Data-driven Confocal Microscopy to Hematoxylin and Eosin Transformation

Digitally Stained Confocal Microscopy through Deep Learning

Sergio García Campderrich

October 23, 2019

Contents

Contents

Contents

Contents

Contents

Introduction

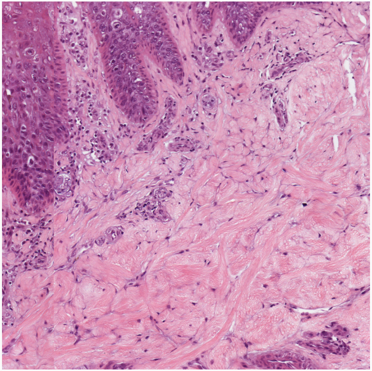
Introduction

Hematoxylin and Eosin (H&E)

Confocal microscopy

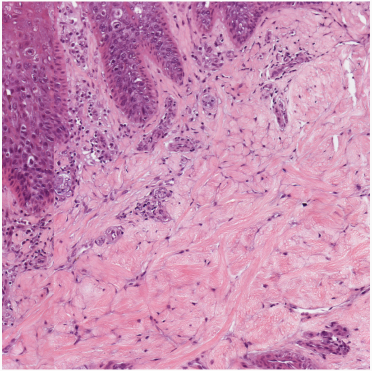
Data-driven transformation

What is it?



H&E stained tissue sample

What is it?

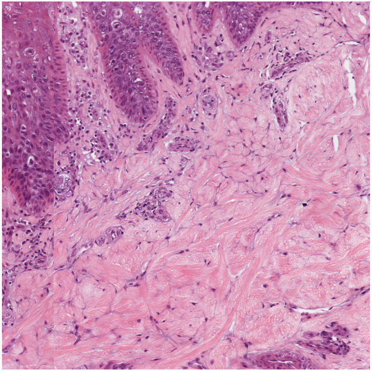


H&E stained tissue sample

Usage

- Hematoxylin and eosin stain (H&E stain) is one of the principal tissue stains used in histology and Mohs sugery.

What is it?

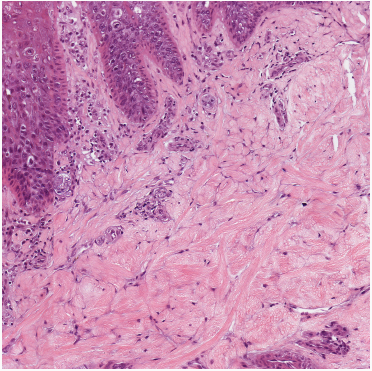


H&E stained tissue sample

Usage

- Hematoxylin and eosin stain (H&E stain) is one of the principal tissue stains used in histology and Mohs sugery.
- The **hematoxylin** stains cell nuclei **blue**, and **eosin** stains the extracellular matrix and cytoplasm **pink**, with other structures taking on different shades, hues, and combinations of these colors.

What is it?

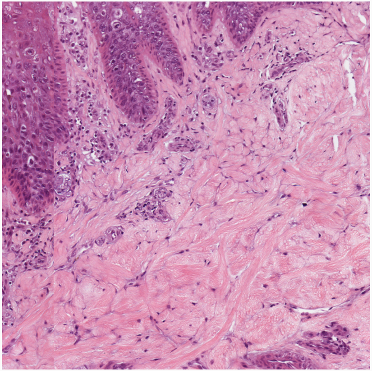


H&E stained tissue sample

It can be a slow process

- Tissues are typically frozen, cut, fixed in alcohol and then stained.

What is it?



H&E stained tissue sample

It can be a slow process

- Tissues are typically frozen, cut, fixed in alcohol and then stained.
- Involves application of hematoxylin mixed with a metallic salt, a rinse in a weak acid solution, followed by bluing in alkaline water. After the application of hematoxylin, the tissue is counterstained with eosin.

Introduction

Hematoxylin and Eosin (H&E)

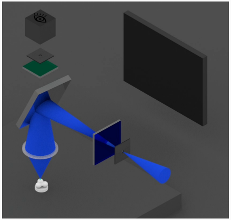
Confocal microscopy

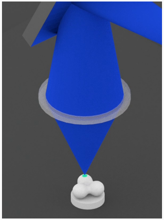
Data-driven transformation

What is it?

CM is an optical imaging technique for increasing optical resolution and contrast of a micrograph by means of using a spatial pinhole to block out-of-focus light in image formation.

It is able to capture multiple two-dimensional images at different depths in a sample (a process known as optical sectioning).



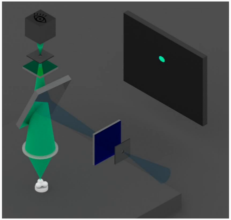


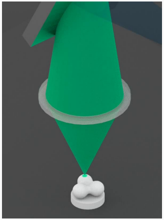
[1]The light beam is focused by a pinhole on a small part of the sample

What is it?

CM is an optical imaging technique for increasing optical resolution and contrast of a micrograph by means of using a spatial pinhole to block out-of-focus light in image formation.

It is able to capture multiple two-dimensional images at different depths in a sample (a process known as optical sectioning).





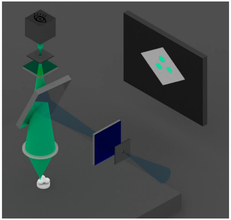
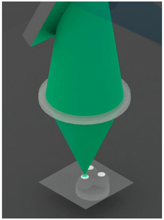
[1]If fluorosphores are present, they emit light

[1]Images extracted from wikimedia.

What is it?

CM is an optical imaging technique for increasing optical resolution and contrast of a micrograph by means of using a spatial pinhole to block out-of-focus light in image formation.

It is able to capture multiple two-dimensional images at different depths in a sample (a process known as optical sectioning).

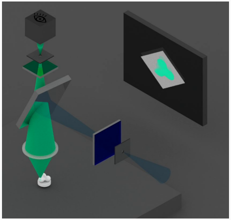
[1]The surface is scanned by moving the sample/light beam

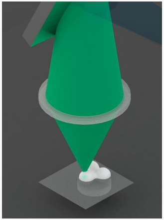
[1]Images extracted from wikimedia.

What is it?

CM is an optical imaging technique for increasing optical resolution and contrast of a micrograph by means of using a spatial pinhole to block out-of-focus light in image formation.

It is able to capture multiple two-dimensional images at different depths in a sample (a process known as optical sectioning).





[1]One can move vertically to obtain images at different heights

[1]Images extracted from wikimedia.

Similarities and dissimilarities with H&E

Similarities

- Usage in histopathology
- Up to cellular level resolution
- Expose similar structures in tissues

Similarities and dissimilarities with H&E

Similarities

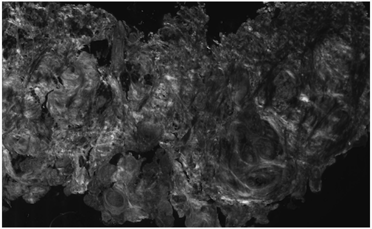
- Usage in histopathology
- Up to cellular level resolution
- Expose similar structures in tissues

Differences

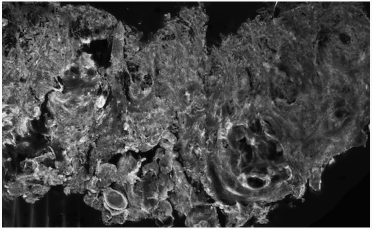
- Specimens can be examined in near real-time without time consuming processing procedures.
- Output largely diers from the standard H&E slides

Modes

In addition to "standard" reflection mode, CM can work in fluorescent mode for specimens stained with fluorochromes.



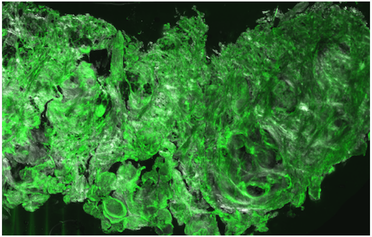
Reflectance mode



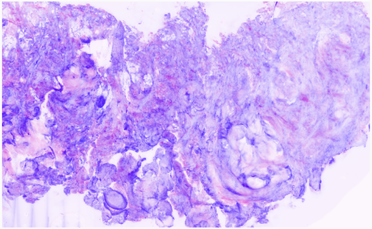
Fluorescence mode

Modes

In addition to “standard” reflection mode, CM can work in fluorescent mode for specimens stained with fluorochromes.



Blend of reflectance and fluorescence mode.



Pseudocolor/falsecolor H&E-like digital stain from reflectance and fluorescence modes

Introduction

Hematoxylin and Eosin (H&E)

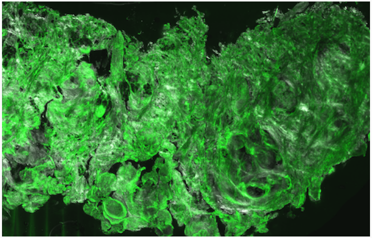
Confocal microscopy

Data-driven transformation

Why?

As confocal micrographs largely differ from the standard H&E slides that pathologists typically use to analyze tissue samples, professionals need to undergo specific training.

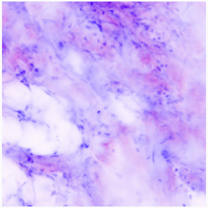
A correctly done CM to H&E mapping should bring the efficiency of CM to untrained pathologists and surgeons.



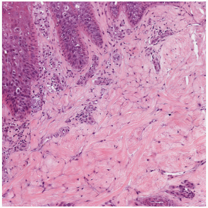
Why?

Standard pseudocolor transformation is a linear transformation that cannot capture the true appearance of H&E stained samples.

Different structures should be mapped to different colours even if the pixel values are the same.

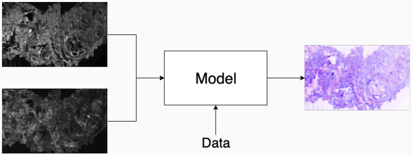


Colors and structures greatly vary from the ones found in H&E slides

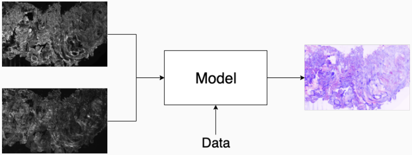


H&E slide

How?



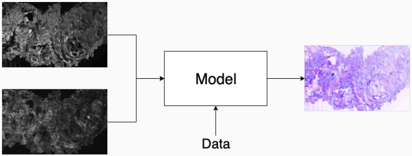
How?



Solution

- Data driven approach to try to find a better transformation

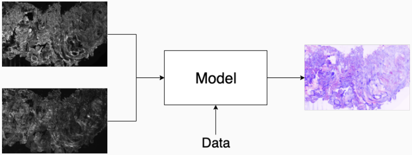
How?



Solution

- Data driven approach to try to find a better transformation
- CNNs are good at *learning* non-linear mappings

How?



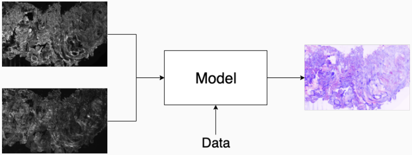
Solution

- Data driven approach to try to find a better transformation
- CNNs are good at *learning* non-linear mappings

But

- Traditional solutions require pairs of CM and H&E slides, which would be very difficult

How?



Solution

- Data driven approach to try to find a better transformation
- CNNs are good at *learning* non-linear mappings
- Use CycleGANs framework

But

- Traditional solutions require pairs of CM and H&E slides, which would be very difficult

Theoric background

Theoric background

Artificial neural networks

Generative models

Convolutional neural networks (CNNs)

Short description

- Convolution is fast way to apply a linear translation-invariant transformation

Convolutional neural networks (CNNs)

Short description

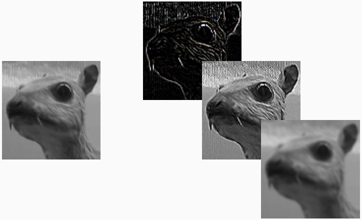
- Convolution is fast way to apply a linear translation-invariant transformation

Convolutional neural networks (CNNs)

Short description

- Convolution is fast way to apply a linear translation-invariant transformation

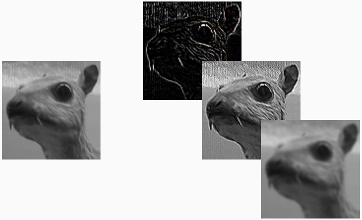
Convolutional neural networks (CNNs)

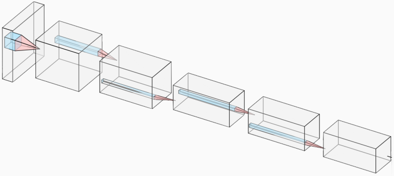


Short description

- Convolution is fast way to apply a linear translation-invariant transformation

Convolutional neural networks (CNNs)





Short description

- Convolution is fast way to apply a linear translation-invariant transformation
- A layer contains multiple filters
- A layer's output is the concatenation of each resulting convolution

Training

Define an objective function and optimize it

$$\theta^* = \underset{\theta}{\operatorname{argmin}}\, \mathcal{J}\left((\mathbf{X}, \mathbf{Y}), f_\theta\right)$$

Neural networks are fully differentiable by construction, so we can use gradient-based methods.

Training

Define an objective function and optimize it

$$\theta^* = \underset{\theta}{\operatorname{argmin}}\, \mathcal{J}\left((\mathbf{X}, \mathbf{Y}), f_\theta\right)$$

Neural networks are fully differentiable by construction, so we can use gradient-based methods.

But
Defining an objective function for our problem is straightforward.

We want to *generate* images that *look like* H&E and are *realistic*.

Theoric background

Artificial neural networks

Generative models

Generative adversarial networks (GANs)

Framework for estimating generative models via an adversarial process, in which two models are trained: a generative model G that captures the data distribution, and a discriminative model D.

Two-player minimax game

$$\min_G \max_D \mathbb{E}_{x \sim p_{data}(x)}\{\log D(x)\} + \mathbb{E}_{z \sim p_z(z)}\{\log(1 - D(G(z)))\}$$

Generative adversarial networks (GANs)

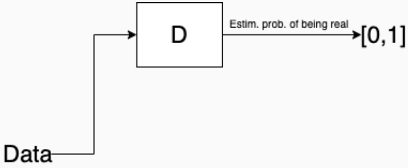
Framework for estimating generative models via an adversarial process, in which two models are trained: a generative model G that captures the data distribution, and a discriminative model D.

Two-player minimax game

$$\min_G \max_D \mathbb{E}_{x \sim p_{data}(x)}\{\log D(x)\} + \mathbb{E}_{z \sim p_z(z)}\{\log(1 - D(G(z)))\}$$

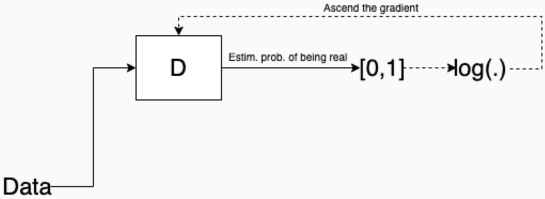
Intuitively
Two models "fight" one against the other (over the discriminator's loss) in such a way that they improve each other, until the generator's samples are indistiguishable from data.

Visual representation



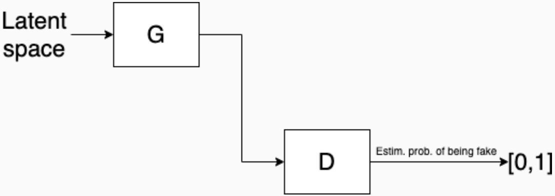
Discriminator predicts the probability of being "real" of a sample drawn from the dataset

Visual representation



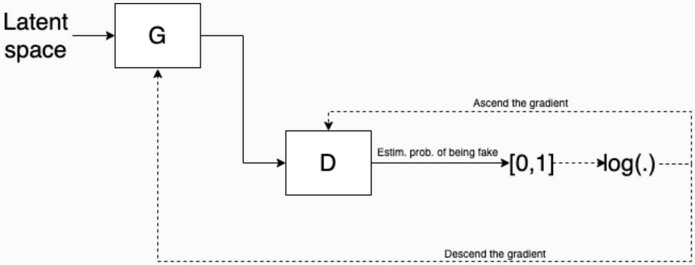
The discriminator's parameters are updated to increase the log-likelihood

Visual representation



Discriminator predicts the probability of being “fake” of a sample drawn from the generator

Visual representation



The discriminator's parameters are updated to increase the log-likelihood.
The generator's parameters are updated to decrese the log-likelihood of the discriminator

Cicle-consistent GANs (CycleGANs)

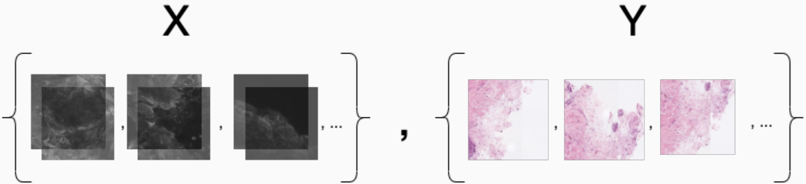
Objective

- Framework for image-to-image translation based on GANs, where mapping is learned with unpaired data

Cicle-consistent GANs (CycleGANs)

Objective

- Framework for image-to-image translation based on GANs, where mapping is learned with unpaired data
- Unpaired data consists of a source set **X** and a target set **Y** with no information provided as to which $\mathbf{x}_i$ matches which $\mathbf{y}_i$



Cicle-consistent GANs (CycleGANs)

Objective

- Framework for image-to-image translation based on GANs, where mapping is learned with unpaired data
- Unpaired data consists of a source set $\mathbf{X}$ and a target set $\mathbf{Y}$ with no information provided as to which $\mathbf{x}_i$ matches which $\mathbf{y}_i$

Why not "simple" GANs?

- Simply using the discriminator loss is not sufficient as it leads to the problem known as mode-collapse where the generator ignores the source

Cicle-consistent GANs (CycleGANs)

Objective

- Framework for image-to-image translation based on GANs, where mapping is learned with unpaired data
- Unpaired data consists of a source set $\mathbf{X}$ and a target set $\mathbf{Y}$ with no information provided as to which $\mathbf{x}_i$ matches which $\mathbf{y}_i$

Why not "simple" GANs?

- Simply using the discriminator loss is not sufficient as it leads to the problem known as mode-collapse where the generator ignores the source
- This problem is addressed by "encouraging" the mapping to be cycle-consistent, i.e.: $G_{Y \to X}(G_{X \to Y}(\mathbf{x})) \approx \mathbf{x}$.

Cycle-consistent GANs (CycleGANs)

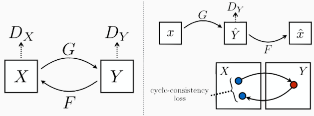
Two pairs of generator-discriminator are trained

- $G_{X \to Y}$ maps from domain X to domain Y. D_Y discriminates in domain Y.
- $G_{Y \to X}$ maps from domain Y to domain X. D_X discriminates in domain X.

Cycle-consistent GANs (CycleGANs)

Two pairs of generator-discriminator are trained

- $G_{X \to Y}$ maps from domain X to domain Y. D_Y discriminates in domain Y.
- $G_{Y \to X}$ maps from domain Y to domain X. D_X discriminates in domain X.



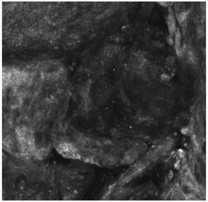
Modified loss for $G_{X \to Y}$ (analogous for $G_{Y \to X}$):

$$\mathcal{L}_{G_{X \to Y}} = \log(1 - D_Y(G_{X \to Y}(\mathbf{x}))) + \lambda \left\| \mathbf{x} - G_{Y \to X}(G_{X \to Y}(\mathbf{x})) \right\|_1$$

Methodology

Speckle noise

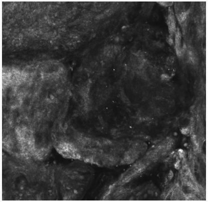
The reflectance CM image is corrupted by speckle:
Scattered signals add constructively and destructively depending on the relative phases of each scattered waveform. Speckle results from these patterns of constructive and destructive interference shown as bright and dark dots in the image.



Example of speckled RCM image

Speckle noise

The reflectance CM image is corrupted by speckle:
Scattered signals add constructively and destructively depending on the relative phases of each scattered waveform. Speckle results from these patterns of constructive and destructive interference shown as bright and dark dots in the image.



Example of speckled RCM image

Speckle model

$$Y = X \odot F$$

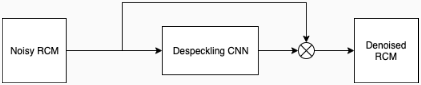
$$F \sim \Gamma(k = L, \theta = \frac{1}{L})$$

Despeckling network

A CNN is used to filter speckle noise prior to the stain transformation.

Despeckling network

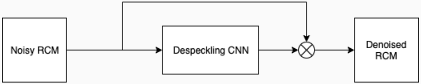
A CNN is used to filter speckle noise prior to the stain transformation.



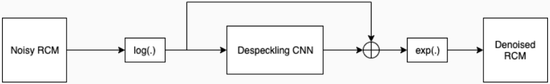
Model with multiplicative skip-connection

Despeckling network

A CNN is used to filter speckle noise prior to the stain transformation.



Model with multiplicative skip-connection



Model with additive skip-connection

Training

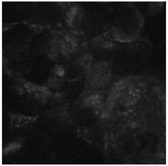
Loss
The models are trained using the MSE loss:

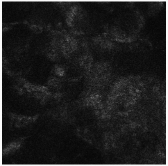
$$\mathcal{L} = \left\| \mathbf{x}_{clean} - f_\theta(\mathbf{x}_{noisy}) \right\|_2$$

Loss
The models are trained using the MSE loss:

$$\mathcal{L} = \left\|\mathbf{x}_{clean} - f_\theta(\mathbf{x}_{noisy})\right\|_2$$

Data
Due to the impossibility to obtain the noisy and clean versions of the the RCM images, artificially contaminated FCM images are used to obtain the $(\mathbf{x}_{clean}, \mathbf{x}_{noisy})$ pairs.





Methodology

Despeckling

Stain

Inference technique

Quality measure

Generators' architecture

Both families follow an encoder-decoder structure, i.e.: a series of convolution layers with down-sampling (encoder) followed by the same number of layers with up-sampling[2] (decoder), presumably the encoder maps the input into a latent representation where semantic transformations can be more easily defined and then the decoder "brings" it back to the image space.

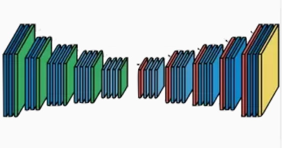
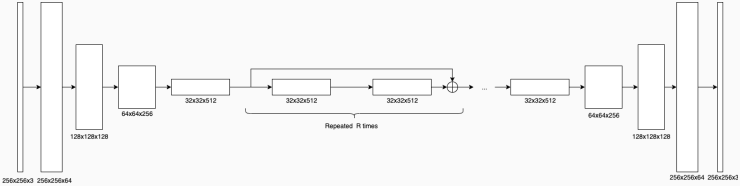


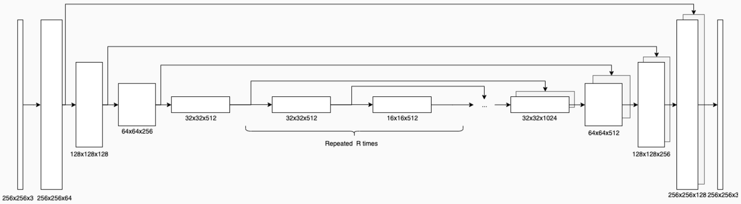
image credits

Residual network



In a residual block, intead of trying to learn a transformation $\mathcal{T}(\mathbf{x})$, the residual $\mathcal{F}(\mathbf{x})$ is learnt so that $\mathcal{T}(\mathbf{x}) = \mathcal{F}(\mathbf{x}) + \mathbf{x}$ the motivation behind this is to avoid the problem know as the degradation problem where deeper networks perform worse than shallower counterparts, when in theory they should at least perfom equally well.

UNet-like network



As a means to obtain low-level information (location, texture, ...) from the encoder, the output from the corresponding encoder layer is concatenated to the output of the previous decoder layer.

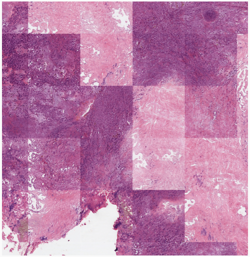
Why?

Need
Whole slide images are too large to fit directly on a GPU, therefore, the inference has to be tile-by-tile to obtain the stain transformed result.

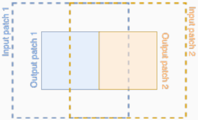
Why?

Need
Whole slide images are too large to fit directly on a GPU, therefore, the inference has to be tile-by-tile to obtain the stain transformed result.

Problem
This introduces artifacts between adjacent tiles in the output due to instance normalization relying on tile statistics.



Method I



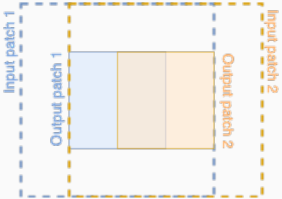
The tiles predictions share context but the outputs do not overlap.

Method I - Example

Independent tile-by-tile inference

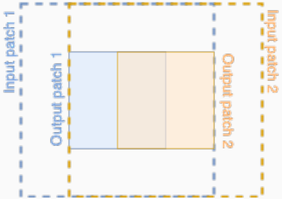
Tile-by-tile inference using method I

Method II

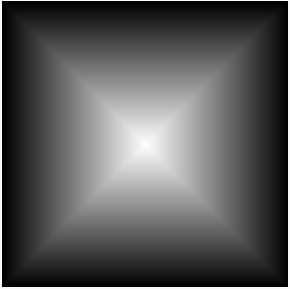


Shifting the window by quarter of its size creates an overlap between the tiles output.

Method II



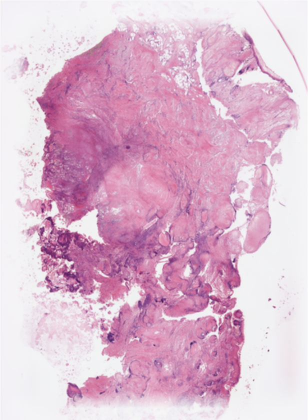
Shifting the window by quarter of its size creates an overlap between the tiles output.



Output is weighted by a “pyramidal” window. Therefore, each pixel value in the final transformation is a weighted sum of 4 predictions.

Method II - Example

Independent tile-by-tile inference

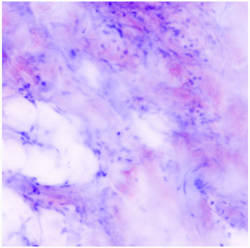


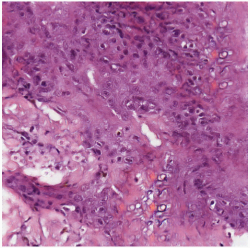
Tile-by-tile inference using method II

Methodology

- Despeckling
- Stain
- Inference technique
- Quality measure

Two metrics are used to try to measure if the generated samples contain structures that are not present in the source image (popularly known as hallucinations). The metrics are based on comparing the generated image with the equivalent linear staining image (both in gray-scale).





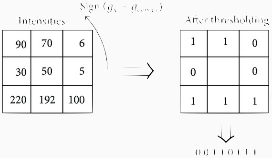
An example of an hallucination

SSIM

To validate the structure integrity of the transformed wholeslides, the SSIM metric is used. SSIM is a perception-based model that considers image degradation as perceived change in structural information.

LBP histogram distance

Use the Local Binary Patterns (LBP) descriptor to measure the distance in a texture sense.



The distance between the result and source is measured using the chi-squared distance between the normalized LBP histograms

Results

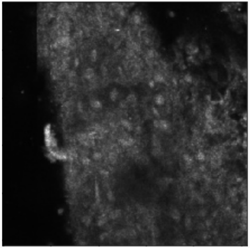
Quantitative results

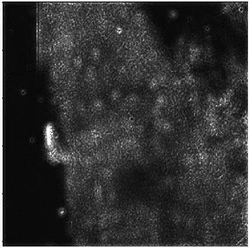
Model	M	K	N	mean $SSIM$
Multiply ($L = 1$)	3	32	5	0.877
Multiply ($L = 1$)	5	64	5	0.728
Multiply ($L = 5$)	3	32	5	0.960
Multiply ($L = 5$)	5	64	5	0.947
Log-Add ($L = 1$)	3	32	5	0.960
Log-Add ($L = 1$)	5	64	5	0.806
Log-Add ($L = 5$)	3	32	5	0.968
Log-Add ($L = 5$)	5	64	5	0.965

Models comparison with different number of layers M, number of filters K and filter size N. The validation set mean $SSIM_{input}$ is 0.414 for $L = 1$ and 0.723 for $L = 5$.

Qualitative results

Good SSIM is obtained when comparing a noise-free FCM and its denoised version; but when applying the DespeckleNN to RCM images, the result looks noisier.





Results

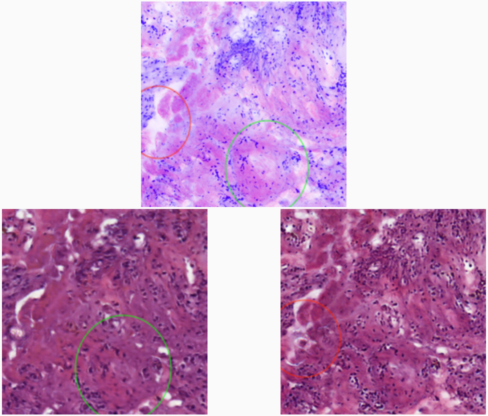
- Despeckling
- Stain
- Inference technique

Quantitative results

Metric	Residual	UNet-like
LBP histogram chi-squared distance	0.0332	0.0183
SSIM	0.5002	0.5570

Metrics mean values on validation set for the two tested models.

Qualitative results



(N) Linear stain (SW) Residual model stain (SE) Unet model stain

The motivation behind using the UNet-like model is to mantain the structure, in practice this generally holds but still some structures are "hallucinated" and nuclei present in the source image are eliminated (evaluation supported by dermatology expert).

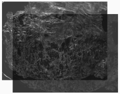
Results

Despeckling

Stain

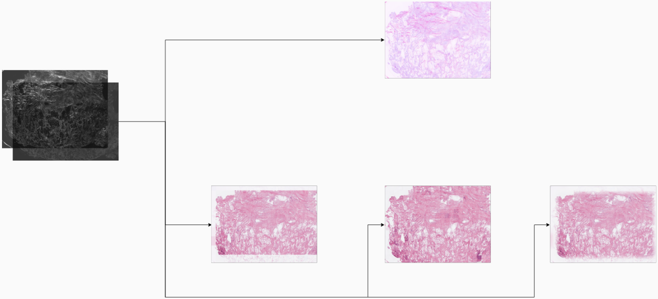
Inference technique

Experiments



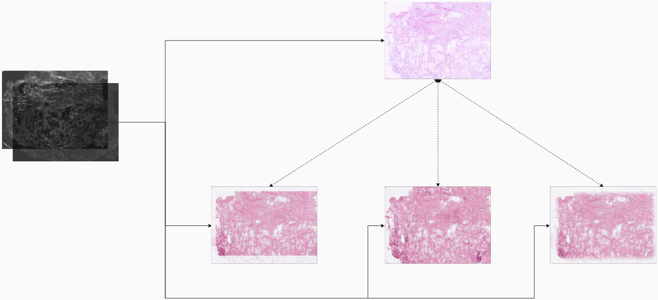
7 whole slides are used to compute the metrics

Experiments



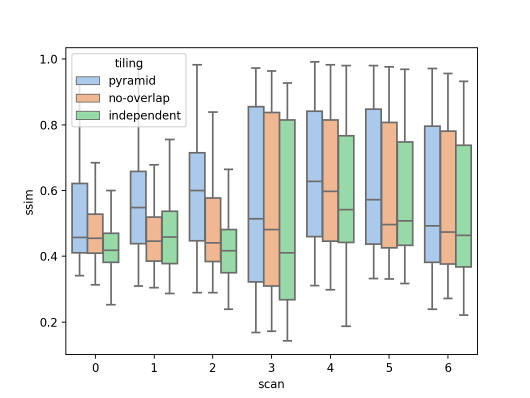
The corresponding linear stain and StainNN transformations (with different inference methods) are applied

Experiments

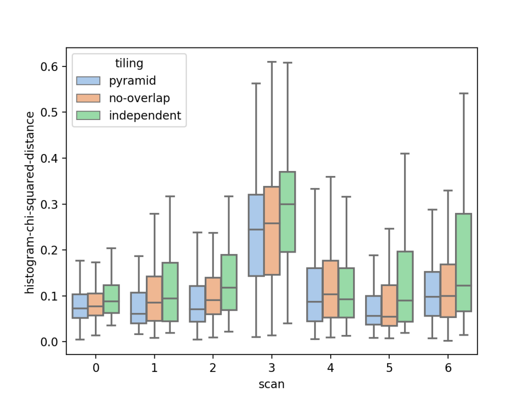


The metrics between the StainNN and linear versions are computed in 500x500 patches

SSIM



The pyramid tiling (method II) distributions in particular has a higher variability compared to the other ones, but the general ditribution is always higher.



The distributions for the pyramid (method II) tend to be in a lower range and have less variability.

Conclusions and future development

Conclusions

- The use of the CycleGANs framework for digitally staining CM slides has been studied, as well as fully-convolutional models for speckle denoising.
- Different inference techniques for whole slides are developed and compared.
- A way of measuring StainNN hallucinations is studied.
- UNet-like architecture is superior to the residual one based on both the LBP and SSIM metrics, but still the structures are not always preserved.

Future development

1. Find a way to train despeckling model that generalizes to RCM.
2. A new stain model should be trained with a loss that further encourages structure integrity:
    - Showing both the input and output to the discriminator.
    - Modify the CycleGAN loss.