



HAL
open science

Mitosis Detection in Breast Cancer with Deep Learning: A New Approach

Abdelwahhab Boudjelal, Abderrahim Elmoataz, Youssef Chahir

► To cite this version:

Abdelwahhab Boudjelal, Abderrahim Elmoataz, Youssef Chahir. Mitosis Detection in Breast Cancer with Deep Learning: A New Approach. 2022 International Conference of Advanced Technology in Electronic and Electrical Engineering (ICATEEE), Nov 2022, M'sila, Algeria. <hal-04980825>

HAL Id: hal-04980825

<https://hal.science/hal-04980825v1>

Submitted on 6 Mar 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Mitosis Detection in Breast Cancer with Deep Learning: A New Approach

Abdelwahhab Boudjelal
Normandie Univ, UNICAEN, ENSICAEN,
CNRS, GREYC, 14000 Caen, France
Electrical Engineering Laboratory LGE
M'sila University, 28000 M'sila
abdelwahhab.boudjelal@unicaen.fr

Abderrahim Elmoataz
Normandie Univ, UNICAEN,
ENSICAEN, CNRS, GREYC
14000 Caen, France
abderrahim.elmoataz-billah@unicaen.fr

Youssef Chahir
Normandie Univ, UNICAEN,
ENSICAEN, CNRS, GREYC
14000 Caen, France
youssef.chahir@unicaen.fr

Abstract—In this work, we propose a new approach for spotting mitoses in breast cancer histology images. This new approach involves integrating two publicly accessible datasets following a normalization procedure of color. The mitotic samples are then enhanced by preserving the context to address class imbalance. After this, the candidate mitotic cells are classified into the target classes using a ResNet classifier. Through this method, we were able to accurately identify mitosis in the combined dataset while attempting to identify it in the images. We demonstrate that our method outperforms all current methods by comparing it to state-of-the-art methods using a public dataset. Our results indicate that the proposed technique can be used to automatically identify mitotic cells in the images of histopathology of breast cancer.

Index Terms—histopathology, ResNet-50, mitosis detection

I. INTRODUCTION

The most frequent malignancy in women is breast cancer, with regard to about 2685,000 deaths and .3 million new cases worldwide in 2020 [1]. While most breast cancers can be cured if detected early, the current gold standard for breast cancer screening, mammography, suffers from low sensitivity, particularly in women with dense breast tissue.

Digital pathology, which uses whole slide images (WSIs) to digitize tissue specimens, has emerged as a promising alternative to traditional microscopy for cancer diagnosis. WSIs of breast cancer can be used to generate mitosis annotations, which are associated with tumor grade and have been shown to be a strong predictor of patient outcomes. However, the task of mitosis detection in breast cancer histologic images is a challenging one, due to the large size and high resolution of WSIs, the large variation in mitosis appearance, and the presence of artifacts such as staining and background clutter.

A number of mitosis detection techniques have been proposed, but most of these techniques are based on heuristic rules that are designed by expert pathologists. These methods are therefore both slow and inaccurate, and often require manual intervention.

Recently, there has been a growing interest in the use of deep learning for the automated detection of mitotic cells in histopathological images [2]–[6]. Deep learning is a powerful tool for automated image analysis and has been shown to be effective for a number of applications such as object detection

[7], classification of image [8], and image segmentation. Cell-level CNNs have been used in automated pathology activities such as mitosis detection and cell nucleus detection [9].

In order to train deep learning based object detection models, a large number of images with annotated object locations are required. However, the collection of such a large quantity of images is extremely expensive and takes a lot of time. Additionally, it is very difficult to annotate images with high accuracy. Thus, how to train deep learning based object detection models with limited annotated images is a challenging problem. Data augmentation is a common approach to train deep learning based object detection models. It consists of applying a set of operations to the training images to create new images. Common data augmentation operations include cropping, rotation, and horizontal flipping. Such operations introduce appearance and spatial variability in the generated images, thereby improving the robustness of the trained deep learning based object detection models. However, the generated images have different appearances and locations because these data augmentation operations do not preserve the object's context.

This work proposes a deep learning based mitosis detection method that is compared with state of the art deep learning and traditional methods. The proposed approach uses a pre-trained ResNet-50 deep convolution architecture and is based on combined datasets.

The paper is structured as follows: First, the proposed approach is explained in Section II, The results of the experiments are shown in Section III, finally the paper is concluded with a plan for further research in Section IV.

II. METHODS

A. Dataset

Both the publicly available image databases MITOS-ATYPIA-14 and ICPR-MITOS 2012 can be used for academic studies. The 2012 ICPR competition saw the debut of the MITOSIS dataset, which includes 50 RGB images, 15 of which are for testing and 35 for training. Each high-power field (HPF) has an area of 512 by 512 per micrometer, suggesting that the image size is 2084 by 2084 per micrometer. Each image has a resolution of 0.2456 micrometers per pixel.

The slides were scanned using two scanners, the Aperio XT (scanner A) and the Scanner of Hamamatsu (scanner H). The dataset of MITOS-ATYPIA-14 from ICPR 2014 was used in a competition in which investigators were asked to assess nuclear atypia and determine the mitotic cells number. The dataset consists of 496 test images obtained from five different breast biopsies and 1,200 training images obtained from sixteen different biopsies. Each image is 1539 by 1376 pixels in size at 40 by 40 magnification, and only the centroid pixels of each mitosis were annotated by the pathologists. It is difficult to obtain remarkable performance with this dataset due to the large differences in the images in the dataset, which depend on the tissue acquisition procedure, staining, and lighting conditions.

B. PROPOSED METHOD

The proposed method for detecting mitosis in images uses color normalization, data augmentation, training and evaluation of a convolutional neural network (CNN), and testing of the algorithm. Figure 1 provides a broad overview of the proposed technique. The four key divisions of the proposed process are color normalization of the image, data augmentation, training and evaluation of the CNN, and testing of the algorithm.

1) Image color normalization:

Histopathology images often contain irregular staining, making it difficult to learn the features required for mitosis detection by DCNN. To solve this problem, we propose a fast and efficient approach for color normalization of the image. We first extract the R, G, and B color channels from the input image I , and then calculate standard deviation and the mean of each channel:

$$\mu_c = \frac{1}{N} \sum_{i=1}^N I_{i,c}, \quad \sigma_c = \sqrt{\frac{1}{N} \sum_{i=1}^N (I_{i,c} - \mu_c)^2}, \quad (1)$$

where N is the pixels number in the image, $c \in \{R, G, B\}$ is the color channel, and $I_{i,c}$ is the intensity of pixel i in channel c . We then standardize the channels using the following transform:

$$I'_{i,c} = \frac{I_{i,c} - \mu_c}{\sigma_c}. \quad (2)$$

The transformed image I' is then augmented and fed into the DCNN.

The color normalization of images from two distinct datasets, ICPR-MITOS 2012 and MITOS-ATYPIA-14, is shown in Figure 2 (a and b). There is a clear difference in the color properties of the images obtained using these two datasets. As seen in Figure 2, color homogeneity is the result of the color normalization of these images (b and c). In contrast to the original images, the output images have an even color scheme.

2) Data Augmentation

The proposed strategy for increasing the number of training samples and reducing the risk of overfitting is to use data augmentation. The augmentation Data is a commonly used strategy to enhance the performance of the convolutional neural network. The proposed method uses context preserving data augmentation to maintain the semantic meaning of the patches. The algorithm from the original image is utilized to retrieve patches from the original image. The boundary pixels are then generated to generate a new patch.

The number of generated patches is controlled by a parameter n . The number of generated patches is given by,

$$N = n^2 \quad (3)$$

The parameter n is chosen to be 2 in the proposed method. The generated patches are then augmented with rotations and flips to obtain the final training set.

3) DCNN Model

We use the ResNet-50 [10] DCNN model for cells classification. ResNet is a popular DCNN architecture that has shown success on a variety of image classification and detection tasks.

There are a few reasons why the ResNet-50 CNN model is a good choice for mitosis detection. First, the model is pre-trained on a large dataset (ImageNet), which means that it already has a good understanding of basic image features. This means that the model can more quickly learn to detect mitosis in new images. Second, the model is relatively shallow, which means that it is less likely to overfit on the relatively small mitosis dataset. Finally, the model has a good reputation for generalizability, meaning that it is likely to perform well on unseen data.

The trained model is what we use. A patch measuring 64 by 64 pixels serves as the network's input. We use the Adam optimizer [11], with a learning rate of 0.0001, to adjust the model on the training set. Training is done over 25 epochs with a batch size of 16.

4) Candidate cell detection

The normalized images are converted to binary images using Otsu's method which is computed separately for each channel. The binary images are then passed through a series of morphological operations to remove the background noise and obtain the cell regions. The operations performed are opening with a disk of radius of 5 pixels, closing with a disk of radius 10 pixels, and opening with a disk of radius 5 pixels. The binary image obtained after all three operations is inverted to obtain the cell regions. The centroid of each cell is found via linked component labeling.

III. EXPERIMENTAL AND RESULTS

A. Experimental Setup

The proposed technique is implemented using Python 3.5. This study uses two Intel Xeon-E5-2620-v4 with 2.1 GHz

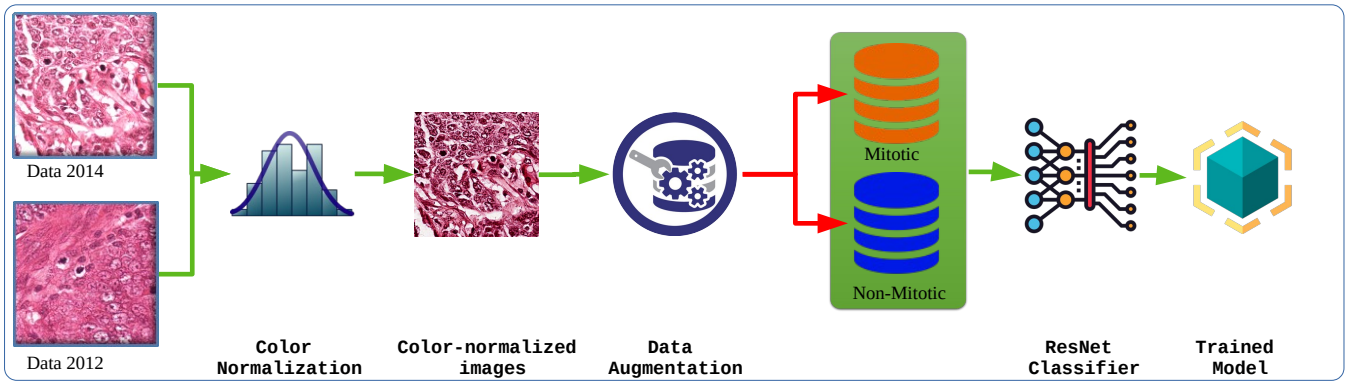


Fig. 1: A graphical outline of the proposed method

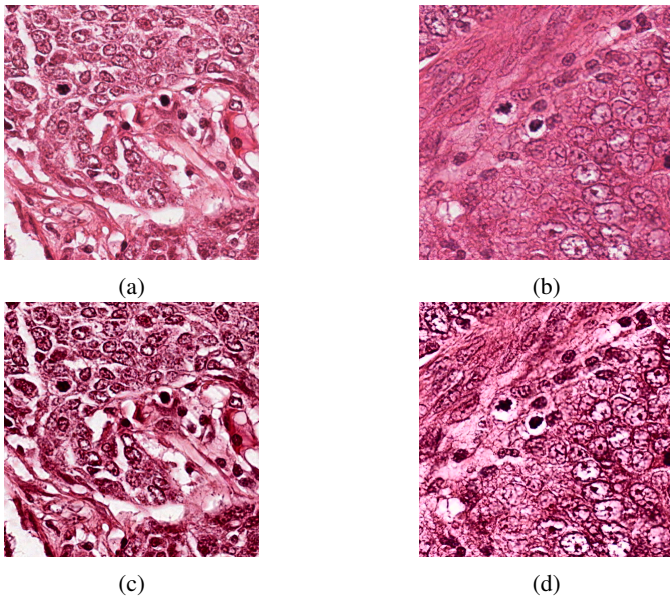


Fig. 2: Examples of the color normalization outputs from the ICPR2012 and MITOS-ATYPIA datasets are shown in images (c, d).

CPUs, 128 GB of RAM, and an MSI GeForce GTX TITAN X with 3072 Cores and 12 GB of RAM as its experimental platform.

B. Evaluation Metrics

The DCNN model is evaluated using four parameters: precision, accuracy, F1 score, and recall. Accuracy is the percentage of images that are correctly classified out of all images. Precision is the proportion of correctly classified mitotic events compared to all projected mitotic events. Recall is the proportion of correctly identified mitotic events compared to all mitotic events. The F1 score is the harmonic mean of recall and accuracy.

C. Results and Discussion

The proposed algorithm for mitotic detection was tested on two publicly available datasets: ICPR-MITOS 2012 [12]

and MITOS-ATYPIA-14 [13]. The evaluation metrics were the same as the ones used in the challenge. Note that the evaluation metric used in the challenge is not the standard one in the field.

The ResNet 50 model was trained on this last combined dataset (see Figure 3).

The model trained on the combined dataset seems to be more stable, based on a graph of the epochs of training versus validation loss and validation accuracy, training loss, and training accuracy.

The algorithm was compared with the top-ranked algorithm from the challenge, VGG16 [8], as well as some other algorithms from the field, namely, Mask-RCNN [14], and StarDist [15].

Our proposed algorithm outperforms all the other algorithms on all two datasets. The results are shown in Table I and Table II.

TABLE I: Results on the dataset of MITOS-ATYPIA-14

| Method | Precision | Recall | F1-score |
|--------------------|--------------|--------------|--------------|
| StarDist [15] | 0.556 | 0.552 | 0.554 |
| VGG16 [8] | 0.489 | 0.701 | 0.577 |
| Mask-RCNN [14] | 0.518 | 0.807 | 0.631 |
| Proposed algorithm | 0.650 | 0.873 | 0.708 |

TABLE II: Results on the ICPR2012 dataset

| Method | Precision | Recall | F1-score |
|--------------------|--------------|--------------|--------------|
| StarDist [15] | 0.576 | 0.540 | 0.563 |
| VGG16 [8] | 0.499 | 0.710 | 0.560 |
| Mask-RCNN [14] | 0.521 | 0.817 | 0.644 |
| Proposed algorithm | 0.641 | 0.859 | 0.691 |

The results indicate that the proposed approach can be used to automatically detection of mitotic cells in images of histopathology of breast cancer.

Our study has a number of limitations. First, our dataset is small. Second, our dataset is imbalanced and includes twice as many images without mitotic cells as images with mitotic cells. Third, our method is tested on a held-out test set and may not generalize to other datasets. Our work in future will focus on increasing the size of the dataset and balancing the

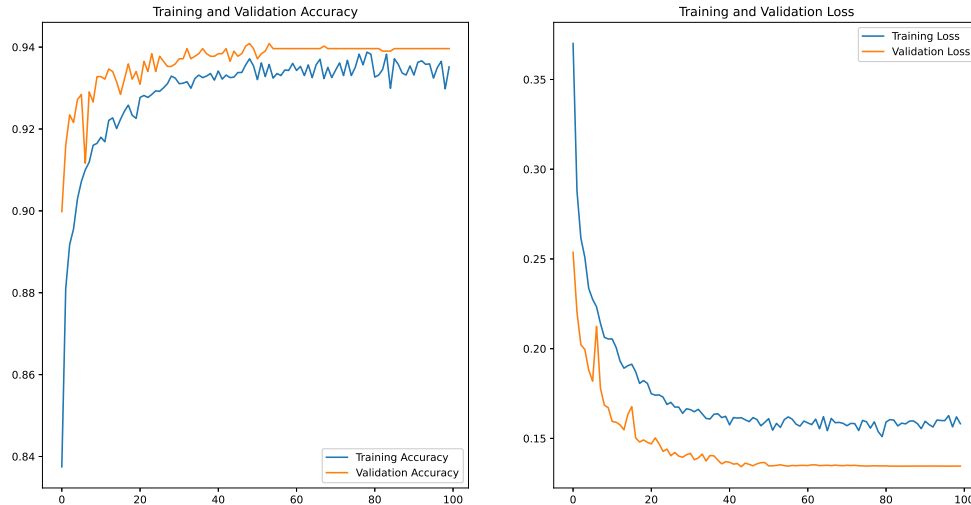


Fig. 3: Learning pattern of the ResNET using combined dataset

dataset. Futur work will also focus on testing the proposed method on other datasets.

ACKNOWLEDGMENT

The first and third authors have been supported by the CPER/FEDER (ID: RIN "Eye Tracking"). The first and second authors have been supported by the European Union (FEDER) and the Normandy Region through the MIDIPATH project.

IV. CONCLUSION

In this work, a simple and fast algorithm for mitotic detection in microscopy images was proposed. The proposed algorithm consists of four main steps: image color normalization, data augmentation, CNN training and image classification. Two publicly accessible datasets were used to evaluate the proposed technique. The proposed technique is designed to take advantage of the high level of detail and resolution in histology images. The proposed technique showed superior performance over all the compared methods in all of the two datasets in terms of recall.

REFERENCES

- [1] S. Lei, R. Zheng, S. Zhang, S. Wang, R. Chen, K. Sun, H. Zeng, J. Zhou, and W. Wei, "Global patterns of breast cancer incidence and mortality: A population-based cancer registry data analysis from 2000 to 2020," *Cancer Communications*, vol. 41, no. 11, pp. 1183–1194, 2021.
- [2] D. Ciregan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 2012, pp. 3642–3649.
- [3] I. O. Sigirci, A. Albayrak, and G. Bilgin, "Detection of mitotic cells in breast cancer histopathological images using deep versus handcrafted features," *Multimedia Tools and Applications*, vol. 81, no. 10, pp. 13 179–13 202, 2022.
- [4] N. Dif and Z. Elberichi, "Deep learning methods for mitosis detection in breast cancer histopathological images: A comprehensive review," *Artificial Intelligence and Machine Learning for Digital Pathology*, pp. 279–306, 2020.
- [5] W. Ba, R. Wang, G. Yin, Z. Song, J. Zou, C. Zhong, J. Yang, G. Yu, H. Yang, L. Zhang *et al.*, "Diagnostic assessment of deep learning for melanocytic lesions using whole-slide pathological images," *Translational oncology*, vol. 14, no. 9, p. 101161, 2021.
- [6] D. K. Das and P. K. Dutta, "Efficient automated detection of mitotic cells from breast histological images using deep convolution neural network with wavelet decomposed patches," *Computers in biology and medicine*, vol. 104, pp. 29–42, 2019.
- [7] S. Ren, K. He, R. Girshick, J. Sun, and F. R-CNN, "Towards real-time object detection with region proposal networks, 2015," *ArXiv*. Available online: <https://arxiv.org/abs/1506.01497> (accessed on June 2021), 2019.
- [8] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [9] J. Xie, R. Girshick, and A. Farhadi, "Unsupervised deep embedding for clustering analysis," in *International conference on machine learning*. PMLR, 2016, pp. 478–487.
- [10] 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*, 2016, pp. 770–778.
- [11] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [12] R. Ludovic, R. Daniel, L. Nicolas, K. Maria, I. Humayun, K. Jacques, C. Frédérique, G. Catherine *et al.*, "Mitosis detection in breast cancer histological images an icpr 2012 contest," *Journal of pathology informatics*, vol. 4, no. 1, p. 8, 2013.
- [13] L. Roux, "Mitosis atypia 14 grand challenge," *Mitosis atypia 14 grand challenge*, 2014.
- [14] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.
- [15] U. Schmidt, M. Weigert, C. Broaddus, and G. Myers, "Cell detection with star-convex polygons," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 265–273.