ABSTRACT
General Instance Re-identification is a very important task in computer vision, which can be widely used in many practical applications, such as person/vehicle re-identification, face recognition, wildlife protection, commodity tracing, snapshots, and so on. To meet the increasing application demand for general instance re-identification, we present FastReID as a widely used software system. In FastReID, the 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 wildlife protection, commodity tracing, snapshots, and so on. To reproduce our project results very easily. The source codes and models have been released with many users and have received over 3000 stars and 800 forks at https://github.com/JDAI-Explore-Academy/FastReID. Moreover, we 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, it can reproduce our project results very easily. The source codes and models have been released with many users and have received over 3000 stars and 800 forks at https://github.com/JDAI-Explore-Academy/FastReID.

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 snapshot 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 own 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 retrieval and face recognition, etc. FastReID provides 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.

2 HIGHLIGHT OF FASTREID

FastReID provides a complete toolkit for training, evaluation, fine-tuning and model deployment. Besides, FastReID provides strong baselines that are capable of achieving state-of-the-art performance on multiple tasks, and it have the following characteristics:

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 practitioners can quickly implement their ideas without writing 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. Deep models are 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 deployment.

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 Cutout, Mixup 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, ResNeXt, ResNetS, MobileNet, ShuffleNet and RegNet. We also add attention-like non-local module and instance batch normalization (IBN) 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 \( X \in \mathbb{R}^{W \times H \times C} \) as input and produces a vector \( f \in \mathbb{R}^{W \times H \times C} \) as an output of the pooling process, where \( W, H, C \) respectively represent the width, the height and the channel of the feature maps.

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. 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 is used to solve internal covariate shift because it is very difficult to train models with saturating nonlinearities.

Reduction layer is aiming to make the high-dimensional feature become the low-dimensional feature, i.e., \( 2048\text{-dim} \rightarrow 512\text{-dim} \).

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

3.5 Loss Function

FastReID implements a series of softmax-style losses such as Cross-entropy loss with Label Smoothing, Arcface loss, Circle loss, CosFace and metric-style Triplet loss.

4 TRAINING STRATEGY

Model training contains many tricks such as learning rate for different iteration, network warm-up and backbone 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., \( 3.5 \times 10^{-5} \) in the initial training and then gradually increase it during the 2k iterations. After that, the learning rate remains at \( 3.5 \times 10^{-4} \) between 2k iterations and 9k iterations. Then, the learning rate starts from \( 3.5 \times 10^{-4} \) and decays to \( 7.7 \times 10^{-7} \) at cosine rule after 9k iterations, the training is finished at 18k iterations.
5 EVALUATION

5.1 Distance Metric.

Euclidean 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 \( x \) for \( x \) with dimension \( w_x \times h_x \times d \), and \( y \) for \( y \) with dimension \( w_y \times h_y \times d \). The total \( N \) spatial features from \( N \) locations are aggregated into a matrix:

\[
X = \{ x_n \}_{n=1}^{N} \in \mathbb{R}^{d \times N},
\]

where \( N = w_x \times h_x \). Likewise, we construct the gallery feature matrix:

\[
Y = \{ y_m \}_{m=1}^{M} \in \mathbb{R}^{d \times M},
\]

where \( M = w_y \times h_y \). Then, \( x_n \) can find the most similar spatial feature in \( Y \) to match, and its matching score \( s_n \). Therefore, we try to obtain the similar scores for all spatial features of \( X \) with respect to \( Y \), 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 [13] and Query Expansion (QE) [3] 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 \( f_q \) and \( m \) similar gallery features are defined as \( f_y \). Then the new query feature is constructed by averaging the verified gallery features and the query feature. So the new query feature \( f_{new} \) can be defined as:

\[
f_{new} = \frac{f_q + \sum_{i=1}^{m} f_y^{(i)}}{m + 1}.
\]

The results 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. 2 shows the ROC curves on the three benchmarking datasets.

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

| Methods                  | Market1501 | DukeMTMC | MSMT17 |
|--------------------------|------------|----------|--------|
|                           | R1 - mAP  | R1 - mAP | R1 - mAP |
| FastReID (ResNet50)      | 95.4       | 89.6     | 83.3     |
| FastReID (ResNet50-ibn)  | 95.7       | 89.3     | 84.0     |
| FastReID (ResNeSt)       | 95.0       | 87.0     | 82.6     |
| FastReID-MGN (ResNet50-ibn) | 95.7 | 89.7    | 85.1 | 65.4 |
| FastReID (ResNet101-ibn) | 96.3       | 90.3     | 85.1     |
| + QE                     | 96.5       | 94.4     | 90.1     |
| + Rerank                 | 96.8       | 95.3     | 94.4     |

After that the new query feature \( f_{new} \) 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 curve (CMC) and the mean Average Precision (mAP). Besides, we also add two metrics: receiver operating characteristic (ROC) curve and mean inverse negative penalty (mINP).

5.4 Visualization

FastReID provide a rank list tool of retrieval result that contributes to checking the problems of our algorithm that we haven’t solved.
6.2 Cross-domain Person Re-identification

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. FastReID proposes 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 bottleneck 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. Table 2 shows the results on several datasets, 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.

6.3 Partial Person Re-identification

Partial person re-identification (re-id) is a challenging problem, where only several partial observations (images) of people are available for matching. Table 3 shows the results on PartialREID, OccludedREID and PartialLIDS datasets. FastReID-DSR can achieve 82.7%(76.8%), 81.6%(70.9%) and 73.1%(79.8) at rank-1/mAP metrics.

6.4 Vehicle Re-identification

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. The state-of-the-art algorithms published during 2015-2019 are listed in Table 3 and Table 4. FastReID achieves the best performance on VeRi, VehicleID and VERI-Wild, respectively.

### Table 2: Performance comparison to the unsupervised cross-domain re-id SOTA methods on three benchmark datasets.

| Methods       | D→M | M→D |
|---------------|-----|-----|
| mAP R1        |     |     |
| FastReID-MLT  | 80.5| 69.2|
| Supervised (BOT) [8] | 85.7| 75.8|

### Table 3: Comparison of the state-of-the-art Partial Person Re-ID methods on the PartialREID, OccludedREID and PartialLIDS datasets.

| Methods          | PartialREID | OccludedREID | PartialLIDS |
|------------------|-------------|--------------|-------------|
| SCPNet [4] (ACCV'18) | 68.3        | -            | -           |
| DSR [6] (CVPR'18)   | 73.7        | 62.8         | 64.3        |
| VPM [10] (CVPR'19)  | 67.7        | 65.5         | 68.1        |
| FPR [7] (ACCV'19)   | 81.0        | 68.0         | 68.1        |
| HOReID [11] (CVPR’20) | 85.3 | 72.6         | -           |
| FastReID-DSR       | 82.7        | 76.8         | 73.1        |

### Table 4: Comparison of the state-of-the-art vehicle Re-ID methods on the VehicleID dataset.

| Methods-vehicle Re-Id | Small | Medium | Large |
|-----------------------|-------|--------|-------|
| mAP R1                |       |        |       |
| FastReID              | 86.6  | 97.9   | 96.0  |
| PRN [5]               | 78.4  | 92.5   | 83.3  |

7 CONCLUSION

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. We’re 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.

ACKNOWLEDGEMENTS

This work was supported by the Beijing Nova Program (No. 20220484063).

### REFERENCES

[1] Saghir Alfasly, Yongjian Hu, Haoliang Li, Tiancai Liang, Xiaofeng Jin, Beibei Liu, and Qingli Zhao. 2019. Multi-label based similarity learning for vehicle re-identification. IEEE Access (2019).
[2] Song Bai, Xiang Bai, and Qi Tian. 2017. Scalable person re-identification on supervised smoothed manifold. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
[3] Ranjita Bhagwan, Kiran Tata, Yuchung Cheng, Stefan Savage, and Geoffrey M Voelker. 2004. Total recall: System support for automated availability management. In Nodi, Vol. 4.
[4] Xing Fan, Hao Luo, Xuan Zhang, Lingxiao He, Chi Zhang, and Wei Jiang. 2018. Sceptr: Spatial-channel parallelism network for joint holistic and partial person re-identification. In Asian Conference on Computer Vision (ACCV).
[5] Bing He, Jia Li, Yifan Zhao, and Yongchong Tian. 2019. Part-regularized near-duplicate vehicle re-identification. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
[6] Lingxiao He, Jian Liang, Haoliang Li, and Zhenan Sun. 2018. Deep spatial feature reconstruction for partial person re-identification. Alignment-free approach. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
[7] Lingxiao He, Yinggang Wang, Wu Liu, He Zhao, Zhenan Sun, and Jiashi Feng. 2019. Foreground-aware pyramid reconstruction for alignment-free occluded person re-identification. In The IEEE International Conference on Computer Vision (ICCV).
[8] Hao Luo, Youzi Gu, Xingyu Liao, Shengq Lai, and Wei Jiang. 2019. Bag of tricks and a strong baseline for deep person re-identification. In IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).
[9] Xuelin Qian, Yanwei Fu, Tao Xiang, Yu-Gang Jiang, and Xiangyang Xue. 2019. Leader-based multi-scale attention deep architecture for person re-identification. IEEE Transactions on Pattern Analysis and Machine Intelligence (2019).
[10] Yifan Sun, Qin Xu, Yali Li, Yikang Li, Shengjin Wang, and Jian Sun. 2019. 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).
[11] Zhezhong Zheng, Lianzhi Zheng, and Yi Yang. 2017. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In The IEEE International Conference on Computer Vision (ICCV).
[12] Zhedong Zheng, Liang Zheng, and Yi Yang. 2017. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In The IEEE International Conference on Computer Vision (ICCV).
[13] Zhun Zhong, Liang Zheng, Donglin Cao, and Shaozi Li. 2017. Re-ranking person re-identification with k-reciprocal encoding. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).