

# H&E stain augmentation improves generalization of convolutional networks for histopathological mitosis detection

David Tellez<sup>1,2</sup>, Maschenka Balkenhol<sup>1,2</sup>, Nico Karssemeijer<sup>1</sup>, Geert Litjens<sup>1,2</sup>, Jeroen van der Laak<sup>1,2</sup>, and Francesco Ciompi<sup>1,2</sup>

<sup>1</sup>Diagnostic Image Analysis Group, Radboud University Medical Center, Nijmegen, The Netherlands

<sup>2</sup>Department of Pathology, Radboud University Medical Center, Nijmegen, The Netherlands

## ABSTRACT

The number of mitotic figures per tumor area observed in hematoxylin and eosin (H&E) histological tissue sections under light microscopy is an important biomarker for breast cancer prognosis. Whole-slide imaging and computational pathology have enabled the development of automatic mitosis detection algorithms based on convolutional neural networks (CNNs). These models can suffer from high generalization error, i.e. trained networks often underperform on datasets originating from pathology laboratories different than the one that provided the training data, mainly due to the presence of inter-laboratory stain variations. We propose a novel data augmentation strategy that exploits the properties of the H&E color space to simulate a broad range of realistic H&E stain variations. To our best knowledge, this is the first time that data augmentation is performed directly in the H&E color space, instead of RGB. The proposed technique uses color deconvolution to transform RGB images into the H&E color space, modifies the H&E color channels stochastically, and projects them back to RGB space. We trained a CNN-based mitosis detector on homogeneous data from a single institution, and tested its performance on an external, multicenter cohort that contained a wide range of unseen H&E stain variations. We compared CNNs trained with and without the proposed augmentation strategy and observed a significant improvement in performance and robustness to unseen stain variations when the new color augmentation technique was included. In essence, we have shown that CNNs can be made robust to inter-lab stain variation by incorporating extensive stain augmentation techniques.

**Keywords:** Mitosis detection, computational pathology, data augmentation, convolutional neural networks

## 1. INTRODUCTION

Mitosis is a crucial phase in the cell cycle of eukaryotic cells where a replicated set of chromosomes is divided into two individual cell nuclei. These chromosomes are visible in hematoxylin and eosin (H&E) stained slides using light microscopy and enable pathologists to identify mitotic figures, i.e. cell nuclei undergoing mitosis. Assessing the number of mitotic figures per mm<sup>2</sup> is part of routine practice of grading breast cancer specimens<sup>1</sup>.

The recent introduction of whole-slide scanners in clinical pathology enables pathologists to make their diagnoses on digitized slides. Moreover, it paves the way for image analysis algorithms like automated mitosis

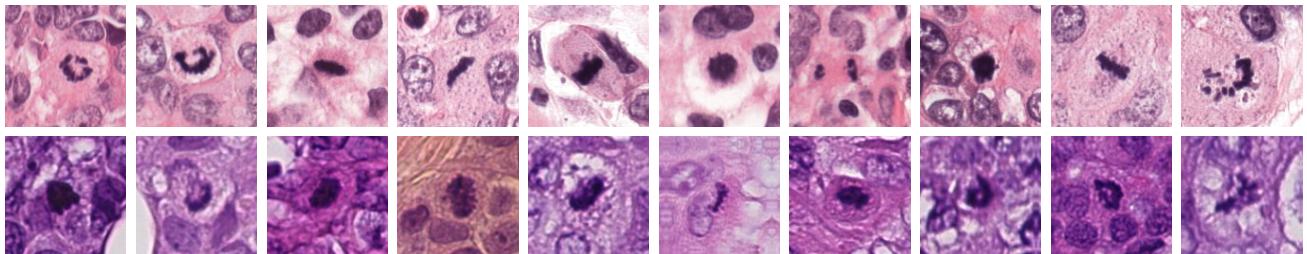


Figure 1. Examples of mitotic figures with variable H&E stain. Top row: patches from the *Radboudumc* dataset used to train the mitosis detector. Bottom row: patches from the *TUPAC* dataset used exclusively to test the mitosis detector.

<b>Function</b>	<b>Filters</b>	<b>Size</b>	<b>Stride</b>
conv	32	3x3	1
conv	32	3x3	2
conv	64	3x3	1
conv	64	3x3	2
conv	128	3x3	1
conv	128	3x3	1
conv	256	3x3	1
conv	256	3x3	1
conv	512	14x14	1
dropout	-	-	-
conv	2	1x1	1

Table 1. Architecture of the CNN.

detection. There have been several attempts to develop automatic mitosis detection methods and a number of public challenges have been organized on the subject<sup>2–5</sup>.

The problem of high generalization error is particularly acute for automatic mitosis detection with convolutional neural networks (CNNs)<sup>6,7</sup>, i.e. trained networks often underperform on datasets originating from pathology laboratories different than the one that provided the training data. The major factor causing this phenomena is the inter-laboratory stain variability, where similar morphological structures present variable appearance in the color space, as illustrated in Figure 1.

Several solutions have been proposed to solve or cope with this problem: (1) building multicenter training datasets, (2) preprocessing training and testing images with stain standardization techniques<sup>8–10</sup>, and (3) including data augmentation techniques during the model training<sup>11,12</sup>.

We argue that designing specific data augmentation strategies for H&E stained tissue images is the most promising approach to reduce the generalization error of these networks, avoiding the elevated costs of assembling a multicenter cohort, and effectively enforcing certain stain invariance into the trained models.

We propose a novel data augmentation algorithm that exploits the intrinsic properties of the H&E color space, and mimics common artifacts that happen during the staining and scanning process. By employing this data augmentation method, we were able to substantially improve the classification performance of a CNN trained for mitosis detection on homogeneous data from our home institution, when tested on external data from multiple centers.

## 2. METHODS

### 2.1 Datasets

In order to train the mitosis detector, we assembled a dataset of 27000 and 1 million patches of regions with and without mitotic figures, respectively, extracted from 18 triple negative breast cancer whole-slide images from the Radboud University Medical Center, and labeled it as the *Radboudumc* dataset. All whole-slide images were stain standardized<sup>8</sup> to minimize stain variations that the detector could learn from.

Additionally, we used a publicly available, multicenter dataset consisting of exhaustively annotated breast cancer tumor regions<sup>5</sup>, labeled as the *TUPAC* dataset. These images were not stain standardized and were used exclusively to evaluate whether our mitosis detector can generalize to unseen stain variations.

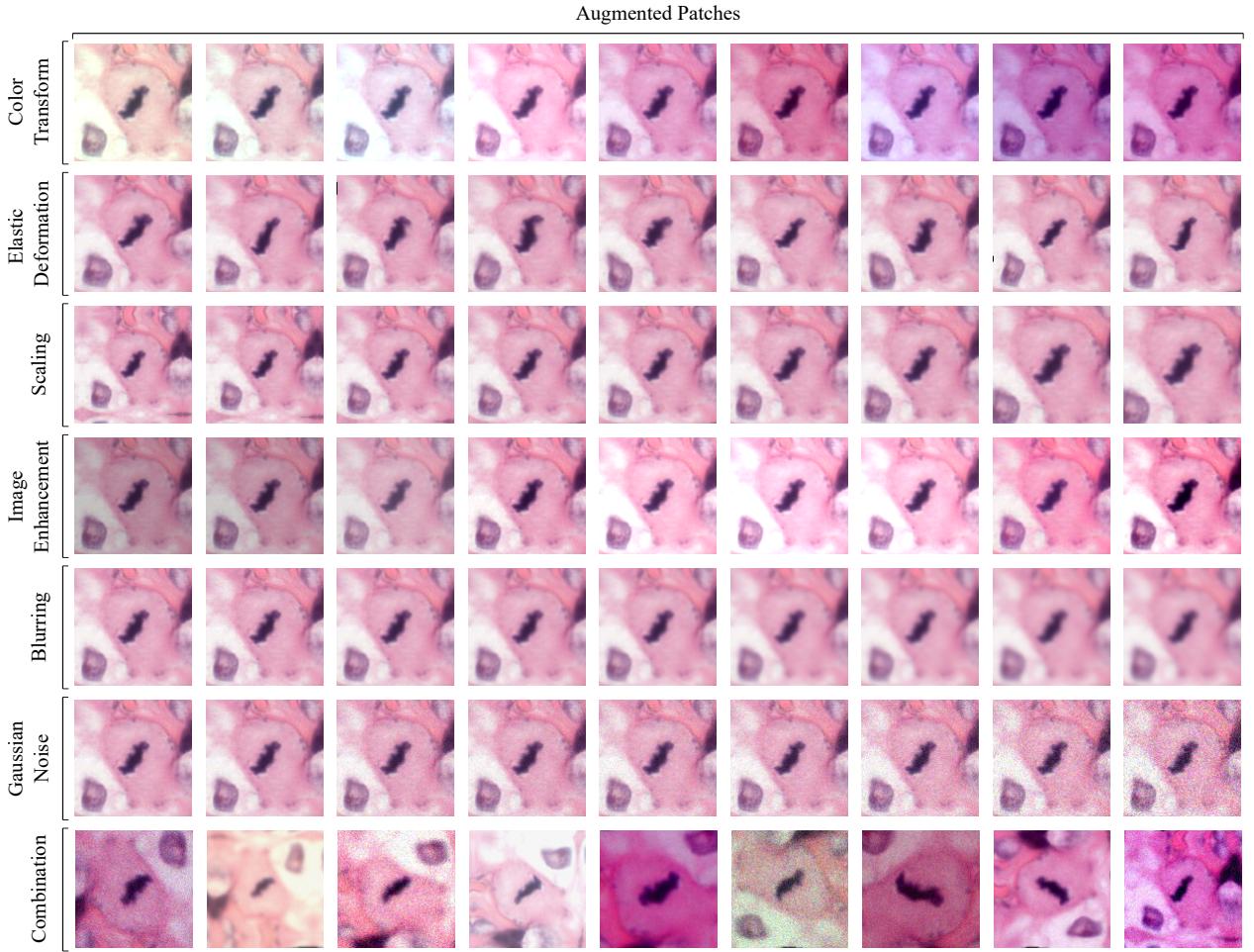


Figure 2. Multiple versions of the same mitotic patch modified with data augmentation routines. The first row corresponds to the novel *color stain augmentation* technique. *Image enhancement* includes brightness, contrast and color intensity changes. The *Combination* row includes rotation and mirroring additionally to the rest of the augmentation techniques.

## 2.2 Mitosis detector

We trained a CNN to classify whether a given tissue patch depicted a mitotic figure in its center or not. Given the large class imbalance found in the *Radboudumc dataset*, it was necessary to identify highly informative negative samples to train an effective mitosis classifier, i.e. negative examples that resembled positive ones, where the network could learn the most.

We proceeded similarly as stated in<sup>6</sup>. First, we built an *easy* training dataset by including all positive patches and a number of uniformly sampled negative patches, and trained an auxiliary network with it. Second, we evaluated all negative patches with this auxiliary network, obtaining a prediction probability for each of them. Finally, we built a *difficult* training dataset by selecting all positive images, and a number of negative patches sampled proportionally to their probability of being mitosis, so that harder samples were chosen more often.

We trained a second and definitive CNN on the *difficult* dataset to effectively distinguish between mitotic and non-mitotic H&E patches. During training, we applied several techniques to augment the dataset on-the-fly, preventing overfitting and improving generalization, that are described in detail in the following section.

Both CNNs, the auxiliary and the final one, were trained to minimize the cross-entropy loss of the network outputs with respect to the patch labels, using stochastic gradient descent with Adam optimization and mini-

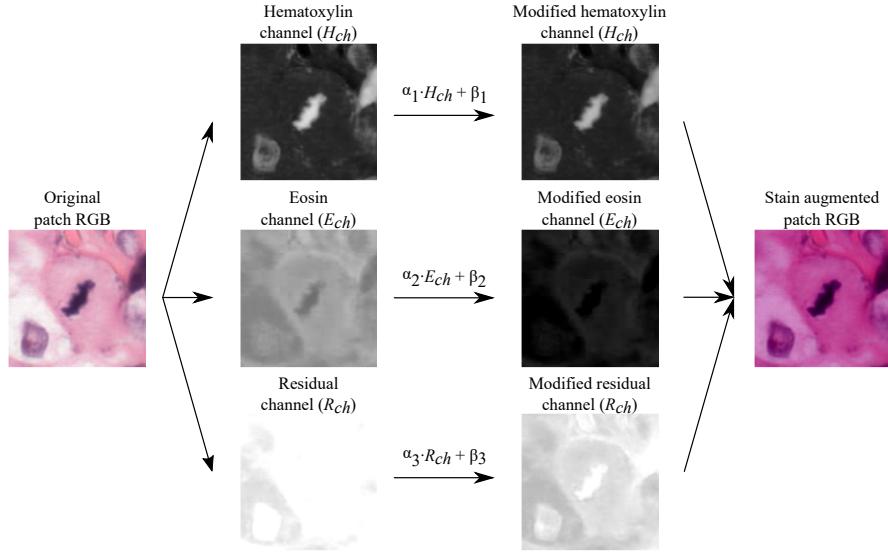


Figure 3. Diagram describing the color stain augmentation algorithm: (1) a RGB patch is transformed to H&E+R color space; (2) each stain channel is individually perturbed with a random factor  $\alpha_i$  and bias  $\beta_i$ ; and (3) the resulting channels are merged and transformed back to RGB color space.

batch of 64 samples. Their architecture is depicted in Table 1. To prevent overfitting, an additional L2 term was added to the network loss, with a factor of  $1 \times 10^{-5}$ . Furthermore, the learning rate was exponentially decreased from  $1 \times 10^{-3}$  to  $3 \times 10^{-5}$  through 20 epochs, defining an epoch as a set of iterations where all unique training samples were sampled once.

### 2.3 Data augmentation

We implemented several data augmentation routines tailored for H&E histopathology imaging that can be grouped in three categories attending to the feature invariance enforced during the network training, illustrated in Figure 2.

**Morphology invariance.** We exploited the fact that mitotic figures can have variable shape and size by augmenting the training patches with rotation, vertical and horizontal mirroring, scaling, elastic deformation<sup>13</sup>, and image translation around the central pixel.

**Stain invariance.** In H&E stain, hematoxylin and eosin are two chemical compounds that dye cell nuclei in blue, and the rest of structures in multiple shades of pink, respectively. The intensity of each color channel is dependent on multiple factors of the staining procedure in the lab. Unfortunately, these two color channels cannot be captured individually with regular slide scanners, instead, RGB images are taken. We used a novel approach to simulate a broad range of realistic H&E stained images by retrieving and modifying the intensity of the hematoxylin and eosin color channels directly (branded as color stain augmentation), as illustrated in the first row of Figure 3.

First, we transformed each patch sample from RGB to H&E color space using a method based on color deconvolution<sup>14</sup>. Second, we modified each channel  $i$  individually, i.e. hematoxylin ( $H_{ch}$ ), eosin ( $E_{ch}$ ) and residual ( $R_{ch}$ ), with random factors  $\alpha_i$  and biases  $\beta_i$  taken from two uniform distributions with ranges [0.95, 1.05] and [-0.05, 0.05], respectively. Finally, we combined and projected the resulting color channels back to RGB.

Additionally, we simulated further alternative stains by modifying image brightness, contrast and color intensity (from grayscale to full-color) using standard image enhancement routines, as illustrated in the fourth row of Figure 3.

Augmentation	Radboudumc	TUPAC
None	0.612	0.156
Rotation (R)	0.653	0.032
Color stain (C)	0.590	<b>0.417</b>
Scaling (S)	0.606	0.199
Elastic deformation (E)	0.616	0.187
Image enhancement (H)	0.606	0.198
Blurring (B)	0.598	0.170
Additive Gaussian noise (G)	0.608	0.154
Combined RSEB	0.653	0.018
Combined RCSEB	0.642	0.412
Combined RCSEHBG	0.628	<b>0.613</b>

Table 2. Quantitative impact of data augmentation. Each row represents a CNN trained with a different augmentation strategy and shows its classification performance on two datasets: the *Radboudumc dataset* (similar to the training dataset) and the *TUPAC dataset* (completely unseen). Image enhancement includes brightness, contrast and color intensity shifts.

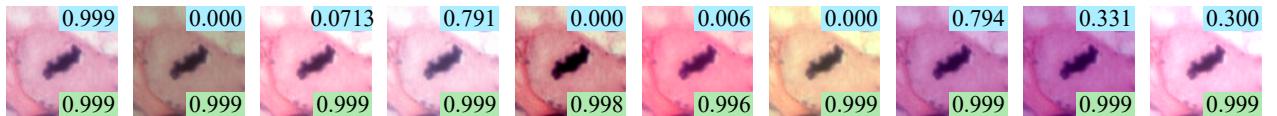


Figure 4. Qualitative analysis of the impact of using color stain augmentation. Numerical results indicate predicted probabilities for several mitotic patches with identical morphology but different stain appearance. Predictions in blue are produced by the *RSEB* CNN (trained without color stain augmentation), and predictions in green are produced by the *RCSEB* CNN (trained with color stain augmentation).

**Artifact invariance.** We mimicked the out of focus artifact of whole-slide scanners with a Gaussian filter with stochastic sigma (also known as blurring), and added Gaussian noise to decrease the signal-to-noise ratio of the images, mimicking image quality related artifacts.

### 3. RESULTS

**Quantitative impact of data augmentation.** In order to further understand the role of data augmentation in the context of mitosis detection, we trained several instances of the final CNN with identical training protocol except for the data augmentation routines used during training. In particular, we tested how each component of the augmentation method, and some combinations of them, influenced the CNN’s ability to detect mitotic figures in external, unseen data with different stain, namely the *TUPAC dataset*. We measured the maximum F1-score obtained by each network on this dataset and reported the values on Table 2. For completeness, we also reported there the maximum F1-score obtained in the validation set of the *Radboudumc dataset*.

The proposed color stain augmentation algorithm was crucial to decrease the generalization error observed when applying the trained CNN to external data. In particular, the performance of the model on the external *TUPAC dataset* improved when this technique was used both alone, and combined with other augmentation strategies.

**Qualitative impact of data augmentation.** Table 2 reveals that our color stain augmentation routine provided a major boost in performance on external data, i.e. images with significantly different H&E stain. We inspected and compared the predicted probabilities generated by two different CNNs when applied to a small set of patches exhibiting identical morphology but different stain, included in Figure 4. These two networks corresponded to *RSEB* and *RCSEB* on Table 2, noticing that the second one was trained using our novel color stain augmentation technique.

Figure 4 illustrates that the network trained without color stain augmentation (*RSEB*) was unable to correctly classify patches with identical morphology but different stain appearance. The network trained with color stain augmentation (*RCSEB*) exhibited a higher degree of invariance to stain changes, suggesting that it had effectively learned to ignore this feature when discerning whether a patch contained a mitotic figure or not.

#### 4. DISCUSSION & CONCLUSION

The novel color stain augmentation routine proposed in this work is crucial to decrease the generalization error observed when applying the trained CNN to external data. The performance of the classifier on the *TUPAC* dataset improves when this technique is used alone and in conjunction with other data augmentation strategies.

It is worth noticing that the best result is obtained when all the augmentation techniques are combined together, suggesting that the lack of generalization exhibited by some of the networks tested in the previous section could be due to not only stain variance but to other factors such as morphological differences and the presence of artifacts in the test images.

We have shown that using the proposed data augmentation strategy enabled our trained CNN to perform well on a large variety of H&E stains not present on the training dataset in the context of mitosis detection. We argue that this strategy could be extended to other applications in computational pathology, potentially reducing the need for stain standardization and the use of expensive multicenter cohorts for training purposes.

To our best knowledge, this is the first time that data augmentation is performed directly in the H&E color space, instead of RGB, in the context of mitosis detection with CNNs. We propose a novel data augmentation algorithm that exploits the properties of the H&E color space and simulates a broad range of realistic H&E stain variations. By using this method, we were able to substantially improve the classification performance of a CNN trained on homogeneous data from our home institution, when tested on external data from multiple centers with a variety of unseen H&E stain variations. As such, we have shown that CNNs can be made robust to inter-lab stain variation by incorporating extensive stain augmentation techniques.

#### ACKNOWLEDGMENTS

We gratefully acknowledge the support of NVIDIA Corporation with the donation of a Titan X GPU card for our research, and the developers of Theano<sup>15</sup> and Lasagne<sup>16</sup>, the open source tools that we used to run our experiments.

#### REFERENCES

- [1] Bloom, H. and Richardson, W., “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).
- [2] Roux, L., Racoceanu, D., Loméne, N., Kulikova, M., Irshad, H., Klossa, J., Capron, F., Genestie, C., Le Naour, G., Gurcan, M. N., et al., “Mitosis detection in breast cancer histological images an icpr 2012 contest,” *Journal of pathology informatics* **4**(1), 8 (2013).
- [3] Veta, M., Van Diest, P. J., Willems, S. M., Wang, H., Madabhushi, A., Cruz-Roa, A., Gonzalez, F., Larsen, A. B., Vestergaard, J. S., Dahl, A. B., et al., “Assessment of algorithms for mitosis detection in breast cancer histopathology images,” *Medical image analysis* **20**(1), 237–248 (2015).
- [4] Roux, L., “Mitosis detection in breast cancer histological images,” (2014). <https://mitos-atypia-14.grand-challenge.org>.
- [5] Veta, M., “Tumor proliferation assessment challenge,” (2016). <http://tupac.tue-image.nl>.
- [6] Cireşan, D. C., Giusti, A., Gambardella, L. M., and Schmidhuber, J., “Mitosis detection in breast cancer histology images with deep neural networks,” in [*International Conference on Medical Image Computing and Computer-assisted Intervention*], 411–418, Springer (2013).
- [7] Goodfellow, I., Bengio, Y., and Courville, A., [Deep Learning], MIT Press (2016). <http://www.deeplearningbook.org>.

- [8] Bejnordi, B. E., Litjens, G., Timofeeva, N., Otte-Höller, I., Homeyer, A., Karssemeijer, N., and van der Laak, J. A., "Stain specific standardization of whole-slide histopathological images," *IEEE transactions on medical imaging* **35**(2), 404–415 (2016).
- [9] Macenko, M., Niethammer, M., Marron, J., Borland, D., Woosley, J. T., Guan, X., Schmitt, C., and Thomas, N. E., "A method for normalizing histology slides for quantitative analysis," in [*Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on*], 1107–1110, IEEE (2009).
- [10] Khan, A. M., Rajpoot, N., Treanor, D., and Magee, D., "A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution," *IEEE Transactions on Biomedical Engineering* **61**(6), 1729–1738 (2014).
- [11] Veta, M., van Diest, P. J., Jiwa, M., Al-Janabi, S., and Pluim, J. P., "Mitosis counting in breast cancer: Object-level interobserver agreement and comparison to an automatic method," *PloS one* **11**(8), e0161286 (2016).
- [12] Liu, Y., Gadepalli, K., Norouzi, M., Dahl, G. E., Kohlberger, T., Boyko, A., Venugopalan, S., Timofeev, A., Nelson, P. Q., Corrado, G. S., et al., "Detecting cancer metastases on gigapixel pathology images," *arXiv preprint arXiv:1703.02442* (2017).
- [13] Simard, P. Y., Steinkraus, D., Platt, J. C., et al., "Best practices for convolutional neural networks applied to visual document analysis," in [*ICDAR*], **3**, 958–962, Citeseer (2003).
- [14] Ruifrok, A. C., Johnston, D. A., et al., "Quantification of histochemical staining by color deconvolution," *Analytical and quantitative cytology and histology* **23**(4), 291–299 (2001).
- [15] Theano Development Team, "Theano: A Python framework for fast computation of mathematical expressions," *arXiv e-prints* **abs/1605.02688** (May 2016).
- [16] Dieleman, S., Schlter, J., Raffel, C., Olson, E., Snderby, S. K., Nouri, D., et al., "Lasagne: First release.," (Aug. 2015).