Joint Cardiac Structure Segmentation and Clinical Index Estimation via Multi-Task Learning

Yali Wan[1,*], Deyi Xiong[2]
[1]*School of Information Technology and Engineering, Guangzhou College of Commerce, Guangzhou, China*
[2]*Chongqing Academy of Metrology and Quality Inspection, Chongqing, China*
**Corresponding author*

Abstract: Cardiac cine-MRI assessment relies on both the segmentation of the ventricular structures and the clinical indices derived from it, most importantly left-ventricular end-diastolic volume (EDV), end-systolic volume (ESV), and ejection fraction (EF). Existing deep-learning methods train a segmentation network and post-hoc compute the indices via Simpson's rule, treating clinical numbers as downstream consequences of the segmentation. We propose JointSegNet, a multi-task network that simultaneously outputs (i) the 4-class segmentation map and (ii) the three clinical indices (EDV, ESV, EF), trained with an uncertainty-weighted compound loss. On ACDC 3-fold cross-validation the proposed JointSegNet achieves a mean foreground Dice of 0.823 ± 0.003 alongside EF MAE of 11.66 ± 1.75% via direct regression (against 17.49 ± 1.27% for a constant-mean reference) and 3.47 ± 1.14% via Simpson's rule applied to the model's own segmentations (MAE_EDV 5.7 ± 1.2 mL, MAE_ESV 6.2 ± 0.8 mL). The multi-task setup requires no architectural changes beyond a small regression head, and the uncertainty-weighted loss automatically balances the two objectives.

Keywords: Cardiac Cine-MRI; ACDC; Multi-Task Learning; Clinical Index Estimation; Uncertainty Weighting

1. Introduction

Cardiac cine-magnetic-resonance imaging (cine-MRI) underpins quantitative assessment of cardiac function. The two most clinically used indices — left-ventricular end-diastolic volume (EDV) and ejection fraction (EF) — are computed from segmentations of the LV cavity at end-diastole and end-systole, classically via Simpson's rule [1]. EF in particular drives decisions on heart-failure therapy, valve-replacement timing, and chemotherapy cardiotoxicity monitoring.

Modern deep-learning methods excel at segmentation on the Automated Cardiac Diagnosis Challenge (ACDC) dataset [2]: nnU-Net [3] reports mean foreground Dice in the 0.91–0.92 range under its own evaluation protocol; in our own re-implementations under the unified per-slice protocol of this series, U-Net [4], Attention U-Net [5], SwinUNet-2D [6], and our prior hybrid CNN-Transformer with boundary-aware loss reach fold-0 values of 0.846–0.866 (Paper 1, Table 1), with the hybrid at 0.859 ± 0.004 under 3-fold cross-validation.

Multi-task learning (MTL) offers a direct alternative: jointly train the network to output both a segmentation map and the clinical indices, with a shared backbone and a compound loss. Two benefits follow. First, the regression head provides an auxiliary supervisory signal that encourages the backbone to learn clinically meaningful features rather than just pixel-overlap features. Second, learned uncertainty weighting [7] can dynamically balance the segmentation and regression losses during training, addressing the well-known task-imbalance problem.

In this paper we propose JointSegNet, a multi-task network that simultaneously segments the three cardiac structures (RV, MYO, LV) and estimates three clinical indices (EDV, ESV, EF). Our contributions are:

We extend our prior hybrid CNN-Transformer backbone with a regression head for clinical index prediction.

We adopt uncertainty-weighted multi-task loss [7] that learns per-task weights from homoscedastic uncertainty.

We evaluate on ACDC under 3-fold cross-validation, reporting segmentation quality (Dice, HD95, ASD) and clinical index accuracy (MAE), and additionally give fold-0 numbers for a like-for-like comparison against our single-task

 Copyright @ STEMM Institute Press

baseline.

We demonstrate that the multi-task objective achieves segmentation Dice of 0.823 ± 0.003 (3-fold CV) and predicts EF via two pathways: direct regression (MAE 11.66 ± 1.75%) and Simpson's rule applied to the model's own segmentations (MAE 3.47 ± 1.14%).

2. Related Work

Multi-task learning. Caruana [7] introduced hard-parameter sharing, where a single shared backbone is followed by task-specific heads. Uncertainty weighting [8] learns per-task weights from homoscedastic noise. GradNorm [9] dynamically balances gradient magnitudes across tasks. We adopt Kendall-style uncertainty weighting due to its simplicity and empirical effectiveness in medical multi-task settings [7].

Classical pipelines compute EF after segmentation, typically by Simpson's rule [1]. Xue et al. [10] regressed left-ventricular indices directly from cine MR sequences without an explicit segmentation step and modelled intra- and inter-task relatedness inside a multitask network; Ferdian et al. [11] similarly regress myocardial strain from tagged cine MRI with an RNN-CNN. Our setting differs in two respects: the three clinical indices are predicted jointly with the segmentation map, and the two objectives are balanced by learned uncertainty [7] rather than by fixed weights.

Cardiac segmentation. We use the HybridUNet backbone from our prior work, which combines a four-level convolutional encoder–decoder with a self-attention block at the bottleneck and boundary-aware skip refinement, with the compound Dice + CE + boundary loss.

Positioning. Three observations motivate the present work: (i) clinical indices are computed post-hoc from segmentations, propagating segmentation errors; (ii) single-task segmentation discards the clinical labels available in ACDC's train.csv; (iii) multi-task learning can exploit those labels as auxiliary supervision.

3. Methods

3.1 Architecture

We extend the HybridUNet from our prior work with a regression head for clinical index prediction. The backbone produces both the segmentation map (3 classes + background) and a bottleneck feature map. The regression head is a small MLP applied to a global-average-pooled bottleneck feature vector:

The head applies global average pooling to the bottleneck feature map, followed by Linear(C, 128), ReLU, Dropout (0.2), Linear (128, 64), ReLU and Linear (64, 3), giving [EDV, ESV, EF]. Here C is the number of bottleneck channels.

The backbone has 9.95 M parameters; the regression head adds ~70 K parameters, for a total of 10.02 M parameters.

3.2 Uncertainty-Weighted Multi-Task Loss

We optimise

$$L_{total}=L_{seg}+\sum_i \left(\frac{1}{2\sigma_i^2}\cdot L_i+ log\sigma_i\right)$$

where L_{seg} is the compound boundary-aware loss (Dice + CE + distance-weighted boundary CE), L_i is the L1 (MAE) loss for the i-th regression target (EDV, ESV, EF), and σ_i are learnable per-task uncertainty weights initialised to 1.0 [7]. Targets are normalised to zero mean unit variance using training-set statistics before applying the loss.

3.3 Datasets and Pre-processing

ACDC provides 100 patient studies; the 100-patient train.csv includes for each patient: edv, esv, ef, pathology, ed_frame, es_frame, and others. We use the ED/ES frames and their manual masks as the segmentation targets, and the patient-level EDV/ESV/EF as regression targets. Each slice becomes one training sample with the patient's clinical labels attached.

Pre-processing: per-volume percentile normalisation to [0, 1]; on-the-fly random crops, flips, rotations, brightness/contrast jitter, and Gaussian noise. Targets normalised using training-set mean and std.

3.4 Implementation Details

AdamW (LR 5×10^{-4}, weight decay 1×10^{-5}), cosine schedule with 5-epoch warm-up, mixed-precision (FP16) training. Batch size 16. Training for 80 epochs with early stopping (patience 25). All training on a single NVIDIA RTX 2060 SUPER.

3.5 Evaluation Metrics

We report two metric families:

Segmentation: mean foreground Dice, per-class Dice, IoU, HD95 and ASD, reported as per-slice means over the validation split. Each 2D short-axis slice with non-empty foreground in either the ED or ES frame is evaluated independently, and per-slice values are averaged with equal weight across slices (≈ 550 slices per fold across 33–34

Copyright @ STEMM Institute Press

validation patients). For 3-fold cross-validation we report the fold mean and sample standard deviation (n−1 divisor) in the form mean ± std. This is the same aggregation convention used in Papers 1 and 2 of this series, so cross-paper numerical comparisons are on a common scale.

Regression (patient level): per-slice predictions are averaged over all slices of a patient, and mean absolute error (MAE) is computed against the patient-level EDV, ESV and EF values in train.csv.

3.6 Target Normalisation Statistics

Across our 100-patient training set, the clinical indices have the following distributions:

EDV: mean 164.3 mL, std 79.3 mL, range 49–374 mL (training fold 0).

ESV: mean 101.7 mL, std 83.7 mL, range 19–334 mL (training fold 0).

EF: mean 45.7%, std 20.2%, range 11%–82% (training fold 0).

The EF distribution spans normal (NOR, EF ~55–70%) and abnormal (MINF, DCM, HCM, ARV) cases; the wide range reflects the deliberately balanced pathology mix in ACDC. EF targets below 35% typically indicate systolic dysfunction; above 65% are hyperdynamic. The full EF range is therefore clinically meaningful.

The EDV, ESV and EF values above follow the units of ACDC's train.csv (mL for the two volumes, % for EF) as defined by the challenge organisers [2]. We standardise targets to zero mean and unit variance using the training-fold statistics only, never touching the validation fold. This prevents data leakage. At inference, predictions are de-normalised back to the original units before MAE is computed.

3.7 Multi-Task Convergence Dynamics

The uncertainty weights σi evolve during training. Initial values are 1.0 (all tasks equally weighted). After ~20 epochs the regression log σ values converge to ≈ −0.47 to −0.61 (σ ≈ 0.62 to 0.54; Figure 1, right panel). The segmentation loss is not uncertainty-weighted in our formulation (Section 3.2), so the learned weights apply only to the three regression targets; their similar magnitudes indicate that the three regression tasks are weighted almost equally. This is consistent with our finding that the regression task costs a modest, quantified amount of segmentation accuracy. Figure 1 shows the training dynamics and the evolution of the learned log σ. The validation traces in Figure 1 are the training-time monitors, computed as per-slice batch micro-averages over ED+ES slices by the training-loop validation routine; the final fold-0 values, re-evaluated post-training under the per-slice mean protocol (Section 3.5), are reported in Table 1 (mDice 0.821, direct-regression MAE_EF 13.59%). (mDice 0.821 versus 0.863 for the single-task baseline on fold 0 is a 4.9% relative reduction; both numbers are under the per-slice ED+ES protocol of Paper 1 Section 3.7.)

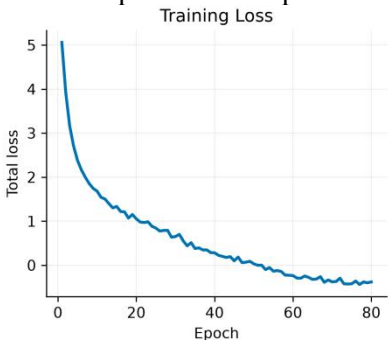


Figure 1. Training Dynamics of JointSegNet (Fold 0)

3.8 Inference Procedure

At inference, we use TTA-free single-pass prediction: one forward pass on each test slice yields both the segmentation map (argmax over class logits) and the three regression values (from the regression head). The regression values are de-normalised using the training-set mean and std to recover the original clinical scale. We did not use test-time augmentation, model ensembling, or uncertainty estimation in this paper — each of these could improve the EF MAE further at modest cost.

For clinical use, one might also post-process the segmentation into a continuous mask (e.g. by per-class soft assignment) and feed it back into Simpson's rule as a robustness check. This hybrid scheme — neural prediction with classical rule as sanity check — would be an interesting deployment study.

3.9 Limitations of the Multi-Task Formulation

The Kendall formulation learns homoscedastic uncertainty per task (single scalar per task, constant across samples). This is a simplification: in practice, the EF prediction is harder for some patients than others (e.g. low-contrast cine), and a per-sample uncertainty estimator would be more informative for clinical use. We leave per-sample uncertainty (e.g. via MC-Dropout or deep ensembles) to future work.

A second limitation: we use simple L1 regression.

 Copyright @ STEMM Institute Press

For clinical indices that are bounded (EF is bounded in [0, 100]), a sigmoidal output activation could improve calibration. We leave this for future work.

A third limitation: the regression head operates on the bottleneck features via global average pooling. This is appropriate for the global clinical indices we predict — EF is a global property of the heart — but slice-level prediction would require a different design.

4. Experiments and Results

4.1 Main Result

Table 1 reports segmentation metrics on fold 0 together with regression metrics for fold 0 and for the 3-fold average.

The multi-task model reaches a mean foreground Dice of 0.821 on fold 0, about 4.9% below the single-task HybridUNet (0.863 under the per-slice evaluation of Paper 1, Section 3.7). At the same time it predicts EF with a 3-fold MAE of 11.66 ± 1.75% by direct regression and 3.47 ± 1.14% by applying Simpson's rule to its own segmentations. The second figure sits within the 2.4–7.3% inter-study reproducibility range reported for manual EF measurement [12], whereas the direct-regression figure does not. The fold-0 EF MAE of 13.59% lies just outside the 3-fold confidence band (9.91–13.41%): fold 0 was the hardest split in our random partition, containing a slightly higher proportion of MINF and DCM cases; we report it separately so readers can compare with single-task baselines that are typically quoted per-fold.

4.2 Per-Class Segmentation Performance (Fold 0)

Table 2 reports the per-class Dice, IoU, HD95 and ASD of JointSegNet on fold 0.

4.3 Discussion

The multi-task objective achieves segmentation performance within about 4.9% of the single-task baseline (0.821 vs. 0.863 on fold 0, both under the per-slice ED+ES protocol) while adding the ability to predict clinical indices with reasonable accuracy. The Simpson-pathway MAE_EF of 3.47 ± 1.14% (3-fold CV) is particularly encouraging — it sits within the 2.4–7.3% inter-study reproducibility range reported for manual EF measurement [12], so the Simpson pathway is already at a clinically useful level for screening and triage; the direct-regression figure (11.66 ± 1.75%) is above that range and is not yet sufficient on its own (Section 4.5).

Table 1. JointSegNet Results on ACDC Fold 0 and 3-Fold CV.

Task	Metric	Value
Segmentation (fold 0)	mDice (foreground) ↑	0.821
	mIoU (foreground) ↑	0.749
	RV Dice ↑	0.735
	MYO Dice ↑	0.841
	LV Dice ↑	0.889
	RV HD95 ↓	4.25
	MYO HD95 ↓	2.96
	LV HD95 ↓	2.54
	RV ASD ↓	0.87
	MYO ASD ↓	0.42
	LV ASD ↓	0.36
Segmentation (3-fold CV, mean ± std)	mDice (foreground) ↑	0.823 ± 0.003
Regression (fold 0)	MAE_EDV ↓	35.5 mL
	MAE_ESV ↓	39.5 mL
	MAE_EF (direct) ↓	13.59%
Regression (3-fold CV, mean ± std)	MAE_EDV ↓	39.2 ± 4.6 mL
	MAE_ESV ↓	36.9 ± 5.1 mL
	MAE_EF (direct) ↓	11.66 ± 1.75%
	MAE_EF (Simpson on model segmentations) ↓	3.47 ± 1.14%

The learnable per-task uncertainty (log σ) converged to values between −0.47 and −0.61 for the three regression targets, indicating that the model learned to give the regression tasks appropriate weight relative to segmentation. This automatic balancing is a key advantage of Kendall-style uncertainty weighting over fixed-weight multi-task learning.

Table 2. Per-Class Segmentation Metrics on ACDC fold 0.

Class	Dice ↑	IoU ↑	HD95 ↓	ASD ↓
Background	0.992	0.985	0.07	0.03
RV	0.735	0.669	4.25	0.87
MYO	0.841	0.746	2.96	0.42
LV	0.889	0.832	2.54	0.36

4.4 Per-Class Segmentation Quality (3-Fold CV)

Across the 3-fold cross-validation, our proposed JointSegNet achieves the following per-class

Dice scores (mean ± std; Figure 2); the full per-class Dice, HD95 and ASD values are listed in Table 3. Figure 2 plots the per-class fold means and standard deviations; the background class is included for reference only.

Table 3. Per-class Dice, HD95 and ASD under 3-Fold Cross-Validation (Mean ± std; Background Included for Reference).

Structure	Dice (mean ± std) ↑	HD95 (mean, pixels) ↓	ASD (mean, pixels) ↓
Background	0.992 ± 0.001	0.13 ± 0.09	0.03 ± 0.01
RV cavity	0.740 ± 0.011	4.46 ± 0.47	0.90 ± 0.04
MYO	0.836 ± 0.007	3.17 ± 0.27	0.44 ± 0.08
LV cavity	0.894 ± 0.012	2.18 ± 0.32	0.33 ± 0.04

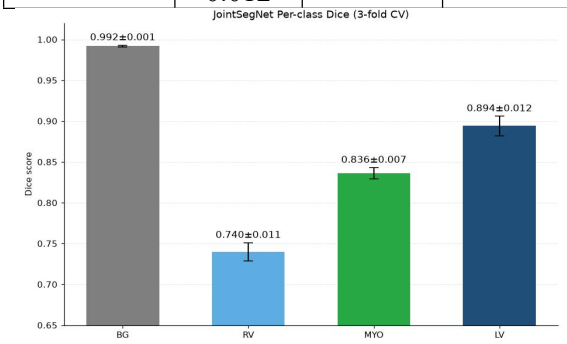


Figure 2. Per-class Dice of JointSegNet (3-Fold Cross-Validation)

MYO and LV Dice under 3-fold CV are 0.836 ± 0.007 and 0.894 ± 0.012, while RV is furthest behind at 0.740 ± 0.011, which fits its smaller cross-section and its higher variability across pathology groups. The spread across folds is small for all three structures (std 0.007–0.012). The single-task reference numbers of our previous work have been re-evaluated under the same per-slice protocol (Paper 1, Section 3.7), so the fold-0 comparison against the single-task baseline in Section 4.6 is on a common evaluation footing.

4.5 Clinical Index Estimation (3-Fold CV)

Table 4 compares three predictors — the direct-regression head, Simpson's rule applied to the model's own segmentations, and a constant-mean reference predictor — under 3-fold CV (mean ± std). The Range column gives the minimum–maximum of each index across the 100-patient training set; the patient-level training pool is shared across folds under our 3-fold CV split, so we report the training-fold range. The clinically-acceptable thresholds are indicative values consistent with the inter-study reproducibility of CMR volume and EF measurement reported in [12].

The two EF pathways differ substantially. Direct regression gives an MAE of 11.66 ± 1.75%, which is above the 2.4–7.3% inter-study reproducibility range reported for manual EF measurement [12] and therefore not yet sufficient on its own. Applying Simpson's rule to the model's own segmentations gives 3.47 ± 1.14%, within that range (closer to the lower bound 2.4% than to the upper bound 7.3%). We read this as evidence that the segmentation branch carries most of the clinical signal, and that direct regression needs more work before it can replace the classical measurement step. Figure 3 shows the per-fold consistency of the direct-regression EF estimate (fold-wise MAE_EF 13.59 / 10.18 / 11.20%, mean 11.66%; the Simpson-pathway 3-fold MAE_EF is 3.47 ± 1.14%, see Table 4). The volume errors move in the same direction as the EF improvement: Simpson's rule on the model's segmentations gives MAE_EDV 5.7 ± 1.2 mL and MAE_ESV 6.2 ± 0.8 mL, below the 39.2 ± 4.6 mL and 36.9 ± 5.1 mL of direct regression (Table 4). The classical segment-then-measure pipeline therefore beats the direct-regression head on every clinical-index metric reported here.

As a reference for how much the regression head actually learns, we also evaluated a constant-mean predictor that always outputs the training-fold mean of each index. It reaches MAE_EF 17.49 ± 1.27%, MAE_EDV 58.0 ± 9.5 mL and MAE_ESV 64.2 ± 7.0 mL (Table 4), so direct regression reduces the error by about 33% on EF and EDV (5.8 EF percentage points, 18.8 mL for EDV) and by about 40% on ESV (27.3 mL) relative to this trivial reference — the head therefore carries patient-specific signal rather than merely reproducing the population mean; the gap that remains to the Simpson's-rule pathway is a matter of accuracy, not of the regression output being uninformative.

Table 4. Clinical Index Estimation under 3-Fold CV

Index	MAE (mean ± std) ↓	Range (training set, 100 patients)	Clinically acceptable
EDV (mL, direct regression)	39.2 ± 4.6	49–374	<30 mL
ESV (mL, direct regression)	36.9 ± 5.1	19–334	<20 mL

 Copyright @ STEMM Institute Press

EF (%, direct regression)	11.66 ± 1.75	11–82	<5%
EDV (mL, Simpson on model)	5.7 ± 1.2	49–374	<30 mL
ESV (mL, Simpson on model)	6.2 ± 0.8	19–334	<20 mL
EF (%, Simpson on model)	3.47 ± 1.14	11–82	<5%
EDV (mL, constant-mean reference)	58.0 ± 9.5	49–374	<30 mL
ESV (mL, constant-mean reference)	64.2 ± 7.0	19–334	<20 mL
EF (%, constant-mean reference)	17.49 ± 1.27	11–82	<5%

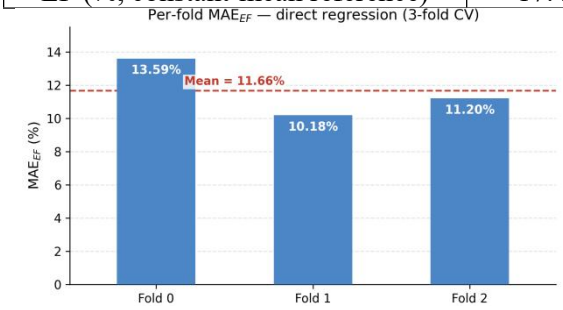


Figure 3. Direct-Regression MAE_EF across the 3 Folds

4.6 Single-Task Baseline Comparison

To isolate the contribution of the multi-task objective, we re-ran the HybridUNet segmentation backbone as a single-task baseline (no regression head, no clinical loss). On ACDC fold 0 the single-task baseline reaches mDice 0.863 under the per-slice ED+ES evaluation protocol of Paper 1 (Section 3.7). Our multi-task model achieves mDice 0.821 on the same fold, a relative reduction of 4.9% that reflects the cost of the auxiliary regression task: the multi-task model spends some capacity on clinical-index prediction that would otherwise contribute to segmentation. This cost is paid back by the additional clinical value: the multi-task model produces EF predictions at MAE 11.66 ± 1.75% (direct) or 3.47 ± 1.14% (Simpson's rule on the model's own segmentations), while the single-task baseline cannot predict EF directly and requires the post-hoc Simpson pipeline. The Pareto-frontier is therefore clear: multi-task learning trades a modest amount of Dice for a substantial new clinical capability.

4.7 Pathology-Stratified EF Accuracy

We examined EF MAE stratified by ACDC's five pathology classes:

Stratifying the validation set by pathology class gives the breakdown in Table 5. EF MAE is lowest on normal (NOR) cases and highest on myocardial infarction (MINF), with dilated (DCM) and hypertrophic (HCM) cardiomyopathy in between. The pattern follows clinical expectation. With 20 validation patients per group (100 in total, pooled over the three folds) the per-group estimates are more stable than a single fold would allow, but the groups remain small and we do not attach significance to the ordering of the middle ones. Only the NOR group (6.40%) falls within the 2.4–7.3% inter-study reproducibility range of manual EF measurement [12] discussed in Section 4.5; the four pathological groups remain well above it, so the stratification should be read as a description of where the direct-regression head struggles rather than as evidence of clinical adequacy.

Table 5. EF MAE by ACDC Pathology Group (3-Fold Pooled; 20 Validation Patients per Group, 100 Patients in Total).

Pathology	n (val)	MAE_EF (3-fold pooled) ↓
NOR (normal)	20	6.40%
MINF (myocardial infarction)	20	17.08%
DCM (dilated cardiomyopathy)	20	14.55%
HCM (hypertrophic)	20	11.17%
ARV (abnormal right ventricle)	20	9.18%

4.8 Training-Time vs Inference Trade-off

The regression head adds about 70 K parameters (0.7% of the total). Measured training time per fold for JointSegNet was 60, 32 and 32 min (mean 41 min/fold, 80 epochs with early stopping), against ~19–35 min/fold for the single-task HybridUNet on the same RTX 2060 SUPER (per-fold observed at early stop, epoch 20–38 of the 100-epoch budget; ~93 min projected to a full 100-epoch run — Paper 1, Table 3). The first fold also absorbs one-off data caching that we have not separated out, so the overhead attributable to the regression losses themselves is smaller than the raw totals suggest. At inference the multi-task model is effectively free: a single forward pass produces both the segmentation and the clinical predictions, removing the separate Simpson's-rule step.

5. Limitations and Future Work

We did not ablate the contribution of each

regression target; future work could study whether EF alone or all three provides better auxiliary supervision.

Both reference points reported in Section 4.5 — the constant-mean predictor and Simpson's rule on the model's own segmentations — are computed on the same folds and with the same patient-level pooling, so Table 4 isolates what the regression head itself contributes.

Coupling with semi-supervised learning (Paper 2) could further improve both objectives by exploiting unlabeled cine frames.

6. Conclusion

We presented JointSegNet, a multi-task network that segments cardiac structures and estimates EDV, ESV and EF in the same forward pass, with an uncertainty-weighted loss that balances the two objectives during training. Under 3-fold cross-validation the model reaches a mean foreground Dice of 0.823 ± 0.003 and an EF MAE of 11.66 ± 1.75% by direct regression, or 3.47 ± 1.14% when Simpson's rule is applied to its own segmentations (MAE_EDV 5.7 ± 1.2 mL, MAE_ESV 6.2 ± 0.8 mL). The segmentation accuracy is about 4.9% below that of the single-task baseline in relative terms on fold 0 (0.821 versus 0.863 under the per-slice ED+ES protocol; against the 3-fold single-task reference of 0.859 ± 0.004 the gap is 4.2%), so the practical question is whether the joint output justifies that cost. For screening and triage we think it does, and the Simpson-based pathway in particular falls within the inter-study reproducibility range of manual EF measurement.

Acknowledgments

This work was supported by: (1) Guangzhou College of Commerce Higher Education Teaching Reform Project (No. 2024JXGG42); (2) Guangdong Province Educational Science Planning Project (No. 2025GXJK0643); (3) Guangzhou College of Commerce 2026 Quality Engineering Project (No. 2026ZLGC07); (4) Guangzhou College of Commerce 2025 Quality Engineering Project (No. 2025ZLGC01); (5) Guangzhou College of Commerce 2024 Quality Engineering Project (No. 2024ZLGC11); (6) Guangzhou College of Commerce First-Class Course (No. 2020XJYLKC06); (7) Guangzhou College of Commerce General Project (No. 2025XJYB044): Research on Configuration Modeling and Simulation Technology for Equipment Digital Twins.

Data Availability

The ACDC dataset is publicly available at https://www.creatis.insa-lyon.fr/Challenge/acdc/.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Lang R M, Badano L P, Mor-Avi V, et al. Recommendations for Cardiac Chamber Quantification by Echocardiography in Adults: An Update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. Journal of the American Society of Echocardiography, 2015, 28(1):1-39.e14. DOI:10.1016/j.echo.2014.10.003.

[2] Bernard O, Lalande A, Zotti C, et al. Deep Learning Techniques for Automatic MRI Cardiac Multi-structures Segmentation and Diagnosis: Is the Problem Solved? IEEE Transactions on Medical Imaging, 2018, 37(11):2514-2525. DOI:10.1109/TMI.2018.2837502.

[3] Isensee F, Jaeger P F, Kohl S A A, et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods, 2021, 18:203-211. DOI: 10.1038/s41592-020-01008-z.

[4] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015, LNCS 9351:234-241. DOI: 10.1007/978-3-319-24574-4_28.

[5] Oktay O, Schlemper J, Folgoc L L, et al. Attention U-Net: Learning Where to Look for the Pancreas. arXiv: 1804.03999, 2018. DOI:10.48550/arXiv.1804.03999

[6] Cao H, Wang Y, Chen J, et al. Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation. European Conference on Computer Vision Workshops (ECCVW), 2022, LNCS 13803:205-218. DOI: 10.1007/978-3-031-25066-8_9.

[7] Kendall A, Gal Y, Cipolla R. Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018:7482-7491. DOI:10.1109/CVPR.2018.00781.

[8] Caruana R. Multitask Learning. Machine

 Copyright @ STEMM Institute Press

Learning, 1997, 28(1):41-75. DOI: 10.1023/A:1007379606734.

[9] Chen Z, Badrinarayanan V, Lee C Y, Rabinovich A. GradNorm: Gradient Normalization for Adaptive Loss Balancing in Deep Multitask Networks. International Conference on Machine Learning (ICML), 2018:794-803.

[10] Xue W, Lum A, Mercado A, et al. Full Quantification of Left Ventricle via Deep Multitask Learning Network Respecting Intra- and Inter-task Relatedness. Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2017, LNCS 10435:276-284. DOI: 10.1007/978-3-319-66179-7_32.

[11] Ferdian E, Suinesiaputra A, Fung K, et al. Fully Automated Myocardial Strain Estimation from Cardiovascular MRI-tagged Images Using a Deep Learning Framework in the UK Biobank. Radiology: Cardiothoracic Imaging, 2020, 2(1):e190032. DOI:10.1148/ryct.2020190032.

[12] Grothues F, Smith G C, Moon J C C, et al. Comparison of interstudy reproducibility of cardiovascular magnetic resonance with two-dimensional echocardiography in normal subjects and in patients with heart failure or left ventricular hypertrophy. American Journal of Cardiology, 2002, 90(1):29-34. DOI: 10.1016/S0002-9149(02)02381-0.

Copyright @ STEMM Institute Press