UNIVERSITY OF TARTU
Institute of Computer Science
Computer Science Curriculum

Denys Krupovych

Deblurring of microscopic 3D spheroid images using GANs

Master's Thesis (30 ECTS)

Supervisor(s):
Dmytro Fishman, PhD
Mikhail Papkov, MSc

Tartu 2023

Deblurring of microscopic 3D spheroid images using GANs

Abstract:

Spheroids are 3D aggregates of cells that have become increasingly important in the study of cancer and drug discovery due to their ability to mimic in real tumour microenvironments. However, spheroid imaging presents several challenges due to its complex structure, irregular shape, and optical properties. In this thesis, we experiment with deep learning approaches to address these challenges and improve the quality of spheroid images. Specifically, we use a modified U-Net architecture and generative adversarial networks (GANs) to generate high-resolution spheroid images. We evaluate our approach on a dataset of spheroids and compare the performance of unsupervised and supervised neural network architectures for the deblurring of spheroid images. Our work provides useful information for further research in spheroid image analysis and has potential applications in cancer diagnosis and drug discovery.

Keywords:

deep learning, convolutional neural networks, spheroids, generative adversarial networks, image-to-image translation, deblurring

CERCS: T111 - Imaging, image processing; P176 - Artificial intelligence; B110 - Bioinformatics, medical informatics, biomathematics biometrics

Mikroskoopiliste 3D-sferoidipiltide hägustamine GAN-ide abil

Lühikokkuvõte:

Sferoidid on 3D rakkude agregaadid, mis on muutunud vähi ja ravimite avastamise uurimisel üha olulisemaks tänu nende võimele imiteerida reaalseid kasvaja mikrokeskkondi. Kuid sferoidne kujutamine esitab mitmeid väljakutseid nende keerulise struktuuri, ebakorrapärase kuju ja optiliste omaduste tõttu. Käesolevas töös katsetame süvaõppe lähenemisviise, et lahendada neid probleeme ja parandada sferoidipiltide kvaliteeti. Täpsemalt, me kasutame modifitseeritud U-Neti arhitektuuri ja generatiivseid võistlusvõrke (GAN-id), et genereerida kõrge resolutsiooniga sferoidipilte . Hindame oma lähenemist spheroidide andmestikele ja võrdleme juhendatud ja järelevalveta närvivõrgu arhitektuuride toimimist sferodipiltide hävitamiseks. Meie töö annab kasulikku teavet sferoidipiltide analüüsi edasiseks uurimiseks ning sellel on potentsiaalseid rakendusi vähi diagnoosimisel ja ravimite avastamisel.

Võtmesõnad:

sügavõpe, konvolutsioonilised närvivõrgud, sferoidid, generatiivsed vastandvõrgud, kujutisest pildiks tõlkimine, hägustamine

CERCS: T111 - Pilditehnika; P176 - Tehisintellekt; B110 - Bioinformaatika, meditsiiniinformaatika, biomatemaatika, biomeetrika

Table of Contents

1 Introduction

In recent years, three-dimensional (3D) cell culture models, such as spheroids, have gained increasing interest in the field of biomedical research due to their ability to mimic the real microenvironment and exhibit physiological and functional properties that cannot be captured with traditional two-dimensional (2D) cell cultures. Spheroids are self-assembled, spherical aggregates of cells with great potential for various applications, including drug screening, tissue engineering, and disease modelling [1].

However, analysing spheroids can be challenging due to their complex structure and optical properties, which presents difficulties for imaging systems. Traditional methods for analyzing spheroids, such as spheroid cleaning protocols and optical cleaning [2] , have limitations in terms of their efficacy and cost. While classical computer vision algorithms, including deconvolution and thresholding-based methods, can also be used, they often require manual application to each individual data sample, which can be time-consuming and inefficient when working with large datasets.

Therefore, there is a need for automated and reliable methods for image deblurring, which is essential to facilitate the use of spheroids in biomedical research. In recent years, deep learning-based approaches, particularly convolutional neural networks (CNNs), have shown great promise in various biomedical image analysis tasks, including image translation and deblurring. These approaches can extract meaningful features from images, enabling accurate and efficient spheroid deblurring.

This thesis explores the potential of deep learning-based approaches for spheroid deblurring. We aim to analyse how neural networks can effectively solve the spheroid deblurring task, thereby accelerating the development of new therapies and improving disease modelling [1]. We evaluate our methods on provided datasets of microscopy images of spheroids and compare the performance of supervised and unsupervised models for deblurring. We investigated whether a model trained on unsupervised data could outperform a model trained on supervised data. Additionally, we investigate the effects of different network architectures and image augmentations on the performance of the proposed models. By doing so, we aim to contribute to developing automated and reliable methods for spheroid deblurring using deep learning-based approaches, thereby facilitating their use in various biomedical applications.

We have formed the thesis structure in the following way:

- Background chapter gives a concise biological context around spheroids and provides a short overview of neural network architectures and their application in image translation tasks;
- Data and Methods chapter describes microscopy image data used for spheroid deblurring, image preprocessing steps that were used for the experiments and the neural network parameters and modifications that were utilised during experiments;
- Experiments and Results chapter, gives an overview on the performed experiments with supervised and unsupervised data and models;
- In Discussion, we describe the limitations of the approaches that we tried and sets directions for potential future work;
- Conclusion summarises the results obtained during the experiments and the thesis itself;

2 Background

In this section, we first present a comprehensive introduction to spheroids, including their culturing process, distinction from 2D cell cultures, and the challenges associated with their analysis. Subsequently, we describe the neural network architectures that are commonly implemented in computer vision tasks and illustrate their relevance in resolving the image deblurring problem through the image translation task.

2.1 Spheroids

Spheroids are three-dimensional cell cultures closely mimicking the real microenvironment, making them important for studying cells' behaviour in isolated surroundings [3]. They are widely used in biomedical applications, such as drug discovery, tissue engineering, and cancer research [4]. Spheroids are formed by aggregating cells that have been cultured in laboratory surrounding under conditions that promote self-assembly. They can be made from different types of cells, including cancer cells, stem cells, and primary cells. They can be cultured using various techniques, including hanging drop, agarose overlay, and non-adherent surface-based methods [3]. These techniques promote cell aggregation and provide a 3D environment that allows cells to interact with each other and the extracellular matrix, forming spheroids. In this thesis, we are working with spheroid cysts. Spheroid cysts differ from spheroid tumours in their cellular composition and structure. Spheroid tumours are characterised by the growth of cells within the spheroid, while spheroid cysts contain a hollow lumen filled with fluid or other materials.

Compared to traditional two-dimensional cell cultures, spheroids offer advantages such as better resemblance to a real cell morphology, higher interactions between cells, and improved cellular functions [4]. They are essential for drug screening and development because they provide a more realistic environment for testing drug efficiency and irradiation response than 2D cell cultures [4]. However, analysing spheroids can be challenging due to their complex structure, heterogeneity, and variability. Traditional approaches for analysing spheroids involve manual analysis, which is time-consuming, subjective, and prone to errors.

Various approaches have been developed to overcome these challenges, including neural network-based methods, chemical and biological assays, and microfluidic devices. Neural networks have been shown to be effective in accurately segmenting spheroids, extracting features, and predicting drug response. Chemical and biological assays can provide information on the viability and function of cells within the spheroid. Microfluidic devices offer precise control over the microenvironment, enabling the manipulation and analysis of spheroids under controlled conditions.

2.2 Biomedical image translation using neural networks

Neural networks have been widely used in computer vision to extract meaningful information from images and perform classification, detection, segmentation, and image synthesis tasks. In biomedical image analysis, computer vision techniques have become increasingly

important to automate image interpretation, enable large-scale analysis, and assist medical diagnosis. Deep learning models such as convolutional neural networks (CNNs) or neural network frameworks such as generative adversarial networks (GANs), and their variants have been successfully applied to various biomedical imaging tasks, including tumour detection and segmentation, medical image super-resolution, and image-to-image translation. These models have shown promising results and the potential to improve the accuracy and efficiency of medical imaging analysis, ultimately leading to better patient care. The neural networks were also successfully applied to the biomedical image translation task. Biomedical image translation involves transforming one type of medical image to another, such as converting MRI images to CT images or converting low-resolution images to high-resolution images. This translation task is important for disease diagnosis, treatment planning, and monitoring. In recent years, various neural network architectures and frameworks, such as GANs and U-Nets, have been used for biomedical image translation tasks with promising results. Generative adversarial networks are the neural network frameworks that comprised of two or even more networks that compete each other. The discriminator is the first network which attempts to distinguish between real and generated images. The second network, known as the generator, aims to produce images that are indistinguishable from genuine images. These frameworks can learn complex and non-linear relationships between images and generate high-quality images that aid medical decision-making. This motivated us to investigate neural network architectures and their potential applications to address the problem of spheroid deblurring.

2.2.1 ResNet

The ResNet is a deep neural network architecture that can effectively train deep neural networks [5]. During training, the accuracy of the network may saturate and then degrade as the network depth increases due to the vanishing gradient problem, where the gradient signal becomes too small to update the weights in the earlier layers during backpropagation. To address this issue, the authors of ResNet architecture proposed a residual learning approach that can be expressed as:

$$\mathcal{H}(x) = \mathcal{F}(x) + x,$$

where x is the input, $\mathcal{F}(x)$ is a residual function, and $\mathcal{H}(x)$ is an identity mapping function. In ResNet, the model is divided into several residual blocks. The architecture of a ResNet block is shown in Figure 1. In this block, the residual function is described by its stacked layers. An identity mapping function uses a shortcut connection (residual connection) and element-wise addition to bypassing one or more layers in the network, allowing gradients to be backpropagated more effectively without adding new parameters and complex computations. The authors conducted ablation studies to evaluate the performance of neural networks with and without residual connections in multiple domains, such as image classification, object detection, and object localisation tasks. The results demonstrated that adding residual connections improves the network's performance in all evaluated domains.

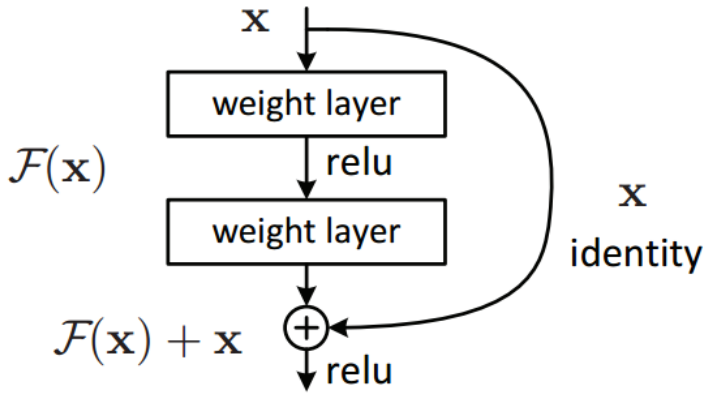

Figure 1. The architecture of the residual block [5]. The input of the layers is represented by x, and the residual function that stacks the network layers is represented by $\mathcal{F}(x)$.

2.2.2 U-Net

The U-Net [6] is a convolutional neural network architecture designed explicitly for biomedical image segmentation tasks with limited training data available. The architecture comprises a contracting path that captures context and a symmetric expanding path that enables precise localisation, as depicted in Figure 2. The contracting path involves multiple unpadded convolutional layers followed by a rectified linear unit (ReLU) and max-pooling layers, resulting in a downsampling approach. The downsampling operation reduces the spatial resolution and doubles the number of feature maps. The feature maps from the contracting path are cropped, copied, and combined with the upsampled expanding path output to improve the localisation after each downsampling step using skip connections. Cropping is necessary to fit the contracting feature map into the expanding feature map. The skip connections between the contracting and expanding paths maintain spatial information and enable precise segmentation. The expanding path involves a convolutional layer that reduces the input feature map, which is then concatenated with the cropped contracting feature map and passed through multiple convolution layers, each followed by a ReLU. The U-Net architecture was evaluated on three biomedical image segmentation tasks: segmentation of neuronal structures in electron microscopic recordings and cell segmentation in light microscopic images. In all tasks, the U-Net outperformed other methods regarding accuracy and error rates. It is worth mentioning that U-Net also required fewer training samples than other methods due to its efficient use of the available training data.

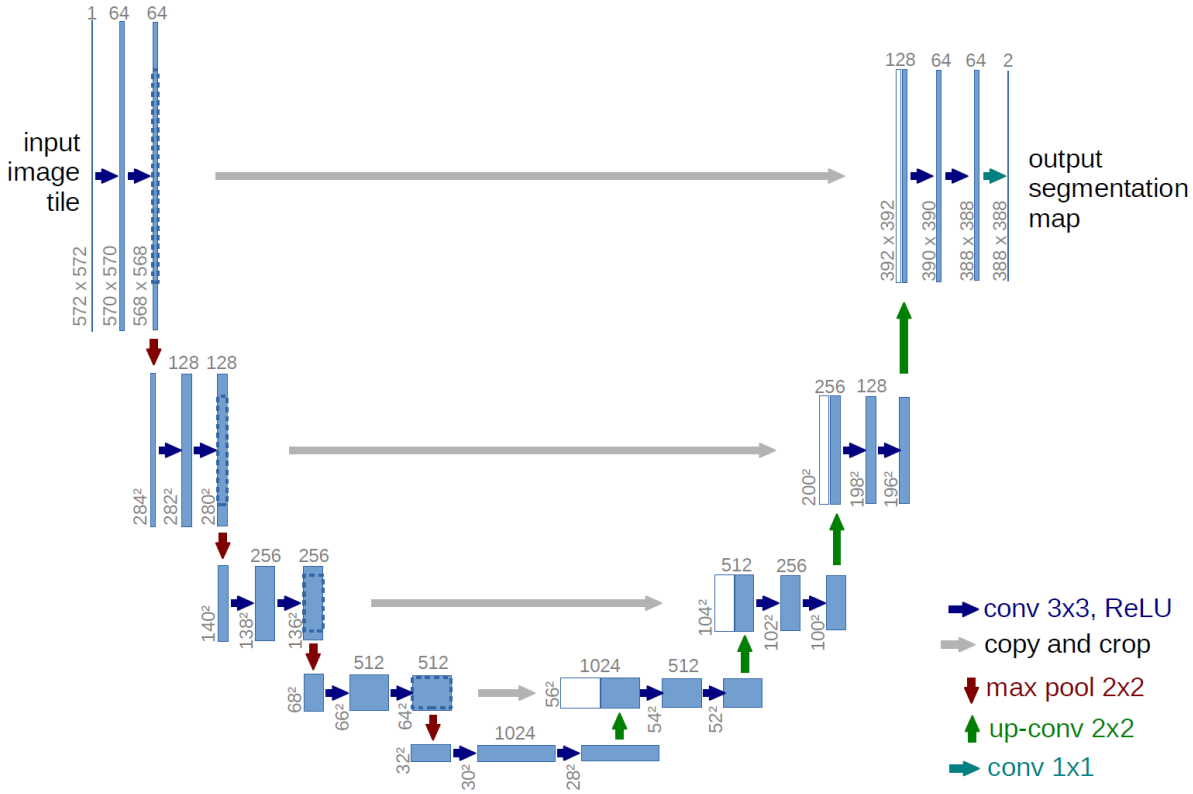

Figure 2. The U-Net architecture [6]. Blue boxes represent multi-channel feature maps, while copied feature maps are depicted by white boxes. Arrows indicate the different operations performed in the architecture. The top of the box indicates the number of channels, while the lower left edge indicates the x-y size.

2.2.3 pix2pix

Pix2pix is a conditional generative adversarial network capable of generating high-quality and realistic image outputs from input images [7]. As a conditional GAN, it learns a mapping between the input image with the random noise and the output image, requiring corresponding pairs for training. Pix2pix consists of a generator and a discriminator network, common elements of any GAN. Figure 3 presents the pix2pix architecture. The discriminator network aims to differentiate the real output image from the generated one by considering the input image. On the other hand, the generator network aims to deceive the discriminator and produce an indistinguishable image from the target image. The proposed pix2pix generator network is based on the U-Net architecture, an encoder-decoder network with skip connections that enable it to capture the input image's low-level and high-level features of the input image. The PatchGAN architecture was proposed as the discriminator network, a model that consists of convolutional layers. This architecture is called PatchGAN because it classifies between the real and generated images using image patches, which can be smaller or have the same size as the original image. The discriminator is applied convolutionally across the whole image, averaging all the responses to provide the output. The model was trained using a combination of adversarial loss and generator loss, which enforces the similarity between the generated and real images. The generator loss is the L1 loss, which

calculates the mean absolute difference between all generated and real output images. The adversarial loss function can be expressed as:

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y}[logD(x, y)] + \mathbb{E}_{x,z}[log(1 - D(x, G(x, z)))],$$

where x is the input image, z is the random noise, y is the output image, G is the generator, D is the discriminator, $\mathbb{E}_{x,y}$ denotes mean likelihood over all real and output images, $\mathbb{E}_{x,z}$ denotes mean likelihood over all real and generated images [7]. Here G tries to minimise the objective function while an adversarial D tries to maximise it.

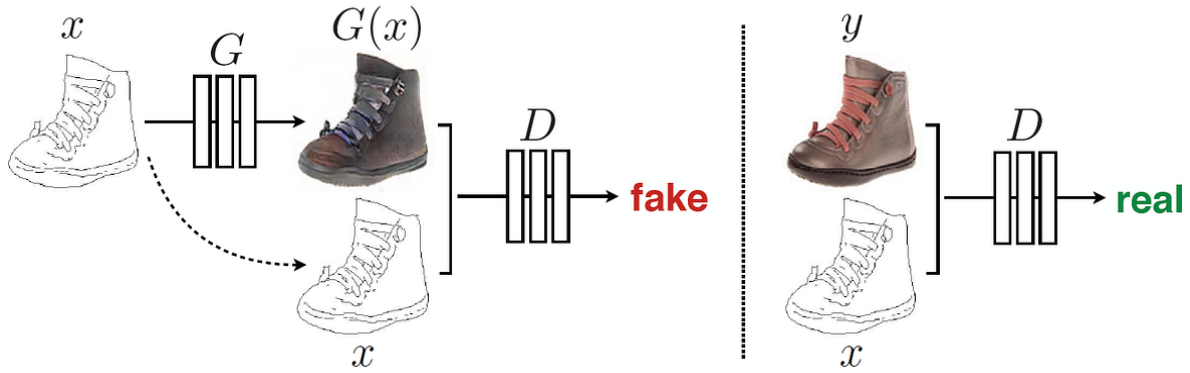

Figure 3. Pix2pix training [7]. The architecture of the model consists of a generator G and discriminator D. The discriminator classifies the target and generated images that are paired with the input image. Meanwhile, the generator attempts to produce an image closely resembling the target image to deceive the discriminator's classification.

Pix2pix was initially proposed as a deep learning-based method for metal artifact reduction (MAR) in computed tomography (CT) images of the ear [8]. MAR is a challenging task encountered in CT imaging due to the presence of metal objects, such as cochlear implants, that cause severe image distortions and artifacts. The dataset used in the study consisted of pre-implantation CT (Pre-CT) and post-implantation CT (Post-CT) images. As a pre-processing step, the Post-CT images were replicated into three channels, and the intensity values in each channel were modified using different percentiles. The authors trained the model using both ResNet-based and U-Net-based generators, with and without pre-processing and used a 70×70 PatchGAN as the discriminator network. The authors trained the model on a dataset of 90 ears and tested it on a dataset of 74 ears. The results demonstrate that the proposed methods outperform the previous MAR method in terms of both visual quality and quantitative metrics. The ablation experiments revealed that both U-Net-based and ResNet-based generators perform better with pre-processing, with the ResNet-based generator with preprocessing outperforming the U-Net-based generator with preprocessing. These results suggest that pix2pix has the potential to improve the diagnostic accuracy of CT imaging and assist in the clinical decision-making process for patients with cochlear implants.

2.2.4 CycleGAN

Similar to pix2pix, CycleGAN [9] is a framework that enables image-to-image translations. One of the benefits of CycleGAN over pix2pix is its capability to perform unsupervised image translations without requiring image pairs for training. Compared to pix2pix, the CycleGAN model comprises two generators that learn how to translate images between both domains and two discriminators that learn how to differentiate generated images for each generator. The CycleGAN architecture is shown in Figure 4 (a). Unsupervised image translation is achieved by modifying the loss function and introducing the cycle consistency loss. The general loss function is computed as the sum of adversarial losses for both domains and the cycle consistency loss. The adversarial loss is computed using the same method as in pix2pix. The cycle consistency loss is a sum of L1 losses that computes the difference between the original and reconstructed images for both domains, as shown in Figure 4 (b) and (c). The reconstructed image is the image that has been translated into the target domain and subsequently translated back into the original domain. The cycle consistency loss aims to ensure that the learned mappings for both domains do not conflict with one another. The authors employed a ResNet-based generator and a 70x70 PatchGAN discriminator for the CycleGAN model based on the performance of the 70x70 PatchGAN in pix2pix.

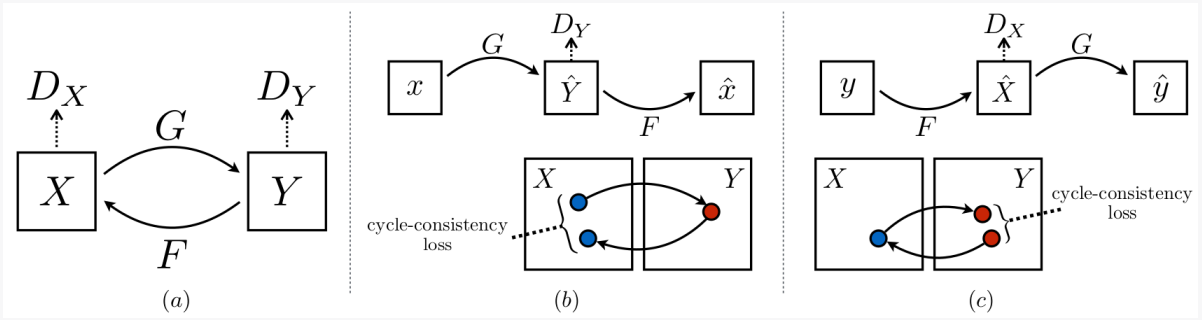

Figure 4. Description of how the CycleGAN works [9]. (a) describes the CycleGAN architecture that consists of two generators and two discriminators. (b) shows the forward cycle-consistency loss from domain X to Y and back. (c) depicts the backward cycle-consistency loss from domain Y to X and back.

CycleGAN has also been utilised in the biomedical domain for synthesising CT images from magnetic resonance (MR) images [10]. This approach addresses the issue of the limited availability of paired MR-CT datasets, which can impede the development and implementation of deep learning techniques for medical imaging. The dataset for the experiments consisted of brain MR and CT images of 24 patients. The dataset was partitioned into a training set containing images from 18 patients and a test set, which included MR-CT pairs from 6 patients. The model employed a ResNet-based generator and a 70x70 PatchGAN discriminator, similar to the approach in the CycleGAN paper [9]. The authors compared the performance of their method to a ResNet generator trained on paired images. The authors compared the synthesised CT and MR images and their respective ground truth images. The results indicate that CycleGAN trained on unpaired data outperforms the ResNet model trained on image pairs regarding image quality, accuracy, and robustness.

3 Data and Methods

This section describes the data that was used and the format of how the images were obtained. We give an overview of the preprocessing steps we performed to form the datasets for the deblurring task. Then we describe the image augmentation approaches that we used in our experiments. After that, we discuss the neural network architectures utilised in this work. Finally, we provide information about the AI-enabled services that assist thesis writing.

3.1 Dataset

Spheroids data were provided by PerkinElmer to the Biomedical Computer Vision Group of the University of Tartu. The spheroids were prepared using DRAQ5 fluorescent dye and imaged at 40x magnification on a 384-well microplate using the Opera Phenix High-Content Screening System. Two distinct datasets were utilised for this study. The first dataset, hereafter referred to as the "unsupervised" dataset, consisted of raw spheroid images and corresponding annotation files. The normalised crops of the front and rear parts of the spheroid from the normalized dataset are shown in Figure 5. The second dataset, hereafter referred to as the "supervised" dataset, comprised image pairs of raw spheroid images and their corrected versions with corresponding annotation files. The corrected images depict spheroids without background blur and highlight cells in certain areas. These corrected images were generated using a semi-automated proprietary algorithm developed and protected by PerkinElmer, which was used as the ground truth for this study. We would like to point out that we had no access to this data before most of our experiments.

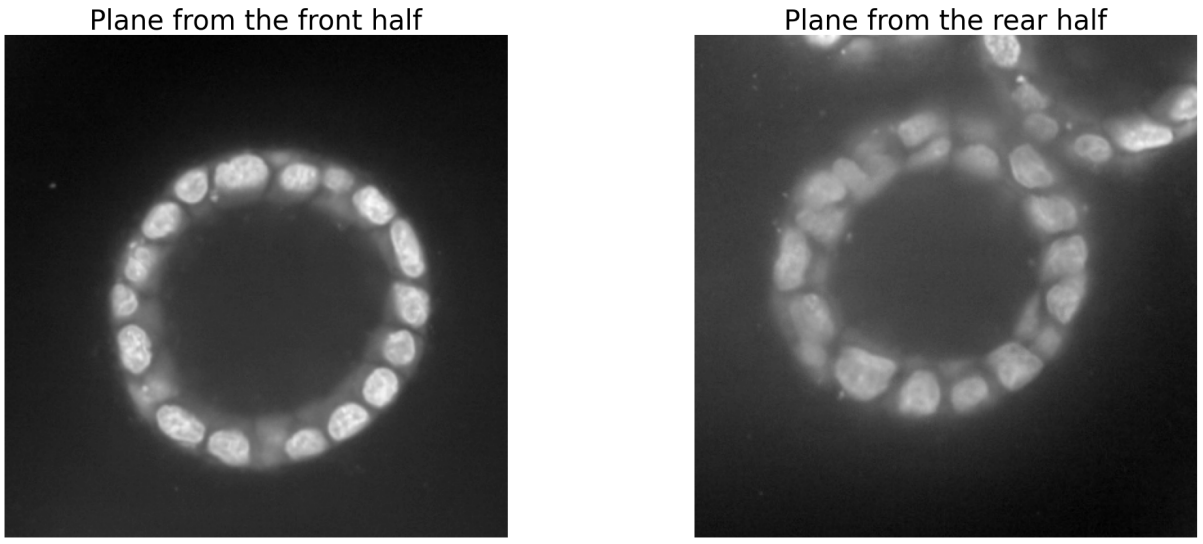

Figure 5. Two spheroid planes from the unsupervised dataset. The left image represents the clear half of the spheroid while the right image is from the blurred half of the spheroid. Both images are located in 30 planes each from the plane with the biggest spheroid diameter.

Both the unsupervised and supervised datasets are three-dimensional stacks of images. We use the term “three-dimensional stack” interchangeably with the term “stack”, while the

words “image” and “plane” are used as synonyms to denote individual images within the stack throughout the thesis. All images are saved in TIFF format with a grayscale 16-bit depth and a size of 1080x1080 pixels. The unsupervised dataset comprises 920 images, forming four stacks of 230 planes each. The supervised dataset consists of 600 images forming a stack of 300 image pairs. The distance between the two images in all stacks is 0.5 μm. The unsupervised dataset contains 31 spheroids, while the supervised dataset contains seven spheroids of varying depths and morphologies. Additionally, the number of spheroids in the unsupervised stacks is uneven, with the first stack containing fifteen spheroids, the second containing thirteen spheroids, and the third and fourth containing one and two spheroids, respectively. As the screening system only images a portion of the well, some spheroids are not fully captured, resulting in variations in the size of the measured spheroids.

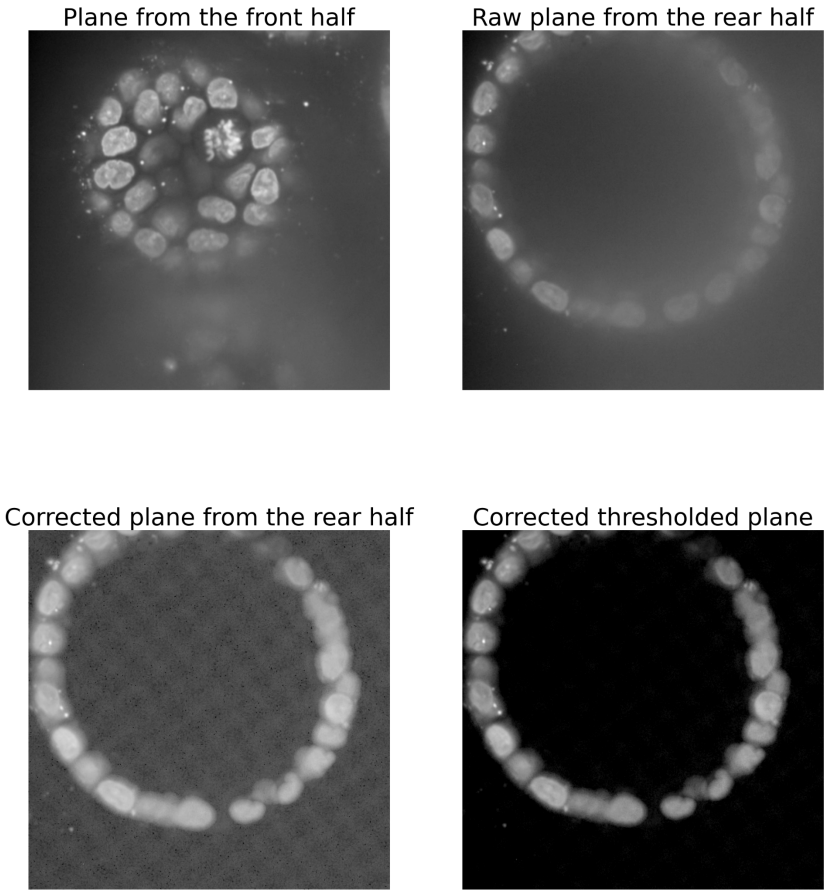

Figure 6. The examples of data in the supervised dataset. The upper left image represent plane that is located on the clear part. Raw plane from the rear half that shows how the spheroid looks without correction and located upper right. Corrected plane that shows how the corrected image looks after normalisation is located at bottom left figure. Our approach on the reduce the visibility of the pattern is located on the bottom right figure.

The obtained corrected images had one feature we observed after normalisation. Due to the correction algorithm outputs, each corrected image had a checkerboard pattern in the background which can be observed in Figure 6 as well as raw images and our attempts to reduce it visibility. This pattern appeared because the correction of spheroids changes the

intensity and signals to get rid of the blur and increase the brightness of the cells. In order to reduce its visibility on the images, we decided to trashhold values that are responsible for the background. This approach does not remove the pattern, but allows to ignore it during analysis. We used this thresholding for the corrected images to make them more contrast in further experiments.

3.2 Preprocessing

We used three steps to create a dataset from the original images. These pre-processing steps were applied to both supervised and unsupervised datasets. The first step is image transformation (normalisation), as the original images are in low contrast. The second step is image segmentation. This step is necessary to collect information about spheroids' sizes, lengths and shapes. Based on the second step, the third step is to crop the images and form the dataset. This is needed to make a dataset bigger and reduce the white space in the image. We experimented with multiple transformation (normalisation) techniques, segmentation methods and crop sizes to make the optimal images for neural network training.

3.2.1 Image transformation (normalisation)

The raw images captured by the imaging system used in this study exhibited low contrast, making it challenging to discern the cells and the surrounding blur. To overcome this limitation, we implemented multiple image transformation techniques to normalise the image values and increase the contrast.

One of the techniques we applied was log transformation, which involved applying the natural logarithm function to each pixel value in the image. By expanding low-intensity pixel values and compressing high-intensity values, this transformation led to an image with improved contrast and dynamic range. While log-transformed images facilitated a more precise structure of the spheroids, they also increased the contrast of the background blur.

We also attempted quantile transformation, which involved reassigning pixel values below and above the distribution quantile to the closest pixel values. This technique limited the pixel value range and allowed the highlighting both high and low pixel values. However, it resulted in glaring cells of the spheroids, which made it challenging to discern the structure of single cells.

Finally, we explored gamma correction, which involved altering the image's brightness and contrast by changing the relationship between the input and output pixel values. This technique relied on the gamma function, which relates the input pixel value to the output brightness degree. The gamma function can be defined by the expression:

$$x_{out} = x_{in}^{\gamma},$$

where x_{in} represents the input image and γ is the gamma value, which is applied as an exponent to the input image to obtain the output image x_{out}. A gamma value of less than one would increase the image's contrast and make it brighter, while a gamma value greater than one would decrease the contrast and make it darker. We used a gamma value of 0.2 to normalise the images for our experiments.

Following experimentation with various normalization techniques, we found that gamma correction was the most effective approach for normalizing the images. Our selection of an image normalization technique was based entirely on qualitative analysis of the normalised images. Subsequently, upon acquiring the supervised dataset, we confirmed our suggestion that gamma correction produced a superior performance by utilizing the Structural Similarity Index Measure (SSIM) on image normalisation techniques. The gamma corrected images contained sufficient information about the background blur and represented the bright enough spheroids. The comparison of all image transformation methods applied to one specific image is depicted in Figure 7.

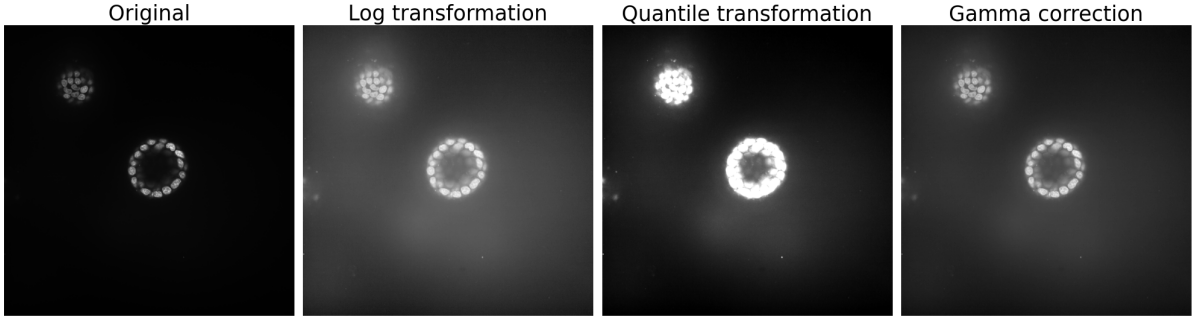

Figure 7. Comparison of the image transformation results. Original — the raw image we obtained from PerkinElmer. Log transformation — the result of logarithmic transformation employed to the raw image. Quantile transformation — the result of Quantile transformation utilised to the raw image. Gamma correction — the result of Gamma correction applied to the raw image.

3.2.2 Segmentation

It was necessary to segment spheroids to obtain crucial information about their size and precise location. Spheroids could be segmented by detecting connected components to thresholded pixel values across the entire image stack. To achieve the segmentation, we tried several thresholding techniques.

The first approach we tried was Otsu's thresholding, which automatically determines the threshold value based on the pixel intensities. Instead of being manually selected, Otsu's threshold value is based on minimising the variance between the background and foreground classes while maximising the between-class variance. This approach produced segmentations that captured all spheroids on the image. However, the generated masks were inaccurate as they included background blur. This led to the blur filling the spaces between the spheroids

and not allowing them to segment as separate entities. An example of how the blur capture affects segmentation is shown in Figure 8, titled “Otsu”.

Subsequently, we attempted to segment spheroids using a watershed algorithm. This technique involves applying gradient transformation to the image to identify local minima, the regions of interest. These regions are marked as the source of “water” that "floods" the surrounding parts of the image until it reaches the "flood" from another water source. The resulting watershed line forms the boundary between water sources and is used to create segmentation masks. Although watershed segmentation is more advanced than previous algorithms, it is susceptible to image noise. Figure 8, titled “Watershed”, depicts the result of the watershed algorithm. In our case, the watershed algorithm could not capture the cells in spheroids as a single entity. Connected component labelling revealed that the spheroids were incorrectly labelled. Hence, we decided to adopt another approach for segmentation.

Finally, we utilised the segmentation masks provided by PerkinElmer, which were generated using a semi-automated proprietary algorithm for segmentation. This algorithm, developed and protected by the company, provided ready-made segmentation masks that eliminated background blur from the spheroid masks and segmented the spheroids, considering their morphology. The visual representation of segmentation provided by PerkinElmer is shown in Figure 8, titled “Proprietary”.

We opted for a proprietary algorithm to segment the connected components of the spheroids in the stack, as it produced the most precise masks. This allowed us to compute all the necessary spheroid measurements accurately. Figure 8 visually compares the segmentation results obtained through all the approaches applied to a specific image.

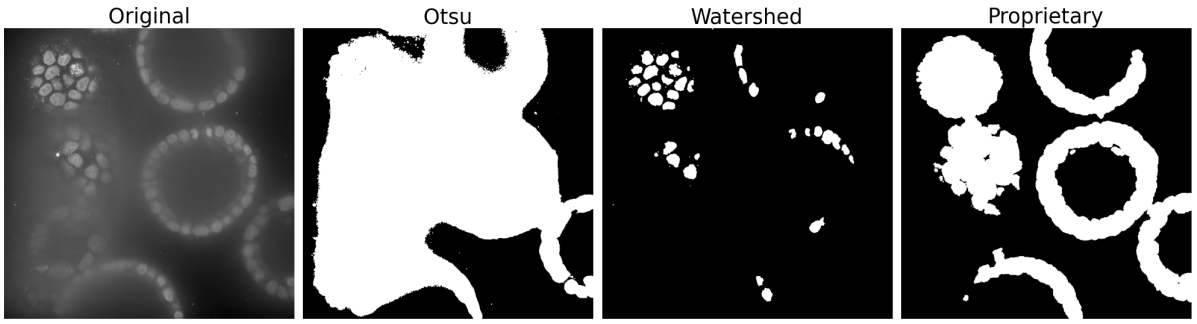

Figure 8. Comparison of all segmentation approaches applied on one normalised image. Original — gamma-corrected image before segmentation. Otsu — the result of Otsu’s thresholding on the image. Watershed — the result of the Watershed algorithm applied to the image. Proprietary — the obtained mask that was segmented using a proprietary algorithm.

3.2.3 Cropping and dataset generation

Our analysis of spheroid sizes revealed that the largest spheroid in the unsupervised dataset was 391x338 pixels, while the largest spheroid in the supervised dataset was 411x502 pixels. To standardise the image sizes, we cropped the original images into 391x391 pixel crops for the unsupervised dataset and 512x512 pixel crops for the supervised dataset. As a part of our

observations, we calculated each spheroid area on the planes. We also calculated the area of each spheroid on its respective planes and identified the plane and corresponding spheroid coordinates where each spheroid had the largest area. The plane with the largest area was called the "biggest plane" of the specific spheroid. For each spheroid, we selected the XY coordinates of the biggest plane as the reference point for image cropping. Including all planes of the spheroid in the dataset during training resulted in noisy images, so we filtered out planes with less than 20% of the size of the biggest plane. This means crops with a spheroid area of fewer than 20 pixels will not be included in the dataset if the biggest plane area is 100 pixels.

For the unsupervised dataset, we divided each spheroid into two parts based on the biggest plane. When the light beams from the fluorescent dye reach the spheroid cells, the cells exhibit absorption and reflection, causing only a fraction of the light to transmit. Consequently, inadequate light reception occurs in the area behind the biggest plane, resulting in bright and well-defined cells in front of the biggest plane. In contrast, the cells behind it appear blurred and indistinguishable. This led us to use the planes before the biggest plane as a “clear” dataset, while the planes behind it were used as a “blurred” dataset.
In contrast, for the supervised dataset, we formed pairs with raw and corrected planes for each spheroid without dividing it into halves, like in the case of the unsupervised dataset.

3.3 Image augmentations

After the first experiments with the unsupervised dataset, we found that the model was sensitive to the spheroid surrounding. This resulted in artifacts such as noise around the spheroid perimeter, hallucinating cells on the image, and adding more blur on the crops. While noise can be avoided simply by removing the first and last planes of the spheroid in the dataset, we assumed that the solution for removing hallucinating cells after generation is to remove all the cells that do not belong to the spheroid itself from the image. As our initial experiments were conducted with a limited number of spheroids, we tried to solve the problem by augmenting the existing crops. To perform these augmentations, we based them on the spheroid masks obtained from the segmentation. The comparison of all implemented augmentations is shown in Figure 9.

The first image augmentation approach that we tried was to “filter out” neighbouring spheroids from crops. We call it “filter out” because we also change the background values around spheroids in this approach. The idea of this approach consists of two parts. The first part is to change the background around the spheroids by decreasing the pixel intensity higher than the mean background intensity. In the second part, we remove all the adjacent spheroids and replace them with pixel values with slight deviations from the mean.

The second augmentation we used was to “blur” surrounding cells. In this approach, we apply heavy Gaussian blur with a sigma value of 431 on the nearby spheroids. This approach does not remove neighbouring spheroids but makes them blurry enough that the model cannot

relate them to the environment of the target spheroid.

The third approach we applied to the spheroids was to “dilate” the target spheroid mask and remove all the background that does not belong to the target mask. In this approach, we use morphological operator dilation to expand the spheroid mask and to capture the surrounding blur. The cells and other backgrounds outside this mask were removed from the image. That augmentation helps to save the surrounding blur we are interested in and remove all the distractions from planes.

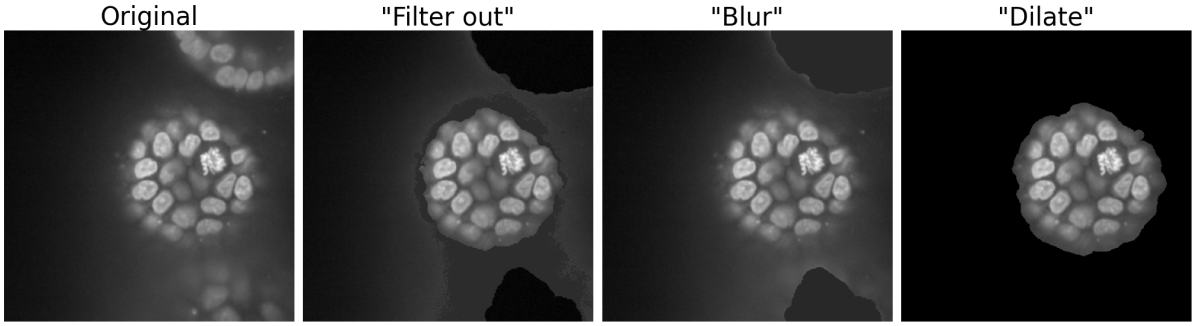

Figure 9. Results of the image augmentations that were applied on crops. Original — gamma-corrected image before applying augmentations. “Filter out” — the approach where the neighbouring spheroids are removed, and background values are changed. “Blur” — the augmentation that applies heavy Gaussian blur on surrounding spheroids. “Dilate” — an approach that extends the spheroid segmentation mask using a morphological dilation operator and removes all the background outside the mask area.

3.4 Neural Networks

This section describes the specifications of the neural networks used to solve the deblurring task in spheroids that we focus on in this thesis. As the main neural network architectures related to our task were described in the Background part, in this section, we discuss the changes we made to the baseline approaches to be able to deblur images of spheroids. The unsupervised dataset was utilized for training both CycleGAN and CycleGAN with conditional planes. CycleGAN with and without conditional planes were also qualitatively evaluated using the unsupervised dataset, while the supervised dataset was later used only for their quantitative evaluation. The 3D-U-Net requires image pairs, so we trained it only using the supervised dataset. As our experiments progressed, we discovered the need to go beyond 2D image translation using CycleGAN and added conditional planes into the model, transforming it into 2.5D image translation. Already having successful examples of U-Net modifications for 3D images, we decided to use a 3D-U-Net on a supervised dataset to capture a full spheroid context during training.

3.4.1 CycleGAN

For our problem, we employed the original implementation of CycleGAN [9] as our baseline. Compared with the neural networks presented below, the architecture of this model remained

unaltered. After conducting initial experiments, we observed that the model's performance was more sensitive to the input data than parameter adjustments. Therefore, we kept the parameters fixed for subsequent experiments. This section outlines the network parameters used for training the CycleGAN model. The Mean Squared Error (MSE) function was used as the objective function, with Adam as the optimizer. All models were trained for 200 epochs. The number of filters in the last convolution layer was set to 64 for both the generator and the discriminator. We used a PatchGAN discriminator with a patch size of 70x70 pixels. For the CycleGAN generator, we adopted a ResNet-based generator [9] architecture consisting of 9 ResNet blocks and downsampling/upsampling operations inside. In this implementation, the ResNet block comprises two padding and two convolutional layers, with the input added to the final convolution layer's output via the residual connection.

3.4.2 CycleGAN with conditional planes

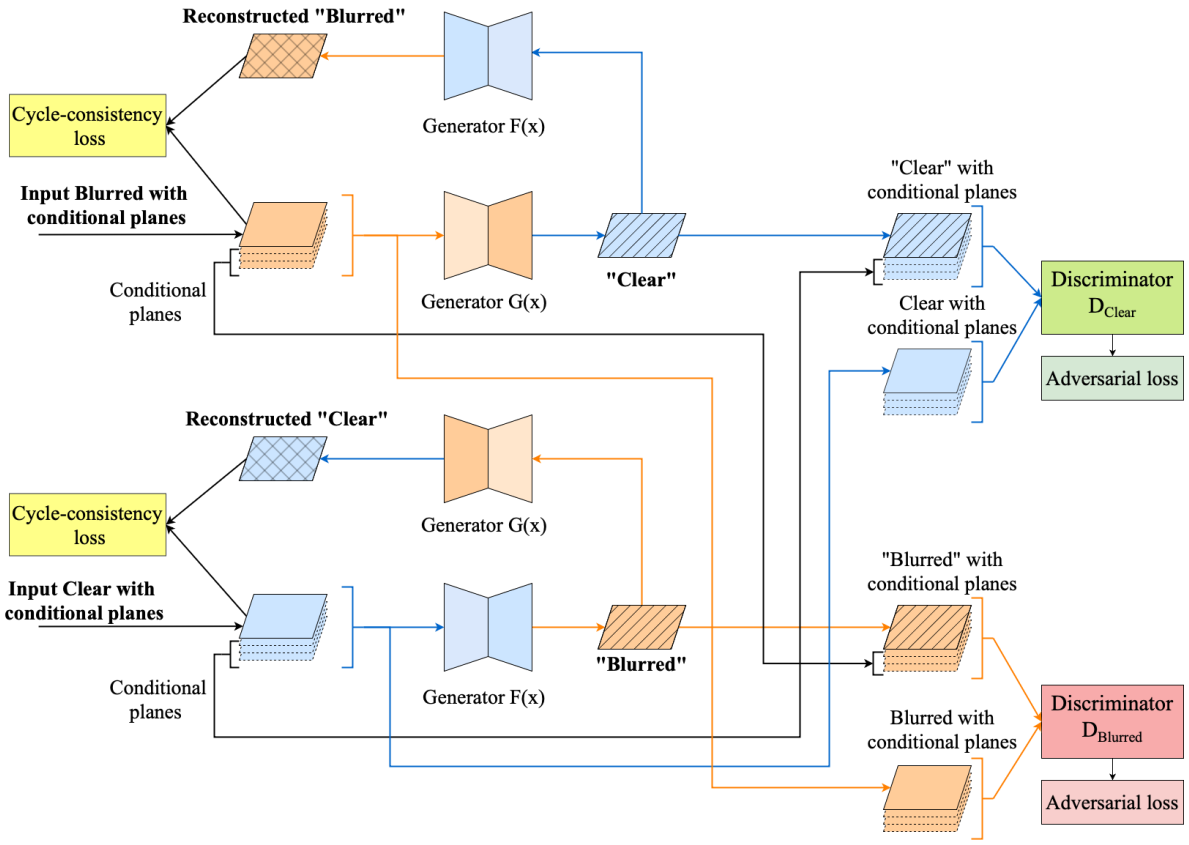

Figure 10. CycleGAN with conditional planes framework. Parallelograms represent images. Parallelograms with hatches on the perimeter represent conditional planes. Parallelograms with diagonal lines inside stand for generated images, while the figures with crossed lines represent reconstructed images. Generator G(x) translates images from the “Blurred” domain into “Clear”, while generator F(x) translates images from the “Clear” domain into “Blurred”. Both discriminators try to distinguish between real and generated images using conditional planes. Cycle-consistency loss is calculated using only the reconstructed plane and original target plane.

Based on our previous observations and results obtained from CycleGAN, it became evident that 2D images cannot capture information regarding how the blur changes as we move from one plane to another. Thus, we hypothesised that context from previous planes might contain information regarding the blur propagation and decided to transition from 2D to 2.5D context. In practical scenarios, the model must deblur the target image using only previous images as contextual planes. Therefore, we utilised images located before the target plane as contextual planes. The proposed idea involves feeding the target plane and a couple of contextual planes to the generator, which would generate only the target image. Similarly, the discriminator also takes contextual and target images, distinguishing the generated image from the real one. However, contextual planes are used solely as additional information, and their output remains unchanged. We employed the same parameters as those used in CycleGAN for our experiments. However, we modified the neural network architecture by changing the generator and discriminator network inputs to feed the target image with the conditional planes and output only the translated target image. During feeding the conditional planes into generators and discriminators, they remain unchanged. The cycle consistency loss was calculated without the conditional planes as they do not alter the results. The architecture is shown in Figure 10. We also modified the image sampling process to avoid selecting images from different spheroids.

3.4.3 3D-U-Net

Having the pairs of raw and corrected images, we were not obliged anymore to cut spheroids into halves to create a dataset. It allowed us to use the whole spheroid to capture more context during the model training.

Initially, a 3D-U-Net was implemented for a volumetric segmentation [11]. The main difference between a standard U-Net and a 3D-U-Net is using 3D convolutional layers instead of 2D. The implementation that we have used enabled us to adopt the original 3D-U-Net for our task. In the image translation task, the network would be trained to learn a mapping between images from raw images to the corrected ones using a loss function that minimises the difference between the predicted and ground truth images. To transform the model from a segmentation task into an image translation task, we changed the loss and objective function to MSE. To produce grayscale images instead of segmentation masks, we removed the final activation function and left only one output channel. The model was trained on 500 epochs. Before the training, the model preprocessed and augmented the images to improve its robustness. As the preprocessing step, the model applies the augmentations such as random flipping of the volumetric image and its random rotation at 90 degrees along the specified axes.

3.5 Evaluation Metrics for neural networks

Evaluating the performance of our models is a crucial step in the improvement of image quality. Two widely used metrics for evaluating the quality of image generation and reconstruction are the Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index

Measure (SSIM). Having a paired supervised dataset, these metrics can be used for evaluating both the models trained on unsupervised and supervised data.

3.5.1 Peak Signal-to-Noise Ratio (PSNR)

Peak Signal-to-Noise Ratio (PSNR) is a metric used to evaluate the quality of images or videos after they have been compressed or distorted. PSNR measures the difference between the original, uncompressed or undistorted image and the compressed or distorted version in terms of mean square error (MSE). It is expressed in decibels (dB).

PSNR is widely used in image and video compression as it provides a simple and objective way to compare the quality of different compression algorithms. It is also used in image restoration to evaluate the effectiveness of denoising, deblurring, or other restoration techniques.However, PSNR has some limitations. It only considers the mean squared error between the original and distorted images and does not consider the human perception of image quality. Therefore, images with high PSNR may not necessarily be perceived as high quality by human observers. Additionally, PSNR is sensitive to small changes in pixel values, which may not be significant to the overall image quality.

3.5.2 Structural Similarity Index Measure (SSIM)

SSIM is a widely used method for measuring the similarity between two images. SSIM measures [12] the quality degradation of a distorted image by comparing it to a reference or original image. It is calculated by considering three components of image information: luminance, contrast, and structure. The luminance component represents the image's overall brightness, the contrast component measures the differences in brightness, and the structural component considers the patterns of pixels in the image.

To calculate SSIM, two windows, one for the reference and the other for the distorted image, are selected. The windows are moved over the entire image, and for each window, the mean, variance, and covariance of the luminance, contrast, and structure components are computed. The SSIM index is then obtained by taking the product of the three components, where the weights are chosen based on the expected visual importance of each component [12]. The SSIM index ranges from -1 to 1, with 1 indicating perfect similarity between two images and -1 indicating maximum dissimilarity. A higher SSIM value indicates that the distorted image is closer to the original image in terms of structural information. Hence, the quality of the image is considered to be better.

SSIM has been shown to correlate well with human perception of image quality, and is widely used in evaluating the performance of image processing algorithms, including neural networks. It has advantages over other metrics like PSNR because it considers the structural information of the image rather than just the pixel-wise differences.

3.6 Thesis Writing

To write this thesis, we also used such services as Grammarly and ChatGPT. Grammarly is an AI-enabled cloud-based typing assistant. It can check the text for grammar, spelling and punctuation, clarity, engagement and delivery mistakes, proposes a replacement for found errors and detects plagiarism [13]. We used Grammarly to correct grammar mistakes, rearranging words in a sentence for better clarity and correctly placing punctuation marks. ChatGPT is a chatbot based on the Generative Pre-trained Transformer 3.5 (GPT-3.5) large language model [14]. It can solve tasks such as question answering, text generation, text summarisation, and text translation. We utilised the ChatGPT to paraphrase the sentences more academically, summarise the articles later checked and used as a basis for the thesis, and answer questions related to the thesis topic.

4 Experiments and Results

In this section, we present the experiments conducted to test our hypotheses, which are grouped based on the neural network architectures employed. The models' performances were evaluated based on the one spheroid from the supervised dataset serving as the test set. This test set comprises 114 pairs of raw and corrected images. The remaining spheroids were used for training the 3D-U-Net and not used elsewhere. The evaluation metrics used were peak signal-to-noise ratio and structural similarity index measure. We used SSIM and PSNR between the corrected and raw images of the test spheroid as the baseline for all models. It is important to note that the experiments conducted on unsupervised data were qualitatively evaluated through visual comparison since numerical measurements were not feasible until the supervised dataset was obtained. Finally, we compared the results of each model and determined which model better solved our task based on the experiments conducted.

4.1 CycleGAN experiments

This section outlines the experiments conducted with CycleGAN and their results. Our primary goal was to investigate how the input data affected the model's performance. Due to limited training data, we employed all the image augmentations presented in Chapter 3.2.2 to determine whether they would enhance the image deblurring process. In subsequent experiments, we continued to use these augmentations for comparative analysis. As previously mentioned in the Data and Methods section, all models were trained using the same parameters. Our experiments mainly focused on comparing the model's performance with increasing amounts of data and data augmentation.

During our experiments with CycleGAN, we did not have access to the supervised dataset. Therefore, our assessments were based on qualitative analysis and visual comparisons of original and generated images. After obtaining the supervised dataset, we evaluated our experiments quantitatively using a test set spheroid.

4.1.1 Initial experiments

In the initial experiments, we selected 12 spheroids with round shapes to train the CycleGAN forming a training set consisting of 655 pairs for the "clear" and "blurred" domains. Our aim was to assess the model's ability to produce satisfactory results using perfect-case data. However, as the target spheroids were still surrounded by neighboring cells that could interfere with the model's performance, we introduced image augmentations to minimize their impact on the target crops. The example of models outputs and comparison with raw and corrected data is shown in Figure 11.

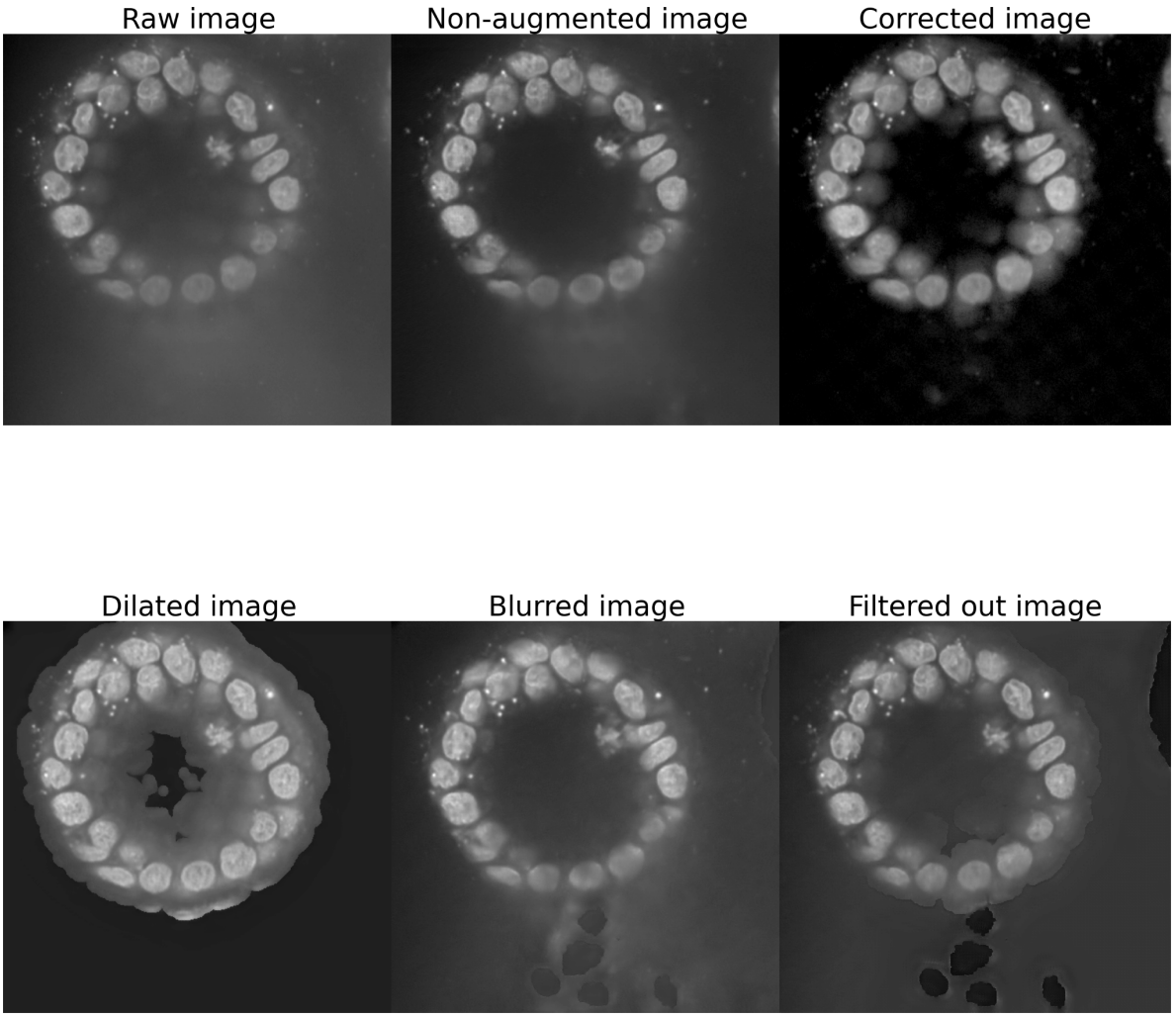

Figure 11. Performance of the initial experiments models. Raw and Corrected images represent the input data and the expected output data. Non-augmented image stands for the model that was trained on data without any augmentations. Dilated, Blurred and Filtered out indicate the augmentation approaches that were utilissed.

The results of the evaluation are depicted in Figure 12, where we trained the models using both raw data and image augmentations. Overall, we observed that at least one method was approaching the baseline performance. The PSNR graph shows that the models trained on non-augmented data and the "dilated" approach outperformed the baseline results on individual planes, whereas the model trained on the other augmentation images barely surpassed the baseline on only a few images. Nevertheless, none of the models completely outperformed the baseline in PSNR. A similar trend was observed for SSIM, where only the models trained on non-augmented and "dilated" approaches managed to outperform the baseline values on some planes. Although the "dilated" and non-augmented trained model demonstrated better SSIM performance, they did not exceed the baseline.

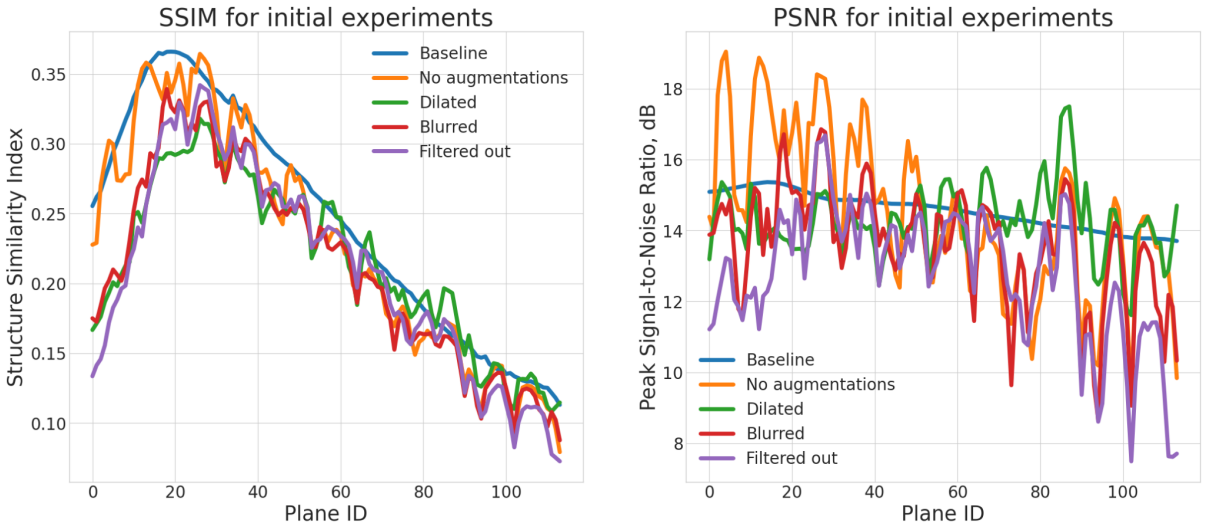

Figure 12. Evaluation metrics for initial experiments. Both graphs show metrics for the models trained on non-augmented, “Blurred”, “Dilated”, and “Filtered out” approaches. The left graph shows the Structure Similarity Index measurements, while the right graph shows the Peak Signal-to-Noise Ratio measurements of the models.

4.1.2 Experiments with central planes

During the initial experiments, it was observed that the blur did not affect the central planes near the largest plane, and these planes in the blurred part of the spheroid appeared as clean as those in the clear part of the spheroid. This observation led to the idea of removing the central planes to prevent model confusion during training. In this regard, we utilized the datasets from previous experiments and removed the three central planes from each spheroid half, reducing the dataset from 655 to 619 image pairs. To maintain the balance between the two domains required for training, we also removed the central planes from the “clean” part of the dataset. However, for these experiments, we did not evaluate the model trained on “dilated” data as it produced blank black images. The results of the models testing is shown in the Figure 13.

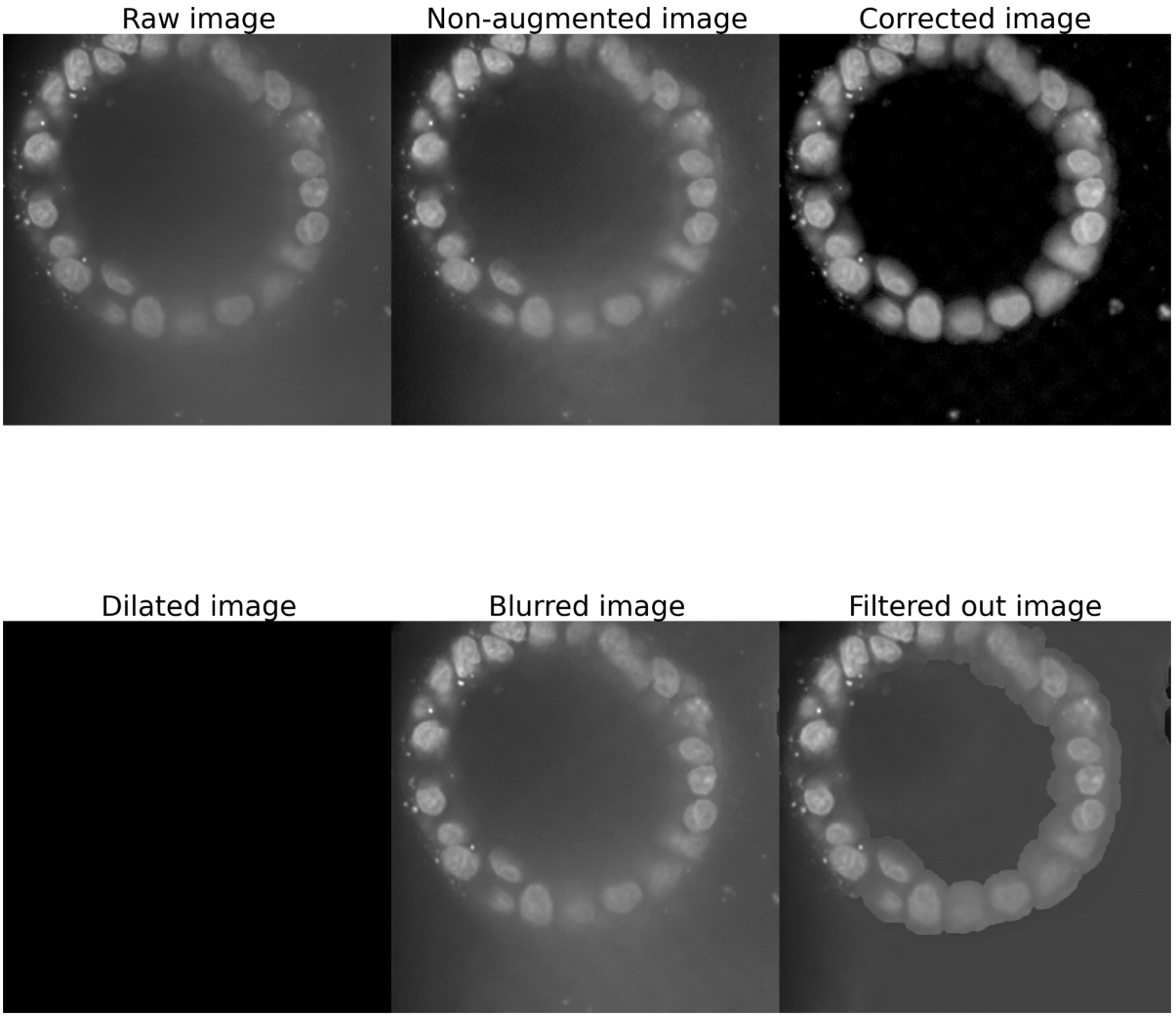

Figure 13. The results of deblurring using the models trained on the dataset without central planes. Raw and Corrected are the images used for our models evaluation. All other images are the results of generation made using CycleGAN. The model trained on Dilated data produced blank images during this experiment.

The results of the models' performance trained on these datasets are presented in Figure 14. It was observed that the performance of the models slightly decreased, as indicated by the PSNR and SSIM results, showing that none of the models outperformed the baseline, and the number of planes where the models exceeded the baseline values decreased. The CycleGAN trained on non-augmented data generated images structurally closer to the baseline but still did not reach its values. The models trained on data with other types of image augmentations also did not improve their results, even if they had exceeded the baseline values in previous experiments. However, the difference in results compared to previous experiments has decreased.

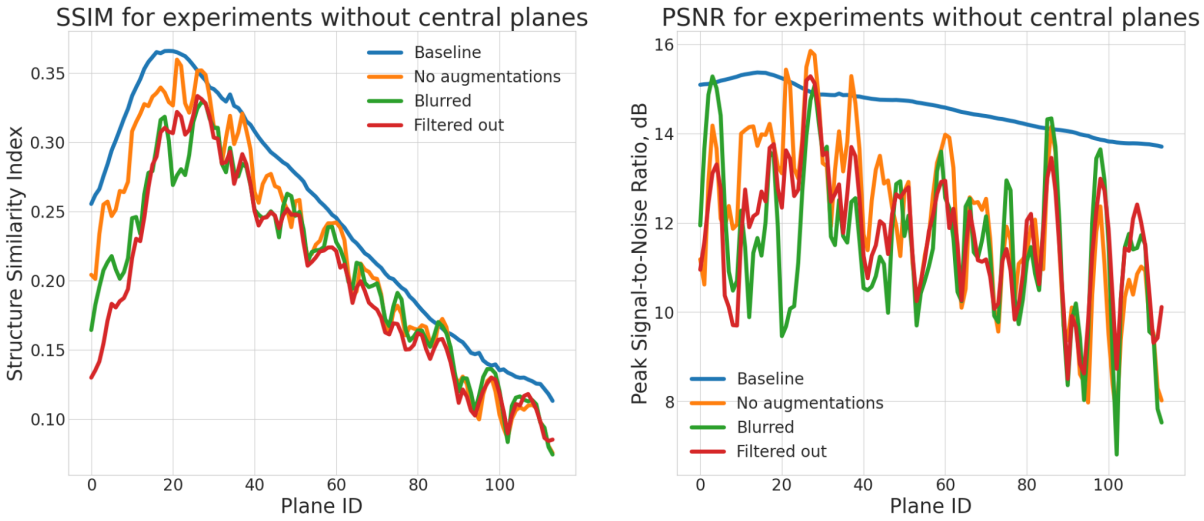

Figure 14. The results of models trained without central planes. SSIM (left) and PSNR (right) show the performance of the models trained on non-augmented, "Blurred", "Dilated", and "Filtered out" pre-processing approaches.

4.1.3 Experiments with a bigger amount of spheroids

The poor results from previous experiments we observed gave us an understanding that the model lacked data. We realised that such a limited amount of data does not provide the model with proper information about the diversity of spheroid shapes and blur behaviour on different cells and planes. It affected the model's robustness as well. This realization led us to increase the number of spheroids used for training from 12 to 28 out of 31 to improve the model's robustness. One spheroid was excluded as it did not have a blurred half, and we did not have the proper amount of planes to make a pair, while the remaining two were used for testing and visual analysis. The increase in the dataset size from 618 to 1295 image pairs allowed for more effective training of the model.

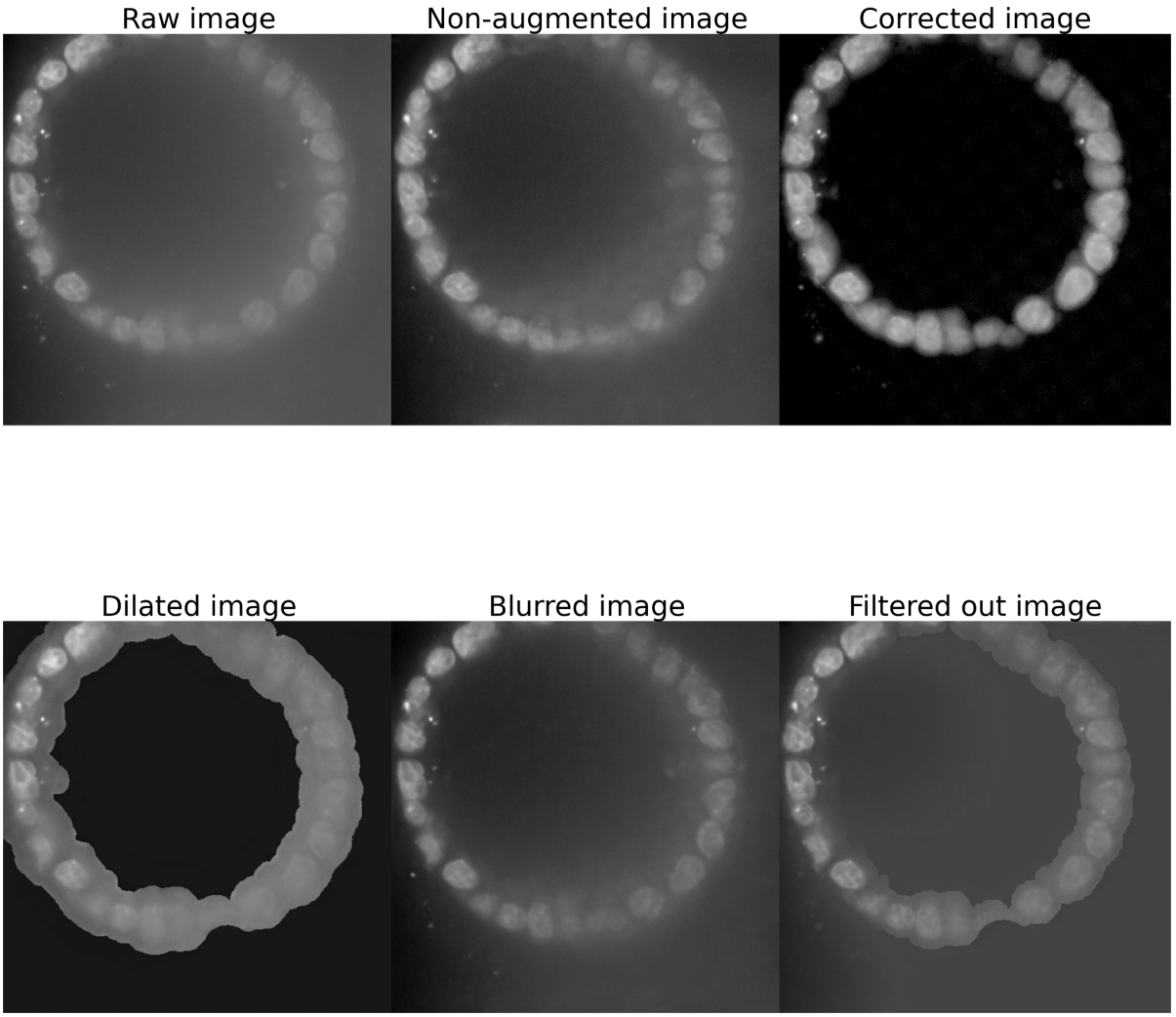

Figure 15. Results of image translation using the models trained on the bigger amount of spheroids. The plane of is located in the rear part of spheroids, so the cells are surrounded by blur. The image translation held using models trained on Dilated, Blurred, Filtered out and non-augmented images.

The impact of the increased dataset size on the performance of the models is presented in Figure 16. While the models did not exceed the SSIM and PSNR threshold, some models trained on "blurred" and "dilated" augmentations generated some images that surpassed the baseline. The best performance in terms of SSIM was achieved by the model trained on a "dilated" augmentation dataset, with a peak value of 0.42 for plane X. The models trained on non-augmented, "blurred," and "filtered out" images showed only a slight improvement. In terms of PSNR, the model trained on "dilated" augmented data outperformed other models, with a peak value of 18.3 compared to peak values of 18.1 for the model with non-augmented data, 17.1 for the model with "blurred" augmented data, and 17 for the model with "filtered out" augmented data.

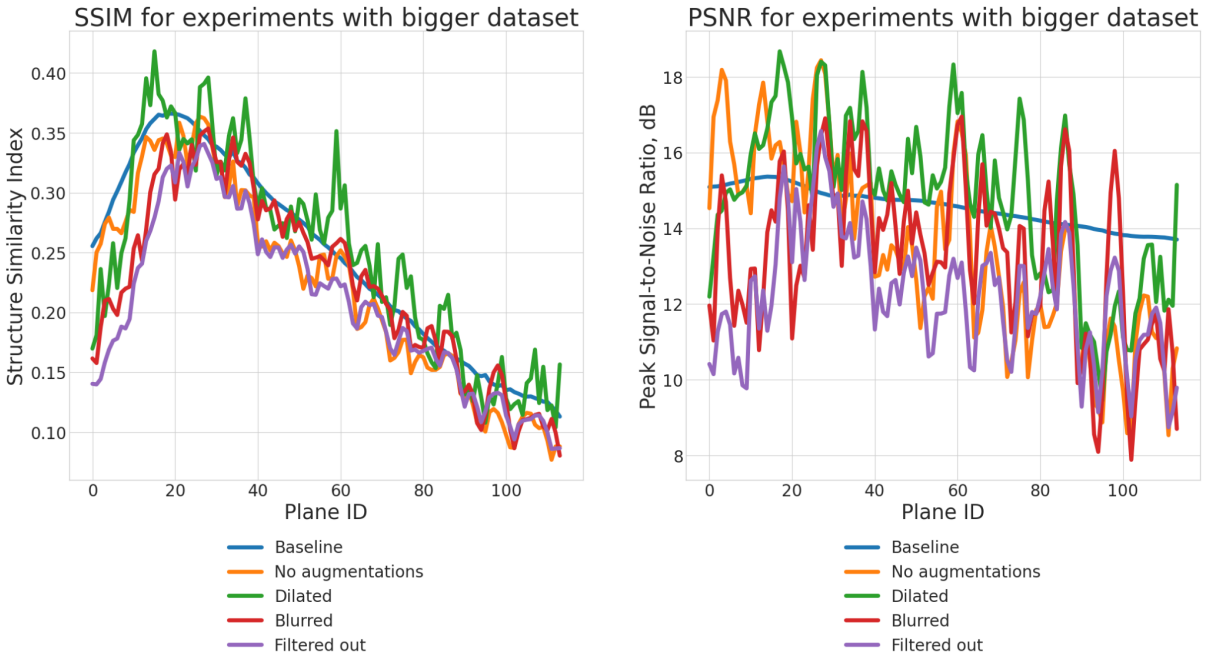

Figure 16. The results of models trained using a larger amount of spheroids. Both graphs show metrics for the models trained on non-augmented, "Blurred", "Dilated", and "Filtered out" approaches. The left graph represents SSIM score, while the right one describes PSNR on the test images.

4.1.4 Experiments with the rebalanced amount of spheroids

In order to ensure the effectiveness of our model deblurring, we realized that our previous testing approach using only two spheroids was insufficient. Moreover, the spheroids in the test set had a round shape and were isolated without neighboring spheroids. To address this, we expanded the training dataset to include 26 spheroids and the test set to include 4 spheroids, with the addition of two spheroids containing neighboring spheroids that were cropped due to their location on the image border. This enabled us to evaluate the model's performance not only on single spheroids but also on images with multiple spheroids adjacent to each other. The comparison of the models performance is shown in Figure 17.

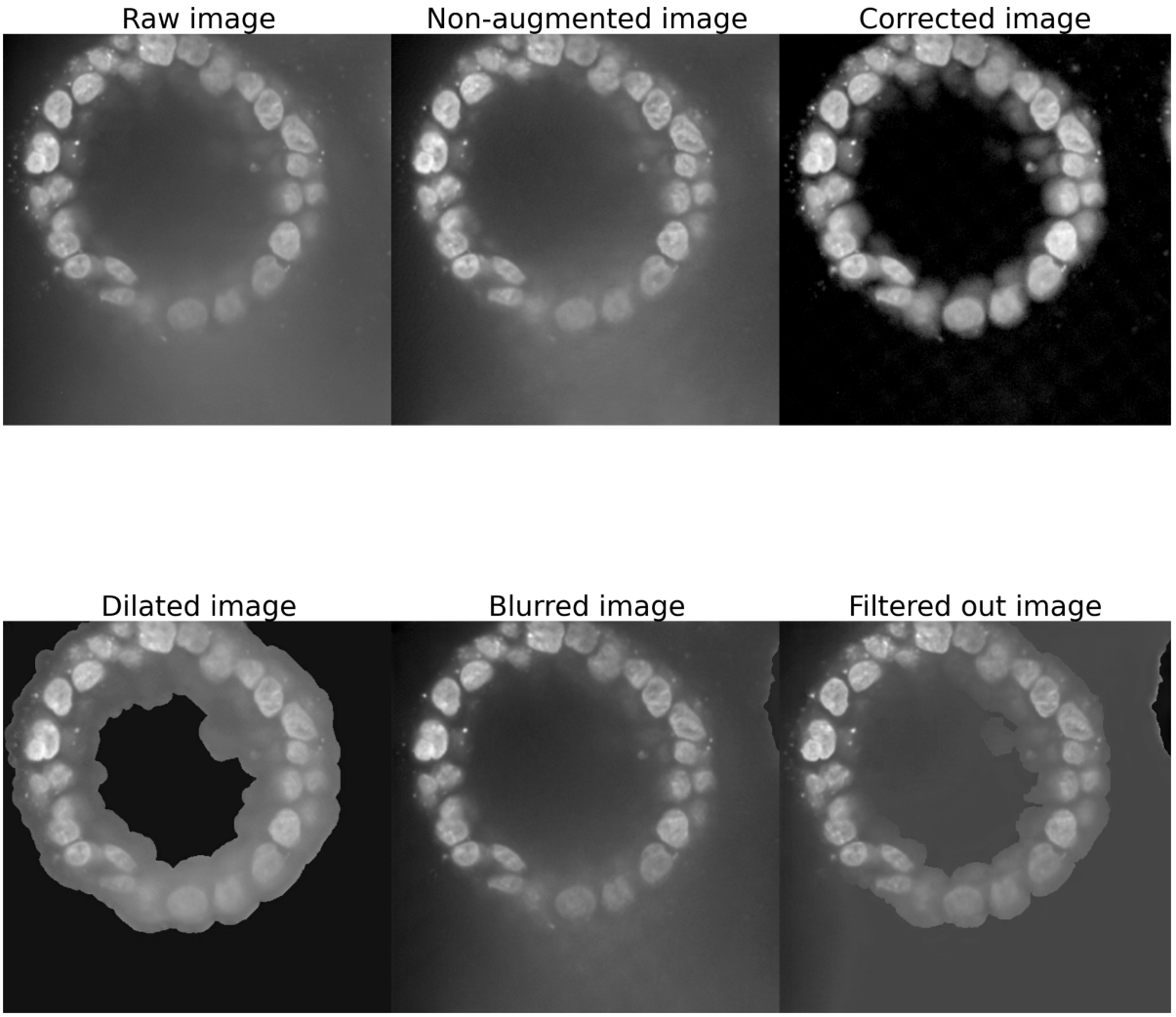

Figure 17. Comparison of the image deblurring attempts using the models trained on the image augmentations and non-augmented images. Raw and Corrected images were used for a better comparison between other results.

However, we observed a slight decrease in performance during these experiments, as shown in Figure 18. The results were similar to our previous experiments, where most models failed to outperform the raw and corrected image measurements on most planes. Notably, the model trained on "dilated" augmentations showed the most significant decrease in performance, possibly due to the fact that this augmentation had the most significant impact on the data and thus affected the model's performance. Additionally, we observed that the results of "dilated" augmented data only increased significantly after we doubled the dataset size. Thus, even a seemingly small change in the dataset could lead to performance degradation.

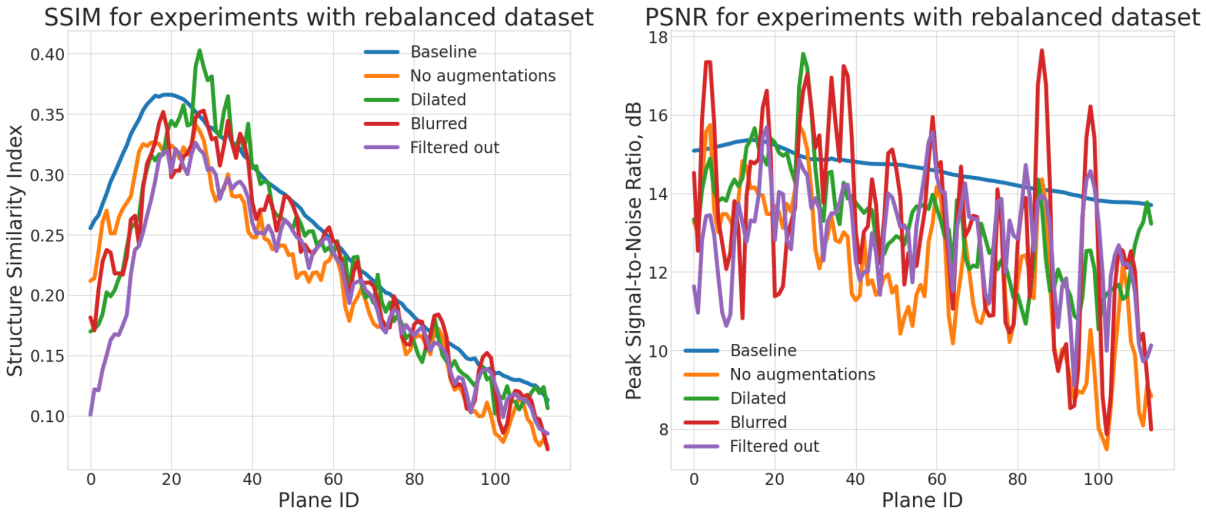

Figure 18. The results of models trained using a rebalanced amount of spheroids. Both left (SSIM) and right (PSNR) graphs show metrics for the models trained using data with non-augmented, “Blurred”, “Dilated”, and “Filtered out” approaches.

4.2 CycleGAN with conditional planes experiments

This section presents the experimental setup using CycleGAN with conditional planes. We utilized the dataset obtained from the experiments with a rebalanced amount of spheroids since it provided a sufficient amount of data for training and testing. We chose the non-augmented dataset because it contains the surrounding area of the spheroid and has more representative data for real-life scenarios. Although the dataset with augmentations showed better results, it lacks the surrounding area of the spheroid. We conducted our initial experiments based on visual comparison and qualitative analysis and later evaluated them on the spheroid from the supervised dataset, similar to the original CycleGAN experiments. In the following sections, we detail the parameters we modified during the experiments, including the number of conditional planes, and improvements to the sampling procedure.

These experiments primarily focused on the impact of the number of conditional planes on the model's performance. An increase in the number of conditional planes led to a decrease in the number of generated images. This reduction occurred because all the planes with an index less than the set number of conditional planes were used as conditional images but not as target ones. We describe the changes in the sampling procedure in the sections where they had an impact on the results.

4.2.1 First experiment

The initial experiment aimed to verify the proper functioning of the model. Specifically, we set the number of conditional planes to three to test the sampling process. The deblurring performance of the model is show in Figure 19. The measurments of this evaluation are presented in Figure 20. The CycleGAN with conditional planes failed to meet the baseline in any of the metrics.

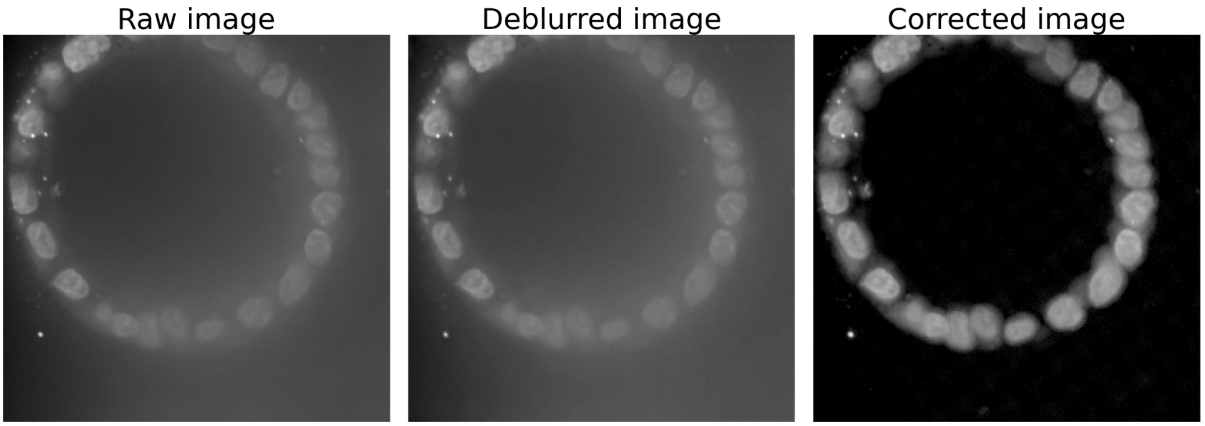

Figure 19. The first deblurring task for the CycleGAN with conditional planes. The deblurred image was generated by model, while the generated image is surrounded by raw image on the left and corrected image on the right.

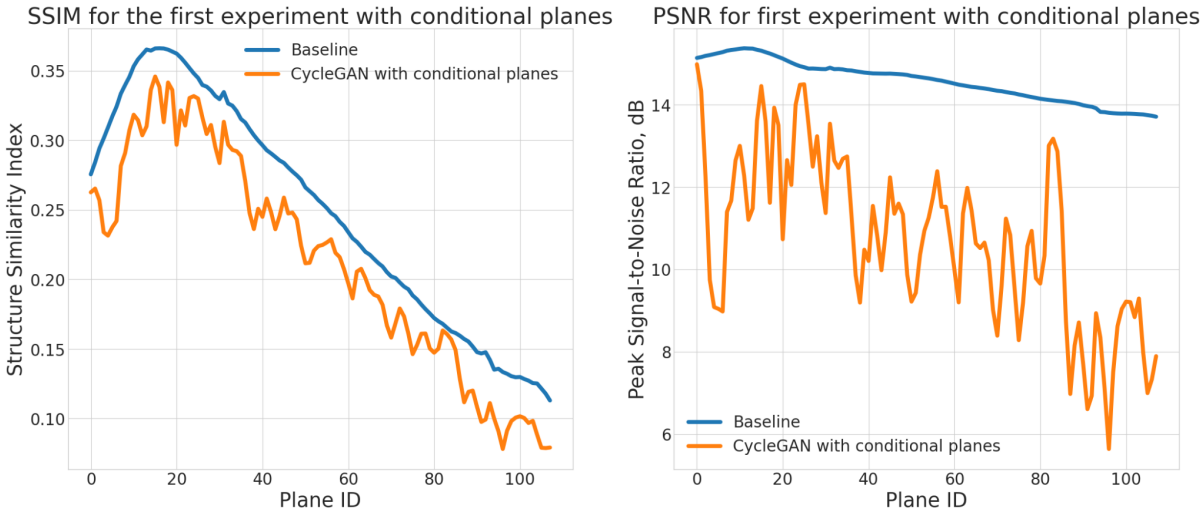

Figure 20. First evaluation of CycleGAN with conditional planes. The baseline is the difference between the corrected and raw images. The left graph represents SSIM while the right represents PSNR.

4.2.2 Experiments with loss weights

Our subsequent experiment aimed to investigate the impact of weight changes in the loss function on the model's performance. We sought to determine whether increasing or decreasing the weights would enhance the model's performance, while the generator and discriminator inputs are altered. For this experiment, we employed a single conditional plane to assess the effect on the minimum number of additional planes. The performance of these models is depicted in Figure 21.

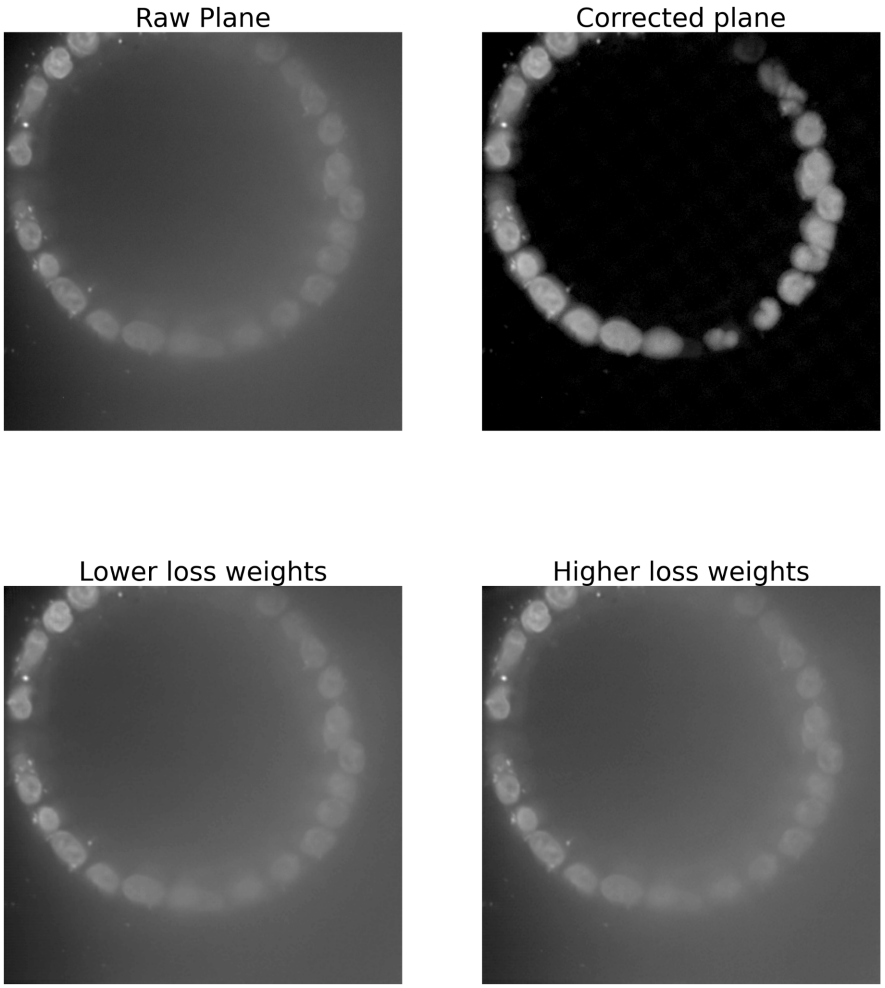

Figure 21. Deblurring performance of the models with different weights for loss function.

In the first approach, we increased the weights from the default value of 10 to 11.5. Conversely, in the second approach, we decreased the weights from the default value of 10 to 8.5. The measurements are presented in Figure 22. The performance of both models was consistent with the first experiment, wherein each model had only one conditional plane. Notably, the difference between the two models with different weights was insignificant. Although the model with smaller weights outperformed the model with larger weights, the improvement was minimal. Nonetheless, neither of the models could generate an image that surpassed the similarity between raw and corrected images.

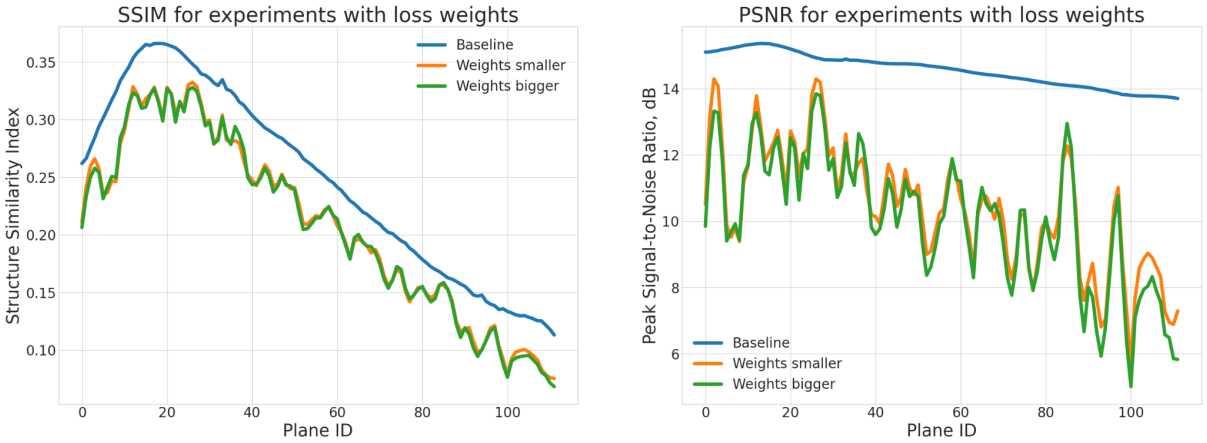

Figure 22. Evaluation of the models with different loss weights. Weights smaller — the model that was trained with smaller weights for cycle loss. Weights bigger — the model that was trained with bigger weights for cycle loss. The left graph shows SSIMs, while the right one shows PSNR of the models and the baseline.

4.2.3 Experiments with a number of conditional planes

In this section, we sought to investigate the influence of the number of conditional planes on the model's performance. To ensure that the sampling procedure did not select parts of other spheroids, we modified the image names to include information about the spheroid and its plane. We implemented a sampling procedure based on their naming convention that checked the plane ID and compared it with the number of conditional planes. We trained the models with one, three, and five conditional planes. The performance of these models on the deblurring task is depicted in Figure 23.

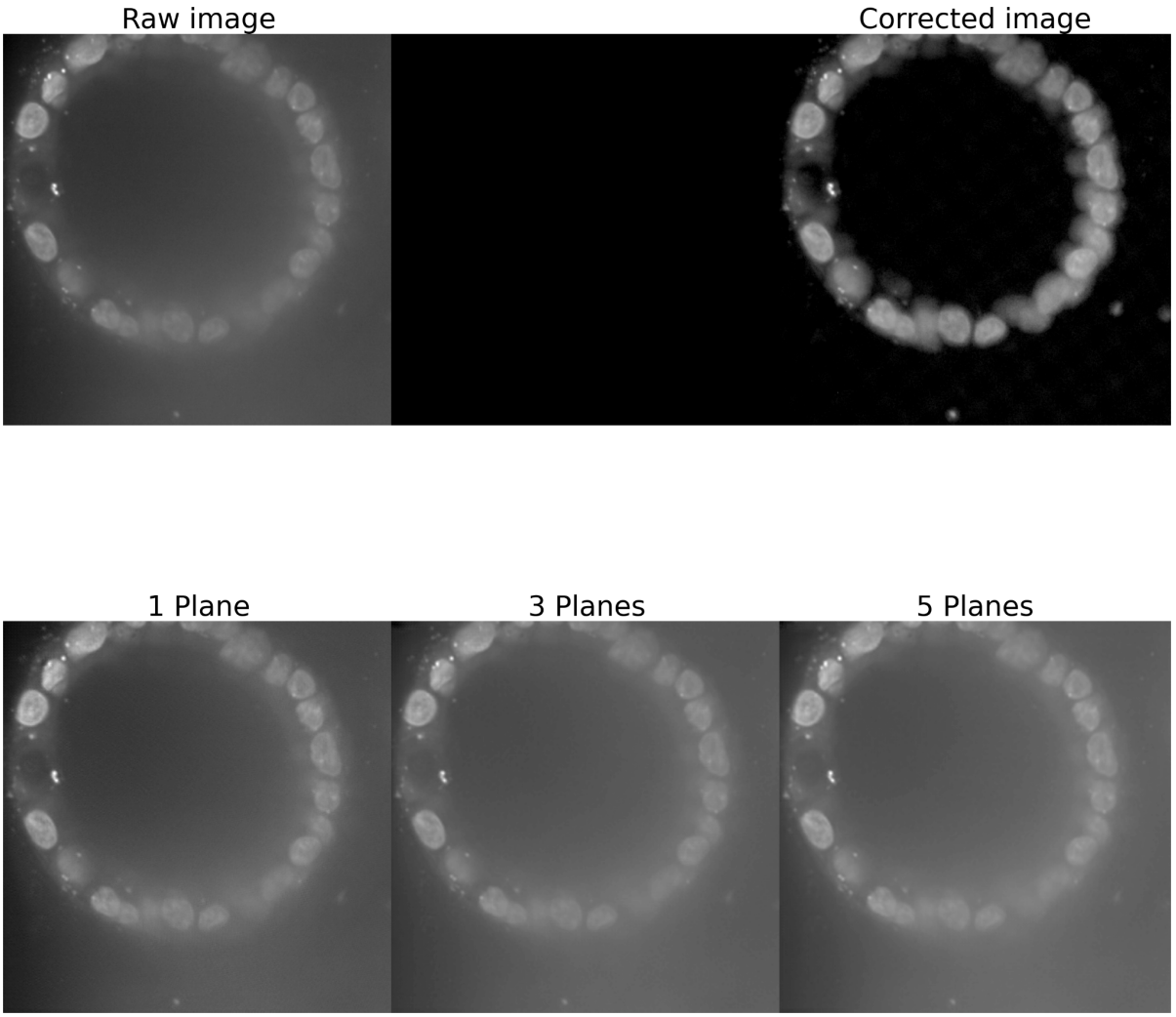

Figure 23. Performance of the models trained on the different number of conditional planes. 1 Plane means that during the sampling model used 1 conditional plane together with target image. The same applies to 3 Planes and 5 Planes with respectful change in number of conditional planes.

The evaluation results of the models are presented in Figure 24. Interestingly, the model trained with three conditional planes produced results similar to those of the first experiment, where three conditional planes were also used for training. It can be observed that the increasing number of conditional planes did not improve the image generation quality beyond a certain point. The results show no significant improvement over the increasing number of conditional planes. Conversely, the model trained with only one conditional plane outperformed the other models but still failed to pass the baseline results. This indicates that increasing the number of planes does not enhance image generation but rather has the opposite effect. Based on the graph lines, we conclude that the optimal number of conditional planes is one. In the following experiments, we explore this assumption further.

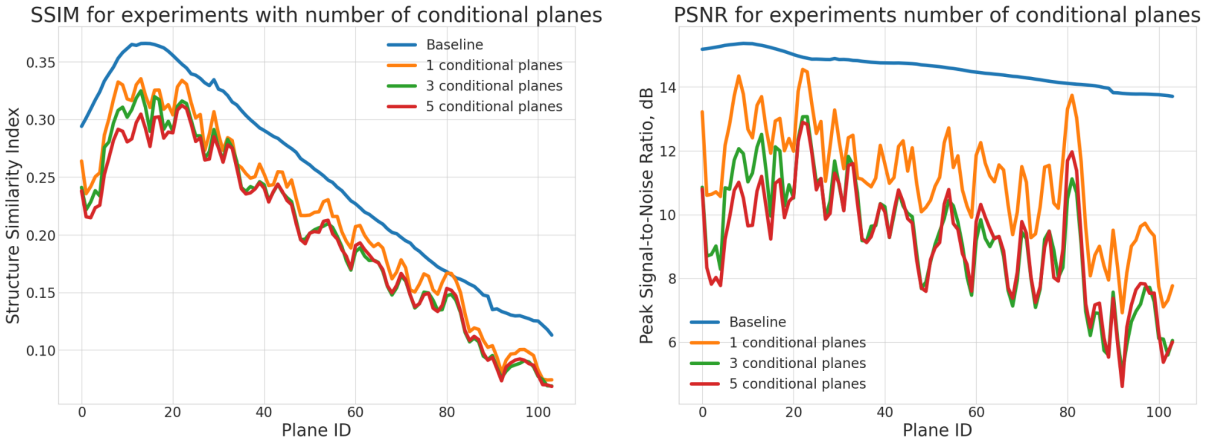

Figure 24. Evaluation of the models trained using a different number of conditional planes. The legend indicates the number of conditional planes used for model training. The left graph shows the Structure Similarity Index measurements, while the right graph shows the Peak Signal-to-Noise Ratio measurements of the models and the baseline.

4.2.4 Experiments with CycleGAN generators

In the original CycleGAN [9] , the ResNet blocks were utilised as the generator. However, we deemed it necessary to explore alternative generator architectures to provide a more comprehensive analysis. Hence, we chose to implement the U-Net generator in combination with the PixelGAN discriminator for the CycleGAN with conditional planes. The PixelGAN discriminator is similar to the PatchGAN discriminator, but instead of comparing image patches of size 70x70 or any other sizes, it compares two images using a patch size of 1 pixel. We trained both models with one conditional plane, with the first model using a 70x70 pixel PatchGAN discriminator, and the second model using the PixelGAN discriminator. The performance of the models are presented in the Figure 25.

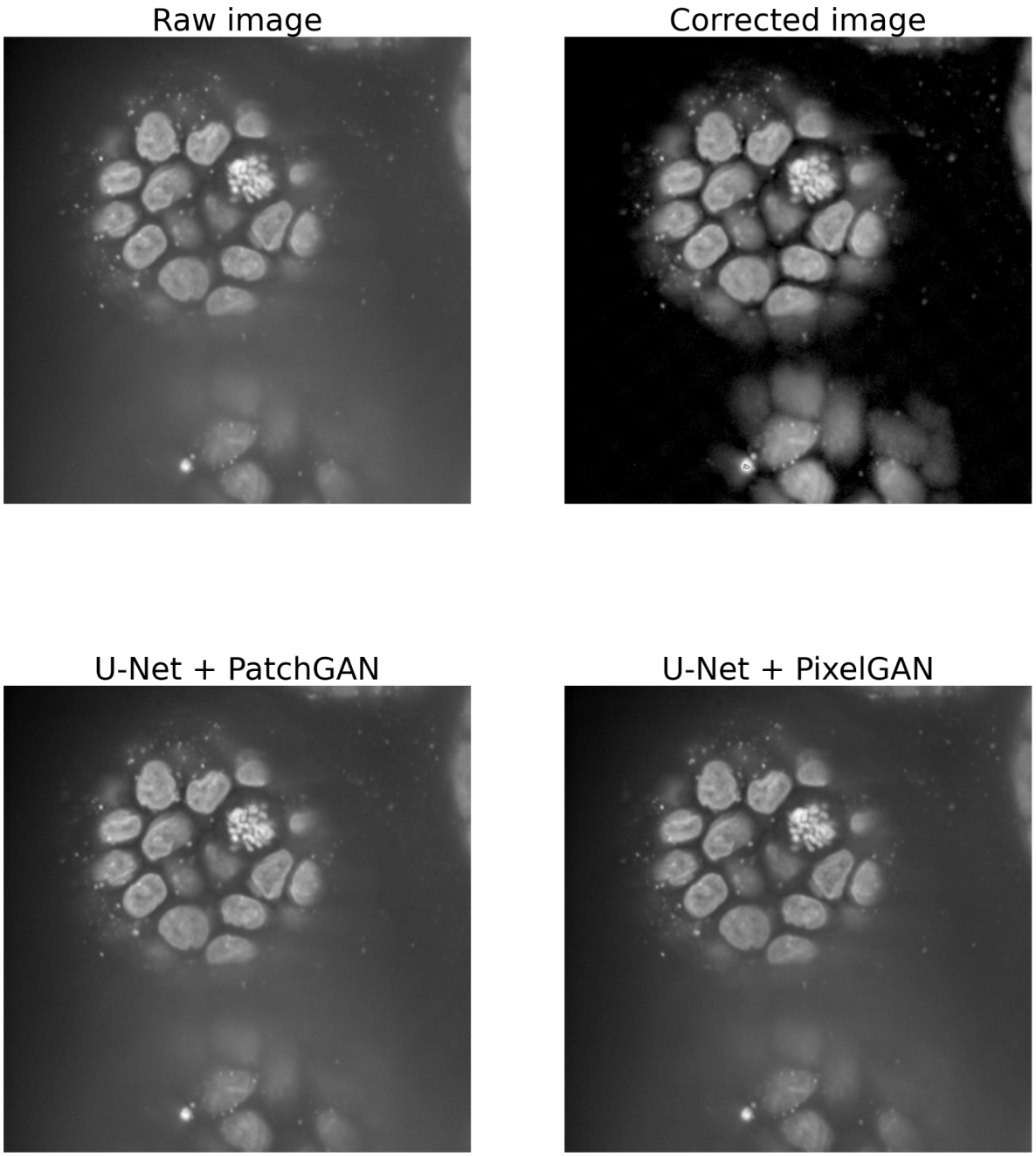

Figure 25. Result of CycleGAN with conditional planes and U-Net-based generators performance. U-Net + PatchGAN means that the model utilises U-Net generator and the PatchGAN with 70x70 pixel patches. U-Net + PatchGAN indicates that the model used U-Net generator with PixelGAN as discriminator.

The evaluation results of these models are presented in Figure 26. While both models were close to surpassing the baseline of SSIM, neither the U-Net generator with PatchGAN discriminator nor the U-Net generator with PixelGAN discriminator could achieve the same results as the PSNR value between the raw and the corrected images. Our measurements indicate that the model with the PixelGAN discriminator slightly outperforms the model with the PatchGAN discriminator, but the difference is not significant.

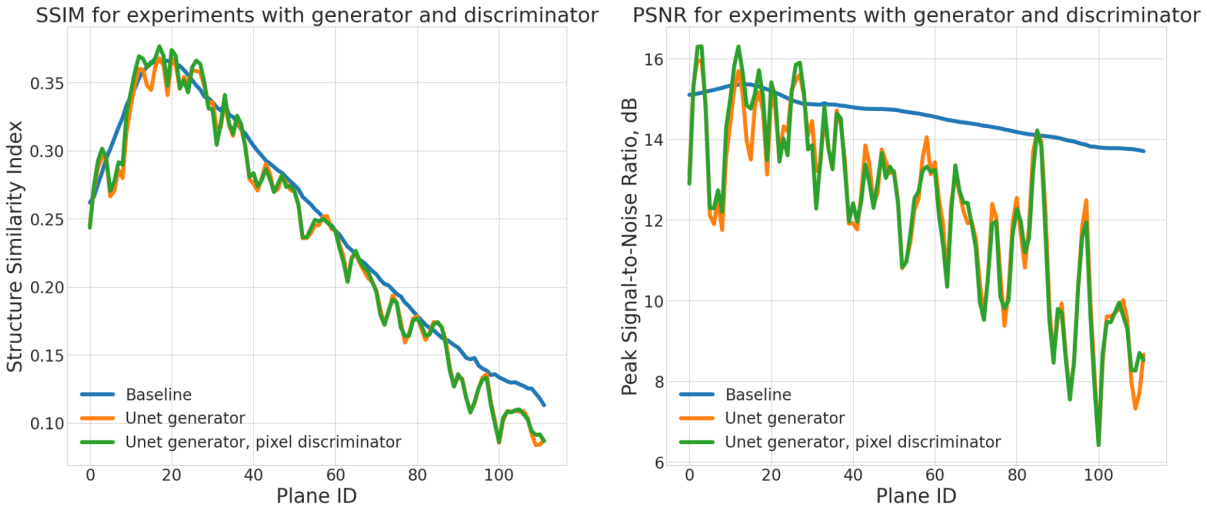

Figure 26. The results of the models' performance were trained using different discriminators. U-Net generator — a model that used U-Net as a generator and 70x70 pixels PatchGAN as a discriminator. U-Net generator, pixel discriminator — a model that utilised U-Net as a generator and PixelGAN as a discriminator. The left graph represents SSIM, while the right shows PSNR.

4.3 3D-U-Net experiment

Since the supervised dataset consists of pairs, the small number of spheroids is compensated by the number of planes from each spheroid. One of the seven spheroids in this dataset was used to run our experiments and test all models. We trained a 3D-U-Net using images from five spheroids and validated the model using images from the remaining spheroid. The training set consisted of 843 image pairs, while the validation set contained 119 image pairs. The visual comparison of the model's performance compared to raw and corrected images is presented in Figure 27.

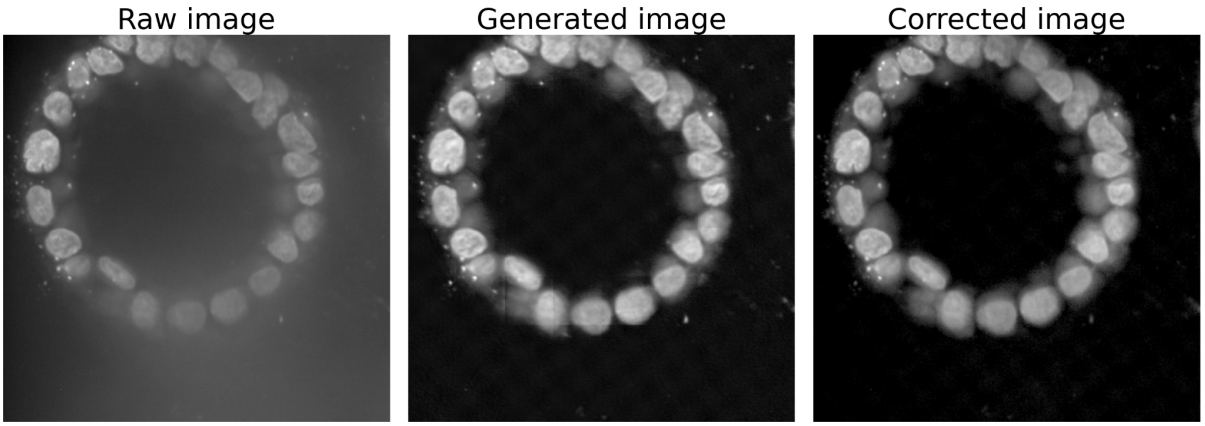

Figure 27. The visual comparison of the 3D-U-Net generation evaluated on the test spheroid. Raw — original image. Generated— result of 3D-U-Net deblurring on raw image. Corrected – the corrected image which is used as ground truth.

The results are summarized in Figure 28. The 3D-U-Net model showed significant improvements over unsupervised models, surpassing the baseline measurements on both metrics. In addition to the quantitative metrics, it is noteworthy to observe the visual enhancement in the quality of spheroids in the generated image compared to the raw image.

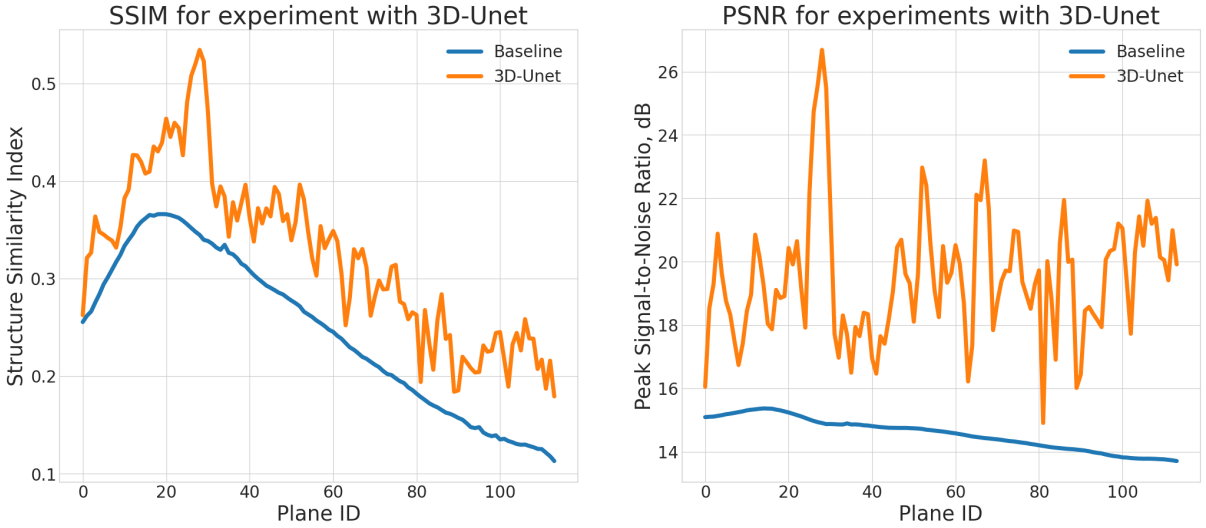

Figure 28. Results of 3D-U-Net performance. The left graph stands for SSIM, while the right graph depicts PSNR.

4.4 Results

After conducting a thorough set of experiments, we have determined that the 3D-U-Net trained on a supervised dataset outperforms the CycleGANs trained on unsupervised data. The experiments conducted in this study have demonstrated that more complex frameworks such as GANs may not always be capable of learning complex mappings between clear and blurred images without corresponding pairs. We have observed that the performance of the unsupervised model improves when one conditional plane is added during training and when using a U-Net-based generator. However, this approach only enables the model to produce images that reach the level of the baseline metric but do not surpass it. One reason for the low quantitative metrics of the unsupervised models is that the metrics are measured using corrected images that contain a background pattern that occurs after the correction procedure. While the 3D-U-Net was trained on pairs where the corrected images contain this pattern, its output also generates this pattern, which increases the similarity of images during the measurements.

In contrast, training on supervised data allows the production of a model capable of handling the deblurring task with a smaller number of corrected spheroids. Furthermore, training on 3D images enables the model to learn more complex mappings across all dimensions compared to using 2D images.

Discussion

For us, the biggest challenge in this work was the data we were working with. Due to the background pattern on the corrected spheroids, the proper evaluation of unsupervised models difficult as the models do not produce this pattern as they did not have such data in the dataset.

We conducted our experiments using the 3D-U-Net on supervised data; however, this architecture is not the only one that employs three-dimensional images to address biomedical image tasks. For example, a modified version of pix2pix known as vox2vox [15] utilized CT images of the brain to generate 3D models of tumors, which is a relevant application in the context of 3D culture deblurring. Utilizing a model that can effectively segment spheroids could enhance research efforts related to spheroids and simplify their analysis.

Another potential approach for image deblurring involves utilizing CycleGAN with a blur kernel [16]. This model was employed for microscopy image deconvolution, where the blur kernel refers to the point spread function (PSF) of the microscope, which characterizes the amount of blurring that occurs when imaging a point source with the microscope. Replacing one of the generators with a physical function or model capable of simulating the blur spread throughout the well may improve the performance of unsupervised models.

Conclusion

Spheroids are widely used in biomedical research due to their complex microenvironments and ability to simulate various biological processes. However, their structure poses challenges for image analysis due to the poor quality of obtained images. Traditional computer vision algorithms may be time-consuming and require a lot of manual fine-tuning to be applied effectively. In this thesis, we conducted a comparative study on unsupervised and supervised approaches for improving the quality of spheroid images using deep learning techniques. Our experiments showed that the supervised approach was more effective in improving visual quality while preserving important structural features. The results also demonstrated that the supervised model outperformed other unsupervised methods in terms of image quality metrics such as PSNR and SSIM.

The main contribution of our thesis is that a simpler model trained on a smaller supervised dataset of corrected and raw image pairs can generate high-quality images better than a more complex framework trained on unsupervised data. Also using 3D model enables learning mapping across z-stack as well. However, there are still limitations associated with the obtained supervised data. For instance, the corrected images contain a pattern generated by the correction algorithm, which can be easily removed by thresholding to eliminate its impact on the analysis. Deblurred spheroid images have the potential to enhance measurement accuracy and reliability in research, which can lead to the discovery of new treatments and therapies.

References

1. Gunti S, Hoke ATK, Vu KP, London NR Jr. Organoid and Spheroid Tumor Models: Techniques and Applications. Cancers. 2021;13. doi:10.3390/cancers13040874

2. Nürnberg E, Vitacolonna M, Klicks J, von Molitor E, Cesetti T, Keller F, et al. Routine Optical Clearing of 3D-Cell Cultures: Simplicity Forward. Front Mol Biosci. 2020;7. doi:10.3389/fmolb.2020.00020

3. In Vitro three-dimensional (3D) cell culture tools for spheroid and organoid models. SLAS Discovery. 2023. doi:10.1016/j.slasd.2023.03.006

4. Kapałczyńska M, Kolenda T, Przybyła W, Zajączkowska M, Teresiak A, Filas V, et al. 2D and 3D cell cultures - a comparison of different types of cancer cell cultures. Arch Med Sci. 2018;14: 910–919.

5. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2015. Available: http://arxiv.org/abs/1512.03385

6. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Springer International Publishing; 2015. pp. 234–241.

7. Isola P, Zhu J-Y, Zhou T, Efros AA. Image-to-Image Translation with Conditional Adversarial Networks. 2016. Available: http://arxiv.org/abs/1611.07004

8. Wang J, Zhao Y, Noble JH, Dawant BM. Conditional Generative Adversarial Networks for Metal Artifact Reduction in CT Images of the Ear. Med Image Comput Comput Assist Interv. 2018;11070: 3–11.

9. Zhu J-Y, Park T, Isola P, Efros AA. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. 2017. Available: http://arxiv.org/abs/1703.10593

10. Wolterink JM, Dinkla AM, Savenije MHF, Seevinck PR, van den Berg CAT, Isgum I. Deep MR to CT Synthesis using Unpaired Data. 2017. Available: http://arxiv.org/abs/1708.01155

11. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. 2016. Available: http://arxiv.org/abs/1606.06650

12. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process. 2004;13: 600–612.

13. How Grammarly Works. [Internet]. Grammarly.com. Available: https://www.grammarly.com/how-grammarly-works

14. Introducing ChatGPT. [Internet]. openai.com. Available: https://openai.com/blog/chatgpt

15. Crimi A, Bakas S. Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain

Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part I. Springer Nature; 2021.

16. CycleGAN With a Blur Kernel for Deconvolution Microscopy: Optimal Transport Geometry. Available: http://dx.doi.org/10.1109/TCI.2020.3006735

Appendix

I. Licence

Non-exclusive licence to reproduce thesis and make thesis public

I, **Denys Krupovych**,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to

reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,

Deblurring of microscopic 3D spheroid images using GANs,

supervised by **Dmytro Fishman** and **Mikhail Papkov**.

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 4.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.

4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Denys Krupovych

08/05/2023