Residual Network (ResNet-18) for Laparoscopic Image Distortion Classification

Mohamed Belmokeddem, Kamila Khemis and Salim Loudjedi

December 19, 2023

Residual network (resnet-18) for laparoscopic image distortion classification

Belmokeddem Mohamed
Biomedical Engineering Department, Faculty of Technology University Abou-bekr Belkaid of Tlemcen
Tlemcen, Algeria
mohamed.belmokeddem@univ-tlemcen.dz

khemis kamila
Biomedical Engineering Department, Faculty of Technology University Abou-bekr Belkaid of Tlemcen
Tlemcen, Algeria
kamila.khemis@univ-tlemcen.dz.

Loudjedi Salim
Surgery B Tlemcen Hospital, Department of medicine University Abou-bekr Belkaid of Tlemcen
Tlemcen, Algeria
ismatsalim.loudjdimouedden@univ-tlemcen.dz

Abstract— **Diminished laparoscopic video quality directly affects a surgeon's visibility and can compromise the outcomes of computational tasks in robot-assisted surgery. To address this challenge, numerous solutions have been proposed based on the detection and classification of laparoscopic video distortions.**
In this work, we propose a method based on Residual networks (ResNet18) for the automatic detection and classification of noise 'NO', smoke 'SM', uneven illumination 'UI', defocus blur 'DB', and motion blur 'MB' in laparoscopic videos. We have obtained an accuracy of 98.75% for training and 97.97% for validation.
The high accuracy scores across the classes emphasize the model's capability to generalize well and make accurate predictions.

Keywords—Laparoscopic video, distortion classification, deep learning.

I. INTRODUCTION

Guaranteeing Adequate Video Quality is a Vital Prerequisite for Uninterrupted Minimally Invasive Surgery (MIS) Procedures. Laparoscopic Videos Are Prone to Various Types of distortions, encompassing Noise, Smoke, Uneven Illumination, Defocus Blur, and Motion Blur. These diverse distortions comprise the five categories of laparoscopic video distortion [1]. In response to the challenge of laparoscopic video distortion, researchers established the Laparoscopic Video Quality Database (LVQ). This database includes a collection of 200 videos, featuring five distinct distortion categories and four different intensity levels [1].

Numerous deep-learning techniques have been applied to detect and categorize these five distortion categories. These methods encompass the utilization of a single Deep Neural Network (DNN) [1], Convolutional Neural Networks like ResNet50, as well as the integration of ResNet and a Fully Connected Neural Network (FCNN) [2].

In this study, we employ Residual Networks (ResNet18) for the automated detection and classification of these distortions. Our computational framework was initially trained on an extended version of the LVQ database, referred to as the LVQ Challenge dataset.

II. MATERIALS AND METHODS

A. Dataset description

This paper utilizes the LVQ database Challenge, as developed by Khan, Beghdadi, Cheikh, et al. [1] [3]; This database consists of 20 reference videos, each spanning 10 seconds in duration. These reference videos have been extracted from the Cholec80 dataset, originally introduced by Twinanda et al. [4]. The Cholec80 dataset encompasses ten distinct categories of scene variations, including bleeding (BL), grasping and burning (GB), multiple instruments (MI), irrigation (IR), clipping (CL), stretching away (SA), cutting (CU), stretching forward (SF), organ extraction (OE), and burning (BU). Each of the reference videos has been intentionally subjected to five different categories of distortions, with four varying levels of severity as shown in Table 1.
This extensive dataset encompasses a total of 400 videos with single distortion and 400 videos with multiple distortions. The five primary categories of distortion within this dataset include smoke, noise, uneven illumination, defocus blur, and motion blur.

In our study, we specifically focused on utilizing the videos from the five single distortion classes to train our model. Each of these single distortion classes consists of 80 videos, resulting in a balanced and comprehensive dataset.

To evaluate the performance of the proposed methodology, we have used a total of 20660 frames (4132 frames for each class) extracted from the five laparoscopic video classes of LVQ dataset with a rate of 5 fps, as shown in Table 2.
We resize video frames to 224*224 pixels to apply them to ResNet-18. Eighty percent, (80%) of image, data from each class (3305 images) was used for training and 20% of images were reserved for testing.

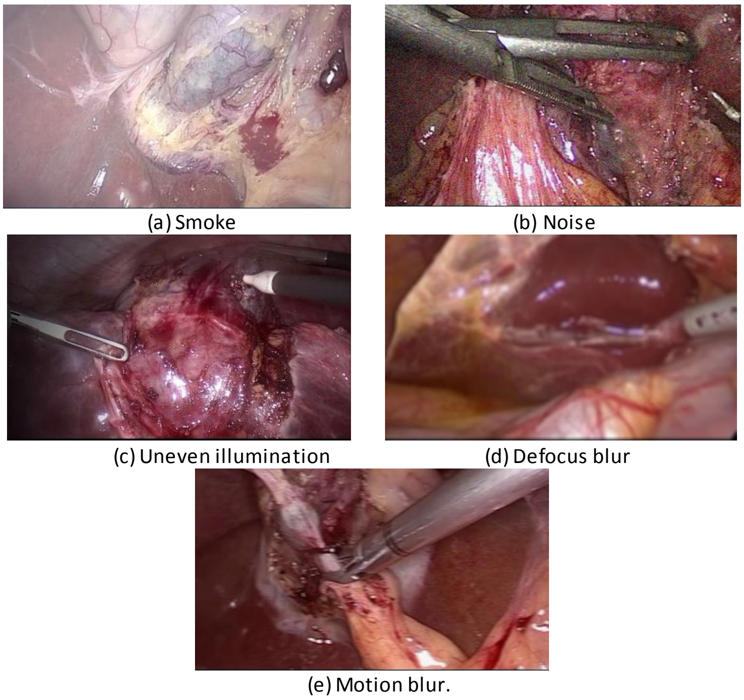

Fig.1. Distorted images extracted from the LVQ database [1]

TABLE I. DATASET DESCRIPTION [1]

Resolution of video	**Frame rate**	**Number of distortions**	**Number of levels**
512 X 288	25 fps	5	4

TABLE II. TRAINING DATA SUMMARY

Dataset	***Number of frames***
Number of defocus blurred frames	4132
Number of motion blurred frames	4132
Number of smoked frames	4132
Number of uneven illumination frames	4132
Number of noised frames	4132
Number of total frames	20660

B. Methods

In this paper, we harness the ResNet architecture to tackle the classification of laparoscopic video distortions. ResNet-18, a convolutional neural network renowned for its depth, consists of 18 layers and boasts 2.37 million trainable parameters. This architecture is renowned for its remarkable generalization capabilities [5]. Given its prowess, we have opted to leverage the ResNet architecture, specifically replacing the output fully connected layer with a new layer containing five units. This alignment with the number of classes within our dataset ensures a tailored approach.

For model training, we utilize a laparoscopic dataset and employ the cross-entropy loss function, a commonly used metric in classification tasks. The cross-entropy loss function is minimized during training. We accomplish this using the Adam optimizer with a learning rate of 0.01. Our training framework is implemented within the PyTorch framework and executed on an NVIDIA T1000 GPU with 12 GB of memory. During model training, we utilize mini-batches with a size of 10 and conduct training over 50 epochs.

III. RESULTS

For evaluating the performance of distortion classification, we have evaluated the classification performance of our proposed architecture. As shown in Fig.2, the matrix of confusion of the classification by class. Table 3 shows the accuracy of the classification of our model by class. The distortions present in the LVQ database are denoted as follows: SM for smoke, WN for noise, UI for uneven illumination, DB for blur due to defocus, and MB for blur due to motion.

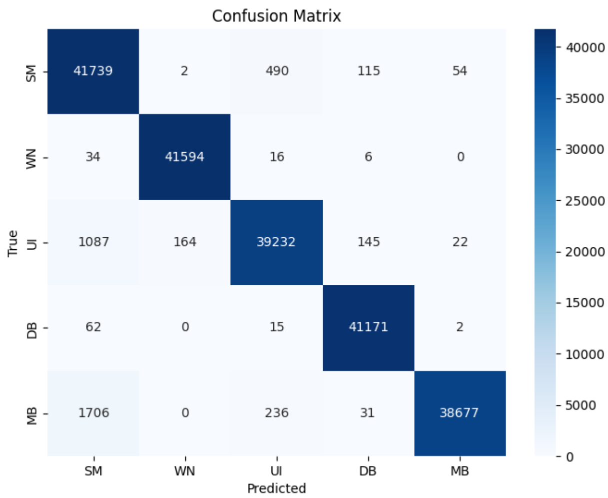

Fig.2. Confusion matrix of the proposed solution

TABLE III. PREDICTION RATE BY DISTORTION'S CLASS

Class	***Smoke***	***Noise***	***Uneven illumination***	***Defocus blur***	***Motion blur***
Prediction rate %	98.44	99.86	96.51	99.80	95.14

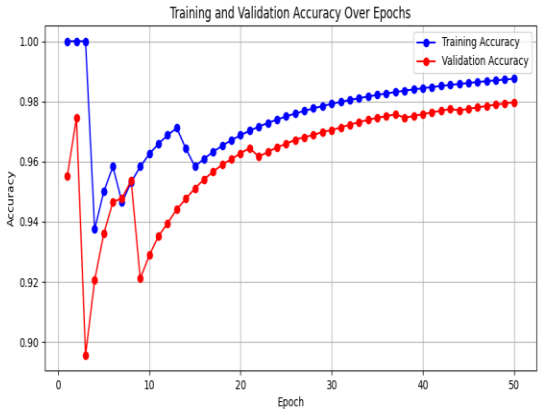

Fig. 3. Accuracy evolution with the number of epochs for our approach

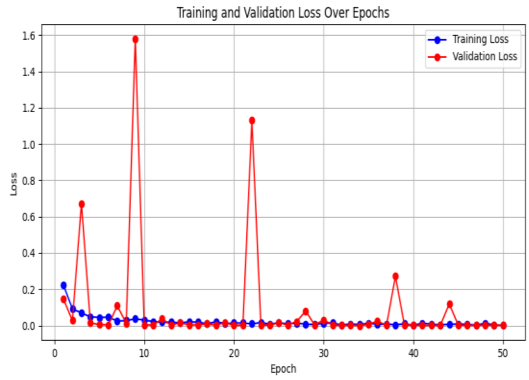

Fig. 4. Loss function evolution with the number of epochs for our approach

IV. DISCUSSION

The results of using this approach are highly encouraging, and exhibit impressive prediction rate scores:

Smoke ('SM'): Achieving a prediction rate of 98.44% for the 'SM' class is remarkable. This high value suggests that the model excels at correctly identifying frames affected by smoke, which is a critical factor in laparoscopic video quality assessment.

Noise ('WN'): With a prediction rate of 99.86% for the 'WN' class, the model shows exceptional performance in recognizing noisy frames. Minimal misclassifications in this category indicate the model's robustness in handling noise-related distortions.

Uneven Illumination ('UI'): A prediction rate of 96.51% for the 'UI' class reflects the model's ability to distinguish frames with uneven illumination. Though slightly lower than the previous classes, this score remains impressive.

Defocus Blur ('DB'): The 'DB' class attains an outstanding prediction rate of 99.80%. This highlights the model's proficiency in identifying frames affected by defocus blur, which is a significant factor in laparoscopic video quality assessment.

Motion Blur ('MB'): With a prediction rate of 95.14% for the 'MB' class, the model demonstrates its capability to classify frames exhibiting motion blur accurately. Although the prediction rate is slightly lower than some other classes, it remains a commendable performance.

These results collectively indicate that the ResNet-18 model is highly effective in recognizing and classifying frames affected by various laparoscopic video distortions. The high prediction rates across the classes emphasize the model's capability to generalize well and make accurate predictions, even when faced with different types of distortions commonly encountered in laparoscopic videos.

Moreover, the balanced distribution of frames across classes, with 4132 samples per class, ensures that the model is not biased toward any specific category. This balanced dataset contributes to the model's capacity to provide reliable predictions across all classes.

To show the interest of this approach, Fig.3 illustrates the accuracy evolution of both training and validation datasets over epochs. Thus, the convenience of curves can be observed. At the number of epochs=50, we have a Training Accuracy of 98.75% and a Validation Accuracy of 97.97%.

Fig. 4 illustrates the loss functions evaluated on the training and validation datasets over epochs. Thus, it can be observed the convergence of the loss validation and loss training to zero.

V. CONCLUSION

The primary objective of this study was the classification of laparoscopic video distortions. To achieve this, we utilized the ResNet18 architecture.

The results obtained, particularly in terms of accuracy, are not only remarkable but also highly encouraging. They provide a strong foundation for expanding the process of detecting distortions in laparoscopic videos.

ACKNOWLEDGMENT

There are no conflicts to declare.

No funding was received to conduct this research and publication of this article.

All authors have seen and approved the manuscript.

REFERENCES

[1] Z.A.Khan et al., "Towards a video quality assessment based framework for enhancement of laparoscopic videos". J. Med. Imaging 11316, 2020.

[2] ZA.Khan, A.Beghdadi, M.Kaaniche, FA.Cheikh, O.Gharbi , "A neural network based framework for effective laparoscopic video quality assessment", Comput. Med. Imaging Graph 101, 2022.

[3] Khan, Z.A., Beghdadi, A., Kaaniche, M. and Cheikh, F.A., 2020, October. Residual networks based distortion classification and ranking for laparoscopic image quality assessment. In 2020 IEEE International Conference on Image Processing (ICIP) (pp. 176-180). IEEE.

[4] Twinanda .P, Shehata S, Mutter D, Marescaux J, Mathelin M, Padoy N. EndoNet: A Deep Architecture for Recognition Tasks on Laparoscopic Videos. IEEE Trans. on Medical Imaging 2016.

[5] He K, Zhang X, Ren S, Sun J, Deep residual learning for image recognition, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778. Doi: 10.1109/CVPR.2016.90.