



Article

# Artificial Intelligence-Based Mitosis Detection in Breast Cancer Histopathology Images Using Faster R-CNN and Deep CNNs

Tahir Mahmood , Muhammad Arsalan , Muhammad Owais , Min Beom Lee and Kang Ryoung Park \*

Division of Electronics and Electrical Engineering, Dongguk University, 30 Pildong-ro 1-gil, Jung-gu, Seoul 04620, Korea; tahirmahmood.cs@gmail.com (T.M.); arsal@dongguk.edu (M.A.); malikowais266@gmail.com (M.O.); mblee@dongguk.edu (M.B.L.)

\* Correspondence: parkgr@dgu.edu; Tel.: +82-10-3111-7022; Fax: +82-2-2277-8735

Received: 3 February 2020; Accepted: 6 March 2020; Published: 10 March 2020



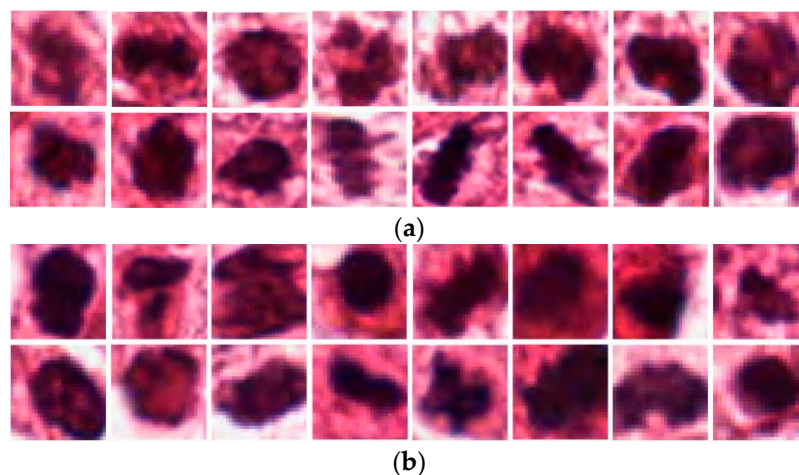
**Abstract:** Breast cancer is the leading cause of mortality in women. Early diagnosis of breast cancer can reduce the mortality rate. In the diagnosis, the mitotic cell count is an important biomarker for predicting the aggressiveness, prognosis, and grade of breast cancer. In general, pathologists manually examine histopathology images under high-resolution microscopes for the detection of mitotic cells. However, because of the minute differences between the mitotic and normal cells, this process is tiresome, time-consuming, and subjective. To overcome these challenges, artificial-intelligence-based (AI-based) techniques have been developed which automatically detect mitotic cells in the histopathology images. Such AI techniques accelerate the diagnosis and can be used as a second-opinion system for a medical doctor. Previously, conventional image-processing techniques were used for the detection of mitotic cells, which have low accuracy and high computational cost. Therefore, a number of deep-learning techniques that demonstrate outstanding performance and low computational cost were recently developed; however, they still require improvement in terms of accuracy and reliability. Therefore, we present a multistage mitotic-cell-detection method based on Faster region convolutional neural network (Faster R-CNN) and deep CNNs. Two open datasets (international conference on pattern recognition (ICPR) 2012 and ICPR 2014 (MITOS-ATYPIA-14)) of breast cancer histopathology were used in our experiments. The experimental results showed that our method achieves the state-of-the-art results of 0.876 precision, 0.841 recall, and 0.858 F1-measure for the ICPR 2012 dataset, and 0.848 precision, 0.583 recall, and 0.691 F1-measure for the ICPR 2014 dataset, which were higher than those obtained using previous methods. Moreover, we tested the generalization capability of our technique by testing on the tumor proliferation assessment challenge 2016 (TUPAC16) dataset and found that our technique also performs well in a cross-dataset experiment which proved the generalization capability of our proposed technique.

**Keywords:** breast cancer; mitotic cell count; artificial intelligence; Faster R-CNN; deep CNNs

## 1. Introduction

Breast cancer is the most common and leading cause of death among women. According to the global cancer project (GLOBOCAN 2012), breast cancer accounts for 25.1% of all cancers in women [1]. Early diagnosis of breast cancer is an important factor for the reduction of the mortality rate because its treatment plan is advised on the basis of the grade and prognosis of the cancer. To determine the grade of breast cancer, the Nottingham grading system has been widely used. According to this system, there are three biomarkers for the grading of breast cancer in histopathology images. These biomarkers are

nuclear atypia, tubule formation, and the mitotic cell count. Among these biomarkers, the mitotic cell count is the most important biomarker as the mitosis cell division process is directly related to the prognosis of tumors [2]. In practice, mitotic cells are generally detected via the visual inspection of the histopathology slide images of the breast under high-resolution microscopes. However, this procedure is tedious, time consuming, and subjective. A low-skilled pathologist could thus arrive at inaccurate detections, which could have serious consequences. Recently, artificial-intelligence (AI) techniques had a great impact on every field of life and even in the medical field. The majority of the processes are now automated and can even be used as a second-opinion system in medical diagnosis. AI techniques [3–10] have been developed previously for solving problems in the medical field. Mitotic-cell detection can also be automated using AI techniques; however, it comprises several challenges. For example, it is difficult to differentiate between mitotic and normal cells without pathological knowledge and the use of high-resolution microscopes because mitotic cells have a texture and morphological features that are similar to normal cells, as shown in Figure 1. Moreover, some of the other organelles of the cell, such as apoptotic cells, have a similar appearance to that of mitotic cells. The mitosis process comprises four stages where each has its own unique characteristics, and thus, a robust technique is required to be developed for detecting diverse mitotic cells. Another major challenge is the maintenance of the standard data-preparation environment. Biopsy, slide preparation, and scanning procedures are required to be performed carefully because a low performance is obtained in the case of issues in data collection, slide preparations, and scanning [11].



**Figure 1.** Examples of (a) mitotic and (b) non-mitotic cells.

Mitotic-cell detection techniques can be categorized into two divisions based on the features extracted from regions of interest (ROIs): handcrafted features and deep features. Handcrafted features are extracted from ROIs by using conventional image-processing techniques. Features such as color, morphology, and texture are extracted, which is followed by classification using machine-learning classification algorithms such as an artificial neural network and a support vector machine (SVM). Previous research [12–16] that has been conducted on this approach has demonstrated good performance and can be used in small-scale applications. In the second approach, deep features are extracted from ROIs by using deep-learning techniques [17–20]. In the mitotic-cell detection task, deep features-based techniques are further divided into three main categories based on the problem formulation. Some researchers consider mitotic-cell detection as a classification task, while others consider it as a semantic segmentation task because of the pixels-based annotations. Few others also consider it to be an object-detection task because the objective was not to determine the shape of the mitotic cells but to count them. In our proposed work, we considered mitotic-cell detection as an object-detection task and proposed a technique that provides the state-of-the-art results.

The rest of this paper is organized as follows. Sections 2 and 3 present the related research works and our contributions, respectively. Sections 4 and 5 respectively present the proposed method and explanations of the experimental setup and performance analysis. Section 6 presents a discussion of the obtained results, while Section 7 presents the conclusion of our research.

## 2. Related Works

Mitotic-cell detection in hematoxylin-and-eosin-stained (H&E-stained) biopsy images have been researched since the invention of whole-slide imaging scanners. In addition, owing to recent developments in AI, a number of techniques have been developed which demonstrate outstanding performance and can be used in real-time applications. Previous research can be divided into two categories: handcrafted-features-based and deep-features-based research. Details of each of the aforementioned categories are provided in the following section.

### 2.1. Mitosis Detection Using Handcrafted Features

Conventional image-processing techniques have been used for the extraction of handcrafted features such as shape, texture, and color, which are followed by the use of machine-learning algorithms for mitotic cell detection. In the extant literature, several techniques comprising the use of handcrafted features have been presented. Irshad presented a technique [12] in which all the expected objects were first segmented, and statistical and morphological features were extracted and classified using a decision-tree classifier [21]. This technique ranked second in the mitosis detection challenge of the international conference on pattern recognition (ICPR) 2012. Tashk et al. presented a technique [13] based on local binary pattern (LBP) and SVM [21] as a classification algorithm. LBP features have high discriminative power and are also invariant to grayscale changes. This technique ranked third in the aforementioned ICPR 2012 challenge. Sommer et al. used shape and intensity features along with texture features in their proposed technique [14] to distinguish between mitotic and non-mitotic cells while using an SVM classifier. They used two open-source biomedical image analysis software: “ilastik” [22] for the segmentation of objects and “CellCognition” [23] for classification into mitotic and non-mitotic cells. This technique comprises a small amount of parameter tuning and no user effort because open source software are used. However, this technique demonstrates a relatively low detection performance as compared to other handcrafted feature-based techniques. Paul et al. focused on the nucleus of the cell in their proposed technique [15]. They used a regenerative random forest tree classifier that demonstrated an excellent performance. However, this technique requires significant computational resources, and thus, it cannot be used in practical clinical application. The majority of handcrafted-features-based methods as presented in Table 1 provide a low detection performance as compared to the recently developed deep-features-based techniques, and they are also computationally expensive owing to the conventional image-processing operations required for the segmentation of objects.

**Table 1.** Comparison of previous studies and proposed method on mitosis detection.

Category	Method	Datasets	Strength	Weakness
Hand-crafted features	Morphological and statistical features with decision tree classifier [12]	ICPR 2012	Efficient in capturing texture features for mitotic cell segmentation	Low detection performance and computationally expensive
	LBP and SVM classifier [13]	ICPR 2012	High discriminative power, computational simplicity, and invariance to grayscale changes	Affected by rotation and limited structural information capturing
	Shape, texture, and intensity features with SVM classifier [14]	ICPR 2012	Small amount of parameter tuning and low user effort	Low detection performance and object segmentation using open-source software
	Intensity, texture, and regenerative random forest tree classifier [15]	ICPR 2012	Good performance for large data	Computationally expensive and complex due to random forest tree
Deep features	Sliding-window-based classification [17]	ICPR 2012	Good detection performance	Computationally expensive
	Combination of color, texture, and shape features, and CNN features with SVM classifier [18]	ICPR 2012	Easy to accommodate for multi-scanner data without major redesign	Computationally expensive
	Handcrafted and CNN features, random forest classifier, and CNN [19]	ICPR 2012	Fast and high precision	Using fixed global and local threshold in object-detection stage
	FCN model for objects segmentation and CNN for classification [20]	ICPR 2012	Robust, fast, and high precision	Not suitable for weakly annotated datasets, and object detection stage is computationally expensive
	Faster R-CNN-based detection and Resnet-50 for classification [2]	ICPR 2012 ICPR 2014	Good performance and inference time	VGG-16 is used as a feature extraction network of Faster R-CNN, which have the vanishing gradient issue
	Concentric circle approach for objects detection and FCN for segmentation [24]	ICPR 2012 ICPR 2014 TUPAC-16	Good technique for weakly annotated datasets	Low detection performance
	Modified Faster R-CNN with Resnet-101 feature-extraction network [25]	ICPR 2014 TUPAC-16	Less inference time	Resnet-101 can be replaced by shallow network
	Lightweight region-based R-CNN [26]	ICPR 2012 ICPR 2014	No requirement of powerful GPUs	Low detection performance
	Mask R-CNN for object detection and handcrafted and CNN features [27]	ICPR 2012 ICPR 2014	Highest performance and inference time	Using expensive GPUs and intensive training
	Faster R-CNN and score-level fusion of Resnet-50 and Densenet-201 (proposed)	ICPR 2012 ICPR 2014	High detection performance	Long processing time owing to multiple networks and intensive training

ICPR, international conference on pattern recognition; LBP, local binary pattern; SVM, support vector machine; CNN, convolutional neural network; Faster R-CNN, Faster region convolutional neural network; TUPAC, tumor proliferation assessment challenge; VGG, visual geometry group.



## 2.2. Mitosis Detection Using Deep Features

A deep-features-based method is more powerful than a handcrafted-features-based method because it takes into consideration thousands of meaningful features during the training. Cirešan et al. presented a technique [17] based on the sliding window approach for the extraction of deep features from images. This technique ranked first in the ICPR 2012 mitosis-detection contest. The sliding window approach is computationally expensive, and thus, this technique is not suitable for clinical application. Malon et al. combined handcrafted nuclear features and deep features from a convolutional neural network (CNN) [18]. This technique also comprised the use of an additional CNN that reduces the sensitivity of mitotic cells during feature extraction; however, it demonstrates a low performance and high computational complexity. Wang et al. presented a cascaded technique [19] in which two classifiers were used independently. One classifier is trained with handcrafted features and the other is trained with CNN features. In the testing stage, a third classifier is used if the outputs of the two classifiers are different. This technique is fast and computationally inexpensive; however, the ROI-selection performance with conventional image processing is lower than that obtained with the deep-learning technique. Chen et al. presented a two-stage technique [20] in which mitotic cells were segmented by a fully convolutional network (FCN) in stage 1, while in stage 2, all the detected objects were further refined by an additional CNN. Recently, region-based CNNs performed well in a number of computer vision applications. For mitotic-cell detection, Li et al. presented a technique [2] based on Faster region convolutional neural network (Faster R-CNN) [28] and residual network (Resnet)-50 [29]. Faster R-CNN initially detects mitotic cells, which are further refined by Resnet-50. The Faster R-CNN used in this technique comprises visual geometry group (VGG)-16 [30] as a feature-extraction network. This technique provides good results and requires less inference time; however, Faster R-CNN results can be improved by using other feature-extraction networks because VGG-16 has the vanishing gradient issue. Li et al. also presented another technique [24] based on concentric circles for a weakly annotated dataset of ICPR 2014. This approach is good for weakly annotated datasets. Cai et al. used a modified Faster R-CNN in their proposed technique [25]. Resnet-101 was used for the feature extraction of the Faster R-CNN. Although this technique is good, Resnet-101 can be replaced by a shallow network. Li et al. presented a lightweight region-based CNN technique [26] that was developed using standard desktop computers without graphics processing units (GPUs). Based on the mask R-CNN [31], Dodballapur et al. presented a technique [27] and that comprised Resnet-50 as a feature-extraction network. Xception network [32] was used for the reduction of false positives. This technique provides high accuracy; however, owing to its use of expensive GPUs and intensive training, it is not suitable for use in practical clinical applications. Table 1 presents a comparison between existing methods and proposed method for mitotic cell detection.

## 3. Contribution

The major contributions of this work are summarized as follows:

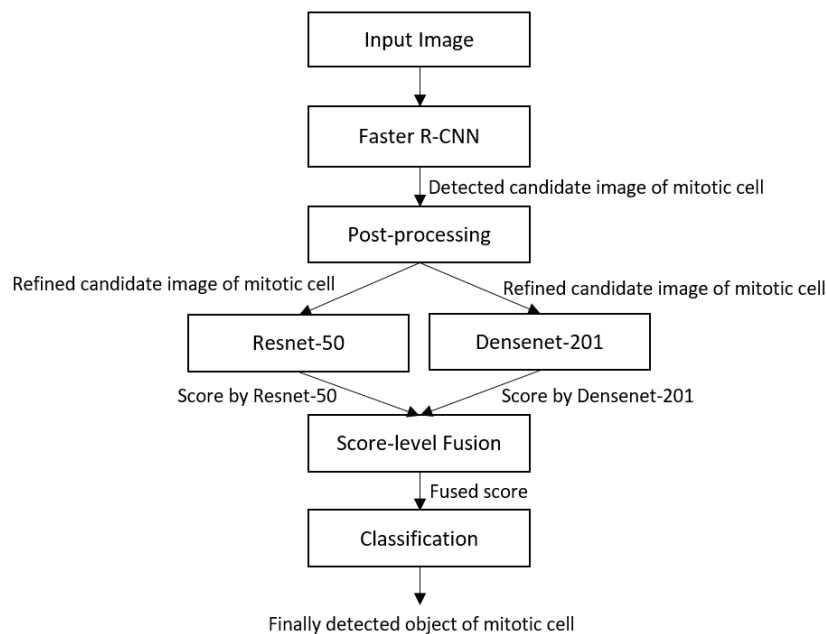
- This proposed technique provides state-of-the-art results in mitosis-detection tasks as per the ICPR 2012 and ICPR 2014 contest datasets.
- Faster R-CNN is used in the first stage in which the primary detection of mitotic cells was performed. We adopt Resnet-50 as a features-extraction network for the first time, thus obtaining better results as compared to the other techniques.
- In the proposed technique, a large number of false positives are produced because of the minute differences between mitotic and non-mitotic objects. To reduce the number of false positives, we perform post-processing on the basis of statistical, texture, shape, and color features.
- To further reduce the number of false positives, we perform a score-level fusion of Resnet-50 and a dense convolutional network (Densenet)-201. This is used for the first time in the mitotic-cells-detection task and it significantly reduces the number of false positives.

- To allow other researchers to perform fair comparisons, our trained models are publicly available in [33].

## 4. Proposed Method

### 4.1. Overview of Proposed Approach

Our proposed technique comprises a multistage mitotic-cell detection framework. There are four main stages in the proposed technique. In stage 1, an image is input into the trained Faster R-CNN detector, which is trained on the training datasets of the ICPR 2012 and ICPR 2014 mitosis-detection data. The detection results obtained using Faster R-CNN are adversely affected by a large number of false positives. Therefore, further post-processing is performed to reduce the number of false positives in stage 2 on the basis of statistical, texture, shape and color features. In stage 3, independently trained Resnet-50 and Densenet-201 scores are fused, and the final classification of the mitotic and non-mitotic cells is performed in stage 4. Figure 2 presents the flow diagram of the proposed technique.



**Figure 2.** Flow diagram of the proposed technique. Faster R-CNN, Faster region convolutional neural network.

### 4.2. Mitotic-Cell Detection Using Faster R-CNN

The image is input into the Faster R-CNN trained network for the initial mitotic-cell detection. The Faster R-CNN is a region-based CNN [28] that was presented in the object-detection competition of the ImageNet Large-Scale Visual Recognition Challenge 2015 [34]. The Faster R-CNN is the combination of three sub-networks, the feature-extraction network, region proposal network (RPN) [35], and classification network, as shown in Figure A1 in the appendix. In the feature-extraction network, a feature map known as an activation map is generated, and appropriate deep features are extracted. Different types of CNNs can be used as a feature extractor depending on the application requirements, and we use a Resnet-50 pre-trained on ImageNet database as a feature extraction network. Table A1 in the appendix presents the detailed architecture of the Resnet-50. The Resnet-50 includes 50 weighted layers that are based on the idea of skipping the blocks of convolution layers by using shortcut connections. In general, the basic blocks known as “bottleneck” blocks follow two design rules: use the same number of filters for the same output feature size and double the filters if the feature size is halved. Moreover, the down-sampling is performed by the convolution layers with a stride of 2, and batch normalization is performed after each convolution and before the rectified-linear-unit (ReLU)

activation. An identity short-cut is used if the input and output have the same dimensions, and the projection shortcut is used to match the dimensions through  $1 \times 1$  convolutions if the dimensions increase [29]. The feature map extracted from the last convolutional layer is forwarded to the RPN and classification networks (ROI pooling network). As only the feature map from the feature-extraction network is necessary for the RPN and classification network instead of the final classification, therefore, only the 49 convolutional layers without average pooling and fully connected layers in Resnet-50 are used in the feature-extraction network, as shown in Table A1 in the appendix.

The RPN is the second part of the Faster R-CNN. It is responsible for the generation of region proposals of various sizes and ratios, which are used in the final classification network. The detailed network architecture is presented in Table A2 in the appendix. In the RPN, anchor boxes of different scales and aspect ratios are initially generated over each pixel of the feature map. In general, nine anchor boxes with scales of 128, 256, and 512 and aspect ratios of 1:1, 1:2, and 2:1 are used. RPN predicts the probability that an anchor box is an object or background. The final list of the proposals is filtered according to the intersection over union (IOU) threshold of 0.8 and non-maximum suppression (NMS) for the target objects. The list of filtered anchor boxes at this stage is the required object proposals, which is forwarded to the next stage. The transformation of the anchor boxes to the final predicted region proposals requires the use of the following Equations (1) and (2). Equation (1) presents the scale invariant translation between the center coordinates, while Equation (2) presents the log-space translation between the width and height.

$$v_x = \frac{x_p - x_a}{w_a}, v_y = \frac{y_p - y_a}{h_a} \tag{1}$$

$$v_w = \log\left(\frac{w_p}{w_a}\right), v_h = \log\left(\frac{h_p}{h_a}\right) \tag{2}$$

where  $v_x, v_y, v_w,$  and  $v_h$  are the bounding box regression vectors, and  $x, y, w,$  and  $h$  are the x and y coordinates of the center, width, and height of each box, respectively. Moreover,  $x_p$  and  $x_a$  are the center coordinate x of each proposal box and anchor box, respectively.

The feature map extracted in step 1 and the region proposals generated in step 2 are input into the classification part of the Faster R-CNN. Table A3 in the appendix presents the details of the classification network used in the Faster R-CNN. In the classification network, the feature map is cropped at a specific point by using region proposals. Each of the cropped feature maps has a different size, and therefore, ROI pooling is applied to obtain a uniform size. The bounding box regression vectors and mitotic cell probabilities are obtained after passing through fully connected layers. The bounding box regression vectors are used for the refinement of proposal boxes into prediction boxes followed by the removal of overlapping boxes via NMS, which results in the final detection results. We use the Faster R-CNN because it provides the highest accuracy and lowest computational complexity as compared to other region-based CNNs such as R-CNN [36] and Fast R-CNN [37]. In R-CNN and Fast R-CNN, region proposals are generated by a selective search algorithm [38] followed by the use of a detection network for classification and bounding boxes regression. The selective search algorithm and detection network are decoupled because false negatives have a direct effect on the detection network. Another disadvantage is the high computational cost and required time of the selective search algorithm. To overcome these shortcomings, the Faster R-CNN is used to replace the selective search by the RPN and shares convolutions across region proposals. Therefore, the Faster R-CNN is computationally inexpensive and has a high accuracy.

In the proposed technique, the Faster R-CNN detects the mitotic cells. During training, the weights are trained to get a minimal loss for each anchor box or proposal in the mini-batch. The loss function of the Faster R-CNN is as follows:

$$L(p, p^*, v, v^*) = L_{cls}(p, p^*) + \sigma p^* L_{reg}(v, v^*) \tag{3}$$

In Equation (3),  $p$  is the probability that an anchor box is the object.  $p^*$  is the ground truth label (mitotic cell = 1, and background = 0), while  $v$  and  $v^*$  are the bounding-box regression vector of the anchor box and its corresponding ground truth, respectively. Similarly,  $L_{cls}$  indicates the classification loss function, and  $L_{reg}$  represents the regression loss function.  $\sigma$  is the weight-balancing parameter used for the weighting of  $cls$  and  $reg$ . During the training, the weights are trained to minimize the loss value, and the Faster R-CNN can accurately predict the position of the mitotic cells.

#### 4.3. False-Positive Mitotic-Cells Removal via Post-Processing

Post-processing comprises the second stage of our proposed technique. In this stage, the mitotic cells detected using the Faster R-CNN are refined on the basis of handcrafted features. After the analysis of the mitotic cells, we observe that there is no appropriate distribution of gray levels. Instead, the mitotic cells usually have a dark blue color with an irregular texture. Therefore, we focus on color and texture in the post-processing. We use statistical LBP [39] histograms of oriented gradients (HOG) [40], and color features for the post-processing. Statistical features have two main sub-divisions: first order and second order. The first order statistical features are related to the gray-level distribution in the image. Single pixel-based estimation is performed to extract in the first order statistical features, while the spatial relationship between the pixels is ignored. Standard deviation, skewness, and kurtosis are examples of first-order statistical features. Histogram-based features are also included in this category. In second-order statistical features, the spatial relationship between the pixels is considered. Co-occurrence or run-length matrices are used for the extraction of the features. Angular second moment, contrast, correlation, homogeneity, and entropy are examples of second-order statistical features.

In our method, we use the first-order statistical features of mean, standard deviation, skewness, and kurtosis. The mean of an image describes the average color of an image, while the standard deviation is the estimation of the underlying brightness probability distribution. Skewness is the description of the darker and lighter colors with respect to the mean, and kurtosis provides information of the uniformity distribution of the intensity distribution. In our method, there is a minute difference between the statistical features of the cells. The optimal threshold is obtained with the training data for each of the features, and each of the detected objects is compared with the threshold for the acceptance or rejection as a mitotic cell. The same procedure is iterated with the remaining post-processing based on the second, third, and fourth features.

The second feature used for the post-processing is the LBP. It is the measure of the local image contrast [41], and it has been used in a number of computer vision applications [42–44]. For the extraction of the LBP feature, each center-pixel value is subtracted from each of the eight adjacent pixels. If the result of the subtraction yields a negative number, the eight adjacent pixels are represented by 0, else it is represented by 1. For labeling this specific pixel, a decimal value is obtained from the conversion of binary code obtained from the concatenation of all the binary codes in a clockwise direction starting from the top left adjacent pixel. Equations (4) and (5) represents the extraction process of the LBP features.

$$LBP_{R,P} = \sum_{i=0}^{P-1} s(g_i - g_c)2^i, \tag{4}$$

$$\text{where } s(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases} \tag{5}$$

In the above equation,  $P$  represents the number of neighboring pixels,  $R$  is the radius of the LBP circle, and  $g_i$  and  $g_c$  are the neighboring pixels and center-pixel gray levels, respectively. The binary value is represented by  $s(x)$ . In our method, the LBP features play an important role because the mitotic cells have an irregular texture and can be differentiated from the normal cells based on the contrast of the ROIs. Therefore, the LBP features are used in the post-processing.

The third post-processing feature is HOG. It is a shape-based feature [45] and has been used in various computer vision applications [46,47]. The target image is pre-processed and resized to a ratio

of 1:2 because the image is required to be divided into  $8 \times 8$  or  $16 \times 16$  patches. The gradient vector of each pixel is calculated along with its magnitude and direction. If we divide the image into cells of  $8 \times 8$  pixels, the magnitude of 64 pixels in each cell is binned and added into nine buckets of unsigned direction. A  $2 \times 2$ -cells block then slides over the image. For each block, four histograms of four cells are concatenated into one dimensional vector of 36 values followed by normalization to obtain a unit weight. The concatenation of all the block vectors is the final HOG feature. Figure A2 in the appendix presents the HOG feature for the mitotic cells. The last set of features includes the color features. A color histogram is used for the extraction of the RGB colors because there exists a diversity in the colors in mitotic cells. In the color histogram, the frequencies of different colors in the image are represented by the discretization of the color values into a number of bins. The frequency of each color in a bin is then represented by the histogram.

Feature extraction is followed by the acceptance or rejection of objects as a mitotic cell based on an optimal threshold of each of the feature. These optimal thresholds are experimentally obtained from the training data. Objects are accepted or rejected as a mitotic cell by comparing with thresholds set for each of the feature. In detail, any candidate is determined as a mitotic cell in the case that its feature value (for example, HOG, LBP, statistical, and color features) is higher than the optimal threshold whereas it is determined as non-mitotic cell if its value is lower than the threshold. Different set of features and thresholds are tried to get the best discriminative rule for the acceptance of a large number of mitotic cells.

#### 4.4. Final Classification of Mitotic Cells via Score-Level Fusion of Two CNNs

In the next stage, the final classification of the mitotic cells is performed via the score-level fusion of Resnet-50 and Densenet-201. The detailed structure of Resnet-50 is presented in Table A1 in the appendix, while the explanations of Resnet-50 are presented in Section 4.2. As the final score should be obtained in this case, average pooling, Softmax, fully connected, and classification layers are included in Resnet-50 in Table A1. Densenet adopts a dense connectivity, which improves the skip connection structure of Resnet. This is the method of concatenating the feature maps of the  $l$ th layer and previous layers based on dense block. Therefore, the input of the  $l$ th layer comprises the concatenated feature map of the previous layers ( $x_0, x_1, \dots, x_{l-1}$ ), as presented in Equation (6) [48]. In Equation (6),  $H_l(\cdot)$  is a function that includes the operations of convolution, pooling, batch normalization (BN), or ReLU. The detailed structure and explanations of Densenet-201 can be referred to in [48].

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]) \quad (6)$$

Score-level fusion is an integration technique in which the scores from multiple modalities are fused to make a decision. In general, the data of a single modality lack uniqueness and non-universality and also comprise noise [49]. Therefore, multi-modal data obtained by the fusion of single modalities have better discrimination abilities and are used in various applications [50,51]. The fusion of the information can usually be performed at three levels: (a) feature-extraction level, (b) matching-score level, and (c) decision level. In the feature-extraction level, a higher-dimension feature vector is obtained by concatenating the features that are obtained from the individual classifier. Feature-reduction techniques are employed for the selection of useful features. The score-level fusion integrates the classifier's scores based on the proximity of the scores, and the decision-level fusion is performed based on the final decision of "Yes" or "No" [52].

In our work, we have used the score-level fusion technique because of its superior performance as compared to other fusion techniques [52]. The pretrained Resnet-50 and Densenet-201 are each trained over the patches of the detected mitotic cells. In the testing phase, an image is passed through trained Faster R-CNN and each of the detected objects are then passed through trained Resnet-50 and Densenet-201 networks. We know that probabilistic scores are produced from the output layer, so we obtained that score for each of the classifiers and fused it together for final classification. In score-level



fusion the match scores are fused together to render a decision about the identity of object. Moreover, as mentioned above for different levels of fusion, we adopted hierarchical score-level fusion. In detail, the score of input candidate is compared with a predetermined threshold in the first hierarchical stage, and that whose score is larger than the threshold is determined as mitotic cell. Then, it is compared with the threshold in the second stage and proceeds until the last stage. After all the stages, the accepted one is determined as a mitotic cell. These thresholds were experimentally determined with training data. It can be seen in Figure A3 of the appendix how positive and negative samples are passed through Resnet-50 and Densenet-201, followed by the score-level fusion and classification.

## 5. Experiments and Performance Analysis

In this section, we present the experimental datasets and environment settings with hardware and software specifications used for performing the experiments. Moreover, the evaluation criteria and performance analysis are also described in detail.

### 5.1. Datasets

In the proposed technique, two publicly available datasets of mitotic-cell detection in histopathology images are used. The details of each dataset are as follows.

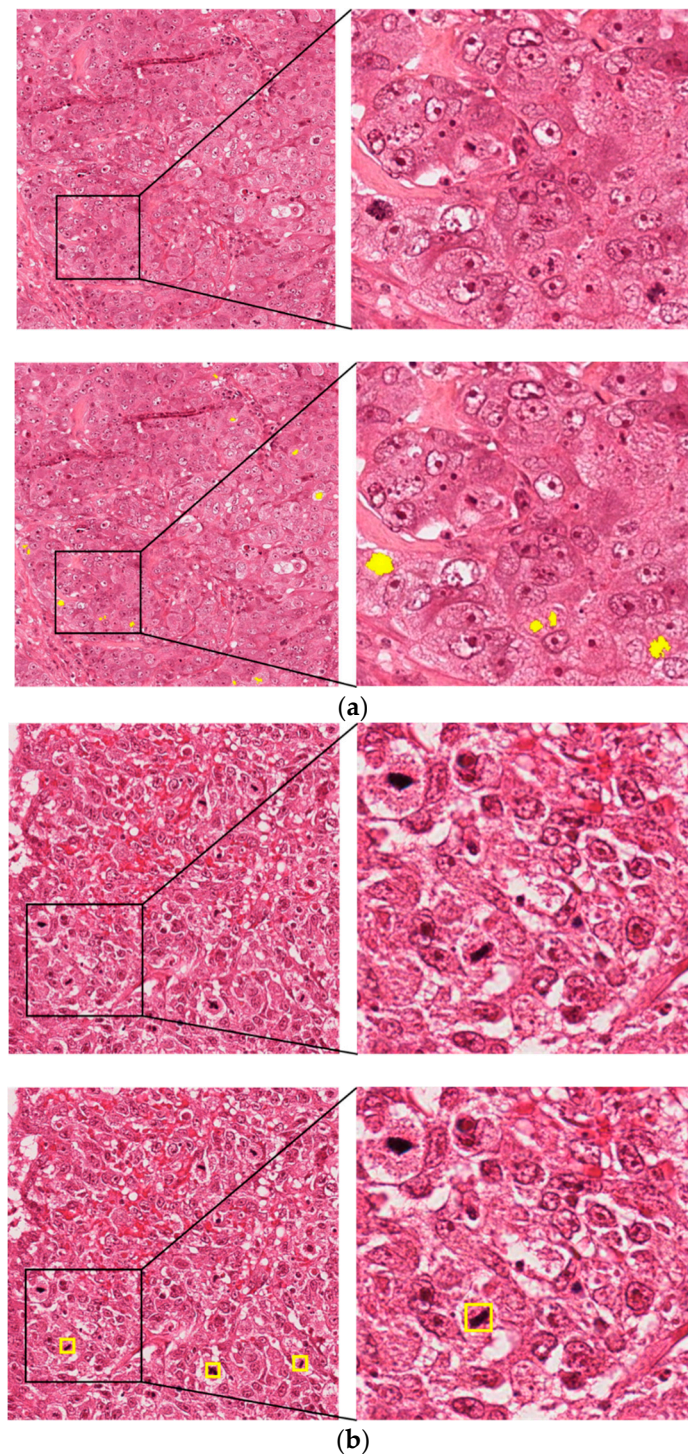
#### 5.1.1. ICPR 2012 MITOSIS Dataset

The ICPR 2012 MITOSIS dataset was introduced in the ICPR 2012 contest [53]. It comprises 50 RGB images of which 35 images are fixed for training and 15 images for testing. For acquiring this dataset, 10 high-power fields (HPFs) of size  $512 \times 512 \mu\text{m}^2$  at  $40\times$  magnification are selected from the biopsy images of five breast-cancer patients. The resolution of each image is  $0.2456 \mu\text{m}$  per pixel, and each of the HPFs has an area of  $512 \times 512 \mu\text{m}$ , which implies that the image size is  $2084 \times 2084$  pixels. Two scanners, namely, Aperio XT (scanner A) scanner and Hamamatsu NanoZoomer Scanner (scanner H) were used. Expert pathologists performed the annotations by mutual consent. There were 226 and 101 mitotic cells in the training and testing sets, respectively. In our experiments, we used Aperio XT scanner images. According to the instructions of the dataset collector, we obtained 35 images containing 226 mitotic cells in the training dataset and 15 images containing 101 mitotic cells in the testing dataset. The image dimension of  $2084 \times 2084$  pixels was too large for training over normal GPUs. Therefore, we densely extracted patches of size  $521 \times 521$  with a pixel difference of 40, which generates 1493 images comprising 1920 mitotic cells. Figure 3a presents sample figures of the ICPR 2012 dataset. The upper-left figure is the original dataset image, and the upper-right image presents the magnified version of a specific part in the image while the lower images are the ground truth images of the upper images.

#### 5.1.2. ICPR 2014 Dataset

The ICPR 2014 dataset was presented in the MITOS-ATYPIA-14 grand challenge [54], in which researchers were required to compete for nuclear atypia scoring and mitotic-cells count. This dataset comprised 1200 training images acquired from 16 different biopsies and 496 testing images acquired from five different breast biopsies. The size of each images is  $1539 \times 1376$  pixels at  $40\times$  magnification, which is much smaller than those of the ICPR 2012 dataset. Pathologists annotated only the centroid pixels of each of the mitosis. This dataset comprises significant variations in the dataset images according to the tissue-acquisition process, staining, and lighting conditions, and thus, it is challenging to achieve outstanding performance. In our experiments, as the ground truths of the testing data are not provided by the organizers, we performed the experiments by splitting the training data into training and validation sets using the same split protocol mentioned in [2,24,25,27] for obtaining a fair comparison. There is no requirement of patch extraction as the image size is much smaller than that of the ICPR 2012 image. Figure 3b presents sample images from the ICPR 2014 dataset along with ground truth images.





**Figure 3.** Examples of (a) ICPR 2012 and (b) ICPR 2014 datasets with ground truth images.

### 5.2. Data Augmentation

Deep-learning networks usually require a sufficient amount of data for complete and efficient training; however, in the majority of cases, a large amount of data is not available. Therefore, the data augmentation technique is used to generate more data from the original data. Conventional image-processing techniques such as translation, rotation, and flipping are applied to generate new images from the original images [55]. In our experiments, we performed data augmentation only for the training of Resnet-50 and Densenet-201. For the training of Faster R-CNN, we used the original

training data instead of the augmented data because sufficient training data can be obtained based on multiple patches of the input image. The detected objects by Faster R-CNN are collected from the training data and resized to  $224 \times 224$  pixels because the input-image size for Resnet-50 and Densenet-201 pretrained on the ImageNet database should be  $224 \times 224$  pixels. Horizontal flipping, vertical flipping, translation, and cropping-resizing operations are then applied over the image to generate an extra augmented image. We performed data augmentation by using horizontal and vertical flipping and by translation at different axes with flipping.

### 5.3. Experimental Setup and Training

#### 5.3.1. Experimental Setup

The proposed technique was implemented in MATLAB R2019a (MathWorks, Inc., Natick, MA, USA) [56] on a desktop computer with a Windows 10 operating system. The desktop computer had a central processing unit with a 3.60-GHz Intel® (Santa Clara, CA, USA) Core-i7-7700 [57], 16-GB random access memory, and an NVIDIA GeForce GTX 1070 GPU [58].

#### 5.3.2. Training

In our method, training was performed at two different stages. In stage 1, we trained the Faster R-CNN, and in stage 2, Resnet-50 and Densenet-201 were trained for the score-level fusion.

##### Training of Faster R-CNN

As a feature-extraction network of Faster R-CNN, Resnet-50 was pretrained on the ImageNet database, and it was further trained with the ICPR 2012 and ICPR 2014 datasets. The end-to-end training method was used for the simultaneous training of the RPN and classification network. Several overlap ratios of bounding boxes were used to obtain better results, and the stochastic gradient descent (SGD) [59] method was used for the optimization. The SGD method efficiently optimizes all the learnable parameters of the model. The batch size, momentum, learning rate, and weight decay are 1, 0.9, 0.003, and 0.0005, respectively. The number of epochs is also a key parameter for training because the network can be under-fitted or over-fitted, and we performed the training for 25 epochs for both datasets.

##### Training of Resnet-50 and Densenet-201

For the training of Resnet-50 and Densenet-201, SGD was also used for optimization. The initial learning rate of 0.001, momentum of 0.9, learning-rate drop factor of 0.1, and mini-batch size of 50 were used for the training.

### 5.4. Performance Evaluation of Proposed Method

#### 5.4.1. Performance Evaluation Metric

The performance of the proposed techniques is measured based on the number of correctly detected mitotic cells. According to the contest criteria, a true positive is defined as a positive that exists less than  $5 \mu\text{m}$  (20 pixels) and  $8 \mu\text{m}$  (32 pixels), respectively, from ground truth position in the ICPR 2012 and ICPR 2014 datasets. Based on these criteria, we identify the true-positive, false-negative, and false-positive cases. A true positive indicates that the ground truth mitotic cell is correctly detected as a mitotic cell, whereas a false negative indicates that the ground truth mitotic cell is not detected as a mitotic cell. A false positive means that the ground truth non-mitotic cell is incorrectly detected as a mitotic cell. Based on these, precision, recall, and F1-measure are used for the evaluation, as shown in Equations (7)–(9).

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (7)$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{8}$$

$$\text{F1 - measure} = \frac{2 \text{ Precision Recall}}{\text{Precision} + \text{Recall}} \tag{9}$$

where *TP* is the number of true positives, *FP* is the number of false positives, and *FN* is the number of false negatives.

Tables 2 and 3 present the comparative accuracies of the proposed method with the state-of-the-art methods used with the ICPR 2012 and ICPR 2014 datasets. Our proposed technique achieved a precision, recall, and F1-measure of 0.876, 0.841, and 0.858, respectively, with the ICPR 2012 dataset, which indicates that the proposed technique outperforms all other state-of-the-art methods. The ICPR 2014 dataset is more complex and challenging as compared to the ICPR 2012 dataset. We achieved a precision, recall, and F1-measure of 0.848, 0.583, and 0.691, respectively with the ICPR 2014 dataset, which indicates that the proposed technique again outperforms all other techniques.

**Table 2.** Comparisons of the proposed method and previous techniques with ICPR 2012 dataset.

Technique	Precision	Recall	F1-Measure
Sommer et al. [14]	0.519	0.798	0.629
Malon et al. [18]	0.747	0.590	0.659
Tashk et al. [13,60]	0.699	0.72	0.709
Irshad [12,60]	0.698	0.74	0.718
Wang et al. [19]	0.84	0.65	0.735
Ciresan et al. [17]	0.88	0.70	0.782
Li et al. [26]	0.78	0.79	0.784
Chen et al. [20]	0.804	0.772	0.788
Li et al. [24]	0.846	0.762	0.802
Paul et al. [15]	0.835	0.811	0.823
Li et al. [2]	0.854	0.812	0.832
Proposed method	0.876	0.841	0.858

**Table 3.** Comparisons of the proposed method and previous techniques with ICPR 2014 dataset (*N.R.* means “not reported”).

Technique	Precision	Recall	F1-Measure
Li et al. [2]	<i>N.R.</i>	<i>N.R.</i>	0.572
Cai et al. [25]	0.53	0.66	0.585
Li et al. [24]	0.495	0.785	0.607
Li et al. [26]	0.654	0.663	0.659
Dodballapur et al. [27]	0.58	0.82	0.68
Proposed method	0.848	0.583	0.691

#### 5.4.2. Ablation Study

We present the ablation study of our proposed technique to gain the deep insights of the improvements caused by the different components in our proposed technique. Tables 4 and 5 present the ablation study on ICPR 2012 and ICPR 2014 datasets. First, in the Faster R-CNN results, it can be observed that the highest recall is obtained by the optimization of anchor sizes and anchor scales of the Faster R-CNN. Second, post-processing impact can be seen on the improvement of precision as false positives are eliminated in the post-processing stage. Third, deep networks Resnet-50 and Densenet-201 improved the performance but were still lower than the state-of-the-art technique. Score-level fusion of these networks followed by classification has a huge impact and it can be seen that our proposed technique outperformed other state-of-the-art techniques for mitotic cell detection after combination of Faster R-CNN, post-processing and score level fusion of deep networks. Regarding the feature-driven pipeline, our method without feature-driven pipeline (FRCNN of Tables 4 and 5)

shows lower accuracy than that with feature-driven pipeline (FRCNN + PP of Tables 4 and 5), which confirms that our feature-driven pipeline (post-processing) is necessary to enhance the accuracy with our deep learning models.

**Table 4.** Quantitative comparison of each component of the proposed method on ICPR 2012 dataset. (FRCNN indicates Faster R-CNN, PP indicates post-processing (feature-driven method), D-net indicates Densenet-201, R-net indicates Resnet-50 and SF indicates score-level fusion of Densenet-201 and Resnet-50.)

Technique	Precision	Recall	F1-Measure
FRCNN	0.540	0.851	0.661
FRCNN + PP	0.641	0.851	0.731
FRCNN + PP + D-net	0.793	0.722	0.756
FRCNN + PP + R-net	0.7692	0.792	0.780
FRCNN + PP + SF (Proposed)	0.876	0.841	0.858

**Table 5.** Quantitative comparison of each component of the proposed method on ICPR 2014 dataset. (FRCNN indicates Faster R-CNN, PP indicates post-processing (feature-driven method), D-net indicates Densenet-201, R-net indicates Resnet-50 and SF indicates score-level fusion of Densenet-201 and Resnet-50.)

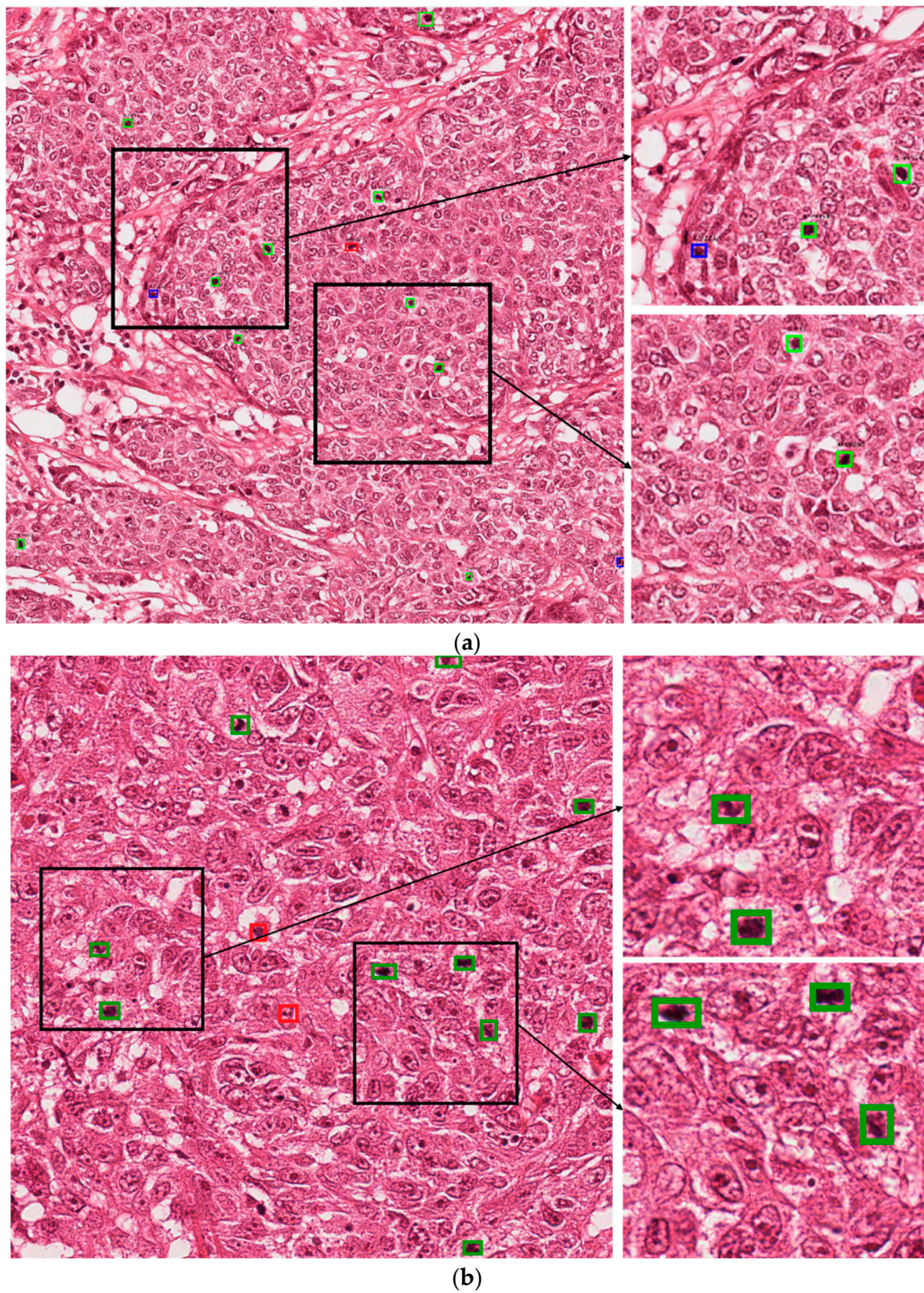
Technique	Precision	Recall	F1-Measure
FRCNN	0.521	0.641	0.575
FRCNN + PP	0.536	0.64	0.584
FRCNN + PP + D-net	0.674	0.599	0.634
FRCNN + PP + R-net	0.689	0.586	0.633
FRCNN + PP + SF (Proposed)	0.848	0.583	0.691

#### 5.4.3. Correct and Incorrect Detection Cases with Proposed Method

Figure 4 presents the correct detection cases for the proposed technique. In these figures, the green boxes represent true positives and the red box represent false positives, while the blue box represents false negatives. As shown in Figure 4, we confirm that our method can correctly detect the mitotic cells even in cases wherein it is difficult to discriminate the mitotic and non-mitotic cells based on human observation. Figure 5 presents the incorrect detection cases for the proposed technique. As shown in Figure 5, the incorrect detections mainly occur in cases where mitotic cells are either very small in size or have the same features as those of normal cells.

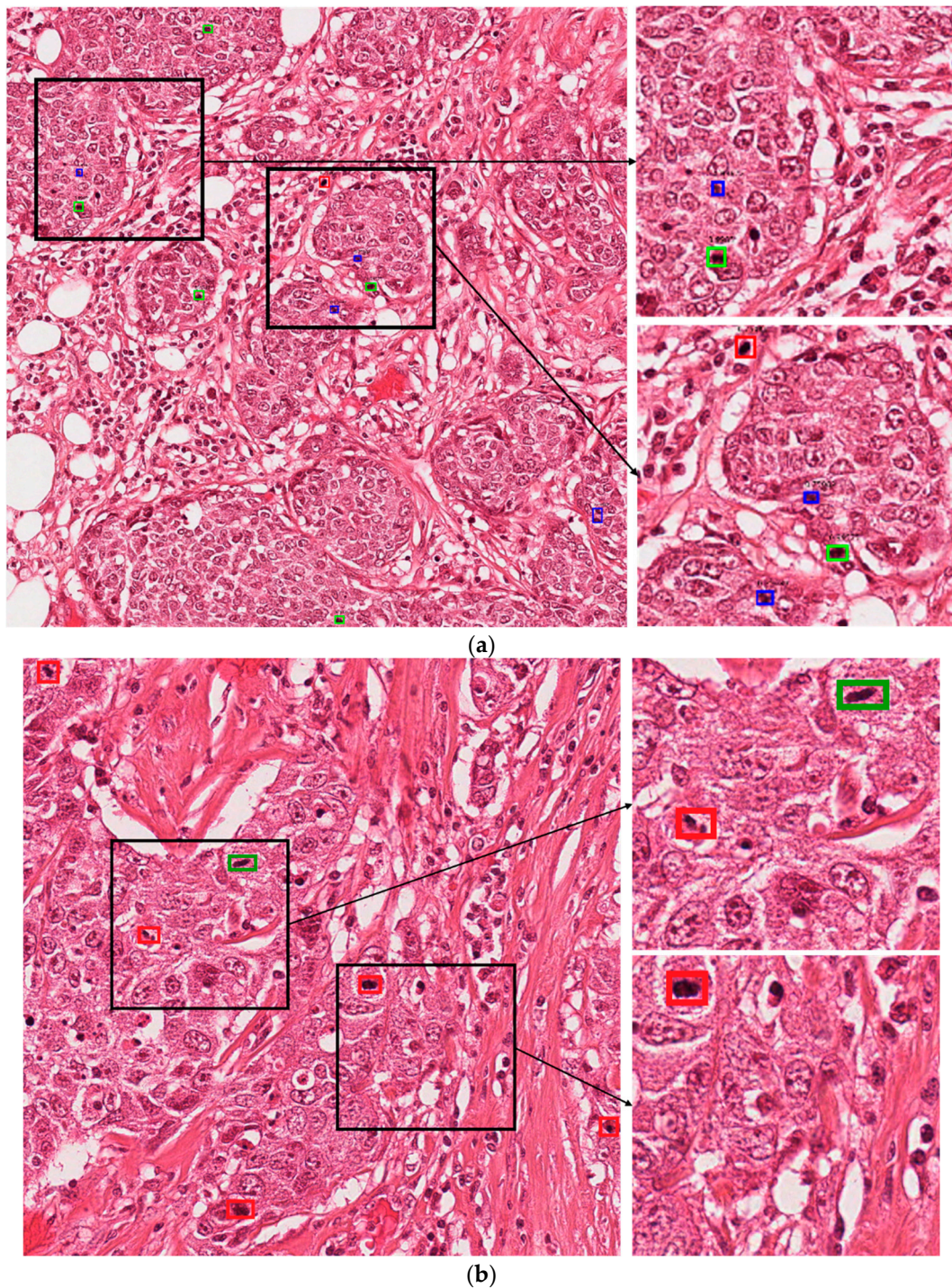
In clinical application, these mitotic cells are detected by a pathologist by observation of histopathology images under a high-resolution microscope. According to the Nottingham grading system, a pathologist must assign a score to three biomarkers for grading of the breast cancer. These biomarkers are tubule formation, nuclear pleomorphism and mitotic cell count [2]. A score from 1 to 3 is assigned to each of the biomarkers which are summed together for calculating the grade of breast cancer. In the case of mitotic cells, mitotic cell count from 10 high power fields are summed together and a score from 1 to 3 is assigned based on the number of mitotic cells. Score 1 is assigned for mitotic cell count 0 to 9, score 2 is assigned for mitotic cell count 10–19, and score 3 is assigned for mitotic cell count 20 or greater than 20 [61]. In the presented test images in Figures 4 and 5, Figure 4a can be assigned a score of 2, while in Figure 5b, score of 1 can be assigned if a pathologist finds the grade of the breast cancer. Along with other two biomarker's score, grade can be finalized.





**Figure 4.** Examples of correct-detection cases of proposed method with image from (a) ICPR 2012 and (b) ICPR 2014 datasets. Green boxes indicate true positives, red boxes indicate false positives, and blue boxes indicate false negatives.





**Figure 5.** Examples of incorrect-detection cases of proposed method with image from (a) ICPR 2012 and (b) ICPR 2014 datasets. Green boxes indicate true positives, red boxes indicate false positives, and blue boxes indicate false negatives.

#### 5.4.4. Cross-Dataset Experiment-TUPAC16

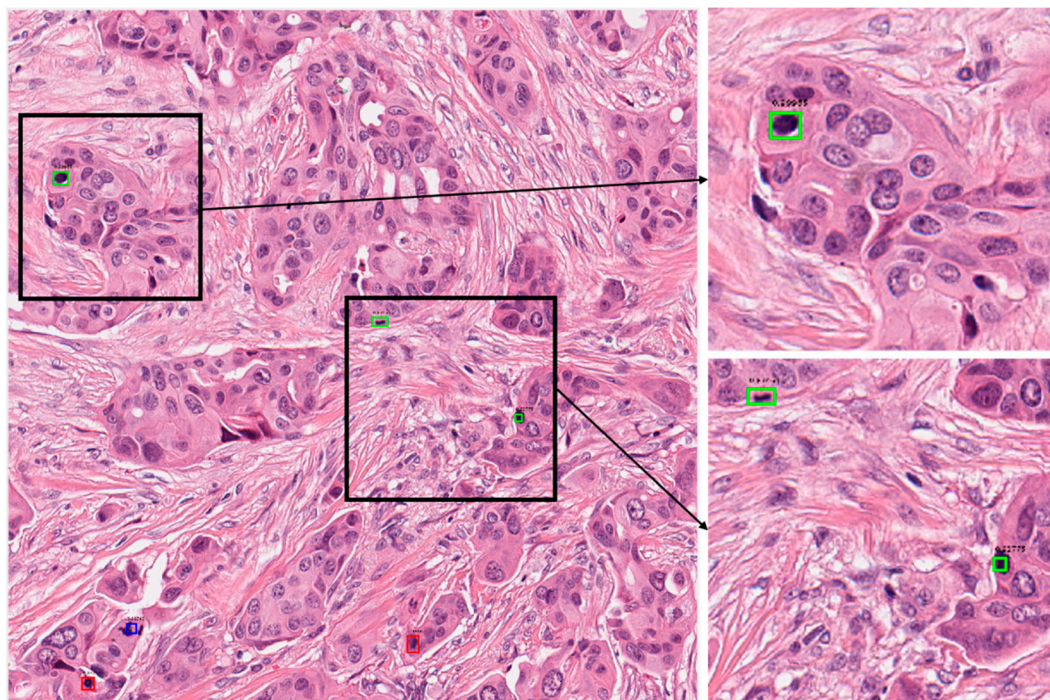
Automatic detection of mitotic cells is a crucial task for deep learning-based techniques. Such techniques should have a good generalization capability because cases could have variations due to tissue compositions and data acquisition environments. We tested the generalization capability of our technique on the TUPAC16 dataset. This dataset consists of cases from three pathology centers. There are 73 breast cancer cases in the train set and 34 cases in the test set. Each of the cases is represented



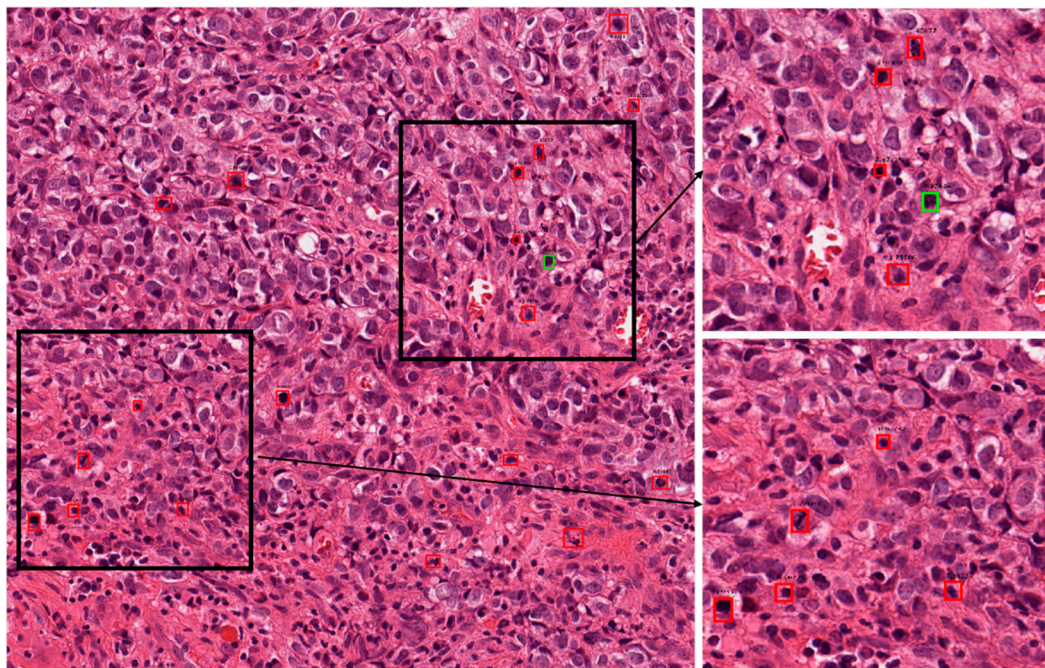
with an area of 2 mm<sup>2</sup>. As the ground truths of the test data are not publicly available, we used training set images for testing the generalization capability of our proposed technique. Table 6 presents the performance of our techniques trained on ICPR2012 and tested on the TUPAC16 dataset. As shown in Table 6, our method shows a little higher accuracy than Akram et al. [62]’s method, but a little lower accuracy than Paeng et al. [63]’s method (the first ranked in TUPAC16 contest). That is because the training and testing datasets for our method are different, whereas Paeng et al. [63]’s method used the same TUPAC16 datasets for training and testing. Therefore, the results on TUPAC16 dataset proved that our proposed technique has generalization ability to the level that it can be used for real-time clinical application as shown in Table 6. There are a lot of variations in staining and tissue composition among cases, but our technique was successful in detection as shown in Figures 6 and 7 where correct and incorrect detections are presented by the proposed technique. It can be observed that our technique had less performance than the state-of-the-art results, but we can state that our technique has a good generalization capability and can be improved further by tuning and training with the TUPAC16 dataset.

**Table 6.** Comparisons of the proposed method on cross-dataset TUPAC16 with ICPR 2012 dataset trained networks (TUPAC16, tumor proliferation assessment challenge 2016. N.R. means “not reported” and proposed method-12 indicates networks trained on ICPR2012 dataset).

Technique	Precision	Recall	F1-Measure
Akram et al. [62]	0.61	0.67	0.64
Paeng et al. [63]	N.R.	N.R.	0.652
Proposed method-12	0.641	0.642	0.642



**Figure 6.** Examples of correct-detection cases of proposed method with image from TUPAC16 dataset. Green boxes indicate true positives, red boxes indicate false positives, and blue boxes indicate false negatives. TUPAC16, tumor proliferation assessment challenge 2016.

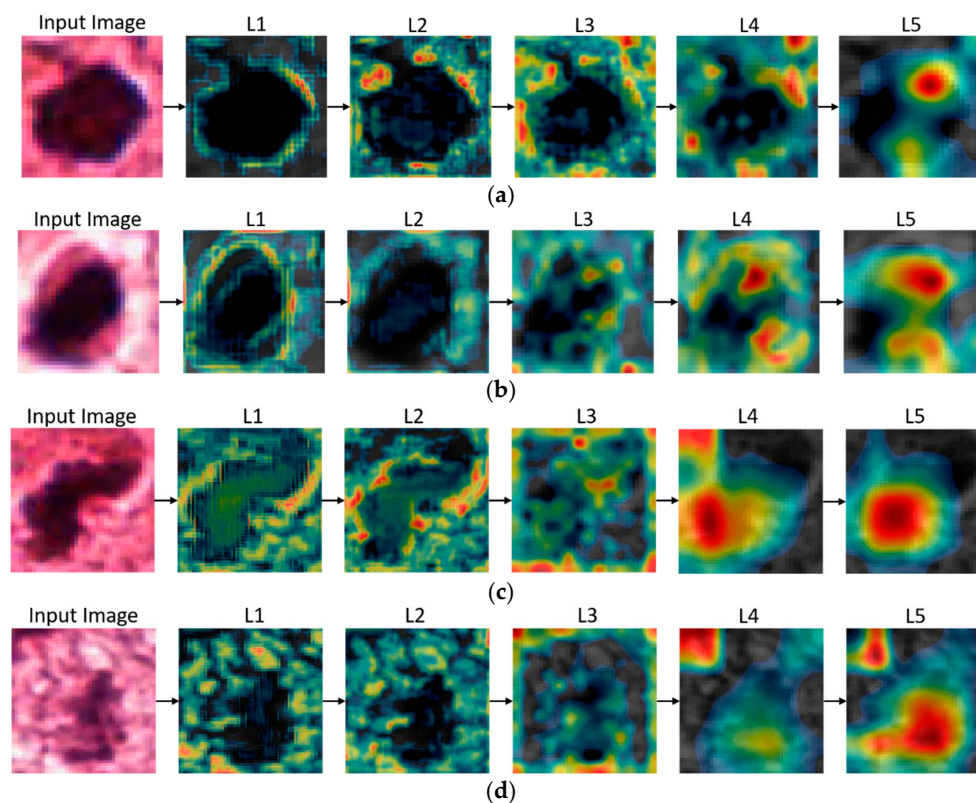


**Figure 7.** Examples of incorrect-detection cases of proposed method with image from TUPAC16 dataset. Green boxes indicate true positives, red boxes indicate false positives, and blue boxes indicate false negatives.

## 6. Activation Maps and Discussion

Deep-learning networks are often considered as “black boxes” because there is no clear explanation behind a specific decision. Hundreds of features are learned during training, which are then used for the decision making. The use of class-specific discriminative regions (activation maps) is one technique that can be used for the investigation of a network to identify the regions of images that are responsible for the decision making. Figure 8 presents the class-specific discriminative regions, also known as activation maps of an image when it passes through the different layers of our trained Resnet-50 and Densenet-201 classifiers. For presentation purpose, we consider the average activation map of a layer, followed by its representation in a pseudo-color scheme, in which red indicates the maximum value and blue indicates the minimum value. In detail, since all the layers of the networks are multi-dimensional along the feature channel axis, for presentation purpose we took the average of the activation maps along the feature channel axes of a layer. This way, we present a single image for a layer. Using these activation maps, we can obtain a visual explanation of the regions of an image that contribute the most in the decision making. It can be easily observed from Figure 8 that although there is no visible difference between the presented cases of mitotic and non-mitotic cells, our trained networks still find deep features to differentiate between the two cells. The earlier layers learn simple features such as color and edges, and the deeper layers learn complex features that have a significant impact on decision making. Moreover, we identified from class-activation maps that our training is not biased towards wrong ROIs. Moreover, as shown in Figure 8, there is noticeable characteristics in activation maps in the case of mitotic cells as compared to those of non-mitotic cells. That is, multiple regions are activated in case of non-mitotic cells, which implies that our trained models have the ability to successfully differentiate between mitotic and non-mitotic cells.





**Figure 8.** Obtained activation maps from different parts of (a), (b) Resnet-50, and (c), (d) Densenet-201 with mitotic and non-mitotic cell images. (a) and (c) comprise mitotic cells whereas (b) and (d) comprise non-mitotic cells. In (a) and (b), L1–L5 are the Resnet-50 layers Conv2-1, Conv3-4, Conv4-1, Conv4-6, and Conv5-3, respectively, as presented in Table A1, whereas L1–L5 in (c) and (d) are the Densenet-201 layers Convolution (1), Dense Block (1), Dense Block (2), Dense Block (3), and Dense Block (4), respectively.

The key observations made in this work are as follows:

- Our proposed-technique results show that recent advances in deep-learning algorithms have decreased the gap between diagnoses performed by human experts and computers. Moreover, a good performance with the ICPR 2012 and ICPR 2014 datasets has proved the generalization capabilities of our proposed technique, and thus, our technique may be used for various lesion detections.
- We have observed that significant variation exists in the sizes of the mitotic cells. Therefore, Faster R-CNN feature-extraction network and anchor-boxes selection play a key role in the detection of mitotic cells. By using Resnet-50 for feature extraction, we successfully extracted efficient features because Resnet-50 uses skip connections, and thus, the mitotic cell's information is not lost. Moreover, we fixed the anchor scale size to 64 instead of 128, 256, or 512 and selected anchor boxes that have an intersection-over-union value less than or equal to 0.5 with ground truths. Therefore, by using Resnet-50 as a feature-extraction network, fixing the anchor-scale size to 64, and limiting the number of anchor boxes, we achieved the state-of-the-art performance.
- We have also observed that Faster R-CNN also depends on the underlying feature-extraction network and RPN. Therefore, in our case Faster R-CNN rapidly converges in only 25 epochs, because of the use of Resnet-50 as a feature-extraction network, the smaller anchor scale, and the limited anchor boxes.
- We have observed that some of the false-positive cases comprise an irregular morphology and dark bluish color and have large variations in texture. These issues can be eliminated by using handcrafted features such as LBP, HOG, and statistical and color features for improving the performance.

- Mitotic-cell-detection techniques [2,24–27] comprise the use of additional classifiers for performance improvement. Although classifiers such as Resnet-50 and Densenet-201 exhibit an outstanding performance owing to Resnet-50's residual learning and skip connection for feature reusability and Densenet-201's feature propagation, feature reusability, and smaller number of parameters, we still observed that the performance can be further improved because single-modality data lack uniqueness and universality. Therefore, in our proposed technique we performed score-level fusion and improved our obtained results as compared to those of state-of-the-art methods.
- Deep-learning networks require a large amount of data for successful training. Owing to the lack of data, some other techniques such as data augmentation are used to increase the data. Data augmentation in the case of mitotic-cell detection is a challenging problem because there are minute differences between mitotic and non-mitotic cells. We observed that the flipping and translation technique for data augmentation produces robust data as proved by the activation maps in Figure 8, where trained classifiers successfully found features in the test data for decision making.

In general, pathologists examine the color and morphological appearances under high-resolution microscopes in order to differentiate between the mitotic and non-mitotic cells. Although pathologists are superior in decision making but these procedures are tiresome and tedious due to extensive similarities among cells. Therefore, AI-based techniques can assist pathologists by pointing out the positive cases. Different AI-based techniques are used to assist pathologists. Our proposed technique can also be used to assist pathologists because we have outperformed the existing techniques for mitotic-cells detection by using a combination of Faster R-CNN, post-processing, and score-level fusion.

## 7. Conclusions

In this paper, we presented a multi-stage mitotic-cell-detection technique based on Faster-RCNN, post-processing, and deep CNNs. Faster RCNN performs the initial detection in stage 1, which is further improved by post-processing and score-level fusion of the deep CNNs in stages 2 and 3, respectively. The performance thus obtained comprises the combined detections of all the stages. The strengths of this work comprise the adoption of Resnet-50 for the feature extraction in the Faster R-CNN, extraction of appropriate features in the post-processing, and score-level fusion of the Resnet-50 and Densenet-201 classifiers. Moreover, our proposed technique outperforms all the existing state-of-the-art techniques on the two open databases of ICPR 2012 and ICPR 2014. This proposed technique can be used for developing an automatic grading system that could serve as a second-opinion system for pathologists. It can also highlight cases that require special attention. This research is useful for pathologists and researchers working in the field of cancer diagnosis based on histology images. Future research work can be based on our publicly available trained models, and a fair comparison of results is also possible.

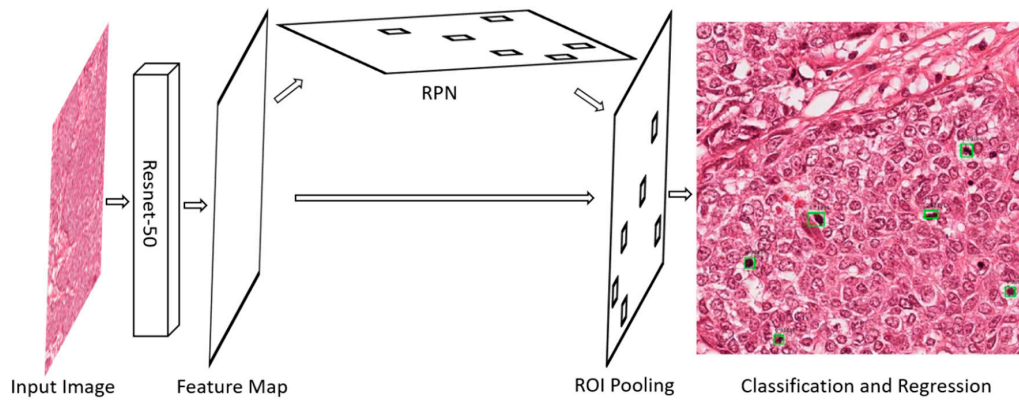
In the future, we intend to improve the accuracies and decrease the computational cost of the presented technique by developing a customized network. We intend to validate our technique using larger databases that provide a diversity of breast cancer cases. We also plan to increase the scope of this technique to applications that combine the biomarkers of nuclear atypia, tubule formation, and mitotic-cell count for the diagnosis and grading of breast cancers.

**Author Contributions:** T.M. and K.R.P. designed the overall system. In addition, they wrote and revised the paper. M.A., M.O., and M.B.L. helped to design comparative analysis and experiments. All authors have read and agreed to the published version of the manuscript.

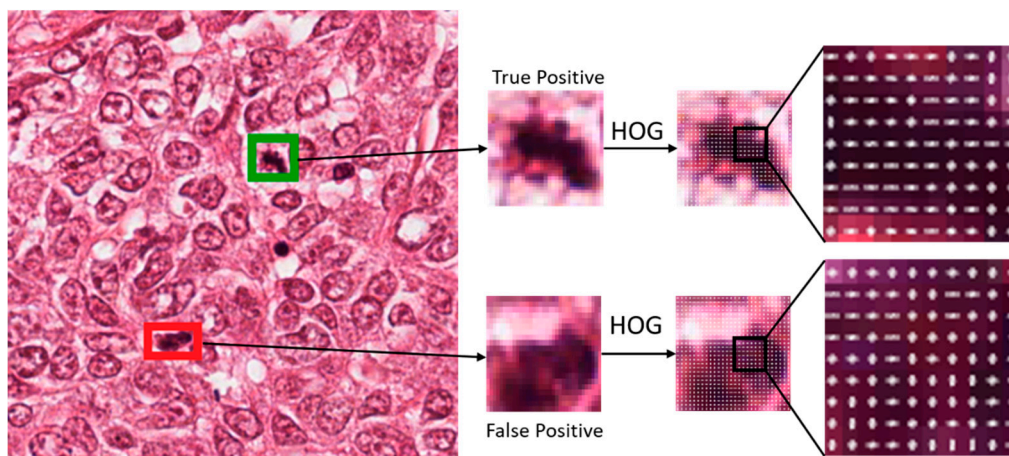
**Acknowledgments:** This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2018R1D1A1B07041921), by the NRF funded by the Ministry of Science and ICT through the Basic Science Research Program (NRF-2019R1F1A1041123), and by the Bio & Medical Technology Development Program of the NRF funded by the Korean government, MSIT (NRF-2016M3A9E1915855).

**Conflicts of Interest:** The authors declare no conflict of interest.

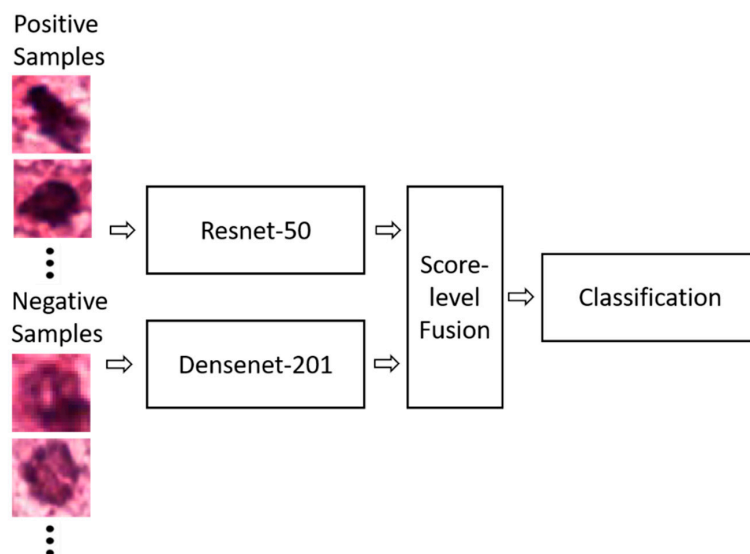
**Appendix A**



**Figure A1.** Faster R-CNN architecture; the feature map is extracted from the input image using the Resnet-50 feature-extraction network followed by the generation of region proposals in the RPN and the final mitotic-cells detection in the classification network [28].



**Figure A2.** HOG features from true-positive (green box) and false-positive (red box) images.



**Figure A3.** Score-level fusion of Resnet-50 and Densenet-201 and classification of mitotic cells.

**Table A1.** Resnet-50 feature-extraction network (Conv means convolutional layer).

Layer Type	Output Size	Numbers of Filters	Kernel Size	Strides	Paddings	Iterations	
Image input layer	224 × 224 × 3						
Conv1	112 × 112 × 64	64	7 × 7 × 3	2	3	1	
Maximum pool	55 × 55 × 64	1	3 × 3	2	0	1	
Conv2	Conv2-1 (1 × 1 Convolutional Mapping)	55 × 55 × 64 64 55 × 55 × 256 256	64 64 256 256	1 × 1 × 64 3 × 3 × 64 1 × 1 × 64 1 × 1 × 64	1 1 1 1	0 1 0 0	1
	Conv2-2-Conv2-3 (Identity Mapping)	55 × 55 × 64 55 × 55 × 64 55 × 55 × 256	64 64 256	1 × 1 × 256 3 × 3 × 64 1 × 1 × 64	1 1 1	0 1 0	2
	Conv3-1 (1 × 1 Convolutional Mapping)	28 × 28 × 128 28 × 28 × 128 28 × 28 × 512 28 × 28 × 512	128 128 512 512	1 × 1 × 256 3 × 3 × 128 1 × 1 × 128 1 × 1 × 256	2 1 1 2	0 1 0 0	1
	Conv3-2-Conv3-4 (Identity Mapping)	28 × 28 × 128 28 × 28 × 128 28 × 28 × 512	128 128 512	1 × 1 × 512 3 × 3 × 128 1 × 1 × 128	1 1 1	0 1 0	3
	Conv4-1 (1 × 1 Convolutional Mapping)	14 × 14 × 256 14 × 14 × 256 14 × 14 × 1024 14 × 14 × 1024	256 256 1024 1024	1 × 1 × 512 1 × 1 × 512 1 × 1 × 512 1 × 1 × 512	2 1 1 2	0 1 0 0	1
Conv4-2-Conv4-6 (Identity Mapping)	14 × 14 × 256 14 × 14 × 256 14 × 14 × 1024	256 256 1024	1 × 1 × 1024 1 × 1 × 256 1 × 1 × 256	1 1 1	0 1 0	5	
Conv5	Conv5-1 (1 × 1 Convolutional Mapping)	7 × 7 × 512 7 × 7 × 512 7 × 7 × 2048 7 × 7 × 2048	512 512 2048 2048	1 × 1 × 1024 3 × 3 × 512 1 × 1 × 512 1 × 1 × 1024	2 1 1 2	0 1 0 0	1
	Conv5-2-Conv5-3 (Identity Mapping)	7 × 7 × 512 7 × 7 × 512 7 × 7 × 2048	512 512 2048	1 × 1 × 2048 3 × 3 × 512 1 × 1 × 512	1 1 1	0 1 0	2

**Table A2.** Region proposal network architecture (CL indicates convolutional layer).

Layer Type	Number of Filters	Output Size	Kernel Size	Strides	Paddings
5_3rd CL Input layer		7 × 7 × 2048			
6th CL (ReLU)	512	7 × 7 × 2048	3 × 3 × 512	1	1
Classification CL (Softmax)	18	7 × 7 × 18	1 × 1 × 512	1	0
6th CL Regression CL	36	7 × 7 × 36	1 × 1 × 512	1	0

**Table A3.** Classification network (ROI coordinate\* comprises x\_min, y\_min, x\_max, and y\_max of ROI of each proposal).

Layer Type	Output Size
5_3rd CL RPN proposal region Input layer	7 × 7 × 2048 (height × width × depth) 300 × 4 (ROI coordinate *)
ROI pooling layer	7 × 7 × 512 (height × width × depth) × 300
1st fully connected layer (ReLU) (Dropout)	4096 × 300
2nd fully connected layer (ReLU) (Dropout)	4096 × 300
Classification convolutional layer (Softmax)	2 × 300
2nd Fully connected layer Regression fully connected layer	4 × 300

RPN, region proposal network.



## References

1. Ghoncheh, M.; Pournamdar, Z.; Salehiniya, H. Incidence and mortality and epidemiology of breast cancer in the world. *Asian Pac. J. Cancer Prev.* **2016**, *17*, 43–46. [[CrossRef](#)]
2. Li, C.; Wang, X.; Liu, W.; Latecki, L.J. DeepMitosis: Mitosis detection via deep detection, verification and segmentation networks. *Med. Image Anal.* **2018**, *45*, 121–133. [[CrossRef](#)] [[PubMed](#)]
3. Arsalan, M.; Kim, D.S.; Lee, M.B.; Owais, M.; Park, K.R. FRED-Net: Fully residual encoder–decoder network for accurate iris segmentation. *Expert Syst. Appl.* **2019**, *122*, 217–241. [[CrossRef](#)]
4. Sajjad, M.; Khan, S.; Muhammad, K.; Wu, W.; Ullah, A.; Baik, S.W. Multi-grade brain tumor classification using deep CNN with extensive data augmentation. *J. Comput. Sci.* **2019**, *30*, 174–182. [[CrossRef](#)]
5. Khan, S.; Islam, N.; Jan, Z.; Ud Din, I.; Rodrigues, J.J.P.C. A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. *Pattern Recognit. Lett.* **2019**, *125*, 1–6. [[CrossRef](#)]
6. Arsalan, M.; Kim, D.S.; Owais, M.; Park, K.R. OR-Skip-Net: Outer residual skip network for skin segmentation in non-ideal situations. *Expert Syst. Appl.* **2020**, *141*, 1–26. [[CrossRef](#)]
7. Lakshmanaprabu, S.K.; Mohanty, S.N.; Shankar, K.; Arunkumar, N.; Ramirez, G. Optimal deep learning model for classification of lung cancer on CT images. *Futur. Gener. Comp. Syst.* **2019**, *92*, 374–382.
8. Owais, M.; Arsalan, M.; Choi, J.; Park, K.R. Effective diagnosis and treatment through content-based medical image retrieval (CBMIR) by using artificial intelligence. *J. Clin. Med.* **2019**, *8*, 462. [[CrossRef](#)]
9. Arsalan, M.; Owais, M.; Mahmood, T.; Cho, S.W.; Park, K.R. Aiding the diagnosis of diabetic and hypertensive retinopathy using artificial intelligence-based semantic segmentation. *J. Clin. Med.* **2019**, *8*, 1446. [[CrossRef](#)]
10. Karim, A.; Mishra, A.; Newton, M.A.H.; Sattar, A. Efficient toxicity prediction via simple features using shallow neural networks and decision trees. *Acs Omega* **2019**, *4*, 1874–1888. [[CrossRef](#)]
11. Veta, M.; van Diest, P.J.; Willems, S.M.; Wang, H.; Madabhushi, A.; Cruz-Roa, A.; Gonzalez, F.; Larsen, A.B.L.; Vestergaard, J.S.; Dahl, A.B.; et al. Assessment of algorithms for mitosis detection in breast cancer histopathology images. *Med. Image Anal.* **2015**, *20*, 237–248. [[CrossRef](#)] [[PubMed](#)]
12. Irshad, H. Automated mitosis detection in histopathology using morphological and multi-channel statistics features. *J. Pathol. Inf.* **2013**, *4*, 1–6. [[CrossRef](#)] [[PubMed](#)]
13. Tashk, A.; Helfroush, M.S.; Danyali, H.; Akbarzadeh, M. An automatic mitosis detection method for breast cancer histopathology slide images based on objective and pixel-wise textural features classification. In Proceedings of the 5th Conference on Information and Knowledge Technology, Shiraz, Iran, 28–30 May 2013; pp. 406–410.
14. Sommer, C.; Fiaschi, L.; Hamprecht, F.A.; Gerlich, D.W. Learning-based mitotic cell detection in histopathological images. In Proceedings of the 21st International Conference on Pattern Recognition, Tsukuba, Japan, 11–15 November 2012; pp. 2306–2309.
15. Paul, A.; Dey, A.; Mukherjee, D.P.; Sivaswamy, J.; Tourani, V. Regenerative random forest with automatic feature selection to detect mitosis in histopathological breast cancer images. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 94–102.
16. Hameed, A.A.; Karlik, B.; Salman, M.S. Back-propagation algorithm with variable adaptive momentum. *Knowl. Based Syst.* **2016**, *114*, 79–87. [[CrossRef](#)]
17. Ciresan, D.C.; Giusti, A.; Gambardella, L.M.; Schmidhuber, J. Mitosis detection in breast cancer histology images with deep neural networks. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention, Nagoya, Japan, 22–26 September 2013; pp. 411–418.
18. Malon, C.D.; Cosatto, E. Classification of mitotic figures with convolutional neural networks and seeded blob features. *J. Pathol. Inform* **2013**, *4*, 1–5. [[CrossRef](#)]
19. Wang, H.; Cruz-Roa, A.; Basavanthally, A.; Gilmore, H.; Shih, N.; Feldman, M.; Tomaszewski, J.; Gonzalez, F.; Madabhushi, A. Cascaded ensemble of convolutional neural networks and handcrafted features for mitosis detection. In Proceedings of the SPIE Medical Imaging, San Diego, CA, USA, 15–20 February 2014; pp. 1–10.
20. Chen, H.; Dou, Q.; Wang, X.; Qin, J.; Heng, P.-A. Mitosis detection in breast cancer histology images via deep cascaded networks. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; pp. 1160–1166.

21. Madzarov, G.; Gjorgjevikj, D.; Chorbev, I. A Multi-class SVM classifier utilizing binary decision tree. *Informatica* **2009**, *33*, 233–241.
22. Sommer, C.; Straehle, C.; Kothe, U.; Hamprecht, F.A. Ilastik: Interactive learning and segmentation toolkit. In Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Chicago, IL, USA, 30 March–2 April 2011; pp. 230–233.
23. Held, M.; Schmitz, M.H.A.; Fischer, B.; Walter, T.; Neumann, B.; Olma, M.H.; Peter, M.; Ellenberg, J.; Gerlich, D.W. CellCognition: time-resolved phenotype annotation in high-throughput live cell imaging. *Nat. Methods* **2010**, *7*, 747–754. [[CrossRef](#)]
24. Li, C.; Wang, X.; Liu, W.; Latecki, L.J.; Wang, B.; Huang, J. Weakly supervised mitosis detection in breast histopathology images using concentric loss. *Med. Image Anal.* **2019**, *53*, 165–178. [[CrossRef](#)]
25. Cai, D.; Sun, X.; Zhou, N.; Han, X.; Yao, J. Efficient mitosis detection in breast cancer histology images by RCNN. In Proceedings of the IEEE 16th International Symposium on Biomedical Imaging, Venice, Italy, 8–11 April 2019; pp. 919–922.
26. Li, Y.; Mercan, E.; Knezevitch, S.; Elmore, J.G.; Shapiro, L.G. Efficient and accurate mitosis detection—A lightweight RCNN approach. In Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods, Funchal, Portugal, 16–18 January 2018; pp. 69–77.
27. Dodballapur, V.; Song, Y.; Huang, H.; Chen, M.; Chrzanowski, W.; Cai, W. Mask-driven mitosis detection in histopathology images. In Proceedings of the IEEE 16th International Symposium on Biomedical Imaging, Venice, Italy, 8–11 April 2019; pp. 1855–1859.
28. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [[CrossRef](#)]
29. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
30. Sengupta, A.; Ye, Y.; Wang, R.; Liu, C.; Roy, K. Going deeper in spiking neural networks: VGG and residual architectures. *Front. Neurosci.* **2019**, *13*, 1–10. [[CrossRef](#)]
31. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
32. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1800–1807.
33. Available online: <http://dm.dgu.edu/link.html> (accessed on 24 November 2019).
34. ImageNet Large Scale Visual Recognition Challenge 2015 (ILSVRC2015). Available online: <http://image-net.org/challenges/LSVRC/2015/> (accessed on 24 November 2019).
35. Zhong, Z.; Sun, L.; Huo, Q. An anchor-free region proposal network for Faster R-CNN-based text detection approaches. *Int. J. Doc. Anal. Recognit.* **2019**, *22*, 315–327. [[CrossRef](#)]
36. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
37. Girshick, R. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
38. Uijlings, J.R.R.; van de Sande, K.E.A.; Gevers, T.; Smeulders, A.W.M. Selective search for object recognition. *Int. J. Comput. Vis.* **2013**, *104*, 154–171. [[CrossRef](#)]
39. Ahonen, T.; Hadid, A.; Pietikäinen, M. Face recognition with local binary patterns. In Proceedings of the European Conference on Computer Vision, Prague, Czech Republic, 11–14 May 2004; pp. 469–481.
40. Mahmood, T.; Ziauddin, S.; Shahid, A.R.; Safi, A. Mitosis detection in breast cancer histopathology images using statistical, color and shape-based features. *J. Med. Imaging Health Inf.* **2018**, *8*, 932–938. [[CrossRef](#)]
41. Ojala, T.; Pietikäinen, M.; Harwood, D. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognit.* **1996**, *29*, 51–59. [[CrossRef](#)]
42. Kalakech, M.; Porebski, A.; Vandenbroucke, N.; Hamad, D. Unsupervised local binary pattern histogram selection scores for color texture classification. *J. Imaging* **2018**, *4*, 112. [[CrossRef](#)]
43. Sicilia, R.; Cordelli, E.; Merone, M.; Luperto, E.; Papalia, R.; Iannello, G.; Soda, P. Early radiomic experiences in classifying prostate cancer aggressiveness using 3D local binary patterns. In Proceedings of the IEEE 32nd International Symposium on Computer-Based Medical Systems, Cordoba, Spain, 5–7 June 2019; pp. 355–360.

44. Nguyen, D.T.; Cho, S.R.; Shin, K.Y.; Bang, J.W.; Park, K.R. Comparative study of human age estimation with or without preclassification of gender and facial expression. *Sci. World J.* **2014**, *2014*, 1–15. [[CrossRef](#)] [[PubMed](#)]
45. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 20–25 June 2005; pp. 1–8.
46. Nazir, M.; Jan, Z.; Sajjad, M. Facial expression recognition using histogram of oriented gradients based transformed features. *Clust. Comput.* **2018**, *21*, 539–548. [[CrossRef](#)]
47. Lee, W.-Y.; Ko, K.-E.; Sim, K.-B. Robust lip detection based on histogram of oriented gradient features and convolutional neural network under effects of light and background. *Optik* **2017**, *136*, 462–469. [[CrossRef](#)]
48. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269.
49. He, M.; Horng, S.; Fan, P.; Run, R.-S.; Chen, R.-J.; Lai, J.-L.; Khan, M.K.; Sentosa, K.O. Performance evaluation of score level fusion in multimodal biometric systems. *Pattern Recognit.* **2010**, *43*, 1789–1800. [[CrossRef](#)]
50. Yılmaz, M.B.; Yanıkoğlu, B. Score level fusion of classifiers in off-line signature verification. *Info. Fusion* **2016**, *32*, 109–119. [[CrossRef](#)]
51. Castrillón-Santana, M.; Lorenzo-Navarro, J.; Ramón-Balmaseda, E. Multi-scale score level fusion of local descriptors for gender classification in the wild. *Multimed. Tools Appl.* **2017**, *76*, 4695–4711. [[CrossRef](#)]
52. Ross, A. and Jain, A. Information fusion in biometrics. *Pattern Recognit. Lett.* **2003**, *24*, 2115–2125. [[CrossRef](#)]
53. Roux, L.; Racoceanu, D.; Loménie, N.; Kulikova, M.; Irshad, H.; Klossa, J.; Capron, F.; Genestie, C.; Le Naour, G.; Gurcan, M.N. Mitosis detection in breast cancer histological images an ICPR 2012 contest. *J. Pathol. Inf.* **2013**, *4*, 1–7. [[CrossRef](#)] [[PubMed](#)]
54. MITOS-ATYPIA-14 Grand Challenge. Available online: <https://mitos-atypia-14.grand-challenge.org/> (accessed on 14 November 2019).
55. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
56. MATLAB R2019a at a Glance. Available online: [https://www.mathworks.com/products/new\\_products/release2019a.html](https://www.mathworks.com/products/new_products/release2019a.html) (accessed on 10 November 2019).
57. Intel®Core i7-3770K Processor. Available online: <http://ark.intel.com/content/www/us/en/ark/products/65523/intel-core-i7-3770k-processor-8m-cache-up-to-3-90-ghz.html> (accessed on 12 November 2019).
58. GeForce GTX 1070. Available online: <https://www.nvidia.com/ko-kr/geforce/products/10series/geforce-gtx-1070-ti/> (accessed on 12 November 2019).
59. Dogo, E.M.; Afolabi, O.J.; Nwulu, N.I.; Twala, B.; Aigbavboa, C.O. A comparative analysis of gradient descent-based optimization algorithms on convolutional neural networks. In Proceedings of the International Conference on Computational Techniques, Electronics, and Mechanical Systems, Belgaum India, 21–22 December 2018; pp. 92–99.
60. Mitosis Detection in Breast Cancer Histological Images (MITOS dataset). Available online: [http://ludo17.free.fr/mitos\\_2012/results.html](http://ludo17.free.fr/mitos_2012/results.html) (accessed on 10 November 2019).
61. Elston, C.W.; Ellis, I.O. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: Experience from a large study with long-term follow-up. *Histopathology* **2002**, *41*, 154–161. [[CrossRef](#)] [[PubMed](#)]
62. Akram, S.U.; Qaiser, T.; Graham, S.; Kannala, J.; Heikkilä, J.; Rajpoot, N. Leveraging unlabeled whole-slide images for mitosis detection. In Proceedings of the International Workshop on Ophthalmic Medical Image Analysis, Granada, Spain, 16–20 September 2018; pp. 69–77.
63. Paeng, K.; Hwang, S.; Park, S.; Kim, M. A unified framework for tumor proliferation score prediction in breast histopathology. In Proceedings of the 3rd International Workshop on Deep Learning in Medical Image Analysis, Quebec City, QC, Canada, 10–14 September 2017; pp. 231–239.

