PHYSPPG: PHYSICS-INFORMED CONTRASTIVE LEARNING FOR PHOTOPLETHYSMOGRAPHY REPRESENTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Photoplethysmography (PPG) enables accessible cardiovascular monitoring, yet motion and acquisition artifacts can obscure the waveform structure needed for reliable physiological inference. We introduce PHYSPPG, a physics-informed contrastive learning framework for robust PPG representation learning. A temporal encoder maps PPG signals to latent trajectories, which are jointly shaped by a contrastive objective and a generalized Euler–Lagrange residual. Learnable kinetic and potential energies together with non-conservative forces provide a low-order dynamical inductive bias without requiring physiological labels or identifying latent coordinates with specific hemodynamic variables. Evaluations with frozen encoders highlight benefits for blood-pressure estimation and heart-rate estimation on WildPPG, a challenging dataset with substantial motion and environmental variation. The learned residual also supports PPG–ECG modality discrimination, providing a complementary diagnostic of compatibility with the learned dynamics. These findings suggest that physics-informed latent dynamics can complement contrastive learning for hemodynamic inference and challenging ambulatory PPG analysis.

1 INTRODUCTION

Photoplethysmography (PPG) measures peripheral blood-volume changes optically and is now a standard sensing modality in clinical monitors and consumer wearables (Allen, 2007). Its accessibility makes PPG attractive for heart-rate (HR), blood-pressure (BP), cardiovascular-risk, sleep, and affective inference. The same measurement mechanism, however, makes PPG sensitive to motion, contact pressure, sensor placement, skin properties, ambient light, and temperature, motivating both multi-sensor artifact reduction and learned signal-quality assessment (Lee et al., 2020; Chatterjee et al., 2022; Zheng et al., 2024). In ambulatory recordings, corruption can overlap the frequency content of the cardiac waveform rather than appearing as easily removable high-frequency noise. WildPPG, for example, captures 216 hours of outdoor recordings across body locations, activities, altitude, and temperature, exposing a substantial gap between controlled and in-the-wild HR estimation (Meier et al., 2024). Semi-naturalistic consumer-wearable collections such as GalaxyPPG further emphasize the need to evaluate beyond tightly controlled acquisition (Park et al., 2025).

Self-supervised representation learning reduces dependence on expensive labels by pretraining encoders on unlabeled signals. Generic objectives such as SimCLR, BYOL, TS2Vec, TF-C, and CoST learn invariance, contextual, or time–frequency structure (Chen et al., 2020; Grill et al., 2020; Yue et al., 2021; Zhang et al., 2022; Woo et al., 2022). Large pretrained models extend this idea across heterogeneous time-series corpora (Goswami et al., 2024; Ansari et al., 2025). For PPG specifically, PaPaGei, AnyPPG, Pulse-PPG, and SIGMA-PPG show that large-scale clinical, wearable, or field recordings can yield broadly transferable encoders (Pillai et al., 2024; Nie et al., 2025; Saha et al., 2025; Guo et al., 2026). These advances establish strong transfer baselines, but their learning signals are primarily statistical: augmented-view agreement, cross-modal alignment, masking, forecasting, or reconstruction. Such objectives need not distinguish a waveform that is statistically plausible from one whose temporal evolution is incompatible with the smooth, dissipative dynamics expected of peripheral hemodynamics.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

We ask whether a dynamical prior can improve PPG representations alongside data-driven learning. A neural encoder maps each segment to a low-dimensional temporal trajectory, on which learnable positive masses, a potential-energy function, and non-conservative forces define a generalized Euler–Lagrange residual. This residual acts as a physics-informed regularizer: it encourages consistency with learned low-order dynamics, while the contrastive objective promotes agreement between views and discrimination between segments. Their joint optimization combines dynamical structure with transferable information.

This leads to PHYSPPG, a physics-informed representation-learning framework. This distinction matters experimentally. We compare PHYSPPG in three separate contexts: controlled self-supervised objectives trained on the same PPG corpus, general pretrained time-series models, and released PPG-specific foundation models. The external pretrained models differ in architecture and pretraining corpus, so that comparison measures downstream transfer utility; the controlled group more directly isolates the learning objective. Across all groups, the downstream data partitions, optimization budget, and trainable probe capacity are matched.

Our contributions are:

- We formulate PPG pretraining as joint contrastive and physics-informed representation learning, using a differentiable generalized Euler–Lagrange residual on a time-resolved latent trajectory.
- We retain temporal latent coordinates instead of imposing the physical constraint on a single pooled vector, allowing the regularizer and downstream temporal probe to access pulse phase and morphology.
- We evaluate against twelve baselines on eight tasks from five datasets under a common downstream protocol, explicitly separating controlled SSL methods, general pretrained time-series models, and PPG-specific foundation models.
- We show a selective empirical pattern: physics regularization is most useful on blood-pressure estimation and in-the-wild HR tasks, while gains are limited on PPG-DaLiA HR and affective tasks. We further demonstrate that the residual can serve as a promising modality-level out-of-distribution (OOD) score.

2 RELATED WORK

Self-supervised time-series representation learning. SimCLR learns augmentation invariance through a temperature-scaled contrastive objective and a nonlinear projection head (Chen et al., 2020); BYOL removes explicit negatives by matching online and momentum-target networks (Grill et al., 2020). TS2Vec learns contextual representations at multiple temporal scales (Yue et al., 2021). TF-C aligns time- and frequency-domain neighborhoods (Zhang et al., 2022), whereas CoST separates seasonal and trend structure using time- and frequency-domain contrastive losses (Woo et al., 2022). TS-TCC provide complementary patch-based and temporal–contextual pretraining strategies (Nie et al., 2022; Eldele et al., 2021). These methods provide complementary generic priors but do not encode a PPG-specific dynamical constraint. MOMENT pretrains across a broad time-series collection for multiple analysis tasks (Goswami et al., 2024); Chronos-2 is designed for universal zero-shot forecasting (Ansari et al., 2025). We include their representations as broad pretrained baselines, while recognizing that forecasting pretraining is not tailored to physiological representation learning.

Pretrained PPG models. Large-scale wearable pretraining has shown that unlabeled PPG contains information useful for diverse health tasks (Abbaspourazad et al., 2023). PaPaGei uses participant-aware and morphology-aware SSL on more than 57,000 hours of public PPG (Pillai et al., 2024). AnyPPG aligns PPG with synchronized ECG and reports more than 100,000 hours of pretraining (Nie et al., 2025). Related PPG-to-ECG translation methods also exploit cross-modal cardiac structure, using Transformer or diffusion-based reconstruction (Lan, 2022; Shome et al., 2023). Pulse-PPG emphasizes uncurated field data (Saha et al., 2025), and SIGMA-PPG combines prior-guided masking with vector-quantized semantic consistency (Guo et al., 2026). These models motivate our PPG-specific transfer group. In contrast, PHYSPPG asks whether an explicit latent-

dynamics regularizer can help a comparatively compact encoder, especially when the observation is corrupted.

Physics-informed learning for physiological signals. Physics-informed neural networks incorporate differential-equation residuals into learning objectives (Raissi et al., 2019; Karniadakis et al., 2021). Energy-based neural dynamics have been formulated through Hamiltonian and Lagrangian parameterizations (Greydanus et al., 2019; Cranmer et al., 2020); generalized Lagrangian networks extend this perspective to non-conservative systems (Xiao et al., 2024). In arterial hemodynamics, Windkessel models describe pressure and flow through low-order resistance, compliance, and inertial components (Westerhof et al., 2009; Charlton et al., 2022), and have also informed PPG-based pressure estimation (Choudhury et al., 2014). Recent physics-informed temporal models have used cardiovascular constraints for task-specific cuffless BP estimation (Wang et al., 2025b). PHYSPPG instead uses a generalized Lagrangian residual during unlabeled pretraining, treating physics as an inductive bias for transferable representation learning.

3 METHOD

3.1 OVERVIEW

PPG is not itself a mechanical coordinate, but an optical observation of blood-volume changes generated by pulsatile cardiovascular dynamics. Its beat-to-beat evolution therefore inherits temporal structure from an underlying pressure–flow system that can be summarized by low-order inertial, compliant, resistive, and driven components (Westerhof et al., 2009; Charlton et al., 2022). This motivates our central modeling premise: although the learned coordinates need not correspond to pressure, flow, or any named physiological variable, a useful PPG representation should admit a compact dynamical description.

We express this prior through generalized Lagrangian mechanics. For a trajectory $\mathbf{q}(t)$, classical mechanics defines the Lagrangian as the difference between kinetic and potential energies, $\mathcal{L} = \mathcal{T} - \mathcal{V}$, and describes each coordinate through the Euler–Lagrange equation

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}_j}\right) - \frac{\partial \mathcal{L}}{\partial q_j} = F_{\mathrm{nc},j}, \qquad j = 1, \ldots, d. \tag{1}$$

The conservative form is recovered when $F_{\mathrm{nc},j} = 0$; the nonzero right-hand side permits dissipation, cardiac drive, and effects not captured by the energy functions. Importantly, Eq. 1 is a structural prior rather than a prescribed physiological equation in our framework. The encoder, latent coordinates, energy functions, and force field are learned jointly, and departures from this equation form a soft training penalty. The prior thus encourages dynamically coherent representations without removing the flexibility of neural representation learning.

Let $\mathbf{x} \in \mathbb{R}^T$ be a normalized PPG segment. Two stochastic views $\tilde{\mathbf{x}}^{(1)}$ and $\tilde{\mathbf{x}}^{(2)}$ are passed through a shared temporal encoder E_θ. A temporal projection produces

$$\mathbf{q}^{(v)} = Q_\phi(E_\theta(\tilde{\mathbf{x}}^{(v)})) \in \mathbb{R}^{d \times T'}, \qquad v \in \{1, 2\}, \tag{2}$$

where d is the latent dimension and T' is the number of latent time steps. We call $\mathbf{q}$ a *physics-structured latent trajectory*: its coordinates are learned features rather than supervised pressure, flow, compliance, or resistance. The trajectory feeds (i) a contrastive projection branch and (ii) a differentiable physics module (Fig. 1).

3.2 LEARNABLE LAGRANGIAN LATENT DYNAMICS

We instantiate this premise with a generalized Lagrangian and a learnable non-conservative force, rather than treating the circulation as an isolated conservative system. For each time point, a mass network predicts a diagonal mass vector with strictly positive components,

$$\mathbf{m}(\mathbf{q}) \in \mathbb{R}^d_{>0}. \tag{3}$$

Positivity is enforced through a softplus-based parameterization. The kinetic and potential energies are

$$\mathcal{T}(\mathbf{q}, \dot{\mathbf{q}}) = \frac{1}{2}\sum_{j=1}^{d} m_j(\mathbf{q})\dot{q}_j^2, \qquad \mathcal{V}(\mathbf{q}) = V_\omega(\mathbf{q}), \tag{4}$$

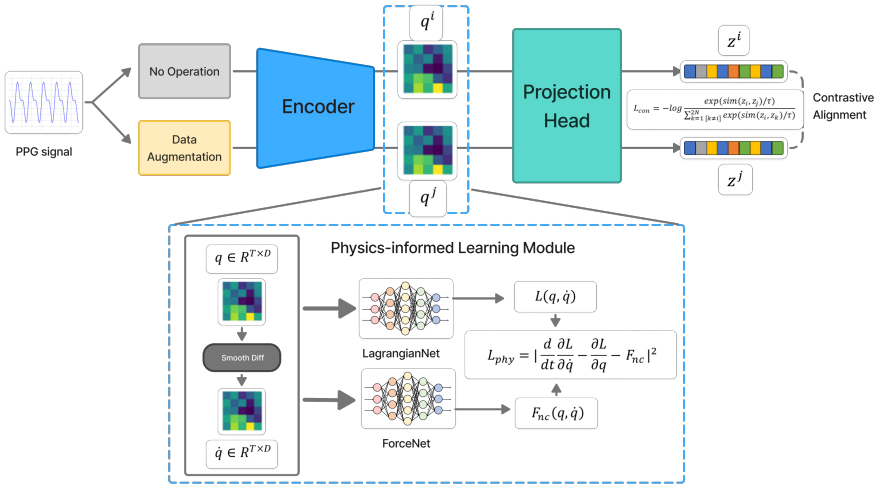


Figure 1: **PhysPPG.** A shared encoder maps PPG views to time-resolved latent trajectories $\mathbf{q}$. The projection head pools these trajectories and produces contrastive embeddings. In parallel, smoothed differentiation and learned Lagrangian and force networks define a generalized Euler–Lagrange residual. The training objective combines contrastive alignment with a soft physics-informed penalty.

and $\mathcal{L} = \mathcal{T} - \mathcal{V}$. A separate network F_η receives $[\mathbf{q}, \dot{\mathbf{q}}]$ and predicts the non-conservative force $\mathbf{F}_{\text{nc}}$. This term permits damping and external drive rather than incorrectly imposing exact energy conservation.

Differentiation amplifies high-frequency noise. We combine smoothing and differentiation using a Gaussian kernel of bandwidth $\sigma > 0$ and its derivative,

$$G_\sigma(s) = \frac{e^{-s^2/(2\sigma^2)}}{\sqrt{2\pi}\sigma}, \qquad D_\sigma(s) = G'_\sigma(s) = -\frac{s}{\sigma^2} G_\sigma(s). \tag{5}$$

For a latent signal, differentiation of its smoothed version gives

$$\dot{\mathbf{q}} \approx \frac{d}{dt}(G_\sigma * \mathbf{q}) = D_\sigma * \mathbf{q}. \tag{6}$$

Here $*$ denotes temporal convolution, applied independently to each latent channel. A finite-support Gaussian-derivative stencil approximates this operator on sampled trajectories; its bandwidth controls the smoothing scale. The fixed linear filter remains differentiable with respect to the encoder output and is also applied to generalized momentum. The partial derivatives of $\mathcal{L}$ treat $\mathbf{q}$ and $\dot{\mathbf{q}}$ as independent arguments. The resulting residual is

$$\mathbf{r}(t) = \frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{q}}}\right) - \frac{\partial \mathcal{L}}{\partial \mathbf{q}} - \mathbf{F}_{\text{nc}}(\mathbf{q}, \dot{\mathbf{q}}), \tag{7}$$

and the physics loss averages its squared norm over batch, time, latent dimensions, and both views:

$$\mathcal{L}_{\text{phy}} = \frac{1}{2} \sum_{v=1}^{2} \mathbb{E}_{b,t,j} \left[r_{b,t,j}^{(v)\,2} \right]. \tag{8}$$

The residual regularizes the temporal representation through the learned energy–force structure.

3.3 Contrastive Representation Objective

For each view, temporal global average pooling followed by a nonlinear projection P_ξ produces $\mathbf{z}^{(v)}$. After ℓ_2 normalization, we optimize a temperature-scaled contrastive objective between the

Table 1: Downstream datasets, tasks, and evaluation protocols.

Dataset	Tasks	Protocol
PPG-BP	SBP, DBP regression	5-fold
PulseDB	SBP, DBP regression	LOSO
WildPPG	HR regression	LOSO
PPG-DaLiA	HR regression	LOSO
WESAD	arousal, valence classification	LOSO

two views of each segment. Writing sim for cosine similarity and $\tau > 0$ for temperature, the per-anchor form is

$$\ell(i,j) = -\log \frac{\exp(\mathrm{sim}(\mathbf{z}_i, \mathbf{z}_j)/\tau)}{\sum_{k \neq i} \exp(\mathrm{sim}(\mathbf{z}_i, \mathbf{z}_k)/\tau)}, \qquad \mathcal{L}_{\mathrm{con}} = \frac{1}{2N} \sum_{i=1}^{N} \left[\ell(i,i') + \ell(i',i)\right]. \tag{9}$$

The complete pretraining objective is

$$\mathcal{L}_{\mathrm{total}} = \lambda_{\mathrm{con}} \mathcal{L}_{\mathrm{con}} + \lambda_{\mathrm{phy}} \mathcal{L}_{\mathrm{phy}}, \tag{10}$$

where $\lambda_{\mathrm{con}} > 0$ and $\lambda_{\mathrm{phy}} \geq 0$ control the contrastive and dynamical contributions. Both terms backpropagate through the shared encoder and temporal projection, so the physics penalty regularizes the representation during pretraining. Setting $\lambda_{\mathrm{phy}} = 0$ recovers the no-physics, SimCLR-style objective while retaining the same temporal representation.

3.4 Residual as a diagnostic score

At inference, Eq. 8 can be evaluated without task labels. We therefore use

$$s(\mathbf{x}) = \mathbb{E}_{t,j}\left[r_{t,j}(\mathbf{x})^2\right] \tag{11}$$

as a post-hoc diagnostic score: inputs that fit the learned PPG dynamics poorly should receive larger scores.

4 Experimental Protocol

4.1 Pretraining and datasets

We pretrain on unlabeled ten-second PPG segments from PulseDB, a collection derived from MIMIC-III and VitalDB with more than five million cleaned segments (Wang et al., 2023). Signals are represented at 125 Hz and normalized independently per segment. The encoder is a one-dimensional ResNet-18 followed by a 3-tap temporal projection to $d = 4$ latent channels. We train for 50 epochs using two independently perturbed views, a 128-dimensional contrastive projection, $\tau = 0.1$, and the loss in Eq. 10 with fixed weights $\lambda_{\mathrm{con}} = 1$ and $\lambda_{\mathrm{phy}} = 0.1$.

We select five datasets to evaluate complementary aspects of generalization across eight tasks (Table 1). PPG-BP tests cross-dataset BP transfer to short fingertip recordings with cuff targets from 219 participants (Liang et al., 2022); PulseDB tests SBP/DBP estimation on unseen subjects within the pretraining source domain (Wang et al., 2023). WildPPG evaluates HR transfer under outdoor motion and environmental variation (Meier et al., 2024), while PPG-DaLiA assesses daily-activity wrist PPG (Reiss et al., 2019). WESAD evaluates transfer beyond HR/BP to affective labels in 15 participants (Schmidt et al., 2018), complementing earlier wearable affect studies (Choi et al., 2023; Li & Washington, 2023; Ninh et al., 2022). All PulseDB downstream test subjects and their recordings are excluded from the pretraining corpus. PulseDB therefore measures subject-disjoint, same-dataset generalization; PPG-BP, WildPPG, PPG-DaLiA, and WESAD assess cross-dataset transfer. These complementary settings probe transferability rather than isolate the causal effect of any single acquisition factor.

4.2 Baselines and downstream evaluation

We organize the twelve baselines into three groups. *General SSL* includes SimCLR, BYOL, TS2Vec, TF-C, and CoST. These methods use the same pretraining data and controlled configuration as

PHYSPPG, with objective-specific components retained. *General pretrained time-series models* includes MOMENT and Chronos-2. *PPG-specific pretrained models* includes PaPaGei-S, PaPaGei-P, AnyPPG, Pulse-PPG, and SIGMA-PPG. For the latter two groups we use released checkpoints and their required input adapters; because architectures and pretraining corpora differ, these comparisons measure transfer utility rather than isolating an objective.

All downstream methods use the same source waveforms, subject/fold partitions, validation rule, training budget, and model-selection metric. Encoders are frozen. Model-specific feature adapters connect native encoder outputs to a shared lightweight temporal probe, and the total number of trainable downstream parameters is matched across methods. Regression uses Huber loss and reports MAE, RMSE, Pearson r, and R^2; classification uses class-weighted cross entropy and reports accuracy and macro-F1. Probes are optimized with AdamW for at most 50 epochs, and checkpoint selection uses validation performance only. We report the mean over held-out subjects/folds; confidence intervals are Student-t 95% intervals. All four regression metrics are reported together in the main tables. All experimental tables report mean $\pm$95% CI half-width to two decimal places. Rankings use unrounded means, so values that appear tied after rounding can have different ranks.

5 RESULTS

5.1 BLOOD-PRESSURE ESTIMATION

The strongest overall results occur on BP estimation (Tables 2 and 3). PHYSPPG is best on all four DBP metrics in PPG-BP, reducing the matched-control MAE from 9.74 to 8.88 and RMSE from 12.27 to 11.16. For PPG-BP SBP, CoST has the lowest MAE (14.43), whereas PHYSPPG has the lowest RMSE (18.27). On PulseDB, PHYSPPG is best on every reported MAE and RMSE: SBP improves from 15.51/19.72 to 15.32/19.17 and DBP from 6.02/10.12 to 5.75/9.90 relative to SimCLR. It is also best on PulseDB SBP correlation and R^2, and on PulseDB DBP R^2 (Table 3). The PPG-BP DBP R^2 remains negative for all methods, indicating poor absolute fit despite PHYSPPG's relative improvement. Negative fold-level R^2 means that squared error exceeds that of the corresponding test-fold target mean, a descriptive reference rather than a deployable predictor.

Blood-pressure estimation depends not only on pulse periodicity but also on waveform morphology and temporal variations associated with vascular mechanics. Preserving a time-resolved latent trajectory and regularizing its evolution through energy–force consistency may help the representation capture these dynamic features, while the contrastive objective maintains discriminative information across samples. This combination therefore provides a plausible task-aligned inductive bias for blood-pressure estimation.

5.2 HEART-RATE ROBUSTNESS

Table 4 shows different transfer behavior on two ambulatory datasets. On PPG-DaLiA, Chronos-2 and AnyPPG obtain the lowest errors, while PHYSPPG ranks ninth on all four metrics. Physics structure is therefore not a universal advantage on PPG-DaLiA. On WildPPG, however, PHYSPPG has the best MAE (24.39), RMSE (30.52), and R^2 (-0.209), and the second-best Pearson correlation (0.150). Against the matched SimCLR control, physics reduces MAE by 3.17% and RMSE by 2.87%. Against AnyPPG, its WildPPG MAE is 24.5% lower (24.39 versus 32.32). However, WildPPG performance remains weak in absolute terms: correlation is low and mean R^2 is negative.

A possible explanation for the WildPPG gains is that dynamical regularization reduces reliance on transient, local, or unstable signal variations. It may instead encourage representations to emphasize pulse structure that persists across time, thereby supporting heart-rate estimation under challenging measurement conditions.

5.3 AFFECTIVE TRANSFER AND SCOPE OF THE INDUCTIVE BIAS

On WESAD, PHYSPPG improves the matched SimCLR point estimate on arousal accuracy and F1 and on valence F1, but not valence accuracy (Table 5). It does not lead either affective task: Chronos-2 is best for arousal accuracy, TS2Vec for arousal F1, CoST for valence accuracy, and AnyPPG for

Table 2: PPG-BP blood-pressure estimation (five folds). MAE/RMSE in mmHg; r and R^2 are dimensionless. Values are mean ±95% CI half-width, both to two decimals. Bold/underline mark first/second by unrounded means; displayed ties may have different ranks.

	SBP				DBP			
Method	MAE ↓	RMSE ↓	r ↑	R^2 ↑	MAE ↓	RMSE ↓	r ↑	R^2 ↑
General SSL								
SimCLR	14.72±3.56	18.59±4.15	0.44±0.17	0.08±0.26	9.74±1.85	12.27±2.29	0.22±0.15	-0.42±0.40
BYOL	14.72±3.68	18.68±4.15	**0.47±0.23**	0.06±0.31	9.78±2.70	12.17±3.28	0.19±0.28	-0.38±0.40
TS2Vec	14.51±3.20	18.61±3.34	0.45±0.14	0.07±0.27	10.09±2.61	12.53±3.17	0.05±0.30	-0.50±0.71
TF-C	14.77±3.20	19.02±3.26	0.44±0.13	0.03±0.21	9.91±1.69	12.37±1.88	0.16±0.21	-0.51±0.82
CoST	**14.43±3.84**	18.52±4.71	0.46±0.19	0.09±0.27	9.59±2.84	11.86±3.45	0.15±0.23	-0.32±0.51
General pretrained time-series models								
MOMENT	16.24±3.03	20.53±3.29	0.38±0.07	-0.13±0.30	10.49±2.40	13.11±3.11	0.11±0.14	-0.63±0.67
Chronos-2	17.02±2.87	21.51±3.38	0.33±0.12	-0.23±0.23	10.49±1.41	13.31±2.01	0.08±0.16	-0.67±0.42
PPG-specific pretrained models								
PaPaGei-S	14.65±3.54	18.73±4.03	0.40±0.13	0.08±0.13	9.65±1.90	12.13±2.38	-0.05±0.13	-0.38±0.32
PaPaGei-P	15.79±3.26	20.28±4.05	0.26±0.22	-0.08±0.14	10.07±1.81	12.67±2.29	0.02±0.16	-0.50±0.33
AnyPPG	15.53±2.75	20.02±3.34	0.32±0.12	-0.07±0.22	9.37±1.62	11.91±2.39	0.14±0.14	-0.32±0.27
Pulse-PPG	15.18±2.75	19.66±3.32	0.39±0.11	-0.03±0.15	10.42±1.86	13.14±2.38	0.06±0.10	-0.62±0.31
SIGMA-PPG	15.67±3.42	19.89±3.57	0.39±0.13	-0.05±0.23	10.29±2.11	13.16±2.52	0.11±0.08	-0.62±0.35
PHYSPPG (ours)	14.56±3.18	**18.27±3.34**	0.45±0.13	**0.12±0.11**	**8.88±2.01**	**11.16±2.41**	**0.29±0.16**	**-0.17±0.36**

Table 3: PulseDB blood-pressure estimation (20 evaluation folds; subject-disjoint, same-domain). MAE/RMSE in mmHg; r and R^2 are dimensionless. Values are mean ±95% CI half-width, both to two decimals. Bold/underline mark first/second by unrounded means; displayed ties may have different ranks.

	SBP				DBP			
Method	MAE ↓	RMSE ↓	r ↑	R^2 ↑	MAE ↓	RMSE ↓	r ↑	R^2 ↑
General SSL								
SimCLR	15.51±1.10	19.72±1.21	0.58±0.04	0.27±0.07	6.02±0.13	10.12±0.10	0.79±0.03	0.59±0.04
BYOL	20.57±1.58	25.20±1.76	0.51±0.05	-0.20±0.12	6.15±0.19	10.24±0.15	0.78±0.03	0.58±0.04
TS2Vec	17.66±1.12	22.33±1.19	0.52±0.04	0.06±0.07	6.46±0.14	10.57±0.16	0.78±0.03	0.55±0.04
TF-C	19.14±1.38	23.83±1.51	0.56±0.04	-0.07±0.09	6.46±0.18	10.49±0.13	0.78±0.03	0.56±0.03
CoST	21.10±1.55	26.19±1.64	0.50±0.04	-0.29±0.11	6.74±0.24	10.81±0.14	0.77±0.03	0.53±0.03
General pretrained time-series models								
MOMENT	19.79±1.40	24.47±1.49	0.51±0.04	-0.13±0.10	6.40±0.26	10.39±0.15	**0.79±0.03**	0.57±0.03
Chronos-2	17.84±1.09	21.89±1.10	0.60±0.03	0.10±0.06	6.26±0.23	10.34±0.13	0.78±0.03	0.57±0.03
PPG-specific pretrained models								
PaPaGei-S	18.23±1.03	22.97±1.08	0.59±0.02	0.01±0.06	9.16±0.55	13.48±0.49	0.61±0.02	0.28±0.02
PaPaGei-P	19.60±1.14	25.21±1.27	0.49±0.03	-0.20±0.08	9.63±0.73	14.21±0.64	0.54±0.02	0.20±0.04
AnyPPG	23.69±1.83	28.72±1.95	0.52±0.04	-0.56±0.14	6.89±0.40	11.18±0.26	0.78±0.03	0.50±0.03
Pulse-PPG	17.47±1.37	21.83±1.46	0.55±0.05	0.10±0.09	6.08±0.24	10.16±0.19	0.78±0.02	0.59±0.03
SIGMA-PPG	17.55±1.05	22.05±1.10	0.57±0.03	0.08±0.06	8.33±0.41	12.44±0.34	0.67±0.02	0.39±0.02
PHYSPPG (ours)	**15.32±0.97**	**19.17±1.04**	**0.61±0.04**	**0.31±0.05**	**5.75±0.10**	**9.90±0.13**	0.79±0.03	**0.60±0.04**

valence F1. These results limit claims of general transfer superiority. The smaller gains on affective labels are consistent with, but do not establish, a task-specific benefit from the dynamical prior.

5.4 MATCHED ABLATION AND RESIDUAL-BASED MODALITY DISCRIMINATION

Across the 28 complete task–metric combinations, PHYSPPG improves over the matched SimCLR/no-physics control on 26 point estimates. It ranks first among all thirteen methods on 16 metrics and in the top three on 20.

We next evaluate whether the learned residual contains information not supplied by the contrastive objective. We score held-out PPG and ECG segments without using ECG during representation pretraining. For the no-physics control, we fit the same dynamics module on frozen PPG features before evaluation. Figure 2 shows that the control residual distributions overlap (AUROC 0.569),

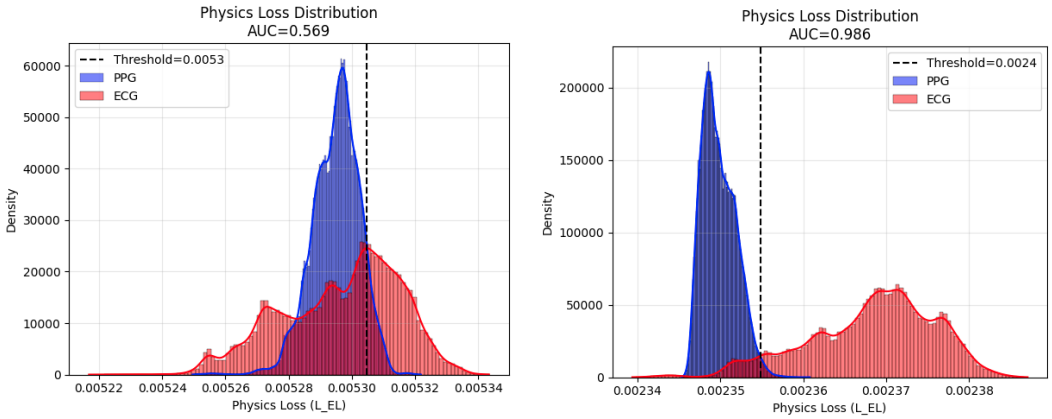
Table 4: Heart-rate estimation. MAE/RMSE in bpm; r and R^2 are dimensionless. Values are mean ±95% CI half-width, both to two decimals. Bold/underline mark first/second by unrounded means; displayed ties may have different ranks.

	PPG-DaLiA				WildPPG			
Method	MAE ↓	RMSE ↓	r ↑	R^2 ↑	MAE ↓	RMSE ↓	r ↑	R^2 ↑
General SSL								
SimCLR	9.81±1.65	12.84±1.79	0.61±0.09	0.19±0.23	25.19±1.95	31.42±1.99	0.13±0.05	-0.29±0.14
BYOL	9.03±1.13	12.21±1.47	0.62±0.08	0.28±0.15	27.68±1.72	33.37±1.73	0.14±0.05	-0.48±0.20
TS2Vec	9.29±1.27	12.36±1.61	0.61±0.08	0.26±0.19	26.61±1.48	32.27±1.57	0.12±0.06	-0.38±0.17
TF-C	8.93±1.30	12.05±1.55	0.63±0.08	0.30±0.17	28.83±1.56	34.42±1.60	0.12±0.03	-0.57±0.20
CoST	9.37±1.52	12.39±1.71	0.62±0.08	0.25±0.21	27.73±1.21	33.57±1.32	0.13±0.04	-0.49±0.16
General pretrained time-series models								
MOMENT	9.23±1.39	12.48±1.69	0.60±0.09	0.24±0.21	30.17±1.18	36.73±1.33	0.11±0.05	-0.80±0.22
Chronos-2	**8.57±1.25**	**11.63±1.53**	**0.65±0.08**	**0.34±0.17**	29.56±1.12	35.44±1.14	0.15±0.03	-0.67±0.19
PPG-specific pretrained models								
PaPaGei-S	10.58±1.81	13.79±2.13	0.54±0.10	0.05±0.32	32.21±0.88	36.41±0.93	0.07±0.03	-0.78±0.22
PaPaGei-P	11.49±2.18	14.76±2.42	0.48±0.09	-0.10±0.42	31.44±0.78	35.81±0.83	0.11±0.02	-0.71±0.20
AnyPPG	8.66±1.05	11.70±1.32	0.63±0.09	0.34±0.14	32.32±1.12	36.96±1.18	0.11±0.03	-0.83±0.24
Pulse-PPG	9.02±1.38	12.15±1.74	0.63±0.08	0.19±0.22	25.95±1.69	32.36±1.65	**0.16±0.04**	-0.37±0.14
SIGMA-PPG	10.13±1.48	13.44±1.76	0.56±0.08	0.28±0.20	31.76±1.12	38.10±1.31	0.07±0.05	-0.96±0.28
PHYSPPG (ours)	9.62±1.55	12.70±1.77	0.61±0.09	0.20±0.23	**24.39±1.86**	**30.52±1.76**	0.15±0.05	**-0.21±0.08**

Table 5: WESAD affective classification (15 held-out subjects). Values are mean ±95% CI half-width, both to two decimals. Bold/underline mark first/second by unrounded means; displayed ties may have different ranks.

	Arousal		Valence	
Method	Acc. ↑	Macro-F1 ↑	Acc. ↑	Macro-F1 ↑
General SSL				
SimCLR	0.64 ± 0.09	0.42 ± 0.06	0.56 ± 0.13	0.34 ± 0.08
BYOL	0.69 ± 0.09	0.44 ± 0.07	0.56 ± 0.09	0.35 ± 0.05
TS2Vec	0.69 ± 0.09	**0.45 ± 0.09**	0.56 ± 0.12	0.33 ± 0.06
TF-C	0.66 ± 0.08	0.45 ± 0.06	0.60 ± 0.09	0.39 ± 0.05
CoST	0.68 ± 0.09	0.44 ± 0.08	**0.61 ± 0.10**	0.39 ± 0.05
General pretrained time-series models				
MOMENT	0.67 ± 0.07	0.45 ± 0.06	0.57 ± 0.08	0.38 ± 0.04
Chronos-2	**0.70 ± 0.08**	0.45 ± 0.07	0.60 ± 0.07	0.40 ± 0.05
PPG-specific pretrained models				
PaPaGei-S	0.58 ± 0.07	0.42 ± 0.06	0.54 ± 0.07	0.37 ± 0.05
PaPaGei-P	0.53 ± 0.07	0.38 ± 0.05	0.46 ± 0.06	0.33 ± 0.04
AnyPPG	0.60 ± 0.08	0.42 ± 0.07	0.58 ± 0.08	**0.41 ± 0.05**
Pulse-PPG	0.68 ± 0.10	0.43 ± 0.07	0.54 ± 0.11	0.31 ± 0.06
SIGMA-PPG	0.64 ± 0.09	0.42 ± 0.06	0.57 ± 0.08	0.39 ± 0.05
PHYSPPG (ours)	0.67 ± 0.09	0.44 ± 0.05	0.55 ± 0.10	0.36 ± 0.07

whereas the PHYSPPG residual separates the modalities with AUROC 0.986. This experiment establishes modality discrimination, not comprehensive artifact detection: ECG is a deliberately distinct periodic signal, and future work must test motion corruption, pathological PPG, and device shifts.



(a) Matched no-physics control (b) PHYSPPG

Figure 2: **Residual-based modality discrimination.** Density estimates of the generalized Euler–Lagrange residual for held-out PPG and ECG segments. The residual learned jointly with the representation sharply separates the two modalities (AUROC 0.986), while a dynamics module fitted after no-physics pretraining does not (AUROC 0.569). AUROC is threshold independent; dashed lines show the operating thresholds used for the displayed analyses.

6 DISCUSSION AND LIMITATIONS

The results support a targeted interpretation of physics-informed representation learning. PHYSPPG is not uniformly best: it trails general and PPG-specific pretrained models on PPG-DaLiA and WESAD. Its largest relative advantages occur on BP estimation and WildPPG.

Several limitations remain. First, the learned generalized coordinates are not identifiable physiological variables. Flexible energy and force networks can explain multiple latent parameterizations, so the residual should be viewed as a regularizer and diagnostic score rather than mechanistic discovery. Second, the OOD analysis separates PPG from ECG; it does not yet establish rejection of motion artifact, arrhythmia, poor contact, or demographic/device shifts. Third, external foundation models necessarily differ in pretraining data, architecture, native input length, and normalization. Their table entries quantify practical frozen transfer under a matched probe, while the matched SSL experiment is the cleaner causal comparison for the physics loss.

7 CONCLUSION

We presented PHYSPPG, a physics-informed representation-learning framework that couples contrastive pretraining with a generalized Euler–Lagrange residual on time-resolved PPG latents. Across eight tasks and twelve diverse baselines, the method is strongest on BP inference and real-world HR estimation and improves 26 of 28 point estimates over its matched no-physics control. It is not universally superior on PPG-DaLiA HR or affective tasks, delineating where the inductive bias is useful. The learned residual also provides a promising modality-level OOD signal. More broadly, the study suggests that structured latent dynamics can make physiological representations more robust and diagnostically informative.

AI USE STATEMENT

Generative AI tools were used to assist translation, language editing, manuscript organization, literature retrieval, static inspection of experimental code, the drafting of portions of the code, and portions of the mathematical proofs. The authors reviewed the AI-assisted text, code, and mathematical arguments, verified numerical values against the experiment tables, checked equations against the implementation, and remain responsible for all claims, citations, proofs, code, and artifacts in this submission.

Ethics statement

This study introduces no new human-subject data collection and uses previously released, de-identified datasets under their respective access conditions and licenses. PPG-based predictions and residual scores are research outputs, not medical diagnoses or certified safety mechanisms. Performance can vary with acquisition hardware, activity, physiology, skin properties, and population composition; clinical use would require prospective validation and subgroup analysis.

Reproducibility statement

Section 3 specifies the objective and differentiable physics module. Section 4 states the datasets, split units, baseline groups, preprocessing, optimization, probe matching, model selection, and metrics. Pretraining code is provided as supplementary material.

References

Salar Abbaspourazad, Oussama Elachqar, Andrew C. Miller, Saba Emrani, Udhyakumar Nallasamy, and Ian Shapiro. Large-scale training of foundation models for wearable biosignals, 2023. URL https://arxiv.org/abs/2312.05409.

John Allen. Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*, 28(3):R1–R39, 2007. ISSN 0967-3334, 1361-6579. doi: 10.1088/0967-3334/28/3/R01. URL https://iopscience.iop.org/article/10.1088/0967-3334/28/3/R01.

Abdul Fatir Ansari, Lorenzo Stella, Caner Turkmen, Xiyuan Zhang, Pedro Mercado, Huibin Shen, Oleksandr Shchur, Syama Sundar Rangapuram, Sebastian Pineda Arango, Shubham Kapoor, Jasper Zschiegner, Danielle C. Maddix, Hao Wang, Michael W. Mahoney, Kari Torkkola, Andrew Gordon Wilson, Michael Bohlke-Schneider, and Yuyang Wang. Chronos: Learning the language of time series, 2024. URL https://arxiv.org/abs/2403.07815.

Abdul Fatir Ansari, Oleksandr Shchur, Jaris Küken, Andreas Auer, Boran Han, Pedro Mercado, Syama Sundar Rangapuram, Huibin Shen, Lorenzo Stella, Xiyuan Zhang, Mononito Goswami, Shubham Kapoor, Danielle C. Maddix, Pablo Guerron, Tony Hu, Junming Yin, Nick Erickson, Prateek Mutalik Desai, Hao Wang, Huzefa Rangwala, George Karypis, Yuyang Wang, and Michael Bohlke-Schneider. Chronos-2: From univariate to universal forecasting, 2025. URL https://arxiv.org/abs/2510.15821.

Ina Indlekofer Attila Reiss. PPG-DaLiA, 2019. URL https://archive.ics.uci.edu/dataset/495.

Rafael Bischof and Michael Kraus. Multi-objective loss balancing for physics-informed deep learning, 2021. URL https://arxiv.org/abs/2110.09813.

Peter H. Charlton, Birutė Paliakaitė, Kristjan Pilt, Martin Bachler, Serena Zanelli, Dániel Kulin, John Allen, Magid Hallab, Elisabetta Bianchini, Christopher C. Mayer, Dimitrios Terentes-Printzios, Verena Dittrich, Bernhard Hametner, Dave Veerasingam, Dejan Žikić, and Vaidotas Marozas. Assessing hemodynamics from the photoplethysmogram to gain insights into vascular age: a review from VascAgeNet. *American Journal of Physiology-Heart and Circulatory Physiology*, 322(4):H493–H522, 2022. ISSN 0363-6135, 1522-1539. doi: 10.1152/ajpheart.00392.2021. URL https://journals.physiology.org/doi/10.1152/ajpheart.00392.2021.

Tamaghno Chatterjee, Aayushman Ghosh, and Sayan Sarkar. Signal quality assessment of photoplethysmogram signals using quantum pattern recognition and lightweight CNN architecture. *arXiv preprint arXiv:2202.00606*, 2022. doi: 10.48550/ARXIV.2202.00606. URL https://arxiv.org/abs/2202.00606.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations, 2020. URL https://arxiv.org/abs/2002.05709.

Zhao Chen, Vijay Badrinarayanan, Chen-Yu Lee, and Andrew Rabinovich. GradNorm: Gradient normalization for adaptive loss balancing in deep multitask networks. 2017. doi: 10.48550/ARXIV.1711.02257. URL https://arxiv.org/abs/1711.02257.

Jiho Choi, Gyutae Hwang, Jun Seong Lee, Moonwook Ryu, and Sang Jun Lee. Weighted knowledge distillation of attention-LRCN for recognizing affective states from PPG signals. *Expert Systems with Applications*, 233:120883, 2023. ISSN 09574174. doi: 10.1016/j.eswa.2023.120883. URL https://linkinghub.elsevier.com/retrieve/pii/S0957417423013854.

Anirban Dutta Choudhury, Rohan Banerjee, Aniruddha Sinha, and Shaswati Kundu. Estimating blood pressure using windkessel model on photoplethysmogram. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4567–4570. IEEE, 2014. ISBN 978-1-4244-7929-0. doi: 10.1109/EMBC.2014.6944640. URL http://ieeexplore.ieee.org/document/6944640/.

Miles Cranmer, Sam Greydanus, Stephan Hoyer, Peter Battaglia, David Spergel, and Shirley Ho. Lagrangian neural networks, 2020. URL https://arxiv.org/abs/2003.04630.

Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, Chee Keong Kwoh, Xiaoli Li, and Cuntai Guan. Time-series representation learning via temporal and contextual contrasting, 2021. URL https://arxiv.org/abs/2106.14112.

William Fedus, Barret Zoph, and Noam Shazeer. Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity, 2021. URL https://arxiv.org/abs/2101.03961.

William Fedus, Jeff Dean, and Barret Zoph. A review of sparse expert models in deep learning, 2022. URL https://arxiv.org/abs/2209.01667.

Márton Á Goda, Peter H Charlton, and Joachim A Behar. pyPPG: a python toolbox for comprehensive photoplethysmography signal analysis. *Physiological Measurement*, 45(4):045001, 2024. ISSN 0967-3334, 1361-6579. doi: 10.1088/1361-6579/ad33a2. URL https://iopscience.iop.org/article/10.1088/1361-6579/ad33a2.

Mononito Goswami, Konrad Szafer, Arjun Choudhry, Yifu Cai, Shuo Li, and Artur Dubrawski. MOMENT: A family of open time-series foundation models, 2024. URL https://arxiv.org/abs/2402.03885.

Sam Greydanus, Misko Dzamba, and Jason Yosinski. Hamiltonian neural networks, 2019. URL http://arxiv.org/abs/1906.01563.

Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H. Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Daniel Guo, Mohammad Gheshlaghi Azar, Bilal Piot, Koray Kavukcuoglu, Rémi Munos, and Michal Valko. Bootstrap your own latent: A new approach to self-supervised learning. In *Advances in Neural Information Processing Systems*, 2020.

Lei Gu, Weiguo Huang, and Guifu Du. A Physics-informed Supervised Contrastive Learning Domain Adaptation Framework for Intelligent Bearing Fault Diagnosis. In *2025 IEEE International Conference on Sensing, Diagnostics, Prognostics, and Control*, pp. 99–104, Chongqing, China, November 2025. IEEE. ISBN 979-8-3315-7739-1. doi: 10.1109/SDPC68151.2025.11347744. URL https://ieeexplore.ieee.org/document/11347744/.

Zongheng Guo, Tao Chen, Yang Jiao, Yi Pan, Xiao Hu, and Manuela Ferrario. SIGMA-PPG: Statistical-prior informed generative masking architecture for PPG foundation model, 2026.

A. Ali Heydari, Craig A. Thompson, and Asif Mehmood. SoftAdapt: Techniques for adaptive loss weighting of neural networks with multi-part loss functions, 2019. URL https://arxiv.org/abs/1912.12355.

Li Jing, Pascal Vincent, Yann LeCun, and Yuandong Tian. Understanding dimensional collapse in contrastive self-supervised learning. 2021. doi: 10.48550/ARXIV.2110.09348. URL https://arxiv.org/abs/2110.09348.

Bharindra Kamanditya, Yunendah Nur Fuadah, Nurul Qashri Mahardika T., and Ki Moo Lim. Continuous blood pressure prediction system using conv-LSTM network on hybrid latent features of photoplethysmogram (PPG) and electrocardiogram (ECG) signals. *Scientific Reports*, 14(1):16450, 2024. ISSN 2045-2322. doi: 10.1038/s41598-024-66514-y. URL https://www.nature.com/articles/s41598-024-66514-y.

George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3:422–440, 2021. doi: 10.1038/s42254-021-00314-5.

Panagiotis Kasnesis, Lazaros Toumanidis, Alessio Burrello, Christos Chatzigeorgiou, and Charalampos Z. Patrikakis. Multi-head cross-attentional PPG and motion signal fusion for heart rate estimation, 2022. URL https://arxiv.org/abs/2210.11415.

Seyed Mehran Kazemi, Rishab Goel, Sepehr Eghbali, Janahan Ramanan, Jaspreet Sahota, Sanjay Thakur, Stella Wu, Cathal Smyth, Pascal Poupart, and Marcus Brubaker. Time2vec: Learning a vector representation of time, 2019. URL https://arxiv.org/abs/1907.05321.

Alex Kendall, Yarin Gal, and Roberto Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics, 2017. URL https://arxiv.org/abs/1705.07115.

Ella Lan. Performer: A novel PPG-to-ECG reconstruction transformer for a digital biomarker of cardiovascular disease detection, 2022. URL https://arxiv.org/abs/2204.11795.

Jongshill Lee, Minseong Kim, Hoon-Ki Park, and In Young Kim. Motion artifact reduction in wearable photoplethysmography based on multi-channel sensors with multiple wavelengths. *Sensors*, 20(5):1493, 2020. ISSN 1424-8220. doi: 10.3390/s20051493. URL https://www.mdpi.com/1424-8220/20/5/1493.

Dmitry Lepikhin, HyoukJoong Lee, Yuanzhong Xu, Dehao Chen, Orhan Firat, Yanping Huang, Maxim Krikun, Noam Shazeer, and Zhifeng Chen. GShard: Scaling giant models with conditional computation and automatic sharding, 2020. URL https://arxiv.org/abs/2006.16668.

Joe Li and Peter Washington. A comparison of personalized and generalized approaches to emotion recognition using consumer wearable devices: Machine learning study, 2023. URL https://arxiv.org/abs/2308.14245.

Xiaoyan Li, Shixin Xu, Faisal Habib, Neda Aminnejad, Arvind Gupta, and Huaxiong Huang. CLEP-GAN: An innovative approach to subject-independent ECG reconstruction from PPG signals, 2025. URL https://arxiv.org/abs/2502.17536.

Yongbo Liang, Guiyong Liu, Zhencheng Chen, and Mohamed Elgendi. PPG-BP database, 2022. URL https://figshare.com/articles/dataset/PPG-BP_Database_zip/5459299.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. iTransformer: Inverted transformers are effective for time series forecasting, 2023. URL https://arxiv.org/abs/2310.06625.

Cooper Lorsung and Amir Barati Farimani. PICL: Physics Informed Contrastive Learning for Partial Differential Equations, 2024. URL https://arxiv.org/abs/2401.16327. Version Number: 4.

Manuel Meier, Berken Utku Demirel, and Christian Holz. WildPPG: A real-world PPG dataset of long continuous recordings, 2024. URL https://arxiv.org/abs/2412.17540.

Guangkun Nie, Xiaocheng Fang, Gongzheng Tang, Yujie Xiao, Jun Li, Bo Liu, Hongyan Li, and Shenda Hong. AnyPPG: An ECG-guided PPG foundation model trained on over 100,000 hours of recordings for holistic health profiling, 2025. URL https://arxiv.org/abs/2511.01747.

Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers, 2022. URL https://arxiv.org/abs/2211.14730.

Van-Tu Ninh, Manh-Duy Nguyen, Sinéad Smyth, Minh-Triet Tran, Graham Healy, Binh T. Nguyen, and Cathal Gurrin. An improved subject-independent stress detection model applied to consumer-grade wearable devices, 2022. URL https://arxiv.org/abs/2203.09663.

Daniel Ocampo, Daniela Posso, Reza Namakian, and Wei Gao. Adaptive loss weighting for machine learning interatomic potentials, 2024. URL https://arxiv.org/abs/2403.18122.

Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding, 2018. URL https://arxiv.org/abs/1807.03748.

Sangjun Park, Dejiang Zheng, and Uichin Lee. GalaxyPPG: A PPG signal dataset collected in semi-naturalistic settings using galaxy watch, 2025. URL https://zenodo.org/doi/10.5281/zenodo.14635823.

Arvind Pillai, Dimitris Spathis, Fahim Kawsar, and Mohammad Malekzadeh. PaPaGei: Open foundation models for optical physiological signals, 2024. URL https://arxiv.org/abs/2410.20542.

Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019. doi: 10.1016/j.jcp.2018.10.045.

Attila Reiss, Ina Indlekofer, Philip Schmidt, and Kristof Van Laerhoven. Deep PPG: Large-scale heart rate estimation with convolutional neural networks. *Sensors*, 19(14):3079, 2019. ISSN 1424-8220. doi: 10.3390/s19143079. URL https://www.mdpi.com/1424-8220/19/14/3079.

Carlos Riquelme, Joan Puigcerver, Basil Mustafa, Maxim Neumann, Rodolphe Jenatton, André Susano Pinto, Daniel Keysers, and Neil Houlsby. Scaling vision with sparse mixture of experts, 2021. URL https://arxiv.org/abs/2106.05974.

Mithun Saha, Maxwell A. Xu, Wanting Mao, Sameer Neupane, James M. Rehg, and Santosh Kumar. Pulse-PPG: An open-source field-trained PPG foundation model for wearable applications across lab and field settings, 2025.

Philip Schmidt, Attila Reiss, Robert Duerichen, Claus Marberger, and Kristof Van Laerhoven. Introducing WESAD, a multimodal dataset for wearable stress and affect detection. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, pp. 400–408. ACM, 2018. ISBN 978-1-4503-5692-3. doi: 10.1145/3242969.3242985. URL https://dl.acm.org/doi/10.1145/3242969.3242985.

Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer, 2017. URL https://arxiv.org/abs/1701.06538.

Debaditya Shome, Pritam Sarkar, and Ali Etemad. Region-disentangled diffusion model for high-fidelity PPG-to-ECG translation, 2023. URL https://arxiv.org/abs/2308.13568.

Juan Vazquez-Rodriguez, Grégoire Lefebvre, Julien Cumin, and James L. Crowley. Transformer-based self-supervised learning for emotion recognition. 2022. doi: 10.48550/ARXIV.2204.05103. URL https://arxiv.org/abs/2204.05103.

Mason Wang and Cheng-Zhi Anna Huang. Latent Fourier transform, 2026. URL https://arxiv.org/abs/2604.17986. ICLR 2026 Oral.

Qi Wang, Hanyang Peng, and Yue Yu. Symphony-MoE: Harmonizing disparate pre-trained models into a coherent mixture-of-experts, 2025a. URL https://arxiv.org/abs/2509.18542.

Rui Wang, Mengshi Qi, Yingxia Shao, Anfu Zhou, and Huadong Ma. PITN: Physics-informed temporal networks for cuffless blood pressure estimation. *IEEE Transactions on Mobile Computing*, pp. 1–14, 2025b. ISSN 1536-1233, 1558-0660, 2161-9875. doi: 10.1109/TMC.2025.3627091. URL https://ieeexplore.ieee.org/document/11223230/.

Tongzhou Wang and Phillip Isola. Understanding contrastive representation learning through alignment and uniformity on the hypersphere, 2020. URL https://arxiv.org/abs/2005.10242.

Weinan Wang, Pedram Mohseni, Kevin L. Kilgore, and Laleh Najafizadeh. PulseDB: A large, cleaned dataset based on MIMIC-III and VitalDB for benchmarking cuff-less blood pressure estimation methods. *Frontiers in Digital Health*, 4:1090854, 2023. ISSN 2673-253X. doi: 10.3389/fdgth.2022.1090854. URL https://www.frontiersin.org/articles/10.3389/fdgth.2022.1090854/full.

Nico Westerhof, Jan-Willem Lankhaar, and Berend E. Westerhof. The arterial windkessel. *Medical & Biological Engineering & Computing*, 47(2):131–141, 2009. ISSN 0140-0118, 1741-0444. doi: 10.1007/s11517-008-0359-2. URL http://link.springer.com/10.1007/s11517-008-0359-2.

Gerald Woo, Chenghao Liu, Doyen Sahoo, Akshat Kumar, and Steven Hoi. CoST: Contrastive learning of disentangled seasonal-trend representations for time series forecasting, 2022. URL https://arxiv.org/abs/2202.01575.

Shanshan Xiao, Jiawei Zhang, and Yifa Tang. Generalized lagrangian neural networks, 2024. URL https://arxiv.org/abs/2401.03728.

Zhihan Yue, Yujing Wang, Juanyong Duan, Tianmeng Yang, Congrui Huang, Yunhai Tong, and Bixiong Xu. TS2vec: Towards universal representation of time series, 2021. URL https://arxiv.org/abs/2106.10466.

Xiang Zhang, Ziyuan Zhao, Theodoros Tsiligkaridis, and Marinka Zitnik. Self-supervised contrastive pre-training for time series via time-frequency consistency. *arXiv preprint arXiv:2206.08496*, 2022. doi: 10.48550/ARXIV.2206.08496. URL https://arxiv.org/abs/2206.08496.

Yali Zheng, Chen Wu, Peizheng Cai, Zhiqiang Zhong, Hongda Huang, and Yuqi Jiang. Tiny-PPG: A lightweight deep neural network for real-time detection of motion artifacts in photoplethysmogram signals on edge devices. *Internet of Things*, 25:101007, 2024. ISSN 25426605. doi: 10.1016/j.iot.2023.101007. URL https://linkinghub.elsevier.com/retrieve/pii/S254266052300330X.

A Implementation Details

The encoder receives a tensor in $\mathbb{R}^{B\times 1\times 1250}$. A one-dimensional ResNet-18 produces 512 temporal channels and a kernel-size-three convolution maps them to four latent coordinates. The smoothed differentiator uses a seven-tap Gaussian derivative kernel with standard deviation one and is applied independently to each latent channel. The numerical derivative uses a unit latent time step ($\Delta t = 1$); latent derivatives are not calibrated physical velocities in seconds. The finite filter uses same-length zero padding. The operator in Section 3 uses mathematical convolution; a cross-correlation implementation requires reversed kernel order to preserve the derivative sign convention. The potential and mass networks are two-hidden-layer MLPs with 128 SiLU units per layer; the force network has the same hidden width and receives the concatenation of position and velocity. Positive mass uses a softplus transform plus 10^{-4}. The contrastive head applies temporal mean pooling, a two-layer MLP with 128 hidden units, batch normalization and ReLU, and a 128-dimensional output. Gaussian perturbation with standard deviation 0.05 is independently applied to each view with probability 0.5.

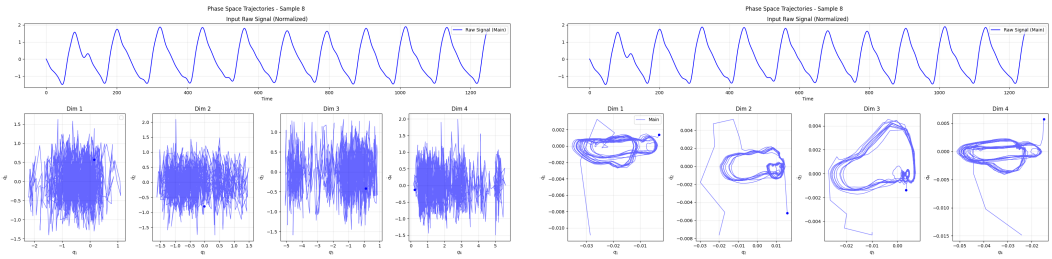
For downstream regression, the common temporal probe uses three one-dimensional convolutional blocks with 32, 64, and 128 channels, followed by adaptive average pooling and a 128–64–1 regressor with dropout 0.2. Classification replaces the last unit with the task's class count. Model-specific adapters are adjusted to match the total trainable downstream parameter count. Regression uses Huber loss; classification uses class-balanced cross entropy. The optimizer is AdamW with learning rate 10^{-3} and weight decay 10^{-4}, batch size 128, and 50 epochs. Three non-test subjects are reserved for validation where the split permits; the best checkpoint is selected by validation MAE for regression and validation accuracy for classification. The test set does not determine checkpoint selection.

B Matched Physics Ablation with Uncertainty

Table 6: Matched no-physics contrastive control versus PhysPPG. Values are mean ±95% CI half-width over held-out subjects/folds, rounded to two decimals. Arrows indicate the preferred direction.

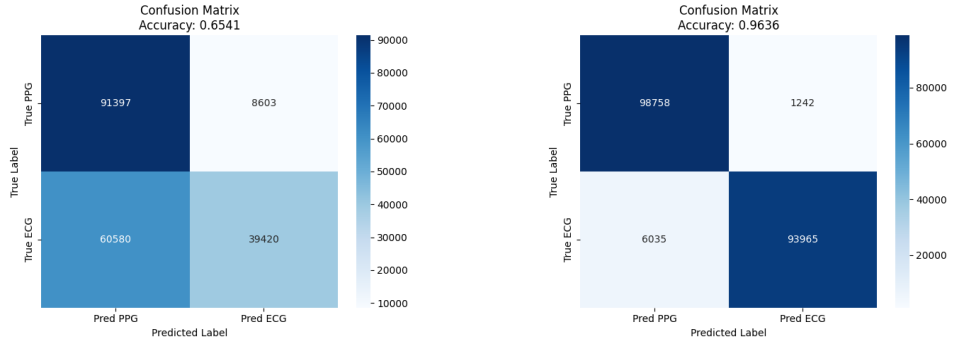
Task	Metric	SimCLR	PhysPPG	Task	Metric	SimCLR	PhysPPG
PPG-BP SBP	MAE ↓	14.72 ± 3.56	14.56 ± 3.18	PPG-BP DBP	MAE ↓	9.74 ± 1.85	8.88 ± 2.01
	RMSE ↓	18.59 ± 4.15	18.27 ± 3.34		RMSE ↓	12.27 ± 2.29	11.16 ± 2.41
	r ↑	0.44 ± 0.17	0.45 ± 0.13		r ↑	0.22 ± 0.15	0.29 ± 0.16
	R^2 ↑	0.08 ± 0.26	0.12 ± 0.11		R^2 ↑	-0.42 ± 0.40	-0.17 ± 0.36
PulseDB SBP	MAE ↓	15.51 ± 1.10	15.32 ± 0.97	PulseDB DBP	MAE ↓	6.02 ± 0.13	5.75 ± 0.10
	RMSE ↓	19.72 ± 1.21	19.17 ± 1.04		RMSE ↓	10.12 ± 0.10	9.90 ± 0.13
	r ↑	0.58 ± 0.04	0.61 ± 0.04		r ↑	0.79 ± 0.03	0.79 ± 0.03
	R^2 ↑	0.27 ± 0.07	0.31 ± 0.05		R^2 ↑	0.59 ± 0.04	0.60 ± 0.04
PPG-DaLiA HR	MAE ↓	9.81 ± 1.65	9.62 ± 1.55	WildPPG HR	MAE ↓	25.19 ± 1.95	24.39 ± 1.86
	RMSE ↓	12.84 ± 1.79	12.70 ± 1.77		RMSE ↓	31.42 ± 1.99	30.52 ± 1.76
	r ↑	0.61 ± 0.09	0.61 ± 0.09		r ↑	0.13 ± 0.05	0.15 ± 0.05
	R^2 ↑	0.19 ± 0.23	0.20 ± 0.23		R^2 ↑	-0.29 ± 0.14	-0.21 ± 0.08
WESAD arousal	Acc. ↑	0.64 ± 0.09	0.67 ± 0.09	WESAD valence	Acc. ↑	0.56 ± 0.13	0.55 ± 0.10
	F1 ↑	0.42 ± 0.06	0.44 ± 0.05		F1 ↑	0.34 ± 0.08	0.36 ± 0.07

C QUALITATIVE LATENT DYNAMICS



(a) Matched no-physics control (b) PHYSPPG

Figure 3: Illustrative phase portraits $(q_j, \dot{q}_j)$ for one PPG segment. Physics regularization produces smoother recurring trajectories than the no-physics control. This single example is qualitative and is not used as quantitative evidence of physiological identifiability.



(a) No-physics residual classifier (b) PHYSPPG residual classifier

Figure 4: Confusion matrices at the operating thresholds shown in Fig. 2. Accuracy increases from 0.654 for the no-physics control to 0.964 for PHYSPPG. Threshold-independent AUROC is the primary metric.

D LAGRANGIAN DERIVATION OF REPRESENTATIVE WINDKESSEL MODELS

Windkessel models describe lumped arterial energy storage and dissipation through compliance, resistance, and, in some extensions, inertance (Westerhof et al., 2009). Here we derive representative circuits using the Lagrange–d'Alembert equation with Rayleigh dissipation. This establishes a structural connection to the generalized Euler–Lagrange formalism, not an identification of the learned PPG latents with a particular cardiovascular model.

D.1 COORDINATES, ASSUMPTIONS, AND ENERGY BALANCE

To distinguish physical coordinates from the learned $\mathbf{q}$, let u_1 and u_2 be cumulative inlet and peripheral outlet volumes, with $Q = \dot{u}_1$ and $Q_{\text{out}} = \dot{u}_2$. Their difference $v = u_1 - u_2$ is the volume stored in the compliance relative to a reference state. Let $p_c = v/C$ be the compliant-node pressure, and let P be inlet pressure; both are measured relative to a constant downstream reference pressure. Assume constant $C > 0$, peripheral resistance $R_2 > 0$, proximal resistance $R_1 \geq 0$, and inertance $I \geq 0$. The symbol I avoids confusion with the Lagrangian $\mathcal{L}$.

The circuit consists of a proximal series resistance and, when present, a series inertance, feeding a parallel compliance–peripheral-resistance load. The energy and Rayleigh function are

$$T = \frac{1}{2} I \dot{u}_1^2, \qquad V = \frac{(u_1 - u_2)^2}{2C}, \qquad \mathcal{D} = \frac{1}{2} R_1 \dot{u}_1^2 + \frac{1}{2} R_2 \dot{u}_2^2. \tag{12}$$

Here T and V have units of pressure times volume, while $\mathcal{D}$ has units of power. The pressure input performs virtual work $P\,\delta u_1$. With $\mathcal{L} = T - V$, the equations are

$$\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{u}_i} - \frac{\partial \mathcal{L}}{\partial u_i} + \frac{\partial \mathcal{D}}{\partial \dot{u}_i} = F_i, \qquad (F_1, F_2) = (P, 0). \tag{13}$$

The partial derivative with respect to position holds velocity fixed. Substitution gives

$$I\ddot{u}_1 + p_c + R_1 \dot{u}_1 = P, \qquad -p_c + R_2 \dot{u}_2 = 0. \tag{14}$$

Since $C\dot{p}_c = \dot{u}_1 - \dot{u}_2$, these equations are equivalent to

$$P = I\dot{Q} + R_1 Q + p_c, \qquad C\dot{p}_c = Q - \frac{p_c}{R_2}. \tag{15}$$

Pressure and flow are conjugate port variables: either can be prescribed as a boundary input, with the other determined by the model. They are not arbitrary independent inputs.

The two volume coordinates do not imply two inertial degrees of freedom. Their kinetic Hessian is $\mathrm{diag}(I, 0)$; it vanishes when $I = 0$. These are singular Lagrangian descriptions with algebraic constraints. A common constant shift of u_1, u_2 leaves all observables unchanged. Such coordinate redundancy is distinct from collapse of a learned representation.

D.2 TWO-ELEMENT WINDKESSEL

Set $I = R_1 = 0$ and $R_2 = R$. Then $T = 0$, $\mathcal{L} = -(u_1 - u_2)^2/(2C)$, and $\mathcal{D} = R\dot{u}_2^2/2$. Equation (14) reduces to

$$\frac{u_1 - u_2}{C} = P, \qquad R\dot{u}_2 = P. \tag{16}$$

Differentiating $u_2 = u_1 - CP$ yields $\dot{u}_2 = Q - C\dot{P}$. Consequently,

$$\boxed{Q = \frac{P}{R} + C\dot{P}.} \tag{17}$$

This is the parallel resistance–compliance Windkessel relation.

D.3 THREE-ELEMENT WINDKESSEL

Set $I = 0$ and retain $R_1 > 0$, representing proximal characteristic impedance by a resistive element. Now $P = p_c + R_1 Q$, whereas the peripheral balance remains $C\dot{p}_c = Q - p_c/R_2$. Thus

$$p_c = P - R_1 Q, \qquad \dot{p}_c = \dot{P} - R_1 \dot{Q}, \tag{18}$$

$$C(\dot{P} - R_1 \dot{Q}) = Q - \frac{P - R_1 Q}{R_2}. \tag{19}$$

Rearrangement recovers the standard three-element relation:

$$\boxed{\left(1+\frac{R_1}{R_2}\right)Q + CR_1\dot{Q} = \frac{P}{R_2} + C\dot{P}.} \tag{20}$$

D.4 Four-element extension with series inertance

For $I > 0$, the inlet inertance stores kinetic energy $IQ^2/2$. From Eq. (15),

$$p_c = P - R_1 Q - I\dot{Q}, \qquad \dot{p}_c = \dot{P} - R_1\dot{Q} - I\ddot{Q}. \tag{21}$$

Inserting these expressions into the compliant-node balance gives

$$C(\dot{P} - R_1\dot{Q} - I\ddot{Q}) = Q - \frac{P - R_1 Q - I\dot{Q}}{R_2}, \tag{22}$$

and hence

$$\boxed{\left(1+\frac{R_1}{R_2}\right)Q + \left(CR_1 + \frac{I}{R_2}\right)\dot{Q} + IC\ddot{Q} = \frac{P}{R_2} + C\dot{P}.} \tag{23}$$

This equation is specific to the stated *series-inertance* topology. Four-element Windkessel formulations with different inertance placements need not have the same input–output equation. Setting $I = 0$ recovers Eq. (20), and additionally setting $R_1 = 0$ recovers Eq. (17).

D.5 What this connection supports in representation learning

For the circuit above, the stored energy $H = T + V = IQ^2/2 + Cp_c^2/2$ obeys

$$\dot{H} = Q(P - R_1 Q - p_c) + p_c(Q - p_c/R_2) \tag{24}$$

$$= PQ - R_1 Q^2 - \frac{p_c^2}{R_2} = PQ - 2\mathcal{D}. \tag{25}$$

Thus changes in stored energy equal supplied hydraulic power minus non-negative resistive dissipation. Moving the Rayleigh term to the right-hand side of Eq. (13) gives the form used in Eq. 7, with

$$\mathbf{F}_{\mathrm{nc}} = \mathbf{F}_{\mathrm{ext}} - \nabla_{\dot{\mathbf{u}}}\mathcal{D}. \tag{26}$$

This is the relevant bridge: classical lumped hemodynamics admits an energy–force description, making that description a physiologically motivated inductive bias for temporal representations.

There are important limits to this argument. PPG is an optical observation associated with peripheral blood-volume changes, not a direct measurement of P, Q, or $\mathbf{u}$ (Allen, 2007; Charlton et al., 2022). The encoder's coordinates are unconstrained learned features and need not be an invertible coordinate transformation of the Windkessel state. Moreover, the strictly positive diagonal latent mass differs from the singular mass matrices above, and the unconstrained force network is not restricted to a Rayleigh dissipative form. Equation (25) therefore does *not* establish passivity of the learned network. The derivation motivates energy storage, forcing, and dissipation as modeling ingredients; it does not prove that PhysPPG recovers Windkessel parameters or that the regularizer necessarily improves downstream transfer. Those are separate empirical questions.

E Non-Collapse Analysis and Latent-Magnitude Diagnostics

We analyze two relevant collapse signatures: complete representation collapse, in which all inputs have the same latent trajectory, and zero collapse, in which the latent magnitude vanishes. The contrastive objective directly penalizes complete collapse by assigning it a higher loss than a representation with positive separation between positive and negative pairs. We then use the latent-magnitude trace to verify that the learned trajectories remain away from the zero representation during training.

E.1 Cost of complete collapse under the contrastive objective

The contrastive embedding is computed from the same latent trajectory that enters the physics module, without an input-dependent bypass:

$$\bar{\mathbf{q}}_i = \frac{1}{T'} \sum_t \mathbf{q}_i(t), \qquad \mathbf{z}_i = P_\xi(\bar{\mathbf{q}}_i). \tag{27}$$

At complete collapse, $\mathbf{q}_i(t) = \mathbf{q}_0(t)$ for every view of every input, so a shared deterministic projection maps them to identical embeddings. This also holds for identical inputs to the projection head's batch normalization in a common batch. If the common embedding is zero, an epsilon-stabilized normalization still assigns equal logits to all candidates; ideal cosine similarity at zero is otherwise undefined.

For the standard InfoNCE/NT-Xent objective (Oord et al., 2018; Chen et al., 2020), a batch of $N \geq 2$ segments supplies $2N$ views. Each anchor has $K = 2N - 1$ candidates: one positive and $K - 1$ negatives, excluding itself. With temperature $\tau > 0$,

$$\ell_i = -\log \frac{\exp(s_i^+/\tau)}{\exp(s_i^+/\tau) + \sum_{j \in \mathcal{N}(i)} \exp(s_{ij}/\tau)}. \tag{28}$$

All similarities coincide at complete collapse. Therefore,

$$p_i^+ = \frac{1}{K}, \qquad \mathcal{L}_{\text{con}}^{\text{collapse}} = \log K. \tag{29}$$

Conditional loss-gap result. Suppose an attainable non-collapsed representation satisfies $s_i^+ - s_{ij} \geq \delta > 0$ for every negative of every anchor in the evaluated batch. Then

$$\ell_i = \log \left(1 + \sum_{j \in \mathcal{N}(i)} \exp((s_{ij} - s_i^+)/\tau) \right) \tag{30}$$

$$\leq \log \left(1 + (K-1) e^{-\delta/\tau} \right) < \log K. \tag{31}$$

The positive-alignment example $s_i^+ = 1$ and $s_{ij} \leq a < 1$ corresponds to $\delta = 1 - a$. The same bound applies to the batch average. Thus, whenever positive pairs achieve a non-zero similarity margin over negatives, complete collapse has strictly higher contrastive loss and is not a minimizer of the contrastive objective. This agrees with the alignment–uniformity perspective (Wang & Isola, 2020).

Candidate counting matters. Equation (29) assumes one occurrence of the positive in the denominator. More generally, with ν identical positive-logit occurrences, one designated numerator, and M negative logits, the collapsed loss is $\log(\nu + M)$ and the separated bound is $\log(\nu + Me^{-\delta/\tau})$. Thus a duplicated-positive variant ($\nu = 2$, $M = 2N - 2$) has collapsed loss $\log(2N)$, not $\log(2N - 1)$. The exact numerical bound must follow the implemented candidate construction; duplication does not make the variant identical to standard NT-Xent.

E.2 Extension to the joint objective

Let superscripts c and n denote a collapsed and an attainable non-collapsed model, respectively, including their dynamics and projection parameters. For Eq. 10,

$$\mathcal{L}_{\text{total}}^{\text{n}} - \mathcal{L}_{\text{total}}^{\text{c}} = -(\log K - \mathcal{L}_{\text{con}}^{\text{n}}) + \lambda_{\text{phy}}(\mathcal{L}_{\text{phy}}^{\text{n}} - \mathcal{L}_{\text{phy}}^{\text{c}}). \tag{32}$$

Consequently the non-collapsed model is preferred whenever

$$\boxed{\log K - \mathcal{L}_{\text{con}}^{\text{n}} > \lambda_{\text{phy}}(\mathcal{L}_{\text{phy}}^{\text{n}} - \mathcal{L}_{\text{phy}}^{\text{c}}).} \tag{33}$$

If $\mathcal{L}_{\text{phy}}^{\text{n}} - \mathcal{L}_{\text{phy}}^{\text{c}} \leq 0$, any positive contrastive gap suffices for $\lambda_{\text{phy}} \geq 0$. Otherwise, the admissible weight is smaller than the contrastive gap divided by the physics-loss increase. A sufficient comparison against a completely collapsed model is $\mathcal{L}_{\text{con}}^{\text{n}} + \lambda_{\text{phy}} \mathcal{L}_{\text{phy}}^{\text{n}} < \log K$, since squared physics residuals are non-negative. These statements use the objective convention $\lambda_{\text{con}} = 1$ and the reported auxiliary weight $\lambda_{\text{phy}} = 0.1$.

E.3 EMPIRICAL LATENT-MAGNITUDE DIAGNOSTIC

For a mini-batch trajectory tensor with dimensions $B \times d \times T'$, define

$$Q_{\mathrm{mag}} = \frac{1}{BdT'} \sum_{b,j,t} |q_{b,j,t}|. \tag{34}$$

Complete zero collapse implies $Q_{\mathrm{mag}} \to 0$. Figure 5 shows that Q_{mag} remains stably away from zero throughout the recorded training steps, with values approximately between 0.75 and 1.15. The learned latent trajectories therefore maintain a non-vanishing amplitude during training.

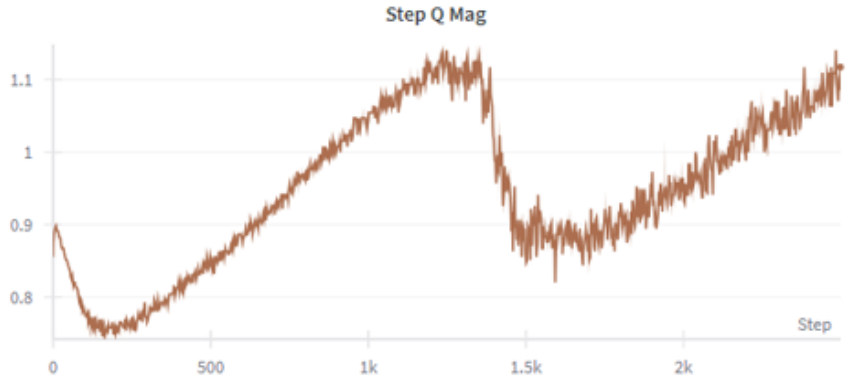


Figure 5: Training trace of the latent-magnitude statistic Q_{mag}. The statistic remains stably non-zero throughout training, showing that the learned latent trajectories maintain a non-vanishing amplitude rather than collapsing to the zero representation.

For scale reference, if $q \sim \mathcal{N}(0, \sigma^2)$, then

$$\mathbb{E}|q| = \frac{2}{\sqrt{2\pi}\sigma} \int_0^\infty x e^{-x^2/(2\sigma^2)}\, dx = \sigma\sqrt{\frac{2}{\pi}}. \tag{35}$$

The standard-Gaussian value is approximately 0.798, providing a familiar reference scale for the observed trajectory. Together, the contrastive loss-gap analysis and the consistently non-zero Q_{mag} trace show that the learned representation avoids both complete identical-representation collapse and zero-amplitude collapse in the analyzed training run.