IE:5995 Deep Learning Progress Report: Articular Cartilage Segmentation

Linjun Yang and Bakir Hajdarevic
University of Iowa
Iowa City, IA 52242
linjun-yang@uiowa.edu, bakir-hajdarevic@uiowa.edu

Abstract

The gold standard for evaluating articular cartilage tissue in histology images relies heavily on human-eye observations. However, this method is not only subjective but also time consuming. In this project, we propose an automated evaluation system using state-of-art deep learning methods. Our approach consists of first segmenting stained cartilage tissue using the fully convolutional neural network, U-Net [1]. This is followed by instance segmentation of articular cartilage cells. In this step, we compare the results of the U-Net and a deep contour-aware network (DCAN) with both achieving similar performance [2]. We hope our method will alleviate histology based approaches for the evaluation of the articular cartilage.

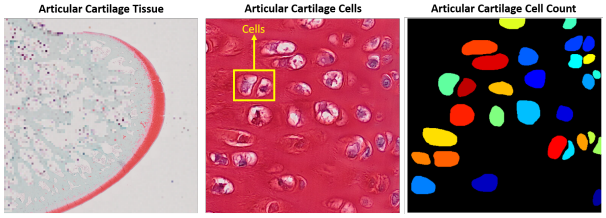

Figure 1. Overview of the main goals of this project. First, a method is required for segmenting the red stained articular cartilage tissue. Within this segmented area, articular cartilage cells are segmented. The resulting binary mask of the segmented cells are then counted. This cell count provides valuable insight in the health of articular cartilage tissue. An automated version of this process will reduce the time and improve the accuracy over human eye cell counting methods.

1. Introduction

Study of osteoarthritis (OA) on articular joints typically relies on the use of animal subjects that are injected with the OA disease. Thereafter, evaluations of healthy and diseased tissue are performed using biochemical based analysis methods. Histology based approaches are considered as the golden standard for the evaluation of the articular cartilage, which is the focus when studying the degenerative joint. This method starts with a section of the target tissue stained using cartilage-cell targeted biochemicals. The stained tissue section is then loaded onto a glass slide and covered with a glass cover. By observing the resulting specimen slide under a microscope, the researcher can obtain crucial tissue information including tissue structure, composition, and cellularity. An evaluation of the tissue health is then performed using several evaluation practices. Most of these require an experienced observer to assess the specimen and provide a set of scores as an assessment of the tissue. However, this human based evaluation method is highly subjective and will thus bias the scores. As an alternative, a fully automated evaluation system of the microscopic images may benefit OA research by providing a fast and accurate evaluation of the cartilage health. Figure 1 provides an overview of the main goals of this project.

2. Related Work

The recently proposed neural network, U-Net, has been proposed and shown to perform very well for biomedical image segmentation, especially for tissue and cellular segmentation [1]. The entire architecture can be interpreted as an encoder-decoder structure. As data is processed deep in the network, higher level features can be learned and then upsampled through the decoder process. The skip connections between the encoder and the decoder combine information from higher resolution to the information decoded from those high level features. These also prevent the issue of vanishing gradient from a training perspective. Data augmentation applied in their original network may help prevent overfitting. The U-Net be slightly altered and used for both the cartilage tissue and cell segmentations in this project.

One of the challenges we faced with segmented cells using a deep learning approach was distinguishing the edges between clusters of cells. To address this challenge, we turned to implmenting a deep contour-aware netowork (DCAN) that was successfully tried in segmenting glands

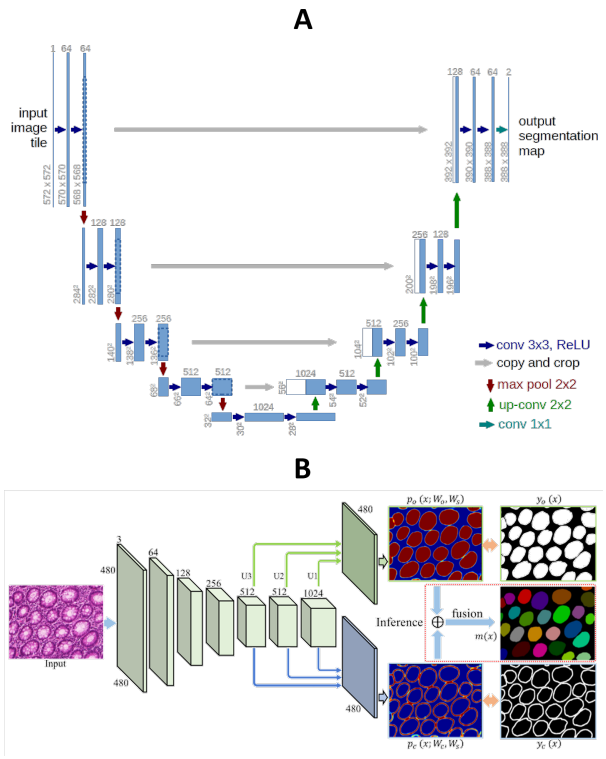

Figure 2. The architecture of U-Net (A) [1] and DCAN (B) [2].

and nuclei [2]. This network functions by combining multi-level contextual features with auxiliary supervision with cell-object contours information [2]. Figure 2 provides the architecture of both the DCAN and U-Net.

3. Data

All articular cartilage images used were given by the University of Iowa's Orthopedic Biomechanics Lab. The MATLAB application, Image Segmenter (MATLAB v9.0, The Mathworks Inc.), was used to generate manually segmented "ground truth" masks. Table 1 lists how images were partitioned between training, validation and testing for segmentation tasks. A variety of healthy and injurious cartilage cross sectional images were used. More unhealthy images were used as injurious cartilage tissue does not stain as well as healthy. By using a variety of ground truth images, we hoped to avoid the effects of class imbalance. For tissue segmentation, we split the data into approximately 47 percent training and about 13.3 percent for validation and testing. During the cell segmentation task, the data was split into 80 percent training, 10 percent validation, and 10 percent testing.

	Training	Validation	Testing
Tissue	35	10	10
Cell	100	20	20

Table 1. Summary of how the collection of ground truth images were partitioned for both tasks. A variety of healthy and injurious cartilage cross sectional images were used. More unhealthy images were used as injurious cartilage tissue does not stain as well as healthy to help mitigate the effects of class imbalance.

Augmentation Type	Change
Hue Adjustment	Adjust image hue in range (0.5,1.0)
Brightness Adjustment	Adjust brightness in (0.9,1.0)
Elastic Deformation	Scale images in x and y direction
Image Gaussian Blurring	Convolve image with gaussian function
Image Reflection	Flip image vertically/horizontally
Image Rotation	Rotate image (0,90) degrees

Table 2. Summary of the 6 various data augmentation methods used for both segmentation tasks. Blurring was not used for tissue segmentation.

3.1. Data Augmentation

As all the images in our dataset were acquired using a microscope, lighting conditions and contrast vary slightly between images. To mitigate potential detrimental effects of these subtle discrepancies, we have applied a data augmentation approach for adjusting the color of our images as well as performing a number of other transformations. Table 2 describes the data augmentation approaches used. These include linear transformation, brightness transformation, elastic deformation as well as hue adjustment. The type of data augmentation used for both tasks was essentially the same except we omitted blurring for the tissue segmentation data set.

4. Methods

The processing pipeline develoepd for this project is shown in Figure 3. Overall, two segmentation tasks are performed: one for articular cartilage tissue and another that segments articular cartilage cells.

4.1. Tissue Segmentation

We have made several changes to the original U-Net model. The original architecture takes input of size (572, 572, 1) and output image of shape (388, 388, 2). The image size shrinking during each convolution constraints input size selection accordingly for the correct calculation in the model. In our adapted model, modifications are made as follow. A simpler network is built where the input size is (512, 512, 3). The size of the input image for each convolution operation is preserved. This leads to a much more symmetric U-Net architecture. Therefore, cropping of the activation map from the contracting path is not needed when

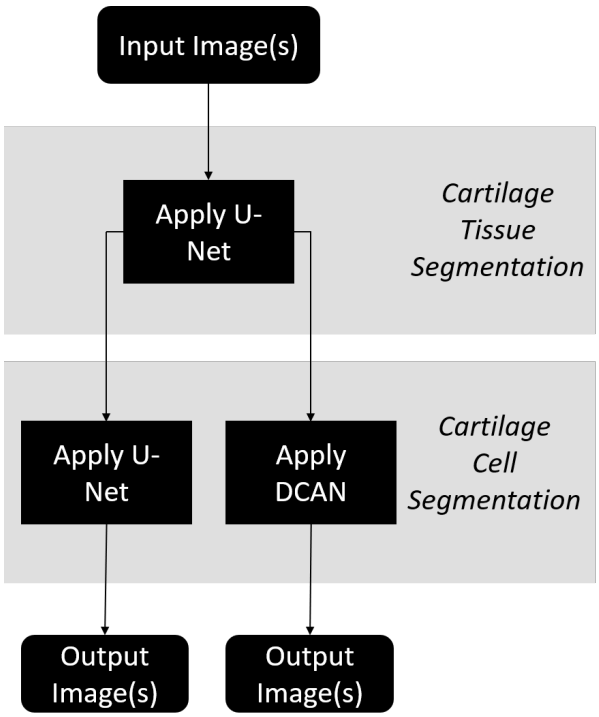

Figure 3. Overview of the processing pipeline developed for instance segmentation of articular cartilage cells. Two different neural networks were compared for optimal cell segmentation.

doing the concatenation along the expansive path. Before the rectified linear unit (ReLU) activation, the convoluted image is applied with Batch Normalization (BN) for each channel. Batch normalization layer is also added after each concatenation operation. Accordingly, this can deal with the covariant shift and also improve the network training progress. At the output layer, we use sigmoid activation function since we only have two classes and that gives output of shape (512, 512, 1). Adam optimizer is used to train the model. The loss to be optimized is a combination of binary cross entropy loss and dice loss [3]. This combined loss is expected to deal with class imbalance problem.

4.2. Cell Segmentation

A similar adaption to the U-Net is applied for the task of cell segmentation. The input size is (256,256,3). The loss weight map is precomputed and applied to account for the class imbalance as well as to force the model to learn to separate the cells that are close to each other. The weighted binary cross entropy loss is optimized using Adam optimizer.

We adapted a similar network as the original cited DCAN. Auxiliary supervision was accomplished by concatenating three output images into one output image. Input and output images had sizes of (512, 512) for all channels. As opposed to the U-Net, in addition to whole object masks

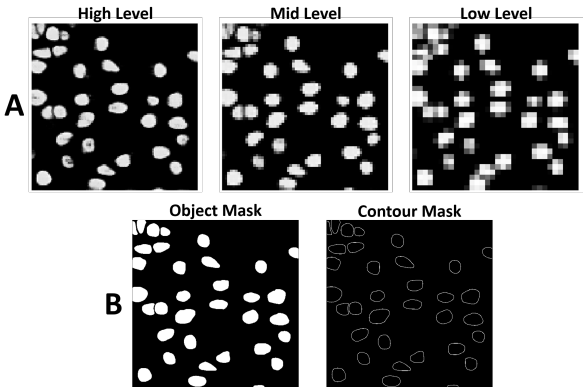

Figure 4. Examples segmentation results of tbe auxiliary supervision (A) and the object and contour masks input into DCAN (B).

we also input contour object masks to supplement the training. Supplementing the addition of contour masks would in theory help with the vanishing gradient problem by providing edge information of cells in close proximity. Figure 4 provides examples of the theory of multi-level contextual features with auxiliary supervision as well as the differences between an object mask and an object contours mask. Finally, a binary cross entropy loss function was used for both the U-Net and DCAN.

5. Results and Discussion

5.1. Tissue Segmentation

Figure 5 provides several test example results from the tissue segmentation task. Figure 6 provides the epoch training history of the tissue segmentation task. An average dice coefficient of 0.9032 is achieved. The lowest value of 0.5285 is obtained from the image in the 3rd row in the figure 5. All other (9 images) test predictions had values above 0.9. The resulting model gives an overall good segmentation of the cartilage tissue while performing poorly on a few images that are not stained well. This poor staining is expected in tissues that have sustained some sort of injury and as a result are inflamed. Furthermore, the poor segmentation for these type of images may be attributed to how the data set contains only a small portion of images with poor staining. We may equivalently characterize this as a type of class imbalance problem. By re-training the network with a larger number of images of injured cartilage tissue, we may potential alleviate the problems imposed by class imbalance.

5.2. Cell Segmentation

Table 3 provides a summary of the results of inputting the network with test images (20). Shown are the average dice coefficient and F1 scores computed for the test images.

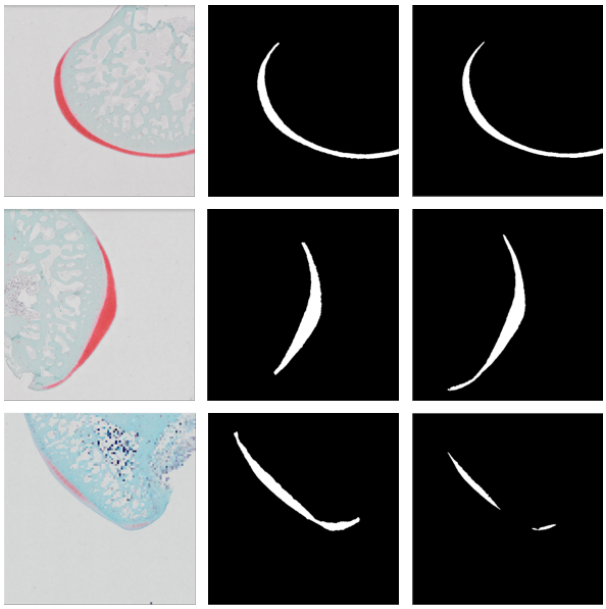

Figure 5. Examples of input images, ground truth segmented images, and predicted segmentations output from the implemented U-Net. The first two rows correspond to successful segmentation predictions. The last row corresponds to a failed segmented prediction which may be attributed to the lack of the red stain in the input image.

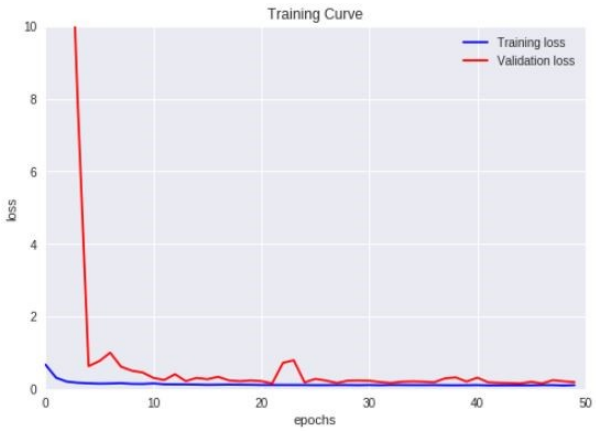

Figure 6. The training progress of both the training loss and validation loss as a function of epoch number for the tissue segmentation task. Both losses decreased to 0 after about 10 epochs.

As seen for tissue segmentation, a class imbalance issue occurs for cell segmentation. U-Net achieves an average dice coefficient of 0.8846 and a F1 score of 0.8469. Its prediction on some cells with less stained appearance is poor and results are low in metrics value (lowest 0.4898 for F1 and 0.4809 for dice, respectively). The segmentation results

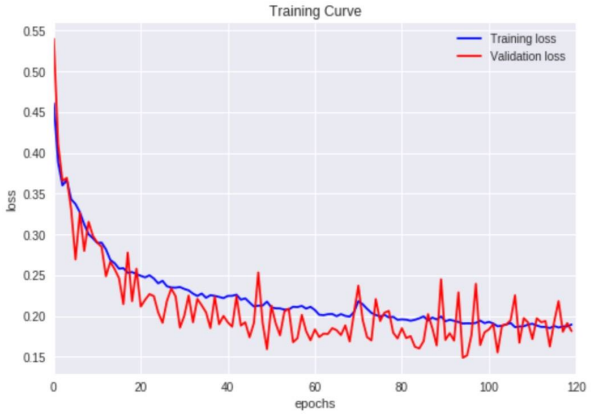

Figure 7. The training progress of both the training loss and validation loss as a function of epoch number for the U-Net cell segmentation task. Both losses decreased to 0 after about 10 epochs.

	U-Net (Tissue)	U-Net (Cell)	DCAN
Dice Coefficient	0.9032	0.8846	0.6483
F1 Score	-	0.8469	0.6749

Table 3. Shown are the average dice coefficient and F1 scores computed for the test images of both the U-Net and DCAN.

for DCAN were worse, yielding average dice loss and F1 scores of 0.6483 and 0.6749, respectively. Maximum and minimum dice loss values were 0.8761 and 0.0772. Similar metric were obtained for the F1 score: 0.9054 and 0.0645. We also observed that DCAN also seems to perform poorly on images where there are a few number of cells. Figures 7 and 8 provide the epoch history for both training and validation loss for U-Net and DCAN , respectively. Figure 9 provides example test results of the cell segmentation task using either U-Net and DCAN. The different colors correspond to different cells detected during post processing of the resulting segmentation masks. This was accomplished using MATLAB software (MATLAB v9.0, The Mathworks Inc.).

Similar to the tissue segmentation task, a class imbalance issue was observed for the training of both U-Net and DCAN. This problem may also alleviated by using a larger number of images as well using a variety of cartilage tissue types and areas (i.e. middle of cartilage tissue versus edge of tissue). Furthermore, we should also emphasize on training the network with greater number of images that contain a fewer number of cells but are still clustered. Using this approach, we may observe an improvement in the DCAN's performance as it seemed to struggle with images of the aforementioned type.

6. Conclusion

We have shown our method for instance segmentation of articular cartilage cells using a deep learning approach. Us-

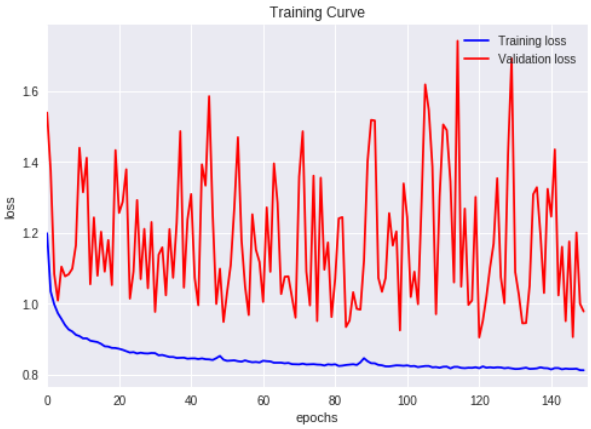

Figure 8. The training progress of both the training loss and validation loss as a function of epoch number for the DCAN cell segmentation task. The training loss decreases after 10 epochs but the validation loss is sporadic throughout

ing U-Net is sufficient for segmenting both articular cartilage cells and tissues as it yielded high dice coefficient and F1 scores. The implemented DCAN performed poorly in comparison to the U-Net. For both training steps in the processing pipeline, the class imbalance is expected to be the primary culprit of the poor model performance for certain special cases. By adding more special cases for the training set as well as the validation set should mitigate the detrimental effects of class imbalance.

7. Future Work

A number of other future works remain to be done. This includes increasing the number of ground truth images input into the processinge pipeline to help avoid the class imbalance issue. The primary future work is that a method must be incorporated into the pipeline for counting the number of cells from the final segmentation task. Figure 10 provides a potential processing pipeline that uses the output mask from the tissue segmentation task into smaller images. These smaller images will then each be input into the network for cell segmentation. Once all input images have been processed, the resulting segmentations are conjoined together in order to perform a cell count of the articular cartilage tissue. This count will aid in characterizing and quantifying histologic articular cartilage tissue.

8. Acknowledgements

We would like to acknowledge and thank Professors Baek and Goetz for their invaluable assistance and feedback throughout the duration of this project. In addition, we would like to thank our various fellow peers that have assisted us throughout the semester.

9. References

[1] Ronneberger, Olaf et al. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015.

[2] Chen, Hao et al. "DCAN: Deep contour-aware networks for object instance segmentation from histology images," Medical Image Analysis, vol. 36, pp. 135-146, 2017.

[3] Milletari, Fausto et al. "V-net: Fully convolutional neural networks for volumetric medical image segmentation." 3D Vision (3DV), 2016 Fourth International Conference on. IEEE, 2016.

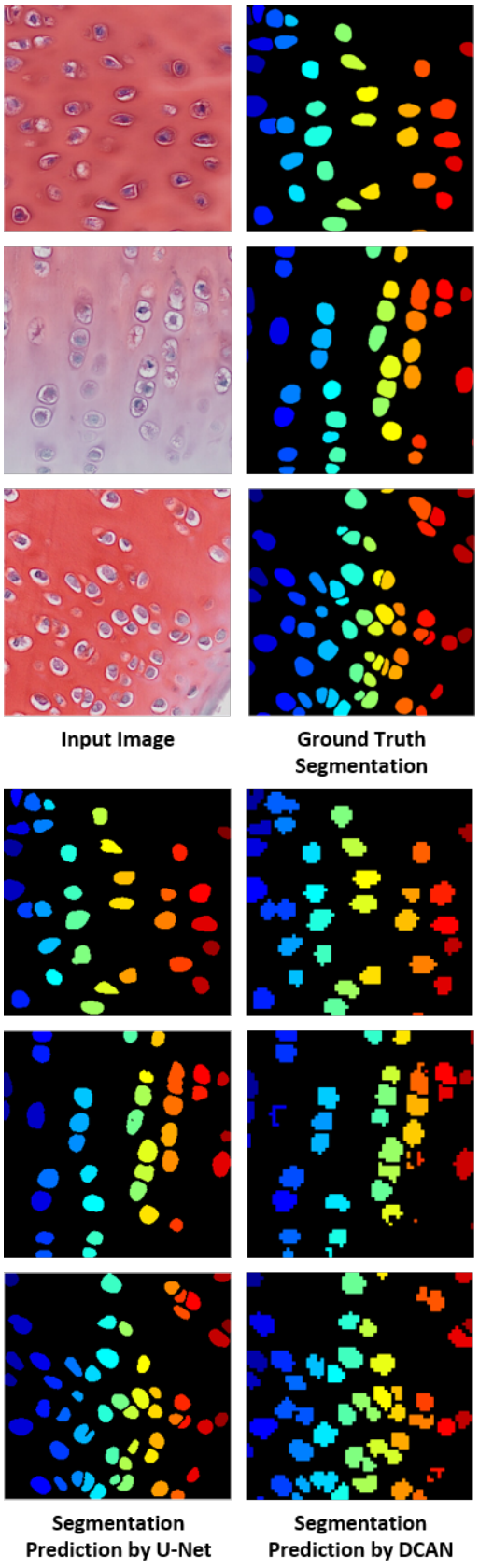

Figure 9. Example results for several test inputs to a cell segmentation network (U-Net and DCAN). The different colors correspond to different cells detected during post processing of the resulting segmentation masks.

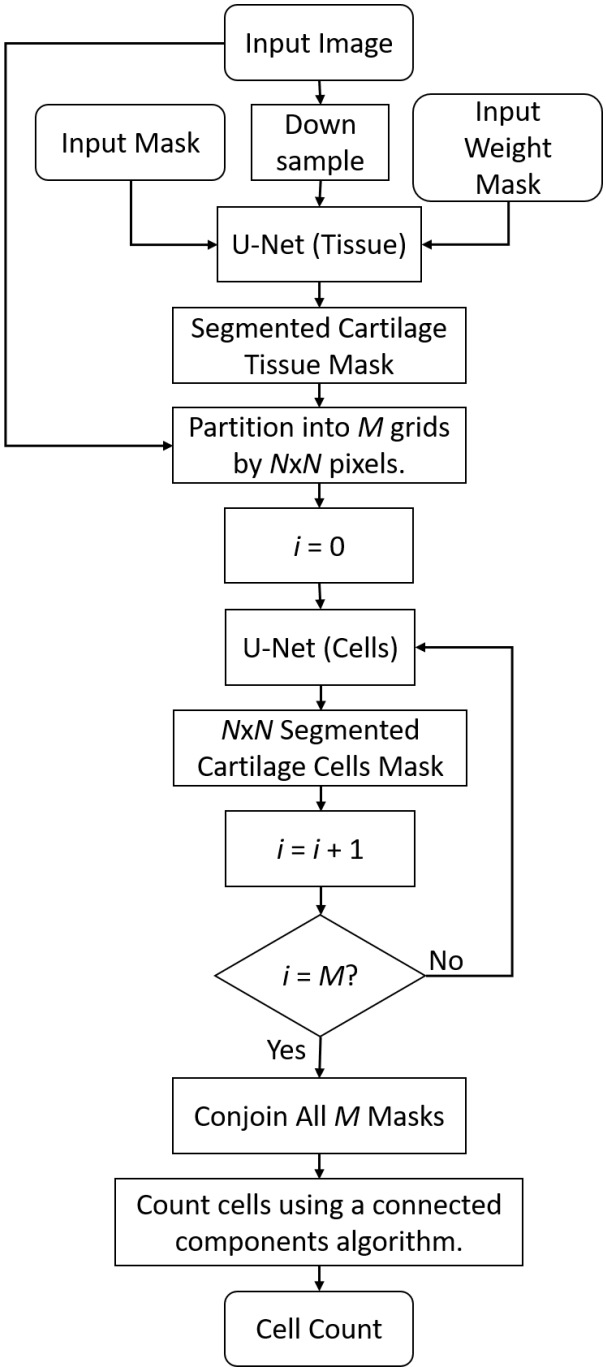

Figure 10. The proposed processing pipeline as part of the future work. Cartilage tissue image segmentation results are partitioned into smaller chunks that are used to train the cell segmentation neural network. Image results from this network are conjoined back into the original image. A method for counting the number of cells in the binary mask is applied.