LEARNING TRANSFERABLE NOISE-ENVIRONMENT REPRESENTATIONS FOR CROSS-DOMAIN TIME-SERIES RESTORATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Time-series restoration under domain shift is challenging because degradation patterns vary across sensing environments, while conventional models learn a single mapping from corrupted to clean observations. We study whether observable background context can provide transferable representations of degradation conditions. We formulate restoration as noise-environment- conditioned learning and introduce PCD-Net, which encodes background segments into environment representations and learns a differentiable routing distribution over a prototype memory to condition waveform reconstruction through cross-attention. This separation between environment representation and restoration enables selective cross-domain adaptation, where degradation-related representations are updated together with a lightweight reconstruction adjustment module while the general restoration backbone is preserved. Experiments across seismic, mine microseismic, and industrial waveform domains show that the learned environment representations improve restoration under varying noise conditions and provide an effective basis for parameter-efficient cross-domain adaptation. These results suggest that explicitly modeling observable degradation context can enhance the transferability of time-series restoration models.

1 INTRODUCTION

Time-series restoration is fundamental to sensing systems that support event detection, anomaly identification, and safety assessment. However, sensing conditions are rarely stationary. Changes in sensors, operating states, acquisition environments, and external interference can substantially alter degradation patterns across domains (Bergen et al., 2019; Sheng et al., 2025; Zhang & Zhang, 2024; Liu et al., 2024; Ansari et al., 2024; Shi et al., 2025). A model trained under one environment may therefore suffer performance degradation when deployed under another, even when the underlying signal structure remains similar. As illustrated in Figure 1(a), conventional restoration methods typically learn a fixed corrupted-to-clean mapping and are sensitive to shifts in degradation distributions.

Most existing restoration approaches infer degradation characteristics directly from corrupted observations and optimize a global restoration function. Although deep restoration and representation learning methods have achieved strong performance through nonlinear modeling and transferable feature learning (Zamir et al., 2022; Zhou et al., 2021; Zhang et al., 2017; Yue et al., 2022; Batson & Royer, 2019; Wu et al., 2023), they generally do not explicitly model the environmental factors that generate degradation. As a result, their performance may deteriorate when the degradation distribution changes across sensing environments.

Continuous monitoring systems provide additional contextual information that is often overlooked. Background segments recorded before a target event contain limited target information but preserve local sensing characteristics, including interference patterns, sensor responses, and noise statistics (Sun et al., 2026; Yang et al., 2026). We therefore consider an event-centered continuous monitoring setting, where an event-free background segment is naturally available before restoration. Rather than treating this background observation as irrelevant noise, we use it as observable

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.

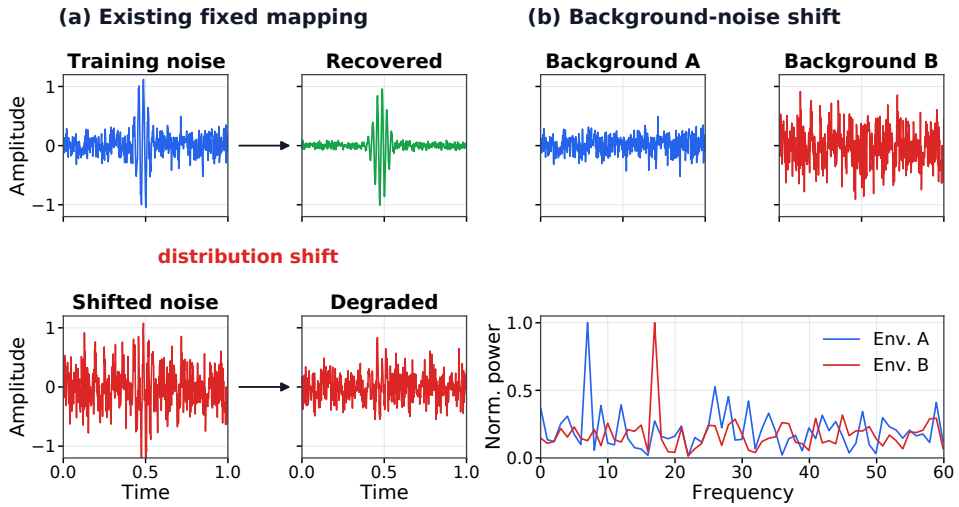


Figure 1: Motivation of noise-environment-conditioned restoration. (a) A conventional model learns a fixed noisy-to-clean mapping and may degrade when the noise environment changes. (b) Background segments from different environments exhibit distinct waveform and spectral characteristics, providing observable evidence of the current noise environment.

evidence of the current degradation environment and formulate restoration as a noise-environment-conditioned learning problem.

A key challenge is how to represent such environmental context in a transferable form. Directly concatenating background features with corrupted observations does not explicitly separate environment-dependent degradation from restoration knowledge. Inspired by prototype and memory-based representation learning (van den Oord et al., 2017; Santoro et al., 2016; Chen et al., 2020; He et al., 2020; Snell et al., 2017), we introduce prototypes that characterize latent degradation environments and learn a differentiable routing mechanism to dynamically condition restoration.

Based on this idea, we introduce **PCD-Net** (Prototype-Conditioned Denoising Network), a framework for learning transferable noise-environment representations for cross-domain time-series restoration. PCD-Net encodes background observations into environment descriptors, learns a differentiable routing distribution over prototype memory, and injects environment information into waveform reconstruction through prototype-conditioned cross-attention. This design separates environment-dependent degradation representations from transferable waveform restoration knowledge.

The same separation enables parameter-efficient cross-domain adaptation. Instead of updating the entire model under domain shift, PCD-Net adapts environment-related components, including the context encoder, prototype memory, and routing mechanism, while preserving the general restoration backbone. We evaluate PCD-Net on seismic, mine microseismic, and industrial waveform datasets with substantially different noise characteristics.

Our main contributions are:

- We formulate cross-domain time-series restoration as **noise-environment-conditioned learning**, using naturally available background observations as explicit evidence of the degradation process.
- We introduce a **differentiable prototype-conditioned environment representation**, where background-derived routing distributions dynamically condition waveform restoration through cross-attention.
- We develop a **parameter-efficient selective adaptation strategy** that updates domain-specific degradation representations while preserving transferable restoration knowledge.

2 Related Work

2.1 Time-Series Restoration and Representation Learning

Time-series restoration has traditionally relied on signal-processing assumptions, such as spectral separation and handcrafted filtering strategies. Although effective under stationary conditions, these approaches often struggle when noise characteristics vary across sensing environments (Huang et al., 1998; Vincent et al., 2008). Deep learning methods learn nonlinear restoration mappings from corrupted observations and have achieved improved waveform reconstruction using convolutional, recurrent, and attention-based architectures (Zhang et al., 2017; Lehtinen et al., 2018; Zhu et al., 2019; Luo et al., 2020; Zamir et al., 2022; Batson & Royer, 2019).

Recent studies have shifted from task-specific restoration toward learning generalizable temporal representations. Self-supervised approaches, including contrastive and predictive learning, learn transferable representations without requiring extensive clean annotations (Yue et al., 2022; Eldele et al., 2021; Chen et al., 2020; He et al., 2020; Tonekaboni et al., 2021; Woo et al., 2022). Large-scale time-series models further explore pretraining strategies for transferable sequence representations (Liu et al., 2024; Ansari et al., 2024; Shi et al., 2025; Zhang & Zhang, 2024; Zhou et al., 2021). However, these methods mainly focus on temporal structures and do not explicitly model environmental factors causing degradation. PCD-Net addresses this limitation by learning degradation environment representations from observable background context.

2.2 Domain Adaptation under Distribution Shift

Domain adaptation improves model generalization when training and deployment distributions differ. Existing approaches commonly reduce distribution gaps through feature alignment, adversarial learning, or test-time adaptation strategies (Sun & Saenko, 2016; Wang et al., 2021; Ganin et al., 2016; He et al., 2023). Recent time-series studies further investigate domain-specific representation learning and adaptation under heterogeneous temporal distributions (Deng et al., 2024; Kim et al., 2025; Patel et al., 2025).

However, most adaptation methods treat domain shift as a general feature distribution discrepancy and do not explicitly identify the source of degradation variation. In sensing systems, domain shifts are often associated with changes in noise conditions, sensor responses, and operating environments. PCD-Net separates environment-dependent degradation representations from transferable restoration knowledge and performs selective adaptation of environment-related components with lightweight reconstruction adjustment.

2.3 Prototype-Based Environment Representation Learning

Prototype-based learning provides compact representations by organizing recurring patterns in latent space. Prototypical Networks use reference embeddings for representation comparison, while memory-based architectures store learnable patterns for adaptive inference (Snell et al., 2017; van den Oord et al., 2017; Santoro et al., 2016).

Recent restoration studies have explored prototype-based degradation modeling. For example, dual-level prototype learning represents complex degradations through multiple prototype representations and integrates them into Transformer-based restoration modules (Wang et al., 2025). Degradation-aware adaptation methods further investigate selective updating of degradation-related parameters to preserve pretrained knowledge (Tang et al., 2026). Nevertheless, existing prototype-based approaches typically infer degradation patterns from corrupted observations or focus on task-specific clusters. Different from these methods, PCD-Net learns degradation representations from naturally available background observations and introduces differentiable routing over prototype memory to condition restoration under changing sensing environments.

3 METHODOLOGY

3.1 PROBLEM FORMULATION

Let $x \in \mathbb{R}^{C\times T}$ denote a corrupted time-series segment and $y \in \mathbb{R}^{C\times T}$ its corresponding clean signal, where C and T represent the number of channels and temporal samples, respectively.

Most existing restoration methods learn a direct mapping from corrupted observations to clean signals:

$$\hat{y} = G_\phi(x), \tag{1}$$

where G_ϕ denotes the restoration model. Such a formulation assumes that the degradation patterns observed during training remain consistent during deployment. However, practical sensing systems often operate under changing conditions, where variations in sensors, acquisition environments, and external interference introduce different degradation distributions. Consequently, a fixed corrupted-to-clean mapping may fail to generalize across domains.

In continuous monitoring scenarios, background observations before the target event are often naturally available. We denote the background segment as $b \in \mathbb{R}^{C\times T_b}$. Although b contains limited information about the target signal itself, it reflects the local sensing condition and provides observable evidence of the degradation environment affecting x.

Therefore, we formulate time-series restoration as a noise-environment-conditioned learning problem:

$$\hat{y} = F_\theta(x, b), \tag{2}$$

where the background segment is introduced as environmental context to guide waveform restoration.

The objective of PCD-Net is not to directly reconstruct the target signal from the background observation, but to learn a transferable representation of the underlying degradation environment and use it to condition restoration. Different from conventional approaches that learn a single global restoration mapping, PCD-Net explicitly models environment-dependent degradation variations and learns how such variations can guide adaptive waveform reconstruction.

3.2 NOISE-ENVIRONMENT CONDITIONED RESTORATION

PCD-Net learns transferable noise-environment representations during source-domain training and uses them to condition waveform restoration across different domains. As illustrated in Figure 2, the framework consists of two interacting pathways: an environment representation pathway and a waveform restoration pathway.

The environment representation pathway extracts degradation-related information from pre-event background observations. Given a background segment b, the context encoder produces an environment descriptor:

$$z_b = E_b(b), \tag{3}$$

where $E_b(\cdot)$ encodes characteristics of the current sensing condition. Unlike conventional restoration methods that infer degradation information only from corrupted observations, PCD-Net explicitly models the observable environmental context associated with signal degradation.

To represent recurring degradation patterns across different sensing environments, we introduce a learnable prototype memory:

$$P = \{p_1, p_2, \ldots, p_K\}, \tag{4}$$

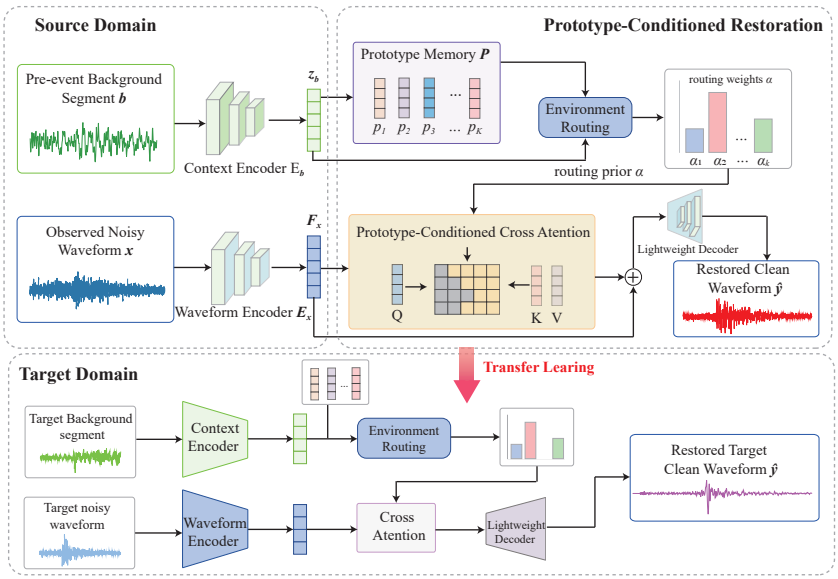


Figure 2: Overview of PCD-Net. Background features are encoded into environment representations and routed through a differentiable prototype memory to guide prototype-conditioned restoration. During target adaptation, environment-related modules and a lightweight reconstruction adjustment module are updated while the general restoration backbone remains fixed.

where each prototype represents a latent noise-environment characteristic rather than a semantic category. These prototypes provide reusable environmental representations that can be transferred across domains.

Given the environment descriptor z_b, PCD-Net first computes the similarity between the current environment and each prototype:

$$s_k = \frac{z_b^T p_k}{\|z_b\|_2 \|p_k\|_2}, \tag{5}$$

where s_k denotes the compatibility between the observed background condition and the k-th prototype.

Instead of performing non-differentiable hard prototype selection, we introduce a differentiable environment routing mechanism. The similarity scores are converted into a routing distribution over all prototypes:

$$\alpha_k = \frac{\exp(s_k/\tau)}{\sum_{j=1}^{K} \exp(s_j/\tau)}, \tag{6}$$

where τ is a temperature parameter controlling the sharpness of the routing distribution. The routing weights $\alpha = \{\alpha_1, \ldots, \alpha_K\}$ provide a continuous representation of the current degradation environment and allow gradients from the restoration objective to propagate through the environment representation pathway. Because the routing distribution assigns a non-zero weight to every prototype, all prototypes receive gradients and none collapse to a single degenerate state.

Meanwhile, the waveform restoration pathway extracts hierarchical features from the corrupted waveform:

$$F_x = E_x(x), \tag{7}$$

where $E_x(\cdot)$ denotes the waveform encoder.

Rather than directly concatenating background features with waveform features, PCD-Net introduces environment-conditioned prototype interactions through cross-attention. The waveform features act as queries, while the prototype memory provides keys and values:

$$Q = W_q F_x, \quad K = W_k P, \quad V = W_v P. \tag{8}$$

To incorporate the environment routing prior into the prototype interaction, we modify the attention operation as:

$$A = \mathrm{Softmax}\left(\frac{QK^T}{\sqrt{d}} + \lambda_r \log(\alpha + \epsilon)\right) V, \tag{9}$$

where d denotes the feature dimension, λ_r controls the contribution of the environment routing prior, and ϵ is a small constant for numerical stability. Since the attention logits have a temporal-prototype structure, the routing prior is broadcast along the temporal dimension before being added to the attention logits. The routing distribution biases the attention operation toward prototypes consistent with the current degradation condition while maintaining a fully differentiable prototype conditioning process.

The attended environment information is integrated with waveform features and passed to the reconstruction decoder:

$$\hat{y} = D(F_x + A), \tag{10}$$

where $D(\cdot)$ denotes the reconstruction decoder.

Through this design, PCD-Net explicitly separates environment-dependent degradation representations from transferable waveform restoration knowledge. The context encoder, prototype memory, and routing mechanism learn how degradation conditions vary across sensing environments, while the restoration pathway preserves general waveform recovery ability. A lightweight decoder calibration module further allows efficient adjustment of target-domain reconstruction characteristics without updating the entire restoration backbone.

3.3 Selective Environment Representation Adaptation

Although PCD-Net learns environment-conditioned restoration from the source domain, degradation characteristics may vary across different sensing environments. Changes in acquisition systems, operating conditions, and background interference can introduce domain-specific variations. Directly fine-tuning the entire restoration network may modify transferable waveform reconstruction knowledge and increase the risk of overfitting when only limited target-domain observations are available.

To address this issue, PCD-Net performs selective cross-domain adaptation by updating environment-related representations together with a small number of reconstruction adjustment parameters, while preserving the general restoration backbone. We decompose the model parameters into three groups:

$$\theta = \theta_{\mathrm{rec}} \cup \theta_{\mathrm{env}} \cup \theta_{\mathrm{cal}}, \tag{11}$$

where θ_{rec} denotes transferable restoration parameters, θ_{env} represents environment adaptation parameters, and θ_{cal} denotes lightweight reconstruction adjustment parameters.

Specifically, θ_{rec} includes the waveform encoder E_x and the main decoder backbone D. These components capture general waveform reconstruction knowledge learned from the source domain and remain fixed during target adaptation:

$$\theta^t_{\mathrm{rec}} = \theta^{st}_{\mathrm{rec}}. \tag{12}$$

In contrast, θ_{env} consists of the context encoder E_b, prototype memory P, and routing-related parameters. These components characterize the degradation environment of the target domain and are updated using target-domain restoration objectives:

$$\theta_{\text{env}}^{t} = \theta_{\text{env}}^{st} - \eta \nabla_{\theta_{\text{env}}} \mathcal{L}_t, \tag{13}$$

where θ^{st} denotes the source-domain training checkpoint, η represents the adaptation learning rate, and $\mathcal{L}_t$ denotes the target-domain restoration objective.

To further compensate for target-specific reconstruction variations without modifying the complete decoder, PCD-Net introduces a lightweight decoder calibration module. The calibration parameters θ_{cal} are inserted into the reconstruction pathway as low-dimensional feature adapters, which adjust target-domain restoration characteristics while preserving the pretrained decoder backbone. The adapter parameters are optimized as:

$$\theta_{\text{cal}}^{t} = \theta_{\text{cal}}^{st} - \eta \nabla_{\theta_{\text{cal}}} \mathcal{L}_t. \tag{14}$$

Because the proposed environment routing mechanism is fully differentiable, the target-domain restoration loss can directly update the context encoder, prototype memory, routing mechanism, and lightweight reconstruction adjustment module. Consequently, PCD-Net adapts both the degradation representation and the target-specific reconstruction characteristics without relearning the entire restoration model.

By preserving the restoration backbone while selectively adapting environment-related representations and lightweight reconstruction adjustment parameters, PCD-Net separates transferable waveform recovery knowledge from domain-specific degradation variations. This strategy enables efficient cross-domain restoration with fewer trainable parameters compared with full model fine-tuning.

4 Experiments

4.1 Experimental Setup

Datasets. We evaluate PCD-Net on three heterogeneous waveform datasets. A public seismic dataset is used as the source domain, while mine microseismic and industrial waveform datasets are used as target domains for cross-domain evaluation.

Baselines. We compare PCD-Net with classical methods (Bandpass and Wavelet) and deep restoration models, including DeepDenoiser, DPRNN, and Restormer1D.

Metrics. We report ΔSNR, CC, RMSE, and ST-MAE to evaluate restoration quality.

Implementation Details. All experiments are conducted on an NVIDIA RTX 4060 Laptop GPU. PCD-Net is optimized using AdamW with cosine learning-rate decay. The prototype memory size and routing temperature are selected according to validation performance.

4.2 Restoration Performance and Environment Representation Analysis

We first evaluate whether modeling degradation environments improves restoration. Table 1 summarizes comparisons on the STEAD test set. Each experiment is repeated with five random seeds, and mean and standard deviation are reported.

PCD-Net achieves the best performance across all metrics. The improvement over fixed corrupted-to-clean mappings demonstrates that background-derived environment representations provide useful degradation information for adaptive restoration.

Figure 3 presents qualitative waveform and time-frequency comparisons. PCD-Net produces cleaner reconstructed signals while preserving important temporal structures under complex degradation conditions. These results demonstrate that environment-conditioned restoration provides more effective degradation-aware guidance compared with directly learning a global restoration mapping.

Table 1: Multi-seed stability on the STEAD test set.

Method	ΔSNR (dB)↑	CC↑	RMSE↓	ST-MAE↓
Bandpass	1.2459 ± 0.0867	0.5502 ± 0.0038	0.0912 ± 0.0006	0.0684 ± 0.0005
Wavelet	3.6416 ± 0.0074	0.5808 ± 0.0007	0.0660 ± 0.0001	0.0451 ± 0.0001
DeepDenoiser	2.0439 ± 0.0120	0.5418 ± 0.0003	0.0806 ± 0.0004	0.0568 ± 0.0003
DPRNN	2.2270 ± 0.0030	0.4035 ± 0.0005	0.0671 ± 0.0001	0.0359 ± 0.0000
Restormer1D	5.9341 ± 0.0300	0.6871 ± 0.0026	0.0478 ± 0.0003	0.0303 ± 0.0002
ProtoCond-TCN	6.1816 ± 0.0160	0.7264 ± 0.0005	0.0482 ± 0.0005	0.0325 ± 0.0005
PCD-Net	$\mathbf{7.7548 \pm 0.0134}$	$\mathbf{0.7633 \pm 0.0015}$	$\mathbf{0.0406 \pm 0.0004}$	$\mathbf{0.0272 \pm 0.0002}$

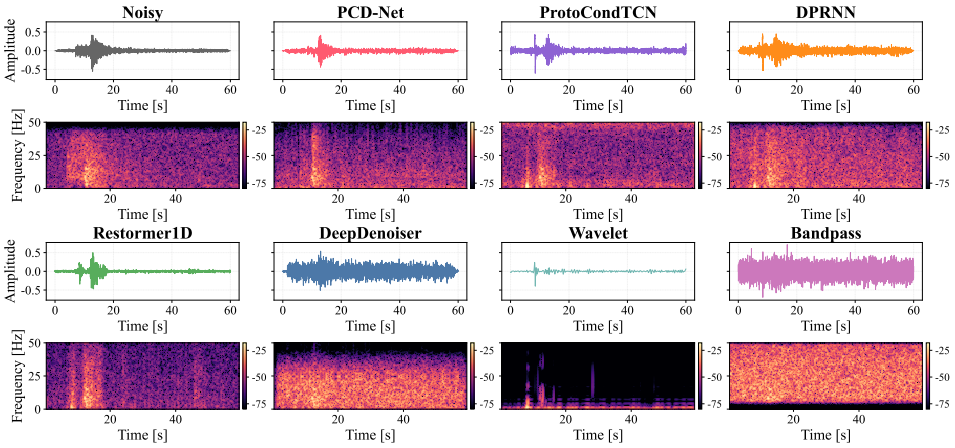


Figure 3: Qualitative comparison of waveform restoration results. PCD-Net better preserves waveform structures and removes environment-dependent interference under noisy conditions.

We further evaluate robustness under different input SNR ranges. As shown in Figure 4, PCD-Net consistently achieves the largest SNR improvement across all ranges. The larger improvement under low-SNR conditions indicates that environment representations provide more effective guidance when degradation is severe.

The ablation results in Table 2 verify the contribution of each component. Removing the background environment representation causes the largest performance drop, confirming that observable degradation context provides essential information for environment-conditioned restoration.

Removing differentiable routing further reduces performance, demonstrating that adaptive assignment of environment representations is important for effective prototype conditioning. The performance decrease after removing prototype memory indicates that learnable prototypes provide useful latent representations for capturing different degradation conditions.

Finally, removing the lightweight decoder adapter results in a smaller performance decrease. This indicates that the adapter provides additional target-domain reconstruction adjustment, while the primary improvement comes from environment-aware representation learning and routing.

4.3 Cross-Domain Adaptation Analysis

To evaluate the effectiveness of the proposed selective adaptation strategy under domain shift, we conduct additional experiments on the industrial target domain. Unlike source-domain restoration, target-domain deployment involves different sensing conditions and degradation characteristics. Therefore, we compare four adaptation strategies: frozen inference, lightweight reconstruction adjustment, selective adaptation, and full-model fine-tuning.

As shown in Table 3, the frozen source-domain model suffers from substantial performance degradation when directly transferred to the industrial target domain, indicating a significant mismatch between source and target degradation conditions. Updating only the lightweight reconstruction

Table 2: Ablation study of PCD-Net components.

Variant	ΔSNR	CC	RMSE	ST-MAE
w/o Background	5.7306	0.6628	0.0607	0.0474
w/o Routing	6.7970	0.6919	0.0605	0.0471
w/o Prototype	7.1653	0.7032	0.0406	0.0272
w/o Decoder Adapter	7.1170	0.7626	0.0406	0.0274
PCD-Net	**7.8554**	**0.7667**	**0.0402**	**0.0269**

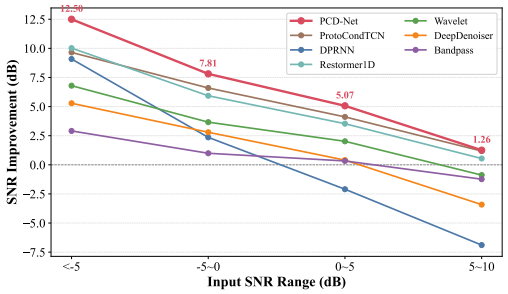


Figure 4: SNR improvement across different input-SNR ranges.

Table 3: Cross-domain adaptation strategies on the industrial target domain.

Strategy	ΔSNR	CC	RMSE	ST-MAE
Frozen	6.0261	0.6605	0.2500	0.1884
Decoder Adjustment	6.0915	0.6710	0.2490	0.1881
Selective Adaptation	9.6455	0.8001	0.1588	0.1203
Full Fine-tuning	**10.8398**	**0.8311**	**0.1389**	**0.1051**

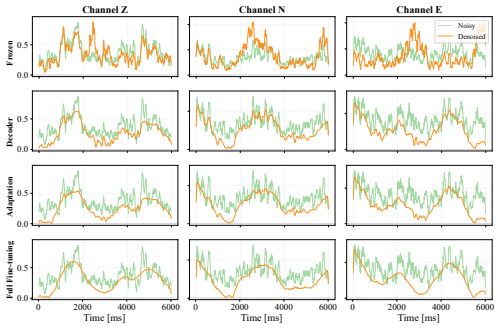


Figure 5: Temporal fidelity comparison on the industrial target domain measured by ST-MAE.

adjustment module provides limited improvement, suggesting that adapting reconstruction features alone is insufficient to address large degradation variations.

In contrast, selective adaptation, which updates environment-related representations together with lightweight reconstruction adjustment, achieves a substantial improvement over frozen inference and decoder-only adjustment. This result demonstrates that adapting degradation environment representations is important for capturing target-domain variations, while lightweight reconstruction adjustment provides additional refinement.

Full fine-tuning achieves the highest restoration performance since all model parameters are optimized for the target domain. However, selective adaptation preserves the transferable restoration backbone and updates only domain-relevant components, providing a more parameter-efficient adaptation strategy.

To further evaluate cross-domain waveform preservation, we report a temporal fidelity comparison on the industrial target domain. As shown in Figure 5, selective adaptation reduces ST-MAE compared with non-adapted models, indicating that environment-aware adaptation improves temporal structure preservation under domain shift. Additional sensitivity analysis of prototype number and routing temperature is provided in the appendix.

5 Conclusion

In this paper, we introduce PCD-Net, a framework for learning transferable noise-environment representations for cross-domain time-series restoration. We reformulate restoration as a noise-environment-conditioned learning problem, where pre-event background observations provide contextual evidence of the underlying degradation condition. PCD-Net learns compact degradation representations through a learnable prototype memory and employs a differentiable routing mechanism to dynamically condition waveform reconstruction via prototype-conditioned cross-attention. Furthermore, we introduce a parameter-efficient selective adaptation strategy that updates domain-specific environment representations together with lightweight reconstruction adjustment parameters, while preserving transferable restoration knowledge. Experiments on seismic, mine microseismic, and industrial waveform datasets demonstrate that the learned environment representations improve restoration quality and enable effective adaptation across heterogeneous noise environments.

AI USE STATEMENT

In this work, we used generative AI tools for language polishing of the manuscript. We have not used generative AI tools for data collection. We have reviewed all AI-assisted work, and the technical content, experimental results, and claims in this paper were verified by the authors. We take responsibility for the final content of this work, including text, claims and artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

To ensure the reproducibility of the reported results, we provide the complete implementation details in the appendix. Appendix B specifies the datasets and their preprocessing (including waveform length and background segment configuration), the baseline models, the evaluation metrics, and the training setup such as optimizer, batch size, and hardware. Appendix A presents the full training objective, the quality-aware training target, and the adaptive inference refinement procedure. Algorithm 1 summarizes the selective target-domain adaptation strategy. Appendix C reports hyperparameter sensitivity and complexity analyses.

ETHICS STATEMENT

We have carefully reviewed the ICLR Code of Ethics and confirm that this work adheres to its principles. The model checkpoints and datasets utilized in our work are publicly available and standard benchmarks in the field. To the best of our knowledge, our research does not introduce new ethical concerns.

REFERENCES

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. *Transactions on Machine Learning Research*, 2024.

Joshua Batson and Loic Royer. Noise2self: Blind denoising by self-supervision. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pp. 524–533, 2019.

Karianne J. Bergen, Paul A. Johnson, Maarten V. de Hoop, and Gregory C. Beroza. Machine learning for data-driven discovery in solid earth geoscience. *Science*, 363:eaau0323, 2019. doi: 10.1126/science.aau0323.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pp. 1597–1607, 2020.

Songgaojun Deng, Olivier Sprangers, Ming Li, Sebastian Schelter, and Maarten de Rijke. Domain generalization in time series forecasting. *ACM Trans. Knowl. Discov. Data*, 18, 2024. ISSN 1556-4681. doi: 10.1145/3643035.

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. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, pp. 2352–2359, 2021. doi: 10.24963/ijcai.2021/324.

Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario March, and Victor Lempitsky. Domain-adversarial training of neural networks. In *JMLR*, 2016.

Huan He, Owen Queen, Teddy Koker, Consuelo Cuevas, Theodoros Tsiligkaridis, and M. Zitnik. Domain adaptation for time series under feature and label shifts. In *ICML*, 2023.

Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9729–9738, 2020.

Norden E. Huang, Zheng Shen, Steven R. Long, Manli C. Wu, Hsing H. Shih, Quanan Zheng, Nai-Chyuan Yen, Chi-Chao Tung, and Henry H. Liu. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society A*, 454:903–995, 1998. doi: 10.1098/rspa.1998.0193.

HyunGi Kim, Siwon Kim, Jisoo Mok, and Sungroh Yoon. Battling the non-stationarity in time series forecasting via test-time adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 17868–17876, 2025.

Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2Noise: Learning image restoration without clean data. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, pp. 2965–2974, 2018.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. iTransformer: Inverted transformers are effective for time series forecasting. In *International Conference on Learning Representations*, 2024.

Yi Luo, Zhuo Chen, and Takuya Yoshioka. Dual-path RNN: Efficient long sequence modeling for time-domain single-channel speech separation. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020.

Gaurav Patel, Christopher Sandino, Behrooz Mahasseni, Ellen L Zippi, Erdrin Azemi, Ali Moin, and Juri Minxha. Efficient source-free time-series adaptation via parameter subspace disentanglement. In *ICLR*, 2025.

Adam Santoro, Sergey Bartunov, Matthew Botvinick, Daan Wierstra, and Timothy Lillicrap. Meta-learning with memory-augmented neural networks. In *International Conference on Machine Learning*, pp. 1842–1850, 2016.

Guanqun Sheng, Junyi Pang, Wei Chen, and Xingong Tang. Dcaf-net: An intelligent denoising method for microseismic signals based on the combination of time and wavelet domains. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–17, 2025. doi: 10.1109/TGRS.2025.3592196.

Xiaoming Shi, Shiyu Wang, Yuqi Nie, Dianqi Li, Zhou Ye, Qingsong Wen, and Ming Jin. Time-MoE: Billion-scale time series foundation models with mixture of experts. In *International Conference on Learning Representations*, 2025.

Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *European Conference on Computer Vision Workshops*, pp. 443–450, 2016.

Wei Sun, Shengbao Yu, and Junqiu Wang. WTTnet: A network combining wavelet transform and transformer for denoising microseismic signal. *Frontiers in Earth Science*, 14:1790572, 2026.

Ni Tang, Shenghao Nie, Xiaotong Luo, Yuan Xie, and Yanyun Qu. Degradation-consistent test-time adaptation for all-in-one image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026.

Sana Tonekaboni, David Eytan, and Anna Goldenberg. Unsupervised representation learning for time series with temporal neighborhood coding. In *ICLR*, 2021.

Aaron van den Oord, Oriol Vinyals, and Koray Kavukcuoglu. Neural discrete representation learning. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th International Conference on Machine Learning*, pp. 1096–1103, 2008.

Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. In *International Conference on Learning Representations*, 2021.

Zhongze Wang, Haitao Zhao, Lujian Yao, Jingchao Peng, and Kaijie Zhao. Dual-level prototype learning for composite degraded image restoration. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2025. doi: 10.1109/ICCV51701.2025.01300.

Gerald Woo, Chenghao Liu, Doyen Sahoo, Akshat Kumar, and Steven Hoi. Cost: Contrastive learning of disentangled seasonal-trend representations for time series forecasting. In *ICLR*, 2022.

Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. TimesNet: Temporal 2d-variation modeling for general time series analysis. In *International Conference on Learning Representations*, 2023.

Zesheng Yang, Qingfeng Xue, Xiaoning Wang, and Tao Wang. TFDenoiser-Edge: A hybrid convolutional neural network–transformer framework for real-time seismic denoising on edge devices in extreme environments. *Journal of Seismic Exploration*, 35:025010138, 2026.

Zhihan Yue, Yujing Wang, Juanyong Duan, Tianmeng Yang, Congrui Huang, Yunhai Tong, and Bixiong Xu. TS2Vec: Towards universal representation of time series. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 8980–8987, 2022. doi: 10.1609/aaai.v36i8.20881.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*, 2022.

Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*, 26:3142–3155, 2017.

Xiong Zhang and Miao Zhang. Universal neural networks for real-time earthquake early warning trained with generalized earthquakes. *Communications Earth & Environment*, 5:528, 2024.

Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 11106–11115, 2021.

Weiqiang Zhu, S. Mostafa Mousavi, and Gregory C. Beroza. Seismic signal denoising and decomposition using deep neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 57 (11):9476–9488, 2019.

Appendix

A Additional Formulation Details

A.1 Differentiable Environment Routing

The main text introduces a differentiable routing mechanism over a learnable prototype memory to represent the current degradation environment. Given a background segment b, the context encoder produces an environment descriptor:

$$z_b = E_b(b), \tag{15}$$

where $E_b(\cdot)$ extracts degradation-related characteristics from the observed sensing environment.

The prototype memory is defined as:

$$\mathcal{P} = \{p_1, p_2, \ldots, p_K\}, \tag{16}$$

where each prototype represents a latent degradation environment rather than a semantic category. Given the environment descriptor z_b, PCD-Net computes the compatibility between the current environment and each prototype using cosine similarity:

$$s_k = \frac{z_b^T p_k}{\|z_b\|_2 \|p_k\|_2}, \tag{17}$$

where s_k denotes the similarity score between the observed background condition and the k-th prototype.

Different from hard prototype selection, which introduces a non-differentiable operation and prevents gradients from propagating through the prototype selection process, PCD-Net adopts a differentiable routing mechanism. The similarity scores are converted into a continuous routing distribution:

$$\alpha_k = \frac{\exp(s_k/\tau)}{\sum_{j=1}^{K} \exp(s_j/\tau)}, \tag{18}$$

where τ is a temperature parameter controlling the sharpness of the routing distribution. The routing weights $\alpha = \{\alpha_1, \ldots, \alpha_K\}$ provide a soft representation of the current degradation environment and enable end-to-end optimization of the environment representation pathway.

During prototype-conditioned cross-attention, the waveform features serve as queries, while the prototype memory provides keys and values:

$$Q = W_q F_x, \quad K = W_k \mathcal{P}, \quad V = W_v \mathcal{P}. \tag{19}$$

The routing distribution is incorporated into the attention logits:

$$A = \mathrm{Softmax}\left(\frac{QK^T}{\sqrt{d}} + \lambda_r \log(\alpha + \epsilon)\right) V, \tag{20}$$

where d denotes the feature dimension, λ_r controls the contribution of the routing prior, and ϵ is used for numerical stability.

Since the attention matrix has a temporal-prototype structure, the routing weights are broadcast along the temporal dimension before being added to the attention logits. This allows the restoration pathway to dynamically focus on prototype representations consistent with the current degradation environment while maintaining a fully differentiable optimization process.

A.2 Selective Adaptation Objective

During cross-domain adaptation, PCD-Net separates parameters according to their roles in restoration and environment adaptation:

$$\theta = \theta_{\mathrm{rec}} \cup \theta_{\mathrm{env}} \cup \theta_{\mathrm{cal}}, \tag{21}$$

where θ_{rec} denotes the transferable restoration backbone, θ_{env} represents environment adaptation parameters, and θ_{cal} denotes lightweight decoder reconstruction adjustment parameters.

Specifically, θ_{rec} contains the waveform encoder and the main decoder backbone. These components capture general waveform restoration knowledge learned from the source domain and remain fixed during target-domain adaptation:

$$\theta^{t}_{\mathrm{rec}} = \theta^{st}_{\mathrm{rec}}. \tag{22}$$

The environment-related parameters include the context encoder, prototype memory, and routing mechanism:

$$\theta_{\mathrm{env}} = \{E_b, \mathcal{P}, \theta_{\mathrm{route}}\}. \tag{23}$$

These parameters are updated to adapt the degradation representation to the target sensing environment:

$$\theta^{t}_{\mathrm{env}} = \theta^{st}_{\mathrm{env}} - \eta \nabla_{\theta_{\mathrm{env}}} \mathcal{L}_t, \tag{24}$$

where η denotes the adaptation learning rate and $\mathcal{L}_t$ represents the target-domain restoration objective.

To further compensate for target-specific reconstruction variations, PCD-Net introduces a lightweight decoder reconstruction adjustment module. Instead of updating the entire decoder, the reconstruction adjustment module performs low-dimensional feature adjustment with a small number of additional parameters:

$$\theta^{t}_{\mathrm{cal}} = \theta^{st}_{\mathrm{cal}} - \eta \nabla_{\theta_{\mathrm{cal}}} \mathcal{L}_t. \tag{25}$$

Therefore, selective adaptation updates environment representations and lightweight reconstruction adjustment while preserving the general restoration backbone. This design enables efficient adaptation under domain shift with fewer trainable parameters than full-model fine-tuning.

A.3 Lightweight Decoder reconstruction adjustment Module

The decoder reconstruction adjustment module is designed to adjust target-domain reconstruction characteristics without modifying the complete restoration decoder. Given decoder features h_d, the adapter applies a bottleneck transformation:

$$h^{'}_{d} = h_d + W_{up}\sigma(W_{down}h_d), \tag{26}$$

where W_{down} and W_{up} project the decoder features into and back from a lower-dimensional latent space, respectively, and $\sigma(\cdot)$ denotes the nonlinear activation function.

The bottleneck dimension is smaller than the original feature dimension, allowing the adapter to introduce only a limited number of additional trainable parameters. During adaptation, only the adapter parameters are optimized, while the original decoder backbone remains unchanged.

This lightweight reconstruction adjustment mechanism preserves the transferable restoration capability learned from the source domain while allowing efficient adjustment to target-domain reconstruction characteristics.

A.4 Hyperparameter Analysis

We further analyze the influence of the prototype memory size K and routing temperature τ on restoration performance. The prototype number controls the capacity of the degradation environment representation space, while the temperature determines the sharpness of the routing distribution.

Different values of K and τ are evaluated on the validation set. The results demonstrate that PCD-Net maintains stable performance across a range of configurations, indicating that the proposed differentiable environment routing mechanism is robust to reasonable hyperparameter variations.

B Experimental Setup

B.1 Datasets

We evaluate PCD-Net on three heterogeneous one-dimensional waveform datasets: the STEAD earthquake dataset, a mine microseismic dataset, and an industrial non-natural waveform dataset.

STEAD. STEAD is used as the source-domain dataset for pretraining. It contains a large number of three-component earthquake waveforms with diverse signal and noise conditions. In this work, STEAD provides diverse signal-degradation combinations for learning general waveform restoration knowledge.

Mine microseismic dataset. The mine microseismic dataset is used as a target-domain dataset for evaluating cross-domain adaptation. Compared with STEAD, this dataset contains weaker events, more complex industrial background interference, and substantial distribution shifts.

Industrial non-natural waveform dataset. The industrial non-natural waveform dataset is used as another target domain. It contains waveform patterns and degradation characteristics that differ from natural seismic observations and is used to further evaluate the transferability and robustness of the proposed method.

For all datasets, each input sample is represented as a three-component waveform. Unless otherwise specified, the waveform length is fixed to 6000 samples, corresponding to 60 s at 100 Hz. The event-preceding background segment contains 400 samples and is used as observable evidence of the current degradation environment.

B.2 Baseline Models

We compare PCD-Net with traditional signal-processing methods, representative deep learning-based restoration models, and a mechanism-related prototype-conditioned baseline.

Bandpass: a conventional bandpass filtering method. It suppresses frequencies outside a predefined passband but cannot adapt to non-stationary degradation patterns.

Wavelet: a wavelet-thresholding denoising method that performs multi-scale decomposition and threshold-based coefficient shrinkage.

DeepDenoiser: a convolutional neural network-based seismic denoising model that learns a nonlinear mapping from corrupted waveforms to clean signals.

DPRNN: a dual-path recurrent neural network that improves temporal modeling through intra-segment and inter-segment sequence modeling.

Restormer1D: a one-dimensional adaptation of the Restormer architecture that uses attention-based feature aggregation for waveform restoration.

ProtoCond-TCN: a mechanism-related baseline that encodes the pre-event background segment, learns a prototype-conditioned representation from a learnable prototype memory, and injects the resulting environmental condition into a temporal convolutional denoising network.

PCD-Net: the proposed framework, which learns transferable degradation-environment representations from background observations, performs differentiable routing over a prototype memory, and conditions waveform restoration through prototype-conditioned cross-attention.

B.3 Evaluation Metrics

We report four evaluation metrics: SNR improvement, correlation coefficient, root mean squared error, and spectral-temporal mean absolute error.

The SNR improvement is computed as

$$\Delta\mathrm{SNR} = \mathrm{SNR}(\hat{y}, y) - \mathrm{SNR}(x, y). \tag{27}$$

The correlation coefficient is defined as

$$\mathrm{CC} = \frac{\sum_i (\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})}{\sqrt{\sum_i (\hat{y}_i - \bar{\hat{y}})^2}\sqrt{\sum_i (y_i - \bar{y})^2}}. \tag{28}$$

The RMSE is computed by

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_i (\hat{y}_i - y_i)^2}. \tag{29}$$

The spectral-temporal mean absolute error (ST-MAE) evaluates reconstruction error in the time-frequency domain:

$$\text{ST-MAE} = \frac{1}{|\Omega|}\sum_{(f,t)\in\Omega} \left|S_{\hat{y}}(f,t) - S_y(f,t)\right|, \tag{30}$$

where $S_{\hat{y}}$ and S_y denote the time-frequency representations of the recovered and clean signals, respectively.

B.4 Implementation Details

PCD-Net is first pretrained on STEAD and then adapted to target-domain data using selective cross-domain adaptation. Unless otherwise specified, all methods are trained and evaluated under the same data split, input length, optimizer, and early-stopping criterion.

For transfer experiments, the STEAD-pretrained checkpoint is used as the initialization. We evaluate four adaptation strategies:

(1) **Frozen**, where all parameters remain unchanged;

(2) **Decoder Adjustment**, where only the lightweight decoder reconstruction adjustment module is updated;

(3) **Selective Adaptation**, where environment-related modules and the lightweight decoder reconstruction adjustment module are updated while preserving the restoration backbone;

(4) **Full Fine-tuning**, where all model parameters are optimized.

The optimizer is AdamW with cosine learning-rate decay. The batch size is set to 2 unless limited by GPU memory. All experiments are conducted on an NVIDIA RTX 4060 Laptop GPU.

B.5 Target-Domain Adaptation

Algorithm 1 summarizes the selective target-domain adaptation procedure. The model is initialized from the STEAD-pretrained source checkpoint θ_s. According to the selected adaptation strategy, only the corresponding modules are updated while the remaining parameters are frozen.

In selective adaptation, environment-related modules, including the context encoder, prototype memory, and routing mechanism, are updated together with the lightweight decoder reconstruction adjustment module. The waveform encoder and the main decoder backbone remain fixed to preserve transferable restoration knowledge.

Algorithm 1 Selective target-domain adaptation with strategy-dependent parameter updates. The algorithm initializes from a source-domain checkpoint and updates only the modules enabled by the selected adaptation strategy.

Input: source checkpoint θ_s, target set $\mathcal{D}_t$, adaptation strategy m, epochs E
Output: adapted parameters θ_t
Initialize $\theta_t \leftarrow \theta_s$
Enable trainable modules according to adaptation strategy m
for epoch $= 1$ to E **do**
  for mini-batch (x, b, y) in $\mathcal{D}_t$ **do**
    $\hat{y} \leftarrow F_{\theta_t}(x, b)$
    Compute the target-domain restoration loss
    Update only trainable parameters in θ_t
  end for
  Evaluate on the validation split for early stopping
end for
return θ_t

Table 4: Comparison of different cross-domain adaptation strategies on the mine microseismic dataset.

Strategy	ΔSNR (dB)↑	CC↑	RMSE↓	ST-MAE↓
Frozen	5.6922	0.6471	0.2712	0.2055
Decoder Adjustment	8.9254	0.7775	0.1906	0.1450
Selective Adaptation	9.4749	0.7874	0.1694	0.1283
Full Fine-tuning	**10.8595**	**0.8257**	**0.1470**	**0.1113**

C Additional Experimental Results

C.1 Additional Cross-Domain Results

We provide additional cross-domain adaptation results on the mine microseismic dataset, which serves as another target domain for evaluating the transferability of PCD-Net under different sensing conditions.

Table 4 compares different adaptation strategies. The frozen source-domain model suffers from performance degradation when directly transferred to the target domain, demonstrating the existence of degradation distribution shifts between source and target environments. Decoder reconstruction adjustment alone improves reconstruction quality by adjusting target-specific restoration characteristics. However, adapting environment representations together with lightweight decoder reconstruction adjustment achieves further improvements, demonstrating that modeling target-domain degradation conditions is important for effective cross-domain restoration.

Although full fine-tuning achieves the highest restoration performance by optimizing all parameters, selective adaptation provides competitive performance while preserving the transferable restoration backbone and updating only domain-relevant components.

Figure 7 further visualizes the local temporal reconstruction error. Selective adaptation achieves lower temporal errors than frozen transfer and decoder-only calibration, indicating that adapting environment representations improves not only noise suppression but also temporal structure preservation.

C.2 Routing Hyperparameter Sensitivity

We analyze the influence of two key routing hyperparameters: the prototype memory size K and the routing temperature τ.

The prototype number K determines the capacity of the degradation environment representation space, while the temperature τ controls the sharpness of the differentiable routing distribution. Dif-

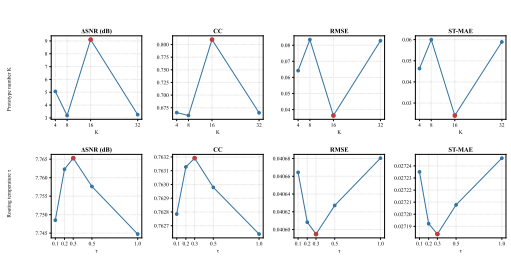


Figure 6: Sensitivity analysis of the prototype memory size K and routing temperature τ.

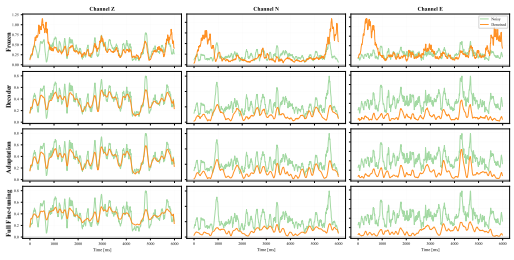


Figure 7: Local temporal-error comparison of different adaptation strategies on the mine microseismic dataset.

Table 5: Computational complexity comparison.

Method	FLOPs (G)	Lat. (ms)	Thr. (samp./s)	RTF (10^{-4})
Bandpass	–	6.140	162.879	1.023
Wavelet	–	2.274	439.844	0.379
DeepDenoiser	1.713	5.410	184.854	0.902
DPRNN	2.809	6.387	156.572	1.064
Restormer1D	8.048	27.732	36.059	4.622
ProtoCond-TCN	5.808	10.669	93.728	1.778
PCD-Net	3.097	8.956	111.661	1.493

ferent configurations are evaluated on the validation set to analyze the robustness of the proposed environment routing mechanism.

As shown in Fig. 6, both K and τ influence restoration performance. A small prototype memory limits the diversity of represented degradation environments, whereas an excessively large memory may introduce redundant environment patterns. The routing temperature controls the balance between smooth environment assignment and selective prototype activation. The selected configuration provides an effective trade-off between representation capacity and routing stability.

C.3 Complexity and Inference Efficiency

Table 5 reports the computational complexity and inference efficiency of different restoration methods. Although PCD-Net introduces an additional environment representation pathway and prototype-conditioned interaction module, it maintains practical inference efficiency for waveform restoration.