NUST

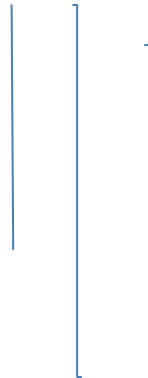## NATIONAL UNIVERSITY OF SCIENCES & TECHNOLOGY

MICROSCOPE BASED MEDICAL IMAGE SCANNER WITH DEEP LEARNING APPLICATIONS

PROJECT BY:

TALHA MEHBOOB

MOMIN AHMAD KHAN

MUHAMMAD UZAIR KHATTAK

MUHAMMAD ZEESHAN KARAMAT

ADVISOR:

MR. IMRAN ABEEL

CO-ADVISOR:

DR. FAISAL SHAFAIT

Dedication

This project is wholeheartedly dedicated to our dear parents, who have been our source of inspiration and gave us strength when we thought of giving up. Also, to our brothers, sisters, relatives, mentors, friends, and classmates who encouraged us and provided us advice which enabled us to finish this project. And finally, we dedicate this project to Almighty Allah, for the guidance, strength, power of mind, protection, and skills.

<u>Acknowledgements</u>

We acknowledge the guidance and help of the following people. Their contribution plays an immense role in the completion of our project. We thank them for the time and energy they have invested in this project.

- Sir Imran Abeel: Our advisor
- Sir Faisal Shafait: Our co-advisor
- Sir Hasan Aqeel: Our mentor
- Sir Yasir: Lab Technician at Basic Electronics Lab
- Syed Wajih Hasnain: Graduate Researcher at SIGMA Lab
- Muhammad Zaid Mughal: Our batchmate and dear friend

Contents

List of Figures

List of Tables

Abstract

D igital pathology and deep learning algorithms are gradually paving way for automated diagnosis and have the potential to significantly improve cancer screening, diagnosis, and prognosis. However, the Whole Slide Imaging (WSI) scanners currently available for slide digitization are expensive, and inaccessible for hospitals and laboratories in developing countries. We present "POPScan", a low-cost WSI system designed specifically for resource and budget-constrained healthcare systems of developing countries. The modular and portable nature of "POPScan" allows scaling and sharing of the scanner between different hospitals and labs. We plan to make the complete design open source for modification, improvement, and adaptation to a wide range of microscope and camera models. POPScan consists of stepper motors, a digital camera and 3D-printed components attached to a microscope to convert it into a WSI scanner. The accompanying software has motor speed and direction control, autofocus and stitching algorithms that generate WSI images in SVS format. The overall cost of the POPScan components is less than \$200 (excluding the cost of the microscope). The quality of the generated images was systematically evaluated by three pathologists and found to be of high quality. Furthermore, the scanned slides were also employed for training a deep neural network for epithelium segmentation and the results were found to be comparable with slides scanned by a commercial WSI scanner.

Introduction

Throughout the course of history, Histology has proved itself to be a viable asset for the pathologists to study the root causes and repercussions of diseases by analyzing the structure of tissues and cells present in body [1]. Pathologists have always endeavored to incorporate the modern and novel approaches and technologies to obtain, analyze and process the information derived from the microscopic analysis of cells, tissues, and organs [2]. One of the modern techniques that has altered the way of analysis of pathological test pieces is Digital pathology [3]. This involves acquiring digital images of the specimen which are examined by the pathologists to unveil any abnormality in the patterns and structures of biological tissues and cells.

The cardinal part of digital pathology is the Whole Slide Imaging (WSI) [4], which is the procuring of scanned digitized images of glass slides by using Whole Slide Image Scanners. These devices are comprised of different modules that accomplish scanning over whole tissue slides using various well-defined scanning algorithms and attaining sharp images of small patches employing auto-focusing algorithms [5]. These images obtained of small patches are then stitched together [6] to obtain a large Whole Slide Image of size commonly ranging from 0.2 GB to 1 GB [2].

The significance of Whole slide imaging can be observed as there is scanty number of pathologists available for the expeditious analysis of pathological slides in developing world which create numerous complications in the process of diagnosis and often lead to misidentification of ailment causing health and viability issues of patients [7]. Thus, here the high-resolution digitized images can be transmitted via electronic channels for the recognition of abnormalities in histologic slides and judgmental discussions [8].

The commercially available Whole Slide Image Scanners, which are normally used to obtain scanned virtual images of pathological slides are worth a pretty penny [9], thus they cannot be deployed in locations with fewer and insubstantial resources. The poor infrastructure and lack of availability of adequate capital and funds is the major reason of inaccessibility of such diagnostic devices in developing countries which create a void that devours the certainty of life, for numerous ailments could not be identified timely [10]. Thus, there is a need to substitute these costly diagnostic devices with affordable, low-cost implements that are readily available for

rapid diagnosis of both contagious and non-contagious diseases and that does not require a skilled professional to operate it [11]. Also, the exorbitant prices of premium scanners are inspiring many of researchers to generate affordable, low-priced medical devices such as automated, low-cost microscopes. These implements have shown promising results in numerous realistic problems against their analogues [12] [13] [14] [15].

Literature Review

There is a significant trade-off between ease-of-use and diagnostic performance. The ease of handling the system and cost effectiveness decreases as the degree of automation increases. However, a higher degree of automation increases the diagnostic performance. It is significant to observe that the diagnostic performance of these automated, low-cost implements is considerably reduced when compared to their analogues [16] [17].

A low-cost user-friendly version without auto-focusing has been implemented by [6] [18] [19]. The drawback of such an implementation is that in the presence of 3D tissues or topographic variations [20], manual focusing is required, and since this system provides automated movement only in x-y directions and no such movement in z directions, makes the scanning of 3D tissues via manual movements a tedious and laborious work.

This implementation uses Microsoft Image Composite Editor (ICE) for image stitching. However, ICE has some color consistency issues because of uneven illumination [21]. Stitching results obtained from ICE also have some geometric inaccuracies [22].

In this implementation, it is not possible to detect the 3D tissues which cause the whole slide image to be a bit blurry at some specific locations and the color illumination issues and geometric inaccuracies while stitching using ICE eventually leads to reduced diagnostic performances.

Auto-Scope by [23] moves in x and y directions of a 0.1 microliter blood-stained slide to obtain at least 300 fields of views (FOV) and provides the auto-focus capabilities to obtain the maximum focused image at a particular FOV. Each image corresponding to single field of view is stored and thus used for the detection of Malaria Parasites.

However, there is no mosaicking or stitching of those images corresponding to each FOV to create a whole slide image, which could enable the pathologist to look at all the FOVs at the same time. Also, the whole setup costs around 1500 USD to 4000 USD [23] which is meagre when compared to Standard Whole Slide image scanners, but still very high to be used in resource constraint environments as in developing countries.

Lu et al. [9] present a machine to be used in resource constraint environments at a proposed manufacturing price of around 400 - 600 USD. This model can auto-scan slides in x-y directions in a range of 13 mm and provides z-axis movement for auto-focusing module. The images obtained for each field of view are then stitched together using ImageJ stitching plugin by Fiji to generate a Whole Slide Image. This implementation acquires the "Grid Collection" mode of ImageJ which demands all the tiles to be perfectly rectangular in geometry and requires them to be named after their position in rectangular grid. Due to the imperfect movements of motors during the scanning process, it is not guaranteed to obtain sequential, and grid acquired images which may lead to an arduous situation [9].

Moreover, for some datasets with large scale 3-Dimensional images, the ImageJ (by Fiji) exceeds the physical memory easily [24]. Also, during the stitching of a large number of small images or tiles, it may require the images to be positioned manually [24] which could present a challenge due to the complexity it involves.

"InstantScope" [20] is another effort in the domain to build a hardware to generate the Gigapixel size images of biological slides. This design is primarily focusing to provide a novel auto-focusing module with 2 pinhole modulated cameras installed at the eye-piece lenses, an inspiration they got from professional photography to provide the instant focal point. They system can provide an auto-focusing in the range of millimeters [20]. The system's acquisition time for pathology slide with tissue area being 1.4 cm wide is 90 seconds using the 9-Megapixel CCD Camera and with field of view of 14mm by 8mm, however acquisition time for a 1.5cm blood smeared slide is 16 minutes using the color CMOS sensor with a field of view of 15mm by 15mm. The system is a low-cost implementation, and it uses a commercially available software, "MS ICE" for the stitching of the tiles to generate the Whole Slide Image. However, ICE has some color consistency issues because of uneven illumination [21]. ICE also has some geometric inaccuracies [22]. Also, MS ICE while stitching does not care for the positional information of individual segments [20]. The stitching time for a pathology slide with tissue area being 1.4 cm wide is almost 40 minutes.

Qidwai et al. [25] proposed an image stitching system along with a microscopic setup to provide an alternative to the expensive WSI scanning machines. Its system can stitch and generate WSI's based on scans with 10x, 40x, and 100x magnifications. However, the stepper motors

which are being used have a low resolution in terms of micro-stepping., i.e., step angle of 5.625°/64 steps [25]. Also, the scanning system does not provide automatic auto-focusing, so this focusing is done manually with the help of a person. Alongside with that, this system uses a digital camera, with "Amscope" adaptor, to mount on the microscope. As normal, digital cameras are not natively built for acquiring microscopic images, it can provide some compromise regarding frame's quality.

To the best of our knowledge, there are no works in existing literature that compare the performance of deep learning on smartphone and WSI acquired image datasets. We presented preliminary results on smartphone and WSI datasets in [26] however, only mitosis detection was examined.

Furthermore, due to the non-availability of a WSI scanner at the time, the benchmark WSI slides were taken from a different, publicly available, dataset (breast cancer) [27] than the smartphone slides (OSCC). Some similar efforts do exist, for example: a tele pathology application for smartphone users was presented in [28]; in [29] it was demonstrated that a smartphone camera with an additional zoom lens can act as a cheap alternative for light microscope with only a slight drop in diagnostic performance. Similarly, in [30] it was demonstrated that pathological images captured with smartphones can be used for pathological analyses. However, none of these works examined the impact of low- cost/smartphone-based image acquisition on AI and deep learning algorithms.

Problem Statement

Every year 100,000 people in Pakistan die of cancer [31]. The number is expected to rise and across the globe too. Global cancer cases are projected to rise by 70% within the next two decades. Besides cancer, there is a plethora of other diseases including infectious diseases, blood disorders and AIDS. This is an alarming situation and there is a dire need to develop tools and techniques that aid in the diagnosis and treatment process of cancer.

Pathology is a common technique for diagnosis of cancer. It involves the study of causes and origin of the disease. It is achieved through the examination of body tissues and body fluids. The microscope is perhaps the most important tool used by pathologists.

In an increasingly digital world, more and more processes and techniques are being digitized to improve their quality and standard. The field of pathology is no different and conventional pathology techniques are being gradually replaced by digital ones. One such example is the Whole Slide Imaging (WSI) Scanner. It can take a slide as input and then producing its high-quality digital version. It serves several important functions; the slide can be examined in greater detail at different zoom levels, the slide can be shared over the internet as it is now in digital form, and lastly, the slide can be used for further processing such as applying state-of-the-art Deep Learning techniques for diagnosis of various diseases.

There are two major drawbacks, however. The first one is that the commercially available WSI Scanners are too expensive to be used in a third-world country like Pakistan. A standard WSI Scanner costs around 100,000 USD [32]. This is an exorbitant price tag for a developing country like Pakistan. The second drawback is that there is a lack of specialist pathologists in Pakistan, as in other developing countries.

Therefore, by keeping the above discussion in mind, there is a dire need to fill the void in the current scenario of digital pathology in Pakistan. The solution must tackle both the issues at hand i.e., high cost of commercially available WSI Scanners and the shortage of pathologists in Pakistan.

Methodology

Our Approach

Our approach to tackle the above-mentioned problems is to design a Whole Slide Imaging Scanner that is low in cost and easy to use. We are using our knowledge of electronics, embedded systems, and 3D modelling to design a WSI Scanner by modifying a conventional microscope. To ensure a focused slide image, we incorporate a feature for auto-focusing the slide without human aid. For obtaining the slide in digital form, we are using motor action to enable motion along the 3 axes. In order to condition the image frames and stitch them together, we used several image processing and computer vision techniques. Finally, a graphical user interface (GUI) was developed to operate all these functions from a single platform with ease.
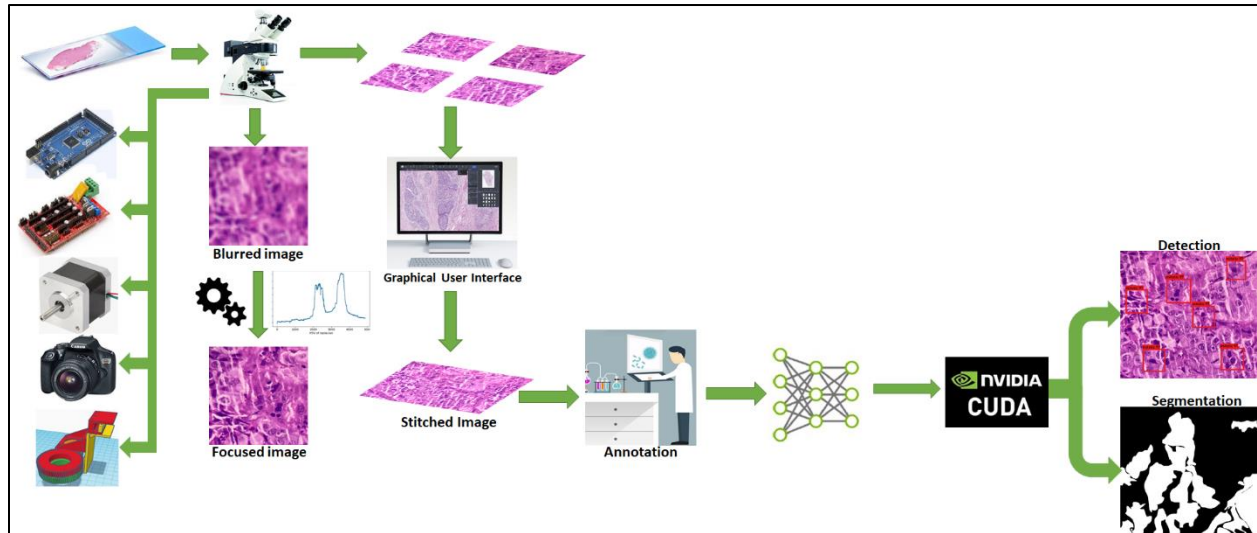
Figure 1: Pipeline of Our Project

Tools/Software/Libraries Used

- Nema 17 Stepper Motors: Calibrated at 24steps/mm for y-axis and 50steps/mm for x-axis.

- Arduino Mega: Used as the interface between motors and our computer by making use of serial communication.

- Ramps 1.4: Arduino Mega Shield used for efficient connections with motor drivers.

- DRV 8825 and A4988: Motor Drivers with the feature to limit current and control motor step size.

14

- 12V Power Supply: Used to provide a constant voltage to our motors.

- Hayear 34MP Microscope Camera: Dedicated camera that can be attached on top of the microscope lens for viewing the slide underneath.

- Arduino IDE: For writing the code for motor action and scanning procedure in C language.

- PyCharm IDE: For writing the codes for GUI, auto-focus, stitching, deep learning in Python.

- tkinter: Python library used in designing GUI.

- skimage: Python library used in design of our custom image-stitching algorithm.

- OpenCV: Python library used in the design of our auto-focus algorithm.

- Darknet: Open-source neural network framework, written in C and CUDA, used to implement YOLOv4 for cancer cell detection.

Novelty

Our implementation has some novelties as compared to the ones discusses above. The system we present is end-to-end with an emphasis on ease-of-use for the user. This is because our system is intended for use in a medical facility, where someone is not concerned with the internal working of our system. To achieve this purpose, we have designed a user-friendly Graphical User Interface (GUI) as a frontend for our system.

In addition to this, our system has auto-focus, machine learning and image stitching capabilities. Among the implementations done by others, which are discussed above, there is none which has all these three features. Thus, we take pride in providing all these features in a single system as it is a novel feature of ours.

Contribution

Originally, we divided our work as follows:

- Momin Ahmad and Uzair Khattak:
    1. Scanning System
    2. Stitching and Blending of Images
    3. Electronics and Circuit Design

- Talha Mehboob and Zeeshan Karamat:
    1. 3D Modelling and Structure Design
    2. Analysis of Images using Deep Learning
    3. Design of Graphical User Interface

However, all of us have worked on every aspect of the project and helped each other out. Also, owing to the special circumstances in wake of COVID-19, we had difficulty in access to the lab. Eventually we worked together at virtual meetings and going to the lab whenever possible.

Hardware Models

Acrylic Model

In order to implement low-cost whole slide image scanner. We decided to design our model using acrylic material. Using rectangular acrylic pieces, we designed our model and attached stepper motors that not only automated the scanning process in x and y axis but also allowed z axis movement. Whole process was automated fully using step motors. At this point we used mobile phone camera with a 20x lens to scan slide and to capture frames. Apparently, the model was running smoothly but when the result was seen on mobile phone screen there were significant vibrations which were not visible by naked eye, but they were significant at 20x. Even air breeze was enough to make it vibrate. These vibrations were due to flexibility of material and jerks produced by motors. Moreover, there was a problem related to the optics of our system. Magnification of microscope with tube length "L", and lens with focal length "f_o" is given by:

$$M = \frac{L}{f_o}$$

Optimal tube length was needed for required magnification which we could not find in this model.
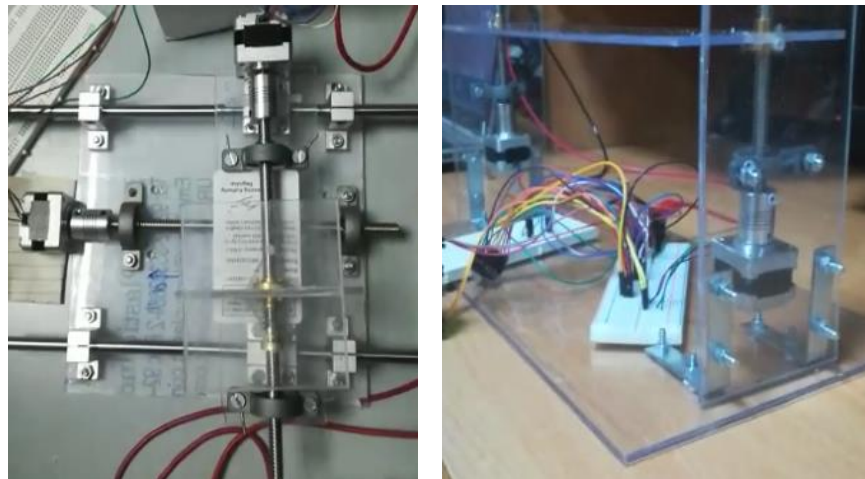
Figure 2: Acrylic Model for our Scanner

Wooden Model

In our second iteration we implemented our design using wood. Wood being more rigid was expected to control vibrations. Vibrations due to external sources were compensated. This model was better than the acrylic model but again there were micro-vibrations due to motors. When scanning tissues, we cannot ignore these micro vibrations. So, we decided to improve our model further by making use of more robust mechanical materials.

Figure 3: Wooden Structure

3D-Printed Model

In our third iteration we went for 3D-printed model. We designed our model in tinker-cad [33]. We printed the model parts by parts. All parts were printed using PLA (Polyacetic Acid) material. All parts were connected to implement our low-cost whole slide image scanner. This model was very efficient against vibrations. There were little to no vibrations due to flexible base. These vibrations have no effect on captured frames. In this model vibrations due to surrounding as well as motors were compensated. The problem with this model was that due to heating of motors the material started to melt. Also, the capture frames started to lose focus after some time of scanning. After research it was found out that due to 3D structure of tissue the focus is lost. Auto focus was designed to solve this problem. The implementation of auto focus required very controlled motion by motors. We tried to implement auto focus by printing a tube with threading so we can have controlled motion in z-axis. This idea was inspired from the working of microscope. After number of iterations when we could not achieve desired results. We decided to improve our model to get rid of the persisting problems.
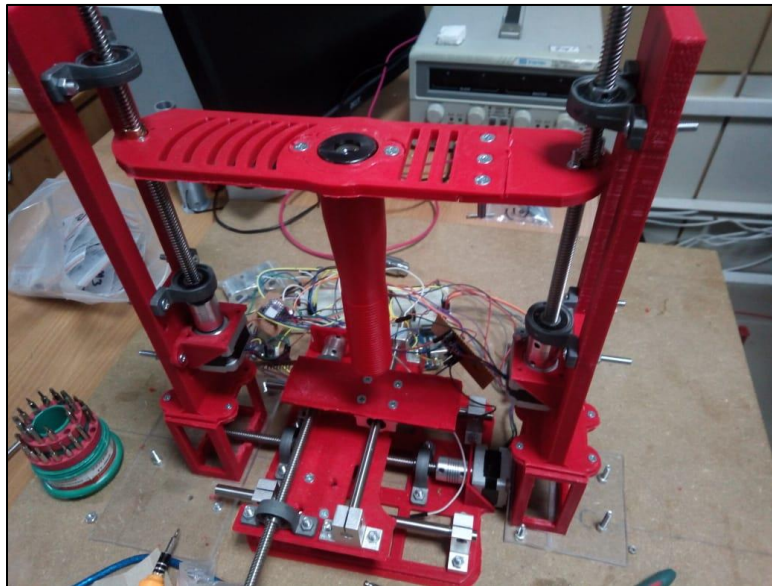
Figure 4: 3D-Printed Structure

19

Automation of Microscope

After the first 3 iterations with custom hardware design for WSI scanner, we observed that this kind of design requires most of the parts to be 3D printed and great levels of precision and accuracy are required in building the mechanical and optical systems to curb the vibrations at the microscopic level.

This motivated us to utilize the optical accuracy of existing machines such as compound microscope. So, we used the 3D-printed hardware to automate the motion of a compound microscope in all 3 dimensions, i.e., x-y plane for scanning and z plane for auto-focus module using stepper motors. For the circuitry part, we used Ramps 1.4, which is an Arduino Mega shield for the stepper motor driver DRV8825 which we are using.
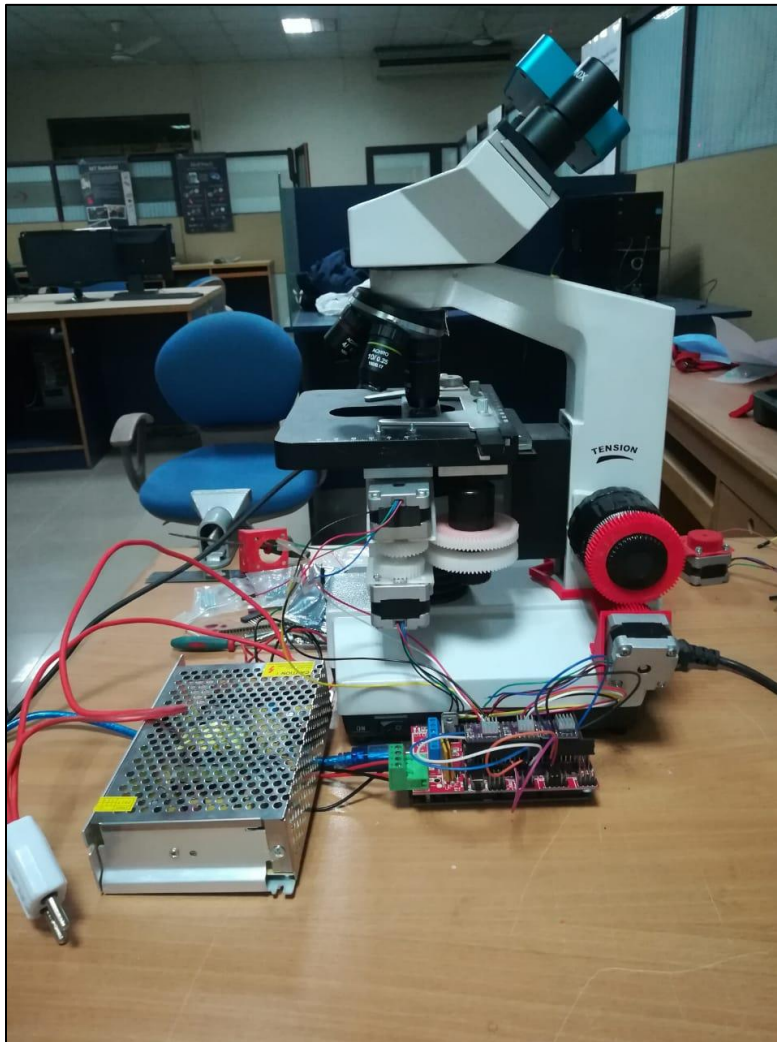
Figure 5: Microscope Automated with Motors and Microcontroller

To control our motors with pinpoint accuracy, we needed to calibrate them first. The ratio of distance moved on slide and steps taken by the motor depends on two factors:

- Stepping mode used for the motor. The stepping mode is configurable and set through dedicated pins on the motor driver. We use 1/32 micro-stepping mode where a step taken by the motor is the $1/32^{th}$ part of a normal full step. Therefore, the motor takes more steps for moving a certain distance in micro-step mode as compared to when it moves in normal full-step mode.
- Size of gears. To be specific, the radii of the two gears which attach to the motor and microscope platform, respectively. Smaller the gear, more precise will be our movement on the slide.
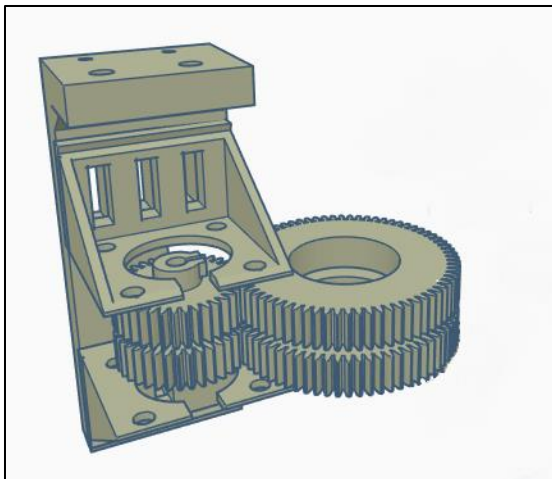
Figure 6: 3D-Printed Gears for Microscope
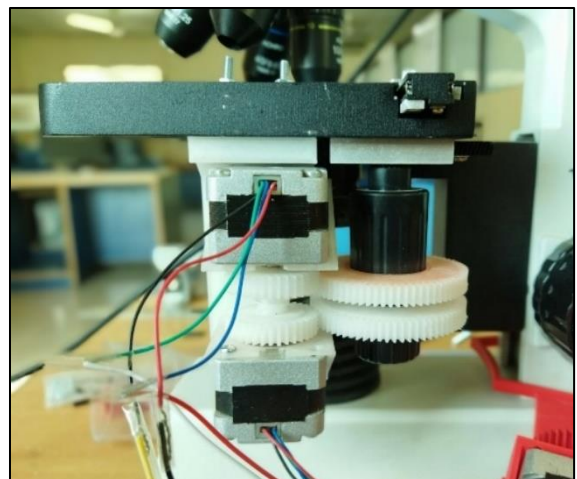
Figure 7: Gear Assembly Coupled with Microscope Knobs

Auto-Focus

The Microscope-based WSI scanner has a complete module of auto-focusing. There are two main reasons for which the auto-focusing module is implemented. As the physical slide is scanned with the help of high-resolution stepper motors in the x and y axes, the microscopic camera records the scan and frames are extracted from it which are to be stitched. During this process, sometimes the camera loses its focus due to the continuous motion of stepper motors and due to sudden change of visual contents in the physical slide (frequently white patches occurs in the slide). Secondly, the tissues in the physical slides that are being scanned are 3D in nature, they have a depth factor associated to them. As we scan along the X and Y axes without, we lose focus because the tissues are 3D in nature. So due to these reasons, we found it customary to design an auto-focusing module on the scanning system.

Selection of Best Metric

The objective was to achieve maximum focus for a given frame in the field of view. Higher the level of focus in a frame, higher will be its sharpness. This sharpness of the image is governed by the number of edges present in that frame. There are several operators that can detect edges e.g., Sobel filter, Canny edge detector, Laplacian etc. These operators return an output matrix (or filtered image) whose sum of elements is proportional to the number of edges in the frame.

Sun et al. [34] provide a comprehensive analysis of the auto-focus metrics and algorithms. We implemented some of the key auto-focus metrics from this above paper and used them on test images to select the best metric for our purpose. We compared these edge detectors by comparing their output when the platform is moving in the z-axis i.e., the frame is going in and out of focus.

For a more unbiased analysis, the metrics performances are again compared on same set of images with same amount of blur for all metrics. A sample biological digital image is used in our experiment from which 7 additional images are synthesized, with a higher level of blurriness in each successive image. These images are plotted below in gray-scale format (Figure 6).

After creating the set of images, each of the selected auto-focus metric was used on those pictures. Firstly, the images were converted into HSV format before using the metrics on them. At first, we were using the standard BGR format which is default in OpenCV, but BGR was not

giving us sharp differences between two images at different blur levels. We tried various color formats and found that the best format for our scenario is the HSV (Hue, Saturation, Value) format. This experiment was carried to find out which filter or detector shows significant change for a small amount of added blurriness in the picture, (or you can say small change in displacement in the Z-direction). Results are plotted for each selected autofocus metric (figures 7-11) and the proposed metric that was used is "Energy of Laplace". The Laplacian method works better is our case since it gives a significant change in output for a blur (or displacement in Z direction). Inbuilt OpenCV Functions were used for Sobel and Canny Filters. The rest of the methods were implemented according to the equations shown below:

$$f_{Norm\_Variance} = \frac{1}{H \cdot W \cdot \bar{\iota}} \sum_{Height} \sum_{Width} (i(x,y) - \bar{\iota})^2$$

$$f_{Brenner} = \sum_{Height} \sum_{Width} (i(x+2,y) - i(x,y))^2$$

Here, i(x,y) denotes a specific pixel at height and width, x and y units from the origin, respectively. $\bar{\iota}$ denotes the mean intensity of all the image pixels. H and W are the image height and width, respectively.

$$L = \begin{bmatrix} -1 & -4 & -1 \\ -4 & -20 & -4 \\ -1 & -4 & -1 \end{bmatrix}$$

$$f_{Energy\_Laplace} = \sum_{Height} \sum_{Width} C(x,y)^2$$

Here, the matrix "L" represents the convolution mask with which the image is convolved to obtain the second derivative "C".

Original   Blur1   Blur2

Blur3   Blur4   Blur5

Blur6   Blur7
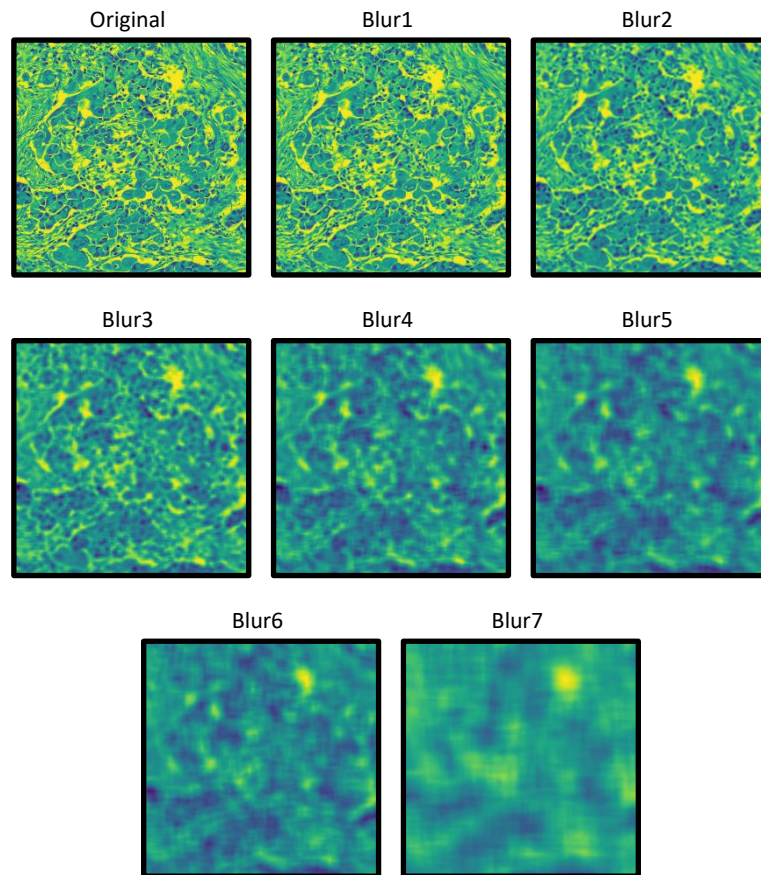
Figure 8 Different Levels of Blur in an Image from a Slide
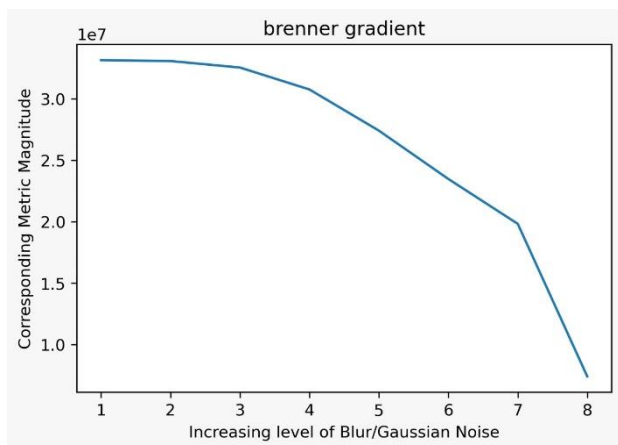
Figure 5: Brenner Gradient
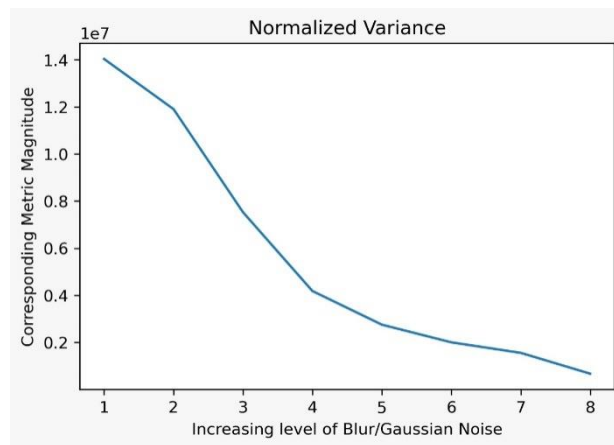
Figure 6: Normalized Variance
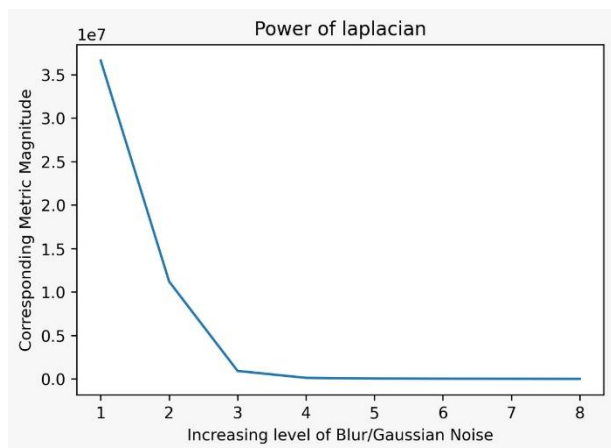
Figure 710: Laplacian Filter
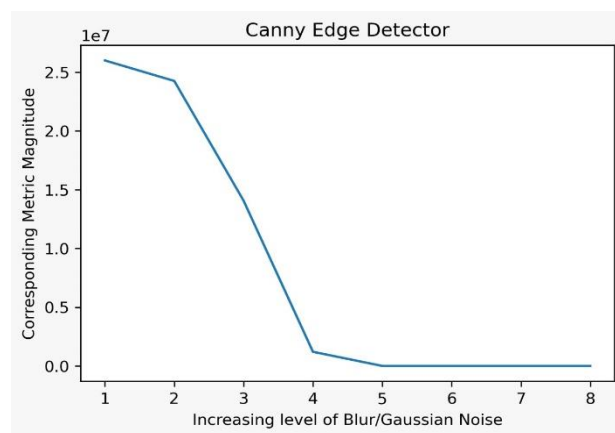
Figure 810: Canny Edge Detector
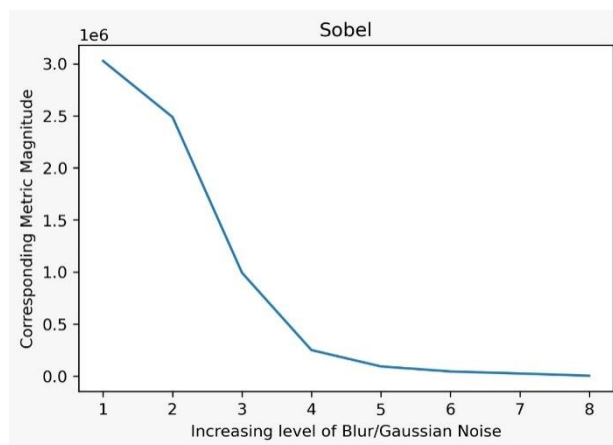
Figure 9: Sobel Filter

Algorithm for Autofocus

For a given frame, there is a unique optimal maximum, with a corresponding z-displacement, where the Laplacian filter sum outputs the maximum value, and this particular z-axis height refers to the full focused and sharp version of that frame.

However, there can exist local maxima somewhere at z-axis position, but it would not correspond to focused image. So, the task is to move the z-axis displacement into the global maximum.

The process of finding the global maximum is quite simple. It was concluded, by a series of experiments, that the global maximum, at a set magnification level, occurs somewhere in the z-displacement of lens greater than a few millimeters, measured with respect to the most bottom possible position of microscope lens. This position corresponds to the location very close to global maxima. At the start of auto-focus algorithm, the z-axis motors start from the bottom-most position and displaces itself a few millimeters upward, and thus it leaves behind the possible local maxima. An initial value of the Laplacian Filter sum is obtained. Then we continue moving in the same direction and again calculate the value of Laplacian Filter sum. Here we take the decision to continue moving in that direction if the sum value is larger than the previous one, or we reverse the direction if the sum value is less than the previous one. Basically, our main objective is to always move in the direction of increasing sum values since it corresponds to a higher level of focus. We perform this decision making at each frame, and after a few iterations we reach our desired point which is the global maxima of focus level.

At start, the stepper motor is configured to use full steps, to minimize the time for searching the maximum focus point, and when it surpasses the global maximum, the symmetric up and down motion by the motor is detected. After the detection of pattern which indicates that the displacement in a to-and-fro fashion around the maximum point, the stepper motor damping state is activated, in which it uses micro stepping to move the lens more precisely towards the global maximum location. Again, after some steps like this, damping pattern is detected but this time the lens has reached at a z-axis position which corresponds to the global maximum, and a sharp, fully focused image is retrieved at that point. This way, the auto-focus module works, and it is integrated during the scanning process in the system.
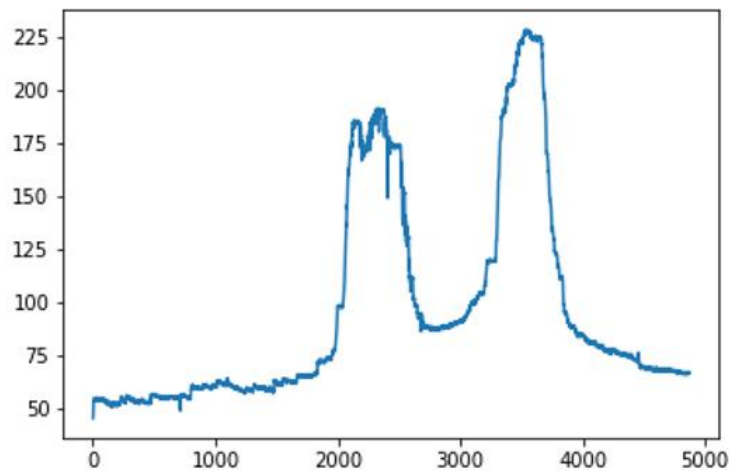
In the above graph, we can see the summary of the auto-focusing process. We initially start from a point which is far from the optimal focus point. Then we move in the direction of increasing focus. Since we are taking large steps, we may cross the local or global maximum. Taking large steps is essential in the beginning as it prevents us from stopping at a local maximum. Only when we are close to the global maximum for the focus level, we take micro-steps so that we do not miss the global maximum this time. This process is analogous to the gradient descent method used in machine learning. It is pertinent to mention that our approach was inspired by the same method.
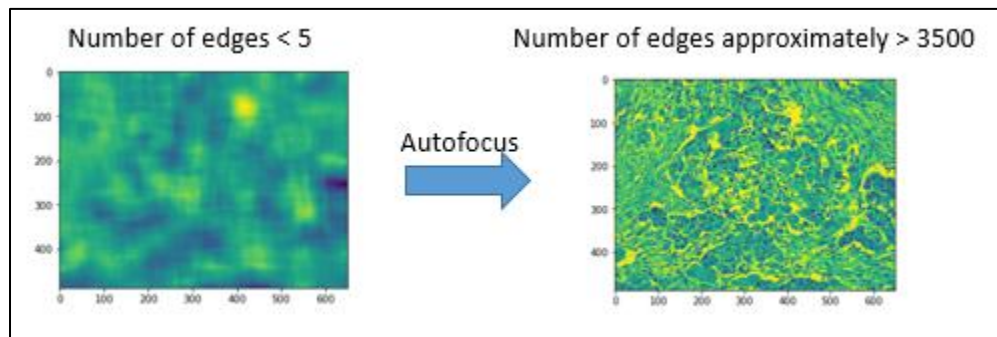
Figure 12: Our Auto-focus Algorithm Ensures High Quality Frames. Level of Focus is proportional to Number of Edges in Frame

Scanning System

The very first process of obtaining a digital Whole Slide Image is the scanning mechanism from which high-quality frames of a particular physical slide under examination are captured. The physical slide, which is to be scanned is placed under the microscope, and then using the Graphical User Interface (GUI), the user is allowed to perform several operations on that slide. When the scanning process is initiated, the slide is automatically displaced into a reset position (Origin) using the controlled stepper motors movement in the x and y direction. From the reset (or initial) position, the slide is further displaced until it reaches a predefined start point location from where the scanning along with image capturing will start. The position of the slide at this predefined start position is necessary because the camera is directly pointing at one corner of the slide at this start position, from where the scanning shall start and thus enabling the system to scan complete slide with no redundant displacements covered by the stepper motors.

From the start point location, the scanning of the physical slide is initiated in a rectangular fashion. The microscopic digital camera is activated, and it captures the high-quality frames as the slide moves under it. The slide is assumed to be divided into horizontal strips (along X direction). So initially, the first horizontal strip is scanned, and then the slide is displaced a little along the Y-axis, enabling scanning of the next horizontal strip. The slide is displaced along the y-axis in such a way that the next horizontal strip to be captured has a significant overlapping area with the previous horizontal strip. In a similar fashion, a complete slide is scanned, and all captured frames are stored to disk. In this scanning process, the auto-focus module is integrated in the system to keep the image frames as focused as possible. At the start of each horizontal strip, the auto-focus module runs first and the frame under the camera at that start point is focused to the maximum. During the action of auto-focus, the image capturing process is paused, so that the frames of that particular still image (at the start point of the horizontal strip) on which the auto-focus module is performing calibration are not stored on disk which are obviously redundant. This integration of auto-focus makes sure that the frame which is being captured does not become blurry due to disturbances in the movement of the slide (because of the stepper motors) and due to arbitrary white patches, that appears in the slide.

The proposed method is completely automatic and does not require any human intervention for its operation. Only the physical slide is placed manually under the microscope and the scanning process is initiated by selecting the "Scan Slide" operation in the GUI.

One important constraint that must be satisfied during the scanning process is that adequate overlapping (common area) between any two successive captured image frames should be observed. This condition must be satisfied since our proposed image stitching algorithm requires two successive frames to have a significant overlap in order to correctly identify their common region and then stitch them together. To satisfy this condition, the image capture rate and the speed of the stepper motors are calibrated such that the consecutive image frames have a 30% overlap between them.
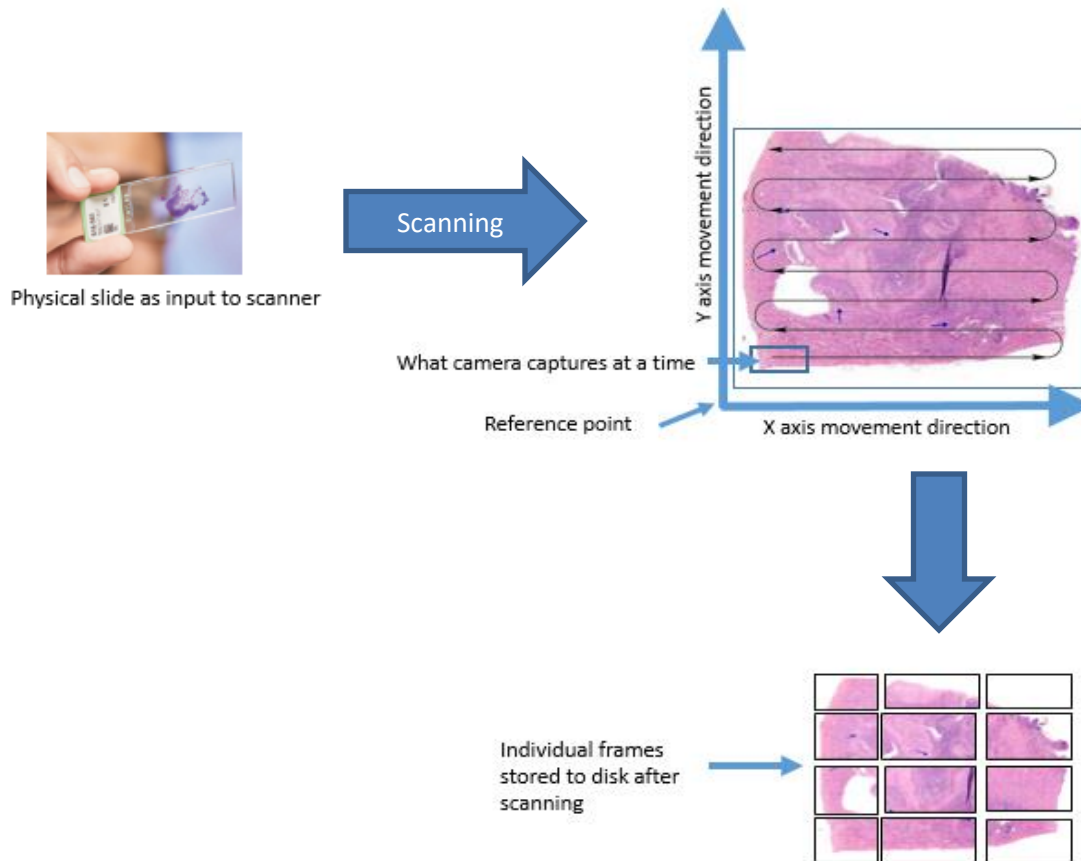
Figure 13: Scanning Pipeline

Stitching and Blending

Stitching

H igh quality frames from our proposed image acquisition pipeline are obtained and stored sequentially, which are then stitched together to form a large whole slide image. The images are planar, so instead of using advanced image registration methods based on SIFT and SURF, offset between two consecutive frames is calculated using cross-correlation method in order to align them. This technique consumes less time and enables real-time image stitching capability in the proposed system [35]. The images need to be in good quality in order to make this technique feasible, and the auto-focusing step ensures that the frames obtained are fully focused, thus meeting the requirement. An empty big 2D composite image array is created before initiating image stitching algorithm. This composite array would store the stitched images and eventually form a complete whole slide image. Two pointers/variables are assigned, last width pixel position and last height pixel position. These variables keep track of the width and height positions in the composite image where the incoming image frame will be pasted. Every time an image is pasted onto composite image, these variables update their values to new pixel location of height and width which corresponds to the height and width position where the upcoming next image frame will be pasted. Four additional cropping pointers are created which keeps tracks of the maximum and minimum pixel positions of height and width used during the entire stitching process. These pointers are used at the end of algorithm to crop-out only the portion of composite image where stitched whole slide image is present, leaving behind the unused portion of composite image.

Figure 14: Cross-Correlation Using SkImage [36]

Two frames in the sequence are cross-correlated and the relative shift values along the X and Y direction are obtained. The first image in the sequence is pasted completely onto the empty array.

This is only performed for single time at the start of algorithm. The shift values from cross-correlation give the exact location of the second frame relative to the first, and the second frame is stitched in the final image using its calculated location pixels. Last width pixel position and last height pixel position pointers are updated after a single frame is pasted onto empty image. Cropping pointers also updates their values if new maximum/minimum width and height locations are used while pasting the frame. This process is repeated for all the sequentially captured frames and the output is a single, large, stitched image. Cropping pointers crop-out only region of stitched image from empty image. For a 15000 x 15000 pixels final image, the process takes around 20 minutes on a normal computer. The resultant stitched image is then saved onto the computer.
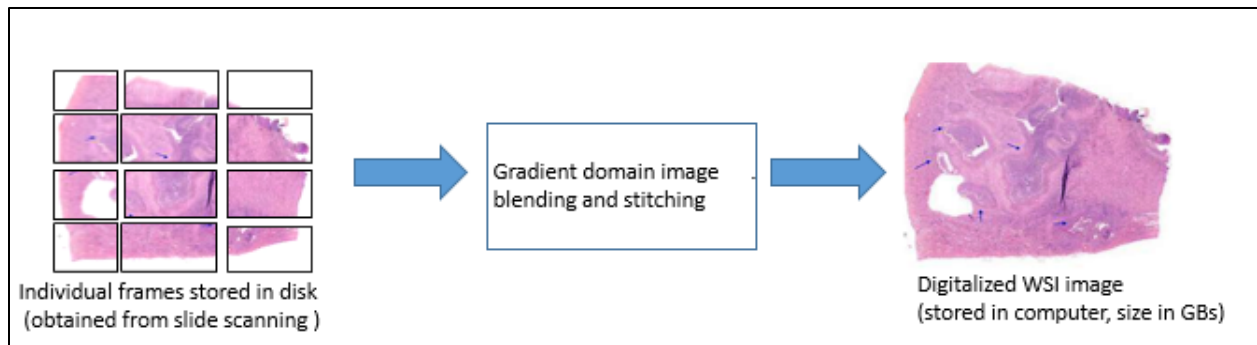
Figure 15: Stitching and Blending Pipeline

Blending

The final image obtained as a result of using the cross-correlation technique has seam lines running across it, because the frames vary slightly in illumination, intensity, and color. When all frames are put together, the seam lines are noticeable. To reduce the effects of seam lines and artifacts, different existing techniques of blending are used and compared.

The first technique uses averaging by assigning a constant weight to each pixel in the overlapping region of the two frames. This averaging acts like a low pass filter and it reduces the sharp edges in the frame transition region. This method is effective in reducing the seam lines and artifacts, but not completely because the transition from one frame to another is not completely smooth. This method is based on the following equation, where A and B are the two images to be stitched and have some common region. α is the constant weight assigned to the overlapping region of the first image and it has a value between 0 and 1. "I" is the resultant image after applying this method. "x" and "y" denote the specific pixel location.

$$I(x, y) = \alpha \cdot A(x, y) + (1 - \alpha) \cdot B(x, y)$$

To alleviate the problem of an abrupt transition at the edge, a weighted-averaging technique is implemented here [37]. It uses a weighted mask that gradually decreases the effect of one frame near the edge and its inverse mask gradually increases the effect of the other frame near the edge. This way, the two frames are stitched together smoothly, and the seam lines vanish. The result from this technique does have very minute seam lines, which can be observed upon very close inspection. This is because the transition between two frames is not completely smooth even in this technique.

To further improve the quality of the final stitched image, we use the technique of Laplacian blending. The first step is generating a Laplacian pyramid for the two frames to be stitched. We generate a 7-level Laplacian pyramid and then blending the frames in the gradient domain [38]. The result obtained after using Laplacian blending does not possess any seam lines, because there is a very smooth transition between the frames being stitched together.
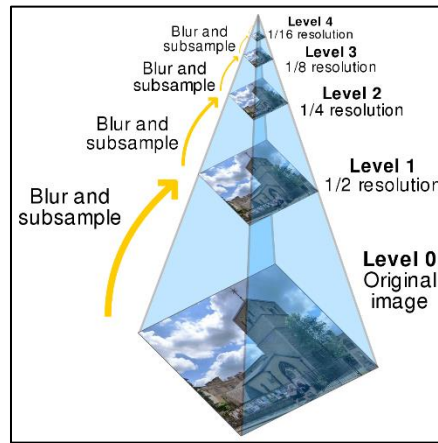
Figure 16: Breakdown of a Laplacian Pyramid [39]

In order to make sure that the frames being stitched have a significant overlap between them, we keep a check in our stitching algorithm that makes sure the overlapping height or width is above a specific threshold. It is observed that if two frames have very less overlapping region then we will have a stronger effect of seam lines in our final output image. This is because the Laplacian mask does not work well on small overlapping regions.

Results

Initially, we obtained stitched images using the Microsoft Image Composite Editor (ICE). We have previously discussed its merits and demerits. The following Whole Slide Image was obtained using ICE on one of our datasets:
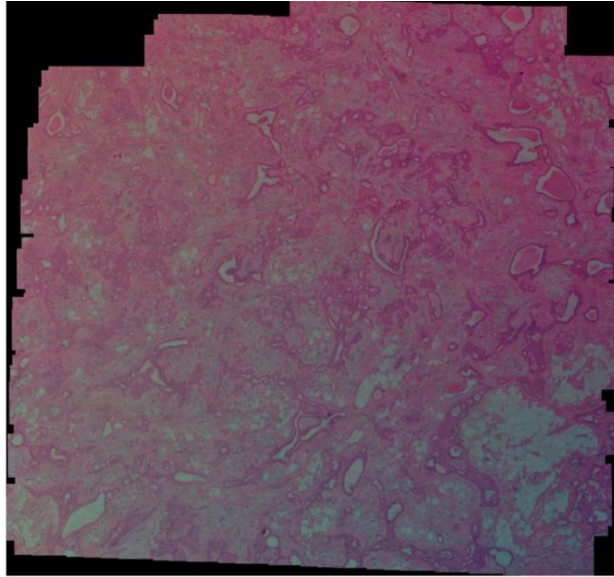
Figure 17: Stitched Image from ICE

The following Image was obtained after applying Linear Blending. The seam lines are evident in this result.
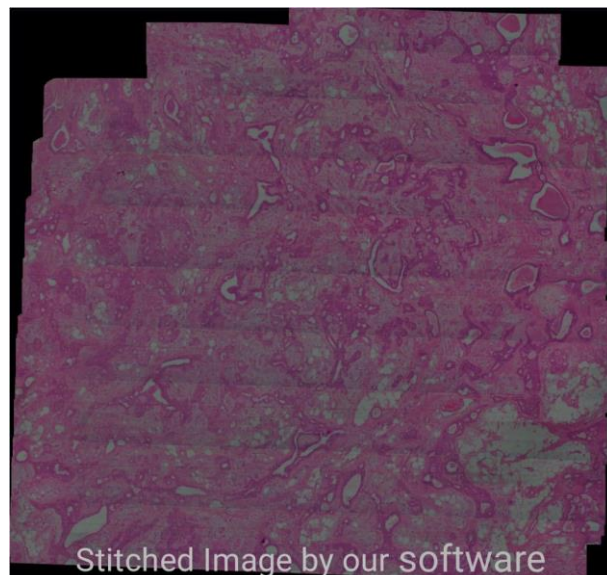
Figure 18: Stitched Image Using Linear Blending

After Linear Blending, we implemented stitching using Gaussian Filters. The results obtained were much better, but not perfect.
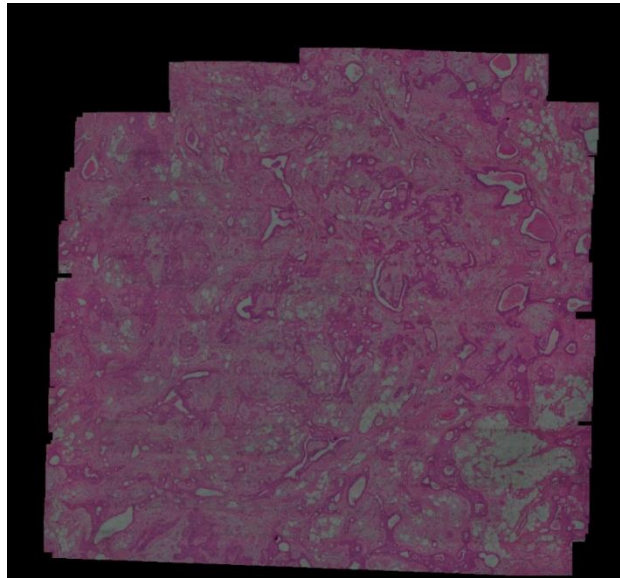
Figure 19: Stitched Image Using Gaussian Blending

Finally, we implemented Laplacian Blending, and as discussed in the previous section, it gave us perfect results as shown below:
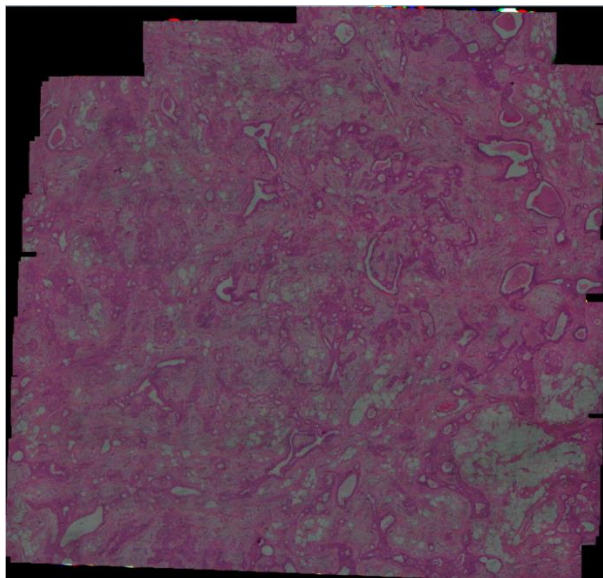
Figure 20: Stitched Image Using Laplacian Blending

GUI

The purpose of our system is to be user-friendly since it is to be used in medical environments, as discussed previously. Therefore, we needed a frontend for our system that acts as an abstraction layer over the underlying complexities. Also, since our project has several modules, such as autofocus, scanning, stitching, and object detection, there was a need to develop a centralized system whereby control could be issued from a single platform. These modules were developed on different platforms and connecting them with a single interface proved to be beneficial and efficient rather than operating them independently.

We chose the "tkinter" package of Python as the tool to develop our Graphical User Interface (GUI) as it is easy to use and offers flexibility in terms of features. We created the user interface such that there is a dedicated button for each function, such as a button for auto-focus and a button for initiating the scanning process. The buttons are arranged in an orderly fashion such that related buttons are grouped together.

Our GUI also has the feature of displaying the Field of View (FOV) of the slide in real-time. A window is part of our GUI, and it serves the purpose of displaying the camera view of the slide. This way, it is easier for us to control every operation from a single platform without resorting to using other software. Previously, we had to either use a camera-compatible software for viewing slides or look at the slide through the microscope lens. Both these methods were inefficient, and the use of GUI resolved these issues.
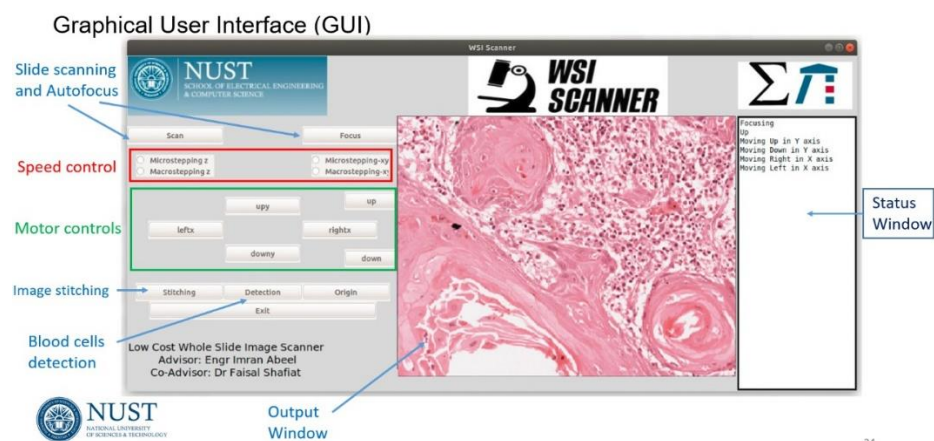
Figure 21: Graphical User Interface

Analysis of Images Using Deep Learning Techniques

Introduction

Digitization of slide data via Whole Slide Imaging (WSI) scanners has introduced significant efficiency into the work- flow of cancer pathology. Pathologists are now able to easily share digital slides and rapidly obtain second opinions. Unlike physical slides, digital slides do not degrade over time and thus are easy to archive. Efficiency can be enhanced further by employing AI based algorithms as second readers and tele pathology solutions have also become much more viable. However, WSI hardware is currently quite expensive and thus beyond the budgetary constraints of healthcare systems in developing countries. A dearth of qualified pathologists further aggravates this problem. Given these limitations, there is thus a need to develop low-cost slide digitization hardware for tele pathology in developing countries. Furthermore, the developed hardware needs to generate high-quality images so that state of the art AI and computer vision algorithms can also be leveraged to reduce pathologist workload. This work examines the viability of using a smartphone camera as the image acquisition hardware in automated pathology applications. More specifically, slides of 12 Oral Squamous Cell Carcinoma (OSCC) cases are first digitized via two different image acquisition devices: a WSI scanner and a smartphone placed directly at the eyepiece of a clinical
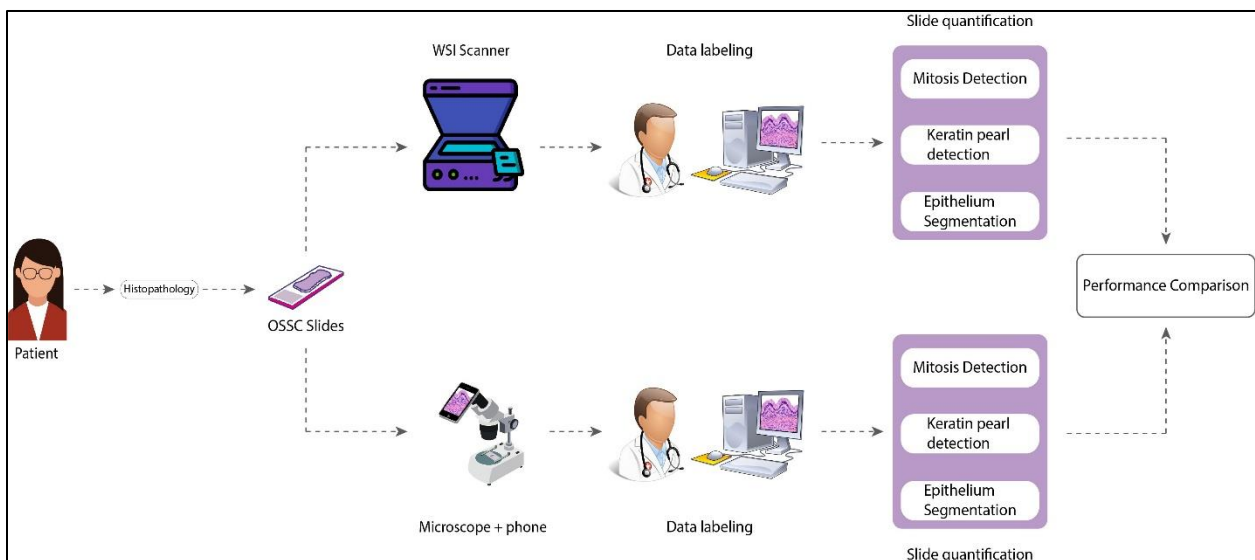
Figure 22: Comparison of WSI and smartphone acquired digital slide images. Identical deep learning models are trained on both image datasets for three tasks: Mitosis detection, Keratin Pearl detection and Epithelium Segmentation.

microscope. After labeling, both image datasets are used to train identical deep neural network models to detect three different features on the slide images. The performance of deep learning

algorithms on WSI images serves as a benchmark against which smartphone acquired images are evaluated. The whole process of training is repeated 10 times and is validated against a random test dataset every time. This tenfold training and validation is to ensure the statistical significance of the machine learning model we have trained. A high-level overview of this work is presented in Fig.23.

One of the limitations of smartphone-based image acquisition is that only a limited field of view can be captured at one time. Furthermore, magnification level also must be selected manually. It is possible, to overcome these limitations by constructing motorized attachments that can automatically move a slide platform under the microscope lens; acquire images of multiple fields of view and then stitch them together using simple image stitching algorithms to generate an image of the whole slide (just like panoramic imaging). Work has been carried out on this implementation as mentioned in the other half of this report however this section is to test whether smartphone acquired image quality is good enough to be used in computer vision applications? Validity of this hypothesis can be used to justify investment of substantial time and effort required for development of smartphone based WSI scanners. On the contrary, if the hypothesis is invalid then the results of this study can be used to examine the viability of (and effort required for) problem- specific improvements in computer vision algorithms that could be used for enhancing performance on smartphone acquired data.

Data Acquisition and Labelling

The dataset for work consisted of slides from 12 cases of OSCC. Our focus on oral cancer stems from the fact that it is one of the most prevalent forms of cancer among adult males in Pakistan [40]. This high rate of incidence of oral cancer is due to the usage of chewable tobacco, betel nut, smoking, and other oral intoxicants being frequently used in the region. The 5-year survival rate of oral cancer is about 40%. Oral Squamous Cell Carcinoma (OSCC) is the most common type of oral cancer accounting for 90% to 95% of the total cases [41].

Slide samples were first digitized by placing a smartphone (an iPhone 8) over the eyepiece of a microscope. The microscope used costs around 1,500 USD and similar models are employed for diagnosis at most pathology labs and clinics across Pakistan. These same slides were then

digitized using a Leica/Aperio CS2 scanner at the University of Sheffield. In our current setup, a smartphone can only capture one field of view (FOV) at one time therefore, we extracted snapshots of the WSI images to construct the WSI dataset. The magnification level and the size of the snapshot were set so that we were able to capture an area whose size was comparable to the FOV of the corresponding smartphone dataset. The statistics of the smartphone and the WSI datasets are summarized in Table 1 and Table 2 respectively. The number of FOVs for each feature in both datasets were kept similar.

Table 1: Summary of Smartphone Dataset

Feature Type	# FOVs	Magnification	Feature Count
Mitosis	165	40x	335
Keratin Pearl	88	20x	423
Epithelium	88	20x	–

Table 2: Summary of WSI Dataset

Feature Type	# FOVs	Magnification	Feature Count
Mitosis	166	40x	331
Keratin Pearl	86	20x	416
Epithelium	86	20x	–

For instance, the number of FOVs containing mitotic figures in the smartphone and the WSI datasets was 165 and 166 respectively; with the total count of mitotic figures being 335 in the smartphone dataset and 331 in the WSI dataset. A total of 423 and 416 instances of Keratin Pearls were observed in the smartphone and WSI datasets, respectively. Epithelial regions do not appear as distinct objects with well- defined boundaries, furthermore, they can traverse multiple FOVs. Consequently, their counts are not presented. The number of FOVs containing Epithelial regions were however, kept similar in both datasets.

Both datasets were labelled by a team of 3 experienced pathologists, including, a board-certified oral pathologist from the United States (A.K.) and a consultant specialist in oral and maxillofacial pathologist from the United Kingdom (S.A.K.). During the labeling process, one of the examiners (A.K.) diligently annotated the three features of interest. The other 2 examiners (XX and S.A.K.) examined the labels drawn by the first examiner and accepted or rejected them. An image was excluded from the dataset if at least 2 out of the 3 examiners did not agreed. The Institution's Ethical Review Board approved use of data for this study.
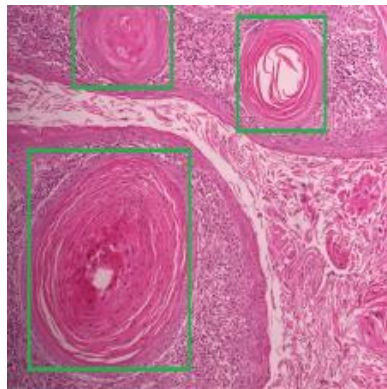
(a) Keratin Pearls labelled in slide image
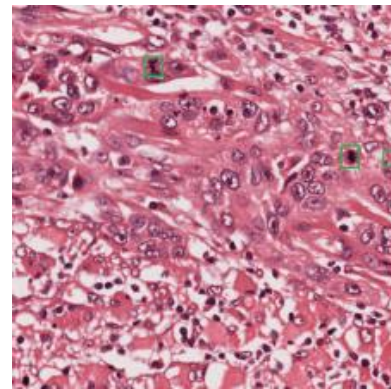
(b) Mitosis labelled in slide image

Figure 23: Example Images from Dataset

Experiments

The experiments for the keratin pearl and epithelium detection were carried out on Google Colab, online GPU services whereas the Mitosis detection were done on NVidia GTX 1060 GPU. To ensure a judicious comparison, identical deep learning pipelines were employed for quantification of each feature on both the smartphone and the WSI dataset. Moreover, we have trained each dataset tenfold and validated on multiple random images each time choses randomly from the parent images. The purpose of this step is to ensure the statistical significance of the results obtained from the microscopic images. The details of each approach are described in the subsequent sections.

Mitosis Detection

The appearance of Mitotic figures in slides is subject to significant variation in shape, size, color, contrast, etc. Furthermore, the size of mitotic figures is small, and they are generally visible at high power fields (HPF). These factors can make mitosis detection quite challenging making this a suitable task to gauge the amount of performance degradation that may be observed when employing low-cost image acquisition hardware.

Several different approaches were examined for mitosis detection. Single Shot Detector (SSD) [42] was tried first however, the performance of SSD was very poor due to the target size being too small.

To achieve maximum mitotic cells detection accuracy, YOLOv4 [43] (You Only Look Once) detector is used. Released in 2020, YOLOv4 is state of the art deep learning AI model that has both high accuracy and real-time compute compatibility, thus very suited for our problem.
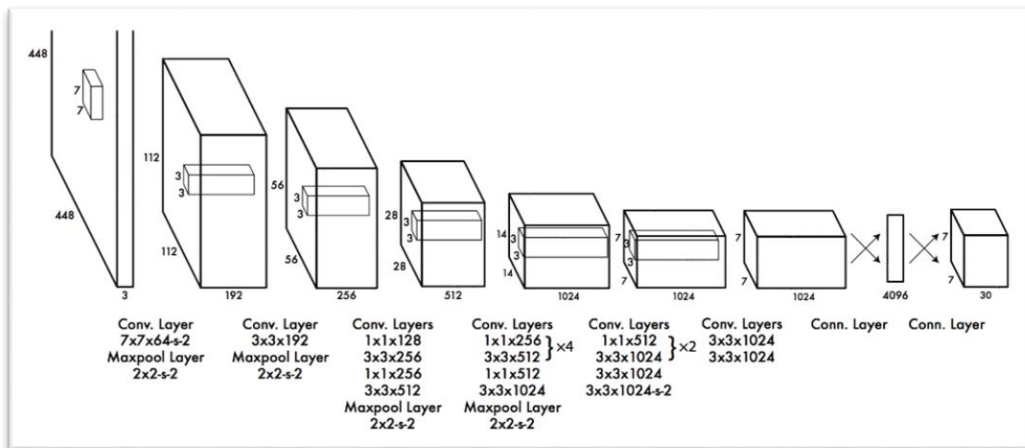
Figure 24: YOLOv4 Architecture

Written in Darknet framework, YOLOv4 utilizes full support of CUDA and CUDNN which enables it to be directly deployed on high end GPU.

Original smart phone images comprised of 165 images which were taken from a smart phone camera deployed on a low-cost medical microscope. Total 335 mitotic cells in this overall dataset are present.

For the WSI scanner dataset, high quality slides were scanned on the Aperio commercial WSI scanner and the dataset comprised of 109 images in total. Total 331 mitotic cells are present in this dataset.

Sophisticated image augmentation techniques were applied to increase the number of images including vertical and horizontal flipping, image rotation, random noise to images and hue increment in images. Python script was used for the data augmentation. This not only increased the overall dataset size, but also added generalization to dataset and the detector would be thus trained on very generic dataset.

Through data-augmentation, the final smart phone image numbers increased to 1155 high quality images with multiple image orientation. Similarly, WSI image dataset increased to 763 high quality images.

Both the smart phone and the commercial WSI scanner mitotic cell images were trained and tested on YOLO detector. Specific model configuration changes were performed in order to allow our dataset to train on the model. Batch size of 4 along with input image size of 416x416 was configured. Total epochs number for the experiment was set to 6000.

After configuring the YOLO model, both datasets were trained independently using 10-fold cross-validation rule. Thus, each training session consisted of 90% data in the train set with 10% data in the validation set respectively.

Ground Truth / Predictions
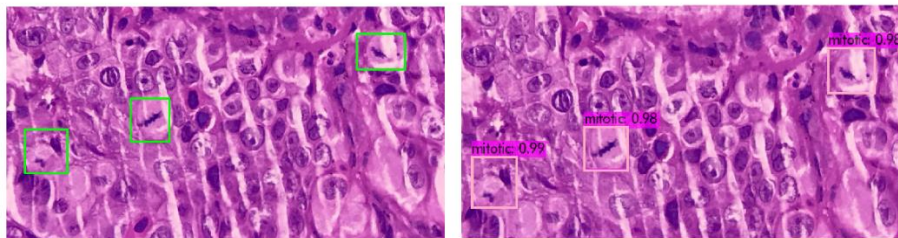
Figure 25: Ground Truth Vs. Prediction Images for Mitotic Figures for Microscopic data
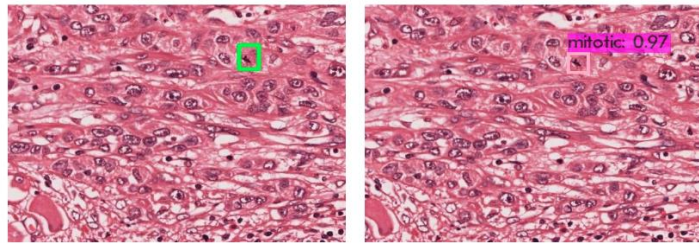
Figure 26: Ground Truth Vs. Prediction Images for Mitotic Figures for WSI data

Keratin Pearl Detection

As compared to mitotic figures, keratin pearls are larger in size and are generally detected at 10x magnification. Keratin pearls also exhibit large variations in size and structure. They can be found in epithelial as well as non-epithelial regions and the appearance can vary a lot. This makes the detection harder. So, in order to segment the Keratin pearl tissues in the images, we used U-net [44] architecture of deep neural networks since it has a proven track record of giving performance in medical imaging tasks with limited training data [45]. FastAI, which uses PyTorch as the back- end, was used to train and test the model on our data. We trained the U-net for 10 epochs first with learning rate 1e-3 and using cyclic learning rate [46] and with a weight decay factor of 1e-2 to avoid overfitting. Then, we trained for a further 25 epochs with learning rate 1e-4 using discriminative learning rates that FastAI provides along with cyclic learning rate. Batch size during the whole training process was 2 and we used ADAM optimizer. The backbone used in this case was resnet34 and the dataset was split into 75/15/10 for training, validation, and testing, respectively.

Ground Truth   Prediction over Ground Truth   Prediction
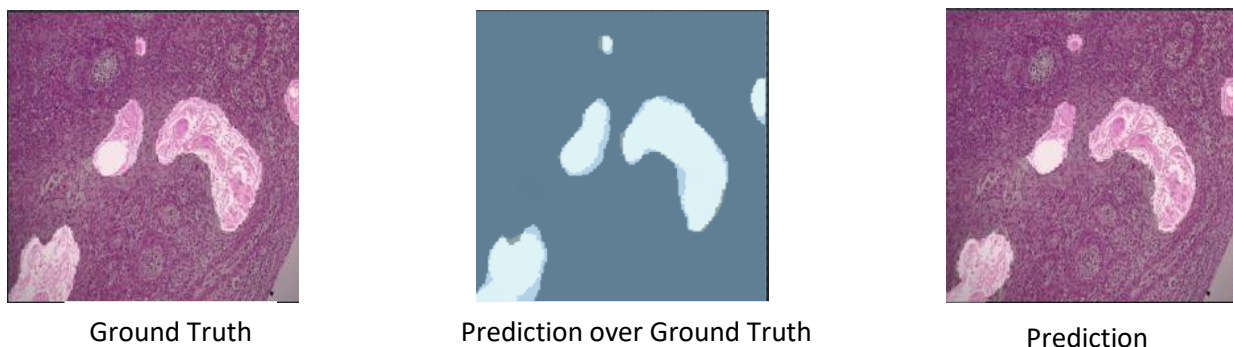
Figure 27: Ground Truth, Prediction over Ground Truth and Prediction Images for Keratin Pearl for microscopic data
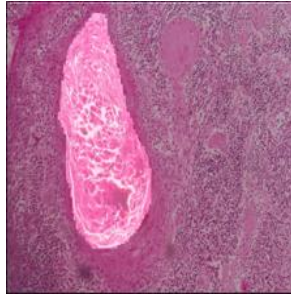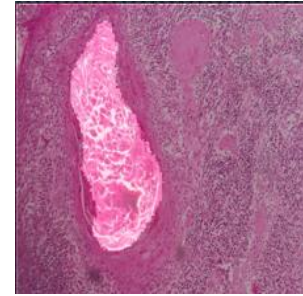
| Ground Truth | Prediction over Ground | Prediction |

Figure 28: Ground Truth, Prediction over Ground Truth and Prediction Images for Keratin Pearl for WSI data

Epithelium Tissues Detection

Epithelium tissues are more irregular in shape, and they are easier to allocate but harder to be completely identified due to its irregular shape. These tissues can also be found on the magnification of 10x, so in this regard they are much similar to the Keratin Pearls. So, based on this similarity, to segment the epithelium tissues in the images, we used U-Net architecture [44] as well. And since we know that the U-net model has a proven track record of giving performance in medical imaging tasks with limited training data [45]. Similar to the Keratin Pearl case, FastAI, which uses PyTorch as the back- end, was used to train and test the model on our data. We trained the U-net for 10 epochs first with learning rate 1e-3 and using cyclic learning rate [46] and with a weight decay factor of $1e^{-2}$ to avoid overfitting. Then, we trained for a further 25 epochs with learning rate 1e-4 using discriminative learning rates that FastAI provides along with cyclic learning rate. Batch size during the whole training process was 2 and we used ADAM optimizer. The backbone used in this case was resnet34 as well and the dataset was split into 75/15/10 for training, validation, and testing, respectively.
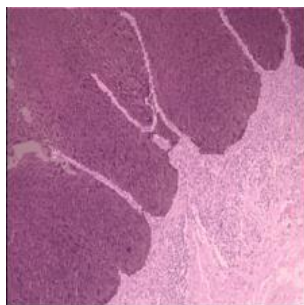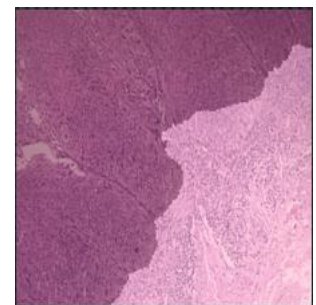
| Ground Truth | Prediction over Ground Truth | Prediction |

Figure 29: Ground Truth, Prediction over Ground Truth and Prediction Images for Epithelium tissues for microscopic data
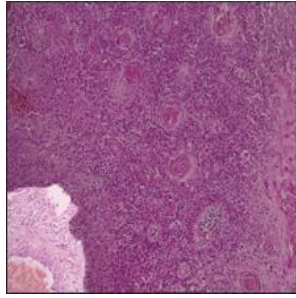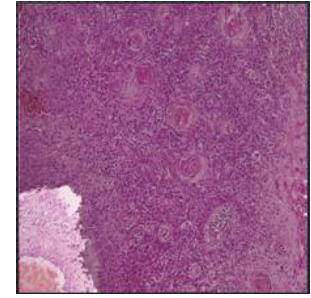
Ground Truth   Prediction over Ground Truth   Prediction

Figure 30: Ground Truth, Prediction over Ground Truth and Prediction Images for Epithelium tissues for WSI data
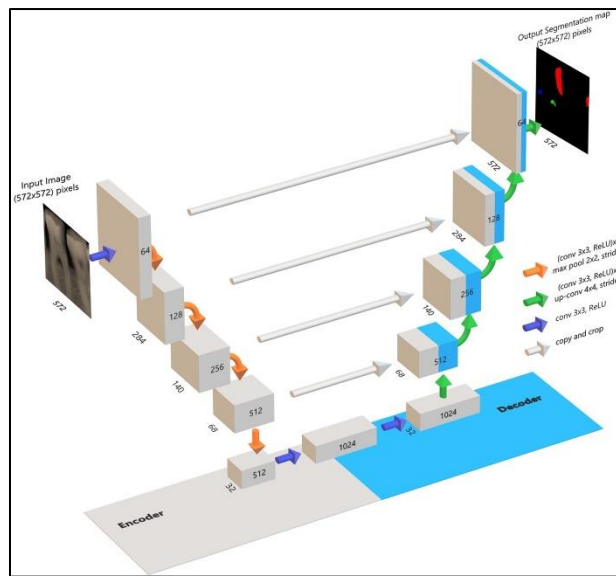
Figure 31: UNet Architecture Employed for Epithelium Segmentation

Results

Since the tasks are of two different natures i.e. object detection and segmentation, we have chosen two different evaluation metrics. For the sake of evaluating the segmentation results, we have chosen Sørensen–Dice coefficient and Intersection over Union IOU, and for the evaluation of object detection results, F1 score and mAP (mean average precision) is chosen. Also, the training is repeated 10 times and is validated against a randomly chosen test images every time. This tenfold training and validation is to ensure the statistical significance of the machine learning model we have trained. Below are given the performance metrics for each of the anomaly that we have detected.

Table 3 shows the results of each iteration of tenfold training that we have done on the two datasets for Keratin Pearl that are WSI, and microscopic datasets and Table 4 represents the mean average of the evaluating metrics.

The microscopic dataset performed better than the WSI dataset for Keratin Pearl detection in Dice Score. As, we can see that the microscopic dataset on mean Dice Score outperforms the WSI data by almost 0.7%, however the WSI datasets performs better than microscopic dataset when it comes to IOU score. Thus, we can say the results are very much comparable to each other which somehow defines the authenticity of our presented method of using the smartphone microscopic setup for the detection of anomalies in place of using the standard scanner.

Table 3: WSI Vs Microscopic Results for Keratin Pearl

KeratinPearl_WSI			KeratinPearl_microscopic		
Iteration No.	Dice Score	IOU	Iteration No.	Dice Score	IOU
1	0.6387	0.7404	1	0.68139	0.7412
2	0.6058	0.7241	2	0.4845	0.6572
3	0.62083	0.7254	4	0.67275	0.73797
4	0.6198	0.73319	7	0.60807	0.70814
5	0.64853	0.74139	6	0.71194	0.75626
6	0.72849	0.78448	5	0.7564	0.7964
7	0.61867	0.709	8	0.56608	0.6948
8	0.7056	0.7772	3	0.7351	0.7782
9	0.7136	0.7736	9	0.7355	0.78782
10	0.6484	0.7364	10	0.67184	0.72583
Mean	**0.654842**	**0.744516**	**Mean**	**0.662357**	**0.738382**

Table 4: Performance of Both Datasets for Keratin Pearl Detection

Dataset	Dice Score	IOU
WSI	0.6548	0.7445
Microscopic	0.6623	0.7383

Table 5 shows the results of each iteration of tenfold training that we have done on the two datasets for Epithelium that are WSI, and microscopic datasets and Table 6 represents the mean average of the evaluating metrics.

Epithelium segmentation use-case is almost a tie. Both the datasets perform very close to each other. This is possibly due to the robustness of the U-Net algorithm itself. Also, because we used cutting-edge techniques like cyclic learning rate and layer specific learning rates. Perhaps the network learns to filter out the noise better. For larger objects smartphones images produced lesser false positives. It is interesting to note that for a seemingly complex task of segmentation, the results are almost comparable. For the naked eye, a difference of 0.002 in dice coefficient does not make much difference. Thus, a network like U-net is easily able to overcome the noise present in smartphone images.

Table 5: WSI vs Microscopic Results for Epithelium Data

Epithelium_WSI			Epithelium_Microscopic		
Iteration No.	Dice Score	IOU	Iteration No.	Dice Score	IOU
1	0.743399	0.69997	1	0.70664	0.632926
2	0.629469	0.615141	2	0.63651	0.6104
3	0.7182053	0.636737	3	0.72183	0.664092
4	0.74621	0.66506	4	0.763648	0.696296
5	0.661219	0.666323	5	0.669136	0.588741
6	0.634951	0.618382	6	0.663101	0.659455
7	0.68919	0.655544	7	0.676487	0.5865176
8	0.76101	0.68076	8	0.74416	0.647035
9	0.737919	0.652218	9	0.68713	0.6606
10	0.774549	0.700168	10	0.8021	0.65727
Mean	**0.7096**	**0.6590**	**Mean**	**0.7070**	**0.6403**

Table 6: Performance of Both Datasets for Mitosis Detection

Dataset	Dice Score	IOU
WSI	0.7096	0.6590
Microscopic	0.7070	0.6403

Our objective was not to create the best object detector or segmentation network. The aim was to compare how pathological images from a smartphone fare against images from an expensive WSI scanner. To prove the significance of our results, we performed the statistical significance test for both Keratin Pearl and Epithelium result iteration data. And, to prove our hypothesis, the results prove to be significant with P score of $0.24 > 0.05$ and $0.09 > 0.05$ for Keratin Pearl and Epithelium respectively, meaning the samples are likely to be drawn from the same distributions proving the significance of the data samples obtained from two above discussed methods.

Table 7 shows the results of each iteration of tenfold training that we have done on the two datasets for Mitotic Figures that are WSI, and microscopic datasets and Table 8 represents the mean average of the evaluating metrics.

The results obtained from the 10-fold cross-validation training process for the mitotic datasets are very much promising. We can see that the mean average precision of the smart phone dataset training is far better than the mitotic WSI dataset training results i.e., 92.29% mAP for smartphone dataset vs 88.33% mAP for the WSI dataset. However, the F1 score of the WSI dataset is greater than the mean F1 score of smart phone dataset. This shows that both datasets can be considered fairly decent when dealing with the mitotic cells. As already evident from the images, we can expect such results because the image quality of smart phone dataset is very high, and all the mitotic figures were strongly visible during annotation process as well.

YOLOv4 model performed very robustly in both datasets and data augmentation techniques also played an important role. Mean average precision and F1-score are suitable metrics for determining model's ability to perform successful object detection.

So, these experiments prove that the image analysis for mitotic cells detection can be securely used in the mitotic cells dataset obtained from smartphone based low cost WSI image acquisition system.

Table 7: WSI vs Microscopic Results for Mitotic Figures

Mitotic_Detection_WSI			Mitotic_Detection_Microscopic		
Iteration No.	F1 Score	mAP	Iteration No.	F1 Score	mAP
1	0.85	82.40%	1	0.94	96.70%
2	0.89	88.18%	2	0.9	94.14%
3	0.87	89.51%	3	0.88	92.85%
4	0.9	92.95%	4	0.84	89.36%
5	0.91	89.74%	5	0.88	91.16%
6	0.88	90.12%	6	0.88	91.15%
7	0.83	82.76%	7	0.86	89.96%
8	0.89	91.87%	8	0.87	93.79%
9	0.88	89.12%	9	0.88	89.71%
10	0.85	86.64%	10	0.92	94.12%
Mean	**0.875**	**88.33%**	**Mean**	**0.885**	**92.29%**

Table 8: Performance of Both datasets for Epithelium Segmentation

Dataset	F1 Score	mAP
WSI	0.875	88.33%
Microscopic	0.885	92.29%

Conclusion and Future Work

In conclusion, we have seen that our results are comparable to those obtained using commercial WSI Scanners. The biggest advantage of our approach is that it is low in cost. We have shown our final stitched images to professional pathologists, and they have expressed positive views over it, stating that not only the images are focused and clear, but also annotatable for Deep Learning purposes. Thus, our WSI system is up to the standards to be deployed for pathological purposes.

Potential future work involves integrating our system with a cloud system to provide remote diagnostic services. We can add a multi-slide scanning feature, like in commercially available WSI Scanners. Currently our scanner is connected and controlled using a laptop. To enclose the system in one package, we can deploy the system on a Linux-based system that supports deep learning. A strong contender for this functionality is the NVIDIA Jetson Nano.

Automatic mitotic detection is critical for cancer grading. Our results demonstrate that deep learning-based detection of mitotic cells does not degrade significantly when using smartphone-based images. Also, our microscopic data has similar performance to the WSI data and this has been proved by statistical significance tests which failed to prove the null hypothesis. This is very encouraging since it opens the doors for smartphone-based tele-pathology in developing countries where most public sector laboratories are underfunded and cannot afford expensive WSI scanners. We believe that larger smartphone-based datasets can further improve performance. In addition to this, explicit modeling of noise inducing factors such as blurring, intensity variations and non-uniform variations can also deliver further performance gains. As part of our future work, we plan to examine performance on more quantification tasks on larger size datasets. We also plan to incorporate the impact of factors such as noise etc.

Digital Pathology holds great potential for the times to come. Recent advancement in machine learning and deep learning techniques indicate a promising future for the field of digital pathology. Also, we believe that the low-cost implement of the WSI Scanner will be a game-changer for developing countries such as Pakistan.

References

[1] Cardesa, Antonio, N. Zidar, L. Alos, A. Nadal, N. Gale and a. G. Klöppel, The Kaiser's cancer revisited: was Virchow totally wrong?, Virchows Archiv 458, no. 6, 2011.

[2] J. D. Webster and a. R. W. Dunstan, Whole-slide imaging and automated image analysis: considerations and opportunities in the practice of pathology, Veterinary pathology 51, no. 1, 2014.

[3] Al-Janabi, Shaimaa, A. Huisman and a. P. J. V. Diest, Digital pathology: current status and future perspectives, Histopathology 61, no. 1, 2012.

[4] Ghaznavi, Farzad, A. Evans, A. Madabhushi and a. M. Feldman, Digital imaging in pathology: whole-slide imaging and beyond, Annual Review of Pathology: Mechanisms of Disease 8, 2013.

[5] R. S. Weinstein, A. R. Graham, L. C. Richter, G. P. Barker, E. A. Krupinski, A. M. Lopez, K. A. Erps, A. K. Bhattacharyya, Y. Yagi and J. R. Gilbertson, Overview of telepathology, virtual microscopy, and whole slide imaging: prospects for the future, Human pathology 40, no. 8, 2009.

[6] Holmström, Oscar, N. Linder, H. Moilanen, A. Suutala, S. Nordling, A. Ståhls, M. Lundin, V. Diwan and J. Lundin, Detection of breast cancer lymph node metastases in frozen sections with a point-of-care low-cost microscope scanner, PloS one 14, no. 3, 2019.

[7] Hitchcock and C. L., The future of telepathology for the developing world, Archives of pathology \& laboratory medicine 135, no. 2, 2011.

[8] Farahani, Navid, Pantanowitz and a. Liron, Overview of telepathology, Surgical Pathology Clinics 8, no. 2, 2015.

[9] Q. Lu, G. Liu, C. Xiao, C. Hu, S. Zhang, R. X. Xu, K. Chu, Q. Xu and Z. J. Smith, A modular, open-source, slide-scanning microscope for diagnostic applications in resource-constrained settings, PloS one 13, no. 3, 2018.

[10] Sharma, Shikha, J. Zapatero-Rodríguez, P. Estrela and R. O'Kennedy, Point-of-care diagnostics in low resource settings: present status and future role of microfluidics, Biosensors 5, no. 3, 2015.

[11] Chu, Kaiqin, Z. J. Smith and a. S. Wachsmann-Hogiu, Development of inexpensive blood imaging systems: where are we now?, Expert review of medical devices 12, no. 5, 2015.

[12] Markiewicz, Katarzyna, J. A. V. Til and M. J. IJzerman, Medical devices early assessment methods: systematic literature review, Int J Technol Assess Health Care 30, no. 2, 2014.

[13] Miller, A. R., G. L. Davis, Z. M. Oden, M. R. Razavi, A. Fateh, M. Ghazanfari and F. Abdolrahimi, Portable, battery-operated, low-cost, bright field and fluorescence microscope, PloS one 5, no. 8, 2010.

[14] Campbell, R. AA, R. W. Eifert and a. G. C. Turner, Openstage: a low-cost motorized microscope stage with sub-micron positioning accuracy, PloS one 9, no. 2, 2014.

[15] Lu, Yujie, Y. Liu and T. K. Lau, Simple, portable, and low-cost microscope based on off-axis digital holography using two spherical waves, Optics letters 39, no. 15, no. 2, 2014.

[16] Ephraim, R. KD, E. Duah, J. S. Cybulski, M. Prakash, M. V. D'Ambrosio, D. A. Fletcher, J. Keiser, J. R. A. Bogoch and a. Isaac, Diagnosis of Schistosoma haematobium infection with a mobile phone-mounted Foldscope and a reversed-lens CellScope in Ghana, The American journal of tropical medicine and hygiene 92, no. 6, 2015.

[17] Bogoch, I. I., J. R. Andrews, B. Speich, J. Utzinger, S. M. Ame and S. M. A. a. J. Keiser, Mobile phone microscopy for the diagnosis of soil-transmitted helminth infections: a proof-of-concept study, The American journal of tropical medicine and hygiene 88, no. 4, 2013.

[18] Holmström, Oscar, N. Linder, B. Ngasala, A. Mårtensson, E. Linder, M. Lundin, H. Moilanen, A. Suutala, V. Diwan and J. Lundin, Point-of-care mobile digital microscopy and deep learning for the detection of soil-transmitted helminths and Schistosoma haematobium, Global health action 10, no. sup3, 2017.

[19] Holmström, Oscar, N. Linder, M. Lundin, H. Moilanen, A. Suutala, R. Turkki, H. Joensuu, J. Isola, V. Diwan and J. Lundin, Quantification of estrogen receptor-alpha expression in human breast carcinomas with a miniaturized, low-cost digital microscope: a comparison with a high-end whole slide-scanner, PLoS One 10, no. 12, 2015.

[20] K. Guo, J. Liao, Z. Bian, X. Heng and G. Zheng, InstantScope: a low-cost whole slide imaging system with instant focal plane detection, Biomedical Optics Express 6, no. 9, 2015.

[21] Li, Jiayuan, Q. Hu and M. Ai, Optimal illumination and color consistency for optical remote-sensing image mosaicking, IEEE Geoscience and Remote Sensing Letters 14, no. 11, 2017.

[22] Gross and J. W., A comparison of orthomosaic software for use with ultra high resolution imagery of a wetland environment, Center for Geographic Information Science and Geography Department, 2015.

[23] Delahunt, C. B., C. Mehanian, L. Hu, S. K. McGuire, C. R. Champlin, M. P. Horning and B. K. W. a. C. M. Thompon, Automated microscopy and machine learning for expert-level malaria field diagnosis, IEEE Global Humanitarian Technology Conference (GHTC), pp. 393-399, 2015.

[24] Y. Yu and H. Peng, Automated high speed stitching of large 3D microscopic images, IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 238-241, 2011.

[25] U. Qidwai and M. Akbar, Image stitching system with scanning microscopy for histopathological applications, IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES) (pp. 522-527), 2016.

[26] M. A. T. Khan, W. Nawaz, H. Rashid, A. Kiyani, S. A. Khurram and H. Aqeel, "Cancer Image Quantification With And Without, Expensive Whole Slide Imaging Scanners," *2019 41st Annual International Conference of the {IEEE} Engineering in Medicine and Biology Society ({EMBC}),* 2019.

[27] L. Roux, D. Racoceanu, N. Lomenie, M. Kulikova, H. Irshad, J. Klossa, F. Capron, C. Genestie, G. L. Naour and M. N. Gurcan, "Mitosis detection in breast cancer histological images An ICPR 2012 contest," *Journal of pathology informatics,* vol. 4, 2013.

[28] Cucor, D. J. Hartman, B. J. Kolowitz, G. R. Lauro, J. S. McHugh, V. Palat, L. Pantanowitz, A. Parwani, A. Reden, S. Sloka, S. A. Yousem, I. Ahmed and B. C. a. I. C., "Pocket pathologist: A mobile application for rapid diagnostic surgical pathology consultation," *Journal of Pathology Informatics,* vol. 5, no. 1, 2014.

[29] Rapini, R. R. Jahan-Tigh, G. M. Chinn and R. P., "A Comparative Study Between Smartphone-Based Microscopy and Conventional Light Microscopy in 1021 Dermatopathology Specimens," *Archives of Pathology Laboratory Medicine,* vol. 140, no. 1, pp. 86-90, 2016.

[30] P. F. Yagi, F. Liu and Yukako, "Evaluation of a smartphone for telepathology: Lessons learned," *Journal of Pathology Informatics,* vol. 6, no. 1, 2015.

[31] [Online]. Available: http://www.cancerindex.org/.

[32] [Online]. Available: https://www.medicalexpo.com/medical-manufacturer/microscope-slide-scanner-31319.html.

[33] [Online]. Available: https://www.tinkercad.com/.

[34] Sun, Yu, S. Duthaler and B. J. Nelson, Autofocusing in computer microscopy: selecting the optimal focus algorithm, Microscopy research and technique 65, no. 3, 2004.

[35] Guizar-Sicairos, Manuel, S. T. Thurman and J. R. Fienup, Efficient subpixel image registration algorithms, Optics letters 33, no. 2, 2008.

[36] [Online]. Available: https://scikit-image.org/docs/0.13.x/auto_examples/transform/plot_register_translation.html.

[37] P. Thevenaz and M. Unser, User-friendly semiautomated assembly of accurate image mosaics in microscopy, Microscopy research and technique 70, no. 2, 2007.

[38] A. Zomet, A. Levin, S. Peleg and Y. Weiss, Seamless image stitching by minimizing false edges, IEEE transactions on image processing 15, no. 4, 2006.

[39] [Online]. Available: https://en.wikipedia.org/wiki/Pyramid_(image_processing).

[40] Badar, S. Mahmood, R. Faraz, A. Yousaf, A. u. Quader, HinaAsif, A. Atif, L. Nadeem, N. Parveen, R. Tanveer and A. H. a. Farhana, "COLLECTIVE CANCER REGISTRY REPORT FROM DECEMBER 1994 TILL DECEMBER 2018, OF THE SHAUKAT KHANUM MEMORIAL CANCER HOSPITAL & RESEARCH CENTER, PAKISTAN," *Shaukat Khanum Memorial Cancer Hospital Research Center,* 2018.

[41] Goiato, D. C. C. Franca, L. M. Monti, A. L. d. Castro, A. M. P. Soubhia, L. E. R. Volpato and S. d. A. a. M. C., "Unusual Presentation of Oral Squamous Cell Carcinoma in a Young Woman," *Sultan Qaboos University Medical Journal,* vol. 12, no. 2, 2012.

[42] Berg, W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed and C.-Y. F. a. A. C., "{SSD}: Single Shot {MultiBox} Detector," in *Computer Vision {ECCV}*, Springer International Publishing, 2016, pp. 21-37.

[43] A. Bochkovskiy, C.-Y. Wang and H.-Y. M. Liao, YOLOv4: Optimal Speed and Accuracy of Object Detection, arXiv, 2020.

[44] Ronneberger, Olaf, Fischer, P. a. Brox and Thomas, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234-241.

[45] H. A. Khan, M. A. Haider, H. A. Ansari, H. Ishaq, A. Kiyani, Sohail, Kanwal, Muhammad, Muhammad and S. A. Khurram, "Automated feature detection in dental periapical radiographs by using deep learning," *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology,* 2020.

[46] L. N. Smith, "Cyclical Learning Rates for Training Neural Networks," in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2017.