Supplementary Material for: "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

A. Datasets and Evaluation Metrics

Experiments are carried out on four widely used person reID benchmarks: Market-1501 (M) [8], DukeMTMC-reID (D) [5], CUHK03 (CU) [2], and MSMT17 (MS) [6], IUSTPersonReID (IUST) [4], Occluded-DukeMTMC (OC-D)[3]. Statistics of these datasets are summarized in Table 1. Performance is reported with mean Average Precision (mAP) and Cumulative Matching Characteristic (CMC) at Rank-1, following the standard protocol.

B. Multi-camera as Multi-Source Case Study

A real-world scenario arises when a new camera is added to a pre-existing video surveillance network. This camera observes the scene from previously uncovered operating conditions (e.g., viewpoint and illumination). To emulate this scenario using the MSMT17 dataset, one of the largest and most challenging person re-ID benchmarks, containing over 15 camera subsets. Cameras 12–15 are designated as source domains, while cameras 6 and 5 as target domains. As we can see, averaging two source backbones, Cams 15+14 already surpasses the strongest single-source baseline on Cam 5 (82.2 mAP vs. 79.3 mAP). These results indicate that complementary viewpoints are exploitable even without target data. Furthermore, unsupervised adaptation boosts every source set. For instance, Cam 12 → Cam 5 rises from 71.4% to 75.1% mAP (+3.7 pp). When multiple experts are available, naive source blending can decrease mAP and R-1 (e.g., Cams 15+14+13 → Cam 5). The proposed SAGE-ReID gate instead improves performance to 84.4% mAP / 93.4% R-1, and to 80.8% mAP / 93.8% R-1 when four experts are fused for Cam 6. Latency scales modestly with the number of active experts, from 0.10ms (one camera) to 0.36ms (four cameras), well below a 30fps real-time budget.

C. Effectiveness of the Gating Mechanism

Table 3 reports the proposed gating function performance versus two possible techniques. All variants first freeze the $\text{ViT}_{\text{B/16}}$ backbones trained on each source with $\mathcal{L}_{\mathcal{S}}$, then attach domain-specific low-rank adapters trained on the target domain with the pseudo-label loss $\mathcal{L}_{\text{UDA}}$. After that, all variants average the frozen source weights to build a shared backbone. Finally, we compare three ways to fuse the low-rank adapters: (i) LoRA-Avg, simply averages the low-rank adapters; (ii) LoRA-MoLE, learns a per-token mixture of experts as in [7]; (iii) SAGE-ReID employs the proposed gating mechanism as described in Sec. 3.2 of the main document. The proposed gate gives the best mAP on all transfers. On MSMT17, SAGE-ReID improves over averaging by +0.9 mAP and over MoLE by +0.3 mAP. On DukeMTMC, SAGE-ReID outperforms both baselines by more than +1.0 mAP. MoLE narrows the gap on the small Market-1501 target but does so at the cost of a large gating tensor that scales quadratically with the number of sources s and linearly with the sequence length κ. Table 4 shows that for s=10 experts, MoLE inflates parameters by $\times 11$, whereas our gate adds only +9 M parameters and +0.1 GFLOPs. The learnable parameters of our gate are of size $\kappa \times d$. Thus, the number of parameters and FLOPs stays fixed no matter how many sources there are. The shared gate, therefore, achieves better performance and cost while avoiding expert collapse observed with naive averaging. These results confirm that the proposed gate is effective and efficient to leverage complementary LoRA experts.

D. Time & Memory & Parameter Efficiency

We profile training-time and memory for the (M + D + CU) → MS transfer. Table 5 reports total/trainable number of parameters, training memory, per-iteration latency, and wall-clock time (batch size 32, 256×128, fp32).

Statistic	Market1501 [8]	DukeMTMC-reID [5]	CUHK03 [2]	MSMT17 [6]	IUSTPersonReID [4]	Occluded-DukeMTMC [3]
# Cameras	6	8	2	15	19	8
# Images	32,217	36,411	28,193	126,441	117,455	36,411
# IDs	1,501	1,404	1,467	4,101	1,847	1221
Train (img / IDs)	12,936 / 751	16,522 / 702	26,263 / 1,367	32,621 / 1,041	72,393 / 1,193	15,618 / 702
Query (img / IDs)	3,368 / 750	2,228 / 702	200 / 100	11,659 / 3,060	1,428 / 654	2,210 / 519
Gallery (img / IDs)	15,913 / 750	17,661 / 702	1,730 / 100	82,161 / 3,060	43,634 / 654	17,661 / 1,110

Table 1. Statistics of the datasets used for training and evaluation. For Occluded-DukeMTMC, total images are computed as Train + Gallery + Query.

	Cam 6			Cam 5		
Source: Model	**mAP**	**R1**	$\mathbf{T_{inf}}$	**mAP**	**R1**	$\mathbf{T_{inf}}$
Cam 12: $\mathcal{L}_S$ + $\text{ViT}_{B/16}$	73.5	92.2	0.10	71.4	87.1	0.10
Cam 12: $\mathcal{L}_{UDA}$ + $\text{ViT}_{B/16}$ (LoRA)	74.1	91.6	0.10	75.1	88.4	0.10
Cam 13: $\mathcal{L}_S$ + $\text{ViT}_{B/16}$	70.2	88.9	0.10	74.7	89.3	0.10
Cam 13: $\mathcal{L}_{UDA}$ + $\text{ViT}_{B/16}$ (LoRA)	71.1	88.9	0.10	76.2	89.8	0.10
Cam 14: $\mathcal{L}_S$ + $\text{ViT}_{B/16}$	74.9	91.4	0.10	79.3	91.2	0.10
Cam 14: $\mathcal{L}_{UDA}$ + $\text{ViT}_{B/16}$ (LoRA)	76.1	91.2	0.10	81.0	91.8	0.10
Cam 15: $\mathcal{L}_S$ + $\text{ViT}_{B/16}$	70.5	90.6	0.10	77.1	90.4	0.10
Cam 15: $\mathcal{L}_{UDA}$ + $\text{ViT}_{B/16}$ (LoRA)	72.1	90.6	0.10	78.8	90.8	0.10
Cams 15 + 14: $\mathcal{L}_S$ + $\text{ViT}_{B/16}$	78.8	92.8	0.26	82.2	92.8	0.26
Cams 15 + 14: Blend + $\mathcal{L}_{UDA}$ + $\text{ViT}_{B/16}$ (LoRA)	78.6	91.8	0.10	82.2	91.9	0.10
Cams 15 + 14: SAGE-ReID	**79.5**	**93.2**	0.26	**83.9**	**92.9**	0.26
Cams 15 + 14 + 13: $\mathcal{L}_S$ + $\text{ViT}_{B/16}$	79.5	93.2	0.32	83.9	93.1	0.32
Cams 15 + 14 + 13: Blend + $\mathcal{L}_{UDA}$ + $\text{ViT}_{B/16}$ (LoRA)	79.8	92.8	0.10	83.5	92.3	0.10
Cams 15 + 14 + 13: SAGE-ReID	**80.2**	**93.4**	0.32	**84.4**	**93.4**	0.32
Cams 15 + 14 + 13 + 12: $\mathcal{L}_S$ + $\text{ViT}_{B/16}$	80.2	93.4	0.36	83.2	92.8	0.36
Cams 15 + 14 + 13 + 12: Blend + $\mathcal{L}_{UDA}$ + $\text{ViT}_{B/16}$ (LoRA)	80.1	93.4	0.10	83.5	92.6	0.10
Cams 15 + 14 + 13 + 12: SAGE-ReID	**80.8**	**93.8**	0.36	**83.9**	**93.1**	0.36

Table 2. Per-camera adaptation on MSMT17. Cams 12–15 are treated as source domains, while Cams 5 and 6 are unseen targets.

	After Adaptation	
Methods	**mAP**	**R-1**
Multi-Source (M + D + CU) → MS:		
$\mathcal{L}_{UDA}$ + $\text{ViT}_{B/16}$ (LoRA) + LoRA Avg	43.2	69.5
$\mathcal{L}_{UDA}$ + $\text{ViT}_{B/16}$ (LoRA) + LoRA MoLE [7]	43.8	69.4
SAGE-ReID	**44.1**	**69.8**
Multi-Source (D + CU + MS) → M:		
$\mathcal{L}_{UDA}$ + $\text{ViT}_{B/16}$ (LoRA) + LoRA Avg	79.5	91.9
$\mathcal{L}_{UDA}$ + $\text{ViT}_{B/16}$ (LoRA) + LoRA MoLE [7]	83.6	93.0
SAGE-ReID	**83.7**	**93.0**
Multi-Source (M + CU + MS) → D:		
$\mathcal{L}_{UDA}$ + $\text{ViT}_{B/16}$ (LoRA) + LoRA Avg	71.7	83.1
$\mathcal{L}_{UDA}$ + $\text{ViT}_{B/16}$ (LoRA) + LoRA MoLE [7]	71.5	83.0
SAGE-ReID	**72.7**	**83.3**

Table 3. Mean average precision (mAP) and Rank-1 (R-1) for three LoRA composition strategies after single-source adaptation.

Stage 1 (single-source target adaptation). SAGE-ReID updates only LoRA modules, yielding 3.6M trainable parameters and 116.6 MB training memory, whereas MSUDA/CDM updates the full backbone (76.8M; 1172.4 MB). This provides a $>10\times$ reduction in training memory. Per-iteration latency is higher for SAGE-ReID (50.5 ms vs. 13.5 ms) and the wall-clock is longer (3 h 48 min vs. 1 h), but the parameter/memory footprint is substantially smaller.

# Experts	Methods	GFLOPs / img	Params (M)
3	LoRA Avg	23.64	90.82
	LoRA MoLE	23.86	187.12
	SAGE-ReID	23.67	101.52
10	LoRA Avg	25.85	99.08
	LoRA MoLE	28.09	1 169.06
	SAGE-ReID	25.95	109.78

Table 4. Computational cost and parameter count of different LoRA-composition schemes. LoRA Avg refers to $\mathcal{L}_{UDA}$ + $\text{ViT}_{B/16}$ (LoRA) + LoRA Avg and LoRA MoLE refers to $\mathcal{L}_{UDA}$ + $\text{ViT}_{B/16}$ (LoRA) + LoRA MoLE in Table 3. "Params" is the *total* number of parameters, including frozen ViT, LoRAs, and the gate module.

Stage 2 (expert merging). The proposed gate trains 10.6M parameters with 168.8 MB training memory and converges in $\approx$1 min. It is much lighter than MoLE (96.2M; 821.8 MB; 148.0 ms/iter) and far lighter and faster than MSUDA (72.4M; 1383.6 MB; 5 h). End-to-end, SAGE-ReID requires $\sim$3.8 h vs. $\sim$6 h for MSUDA.

Scalability and inference cost. Our gate adds $\mathcal{O}(\kappa d)$ parameters per layer and does not grow with the number of sources s, while MoLE scales as $\mathcal{O}(s^2\kappa d)$. In Table 4, with $s = 10$ experts, MoLE's parameters inflate to $\sim$1.17B ($\approx$11$\times$ SAGE-ReID), whereas our gate adds only $+$9M parameters and $+0.1$ GFLOPs; GFLOPs/image remain close across methods (e.g., 25.95 for SAGE-ReID). At inference time, SAGE-ReID performs a single merged forward pass.

E. Analysis of Gating Coefficients

For a target domain instance x_t, the gate outputs a weight $\alpha_{i,\ell,m}(x_t) \in [0, 1]$ for each source domain expert i and layer l. Each Transformer layer contains four modules $m \in \{\text{FC1}, \text{FC2}, \text{QKV}, \text{Proj}\}$. Figure 1 reports a single average per source expert and module by averaging across

	Method	# Total Params (M)	# Trainable Params (M)	Trainable Memory (MB)	ms/iter	Training Time (h/min)
Stage 1	SAGE-ReID	88.4	3.6	116.6	50.5	3 h 48 min
	MSUDA	76.8	76.8	1172.4	13.5	1 h 0 min
	CDM	76.8	76.8	1172.4	–	–
Stage 2	SAGE-ReID	101.5	10.6	168.8	120.1	0 h 1 min
	MoLE	187.1	96.2	821.8	148.0	0 h 1 min
	MSUDA	144.8	72.4	1383.6	65.1	5 h 0 min
	CDM	217.2	108.6	1659.8	–	–

Table 5. Model efficiency: latency is wall-clocked per iteration (batch size 32, image size 256 × 128, fp32). Training time is measured on MSMT17. M = million; MB = megabyte. CDM latency and total training time are not reported due to the unavailability of the code.

target instances and layers:

$$\bar{g}_{i,m} = \frac{1}{n_T\, L} \sum_{\ell=1}^{L} \sum_{t=1}^{n_T} \alpha_{i,\ell,m}(x_t),$$

where n_T is the number of target instances in query set and L is the number of layers that contain module m.

Findings from Fig. 1 (i) *Early modules are selective.* At QKV/FC1, distributions are peaked: for Duke and Market1501 targets, the MSMT17 expert typically receives the largest share, indicating that attention mixing and MLP expansion benefit most from the most diverse source. (ii) *Late modules blend.* At Proj/FC2, weights are more uniform, showing consolidation of cues from multiple sources. (iii) *Large-target behavior.* On MSMT17, coefficients are generally even across experts, with only mild preference at QKV/FC1 and near-uniform mixing at Proj/FC2. Overall, the gate allocates discriminative capacity in early modules and performs stabilizing mixing at later ones, consistent with composing per-module residuals while keeping the backbone and adapters frozen.

F. Analysis of Inter-Domain Synergies

Following the same notation of the previous section, instead of averaging, we concatenate all per-layer, per-instance gating weights and compute correlations on the pooled samples. For each module m, we define $G_m \in \mathbb{R}^{n_T L \times S}$ as:

$$[G_m]_{p,i} \;=\; \alpha_{i,\ell,m}(x^t).$$

We then form the $s \times s$ inter-source correlation matrix C_m (i.e., the Pearson correlation across the $n_T L$ concatenated samples), such that:

$$C_m \;=\; \mathrm{corr}(G_m).$$

Findings (Fig. 2). (i) *Early modules are structured.* At QKV/FC1, correlations form clear "source teams": for Duke as target, {Market, CUHK03} align and contrast with MSMT17; for Market1501, {Duke, CUHK03} align against MSMT17. For MSMT17, the dominant contrast is Market vs. Duke with CUHK03 weakly mediating. (ii) *Late modules blend.* At Proj/FC2, correlations shrink toward 0, indicating more uniform mixing that consolidates cues from multiple sources. Overall, the gate forms target-dependent coalitions in early modules (routing discriminative capacity) and then stabilizes by blending in later modules, evidencing complementarity rather than redundancy among sources.

G. Effect of Occlusion and Source Exclusion

Table 6 evaluates two factors: (i) replacing Duke-DukeMTMC-reID (D) with the more challenging Occluded-DukeMTMC-reID [3] (OC-D), and (ii) a leave-one-out source-exclusion study. We also include IUSTPersonReID (IUST) [4], which introduces a substantial domain shift. First, swapping D for OC-D consistently lowers performance (mAP/R-1), reflecting stronger occlusion and a larger mismatch to the target. Second, the LOO ablation shows that removing any source degrades results, confirming that all sources are complementary. Notably, although OC-D is individually weaker than the other source datasets, excluding it still causes a measurable drop, indicating that it contributes non-redundant information to the merge. In short, OC-D remains a useful component of the multi-source pool.

H. Feature Alignment Analysis

Centered Kernel Alignment (CKA) [1] measures representational similarity (1.0 = identical) between two models' activations on the same inputs. We apply the same averaging procedure as in Sec. E, now on feature maps rather than gating weights. In Table 7 we observe consistently high CKA between each single expert (M, CU, MS) and the gated mixture (M + CU + MS) on QKV/FC1/Proj (typically 0.90–0.96), with lower scores at FC2 (0.63–0.77). This pattern indicates that the gating function largely preserves each expert's representation in the QKV/FC1/Proj layers, while FC2 performs task-specific reshaping. The per-row means (0.84, 0.87, 0.89) further show that alignment holds across all experts, supporting that the gate mixes via soft selection

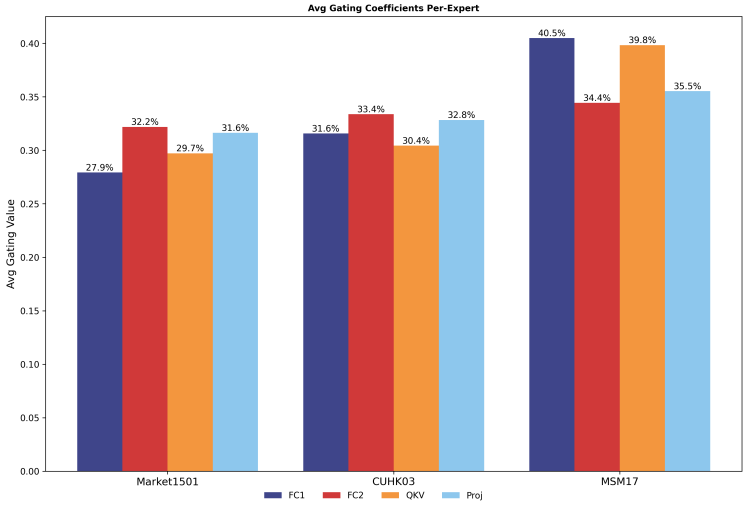
Methods	D		IUST		M	
	mAP	**R1**	**mAP**	**R1**	**mAP**	**R1**
$\mathcal{L}_{\text{UDA}}$ + $\text{ViT}_{\text{B/16}}$ (LoRA)	41.0	65.8	40.1	65.0	39.0	63.5
Multi-Source (Duke + IUST + Market1501):						
SAGE-ReID (LOO = D)			41.7	68.1		
SAGE-ReID (LOO = IUST)			41.6	67.3		
SAGE-ReID (LOO = M)			43.5	69.5		
Ties			34.8	61.8		
SAGE-ReID			**44.5**	**70.2**		
Methods	**OC-D**		**IUST**		**M**	
	mAP	**R1**	**mAP**	**R1**	**mAP**	**R1**
$\mathcal{L}_{\text{UDA}}$ + $\text{ViT}_{\text{B/16}}$ (LoRA) (Source Free)	34.2	61.7	40.1	65.0	39.0	63.5
Multi-Source (OC-Duke + IUST + Market1501):						
SAGE-ReID (LOO = OC-D)			41.7	68.1		
SAGE-ReID (LOO = IUST)			39.5	65.6		
SAGE-ReID (LOO = M)			37.7	65.3		
SAGE-ReID			**42.1**	**68.3**		

Table 6. Impact of removing one source dataset contribution and adding Ocluded-DukeMTMC (OC-D). LOO stands for leave-one-out

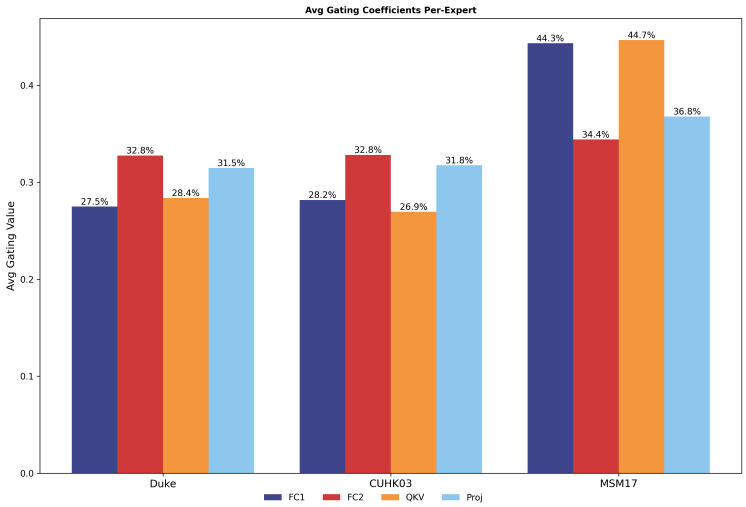
Target	Single / Combo	QKV	Proj	FC1	FC2	Mean
D	M ↔ (M+CU+MS)	0.96	0.91	0.91	0.71	0.84
	CU ↔ (M+CU+MS)	0.92	0.90	0.92	0.63	0.87
	MS ↔ (M+CU+MS)	0.95	0.92	0.93	0.77	0.89

Table 7. **Layer-wise CKA across experts/combinations.** For *Target Duke (D)*, linear CKA is computed by first averaging feature maps over test samples, then averaging across the 12 Transformer blocks for each layer (QKV, Proj, FC1, FC2). **Mean** is the average across the four layers.

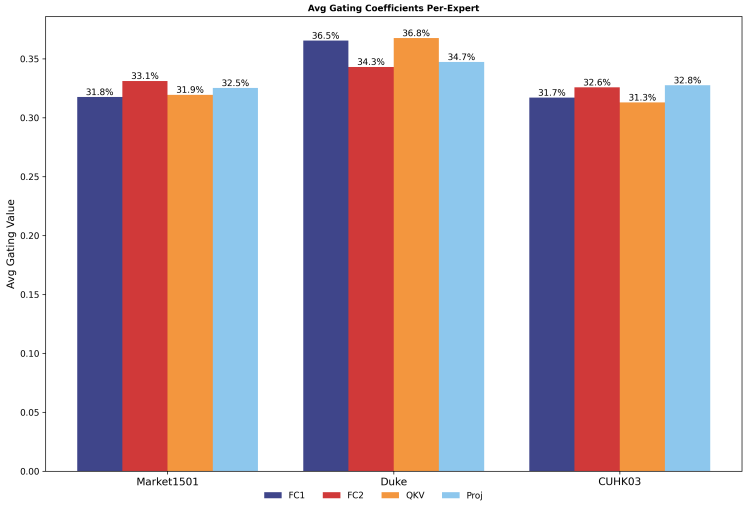
rather than creating a new, conflicting representation.



(a) Average gating coefficient (Target: DukeMTMC-reID)

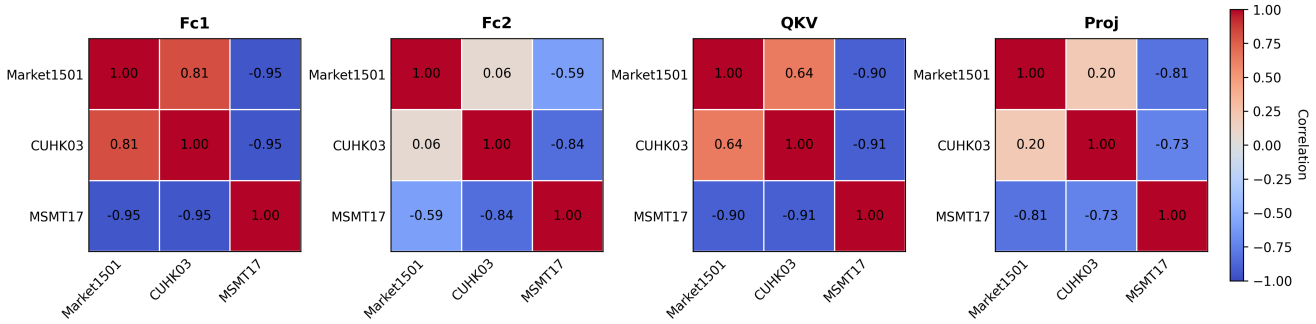


(b) Average Average gating coefficient (Target: Market-1501)

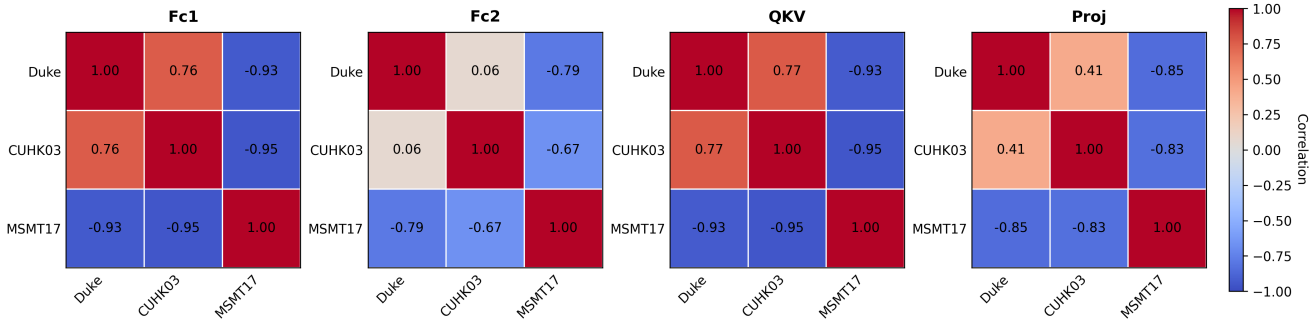


(c) Average Average gating coefficient (Target: MSMT17)

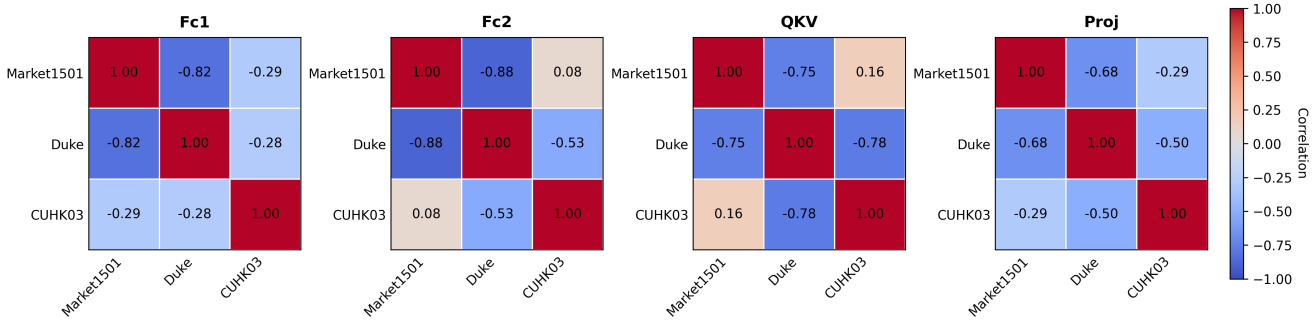
Figure 1. Average gating coefficient $\bar{g}_{i,m}$ per target domain.



(a) Target : DukeMTMC-reID



(b) Target : Market1501



(c) Target : MSMT17

Figure 2. Overall gating-correlation visualization across layers.

References

[1] Simon Kornblith, Mohammad Norouzi, Honglak Lee, and Geoffrey Hinton. Similarity of neural network representations revisited. In *International conference on machine learning*, pages 3519–3529. PMlR, 2019. 3

[2] 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. 1, 2

[3] Jiaxu Miao, Yu Wu, Ping Liu, Yuhang Ding, and Yi Yang. Pose-guided feature alignment for occluded person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019. 1, 2, 3

[4] Alireza Sedighi Moghaddam, Fatemeh Anvari, Mohammadjavad Mirshekari Haghighi, Mohammadali Fakhari, and Mohammad Reza Mohammadi. A culturally-aware benchmark for person re-identification in modest attire. *Engineering Applications of Artificial Intelligence*, 158:111494, 2025. 1, 2, 3

[5] 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. 1, 2

[6] 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

[7] Xun Wu, Shaohan Huang, and Furu Wei. Mixture of lora experts. In *International Conference on Learning Representations (ICLR)*, 2024. 1, 2

[8] 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. 1, 2