NATA-Net: A Network with Adaptive Task-Aware Heads for Universal Cardiac MRI Segmentation

Minh Nhat Trinh[1] and Teresa M Correia[1,2]

[1] Quantitative Bio-imaging Lab, Centro de Ciências do Mar (CCMAR/CIMAR-LA), Campus de Gambelas, Universidade do Algarve, Faro, Portugal
[2] School of Biomedical Engineering and Imaging Sciences, King's College London, London, United Kingdom
tmcorreia@ualg.pt

Abstract. Accurate cardiac magnetic resonance imaging (CMR) segmentation is essential for quantifying clinical biomarkers such as ejection fraction, myocardial mass, and scar burden. However, universal CMR segmentation remains challenging due to variations across pulse sequences, anatomical views, scanners, institutions, and label configurations. To address these challenges, we propose NATA-Net, a lightweight multi-head neural network for multi-sequence and multi-view CMR segmentation. The model uses a shared fully convolutional dense encoder–decoder backbone to learn common cardiac representations, while task-specific prediction heads adapt the output space to different modalities and views. This unified design avoids training separate models for each task and allows all available datasets to be jointly exploited, which is beneficial under limited annotation settings. With only 2.6M parameters and 67 layers, NATA-Net provides an efficient balance between segmentation accuracy, memory usage, and inference speed.
On the Universal Multi-Sequence, Multi-Center and Multi-View CMR Segmentation Challenge validation set, NATA-Net achieved good-quality segmentations across multiple tasks. For Cine, it achieved an average Dice Score (DSC) of 0.91, Hausdorff distance (HD) of 8.67 mm, and Average Surface Distance (ASD) of 0.47 mm. For LGE segmentation, it achieved an average DSC of 0.75 across all left ventricle short-axis and long-axis views and right atrium views. For clinical metric estimation, NATA-Net demonstrated high predictive accuracy, achieving a Pearson Correlation Coefficient (PCC) of 0.89 for Cine ejection fraction prediction and a volumetric PCC of 0.70 for LGE scar mass estimation. Overall, NATA-NET achieved an average challenge score of **0.69**.

Keywords: Cardiac MRI segmentation · Multi-head network · Lightweight segmentation model

1 Introduction

Cardiac magnetic resonance imaging (CMR) is the gold standard for assessing cardiac function and myocardial viability [1]. Accurate segmentation of cardiac

structures is essential for estimating clinical biomarkers such as ejection fraction, myocardial mass, and scar extent. However, manual annotation is time-consuming, observer-dependent, and difficult to scale to large multi-center datasets. These limitations motivate the development of robust automated CMR segmentation methods [2].

Despite recent advances, universal CMR segmentation remains challenging because CMR data are highly heterogeneous. Cine and Late Gadolinium Enhancement (LGE) imaging provide different clinical information and exhibit different intensity distributions and contrast patterns. LGE, in particular, has become increasingly important for atrial cavity and scar analysis, leading to several dedicated datasets and benchmark studies for left and right atrial segmentation [3,4,5]. Moreover, CMR is acquired from multiple views, such as short-axis (SAX), two-chamber (2CH), and four-chamber (4CH) views, where visible structures and label sets may differ. Multi-center acquisitions further introduce variations in scanner vendor, protocol, resolution, and patient populations, making it difficult to train a single model that generalizes across sequences, views, and segmentation targets. These challenges are further emphasized by the recent Universal Multi-Sequence, Multi-Center and Multi-View CMR Segmentation Challenge [6].

Existing segmentation methods, including U-Net variants [7], nnU-Net [8], and Transformer-based models [9], have achieved strong performance in fully supervised settings. However, most methods assume a fixed modality, view, and label space. Semi-supervised and partial-label segmentation methods address data scarcity using pseudo-labeling, consistency learning, teacher–student training, or label-set-aware objectives. Nevertheless, extending these approaches to multi-sequence and multi-view CMR remains non-trivial due to significant domain shifts and inconsistent anatomical label definitions across acquisitions. Thus, a unified homogeneous decoder may suffer from label-space conflicts, whereas training separate models for each task increases computational cost and limits knowledge sharing.

To address these issues, we propose NATA-Net, a lightweight Network with Adaptive Task-Aware heads for universal segmentation of multi-sequence and multi-view CMR data. Inspired by Fully Convolutional DenseNets [10], NATA-Net exploits dense connections for efficient feature reuse and gradient propagation. A shared dense encoder-decoder backbone learns common cardiac representations for universal segmentation of multi-sequence and multi-view CMR data, while task-specific prediction heads adapt the output space to each sequence-view setting. This allows the model to handle heterogeneous CMR acquisitions without training a separate network for every task. NATA-Net is designed to balance segmentation performance and computational efficiency. Instead of training multiple large models independently, it can be trained within a unified framework and activates only the corresponding lightweight head for each input. With 2.6M parameters and 67 layers, the model remains compact while preserving view- and task-specific specialization.

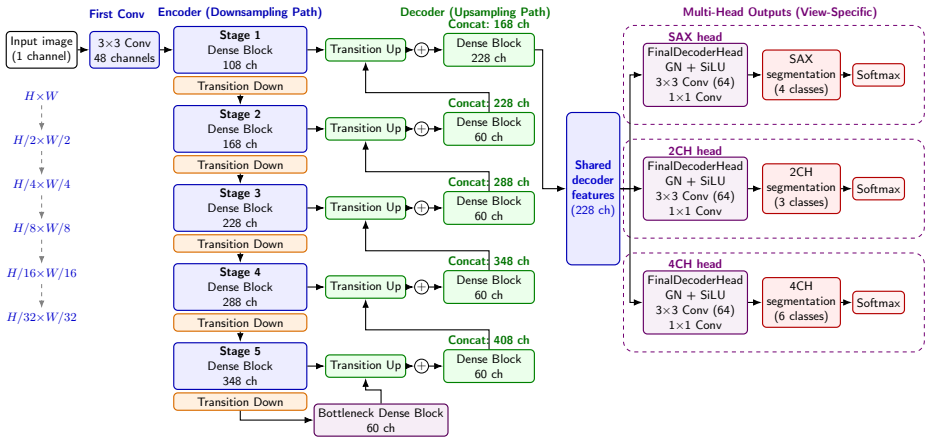

Fig. 1. NATA-NET Architecture: Shared encoder–decoder for universal multi-sequence, multi-view CMR segmentation with task-specific prediction heads.

Our contributions are threefold. First, we propose a lightweight multi-head dense segmentation network for universal multi-sequence and multi-view CMR segmentation. Second, we introduce a shared-backbone and task-specific-head strategy for heterogeneous datasets with different sequences, views, and label spaces. Finally, we evaluate NATA-Net on the Universal Multi-Sequence, Multi-Center and Multi-View CMR Segmentation Challenge [6], targeting both anatomical segmentation and clinically relevant biomarker estimation.

2 Method

Figure 1 illustrates the proposed NATA-Net, a Network with Adaptive Task-Aware Heads for multi-view CINE/LGE segmentation. The network follows an encoder–decoder structure inspired by the Fully Convolutional DenseNets architecture [10], where dense connections are used to promote feature reuse and stable gradient propagation. Given a single-channel CMR slice, an initial 3×3 convolution maps the input to 48 feature channels. Then, the features are processed by a shared dense encoder, a bottleneck block, and a shared dense decoder.

Each dense layer consists of Group Normalization [11], SiLU activation [12], a 3×3 convolution block, and dropout. The growth rate is set to $k = 12$, and the output of each dense layer is concatenated with the previous feature maps within the same dense block. The encoder contains five dense blocks, each followed by a transition-down module with a 1×1 convolution and 2×2 max pooling. The decoder restores spatial resolution using transposed convolutions and skip connections from the encoder, allowing the model to recover fine anatomical features.

To handle different segmentation tasks, NATA-Net adopts a task-dependent multi-head design. The encoder, bottleneck, and decoder are shared across all

inputs to learn common cardiac representations, while the final prediction heads are defined according to the sequence, view, and label set of each task. For example, for the Cine CMR task, the SAX, 2CH, and 4CH heads predict 4, 3, and 6 classes, respectively. For the LGE task, the SAX, 2CH, 4CH, and right atrium (RAS) heads predict 5, 4, 5, and 2 classes, respectively. Each head consists of Group Normalization, SiLU activation, a 3×3 convolution reducing the feature dimension to 64 channels, dropout, and a final 1×1 convolution that maps the features to the required number of classes. During training and inference, the appropriate head is selected based on the input, and softmax is applied along the channel dimension to obtain the final probability map.

The proposed design also aims to improve inference efficiency and reduce computational and memory requirements. Instead of training separate models for different CMR sequences and views, NATA-Net uses one shared backbone with lightweight task-specific prediction heads. This reduces redundant computation and memory usage while maintaining task specialization. The resulting model contains only 2.6M parameters with 67 layers, making it compact and computationally efficient for fast inference in challenging settings with multiple CMR sequences and views.

3 Experiments

3.1 Dataset and Evaluation Measures

The network was trained and tested on the Universal Multi-Sequence, Multi-Center and Multi-View CMR Segmentation Challenge dataset comprising 750 views, acquired using 3.0T CMR scanners with 32-channel coils from three different vendors (Philips Healthcare, Siemens Healthineers and GE Healthcare) [6]. Data consists of paired CMR images and pixel-level segmentation annotations from two core clinical sequences: Cine and LGE imaging. Cine is used for cardiac functional assessment, while LGE is used for myocardial scar evaluation. The dataset is organized into two tasks. Task 1, `CINE_MULTI`, contains 315 multi-slice Cine acquisitions from three views: 105 SAX, 105 2CH, and 105 4CH volumes. These correspond to 47,758 SAX slices, 12,285 2CH slices, and 11,870 4CH slices along the z-axis. Annotated structures include the left ventricular (LV) cavity, LV myocardium, right ventricular (RV) cavity with both atria additionally annotated in the 4CH view.

Task 2, `LGE_MULTI`, contains 200 LGE volumes from four views: 40 SAX, 40 2CH, 40 4CH, and 80 RAS volumes. These correspond to 670 SAX slices, 201 2CH slices, 201 4CH slices, and 1,111 RAS slices along the z-axis. The annotations include the LV, RV, myocardium, myocardial scar, and right atrium, depending on the view. Each task also provides metadata files containing image–label path mappings and clinical information, including left ventricle ejection fraction (LVEF) for Cine SAX and left ventricular myocardial scar mass for LGE SAX.

Segmentation performance is evaluated using the Dice Similarity Coefficient (DSC), computed for each structure and averaged across structures, views, and

patients. For Task 1, boundary accuracy is further assessed using Hausdorff Distance (HD) and Average Symmetric Surface Distance (ASD), both measured in millimeters. In addition to segmentation metrics, clinical metric estimation is evaluated. For Cine MRI, the predicted LVEF is compared with the reference value using the Pearson Correlation Coefficient (PCC). For LGE, scar mass estimation is assessed using volumetric Pearson Correlation Coefficient (vPCC) and Relative Absolute Error (RAE), where RAE is computed over patients with non-zero ground-truth scar mass.

3.2 Preprocessing

All CMR data were converted into a unified 2D slice-based format, and the model was trained on individual slices rather than full 3D volumes. To reduce intensity variations across sequences, centers, and acquisition protocols, intensity normalization was performed before resizing. For Cine MRI, images were clipped at the 0.05th and 99.5th percentiles and normalized to $[0, 1]$ using min–max scaling:

$$\hat{I} = \frac{\text{clip}(I, p_{\text{low}}, p_{\text{high}}) - p_{\text{low}}}{p_{\text{high}} - p_{\text{low}}},$$

where p_{low} and p_{high} denote the lower and upper percentile values.

For LGE MRI, magnitude images were clipped at the 0.01th and 99.9th percentiles and normalized to $[0, 1]$ to preserve scar-related regions while suppressing outliers. Images were then cropped around the foreground (when necessary) and resized to 256×256, with the same transformation applied to the corresponding masks.

3.3 Implementation details

All experiments were conducted on an NVIDIA GeForce RTX 4070 SUPER GPU with 12 GB memory. The network was trained using mixed-precision computation with FP16/FP32 to accelerate training while maintaining numerical stability. Each task was trained using five-fold cross-validation with ratio of 80:20. NATA-Net was optimized with the NAdam optimizer [13] using an initial learning rate of 10^{-3} for approximately 200 epochs. The Focal Active Contour loss [14] was used as the training objective to address class imbalance through focal weighting and improve boundary precision through active contour. This loss encourages both region-level consistency and boundary alignment, which is important for cardiac structures with thin or irregular shapes, such as the myocardium and scar regions. It also places greater emphasis on misclassified or low-confidence pixels, thereby helping alleviate class imbalance problem.

For Task 2, the number of training slices along the z-axis was relatively limited. Therefore, we initialized the Task 2 model using the trained weights from Task 1 as transfer learning. This initialization helped the model converge faster and provided a stronger cardiac feature representation for LGE segmentation. The code is available on https://github.com/QBioImaging/CMRSEG_Challange.

3.4 Post-processing

During inference, we ensembled the five cross-validation models by averaging their softmax probability maps and assigning each pixel location in the reconstructed multi-slice stack to the class with the highest averaged probability. A lightweight 3D morphological post-processing step was then applied to each foreground class. Small connected components were removed, small holes were filled, and 3D morphological closing with a spherical structuring element was used to smooth boundaries. The processed class masks were finally merged into a single multi-class volume, reducing noisy predictions while preserving anatomical consistency.

4 Results and Discussion

The performance of our proposed model was evaluated on the CMR Segmentation Challenge validation set, which served as our test set, through the Codabench [15] platform using organizer-provided ground-truth masks. Figure 2A shows the DSC distribution across different views and anatomical labels for both Cine and LGE segmentation. Overall, NATA-Net achieved high and stable DSC scores for major cardiac structures, although DSC can be less reliable

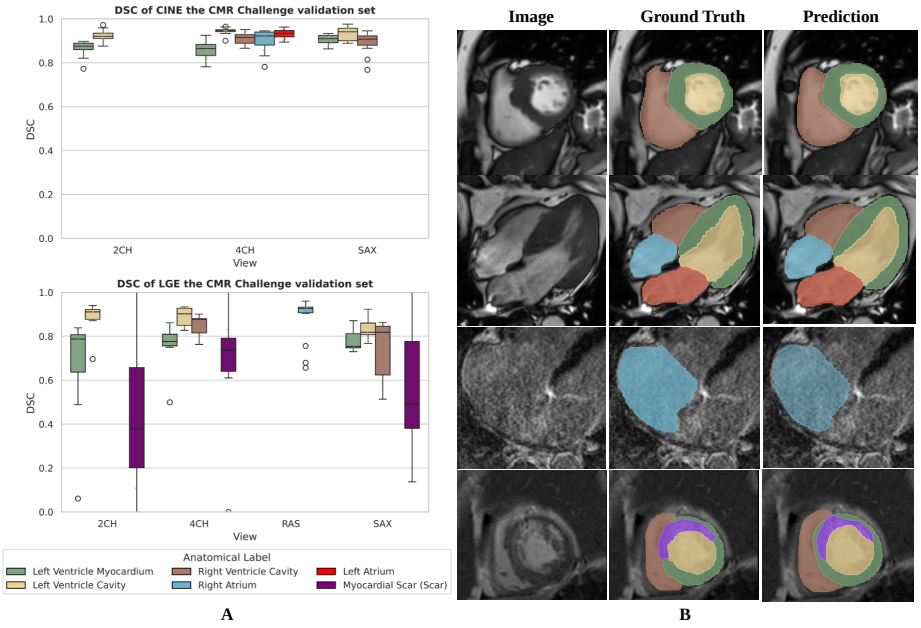

Fig. 2. Segmentation results on the CMR Segmentation Challenge validation set. A) Distribution of DSC scores for Cine and LGE segmentation across different views and anatomical labels. B) Representative qualitative results. From left to right: input Cine/LGE image, ground-truth segmentation, and NATA-Net prediction.

for very small regions, such as small myocardial scars, due to its sensitivity to minor segmentation errors. Figure 2B shows representative qualitative results. The predicted segmentations are visually consistent with the ground truth across different CMR sequences and views, demonstrating that NATA-Net can accurately delineate both large cardiac structures, such as the LV/RV blood pools and atria, and more challenging structures, such as myocardial scar.

Table 1. Segmentation results on the CMR Segmentation Challenge validation set. Values are reported as mean ± standard deviation. Higher DSC indicates better performance, while lower HD and ASD indicate better boundary accuracy.

Task	View	DSC ↑	HD (mm) ↓	ASD (mm) ↓	Samples
Task 1: Cine	SAX	0.91 ± 0.04	10.14 ± 4.26	0.62 ± 0.33	45
	2CH	0.89 ± 0.04	7.42 ± 1.91	0.42 ± 0.11	30
	4CH	0.91 ± 0.04	8.29 ± 5.40	0.39 ± 0.23	75
	Average	0.91 ± 0.04	8.67 ± 4.65	0.47 ± 0.27	150
Task 2: LGE	SAX	0.73 ± 0.20	31.06 ± 34.31	3.34 ± 3.47	28
	2CH	0.66 ± 0.31	20.45 ± 18.84	3.60 ± 7.46	21
	4CH	0.79 ± 0.19	17.31 ± 22.13	0.83 ± 0.96	28
	RAS	0.88 ± 0.10	20.30 ± 12.35	1.46 ± 1.73	14
	Average	0.75 ± 0.23	22.81 ± 25.26	2.34 ± 4.24	91

Table 1 reports the quantitative segmentation performance on the challenge validation set. For Task 1, NATA-Net achieved an average DSC of 0.91, HD of 8.67 mm, and ASD of 0.47 mm across all Cine views. The model obtained consistent DSC scores across SAX, 2CH, and 4CH views. For Task 2, the average DSC was 0.75 across all LGE views, with the highest performance observed on the RAS view.

Table 2. Clinical metric estimation results on the CMR Segmentation Challenge validation set. Higher PCC/vPCC and lower RAE indicate better performance.

Task	Metric	PCC/vPCC ↑	RAE ↓	Samples
Task 1: Cine	LVEF	0.89	0.11	15
Task 2: LGE	Scar Mass	0.70	1.35	7

Table 2 summarizes the clinical metric estimation results. For Cine, the predicted ejection fraction achieved a PCC of 0.89 with the reference values. For LGE, scar mass estimation obtained a vPCC of 0.70. In addition to segmentation accuracy, NATA-Net is computationally efficient. The model contains only 2.6M parameters and requires approximately 4 GB of GPU memory during inference, with an average inference time of about 3 seconds per patient in our validation setting. This low memory footprint and fast runtime make the model practical

for low-end GPU environments and suitable for research or resource-constrained deployment scenarios.

Although NATA-Net achieved promising validation results, several limitations remain. The current framework uses 2D slice-based segmentation, which is memory-efficient but does not fully exploit inter-slice anatomical continuity. This may limit performance for structures with weak boundaries or few annotated slices, especially in LGE. Compared with the Cine dataset, the LGE dataset is much smaller, with 200 files and only 2,183 slices. In particular, the LGE 2CH and 4CH views contain only 201 slices each, making scar segmentation and scar mass estimation more challenging. In addition, the multi-head design relies on predefined task-specific heads and does not explicitly model relationships between different views. Nevertheless, the predicted scar burden showed good agreement with the ground truth, with no significant difference observed between predicted and reference values ($p = 1.00$, Wilcoxon signed-rank test).

Future work will explore lightweight 2.5D or 3D extensions to incorporate volumetric context while preserving computational efficiency. We will also investigate cross-view feature fusion, uncertainty-aware post-processing, and more advanced transfer learning or semi-supervised strategies to improve LGE scar segmentation under limited annotation settings.

5 Conclusion

We presented NATA-Net, a lightweight network for universal multi-sequence and multi-view CMR segmentation. The model combines a shared dense encoder–decoder backbone with task-specific prediction heads, allowing it to learn common cardiac representations while adapting to different Cine and LGE MRI views and label configurations.

On the Universal Multi-Sequence, Multi-Center and Multi-View CMR Segmentation Challenge validation set, NATA-Net achieved competitive segmentation and clinical metric estimation performance. These results suggest that the proposed framework provides a practical balance between accuracy, memory efficiency, and inference speed, making it a compact and effective solution for multi-view CMR segmentation and clinically relevant biomarker estimation.

Acknowledgements This study received funding from the FCT–Portuguese Foundation for Science and Technology through grants UID/04326/2025, UID/PRR/04326/2025, LA/P/0101/2020 (https://doi.org/10.54499/LA/P/0101/2020), 2025.02647.BDANA (https://doi.org/10.54499/2025.02647.BDANA), 'la Caixa' Foundation and FCT, I.P. under project LCF/PR/HR22/00533, NVIDIA GPU Hardware Grant, and EU Horizon Europe project IMAGINE, GA No. 101094250 (https://doi.org/10.3030/101094250).

Disclosure of Interests

The authors declare no competing interests.

References

1. Campello, V.M., Gkontra, P., Izquierdo, C., Martin-Isla, C., Sojoudi, A., Full, P.M., Maier-Hein, K., Zhang, Y., He, Z., Ma, J., et al.: Multi-centre, multi-vendor and multi-disease cardiac segmentation: the m&ms challenge. IEEE Transactions on Medical Imaging **40**(12), 3543–3554 (2021). https://doi.org/10.1109/TMI.2021.3090082 1
2. Bai, W., Sinclair, M., Tarroni, G., Oktay, O., Rajchl, M., Vaillant, G., Lee, A.M., Aung, N., Lukaschuk, E., Sanghvi, M.M., et al.: Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. Journal of cardiovascular magnetic resonance **20**(1), 65 (2018). https://doi.org/https://doi.org/10.1186/s12968-018-0471-x 2
3. Bai, J., Qiu, R., Chen, J., Wang, L., Li, L., Tian, Y., Wang, H., Lu, Y., Zhao, J.: A two-stage method with a shared 3d u-net for left atrial segmentation of late gadolinium-enhanced mri images. Cardiovascular Innovations and Applications **8**(1), 976 (2023). https://doi.org/10.15212/CVIA.2023.0039 2
4. Zhu, J., Bai, J., Zhou, Z., Liang, Y., Chen, Z., Chen, X., Zhang, X.: Ras dataset: a 3d cardiac lge-mri dataset for segmentation of right atrial cavity. Scientific Data **11**(1), 401 (2024). https://doi.org/https://doi.org/10.1038/s41597-024-03253-9 2
5. Bai, J., Zhu, J., Chen, Z., Yang, Z., Lu, Y., Li, L., Li, Q., Wang, W., Zhang, H., Wang, K., et al.: A benchmark framework for the right atrium cavity segmentation from lge-mris. IEEE Transactions on Medical Imaging (2025). https://doi.org/https://doi.org/10.1109/TMI.2025.3590694 2
6. Qu, T., Su, Z., Zhang, H., Bai, J., Zhang, N., Wang, H., Zhou, Z., Zhao, P., Bo, K., Zhao, J., Gan, J., Zhan, Y., Lu, H., Zhang, X., Cai, W., Xun, L., Zhang, H.: Universal multi-sequence, multi-center and multi- view cmr segmentation challenge (Apr 2026). https://doi.org/10.5281/zenodo.19728181 2, 3, 4
7. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241. Springer International Publishing, Cham (2015). https://doi.org/https://doi.org/10.1007/978-3-319-24574-4_28 2
8. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021). https://doi.org/https://doi.org/10.1038/s41592-020-01008-z 2
9. Fan, C., Su, Q., Xiao, Z., Su, H., Hou, A., Luan, B.: ViT-FRD: A vision transformer model for cardiac MRI image segmentation based on feature recombination distillation. IEEE Access **11**, 129763–129772 (2023). https://doi.org/10.1109/access.2023.3302522 2
10. Jégou, S., Drozdzal, M., Vazquez, D., Romero, A., Bengio, Y.: The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 1175–1183 (2017). https://doi.org/10.1109/CVPRW.2017.156 2, 3
11. Wu, Y., He, K.: Group normalization. In: Proceedings of the European conference on computer vision (ECCV). pp. 3–19 (2018). https://doi.org/10.1007/978-3-030-01261-8_1 3

12. Elfwing, S., Uchibe, E., Doya, K.: Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. Neural networks **107**, 3–11 (2018). https://doi.org/https://doi.org/10.1016/j.neunet.2017.12.012 3
13. Dozat, T.: Incorporating nesterov momentum into adam (2016), https://openreview.net/forum?id=OM0jvwB8jIp57ZJjtNEZ 5
14. Trinh, M.N., Nham, D.H.N., Pham, V.T., Tran, T.T.: An attention-pidi-unet and focal active contour loss for biomedical image segmentation. In: 2022 RIVF International Conference on Computing and Communication Technologies (RIVF). pp. 635–640 (2022). https://doi.org/10.1109/RIVF55975.2022.10013852 5
15. Xu, Z., Escalera, S., Pavão, A., Richard, M., Tu, W.W., Yao, Q., Zhao, H., Guyon, I.: Codabench: Flexible, easy-to-use, and reproducible meta-benchmark platform. Patterns **3**(7), 100543 (2022). https://doi.org/10.1016/j.patter.2022.100543 6