LEARNING INSTANCE-ADAPTIVE COUPLED-AXIS RELATION PROMPTS FOR ALL-IN-ONE IMAGE RESTORATION

Anonymous authors
Paper under double-blind review

ABSTRACT

All-in-one image restoration aims to recover images corrupted by unknown and heterogeneous degradations using a single model. Existing methods primarily adapt shared features or parameters through prompts and degradation representations, while leaving the underlying attention interactions generic and often dense. This separation between degradation conditioning and relation modeling provides no structured carrier for expressing instance-specific relation changes. We propose InCaRP, an all-in-one restoration Transformer that learns instance-adaptive coupled-axis relation prompts. Specifically, low-rank relation-space attention performs structured sparsification before correlation by projecting redundant key and value responses directly into a compact set of input-dependent relation carriers, while preserving the full query and output representations. Coupled-axis prompt calibration then uses shared angular retrieval to compose complementary query-axis and relation-axis prompts from paired banks. Normalized interactions between the two prompts synthesize bounded scale and shift fields that calibrate the compact relations before value aggregation. The unified design therefore combines pre-correlation information compression with instance-adaptive relation calibration. Extensive experiments on all-in-one benchmarks demonstrate that our method achieves state-of-the-art performance while maintaining parameter efficiency and shows strong, balanced performance across heterogeneous restoration tasks.

1 INTRODUCTION

Image restoration recovers clean images from observations corrupted by noise, haze, rain, motion blur, or insufficient illumination. CNN and Transformer restorers have achieved strong performance across these tasks (Zamir et al., 2021; Liang et al., 2021; Wang et al., 2022a; Zamir et al., 2022; Chen et al., 2022a). Although their architectures are reusable, conventional task-specific training maintains separate parameters for different degradations and assumes that the restoration task is known. All-in-one restoration instead learns a single blind mapping that infers the required transformation directly from the input. Existing methods enable such adaptation through learned degradation representations (Li et al., 2022; Liu et al., 2022; Zhang et al., 2023), visual, textual, or domain-aware prompts (Potlapalli et al., 2023; Conde et al., 2024; Dong et al., 2026), frequency modulation (Cui et al., 2025), heterogeneous experts (Guo et al., 2024a; Zamfir et al., 2025), or residual assembly (He et al., 2026a). Despite differing implementations, the conditioning paradigms rest on a common premise: different inputs should activate distinct restoration behaviors within a shared parameter budget.

Most adaptation mechanisms answer *what condition should be injected*, but leave *where that condition should act* largely implicit. Feature-level prompt conditioning injects task cues into a feature stream before interactions are formed (Figure 1(a)). Post-correlation sparse adaptation instead constructs a dense relation map first and then applies Top-k selection (Chen et al., 2023) or uses sparse responses to modulate a dense branch (Zhou et al., 2025) (Figure 1(b)). Transposed attention (Zamir et al., 2022) avoids resolution-dependent spatial score maps, but every query response still interacts with the complete set of key responses. Consequently, these strategies either condition features outside relation space or reduce redundancy only after exhaustive interactions have been constructed.

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

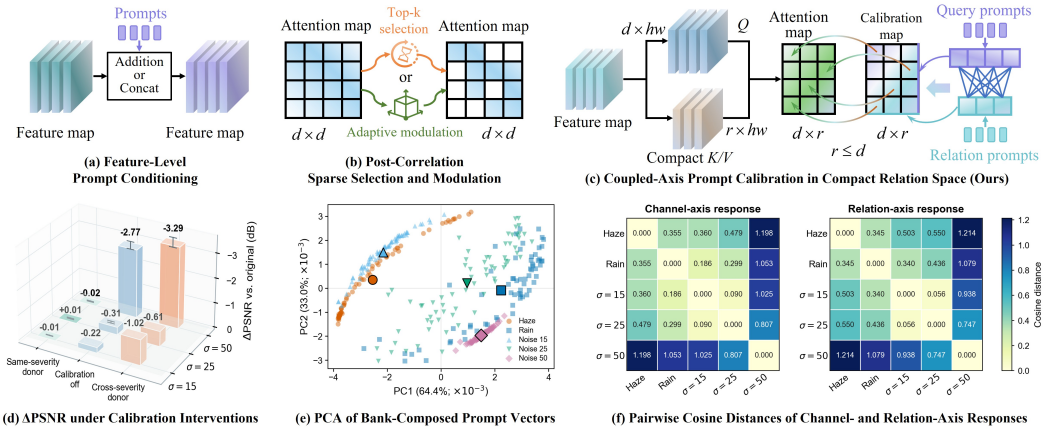


Figure 1: **Conditioning paradigms and prompt-mechanism evidence.** (a) Feature-level prompts condition representations before relation construction. (b) Post-correlation methods sparsify or modulate an already formed dense relation map. (c) InCaRP compresses K/V responses before correlation and calibrates compact logits with paired query-axis and relation-axis prompts. (d) Denoising ΔPSNR when calibration fields at all CAPC blocks are replaced by same-severity or cross-severity donors, or disabled. (e) PCA projection of bank-composed prompt vectors, outlined markers denote degradation-condition centroids. (f) Pairwise cosine distances between degradation-wise channel-axis and relation-axis response centroids.

We address this limitation by conditioning a compact interaction carrier directly. Low-Rank Relation-Space Attention (LRSA) performs structured sparsification before correlation through an asymmetric K/V projection: it condenses redundant key and value responses into r input-dependent carriers while preserving the full query, output, residual, and feed-forward pathways. The resulting rectangular relation logits expose a query-channel axis and a relation-atom axis. Coupled-Axis Prompt Calibration (CAPC) then places degradation conditioning on these logits. Each decoder state produces an image-specific retrieval query, obtains a continuous coordinate through hyperspherical matching, and uses the shared coordinate to compose complementary query-axis and relation-axis prompts from paired banks. Their normalized low-rank interaction synthesizes bounded scale and shift fields that calibrate compact relations before value aggregation. As illustrated in Figure 1(c), InCaRP therefore couples pre-correlation information compression with instance-adaptive relation calibration rather than treating prompting and relation construction as separate operations.

We place LRSA throughout a four-level encoder-decoder Transformer and instantiate CAPC in the decoder to adapt the shared features to each input condition during image reconstruction. The resulting network contains 22.7M parameters. Our contributions are summarized as follows:

- We design LRSA, which performs structured pre-correlation sparsification by projecting redundant K/V responses directly into compact input-dependent carriers while preserving the full query and output pathways.
- We introduce CAPC for instance-adaptive relation prompting. Hyperspherical query-bank retrieval provides a shared degradation coordinate for paired query-axis and relation-axis prompts, which produce bounded affine calibration directly in compact relation space.
- Extensive experiments on all-in-one image restoration benchmarks demonstrate that our method achieves state-of-the-art performance while remaining parameter-efficient and delivering strong, balanced results across heterogeneous restoration tasks.

2 Related Work

Image restoration networks. Learning-based methods exploit data priors for super-resolution (Dong et al., 2014; Liang et al., 2021; Chen et al., 2025), denoising (Zhang et al., 2017; 2018; Guo et al., 2025), dehazing (Li et al., 2017; Song et al., 2023; Jin et al., 2025), deraining (Ren et al., 2019;

Chen et al., 2023; Dong et al., 2025), deblurring (Nah et al., 2017; Kong et al., 2023; Gu et al., 2025), and low-light enhancement (Wei et al., 2018; Cai et al., 2023; Yan et al., 2025). Convolutional (Zamir et al., 2021; Chen et al., 2022a), Transformer (Wang et al., 2022a; Zamir et al., 2022; Chen et al., 2022b), and state-space (Guo et al., 2024b) backbones are reusable across tasks, but task-specific training still requires separate models. All-in-one restoration instead handles heterogeneous degradations with one model.

All-in-one image restoration. Unified weather removal (Li et al., 2020; Valanarasu et al., 2022) and subsequent restorers (Liu et al., 2022; Wang et al., 2023; Yang et al., 2025; Luo et al., 2024; Duan et al., 2024; Guo et al., 2024c) established blind restoration across multiple corruptions. AirNet (Li et al., 2022) learns discriminative degradation representations, while IDR (Zhang et al., 2023) factorizes corruptions into reusable ingredients. Later methods address task interference through collaborative learning (Wu et al., 2024), feature perturbation (Tian et al., 2025), or quality-aware supervision (Zhang et al., 2026). AdaIR (Cui et al., 2025) performs frequency modulation, DSRIR (Wang et al., 2026b) combines distinct subnets, MoCE-IR (Zamfir et al., 2025) routes features to experts, and CEA (He et al., 2026a) generates input-conditioned updates. SLER-IR (Peng et al., 2026) uses hyperspherical cues for expert routing, whereas R2R (Wang et al., 2026a) retrieves clean priors from an external bank. These approaches condition restoration through degradation representations, feature modulation, or computation routing. Our focus is a compact relation carrier jointly calibrated by axis-specific prompts.

Sparse attention and prompt learning. Attention efficiency uses factorization (Wang et al., 2020; Shen et al., 2021) or sparse interactions (Zhou et al., 2024). DRSformer (Chen et al., 2023) retains Top-k correlations, and ASTv2 (Zhou et al., 2025) uses sparse responses to calibrate dense attention. Both act on completed relation maps (Figure 1(b)), LRSA instead projects K/V responses into compact carriers before correlation. Prompt tuning supports input adaptation (Jia et al., 2022; Wang et al., 2022c;b), including visual and multimodal restoration conditioning (Ma et al., 2023; Li et al., 2026; Ai et al., 2024; Yu et al., 2024). PromptIR (Potlapalli et al., 2023) conditions features (Figure 1(a)), CPL (Wu et al., 2026) regularizes specialization, and DATPRL-IR (Dong et al., 2026) composes task and domain prompts. We instead place paired prompts on the two axes of a compact carrier for direct logit calibration (Figure 1(c)).

3 METHOD

3.1 OVERVIEW

In this work, to reduce redundant attention interactions and adapt relation modeling to unknown degradations, we propose InCaRP, coupling compact relation construction with instance-adaptive relation calibration. Given a degraded image $\mathbf{I} \in \mathbb{R}^{3\times H\times W}$, the model learns a blind mapping $\mathbf{J} = \mathcal{F}_\Theta(\mathbf{I})$ without degradation labels at inference. Specifically, Low-Rank Relation-Space Attention (LRSA) compresses K/V responses before correlation into compact carriers for shared aggregation, while Coupled-Axis Prompt Calibration (CAPC) adapts the query and relation axes to each input.

Figure 2 shows the four-level encoder-decoder architecture, which follows the multi-scale organization of restoration Transformers (Zamir et al., 2022). An overlapping convolution embeds $\mathbf{I}$ into shallow features. Relation-Space Transformer Blocks (RSTBs) progressively encode hierarchical representations as spatial resolution decreases and channel width increases. The decoder fuses corresponding skip features and reconstructs clean structures through CAPC-enhanced RSTBs (CAPC-RSTBs). Full-resolution refinement and a final residual prediction produce $\mathbf{J}$.

Each RSTB contains pre-normalized LRSA and a Gated-Dconv Feed-Forward Network (GDFN) (Zamir et al., 2022), whereas a CAPC-RSTB replaces LRSA with CAPC-LRSA. For a block input $\mathbf{X}$, the residual updates are

$$\mathbf{X}_a = \mathbf{X} + \mathcal{A}(\mathrm{LN}(\mathbf{X}))\,, \tag{1}$$

$$\mathbf{X}_o = \mathbf{X}_a + \mathcal{G}(\mathrm{LN}(\mathbf{X}_a))\,, \tag{2}$$

where $\mathcal{A}$ denotes LRSA or CAPC-LRSA, and $\mathcal{G}$ denotes GDFN. The encoder and refinement stages use RSTBs, while every decoder stage uses CAPC-RSTBs. Decoder-only calibration preserves a shared encoder and enables instance-specific relation adaptation during clean-structure reconstruction.

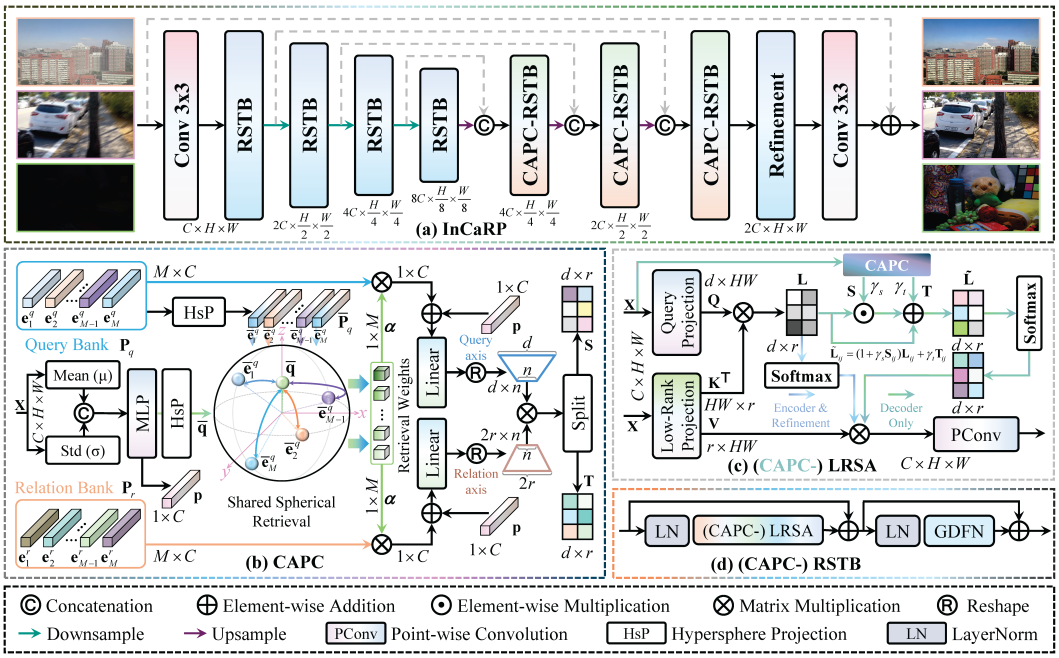


Figure 2: Overview of InCaRP. The upper panel presents the four-level encoder-decoder architecture built from RSTBs and decoder-side CAPC-RSTBs. The lower panels schematically illustrate (b) shared retrieval and paired query-axis/relation-axis prompt composition in CAPC, (c) LRSA/CAPC-LRSA, (d) the corresponding RSTB/CAPC-RSTB structure.

3.2 Low-Rank Relation-Space Attention

The *relation space* of a transposed-attention head is the logit domain formed by query–key correlations. Transposed attention (Zamir et al., 2022) avoids resolution-dependent spatial score maps, yet every query response remains correlated with the complete set of key responses. Many of these responses may encode overlapping spatial evidence, making exhaustive score allocation unnecessary for global aggregation. Top-k attention (Chen et al., 2023) selects entries after constructing the full relation map, while ASTv2 (Zhou et al., 2025) uses sparse responses to calibrate a dense branch. LRSA instead moves sparsification before correlation: a learned asymmetric projection condenses K/V information directly into input-dependent relation carriers, against which the preserved queries form only the interactions needed for aggregation.

Let $\mathbf{X} \in \mathbb{R}^{B\times C\times H\times W}$ be an input feature, h the number of attention heads, $d = C/h$ the per-head channel width, and $N = HW$. Conventional transposed attention projects queries, keys, and values to d responses per head and exhaustively correlates the query and key sets. LRSA retains the d query responses but projects the key and value streams onto a compact basis of relation rank $r < d$. Point-wise convolutions $\mathcal{P}$ mix channels, depth-wise convolutions $\mathcal{D}$ encode local context, and $\mathcal{R}$ reshapes the projections into multiple heads:

$$\begin{aligned} \mathbf{Q} &= \mathcal{R}_{h,d,N}(\mathcal{D}_q(\mathcal{P}_q(\mathbf{X}))) \,, \\ [\mathbf{K}, \mathbf{V}] &= \mathrm{Split}_r[\mathcal{R}_{h,2r,N}(\mathcal{D}_{kv}(\mathcal{P}_{kv}(\mathbf{X})))] \,. \end{aligned} \tag{3}$$

The resulting tensors satisfy $\mathbf{Q} \in \mathbb{R}^{B\times h\times d\times N}$ and $\mathbf{K}, \mathbf{V} \in \mathbb{R}^{B\times h\times r\times N}$. Suppressing batch and head indices, each head constructs a rectangular relation map and aggregates the associated value carriers:

$$\begin{aligned} \mathbf{L} &= \tau \mathbf{Q}\mathbf{K}^\top \in \mathbb{R}^{d\times r}, \quad \mathbf{A}_r = \mathrm{Softmax}(\mathbf{L}), \\ \Delta\mathbf{X} &= \mathbf{A}_r\mathbf{V} \in \mathbb{R}^{d\times N}, \qquad \mathbf{Y} = \mathcal{P}_o\big(\mathcal{R}^{-1}[\Delta\mathbf{X}]\big) \,, \end{aligned} \tag{4}$$

where τ is a learnable head-wise scaling factor, softmax operates over the r carriers, and $\mathcal{P}_o$ denotes the output projection. The K/V projections derive both addressing and response carriers from the current feature rather than a fixed external basis. Each row of $\mathbf{A}_r$ consequently selects a different

mixture of the value responses for one query channel. The resulting K/V carriers serve as compact *relation atoms*, with key carriers determining channel-to-atom affinities and paired value carriers supplying the corresponding spatial information. Full query and output widths are preserved because the bottleneck is confined to the key–value side of global aggregation.

The asymmetric construction also imposes an explicit aggregation constraint. For each head, $\mathrm{rank}(\Delta\mathbf{X}) = \mathrm{rank}(\mathbf{A}_r\mathbf{V}) \leq r$, so the d query-channel updates lie in at most r shared value-response directions. Concatenating h heads gives an attention-update rank of at most hr before output projection. The constraint does not limit the full-width residual or feed-forward pathways and therefore separates representation width from the rank budget assigned to global interaction. Compared with exhaustive transposed attention, the per-head relation memory decreases from $\mathcal{O}(d^2)$ to $\mathcal{O}(dr)$ and the corresponding computation decreases from $\mathcal{O}(Nd^2)$ to $\mathcal{O}(Ndr)$. Beyond reducing redundant score allocation, the compact domain exposes a query-channel axis and a relation-atom axis for the coupled prompt calibration described next.

3.3 Instance-Adaptive Coupled-Axis Prompt Calibration

The rectangular LRSA logits expose two non-interchangeable axes. The d query-channel axis identifies the feature channels requesting global information, while the r relation-atom axis identifies the compact K/V responses supplying that information. A single shared prompt ties the conditioning of both roles, whereas independently retrieved prompts may describe inconsistent degradation states. CAPC therefore derives a query-axis prompt and a relation-axis prompt from one instance-specific retrieval coordinate. Separate prompt contents specialize the two axes, and the shared retrieval distribution keeps them aligned with the same decoder state. The related but non-identical response geometries in Figure 1(f) are consistent with these distinct functional roles.

For every CAPC-RSTB input $\mathbf{X}$, channel-wise mean and standard deviation summarize the current feature distribution. A two-layer projection ϕ with a SiLU activation transforms the statistical descriptor into an image prompt $\mathbf{p}$:

$$\mathbf{s}(\mathbf{X}) = \mathrm{Concat}(\mathrm{Mean}_{H,W}(\mathbf{X}), \mathrm{Std}_{H,W}(\mathbf{X})) \in \mathbb{R}^{2C}, \tag{5}$$

$$\mathbf{p} = \phi(\mathbf{s}(\mathbf{X})) \in \mathbb{R}^{C}. \tag{6}$$

Recomputing $\mathbf{p}$ at every decoder block produces an instance-adaptive and layer-specific condition. The common image prompt supplies degradation evidence for both relation axes, while paired retrieval resolves the evidence into complementary axis-specific contents.

An unnormalized inner product entangles prompt direction with magnitude and may bias retrieval toward large-norm entries. Hypersphere Projection (HsP) instead establishes angular similarity on the unit hypersphere. Let $\mathbf{P}_q = [\mathbf{e}_1^q, \ldots, \mathbf{e}_M^q]^\top$ and $\mathbf{P}_r = [\mathbf{e}_1^r, \ldots, \mathbf{e}_M^r]^\top$ denote paired query-prompt and relation-prompt banks, where $\mathbf{P}_q, \mathbf{P}_r \in \mathbb{R}^{M\times C}$. HsP explicitly normalizes the projected retrieval query and every query-prompt entry before computing the shared retrieval distribution:

$$\overline{\mathbf{q}} = \frac{\mathbf{W}_q\mathbf{p}}{\|\mathbf{W}_q\mathbf{p}\|_2}, \qquad \overline{\mathbf{e}}_m^q = \frac{\mathbf{e}_m^q}{\|\mathbf{e}_m^q\|_2}, \tag{7}$$

$$\alpha_m = \frac{\exp\big(\overline{\mathbf{q}}^\top \overline{\mathbf{e}}_m^q\big)}{\sum_{j=1}^{M} \exp\big(\overline{\mathbf{q}}^\top \overline{\mathbf{e}}_j^q\big)}. \tag{8}$$

The weights $\boldsymbol{\alpha}$ are invariant to the norms of the retrieval query and query-bank entries and define a continuous coordinate over the learned query-prompt directions. A soft mixture allows nearby degradation states to share prompt pairs rather than enforcing a discrete task assignment. Because both relation axes originate from the same feature state, the same coordinate retrieves their complementary prompt contents:

$$\mathbf{p}_q = \mathbf{p} + \sum_{m=1}^{M} \alpha_m \mathbf{e}_m^q, \qquad \mathbf{p}_r = \mathbf{p} + \sum_{m=1}^{M} \alpha_m \mathbf{e}_m^r. \tag{9}$$

The query-axis prompt $\mathbf{p}_q$ augments the image prompt with retrieved query-bank content and controls the d query channels. The relation-axis prompt $\mathbf{p}_r$ incorporates the paired relation-bank content and controls the r relation atoms. Sharing $\boldsymbol{\alpha}$ aligns both prompts with one instance-specific degradation

coordinate, while separate banks retain axis-specific information. Coupling therefore arises from shared retrieval rather than identical prompt content.

Generating an independent scale and shift for every logit would require dense prompt-to-relation prediction. CAPC instead follows the rectangular geometry of LRSA and factorizes calibration along the two axes through a compact dimension n. The query-axis prompt produces a normalized factor for each query channel, while the relation-axis prompt produces normalized scale and shift factors for every relation atom:

$$\mathbf{U} = \mathrm{Norm}_n(g_q(\mathbf{p}_q)) \in \mathbb{R}^{B\times h\times d\times n}, \tag{10}$$

$$[\mathbf{V}_s, \mathbf{V}_t] = \mathrm{Split}_r[\mathrm{Norm}_n(g_r(\mathbf{p}_r))], \quad \mathbf{V}_s, \mathbf{V}_t \in \mathbb{R}^{B\times h\times r\times n}. \tag{11}$$

The operator Norm_n normalizes every factor vector along the n-dimensional coordinate. The query factors $\mathbf{U}$ and the relation-atom factors $\mathbf{V}_s, \mathbf{V}_t$ subsequently interact through angular products, generating scale and shift fields over the rectangular relation map:

$$\mathbf{S} = \mathbf{U}\mathbf{V}_s^\top, \qquad \mathbf{T} = \mathbf{U}\mathbf{V}_t^\top, \qquad \mathbf{S}, \mathbf{T} \in \mathbb{R}^{B\times h\times d\times r}, \tag{12}$$

$$|S_{ij}| \le 1, \qquad |T_{ij}| \le 1, \qquad \mathrm{rank}(\mathbf{S}), \mathrm{rank}(\mathbf{T}) \le n. \tag{13}$$

Each S_{ij} or T_{ij} jointly encodes query channel i and relation atom j as their cosine affinity. Equation (8) retrieves paired prompt contents in C dimensions, whereas Eqs. (11)–(13) synthesize bounded low-rank interactions in n dimensions. The factorized construction follows the rectangular geometry of LRSA and avoids direct dense prompt-to-attention generation (Ha et al., 2017; He et al., 2026a).

Given the base logits $\mathbf{L}$ from Eq. (4), the coupled fields perform instance-conditioned affine calibration before value aggregation:

$$\widetilde{\mathbf{L}} = \mathbf{L} \odot (\mathbf{1} + \gamma_s \mathbf{S}) + \gamma_t \mathbf{T}, \tag{14}$$

$$\mathcal{A}_{\mathrm{rp}} = \mathrm{Softmax}(\widetilde{\mathbf{L}})\mathbf{V}. \tag{15}$$

The multiplicative field adjusts the contrast of existing query–atom matches, whereas the additive field changes their relative preference before softmax. Combining both fields allows degradation conditioning to strengthen, suppress, or reorder compact relations without altering the value carriers. Small initial values for γ_s and γ_t keep each decoder block close to the shared LRSA transformation at the beginning of training and gradually introduce conditional calibration. The resulting design separates three capacities through d, r, and n, corresponding to feature representation, compact relation aggregation, and prompt calibration.

4 Experiments

4.1 Experimental settings

We evaluate InCaRP under the standard AIO-3 and AIO-5 protocols. The two settings progressively expand degradation diversity, allowing us to assess whether compact relation modeling and instance-adaptive calibration maintain accurate, balanced restoration within a single shared model. We next describe the implementation and evaluation protocols.

Implementation Details. We implement the model in PyTorch (Paszke et al., 2019) and train it on one NVIDIA GeForce RTX 5090 GPU. The four encoder stages use $[4, 6, 6, 8]$ RSTBs, widths $[48, 96, 192, 384]$, and $[1, 2, 4, 8]$ heads, the decoder uses $[4, 4, 2]$ CAPC-RSTBs and four refinement RSTBs. Each head sets $d = 48$ and $r = 32$, yielding a 48×32 relation map. The ten CAPC-RSTBs use $M = 8$, $n = 4$, and $\gamma_s = \gamma_t = 10^{-3}$. Both models are trained for 120 epochs on 128×128 crops with rotations and flips and batch size 12, using AdamW (Loshchilov & Hutter, 2019), an initial learning rate of 2×10^{-4}, and cosine annealing (Loshchilov & Hutter, 2017).

Datasets and Evaluation Metrics. Following standard protocols (Li et al., 2022; Potlapalli et al., 2023; Zhang et al., 2023; He et al., 2026a), each model is trained jointly and tested without degradation labels. AIO-3 evaluates dehazing, deraining, and denoising on SOTS (Li et al., 2019), Rain100L (Yang et al., 2020), and BSD68 (Martin et al., 2001) with $\sigma \in \{15, 25, 50\}$. AIO-5 adds GoPro deblurring (Nah et al., 2017) and LOL-v1 low-light enhancement (Wei et al., 2018), using $\sigma = 25$ on BSD68. We report PSNR and SSIM (Wang et al., 2004).

Table 1: Quantitative comparison (PSNR (dB, ↑), SSIM (↑)) under the three-degradation setting. **Best** and second-best results are shown in bold and underlined, respectively.

Method	Venue	Params.	Dehazing		Deraining		Denoising						Average	
			SOTS		Rain100L		BSD68$_{\sigma=15}$		BSD68 $_{\sigma=25}$		BSD68 $_{\sigma=50}$		PSNR	SSIM
DL	TPAMI'2021	2M	26.92	0.931	32.62	0.931	33.05	0.914	30.41	0.861	26.90	0.740	29.98	0.876
AirNet	CVPR'2022	9M	27.94	0.962	34.90	0.967	33.92	0.933	31.26	0.888	28.00	0.797	31.20	0.910
PromptIR	NeurIPS'2023	36M	30.58	0.974	36.37	0.972	33.98	0.933	31.31	0.888	28.06	0.799	32.06	0.913
Gridformer	IJCV'2024	34M	30.37	0.970	37.15	0.972	33.93	0.931	31.37	0.887	28.11	0.801	32.19	0.912
DA-CLIP	ICLR'2024	125M	29.46	0.963	36.28	0.968	30.02	0.821	24.86	0.585	22.29	0.476	28.58	0.763
InstructIR	ECCV'2024	17M	30.22	0.959	37.98	0.978	34.15	0.933	31.52	0.890	28.30	0.803	32.43	0.913
VLU-Net	CVPR'2025	35M	30.71	0.980	38.93	0.984	34.13	0.935	31.48	0.892	28.23	0.804	32.70	0.919
MoCE-IR	CVPR'2025	25M	31.34	0.979	38.57	0.984	34.11	0.932	31.45	0.888	28.18	0.800	32.73	0.917
AdaIR	ICLR'2025	29M	31.06	0.980	38.64	0.983	34.12	0.935	31.45	0.892	28.19	0.802	32.69	0.918
DA-RCOT	TPAMI'2025	50M	31.26	0.977	38.36	0.983	33.98	0.934	31.33	0.890	28.10	0.801	32.60	0.917
R2R	CVPR'2026	20M	31.40	0.977	37.46	0.980	34.10	0.936	31.45	**0.895**	28.22	0.806	32.53	0.918
CPLIR	TPAMI'2026	35M	31.27	0.980	38.77	**0.985**	34.15	0.933	31.50	0.889	28.23	0.800	32.78	0.917
FIT	ICML'2026	–	31.11	0.981	38.92	0.984	34.23	0.936	31.55	0.894	28.32	0.806	32.83	0.920
BaryIR	TPAMI'2026	36M	31.40	0.980	39.02	**0.985**	34.16	0.935	31.54	0.892	28.25	0.802	32.86	0.919
InCaRP (ours)	–	23M	**32.03**	**0.983**	**39.12**	**0.985**	**34.24**	**0.937**	**31.58**	**0.895**	**28.33**	**0.808**	**33.06**	**0.922**

Table 2: Quantitative comparison (PSNR (dB, ↑), SSIM (↑)) under the five-degradation setting. **Best** and second-best results are shown in bold and underlined, respectively.

Method	Venue	Params.	Dehazing		Deraining		Denoising		Deblurring		Low-Light		Average	
			SOTS		Rain100L		BSD68$_{\sigma=25}$		GoPro		LOLv1		PSNR	SSIM
DL	TPAMI'2021	2M	20.54	0.826	21.96	0.762	23.09	0.745	19.86	0.672	19.83	0.712	21.05	0.743
AirNet	CVPR'2022	9M	21.04	0.884	32.98	0.951	30.91	0.882	24.35	0.781	18.18	0.735	25.49	0.847
PromptIR	NeurIPS'2023	36M	26.54	0.949	36.37	0.970	31.47	0.886	28.71	0.881	22.68	0.832	29.15	0.904
DA-CLIP	ICLR'2024	125M	29.78	0.968	35.65	0.962	30.93	0.885	27.31	0.838	21.66	0.828	29.07	0.896
Gridformer	IJCV'2024	34M	26.79	0.951	36.61	0.971	31.45	0.885	29.22	0.884	22.59	0.831	29.33	0.904
InstructIR	ECCV'2024	17M	27.10	0.956	36.84	0.973	31.40	0.873	29.40	0.886	23.00	0.836	29.55	0.908
VLU-Net	CVPR'2025	35M	30.84	0.980	38.54	0.982	31.43	0.891	27.46	0.840	22.29	0.833	30.11	0.905
MoCE-IR	CVPR'2025	25M	30.48	0.974	38.04	0.982	31.34	0.887	30.05	0.899	23.00	0.852	30.58	0.919
AdaIR	ICLR'2025	29M	30.53	0.978	38.02	0.981	31.35	0.888	28.12	0.858	23.00	0.845	30.20	0.910
DA-RCOT	TPAMI'2025	50M	30.96	0.975	37.87	0.980	31.23	0.888	28.68	0.872	23.25	0.836	30.40	0.911
R2R	CVPR'2026	20M	30.64	0.974	36.61	0.975	31.35	0.891	**30.93**	0.911	22.88	0.856	30.48	0.921
CPLIR	TPAMI'2026	35M	30.82	0.978	38.20	**0.983**	31.41	0.887	28.72	0.874	**23.65**	0.855	30.55	0.915
FIT	ICML'2026	–	30.81	0.979	**38.56**	**0.983**	**31.57**	0.893	29.52	0.890	23.13	0.848	30.72	0.919
BaryIR	TPAMI'2026	36M	31.20	0.979	38.10	0.982	31.43	0.891	29.51	0.889	23.37	0.854	30.72	0.919
InCaRP (ours)	–	23M	**32.22**	**0.983**	38.37	**0.983**	31.50	**0.894**	30.92	**0.912**	23.53	**0.858**	**31.31**	**0.926**

4.2 All-in-One Restoration Results

Under the standard AIO-3 and AIO-5 protocols, we evaluate the proposed method on both benchmarks. Across the two settings, we compare against DL (Fan et al., 2021) and representative AiOIR methods, including AirNet (Li et al., 2022), PromptIR (Potlapalli et al., 2023), Gridformer (Wang et al., 2024), DA-CLIP (Luo et al., 2024), InstructIR (Conde et al., 2024), VLU-Net (Zeng et al., 2025), MoCE-IR (Zamfir et al., 2025), AdaIR (Cui et al., 2025), DA-RCOT (Tang et al., 2025), CPLIR (Wu et al., 2026), R2R (Wang et al., 2026a), FIT (He et al., 2026b), and BaryIR (Tang et al., 2026).

Three Degradations. The AIO-3 setting jointly evaluates dehazing, deraining, and Gaussian denoising at three noise levels. As shown in Table 1, our model achieves an average PSNR/SSIM of 33.06 dB/0.922. Compared with the recent BaryIR, which provides the strongest competing average PSNR, the proposed approach improves PSNR by 0.20 dB and SSIM by 0.003. Compared with FIT, the corresponding gains are 0.23 dB and 0.002. The advantages are most pronounced on the two weather-restoration tasks. On SOTS, our framework surpasses BaryIR by 0.63 dB and FIT by 0.92 dB, while on Rain100L, InCaRP exceeds R2R and FIT by 1.66 dB and 0.20 dB, respectively. Improvements across both dehazing and deraining show that the overall gain is not tied to a single favorable task. Moreover, the resulting model achieves the best or tied-best value for all 12 reported metrics, including 10 outright wins and two ties, demonstrating a consistently strong performance profile across the AIO-3 benchmark.

Five Degradations. The more challenging AIO-5 protocol additionally considers motion deblurring and low-light enhancement. As shown in Table 2, our approach improves the average PSNR/SSIM

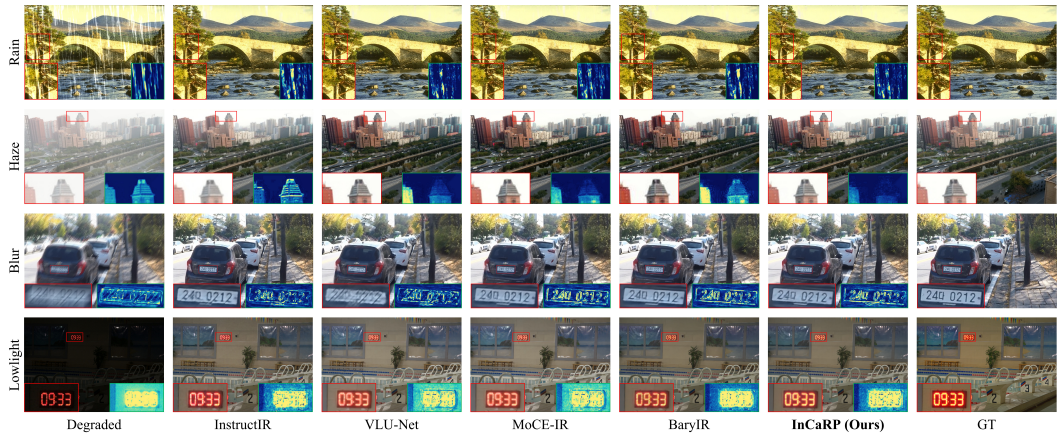


Figure 3: Qualitative comparison across four degradation types with state-of-the-art methods. Red boxes highlight enlarged regions for detailed comparison. Error maps with green boxes visualize the corresponding reconstruction discrepancies, where brighter colors indicate larger errors.

from 30.72 dB/0.919 achieved by the recent FIT to 31.31 dB/0.926, corresponding to gains of 0.59 dB/0.007. The improvements are substantial on SOTS, GoPro, and LOLv1, where the proposed model exceeds FIT by 1.41, 1.40, and 0.40 dB, respectively, demonstrating that the average gain is supported by several distinct restoration tasks. Compared with R2R, InCaRP further improves the average PSNR/SSIM by 0.83 dB/0.005 and obtains gains of 1.58 dB on SOTS and 0.65 dB on LOLv1. Meanwhile, our framework essentially matches the GoPro PSNR of R2R (30.92 vs. 30.93 dB) while attaining a higher SSIM (0.912 vs. 0.911), preserving strong deblurring performance alongside the gains on other tasks. Overall, the proposed method uses only 23M parameters, 13M fewer than BaryIR and 27M fewer than DA-RCOT, while achieving higher average restoration performance on both benchmarks and a more balanced performance profile across tasks.

Qualitative Results. Figure 3 compares representative outputs for deraining, dehazing, deblurring, and low-light enhancement. The enlarged regions show that our model produces clean sky structures after deraining, restores distant building contrast in dehazing, resolves license-plate characters after deblurring, and recovers luminous digits without excessive haloing under low illumination. While competing restorers exhibit varying strengths across degradation types, the proposed approach preserves local structures and global appearance consistently across all four examples. The qualitative behavior accords with the balanced AIO-5 performance reported in Table 2.

Additional results on CDD-11 compositional degradations, together with quantitative analysis, are provided in Appendix B.1.

4.3 Ablation Studies

We conduct ablation studies under the AIO-3 setting to evaluate the contribution of each proposed component. Unless otherwise specified, all variants follow the main training and evaluation settings. Additional analyses are provided in Appendix C.

Core modules and hyperspherical retrieval. Table 3 evaluates LRSA, CAPC, and hyperspherical retrieval against exhaustive transposed attention. LRSA improves the average from 32.09 dB/0.913 to 32.38 dB/0.913. Higher PSNR with unchanged SSIM indicates that compact K/V carriers suppress redundant interactions without disrupting structural reconstruction. CAPC alone reaches 32.52 dB/0.917, showing that relation calibration is effective independently of compression. Their combination obtains 32.96 dB/0.920, exceeding either change alone and suggesting that compact carriers provide a cleaner basis for prompt calibration. Hyperspherical retrieval yields the final 33.06 dB/0.922 by eliminating vector-norm bias in prompt matching.

Relation and prompt capacity. Table 4 varies the bank size M and relation rank r. The steady gains from $r = 16$ to $r = 32$ indicate that an overly narrow carrier set discards useful K/V evidence, whereas the slight decline at $r = 48$ suggests that excessive capacity weakens the intended redundancy

Table 3: Ablation of LRSA, CAPC, and hyperspherical retrieval (HsR).

Variant	LRSA	CAPC	HsR	Avg. P/S
Baseline	–	–	–	32.09/0.913
+ LRSA	✓	–	–	32.38/0.913
+ CAPC	–	✓	✓	32.52/0.917
+ LRSA + CAPC	✓	✓	–	32.96/0.920
Full model	✓	✓	✓	**33.06/0.922**

Table 4: Sensitivity to prompt size M and rank r.

M	r	SOTS	Rain100L	BSD68$_{15-50}$	Avg. P/S
4	32	31.09/0.980	39.09/0.984	31.32/0.878	32.84/0.920
8	16	30.96/0.980	38.95/0.984	31.24/0.876	32.73/0.918
8	24	31.72/0.982	39.08/0.985	31.37/0.879	32.98/0.921
8	32	**32.03/0.983**	**39.12/0.985**	**31.39/0.880**	**33.06/0.922**
8	48	31.89/0.982	39.08/0.985	31.36/0.879	33.01/0.921
16	32	31.99/0.982	39.11/0.985	31.35/0.879	33.03/0.921

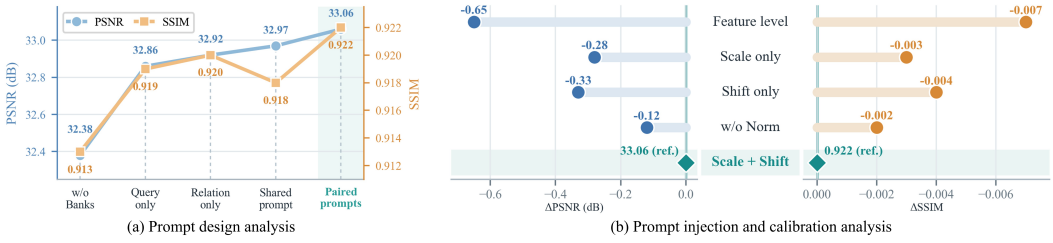


Figure 4: Prompt-related ablations. (a) Effects of prompt-bank construction and query–relation coordination. (b) Effects of feature-level conditioning and alternative logit-calibration designs.

constraint. Similarly, $M = 4$ underrepresents condition diversity, while $M = 16$ adds no benefit and may introduce redundant bank entries. Thus, $(M, r) = (8, 32)$ best balances condition coverage with compact relation modeling.

Design of query and relation prompts. Figure 4(a) shows that query-only and relation-only prompting improve the no-bank result to 32.86 dB/0.919 and 32.92 dB/0.920, confirming that both information-requesting channels and information-supplying atoms require conditioning. The slightly stronger relation-only result highlights the importance of adapting the compact carriers, but neither single axis matches paired prompting. Relative to relation-only, one shared prompt raises PSNR but lowers SSIM, suggesting that identical content conflates the two roles. Paired banks with shared retrieval instead preserve axis-specific contents under one condition coordinate and achieve 33.06 dB/0.922, consistent with Figure 1(e). Appendix C.4 further examines retrieval coordination.

Where and how should the retrieved prompts act? Figure 4(b) shows that feature modulation reaches 32.41 dB/0.915, whereas direct logit calibration obtains 33.06 dB/0.922, indicating that conditioning the interaction pattern is more effective than perturbing its input features. Scale-only and shift-only variants are inferior because one adjusts match contrast and the other changes relative preference; both are needed to reshape relation allocation. The drop without axis-factor normalization further supports bounded, complementary fields. In Figure 1(d), same-severity field replacement changes PSNR by at most 0.02 dB, while disabling calibration or using a cross-severity donor causes drops up to 2.77 and 3.29 dB. The fields therefore encode degradation-aligned relation adjustments rather than image content.

5 Conclusion

This work introduced InCaRP, an all-in-one restoration framework for handling diverse degradations with a single model. Two complementary components underpin the framework. LRSA performs structured sparsification before correlation by projecting redundant key and value responses into compact, input-dependent relation carriers. CAPC then uses paired query-axis and relation-axis prompts to calibrate the resulting interactions during reconstruction. Shared hyperspherical retrieval makes both prompts responsive to the same degraded input while preserving their distinct roles. Experiments on AIO-3 and AIO-5 show stronger average restoration performance than recent methods, with improvements across multiple tasks rather than a single favorable degradation. Component ablations further support the complementary contributions of relation construction and coupled-axis calibration. Together, these results show that the proposed framework achieves a more balanced restoration profile with fewer parameters than several strong competitors.

AI USE STATEMENT

Generative AI tools were used to aid and polish the writing, including manuscript organization, language refinement, and improvements to clarity, and to support the retrieval and discovery of relevant literature. The authors reviewed and revised all AI-assisted text, verified the retrieved references against the original sources, and take full responsibility for the final manuscript.

ETHICS STATEMENT

The authors confirm adherence to the ICLR Code of Ethics.

REPRODUCIBILITY STATEMENT

The architecture and forward computations are specified in Section 3, while Section 4 reports the datasets, evaluation protocol, optimization settings, training schedule, and hardware environment. Appendix A provides the stage-wise network configuration and complete training objective, Appendix B reports the CDD-11 compositional-restoration experiment, Appendix C contains additional ablations and retrieval analyses, and Appendix D discusses the model design, results, limitations, and future directions.

REFERENCES

Yuang Ai, Huaibo Huang, Xiaoqiang Zhou, Jiexiang Wang, and Ran He. Multimodal prompt perceiver: Empower adaptiveness generalizability and fidelity for all-in-one image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 25432–25444, 2024.

Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12504–12513, 2023.

Bin Chen, Gehui Li, Rongyuan Wu, Xindong Zhang, Jie Chen, Jian Zhang, and Lei Zhang. Adversarial diffusion compression for real-world image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 28208–28220, 2025.

Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European Conference on Computer Vision*, pp. 17–33, 2022a.

Xiang Chen, Hao Li, Mingqiang Li, and Jinshan Pan. Learning a sparse transformer network for effective image deraining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5896–5905, 2023.

Zheng Chen, Yulun Zhang, Jinjin Gu, Yongbing Zhang, Linghe Kong, and Xin Yuan. Cross aggregation transformer for image restoration. In *Advances in Neural Information Processing Systems*, volume 35, pp. 25478–25490, 2022b.

Marcos V. Conde, Gregor Geigle, and Radu Timofte. InstructIR: High-quality image restoration following human instructions. In *European Conference on Computer Vision*, pp. 1–21, 2024.

Yuning Cui, Syed Waqas Zamir, Salman Khan, Alois Knoll, Mubarak Shah, and Fahad Shahbaz Khan. AdaIR: Adaptive all-in-one image restoration via frequency mining and modulation. In *International Conference on Learning Representations*, 2025.

Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *European Conference on Computer Vision*, pp. 184–199, 2014.

Guanglu Dong, Tianheng Zheng, Yuanzhouhan Cao, Linbo Qing, and Chao Ren. Channel consistency prior and self-reconstruction strategy based unsupervised image deraining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7469–7479, 2025.

Guanglu Dong, Chunlei Li, Chao Ren, Jingliang Hu, Yilei Shi, Xiao Xiang Zhu, and Lichao Mou. Learning domain-aware task prompt representations for multi-domain all-in-one image restoration. In *International Conference on Learning Representations*, 2026.

Huiyu Duan, Xiongkuo Min, Sijing Wu, Wei Shen, and Guangtao Zhai. UniProcessor: A text-induced unified low-level image processor. In *European Conference on Computer Vision*, pp. 180–199, 2024.

Qingnan Fan, Dongdong Chen, Lu Yuan, Gang Hua, Nenghai Yu, and Baoquan Chen. A general decoupled learning framework for parameterized image operators. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(1):33–47, 2021.

Yubin Gu, Yuan Meng, Jiayi Ji, and Xiaoshuai Sun. ACL: Activating capability of linear attention for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17913–17923, 2025.

Hang Guo, Tao Dai, Yuanchao Bai, Bin Chen, Xudong Ren, Zexuan Zhu, and Shu-Tao Xia. Parameter efficient adaptation for image restoration with heterogeneous mixture-of-experts. In *Advances in Neural Information Processing Systems*, volume 37, pp. 13522–13547, 2024a.

Hang Guo, Jinmin Li, Tao Dai, Zhihao Ouyang, Xudong Ren, and Shu-Tao Xia. MambaIR: A simple baseline for image restoration with state-space model. In *European Conference on Computer Vision*, pp. 222–241, 2024b.

Hang Guo, Yong Guo, Yaohua Zha, Yulun Zhang, Wenbo Li, Tao Dai, Shu-Tao Xia, and Yawei Li. MambaIRv2: Attentive state space restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 28124–28133, 2025.

Yu Guo, Yuan Gao, Yuxu Lu, Huilin Zhu, Ryan Wen Liu, and Shengfeng He. Onerestore: A universal restoration framework for composite degradation. In *European Conference on Computer Vision*, pp. 255–272, 2024c.

David Ha, Andrew Dai, and Quoc V. Le. Hypernetworks. In *International Conference on Learning Representations*, 2017.

Haisen He, Xiangyu Zou, SongLin Dong, Heng Li, Yihong Gong, and Zhiheng Ma. Continuous expert assembly: Instance-conditioned low-rank residuals for all-in-one image restoration. *arXiv preprint arXiv:2605.06127*, 2026a.

Zihao He, Yunfeng Wu, Xinchao Wang, and Songhua Liu. Bend the basics: Degradation-aware deformable tokenization for all-in-one image restoration. In *Proceedings of the International Conference on Machine Learning*, 2026b.

Menglin Jia, Luming Tang, Bor-Chun Chen, Claire Cardie, Serge Belongie, Bharath Hariharan, and Ser-Nam Lim. Visual prompt tuning. In *European Conference on Computer Vision*, pp. 709–727, 2022.

Zhi Jin, Yuwei Qiu, Kaihao Zhang, Hongdong Li, and Wenhan Luo. MB-TaylorFormer V2: Improved multi-branch linear transformer expanded by taylor formula for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(7):5990–6005, 2025.

Lingshun Kong, Jiangxin Dong, Jianjun Ge, Mingqiang Li, and Jinshan Pan. Efficient frequency domain-based transformers for high-quality image deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5886–5895, 2023.

Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. AOD-Net: All-in-one dehazing network. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4780–4788, 2017.

Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28 (1):492–505, 2019.

Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-in-one image restoration for unknown corruption. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17452–17462, 2022.

Ruoteng Li, Robby T. Tan, and Loong-Fah Cheong. All in one bad weather removal using architectural search. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3175–3185, 2020.

Zilong Li, Yiming Lei, Chenglong Ma, Junping Zhang, and Hongming Shan. Prompt-in-prompt learning for all-in-one image restoration. *Pattern Recognition*, 180:114401, 2026.

Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. SwinIR: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pp. 1833–1844, 2021.

Lin Liu, Lingxi Xie, Xiaopeng Zhang, Shanxin Yuan, Xiangyu Chen, Wengang Zhou, Houqiang Li, and Qi Tian. TAPE: Task-agnostic prior embedding for image restoration. In *European Conference on Computer Vision*, pp. 447–464, 2022.

Ilya Loshchilov and Frank Hutter. SGDR: Stochastic gradient descent with warm restarts. In *International Conference on Learning Representations*, 2017.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.

Ziwei Luo, Fredrik K. Gustafsson, Zheng Zhao, Jens Sjolund, and Thomas B. Schon. Controlling vision-language models for multi-task image restoration. In *International Conference on Learning Representations*, 2024.

Jiaqi Ma, Tianheng Cheng, Guoli Wang, Qian Zhang, Xinggang Wang, and Lefei Zhang. ProRes: Exploring degradation-aware visual prompt for universal image restoration. *arXiv preprint arXiv:2306.13653*, 2023.

David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2, pp. 416–423, 2001.

Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 257–265, 2017.

Ozan Özdenizci and Robert Legenstein. Restoring vision in adverse weather conditions with patch-based denoising diffusion models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8):10346–10357, 2023.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Shurui Peng, Xin Lin, Shi Luo, Jincen Ou, Dizhe Zhang, Lu Qi, Truong Nguyen, and Chao Ren. SLER-IR: Spherical layer-wise expert routing for all-in-one image restoration. In *European Conference on Computer Vision*, pp. 472–491, 2026.

Vaishnav Potlapalli, Syed Waqas Zamir, Salman Khan, and Fahad Shahbaz Khan. PromptIR: Prompting for all-in-one image restoration. In *Advances in Neural Information Processing Systems*, volume 36, pp. 71275–71293, 2023.

Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. Progressive image deraining networks: A better and simpler baseline. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3937–3946, 2019.

Zhuoran Shen, Mingyuan Zhang, Haiyu Zhao, Shuai Yi, and Hongsheng Li. Efficient attention: Attention with linear complexities. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 3531–3539, 2021.

Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision transformers for single image dehazing. *IEEE Transactions on Image Processing*, 32:1927–1941, 2023.

Xiaole Tang, Xiang Gu, Xiaoyi He, Xin Hu, and Jian Sun. Degradation-aware residual-conditioned optimal transport for unified image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(8):6764–6779, 2025.

Xiaole Tang, Xiaoyi He, Jiayi Xu, Xiang Gu, and Jian Sun. Learning continuous wasserstein barycenter space for generalized all-in-one image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(7):8395–8410, 2026.

Xiangpeng Tian, Xiangyu Liao, Xiao Liu, Meng Li, and Chao Ren. Degradation-aware feature perturbation for all-in-one image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 28165–28175, 2025.

Jeya Maria Jose Valanarasu, Rajeev Yasarla, and Vishal M. Patel. TransWeather: Transformer-based restoration of images degraded by adverse weather conditions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2353–2363, 2022.

Chenxu Wang, Kai Zhang, and Jian Yang. Retrieve-to-Restore: Efficient all-in-one image restoration with a retrieval-based degradation bank. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1277–1287, June 2026a.

Sinong Wang, Belinda Li, Madian Khabsa, Han Fang, and Hao Ma. Linformer: Self-attention with linear complexity. *arXiv preprint arXiv:2006.04768*, 2020.

Tao Wang, Kaihao Zhang, Ziqian Shao, Wenhan Luo, Bjorn Stenger, Tong Lu, Tae-Kyun Kim, Wei Liu, and Hongdong Li. GridFormer: Residual dense transformer with grid structure for image restoration in adverse weather conditions. *International Journal of Computer Vision*, 132: 4541–4563, 2024.

Xiaofeng Wang, Honggang Chen, Haosong Gou, Jie He, Zhengyong Wang, Xiaohai He, Linbo Qing, and Ray E. Sheriff. RestorNet: An efficient network for multiple degradation image restoration. *Knowledge-Based Systems*, 282:111116, 2023.

Xiaofeng Wang, Xiao Liu, Yutong Yang, Zhengyong Wang, Xiaohai He, Honggang Chen, Yi Li, and Pingyu Wang. DSRIR: Dynamic spatial refinement learning for progressive all-in-one image restoration. *Information Processing & Management*, 63(2):104380, 2026b.

Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17683–17693, 2022a.

Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612, 2004.

Zifeng Wang, Zizhao Zhang, Sayna Ebrahimi, Ruoxi Sun, Han Zhang, Chen-Yu Lee, Xiaoqi Ren, Guolong Su, Vincent Perot, Jennifer Dy, and Tomas Pfister. DualPrompt: Complementary prompting for rehearsal-free continual learning. In *European Conference on Computer Vision*, pp. 631–648, 2022b.

Zifeng Wang, Zizhao Zhang, Chen-Yu Lee, Han Zhang, Ruoxi Sun, Xiaoqi Ren, Guolong Su, Vincent Perot, Jennifer Dy, and Tomas Pfister. Learning to prompt for continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 139–149, 2022c.

Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *British Machine Vision Conference*, 2018.

Gang Wu, Junjun Jiang, Kui Jiang, and Xianming Liu. Harmony in diversity: Improving all-in-one image restoration via multi-task collaboration. In *Proceedings of the ACM International Conference on Multimedia*, pp. 6015–6023, 2024.

Gang Wu, Junjun Jiang, Kui Jiang, Xianming Liu, and Liqiang Nie. Beyond degradation redundancy: Contrastive prompt learning for all-in-one image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(4):4005–4022, 2026.

Qingsen Yan, Yixu Feng, Cheng Zhang, Guansong Pang, Kangbiao Shi, Peng Wu, Wei Dong, Jinqiu Sun, and Yanning Zhang. HVI: A new color space for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5678–5687, 2025.

Wenhan Yang, Robby T. Tan, Jiashi Feng, Zongming Guo, Shuicheng Yan, and Jiaying Liu. Joint rain detection and removal from a single image with contextualized deep networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(6):1377–1393, 2020.

Yutong Yang, Xiaofeng Wang, Xin Lin, and Honggang Chen. DEMNet: A degradation difference enabled multi-stage network for multiple degradation image restoration. *Knowledge-Based Systems*, 318:113426, 2025.

Yongsheng Yu, Ziyun Zeng, Hang Hua, Jianlong Fu, and Jiebo Luo. PromptFix: You prompt and we fix the photo. In *Advances in Neural Information Processing Systems*, volume 37, pp. 40000–40031, 2024.

Eduard Zamfir, Zongwei Wu, Nancy Mehta, Yuedong Tan, Danda Pani Paudel, Yulun Zhang, and Radu Timofte. Complexity experts are task-discriminative learners for any image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12753–12763, 2025.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14821–14831, 2021.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5728–5739, 2022.

Haijin Zeng, Xiangming Wang, Yongyong Chen, Jingyong Su, and Jie Liu. Vision-language gradient descent-driven all-in-one deep unfolding networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7524–7533, 2025.

Jinghao Zhang, Jie Huang, Mingde Yao, Zizheng Yang, Hu Yu, Man Zhou, and Feng Zhao. Ingredient-oriented multi-degradation learning for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5825–5835, 2023.

Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*, 26 (7):3142–3155, 2017.

Kai Zhang, Wangmeng Zuo, and Lei Zhang. FFDNet: Toward a fast and flexible solution for CNN-based image denoising. *IEEE Transactions on Image Processing*, 27(9):4608–4622, 2018.

Xu Zhang, Jiaqi Ma, Guoli Wang, Qian Zhang, Huan Zhang, and Lefei Zhang. Perceive-IR: Learning to perceive degradation better for all-in-one image restoration. *IEEE Transactions on Image Processing*, 35:2018–2033, 2026.

Shihao Zhou, Duosheng Chen, Jinshan Pan, Jinglei Shi, and Jufeng Yang. Adapt or perish: Adaptive sparse transformer with attentive feature refinement for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2952–2963, 2024.

Shihao Zhou, Jinshan Pan, and Jufeng Yang. Learning an adaptive sparse transformer for efficient image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(11): 10344–10360, 2025.

Yurui Zhu, Tianyu Wang, Xueyang Fu, Xuanyu Yang, Xin Guo, Jifeng Dai, Yu Qiao, and Xiaowei Hu. Learning weather-general and weather-specific features for image restoration under multiple adverse weather conditions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21747–21758, 2023.

A Detailed implementation

A.1 Architecture configuration

Table 5 summarizes the stage-wise configuration of the default InCaRP model. The four-level encoder progressively increases the channel width and number of attention heads to capture hierarchical representations, while retaining a relation rank of $r = 32$ in every stage. The decoder reconstructs the image through three CAPC-RSTB stages, placing prompt calibration only in the reconstruction pathway, and four full-resolution RSTBs further refine the output without additional calibration.

Table 5: Configuration of the default InCaRP model.

Stage	Channels	Heads	Blocks	Rank	Prompt calibration
Encoder 1	48	1	4	32	–
Encoder 2	96	2	6	32	–
Encoder 3	192	4	6	32	–
Latent	384	8	8	32	–
Decoder 3	192	4	4	32	✓
Decoder 2	96	2	4	32	✓
Decoder 1	96	1	2	32	✓
Refinement	96	1	4	32	–

A.2 Training objective

For the restored image $\widehat{\mathbf{J}}$ and clean target $\mathbf{J}$, we combine an L_1 reconstruction loss with an FFT loss:

$$\mathcal{L} = \underbrace{\left\|\widehat{\mathbf{J}} - \mathbf{J}\right\|_1}_{\mathcal{L}_1} + 0.1 \underbrace{\left\|\mathcal{F}(\widehat{\mathbf{J}}) - \mathcal{F}(\mathbf{J})\right\|_1}_{\mathcal{L}_{\mathrm{FFT}}}, \tag{16}$$

where $\mathcal{F}$ denotes the two-dimensional Fourier transform represented by its real and imaginary components.

B Additional experimental results

B.1 CDD-11 compositional restoration

CDD-11 (Guo et al., 2024c) evaluates whether one model can restore four single degradations together with five two-degradation and two three-degradation compositions formed from low-light, haze, rain, and snow. We jointly train one model on all 11 degradation conditions for 200 epochs. All remaining implementation settings follow those used for AIO-3 and AIO-5 in Section 4.1. Table 6 compares our method with AirNet (Li et al., 2022), PromptIR (Potlapalli et al., 2023), WGWSNet (Zhu et al., 2023), WeatherDiff (Özdenizci & Legenstein, 2023), OneRestore (Guo et al., 2024c), and MoCE-IR-S (Zamfir et al., 2025). InCaRP attains an overall average PSNR/SSIM of 30.13 dB/0.891. Compared with MoCE-IR-S, the strongest remaining baseline, the proposed approach improves the average PSNR/SSIM by 1.08 dB/0.010, with the largest PSNR gain of 2.28 dB obtained on haze removal.

Table 6: Quantitative comparison on the CDD-11 compositional-degradation benchmark. L, H, R, and S denote low-light, haze, rain, and snow, respectively. Best results are shown in bold.

Method	Venue	Single Degradations				Double Degradations					Triple Degradations		Average
		L	H	R	S	L+H	L+R	L+S	H+R	H+S	L+H+R	L+H+S	P/S
AirNet	CVPR'2022	24.83/0.778	24.21/0.951	26.55/0.891	26.79/0.919	23.23/0.779	22.82/0.710	23.29/0.723	22.21/0.868	23.29/0.901	21.80/0.708	22.24/0.725	23.75/0.814
PromptIR	NeurIPS'2023	26.32/0.805	26.10/0.969	31.56/0.946	31.53/0.960	24.49/0.789	25.05/0.771	24.51/0.761	24.54/0.924	23.70/0.925	23.74/0.752	23.33/0.747	25.90/0.850
WGWSNet	CVPR'2023	24.39/0.774	27.90/0.982	33.15/0.964	34.43/0.973	24.27/0.800	25.06/0.772	24.60/0.765	27.23/0.955	27.65/0.960	23.90/0.772	23.97/0.771	26.96/0.863
WeatherDiff	TPAMI'2023	23.58/0.763	21.99/0.904	24.85/0.885	24.80/0.888	21.83/0.756	22.69/0.730	22.12/0.707	21.25/0.868	21.99/0.868	21.23/0.716	21.04/0.698	22.49/0.799
OneRestore	ECCV'2024	26.48/0.826	32.52/0.990	33.40/0.964	34.31/0.973	25.79/0.822	25.58/0.799	25.19/0.789	29.99/0.957	30.21/0.964	24.78/0.788	24.90/0.791	28.47/0.878
MoCE-IR-S	CVPR'2025	27.26/0.824	32.66/0.990	34.31/0.970	35.91/**0.980**	26.24/0.817	26.25/0.800	26.04/0.793	29.93/0.964	30.19/0.970	25.41/0.789	25.39/0.790	29.05/0.881
InCaRP (ours)	–	**27.68/0.837**	**34.94/0.993**	**35.17/0.972**	**36.79/0.980**	**27.02/0.834**	**26.83/0.816**	**26.76/0.810**	**31.83/0.969**	**31.98/0.973**	**26.19/0.807**	**26.26/0.807**	**30.13/0.891**

Table 7: Pre-correlation relation sparsification on AIO-3. Widths refer to each attention head. Top-r selects r entries after forming the full relation matrix, whereas LRSA projects K/V responses directly into r carriers before correlation. Parameters report model size.

Variant	Q width	K/V width	Score size	Selection	SOTS	Rain100L	Denoise	Avg. P/S	Params. (M)
Full transposed attention	48	48	48×48	None	29.71/0.976	37.47/0.978	31.09/0.871	32.09/0.913	21.25
Post-hoc Top-32	48	48	48×48	Top-32	29.52/0.972	38.15/0.981	31.17/0.875	32.23/0.916	21.25
Symmetric bottleneck	32	32	32×32	None	28.81/0.967	37.53/0.978	31.07/0.873	31.91/0.913	20.09
LRSA, $r = 16$	48	16	48×16	None	30.96/0.980	38.95/0.984	31.24/0.876	32.73/0.918	21.09
LRSA, $r = 24$	48	24	48×24	None	31.72/0.982	39.08/0.985	31.37/0.879	32.98/0.921	21.93
LRSA, $r = 32$	48	32	48×32	None	**32.03/0.983**	**39.12/0.985**	**31.39/0.880**	**33.06/0.922**	22.77
LRSA, $r = 48$	48	48	48×48	None	31.89/0.982	39.08/0.985	31.36/0.879	33.01/0.921	24.44

C Additional ablations and analyses

Unless otherwise stated, every trainable ablation is optimized from scratch using the complete AIO-3 training protocol, with the same data, augmentation, loss, optimizer, learning-rate schedule, number of training epochs, and evaluation code. SOTS, Rain100L, and Denoise denote dehazing, deraining, and the mean over BSD68 at $\sigma \in \{15, 25, 50\}$, respectively. Avg. P/S averages PSNR and SSIM over the five AIO-3 test conditions.

C.1 Pre-correlation relation sparsification

The component study in the main paper compares exhaustive transposed attention with the proposed pre-correlation sparsification. Table 7 adds two structural controls and varies the relation rank within a single experiment. The symmetric bottleneck reduces the query, key, and value widths to r, thereby compressing the complete representation pathway. Post-hoc Top-r attention retains the full $d \times d$ score matrix but preserves only the largest r entries in each query row. In contrast, LRSA preserves the d-dimensional query and output pathways while directly projecting the K/V responses into r input-dependent carriers before correlation. The table therefore distinguishes structured sparsity over the response basis from generic width reduction and post-hoc entry selection. Both controls underperform LRSA, while $r = 32$ gives the best accuracy–capacity balance.

C.2 Retrieval dynamics across decoder depth

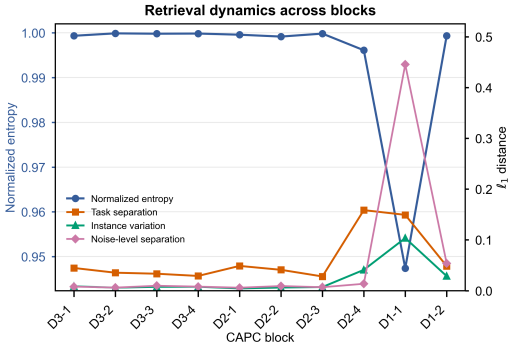


Figure 5: Shared prompt-pair retrieval across decoder blocks. Normalized entropy quantifies mixture concentration, while task separation, instance variation, and noise-level separation measure ℓ_1 distances among task centroids, individual samples, and noise-level centroids, respectively.

Figure 5 tracks the shared query–relation prompt-pair retrieval throughout the decoder. Coarse and intermediate blocks retain high-entropy mixtures with limited separation, consistent with broadly shared prompt-pair support. Condition-dependent behavior becomes more pronounced at selected late blocks, particularly D2-4 and D1-1. D1-1 reaches the lowest normalized entropy of 0.947 and the largest noise-level separation of 0.445, whereas the recovery at D1-2 shows that retrieval specialization does not increase monotonically with depth. These dynamics indicate that concentrated conditional

Table 8: Sensitivity to calibration-factor dimension n on AIO-3. The prompt-bank size and relation rank remain $M = 8$ and $r = 32$.

n	SOTS	Rain100L	Denoise	Avg. P/S	Params.
1	31.22/0.982	38.88/0.984	31.32/0.878	32.81/0.920	21.38
2	31.85/0.981	39.08/0.985	31.29/0.876	32.96/0.919	21.84
4	**32.03/0.983**	**39.12/0.985**	**31.39/0.880**	**33.06/0.922**	22.77
8	31.89/0.983	39.12/0.985	31.36/0.879	33.02/0.921	24.61

Table 9: Shared versus independent prompt-pair retrieval on AIO-3.

Retrieval variant	Avg. P/S
Independent retrieval for each bank	32.73/0.918
Shared hyperspherical retrieval (ours)	32.81/0.919

composition is localized to particular high-resolution blocks rather than repeated uniformly throughout the decoder.

C.3 Calibration-factor dimension

The factor dimension n controls the rank upper bound of the prompt-generated scale and shift fields without changing the relation rank r. Table 8 varies $n \in \{1, 2, 4, 8\}$ while fixing $M = 8$, $r = 32$, and all remaining settings. Increasing n from 1 to 4 raises the average from 32.81 dB/0.920 to 33.06 dB/0.922, showing that a one-dimensional factor is too restrictive to express diverse query–atom adjustments. The largest improvement occurs on SOTS ($+0.81$ dB), whereas denoising changes by only 0.07 dB, indicating that the required calibration capacity varies across degradation structures. Expanding n to 8 adds 1.84M parameters over $n = 4$ but slightly reduces the average, which suggests that the factorization has already saturated and additional degrees of freedom introduce redundancy rather than useful calibration. We therefore use $n = 4$ as the accuracy–capacity balance.

C.4 Shared prompt-pair retrieval

Table 9 compares shared and independent retrieval while retaining distinct query and relation banks, hyperspherical matching, and the same normalized affine calibration. Independent retrieval composes the two prompts with separate weight distributions, which allows their mixture coordinates to describe different input conditions. Sharing the weights instead aligns both axes to one condition coordinate while the separate banks retain their specialized contents. This coordination improves the average from 32.73 dB/0.918 to 32.81 dB/0.919. Because bank capacity and calibration remain unchanged, the gain can be attributed to query–relation alignment rather than additional parameters. Together with the normalization ablation in Table 3, the result supports shared hyperspherical retrieval as a coordination mechanism for paired prompts.

D Limitations and Analysis

D.1 Model-design analysis

The two components of InCaRP address different aspects of all-in-one restoration. LRSA preserves the full query and output pathways while projecting redundant K/V responses into a compact carrier set, thereby realizing structured sparsity before correlation instead of selecting entries from an already formed dense map. CAPC then adapts this carrier through paired query-axis and relation-axis prompts. Sharing the retrieval weights aligns the two prompts to the same input condition, whereas their separate banks and projections preserve their distinct functional roles. This division is consistent with the component ablation: LRSA alone improves the baseline from 32.09 dB/0.913 to 32.38 dB/0.913, and coupling it with CAPC and hyperspherical retrieval further raises the result to 33.06 dB/0.922. The prompt ablations provide complementary evidence, since paired prompts outperform no-bank, single-axis, and shared-prompt alternatives.

D.2 Result analysis

The quantitative results reveal two useful characteristics. First, the gains extend across benchmark complexity: the complete model reaches 33.06 dB/0.922 on AIO-3, 31.31 dB/0.926 on AIO-5, and 30.13 dB/0.891 on CDD-11. Second, the improvement is not produced by sacrificing one task for another. On AIO-3, InCaRP obtains the best or tied-best value for every reported metric, with particularly clear gains on dehazing and deraining. On AIO-5, it improves the strongest competing average from 30.72 dB/0.919 to 31.31 dB/0.926 and remains competitive on tasks where another method attains the highest individual PSNR. Together with the qualitative comparisons and error maps, these results indicate that compact relation construction and coupled-axis calibration yield a more balanced restoration profile while using only 22.7M parameters.

D.3 Limitations and future directions

The current design uses a fixed relation rank across inputs and network stages, although the capacity required to represent useful interactions may vary with degradation type, severity, and feature resolution. Moreover, the retrieval query is derived from global channel statistics. This compact descriptor is effective for overall degradation conditioning but may underrepresent strongly localized artifacts or complex spatial mixtures. The per-task gains on AIO-5 are also not uniform, suggesting that the benefit of relation compression and global prompt retrieval depends on the structure of each degradation. Future work could therefore investigate input- or stage-adaptive rank allocation and spatially varying prompt retrieval while retaining the compact relation carrier and coordinated query–relation calibration.