Robustness of U-Net based models to common image artefacts

Andrea Oliveri, Célina Chkroun, Bernardo Conde
EPFL, Switzerland

***Abstract*—In this paper we analyzed the robustness of the U-Net model, included in the DeepImageJ plugin, normally used for pancreatic cell segmentation. We observed that this U-Net was extremely sensitive to gaussian noise in the image and changes in the cell size, as well as being less performant when faced with non-uniformly illuminated images. Deep learning models such as this one have very useful applications in cellular optical microscopy, but become unusable when such artefacts are present in the images to be segmented. After analyzing the impact of such common artefacts in optical microscopy, we proposed an improved version of this U-Net architecture, which was considerably more resistant to some of these common artefacts.**

I. Introduction

Image segmentation of cells is an extremely important problem in biological optimal microscopy [1]. The state-of-the-art techniques employed for this task use Deep Learning models [2], with one of the more effective ones at the time of this writing being U-Net [2]. Such deep learning models are known to perform very well on images similar to those used to train them, but may fail in generalizing to images visually-perturbed by artefacts common in real-life data [3]. This effect is amplified when using relatively small datasets for training, such as those often available in this research field [3]. In this paper we report our findings on the behaviour of an existing and widely used U-Net model, part of the DeepImageJ plugin and used for pancreatic cell segmentation, whose original implementation can be found at [4]. We test its behaviour when confronted with images containing strong non-uniform illumination, large gaussian noise and different cell sizes. We also show that, by retraining the U-Net with these types of perturbed images included in the training set, we can considerably increase its robustness against these common image artefacts.

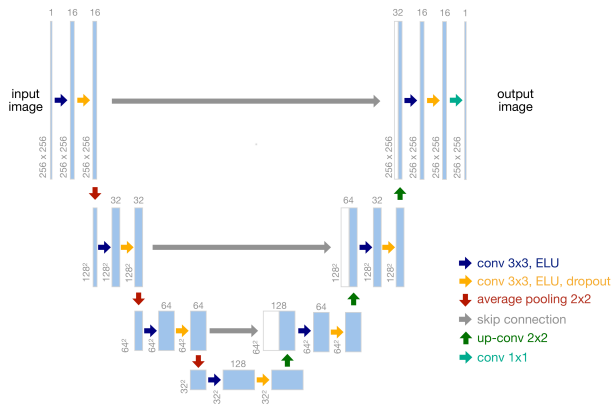

Figure 1. U-Net architecture used for this project. Image taken from [4].

II. Models and Methods

To achieve our goal of investigating the robustness of the U-Net architecture to different artefacts applied to images, we started from its original implementation [4]. We re-implemented the architecture shown in Figure 1 in Keras-Tensorflow, and re-used the original hyper-parameters and dataset.

The architecture shown in Figure 1 consists of two paths: a contracting path (on the left) and an expansive path (on the right). In the contracting path, each step consists of two convolutional layers followed by exponential linear unit (ELU) activations, the latter of which is also followed by a dropout layer and a 2×2 average pooling layer, with stride size of 2 to down-sample the data. Across different steps, the number of channels of the features is doubled. In the expansive path, each step performs a concatenation with the output of a layer in the contracting path, and that of a 2×2 transposed convolution, which reduces by a factor of 2 the number of features, but also increases the size of the image by the same factor. The result of the concatenation is then fed into two convolutional layers, each also using ELU activations, and then into a dropout layer. In the last step of the expansive path, a final 1×1 convolution reduces the channels of the output image into one channel only, and yields a segmentation mask having the same size as the input image.

This U-Net architecture uses ELU as an activation function after each convolutional layer. As it has been shown in previous works, this allows for faster training by removing the need of a Batch Normalization layer, and potentially better-performing models than rectified linear unit (ReLU) [5] [6]. The initialization of the convolution layers was done using He initialization as previous research showed its benefits over other initialization techniques [7] [8]. Dropout layers were added after each two consecutive convolutional layers, in order to reduce the likelihood of overfitting our training data [9]. We used the Adam optimizer, with default parameters, as they were observed to converge after only around 100 epochs. A batch size of 2 was chosen mainly due to limitations in memory available in the GPU used for training, but as shown in previous publications [10], despite it increasing the training time, this choice often allows us to obtain more robust models when using small training datasets such as ours.

The model was trained using the Model Checkpoint callback, as this prevents the model from overfitting the training data, and ultimately only stores the model which performs best on the validation set [11]. Early Stopping callback was also used, and was given a large patience value, as its role was essentially just to stop running after convergence rather than avoiding overfitting. The number of epochs chosen to train the model was chosen to be sufficiently large for Early Stop to consistently kick in before all epochs were run.

Three different data pre-processing techniques were applied to the training images, and three independent U-Net models were trained with same parameters and architecture. The first U-Net was trained using the same image pre-treatement as in [4], which consisted in cropping random patches, with the correct size, from the training image. The second U-Net was trained using simple data augmentation techniques, more specifically: rotations by multiples of 90 degrees, horizontal and vertical flips. These same transformations were also applied to the masks to guarantee coherence between training inputs and outputs. Finally, the third U-Net was trained using images with the same data augmentation techniques as the previous U-Net, but random 2D gaussians and gaussian noise were also added. A maximum distortion value was chosen so that the obtained images could still be considered to have a reasonable amount of artefacts, but it was a much lower value than the maximum amount of distortion used during the evaluation of the models.

Our study uses the same dataset as [4], which was provided at

[12] by Dr. T. Becker. Fraunhofer Institution for Marine Biotechnology, Lübeck, Germany in the scope of the Cell Tracking Challenge. The dataset is made up of 8-bit gray-scale 2D images of size 720×576 whose acquisition details can be found at [12]. It should be noted that only the images for which a gold-reference tracking annotation was provided were used in this paper.

The total number of images with a gold-reference annotation was 404. These images were split into 202 images used to generate the training set, 32 images used to generate the validation set and 170 used to generate the test set. To account for the difference in shape between the images and the input layer of the U-Net, 6 patches of size 256×256 were randomly extracted from each image.

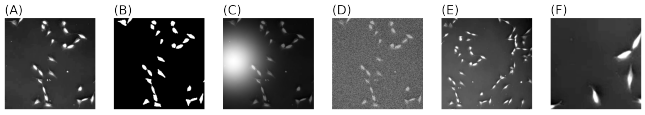

Figure 2. Example of a test patch and its corresponding mask, as well as an example of the artificial artefacts. (A) Example of patch from test image with no distortion. (B) True segmentation mask corresponding to (A). (C) Artificially non-uniformly illuminated version of (A). (D) Artificially gaussian noisy version of (A). (E, F) Example of patches resized to have smaller (respectively larger) cells than those used for training.

To evaluate the performance of the original U-Net on distorted data, we focused on three common types of artefacts: non-uniform illumination, gaussian noise and cells of different sizes. We took images that our models were not trained with and artificially applied distortions of varying intensity to simulate these artefacts. To simulate a non-uniform illumination, we added to the images randomly-centered 2D gaussians, of standard deviation proportional to the image pixel size and with tunable amplitude. To simulate a noisy image, gaussian noise with zero-mean and tunable standard deviation was randomly generated and added to the original image. In both cases, the addition was performed using numbers with a larger bit size than the original image, as to avoid saturation of the image, and the intermediate result was then normalised to have values in the 8-bits range: $[0, 255]$. In order to simulate cells having a larger or smaller size than in the original image, a section of the image of size chosen accordingly to the desired zooming factor was cropped and then reshaped using cubic interpolation into a 256×256 image. Zooming factors as low as 0.5 could be achieved, thanks to the fact the original images were considerably larger than the input size of the U-Net. To reduce the risk of having a large amount of patches with no cells in them when at large zoom factors, we changed the number of patches we take from the image proportionately to the ratio of surface of the extracted patches and original image, clipping when reaching a too large number of patches per image due to computational limitations. Examples of such artificial generated artefacts can be seen in Figure 2.

To evaluate the models, we applied an increasing amount of distortion to all the test patches and predicted the segmentation masks with each model. The segmentation masks of all models, which had values in range $[0, 1]$, were binarized using a threshold at 0.5. From those outputs we computed the binary pixel-wise accuracy, the Jaccard score, the number of different cells predicted (this metric was calculated using 8-way connected components analysis), the precision (measuring the strength of the true positive prediction) and the recall (often also called sensitivity, or true positive rate) for all of the distortion's intensities. All of these metrics were computed over a large number of patches and combined to obtain robust measures (averaged for accuracy, Jaccard, precision and recall or summed for the number of cells measure). In the results section, the Jaccard score was omitted, since it did not provide any additional information over the accuracy measure.

When evaluating the models for the non-uniform illumination and gaussian noise artefacts, the deep learning models were also compared to classical image-processing segmentation techniques, consisting in applying a filter to remove the distortion (in the case of non-uniform illumination, a Difference of Gaussians (DoG) was used and calibrated by empirically measuring the cell sizes in pixels from the images, while for the gaussian noise, a state-of-the-art Non-local Means Denoising filter was used) followed by Otsu's thresholding.

III. Results

As we can see from Figures 3, 4 and 5, when no distortions are applied, the number of cells segmented by all models is larger than the number of cells actually present in the ground-truth segmentation masks. This is mainly due to the presence of regions of the images, such as the one visible in Figure 6 (A), that cause the models to incorrectly segment a cell at that spot (false positives). This is confirmed by the fact that the precision (measuring the strength of the true positive prediction) of none of the models is exactly at 1, even in absence of deformation.

As we can see from Figure 3, a large amount of non-uniform illumination causes the U-Net models to tend towards predicting only background. Indeed, as accuracies converge to that of only-background prediction, the number of cells tends to a value close to zero and recall (sensitivity) experiences a large drop for increasing amounts of distortions. The observed effect, which can be seen in Figure 6 (E), is that the increasing amounts of non-uniform illumination cause the cells' predictions to shrink more and more, until only background is predicted. As for the image processing method using a DoG followed by Otsu's thresholding, we can observe that it provides stable predictions up to a certain level of distortion, after which its accuracy and precision greatly drop because, as can be seen in Figure 6 (E), it starts to detect the center of the gaussian as a cell. As can also be seen in Figure 6 (E), the cells sufficiently far from the center of the gaussian are still segmented relatively well, causing the relatively small drop in recall and number of predicted cells of this model over the tested range of distortion. The maximum accuracy achieved by the image processing method is smaller to that of the deep learning models because, as can be seen from Figure 6 (D), its predictions for the cells are rounded up, thus not corresponding perfectly with the true segmentation mask. However, for a very tiny window of distortion amplitudes, the image processing method manages to perform better than the U-Net models not trained with images containing this type of distortion. Nevertheless, it must be noted that the U-Net model trained with images containing this type of distortion performs considerably better than all the other models, over the whole range of distortions tested.

For the gaussian noise artefact, the results in Figure 4 show that the U-Nets not trained with distorted data and the image processing method start failing relatively quickly. Examples of such can be observed in Figures 6 (F) and (G) for the U-Net and image processing method respectively. Indeed, as shown by the graph in Figure 4, they experience a quick drop in accuracy and precision, falling even below the only-background prediction, thus meaning that they start segmenting regions of the image where there are actually no cells. This effect gives rise to a sharp increase in the number of segmented cells, as can be seen in Figure 4 (B). The U-Net trained with distorted images, however, converges again towards predicting all background, with a relatively low decrease in precision, but a large fall in recall, as it starts to miss the cells

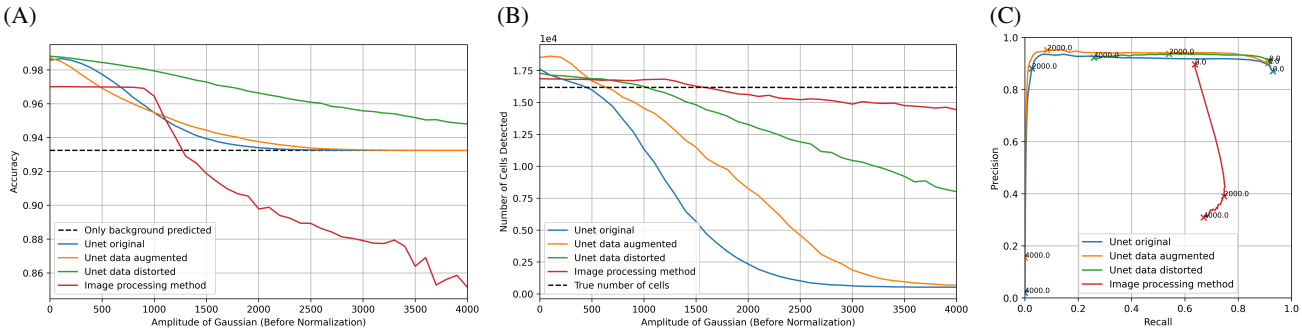

Figure 3. Evaluation of the robustness against an added gaussian on the input test images, which simulates a non-uniform illumination, of the original U-Net, the U-Net retrained with simple data augmentation (random rotation by 90° and random flips of training images), the U-Net retrained with the same simple data augmentation and also with random added distortions (addition of a gaussian to simulate non-uniform illumination and addition of a gaussian noise) and an image processing method to segment cells that uses a Difference of Gaussians filter followed by Otsu's thresholding. The different levels of degradation were simulated by the modulation of the amplitude of the added gaussian. (A) Accuracy of the different U-Nets and the image processing method in function of the amplitude of the added gaussian. The accuracy of only background being predicted is shown by the dotted line. (B) Total number of cells in the images predicted by the different U-Nets and the image processing method in function of the amplitude of the added gaussian. The real total number of cells to detect is shown by the dotted line. (C) Precision-Recall curve of the predictions made by the different U-Nets and the image processing method when different levels of distortion are applied.

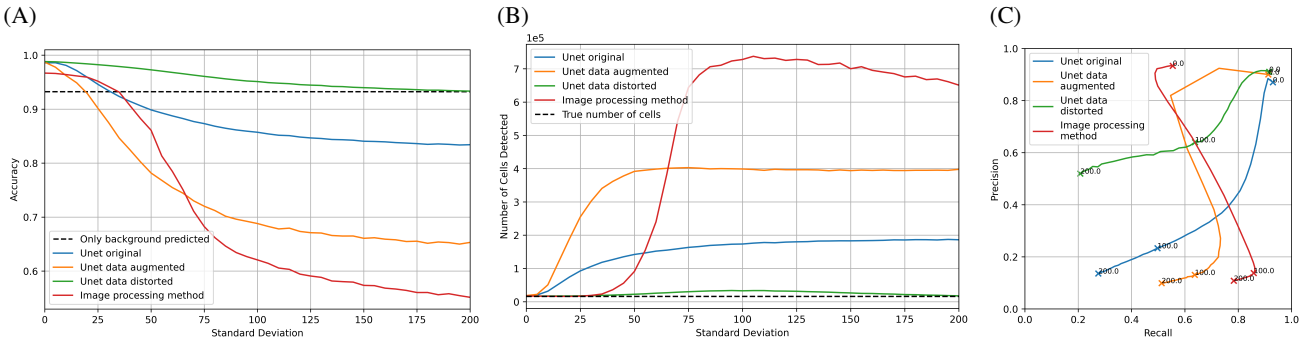

Figure 4. Evaluation of the robustness against an added gaussian noise on the input test images of the original U-Net, the U-Net retrained with simple data augmentation (random rotation by 90° and random flips of training images), the U-Net retrained with the same simple data augmentation and also with random added distortions (addition of a gaussian to simulate non-uniform illumination and addition of a gaussian noise) and an image processing method to segment cells that uses a Non-local Means Denoising filter followed by Otsu's thresholding. The different levels of degradation were simulated by the modulation of the standard deviation of the added gaussian noise. (A) Accuracy of the different U-Nets and the image processing method in function of the standard deviation of the added gaussian noise. The accuracy when only background is predicted is shown by the dotted line. (B) Total number of cells in the images predicted by the different U-Nets and the image processing method in function of the standard deviation of the added gaussian noise. The real total number of cells to detect is shown by the dotted line. (C) Precision-Recall curve of the predictions made by the different U-Nets and the image processing method when different levels of distortion are applied.

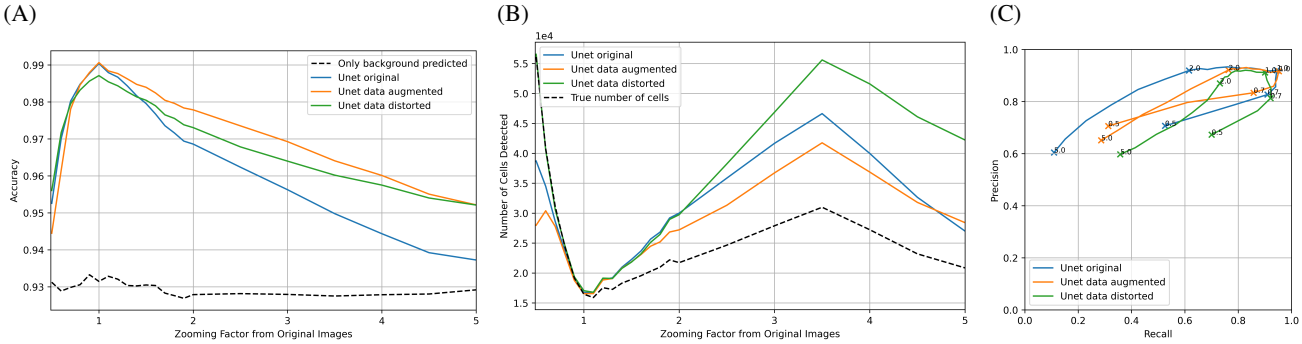

Figure 5. Evaluation of the robustness against a change of the cell size of the original U-Net, the U-Net retrained with simple data augmentation (random rotation by 90° and random flips of training images) and the U-Net retrained with the same simple data augmentation and also with random added distortions (addition of a gaussian to simulate non-uniform illumination and addition of a gaussian noise). The different levels of degradation were simulated by taking patches of different sizes (smaller patches correspond to larger zooming factors) and reshaping them via cubic interpolation to meet the constant required size of input images of the U-Net: 256×256. (A) Accuracy of the different U-Nets in function of the zooming factor. The accuracy when only background is predicted is shown by the dotted line. (B) Total number of cells in the images predicted by the different U-Nets in function of the zooming factor. The real total number of cells to detect is shown by the dotted line. (C) Precision-Recall curve of the predictions made by the different U-Nets when different zooming factor are applied.

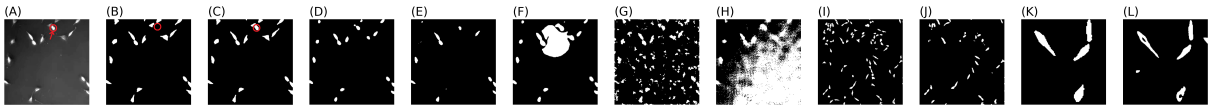

Figure 6. Example of predictions from different models under different conditions. (A) Original image used for segmentation in (C-H). Circled in red is a region of the image causing false positives in the model predictions. (B) Original segmentation mask of (A). (C) Example of segmentation mask generated by U-Net models from (A) without any distortions applied. (D) Example of segmentation mask from the image processing method using a Difference of Gaussians filter and Otsu's thresholding from (A) without any distortions applied. (E) Example of segmentation mask generated by U-Net models when a large non-uniform illumination is applied to (A). (F) Example of segmentation failure of same model as in (D) when a large non-uniform illumination is applied to (A). (G) Example of segmentation failure of the original U-Net and U-Net retrained with simple data augmentation (random rotation by 90° and random flips of training images) when gaussian noise is applied to (A). (H) Example of segmentation failure of the image processing method using a Non-local Means Denoising filter and Otsu's thresholding when gaussian noise is applied to (A). (I) Original segmentation mask of image used for segmentation (J), characterized by cells smaller than those the U-Nets were trained with. (J) Example of segmentation mask generated by U-Net models from image corresponding to mask (I), in which some small cells are not segmented. (K) Original segmentation mask of image used for segmentation (L), characterized by cells larger than those the U-Nets were trained with. (L) Example of segmentation mask generated by U-Net models from image corresponding to mask (K), in which some parts of large cells are not correctly segmented.

behind the noise. It must be noted that the U-Net model trained with images containing this type of distortion performs considerably better than all the other models over the whole range of distortion tested.

Finally, the last artefact consisted in using cells considerably smaller and larger than those the U-Nets were trained with. In Figure 5, we can observe that the accuracy and recall of all the U-Nets drops quickly when cells are much smaller than those they were trained with (i.e for zooming factor smaller than 1) because, as seen in Figures 5 (B) and 6 (J), the models do not segment certain cells at all, or even segment them incorrectly (for example, in the case of the U-Net trained with distorted data, breaking them into pieces). In Figure 5, we can observe that the accuracy and recall of all the U-Nets also drops when cells are much larger than those they were trained with (i.e for zooming factor larger than 1). Additionally, Figure 5 (B) shows an increase in the number of predicted cells compared to the true number of cells in the masks. This is due to, as can be seen in Figure 6 (L), the U-Net models not segmenting correctly the interior of large cells, which sometimes leads to them breaking into separated segmentations. This effect of splitting large cells into several small ones is particularly pronounced in the U-Net trained with distorted data.

IV. Discussion

In the results section, we showed that some images contained artefacts similar in shape and color to cells (possibly due to contamination in the sample, bubbles in the substrate, ...), which the different models often segmented as cells, therefore leading to false positives. These artefacts are very hard to distinguish from the actual cells once the photo is taken, but their presence could be attenuated by some laboratory practical techniques.

We also showed that the DoG, which is a band-pass filter, produces segmentation masks in which the cells are rounded up. This is due to different cells not always having a shape and size that perfectly fits in the band-pass of the filter, filtering out some parts of the borders. Additionally, as it is not an ideal filter, when applying large amounts of non-uniform illumination, it can't completely filter out the distortion, thus leading to the thresholding detecting the center of the gaussian as a point of interest. Similarly, the Non-Local Means Denoiser is not perfect and also eventually fails, causing filtering artefacts that get segmented as cells.

Despite the exact internal functioning of the U-Nets not being completely well-defined, we can make some hypothesis on how they react to different spatial frequencies in the image.

When confronted with non-uniform illumination (which is a low-frequency perturbation), they all exhibit a similar behaviour (even though the U-Net trained using data containing this deformation is much more resistant): the size of the predicted cells gradually decreases for increasing amounts of non-uniform illumination. This effect is justified by the fact that all the U-Nets likely act as a high-pass filter in order to detect the borders of the cells and filter the low-frequency, non-uniform illumination that is already present in the training set. Therefore, for large non-uniform illuminations, the contrast between nearby parts of the image decreases, as can be seen in Figure 2 (C), becoming increasingly difficult to precisely distinguish the cells from the background.

Both U-Net models that were not trained with distorted data easily segment non-existing cells when faced with gaussian noise. This behaviour is coherent with the intuition that they developed some kind of a high-pass filter, as previously discussed. On the other hand, the U-Net model trained with data containing this perturbation filters the high-frequency noise very effectively, suggesting that it also developed some kind of a low-pass filter (therefore making an overall band-pass filter, capable of filtering both very low and very high frequency perturbations).

When faced with cells of different sizes compared to those they were trained with, all of them show similar behaviours: they do not detect relatively small cells and often break up large ones into multiple cells. This first effect is justified by the U-Net filters expecting cells of a particular size, and therefore not outputting a sufficiently large signal when smaller ones are present. The second effect can be linked to the high-pass behaviour discussed previously: the center of large cells is locally at a low frequency and, therefore, gets filtered by the U-Net models. This effect is much more pronounced in the U-Net trained with data containing non-uniform illumination.

The greatly improved resistance of the U-Net model trained with non-uniform illumination and gaussian noise perturbations proves that the training of this more robust model was successful.

V. Summary

To conclude, we successfully analyzed the robustness of the U-Net included in the DeepImageJ package, used for pancreatic cell segmentation, to some common image artefacts found in optical microscopy. We also compared it to other image processing methods used for segmentation. And finally, we achieve retraining the U-Net with distorted data and showed that this led to a considerable increase in its robustness against these common artefacts.

REFERENCES

[1] S. Dimopoulos, C. E. Mayer, F. Rudolf, and J. Stelling, "Accurate cell segmentation in microscopy images using membrane patterns," *Bioinformatics*, vol. 30, no. 18, pp. 2644–2651, Sep. 2014. [Online]. Available: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btu302

[2] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *arXiv:1505.04597 [cs]*, May 2015, arXiv: 1505.04597. [Online]. Available: http://arxiv.org/abs/1505.04597

[3] H. Noh, T. You, J. Mun, and B. Han, "Regularizing Deep Neural Networks by Noise: Its Interpretation and Optimization," *arXiv:1710.05179 [cs]*, Nov. 2017, arXiv: 1710.05179. [Online]. Available: http://arxiv.org/abs/1710.05179

[4] I. Arganda-Carreras and E. Gómez-de Mariscal, "Deep Learning example: U-Net for binary segmentation." [Online]. Available: https://colab.research.google.com/github/deepimagej/models/blob/master/u-net_pancreatic_segmentation/U_Net_PhC_C2DL_PSC_segmentation.ipynb

[5] D. Pedamonti, "Comparison of non-linear activation functions for deep neural networks on MNIST classification task," *arXiv:1804.02763 [cs, stat]*, Apr. 2018, arXiv: 1804.02763. [Online]. Available: http://arxiv.org/abs/1804.02763

[6] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)," *arXiv:1511.07289 [cs]*, Feb. 2016, arXiv: 1511.07289. [Online]. Available: http://arxiv.org/abs/1511.07289

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," in *2015 IEEE International Conference on Computer Vision (ICCV)*. Santiago, Chile: IEEE, Dec. 2015, pp. 1026–1034. [Online]. Available: http://ieeexplore.ieee.org/document/7410480/

[8] D. Godoy, "Hyper-parameters in Action! Part II - Weight Initializers," Dec. 2018. [Online]. Available: https://towardsdatascience.com/hyper-parameters-in-action-part-ii-weight-initializers-35aee1a28404

[9] J. Brownlee, "A Gentle Introduction to Dropout for Regularizing Deep Neural Networks," Dec. 2018. [Online]. Available: https://machinelearningmastery.com/dropout-for-regularizing-deep-neural-networks/

[10] D. Masters and C. Luschi, "Revisiting Small Batch Training for Deep Neural Networks," *arXiv:1804.07612 [cs, stat]*, Apr. 2018, arXiv: 1804.07612. [Online]. Available: http://arxiv.org/abs/1804.07612

[11] "Avoid wasting resources with EarlyStopping and ModelCheckpoint in Keras," May 2019. [Online]. Available: https://www.machinecurve.com/index.php/2019/05/30/avoid-wasting-resources-with-earlystopping-and-modelcheckpoint-in-keras/

[12] "2D+Time Datasets – Cell Tracking Challenge." [Online]. Available: http://celltrackingchallenge.net/2d-datasets/