



Image Magnification Regression Using DenseNet for Exploiting Histopathology Open Access Content

Sebastian Otálora^{1,2}(✉), Manfredo Atzori², Vincent Andrearczyk²,
and Henning Müller^{1,2}

¹ University of Geneva (UNIGE), Geneva, Switzerland

² University of Applied Sciences Western Switzerland (HES-SO), Sierre, Switzerland
juan.otaloramontenegro@hevs.ch

Abstract. Open access medical content databases such as PubMed Central and TCGA offer possibilities to obtain large amounts of images for training deep learning models. Nevertheless, accurate labeling of large-scale medical datasets is not available and poses challenging tasks for using such datasets. Predicting unknown magnification levels and standardize staining procedures is a necessary preprocessing step for using this data in retrieval and classification tasks. In this paper, a CNN-based regression approach to learn the magnification of histopathology images is presented, comparing two deep learning architectures tailored to regress the magnification. A comparison of the performance of the models is done in a dataset of 34,441 breast cancer patches with several magnifications. The best model, a fusion of DenseNet-based CNNs, obtained a kappa score of 0.888. The methods are also evaluated qualitatively on a set of images from biomedical journals and TCGA prostate patches.

1 Introduction

Pathologists analyze biopsies looking for structural patterns such as nuclei and gland deformations to grade various types of cancer and to describe the structures in the images for later writing the pathology report. These visual patterns are traditionally inspected using a light microscope at a certain magnification level but also increasingly through digital biopsy slides, namely Whole Slide Images (WSIs).

Deep Learning (DL) models and, in particular, Convolutional Neural Networks (CNNs) learn high-level discriminative features for digital pathology tasks [2, 4, 7] such as classification and content-based image retrieval [5]. Most supervised DL models require thousands or even hundreds of thousands of manually annotated patches when building a model from scratch, which is extremely difficult to obtain for medical data. Given the availability of open access data repositories such as the cancer genome atlas (TCGA¹), digital teaching files and

¹ <http://cancergenome.nih.gov/>.

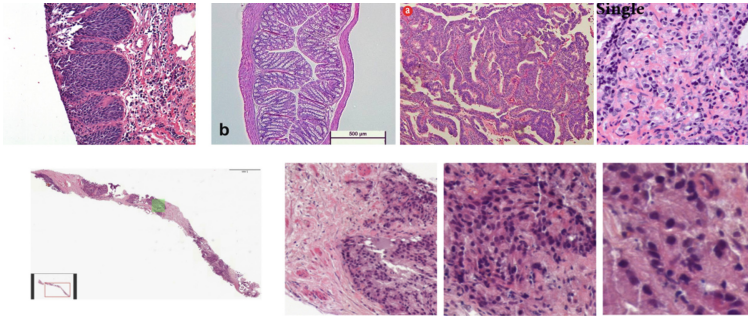


Fig. 1. Top row: PubMed Central images from the category *Light Microscope*. In their captions, the magnifications were identified as $5\times$, $5\times$, $10\times$ and $20\times$ respectively. Bottom row: Three Prostate biopsy patches extracted at a $5\times$, $10\times$ and $20\times$, respectively.

PubMed Central (PMC²), an open question is how to use these datasets for leveraging useful knowledge from them effectively, since they offer an attractive possibility to obtain large amounts of relevant medical images for training models, and ultimately solving concrete medical inquiries. In TCGA, the available WSIs lack of local annotations, but the magnification information is provided in the WSI file. In PMC, the challenge is bigger since there is a wide variety of organs and species (humans, macaques and mice), staining procedures and slide preparation methods and also unknown magnification levels of the images. Example images are shown in Fig. 1. All these factors vary strongly among digital pathology images and even more after figure editing, for example when writing a scientific publication or after publishing an article. Raw data of the WSIs from where the images come are never available.

Several authors have studied the influence of the magnification level for WSI classification, nuclei detection and segmentation with interesting findings. Bayramoglu et al. [1] trained a multitask CNN to predict both malignancy and image magnification level simultaneously, showing that the network trained with multiple magnifications outperforms the single magnification one, they also encourage to regress the magnification level instead of limiting a classifier to a discrete set of levels. Janowczyk et al. [4], trained a standard CNN for nuclei segmentation based on the AlexNet architecture, forcing the network to learn better boundaries, they discuss the need for re-training of models for each magnification level. Kumar et al. [6] designed a CNN that outputs a 3-class probability for each pixel (background, boundary and inside nuclei) and evaluate their method on several tissue types outperforming CellProfiler and Fiji in a fixed magnification level. Otálora et al. [8] trained a deep CNN to predict three fixed levels of magnification and evaluated in a single type of tissue, their results show that a pretrained network has better overall performance; however, in content-based retrieval tasks, where the query pattern could be in any type of tissue at any magnification, this classifier is of limited usability.

² <https://www.ncbi.nlm.nih.gov/pmc/>, URLs as of 3 January 2018.

The objective of this paper is to tackle the variability in scale using a regression approach based on deep CNNs tailored to regress directly the magnification level. The proposed approach is tested on different type of tissues in open access datasets showing the generalization of the method, an exploration of the combination of different regression approaches led to a good quantitative performance of magnification prediction.

2 Methods

Regressing Nuclei Average Area: The average nuclei area in terms of pixels can provide an estimate of the magnification of an image, this regressor can be used for computing differences between nuclei areas of different kind of tissues as shown in the results section, nevertheless this depends on the cell type and disease. This regression has the advantage that bypasses the problem of nuclei segmentation at test time, even though the annotated masks are still needed for computing the average area ground truth. In both architectures, the last layer is designed to output only a real number, i.e., the nuclei average area, in order to minimize the mean squared error between the ground truth and predicted average areas. Predicting the magnification with an average area is done by computing the closest magnification mean area using the mean of the nucleus areas in the training set patches and then assigning its correspondent magnification. i.e. if the predicted area is 650 pixels, the magnification assigned will be $30\times$.

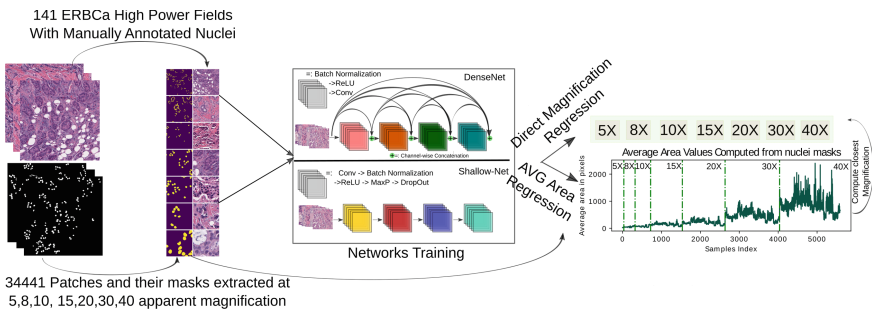


Fig. 2. Overall schema of our magnification regression approach. For the regression of the average area, the segmentation masks are used to compute the average area of nuclei. Instead of having a unit layer at the end of each architecture that regres to the magnification (top branch), a unit outputs the average area (bottom branch).

A comparison of two different CNN architectures is done in the two scenarios of direct magnification and average area nuclei regression, as shown in Fig. 2. The first architecture is the state-of-the-art DenseNet architecture [3] that features a dense connectivity pattern among its layers. DenseNet introduces direct

connections between any two subsequent layers with the same feature map size. The main advantage of DenseNet over other very deep architectures is that it reuses information at multiple levels without drastically increasing the number of parameters in the model, particularly with the inclusion of bottleneck and compression layers. For the second architecture, a relatively shallow network, named **ShallowNet**, is designed. It consists of 4 consecutive blocks of convolution, batch normalization, rectified linear units and dropout with a probability of 0.25. The comparison of the two architectures assesses the performance gain in deeper and more complex architectures versus a more parameter-efficient one. In the case of direct regression of the magnification, the last output unit of the two networks is set to predict the magnification value of the patch directly, without computing the area of the nuclei in the segmentation mask. The regressed magnification is mapped to the closest magnification by calculating the minimum absolute value between the prediction and the magnification classes. The details of the two DL architectures are:

DenseNet-BC 121: The chosen architecture is the 121-layer variation of DenseNet with 7 million parameters and perform experiments fine-tuning all the layers from pre-trained ImageNet weights and training the weights from scratch.

ShallowNet: A 4-layer CNN consisting of 3×3 convolutional kernels, followed by batch normalization, ReLU activation, dropout of 0.25 and max-pooling of a 2×2 neighborhood, ending in a dense layer with a linear activation that is expected to output the average area of the nuclei in the patch. This designed network has 2.7 million parameters.

As baseline for the area regression the DL nuclei segmentation method of Kumar et al. [6] is chosen. Since the calculated average nuclei area is needed for comparing it with the regression approach, we added the first and second output probability maps of their network that corresponds to the probabilities of pixels belonging to the inner nuclei and their boundary. An Otsu threshold is computed from this output to obtain a binary mask from which the average area of the nuclei is calculated using the resulting blobs. All the nuclei that were on the edge of the patches were removed to have a more robust prediction. Also, detected areas of less than 20 pixels are not taken into account since in the ground truth the minimum nuclei area at $5 \times$ was 24 pixels. Even though this was not a fair comparison, since the model of Kumar was trained for a single magnification, this highlights the advantage of having a flexible area regressor.

2.1 Datasets

The data used for training in our approach is the publicly available dataset used for nuclei segmentation in [4], that allows to confidently estimate via manual annotations the ground-truth nuclei average area, and also downsampling the original image and masks to obtain the different magnification levels. This dataset consists of 141 images and masks of 2000×2000 pixels @ $40 \times$ ROIs of

Table 1. Number of ERBCa patches extracted per magnification and partition

Partition/#patches	5X	8X	10X	15X	20X	30X	40X	Total
Train	47	1087	2026	3646	4501	5368	5819	22494
Validation	4	294	585	1066	1302	1527	1575	6353
Test	18	214	440	864	1099	1401	1558	5594
<i>Total</i>	69	1595	3051	5576	6902	8296	8952	34441

estrogen receptor-positive breast cancer (ERBCa). The images contain a subset (not all nuclei in the images were annotated) of 12,000 manually annotated nuclei. We extracted 34441 patches for 5, 8, 10, 15, 20, 30, and 40 \times magnifications. The number of patches per magnification was kept within the same ranges when possible, i.e., for 5 \times and 8 \times is not possible to extract as many patches as in 30 or 40 \times due to the large area covered by the lower magnifications. The patches are separated into training (94), validation (27) and test (20) partitions checking that all the patches from a single image were in the same partition. In each patch, the condition that at least 3 complete nuclei were present is ensured. The complete distribution of patches is shown in Table 1 and example patches with masks are displayed in the input of the networks in Fig. 2.

For assessing the generalization of the approach, we tested it on two external open access databases: TCGA patches and PubMed Central histopathology images. The best trained model was tested on the test partition of 99125 patches used in the evaluation of the method reported in [5]. The patches corresponds to areas of low (45081) and high (54044) grade prostate cancer, with reported Gleason scores 6-7 and 8-9-10 respectively, at 20 \times magnification. For the PMC set, a total of 5,764,238 images with captions were crawled. A standard multi-modal CNN architecture was used for the captions and images to identify the image modality, e.g. light microscopy, x-ray, MRI, etc. The classification process led to a total of 291 prostate histopathology images.

3 Results

In Table 2 the magnification prediction results in the ERBCa test patches set are summarized. The DenseNet architecture trained from scratch to regress magnification led to a better classification performance than any other method separately, this is likely due to two factors: First, since is not computing an intermediate area, the network is less prone to introduce noise of overlapping classes measured by the area, and secondly, since it doesn't start with pre-trained weights from Imagenet, it has more flexibility to learn appropriate filters for histopathology images. Three combinations of both approaches were explored: Concatenation of the feature vectors, using the average-area learned weights to then fine-tune to regress magnification, and linearly combining the magnifications predicted by the area and the direct approaches, i.e.: $\alpha \times \text{Densenet}_{area} + 1 - \alpha \times \text{Densenet}_{magnif.}$. From this experiments, the first two

Table 2. Left: classification results in the test set for all the compared methods. Right: confusion matrix for the best model: linear mixture of DenseNet models ($\alpha = 0.2$).

Method	Kappa	F1-Score		5×	8×	10×	15×	20×	30×	40×
ShallowNet	0.720	0.779	5×	13	1	4	0	0	0	0
DenseNet	0.888	0.911	8×	0	169	45	0	0	0	0
Pre-trained Densenet	0.616	0.693	10×	0	46	331	60	0	0	0
(Area) Kumar	0.079	0.037	15×	0	0	3	748	113	0	0
(Area) ShallowNet	0.506	0.619	20×	0	0	0	1	1097	1	0
(Area) DenseNet	0.713	0.772	30×	0	0	0	0	15	1364	22
(Area) Pre-trained Densenet	0.734	0.789	40×	0	0	0	0	0	177	1381
(Area) GT Area	0.615	0.692								
Fusion (α)area + (1- α)direct	0.888	0.912								
Fusion (PMC 3-Class subset)	0.16	0.392								

did not show any significant improvement in the test set over the two approaches separately, thus not reported here. The third one led to a slightly better performance than the direct approach, using an alpha value of 0.2, and is the model which is reported in the confusion matrix for Table 2. In the area regression scenario, the two DL regressors presented are consistently closer to the ground truth average area than the baseline method. The baseline works very well on the lower-medium magnifications but fails at capturing the changes in big nuclei. Examples of patches are presented in Fig. 4. The class-activation maps are computed using the Grad-CAM method in the last dense layer as implemented in Keras-vis³. Both networks were implemented in Keras and optimized using Adam with initial learning rates explored logarithmically between 0.01 and 10^{-9} . The best learning rates were found to be 0.01, 0.0001, and 0.0001 for the ImageNet pre-trained and trained from scratch DenseNet and ShallowNet respectively. The best performing model on the test ERBCa patches was also evaluated on TCGA-PRAD and PMC databases.

PMC Histopathology Prostate Images: Since the PMC images are directly from articles the size and resolution of the images varies widely. Only the central 224×224 pixels in RGB channels were considered as this is the input size for our network. The predictions for most image patches are accurate as shown in Fig. 4, even with unknown stainings as the first two examples show. The lower magnifications (very small nuclei) are more challenging and, as a result, some of the predictions for those images are not correct as shown in the bottom-right images. A random selection of 55 images from were the magnification are available from the captions were selected to perform a quantitative test. In Fig. 3 a t-SNE embedding shows how the images at 20× tend to cluster in a single part of the feature space, whereas the 10 and particularly 5× images are more spread across since their differences with closer magnifications are more subtle, as also seen in the quantitative results in the ERBCa patches.

³ keras-vis [https://github.com/raghakot/keras-vis\(2017\)](https://github.com/raghakot/keras-vis(2017)).

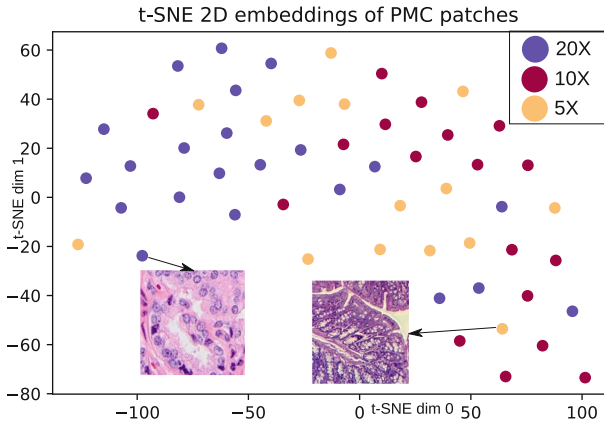


Fig. 3. 2D t-SNE embedding from the 1024-dimensional representation of 55 randomly selected PMC prostate images.

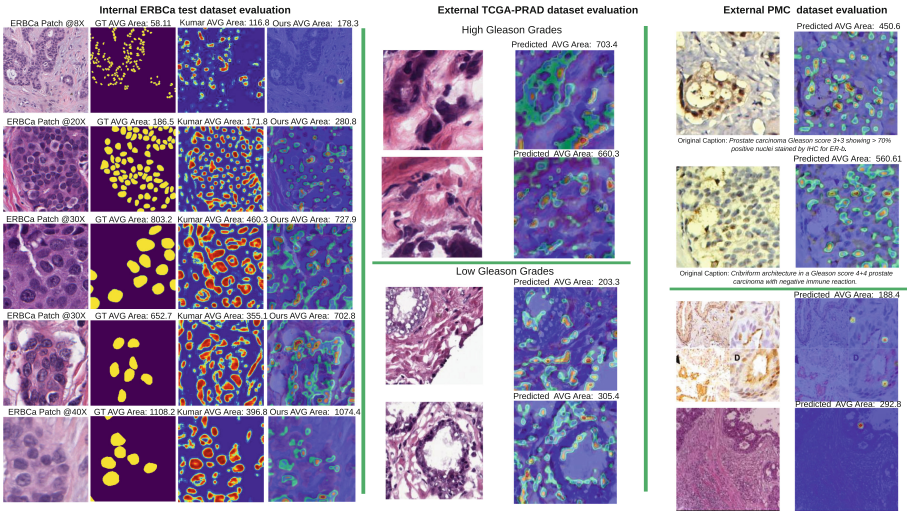


Fig. 4. Predictions of the regressor in all the datasets. Left: ERBCa test patches at several magnifications, our model is more robust in higher magnifications than the automatic segmentation. Middle: TCGA-PRAD patches from low and high cancer grades, showing dissimilar predicted average areas. Right: PMC prostate images: the first two predictions are consistent with manual annotations, the bottom ones are incorrect.

In the TCGA-PRAD dataset 92% of the patches were classified correctly at 20× using the area-magnification fusion approach.

4 Conclusion

In this paper, two CNN architectures were trained to regress the magnification level in histopathology images, using a direct regression approach or by first learning the average area of the nuclei. For internal evaluation, the best magnification regressor was a linear combination of two DenseNets: One trained to regress the area and the other to regress the magnification, this model had the best performance in terms of Kappa and F1-score, suggesting that a complementarity between the two models exists. In the case of the area regression a comparison is done with a state-of-the-art DL segmentation method, showing better overall performance as measured in MAE and F1-scores. Finally, the predictions of our model on the TCGA and PMC databases were accurate for a subset of filtered prostate images. Our model was able to generalize to several tissue types and provides useful information for exploiting the content in open access databases of histopathology images.

References

1. Bayramoglu, N., Kannala, J., Heikkilä, J.: Deep learning for magnification independent breast cancer histopathology image classification. In: 2016 23rd International Conference on Pattern Recognition (ICPR), pp. 2440–2445. IEEE (2016)
2. Cruz-Roa, A., et al.: Accurate and reproducible invasive breast cancer detection in whole-slide images: a deep learning approach for quantifying tumor extent. *Sci. Rep.* **7**, 46450 (2017)
3. Huang, G., Liu, Z., Weinberger, K.Q., van der Maaten, L.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2017)
4. Janowczyk, A., Madabhushi, A.: Deep learning for digital pathology image analysis: a comprehensive tutorial with selected use cases. *J. Pathol. Inf.* **7**, 29 (2016)
5. Jimenez-del-Toro, O., Otálora, S., Atzori, M., Müller, H.: Deep multimodal case-based retrieval for large histopathology datasets. In: Wu, G., Munsell, B.C., Zhan, Y., Bai, W., Sanroma, G., Coupé, P. (eds.) Patch-MI 2017. LNCS, vol. 10530, pp. 149–157. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67434-6_17
6. Kumar, N., Verma, R., Sharma, S., Bhargava, S., Vahadane, A., Sethi, A.: A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE Trans. Med. Imaging* **36**(7), 1550–1560 (2017)
7. Litjens, G., et al.: Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Sci. Rep.* **6**, 26286 (2016)
8. Otálora, S., Perdomo, O., Atzori, M., Andresson, M., Hedlund, M., Müller, H.: Determining the scale of image patches using a deep learning approach. In: IEEE 15th International Symposium on Biomedical Imaging, ISBI 2018. IEEE, April 2018