

Long axis cardiac MRI segmentation using anatomically-guided UNets and transfer learning

Andre Von Zuben¹[0000–0003–4433–2608], Emily Whitt²[0000–0002–6172–485X],
Felipe A. C. Viana¹[0000–0002–2196–7603], and Luigi E.
Perotti¹[0000–0002–9010–2144]

¹ Department of Mechanical and Aerospace Engineering, University of Central
Florida, Orlando, FL, USA avzuben@knights.ucf.edu

² Burnett School of Biomedical Sciences, University of Central Florida, Orlando, FL,
USA

Abstract. In this work we present a machine learning model to segment long axis magnetic resonance images of the left ventricle (LV) and address the challenges encountered when, in doing so, a small training dataset is used. Our approach is based on a heart locator and an anatomically guided UNet model in which the UNet is followed by a B-Spline head to condition training. The model is developed using transfer learning, which enabled the training and testing of the proposed strategy from a small swine dataset. The segmented LV cavity and myocardium in the long axis view show good agreement with ground truth segmentations at different cardiac phases based on the Dice similarity coefficient. In addition the model provides a measure of segmentations’ uncertainty, which can then be incorporated while developing LV computational models and indices of cardiac performance based on the segmented images. Finally, several challenges related to long axis, as opposed to short axis, image segmentation are highlighted, including proposed solutions.

Keywords: Cardiac image segmentation · Long axis MRI · Machine learning · Anatomically-guided UNet.

1 Introduction

Subject specific computational cardiology has the potential to significantly improve diagnosis, prognosis, and therapy planning for patients affected by cardiac diseases. One of the current challenges in deploying computational models to the clinic consists in quickly generating accurate anatomical models from imaging data. In recent years, machine learning (ML) has made it possible to automatically process imaging data. For example, in the context of cardiac magnetic resonance imaging (cMRI), ML models have been extensively used to segment cine MRI data to compute indices of cardiac function.

Currently, most of the ML effort has been directed toward segmenting the left ventricle (LV) myocardium (LVM) and cavity (LVC) from slices acquired in a short axis view. However, there could be significant benefits from building LV

models based on images acquired in the long axis view. For example, in contrast to short axis based models, the base and the apex of the LV are clearly identified in the long axis views. Furthermore, fewer long axis versus short axis slices may be needed to build a full LV model, potentially shortening the acquisition time.

In this work we propose a ML model to segment the LVM and LVC from long axis (LA) images. To enable the development of this new model based on a significantly smaller and swine dataset, we apply transfer learning [9] to our previous model [13], which was developed to segment short axis (SA) images and applied to the ACDC [1] human dataset. Therefore, the current study applies transfer learning across species (human to swine) and between different image views (short axis to long axis). In the following, after describing the key features of our algorithm, we present representative results and the new challenges encountered when segmenting long axis images.

2 Methods

2.1 Long axis cine MRI dataset

We tested the proposed ML segmentation model using MRI data acquired in nine (N=9) healthy swine subjects. All animal experiments were approved by the UCLA Institutional Animal Care and Use Committee (ARC protocol # 2015-124). Subjects were imaged using a 3T MRI scanner (Prisma, Siemens) and, among other data, balanced steady state free precession (bSSFP) 2D cine MRI data was acquired in short and long axis views. Here, we focus on images acquired along 6 long axis planes approximately 30° apart. The in plane spatial image resolution is $1.18 \times 1.18 \text{ mm}^2$. For each subject and long axis location, three cardiac phases (the beginning, mid, and end of systole) are selected and segmented. In total, our dataset consists of 162 unique images. In each segmentation, the LVM and LVC are identified. Data augmentation via elastic deformation is included similarly to our previous work [12]. Via data augmentation, we generated ten variations for each image, leading to a total of 1782 images including the original images.

2.2 Deep Learning Segmentation Strategy

Model Architecture. The proposed pipeline illustrated in Fig. 1 is composed of three neural networks designed to perform the following tasks:

1. Localization of the LVC: the LVC-Locator consists in a traditional UNet.
2. Anatomically-guided feature extraction: two anatomically-guided deep neural networks (UNet_{AG}) generate the contours of the LVC and LVM, i.e., the LV endocardial and epicardial walls.

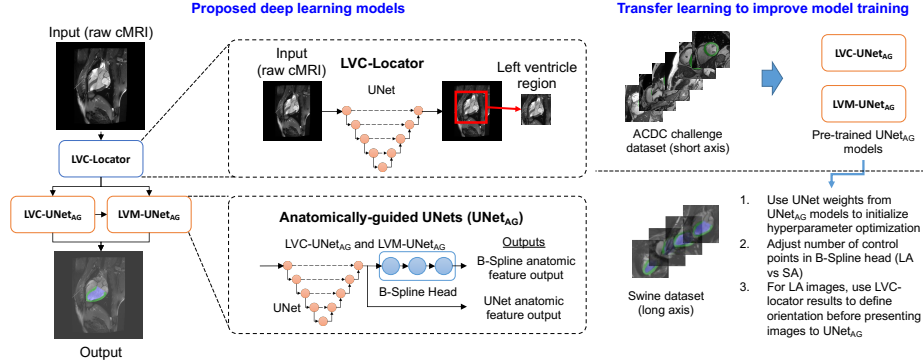


Fig. 1. Proposed pipeline of deep learning models for LV long axis segmentation.

The *LVC locator network* is a UNet [11, 6] that receives a 352×352 pixels image as input and outputs each pixel’s probability to be part of the heart. In training, a combination of the Dice similarity coefficient (DSC) loss and the binary cross-entropy loss is used as the loss function. The process to locate the region of the image containing the LV includes identifying the pixels with the highest and lowest horizontal and vertical locations, calculating their midpoint, and extracting a 144×144 -pixel region centered at that midpoint. Before extracting the region of interest, the image is rotated using the singular value decomposition of the UNet output to align the LV long axis along the vertical or horizontal directions. This step simplifies the implementation of the B-Spline layers in the UNet_{AG}.

The *anatomically-guided UNet (UNet_{AG}) model* is a specialized implementation of the UNet proposed in [13] in which the UNet is followed by a B-Spline head. In our UNet implementation, the encoder contains four convolutional blocks (2D convolutional layers and batch normalization) and the decoder contains the four respective transposed convolutional blocks. The B-Spline head is composed of three layers: a contour detection, a B-Spline, and a perimeter-filling layer. As further detailed in [13], the B-Spline head conditions the training of the UNet by returning a smooth and constrained region with high probability of containing the region of interest. The LVM-UNet_{AG} receives the LVC-UNet_{AG} output as an additional channel, enhancing the overall left ventricle contours generated by the pipeline and leading to a robust wall thickness estimation. These characteristics, combined with the smooth and continuous nature of the B-Spline head (i.e., the B-Spline penalizes non-physiological protrusions, jaggedness, and discontinuities), provide lower segmentation uncertainty at the endocardial and epicardial walls. Furthermore, to increase robustness during training and prediction, four copies of the same image are passed to the UNet_{AG} (each copy corresponds to the original image rotated by 0° , 90° , 180° , and 270°): the probabilities computed for each image are then averaged to compute the final segmentation.

Transfer Learning (TL). TL is adopted to implement the UNet_{AG} such that:

- The UNet portion has the same architecture presented in [13] and the weights generated during training with the ACDC dataset [1] will be used as initial guess for the training with the swine dataset. This approach will lead to faster convergence during training and robustness with respect to the reduced dataset.
- The B-Spline head for the long axis will have 45 control points, as opposed to 20 found in [13]. The larger number of control points will allow the UNet_{AG} to conform to the LA geometry, which departs from the circular-like SA geometry.

The choice of 45 B-Spline control points is motivated to promote smoothness of the segmentation while enabling flexibility to match the long axis LV geometry. We conducted studies using 20, 45, 90, 180, and 360 control points and observed only a moderate increase in DSC, which reaches a plateau at ≈ 45 control points. Additionally, a lower number of control points limits memory usage.

Ablation study. Given the size of the swine dataset, a 3-fold cross-validation strategy was used to assess the quality of the resulting models. The data was split into 3 groups, each containing the images for 3 subjects: two groups are used for training and the remaining one for validation. Using this approach, we perform an ablation study by varying the initial weights used in training and the utilization of the B-Spline head:

- *Case #1:* UNet_{AG} that uses the UNet transferred from [13] to initialize the training and B-Spline head with 45 control points.
- *Case #2:* UNet_{AG} that uses the UNet transferred from [13] to initialize the training but **without** the B-Spline head.
- *Case #3:* UNet_{AG} with UNet trained from initially randomized weights (Glorot Uniform initialization function) and B-Spline head with 45 control points.
- *Case #4:* UNet_{AG} with UNet trained from initially randomized weights (Glorot Uniform initialization function) **without** the B-Spline head.

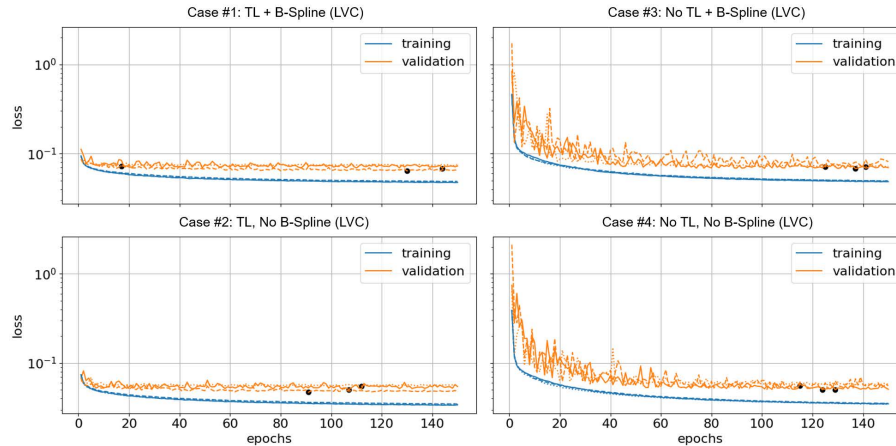
3 Results

First, we summarize in Tab. 1 the model size and computational cost associated with training each network of the pipeline illustrated in Fig. 1. In our study, both the model and the data fit well in the GPU memory. The LVC-Locator is a straightforward segmentation model with low accuracy requirement (other than the LVC rough segmentation and the localization of the region of interest); therefore, we do not report further results on it.

Figure 2 shows the convergence of the loss function throughout training of the LVC-UNet_{AG} for each one of the cases detailed in the “Ablation study” subsection. Similar results were observed for the LVM-UNet_{AG} (data not shown).

Table 1. Trainable parameters and training computational cost. All networks were trained on a GPU NVIDIA Tesla P100 with 16 GB of memory.

Model	Trainable parameters	Time per epoch [s]
LVC-Locator	17,660,694	66
LVC-UNet _{AG}	17,660,694	72
LVM-UNet _{AG}	17,661,126	114

**Fig. 2.** Loss vs. epochs for LVC-UNet_{AG} models during training and validation. Black circular markers correspond to the lowest validation loss within 150 epochs.

Next, in Tab. 2, we report the cross-validation performance of the model in terms of DSC for segmentations obtained at the beginning, mid, and end of systole for the swine dataset. The obtained DSC values are comparable with, although slightly lower than, the results reported for SA image segmentation computed from the large ACDC human dataset [1] in our previous work [13].

Figure 3 illustrates the image segmentations obtained with the proposed UNet_{AG} transferred from [13] and using the B-Spline head (case #1). The right and left panels present results for end diastole (ED), mid systole (MS), and end systole (ES) for two subjects in the validation fold. These cases represent different levels of complexity due to contrast between the myocardium and the background, the presence of image artifacts, and different views of the ventricles, including the prominence of the right ventricle.

We conclude by illustrating two challenges specific to LA images' segmentation (Fig. 4). First, the LV myocardium prediction may be closed at the base (Fig. 4, top). Second, the left and right ventricles may be connected due to low contrast or artifacts in the image close to the basal plane (Fig. 4, bottom).

Table 2. Comparison of DSC values obtained in the ablation study and values obtained for SA segmentation using our previous model [13] and the large human ACDC dataset. TL: transfer learning; ED: end diastole; MS: mid systole; ES: end systole.

Method	LVC			LVM		
	ED	MS	ES	ED	MS	ES
TL + B-Spline	0.924	0.903	0.866	0.821	0.824	0.815
TL - No B-Spline	0.926	0.907	0.867	0.825	0.828	0.817
No TL + B-Spline	0.921	0.906	0.858	0.801	0.818	0.799
No TL - No B-Spline	0.923	0.910	0.859	0.801	0.814	0.788
ACDC Dataset (short axis only)	0.951	-	0.859	0.866	-	0.874

4 Discussion

In the ablation study, we proposed investigating the performance of transferring the models from [13] as well as the benefits of the B-Spline head. From the analysis of the cross-validation study reported in Tables 1 and 2 as well as Fig. 2, we conclude:

- While the cost-per-epoch is the same for all cases, transferring the models from [13] helps significantly, as the optimization starts at a significantly lower loss function value. We opted for training all the tested cases for the same number of epochs (150) for a fairer comparison. Within 150 epochs, we chose the lowest validation loss (marked with black dots in Fig. 2) to select the weights of the models. We also note that the final validation loss obtained when TL is implemented is lower than the final validation loss obtained without TL, although this difference is small. In practice, other approaches requiring fewer epochs to converge may be used. For example, a different strategy consists in terminating the optimization procedure when the validation loss stabilizes (low noise). Such alternative strategy would benefit even more from the use of transfer learning as, in this case, the validation loss noise is low from the start.
- Given that the original model [13] was trained on the much larger ACDC dataset **with** the B-Spline head, it is difficult to isolate the benefit of the B-Spline head only when using TL. Indeed, the transferred model could have benefited from the B-Spline head in both cases #1 and #2, although case #2 does not include the B-Spline head during the fine tuning using the swine dataset. In addition, the models trained without TL were based only on the small swine dataset, making it difficult to conclusively analyze the role of the B-Spline head on convergence from scratch. However, we highlight that the main benefit of the B-Spline head is not in the speed of training (although it is beneficial), but consists in regularizing the UNet output and in generating smoother contours.

Given the practical advantage during training and the DSC results shown in Tab. 2, we proceeded to report the predictions’ results of the UNet_{AG} transferred

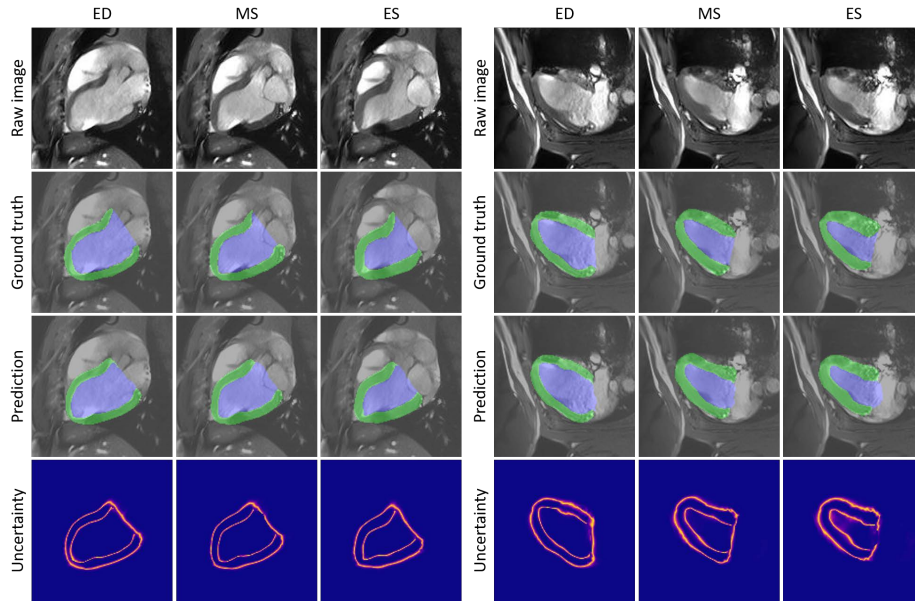


Fig. 3. Raw image, ground truth, prediction, and uncertainty results obtained in case #1. In the ground truth and prediction images, LVM and LVC are shown in green and blue, respectively. The prediction row is obtained by selecting the 50th percentiles of the UNet_{AG} probabilities. Uncertainty results illustrate the 95% prediction interval.

from [13] and using the B-Spline head (case #1). Overall, the obtained segmentations agree well with the ground truth regions (Fig. 3) at the beginning, mid, and end of systole. The LVM regions are smooth and anatomically correct, with reduced jaggedness due to image resolution and artifacts. The LVC predictions also present good agreement with the ground truth segmentations. The LVM and LVC predictions are paired with their uncertainty estimates (Fig. 3, bottom). Overall, the uncertainty estimates present small bands (few pixels across), which indicates prediction robustness with respect to the dataset. Larger uncertainties are present in regions with lower contrast and image artifacts, for example close to the apex and base of the LV.

Our proposed pipeline starts with an LVC-Locator network used to identify the region of interest (ROI). Different techniques were recently introduced for ROI isolation based on attention mechanisms [7, 4, 2], bounding boxes [10, 3], or spatiotemporal statistical analysis [5]. Although these methods are both robust and lightweight, their outputs do not carry enough information for our model. Instead, we chose to employ a UNet to compute a rough preliminary segmentation of the LVC so that we can find the ROI and, at the same time, utilize the preliminary segmentation to rotate the input images and align the LV long axis along the vertical or horizontal direction.

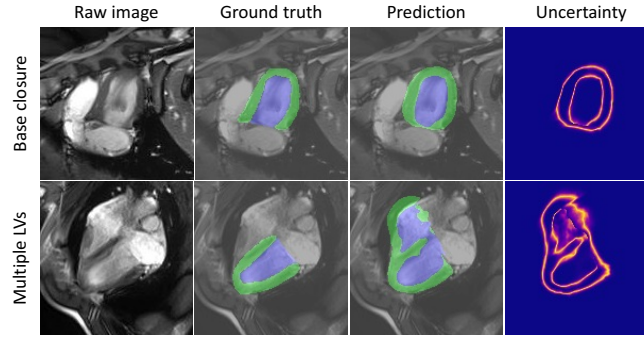


Fig. 4. Examples of specific challenges found during long axis segmentation. Top row: the LV myocardium prediction is incorrectly closed at the base. Bottom row: the left and right ventricles are incorrectly connected due to low image contrast and artifacts near the LV base.

Subsequently, the anatomically-guided deep learning models segment the LVC and LVM. The current cascade approach (the LVM-UNet_{AG} is informed by the LVC-UNet_{AG}) and the use of the B-Spline head lead to lower uncertainty regarding the location of the epicardial and endocardial outlines. An alternative approach consists in connecting two B-Spline heads to the same UNet backbone. Although losing the cascade approach would likely increase the uncertainty of epicardial and endocardial walls, this negative effect may be compensated, at least in part, by the presence of the two B-Spline heads, making this another approach to consider.

The proposed methodology also presents limitations. The first open challenge regards the possible closure of the LV base in the prediction (Fig. 4, top). A strategy to correct this problem consists in repeating the training while including a classifier to identify if the shape of the predicted LVM is physiologically correct (e.g., it has a ‘U’ versus an ‘O’ shape). Another strategy to reinforce a physiologically correct LV anatomy could follow the work presented in [8]. Additionally, one could segment the LVM first and then use the LVM segmentation as starting point to estimate the LVC. As this strategy reverses the current order (the current pipeline uses the LVC segmentation as an additional input channel for the LVM-UNet_{AG}), transfer learning could not be directly applied. However, predicting the LVC from the LVM could eliminate the over-prediction at the LVM base. A second challenge with our current approach consists in occasionally predicting multiple LV chambers (Fig. 4, bottom). As multiple chambers are usually predicted only in one cardiac phase, presenting the UNet_{AG} with images from different phases (e.g., ED, MS, and ES) could benefit the predictions. However, this solution may not be feasible with small datasets as multiple images could not be segmented independently, effectively reducing further the size of the dataset. Several of these approaches are currently being implemented to improve the proposed model for LA MRI segmentation.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant Number 2205043.

References

1. Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.A., Cetin, I., Lekadir, K., Camara, O., Gonzalez Ballester, M.A., Sanroma, G., Napel, S., Petersen, S., Tziritas, G., Grinias, E., Khened, M., Kollerathu, V.A., Krishnamurthi, G., Rohé, M.M., Pennec, X., Sermesant, M., Isensee, F., Jäger, P., Maier-Hein, K.H., Full, P.M., Wolf, I., Engelhardt, S., Baumgartner, C.F., Koch, L.M., Wolterink, J.M., Išgum, I., Jang, Y., Hong, Y., Patravali, J., Jain, S., Humbert, O., Jodoin, P.M.: 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)
2. Cheng, H., Lu, J., Luo, M., Liu, W., Zhang, K.: PTANet: Triple Attention Network for point cloud semantic segmentation. *Engineering Applications of Artificial Intelligence* **102**, 104239 (2021)
3. Ferdian, E., Suinesiaputra, A., Fung, K., Aung, N., Lukaschuk, E., Barutcu, A., Maclean, E., Paiva, J., Piechnik, S.K., Neubauer, S., Steffen, E.P., Alistair, A.Y.: Fully Automated Myocardial Strain Estimation from Cardiovascular MRI-tagged Images Using a Deep Learning Framework in the UK Biobank. *Radiology: Cardiothoracic Imaging* **2**(1), e190032 (2020)
4. Islam, M., Vibashan, V.S., Jose, V.J.M., Wijethilake, N., Utkarsh, U., Ren, H.: Brain Tumor Segmentation and Survival Prediction Using 3D Attention UNet. In: Crimi, A., Bakas, S. (eds.) *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. pp. 262–272. Springer International Publishing, Cham (2020)
5. Khened, M., Kollerathu, V.A., 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)
6. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A., van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. *Medical Image Analysis* **42**, 60–88 (2017)
7. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., Glocker, B., Rueckert, D.: Attention U-Net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999* (2018)
8. Popescu, D.M., Abramson, H.G., Yu, R., Lai, C., Shade, J.K., Wu, K.C., Maggioni, M., Trayanova, N.A.: Anatomically informed deep learning on contrast-enhanced cardiac magnetic resonance imaging for scar segmentation and clinical feature extraction. *Cardiovascular Digital Health Journal* **3**(1), 2–13 (2022)
9. Raghu, M., Zhang, C., Kleinberg, J., Bengio, S.: Transfusion: Understanding transfer learning for medical imaging. In: Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R. (eds.) *Advances in Neural Information Processing Systems*. vol. 32. Curran Associates, Inc. (2019)
10. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 779–788 (2016)

11. 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)
12. Zuben, A.V., Heckman, K., Viana, F.A.C., Perotti, L.E.: A multi-step machine learning approach for short axis MR images segmentation. In: *Functional Imaging and Modeling of the Heart: 11th International Conference, FIMH 2021*. pp. 122–133. Springer (2021)
13. Zuben, A.V., Perotti, L.E., Viana, F.A.C.: Anatomically-guided deep learning for left ventricle geometry generation with uncertainty quantification based on short-axis MR images. *Engineering Applications of Artificial Intelligence* **121**, 106012 (2023)