

PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

CondenseUNet: a memory-efficient condensely-connected architecture for bi-ventricular blood pool and myocardium segmentation

Hasan, S. M. Kamrul, Linte, Cristian

S. M. Kamrul Hasan, Cristian A. Linte, "CondenseUNet: a memory-efficient condensely-connected architecture for bi-ventricular blood pool and myocardium segmentation," Proc. SPIE 11315, Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling, 113151J (16 March 2020); doi: 10.1117/12.2550640

SPIE.

Event: SPIE Medical Imaging, 2020, Houston, Texas, United States

CondenseUNet: A memory-efficient condensely-connected architecture for bi-ventricular blood pool and myocardium segmentation

S. M. Kamrul Hasan^{a(✉)} and Cristian A. Linte^{a,b}

^aCenter for Imaging Science, Rochester Institute of Technology, NY, USA

^bBiomedical Engineering, Rochester Institute of Technology, NY, USA

ABSTRACT

With the advent of Cardiac Cine Magnetic Resonance (CMR) Imaging, there has been a paradigm shift in medical technology, thanks to its capability of imaging different structures within the heart without ionizing radiation. However, it is very challenging to conduct pre-operative planning of minimally invasive cardiac procedures without accurate segmentation and identification of the left ventricle (LV), right ventricle (RV) blood-pool, and LV-myocardium. Manual segmentation of those structures, nevertheless, is time-consuming and often prone to error and biased outcomes. Hence, automatic and computationally efficient segmentation techniques are paramount. In this work, we propose a novel memory-efficient Convolutional Neural Network (CNN) architecture as a modification of both CondenseNet, as well as DenseNet for ventricular blood-pool segmentation by introducing a bottleneck block and an upsampling path. Our experiments show that the proposed architecture runs on the Automated Cardiac Diagnosis Challenge (ACDC) dataset using half (50%) the memory requirement of DenseNet and one-twelfth ($\sim 8\%$) of the memory requirements of U-Net, while still maintaining excellent accuracy of cardiac segmentation. We validated the framework on the ACDC dataset featuring one healthy and four pathology groups whose heart images were acquired throughout the cardiac cycle and achieved the mean dice scores of 96.78% (LV blood-pool), 93.46% (RV blood-pool) and 90.1% (LV-Myocardium). These results are promising and promote the proposed methods as a competitive tool for cardiac image segmentation and clinical parameter estimation that has the potential to provide fast and accurate results, as needed for pre-procedural planning and / or pre-operative applications.

Keywords: Cine MR image, cardiac imaging, deep learning, image segmentation, learned group convolution, ventricle blood-pool, myocardium, weight pruning

1. INTRODUCTION

Cardiovascular diseases (CVDs) are the leading cause of death for both men and women in the United States (US) according to the American Heart Association and someone dies from distinct form of CVDs in every 38 seconds, based on 2016 data *. Even the number is set to reach 130 million by the year 2035 as projected by the American Heart Association.¹ Nevertheless, significant scientific advances revolutionized the cardiovascular research landscape on the promise of strategical improvement to augment the diagnosis of the patients suffering from CVDs.

Cardiac Magnetic Resonance Imaging (CMRI) has made a significant paradigm shift in medical imaging through the noninvasive quantification of the volumetric changes in the heart throughout the cardiac cycle. To further study the anatomical structure and functional characteristics of the heart, dynamic short-axis cine cardiac MRI has been used extensively in recent decades. From the machine learning perspective, cardiac image segmentation is a multi-class classification problem aiming to assign each voxel, a target label. Previously, traditional machine learning techniques had been shown to achieve good performance in cardiac image segmentation.²

(✉) Further author information:

S. M. Kamrul Hasan (E-mail: sh3190@rit.edu)

Cristian A. Linte (E-mail: calbme@rit.edu)

*<https://newsroom.heart.org/news/nearly-half-of-all-u-s-adults-have-some-form-of-cardiovascular-disease?>

Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling, edited by
Baowei Fei, Cristian A. Linte, Proc. of SPIE Vol. 11315, 113151J · © 2020 SPIE
CCC code: 1605-7422/20/\$21 · doi: 10.1117/12.2550640

Proc. of SPIE Vol. 11315 113151J-1

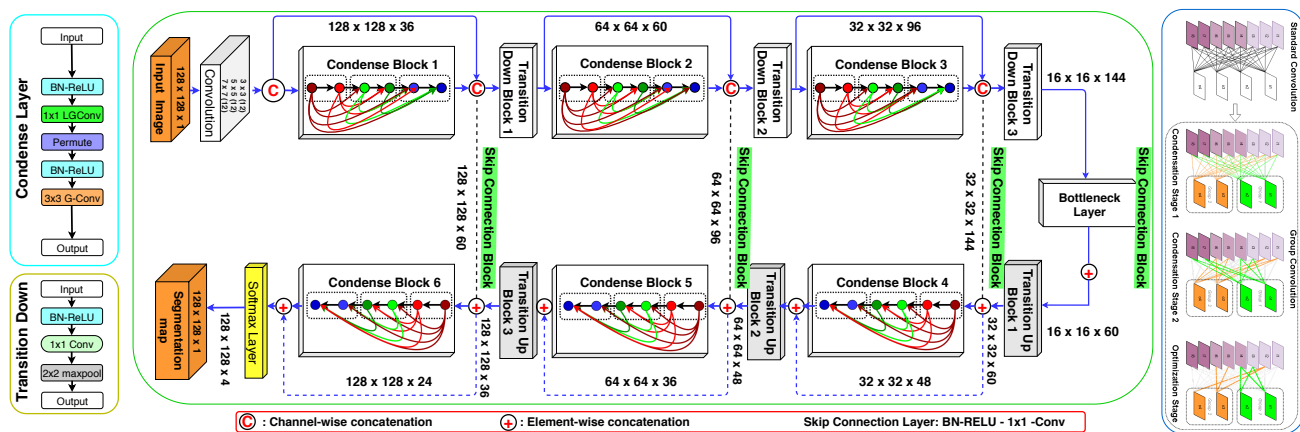


Figure 1: *CondenseUNet* architecture is comprised of a downsampling path and an upsampling path. Each Condense block (CDB) consists of 3 Layers with a growth rate of $k = 16$. The transformations within each CDB layer are mentioned in cyan rectangular box and transition-down block in the yellow rectangular box. The concept of Learned Group Convolution (LGConv) is illustrated in the blue rectangular box and explained in Section 2.1. The element-wise addition of Feature maps from both up and down-sampling path is matched by skip-connection block (red rectangular box). The gray blocks represent the transition-up operation which consists of 3×3 transposed convolution operation.

However, they require both prior information and manual interaction. However, manual segmentation can be susceptible to inter / intra-observer variability, hence the attempt to solve this problem automatically is desirable in the clinical setting.

The superior performance of Convolutional Neural Networks (CNNs) in solving high-level computer vision tasks has triggered the medical imaging research community to develop a powerful machine learning tool for medical image segmentation in learning intricate features in an end-to-end manner.³

Jain *et al.*⁴ designed a CNN model for cardiac image segmentation using a 2D and 3D segmentation pipeline. Isensee *et al.*⁵ proposed to segment bi-ventricle and myocardium using an ensemble of modified 2D and 3D U-Net. Wolterink *et al.*⁶ designed a deep neural network for automatic cardiac segmentation, as well as disease classification from the cardiac features. Baumgartner *et al.*⁷ explored various 2D and 3D convolution neural networks for the segmentation of the left (LV) and right (RV) ventricular cavities and the myocardium. Khened *et al.*⁸ employed a multi-scale residual DenseNet model to automatically segment the cardiac structure from cine MRI sequence. Although these methods were successful for cardiac segmentation, the use of deep model compression tasks for medical image segmentation is still rarely reported.

Although the first introduction of group convolution in AlexNet⁹ has been well illustrated its efficacy in recent network design, the pre-defined use of filters in each group convolution¹⁰ restricts its representation capability. Therefore, in this work, we propose to combine both learned-group convolution and weight-pruning technique with the segmentation problem in a fully convolutional setting. It is expected that *CondenseUNet* will significantly improve the performance of the network in segmenting different cardiac structures (LV, RV, and LV-Myocardium) from both end-diastole and end-systole images, while also significantly reducing the number of model parameters. The proposed network architecture is shown in Figure 1.

2. METHODOLOGY

2.1 Network Architecture

CondenseUNet is both a modification of DenseNet,¹¹ as well as a combination of CondenseNet¹² and U-Net.¹³ Our proposed *CondenseUNet* framework substitutes the concept of both standard convolution and group convolution (G-Conv) with learned group-convolution (LG-Conv). The standard *Group Convolution (GConv)*¹⁰

partitions the input features into disjoint groups. Given an input tensor of shape (I_i, H, W) , number of output channels O_o , and the partitioned group M , we apply GConv layers between input partition $(\frac{I_i}{M}, H, W)$ and $(\frac{O_o}{M}, \frac{I_i}{M}, \frac{K}{h}, \frac{K}{w})$ weight group; where $(\frac{K}{h}, \frac{K}{w})$ is the shape of the kernel, leading to the computational cost, $\mathcal{O}(\frac{I_i X O_o}{M})$. While the standard convolution needs an increased level of computation, i.e. $\mathcal{O}(I_i \times O_o)$, and concurrently, the pre-defined use of filters in each group convolution¹⁰ restricts its representation capability, these aforementioned problems are mitigated by introducing *LG-Conv* that learns group convolution dynamically during training through a multi-stage scheme, illustrated in the blue rectangular box in Figure 1. Before training, output filters are split into equal-sized M groups, where each group has its own weight. Thus, each group can select its own set of relevant input features, assisting the system to predict most relevant features at the relevant connections. This multi-stage pipeline consists of *multi-condensing* stages followed by the *optimization* stage.

In the first half of the pipeline, training is initiated by calculating the magnitude of the weights for each incoming feature, which are then averaged. After that, the low-magnitude weighted column is screened out from the features. Thus, a fraction of $(C - 1)/C$ is truncated after each of the $C - 1$ condensing stages.

The second part of the pipeline is where all training occurs. This stage is focused on finding the optimal permutation connection that will share a similar sparsity pattern, to mitigate any negative effects on accuracy induced by the pruning process. As mentioned by Huang *et al.*, in their paper on CondenseNet,¹² both L_1 and L_2 regularizations are efficient for solving the overfitting problem, but they do not perform well as far for network optimization. To restore performance, we introduce group-lasso, an efficient regularizer that is a natural generalization of the standard lasso (least absolute shrinkage and selection operator) objective. Additionally, the group-lasso regularizer encourages group-level sparsity at the factor level by forcing all outgoing connections from a single neuron (corresponding to a group) to be either simultaneously zero or not.

As *CondenseUNet* is based on both U-Net and DenseNet, it is comprised of a down-sampling and up-sampling path. The down-sampling path is similar to CondenseNet and the up-sampling path is comprised of transposed convolutions, condense blocks and skip-connections with a soft-max layer to generate the image mask. Concatenation in the skip-layer has been replaced by an element-wise addition operation to mitigate the problem of the feature-map explosion. We employ a number of layers per block as 2,3,4,5,4,3,2 with 32 initial feature maps, 3 max-pooling layers, a growth rate of $k = 16$, and condensation factor, $C = 4$ (Figure 1).

2.2 Cardiac MRI Data

For this study, we used the ACDC dataset[†], which is composed of short-axis cardiac cine-MR images acquired from 100 different patients divided into 5 evenly distributed subgroups according to their cardiac condition: normal- NOR, myocardial infarction- MINF, dilated cardiomyopathy- DCM, hypertrophic cardiomyopathy- HCM, and abnormal right ventricle- ARV, available as a part of the STACOM 2017 ACDC challenge.¹⁴

2.3 Data Pre-processing

To solve the class-imbalance problem in multi-slice cardiac MR images, a patch of size 128×128 was extracted around the LV center from a full-sized cardiac MR and slice-wise normalization of voxel intensities were performed. The training dataset was divided into 70% training data, 15% validation data, and 15% testing data with five non-overlapping folds for cross-validation. We heavily augment the ACDC dataset through both affine and elastic transformations, including several operations: (i) re-scaling: random zoom factor ranging $0.8 \sim 1.2$, (ii) translation: random shift ranging $-5 \sim 5mm$, (iii) rotation, and (iv) Gaussian noise addition.

2.4 Loss Function and Evaluation Metrics

We use a dual-loss function that incorporates both Cross-Entropy and Dice loss which is formulated in the equation 1 and also illustrated in our paper,¹⁵U-NetPlus:

$$\mathcal{L}_{Dual-Loss} = \alpha \cdot \mathcal{L}_{Entropy}(x, y_1 | \theta) + \beta \cdot (1 - \mathcal{L}_{Dice}(x, y_1 | \theta)) \quad (1)$$

[†]<https://www.creatis.insa-lyon.fr/Challenge/acdc/databases.html>

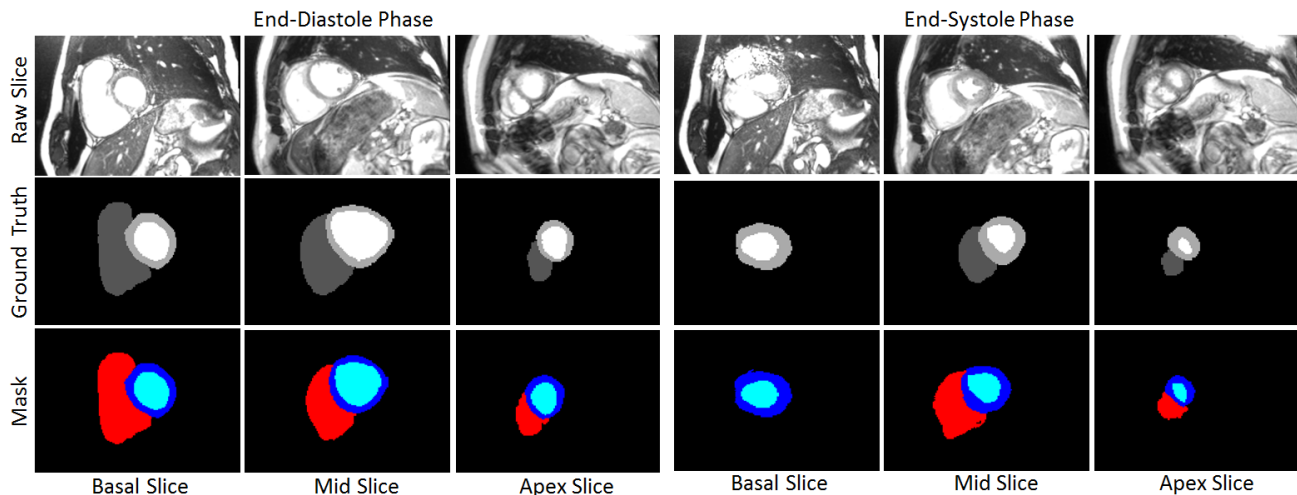


Figure 2: Segmentation results at both ED and ES phases from the base (high slice index) to apex (low slice index) showing RV in red, LV in cyan, and Myo in blue.

Table 1: Comparison of the segmentation results: Dice scores (%) on ACDC dataset for RV blood-pool, LV-myocardium, and LV blood-pool, obtained from all best networks against proposed model **CondenseUNet**: **CUNet**. UNet by [Jain *et al.*],¹⁸ MUNet: Modified 3D UNet by [Baumgartner *et al.*],⁷ MNet: Modified M-Net by [Jang *et al.*],¹⁹ and DNet: DenseNet by [Khened *et al.*]⁸

(a) Right Ventricle (RV) Blood-pool Segmentation Evaluation

	End Diastole (ED)						End Systole (ES)					
	UNet	DCN	MUNet	MNet	DNet	CUNet	UNet	DCN	MUNet	MNet	DNet	CUNet
Dice	91.0	92.0	93.20	92.90	93.50	93.46	83.0	84.0	88.30	88.50	87.90	87.60

(b) LV-Myocardium Segmentation Evaluation

	End Diastole (ED)						End Systole (ES)					
	UNet	DCN	MUNet	MNet	DNet	CUNet	UNet	DCN	MUNet	MNet	DNet	CUNet
Dice	86.0	86.0	89.2	87.0	88.9	89.43	88.0	88.0	90.1	88.8	89.8	90.1

(c) Left Ventricle (LV) Blood-pool Segmentation Evaluation

	End Diastole (ED)						End Systole (ES)					
	UNet	DCN	MUNet	MNet	DNet	CUNet	UNet	DCN	MUNet	MNet	DNet	CUNet
Dice	95.0	96.0	96.3	96.1	96.4	96.78	90.0	91.0	91.1	91.5	91.7	95.1

where $\mathcal{L}_{Entropy}$ is the cross-entropy loss and \mathcal{L}_{Dice} is the dice loss. The symbol, θ denotes the network weights, x denotes the training samples, y_1 be the label map of the training set, α and β denotes the weighting parameters. Additionally, we evaluate several clinical indices, including myocardial mass and ejection fraction.¹⁶ All assessment metrics were evaluated and compared between our proposed method and the ground truth.

2.5 Network Training and Testing

The Networks implemented in Tensorflow[‡] were initialized with He normal initializer¹⁷ and trained for 200 epochs with a batch size of 16. We used the Adam optimizer with a learning rate of 0.0001. The standard DenseNet has a growth rate of $k = 12$; in our case, k was 16. All experiments were conducted on a machine equipped with an NVIDIA RTX 2080 Ti GPU (11GB memory).

[‡]<https://www.tensorflow.org/>

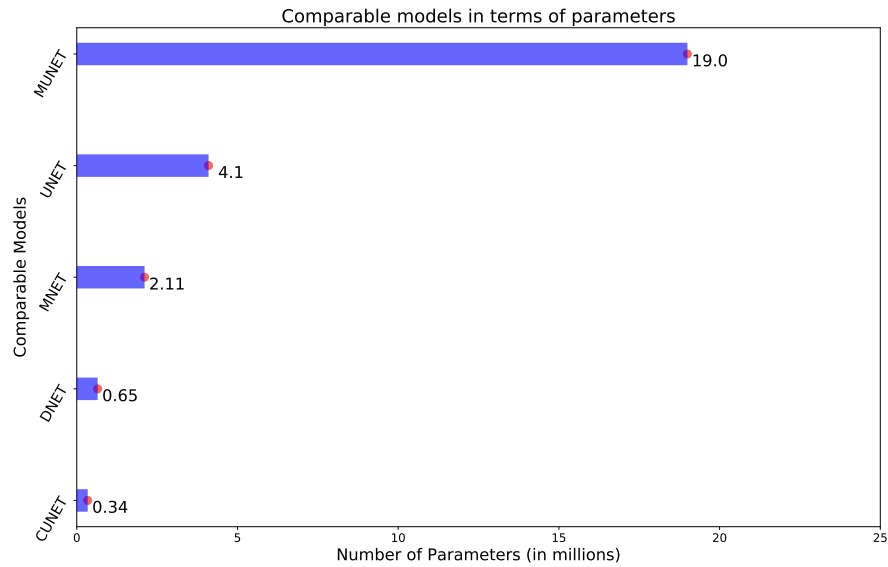


Figure 3: Comparison of *CondenseUNet* with other convolutional networks on the ACDC dataset in terms of number of model parameters ($\times 10^6$) – MUNet: Modified 3D UNet, MNet: Modified M-Net, DNet: DenseNet, CUNet: CondenseUNet.

3. RESULTS

In this section we evaluated the performance of our proposed *CondenseUNet* in terms of geometric scores, and number of parameters. Table 1 summarizes the Dice score, which measure the accuracy of the segmentation of the left and right ventricle blood-pool in diastole and systole and the left ventricle myocardium vis-a-vis their corresponding ground truth segmentation. The table also shows that the segmentation accuracy is generally better than the best comparable architecture for both LV and LV-myocardium in both ED and ES phase. An illustration of the qualitative segmentation results in both end-diastole (ED) and end-systole (ES) is provided in Figure 2. As shown, the segmentation mask for the ventricles and LV-myocardium performs best in both end-diastole and end-systole phase. Nevertheless, with the exception of several instance of the RV blood-pool segmentation, *CondenseUNet* yields the highest segmentation performance, on average, and significantly surpasses all other network architectures in terms of the number of model parameters.

Figure 3 plots the number of parameters (in millions) for all the comparable architectures. It shows that our proposed architecture (CUNet) runs on the ACDC dataset using half (50%) the memory requirement of DenseNet and one-twelfth ($\sim 8\%$) of the memory requirements of U-Net, while still achieving comparable segmentation performance.

4. CONCLUSION AND FUTURE WORK

In this paper, we propose a new paradigm for accurate LV, RV blood pool and myocardium segmentation from cine cardiac MR images by combining the memory-efficient CondenseNet architecture with the modified U-Net model. The capability of our network to learn the group structure allows multiple groups to re-use the same features via dense connectivity. Moreover, the integration of efficient weight pruning with a simple regularizer leads to high computational savings without compromising the accuracy of the segmentation and the fidelity of the estimated clinical parameters. The computationally efficient and accurate segmentation masks obtained from our proposed method can be used for pre-operative generation of subject-specific LV, RV and LV-myocardium models for surgical planning, navigation, and guidance applications.

Our proposed work reveals that a properly designed condensely connected network, when trained in the U-Net shaped framework, produces significantly higher performance with fewer trainable parameters. According to DICE metrics, we are achieving 96.78% DICE for LV blood pool in ED and 95.1% in ES phase, which showed at least 0.4% improvement in ED phase and 3.7% improvement in ES phase over the current methods, as well as more than 5% improvement over the standard U-Net architecture. We observed that the segmentation results for RV have not improved significantly beyond those of the LV or myocardium.

An alternative solution for better segmentation of the RV would be to perform architecture-agnostic efficient inference with cosine learning rate, as well as slice refinement. For future work, we will explore an additional attention analysis to further analyze the segmentation performance. This study will help us understand and visualize where our algorithm “looks” in an image by using a novel image saliency technique. Therefore, this computationally efficient segmentation model can be used for the generation of subject-specific models of the cardiac anatomy for computer-aided diagnosis or minimally invasive therapy planning.

ACKNOWLEDGMENTS

Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award No. R35GM128877 and by the Office of Advanced Cyber infrastructure of the National Science Foundation under Award No. 1808530.

REFERENCES

- [1] Benjamin, E. J., Muntner, P., and Bittencourt, M. S., “Heart disease and stroke statistics-2019 update: a report from the American Heart Association,” *Circulation* **139**(10), e56–e528 (2019).
- [2] Peng, P., Lekadir, K., Gooya, A., Shao, L., Petersen, S. E., and Frangi, A. F., “A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging,” *Magnetic Resonance Materials in Physics, Biology and Medicine* **29**(2), 155–195 (2016).
- [3] He, K., Gkioxari, G., Dollár, P., and Girshick, R., “Mask R-CNN,” in [*Proceedings of the IEEE International Conference on Computer Vision*], 2961–2969 (2017).
- [4] Jain, V., Bansal, P., Singh, A. K., and Srivastava, R., “Efficient single image super resolution using enhanced learned group convolutions,” in [*International Conference on Neural Information Processing*], 466–475, Springer (2018).
- [5] Isensee, F., Jaeger, P. F., Full, P. M., Wolf, I., Engelhardt, S., and Maier-Hein, K. H., “Automatic cardiac disease assessment on cine-MRI via time-series segmentation and domain specific features,” in [*International Workshop on Statistical Atlases and Computational Models of the Heart*], 120–129, Springer (2017).
- [6] Wolterink, J. M., Leiner, T., Viergever, M. A., and Išgum, I., “Automatic segmentation and disease classification using cardiac cine MR images,” in [*International Workshop on Statistical Atlases and Computational Models of the Heart*], 101–110, Springer (2017).
- [7] Baumgartner, C. F., Koch, L. M., Pollefeys, M., and Konukoglu, E., “An exploration of 2D and 3D deep learning techniques for cardiac MR image segmentation,” in [*International Workshop on Statistical Atlases and Computational Models of the Heart*], 111–119, Springer (2017).
- [8] Khened, M., Kollerathu, V. A., and Krishnamurthi, G., “Fully convolutional multi-scale residual DenseNets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers,” *Medical Image Analysis* **51**, 21–45 (2019).
- [9] Krizhevsky, A., Sutskever, I., and Hinton, G. E., “Imagenet classification with deep convolutional neural networks,” in [*Advances in Neural Information Processing Systems*], 1097–1105 (2012).
- [10] Xie, S., Girshick, R., Dollár, P., Tu, Z., and He, K., “Aggregated residual transformations for deep neural networks,” in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], 1492–1500 (2017).
- [11] Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q., “Densely connected convolutional networks,” in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], 4700–4708 (2017).

- [12] Huang, G., Liu, S., Van der Maaten, L., and Weinberger, K. Q., “CondenseNet: An efficient DenseNet using learned group convolutions,” in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], 2752–2761 (2018).
- [13] Ronneberger, O., Fischer, P., and Brox, T., “U-Net: Convolutional networks for biomedical image segmentation,” in [*International Conference on Medical Image Computing and Computer-Assisted Intervention*], 234–241, Springer (2015).
- [14] Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.-A., Cetin, I., Lekadir, K., Camara, O., Ballester, M. A. G., et al., “Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved?,” *IEEE Transactions on Medical Imaging* **37**(11), 2514–2525 (2018).
- [15] Kamrul Hasan, S. M. and Linte, C. A., “U-NetPlus: A modified encoder-decoder U-Net architecture for semantic and instance segmentation of surgical instruments from laparoscopic images,” in [*41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*], 7205–7211 (2019).
- [16] Waghlikar, K. B., Fischer, C. M., Goodson, A., Herrick, C. D., Rees, M., Toscano, E., MacRae, C. A., Scirica, B. M., Desai, A. S., and Murphy, S. N., “Extraction of ejection fraction from echocardiography notes for constructing a cohort of patients having heart failure with reduced ejection fraction (hfrf),” *Journal of Medical Systems* **42**(11), 209 (2018).
- [17] He, K., Zhang, X., Ren, S., and Sun, J., “Delving deep into rectifiers: surpassing human-level performance on ImageNet classification,” in [*Proceedings of the IEEE International Conference on Computer Vision*], 1026–1034 (2015).
- [18] Patravali, J., Jain, S., and Chilamkurthy, S., “2D-3D fully convolutional neural networks for cardiac MR segmentation,” in [*International Workshop on Statistical Atlases and Computational Models of the Heart*], 130–139, Springer (2017).
- [19] Jang, Y., Hong, Y., Ha, S., Kim, S., and Chang, H.-J., “Automatic segmentation of LV and RV in cardiac MRI,” in [*International Workshop on Statistical Atlases and Computational Models of the Heart*], 161–169, Springer (2017).