

Improvement of mitosis detection through the combination of PHH3 and HE features

Santiago López-Tapia¹, Cristobal Olivencia¹, José Aneiros-Fernández², and Nicolás Pérez de la Blanca^{1*}

¹ Computer Science and Artificial Intelligence Department, University of Granada, Granada, SPAIN,

sltapia@decsai.ugr.es, cristoly94@gmail.com, nicolas@ugr.es,

² Intercenter Unit of Pathological Anatomy, San Cecilio University Hospital, Granada, Spain

janeirosf@hotmail.com

Abstract. Mitosis detection in hematoxylin and eosin (H&E) images is prone to error due to the unspecificity of the stain for this purpose. Alternatively, the immunohistochemistry phospho-histone H3 (PHH3) stain has improved the task with a significant reduction of the false negatives. These facts point out on the interest in combining features from both stains to improve mitosis detection. Here we propose an algorithm that, taking as input a pair of whole-slides images (WSI) scanned from the same slide and stained with H&E and PHH3 respectively, find the matching between the stains of the same object. This allows to use both stains in the detection stage. Linear filtering in combination with local search based on a kd-tree structure is used to find potential matches between objects. A Siamese convolutional neural network (SCNN) is trained to detect the correct matches and a CNN model is trained for mitosis detection from matches. At the best of our knowledge, this is the first time that mitosis detection in WSI is assessed combining two stains. The experiments show a strong improvement of the detection F1-score when H&E and PHH3 are used jointly compared to the single stain F1-scores.

Keywords: Mitosis detection, WSI, PHH3 and HE, Siamese CNN

1 Introduction

The quantification of mitosis in histopathological tissues and specifically its ratio per square millimeter is one of the most relevant factors in the prognosis of cancer. Unfortunately, the process of mitosis detection on images stained with standard hematoxylin and eosin (H&E) is difficult and prone to errors due to multiple factors consequence of its unspecificity [16]. H&E staining only helps indirectly to mitosis identification, being the hyperchromaticity induced on the

* 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

mitotic cell nucleus one of its most salient features. Unfortunately, many other tissue parts are stained with a similar color too.

Phospho-histone H3 (PHH3) is a well-known immunomarker, specific for cells undergoing mitoses [14]. This fact causes PHH3 to improve the inter-observer variability of the mitosis count by a decrease in false negatives, but at the same time is prone to false positives as for instance in inter-phase tumor cells with phosphorylated core protein H3. The staining with PHH3 has meant an important improvement in mitosis detection for many type of cancers [12, 4].

The technology for the whole scanning of tissue slides (WSI) is able of digitizing a slide at resolutions of 0.25 – 0.16 microns per pixel, which means image sizes of 10^{10} pixels. In this setting, the task of mitosis detection can only be addressed using accurate and efficient algorithms. The convolutional neuronal network (CNN) models have demonstrated, in recent years, a clear superiority over traditional approaches in this task. [10, 6]. Here we focus on these kind of models.

An issue that remains to be explored in some detail is the relevance of the combination of stains in the mitosis detection process. Recently, in [15] an interesting approach taking advantage of the properties of both stains, H&E and PHH3, to build a mitosis detector on H&E has been proposed. This approach uses the PHH3 information to locate ground-truth mitosis on WSI but the goal is a classifier on H&E. Although the approach means an important step in the detection of mitosis in WSI, several issues still remain open. First, to design a simple training model taking advantage of both stains simultaneously. Second, the labeling process should take into account both stains. Fig.1 shows some cases of mitosis where the labeling from a single stain is misleading. Finally, assessing the contribution of trained detectors with both stains is a relevant issue to improve routine in daily practice.

In contrast to the above discussed approach, here we propose the simultaneous use of both stains in the labeling and detection stages. To do that we stain twice each slide taking advantage of the property of the antigenic recovering of the immunohistochemistry for destaining the H&E. This strategy reports important benefits: (i) better labeling, (ii) training dataset with both stains, (iii) improvement in detection score. The two most important challenges in our approach are a fast search for potential correct matches and an assessment model for matches.

Our main contributions in this paper are: (i) a fast and efficient technique to generate matching between both stains of the same object, (ii) the proposal of a SCNN model to validate the matches; (iii) we show that training from both stains means a clear improvement in detection score compared to use of only one. Finally, we emphasize that our searching algorithm makes very easy the labeling of pairs.

The rest of the paper is as follows. Section.2 defines the problem. Section.3 discuss the proposed approach. Section.4 shows the experimental results, and in Section.5 the discussion and conclusions are presented.

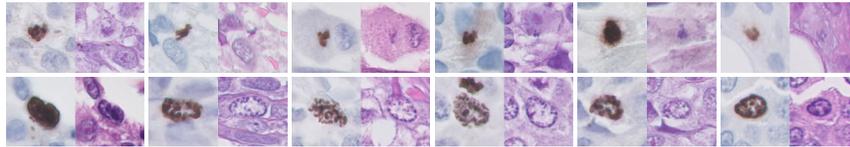


Fig. 1: This figure shows by rows examples of two difficult scenarios regarding mitosis detection in H&E or PHH3. The first row shows examples where the mitoses are very difficult to detect in H&E but can be easily detected on PHH3. The second row shows examples where the PHH3 stain indicates positive mitosis but the H&E stain shows that it is not.

2 Problem definition

To begin with we focus on the automatic object matching between stains of the same histological tissue. The relevance of this task is due to the lack of consensus between pathologists when they are asked to label a set of cells as mitosis or no-mitosis in H&E images. In the MYTHOS-ATYPIA challenge[5], for instance, multi-labels had to be considered. Daily practice has shown that many ambiguities can be solve when both stains are observed together. Fig.1 shows some examples. This has motivated the interest to know how much a detector can improve when training with both stains. The automatic identification of correct matches between stains it is not a straightforward task. The manipulation of the slide in the double staining process, that is, staining with H&E and scanning, destaining, and restaining again with PHH3 and new scanning, introduce small local deformations on the tissue that makes impossible automatic matching of the images using geometrical registering. In addition, the different response of the tissue to each one of the stains also introduce strong differences in the shape and color of the surfaces of the cells as shows Fig.1. To overcome all these deformations, we propose a search strategy to extract possible matches and a similarity distance to find correct matches. For this latter task, we propose a Siamese CNN (SCNN) [3, 8] since the CNN models have shown to be very efficient in extracting similar features from images, that being visually different, are similar in a some semantic context. At one last step, the correct matches are assessed, for mitosis presence, by a CNN classifier.

3 Methodology

3.1 Matches extraction

Let's denote by $p\text{-WSI}=(I_{\text{PHH3}}, I_{\text{HE}})$ the two WSI images of the same slide with different stain. We extract the objects present in each image applying standard cell detection functions, [1], and eliminating all those objects with a size greater than a preset threshold. For this, we use the hematoxiline and DAB bands of the H&E and PHH3 images respectively. The center of mass of the remaining connected components (CC) is computed. A kd-tree data structure (KdT) [2] is fed with the coordinates of the centers of the H&E image. The centers of the PHH3

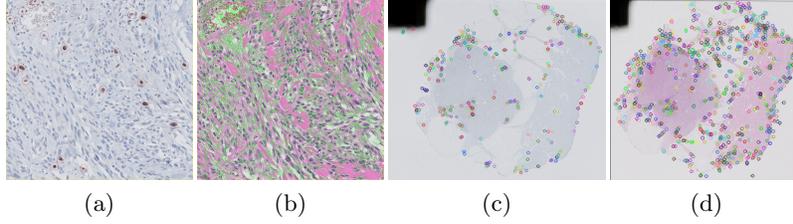


Fig. 2: Images (a) and (b) show, inside circles, objects detected on PHH3 and H&E respectively. Images (c) and (d) show, in circles of color, SURF points detected in PHH3 and H&E respectively. (Best see it at higher magnification)

image are saved as a list of points, L_{DAB} . In order to reduce the number of pair to analyze we take advantage of the specificity of the PHH3 stain to identify the potential mitosis presents in the image. To this end, each vector of coordinates in the DAB list is used as query to the KdT to retrieve matching candidates from the H&E image. Fig.2(a,b)) shows an example of how unbalanced is the number of detections in both stains. In order to make easier the searching process we register the bounding boxes of the tissue area in both images through an affine transformation, $\mathcal{A} : I_{PHH3} \rightarrow I_{HE}$, estimate from SURF points [2] detected from grey levels after sub-sampling the image by a factor of ten. For each point $x \in L_{DAB}$, its coordinates are projected onto the axes of H&E by the affine transformation, $y = \mathcal{A}x$, and all points $z \in \text{KdT}$ such that $distance(y, z) < thr$ are extracted, where thr is a prefixed threshold. Let's denote by p-center the pair formed by the coordinates of the query-point, x , and the coordinates of anyone of its matches. For each p-center, image-patches of size 80×80 pixels centered on them are extracted from the images. Let's denote them as p-patch. These p-patch are assessed by the SCNN that output a similarity distance in terms of a probability. For each x the p-patch with maximum probability is considered the true match. Let's denote a correct p-patch as p-match. In summary, our matching algorithms is as follows:

ALGORITHM: $MS(H\&E, PHH3, T, \mathcal{P}_{HE}, \mathcal{P}_{PHH3})$

Input:

- $(H\&E, PHH3)$: WSI of the same slide
- T : distance-threshold for searching
- \mathcal{P}_{HE} : list of coordinates of the object centers detected in H&E
- \mathcal{P}_{PHH3} : list of coordinates of the object centers detected in in PHH3

Preprocessing:

- Build a KdT from \mathcal{P}_{HE} .
- Compute SURF points: $SURF_{HE}, SURF_{PHH3}$
- Compute Global affine transformation: $\mathcal{A} : SURF_{PHH3} \rightarrow SURF_{HE}$.

Correspondences:

For each item $p \in \mathcal{P}_{PHH3}$

- 1.- Compute $\hat{p} = \mathcal{A}p$
- 2.- Compute $\mathcal{P}_{KdT}(p) = \{q | q \in \text{KdT}, distance(\hat{p}, q) < T\}$

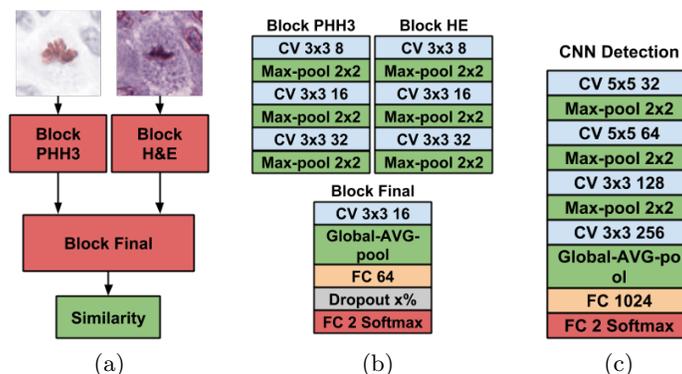


Fig. 3: Siamese architecture: (a) shows the global network design composed of two parallel branches to process each one of the images. After the feature extraction a function of the feature vectors compute the similarity between the images. In (b) we show the three main blocks that compose the model. CV correspond to Convolution and ReLU activation and FC to full connected layer followed by ReLU. We use batch normalization before each ReLU. In (c) the architecture of the CNN model used for mitosis detection is shown.

3.- Extract patches $\{o_q\}$ centered in $q \in \mathcal{P}_{\text{KdT}}(p)$

4.- Compute $\hat{q} = \underset{q \in \mathcal{P}_{\text{KdT}}(p)}{\text{argmax}} \text{Similarity}_{SCCN}(o_p, o_q)$

5.- Output $(o_{\hat{q}}, o_p)$

where Similarity_{SCCN} denote the probability computed by the Siamese network.

3.2 Dataset and Labeling

Two datasets of p-match have been created. The first dataset, DS1, is defined by 57k (1k=1000) p-match extracted after staining and scanning 48 slides, 30 of skin cancer (melanoma) and 18 of breast cancer. The second dataset, DS2, is defined by 11k p-match of mitosis and 75k p-patch no mitosis extracted from 17 slides of melanoma. The slides were scanned with a Philips Ultra-Fast Scanner at a spatial resolution of 0.25 microns per pixel. All p-patch were labeled by a senior pathologist of the Saint Cecilio University Hospital in Granada, who annotated a percentage of the correct matches on each p-WSI. An interactive software which iterates showing p-patches and their surrounding areas was used for this task. A p-patch is tagged with a maximum of two clicks: one click to decide correspondence vs. no correspondence and another click to decide mitosis vs. non-mitosis. This is a very simple routine that allows to label many pairs in a short period of time.

3.3 Training

Our specific SCNN model is shown in Fig.3(a-b). It can be observed that Block-PHH3 y Block-H&E share the same architecture based on a standard Lenet

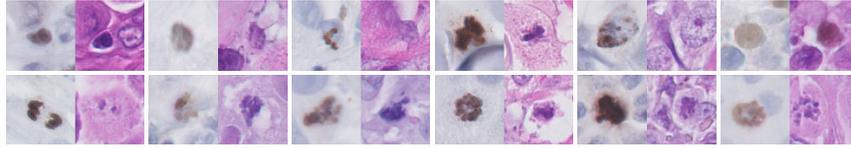


Fig. 4: This Fig. show some examples of the errors of the algorithm-MS. The first row shows examples of false negative p-match. The second row shows examples of false positives p-match. See the pair as (PHH3,HE)

model of CNN [9]. Block-Final processes the features from the input blocks to learn the similarities. The network is trained during 100 epochs using a batch size of 128 with Adam[7] optimizer and initial learning rate of 0.0002. We reduce the learning rate by a factor of 10 each 10 epochs if the training loss has not been reduced. The training stops if the loss keeps without reducing after another 20 epochs. The networks outputs the probability of a p-patch, {he, phh3}, of being a p-match. We train the network to minimize the binary cross entropy loss $\mathcal{L}(\cdot, \cdot)$, defined for each sample as,

$$\mathcal{L}(\text{he, phh3}) = -y(\log(f_{\theta}(\text{he, phh3}) + (1 - y) \log(1 - f_{\theta}(\text{he, phh3})))$$

where $y \in \{+1, -1\}$ represents the image-pair’s label and f_{θ} represents the function computed by our SCNN. To regularize the model, we use L2-weight decay of strength 1.0 on the parameters of the network and Dropout[13] with probability of 0.3 before the last full connected layer. The CNN used for mitosis detection from p-match is shown in Fig. 3(c). We minimize the binary cross-entropy loss function using the Adam[7] optimizer with learning rate set to 0.001 during the first 50 epochs, then reduced to 0.0001 for 25 epochs and finally set to 0.00001 for another 25 epochs. We set the weight decay parameter to 0.0001 and use Dropout of 0.5 before each non-linearity except before the Softmax layer. Also, we use data augmentation on the p-match by rotating the input patches by 90° , 180° and 270° and performing horizontal and vertical flips. We also add Gaussian noise with $\sigma = 0.0001$ to the input.

4 Experimental results

We assess the performance of our algorithm-MS by cross-validation. To do this, we define five folds from the set of 48 p-WSI. On each fold 43 p-WSI are used for training and 5 for testing. In total we use 25 different p-WSI in testing. On each fold the set of p-match, extracted from each image, is used according to the role of the image in that fold. Table.1 shows the number of p-match used in training and testing for each fold. The items for the negative class are generated by random combinations of the p-match items. We generate as many negative item as there are p-match. The test with each fold begins by detecting and extracting the coordinates of the centers of the objects in the p-WSI test. We use cell detection routines of the QuPath[1] free software to extract the center of the object on each p-WSI. The kd-tree structure is build using [2]. From them the set of p-patch is estimated. Eventually, the p-patch are assessed by

| | Fold-1 | Fold-2 | Fold-3 | Fold-4 | Fold-5 |
|------------------------|--------|--------|--------|--------|--------|
| Train matches | 56.6k | 52.5k | 47k | 54k | 45.6k |
| Valid. matches | 721 | 4.8k | 10.4k | 3.4k | 11.7k |
| Valid.Accuracy (80×80) | 98.6% | 99.9% | 100% | 99.9% | 99.6% |

| Patches | H&E | PHH3 | H&E+PHH3 |
|--------------------|-------------|-------------|-------------|
| Detection F1-score | 73.3 ± 0.5% | 77.6 ± 0.2% | 80.7 ± 0.4% |

Table 1: Top: results of the correspondence experiment. 1k=1000. The second row shows the number of corresponding pairs used, in each fold, in training and validation respectively. The third row shows the validation accuracy in each fold for patches of 80×80 pixels. Bottom: detection F1-score using the different stains.

the SCNN. In this experiment what we measure is the accuracy of the p-match test elements (see Table.1(top)). In order to evaluate the effect of the number of p-match in the testing matching error, we design the folds to cover a broad range of values in testing. A value of $thr=60$ is used as searching distance in the KdT. The average query time per image is about 3s. Third row in Table.1 shows the accuracy achieved on each fold. The estimated accuracy of the algorithm-MS for matches is $99.6\% \pm 0.58$. Fig..4 shows some examples of p-match errors from SCNN. We assess the H&E+PHH3 improvement versus the single stains, on dataset labeled from both stains, using the detector architecture shown in Fig. 3(c). We select this architecture for two reasons. First, it represents an adaptation of Lenet model which is the most popular CNN used for mitosis detection. Second, our dataset is filtered by the matching algorithm that removes much of the false positives. This makes unnecessary a complex architecture for this task. In a first experiment we train and test our detector using each one of the components, H&E and PHH3, of the p-WSI. In the second experiment we use full p-WSI. In all cases the color of the images was normalized using the algorithm given in [11]. From the dataset, DS2, we constructed 5 partitions of WSIs and used them for cross-validation. Table.1 in the bottom shows the detection F1-score achieved by our detector using patches from H&E, PHH3 and H&E+PHH3 respectively. The result shows that using together both stains greatly improve the F1 score with respect to only using one. To evaluate the impact of the p-match errors in detection we test our CNN with the same image dataset but computing the p-match using the algorithm-MS. In this case an F1 score of $80.1 \pm 0.4\%$ is achieved, which means a drop of only 0.6 points.

5 Discussion and conclusions

The proposed approach shows that both stains H&E and PHH3 when used together make a significant contribution to the detection of mitosis. In addition, our approach contributes with a new technique for the labeling of mitosis using both stains simultaneously. The size of the datasets makes our results preliminary but also reliable. It remains to be done a full evaluation of the matching errors and the influence of the detector. The help of our algorithm-MS in the complete

labeling of p-WSI opens the door to create larger and more challenging training data sets to evaluate new algorithms. This will be one goal for future work.

References

1. Bankhead, P., Loughrey, M.B., and col.: QuPath: Open source software for digital pathology image analysis. *Scientific Reports* 7(1), 16878, <https://doi.org/10.1038/s41598-017-17204-5>
2. Bradski, G.: The OpenCV Library. *Dr. Dobb's Journal of Software Tools* (2000)
3. Chopra, S., Hadsell, R., LeCun, Y.: Learning a similarity metric discriminatively with application to face verification. In: *Computer Vision and Pattern Recognition*. vol. 1, pp. 539–546 (2005)
4. Dessauvagie, B.F., Thomas, C., Robinson, C., Frost, F.A., Harvey, J., Sterrett, G.F.: Validation of mitosis counting by automated phosphohistone h3 (phh3) digital image analysis in a breast carcinoma tissue microarray. *Pathology* 47(4), 329–334 (Jun 2015)
5. ICPR: <https://mitos-atypia-14.grand-challenge.org/> (2014)
6. Janowczyk, A., Madabhushi, A.: Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *J Pathol Inform* 7(1), 29–29 (Jan 2016)
7. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *CoRR* abs/1412.6980 (2014), <http://arxiv.org/abs/1412.6980>
8. Koch, G., Zemel, R., Salakhutdinov, R.: Siamese neural networks for one-shot image recognition. In: *ICML workshop on Deep Learning* (2015)
9. LeCun, Y., Bengio, Y.: Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks* 3361(10) (1995)
10. Litjens, G.J.S., Kooi, T., and col.: A survey on deep learning in medical image analysis. *CoRR* abs/1702.05747 (2017), <http://arxiv.org/abs/1702.05747>
11. 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)
12. Nielsen, P.S., Riber-Hansen, R., Jensen, T.O., Schmidt, H., Steiniche, T.: Proliferation indices of phosphohistone h3 and ki67: strong prognostic markers in a consecutive cohort with stage i/ii melanoma. *Modern Pathology* 26, 404 (Nov 2012)
13. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research* 15, 1929–1958 (2014)
14. Tapia, C., Kutzner, H., Mentzel, T., Savic, S., Baumhoer, D., Glatz, K.: Two mitosis-specific antibodies, mpm-2 and phospho-histone h3 (ser28), allow rapid and precise determination of mitotic activity. *Am J Surg Pathol* 30(1), 83–9 (2006)
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)