To reject or not to reject, that is the question: A new perspective on counting fragments in Whole-Slide-Images

Ana Beatriz Vieira[1,2]
anabeatrizvieira@tecnico.ulisboa.pt

Tomé Albuquerque[3,4]
tome.m.albuquerque@inesctec.pt

Diana Montezuma[5], Liliana Ribeiro[5], Domingos Oliveira[5], Isabel Macedo Pinto[5]
diana.felizardo@impdiagnostics.com

Jaime S. Cardoso[3,4]
jaime.cardoso@inesctec.pt

Arlindo L. Oliveira[1,2]
arlindo.oliveira@tecnico.ulisboa.pt

[1] Instituto Superior Técnico
Lisbon, Portugal

[2] INESC-ID
Lisbon, Portugal

[3] Faculdade de Engenharia da Universidade do Porto
Porto, Portugal

[4] INESC-TEC
Porto, Portugal

[5] IMP Diagnostics
Porto, Portugal

Abstract

Quality control of medical images plays an important role in digital pathology since verifying that the images meet all requirements can imply manual analysis. Manual assessment of pathology specimen fragments is intended to ensure that the number of fragments described in the macroscopic report corresponds to the number of fragments present on the slide, avoiding the loss of valuable material during grossing. However, this process is currently performed manually and is time-consuming and subjective. We applied an object detection model, YOLOv5, to detect fragments and sets in Whole-Slide-Images dataset from IMP Diagnostics. Subsequently, we counted the final number of fragments by dividing the number of fragments by the number of sets. We decided to add a reject option when the confidence was low, based on the value of the division of fragments and sets, forcing the rejection of the sample if this number is not an integer. When tested on a set of 700 images, the model achieves an overall accuracy of 87.9% (without rejection), which increases to 92.8% if we reject 10.9% of the samples. The reject option allows the model to exclude the samples with the lowest confidence value and, therefore, which give rise to the most doubts in the count. But to what extent is rejecting samples beneficial?

1 Introduction

Digital pathology (DP) has become a game-changer, introducing new ways of improving diagnostic accuracy and simplifying healthcare processes. A crucial component of this field is quality control, a meticulous process that guarantees the reliability and trustworthiness of pathological analyses. An important quality control process in pathology laboratories involves the precise counting of pathological fragments present on microscopic slides. This counting procedure aims to ensure that the number of fragments described in the macroscopic report corresponds to the number of fragments observed on the slide [1, 4]. Since this is a manual and time-consuming procedure, we propose an automated system to replace this manual step. Figure 1 represents the manual and automatic assessment.

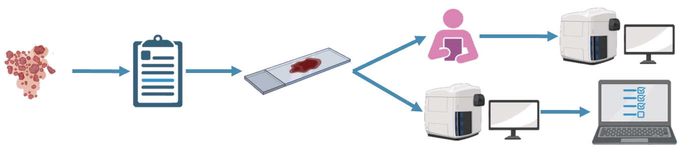

Figure 1: Illustration of manual and automatic assessment.

The aim is to train a model to learn with the different tissue placement combinations, in order to detect and count the number of fragments and sets present in each Whole-Slide-Image (WSI), which is obtained after the slide scanning. The WSI is a high-resolution digital file that is usually obtained by sequentially capturing small blocks or strips of high-resolution images, which are then assembled to create a complete picture of a histological section, reproducing the glass slide [2]. Figure 2 shows an example of different combinations of our WSIs.

The repeating sets (cases 2.b and 2.c) allow to increase the observed material (more cuts per fragment), and are a common practice in many pathology laboratories. The main difference between our approach and previous work is the inclusion of a rejection option, which allows the model to reject samples when confidence is low [3], so that images need to be reviewed and classified manually.

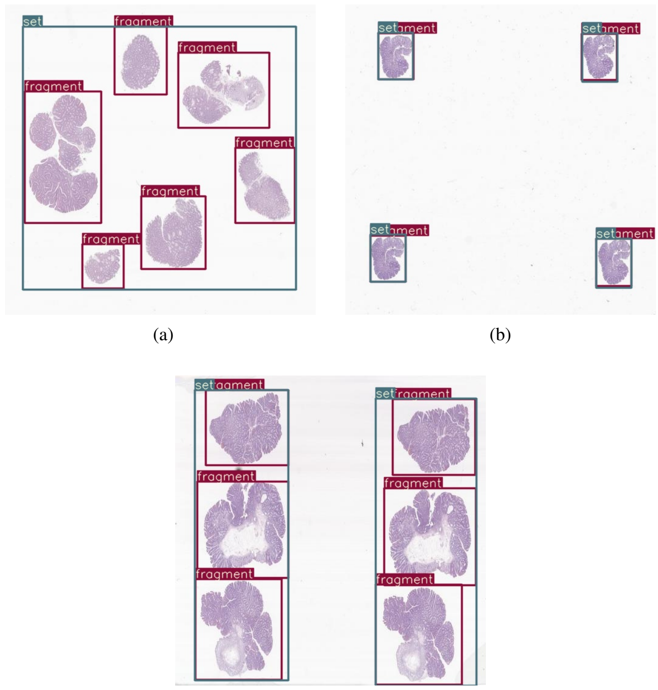

Figure 2: Representation of different combinations of tissue slides: (a) one set with different fragments; (b) several sets with only one fragment; (c) two equal sets of three different fragments.

2 Methodology

2.1 Dataset

Our dataset consists of 3254 Whole-Slide-Images from IMP Diagnostics archive, a labeled dataset of different pathology samples, digitized with 2 Leica GT450 WSI scanners, at 40x equivalent magnification. The dataset was divided into train, validation, and test subsets with 2054, 500, and 700 samples of size 512x512px, respectively. To overcome the relative small amount of data in our dataset, different types of data augmentation were used during the training - vertical and horizontal flipping, and HSV (Hue, Saturation, Value) augmentation. In addition, the images were re-scaled from the range of 0-255 to 0-1.

2.2 Detection

Since to count objects we first need to identify them, we decided to divide the work into two phases: detection and counting. For the first stage of the proposed work, we applied a state-of-the-art object detection model, YOLOv5 to detect fragments and sets in each image. To feed the network, each input image is spatially annotated with the measurements of the bounding boxes (height, width, and center coordinates) and the class (0: fragment, 1: set). Each output image is followed by a file containing the information mentioned above and the probability value of the object's detection.

2.3 Counting

For the second stage, we submitted the output images from the YOLOv5 model to multiple rules to improve the counting performance and classify them into one of the ten existing classes. This counting step is performed by dividing the number of fragments by the number of sets, in each image. Finally, in order to increase the accuracy of the counting process, we added a rejection option when the confidence was low. The rule to reject the automatic counting is based on dividing the number of fragments by the number of sets. If this number is not an integer, indicating an inconsistency in the number of fragments per set, a warning is given, and that sample is not classified by the model and must be reviewed and classified manually. Figure 3 shows the main steps of our work.

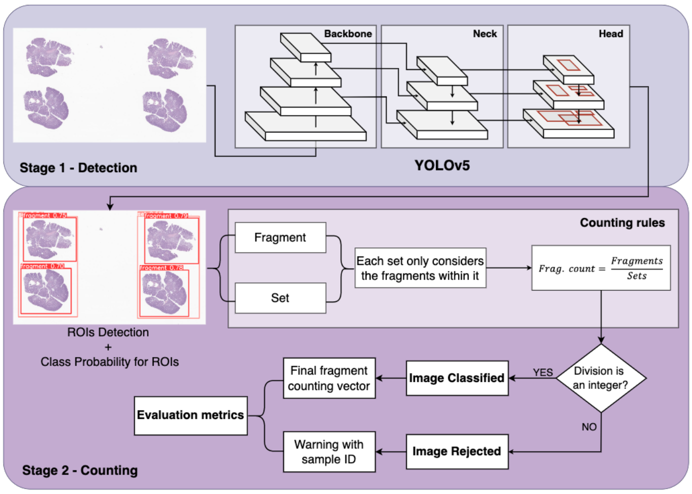

Figure 3: Stages of the proposed method.

3 Results and Discussion

The model achieved an overall accuracy of 87.9%, without rejection, i.e., when all 700 samples were classified. On the other hand, by rejecting 10.9% of the samples, the accuracy increases to 92.8%, meaning that 624 images would be classified and 76 rejected. Other evaluation metrics are shown in Table 1, which demonstrates that the task benefits from applying the reject option.

Table 1: Evaluation metrics for fragment counting.

Rejection	Accuracy	MAE	MSE	F1-Score	Precision
No	87.9%	0.187	0.610	88.1%	88.5%
Yes	92.8%	0.135	0.583	92.8%	93.0%

In order to understand how beneficial it is to apply the rejection method, it was decided to apply a threshold between 0 and 1 to the YOLOv5 model, so that we could fine-tune the final score according to the desired accuracy. Figure 4 shows the relation between rejection rate and accuracy according to the applied threshold. This threshold value is applied at the counting stage and allows only images whose confidence value for detecting each object (from the YOLOv5 model) is higher than this threshold to be considered for counting.

As expected the accuracy increases with the threshold score. The tradeoff is that the number of classified samples becomes smaller. It is important to determine whether this trade-off is beneficial and applicable, according to the task and requirements.

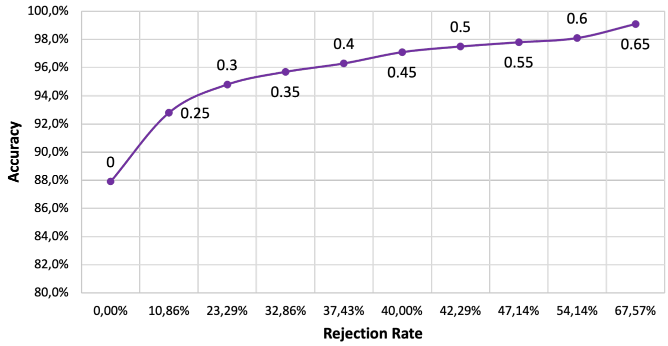

Figure 4: Trade-off between rejection rate and accuracy. The threshold value used in this work was 0.25, which results in an accuracy of 92.8% and a rejection of 10.9%.

4 Conclusions and Future Work

The obtained results are relevant because they highlight the importance of the use of a rejection option, which improves accuracy on the automated reviewed cases while still enabling the reduction of the manual workload. In future work, we will further improve the model's accuracy, by a second counting round of the rejected and misclassified fragments. In addition, we intend to calculate the uncertainty in the fragment count and use it to improve the decision-making of the rejection option, with the aim of rejecting images according to the uncertainty of the count (through a threshold value), making the method more robust and reliable.

Finally, we want to increase the the number of labeled examples in the dataset, since the complexity of WSIs implies that a larger and more diverse dataset leads to better performance of the detection model and, consequently, to a more accurate fragment count.

Acknowledgments

This work was partially supported by national funds provided by Fundação para a Ciência e Tecnologia (FCT), under projects PRELUNA, PTDC/CCI-INF/4703/2021) and UIDB/50021/2020. In addition, IMP Diagnostics also provided funding support for this study. Tomé Albuquerque was supported by Ph.D. grant 2021.05102.BD, provided by FCT.

References

[1] Tomé Albuquerque, Ana Moreira, Beatriz Barros, Diana Montezuma, Sara P. Oliveira, Pedro C. Neto, João Monteiro, Liliana Ribeiro, Sofia Gonçalves, Ana Monteiro, Isabel M. Pinto, and Jaime S. Cardoso. Quality control in digital pathology: Automatic fragment detection and counting. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 588–593, 2022. doi: 10.1109/EMBC48229.2022.9871208.

[2] Neeta Kumar, Ruchika Gupta, and Sanjay Gupta. Whole slide imaging (wsi) in pathology: Current perspectives and future directions. *Journal of Digital Imaging*, 33, 05 2020. doi: 10.1007/s10278-020-00351-z.

[3] Sarah Laroui, Xavier Descombes, Aurélia Vernay, Florent Villiers, François Villalba, and Eric Debreuve. How to define a rejection class based on model learning? In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 569–576, 2021. doi: 10.1109/ICPR48806.2021.9412381.

[4] Ana Beatriz Vieira, Diana Montezuma, Tomé Albuquerque, Liliana Ribeiro, Filipa Rebolo, Domingos Oliveira, João Monteiro, Sofia Gonçalves, Isabel Macedo Pinto, Jaime Cardoso, and Arlindo Oliveira. Automatic detection and counting of pathological specimens with a rejection option. In *35th European Congress of Pathology*, 2023.