Automatically Monitoring the Colonoscopic video quality

Priyansh Shah

***Abstract*— Colonoscopy is the primary method for screening and prevention of colon cancer. The goal of colonoscopy is to find and remove polyps before they develop into colon cancer. However, colonoscopy is not a perfect procedure. The existing guidelines for a good colonoscopy suggest to maintain a minimum withdrawal time of 6 minutes. However, it is not adequate to guarantee the quality of colonoscopy. This report presents two deep learning architectures to measure the informativeness of colonoscopic images. Results show that both these architectures perform well on the data.**

I. INTRODUCTION

Colonoscopy is the primary method for screening and prevention of colon cancer, during which a tiny camera is inserted and guided through the colon. The goal of the colonoscopy is to find and remove polyps before they develop into colon cancer. Colonoscopy has led to a significant decline in the incidence of colon cancer and hence has proved to be effective. However, polyps are missed during the colonoscopy procedure due to various human factors like attentiveness, diligence, etc. Therefore, to counter this and ensure better quality colonoscopy, a number of guidelines have been established. Among them is to maintain a minimum withdrawal time of 6 minutes. However, such a constraint does not guarantee a quality inspection. For instance, it might be the case that colonoscopist spent majority of time in one part and performed quick examination in other parts. This is where Computer Aided Diagnosis (CAD) can be used to automate the monitoring of the quality of the colonoscopy video. The output of CAD ideally is shown in real-time on the monitor and therefore assist in decision-making. This paper presents a method of using neural network based architecture to classify the images. This paper reports the results for Resnet50 and MobileNetV2 architectures. The remainder of the report is organized as follows: Section II presents the related work; Section III describes about the data used and various pre-processing steps taken; Section IV presents the implementation and results; Section V presents the conclusions and outlines the future work.

II. RELATED WORK

There are different methods for automatically assessing the image quality of colonoscopic images. Nima et al. [2] suggested a method based on global and local image features that are extracted by histogram pooling over the entire image reconstruction error and region-based energy pooling, i.e. l2 -norm of reconstruction error. In [3] the authors, get texture related features using Local Binary patterns on frequency domain. These are then used along with SVM classifier to classify informative vs non-informative frames. Park et al. [4] address this challenge by presenting an algorithm based on a hidden Markov model (HMM) in combination with two measures of data quality to filter out uninformative frames. A two-level framework based on an embedded hidden Markov model (EHHM) is used to incorporate the proposed quality assessment algorithm into a complete, automated diagnostic image analysis system for colonoscopy video.

Different from the above approaches, this paper considers using an end to end deep learning neural network model to classify images into two categories i.e. blurry and clear.

III. DATA

The database contains 1062 images out of which 829 are clear images and 233 are blurry images. Figures 1 and 2 show example images from the dataset.

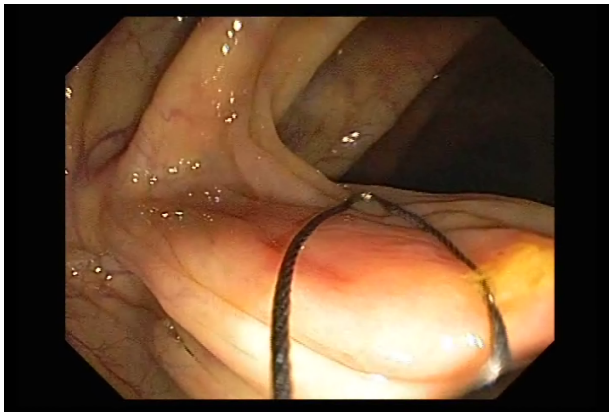


Fig. 1: Example Clear image from Dataset

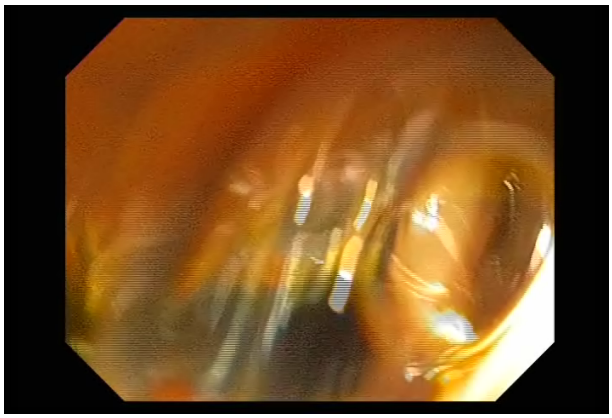
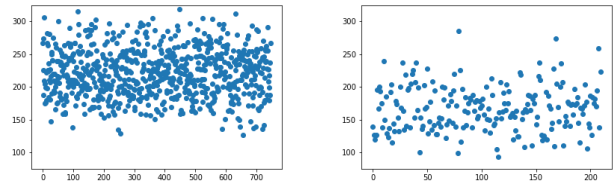
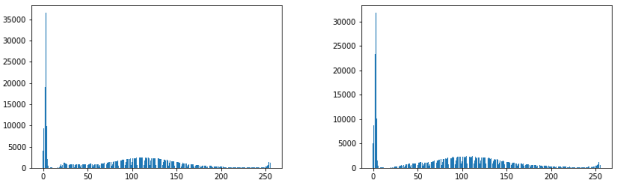


Fig. 2: Example Blurry Image from Dataset



(a) Clear Sizes (b) Blurry Sizes

Fig. 3: File size comparison



(a) Clear Images histogram (b) Blurry Images histogram

Fig. 4: Image histogram comparison

A. Data Exploration

This paper first explores the data to understand the dataset and potentially find some patterns which may lead to solve the problem better.

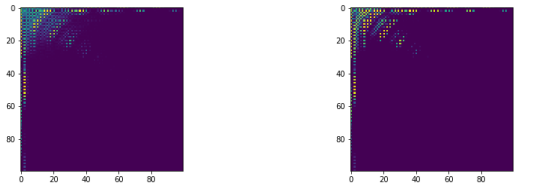
Figures 3a and 3b show the comparison of file sizes between the two types of images. The intuition was that blurry images inherently should have lesser information so should be lesser in size. As seen from the figures this is true to some extent though there is some overlap. Mean size for clear images is 220.8 Kb and for blurry images it is 167Kb.

Figures 4a and 4b show the comparison of averaged histograms for clear and blurry images. The thought process behind this was that blurry images should show little variability in pixel values compared to clear images. As seen from the figures, this does not seem to be true. Both histograms are very similar.

Figures 5a and 5b show comparison of averaged DCT for clear and blurry images. The intuition here was that blurry images should contain much more low frequency data than clear images so I wanted to check if that could be used as a feature. As the figures show, blurry images do have more intensity in lower frequencies however, there is not much difference overall.

B. Data Augmentation

As seen from the number of images, the dataset is highly unbalanced. Therefore, it was necessary to generate more blurry images so that the model properly learns to differentiate the two classes of images. Figure 6 shows an example where a Gaussian blur was introduced to a clear image from the dataset to create a blurred image. In addition, more images were generated by rotation to make sure the model generalizes better to new data and is rotation invariant.



(a) Clear Images DCT (b) Blurry Images DCT

Fig. 5: DCT comparison

Following this procedure the number of blurry images almost doubled.

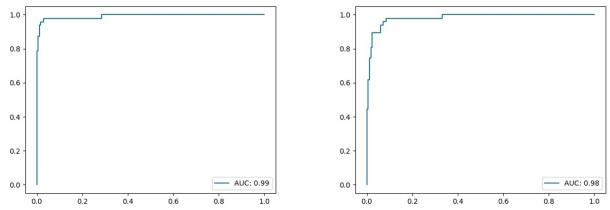
Fig. 6: Gaussian blur

IV. IMPLEMENTATION AND RESULTS

A. CNN

Convolutional layers are responsible for detecting certain local features in all locations of the input images. To detect local structures, each node in a convolutional layer is connected to only a small subset of spatially connected neurons in the input image channels. To enable the search for the same local feature throughout the input channels, the connection weights are shared between the nodes in the convolutional layers. Each set of shared weights is called a kernel, or a convolution kernel. Thus, a convolutional layer with kernels learns to detect local features whose strength across the input images is visible in the resulting feature maps. To reduce computational complexity and achieve a hierarchical set of image features, each sequence of convolution layers is followed by a pooling layer. CNNs consist of several pairs of convolutional and pooling layers, followed by a number of consecutive fully connected layers, and finally a softmax layer to generate the desired outputs.

The early layers of a CNN learn low level image features, which are applicable to most vision tasks, but the late layers learn high-level features, which are specific to the application at hand.



(a) ResNet50 (b) MobileNetV2

Fig. 7: ROC curve for different architectures.

B. Resnet

In residual learning framework the layers are explicitly reformulated as learning residual functions with reference to the layer inputs, instead of learning unreferenced functions. Residual networks [6] are easier to optimize, and can gain accuracy from considerably increased depth.

ResNet models are implemented with single-layer skips. This is done to avoid the problem of vanishing gradients, by reusing activations from a previous layer until the adjacent layer learns its weights.

However with Resnet, the layers might be massive and hence the network is prone to overfit the data.

C. MobileNet

MobileNets are a class of efficient models used for mobile and embedded vision applications. These are based on a streamlined architecture that uses depth-wise separable convolutions to build light weight deep neural networks. These are small, low-latency, low-power models that are parametrized to meet the resource constraints of a variety of use cases. They are built for classification, detection, embeddings and segmentation similar to other popular large scale models. [5] introduce two simple global hyper-parameters that efficiently trade off between latency and accuracy. These hyper-parameters allow the model builder to choose the right sized model for their application based on the constraints of the problem.

Table I and Figures 7a, 7b show the accuracy and ROC curve respectively using MobileNetV2 and ResNet50.

Algorithm	Accuracy
ResNet	0.9148
MobileNetV2	0.9626

TABLE I: Accuracy on Different Architectures

V. CONCLUSIONS AND FUTURE WORK

Distinguishing clear from blurry images in colonoscopy is important to create a robust measure for quality of the procedure. The dataset consists of 829 clear images and 233 blurry images. Since the data is imbalanced this paper used gaussian blur to create more blurry images. In addition, more images were created for both the classes using rotation for better generalization. This paper experiments with some state of the art deep neural architectures which have shown good performance in the literature on image classification tasks. Resnet50 achieves an accuracy of 0.9148 and Mobile net gets accuracy of 0.9626. Since, Resnet50 is more complex architecture I believe it's performance is slightly less because it likely overfits the data.

For future work, I will consider on improving the ResNet model to avoid overfitting. I will also explore the results in more depth, calculating confusion metrics to analyze the number of false positives and false negatives which will give more insight into where the model is failing. Finally, I will also provide a detection latency analysis of the system as I want the approach to perform in real time in real-world setting to assist the clinical doctors. Lastly, I will start working on the second task which is automatic polyp detection and integrate both these tasks.

REFERENCES

[1] C. N. Fischer, R. K. Cytron & R. J. LeBlanc. ÒCrafting a CompilerÓ, 2nd ed.,Boston: Pearson Education, 2010.

[2] N. Tajbakhsh, "Automatic assessment of image informativeness in colonoscopy," in Abdominal Imaging. Computational and Clinical Applications. New York: Springer, 2014, pp. 151–158.

[3] Ballesteros C., Trujillo M., Mazo C., Chaves D., Hoyos J. (2017) Automatic Classification of Non-informative Frames in Colonoscopy Videos Using Texture Analysis. In: Beltrán-Castañón C., Nyström I., Famili F. (eds) Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. CIARP 2016. Lecture Notes in Computer Science, vol 10125. Springer, Cham

[4] Park, S.Y., Sargent, D., Spofford, I., Vosburgh, K.: Colonoscopy video quality assessment using hidden markov random fields. In: SPIE Medical Imaging, pp. 79632P–79632P. International Society for Optics and Photonics (2011)

[5] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications," in arXiv:1704.04861, 2017.

[6] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. InCVPR, 2016.