FastReID: A Pytorch Toolbox for General Instance Re-identification

Lingxiao He*†, Xingyu Liao*, Wu Liu†, Xinchen Liu, Peng Cheng and Tao Mei
JD AI Research
{helingxiao3, liaoxingyu5, liuwu1, liuxinchen1, chengpeng8, tmei}@jd.com

Abstract

General Instance Re-identification is a very important task in the computer vision, which can be widely used in many practical applications, such as person/vehicle re-identification, face recognition, wildlife protection, commodity tracing, and snapshop, etc.. To meet the increasing application demand for general instance re-identification, we present FastReID as a widely used software system in JD AI Research. In FastReID, highly modular and extensible design makes it easy for the researcher to achieve new research ideas. Friendly manageable system configuration and engineering deployment functions allow practitioners to quickly deploy models into productions. We have implemented some state-of-the-art projects, including person re-id, partial re-id, cross-domain re-id and vehicle re-id, and plan to release these pre-trained models on multiple benchmark datasets. FastReID is by far the most general and high-performance toolbox that supports single and multiple GPU servers, you can reproduce our project results very easily and are very welcome to use it, the code and models are available at https: `https://github.com/JDAI-CV/fast-reid`.

1. Introduction

General instance re-identification (re-id), as an instance-centric AI technique, aiming at finding a certain person/vehicle/face/object of interest in a large amount of videos. It facilitates various applications that require painful and boring video watching, including searching for video shots related to an actor of interest from TV series, a lost child in a shopping mall from camera videos, a suspect vehicle from a city surveillance system. Moreover, the General instance re-identification technique is also used for snapshop in e-commerce platforms, commodity tracing in merchandise security and wildlife protection. Many researchers realize a task based on open source code, less extensible and reusable modification make it difficult to reproduce the results. Besides, there often exists a gap between academic research and practical applications, which makes it difficult for academic research techniques to be quickly transferred to productions.

To accelerate progress in the community of general instance re-identification including researchers and practitioners in academia and industry, we now release a unified instance re-id library named FastReID. We have introduced a stronger modular, extensible design that allows researchers and practitioners easily to plug their oven designed module without repeatedly rewriting codebase, into a re-id system for further rapidly moving research ideas into production models. Manageable system configuration makes it more flexible and extensible, which is easily extended to a range of tasks, such as general image retrieve and face recognition, etc. Based on FastReID, we provide many state-of-the-art pre-trained models on multiple tasks about person re-id, cross-domain person re-id, partial person re-id and vehicle re-id, and in the future we will release face recognition and object retrieval models. Besides, we hope that the library can provide a fair comparison between different approaches.

Recently, FastReID has become one of the widely used open-source library in JD AI Research. We will continually refine it and add new features to it. We warmly welcome individuals, labs to use our open-source library and look forward to cooperating with you to jointly accelerate AI Research and achieve technological breakthroughs.

2. Highlight of FastReID

FastReID provides a complete toolkit for training, evaluation, finetuning and model deployment. Besides, FastReID provides strong baselines that are capable of achieving state-of-the-art performance on multiple tasks.

Modular and extensible design. In FastReID, we introduce a modular design that allows users to plug custom-designed modules into almost any part of the re-identification system. Therefore, many new researchers and

*Authors contributed equally; †Corresponding author

Acknowledgements Thanks to Kecheng Zheng (*zkcys001@mail.ustc.edu.cn*), Jinkai Zheng (*zhengjinkai3@hdu.edu.cn*) and Boqiang Xu (*boqiang.xu@cripac.ia.ac.cn*), partial works were done when they interned at JD AI Research.

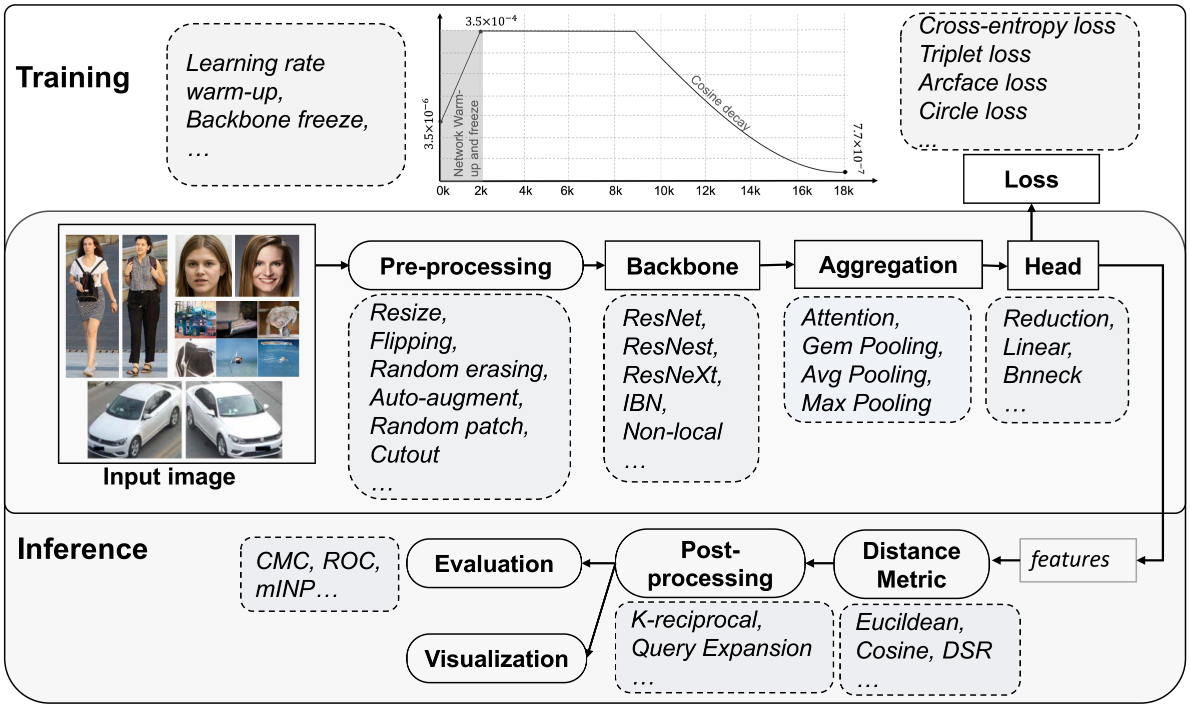

Figure 1. The Pipeline of FastReID library.

practitioners can quickly implement their ideas without rewriting hundreds of thousands of lines of code.

Manageable system configuration. FastReID implemented in PyTorch is able to provide fast training on multi-GPU servers. Model definitions, training and testing are written as YAML files. FastReID supports many optional components, such as backbone, head aggregation layer and loss function, and training strategy.

Richer evaluation system. At present, many researchers only provide a single CMC evaluation index. To meet the requirement of model deployment in practical scenarios, FastReID provides more abundant evaluation indexes, e.g., ROC and mINP, which can better reflect the performance of models.

Engineering deployment. Too deep model is hard to deploy in edge computing hardware and AI chips due to time-consuming inference and unrealizable layers. FastReID implements the knowledge distillation module to obtain a more precise and efficient lightweight model. Also, FastReID provides a conversion tool, e.g., PyTorch→Caffe and PyTorch→TensorRT to achieve fast model deployment.

State-of-the-art pre-trained models. FastReID provides state-of-the-art inference models including person re-id, partial re-id, cross-domain re-id and vehicle re-id. We plan to release these pre-trained models. FastReID is very easy to extend to general object retrieval and face recognition. We hope that a common software advanced new ideas to applications.

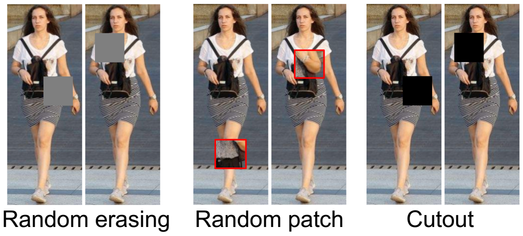
Random erasing Random patch Cutout

Figure 2. Image pre-processing.

3. Architecture of FastReID

In this section, we elaborate on the pipeline of FastReID as shown in Fig. 1. The whole pipeline consists of four modules: image pre-processing, backbone, aggregation and head, we will introduce them in detail one by one.

3.1. Image Pre-processing

The collected images are of different sizes, we first resize the images to fixed-size images. And images can be packaged into batches and then input into the network. To obtain a more robust model, **flipping** as a data augmentation method by mirroring the source images to make data more diverse. **Random erasing**, **Random patch** [1] and **Cutout** [2] are also augmentation methods that randomly selects a rectangle region in an image and erases its pixels with random values, another image patch and zero values, making the model effectively reduce the risk of over-fitting and robust to occlusion. **Auto-augment** is based on automl technique to achieve effective data augmentation for improving the robustness of feature representation. It uses an auto search algorithm to find the fusion policy about multiple image processing functions such as translation, rotation and shearing.

3.2. Backbone

Backbone is the network that infers an image to feature maps, such as a ResNet without the last average pooling layer. FastReID achieves three different backbones including ResNet [3], ResNeXt [4] and ResNeSt [5]. We also add attention-like non-local [6] module and instance batch normalization (IBN) [7] module into backbones to learn more robust feature.

3.3. Aggregation

The aggregation layer aims to aggregate feature maps generated by the backbone into a global feature. We will introduce four aggregation methods: max pooling, average pooling, GeM pooling and attention pooling. The pooling layer takes $\mathbf{X} \in \mathbb{R}^{W \times H \times C}$ as input and produces a vector $\mathbf{f} \in \mathbb{R}^{1 \times 1 \times C}$ as an output of the pooling process, where W, H, C respectively represent the width, the height and the channel of the feature maps. The global vector $\mathbf{f} = [f_1, ..., f_c, ..., f_C]$ in the case of the max pooling, average pooling, GeM pooling and attention pooling of are respectively given by

$$\text{Max Pooling}: f_c = \max_{x \in \mathbf{X}_c} x \tag{1}$$

$$\text{Avg Pooling}: f_c = \frac{1}{|\mathbf{X}_c|} \sum_{x \in \mathbf{X}_c} x \tag{2}$$

$$\text{Gem Pooling}: f_c = (\frac{1}{|\mathbf{X}_c|} \sum_{x \in \mathbf{X}_c} x^{\alpha})^{\frac{1}{\alpha}} \tag{3}$$

$$\text{Attention Pooling}: f_c = \frac{1}{|\mathbf{X}_c * \mathbf{W}_c|} \sum_{x \in \mathbf{X}_c, w \in \mathbf{W}_c} w * x \tag{4}$$

where α is control coefficient and $\mathbf{W}_c$ are the softmax attention weights.

3.4. Head

Head is the part of addressing the global vector generated by aggregation module, including batch normalization (BN) head, Linear head and Reduction head. Three types of the head are shown in Fig. 3, the linear head only contains a decision layer, the BN head contains a bn layer and a decision layer and the reduction head contains conv+bn+relu+dropout operation, a reduction layer and a decision layer.

Batch Normalization [8] is used to solve internal covariate shift because it is very difficult to train models with saturating non-linearities. Given a batch of feature vector $\mathbf{f} \in \mathbb{R}^{m \times C}$ (m is the sample number in a batch), then the bn feature vector $\mathbf{f}_{bn} \in \mathbb{R}^{m \times C}$ can be computed as

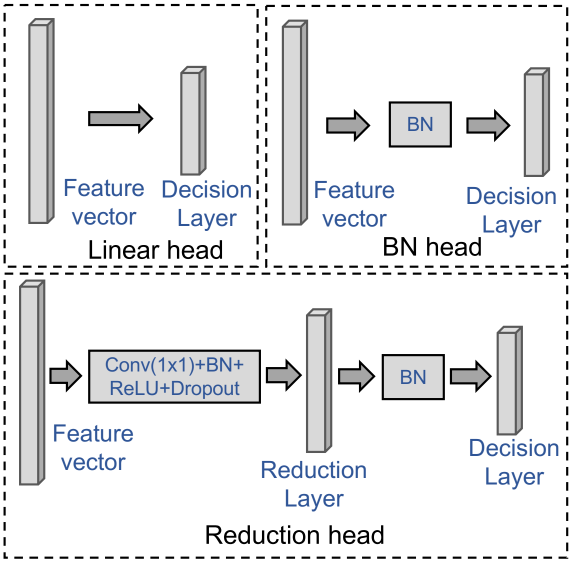

Figure 3. Different heads that implemented in FastReID

$$\begin{cases} \mu = \frac{1}{m} \sum_{i=1}^{m} \mathbf{f}_i, \\ \sigma^2 = \frac{1}{m} \sum_{i=1}^{m} (\mathbf{f}_i - \mu)^2, \\ \mathbf{f}_{bn} = \gamma \cdot \frac{\mathbf{f} - \mu}{\sqrt{\sigma^2 + \epsilon}} + \beta \end{cases} \tag{5}$$

where γ and β are trainable scale and shift parameters, and ϵ is a constant added to the mini-batch variance for numerical stability.

Reduction layer is aiming to make the high-dimensional feature become the low-dimensional feature, i.e., 2048-dim→512-dim.

Decision layer outputs the probability of different categories to distinguish different categories for the following model training.

4. Training

4.1. Loss Function

Four different loss functions are implemented in FastReID.

Cross-entropy loss is usually used for one-of-many classification, which can be defined as

$$\mathcal{L}_{ce} = \sum_{i=1}^{C} y_i \log \hat{y_i} + (1 - y_i) \log(1 - \hat{y_i}), \tag{6}$$

where $\hat{y_i} = \frac{e^{\mathbf{W}_i^T \mathbf{f}}}{\sum_{i=1}^{C} e^{\mathbf{W}_i^T \mathbf{f}}}$. Cross-entropy loss makes the predicted logit values to approximate to the ground truth. It encourages the differences between the largest logit and all others to become large, and this, combined with the bounded gradient reduces the ability of the model to adapt, resulting in a model too confident about its predictions. This, in turn, can lead to over-fitting. To build a robust model that can generalize well, **Label Smoothing** is proposed by Google Brain to address the problem. It encourages the activations of the penultimate layer to be close to the template of the correct class and equally distant to the templates of the incorrect classes. So the ground truth label $\mathbf{y}$ in cross-entropy loss can be defined as $y_i(j = c) = 1 - \delta$ and $y_i(j \neq c) = \frac{\delta}{C-1}$.

Arcface loss [9] maps cartesian coordinates to spherical coordinates. It transforms the logit as $\mathbf{W}_i^T \mathbf{f} = \|\mathbf{W}_i\| \|\mathbf{f}\| \cos\theta_i$, where θ_i is the angle between the weight $\mathbf{W}_i$ and the feature $\mathbf{f}$. It fixes the individual weight $\|\mathbf{W}_i\| = 1$ by l_2 normalisation and also fixes the embedding feature $\mathbf{f}$ by l2 normalisation and re-scale it to s, so $\hat{y_i} = \frac{e^{s\cos\theta_i}}{\sum_{i=1}^{C} e^{s\cos\theta_i}}$. To simultaneously enhace the intra-class compactness and inter-class discrepancy, Arcface adds an additive angular margin penalty m in the intra-class measure. So $\hat{y_i}$ can rewritten as $\hat{y_i} = \frac{e^{s\cos(\theta_i + m)}}{e^{s\cos(\theta_i + m)} + \sum_{i=1, i\neq c}^{C-1} e^{s\cos\theta_i}}$.

Circle loss. The derivation process of circle loss is not described here in detail, it can refer to [10].

Triplet loss ensures that an image of a specific person is closer to all other images of the same person than to any images of other persons, which wants to make an image x_i^a (anchor) of a specific person closer to all other images x_i^p (positive) of the same person than to any image x_i^n (negative) of any other person in the image embedding space. Thus, we want $D(x_i^a, x_i^p) + m < D(x_i^a, x_i^n)$, where $D(:,:)$ is measure distance between a pair of person images. Then the *Triplet Loss* with N samples is defined as $\sum_{i=1}^{N}[m + D(g_i^a, g_i^p) - D(g_i^a, g_i^n)]$, where m is a margin that is enforced between a pair of positive and negative.

4.2. Training Strategy

Fig. 4 shows the train strategy that contains many tricks including learning rate for different iteration, network warm-up and freeze.

Learning rate warm-up helps to slow down the premature over-fitting of the mini-batch in the initial stage of the model training. Also, it helps to maintain the stability of the deep layer of the model. Therefore, we will give a very small learning rate, e.g., 3.5×10^{-5} in the initial training and then gradually increase it during the 2k iterations. After that, the learning rate remains at 3.5^{-4} between 2k iterations and 9k iterations. Then, the learning rate starts from 3.5×10^{-4} and decays to 7.7×10^{-7} at cosine rule after 9k iterations, the training is finished at 18k iterations.

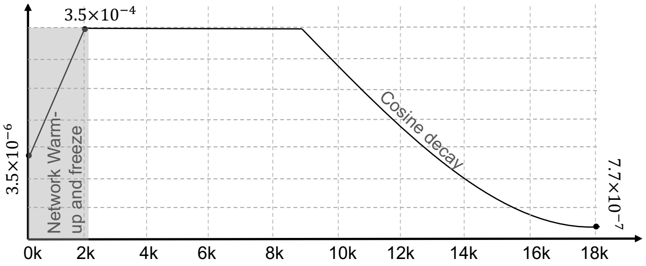

Figure 4. Learning rate curve as a function of the number of iteration

Backbone freeze. To re-train a classification network to meet the requirement of our tasks, we use the collected data from the tasks to fine-tune on the ImageNet pre-trained model. Generally, we add a classifier that collected the network such as ResNet, and the classifier parameters are randomly initialized. To initialize the parameters of the classifier better, we only train the classifier parameters while freezing the network parameters without updating at the beginning of the training (2k iterations). After 2k iterations, we will free the network parameter for end-to-end training.

5. Testing

5.1. Distance Metric.

Eucildean and cosine measure are implemented in FastReID. And we also implement a local matching method: deep spatial reconstruction (DSR).

Deep spatial reconstruction. Suppose there is a pair of person images x and y. Denote the spatial features map from backbone as $\mathbf{x}$ for x with dimension dimension $w_x \times h_x \times d$, and $\mathbf{y}$ for y with dimension $w_y \times h_y \times d$. The total N spatial features from N locations are aggregated into a matrix $\mathbf{X} = [\mathbf{x}_n]_{n=1}^{N} \in \mathbb{R}^{d\times N}$, where $N = w_x \times h_x$. Likewise, we construct the gallery feature matrix $\mathbf{Y} = \{\mathbf{y}_m\}_{m=1}^{M} \in \mathbb{R}^{d\times M}$, $M = w_y \times h_y$. Then, $\mathbf{x}_n$ can find the most similar spatial feature in $\mathbf{Y}$ to match, and its matching score s_n. Therefore, we try to obtain the similar scores for all spatial features of $\mathbf{X}$ with respect to $\mathbf{Y}$, and the final matching score can be defined as $s = \sum_{n=1}^{N} s_n$.

5.2. Post-processing.

Two re-rank methods: K-reciprocal coding [11] and Query Expansion (QE) [12] are implemented in FastReID.

Query expansion. Given a query image, and use it to find m similar gallery images. The query feature is defined as $\mathbf{f}_q$ and m similar gallery features are defined as $\mathbf{f}_g$. Then the new query feature is constructed by averaging the verified gallery features and the query feature. So the new query

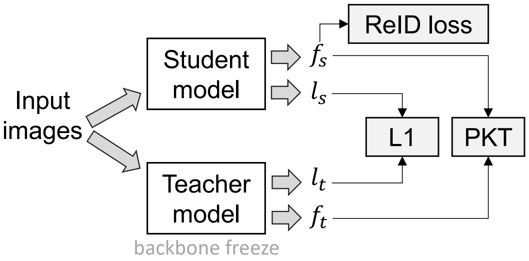

Figure 5. Illustration of knowledge distillation module

feature $\mathbf{f}_{newq}$ can be defined as

$$\mathbf{f}_{q_{new}} = \frac{\mathbf{f}_q + \sum_{i=1}^{m} \mathbf{f}_g^{(i)}}{m+1}. \tag{7}$$

After that the new query feature $\mathbf{f}_{q_{new}}$ is used for following image retrieve. QE can be easily used for practical scenarios.

5.3. Evaluation

For performance evaluation, we employ the standard metrics as in most person re-identification literature, namely the cumulative matching cure (CMC) and the mean Average Precision (mAP). Besides, we also add two metrics: receiver operating characteristic (ROC) curve and mean inverse negative penalty (mINP) [13].

5.4. Visualization

We provide a rank list tool of retrieval result that contributes to checking the problems of our algorithm that we haven't solved.

6. Deployment

In general, the deeper the model, the better the performance. However, too deep a model is not easy to deploy in edge computing hardware and AI chips since 1) it needs time-consuming inference; 2) many layers are difficult to implement on AI chips. Considering these reasons, we implement the knowledge distillation module in FastReID to achieve a high-precision, high-efficiency lightweight model.

As shown in Fig. 5, given a pre-trained student model and a pre-trained teacher model on reid datasets, the teacher model is a deeper model with non-local module, ibn module and some useful tricks. The student model is simple and shallow. We adopt two-stream way to train the student model with teacher backbone freezing. The student and teacher models respectively output classifier logits $\mathbf{l}_s$, $\mathbf{l}_t$ and features $\mathbf{f}_s$, $\mathbf{f}_t$. We want student model to learn classification ability as much as possible about the teacher model, the logit learning can be defined as

$$\mathcal{L}_{logit} = \|\mathbf{l}_s - \mathbf{l}_t\|_1. \tag{8}$$

Table 1. Performance comparison on Market1501, DukeMTMC and MSMT17 datasets.

Methods	**Market1501**		**DukeMTMC**		**MSMT17**	
	R1	mAP	R1	mAP	R1	mAP
SPReID [14] (CVPR'18)	92.5	81.3	84.4	70.1	-	-
PCB [15] (ECCV'18)	92.3	77.4	81.8	66.1	-	-
AANet [16] (CVPR'19)	93.9	83.4	87.7	74.3	-	-
IANet [17] (CVPR'19)	94.4	83.1	87.1	73.4	75.5	45.8
CAMA [18] (CVPR'19)	94.7	84.5	85.8	72.9	-	-
DGNet [19] (CVPR'19)	94.8	86.0	86.6	74.8	-	-
DSAP [20] (CVPR'19)	95.7	87.6	86.2	74.3	-	-
Pyramid [19] (CVPR'19)	95.7	88.2	89.0	79.0	-	-
Auto-ReID [21] (ICCV'19)	94.5	85.1	-	-	78.2	52.5
OSNet [1] (ICCV'19)	94.8	84.9	88.6	73.5	78.7	52.9
MHN [22] (ICCV'19)	95.1	85.0	89.1	77.2	-	-
P^2-Net [23] (ICCV'19)	95.2	85.6	86.5	75.1	-	-
BDB [24] (ICCV'19)	95.3	86.7	89.0	76.0	-	-
FPR [25] (ICCV'19)	95.4	86.6	88.6	78.4	-	-
ABDNet [22] (ICCV'19)	95.6	88.3	89.0	78.6	82.3	60.8
SONA [26] (ICCV'19)	95.7	88.7	89.3	78.1	-	-
SCAL [22] (ICCV'19)	95.8	89.3	89.0	79.6	-	-
CAR [1] (ICCV'19)	96.1	84.7	86.3	73.1	-	-
Circle Loss [10] (CVPR'20)	96.1	87.4	-	-	76.9	52.1
FastReID (ResNet50)	95.4	88.2	89.6	79.8	83.3	59.9
FastReID (ResNet50-ibn)	95.7	89.3	91.3	81.6	84.0	61.2
FastReID (ResNeSt)	95.0	87.0	90.5	79.1	82.6	58.2
FastReID-MGN (ResNet50-ibn)	95.7	89.7	91.6	82.1	**85.1**	**65.4**
FastReID (ResNet101-ibn)	**96.3**	**90.3**	**92.4**	**83.2**	**85.1**	63.3
+ QE	**96.5**	**94.4**	**93.4**	**90.1**	**87.9**	**76.9**
+ Rerank	**96.8**	**95.3**	**94.4**	**92.2**	**-**	**-**

In order to ensure the consistency of student model and teacher model in the feature space distribution, probabilistic knowledge transfer model based on Kullback-Leibler divergence is used for optimizing the student model:

$$\begin{cases} \mathcal{L}_{PKT} = \sum_{i=1}^{N} \sum_{j=1, i \neq j}^{N} p_{j|i} \log(\frac{p_{j|i}}{p_{i|j}}) \\ p_{i|j} = \frac{K(\mathbf{f}_s^i, \mathbf{f}_s^j)}{\sum_{j=1, i \neq j}^{N} K(\mathbf{f}_s^i, \mathbf{f}_s^j)} \\ p_{j|i} = \frac{K(\mathbf{f}_t^i, \mathbf{f}_t^j)}{\sum_{j=1, i \neq j}^{N} K(\mathbf{f}_t^{i),\mathbf{f}_t^{j}}} \end{cases} \tag{9}$$

where $K(:,:)$ is cosine similarity measure.

At the same time, the student model needs ReID loss $\mathcal{L}_{reid}$ to optimize the entire network. Therefore, the total loss is:

$$\mathcal{L}_{kd} = \mathcal{L}_{logit} + \alpha \mathcal{L}_{PKT} + \mathcal{L}_{reid}. \tag{10}$$

After finish training, the $\mathbf{f}_s$ is used for inference.

We also provide model conversion tool (**PyTorch** → **Caffe** and **PyTorch** → **TensorRT**) in the FastReID library.

7. Projects

7.1. Person Re-identification

Datasets. Three person re-id benchmarking datasets: Market1501 [27], DukeMTMC [28], MSMT17 [29] are used for evaluating the FastReID. We won't go into the details of the database here.

Table 2. Ablation Studies of FastReID on DukeMTMC. (ResNet50, 384×128).

Bag-of-Tricks	IBN	Auto-Augment	Soft Margin	Non-Local	Gem Pooling	Circle Loss	Backbone Freeze	Cosine Lr Scheduler	R1	mAP	mINP
√									85.5	75.2	37.9
√	√								89.2	79.1	43.9
√		√							84.9	72.8	34.5
√			√						86.1	76.3	39.0
√				√					87.3	77.6	42.0
√					√				87.4	77.1	40.3
√						√			88.7	78.3	41.8
√							√		85.9	74.7	36.4
√						√		√	88.8	77.8	40.3
√						√	√		89.5	78.3	41.6
√						√	√	√	89.5	78.5	42.5
√	√	√	√	√	√	√	√	√	91.3	81.6	47.6

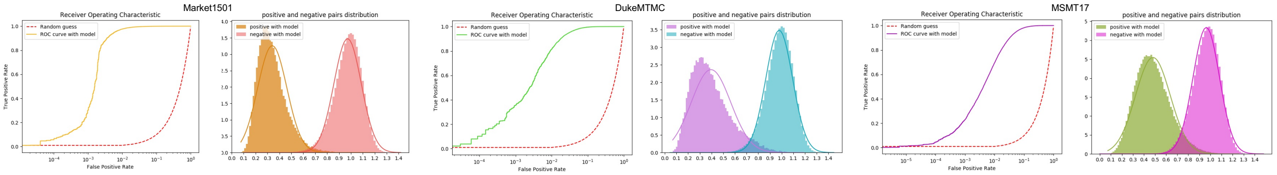

Figure 6. ROC curves and distribution curves between intra-class and inter-class samples on three benchmarking datasets for FastReID (ResNet101-ibn)

FastReID Setting. We use flipping, random erasing and auto-augment to process the training image. The IBN-ResNet101 with a Non-local module is used as the backbone. The gem pooling and bnneck are used as the head and aggregation layer, respectively. For the batch hard triplet loss function, one batch consists of 4 subjects, and each subject has 16 different images, and we use circle loss and triplet loss to train the whole network.

Result. The state-of-the-art algorithms published in CVPR, ICCV, ECCV during 2018-2020 are listed in Table 1, FastReID achieves the best performance on Market1501 96.3%(90.3%), DukeMTMC 92.4%(83.2%) and MSMT17 85.1%(65.4%) at rank-1/mAP accuracy, respectively. Fig. 6 shows the ROC curves on the three benchmarking datasets.

7.2. Cross-domain Person Re-identification

Problem definition. Cross-domain person re-identification aims at adapting the model trained on a labeled source domain dataset to another target domain dataset without any annotation.

Setting. We propose a cross-domain method FastReID-MLT that adopts mixture label transport to learn pseudo label by multi-granularity strategy. We first train a model with a source-domain dataset and then finetune on the pre-trained model with pseudo labels of the target-domain dataset. FastReID-MLT is implemented by ResNet50 backbone, gem pooling and bnneck head. For the batch hard triplet loss function, one batch consists of 4 subjects, and each subject has 16 different images, and we use circle loss and triplet loss to train the whole network. Detailed configuration can be found on the GitHub website. The framework of FastReID-MLT is shown in Fig. 7.

Result. Table 3 shows the results on several datasets,

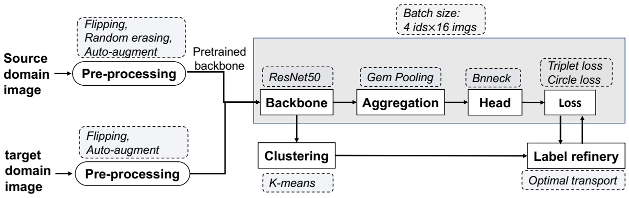

Figure 7. Framework of FastReID-MLT

Table 3. Performance comparison to the unsupervised cross-domain re-id SOTA methods on three benchmark datasets. "BOT" denotes to the bag of tricks method, which is a strong baseline in the ReID task. M: Market1501, D: DukeMTMC, MS: MSMT17.

Methods	D→M		M→D	
	mAP	R1	mAP	R1
TJ-AIDL [30] (CVPR'18)	26.5	58.2	23.0	44.3
SPGAN [31] (CVPR'18)	22.8	51.5	22.3	41.1
ATNet [32](CVPR'19)	25.6	55.7	24.9	45.1
SPGAN+LMP [33](CVPR'18)	26.7	57.7	26.2	46.4
HHL [34] (ECCV'18)	31.4	62.2	27.2	46.9
ARN [35] (CVPR'18-WS)	39.4	70.3	33.4	60.2
ECN [36] (CVPR'19)	43.0	75.1	40.4	63.3
UCDA [37] (ICCV'19)	30.9	60.4	31.0	47.7
PDA-Net [38] (ICCV'19)	47.6	75.2	45.1	63.2
PCB-PAST [39] (ICCV'19)	54.6	78.4	54.3	72.4
SSG [40] (ICCV'19)	58.3	80.0	53.4	73.0
MPLP+MMCL [41] (CVPR'20)	60.4	84.4	51.4	72.4
AD-Cluster [42] (CVPR'20)	68.3	86.7	54.1	72.6
MMT [43] (ICLR'20)	71.2	87.7	65.1	78.0
FastReID-MLT	**80.5**	**92.7**	**69.2**	**82.7**
Supervised learning (BOT [44])	85.7	94.1	75.8	86.2

Methods	M→MS		D→MS	
	mAP	R1	mAP	R1
PTGAN [45] (CVPR'18)	2.9	10.2	3.3	11.8
ENC [36] (CVPR'19)	8.5	25.3	10.2	30.2
SSG [40] (ICCV'19)	13.2	31.6	13.3	32.2
DAAM [46] (AAAI'20)	20.8	44.5	21.6	46.7
MMT [43] (ICLR'20)	22.9	49.2	23.3	50.1
FastReID-MLT	**26.5**	**56.6**	**27.7**	**59.5**
Supervised learning (BOT [44])	48.3	72.3	48.3	72.3

Table 4. Comparison of the state-of-the-art Partial Person Re-ID methods on the PartialREID, OccludedREID and PartialiLIDS datastes.

Methods	PartialREID		OccludedREID		PartialiLIDS	
	R1	mAP	R1	mAP	R1	mAP
PCB [15] (ECCV'18)	56.3	54.7	41.3	38.9	46.8	40.2
SCPNet [47] (ACCV'18)	68.3	-	-	-	-	-
DSR [48] (CVPR'18)	73.7	68.1	72.8	62.8	64.3	58.1
VPM [49] (CVPR'19)	67.7	-	-	-	65.5	-
FPR [50] (ICCV'19)	81.0	76.6	78.3	68.0	68.1	61.8
HOReID [51] (CVPR'20)	**85.3**	-	80.3	70.2	72.6	-
FastReID-DSR	82.7	**76.8**	**81.6**	**70.9**	**73.1**	**79.8**

FastReID-MLT can achieve 92.7%(80.5%), 82.7%(69.2%) under D→M, M→D settings. The result is close to supervised learning results.

7.3. Partial Person Re-identification

Problem definition. Partial person re-identification (re-id) is a challenging problem, where only several partial observations (images) of people are available for matching.
Setting. The setting as shown in Fig. 8.

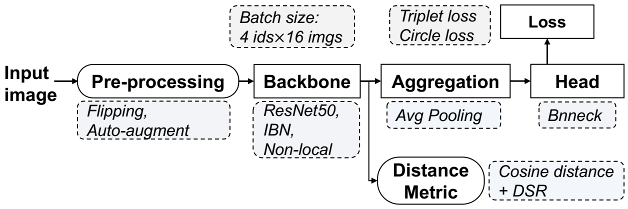

Figure 8. Framework of FastReID-DSR

Result. Table 5 shows the results on PartialREID, OccludedREID and PartialiLIDS datasets. FastReID-DSR can achieve 82.7% (76.8%), 81.6% (70.9%) and 73.1% (79.8) at rank-1/mAP metrics.

7.4. Vehicle Re-identification

Datasets. Three vehicle re-id benchmarking datasets: VeRi, VehicleID and VERI-Wild are used for evaluating the proposed FastReIDin the FastReID. We won't go into the details of the database here.
Settings. The setting as shown in Fig. 9.
Result. The state-of-the-art algorithms published during 2015-2019 are listed in Table 5, Table 6, Table 7. FastReID achieves the best performance on VeRi, VehicleID and VERI-Wild, respectively.

8. Conclusion

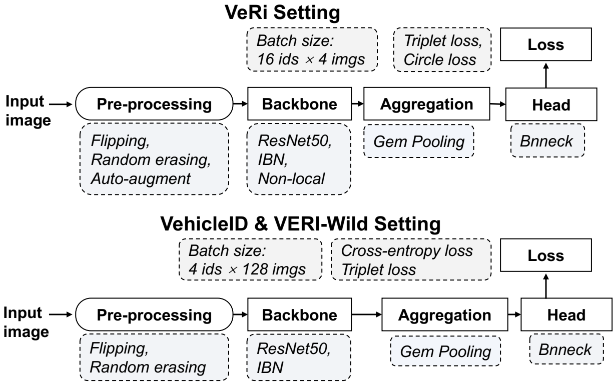

Figure 9. Framework of FastReID on VehicleID and VERI-Wild

Table 5. Comparison of the state-of-the-art vehicle Re-Id methods on the VeRi dataset.

Methods	mAP (%)	R-1 (%)	R-5 (%)
Siamese-CNN [52] (ICCV'17)	54.2	79.3	88.9
FDA-Net [53] (CVPR'19)	55.5	84.3	92.4
Siamese-CNN+ST [52] (ICCV'17)	58.3	83.5	90.0
PROVID [54] (TMM'18)	53.4	81.6	95.1
PRN [55] (CVPR'19)	70.2	92.2	97.9
PAMTRI [56](ICCV'19)	71.8	92.9	97.0
PRN [55] (CVPR'19)	74.3	94.3	98.9
FastReID	**81.9**	**97.0**	**99.0**

Table 6. Comparison of the state-of-the-art vehicle Re-Id methods on the VehicleID dataset.

Methods	Small		Medium		Large	
	R-1	R-5	R-1	R-5	R-1	R-5
DRDL [57]	48.9	73.5	42.8	66.8	38.2	61.6
NuFACT [54]	48.9	69.5	43.6	65.3	38.6	60.7
VAMI [58]	63.1	83.3	52.9	75.1	47.3	70.3
FDA-Net [53]	-	-	59.8	77.1	55.5	74.7
AAVER [59]	74.7	93.8	68.6	90.0	63.5	85.6
OIFE [60]	-	-	-	-	67.0	82.9
PRN [55]	78.4	92.3	75.0	88.3	74.2	86.4
FastReID	**86.6**	**97.9**	**82.9**	**96.0**	**80.6**	**93.9**

Table 7. Comparison of the state-of-the-art vehicle Re-Id methods on the VERI-Wild dataset.

Methods	Small		Medium		Large	
	mAP	R-1	mAP	R-1	mAP	R-1
GoogLeNet [61]	24.3	57.2	24.2	53.2	21.5	44.6
DRDL [57]	22.5	57.0	19.3	51.9	14.8	44.6
FDA-Net [53]	35.1	64.0	29.8	57.8	22.8	49.4
MLSL [62]	46.3	86.0	42.4	83.0	36.6	77.5
FastReID	**87.7**	**96.4**	**83.5**	**95.1**	**77.3**	**92.5**

This paper introduces a open source library namely FastReID for general instance re-identification. Experimental results demonstrated the versatility and effectiveness of FastReID on multiple tasks, such as person re-identification and vehicle re-identification. Were sharing FastReID because open source research platforms are critical to the rapid advances in AI made by the entire community, including researchers and practitioners in academia and industry. We hope that releasing FastReID will continue to accelerate progress in the area of general instance re-identification. We also look forward to collaborating with learning from each other for advancing the development of computer vision.

References

[1] Kaiyang Zhou, Yongxin Yang, Andrea Cavallaro, and Tao Xiang. Omni-scale feature learning for person re-identification. In *The IEEE International Conference on Computer Vision (ICCV)*, 2019. 2, 5

[2] Terrance DeVries and Graham W Taylor. Improved regularization of convolutional neural networks with cutout. *arXiv preprint arXiv:1708.04552*, 2017. 2

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 3

[4] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. In *The IEEE conference on computer vision and pattern recognition (CVPR)*, 2017. 3

[5] Hang Zhang, Chongruo Wu, Zhongyue Zhang, Yi Zhu, Zhi Zhang, Haibin Lin, Yue Sun, Tong He, Jonas Mueller, R Manmatha, et al. Resnest: Split-attention networks. *arXiv preprint arXiv:2004.08955*, 2020. 3

[6] Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He. Non-local neural networks. In *The IEEE conference on computer vision and pattern recognition (CVPR)*, 2018. 3

[7] Xingang Pan, Ping Luo, Jianping Shi, and Xiaoou Tang. Two at once: Enhancing learning and generalization capacities via ibn-net. In *The European Conference on Computer Vision (ECCV)*, 2018. 3

[8] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015. 3

[9] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 4

[10] Yifan Sun, Changmao Cheng, Yuhan Zhang, Chi Zhang, Liang Zheng, Zhongdao Wang, and Yichen Wei. Circle loss: A unified perspective of pair similarity optimization. 2020. 4, 5

[11] Zhun Zhong, Liang Zheng, Donglin Cao, and Shaozi Li. Re-ranking person re-identification with k-reciprocal encoding. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 4

[12] Ranjita Bhagwan, Kiran Tati, Yuchung Cheng, Stefan Savage, and Geoffrey M Voelker. Total recall: System support for automated availability management. In *Nsdi*, volume 4, 2004. 4

[13] Mang Ye, Jianbing Shen, Gaojie Lin, Tao Xiang, Ling Shao, and Steven CH Hoi. Deep learning for person re-identification: A survey and outlook. *arXiv preprint arXiv:2001.04193*, 2020. 5

[14] Mahdi M. Kalayeh, Emrah Basaran, Muhittin Gkmen, Mustafa E. Kamasak, and Mubarak Shah. Human semantic parsing for person re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 5

[15] 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)*, 2018. 5, 7

[16] Chiat-Pin Tay, Sharmili Roy, and Kim-Hui Yap. Aanet: Attribute attention network for person re-identifications. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 5

[17] Ruibing Hou, Bingpeng Ma, Hong Chang, Xinqian Gu, Shiguang Shan, and Xilin Chen. Interaction-and-aggregation network for person re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 5

[18] Wenjie Yang, Houjing Huang, Zhang Zhang, Xiaotang Chen, Kaiqi Huang, and Shu Zhang. Towards rich feature discovery with class activation maps augmentation for person re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 5

[19] Feng Zheng, Cheng Deng, Xing Sun, Xinyang Jiang, Xiaowei Guo, Zongqiao Yu, Feiyue Huang, and Rongrong Ji. Pyramidal person re-identification via multi-loss dynamic training. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 5

[20] Zhizheng Zhang, Cuiling Lan, Wenjun Zeng, and Zhibo Chen. Densely semantically aligned person re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 5

[21] Ruijie Quan, Xuanyi Dong, Yu Wu, Linchao Zhu, and Yi Yang. Auto-reid: Searching for a part-aware convnet for person re-identification. In *The IEEE International Conference on Computer Vision (ICCV)*, 2019. 5

[22] Binghui Chen, Weihong Deng, and Jiani Hu. Mixed high-order attention network for person re-identification. In *The IEEE International Conference on Computer Vision (ICCV)*, 2019. 5

[23] Jianyuan Guo, Yuhui Yuan, Lang Huang, Chao Zhang, Jin-Ge Yao, and Kai Han. Beyond human parts: Dual part-aligned representations for person re-identification. In *The IEEE International Conference on Computer Vision (ICCV)*, 2019. 5

[24] Zuozhuo Dai, Mingqiang Chen, Xiaodong Gu, Siyu Zhu, and Ping Tan. Batch dropblock network for person re-identification and beyond. In *The IEEE International Conference on Computer Vision (ICCV)*, 2019. 5

[25] Lingxiao He, Yinggang Wang, Wu Liu, He Zhao, Zhenan Sun, and Jiashi Feng. Foreground-aware pyramid reconstruction for alignment-free occluded person re-identification. In *The IEEE International Conference on Computer Vision (ICCV)*, 2019. 5

[26] Bryan (Ning) Xia, Yuan Gong, Yizhe Zhang, and Christian Poellabauer. Second-order non-local attention networks for person re-identification. In *The IEEE International Conference on Computer Vision (ICCV)*, 2019. 5

[27] Song Bai, Xiang Bai, and Qi Tian. Scalable person re-identification on supervised smoothed manifold. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 5

[28] 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)*, 2017. 5

[29] Xuelin Qian, Yanwei Fu, Tao Xiang, Yu-Gang Jiang, and Xiangyang Xue. Leader-based multi-scale attention deep architecture for person re-identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. 5

[30] Jingya Wang, Xiatian Zhu, Shaogang Gong, and Wei Li. Transferable joint attribute-identity deep learning for unsupervised person re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 6

[31] 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)*, 2018. 6

[32] Jiawei Liu, Zheng-Jun Zha, Di Chen, Richang Hong, and Meng Wang. Adaptive transfer network for cross-domain person re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 6

[33] 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)*, 2018. 6

[34] 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)*, 2018. 6

[35] Yu-Jhe Li, Fu-En Yang, Yen-Cheng Liu, Yu-Ying Yeh, Xiaofei Du, and Yu-Chiang Frank Wang. Adaptation and re-identification network: An unsupervised deep transfer learning approach to person re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition Workshops CVPRW)*, 2018. 6

[36] Zhun Zhong, Liang Zheng, Zhiming Luo, Shaozi Li, and Yi Yang. Invariance matters: Exemplar memory for domain adaptive person re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 6

[37] Lei Qi, Lei Wang, Jing Huo, Luping Zhou, Yinghuan Shi, and Yang Gao. A novel unsupervised camera-aware domain adaptation framework for person re-identification. *The IEEE International Conference on Computer Vision (ICCV)*, 2019. 6

[38] Yu-Jhe Li, Ci-Siang Lin, Yan-Bo Lin, and Yu-Chiang Frank Wang. Cross-dataset person re-identification via unsupervised pose disentanglement and adaptation. *The IEEE International Conference on Computer Vision (ICCV)*, 2019. 6

[39] Xinyu Zhang, Jiewei Cao, Chunhua Shen, and Mingyu You. Self-training with progressive augmentation for unsupervised cross-domain person re-identification. *The IEEE International Conference on Computer Vision (ICCV)*, 2019. 6

[40] Fu Yang, Wei Yunchao, Wang Guanshuo, Zhou Yuqian, Shi Honghui, and Huang Thomas. Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification. *The IEEE International Conference on Computer Vision (ICCV)*, 2019. 6

[41] Dongkai Wang and Shiliang Zhang. Unsupervised person re-identification via multi-label classification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 6

[42] Yunpeng Zhai, Shijian Lu, Qixiang Ye, Xuebo Shan, Jie Chen, Rongrong Ji, and Yonghong Tian. Ad-cluster: Augmented discriminative clustering for domain adaptive person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 6

[43] 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. 6

[44] Hao Luo, Youzhi Gu, Xingyu Liao, Shenqi Lai, and Wei Jiang. Bag of tricks and a strong baseline for deep person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019. 6

[45] 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)*, 2018. 6

[46] Yi Jin Yidong Li Junliang Xing Shiming Ge Yangru Huang, Peixi Peng. Domain adaptive attention model for unsupervised cross-domain person re-identification. In *The AAAI Conference on Artificial Intelligence*, 2020. 6

[47] Xing Fan, Hao Luo, Xuan Zhang, Lingxiao He, Chi Zhang, and Wei Jiang. Scpnet: Spatial-channel parallelism network for joint holistic and partial person re-identification. In *Asian Conference on Computer Vision (ACCV)*, 2018. 7

[48] Lingxiao He, Jian Liang, Haiqing Li, and Zhenan Sun. Deep spatial feature reconstruction for partial person re-identification: Alignment-free approach. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 7

[49] Yifan Sun, Qin Xu, Yali Li, Chi Zhang, Yikang Li, Shengjin Wang, and Jian Sun. Perceive where to focus: Learning visibility-aware part-level features for partial person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 7

[50] Lingxiao He, Yinggang Wang, Wu Liu, He Zhao, Zhenan Sun, and Jiashi Feng. Foreground-aware pyramid reconstruction for alignment-free occluded person re-identification. In *IEEE International Conference on Computer Vision (ICCV)*, 2019. 7

[51] Guan'an Wang, Shuo Yang, Huanyu Liu, Zhicheng Wang, Yang Yang, Shuliang Wang, Gang Yu, Jian Sun, et al. High-order information matters: Learning relation and topology for occluded person re-identification. 2020. 7

[52] Yantao Shen, Tong Xiao, Hongsheng Li, Shuai Yi, and Xiaogang Wang. Learning deep neural networks for vehicle re-id with visual-spatio-temporal path proposals. In *IEEE International Conference on Computer Vision (CVPR)*, 2017. 7

[53] Yihang Lou, Yan Bai, Jun Liu, Shiqi Wang, and Lingyu Duan. Veri-wild: A large dataset and a new method for vehicle re-identification in the wild. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 7

[54] Xinchen Liu, Wu Liu, Tao Mei, and Huadong Ma. PROVID: progressive and multimodal vehicle reidentification for large-scale urban surveillance. *IEEE Transactions on Multimedia*, 2018. 7

[55] Bing He, Jia Li, Yifan Zhao, and Yonghong Tian. Part-regularized near-duplicate vehicle re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 7

[56] Zheng Tang, Milind Naphade, Stan Birchfield, Jonathan Tremblay, William Hodge, Ratnesh Kumar, Shuo Wang, and Xiaodong Yang. Pamtri: Pose-aware multi-task learning for vehicle re-identification using highly randomized synthetic data. In *IEEE International Conference on Computer Vision (ICCV)*, 2019. 7

[57] Hongye Liu, Yonghong Tian, Yaowei Yang, Lu Pang, and Tiejun Huang. Deep relative distance learning: Tell the difference between similar vehicles. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 7

[58] Yi Zhou and Ling Shao. Aware attentive multi-view inference for vehicle re-identification. In *IEEE International Conference on Computer Vision (CVPR)*, 2018. 7

[59] Pirazh Khorramshahi, Amit Kumar, Neehar Peri, Sai Saketh Rambhatla, Jun-Cheng Chen, and Rama Chellappa. A dual-path model with adaptive attention for vehicle re-identification. In *IEEE International Conference on Computer Vision (ICCV)*, 2019. 7

[60] Zhongdao Wang, Luming Tang, Xihui Liu, Zhuliang Yao, Shuai Yi, Jing Shao, Junjie Yan, Shengjin Wang, Hongsheng Li, and Xiaogang Wang. Orientation invariant feature embedding and spatial temporal regularization for vehicle re-identification. In *IEEE International Conference on Computer Vision (CVPR*, 2017. 7

[61] Linjie Yang, Ping Luo, Chen Change Loy, and Xiaoou Tang. A large-scale car dataset for fine-grained categorization and verification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. 7

[62] Saghir Alfasly, Yongjian Hu, Haoliang Li, Tiancai Liang, Xiaofeng Jin, Beibei Liu, and Qingli Zhao. Multi-label-based similarity learning for vehicle re-identification. *IEEE Access*, 2019. 7