IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 24, 2021, accepted August 13, 2021, date of publication August 17, 2021, date of current version August 24, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3105454

Transfer Learning and Decision Fusion for Real Time Distortion Classification in Laparoscopic Videos

NOUAR ALDAHOUL[1,2], HEZERUL ABDUL KARIM[1], (Senior Member, IEEE), MYLES JOSHUA TOLEDO TAN[2,3,4], (Member, IEEE), AND JAMIE LEDESMA FERMIN[2,3], (Student Member, IEEE)

[1]Faculty of Engineering, Multimedia University, Cyberjaya 63100, Malaysia
[2]Yo-Vivo Corporation, Bacolod 6100, Philippines
[3]College of Engineering and Technology, University of St. La Salle, Bacolod 6100, Philippines
[4]Department of Natural Sciences, University of St. La Salle, Bacolod 6100, Philippines

Corresponding author: Nouar Aldahoul (nouar.aldahoul@live.iium.edu.my)

This research project was funded by Multimedia University, Malaysia.

ABSTRACT Laparoscopic surgery is a surgical procedure performed by inserting narrow tubes into the abdomen without making large incisions in the skin. It is done with the aid of a video camera. Laparoscopic videos are affected by various distortions during surgery which lead to loss of visual quality. Identification of these distortions is the primary requisite in automated video enhancement systems used to classify the distortions correctly and accordingly select the proper algorithm to enhance video quality. In addition to high accuracy, the speed of distortion classification should be high, and the system must consider real-time conditions. This paper aims to address the issues faced by similar methods by developing a fast and accurate deep learning model for distortion classification. The dataset proposed by the ICIP2020 conference challenge was used for training and evaluation of the proposed method. This challenging dataset contains videos that have five types of distortions such as noise, smoke, uneven illumination, defocus blur, and motion blur with four levels of intensity. This paper discusses the proposed solution which received the first prize in the ICIP2020 challenge. The solution utilized a transfer learning approach to transfer representation from the domain of natural images to the domain of laparoscopic videos. We used a pre-trained ResNet50 convolutional neural network (CNN) to extract informative features that were mapped by support vector machine (SVM) classifiers to various distortion categories. In this work, the problem of multiple distortions in the same video was formulated as a multi-label distortion classification problem. The approach of transfer learning with decision fusion was applied and was found to outperform other solutions in terms of accuracy (83%), F1 score of a single distortion (94.7%), and F1 score of single and multiple distortions (94.9%). In addition, the proposed solution can run in real time with an inference speed of 20 frames per second (FPS).

INDEX TERMS Decision fusion, distortion classification, laparoscopic video, multilabel classification, real time, transfer learning.

I. INTRODUCTION

A. LAPAROSCOPIC SURGERY

If one were to investigate the long history of Western medicine, one would discover that the first few ideas that led to the conception of minimally invasive surgery were born in as early as 400 B.C. when Hippocrates described the use of a speculum for the examination of haemorrhoids [1]. Two millennia later, Ephraim McDowell removed an ovarian tumour during the first successful elective laparotomy, i.e. an open surgery of the abdomen [2]. Laparotomies have remained the gold standard since then until not very long ago [3] with some surgeons arguing for its superiority for certain procedures [4]–[6]. Some experts in the 1990s regarded

The associate editor coordinating the review of this manuscript and approving it for publication was Juan Liu.

laparoscopic surgery as the gold standard for procedures [7], such as cholecystectomies in the treatment of symptomatic cholelithiasis [8], adrenalectomies [9] and endorectal colon pull-through for the treatment of Hirschsprung's disease [10]; while some recognized the potential for it to become so for appendectomies [11]. The most cited reasons for the superiority of laparoscopic procedures over laparotomies include less infections, less blood loss, less pain, lower perioperative morbidity, shorter hospital stay, faster recovery, and earlier return to work without any significant difference in surgical outcomes [3], [8]–[21].

The design of a video quality assessment (VQA) system for laparoscopic surgery of superior calibre is highly essential. Very often, laparoscopic videos are obscured by distortions that degrade the visibility of the patient's anatomy and thus, the overall quality of the laparoscopic or robot-assisted procedure [22], [23]. These distortions commonly arise in the form of noise, smoke, uneven illumination, and blur, which are all concomitant artifacts that come with the operation of the surgical equipment involved in minimally invasive surgery [24]. However, most of the existing solutions developed to address the issues associated with these distortions rely on augmenting the hardware using one or a combination of the many means enumerated in [25]. Unfortunately, these means consume a significant amount of time to carry out, yet despite this, cannot guarantee robust improvements in laparoscopic video quality. Therefore, we believe that an automated video enhancement system will prove to be the key solution to the problem presented.

B. CONVOLUTIONAL NEURAL NETWORKS

Convolutional neural networks (CNNs) [26], [27] are arguably the most notable approach to Deep Learning [28]. CNNs are made up of two core layers, namely the convolutional layer and the pooling layer. Units in the convolutional layer are arranged in feature maps, with each unit linked by a set of weights to local patches in the feature maps of the previous layer. The local weighted sum is inputted to an activation function. The pooling layer then combines semantically similar features [28].

In 2012, the ImageNet project which runs the annual ImageNet Large Scale Visual Recognition Challenge (ILSVRC) witnessed the success of AlexNet, which achieved a historically low top-5 error rate for classification tasks [29], [30]. AlexNet, which used purely supervised learning with no unsupervised pre-training, was considered a major improvement to its CNN predecessors [29]. In 2014, ILSVRC witnessed two great improvements to AlexNet; namely the visual geometry group (VGG) model, which won second place (top-5 error rate 7.3%), and the GoogLeNet model, which won first place (top-5 error rate 6.7%). Both cut the top-5 error rate of their predecessor, AlexNet (top-5 error rate 16.4%) to less than half [28]. The VGG model created by the Visual Geometry Group from the University of Oxford used a very deep CNN with 16 weight layers, 13 of which were convolutional layers with 3×3 filters, while three were fully connected layers [31]. GoogLeNet, on the other hand, made use of a CNN architecture codenamed Inception, whose objective was to determine how an optimal local sparse structure of a convolutional vision network could be estimated. It made use of 22 layers and nine Inception modules and employed average pooling instead of fully connected layers for classification [32]. Moreover, despite its much lower top-5 error rate, GoogLeNet uses 12 times fewer parameters than its predecessor, AlexNet [28]. In 2015, ILSVRC witnessed a remarkable improvement which brought its predecessor's top-5 error rate down by close to half. That year, the residual networks (ResNet; top-5 error rate 3.6%) [33] by Microsoft Research, with 152 layers and a lower complexity than VGG, won first place [28]. This model introduced deep residual learning to address the problem of degradation. Like VGG, its convolutional layers had 3×3 filters. Moreover, it ended with global average pooling and 1000-way fully-connected layers with softmax [33].

C. THE ICIP2020 CHALLENGE

The challenge of distortion classification in laparoscopic video was proposed for the 2020 IEEE International Conference on Image Processing (ICIP) wherein participants were tasked to classify distortions in laparoscopic videos. The objective of this challenge was to address the problem in VQA by constructing a rapid, integrated, and effective algorithm for real-time classification of laparoscopic video distortions [34]. This paper aims to present the work that went into solving the problem introduced by the challenge and the proposed solution that was submitted and which subsequently received the first prize. Figure 1 provides a diagram that describes the process for automating video-guided surgery. Blocks represent the two stages of the automation process, namely distortion identification and quality enhancement.

D. RELATED WORKS

To the best of our knowledge, there has not been a significant amount of work done to address distortions in laparoscopic videos in real-time despite the apparent necessity. But before we begin to give you a glimpse into the work that we have completed, let us talk about some of the more foundational research upon which our work is built and understand some of the motivations for this proposed method.

In as early as 2012, an earlier hybrid system aimed at classifying distortion and evaluating image quality for media applications was developed. It used traditional image quality measures (IQMs) as features and simple linear discriminant analysis as its classifier [35].

Subsequently, in 2013, the distortion/content-dependent video quality index (DCVQI), which at that time, was a new video quality index, and although still for digital video applications outside of medicine, was proposed. This approach classified distortions (e.g., blurring, ringing, jitter) as either local or global [36].

A fairly more recent application of VQA in medicine was suggested in 2017 for use in telesurgery, wherein a surgical

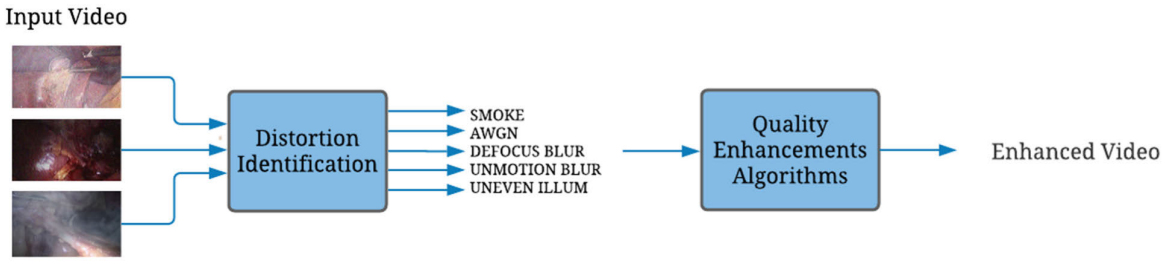


FIGURE 1. Automated video guided surgery with two blocks: distortion identification and quality enhancement.

operation would be performed by a surgeon at the patient's side with an expert surgeon guiding the procedure remotely. For this to be possible, a reliable transmission of surgical video over large distances is necessary. This, however, is susceptible to signal distortions that may arise as a result of data compression artifacts and transmission errors. These distortions significantly affect the video quality perceived by the surgeon. A qualitative assessment of the application followed and confirmed the existence of distortions. These assessment, however, was not real-time and solely provided direction for future research and an appreciation for the need to address distortions in surgical videos [37]. Around the same time that year, work to classify semantic shots from laparoscopic gynaecologic surgery for purposes of medical education, medical research and everyday surgical documentation using single-frame CNNs was carried out by another team from Austria. High-level features were also extracted from the AlexNet architecture with weights from a pre-trained model and were subsequently fed into a support-vector machine (SVM) classifier. Through this work, they found that previous advances in general image classification methods transfer nicely into surgical video classification [38]. Hence, the potential for work to classify distortions for the eventual goal of enhancing surgical videos is vast.

In 2018, a method aimed at removing surgical smoke in laparoscopic images was developed by Wang *et al.* [39]. Here, a smoke-distorted image was separated into two parts – direct attenuation and smoke veil. A cost function defined based on the observation that a smoke veil has low contrast and low inter-channel differences was calculated using an augmented Lagrangian method.

A more general-purpose approach to VQA was proposed in 2019 involving a new architecture for no-reference assessments based on features from pretrained CNNs, transfer learning, temporal pooling of deep features, and regression. Here, solutions were obtained by solely applying temporally pooled deep features without any manually derived features [40].

A recent work in 2020 targeted VQA by detecting and identifying distortions and their severity automatically. Here, the authors constructed a laparoscopic video quality database with a set of 200 videos, with five types of distortions and four levels of intensity for this purpose [22]. A distortion-specific classification method was used for each type of distortion such as a fast noise variance estimator with a threshold for noise distortion, statistics of the luminance component of an image for uneven illumination distortion, a saturation analysis (SAN) classifier for smoke distortion, and a perceptual blur index (PBI) with a threshold classifier for blur distortion [22]. As a replacement for traditional methods that depend on the distortion categories for coefficients modelling to extract specific features from the images, a single deep neural network was proposed to solve the two important problems of distortion classification and quality ranking [41]. A set of 20000 images (1000 images per class) were selected from the Cholec80 laparoscopic dataset with 20 classes (five distortions, each with four levels of severity) [41]. Both previous works focused on a single distortion classification problem. The problem of multiple distortions has yet to be investigated in laparoscopic videos.

In this work, we propose a computational framework for automatic evaluation of video quality for laparoscopic surgery. This work solely focuses on the module of distortion detection/classification which is the first stage of the VQA system. This paper aims to address the multi-label distortion classification problem by developing an accurate and real-time solution.

This paper has several contributions as follows:

- A novel distortions classification system that was utilized to identify multiple distortions available in the same laparoscopic video for quality assessment.
- A real-time solution that can run with a speed of 20 FPS to identify distortions in live-captured videos.
- To the best of our knowledge, this is the first paper that targets multi-label distortion classification with a novel laparoscopic video dataset proposed by ICIP2020 challenge organizers.
- We validated the concept of domain generalization and adaptation from the domain of natural photos to the medical imaging domain. This approach transfers deep representation from ImageNet to laparoscopic videos.
- The decision fusion approach had a significant impact on classification performance. An ablation study was performed to check the improvement that fusion can contribute to the model's performance in terms of accuracy, and F1 score.

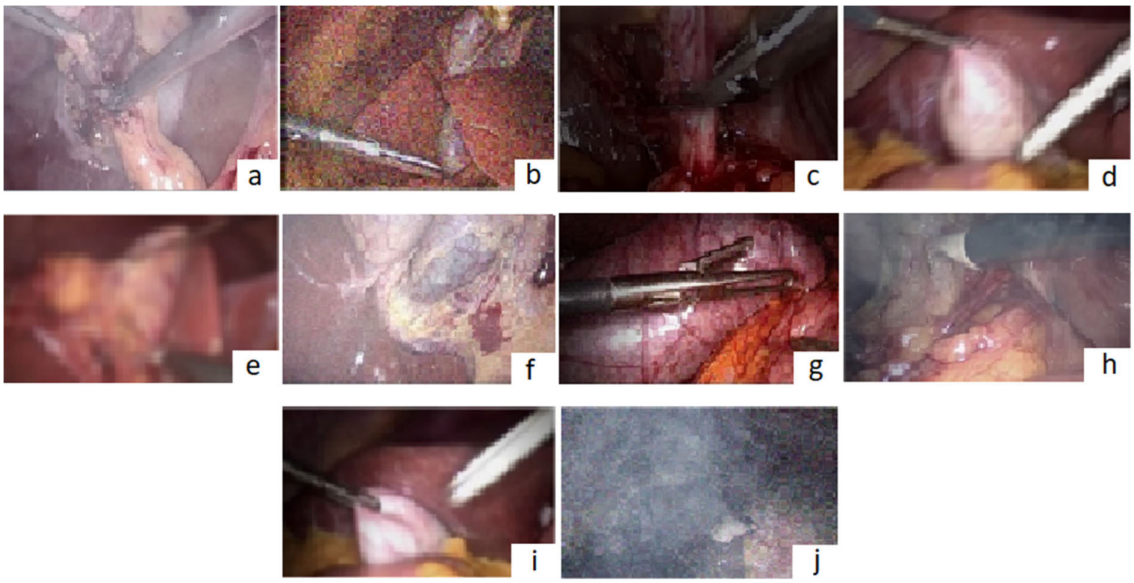


FIGURE 2. Various samples from the laparoscopic dataset [34] including single distortion: (a) smoke, (b) AWGN, (c) uneven illumination, (d) motion blur, (e) defocus blur, and multiple distortion: (f) noise-smoke, (g) noise-uneven, (h) smoke-uneven, (i) defocus-uneven, and (j) uneven-smoke-noise.

- The proposed solution received the first prize in the ICIP2020 challenge and outperformed other solutions in terms of accuracy (83%), F1 score of single distortion (94.7 %), and F1 score of single distortion and multi-distortion (94.9 %) and inference speed of 20 FPS.

This paper is organized as follows: II. MATERIALS AND METHODS describes the datasets; and demonstrates the CNN for feature learning, and the Pre-trained CNN for feature extraction and transfer learning. In III. RESULTS AND DISCUSSION, the experimental setup and results are discussed. Finally, IV. CONCLUSION AND FUTURE WORK summarizes the outcome and significance of this work and gives readers a glimpse into potential improvements to our work in the future.

II. MATERIALS AND METHODS

In this section, the dataset used in this work is described. Additionally, the methods of feature learning, feature extraction, and transfer learning are demonstrated to shed light on the functionalities, advantages, and drawbacks of each. Furthermore, the approach of decision fusion and the proposed solution is discussed in detail.

A. DATASET OVERVIEW

This section demonstrates an extended version of the laparoscopic video quality (LVQ) database [22] that includes laparoscopic videos. It was proposed in the ICIP2020 challenge [34]. These videos have been carefully selected from the Cholec80 public dataset that has 80 different videos of cholecystectomies [42]. The extracted videos were selected considering multiple variations of scene content and have been distorted with either single or multiple distortions simultaneously at different levels. The five main distortions were additive white gaussian noise (AWGN), smoke, defocus blur, motion blur, and uneven illumination. Figure 2 illustrates a few samples of laparoscopic video frames. A summary of the dataset including the number of samples in the training and testing stages is shown in Table 1. In addition, the distortion categories including single and multiple distortions are listed in Table 2. There were two challenges in this dataset. The first challenge was an imbalanced dataset with various videos for each distortion (400 uneven illumination videos, 320 smoke videos, 300 AWGN videos, 160 defocus blur videos, and 80 motion blur videos) [34]. The second challenge was that each video was distorted by single or multiple distortions and thus, the problem of distortion

TABLE 1. Dataset summary.

Dataset	Number of Videos
Training Samples	800
Testing Samples	200
Total	1000

TABLE 2. Description of distortion categories [34].

Categories and Types of Distortions	
Single Distortion	**Multiple Distortion**
Noise	Defocus blur + Uneven illumination
Defocus blur	Noise + Smoke
Motion blur	Noise + Uneven illumination
Smoke	Smoke + Uneven illumination
Uneven illumination	Noise + Smoke + Uneven illumination

TABLE 3. Dataset description [34].

Number of reference videos (train + test)	20 + 5
Duration of each video	10 seconds
Resolution of each video	512 × 288
Frame rate	25 fps
Number of distortions	5
Number of levels	4
Number of distorted videos for training	800 (400 with single distortion only and 400 with multiple distortions)
Number of distorted videos for test	200 (80 with single distortion only and 120 with multiple distortions)

classification was formulated as a multi-label classification problem. Table 3 provides the description of the dataset. The training was done by frame to have a total of 20660 frames for model training (4132 frames for each class). In the testing stage, the majority voting method was used to select the class that represented the whole video. To balance the videos of defocus and motion blur, data were augmented by flipping the extracted frames vertically.

B. THE PROPOSED SOLUTION

1) SUPERVISED DEEP CNN LEARNING

An end-to-end learning mode was used to fine-tune the parameters of the whole CNN. During training, the images and labels were available. The network consists of convolutional, pooling, batch normalization, dropout, and fully connected (top) layers. This stage of feature learning aims to fit the large-scale ImageNet dataset and produce 1000 categories. The network parameters were tuned by the stochastic gradient descent (SGD) algorithm. This learning scheme is useful when large-scale datasets are available to overcome the overfitting problem. At the end of the training, the optimal parameters generated were ready to be transferred to a new small-scale dataset [43], [44].

2) PRETRAINED CNN AND CNN BASED FEATURE EXTRACTION

Transfer learning was demonstrated by training ResNet50, a deep CNN [33] with a large-scale dataset such as ImageNet [44]. It was used with a novel small-scale dataset such as that of the laparoscopic videos [34]. The objective is to utilize the parameters of the first layers of the CNN after removing top layers to extract spatial features from the laparoscopic frames and to transfer these features from natural images to medical images. Transfer learning was found to solve the lack of availability of large-scale data, which is a common problem in medical applications. Transfer learning can also speed up the learning process by tuning only a few top layers to fit the limited amount of new data.

3) RESIDUAL NETWORK

Many times, very deep CNNs suffer from the gradient vanishing problem which leads to a drop in accuracy [33]. To address this problem, the residual network (ResNet) uses skip connections instead of direct stacked layers [33]. ResNet is a well-known deep neural network with high generalization ability used for image recognition. Residual networks have various versions with different numbers of layers, such as ResNet50 with 50 layers and over 23 million trainable parameters [33]. In this paper, ResNet50 was utilized after removing the top layers to extract the features from laparoscopic frames before being applied to SVM classifiers.

4) SUPPORT VECTOR MACHINE (SVM)

SVM is a supervised learning classifier that has either a linear kernel function or a non-linear kernel function (e.g. Gaussian kernel, polynomial kernel) [45]. The inputs into the SVM are feature vectors extracted from ResNet50. The objective of an SVM is to separate the feature vectors to maximise the margins from both vectors. In this paper, various kernels of SVM were evaluated and a Gaussian kernel was selected to give the best performance in terms of accuracy. SVM was trained to fit the extracted features and map them to specific categories. Two types of SVMs were utilized in this work:

a: SVM-BASED MULTI-CLASS CLASSIFIER

This SVM was used to classify distortions in video frames into one of five categories. This worked well in videos affected by single distortions. For videos with multiple distortions, only one dominant distortion with the highest probability was produced. We utilized this classifier for defocus and motion blur. Few videos in the training set (80 videos) have single motion blur distortions. In other words, in the training set, there were no videos with motion blur that had other types of distortions in the same video. However, these types of videos existed in the testing sets. Therefore, an SVM-based five-class classifier was used to detect motion or defocus blur that existed along with other types of distortions in the video frame.

b: SVM-BASED BINARY CLASSIFIER

In multi-label or multi-distortion classification scenarios, five binary SVMs were used to classify each type of distortion. For example, to detect smoke distortion, a binary SVM was used to classify video frames into smoke or non-smoke classes. This set of five SVM binary classifiers works well in videos affected by multiple distortions to address multi-label classification problems. In the proposed solution, we utilized only three binary classifiers for noise, smoke, and uneven illumination distortions because the other two distortions (defocus and motion blur) were detected by SVM-based multi-class classifiers.

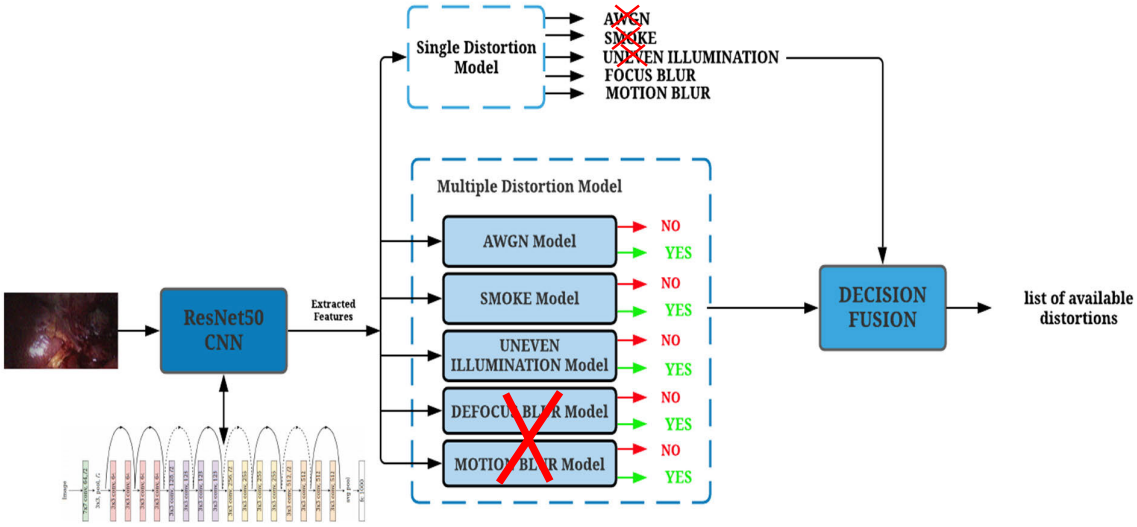


FIGURE 3. The detailed block diagram of the proposed distortion classification solution.

5) DECISION FUSION

This approach aims to combine decisions or predicted classes from several SVMs to make the final decision and list all available distortions in the laparoscopic video.

The five-class SVM produced a class C_{dis}, where $C_{dis} \in \{1, 2, 3, 4, 5\}$. On the other hand, five binary SVMs produced five classes: C_n, C_s, C_u, C_{db}, C_{mb}, where $C_n \in \{0, 1\}$, $C_s \in \{0, 1\}$, $C_u \in \{0, 1\}$, $C_{db} \in \{0, 1\}$, $C_{mb} \in \{0, 1\}$. The final decision is defined as follows:

1) Define empty distortion list
2) If ($C_{dis} == 4$):
   Add defocus blur to distortion list
   else if ($C_{dis} == 5$):
   Add motion blur to distortion list
   if ($C_n == 1$):
   Add noise to distortion list
   if ($C_s == 1$):
   Add smoke to distortion list
   if ($C_u == 1$):
   Add uneven illumination to distortion list

In this paper, the laparoscopic video was converted into frames by sampling at a rate of five frames per second. The length of laparoscopic videos in the dataset is 10 seconds, and thus 50 frames were used to represent one video. The video frames were resized to 224 × 224 pixels to be applied on the ResNet50 CNN input. This was done to extract the features from the frames. The number of elements in the feature vector is 2048. The majority voting method was used to label the videos with a distortion class that is frequently visible in the video frames. For instance, if 30 frames in the video contain noise, and 20 contain no noise, the video would be classified as having noise distortion.

The detailed block diagram and flow chart of the proposed solution for laparoscopic distortion classification are illustrated in Figure 3 and Figure 4, respectively.

III. RESULTS AND DISCUSSION

In this section, the experimental setup, performance metrics, and results are discussed in detail.

A. EXPERIMENTAL SETUP

The experiments were performed on a desktop computer (CPU: Intel Core i7 - 8700 @ 3.19 GHz; GPU: NVIDIA GeForce GTX 1080 Ti with 11 GB; and RAM of 64 GB) running on Windows 10 ×64. Various frameworks such as OpenCV [46], Scikit Learn [47], TensorFlow [48], and Keras [48] were utilized with Python.

B. PERFORMANCE METRICS

Several performance metrics were used to validate the performance of the proposed solution. Accuracy, precision, recall, and F1 score, are significant indicators that should be considered in validating the results of the distortion classification task.

The summary of performance metrics is as follows:

1) **Accuracy** is a measure that calculates the number of samples predicted correctly over all available samples.

$$\frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

where TP is True Positive, TN is True Negative, FP is False Positive, and FN is False Negative.

In the ICIP2020 challenge, accuracy was considered as one of the evaluation metrics to rank the top solutions. The final accuracy considers the accuracy of single and multiple distortions. If a video has multiple distortions and the set of

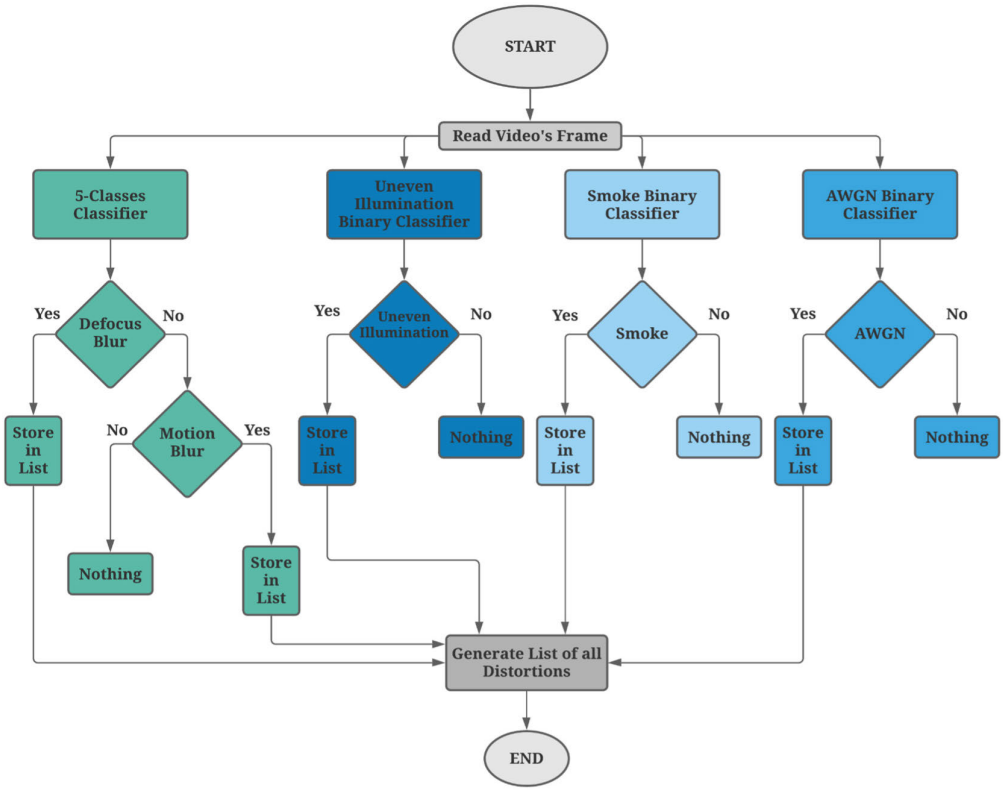


FIGURE 4. Flow chart of the proposed solution for laparoscopic distortion classification.

predicted classes belongs to the set of actual classes, the video is said to have been classified correctly. On the other hand, if one of the actual distortions is not available in the predicted set, then the video is said to have been misclassified.

2) **Precision** is a measure that calculates the number of samples predicted correctly to have specific distortions over the number of all samples predicted to exhibit specific distortions, both correctly and incorrectly.

$$\frac{TP}{TP + FP} \tag{2}$$

3) **Recall (Sensitivity)** is a measure that calculates the number of samples predicted correctly with specific distortions over the number of all samples with actual distortions.

$$\frac{TP}{TP + FN} \tag{3}$$

4) **F1 score** is a metric that summarizes recall and precision into a single quantity.

$$\frac{2 \times precision \times recall}{precision + recall} \tag{4}$$

1) **False positive rate or false alarm** is the ratio between the number of samples with no specific distortions wrongly predicted to exhibit a distortion and the total number of actual samples with no distortion.

$$\frac{FP}{FP + TN} \tag{5}$$

In the ICIP2020 challenge, F1 score was also considered as one of the evaluation metrics to rank the top solutions. There are two types: F1 score for single distortion if a video has only a single distortion, and F1 score for single + multi distortions if a video has single or multiple distortions.

C. EXPERIMENTAL RESULTS

This section discusses the evaluation results of the proposed method and other existing methods, such as the multi-label classification model which includes ResNet50 with five binary SVM classifiers, one for each distortion type. After that, the comparison was done with other solutions presented in the ICIP challenge in terms of accuracy, F1 score, and inference speed.

First, the results of the proposed solution evaluated on 200 videos in the testing set for each type of distortion were shown in Figure 5. The confusion matrix of each distortion including AWGN noise, defocus blur, motion blur, smoke, and uneven illumination was also illustrated. The proposed method was able to detect noise and defocus blur with accuracies of 100%. On the other hand, motion blur had the lowest detection accuracy because of the lack of training samples that had a combination of motion blur and other distortion

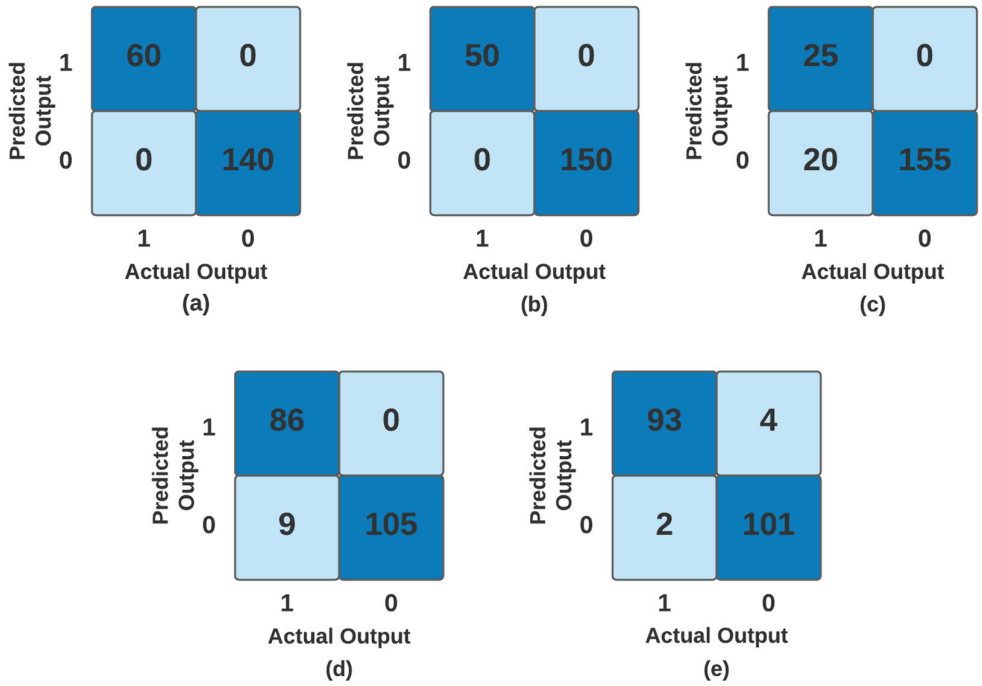


FIGURE 5. Confusion matrix of the proposed solution for each distortion: (a) AWGN noise, (b) defocus blur, (c) motion blur, (d) smoke, and (e) uneven illumination.

TABLE 4. Performance metrics of the proposed solution for each distortion.

Distortion	Accuracy %	Recall %	Precision %	F1-Score %	False Positive Rate %
AWGN noise	100	100	100	100	0
Defocus blur	100	100	100	100	0
Motion blur	90	55.55	100	71.42	0
Smoke	95.5	90.53	100	95.03	0
Uneven illumination	97	97.89	95.88	96.87	3.8
Average	**96.5**	**88.79**	**99.18**	**92.66**	**0.76**

types. However, testing samples had combinations of this nature. Table 4 shows the performance metrics of the proposed method for each distortion in terms of accuracy, recall, precision, F1-score, and false alarm rate. The advantage of the proposed solution is its very high value of average precision (99.18%) and low average false alarm rate (0.76 %).

It is obvious that the deep representation extracted from medical images such as laparoscopic frames utilizing the ResNet50 CNN is appropriate to be classified into various types of distortions.

D. ABLATION STUDY

This study aimed to validate the significance of the proposed solution summarized by the decision fusion of four SVMs including three binary SVMs and one multi-class SVM. A comparison with a multi-label classification model summarized by five binary SVMs was carried out to validate the results.

Figure 6 and Table 5 show the confusion matrix and performance metrics of multi-label classification model for each distortion, respectively.

The comparison between Figure 5 and Figure 6 from one side and between Table 4 and Table 5 from the second side was done. The proposed method was found to outperform multi-label classification models in terms of average accuracy by 3.5%, average F1 score by 2.4%, and average false alarm rate by 6%. The drawback of the multi-label classification model is the high average false alarm (7.06%) because the model detected distortions that are not existent in the videos. The defocus blur distortion was behind the drop in accuracy and the increase in false alarm rate in multi-label classification models. The previous problem was caused by the lack of samples with a combination of defocus blur and other distortions in the training set.

The previous results compared the proposed solution with a multi-label classification model by considering whether each

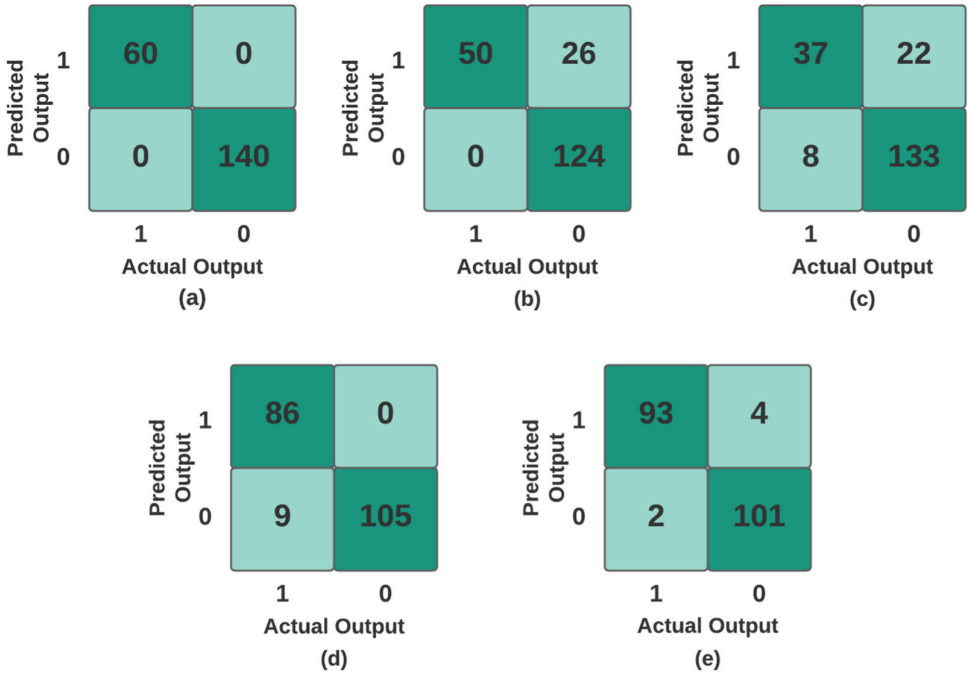


FIGURE 6. Confusion matrix of multi-label classification models for each distortion: (a) AWGN noise, (b) defocus blur, (c) motion blur, (d) smoke, and (e) uneven illumination.

TABLE 5. Performancemetrics of multi-label classification model for each distortion.

Distortion	Accuracy %	Recall %	Precision %	F1-Score %	False Positive Rate %
AWGN noise	100	100	100	100	0
Defocus blur	87	100	78.95	88.24	17.33
Motion blur	85	82.22	62.71	71.15	14.19
Smoke	95.5	90.53	100	95.03	0
Uneven illumination	97	97.89	95.88	96.87	3.8
Average	**92.9**	**94.13**	**87.51**	**90.26**	**7.06**

distortion is existent in each of the 200 testing videos. To give more insight into the performance of the proposed solution and its ability to address the problem of multi-label or multi-distortion classification, accuracy and F1 score were calculated considering all distortions available simultaneously in the video as shown in Table 6. For instance, if the actual video had three distortions, but the method was able to predict only two distortions, this video would be added to the list of videos that were misclassified. The baseline method was also a multi-label classification model (Resnet50 + Five Binary SVMs). The proposed decision fusion of Resnet50 + 4 SVMs was able to outperform the baseline in terms of accuracy that arrived at 83% compared with the 69.5% accuracy of the baseline. Additionally, the F1 score of single distortion was 94.7% which is higher than 90.91% of the baseline. Finally, the F1 score of both single and multi-distortions was found to be 94.9% compared with 90.18% of the baseline. The best values are shown in bold.

After comparing the proposed method with the multi-label classification model, the next step was to compare it with other solutions presented in the ICIP challenge in terms of accuracy, F1 score, and inference speed. The existing solutions were considered as baseline methods and summarized as follows:

1) THE FIRST BASELINE [49]

The solution that received the second prize used a VGG16 CNN [31] to extract features. The image was divided into two patches: Right and Left. These two patches were applied to the VGG16 CNN to extract feature maps. The feature maps extracted from the two patches using VGG16 were combined in one feature vector that was applied to a fully connected neural network. The neural network includes two hidden layers with 4096 nodes, two batch normalization layers, and two dropout layers. The output of this network were five types of distortions. The hyperparameters include binary

TABLE 6. Comparison between the proposed solution and the multi-label classification model.

Method	Accuracy %	F1-score (Single Distortion) %	F1-score (Single + Multi Distortions) %
Decision Fusion Resnet50+ 4 SVMs (Proposed)	**83.0**	**94.7**	**94.9**
Resnet50+ Five Binary SVMs (Baseline)	69.5	90.91	90.18

TABLE 7. Comparison between the proposed solution and other existing methods in terms of accuracy, F1-score of single-distortion, and F1-score of single + multi-distortions [34], [49].

Method	F1-score (Single + Multi Distortions) %	F1-score (Single Distortion) %	Accuracy %
Decision Fusion ResNet50 + 4 SVMs (Proposed)	**94.9**	**94.7**	**83.0**
First baseline VGG16 + many FM + FC [34,49]	94.1	93.3	81.5
Second baseline VGG16 + 5 FC [34,49]	93.3	90.7	78.0
Baseline [34]	91.5	88.0	76.5
Baseline [34]	85.4	98.7	58.0
Baseline [34]	83.2	89.3	57.0

cross entropy as a loss function, and Adam as an optimizer. In addition, a downhill simplex algorithm was used to find the optimal threshold to optimize the classification stage. This description of the solution was given by the winners in the ICIP challenge presentation event.

2) THE SECOND BASELINE [49]

The solution that received the third prize used a deep multitask learning model. It includes one shared feature extraction block and five independent binary classifiers (one for each distortion type). The shared feature extraction used the VGG16 CNN [31] pre-trained with ImageNet after removing fully connected layers from the top layers. In addition, five blocks of binary classifiers were added after the feature extraction block. Each classifier has two fully connected layers with 512 nodes and one node in the output layer with a sigmoid activation function. The whole network was trained in an end-to-end theme after freezing parameters of the feature extraction block and fine-tuning only parameters of layers in the five classifiers. The total loss function is the sum of the binary cross entropy loss functions for each classifier. This description of the solution was also given by winners in the ICIP challenge presentation event.

The performance metrics of the proposed solution and the baselines mentioned were shown in descending order (from the top to the last) in Table 7. The best values are shown in bold. These metrics were also shown in the leader board of the ICIP2020 challenge [34]. Unfortunately, not all methods were described. Only the methods of the top three solutions were presented in the event.

After evaluating the proposed method in terms of accuracy and F1 score, two other evaluation metrics that were used to rank the top solutions in the ICIP2020 challenge were inference time and speed. Table 8 shows the inference time and speed of the proposed solution which took 0.05 seconds calculated from the time step of applying the video frame at the input and extracting the features, until the time step of producing the distortion types at the output. Therefore, the model was able to scan video frames with a speed of 20 FPS. The comparison with other top solutions is shown in Table 8. Figure 7 shows a few screenshots of the graphical user interface (GUI) of the demonstration.

TABLE 8. Inference time and speed of the proposed solution compared with other methods [34].

Method	Average Time Per Frame (Seconds)	Speed (FPS)
Decision Fusion (Resnet50 + 4 SVMs) (Proposed)	**0.05**	**20**
VGG16 + many FM + FC [34,49] (Baseline)	0.104	~ 9.62
VGG16 + 5 FC [34,49] (Baseline)	0.05	20

In summary of the results, the work presented in this paper was found to produce a solution that is fast and accurate when used for distortion classification in laparoscopic videos.

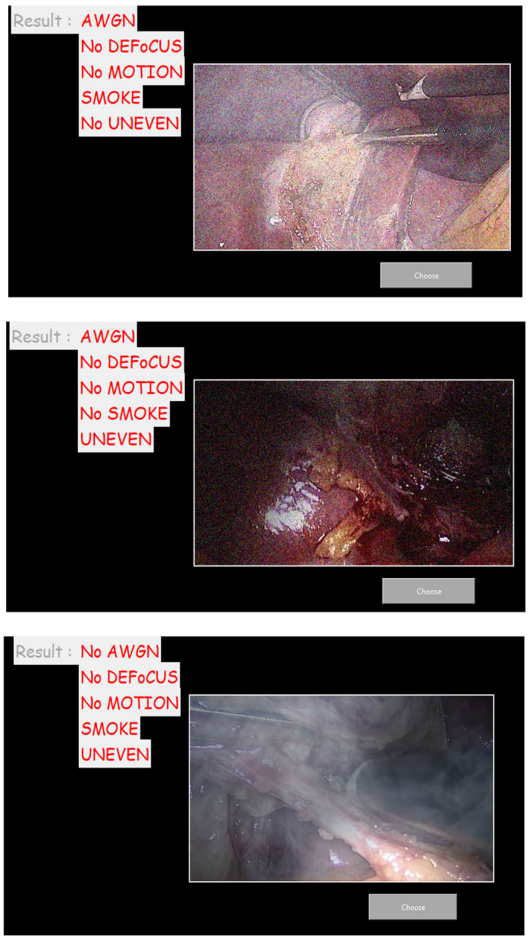


FIGURE 7. Graphical user interface (GUI) of the proposed solution.

The advantages of the proposed solution are as follows:

1. It is robust in automatically classifying five types of distortions including AWGN noise, smoke, uneven illumination, defocus blur, and motion blur that exist simultaneously in multi-distortion laparoscopic videos.
2. It runs in real-time at a speed of 20 FPS. In other words, only one frame is enough to classify the distortion. This helps to identify distortions in live-captured laparoscopic videos.

The limitation of the proposed solution is its inability to recognize ''Defocus Blur'' and ''Motion Blur'' simultaneously. The reason is that ''Defocus Blur'', and ''Motion Blur'' distortions were produced at the output of the multi-class classifier. Nonetheless, the proposed solution was found to outperform other methods.

IV. CONCLUSION AND FUTURE WORK

In this paper, a problem of multi-label distortion classification in laparoscopic videos was addressed.

We utilized a decision fusion approach, which combines decisions from multiple classifiers including multi-class classifiers and three binary classifiers. A pre-trained ResNet50 CNN was utilized to transfer representation from ImageNet natural images to the frames of laparoscopic videos. The CNN-based extracted features were classified by a multiclass SVM into one of five categories or distortions and by a binary classifier into one of two categories, such as smoke and no-smoke. The proposed decision fusion approach was found to improve the performance of distortion classification in terms of Accuracy (83%), F1 score of single distortion (94.7%), and F1 score of single and multi-distortion (94.9%). The performance was evaluated on a laparoscopic video dataset that included 800 videos for training and validation and 200 videos for testing.

This work focused on the utilization of a ResNet50 CNN to extract spatial features from still laparoscopic frames. In addition, it used SVM for frame-level classification. There are still opportunities to utilize other CNNs such as EfficientNet [50] which was found to outperform ResNet50 in some applications [51]. Additionally, temporal information can also be extracted from laparoscopic videos utilizing long short-term memory (LSTM) [52].

The current solution transfers deep representation from natural images of the ImageNet dataset that contains 1000 categories to laparoscopic videos. Hence in the future, we can further improve performance by fine-tuning more layers of the ResNet CNN with laparoscopic images, but this entails collection of more laparoscopic videos affected by distortions.

The solution proposed in this paper can classify five types of distortions in laparoscopic videos. Hence, in the future, we intend to enhance this work to consider the problem of distortion ranking to identify the intensity or levels of severity of each distortion. This is important for quality score evaluation in VQA systems.

ACKNOWLEDGMENT

The authors would like to acknowledge the contributions of each author. Conceptualization was done by Nouar Aldahoul and Jamie Ledesma Fermin; data curation by Nouar Aldahoul; formal analysis by Nouar Aldahoul; funding acquisition by Hezerul Abdul Karim; investigation by Nouar Aldahoul and Myles Joshua Toledo Tan; methodology by Nouar Aldahoul and Hezerul Abdul Karim; project administration by Hezerul Abdul Karim; software by Nouar Aldahoul; validation by Nouar Aldahoul and Myles Joshua Toledo Tan; visualization by Nouar Aldahoul; writing–original draft preparation by Nouar Aldahoul and Jamie Ledesma Fermin; and writing–review and editing by Nouar Aldahoul, Hezerul Abdul Karim, Myles Joshua Toledo Tan, and Jamie Ledesma Fermin. The research work and solution presented in this article won the first prize in the ICIP2020 Challenge that was organized by Université Sorbonne Paris Nord, France; Norwegian University of Science and Technology, Norway; and Oslo University Hospital, Norway.

REFERENCES

[1] S. D. St Peter and G. W. Holcomb, ''History of minimally invasive surgery,'' in *Atlas Pediatric Laparoscopy Thoracoscopy*, G. W. Holcomb, K. E. Georgeson S. S. Rothenerg, Eds. Philadelphia, PA, USA: Elsevier, 2008, pp. 1–5.

[2] H. Ellis, "The first successful elective laparotomy," *J. Perioperative Pract.*, vol. 25, no. 10, pp. 207–208, Oct. 2015, doi: 10.1177/175045891502501005.

[3] F. Keus, J. A. F. de Jong, H. G. Gooszen, and C. J. H. M. van Laarhoven, "Laparoscopic versus open cholecystectomy for patients with symptomatic cholecystolithiasis," *Cochrane Database Syst. Rev.*, vol. 4, Oct. 2006, Art. no. CD006231, doi: 10.1002/14651858.CD006231.

[4] J. M. Cozar and M. Tallada, "Open partial nephrectomy in renal cancer: A feasible gold standard technique in all hospitals," *Adv. Urol.*, vol. 2008, Jan. 2008, Art. no. 916463, doi: 10.1155/2008/916463.

[5] K. Vagenas, P. Spyrakopoulos, M. Karanikolas, G. Sakelaropoulos, I. Maroulis, and D. Karavias, "Mini-laparotomy cholecystectomy versus laparoscopic cholecystectomy: Which way to go?" *Surg. Laparoscopy, Endoscopy Percutaneous Techn.*, vol. 16, no. 5, pp. 321–324, Oct. 2006, doi: 10.1097/01.sle.0000213720.42215.7b.

[6] V. Kuzinkovas and R. Singhal, "Laparotomy: Still the gold standard," *Minerva Chirurgica*, vol. 63, no. 1, p. 69, Feb. 2008.

[7] A. Ng, N. Wang, and M. Tran, "Minimally invasive surgery: Early concepts to gold standards," *Brit. J. Hospital Med.*, vol. 80, no. 9, pp. 494–495, Sep. 2019, doi: 10.12968/hmed.2019.80.9.494.

[8] N. J. Soper, P. T. Stockmann, D. L. Dunnegan, and S. W. Ashley, "Laparoscopic cholecystectomy the new'gold standard?" *Arch. Surg.*, vol. 127, no. 8, pp. 917–921 and 921–923, Aug. 1992, doi: 10.1001/archsurg.1992.01420080051008.

[9] C. D. Smith, C. J. Weber, and J. R. Amerson, "Laparoscopic adrenalectomy: New gold standard," *World J. Surg.*, vol. 23, no. 4, pp. 389–396, Apr. 1999, doi: 10.1007/pl00012314.

[10] K. E. Georgeson, R. D. Cohen, A. Hebra, J. Z. Jona, D. M. Powell, S. S. Rothenberg, and E. P. Tagge, "Primary laparoscopic-assisted endorectal colon pull-through for Hirschsprung's disease: A new gold standard," *Ann. Surg.*, vol. 229, no. 5, p. 678, May 1999, doi: 10.1097/00000658-199905000-00010.

[11] M. Heinzelmann, H. P. Simmen, A. S. Cummins, and F. Largiadár, "Is laparoscopic appendectomy the new'gold standard?" *Arch. Surg.*, vol. 130, no. 7, pp. 782–785, Jul. 1995, doi: 10.1001/archsurg.1995.01430070104022.

[12] X. Li, J. Zhang, L. Sang, W. Zhang, Z. Chu, X. Li, and Y. Liu, "Laparoscopic versus conventional appendectomy meta-analysis of randomized controlled trials," *BMC Gastroenterol.*, vol. 10, no. 1, p. 129, Nov. 2010, doi: 10.1186/1471-230X-10-129.

[13] N. Butler, S. Collins, B. Memon, and M. A. Memon, "Minimally invasive esophagectomy: Current status and future direction," *Surg. Endoscopy*, vol. 25, no. 7, pp. 2071–2083, Jul. 2011, doi: 10.1007/s00464-010-1511-2.

[14] M. Watanabe, Y. Baba, Y. Nagai, and H. Baba, "Minimally invasive esophagectomy for esophageal cancer: An updated review," *Surg. Today*, vol. 43, no. 3, pp. 237–244, Mar. 2013, doi: 10.1007/s00595-012-0300-z.

[15] M. J. Peters, A. Mukhtar, R. M. Yunus, S. Khan, J. Pappalardo, B. Memon, and M. A. Memon, "Meta-analysis of randomized clinical trials comparing open and laparoscopic anti-reflux surgery," *Amer. J. Gastroenterol.*, vol. 104, no. 6, pp. 1548–1561, Jun. 2009, doi: 10.1038/ajg.2009.176.

[16] M. A. Memon, M. S. Subramanya, M. B. Hossain, R. M. Yunus, S. Khan, and B. Memon, "Laparoscopic anterior versus posterior fundoplication for gastro-esophageal reflux disease: A meta-analysis and systematic review," *World J. Surg.*, vol. 39, no. 4, pp. 981–996, Apr. 2015, doi: 10.1007/s00268-014-2889-0.

[17] L. S. Feldman, "Laparoscopic splenectomy: Standardized approach," *World J. Surg.*, vol. 35, no. 7, pp. 1487–1495, Jul. 2011, doi: 10.1007/s00268-011-1059-x.

[18] D. Qian, Z. He, J. Hua, J. Gong, S. Lin, and Z. Song, "Hand-assisted versus conventional laparoscopic splenectomy: A systematic review and meta-analysis: HALS versus CLS in splenectomy," *ANZ J. Surg.*, vol. 84, no. 12, pp. 915–920, Dec. 2014, doi: 10.1111/ans.12597.

[19] A. Herrmann and R. L. De Wilde, "Laparoscopic myomectomy—The gold standard," *Gynecol. Minimally Invasive Therapy*, vol. 3, no. 2, pp. 31–38, May 2014, doi: 10.1016/j.gmit.2014.02.001.

[20] D. E. Pittaway, P. Takacs, and P. Bauguess, "Laparoscopic adnexectomy: A comparison with laparotomy," *Amer. J. Obstetrics Gynecol.*, vol. 171, no. 2, pp. 385–391, Aug. 1994, doi: 10.1016/S0002-9378(94)70039-7.

[21] L. Chen, J. Ding, and K. Hua, "Comparative analysis of laparoscopy versus laparotomy in the management of ovarian cyst during pregnancy," *J. Obstetrics Gynaecol. Res.*, vol. 40, no. 3, pp. 763–769, Mar. 2014, doi: 10.1111/jog.12228.

[22] Z. A. Khan, A. Beghdadi, F. A. Cheikh, M. Kaaniche, E. Pelanis, R. Palomar, Å. A. Fretland, B. Edwin, and O. J. Elle, "Towards a video quality assessment based framework for enhancement of laparoscopic videos," *Proc. SPIE Med. Imag. Image Perception, Observer Perform. Technol. Assessment*, vol. 11316, Mar. 2020, Art. no. 113160P, doi: 10.1117/12.2549266.

[23] J. Zhou and S. Payandeh, "Visual tracking of laparoscopic instruments," *J. Autom. Control Eng.*, vol. 2, no. 3, pp. 234–241, 2014, doi: 10.12720/joace.2.3.234-241.

[24] E. G. G. Verdaasdonk, L. P. S. Stassen, M. van der Elst, T. M. Karsten, and J. Dankelman, "Problems with technical equipment during laparoscopic surgery: An observational study," *Surg. Endoscopy*, vol. 21, no. 2, pp. 275–279, Feb. 2007, doi: 10.1007/s00464-006-0019-2.

[25] M. Siddaiah-Subramanya, M. Nyandowe, and K. W. Tiang, "Technical problems during laparoscopy: A systematic method of troubleshooting for surgeons," *Innov. Surg. Sci.*, vol. 2, no. 4, pp. 233–237, Aug. 2017, doi: 10.1515/iss-2017-0031.

[26] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, pp. 541–551, Jan. 1989.

[27] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[28] M. Pak and S. Kim, "A review of deep learning in image recognition," in *Proc. 4th Int. Conf. Comput. Appl. Inf. Process. Technol. (CAIPT)*, Aug. 2017, pp. 1–3, doi: 10.1109/CAIPT.2017.8320684.

[29] N. Aloysius and M. Geetha, "A review on deep convolutional neural networks," in *Proc. Int. Conf. Commun. Signal Process. (ICCSP)*, Apr. 2017, pp. 588–592, doi: 10.1109/ICCSP.2017.8286426.

[30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. 25th Int. Conf. Neural Inf. Process. Syst.*, vol. 1, Red Hook, NY, USA, 2012, pp. 1097–1105.

[31] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[32] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE erence Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[34] A. Beghdadi. (2020). *Real-Time Distortion Classification in Laparoscopic Videos ICIP*. Accessed: Jul. 3, 2021. [Online]. Available: https://2020.ieeeicip.org/challenge/real-time-distortion-classification-in-laparoscopic-videos/

[35] A. Chetouani, A. Beghdadi, and M. Deriche, "A hybrid system for distortion classification and image quality evaluation," *Signal Process., Image Commun.*, vol. 27, no. 9, pp. 948–960, Oct. 2012, doi: 10.1016/j.image.2012.06.001.

[36] S. Hu, L. Deng, and C.-C.-J. Kuo, "A new distortion/content-dependent video quality index (DCVQI)," in *Proc. Picture Coding Symp. (PCS)*, Dec. 2013, pp. 197–200, doi: 10.1109/PCS.2013.6737717.

[37] L. Lévêque, W. Zhang, C. Cavaro-Ménard, P. Le Callet, and H. Liu, "Study of video quality assessment for telesurgery," *IEEE Access*, vol. 5, pp. 9990–9999, 2017, doi: 10.1109/ACCESS.2017.2704285.

[38] S. Petscharnig and K. Schöffmann, "Learning laparoscopic video shot classification for gynecological surgery," *Multimedia Tools Appl.*, vol. 77, no. 7, pp. 8061–8079, Apr. 2018, doi: 10.1007/s11042-017-4699-5.

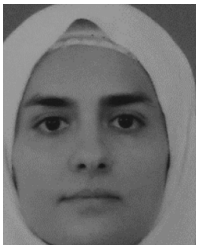
[39] C. Wang, F. Alaya Cheikh, M. Kaaniche, A. Beghdadi, and O. J. Elle, "Variational based smoke removal in laparoscopic images," *Biomed. Eng. Line*, vol. 17, no. 1, Dec. 2018, Art. no. 139, doi: 10.1186/s12938-018-0590-5.

[40] D. Varga, "No-reference video quality assessment based on the temporal pooling of deep features," *Neural Process. Lett.*, vol. 50, no. 3, pp. 2595–2608, Dec. 2019, doi: 10.1007/s11063-019-10036-6.

[41] Z. A. Khan, A. Beghdadi, M. Kaaniche, and F. A. Cheikh, "Residual networks based distortion classification and ranking for laparoscopic image quality assessment," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2020, pp. 176–180, doi: 10.1109/ICIP40778.2020.9191111.

[42] A. P. Twinanda, S. Shehata, D. Mutter, J. Marescaux, M. de Mathelin, and N. Padoy, "EndoNet: A deep architecture for recognition tasks on laparoscopic videos," *IEEE Trans. Med. Imag.*, vol. 36, no. 1, pp. 86–97, Jan. 2016.

[43] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015, doi: 10.1038/nature14539.
[44] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255, doi: 10.1109/CVPR.2009.5206848.
[45] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intell. Syst. Appl.*, vol. 13, no. 4, pp. 18–28, Jul. 1998.
[46] (Jun. 2020). *Open CV*. [Online]. Available: https://docs.opencv.org/master/
[47] (Jun. 2020). *Scikit-Learn: Machine Learning in Python-Scikit-Learn 0.24.0 Documentation*. [Online]. Available: https://scikit-learn.org/stable/.
[48] (Jun. 2020). *TensorFlow Core V2.5.0 Documentation*. [Online]. Available: https://www.tensorflow.org/api_docs/python/tf/keras/applications/resnet
[49] (2020). *ICIP2020 Challenge Presentation Event*. [Online]. Available: https://drive.google.com/file/d/1QI4-8vzjlcq6tsBYzDv7LWDtN8PeTZCB/view?usp=sharing
[50] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," 2019, *arXiv:1905.11946*. [Online]. Available: http://arxiv.org/abs/1905.11946
[51] N. Aldahoul, H. A. Karim, M. H. L. Abdullah, and A. S. Ba Wazir, "An evaluation of traditional and CNN-based feature descriptors for cartoon pornography detection," *IEEE Access*, vol. 9, pp. 39910–39925, 2021, doi: 10.1109/ACCESS.2021.3064392.
[52] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.



NOUAR ALDAHOUL received the B.Eng. and M.Eng. degrees in computer engineering from Damascus University, in 2008 and 2012, respectively, and the Ph.D. degree in machine learning from International Islamic University Malaysia, in 2019. She is currently a Researcher with the Faculty of Engineering, Multimedia University, Malaysia. Her main research interests include deep learning, computer vision, and the Internet of Things. She was a recipient of several awards, such as the ICIP2020 Challenge Award and the Best Conference Paper Award.

HEZERUL ABDUL KARIM (Senior Member, IEEE) received the B.Eng. degree in electronics from the University of Wales Swansea, U.K., in 1998, with a focus on communications, the M.Eng. degree in science from Multimedia University, Malaysia, in 2003, and the Ph.D. degree from the University of Surrey, U.K., in 2008. He is currently an Associate Professor with the Faculty of Engineering, Multimedia University. His research interests include telemetry, error resilience and multiple description video coding for 2D/3D image/video coding and transmission, and content-based image/video recognition. He is also serving as a Treasurer for the IEEE Signal Processing Society Malaysia Chapter.

MYLES JOSHUA TOLEDO TAN (Member, IEEE) was born in Bacolod, Philippines, in 1996. He received the B.S. degree *(summa cum laude)* in biomedical engineering from University at Buffalo, The State University of New York, in 2017, and the M.S. degree in applied biomedical engineering from Johns Hopkins University, Baltimore, MD, USA, in 2018. He has been an Assistant Professor of chemical engineering with the University of St. La Salle (USLS), since 2018, where he was appointed as an Assistant Professor of natural sciences, in 2020. He has been actively involved in the education and training of students with the Department of Electronics Engineering and the Department of Electrical Engineering, USLS. He also leads Tan Research Group. His research interests include biomedical signal processing, medical imaging, deep learning, and engineering and mathematics education. He is a member of the Institute of Physics, U.K., and Tau Beta Pi—The Engineering Honor Society (USA); and an Associate Member of the Institute of Mathematics and its Applications, U.K. He was a recipient of the Tau Beta Pi Engineering Honor Society Record Scholarship and Grace Capen Award.

JAMIE LEDESMA FERMIN (Student Member, IEEE) was born in Bacolod, Philippines, in 1999. He is currently pursuing the bachelor's degree with the Electronics Engineering Program, University of St. La Salle (USLS), Bacolod. He has been an Undergraduate Research Assistant with Tan Research Group, USLS, since 2018. His research interests include biomedical signal processing, medical imaging, and deep learning. He is a Student Member of the Institute of Mathematics and its Applications, U.K.