Video Content Analysis Of Colonoscopy Video

by

Dawit Weldeghergish

A thesis submitted in fulfillment for the
Master's Degree

in the
3D Multimedia Technology
University of Jean Monnet

Academic Supervisors:

Prof. Alain Tremeau

Prof. Faouzi Cheikh

October 2017

A thesis submitted in fulfillment

of the requirements for Master of Science(3DMT).

Done in

Host Supervisor:

Prof.Yang Cai

Abstract

3D Multimedia Technology

University of Jean Monnet

Masters Degree

by Dawit Weldeghergish

Recently Video analysis and getting knowledge from videos has been use full technique in today's health care technology.Endoscope is one of today's technology that has get a grate attention in medical world.Many open surgical procedures now are being converted to endoscopic procedures including resection of gallbladders, retrieval of donor kidneys, resection of tumors of colon and pancreas, correction of hiatal hernias, coronary artery bypass grafting and minimal invasive neurosurgery's (i.e., video endoscopic neurosurgery).Endoscopic procedure is performed by endoscope with a tiny video camera at the tip of the endoscope,which generates a video signal of the interior of the human body.Despite a grate medical knowledge in colonscopy a little work has been so far done on developing a system that analysis the content of the video and provide user-friendly and efficient access to the medical, scientific, or educational content on such videos.

In this thesis we focus on analyzing of the content of a colonscopy video mainly focusing on stool and colon distention.There are other information's that could be extracted from colonscopy video that happen during colonscopy procedure including uninformative frame detection,surface area evaluation,biopsy,tissue specimen detection,polyp detection,clarity,audio,bleeding detection.As for thesis we have limited time we focus on stool detection and classification,colon distention and classification. Color based object recognition has shown a significant role in the field of medical imaging technology.In our study made on the previous algorithm implemented using color feature has also shown a significant accuracy in stool detection from a colonscopy video.We continue to work over this feature to learn over standard bowel preparation methods like Boston bowel preparation and we used SVM classifier to learn over the selected color features.

As recently deep learning have shown a significant role on learning a good representative features in images classification especially through Convolutional Neural Networks (CNNs), In this thesis we explore small deep leaning networks for colon distention classification.However CNN training for automated colon distention classification still provides a challenge due to the lack of large and publicly available annotated databases.we compared our results with previously implemented based on classical shape feature ,further more we explored transfer learning mainly off-the-shelf training and compared the

results with CNNs trained from scratch and using classical features. ...

Acknowledgements

First of all I would like to thank God for keeping me healthy and giving me this opportunity.I would also like to extend My heartfelt appreciation to my programme coordinator, Professor Alain Treamue, for his keen cooperation and understanding my problem during working on this project.I would also specifically like to thank Prof.Yang Cai and Prof.Faouzi Cheikh for supervising my research on this project and providing resources for the experiments.Also I would like to express my deepest appreciation to GI doctors Dr.Thakkar and Dr.Bahrat in West Penn hospital at Pittsburgh USA for providing us data and helping us in annotating the data.

Additionally, I thank all the people at CyLab CMU for the perfect working atmosphere. Lastly, I thank my family and friends for their love and support.. . .

Contents

List of Figures

Chapter 1

Introduction

Colorectal cancer is currently the second leading cause of cancer related deaths in the United States, just behind lung cancer[1]. In 2003, more than 143,000 people were diagnosed with colorectal cancer, and nearly 56,000 people died from it. The standard procedure for identification and removal of colorectal cancer is a colonoscopy. In this procedure, an endoscope, which has a small video camera with a wide-angle lens on the tip of it, is inserted into the rectum. A typical colonscopic Procesdure can be seen in Figure 1. Typically, the endoscope is inserted through the entire length of the large intestine to where it reaches the terminal ileum, which is where the large intestine reaches the small intestine.

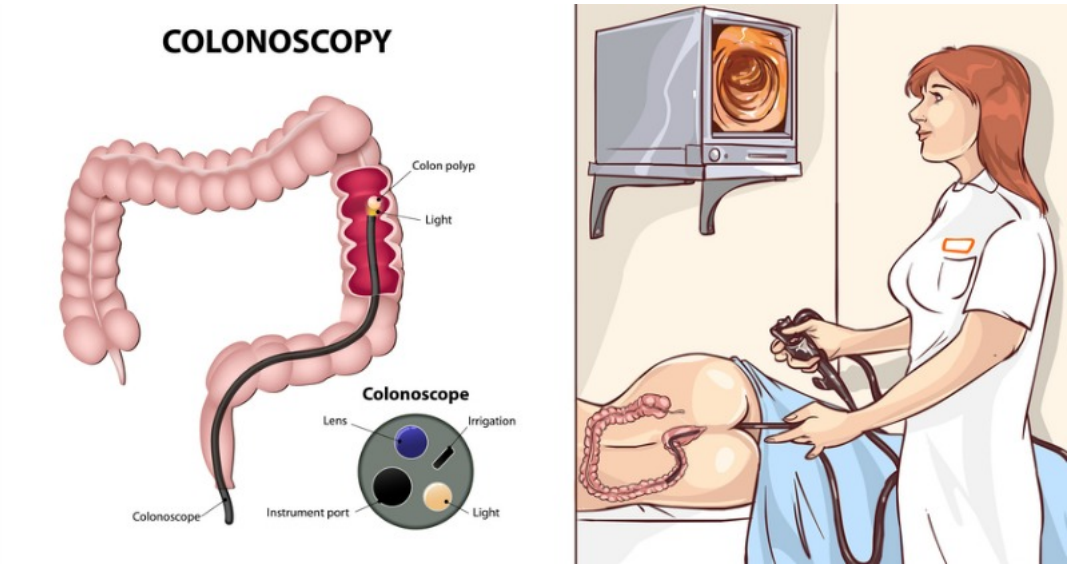

FIGURE 1.1: Depiction of a Colonscopic Procedure

Then,the endoscope is slowly removed from the colon, while the endoscopist carefully looks at the walls of the colon through the screen. Any abnormality seen during the procedure might be considered as a polyp,the physicians(GI doctors) must decide whether or not to remove the abnormality seen.

Automatic analysis of colonoscopy semantics from colonoscopy videos is a very important research problem as it can create platform for improving and assessing endoscopists procedural skills.Secondly it is use full to use it as educational resources for endoscopist, Currently, endoscopists typically capture images of interest using proprietary software or occasionally record the entire procedure onto a VHS tape.Although the captured analog video allows post-procedure analysis of the entire colonoscopic procedure, images in the VHS tape are of relative poor quality. It is also time consuming to locate a few interesting images within the entire video.

1.1 Problem Description

While colonoscopy has contributed to a decline in the number of colorectal cancer-related deaths, recent data suggest that there is still a significant miss-rate for the detection of even large polyps and cancers. In patients undergoing colonoscopy, 22% to 28% of polyps are missed. It has been indicated that on the study made for Factors influencing the miss rate of polyps [2] colorectal cancer detected within 3 years after colonoscopy may originate from missed lesions.The study further indicated that risk of missing a polyp is related to patient factors, patients with more than two polyps have high risk of missing rate.Other factors may also influence polyp detection rate,such as speed of withdrawal using complete range of motion of the endoscope,study indicated that endoscopists with mean withdrawal times 8 minutes had higher Polyp detection rate[3], bowel preparation (or obstruction), distention of the colon ,bleeding and especially endoscopist experience.In order to improve the quality of the average colonoscopy,some quality metrics can be used to objectively determine if a physician did a thorough colon inspection. So conducting an automated colonoscopy video assessment for every procedure in real time will be help full for the surgeons.

1.1.1 Contribution

Colonscopic video content analysis includes a number of analysis, like informative and uninformative frame classification,withdrawal time,Surface Area covered in the colon, Distention,Preparation, Clarity and audio analysis.As the time is limited in this thesis we focused on preparation and Distention evaluation based on the content of the video.

The first thing we try to solve is the preparation detection(stool detection in colon), this is mainly to check if a colon is prepared well before the screening.This is typical color detection problem as our colonoscopy images are modeled in RGB color space in which each color band is represented with 8-bit ranging from 0 to 255, and giving us a total of 255^3 potential colors.Even though the color of the stool vary from light tan to green, we have observed that there is a range of colors that is unique to stool.We modeled the stool color using a well known color feature called color histogram, in addition to this we compared HSV,LAB and LUV color spaces.Finaly we have used Support Vector Machine module to clasify the images in to four class given as very-poor,poor,sub optimal and Optimal.

The second thing we have solved is to classify between distended and closed frames.For this classification we have tested two approaches to improve the score of classification.

- Testing the previous implementation, which is based on the idea that frames containing image of open or closed distention can be categorized by detection of smooth or curvy contours.Segmentation of the region of interest was applied, followed by extract major smooth contours from the frame.To make the actual classification then it was defined shape features of this blobs based on its area perimeter,compactness and elongatedness.This descriptors of the blobs are then our features for classification.We learn over this features and tuned our parameters then we defined thresholds that can classify this frames in video into closed and open distention.

- Deep learning and Transfer learning is another approach we have explored to classify between distended and closed colon frames.where we have sub divided the open frames and closed frames in to four classes and we tend to learn over this classes of Very Poor(where the image is totally collapsed), Poor(Partially collapsed),Sub optimal(where Partially Distended) and Optimal (totally Distended).In this case,we

tested small networks training from scratch and pre-trained CNN over large data set like image net, the last or next-to-last linear fully connected layer is removed and the remaining pre-trained CNN is used as a feature extractor to generate a feature vector for each input image from a different database. These feature vectors can be used to train a new classifier (such as a support vector machine, SVM) to classify the images correctly.The classification has shown noticeable improvement from the previous technique used.

Chapter 2

Background and Related Works

A Colon is a 150cm muscular tube.A normal colon consists of six parts: cecum with appendix, ascending colon, transverse colon, descending colon, sigmoid and rectum.

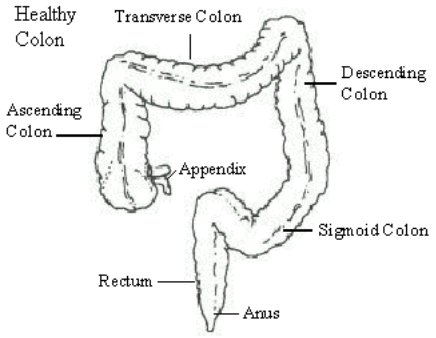

FIGURE 2.1: colon endoscopic segments

Colonoscopy is the primary method for colon cancer screening and prevention, during which a tiny camera is navigated into the colon in order to find and remove polyps.During the colonoscopic procedure, a flexible endoscope (a flexible tube with a tiny video camera at the tip) is advanced under direct vision via the anus into the rectum and then gradually into the most proximal part of the colon or the terminal ileum. Colonscopic procedure has two phases the insertion phase and the withdrawal phase. During the insertion phase, the endoscopist rapidly advances the tip of the endoscope to the most proximal location possible (cecum or terminal ileum). Careful mucosal examination

starts when the doctor reaches the cecum or ileum, diagnostic and therapeutic operations are typically performed during the withdrawal phase when the endoscope is gradually withdrawn.

2.1 Diagnostic and Therapeutic Operations

An endoscope has instrument channels that allow the insertion of flexible accessories such as biopsy forceps, cytology brushes, sclerotherapy needles, and diathermy snares from a port on the endoscope control head through the shaft and into the field of view. These instruments are used for tissue-sampling, other diagnostic and therapeutic procedures. Biopsy forceps used for tissue sampling consist of a pair of sharpened cups,a spiral metal cable, and a control handle. The tissue specimen is used for microscopic examining for its structure or for searching for the presence of infectious agents such as Helicobacter pylori. [4].

FIGURE 2.2: Examples of instruments

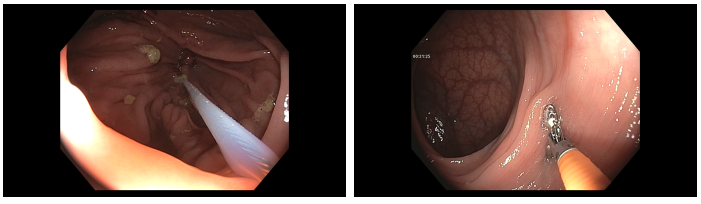

FIGURE 2.3: Colon images with surgical instruments during colonoscopic procedure

2.2 Colon Distention

Distention in colon is determine whether a doctor is able to visualize the hole part of the colon,The ideal agent for colonic luminal expansion would facilitate cecal intubation, provide excellent mucosal visualization, limit intra and post procedure pain, and would be safe and inexpensive.Air insufflation is the most commonly used technique for luminal distention since the advent of colonoscopy in the late 1960s.Colon distention can be classified as an open or closed colon.There are two methods of insufflation of colon presented in [5] using air,CO_2,water and other agents like Helium, argon, nitrogen, and xenon which have issues related to absorb-ability, availability, and expense significantly limit their application to colonoscopy, among this methods the most common method is using CO_2 as this has shown significant reduction in abdominal pain intra- and post-procedural up to 24 hours compared with air insufflation and it also appears to benefit patients undergoing some lengthier upper endoscopic procedures with respect to less post-procedure pain. As explained in the journal "Methods of luminal distention for colonoscopy".During the insertion phase of colonoscopy, at least partial distention of the lumen is needed to allow adequate visualization to safely direct the instrument to the cecum.During withdrawal, a greater degree of luminal distention is desired to allow optimal inspection of the colonic mucosa [5].

2.2.1 Open

In case of Open colon we tend to look in the image for smooth edges, the edges that are curvy enough and are not winding in various directions (as it happens in case of contours (edges) present in Closed colon. Usually they can be described instinctively as rings or parts of rings (arches), which are not 100% circular, but rather triangular in shape with highly circular segments.Typical example of distended colons is shown figure 2.4.

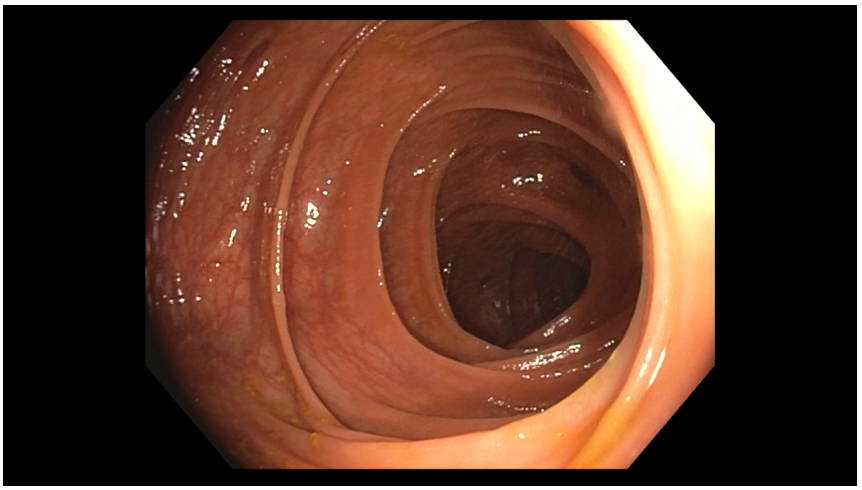

FIGURE 2.4: Open Colon

2.2.2 Closed colon

Closed colon is when good open colon is closing or closed and therefore we do not see all the surface of the colon, which means the doctors cannot properly examine the patient. To expand the colon, they inject CO2 into the colon, which causes the closed colon to become open one and allow further examination.Typical example of collapsed colon is shown in figure 2.5.

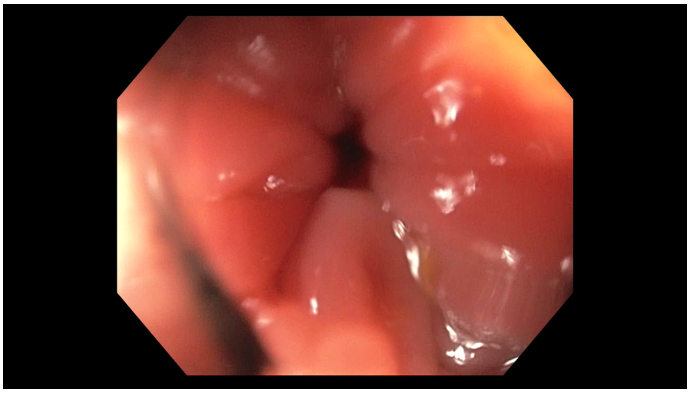

FIGURE 2.5: Closed Colon

2.3 Colon Preparation

The diagnostic accuracy of colonoscopy depends on the quality of bowel preparation[6]. Inadequate cleansing can result in missed pathologic lesions.The ideal preparation method

would reliably empty the colon of all fecal material, and have little effect on the gross or the microscopic appearance of the mucosa. It would require a relatively short period for ingestion and evacuation, cause little patient discomfort, and produce no significant fluid-electrolyte shifts. It also should maximize the detection of colonic disease including polyps and carcinoma.

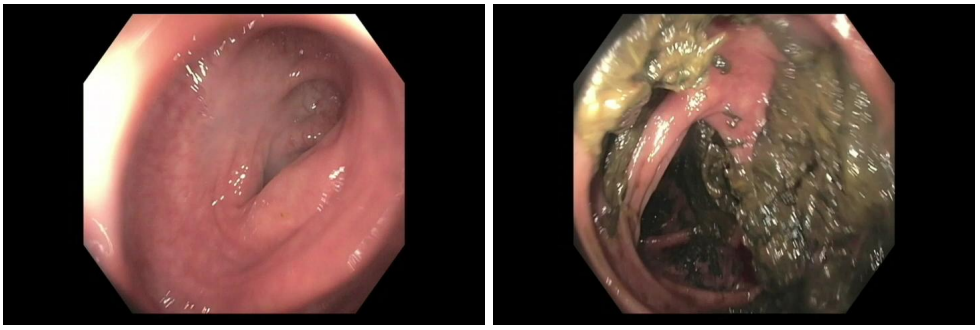

FIGURE 2.6: prepared colon and unprepared colon

2.4 Related Works

Despite intensive research in medical imaging in recent years, research on image analysis for colonoscopy videos has been minimal,to the extent of our knowledge we did not find any paper published regarding distention classification of clonic images.A lot of research in the area of frame classification of informative and uninformative frames has been published[7–9].In the paper Informative frame classification for endoscopy video[7],have compared frame classification based on edge and clustering.The features used for classification where obtained by constructing Fourier Transform of a gray- scale image and Based on the contents of the image, the frequency spectrums generate different pattern,where the informative frames has a lot of clear edge information so its spectrum of the informative frame does not show prominent components along the 45° directions because it has a wider range of bandwidths,from low to high frequencies .A well-known statistical approach gray-level co-occurrence matrices(GLCM) was used to analyze texture of the images frequency spectrum.In addition to this the paper has shown techniques for Specular reflection detection using multiple thresholds adaptively instead of defining global threshold, which is less sensitive to the thresholds.In comparison to simple thresholds methods the method proposed on this paper has shown an improvement by 0.9% of accuracy.Finally the paper has concluded that ,Overall the clustering-based technique generates better performance results than the edge-based technique and the performance of both the edge-based technique and the clustering-based technique increases when corrections for the specular reflection areas are incorporated into the computation. Similarly based on the idea of informative frame has larger amount of edges than non informative frames Cristian Ballesteros, Maria Trujillo and Claudia Mazo [8], proposed an automatic classification of non-Informative frames in colonoscopy videos using sobel edge detector with and without the Brightness Segmentation (BS) using a threshold equal to 2.5.

Classification of tissues in colon biopsies are described in [9–11] .Based on morphology of the glandular cells of the tissue region, morphological features are described by Khalid Masood[9], which describe the shape, size,orientation and other geometrical attributes of the cellular components, are extracted to discriminate between two classes of input data.GLCM was also used to analyze the morphological features and the experiments with Super vector machine using Gaussian kernel have the best classification accuracy of 100 percent for a threshold of 60 percent correct patches.Similar morphological feature

extraction system for Colon tissue classification has also been shown by Alexandre Cecilien Dufour in the paper hyper spectral colon tissue classification using morphological analysis[10].Recently in [11] multispectral texture features have proposed for differentiating CRC tissues in colon biopsies.In this technique GLCM and discrete wavelet features are extracted from microscopic images corresponding to different spectral bands. These features are then used as input to a random forest model, for the classification of samples into four classes: stroma, benign hyperplasia, intra-epithelial neoplasia and carcinoma. A combination of three well know type of features Local binary patterns by Ojala [12] ,Local ternary patterns by X. Tan and B. Triggs, [13] and Haralick texture patterns by R.M. Haralick [14], were proposed by Saima Rathore [15]and the result has shown that hybrid features produce superior results compared to individual features. Segmentation accuracy for hybrid features is 98.8%, which is superior compared to individual best of 95.8%, 98.2% and 94.5% for LBP, LTP and Haralick features, respectively.

Detection of stool in colonscopyic video is also another task in video analysis where a number of techniques are proposed[16, 17], as every patient before colon screening undergoes through bowel preparation because it creates problem in the visibility of polyps , its important to automatically analyze if the preparation was good.Marius George Linguraru has shown automatic stool subtraction algorithm from CT Colonography[16].To adjust the variable intensities that might occur in different CT images they have utilized expectation maximization algorithm which eventually resulted in dividing the the image into four class air, tissue, stool, and unclassified residuals. Later Sae Hwang and JungHwan Oh[17] proposed stool detection based on the color features using the Support Vector Machine.They have compared color histogram methods among three color spaces, RGB,HSV and L*u*v* color spaces and performance of CIE L*u*v* color space was higher than the other color spaces.

Coloscopic video has two phases insertion and withdrawal phase, the main important phase is withdrawal phase, but during coloscopic video analysis we need to tell on which phase the procedure is at a given time, which also the phases are sub divided in to different parts of the colon such as ascending,descending and traversal colon.JungHwan Oh on his paper blurry frame detection and shot segmentation in Colonoscopy Videos [18]proposed a technique for identifying the borders in the shots based on calculate the difference between the two consecutive frames and checking if it is larger than a shot segmentation threshold,which is indicating that there is a shot boundary between

the consecutive frames.Another technique introduced in this paper was to detect blurry images using Canny Edge Detector,where blurred images result in blurred edges and this is due to the discontinuity of the edge pixels, so to differentiate between blurred images and clear images, they defined an isolated pixel ratio and if this ratio is grater than a upper threshold it will be declared as a blurred image, if it is lower than minimum threshold it will be classified as clear, if it is in between it will be classified as ambiguous.[19, 20] focused on lumen identification given that the image is known to have the lumen.. Tian et al. proposed adaptive progressive thresholding(APT) technique for detecting the preliminary ROI from a grey-level endoscopic image followed by enhancement of the boundary using iris filter[19].Kumar proposed a technique for the extraction of lumen region and boundary from the GI images using INS(Integrated Neighbourhood Search) [20]. Several algorithms and techniques have been proposed for detection of polyps in colonscopic videos[21–26].Nima Tajbakhsh proposed hybrid context-shape approach for polyp detection as pure shape-based approach may mislead a polyp detector towards other polyp-like structures[21],again in [22] Nima proposed Channel fusion by stacking color,shape,and temporal patches for each polyp followed by training one CNN using those features.Sungheon Park has also shown that it possible to learn a use full hierarchical features using CNN(Convolutional neural network) for polyp detection[23].Polyp detection using ellipse fitting was also proposed by Sae Hwang, which it was observed that all detected ellipse does not represent the polyp structure and they have introduced another three techniques using Curve Direction and Curvature,edge distance and intensity value to filter out the actual polyps[24].Lequan Yu proposed 3D CNN corp-orated with online and offline integration strategy for automated detection of polyps from colonoscopy videos[25].Alaa El Khatib [26] has also compared the different algorithms used in feature extraction for automatic polyp detection and they found that HOG(encode local shape information) and LBP(encode local texture information) (SVM) generally have the best performance. Recently the state of art algorithms has been over excitement generated by deep learning is a very promising direction where massively trained neural network based classifiers can be used to better differentiate polyp frames from normal frames. However, deep learning networks in general require a huge amount of training data, in particular labeled data of positive (polyp frame) and negative (normal frame) samples.

2.5 Object Detection and Recognition

Object recognition and Detection is one of the most relevant and challenging problem in computer vision.Object detection and recognition is a big part of peoples lives,We, as human beings, constantly recognize various objects such as people, buildings, and automobiles.Yet it remains a mystery how we detect objects accurately and with little apparent effort. Comprehensive explanations have defied psychologists and physiologists for more than a century.

Generally most of the object recognition methods used two approaches feature-based approach which tends to extract features from the image followed by matching or classification algorithms and by directly using the images.Common methods of feature extraction include those based on visual features (e.g.,edges, shapes,textures, segmentation, and gradient peaks), statistical characteristics (e.g minima, medians, and histograms), and transformation features (e.g. Hough transforms, wavelet transforms, and Gabor transforms) [27].There are also several feature extraction methods proposed and the two of standard vector-based techniques,SIFT and SURF, are applied to detect and describe local features.In object recognition we have methods of matching the features we selected this ranges from template matching to highly sophisticated models. In using Template matching it creates a template of an object and it measures the similarity between the interested image and the template.similarity measures can be cross correlation or sum of absolute differences.cross-correlation measure the similarity of two series as a function of the displacement of one relative to the other.SAD(Sum of absolute Difference) algorithm is a well known matching cost computation algorithm ,it considers the absolute difference between the intensity of each pixel in the reference block and that of the corresponding pixel in the target block[28].Instead of searching through all possible locations like in simple template matching methods,models based approach uses statistical models and various heuristics to guide and improve the search.Statistical models approximates reality and optionally to make predictions from this approximation.There are also several approaches to statistical modeling of object classes that are highly dedicated to a particular category of objects,such as handwritten characters[29].classification algorithms are one of the statistical models used classify an object to its respective group.Support vector machines,Linear classifiers,Quadratic classifiers,k-nearest neighbor and Neural networks

are some of the classifiers adopted by most of the classification problems. The classification or learning algorithms are basically of three kinds, supervised,semi-supervised and unsupervised learning. for supervising learning is where we have the ground truth and to learn the mapping function from the input to the output example are SVM,Linear regression,Random Forest.Unsupervised learning is where you only have input data and we don't know the ground truth where we try to learn the underlying structure or distribution in the data in order to learn more about the data typical classical algorithm is k-means. semi-supervised learning is the combination of supervised and unsupervised learning this is mostly when our data is semi labeled.

Another class of recognition methods processes images directly instead of extracting features first.A typical classic example will be eigenfaces,which are the principal components of initial training set of images for Face Detection[30].Principal Component Analysis is a popular technique for data compression and has been successfully used in many computer vision task like face recognition.In the learning stage, principal component analysis is performed on all classes or in the training data.Based on the fact that if we have a set of Principal Components that were obtained from one class only these must reconstruct better the images of this class than images in other class and viceversa,if we have a set of Principal Component obtained from images of anything except this class the reconstruction of the class will not be as good as the first result obtained. When we want to classify the images we calculate the reconstruction error[31].

2.6 Super Vector Machine

Support Vector Machine(SVM) is a supervised machine learning algorithm based on the concept of decision planes that define decision boundaries.SVM's are universal learners.They can learn linear separable functions, and by simple plug-in of an appropriate kernel function they can be used to learn polynomial classifiers,radial base function and three layer sigmoid neural nets.Data that is completely linearly separable can be separated by hard margin SVM,This is not really useful in practice, in 1995, Cortes and Vapnik [32]proposed the idea of a "soft margin" SVM that allows some examples to be "ignored" or placed on the wrong side of the margin.Given a feature vector $x_i \varepsilon R^n, i = 1, ...n$ of two classes and an indicator vector $y \varepsilon R^l$ such that $y_i \varepsilon -1, 1$,the

optimization problem for the hard margin SVM is given by.

$$\begin{aligned} \min_{\omega,b,\xi} \quad & \frac{1}{2}\|\omega\|_2^2 \\ \text{subject to} \quad & y_i(\omega^t(x_i)+b) \geq 1, \end{aligned} \tag{2.1}$$

Most of the time our data is not linearly separable so our above optimization problem is in feasible, meaning that there are no parameters ω,b that give us a solution to the problem.To overcome this problem Cortes and Vapnik[32] introduced slack variables,which are variables that allow us to relax the constraint.We can define a slack variable as a value ξ that, roughly, indicates how much we must move our point so that it is correctly and confidently classified. This makes sense - small slack variables means that we have a correct classification but are not very confident, while large slack variables means that we have not classified the point correctly and we have a variable C that controls how much we penalize our use of slack variables, as we increase C we penalize our slack variables more. The optimization problem for the Soft margin SVM is given by.

$$\begin{aligned} \min_{\omega,b,\xi} \quad & \frac{1}{2}\omega^t\omega + C\sum_{i=1}^{l}\xi \\ \text{subject to} \quad & y_i(\omega^t\phi(x_i)+b) \geq 1-\xi_i, \\ & \xi \geq 0, i = 1....l \end{aligned} \tag{2.2}$$

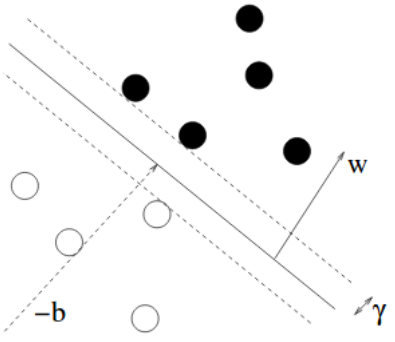

FIGURE 2.7: Here (ω,-b) defines the separable hyper-plane and γ is the size of the Margin

If the prediction is correct, no adjustments are made. If the prediction is wrong,the parameters,describing a hyper-plane, are moved in the direction of the point where the mistake occurred.The output scale of a SVM is determined so that outputs for the support vectors are +1 or -1[33].

2.7 Deep learning

In recent years there has been an increased interest in machine learning techniques that is based on raw data input, instead of hand crafted features. More importantly, deep learning is also allowing the implementation of new applications that are more focused on high-level classifications that do not depend on lesion segmentation.Deep learning is the implementation of neural networks with more than a single hidden layer of neurons.

2.7.1 Layers

2.7.1.1 Convolutional Layer

Given a two-dimensional image, I, and a small matrix, K of size h x w, (known as a convolution kernel), which we assume encodes a way of extracting an interesting image feature, we compute the convolved image, I*K, by overlaying the kernel on top of the image in all possible ways, and recording the sum of element wise products between the image and the kernel.So now we have a single number for each K,this number is just representative of when the filter is at the top left of the image. Now, we repeat this process for every location on the input volume[34].

$$(I * K)_x y = \sum_{i=1}^{h} \sum_{i=1}^{w} K_i j * I_x + i - 1, y + j - 1 \tag{2.3}$$

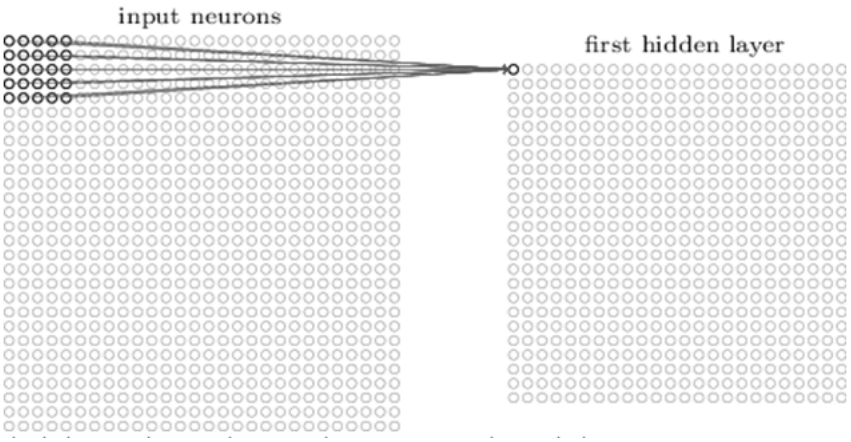

FIGURE 2.8: 5x5 kernel over 32x32 image resulting in 28x28 activation map

2.7.2 Connected Layer

The output activation map of the first layer will be an input to the next layer in the network of CNN.A classic CNN architecture would look like this

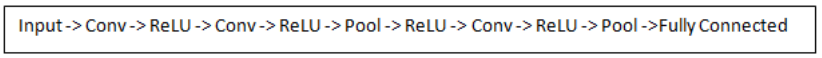

FIGURE 2.9: CNN Structure

The first layers of the network mostly detect low level features such as edges and curves.In order to detect high level features and to predict whether an image is a type of object, we need the network to be able to recognize higher level features of the object.Another interesting thing to note is that as you go deeper into the network, the filters begin to have a larger and larger receptive field, which means that they are able to consider information from a larger area of the original input volume .Fully connected basically takes an input volume and outputs an N dimensional vector where N is the number of classes that the program has to choose from.FC layer looks at what high level features most strongly correlate to a particular class and has particular weights so that when you compute the products between the weights and the previous layer, you get the correct probabilities for the different classes[34].

2.7.3 Rectified Linear Unit

RLU has the following mathematical formula

$$y = max(0, x) \tag{2.4}$$

In the past, nonlinear functions like tanh and sigmoid were used, but researchers found out that ReLU layers work far better because the network is able to train a lot faster.In fact, it does not suffer from the vanishing or exploding gradient. However, the ReLU removes all the negative information's and thus appears not suited for all data-sets and architectures[35].

2.7.4 Spatial Pooling

This is also refereed as down sampling layer.The operation performed by this layer is also called down-sampling, as the reduction of size leads to loss of information as well, however the loos is important to the network as it reduces the computational time and it is one way of fighting over-fitting. he spatial pooling layer is defined by its aggregation function, the height and width dimensions of the area where it is applied, and the properties of the convolution (e.g.padding,stride).

2.7.5 Batch Normalization

This adds a normalization step (shifting inputs to zero-mean and unit variance) to make the inputs of each trainable layers comparable across features. By doing this it ensures a high learning rate while keeping the network learning.

2.7.6 Convolutional Architectures

A lot of convolutional architectures have been developed from the 1990s.We will look on to some of the architectures proposed and each one gives an advance over the previous one in visual recognition.

2.7.7 CNN LeNet-5

In 1998 Yann LeCun [36]proposed one of the successful CNN based application for Digit recognition.In this CNN architecture three sequences of layers were used , convolution,pooling and non-linearity which are the most common ingredients of CNN architectures.Inputs where normalized using mean and standard deviation to accelerate training,average pooling was taken as pooling function,tangential or sigmoid as non linearity function.Finally as we can see from the figure 2.10 a fully connected high level feature map layer as a classifier is connected in the network.

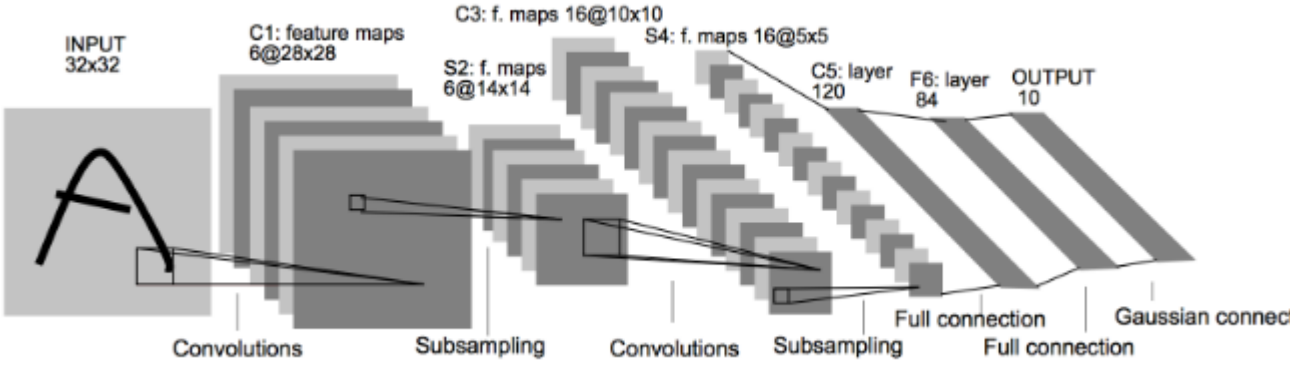

FIGURE 2.10: Architecture of LeNet-5

2.7.8 Alex Net

This CNN architecture is one of the popular architecture proposed by Alex Krizhevsky [35] in 2012 for image-Net data-set of over 15 million labeled high-resolution images belonging to roughly 22,000 categories.This model has significantly outperformed the other hand crafted models (accuracy top 5 of 84% compared to the second runner-up with 74%).The network was made up of 5 conv layers, max-pooling layers, dropout layers, and 3 fully connected layers,ReLU as non-linearity function.Recently the authors provided a multi-GPUs implementation in CUDA to bypass the memory needs.

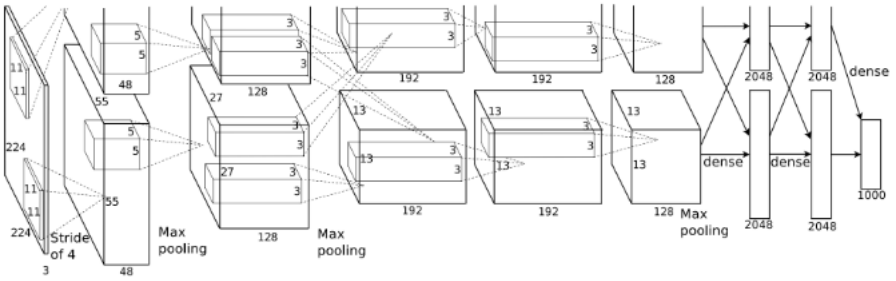

FIGURE 2.11: Architecture of Alex-net

2.7.9 VGG net

Later on 2014 Karen Simonyan and Andrew Zisserman [37]of the University of Oxford created a 19 layer CNN that strictly used 3x3 filters with stride and pad of 1, along with 2x2 max-pooling layers with stride 2 three Fully-Connected (FC) layers.Their main contribution was to show that depth is a critical component for good performance.The other greater idea that shown on vgg architectures was that the insight of using multiple 3x3 convolution in sequence can eliminate the effect of larger receptive fields.This was also tested and developed for image net data-set.All hidden layers are equipped with the rectification using RLU non linearity functions .However VGG nets has come with high memory cost and require lot of computation.

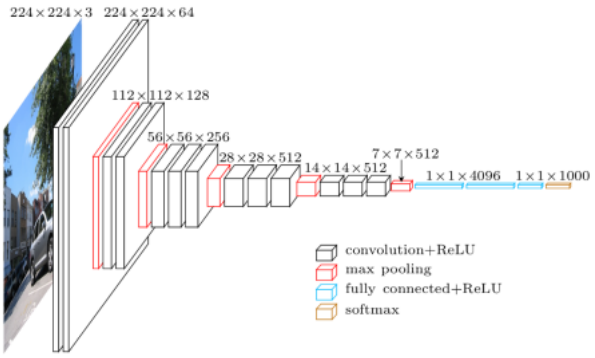

FIGURE 2.12: Vgg Architecture with 16 weight layers

2.7.10 Google Net or Inception Models

Going deeper with convolutions Google come up with an Inception architecture model later in 2015[38, 39] which they upgraded the models by rethinking over the inception. Its main contribution was the development of an Inception Module that dramatically reduced the number of parameters.This model has employed only 5 million parameters, which represented a 12x reduction with respect to its predecessor AlexNet,which used 60 million parameters.Further more, VGGNet employed about 3x more parameters than AlexNet.Pooling operations have been essential for the success of convolutional networks,Google net has also used this technique at the top of the convolutional layers instead of fully connected layers.This has also eliminated large amount of parameters which ultimately resulted google net to be used under strict constraints on memory and

computational budget.In google net it is clear that there is parallelism its not sequential.The figure 2.13 shows us the inception module proposed in google net.

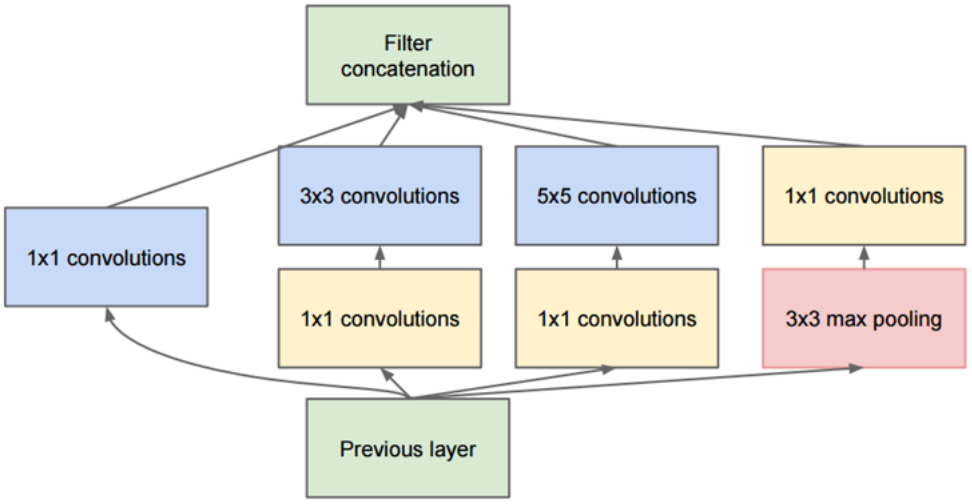

FIGURE 2.13: Inception module with dimensionality reduction

The bottom green box is our input and the top one is the output of the model.Inception module allows you to do is perform all of these operations in parallel. In fact, this was exactly the nave idea that the authors came up with shown in fig 2.13, but why it didn't work even though it might cover the optimal sparse structure is the pooling layer with outputs of the convolutional layers would lead to an inevitable increase in the number of outputs from stage to stage which eventually will increase the computation within few stages.The most recent architecture available is InceptionV3 [39]. Notably, it uses batch normalization

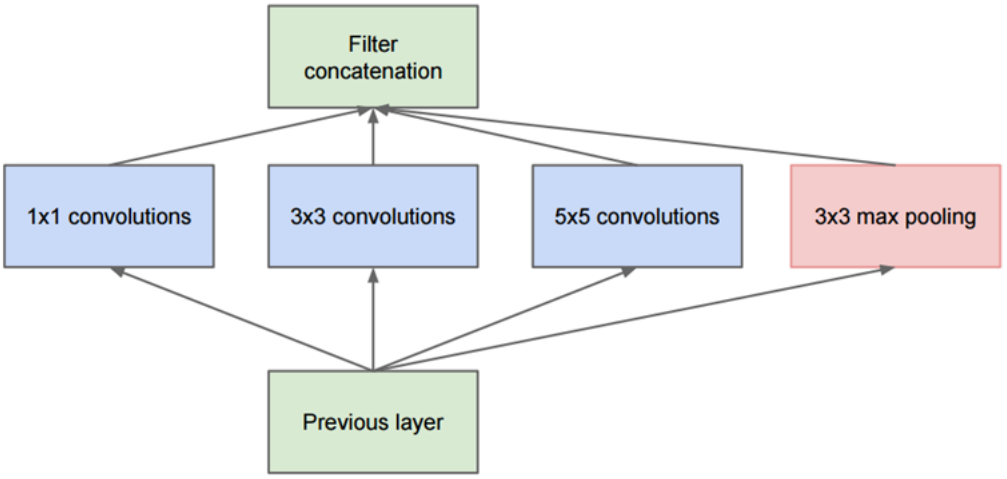

FIGURE 2.14: Inception module, nave version

2.7.11 ResNet

More deeper network with 152 layers called Deep Residual Learning [40]was proposed in 2015 by Microsoft in China. This was the winner architecture of ILSVRC2015 with an incredible error rate of 3.6%.In our traditional CNN we have input x that goes through conv-rlu-conv series and results in F(x), the idea brought in this network is that adding this output to the original input, H(x)=F(x)+x were in traditional out F(x) will be H(x).The hypothesis is that it is easier to optimize the residual mapping than to optimize the original, unreferenced mapping.

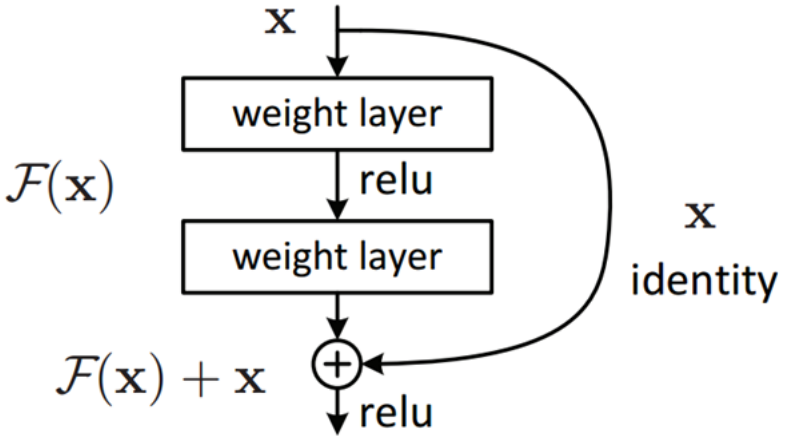

FIGURE 2.15: Residual learning: a building block

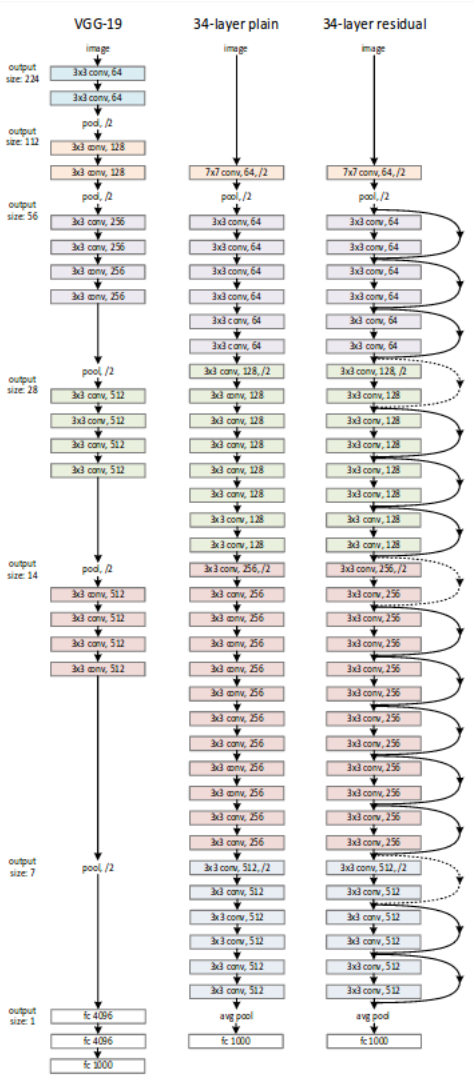

FIGURE 2.16: Network architectures for Image-Net.Left: the VGG-19 model [41] (19.6 billion FLOPs) as a reference.Middle: a plain network with 34 parameter layers (3.6 billion FLOPs).Right: a residual network with 34 parameter layer

The bar chart below shows the error rate comparison of the networks developed for image-Net.

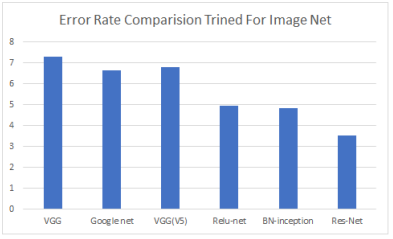

FIGURE 2.17: Error Comparison

2.7.12 Training

2.7.12.1 From Scratch

Initialization

After the success of CNNs in 2012 initialization with Gaussian noise with mean equal to zero and standard deviation set to 0.01 and adding bias equal to one for some layers become very popular, however this has become not possible to train CNNs from scratch using this initialization for deep CNNs, the main problem described was,If each layer, not properly initialized, scales input by k, the final scale would be k_L, where L is a number of layers. Values of k ¿ 1 lead to extremely large values of output layers, k ¡1 leads to a diminishing signal and gradient. In 2016 a new popular approach for deep CNNs has been proposed by Dmytro Mishkin and Jiri Matas [41]initialization with Layer-sequential unit-variance(LSUV) viewed as an orthonormal initialization combined with batch normalization performed only on the first mini-batch.The author farther described that such normalization is sufficient and computationally highly efficient in comparison with full batch normalization.More information can be found in the chapter 3 of Michael Nielsens book [34].

Loss Function

A loss function can be defined in many different ways but a common one is MSE (mean squared error), which is times (actual - predicted) squared, which is similar to Stochastic Gradient Descent (SGD),where it can find efficient weights and biases regarding the training set .In the first training the loos will be high but our ultimate goal is to predicted label (output of the ConvNet) is the same as the training label. The mean square error is given by

$$Loss(x, y) = \frac{1}{n} \sum_{i}^{n} |x_i - y_i|^2 \tag{2.5}$$

Back Propagation

In back propagation now, we perform a backward pass through the network, which is determining which weights contributed most to the loss and finding ways to adjust them so that the loss decreases.The goal of back propagation is to compute the partial derivatives $\frac{\partial C}{\partial w}$ and $\frac{\partial C}{\partial b}$ of the cost function C with respect to any weight w or bias b in

the network.Back propagation also help us in understanding the change of weight and bias in our network by computing the error in j neuron in the l layer.Once we compute this derivative, we then go to the last step which is the weight update. This is where we take all the weights of the filters and update them so that they change in the opposite direction of the gradient.For detail explanation we can refer to chapter 2 of Michael Nielsens book [34]

Optimization

The classical way of converting machine learning problem to optimization problem is by replacing the true distribution by the empirical distribution and then we tend to optimize the empirical risk hoping that the risk decreases signicantly.However this technique is prone to overfilling , The most effective modern optimization algorithms are based on gradient descent.Stochastic gradient descent(SGD) is one of the gradient descent techniques used to update our parameters and its variants.We have a learning rate which is critical parameter for SGD chosen by the programmer, but it is usually best to choose it by monitoring learning curves that plot the objective function as a function of time[42].A high learning rate means that bigger steps are taken in the weight updates and thus, it may take less time for the model to converge on an optimal set of weights. However, a learning rate that is too high could result in jumps that are too large and not precise enough to reach the optimal point.

2.7.13 Testing

Finally we need to test our trained model using a test set, which our model did not train with. Let see the main points why we evaluate the predictive performance of a model:

- To be able to estimate the general performance of our model.
- To improve the performance by tweaking the learning rate and other parameters ,selecting the best performing model from a given hypothesis space.
- We want to identify the machine learning algorithm that is best-suited for the problem at hand; thus, we want to compare different algorithms, selecting the best-performing one as well as the best performing model from the algorithms hypothesis space.

we can use several ways to achieve the above goals and to test our model, 0-1 loss prediction accuracy, confusion matrix, ROC curves,F1 scores,train and validation curves.

2.7.13.1 Transfer Learning

Most of our deep learning models to work efficiently requires huge amount of labeled data, in reality most of the time we don't have enough variety labeled data.Instead of training new deep learning model from scratch transfer learning help us to transfer knowledge from the pre-trained model and to generalize over the new task.According to Pan and Yang Given[43] a source domain D_s and learning task T_s, a target domain D_T and learning task T_T transfer learning aims to help improve the learning of the target predictive function f_T in D_T using the knowledge in D_s and T_s where D_s is different from D_T or T_s is different from T_T.

Off The Shelf Features

We can use a pre-trained model as a feature extraction mechanism. What we can do is that we can remove the output layer(the one which gives the probabilities for being in each of the 1000 classes) and then use the entire network as a fixed feature extractor for the new data set.Finally, a classifier is trained and tested on the features. Typically, the later is a Support Vector Machine with a linear kernel.

Fine Tuning

Training a pre-trained model partially by freezing the weights of initial layers and retrain only the higher layers. We can try and test as to how many layers to be frozen and how many to be trained.By resetting and applying smaller learning rate to a pre-trained model we can generalize the features to the new data set. The more different is the new data set from the original data set, more parameters/layers must be reset.

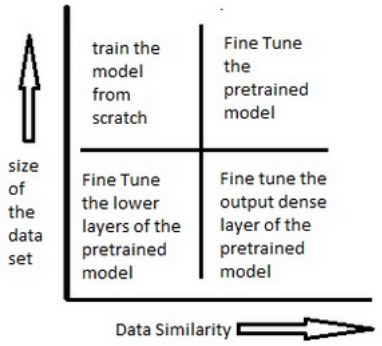

FIGURE 2.18: How to tune a pretrained model

2.8 Feature Extraction

A feature is a characteristic that can capture a certain visual property of an image either globally for the entire image or locally for regions or objects. Broadly features can be classified in to two class low-level(edges,corners,color) and high level features(human interpretation when looking into an image). We call these low-level features because most of them are extracted directly from digital representations of objects in the database and have little or nothing to do with human perception.

2.8.1 Edges and corners

Harris and Stephen where the first to come up with successful corner detection as a feature from images.But since it was only detecting corners, his work suffered from a lack of connectivity of feature-points which represented a major limitation for obtaining major level descriptors such as surfaces and objects.Canny and Sobel are another successful edge detectors[44] .Later high speed corner detection features from Accelerated Segment Test was proposed[45].In the comparative study made in [45] based on accuracy,consistency and speed it was shown that most of these detectors satisfied one of the criterion's but failed in the others.

2.8.2 Texture based

Approaches to texture analysis are usually categorized into structural,statistical, model based and transformation.

Local Binary Pattern(LBP)

LBP was introduced by Ojala [46] in 1994 .A 3x3 mask is used against the neighborhood pixels to define a particular texture and evaluate a local binary pattern (LBP).Mathematically given by

$$LBP(P,R) = \sum_{p=0}^{p-1} s(q_p - q_c)2^p \tag{2.6}$$

where P indicates the number of pixels in the neighborhood, and s(x)=1 when x≥0, else 0.Histogram of the obtained binary numbers is computed to represent the texture of the image.

Gabor filter

Gabor filter is one of the popular technique used as texture feature extraction [47]. Gabor filter-based feature extractor is in fact a Gabor filter bank consisting of filters with different frequencies and orientations. A common practice in Gabor filter design is to first define the highest frequency, the total number of frequencies n_f and the total number of orientations n_o, and then create filters based on combination of frequency and orientation parameters.

SIFT

SIFT is proposed by by David Lowe in 1999 it computes over local regions, usually centered on feature points but sometimes also densely sampled for object recognition task.SIFT feature is based on some of the appearance of the object's interest in the point of interest and has nothing to do with the image size and rotation.This algorithm simulates the multi-scale features of image data in different scale spaces,then the image is filtered and smoothed by using different Gaussian functions.Difference of Gaussians(DOG) is used to remove bad feature points,as DOG function has a strong edge response at the edge of the image this will not give stable features.Hesian matrix is applied to stabilize the feature pints and to make scale invariant gradient direction distribution of the neighborhood pixel of the key point is applied.Finally a block gradient histogram is used to generate a unique feature vector[48].

SURF(Speed-ed Up Robust Features)

SURF algorithm based on the same principles and steps as SIFT, but details in each step are different.SURF feature extraction is a scale and image rotation invariant detectors

and descriptors.The SURF descriptor is based on similar properties with a complexity stripped down even further. The first step consists of fixing a reproducible orientation based on information from a circular region around the interest point. In order to be invariant to rotation, they identify a reproducible orientation for the interest points. Then, for the extraction of the descriptor, the first step consists of constructing a square region centered around the interest point, and oriented along the orientation selected[49].

2.8.3 Color Feature

Human beings see all light as combination of three distinct colors (i.e. Red, Green, and Blue). Most commercially sold digital cameras, designed to see the world as we see it, are using the filters specially designed to respond to frequencies of light corresponding to these three colors. Digital color images including our colonoscopy images are modeled in RGB color space in which each color band is represented with 8-bit ranging from 0 to 255, and giving us a total of 255 by 3 potential colors. when thinking about extracting color features the most key components of color feature are color space,color quantification and then similarity measure.

2.8.4 Color space

Color space is a simple mathematical model where it describes a rage of colors as numbers,each color in the system is represented by a single pixel(e.g RGB color space uses three numbers to represent a color).However, all visible colors could not be specified with positive values of red, green and blue components CIE(International Commission on Illumination) come with a new representation of three standard primaries (X, Y, and Z) to replace red, green, and blue.

CIELAB

Lab Color space is defined by the CIE, based on one channel for luminance (lightness) (L) and two color channels (a and b). Lab colour spaces can be computed via simple formulas from the XYZ space, but is more perceptually uniform than XYZ meaning any change in the color value will reflect the same importance in the visual perception.The L*a*b* colour space is derived from the CIE XY tristimulus values. The L*a*b* space consists of a luminosity 'L*' layer,chromaticity layer 'a*' indicating where colour falls along the

red-green axis, and chromaticity layer 'b*' indicating where the colour falls along the blue-yellow axis. The approach is to choose a small sample region for each colour and to calculate each sample region's average colour in 'a*b*' space[50]. Mathematically the conversion of RGB to XYZ can be given by :

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 2.768892 & 1.751748 & 1.130160 \\ 1.000000 & 4.590700 & 0.060100 \\ 0 & 0.056508 & 5.597292 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

XYZ to Lab

L = $116f_y$-16
a = $500(f_x$-$f_y)$
b = $200(f_y$-$f_z)$

$$f_x = \begin{cases} \sqrt[3]{x_r} & if x_r > \varepsilon \\ \frac{kx_r+16}{116} & otherwise \end{cases}$$

$$f_y = \begin{cases} \sqrt[3]{y_r} & if y_r > \varepsilon \\ \frac{ky_r+16}{116} & otherwise \end{cases}$$

$$f_z = \begin{cases} \sqrt[3]{z_r} & if z_r > \varepsilon \\ \frac{kz_r+16}{116} & otherwise \end{cases}$$

$x_r = \frac{X}{X_r}$
$y_r = \frac{Y}{Y_r}$
$z_r = \frac{Z}{Z_r}$

$$\varepsilon = \begin{cases} 0.008856 & Actual CIE standard \\ 216/24389 & Intent of the CIE standard \end{cases}$$

$$k = \begin{cases} 903.3 & ActualCIEstandard \\ 24389/27 & IntentoftheCIEstandard \end{cases}$$

CIELUV

This is another color space defined by CIE, which also a transformation of the CIEXYZ color space its mathematical expression is

$$L = \begin{cases} 116\sqrt[3]{y_r} - 16 & if y_r > \varepsilon \\ ky_r & otherwise \end{cases}$$

$u = 13L(u' - u'_r)$
$v = 13L(v' - v'_r)$
where

$y_r = \frac{Y}{Y_r}$
$u' = \frac{4X}{X+15Y+3Z}$
$v' = \frac{9Y}{X+15Y+3Z}$
$u'_r = \frac{4X_r}{X_r+15Y_r+3Z_r}$
$v'_r = \frac{9Y_r}{X_r+15Y_r+3Z_r}$

$$\varepsilon = \begin{cases} 0.008856 & ActualCIEstandard \\ 216/24389 & IntentoftheCIEstandard \end{cases}$$

$$k = \begin{cases} 903.3 & ActualCIEstandard \\ 24389/27 & IntentoftheCIEstandard \end{cases}$$

As we can see from the mathematical expressions of both formulas for Lab and Luv,We have the same value of L when y_r is smaller than ε.

HSV

HSV space is frequently used in computer graphics and is a rather intuitive way of describing color. The three color components are hue, saturation (lightness) and value (brightness). The hue is invariant to the changes in illumination and camera direction. RGB coordinates can be easily translated to the HSV coordinates by the following formula[51].

$$H = \begin{cases} undefined & \text{if max(R,G,B)=min(R,G,B)} \\ 60.(G-B)/max(R,G,B) - min(R,G,B) + 0 & \text{max(R,G,B)=R and G>=B} \\ 60.(G-B)/max(R,G,B) - min(R,G,B) + 360 & \text{max(R,G,B)=R and G<B} \\ 60.(G-B)/max(R,G,B) - min(R,G,B) + 120 & \text{max(R,G,B)=G} \\ 60.(G-B)/max(R,G,B) - min(R,G,B) + 240 & \text{max(R,G,B)=B} \end{cases}$$

$$S = \begin{cases} 0 & \text{if max(R,G,B)=0} \\ 1 - min(R,G,B)/max(R,G,B) & \text{otherwise} \end{cases}$$

V= max(R,G,B)

2.8.4.1 Color Histogram

The color histogram is easy to compute and effective in characterizing both the global and local distribution of color in an image. In addition, it is robust to translation and rotation about the view axis and changes only slowly with the scale, occlusion and viewing angle. Since any pixel in the image can be described by three components in a certain color space, a histogram, i.e., the distribution of the number of pixels for each quantized bin, can be defined for each component. Clearly, the more bins a color histogram contains, the more discrimination power it has.

Chapter 3

Materials and Methods

3.1 SVM For Stool Classification

3.1.1 Dataset

The gradual advances in the development of endoscopic treatment of gastrointestinal tumors, early detection and accurate diagnosis of tumors have been increasing in importance. Under such circumstances, new technological instruments and image enhancement technologies have been an important object of research in clinical decision support system area.With high-magnification colonoscopies it is possible to acquire images up to 150-fold magnified, revealing the fine surface structure of the mucosa as well as small lesions.

In this work we have used the i-scan technology based data set where this is an image processing technique which has three modes of image enhancement ,surface enhancement (SE; enhancement of the structure through recognition of the edges); contrast enhancement (CE; enhancement of depressed areas and differences in structure through colored presentation of low density areas); and tone enhancement (TE; enhancement tailored to individual organs through modification of the combination of RGB components for each pixel).

As there is no publicly available detest for bowel preparation, We have prepared our own data-set with 2000 images for learning over bowel preparation around 500 images for each class,the data was prepared with the help of gastroenterologists from west Penn hospital

in Pittsburgh.Among them 1300 images where prepared for training and 700 images for testing.The classification system selected was Boston Bowel Preparation,where this classification system tends to classify the images in to four classes give numbers from 0 to 4.According to Boston Bowel Preparation, colon preparation is divided in to four class where the figure 3.1 below shows the four examples of the classes.

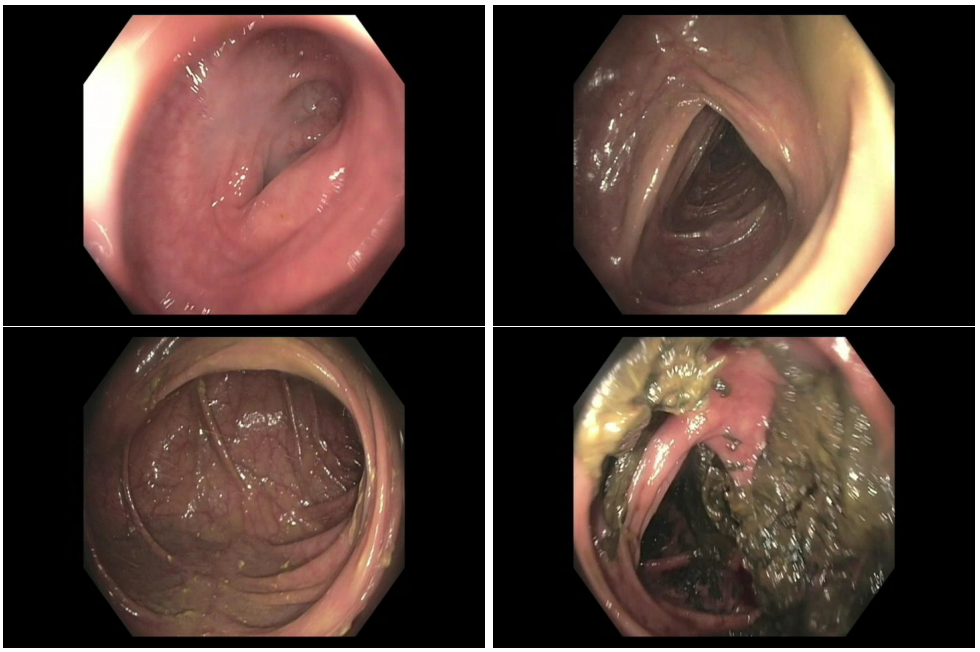

FIGURE 3.1: The 4th images shows very-poor,next left side to it poor ,next shows sub-optimal and the first image shows Optimal given rate from 0 to 3 consecutively.

0(Very-poor) = Unprepared colon segment with mucosa not seen due to solid stool that cannot be cleared.

1(Poor) = Portion of mucosa of the colon segment seen, but other areas of the colon segment not well seen due to staining, residual stool and/or opaque liquid.

2(sub-optimal) = Minor amount of residual staining, small fragments of stool and/or opaque liquid, but mucosa of colon segment seen well.

3 (Optimal)= Entire mucosa of colon segment seen well with no residual staining,small fragments of stool or opaque liquid.

The wording of the scale was finalized after incorporating feedback from three colleagues experienced in colonoscopy [52]

Stool varies in consistency from solid lumps to transparent water-diluted fluid.Our training data has tried to include all the cases according the Boston preparation score.

3.1.2 SVM And Medical Imaging

Support vector machine (SVM) learning has drawn considerable attention in the medical imaging technology due to its solid theoretical foundation and sensitivity to overfilling. Indeed, also it has been reported that SVM-based approaches are able to significantly outperform competing methods in many applications digit recognition,object recognition,face detection in image,hand written recognition and text categorization.SVM gets advantage by focusing on the training examples that are most difficult to classify.SVM has shown good result on detection of stool in the previous study's made by Sae Hwang [18], in object detection from medical images such as micro calcification [53].Recently svm multi-class with rbf kernel was also used by Shujin Zhu and Tizhoosh[54] to classify medical images in high speed and low memory requirement.

In this study color histograms a well known color features where feed to the svm kernel.As this features are not linearly separable so we used the rbf kernel to find a separable line between each of four class according to Boston bowel preparation.Polynomial kernel most of the time in practice are less use full due to performance and predictive reasons.so the best approach we come up with is to use linear SVM if your problem is linearly separable or to use rbf kernels when your problem is not linearly separable.

RBF kernels actually creates non linear combinations of our features to uplift our samples on to higher dimensional feature space where they can be linearly separated.

3.1.3 Color Features

Like many other computer vision algorithm in object recognition, first we need to extract some features from the input image.Our colonscopic images are modeled by RGB color space in which each color band is represented with 8-bit ranging from 0 to 255, and giving us a total of 255 by 3 potential colors.As we can see in the above pictures, it has some unique colors that can differentiate the stool color from the colon color.To model this color we have tested HSV,RGB,Luv color spaces in which we modelize the feature using the well known color feature color histogram.Generally there are two ways of extracting features which are global feature extraction and local feature extraction.In both cases the output of feature extraction would be a vector representation of the image. To find such a representation histograms are used extensively in computer vision.

For histogram computation, we quantize the color space into a number (k) of bins.Histograms also help us to see the stool color distribution in an image.Color histogram divides the color space into eight bins. The first color component from each color space is quantized into eight bins and the rest of the two color components are quantized into 8 bins; therefore we obtain total of 8x8x8 =512 bins.

The most important colors are yellow,pink and brown that helps us to differentiate the level of preparation, if we have more amount of brown color this shows that it totally bud and rated as 0 if it has yellow color it rated as optimal given score of 2 ,if it is totally pink it is rated as completely clean it is scored as 3.The most difficult ones are to differentiate between sub optima which are give a rate of 1 in Boston rating for preparation and optimal, as we have characterized to be if the color is combination of brown and yellow, meaning if we find a portion of yellow and a portion of brown we categorize it as sub-optimal(score of 1).This detection of stool is performed during insertion period as, which evaluates how was the patients preparation for the surgery ,the doctor has an option to clean using water the unseen part during withdrawal phase.

3.2 CNNs and Transfer Learning For Distention Classification

3.2.1 CNN and Medical Images

In recent years deep learning has shown an incredible progress in different fields of medical imaging Tumor detection,tracking tumor development,blood flow quantification and visualization,medical interpretation and diabetic retinopathy.However lack of large, annotated, and publicly available medical image databases has put a grate challenges in performing deep learning algorithms that do a high-level representations of knowledge through a large volume of annotated data.

Most recent research has addressed Deep Learning techniques in medical imaging by training the models from scratch[55],Such CNNs are often integrated into existing image analysis pipelines and replace traditional handcrafted machine learning methods. This is the approach followed by the largest group of papers in this survey made on [55].Further the survey has shown that most of the groups of researchers have shown different results using the same type of architecture and network.A key aspect that is often overlooked on the survey is that expert knowledge about the task to be solved can provide advantages that go beyond adding more layers to a CNN.Most of the successfully deep learning models have shown that using pre-processing and data augmentation can lead to a better trained models. For our distention classification we have also used some architectures to train from scratch.

Other recent studies shown the use of pre-trained CNNs as feature extractors and knowledge transfer from natural images to the medical imaging domain using off-the-shelf CNNs. Moreover, in [56], Van Ginneken et al. show that the combination of CNNs features and classical features for pulmonary nodule detection can improve the performance of the model.

Despite the difference between natural images and medical images researchers has shown some promising results by transferring knowledge using pre-trained models called off-the-shelf training,example [57, 58].we have also used and compared this techniques with training from scratch for colon distention classification.

3.2.2 Data

High resolution acquisition devices in colonscopic procedure has helped a lot of research and video analysis to be made in coloscopic videos.In this work we have use the same i-scan technology as in the data acquired for bowel preparation, from the same patients too. The data for distention was prepared in such a way that, how far the colon is distended , meaning instead of preparing data totally closed and totally open ones we prepared the same way to bowel preparation rating the distention from 0 to 3 (from Optimal to collapsed colon).

0(Collapsed) = The colon is completely collapsed or closed so no optimal inspection of the colonic mucosa can be achieved.

1(Partially collapsed) = The colon is partially opened but it is not possible to make optimal inspection.

2(Partially opened) = The colon is more opened and it possible to make inspection.

3(Optimal) = The colon is totally opened and its optimal to make inspection of the colonic mucosa and its widely open we can see the rings. Usually they can be described instinctively as rings or parts of rings (arches), which are not 100% circular, but rather triangular in shape with highly circular segment

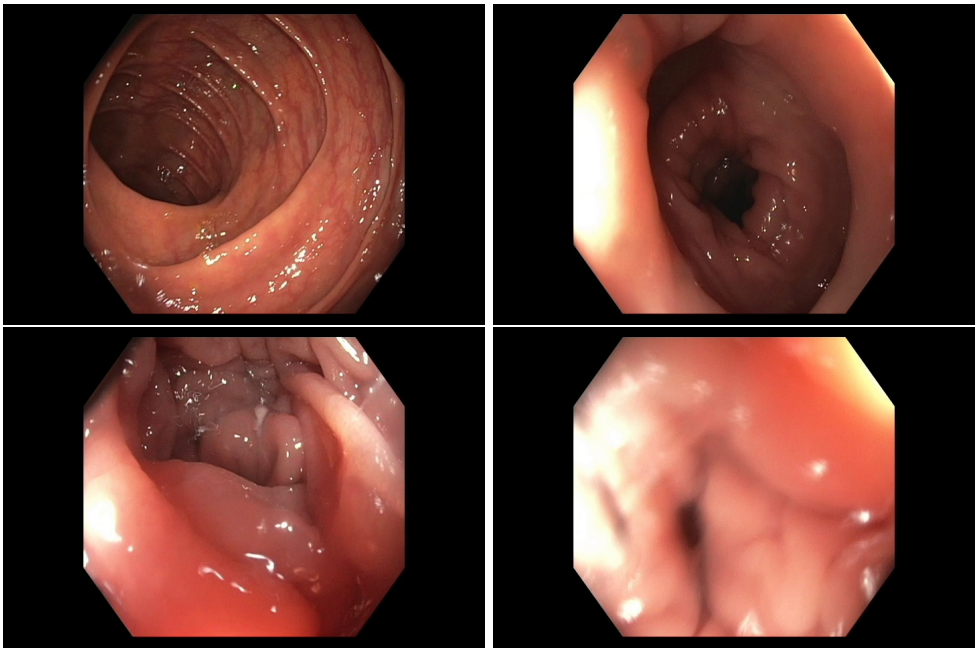

FIGURE 3.2: The 4th images shows bud,next left side to it shows sub-optimal ,next shows optimal and the first image shows good given rate from 0 to 3 consecutively.

The number of images for training and testing are shown in the table below according to the class of distention

	0(Collapsed)	1(Partially collapsed)	2(Partially opened)	3(Optimal)	Total
Training	238	171	137	118	664
Testing	100	100	100	100	400
Total	338	271	237	218	1064

Each class has been characterized that it contains different kinds of shapes that represent the given class, for example class 0 will contain different shapes of closeness and this closeness is the most frequent observed totally collapsed colon in most of the videos of colonoscopy operational shots, the same applies for other class .For more illustration lets see the figure 3.3 to 3.6.

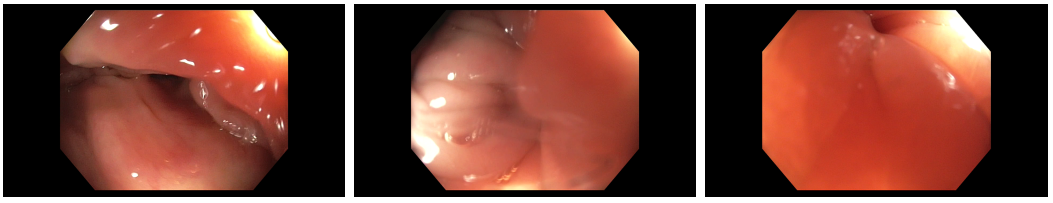

FIGURE 3.3: Different kinds of class 0(Collapsed colon)

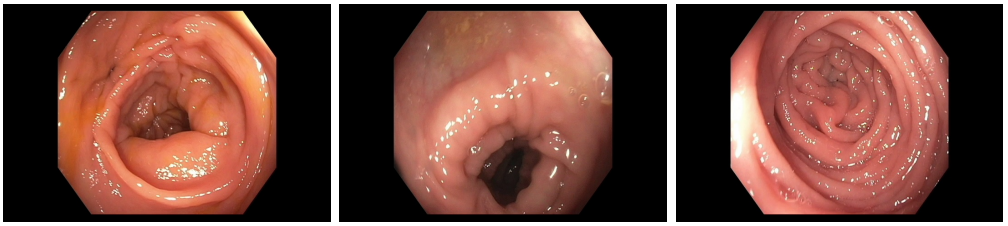

FIGURE 3.4: Different kinds of class 1(Partially Collapsed colon)

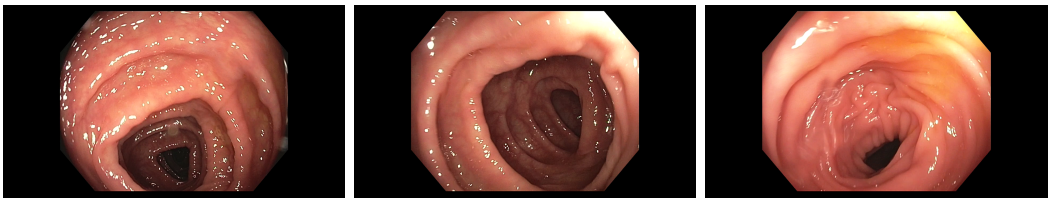

FIGURE 3.5: Different kinds of class 2(Partially Open colon)

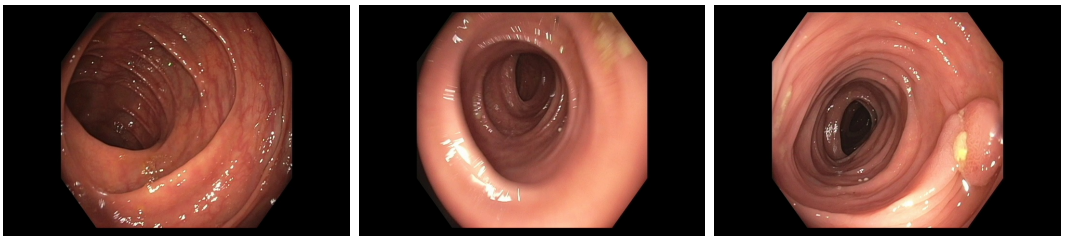

FIGURE 3.6: Different kinds of class 3(Open colon)

As we can observe in the images as we move from top to down we can see the level of openness increases,which means the level of comfort ability for the doctor to make the procedure and as we move from left to right we can see that in each class we have different shapes that humans can tell that it is close but difficult for computers.

3.2.3 CNN Techniques

CNNs are generally trained using large and diverse annotated data,but this is not always true specially in medical images and typically in our case of colon distention.CNNs are defined for solving computer vision problems in order to learn representative features with good generalization to outside world.To avoid over -fitting we used the well known systems of fighting data augmentation and normalization. In this experiment normalization is done by subtracting mean and dividing each input by its standard deviation.Data augmentation is performed by flipping and rotating the image 90^o to the left and right.

To train and to test our CNNs from scratch we have use the data described previously,based on the work of [59] trained with 128 x 128 sub images.As we are interested in the shape information, meaning for example in case of closed colon the edges we see on the frame are very often shorter, less clean, less curvy and definitely more winding that the ones in open colon.This information is mainly we will find not in the entire of the image but in the part of the images.

For the second experiment we have used different architecture than training from scratch,as we have few data to train, even though we have used data augmentation to overcome over-fitting and to improve robustness of the model there is highly likely we have over-fitted to our data set.So we explore off-the-shelf CNN trained to perform classification on the Image-Net challenge data.This is a part of transfer learning that uses pre-trained

model as a feature extraction,removing the last fully-connected layer (this layers outputs are the 1000 class scores for a different task like Image-Net), then treat the rest of the ConvNet as a fixed feature extractor for the new data-set and learn a linear classifier(SVM) over this features. we have tested recent deep learning architectures which are deep and have shown lower error rate.VGG19,VGG16,resnet,inception v3 which takes input of 224 x 224x3 .

1. Vgg-19 : This model is 19 layer network which uses only 33 convolutional layers stacked on top of each other in increasing depth. Reducing volume size is handled by max pooling. Two fully-connected layers, each with 4,096 nodes are then followed by a soft max classifier.

2. Vgg-16 : This model is 16 layer network which is proposed by the same group, with 3x3 convlutional layers stacked on the top of each other in increasing depth.

3. Resnet: This is much more deeper model than VGG models proposed by Microsoft researchers in 2015 which has shown an error rate of much smaller and state of the art in deep learning models proposed so far with an error rate of 3.6

4. Inception V3: This is based on parallel feature extraction system by computing 1x1,3x3 and 5x5 convolutions and stacking them together along the channel dimension and before being fed into the next layer in the network.The original incarnation of this architecture was called GoogLeNet, but subsequent manifestations have simply been called Inception VN where N refers to the version number put out by Google.

We formed a feature vector using this pre-trained models, as the images used for training this pre-trained models is to much different from the colonic images we have, so its recommended to use the previous layer of the pre-trained models.In deep learning models the last layers learn high level features of the data that we use to train the model,but the previous layers have rich low level features which may be also help-full in our clonic images. so we tested feature extraction performance using the previous layers in each of the above pre-trained models.For example feature vectors obtained from VGG19 from the previous layer have a size of 512x1.we have used keras with Tenserflow back-end in python environment to train all CNNs and to transfer learning from pre-trained models.

Chapter 4

Results and Discussion

4.1 Trained SVM for stool Classification

In this part of stool detection or colon preparation evaluation we have used the color feature as the main tool to classify the frames into four class of Boston bowel preparation.Colors like brown and yellow will represent the stool and pink will represent the color of the colon under the given color space.In the previous implementation before I was assigned in this project a global threshold system was implemented, meaning after finding the pixels with stool and with out stool, its ratio of the area that have stool and that doesn't contain stool in hsv color space is calculated, then if it's greater than the global threshold it said to be the specific frame has a stool(that is more yellow or brown pixels than pink pixels).The classification was based on finding a stool in specific frame without classifying it into some preparation format like Boston bowel class of preparation.However this implementation even though it's simple and classic way of object detection system it has shown good performance scoring of around 89% in correlation to GI doctors.The correlation is shown the figure bellow,We have prepared 15 videos to test this previous algorithm,the videos were evaluated from 0% to 100% scale by the computer and the Endoscopist ,as we can observe in the distribution of the points in the graph the algorithm has higher correlation in videos that are prepared well(i.e that doesn't contain stool).It has shown less performance in the videos that are moderate.

To make this algorithm more robust and to instead of binary classification of the frames into prepared and unprepared,we wont to learn over some predefined bowel preparation

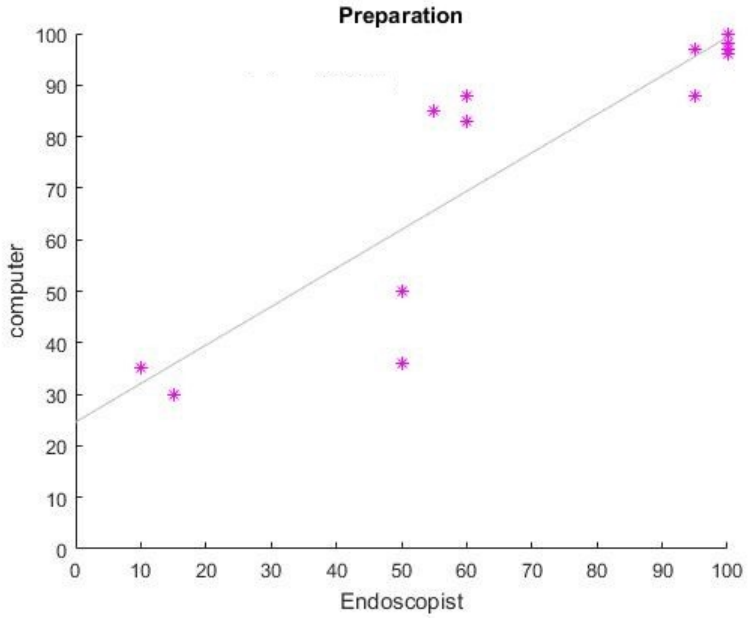

FIGURE 4.1: Correlation between endoscopist and computer

systems like Boston bowel preparation. As the color features have already shown good performance in the previous study made and test we made ,we decide to continue learning over this model and we propose SVM to learn classifiers among the classes.

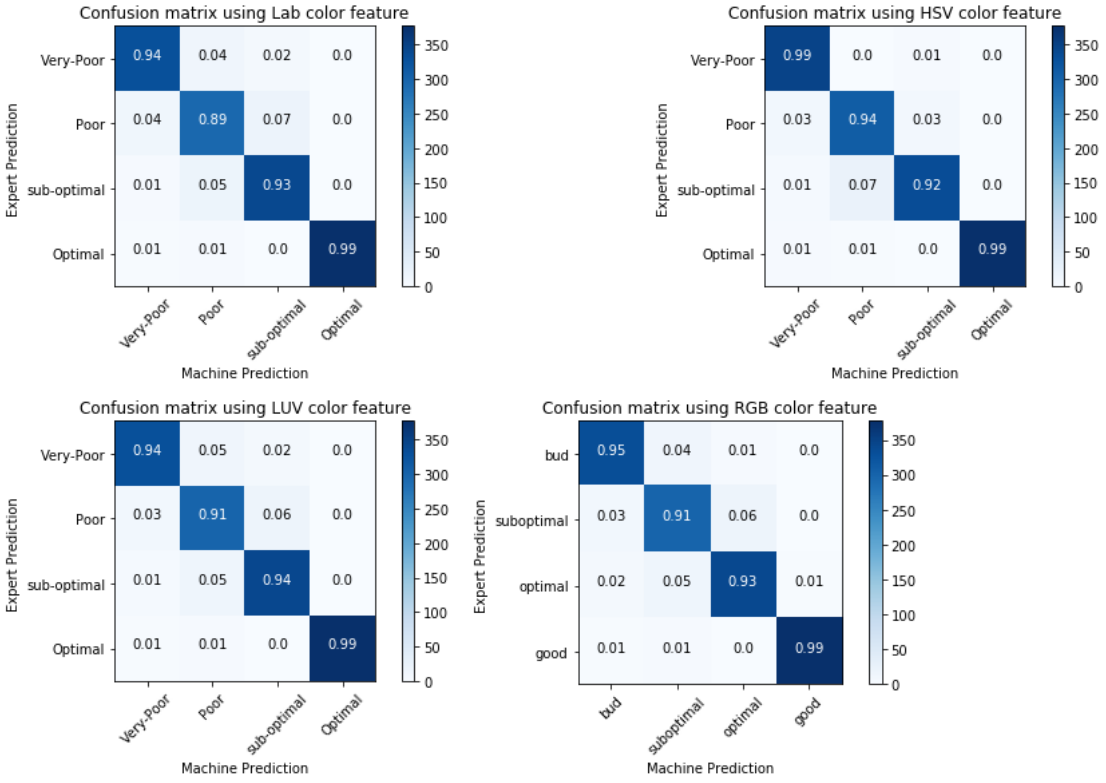

FIGURE 4.2: Stool classification using color features

When using SVM choosing kernel function and related parameters is very important and necessary,gamma parameter defines how far the influence of a single training example reaches, with low values meaning far and high values meaning close.C parameter trades off miss classification of training examples against simplicity of the decision surface. A low C makes the decision surface smooth, while a high C aims at classifying all training examples correctly by giving the model freedom to select more samples as support vectors. To correctly identify gamma and C several methods of optimization are proposed Genetic Algorithm (GA), k-Cross Validation (k-CV), Grid Search (GS) and etc. By comparison, the idea of GS is intuitive, which can search until get the optimized parameters of SVM. we have used Grid search as an optimization technique.The cross validated SVM parameter (C and gamma) used are 100 and 10.

As can be observed from the confusion matrices on average the result obtained from HSV is with accuracy of 96% while using Luv is 95% ,using Lab 93.4%and rgb is 94%.This experiment has shown that HSV color space has more discriminant power between pink and yellow or between pink and brown.

The results obtained from different color spaces,by using different color bins and the classification accuracy's are shown in the table below.

Bins	HSV			Luv			Lab			RGB		
-	Precision	recall	f1-score	Precision	recall	f1-score	Precision	recall	f1-score	Precision	recall	f1-score
8	96.00	96.00	96.00	92.00	92.00	92.00	92.00	92.00	92.00	96.00	96.00	96.00
16	97.00	97.00	97.00	97.00	97.00	97.00	97.00	97.00	97.00	96.00	96.00	96.00
32	96.00	96.00	96.00	96.00	96.00	96.00	97.00	97.00	97.00	95.00	95.00	95.00
64	96.00	96.00	96.00	95.00	95.00	95.00	96.00	95.00	95.00	96.00	96.00	96.00

4.2 Transfer Learning And Deep Learning For Distention Classification

4.2.1 Previous Implementation

In the previous implementation of colon distention classification using classical features such as shape features which includes, defining the area,perimeter,elogatedness,compactness of the detected blob where implemented to classify between closed and open colons in CMU CayLab,The proposed and implemented method before I was assigned to this project was shown in the figure 4.3 We have tested this algorithm by preparing 18

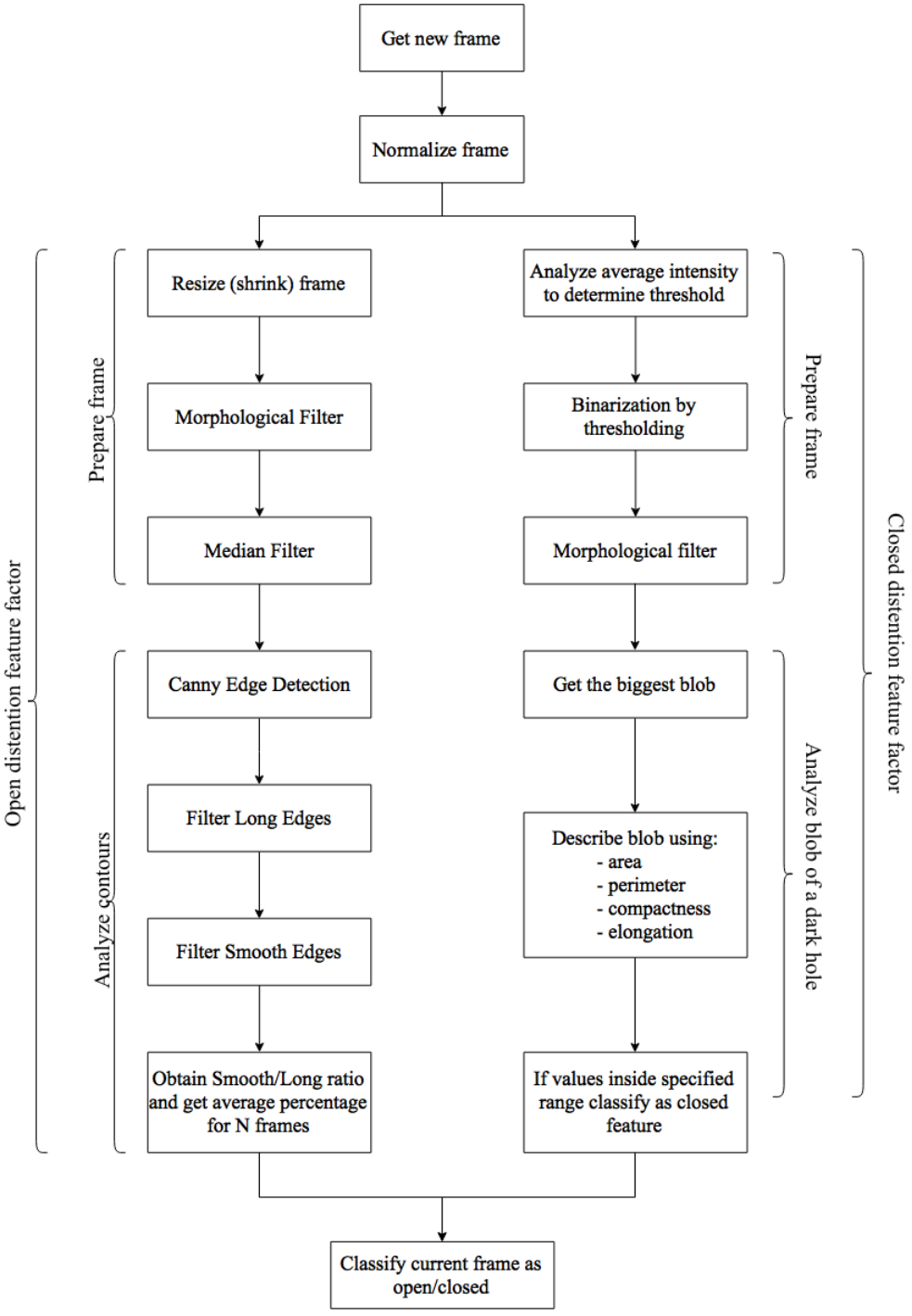

FIGURE 4.3: Previous method proposed for distention classification

videos of 5cm distance in the colon, that are prepared as videos that contain frames of poor,very-poor,sub-optimal and optimally distended colons.We made a correlation score method to see how the result is correlated to computers score and the result we found was around 67% .The problem we learned was the biggest blob area found for closed

frames and open frames is always in a certain range, meaning it is not linearly classifiable, we have frames that are totally closed but results in big area which we expect it to be only happening to open colons.The second problem is the kind of opening and closing of the colon experienced in several frames of the video is not the same for all closed frames and open frames,as a result the blob shape parameters we are getting may not be the same as we expect.

4.2.2 CNNs Trained from Scratch

In training CNNs from scratch we have tested three small networks based on the work done in [59],as bigger networks require more data a grate amount of computation in training .This networks has shown significant accuracy in detection of polyps and as we are using the same kind of images which are clonic images,we get motivated to test those architectures and to train them to learn over distention classification.we have used Gpu GeForce GTX 1070 for training the network .This architectures are designed in such a way that changing the size and number of filters as well as the number of units in the fully connected layer.We have divided the data in to training,validation and evaluation .We can see the architectures and their corresponding accuracy for distention classification of input size of 224 x224 x 3 in % in the table below.

Network index	Number of convolutional filters/size			Connected layer	Acc
	Layer 1	Layer 2	Layer 3		
CNN-01	48/11x11	72/5x5	512/6x6	512	85.00
CNN-02	24/11x11	48/5x5	1024/6x6	1024	83.00
CNN-03	48/11x11	72/5x5	1024/6x6	1024	87.00

Each convolutional layers are followed by max pooling,with stride of 1,Relu activation function, and padding.We have also added dropout regularization technique to avoid over fitting,randomly 25% of the neurons are ignored during training and updates are not applied to the neuron on the backward pass. In the above experiment the accuracy is the validation accuracy of the networks.We have trained soft-max classifier over this fully connected features which is the generalization of logistic function . CNN-03 has achieved better accuracy and the confusion matrix of this architecture is given in figure 4.4.

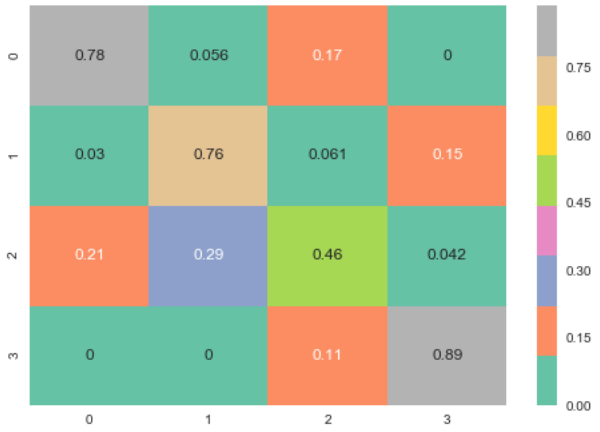

FIGURE 4.4: CNN-03 architecture confusion matrix result, 0-represents very-poor,1-represents poor,2-represents sub-optimal,3-represents optimal

The training accuracy obtained from CNN-03 looks good but when we try to see the result of every class accuracy with data prepared for evaluation,we have very bud accuracy for class sub-optimal openness of the colon, this is also even some times difficult to differentiate by using human eyes.As a result we can observe the confusion matrix for the new test data-set it has shown an average accuracy of 72.25%,further more it has shown that for the class of partial opening 46% has correctly classified as partial distended colon, but 42% has been identified as totally distended colons.The problem of lack of data still is an issue.Therefore, in order to try solving this problem, we also propose the use of transfer learning by pretrained CNNs that will be also explained in the next section.

4.2.3 Pretrained CNNs

In this section, we present using a pre-trained CNN, the last or next to the last linear fully connected layer is removed and the remaining pre-trained CNN is used as a feature extractor to generate a feature vector for each input image from a different database. These feature vectors can be used to train a new classifier (such as a support vector machine, SVM) to classify the images correctly. If the original database is similar to the target database, the probability that the high level features describe the image correctly is high and relevant to this new database. If the target database is not so similar to the original, it can be more appropriate to use higher level features, that is, features from previous layers of CNN.We used the same training data set used in training CNNs from

scratch.In this experiment the images are re-sized to 224 x 224 for Vgg's,Resnet and 299 x 299 for InceptionV3.

Network	precision	recall	f1-score
Vgg-16	78.00	79.00	78.00
Vgg-19	81.00	81.00	81.00
Resnet50	70.00	70.00	70.00
InceptionV3	85.00	83.00	83.00

The above result is obtained from the feature extracted from CNN prior to the full connected layer,it contains 512 feature vector.As we can observe from the results inceptionV3 or Google net has shown better accuracy than other networks.We can also see in the figure 4.5 bellow the confusion matrix obtained from Google net using the evaluation data-set.We can see in the confusion matrix it has shown better prediction of class 2 which is partial opening of the colon, training from scratch as shown in the previous section 46% of prediction where as using pre-trained Google net it has shown 64% accuracy.

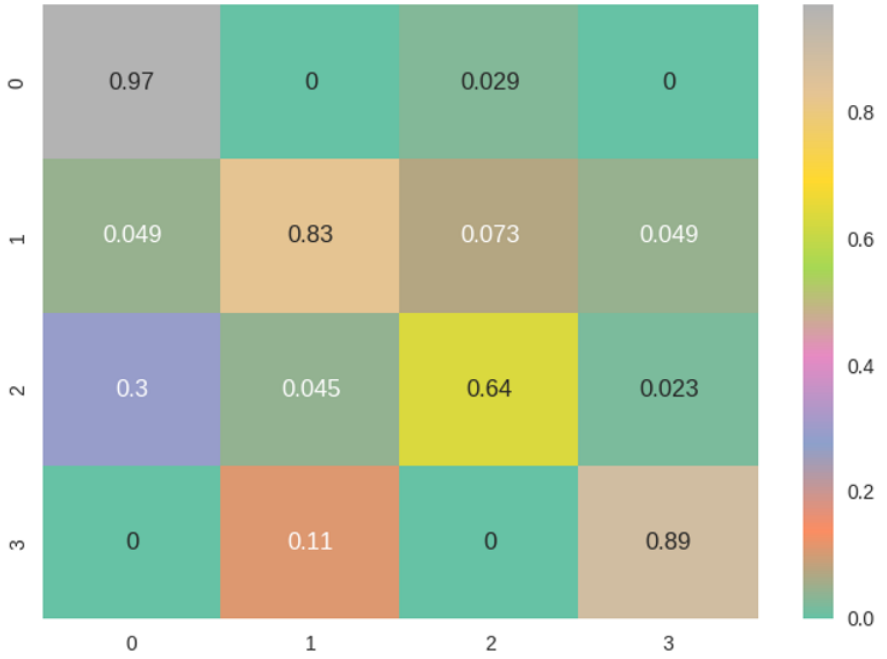

FIGURE 4.5: Confusion matrix using InceptionV3 CNN, 0(Very-poor),1(poor),2(sub-optimal),3(optimal)

Class	precision				Recall				f1 score			
	Vgg-16	Vgg-19	Resne50	InceptionV3	Vgg-16	Vgg-19	Resne50	InceptionV3	Vgg-16	Vgg-19	Resne50	InceptionV3
Very-poor(0)	82.00	82.00	75.00	84.00	85.00	85.00	78.00	89.00	84.00	84.00	74.00	89.00
Poor(1)	77.00	78.00	68.00	87.00	80.00	76.00	66.00	80.00	79.00	77.00	67.00	84.00
sub-optima(2)l	73.00	66.00	72.00	82.00	53.00	52.00	52.00	64.00	62.00	58.00	61.00	72.00
Optimal(3)	83.00	83.00	78.00	81.00	93.00	94.00	94.00	97.00	88.00	88.00	85.00	89.00

Each class score of precision,recall and f1 score are shown in the above table

In this work we have seen that when comparing small deep learning networks for small data set and transfer learning,transfer learning have shown better performance.Even though most papers have claimed that if there is huge and variety amount of data more deep networks will be more sensitive to over-fitting and give high accuracy,in this thesis we have seen that transfer learning has also grate role in overcoming the problem of availability of huge data.

Chapter 5

Conclusion

In this work we have explored automatic recognition of stool and distention classification as part of analysis of colonscopic videos, which helps us to automatic discovery of the medical knowledge by parsing the colonoscopy videos into semantic units is highly desirable and very useful for procedure quality measurement,improving endoscopists procedural skills and educational activities(presentations,teaching of fellows, manuscripts, etc.).The presence of stool in most of the frames of the video will indicate that the colon preparation was bud and it will not be appropriate to go through the procedure of inspection of colon for finding polyps.The proper distention of the colon,meaning we have more frames classified as optimal distended colons will also indicate that its easy to explore the surface of the colon ,during the inspection for polyp identification by the GI doctors.

In this work we have defined two datasets of operational shots: a colonscopic seen with frames of stool at different level of preparation according to Boston bowel preparation,a colonscopic seen with frames that show different levels of distention given score of 0 to 3(very-poor to optimal).Because of the unique characteristics of colonoscopy frames novel and robust algorithms for detecting of stool and distention classifications where explored.For stool detection we proposed feature based technique, in this technique there are two major steps used extracting color features in different color spaces where hsv color space has shown better performance than other color spaces like (RGB,Luv,Lab) and learning using SVM multi-class classifier using RBF kernel over the features extracted.In this technique even though previous methods have shown significant results for frame

classification frames that contain stool and that are super clear, the algorithms where not robust, meaning the doctors has indicated that they can make the procedure even though it is not super clean so the algorithms need to be flexible in analyzing the video,for example the doctor can still make colon inspection if the preparation level is "2" it necessary need not be level "3", so during analysis it use-full to evaluate in this way.

For distention classification we tested the previous classical algorithm which used shape features as means of classification of distended and closed(collapsed) colon,In this technique used the main features are area,perimeter,elongatedness,compactness of the detected blob combining this features and learning method actually was made manually and it has show an accuracy of 67%.Depending on this and exploring that colon frames have variety types of distention we proposed to see the state of the art classification algorithms such as Deep learning and transfer learning. we explored deep learning trained from scratch and using by transfer learning methods with pre-trained networks(off-the-shelf training).We have compared both of this techniques and we have seen that off-the-shelf training which is made by removing the fully connected layer and making the remaining network as a feature extraction model, then training a linear classifier over over this features.On this techniques we have compared 4 recent much deeper networks(Vgg-16,Vgg-19,Resnet50,InceptionV3) and InceptionV3 has shown an an accuracy of 85% and generally also better class wise prediction as shown in the previous chapter using the confusion matrix.

Chapter 6

Future Work

Our feature work will include increasing the performance of Distention and testing this algorithm on different patients live and recorded video.As transfer learning are showing significant impact in learning and classification of different scope of problems for future it would be interesting to see if the performance can be increased by finding a pre-trained models on colonscopic,such as networks trained over data set prepared for polyp detection and use this networks to trasfer the weights they learned and use them for different analysis that can be performed in colonscopic videos.Combining classical shape features with deep learning features and training them using SVM would also be another way to boost the performance of the algorithm.

For stool detection this has shown good performance so far and for future work it needs to be tested in live and try to see if we can have the same performance.As with the previous algorithms proposed we have tested it in live and it has shown good performance, which indicated that color features have promising discriminant power..

As colonscopic video analysis is not only this two subjects, it has more information that can be gained from the video or can be evaluated automatically and become help full for the GI doctors,such as surface area evaluation,uninformative frame detection,end of colon(cecum) detection,audio analysis,video clarity and others would also be future works of this thesis.

Bibliography

[1] Peter G. Sharma, Robyn S. Rossos. A review on the quality of colonoscopy reporting. *Journal Article*, 6, 2016. URL https://www.hindawi.com/journals/cjgh/2016/9423142/cta/.

[2] F. P. Vleggaar P. D. Siersema A. M. Leufkens, M. G. H. van Oijen. Factors influencing the miss rate of polyps in a back-to-back colonoscopy study. *Thieme Endoscopy*, March 2012. URL https://www.thieme-connect.com/DOI/DOI?10.1055/s-0031-1291666.

[3] Jacobson BC Calderwood AH. Colonoscopy quality: Metrics and implementation. *Gastroenterology clinics of North America*, September 2013.

[4] MS. Cappell and D. Friedel. Computer-aided detection of diagnostic and therapeutic operations in colonoscopy videos. *IEEE International Conference on, pp. 648-653*, 2009. URL http://ieeexplore.ieee.org/document/4237329.

[5] Methods of luminal distention for colonoscopy. 2013. URL http://www.giejournal.org/article/S0016-5107(12)02752-6/pdf.

[6] MS. Cappell and D. Friedel. The role of sigmoidoscopy and colonoscopy in the diagnosis and management of lower gastrointestinal disorders: endoscopic findings, therapy, and complications,. *The new england journal of medicine*, July 2002. URL http://www.medical.theclinics.com/article/S0025-7125(02)00077-9/pdf.

[7] SaeHwangb.JeongKyuLeeb.Piet C.de Groend JungHwanOh. Informative frame classification for endoscopy video. *IEEE International Conference on, pp. 648-653*, 2009. URL http://www.sciencedirect.com/science/article/pii/S136184150600079X?via.

[8] Maria Trujillo Cristian Ballesteros and Claudia Mazo. Automatic classification of non-informative frames in colonoscopy videos. *IEEE International Conference on, pp. 648-653*, 2015. URL http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7818267.

[9] Maria Trujillo Cristian Ballesteros and Claudia Mazo. Co-occurrence and morphological analysis for colon tissue biopsy classification. URL http://www.dcs.warwick.ac.uk/~nasir/papers/fit06.pdf.

[10] Christel Ducroz Robin Tournemenne Beryl Cummings Roman Thibeaux Nancy Guillen Alfred O. Hero Jean-Christophe Olivo-Marin Alexandre Cecilien Dufour, Tzu-Yu Liu. Hyperspectral colon tissue classification using morphological analysis. *Annual IEEE Conference, pp. 1-4*, 2008. URL http://ieeexplore.ieee.org/document/4136915.

[11] Lama Hassan Ahmad Chaddad, Christian Desrosiers. Multispectral texture analysis of histopathological abnormalities in colorectal tissues. *Annual IEEE Conference*, 2016. URL http://ieeexplore.ieee.org/document/7532835/.

[12] M. Pietikainen T. Ojala and D. Harwoo. A comparative study of texture measures with classification based on feature distribution. *Pattern Recognition*, 1996.

[13] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *Proceedings of the 3rd International Conference on Analysis and Modeling of Faces and Gestures*, 2007.

[14] XR. M. Haralick. Statistical and structural approaches to texture. *Proceedings of the 3rd International Conference on Analysis and Modeling of Faces and Gestures*, 1979. URL http://ieeexplore.ieee.org/abstract/document/1455597.

[15] Mutawarra Hussain Abdul Jalil Saima Rathore, Muhammad Aksam Iftikhar. A novel approach for ensemble clustering of colon biopsy images. *2013 11th International Conference on Frontiers of Information Technology*, 1979. URL http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6717220.

[16] Robert L. Van Uitert Jiamin Liu Joel G. Fletcher Armando Manduca Marius George Linguraru, Shan Zhao and Ronald M. Summers. Cad of colon cancer

on ct colonography cases without cathartic bowel preparation. *Annual International IEEE EMBS Conference*, 2008. URL http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4649833.

[17] Wallapak Tavanapong Sae Hwang, JungHwan Oh. Stool detection in colonoscopy videos. *Annual International IEEE EMBS Conference*, 2008. URL http://ieeexplore.ieee.org/document/4649835/authors.

[18] Wallapak Tavanapong Piet C. de Groen JungHwan Oh, Sae Hwang and Johnny Wong. Blurry frame detection and shot segmentation in colonoscopy videos. 2008. URL http://proceedings.spiedigitallibrary.org/.

[19] T.and Vijayan Asari K. Tian, H. Srikanthan. Automatic segmentation algorithm for the extraction of lumen region and boundary from endoscopic images. *Medical and Biological Engineering and Computing*, 2000. URL https://rd.springer.com/article/10.1007/BF02345260.

[20] S.Kumar.K.V.Asari' and D .Radhakrishnan'. Online extraction of lumen region and boundary from endoscopic images using a quad structur. *IET Conference Publications*, 1999. URL http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=791175.

[21] Jianming Liang Nima Tajbakhsh, Suryakanth R. Gurudu. Automated polyp detection in colonoscopy videos using shape and context information. *IEEE Transactions on Medical Imaging*, 2015. URL http://ieeexplore.ieee.org/document/7294676/.

[22] Jianming Liang Nima Tajbakhsh, Suryakanth R. Gurudu. Automatic polyp detection in colonoscopy videos using an ensemble of convolutional neural networks. *IEEE Transactions on Medical Imaging*, 2015. URL http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7163821.

[23] Myunggi Lee Sungheon Park and Nojun Kwak. Polyp detection in colonoscopy videos using deeply-learned hierarchical features. *IEEE Transactions on Medical Imaging*, 2015. URL http://mipal.snu.ac.kr/images/0/0b/Polyp_short_report.pdf.

[24] Wallapak Tavanapong Johnny Wong Piet C. de Groen ae Hwang, JungHwan Oh. Polyp detection in colonoscopy video using elliptical shape

feature. 2006. URL https://ai2-s2-pdfs.s3.amazonaws.com/1a95/e18120f80c49166fe730e8492b18663e452d.pdf.

[25] Wallapak Tavanapong Johnny Wong Piet C. de Groen Ae Hwang, JungHwan Oh. Integrating online and offline three-dimensional deep learning for automated polyp detection in colonoscopy videos. *IOMEDICAL AND HEALTH INFORMATICS*, 2016. URL http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7776845&tag=1.

[26] Kimmo Kansanen Tor Audun Ramstad Seyyed Hamed Fouladi, Ilangko Balasingham. Extracting remote photoplethysmogram signal from endoscopy videos for vessel and capillary density recognition. *IEEE 38th Annual International Conference*, 2016. URL http://ieeexplore.ieee.org/document/7590681/.

[27] Haidi Ibrahim Rostam Affendi Hamzah and Anwar Hasni Abu Hassan. Stereo matching algorithm based on illumination control to improve the accuracy. *Image Analysis Stereology*, 2016.

[28] Template matching using fast normalized cross correlation. URL https://isas.uka.de/Publikationen/SPIE01_BriechleHanebeck_CrossCorr.pdf.

[29] C. K. Williams M. Revow and G. E. Hinton. Sharing visual features for multiclass and multiview object detect ion. *Massachusetts Institute of Technology*, 2004.

[30] Alex Matthew, Turk. Face recognition using eigenfaces. *Massachusetts Institute of Technology*. URL https://www.cs.ucsb.edu/~mturk/Papers/mturk-CVPR91.pdf.

[31] Olac Fuentes Luis Malago, Borja. Object detection using image reconstruction with pca. *Image and Vision Computing xxx (2007) xxxxx*, 2007. URL https://pdfs.semanticscholar.org/f1d1/60add7d1d1d7ffe200a7a2b1275dd3879223.pdf.

[32] CORINNA CORTES VLADIMIR VAPNIK. Support-vector networks. 1995. URL http://image.diku.dk/imagecanon/material/cortes_vapnik95.pdf.

[33] Chih-Chung Chang and Chih-Jen Lin. A library for support vector machines. *Advances in Neural Information Processing Systems*, 2001. URL http://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf.

[34] M.Nielsen. Neural networks and deep learning is a free online book. 2017. URL http://neuralnetworksanddeeplearning.com.

[35] Ilya Sutskever Alex Krizhevsky and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural network. 2012. URL https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

[36] Yoshua Bengio Yann LeCun, L eon Bottou and Patrick Haffner. Gradient-based learning applied to document recognition. 1998. URL http://yann.lecun.com/exdb/publis/pdf/lecun-01a.pdf.

[37] Karen Simonyan and Andrew Zisserman. Verry deep convolutional network for large-scale image recognition. 2014. URL https://arxiv.org/pdf/1409.1556.pdf.

[38] Yangqing Jia Pierre Sermanet Scott Reed Dragomir Anguelov Dumitru Erhan Vincent Vanhoucke Andrew Rabinovich Google Inc.University of North Carolina Chapel Hill University of Michigan Ann Arbor Magic Leap Inc. Christian Szegedy, Wei Liu. Going deeper with convolutions. 2015. URL http://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/Szegedy_Going_Deeper_With_2015_CVPR_paper.pdf.

[39] Sergey Ioffe Jonathon Shlens Christian Szegedy, Vincent Vanhoucke. Rethinking the inception architecture for computer vision. 2015. URL https://arxiv.org/pdf/1512.00567.pdf.

[40] Shaoqing Ren Jian Sun Kaiming He, Xiangyu Zhang. Deep residual learning for image recognition. 2015. URL https://arxiv.org/pdf/1512.03385.pdf.

[41] Jiri Matas Dmytro Mishkin. All you need is a good init. 2016. URL https://arxiv.org/pdf/1511.06422.pdf.

[42] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep learning. 2016. URL http://www.deeplearningbook.org/contents/optimization.html.

[43] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. 2009. URL http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5288526.

[44] meky El-gayarOpens, SolimanOpens. A comparative study of image low level feature extraction algorithms. 2013. URL http://www.sciencedirect.com/science/article/pii/S1110866513000248#b0190.

[45] Tom Drummond Edward Rosten, Reid Porter. Faster and better: A machine learning approach to corner detection. 2010. URL http://ieeexplore.ieee.org/document/4674368/.

[46] D. Harwood T. Ojala, M. Pietikainen. Performance evaluation of texture measures with classification based on kullback discrimination of distributions in pattern recognition. 1994. URL http://www.sciencedirect.com/science/article/pii/S0167865502000569.

[47] M. Idrissa and M. Acheroy. Texture classification using gabor filters. 2002. URL http://www.sciencedirect.com/science/article/pii/S0167865502000569.

[48] A. Koike H. Sankoh A. Suda, H. Murakami and S. Naito. Performance comparison between sift and surf for feature points matching in dynamic calibration with zoom camera. 2012.

[49] Tinne Tuytelaars Herbert Bay and Luc Van Gool. Surf:speeded up robust features.

[50] Ashwini Verma Vivek Singh Rathore1, Messala Sudhir Kumar. Colour based image segmentation using l*a*b* colour space based on genetic algorithm. 2012. URL http://www.ijetae.com/files/Volume2Issue6/IJETAE_0612_28.pdf.

[51] Budapest Tech. Color content based image classification. 2007. URL http://users.nik.uni-obuda.hu/sergyan/Publications/2007_SAMI.pdf.

[52] Audrey H.Calderwood MD Gheorghe Doros PhD Oren K. Fix MD MSc and Brian C. Jacobson MD MPH FASGE Edwin J.Lai, MD. he boston bowel preparation scale:a valid and reliable instrument for colonoscopy-oriented research. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2763922/.

[53] M.N. Wernick El-Naqa, Yongyi Yang. A support vector machine approach for detection of microcalcifications. 2002. URL http://ieeexplore.ieee.org/document/1176643/.

[54] H.R. Tizhoosh Shujin Zhu. Radon features and barcodes for medical image retrieval via svm. 2016. URL http://ieeexplore.ieee.org/document/7727867/authors.

[55] Babak Ehteshami Bejnordi Arnaud Arindra Adiyoso Setio Francesco Ciompi Mohsen Ghafoorian Jeroen A.W.M. van der Laak Bram van Ginneken Clara I. Sanchez Geert Litjens, Thijs Kooi. A survey on deep learning in medical image analysis. 2017. URL https://arxiv.org/pdf/1702.05747.pdf.

[56] Jacobs C. Ciompi F Van Ginneken B., Setio A. A. A. Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans. 2015. URL http://ieeexplore.ieee.org/document/1176643/.

[57] Jacobs C. Ciompi F Van Ginneken B., Setio A. A. A. Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans. 2015. URL http://ieeexplore.ieee.org/document/1176643/.

[58] Suryakanth R. Gurudu Nima Tajbakhsh, Jae Y. Shin. Convolutional neural networks for medical image analysis: Full training or fine tuning. 2016. URL http://ieeexplore.ieee.org/document/7426826/.

[59] Georg Wimmer Eduardo Ribeiro, Andreas Uhl and Michael Hfner.