ISC-LoRA: Intrinsic Spectral Continuation for Multitask Dense Prediction

Anonymous authors
Paper under double-blind review

Abstract

Multitask learning with low-rank adaptation (LoRA) efficiently adapts a shared visual model to complementary dense prediction tasks, including semantic segmentation, boundary detection, and geometry estimation. Reweighting the visual features based on their spatial distribution adaptive to the specific task and inputs can further improve these predictions. To do this, multitask LoRA methods often separate low-rank updates into shared and task-private branches. The low-dimensional features in each private branch retain the spatial grid of the current input and can reflect task-specific spatial variation. However, the spatial context does not directly indicate which frequencies should be emphasized or how strongly, limiting the adaptation capability to the specific input or task. Building on this structure, we introduce Intrinsic Spectral Continuation LoRA (ISC-LoRA), a subspace-conditioned training scheme that derives a bounded response over the full *spatial-frequency* grid from each private branch's current features and subspace. This response is recomputed for each input and private branch, adapting the emphasis of spatial frequencies to each task and image without predefined frequency bands. To make the response intrinsic to this subspace rather than to the particular LoRA rank coordinates used to represent it, our method combines frequency preferences along the principal directions of feature variation using geometry-aware regularization and covariance normalization. We further prove that ISC-LoRA can reweight the spatial frequencies of the private output within a controlled range, with the response determined by the current input and task-private subspace, independent of the chosen LoRA rank basis. Notably, its continuation schedule reduces the modulation strength to zero, making the spectral operator the identity while retaining the learned LoRA factors. The resulting adapters can therefore be evaluated and deployed using standard LoRA computation, without spectral operations. Experiments on PASCAL-Context and NYUD-v2 show that ISC-LoRA outperforms representative multitask adaptation methods. Beyond natural images, evaluations on ten two-dimensional Cell Tracking Challenge (CTC) datasets further demonstrate its advantages over representative adaptation baselines in cell segmentation, detection, and tracking. Our code and checkpoints are available at `https://anonymous.4open.science/r/ISC-LoRA`.

1 Introduction

Multitask dense prediction seeks complementary descriptions of the same image, from semantic regions and object boundaries to scene geometry. Pretrained visual models provide a common starting point for these tasks (Ranftl et al., 2021; Liu et al., 2021; Oquab et al., 2024; Siméoni et al., 2026). To adapt these models efficiently, low-rank adaptation (LoRA) (Hu et al., 2022) learns compact factorized updates while freezing the pretrained weights. Multitask LoRA methods differ in how they combine shared and task-specific updates (Agiza et al., 2024; Fontana et al., 2026; Gangwar et al., 2026). In the shared and task-private formulation considered here, the shared branch learns updates used by all tasks, while the private branches meet each task's adaptation needs. Each private branch contains a low-dimensional bottleneck that retains the spatial grid of its input feature map.

For the same image, task-specific supervision can lead private branches to learn different representations (Agiza et al., 2024). In frequency terms, spatial variation ranges from slowly varying structure at lower spatial frequencies to rapid changes such as boundaries and fine detail at higher

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.

spatial frequencies (Ly & Nguyen, 2024; Bi et al., 2024). Input context can also guide task-specific feature adaptation (Baek et al., 2025). The bottleneck outputs can thus serve as task- and input-dependent spatial context for spectral adaptation. However, this context does not directly indicate which frequencies should be emphasized or how strongly, which would otherwise help further enhance important features in the images.

Frequency-based visual adaptation offers ways to shape this response. FreqFit (Ly & Nguyen, 2024) learns a spectral filter for visual features, while FADA (Bi et al., 2024) processes low- and high-frequency components separately. FAAR (Fontana et al., 2026) combines task-specific spectral filters with input-specific cross-task context in a multitask decoder, and Freq-LoRA (Xia et al., 2026) uses image statistics to gate a learned set of frequency bands. Building on these approaches, we seek to use the spatial context within each task-private LoRA bottleneck to derive a response over the full spatial-frequency grid without predefined frequency bands.

We therefore reinterpret each task-private branch in terms of the feature subspace it selects. Directions within this subspace may favor different spatial scales, and representational similarity measures aggregate direction-wise scores using either unweighted or importance-weighted averages (Klabunde et al., 2025). In our setting, uniform averaging gives low-variation directions equal influence, direct energy weighting can let a single high-energy direction dominate, and learned aggregation introduces additional parameters. Building on this view, we introduce Intrinsic Spectral Continuation LoRA (ISC-LoRA) for multi-task dense prediction. ISC-LoRA combines frequency preferences along the principal directions of feature variation using geometry-aware regularization and covariance normalization, yielding a bounded spectral response determined by the current input and private subspace and invariant to the rank basis, as illustrated in Figure 1.

ISC-LoRA uses the resulting scalar to set a bounded spectral gain shared across the private branch's rank channels. Its sign determines whether higher or lower frequencies are emphasized, while its magnitude and the continuation strength set the size of the change. During backpropagation, we hold the measured scalar fixed, so supervision updates the adapter through the selected response without differentiating through the measurement. We show that ISC-LoRA can reweight the spatial frequencies of the private output within a controlled range. A continuation schedule gradually reduces the modulation strength to zero as training proceeds. At that endpoint, the spectral operator becomes the identity while the learned LoRA factors remain. The final adapter therefore uses ordinary LoRA computation without additional trainable parameters or inference operations.

We evaluate the learned adapters on PASCAL-Context (Mottaghi et al., 2014) and NYUD-v2 (Silberman et al., 2012) after the spectral response has been removed. This tests whether temporary conditioning benefits the final multitask predictor. We also compare ISC-LoRA with representative multitask adaptation methods and use ablation studies to examine its main design choices. Beyond natural-image dense prediction, we extend our evaluation to cellular microscopy on ten two-dimensional Cell Tracking Challenge (CTC) datasets (Maška et al., 2023), reporting segmentation, detection, and tracking metrics.

Our contributions are summarized as follows:

- **Subspace-conditioned spectral adaptation.** We introduce ISC-LoRA, which uses the spatial information retained by each task-private LoRA branch as context to derive an input- and task-dependent response over the full spatial-frequency grid, without predefined frequency bands. This allows each branch to adjust the balance between lower-frequency structure and higher-frequency detail.
- **Intrinsic and controlled response selection.** We show that raw channel-energy summaries can select different responses after a rank-basis change that preserves the unmodulated LoRA update. ISC-LoRA instead combines geometry-aware regularization with covariance-normalized aggregation of directional frequency preferences, so the response depends on the current input and represented private subspace rather than arbitrary rank coordinates. We prove rank-basis invariance of the scalar and modulated private output and bound the resulting private-output perturbation.
- **Spectral continuation to an identity endpoint.** We gradually withdraw the spectral response until the operator is exactly the identity while retaining the LoRA factors learned under conditioning. The resulting adapter can therefore be evaluated and deployed using standard LoRA computation.

2 Related Work

Parameter-efficient multitask adaptation. Parameter-efficient multitask methods differ in how they share and specialize adaptation capacity. VMT-Adapter (Xin et al., 2024) combines shared projections with task-specific knowledge extraction, while MTLoRA (Agiza et al., 2024) separates low-rank updates into task-agnostic and task-specific components. TGLoRA (Gangwar et al., 2026) shares adapters in early layers and progressively specializes them, using gradient-based task similarity to group tasks. mtLoRA (Tian et al., 2026) combines singular-value-aware regularization, block-level adaptation, and dimension-specific routing. Together, these methods organize shared and task-specific adaptation capacity. ISC-LoRA complements these designs by using each private branch's spatial information as input- and task-dependent context.

Spectral and input-dependent adaptation. FreqFit (Ly & Nguyen, 2024; 2025) learns spectral filters between Transformer blocks, directly adapting interactions across spatial locations. FADA (Bi et al., 2024) uses low- and high-frequency branches to adapt vision foundation models for domain-generalized semantic segmentation. FAAR (Fontana et al., 2026) combines task-specific channel-wise spectral filters with cross-task frequency-consensus alignment for multitask dense prediction. Freq-LoRA (Xia et al., 2026) gates learned Gaussian frequency bands using per-image DCT statistics. Input-dependent adaptation also appears in the dynamic task filters of TADFormer (Baek et al., 2025) and the effective-rank gates of FouRA (Borse et al., 2024). Together, these methods establish spectral processing and input-dependent control as complementary directions for visual adaptation. ISC-LoRA derives the signal for its spatial response from each task-private low-rank subspace.

Factor geometry in low-rank adaptation. LoRA-Pro (Wang et al., 2025) relates factor gradients to the equivalent weight-space gradient and adjusts them to reduce its discrepancy from the full fine-tuning gradient. LoRA-RITE (Yen et al., 2025) addresses transformation dependence in adaptive factor updates. Together, these studies address factorized optimization through factor gradients and the transformation dependence of weight updates. ISC-LoRA instead asks whether equivalent rank bases select the same spectral response from task-private features.

3 Method

3.1 Private subspaces as spatial context

ISC-LoRA uses spatial-frequency variation within each task-private branch as context for its own adaptation. This context selects a bounded spectral response during training. Continuation then withdraws the response until the conditioning operator becomes the identity, and evaluation uses the resulting adapter with the response disabled. We first define the conditioning operator $\mathcal{R}_\varepsilon$ and then construct the scalar $\kappa(P, A)$ that controls its spectral response.

For one sample at a task-private insertion site, let $X \in \mathbb{R}^{d_{\text{in}} \times h \times w}$ be the adapter input feature map. The task-private down-projection $A \in \mathbb{R}^{r \times d_{\text{in}}}$ and up-projection $B \in \mathbb{R}^{d_{\text{out}} \times r}$ act on the channel dimension at each spatial position. Here, d_{in} and d_{out} are the input and output channel counts, r is the LoRA rank, and h and w are the spatial dimensions. The tensor $P = AX \in \mathbb{R}^{r \times h \times w}$ contains the spatial bottleneck features, and $BP \in \mathbb{R}^{d_{\text{out}} \times h \times w}$ is the private branch output; the usual LoRA scale is omitted for clarity. ISC-LoRA replaces this output by $B\mathcal{R}_\varepsilon(P; A)$ while leaving the pretrained and shared paths unchanged. Figure 1(a) shows this interface, and panel (b) expands ISC into an identity path and a controlled spectral residual.

For a channel map $U \in \mathbb{R}^{c \times h \times w}$, let $\mu_U \in \mathbb{R}^c$ be its channelwise spatial mean, broadcast over the grid. The conditioning operator and its channelwise spatial transform are

$$\mathcal{R}_\varepsilon(P; A) = P + \varepsilon\kappa(P, A)\mathcal{T}(P), \qquad \mathcal{T}(U) = \mathcal{F}^{-1}[s(k)\mathcal{F}(U - \mu_U)(k)]\,. \tag{1}$$

Here, $\mathcal{F}$ denotes the orthonormal spatial Fourier transform, and $k \in \Omega$ indexes the full discrete spatial-frequency grid. When $U = P$, we write $\mu = \mu_P$ for brevity. The fixed radial response is $s(k) = 2q(k) - 1$, where $q(k) \in [0, 1]$. The coordinate q orders frequencies by distance from the Fourier origin and assigns each radial shell the midpoint of its cumulative grid measure; conjugate frequencies receive the same value. Appendix C gives the precise discrete definition. The residual term in Eq. (1) contains three factors with distinct roles. Applying $\mathcal{T}$ to P supplies the spatial

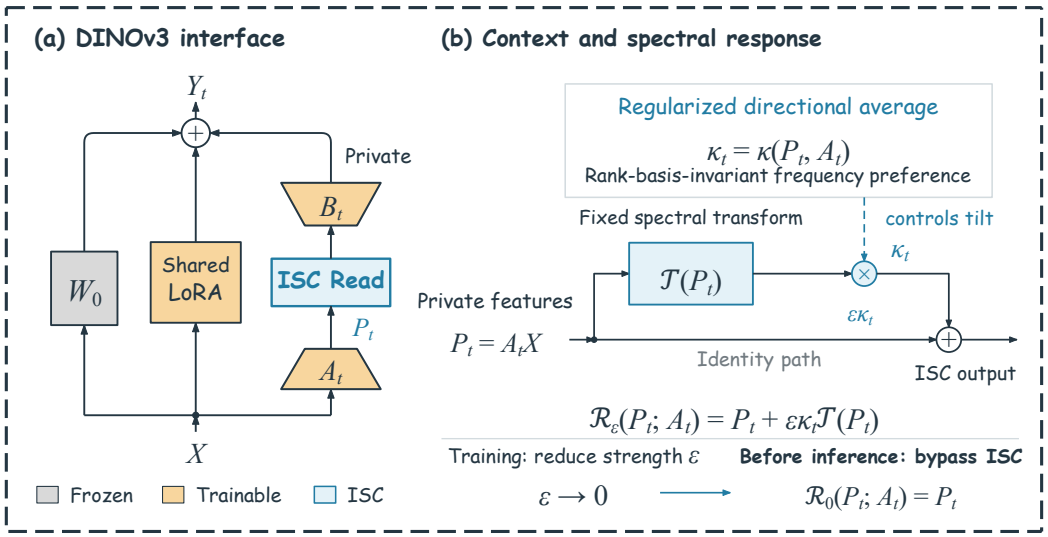

Figure 1: **ISC-LoRA uses task-private subspaces as spatial context during training.** (a) For task t, ISC is placed between the task-private factors A_t and B_t, alongside the frozen projection and shared LoRA branch. (b) For $P_t = A_t X$, a geometry-regularized average of directional frequency preferences produces the per-sample scalar κ_t. By Theorem 3.1, κ_t is invariant to invertible rank-basis changes and is detached during backpropagation. The product $\varepsilon\kappa_t$ controls the fixed, mean-centered transform $\mathcal{T}$, yielding $\mathcal{R}_\varepsilon(P_t; A_t) = P_t + \varepsilon\kappa_t\mathcal{T}(P_t)$. Continuation sets $\varepsilon = 0$ before inference, restoring ordinary LoRA; ISC introduces no trainable parameters or inference operations.

residual, $\kappa(P, A)$ sets its input-dependent sign and strength, and $\varepsilon \in [0, 1)$ bounds the overall intervention. Equivalently, each centered spectral coefficient receives the gain $g(k) = 1 + \varepsilon\kappa s(k)$. For $\kappa > 0$, the gain increases with $q(k)$, emphasizing higher frequencies relative to lower ones. For $\kappa < 0$, this pattern is reversed. The operator $\mathcal{R}_\varepsilon$ preserves the spatial mean.

Because the same gain is shared across rank channels, the modulation can also be expressed directly at the private output. The transform $\mathcal{T}$ acts only on spatial positions and uses the same multiplier for every channel, so it commutes with the pointwise linear map B. Writing $Z = BP$ gives

$$B\mathcal{R}_\varepsilon(P; A) = Z + \varepsilon\kappa(P, A)\mathcal{T}(Z). \tag{2}$$

Thus, $\kappa(P, A)$ selects the response applied to the private output $Z = BP$. Each task-private branch has its own A and hence its own bottleneck $P = AX$, so the selected response may differ across branches for the same image.

3.2 From subspace geometry to directional preference

We now construct the scalar $\kappa(P, A)$ that controls the private response in Eqs. (1) and (2). Changing the rank basis can mix or rescale the channels of P without changing BP, provided B is transformed accordingly. The measurement should therefore be insensitive to this coordinate choice. For full-row-rank A, we define

$$H_A = AA^\top, \qquad Q = H_A^{-1/2}A, \qquad QQ^\top = I_r. \tag{3}$$

Here H_A is the row Gram matrix of A, $H_A^{-1/2}$ denotes its symmetric positive-definite inverse square root, and I_r is the $r \times r$ identity matrix. The rows of Q form an orthonormal basis for the same input subspace. Replacing A by GA for an invertible G changes these coordinates only by an orthogonal transformation. This removes arbitrary scaling of the basis vectors without forcing the feature variances along different directions to be equal.

Let $S(k) = \mathcal{F}(P - \mu_P)(k) \in \mathbb{C}^r$, and let $\langle\cdot\rangle_\Omega$ denote the average over $k \in \Omega$. We describe the spatial variation using the covariance $C \in \mathbb{R}^{r\times r}$ and the frequency-weighted covariance $F \in \mathbb{R}^{r\times r}$,

defined by

$$C = \langle \mathrm{Re}[S(k)S(k)^*] \rangle_\Omega , \qquad F = \langle q(k)\, \mathrm{Re}[S(k)S(k)^*] \rangle_\Omega . \tag{4}$$

Here $S(k)^*$ denotes the conjugate transpose of $S(k)$, Re takes the real part, and $\preceq$ denotes the positive-semidefinite order. Since $q(k) \in [0,1]$, we have $0 \preceq F \preceq C$. In the orthonormal coordinates of Eq. (3), these moments become $\widehat{C} = H_A^{-1/2} C H_A^{-1/2}$ and $\widehat{F} = H_A^{-1/2} F H_A^{-1/2}$. They characterize the variance and frequency preference of each principal direction of feature variation. Assuming nonzero centered energy, write $\widehat{C}u_i = c_i u_i$ for orthonormal u_i and define

$$a = \frac{1}{r}\,\mathrm{tr}(\widehat{C}) > 0, \qquad p_i = \frac{u_i^\top \widehat{F} u_i}{c_i} \quad (c_i > 0). \tag{5}$$

Here tr denotes the matrix trace, and $a = r^{-1}\sum_{i=1}^{r} c_i$ is the mean variance across the r principal directions. The vector u_i specifies a direction in the rank-coordinate space rather than in the image plane. Its eigenvalue c_i measures variance, while $p_i \in [0,1]$ is the energy-weighted mean of the frequency coordinate q along u_i. Thus, the sign of $2p_i - 1$ indicates whether this directional mean lies above or below $1/2$. ISC-LoRA then forms a regularized average of these directional preferences.

$$\kappa(P, A) = \frac{1}{r} \sum_{c_i > 0} \frac{c_i}{c_i + \lambda a}(2p_i - 1), \qquad \lambda > 0. \tag{6}$$

The positive parameter λ controls the strength of regularization. If a positive eigenvalue is repeated, the individual p_i values may depend on the chosen orthonormal basis of its eigenspace. The shrinkage coefficient is constant within that eigenspace, and the sum of the corresponding Rayleigh quotients is basis independent, so its total contribution to Eq. (6) is unchanged. Each direction contributes at most $1/r$ in magnitude. Increasing its variance therefore does not give it influence proportional to its energy, while the shrinkage factor attenuates directions with little variance relative to a. Variance still affects the signal through shrinkage factors normalized by their common scale a. Consequently, its sign can differ from that produced by direct spectral-energy pooling.

An energy centroid can be dominated by a few high-variance directions even when several other directions express a different frequency preference. ISC-LoRA instead uses the average in Eq. (6) to select which frequencies to emphasize during training. Preferences with the same sign reinforce one another, whereas opposing signs can cancel. Their aggregate determines a single shared radial gain; the operator does not assign a separate gain to each covariance direction.

The directional expression also gives a compact, basis-independent computation. Adding the same regularization λa to every orthonormal coordinate gives the ridge $\lambda a I_r$, which becomes $\lambda a H_A$ in the original coordinates. With

$$\begin{aligned} R_\lambda &= \lambda a H_A, & D &= C + R_\lambda, & N &= F + \tfrac{1}{2} R_\lambda, \\ \kappa(P, A) &= \frac{2}{r}\,\mathrm{tr}(ND^{-1}) - 1, & a &= \frac{1}{r}\,\mathrm{tr}(H_A^{-1} C), \end{aligned} \tag{7}$$

we recover exactly the average in Eq. (6). The half-weighted ridge in N assigns the neutral centroid $1/2$ to directions with zero sample variance. The implementation evaluates the traces using small matrix solves rather than explicit matrix inverses. When C is invertible, the unregularized trace at $\lambda = 0$ is already rank-basis invariant. With positive regularization, the H_A-based ridge preserves this consistency even when the sample covariance is singular.

The geometry and the aggregation rule have separate roles. Regularization based on H_A preserves rank-basis invariance, while covariance normalization determines how strongly each direction influences the signal. A geometry-normalized energy centroid, denoted κ_H in Appendix A, is also invariant but weights directional preferences in proportion to their variance. It agrees with ISC-LoRA at rank one or when all r covariance eigenvalues are equal and positive. When directional energies differ, the two summaries assign different weights; heterogeneous preferences can then change both the magnitude and the sign of the selected response. Appendix A separates the effects of directional weighting and overall shrinkage.

3.3 Consistent and bounded private responses

We next establish two complementary properties. The theorem gives the central structural guarantee that the modulated private output remains unchanged under a change of rank basis. The subsequent

proposition bounds the response and establishes its identity endpoint. Both results hold under exact arithmetic, assuming that A has full row rank, the centered energy is nonzero, and $\lambda > 0$. Neither requires a rank condition on B. We use $\mathrm{GL}(r)$ to denote the general linear group of invertible real $r \times r$ matrices, and $\|\cdot\|_F$ to denote the Frobenius norm over all entries.

Theorem 3.1 (Forward rank-basis invariance). *For $G \in \mathrm{GL}(r)$, let $A' = GA$, $B' = BG^{-1}$, and $P' = GP$. Then*

$$\kappa(P', A') = \kappa(P, A), \quad \mathcal{R}_\varepsilon(P'; A') = G\mathcal{R}_\varepsilon(P; A), \quad B'\mathcal{R}_\varepsilon(P'; A') = B\mathcal{R}_\varepsilon(P; A). \tag{8}$$

Under the basis change, a is invariant, and each of C, F, H_A, and R_λ transforms by the congruence $M \mapsto GMG^\top$, whereas ND^{-1} transforms by the similarity $ND^{-1} \mapsto GND^{-1}G^{-1}$. The trace is therefore preserved, and the shared spatial gain gives the same private output after applying the correspondingly transformed up-projection B'. Thus, *intrinsic* refers to the choice of rank coordinates within the selected subspace. Equal products BA involving different selected subspaces fall outside this equivalence, as illustrated in Appendix D. This invariance concerns the forward response at a fixed parameter state; equivalent parameterizations need not follow the same optimization trajectory.

Proposition 3.2 (Bounded response and identity endpoint). *For $0 \le \varepsilon < 1$, $|\kappa| \le 1$, and for every $k \in \Omega$, the centered spectral gain satisfies $1 - \varepsilon \le 1 + \varepsilon\kappa s(k) \le 1 + \varepsilon$. Moreover,*

$$\begin{aligned} \|\mathcal{R}_\varepsilon(P; A) - P\|_F &\le \varepsilon \|P - \mu_P\|_F, \\ \|B(\mathcal{R}_\varepsilon(P; A) - P)\|_F &\le \varepsilon \|B(P - \mu_P)\|_F. \end{aligned} \tag{9}$$

The operator preserves the spatial mean and satisfies $\mathcal{R}_0(P; A) = P$.

The bound on κ follows from $0 \preceq N \preceq D$, and Parseval's identity transfers the resulting gain bound to the spatial response. Because the gain is shared across channels, the second inequality directly bounds the perturbation by the centered component of the private output. Together, the theorem and proposition establish rank-basis-consistent response selection and bound the resulting perturbation at each private insertion site. Appendix B provides the complete proofs.

3.4 Using the context during supervised adaptation

For each update, the measured statistic sets the spectral gain but is not itself differentiated through. We therefore hold it fixed during backpropagation, as indicated by the dashed control arrow in Figure 1(b), and recompute it on the next forward pass. Let $\mathrm{stopgrad}(z)$ return z on the forward pass and have zero derivative on the backward pass. For $\bar{\kappa} = \mathrm{stopgrad}(\kappa(P, A))$ and a feature perturbation δP, the local differential used on the feature path is

$$J_{\mathrm{used}}[\delta P] = \delta P + \varepsilon\bar{\kappa}\mathcal{T}(\delta P). \tag{10}$$

The supervised-loss gradient still passes through $P = AX$ and B, so both LoRA factors remain trainable. We detach the spectral moments and adapter geometry used to compute κ, removing the term $\varepsilon\mathcal{T}(P)\, d\kappa$ that differentiating through response selection would introduce. The next forward pass recomputes the context from the updated factors and features.

Continuation makes this conditioning temporary. Let $\xi \in [0, 1]$ denote normalized withdrawal progress, with $\xi = 0$ and $\xi = 1$ at the first and last updates, respectively. We use a cosine schedule for the modulation strength.

$$\varepsilon(\xi) = \frac{\varepsilon_{\max}}{2}\big(1 + \cos(\pi\xi)\big), \qquad 0 \le \varepsilon_{\max} < 1. \tag{11}$$

Here $\varepsilon_{\max}$ is the value of ε at the start of withdrawal. At the end of the interval, $\varepsilon = 0$ and the operator is exactly the identity. Any remaining training with the identity operator is specified by the experimental protocol; some protocols end when withdrawal is complete. The method introduces no trainable parameters. At evaluation, $\mathcal{R}_0(P; A) = P$, so the learned LoRA factors are retained and the ISC operator adds no inference operations beyond ordinary LoRA. Evaluation therefore measures the adaptation retained in the LoRA factors after temporary conditioning, rather than the immediate forward effect of the response.

Table 1: Swin-T/ImageNet-1K results on PASCAL-Context, three-task NYUD-v2, and CTC.

Method	PASCAL-Context				NYUD-v2 (3 tasks)			CTC (local)		
	SEG	PART mIoU (%) ↑	SAL	Norm. mErr (°) ↓	SEG mIoU (%) ↑	Depth RMSE ↓	Norm. mErr (°) ↓	SEG↑	DET↑	TRA↑
MTLoRA	67.90	59.84	65.40	16.60	41.52	**0.62**	24.99	0.668	0.845	0.834
DiTASK	70.09	59.03	64.55	17.47	41.13	0.65	27.25	0.663	0.848	0.837
TADFormer	70.82	60.45	65.88	16.48	40.85	0.64	27.48	0.672	0.853	0.843
TGLoRA	70.53	60.96	66.12	16.43	41.84	**0.62**	24.38	0.679	0.855	0.849
FAAR	**72.02**	61.25	66.11	16.35	41.27	0.63	26.35	0.682	0.856	0.847
Free Sinewich	71.25	61.38	66.24	16.14	41.93	**0.62**	24.38	0.655	0.845	0.835
ISC-LoRA	71.53	**61.39**	**66.25**	**16.12**	**41.95**	**0.62**	**24.35**	**0.684**	**0.863**	**0.851**

The conditioning statistic, fixed-context gradient rule, and continuation schedule define ISC-LoRA. Appendix E gives the full algorithm, numerical conventions, and the additional computation required while conditioning is active.

4 Experiments

The empirical question is whether task-private spectral context improves the ordinary LoRA adapter left at the identity endpoint. We report task-wise comparisons and matched response-off ablations, then examine a training trajectory and a same-state conditioner probe. Boldface marks the best displayed value in each column and block, including ties.

4.1 Experimental setup

Datasets and evaluation protocols. We evaluate PASCAL-Context (Mottaghi et al., 2014) and NYUD-v2 (Silberman et al., 2012) for multitask dense prediction. PASCAL-Context uses four tasks; NYUD-v2 uses separately reported three- and four-task settings, with edge detection added in the latter. For ten two-dimensional Cell Tracking Challenge (CTC) datasets (Maška et al., 2023), we train on public sequence 01 and evaluate locally on public sequence 02 using official SEG, DET, and TRA measures. The CTC tables report equal-weight macro-averages over the ten datasets. Task definitions and evaluation details are given in Appendices E.4 and E.5.

Model and training configurations. Our primary backbone is ImageNet-1K-pretrained Swin-T (Liu et al., 2021); we also evaluate ImageNet-22K-pretrained Swin-T and DINOv3 ViT-L/16 and ViT-7B/16 (Siméoni et al., 2026). All backbones remain frozen. We evaluate ISC-LoRA after removing the training-time spectral response, using standard LoRA inference. Appendix E.4 gives adapter ranks and scaling, DINOv3 insertion sites, heads, input sizes, and the LoRA factor learning-rate ratio; the CTC decoder and association protocol are in Appendix E.5.

4.2 Main Results

Multitask adaptation with Swin-T. Table 1 compares ISC-LoRA on Swin-T/ImageNet-1K with low-rank multitask adapters (MTLoRA (Agiza et al., 2024) and TGLoRA (Gangwar et al., 2026)), task-adaptive methods (DiTASK (Mantri et al., 2025), TADFormer (Baek et al., 2025)), frequency-aware feature adaptation (FAAR (Fontana et al., 2026)), and sinusoidal weight modulation (Free Sinewich (Liu et al., 2026)). ISC-LoRA has the highest displayed PASCAL part and saliency mIoU, and the lowest normal error. FAAR leads semantic mIoU (72.02 versus 71.53). On three-task NYUD-v2, ISC-LoRA has the highest segmentation mIoU and lowest normal error, while depth RMSE ties at 0.62. Because the methods differ in adaptation design, these rankings alone do not isolate the effect of ISC's task-private spectral conditioning. Table 3 therefore compares ISC-LoRA with ordinary LoRA within matched settings.

Different pretrained representations. Table 2 extends the comparison to ImageNet-22K-pretrained Swin-T and DINOv3 ViT-L/16 and ViT-7B/16. The Swin-T block revisits MTLoRA,

Table 2: Dense prediction results for ImageNet-22K-pretrained Swin-T and DINOv3 ViT-L/16 and ViT-7B/16. Results are grouped by backbone; NYUD-v2 uses four jointly trained tasks.

	PASCAL-Context				NYUD-v2 (4 tasks)				CTC (local)		
Method	SEG↑	PART↑ mIoU (%)	SAL↑	Norm.↓ Error	SEG↑ mIoU (%)	Depth↓ RMSE	Norm.↓ Error	Edge↑ ODS-F	SEG↑	DET↑	TRA↑
Swin-T / ImageNet-22K											
MTLoRA	68.71	61.32	65.37	16.44	41.11	0.65	24.55	0.583	0.680	0.852	0.842
DiTASK	69.66	62.02	65.00	17.10	37.36	0.75	28.63	0.571	0.657	0.853	0.838
TADFormer	**72.47**	62.00	65.99	16.48	41.37	**0.64**	24.54	0.585	0.682	0.861	0.849
Free Sinewich	72.43	62.18	66.04	16.42	42.26	**0.64**	23.41	0.594	0.682	0.860	0.847
ISC-LoRA	72.46	**62.25**	**66.14**	**16.41**	**42.33**	**0.64**	**23.36**	**0.597**	**0.685**	**0.866**	**0.854**
DINOv3 ViT-L/16											
CvM	84.25	77.63	65.28	15.37	64.98	0.39	15.97	**0.816**	0.677	0.866	0.857
M2H-MX	84.72	**78.63**	65.30	15.19	65.32	0.38	15.54	0.813	0.726	0.872	0.868
ISC-LoRA	**84.75**	78.61	**65.33**	**15.16**	**66.97**	**0.37**	**15.52**	**0.816**	**0.735**	**0.881**	**0.875**
DINOv3 ViT-7B/16											
CvM	84.96	80.19	65.41	14.99	65.32	0.35	14.74	0.823	0.715	0.883	0.892
M2H-MX	85.37	80.41	65.49	14.72	66.39	**0.33**	13.77	0.825	0.740	0.878	0.881
ISC-LoRA	**85.65**	**80.66**	**65.84**	**14.04**	**67.85**	**0.33**	**13.69**	**0.827**	**0.742**	**0.890**	**0.894**

Table 3: NYUD-v2 ablation results for ordinary LoRA and conditioner variants.

	Swin-T / 1K, NYUD (3 tasks)			DINOv3-7B, NYUD (4 tasks)			
Conditioning	SEG↑	Depth↓	Norm.↓	SEG↑	Depth↓	Norm.↓	Edge↑
Ordinary LoRA	41.315	0.617	25.073	67.669	0.338	13.921	0.811
H-energy	41.732	0.619	24.445	67.768	0.334	13.911	0.825
Shrinkage-matched H-energy	41.490	0.617	25.075	67.818	0.333	13.720	0.820
Euclidean metric	41.488	0.617	24.889	67.760	0.335	13.706	**0.827**
Reversed ISC	41.787	0.617	24.659	67.806	0.334	13.723	0.825
ISC-LoRA	**41.951**	**0.616**	**24.353**	**67.851**	**0.331**	**13.692**	**0.827**

DiTASK, TADFormer, and Free Sinewich. The DINOv3 blocks report scores from our reproductions of CvM (Wang et al., 2026) (cross-view module) and M2H-MX (Udugama et al., 2026) (multi-scale decoding and cross-task refinement). ISC-LoRA leads on several tasks across the Swin-T and DINOv3 settings, although the backbone architectures and pretraining datasets differ. These results across backbones suggest that ISC-LoRA's task-wise advantages are not confined to a single pretrained representation.

Cell segmentation, detection, and tracking. The CTC columns evaluate adapters using a shared flow and probability prediction head and fixed instance reconstruction and association procedures. In each backbone block, ISC-LoRA has the highest displayed local ten-dataset macro-averages on sequence 02 for SEG, DET, and TRA. All three metrics come from the same instance and track processing pipeline, so their aligned rankings describe end-to-end system performance rather than three independent learned effects.

4.3 Ablation Study

The preceding tables compare several methods. Table 3 tests whether training-time conditioning changes the response-off predictor in two NYUD-v2 settings. Within each setting, ordinary LoRA and ISC-LoRA share the backbone, adapter configuration, head architecture, optimizer, and training budget. The selected ISC checkpoints improve all seven metrics. Both methods use standard LoRA inference at evaluation, so the differences reflect trained adapters and task heads.

Geometric regularization and directional aggregation. These controls separate design properties that score rankings alone cannot distinguish. H-energy weights directions by energy, so a

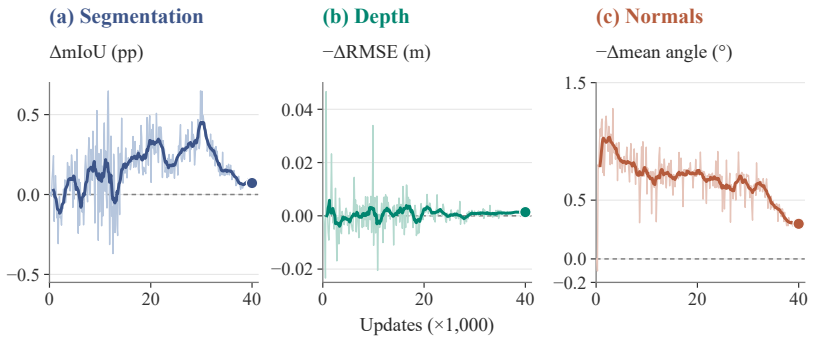

Figure 2: Task-wise validation gaps during Swin-T training on three-task NYUD-v2. Positive values favor ISC-LoRA. Pale lines show observations; dark lines show trends.

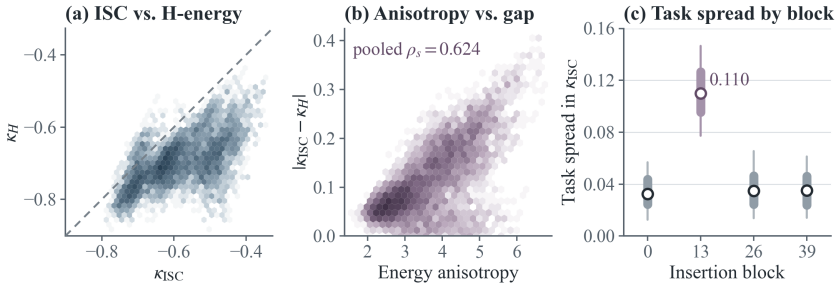

Figure 3: Variation of the response coefficient. Fixed training features from DINOv3 ViT-7B/16 on four-task NYUD-v2. The panels compare the two coefficients, relate their gap to energy anisotropy, and show how ISC's coefficient varies across task branches and insertion sites.

few high-energy directions may obscure frequency preferences in lower-energy directions. The shrinkage-matched variant matches ISC's scalar prefactor at a fixed state but retains this weighting. Euclidean-metric regularization depends on the coordinate scaling of the LoRA factors, so equivalent rank-basis parameterizations of the same unconditioned update can produce different conditioning responses.

Performance trajectories and conditioner analysis. Figure 2 tracks task-wise validation gaps over training. ISC-LoRA's spectral response follows a continuation schedule during training. Positive values in all three panels favor ISC-LoRA. The segmentation gap fluctuates but remains positive late in training; the normal gap narrows but remains positive at the final update; the depth gap is small. Thus, the performance gaps vary across tasks and training stages.

Figure 3 examines ISC's training-time conditioning signal across feature geometry and task branches. (a) compares ISC and H-energy on the same ISC-LoRA states, showing agreement on tilt direction but different coefficient magnitudes. (b) relates their gap to energy anisotropy: in the pooled records, the gap tends to grow as energy becomes concentrated in fewer directions. (c) contrasts the four task branches for the same input, showing that they can produce different coefficients.

5 Conclusion

We introduced ISC-LoRA, which derives a bounded spatial-frequency response from each input and task-private LoRA subspace during training. Its construction is invariant to LoRA rank-basis changes, and continuation returns the spectral operator to the identity for standard LoRA inference. Experiments on PASCAL-Context, NYUD-v2, and ten CTC datasets show task-wise gains across natural-image and cell-analysis tasks. These results support task-private spectral context as a useful training signal for multitask dense prediction.

AI USE STATEMENT

LLMs were used solely to assist with manuscript writing. They were not used to derive proofs or conduct analyses without human verification. All LLM contributions are acknowledged here and in the submission form, in accordance with ICLR's disclosure requirements.

ETHICS STATEMENT

We use publicly available datasets and pretrained models and collect no new human-participant data.

REPRODUCIBILITY STATEMENT

The anonymous repository linked in the abstract provides the ISC-LoRA implementation and released checkpoints. Appendix B gives the proofs, and Appendices E.4 and E.5 describe the experimental configurations and evaluation protocols.

REFERENCES

Ahmed Agiza, Marina Neseem, and Sherief Reda. MTLoRA: Low-rank adaptation approach for efficient multi-task learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16196–16205, 2024.

Seungmin Baek, Soyul Lee, Hayeon Jo, Hyesong Choi, and Dongbo Min. TADFormer: Task-adaptive dynamic transformer for efficient multi-task learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14858–14868, 2025.

Qi Bi, Jingjun Yi, Hao Zheng, Haolan Zhan, Yawen Huang, Wei Ji, Yuexiang Li, and Yefeng Zheng. Learning frequency-adapted vision foundation model for domain generalized semantic segmentation. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-2983.

Shubhankar Borse, Shreya Kadambi, Nilesh Prasad Pandey, Kartikeya Bhardwaj, Viswanath Ganapathy, Sweta Priyadarshi, Risheek Garrepalli, Rafael Esteves, Munawar Hayat, and Fatih Porikli. FouRA: Fourier low-rank adaptation. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Maxime Fontana, Michael Spratling, and Miaojing Shi. FAAR: Efficient frequency-aware multi-task fine-tuning via automatic rank selection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 31135–31144, 2026.

Neeraj Gangwar, Anshuka Rangi, Rishabh Deshmukh, Holakou Rahmanian, Yesh Dattatreya, and Nickvash Kani. Parameter-efficient multi-task learning via progressive task-specific adaptation. In *Proceedings of The 29th International Conference on Artificial Intelligence and Statistics*, volume 300 of *Proceedings of Machine Learning Research*, pp. 3916–3924. PMLR, 2026.

Soufiane Hayou, Nikhil Ghosh, and Bin Yu. LoRA+: Efficient low rank adaptation of large models. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 17783–17806, 2024.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022.

Max Klabunde, Tobias Schumacher, Markus Strohmaier, and Florian Lemmerich. Similarity of neural network models: A survey of functional and representational measures. *ACM Computing Surveys*, 57(9):1–52, 2025. doi: 10.1145/3728458.

Shih-Wen Liu, Yen-Chang Chen, Wei-Ta Chu, Fu-En Yang, and Yu-Chiang Frank Wang. Frequency switching mechanism for parameter-efficient multi-task learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20273–20282, 2026.

Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin Transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 10012–10022, October 2021.

Son T. Ly and Hien V. Nguyen. Frequency strikes back: Boosting parameter-efficient foundation model adaptation for medical imaging. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2025*, volume 15965 of *Lecture Notes in Computer Science*, pp. 258–267. Springer Nature Switzerland, 2025. doi: 10.1007/978-3-032-04978-0_25.

Son Thai Ly and Hien V. Nguyen. Enhancing parameter-efficient fine-tuning of vision transformers through frequency-based adaptation, 2024. arXiv:2411.19297.

Krishna Sri Ipsit Mantri, Carola-Bibiane Schönlieb, Bruno Ribeiro, Chaim Baskin, and Moshe Eliasof. DiTASK: Multi-task fine-tuning with diffeomorphic transformations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 25218–25229, 2025.

Martin Maška, Vladimír Ulman, Pablo Delgado-Rodriguez, Estibaliz Gómez-de Mariscal, Tereza Nečasová, Fidel A Guerrero Peña, Tsang Ing Ren, Elliot M Meyerowitz, Tim Scherr, Katharina Löffler, et al. The Cell Tracking Challenge: 10 years of objective benchmarking. *Nature Methods*, 20(7):1010–1020, 2023.

Roozbeh Mottaghi, Xianjie Chen, Xiaobai Liu, Nam-Gyu Cho, Seong-Whan Lee, Sanja Fidler, Raquel Urtasun, and Alan Yuille. The role of context for object detection and semantic segmentation in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 891–898, 2014.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy V. Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Mido Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Hervé Jégou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. DINOv2: Learning robust visual features without supervision. *Transactions on Machine Learning Research*, 2024.

René Ranftl, Alexey Bochkovskiy, and Vladlen Koltun. Vision transformers for dense prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12179–12188, 2021.

Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. Indoor segmentation and support inference from RGBD images. In *Computer Vision – ECCV 2012*, volume 7576 of *Lecture Notes in Computer Science*, pp. 746–760. Springer Berlin Heidelberg, 2012. doi: 10.1007/978-3-642-33715-4_54.

Oriane Siméoni, Huy V. Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seung Eun Yi, Michael Ramamonjisoa, Francisco Massa, Daniel HAZIZA, Luca Wehrstedt, Jianyuan Wang, Timothée Darcet, Théo Moutakanni, Leonel Sentana, Claire Roberts, Andrea Vedaldi, Jamie Tolan, John Brandt, Camille Couprie, Julien Mairal, Herve Jegou, Patrick Labatut, and Piotr Bojanowski. DINOv3. *Transactions on Machine Learning Research*, 2026. ISSN 2835-8856. Featured Certification.

Carsen Stringer, Tim Wang, Michalis Michaelos, and Marius Pachitariu. Cellpose: A generalist algorithm for cellular segmentation. *Nature Methods*, 18(1):100–106, 2021.

Zichen Tian, Antoine Ledent, and Qianru Sun. Scalable multi-task low-rank model adaptation. In *International Conference on Learning Representations*, 2026.

U. V. B. L. Udugama, George Vosselman, and Francesco Nex. M2H-MX: Multi-task semantic and geometric perception for real-time monocular 3D scene graph construction. In *ICRA 2026 Workshop on Semantics for Reliable Robot Autonomy: From Environment Understanding and Reasoning to Safe Interaction*, 2026. Workshop paper; preprint arXiv:2603.29236.

Xiaoye Wang, Chen Tang, Xiangyu Yue, and Wei-Hong Li. 3D-Aware multi-task learning with cross-view correlations for dense scene understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5793–5803, 2026.

Zhengbo Wang, Jian Liang, Ran He, Zilei Wang, and Tieniu Tan. LoRA-Pro: Are low-rank adapters properly optimized? In *The Thirteenth International Conference on Learning Representations*, 2025.

Yingwei Xia, Tian Yu, Wang Xi, Fan Wang, Yong Liu, Nanhao Liang, and Wen Zhang. Freq-LoRA: Frequency-domain low-rank adaptation for weather-robust aircraft segmentation in EO remote sensing. *Remote Sensing*, 18(16):2674, 2026. doi: 10.3390/rs18162674.

Yi Xin, Junlong Du, Qiang Wang, Zhiwen Lin, and Ke Yan. VMT-Adapter: Parameter-efficient transfer learning for multi-task dense scene understanding. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(14):16085–16093, 2024. doi: 10.1609/aaai.v38i14.29541.

Haiming Yao, Wei Luo, Qiyu Chen, Jianxing Liao, and Wei You. Parameter-, memory-, time-efficient multi-task dense vision adaptation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(14):11848–11856, 2026. doi: 10.1609/aaai.v40i14.38171.

Jui-Nan Yen, Si Si, Zhao Meng, Felix Yu, Sai Surya Duvvuri, Inderjit S. Dhillon, Cho-Jui Hsieh, and Sanjiv Kumar. LoRA done RITE: Robust invariant transformation equilibration for LoRA optimization. In *International Conference on Learning Representations*, 2025.

ISC-LoRA: Intrinsic Spectral Continuation for Multitask Dense Prediction

———— Appendix ————

The structure of the Appendix is as follows,

- Appendix A: directional interpretation and energy-based alternatives.
- Appendix B: proofs of rank-basis invariance and response bounds.
- Appendix C: the discrete spatial-frequency coordinate.
- Appendix D: the scope of the rank-basis statement.
- Appendix E: algorithm, experimental configurations, numerical conventions, and evaluation scope.

A Interpreting the conditioning signal

This appendix first derives the directional interpretation of the ISC statistic, then relates it to energy-based alternatives. The distinction is useful throughout the method: subspace geometry determines which coordinate changes leave the signal unchanged, whereas covariance normalization determines how the measured directions contribute to that signal.

A.1 From orthonormal coordinates to the trace expression

Let $H_A = AA^\top \succ 0$ and $Q = H_A^{-1/2}A$. Since $QQ^\top = I_r$, the features $H_A^{-1/2}P = QX$ describe the selected input subspace in orthonormal row coordinates. For $A' = GA$, define

$$O = H_{A'}^{-1/2} G H_A^{1/2}. \tag{12}$$

Then $OO^\top = I_r$ and $Q' = OQ$. Consequently, the corresponding moments $\widehat{C} = H_A^{-1/2} C H_A^{-1/2}$ and $\widehat{F} = H_A^{-1/2} F H_A^{-1/2}$ change only by orthogonal congruence. This construction removes arbitrary coordinate stretching within the selected subspace; it does not whiten the sample covariance.

Write $\widehat{C}u_i = c_i u_i$ in an orthonormal eigenbasis, and let $a = r^{-1}\sum_i c_i > 0$. For $c_i > 0$, set $p_i = u_i^\top \widehat{F} u_i / c_i$. The order $0 \preceq \widehat{F} \preceq \widehat{C}$ gives $p_i \in [0,1]$. Transforming the trace into these coordinates yields

$$\begin{aligned}\kappa &= \frac{2}{r}\operatorname{tr}\Big[(\widehat{F} + \tfrac{1}{2}\lambda a I_r)(\widehat{C} + \lambda a I_r)^{-1}\Big] - 1 \\ &= \frac{1}{r}\sum_{i=1}^{r} \frac{2u_i^\top \widehat{F} u_i - c_i}{c_i + \lambda a} = \frac{1}{r}\sum_{c_i>0} \frac{c_i}{c_i + \lambda a}(2p_i - 1).\end{aligned} \tag{13}$$

When $c_i = 0$, the same matrix order implies $u_i^\top \widehat{F} u_i = 0$, so that direction contributes zero. The half-weight ridge in the numerator gives this neutrality in the trace expression: it cancels the denominator's ridge when the centered directional preference is formed.

Equation (13) separates preference from influence. The term $2p_i - 1$ describes where a direction's variation lies on the fixed radial coordinate, while $c_i/(c_i + \lambda a)$ shrinks directions with little variance relative to the common scale. Each signed term contributes at most $1/r$ in magnitude. Variance nevertheless remains relevant: changing one eigenvalue also changes a, thereby changing the shrinkage of the other directions. ISC thus controls direct energetic dominance without assigning equal influence to every direction.

If an eigenvalue is repeated, individual p_i values can depend on the eigenbasis within its eigenspace. Their sum, and hence the contribution of that eigenspace, is unchanged. Opposing directional preferences can also cancel. A small κ may therefore accompany pronounced spatial structure; it describes the aggregate preference, rather than the relevance of individual features to a task.

A.2 SEPARATING GEOMETRY FROM AGGREGATION

An energy centroid in the same subspace geometry gives the alternative

$$\kappa_H = \frac{1}{1+\lambda}\left(\frac{2\,\mathrm{tr}(H_A^{-1}F)}{\mathrm{tr}(H_A^{-1}C)} - 1\right) = \frac{1}{1+\lambda}\sum_{c_i>0}\frac{c_i}{\sum_j c_j}(2p_i - 1). \tag{14}$$

Both traces in the ratio are rank-basis invariant. Unlike ISC, this summary allocates influence in proportion to directional energy. The two constructions coincide at rank one, or when all r covariance eigenvalues are equal and positive. These are sufficient conditions for agreement, not necessary ones. Unequal directional energies can change the response magnitude; heterogeneous frequency preferences can also change its sign.

The comparison also changes the total weight assigned to the preferences. Let

$$h_i = \frac{c_i}{c_i + \lambda a}, \qquad \tau = \frac{1}{r}\sum_i h_i, \qquad w_i = \frac{h_i}{r\tau}, \qquad \kappa = \tau\sum_{c_i>0} w_i(2p_i - 1). \tag{15}$$

Here $\tau > 0$ and $\sum_i w_i = 1$. Since $t/(t+\lambda a)$ is strictly concave for $t \geq 0$, Jensen's inequality gives

$$0 < \tau \leq \frac{a}{a+\lambda a} = \frac{1}{1+\lambda}, \tag{16}$$

with equality when all r covariance eigenvalues are equal. H-energy uses the fixed prefactor $1/(1+\lambda)$ and weights $c_i/(ra)$. ISC therefore changes both relative directional influence and overall shrinkage. Matching nominal strength alone does not isolate these effects; the realized output perturbation also depends on the private spectrum.

The shrinkage-matched H-energy control in Table 3 keeps H-energy's directional weights but replaces its fixed prefactor with the ISC value τ measured at the same state:

$$\kappa_{H,\tau} = \tau\sum_{c_i>0}\frac{c_i}{\sum_j c_j}(2p_i - 1). \tag{17}$$

It therefore matches ISC's scalar shrinkage while retaining energy-weighted aggregation.

For a concrete example, take $H_A = I_3$, directional energies $(100, 10, 10)$, and frequency coordinates $(3/16, 13/16, 13/16)$. With $\lambda = 0.01$, the undamped energy statistic is -0.416667, $\kappa_H \approx -0.412541$, and $\kappa \approx +0.193138$. The first direction carries 83.3% of the energy, yet the other two determine the sign of the directional average. This algebraic example explains why following κ need not reinforce the dominant raw spectral energy. Whether that choice benefits supervision remains an empirical question.

The geometry of the regularizer can be isolated by replacing H_A with I_r:

$$a_E = \frac{\mathrm{tr}(C)}{r}, \qquad R_E = \lambda a_E I_r, \qquad \kappa_E = \frac{2}{r}\,\mathrm{tr}\big[(F + R_E/2)(C + R_E)^{-1}\big] - 1. \tag{18}$$

This preserves the normalized-trace aggregation while changing the positive ridge. If C is invertible and $\lambda = 0$, the unregularized statistic $2\,\mathrm{tr}(FC^{-1})/r - 1$ is already invariant: congruence of F, C becomes similarity of FC^{-1}. Thus H_A is needed here to make positive regularization respect the same geometry, including when C is singular.

The Euclidean-energy statistic changes both the geometry and the aggregation:

$$\kappa_{EE} = \frac{1}{1+\lambda}\left(\frac{2\,\mathrm{tr}(F)}{\mathrm{tr}(C)} - 1\right). \tag{19}$$

These alternatives distinguish the choices made by the ISC construction. A common nominal strength ε does not imply equal output perturbations: for a nonzero private output, $\|B(\mathcal{R} - P)\|_F/\|BP\|_F$ describes the realized relative change. This ratio is undefined when its denominator is zero. The reversed-ISC control in Table 3 uses $-\kappa$ in place of κ with the same radial coordinate, nominal ε, and continuation schedule.

B PROOFS OF THE OPERATOR PROPERTIES

Throughout this section, assume exact arithmetic, full row rank of A, nonzero centered energy, and $\lambda > 0$. Then $H_A \succ 0$, $a > 0$, and $D = C + \lambda a H_A \succ 0$. The sample covariance may be singular, and B need not have full column rank.

B.1 FORWARD RANK-BASIS INVARIANCE

For any real $G \in \mathrm{GL}(r)$, set $A' = GA$, $P' = GP$, and $B' = BG^{-1}$. Centering and the spatial Fourier transform commute with G, so $\mu' = G\mu$ and $S'(k) = GS(k)$. It follows that

$$H_{A'} = GH_AG^\top, \qquad C' = GCG^\top, \qquad F' = GFG^\top. \tag{20}$$

The average subspace variance is unchanged, because

$$a' = \frac{1}{r}\operatorname{tr}\big(G^{-\top}H_A^{-1}CG^\top\big) = \frac{1}{r}\operatorname{tr}(H_A^{-1}C) = a. \tag{21}$$

Consequently, $R'_\lambda = GR_\lambda G^\top$, $D' = GDG^\top$, and $N' = GNG^\top$. In particular,

$$N'D'^{-1} = GND^{-1}G^{-1}, \qquad \kappa(P'; A') = \kappa(P; A). \tag{22}$$

The fixed spatial operator acts identically on every rank channel, giving $\mathcal{T}(GP) = G\mathcal{T}(P)$. Therefore

$$\mathcal{R}_\varepsilon(P'; A') = GP + \varepsilon\kappa(P; A)G\mathcal{T}(P) = G\mathcal{R}_\varepsilon(P; A), \tag{23}$$

$$B'\mathcal{R}_\varepsilon(P'; A') = B\mathcal{R}_\varepsilon(P; A). \tag{24}$$

The inverse transformation in B' cancels the coordinate change in the conditioned features. Any unmodulated prefix tokens satisfy the same identity through the ordinary product $B'A' = BA$. Stop-gradient leaves these forward equalities unchanged; it determines the backward rule discussed in Appendix E.

B.2 GAIN AND OUTPUT PERTURBATION BOUNDS

For a real vector v, $v^\top \operatorname{Re}[S(k)S(k)^*]v = |v^\top S(k)|^2 \geq 0$. Since $q(k) \in [0, 1]$, we have $0 \preceq F \preceq C$, and hence $0 \preceq N \preceq D$. All eigenvalues of $D^{-1/2}ND^{-1/2}$ lie in $[0, 1]$. Its trace equals that of ND^{-1}, yielding $|\kappa| \leq 1$. Together with $|s(k)| \leq 1$, this gives $1 - \varepsilon \leq g(k) \leq 1 + \varepsilon$.

Using the orthonormal Fourier transform and the conjugate multiplicities $m(k)$ from Appendix C, Parseval's identity gives

$$\|\mathcal{R}_\varepsilon(P; A) - P\|_F^2 = \varepsilon^2\kappa^2 \sum_{k\in\Omega^+} m(k)s(k)^2\|S(k)\|_2^2 \leq \varepsilon^2\|P - \mu\|_F^2, \tag{25}$$

$$\|B(\mathcal{R}_\varepsilon(P; A) - P)\|_F^2 = \varepsilon^2\kappa^2 \sum_{k\in\Omega^+} m(k)s(k)^2\|BS(k)\|_2^2 \leq \varepsilon^2\|B(P - \mu)\|_F^2. \tag{26}$$

Taking square roots proves both bounds. The second follows directly because $B\mathcal{T}(P) = \mathcal{T}(BP)$; no additional operator-norm factor for B is needed. A LoRA scale multiplying the private branch multiplies both sides by its absolute value.

Centering makes the DC coefficient zero, and a scalar Fourier multiplier leaves that coefficient zero. Adding μ therefore restores the original spatial mean. Setting $\varepsilon = 0$ gives $\mathcal{R}_0(P; A) = P$ exactly. These statements concern the response at one private insertion for one input. They neither identify optimizer trajectories under a change of factor coordinates nor bound the change in the entire network's predictions. In particular, the input dependence of κ is not covered by a two-input Lipschitz or global-invertibility claim.

C THE DISCRETE SPATIAL COORDINATE

The fixed coordinate assigns a radial order to every Fourier location, without partitioning the spectrum into a fixed number of bands. For an $h \times w$ map, the stored real Fourier grid is

$\Omega^+_{h,w} = \{0, \dots, h-1\} \times \{0, \dots, \lfloor w/2 \rfloor\}$. Its radius and conjugate multiplicity are

$$\rho(k)^2 = \frac{\min(k_y, h-k_y)^2}{h^2} + \frac{k_x^2}{w^2}, \qquad m(k) = \begin{cases} 1, & k_x = 0 \text{ or } (w \text{ even and } k_x = w/2), \\ 2, & \text{otherwise.} \end{cases} \tag{27}$$

The total measure is $M = \sum_{k \in \Omega^+} m(k) = hw$. We assign each radial shell the midpoint of its cumulative grid measure:

$$q(k) = \frac{\sum_{\rho(j) < \rho(k)} m(j) + \frac{1}{2} \sum_{\rho(j) = \rho(k)} m(j)}{M}, \qquad s(k) = 2q(k) - 1. \tag{28}$$

Equal-radius frequencies share a coordinate, with exact shell equality determined by the integer key $\min(k_y, h-k_y)^2 w^2 + k_x^2 h^2$. This avoids floating-point equality tests. Conjugate multiplicity recovers the moments of the full Fourier grid from the stored half-spectrum:

$$C = \frac{1}{M} \sum_{k \in \Omega^+} m(k) \operatorname{Re}[S(k)S(k)^*], \qquad F = \frac{1}{M} \sum_{k \in \Omega^+} m(k) q(k) \operatorname{Re}[S(k)S(k)^*]. \tag{29}$$

To see why the coordinate has a neutral midpoint, let b_j denote successive shell masses and $d_j = \sum_{i<j} b_i$ their preceding cumulative masses. Then

$$\sum_j b_j (d_j + b_j/2) = \sum_{i<j} b_i b_j + \frac{1}{2} \sum_j b_j^2 = \frac{1}{2} \left(\sum_j b_j \right)^2 = \frac{M^2}{2}. \tag{30}$$

Thus the grid-measure averages of q and s are $1/2$ and zero, respectively. This neutrality concerns the grid measure, rather than the energy distribution of a particular sample. The coordinate orders relative radial frequencies at each resolution; it does not identify physical wavelengths across resolutions. Mean preservation follows separately from centering and restoring the DC component, as established in Appendix B.

D THE SCOPE OF RANK-BASIS INVARIANCE

The distinction between changing coordinates and changing the selected subspace can be made explicit with small spatial examples. They also show why a condition measured before B can differ from one measured on the current private output.

D.1 ONE UPDATE, DIFFERENT ENERGY-BASED INTERVENTIONS

On a 4×4 grid, let $P_1(y, x) = \sqrt{2} \cos(\pi x/2)$ and $P_2(y, x) = \sqrt{2} \cos(\pi x/2 + \pi y)$. With $A = B = I_2$ and $X = P$ after flattening, the orthonormal transform gives $C = I_2$, $F = \operatorname{diag}(3/16, 13/16)$, and $H_A = I_2$. The coordinates follow directly from Eq. (28).

For $G_1 = \operatorname{diag}(10, 1)$, use $A' = G_1 A$ and $B' = B G_1^{-1}$. Although $B'A' = BA$, the moments become $C' = \operatorname{diag}(100, 1)$ and $F' = \operatorname{diag}(300/16, 13/16)$. The complementary rescaling $G_2 = \operatorname{diag}(1, 10)$ favors the other coordinate. At $\lambda = 0.01$, the three statistics take the values

Conditioner	Original basis	Basis G_1	Basis G_2
ISC	0	0	0
Euclidean metric	0	-0.103289	$+0.103289$
Euclidean energy	0	-0.606558	$+0.606558$

The compensated factorizations represent the same unconditioned update, yet the Euclidean statistics select opposite interventions. Multiplication by B' cancels the coordinate scaling but cannot cancel the changed scalar gain. This example establishes the ambiguity of the specified conditioning rules; it makes no assumption that every spectral adaptation method uses either rule.

D.2 Equal products with different selected subspaces

Let $A_1 = [e_1, e_2]^\top$, $A_2 = [e_1, e_3]^\top$, and $B = [1\ \ 0]$, where the e_j are the standard basis vectors of $\mathbb{R}^3$. Both A_i have full row rank and $BA_1 = BA_2 = e_1^\top$, but their row spaces differ. No invertible G maps A_1 to A_2.

Take the first two input channels from the preceding example and set $X_3(y, x) = \sqrt{2}\cos(\pi y/2)$. Both selected subspaces have $H = I_2$ and $C = I_2$, while $F_1 = \mathrm{diag}(3/16, 13/16)$ and $F_2 = \mathrm{diag}(3/16, 3/16)$. At $\lambda = 0.01$, $\kappa_1 = 0$ and $\kappa_2 = -125/202$. For $\varepsilon = 0.1$, the private outputs are X_1 and $(1 + 125/3232)X_1$, respectively. Equal products therefore need not give equal conditioned outputs when the unused directions of A differ.

If the product itself has rank r, its row space equals that of every size-r full-rank factor A. Any two such row bases satisfy $A_2 = GA_1$ for an invertible G; equality of the products then gives $B_2 = B_1G^{-1}$. The theorem covers this case. With rank-deficient B, it instead guarantees consistency within a fixed rank-basis orbit, which is the meaning of intrinsic conditioning used here.

D.3 Subspace context at zero readout

Let $R = \mathcal{R}_\varepsilon(P; A)$ and let E be the upstream gradient of BR, with spatial locations flattened into columns. At $B = 0$, both BR and BP vanish, but

$$\nabla_B L = ER^\top = EP^\top + \varepsilon\kappa E\mathcal{T}(P)^\top. \tag{31}$$

Thus the selected subspace can supply context for the first update of B before it produces a nonzero output. A normalized energy statistic based solely on BP would encounter zero energy at this point. This explains a consequence of measuring the available subspace rather than the current readout; it does not assert that all retained directions are useful, or establish superiority over every output-conditioned construction.

E Algorithm and implementation

For completeness, flatten the spatial positions into $X \in \mathbb{R}^{d_{\mathrm{in}} \times n}$, where $n = hw$. The task-private factors are $A_t \in \mathbb{R}^{r_t \times d_{\mathrm{in}}}$ and $B_t \in \mathbb{R}^{d_{\mathrm{out}} \times r_t}$. The operator reshapes A_tX into an $r_t \times h \times w$ map and restores the matrix layout afterwards. With these reshapes implicit, the adapted layer is

$$Y_t = W_0X + \gamma_{\mathrm{sh}}B_{\mathrm{sh}}A_{\mathrm{sh}}X + \gamma_tB_t\mathcal{R}_\varepsilon(A_tX; A_t). \tag{32}$$

Here $W_0 \in \mathbb{R}^{d_{\mathrm{out}} \times d_{\mathrm{in}}}$ is frozen, $\gamma_{\mathrm{sh}}, \gamma_t$ are the LoRA scales, and the shared factors have dimensions $A_{\mathrm{sh}} \in \mathbb{R}^{r_{\mathrm{sh}} \times d_{\mathrm{in}}}$ and $B_{\mathrm{sh}} \in \mathbb{R}^{d_{\mathrm{out}} \times r_{\mathrm{sh}}}$. The shared term is omitted in architectures without that branch. Task and insertion indices are suppressed in the operator analysis, which applies independently to each sample and private spatial map.

Algorithm 1 connects this layer form to the training rule. The spectrum remains on the live feature path, whereas measurement of κ is detached. Pale blue marks the ISC-specific context, response, and identity exit.

Algorithm 1 ISC-LoRA conditioning of one spatial map

Require: $P \in \mathbb{R}^{r\times h\times w}$, $A \in \mathbb{R}^{r\times d_{\text{in}}}$, $0 \le \varepsilon < 1$, $\lambda > 0$

if $\varepsilon = 0$ **then return** P
$P_f \leftarrow \mathrm{FP32}(P)$; $\mu \leftarrow \mathrm{Mean}_{h,w}(P_f)$
$S \leftarrow \mathrm{rFFT2}(P_f - \mu, \mathrm{norm} = \texttt{ortho})$ ▷ live feature path
Retrieve q, m from Eq. (28); $M \leftarrow hw$
begin stop-gradient
  $\overline{S} \leftarrow \mathrm{stopgrad}(S)$; $\overline{A} \leftarrow \mathrm{stopgrad}(\mathrm{FP32}(A))$
  $C \leftarrow M^{-1}\sum_k m(k)\,\mathrm{Re}[\overline{S}(k)\overline{S}(k)^*]$
  $F \leftarrow M^{-1}\sum_k m(k)q(k)\,\mathrm{Re}[\overline{S}(k)\overline{S}(k)^*]$
  Symmetrize C, F; $H_A \leftarrow \overline{A}\overline{A}^\top$
  $(L_A, \mathrm{ok}_A) \leftarrow \mathrm{SafeChol}(H_A)$; $\kappa \leftarrow 0$
  if ok_A and C, F are finite **then**
    $a \leftarrow \mathrm{tr}(\mathrm{CholSolve}(L_A, C))/r$
    if a is finite and $a > 10^{-8}$ **then**
      $R_\lambda \leftarrow \lambda a H_A$; $D \leftarrow C + R_\lambda$; $N \leftarrow F + R_\lambda/2$
      $(L_D, \mathrm{ok}_D) \leftarrow \mathrm{SafeChol}(D)$
      if ok_D **then**
        $U \leftarrow \mathrm{CholSolve}(L_D, N^\top)^\top$ ▷ solve $UD = N$
        $\kappa \leftarrow 2\,\mathrm{tr}(U)/r - 1$
      end if
    end if
  end if
end stop-gradient
$S_\varepsilon(k) \leftarrow [1 + \varepsilon\kappa(2q(k) - 1)]S(k)$
return $\mathrm{Cast}(\mathrm{irFFT2}(S_\varepsilon, \mathrm{size} = (h, w), \mathrm{norm} = \texttt{ortho}) + \mu, \mathrm{dtype}(P))$

E.1 Supervision with a fixed measured context

At a differentiable point of the undetached construction, regard P and A as independent arguments. Its differential would be

$$\delta\mathcal{R} = (I + \varepsilon\kappa\mathcal{T})\delta P + \varepsilon\mathcal{T}(P)\big(D_P\kappa[\delta P] + D_A\kappa[\delta A]\big). \tag{33}$$

ISC removes the second term by setting $\bar{\kappa} = \mathrm{stopgrad}(\kappa)$. It retains $\delta P = \delta A\, X + A\,\delta X$ through the live path, so the trainable factors still receive supervised gradients. The real, conjugate-symmetric Fourier multiplier, including centering, is self-adjoint. Its retained backward action is therefore $I + \varepsilon\bar{\kappa}\mathcal{T}$, just as its local forward action on feature perturbations is. Each update uses the current measurement as a fixed spatial context and measures it again on the next forward pass. Allowing gradients through that measurement would define a different training rule.

E.2 Numerical conventions

The transforms, moments, and solves use FP32 with autocasting disabled. In Algorithm 1, $\mathrm{SafeChol}$ requires a successful Cholesky factorization and finite factor entries; $\mathrm{CholSolve}(L, V)$ solves $LL^\top U = V$. The construction of H_A and its factorization can be shared across samples in a batch. The denominator solve and energy validity test remain sample-specific.

The reference implementation assigns $\kappa = 0$ if metric factorization fails, or for a sample with nonfinite moments, invalid energy, or a failed denominator factorization. This recovers identity in exact arithmetic for finite features, with Fourier round-off in the numerical path. The $\varepsilon = 0$ branch instead returns P before transforms or statistics.

The zero-conditioner fallback does not sanitize nonfinite input spectra, and the implementation does not separately check the final solved scalar for finiteness. The theorem therefore concerns the stated mathematical domain, rather than arbitrary invalid numerical inputs. Finite-precision factorization decisions and the absolute energy threshold are likewise outside its coordinate-invariance guarantee.

E.3 CONTINUATION, SPATIAL LAYOUT, AND COMPUTATION

During the continuation interval, each supervised update uses the scheduled ε, remeasures its detached contexts, and updates the live trainable parameters. Once ε reaches zero, the remaining training and evaluation use the identity branch. This response schedule is independent of the optimizer's learning-rate schedule. Head-only initialization, any shared anchor, head freezing during the final phase, and dynamic task weighting belong to the common experimental recipe. Using the same dynamic weighting rule in two runs does not require their weight trajectories to remain equal after their losses diverge.

The operator acts on complete spatial patch maps and leaves class or register tokens unchanged. At a windowed Swin insertion, the complete stage map is reconstructed by undoing the window arrangement and shift; these are restored after conditioning. This preserves the full-map definition of the operator, since window-local spectra would define a different spatial response.

For a batch of b samples, $n = hw$ spatial tokens, and one insertion site, the additional work is

$$O\big(r^2 d_{\text{in}} + b[rn \log n + r^2 n + r^3]\big)\,. \tag{34}$$

The terms account for the shared metric, spatial transforms, matrix moments, and per-sample solves. The eigendecomposition used to interpret κ is not computed by the algorithm. Scalar modulation therefore retains a training cost from measurement, while the exact identity branch removes that cost at deployment.

E.4 EXPERIMENTAL CONFIGURATION AND EVIDENCE

Tasks and adapter configuration. PASCAL-Context includes semantic segmentation, human-part segmentation, saliency estimation, and surface-normal estimation. The three-task NYUD-v2 setting includes semantic segmentation, depth, and surface normals; the four-task setting additionally trains edge detection. In the primary Swin-T setting, the shared adapter has rank 64 and each task-private adapter has rank 4, with both branch outputs scaled by 4 (Agiza et al., 2024). DINOv3 uses shared and task-private rank-8 adapters in the QKV and attention-output projections, with branch scale 2. We use $\lambda = 0.01$, $\varepsilon_{\max} = 0.1$, and the DINOv3 LoRA factor learning-rate ratio $\eta_B/\eta_A = 16$ (Hayou et al., 2024).

Prediction heads and inputs. Swin-T uses task-specific high-resolution heads and 448×448 natural-image inputs. DINOv3 uses 512×512 PASCAL-Context crops and 448×576 NYUD-v2 crops. Its PASCAL semantic and part segmentation heads and NYUD-v2 depth head follow DPT (Ranftl et al., 2021); the remaining PASCAL tasks use dense heads, and the remaining NYUD-v2 tasks use linear heads. CTC uses balanced 384×384 training patches and the shared three-channel prediction head described in Appendix E.5.

Training and checkpoint selection. For Swin-T, we use a 300-epoch training budget, following MTLoRA and TGLoRA (Agiza et al., 2024; Gangwar et al., 2026). For DINOv3, we use a budget of 40,000 optimizer updates. Table 4 summarizes the matched NYUD-v2 settings. Within each matched setting, ordinary LoRA, ISC-LoRA, and the conditioner variants use the same training budget and checkpoint candidates. For each independently trained model in these comparisons, we select one checkpoint using an aggregate of the validation metrics across all tasks in that setting. All scores from our experiments, including the ablations, use a single checkpoint per trained model rather than separate task-wise best checkpoints. For ISC-LoRA and its variants, candidate checkpoints are evaluated with the spectral response removed. The CTC experiments use sequence 01 for development and training and reserve sequence 02 for final evaluation, without selecting checkpoints on sequence 02.

Scope of the reported comparisons. In Table 1, the PASCAL-Context comparator scores are taken from published studies. For three-task NYUD-v2, the MTLoRA and TGLoRA scores are taken from TGLoRA (Gangwar et al., 2026), and the DiTASK, TADFormer, and FAAR scores from FAAR (Fontana et al., 2026); the Free Sinewich scores are our reproduction. In the Swin-T/ImageNet-22K block of Table 2, the PASCAL-Context MTLoRA and TADFormer scores are taken from Yao et al. (Yao et al., 2026), the DiTASK scores from DiTASK (Mantri et al., 2025),

Table 4: Training configurations for the NYUD-v2 experiments. IN1K and IN22K denote ImageNet-1K and ImageNet-22K pretraining, respectively. NYUD-3 comprises semantic segmentation, depth, and surface normals; NYUD-4 additionally includes edge detection. The listed DINOv3 hyperparameters apply to both ViT-L/16 and ViT-7B/16.

Configuration	Swin-T / IN1K	Swin-T / IN22K	DINOv3
Backbone	Swin-T	Swin-T	ViT-L/16 and ViT-7B/16
Task setting	NYUD-3	NYUD-4	NYUD-4
Input size	448×448	448×448	448×576
Training epochs	300	300	201
Optimizer updates	39,600	39,600	40,000
Effective batch size	6	6	4
Optimizer	AdamW	AdamW	AdamW
Adam coefficients	$(0.9, 0.999)$	$(0.9, 0.999)$	$(0.9, 0.999)$
Optimizer ϵ	10^{-8}	10^{-8}	10^{-8}
LoRA learning rates	10^{-4}	10^{-4}	$\eta_A = 10^{-5}$, $\eta_B = 1.6 \times 10^{-4}$
Prediction-head learning rates	10^{-4}	10^{-4}	Linear: 3×10^{-4}; DPT: 8×10^{-5}
Learning-rate schedule	Cosine decay	Cosine decay	Polynomial decay, power 0.9
Minimum learning rate	10^{-6}	10^{-6}	0
Weight decay	0.05	0.05	10^{-4}
Gradient clipping	Global norm 5	Global norm 5	Global norm 10
Shared / private rank	64 / 4	64 / 4	8 / 8
Shared / private output scale	4 / 4	4 / 4	2 / 2
LoRA dropout	0.05	0.05	0
Prediction heads	High-resolution heads	High-resolution heads	DPT for depth; linear heads otherwise

and the Free Sinewich scores from our reproduction. The four-task NYUD-v2 comparator scores in this block are taken from Free Sinewich (Liu et al., 2026). The published four-task NYUD-v2 depth and edge scores are divided by 100 for display. The DINOv3 CvM and M2H-MX scores, all CTC scores, and all ISC-LoRA and Table 3 scores are from our experiments.

E.5 CTC PREDICTION AND ASSOCIATION PROTOCOL

The cellular extension covers ten two-dimensional Cell Tracking Challenge (CTC) datasets (Maška et al., 2023): BF-C2DL-HSC, BF-C2DL-MuSC, DIC-C2DH-HeLa, Fluo-C2DL-Huh7, Fluo-C2DL-MSC, Fluo-N2DH-GOWT1, Fluo-N2DH-SIM+, Fluo-N2DL-HeLa, PhC-C2DH-U373, and PhC-C2DL-PSC. We use sequence 01 for development and training and reserve sequence 02 for local evaluation with the official CTC executables. Both sequences belong to the public training release, so these results are separate from the hidden-test leaderboard. Input scales, reconstruction thresholds, and association settings are fixed using sequence 01. After development, each model is refitted using all sequence-01 frames with valid shape supervision and then evaluated on sequence 02 without further tuning.

Each dataset activates its dataset-private LoRA branch, while all branches within a model share a trainable prediction head. Following the Cellpose representation (Stringer et al., 2021), the head predicts horizontal and vertical diffusion flows together with a cell-probability logit. At inference, each frame is processed using overlapping 384×384 tiles with a 64-pixel overlap. Tile predictions are combined using a Hann-weighted average and restored to the native image resolution. Instance masks are reconstructed with the Cellpose dynamics operator using a cell-probability threshold of 0, a flow threshold of 0.4, 200 dynamics iterations, and a minimum instance size of 15 pixels. All reported CTC reproductions use the same terminal three-channel decoder architecture and the same reconstruction settings while retaining their method-specific feature transformations.

For temporal association, we apply the same deterministic linker to every method and seed. Between each pair of consecutive frames, we compute the intersection-over-union (IoU) matrix between predicted instances and obtain a one-to-one assignment using the Hungarian algorithm. Matches with an IoU below 0.05 are discarded. Each unmatched instance in the current frame starts a new track, whereas an unmatched instance from the previous frame terminates its track. The linker does not reconnect tracks across temporal gaps or model cell divisions; all newly created tracks therefore receive a parent identifier of zero. This fixed association stage prevents method-specific tuning of the tracking procedure.

We compute SEG, DET, and TRA using the official CTC evaluation programs (Maška et al., 2023). SEG evaluates the overlap between predicted and reference instance masks, DET evaluates object detection using the normalized Acyclic Oriented Graph Matching measure for detection (AOGM-D), and TRA evaluates the predicted tracking graph using normalized Acyclic Oriented Graph Matching (AOGM). All three scores lie in $[0, 1]$, with higher values indicating better performance. For metric M and seed s, we first compute the equal-weight macro-average over the ten datasets,

$$\overline{M}_s = \frac{1}{10} \sum_{d=1}^{10} M_{d,s}. \tag{35}$$

All displayed CTC methods use the same terminal prediction, reconstruction, association, and evaluation protocol described above.