Tumor Proliferation Assessment of Whole Slide Images

Mikael Rousson[a], Martin Hedlund[a], Mats Andersson[a], Ludwig Jacobsson[a], Gunnar Lathen[a],
Bjorn Norell[a], Oscar Jimenez-del-Toro[b], Henning Mueller[b], and Manfredo Atzori[b]

[a]ContextVision AB, Linkoping, Sweden; [b]University of Applied Sciences Western Switzerland (HES–SO), Sierre;

ABSTRACT

Grading whole slide images (WSIs) from patient tissue samples is an important task in digital pathology, particularly for diagnosis and treatment planning. However, this visual inspection task, performed by pathologists, is inherently subjective and has limited reproducibility. Moreover, grading of WSIs is time consuming and expensive. Designing a robust and automatic solution for quantitative decision support can improve the objectivity and reproducibility of this task. This paper presents a fully automatic pipeline for tumor proliferation assessment based on mitosis counting. The approach consists of three steps: i) region of interest selection based on tumor color characteristics, ii) mitosis counting using a deep network based detector, and iii) grade prediction from ROI mitosis counts. The full strategy was submitted and evaluated during the Tumor Proliferation Assessment Challenge (TUPAC) 2016. TUPAC is the first digital pathology challenge grading whole slide images, thus mimicking more closely a real case scenario. The pipeline is extremely fast and obtained the 2nd place for the tumor proliferation assessment task and the 3rd place in the mitosis counting task, among 17 participants. The performance of this fully automatic method is similar to the performance of pathologists and this shows the high quality of automatic solutions for decision support.

Keywords: tumor proliferation grading, deep learning, mitosis detection, digital pathology, whole slide imaging

1. INTRODUCTION

The visual inspection of histopathology tissue samples is the standard practice to provide the final diagnosis of several diseases and most cancers[7,14]). Histopathology digitalization opens the possibility to develop computer based image analysis methods that can support and accelerate the work of pathologists with objective and quantitative image features.[7] Several detection, segmentation and classification algorithms have been proposed for histopathology images using traditional machine learning approaches, deep learning approaches or the fusion of the two.[9,12]

Histological tumor grading systems measure how closely the different parts of a tumor correspond to specific parameters when examined under a microscope. Tumor grading systems are used to provide an estimation of the patient's prognosis and to define treatment plans. The Bloom and Richardson grading system[2] is the most widely used histological grading system for breast cancer and it is based on three main parameters: mitotic activity, nuclear pleomorphism and degree of tubule formation.

Tumor proliferation speed as a biomarker for breast cancer is highly correlated with patient prognosis and an important feature for treatment planning.[17] The proliferation speed is commonly assessed visually by pathologists by counting the dividing cell nuclei (mitotic figures) in hematoxylin & eosin stained histological slides, a high count in a specific area indicating a high tumor proliferation speed. Although this task is routinely performed in pathology departments, it suffers from poor inter– and intra– observer agreement, particularly between non–specialized pathologists.[18] This is due to the highly subjective nature and difficulty of the task that requires concentration and changes in scales of the very large WSIs (Whole Slice Images).

The drawbacks of the current manual practice have created an interest in automatic tumor proliferation assessment for decision support. Just as the visual inspection of whole slide images by pathologists, automatically assessing tumor proliferation faces several challenges. Histological slides require an analysis at very high resolution, since mitotic figures occur on the cellular level. Assessing the slide requires attention to image structures

---

Further author information: Send correspondence to H.M. Email: henning.mueller@hevs.ch

of approximately 10^4 pixels (a 100×100 pixel neighborhood) within large WSIs containing 10^{10} pixels. The size of the images and the nature of the assessment lead to a task with weak supervision as a result of the large dimensionality reduction.[16] Automatic methods have aimed at reducing the subjectivity of current practices and limiting the time for slide assessment. Recent advances using machine learning methods, such as convolutional neural networks for image analysis, have obtained promising results addressing these tasks.[5, 8]

The success and development of machine learning algorithms are largely driven by the availability and quality of annotated data.[9] In recent years, several challenges such as the Mitosis detection in breast cancer histological images (MITOS) 2012[13] and the Assessment of mitosis detection algorithms (AMIDA) 2013 challenge[18] were organized to push scientific research forward in the field and include the provision of annotated data. These first challenges focused on predictions at local fields within the slides including the detection of mitotic figures and segmentation of tumors. Although the state–of–the–art for these tasks is approaching and sometimes even surpassing[1] human–level performance, these results were achieved in relatively controlled conditions and were not focused on the task of slide or patient assessment. Images were cropped from whole slide images (WSI) to provide only the region of interest with slides generated from a limited number of sources resulting in low variability in staining. The Tumor Proliferation Assessment Challenge (TUPAC) 2016* is the first case in which whole slide assessment is attempted, requesting a global tumor proliferation grade for each slide.

We propose a whole slide grading strategy that includes a mitosis detection step and a novel approach to apply this detector to whole slide images in a robust way. This new network architecture requires fewer parameters and a much lower computation time without compromising accuracy. Our model proved to be robust to the variation of slides that exists in pathological practice. The approach achieved high scores on the final evaluation task of TUPAC 2016 compared to the other approaches submitted.

2. METHODS

The proposed tumor proliferation assessment strategy mimics how a pathologist grades the slides. First, the analysis is targeted to the invasive areas of the WSI where tumor cells are most frequent. These areas are referred to as *Regions of Interest (ROIs)*. A selection of ROIs from the WSI is then evaluated by a mitosis detector. Finally, the detections within the ROIs are combined to estimate the proliferation grade of the WSI.

2.1. Data

The algorithm was trained and tested on the publicly available TUPAC 2016 challenge data. The TUPAC data* include 500 diverse breast cancer histological WSIs (usually around $0.25\,\mu$m per pixel and $1 - 10$ Gigapixels) for training, annotated according to tumor proliferation grade (1, 2 or 3). In the 148 training cases, Regions of Interest were manually annotated by pathologists and they routinely performed the mitosis counting in these areas. The test set consists of 321 WSIs from different breast cancer cases.

An auxiliary data set (AMIDA13†), also provided during the TUPAC 2016 challenge, was used to train the mitosis detector for the full pipeline. The AMIDA data[18] consists of 656 images with annotated mitoses, each image representing one high power field (2000×2000 pixels at $0.25\,\mu$m/pixel, covering an area of 0.5×0.5 mm^2).

2.2. Region of Interest Selection

For a subset of the data in the TUPAC 2016 challenge, annotated ROIs made by expert pathologists were provided. From the TUPAC data annotations it could be deduced that the invasive parts of the WSIs are in general characterized by having a slightly darker and more blueish appearance in relation to other tissue Fig. 2. The first step towards selecting ROIs is to evaluate the probability that diverse parts of the image include the most invasive cancer regions. A map related to the probability of invasion was estimated according to the annotations performed by the TUPAC pathologists. The map, h, was computed from the RGB components of
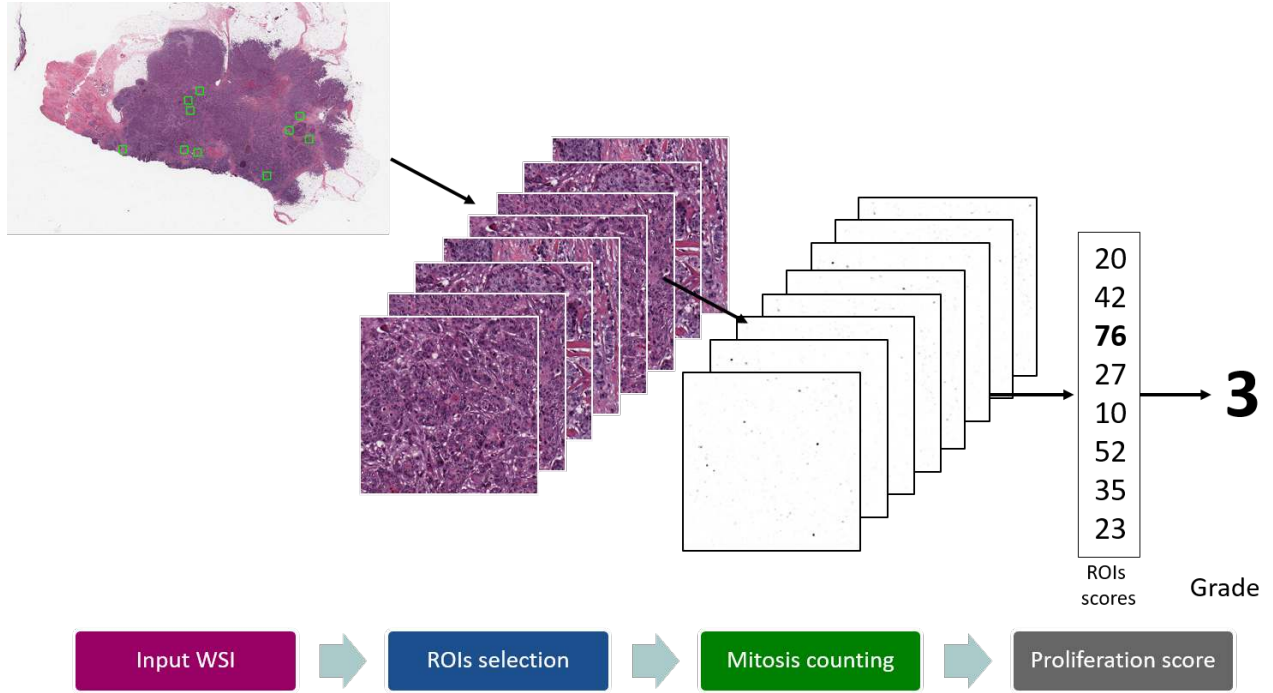

Figure 1. Our pipeline from WSI to grade is composed of three steps: ROI selection, mitosis detection and proliferation score estimation.

the WSI using a combination of two mappings, (h_1, h_2), targeted to highlight the invasive tumor area and to restrain areas containing fat, connective tissue and background.

$$
\begin{aligned}
h_1 &= \tanh\left(\frac{2(B-R)}{G+1} + 0.5\right) + 0.5 \quad h_2 = \tanh\left(\frac{640 - R - G - B}{200}\right) + 0.5 \\
h &= 0.5 \tanh\left(h_1 \, h_2 - 1.75\right) + 0.5
\end{aligned}
$$

Here the R, G and B color components are in the range of $[0, 255]$ and the heat map is limited to the range $[0, 1]$. The parameters in the presented maps were experimentally defined according to the TUPAC data. The proposed mapping is built from intuition with simplicity in mind and proved to be stable enough against major stain variations. Nevertheless, the heat map image is not quantitative. The magnitude of the heat maps from two WSIs originating from different labs cannot be compared directly due to variations in staining and scanner settings. The purpose of the heat map is to guide the ROI selection within each WSI. Fig. 2 represents a qualitative evaluation of this phase, showing how the ROIs defined by the TUPAC 2016 pathologists matched high intensity points.

As a final step in the heat map generation a low pass filter is applied to suppress local variations to a range that corresponds to the spatial extent of the ROI. The position of the first ROI is obtained from the location of the global maximum. The heat map image is then de–emphasized near this position by a long tail filter function that sets the magnitude close to zero in the immediate vicinity of the ROI position while also slightly reducing the magnitude in the surrounding area. The objective for this is to complete a sparse sampling distribution of all the high intensity areas in the heat map. The ROIs are obtained recursively by finding the position of the global maxima in the map after reducing the magnitude of the areas close to the previous maximum using local filters. The right part of Fig. 2 shows the map after 7 extracted ROIs. Note the even distribution of the extracted ROIs with respect to the magnitude of the heat map. For each WSI, 16 ROIs were sequentially generated with a size of $2\,\mathrm{mm}^2$ each. After optimizing the algorithm in training, only the first 3 areas were evaluated for the mitoses detection task in the TUPAC challenge.
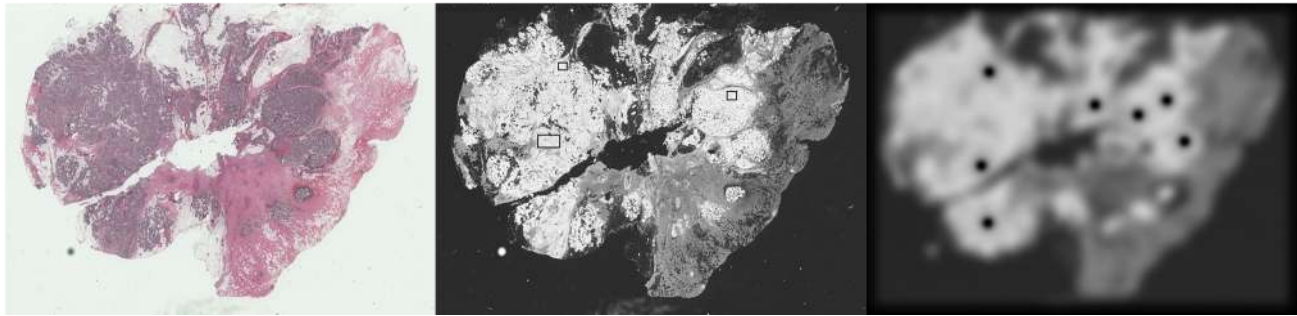
Figure 2. Left: original WSI. Middle: associated heat map, h, and the three ROIs suggested by the pathologists. Right: State of the heat map after generating 7 ROIs.

Table 1. Network architecture for input patches of shape $(3, 33, 33)$.

Layer type	Output shape	Param #
1×1 Conv2D+Linear	(1, 33, 33)	4
3×3 Conv2D+ReLU + 2×2 MaxPool2D Stride-2	(8, 15, 15)	80
3×3 Conv2D+ReLU + 2×2 MaxPool2D Stride-1	(16, 12, 12)	1168
3×3 Conv2D+ReLU + 2×2 MaxPool2D Stride-1	(16, 9, 9)	2320
3×3 Conv2D+ReLU + 2×2 MaxPool2D Stride-1	(16, 6, 6)	2320
3×3 Conv2D+ReLU + 2×2 MaxPool2D Stride-1	(16, 3, 3)	2320
2×2 Con2D+ReLU	(16, 2, 2)	1040
2×2 Con2D+ReLU	(100, 1, 1)	6500
1×1 Con2D+Softmax	(2, 1, 1)	202
		15 954

2.3. Mitosis Detection

Network Architecture   The network architecture proposed by Ciresan et al.[5] is a typical image classification network, i.e. for each input image, it returns a predicted class. Image classification networks used for detection often formulate the problem as a patch classification problem, where positive patches are centered on mitoses and negative patches have no mitoses.

The work in Shelhamer et al.[15] showed that this type of network can be seen as fully–convolutional and can be applied across images with bigger size than the training patches. As a result, a single pass through the network provides predictions for multiple patches. The spatial density of the predicted patches depends on the overall stride of the network, which is commonly equal to $2^{\text{nb pooling layers}}$ (each pooling layer having a stride of 2 pixels). The original network had 5 pooling layers and a resulting total stride of 32 pixels. The large stride becomes a problem when the goal is to efficiently compute the mitosis score for each pixel of an image. The network would need to be applied to 32×32 shifted versions of the image.

An optimized implementation was introduced in Masci et al.[11] to minimize this overhead, but here we consider a different solution that works with standard deep learning implementations. Instead of using a stride 2 for all max pooling layers, the stride is fixed to 1 for all layers except for the first one. This makes the network robust to small translations due to pooling, respecting the small size of nuclei. While this does not reduce the size of the image by a factor of 2, we still get a reduction of 1 pixel and keep the robustness to small translations due to pooling. As a consequence, the total stride of the network is 2 and a single pass gives a detection image with half the resolution of the input.

Training Strategy   To train the network, patches with mitoses (positive patches) are obtained by cropping areas centered on mitoses annotated by pathologists. Negative patches (patches without mitoses) are randomly sampled in areas that do not overlap with any mitosis. This is strongly unbalanced, which is the reason for

negative patches having an impact on the detector performance. Random sampling can provide negative samples that are easy to discriminate from positive samples and the final detector has a high false positive rate. This is a common problem in object detection and a solution is to train a first detector on random negatives and use this detector to sample new "hard" negatives.[19]

The detector used here (Algorithm 1) updates its training set every time it starts over–fitting by adding false positive patches to the negative set. After each epoch, we test if the classifier has reached a target objective. When the objective is reached, the classifier is applied to the original images and picks false detections (no overlap with annotated mitoses). The new patches are then added to the training set and the training process is resumed until the objective is reached again. The process is iterated until the number of hard negatives in the original training images is below an experimentally defined threshold.

Algorithm 1: Mitoses detector training strategy with hard online negative mining.

Data: AMIDA13 data set: ROIs with annotated mitosis coordinates
Result: Mitosis coordinates on the AMIDA test set
Sample positive and random negative patches;
for $i = 0$ **to** max_{epochs} **do**
    Train patch detector on one epoch;
    if $TP > TP_{min}$ *and* $FP < FP_{max}$ **then**
        Run detector on original training images;
        if *Number of false detection below a threshold* **then**
            Stop training;
        end
        Add false detection patches to training set;
    end
end

Data Augmentation Data augmentation allows to increase the amount of training data, which is important with relatively few annotated mitoses for training a neural network (< 2000). Random geometric transformations were applied to each sample during training. The augmentations include translation (+/-10% of the patch size), +/-90 degrees rotations and vertical/horizontal flipping.[6]

Augmentation is also important at inference time.[6] By combining rotations and flipping, we get 8 predictions for the same input that are combined by pixel–wise averaging. Data augmentation during training and inference has a dramatic impact on the generalization of the detector. This is discussed in Section 3.

2.4. Whole Slide Grading

As done by pathologists, a fixed number of ROIs are extracted and the mitoses are detected for each ROI.[3] The combination of predictions is then used to predict the overall grade. By simply counting the number of detections in each patch and using a compounding strategy similar to the clinical process the slide grade can be assessed. However, a different approach was chosen since all selected ROIs may not correspond to invasive tumor areas if the slide tumor area is limited. We compared several strategies using cross–validation (see 3.2 for details). The selected heuristic to estimate the final grade for each slide is to count the number of pixels with a detection score above a threshold for each ROI and then take the maximum of these counts across the patches. This gives a single score for the slide. Finally, we assume that the test and training sets are sampled from the same data distribution. This assumption lets us quantify the scores to: grade 1, 2 or 3.

3. EVALUATION

3.1. Mitosis Detection

AMIDA13 By training our detector on the same training/test split as the one used in the AMIDA13 challenge we were able to compare our results to the best submissions (see Table 2). For the selection of hard negatives

Table 2. Mitoses detector evaluation on AMIDA13 test set. A: Color input without channel reduction; B: Grayscale input; C: Color input converted to a single channel in the network; D: Same as C but without augmentation; E: Ensemble of 7 networks C with different seeds for weight initialization.

Method	F_1–score[‡].	sec./mm^2
[5] Ensemble of 3 networks	0.611	1920.0[§]
A: col-0.5mpp	0.552	0.9
B: gray-0.5mpp	0.494	0.6
C: c2g-0.5mpp	0.582	0.9
D: c2g-0.5mpp (no aug.)	0.420	0.2
E: Ensemble 7×C	**0.624**	6.2

during the detector training (see Algorithm 1), a target of 50% for the true positives and 5% for false positives were used. The idea is to be able to reject almost all negatives before looking for harder ones while detecting at least half of the positives. The network was trained for 500 epochs with an Adam optimizer.[10]

At inference time the detector is applied to all 8 augmentations of the data and the final prediction is an ensemble of these results. Those augmentations give a noticeable increase of 50% for the final F_1–score as shown in Table 2. Despite our changes for better efficiency, our detector is still competitive in terms of accuracy with the best published results.[4, 5]

TUPAC 2016 challenge   The proposed mitosis detection approach was also evaluated during the TUPAC 2016 challenge. The submission was labeled as team 'Contextvision, Sweden (SLDESUTO-BOX)' and the results are publicly available in the TUPAC 2016 website[¶]. The top five submission results for the 'Task 3: Mitosis Detection' are shown in Table 3. Our proposed approach obtained the 3rd best result in the challenge for this task and unlike other submitted approaches was trained using only the data set from the TUPAC 2016 challenge.

Table 3. The five best mitosis detection approaches submitted to the TUPAC 2016 challenge. The proposed mitosis detection strategy (Contextvision, Sweden (SLDESUTO–BOX)) obtained the third best F_1–score. The column '*AddData*' shows if the submitted algorithm was trained using additional data to provided the TUPAC 2016 data set.

Task 3: Mitosis detection		
Team	F_1–*score*	*AddData*
Lunit Inc., Korea	0.652	No
IBM Zurich Research and Brazil	0.648	Yes
Contextvision, Sweden (i.e. SLDESUTO–BOX)	**0.616**	**No**
The Chinese University of Hong Kong	0.601	Yes
Microsoft Research Asia	0.596	Yes

3.2. Whole Slide Grading

Cross–Validation   We used the TUPAC 2016 training set[*] for the optimization of the grading algorithm. In order to deduce the grade from the mitosis detection maps we created 10 random splits of the 500 annotated slides with an 80/20 training/test ratio and used a cross–validation strategy on these splits for validation. We noticed that WSIs in the data set have two typical color appearances ("blueish" or "pinkish"). Therefore, we decided to cluster them into two groups and optimize classification thresholds separately for each type. Figure 2 shows an example of grading with details about each step of the pipeline. The WSIs were categorized into blue

---

[‡] F_1–score is defined as $2RP/(R+P)$ with recall $R = N_{tp}/(N_{tp}+N_{fn})$ and precision $P = N_{tp}/(N_{tp}+N_{fp})$ and where N_{tp}, N_{fp} and N_{fn} are the number of true positives, false positives and false negatives, respectively.

[§] Patch–based GPU implementation reported in.[5]

[¶] TUPAC 2016 results, `http://tupac.tue-image.nl/node/62`, as of 15. Dec. 2017

and non–blue classes based on their overall blue/red color cast and the corresponding detection threshold was adapted for the two classes.

The final technique (i.e. the one submitted to the TUPAC 2016 challenge) is an ensemble of two predictions. One is based on a single detector and the other is using detectors for blue and non–blue WSIs. The two predictions were combined by a weighted average. The weights for the ensemble and thresholds were optimized by cross–validation. Figure 2 shows an example of grading with details about each step of the pipeline.

We obtained a cross–validation weighted Kappa of 0.47 ± 0.1 with a single mitosis detector and 0.51 ± 0.1 with a combination of two detectors (by averaging detection maps).

TUPAC 2016 challenge The proposed mitosis detection approach was evaluated during the TUPAC 2016 challenge on the TUPAC test set. The top five submission results for the 'Task 1: Prediction of the proliferation score based on mitosis counting' are shown in Table 3. Our proposed approach obtained the 2nd best results in the challenge for this task and it was trained only with the TUPAC 2016 data. This compares favorably to the best performing methods in the TUPAC challenge where the winning team got a Kappa of 0.567 vs. 0.534 of our method.

Table 4. Top five tumor proliferation assessment approaches submitted to the TUPAC 2016 challenge. The proposed WSI grading strategy (i.e. Contextvision, Sweden (SLDESUTO–BOX)) obtained the second best quadratic weighted Cohen's kappa between the predicted and ground truth proliferation scores. The column '*AddData*' shows if the submitted algorithm was trained using additional data compared to the data provided the TUPAC 2016 data set.

Task 1: Prediction of the proliferation score based on mitosis counting		
Team	*Weighted kappa*	*AddData*
Lunit Inc., Korea	0.567	No
Contextvision, Sweden (SLDESUTO-BOX)	**0.534**	**No**
Sectra, Sweden	0.462	Yes
University of Heidelberg Germany	0.417	No
IBM Zurich Research and Brazil	0.385	Yes

4. DISCUSSION

The proposed method is heavily influenced by the work-flow of pathologists and consists of three parts. First, the regions of interests for tumor analysis in each slide are selected using a novel approach that has proven robust to slide staining variations. The regions of interest are then analyzed by a deep convolutional network as a mitotic figure detector. The detector is trained using a previous data set of a mitotic figure detection challenge, the AMIDA13 challenge. Dividing cell nuclei are annotated. The final assessment is a combination of the results from the mitosis detection in the regions of interest.

The pipeline proves robust and obtained very good results on the TUPAC 2016 challenge among 17 participants. Our model for mitosis detection had a similar F_1–score as the best result from the AMIDA13 challenge but on the data set of TUPAC 2016 that more closely resembles the histological data found in reality. The approach is simple and fast when compared to other attempts on the TUPAC data set and nevertheless results in similar F_1 and weighted Cohen's Kappa scores.

The developed mitosis detector is based on the convolutional neural network proposed by Ciresan et al.[5] We designed the mitosis detector, keeping in mind that its objective is to be applied to large sections of a whole slide image. The tumor proliferation scores estimated by the resulting model on the TUPAC test data correspond to a moderate to good conformity for a human expert pathologist. Furthermore, our mitosis detector outperforms the winning contributions of the AMIDA13 challenge when evaluated on the same test data used in the competition. The quality and performance of these algorithms have reached a level for them to have a profound impact in the pathology work–flow and prognostic reproducibility.

Future challenges include solving the subjectivity of annotations and variations in specimen staining (due to different practices regarding staining and slide preparation). We are currently looking at different improvements to the approach. The most prominent is to improve the mitoses detector by using color normalization and additional augmentation methods.

5. CONCLUSION

This paper presents a fully automatic pipeline for WSI grading using a new network architecture with fewer parameters and lower computation time. The proposed method is similar to the work–flow pathologists perform when analyzing these samples. This grading strategy was evaluated using data from two open challenges, namely AMIDA13 and TUPAC 2016. Our simple and fast approach for mitosis detection has a similar F_1–score and weighted Cohen kappa compared to the best results submitted to the AMIDA13 challenge. Moreover, the TUPAC 2016 challenge more closely resembled the tasks pathologists deal with in real conditions as the algorithms were tested in whole slide images. Our full tumor proliferation assessment algorithm obtained top scores in this evaluation with a low computational time.

ACKNOWLEDGMENTS

REFERENCES

1. Babak Ehteshami Bejnordi, Guido Zuidhof, Maschenka Balkenhol, Meyke Hermsen, Peter Bult, Bram van Ginneken, Geert Karssemeijer, Nico Litjens, and Jeroen van der Laak. Context-aware stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images. *Journal of Medical Imaging*, 4:4 – 4 – 8, 2017.
2. H.J.G. Bloom and W.W. Richardson. Histological grading and prognosis in breast cancer: a study of 1409 cases of which 359 have been followed for 15 years. *British journal of cancer*, 11(3):359, 1957.
3. H.R. Champion and I.W.J. Wallace. Breast cancer grading. *British journal of cancer*, 25(3):441–448, 1971.
4. Hao Chen, Qi Dou, Xi Wang, Jing Qin, and Pheng-Ann Heng. Mitosis detection in breast cancer histology images via deep cascaded networks. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pages 1160–1166. AAAI Press, 2016.
5. Dan C. Cireşan, Alessandro Giusti, Luca M. Gambardella, and Jürgen Schmidhuber. Mitosis detection in breast cancer histology images with deep neural networks. In *International Conference on Medical Image Computing and Computer-assisted Intervention*, pages 411–418. Springer, 2013.
6. Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.
7. M. N. Gurcan, L. E. Boucheron, A. Can, Anant Madabhushi, N. M. Rajpoot, and B. Yener. Histopathological image analysis: A review. *IEEE Reviews in Biomedical Engineering*, 2:147–171, 2009.
8. Oscar Jimenez-del-Toro, Manfredo Atzori, Sebastian Otálora, Mats Andersson, Kristian Eurén, Martin Hedlund, Peter Rönnquist, and Henning Müller. Convolutional neural networks for an automatic classification of prostate tissue slides with high–grade gleason score. In *Proceedings of SPIE Medical Imaging*, volume 10140, pages 101400O–101400O–9. International Society for Optics and Photonics, 2017.
9. Oscar Jimenez-del-Toro, Sebastian Otálora, Mats Andersson, Kristian Eurén, Martin Hedlund, Mikael Rousson, Henning Müller, and Manfredo Atzori. Analysis of histopathology images: From traditional machine learning to deep learning. In Adrien Depeursinge, Omar S. Al-Kadi, and J. Ross Mitchell, editors, *Biomedical Texture Analysis: Fundamentals, Applications, Tools, and Challenges Ahead*. Elsevier, 2017.
10. Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
11. Jonathan Masci, Alessandro Giusti, Dan Ciresan, Gabriel Fricout, and Jurgen Schmidhuber. A fast learning algorithm for image segmentation with max-pooling convolutional networks. In *Image Processing (ICIP), 2013 20th IEEE International Conference on*, pages 2713–2717. IEEE, 2013.

12. Sebastian Otálora, Angel Cruz-Roa, John Arevalo, Manfredo Atzori, Anant Mandabhushi, Alexander Judkins, Fabio González, Henning Müller, and Adrien Depeursinge. Combining unsupervised feature learning and riesz wavelets for histopathology image representation: Application to identifying anaplastic medulloblastoma. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention — MICCAI 2015*, volume 9349 of *Lecture Notes in Computer Science*, pages 581–588. Springer International Publishing, Oct 2015.

13. Ludovic Roux, Daniel Racoceanu, Nicolas Loménie, Maria Kulikova, Humayun Irshad, Jacques Klossa, Frédérique Capron, Catherine Genestie, Gilles Naour, and Metin Gurcan. Mitosis detection in breast cancer histological images An ICPR 2012 contest. *Journal of Pathology Informatics*, 4(1):8, 2013.

14. Raphael Rubin, David S. Strayer, and Emanuel Rubin, editors. *Rubin's pathology: Clinicopathologic Foundations of Medicine*. Lippincott Williams & Wilkins, 2008.

15. Evan Shelhamer, Jonathon Long, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 2016.

16. Paul J. van Diest, Jan P. A. Baak, Erna Matze-Cok, E.C.M. Wisse-Brekelmans, C.M. van Galen, P.H.J. Kurver, S.M. Bellot, J. Fijnheer, L.H.M. Van Gorp, and W.S. Kwee. Reproducibility of mitosis counting in 2,469 breast cancer specimens: results from the multicenter morphometric mammary carcinoma project. *Human pathology*, 23(6):603–607, 1992.

17. Paul J. van Diest, Elsken van der Wall, and Jan P.A. Baak. Prognostic value of proliferation in invasive breast cancer: a review. *Journal of Clinical Pathology*, 57(7):675–681, 2004.

18. Mitko Veta, Paul J. Van Diest, Stefan M. Willems, Haibo Wang, Anant Madabhushi, Angel Cruz-Roa, Fabio Gonzalez, Anders B.L. Larsen, Jacob S. Vestergaard, Anders B Dahl, et al. Assessment of algorithms for mitosis detection in breast cancer histopathology images. *Medical Image Analysis*, 20(1):237–248, 2015.

19. Paul Viola and Michael J. Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, 2004.