Visual quality of histopathological images impact on lymphocytes detection using state-of-the-art algorithms

A. Polejowska, M. Sobotka, M. Kalinowski, M. Kordowski, T. Neumann

Department of Biomedical Engineering
Faculty of Electronics, Telecommunications and Informatics
Gdańsk University of Technology
Gdańsk, Poland

Abstract—**Lymphocytes, one of the leukocytes, i.e. white blood cells play an active role in the immune system. In diagnosing and treating diseases, histopathological images provide key information. The number of lymphocytes in the image, their distribution and type may indicate the presence of a specific lesion. The detection and quantification of lymphocytes not only supports the work of histopathologists, but also gives a chance to monitor diseases or analyze the general condition of the immune system. In this study, the impact of visual quality on the performance of state-of-the-art algorithms for detecting lymphocytes in histopathological images was examined. Two datasets were utilized, and image modifications such as blur, noise, sharpness, saturation, brightness and contrast were applied to evaluate the performance of YOLOv5 and Detectron2 deep learning models. It was found that visual quality has a significant impact on the performance of these algorithms, and that high-quality images are necessary for accurate detection of lymphocytes. These findings have significant implications for the use of computational methods in digital pathology. The need for careful attention to image quality in order to achieve reliable results is highlighted.**

Index Terms—**histopathological images, digital pathology, lymphocyte detection, deep learning, state-of-the-art algorithms, YOLOv5, Detectron2, image quality, blur, noise, sharpness, saturation, brightness, contrast**

I. INTRODUCTION

Histopathological images, which are images of tissue samples taken from a patient's body, are an invaluable tool in the diagnosis and treatment of diseases. Accurate detection of lymphocytes, which are a type of immune cell, in these images is critical for the diagnosis of certain conditions [11]. However, the visual quality of histopathological images can greatly affect the accuracy of algorithms in detecting lymphocytes.

In this research, the impact of visual quality on the performance of state-of-the-art algorithms on detecting lymphocytes in histopathological images was investigated. Two datasets were employed and image modifications, including blur, noise, sharpness, saturation, brightness and contrast, were systematically applied to evaluate the performance of YOLOv5 [13] and Detectron2 [25] deep learning models. The effect of visual quality on the accuracy of these algorithms and any potential issues that may arise when using them to analyze histopathological images were determined.

Understanding the effect of visual quality on the performance of algorithms for detecting lymphocytes in histopathological images is crucial for advancing the application of computational methods in digital pathology. The findings of this study have significant implications for the implementation of computational methods in pathology and underscore the importance of ensuring image quality for achieving reliable results.

II. RELATED WORK

In the field of medical image analysis, the impact of image quality on the performance of deep learning algorithms has received increasing attention. While extensive research has been conducted on natural images, there is still limited knowledge of the effect of image quality on medical images, particularly within the domain of digital pathology and the crucial task of detecting lymphocytes.

In recent years, various studies have examined the effect of image quality on deep neural networks in the natural image domain. One study that examined the effect of image quality on deep neural networks was published by Samuel Dodge and Lina Karam [4]. The authors investigated how blur, noise, and compression affect the performance of image classification tasks. They found that blur and noise had a significantly negative impact on networks' performance, while the effect of contrast and compression was less significant.

Other researchers have focused on improving the performance of deep learning algorithms using various image modification techniques. The literature has demonstrated the criticality of properly preparing histopathological images prior to utilizing neural network models, as the quality of the input data significantly impacts the performance of the model [22, 17]. The work by Guerrero and Oliveira [8] proposed a preprocessing stage with stain normalization to enhance the accuracy of deep learning models in the lymphocytes detection task. The proposed method was found to be one of the best-ranked results in the state of the art. This research emphasizes how preprocessing histopathological images can boost the performance of deep learning models.

In the field of digital pathology, researchers have also focused on developing specific deep learning models for lymphocyte detection. One such study (YOLLO) [20], proposed a

model tailored for detection of lymphocytes in histopathology whole slide images (WSIs) stained with immunohistochemistry (IHC). The authors made modifications to the original YOLO architecture, including simplifying the architecture and implementing a guided sampling strategy. As a result, detection and processing times were sped up. This research highlights the potential of using object detection methods for lymphocyte detection in histopathology. The study also emphasizes the importance of considering the specific characteristics of the image and the task at hand when designing deep learning algorithms.

III. EXPERIMENTAL SETUP

In the following, the experimental setup utilized to conduct experiments on the selected model architectures is described. Details regarding the dataset and the data annotation that was carried out to train models in an appropriate data format are presented. The neural network architectures implemented in the project are also outlined, and the results of their detection predictions on degraded images are compared.

A. Datasets

In this research, the influence of visual parameters on the quality of lymphocyte detection in histopathological images was examined by utilizing two datasets. The first dataset employed was the Leukocyte Images for Segmentation and Classification (LISC) [19], which contains hematological images taken from the peripheral blood of healthy individuals. These images were classified into five classes of leukocyte images, with a dimension of 720x576x3 pixels. The lymphocyte class consisted of 52 images, which were provided with ground truth for all images.

For the second dataset, a dataset from Andrew Janowczyk [12] was selected. This dataset comprises 100 images of ER + BCa images scanned at 20x, with each image measuring 100x100 pixels. An expert pathologist identified the lymphocyte centers.

The datasets were combined, and each image was resized to 256x256x3 pixels. Lymphocytes in the images were manually annotated using the Roboflow platform [5] based on the annotations of specialists. The final dataset contains a total of 2858 lymphocyte determinations, which were split into: 53% training set, 26% test set and 21% validation set.

B. Data preprocessing

Manual annotation of lymphocyte centers was conducted by an experienced pathologist to obtain raw ground truth images. In order to conform to the dimensions of the images on which the model was trained, a process of resizing all images to 256x256 pixels was performed. During this process, the original image scale was maintained. Subsequently, manual annotation of 152 lymphocyte images was carried out.

A Reinhard normalization method was used to normalize the lymphocyte test dataset images. The Reinhard algorithm was employed to match the color distribution of the source image to the color distribution of the target image, which was
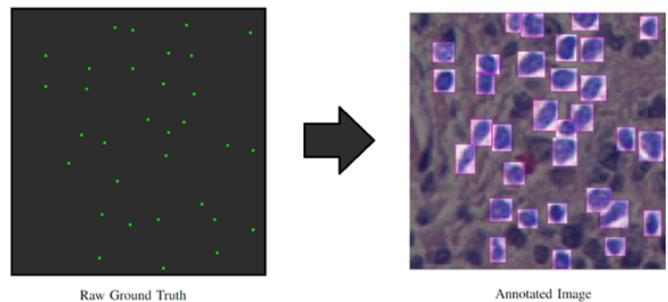
Fig. 1. A raw image of ground truth (from pathologist) and our annotated image using Roboflow platform.
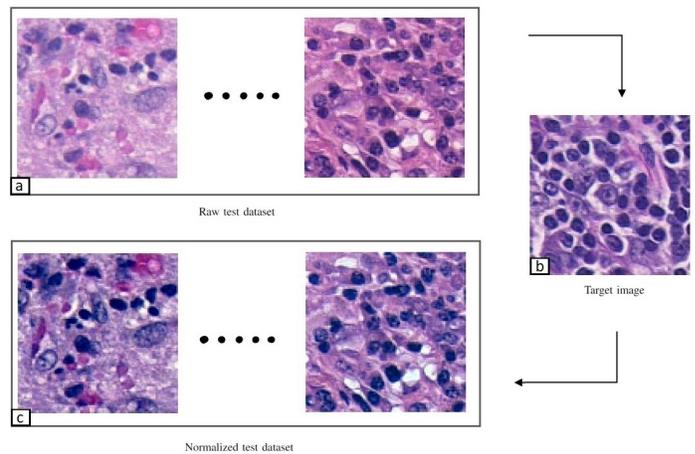
Fig. 2. a) Raw test dataset with different stains; b) Target color base image; c) Normalized test dataset.

achieved through equalizing the mean and standard deviation of the pixel values in each channel [3].

The impact of stain normalization on cell lymphocyte detection was quantified by comparing the testing metrics before and after the implementation of the Reinhard normalization algorithm on the test dataset (Table I).

TABLE I
COMPARING METRICS ON TEST DATASET WITH REINHARD
NORMALIZATION

	Original staining	Reinhard normalized staining
YOLOv5 AP	84.46%	83.34%
Detectron2 AP	83.16%	83.43%
YOLOv5 F1	87.62%	88.78%
Detectron2 F1	91.92%	91.26%

C. Models

In this paper, two machine learning frameworks were considered.

1) YOLOv5: the architecture was developed to predict detected objects in real time. In this study, YOLOv5 is used as it was adapted to be a part of the Python Pytorch library. It consists of three modules: the Backbone, Neck and Head. The Backbone module is mainly used to extract features from input image. It aggregates and forms image
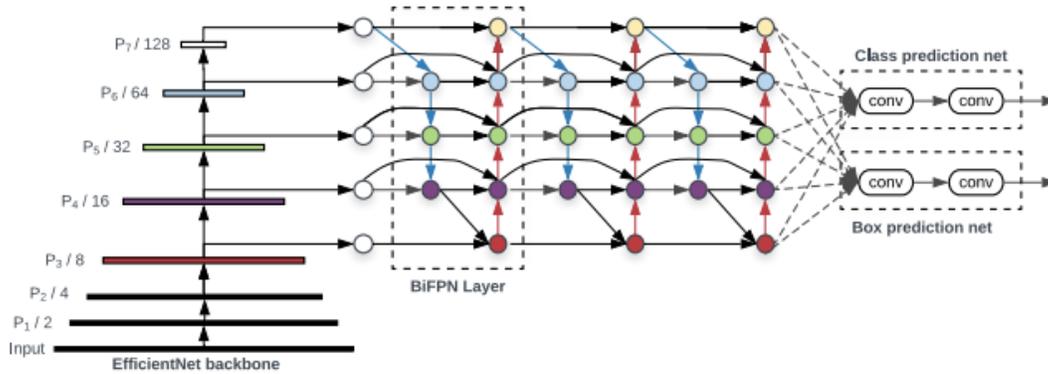
Fig. 3. The YOLOv5 model architecture [23].

features at different granularities [23]. Model Neck is used to generate feature pyramids which helps to identify the same object with different sizes and scales. It consists of a series of layers to mix and combine image features. The Head module performs final object detection. It predicts anchor boxes on features generating output vectors. Output vectors contain class probabilities, objectness scores and bounding box coordinates. YOLO family networks loss function is calculated on objectness score, class probability score and bounding box regression [24]. The YOLO neural network performs feature extraction and object classification methods at the same time during inference. In the YOLO network, images are divided into NxN grids. Candidate boxes are equally distributed on the x and y axis. The candidate boxes have object detection and predict the confidence of the existence of the object in each candidate box. Confidence scores reflect how confident the model is that the box contains an object and also how accurate is the prediction of the box. Authors defined confidence as Pr(Object) x IOU(truth prediction) [18]. If no object exists in that cell, the confidence score should be zero. In other cases, the confidence score should equal the intersection over union (IOU) between the predicted box and the ground truth.

Each bounding box consists of five predictions: x, y, w, h, and confidence. The (x, y) coordinates represent the center of the box relative to the bounds of the grid cell. The width and height are predicted relative to the whole image [18]. Each grid cell predicts C conditional class probabilities. These probabilities are conditioned on the grid cell containing an object. There is only one set of class probabilities predicted per grid cell, regardless of the number of boxes B. Scores of probability of object occurence and how well the predicted box fits the ground truth box are encoded into one result [18]. YOLO algorithm was used in the following application on lymphocyte detection [20]. There are five versions of YOLO v5 models. In this paper the model "s" is used to obtain maximum time performance during inference to compare faster model with other approaches. The weights file is about 14 MB in size, which is relatively small, however it has worse performance.

2) Detectron2: In order to compare the impact of image degradation on object detection between different networks the Detectron2 framework was used which offers various models. The RetinaNet model was selected [14]. It is an improvement of two existing models: Feature Pyramid Networks (FPN) and Focal Loss. FPN networks are based on object detection at varying scales in an image. Featurized image pyramids are feature pyramids built upon image pyramids. Images are subsampled into lower resolution with smaller size. This process is compute and memory intensive, which is why hand-engineered features were replaced by convolutional neural networks (CNNs). In a CNN architecture after every convolutional block feature maps decreases forming pyramidal structure. Pyramidal structure is needed to get the most accurate results. FPN combines low-resolution semantically strong features with high-resolution semantically weak features [14]. This is achieved by creating a top-down pathway with lateral connections to bottom-up convolutional layers. RetinaNet consists of four elements: the bottom-up pathway, top-down pathway and lateral connections, classification subnetwork and regression subnetwork. The bottom-up pathway is the feed-forward computation of the backbone convolutional network, which computes a feature hierarchy consisting of feature maps at several scales [14]. The top-down pathway upsamples feature maps from higher pyramid levels. The lateral connections merge top-down layers with bottom-top layers with the same spatial size. Classification subnetwork predicts probability of every object detected at each stage. Regression subnetwork regresses offset for predicted box with ground truth object. Another important element of RetinaNet is Focal Loss. The Focal Loss is a new loss function that acts as a more effective alternative to previous approaches for dealing with class imbalance [15]. It assigns more weight to hard or easily misclassified examples. Focal loss is an extension of the cross entropy loss function that would down-weight easy examples and focus training on hard negatives. In this project weights of RetinaNet model are about 797 MB. The difference between size of compared models matter in time of inference. The YOLOv5 algorithm inference is much faster but RetinaNet from Detectron2 framework has more accurate prediction.
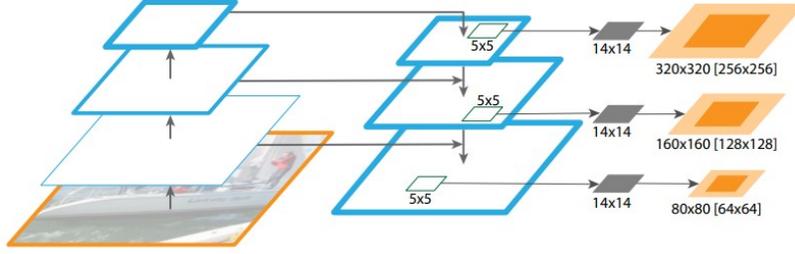
Fig. 4. The RetinaNet model architecture [14].

D. Application

In order to systematically evaluate the impact of different image imperfections on the performance of lymphocyte detection algorithms, a specialized web application was developed that facilitates the easy modification of various image parameters and observation of the resulting detections. The user-friendly interface of the application enables the quick and easy application of different image degradation techniques, such as blur, noise, and contrast adjustments, and immediately allows for the evaluation of the effect on the performance of the YOLOv5 and Detectron2 models.

In addition to providing a convenient platform for conducting experiments, the application is also integrated with an experiment tracking platform [1]. This allows for accurate and comprehensive tracking of the performance of the models under various conditions and the identification of patterns and trends that may not be immediately apparent from visual inspection alone. Through the use of this application, valuable insights into the factors that influence the performance of lymphocyte detection algorithms were gained and strategies for improving their accuracy and robustness were identified.

IV. IMAGE DEGRADATION

The purpose of this section is to describe the image degradation techniques that were applied to the test dataset in order to evaluate the performance of the YOLOv5 and Detectron2 models in the presence of various types of image imperfections. These imperfections included blur, noise, contrast, sharpness, and brightness modifications. In order to apply these modifications, various image processing techniques and algorithms were used. The results of these modifications were then used to evaluate the performance of the models.

A. Gaussian Blur

Blur is a common issue in image processing, and it can significantly affect the performance of object detection algorithms. Blur can be caused by various factors, such as camera shake, out-of-focus subjects, or poor lighting conditions. In the field of medical imaging, blur can be particularly problematic during the digitization process of histopathological samples, as the camera focus may vary in natural settings. A study published by Arizona State University [4] found that blur has a negative impact on image classification, particularly when

compared to other degradation methods. The authors of this study found that blur and noise had the greatest influence on the performance of convolutional neural networks for classification tasks. It is believed that blur interferes with the ability of convolutional layers to locate edges at early stages of feature abstraction, leading to inaccurate feature extraction at the beginning of the network training process.

To mitigate the negative effects of blur on object detection, researchers have proposed various approaches, such as augmenting the training dataset with blurred images [16]. By increasing the robustness of the model to blur, it is expected to improve the performance of object detection in real-world scenarios, where blur is a common issue.

To simulate blurring and examine its impact on lymphocyte detection, the Gaussian kernel was used, a method as described in Gwosdek et al. [9]. This method convolves the image with a Gaussian kernel, which is a matrix of weights generated based on a specified standard deviation. The size of the kernel is determined by the radius parameter, which specifies the standard deviation of the Gaussian distribution. A larger radius will result in a larger kernel and more smoothing, while a smaller radius will result in a smaller kernel and less smoothing. In our experiments, the kernel standard deviation was varied to study the effect of different levels of blur on the performance of the object detection algorithms. Figure 5 shows the same image with different levels of blur applied, demonstrating the visual impact of this image degradation method.

B. Noise

The next image degradation parameter that was tested is noise. Noise causes unwanted effects on the image such as artifacts, unrealistic edges, invisible lines, corners, blurred objects and distorts background scenes. In this paper, the influence of three selected noises - Gaussian noise, Speckle noise and Salt and Pepper noise on the performance of state-of-the-art algorithms for detecting lymphocytes in histopathological images was examined. These three types help to simulate noisy images taken by non-high quality or malfunctioning cameras. The range of noise parameters was increased as in the presented work [21].

Gaussian noise: is statistical noise with a probability density function equal to the normal distribution, also known as the Gaussian distribution. The random Gaussian function is
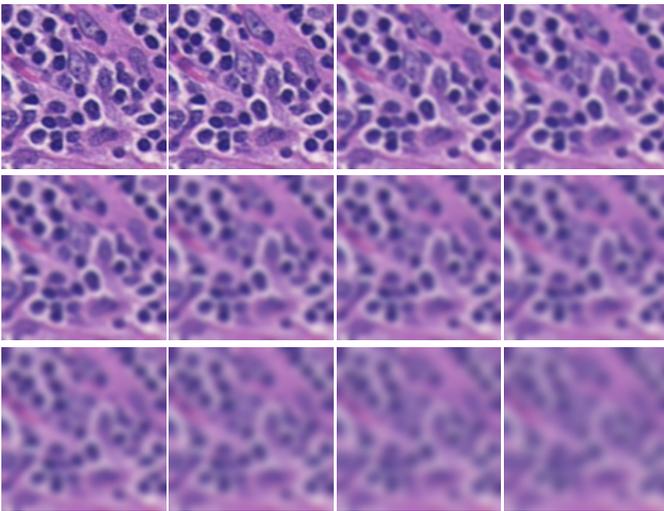
Fig. 5. Example images from the test dataset with different levels of Gaussian blur applied - radius size: 1.0, 1.5, 2.5, 3.0, 3.5, 4.5, 5.0, 5.5, 6.0, 7.0, 8.0 and 9.0

added to the image function to generate this noise. It is also called electronic noise because it is generated in amplifiers or detectors. The source is the thermal vibrations of atoms and the discrete nature of the radiation of warm objects [2]. The range used to add Gaussian noise was started with variance 0, which is the original image, and then increased to variance 2.56 using log space. The results of the image with noise are shown in Figure 6.
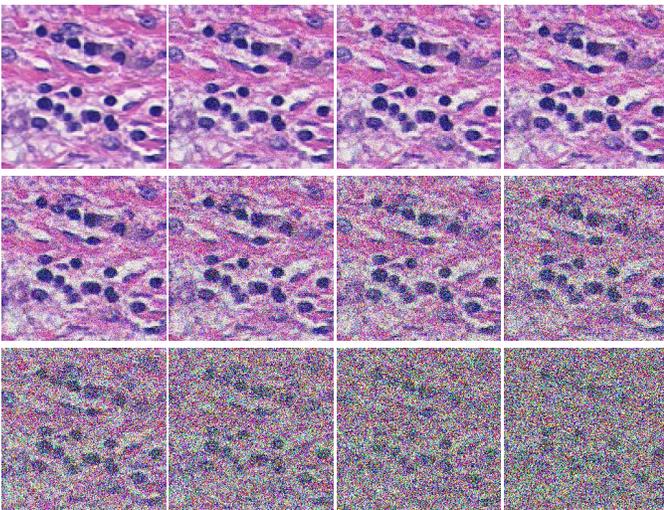
Fig. 6. Example images from the test dataset with different levels of Gaussian noise applied - variance values: 0, 0.01, 0.02, 0.03, 0.05, 0.09, 0.16, 0.28, 0.49, 0.84, 1.47 and 2.56.

Speckle noise: multiplicative noise. It can be generated by multiplying random pixel values by different pixels of the image. The presence of this noise is a fundamental problem in optical holography and digital image reconstruction. Speckle noise can appear in an image similarly to Gaussian noise, and its probability density function follows gamma distribution [2]. Like in the Gaussian noise equation, the variance is a variable that has been changed to control how much noise is

present in the image, which ranges from 0 to 15 as shown in Figure 7.
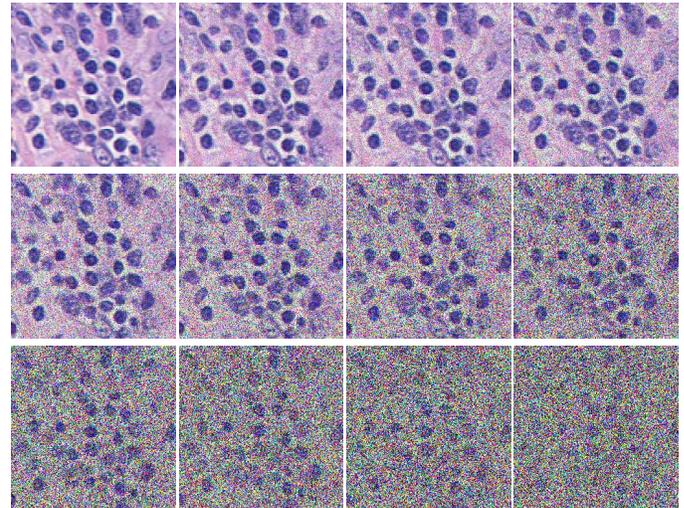
Fig. 7. Example images from the test dataset with different levels of Speckle noise applied - variance values: 0, 0.03, 0.06, 0.1, 0.19, 0.36, 0.67, 1.25, 2.32, 4.33, 8.06 and 15.0

Salt and Pepper Noise: is also known as data drop noise, because it causes the original data values to drop. S&P noise is randomly distributed over the image and can only be a minimum (pepper) or maximum (salt) value in a typical image range [0, 255] [21]. The amount of S&P noise within the images was modified using a random, uniform distribution of half salted pixels and half peppered pixels. As shown in the Figure 8, the noise level started with 0% of noisy pixels and increased to 50% over time.
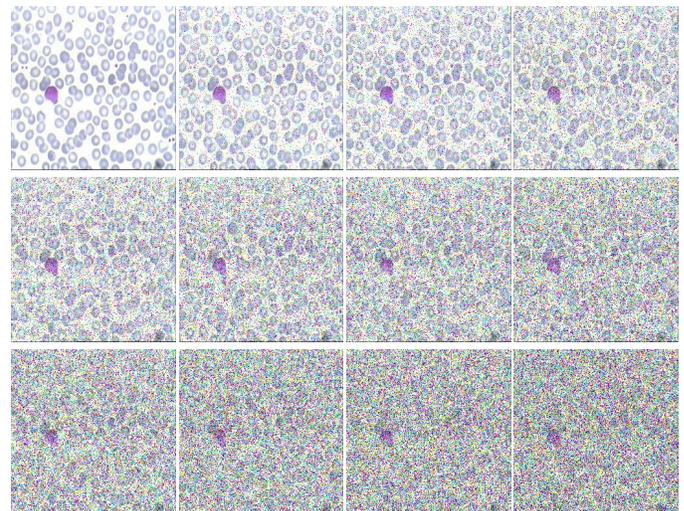
Fig. 8. Example images from the test dataset with different levels of Salt and Pepper noise applied - the percentages used in this range: 0%, 5%, 9%, 14%, 18%, 23%, 27%, 32%, 36%, 41%, 45% and 50%.

C. Sharpness

Sharpness is inversely related to blur which is typically determined by the spread of edges in the spatial domain,

and accordingly the attenuation of high frequency components [26]. Digital images have limitations of distinguishment lines of contrast with such clarity and sharpness as human eye. Sensors of digital cameraa are limited by the number of pixels and their range of values. Sharpness is a combination of two factors: resolution and acutance. Resolution is straightforward and not subjective. It depends on the number of pixels in the image file. All other factors equal, the higher the resolution of the image, the sharper it can be. Acutance is a subjective measure of the contrast at an edge. Image sharpening is performed by applying image convolution with mask with specified parameters mentioned in [6] or by applying algorithm presented in [7]. To perform image sharpening the Python Pillow ImageEnhance module was used with specified sharpening factors. The factor value ranged from 1.0 to 12.0 with a 1.0 step. As in previous parameter, library component preserves the original image with factor value 1.0. Models evaluation is presented below.
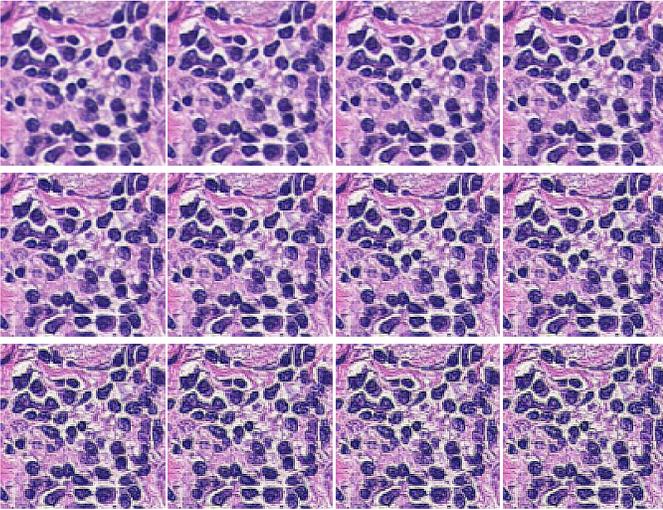
Fig. 9. Examples image from the test dataset with different levels of of sharpness applied - sharpness levels: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12

D. Saturation

Another image parameter considered in data augmentation is color saturation. It adjusts how vibrant the image is. Saturation is the amount of gray mixed in with the pure colors in each pixel. A fully desaturated image is grayscale and a highly saturated image contains more pure colors. An increased saturation shifts colors more towards the primary colors while decreased saturation mutes colors up to grayscale. Adjusting saturation to images during the data augmentation process can help models perform when the colors are affected by weather conditions. Selected models were evaluated on a modified test dataset. FFor this purpose the Pillow ImageEnhance module was used to change the saturation of images. Exact factor values applied to methods and model results are given below. The factor value ranged from 0.25 to 1.0 with a 0.25 step. A value of 1.0 preserves the original image. Then, the factor was increased to 5.0 with a 1.0 step.

E. Brightness

Image brightness is one of the simplest pixel value modification to perform on. Modifying brightness can be accomplished by multiplication between pixels values with a specified ratio. Discrimination between different intensity levels on image is an important consideration. The range of light intensity levels to which the human visual system can adapt is wide from the scotopic threshold to the glare limit [7]. Computer colors representation are more limited. The proper settings of image brightness modification can lead to increased model ability to generalization and birightness independence. Brightness manipulation on images allows us to generate new images samples in dataset. This allows us to generate images with a range from black to white. Spot brightness modification can lead to add shadow or bright spots to simulate complicated light condition in real situation. To perform brightness modification the Pillow module was used to manipulate brightness over dataset images to analyze how this parameter affects lymphocyte detection between two selected models. The factor value ranged from 1.0 to 9.0 with a 1.0 step. As in previous parameter, library component preserves the original image with factor value 1.0. Models evaluation is presented below.
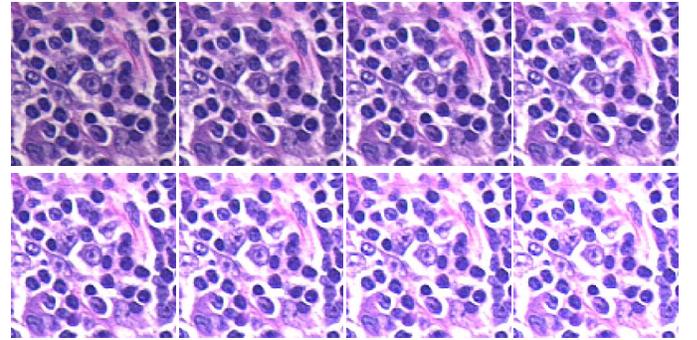
Fig. 10. Example images from the test dataset with different levels of brightness applied - the percentages used in this range: 0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%.

F. Contrast

Contrast is the electro-optical parameter of the image, which determines the ratio of the maximum luminance to the minimum luminance. When the contrast adjustment is raised, the image will have a higher percentage of dark pixels and whites (middle pixel values are eliminated). Contrast reduction is obtained by blending the input image with a gray image [10].

$$output = (1 - alpha) \cdot image + alpha \cdot image \quad (1)$$

This equation was used to change images contrast. Alpha is a key factor, that has been changed in range from 0.1 to 3.0. The experiment was started with an alpha value 0.1, next from an alpha value of 0.2 to 2.0 in steps of 0.2. Then, the alpha value was increased to 2.5, where last step was 3.0. An alpha value of 1.0 is the same as in other image degradation experiments present the original image.
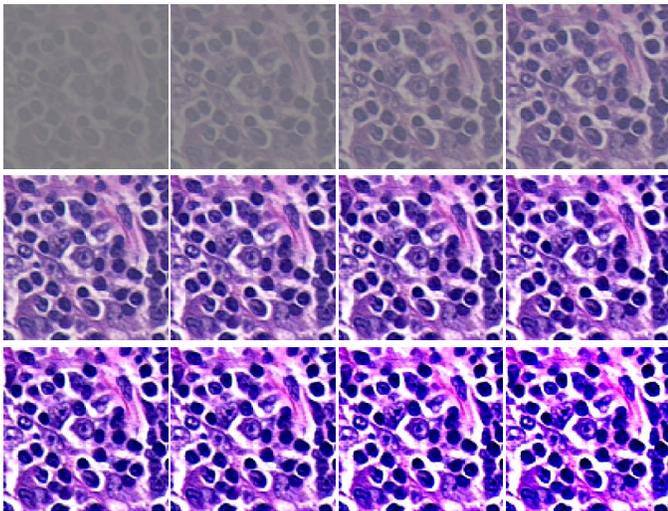
Fig. 11. Example image from the test dataset with different levels of contrast applied - contrast value: 0.1, 0.2, 0.4, 0.6, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0, 2.5, 3.0.

V. EXPERIMENTS AND RESULTS

The purpose of the experiments in this study was to investigate the impact of various image imperfections on the performance of state-of-the-art object detection algorithms. To do this, a test dataset of 32 images was selected and various modifications were applied to them, including blur, noise, sharpness, saturation, brightness and contrast. The remaining original data was used for training both the YOLOv5 and Detectron2 models. The performance of the models was then measured using metrics such as Average Precision (AP) and F1-score. On the test dataset, both models were predicted with a confidence threshold of 0.2. The average precision metric was calculated with the IoU parameter set to 0.7.

In the following subsections, the results of these experiments and discuss their implications for object detection in the context of histopathological images are presented.

1) Gaussian Blur: The results of applying Gaussian blur on the test dataset are shown in Table II. The blur radius size is varied from 0.5 to 9, and for each size, the Average Precision (AP) and F1-score are calculated for both models. As the radius size of the Gaussian blur increases, the average precision and F1-score of both YOLOv5 and Detectron2 decrease starting from radius equal to 1. This suggests that as the level of blur in the images increases, the ability of both models to accurately detect objects decreases. This is consistent with previous research that has shown that blur can significantly degrade the performance of deep neural networks for image classification tasks [4].

It is worth noting that both models experienced a more significant decrease in performance at radius sizes above 5, with YOLOv5 reaching 0% average precision at radius size 9 and Detectron2 reaching 0% at radius size 6. This suggests that there is a threshold for the level of blur that the models can tolerate before experiencing a significant drop in performance.

2) Noise: The results of the applied noise are presented in the following tables – Gaussian Noise Table III, Speckle Noise Table IV and Salt and Pepper Noise Table V. For higher values of overlaid noise degradation, a significant visual decrease in image quality and loss of important image features, i. e. loss of lymphocyte cell edges was observed due to introduced disturbances. After the experiments, an overall decrease in lymphocyte detection capacity for both models was observed. Compared to Retnianet, the YOLO model achieves both higher values of Average Precision and F1-score.

TABLE II
METRICS ON TEST DATASET WITH GAUSSIAN BLUR APPLIED

Radius size	YOLOv5 AP	Detectron2 AP	YOLOv5 F1	Detectron2 F1
0.5	83.82%	82.20%	87.65%	91.86%
1	81.02%	82.63%	88.63%	91.24%
1.5	80.84%	83.25%	84.10%	89.85%
2	76.87%	77.56%	80.67%	85.29%
2.5	65.16%	72.87%	77.02%	81.34%
3	59.24%	63.57%	68.40%	65.90%
3.5	53.04%	48.43%	62.41%	55.15%
4	45.33%	39.46%	46.46%	28.66%
4.5	32.42%	16.15%	35.80%	4.49%
5	25.17%	1.66%	21.64%	4.02%
5.5	20.49%	1.65%	4.25%	1.91%
6	18.25%	0%	0.03%	0.06%
6.5	17.71%	0%	0%	0.01%
7	15.76%	0%	0%	0%
7.5	6.25%	0%	0%	0%
8	3.13%	0%	0%	0%
8.5	3.12%	0%	0%	0%
9	0%	0%	0%	0%

TABLE III
METRICS ON TEST DATASET WITH GAUSSIAN NOISE APPLIED

Variance	YOLOv5 AP	Detectron2 AP	YOLOv5 F1	Detectron2 F1
0	83.24%	83.55%	87.56%	92.00%
0.01	83.37%	74.91%	87.76%	83.28%
0.02	79.32%	64.48%	87.48%	78.19%
0.03	78.47%	70.04%	85.79%	80.48%
0.05	74.51%	50.94%	83.25%	66.05%
0.09	63.92%	25.77%	64.70%	31.74%
0.16	32.65%	16.44%	33.76%	12.44%
0.28	2.05%	1.47%	7.43%	1.25%
0.49	1.83%	0%	0.60%	0%
0.84	0.10%	0%	0.17%	0%
1.47	0%	0%	0%	0%
2.56	0%	0%	0%	0%

TABLE IV
METRICS ON TEST DATASET WITH SPECKLE NOISE APPLIED

Variance	YOLOv5 AP	Detectron2 AP	YOLOv5 F1	Detectron2 F1
0	83.24%	83.55%	87.56%	92.00%
0.03	82.21%	69.70%	87.05%	83.43%
0.06	78.34%	67.45%	87.43%	76.12%
0.1	76.36%	44.99%	80.63%	65.84%
0.19	54.70%	30.96%	56.90%	38.87%
0.36	21.09%	16.93%	36.96%	29.92%
0.67	13.17%	10.91%	26.46%	19.62%
1.25	6.04%	5.00%	18.46%	7.90%
2.32	1.09%	1.20%	8.93%	2.46%
4.33	0.46%	0%	2.37%	0%
8.06	0.01%	0%	0.48%	0%
15	0%	0%	0%	0%

TABLE V
METRICS ON TEST DATASET WITH SALT AND PEPPER NOISE APPLIED

Noise range	YOLOv5 AP	Detectron2 AP	YOLOv5 F1	Detectron2 F1
0%	83.24%	83.55%	87.56%	92.00%
5%	80.93%	67.11%	86.85%	82.17%
9%	79.38%	48.10%	85.05%	55.40%
14%	66.62%	30.62%	77.77%	35.00%
18%	66.84%	22.00%	63.10%	29.58%
23%	37.30%	15.47%	52.51%	16.58%
27%	32.70%	7.63%	36.29%	8.89%
32%	23.06%	4.50%	17.56%	4.31%
36%	16.17%	1.96%	13.12%	2.30%
41%	1.06%	0.94%	5.70%	0.45%
45%	1.09%	0.33%	3.17%	0.60%
50%	0.23%	0%	0.70%	0%

3) Sharpness: The results of images sharpness modification are shown in Table VI. To perform image sharpening the sharpness factor value was ranged from 1.0 to 12.0 with a 1.0 step. The value of factor 1.0 preserves quality of the original image. Both models lose their detection capability as the parameter increases. The YOLO model has a greater Average Precision value over test dataset modifications. The Retinanet model has a higher F1 score during this test.

TABLE VI
METRICS ON TEST DATASET WITH SHARPNESS MODIFICATION APPLIED

Sharpness level	YOLOv5 AP	Detectron2 AP	YOLOv5 F1	Detectron2 F1
1	83.20%	83.58%	87.56%	92.05%
2	83.01%	82.13%	87.38%	91.33%
3	83.67%	80.90%	85.67%	90.23%
4	82.31%	75.01%	85.11%	87.92%
5	81.93%	73.02%	83.33%	88.19%
6	75.10%	73.03%	83.24%	84.91%
7	74.00%	71.24%	81.67%	84.09%
8	68.60%	70.98%	80.95%	83.80%
9	67.46%	69.78%	78.17%	81.04%
10	59.95%	65.88%	74.04%	77.40%
11	58.50%	63.52%	70.79%	77.14%
12	52.57%	56.98%	66.48%	74.92%

4) Saturation: The results of image saturation modification are shown in Table VII. This test was performed by increasing the saturation factor value from 0.25 to 1.0 with a step of 0.25. The value of 1.0 preserves the original quality of the image. Then, the factor value was increased to 5.0 with a 1.0 step. The detection capability loss is observed with both an increase and a decrease in the parameter value. Both metrics were higher for the Retinanet model for parameter values greater than 0.75.

5) Brightness: The results of images brightness modification are shown in Table VIII. This test was performed by increasing the saturation factor from 1.0 to 8.0 at step 1.0. The factor value of 1.0 preserves the original quality of the image. The metrics of both models decrease as the parameter increases. With a parameter value of 8, both models has Average Precision value less than 50%. The YOLO model has higher Average Precision value and RetinaNet has higher F1 score during this test.

6) Contrast: The results of modifying image contrast on the test dataset are shown in Table IX. The Alpha parameter

TABLE VII
METRICS ON TEST DATASET WITH SATURATION MODIFICATION APPLIED

Saturation level	YOLOv5 AP	Detectron2 AP	YOLOv5 F1	Detectron2 F1
0.25	74.40%	69.18%	86.91%	73.89%
0.5	80.61%	75.87%	88.00%	91.49%
0.75	83.25%	83.25%	88.62%	92.04%
1	83.20%	83.58%	87.56%	92.05%
2	76.33%	78.96%	74.60%	79.80%
3	67.86%	74.54%	49.70%	68.56%
4	44.47%	73.38%	45.10%	59.21%
5	33.95%	71.07%	38.93%	52.34%

TABLE VIII
METRICS ON TEST DATASET WITH BRIGHTNESS MODIFICATION APPLIED

Brightness level	YOLOv5 AP	Detectron2 AP	YOLOv5 F1	Detectron2 F1
1	83.20%	83.58%	87.56%	92.05%
2	82.01%	78.42%	88.21%	91.74%
3	79.26%	75.07%	91.20%	89.95%
4	75.79%	76.79%	90.97%	87.39%
5	71.59%	60.57%	86.82%	85.17%
6	68.31%	58.42%	79.38%	82.17%
7	55.63%	51.39%	67.68%	72.40%
8	49.16%	39.57%	61.71%	59.89%

is in range from 0.1 to 3.0, and for each iteration, the Average Precision and F1-score are calculated for both models. When the contrast value decreases below 1.0, the AP and F1-score of both YOLOv5 and Detectron2 also decreases. This suggests that if the contrast decreases, both models have more and more difficulty in lymphocytes predicting. For Alpha parameter = 0.1 both models were unable to detect lymphocytes.

It is worth noting that both models F1 scores remained at about the same level when the contrast was increased. At the same time, AP metrics of both models scores noticeably deteriorated.

TABLE IX
METRICS ON TEST DATASET WITH DIFFRENT CONTRAST

Contrast level	YOLOv5 AP	Detectron2 AP	YOLOv5 F1	Detectron2 F1
0.1	0%	0%	0%	0%
0.2	0%	8.93%	0%	0.59%
0.4	49.16%	64.78%	30.54%	75.11%
0.6	78.18%	78.84%	80.15%	88.91%
0.8	83.90%	82.72%	86.85%	90.77%
1	83.43%	83.16%	87.62%	91.92%
1.2	80.59%	82.12%	87.29%	91.75%
1.4	78.15%	79.14%	86.36%	91.52%
1.6	75.94%	76.73%	85.74%	91.40%
1.8	77.80%	80.24%	87.74%	91.03%
2.0	77.75%	74.04%	87.44%	91.04%
2.5	72.90%	72.70%	88.15%	90.88%
3.0	69.91%	69.43%	88.36%	89.97%

VI. DISCUSSION AND CONCLUSION

In this study, the effect of image quality characteristics on the detection of lymphocytes in histopathological images using state-of-the-art algorithms was evaluated. The results indicated that parameters such as image blur, noise, contrast,

brightness, saturation, and sharpness have a significant impact on the performance of these algorithms. Specifically, it was observed that images with higher levels of blur, lower contrast, and lower sharpness resulted in reduced detection rates of lymphocytes. A higher AP metric was for the YOLO network and a higher F1 metric was for RetinaNet from Detectron2. The RetinaNet model detects objects with greater precision on the test dataset but with worse inference performance than the YOLO network. The use of a specific model depends on the purpose of the application and the conditions under which it could be launched.

Furthermore, the results demonstrated that the performance of the algorithms varied depending on the specific method employed. For instance, the algorithm that performed optimally on images with high blur was not the same as the algorithm that performed optimally on images with low contrast. This highlights the importance of considering the specific visual quality characteristics of an image when selecting an algorithm for lymphocytes detection.

These findings have significant implications in the field of histopathological image analysis. They underscore the importance of maintaining optimal image quality during the digitization process of histopathological samples, such as by ensuring proper focus of the camera to minimize blur. Poor image quality may compromise the accuracy of object detection models, leading to unreliable results in lymphocyte detection task.

References

[1] Lukas Biewald. *Experiment Tracking with Weights and Biases*. Accessed on 26.01.2023. (2020). URL: https://www.wandb.com/.

[2] Ajay Kumar Boyat and Brijendra Kumar Joshi. "A Review Paper: Noise Models in Digital Image Processing". In: (2015). Computer Vision and Pattern Recognition. DOI: 10.48550/ARXIV.1505.03489. URL: https://arxiv.org/abs/1505.03489.

[3] Elzbieta Budginaitė et al. "Deep learning model for cell nuclei segmentation and lymphocyte identification in whole slide histology images". In: *Informatica* 32.1 (2021), pp. 23–40.

[4] Samuel F. Dodge and Lina J. Karam. "Understanding How Image Quality Affects Deep Neural Networks". In: *CoRR* abs/1604.04004 (2016). pp. 1-6. arXiv: 1604.04004. URL: http://arxiv.org/abs/1604.04004.

[5] Solawetz J. Dwyer B. Nelson J. *al. Roboflow (Version 1.0) [Software]*. Accessed on 26.01.2023. (2022). URL: https://roboflow.com.

[6] Jason Fitzpatrick. "How Digital Image Sharpening Works, and Why You Should Use It". In: (2016).

[7] Rafael C Gonzalez, Richard E Woods, and SL Eddins. "Image segmentation". In: *Digital image processing* 2 (2002), pp. 331–390.

[8] Rodrigo Escobar Díaz Guerrero and José Luís Oliveira. "Improvements in lymphocytes detection using deep learning with a preprocessing stage". In: *2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS)*. (2021), pp. 178–182. DOI: 10.1109/CBMS52027.2021.00068.

[9] Pascal Gwosdek et al. "Theoretical foundations of gaussian convolution by extended box filtering". In: *International Conference on Scale Space and Variational Methods in Computer Vision*. Springer. (2011), pp. 447–458.

[10] Paul Haeberli and Douglas Voorhies. "Image processing by linear interpolation and extrapolation". In: *IRIS Universe Magazine* 28 (1994), pp. 8–9.

[11] Shona Hendry et al. "Assessing tumor infiltrating lymphocytes in solid tumors: a practical review for pathologists and proposal for a standardized method from the International Immuno-Oncology Biomarkers Working Group: Part 2: TILs in melanoma, gastrointestinal tract carcinomas, non-small cell lung carcinoma and mesothelioma, endometrial and ovarian carcinomas, squamous cell carcinoma of the head and neck, genitourinary carcinomas, and primary brain tumors". In: *Advances in anatomic pathology* 24.6 (2017), p. 311.

[12] Andrew Janowczyk. *Use case 4: Lymphocyte detection*. http://www.andrewjanowczyk.com/use-case-4-lymphocyte-detection/. Accessed on 26.01.2023. (2015).

[13] Glenn Jocher et al. *ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation*. Version v7.0. (2022). DOI: 10.5281/zenodo.7347926. URL: https://doi.org/10.5281/zenodo.7347926.

[14] Tsung-Yi Lin et al. "Feature pyramid networks for object detection". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. (2017), pp. 2117–2125.

[15] Tsung-Yi Lin et al. "Focal loss for dense object detection". In: *Proceedings of the IEEE international conference on computer vision*. (2017), pp. 2980–2988.

[16] Chengji Liu et al. "Object detection based on YOLO network". In: *2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC)*. IEEE. (2018), pp. 799–803.

[17] Şaban Öztürk and Bayram Akdemir. "Effects of histopathological image pre-processing on convolutional neural networks". In: *Procedia computer science* 132 (2018), pp. 396–403.

[18] Joseph Redmon et al. "You only look once: Unified, real-time object detection". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. (2016), pp. 779–788.

[19] H. Rezatofighi S.H. Soltanian-Zadeh. *Automatic recognition of five types of white blood cells in peripheral blood*. http://users.cecs.anu.edu.au/ hrezatofighi/Data/Leukocyte%20Data.htm. (2011).

[20] Mart van Rijthoven et al. "You only look on lymphocytes once". In: (2018). Medical Imaging with Deep Learning, Amsterdam.

[21] Matthew Seals. *On the Robustness of Object Detection Based Deep Learning Models*. Master's Thesis, University of Tennessee. (2019). URL: https://trace.tennessee.edu/utk_gradthes/5487.

[22] Nilgün Şengöz et al. "Importance of Preprocessing in Histopathology Image Classification Using Deep Convolutional Neural Network". In: *Advances in Artificial Intelligence Research* 2.1 (2022), pp. 1–6.

[23] Mingxing Tan, Ruoming Pang, and Quoc V. Le. "EfficientDet: Scalable and Efficient Object Detection". In: (2019). DOI: 10.48550/ARXIV.1911.09070. URL: https://arxiv.org/abs/1911.09070.

[24] Chien-Yao Wang et al. "CSPNet: A new backbone that can enhance learning capability of cnn". In: (2020), pp. 390–391.

[25] Yuxin Wu et al. *Detectron2*. https://github.com/facebookresearch/detectron2. Accessed on 26.01.2023. (2019).

[26] Shaode Yu et al. "A shallow convolutional neural network for blind image sharpness assessment". In: *PloS one* 12.5 (2017), e0176632.