

Fast Fourier Transformation for Optimizing Convolutional Neural Networks in Object Recognition

Varsha Nair¹, Moitrayee Chatterjee¹, Neda Tavakoli², Akbar Siami Namin¹, and Craig Snoeyink³

^{1,2}Department of Computer Science, ³Department of Mechanical and Aerospace Engineering

¹Texas Tech University, ²Georgia Institute of Technology, ³State University of New York at Buffalo
 {varsha.nair, moitrayee.chatterjee, akbar.namin}@ttu.edu; neda.tavakoli@gatech.edu; craigsno@buffalo.edu

Abstract

This paper proposes to use Fast Fourier Transformation-based U-Net (a refined “fully convolutional networks”) and perform image convolution in neural networks. Leveraging the Fast Fourier Transformation, it reduces the image convolution costs involved in the Convolutional Neural Networks (CNNs) and thus reduces the overall computational costs. The proposed model identifies the object information from the images. We apply the Fast Fourier transform algorithm on an image data set to obtain more accessible information about the image data, before segmenting them through the U-Net architecture. More specifically, we implement the FFT-based convolutional neural network to improve the training time of the network. The proposed approach was applied to publicly available Broad Bioimage Benchmark Collection (BBBC) dataset. Our model demonstrated improvement in training time during convolution from 600 – 700 ms/step to 400 – 500 ms/step. We evaluated the accuracy of our model using Intersection over Union (IoU) metric showing significant improvements¹.

Index Terms

Object Recognition, CNN, Fast Fourier Transformation, Image Processing.

I. INTRODUCTION

Computer vision follows a sequence of operations to identify the characteristics of an image to perceive and analyze. Computer vision can be extensively used to solve a wide variety of problems in the field of remote sensing, spectroscopy, medical imaging, image processing, meteorology, astronomy and microscopy. Image recognition is an exciting research fields in computer vision, for which researchers predict that the global market will reach \$38.92 billion by 2021 [1].

The application of deep learning approaches for image enhancement, restoration and morphing [11], [12]. has significantly improved the data (or features) extraction from images. Moreover, Convolutional Neural Networks (CNNs) have been extremely successful in image processing. However, CNNs are computationally expensive due to computation costs of convolutions. By transforming images and CNN kernels into frequency space using Fast Fourier Transformation (FFT), a convolution operator becomes a single element-wise multiplication. Hence, FFT is used tremendously for image convolution. But, there is an overhead cost involved in performing FFT. This type of transformation is efficient in reducing the convolution cost only when the convolution is large. Recently, convolutions of CNNs have been optimized using FFT [25] and being adapted in mobile and embedding systems. Here, we apply FFT to the input data only and demonstrate the performance improvement on detecting objects.

Convolutional Networks [13] have existed for some times but the required size of training sets and the network size restricted their wide adoptions. On the other hand, their deep variations have shown outstanding performance in visual object recognition tasks [12], through supervised training.

The object identification algorithms typically use extracted features and learning algorithms to apprehend instances of an object category. Currently, most of the object detection models specify where object is located in image with a bounding box [10], [17].

This study attempts to identify object position using a FFT-based convolutional neural network to detect the position by image segmentation rather than a bounding box. One of the major issues with using CNN for image dataset is the training time. This study focuses on lesion detection in image data set using the U-Net model and thus aims at improving the training time by implementing Fast Fourier Transformation layer to the U-Net model. This paper proposes to accomplish this by using a Fast Fourier convolution with fully CNN. The proposed model detects objects from an image, utilizing the scaling properties of FFT that improves the convolution time. The key contributions of this paper are as follows:

¹This paper is the pre-print of a paper to appear in the proceedings of the IEEE International Conference on Machine Learning Applications (ICMLA'20) entitled: “Optimizing CNN using Fast Fourier Transformation for Object Recognition.”

- Introducing Fast Fourier Transformation-based deep learning algorithm (i.e., FFT-based CNN) in the context of object recognition for optimization purposes,
- Implementing FFT on an U-Net architecture and capturing key features, and
- Evaluating the proposed Fast Fourier Transformation-based CNN through a set of experiments using bio-image dataset in the domain of health analytic.

The rest of the paper is organized as follows: we review the existing work in object identification in Section II and then provide explanation of the ideas and theories used in our implementation in Section III. Section IV presents the methodology and generic algorithm for proposed approach. Section V gives an insight to the experiments and the results. We concluded the paper in Section VI.

II. RELATED WORK

To incorporate the contextual information for lesion recognition Gao et al. [8] proposed a fusion of CNN (convolutional Neural Networks) and RNN (Recurrent Neural Networks) based approach. Shen et al. [21] implemented three CNNs for addressing the problem of object recognition. While these approaches are targeted for object identification for 2D images, Setio et al. [20] proposed a multi-stream CNN for object classification in 3D images.

For quantitative analysis of clinical characteristics, the bio-image needs segmentation to identify the pixels that represent the object of interest. Ronneberger et. al. [18] proposed U-Net for bio medical image segmentation. It is a refined architecture developed from fully-convolutional neural network which is used for fast and precise segmentation of images. It was named on the fact that its architecture is in the shape of the letter “U”. The U-Net has been further extended [5] for 3D image segmentation. V-Net [15] is another variant of U-Net for 3D images. RNNs are also used for image segmentations [27]. Stollenga et al. [22] implemented a 3D LSTM-RNN with convolutional layer. Andermatt et al. [2] proposed gated RNN for segmenting 3D bio-images. Chen et al. [3] combined RNN with U-Net like architecture for segmentation. Convolutional Neural Networks also have been used in many other application domains such as mapping genome data into images [23].

III. THEORETICAL BACKGROUND

This section discusses the two major techniques that we used for the convolution and image processing: (1) Fast Fourier Transformation and (2) U-Net.

A. Fast Fourier Transformation

In image processing, FFT is applied on an image to decompose it into real and imaginary components, representing the image in a frequency domain [9]. The FFT of a 2D image [16] can be calculated using Equations (1) and (2):

$$F(x, y) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n) e^{-i \times 2 \times \pi \left(x \frac{m}{M} + y \frac{n}{N} \right)} \quad (1)$$

$$f(x, y) = \frac{1}{M \cdot N} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} F(m, n) e^{i \times 2 \times \pi \left(x \frac{m}{M} + y \frac{n}{N} \right)} \quad (2)$$

where $f(m, n)$ represents the pixel at position (m, n) , whereas $F(x, y)$ is the function to represent the image in the frequency domain pertaining to the position x and y , $M \times N$ represents dimension of the image, and i is $\sqrt{-1}$ [9].

In our work, we used a tensorflow-based implementation of Fast Fourier Transform in order to produce a transformed feature map in Fourier domain, which is then provided to U-Net for object segmentation.

B. U-Net Architecture

Ronneberger et al. proposed the U-Net architecture [18] consists of a contracting path to apprehend the context and a symmetric expanding path that enables precise localization. This network is trained end-to-end using very few images and outperformed sliding-window convolutional network like OverFeat [19].

In Figure 1, the architecture of U-net is presented where each blue box is a multi-channel feature map. The number of channels is indicated on the top of each box. White boxes are mapping of copied features. The arrows denote the various operations within the network.

IV. METHODOLOGY

This section presents the architecture of the proposed model along with the developed algorithm. This work implements the U-Net [18] with Fast Fourier Transform-based NN. The neural network implements the Fast Fourier Transform for the convolution between image and the kernel (i.e., mask). We trained the NN with labeled dataset [14] which consists of synthetic cell images and masks. Each pixel on the image is classified as either being part of a cell or not. The image is a matrix that has values ranging between 0 and 255 (RGB). The mask (i.e., kernel) is a 2-D matrix with white and black values. Each pixel in the predicted mask represents the probability that the pixel is part of a cell. The output of the prediction model is clustered to identify the object location in a 2-D environment.

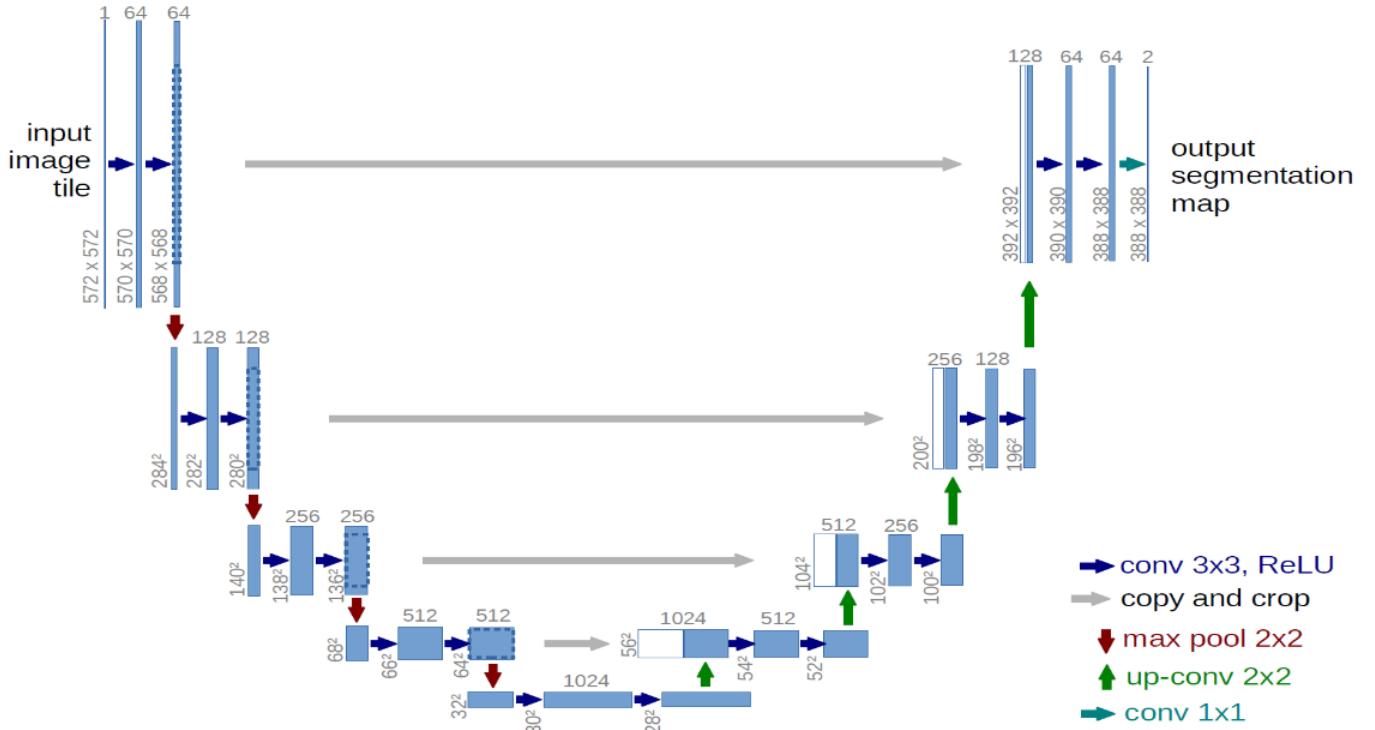


Fig. 1: Architecture of U-Net [18] on lowest resolution of 32×32 pixels.

A. Model Overview

The proposed architecture is built upon the U-Net, which is a fully Convolutional Network. The U-Net is modified to yield better segmentation in medical imaging. The proposed model consists of the contracting path and an expansion path. These two together form a U-shaped architecture. The contracting path comprises of a fully connected convolutional network and includes repeated application of convolutions, each of which is followed by an exponential Linear Unit (eLU) and max-pooling operation. The contraction path decreases the spatial information, while increasing the feature information. The expansive path combines the feature and spatial information through a sequence of up-convolutions and concatenations with high-resolution features from the contracting path. The up-convolution in this experiment is de-convolution operation, which is used for up-sampling. In this model, we incorporate dropout in every block to avoid overfitting. Figure 2 displays the architecture of the proposed model.

It consists of:

- 1) **Input Layer.** This layer takes the input image and performs Fast Fourier convolution by applying the Keras-based FFT function [4]. The input layer is composed of:
 - a) A lambda layer with Fast Fourier Transform
 - b) A 3×3 Convolution layer and activation function, and
 - c) A lambda layer with Inverse Fast Fourier Transform.
- 2) **Contracting Path.** The Contracting layer consists of four blocks where each block is a group of operations bound together logically. In this part of the architecture, each block entails:
 - a) Two 3×3 Convolution layers with activation function, and
 - b) A 2×2 Max pooling layer.

This path applies 3×3 convolution repeatedly. Each convolution is followed by exponential Linear Unit (eLU) and a 2×2 max pooling operation with stride 2 for down-sampling. Here, at every max pooling step the feature channels are doubled.
- 3) **Bottleneck.** The bottleneck is the connection between contracting and expanding paths. It consists of two convolutional layers with dropout:
 - a) Two 3×3 Convolution layers with activation functions, and
 - b) A 2×2 Convolution layer.
- 4) **Expanding.** Every step in the expansive path consists of an up-sampling of the feature map, followed by a 2×2 convolution to reduce the number of features. A concatenation is applied with the corresponding feature map from the

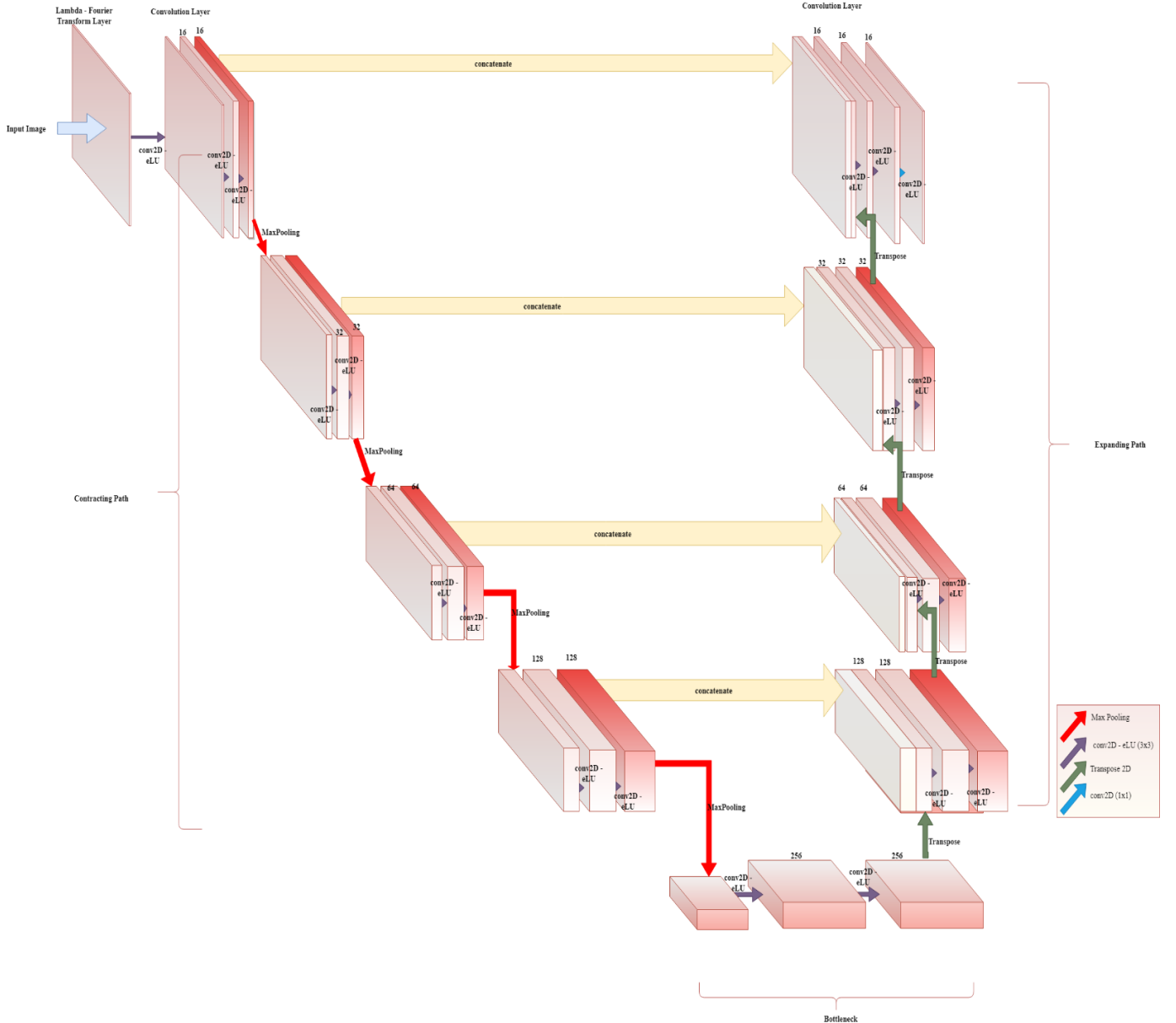


Fig. 2: Architecture of the proposed model.

contracting path. Furthermore, two 3×3 convolutions with eLU are applied. It consists of 4 blocks where each comprises of:

- A de-convolution layer with stride 2,
- A concatenation with the corresponding cropped feature map from the contracting path,
- Two 3×3 convolution layers with activation functions,
- An output layer with 1×1 convolutional layer, and
- At the final layer, a 1×1 convolution layer is used to map each feature vector to the input class.

B. Process Flow

The Figure 3 provides a process flow of the proposed approach. The proposed architecture is built upon the U-Net, which is a fully Convolutional Network. The U-Net is modified to yield better segmentation in medical imaging. After importing the training dataset with image and mask for training the neural network, we resized the dimensions of the input image to 256×256 limit. Then a tensorflow spectral function is applied to the images to compute 2-D Fast Fourier Transform and Inverse Fast Fourier Transform. This tensorflow-based FFT layer is added to the neural network. The convolution of input image and mask is then implemented using a 2-D convolutional layer, which performs element-wise multiplications between

the Fourier transform of the input and mask. Subsequently, an Inverse Fourier Transform layer is applied to retain the original dimensions.

The purpose of pooling in the contraction path is to reduce the size of the feature channels so that we have fewer parameters in the network. Therefore, the NN retains the most relevant features (max valued pixels) from each region and discards the rest of pixels, and it is achieved through pooling in the contraction path. The output of this layer is provided to a flattening layer, which flattens all feature structure to create a single long feature vector to be used by the output layer for object identification.

In this model, we incorporate dropout in every block to avoid overfitting. Figure 2 displays the architecture of the proposed model.

C. Drop Out

Dropout is the technique of ignoring randomly selected nodes during training to prevent overfitting in neural networks. They are “dropped-out” randomly. For this experiment dropout is set on each block in the architecture. In Figure 3, the dropout rate is set as 0.1 for the blocks c1, c2, c8 and c9 and 0.2 for the blocks c3, c4, c5, c6 and c7.

D. Training, Validation, and Testing

The model is trained on the labeled data set where each image has a mask. The mask helps to identify the position of the object of interest in the main image. Additionally, validation and model fitting set is used for unbiased evaluation of the model on the training dataset. The model is tested against validation test data so as to ensure to calculate the model metrics. It is used to avoid overfitting. In this experiment, the validation test data consists of image and mask where the mask is compared with prediction to generate intersection over union scale.

After training the model, a scaled test set is used for an unbiased performance evaluation. The proposed model predicts the test mask from the test image dataset, which consists of the position of the objects in the image. The test mask is resized and local maxima are calculated on the test mask to capture the high intensity points. Further, DBScan clustering algorithm is applied to aggregate the local maxima in order to capture the object location and the number of objects in the image [7].

V. CASE STUDY

This section presents the results of a case study and implementation of the proposed FFT-based neural network algorithm. In this study, the accuracy of the result from the CNN based model is compared to the result of a non-CNN based system.

A. Experimental Procedure and Data Collection

Out of the total 19,200 simulated cell microscopy images available in the BBBC005v1 [14] dataset, the training set comprises of 3,600 images and 1,200 masks while test set comprises of 85 images. The validation set comprises of 47 images and 47 masks from the training set to ensure the accuracy of the model by fitting better models. The model uses these processed dataset and predicts the mask, which is then further processed by resizing and calculating the local maxima. The processed image is clustered in order to identify the object position and number in the image.

B. Results

The experiments are performed in 1, 10, 50 and 100 epochs on FFT-based and non-FFT based algorithms. For the sake of space limit, we discuss the results for 100 epochs only. The output obtained from the training is processed and clustered to identify the objects in the test image. For all the discussions on results, Figure 4 is used as one of the test images.

The images shown in Figures 5a and 5b are the output results when Figure 4 is used as the test image. The mask prediction obtained using FFT-based algorithm is presented in Figure 5a and Figure 5b, where the Figure 5a is the predicted mask and Figure 6a is the unsampled prediction. The prediction obtained from model without FFT is shown in Figure 5b and unsampled prediction is Figure 6b.

Table I reports the performance metrics obtained for model in 100 epochs with and without FFT, respectively. The evaluation metrics used in this experiment is Intersection Over Union (IoU) or the Jaccard index, a metric for measuring the accuracy of detecting objects. The average IoU obtained when FFT-based Neural Network is used is 0.83; whereas, when FFT is not employed, the average IoU obtained is 0.52. As the results suggest, the accuracy of object detection is much higher when FFT is used.

TABLE I: Metric and cost function value for the experiment.

FFT	Avg Valid. IoU	DICE Value	Mean IoU	CV. Val Loss	CV. Val mean IoU
With	0.83	-0.8392	0.8747	-0.8753	0.8748
W/o	0.52	-0.5631	0.6815	-0.6025	0.6815

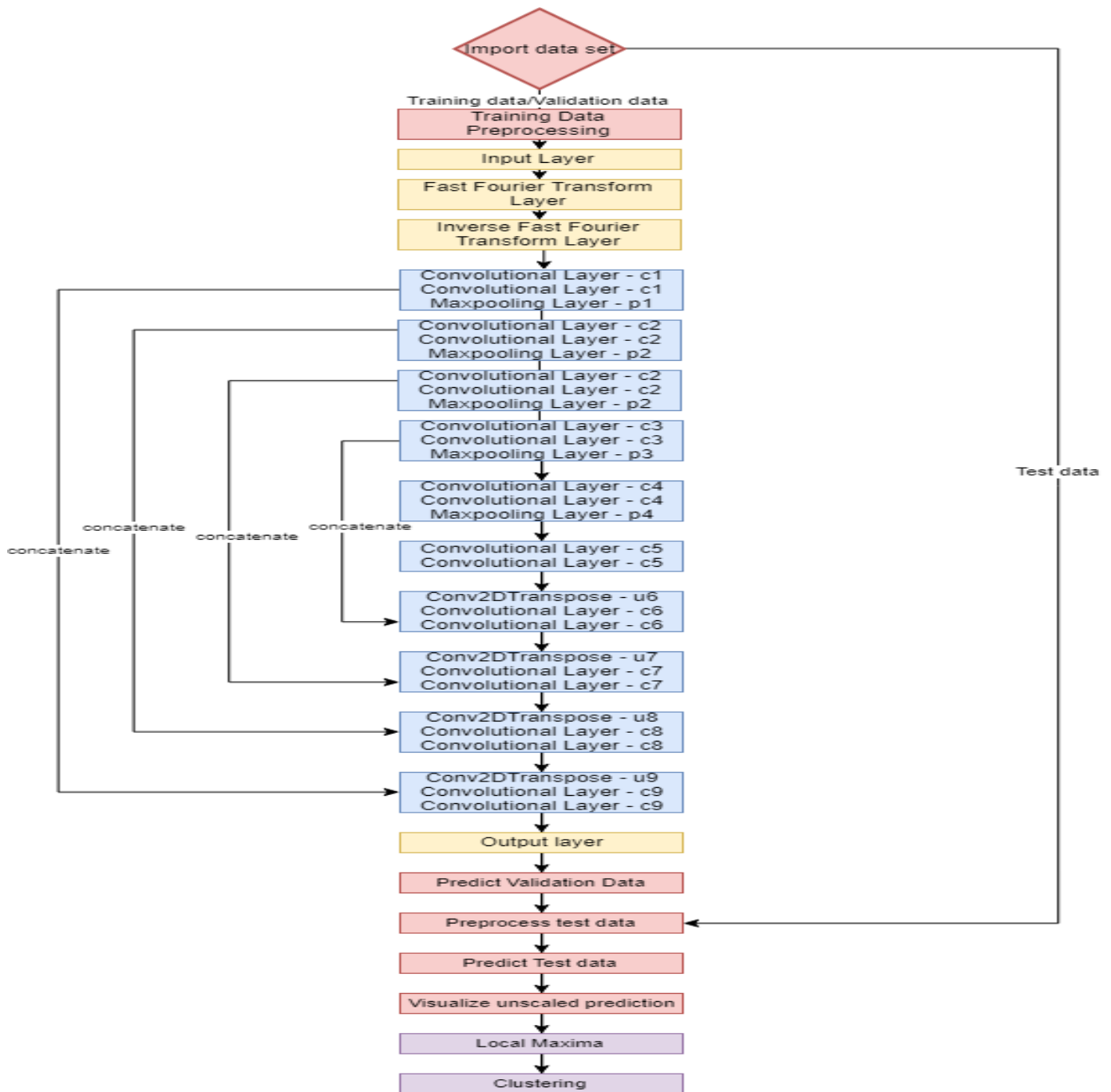


Fig. 3: Process flow of the proposed architecture.

The loss function used in this experiment is DICE value [6] to assess the segmentation. The DICE value obtained with FFT based Algorithm is -0.8392 and without FFT based algorithm is -0.5631 , showing a better similarity achieved by FFT. The Mean IoU for the final iteration obtained is 0.8747 for FFT based algorithm and 0.6815 for non-FFT based algorithm.

For algorithm with FFT, cross-validation loss obtained is -0.8753 and for non-FFT based algorithm it is -0.6025 . The cross-validation mean IoU is 0.8748 and 0.6815 for FFT based algorithm and non-FFT based algorithm, respectively.

Figures 7a and 7b illustrate the plots for loss and mean IoU obtained on the training and validation phases with FFT and without FFT, respectively. As log-loss plots illustrate the log-loss values are approaching to -0.9 and -0.6 , respectively for with and without FFT, indicating the reduction in loss values. Moreover, it is also observable from the plots that the log-loss values are very stable and consistent when FFT-based approach is implemented; whereas, the log-loss values when FFT is not applied fluctuate between different epochs.

In the mean IoU metric plot, the X-axis consists the number of epochs and Y-axis consists of the mean IoU value. The graph consists of plot for mean IoU value obtained for each epoch. Both FFT-based and non-FFT-based approaches show similar

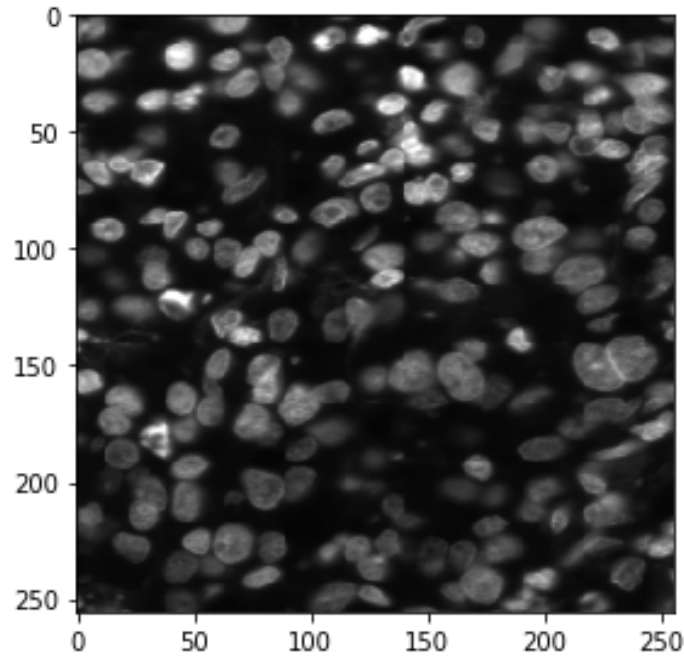


Fig. 4: Original image.

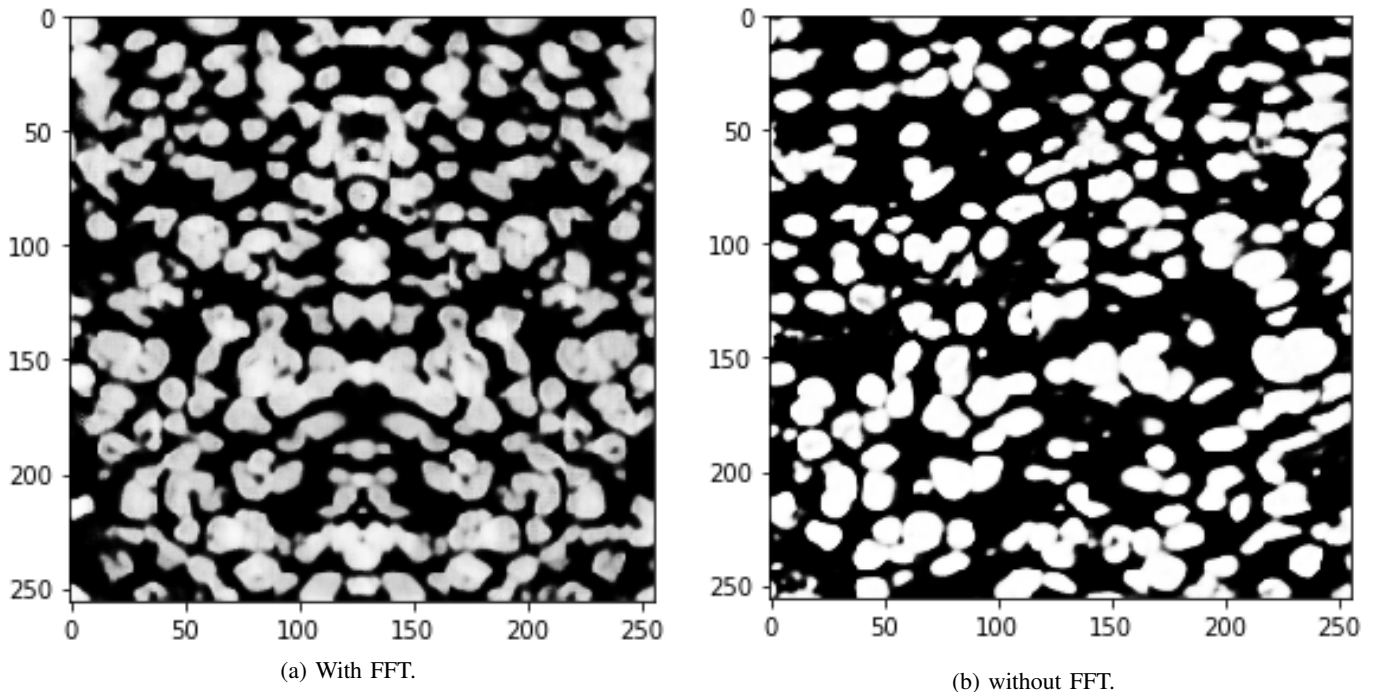


Fig. 5: Mask predicted.

trends. However, as the scale of y-index indicates, the mean of IoU when FFT is utilized is much higher than compared to cases when FFT is not applied.

C. Generate Local Maxima and Clustering Using DBScan

The input image for this experiment is obtained after the prediction of the test image from the model (Figure 5a). The resultant image, Figure 5a, is resized and used as the input. The local maxima are calculated based on the brightness of the

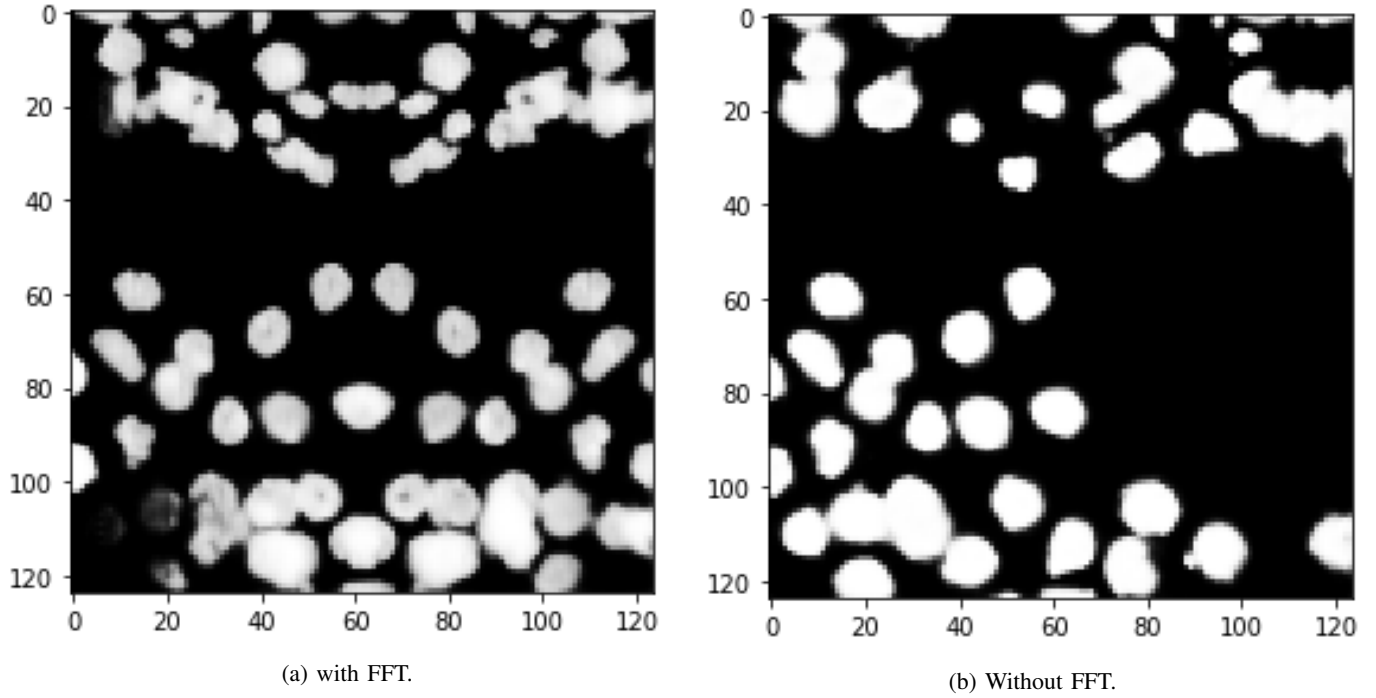


Fig. 6: Unsampled prediction.

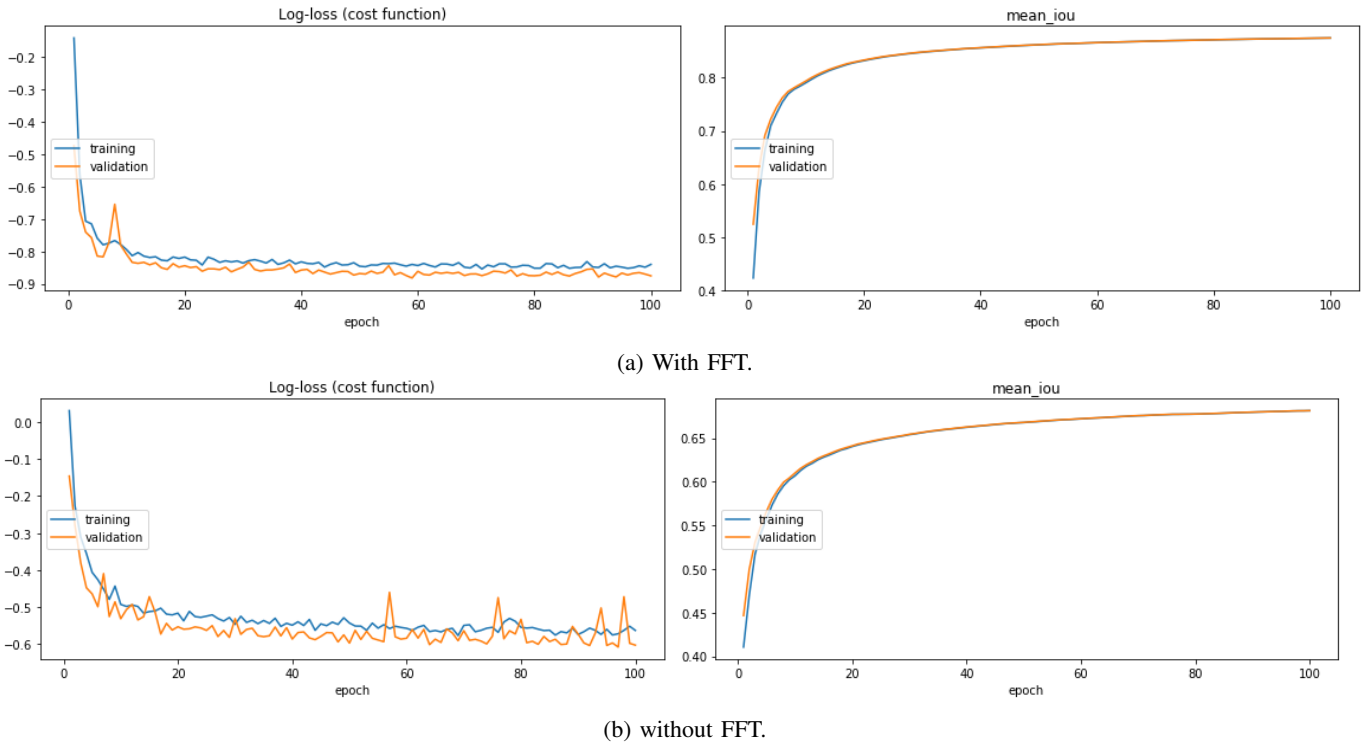


Fig. 7: Log-Loss and mean IoU metric plots for 100 epochs.

pixel in the resized image, Figure 8, which will help in identifying peaks using the peak local maxima from scikit-image [24].

In this experiment, Figure 8 (left) is used as the input image to detect the local maxima by applying maximum filter to the original image where Figure 8 (middle) is the image after maximum filter is applied. This helps in accentuating the high intensity region of the image. The local maxima function is applied to Figure 8 (middle). The resultant image is Figure 8 (right) where the high intensity pixels are identified using local maxima.

The output image obtained with local maxima identified, 8 (right), are then clustered using DBScan to identify the number

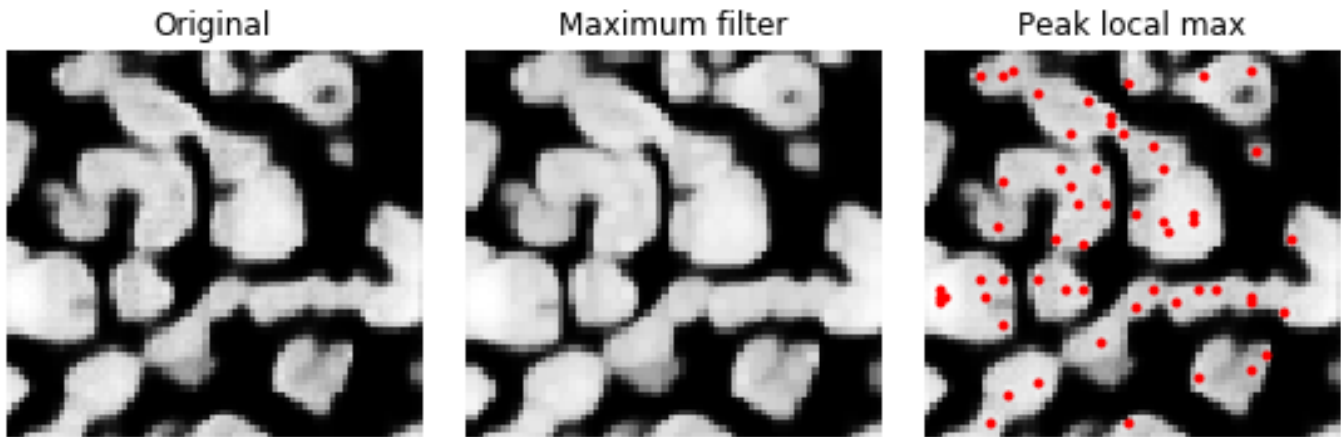


Fig. 8: Original image, Original image with maximum filter, Image with local peak maxima.

of objects by the number of clusters. Figure 9 shows the result of applying clustering using DBScan on Figure 8. The clustering algorithm has identified 20 clusters in the image.

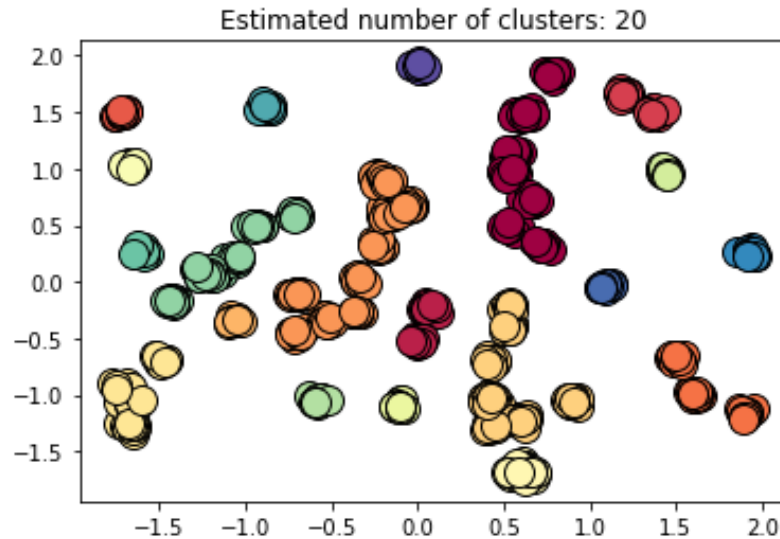


Fig. 9: DBScan cluster.

VI. CONCLUSION AND FUTURE WORK

In this work, an FFT-based convolutional neural network was successfully trained to identify the objects present in an image. The proposed FFT based convolutional neural network was able to improve the training time during convolution by reducing the number of steps needed to learn the feature and train the network. By comparing the Intersection Over Union (IoU) scores, we were able to determine the efficiency of the detection. The FFT was able to improve the training time during convolution from 600-700 ms/step to 400-500 ms/step. We compared FFT based model with non-FFT based model there and noticed an improvement in the Intersection Over Union (IoU) scores. In our work, for 100 epoch FFT based convolution IoU score was 0.83 and for non-FFT based model the IoU score was 0.52. The comparison of IoU scores implies that FFT based CNN has a better accuracy when compared to non-FFT CNN.

In conclusion we can say FFT-based convolution implemented in a CNN architecture can improve the convolution time and accuracy. This work can be done by comparing the model using multiple metrics and comparing the current model with many different object detection models. It can also include comparing the effects of the model using multiple datasets with varied objects. This study can be further improved by finding the object position on a 3D environment to identify its position more precisely.

These results are promising and warrant further investigation. The DFFT can be implemented as a single layer NN [26], so it would seem that U-Net would be capable of learning this operation if it were useful for segmentation. Why U-Net is unable

to transform or utilize the frequency domain of an image on its own is unclear but it is clear that performing the FFT at the onset is advantageous for the learning process.

ACKNOWLEDGMENT

This research work is supported by National Institute of Health (NIH) under Grants No: R15GM120669. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

- [1] Image recognition in retail market worth \$3.7 billion by 2025.
- [2] Simon Andermatt, Simon Pezold, and Philippe Cattin. Multi-dimensional gated recurrent units for the segmentation of biomedical 3d-data. In *Deep learning and data labeling for medical applications*, pages 142–151. Springer, 2016.
- [3] Jianxu Chen, Lin Yang, Yizhe Zhang, Mark Alber, and Danny Z Chen. Combining fully convolutional and recurrent neural networks for 3d biomedical image segmentation. In *Advances in neural information processing systems*, pages 3036–3044, 2016.
- [4] François Chollet et al. Keras, 2015.
- [5] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention*, pages 424–432, 2016.
- [6] Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.
- [7] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, KDD’96*, pages 226–231, Portland, Oregon, August 1996. AAAI Press.
- [8] Xinting Gao, Stephen Lin, and Tien Yin Wong. Automatic feature learning to grade nuclear cataracts based on deep learning. *IEEE Transactions on Biomedical Engineering*, 62(11):2693–2701, 2015.
- [9] Joseph W. Goodman. *Introduction to Fourier Optics*. W.H. Freeman, Macmillan Learning, fourth edition edition.
- [10] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-CNN. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.
- [11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [13] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989.
- [14] Vebjorn Ljosa, Katherine L Sokolnicki, and Anne E Carpenter. Annotated high-throughput microscopy image sets for validation. *Nature methods*, 9(7):637–637, 2012.
- [15] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. IEEE, 2016.
- [16] Raghu Muthyalampalli. Implementation of fast fourier transform for image processing in directx 10, 2009.
- [17] Joseph Redmon and Ali Farhadi. Yolo9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7263–7271, 2017.
- [18] O. Ronneberger, P.Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of LNCS, pages 234–241. Springer, 2015.
- [19] Pierre Sermanet, David Eigen, Xiang Zhang, Michaël Mathieu, Rob Fergus, and Yann LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv:1312.6229*, 2013.
- [20] Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Geert Litjens, Paul Gerke, Colin Jacobs, Sarah J Van Riel, Mathilde Marie Winkler Wille, Matiullah Naqibullah, Clara I Sánchez, and Bram van Ginneken. Pulmonary nodule detection in ct images: false positive reduction using multi-view convolutional networks. *IEEE transactions on medical imaging*, 35(5):1160–1169, 2016.
- [21] Wei Shen, Feng Yang, Wei Mu, Caiyun Yang, Xin Yang, and Jie Tian. Automatic localization of vertebrae based on convolutional neural networks. In *Medical Imaging 2015: Image Processing*, volume 9413, page 94132E. International Society for Optics and Photonics, 2015.
- [22] Marijn F Stollenga, Wonmin Byeon, Marcus Liwicki, and Juergen Schmidhuber. Parallel multi-dimensional lstm, with application to fast biomedical volumetric image segmentation. In *Advances in neural information processing systems*, pages 2998–3006, 2015.
- [23] Neda Tavakoli. Seq2image: Sequence analysis using visualization and deep convolutional neural network. In *IEEE COMPSAC*, pages 1332–1337, 2020.
- [24] Stéfan van der Walt, Johannes L. Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D. Warner, Neil Yager, Emmanuelle Goullart, Tony Yu, and the scikit-image contributors. scikit-image: image processing in Python. *PeerJ*, 2:e453, 6 2014.
- [25] Artem Vasilyev. CNN optimizations for embedded systems and fft. *Stanford University Report*, 2015.
- [26] Rosemarie Velik. Discrete Fourier Transform Computation Using Neural Networks. In *2008 International Conference on Computational Intelligence and Security*, volume 1, pages 120–123, December 2008.
- [27] Yuanpu Xie, Zizhao Zhang, Manish Sapkota, and Lin Yang. Spatial clockwork recurrent neural network for muscle perimysium segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 185–193. Springer, 2016.