rPPG-VQA: A Video Quality Assessment Framework for Unsupervised rPPG Training

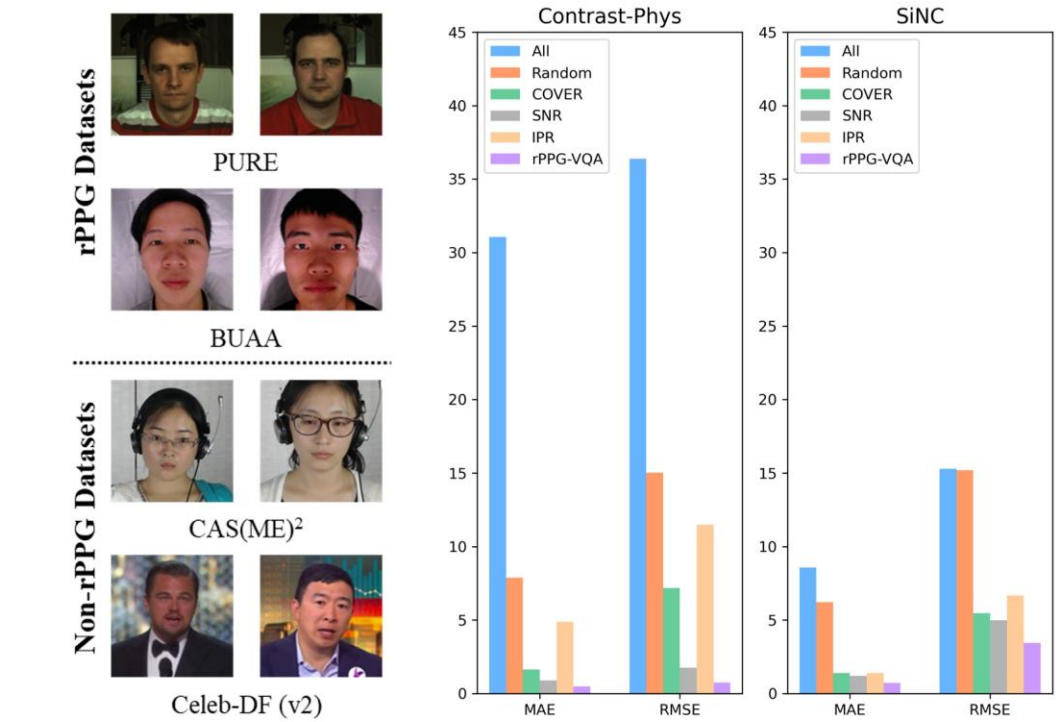
Tianyang Dai, Ming Chang, Yan Chen, Yang Hu*

University of Science and Technology of China

2026 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2026)

Motivation & Problem

- **Remote Photoplethysmography (rPPG):** Contactless heart rate measurement.
- **The Promise:** Unsupervised learning reduces reliance on expensive labeled data.
- **The Problem:** Training on low-quality "in-the-wild" videos degrades model performance.
- **The Missing Step:** Evaluating video suitability before training.



(a) rPPG datasets vs. non-rPPG datasets (b) Quantitative Comparison on PURE

Limitations of Prior Work

- **Traditional VQA:** Designed for human vision system (HVS), not machine vision system (MVS).
- **Disconnect:** Visually clear ≠ physiologically viable.
- **Standalone SNR metrics:** Easily deceived by periodic artifacts (e.g., flashing lights).

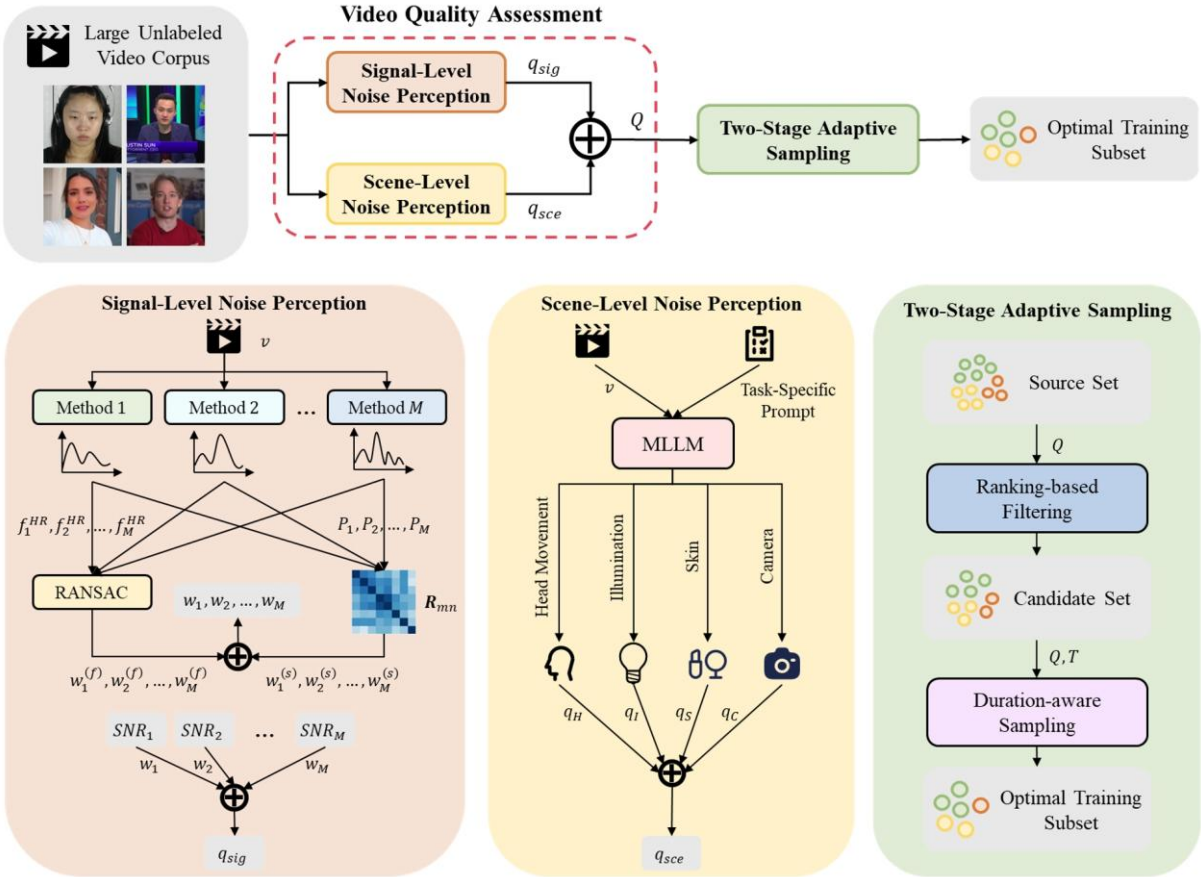
(a) (b)

Key Contribution

1. **rPPG-VQA Framework:** A dual-branch method.
   - Signal-level integrity + Scene-level reasoning.
2. **Two-Stage Adaptive Sampling (TAS):** Curates optimal datasets balancing quality and diversity.
3. **Result:** Substantial accuracy improvements on standard benchmarks.

Method Overview

- **Input:** Unlabeled “in-the-wild” video pool.
- **Branch 1:** Signal-Level Noise Perception (Consensus of multiple methods).
- **Branch 2:** Scene-Level Noise Perception (MLLM-based reasoning).
- **Output:** Unified Quality Score → Target Training Set.



Method Details

- **Signal-Level:** Multi-method consensus (Frequency + Spectral correlation).
  - ✓ Rewards agreement, penalizes outliers (RANSAC).
  - ✓ $w_{i,m}^{(f)} = \exp\left(-\frac{\left(f_{i,m}^{HR} - f_{i,cons}^{HR}\right)^2}{2\sigma_f^2}\right)$
  - ✓ $w_{i,m}^{(s)} = \frac{1}{M-1}\sum_{n \neq m} \mathbf{R}_{i,mn}^2$
- **Scene-Level:** Multimodal large language model (MLLM).
  - ✓ Scores: Head Movement, Illumination, Skin Noise, Camera Noise.

Dimension	Score
Head Movement Noise	$\{0, 1, 2, 3\}$
Illumination Noise	$\{0, 1, 2, 3\}$
Skin Noise	$\{0, 1, 2\}$
Camera Noise	$\gamma_i \cdot \{0, 1, 2\}$

Experiments & Results

Training Set	Sampling	Quality Criterion	Gideon21			Contrast-Phys			SiNC		
			MAE↓	RMSE↓	R↑	MAE↓	RMSE↓	R↑	MAE↓	RMSE↓	R↑
CAS(ME)^2	-	-	1.80	2.24	0.99	1.72	2.74	0.98	2.00	8.16	0.93
Celeb-DF (v2)	-	-	11.94	13.92	0.36	37.34	40.82	-0.14	23.12	28.45	0.03
CAS(ME)^2 + Celeb-DF (v2)	-	-	6.91	12.26	0.57	31.07	36.37	-0.30	8.58	15.28	0.39
	Random	-	4.04	12.02	0.64	7.86	15.02	0.39	6.20	15.20	0.56
	Top/Bottom K	COVER	0.46	0.60	**1.00**	2.56	10.07	0.72	1.69	6.76	0.88
		SNR	0.47	0.58	**1.00**	2.34	8.32	0.82	0.76	4.23	0.95
		IPR	0.57	0.75	**1.00**	4.63	10.53	0.71	1.52	6.54	0.89
		rPPG-VQA	0.42	0.50	**1.00**	1.36	4.05	0.96	0.87	4.33	0.95
	WRS	COVER	0.50	0.58	**1.00**	1.64	7.17	0.86	1.39	5.46	0.92
		SNR	0.42	0.55	**1.00**	0.88	1.76	0.99	1.19	4.97	0.94
		IPR	0.57	0.72	**1.00**	4.89	11.49	0.67	1.39	6.67	0.88
		rPPG-VQA	0.37	0.47	**1.00**	0.71	1.60	0.99	0.74	4.12	0.96
	TAS	**rPPG-VQA**	**0.36**	**0.44**	1.00	**0.47**	**0.74**	**1.00**	**0.72**	**3.43**	**0.97**

- **Datasets:** CAS(ME)^2 + Celeb-DF (v2) (Training) → PURE & BUAA (Testing).
- **Key Finding:** "Wild" data is a double-edged sword.
- **Performance:** MAE dropped from 11.94 to **0.36** (PURE dataset, Gideon21 method).
- Cross-dataset generalization heavily improved.

Ablation

Frequency Consistency Weight	Spectral Correlation Weight	MAE↓	RMSE↓	R↑
✗	✗	0.82	1.47	**1.00**
✓	✗	0.52	0.79	**1.00**
✗	✓	0.57	0.90	**1.00**
✓	✓	**0.47**	**0.74**	**1.00**

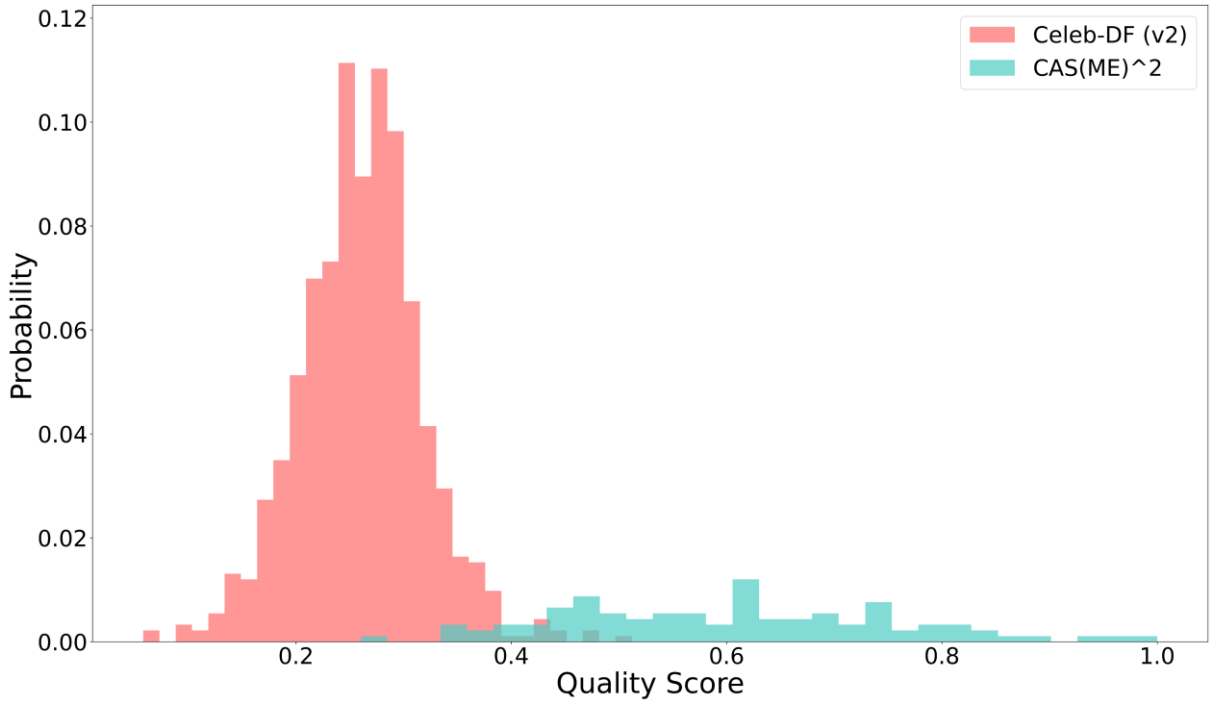
Ranking-based Filtering	Duration-aware Sampling	MAE↓	RMSE↓	R↑
✗	✗	0.71	1.60	0.99
✓	✗	0.70	1.22	**1.00**
✗	✓	0.49	0.82	**1.00**
✓	✓	**0.47**	**0.74**	**1.00**

α	MAE↓	RMSE↓	R↑
0.0	6.91	14.06	0.49
0.2	5.59	12.11	0.63
0.4	4.39	10.81	0.70
0.6	1.52	6.54	0.89
0.8	**0.47**	**0.74**	**1.00**
1.0	0.78	1.15	**1.00**

- **Why TAS?:** Beats Top-K by balancing quality and duration.
- **Fusion Weight (α):** Optimum at $\alpha = 0.8$.
- **Conclusion:** Signal analysis and scene reasoning are strictly complementary.

Visualization

- **High Quality Score (0.6 - 1.0):** Stable illumination, clear face, real physiological signal.
- **Low Quality Score (0.0 - 0.4):** Severe motion, deepfakes, extreme lighting.



Conclusion & Take-aways

- **Take-away 1:** VQA tailored for rPPG is indispensable for unsupervised learning.
- **Take-away 2:** rPPG-VQA bridges signal integrity and MLLM scene reasoning.
- **Code:** https://github.com/Tianyang-Dai/rPPG-VQA.