This WACV paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Low-Rank Expert Merging for Multi-Source Domain Adaptation in Person Re-Identification

Taha Mustapha Nehdi, Nairouz Mrabah, Atif Belal, Marco Pedersoli, Eric Granger
LIVIA, ILLS, Dept. of Systems Engineering, ETS Montreal, Canada
taha-mustapha.nehdi.1@ens.etsmtl.ca, nairouz.mrabah@gmail.com, atif.belal.1@ens.etsmtl.ca,
{marco.pedersoli, eric.granger}@etsmtl.ca

Abstract

Adapting person re-identification (reID) models to new target environments remains a challenging problem that is typically addressed using unsupervised domain adaptation (UDA) methods. Recent works show that when labeled data originates from several distinct sources (e.g., datasets and cameras), considering each source separately and applying multi-source domain adaptation (MSDA) typically yields higher accuracy and robustness compared to blending the sources and performing conventional UDA. However, state-of-the-art MSDA methods learn domain-specific backbone models or require access to source domain data during adaptation, resulting in significant growth in training parameters and computational cost. In this paper, a ***S****ource-free* ***A****daptive* ***G****ated* ***E****xperts* ***(SAGE-reID)*** *method is introduced for person reID. Our SAGE-reID is a cost-effective, source-free MSDA method that first trains individual source-specific low-rank adapters (LoRA) through source-free UDA. Then, a lightweight gating network is introduced and trained to dynamically assign optimal merging weights for fusion of LoRA experts, enabling effective cross-domain knowledge transfer. While the number of backbone parameters remains constant across source domains, LoRA experts scale linearly but remain negligible in size ($\leqslant 2\%$ per source), reducing both the memory consumption and risk of overfitting. Extensive experiments conducted on three challenging benchmarks – Market-1501, DukeMTMC-reID, and MSMT17 – indicate that SAGE-reID can outperform state-of-the-art methods while remaining computationally efficient. Our code is available:* https://github.com/nehdiii/SAGE-reID

1. Introduction

Person re-identification (reID) aims to match query images of individuals captured over a network of non-overlapping cameras against instances in a gallery. It has a wide-ranging applications in surveillance [41], smart city [21], and autonomous driving [34]. In recent years, supervised person reID systems have achieved remarkable progress [15, 24, 39, 43]. However, they often suffer from poor generalization when applied in unseen operational target domains due to the distribution shift between the source (lab training) and operational target (testing) data. Collecting large-scale labeled reID datasets is impractical due to the annotation cost. To mitigate this, several unsupervised domain adaptation (UDA) methods have been proposed [14, 23, 33, 46, 47], transferring knowledge from labeled source domains to unlabeled target domains. UDA methods for reID [4, 11, 12, 16, 20, 23, 26, 45–47, 51] typically adapt a pretrained model on a single labeled source domain dataset and unlabeled target domain dataset. They only rely on supervision from only one source dataset, leaving the rich variation of other source domains untapped.

MSDA for person reID was introduced [1] to leverage supervision across varying environmental conditions, like

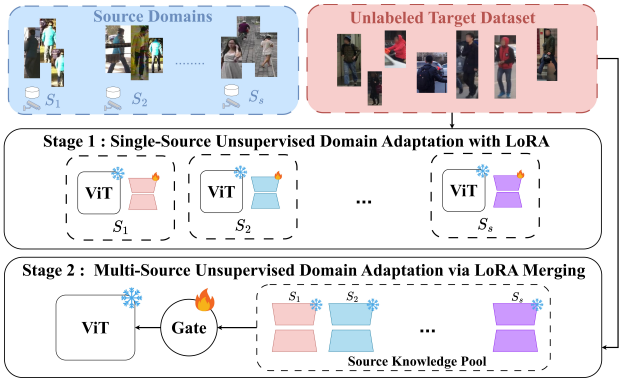


Figure 1. Illustration of our **SAGE-reID** method. **Stage 1:** Single-source UDA with LORA: each pre-trained source-specific model is augmented with low-rank adapters, and is finetuned on the target domain. **Stage 2:** LoRA fusion: a lightweight gating network predicts a linear combination of weights for each adapter and merges them on-the-fly in a single forward pass.

background, camera viewpoint, illumination, and occlusion. By integrating complementary knowledge from multiple labeled source domains, MSDA produces more robust and discriminative target-domain representations than UDA approaches [1, 36]. Early MSDA methods assume that all domains share the same label set [3, 9, 29, 40]. However, in person reID, each dataset has distinct identities, so this shared-label premise fails and often causes negative transfer. To address this, RDSBN [1] employs a two-stage pipeline: first, Rectified Domain-Specific Batch Normalization (RDSBN) adjusts each sample's features to remove dataset bias; second, a graph-based Multi-Domain Information Fusion (MDIF) module merges knowledge from all source domains. However, MDIF requires access to source data during adaptation, and this is often infeasible in real-world deployments due to privacy regulations, proprietary restrictions, and bandwidth constraints. CDR [36] is the only source-free MSDA method for person reID. Given pre-trained expert models on each source domain, CDR adapts each expert to the target by self-training its mean-teacher model through clustering-based pseudo-labels. To ensure consistent features from each expert, a dual similarity consistency and decorrelation losses are used. However, its distillation stage requires full-backbone training, and during inference, each image is processed through all experts, resulting in substantial computational and memory costs that grow linearly with the number of sources.

To overcome these limitations, we introduce **S**ource-free **A**daptive **G**ated **E**xperts for person reID **(SAGE-reID)**, an MSDA framework for person reID. Starting from s independently pre-trained source models, we adapt each one to the unlabeled target domain by only fine-tuning a lightweight low-rank adapter (LoRA) [17] appended to its frozen backbone (Fig. 1a). Pseudo-labels on the target domain are generated via DBSCAN clustering and are iteratively refined until convergence [7]. These pseudo-labels are used to train each LoRA expert independently, aligning the model to the target distribution. In the subsequent domain-fusion stage, all backbone and LoRA parameters are kept frozen. A "fused" backbone is initialized by averaging the weights of s source backbones corresponding to s source domains, while each LoRA expert retains its target-adapted weights. A compact gating network then predicts a linear combination of weights per sample and performs only a single forward pass to merge the s LoRA experts on-the-fly (Fig. 1b). This dynamic composition narrows the inter-domain gaps and reduces the computational and memory cost.

The contributions of this paper are summarized as follows. **(1)** A source-free MSDA method called **SAGE-reID** is introduced that merges low-rank adapters to adapt each pre-trained expert to the target distribution, leveraging low-rank adapters to elegantly and efficiently capture target-domain characteristics. **(2)** A lightweight gating module is designed that learns, in a "single" forward pass, an optimal linear combination of weights to merge the LoRA experts on-the-fly. **(3)** Extensive experiments are conducted on standard image-based person reID benchmarks, show that a significant gain in adaptation performance can be achieved with a cost-effective method.

2. Related Works

UDA for Person ReID. UDA methods for person ReID aim to mitigate the domain gap between the source and the target domains. These methods can be categorized into two groups: generative methods and self-training. The first category of approaches [5, 33] employs generative models, like Generative Adversarial Networks (GANs) [13] to translate source images into target style images with identity-preserving techniques [5]. The generated images are then used to train a re-ID model using supervised learning. These approaches often suffer from unstable training of GANs [22] and involve large parameter counts. The second line of work [4, 12, 16, 20, 23, 26, 45, 46, 50] relies on pseudo-label self-training, where a model trained on the source domain extracts features from the unlabeled target domain. These features are clustered, assigning pseudo-identity labels to each cluster. The quality of pseudo-labels significantly impacts re-ID performance on the target domain. Several studies [45, 46, 50] refine these labels by ensembling multiple re-ID models and enforcing prediction consistency [45, 46]; to mitigate camera bias, where clusters may contain images from a single camera, they employ cluster-weighting techniques [23]. Rather than adapting in one step, [4] progressively bridges domains by generating intermediate representations through cross-domain feature mixing [32]. Despite their impressive performance, all UDA methods use only a single source domain, underutilizing the abundance of available labeled datasets.

MSDA for Person ReID. MSDA aims to adapt the model using multiple labeled source datasets and an unlabeled target dataset. Only two work addresses MSDA for person reID: MSUDA [1] and CDR [36]. MSUDA [1] applies Rectification Domain-Specific Batch Norm (RDSBN) to neutralize dataset-specific artifacts and uses a graph convolution network to blend features via domain-agent nodes, yielding substantial gains over single-source baselines. However, MSUDA requires access to source data during adaptation, which might not be available due to privacy reasons. Source-free MSDA (MSFDA) eliminates the need for source dataset access at adaptation time by leveraging only pre-trained source models and unlabeled target samples. CDR [36] assigns each source domain its mean-teacher expert. After adapting each expert to the target domain, the features from all experts are concatenated to form a single embedding, and dual-similarity distillation and representation-decorrelation losses are employed to en-

force complementary information. While [36] achieves high performance, it trains all the parameters of each expert, causing computational cost to scale linearly with the number of sources. In contrast, our method only trains LoRA parameters, resulting in a significant reduction in computational overhead for the re-ID model regardless of the number of sources.

Merging Low-Rank Experts. Merging refers to the process of integrating separately fine-tuned low-rank adapters into a single model instance without full retraining. A conventional strategy for multi-skill learning involves training on a concatenation of datasets, often referred to as data mixing. While conceptually straightforward, this approach becomes increasingly impractical as the number of skills grows, since each new skill addition requires full model retraining on an updated mixture of datasets. To overcome this limitation, recent methods explore model merging as a scalable alternative. Rather than retraining, these techniques first independently fine-tune LoRA modules per skill and then merge them for composition. One family of approaches, known as the concatenation of LoRAs, linearly combines LoRA updates via layer-wise coefficients [2, 8, 27, 28, 35]. A second class of methods focuses on linear merging of LoRA parameters, which combines LoRA matrices before computing the final weight update. Examples include TIES [38] and DARE [42], which incorporate sparsity and sign consistency to improve merge quality, and LoRA Hub [18], which assigns static coefficients based on a few-shot target task. While these methods have primarily been studied for multi-task performance, we investigate their potential in MSDA.

3. Proposed Method

Consider s labeled source domains $\{\mathcal{S}^{(i)}\}_{i=1}^{s}$ and an unlabeled target domain $\mathcal{T}$. The i^{th} source domain holds $n_{S^{(i)}}$ image–label pairs $\mathcal{S}^{(i)} = \{(x_j^{(i)}, y_j^{(i)})\}_{j=1}^{n_{S^{(i)}}}$. The image $x_j^{(i)}$ denotes the j^{th} observation for the i^{th} source $\mathcal{S}^{(i)}$, and $y_j^{(i)}$ is its corresponding person identity. The target set contains n_T unlabeled samples $\mathcal{T} = \{x_j\}_{j=1}^{n_T}$.

In source-free adaptation, the source images are unavailable. Instead, the pretrained parameters of the s source models are used. Let L be the layer count and denote by $\mathbf{W}_0^{(i,\ell)}$ the frozen weight matrix of layer ℓ in source model i, with $i \in \{1, \dots, s\}$ and $\ell \in \{1, \dots, L\}$. We define the parameters of backbone i as $\mathcal{W}_0^{(i)} = \{\mathbf{W}_0^{(i,\ell)}\}_{\ell=1}^{L}$ and define the complete source prior as $\mathbb{W}_0 = \bigcup_{i=1}^{s} \mathcal{W}_0^{(i)}$. The goal is to exploit $\mathbb{W}_0$ and $\mathcal{T}$ to train a target model without revisiting any source data.

Low-rank adaptation. LoRA [17] augments each frozen weight matrix $\mathbf{W}_0$ with a trainable low-rank residual $\Delta\mathbf{W}$. The residual factorizes as $\Delta\mathbf{W} = \mathbf{A} \cdot \mathbf{B}$, where $\mathbf{A} \in \mathbb{R}^{n\times r}$ and $\mathbf{B} \in \mathbb{R}^{r\times m}$ with $r \ll \min(n, m)$. This low-rank factorization makes the adaptation fast and memory-efficient. The adapted weights are expressed as follows:

$$\mathbf{W} = \mathbf{W}_0 + \mathbf{A} \cdot \mathbf{B}. \tag{1}$$

During fine-tuning, only the parameters $\mathbf{A}$ and $\mathbf{B}$ of each layer are updated, while $\mathbf{W}_0$ stays fixed. This results in a parameter complexity of $\mathcal{O}\big(r(n+m)\big)$ for each layer, which is drastically lower than the $\mathcal{O}(nm)$ cost of full fine-tuning.

3.1. Training Overview.

This subsection describes the two-stage pipeline of our method, SAGE-reID. In the first stage, we start from the s source backbones, each pre-trained with cross-entropy and triplet losses on its source domain. For every source backbone i, we freeze all weights $\mathcal{W}_0^{(i)}$ and attach one LoRA pair $(\mathbf{A}^{(i,\ell)}, \mathbf{B}^{(i,\ell)})$ to every layer ℓ. Then, source-free UDA is performed on $\mathcal{T}$ via pseudo-supervision similar to [7]. Only the LoRA pairs are trained for each backbone, which produces s lightweight experts aligned to the target domain. In the second stage, a shared backbone is formed by averaging the weights of the source models $\overline{\mathcal{W}}_0 = \{\overline{\mathbf{W}}_0^{(\ell)}\}_{\ell=1}^{L}$, such that $\overline{\mathbf{W}}_0^{(\ell)} = \frac{1}{s}\sum_{i=1}^{s} \mathbf{W}_0^{(i,\ell)}$. Then, we fuse the experts through a linear combination:

$$\Delta\overline{\mathbf{W}}^{(\ell)} = \sum_{i=1}^{s} \alpha_{i,\ell}\, \Delta\overline{\mathbf{W}}^{(i,\ell)} = \sum_{i=1}^{s} \alpha_{i,\ell}\, \mathbf{A}^{(i,\ell)} \cdot \mathbf{B}^{(i,\ell)}, \tag{2}$$

where $\alpha_{i,\ell} \in [0, 1]$ is a learnable coefficient for the i^{th} source and ℓ^{th} layer. A single gradient step on the UDA loss estimates these coefficients while keeping the shared backbone and all LoRA parameters fixed. Figure 2 outlines the second stage of SAGE-reID. The next subsections elaborate on these two stages in full detail.

3.2. First Stage

Each source encoder is a Vision Transformer $f_{\mathcal{W}_0^{(i)}}$ followed by a linear classifier $g_{\Theta^{(i)}} : \mathbb{R}^D \to \mathbb{R}^{C^{(i)}}$, where $\mathbf{\Theta}^{(i)}$ denotes the classifier weights, D is the latent space dimension, and $C^{(i)}$ the number of identities in $\mathcal{S}^{(i)}$. For an image $x_j^{(i)}$, we extract the feature $z_j^{(i)} = f_{\mathcal{W}_0^{(i)}}(x_j^{(i)}) \in \mathbb{R}^D$. Backbone and classifier are trained to minimize the joint loss:

$$\begin{aligned}\mathcal{L}_{\mathcal{S}^{(i)}} = &\sum_{j=1}^{n_{S^{(i)}}} \mathcal{L}_{\text{CE}}\big(g_{\Theta^{(i)}}(z_j^{(i)}),\, y_j^{(i)}\big) \\ &+ \lambda \sum_{(a,p,n)\in\mathcal{P}^{(i)}} \mathcal{L}_{\text{Tri}}\big(z_a^{(i)}, z_p^{(i)}, z_n^{(i)}\big),\end{aligned} \tag{3}$$

$$\mathcal{L}_{\text{CE}}(\mathbf{u}, y) = -\log \frac{\exp(u_y)}{\sum_{c=1}^{C^{(i)}} \exp(u_c)}, \qquad \mathbf{u} = g_{\Theta^{(i)}}(z),$$

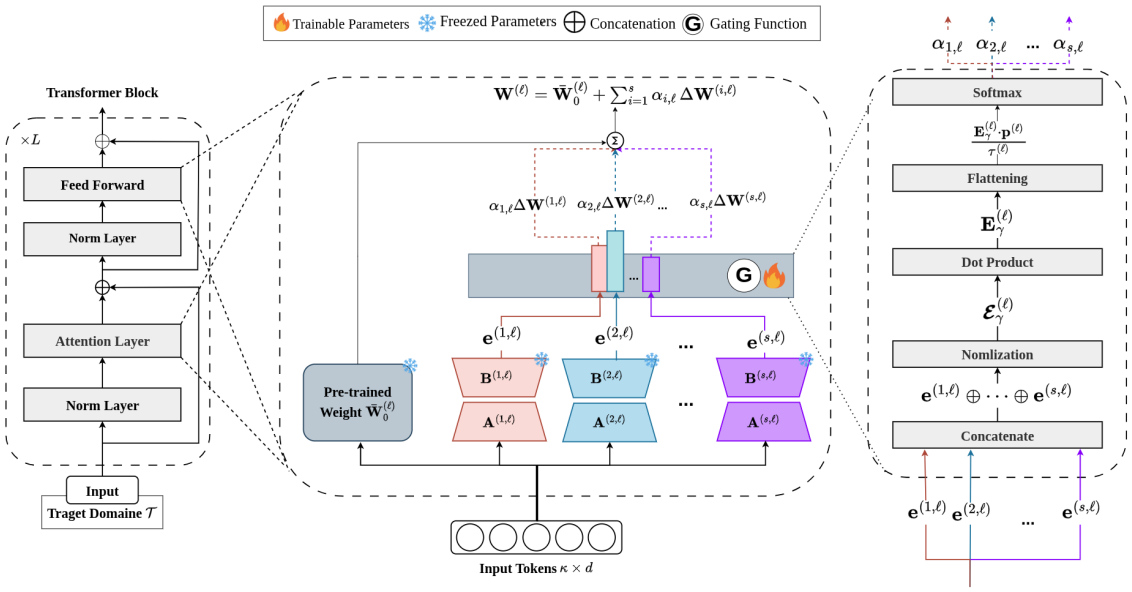


Figure 2. Illustration of the proposed low-rank expert merging stage (Stage 2) in our MSDA framework. Each source-specific LoRA expert, trained independently in Stage 1, provides a low-rank residual update. During training, a lightweight gating network dynamically predicts linear combination of coefficients for each expert at every transformer layer based on the input features. These coefficients are then used to combine the residuals, enabling fusion of multi-source knowledge through a single forward pass. The backbone and LoRA weights remain frozen, making this stage computationally light while enhancing generalization to the target domain.

$\mathcal{L}_{\text{Tri}}(z_a, z_p, z_n) = \max\{0,\ m + \|z_a - z_p\|_2 - \|z_a - z_n\|_2\}$,

where $\mathcal{P}^{(i)}$ is the set of mined anchor–positive–negative triplets, the scalar $m \in \mathbb{R}_+^*$ is the margin value, and the scalar $\lambda \in \mathbb{R}_+$ balances classification and metric objectives.

After pre-training, we append a trainable low-rank adapter $\Delta\mathbf{W}^{(i,\ell)} = \mathbf{A}^{(i,\ell)} \cdot \mathbf{B}^{(i,\ell)}$, with rank $r \ll \min(n, m)$, for every layer ℓ. The adapted model after integrating the low-rank modules is denoted $f_{\mathcal{W}_0^{(i)} + \Delta\mathcal{W}^{(i)}}$, where $\Delta\mathcal{W}^{(i)} = \{\Delta\mathbf{W}^{(i,\ell)}\}_{\ell=1}^{L}$. For every backbone i, we freeze $\mathcal{W}_0^{(i)}$ and update $\Delta\mathcal{W}^{(i)}$ and the classification head.

At each epoch, we extract target features $z_j^{(i)} = f_{\mathcal{W}_0^{(i)} + \Delta\mathcal{W}^{(i)}}(x_j)$ using backbone i and cluster them with DBSCAN to obtain pseudo-labels $\hat{y}_j^{(i)}$. The low-rank adapters are trained to minimize the unsupervised losses:

$$\mathcal{L}_{\text{UDA}}^{(i)} = \sum_{j=1}^{n_T} \mathcal{L}_{\text{CE}}\big(g_{\Theta_{\mathcal{T}}^{(i)}}(z_j^{(i)}),\ \hat{y}_j^{(i)}\big) + \lambda \sum_{(a,p,n)\in\hat{\mathcal{P}}^{(i)}} \mathcal{L}_{\text{Tri}}\big(z_a^{(i)}, z_p^{(i)}, z_n^{(i)}\big), \tag{4}$$

where $\hat{\mathcal{P}}^{(i)}$ is the set of anchor–positive–negative triplets formed under the current pseudo-labels for backbone i, and $\Theta_{\mathcal{T}}^{(i)}$ denotes the target classification head for backbone i. Each expert keeps its source prior while gradually aligning to the target distribution through the low-rank residuals.

3.3. Second Stage

After the s experts finish target adaptation, we build a shared backbone by averaging the frozen source weights:

$$\overline{\mathbf{W}}_0^{(\ell)} = \frac{1}{s}\sum_{i=1}^{s} \mathbf{W}_0^{(i,\ell)}, \qquad \overline{\mathcal{W}}_0 = \{\overline{\mathbf{W}}_0^{(\ell)}\}_{\ell=1}^{L}. \tag{5}$$

This shared backbone prevents drift when the low-rank residuals are merged in the next step.

Low-rank Expert Merging. The LoRA experts capture complementary biases toward the target domain. We fuse them through a lightweight, input-adaptive gate. For each layer ℓ, we keep the averaged backbone weight $\overline{\mathbf{W}}_0^{(\ell)}$ and LoRA parameters frozen. Then, we combine the experts through the coefficients $\alpha_{i,\ell} \in [0, 1]$ with $\sum_{i=1}^{s} \alpha_{i,\ell} = 1$. Thus, we obtain the weight matrix $\mathbf{W}^{(\ell)}$:

$$\mathbf{W}^{(\ell)} = \overline{\mathbf{W}}_0^{(\ell)} + \sum_{i=1}^{s} \alpha_{i,\ell}\, \Delta\mathbf{W}^{(i,\ell)}, \tag{6}$$

where $\Delta\mathbf{W}^{(i,\ell)} = \mathbf{A}^{(i,\ell)} \cdot \mathbf{B}^{(i,\ell)}$ denotes the low-rank update provided by expert i at layer ℓ.

Gating Mechanism. Let the input activation to layer ℓ through the shared backbone be $\mathbf{x}^{[\ell]} \in \mathbb{R}^{\kappa\times d}$ (we omit the batch index for clarity), with κ denoting the number of tokens and d the space dimension. Each expert produces a residual response $\mathbf{e}^{(i,\ell)}$ expressed as follows:

$$\mathbf{e}^{(i,\ell)} = \mathbf{x}^{[\ell]} \cdot \Delta\mathbf{W}^{(i,\ell)} \in \mathbb{R}^{\kappa\times d}. \tag{7}$$

To align the scale of residuals across experts trained on diverse sources, we concatenate and ℓ_2-normalize them:

$$\boldsymbol{\mathcal{E}}_{\gamma}^{(\ell)} = \mathrm{Norm}\big(\mathbf{e}^{(1,\ell)} \oplus \cdots \oplus \mathbf{e}^{(s,\ell)}\big) \in \mathbb{R}^{s\times\kappa\times d}, \quad (8)$$

where $\oplus$ is the concatenation operation. The normalization operation is applied with respect to the feature dimension d. We then flatten each expert residual to form a matrix $\mathbf{E}_{\gamma}^{(\ell)} \in \mathbb{R}^{s\times(\kappa d)}$. Finally, we compute the gating coefficient vector $\boldsymbol{\alpha}_\ell$ for layer ℓ using a trainable projection vector $\mathbf{p}^{(\ell)} \in \mathbb{R}^{\kappa d}$ and a learnable temperature value $\tau^{(\ell)} \in \mathbb{R}_+^*$ as expressed:

$$\boldsymbol{\alpha}_\ell = \mathrm{Softmax}\left(\frac{\mathbf{E}_{\gamma}^{(\ell)} \cdot \mathbf{p}^{(\ell)}}{\tau^{(\ell)}}\right) = \big[\alpha_{1,\ell}, \cdots, \alpha_{s,\ell}\big] \in \mathbb{R}^s. \quad (9)$$

Gate Training. To preserve both general and domain-specific knowledge, we freeze the shared backbone and LoRA parameters during the gate training process and optimize the gating coefficients $\mathbf{p}^{(\ell)}$ and temperature $\tau^{(\ell)}$ for each layer ℓ along with the target classification head $g_{\Theta_\mathcal{T}}$, where $\Theta_\mathcal{T}$ denotes the classifier weight matrix.

We denote the final model after averaging the backbones and integrating the gating mechanism as $f_{\overline{\mathcal{W}}}$, where $\overline{\mathcal{W}} = \{\mathbf{W}^{(\ell)}\}_{\ell=1}^{L}$ and the weight matrix $\mathbf{W}^{(\ell)}$ for layer ℓ is defined in Eq. (6). The target features $z_j = f_{\overline{\mathcal{W}}}(x_j)$ are extracted using the shared backbone and merged LoRAs, and cluster these features with DBSCAN to obtain pseudo-labels $\hat{y}_j$. The gating mechanism is trained to minimize the loss:

$$\mathcal{L}_{\text{UDA}} = \sum_{j=1}^{n_\mathcal{T}} \mathcal{L}_{\text{CE}}\big(g_{\Theta_\mathcal{T}}(z_j),\ \hat{y}_j\big) + \lambda \sum_{(a,p,n)\in\hat{\mathcal{P}}} \mathcal{L}_{\text{Tri}}\big(z_a, z_p, z_n\big), \quad (10)$$

where $\hat{\mathcal{P}}$ is the set of anchor–positive–negative triplets formed under the current pseudo-labels. Because of the small number of parameters, the gate converges quickly, requiring training for only one epoch on the target dataset. Each layer learns its own gating parameters $(\mathbf{p}^{(\ell)},\ \tau^{(\ell)})$. Thus, the model can adaptively combine expert contributions in a depth-aware fashion.

Parameter Cost. Unlike the expert-specific gating of [35], which flattens all dimensions in $\boldsymbol{\mathcal{E}}_{\gamma}^{(\ell)}$ and requires s projections of dimension $s\kappa d$ per layer, we use a *single* projection vector $\mathbf{p}^{(\ell)} \in \mathbb{R}^{\kappa d}$ per layer. This reduces the memory and computation to $\mathcal{O}(\kappa d)$ instead of $\mathcal{O}(s^2\kappa d)$ per layer while preserving sufficient expressiveness for multi-source domain adaptation.

4. Results and Discussion

4.1. Experimental Methodology

Datasets and Evaluation Metrics. Experiments are carried out on four widely used person reID benchmarks: Market-1501 (M) [49], DukeMTMC-reID (D) [30], CUHK03 (CU) [25], and MSMT17 (MS) [33]. Statistics of these datasets are provided in Appendix A (Suppl. Mat.). Performance is reported with mean Average Precision (mAP) and Cumulative Matching Characteristic (CMC) at Rank-1, following the standard evaluation protocol.

Implementation Details. We use the Vision Transformer (ViT) [6] pre-trained on ImageNet-21K and fine-tuned on ImageNet-1K as a backbone network, following the protocol in [15]. During the first stage, each backbone is trained on a source dataset, followed by a 1-D BatchNorm, an L_2-normalization layer, and a classification head. After that, LoRA adapters are inserted into all projection layers of each transformer block (query, key, value, output, up, and down projections). For Market-1501 and DukeMTMC, the rank and scaling coefficient of the low-rank adapters are set to r=4 and β=16, respectively. These hyperparameters are set to r=8 and β=32 for MSMT17. A LoRA dropout of 0.05 is applied across all datasets. All backbones are trained for 70 epochs with SGD and a fixed learning rate of 0.008. Input images are resized to 256×128 and augmented with random horizontal flip, random crop, and random erasing following the protocol in [52]. Each mini-batch consists of 64 images from 16 identities (4 images per identity). The hyperparameters remain unchanged for the second stage, except for the number of epochs, which is reduced to one. All experiments are conducted on two A100 GPUs.

Baselines Methods. Table 1 groups the competing approaches into four tiers that progressively relax domain-adaptation assumptions: (i) *Source-Only (No Adaptation).* A $\text{ViT}_{\text{B/16}}$ backbone is trained on a single labelled source with the supervised loss $\mathcal{L}_\mathcal{S}$ of Eq. (3) and evaluated directly on the target domain. This baseline constitutes a lower bound because no adaptation is performed on the target; (ii) *Single-Source UDA.* It includes classical unsupervised domain adaptation (UDA) methods, such as MMT [11], UNRN [46], GLT [48], IDM [4], SpCL [12], and PUL [7]. MMT, UNRN, GLT, and PUL adhere to a two-stage paradigm, first pre-training on a single labeled source dataset and then fine-tuning on the unlabeled target. In contrast, IDM and SpCL employ joint training, optimizing on both source and target images within every training mini-batch rather than running a separate source-only pre-training step. We add a baseline, $\mathcal{L}_{\text{UDA}}$ + $\text{ViT}_{\text{B/16}}$ (LoRA), which freezes a $\text{ViT}_{\text{B/16}}$ backbone after source pretraining, inserts LoRA adapters, and optimises these adapters with the unsupervised loss of Eq. (4); (iii) *Multi-Source UDA (MSDA).* Approaches such as MSUDA [1] and CDM [36] exploit multiple heterogeneous sources during training to capture complementary domain cues before adapting to the target. To isolate the benefit of multi-source awareness, we include, blending + $\mathcal{L}_{\text{UDA}}$ + $\text{ViT}_{\text{B/16}}$ (LoRA), a baseline that simply pools all sources, pretrains on the

merged dataset and then performs low-rank adaptation. Our full model, SAGE-reID, adds the second-stage merging and gating mechanism described in Sec. 3.3; (iv) *Supervised Methods.* The final block reports fully-supervised ResNet-50 and $\text{ViT}_{\text{B/16}}$ trained directly on target labels. This category indicates the upper bound performance achievable with complete supervision.

4.2. Comparison with State-of-the-Art Methods

MSMT17. In this benchmark, Market1501 (M), DukeMTMC-reID (D), and CUHK03 (CU) are used as source domains, while MSMT17 (MS) is used as the target domain. This setup is particularly challenging due to the large scale of the target dataset. In the single-source UDA setting, our LoRA-equipped ViT attains 39.0% mAP for M→MS and 41.0% mAP for D→MS, exceeding IDM by 5.5 pp and 7.5 pp, respectively. When all three sources are leveraged simultaneously, the full SAGE-reID pipeline achieves 44.1% mAP and 69.8% Rank-1, outperforming MSUDA by +9.2 mAP and +5.1 Rank-1. These results demonstrate that (i) parameter-efficient LoRA adapters enable Vision Transformers to generalize in large-scale re-ID transfer, and (ii) the proposed expert-merging stage effectively consolidates heterogeneous source knowledge.

DukeMTMC-reID. In this setting, Market1501 (M), MSMT17 (MS), and CUHK03 (CU) datasets are used as source domains, while the DukeMTMC-reID (D) dataset is used as the target domain. As shown in Table 1 (middle), under the UDA setting, replacing the ResNet50 backbone of PUL with a frozen $\text{ViT}_{\text{B/16}}$ + LoRA raises mAP from 60.1% to 69.2% for M→D and from 61.2% to 71.2% for MS→D). These results can be explained by LoRA capacity to alleviate overfitting and confirm the benefit of low-rank adaptation even without cross-domain fusion. In the *source-free* MSDA setting, the complete SAGE-reID pipeline yields 72.7% mAP and 84.1% Rank-1 and improves over the previous SOTA method, CDM, by +1.9 mAP and +1.8 Rank-1. This margin showcases the effectiveness of the proposed expert averaging and lightweight gating mechanism.

Market1501. In this setting, DukeMTMC-reID (D), MSMT17 (MS), and CUHK03 (CU) are used as the source domains, while the Market1501 (M) dataset serves as the target domain. This scenario is challenging due to the small scale of the target dataset, which can cause strong overfitting, particularly when no source images are available during adaptation. As shown in Table 1 (left), replacing PUL's ResNet-50 with a frozen $\text{ViT}_{\text{B/16}}$ + LoRA lifts mAP from 72.4% to 81.5% for D → M and from 73.5% to 82.6% for MS → M. Rank-1 shows similar gains. These results highlight the capacity of LoRA in reducing overfitting, particularly for a small training dataset. IDM (single-source) and MSUDA (multi-source) score higher than our methods, but both of them retain and replay the source images during adaptation. This huge extra data regularizes the model and counters overfitting on Market1501. In contrast, CDM and SAGE-reID are source-free. Among source-free MSDA approaches, SAGE-reID yields the best scores (83.7% mAP and 93.0% Rank-1), outperforming CDM by +2.5 mAP. Furthermore, our UDA scores match IDM and trail MSUDA by only 2.3 mAP, confirming the effectiveness of the proposed LoRA-gate strategy without relying on source data.

4.3. Ablation Study

Importance of Source Knowledge Separation. A central challenge in MSDA is how to exploit source-domain knowledge so that each domain contributes effectively to the target domain. Table 2 includes four baselines and shows results with respect to the number of source datasets. The four baselines are: (i) Blend + $\mathcal{L}_{\mathcal{S}}$ + $\text{ViT}_{\text{B/16}}$, all source images are pooled, and a single $\text{ViT}_{\text{B/16}}$ is trained with the supervised loss $\mathcal{L}_{\mathcal{S}}$ without any target fine-tuning; (ii) Avg + $\mathcal{L}_{\mathcal{S}}$ + $\text{ViT}_{\text{B/16}}$, a separate backbone is trained on each source. The s trained networks are merged by simple weight averaging without any target fine-tuning; (iii) Blend + $\mathcal{L}_{\text{UDA}}$ + $\text{ViT}_{\text{B/16}}$ (LoRA), the blended model of (i) is taken as frozen backbone. LoRA adapters are inserted and updated on the target via the unsupervised loss $\mathcal{L}_{\text{UDA}}$; (iv) Avg + $\mathcal{L}_{\text{UDA}}$ + $\text{ViT}_{\text{B/16}}$ (LoRA), identical to (iii) but the averaged backbone of (ii) is used instead of blending.

When only supervised pretraining is allowed, averaging the backbones (row 2) surpasses the blending-based model (row 1) by a significant margin, e.g. +3.3 mAP for two sources and +4 mAP for three sources. These results confirm that training with the sources independently avoids negative interference. After unsupervised adaptation, both LoRA variants boost mAP by more than 25 pp in mAP over the supervised counterparts, showing that lightweight target tuning is crucial. As we can see from the same table, initializing the adapters from the averaged backbone (row 4) is consistently better than the blending-based model (row 3). Furthermore, blending can even degrade performance when the number of sources increases. This result confirms again the negative interference caused by blending the sources.

Our model (row 5) adds a gating mechanism to fuse the LoRA experts. This yields another gain of 0.7–1.4 mAP over the best baseline and reaches 44.1%/69.8% (mAP/R-1) with three sources. These results show the effectiveness of source separation followed by learned re-composition.

Effectiveness of Low-Rank Adaptation. Table 3 disentangles the contribution of (i) the backbone architecture, (ii) parameter–efficient low-rank adapters, and (iii) the proposed gating fusion. All variants share the same two-stage protocol: supervised pre-training on the source set with $\mathcal{L}_{\mathcal{S}}$, followed by source–free adaptation on the target with the pseudo-label loss $\mathcal{L}_{\text{UDA}}$. First, we observe that replac-

Methods	Target: Market1501			Target: DukeMTMC-reID			Target: MSMT17		
	Source	mAP	R-1	Source	mAP	R-1	Source	mAP	R-1
· ***Lower Bound: Source-Only (No Adaptation)***									
$\mathcal{L}_S$ + ViT$_{B/16}$	D	34.0	61.3	M	42.4	60.9	M	12.6	32.0
$\mathcal{L}_S$ + ViT$_{B/16}$	MS	40.1	66.9	MS	51.7	68.7	D	13.1	34.5
$\mathcal{L}_S$ + ViT$_{B/16}$	D + MS + CU	**49.5**	**75.3**	M + MS + CU	**56.0**	**71.7**	D + M + CU	**21.0**	**46.4**
· ***UDA Methods***									
MMT [11]	D	71.2	87.7	M	65.1	78.0	M	24.0	50.1
UNRN [46]	D	78.1	91.9	M	69.1	82.0	M	34.2	65.2
GLT [48]	D	79.5	92.2	M	69.2	82.0	M	26.5	56.6
IDM [4]	D	**82.8**	**93.2**	M	**70.5**	**83.6**	M	33.5	61.3
SpCL [12]	–	–	–	–	–	–	M	26.8	53.7
PUL [7]	D	72.4	87.7	M	60.1	76.1	M	12.0	27.5
$\mathcal{L}_{UDA}$ + ViT$_{B/16}$ (LoRA)	D	81.5	92.5	M	69.2	81.1	M	**39.0**	**63.5**
SpCL [12]	MS	77.5	89.7	–	–	–	–	–	–
GLT [48]	–	–	–	–	–	–	–	27.7	59.5
PUL [7]	MS	73.5	88.7	MS	61.2	76.0	D	12.8	29.5
$\mathcal{L}_{UDA}$ + ViT$_{B/16}$ (LoRA)	MS	**82.6**	**92.8**	MS	**71.2**	**82.5**	D	**41.0**	**65.8**
· ***MSDA Methods***									
MSUDA [1]	D + MS + CU	**86.0**	**94.8**	M + MS + CU	**68.9**	**82.1**	D + M + CU	**34.9**	**64.7**
· ***MSFDA: Source-free MSDA Methods***									
CDM [36]	D + MS + CU	81.2	92.9	M + MS + CU	70.8	82.3	D + M + CU	32.9	63.8
Blend + $\mathcal{L}_{UDA}$ + ViT$_{B/16}$ (LoRA)	D + MS + CU	80.5	92.3	M + MS + CU	65.1	79.6	D + M + CU	42.7	67.4
SAGE-ReID	D + MS + CU	**83.7**	**93.0**	M + MS + CU	**72.7**	**84.1**	D + M + CU	**44.1**	**69.8**
· ***Upper Bound: Supervised Methods***									
ALDER + ResNet50 [44]	–	**88.9**	**95.6**	–	78.9	89.9	–	59.1	82.2
DCAL + ViT$_{B/16}$ [53]	–	87.5	94.7	–	**80.1**	**89.0**	–	**64.0**	**83.1**

Table 1. Performance of SAGE-ReID against state-of-the-art adaptation methods on three targets: Market-1501 (left), DukeMTMC-reID (centre), and MSMT17 (right). The table groups approaches into *Source-Only*, *Single-Source UDA*, *Multi-Source UDA (MSDA)*, and *Supervised Methods*. Best scores in each block are in **bold** and grey rows denote our methods.

Methods	D		D+CU		D+CU+M	
	mAP	R-1	mAP	R-1	mAP	R-1
Blend + $\mathcal{L}_S$ + ViT$_{B/16}$	13.1	34.5	15.8	37.5	17.0	39.9
Avg + $\mathcal{L}_S$ + ViT$_{B/16}$	**13.1**	**34.5**	**19.1**	**44.3**	**21.0**	**46.4**
Blend + $\mathcal{L}_{UDA}$ + ViT$_{B/16}$ (LoRA)	41.0	65.8	40.3	65.3	42.7	67.4
Avg + $\mathcal{L}_{UDA}$ + ViT$_{B/16}$ (LoRA)	41.0	65.8	41.5	67.4	43.4	68.4
SAGE-ReID	**41.0**	**65.8**	**43.5**	**69.3**	**44.1**	**69.8**

Table 2. Impact of source knowledge separation on MSMT17 with different numbers of source domains using ViT$_{B/16}$ backbone.

ing the fully fine-tuned CNN baseline ($\mathcal{L}_{UDA}$ + ResNet50) with a fully fine-tuned Vision Transformer ($\mathcal{L}_{UDA}$ + ViT$_{B/16}$) yields large accuracy gains on every transfer path. For the challenging M → MS task, mAP rises from 12.0 % to 39.4 %, while for M → D, it jumps from 60.1 % to 68.1 %. These results confirm the stronger representation capacity of ViT even when source data are no longer accessible during adaptation. Second, inserting low-rank adapters and freezing the ViT weights ($\mathcal{L}_{UDA}$ + ViT$_{B/16}$(LoRA)) further improves or matches the fully tuneable ViT in five out of the nine single-source transfers. The benefit is most visible when the target is small: on MS → M mAP grows from 76.7 % to 82.6 %, and on CU → M from 79.1 % to 81.6 %. These results support the finding that LoRA mitigates over-fitting by restricting the number of trainable parameters. Third, SAGE-reID merges three LoRA experts with the lightweight gate of Sec. 3.3. The fused model surpasses the best single-source LoRA on every benchmark: +3.1 mAP / +4.0 R-1 on MSMT17, +1.1 mAP / +0.2 R-1 on Market1501, and +1.5 mAP / +0.8 R-1 on DukeMTMC. These steady but non-trivial margins indicate that the gate learns to combine complementary domain-specific adaptations, which can not be obtained by naive weight averaging.

Comparison with MSUDA. MSUDA [1] proposes two architectural enhancements: (i) RDSBN, an improved domain-specific batch-normalization layer that strengthens domain-specific learning, and (ii) MDIF, a feature-fusion module. Because the original model was trained with the MMT loss [11], we re-evaluate it with the loss function $\mathcal{L}_{UDA}$ for a fair comparison. The results are presented in Table 4. Our method outperforms MSUDA by +3.3 mAP

Methods	M		D		CU	
	mAP	R-1	mAP	R-1	mAP	R-1
Target: MSMT17						
$\mathcal{L}_{\text{UDA}}$ + ResNet50	12.0	27.5	12.8	29.5	11.5	26.9
$\mathcal{L}_{\text{UDA}}$ + ViT$_{\text{B/16}}$	39.4	65.5	39.5	65.9	36.7	62.8
$\mathcal{L}_{\text{UDA}}$ + ViT$_{\text{B/16}}$ (LoRA)	39.0	63.5	41.0	65.8	38.3	64.4
Multi-Source (M + D + CU):						
SAGE-ReID			**44.1**	**69.8**		
Methods	**D**		**CU**		**MS**	
	mAP	**R-1**	**mAP**	**R-1**	**mAP**	**R-1**
Target: Market1501						
$\mathcal{L}_{\text{UDA}}$ + ResNet50	72.4	87.7	72.5	87.6	73.5	88.7
$\mathcal{L}_{\text{UDA}}$ + ViT$_{\text{B/16}}$	72.6	88.0	79.1	91.7	76.7	89.8
$\mathcal{L}_{\text{UDA}}$ + ViT$_{\text{B/16}}$ (LoRA)	81.5	92.5	81.6	92.0	82.6	92.8
Multi-Source (D + CU + MS):						
SAGE-ReID			**83.7**	**93.0**		
Methods	**M**		**CU**		**MS**	
	mAP	**R-1**	**mAP**	**R-1**	**mAP**	**R-1**
Target: DukeMTMC-reID						
$\mathcal{L}_{\text{UDA}}$ + ResNet50	60.1	76.1	60.3	76.1	61.2	76.0
$\mathcal{L}_{\text{UDA}}$ + ViT$_{\text{B/16}}$	68.1	80.7	65.4	78.8	68.1	80.9
$\mathcal{L}_{\text{UDA}}$ + ViT$_{\text{B/16}}$ (LoRA)	69.2	81.1	70.4	82.4	71.2	82.5
Multi-Source (M + CU + MS):						
SAGE-ReID			**72.7**	**83.3**		

Table 3. Impact of LoRA adapters, Vision Transformers, and the gating mechanism. All results are reported after adaptation, and no source images are used during fine-tuning.

Methods	D + CU → M		M + CU → D	
	mAP	R-1	mAP	R-1
$\mathcal{L}_{\text{UDA}}$ + DSBN*	74.1	89.2	60.0	76.2
$\mathcal{L}_{\text{UDA}}$ + MSUDA*	79.4	92.1	65.0	79.1
SAGE-ReID	**82.7**	**92.8**	**72.2**	**83.2**

Table 4. Multi-source adaptation under $\mathcal{L}_{\text{UDA}}$ with a ViT-B/16 backbone. *Results copied from the original MSUDA paper [1].

and +0.7 top-1 accuracy and +7.2 mAP and +4.1 top-1 accuracy in D + CU → M and M + CU → D, respectively, which shows the effectiveness of the proposed method.

Comparison with LoRA Composition and Weight Merging. In Table 5, we benchmark state-of-the-art merging families under our MSFDA protocol with identical ViT-B/16 backbones, $\mathcal{L}_{\text{UDA}}$ pseudo-supervision, and training budgets (all methods fuse the same target-adapted LoRA experts) on two (D + CU + M) → MS and (CU + M + MS) → D transfers. Task Arithmetic (TA) [19] creates task vectors (difference between the finetuned and pretrained weights) then merges them by summation, yielding moderate results (33.5/59.4 on MS; 66.2/80.8 on D). TIES [37] resolves sign conflicts in task-vector space but underperforms TA on MS (23.1/49.2) while remaining competitive on D (57.3/72.8). KnOTS [31] observes that LoRA updates lie in a different subspace than full-weight task vectors. To address this problem, KnOTS aligns the LoRAs into a shared subspace, improving both TA and TIES (KnOTS–TA: 34.8/61.6 on MS, 67.3/80.5 on D; KnOTS–TIES: 24.8/51.6 on MS, 59.2/73.6 on D). TSV-M [10] further reduces interference via singular-vector decorrelation between task vectors. Thus, the merged directions for TSV-M interfere less (36.6/62.9 on MS; 70.0/80.3 on D). Among LoRA-focused composition methods with few-shot target supervision (pseudo-labels in our case), LoRAHub [18] learns global static and layer-shared coefficients and surpasses data-free approaches (40.4/67.4 on MS; 69.4/82.3 on D), while MoLE [35] performs per-token expert gating (i.e., for each token embedding, the gate predicts soft weights across LoRA experts) and yields strong results (43.8/69.4 on MS and 71.5/83.0 on D). However, our gating attains the best results on both transfers with a margin over MoLE of +1.2/+0.3 (mAP/R-1) on (CU+M+MS)→D (Table 5). The gains indicate that the input-aware composition of LoRA residuals is more effective than static or weight-space merges, while retaining favorable compute/parameter cost (see Suppl. Tables 4 & 5).

Method	Venue	(D + CU + M) → MS		(CU + M + MS) → D	
		mAP	R-1	mAP	R-1
TIES	NeurIPS'24	23.1	49.2	57.3	72.8
KnOTS–TIES	ICLR'25	24.8	51.6	59.2	73.6
TA	ICLR'23	33.5	59.4	66.2	80.8
KnOTS–TA	ICLR'25	34.8	61.6	67.3	80.5
TSV-M	CVPR'25	36.6	62.9	70.0	80.3
LoRAHub	COLM'24	40.4	67.4	69.4	82.3
MoLE	ICLR'24	43.8	69.4	71.5	83.0
SAGE-ReID	-	**44.1**	**69.8**	**72.7**	**83.3**

Table 5. LoRA and weight merging methods under our MSFDA protocol (all methods fuse the same target-adapted LoRA experts).

5. Conclusion

Current MSDA methods for person re-identification require access to source-domain data during adaptation and train a full backbone for each source. In this paper, we introduced SAGE-reID, a source-free MSDA framework that fine-tunes only lightweight low-rank adapters (LoRA) per source model and leverages a compact, learnable gating network to fuse them on-the-fly in a single forward pass. This design requires training only a few parameters, significantly reducing the computational and memory costs, and eliminating the need for source data at adaptation time. Comprehensive experiments on Market-1501, DukeMTMC-reID, and MSMT17 demonstrate that SAGE-reID consistently outperforms state-of-the-art MSDA approaches in both accuracy and efficiency. A potential limitation is the naive averaging of source backbones for initialization. Future work will explore fully unsupervised initialization schemes.

Acknowledgements This work was supported by Distech Controls Inc., the Natural Sciences and Engineering Research Council of Canada, the Digital Research Alliance of Canada, and MITACS.

References

[1] Zechen Bai, Zhigang Wang, Jian Wang, Di Hu, and Errui Ding. Unsupervised multi-source domain adaptation for person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12914–12923, 2021. 1, 2, 5, 7, 8

[2] Eric L. Buehler and Markus J. Buehler. X-lora: Mixture of low-rank adapter experts, a flexible framework for large language models with applications in protein mechanics and molecular design. *APL Machine Learning*, 2(2):026119, 2024. 3

[3] Woong-Gi Chang, Tackgeun You, Seonguk Seo, Suha Kwak, and Bohyung Han. Domain-specific batch normalization for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7354–7362, 2019. 2

[4] Yongxing Dai, Jun Liu, Yifan Sun, Zekun Tong, Chi Zhang, and Ling-Yu Duan. Idm: An intermediate domain module for domain adaptive person re-id. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 11864–11874, 2021. 1, 2, 5, 7

[5] Weijian Deng, Liang Zheng, Qixiang Ye, Guoliang Kang, Yi Yang, and Jianbin Jiao. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 2

[6] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations (ICLR)*, 2021. 5

[7] Hehe Fan, Liang Zheng, Chenggang Yan, and Yi Yang. Unsupervised person re-identification: Clustering and fine-tuning. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 14(4):1–18, 2018. 2, 3, 5, 7

[8] Wenfeng Feng, Chuzhan Hao, Yuewei Zhang, Yu Han, and Hao Wang. Mixture-of-loras: An efficient multitask tuning method for large language models. In *Proceedings of the Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING)*, pages 11371–11380, 2024. 3

[9] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 1180–1189, 2015. 2

[10] Antonio Andrea Gargiulo, Donato Crisostomi, Maria Sofia Bucarelli, Simone Scardapane, Fabrizio Silvestri, and Emanuele Rodolà. Task singular vectors: Reducing task interference in model merging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18695–18705, 2025. 8

[11] Yixiao Ge, Dapeng Chen, and Hongsheng Li. Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification. In *International Conference on Learning Representations (ICLR)*, 2020. 1, 5, 7

[12] Yixiao Ge, Feng Zhu, Dapeng Chen, Rui Zhao, and hongsheng Li. Self-paced contrastive learning with hybrid memory for domain adaptive object re-id. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020. 1, 2, 5, 7

[13] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 2672–2680, 2014. 2

[14] Alexis Guichemerre, Soufiane Belharbi, Tsiry Mayet, Shakeeb Murtaza, Pourya Shamsolmoali, Luke Mccaffrey, and Eric Granger. Source-free domain adaptation of weakly-supervised object localization models for histology. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 33–43, 2024. 1

[15] Shuting He, Hao Luo, Pichao Wang, Fan Wang, Hao Li, and Wei Jiang. Transreid: Transformer-based object re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 15013–15022, 2021. 1, 5

[16] Tao He, Leqi Shen, Yuchen Guo, Guiguang Ding, and Zhenhua Guo. Secret: Self-consistent pseudo label refinement for unsupervised domain adaptive person re-identification. In *Proceedings of the AAAI conference on Artificial Intelligence (AAAI)*, pages 879–887, 2022. 1, 2

[17] Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2022. 2, 3

[18] Chengsong Huang, Qian Liu, Bill Yuchen Lin, Tianyu Pang, Chao Du, and Min Lin. Lorahub: Efficient cross-task generalization via dynamic loRA composition. In *Conference on Language Modeling (COLM)*, 2024. 3, 8

[19] Gabriel Ilharco, Marco Tulio Ribeiro, Mitchell Wortsman, Suchin Gururangan, Ludwig Schmidt, Hannaneh Hajishirzi, and Ali Farhadi. Editing models with task arithmetic. *arXiv preprint arXiv:2212.04089*, 2022. 8

[20] Takashi Isobe, Dong Li, Lu Tian, Weihua Chen, Yi Shan, and Shengjin Wang. Towards discriminative representation learning for unsupervised person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8526–8536, 2021. 1, 2

[21] Samee Ullah Khan, Tanveer Hussain, Amin Ullah, and Sung Wook Baik. Deep-reid: Deep features and autoencoder assisted image patching strategy for person re-identification in smart cities surveillance. *Multimedia Tools and Applications*, 83(5):15079–15100, 2024. 1

[22] Naveen Kodali, Jacob Abernethy, James Hays, and Zsolt Kira. On convergence and stability of gans. *arXiv preprint arXiv:1705.07215*, 2017. 2

[23] Geon Lee, Sanghoon Lee, Dohyung Kim, Younghoon Shin, Yongsang Yoon, and Bumsub Ham. Camera-driven representation learning for unsupervised domain adaptive per-

son re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 11453–11462, 2023. 1, 2

[24] Hanjun Li, Gaojie Wu, and Wei-Shi Zheng. Combined depth space based architecture search for person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6729–6738, 2021. 1

[25] 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 (CVPR)*, 2014. 5

[26] Zongyi Li, Yuxuan Shi, Hefei Ling, Jiazhong Chen, Qian Wang, and Fengfan Zhou. Reliability exploration with self-ensemble learning for domain adaptive person re-identification. In *Proceedings of the AAAI conference on Artificial Intelligence (AAAI)*, pages 1527–1535, 2022. 1, 2

[27] Tongxu Luo, Jiahe Lei, Fangyu Lei, Weihao Liu, Shizhu He, Jun Zhao, and Kang Liu. Moelora: Contrastive learning guided mixture of experts on parameter-efficient fine-tuning for large language models. *arXiv preprint arXiv:2402.12851*, 2024. 3

[28] Mohammed Muqeeth, Haokun Liu, Yufan Liu, and Colin Raffel. Learning to route among specialized experts for zero-shot generalization. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 1496–1513, 2024. 3

[29] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019. 2

[30] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pages 17–35, 2016. 5

[31] George Stoica, Pratik Ramesh, Boglarka Ecsedi, Leshem Choshen, and Judy Hoffman. Model merging with svd to tie the knots. *arXiv preprint arXiv:2410.19735*, 2024. 8

[32] Vikas Verma, Alex Lamb, Christopher Beckham, Amir Najafi, Ioannis Mitliagkas, David Lopez-Paz, and Yoshua Bengio. Manifold mixup: Better representations by interpolating hidden states. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 6438–6447, 2019. 2

[33] 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 (CVPR)*, 2018. 1, 2, 5

[34] Kelvin Wong, Shenlong Wang, Mengye Ren, Ming Liang, and Raquel Urtasun. Identifying unknown instances for autonomous driving. In *Proceedings of the Conference on Robot Learning (CoRL)*, pages 384–393, 2020. 1

[35] Xun Wu, Shaohan Huang, and Furu Wei. Mixture of lora experts. In *International Conference on Learning Representations (ICLR)*, 2024. 3, 5, 8

[36] Yuqiao Xian, Yi-Xing Peng, Xing Sun, and Wei-Shi Zheng. Distilling consistent relations for multi-source domain adaptive person re-identification. *Pattern Recognition (PR)*, 157: 110821, 2025. 2, 3, 5, 7

[37] Prateek Yadav, Derek Tam, Leshem Choshen, Colin A Raffel, and Mohit Bansal. Ties-merging: Resolving interference when merging models. In *Advances in Neural Information Processing Systems*, pages 7093–7115. Curran Associates, Inc., 2023. 8

[38] Prateek Yadav, Derek Tam, Leshem Choshen, Colin A Raffel, and Mohit Bansal. Ties-merging: Resolving interference when merging models. *Advances in Neural Information Processing Systems*, 36:7093–7115, 2023. 3

[39] Jinxi Yang, He Li, Bo Du, and Mang Ye. Cheb-gr: Rethinking k-nearest neighbor search in re-ranking for person re-identification. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, pages 19261–19270, 2025. 1

[40] Luyu Yang, Yogesh Balaji, Ser-Nam Lim, and Abhinav Shrivastava. Curriculum manager for source selection in multi-source domain adaptation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 608–624, 2020. 2

[41] Mang Ye, Jianbing Shen, Gaojie Lin, Tao Xiang, Ling Shao, and Steven CH Hoi. Deep learning for person re-identification: A survey and outlook. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 44(6): 2872–2893, 2021. 1

[42] Le Yu, Bowen Yu, Haiyang Yu, Fei Huang, and Yongbin Li. Language models are super mario: Absorbing abilities from homologous models as a free lunch. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2024. 3

[43] Guiwei Zhang, Yongfei Zhang, Tianyu Zhang, Bo Li, and Shiliang Pu. Pha: Patch-wise high-frequency augmentation for transformer-based person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14133–14142, 2023. 1

[44] Quan Zhang, Jianhuang Lai, Zhanxiang Feng, and Xiaohua Xie. Seeing like a human: Asynchronous learning with dynamic progressive refinement for person re-identification. *IEEE Transactions on Image Processing (TIP)*, 31:352–365, 2021. 7

[45] Fang Zhao, Shengcai Liao, Guo-Sen Xie, Jian Zhao, Kaihao Zhang, and Ling Shao. Unsupervised domain adaptation with noise resistible mutual-training for person re-identification. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 526–544, 2020. 1, 2

[46] Kecheng Zheng, Cuiling Lan, Wenjun Zeng, Zhizheng Zhang, and Zheng-Jun Zha. Exploiting sample uncertainty for domain adaptive person re-identification. In *Proceedings of the AAAI conference on Artificial Intelligence (AAAI)*, pages 3538–3546, 2021. 1, 2, 5, 7

[47] Kecheng Zheng, Wu Liu, Lingxiao He, Tao Mei, Jiebo Luo, and Zheng-Jun Zha. Group-aware label transfer for domain adaptive person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5310–5319, 2021. 1

[48] Kecheng Zheng, Wu Liu, Lingxiao He, Tao Mei, Jiebo Luo, and Zheng-Jun Zha. Group-aware label transfer for domain adaptive person re-identification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5310–5319, 2021. 5, 7

[49] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2015. 5

[50] Yi Zheng, Shixiang Tang, Guolong Teng, Yixiao Ge, Kaijian Liu, Jing Qin, Donglian Qi, and Dapeng Chen. Online pseudo label generation by hierarchical cluster dynamics for adaptive person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8371–8381, 2021. 2

[51] Yi Zheng, Shixiang Tang, Guolong Teng, Yixiao Ge, Kaijian Liu, Jing Qin, Donglian Qi, and Dapeng Chen. Online pseudo label generation by hierarchical cluster dynamics for adaptive person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8371–8381, 2021. 1

[52] Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. Random erasing data augmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pages 13001–13008, 2020. 5

[53] Haowei Zhu, Wenjing Ke, Dong Li, Ji Liu, Lu Tian, and Yi Shan. Dual cross-attention learning for fine-grained visual categorization and object re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4692–4702, 2022. 7