

# VISUAL ENHANCEMENT OF IMAGE STACKS FROM A Z-SCANNING MICROSCOPE USING FOCAL LENGTHS WEIGHTED GANS

## ABSTRACT

In order to perform the z-stack imaging of a 1-micron-thick human brain slice without blurring and to improve the z-resolution, we propose an approach to train a 3D convolutional neural network to enhance the z-stacks' visual performance without 3D ground truth images. We implement two 3D generators and two 2D discriminators and formulate the adversarial and the consistency losses with distinct weights along the stacked direction according to the focal lengths. Our network learns to perform an intensity transformation from image stacks to volumetric 3D patches, in which the cross sections are indistinguishable from each other and have an image sharpness similar to that of the sharp image from the focal plane.

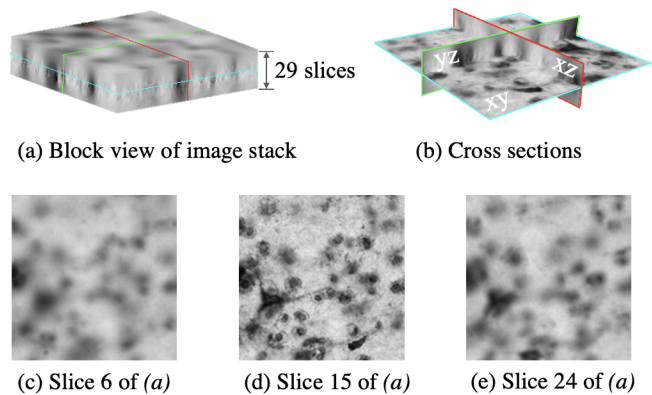
**Index Terms**— Deblurring, z-scan, light microscopy, weighted gans, weighted adversarial loss

## 1. INTRODUCTION

Measuring the 3D distribution of neuronal cells across the entire human brain is an important basis for understanding the architecture and function of this most complex organ. The development of technologies that would allow us to build a whole brain human atlas at the level of individual cells is an active field of research. Among the promising imaging methods are tissue clearing [1, 2, 3], light sheet imaging [4], x-ray tomography [5], optical coherence tomography [6] and 3D reconstruction from histology [7, 8, 9].

3D reconstruction from stained whole-brain histological sections as in [7] has the advantage of building on well-established and understood methods. To reach the single cell resolution, however, extremely precise alignment of consecutive sections [10] is required, as well as some form of sub-sampling to increase the z resolution, which is limited by the minimum section thickness. Modern whole slide scanners allow z-stack scanning at high-throughput, delivering for each individual physical tissue section a series of high-resolution 2D images. These images represent the gradual optical depth by focusing through the section. These stacks allow appreciation of the 3D shape and z location of each cell body. However, the z direction suffers from typical blurring effects, as depicted in Fig. 1 (b).

An established approach to reduce blurring in such z-stacks is 3D deconvolution with an appropriate kernel func-



**Fig. 1:** Small region of interest (Roi) extracted from a real image stack: (a): block view of a stack, which is made up by 29 slices; (b): cross-sectional view with two vertical cross-sections (xz, yz) and one horizontal section (xy); ((c),(d),(e): slices 6,15 and 24 of image stack.

tion, where the imaging processing is modelled by 3D convolution of the attenuation distribution of the tissue with the 3D point spread function (PSF) of the microscope. Given an appropriate model and parameters for the PSF, deconvolution algorithms can recover an estimate of the original attenuation distribution, i.e., of the desired sharp image. Unfortunately, the estimation of a suitable PSF is non-trivial - whether using experimental, numerical or blind estimation methods-, and often, the results are not fully satisfactory, especially in case of thin samples in a light microscope.

In this work, we look into the feasibility of a different approach to the deblurring of z-stacks, which uses Generative Adversarial Networks (GANs) to learn a visual enhancement directly from the data in an unsupervised fashion, and without specific modelling of the imaging physics. Motivated by the recent success of CycleGANs [11] in the translation of images between different classes of texture distributions, we aim to train a 3D convolutional neural network to visually enhance the z-stacks without access to ground truth images. Our approach differs from those employed in conventional GANs in that we formulate the loss and perform the training based on 2D image patches, while the network learns to perform an intensity transformation of volumetric 3D patches.

The main idea is as follows: We exploit the fact that the blurriness of the z-stacks is mostly observed in vertical *cross-*

sections (2D image patches parallel to the z axis; see Fig. 3) as well as in the *horizontal sections* (2D image patches parallel to the x/y plane) that are close to the top or bottom of the stack, demonstrated in Fig. 1 (c) and Fig. 1. The 2D image patches from horizontal planes in the centre of the stack are usually sharp, cf. Fig. 1 (d). We thus formulate the task of image enhancement as follows:

1. Modify image intensities of the z-stack such that the textural properties of the vertical cross-sections and horizontal sections near the top and bottom of the output z-stack are indistinguishable from those of the horizontal sections close to the centre of the input stack.
2. Leave the cellular structures overall consistent, i.e., i) do not introduce new cellular structures, ii) do not suppress existing cellular structures, and iii) leave existing structures centered at the same location in the stack.

We present a novel network architecture and training strategy for the visual enhancement of z-stacks (ZSE-GAN) that implements the above idea, and we investigate the extent to which it succeeds at the above tasks.

## 2. MATERIALS AND METHODS

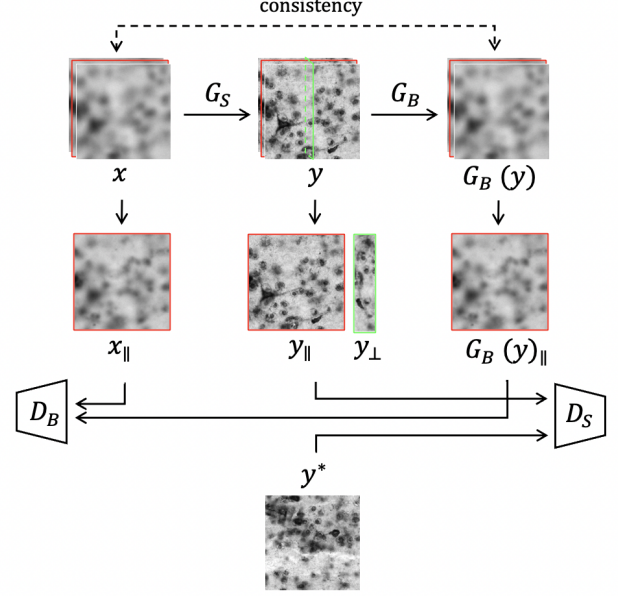
### 2.1. Datasets

For this project, we z-scanned of a whole section of human brain, acquired using the *Huron Tissuescope LE120* device (Huron Technologies International Inc.<sup>1</sup>). We extracted 2000 regions of interest from one brain section scan, each representing a z-scan stack with 29 optical planes sampled in a local region of interest, and represented as a volumetric image patch. The dimensions of each volumetric patch were  $128 \times 128 \times 29$  voxels, with an isotropic spatial resolution of 1 micron. When the microscope’s camera scans the brain section with the proper depth of focus, in-focus images can be obtained. We cropped 2000 image patches with dimensions  $128 \times 128$  from the 2D in-focus image. The validation setup is based on a random split of the volumetric images dataset and in-focus images datasets in 80% for training and 20% for testing.

### 2.2. Method

Our goal is to translate an image stack sample  $x$  in set  $X$  to the desired sharp volumetric image  $y$  in set  $Y$ . Set  $X$  refers to an image set of z-scan stacks. Each stack is a series of images scanned perpendicular to the brain slide. Set  $Y$  is a volumetric images set of translated  $x$  from  $X$ , in which the cross-sections of  $y$  have a sharpness similar to that of the in-focus image. An intersection of a plane  $\mathbb{P}$  and a stack  $x$  from  $X$  is denoted as cross section  $x_{\mathbb{P}}$ . The horizontal cross-sectional image  $x_{\parallel}$  or vertical cross-sectional image  $x_{\perp}$  is obtained if  $\mathbb{P}$  is parallel or perpendicular to the brain slide, respectively.

<sup>1</sup><https://hurondigitalpathology.com>



**Fig. 2:** ZSE-GAN architecture. There are two 3D generators  $G_S$  and  $G_B$ , which are associated with 2D adversarial discriminators  $D_S$  and  $D_B$  in the networks.

Our model includes two 3D translators  $G_S : X \rightarrow Y$  and  $G_B : Y \rightarrow X$ , as well as two 2D discriminators,  $D_S$  and  $D_B$ . For  $x \in X$  and  $y \in Y$ ,  $D_B$  aims to differentiate between images  $x_{\parallel}$  and  $G_B(y)_{\parallel}$ , whereas  $D_S$  works analogously for the images  $y_{\parallel}$ ,  $y_{\perp}$  and image  $y^*$ , where  $y^*$  is drawn from the 2D sharp images scanned from the focal plane, while  $y_{\parallel}$ ,  $y_{\perp}$  and  $G_B(y)_{\parallel}$  denote the the horizontal and vertical cross-sections of  $y$  and horizontal cross-section of  $G_B(y)$ , respectively. The architecture of our network is illustrated in Fig. 2.

#### 2.2.1. Anisotropic Adversarial Loss

For the mapping  $G_S : X \rightarrow Y$  and its corresponding discriminator  $D_S$ , we use the objective function

$$\mathcal{L}_{\text{GAN}}(G_S, D_S, X, Y) = \mathbb{E}_{D_S} + \mathbb{E}_{G_S}, \quad (1)$$

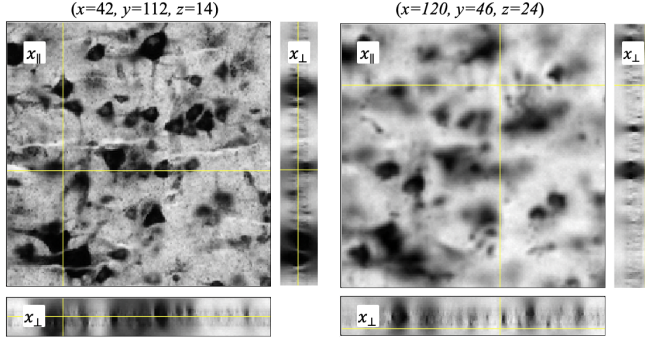
in which

$$\mathbb{E}_{D_S} = \mathbb{E}[\log(D_S(y_{\parallel})) \cdot D_S(y_{\perp})^p] \quad (2)$$

$$\mathbb{E}_{G_S} = \mathbb{E}[\log((1 - D_S(y_{\parallel})) \cdot (1 - D_S(y_{\perp}))^p)], \quad (3)$$

where  $G_S$  tries to generate isotropic volume  $y := G_S(x)$  whose cross sections  $y_{\mathbb{P}}$  are indistinguishable from the in-focus images  $y^*$ , while  $D_S$  tries to distinguish  $y_{\mathbb{P}}$  from  $y^*$ . As shown in Fig. 3, different vertical cross sections appear to have similar blurriness and distortion. Compared to the horizontal cross-section, the vertical cross sections not only are blurry but also lack cell structures. Instead of cell structures,

there are some dark spots in the vertical cross-section. Thus, we introduced a weight  $p$ , in order to reinforce a certain amount of learning to optimize the vertical cross-sections.



**Fig. 3:** Orthogonal view of an image stack with dimension  $200 \times 200 \times 29$ .

Mapping  $G_B : Y \rightarrow X$  is a convolutional blur generator, which reconstructs the translated image  $y$  back to the image stack  $x$ . Discriminator  $D_B$  differentiates between  $x_{\parallel}$  and  $G_B(y)_{\parallel}$ . We have a similar objective function as (1):

$$\mathcal{L}_{\text{GAN}}(G_B, D_B, X, Y) = \mathbb{E}_{D_B} + \mathbb{E}_{G_B} \quad (4)$$

but with

$$\mathbb{E}_{D_B} = \mathbb{E}[\log D_B(G_B(y)_{\parallel})^{\alpha(x_{\parallel})}] \quad (5)$$

$$\mathbb{E}_{G_B} = \mathbb{E}[\log(1 - D_B(G_B(y)_{\parallel})^{\alpha(x_{\parallel})})]. \quad (6)$$

Blur generator  $G_B$  translates the isotropic sharp 3D image  $y$  back to image  $x$ . Since image stack  $x$  is formed by 29 discrete scanned images,  $x_{\perp}$  is not smooth; however, the desired  $y$  is smooth (homogeneous) for any arbitrary  $y_{\perp}$ . For this reason, we trained  $G_B$  to only satisfy the condition that  $G_B(y)_{\parallel}$  is indistinguishable from the corresponding  $x_{\parallel}$ .

The sharpness of the slices which make up the image stack in the set  $X$  gradually increases from the top slice to the slices near to the focal plane and decrease from these slices to the bottom slice. Therefore, we defined a predicted coefficient function  $\alpha(x_{\parallel})$  to represent the sharpness level of the slice (image) in the stack that feeds to the discriminator  $D_B$ , i.e.,  $D_B$  regards the images near the centre of image stack or image volume as less blurry than the outer ones.

### 2.2.2. Cycle Consistency Loss

The optimal mapping between  $X$  and  $Y$  is theoretically bijective, i.e., the features in the translated image  $y$  match the features in the image stack  $x$ . We modified this idea to use a consistency loss. For each  $x \in X$  and  $y \in Y$ , the consistency satisfies the following cycle:

$$x \approx G_B \circ G_S(x). \quad (7)$$

We, therefore, adopt the consistency loss as follows:

$$\mathcal{L}_{\text{CON}}(G_S, G_B) = \mathbb{E}[\|x - G_B \circ G_S(x)\|_1]. \quad (8)$$

In the end, we train our generators under the supervision of a combination of adversarial and consistency losses:

$$\begin{aligned} &\mathcal{L}(G_S, G_B, D_S, D_B) \\ &= \mathcal{L}_{\text{GAN}}(G_S, D_S) + \mathcal{L}_{\text{GAN}}(G_B, D_B) + \lambda \mathcal{L}_{\text{CON}}(G_S, G_B), \end{aligned} \quad (9)$$

in which,  $\lambda$  is a parameter of consistency against the adversarial training. The final objective function is defined as follows:

$$G_S^*, G_B^* = \arg \min_{G_S, G_B} \max_{D_S, D_B} \mathcal{L}(G_S, G_B, D_S, D_B). \quad (10)$$

## 3. EXPERIMENTS AND RESULTS

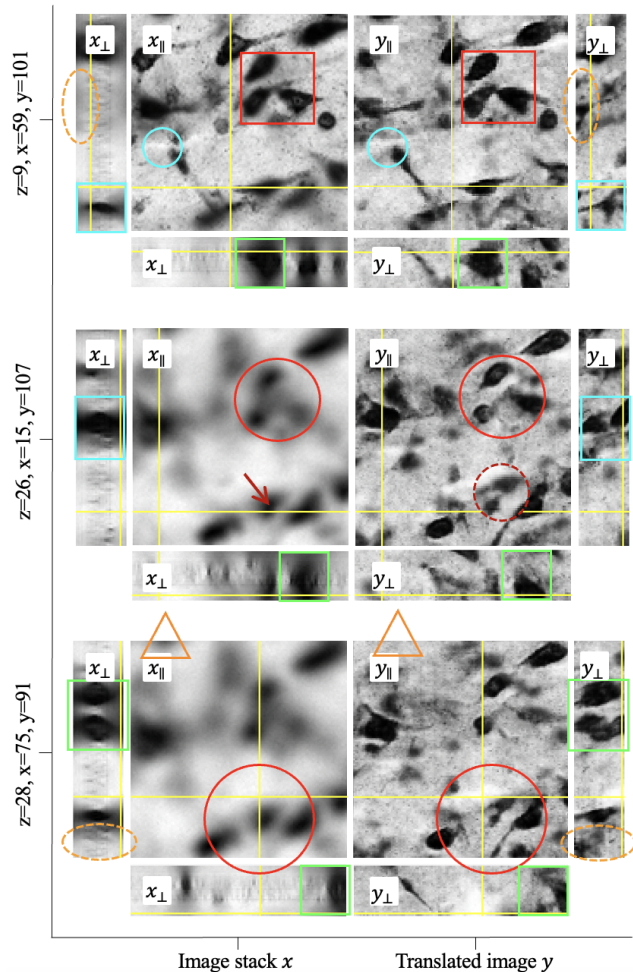
Our model was implemented using Pytorch. We qualitatively compare the result generated by ZSE-GAN with the input image stack here for a visual evaluation. We use the *Structural Similarity Index Measure* (SSIM) [12] as a reference of similarity between  $x_{\parallel}$  and  $y_{\parallel}$ . In order to assess whether the sharpness of  $y_{\perp}$  is improved and indistinguishable from the sharp images from the focal plane, we analyse the statistical gradient magnitude and FM [13]. For an evaluation of the  $y_{\perp}$  that has an image quality similar to that of the sharp images from the focal plane, we adopt the no-reference image metric PIQE [14].

### 3.1. Visual result

The third and fourth columns in Fig. 4 show the cross-sections (with notations  $y_{\parallel}$  and  $y_{\perp}$ ) of the translated 3D image volume slicing in three different orthogonal planes. The ROIs are marked with different colours and shapes. As a reference, the same locations are indicated in the input image stack  $x$  extracted from one particular sample of a z-stack. As can be seen in the example illustrated in Fig. 4, ZSE-GAN produces plausible structures: The instances of cells match the inputs and become clear. We observed three levels of blurry patterns in  $x$ , which appear visually sharper in  $y$  in the same position:

1. The blurry patterns, which still have some shapes or unclear structures of cells, are translated to sharp ones (marked with red circle) in  $y$  and tend to produce fine textures, which are pointed out using red arrows and shown in the red dotted circle for the enlarged parts.
2. Fig. 4 shows that the structures and textures are monotone and blurry in  $x_{\perp}$ . The unstructured dark patches marked with green squares in  $x_{\perp}$  are translated to various cell structures in the same positions in  $y_{\perp}$ . The dark patches from  $x_{\perp}$  mostly appear in similar flat oval shapes, whereas they are translated to different shapes and amounts of cells (marked with blue squares) with proper sharpness at the





**Fig. 4:** Orthogonal view of input image stack  $x$  and translated Result  $y$ .

3. However, some of the blurry patterns in  $x_{||}$ , which possibly come from other images because of the transparency of the brain tissue, get removed in  $y_{||}$ , which are marked with orange triangles.

Not only are there blurry patterns in  $x$  but also the sharpness of some images is very high; for example, the cells (marked with red squares in Fig. 4) in  $x_{\parallel(z=9)}$ , which are near to the focal plane. After implementing the ZSC-GAN, the blurry patterns become sharp, or fine cell structures are created from the blurry places in the same position, or the spots from the images in other layers vanish. But the sharp cells should maintain their sharpness and structure in the same position, examples are marked with red square, in which, the translated  $y_{\parallel}$  preserves its sharpness and structure in the same position as  $x_{\parallel}$ . There are also not always cell structures everywhere in ROIs. The white pattern should produce nothing after our image enhancement: Some blank places marked with blue

circles in  $x$  stay blank but with sharper edges.

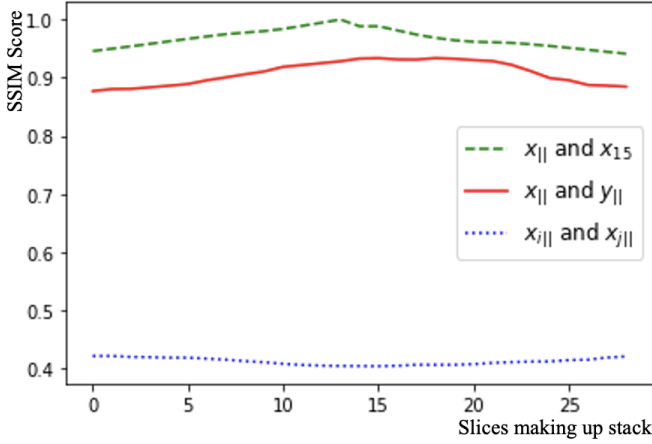
### 3.2. Experimental measurements

We test our proposed method in two categories: similarity of  $x_{\parallel}$  and  $y_{\parallel}$  and indistinguishability of  $y_{\parallel}$  and  $y_{\perp}$  from the sharp image  $x_{15}$  from the focal plane. The outermost images of an image stack show a lack of more precise information as they are farther away from the focal plane. Because of the transparency of tissue, the outer images possibly even contain some textures from others. Given an image stack  $x$ , if  $x_{\parallel}$  is near to the sharp image from the focal plane, then the desired  $y_{\parallel}$  in the translated image  $y$  should be highly consistent to  $x_{\parallel}$ . However, the generator  $G_S$  learns to create cell structures from the dark pattern in the blurry outermost images. Thus, there should be a loose similarity between  $x_{\parallel}$  and  $y_{\parallel}$ , if  $x_{\parallel}$  is far away from the central image. Besides the conditional similarity,  $y_{\vee}$  from the desired translated  $y$  should have a sharpness and image quality similar to those of the sharp image from the focal plane.

### 3.2.1. Similarity measure

We cropped 449 ROIs for measuring the correspondence between input image stacks and the translated ones. The expected similarity between  $x_{\parallel}$  and  $y_{\parallel}$  should be sufficiently high and slightly increase from the outermost to the centre. Since the blurring has less of an impact on the more uniform regions (the obvious cell structures or blank areas) of the image, in order to measure the structural similarity between the blurry input  $x$  and the sharp output  $y$ , the similarity metric used here shall achieve high score for measuring the image and its blurring version. SSIM metric is based on visible structures in the image. The value of SSIM metric is supposed to be in the range  $[0, 1]$ . The value 1 indicates the best similarity and happens if the two images are equivalent.

In Fig. 5, the SSIM score (indicated as blue dotted line) of comparison between two different image stacks  $x_i$  and  $x_j$  is small for all the slices making up the stack, which corresponds to poor similarity. We compare the similarity of the images making up the image stack with its sharp image from the focal plane. The sharp image of is supposed to be the central image  $x_{15}$  of the image stack. As shown in Fig. 5, the SSIM score reaches 1 in the middle of the green dotted line and are larger than 0.9 for all the slices, when we measure the similarity between all the slices and the sharp one of the same input image stack. Since the images of an image stack are more likely to be similar to the sharp image of the same image stack, we calculate it as reference. Both of the red and green dotted lines show the SSIM scores are high and decrease slightly from the centre to the outermost. Both of them show . But compared to the SSIM score between  $x_{||}$  and  $x_{15}$ , the red line Fig. 5 displays visibly poorer similarity than the green dotted line shows. Because  $y$  shall not only become



**Fig. 5:** SSIM score between input  $x$  and its translated output  $y$ , two arbitrary inputs  $x_i$  and  $x_j$ , and every slices of input  $x$  and the sharp slice of the same input

sharper than  $x$ , but also create cell structures from blurry pattern and remove the pattern coming from transparency. The average SSIM score of the 449 ROIs is displayed in Table 1.

For our experiment, the ground-truth data are not available. Furthermore, there is no conventional GAN model in comparison to ours for reference. We need to estimate whether  $y_{||}$  and  $y_{\perp}$  have a sharpness and an image quality similar to those of the in-focus image.

### 3.2.1.1. Sharpness measurements

We chose 1000 ROIs for analysis the statistical gradients of the input image stacks and the generated images. As displayed in the upper left corner of Fig. 6, the gradient magnitude histogram of  $x_{||}(z=21, z=8, z=25)$ , which are away from the centre of the stack, is dominated by smaller gradients compared to the central image  $x_{||}(z=15)$ . As  $x_{||}(z=25, z=8 \text{ or } z=21)$  are farther away from the central image  $x_{||}(z=15)$ , the cyan dotted curve followed by the green, blue and black dotted curves has higher value around gradient zero and a steeper slope successively. After the image enhancement translation, the gradient magnitude histogram curves of  $y_{||}(z=25, z=8 \text{ or } z=21)$  merge to the curve of central image  $x_{||}(z=15)$ , which is illustrated in the upper right corner of Fig. 6.

Chosen two different  $y_{\perp}(x=25, y=25)$  have identical gradient magnitude histogram curves, but higher value around gradient zero in comparison to the  $x_{||}(z=15)$ , which is marked with a back dotted curve in the middle left diagram of Fig. 6. The same as the result from  $y_{||}$ , the curves merge to the central one after image enhancement. The output histogram confirms our expectation that the gradient magnitude histograms of  $y_{\vee}$  become similar to the  $x_{||}(z=15)$ , and to reflect the enhanced sharpness overall. The average absolute gradient values of those  $y_{||}$  and  $y_{\perp}$  is displayed in Table 2, which obtained the result which coincides with what gradient magnitude histogram

Slice	1	2	3	4	5
$x_{  }/x_{15}$	0.95	0.95	0.95	0.96	0.96
$x_{  }/y_{  }$	0.88	0.88	0.88	0.88	0.89
$x_{i  }/x_{j  }$	0.41	0.41	0.41	0.41	0.4
Slice	6	7	8	9	10
$x_{  }/x_{15}$	0.97	0.97	0.97	0.98	0.98
$x_{  }/y_{  }$	0.89	0.90	0.90	0.91	0.91
$x_{i  }/x_{j  }$	0.41	0.41	0.41	0.41	0.41
Slice	11	12	13	14	15
$x_{  }/x_{15}$	0.98	0.99	0.99	1.00	0.99
$x_{  }/y_{  }$	0.92	0.92	0.92	0.93	0.93
$x_{i  }/x_{j  }$	0.41	0.41	0.41	0.41	0.41
Slice	16	17	18	19	20
$x_{  }/x_{15}$	0.99	0.98	0.97	0.97	0.96
$x_{  }/y_{  }$	0.93	0.93	0.93	0.93	0.93
$x_{i  }/x_{j  }$	0.41	0.41	0.41	0.41	0.40
Slice	21	22	23	24	25
$x_{  }/x_{15}$	0.96	0.96	0.96	0.96	0.95
$x_{  }/y_{  }$	0.93	0.93	0.92	0.91	0.90
$x_{i  }/x_{j  }$	0.40	0.40	0.40	0.39	0.39
Slice	26	27	28	29	
$x_{  }/x_{15}$	0.95	0.95	0.94	0.94	
$x_{  }/y_{  }$	0.90	0.89	0.89	0.88	
$x_{i  }/x_{j  }$	0.39	0.39	0.40	0.41	

**Table 1:** Image quality and sharpness scores of  $y_{\vee}$ .

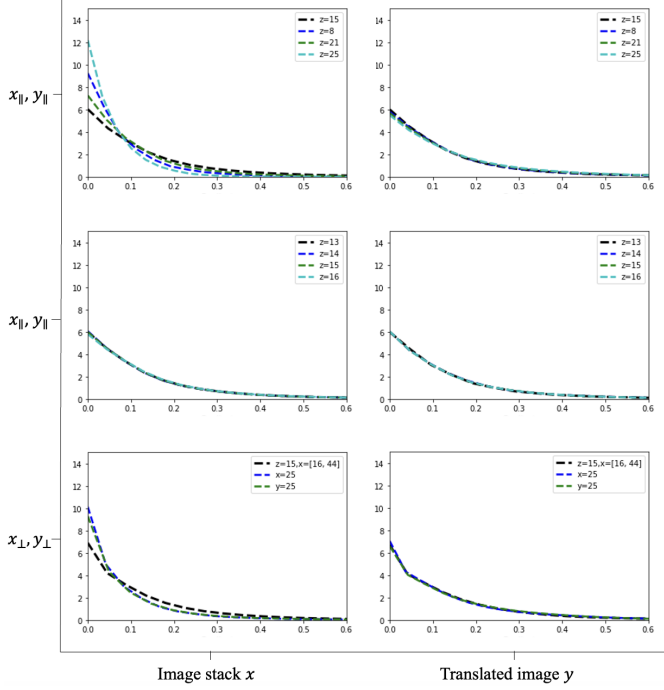
shows.

Average gradient					
	$x_{  }$	$y_{  }$		$x_{\perp}$	$x_{\perp}$
$z=25$	0.0537	0.1315	$x=24$	0.0841	0.1307
$z=22$	0.0786	0.1300	$y=24$	0.0885	0.1217
$z=8$	0.0894	0.1268	$z=15$	0.1220	0.1240
$z=15$	0.1293	0.1310			

**Table 2:** Average absolute gradient values of 1000 ROIs.

Analysing gradient frequency is evidence for the enhancement of the sharpness after translating  $x$  to  $y$ . However, the  $x_{||}(z=12, z=13, z=14)$  near the centre have no distinct difference from the  $x_{||}(z=15)$ . This phenomenon is displayed in the last row of Fig. 6. For the less blurry images, the gradient magnitude histogram alone cannot sufficiently be the assessment for either sharpness improvement or indistinguishable sharpness. Thus, we adopt another measurement to assess the enhancement of sharpness. Depending on the depth of focus, the sharpness of  $x_{||}$  becomes visibly lower from the centre to the outermost. We need a measurement to represent this tendency in order to proceed to the next step to assess the improvement of  $y$ . For measurement FM, the scores of  $x_{||}$  monotonically increase from  $x_{||}(z=1)$  to  $x_{||}(z=15)$  and monotonically decrease after  $x_{||}(z=15)$  until the last.

Table 3 bellowed reports the metric value of FM for  $y_{||}$



**Fig. 6:** Gradient frequency of cross sections.

and  $y_{\perp}$ . FM mostly achieves larger value than the reference in-focus image  $y_{||}(z=15)$ , what is illustrated in the upper left corner of Fig. 7 intuitively. The graph of the sharpness metric of  $x_{||}$  marked with blue curve has a similar shape as an opening-downwards parabolas, which satisfies the sharpness changing direction of the original image stack. The dark red bars, which indicate the FM values of corresponding translated  $y_{||}$ , are above the FM metric curve of  $x_{||}$ . The FM score has the largest value around  $y_{||}(z=15)$ , but the second largest around the outermost; the images after  $y_{||}(z=20)$  have lower sharpness. FM was designed in frequency domain and based on the amount of blurriness in the image. Compare to the gradient magnitude histogram measure, it is thus sensible enough to detect gradually changing sharpness by setting the proper threshold.

In the upper left corner of Fig. 7 is displayed the graphic result of  $y_{||}$ . Since the outer images have less fine texture (cf. the  $y_{||}(z=4)$  in the right upper corner of Fig. 4), the amount of blurriness is also possibly less than that of those images with fine textures. The FM value of outer  $y_{||}$  is possibly large, though the sharpness is not surely higher than the others.

The FM value of  $y_{\perp}$  is more or less similar to the value of  $x_{\perp}(z=15)$ ; see Table 3. The diagram in the left lower corner of Fig. 7 demonstrates the FM values of  $y_{\perp}$  marked with read dots are fluctuated along the value of  $x_{\perp}(z=15)$  marked with a red dashed line. The FM values of both  $x_{\perp}$  and  $y_{\perp}$  are steady, i.e., the sharpness of  $x_{\perp}$  or  $y_{\perp}$  is indistinguishable from other vertical cross-sections in  $x$  or  $y$  according to the FM metric. But the value of  $y_{\perp}$  is much larger than  $x_{\perp}$ 's, i.e.,

the sharpness of  $x_{\perp}$  gets improved after image enhancement.

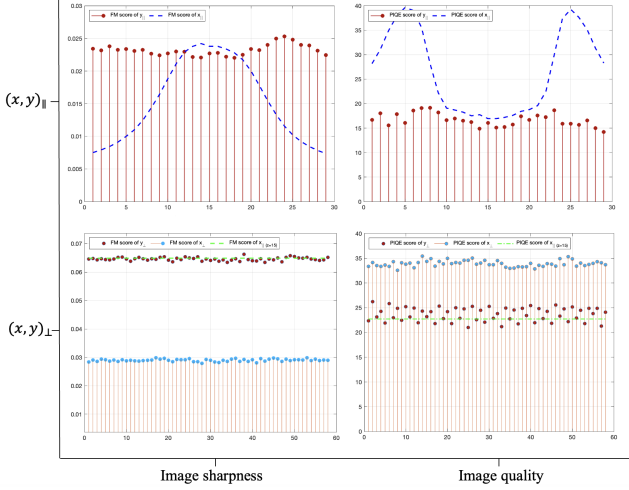
Slice	1	2	3	4	5
PIQE( $y$ )	16.69	18.04	15.58	17.84	16.10
PIQE( $x$ )	28.18	31.06	34.91	37.95	39.65
FM( $y$ )	0.0234	0.0232	0.0238	0.0232	0.0234
FM( $x$ )	0.0075	0.0079	0.0084	0.0091	0.0100
Slice	6	7	8	9	10
PIQE( $y$ )	18.59	19.09	19.14	18.20	16.6
PIQE( $x$ )	39.03	35.49	28.27	22.24	19.10
FM( $y$ )	0.0231	0.0232	0.0227	0.0224	0.0227
FM( $x$ )	0.0110	0.0125	0.0143	0.0164	0.0188
Slice	11	12	13	14	15
PIQE( $y$ )	16.99	16.50	16.24	14.88	16.09
PIQE( $x$ )	18.71	18.27	17.53	17.75	16.96
FM( $y$ )	0.0230	0.0230	0.0222	0.0221	0.0227
FM( $x$ )	0.0211	0.0229	0.0238	0.0242	0.0238
Slice	16	17	18	19	20
PIQE( $y$ )	15.14	15.23	15.73	17.40	16.66
PIQE( $x$ )	16.96	17.20	17.61	18.44	18.79
FM( $y$ )	0.0228	0.0222	0.0220	0.0225	0.0234
FM( $x$ )	0.0238	0.0234	0.0228	0.0217	0.0200
Slice	21	22	23	24	25
PIQE( $y$ )	17.58	17.26	18.63	15.93	15.91
PIQE( $x$ )	19.60	22.46	29.88	37.36	39.27
FM( $y$ )	0.0232	0.0240	0.0250	0.0253	0.0248
FM( $x$ )	0.0178	0.0155	0.0134	0.0116	0.0102
Slice	26	27	28	29	$y_{\perp}$
PIQE( $y$ )	15.71	16.59	15.03	14.21	22.73
PIQE( $x$ )	37.61	35.09	31.31	28.30	33.91
FM( $y$ )	0.0240	0.0239	0.0231	0.0224	0.0649
FM( $x$ )	0.0092	0.0084	0.0078	0.0074	0.0287

**Table 3:** Image quality and sharpness scores of  $y_{\perp}$ .

### 3.2.1.2. Image quality metrics

We want to measure the quality of the image  $y$  translated by ZSE-GAN without information of any actual sharp image stack. The  $y_{||}$  and  $y_{\perp}$  should have similar image quality with the in-focus images, which are near to the image  $x_{\perp}(z=15)$ .

The image quality metric we adopt here is PIQE, which was designed by assigning the significant local regions different distortion levels and calculating the overall perceptual quality of the image. The score of PIQE for the enhanced image  $y$  are summarized in Table 3. The PIQE score of  $x_{||}$  increases from the outermost images to the images  $x_{||}(z=5, z=25)$ , and decreases until the supposed in-focus image  $x_{||}(z=15)$ . The performance passes the expected image quality of original input stack  $x_{||}$  (cf. the blue curve in the upper right corner of Fig. 7). After the image enhancement, the PIQE score of  $y_{||}$  has a value similar to that of  $x_{||}(z=15)$ . The relatively better and worse scores of PIQE are from im-



**Fig. 7:** Illustration of image sharpness and quality.

ages around  $x_{||}(z=4)$  and  $x_{||}(z=25)$  respectively. The PIQE scores of  $x_{\perp}$  are around 30 to 40 and slope up and down obviously, while the scores of  $y_{\perp}$  are around 20 to 25 and gently fluctuated.

#### 4. CONCLUSION AND DISCUSSION

Thus far, we have presented the qualitative and quantitative output of ZSE-GAN. We have shown in this work that the focus-depth-based ZSE-GAN framework can be utilized to learn a translation of z-stack images. This can be achieved by using 2D discriminator to encourage the 3D generator to produce sharper 3D isotropic volumetric images from the corresponding image stacks. Such a process further entails the use of different consistency weights for the stacked images to preserve the correspondence in different levels.

Since the image stack  $x$  is scanned from a brain slide, and the tissue was partly transparent, some blurry patches in an image  $x_{||}$  may belong to the images above or below. Although we hypothesize our model deals with an image translation problem, our desired result is getting rid of the excess textures from other images; and generating the necessary textures in the very blurry images. As seen from the above visual result in Fig. 4, our ZSE-GAN accomplished this purpose: cell structures are detected from most blurry patches; simultaneously, some dark patches are erased if those patches are supposed to come from other images because of transparency. The translated image  $y$  preserves the significant structures from  $x$ . The MI score and Hellinger distance were calculated between  $x$  and  $y$ . The sharp images near  $x_{||}(z=15)$  have much better consistency to the images near  $y_{||}(z=15)$  in the translated result than the outer blurry images, which conforms to the weighted consistency of our network.

The sharpness of  $x_{||}$  is obviously not identical to each other. The frequencies of the gradients, FM score and  $\Delta\text{DOM}$  score are highly consistent for  $y_{||}$  of the translated images and

very similar to the in-focus images. The appearances of  $x_{||}$  and  $x_{\perp}$  are different. This difference can be attributed to the fact that  $x_{\perp}$  is rather blurry and simultaneously, has artificial effects (discrete or unsmooth). The  $x_{\perp}$  from original image stack has similar low sharpness. The ZSE-GAN succeeds in generating cell structures and textures from  $x_{\perp}$ , which are nearly as sharp as the in-focus images.

Currently, we haven't found a clearly defined quantitative evaluation to address the visual enhancement problem of image stacks. It is hard to decide which fine structures should be cleared away and what should be created from the noisy area. The reason why we adopted the PIQE as image quality reference for our experiments is that they can roughly simulate the quality of the images making up the image stack. The images  $x_{||}(z=5\dots25)$  are scanned by focusing on the brain slice. The brain slice is uneven on the surface. In order to contain the most data information of the slice, the images  $x_{||}(z=1\dots4, 26\dots29)$  are scanned by focusing above the actual slice. Thus, the outermost images possibly have better image quality than the neighboring images. That is why there are two humps around  $x_{||}(z=5)$  and  $x_{||}(z=25)$  in the graph of PIQE for  $x_{||}$ . The image quality of both  $x_{||}$  and  $x_{\perp}$  get obviously improved and come closer to that of the in-focus image.

**Compliance with ethical standards** The studies carried out require no ethical approvals. The human post mortem brains for these studies were obtained by the body donor programs of the Anatomical Institute of University of Düsseldorf, Rostock and Aachen or from collaboration between Anatomical Institute and departments of pathology in accordance with legal requirements. All body donors have signed a declarations of agreement.

**Acknowledgments** This work was supported by the Priority Programme "Computational Connectomics" (SPP 2041) of the German Research Foundation DFG, and received funding from the Helmholtz Association's Initiative and Networking Fund through the Helmholtz International BigBrain Analytics and Learning Laboratory (HIBALL) under the Helmholtz International Lab grant agreement InterLabs-0015. Computing time was granted through JARA-HPC on the supercomputer JURECA at Jülich Supercomputing Centre (JSC) as part of the project CJINM16. The authors declare no competing interests.

## 5. REFERENCES

- [1] T. C. Murakami et al., “A three-dimensional single-cell-resolution whole-brain atlas using CUBIC-X expansion microscopy and tissue clearing,” *Nature Neuroscience*, vol. 21, no. 4, pp. 625–637, Apr. 2018.
- [2] S. Zhao et al., “Cellular and Molecular Probing of Intact Human Organs,” *Cell*, vol. 180, no. 4, pp. 796–812.e19, Feb. 2020.
- [3] I. Costantini et al., “A new versatile clearing method for brain imaging,” in *Optical Techniques in Neurosurgery, Neurophotonics, and Optogenetics II*, Mar. 2015, vol. 9305, p. 930514, International Society for Optics and Photonics.
- [4] L. Pesce et al., “Fast volumetric mapping of human brain slices,” in *Neurophotonics*, Apr. 2020, vol. 11360, p. 113600A, International Society for Optics and Photonics.
- [5] M. de Castro Fonseca et al., “High-resolution synchrotron-based X-ray microtomography as a tool to unveil the three-dimensional neuronal architecture of the brain,” *Scientific Reports*, vol. 8, no. 1, pp. 12074, Aug. 2018.
- [6] H. Wang et al., “Characterizing the optical properties of human brain tissue with high numerical aperture optical coherence tomography,” *Biomedical Optics Express*, vol. 8, no. 12, pp. 5617–5636, Dec. 2017.
- [7] K. Amunts et al., “BigBrain: An Ultrahigh-Resolution 3D Human Brain Model,” *Science*, vol. 340, no. 6139, pp. 1472–1475, June 2013.
- [8] M. Mancini et al., “A multimodal computational pipeline for 3D histology of the human brain,” *Scientific Reports*, vol. 10, no. 1, pp. 13839, Aug. 2020.
- [9] J. Pichat et al., “A Survey of Methods for 3D Histology Reconstruction,” *Medical Image Analysis*, vol. 46, pp. 73–105, May 2018.
- [10] T. Dickscheid et al., “Towards 3D Reconstruction of Neuronal Cell Distributions from Histological Human Brain Sections,” *Advances in Parallel Computing*, vol. 34, no. Future Trends of HPC in a Disruptive Scenario, pp. 223–239, 2019.
- [11] J.-Y. Zhu et al., “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks,” *arXiv:1703.10593 [cs]*, Nov. 2018.
- [12] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [13] Kanjar De and V. Masilamani, “Image sharpness measure for blurred images in frequency domain,” *Procedia Engineering*, vol. 64, pp. 149 – 158, 2013, International Conference on Design and Manufacturing (IConDM2013).
- [14] Venkatanath N, Praneeth D, Maruthi Chandrasekhar Bh, S. S. Channappayya, and S. S. Medasani, “Blind image quality evaluation using perception based features,” in *2015 Twenty First National Conference on Communications (NCC)*, Feb 2015, pp. 1–6.