Article

Automated Left Ventricle Segmentation in Echocardiography Using YOLO: A Deep Learning Approach for Enhanced Cardiac Function Assessment

Madankumar Balasubramani [1] [ID], Chih-Wei Sung [2,3] [ID], Mu-Yang Hsieh [4], Edward Pei-Chuan Huang [2,3,*], Jiann-Shing Shieh [1,*] and Maysam F. Abbod [5] [ID]

1   Department of Mechanical Engineering, Yuan Ze University, Chung-Li District, Taoyuan City 320, Taiwan
2   Department of Emergency Medicine, National Taiwan University Hospital Hsin-Chu Branch, Hsinchu City 300, Taiwan
3   Department of Emergency Medicine, College of Medicine, National Taiwan University, Taipei City 100, Taiwan
4   Division of Cardiology, Department of Internal Medicine, National Taiwan University Hospital Hsin-Chu Branch, Hsinchu City 300, Taiwan; myhsieh1128@ntuh.gov.tw
5   Department of Electronics and Electrical Engineering, Brunel University London, Uxbridge UB8 3PH, UK; maysam.abbod@brunel.ac.uk
*   Correspondence: edwardhuang@ntu.edu.tw (E.P.-C.H.); jsshieh@saturn.yzu.edu.tw (J.-S.S.)

Abstract: Accurate segmentation of the left ventricle (LV) using echocardiogram (Echo) images is essential for cardiovascular analysis. Conventional techniques are labor-intensive and exhibit inter-observer variability. Deep learning has emerged as a powerful tool for automated medical image segmentation, offering advantages in speed and potentially superior accuracy. This study explores the efficacy of employing a YOLO (You Only Look Once) segmentation model for automated LV segmentation in Echo images. YOLO, a cutting-edge object detection model, achieves exceptional speed–accuracy balance through its well-designed architecture. It utilizes efficient dilated convolutional layers and bottleneck blocks for feature extraction while incorporating innovations like path aggregation and spatial attention mechanisms. These attributes make YOLO a compelling candidate for adaptation to LV segmentation in Echo images. We posit that by fine-tuning a pre-trained YOLO-based model on a well-annotated Echo image dataset, we can leverage the model's strengths in real-time processing and precise object localization to achieve robust LV segmentation. The proposed approach entails fine-tuning a pre-trained YOLO model on a rigorously labeled Echo image dataset. Model performance has been evaluated using established metrics such as mean Average Precision (mAP) at an Intersection over Union (IoU) threshold of 50% (mAP50) with 98.31% and across a range of IoU thresholds from 50% to 95% (mAP50:95) with 75.27%. Successful implementation of YOLO for LV segmentation has the potential to significantly expedite and standardize Echo image analysis. This advancement could translate to improved clinical decision-making and enhanced patient care.

Keywords: left ventricle (LV); echocardiogram; deep learning; segmentation; feature extraction; dilated convolution; YOLO

1. Introduction

Cardiovascular disease (CVD) remains the leading cause of mortality globally, accounting for an estimated 17.9 million deaths annually [1]. Accurate assessment of cardiac function is crucial for timely diagnosis, risk stratification, and guiding treatment decisions in patients with CVD [2]. Echocardiography, a non-invasive ultrasound imaging modality, plays a pivotal role in cardiac evaluation. Left ventricle (LV) segmentation, the process of delineating the boundaries of the LV chamber in an echocardiographic image, is a fundamental step in quantifying various cardiac parameters, including left ventricular ejection fraction (LVEF), a key indicator of cardiac pumping efficiency [3,4]. Traditionally,

LV segmentation is performed manually by trained cardiologists, a time-consuming and subjective process prone to inter-observer variability [5].

The limitations of manual LV segmentation have fueled the exploration of automated segmentation techniques using machine learning and, more recently, deep learning approaches [6]. Deep learning, particularly convolutional neural networks (CNNs), have demonstrated remarkable success in various medical image analysis tasks, including LV segmentation in echocardiography [7,8].

Despite the advancements in deep learning-based LV segmentation, several challenges persist. Echocardiographic images exhibit significant variability in terms of image quality, acquisition view (apical and parasternal), and patient characteristics, posing challenges for models to generalize effectively [9]. Traditional CNN architectures with limited receptive field sizes may struggle to capture long-range dependencies and contextual information crucial for accurate segmentation, especially in cases where the LV occupies a significant portion of the image [10].

The background region often dominates echocardiographic images compared to the LV area [11]. This class imbalance can lead to models prioritizing background segmentation over accurate LV delineation [12]. Real-world clinical applications necessitate fast and efficient segmentation models to enable seamless integration into clinical workflows [13]. The proposed work in this paper aims to address the limitations of existing left ventricular (LV) segmentation methods by developing a robust, accurate, and efficient deep learning-based approach. Current techniques struggle with handling anatomical variability, pathological cases, image quality variations, and incorporating prior knowledge. The research gap lies in leveraging deep learning to achieve accurate and generalizable LV segmentation while balancing computational efficiency. The motivation is to provide a reliable and clinically applicable solution that can improve diagnosis, treatment planning, and monitoring of cardiovascular diseases. By overcoming the challenges of anatomical variability, pathological cases, and image artifacts, the proposed method aims to enhance the accuracy, robustness, and generalization capability of LV segmentation, ultimately contributing to better patient care and clinical decision-making in cardiovascular imaging. This research aim is to investigate the potential of YOLOv8 for automatic LV segmentation in echocardiography [14,15]. We hypothesize that by leveraging YOLOv8's strengths, we can develop a model that offers the following advantages:

- Improved Accuracy: Achieve high Dice similarity coefficient (DSC) and Intersection over Union (IoU) metrics, indicating accurate LV delineation.
- Enhanced Generalizability: Demonstrate robust performance across diverse echocardiographic images with varying acquisition views and patient characteristics.
- Computational Efficiency: Maintain faster inference times compared to traditional CNN-based segmentation models.

After reviewing the relevant literature in Section 2, we present technical details of our approach in Section 3. The experimental evaluation comparing our method against current state-of-the-art techniques is then described in Section 4. Finally, we discuss the implications of our results and outline promising directions for future research in Section 5.

2. Related Work

Echocardiograms, ultrasound images of the heart, have become a cornerstone of cardiovascular diagnosis and management [16]. Traditionally, analyzing these images has relied on manual segmentation of anatomical structures, a time-consuming and subjective process [17]. However, the emergence of deep learning has ushered in a new era of automated segmentation, offering the potential to revolutionize how we analyze echocardiograms [18,19].

Deep learning techniques have significantly improved the accuracy and efficiency of echocardiogram analysis. Encoder–decoder networks augmented with attention mechanisms have allowed models to focus on critical regions within the image, leading to improved segmentation accuracy for structures like the LV [20,21]. Advancements in archi-

tectures, such as residual connections and dense blocks, have improved training efficiency and potentially yielded better segmentation performance [22]. A key challenge in this field is the limited availability of large, annotated datasets. Several approaches have been developed to address this:

- Data Augmentation: Strategies like random flipping and scaling have been implemented to increase training data variability and enhance model generalizability [23–26].
- Weakly Supervised Learning: These techniques utilize less-laborious annotations (e.g., bounding boxes instead of pixel-wise masks) to reduce the annotation burden while maintaining acceptable performance [27–29].
- Generative Adversarial Networks (GANs): GANs are being investigated for generating synthetic echocardiogram data to complement real data during training [30].
- Transfer Learning: Techniques for handling limited data and transfer learning from related tasks are crucial areas of ongoing research [31,32].

LV segmentation serves as a prime example of advancements in this field. Notable approaches include:

- Cascade CNN architectures and multi-scale attention networks, achieving high Dice coefficients and IoU metrics [33–35].
- TCSEgNet: A two-chamber segmentation network leveraging temporal context [36].
- UniLVSeg: Investigating both 3D segmentation and 2D super image approaches using weakly and self-supervised training [37].
- MFP-Unet: A novel architecture addressing shortcomings of traditional U-net models in LV segmentation [38].
- EchoNet Dynamic: Designed for accurate and efficient LV segmentation [39].

Motion artifacts, particularly from respiration, pose significant challenges. The I2I-cVAE method introduced a 3D image-to-image deep learning network with a conditional variational autoencoder to account for respiratory motion in whole-heart reconstructions [40]. While deep learning offers immense potential, several challenges remain:

- Robustness to image quality variations: Models need to handle variations in image quality due to acquisition protocols and patient factors [41,42].
- Interpretability and explainability: Understanding how models arrive at segmentation results is crucial for clinical trust and adoption [43].
- Real-time segmentation: Optimizing models for faster inference while maintaining accuracy could significantly improve clinical workflow efficiency [44,45].
- Clinical integration: Seamless integration into existing clinical software and workflows is essential for practical adoption.

The potential benefits of deep learning-based segmentation in clinical practice are significant, including more efficient and objective quantification of anatomical volumes, improved diagnostic accuracy, guidance for minimally invasive procedures, and facilitation of personalized treatment plans [46–50]. As research continues to address these challenges and explore new avenues, we can expect further advancements in this field [51]. Deep learning-based segmentation holds the potential to transform echocardiogram analysis, leading to more efficient workflows, improved clinical decision-making, and ultimately, better patient outcomes [52].

This work builds upon these advancements by YOLOv8n-seg, the latest in the YOLO (You Only Look Once) family, which can be adapted for left ventricle (LV) segmentation in echocardiograms. This approach combines YOLO's speed and efficiency with instance segmentation capabilities. YOLOv8n-seg uses a CSPDarknet backbone for feature extraction, a feature pyramid network for multi-scale representation, and a segmentation head alongside detection heads. This architecture is well-suited for real-time processing of echocardiogram video streams and can handle varying LV sizes and orientations. Key advantages include single-stage detection and segmentation, making it efficient for identifying and delineating

the LV. However, adapting to echocardiogram-specific characteristics and handling image quality variations remain challenges.

Training requires annotated echocardiogram datasets with LV segmentation masks. Data augmentation and transfer learning from pre-trained weights are crucial due to limited medical imaging datasets. Post-processing may involve refining segmentation masks and integrating temporal information for video sequences. Evaluation typically uses IoU (Intersection over Union) and Dice coefficient metrics, with speed and computational efficiency also being important factors. Clinical integration could enable real-time LV segmentation during echocardiogram acquisition, potentially assisting in automated calculation of LV volumes and ejection fraction.

Challenges include adapting to ultrasound-specific artifacts and ensuring consistent performance across different echocardiogram views. However, this approach leverages YOLOv8n-seg's efficiency and accuracy for LV segmentation, offering a balance of speed and precision suitable for clinical applications. This method represents a promising direction in automating and improving the accuracy of LV analysis in echocardiograms, potentially enhancing diagnostic capabilities and workflow efficiency in cardiology.

3. Materials and Methods

3.1. Dataset

The dataset used in this study comprised echocardiographic images obtained from patients presented to the emergency department National Taiwan University Hospital, Hsinchu branch, Taiwan (R.O.C.). Expert cardiologists meticulously annotated the left LV boundaries in these images to serve as ground truth labels for training and evaluating deep learning models for LV instance segmentation. The Institutional Review Board (IRB) has granted approval for this study under reference number 110-069-E.

A total of 4781 echocardiographic image cases were selected and split into three subsets: 3877 for training, 477 for validation, and 427 reserved for testing. The partitioning into these mutually exclusive subsets was performed using a random sampling approach to ensure an unbiased distribution across the training, validation, and testing data. The images were acquired using a variety of echocardiographic modalities (2D, 3D, stress Echo) and imaging planes, reflecting the diverse range of data encountered in real-world clinical settings.

The dataset intentionally included cases with varying image quality, noise levels, artifacts, and other inherent challenges typical of emergency echocardiography. For each case, expert cardiologists manually delineated the endocardial borders of the LV cavity throughout the full cardiac cycle, generating precise instance-level segmentation masks. This labor-intensive annotation process leveraged decades of specialist experience to ensure highly accurate ground truth labels. By rigorously annotating this large, multi-institutional dataset, this study aimed to develop and validate deep learning algorithms capable of providing robust, automated LV segmentation to aid cardiovascular imaging interpretation and quantification in emergency settings.

3.2. YOLOv8's Architecture

YOLOv8 was developed by Ultralytics [53] and utilizes a convolutional neural network that can be divided into two main parts: the backbone and the head, as shown in Figure 1. A modified version of the Cross Stage Partial (CSP) Darknet53 [54] architecture forms the backbone of YOLOv8. This architecture consists of 53 convolutional layers and employs cross-stage partial connections to improve information flow between the different layers.

The backbone is the base convolutional neural network (CNN) that extracts features from the input image. The backbone takes the input image and produces a feature map, which is a tensor containing high-level representations of the image at different spatial locations.
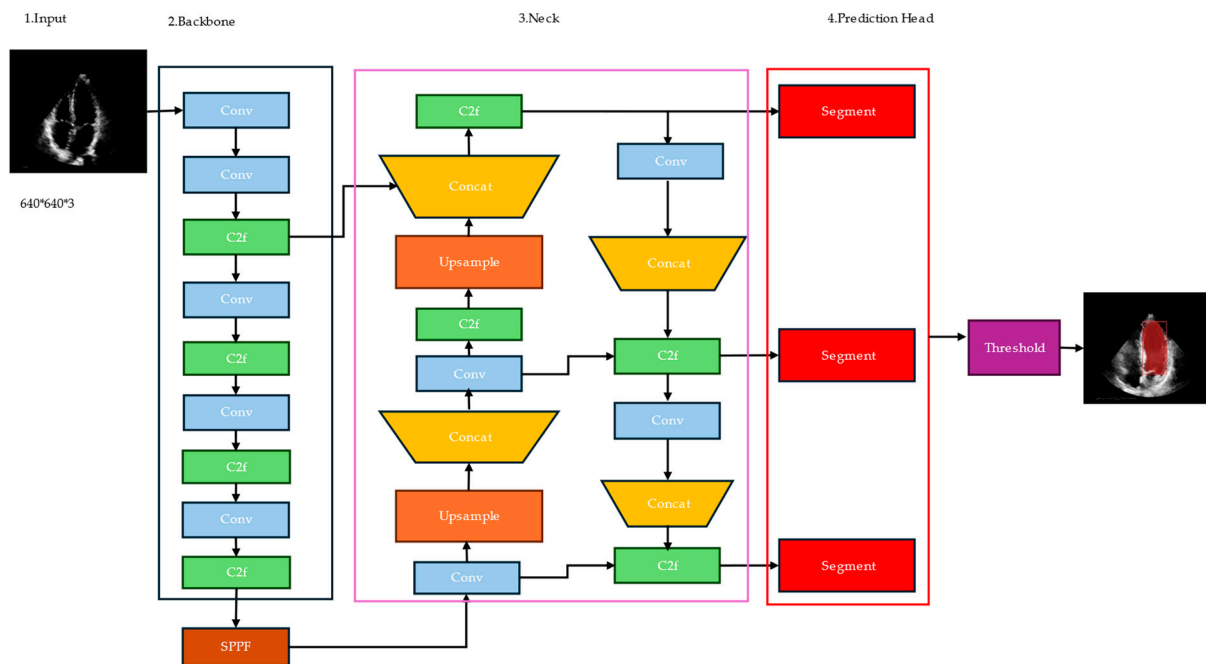
Figure 1. Architecture of YOLOv8.

The neck, also known as the feature pyramid network (FPN), is responsible for combining features from different scales of the backbone's output. The purpose of the neck is to capture both high-resolution and low-resolution information from the input image, which is essential for accurate segmentation. FPN consists of a top-down pathway and a bottom-up pathway. The top-down pathway takes the high-level semantic features from the backbone and up-samples them to higher resolutions, while the bottom-up pathway takes the low-level features and combines them with the up-sampled features. Compared to its predecessors, YOLOv8 introduces several improvements that make it particularly well-suited for medical image analysis tasks like left ventricle segmentation.

- Multi-Scale Feature Extraction: The YOLOv8n-seg model employs a feature pyramid network (FPN) backbone that extracts feature maps at multiple scales from the input image, capturing both high-resolution and low-resolution information.
- Fusion of Multi-Scale Features: The model's decoder component fuses the multi-scale feature maps from the FPN backbone through a top-down and bottom-up pathway. This fusion process combines the high-resolution features, which capture fine-grained details, with the low-resolution features, which encode global contextual information.
- Benefits for Left Ventricle Segmentation: The fusion of multi-scale features is particularly beneficial for the left ventricle segmentation task. The high-resolution features help in accurately delineating the intricate boundaries and shape of the left ventricle, while the low-resolution features provide contextual information about the surrounding anatomical structures, aiding in distinguishing the left ventricle from other cardiac structures.

This fusion of features from different scales helps the model better localize objects and capture fine-grained details. The head is the final component of the YOLO segmentation model, responsible for generating the segmentation masks. It takes the combined features from the neck and applies a series of convolutional layers to produce the final segmentation output. In the YOLOv8n-seg segmentation model, the head typically consists of multiple branches, each responsible for predicting the segmentation masks at a different scale. This multi-scale prediction allows the model to capture objects of varying sizes and resolutions. Each branch in the head outputs a tensor with a specific number of channels, where each

channel corresponds to a different class or object category. The values in these channels represent the confidence scores or probabilities of each pixel belonging to a particular class.

3.3. Proposed Architecture

Our proposed architecture for instance segmentation utilizes several modules and techniques to achieve accurate object detection and segmentation, as shown in Figure 2. Among them, dilated convolution, c2f, SPPF, and segment modules play crucial roles. While YOLOv8-seg demonstrates remarkable capabilities in object detection, it can encounter difficulties when identifying specific boundaries within complex medical images. This is particularly true for tasks like left ventricle segmentation, where accurately delineating the border (inner lining of the left ventricle) is crucial. Here follows a breakdown of the challenges:
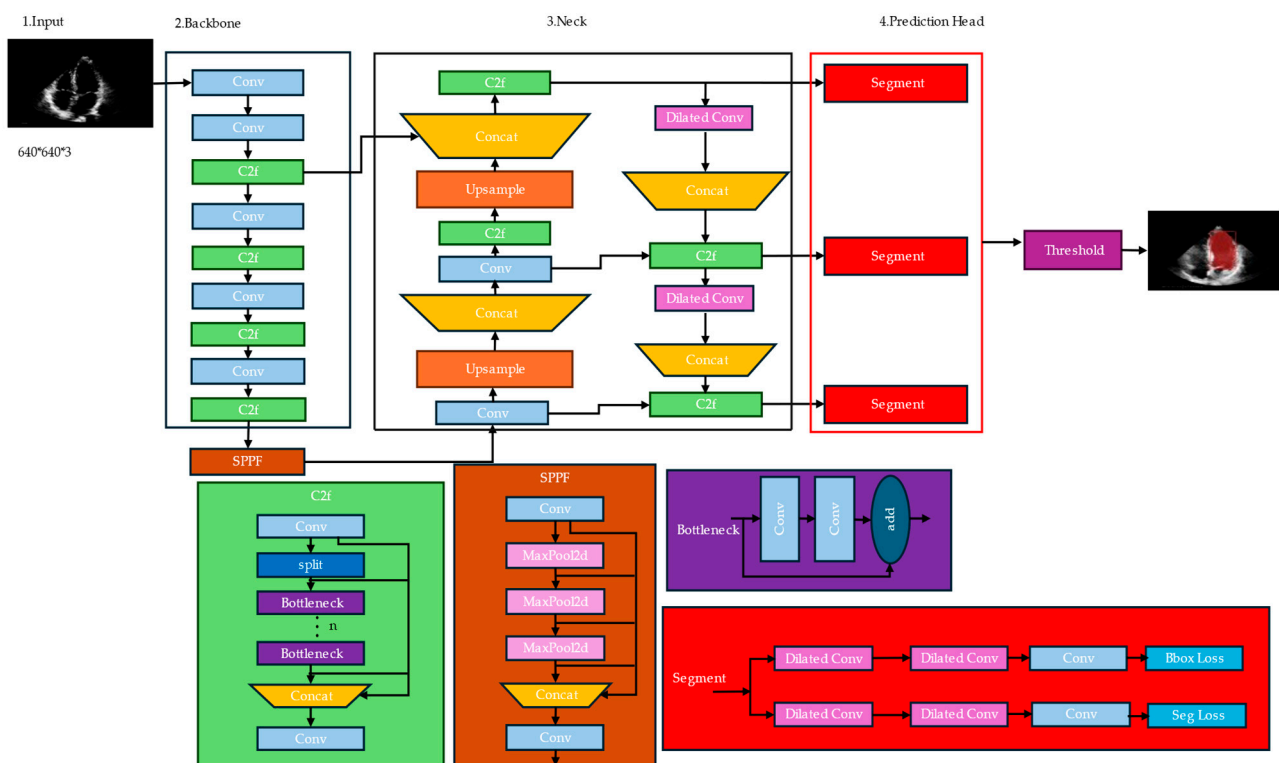

Figure 2. Improved architecture of YOLOv8n-seg for left ventricle segmentation.

- Information loss during feature extraction: The process of extracting features from an image can lead to the loss of crucial details, especially for intricate structures like the endocardial border. As the network processes information at deeper layers, it prioritizes prominent features, potentially neglecting the finer details that define the endocardial border in complex images.
- Difficulties with overlapping structures: The endocardial border can be obscured or overlapped by other structures within the left ventricle, such as papillary muscles or trabeculae. This overlap makes it challenging for the network to accurately distinguish and locate the precise boundary of the left ventricle.

These challenges can lead to inaccurate segmentation of the left ventricle, impacting downstream medical applications. To address these challenges, we propose a novel segmentation algorithm that significantly improves the segmentation of small structures like the left ventricle while maintaining accuracy for larger ones. Here are the key components:

- Enhanced down-sampling: We introduce a new down-sampling module that utilizes depth-wise separable convolution followed by a combination of MaxPooling and a 3×3 convolution with stride = 2. This concatenation approach effectively recovers

information lost during the down-sampling process. Consequently, it preserves a more complete picture of the context throughout feature extraction, leading to better preservation of the left ventricle's features.

- Improved feature fusion: The network incorporates an enhanced feature fusion method. This method facilitates a better integration of shallow and deep information. By combining low-level details with high-level semantic understanding, the network retains more comprehensive information about the left ventricle. This improves segmentation accuracy by reducing the issue of overlooking small structures due to the dominance of larger features.

3.3.1. Dilated Convolution

Dilated convolution [55–58] is employed in CNNs for tasks such as image segmentation and object detection. Unlike traditional convolution operations, where the kernel weights are applied contiguously across the input, dilated convolution introduces strategically spaced gaps or holes between the kernel elements as shown in Figure 3. In YOLOv8n-seg, they likely play a crucial role in improving LV segmentation performance by capturing fine-grained details while maintaining spatial resolution. This approach allows the network to effectively handle multi-scale features, enhancing object detection and segmentation across various sizes. By employing dilated convolutions, YOLOv8n-seg can potentially achieve a better balance between local and global context understanding, leading to more accurate and detailed segmentation results.
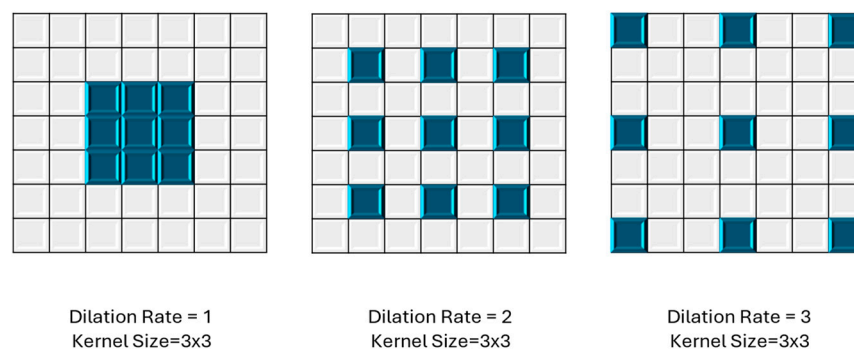

Dilation Rate = 1
Kernel Size=3x3

Dilation Rate = 2
Kernel Size=3x3

Dilation Rate = 3
Kernel Size=3x3

Figure 3. Different dilation rates in dilated convolutions, using a 3×3 kernel size.

3.3.2. C2F (Class-to-Fortitude) Module

The C2F module is responsible for enhancing the model's ability to distinguish between different object classes. It takes the class predictions (output of the classification head) and the bounding box predictions (output of the regression head) as inputs. The C2F module then applies a series of convolutions and up-sampling operations to generate feature maps that are fused with the segmentation features. This fusion helps the segmentation head better differentiate between object instances of different classes, leading to improved instance segmentation performance.

3.3.3. Spatial Pyramid Pooling Fortitude Module

The SPPF module is an extension of the Spatial Pyramid Pooling Fortitude (SPPF) technique, which has been widely used in object detection models like Faster R-CNN [59] and YOLOv5 [53]. The SPPF module is designed to capture multi-scale features by applying parallel pooling operations at different kernel sizes and strides. This allows the model to effectively handle objects of varying scales and sizes within an image.

In the context of instance segmentation, the SPPF module takes the feature maps from the backbone network and applies parallel pooling operations at different scales. The resulting feature maps are then concatenated and passed through a series of convolutions to generate enhanced feature representations. These multi-scale features are then fused

with the segmentation features, aiding the model in detecting and segmenting objects at various scales.

3.3.4. Segmentation Module

The segmentation module is responsible for generating instance segmentation masks from the enhanced feature maps produced by the C2F and SPPF modules. It consists of a set of parallel convolutional layers that generate a fixed number of prototype masks (typically 32 or 64). These prototype masks are then combined with the bounding box predictions and segmentation predictions to produce instance-specific segmentation masks. The key idea behind this module is to learn a set of mask representations during training. During inference, these are adaptively combined and scaled based on the predicted bounding boxes and classes to generate final instance segmentation masks. This approach allows the model to generate high-quality segmentation masks without the need for explicit pixel-wise labeling of instances during training, which can be time-consuming and laborious.

The Darknet Bottleneck module is a key component of the backbone network used in YOLOv8n-seg. It is inspired by the Darknet-53 architecture, which was initially introduced in the YOLOv3 object detection model. This module is designed to balance the trade-off between computational efficiency and representational capacity, allowing for efficient feature extraction while preserving important spatial and semantic information. The Darknet Bottleneck module consists of the following components:

- Convolutional layer: The input feature maps are first processed by a convolutional layer with a 1×1 kernel size. This layer serves as a dimensionality reduction step, reducing the number of channels in the feature maps. This operation is computationally efficient and helps reduce the overall computational complexity of the network.
- Batch normalization and activation: After the convolutional layer, batch normalization is applied to stabilize the training process and improve convergence. This is followed by an activation function, typically the leaky Rectified Linear Unit (ReLU), which introduces non-linearity into the feature representations.
- Convolutional layer with bottleneck: The next step involves a convolutional layer with a 3×3 kernel size, which is the main feature extraction component of the module. However, instead of using the full number of channels, a bottleneck approach is employed. The number of channels in this layer is typically set to a lower value (e.g., one-quarter or one-half of the input channels) to reduce computational complexity while still capturing important spatial and semantic information.

The Darknet Bottleneck module is repeated multiple times within the backbone network, with the number of repetitions determined by the specific architecture (e.g., Darknet-53 in YOLOv8). This modular design allows for efficient feature extraction while maintaining a balance between computational complexity and representational capacity.

By incorporating the Darknet Bottleneck module into the YOLOv8-seg backbone network, the model can efficiently extract rich feature representations from the input images, which are then used by the subsequent modules for accurate instance segmentation. For medical image segmentation tasks, a combination of Robust TLoss [60] (Truncated L1 Loss) and Dice Loss [61] can be used as the loss function. This combined loss function [62] aims to leverage the strengths of both loss functions: Robust TLoss for handling outliers and Dice Loss for optimizing the overlap between the predicted and ground truth segmentation masks.

Equation (1) for the combined loss function can be written as follows:

$$T_{Seg} = \alpha \times Robust\ TLoss(X, Y) + \beta \times Dice\ Loss(X, Y) \tag{1}$$

where

- 'X' is the ground truth segmentation mask;
- 'Y' is the predicted segmentation mask;

- 'Robust TLoss(X, Y)' is the Robust Truncated L1 Loss between the ground truth and predicted masks;
- 'Dice Loss(X, Y)' is the Dice Loss between the ground truth and predicted masks; and
- 'α' and 'β' are weights that control the relative importance of the two loss terms.

The Robust TLoss term is calculated using Equation (2):

$$Robust\ TLoss(X, Y) = \begin{cases} \left(\frac{1}{2}\right) \times (X - Y)^2 & if\ |X - Y| < 1 \\ |X - Y| - \left(\frac{1}{2}\right) & otherwise \end{cases} \tag{2}$$

This term helps to handle outliers and large errors in the segmentation masks, while still being differentiable for small errors.

The Dice Loss term is calculated using Equation (3):

$$Dice\ Loss(X, Y) = 1 - (2 \times |X \cap Y| + \varepsilon)/(|X| + |Y| + \varepsilon) \tag{3}$$

where 'ε' is a small constant value (e.g., 1×10^{-5}) added for numerical stability.

By combining these two loss terms, the overall loss function aims to achieve a balance between handling outliers and optimizing the overlap between the predicted and ground truth segmentation masks. The weights 'α' and 'β' are tuned based on the LV segmentation task and the characteristics of the dataset. The weight selection process often involves hyperparameter tuning and empirical evaluation. Common strategies include grid search, random search, or more advanced techniques like Bayesian optimization. Bayesian optimization provides a powerful and sample-efficient approach for tuning the α and β weights in the combined loss function for LV segmentation, especially when computational resources are limited or the hyperparameter space is complex.

It is important to note that the optimal weight selection may depend on factors such as the dataset, task, and model architecture. Experimentation and validation on held-out data are crucial to ensure that the combined loss function effectively leverages the strengths of both loss functions and leads to improved model performance.

This combined loss function has been used in various medical image segmentation applications, such as tumor segmentation, organ segmentation, and lesion detection, and has shown improved performance compared to using either loss function alone.

YOLOv8n-seg's ability to perform detections in real-time is crucial for potential clinical applications where fast segmentation results are essential for diagnosis and treatment decisions. YOLO is a well-known architecture with a large user base. This provides access to readily available implementation resources and facilitates troubleshooting if needed. YOLOv8n-seg offers multiple versions catering to different needs. This allows for choosing a configuration that balances accuracy and resource requirements for deployment in clinical settings. The hyperparameters of YOLOv8n-seg are listed in Table 1.

Table 1. Hyperparameters of Yolov8n-seg model.

Hyperparameter	Value	Description
model	yolov8n-seg.pt	Pre-trained YOLOv8n segmentation model.
data	coco128.yaml	Data configuration file for the COCO dataset.
epochs	200	Number of training epochs.
batch	16	Batch size for training.
imgsz	640	Input image size.
optimizer	SGD	Optimization algorithm (Stochastic Gradient Descent).
lr0	0.01	Initial learning rate.
lrf	0.01	Final learning rate.
momentum	0.937	Momentum value for the optimizer.
weight_decay	0.0005	Weight decay regularization value.

Table 1. *Cont.*

Hyperparameter	Value	Description
warmup_epochs	3.0	Number of warmup epochs.
warmup_momentum	0.8	Momentum value during warmup.
cosine	True	Whether to use cosine annealing for the learning rate scheduler.
box	7.5	Box loss weight.
cls	0.5	Classification loss weight.
mask	2.0	Mask (segmentation) loss weight.

3.4. Evaluation Metric

3.4.1. Dice Similarity Coefficient (DSC):

In medical image analysis and computer vision, DSC is widely used to evaluate the performance of segmentation algorithms. It helps to assess how well an algorithm can delineate a specific region of interest (ROI) in an image, such as the left ventricle.

Equation (4) is as follows:

$$DSC = \frac{2 \times |A \cap B|}{|A| + |B|} \tag{4}$$

where A represents the predicted area of the mask and B represents the ground truth containing the object. A higher DSC value signifies a greater similarity between the two sets being compared. In image segmentation tasks, a high DSC indicates that the algorithm successfully segmented the target region with minimal errors (omission or inclusion of unnecessary pixels).

3.4.2. Intersection over Union

IoU is a classical metric for evaluating the performance of the model for object detection. It calculates the ratio of the overlap and union between the predicted bounding box and the ground truth bounding box, which measures the intersection of these two bounding boxes. The IoU is represented by Equation (5):

$$IoU(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{5}$$

where A represents the predicted area of the mask and B represents the ground truth containing the object. The performance of the model improves as the IoU value increases, with higher IoU values indicating less difference between the generated candidate and ground truth bounding boxes.

3.4.3. Mean Average Precision (mAP)

Mean Average Precision (mAP) considers both precision (correctly identified objects) and recall (detecting all true positives) across different IoU thresholds. It provides a comprehensive overview of the model's detection performance. The mAP@50 metric specifically refers to the mAP calculated using an IoU threshold of 0.5. In simpler terms, it represents the average precision of detections where the predicted bounding box or segmentation mask overlaps with the ground truth by at least 50%. This is a commonly used threshold for evaluating object detection models. The mAP@0.5:0.95 metric represents the mean average precision across a range of IoU thresholds, typically from 0.5 to 0.95 with increments of 0.05. It provides a more in-depth analysis of the model's performance under varying degrees of overlap between predictions and ground truth. A higher average mAP across this range suggests the model performs well even with less perfect overlaps.

3.4.4. Precision-Recall Curve

Precision-Recall Curve (P-R Curve) is a curve with recall as the *x*-axis and precision as the *y*-axis. Each point represents a different threshold value, and all points are con-

nected as a curve. The recall I and precision (P) are calculated according to the following Equations (6) and (7):

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

where True Positive (*TP*) denotes a prediction result as a positive class that is judged to be true; False Positive (*FP*) denotes a prediction result as a positive class that is judged to be false, and False Negative (*FN*) denotes a prediction result as a negative class that is judged to be false.

4. Results and Discussion

In this study, the training and testing for the network were conducted on a workstation with an Intel Core i7-11700K @3.6 GHZ, NVIDIA GeForce RTX 4090 24G GPU, and Windows 11 operating system. A pre-trained YOLOv8 model was used in the training process from MS COCO [63] (Microsoft Common Objects in Context) val2017.

Table 2 compares five YOLOv8-seg models (likely denoted by size—n, s, m, l, and x—signifying increasing complexity) across various metrics. These models are evaluated on their ability to identify the LV in medical images. Precision and recall measure the models' effectiveness in correctly classifying pixels, while mAP50 and mAP50-95 assess overall detection accuracy with varying strictness (Intersection over Union thresholds).

Table 2. Performance comparison of various YOLO models.

Model	Size (Pixels)	Precision	Recall	mAP50	mAP50-95	Params (M)	FLOPs (B)
YOLOv8n-seg	416	0.97247	0.95840	0.96064	0.75742	3.4	12.6
YOLOv8s-seg	416	0.97306	0.96771	0.97887	0.75604	11.8	42.6
YOLOv8m-seg	416	0.97363	0.97692	0.97957	0.75818	27.3	110.2
YOLOv8l-seg	416	0.97338	0.97899	0.97964	0.75626	46	220.5
YOLOv8x-seg	416	0.97572	0.97907	0.98005	0.75784	71.8	344.1
YOLOv8n-seg	640	0.97448	0.97456	0.97973	0.75875	3.4	12.6
YOLOv8s-seg	640	0.97651	0.97571	0.98164	0.76066	11.8	42.6
YOLOv8m-seg	640	0.9768	0.97894	0.98271	0.75816	27.3	110.2
YOLOv8l-seg	640	0.97583	0.97770	0.98263	0.75821	46	220.5
YOLOv8x-seg	640	0.97654	0.97921	0.98269	0.75852	71.8	344.1
YOLOv8n-seg	1280	0.97651	0.97907	0.98154	0.75671	3.4	12.6
YOLOv8s-seg	1280	0.97654	0.97907	0.97932	0.75164	11.8	42.6
YOLOv8m-seg	1280	0.97657	0.97907	0.98108	0.75491	27.3	110.2
YOLOv8l-seg	1280	0.9766	0.97907	0.98126	0.75542	46	220.5
YOLOv8x-seg	1280	0.97661	0.97907	0.98071	0.75409	71.8	344.1

A crucial aspect of the table is the inclusion of model complexity parameters. The number of trainable parameters (millions) and FLOPs (floating-point operations) required per image indicate the model's computational burden. Generally, larger models with more parameters (l, x) tend to achieve higher mAP scores, signifying better segmentation accuracy. However, this comes at the cost of increased computational demand, reflected in higher FLOPs.

Table 2 also explores the impact of input image size. By comparing models trained on different resolutions (416, 640, and 1280 pixels), we can observe that for some models, increasing the resolution can lead to marginal improvements in segmentation accuracy. This is likely because higher resolution images provide more detailed information about the LV structure.

The key takeaway from this analysis is the interplay between model complexity and computational efficiency. While larger models consistently outperform smaller ones in terms of mAP scores, this comes at a significant cost. Their increased number of parameters

translates to higher computational demands, making them less suitable for real-time applications or deployment on devices with limited resources.

For instance, YOLOv8x-seg achieves the highest mAP scores across all resolutions but requires over 71 million parameters and 344 billion FLOPs per image. Conversely, YOLOv8n-seg, with only 3.4 million parameters and 12.6 billion FLOPs, offers a faster and more resource-friendly solution, albeit with a slight decrease in segmentation accuracy. If prioritizing the highest segmentation accuracy is paramount, a larger model (l or x) might be preferable despite its computational demands. This scenario could be relevant in research settings where precise LV measurements are critical.

However, for real-time applications or deployment on mobile devices, computational efficiency becomes a major concern. In such cases, a smaller model (n or s) might be a more suitable choice. While sacrificing some accuracy, these models offer faster processing speeds and lower memory requirements, making them ideal for resource-constrained environments.

Beyond the metrics presented in Table 2, several other factors influence model selection. The available computational resources and memory constraints of the deployment environment play a crucial role. Additionally, the quality and quantity of training data can significantly impact model performance. Following the training, a confusion matrix was generated by analyzing the results matrix, as shown in Figure 4. Based on the confusion matrix, the classification model appears to be performing reasonably well for the positive class, but there is one instance misclassified as a false positive and one instance misclassified as a false negative. Henceforth, while the model seems to be doing a decent job, especially for the majority LV class, there are a few misclassified instances for both classes that could potentially be improved upon by further tuning or adjusting the model's parameters or features.

Figure 4. Confusion matrix.

The recall and precision are of paramount importance, and the pursuit of enhanced performance is a constant endeavor. Table 3 presents a comprehensive ablation of four models: Yolov8n-seg, RobustTLoss, DiceLoss, and the proposed model, each vying for superiority across the metrics of precision, recall, mAP50, and mAP50-95.

The Yolov8n model serves as the baseline, with respectable scores of 0.93532 for precision, 0.94171 for recall, 0.95777 for mAP50, and 0.62485 for mAP50-95. However, the introduction of the RobustTLoss and DiceLoss functions demonstrates the potential for improvement. The RobustTLoss model outperforms the baseline across all metrics, boasting a precision of 0.95284, recall of 0.95581, mAP50 of 0.96487, and mAP50-95 of

0.61742. Similarly, the DiceLoss model exhibits superior precision (0.94971) and mAP50 (0.96788) compared to baseline, although its recall (0.93488) and mAP50-95 (0.60437) trail slightly behind. While both the RobustTLoss and DiceLoss models demonstrate their prowess, the true standout is the proposed model. With a remarkable precision of 0.98359, recall of 0.97561, mAP50 of 0.9831, and a staggering mAP50-95 of 0.7527, this model sets a new benchmark for object detection performance. Its ability to strike an exquisite balance between precision and recall, coupled with its exceptional mAP50 and mAP50-95 scores, is a testament to the innovative techniques employed in its development.

Table 3. Comparison of proposed model with various loss functions.

Model	Precision	Recall	mAP50	mAP50-95
Yolov8n-Seg	0.93532	0.94171	0.95777	0.62485
Yolov8n-Seg +RobustTLoss [60]	0.95284	0.95581	0.96487	0.61742
Yolov8n-Seg +DiceLoss [61]	0.94971	0.93488	0.96788	0.60437
Ours	0.98359	0.97561	0.9831	0.7527

The proposed model's superiority is particularly evident in the mAP50-95 metric, which evaluates the model's performance across various IoU thresholds. Its score of 0.75876 is a significant improvement over the baseline and the other models, indicating robust and consistent object detection capabilities across a wide range of scenarios. In Figure 5, the mAP50 (mean Average Precision at IoU = 0.5) metric evaluates how well the model's predicted segmentation masks overlap with the ground truth masks, considering an IoU threshold of 0.5. mAP50-95 assesses this overlap across a range of IoU thresholds from 0.5 to 0.95, providing a more comprehensive evaluation of segmentation quality. In this context, the consistent and high performance of the (B) model across precision, recall, mAP50, and mAP50-95 suggests that it is a reliable and accurate approach for left ventricle segmentation. However, model M, while competitive in some metrics like precision and mAP50, exhibits more variability and lower performance in recall and mAP50-95, indicating potential challenges in capturing all relevant left ventricle regions or handling different IoU thresholds.
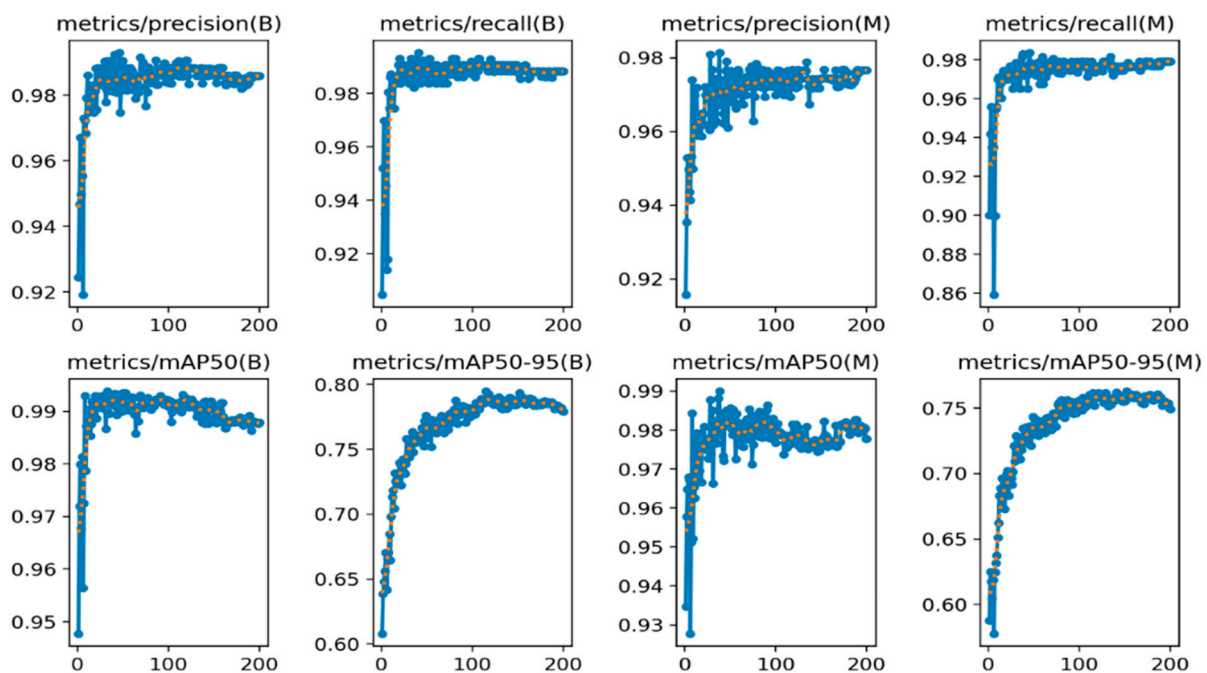

Figure 5. mAP 50 and mAP50:95 for 200 epochs. Training curve denoted by blue dotted line and validation curve denoted by orange dotted line.

In this work, Table 4 compares the performance of the proposed deep learning model for left ventricular (LV) segmentation against several state-of-the-art (SOTA) methods, including SegNet [64], DeepLabv3 [65], MFP-UNet (2019), Echonet (2020), SegAN (2021), and TC-SegNet (2023). We evaluate these models with in-house test set, which exhibits significant variations in image quality, anatomical structures, and pathological conditions. To quantitatively evaluate the performance of our model against the SOTA models, we report the mean Dice similarity coefficient (DSC) and Intersection over Union (IoU) on our test set. Our proposed model achieves a mean DSC of 0.9800 and an IoU of 0.7600, outperforming SegNet (DSC: 0.7651, IoU:0.6195), DeepLabv3 (DSC: 0.7890, IoU:0.7200), MFP-Unet (DSC: 0.7832, IoU:0.7390), EchoNet (DSC: 0.9200), SegAN (2021) (DSC: 0.8566, IoU:0.8122), and TC-SegNet (2023) (DSC: 0.9559, IoU:0.8882). Our qualitative results further demonstrate the robustness and accuracy of our model in handling challenging cases, such as poor image quality, anatomical variations, and pathological conditions. In contrast, SegNet, DeepLabv3, and EchoNet often struggle with these challenging scenarios, leading to inaccurate or incomplete segmentations. Deep learning architectures for left ventricle (LV) segmentation in echocardiograms have evolved significantly, with a trend towards efficiency and reduced parameter counts. SegNet, an early encoder–decoder network, contains approximately 29.5 million parameters. The MFP-UNet (2019) reduced this to about 7.8 million using a multi-scale feature pyramid design. EchoNet (2020), developed specifically for echocardiogram analysis, was further optimized to around 6.5 million parameters. SegAN, a GAN-based approach from 2021, typically has 8–10 million parameters, varying with architectural choices. DeepLabv3, while not exclusive to echocardiograms, is a powerful semantic segmentation model with 40–60 million parameters, depending on its backbone. The recent TC-SegNet (2023) demonstrates impressive efficiency with only about 5.2 million parameters. For comparison, YOLOv8n-seg adapted for segmentation has around 3.2 million parameters, showcasing potential for even lighter models. This evolution reflects a shift towards balancing high accuracy with lower computational demands, which is particularly vital in medical imaging. The progression of these architectures underscores the field's commitment to developing powerful yet resource-efficient solutions for echocardiogram analysis, potentially improving real-time performance and broader applicability in clinical practice.

Table 4. Performance comparison of SOTA methods.

Model	Number of Parameters (in Millions)	DSC	IoU
Deeplabv3	40	0.7890	0.7200
SegNet	29.5	0.7651	0.6195
MFP-UNet (2019)	7.8	0.7832	0.7390
Echonet (2020)	6.5	0.9200	-
SegAN (2021)	9.5	0.8566	0.8122
TC-SegNet (2023)	5.2	0.9559	0.8882
Ours	3.2	0.9800	0.7600

In Figure 6, it can be evidently seen that the proposed model has outperformed on different kinds of shapes and contours for left ventricle segmentation. Accurate and robust left ventricular segmentation from Echo images is crucial for various clinical applications, such as assessing cardiac function, monitoring disease progression, and guiding treatment decisions. Continued refinement and validation of automated segmentation algorithms are essential to ensure reliable and clinically meaningful results.
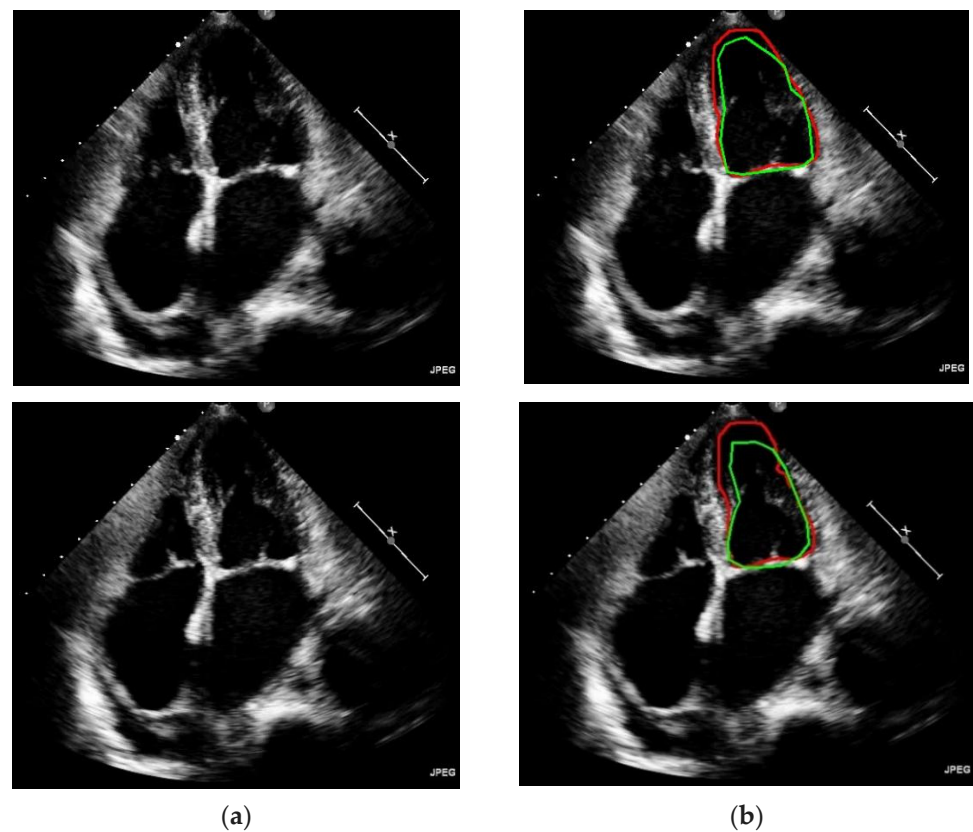
Figure 6. Predicted left ventricle segmentation. (**a**) Original input image (**b**) Ground truth annotated by the experts are denoted by green colored line, and segmentation of frames by model are denoted by red colored line.

5. Conclusions

Accurate and reliable left ventricle segmentation from echocardiogram images is a crucial task in cardiovascular imaging and diagnosis. This study demonstrates the potential of automated segmentation models to capture the overall shape and location of the left ventricular cavity with a highest mAP50 of 98.31% and mAP50:95 of 75.27%. However, some instances of misalignment or inaccuracies highlight the need for further refinement and validation of these algorithms.

Precise delineation of the left ventricular boundaries is essential for quantitative assessments, such as measuring volumes, ejection fraction, and regional wall motion abnormalities. These measurements play a vital role in diagnosing and monitoring various cardiovascular conditions, including heart failure, valvular diseases, and myocardial infarctions.

To advance the field of left ventricle echocardiogram segmentation, several areas of future work can be explored:

- Larger and more diverse datasets: Training segmentation models on larger and more diverse datasets, encompassing various pathologies, imaging modalities, and acquisition protocols, can enhance their generalization capabilities and robustness.
- Incorporation of temporal information: Echocardiograms capture dynamic cardiac cycles. Leveraging temporal information by integrating recurrent neural networks or temporal modeling techniques could improve segmentation accuracy and consistency across frames.
- Uncertainty quantification: Developing methods to quantify the uncertainty or confidence of segmentation predictions can provide valuable insights for clinicians and aid in decision-making processes.

This work has demonstrated YOLOv8n-seg's effectiveness for left ventricle segmentation from echocardiograms. However, ethical risks must be considered. Potential model

bias from unrepresentative training data could lead to errors and discrimination. Robust de-identification and security are crucial to protect patients' private health data. Clear policies governing appropriate use are needed to prevent misuse, such as unauthorized surveillance. As a powerful anatomical mapping tool, safeguards against weaponization or exploiting individuals are required. While promising, the development and deployment of such medical AI must involve multistakeholder collaboration to implement governance mitigating ethical risks. Only then can its benefits be fully realized while protecting against misuse. By addressing these challenges and opportunities, the field of left ventricle echocardiogram segmentation can advance towards more accurate, reliable, and clinically applicable solutions, ultimately improving cardiovascular care and patient outcomes.

Author Contributions: M.B. contributed to the development of the segmentation algorithm and the implementation of the deep learning model. C.-W.S. assisted in the data preprocessing and augmentation techniques for the echocardiogram images. M.-Y.H. played a role in the evaluation of the segmentation results and the calculation of performance metrics. E.P.-C.H. provided guidance and expertise in the field of cardiovascular imaging and interpretation of echocardiograms. M.F.A. contributed to the study design and the clinical validation of the segmentation results. J.-S.S. supervised the overall research project, provided valuable insights, and helped in the analysis of the results. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Approval was granted under reference number 110-069-E.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Gaziano, T.A. Cardiovascular diseases worldwide. *Public Health Approach Cardiovasc. Dis. Prev. Manag* **2022**, *1*, 8–18.
2. Lloyd-Jones, D.M.; Braun, L.T.; Ndumele, C.E.; Smith, S.C., Jr.; Sperling, L.S.; Virani, S.S.; Blumenthal, R.S. Use of risk assessment tools to guide decision-making in the primary prevention of atherosclerotic cardiovascular disease: A special report from the American Heart Association and American College of Cardiology. *Circulation* **2019**, *139*, e1162–e1177. [CrossRef] [PubMed]
3. Chen, R.; Zhu, M.; Sahn, D.J.; Ashraf, M. Non-invasive evaluation of heart function with four-dimensional echocardiography. *PLoS ONE* **2016**, *11*, e0154996. [CrossRef]
4. Cacciapuoti, F. The role of echocardiography in the non-invasive diagnosis of cardiac amyloidosis. *J. Echocardiogr.* **2015**, *13*, 84–89. [CrossRef]
5. Karim, R.; Bhagirath, P.; Claus, P.; Housden, R.J.; Chen, Z.; Karimaghaloo, Z.; Sohn, H.-M.; Rodríguez, L.L.; Vera, S.; Albà, X. Evaluation of state-of-the-art segmentation algorithms for left ventricle infarct from late Gadolinium enhancement MR images. *Med. Image Anal.* **2016**, *30*, 95–107. [CrossRef]
6. Slomka, P.J.; Dey, D.; Sitek, A.; Motwani, M.; Berman, D.S.; Germano, G. Cardiac imaging: Working towards fully-automated machine analysis & interpretation. *Expert Rev. Med. Devices* **2017**, *14*, 197–212.
7. Kim, T.; Hedayat, M.; Vaitkus, V.V.; Belohlavek, M.; Krishnamurthy, V.; Borazjani, I. Automatic segmentation of the left ventricle in echocardiographic images using convolutional neural networks. *Quant. Imaging Med. Surg.* **2021**, *11*, 1763. [CrossRef] [PubMed]
8. Wong, K.K.; Fortino, G.; Abbott, D. Deep learning-based cardiovascular image diagnosis: A promising challenge. *Future Gener. Comput. Syst.* **2020**, *110*, 802–811. [CrossRef]
9. Zolgharni, M. Automated Assessment of Echocardiographic Image Quality Using Deep Convolutional Neural Networks. Ph.D. Thesis, The University of West London, London, UK, 2022.
10. Luo, X.; Zhang, H.; Huang, X.; Gong, H.; Zhang, J. DBNet-SI: Dual branch network of shift window attention and inception structure for skin lesion segmentation. *Comput. Biol. Med.* **2024**, *170*, 108090. [CrossRef]
11. Palmieri, V.; Dahlöf, B.; DeQuattro, V.; Sharpe, N.; Bella, J.N.; de Simone, G.; Paranicas, M.; Fishman, D.; Devereux, R.B. Reliability of echocardiographic assessment of left ventricular structure and function: The PRESERVE study. *J. Am. Coll. Cardiol.* **1999**, *34*, 1625–1632. [CrossRef]
12. Jin, X.; Thomas, M.A.; Dise, J.; Kavanaugh, J.; Hilliard, J.; Zoberi, I.; Robinson, C.G.; Hugo, G.D. Robustness of deep learning segmentation of cardiac substructures in noncontrast computed tomography for breast cancer radiotherapy. *Med. Phys.* **2021**, *48*, 7172–7188. [CrossRef] [PubMed]
13. Chaudhari, A.S.; Sandino, C.M.; Cole, E.K.; Larson, D.B.; Gold, G.E.; Vasanawala, S.S.; Lungren, M.P.; Hargreaves, B.A.; Langlotz, C.P. Prospective deployment of deep learning in MRI: A framework for important considerations, challenges, and recommendations for best practices. *J. Magn. Reson. Imaging* **2021**, *54*, 357–371. [CrossRef] [PubMed]

14. He, J.; Yang, L.; Liang, B.; Li, S.; Xu, C. Fetal cardiac ultrasound standard section detection model based on multitask learning and mixed attention mechanism. *Neurocomputing* **2024**, *579*, 127443. [CrossRef]

15. Ragab, M.G.; Abdulkader, S.J.; Muneer, A.; Alqushaibi, A.; Sumiea, E.H.; Qureshi, R.; Al-Selwi, S.M.; Alhussian, H. A Comprehensive Systematic Review of YOLO for Medical Object Detection (2018 to 2023). *IEEE Access* **2024**, *12*, 57815–57836. [CrossRef]

16. Kimura, B.J. Point-of-care cardiac ultrasound techniques in the physical examination: Better at the bedside. *Heart* **2017**, *103*, 987–994. [CrossRef] [PubMed]

17. González-Villà, S.; Oliver, A.; Valverde, S.; Wang, L.; Zwiggelaar, R.; Lladó, X. A review on brain structures segmentation in magnetic resonance imaging. *Artif. Intell. Med.* **2016**, *73*, 45–69. [CrossRef] [PubMed]

18. Gopalan, D.; Gibbs, J.S.R. From Early Morphometrics to Machine Learning—What Future for Cardiovascular Imaging of the Pulmonary Circulation? *Diagnostics* **2020**, *10*, 1004. [CrossRef]

19. Lang, R.M.; Addetia, K.; Narang, A.; Mor-Avi, V. 3-Dimensional echocardiography: Latest developments and future directions. *JACC Cardiovasc. Imaging* **2018**, *11*, 1854–1878. [CrossRef]

20. Minaee, S.; Boykov, Y.; Porikli, F.; Plaza, A.; Kehtarnavaz, N.; Terzopoulos, D. Image segmentation using deep learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 3523–3542. [CrossRef]

21. Li, X.; Li, M.; Yan, P.; Li, G.; Jiang, Y.; Luo, H.; Yin, S. Deep learning attention mechanism in medical image analysis: Basics and beyonds. *Int. J. Netw. Dyn. Intell.* **2023**, *2*, 93–116. [CrossRef]

22. Yuan, X.; Shi, J.; Gu, L. A review of deep learning methods for semantic segmentation of remote sensing imagery. *Expert Syst. Appl.* **2021**, *169*, 114417. [CrossRef]

23. Oza, P.; Sharma, P.; Patel, S.; Adedoyin, F.; Bruno, A. Image augmentation techniques for mammogram analysis. *J. Imaging* **2022**, *8*, 141. [CrossRef] [PubMed]

24. Gandhi, S.; Mosleh, W.; Shen, J.; Chow, C.M. Automation, machine learning, and artificial intelligence in echocardiography: A brave new world. *Echocardiography* **2018**, *35*, 1402–1418. [CrossRef] [PubMed]

25. Chen, C.; Qin, C.; Qiu, H.; Tarroni, G.; Duan, J.; Bai, W.; Rueckert, D. Deep learning for cardiac image segmentation: A review. *Front. Cardiovasc. Med.* **2020**, *7*, 25. [CrossRef] [PubMed]

26. Suri, J.S.; Liu, K.; Singh, S.; Laxminarayan, S.N.; Zeng, X.; Reden, L. Shape recovery algorithms using level sets in 2-D/3-D medical imagery: A state-of-the-art review. *IEEE Trans. Inf. Technol. Biomed.* **2002**, *6*, 8–28. [CrossRef] [PubMed]

27. Tajbakhsh, N.; Jeyaseelan, L.; Li, Q.; Chiang, J.N.; Wu, Z.; Ding, X. Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Med. Image Anal.* **2020**, *63*, 101693. [CrossRef] [PubMed]

28. Yang, L.; Zhang, Y.; Chen, J.; Zhang, S.; Chen, D.Z. Suggestive annotation: A deep active learning framework for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, 11–13 September 2017; Proceedings, Part III 20. Springer: Berlin/Heidelberg, Germany, 2017; pp. 399–407.

29. Greenwald, N.F.; Miller, G.; Moen, E.; Kong, A.; Kagel, A.; Dougherty, T.; Fullaway, C.C.; McIntosh, B.J.; Leow, K.X.; Schwartz, M.S. Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nat. Biotechnol.* **2022**, *40*, 555–565. [CrossRef] [PubMed]

30. Kulik, D. Synthetic Ultrasound Video Generation with Generative Adversarial Networks. Master's Thesis, Carleton University, Ottawa, ON, Canada, 2023.

31. Niu, S.; Liu, Y.; Wang, J.; Song, H. A decade survey of transfer learning (2010–2020). *IEEE Trans. Artif. Intell.* **2020**, *1*, 151–166. [CrossRef]

32. Zhu, Z.; Lin, K.; Jain, A.K.; Zhou, J. Transfer learning in deep reinforcement learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 13344–13362. [CrossRef]

33. Xiong, Z.; Xia, Q.; Hu, Z.; Huang, N.; Bian, C.; Zheng, Y.; Vesal, S.; Ravikumar, N.; Maier, A.; Yang, X. A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging. *Med. Image Anal.* **2021**, *67*, 101832. [CrossRef]

34. Ullah, Z.; Usman, M.; Jeon, M.; Gwak, J. Cascade multiscale residual attention cnns with adaptive roi for automatic brain tumor segmentation. *Inf. Sci.* **2022**, *608*, 1541–1556. [CrossRef]

35. Valindria, V.V.; Lavdas, I.; Bai, W.; Kamnitsas, K.; Aboagye, E.O.; Rockall, A.G.; Rueckert, D.; Glocker, B. Reverse classification accuracy: Predicting segmentation performance in the absence of ground truth. *IEEE Trans. Med. Imaging* **2017**, *36*, 1597–1606. [CrossRef]

36. Lal, S. TC-SegNet: Robust deep learning network for fully automatic two-chamber segmentation of two-dimensional echocardiography. *Multimed. Tools Appl.* **2024**, *83*, 6093–6111. [CrossRef] [PubMed]

37. Maani, F.; Ukaye, A.; Saadi, N.; Saeed, N.; Yaqub, M. UniLVSeg: Unified Left Ventricular Segmentation with Sparsely Annotated Echocardiogram Videos through Self-Supervised Temporal Masking and Weakly Supervised Training. *arXiv* **2023**, arXiv:2310.00454.

38. Moradi, S.; Oghli, M.G.; Alizadehasl, A.; Shiri, I.; Oveisi, N.; Oveisi, M.; Maleki, M.; Dhooge, J. MFP-Unet: A novel deep learning based approach for left ventricle segmentation in echocardiography. *Phys. Medica* **2019**, *67*, 58–69. [CrossRef]

39. Ouyang, D.; He, B.; Ghorbani, A.; Yuan, N.; Ebinger, J.; Langlotz, C.P.; Heidenreich, P.A.; Harrington, R.A.; Liang, D.H.; Ashley, E.A. Video-based AI for beat-to-beat assessment of cardiac function. *Nature* **2020**, *580*, 252–256. [CrossRef] [PubMed]

40. Yang, Y.; Shah, Z.; Jacob, A.J.; Hair, J.; Chitiboi, T.; Passerini, T.; Yerly, J.; Di Sopra, L.; Piccini, D.; Hosseini, Z. Deep learning-based left ventricular segmentation demonstrates improved performance on respiratory motion-resolved whole-heart reconstructions. *Front. Radiol.* **2023**, *3*, 1144004. [CrossRef]

41. Song, Y.; Ren, S.; Lu, Y.; Fu, X.; Wong, K.K. Deep learning-based automatic segmentation of images in cardiac radiography: A promising challenge. *Comput. Methods Programs Biomed.* **2022**, *220*, 106821. [CrossRef] [PubMed]

42. Petersen, E.; Feragen, A.; da Costa Zemsch, M.L.; Henriksen, A.; Wiese Christensen, O.E.; Ganz, M.; Initiative, A.s.D.N. Feature robustness and sex differences in medical imaging: A case study in MRI-based Alzheimer's disease detection. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Singapore, 18–22 September 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 88–98.

43. Zhang, Y.; Liao, Q.V.; Bellamy, R.K. Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, Barcelona, Spain, 27–30 January 2020; pp. 295–305.

44. Dergachyova, O.; Bouget, D.; Huaulmé, A.; Morandi, X.; Jannin, P. Automatic data-driven real-time segmentation and recognition of surgical workflow. *Int. J. Comput. Assist. Radiol. Surg.* **2016**, *11*, 1081–1089. [CrossRef]

45. Jacob, C.; Sanchez-Vazquez, A.; Ivory, C. Factors impacting clinicians' adoption of a clinical photo documentation app and its implications for clinical workflows and quality of care: Qualitative case study. *JMIR Mhealth Uhealth* **2020**, *8*, e20203. [CrossRef]

46. Xu, P.; Liu, H. Simultaneous reconstruction and segmentation of MRI image by manifold learning. In Proceedings of the 2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), Manchester, UK, 26 October–2 November 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–5.

47. Peng, P.; Lekadir, K.; Gooya, A.; Shao, L.; Petersen, S.E.; Frangi, A.F. A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging. *Magn. Reson. Mater. Phys. Biol. Med.* **2016**, *29*, 155–195. [CrossRef] [PubMed]

48. Lin, A.; Kolossváry, M.; Išgum, I.; Maurovich-Horvat, P.; Slomka, P.J.; Dey, D. Artificial intelligence: Improving the efficiency of cardiovascular imaging. *Expert Rev. Med. Devices* **2020**, *17*, 565–577. [CrossRef] [PubMed]

49. Peirlinck, M.; Costabal, F.S.; Yao, J.; Guccione, J.; Tripathy, S.; Wang, Y.; Ozturk, D.; Segars, P.; Morrison, T.; Levine, S. Precision medicine in human heart modeling: Perspectives, challenges, and opportunities. *Biomech. Model. Mechanobiol.* **2021**, *20*, 803–831. [CrossRef] [PubMed]

50. Madani, A.; Ong, J.R.; Tibrewal, A.; Mofrad, M.R. Deep echocardiography: Data-efficient supervised and semi-supervised deep learning towards automated diagnosis of cardiac disease. *NPJ Digit. Med.* **2018**, *1*, 59. [CrossRef] [PubMed]

51. Brown, M.E.; Mitchell, M.S. Ethical and unethical leadership: Exploring new avenues for future research. *Bus. Ethics Q.* **2010**, *20*, 583–616. [CrossRef]

52. de Siqueira, V.S.; Borges, M.M.; Furtado, R.G.; Dourado, C.N.; da Costa, R.M. Artificial intelligence applied to support medical decisions for the automatic analysis of echocardiogram images: A systematic review. *Artif. Intell. Med.* **2021**, *120*, 102165. [CrossRef] [PubMed]

53. Jocher, G.; Chaurasia, A.; Jing, Q. Ultralytics YOLO. 2023. Available online: https://ultralytics.com (accessed on 15 January 2024).

54. Wang, C.-Y.; Liao, H.-Y.M.; Wu, Y.-H.; Chen, P.-Y.; Hsieh, J.-W.; Yeh, I.-H. CSPNet: A new backbone that can enhance learning capability of CNN. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 390–391.

55. Holschneider, M.; Kronland-Martinet, R.; Morlet, J.; Tchamitchian, P. A real-time algorithm for signal analysis with the help of the wavelet transform. In Proceedings of the Wavelets: Time-Frequency Methods and Phase Space Proceedings of the International Conference, Marseille, France, 14–18 December 1987; Springer: Berlin/Heidelberg, Germany, 1990; pp. 286–297.

56. Giusti, A.; Cireşan, D.C.; Masci, J.; Gambardella, L.M.; Schmidhuber, J. Fast image scanning with deep max-pooling convolutional neural networks. In Proceedings of the 2013 IEEE International Conference on Image Processing, Melbourne, VIC, Australia, 15–18 September 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 4034–4038.

57. Sermanet, P.; Eigen, D.; Zhang, X.; Mathieu, M.; Fergus, R.; LeCun, Y. Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv* **2013**, arXiv:1312.6229.

58. Papandreou, G.; Kokkinos, I.; Savalle, P.-A. Modeling local and global deformations in deep learning: Epitomic convolution, multiple instance learning, and sliding window detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 390–399.

59. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.

60. Gonzalez-Jimenez, A.; Lionetti, S.; Gottfrois, P.; Gröger, F.; Pouly, M.; Navarini, A.A. Robust t-loss for medical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Vancouver, BC, Canada, 8–12 October 2023; Springer: Berlin/Heidelberg, Germany, 2023; pp. 714–724.

61. Milletari, F.; Navab, N.; Ahmadi, S.-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 565–571.

62. Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-IoU loss: Faster and better learning for bounding box regression. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 12993–13000.

63. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, 6–12 September 2014; Proceedings, Part V 13. Springer: Berlin/Heidelberg, Germany, 2014; pp. 740–755.

64. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef]

65. Chen, L.-C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587.