

U-Net Based Multi-Output Network for Lung Disease Segmentation and Classification using Chest X-ray Dataset

Jaiden X. Schraut

September 6, 2022

Abstract

Medical Imaging Segmentation of Chest X-rays is used for the purpose of identification and differentiation of lung cancer, pneumonia, COVID-19, and similar respiratory diseases. Widespread application of computer-supported perception methods into the diagnostic pipeline has been demonstrated to increase prognostic accuracy, and aid doctors in efficiently treating patients. Modern models attempt the task of segmentation and classification separately, and improve diagnostic efficiency; however, to further enhance this process, this paper proposes a multi-output network which follows a U-Net architecture for image segmentation output and features an additional CNN module for auxiliary classification output. The proposed model achieves a final Jaccard Index of .9634 for image segmentation and a final accuracy of .9600 for classification, on the COVID-19 RADIOGRAPHY DATABASE.

Keywords Chest X-ray, Deep Learning, Image Segmentation, Image Classification

1

Introduction

Medical imaging is a vital and necessary tool for understanding the structure of the human body in diagnosis, treatment, research, and clinical contexts.[1, 9]. Historically, chest X-ray has been the most commonly utilized radiological examination. In 2006, approximately 129 million chest x-rays were obtained in the United States alone [13]. The high demand can be attributed to the low cost, and widespread availability of X-rays as compared to alternative imaging tests, such as a diagnostic ultrasound, computed tomography (CT), or magnetic resonance imaging (MRI) [5]. As such, the chest x-ray is typically the first medical image captured, and remains vital in prognosis and treatment thereafter [15]. In order to accurately diagnose respiratory disease, lung segmentation of chest X-rays is necessary. Medical imaging segmentation plays a critical role in analyzing medical imaging through feature extraction, a process which partitions an image by identifying homogeneous properties [8]. A region may be divided by brightness, and texture in the characteristics of adjacent pixels. Pixels in images take values between 0-255 in greyscale. Radiologists perform manual segmentation, a time-consuming and arduous task,

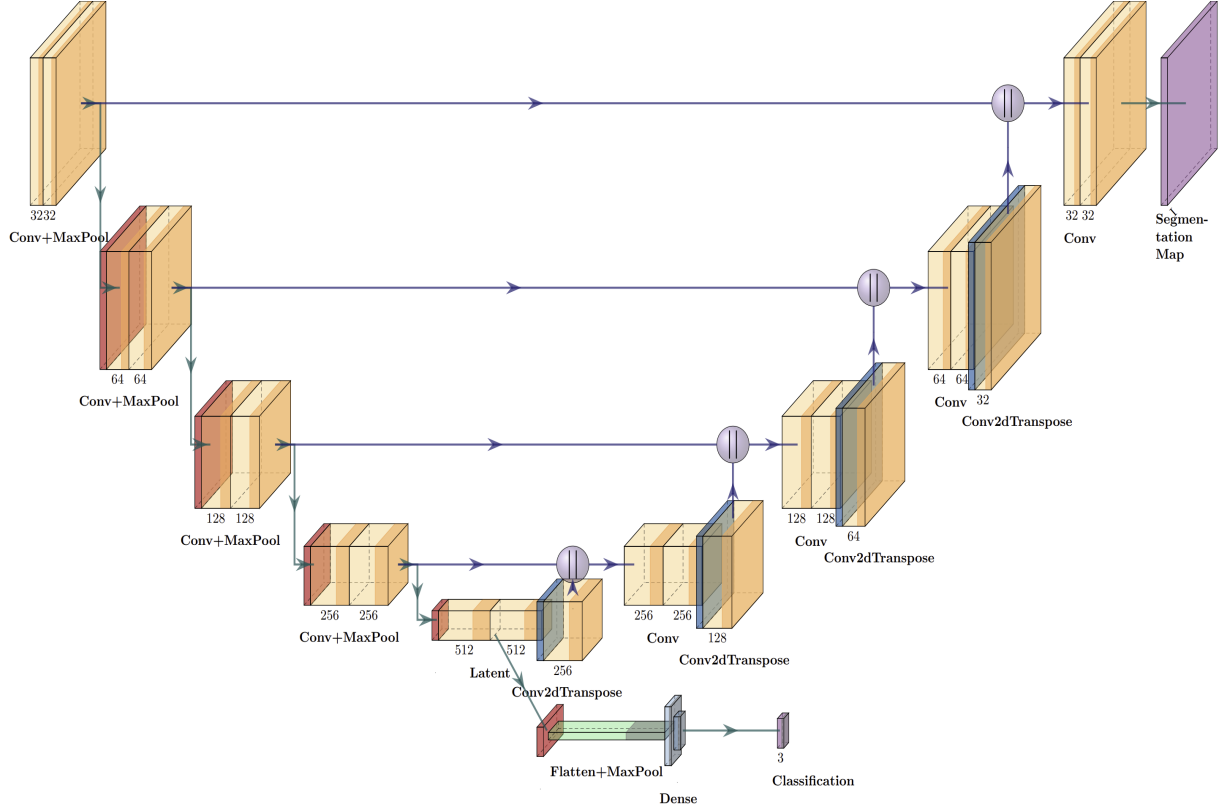
which suffers from high-observer variability due to conflicting interpretation[7, 2]. In lung segmentation, it is very difficult to identify small or subtle abnormalities, or to precisely differentiate between pathological patterns of diseases [3]. Older chest x-ray images were sized 128 by 128, with modern imaging having significantly higher resolution. Even at this relatively low resolution, ($128 * 128 = 16,384$ pixels) it is considered high dimensional data for human eyes to precisely observe. Differentiation of non-COVID viral pneumonia and COVID-19, is important for determining appropriate treatment [6]. The inherent difficulty in human interpretation of chest X-ray analysis has led researchers to pursue automated segmentation algorithms for this purpose. Similarly, deep convolutional neural networks (CNN)[10] have demonstrated effective image classification, image segmentation, and semantic segmentation, a process of classifying each individual pixel of an image[4, 18, 12]. Fully Convolutional Networks (FCN)[12], a type of CNN, have been used extensively in modern semantic segmentation algorithms. In [16], the researchers advocate for U-Net, an encoder-decoder network of FCN for biomedical image segmentation. Since its introduction, the U-Net architecture has demonstrated significant success, thus, recent studies have focused on further developing and applying this architecture rather than proposing new architectures and concepts [14]. In the future, it is reasonable to predict that demand for comprehensive medical image analysis may require simultaneous segmentation and classification to reduce doctor's workloads. This paper proposes a modification to the existing U-Net model which accomplishes the task of classification and segmentation concurrently.

Proposed Methods

In recent years, the U-Net architecture [16] has been recognized as one of the leading methods in medical image segmentation. The U-Net architecture builds on the Fully Convolutional Network architecture, by implementing upsampling operators in the contracting network, and symmetry in the contracting and expansive paths. The name comes from the U-shaped architecture represented by the contracting and expansive paths. Upsampling operators increase the output resolution to allow for precise segmentation. This paper proposes to update "upsampling" with "Conv2DTranspose" action. The symmetry within the model supports an overlap-tile strategy, which allows the network to learn efficiently even with little training data, a persistent issue in the field of medical imaging. U-Net also supports multi-scale prediction and deep supervision, as U-Net uses skip connection

¹J. Schraut is with the Department of Computer Science, College of Literature, Science, and the Arts, University of Michigan, Ann Arbor, MI, 48103 USA (contact: x.schraut@gmail.com)

Figure 1: **Diagram of proposed method.** This is the executive diagram of the workflow for this paper.



across the model, more low-level features have presence in the segmented output. U-Net was chosen both for the task of lung segmentation, and for the ease of implementation in the novelty of this model, an auxiliary classification output. The innovation here is the integration of a CNN [10] to the latent layer of the U-Net. It can be said that this model shows similarity to the Siamese network architecture as both the U-Net and CNN take advantage of the contracting path for feature extraction. While the model only takes one input, it's worth mentioning that the model utilizes identical weights and biases to generate output. A traditional CNN is composed of a similar feature extraction module, a classification module and a probabilistic distribution to display output. The classification element of the multi-output approach might achieve a potential benefit from backpropagation through the segmentation element. The contracting path serves both the U-Net and the CNN, similar to the concept of a siamese network.

Proposed Architecture

The proposed network architecture is shown in figure 1. Much like the standard U-Net model, the contracting path is composed of multiple convolutional blocks each accompanied by a 2x2 max pooling layer. Each successive layer has stride 2 for downsampling, which doubles the number of feature channels. At the latent layer, the model splits into a max pooling layer, a flatten layer, and 3 dense layers for classifi-

cation. All dense layers have ReLU activation function (1), with the exception of the final dense layer. The nature of the classification problem requires the final dense layer to be a softmax activation function, as there are multiple classes. On the expansive path of the U-Net model, following the principles of symmetry, an equal amount of convolutional blocks accompanied by a 2x2 Conv2dTranspose layer, with stride 2, followed by a concatenate layer. The Conv2dTranspose layer behaves similarly to an inverse convolutional layer with a 2x2 stride. The layer has an upsampling effect while interpreting the input data to assure detail. The concatenate layer combines the cropped feature map from the opposite convolutional block in the contracting path. This is performed as edge pixel data is lost in the process. The final layer is a 1x1 convolutional layer connect to the component feature vectors to relevant classes. Each convolutional block consists of two 3x3 unpadded convolutional layers, two Batch Normalization layers, and two ReLU activation layers for the contracting path, and two Leaky ReLU activation layers for the expansive path. Batch Normalization is used in this model to further enhance model performance by stabilizing the learning process through normalizing activation vectors without compromising on training convergence. The U-Net based multi-output architecture was realized through the Tensorflow and Keras Python libraries.

Activation Functions

In order to improve the performance of the model, the effects of several activation functions were compared, prompted by the work of [19]. The authors investigated the differences in performance of activation functions in image analysis. The study concluded that in the task of medical image analysis, there may be marginal improvement in results based on selected activation function. In response to their findings, the model was evaluated separately with each of these functions in image segmentation. The activation functions were implemented in the decoder block of the model.

$$\text{ReLU}(z) = \max(z, 0) \quad (1)$$

$$\text{Leaky ReLU}(z) = \max(\alpha z, z) \quad (2)$$

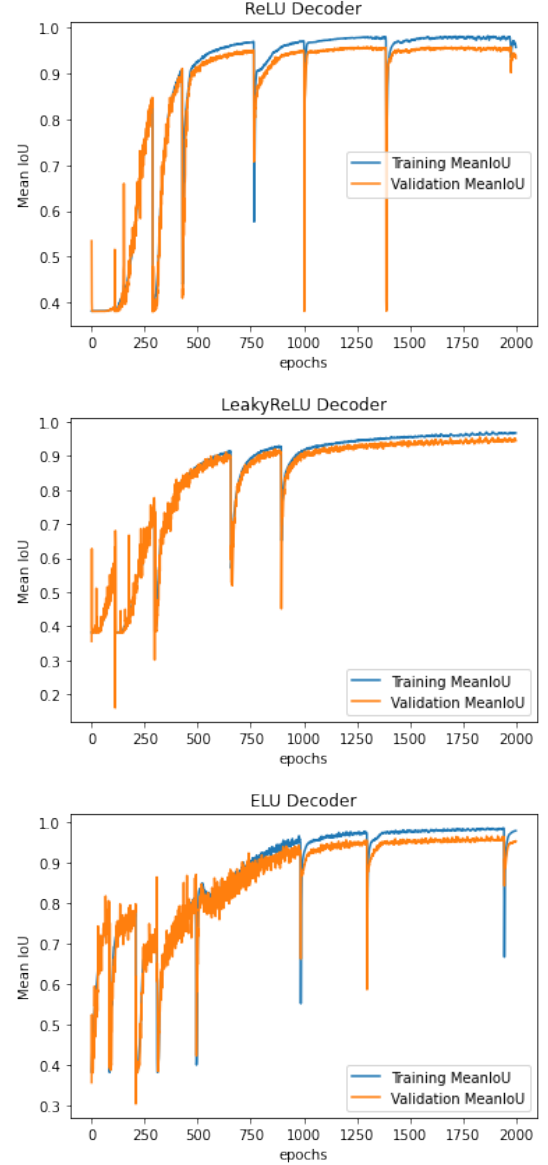
where α is a tuning parameter.

$$\text{ELU}(z) = \max(z, \alpha(e^z - 1)) \quad (3)$$

The standard rectified linear unit (see equation 1) activation function takes an input and allows the information to pass through if it is non-negative. Updated versions of ReLU can also be used. Leaky rectified linear unit (see equation 2) is one of its versions and it lets information pass through with a scaling factor of α . Exponential linear unit (see equation 3) is another version of ReLU with exponential scaling. ReLU is the predominant activation function for training deep and traditional neural networks. The ReLU activation function accelerates the training rate of deep and traditional neural networks compared to conventional activation functions as the derivative of ReLU is 1 for a positive input. The network saves time for computing error metrics in the process of training. As well as increasing computational efficiency, ReLU does not prompt the vanishing gradient problem when the model increases layers. The vanishing gradient problem occurs when the partial derivative of the loss function approaches zero. In neural networks that rely on backpropagation or gradient-based learning, the partial gradient effects the weights proportionally to the value of itself. A value approaching zero will become so insignificant that the model may prevent the weight from changing its value, stopping the model from training altogether. Hyperbolic tangent and sigmoid activation functions are known to be susceptible to this problem as well. Leaky ReLU is a modification of ReLU, producing small output values given a negative input in comparison to a value of zero given by the ReLU function in the same scenario. This modification prevents the dying ReLU problem, where a neuron may learn a large negative bias and continually output the same value. This neuron is 'dead,' and will now have no future effect on the model, as it is improbable to learn when the function gradient is at 0. The nonzero output value given a negative input gives a 'dead' neuron a chance to become active. The ELU activation function is an identity function for non-negative inputs, like ReLU. ELU differs from ReLU due to an α constant which determines function smoothness for negative inputs. ELU does not suffer from the dying neuron problem and the vanishing gradient problem, while generalizing better. ELU tends to have a comparatively faster convergence

time than ReLU, though it is slower to compute due to the non-linearity calculations necessary for negative inputs.

Figure 2: Comparative Activation Performance.



The model was trained on the COVID-19 Radiography Dataset for 2000 epochs with differing activation functions in the expansive path in order to reveal which activation function produced superior performance. ELU, in this experiment, had a slower convergence than ReLU and Leaky ReLU, while ReLU and Leaky ReLU had similar convergence speeds, at around 600 epochs. Final validation results indicate that no model is significantly better or worse, with Leaky ReLU outperforming both other activation functions by a marginal amount.

Figure 3: **Confusion Matrix.** This figure presents the confusion matrix which is defined using true positive (TP), true negative (TN), false negative (FN), and false positive (FP).

		Prediction outcome		
		p	n	total
actual value	p'	True Positive (TP)	False Negative (FN)	P'
	n'	False Positive (FP)	True Negative (TN)	N'
total		P	N	

Loss

The Intersection over Union (IoU) (Eq. 4), commonly known as the Jaccard index, is the area of overlap between the predicted segmentation divided by the area of union between the predicted segmentation and the ground truth. For application in image segmentation, the mean IoU of the image is calculated by taking the IoU of each class and averaging them.

$$\mathcal{J}(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (4)$$

where A and B are two sets.

This metric represents the ratio between the model interpretation of the lungs in comparison to the actual lungs in a given input chest X-ray. The metric used to evaluate the classification module of the model is accuracy, formally defined (Eq. 5). Accuracy is formally defined as the number of correct predictions over total predictions. The confusion matrix is defined below where the prediction outcome can be positive or negatives and the actual value is also positive or negative. In the segmentation component, the mask is binary and the binary values are treated as positive and negative, respectively. In classification, if the model predicts the identity of the image correctly, it is considered a true positive 3.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

Training

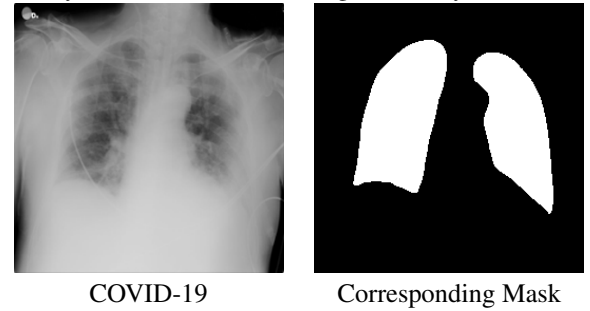
The model was trained with 900 COVID-19 chest X-rays, 900 Viral-pneumonia X-rays, and 900 Normal chest X-rays. Images were resized to (128 x 128), and 10% of the data was used for model validation. Adam was chosen as the model optimizer function. The model was trained on a batch size of 60, with a maximum of 2,000 epochs, on a Tesla P100-PCIE-16GB. The model tends to marginally improve, warranting the large amount of training epochs. The model was

tested with 100 COVID-19 chest X-rays, 100 Pneumonia X-rays, and 100 Normal chest X-rays.

Application

The COVID-19 Radiography Database, created by researchers from Qatar University, Doha, Qatar, and the University of Dhaka, Bangladesh along with their collaborators from Pakistan and Malaysia in collaboration with medical doctors, was used for training, validation and testing. This is a public dataset featuring 3616 COVID-19 positive cases along with 10,192 Normal, and 1345 Viral Pneumonia images. All images are available in PNG format in 299 x 299 resolution. Lung segmentation masks are included with corresponding chest X-ray images. The segmentation masks were manually generated by the researchers.

Figure 4: **A Sample of Chest X-ray Data.** The original data is 299 by 299 and the resized image is 128 by 128.



Results & Discussion

Table 1: **Experimental Results.** This table summarizes the results of the experiment.

Activation Decoder	Segmentation Jaccard Index	Classification Accuracy
ReLU	0.9521	0.9533
Leaky ReLU	0.9634	0.9600
ELU	0.9588	0.9467
[11]	0.9770	-
[17]	-	0.9590

The concept of multi-output models in AI is far from new. However, in the field of medical imaging, there does not yet seem to exist a similar application of a multi-output concept. Hence, there lacks appropriate candidates for comparison of model performance. Of all the models tested, the implementation of Leaky ReLU (Table: 1) in the expansive path produced the best results by a small margin in both the classification and segmentation tests. Change of activation function did not significantly improve the model.

The U-Net module of the multi-output architecture displays satisfactory lung segmentation results. State-of-the-art U-Net models[11] achieve similar Jaccard Index scores

Table 2: Comparison of time consumption for training.

Model	Training Time for 100 epochs
U-net (alone)	11min 39sec
CNN (alone)	2min 37sec
Combined	14min 16sec
Proposed model	11min 44sec

(Table: 1) in the task of chest X-ray lung segmentation, which provides ample support for the multi-output architecture concept. The same can also be said for modern classification techniques. Findings from the researchers in [17] also show similar testing accuracy scores (Table: 1) in lung disease classification. These findings corroborate the classification module of the multi-output architecture concept. The scores achieved in segmentation and classification can be further improved. To justify use of the proposed network, training times were of standalone CNN and U-Net models were added together and compared to the training time of the proposed network (Table: 2). All models were tested on the same hardware(see section: Training) with the same dataset(COVID-19 Radiography). The multi-output architecture demonstrates better training times than the combined training time of the CNN and U-Net model. The U-Net model was made identically to the proposed network (1), without the auxiliary classification output. These measures ensure that there are no extraneous variables which may effect the outcome of net training time between models. A comprehensive multi-output network will train faster than separate models which individually complete the same tasks. While, the model does not surpass other established models in medical imaging analysis, the proposed model shows better training time and performance on par with these models. At the moment, the exclusive function of the classification module is to identify a disease through lung segmentation of a given chest X-ray. In future work, training on full-resolution images or additional classification tasks can be implemented, such as severity grading of given diseases. Different classification models such as AlexNet, VGG16, or Google Inception could be implemented in a similar way to an existing segmentation architecture for better performance.

Conclusion

This paper has proposed a model architecture with concurrent image segmentation and classification output. The model is based on U-Net and CNN architecture, with Conv2dTranspose and Leaky ReLU to optimize the expansive path. All convolutional blocks utilize batch normalization. The proposed method achieves a Jaccard Index score of 0.9634 and a accuracy score of 0.9600 on the COVID-19 Radiography database for evaluating the segmentation and classification modules, respectively. As proof of concept, these findings are consistent with modern medical image analysis models, while demonstrating improved model training time. Though the concept of a concurrent segmen-

tation and classification architecture in medical imaging is relatively new, there is still much work to be done. The applications of deep learning technology can be adapted to other medical imaging techniques such as CT scan, MRI, diagnostic ultrasound, and others. There is much potential to benefit the field of medical image analysis by assisting human doctors with AI-enhanced solutions.

References

- [1] A. Abedalla, M. Abdullah, M. Al-Ayyoub, and E. Benkhelifa. Chest x-ray pneumothorax segmentation using u-net with efficientnet and resnet architectures. *PeerJ Computer Science*, 7:e607, 2021.
- [2] Y. Balabanova, R. Coker, I. Fedorin, S. Zakharova, S. Plavinskij, N. Krukov, R. Atun, and F. Drobniowski. Variability in interpretation of chest radiographs among russian clinicians and implications for screening programmes: observational study. *Bmj*, 331(7513):379–382, 2005.
- [3] E. Çallı, E. Sogancioglu, B. van Ginneken, K. G. van Leeuwen, and K. Murphy. Deep learning for chest x-ray analysis: A survey. *Medical Image Analysis*, 72:102125, 2021.
- [4] D. Ciresan, A. Giusti, L. Gambardella, and J. Schmidhuber. Deep neural networks segment neuronal membranes in electron microscopy images. *Advances in neural information processing systems*, 25, 2012.
- [5] N. England and N. Improvement. Diagnostic imaging dataset statistical release. *London: Department of Health*, 421, 2016.
- [6] A. Eslambolchi, A. Maliglig, A. Gupta, and A. Gholamrezanezhad. Covid-19 or non-covid viral pneumonia: How to differentiate based on the radiologic findings? *World Journal of Radiology*, 12(12):289, 2020.
- [7] M. S. Fasihi and W. B. Mikhael. Overview of current biomedical image segmentation methods. In *2016 International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 803–808. IEEE, 2016.
- [8] N. E. A. Khalid, S. Ibrahim, and P. Haniff. Mri brain abnormalities segmentation using k-nearest neighbors(k-nn). *International Journal on Computer Science and Engineering*, 3(2):980–990, 2011.
- [9] E. A. Krupinski. The importance of perception research in medical imaging. *Radiation medicine*, 18(6):329–334, 2000.
- [10] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [11] W. Liu, J. Luo, Y. Yang, W. Wang, J. Deng, and L. Yu. Automatic lung segmentation in chest x-ray images using improved u-net. *Scientific Reports*, 12(1):1–10, 2022.

- [12] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.
- [13] F. A. Mettler Jr, M. Bhargavan, K. Faulkner, D. B. Gilley, J. E. Gray, G. S. Ibbott, J. A. Lipoti, M. Mahesh, J. L. McCrohan, M. G. Stabin, et al. Radiologic and nuclear medicine studies in the united states and worldwide: frequency, radiation dose, and comparison with other radiation sources—1950–2007. *Radiology*, 253(2):520–531, 2009.
- [14] A. S. Panayides, A. Amini, N. D. Filipovic, A. Sharma, S. A. Tsaftaris, A. Young, D. Foran, N. Do, S. Golemati, T. Kurc, et al. Ai in medical imaging informatics: current challenges and future directions. *IEEE Journal of Biomedical and Health Informatics*, 24(7):1837–1857, 2020.
- [15] S. Raoof, D. Feigin, A. Sung, S. Raoof, L. Irugulpati, and E. C. Rosenow III. Interpretation of plain chest roentgenogram. *Chest*, 141(2):545–558, 2012.
- [16] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [17] A. Shelke, M. Inamdar, V. Shah, A. Tiwari, A. Husain, T. Chafekar, and N. Mehendale. Chest x-ray classification using deep learning for automated covid-19 screening. *SN computer science*, 2(4):1–9, 2021.
- [18] P. V. Tran. A fully convolutional neural network for cardiac segmentation in short-axis mri. *arXiv preprint arXiv:1604.00494*, 2016.
- [19] B. Xu, N. Wang, T. Chen, and M. Li. Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv:1505.00853*, 2015.