Appendix for Q-PART

Abstract. In this appendix, we provides supplementary materials and detailed explanations for the main paper. Appendix A presents a comprehensive mathematical proof of Theorem 1. Appendix B elaborates on our algorithm implementation, providing detailed descriptions of both training and test-time adaptation phases of the proposed Q-PART framework. Appendix C contains additional experimental details, including formal definitions of evaluation metrics, detailed implementation of baseline methods, and complete ablation study results across all age cohorts.

A. Proof of Theorem 1

Proof: We begin by computing the expected variance loss $\mathbb{E}[L_{\text{var}}]$. The variance loss is defined as

$$L_{\text{var}} = \frac{1}{K}\sum_{k=1}^{K}(\hat{y}_k - \bar{y})^2. \tag{11}$$

Taking the expected value

$$\mathbb{E}[L_{\text{var}}] = \frac{1}{K}\sum_{k=1}^{K}\mathbb{E}\left[(\hat{y}_k - \bar{y})^2\right]. \tag{12}$$

Since $\mathbb{E}\left[\hat{y}_k - \bar{y}\right] = \mathbb{E}[\hat{y}_k] - \mathbb{E}[\bar{y}] = 0$, each term $\mathbb{E}\left[(\hat{y}_k - \bar{y})^2\right]$ can be expressed using the variance formula:

$$\mathbb{E}\left[(\hat{y}_k - \bar{y})^2\right] = \text{Var}(\hat{y}_k - \bar{y}). \tag{13}$$

Then, noted that:

$$\text{Var}(\hat{y}_k - \bar{y}) = \text{Var}(\hat{y}_k) + \text{Var}(\bar{y}) - 2\,\text{Cov}(\hat{y}_k, \bar{y}). \tag{14}$$

Set the variance of the predictions from the augmented samples $\hat{y}_k$ as σ^2,

$$\text{Var}(\hat{y}_k) = \sigma^2. \tag{15}$$

Under Assumption 2 (Augmentation Independence), the predictions $\hat{y}_k$ are independent for different k. Compute the $\text{Var}(\bar{y})$

$$\begin{aligned}\text{Var}(\bar{y}) = \text{Var}\left(\frac{1}{K}\sum_{k=1}^{K}\hat{y}_k\right) &= \frac{1}{K^2}\sum_{k=1}^{K}\text{Var}(\hat{y}_k) \\ &= \frac{1}{K^2}\cdot K\cdot\sigma^2 = \frac{\sigma^2}{K}.\end{aligned} \tag{16}$$

Now, compute the covariance $\text{Cov}(\hat{y}_k, \bar{y})$:

$$\begin{aligned}\text{Cov}(\hat{y}_k, \bar{y}) &= \text{Cov}\left(\hat{y}_k, \frac{1}{K}\sum_{j=1}^{K}\hat{y}_j\right) \\ &= \frac{1}{K}\text{Cov}(\hat{y}_k, \hat{y}_k) + \frac{1}{K}\sum_{j\neq k}\text{Cov}(\hat{y}_k, \hat{y}_j).\end{aligned} \tag{17}$$

Since $\text{Cov}(\hat{y}_k, \hat{y}_k) = \text{Var}(\hat{y}_k) = \sigma^2$, and under Assumption 2, for $j \neq k$, $\text{Cov}(\hat{y}_k, \hat{y}_j) = 0$, we have:

$$\text{Cov}(\hat{y}_k, \bar{y}) = \frac{1}{K}\sigma^2. \tag{18}$$

Now, compute $\text{Var}(\hat{y}_k - \bar{y})$:

$$\begin{aligned}\text{Var}(\hat{y}_k - \bar{y}) &= \sigma^2 + \frac{\sigma^2}{K} - 2\left(\frac{\sigma^2}{K}\right) &(19)\\ &= \sigma^2\left(1 + \frac{1}{K} - \frac{2}{K}\right) &(20)\\ &= \sigma^2\left(1 - \frac{1}{K}\right). &(21)\end{aligned}$$

Thus,

$$\mathbb{E}[L_{\text{var}}] = \frac{1}{K}\cdot K\cdot\sigma^2\left(1 - \frac{1}{K}\right) = \sigma^2\left(1 - \frac{1}{K}\right). \tag{22}$$

Simplify:

$$\mathbb{E}[L_{\text{var}}] = \sigma^2\left(\frac{K-1}{K}\right). \tag{23}$$

The expected regression loss is:

$$\mathbb{E}[L_{\text{reg}}] = \mathbb{E}\left[(\hat{y} - y)^2\right] = \mathbb{E}\left[(\bar{y} - y)^2\right]. \tag{24}$$

Since we can obtain the prediction $\hat{y}$ for the test sample $\boldsymbol{x}_{\text{test}}$ with the average over all augmented versions:

$$\hat{y} = \bar{y} = \frac{1}{K}\sum_{k=1}^{K}\hat{y}_k. \tag{25}$$

Under Assumption 1 (Unbiased Augmentation):

$$\mathbb{E}_{\mathcal{T}_{\text{aug}}}[\hat{y}_k] = y, \quad \forall k. \tag{26}$$

the expected value of $\bar{y}$ is equal to y,

$$\mathbb{E}[\bar{y}] = \mathbb{E}\left[\frac{1}{K}\sum_{k=1}^{K}\hat{y}_k\right] = \frac{1}{K}\sum_{k=1}^{K}\mathbb{E}[\hat{y}_k] = y. \tag{27}$$

Therefore, we have:

$$\mathbb{E}[L_{\text{reg}}] = \mathbb{E}\left[(\bar{y} - \mathbb{E}[\bar{y}])^2\right] = \text{Var}(\bar{y}) = \frac{\sigma^2}{K}. \tag{28}$$

So far, we have:

$$\begin{aligned}\mathbb{E}[L_{\text{reg}}] &= \frac{\sigma^2}{K}, &(29)\\ \mathbb{E}[L_{\text{var}}] &= \sigma^2\left(\frac{K-1}{K}\right). &(30)\end{aligned}$$

Substitute Equation 29 into Equation 30:

$$\mathbb{E}[L_{\text{reg}}] = \frac{\sigma^2}{K} = \frac{1}{K} \cdot \mathbb{E}[L_{\text{var}}] \left(\frac{K}{K-1} \right) = \frac{\mathbb{E}[L_{\text{var}}]}{K-1}. \quad (31)$$

Recall that for $K \geq 2$, the following inequality holds:

$$\frac{1}{K-1} \leq \frac{2}{K}. \quad (32)$$

Therefore, we have:

$$\mathbb{E}[L_{\text{reg}}] = \frac{\mathbb{E}[L_{\text{var}}]}{K-1} \leq \frac{2\,\mathbb{E}[L_{\text{var}}]}{K}. \quad (33)$$

□

Under Assumptions 1 and 2, we have established that minimizing the variance loss $\mathbb{E}[L_{\text{var}}]$ effectively reduces the expected regression error $\mathbb{E}[L_{\text{reg}}]$, with the bound improving as the number of augmentations K increases. This theoretical result justifies the effectiveness of our variance minimization strategy in test-time training.

B. Algorithm

As shown in Algorithm 1, our approach consists of two phases. During training, we first extract features through an encoder and decompose them into periodic and aperiodic components. The periodic component is modeled using sinusoidal functions with learned parameters (frequency, phase, bias, and velocity), while the aperiodic component is captured through a continuous-time framework using cubic spline interpolation. During test-time adaptation, we generate K augmented views of each test sample and employ a differential learning rate strategy: applying smaller learning rates to periodic components' batch normalization parameters to maintain stable cardiac patterns, while using larger learning rates for aperiodic components to enable flexible patient-specific adaptation.

Algorithm 1 Test-time Training for Quasi-Period Network

Require: Training data $\mathcal{D}_{train} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{N}$, test sample $\boldsymbol{x}_{test}$
Ensure: Adapted model parameters θ

// Training Phase
for each training iteration **do**
  $\boldsymbol{z} \leftarrow \text{Enc}(\boldsymbol{x})$ ▷ Initial feature extraction
  // Periodic Component
  $f, \phi, b, v \leftarrow p(\boldsymbol{z})$ ▷ Extract periodic parameters
  $\hat{\boldsymbol{z}}^{period} \leftarrow \cos(2\pi(ft - \phi)) + \sin(2\pi(ft - \phi)) + vt + b$
  // Aperiodic Component
  $\boldsymbol{z}' \leftarrow \boldsymbol{z} - \hat{\boldsymbol{z}}^{period}$ ▷ Residual features
  $\boldsymbol{V}(t) \leftarrow \text{CubicSpline}(\boldsymbol{z}'_t)$ ▷ Continuous path
  $\hat{\boldsymbol{z}}^{aperiod} \leftarrow \boldsymbol{z}'_0 + \int_0^T f_\theta(\hat{\boldsymbol{z}}_t^{aperiod}, t) d\boldsymbol{V}(t)$
  // Loss Computation
  Calculate training loss $\mathcal{L}_{total}$ using Eq. 6
  Update model parameters
end for
// Test-time Training Phase
for each test sample $\boldsymbol{x}_{test}$ **do**
  // Generate Augmented Samples
  $\{\boldsymbol{x}_{test}^k\}_{k=1}^{K} \leftarrow \mathcal{T}_{aug}(\boldsymbol{x}_{test})$
  for each adaptation iteration **do**
    for $k = 1$ to K **do**
      Forward pass $\boldsymbol{x}_{test}^k$ through network
      Compute predictions $\hat{y}_k$
    end for
    // Compute Test-time Losses Calculate test-time loss $\mathcal{L}_{test}$ using Eq. 9
    // Differential Adaptation
    Update periodic BN parameters with small learning rate
    Update aperiodic BN parameters with large learning rate
  end for
end for

C. Experiment

C.1. Evaluation Metrics

MAE measures the average absolute differences between predicted and ground truth LVEF values. RMSE emphasizes larger prediction errors by computing the square root of the mean squared differences. MAPE calculates the percentage error relative to the ground truth value, providing a scale-independent assessment of model performance. These metrics are formally defined as:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|, \quad (34)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}, \quad (35)$$

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^{n} |\frac{y_i - \hat{y}_i}{y_i}|, \quad (36)$$

where y_i and $\hat{y}_i$ denote the ground truth and predicted LVEF values respectively, and n is the number of test samples.

As for AUROC, we follow the clinical guidelines and set four critical LVEF thresholds: 35%, 40%, 45%, and 50%. These thresholds are clinically significant as they correspond to different levels of cardiac dysfunction. For each threshold, we compute the AUROC score by treating LVEF prediction as a binary classification problem, where values below the threshold indicate potential cardiac dysfunction. The mean AUROC across all thresholds provides a compre-

Table 4. **All Ablation Study Results of Key Components.** Analysis of three key components: QP-Net (Quasi-Period Network), LR (Learning Rate Strategy), and VM (Variance Minimization).

QP-Net	LR	VM	Pre-School			School Age			Adolescence		
			MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE
			8.129	0.1681	10.09	8.267	0.1454	9.999	8.222	0.1498	10.2
✓			7.949	0.1666	9.909	7.400	0.1368	8.959	7.524	0.1451	9.488
✓	✓		7.909	0.1656	9.878	7.480	0.1331	9.031	7.501	0.1403	9.383
		✓	7.842	0.1640	9.714	6.683	0.1260	8.767	7.146	0.1390	9.014
✓		✓	7.283	0.1619	9.307	6.708	0.1243	8.437	6.988	0.1344	9.002
✓	✓	✓	7.235	0.1611	9.290	6.706	0.1244	8.432	6.980	0.1344	8.950

hensive assessment of the model's ability to identify clinically relevant cardiac conditions.

C.1.1 Baseline Implementation

Segmentation-based methods follow a two-step process: first segmenting the left ventricle in each frame of the echocardiogram video, then calculating LVEF based on the end-diastolic volume (EDV) and end-systolic volume (ESV). Specifically, after obtaining segmentation masks for all frames, the frames with maximum and minimum left ventricular volumes are identified as end-diastolic and end-systolic frames, respectively. LVEF is then calculated using the following formula:

$$\mathrm{LVEF} = \frac{\mathrm{EDV} - \mathrm{ESV}}{\mathrm{EDV}} \times 100\%, \tag{37}$$

where EDV and ESV are computed from the segmentation masks using standard clinical volume estimation methods.

Vision-language models approach LVEF prediction as a cross-modal similarity task. These methods first encode the echocardiogram video into visual tokens through a vision encoder. Simultaneously, they construct a series of language tokens representing different LVEF values (e.g., "The left ventricular ejection fraction is X percent", where X ranges from 0 to 100). The predicted LVEF is determined by finding the language token that exhibits the highest similarity score with the visual tokens in the joint embedding space.

C.1.2 All results for Table 3

We show the all results from three cohorts in Table 4.