Data augmentation based on spatial deformations for histopathology: An evaluation in the context of glomeruli segmentation

Florian Allender, Rémi Allègre, Cédric Wemmert, Jean-Michel Dischler

Data augmentation based on spatial deformations for histopathology: An evaluation in the context of glomeruli segmentation

Florian Allender, Rémi Allègre[1], Cédric Wemmert, Jean-Michel Dischler

ICube, Université de Strasbourg
CNRS, France

Abstract

Background and Objective: The effective application of deep learning to digital histopathology is hampered by the shortage of high-quality annotated images. In this paper we focus on the supervised segmentation of glomerular structures in patches of whole slide images of renal histopathological slides. Considering a U-Net model employed for segmentation, our goal is to evaluate the impact of augmenting training data with *random spatial deformations*.
Methods: We define a random spatial deformation as a vector field resulting from the realization of a stochastic process with uniform or Gaussian distribution. We perform a comparative study of state-of-the-art models that produce this type of deformation, including Random Displacement Fields and deformations based on control points. We introduce a method based on the detection of cell nuclei centers for the automatic positioning of scattered control points that simplifies the control of deformations w.r.t. previous approaches.
Results: We show that augmenting training data with spatially deformed images yields an improvement of up to 0.23 in average Dice score, with respect to training with no augmentation. We demonstrate that deformations with relatively strong distortions yield the best performance increase, while previous work only report the use of deformations with low distortions. The selected deformation models yield similar performance increase, provided that their parameters are properly adjusted. We provide bounds on the optimal parameter values, obtained through parameter sampling, which is achieved in a lower computational complexity with our single-parameter method. The paper is accompanied by a framework for evaluating the impact of random spatial deformations on the performance of any U-Net segmentation model.
Conclusions: To our knowledge, this study is the first to evaluate the impact of random spatial deformations on the segmentation of histopathological images. Our study and framework provide tools to help practitioners and researchers to make a better usage of random spatial deformations when training deep models for segmentation.

Keywords: Histopathological images, glomeruli segmentation, data augmentation, random spatial deformations

1. Introduction

In the field of digital pathology, deep learning techniques for whole slide image (WSI) analysis have attracted considerable attention in recent years [1, 2, 3, 4, 5, 6]. Deep models based on Convolutional Neural Networks (CNN) have the ability to automatically identify microscopic structures, which brings significant support to diagnosis. However, these models are based on learning patterns in images and require a large amount of annotated data to be generalizable. This requirement is a challenge, especially for application of deep models to diseases where few samples are available. This is the case for renal pathology on which we focus on [7]. *Data augmentation* [8] tackles this limitation by expanding the variety of training datasets in an artificial manner, which may involve various image manipulations such as spatial and colorimetric transformations, noise injection, image mixing, or synthetic image generation using generative models. While data augmentation has proven to improve the performance of deep models for various applications, augmentations are domain-dependent and require data-specific design [8]. Automatically finding optimal augmentation policies has been investigated recently for classification, detection or segmentation tasks, using meta-learning schemes or adversarial training (see e.g. *AutoAugment* [9, 10]). These methods still require to select appropriate models for transformations.

In this paper we consider the supervised segmentation of glomerular structures in WSI patches of renal histopathological slides. These complex structures are responsible for blood filtration and production of urine. Renal histopathology is crucial in the study of kidney diseases, and especially transplant rejection, which occurs with an incidence of 7.9% in the first year [11]. Glomeruli are ball-shaped structures that incorporate multiple substructures: membranes, capillaries, mesangial and endothelial cells, podocytes. The appearance of glomeruli can thus exhibit great variability in WSIs (Fig. 1). Besides sectioning, this variability also stems from stainings, patients, procedures in laboratories, and/or microscope and imaging devices, making the robust detection and segmentation of glomeruli challenging.

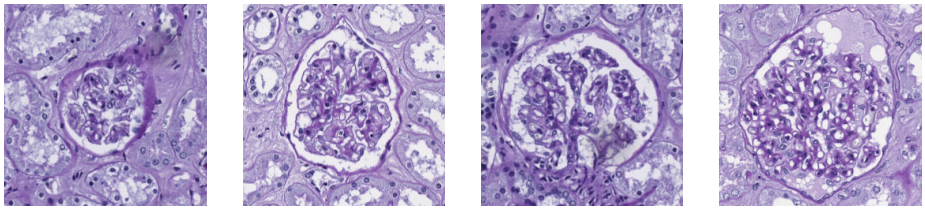

Figure 1: Examples of glomeruli in patches of WSIs, with variable shapes (Periodic acid-Schiff staining).

While there has been recent advances regarding invariance to staining variability [12, 13, 14], we focus here on the variability in shapes, i.e. *geometric* variability, which has not yet been addressed thoroughly in the Medical Imaging literature. We rely on a standard U-Net model [15] for segmentation and evaluate the impact of augmenting training data with *random spatial deformations*. We define a random spatial

[1]Corresponding author.

deformation as a vector field resulting from the filtering of a *white noise*, i.e. the realization of a stochastic process with uniform or Gaussian distribution. Random spatial deformations can be implemented in many different ways, and a vast variety of models has been proposed in the literature, differing on how the vector field is generated.

Considering a selection of state-of-the-art models for random spatial deformations, we address the following questions: (1) Among the considered deformation models, which are best suitable for data augmentation, with the objective of segmenting glomeruli in WSI patches using U-Net? (2) Which range of parameter values are adequate? (3) How can we improve deformation models to achieve better segmentation results? In this context, our contributions are the following:

- We propose a comparative study of a selection of state-of-the-art random spatial deformation models, including Random Displacement Fields and deformations based on control points.
- We provide appropriate range of parameter values for each of the considered random spatial deformation models.
- We propose an original method based on the detection of cell nuclei centers for the automatic positioning of scattered control points, that simplifies the control of random deformations and reduces the computational complexity of finding best parameter values.
- We provide a complete framework for identifying which deformation models and range of parameter values are the best choices for a given application.

In the following section, we review related work (Section 2). Next, we provide a description of the spatial deformation models considered in this paper (Section 3). Our evaluation protocol is then developed (Section 4), before we present and discuss our experimental results (Section 5 and 6). Complementary results discussed in the paper are provided as a Supplemental material.

2. Related work

Glomeruli segmentation. We consider the *semantic* segmentation of glomeruli from tissue (i.e. background) in patches extracted from WSIs [16, 13, 17, 18, 19, 20], which consists in solving a two-class problem at pixel level for each input patch. We suppose that each patch contains either a single glomerulus surrounded by tissue, or partial glomeruli, or tissue only. The U-Net model [15] is the state-of-the-art CNN architecture for segmentation in Biomedical Imaging [2], especially for glomeruli segmentation in WSI patches [16, 18, 19]. Some variants like U-Net++ [21, 22] or SE U-Net [23] result in slight improvements, as reported in [13]. Other architectures like Seg-Net [17] or U-Net with a ResNet34 backbone [20], pre-trained on the ImageNet dataset, have been proposed. Both works report improved segmentation results w.r.t. U-Net, Salvi et al. [20] adding a post-processing stage based on cell nuclei and lumen detection to refine the segmentation results. These works only consider flipping, rotation and scaling as geometric transformations. In this paper, we choose to rely on the basic, well established and still widely used standard U-Net model, in order to study and understand the benefits of random spatial deformations as an augmentation. Note also that we do not use a pre-trained model.

Augmentations	[16]	[18]	[19]	[20]	[24]	[26]	[27]	[29]	[38]	[31]	[36]
Rotations	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Mirroring	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Shift/crops		✓	✓						✓		✓
Scaling		✓	✓	✓	✓	✓			✓		
Shearing									✓		✓
Elastic deformations	✓	✓	✓		✓	✓	✓				✓
Blur	✓	✓	✓		✓	✓					✓
Additive noise		✓			✓	✓					
Cutout			✓								
Color shifting	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓

Table 1: Augmentations used in a selection of papers from the digital histopathology field.

Glomeruli segmentation in patches of WSIs is different from direct segmentation of WSIs, as achieved e.g. in [24, 25, 26, 27], where modified U-Net architectures are used to account for high-resolution input images. Also note that segmentation of glomeruli in WSI patches is different from glomeruli detection, which is achieved with CNN architectures conceived for object detection like Faster RCNN or YOLOv3 in the most recent works [28, 29, 13, 30, 31].

Data augmentation for histopathology. Data augmentation has been introduced in deep learning to prevent overfitting issues that arise from small and imbalanced datasets. A complete review is beyond the scope of this paper, the reader can refer to [8] for a survey. There exist two main approaches: 1) The *transformation* approach consists in applying individual (or compositions of) transformations to training datasets, or in some cases, to test datasets; 2) The *generative* approach consists in generating new data by means of a trained generative model, e.g. a Generative Adversarial Network (GAN). We focus here on augmentations of the *first* category, applied to training datasets. Geometric transformations like affine transformations and spatial deformations have become usual for almost all applications, besides color shifting, addition of noise, or image mixing or blending, that are part of libraries like *Augmentor* [32]. In the field of digital histopathology, common augmentations include geometric transformations and specific transformations related to staining variations [33, 34, 35, 36]. Those types of augmentations have been especially used for glomeruli detection and segmentation [16, 26, 24, 31, 18, 19, 36, 20]. See Table 2 for a summary of the augmentation methods used in these papers. One can note that rotations and mirroring are systematically used, and that staining transformations and elastic deformations are among the most commonly used augmentation methods. Elastic deformations refer to random spatial deformations based on control points on a regular grid, as typically proposed by Ronneberger et al. [15]. The benefits of staining transformations [37, 12, 14, 36] and affine transformations [36] have been quantified. The impact of using different augmentations for the segmentation of WSIs of rat organs is evaluated in [27], including affine transformations and elastic deformations, but only with a binary approach (with/without). In this paper, we propose to fill the gap regarding *random spatial deformations* and the adjustment of the parameter values for the case of glomeruli segmentation in WSI patches.

Augmentations	Bel et al, 2018	Merveille et al, 2021	Davis et al, 2021	Salvi et al, 2021	Altini et al, 2020
Rotations	✓	✓	✓	✓	✓
Mirroring	✓	✓	✓	✓	✓
Shift/crops		✓	✓		
Scaling		✓	✓	✓	✓
Elastic deformations	✓	✓	✓		✓
Blur	✓	✓	✓		✓
Additive noise		✓			✓
Cutout			✓		
Color shifting	✓	✓	✓		✓

Table 2: Augmentations used in a selection of papers from the digital histopathology field.

Automatic data augmentation. Some recent works attempt to find optimal data augmentation policies automatically, through meta-learning schemes [9, 39, 40], adversarial training [10] or a combination for better performance [41]. [36] apply the RandAugment methodology [40] in the context of a WSI patch classification task for breast cancer detection. In this paper, we evaluate the impact of data augmentation by considering a single deformation model at a time, i.e. we do not mix deformed images obtained with different models. Therefore we use a simple meta-learning scheme that performs parameter sampling inspired by RandAugment [40], in order to find the parameter values that achieve the best segmentation performance.

Spatial deformation models. The most basic spatial deformation model consists of *linear* transformations, that include affine and projective transformations. While linear transformations are ubiquitous in data augmentation policies [8], often only consisting of flips and rotations, they enable limited variations in image content.

Non-linear models have been developed to increase the variety and locality of spatial deformations, some of them describing the deformations in terms of a physical process, and known in the literature as *elastic deformations*, *non-rigid deformations*, or *domain warping* (see [42] for a survey). In the Medical Imaging community, non-linear deformations have been widely employed for image registration [43]. They attracted attention more recently for the purpose of improving the training of deep learning models. Chlap et al. [44] surveyed deformation models in the context of data augmentation for CT and MRI images. In the field of digital histopathology, correspondences cannot be established between images, which makes image registration or statistical shape models inadequate to produce variations of training images. The alternative is to generate *random* deformations, i.e. vector fields resulting from the realization of a stochastic process with controlled statistics.

Previous work in the field of digital histopathology [16, 26, 24, 31, 18, 19] report the use of two random spatial deformation models involved in data augmentation for the segmentation of WSIs patches: Random Displacement Fields as proposed in [45], and grid-based deformations, as suggested in the original U-Net paper [15]. Two choices have to be done: 1) the deformation model, and 2) the parameter values. In this paper, we review these models (Section 3), and we evaluate their impact when used for data augmentation in the context for glomeruli segmentation (Section 4). We note that the influence of the parameter values chosen in previous work is not evaluated, and that these values result in deformed images that are *visually close* to the available training ones. The search for optimal parameter values for the data augmentation task has not yet been investigated. In the present paper, we explore the appropriate values for each

of the considered deformation models.

3. Random spatial deformations

3.1. Background and notations

A 2D spatial deformation is a continuous function $\mathcal{D} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ that maps input coordinates $\mathbf{x} = [x, y]$ to transformed coordinates $\left[\mathcal{D}_x(\mathbf{x}), \mathcal{D}_y(\mathbf{y})\right]$, possibly not bijective. Applying deformations to images involves a *sampling* (i.e. *continuous-to-discrete*) problem solving. Transformed coordinates may not correspond to pixel locations, and may not cover the whole spatial image domain. Therefore, the transformation $\mathcal{D}$ is usually defined in an inverse fashion (a.k.a. *reverse warping*), and computed from the pixel grid of the *output* image to the spatial domain of the *input* image. In what follows, we will consider square images of size N^2. The pixel grid of the output (deformed) image I will be denoted as $\mathbf{P} = \{\mathbf{p}_i\}_{i=1}^{N^2}$, where $\mathbf{p}_i = [x_i, y_i]$ are the pixel coordinates. The pixel grid of the associated input image I' will be denoted as $\mathbf{P'} = \{\mathbf{p}'_i\}_{i=1}^{N^2}$. In our context, we can write:

$$\mathcal{D}(\mathbf{P}) = \mathbf{P} + \overrightarrow{\mathbf{D}} \tag{1}$$

where $\overrightarrow{\mathbf{D}} = \{\vec{\mathbf{d}}_i\}_{i=1}^{N^2}$ is a set of displacement vectors, or *flow field*. Following this approach, pixel values in I are obtained by interpolating pixel values in I', considering the pixel locations that are the closest from the deformed coordinates $\mathcal{D}(\mathbf{P})$ in the spatial domain of I', i.e.:

$$I = \mathcal{S}(\mathcal{D}(\mathbf{P}), I') = \mathcal{S}(\mathbf{P} + \overrightarrow{\mathbf{D}}, I') \tag{2}$$

where $\mathcal{S}$ is a continuous interpolation function. Common choices for $\mathcal{S}$ are bilinear interpolation, and higher-order interpolation such as bicubic or spline interpolation. We consider that this choice is not critical for our study, and rely either on bilinear, or bicubic interpolation.

Note that some locations in $\mathcal{D}(\mathbf{P})$ may fall outside the bounds of the input pixel grid. We solve this issue by extending the input image with symmetric padding along the edges of the input image.

3.2. Characterization of random deformations

Definition. A random spatial deformation model is defined by a specific method to generate a flow field $\overrightarrow{\mathbf{D}}$ from a set of random values, so that $\overrightarrow{\mathbf{D}}$ is a multivariate Gaussian random field where the displacement vectors on sampling (i.e. pixel) locations are spatially correlated. Those spatial correlations may arise from filtering (e.g. for Random Displacement Fields), or from the distance to control points and associated interpolation function (e.g. for grid-based deformations).

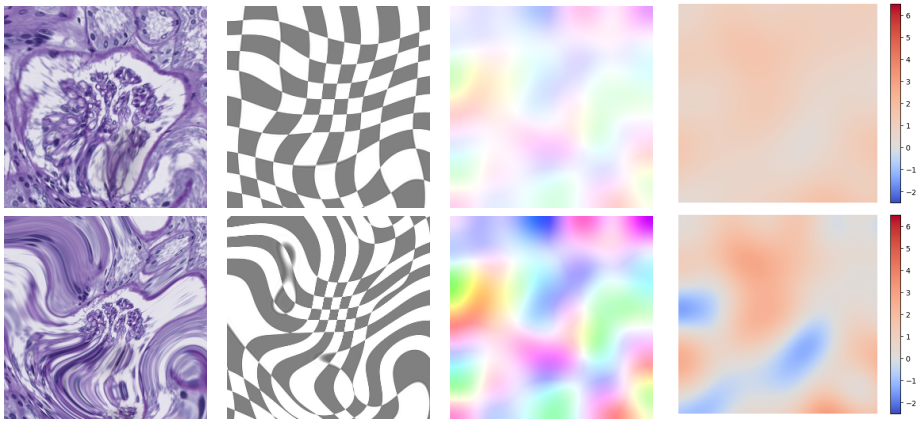

Figure 2: Two examples of deformations applied to a WSI patch containing a single glomerulus, with plots of the corresponding values of the determinant of the Jacobian matrix J. The first row shows a deformation with moderate distortions, whereas the second row shows a deformation with foldover, which is reflected by negative values of J. (a): Deformed image; (b): Deformation applied to a checkerboard; (c): HSV color-coded flow field; (d): Plot of J.

Geometric properties. Random deformations may preserve the shapes and textures contained in an image only at some degree, making the result more or less "realistic", or "distorted". While previous work consider augmenting data with visually realistic deformed images, our goal is to compare the deformation models and determine the optimal amount of distortion for the training of the segmentation model. For this purpose, we propose to characterize random deformations from a geometric point of view.

The ability of random deformations to preserve the shapes and textures contained in an image is directly related to the preservation of *distances* and *angles*. One special case of distortion is *foldover*, that occurs when the orientation between pairs of pixel locations is flipped. The occurrence of foldover can be seen as an upper limit of distortion, since the induced topological inconsistencies may cause profound changes in the local appearance of the deformed images.

The amount of local distortion induced by a flow field can be characterized by the Jacobian matrix of the deformation function $\mathcal{D}$ at every point $\mathbf{p}_i \in \mathbf{P}$, denoted by $\nabla\mathcal{D}(\mathbf{p}_i)$. The determinant of the Jacobian matrix, i.e. $\mathrm{J}(\mathbf{p}_i) = \det(\nabla\mathcal{D}(\mathbf{p}_i))$, encodes the local scale factor by which the distances are shrunk ($|\mathrm{J}(\mathbf{p})| \in [0, 1)$) or expanded ($|\mathrm{J}(\mathbf{p})| > 1$), and its sign reveals whether the deformation produces foldover ($\mathrm{J}(\mathbf{p}) \leq 0$) or not ($\mathrm{J}(\mathbf{p}) > 0$). The distances are preserved when $|\mathrm{J}(\mathbf{p})| = 1$. The less the displacement vectors at neighbor pixel locations are correlated, the more foldover is thus likely to occur locally. Fig 2 shows plots of J for two examples of deformations obtained with the Random Displacement Fields model. The case of the second line exhibits foldover: J is negative at some locations of the flow field, representing about 20% of the pixel locations.

Another feature of flow fields involved in shape and texture preservation is the local scaling of *angles*. A deformation that preserves angles is said to be *conformal*,

and its Jacobian matrix is a rotation matrix multiplied by a scalar. We evaluate the closeness of a deformation to a conformal deformation by computing the dot product of the normalized components of the Jacobian matrix, i.e. $S(\mathbf{p}) = \frac{\mathcal{D}_x(\mathbf{p})}{\|\mathcal{D}_x(\mathbf{p})\|} \cdot \frac{\mathcal{D}_y(\mathbf{p})}{\|\mathcal{D}_y(\mathbf{p})\|}$. The value of $S(\mathbf{p})$ ranges between -1 (maximum angle expansion) and 1 (maximum angle shrinking). The closest to 0, the highest the angle preservation.

For the selected deformation models and varying parameter values, we examine the distribution of the previous metrics, calculated on 1k randomly generated flow fields in the following form:

- Foldover rate, computed as: $|\{\mathbf{p}_i \in \mathbf{P} \text{ s.t. } J(\mathbf{p}_i) \leq 0\}_{i=1}^{N^2}|/N^2$;
- Preservation of distance, computed as the histogram of $\{J(\mathbf{p}_i), \mathbf{p}_i \in \mathbf{P}\}$;
- Preservation of angles, computed as the histogram of $\{S(\mathbf{p}_i), \mathbf{p}_i \in \mathbf{P}\}$.

Fig. 3 shows some plots of these distributions for the Random Displacement Fields model. Additional plots are available in the Supplemental material.

3.3. Deformation models

In this section, we review Random Displacement Fields and deformations based on control points for generating random deformations. We also introduce a novel approach based on unstructured control points located at detected cell nuclei centers.

3.3.1. Random Displacement Fields

Random Displacement Fields (RDF) have been proposed by Simard et al. [45] to simulate smooth deformations. First, a displacement field $\overrightarrow{\mathbf{V}} = \{\vec{\mathbf{v}}_i\}_{i=1}^{N^2}$ is generated, where $\mathbf{v}_i = [v_x, v_y]$ is such that $v_x = \text{rand}(-1, +1)$ and $v_y = \text{rand}(-1, +1)$. The function $\text{rand}(-1, +1)$ returns uniformly distributed pseudo-random numbers between -1 and +1. Then, $\overrightarrow{\mathbf{V}}$ is convolved with a Gaussian filter $\mathcal{G}_\sigma$ of standard deviation σ, where σ acts as an "elasticity" coefficient (small values result in a completely random field, while large values tend to produce rigid deformations). The RDF is finally multiplied by a scaling factor α that controls the amplitude of the deformation. The flow field is therefore obtained as:

$$\overrightarrow{\mathbf{D}} = \alpha\, \mathcal{G}_\sigma(\overrightarrow{\mathbf{V}}) \tag{3}$$

The only work related to glomeruli segmentation that mentions the usage of RDF for data augmentation is the work by Merveille et al. [18], where $\sigma = 10$ and $\alpha = 100$, for patches of size $N = 508$. This choice is not discussed by the authors. The sampling resolution w.r.t. tissue size is the same as in our experiments (0.506 µm/pixel), making it possible to deform our patches of size $N = 256$ with the same impact. These values produce results that are visually close to the original images, as shown in Fig. 4.

The curve plot of Fig. 3(a), first row, illustrates the emergence of foldover for the cases $\sigma = 10$, that occurs starting from $\alpha \approx 150$ (see the previous Section for interpretation). The foldover rate remains low (less than 1%) for values of $150 < \alpha < 200$, and only affects 10% of the deformation fields. Distances remain well preserved, while the scale factor of angles approximately ranges between -0.5 and 0.5. These elements question about the benefit of larger values of α w.r.t. σ, that may contribute favorably

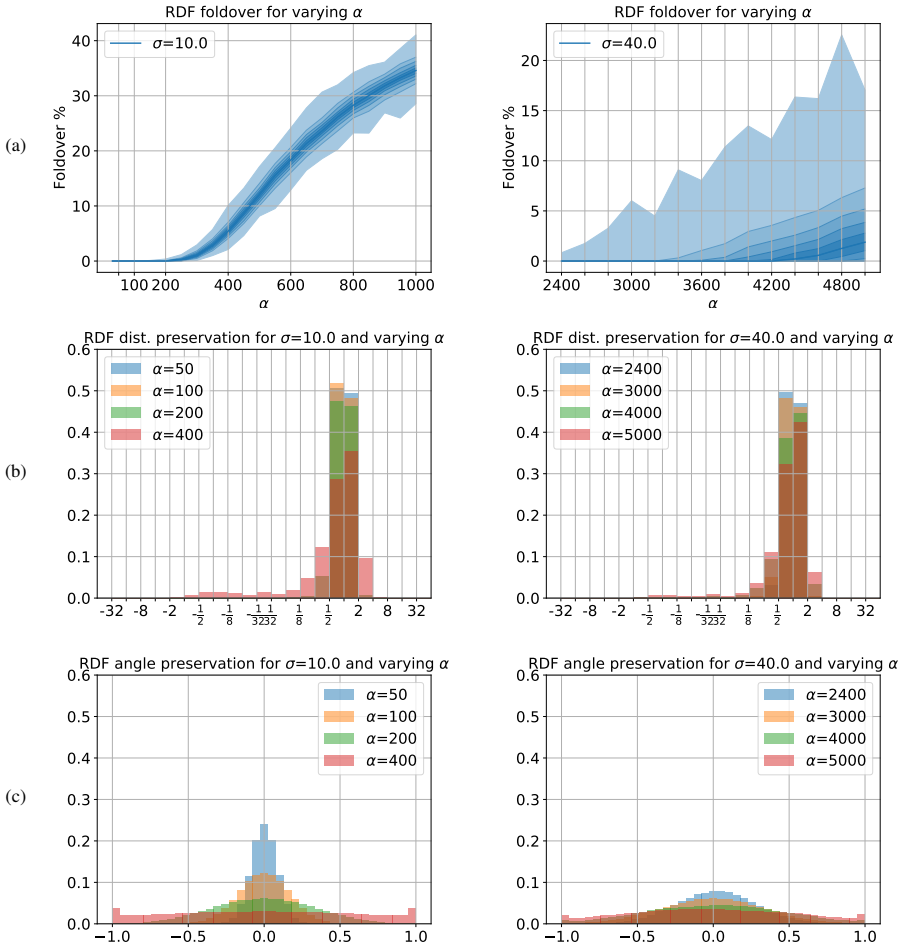

Figure 3: Characterization of the flow fields generated with Random Displacement Fields, based on 1k randomly generated flow fields, for $\sigma = 10$ (first column) and $\sigma = 40$ (second column) and increasing values of α. (a): Foldover rate (the central curve represents the median of the distribution, while surrounding curves correspond to percentiles); (b): Preservation of distances (symmetric $\log_2$ scale is used for bin edges); (c): Preservation of angles. Plots for random flow fields generated with the other models considered in this paper are provided in the Supplemental material.

to data augmentation, and thus are to be considered for a study of the optimal parameter values. Fig. 4 shows deformation results for $\sigma = 10$ and $\sigma = 40$, with varying α. This figure, besides Fig. 3 shows that deformations with low distortion can be produced with larger values of σ, and appropriate values of α, such as $\alpha = 2400$ for $\sigma = 40$. This can be explained by the band-limiting effect of the Gaussian filter, that increases with σ, and requires larger values of α to produce distortions.

Controlling the amount of distortion is a difficult task with this model, since there is no intuitive idea of the amplitude of the deformations induced by a value of α for

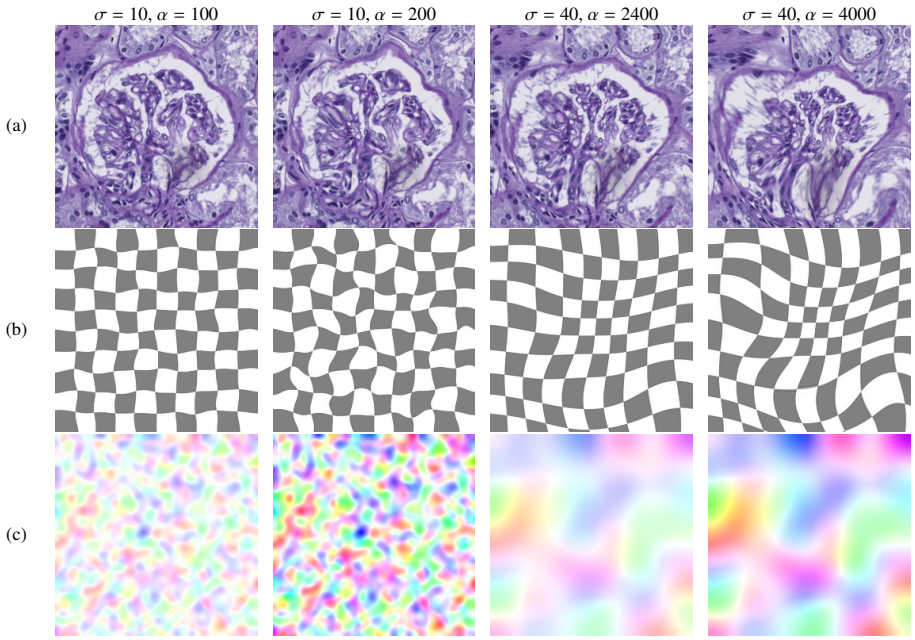

Figure 4: Examples of RDF deformations applied to a WSI patch containing a single glomerulus, for varying values of σ and α. Note that the directions of displacement vectors are identical in the four examples. (a): Deformed patch; (b): Deformation applied to a checkerboard; (c): HSV color-coded flow field.

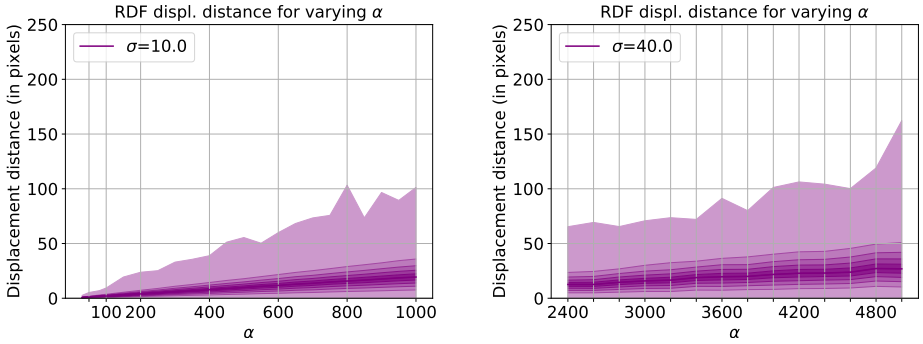

Figure 5: Distribution of displacement distances for Random Displacement Fields, for varying parameter values. For each set of parameter values, the displacement distances have been sampled from a random set of 4 flow fields of size $N = 256$ (i.e. about 262k measurements). The central curve represents the median of the distribution, while surrounding curves correspond to percentiles.

a given value of σ. Fig. 5 shows that, for a given value of σ, a linear relation exists between α and the displacement distance computed as the norm of the displacement vectors, with a slope that is however not proportional to σ. The following section focuses on deformations based on control points, that remedy to this shortcoming.

3.3.2. Deformations based on control points

Deformations based on control points are highly common, and a number of interpolation methods have been adapted to this task. Given a set of source control points $\mathbf{C}' = \{\mathbf{c}'_k\}_{k=1}^K$, target locations $\mathbf{C} = \{\mathbf{c}_k\}_{k=1}^K$ are computed with random displacements. As suggested by the authors of U-Net [15], we consider here that the local magnitude of the displacements follows a Normal distribution with zero mean and standard deviation σ, i.e.:

$$\mathbf{C} = \mathbf{C}' + \sigma \overrightarrow{\mathbf{V}} \tag{4}$$

where $\overrightarrow{\mathbf{V}} = \{\vec{\mathbf{v}}_k\}_{k=1}^K$ is such that $\mathbf{v}_k \sim \mathcal{N}(0, 1)$. The flow field $\overrightarrow{\mathbf{D}}$ is computed by interpolating the displacement vectors between the target control points $\mathbf{C}$, i.e.

$$\overrightarrow{\mathbf{D}} = \mathcal{T}(\mathbf{P}, -\overrightarrow{\mathbf{V}}) \tag{5}$$

where $\mathcal{T}$ is an interpolation function. If topological relations between control points are available, e.g. in the form of a regular grid, common choices are bilinear, or bicubic interpolation as suggested by the authors of the U-Net model [15]. For a higher degree of smoothness and unstructured control points respectively, we consider *spline* and *Moving Least Squares* (MLS) interpolation, that we detail below.

In this approach, the source control points $\mathbf{C}'$ and the scaling σ are the user-controlled parameters. We detail the available options to generate control points later on.

Spline interpolation. Polyharmonic splines can be applied to structured or unstructured control points. The interpolating function has the form:

$$f(\mathbf{p}) = \sum_{k=1}^{K} w_k \, \phi(\|\mathbf{p} - \mathbf{c}_k\|) + \mathbf{a}^\top \begin{bmatrix} \mathbf{p} \\ 1 \end{bmatrix} \tag{6}$$

where the first term is a weighted sum of polyharmonic Radial Basis Function (RBF) terms, with centers $\mathbf{C}$, and the second term is a polynomial linear term. In this paper we consider the two following choices for ϕ:

- $\phi(r) = r^2 \ln(r)$, which corresponds to the well-known Thin Plate Spline (TPS) [46];
- $\phi(r) = r^3$, which is the cubic spline.

The weights $\{w_k\}_{k=1}^K$ and the coefficients of the polynomial $\mathbf{a}$ are estimated such that the interpolating function exactly fits the value of the function at the control points $\mathbf{C}$, which is achieved by solving a linear system.

Moving Least Squares deformations. Moving Least Squares (MLS) deformations have been introduced by Schaefer et al. in [47]. Given the set of source control points $\mathbf{C}'$ and target points $\mathbf{C}$, MLS deformation finds a transformation function $\mathcal{D}$ that satisfies three conditions: (1) Interpolation: the control points $\mathbf{C}$ map to $\mathbf{C}'$, i.e. $\mathcal{D}(\mathbf{C}) = \mathbf{C}'$; (2) Smoothness: $\mathcal{D}$ should produce smooth deformations; (3) Identity: if $\mathbf{C}' = \mathbf{C}$, then $\mathcal{D}$ should be the identity function, i.e. $\mathcal{D}(\mathbf{C}) = \mathbf{C}$. For any input pixel location $\mathbf{p}$ to be transformed, $\mathcal{D}(\mathbf{p})$ is constructed using Moving Least Squares to minimize

$$\sum_{k=1}^{K} w_k \|(\mathbf{c}_k - \bar{\mathbf{c}})\mathbf{M} - (\mathbf{c}'_k - \bar{\mathbf{c}}')\|^2, \tag{7}$$

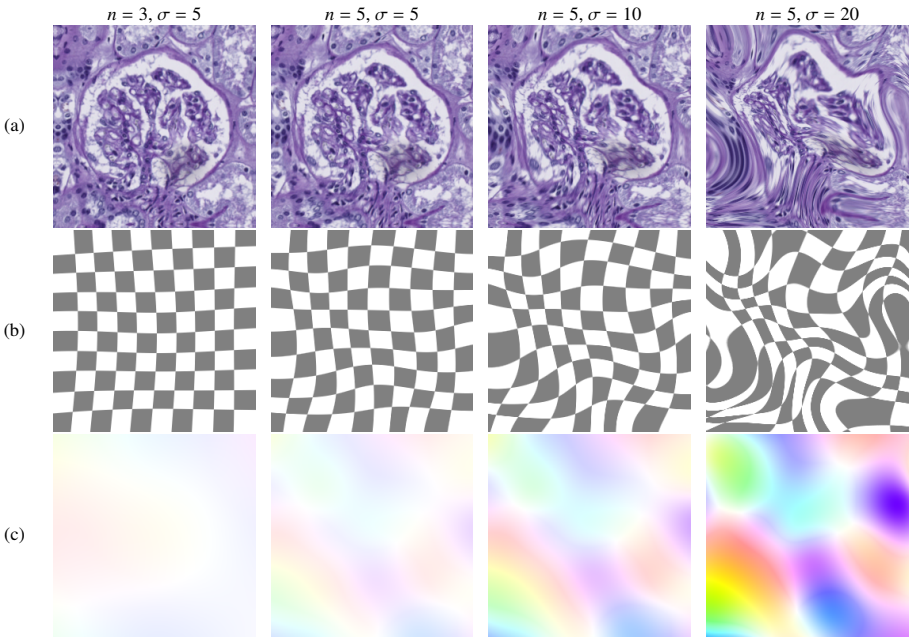

Figure 6: Examples of grid-based deformations with order-3 interpolation applied to a WSI patch containing a single glomerulus, for varying values of n and σ. Note that the directions of displacement vectors are identical in the three examples where $n = 5$. (a): Deformed patch; (b): Deformation applied to a checkerboard; (c): HSV color-coded flow field.

where $\mathbf{M}$ is an affine transformation matrix, and weights w_k are defined by $w_k = \frac{1}{\|\mathbf{c}_k - \mathbf{p}\|^4}$. The points $\bar{\mathbf{c}}$ and $\bar{\mathbf{c}}'$ are the weighted centroids of the target and source control points, respectively. We focus here on the *as-rigid-as-possible* version of MLS deformations (MLS-ARAP), where the constraint $\mathbf{M}^\top \mathbf{M} = \mathbf{I}$ is introduced (i.e. the matrix $\mathbf{M}$ contains no scaling). Further mathematical details are provided in Appendix A.

The goal of MLS-ARAP is to produce a realistic result by preserving as much as possible the shapes and textures in the input image when moving the control points. For this purpose, control points should be placed on *rigid* parts, whereas other areas are considered as flexible.

Control points on a regular grid. Control points on a regular grid are easily generated by computing a tessellation of the input image. This method can thus be used to deform any type of image content in a random fashion, and we call this kind of method Grid-Based Deformation (GBD). The spacing between grid points determines how much the displacement vectors at neighbor pixel locations are correlated, and hence fixes the maximum amount of distortion of the flow field. The value of the scaling factor σ determines the amount of distortion.

The grid points spacing is generally set by choosing the grid size for a given image size, which we will denote by $n^2 = K$. The corresponding distance in pixels between neighbor control points is thus $d = N/(n-1)$. For the same value of d, a large scaling

factor σ will yield a higher amount of distortion than a small scaling factor (see Supplemental material for corresponding plots of foldover rate, preservation of distances and preservation of angles).

The maximum displacement in pixels along one axis for a given value of σ obey the three-sigma rule, i.e. about 68% of the displacements will be less than σ pixels, while 28% will be comprised in $[\sigma, 2\sigma]$, and 4% will fall in $[2\sigma, 3\sigma]$, which offers a more intuitive control than Random Displacement Fields.

Fig. 6 shows examples of deformations based on control points on a regular grid using a cubic spline (GBD order-3, or GBD-3) interpolation, including the parameter values proposed in the U-Net paper [15], scaled to our patch size, which gives $n = 3$ and $\sigma = 5$. The parameter values proposed in the U-Net paper result in very moderate distortions, which leads us to consider stronger deformations. The other examples use $n = 5$ with σ ranging from 5 to 20, yielding increasing distortions.

Fig. 3 and 4 of the Supplemental material show the behavior of random grid-based deformations with TPS (order-2) and cubic (order-3) interpolation, respectively, with varying values of n and σ, for an image size $N = 256$. Almost no difference is observed between TPS and cubic interpolation. It can be seen that foldover starts to occur with increasing probability among random deformation fields when $\sigma \approx 10$ for $n = 3$ ($d \approx 85$), and $\sigma > 5$ for $n = 5$ ($d \approx 51$) and $n = 10$ ($d \approx 26$). A simple way to drastically limit foldover would be to consider $d/2$ as an upper bound for σ. However, this approach excludes some foldover-free deformations that may contribute to the data augmentation. For example, approximately 30% of the deformations with $n = 3$ and $\sigma = 50$ are foldover-free (Fig. 3 and 4 of the Supplemental material). In our study on the performance of data augmentation for the glomeruli segmentation task, we examine the impact of foldover-free deformations vs. deformations containing foldover.

Unstructured control points. Applying deformations based on unstructured control points requires to place source control points $\mathbf{C}'$ at appropriate locations in the input image. Instead of positioning them at random positions, we propose to perform a content-aware choice, and utilize detected cell nuclei centers. We describe the method in the following subsection.

3.3.3. Cell nuclei-based deformations

In this section, we describe our novel deformation method based on detected cell nuclei centers and MLS-ARAP deformations. The motivation is that the appearance of cell nuclei centers is almost invariant across WSI patches. That is, they are assumed to be *rigid* parts in the content of the patches, and our goal is to preserve their shape through a MLS-ARAP interpolation. We proceed in three steps, given a WSI patch to be deformed:

1. We detect cell nuclei centers in the patch, that become the source control points $\mathbf{C}'$ (this step has to be done only once for a given patch).
2. We compute displacement vectors at source control points to obtain the target control points $\mathbf{C}$.
3. We interpolate the displacement vectors $\{\overrightarrow{\mathbf{c}'_i\mathbf{c}_i}\}_{i=1}^{K}$ between source and target control points to get output pixel locations $\mathbf{P}$.

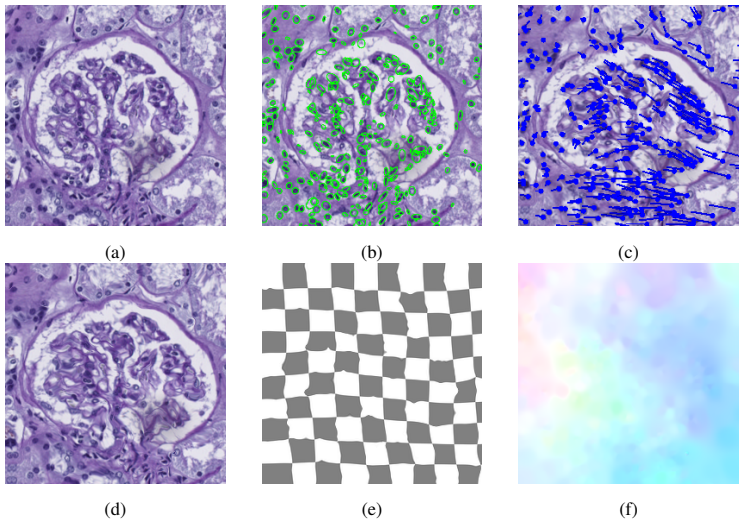

Figure 7: Example of CNB-MVN-MLS deformation applied to a WSI patch containing a single glomerulus, for $K = 310$ and $\sigma = 12$; (a): Input WSI patch; (b): Detected cell nuclei; (c): Cell nuclei displacements; (d): Deformed image; (e): Deformation applied to a checkerboard; (f): HSV color-coded flow field.

We use the method described in [48] for cell nuclei detection in WSI patches, although any other method could be used. Since cell nuclei are irregularly spaced, moving control points independently is prone to foldover. Instead, we obtain the displacement vectors of control points by sampling a multivariate Normal distribution $\mathcal{N}(0, \Sigma)$, where Σ is a covariance matrix of size $K \times K$ that takes account of the distances between detected cell nuclei centers. More precisely, we define the entries of Σ as follows:

$$\Sigma_{i,j} = \sigma \exp\left(-\frac{\|\mathbf{c}'_i \mathbf{c}'_j\|}{\sqrt{2}N}\right) \tag{8}$$

for $i, j \in [1, K]$, where σ represents a scaling factor that controls the maximum amplitude of the deformations. Since cell nuclei are spread over the whole image in each patch, we use the size of the image diagonal $\sqrt{2}N$ to normalize the distances. The user-controlled parameters therefore only consist in the scaling coefficient σ. We call this method CNB-MVN-MLS for short (Cell Nuclei Based, Multivariate Normal, MLS-ARAP). Having a single parameter is advantageous over the other considered methods for the purpose of parameter estimation. A result obtained with this method is shown in Fig. 7, where $K = 310$, and $\sigma = 12$. In our patches, the number of detected nuclei cell centers in a single WSI patch varies between 250 and 320. Additional results are presented in Fig. 8. It can be seen that the obtained deformations are richer than deformations produced with RDF or grid-based deformations: they incorporate *low* and *high frequencies*, which is a consequence of *both* the spatial correlation of control points and the rigidity constraint. Distortions have a behavior that is similar to the other examined deformation models (see Supplemental material Fig. 5 for foldover and distance/angle preservation plots).

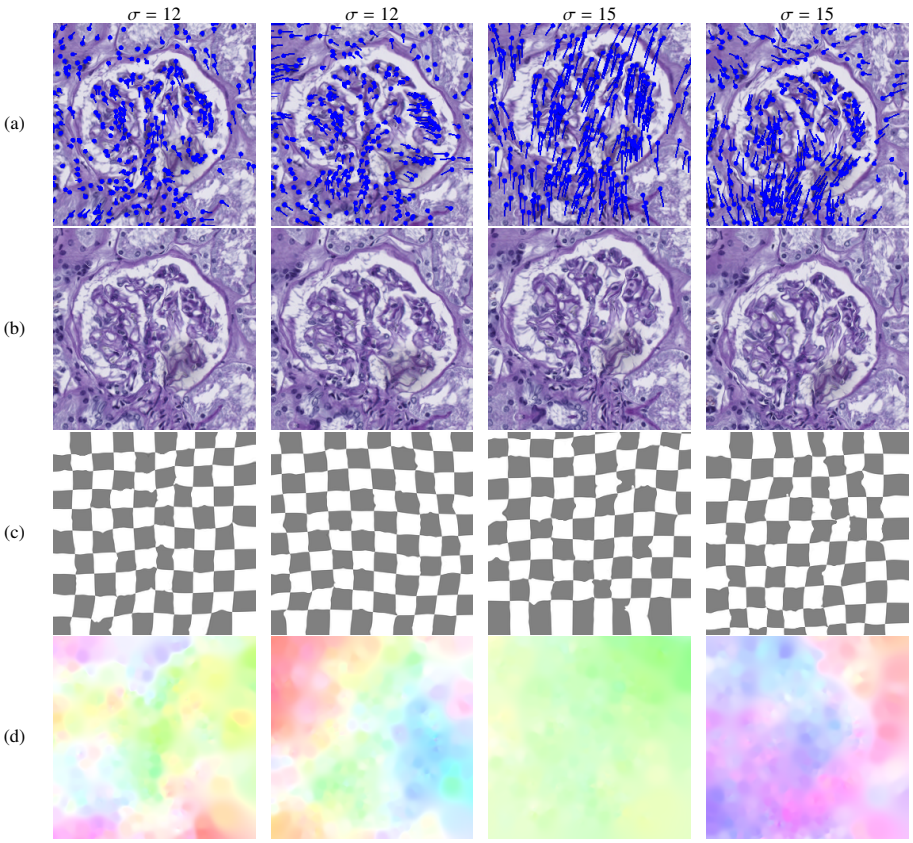

Figure 8: Additional examples of CNB-MVN-MLS deformations applied to a WSI patch containing a single glomerulus, for $K = 310$ and two values of σ; (a): Cell nuclei displacements; (b): Deformed image; (c): Deformation applied to a checkerboard; (d): HSV color-coded flow field.

4. Evaluation protocol and implementation

In this section, we present our training protocol, followed by the details of our implementation.

4.1. Evaluation protocol

Glomeruli segmentation is formulated as a two classes problem where each pixel of a WSI patch receives a label *glomeruli*, or a label *tissue*. Below we detail the dataset used to train the U-Net segmentation model, as well as our baseline and approach to assess the impact of random spatial deformations as an augmentation technique.

Dataset. The input to the U-Net are patches sampled from WSIs, such that each patch may either contain a complete glomerulus, or up to three partial glomeruli, or no glomeruli (i.e., tissue). We have at our disposal several WSIs from different patients, acquired with the same sampling resolution (0.506 μm/pixel), and distributed between

Patches	Train (total)	Validation (total)	Test (total)
Glomeruli	20 (660)	10 (589)	100 (1 091)
No glomeruli	20 (4 620)	10 (4 123)	100 (7 637)

Table 3: Number of patches used in each dataset. The number in parentheses is the total number of patches available in the original dataset.

training, validation and test, from which we extract 512×512 patches with and without glomeruli. The composition of this dataset is summarized in Table 3. We select randomly 40 patches from the training base, 20 patches from the validation base and 100 patches from the test base, half with glomeruli and the other half without glomeruli (Table 3). We purposely choose a small number of examples for this experiment in order to make it easier to highlight the differences in the results. From each patch, we randomly extract 5 crops of size 256×256. Those new patches are the ones presented to the U-Net. The test set remains the same for all subsequent experiments.

Baseline. Our baseline is composed of all the 90° rotations and flips we can apply to an image. Since the sampling resolution is the same for all patches, we did not include global scaling in our baseline. The observed objects are at the same scale in all images, even if their sizes can differ. These local variations in size are simulated in a more faithful way by spatial deformations.

Application of deformations. To assess the impact of using random spatial deformations as augmentation techniques on segmentation results using a U-Net, we propose a protocol that aims to provide a fair comparison of the different deformation models. To do so, we have to make sure that the parameter values used for each model are, as far as possible, close to optimal parameter values. We use a meta-learning scheme akin to RandAugment [40]. For each model we consider the (one or two) parameter(s) described in section 3, for which we hand-select a few values. Those values are chosen to cover different cases, with varying distortion amounts. However, due to the computational cost of meta-learning (a single training takes approximately two hours on an NVidia P100), we do not search for the exact set of optimal parameter values, but rather aim to find a good range of parameter values, and highlight the influence of the characteristics of the deformations. We performed 5 training for each set of parameter values in a grid-search manner, and averaged the results. The selected parameter values for the chosen deformation models are summarized below.

- RDF (σ, α)
  - $\sigma = \{5, 10, 20\}$
  - $\alpha = \{100, 200, 400, 800\}$
- GBD order-2 (n, σ)
  - $n = \{3, 5, 10\}$
  - $\sigma = \{5, 10, 20, 50\}$
- GBD order-3 (n, σ)
  - $n = \{3, 5, 10\}$
  - $\sigma = \{5, 10, 20, 50\}$
- CNB-MVN-MLS (σ)
  - $\sigma = \{5, 10, 15, 30, 100\}$

The augmentations are computed offline *before training*, which means that the training time is the same for all methods. For a given deformation method and a given set of parameter values, we apply 10 deformations to each *training* image. Each of

those 10 deformations is randomly sampled, meaning that we sample 2000 flow fields for each set of parameter values and each deformation model. To the training and validation sets, we then apply our baseline, thus multiplying the training dataset size by 7 (14,000 images in total for the training set).

Since we have a low number of repetitions (5 trainings per experiment), we use the Wilcoxon–Mann–Whitney statistical test [49, 50] to give an indication on whether or not we can consider two experiments as equivalent w.r.t their average Dice scores (see Appendix B for p-values).

4.2. Implementation

We implemented the U-Net in Python using TensorFlow 2, as described in the original paper [15] with only one convolution layer per stage, using transposed convolutions for upsampling, and cross entropy loss, which is a standard choice for stable training (see e.g. [51] for a review of loss functions). The contracting path is composed of 6 convolutional layers and the expanding path of 5 convolutional layers, with 64 filters in the first layer. The number of filters doubles at each stage, with a maximum of 512. We use the Adam optimizer [52] with a 4.10^{-6} learning rate.

The deformation models described in Section 3 were also implemented in Python. For Random Displacement Fields and Grid-Based Deformations, we use the TensorFlow Addons Image module, that runs on the GPU. We implemented the deformation models as CPU code for CNB-MVN-MLS deformations, although they could be implemented on the GPU either. The time required to sample a flow field and deform a 256×256 image for CNB-MVN-MLS deformations with our CPU codes is about 1.5 s, whereas RDF and GDB take 60 ms per image. Similar performance (60 ms per image) would be expected for for CNB-MVN-MLS deformations running on the GPU.

We make our code publicly available through a GitHub repository[2]. We provide the complete framework for evaluating the impact of the deformation models that we considered as augmentations for any 2D image segmentation task based on a U-Net.

5. Experiments and results

We first analyze the segmentation performance w.r.t. augmentation with spatial deformations as described in the protocol presented in the previous section. Then we study the influence of geometric distortions on the obtained performance results.

5.1. Segmentation results

We ran each training for 30 epochs. When using a larger number of epochs, we have verified experimentally that the validation loss was increasing for each training. We kept the trained segmentation model with the lowest validation loss. At test time, we evaluated the model on four metrics: Dice score (or F1-score), precision, recall and specificity. All are averaged over 5 repetitions. We recall the formulas of those metrics:

[2] `https://github.com/ASTex-ICube/spatial_deformations_aug`

σ \ α	100	200	400	800
5	0.7586 ± 0.0233	0.7275 ± 0.0387	0.5342 ± 0.0399	0.4807 ± 0.0448
10	0.7250 ± 0.0523	0.7748 ± 0.0184	0.7892 ± 0.0370	0.7787 ± 0.0175
20	0.6653 ± 0.0519	0.7367 ± 0.0481	0.7729 ± 0.0292	0.7859 ± 0.0256

σ \ α	2400	5000
40	0.7554 ± 0.0292	0.6899 ± 0.0223

(a) Average Dice score w.r.t. to the RDF parameter values

n \ σ	5	10	20	50
3	0.7884 ± 0.0224	0.8176 ± 0.0383	0.8434 ± 0.0181	0.8502 ± 0.0158
5	0.8014 ± 0.0116	0.8320 ± 0.0243	0.8541 ± 0.0125	0.7888 ± 0.0351
10	0.8028 ± 0.0202	0.8402 ± 0.0317	0.7648 ± 0.0321	0.7171 ± 0.0473

(b) Average Dice score w.r.t. to the GBD-2 parameter values

n \ σ	5	10	20	50
3	0.7688 ± 0.0135	0.8177 ± 0.0184	0.8593 ± 0.0162	0.8510 ± 0.0139
5	0.7931 ± 0.0207	0.8407 ± 0.0344	0.8114 ± 0.0319	0.7381 ± 0.0542
10	0.8113 ± 0.0311	0.8448 ± 0.0165	0.7506 ± 0.0328	0.7011 ± 0.0197

(c) Average Dice score w.r.t. to the GBD-3 parameter values

σ	5	10	15	30	100
	0.8212 ± 0.0116	0.8501 ± 0.0121	0.8525 ± 0.0182	0.8337 ± 0.0161	0.70417 ± 0.0298

(d) Average Dice score w.r.t. to the CNB-MVN-MLS parameter values

Table 4: U-Net glomeruli segmentation performance considering the average Dice score for the parameter values of the deformation models considered in the paper.

- Dice score $= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$
- Precision $= \frac{|\text{True positives}|}{|\text{True positives}| + |\text{False positives}|}$
- Recall $= \frac{|\text{True positives}|}{|\text{True positives}| + |\text{False negatives}|}$
- Specificity $= \frac{|\text{True negatives}|}{|\text{True negatives}| + |\text{False positives}|}$

For simplicity, we will focus only on the average Dice score, but the results for all metrics can be found in the Supplementary material. We present the average Dice score for all deformation models in Table 4. We highlight in red the best score, and in blue the scores that are equivalent to the best one w.r.t. the Wilcoxon test. Note that there can be a difference of up to 0.06 in average Dice score for two equivalent experiments in the sense of the Wilcoxon test, meaning we cannot consider the sets of parameter values corresponding to these experiments as interchangeable. However, the Wilcoxon test gives us an interval where we can safely assume the best parameter values are w.r.t. the chosen value ranges, thus potentially helping us refine the grid search. In particular, in order to provide the best bounds possible, we tried to have non-equivalent parameter values (i.e. values not highlighted) at the extremities of each line with the best average Dice score. This is the reason why we conducted some additional experiments w.r.t. the ranges of values defined in the protocol. We also give the average Dice score when only the baseline (i.e. no deformation) is used for data augmentation: 0.6231 ± 0.0401. It should be noticed that there may be multiple local maxima in the landscape of the average Dice scores. The chosen ranges of parameter values determine some maxima that are not guaranteed to be absolute.

When focusing on a single deformation model at a time, the impact of the parameter values clearly appears. For example, in the case of the RDF method, some parameter values are detrimental, such as $\sigma = 20$ and $\alpha = 50$, while others can boost the average Dice score from 0.62 to more than 0.75. All the selected parameter values for the GBD-3 method improve the results, the values corresponding to the highest average Dice score being $n = 3$ and $\sigma = 20$.

We summarize below the best average Dice scores for each deformation model and corresponding parameter values, obtained from the sample values described in Section 4.1:

- Baseline: 0.6231 ± 0.0401
- RDF ($\sigma = 10$, $\alpha = 400$): 0.7892 ± 0.0370

- GBD order-2 ($n = 5, \sigma = 20$): 0.8541 ± 0.0125
- GBD order-3 ($n = 3, \sigma = 20$): 0.8593 ± 0.0162
- CNB-MVN-MLS ($\sigma = 15$): 0.8525 ± 0.0182

The best average Dice scores are obtained with the GBD with order-3 interpolation method, but we note that aside from the RDF method, all results are close. Considering the variance and the small number of repetitions, we can state that all methods (aside RDF) are equivalent in performance. This is corroborated by the Wilcoxon test, that we also applied to the best average dice scores obtained with the different deformation models. Note that this does not mean that the RDF method is necessary inferior: we could not find parameter values that provide the same performance as the other methods, which indicates that they either do not exist or are more difficult to find. This corroborates the observation made in Section 3 that the RDF model is more difficult to control than the other deformation models. We also note that the change of the interpolation method from order-2 to order-3 for GBD does not provide a noticeable change in the average Dice score (0.005 gap).

The CNB-MVN-MLS model perfoms on par with the other models, but with a single parameter to control, which significantly reduces the computational complexity of finding a best parameter value.

This leads us to the conclusion that changing methods is not the most decisive factor and that an optimization of the parameter values is crucial to obtain the best segmentation results. Since the maximum reported average Dice scores are close from one deformation model to the other, we can also state that the obtained values are close to a global maximum with good confidence. The observed similar performance of the different deformation models is consistent with the fact that they all rely on the sampling of a filtered white noise, as detailed in 3. A significant advantage of the CNB-MVN-MLS model lies in its control through a single parameter.

For more completeness, we made a complementary experiment mixing deformations with several parameter values. We first examined picking the parameter values of the GBD order-3 method at random. We fixed $n = 3$ and uniformly sampled σ among the values 10, 20 and 50, for each of the 2000 transformations we apply in the protocol. We obtain an average Dice score of 0.8116 ± 0.0287, the same as for $(n = 3, \sigma = 10)$. Thus mixing the parameter values does not seem to provide additional diversity and seems to hold the performance to the lowest score among the scores corresponding to single parameter values.

5.2. Impact of distortions

When looking at the deformed images, we observe that deformations with moderate distorsions, which are often used to keep the examples "realistic", perform lesser than deformations involving stronger distortions that may look unrealistic. As presented in Section 3, deformations with a high amount of distortions implies a lesser preservation of distances and angles in the result images. Varying the angles have important consequences on the shapes, which leads to better generalizability (or robustness) of the segmentation model. As one could expect, deformations provoking foldover result in a degradation of the segmentation performance as the foldover rate increases. However, deformations containing moderate foldover can still yield better performance

than the baseline. As an example, GBD-3 deformations with $n = 3$ and $\sigma = 5$ induce very limited distortion, and lead to +0.15 in average Dice score in comparison with the baseline. With the same model, deformations with $n = 3$ and $\sigma = 50$ yield a higher amount of distortion, with 1/3 of deformations containing foldover (see Fig. 3 of the Supplemental material), and lead to a gain of +0.23. Deformation with $n = 10$ and $\sigma = 50$ provoke a high foldover rate (more than 40%), yielding a +0.08 amelioration in average Dice score.

Our interpretation is that foldover creates new patterns that are not found in the original images, thus introducing variety that may be beneficial for training the U-Net model. However, if those new patterns are dominant, the original appearance of the image (especially the topological structure) may be completely lost, yielding to poor contribution to training, or even degrading the segmentation performance.

5.3. Limitations

Since they are based on the RandAugment strategy [40], our proposed protocol are computationally expensive. We had to perform 420 trainings, which amounts to approximately 840 GPU hours on an NVidia Tesla P100 SXM2 with 16 GB RAM. While our parameter sampling choices and statistical tests yield only bounds on the best parameter values, estimating tighter bounds or exact best parameter values for each method would require additional computational resources with this approach. In particular, more repetitions for each set of parameter values could provide a better separation between close parameter values, or conversely show that they are equivalent. We provide our complete evaluation framework to the end of both enabling parameter values refinement, depending on the available computing power, and evaluating data augmentation based on spatial deformation for other datasets or applications.

It is worth mentioning that our conclusions cannot be generalized to other dataset sizes or to other dataset types, i.e. other images. As we know from [40], the size of the training dataset has an impact on the optimization process of a deep model hyperparameters. In order to exhibit a difference between the performance gain enabled by the deformation models and their parameter values, we used only 20 images with glomeruli. It helped us show the impact of deformations as an augmentation technique with a potential gain of more than 0.20 in average Dice score, but at the cost of a relatively high variance across trainings. Thus the low number of images is another factor, in conjunction with the small number of repetitions, that contributes to the large ranges obtained for the estimation of best parameter values.

Regarding other dataset types, the impact of the deformation models may be different, since changing the images changes the shapes and textures. We did not conducted experiments with images other than renal WSI patches, but the framework we provide allows for investigations on any image type and dataset size.

6. Conclusion and future work

In this paper, we evaluated the impact of random spatial deformations used as augmentation in the case of renal histopathology patches segmentation using a U-Net

model. We focused our study on spatial deformations based on a multivariate Gaussian random field. We performed a comparative study of state-of-the-art deformation models based on this approach used in the field of digital histopathology, considering geometric aspects. We proposed an original content-aware random deformation model named CNB-MVN-MLS based on the detection and displacement of cell nuclei centers that involves a single parameter.

We compared the performance offered by these deformation models when used for augmentation in the considered segmentation task. We used a meta learning scheme to optimize their parameter values and find appropriate bounds. We found out that the different models show similar performance (measured in average Dice score) and that using the good range of parameter values can boost the average Dice score by up to 0.23 on our application. We showed that deformations with relatively strong distortions yield the best performance increase for segmentation, while previous work only report the use of deformations with low distortions. Our CNB-MVN-MLS model performs on par with the other considered deformation models but with only one parameter to control, improving the performance of best parameter estimation.

In future work, we would like to develop an online protocol, and make a comparison with the proposed offline protocol. By using more images and/or trainings, we could also lower the variance across training and thus provide better bounds for the optimal parameters. We also plan to extend the evaluation to alternative pre-trained network architectures for segmentation like Seg-Net [17] or U-Net with a ResNet34 backbone [20].

Following the direction of work of our content-aware CNB-MVN-MLS random deformation model, we would like to further develop models with an improved correlation between the displacement vectors and the content of images. Methods based on physical simulation [42] take account of material and mechanical properties, but at a greater computational cost for the generation of deformation fields, which would require to identify a trade-off.

7. Acknowledgments

The WSI patches used in our experiments are courtesy of the ERACoSysMed project "SysMIFTA", co-funded by EU H2020 and the national funding agencies German Ministry of Education and Research (BMBF) project management PTJ (FKZ: 031L-0085A), and Agence National de la Recherche (ANR), project number ANR-15-CMED-0004.

The authors would like to acknowledge the High Performance Computing Center ROMEO hosted by the University of Reims Champagne-Ardenne for access to computing resources.

8. Declaration of competing interest

The authors declare that they have no conflict of interest.

Appendix A. As-rigid-as-possible MLS deformations

Referring to [47] and 3.3.2, let $\bar{\mathbf{c}}$ and $\bar{\mathbf{c}}'$ denote the weighted centroids of the target and source control points, computed as $\bar{\mathbf{c}} = \frac{\sum_k w_k \mathbf{c}_k}{\sum_k w_k}$ and $\bar{\mathbf{c}}' = \frac{\sum_k w_k \mathbf{c}'_k}{\sum_k w_k}$. The corresponding transformation function for MLS-ARAP is defined as follows:

$$\mathcal{D}(\mathbf{p}) = \|\mathbf{p} - \bar{\mathbf{c}}\| \frac{\overrightarrow{\mathbf{D}}(\mathbf{p})}{\left\|\overrightarrow{\mathbf{D}}(\mathbf{p})\right\|} + \bar{\mathbf{c}}', \tag{A.1}$$

where $\overrightarrow{\mathbf{D}}(\mathbf{p}) = \sum_k (\mathbf{c}'_k - \bar{\mathbf{c}}')\mathbf{A}_k$ with:

$$\mathbf{A}_k = w_k \begin{bmatrix} \mathbf{c}_k - \bar{\mathbf{c}} \\ (\mathbf{c}'_k - \bar{\mathbf{c}}')^{\perp} \end{bmatrix} \begin{bmatrix} \mathbf{p} - \bar{\mathbf{c}} \\ -(\mathbf{p} - \bar{\mathbf{c}})^{\perp} \end{bmatrix}^{\top} \tag{A.2}$$

where $\perp$ is an operator on 2D vectors such that $[x, y]^{\perp} = [-y, x]$.

Appendix B. Wilcoxon-Mann-Whitney statistical test

Considering the Dice scores of two experiments as independent observations from two distributions X and Y, we test the null hypothesis H0, i.e. the distributions of both populations are equal. If we note $X_1, ..., X_n$ and $Y_1, ..., Y_m$ the samples drawn from each distribution, we can define the U-statistic as

$$U = \sum_{i=1}^{n} \sum_{j=1}^{m} S(X_i, Y_j), \tag{B.1}$$

with

$$S(X, Y) = \begin{cases} 1 & \text{if } X > Y, \\ \frac{1}{2} & \text{if } Y = X, \\ 0 & \text{if } X < Y. \end{cases} \tag{B.2}$$

With $n = m = 5$, we can reject the null hypothesis when $U \leq 2$ or $U \geq 23$ according to tables in [49]. We give below the associated p-values for every comparison we made.

Appendix B.1. Comparison between deformation models

We present in Table 5 ((a) and (b)) of the Supplemental material the comparison between the considered deformations models, under two different hypothesis and using the best set of parameters for each model. Table 5(a) gives the p-values when testing H0: the baseline Dice score is higher (one-sided test), (b) gives the p-values when testing H0: the two considered distributions are equal.

Appendix B.2. Comparison between parameter values for each deformation model

Table 5 ((b) to (f)) of the Supplemental material presents the comparison between the parameter values for the considered deformation models individually, under the hypothesis that two distributions are equal (two-sided test). A p-value greater then 0.05 means that the two sets of parameter values yield significantly close results.

References

[1] G. Litjens, C. I. Sánchez, N. Timofeeva, M. Hermsen, I. Nagtegaal, I. Kovacs, C. Hulsbergen-van de Kaa, P. Bult, B. van Ginneken, J. van der Laak, Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis, Scientific reports 6 (2016) 26286. doi:10.1038/srep26286.

[2] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, C. I. Sánchez, A survey on deep learning in medical image analysis, Medical Image Analysis 42 (2017) 60–88. doi:10.1016/j.media.2017.07.005.

[3] D. Komura, S. Ishikawa, Machine learning methods for histopathological image analysis, Computational and Structural Biotechnology Journal 16 (2018) 34–42. doi:10.1016/j.csbj.2018.01.001.

[4] N. Dimitriou, O. Arandjelović, P. D. Caie, Deep learning for whole slide image analysis: An overview, Frontiers in Medicine 6 (2019) 264. doi:10.3389/fmed.2019.00264.

[5] C. L. Srinidhi, O. Ciga, A. L. Martel, Deep neural network models for computational histopathology: A survey, Medical Image Analysis 67 (2021) 101813. doi:10.1016/j.media.2020.101813.

[6] C. Wemmert, J. Weber, F. Feuerhake, G. Forestier, Deep learning for histopathological image analysis, Deep Learning for Biomedical Data Analysis (2021) 153–169doi:10.1007/978-3-030-71676-9_7.

[7] L. Barisoni, K. J. Lafata, S. M. Hewitt, A. Madabhushi, U. G. J. Balis, Digital pathology and computational image analysis in nephropathology, Nature Reviews Nephrology 16 (11) (2020) 669–685. doi:10.1038/s41581-020-0321-6.

[8] C. Shorten, T. M. Khoshgoftaar, A survey on image data augmentation for deep learning, Journal of Big Data 6 (1) (2019) 60. doi:10.1186/s40537-019-0197-0.

[9] E. D. Cubuk, B. Zoph, D. Mané, V. Vasudevan, Q. V. Le, AutoAugment: Learning augmentation strategies from data, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 113–123. doi:10.1109/CVPR.2019.00020.

[10] X. Zhang, Q. Wang, J. Zhang, Z. Zhong, Adversarial AutoAugment, in: Proc. International Conference on Learning Representations, 2020.

[11] R. H. Naik, S. H. Shawar, Renal transplantation rejection, StatPearls [Internet]PMID: 31971715.

[12] T. Lampert, O. Merveille, J. Schmitz, G. Forestier, F. Feuerhake, C. Wemmert, Strategies for training stain invariant CNNS, in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), 2019, pp. 905–909. doi:10.1109/ISBI.2019.8759266.

[13] R. Liu, L. Wang, J. He, W. Chen, Towards staining independent segmentation of glomerulus from histopathological images of kidney, bioRxivdoi:10.1101/821181.

[14] J. Vasiljević, F. Feuerhake, C. Wemmert, T. Lampert, Self adversarial attack as an augmentation method for immunohistochemical stainings, 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI).

[15] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Cham, 2015, pp. 234–241.

[16] T. de Bel, M. Hermsen, B. Smeets, L. Hilbrands, J. van der Laak, G. Litjens, Automatic segmentation of histopathological slides of renal tissue using deep learning, in: Medical Imaging 2018: Digital Pathology, Vol. 10581, International Society for Optics and Photonics, 2018, pp. 285 – 290.

[17] G. Bueno, M. M. Fernandez-Carrobles, L. Gonzalez-Lopez, O. Deniz, Glomerulosclerosis identification in whole slide images using semantic segmentation, Computer Methods and Programs in Biomedicine 184 (2020) 105273. doi:https://doi.org/10.1016/j.cmpb.2019.105273.
URL https://www.sciencedirect.com/science/article/pii/S0169260719311381

[18] O. Merveille, T. Lampert, J. Schmitz, G. Forestier, F. Feuerhake, C. Wemmert, An automatic framework for fusing information from differently stained consecutive digital whole slide images: A case study in renal histology, Computer Methods and Programs in Biomedicine 208 (2021) 106157. doi:10.1016/j.cmpb.2021.106157.

[19] R. C. Davis, X. Li, Y. Xu, Z. Wang, N. Souma, G. Sotolongo, J. Bell, M. Ellis, D. Howell, X. Shen, K. Lafata, L. Barisoni, Deep learning segmentation of glomeruli on kidney donor frozen sections, medRxivdoi:10.1101/2021.09.16.21263707.

[20] M. Salvi, A. Mogetta, A. Gambella, L. Molinaro, A. Barreca, M. Papotti, F. Molinari, Automated assessment of glomerulosclerosis and tubular atrophy using deep learning, Computerized Medical Imaging and Graphics 90 (2021) 101930. doi:https://doi.org/10.1016/j.compmedimag.2021.101930.
URL https://www.sciencedirect.com/science/article/pii/S0895611121000793

[21] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, J. Liang, UNet++: A nested U-Net architecture for medical image segmentation, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Cham, 2018, pp. 3–11.

[22] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, J. Liang, UNet++: Redesigning skip connections to exploit multiscale features in image segmentation, IEEE Transactions on Medical Imaging 39 (6) (2020) 1856–1867. doi:10.1109/TMI.2019.2959609.

[23] A. G. Roy, N. Navab, C. Wachinger, Concurrent spatial and channel 'squeeze & excitation' in fully convolutional networks, in: Medical Image Computing and Computer Assisted Intervention – MICCAI 2018, Cham, 2018, pp. 421–429.

[24] M. Hermsen, T. de Bel, M. den Boer, E. J. Steenbergen, J. Kers, S. Florquin, J. J. T. H. Roelofs, M. D. Stegall, M. P. Alexander, B. H. Smith, B. Smeets, L. B. Hilbrands, J. A. W. M. van der Laak, Deep Learning–Based Histopathologic Assessment of Kidney Tissue, Journal of the American Society of Nephrology 30 (10) (2019) 1968–1979. doi:10.1681/ASN.2019020144.

[25] M. Gadermayr, A.-K. Dombrowski, B. M. Klinkhammer, P. Boor, D. Merhof, Cnn cascades for segmenting sparse objects in gigapixel whole slide images, Computerized Medical Imaging and Graphics 71 (2019) 40–48. doi:10.1016/j.compmedimag.2018.11.002.

[26] N. Altini, G. D. Cascarano, A. Brunetti, F. Marino, M. T. Rocchetti, S. Matino, U. Venere, M. Rossini, F. Pesce, L. Gesualdo, V. Bevilacqua, Semantic segmentation framework for glomeruli detection and classification in kidney histological sections, Electronics 9 (3). doi:10.3390/electronics9030503.

[27] J. Kuklyte, J. Fitzgerald, S. Nelissen, H. Wei, A. Whelan, A. Power, A. Ahmad, M. Miarka, M. Gregson, M. Maxwell, R. Raji, J. Lenihan, E. Finn-Moloney, M. Rafferty, M. Cary, E. Barale-Thomas, D. O'Shea, Evaluation of the Use of Single- and Multi-Magnification Convolutional Neural Networks for the Determination and Quantitation of Lesions in Nonclinical Pathology Studies, Toxicologic Pathology 49 (4) (2021) 815–842, pMID: 33618634. doi:10.1177/0192623320986423.

[28] M. Temerinac-Ott, G. Forestier, J. Schmitz, M. Hermsen, J. Bräsen, F. Feuerhake, C. Wemmert, Detection of glomeruli in renal pathology by mutual comparison of multiple staining modalities, in: Proceedings of the 10th International Symposium on Image and Signal Processing and Analysis, 2017, pp. 19–24. doi:10.1109/ISPA.2017.8073562.

[29] J. Gallego, A. Pedraza, S. Lopez, G. Steiner, L. Gonzalez, A. Laurinavicius, G. Bueno, Glomerulus classification and detection based on convolutional neural networks, Journal of Imaging 4 (1). doi:10.3390/jimaging4010020.

[30] S. Kannan, L. A. Morgan, B. Liang, M. G. Cheung, C. Q. Lin, D. Mun, R. G. Nader, M. E. Belghasem, J. M. Henderson, J. M. Francis, V. C. Chitalia, V. B. Kolachalama, Segmentation of glomeruli within trichrome images using deep learning, Kidney International Reports 4 (7) (2019) 955–962. doi:10.1016/j.ekir.2019.04.008.

[31] R. Heckenauer, J. Weber, C. Wemmert, F. Feuerhake, M. Hassenforder, P.-A. Muller, G. Forestier, Real-time detection of glomeruli in renal pathology, in: 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS), 2020, pp. 350–355. doi:10.1109/CBMS49503.2020.00072.

[32] M. D. Bloice, P. M. Roth, A. Holzinger, Biomedical image augmentation using Augmentor, Bioinformatics 35.21 (2019) 4522–4524.

[33] Y.-R. Van Eycke, A. Foucart, C. Decaestecker, Strategies to reduce the expert supervision required for deep learning-based segmentation of histopathological images, Frontiers in Medicine 6 (2019) 222. doi:10.3389/fmed.2019.00222.

[34] Y. Xiao, E. Decencière, S. Velasco-Forero, H. Burdin, T. Bornschlögl, F. Bernerd, E. Warrick, T. Baldeweck, A new color augmentation method for deep learning segmentation of histological images, in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), 2019, pp. 886–890. doi:10.1109/ISBI.2019.8759591.

[35] S. T. Mpinda Ataky, J. de Matos, A. d. S. Britto, L. E. S. Oliveira, A. L. Koerich, Data augmentation for histopathological images based on gaussian-laplacian pyramid blending, in: 2020 International Joint Conference on Neural Networks (IJCNN), 2020, pp. 1–8. doi:10.1109/IJCNN48605.2020.9206855.

[36] K. Faryna, J. van der Laak, G. Litjens, Tailoring automated data augmentation to H&E-stained histopathology, in: Proc. Conference on Medical Imaging with Deep Learning, Vol. 143 of Proc. Machine Learning Research, 2021, pp. 168–178.
URL https://proceedings.mlr.press/v143/faryna21a.html

[37] D. Tellez, G. Litjens, P. Bándi, W. Bulten, J.-M. Bokhorst, F. Ciompi, J. van der Laak, Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology, Medical Image Analysis 58 (2019) 101544. doi:10.1016/j.media.2019.101544.

[38] A. Kanazawa, D. W. Jacobs, M. Chandraker, Warpnet: Weakly supervised matching for single-view reconstruction, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 3253–3261. doi:10.1109/CVPR.2016.354.

[39] S. Lim, I. Kim, T. Kim, C. Kim, S. Kim, Fast AutoAugment, in: Advances in Neural Information Processing Systems, Vol. 32, 2019, pp. 6665—-6675.

[40] E. D. Cubuk, B. Zoph, J. Shlens, Q. V. Le, Randaugment: Practical automated data augmentation with a reduced search space, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020, pp. 3008–3017. doi:10.1109/CVPRW50498.2020.00359.

[41] Z. Tang, Y. Gao, L. Karlinsky, P. Sattigeri, R. Feris, D. Metaxas, Onlineaugment: Online data augmentation with less domain knowledge, in: Computer Vision – ECCV 2020, Cham, 2020, pp. 313–329.

[42] A. Nealen, M. Müller, R. Keiser, E. Boxerman, M. Carlson, Physically based deformable models in computer graphics, Computer Graphics Forum 25 (4) (2006) 809–836. doi:10.1111/j.1467-8659.2006.01000.x.

[43] M. Wang, P. Li, A review of deformation models in medical image registration, Journal of Medical and Biological Engineering 39 (1) (2019) 1–17. doi:10.1007/s40846-018-0390-1.

[44] P. Chlap, H. Min, N. Vandenberg, J. Dowling, L. Holloway, A. Haworth, A review of medical image data augmentation techniques for deep learning applications, Journal of Medical Imaging and Radiation Oncology 65 (5) (2021) 545–563. doi:10.1111/1754-9485.13261.

[45] P. Y. Simard, D. Steinkraus, J. C. Platt, Best practices for convolutional neural networks applied to visual document analysis, in: Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings., 2003, pp. 958–963. doi:10.1109/ICDAR.2003.1227801.

[46] F. Bookstein, Principal warps: thin-plate splines and the decomposition of deformations, IEEE Transactions on Pattern Analysis and Machine Intelligence 11 (6) (1989) 567–585. doi:10.1109/34.24792.

[47] S. Schaefer, T. McPhail, J. Warren, Image deformation using moving least squares, ACM Trans. Graph. 25 (3) (2006) 533–540. doi:10.1145/1141911.1141920.

[48] F. Mahmood, D. Borders, R. J. Chen, G. N. Mckay, K. J. Salimian, A. Baras, N. J. Durr, Deep adversarial training for multi-organ nuclei segmentation in histopathology images, IEEE Transactions on Medical Imaging 39 (11) (2020) 3257–3267. doi:10.1109/TMI.2019.2927182.

[49] F. Wilcoxon, Individual comparisons by ranking methods, Biometrics Bulletin 1 (6) (1945) 80–83.

[50] H. B. Mann, D. R. Whitney, On a test of whether one of two random variables is stochastically larger than the other, The Annals of Mathematical Statistics 18 (1) (1947) 50 – 60. doi:10.1214/aoms/1177730491.

[51] S. Asgari Taghanaki, K. Abhishek, J. P. Cohen, J. Cohen-Adad, G. Hamarneh, Deep semantic segmentation of natural and medical images: a review, Artificial Intelligence Review 54 (1) (2021) 137–178. doi:10.1007/s10462-020-09854-1.
URL https://doi.org/10.1007/s10462-020-09854-1

[52] D. Kingma, J. Ba, Adam: A method for stochastic optimization, International Conference on Learning Representations.