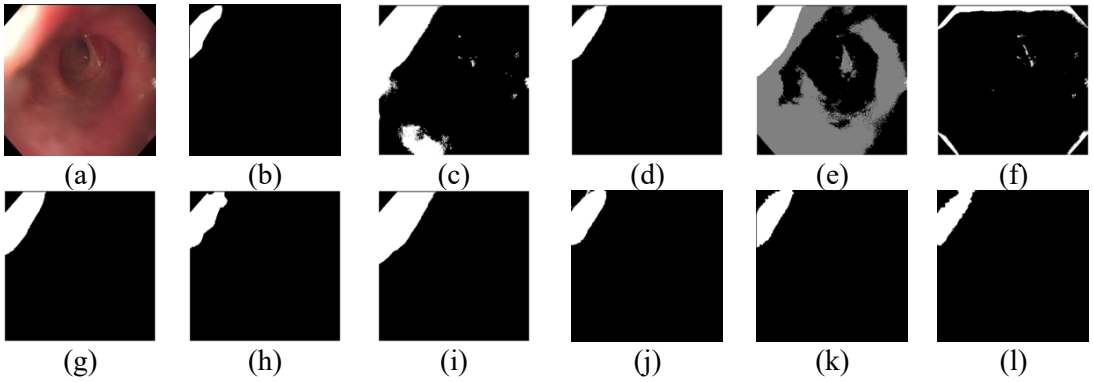

Figure 5.5 Segmentation of Saturation Artefact

(a) Original Image (b) Mask (c) Otsu (d) Gray Threshold (e) Multi Threshold (f) Adaptive Threshold (g) Active Contour (h) Lazy Snapping (i) Super Pixel (j) U-Net (k) FPN (l) Link-Net

Table 5.1 Performance Measures of Various Algorithms for the Saturation Artefact

Image	Otsu	Gray Threshold	Multi Threshold	Adaptive Threshold	Active Contour	Lazy Snapping	Super Pixel	U-Net	FPN	Link-Net
Specificity										
1	0.8504	0.9686	0.4590	0.9363	0.9508	0.9628	0.9130	0.9865	0.9863	0.9916
2	0.7981	0.9773	0.4913	0.9532	0.9543	0.9641	0.9130	0.9940	0.9896	0.9762
3	0.7763	0.9602	0.4189	0.8682	0.9119	0.8430	0.8465	0.9926	0.9779	0.9863
Accuracy										
1	0.8555	0.9694	0.4782	0.9108	0.9508	0.9639	0.9159	0.9870	0.9868	0.9915
2	0.8064	0.9765	0.5136	0.9175	0.9554	0.9650	0.9162	0.9941	0.9898	0.9770
3	0.7916	0.9546	0.4610	0.8314	0.9160	0.8542	0.8530	0.9833	0.9725	0.9789
Sensitivity										

1	0.9940	0.9903	0.9964	0.2231	0.9914	0.9934	0.9936	1.0000	0.9994	0.9883
2	0.9814	0.9586	0.9824	0.1657	0.9797	0.9847	0.9821	0.9968	0.9955	0.9936
3	0.9749	0.8874	0.9678	0.3888	0.9650	0.9896	0.9320	0.8715	0.9068	0.8903
Dice										
1	0.3295	0.6982	0.1200	0.1516	0.5901	0.6625	0.4578	0.8462	0.8437	0.8928
2	0.3152	0.7872	0.1550	0.1542	0.6662	0.7188	0.5154	0.9387	0.8990	0.7968
3	0.4176	0.7500	0.2159	0.2612	0.6377	0.5100	0.4930	0.8892	0.8346	0.8663
Jaccard										
1	0.1973	0.5363	0.0634	0.0820	0.4185	0.4954	0.2968	0.7333	0.7297	0.8064
2	0.1871	0.6490	0.0840	0.0835	0.4995	0.5610	0.3472	0.8844	0.8166	0.6623
3	0.2639	0.6000	0.1210	0.1502	0.4689	0.3422	0.3271	0.8005	0.7162	0.7642

It is evident from the simulation results that algorithms such as, gray threshold and active contour perform better, yet the performance of Link-Net is better across all means for the test image chosen. Hence for outlining the artefact saturation, Link-Net can be preferred.

- **Segmentation of the specular reflection artefact**

Specular reflections can be differentiated from background by its property of having tiny bright pixel areas. It is well known that endoscopic images are most popularly affected by specular reflections. The artefact spreads across the frames. Figure 5.6 (a) and (b) shows the input test image and its corresponding binary mask for specular reflections. Figure 5.6 (c) – (l) shows the segmented output obtained from all the traditional and DL based algorithm under discussion. Outputs of algorithms like active contour in Figure 5.6 (g) and adaptive and multi threshold in Figure 5.6 (e) and (f) seem to segment the artefact well. Other algorithms such as, Otsu, Gray threshold and super pixel given in the Figure 5.6 (c), (d), (h) and (i) segmented specular highlight present in few of the test images but not on all. It is observed that the performance of the algorithms is commendable only when there is a moderate to impressive difference in pixel intensities. The training set of DL algorithms are bagged with endoscopic images affected by specular highlights. The images are carefully chosen to ensure that the trained detector can handle all variations of the artefact. Hence the DL based trained model performed well as shown in the Figure 5.6 (j)-(l). The performance scores of various algorithms for specular reflections artefact are measured and tabulated in the Table 5.2.

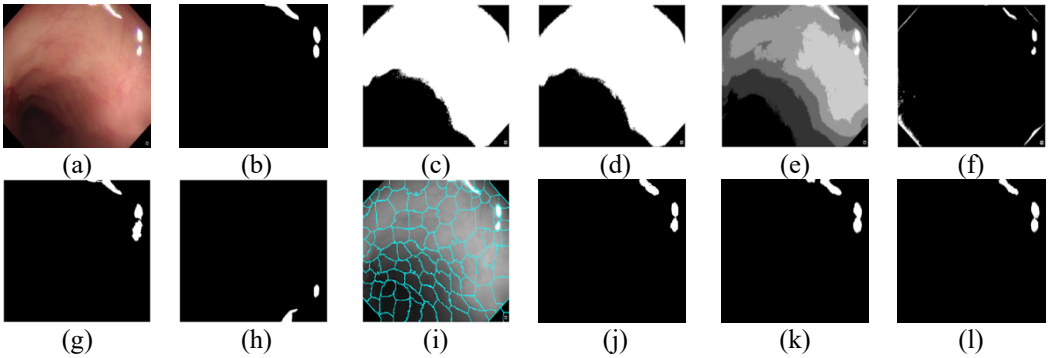

Figure 5.6 Segmentation of Specular Reflections Artefact
(a) Original Image (b) Mask (c) Otsu (d) Gray Threshold (e) Multi Threshold (f) Adaptive Threshold (g) Active Contour (h) Lazy Snapping (i) Super Pixel (j) U-Net (k) FPN (l) Link-Net

Table 5.2 Performance Measures of Various Algorithms for the Specular Reflections Artefact

Image	Otsu	Gray Threshold	Multi Threshold	Adaptive Threshold	Active Contour	Lazy Snapping	Super Pixel	U-Net	FPN	Link-Net
Specificity										
1	0.9637	0.9582	0.9583	0.9614	0.9644	0.957	0.7735	0.9621	0.9709	0.9675
2	0.3381	0.3484	0.5186	0.9748	0.9728	0.9779	0.3664	0.9965	0.9957	0.9969
3	0.4556	0.468	0.4481	0.9718	0.9743	0.9589	0.5553	0.994	0.9946	0.9926
Accuracy										
1	0.9606	0.9553	0.9554	0.9064	0.9608	0.9487	0.7811	0.9033	0.911	0.9343
2	0.3478	0.358	0.5254	0.9667	0.9713	0.9706	0.3747	0.9933	0.9947	0.9934
3	0.458	0.4704	0.4506	0.9704	0.9739	0.9583	0.557	0.994	0.9946	0.9926
Sensitivity										
1	0.9205	0.9183	0.9172	0.1942	0.9142	0.8409	0.8789	0.1231	0.1172	0.494
2	1.0000	1.0000	0.9822	0.4223	0.8691	0.4805	0.9356	0.7796	0.9257	0.7574
3	1.0000	1.0000	1.000	0.6700	0.8864	0.8256	0.9274	0.9962	0.9985	0.9863
Dice										
1	0.7704	0.7467	0.7469	0.2294	0.7699	0.7017	0.3655	0.1515	0.156	0.5134
2	0.0430	0.0437	0.0572	0.2711	0.47	0.3243	0.042	0.7745	0.8359	0.7697
3	0.0163	0.0167	0.0161	0.1693	0.2338	0.1513	0.0185	0.5896	0.6139	0.5341
Jaccard										
1	0.6624	0.5957	0.596	0.1296	0.6259	0.5405	0.2236	0.0819	0.0846	0.3453
2	0.0220	0.0223	0.0295	0.1568	0.3072	0.1935	0.0215	0.6319	0.718	0.6257
3	0.0082	0.0084	0.0081	0.0925	0.1324	0.0818	0.0093	0.418	0.4429	0.3643

The specular reflections are easy to segment due its nature of showing vast difference between the foreground and background. From the results it is proved that even a simple traditional algorithm can perform well. But at the same time not all traditional algorithms are good at segmenting the artefact. When the results of DL based algorithms are focused, all the

network segmented well. Hence few traditional algorithms and all DL algorithms can be preferred for the segmentation of specular reflections.

- **Segmentation of the instrument artefact**

Endoscopic instruments have a clear and defined boundary. Figure 5.7 (a) and (b) shows the input test image and its corresponding binary mask for the instrument artefact. Generally, all the algorithm tends to segment the artefact well. On the other hand, when the specular reflections overlap the instrument, it becomes tedious to segment the boundaries. It is evident from the results presented in the Figure 5.7 (c) - (i). With the presence of specular reflection over the instrument region, traditional algorithms give more priority to saturated regions hence output is not as expected. DL algorithms are trained with perfect ground truth masks; hence a trained segmentation algorithm segment the boundary as expected. The results of DL based algorithms are presented in Figure 5.7 (j) – (l). The accuracy of segmentation boundary is comparatively better than all the traditional algorithms. Table 5.3 lists all the performance scores measured for all three test images for the instrument artefact.

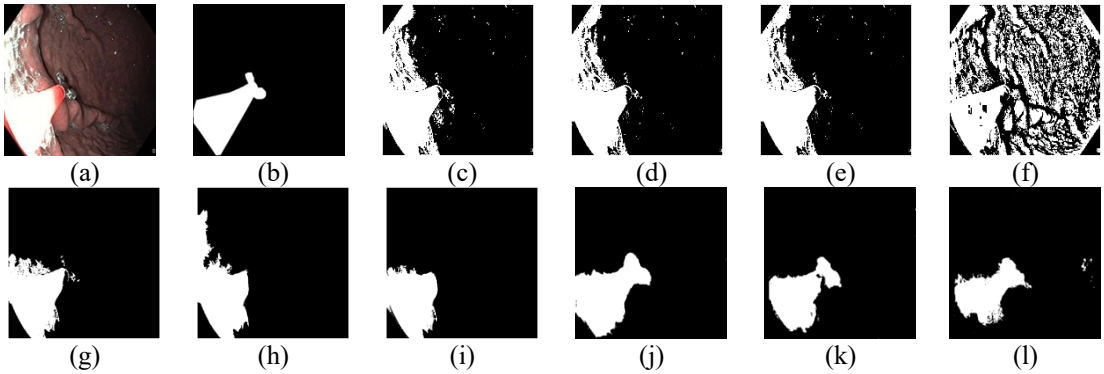

Figure 5.7 Segmentation of Instrument Artefact
(a) Original Image (b) Mask (c) Otsu (d) Gray Threshold (e) Multi Threshold (f) Adaptive Threshold (g) Active Contour(h) Lazy Snapping (i) Super Pixel (j) U-Net (k) FPN (l) Link-Net

Table 5.3 Performance Measures of Various Algorithms for the Instrument Artefact

Image	Otsu	Gray Threshold	Multi Threshold	Adaptive Threshold	Active Contour	Lazy Snapping	Super Pixel	U-Net	FPN	Link-Net
Specificity										
1	0.8251	0.8432	0.8432	0.4645	0.9285	0.9103	0.9447	0.962	0.9751	0.9763
2	0.834	0.834	0.834	0.9275	0.9705	0.7174	0.9618	0.9899	0.9919	0.9864
3	0.4921	0.4921	0.4921	0.9504	0.9766	0.7798	0.9219	0.9506	0.9798	0.9895
Accuracy										
1	0.8344	0.8481	0.8481	0.5311	0.9231	0.8997	0.9305	0.9658	0.9602	0.9515
2	0.8043	0.8043	0.8043	0.8925	0.9521	0.7298	0.938	0.9904	0.9849	0.9793
3	0.4887	0.4887	0.4887	0.9066	0.932	0.7499	0.9131	0.9385	0.973	0.9685
Sensitivity										
1	0.9069	0.8861	0.8861	0.8742	0.8809	0.8174	0.8199	0.9957	0.8449	0.759
2	0.2323	0.2323	0.2323	0.2188	0.5983	0.9687	0.4805	0.9999	0.8484	0.8437
3	0.439	0.439	0.439	0.2686	0.2827	0.3134	0.7852	0.7617	0.8749	0.6636
Dice										
1	0.5558	0.5713	0.5713	0.2901	0.7235	0.6506	0.7293	0.8694	0.8291	0.7814
2	0.1049	0.1049	0.1049	0.1673	0.5525	0.2614	0.4335	0.9116	0.8469	0.8013
3	0.0993	0.0993	0.0993	0.2697	0.3482	0.1386	0.5372	0.6139	0.8065	0.7303
Jaccard										
1	0.3849	0.3999	0.3999	0.1697	0.5668	0.4822	0.574	0.769	0.7081	0.6413
2	0.0553	0.0553	0.0553	0.0913	0.3817	0.1504	0.2767	0.8376	0.7345	0.6684
3	0.056	0.056	0.056	0.1559	0.2108	0.0745	0.3673	0.4429	0.6758	0.5752

It can be concluded from the simulation study that, none of the traditional algorithm segment the artefact instrument with better accuracy. Since the algorithm works based on intensity values of pixel in the given image most of instrument region is covered by saturation. The algorithm mis-concluded and segmented the artefact saturation instead of instrument. The DL based algorithms suffered since the artefacts overlap each other and the features of instruments is too tough to classify between the foreground and the background.

- **Segmentation of bubble artefact**

Bubbles in endoscopic images is an artefact. The artefact by itself does not hold any colour. It is a transparent air sac covered by intestinal fluid. This fluid reflects light that falls on the surface thereby giving way to many other artefacts. The traditional algorithm tries to segment the light that falls on the area, not the bubbles. Most of the algorithms fail to segment the region accurately due to its complex structure and colourless feature. Figure 5.8 (a) and (b) shows the input image and its corresponding binary mask for bubbles artefact. Figure 5.8 (c) – (l) depicts the segmentation output of all the algorithms under study. The results of traditional algorithms presented in Figure 5.8 (c) – (i) shows that none of the traditional algorithm segment the boundaries of the bubbles artefact. Consequently, Figure 5.8 (j) – (l) presents the results of DL - based algorithms. It is clear that DL – based algorithm tried to segment the artefact, yet U-Net and Link-Net over segmented the artefact as shown in Figure 5.8 (j) and (l), whereas FPN under segmented the artefact as shown in the Figure 5.8 (k). Table 5.4 tabulates the performance scores of both traditional and DL algorithms.

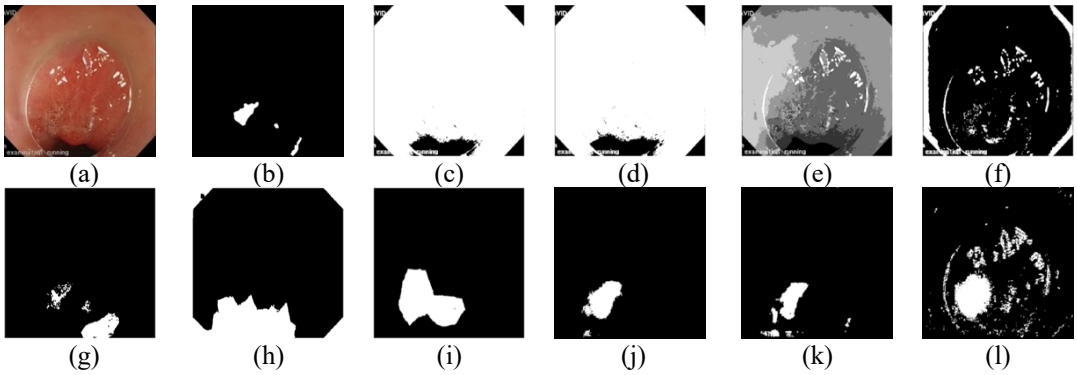

Figure 5.8. Segmentation of Bubbles Artefact

(a) Original Image (b) Mask (c) Otsu (d) Gray Threshold (e) Multi Threshold (f) Adaptive Threshold (g) Active Contour (h) Lazy Snapping (i) Super Pixel (j) U-Net (k) FPN (l) Link-Net

Segmenting the artefact bubble is challenging due to its transparent nature. The pixel value of the foreground and the background cannot be classified in much of the cases. Even DL based algorithms struggle to attain the same which is evident from the results

Table 5.4 Performance Measures of Various Algorithms for the bubbles Artefact

Image	Otsu	Gray Threshold	Multi Threshold	Adaptive Threshold	Active Contour	Lazy Snapping	Super Pixel	U-Net	FPN	Link-Net
Specificity										
1	0.7700	0.083	0.428	0.8782	0.9583	0.8306	0.8958	0.9733	0.9791	0.9046
2	0.6436	0.6592	0.6187	0.9344	0.9451	0.8119	0.8938	0.8725	0.9581	0.8808
3	0.6684	0.6684	0.6244	0.9279	0.9016	0.5024	0.8349	0.9389	0.948	0.8993
Accuracy										
1	0.0914	0.097	0.4271	0.8671	0.9516	0.8205	0.893	0.9703	0.9766	0.9042
2	0.6378	0.6522	0.6137	0.9162	0.9348	0.8034	0.8905	0.8748	0.9579	0.8706
3	0.6108	0.6108	0.5857	0.8282	0.9002	0.5375	0.8323	0.9198	0.9296	0.8248
Sensitivity										
1	0.9792	0.963	0.3706	0.1776	0.5382	0.1976	0.7203	0.7864	0.8214	0.8802
2	0.3490	0.2992	0.3628	0.1553	0.4204	0.3785	0.7268	0.9895	0.9496	0.3593
3	0.1677	0.1677	0.2879	0.0606	0.8895	0.8077	0.8127	0.7733	0.7884	0.2512
Dice										
1	0.0332	0.0329	0.0202	0.0409	0.2619	0.0339	0.1767	0.4578	0.5277	0.2266
2	0.0362	0.0325	0.0354	0.0698	0.2011	0.0699	0.2058	0.2357	0.4682	0.0978
3	0.0902	0.0902	0.1378	0.7050	0.6720	0.2865	0.5270	0.6893	0.7204	0.2479
Jaccard										
1	0.0169	0.0167	0.0102	0.0209	0.1507	0.0173	0.0969	0.2969	0.3584	0.1278
2	0.0185	0.0165	0.0180	0.0362	0.1118	0.0362	0.1147	0.1336	0.3056	0.0514
3	0.0472	0.0472	0.0740	0.0390	0.5061	0.1672	0.3578	0.5259	0.563	0.1415

- **Segmentation of the miscellaneous artefacts**

Figure 5.9 (a) and (b) shows the input test image and its corresponding binary mask for miscellaneous artefacts. Each of them has diverse features. Endoscopic miscellaneous artefacts cover chromatic aberration, debris, text information present in the image and few imaging artefacts. Text information have sharp and clear boundaries but in few images the text information is tiny. Chromatic aberration has wide range of colours spread across the affected region. Similarly, when an organ is imaged with saturation or chromatic aberration or minor movement artefacts occur. Hence segmenting the artefact using traditional algorithms succeed when the endoscopic images are affected with mild chromatic aberration where the values of pixels between the background and the affected region are differentiable. In most of the cases the colour of the artefact and background are the same. In such cases algorithm do not perform well against the variations in artefact as seen in Figure 5.9 (c)-(f). Active contour, lazy snapping and super pixel algorithms segment well as evident from the result presented in Figure 5.9 (g)-(i). DL-based algorithm with its trained knowledge on all variations in miscellaneous artefacts are able to segment the region but in few cases the

algorithm over-segments the affected area. Table 5.5 given holds the performance scores of all traditional and DL-based algorithms for the miscellaneous artefact.

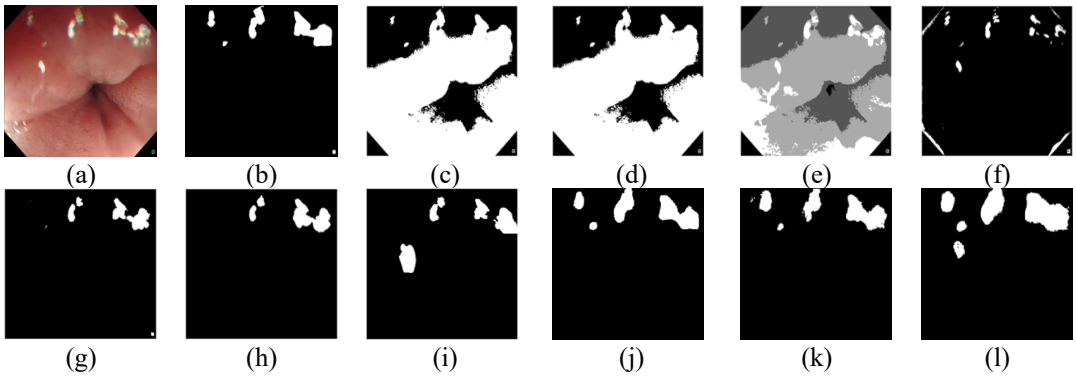

Figure 5.9 Segmentation of Miscellaneous Artefacts

(a) Original Image (b) Mask (c) Otsu (d) Gray Threshold (e) Multi Threshold (f) Adaptive Threshold (g) Active Contour (h) Lazy Snapping (i) Super Pixel (j) U-Net (k) FPN (l) Link-Net

Table 5.5 Performance Measures of Various Algorithms for the Miscellaneous Artefacts

Image	Otsu	Gray Threshold	Multi Threshold	Adaptive Threshold	Active Contour	Lazy Snapping	Super Pixel	U-Net	FPN	Link-Net
Specificity										
1	0.3496	0.3496	0.3536	0.9638	0.976	0.9698	0.9612	0.9834	0.9828	0.9502
2	0.5182	0.5298	0.7086	0.9668	0.9663	0.9212	0.9469	0.9763	0.9814	0.9411
3	0.295	0.3078	0.3712	0.9553	0.9526	0.9777	0.9829	0.9923	0.9924	0.9865
Accuracy										
1	0.3698	0.3698	0.3731	0.9339	0.9591	0.9534	0.9459	0.9819	0.9806	0.9519
2	0.5228	0.5331	0.6958	0.936	0.9418	0.8876	0.9272	0.969	0.9721	0.9431
3	0.324	0.3361	0.3962	0.9167	0.9508	0.9724	0.9808	0.9551	0.9527	0.9513
Sensitivity										
1	0.7936	0.7936	0.7822	0.3085	0.607	0.6094	0.626	0.9498	0.9337	0.9881
2	0.6111	0.5965	0.4476	0.3352	0.465	0.2318	0.5448	0.8273	0.7897	0.9818
3	0.9341	0.9329	0.9235	0.103	0.9123	0.8611	0.9351	0.1715	0.1171	0.2102
Dice										
1	0.1031	0.1031	0.1022	0.2986	0.5754	0.5439	0.5135	0.827	0.8142	0.6521
2	0.1111	0.1108	0.1256	0.3383	0.4382	0.1675	0.4222	0.7228	0.7339	0.6274
3	0.1113	0.1129	0.1217	0.1007	0.627	0.7389	0.815	0.2573	0.1823	0.2812
Jaccard										
1	0.0543	0.0543	0.0539	0.1755	0.4039	0.3735	0.3454	0.705	0.6866	0.4838
2	0.0588	0.0587	0.067	0.2036	0.2806	0.0914	0.2676	0.6417	0.5659	0.5797
3	0.0589	0.0599	0.0648	0.053	0.4566	0.5859	0.6877	0.1476	0.1009	0.1636

Miscellaneous artefacts are the toughest to segment. The nature of the artefact varies from one image to the other. For instance, specular reflection can be identified based on the bright pixel spots and similarly the saturated region. In case of miscellaneous artefacts, it includes chromatic aberration, debris, text information in the image and few imaging artefacts. Each of them has different features. Hence the performance of the DL algorithms is better when compared to traditional algorithms.

In summary, few traditional models such as, active contour and lazy snapping perform well across some of the artefacts. At the same time, threshold-based segmentations are likely to perform well when there is a significant difference in pixel intensity between the class and the background. DL-based models have a balanced segmentation performance through all the endoscopic artefacts. Traditional algorithms consumed ample time for selecting the suitable threshold voltage and the seed to segment every artefact. In comparison, DL models after training segment the artefacts accurately than traditional algorithms. The training of the DL algorithm consumes resources, cost and time but once trained, the prediction happens in less than a second. This trained model could be implemented in the endoscopic imaging pipeline, along with other modules like artefact restoration, which could ease the diagnosis and therapeutic process and aid clinicians with better viewability and lesser procedure time.

5.7 ARTEFACT SEGMENTATION MODELS FOR ENSEMBLE

A pre-trained deep learning network from segmentation model API is preferred. Many combinations of networks are trained for this research. The best-performing network is chosen for the final ensemble. The combination of the network with the backbone trained for this research includes:

Method 1: Link-Net with EfficientnetB3 as the backbone.

Method 2: Link-Net with Inceptionv3 as the backbone.

Method 3: U-Net with MobileNet as the backbone.

Method 4: U-Net with EfficientnetB3 as the backbone.

Method 5: U-Net with SEResNeXt101 as the backbone.

Method 6: PSPNet with Resnet 34 as the backbone.

Method 7: U-Net with ResNeXt101 as the backbone.

Among the seven mentioned networks, the three best-performing networks are chosen. Uniformity is maintained across training all the models. Standard evaluation metrics such as, Jaccard and F2 score are used to compare the performance. After performance evaluation it is concluded to adopt U-Net with EfficientnetB3 backbone, U-Net with SEResNeXt101 backbone and Link-Net with EfficientnetB3 backbone. The performance evaluation is detailed in the Figure 5.10.

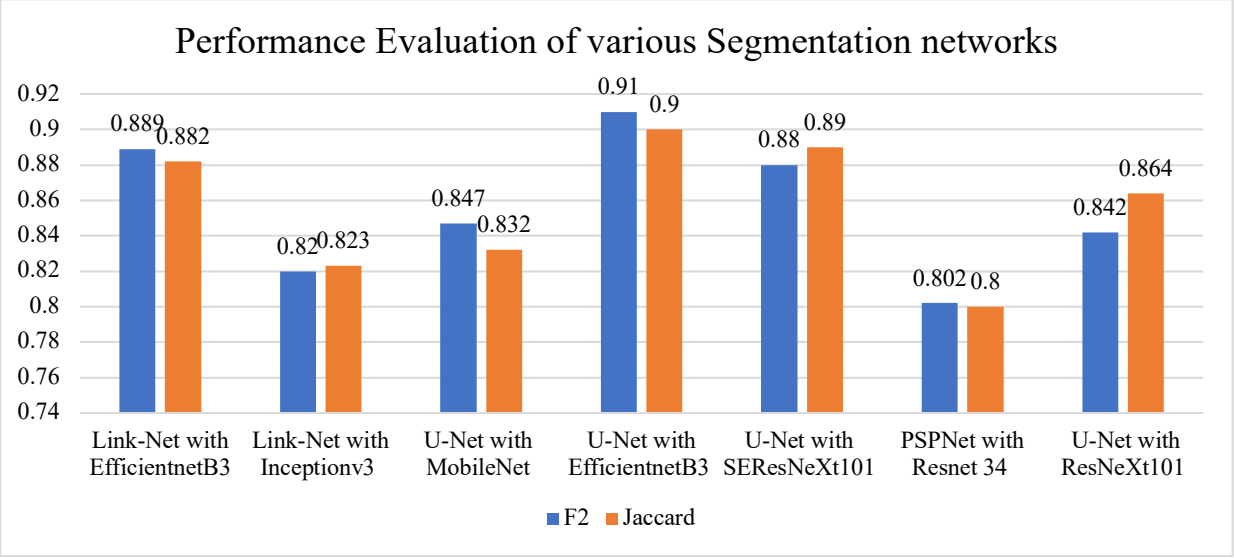

Figure 5.10 Performance Evaluation of Various Deep Learning based Image Segmentation Networks

It is evident from the figure that the performance of U-Net with Efficient Net B3, U-Net with SEREsNeXt101 and Link-Net with Efficient-net B3 is found to be competitively better than other competing algorithms. Hence top performing algorithms are chosen. The architecture and training strategy of the chosen network is discussed in detail in the following sub sections.

5.7.1 U-Net and Link-Net for Ensemble

The adapted U-Net and Link-Net architectures are shown in Figures 5.11 and 5.12. The U-Net architecture, bags the name 'U' because of its structure. The encoder and decoder path are much similar and the concatenation operation between the paths resulted in 'U' shape. The contacting path of the network consists of repeated convolutions. Each of the convolution operation is followed by ReLU and a max pooling layer. In this convolutional path an input image is taken and its spatial resolution is reduced when the image passes from one stage to the other and the feature size increases. The expansion path concatenates the feature information and the spatial information. Hence it results in better resolution output images. The images in the dataset belongs to diverse resolution and size. Hence all the images are resized to 224x224. DL segmentation models are designed to accept an image size of 224x224.

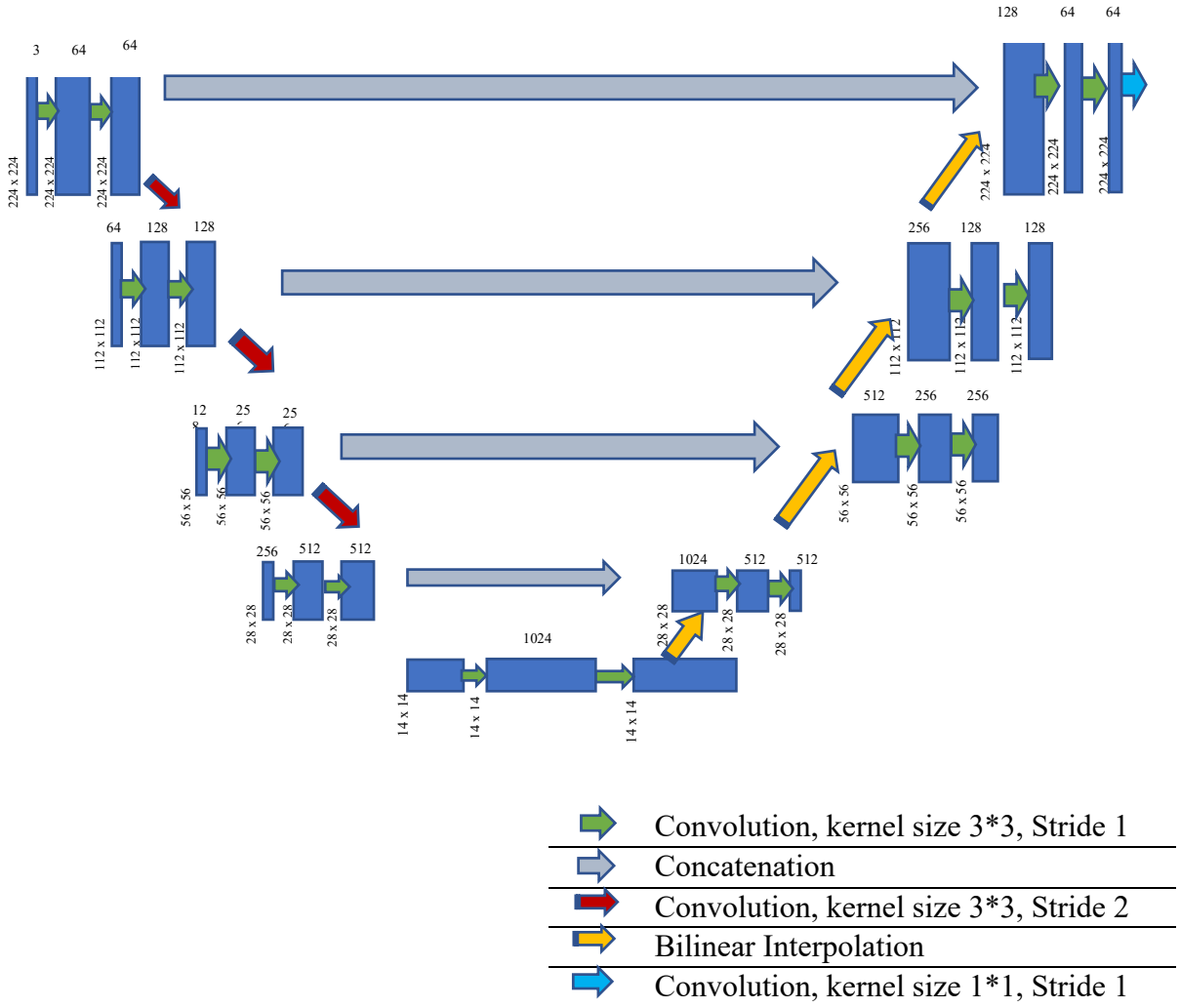

Figure 5.11 U-Net Architecture Adopted for Artefact Segmentation

Link-Net has two main stages, encoder stage on the left and decoder stage presented on the right as given in the Figure 5.12. In the encoder stage, convolutional and max pool layers perform convolution on input image. The size of the kernel is 7 x 7. A stride of 2 is fixed. Along with convolution the layer also performs max pooling operation. The parameters for max pool include area of 3 x 3 and stride of 2. The rest of the encoder blocks are the residual blocks.

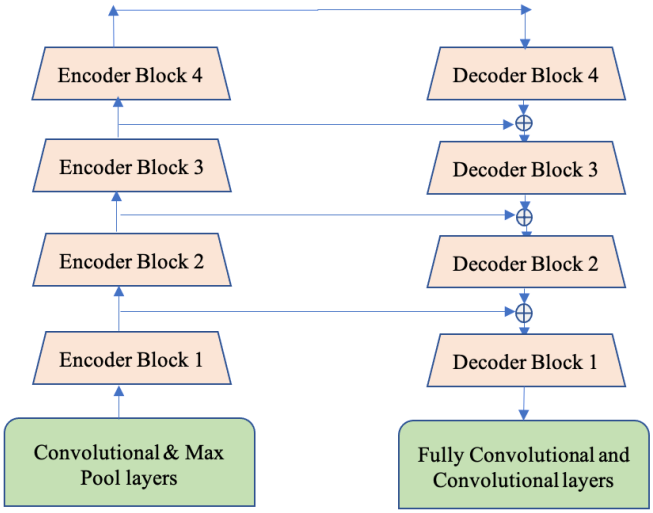

Figure 5.12 Link-Net Architecture Adopted for Artefact Segmentation

The structure of the encoder blocks used are shown in the Figure 5.13. The construction of decoder blocks is portrayed in the Figure 5.14. In the Link-Net architecture batch normalization is used between each of the convolutional layer. Every convolutional layer is followed by ReLU.

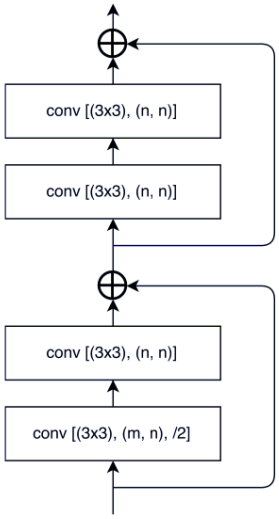

Figure 5.13 Encoder Blocks of Link-Net

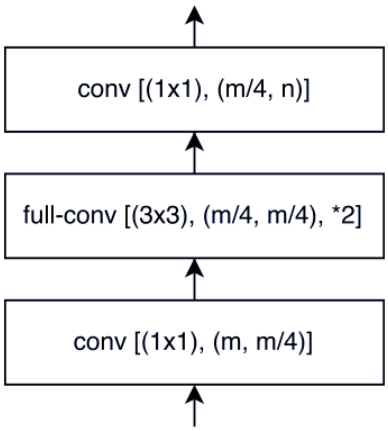

Figure 5.14 Convolutional Layers in Decoder Blocks of Link-Net

5.7.2 Training of U-Net and Link-Net for Artefact Segmentation

All the three networks are trained for 300 epochs with the same training set of images and tested with the images from the test set. Similarly, the optimizer, learning rate and all the other associated parameters are kept constant across all the models. The performance of each model is evaluated after 300 epochs. The model is evaluated based on the metrics Jaccard and F2 score. The initial parameters set to train each of the three models are described in the Table 5.6. Dice and focal loss are considered for all the models. Each of the three models holds 3.4, 4.02,3.90 million learnable parameters.

Table 5.6 Input Parameters Set to Train Segmentation Model

Parameter	U-Net with Efficient-net B3	U-Net with SEResNeXt101	Link-Net with Efficient-net B3
Learning Rate	0.0001	0.0001	0.0001
Encoder Weights	ImageNet	ImageNet	ImageNet
Data Augmentation	Yes	Yes	Yes
Activation	softmax	softmax	softmax
Number of Epochs	300	300	300
Class	5	5	5
Optimizer	adam	adam	adam

The augmented dataset is considered for training. Images from the test set are allowed to pass through the trained model. The predictions of the individual model are tested for its performance measures.

5.7.3 Ensemble Model for Artefact Segmentation

The proposed ensemble model combines three DL models: U-Net with Efficientnetb3 backbone, U-Net with SEResNext101 backbone and Link-Net with Efficientnetb3 backbone. The input images collected from the public dataset is initially resized to 224 x 224 x 3 and divided into train and test set. The train set is data augmented and passed into all the three models. All three models are trained using the expanded training set and assessed individually. After evaluation of all three models, they are combined for final predictions. The binary mask of all three models is averaged based on classes. The averaged output acts as the final prediction of the model. The final averaged mask is binarized using Otsu thresholding method. Figure 5.15 shows the proposed artefact segmentation model.

The algorithmic steps of the proposed ensemble model are described:

Step 1: A sample image along with its binary mask is selected from the test dataset.

Step 2: The test image is passed in to all the three trained models.

Step 3: The trained model predicts and segment the artefacts present.

Step 4: The binary mask for the three artefacts is obtained and each mask corresponding to every artefact is averaged using OpenCV.

Step 5: The averaged mask is binarized with the help of Otsu binarization. An optimum threshold (t) value is identified using Otsu's algorithm. The value is expected to minimize the weighted within class variance. This algorithm tries to find an optimum values between the peaks. The relation is described from the equation 5.8 to 5.11.

$$\sigma_w^2(t) = q_1(t)\sigma_1^2(t) + q_{2\,(t)}\sigma_2^2(t) \tag{5.8}$$

$$q_1(t) = \sum_{i=1}^{t} P(i) \quad \text{and} \quad q_2(t) = \sum_{i=t+1}^{I} P(i) \tag{5.9}$$

$$\mu_1(t) = \sum_{i=1}^{t} \frac{iP(i)}{q_1(t)} \quad \text{and} \quad \mu_2(t) = \sum_{i=t+1}^{I} \frac{iP(i)}{q_2(t)} \tag{5.10}$$

$$\sigma_1^2(t) = \sum_{i=1}^{t} \left[i - \mu_1(t)\right]^2 \frac{P(i)}{q_1(t)} \text{ and } \sigma_2^2(t) = \sum_{i=t+1}^{I} \left[i - \mu_2(t)\right]^2 \frac{P(i)}{q_2(t)} \tag{5.11}$$

Step 6: The result after Otsu's binarization is the final mask. Hence the result has 5 individual binary masks for each one for one artefact as shown in the Figure 5.15.

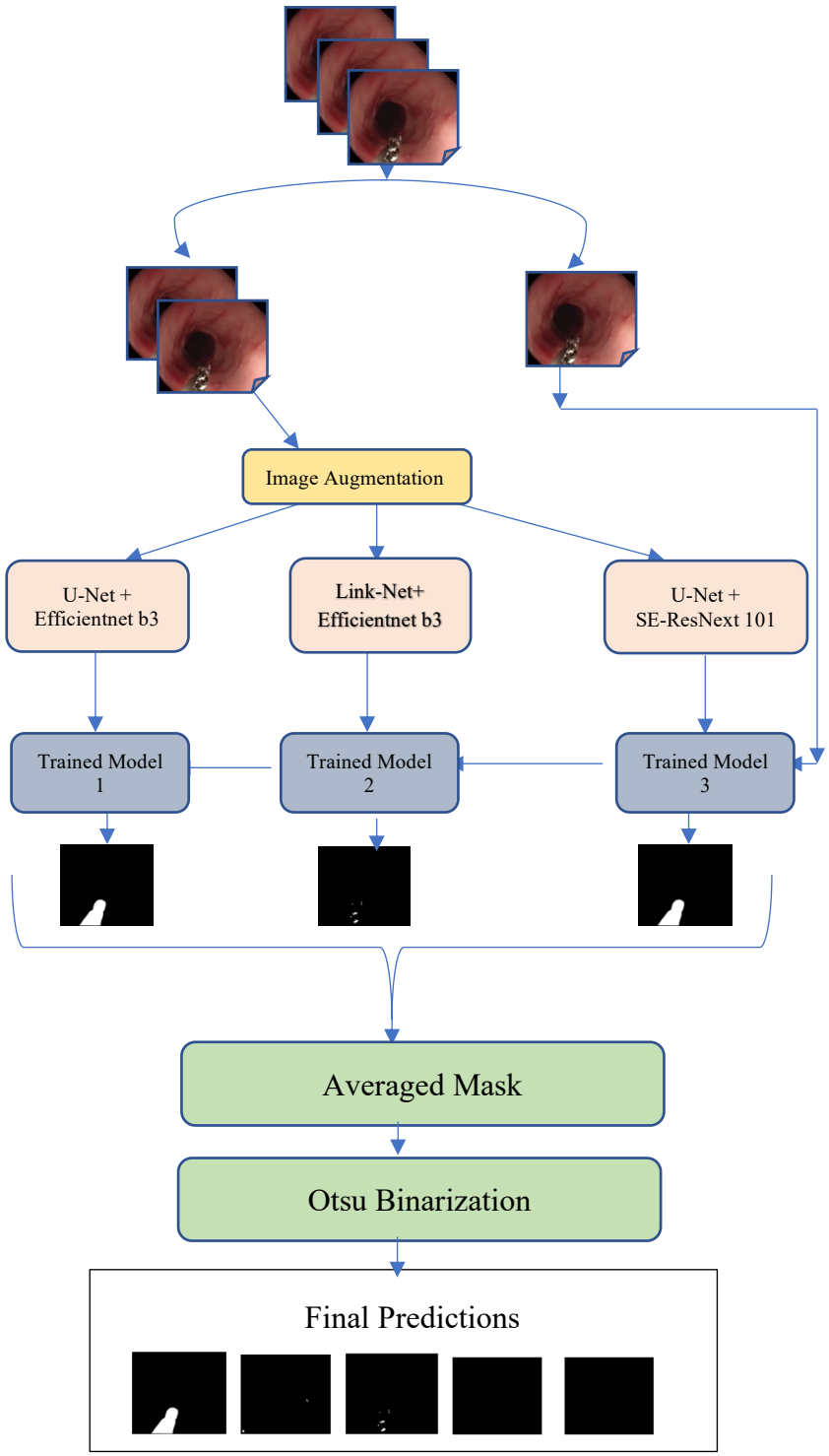

Figure 5.15 Proposed Artefact Segmentation Model

5.7.4 Performance Metrics

The ensemble model proposed in this research is evaluated with two common metrics, Jaccard and F2 score. The formula used to calculate the metric is discussed in the section 5.6.2.

5.7.5 Simulation Results of Proposed DL Based Ensemble Segmentation Model

Combinations of DNN with various backbones are trained for artefact segmentation. Networks that performed best are combined together for the proposed model. The algorithm's performance is assessed against existing results using standard performance evaluation metrics such as, dice score and F2 score. The metric is chosen for comparison from the literature. The proposed model incorporates three learners whose predictions are averaged for final predictions. All three base learners are trained equally with the same train set and standard input parameters. The images in the dataset are found to be very limited. Therefore, data augmentation techniques are employed. The methods include random Gaussian noise, random cropping, varying hue, saturation, brightness, image blurring, image sharpening and flipping. After training for 300 epochs, the model performance is evaluated, tuned and retrained for best results. After finalizing the individual results, all three models are combined. Now the input image is passed on through the proposed model, which holds three learners. Each predicts the binary mask; all three masks are averaged for the final performance. The average mask is converted back to binary form with the help of the traditional Otsu method. Figure 5.16 (a) shows the input test image. Figure 5.16 (b) – (f) shows the original ground truth mask given in the dataset for each artefact. Figure 5.16 (g) – (k) shows the predicted segmented output for each of the artefact. The input image is affected by artefacts such as, saturation and miscellaneous artefacts; hence the proposed model segments the two artefacts. The rest of the masks are left blank as there no instrument, specular reflection or bubbles found in the input test image

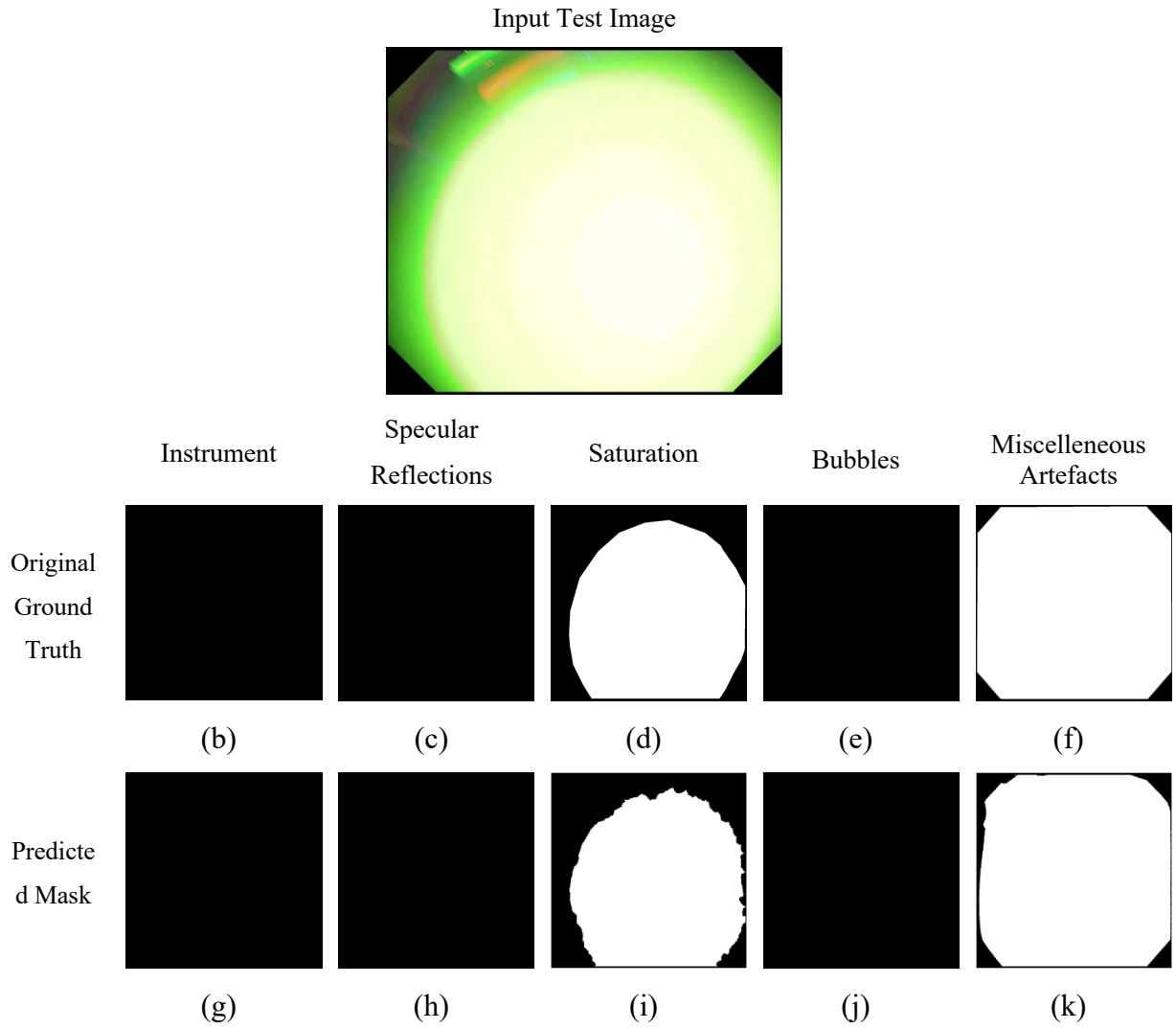

Figure 5.16 (a) - (k) Results of Proposed Segmentation Algorithm

The results presented in Figure 5.17 are obtained from the proposed model and compared with results from the literature. The proposed algorithm is compared against standard networks such as, mask aided R-CNN, DeepLabv3+ and U-Net++. It is apparent from the graphical representation that the proposed model performs relatively better than previous models. The training strategy, choice of backbone aided to improve the performance. The **Jaccard score** of the proposed ensemble model is equal to **0.753** which is 17.36% higher than the existing model proposed by S. Yang & Cochran(2019). The **F2 score** of the proposed model is equal to **0.796** which is 17.42% higher than the existing literature results. Hence such models can preferably be used in the CAD system which assists further processing techniques.

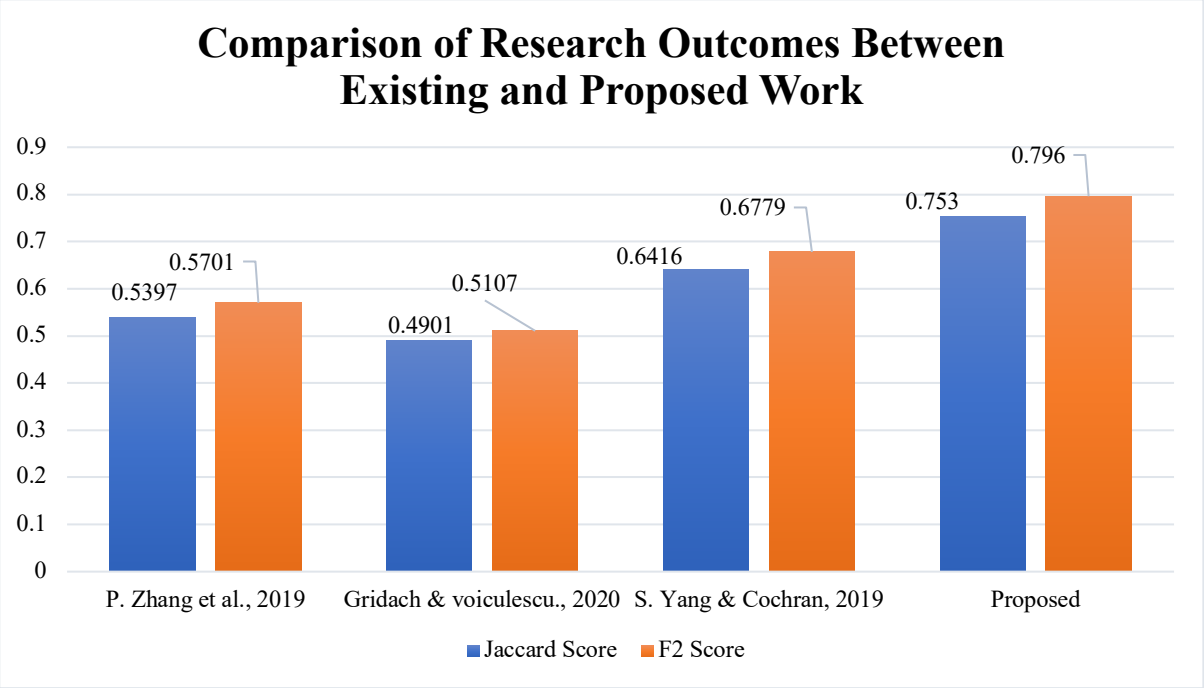

Figure 5.17 Comparison of Research Outcomes Between Existing and Proposed Work

5.8 CHAPTER SUMMARY

A DL -based segmentation model from segmentation model API is chosen. The chosen models are trained with images from EAD dataset. The dataset holds images for artefact segmentation along with binary masks for five commonly occurring artefacts. It includes specular reflections, saturation, bubbles, instruments and miscellaneous artefacts. The data augmentation techniques from albumentations library are adopted to boost the dataset size. The DL-based networks such as, U-Net and Link-Net are chosen with various backbones among the trained network. The input parameters are set common across all models. The performance of the network is estimated using traditional metrics such as IoU, F2 Score, precision and recall. The top three well performing networks are combined for final proposed model.

The proposed model combines the predictions of all three models. The mask of individual artefact is averaged and binarized using Otsu's algorithm. From the research findings it is evident that the combination of DL based segmentation algorithms such as, U-Net with EfficientNet B3 backbone, Link-Net with Efficient-Net b3 backbone and U-net with SERESNeXt101 backbone perform well across segmentation of all artefacts in endoscopic images. The efficient training strategy followed resulted with improved performance across

all models. From the results, it is evident that the proposed DL-based algorithms perform well across all artefacts and the proposed ensemble model showcase 17.36% and 17.42% improvement in Jaccard and F2 score respectively than literature results. Hence such algorithms can be incorporated in the pre-processing pipeline for designing futuristic CAD systems.