

# Defocus Blur Synthesis and Deblurring through Interpolation in the Latent Space

IOANA-GEORGIANA MAZILU, University of Twente, The Netherlands

High quality images play an important role in medical diagnosis and analysis. To ensure that microscopic imaging results are qualitative, most microscope systems nowadays are equipped with autofocus hardware or software-based components. Nonetheless, there are cases when the optimal focal distance is not correctly identified and images present out-of-focus areas or are completely affected by defocus blur. In this paper, we investigate a generative model with twofold applicability. It can be used for recovering from defocus blur as well as for synthesizing defocus blur for data augmentation purposes. Both these tasks are achieved through interpolation in the latent space of an autoencoder. We apply two forms of linear enforcement to the latent space of an autoencoder trained to synthesize defocus blur in microscopy images. We evaluate the models and find that the regularized autoencoders outperform the baseline model in terms of synthesizing blur and deblurring images.

Additional Key Words and Phrases: latent space, linear interpolation, autoencoder, deblurring, microscopy, defocus blur synthesis

## 1 INTRODUCTION

In the medical domain, access to high-quality images is critical for professionals and programs since it affects image analysis and diagnosis. Computer vision is widely used in medical image processing, especially ever since research in deep learning has led to advancements in the performance of tasks such as cell segmentation [21], disease classification [33], image denoising [17], and image enhancement [9], [3].

A commonly used method of acquiring high resolution cell images is through the use of whole slide scanners [14]. Such systems have software or hardware-based autofocus components used to capture microscopy images at an optimal focal distance. This task is often impeded by the structure of the tissue that is scanned, as not all the cells may lie at an equal distance from the lens. Therefore, some images produced this way are prone to being affected by out-of-focus areas [8], which negatively impact their analysis.

Numerous deblurring solutions based on deep learning have emerged in the past years. Generally, these solutions can be divided in two categories, those using a blur kernel for the deblur operation, and kernel-free approaches. Kernel-dependent solutions may require computing this kernel, which is often a complex operation and could negatively impact the deblurring process if not correctly estimated [30]. Other works assume that the blurring kernel is known, which makes the applicability of these solutions limited. Deblurring through depth map estimation has also been proposed, but this is rather a computationally heavy task [23, 35, 38]. For the kernel-free solutions, recovering the sharp image is often

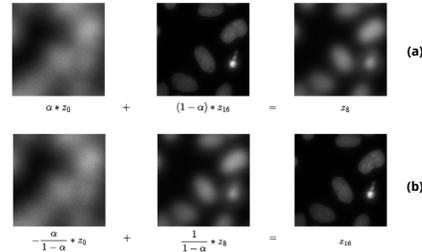


Fig. 1. Blurring (a) and deblurring (b) through interpolation in the latent space. The  $\alpha$  parameter controls the level of synthesized blur or the sharpness of the deblurred image. The  $z$ -notation indicates that images in the last column are obtained from interpolated latent representations, whereas the digits indicate the focal length of the picture. Smaller digits correspond to higher levels of defocus blur.

achieved by training complex networks from an architectural and computational standpoint. Such solutions usually involve the use of adversarial training [27] or multiscale networks [18, 34]. Finally, most papers focus on deblurring solutions applied to real-world data sets, while microscopy images are utilized in only a few studies [6, 13, 22].

We propose a generative model based on linear interpolation in the latent space, with twofold applicability. Firstly, it can be used as a data-augmentation technique to generate images with different blur levels. Images generated this way could be used to assess or enhance the robustness of certain models against noise induced by defocus blur. The blur is synthesized through linear interpolation between latent representations of images taken at different focal distances. Secondly, we employ the model for deblurring images by reversing the interpolation operation. We reconstruct an in-focus image through linear interpolation between the representations of two out-of-focus images, captured at different focal planes, as displayed in Figure 1. The main advantage of our model is its simple and cost-effective architecture. We bypass any complex operations such as depth map or blur kernel estimation and the training procedure is stable, compared to that of a GAN or VAE.

We train a model to synthesize blur through linear interpolation in the latent space. The embedded linear relationship between latent representations of images captured at different focal distances should enable the model to disentangle blur level from image content. The choice of using linear interpolation to perform blur synthesis has two advantages. Firstly, it enables the model to learn a reversible mapping from two out-of-focus images to a more in-focus image. Therefore, we can leverage the representations of out-of-focus images to recover the corresponding in-focus image. Secondly, this operation allows us to control the blur synthesis and blur removal through a single interpolation parameter,  $\alpha$ , ranging from 0 to 1.

TScIT 37, July 8, 2022, Enschede, The Netherlands

© 2022 University of Twente, Faculty of Electrical Engineering, Mathematics and Computer Science.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

The choice of  $\alpha$  should be fairly intuitive given two input images, and easy to fine-tune depending on the expected target image.

The results of the experiments prove the feasibility of recovering in-focus images through interpolation in the latent space. The proposed regularized autoencoders show improved blur synthesis ability compared to a vanilla autoencoder. One of the regularization methods increases the degree of linear dependence between latent representations of blur levels. Therefore, linear traversals in the latent space between two given points provide better results compared to the baseline in terms of replicating defocus blur and deblurring images.

To guide this study, we identify two research questions and two sub-questions.

**RQ1** To what extent can image representations interpolated in the latent space be decoded to generate images with intermediate blur levels?

**RQ1.1** How does a vanilla autoencoder optimized only for image reconstruction represent different blur levels in the latent space?

**RQ1.2** How can we impose a linear relationship on representations of different blur levels in the latent space of an autoencoder?

**RQ2** How suitable is the constrained latent space for recovering a sharp image through linear interpolation between the latent codes of images captured at different focal distances?

To answer these questions, we train three autoencoders with different types of regularization or no regularization. These models are described in detail in Section 3.

The paper is organized as follows. Section 2 gives an overview of papers on topics of interest to this work. Next, Section 3 describes the blurring and deblurring tasks, the proposed models and evaluation metrics. The data set, and training details are provided in Section 4 alongside the experiments results. In Section 5 we discuss the results and suggest further aspects that can be investigated to potentially improve the proposed solution. Finally, in Section 6 we outline the conclusions we draw from this research.

## 2 RELATED WORK

Literature relevant for this paper can be divided in three categories. Firstly, solutions for deblurring and super-resolution are discussed. Next, an overview of studies on regularizing the learned representations of autoencoders is provided. Finally, research on blurring techniques is reviewed.

### 2.1 Deblurring and Super-Resolution

Recent image deblurring solutions are based on convolutional neural networks and employ residual learning, multiscale training, GAN-based models or a combination of these. Quan et al. [30] propose a non-blind deblurring network based on a scale-recurrent attention module. In [34] and [1], the authors use multiscale U-net architectures with residual learning, for recovery from defocus blur and image super-resolution tasks. Residual learning is further employed by Pham et al. [29] and Mahapatra et al. [24] for the super-resolution task, and by Nimisha et al. [27] for image deblurring. All of these

papers rely on local or global residual connections, which are useful for recovering information that may be lost through down sampling, as well as for optimizing the training process for the models [25]. The authors of [15] propose a defocus map estimation model and assess its performance by using the predicted maps for enhancing blur in images and for recovering a sharp image from a blurred input. The defocus map is used to compute the pixel-wise blur level and magnify it by a given factor, while for the deblurring task they use the defocus map to estimate a per-pixel Gaussian kernel and perform a hyper-Laplacian deconvolution. They do not quantitatively measure the performance of blurring and deblurring images using the estimated defocus maps they provide, so it is not clear how well this method works and whether it is suitable for microscopic images as well.

Jiang et al. [11] and Zhang et al. [37] tackle multi-cause blur and propose deep-learning models that can recover sharp images affected by either motion or defocus blur. Both solutions have the disadvantage of being too slow for real-time usage, predicting the sharp image in 1.5 and 1.67 seconds, respectively.

Other papers take a different approach by processing the image at a patch-level, to classify images based on blur quality [32, 36] or to remove blur from images [6, 18]. Their motivation is based on the idea that the blur level is invariant throughout the image and that it is well-defined only in the foreground, which leads to a patch-level approach having a more powerful blur-level assessment and blur removal quality. In [16], a new type of filter is proposed to handle spatially-varying and large defocus blur, based on stacks of pixel-wise deblurring filters. The pixel-wise approach handles spatially variant blur, while stacks of small, separable filters aid recovery from large defocus levels. Despite showing very good results, their model is sensible to bright patches which are usually present in microscopy images.

### 2.2 Regularized autoencoders

Vanilla autoencoders are prone to model a latent space which does not reflect the real distribution of the data and cannot capture the relationship between data points [4]. Several solutions based on adversarial training of plain autoencoders have emerged. Oring et al. [28] propose a regularized version of an autoencoder, based on several loss terms used to impose smoothness and convexity on the latent space distribution. Similarly to our method, they perform interpolation between latent representations of the same object, captured at different rotation angles. Hence, they generate a smooth transition from one position to another by traversing the latent space between representations of the same object at different rotation angles. A shortcoming of their study is that they only test the proposed method on a simple dataset of rotating objects, so it is unknown whether the approach performs well on more complex data distributions. In [31], two autoencoders are used, one as a generator and one as a discriminator, where the latter is used to train the generator in an adversarial manner to reconstruct realistic images, which are pixel-wise close to the original input. Other papers perform adversarial training on VAEs to ensure that generated images are realistic. The authors of [5] use a discriminator network which predicts the probability that a generated image belongs to

the original manifold. In [2] and [10] a slightly different approach is employed. A critic network is used, which predicts the interpolation parameter corresponding to a generated image. The generator must minimize the output of the critic, thus ensuring that generated images are realistic.

### 2.3 Generating blur

Defocus blur generation techniques can be split in two categories: solutions that involve a deblurring step prior to the blur generation and direct image-to-image blurring solutions. The former approach is applied by the authors of [16] and [26], who use a reblurring loss to train their image deblurring network. Lee et al. [16] achieve reblurring by reverting a deblurring function. However, this method is clearly dependent on a deblurring function, which requires additional steps to be computed and may affect blurring performance if not correctly estimated. The approach of Nah et al. [26] is to recover the initial blur kernel from the remaining kernel, after a deblurring operation was performed. For direct image-to-image blurring, the most frequently used solution is to perform a convolution operation on the input image, using a Gaussian filter. This process results in smoother edges and surfaces, and, depending on the chosen kernel size and  $\sigma$  parameter, can generate different levels of blur. The kernel size controls the extent of smoothing, as each pixel is blurred using a rectangular kernel, and larger kernels lead to more smoothing. The  $\sigma$  parameter is the variance of the Gaussian distribution from which the elements of the kernel are sampled. It also has an influence on the extent to which high frequency features around the pixel are reduced. Gaussian kernels are employed in [15], which synthesize defocus blur in images based on a discrete defocus map. This approach is useful for images with complex scenes, where object positioning and the associated depth map are important.

## 3 METHODOLOGY

### 3.1 Task description

*3.1.1 Mimic defocus blur.* We want to impose a linear dependence relation between latent representations of images with different focal lengths. To achieve this, we train a model to synthesize blur through linear interpolation in the latent space. The goal is to reconstruct an image which has a defocus level intermediate to those of the two input images. Let  $x_1$  and  $x_2$  be two images of a microscopic slide captured at different focal distances, where  $x_2$  has a higher blur level compared to  $x_1$ . Additionally, let  $z_1$  and  $z_2$  be the latent codes of  $x_1$  and  $x_2$ . We define the reconstructed interpolated image as:

$$\tilde{x}'_3 = G(z'_3) \quad (1)$$

where

$$z'_3 = \alpha z_1 + (1 - \alpha) z_2 \quad (2)$$

In Equation 1,  $G$  is a decoder, trained to reconstruct the input images encoded by network  $E$  or resulted from interpolations in the latent space of  $E$ . In Equation 2,  $z'_3$  is a latent representation obtained from interpolation and  $\alpha$  is the interpolation parameter which controls the level of blur in the interpolated image based on the input images as follows: if  $\alpha=0$ , the interpolated image is simply  $x_2$ , if  $\alpha=1$  the interpolated image is  $x_1$  and as  $\alpha$  increases from 0 to 1, the blur level of the generated images will decrease. Note that the notations  $z_3$  and

$\tilde{x}_3$  are reserved for the image representation and reconstruction of the ground-truth image  $x_3$ . The corresponding interpolated image representation of  $x_3$  and its reconstruction are denoted as  $z'_3$  and  $\tilde{x}'_3$ .

*3.1.2 Deblurring through linear interpolation in the latent space.* The latent space of a model trained on the blur synthesis task is used to perform deblurring through linear interpolation in the latent space. The goal is to reconstruct the sharp image corresponding to z-stack 16, starting from the most out-of-focus image with z-stack 0 and using a second image, in the range from z-stack 2 to z-stack 14. The linear interpolation for this operation is shown in Equation 3, where  $\alpha$  is the interpolation parameter used in the blurring operation described in Equation 2,  $z'_3$  and  $z_1$  are two latent representations of blurred images and  $z_2$  is the interpolated latent representation corresponding to the sharp image with z-stack 16.

$$z_2 = \frac{1}{1 - \alpha} z'_3 - \frac{\alpha}{1 - \alpha} z_1 \quad (3)$$

## 3.2 Proposed models

*3.2.1 Baseline model.* We use a vanilla AE trained for image reconstruction as the baseline for our experiments. Using this type of model allows us to investigate how well blur can be disentangled from image content in the latent space, without explicitly optimizing the model for this task. We expect that the image representations will not be suitable for synthesizing blur through linear interpolation, since the model can learn arbitrary representations which minimize the loss function without learning factors of variation in the data.

*3.2.2 Regularized models.* We propose a modified optimization objective for the autoencoder such that the latent representations reflect the variation in blur level. In this way, transitions from in-focus images to out-of-focus images are represented in the latent space as linearly dependent vectors. We introduce two differently regularized models.

*Weakly-regularized AE.* The first model is an autoencoder trained for the image reconstruction task, with a weak form of regularization in the latent space. The model receives as input three images of one slide, captured at different focal planes. Its training objective is to minimize the loss function shown in Equation 4. The first term ensures pixel-wise similarity between the reconstructions and the input images with the lowest and highest z-stack levels. The third term is the weak-regularization used to ensure a pixel-wise similar reconstruction of the interpolated image  $\tilde{x}'_3$  and its corresponding ground-truth  $x_3$ , the third input image with median z-stack. This last term is a form of weak regularization since it enforces reconstructed images to be visually similar to their ground-truth versions, but it does not impose any constraints on how the latent space is modeled. We expect this weak form of regularization to help the model enforce a linear relation between the latent representations of the inputs. Pixel-wise similarity between two images is computed using the L1 loss function.

$$L_{weak} = \frac{1}{2}((x_1 - \tilde{x}_1) + (x_2 - \tilde{x}_2)) + (x_3 - \tilde{x}'_3) \quad (4)$$

*Strongly-regularized AE.* The last model we investigate employs a stronger form of regularization and is a fine-tuned version of the weakly-regularized model. Essentially, we want to evaluate the added value of using a regularization term directly in the latent space of the autoencoder. The strong regularization term adds a constraint that the distance between the linearly interpolated image representation and the ground-truth image representation must be minimized. The new loss function is shown in Equation 5. The choice for this constraint directly on the latent representations is motivated by the idea that the latent vectors capture high level features which have a more significant impact on the reconstruction than minimization of a per-pixel loss at the whole image level.

$$L_{strong} = L_{weak} + (\tilde{z}_3 - \tilde{z}'_3) \quad (5)$$

### 3.3 Evaluation metrics

The goal of this study is to develop a model that allows blur synthesis and deblurring through linear interpolation in the latent space. There are three aspects that must be evaluated, namely geometric properties of the latent space, model ability to synthesize blur and ability to deblur images.

*3.3.1 Assessing linear dependence in the latent space.* We employ both visual inspection and metric-based evaluation to assess the degree of linear dependence between latent representations of blur levels. We quantify the degree of linear dependence between image representations based on two geometric properties. Firstly, the latent vectors should have the same direction. This can be measured using a cosine-similarity based metric. We call this the Linear Dependence Score (LDS) and its formula is shown in Equation 6, where  $M$  is the number of images in the test set,  $N$  is the number of representations obtained through linear interpolation between two fixed latent points. Finally,  $z'_n$  and  $z_n$  are the interpolated latent representation of an image and its corresponding ground truth latent representation.

$$LDS = \frac{1}{M} \sum_{i=1}^M \frac{1}{N} \sum_{n=1}^N \cos\_sim(z'_n, z_n) \quad (6)$$

Using linear interpolation to compute latent representations of images forces vectors belonging to a transition from in-focus to out-of-focus to be equally distanced from one another. The metric we use to measure this second property of the latent space structure is shown in Equation 7, and we call it the Average Pairwise Distance (APD), where  $M$  and  $N$  are the same as defined above and  $z_n$  and  $z_{n+1}$  are representations of two consecutive images in terms of blur level. Essentially, APD is a measure of the overlap between original and linearly interpolated image representations.

$$APD = \frac{1}{M} \sum_{i=1}^M \frac{1}{N+1} \sum_{n=0}^{N-1} |d(z_n, z_{n+1}) - d(z'_n, z'_{n+1})| \quad (7)$$

For visually evaluating the linear dependence, the latent representations modeled by the autoencoders must be inspected. This is achieved by mapping the latent representations to a 2D space. We use PCA, a dimensional reduction technique which preserves the linear relationship between data points [19].

*3.3.2 Blur synthesis.* The quality of the reproduced blur is quantified by comparing the reconstructions of original image representations and those of interpolated image representations using Peak Signal-to-Noise-Ratio (PSNR), the Fréchet Inception Distance (FID) [7], and the Focal Frequency Loss (FFL) proposed by Jiang et al. [12]. The scores are computed between generated images from linearly interpolated representations and the reconstruction of the corresponding ground-truth image. The PSNR score computed this way will be later referred to as PSNR-blur. Reconstruction fidelity against the original images is measured with PSNR, referred to later as *PSNR<sub>blur</sub> - quality*.

*3.3.3 Deblurring.* To assess the performance of the models on recovering from defocus blur, we use PSNR. We call this metric PSNR-deblur, and we compute it the same way as PSNR-blur, hence measuring the similarity between reconstructions of sharp images and the corresponding deblurred version obtained through interpolation in the latent space. Reconstruction fidelity against the original sharp images is measured with PSNR, referred to later as *PSNR<sub>deblur</sub> - quality*.

## 4 EXPERIMENTS AND RESULTS

### 4.1 Data set

The data set used to train and evaluate the proposed models is the BBBC006v1 collection obtained from the Broad Bioimage Benchmark Collection [20]. It consists of 768 in-focus images (z-stack 16), and for each in-focus image 16 versions taken below the focal plane (z-stack 0 to 15), and 17 images taken above the focal plane (z-stack 17 to 33). The image with z-stack 16 was captured at an optimal focal distance auto-detected by the microscope. For the experiments we conduct, only the images with even z-stack values from 0 to 16 are used. The images are 696x520 pixels.

It is important to note that the data set includes two different types of slides. The w1 slides stained with Hoechst 33342 markers represent nuclei, while the w2 slides stained with phalloidin represent the cell-structures. The experiments are conducted separately on the w1 and w2 slides, due to the difference in their underlying distributions. The data is split in a training, a validation and a testing set, in a 70:10:20 ratio. The images are ordered alphabetically before the split, with the order of the splits being test-validation-train.

Since the model should learn to map representations of images taken at different focal distances as linearly dependent, we use triplets spanning the range from z-stack 16 to z-stack 0. This way, the autoencoder sees all these blur levels, either as inputs or as interpolated representations. We set  $\alpha = 0.5$  for the training phase

Table 1. Blur synthesis (FFL, FID, PSNR-blur), deblurring (PSNR-deblur), and image quality ( $PSNR_{blur} - quality$ ,  $PSNR_{deblur} - quality$ ) results. The arrows indicate whether a lower or a higher score is desirable. The best scores are highlighted.

		W1 slides			W2 slides		
		Baseline	Weak	Strong	Baseline	Weak	Strong
Blur synthesis	FFL ( $10^{-4}$ )↓	6.48	5.88	<b>5.27</b>	4.70	4.33	<b>4.31</b>
	FID ↓	11.07	6.58	<b>5.99</b>	13.96	11.70	<b>11.17</b>
	PSNR-blur ↑	31.22	31.36	<b>31.95</b>	33.06	33.43	<b>33.57</b>
Image reconstruction quality	$PSNR_{blur} - quality$ ↑	<b>29.01</b>	26.98	27.81	<b>28.39</b>	28.22	27.96
	$PSNR_{deblur} - quality$ ↑	23.60	<b>24.56</b>	23.83	21.86	<b>21.88</b>	21.68
Deblurring	PSNR-deblur ↑	22.07	<b>22.48</b>	22.22	24.12	24.41	<b>24.72</b>

Table 2. We quantify the linear dependence of the image representations with LDS, a cosine similarity-based metric which measures how close to pointing in the same direction the latent vectors are and APD, which is a measure of the overlap between original image representations and their corresponding interpolated representations. The arrows indicate whether a lower or a higher score is desirable. The best scores are highlighted.

		W1 slides			W2 slides		
		Baseline	Weak	Strong	Baseline	Weak	Strong
LDS ↑		0.89	0.89	<b>0.95</b>	0.97	0.96	<b>0.98</b>
APD ↓		0.02	0.02	<b>0.0003</b>	0.005	0.005	<b>0.0002</b>

and use the following triplets in the format (z-stack\_left, z-stack\_target, z-stack\_right): (0, 8, 16), (0, 4, 8), (8, 12, 16), (0, 2, 4), (4, 6, 8), (8, 10, 12), (12, 14, 16).

## 4.2 Training and architecture

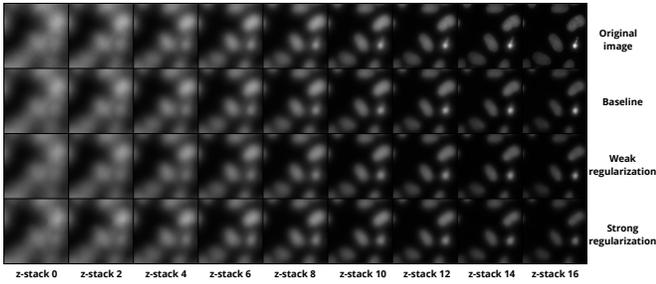
We use a convolutional autoencoder with 5 layers in the down-sampling part and 6 layers in the up-sampling part. Each layer in the encoder consists of a two-strided convolution with a kernel size of 3, batch normalization and Leaky ReLU activation. In the decoder part, the layers are symmetrical, with transposed convolution replacing convolution operations. The last layer before the output is a convolution with kernel-size 3, followed by a Sigmoid activation to ensure the resulting image pixels are within the range [0, 1]. The encoder layers have 64, 128, 256, 512 and 1024 output filters.

The models are trained for 40 epochs, with a batch size of 40. Adam optimizer is used with a learning rate of 0.0001 and default values for its other parameters. Additionally, a step scheduler is used to half the learning rate every 15 epochs. Originally the images are 696×520 pixels, in 16 bit TIFF format with LZW compression. We convert them to an 8 bit format as a pre-processing step and augment the dataset by generating from each image 10 crops of size 128×128 (4 corner crops, 1 center crop and their corresponding horizontally flipped versions). The code is written in PyTorch 1.11.0+cu113 and makes use of the PyTorch Lightning framework. The hardware used to train the models consists of two Tesla T4 GPUs.

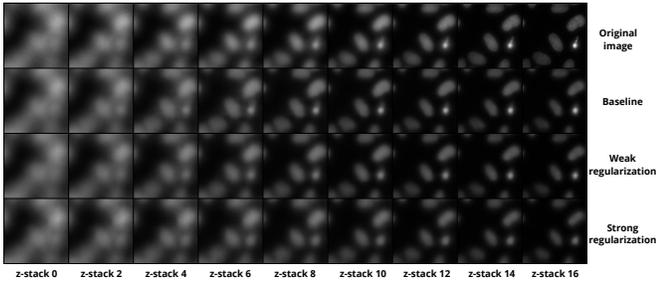
## 4.3 Results

We conduct a total of 6 experiments, 3 for each of the w1 and w2 slides image sets. Each experiment uses one of the proposed models, the baseline and two regularized versions of it, and they are trained to synthesize defocus blur in microscopic images, as described in Section 3.1.1. We evaluate the performance of these models on defocus blur synthesis and image deblurring.

**4.3.1 Comparative analysis of the learned latent space.** The geometric properties of the latent space learned in the different experiments are assessed through LDS and APD, described in detail in Section 3.3.1. These results are displayed in Table 2. To understand the effect of the regularization on the latent space distribution, we visualize 2D projections of image representations as shown in Figure 5. In line with expectations, the vanilla AE baseline does not model different blur levels as linearly dependent in the latent space. There is however a substantial difference between the w1 and w2 slides, the latter achieving a higher degree of linear dependence of 0.97 compared to 0.89 as measured by LDS, and of 0.005 compared to 0.02 as measured by APD. The weak regularization has no effect on the geometric properties of the latent space, as the reported scores indicate. The stronger form of regularization, however, leads to substantial changes in the distribution of the latent representations. For the w1 slides, both the direction of the latent vectors, as measured by LDS and the level of overlap between interpolated and original image representations, as measured by APD improve. LDS increases from 0.89 with the baseline model to 0.95 with the strong



(a) A transition from z-stack 0 to z-stack 16. In this order: original images, original image reconstructions with the baseline, weakly-regularized, and strongly-regularized models.

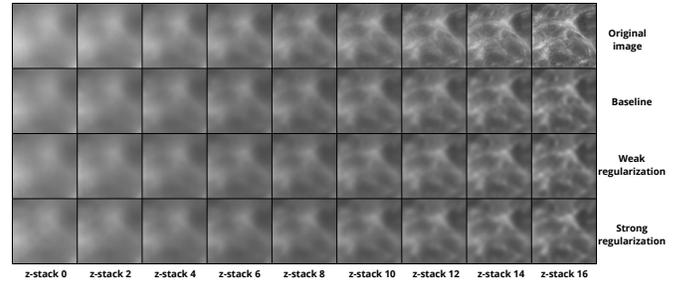


(b) A transition from z-stack 0 to z-stack 16. In this order: original images, reconstructions of interpolated image representations using the baseline, weakly-regularized, and strongly-regularized models.

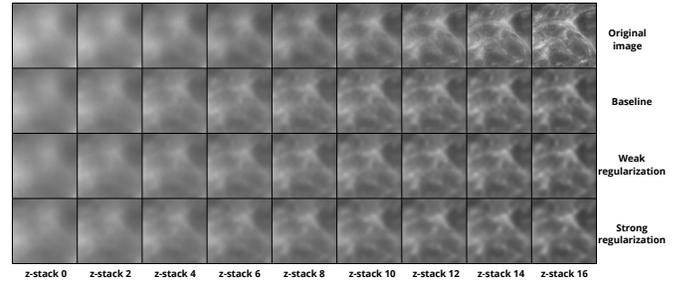
Fig. 2. Results of original image reconstructions (a) compared to synthetic blur (b) for w1 slides. Images from z-stack 2 to z-stack 14 are obtained as  $\alpha z_{16} + (1 - \alpha)z_0$ , with  $\alpha$  increasing from 0.125 to 0.875, from left to right.

regularization, while APD decreases from 0.02 to 0.0003. For the w2 slides, the improvement in latent vector orientation is marginal, from 0.97 to 0.98, but the APD score improves from 0.005 to 0.0002.

**4.3.2 Blur synthesis.** In Table 1, FID, FFL and PSNR-blur are used to measure the visual quality of the replicated blur. Visualizations of blur synthesized through linear interpolation, as well as reconstructions of the corresponding ground-truth images, are shown in Figures 2 and 3. For the w1 experiments with the baseline model, reconstructions from linear traversals of the latent space between two points result in semantically meaningful images in terms of image content. However, as shown in Figure 4, a blending effect between the two input images is visible, rather than a defocus blur effect. For the w2 slides, it is harder to tell whether blur is properly mimicked, given the more complex appearance of images. Nonetheless, quantitative results indicate that the baseline underperforms compared to the regularized models, for both w1 and w2 slides. The ability to synthesize realistic defocus blur improves with the addition of the weak regularization, which helps to reduce the blending effect. The strong regularization performs the best, as the latent vectors achieve a higher degree of linear dependence, so the transitions from sharp to in-focus are smoother, while the blur effect is still replicated. The scores in Table 1 confirm the results from visual inspection. The baseline underperforms compared to the regularized models, the



(a) A transition from z-stack 0 to z-stack 16. In this order: original images, original image reconstructions with the baseline, weakly-regularized, and strongly-regularized models.



(b) A transition from z-stack 0 to z-stack 16. In this order: original images, reconstructions of interpolated image representations using the baseline, weakly-regularized, and strongly-regularized models.

Fig. 3. Results of original image reconstructions (a) compared to synthetic blur (b) for w1 slides. Images from z-stack 2 to z-stack 14 are obtained as  $\alpha z_{16} + (1 - \alpha)z_0$ , with  $\alpha$  increasing from 0.125 to 0.875, from left to right.



Fig. 4. Blur effect to be replicated (Row 1) and blending effect resulting in reconstructions of interpolated image representations using the baseline model (Row 2)

weak regularization drastically reduced FID to 6.58 from 11.07 for the w1 slides, while PSNR-blur only slightly increases. The strong regularization achieves the highest scores, improving by 0.7dB (w1 slides) and 0.5dB (w2 slides) compared to the baseline.

**4.3.3 Deblurring.** The quality of the deblurred images is measured with PSNR-deblur. Figures 7 and 8 display examples of deblurred images using the proposed models, with two different values for  $\alpha$ . For each experiment, we assess the underlying model's performance on the deblurring task, according to the procedure described in Section 3.1.2. For the w1 slides, the PSNR-deblur value indicates

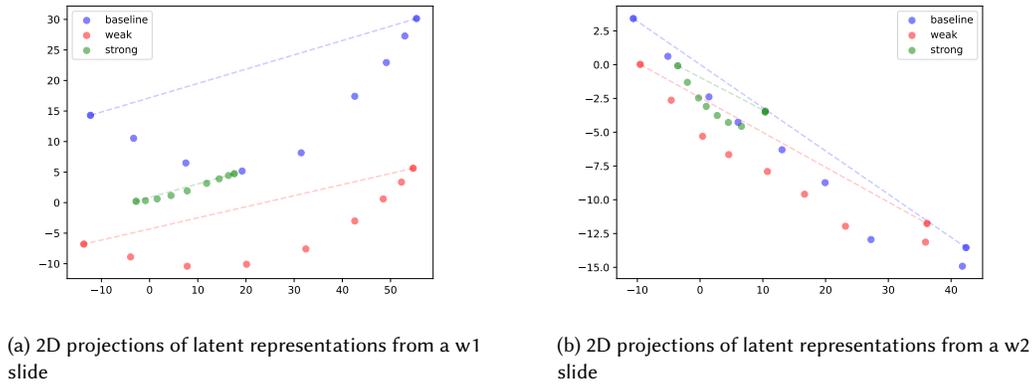


Fig. 5. 2D projections of latent representations obtained from the proposed models. Latent representations of the original images (dots), linear trajectory between z-stack 0 and z-stack 16 (dashed line).

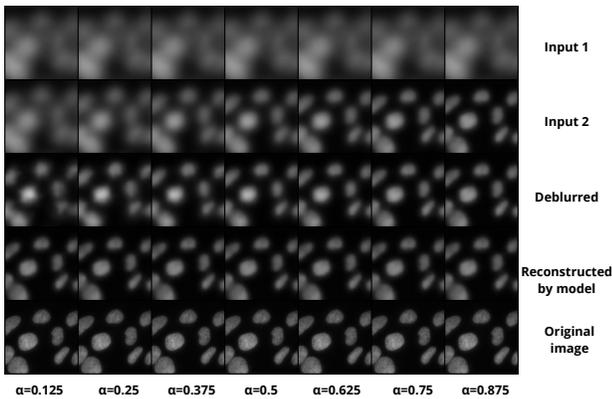


Fig. 6. Effect of the blur level in the input images used for deblurring on the recovered sharp image, when the optimal interpolation parameter  $\alpha$  is known. Results obtained with the weakly-regularized model.

that the regularized autoencoders outperform the baseline model, with the weak regularization achieving the best score.

It was observed that for images whose representations followed a curvilinear path in the latent space of the baseline model, the imposed linear dependence by the strong regularization caused the deblurred images to be affected by artifacts. When the blur levels were naturally mapped to more linearly dependent representations, the regularization helped enhance this relation, without trading off image quality. For the w2 slides, the strong regularization performs the best, with PSNR-deblur increasing by 0.6dB, while the overall image quality only slightly degrades by 0.18dB.

## 5 DISCUSSION AND FUTURE WORK

Overall, the experiments suggest the feasibility of blur synthesis and blur removal through interpolation in the latent space using the proposed regularized autoencoders. For blur generation, the baseline model is not suitable, as linear interpolations result in a blending

effect. In principle, this can be easily explained by the larger gap between the original latent representations of the images and the interpolated ones in the latent space of the baseline model. It is important to note that the ability to replicate blur of the regularized models is achieved at the cost of image reconstruction quality, and significant details are lost. Projections of the latent vectors obtained with the strong regularization show that the imposed linearity forces the representations to be more clustered. In turn, this may be one of the factors leading to degraded reconstructions when using this model, as highlighted in Figure 7b. Therefore, although the strong regularization achieves the desired geometric properties in the latent space, the reconstruction fidelity compared to the ground-truth is very low.

A similar effect is observed for the task of deblurring. For the w1 slides, the deblurred images using the strongly-regularized latent space are too degraded in some cases (see Figure 7b). For the w2 slides, the weak regularization maximizes image quality, while strong regularization outperforms in terms of deblurring. The weak regularization offers a better trade-off between the desired properties of images and latent space for the w2 slides.

Another interesting aspect applicable for both w1 and w2 slides is that even when the deblur operation uses two images with high levels of blur, the recovered details are still significant (see Figures 7a and 8a). Figure 6 shows how the choice of input images affects the deblurring process. We fix one image at z-stack 0 and vary the other one from z-stack 2 to z-stack 14. These results are in line with those from a similar study [22], where the level of detail recovered in the deblurred images decreases with an increase in the focal distance at which slides are captured.

With the baseline and weakly-regularized model trained on the w1 slides, we observe that some 2D projections of latent vectors follow a curvilinear trajectory, while others are very close to being linearly dependent. The curvature of these representations indicates that there may be two directions in the latent space, one corresponding to the variation in blur and one corresponding to the variation in image content. Therefore, from z-stack 0 to z-stack 8 or 10, blur is linearly represented in the latent space. From these z-stacks to z-stack

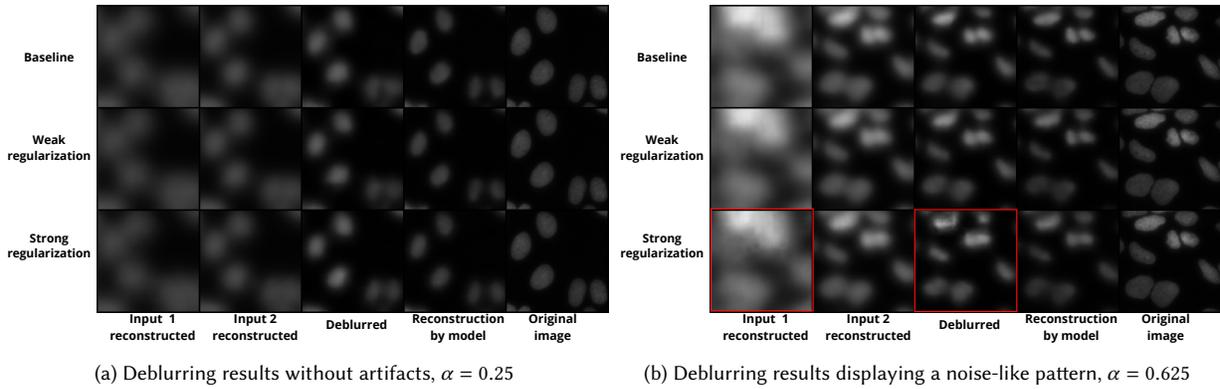


Fig. 7. Deblurring examples for w1 slides, where no artifacts appear (a) and where artifacts appear (b). From left to right: the input images used for deblurring reconstructed by each model, the deblurred image, the corresponding sharp image reconstructed by each model, and the ground-truth in-focus image.

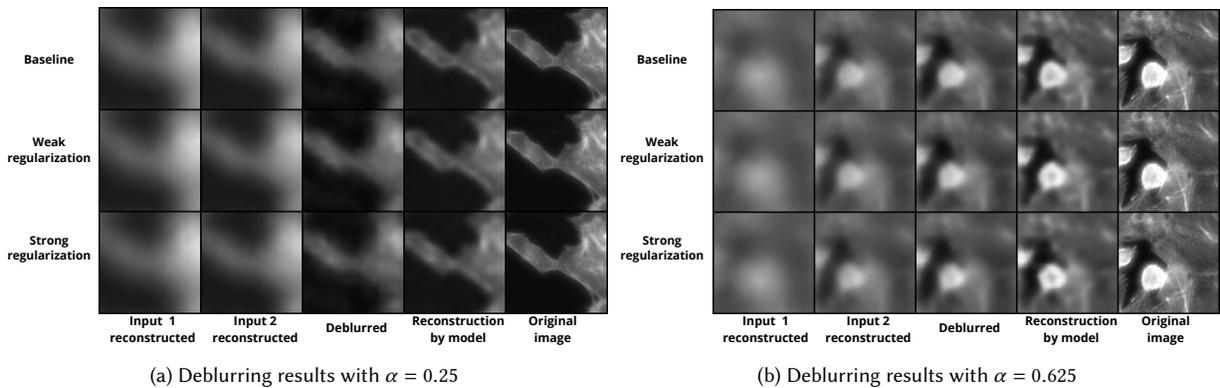


Fig. 8. Deblurring examples for w2 slides. From left to right: the input images used for deblurring reconstructed by each model, the deblurred image, the corresponding sharp image reconstructed by each model, and the ground-truth in-focus image.

16, the direction of the representations changes, and modifications in image content are captured by linearly dependent representations.

Due to time constraints, the architecture chosen for the experiments is fairly simple and leads to reconstructions with considerable loss of detail. This influences the blurring task slightly, but has a more considerable negative effect on the deblurring task. Therefore, it would be interesting to study the effects of the regularization using a more powerful architecture, such that image reconstruction quality will not affect the results to this extent.

Moreover, the experiments we conduct are limited to a set of images captured above the optimal focal distance, but the dataset we use provides images taken below the optimal focal plane as well. A potential future direction is to perform the interpolation between images taken both above and below this optimal point. This would reveal whether the different information contained in these images improves performance of the models on the deblurring task.

Lastly, the process of fine-tuning  $\alpha$  for deblurring should be assessed, to develop a better understanding of how the unknown relation between the blur intensities of the available images affects the recovery of a sharp image.

## 6 CONCLUSIONS

High quality images are crucial in the medical domain to ensure that noise, such as defocus blur, does not affect image analysis and medical diagnosis. In this paper, we investigate the feasibility of a generative model based on linear interpolation in the latent space, for defocus blur synthesis and recovery from defocus blur.

We conduct experiments to analyze how a linear relationship can be enforced on the latent representations learned by a convolutional vanilla autoencoder. We find that the regularized models which we investigate perform well on mimicking the defocus blur effect and recovering a sharp image from two blurred inputs. However, the regularization determines low reconstruction fidelity compared to the ground-truth images. The strong regularization performs best under most metrics, but visually, the weak regularization outperforms in some cases.

## ACKNOWLEDGMENTS

I would like to thank my supervisors, Nicola Strisciuglio and Shunxin Wang for the support, insights and feedback they provided during the course of this research.

## REFERENCES

- [1] Nicolas Basty and Vicente Grau. 2018. *Super Resolution of Cardiac Cine MRI Sequences Using Deep Learning*. 23–31. [https://doi.org/10.1007/978-3-030-00946-5\\_3](https://doi.org/10.1007/978-3-030-00946-5_3)
- [2] David Berthelot, Colin Raffel, Aurko Roy, and Ian J. Goodfellow. 2018. Understanding and Improving Interpolation in Autoencoders via an Adversarial Regularizer. *CoRR abs/1807.07543* (2018). <http://arxiv.org/abs/1807.07543>
- [3] Songkui Chen, Daming Shi, Muhammad Sadiq, and Xiaochun Cheng. 2020. Image Denoising With Generative Adversarial Networks and its Application to Cell Image Enhancement. *IEEE Access* 8 (2020), 82819–82831. <https://doi.org/10.1109/ACCESS.2020.2988284>
- [4] Marta Garnelo and Murray Shanahan. 2019. Reconciling deep learning with symbolic artificial intelligence: representing objects and relations. *Current Opinion in Behavioral Sciences* 29 (2019), 17–23. <https://doi.org/10.1016/j.cobeha.2018.12.010>
- [5] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative Adversarial Networks. <https://doi.org/10.48550/ARXIV.1406.2661>
- [6] Da He, De Cai, Jiasheng Zhou, Jiajia Luo, and Sung-Liang Chen. 2019. Adaptive Weighting Depth-variant Deconvolution of Fluorescence Microscopy Images with Convolutional Neural Network. <https://doi.org/10.48550/ARXIV.1907.03217>
- [7] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. 2017. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. In *Advances in Neural Information Processing Systems*, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), Vol. 30. Curran Associates, Inc. <https://proceedings.neurips.cc/paper/2017/file/8a1d694707eb0fefe65871369074926d-Paper.pdf>
- [8] Md Shakhawat Hossain, Toyama Nakamura, Fumikazu Kimura, Yukako Yagi, and Masahiro Yamaguchi. 2018. Practical image quality evaluation for whole slide imaging scanner. In *Biomedical Imaging and Sensing Conference*, Vol. 10711. International Society for Optics and Photonics, 203–206. <https://doi.org/10.1117/1.2316764>
- [9] Lejia Hu, Shuwen Hu, Wei Gong, and Ke Si. 2021. Image enhancement for fluorescence microscopy based on deep learning with prior knowledge of aberration. *Opt. Lett.* 46, 9 (May 2021), 2055–2058. <https://doi.org/10.1364/OL.418997>
- [10] Akshita Jha, Bhanukiran Vinzamuri, and Chandan K. Reddy. 2021. Fair Representation Learning using Interpolation Enabled Disentanglement. *CoRR abs/2108.00295* (2021). <https://arxiv.org/abs/2108.00295>
- [11] Cheng Jiang, Jun Liao, Pei Dong, Zhaoxuan Ma, De Cai, Guoan Zheng, Yueping Liu, Hong Bu, and Jianhua Yao. 2020. Blind deblurring for microscopic pathology images using deep learning networks. <https://doi.org/10.48550/ARXIV.2011.11879>
- [12] Liming Jiang, Bo Dai, Wayne Wu, and Chen Change Loy. 2020. Focal Frequency Loss for Generative Models. *CoRR abs/2012.12821* (2020). <https://doi.org/10.48550/ARXIV.2012.12821>
- [13] Shaowei Jiang, Jun Liao, Zichao Bian, Kaikai Guo, Yongbing Zhang, and Guoan Zheng. 2018. Transform- and multi-domain deep learning for single-frame rapid autofocusing in whole slide imaging. *Biomedical Optics Express* 9 (04 2018), 1601. <https://doi.org/10.1364/BOE.9.001601>
- [14] Neeta Kumar, Ruchika Gupta, and Sanjay Gupta. 2020. Whole Slide Imaging (WSI) in Pathology: Current Perspectives and Future Directions. *Journal of Digital Imaging* 33 (05 2020), 1034–1040. <https://doi.org/10.1007/s10278-020-00351-z>
- [15] Junyong Lee, Sungkil Lee, Sunghyun Cho, and Seungyong Lee. 2019. Deep Defocus Map Estimation Using Domain Adaptation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 12214–12222. <https://doi.org/10.1109/CVPR.2019.01250>
- [16] Junyong Lee, Hyeonsek Son, Jaesung Rim, Sunghyun Cho, and Seungyong Lee. 2021. Iterative Filter Adaptive Network for Single Image Defocus Deblurring. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. *CoRR abs/2108.13610*. <https://doi.org/10.1109/cvpr46437.2021.00207>
- [17] Sehyung Lee, Makiko Negishi, Hidetoshi Urakubo, Haruo Kasai, and Shin Ishii. 2020. Mu-net: Multi-scale U-net for two-photon microscopy image denoising and restoration. *Neural Networks* 125 (2020), 92–103. <https://doi.org/10.1016/j.neunet.2020.01.026>
- [18] Pengwei Liang, Junjun Jiang, Xianming Liu, and Jiayi Ma. 2021. BaMBNet: A Blur-aware Multi-branch Network for Defocus Deblurring. <https://doi.org/10.48550/ARXIV.2105.14766>
- [19] Yang Liu, Eunice Jun, Qisheng Li, and Jeffrey Heer. 2019. Latent Space Cartography: Visual Analysis of Vector Space Embeddings. *Computer Graphics Forum* 38 (2019).
- [20] Vejbjorn Ljosa, Katherine Sokolnicki, and Anne Carpenter. 2012. Annotated high-throughput microscopy image sets for validation. *Nature methods* 9 (06 2012), 637. <https://doi.org/10.1038/nmeth.2083>
- [21] Jean-Baptiste Lugagne, Haonan Lin, and Mary J. Dunlop. 2020. DeLTA: Automated cell segmentation, tracking, and lineage reconstruction using deep learning. *PLoS Computational Biology* 16 (04 2020), 1–18. <https://doi.org/10.1371/journal.pcbi.1007673>
- [22] Yilin Luo, Luzhe Huang, Yair Rivenson, and Aydogan Ozcan. 2021. Single-Shot Autofocusing of Microscopy Images Using Deep Learning. *ACS Photonics* 8, 2 (Jan 2021), 625–638. <https://doi.org/10.1021/acsp Photonics.0c01774>
- [23] Haoyu Ma, Shaojun Liu, Qingmin Liao, Juncheng Zhang, and Jing-Hao Xue. 2022. Defocus Image Deblurring Network With Defocus Map Estimation as Auxiliary Task. *IEEE Transactions on Image Processing* 31 (2022), 216–226. <https://doi.org/10.1109/TIP.2021.3127850>
- [24] Dwarikanath Mahapatra and Behzad Bozorgtabar. 2019. Progressive Generative Adversarial Networks for Medical Image Super resolution. *CoRR abs/1902.02144* (2019). <http://arxiv.org/abs/1902.02144>
- [25] Xiao-Jiao Mao, Chunhua Shen, and Yu-Bin Yang. 2016. Image Denoising Using Very Deep Fully Convolutional Encoder-Decoder Networks with Symmetric Skip Connections. *CoRR abs/1603.09056* (2016). <http://arxiv.org/abs/1603.09056>
- [26] Seungjun Nah, Sanghyun Son, Jaerin Lee, and Kyoung Mu Lee. 2021. Clean Images are Hard to Reblur: A New Clue for Deblurring. <https://doi.org/10.48550/ARXIV.2104.12665>
- [27] T M Nimisha, Akash Kumar Singh, and A N Rajagopalan. 2017. Blur-Invariant Deep Learning for Blind-Deblurring. In *2017 IEEE International Conference on Computer Vision (ICCV)*. 4762–4770. <https://doi.org/10.1109/ICCV.2017.509>
- [28] Alon Oring, Zohar Yakhini, and Yacov Hel-Or. 2021. Autoencoder Image Interpolation by Shaping the Latent Space. In *Proceedings of the 38th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 139)*. PMLR, 8281–8290. <https://proceedings.mlr.press/v139/oring21a.html>
- [29] Chi-Hieu Pham, Aurélien Ducournau, Ronan Fablet, and François Rousseau. 2017. Brain MRI super-resolution using deep 3D convolutional networks. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. 197–200. <https://doi.org/10.1109/ISBI.2017.7950500>
- [30] Yuhui Quan, Zicong Wu, and Hui Ji. 2021. Gaussian Kernel Mixture Network for Single Image Defocus Deblurring. *CoRR abs/2111.00454* (2021). <https://arxiv.org/abs/2111.00454>
- [31] Tim Sainburg, Marvin Thielk, Brad Theilman, Benjamin Migliori, and Timothy Gentner. 2018. Generative adversarial interpolative autoencoding: adversarial training on latent space interpolations encourage convex latent distributions. *CoRR abs/1807.06650* (2018). <http://arxiv.org/abs/1807.06650>
- [32] Caglar Senaras, M. Khalid Khan Niazi, Gerard Lozanski, and Metin N. Gurcan. 2018. DeepFocus: Detection of out-of-focus regions in whole slide digital images using deep learning. *PLoS ONE* 13 (10 2018), 1–13. <https://doi.org/10.1371/journal.pone.0205387>
- [33] Sergio Varela-Santos and Patricia Melin. 2021. A new modular neural network approach with fuzzy response integration for lung disease classification based on multiple objective feature optimization in chest X-ray images. *Expert Systems with Applications* 168 (2021), 114361. <https://doi.org/10.1016/j.eswa.2020.114361>
- [34] Jiahe Wang and Boran Han. 2022. Defocus Deblur Microscopy via feature interactive coarse-to-fine network. <https://doi.org/10.48550/ARXIV.2201.02876>
- [35] Shumian Xin, Neal Wadhwa, Tianfan Xue, Jonathan T. Barron, Pratul P. Srinivasan, Jiawen Chen, Ioannis Gkioulekas, and Rahul Garg. 2021. Defocus Map Estimation and Deblurring From a Single Dual-Pixel Image. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 2228–2238. <https://doi.org/10.1109/ICCV48922.2021.00223>
- [36] Samuel J. Yang, Marc Berndl, D. Ando, Mariya Barch, Arunachalam Narayanaswamy, Eric Christiansen, Stephan Hoyer, Chris Roat, Jane Hung, Curtis Rueden, Asim Shankar, Steven Finkbeiner, and Philip Nelson. 2018. Assessing microscope image focus quality with deep learning. *BMC Bioinformatics* 19 (03 2018). <https://doi.org/10.1186/s12859-018-2087-4>
- [37] Kaihao Zhang, Wenhan Luo, Boheng Chen, Wenqi Ren, Björn Stenger, Wei Liu, Hongdong Li, and Ming-Hsuan Yang. 2021. Benchmarking Deep Deblurring Algorithms: A Large-Scale Multi-Cause Dataset and A New Baseline Model. *CoRR abs/2112.00234* (2021). <https://arxiv.org/abs/2112.00234>
- [38] Xinxin Zhang, Ronggang Wang, Xiubao Jiang, Wenmin Wang, and Wen Gao. 2016. Spatially variant defocus blur map estimation and deblurring from a single image. *Journal of Visual Communication and Image Representation* 35 (2016), 257–264. <https://doi.org/10.1016/j.jvcir.2016.01.002>