Cancer Detection Assistant: Detecting Mitoses in Histopathological Breast Images

Alexander Busch Tolga Akar Jorge Orozco Alexandre Carey
Artus Krohn-Grimberghe Hadi Jomaa

Department of computer science, Universitätsplatz 1, Universität Hildesheim, 31141 Hildesheim, Germany
{buschal,akarto,orozco,careya}@uni-hildesheim.de
artus@lytiq.de
hsjomaa@ismll.de

Abstract

Breast cancer is one of the most prevalent types of disease in the modern days. Because its early diagnosis is imperative to procure an effective treatmeant, many technological and scientific efforts are being put to this task. As a result, object detection and image classification with deep neural networks are being explored. We continue the work by [1], using a similar architecture for patch extraction with optimized heuristics, and complement it with a ResNet18 CNN to perform patch classification. We demonstrate empirically that a superior F1-score is achieved on the ICPR14 dataset compared to the state of the art. The main contributions of this work are reducing the regions of interest of the patch generator, using a pre-trained ResNet18 model instead of a custom CNN improving the total F1 score of the model, minimizing error types, and providing masks for the areas of interest that can be used in the field of semantic segmentation through our heuristics.

1 Introduction

The first step towards breast cancer diagnosis is to perform a biopsy. In this process, a small breast tissue is removed and then analyzed by a medical expert to detect abnormal cell formations [2]. More accurately, a stained high-power field image (HPF, around 4500x4500 pixels) is explored to count cells undergoing mitosis as this is the main indicator of the presence of cancer.

The analysis process is challenging for the experts for multiple reasons. First, mitoses are only a couple of pixels wide. Because of the organic nature of the biopsy, the images tend to be noisy and irregular, creating a bigger challenge for the pathologists. Added to this fact, in some hospitals the specialists must cover a great number of patients during their workday. They must be able to provide a diagnosis in under 4 minutes per HPF on average if they want to go through all their patients. This challenging process of analyzing a high-resolution image is performed manually and under tight time constraints, making it prone to human error. Furthermore, a correct diagnosis is imperative. In other words, the results have a very low tolerance to mistakes, either for false positives or false negatives. Our goal is to help pathologists during the identification stage of this process, not to replace them. We believe that a decent machine learning algorithm (utilizing computer vision) can lower the burden of pathologists by proposing them candidate cells that could be mitotic. Naturally, the final say whether those proposals are actually mitotic or non-mitotic will belong to the specialists.

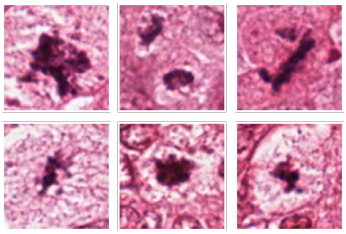


Figure 1: Comparison of mitotic (top), and non-mitotic (bottom) cells. Reprinted from [3].

In general, computer vision and object detection have come a long way since early 2010s, after the introduction of convolutional neural networks (CNNs) into this field. With models such as AlexNet[4], ResNet[5], YOLO[6] or RetinaNet[7] have achieved impressive results in traditional object detection tasks using variety

of different CNN architectures. However, this is not yet the case in some types of biomedical images, as in our case. Unlike the medical images we use for mitotic cell detection, traditional objection detection tries to locate much bigger and distinct objects, compared to a tiny cell on a stained histopathological image. For untrained individuals, it is almost impossible to identify which cell is mitotic or not, whereas even toddlers can point out a coffee mug or a bottle on an image without breaking a sweat. Figure 1 shows an example of the complex task that is to properly identify mitotic from non-mitotic cells.

In order to use these techniques for cancer detection, the HPF images firstly have to be sliced to more manageable sizes that can be then processed. Then, the object detection algorithm should first identify possible objects (proposals), that are then fed to the classifier to select mitotic from non-mitotic. These proposals tend to be numerous due to the nature of the images. To solve these problems, we propose the Cancer Detection Assistant (CDA) algorithm, which combines the use of heuristics to select proposals, and a ResNet18 architecture to perform the classification. The contributions of this study can be presented fourfold:

- CDA reduces the amount of regions of interest, from 240 to 44, by applying a novel set of heuristics before passing to the classifier, as compared to the work by [1]
- The algorithm presented uses a CNN architecture that has achieved great results in the field of object detection, as opposed to the custom architecture used in [1]
- Using the proposed architecture, the F1-score is improved to **0.602** on the ICRP14 dataset.
- Our heuristics provide masks that can be used as ground-truth for future work in the field of semantic segmentation.

This paper is presented as follow: Section 2 provides an overview of the explored literature. Then, Section 3 shows the explanation of our proposed CDA algorithm. Section 5 contains the experimental results achieved, and finally a discussion of the presented architecture, possible future work, and conclusion are shown is sections 6 and 7.

2 Related Work

A lot of previous work has investigated the automatic detection of mitosis in large and heavy histopathological images. There are two main types of detection: the detection based on handcrafted features([8],[9],[10],[11],[12],[13],[14],[15],[16]) and the one based on CNN features.

Handcrafted features approach are used a lot for mitotic cells detection and medical area because the appearance of cells can be described with shape, statistics and texture. These features can be defined by a pathologist or using the statistics given by the analysis of mitotic cells. In the feature based detection, [17] designed a model using a deep neural network and handcrafted features. They focused on the morphological, textural and intensity features to build their model. They showed that 71x71pixels is the right size for capturing fully a mitotic cell. Their method allows to detect mitotic patches from whole slide HPF images.

CNN features are proven to be a robust way to automatically learn the appearance of the mitotic cells. The drawback of the CNN features methods are the heavy computation and the long training time that they require. IDSIA[18] considered mitosis detection as pixel classification problem. They applied a plain CNN which acts as a sliding window. Their method won the ICPR12 contest and achieved a F1-score of 0.782 on ICPR12. The difficulty of their method is its application in clinic environment due to its intensive computation. [19] applied a Deep Cascaded Network for mitosis detection. Their method was built on two CNNs: a Coarse Retrieval Model based on a Fully Connected Network(FCN) which scans the whole slide image for retrieving the region of interest(ROI) and a Fine Discrimination Model based on a pre-trained CNN which categorizes cells as mitotic or non-mitotic. The two networks were trained independently and this disturbs their combination together. They extended their work by training 3 different Fine Discrimination Model with different sizes in order to average the output. They achieved a F1-score of 0.788 for MITOS(ICPR12) and 0.482 for ICPR14(on the test set provided by the challenge) with their model averaging method. The other improvement was the computational time which was about 0.5 seconds for both networks and was 60x faster than the state of the art on ICPR12[18].

[20] proposed an architecture with different parts: they designed a FCN for improving the given center-based annotations; they implemented a detection network which combines a Region Proposal Network and a Region-Based Classifier for detecting mitosis using features information. They created a detection network as a last part of their method to remove false positive. Their method achieved a F1-score of 0.832 on ICPR12 and 0.437 on ICPR14 (on the test set provided by the challenge).

[1] proposed an interesting approach using Haar wavelet decomposition and a deep CNN for cell classification. The utilization of this wavelet decomposition

reduced the convolution time of the network by reducing the size and resolution of the images. This method achieved a F1-score of 0.84 on MITOS(ICPR12) and 0.560 on ICPR14(on the test set provided by the challenge). Some other recent methods used CNN features method to apply on this task([21][22]). [23] proposed the most recent framework (PartMitosis) which is based on two deep FCNs. The first FCN is trained using weak labels while the other one is trained using strong labels. The output of the two networks are fused to acquire the mitosis detection. Their method achieved a F1-score of 0.575 on ICPR14(on the test set provided by the challenge). The Chinese University of Hong Kong(CUHK) won the MITOS-ATYPIA-14 contest with a F1-score of 0.356 but their paper could not be found.

3 Methodology

Our approach is based on the promising work of [1]. The main steps of their method include stain normalization, heuristics for proposal generation, patch extraction of 81x81 px crops, and classification. However, we optimized the set of heuristics to detect the candidate cells, omitted the decomposition, and intensified the CNN (see Figure 2). Before initiating our networks, we transformed the format of the used data into COCO format [24]. This allows a standardized procedure of using further components (see Experiments section).

3.1 Proposal generator and mask generator

The heuristics can be split into two parts, proposal generator and on top mask generator. First, the original image is Macenko-stain-normalized [25]. Stain normalization is useful to level out the natural variations during the manual staining process. Manual processes are prone to slight differences in amount and application of the stain solution, lighting, and storage time. The Macenko method is a singular value decomposition (SVD) geodesic based method. Staining with Hematoxylin & Eosin (H&E) aims to change the color of nuclei in dark purple (Hematoxylin) and the surrounding plasma in light colors (Eosin). Each RGB-value of a pixel can then be mapped into the H and E components (see equation 1). For Macenko, the 2 values resulting from the SVD represent the H and E values the transformation-matrix S maps the image to. Estimating this matrix and normalizing the H and E dimensions is how stain normalization is applied [25].

$$S = \begin{pmatrix} H_R & H_G & H_B \\ E_R & E_G & E_B \end{pmatrix} \tag{1}$$

After stain normalization, transformation into a single-channel blue ratio image is performed, because

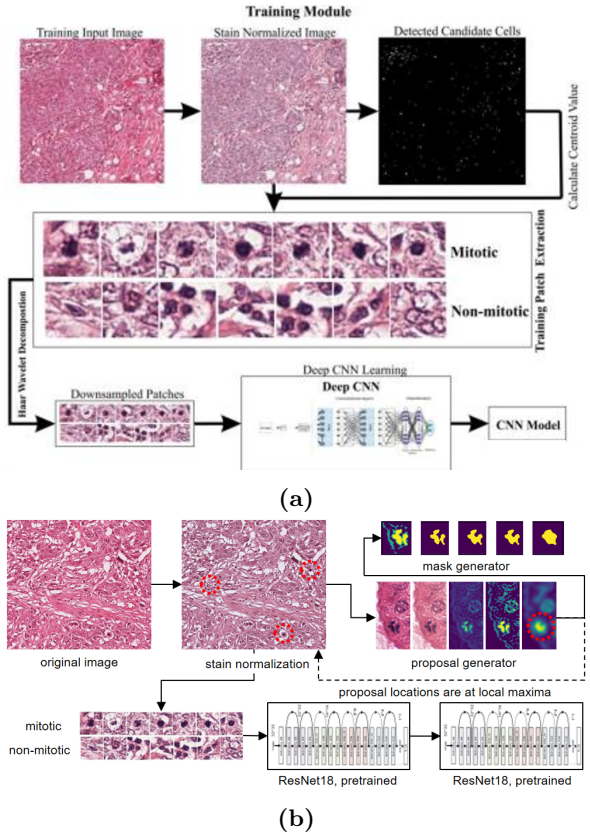


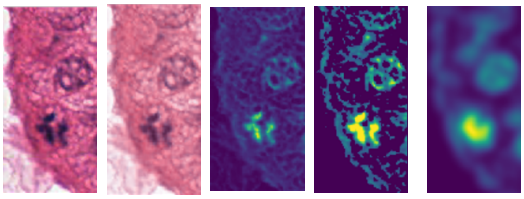
Figure 2: (2a: pipeline from the training image, over patch creation, to a fully trained CNN network. Reprinted from [1]. (2b): Our modifications focusing on the heuristics for detecting candidate cells and the network. (The additional mask generator is not displayed because it belongs to future work and not included in the pipeline)

mitotic cells are found to be highly visible in blue ratio images [10, 26]. The equation for each RGB pixel is:

$$blue_ratio = \frac{100 * B}{1 + R + G} * \frac{256}{1 + R + G + B} \tag{2}$$

Next, Otsu's thresholding [26] based on the blue ratio histogram is performed. This specific thresholding analyzes the histogram of the new height x width x 1 pixel values and cuts at an optimal count. This adaptive thresholding is beneficial for differences between images. For increasing the robustness, the values are cut off at half and double the threshold value. This way outlier values are eliminated, an increased certainty in maintaining all relevant pixels is introduced, and the mitotic regions are kept large for being emphasized in the successive smoothing.
The multi-stage Gaussian smoothing is applied. The smoothing then highlights heavily stained and large regions based on the amount and intensity of the surrounding pixels. The resulting peak value of that re-

gion represents the amount, the intensity, and the center of the region. All local maximum values are ultimately interpreted as centers of regions of interest. The smoothing and locating of the local maxima is one of our extensions to the approach of [1].



(a) (b) (c) (d) (e)

Figure 3: Main phases of the proposal generating heuristics: (3a) is a sample crop including a mitotic region, (3b) is the Macenko stain normalization crop, (3c) is the blue ration image, (3d) is the Otsu's thresholding and normalization sample, (3e) is the Gaussian smoothing sample.

Each 71x71 px region around those local maxima is then individually smoothed and thresholded again for a more specific separation of the pixels for creating a mask. The resulting mask is filled and smaller unconnected pixel values are erased. This is how the mask generator creates a fitting mask for each proposal (see Figure 4). The mask is not necessary for an image classifier nor for our pipeline, but may be used as ground truth for an object detector or for further exploratory methods.

(a) (b) (c)

Figure 4: Main phases of the mask generating heuristics: (4a) applies Gaussian smoothing to the output of (3d), (4b) applies individual Otsu's thresholding per crop, (4c) is the dilation and deletion of small unconnected mask blobs.

We decided to extract patches with the size of 71x71 px, whereas [1] used 81x81 px. Investigations showed that the cells fit into 71x71 px and 81x81 px thus only is advantageous if downsized. While not downsizing, we trade the prevented information loss against possible noise reduction by the Haar Wavelet decomposition.

3.2 Classification network

[1] fed the downsized patches into a custom built CNN. We use a ResNet18 [5] pre-trained on ImageNet [27] data and apply transfer learning. This ResNet18 consists of 5 convolutional stages and a head network. The head network consists of average pooling, a fully connected feed forward network, and a softmax layer. The convolutional stages contain convolutional layers, ReLu activations, and a skip connection. Greater ResNets showed only slightly better results and slowed down the training and experimenting steps. This is the reason why we decided to maintain this architecture. This also indicates that for this kind of CNN, the complexity of the task is solvable and further advancement lay in the data. It is interesting to notice that a pre-trained network on everyday objects performs better than one with random initialization. Further details about the training process can be found in the Experimental Results section.

4 Data

4.1 Dataset

Many public datasets are available ([28, 3, 29, 30]) and a lot of work have investigated these datasets achieving interesting results. Regarding the dataset complexity, TUPAC16 contains more samples than the ICPR14 dataset but the annotations were less precise. ICPR14 stays the most challenging and more complicate dataset to cope with: this is why we decided our research to focus on it.
The ICPR14 breast cancer dataset is a public dataset used in the Mitos Atypia 14 Grand Challenge [3]. It contains 1136 histopathological images at 40x magnification ($377.8245 \times 337.808\ \mu m^2$) with mitosis centroid annotations for each one of them. The annotations were made by 3 pathologists using majority voting, the cells were annotated as mitotic or non-mitotic. These biopsy images are scanned by two scanners, namely Aperio Scanscope XT and Hamamatsu Nanozoomer 2.0-HT. To reduce the complexity of the task, we decided to use Aperio scanner samples. There are multiple mitosis in slides and the whole dataset contains 749 mitosis in total. The ICPR14 dataset, even though it does not contain much images, is a quality dataset: the work on different public datasets showed us that ICPR14 annotations are precise while other datasets might be less precise and error prone. Moreover, ICPR14 is used in many papers which allows us to compare our results with other methods.

4.2 Data Preprocessing

Figure (5a) shows how close are the centroids of proposals and the ground truth values. The graph indicates that our proposals are quite accurate compared to the actual annotations, majority of them being in range of 10 pixels away. In our inspections, we noticed that values bigger than 12 pixels can be reduced to non-centric annotations, i.e. the annotations that are

Table 1: Dataset before and after applying preprocessing and heuristics

# of	Before slides	After mitotic patches	After background patches
Training Set	784	522	21,381
Validation Set	416	227	13,911
Test Set	496	203	11,947
Resolution	1539x1376 px	71x71 px	71x71 px
Total	**1695**	**952**	**47,239**

not exactly pointing the cell nuclei as it should be.

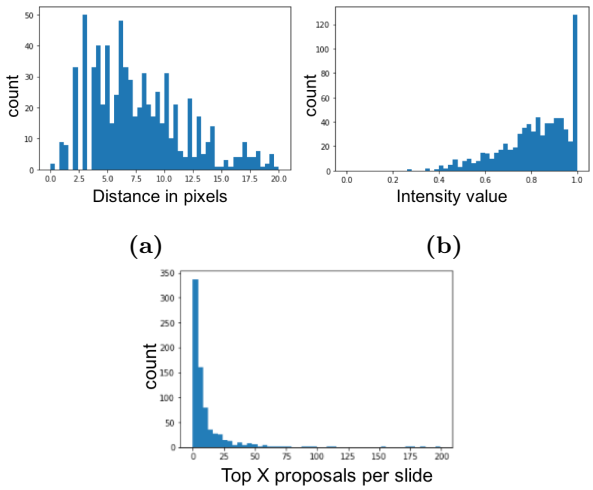


Figure 5: (5a): distance between ground truth mitosis and closest proposal in pixel. (5b): intensity of mitoses. (5c): number of mitoses in top X proposals.

Figure (5b) contains the intensities the local maximums of each pixel. Many proposals share the last bin of the histogram. In other words, they are often close to 1.0, which is an expected result considering the processing of the original image. Figure (5c) demonstrates at what position a mitosis is located in a descending sorted array of the local maximum pixel values. This is a rather important graph, since it proves that 95% of the mitoses lay within the top 44 proposals. We re-implemented the patch extraction of [1]'s approach. and found that it extracts 240 patches on average. The number of proposals in our method is significantly lower than their approach [1].

Table 1 shows that our dataset is highly imbalanced, therefore we decided to over sample the mitotic patches to balance the classes. With over sampling the positive samples, the size of our training set is almost doubled, increasing to 42,762 total patches. Balancing the classes is vital for our case, since we would not prefer our model to be biased against making mitotic predictions. In addition to over sampling the imbalanced class, we employed several probabilistic data augmentations to further enlarge our training set. These random transformations are up to 360° rotation, horizontal and vertical flipping, slight changes in lighting, zooming, and symmetric warping.

5 Experimental Results

In this section, we describe our experiments on the ICPR14 dataset. Our method was applied using the following setup. First, we apply our optimized heuristics to the data and extract patches with the size of 71x71 px. Then, these candidate patches are used to train our deep CNN model. Through our experiments, we prove that our proposed model with optimized heuristics can outperform the current state-of-the-art by achieving a higher score in significantly fewer number of extracted patches.

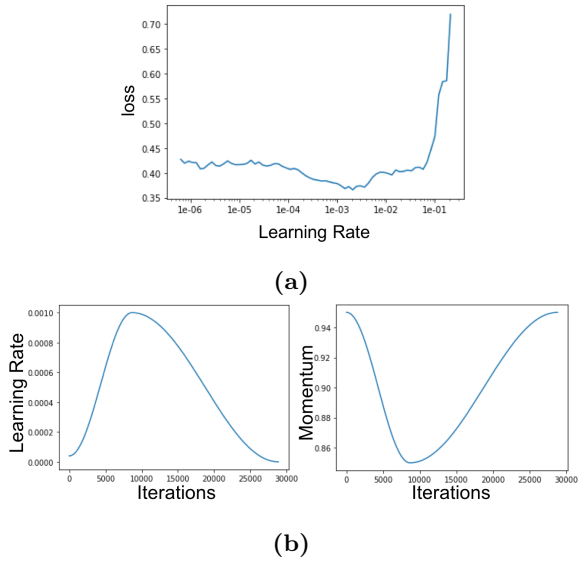


Figure 6: (6a): result of the learning rate finder. (6b): changes of the learning rate and the momentum during the training.

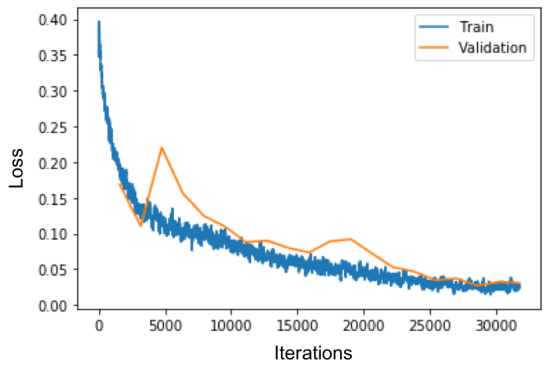
5.1 Hyper-parameter Tuning

To find the best set of hyper-parameters, we trained our model using the training set and evaluated on the validation set. After finding the optimal hyper-parameters, we combined the training and validation sets to train our model with the optimal hyper-parameters, and then evaluate it on the test set. This approach ensures that the official test set is not a part of the training process in any way.

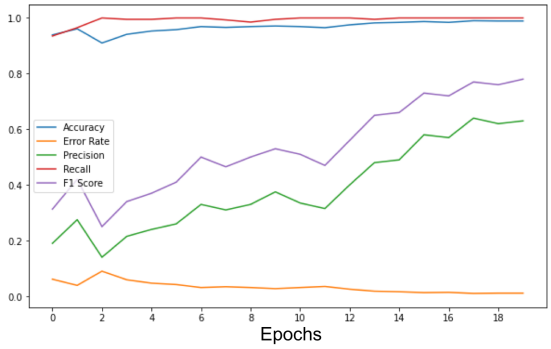
The first step in our learning phase is to train only the last layers of the pre-trained ResNet18 model. Even though one may think that using a model pre-trained on ImageNet[27] dataset may not be helpful

for biomedical computer vision tasks, we found out that using transfer learning helps increasing the performance of our model. As a benchmark, we trained our model from scratch and achieved approximately 0.05 lower F1-score. The reason behind this is that our model does not need to learn low level features (e.g. simple lines and shapes) from scratch. Our initial training with the frozen model converged after 2 epochs, and achieved 0.125 F1-Score, which is far below a desirable outcome. To increase the performance, we unfroze the model and ran a learning rate finder that mocks a training to find an optimal learning rate range. According to the learning rate graph (Figure 6a), the loss curve is steepest between 10^{-3} and 10^{-4}, therefore we chose that range as our learning rate slice for the one-cycle policy [31]. The change in the learning rate during the training can be observed in (Figure 6b). The momentum behaves exactly the opposite of learning rate, i.e. increases when the learning rate is decreasing and vice versa. The momentum values in the optimal training protocol are in between 0.85 and 0.95. We used dropout rate of 0.5 as well during training to prevent over-fitting. With these hyper-parameter setup, our model converged after 30 epochs, achieving more or less 0.5 F1-score on the validation set.

Once we have found out the best training protocol, we started our actual learning phase from the beginning with the combined dataset. Following the same steps described above, we achieved 0.99 accuracy, 0.65 precision, 1.0 recall and nearly 0.8 F1-Score at the end of the training, converging after 20 epochs. On the test dataset, our model outperformed the state-of-the-art[1] by achieving **0.626 recall** and **0.586 F1-score**. By default, the decision threshold is 0.5 for predicting positive class. Naturally, changing this value will cause a change in the number of true and false predictions, directly affecting the F1-score. The aim of this study is to maximize the F1-score, so we evaluated how the decision threshold should be shifted to improve the results. After several trials on the validation set, we empirically found out that increasing this threshold improves the F1-score, due to returning more background predictions. However, this shift in the threshold causes recall to drop (which could be undesirable for such a sensitive task) for the same reason. It must be noted that this value is very sensitive to the trained weights and the provided data. The final result of our proposed model with 0.8 prediction threshold on the ICPR14 test set shows a **0.602 F1-score**, further improving the performance. To the best of our knowledge, this is the highest F1-score achieved on the official test set of the ICPR14 dataset. The comparison of the proposed method and other works that are tested on official ICPR14 test dataset can be seen on Table 2.



(a)



(b)

Figure 7: (7a): convergence of the final model. (7b): evaluation metrics (The increase in performance is due to integrating the validation set to the training set.).

As an alternative to our main approach, we tried using Mask R-CNN [32] as a region proposal network, instead of using the optimized heuristics. However, we achieved inferior performance compared to our main methodology, therefore we decided to exclude it from our work.

We implemented our code using PyTorch and the fast.ai library. We ran our experiments on the virtual machines provided by Google Colaboratory. The final training (i.e. train and validation combined dataset, 20 epochs) took less than 2 hours to finish using the NVIDIA Tesla T4 GPU provided by Google Colaboratory in our session, which has 16GB GDDR6, 300 GB/sec GPU memory and 2560 CUDA® cores. The inference time of the model is 9 seconds on average with the same technical specifications.

6 Discussion and Future Work

One of the first approaches that we took was to try to replace the heuristics with a network layer that could learn how to identify proposals. One of the approaches used was Mask-RCNN [32] due to its flexibility. Still, our experiments showed that the modified heuristics

Table 2: Performance comparison of the reported methods on the testing dataset of MITOSIS-ATYPIA-14 Contest

Model	Precision	Recall	F1-Score
CUHK	0.448	0.300	0.356
DeepMitosis[20]	0.431	0.443	0.437
Deep cascade net[19]	0.411	0.478	0.442
CNN with Haar decomposition[1]	0.544	0.576	0.560
SegMitos-random[22]	0.637	0.502	0.562
PartMitosis[23]	**0.664**	0.507	0.575
Proposed CDA model (0.5 threshold)	0.558	**0.626**	**0.586**
Proposed CDA model (0.8 threshold)	0.629	0.576	**0.602**

obtained more accurate proposals for the images.

On the other side, we found that the annotations for the data sometimes did not match the actual mitoses. It is to our understanding, that the specialists annotating the data would place a pixel mark as reference to find the mitotic cell. But sometimes these annotations would be off of the actual mitosis by some pixels. The mark would be sufficient for the human eye to quickly identify the position of a mitosis, but it is sub-optimal for an automated process that need the exact position of the cell.

When evaluating possible architectures for the classifier, other bigger depths of the ResNet models (ResNet34, ResNet50) were tested . Our experiments showed that increasing the depth did not yield to a significant improvement of the metrics, but only raised the execution time and complexity. In the end, we decided to keep the simplest architecture (ResNet18).

As for future work, the images in the dataset that were used for this study present only cropped sections from a complete HPF image. When trying to handle complete microscopy images, (such as MRSX formatted files generated from a microscope), more adjustments to the images may be needed to account for different types of cellular tissue, distinct colors, or irregular shapes of images. This work can be further expanded to account for such elements.

Also, different datasets account for various conditions in which the images are taken (lighting, color, etc.) The proposed architecture could be trained combining multiple datasets, under many schemes: as transfer learning, combining them all in the training set, etc. It may be possible to create a more robust model through this means.

On the other hand, a more thorough look can be taken into the image staining workflow. Normally, the staining of tissues is made with hematoxylin and eosin (H&E) because it is easier for the human eye to identify possible mitotic cells. Other dyes could be more beneficial to machine learning models, as they may expose information not perceptible to the human eye

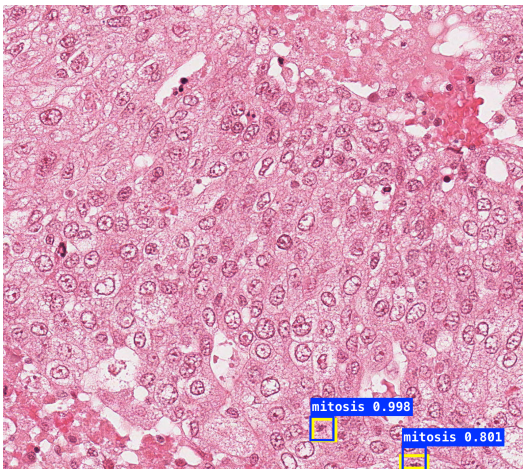


Figure 8: Predictions (blue) and ground truth (yellow) on a sample image.

7 Conclusion

In this paper we presented CDA, a novel way to combine image modification heuristics and deep architectures to achieve state of the art results in the task of object detection. The algorithm successfully reduces the number of proposal areas, incorporates a widely used CNN architecture in object detection, and outperforms the state of the art in mitosis detection. Moreover, the heuristics used produce masks that can be used as ground truth for semantic segmentation tasks.

In order to further contribute to the community dealing with this important problem of cancer detection, all the code to reproduce the experiments, including graphs, and a Jupyter Notebook with a simple service to get an annotated image will be made available on GitHub under `https://github.com/lostensores` TBD.

7.1 Acknowledgements

The working team would like to give their thanks to Prof. Dr. Artus Krohn-Grimberghe for his guidance, to Dr. Ludovic Roux for providing the ICPR14 test set annotations, and the staff at the Information Systems and Machine Learning Lab (ISMLL) from the University of Hildesheim for making this work possible.

References

[1] D. K. Das and P. K. Dutta, "Efficient automated detection of mitotic cells from breast histological images using deep convolution neutral network with wavelet decomposed patches," *Computers in biology and medicine*, vol. 104, pp. 29–42, 2019.

[2] A. S. of Clinical Oncology (ASCO), "Biopsy for diagnosing cancer." `https:`

//www.cancer.net/navigating-cancer-care/diagnosing-cancer/tests-and-procedures/biopsy. Accessed: 11.03.2019.

[3] "Mitos-atypia-14 contest home page." https://mitos-atypia-14.grand-challenge.org/. Accessed: 2020-03-11.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, 2016.

[7] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988, 2017.

[8] M. Veta, P. J. van Diest, S. M. Willems, H. Wang, A. Madabhushi, A. Cruz-Roa, F. Gonzalez, A. B. Larsen, J. S. Vestergaard, A. B. Dahl, D. C. Cireşan, J. Schmidhuber, A. Giusti, L. M. Gambardella, F. B. Tek, T. Walter, C. W. Wang, S. Kondo, B. J. Matuszewski, F. Precioso, V. Snell, J. Kittler, T. E. de Campos, A. M. Khan, N. M. Rajpoot, E. Arkoumani, M. M. Lacle, M. A. Viergever, and J. P. Pluim, "Assessment of algorithms for mitosis detection in breast cancer histopathology images," *Medical Image Analysis*, vol. 20, no. 1, pp. 237–248, 2015.

[9] A. M. Khan, H. El-Daly, and N. M. Rajpoot, "A Gamma-Gaussian mixture model for detection of mitotic cells in breast cancer histopathology images," *Proceedings - International Conference on Pattern Recognition*, no. Icpr, pp. 149–152, 2012.

[10] H. Wang, A. Cruz-Roa, A. Basavanhally, H. Gilmore, N. Shih, M. Feldman, J. Tomaszewski, F. Gonzalez, and A. Madabhushi, "Cascaded ensemble of convolutional neural networks and handcrafted features for mitosis detection," in *Medical Imaging 2014: Digital Pathology*, vol. 9041, p. 90410B, International Society for Optics and Photonics, 2014.

[11] C. Sommer, L. Fiaschi, F. A. Hamprecht, and D. W. Gerlich, "Learning-based mitotic cell detection in histopathological images," *Proceedings - International Conference on Pattern Recognition*, no. Icpr, pp. 2306–2309, 2012.

[12] C. H. Huang and H. K. Lee, "Automated mitosis detection based on eXclusive Independent Component Analysis," *Proceedings - International Conference on Pattern Recognition*, no. Icpr, pp. 1856–1859, 2012.

[13] C. Malon and E. Cosatto, "Classification of mitotic figures with convolutional neural networks and seeded blob features," *Journal of Pathology Informatics*, vol. 4, no. 1, p. 9, 2013.

[14] H. Irshad, "Automated mitosis detection in histopathology using morphological and multichannel statistics features," *Journal of Pathology Informatics*, vol. 4, no. 1, p. 10, 2013.

[15] F. Tek, "Mitosis detection using generic features and an ensemble of cascade adaboosts," *Journal of Pathology Informatics*, vol. 4, no. 1, p. 12, 2013.

[16] A. Paul and D. P. Mukherjee, "Mitosis detection for invasive breast cancer grading in histopathological images," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 4041–4054, 2015.

[17] M. Saha, C. Chakraborty, and D. Racoceanu, "Efficient deep learning model for mitosis detection using breast histopathology images," *Computerized Medical Imaging and Graphics*, vol. 64, pp. 29–40, 2018.

[18] D. C. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images," *International Conference on Medical Image Computing and Computer-assisted Intervention. Springer, Berlin, Heidelberg*, pp. 411–418, 2013.

[19] H. Chen, Q. Dou, X. Wang, J. Qin, and P. A. Heng, "Mitosis detection in breast cancer histology images via deep cascaded networks," *30th AAAI Conference on Artificial Intelligence, AAAI 2016*, pp. 1160–1166, 2016.

[20] C. Li, X. Wang, W. Liu, and L. J. Latecki, "Deepmitosis: Mitosis detection via deep detection, verification and segmentation networks," *Medical Image Analysis*, vol. 45, pp. 121–133, 2018.

[21] D. Cai, X. Sun, N. Zhou, X. Han, and J. Yao, "Efficient mitosis detection in breast cancer histology images by rcnn," in *2019 IEEE 16th International*

Symposium on Biomedical Imaging (ISBI 2019), pp. 919–922, April 2019.

[22] C. Li, X. Wang, W. Liu, L. J. Latecki, B. Wang, and J. Huang, "Weakly supervised mitosis detection in breast histopathology images using concentric loss," *Medical Image Analysis*, vol. 53, pp. 165–178, 2019.

[23] M. Sebai, T. Wang, and S. A. Al-Fadhli, "PartMitosis: A Partially Supervised Deep Learning Framework for Mitosis Detection in Breast Cancer Histopathology Images," *IEEE Access*, vol. 8, pp. 45133–45147, 2020.

[24] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*, pp. 740–755, Springer, 2014.

[25] M. Macenko, M. Niethammer, J. S. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt, and N. E. Thomas, "A method for normalizing histology slides for quantitative analysis," in *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 1107–1110, IEEE, 2009.

[26] R. C. Gonzales and R. E. Woods, "Digital image processing," 2002.

[27] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, Ieee, 2009.

[28] "Mitosis Detection in Breast Cancer Histological Images (MITOS dataset)." `http://ludo17.free.fr/mitos_2012/index.html`.

[29] "Classification of mitosis from the AMIDA13 dataset." `https://ismi-amida13.grand-challenge.org/Introduction/`.

[30] "Tumor Proliferation Assessment Challenge 2016." `http://tupac.tue-image.nl/node/3`.

[31] L. N. Smith, "Cyclical learning rates for training neural networks," in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 464–472, IEEE, 2017.

[32] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017.