Predictive Spectral Calibration for Source-Free Test-Time Regression

Tuan-Kiet Nguyen Viet
Hanoi University of Science and Technology
kiet.nvt220032@hust.edu.vn

Thanh-Trung Huynh
VinUniversity
trung.ht@vinuni.edu.vn

Huy-Hieu Pham
VinUni-Illinois Smart Health Center, VinUniversity
hieu.ph@vinuni.edu.vn

Abstract

Test-time adaptation (TTA) for image regression has received far less attention than its classification counterpart. Methods designed for classification often depend on classification-specific objectives and decision boundaries, making them difficult to transfer directly to continuous regression targets. Recent progress revisits regression TTA through subspace alignment, showing that simple source-guided alignment can be both practical and effective. Building on this line of work, we propose Predictive Spectral Calibration (PSC), a source-free framework that extends subspace alignment to block spectral matching. Instead of relying on a fixed support subspace alone, PSC jointly aligns target features within the source predictive support and calibrates residual spectral slack in the orthogonal complement. PSC remains simple to implement, model-agnostic, and compatible with off-the-shelf pretrained regressors. Experiments on multiple image regression benchmarks show consistent improvements over strong baselines, with particularly clear gains under severe distribution shifts.

1. Introduction

Modern vision systems are increasingly deployed in settings where data shifts are the rule rather than the exception: lighting changes across cameras, weather changes across time, and sensor characteristics vary across devices [2, 8, 9]. In these scenarios, test-time adaptation (TTA) [15, 16] is appealing because it updates a pretrained model using only unlabeled target samples observed at inference. While this paradigm has matured rapidly for classification [10, 17], progress in image regression remains comparatively limited [1], despite the central role of regression tasks such as age estimation, depth prediction, and pose estimation.

A key reason is structural mismatch. Many successful classification TTA objectives rely on confidence sharpening [13], entropy minimization [16], or pseudo-label dynamics [7] that are naturally defined over discrete classes, but become less stable for continuous targets. Recent subspace-alignment methods, especially Significant-Subspace Alignment (SSA) [1], show that constraining adaptation within source-informed feature subspaces can be robust and practical. Building on this line, we introduce Predictive Spectral Calibration (PSC), which generalizes subspace alignment by jointly modeling predictive-support statistics and residual spectral behavior under domain shift.

In summary, our contributions are threefold: First, we present PSC as a practical source-free framework for test-time adaptation in image regression, designed to remain robust under distribution shift while staying compatible with pretrained regressors. Second, we provide a clear theoretical account of why the framework is reliable, including identifiability of key predictive statistics and guarantees on prediction drift under shift. Third, we conduct broad empirical evaluation across multiple benchmarks and corruption settings, where PSC consistently improves over strong baselines, with especially clear gains in severe-shift scenarios.

2. Preliminaries

2.1. Problem Setting

Image Regression. Let $\mathcal{X}$ denote the image space and $\mathcal{Y} = \mathbb{R}$ the continuous target space. Given a labeled source dataset $\mathcal{D}^s = \{(\mathbf{x}_i^s, y_i^s)\}_{i=1}^{N_s} \subset \mathcal{X} \times \mathcal{Y}$, we train a regression model $f_\theta : \mathcal{X} \to \mathcal{Y}$, parameterized by θ, by minimizing the supervised loss

$$\mathcal{L}(\theta) = \frac{1}{N_s} \sum_{i=1}^{N_s} \ell(f_\theta(\mathbf{x}_i), y_i)\,, \tag{1}$$

where ℓ is a point-wise regression loss, e.g., MAE or MSE. Typical image regression tasks include age estimation from facial images, depth estimation, and head-pose regression.

Test-Time Adaptation. Test-time adaptation (TTA) studies how a source-trained model can remain reliable after deployment under target-domain shift, using only unlabeled target samples at inference time. Formally, given an unlabeled target set $\mathcal{D}^t = \{\mathbf{x}_i^t\}_{i=1}^{N_t}$, TTA updates the model parameters online during testing. The objective is to improve target-domain predictive performance while strictly avoiding access to source data $\mathcal{D}^s$.

Source-Free Domain Adaptation. Domain adaptation (DA) seeks to handle distribution shift between a labeled source domain and a target domain. Since many DA methods require source data during adaptation, they are not directly applicable to TTA, where only a pretrained model and unlabeled target samples are available. Source-free domain adaptation (SFDA) addresses this limitation. Widely used methods include Domain-Adversarial Neural Networks (DANN) [6], Representation Subspace Distance (RSD) [4], and Feature Restoration (FR) [5].

Self-Supervised Test-Time Training. Self-supervised test-time training (TTT) improves robustness to distribution shift by updating a model on unlabeled target samples during inference, using self-supervised consistency [14] or reconstruction-style signals [11] available at test time. In this work, we re-implement two well-known variants for the regression setting: Test-Time Training (TTT) [15] and Activation Matching (AM) [12].

Feature Alignment. Feature alignment (FA) methods address distribution shifts by matching the statistics of target features to those computed from the source domain. These approaches typically store source feature statistics during training and align the target feature distribution during testing. Representative examples include Batch-Normalization Adaptation (BNA) [3] and Significant-Subspace Alignment (SSA) [1]. Because they operate directly in the feature space without requiring additional labels or auxiliary tasks, feature alignment methods are particularly suitable for regression-based TTA.

2.2. Significant-Subspace Alignment

Significant-Subspace Alignment (SSA) [1] adapts a regression model by aligning target features to source feature statistics inside a source-informed low-dimensional subspace. Let the regression model be decomposed as

$$f_\theta(\mathbf{x}) = (h_\psi \circ g_\phi)(\mathbf{x}), \quad \theta = (\psi, \phi) \tag{2}$$

where $g_\phi : \mathcal{X} \to \mathbb{R}^D$ is the feature extractor and $h_\psi(\mathbf{z}) = \mathbf{w}^\top \mathbf{z} + b$ is a linear regressor.

Subspace Construction. Given source features $\{\mathbf{z}_i^s\}_{i=1}^{N_s}$, where $\mathbf{z}_i^s = g_\phi(\mathbf{x}_i^s)$, SSA computes source statistics $(\boldsymbol{\mu}^s, \boldsymbol{\Sigma}^s)$. Let $\{(\lambda_k^s, \mathbf{v}_k^s)\}_{k=1}^D$ be the eigenpairs of $\boldsymbol{\Sigma}^s$, sorted as $\lambda_1^s \geq \cdots \geq \lambda_D^s$. SSA keeps the top-$K$ components and defines

$$\begin{aligned} \mathbf{V}^s &= [\mathbf{v}_1^s, \ldots, \mathbf{v}_K^s]^\top \in \mathbb{R}^{K \times D}, \\ \boldsymbol{\Lambda}^s &= \mathrm{diag}(\lambda_1^s, \ldots, \lambda_K^s) \in \mathbb{R}^{K \times K}. \end{aligned} \tag{3}$$

The projected source feature is $\mathbf{u}_i^s = \mathbf{V}^s(\mathbf{z}_i^s - \boldsymbol{\mu}^s)$, and SSA assumes $\mathbf{u}_i^s \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}^s)$.

Test-Time Objective. For a target mini-batch $\mathcal{B}^t$, for each $\mathbf{z}_i^t \in \mathcal{B}^t$, define the projected feature $\mathbf{u}_i^t = \mathbf{V}^s(\mathbf{z}_i^t - \boldsymbol{\mu}^s)$. Let $\tilde{\boldsymbol{\mu}}^t \in \mathbb{R}^K$ and $(\tilde{\boldsymbol{\sigma}}^2)^t \in \mathbb{R}^K$ denote the mean and variance of $\mathbf{u}_i^t$. SSA minimizes symmetric KL (SKL):

$$\begin{aligned} \mathcal{L}_{\mathrm{SSA}} &= \sum_{k=1}^K \alpha_k \mathrm{SKL}\big(\mathcal{N}(0, \lambda_k^s) \,\|\, \mathcal{N}(\tilde{\mu}_k^t, (\tilde{\sigma}_k^t)^2)\big) \\ &= \frac{1}{2} \sum_{k=1}^K \alpha_k \Big[\frac{(\tilde{\mu}_k^t)^2 + \lambda_k^s}{(\tilde{\sigma}_k^t)^2} + \frac{(\tilde{\mu}_k^t)^2 + (\tilde{\sigma}_k^t)^2}{\lambda_k^s} - 2 \Big] \end{aligned} \tag{4}$$

where $\alpha_k = 1 + |\mathbf{w}^\top \mathbf{v}_k^s|$, so SSA assigns larger importance to source directions that have stronger influence on the prediction. SSA is effective because it restricts adaptation to the predictive support inferred from source features. However, it only models the source distribution *inside* the selected support subspace and leaves the complementary residual space unmodeled. This can be restrictive under target shift, where adaptation quality may also depend on controlling how much target features leak outside the source-informed predictive support.

3. Methodology

We propose Predictive Spectral Calibration (PSC) as a source-free TTA method that extends SSA from subspace-only alignment to full *block spectral matching*.

3.1. Block Spectral Source Model

Let $\mathbf{z}^s = g_\phi(\mathbf{x}^s) \in \mathbb{R}^D$ be source features with mean $\boldsymbol{\mu}^s$ and covariance $\boldsymbol{\Sigma}^s$. We reuse the source subspace basis $\mathbf{V}^s$ defined in Eq. (3); by construction, $\mathbf{V}^s \mathbf{V}^{s\top} = \mathbf{I}_K$. Define

$$\mathbf{P}_\parallel^s = \mathbf{V}^{s\top} \mathbf{V}^s \in \mathbb{R}^D, \quad \mathbf{P}_\perp^s = \mathbf{I}_D - \mathbf{P}_\parallel^s \in \mathbb{R}^D. \tag{5}$$

PSC approximates source covariance as

$$\boldsymbol{\Sigma}^s \approx \mathbf{V}^{s\top} \boldsymbol{\Lambda}^s \mathbf{V}^s + \tau \mathbf{P}_\perp^s, \tag{6}$$

where the residual spectral floor is

$$\tau = \frac{1}{D-K} \sum_{k=K+1}^D \lambda_k^s. \tag{7}$$

Eq. (6) explicitly separates an anisotropic predictive block (support) and an isotropic residual block (complement).

Sample Decomposition. For any sample $\mathbf{x}_i$, define

$$\overline{\mathbf{z}}_i = g_\phi(\mathbf{x}_i) - \boldsymbol{\mu}^s. \tag{8}$$

PSC decomposes $\overline{\mathbf{z}}_i$ as

$$\mathbf{u}_i = \mathbf{V}^s \overline{\mathbf{z}}_i, \quad \mathbf{r}_i = \mathbf{P}_\perp^s \overline{\mathbf{z}}_i = \overline{\mathbf{z}}_i - \mathbf{V}^{s\top} \mathbf{u}_i. \tag{9}$$

For source samples, this yields a block Gaussian model:

$$\mathbf{u}_i^s \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}^s), \quad \mathbf{r}_i^s \sim \mathcal{N}(\mathbf{0}, \tau \mathbf{P}_\perp^s). \tag{10}$$

Here $\mathbf{u}_i$ captures the component inside predictive support, while $\mathbf{r}_i$ represents *spectral slack*, i.e., residual feature mass in the orthogonal complement outside that support.

3.2. Support-Space Spectral Matching

Inside the support, PSC uses a fixed rank-K^2 probe bank

$$\mathcal{Q} = \{e_i\}_{i=1}^{K} \cup \{q_{ij}^{+}, q_{ij}^{-}\}_{1\le i<j\le K}, \quad |\mathcal{Q}| = K^2 \tag{11}$$

where

$$q_{ij}^{+} = \frac{e_i + e_j}{\sqrt{2}}, \qquad q_{ij}^{-} = \frac{e_i - e_j}{\sqrt{2}}, \tag{12}$$

and $\{e_i\}_{i=1}^{K}$ denotes the canonical basis of $\mathbb{R}^K$.

For a target mini-batch $\mathcal{B}^t = \{\mathbf{x}_i^t\}_{i=1}^{B}$, we compute projected support features as in Eq. (9):

$$\mathbf{u}_i^t = \mathbf{V}^s\big(g_\phi(\mathbf{x}_i^t) - \boldsymbol{\mu}^s\big). \tag{13}$$

For each probe $\mathbf{q} \in \mathcal{Q}$, let $p_i(\mathbf{q}) = \mathbf{q}^\top \mathbf{u}_i^t$, and denote by $\hat{\mu}_{\mathbf{q}}$ and $\hat{\sigma}_{\mathbf{q}}^2$ the empirical mean and variance of $\{p_i(\mathbf{q})\}_{i=1}^{B}$. The source variance along $\mathbf{q}$ is

$$(\sigma_{\mathbf{q}}^s)^2 = \mathbf{q}^\top \boldsymbol{\Lambda}^s \mathbf{q}. \tag{14}$$

Proposition 1 (Identifiability from the K^2 probe bank)**.** *Assume $\mathbf{u}^s \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}^s)$ and $\mathbf{u}^t \sim \mathcal{N}(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t)$ in the support subspace. If, for every $\mathbf{q} \in \mathcal{Q}$, $\mathbf{q}^\top \mathbf{u}^t \overset{d}{=} \mathbf{q}^\top \mathbf{u}^s$, then $\boldsymbol{\mu}^t = \mathbf{0}$ and $\boldsymbol{\Sigma}^t = \boldsymbol{\Lambda}^s$. Equivalently, $\mathbf{u}^t \overset{d}{=} \mathbf{u}^s$.*

Proof. The claim follows directly from Proposition 3 in the Appendix, which proves that the K^2 probe bank uniquely identifies the full first- and second-order structure in the support subspace. □

To make alignment prediction-aware, we project the regression head into support coordinates:

$$\mathbf{a} = \mathbf{V}^s \mathbf{w} \in \mathbb{R}^K, \qquad \beta_{\mathbf{q}} = \big(|\mathbf{a}^\top \mathbf{q}| + c\big)^\gamma, \tag{15}$$

with hyperparameters $c > 0$ and $\gamma > 0$. The support loss is

$$\begin{aligned}\mathcal{L}_{\text{sup}} &= \frac{1}{K^2}\sum_{\mathbf{q}\in\mathcal{Q}} \beta_{\mathbf{q}} \mathrm{SKL}\big(\mathcal{N}(0, (\sigma_{\mathbf{q}}^s)^2) \,\|\, \mathcal{N}(\hat{\mu}_{\mathbf{q}}, \hat{\sigma}_{\mathbf{q}}^2)\big) \\ &= \frac{1}{2K^2}\sum_{\mathbf{q}\in\mathcal{Q}} \beta_{\mathbf{q}} \left[\frac{\hat{\mu}_{\mathbf{q}}^2 + \hat{\sigma}_{\mathbf{q}}^2}{(\sigma_{\mathbf{q}}^s)^2} + \frac{\hat{\mu}_{\mathbf{q}}^2 + (\sigma_{\mathbf{q}}^s)^2}{\hat{\sigma}_{\mathbf{q}}^2} - 2\right].\end{aligned} \tag{16}$$

3.3. Residual-Space Spectral Matching

For the same mini-batch, define residual statistics

$$\hat{\boldsymbol{\mu}}_\perp = \frac{1}{B}\sum_{i=1}^{B} \mathbf{r}_i, \quad \hat{\nu}_\perp = \frac{1}{B(D-K)}\sum_{i=1}^{B} \|\mathbf{r}_i - \hat{\boldsymbol{\mu}}_\perp\|_2^2. \tag{17}$$

Under Eq. (10), we work on the residual subspace

$$\mathcal{S}_\perp = \mathrm{Im}(\mathbf{P}_\perp^s), \qquad \dim(\mathcal{S}_\perp) = D - K. \tag{18}$$

Choose an orthonormal basis $\mathbf{U}_\perp \in \mathbb{R}^{D\times(D-K)}$ of $\mathcal{S}_\perp$:

$$\mathbf{U}_\perp^\top \mathbf{U}_\perp = \mathbf{I}_{D-K}, \qquad \mathbf{U}_\perp \mathbf{U}_\perp^\top = \mathbf{P}_\perp^s. \tag{19}$$

For any residual vector $\mathbf{r} \in \mathcal{S}_\perp$, the coordinate transform is

$$\tilde{\mathbf{r}} = \mathbf{U}_\perp^\top \mathbf{r}, \qquad \mathbf{r} = \mathbf{U}_\perp \tilde{\mathbf{r}}. \tag{20}$$

The ambient-space residual model is

$$\mathbf{r}^s \sim \mathcal{N}(\mathbf{0}, \tau \mathbf{P}_\perp^s), \qquad \mathbf{r}^t \approx \mathcal{N}(\hat{\boldsymbol{\mu}}_\perp, \hat{\nu}_\perp \mathbf{P}_\perp^s). \tag{21}$$

Table 1. Test scores for TTA from source SVHN to target MNIST. The best scores are highlighted in blue.

Type	Method	Venue	R^2(↑)	RMSE (↓)	MAE (↓)
	Source		0.239	2.527	1.909
DA	DANN [6]	JMLR 2016	0.035	2.653	1.856
	RSD [4]	ICML 2021	0.049	2.777	2.401
	FR [5]	ICLR 2022	0.356	2.323	1.554
SS	TTT [15]	ICML 2020	0.286	2.447	1.950
	AM [12]	CVPR 2023	0.276	2.465	1.637
FA	BNA [3]	WACV 2021	0.341	2.349	1.581
	SSA [1]	ICLR 2025	0.452	2.144	1.342
	PSC ($\lambda = 0$)		0.473	2.102	1.310
	PSC ($\lambda = 1$)		0.457	2.134	1.312
	Oracle		0.866	1.060	0.558

Equivalently, in complement coordinates,

$$\tilde{\mathbf{r}}^s \sim \mathcal{N}(\mathbf{0}, \tau \mathbf{I}_{D-K}), \qquad \tilde{\mathbf{r}}^t \approx \mathcal{N}(\hat{\tilde{\boldsymbol{\mu}}}_\perp, \hat{\nu}_\perp \mathbf{I}_{D-K}), \tag{22}$$

where $\hat{\tilde{\boldsymbol{\mu}}}_\perp = \mathbf{U}_\perp^\top \hat{\boldsymbol{\mu}}_\perp$. The residual loss is defined as:

$$\begin{aligned}\mathcal{L}_{\text{res}} &= \mathrm{SKL}\Big(\mathcal{N}(\mathbf{0}, \tau \mathbf{I}_{D-K}), \mathcal{N}(\hat{\tilde{\boldsymbol{\mu}}}_\perp, \hat{\nu}_\perp \mathbf{I}_{D-K})\Big) \\ &= \frac{1}{2}\left[\frac{\|\hat{\boldsymbol{\mu}}_\perp\|_2^2}{D-K}\left(\frac{1}{\tau} + \frac{1}{\hat{\nu}_\perp}\right) + \frac{\tau}{\hat{\nu}_\perp} + \frac{\hat{\nu}_\perp}{\tau} - 2\right].\end{aligned} \tag{23}$$

because $\|\hat{\tilde{\boldsymbol{\mu}}}_\perp\| = \|\hat{\boldsymbol{\mu}}_\perp\|$. This term explicitly controls feature leakage outside predictive support.

3.4. Test-Time Objective

PSC combines both blocks into one objective:

$$\mathcal{L}_{\text{PSC}} = \mathcal{L}_{\text{sup}} + \lambda \mathcal{L}_{\text{res}}, \tag{24}$$

where $\lambda > 0$ balances support and residual alignment.

Relation to SSA. PSC is a strict generalization of SSA in three senses. First, if probes are canonical basis vectors, support matching reduces to axis-wise alignment used by SSA. Second, if $\lambda = 0$, PSC collapses to support-only matching, again recovering the SSA-style objective. Third, for general probe banks and $\lambda > 0$, PSC additionally constrains residual-space statistics, which SSA does not model. Hence, SSA is a special case of PSC, while PSC can exploit richer directional statistics and out-of-subspace control under domain shift.

At test time, we keep source statistics $\{\boldsymbol{\mu}^s, \mathbf{V}^s, \boldsymbol{\Lambda}^s, \tau\}$ fixed and update the feature-extractor parameters ϕ while freezing the head parameters ψ:

$$\phi^* = \arg\min_{\phi} \mathcal{L}_{\text{PSC}}. \tag{25}$$

Following standard regression TTA practice, this is implemented by updating only affine parameters in normalization layers of g_ϕ.

Table 2. Test R^2 scores on UTKFace under 13 corruption types (higher is better). The best scores are highlighted in blue.

Method	Gaussian noise	Shot noise	Impulse noise	Defocus blur	Motion blur	Zoom blur	Snow	Fog	Brightness	Contrast	Elastic transform	Pixelate	JPEG compression	Mean
Source	−2.557	−0.715	−3.897	−0.023	0.594	0.651	−0.185	0.229	0.263	0.615	0.750	−2.101	−0.190	–
DANN [6]	0.512	0.784	0.281	0.532	0.772	0.786	0.224	0.528	0.107	0.907	0.827	0.466	0.631	0.566
RSD [4]	0.302	0.577	0.129	0.191	0.461	0.384	0.173	0.347	0.245	0.641	0.500	0.270	0.280	0.346
FR [5]	0.499	0.860	0.159	0.455	0.752	0.710	0.227	0.551	0.435	0.943	0.790	0.282	0.585	0.558
TTT [15]	0.564	0.779	0.243	0.387	0.683	0.730	0.148	0.501	0.382	0.858	0.722	0.415	0.531	0.534
AM [12]	0.524	0.764	0.282	0.355	0.470	0.619	0.269	0.535	0.064	0.554	0.710	0.445	0.484	0.467
BNA [3]	0.493	0.858	0.151	0.461	0.754	0.725	0.246	0.555	0.472	0.942	0.800	0.279	0.586	0.563
SSA [1]	0.589	0.860	0.241	0.491	0.753	0.754	0.260	0.552	0.423	0.934	0.803	0.334	0.590	0.583
PSC ($\lambda = 0$)	0.621	0.866	0.265	0.530	0.774	0.767	0.263	0.593	0.474	0.939	0.814	0.332	0.618	0.604
PSC ($\lambda = 1$)	0.605	0.863	0.288	0.516	0.763	0.760	0.343	0.581	0.554	0.940	0.798	0.416	0.624	0.619
Oracle	0.688	0.899	0.369	0.611	0.831	0.832	0.513	0.648	0.614	0.959	0.839	0.576	0.679	0.697

Proposition 2 (PSC controls predictive mean drift)**.** *Let $\mathbf{x} \in \mathcal{X}$ and $\mathbf{z} = g_\phi(\mathbf{x}) \in \mathbb{R}^D$ denote its feature representation. Define the centered feature $\bar{\mathbf{z}} = \mathbf{z} - \boldsymbol{\mu}^s$, the support projection $\mathbf{u} = \mathbf{V}^s\bar{\mathbf{z}}$, and the residual projection $\mathbf{r} = \mathbf{P}^s_\perp\bar{\mathbf{z}}$. For the linear head $h_\psi(\mathbf{z}) = \mathbf{w}^\top\mathbf{z} + b$, decompose*

$$\mathbf{w} = \mathbf{V}^{s\top}\mathbf{a} + \mathbf{w}_\perp, \qquad \mathbf{a} = \mathbf{V}^s\mathbf{w}, \qquad \mathbf{w}_\perp = \mathbf{P}^s_\perp\mathbf{w}.$$

Assume block Gaussian models

$$\mathbf{u}^s \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}^s), \quad \mathbf{r}^s \sim \mathcal{N}(\mathbf{0}, \tau\mathbf{P}^s_\perp),$$
$$\mathbf{u}^t \sim \mathcal{N}(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t), \quad \mathbf{r}^t \sim \mathcal{N}(\boldsymbol{\mu}^t_\perp, \nu^t\mathbf{P}^s_\perp).$$

Define

$$\mathcal{D}_{\mathrm{PSC}} = \mathrm{SKL}\big(\mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}^s), \mathcal{N}(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t)\big) + \mathrm{SKL}\big(\mathcal{N}(\mathbf{0}, \tau\mathbf{P}^s_\perp), \mathcal{N}(\boldsymbol{\mu}^t_\perp, \nu^t\mathbf{P}^s_\perp)\big).$$

Then for $\hat{y} = h_\psi(\mathbf{z})$, the predictive mean drift satisfies

$$\big|\mathbb{E}^t[\hat{y}] - \mathbb{E}^s[\hat{y}]\big| \le \sqrt{2\mathcal{D}_{\mathrm{PSC}}}\sqrt{\mathbf{a}^\top\boldsymbol{\Lambda}^s\mathbf{a} + \tau\|\mathbf{w}_\perp\|_2^2}.$$

Proof. Write the mean drift as $\Delta_\mu = \mathbb{E}^t[\hat{y}] - \mathbb{E}^s[\hat{y}] = \mathbf{a}^\top\boldsymbol{\mu}^t + \mathbf{w}_\perp^\top\boldsymbol{\mu}^t_\perp$ under the support/residual decomposition. Applying Cauchy–Schwarz in the source metrics induced by $(\boldsymbol{\Lambda}^s, \tau\mathbf{P}^s_\perp)$, and using the standard Gaussian SKL lower bounds, yields $|\Delta_\mu| \le \sqrt{2\mathcal{D}_{\mathrm{PSC}}}\sqrt{\mathbf{a}^\top\boldsymbol{\Lambda}^s\mathbf{a} + \tau\|\mathbf{w}_\perp\|_2^2}$. The detailed derivation is provided in the Appendix (see Proposition 4 and the subsequent detailed proof). □

Meaning for PSC. This proposition links PSC's optimization target to predictive robustness: reducing $\mathcal{D}_{\mathrm{PSC}}$ directly tightens an upper bound on target-domain mean prediction drift. The bound separates *distribution mismatch* (first factor) from *model sensitivity* along support/residual directions (second factor), explaining why jointly aligning support statistics and residual slack is beneficial under shift.

4. Experiments

Experimental design. We evaluate PSC under two shift regimes. First, we use cross-domain transfer from SVHN to MNIST to test adaptation under a large appearance gap. Second, we use UTKFace with 13 corruption types to test robustness to structured low-level perturbations. The tables include two references: *Source* denotes the source-trained model without TTA, while *Oracle* is a target-supervised upper bound using target labels.

Effect of λ across settings. The preferred λ depends on the shift type. For SVHN→MNIST, $\lambda = 0$ is typically better because the cross-domain gap is strong and heterogeneous; imposing residual-space matching can over-constrain adaptation and reduce flexibility needed to absorb target-specific changes. For UTKFace corruptions, $\lambda = 1$ is typically better because the shift is mainly nuisance corruption around the same underlying semantics, so residual-space regularization helps suppress out-of-support noise leakage. In short, stronger semantic transfer favors support-focused alignment, while corruption robustness benefits from explicit residual control.

5. Conclusion

We introduced Predictive Spectral Calibration (PSC), a source-free test-time adaptation framework for image regression that combines support-space alignment with residual-space calibration. Across cross-domain transfer and corruption settings, PSC shows consistent improvements over strong baselines while remaining simple and practical to deploy. We view this study as a *preliminary idea* toward more principled regression TTA, and we hope it motivates further work on stronger objectives, broader backbones, and larger-scale benchmarks.

References

[1] Kazuki Adachi, Shin'ya Yamaguchi, Atsutoshi Kumagai, and Tomoki Hamagami. Test-time adaptation for regression by subspace alignment. In *The Thirteenth International Conference on Learning Representations*, 2025. 1, 2, 3, 4

[2] Eunsu Baek, Keondo Park, Jiyoon Kim, and Hyung-Sin Kim. Unexplored faces of robustness and out-of-distribution: Covariate shifts in environment and sensor domains. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22294–22303, 2024. 1

[3] Philipp Benz, Chaoning Zhang, Adil Karjauv, and In So Kweon. Revisiting batch normalization for improving corruption robustness. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 494–503, 2021. 2, 3, 4

[4] Xinyang Chen, Sinan Wang, Jianmin Wang, and Mingsheng Long. Representation subspace distance for domain adaptation regression. In *International Conference on Machine Learning*, pages 1749–1759, 2021. 2, 3, 4

[5] Cian Eastwood, Ian Mason, Chris Williams, and Bernhard Schölkopf. Source-free adaptation to measurement shift via bottom-up feature restoration. In *International Conference on Learning Representations*, 2022. 2, 3, 4

[6] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario March, and Victor Lempitsky. Domain-adversarial training of neural networks. *Journal of machine learning research*, 17(59):1–35, 2016. 2, 3, 4

[7] Sachin Goyal, Mingjie Sun, Aditi Raghunathan, and J Zico Kolter. Test time adaptation via conjugate pseudo-labels. *Advances in Neural Information Processing Systems*, 35:6204–6218, 2022. 1

[8] Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. In *International Conference on Learning Representations*, 2019. 1

[9] Pang Wei Koh, Shiori Sagawa, Henrik Marklund, Sang Michael Xie, Marvin Zhang, Akshay Balsubramani, Weihua Hu, Michihiro Yasunaga, Richard Lanas Phillips, Irena Gao, et al. Wilds: A benchmark of in-the-wild distribution shifts. In *International conference on machine learning*, pages 5637–5664. PMLR, 2021. 1

[10] Jian Liang, Ran He, and Tieniu Tan. A comprehensive survey on test-time adaptation under distribution shifts. *International Journal of Computer Vision*, 133(1):31–64, 2025. 1

[11] Yuejiang Liu, Parth Kothari, Bastien van Delft, Baptiste Bellot-Gurlet, Taylor Mordan, and Alexandre Alahi. Ttt++: When does self-supervised test-time training fail or thrive? In *Advances in Neural Information Processing Systems*, pages 21808–21820. Curran Associates, Inc., 2021. 2

[12] M. Jehanzeb Mirza, Pol Jane Soneira, Wei Lin, Mateusz Kozinski, Horst Possegger, and Horst Bischof. Actmad: Activation matching to align distributions for test-time training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 2, 3, 4

[13] Chaithanya Kumar Mummadi, Robin Hutmacher, Kilian Rambach, Evgeny Levinkov, Thomas Brox, and Jan Hendrik Metzen. Test-time adaptation to distribution shift by confidence maximization and input transformation. *arXiv preprint arXiv:2106.14999*, 2021. 1

[14] Samarth Sinha, Peter Gehler, Francesco Locatello, and Bernt Schiele. Test: Test-time self-training under distribution shift. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 2759–2769, 2023. 2

[15] Yu Sun, Xiaolong Wang, Zhuang Liu, John Miller, Alexei Efros, and Moritz Hardt. Test-time training with self-supervision for generalization under distribution shifts. In *International conference on machine learning*, pages 9229–9248. PMLR, 2020. 1, 2, 3, 4

[16] 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. 1

[17] Zixin Wang, Yadan Luo, Liang Zheng, Zhuoxiao Chen, Sen Wang, and Zi Huang. In search of lost online test-time adaptation: A survey. *International Journal of Computer Vision*, 133(3):1106–1139, 2025. 1

Predictive Spectral Calibration for Source-Free Test-Time Regression

Supplementary Material

A. Theoretical Details

A.1. Detailed Proof of Propositon 1

Proposition 3 (Identifiability of subspace first- and second-order structure from the K^2 probe bank)**.** *Let $u \in \mathbb{R}^K$ be a random vector with mean $\mu \in \mathbb{R}^K$ and covariance matrix $\Sigma \in \mathbb{R}^{K\times K}$. Define the probe bank*

$$\mathcal{Q} = \{e_i\}_{i=1}^K \cup \{q_{ij}^+, q_{ij}^-\}_{1\le i<j\le K}, \qquad q_{ij}^{\pm} = \frac{e_i \pm e_j}{\sqrt{2}}, \tag{26}$$

where $\{e_i\}_{i=1}^K$ is the canonical basis of $\mathbb{R}^K$. Then the projected first- and second-order moments

$$\big\{\mathbb{E}[q^\top u],\ \mathrm{Var}(q^\top u)\big\}_{q\in\mathcal{Q}} \tag{27}$$

uniquely determine (μ, Σ). More explicitly, $\mu_i = \mathbb{E}[e_i^\top u]$, $\Sigma_{ii} = \mathrm{Var}(e_i^\top u)$, and for each $1 \le i < j \le K$,

$$\Sigma_{ij} = \frac{1}{2}\left(\mathrm{Var}\big((q_{ij}^+)^\top u\big) - \mathrm{Var}\big((q_{ij}^-)^\top u\big)\right). \tag{28}$$

Hence, the K^2 probes recover the complete first- and second-order structure of u in the K-dimensional subspace.

Proof. For any $q \in \mathbb{R}^K$, we have $\mathbb{E}[q^\top u] = q^\top \mu$ and $\mathrm{Var}(q^\top u) = q^\top \Sigma q$. Taking $q = e_i$ gives

$$\mu_i = \mathbb{E}[e_i^\top u], \qquad \Sigma_{ii} = \mathrm{Var}(e_i^\top u), \tag{29}$$

so all entries of μ and all diagonal entries of Σ are identified.
Now fix $1 \le i < j \le K$. Using $q_{ij}^{\pm} = (e_i \pm e_j)/\sqrt{2}$,

$$\mathrm{Var}\big((q_{ij}^+)^\top u\big) = \frac{1}{2}\left(\Sigma_{ii} + \Sigma_{jj} + 2\Sigma_{ij}\right), \qquad \mathrm{Var}\big((q_{ij}^-)^\top u\big) = \frac{1}{2}\left(\Sigma_{ii} + \Sigma_{jj} - 2\Sigma_{ij}\right). \tag{30}$$

Subtracting yields

$$\Sigma_{ij} = \frac{1}{2}\left(\mathrm{Var}\big((q_{ij}^+)^\top u\big) - \mathrm{Var}\big((q_{ij}^-)^\top u\big)\right). \tag{31}$$

Thus every off-diagonal entry is identified. Therefore all entries of (μ, Σ) are uniquely recovered from $\{\mathbb{E}[q^\top u], \mathrm{Var}(q^\top u)\}_{q\in\mathcal{Q}}$. □

A.2. Detailed Proof of Proposition 2

Proposition 4 (PSC controls predictive mean drift)**.** *Let $\mathbf{x} \in \mathcal{X}$ and $\mathbf{z} = g_\phi(\mathbf{x}) \in \mathbb{R}^D$ denote the feature representation. Define*

$$\bar{\mathbf{z}} := \mathbf{z} - \boldsymbol{\mu}^s, \qquad \mathbf{u} := \mathbf{V}^s\bar{\mathbf{z}}, \qquad \mathbf{r} := \mathbf{P}_\perp^s\bar{\mathbf{z}}.$$

For the linear regressor $h_\psi(\mathbf{z}) = \mathbf{w}^\top\mathbf{z} + b$, set

$$\mathbf{a} := \mathbf{V}^s\mathbf{w}, \qquad \mathbf{w}_\perp := \mathbf{P}_\perp^s\mathbf{w}, \qquad \mathbf{w} = \mathbf{V}^{s\top}\mathbf{a} + \mathbf{w}_\perp.$$

Assume the source and target block models

$$\mathbf{u}^s \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}^s), \quad \mathbf{r}^s \sim \mathcal{N}(\mathbf{0}, \tau\mathbf{P}_\perp^s), \quad \mathbf{u}^t \sim \mathcal{N}(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t), \quad \mathbf{r}^t \sim \mathcal{N}(\boldsymbol{\mu}_\perp^t, \nu^t\mathbf{P}_\perp^s),$$

with $\boldsymbol{\Lambda}^s \succ \mathbf{0}$, $\boldsymbol{\Sigma}^t \succ \mathbf{0}$, and $\tau, \nu^t > 0$. Define

$$\mathcal{D}_{\mathrm{PSC}} := \mathrm{SKL}\big(\mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}^s), \mathcal{N}(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t)\big) + \mathrm{SKL}\big(\mathcal{N}(\mathbf{0}, \tau\mathbf{P}_\perp^s), \mathcal{N}(\boldsymbol{\mu}_\perp^t, \nu^t\mathbf{P}_\perp^s)\big),$$

where $\mathrm{SKL}(P,Q) := D_{\mathrm{KL}}(P\|Q) + D_{\mathrm{KL}}(Q\|P)$.
Then the predictive mean drift $\Delta_\mu := \mathbb{E}^t[\hat{y}] - \mathbb{E}^s[\hat{y}]$ *with* $\hat{y} = h_\psi(\mathbf{z})$ *satisfies*

$$|\Delta_\mu| \le \sqrt{2\,\mathcal{D}_{\mathrm{PSC}}}\sqrt{\mathbf{a}^\top \boldsymbol{\Lambda}^s \mathbf{a} + \tau\|\mathbf{w}_\perp\|_2^2}.$$

Consequently, the looser bound

$$|\Delta_\mu| \le \sqrt{2\,\mathcal{D}_{\mathrm{PSC}}}\left(\sqrt{\mathbf{a}^\top \boldsymbol{\Lambda}^s \mathbf{a}} + \sqrt{\tau}\,\|\mathbf{w}_\perp\|_2\right)$$

also holds.

Proof. Using $\mathbf{z} = \boldsymbol{\mu}^s + \bar{\mathbf{z}}$ and $\mathbf{w} = \mathbf{V}^{s\top}\mathbf{a} + \mathbf{w}_\perp$,

$$\hat{y} = \mathbf{w}^\top \mathbf{z} + b = \mathbf{w}^\top \boldsymbol{\mu}^s + b + \mathbf{a}^\top \mathbf{u} + \mathbf{w}_\perp^\top \mathbf{r}.$$

Since $\mathbb{E}^s[\mathbf{u}] = \mathbf{0}$, $\mathbb{E}^s[\mathbf{r}] = \mathbf{0}$, $\mathbb{E}^t[\mathbf{u}] = \boldsymbol{\mu}^t$, and $\mathbb{E}^t[\mathbf{r}] = \boldsymbol{\mu}_\perp^t$,

$$\Delta_\mu = \mathbb{E}^t[\hat{y}] - \mathbb{E}^s[\hat{y}] = \mathbf{a}^\top \boldsymbol{\mu}^t + \mathbf{w}_\perp^\top \boldsymbol{\mu}_\perp^t.$$

Define

$$A := \sqrt{\mathbf{a}^\top \boldsymbol{\Lambda}^s \mathbf{a}}, \quad x := \sqrt{(\boldsymbol{\mu}^t)^\top (\boldsymbol{\Lambda}^s)^{-1} \boldsymbol{\mu}^t}, \quad B := \sqrt{\tau}\,\|\mathbf{w}_\perp\|_2, \quad y := \sqrt{\frac{\|\boldsymbol{\mu}_\perp^t\|_2^2}{\tau}}.$$

By Cauchy–Schwarz,

$$\left|\mathbf{a}^\top \boldsymbol{\mu}^t\right| = \left|((\boldsymbol{\Lambda}^s)^{1/2}\mathbf{a})^\top ((\boldsymbol{\Lambda}^s)^{-1/2}\boldsymbol{\mu}^t)\right| \le Ax,$$

and

$$\left|\mathbf{w}_\perp^\top \boldsymbol{\mu}_\perp^t\right| \le \|\mathbf{w}_\perp\|_2\,\|\boldsymbol{\mu}_\perp^t\|_2 = By.$$

Hence

$$|\Delta_\mu| \le Ax + By \le \sqrt{A^2 + B^2}\,\sqrt{x^2 + y^2}.$$

It remains to prove $x^2 + y^2 \le 2\mathcal{D}_{\mathrm{PSC}}$.
For the support block,

$$\mathrm{SKL}\big(\mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}^s), \mathcal{N}(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t)\big) = \frac{1}{2}\Big[(\boldsymbol{\mu}^t)^\top \big((\boldsymbol{\Lambda}^s)^{-1} + (\boldsymbol{\Sigma}^t)^{-1}\big)\boldsymbol{\mu}^t + \mathrm{tr}\big(\boldsymbol{\Lambda}^s(\boldsymbol{\Sigma}^t)^{-1} + \boldsymbol{\Sigma}^t(\boldsymbol{\Lambda}^s)^{-1} - 2\mathbf{I}_K\big)\Big].$$

Since $(\boldsymbol{\Sigma}^t)^{-1} \succ \mathbf{0}$ and $\mathrm{tr}(\mathbf{M} + \mathbf{M}^{-1} - 2\mathbf{I}_K) \ge 0$ for any $\mathbf{M} \succ \mathbf{0}$,

$$\mathrm{SKL}\big(\mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}^s), \mathcal{N}(\boldsymbol{\mu}^t, \boldsymbol{\Sigma}^t)\big) \ge \frac{1}{2}(\boldsymbol{\mu}^t)^\top (\boldsymbol{\Lambda}^s)^{-1}\boldsymbol{\mu}^t = \frac{1}{2}x^2.$$

For the residual block, let $\mathbf{U}_\perp \in \mathbb{R}^{D\times(D-K)}$ satisfy

$$\mathbf{U}_\perp^\top \mathbf{U}_\perp = \mathbf{I}_{D-K}, \qquad \mathbf{U}_\perp \mathbf{U}_\perp^\top = \mathbf{P}_\perp^s.$$

Define $\tilde{\mathbf{r}}^s := \mathbf{U}_\perp^\top \mathbf{r}^s$, $\tilde{\mathbf{r}}^t := \mathbf{U}_\perp^\top \mathbf{r}^t$, and $\tilde{\boldsymbol{\mu}}_\perp^t := \mathbf{U}_\perp^\top \boldsymbol{\mu}_\perp^t$. Then

$$\tilde{\mathbf{r}}^s \sim \mathcal{N}(\mathbf{0}, \tau\mathbf{I}_{D-K}), \qquad \tilde{\mathbf{r}}^t \sim \mathcal{N}(\tilde{\boldsymbol{\mu}}_\perp^t, \nu^t\mathbf{I}_{D-K}),$$

and

$$\mathrm{SKL}\big(\mathcal{N}(\mathbf{0}, \tau\mathbf{P}_\perp^s), \mathcal{N}(\boldsymbol{\mu}_\perp^t, \nu^t\mathbf{P}_\perp^s)\big) = \frac{1}{2}\Big[\left(\frac{1}{\tau} + \frac{1}{\nu^t}\right)\|\tilde{\boldsymbol{\mu}}_\perp^t\|_2^2 + (D-K)\left(\frac{\tau}{\nu^t} + \frac{\nu^t}{\tau} - 2\right)\Big].$$

Because $\nu^t > 0$ and $\frac{\tau}{\nu^t} + \frac{\nu^t}{\tau} - 2 = \frac{(\tau-\nu^t)^2}{\tau\nu^t} \ge 0$,

$$\mathrm{SKL}\big(\mathcal{N}(\mathbf{0}, \tau\mathbf{P}_\perp^s), \mathcal{N}(\boldsymbol{\mu}_\perp^t, \nu^t\mathbf{P}_\perp^s)\big) \ge \frac{1}{2}\,\frac{\|\tilde{\boldsymbol{\mu}}_\perp^t\|_2^2}{\tau} = \frac{1}{2}\,\frac{\|\boldsymbol{\mu}_\perp^t\|_2^2}{\tau} = \frac{1}{2}y^2.$$

Summing the two bounds yields

$$\mathcal{D}_{\mathrm{PSC}} \ge \frac{1}{2}(x^2 + y^2), \qquad x^2 + y^2 \le 2\mathcal{D}_{\mathrm{PSC}}.$$

Therefore,

$$|\Delta_\mu| \le \sqrt{2\,\mathcal{D}_{\mathrm{PSC}}}\sqrt{\mathbf{a}^\top \boldsymbol{\Lambda}^s \mathbf{a} + \tau\|\mathbf{w}_\perp\|_2^2}.$$

Finally, $\sqrt{A^2 + B^2} \le A + B$ gives the looser bound. □