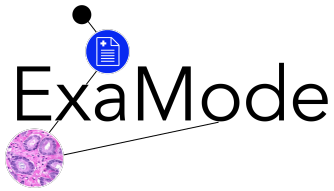


“D3.3”

Tools to extract homogeneous representations of heterogeneous colour visual information

Version: 1.0

Last Update: 29/12/20

Distribution Level: *PU*

Distribution level

PU = Public,
RE = Restricted to a group of the specified Consortium,
PP = Restricted to other program participants (including Commission Services),
CO= Confidential, only for members of the ExaMode Consortium (including the Commission Services)

Project supported by European Union
Horizon 2020 grant agreement 825292

The ExaMode Project Consortium groups the following Organizations:

Partner Name	**Short name**	**Country**
HAUTE ECOLE SPECIALISEE DE SUISSE OCCIDENTALE	HES-SO	Switzerland
UNIVERSITA DEGLI STUDI DI PADOVA	UNIPD	Italy
SIRMA AI	SIRMA AI	Bulgaria
STICHTING KATHOLIEKE UNIVERSITEIT	RADBOUDUMC	Netherlands
MICROSCOPEIT SP ZOO	MICROSCOPEIT	Poland
AZIENDA OSPEDALIERA PER L'EMERGENZA CANNIZZARO	AOEC	Italy
SURFSARA BV	SURFSARA BV	Netherlands

Document Identity

Creation Date: Last Update:	19/11/20 29/12/20

Revision History

Version	**Edition**	**Author(s)**	**Date**
0	0	Sebastian Otálora, Damian Podareanu	20/11/2020
Comments:	Defined libraries structure, content & means of distribution		
0	1	Sebastian Otálora	20/11/2020
Comments:	Defined deliverable structure & content, Included Executive Summary, stainlib methods		
0	2	Sebastian Otálora	26/11/2020
Comments:	Included stainlib examples, and stain augmentation methods, updated references		
0	3	Manfredo Atzori	27/11/2020
Comments:	Proofread the content		
0	4	Sebastian Otálora	30/11/2020
Comments:	Updated version taking into account Manfredo's comments		
0	5	Manfredo Atzori	01/12/2020
Comments:	Final proofread and check		
0	6	Francesco Ciompi	04/12/2020

Comments:	Add small changes to the text	
0	7	Sebastian Otálora
Comments:	Addressing reviewers comments	
0	8	Damian Podareanu
Comments:	Check & links added	
1	0	Manfredo Atzori
Comments:	Final formal check	

Executive summary

The automatic analysis of histopathology images stained with H&E helps clinicians in diagnosing and quantifying diseases. Computer vision methods based on deep learning have shown promising results in computational pathology. Nevertheless, the variability and heterogeneity in histopathology images due to batch effects and differences in preparation in different pathology laboratories and even in different days of the week, produce changes in tissue appearance and, consequently, to the digitized whole-slide image, impeding the application of the trained models in clinical scenarios where there is high variability in the images. In ExaMode, one of the objectives is to reduce the heterogeneity in visual information, particularly with Deliverable 3.3.

Deliverable 3.3 comprises a set of tools to extract homogeneous representations of heterogeneous colour visual information. This deliverable contains stain normalization and color augmentation techniques proposed in the literature, in a unified code source with homogeneous implementations of the most commonly used methods. We have made publicly available a library including these methods together with repositories for other state-of-the art techniques to deal with the heterogeneity of color information in histopathology images.

The tools presented in this deliverable are tested in order to verify Milestone 6 "Image content based knowledge discovery tools prototype test".

The deliverable report will be submitted as a scientific journal article.

Table of Contents

Table of Figures

Index of Tables

List of abbreviations

CNN Convolutional Neural Network

PMC PubMed Central

RGB Red Green Blue

H&E Haematoxylin and Eosin dyes

WSI Whole Slide Image

NMI Normalized Median Intensity

SD Standard Deviation

CV Coefficient of Variation

TCGA The Cancer Genome Atlas

AOEC Azienda Ospedaliera Cannizzaro – Catania

RUMS Radboud University Medical Center

DCGMM Deep Convolutional Gaussian Mixture Model

1 Introduction

During the last decade, the automatic analysis of digital pathology images has become more precise, thanks to the steady development of deep learning algorithms based on deep convolutional neural networks (CNN). Shifting from handcrafted features towards the end–to–end architectures made it possible to detect cancer in histopathology images at the image region and the whole-slide-image level. Methods have become more precise, achieving, in some cases, a performance that is comparable to pathologists for specific tasks [1].

Despite the remarkable performance of the methods, there are still technical barriers that prevent the translation of these advances into clinical applications. A useful deployed deep learning method should be able to cope with the heterogeneity in the color of the images that arise from the chemicals that stain them. Two of the most typical chemicals used in pathology routine to stain tissue slides are Hematoxylin and Eosin. These two chemicals highlight the nuclei with a dark purple color (Hematoxylin) and the cytoplasm with a light pink one (Eosin).

One of the most important factors preventing the application of machine learning methods to clinical practice is related to the heterogeneity of Hematoxylin and Eosin (H&E) images due to tissue preparation and several parameters involved in the tissue preparation and digital scanning process (temperature of the tissue, the thickness of the cuts, the image sensor of the digital camera, among others). An example of stain variation in train and test sets, and its impact in the performance of a CNN model trained only with partial variations is depicted in Figure 1.

Two approaches are mostly applied to keep into account such variations when training CNN models. First, methods that homogenize an input H&E image given a target image template are known as "stain normalization", where the aim is to match the input color distributions (or H&E concentrations) with the one given in a target image. The second approach refers to stain or color augmentation methods, where new synthetic samples are created to increase the size of the training datasets, making them more robust to color variance.

Several image processing and machine learning techniques reported in the literature deal with color heterogeneity, improving classification, and segmentation performance for various tissue types [2,3,4,5]. While the specific normalization technique depends on the task to solve [2,5], recent works have reported consistent improvements in performance and robustness to external datasets employing color augmentation techniques [2,4] or a combination of both techniques [5].

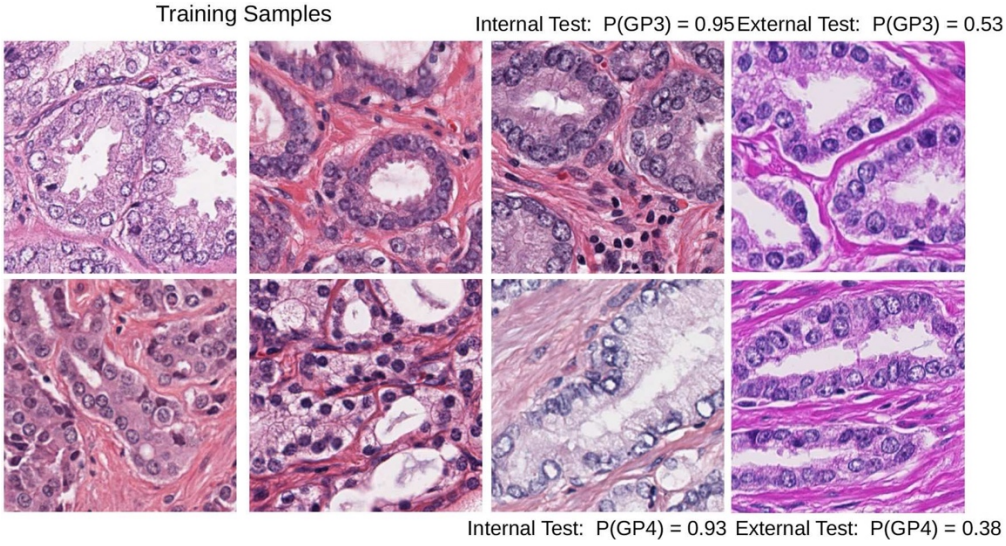


Figure 1 : Example of a model misclassification due to stain variation in the test set [2]. P(GP3) and P(GP4) stands for probability of the image region to contain Gleason patterns 3 and 4 respectively.

Few libraries comprise multiple methods for H&E image normalization and augmentation. Furthermore, there are just a handful of methods that tackle the problem of colour heterogeneity in H&E images using modern machine and deep learning techniques. The codebase from articles in literature is mostly in self-contained repositories, and its evaluation is usually in an ad-hoc task for specific datasets. The lack of libraries with multiple methods included, limits the possibilities to evaluate the best strategy to deal with color heterogeneity for a new dataset. Among the few existing software tools to deal with color heterogeneity, QuPath, Staintools, and HistomicksTK, are among the most popular.

QuPath (https://qupath.github.io/.) is an open and extensible software platform for WSI analysis. It includes methods for estimating and setting stain vectors. There are also scripts[1] created for running specific color normalization methods within QuPath. Due to the big codebase of QuPath, it might be challenging to run a classification or segmentation model without having to write a considerable amount of scripts to have a full pipeline, taking into account datasets with considerable color heterogeneity.

Staintools is a set of tools for tissue image stain normalization and augmentation in python 3. It contains implementations that follow the same coding style of scikit-learn, where the methods are made to fit or train a model. It is open-source and can be downloaded from the Github repository: https://github.com/Peter554/StainTools. The library contains two extractive normalization methods (Macenko, Vahadane), and the only augmentation techniques included in staintools are based on the same extractive methods by modifying the estimated stain concentrations.

HistomicksTK is a python toolkit for histopathology image analysis, it contains several methods for stain normalization and color augmentation based on the stain perturbation methods from the work of Tellez et al.[10]: It can be downloaded from https://pypi.org/project/histomicstk/.

Our aim to develop and test tools to extract homogeneous representations of heterogeneous colour visual information was achieved with this deliverable in two ways. First, we provide the community with an easy-to-use library (stainlib) that includes the most commonly used methods for colour augmentation and normalisation of histopathology images, having as input local image regions. Second, we increase the number of available methods that tackle heterogeneity of colour in H&E images with modern machine and deep learning techniques (iResFlow, stain invariant networks).

2 Stainlib

We developed stainlib as part of deliverable 3.3. Stainlib is a python library containing methods for H&E image normalization and augmentation. Our objective was to develop an easy-to-use python 3 library that includes the most commonly used methods for color augmentation and normalisation of histopathology images, having as input local image regions. In Figure 2 the structure and methods included in the library are displayed. The library can be downloaded

from github[1]. The library can be deployed as Python package directly from the repository or as Docker container that can be downloaded from the following link[2] (the color folder).

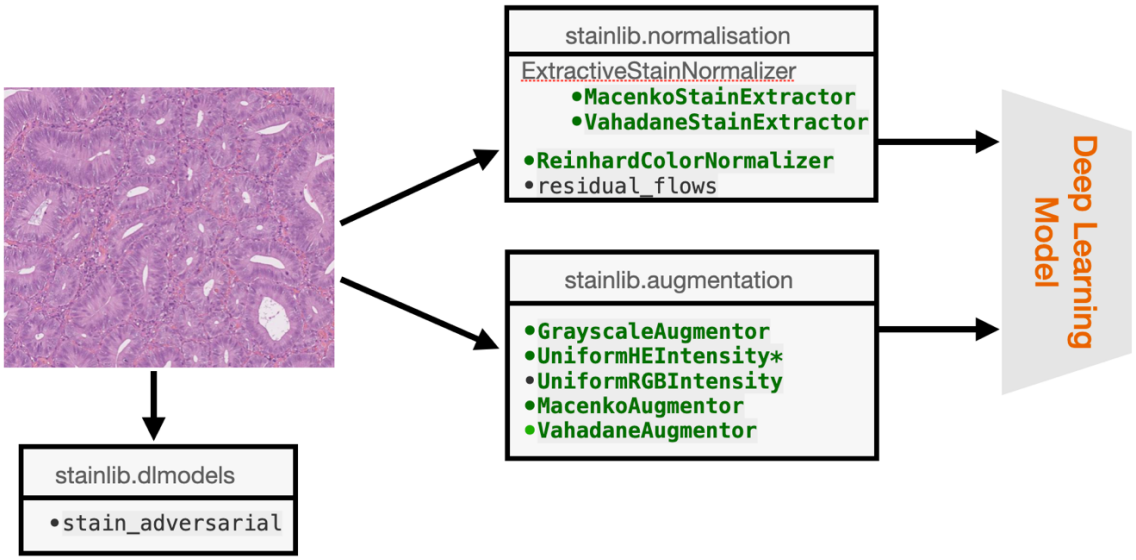


Figure 2: Implemented methods in stainlib for stain normalization and augmentation. The version 1.0 of the library includes all the methods in green and linked implementations of the residual_flows and stain_adversarial methods. For the next version of the library, further state-of-the-art methods developed in ExaMode will be included.

2.1 Stainlib.normalisation

All the pixels of the 3D H&E tissue slides are represented, by digital tissue scanners, in a computer as 2D digital images in the RGB space. Ideally, each pixel should contain a composition of the color representation of Hematoxylin, Eosin, and background. Images acquired from the same center and using the same preparation parameters should share similar stain absorbance coefficients, which can be written as the linear transformation (omitting background that should be close to 255 for the three channels):

Where the first-row vector corresponds to the RGB components of hematoxylin and the second one to the components of Eosin. In staining normalization methods, the aim is to estimate the individual staining absorbance coefficients of the image S and quantify the absorbed light C by the tissue when it was scanned, which is the value in the H&E space of each pixel. The Beer-Lambert law provides a way to estimate them in the optical density space, given the original pixel content for the *k*-channel :

Where *k* ranges in the three RGB channels, $S \in [0,+\infty]^{3\times 2}$ is the matrix of absorbance coefficients, $\boldsymbol{C} \in [0,+\infty]^2$ is the vector of the two staining concentration coefficients and the background value.

[1] https://github.com/sebastianffx/stainlib/
[2] https://surfdrive.surf.nl/files/index.php/s/PBBnjwzwMragAGd/

There are several well-known stain extraction methods that provide an estimation of *S*. In the widely used method of Macenko [6], this matrix is computed by calculating a plane using the two largest singular value decomposition vectors of the image and then projecting the data into this plane and clipping extreme values. In the method of Vahadane [7] this estimation is done by learning a sparse non-negative matrix factorization. In stainlib we include implementations of the Vahadane and Macenko methods using as source code base the implementations in the staintools library[3].

In the Reinhard method [8], the color histogram of the source image (in the LAB color space) is matched with the target image, despite its simplicity and its original domain of application of natural images, it yields good results in histopathology images as well. We have included Reinhard normalization method in stainlib.

2.2 Stainlib.augmentation

Deep learning classification and segmentation models for histopathology yield better results when data augmentation is used [2,3,4,9]. This might be intuitive in the context of training deep learning models, where the larger the amount of data the model is fed with, the more variations the model is exposed to. Therefore, it is more robust to changes in appearance in the test set. When there is a wide range of images with variations in color and sources of preparation (pathology laboratories) included in the training set, the models are more likely to output the correct prediction for new samples. Such range of variations could be synthetically simulated, especially for the color variations due to the stain concentrations.

In stainlib we included five stain augmentation methods: 1) grayscale transformation, 2) shifts of the stain concentrations in H&E space, 3) shifts of the stain concentrations in RGB space, 4) shifts of the stain concentration matrix using the Macenko method and 5) shifts of the stain concentration matrix using the Vahadane method.

For the RGB to grayscale transformation, the method rgb2gray from the scikit-image library is used, and realistic samples are generated using random uniform variations of the grayscale values within a fixed range as follows:

Where and are values drawn from a uniform distribution in the [0,1] range. Similarly, for the shifts of the RGB channels, we generate new samples as follows for each channel *c* as follows:

Again, and are values drawn from a uniform distribution in the [0,1] range.

In the case of the Macenko and Vahadane augmentation, we use the methods' estimated stain H&E concentration matrix and shift it as follows:

Where the values and are drawn from a uniform distribution in the [1-,1+] and [1-,1+] range. Values of =0.2 and =0.2 were set as default values as they yield good qualitative & quantitative results.

[3] Staintools github repository : https://github.com/Peter554/StainTools/tree/master/staintools

In the case of the augmentations based on the shifts of the stain concentrations in H&E space, we followed the implementation of Tellez et al [10].
In section 5.1 we present a qualitative evaluation of the normalization and augmentation methods included in stainlib, quantitative evaluation of this methods has also been done in the research products from the members of our consortium in [2,3,4].

3 iResFlow: Residual Flows

Flow-based generative models parameterize probability distributions through an invertible transformation and can be trained by maximum likelihood. Invertible residual networks provide a flexible family of transformations where only Lipschitz conditions rather than strict architectural constraints are needed for enforcing invertibility [11]. We have implemented the method to compute an invertible mapping of the color information in histopathology WSI patches, we named it iResFlow, the source code and instructions to use the method can be found in the following GitHub repository: https://github.com/sara-nl/color-information. Evaluation of iResFlow is done in section 5.3.

4 Stain Invariant Neural Networks

Domain-invariant training of CNN's is a promising technique to address training a single model for different domains. It includes the source domain information to guide the training towards domain-invariant features, achieving state-of-the-art results in classification tasks. In the case of training classification models with histopathology images, the domain represents the center where the tissue preparation characteristics are similar (e.g., hospital A, hospitals B), where the differences are visible in the stain concentrations (stain invariance). This technique shows excellent generalization performance to external test sets when, and even more, when combined with data augmentation techniques [4,13].

To explicitly write all the possible variations that lead to changes in the appearance of H&E images is infeasible. Stain invariant training of CNN models aims at detaching the domain or center information, where the changes in appearance originate, from the features that the model learns. In Figure 3, the stain invariant model inner working is displayed in a small neural network example. The CNN has shared features for both the mitosis/no-mitosis classifier and for the domain classifier, the main difference with a multi-task CNN, is that the gradients from the domain branch are reversed. The stain-invariant model will penalize when the learned features help classify the domain, guiding them towards features that do not consider the domain information.

Evaluation of stain invariant models is presented in section 5.4. and is described in detail in previously published journal article from the members of the consortium [4].

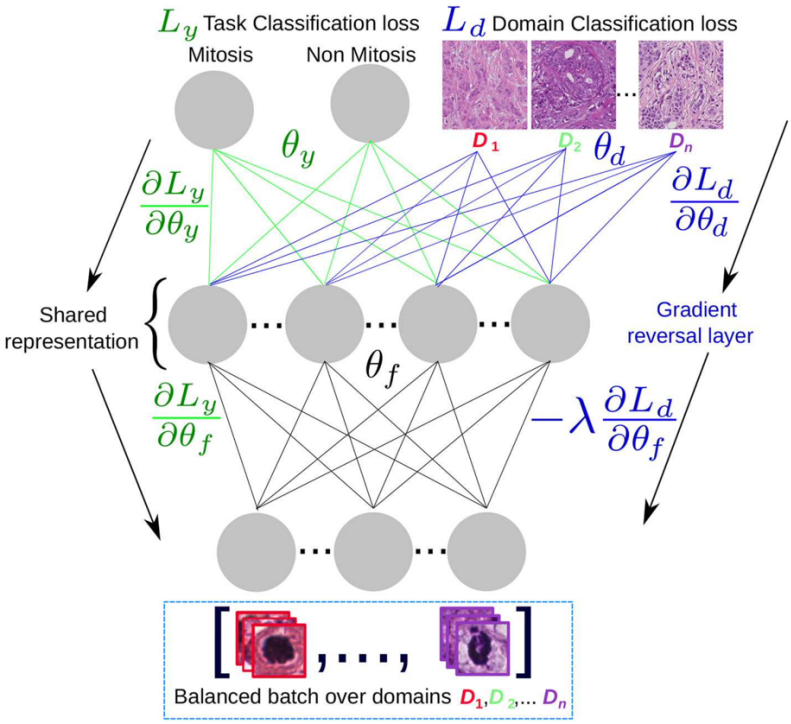


Figure 3: Domain adversarial scheme: A domain-balanced batch of images is passed as input to the network that has two types of outputs: the task classification output and the domain classification output. The shared representation θ_f is optimal for the task classification and unable to discriminate between the *n* domains.

5 Results and evaluation of the tools

Milestone 6 " Image content based knowledge discovery tools prototype test" is verified in this section. Here, we provide qualitative examples of normalisation and augmentation for stainlib (sections 5.1 and 5.2), and a summary of quantitative examples for iResFlow and stain invariant models in sections 5.3 and 5.4, respectively. Regarding the quantitative evaluation of the implemented methods, this has been already carried out by the members of the ExaMode consortium and disseminated in different research articles [3,4,5].

5.1 Stainlib normalisation examples

In this subsection we test qualitatively the normalization methods implemented in stainlib. We have gathered images from different sources and stain concentration from wikipedia with licence attribution-share alike 3.0 unported (Figure 4), the original image sources and python code is provided in the form of a jupyter notebook as annex.

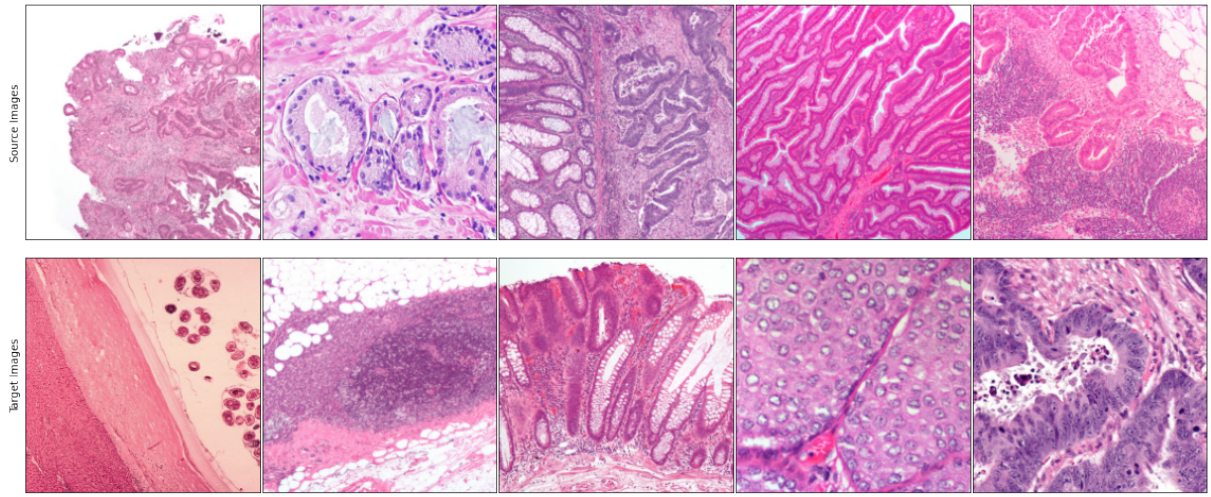


Figure 4: H&E images from used as a source to match the stain distributions in the target images. We choose images at different magnifications and different organs to evaluate the quality in heterogeneous scenarios.

Figure 5 and 6 show the results for the Vahadane and Macenko methods. Being both extractive methods, the results are similar, being the images normalized with Macenko methods slightly darker, or with more stain concentration than with the Vahadane method.

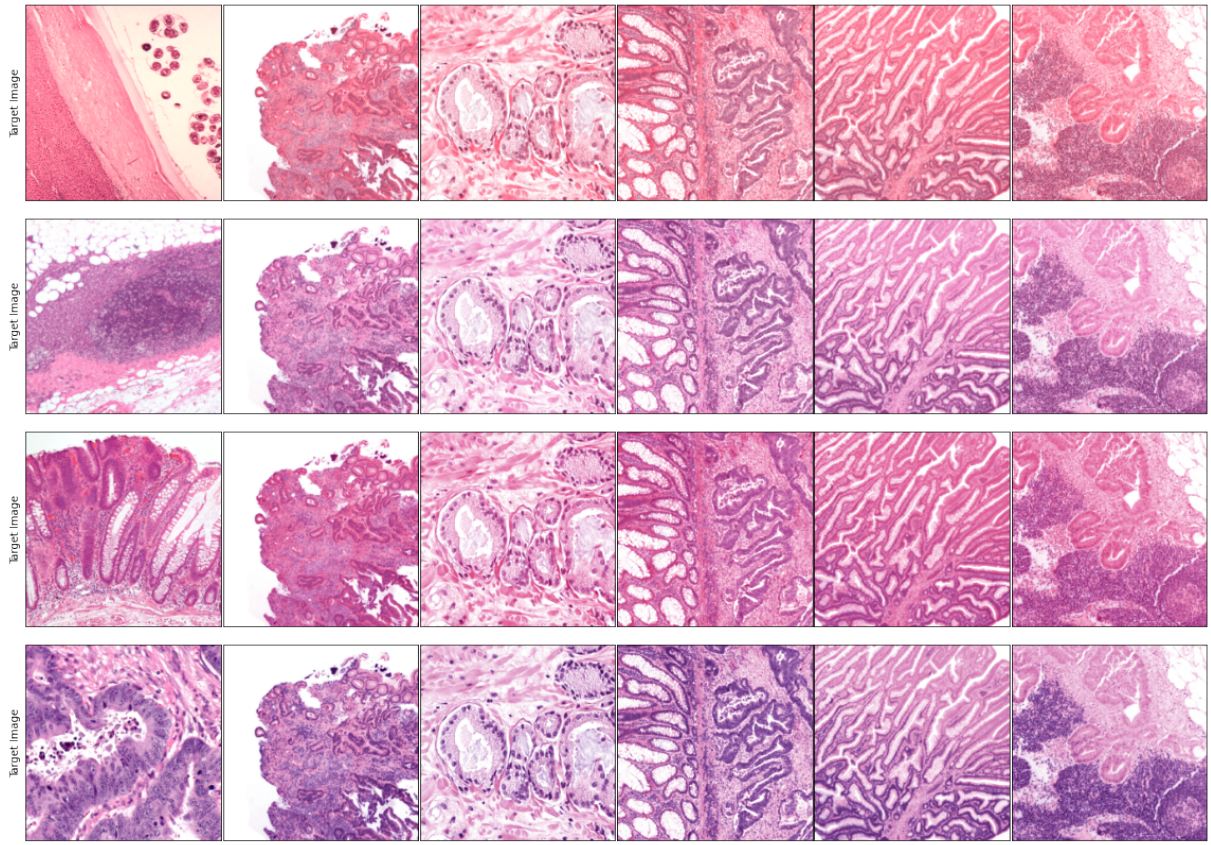


Figure 5: Results of normalisation using the extractive normalisation method of Vahadane.

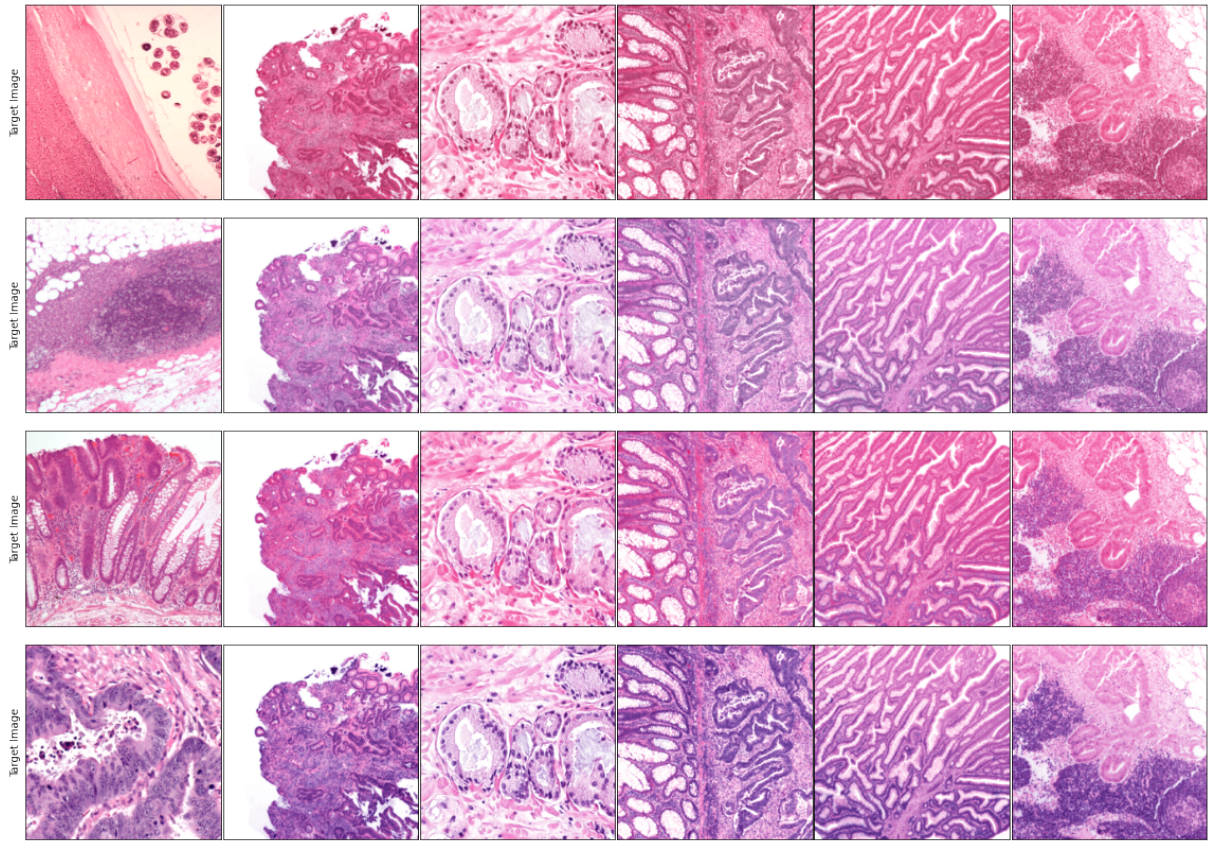


Figure 6: Results of normalisation using the extractive normalisation method of Macenko.

In figure 7, the results for the Reinhard normalisation method are presented. Because the method aims at directly matching the color histogram of the target image, the background is matched with the lightest colour in the target image, which produces unrealistic images. To alleviate this, we included a tissue detector method, that masks the tissue content from the background. This is now included as a parameter in the call for transforming new images using the Reinhard method[4].

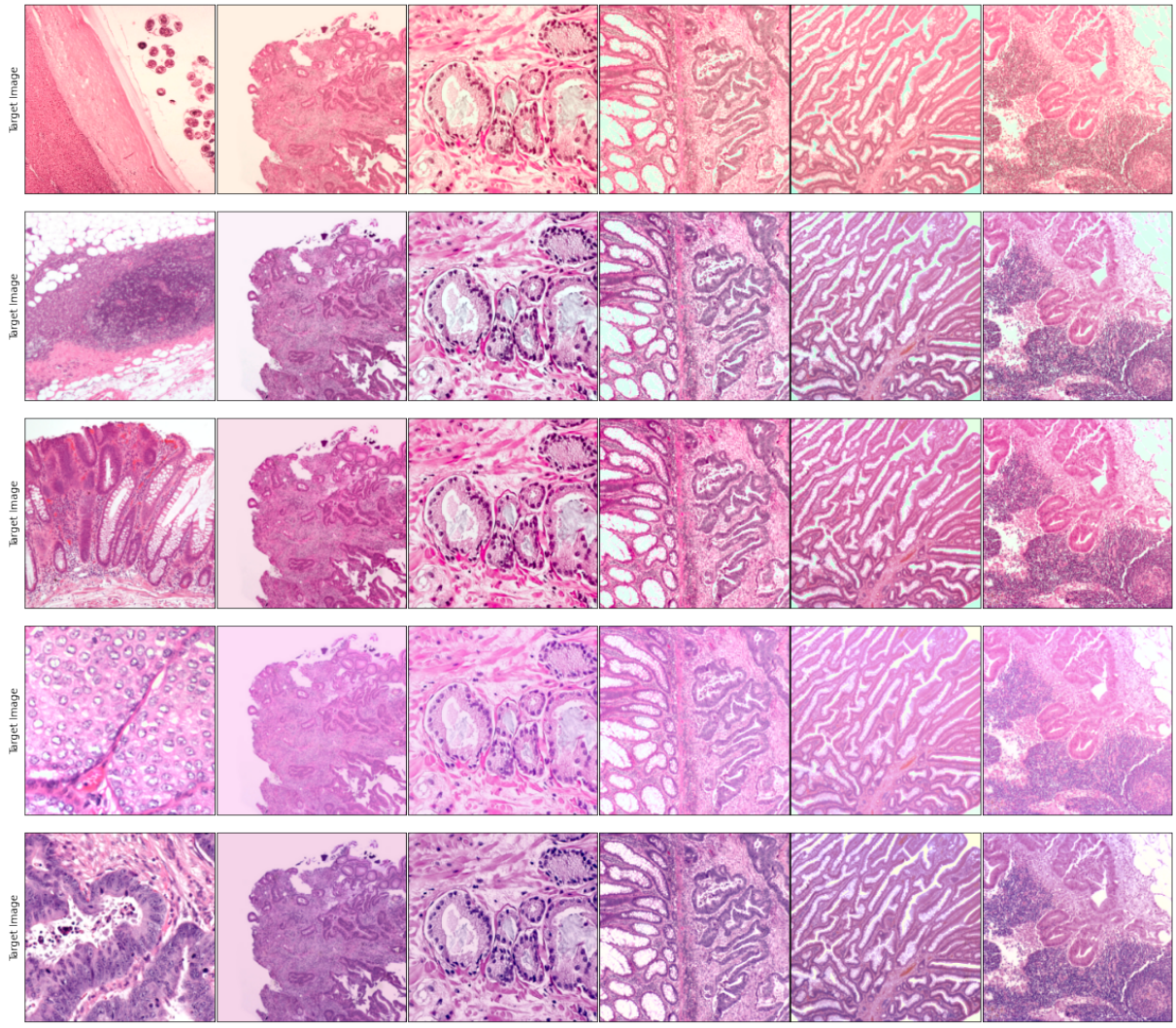


Figure 7: Results of normalisation using the extractive normalisation method of Reinhard.

[4] https://github.com/sebastianffx/stainlib/blob/main/normalization/normalizer.py

5.2 Augmentation

In Figure 8, ten augmented images are displayed for each method. For the grayscale image only the grey channel intensity is modified. In the second row, we show augmentations in the Hematoxylin, Eosin and DAB space (HED) using the Light HED augmentation method. The images generated with the Light HED augmentation method are sampled according to an uniform distribution in the H&E space, as described and evaluated in [2]. Finally, for the Macenko and Vahadane methods, the augmented versions are generated perturbing the estimated stains, as in the case for stain normalization, also here the images for both methods look similar.

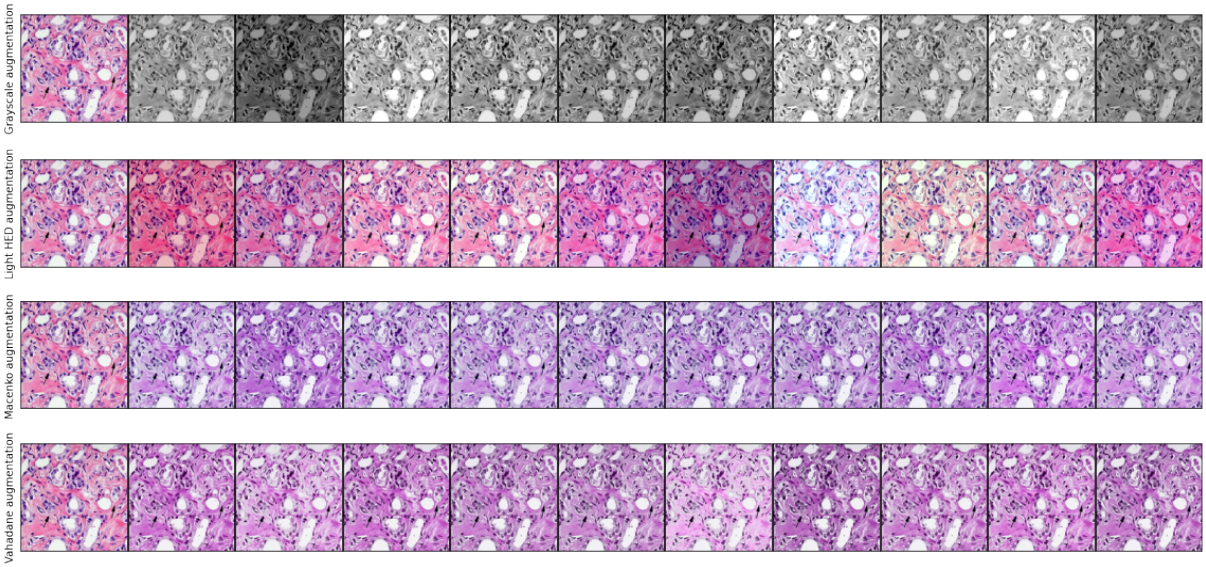


Figure 8: Results of data augmentation using the specified methods in section 2.2.

5.3 iResFlow evaluation

Table 1. Analysis of Normalized Median Intensity (NMI - scores). Circles are outliers. iResFlow has a bigger median performance despite having a SD bigger than DCGMM[14].

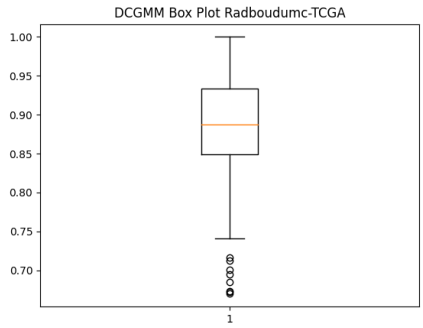
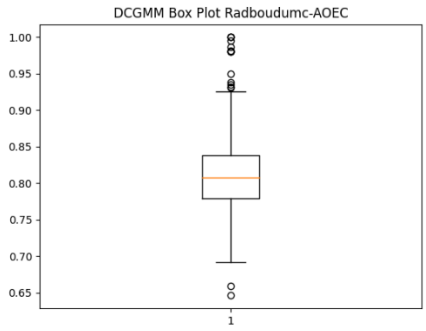
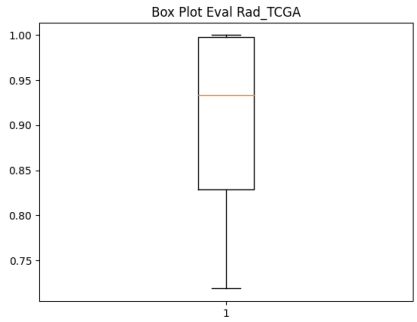
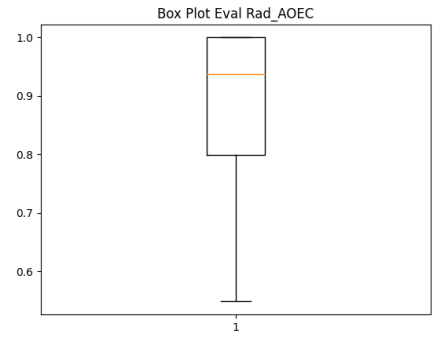
	RUMC (source) - TCGA (target)	RUMC (template) - AOEC (target)
DCGMM		
iResFlow		

Table 2. Results table for NMI SD and NMI CV metrics based on 5 runs of 100 256 x 256 patches of RUMC as template and TCGA as target data.

Model	NMI - Standard Deviation	NMI - Coefficient of Variation
DCGMM	0.0686 +- 0.0065	0.0776 +- 0.0110
iResFlow	0.0381 +- 0.0094	0.0425 +- 0.0148

Table 3. Results table for NMI SD and NMI CV metrics based on 5 runs of 100 256 x 256 patches of RUMC as template and AOEC as target data.

Model	NMI - Standard Deviation	NMI - Coefficient of Variation
DCGMM	0.0547 +- 0.0222	0.0670 +- 0.0249
iResFlow	0.0497 +- 0.0126	0.0563 +- 0.0170

5.4 Stain Invariant Neural Networks Evaluation

In Table 4 are the results for a mitosis detection task. The results indicate that domain adversarial is able to generalize better than stain normalization and color augmentation in the external test dataset. Detailed evaluation and qualitative results of the invariant features are presented in [4].

Table 4. Results table for stain invariant, stain normalisation, stain augmentation and their combinations in the TUPAC dataset of mitosis detection, adapted from [4]. The results show that using domain adversarial training is beneficial for generalizing to external test sets.

CNN model combinations	**Baseline CNN**						
Color augmentation		✓			✓	✓	✓
Staining normalization			✓		✓		✓
Domain adversarial				✓		✓	✓
Internal test set (F1-score)	0.8088 ± 0.02	**0.8117 ± 0.001**	0.7630 ± 0.04	0.6950 ± 0.379	0.7787 ± 0.03	0.6985 ± 0.01	0.6945 ± 0.02
External test set (F1-score)	0.71173 ± 0.02	0.7306 ± 0.07	0.5424 ± 0.01	**0.8236 ± 0.071**	0.5963 ± 0.1	0.6740 ± 0.01	0.5742 ± 0.009
Internal test set (AUC)	0.9596 ± 0.006	**0.9631 ± 0.005**	0.9351 ± 0.001	0.8972 ± 0.011	0.9503 ± 0.01	0.9030 ± 0.002	0.8871 ± 0.02
External test set (AUC)	0.8014 ± 0.01	0.8270 ± 0.06	0.848 ± 0.075	**0.9146 ± 0.003**	0.7925 ± 0.06	0.8446 ± 0.004	0.8255 ± 0.06

6 Discussion & Conclusion

In deliverable 3.3 we implemented and tested novel tools that extract homogeneous representations of heterogeneous colour visual information from H&E images: stainlib, iResFlow, and Stain Invariant networks. In stainlib, we have included the most commonly used methods not only for stain normalisation but also for augmentation. With iResFlow a novel method based on the maximum likelihood showed better results than state-of-the-art deep generative models. We anticipate continuum update of stainlib, making it efficient and useful not only for normalising image regions but also WSIs. The source-code for all the tools is now fully accessible in their repositories. We are confident that this resources will allow to build more robust models that generalize to unseen images from heterogeneous sources.

7 References

[1] Echle, A., Rindtorff, N.T., Brinker, T.J. *et al.* Deep learning in cancer pathology: a new generation of clinical biomarkers. *Br J Cancer* (2020). https://doi.org/10.1038/s41416-020-01122-x
[2] Tellez, David, et al. "Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology." Medical image analysis 58 (2019): 101544.
[3] Khan, Amjad, et al. "Generalizing convolution neural networks on stain color heterogeneous data for computational pathology." Medical Imaging 2020: Digital Pathology. Vol. 11320. International Society for Optics and Photonics, 2020.
[4] Otálora, Sebastian, et al. "Staining invariant features for improving generalization of deep convolutional neural networks in computational pathology." Frontiers in Bioengineering and Biotechnology 7 (2019): 198.
[5] Ciompi, Francesco, et al. "The importance of stain normalization in colorectal tissue classification with convolutional networks." 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). IEEE, 2017.
[6] Macenko, Marc, et al. "A method for normalizing histology slides for quantitative analysis." 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. IEEE, 2009.
[7] Vahadane, Abhishek, et al. "Structure-preserving color normalization and sparse stain separation for histological images." IEEE transactions on medical imaging 35.8 (2016): 1962-1971.
[8] Reinhard, Erik, et al. "Color transfer between images." IEEE Computer graphics and applications 21.5 (2001): 34-41.
[9] Komura, Daisuke, and Shumpei Ishikawa. "Machine learning methods for histopathological image analysis." Computational and structural biotechnology journal 16 (2018): 34-42.
[10] Tellez, David, et al. "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 (2018): 2126-2136.
[11] Chen, Ricky TQ, et al. "Residual flows for invertible generative modeling." Advances in Neural Information Processing Systems. 2019.
[12] Magee, Derek, et al. "Colour normalisation in digital histopathology images." Proc Optical Tissue Image analysis in Microscopy, Histopathology and Endoscopy (MICCAI Workshop). Vol. 100. Daniel Elson, 2009.
[13] Lafarge, Maxime, et al. "Learning domain-invariant representations of histological images." Frontiers in medicine 6 (2019): 162.
[14] Zanjani, Farhad Ghazvinian, and Svitlana Zinger. "Deep convolutional gaussian mixture model for stain-color normalization of histopathological images." International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2018.

Annexes

- Jupyter notebook with the code to reproduce the qualitative of the stain augmentation example figures of this document, accessible online: https://github.com/sebastianffx/stainlib/blob/main/stainlib_augmentation.ipynb
- Jupyter notebook with the code to reproduce the qualitative of the stain augmentation example figures of this document, accessible online: https://github.com/sebastianffx/stainlib/blob/main/stainlib_normalization.ipynb