

# A fast pyramidal Bayesian model for mitosis detection in whole-slide images

Santiago López-Tapia<sup>1</sup>, José Aneiros-Fernández<sup>2</sup>, and Nicolás Pérez de la Blanca<sup>1\*</sup>

<sup>1</sup> Computer Science and Artificial Intelligence Department, University of Granada, Granada, Spain,

`sltapia@decsai.ugr.es`, `nicolas@ugr.es`,

<sup>2</sup> Intercenter Unit of Pathological Anatomy, San Cecilio University Hospital, Granada, Spain

`janeirosf@hotmail.com`

**Abstract.** Mitosis detection in Hematoxylin and Eosin images and its quantification for  $\text{mm}^2$  is currently one of the most valuable prognostic indicators for some types of cancer and specifically for the breast cancer. In whole-slide images the main goal is to detect its presence on the full image. This paper makes several contributions to the mitosis detection task in whole-slide in order to improve the current state of the art and efficiency. A new coarse to fine pyramidal model to detect mitosis is proposed. On each pyramid level a Bayesian convolutional neural network is trained to compute class prediction and uncertainty on each pixel. This information is propagated top-down on the pyramid as a constraining mechanism from the above layers. To cope with local tissue and cell shape deformations geometric invariance is also introduced as a part of the model. The model achieves an F1-score of 82.6% on the MITOS ICPR-2012 test dataset when trained with samples from skin tissue. This is competitive with the current state of the art. In average a whole-slide is analyzed in less than 20 seconds. A new dataset of 8236 mitoses from skin tissue has been created to train our models.

**Keywords:** Mitosis detection, Pyramid, Bayesian model, Multiscale processing

## 1 Introduction

The quantification of mitotic cells in Hematoxylin and Eosin (H&E) images and more specifically its density per square millimeter is one of the current most stronger markers in cancer prognosis.

The advent of the high-resolution scanner technology to the computational pathology field has allowed to obtain digital whole-slides images (WSI). Nevertheless, the huge size of the images and the computing time of the current

---

\* This paper has been supported by the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund (FEDER) under the grant DPI2016-77869-C2-2-R

detection algorithms impose in practice a partial rather than a fully image detection and counting.

Several difficulties can be identified as responsible of the current low detection rate on H&E stained images. On one hand, the variability in RGB color map due to different stain intensities and scanners technology [12, 9]. On the other hand, the presence of very hard false positives due to Hematoxylin staining of non-cells tissue also makes harder the detection process. In addition, the mitosis undergoes four different stages with different shapes and appearances. This geometric variability and the low number of mitosis pixels per WSI also represent a new source of false positive. These difficulties all together make the design of an efficient and accurate mitosis detection algorithms a challenge task [16].

Different Challenges such as TUPAC-2016[10], MITOS-ATYPIA [5] and MITOS ICPR-2012 [13] have been organized in the last years to foster the detection algorithms. But the contributed datasets from them all are too small and only from breast cancer tissue. Currently, there are no other larger open access mitosis datasets. We have created a mitosis dataset from skin cancer images to train our model. In this type of cancer mitosis detection is also a very relevant prognostic indicator[14]. In order to compare our model with other results in the literature we have tested with MITOS ICPR-2012.

## 2 Related works

Many contributions to the use of CNN model for mitosis detection have been proposed since the ICPR-2012 challenge MITOS ICPR-2012 [13] was available [3, 17, 2]. The best result from all these approaches is an F1 score of 78.8%. In [8] an adaptation to the general object detection framework from CNN, Faster R-CNN, is proposed. They focus on the use of very deep architectures for mitosis detection achieving an F1-score of 83.2% in MITOS ICPR-2012. More recently in [15] a new way of approaching the detection task is proposed. They stain twice each slide using Phospho-histone H3 (PHH3) and H&E and leverage on the complementary properties of these stains to improve the detection. They succeed in removing many of the false positives but at the cost of a very complex processing. Our method addresses a similar goal but from a pyramidal approach. All mentioned approaches exploit the depth increment in the architectures as the main mechanism to generate good features. In [19] an approach inspired in Wide Residual networks (WRN) [18] focus on the wide of the layer, instead of the number of layers. This fact simplifies the architecture making it more efficient at test time and easier of training. They reached an F1 score of 64.8% in the challenge TUPAC-2016 [10], which is a result competitive with the state of the art for this dataset. Our architectures are inspired by this network.

The feature extraction stage of all above approaches either use the 40x scale or use a fine to coarse feature pyramid starting in 40x. In both cases the highest resolution scale is the input information. In contrast, here we propose a coarse to fine approach in a top-down pass through a pyramid representing three scales of

the image. We find benefits in both efficiency and accuracy. The standard CNN models lack uncertainty measurements about the predictions as well as specific layers to obtain invariance to geometric deformations. The use of a Bayesian approach to CNN allows us to compute uncertainty in a natural way. On each internal pyramid level, prediction and uncertainty from the above levels are used as input to improve the final model prediction. We find that information from lower resolutions allow us to constraint the optimization process at the highest resolution. In addition, and to cope with both the cell shape variability induced by the phases of the mitosis and the tissue local deformations, our model incorporate specific layers to compute geometric invariant features [6].

In summary our contributions are: a) A new and fast pyramidal mitosis detection algorithm for WSI achieving a F1-score competitive with the state of the art on MITOS ICPR-2012 dataset; b) A new information propagation mechanism between scales from a cascade of Bayesian CNN model; c) The use of uncertainty and geometric invariance to improve the detection score; d) A model able of learning knowledge transfer between tissues; f) Mitosis detection time on WSI faster than ever before.

The rest of the paper is as follows. Section.3 we describes the model. Section.4 describes the training and test stages. Section.5 shows the experiment and Section.6 show the discussion and conclusions.

### 3 MODEL DESCRIPTION

Our model is defined as a forward cascade of classifiers applied on a course-to-fine image pyramid build from a WSI at three magnification scales 10x, 20x and 40x. We assume 40x represents the sample image and the lowest pyramid level. Fig.1 shows a diagram of the architecture. On each pyramid level a Bayesian CNN classifier inspired in the design of a Wide Residual Network [18] is trained. The three classifiers in the cascade output a mask of detected mitosis, a feature map, and the uncertainty per feature in terms of standard deviation as shown in Fig.1.

These feature maps are used by the next detectors, top-down, as soft constraints to focus the training on the most difficult negative samples (see Fig.1). We call this model PB-CNN. Furthermore, to make the model resistant to local and shape deformations, appearing by both the process of collecting and staining tissue and cells shape deformation, a Spatial Transforming Layer[6] is applied before the residual blocks 4th and 7th in scale x40 (see Fig.2). We call this model PB-CNN-STP. On the output of last classifier, we put to zero the predictions of those pixels which uncertainty is higher than a threshold fixed in training. The experiments show that these higher values are usually associated to WSI artifacts of low frequency in the training dataset. Finally, a non-maximum suppression step is carried out to keep, in cluster of overlapping regions, only the one with the highest probability. Our final output is a list of coordinates joint to their corresponding probability and standard deviation. As it can be seen in Fig.2, we use a late fusion criteria incorporating feature maps and uncertainty,

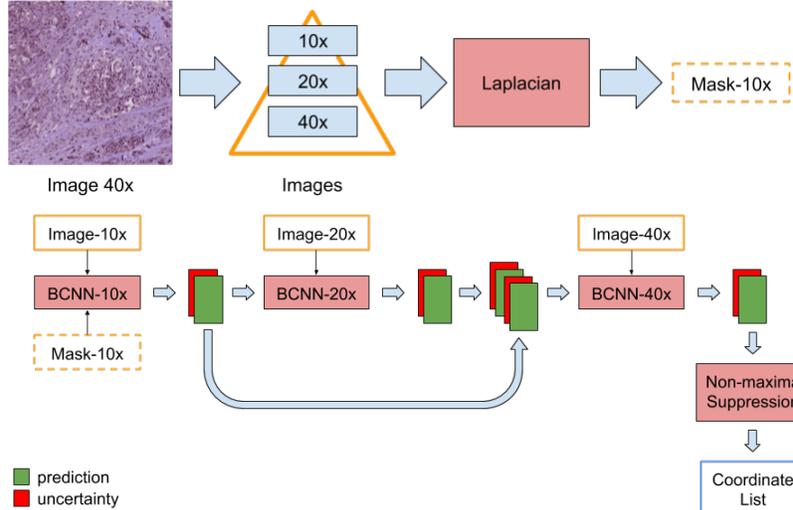


Fig. 1: Diagram showing the pyramidal model and the cascade of classifier to process the pyramid. The top of the figure shows, in this order, the input image, the pyramid building and the computation of the initial mask at 10x. The bottom shows how the first two pyramid levels provide input information to the third pyramid level. The result is the output of a non-maxima suppression process. See details in the text.

of the above levels, at the end of network. We have found in our experiments that this late fusion of features provides better results than doing it earlier.

The architecture of our detector is shown in Fig.2. The architecture is a Wide Residual Network [18] that uses three Wide Residual Units (WRU) (see right block). To reduce the spatial size of processed patch, we use a stride of 2 at certain layers (indicated by  $/2$  in the figure), in the case of the WRU block, the stride is applied at the first convolutional layer. The same architecture is used for PB-CNN-SPT adding a Spatial Transforming Layer[6] at the scale 40x as indicated previously.

## 4 Training and Test

### 4.1 Dataset

The dataset is created from 22 WSI of melanoma skin cancer. The images were acquired using a scanner Philips with a resolution of 0.25 micron per pixel. A senior pathologist of the Unit of Computational Pathology of the University Hospital San Cecilio in Granada labeled the WSI at 40x by indicating the center of the mitosis. 8236 were annotated mitosis.

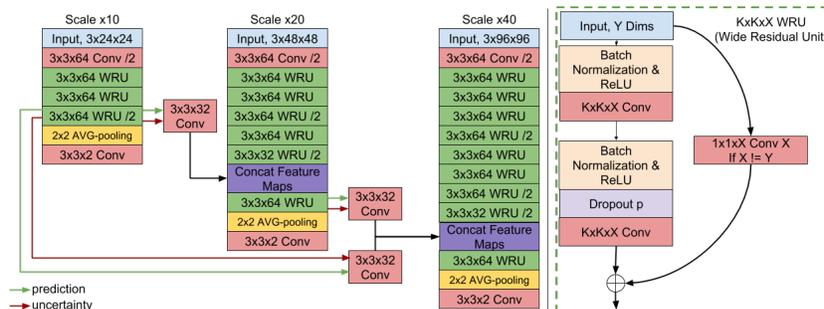


Fig. 2: Network architecture used for PB-CNN (see left figure).  $K \times K \times N$  Conv indicates a convolution layer with  $K \times K$  kernel and  $N$  filters.  $/2$  indicates that the convolution layer uses a stride of 2.  $2 \times 2$  AVG-pooling indicates an average pooling layer of kernel  $2 \times 2$ .

## 4.2 Input mask at 10x

To detect initial relevant regions at 10x we apply a Laplacian of the Gaussian (LoG) filter with  $\sigma = 9$  over the Hematoxylin band obtained by color deconvolution [9]. Then a thresholding for negative values that are less than  $-0.28$  is applied. We select the windows centered at each connected component as possible candidates to contain a mitosis. A window of size  $24 \times 24$  pixels is used.

## 4.3 Learning and testing

Let's denote by GTL the pyramid ground-truth labeling defined by the coordinates of mitosis centers in all scales. A strong labeling pyramid (MGTL) is generated, by labeling as 1 those windows inside circles of fixed radius centered at the GTL's mitosis centers. Radius of 96, 48 and 24 pixels are used for 40x, 20x and 10x scale respectively. The windows of label 0 on each level are computed in runtime as the difference between the MGTL mask and the mask obtained by thresholding and extrapolating the predicted probabilities from the above pyramid level, let denote it as PR. The used threshold is fixed in training time in order to keep all GTL windows within the class 1. The label 0 at each level represents hard false positive, since not being mitosis were predicted as such with high probability by the above level. In the case of the first level (10x), the mask computed in section 4.2 is used as PR. At each pyramid level the negative training samples are obtained by sampling of the mask of class 0. The positive samples are patches centered at the coordinates indicated in GTL. The patch size used is  $24 \times 24$ ,  $48 \times 48$  and  $96 \times 96$  for scales 10x, 20x and 40x, respectively. Before extracting the patches for training, we use the stain normalization algorithm proposed in [12] to reduce the variation in the training dataset. This normalization is also used during testing before the WSI is processed.

Each classifier in the pyramid is trained for 90 epochs with the Adam optimizer [7] to minimized the binary cross-entropy loss:  $\text{BCE}(y, p) = y \log(p) + (1 - y) \log(1 - p)$  where  $y$  is the label and  $p$  the predicted probability of the sample

being mitosis. The probability value of the dropout used in all the models was 0.4 and the weight decay was set to  $10^{-4}$ . The learning rate was set to  $10^{-3}$  for PB-CNN and  $5 \cdot 10^{-4}$  for the scale 40x of PB-CNN-STP; in both cases was divided by 10 each 30 epochs. Each batch was constructed by randomly sampling 32 positive samples and 32 negative samples. Each epoch samples  $10^6$  batches.

Data augmentation has been applied from random rotations and mirroring. We also apply random shifting up to 4, 8 and 16 pixels for scales 10x, 20x and 40x respectively, as well as random scaling by a factor sampled in the range [0.75, 1.25]. Additive Gaussian noise with 0.05 of standard deviation was also added to the input. Finally, in order to introduce robustness to color variation, we use the stain augmentation process proposed in [8] with  $\alpha$  and  $\beta$  parameters sampled in the ranges [0.995, 1.05] and [-0.05, 0.05] respectively.

We implement the Bayesian approach according to [4]. For it, we sample the dropout units from a Bernoulli distribution with probability  $p = 0.4$ . Once trained, the prediction and uncertainty of the network per each input image are computed as the average of the values of 10 new samples of the dropout units after weight adaptation by the forward pass.

Finally, we have found necessary to apply a high Dropout rate to the feature maps of previous levels at the beginning of the training process. This was done in order to force not to rely too much in previous predictions and extract useful information from the current scale. We set this dropout rate to 0.8 and linearly decrease it to 0 at epoch 40.

## 5 EXPERIMENTS

To demonstrate the benefits of our proposed PB-CNN, we first test it on our dataset conformed by 22 WSIs. We separate the WSIs in training and test sets by randomly selecting 5 WSIs as the test set and leaving the remaining 17 ones for training. We have 7133 mitoses for training and 1103 for testing. At the second

<b>Method</b>	<b>PBCNN-STP</b>	<b>DeepDet[8]</b>	<b>RR[11]</b>	<b>CasNN[1]</b>	
<b>F1 score</b>	82.6%	83.2%	82.3%	78.8%	
<b>Method</b>	<b>PBCNN10x</b>	<b>PBCNN20x</b>	<b>PBCNN</b>	<b>PBCNN-STP</b>	<b>WRCNN40x</b>
<b>F1-score</b>	62.8%	72.5%	78.1%	81.3%	71.2%
<b>Ave.Time WSI</b>	$27 \pm 11$	$28 \pm 10$	$29 \pm 11$	$31 \pm 11$	$56 \pm 23$

Table 1: Two first rows show a comparison with state of the art methods on ICPR-2012 MITOSIS test set [13]. Last three rows show a comparison on our test dataset of 5 WSI. Evolution of the F1-score and processing time are shown by scales. Results for the times were calculated applying sliding window on each pixel and using a Nvidia Titan X.

row of Table 1 we show the F1-score and time increase of adding each level of the pyramid, as well as using the Spatial Transforming Layer[6]. All models were tested using the same framework and the same computer with a Nvidia Titan X with 12GB of RAM. As the table shows, each level comes with a significant

increase in performance at the cost of a small increase in computational time. Adding the Spatial Transforming Layer [6] we get an increase of 3.2% in F1-score at the cost of a slightly impact on the processing time. For the sake of comparison, we train a Bayesian Wide Residual Network identical to the one used on the 40x scale only, we call it WR-CNN-40x. The training process was the same as described for our PB-CNN in Section. 4.3, although we change the dropout probability to 0.3 since we find it gives better results. The results of this WR-CNN-40x are show in the two first rows of the Table 1. The propose PB-CNN is almost two times faster and gets a significant better F1-score than this WR-CNN-40x, showing that the increase obtained is due to the pyramid architecture.

In order to compare our models with other in the literature, we train our best performing model PB-CNN-STP with our 22 WSI and test it on MITOS-ICPR2012 test set containing images produced by the Aperio XT scanner. Then, we extract the features provide by each scale before the last classification layer and train a Random Forest classifier on the training dataset of the Aperio XT scanner. Table 1 shows the results in comparison with other state of the art methods. We can see that the best of our proposed method get a competitive result against current state of the art in F1-score, despite being trained on WSI of a different tissue.

## 6 Discussion and conclusions

A new coarse to fine cascade of CNN Bayesian models for mitosis detection has been proposed. The new mechanism of information propagation from top to bottom, using the uncertainty of the prediction, allow to get results competitive with the state of the art on MITOS ICPR-2012 dataset. To the best of our knowledge, this is the first time that a coarse to fine approach combined with uncertainty is used in mitosis detection. In our experiments, the Bayesian pyramid approach reduces the computation time by a factor of two and increases by 7% the F1-score with respect to the same CNN architecture applied only over the 40x scale. We have also shown the benefits of using Spatial Transforming Layers to deal with local geometric deformations. On our dataset this invariance increases the F1-score score by a 3.2%. It is also remarkable that our architecture is trained with samples from a different tissue than breast cancer. This shows that our model is able of learning useful mitosis features for the transfer of learning between tissues. Regarding efficiency, the times measured on whole WSI make our method a good candidate for daily clinic. More experiments on harder databases have to be carried out in order to assess the good properties pointed out for the model. The addition of new input information from immunohistochemistry stains is also other relevant issue for future work.

## References

1. 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 (AAAI-16). pp. 1160–1166 (2016)
2. Chen, H., Wang, X., Heng, P.A.: Automated mitosis detection with deep regression networks. In: IEEE Int Symp Biomedical Imaging. pp. 1204—1207 (2016)
  3. Cireşan, D.C., Giusti, A., Gambardella, L.M., Schmidhuber, J.: Mitosis detection in breast cancer histology images with deep neural networks. In: International Conference on Medical Image Computing and Computer-assisted Intervention. pp. 411–418. Springer (2013)
  4. Gal, Y., Ghahramani, Z.: Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In: Proceedings of the 33rd International Conference on Machine Learning (ICML-16) (2016)
  5. ICPR: <https://mitos-atypia-14.grand-challenge.org/> (2014)
  6. Jaderberg, M., Simonyan, K., Zisserman, A., kavukcuoglu, k.: Spatial transformer networks. In: Advances in Neural Information Processing Systems 28. pp. 2017–2025 (2015)
  7. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. CoRR abs/1412.6980 (2014), <http://arxiv.org/abs/1412.6980>
  8. Li, C., Wanga, X., Liua, W., Latecki, L.J.: Deepmitosis: Mitosis detection via deep detection, verification and segmentation networks. Medical Image Analysis pp. 121–133 (2018)
  9. Macenko, M., Niethammer, M., Marron, J.S., Borland, D., Woosley, J.T., Guan, X., Schmitt, C., Thomas, N.E.: A method for normalizing histology slides for quantitative analysis. In: IEEE International Symposium on Biomedical Imaging: From Nano to Macro. pp. 1107—1110 (2009)
  10. MICCAI: <http://tupac.tue-image.nl/> (2016)
  11. Paul, A., Dey, A., and col.: Regenerative random forest with automatic feature selection to detect mitosis in histopathological breast cancer images. In: MICCAI 2015. pp. 94–102. Springer (2015)
  12. Reinhard, E., Ashikhmin, M., Gooch, B., Shirley, P.: Color transfer between images. IEEE Computer Graphics and Applications 21(5), 34–41 (July 2001)
  13. Roux, L., Racoceanu, D., Lomenie, N., Kulikova, M., Irshad, H., Klossa, J., Capron, F., Genestie, C., Naour, G.L., Gurcan, M.N.: Mitosis detection in breast cancer histological images an icpr 2012 contest. J Pathol Inform (2013)
  14. Tejera-Vaquerizo, A., Pérez-Cabello, G., and col.: Is mitotic rate still useful in the management of patients with thin melanoma? Journal of the European Academy of Dermatology and Venereology 31(12), 2025–2029 (2017)
  15. Tellez, D., Balkenhol, M., and col.: Whole-slide mitosis detection in "h&e" breast histology using phh3 as a reference to train distilled stain-invariant convolutional networks. IEEE Transactions on Medical Imaging 37(9), 2126–2136 (Sept 2018)
  16. Veta, M., van Diest, P.J., and col.: Assessment of algorithms for mitosis detection in breast cancer histopathology images. Medical Image Analysis 20(1), 237 – 248 (2015)
  17. Wang, H., Cruz-Roa, A., Basavanhally, A., Gilmore, H., Shih, N., Feldman, M., Tomaszewski, J., González, F.A., Madabhushi, A.: Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. Jour. Medical Imaging 1 (2014)
  18. Zagoruyko, S., Komodakis, N.: Wide residual networks. In: Proceedings of the British Machine Vision Conference (BMVC). pp. 87.1–87.12 (September 2016)
  19. Zerhouni, E., Lányi, D., Viana, M., Gabrani, M.: Wide residual networks for mitosis detection. In: IEEE 14th International Symposium on Biomedical Imaging (ISBI). pp. 924–928 (2017)