Supplementary Materials: Calibrated Domain-Invariant Learning for Highly Generalizable Large Scale Re-Identification

Ye Yuan[1], Wuyang Chen[1], Tianlong Chen[1], Yang Yang[2], Zhou Ren[3], Zhangyang Wang[1] and Gang Hua[3]
{ye.yuan, wuyang.chen, wiwjp619, atlaswang}@tamu.edu,
yang.yang2@walmart.com, {renzhou200622, ganghua}@gmail.com
[1]Department of Computer Science and Engineering, Texas A&M University
[2]Walmart Technology, [3]Wormpex AI Research
https://github.com/TAMU-VITA/ADIN

1. Training Details

To start our adversarial domain-invariant learning (ADIN), we initialize f_E, f_I and f_N with pretrained weights. We first jointly pretrain the feature extractor f_E and identity classifier f_I, and pretrain f_N by fixing f_E. In ADIN we alternative between optimizing two sub-problems:

$$\min_{f_E, f_N} L_{adv}(f_N(f_E(X))), \min_{f_E, f_I} L_I(f_I(f_E(X)), Y_I). \quad (1)$$

In each alternating round, we optimize the first objective until the validation error of identity classification reducing below a pre-set $threshold_{I-target}$. We then switch to optimizing the second objective, meanwhile monitoring the resulting changes on the identity classification validation error (since f_E is updated): if it drops below another pre-set $threshold_{I-trigger}$, we will switch back to the first object and start the next round of alternations.

Besides, to meet the "hidden constraint" and avoid too weak f_N during training, we will periodically replace the current weights in f_I and f_N with random weights, and re-train them on $f_E(X)$ by:

$$\min_{f_N} L_N(f_N(f_E(X)), Y_I), \min_{f_I} L_I(f_I(f_E(X)), Y_I). \quad (2)$$

with f_E being unaffected and fixed. We then re-start alternations from the new "pre-trained" initialization. This empirical trick is found to strengthen the generalization of f_E, potentially because it gets rid of some trivial local minima. The training procedure is summarized in Algorithm 1.

2. Network Structure of Dual-branch Backbone

For our dual-branch backbone, the first four blocks share the same design as in ResNet50. After the forth block, the network was split into a global and a local branch. In the global branch, the feature passes a global average-pooling and then is fed into the classifier. In the local branch, feature is horizontally partitioned into two equal parts, where each part adopts a separate global average-pooling layer and classifier. During inference the outputs from two branches are concatenated together as the final feature for image retrieval.

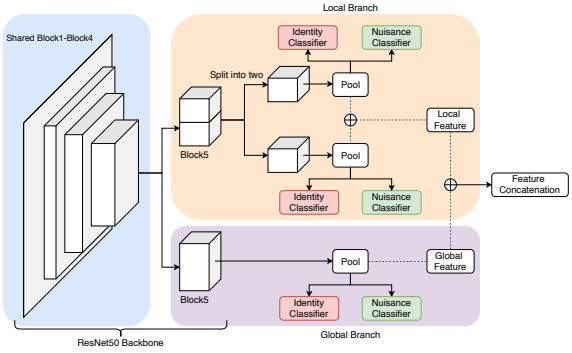

Figure 1: Overview of the dual-branch backbone.

3. Additional Experiment Results

3.1. Ablation study of the adversarial loss L_{adv}

Table 1 displays a full step-by-step comparison for different adversarial loss L_{adv}, together with state-of-the-art ReID models and domain adaptation methods. We use the direct transfer performance from DukeMTMC-ReID (source domain) to Market1501 (target domain) as the indicator.

Compared with the baseline and Reverse Gradient, our proposed calibrated negative loss (CaNE) contributes to both stable adversarial training (robustness against gradient vanishing/explosion and classifier's loss magnitude), and CaNE

Algorithm 1 The training strategy of ADIN

Given pre-trained feature extractor f_E, identity classifier f_I and nuisances classifier f_N
$val_I, val_N \leftarrow$ identity classifier validation accuracy, nuisances classifier validation accuracy.
for number of training epoches **do**
  if $val_I < threshold_{I-trigger}$ **then** ▷ Avoid weak identity recognition performance
    while $val_I \leq threshold_{I-target}$ **do**
      for number of batches **do**
        Sample minibatch of m examples $\{X_1, ..., X_m\}$
        Jointly update the f_E and the f_I by descending its stochastic gradient with loss L_I
      end for
      $val_I \leftarrow$ identity classifier validation accuracy.
    end while
  else if $val_N > threshold_N$ **then** ▷ Suppress nuisance discriminator performance
    Feed all training examples $\{X_1, ..., X_n\}$ into the model
    Jointly update f_E and f_N by descending its gradient with the adversarial loss L_{adv}
  else ▷ Further boost identity recognition performance
    for number of batches **do**
      Sample minibatch of m examples $\{X_1, ..., X_m\}$
      Jointly update f_E and f_I by descending its stochastic gradient with loss L_I
    end for
  end if
  Re-initialize f_I, f_N ▷ Empirically restart f_I f_N every iteration to avoid overfitting extracted features
  Train f_I, f_N by descending its stochastic gradient with classification loss L_I, f_N correspondingly
  $val_I, val_N \leftarrow$ identity classifier validation accuracy, nuisances classifier validation accuracy.
end for

Table 1: Ablation study of adversarial loss: single-dataset performance on DukeMTMC-ReID [1, 2] and direct transfer of DukeMTMC-ReID → Market1501 [3]. * indicates method using images from both source and target domain.

	DukeMTMC-ReID				DukeMTMC-ReID → Market1501			
	top1	top5	top10	mAP	top1	top5	top10	mAP
Spatial-Attention[4]	83.6	91.8	**94.3**	70.1	49.8	67.4	73.8	24.2
PCB[5]	**84.2**	91.7	93.4	69.7	53.9	70.5	76.8	26.5
RPP[5]	84.1	**92.5**	**94.3**	**71.3**	53.1	71.3	76.6	25.7
MGN[6]	55.5	70.2	76.8	35.1	48.7	66.9	73.7	25.1
CycleGAN [7] *	-	-	-	-	48.1	66.2	72.7	20.7
SPGAN [7] *	-	-	-	-	58.1	76.0	82.7	26.9
HHL [8] *	-	-	-	-	**62.2**	**78.8**	**84.0**	**31.4**
ResNet50 (baseline)	76.6	88.0	91.5	58.1	46.8	63.5	70.3	19.0
ResNet50 + Reverse Gradient	Unable to converge							
ResNet50 + NE	76.7	87.9	91.2	57.5	48.8	66.2	72.7	20.4
ResNet50 + CaNE	74.8	86.0	89.1	54.9	51.7	68.6	76.0	22.1
Dual-branch	82.1	91.3	93.0	66.8	54.8	71.7	77.6	25.9
Dual-branch + Reverse Gradient	Unable to converge							
Dual-branch + NE	81.8	90.8	93.3	66.4	55.9	72.5	78.6	26.5
Dual-branch + CaNE	80.7	90.0	91.9	63.8	57.2	73.0	80.0	27.4

is also attentive adversarial effects w.r.t. different nuisances frequencies (attention to sampling imbalance in nuisances), outperforming the negative entropy (NE) loss. Our ADIN framework with CaNE loss not only improve the generalizability of basic backbones like ResNet50, but also boosts more powerful ones like Dual-branch, indicating that ADIN is a general effective adversarial learning framework towards generalizability. Impressively, the Dual-branch backbone equipped with ADIN and CaNE outperforms even domain adaptation methods (HHL [8], SPGAN [7]), where extra target source images and fine-tuning/retraining are required but not in our case. We observed that the ADIN causes a bit decrease in single-dataset accuracy. This is because models with our adversarial training no longer overfitting current small-scale dataset.

3.2. Single-dataset and Direct Transfer Performance without Retraining or Adaption

Here we include the full detailed results. We evaluate three direct transfer cases, two on person ReID: MSMT17 → DukeMTMC-ReID, MSMT-17 → Market1501; and one on vehicle ReID: VeRi-776 [10] → VehicleID[11]. As comparison baselines, we train the same dual-branch backbones (without any adversarial learning) on the source datasets, and test their direct transfer performance too. We further compare with existing strong competitors: one person ReID, a CycleGAN baseline as adopted by [7] for learning an unsupervised data-level domain mapping, two state-of-the-art domain adaptation methods SPGAN [7] and HHL [8], the latter reporting the current best transfer results between DukeMTMC-ReID and Market1501; on vehicle ReID the DAVR [13] which reported the current best transfer results from VeRi-776 [10] to VehicleID [11]. Note that CycleGAN, SPGAN and HHL all need to use (unlabeled) target domain data and perform extra (re-)training for the source domain models: the comparisons are thus apparently to our competitors' advantage.

As can be seen from Tables 2 and 3, while baselines without adversarial learning fail to transfer well as expected, ADIN demonstrates highly impressive results on all three transfer cases. In particular, by training on MSMT17 and

Table 2: Single-dataset performance on MSMT17 [9] and direct transfer of MSMT17 → DukeMTMC-ReID and MSMT17 → Market1501. * indicates method using images from both source and target domain.

	MSMT17				MSMT17 → DukeMTMC-ReID				MSMT17 → Market1501			
	top1	top5	top10	mAP	top1	top5	top10	mAP	top1	top5	top10	mAP
Spatial-Attention[4]	68.7	81.5	85.7	41.8	52.2	68.1	74.1	32.9	49.7	68.9	75.5	25.1
PCB[5]	68.6	81.3	85.8	41.8	54.4	69.6	75.4	34.6	52.7	71.3	77.5	26.7
RPP[5]	73.1	**84.5**	**88.1**	**46.4**	56.7	71.4	76.8	36.7	50.2	70.7	77.5	26.3
MGN[6]	71.7	83.3	87.1	45.7	55.5	70.2	76.8	35.1	48.7	66.9	73.7	25.1
CycleGAN [7] *	-	-	-	-	48.1	66.2	72.7	20.7	38.5	54.6	60.8	19.9
SPGAN [7] *	-	-	-	-	58.1	76.0	82.7	26.9	46.9	62.6	68.5	26.4
HHL [8] *	-	-	-	-	**62.2**	**78.8**	**84.0**	31.4	46.9	61.0	66.7	27.2
ResNet50 (baseline)	63.2	76.7	81.6	31.9	49.7	65.7	71.0	28.2	47.7	64.3	71.5	21.2
ResNet50 + CaNE	62.7	76.5	81.0	30.8	52.6	67.9	73.2	30.4	50.1	66.4	73.5	22.5
Dual-branch	**73.5**	84.3	**88.1**	45.1	59.5	73.5	78.8	38.4	57.8	73.9	80.6	29.4
Dual-branch + CaNE	73.3	**84.5**	87.8	43.0	60.7	74.7	79.5	**39.1**	**59.1**	**75.4**	**81.7**	**30.3**

Table 3: Direct transfer performance between VeRi-776 [10] and VehicleID[11]. * indicates method using images from both source and target domain.

Method	Test size = 800				Test size = 1600				Test size = 2400				Test size = 3200			
	top1	top5	top10	mAP	top1	top5	top10	mAP	top1	top5	top10	mAP	top1	top5	top10	mAP
RAM[12]	34.0	53.7	61.9	43.3	30.5	49.5	56.4	39.5	26.6	43.1	51.1	34.8	24.5	40.3	48.2	32.4
Spatial-Attention[4]	42.4	61.1	69.3	51.5	39.5	57.2	64.2	47.9	36.0	52.7	60.2	44.2	33.7	49.6	56.9	41.6
PCB[5]	43.7	63.1	70.8	53.0	41.3	58.8	65.2	49.7	37.5	53.9	61.5	45.6	35.4	51.4	58.3	43.2
RPP[5]	44.5	63.1	70.1	53.2	40.6	58.4	65.2	49.1	37.0	54.1	61.8	45.3	35.0	51.1	58.4	42.9
MGN[6]	44.6	70.5	**79.7**	56.5	39.9	62.4	72.3	50.6	36.2	58.1	68.0	46.6	32.7	53.1	63.0	42.7
DAVR[13]*	49.5	68.7	–	54.0	45.2	64.0	–	49.7	40.7	59.0	–	45.2	38.7	55.9	–	42.9
ResNet50 (baseline)	44.7	62.5	69.0	48.9	42.3	58.5	64.5	46.2	38.1	54.8	61.7	42.1	36.1	52.2	58.9	39.9
ResNet50 + CaNE	46.0	63.6	69.9	50.2	43.3	59.7	65.6	47.2	38.8	56.0	63.1	42.9	37.0	53.4	60.0	40.9
Dual-branch	51.2	70.3	77.8	55.7	47.3	65.3	72.2	51.6	44.2	62.1	69.7	48.4	41.2	57.9	64.9	45.3
Dual-branch + CaNE	**52.9**	**72.1**	79.4	**57.4**	**48.7**	**67.3**	**74.0**	**53.1**	**45.0**	**64.0**	**71.3**	**49.5**	**42.1**	**59.5**	**66.6**	**46.3**

Table 4: Single-dataset performance on VeRi-776 [10].

	VeRI-776 → VeRi-776			
	top1	top5	top10	mAP
MAA[14]	88.0	94.6	-	58.1
QD-DLF[15]	88.5	94.5	-	61.8
RAM[12]	88.6	94.0	-	61.5
GAN+LSRO[16]	87.7	93.9	-	58.2
BS[17]	90.2	96.4	-	67.6
Spatial-Attention[4]	93.4	96.8	98.2	70.5
PCB[5]	92.4	96.7	98.3	69.4
RPP[5]	93.5	96.9	97.9	69.7
MGN[6]	94.9	97.0	97.5	78.7
ResNet50 (baseline)	91.1	95.5	97.3	60.2
ResNet50 + CaNE	90.7	95.41	96.9	59.5
Dual-branch (baseline)	**95.0**	**97.9**	**98.9**	**73.5**
Dual-branch + CaNE	93.1	96.5	97.9	69.4

directly transferring, ADIN not only surpasses the direct transfer results from other methods but also outperforms state-of-the-art ReID domain adaption models (HHL [8], SPGAN[7], DAVR [13]) , while costing literally no hassle such as (re-)training.

However, in contrast to our ADIN, we find other (single-dataset) top-performers generalize very poorly to unseen domains, indicating the misaligned goal between overfitting small-scale single dataset and generalizing to large-scale unseen scenarios in real life. We believe the effective direct transfer is the right choice for evaluating and promoting larger-scale ReID practice, and hope our proposals and arguments could invoke more discussions in the community. Again, we observed that the ADIN causes a bit decrease in single-dataset accuracy, since models with our adversarial training no longer overfitting current small-scale dataset.

References

[1] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. In *The European Conference on Computer Vision (ECCV)*, September 2016.

[2] Zhedong Zheng, Liang Zheng, and Yi Yang. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[3] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *The IEEE International Conference on Computer Vision (ICCV)*, December 2015.

[4] Haoran Wang, Yue Fan, Zexin Wang, Licheng Jiao, and Bernt Schiele. Parameter-free spatial attention network for person re-identification. *arXiv preprint arXiv:1811.12150*, 2018.

[5] Yifan Sun, Liang Zheng, Yi Yang, Qi Tian, and Shengjin Wang. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In *The European Conference on Computer Vision (ECCV)*, September 2018.

[6] Guanshuo Wang, Yufeng Yuan, Xiong Chen, Jiwei Li, and Xi Zhou. Learning discriminative features with multiple

granularities for person re-identification. In *2018 ACM Multimedia Conference on Multimedia Conference*, pages 274–282. ACM, 2018.

[7] 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 *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[8] Zhun Zhong, Liang Zheng, Shaozi Li, and Yi Yang. Generalizing a person retrieval model hetero- and homogeneously. In *The European Conference on Computer Vision (ECCV)*, September 2018.

[9] Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. Person transfer gan to bridge domain gap for person re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[10] Xinchen Liu, Wu Liu, Huadong Ma, and Huiyuan Fu. Large-scale vehicle re-identification in urban surveillance videos. In *2016 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2016.

[11] Hongye Liu, Yonghong Tian, Yaowei Yang, Lu Pang, and Tiejun Huang. Deep relative distance learning: Tell the difference between similar vehicles. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2167–2175, 2016.

[12] Xiaobin Liu, Shiliang Zhang, Qingming Huang, and Wen Gao. Ram: a region-aware deep model for vehicle re-identification. In *2018 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2018.

[13] Jinjia Peng, Huibing Wang, and Xianping Fu. Cross domain knowledge learning with dual-branch adversarial network for vehicle re-identification. *arXiv preprint arXiv:1905.00006*, 2019.

[14] Na Jiang, Yue Xu, Zhong Zhou, and Wei Wu. Multi-attribute driven vehicle re-identification with spatial-temporal re-ranking. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 858–862. IEEE, 2018.

[15] Jianqing Zhu, Huanqiang Zeng, Jingchang Huang, Shengcai Liao, Zhen Lei, Canhui Cai, and Lixin Zheng. Vehicle re-identification using quadruple directional deep learning features. *IEEE Transactions on Intelligent Transportation Systems*, 2019.

[16] Fangyu Wu, Shiyang Yan, Jeremy S Smith, and Bailing Zhang. Joint semi-supervised learning and re-ranking for vehicle re-identification. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 278–283. IEEE, 2018.

[17] Ratnesh Kumar, Edwin Weill, Farzin Aghdasi, and Parthsarathy Sriram. Vehicle re-identification: an efficient baseline using triplet embedding. *arXiv preprint arXiv:1901.01015*, 2019.