FREQUENCY-STRUCTURED RESPONSE CONSOLIDATION FOR LIFELONG PERSON RE-IDENTIFICATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Lifelong person re-identification (LReID) learns from new domains while retaining previously acquired knowledge, with catastrophic forgetting remaining its central challenge. Existing data replay methods raise privacy concerns by storing historical exemplars, while knowledge distillation methods face cumulative forgetting of undistilled knowledge. Historical style replay transfers previous domain styles to current images, producing paired current-style and historical-style views and thereby avoiding such privacy concerns. However, since domain styles only partially represent previous domains, only limited historical knowledge can be recovered at each stage. Subsequent domain learning may progressively weaken earlier-domain knowledge, leading to catastrophic forgetting. To address this problem, we propose Frequency-Structured Response Consolidation (FSRC), which combines structured frequency modulation with cross-style response alignment to promote knowledge reuse and consolidation. Specifically, Structured Frequency Modulation (SFM) is embedded into the model to selectively regulate frequency-dependent changes in feature responses induced by historical-style variations, allowing the model to adapt to evolving domain styles and facilitating cross-domain knowledge reuse. Meanwhile, we introduce Cross-Style Alignment (CSA) to align paired feature responses of the same pedestrian under current and historical styles, thereby constraining subsequent representation updates to reduce interference with previously acquired knowledge and support knowledge consolidation. Extensive experiments on the standard LReID benchmark under two training orders demonstrate that FSRC surpasses state-of-the-art methods in both seen and unseen domains.

1 INTRODUCTION

Person re-identification (ReID) aims to retrieve the same individual across non-overlapping camera views. However, real-world surveillance environments continually evolve as camera configurations, illumination conditions, scene backgrounds, and pedestrian appearances change. This dynamic setting motivates lifelong person re-identification (LReID) (Pu et al., 2021), where a model learns from incoming domains while preserving earlier-domain retrieval performance. Yet continual adaptation to new domains may substantially degrade performance on earlier domains, making catastrophic forgetting a central challenge in LReID (Li & Hoiem, 2017), as depicted in Fig. 1(a).

Existing LReID methods mainly mitigate catastrophic forgetting through data replay and knowledge distillation. Data replay stores historical exemplars and raises privacy concerns (Wu & Gong, 2021; Yu et al., 2023), as illustrated in Fig. 1(b). In knowledge distillation, undistilled historical knowledge remains vulnerable to cumulative forgetting (Xu et al., 2025; Zhao et al., 2026). To alleviate privacy concerns associated with storing real historical samples, existing style replay methods transfer previous-domain styles to current images, generating paired current- and historical-style views of the same source image (Xu et al., 2025; Chen et al., 2025). However, because previous domains are represented through style cues alone, these views provide only partial historical evidence, as depicted in Fig. 1(c), allowing only limited historical knowledge to be recovered at each stage. As the model learns subsequent domains, knowledge from earlier domains may still be progressively weakened, leading to catastrophic forgetting. More importantly, the paired views expose how historical styles alter the same pedestrian, yet they are mainly used as replay samples, leaving their cross-style relation underexplored. Existing frequency-domain studies show that domain

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.

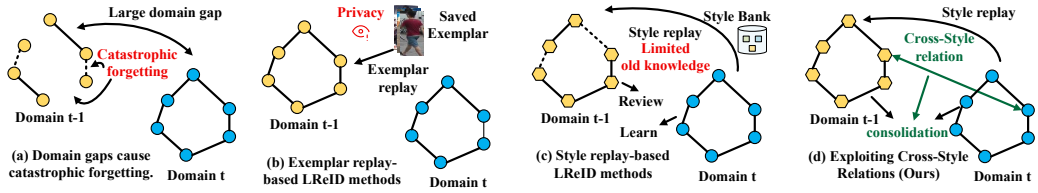


Figure 1: (a) Large domain gaps disrupt earlier-domain identity relations, causing catastrophic forgetting. (b) Exemplar replay stores historical samples, raising privacy concerns and incurring storage overhead. (c) Historical style replay rehearses earlier-domain styles but provides only partial historical evidence. (d) FSRC exploits same-identity cross-style relations to consolidate knowledge, thereby enhancing anti-forgetting capacity. Nodes denote identity representations; hexagons indicate historical-style representations, and dashed links indicate weakened relations.

variation affects spectral components nonuniformly (Xu et al., 2021; Huang et al., 2021) and that these components differ in cross-domain transferability (Lin et al., 2023). Together, these findings and the exact same-source pairing motivate two complementary cues for using limited historical evidence more fully: the frequency-dependent response changes induced by historical styles and the cross-style correspondence between paired views.

To exploit these cues, we propose Frequency-Structured Response Consolidation (FSRC) to promote knowledge reuse and consolidation across domain updates, as illustrated in Fig. 1(d). Specifically, Structured Frequency Modulation (SFM) is introduced to regulate the frequency-dependent response changes induced by historical styles. The learned regulation is retained in the continually updated model, facilitating adaptation to evolving domain styles and cross-domain knowledge reuse across domain updates. Meanwhile, Cross-Style Alignment (CSA) is designed to reduce discrepancies between the regulated responses of same-source current- and historical-style views. Subsequent representation updates are constrained by this paired supervision, thereby alleviating interference with previously acquired knowledge and supporting knowledge consolidation. Together, the two complementary mechanisms make fuller use of the limited historical evidence provided by style replay. The contributions of this work are summarized as follows:

- We identify the underexplored potential of same-source current- and historical-style pairs for knowledge consolidation from the limited evidence provided by style replay, and propose FSRC, a novel LReID framework.
- To facilitate adaptation and cross-domain knowledge reuse, SFM regulates frequency-dependent responses to historical styles and retains the learned regulation across sequential domain updates.
- CSA constrains representation updates under historical-style variations by aligning same-source current- and historical-style representations at both holistic and radial-band levels.

2 Related Work

2.1 Lifelong Person Re-Identification

Replay-based LReID. Replay-based methods mitigate forgetting by storing and replaying historical samples (Chen et al., 2022; Huang et al., 2022; Wu & Gong, 2021). For instance, exemplar replay has been combined with feature-space alignment to reduce cross-domain discrepancies (Ge et al., 2022). The old model is gradually updated with new-domain data during replay to refresh previous knowledge (Yu et al., 2023).

Distillation-based LReID. Distillation-based methods transfer previous knowledge by encouraging consistency between old and new models (Ge et al., 2022; Huang et al., 2023; Pu et al., 2023; Zhao et al., 2021). For instance, classification predictions can be aligned (Pu et al., 2021), while spatial and relational knowledge can be transferred at the patch level (Sun & Mu, 2022). Additionally, erroneous knowledge can be filtered and rectified before transfer (Xu et al., 2024c). Recent exemplar-free LReID methods further preserve knowledge through adaptive model consolidation or by rehearsing

previous-domain distributions and styles, without retaining real historical samples (Li et al., 2024; Xu et al., 2025; Chen et al., 2025).

2.2 Frequency-Domain Learning

Frequency-domain learning decomposes visual signals into spectral components to characterize model behavior and cross-domain variation. Prior studies show that different frequency components exhibit distinct generalization and robustness properties (Wang et al., 2020; Chen et al., 2021). In domain adaptation and generalization, Fourier amplitude transfer and frequency randomization have been used to manipulate domain variation (Yang & Soatto, 2020; Xu et al., 2021; Huang et al., 2021), while frequency filtering suppresses domain-sensitive components (Lin et al., 2023). Frequency-domain modeling has also been explored in ReID, where spectral decomposition and filtering facilitate generalizable representation learning (Peng et al., 2024). Recent work further introduces frequency-domain representations into rehearsal-based continual learning to improve training and storage efficiency (Liu et al., 2024).

Although these methods address knowledge preservation and domain variation from complementary perspectives, existing LReID methods based on historical style replay retain only style information from previous domains and therefore provide only partial historical evidence. Existing frequency-domain methods mainly exploit spectral information for robustness, cross-domain generalization, filtering, or generic continual learning, while the potential of frequency information for knowledge consolidation across sequential domain updates remains underexplored. Addressing this gap requires a framework that exploits frequency-dependent responses and same-source cross-style relations to make fuller use of this evidence for knowledge reuse and consolidation.

3 Method

3.1 Framework Overview

We consider exemplar-free lifelong person re-identification, where a model continually learns from a sequence of ReID domains $\mathcal{D} = \{\mathcal{D}_t\}_{t=1}^{T}$, where T is the number of sequential domains and $t \in \{1, \ldots, T\}$ indexes the learning stage. Each domain is denoted by $\mathcal{D}_t = \{(x_i^t, y_i^t)\}_{i=1}^{n_t}$, where $i \in \{1, \ldots, n_t\}$ indexes the training images, and x_i^t, y_i^t, and n_t denote the i-th image, its identity label, and the number of training images in domain t, respectively. The identity sets of different domains are mutually disjoint. At training step t, only $\mathcal{D}_t$ is accessible, and the preceding model $\mathcal{M}_{t-1}$ is updated without storing images from earlier domains. The resulting model is expected to retain identity knowledge on seen domains and generalize to unseen domains.

Following ADSR (Chen et al., 2025), retained historical style statistics are transferred to current-domain images without accessing old exemplars. Each original image x_i^o is paired with a historical-style view x_i^r, where superscripts o and r denote the original and historical-style views, respectively. Despite appearance differences, the paired views share the same identity label and provide cross-style supervision.

As illustrated in Fig. 2, Frequency-Structured Response Consolidation (FSRC) comprises Structured Frequency Modulation (SFM) and Cross-Style Alignment (CSA). SFM performs sample-conditioned spectral modulation at an intermediate backbone stage, while CSA constrains paired post-SFM representations at holistic and radial-band levels.

3.2 Structured Frequency Modulation

Historical-style variations induce frequency-dependent response changes that differ across samples, motivating sample-conditioned regulation for adaptation to evolving domain styles and cross-domain knowledge reuse. SFM derives a dense mask from each intermediate spectrum to selectively attenuate or amplify spectral responses and injects the resulting change through a residual pathway.

For sample i under view $v \in \{o, r\}$, let $X_{i,\ell}^v = E_{\leq \ell}(x_i^v) \in \mathbb{R}^{C \times H \times W}$ denote its feature map at backbone stage ℓ, where $E_{\leq \ell}$ is the backbone up to stage ℓ, and C, H, and W denote the channel count, height, and width, respectively. A lightweight predictor h_θ receives the real and imaginary spectral components and generates a sample-specific mask shared across channels and varying

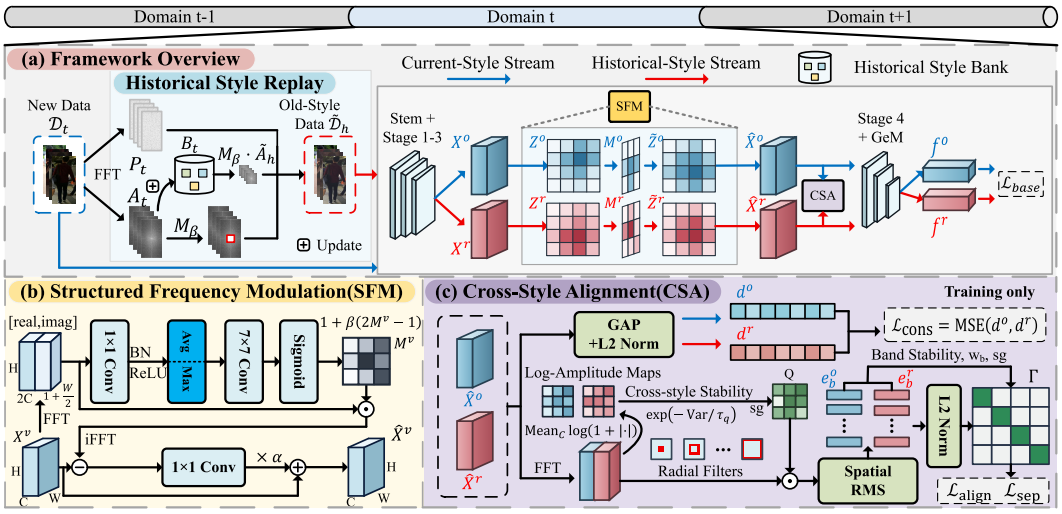


Figure 2: Overall framework of the proposed Frequency-Structured Response Consolidation (FSRC). (a) Framework overview: Historical Style Replay constructs paired original and historical-style views from the current domain, which are processed by a shared backbone equipped with SFM; CSA supplies training-only cross-style supervision. (b) Structured Frequency Modulation: SFM predicts a sample-adaptive mask in the frequency domain and injects the inverse-transformed modulation as a residual correction. (c) Cross-Style Alignment: CSA constrains paired post-SFM representations through holistic consistency and stability-weighted radial-band alignment and separation objectives.

across spectral locations:

$$M^v_{i,\ell} = h_\theta\big(\big[\mathrm{Re}\big(\mathcal{F}_r(X^v_{i,\ell})\big), \mathrm{Im}\big(\mathcal{F}_r(X^v_{i,\ell})\big)\big]\big), \tag{1}$$

where $M^v_{i,\ell}$ is the predicted modulation mask, h_θ is the mask predictor parameterized by θ, and $\mathcal{F}_r$ is the two-dimensional real Fourier transform. The operators $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ extract the real and imaginary components, respectively, and $[\cdot,\cdot]$ denotes channel-wise concatenation. The predictor maps the two spectral components to 128 channels using a 1×1 Conv–BN–ReLU block, aggregates channel-wise average and maximum responses, and applies a 7×7 sigmoid head.

Let $Z^v_{i,\ell} = \mathcal{F}_r(X^v_{i,\ell})$. We center the modulation around the identity mapping so that the mask can express both attenuation and amplification:

$$\widetilde{Z}^v_{i,\ell} = \big[1 + \beta\,\big(2M^v_{i,\ell} - 1\big)\big] \odot Z^v_{i,\ell}, \tag{2}$$

where $\widetilde{Z}^v_{i,\ell}$ is the modulated spectrum, β controls the modulation range, and $\odot$ denotes element-wise multiplication. Mask values below and above 0.5 attenuate and amplify the corresponding spectral responses, respectively.

The modulated spectrum is returned to the spatial domain and injected as a residual correction:

$$\widehat{X}^v_{i,\ell} = X^v_{i,\ell} + \alpha\,\psi\Big(\mathcal{F}_r^{-1}(\widetilde{Z}^v_{i,\ell}) - X^v_{i,\ell}\Big), \tag{3}$$

where $\widehat{X}^v_{i,\ell}$ is the post-SFM feature map, $\mathcal{F}_r^{-1}$ is the inverse real Fourier transform, ψ is a zero-initialized 1×1 convolution, and α controls the contribution of the projected spectral correction. The augmented backbone thus starts from its original representation and progressively learns sample-adaptive spectral corrections. The resulting map is consumed by the remaining backbone and exposed to CSA during training.

3.3 Cross-Style Alignment

To preserve identity responses across historical-style variations, subsequent representation updates are constrained through CSA. Specifically, the constraint is imposed at two complementary levels:

holistic identity responses are anchored, while stability-weighted radial-band alignment is applied across paired original and historical-style views.

Holistic cross-style consistency. We use the normalized global descriptor of each post-SFM map to anchor the overall identity response across styles:

$$d^v_{i,\ell} = \frac{\mathrm{GAP}(\widehat{X}^v_{i,\ell})}{\|\,\mathrm{GAP}(\widehat{X}^v_{i,\ell})\|_2}, \qquad \mathcal{L}_{\mathrm{cons}} = \frac{1}{NC}\sum_{i=1}^{N}\left\|d^o_{i,\ell} - d^r_{i,\ell}\right\|_2^2, \tag{4}$$

where $d^v_{i,\ell} \in \mathbb{R}^C$ is the normalized global descriptor, $\mathrm{GAP}(\cdot)$ denotes global average pooling, $\|\cdot\|_2$ denotes the Euclidean norm, N is the batch size, and $\mathcal{L}_{\mathrm{cons}}$ is the holistic cross-style consistency loss. Because spatial averaging cannot characterize how responses are distributed across spectral bands, we complement this holistic anchor with a band-level representation that retains nonconstant frequency responses.

Spectral stability. Cross-style variation at each spectral location provides a direct cue for identifying stable responses. Let $\widehat{Z}^v_{i,\ell} = \mathcal{F}_r(\widehat{X}^v_{i,\ell})$ denote the post-SFM spectrum. Its channel-averaged log-amplitude map is defined as $a^v_{i,\ell} = C^{-1}\sum_{c=1}^{C}\log(1 + |\widehat{Z}^v_{i,\ell,c}|)$, where $c \in \{1,\dots,C\}$ indexes the feature channels, $|\cdot|$ denotes the complex magnitude, and $\log$ is the natural logarithm. Its agreement across the paired views yields a pointwise stability map:

$$Q_{i,\ell} = \mathrm{sg}\left[\exp\left(-\tau_q^{-1}\,\mathrm{Var}_{v\in\{o,r\}}\left(a^v_{i,\ell}\right)\right)\right], \tag{5}$$

where $Q_{i,\ell}$ is the pointwise spectral-stability map, $\mathrm{Var}_{v\in\{o,r\}}(\cdot)$ denotes population variance over the paired views, $\mathrm{sg}[\cdot]$ denotes the stop-gradient operator, $\exp(\cdot)$ is the exponential function, and τ_q is the pointwise-stability temperature. Accordingly, $Q_{i,\ell}$ gives greater weight to locations with more consistent cross-style responses in subsequent band descriptors.

The normalized radial-frequency range is uniformly partitioned into B disjoint bands represented by binary masks $\{R_b\}_{b=1}^{B}$, where $b \in \{1,\dots,B\}$ indexes the radial bands and R_b selects band b. For each channel, the representation of a band is defined by the Spatial-RMS energy of its stability-weighted, band-limited response:

$$e^v_{i,\ell,b,c} = \frac{1}{\sqrt{HW}}\left\|\mathcal{F}_r^{-1}\Big((Q_{i,\ell}\odot R_b)\odot \widehat{Z}^v_{i,\ell,c}\Big)\right\|_2, \tag{6}$$

where $e^v_{i,\ell,b,c}$ is the Spatial-RMS energy of channel c in radial band b, and $1/\sqrt{HW}$ converts the spatial Euclidean energy into Spatial-RMS energy. The vector $e^v_{i,\ell,b} \in \mathbb{R}^C$ collects the channel energies of the complete band-limited response. Spatial-RMS retains nonconstant responses whose spatial mean vanishes, allowing different radial bands to carry information beyond the holistic GAP descriptor.

The stage index ℓ is fixed throughout CSA and is omitted below. We also estimate the stability of each sample–band pair from the cross-style variation of its energy descriptor:

$$\nu_{i,b} = \frac{1}{C}\sum_{c=1}^{C}\mathrm{Var}_{v\in\{o,r\}}\left(e^v_{i,b,c}\right), \qquad w_{i,b} = \mathrm{sg}\left[\exp\left(-\frac{\nu_{i,b}}{\tau_w}\right)\right], \tag{7}$$

where $\nu_{i,b}$ is the channel-averaged cross-style variance of the energy descriptor, $w_{i,b}$ is the resulting sample–band stability weight, and τ_w is the sample–band stability temperature. Unlike the location map Q, which constructs the representation, $w_{i,b}$ acts at the objective level by controlling how strongly a sample–band pair contributes to cross-style alignment.

Stability-weighted band alignment. Each band descriptor is first normalized as $\bar{e}^v_{i,b} = e^v_{i,b}/\|e^v_{i,b}\|_2$, after which a weighted cross-style band-similarity matrix Γ is constructed. Its (b,k)-th entry is defined as

$$\Gamma_{b,k} = \frac{\sum_{i=1}^{N}\omega_{i,b,k}\left\langle \bar{e}^o_{i,b}, \bar{e}^r_{i,k}\right\rangle}{\sum_{i=1}^{N}\omega_{i,b,k}}, \qquad \omega_{i,b,k} = \frac{w_{i,b} + w_{i,k}}{2}, \tag{8}$$

where $b, k \in \{1, \ldots, B\}$ index the original-view row band and historical-style column band, respectively, $\omega_{i,b,k}$ is their mean stability weight for sample i, and $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product. The diagonal entries quantify cross-style similarity between corresponding bands, whereas the off-diagonal entries quantify similarity between the channel-energy profiles of different bands. Accordingly, we define the alignment and separation objectives as

$$\mathcal{L}_{\text{align}} = \frac{1}{B}\sum_{b=1}^{B}(1-\Gamma_{b,b})^2, \qquad \mathcal{L}_{\text{sep}} = \frac{1}{B(B-1)}\sum_{b\neq k}\Gamma_{b,k}^2, \tag{9}$$
$$\mathcal{L}_{\text{csa}} = \lambda_c\mathcal{L}_{\text{cons}} + \lambda_a\mathcal{L}_{\text{align}} + \lambda_s\mathcal{L}_{\text{sep}},$$

where $\mathcal{L}_{\text{align}}$, $\mathcal{L}_{\text{sep}}$, and $\mathcal{L}_{\text{csa}}$ denote the corresponding-band alignment loss, cross-band separation loss, and total CSA loss, respectively. The summation $\sum_{b\neq k}$ ranges over all ordered pairs of distinct bands, and λ_c, λ_a, and λ_s weight the holistic-consistency, band-alignment, and band-separation terms, respectively. The corresponding-band term aligns stable spectral evidence shared by the paired views, and the weak cross-band term encourages a diagonally dominant band-similarity structure. Gradients propagate through the band descriptors and the SFM pathway. Detaching the stability estimators prevents a trivial loss reduction through shrinking the stability weights.

3.4 Joint Optimization and Inference

We retain the cross-entropy loss, triplet loss, and similarity-preserving knowledge distillation loss of ADSR (Chen et al., 2025), collectively denoted by $\mathcal{L}_{\text{base}}$. CSA adds paired intermediate supervision:

$$\mathcal{L} = \mathcal{L}_{\text{base}} + \mathcal{L}_{\text{csa}}, \tag{10}$$

where $\mathcal{L}$ is the complete training objective. The holistic term is applied whenever paired original and historical-style views are available. At inference, historical-style view generation and CSA are discarded; the backbone with SFM directly extracts the retrieval representation.

4 Experiments

4.1 Experimental Settings

Datasets. All experiments are conducted on the standard LReID benchmark (Pu et al., 2021). Market-1501 (Zheng et al., 2015), CUHK-SYSU (Xiao et al., 2017), DukeMTMC-reID (Ristani et al., 2016), MSMT17 (Wei et al., 2018), and CUHK03 (Li et al., 2014) are used as the five seen domains for anti-forgetting evaluation, while CUHK01 (Li et al., 2013), CUHK02 (Li & Wang, 2013), GRID (Loy et al., 2010), SenseReID (Zhao et al., 2017), VIPeR (Gray & Tao, 2008), i-LIDS (Home Office Scientific Development Branch, 2006), and PRID (Hirzer et al., 2011) form the seven unseen domains for generalization evaluation. Following the standard protocol, 500 training identities from each seen domain are used, and two training orders are adopted. Order-1 is Market→SYSU→Duke→MSMT→CUHK03, while Order-2 is Duke→MSMT→Market→SYSU→CUHK03.

Evaluation metrics. Mean average precision (mAP) and Rank-1 accuracy (R@1) are reported. Seen-Avg and Unseen-Avg denote the unweighted averages over the five seen and seven unseen domains, respectively, measuring resistance to forgetting and cross-domain generalization. Unless stated otherwise, results are reported after the final stage.

Implementation details. The backbone is an ImageNet-pretrained ResNet-50 (He et al., 2016), and input images are resized to 256×128. Random horizontal flipping, padding followed by random cropping, and random erasing are used for augmentation. Each mini-batch contains 25 identities and four images per identity. The first domain is trained for 80 epochs and each subsequent domain for 60 epochs. SGD is used with a learning rate of 0.008, momentum of 0.9, and weight decay of 10^{-4}; the learning rate is multiplied by 0.1 at epoch 30. SFM is inserted at Layer3, and CSA operates on the corresponding post-SFM responses using four radial bands. The loss weights $(\lambda_c, \lambda_a, \lambda_s)$ are set to $(0.5, 0.02, 10^{-5})$, and the stability temperatures (τ_q, τ_w) to $(3\times10^{-4}, 10^{-4})$. At each stage, (α, β) are linearly scheduled from $(0.25, 0)$ to $(0.75, 0.75)$ during a 10-epoch warm-up, after

Table 1: Seen-domain anti-forgetting and unseen-domain generalization under Training Order-1.

Method	Market		SYSU		Duke		MSMT		CUHK03		Seen		Unseen	
	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1
Joint-Train	78.9	90.9	86.7	88.2	71.2	82.9	36.2	61.2	61.2	63.4	66.8	77.3	59.4	52.6
AKA	58.1	77.4	72.5	74.8	28.6	45.1	6.1	16.2	38.7	40.4	40.8	50.8	47.6	42.6
PatchKD	**68.5**	**85.7**	75.6	78.6	33.8	50.4	6.5	17.0	34.1	36.8	43.7	53.7	49.1	45.4
CKP	53.8	76.0	81.2	83.0	49.7	67.0	18.4	40.8	44.1	45.8	49.4	62.5	58.0	51.0
LSTKC	54.7	76.0	81.1	83.4	49.4	66.2	20.0	43.2	44.7	46.5	50.0	63.1	57.0	49.9
DKP	60.3	80.6	83.6	85.4	51.6	68.4	19.7	41.8	43.6	44.2	51.8	64.1	59.2	51.6
DASK	61.2	82.3	81.9	83.7	58.5	74.6	29.1	57.6	46.2	48.1	55.4	69.3	65.3	58.4
PKA	57.7	80.2	**85.0**	**86.9**	59.4	75.6	**31.0**	56.1	44.1	44.6	55.4	68.7	62.3	55.3
ADSR	64.0	83.1	83.2	85.0	58.2	74.6	30.8	**59.2**	**48.7**	**49.9**	57.0	70.4	67.4	60.7
FSRC	66.2	84.9	84.7	86.3	**61.2**	**76.6**	30.2	58.5	**48.7**	49.4	**58.2**	**71.2**	**68.3**	**62.2**

Table 2: Seen-domain anti-forgetting and unseen-domain generalization under Training Order-2.

Method	Duke		MSMT		Market		SYSU		CUHK03		Seen		Unseen	
	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1	mAP	R@1
Joint-Train	71.2	82.9	36.2	61.2	78.9	90.9	86.7	88.2	61.2	63.4	66.8	77.3	59.4	52.6
AKA	42.2	60.1	5.4	15.1	37.2	59.8	71.2	73.9	36.9	37.9	38.6	49.4	46.0	41.7
PatchKD	58.3	74.1	6.4	17.4	43.2	67.4	74.5	76.9	33.7	34.8	43.2	54.1	48.6	44.1
CKP	49.4	67.0	14.5	33.8	56.0	77.6	83.2	84.9	45.3	47.1	49.7	62.1	57.2	50.0
LSTKC	49.9	67.6	14.6	34.0	55.1	76.7	82.3	83.8	46.3	48.1	49.6	62.1	57.6	49.6
DKP	53.4	70.5	14.5	33.3	60.6	81.0	83.0	84.9	45.0	46.1	51.3	63.2	59.0	51.6
DASK	55.7	74.4	25.2	51.9	71.6	87.7	84.8	86.2	48.4	49.8	57.1	70.0	65.5	57.9
PKA	54.2	70.6	24.3	48.0	68.6	85.3	85.0	86.8	43.8	46.9	55.2	67.5	60.3	52.8
ADSR	57.4	74.1	26.3	53.6	72.2	**88.2**	85.6	87.0	50.0	50.7	58.3	70.7	67.3	59.6
FSRC	**60.8**	**76.1**	**26.4**	**54.2**	**72.5**	**88.2**	**85.8**	**87.1**	**50.5**	**51.3**	**59.2**	**71.4**	**68.1**	**61.7**

which the spectral CSA terms are enabled. Following ADSR (Chen et al., 2025), we fuse the current and preceding-stage model parameters after each of stages 2–5, assigning a weight of 0.5 to each model, and use the fused model for evaluation and subsequent updates. A matched reproduction of ADSR (Chen et al., 2025) without SFM or CSA serves as the Baseline for controlled analyses, using the same batch size and optimization settings as FSRC. All experiments are conducted in PyTorch on a single NVIDIA RTX 5090 GPU with a random seed of 0.

4.2 Comparison with State-of-the-Art Methods

We focus on comparisons with ResNet-50-based exemplar-free LReID methods, including AKA (Pu et al., 2021), PatchKD (Sun & Mu, 2022), CKP (Xu et al., 2024a), LSTKC (Xu et al., 2024c), DKP (Xu et al., 2024b), DASK (Xu et al., 2025), PKA (Lu et al., 2025), and ADSR (Chen et al., 2025). Joint-Train, which accesses all seen domains simultaneously, is included as an upper-bound reference. Results for the compared methods are drawn from published studies or reproduced with implementations released by the original authors. Tables 1 and 2 report the per-domain and average performance under Order-1 and Order-2, respectively. Across both tables, boldface and underlining respectively indicate the best and second-best results.

4.3 Seen-Domain Performance Evaluation

FSRC ranks first in Seen-Avg under both orders, reaching 58.2%/71.2% mAP/R@1 under Order-1 and 59.2%/71.4% under Order-2. It exceeds ADSR by 1.2%/0.8% and 0.9%/0.7%, respectively. Under Order-1, ADSR remains slightly stronger on MSMT17 and CUHK03, whereas FSRC performs better on Market-1501, CUHK-SYSU, and DukeMTMC-reID, yielding the best average. Results

Table 3: Ablation study on different components.

Baseline	SFM	CSA	Seen-Avg mAP	Seen-Avg R@1	Unseen-Avg mAP	Unseen-Avg R@1
✓			56.7	70.1	66.5	59.8
✓	✓		57.5	71.2	67.8	61.1
✓		✓	57.1	70.3	66.8	59.9
✓	✓	✓	**58.2**	**71.2**	**68.3**	**62.2**

Table 4: Effect of incorporating FSRC into ADSR and DASK.

Method	Seen-Avg mAP	Seen-Avg R@1	Unseen-Avg mAP	Unseen-Avg R@1
ADSR	56.7	70.1	66.5	59.8
ADSR + FSRC	**58.2**	**71.2**	**68.3**	**62.2**
DASK	56.5	70.1	65.7	58.9
DASK + FSRC	**57.7**	**71.2**	**67.1**	**60.2**

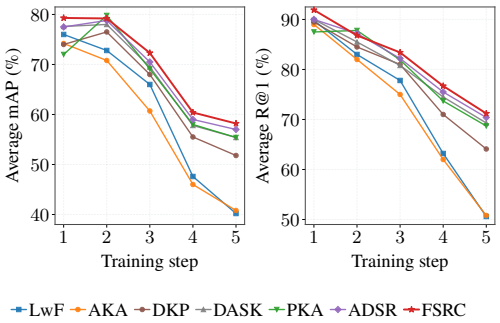


Figure 3: Anti-forgetting tendency on seen domains.

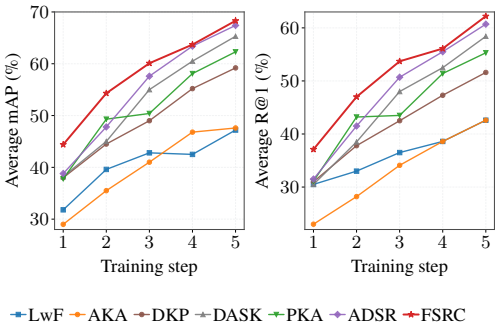


Figure 4: Generalization tendency on unseen domains.

across both orders demonstrate a better balance between knowledge acquisition and retention. Figure 3 further reports average mAP and R@1 over the domains encountered by each Order-1 stage. FSRC leads both metrics at four of the five stages and remains competitive at the second stage, supporting effective knowledge consolidation.

We additionally evaluate average forgetting (AF) to assess historical-domain retention. Relative to ADSR, FSRC reduces AF(mAP) by 0.40 and 1.20 points under Order-1 and Order-2, respectively, indicating less forgetting across sequential updates. Appendix B.2 provides the metric definition and complete results.

4.4 Unseen-Domain Generalization Evaluation

As shown in the Unseen-Avg columns of Tables 1 and 2, FSRC also achieves the best generalization performance among the compared methods under both orders, reaching 68.3%/62.2% mAP/R@1 under Order-1 and 68.1%/61.7% under Order-2, outperforming ADSR by 0.9%/1.5% and 0.8%/2.1%, respectively. Figure 4 further shows that FSRC leads in both mAP and R@1 throughout all five stages, demonstrating consistently stronger generalization during sequential learning.

4.5 Ablation Study

Effectiveness of SFM. Under Order-1, Table 3 presents ablation results for SFM and CSA on the ADSR-based Baseline. SFM improves Baseline by 0.8/1.1 and 1.3/1.3 mAP/R@1 points on seen and unseen domains, respectively. These gains confirm more effective cross-domain knowledge reuse.

Effectiveness of CSA. As an additional control without SFM, CSA is applied to the paired raw Layer3 responses. This control improves Baseline by 0.4%/0.2% and 0.3%/0.1% mAP/R@1 on seen and unseen domains, respectively. When applied to post-SFM responses, CSA further raises unseen-domain mAP/R@1 by 0.5%/1.1%. The full FSRC improves Baseline by 1.5%/1.1% on seen domains and 1.8%/2.4% on unseen domains, demonstrating that response alignment complements SFM in supporting knowledge consolidation across subsequent updates.

Table 5: Component-wise resource overhead.

	Baseline	+ SFM	+ CSA	FSRC
Params (M)	23.51	24.82	23.51	24.82
MACs (G)	4.05	4.21	4.05	4.21
Peak Mem. (GiB)	19.57	20.58	19.57	20.69
Train (s/batch)	0.268	0.287	0.273	0.291
Inf. (ms/img)	2.44	3.10	2.44	3.10

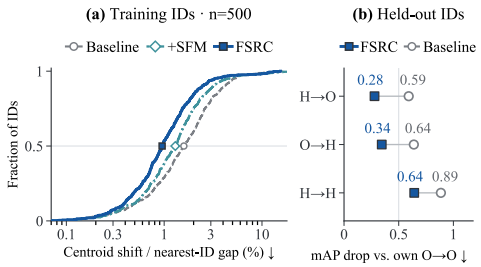


Figure 5: Cross-style consistency: (a) normalized centroid shifts on training identities; (b) cross-style mAP drops on held-out identities. Lower is better.

Applicability across style-replay mechanisms. Adding FSRC to ADSR and DASK improves all four final averages for both mechanisms, supporting its applicability across the two evaluated style-replay approaches. All entries in Table 4 are local reproductions. ADSR and DASK use batch sizes 100 and 128, respectively, while the paired variants within each replay mechanism share the same training protocol.

Visualization results. To better understand how FSRC improves cross-style consistency, Figure 5 visualizes the cross-style organization of training identities and the retrieval sensitivity of held-out identities at the final Order-1 CUHK03 stage. As shown in Figure 5(a), the median cross-style centroid shift normalized by nearest-identity separation decreases from 1.59% for Baseline to 1.30% with SFM and 0.95% for FSRC, with FSRC yielding a lower shift than Baseline for 433 of 500 identities. Figure 5(b) further shows smaller mAP drops for FSRC under $H \rightarrow O$, $O \rightarrow H$, and $H \rightarrow H$, where O and H denote original- and historical-style views, respectively. The drops are measured against each model's own $O \rightarrow O$ result. These results show that SFM and CSA progressively improve cross-style organization, while the full FSRC reduces sensitivity to historical-style interventions on held-out identities.

4.6 Efficiency Analysis

Table 5 reports component-wise computational cost under Order-1. Training latency is measured on cached real batches, including replay, after 10 warm-up iterations and averaged over three 30-iteration runs per stage, with data loading excluded. Relative to Baseline, FSRC increases parameters, MACs, peak allocated GPU memory, and training latency by 5.6%, 3.8%, 5.7%, and 8.7%, respectively. Batch-1 inference averages three 300-image runs after 100 warm-up images, increasing from 2.44 to 3.10 ms/image. CSA adds no inference-time operation; hence, Baseline/+CSA and +SFM/FSRC share their respective deployment graphs.

5 Conclusion

We propose Frequency-Structured Response Consolidation (FSRC) to promote knowledge consolidation from the limited historical evidence provided by style replay across sequential domain updates. FSRC consists of Structured Frequency Modulation (SFM) and Cross-Style Alignment (CSA). To facilitate adaptation to evolving domain styles, SFM is introduced to selectively regulate frequency-dependent response changes induced by historical-style variations, while CSA is designed to align paired original and historical-style responses, thereby reducing interference from subsequent domain updates. Together, these components make fuller use of frequency-dependent responses and same-source cross-style relations for knowledge reuse and consolidation. Experimental results demonstrate that FSRC achieves the best seen- and unseen-domain average performance among the compared ResNet-50-based exemplar-free LReID methods. In future work, we plan to extend the proposed FSRC framework to re-indexing-free LReID.

AI USE STATEMENT

No generative AI tools were used in conducting this research or preparing this manuscript.

REPRODUCIBILITY STATEMENT

Section 3 describes the proposed SFM and CSA modules, and Section 4.1 specifies the datasets, training orders, evaluation metrics, and implementation settings. The appendix provides further benchmark details, parameter analyses, and additional ablation studies. The supplementary code package includes implementations of SFM and CSA, integration examples, data preprocessing utilities, and retrieval metrics, together with environment requirements and a reproducibility guide.

REFERENCES

Guangyao Chen, Peixi Peng, Li Ma, Jia Li, Lin Du, and Yonghong Tian. Amplitude-phase recombination: Rethinking robustness of convolutional neural networks in frequency domain. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 448–457, 2021. doi: 10.1109/ICCV48922.2021.00051.

Hao Chen, Benoit Lagadec, and Francois Bremond. Unsupervised lifelong person re-identification via contrastive rehearsal. *arXiv preprint arXiv:2203.06468*, 2022. doi: 10.48550/arXiv.2203.06468.

Long Chen, De Cheng, Shizhou Zhang, Yinghui Xing, Di Xu, and Yanning Zhang. Amplitude-aware domain style replay for lifelong person re-identification. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 2026–2035, 2025. doi: 10.1145/3746027.3755616.

Wenhang Ge, Junlong Du, Ancong Wu, Yuqiao Xian, Ke Yan, Feiyue Huang, and Wei-Shi Zheng. Lifelong person re-identification by pseudo task knowledge preservation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 688–696, 2022. doi: 10.1609/aaai.v36i1.19949.

Douglas Gray and Hai Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *Proceedings of the European Conference on Computer Vision*, pp. 262–275, 2008.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.

Martin Hirzer, Csaba Beleznai, Peter M. Roth, and Horst Bischof. Person re-identification by descriptive and discriminative classification. In *Proceedings of the Scandinavian Conference on Image Analysis*, pp. 91–102, 2011.

Home Office Scientific Development Branch. Imagery library for intelligent detection systems (i-LIDS). In *2006 IET Conference on Crime and Security*, pp. 445–448, 2006.

Jiaxing Huang, Dayan Guan, Aoran Xiao, and Shijian Lu. FSDR: Frequency space domain randomization for domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6891–6902, 2021.

Jinze Huang, Xiaohan Yu, Dong An, Yaoguang Wei, Xiao Bai, Jin Zheng, Chen Wang, and Jun Zhou. Learning consistent region features for lifelong person re-identification. *Pattern Recognition*, 144:109837, 2023. doi: 10.1016/j.patcog.2023.109837.

Zhipeng Huang, Zhizheng Zhang, Cuiling Lan, Wenjun Zeng, Peng Chu, Quanzeng You, Jiang Wang, Zicheng Liu, and Zheng-Jun Zha. Lifelong unsupervised domain adaptive person re-identification with coordinated anti-forgetting and adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14268–14277, 2022. doi: 10.1109/CVPR52688.2022.01389.

Qiwei Li, Kunlun Xu, Yuxin Peng, and Jiahuan Zhou. Exemplar-free lifelong person re-identification via prompt-guided adaptive knowledge consolidation. *International Journal of Computer Vision*, 132(11):4850–4865, 2024. doi: 10.1007/s11263-024-02110-x.

Wei Li and Xiaogang Wang. Locally aligned feature transforms across views. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3594–3601, 2013.

Wei Li, Rui Zhao, and Xiaogang Wang. Human reidentification with transferred metric learning. In *Computer Vision – ACCV 2012*, pp. 31–44, 2013.

Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang. DeepReID: Deep filter pairing neural network for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 152–159, 2014.

Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(12):2935–2947, 2017.

Shiqi Lin, Zhizheng Zhang, Zhipeng Huang, Yan Lu, Cuiling Lan, Peng Chu, Quanzeng You, Jiang Wang, Zicheng Liu, Amey Parulkar, Viraj Navkal, and Zhibo Chen. Deep frequency filtering for domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11797–11807, 2023.

Ruiqi Liu, Boyu Diao, Libo Huang, Zijia An, Zhulin An, and Yongjun Xu. Continual learning in the frequency domain. In *Advances in Neural Information Processing Systems*, 2024.

Chen Change Loy, Tao Xiang, and Shaogang Gong. Time-delayed correlation analysis for multi-camera activity understanding. *International Journal of Computer Vision*, 90(1):106–129, 2010.

Zhijie Lu, Wuxuan Shi, He Li, and Mang Ye. Prototype-guided knowledge propagation with adaptive learning for lifelong person re-identification. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pp. 5851–5859, 2025. doi: 10.24963/ijcai.2025/651.

Jinjia Peng, Pengpeng Song, Hui Li, and Huibing Wang. ReFID: Reciprocal frequency-aware generalizable person re-identification via decomposition and filtering. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 20(7):1–20, 2024. doi: 10.1145/3643684.

Nan Pu, Wei Chen, Yu Liu, Erwin M. Bakker, and Michael S. Lew. Lifelong person re-identification via adaptive knowledge accumulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7901–7910, 2021.

Nan Pu, Zhun Zhong, Nicu Sebe, and Michael S. Lew. A memorizing and generalizing framework for lifelong person re-identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(11):13567–13585, 2023. doi: 10.1109/TPAMI.2023.3297058.

Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. In *Computer Vision – ECCV 2016 Workshops*, pp. 17–35, 2016.

Zhicheng Sun and Yadong Mu. Patch-based knowledge distillation for lifelong person re-identification. In *Proceedings of the 30th ACM International Conference on Multimedia*, pp. 696–707, 2022. doi: 10.1145/3503161.3548179.

Haohan Wang, Xindi Wu, Zeyi Huang, and Eric P. Xing. High-frequency component helps explain the generalization of convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8681–8691, 2020. doi: 10.1109/CVPR42600.2020.00871.

Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. Person transfer GAN to bridge domain gap for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 79–88, 2018.

Guile Wu and Shaogang Gong. Generalising without forgetting for lifelong person re-identification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 2889–2897, 2021. doi: 10.1609/aaai.v35i4.16395.

Tong Xiao, Shuang Li, Bochao Wang, Liang Lin, and Xiaogang Wang. Joint detection and identification feature learning for person search. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3415–3424, 2017.

Kunlun Xu, Haozhuo Zhang, Yu Li, Yuxin Peng, and Jiahuan Zhou. Mitigate catastrophic remembering via continual knowledge purification for noisy lifelong person re-identification. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 5790–5799, 2024a.

Kunlun Xu, Xu Zou, Yuxin Peng, and Jiahuan Zhou. Distribution-aware knowledge prototyping for non-exemplar lifelong person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16604–16613, 2024b.

Kunlun Xu, Xu Zou, and Jiahuan Zhou. LSTKC: Long short-term knowledge consolidation for lifelong person re-identification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 16202–16210, 2024c. doi: 10.1609/aaai.v38i14.29554.

Kunlun Xu, Chenghao Jiang, Peixi Xiong, Yuxin Peng, and Jiahuan Zhou. DASK: Distribution rehearsing via adaptive style kernel learning for exemplar-free lifelong person re-identification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 8915–8923, 2025. doi: 10.1609/aaai.v39i9.32964.

Qinwei Xu, Ruipeng Zhang, Ya Zhang, Yanfeng Wang, and Qi Tian. A fourier-based framework for domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14383–14392, 2021.

Yanchao Yang and Stefano Soatto. FDA: Fourier domain adaptation for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4085–4095, 2020.

Chunlin Yu, Ye Shi, Zimo Liu, Shenghua Gao, and Jingya Wang. Lifelong person re-identification via knowledge refreshing and consolidation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 3295–3303, 2023. doi: 10.1609/aaai.v37i3.25436.

Bo Zhao, Shixiang Tang, Dapeng Chen, Hakan Bilen, and Rui Zhao. Continual representation learning for biometric identification. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*, pp. 1197–1207, 2021. doi: 10.1109/WACV48630.2021.00124.

Haiyu Zhao, Maoqing Tian, Shuyang Sun, Jing Shao, Junjie Yan, Shuai Yi, Xiaogang Wang, and Xiaoou Tang. Spindle net: Person re-identification with human body region guided feature decomposition and fusion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1077–1085, 2017.

Jiaqi Zhao, Jie Luo, Yong Zhou, Wen-Liang Du, Xixi Li, and Rui Yao. Unified representation causal prompt distillation for re-inference-free lifelong person re-identification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 13144–13152, 2026. doi: 10.1609/aaai.v40i15.38315.

Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1116–1124, 2015.

APPENDIX

A OVERVIEW

This supplementary material provides the detailed LReID benchmark statistics, parameter analysis, forgetting and stage-wise evaluations, and controlled analyses of the proposed SFM and CSA. We also specify the protocol used for the person retrieval comparison on seen and unseen domains. These results extend the evidence in the main paper without repeating its principal comparisons.

B ADDITIONAL EXPERIMENTAL RESULTS

B.1 PARAMETER ANALYSIS

We vary one hyperparameter at a time on Order-1 and keep the remaining settings fixed. Figure 6 reports all five values considered for each parameter.

SFM modulation range $\beta_{\max}$. The parameter $\beta_{\max}$ controls the maximum response modulation applied by SFM. As shown in Figure 6(a), the seen-domain scores increase gradually as the modulation range grows, while the unseen-domain performance reaches its strongest overall result at $\beta_{\max} = 0.75$. A wider range increases SFM's capacity to adapt the spectral statistics of the observed domains; once the modulation becomes excessive, this adaptation grows too specific to the observed styles and transfers less reliably to unseen domains. We therefore use $\beta_{\max} = 0.75$ in all experiments.

CSA alignment weight λ_a. The parameter λ_a controls the strength of radial-band alignment. As shown in Figure 6(b), the seen-domain results remain stable across the tested range, while the unseen metrics are more sensitive to this weight. The setting $\lambda_a = 0.02$ achieves the highest unseen R@1 and remains close to the best unseen mAP. The weight balances the removal of cross-style band shifts against the preservation of identity-bearing differences between bands: weak alignment leaves style variation, whereas excessive alignment suppresses useful spectral contrast. We set $\lambda_a = 0.02$ by default.

CSA stability temperature τ_w. The temperature τ_w determines the sharpness of the sample-band stability weights. Figure 6(c) shows stable seen-domain performance throughout the tested range, together with a clearer change on unseen domains. An overly sharp distribution overlooks complementary bands, whereas an overly flat distribution restores style-sensitive bands; both weaken the cross-style evidence used for transfer and therefore cause clearer variations on unseen domains. The moderate setting $\tau_w = 10^{-4}$ gives the best overall seen–unseen balance and is used throughout the paper.

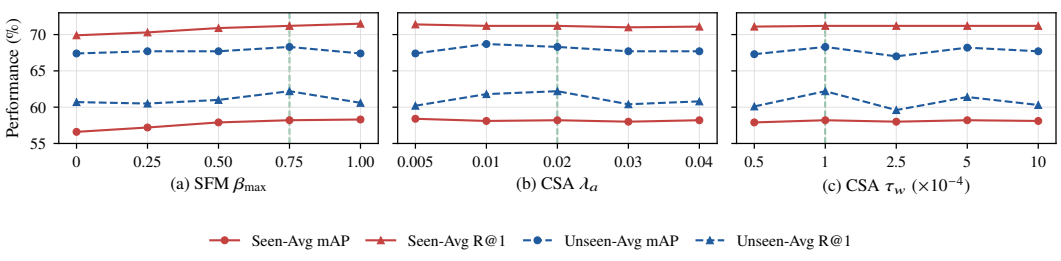


Figure 6: Parameter analysis on Order-1. Each panel varies one parameter while keeping the remaining settings fixed. Dashed vertical lines indicate the default values used in the paper.

B.2 Comparison on AF Performance

Let $A^m_{t,i}$ denote metric m on domain i after training stage t. For a sequence of T domains, we compute Average Forgetting as

$$\mathrm{AF}_m = \frac{1}{T-1}\sum_{i=1}^{T-1}\left(\max_{t\in\{i,\ldots,T\}} A^m_{t,i} - A^m_{T,i}\right). \tag{11}$$

A lower AF indicates that less previously acquired knowledge is forgotten. Table 6 compares ADSR and FSRC on both training orders.

Table 6: Average Forgetting on Order-1 and Order-2. Lower values indicate less forgetting.

Order	**Method**	**AF(mAP)↓**	**AF(R@1)↓**
O1	ADSR	4.55	2.38
	FSRC	**4.15**	**2.30**
O2	ADSR	4.68	4.15
	FSRC	**3.48**	**2.98**

FSRC reduces AF(mAP) by 0.40 and 1.20 points on Order-1 and Order-2, respectively. The corresponding AF(R@1) reductions are 0.08 and 1.18 points. The consistent reductions across both metrics show that the final model retains historical-domain performance more effectively, with a particularly clear gain on Order-2.

C Additional Ablation Studies

C.1 SFM Insertion Layer

We compare four SFM placements while keeping all remaining settings fixed. The insertion stage determines the information on which SFM estimates and applies its spectral correction. Earlier responses retain abundant local texture but provide weaker identity abstraction, making the modulation more sensitive to low-level appearance variations. Later responses are more identity-oriented but spatially coarser, leaving less frequency structure to regulate. Applying SFM at both Layer2 and Layer3 further increases correction capacity and improves fitting to the observed domains, yet the repeated modulation is less transferable to unseen styles. Layer3 provides the most effective compromise between spatial-frequency detail and identity semantics, which explains its strongest overall unseen performance in Table 7.

Table 7: Influence of the SFM insertion stage on Order-1.

Insertion Stage			**Seen-Avg**		**Unseen-Avg**	
L2	**L3**	**L4**	**mAP**	**R@1**	**mAP**	**R@1**
✓			57.7	71.0	67.8	61.8
	✓		**58.2**	71.2	**68.3**	**62.2**
✓	✓		**58.2**	**71.6**	67.7	60.9
		✓	56.3	69.3	66.7	59.7

C.2 Ablation on CSA Components

CSA constrains the same original–historical-style pair at two complementary levels. Holistic consistency matches the pooled identity response and therefore corrects an aggregate cross-style displacement, but it cannot describe how the response is redistributed across radial frequencies. The holistic-only row in Table 8 consequently improves the observed-domain representation without strengthening unseen retrieval. Adding radial-band CSA preserves the correspondence of frequency-localized evidence and produces the clear improvement on unseen domains, while changing the seen-domain averages only slightly.

Table 8: Ablation of CSA components on Order-1. Hol. and Rad. denote holistic and radial-band consistency; checks denote active components. All rows use SFM.

Components					Seen-Avg		Unseen-Avg	
Hol.	**Rad.**	Q	w	**Sep.**	**mAP**	**R@1**	**mAP**	**R@1**
					57.5	71.2	67.8	61.1
✓					58.3	71.3	67.8	60.8
✓	✓	✓	✓	✓	58.2	71.2	**68.3**	**62.2**
✓	✓		✓	✓	**58.4**	**71.5**	68.2	61.8
✓	✓	✓		✓	58.1	71.0	68.0	60.3
✓	✓	✓	✓		58.1	70.8	67.6	60.2

The remaining rows examine how radial-band evidence is selected and organized. The positionwise weight Q suppresses frequency locations that vary across the paired styles, while the sample-band weight w prevents unreliable bands from dominating the alignment of an individual sample. Removing either estimator admits less stable evidence, with the larger Rank-1 drop without w showing the importance of sample-specific band selection. Finally, alignment alone can make different bands respond too similarly; the weak separation term preserves a diagonally dominant correspondence between matched bands. Its removal causes the largest joint decline in unseen mAP and R@1. Together, these results explain why the complete CSA provides the strongest overall unseen result.

C.3 Inference Analysis of SFM

To isolate the inference contribution of SFM, we evaluate the same final Order-1 checkpoint with SFM enabled and disabled. As shown in Table 9, enabling SFM improves all four averages, with gains of $3.36/3.11$ points on seen domains and $3.45/4.26$ points on unseen domains in mAP/R@1. This same-checkpoint intervention confirms that the learned modulation remains active in the final retrieval path.

Table 9: Same-checkpoint inference analysis of SFM on Order-1.

Inference Path		Seen-Avg		Unseen-Avg	
Bypass	**Enabled**	**mAP**	**R@1**	**mAP**	**R@1**
✓		54.85	68.06	64.90	57.97
	✓	**58.21**	**71.17**	**68.34**	**62.23**

We further apply a fixed permutation to the frequency positions while preserving the learned modulation values. On a matched panel of six seen-domain and four unseen-domain evaluations, the native routing outperforms the permuted routing by $5.57/4.63$ points on the seen evaluations and $2.94/3.47$ points on the unseen evaluations in mAP/R@1. The degradation under position permutation shows that the gain depends on the learned association between modulation values and frequency positions.

D Details of the LReID Benchmark

We follow the standard 12-dataset LReID benchmark (Pu et al., 2021), comprising five seen and seven unseen datasets listed in Table 10. "Original identities" reports the official dataset partitions, whereas "LReID identities" reports the benchmark subsets. Each seen dataset contributes 500 training identities, and unseen datasets are used only for testing. CUHK-SYSU was originally introduced for person search. Following the LReID protocol, individual images are cropped using ground-truth person bounding boxes, and only identities containing at least four bounding boxes are retained.

Table 10: Identity statistics of the datasets used in the LReID benchmark. A dash indicates that a dataset is used only for testing.

Type	Dataset	Original Identities			LReID Identities		
		Train	Query	Gallery	Train	Query	Gallery
Seen	CUHK03 (Li et al., 2014)	767	700	700	500	700	700
	Market-1501 (Zheng et al., 2015)	751	750	751	500	751	751
	DukeMTMC-reID (Ristani et al., 2016)	702	702	1110	500	702	1110
	CUHK-SYSU (Xiao et al., 2017)	942	2900	2900	500	2900	2900
	MSMT17-V2 (Wei et al., 2018)	1041	3060	3060	500	3060	3060
Unseen	i-LIDS (Home Office Scientific Development Branch, 2006)	243	60	60	–	60	60
	VIPeR (Gray & Tao, 2008)	316	316	316	–	316	316
	GRID (Loy et al., 2010)	125	125	126	–	125	126
	PRID (Hirzer et al., 2011)	100	100	649	–	100	649
	CUHK01 (Li et al., 2013)	485	486	486	–	486	486
	CUHK02 (Li & Wang, 2013)	1677	239	239	–	239	239
	SenseReID (Zhao et al., 2017)	1718	521	1718	–	521	1718

E Person Retrieval Results

The qualitative comparison uses the final Order-1 checkpoints of ADSR and FSRC. It covers two seen domains, CUHK-SYSU and MSMT17-V2, and two unseen domains, CUHK01 and SenseReID. For every query, both methods use the same query image and gallery candidates, and the Top-10 gallery images are arranged from left to right in descending order of similarity. Query images are marked with black borders, while correct and incorrect gallery matches are marked with green and red borders, respectively. Both methods retrieve from the same complete gallery for each query, so the displayed differences reflect changes in ranking under identical candidates.

Figure 7 shows the results on seen domains. In the displayed CUHK-SYSU cases, ADSR frequently ranks pedestrians with similar clothing ahead of the true match when viewpoint, scale, and background change. FSRC moves valid matches into the Top-10 list and generally places them closer to the front. The contrast is clearer on MSMT17-V2, where cross-camera illumination changes are stronger. FSRC returns several correct instances for most displayed queries, whereas ADSR remains distracted by identities sharing the dominant clothing color. The improvement indicates that FSRC retains identity-relevant structure across the style changes encountered during lifelong training.

Figure 8 extends the comparison to unseen domains. On CUHK01, FSRC recovers cross-view matches despite the highly similar dark clothing shared by many gallery pedestrians. On SenseReID, the queries contain stronger blur, occlusion, and resolution changes. Correct matches still move from outside the Top-10 under ADSR into the displayed ranking under FSRC, showing that the learned response structure transfers beyond the training domains. Some false matches remain near the front when the person occupies few pixels or distinctive details are occluded, which identifies the main failure case shared by both methods.

The matched protocol also permits two complementary observations. An earlier green border measures how rapidly the first valid identity is recovered, whereas several green borders within one row indicate that different views of the same identity remain close in the embedding space. The MSMT17-V2 examples show the latter effect particularly clearly: the FSRC ranking often contains a sequence of correct cross-camera observations instead of a single isolated match. On CUHK-SYSU and the two unseen datasets, where fewer positives may enter the displayed list, the consistent advance of the first correct result provides the more informative comparison.

The seen and unseen panels further describe different requirements of the final lifelong model. CUHK-SYSU and MSMT17-V2 examine whether identity evidence from observed domains remains accessible after all sequential updates. CUHK01 and SenseReID examine whether the same representation supports identities that were never used for training. Improvements in both groups

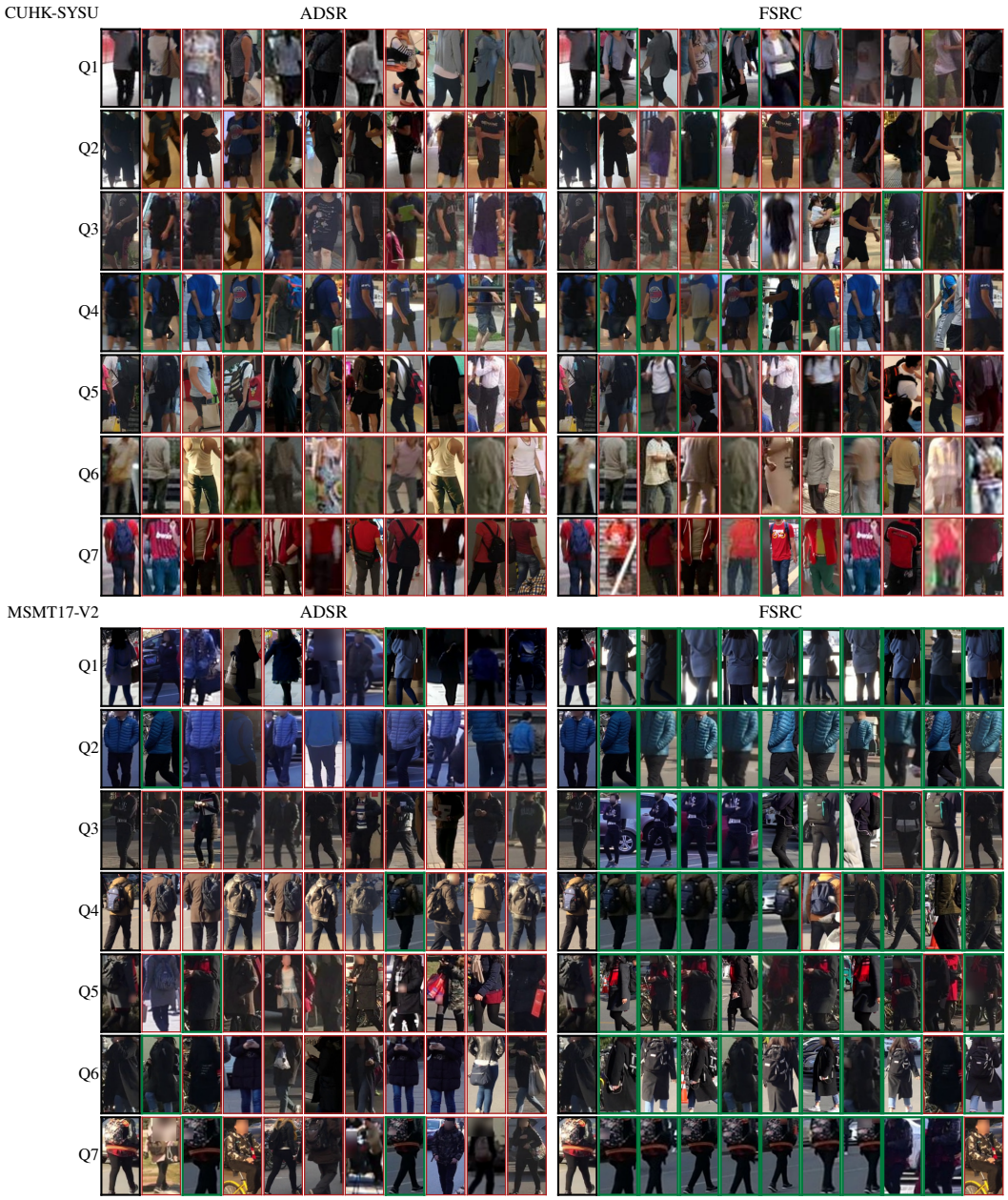


Figure 7: Person retrieval results on the seen domains CUHK-SYSU and MSMT17-V2. ADSR and FSRC use the same query and gallery for each comparison. Black borders denote queries; thick green and thin red borders denote correct and incorrect matches, respectively.

connect the change to historical-domain retention and cross-domain retrieval under the same final model.

Overall, the retrieval results complement the quantitative comparison by showing where the ranking changes occur. The largest improvements arise when global appearance is ambiguous but local identity evidence remains available across cameras. The remaining errors under severe visual degradation also show that FSRC improves the use of transferable evidence without eliminating the intrinsic ambiguity of low-quality person images.



Figure 8: Person retrieval results on the unseen domains CUHK01 and SenseReID. The visualization follows the protocol of Figure 7.