# MoRe: A Large-Scale Motorcycle Re-Identification Dataset

Augusto Figueiredo[1] Johnata Brayan[1] Renan Oliveira Reis[1] Raphael Prates[2]
William Robson Schwartz[1]
Universidade Federal de Minas Gerais (UFMG), Brazil[1]
Universidade Estadual de Campinas (UNICAMP), Brazil[2]
{augusto.maillo, johnatabrayan, renanreis1, pratesufop}@gmail.com, william@dcc.ufmg.br

## Abstract

*Motorcycles are often related to transit and criminal issues due to its abundance in the transit. Despite its importance, motorcycles are a seldom addressed problem in the computer vision community. We credit this problem to the lack of large-scale datasets and strong baseline models. Therefore, we present the first large-scale Motorcycles Re-Identification (MoRe) dataset. MoRe consists of 3,827 individuals (i.e., the set of motorbikes and motorcyclist) captured by ten surveillance cameras placed in Brazil's urban traffic scenarios. Furthermore, we evaluate a deep learning model trained using well-known training tricks from the object re-identification literature to present a strong baseline for the motorcycle re-identification (ReID) problem. More importantly, we highlight some crucial problems in this topic as the influence of distractors and the domain shift. Experimental results demonstrate the effectiveness of the strong baseline model with an increase of at least 19.27 p.p. in the rank-1 when compared to the state-of-the-art in the BPReID dataset. Finally, we present some insights regarding the information learned by the strong baseline model when computing the similarities between motorcycle images.*[1]

## 1. Introduction

Object Re-Identification (ReID) is an important task for smart surveillance systems. For instance, a Re-ID system can be applied to maintain a unique identity for objects, such as cars, motorcycles and people, as they move among surveillance cameras without superposing camera views. Therefore, it is a crucial stage to monitor objects that are not constrained to a single camera-view. In recent years, we have witnessed a rapid development of ReID systems focused mainly on persons [44, 40, 41, 27, 22, 21, 20, 5, 3] and four-wheel vehicles [19, 47, 41, 7, 46, 45, 2, 25, 15]. The importance of these objects is related to the increasing demand for public safety and traffic control. However, in this context, there is an important object that is overlooked by the ReID community: the motorcycles. In this work, we term motorcycle as the set motorbike and motorcyclist.

Motorcycles are abundant objects in the transit and often related with traffic and criminal law breaking in many countries. For instance, motorcycles are used to perform rob-and-run crimes [39] and related to an increasing number of accidents that result in severe injuries and deaths [9]. Therefore, it is crucial to identify the motorcycles and monitor its behavior in wide areas, which is an overwhelming work. Nowadays, surveillance cameras are present almost everywhere and can assist the security personnel to efficiently attain this work. Nonetheless, it is a challenge task due to problems as the low-resolution images, occlusion by other vehicles, motion blur and drastic changes in the illuminations conditions. More importantly, off-the-shelf license plate recognition systems only work in specific camera-views as most of the motorcycles only have license plate at the rear of the vehicle. Therefore, it is important to develop a motorcycle re-identification system based on appearance features.

Motorcycle ReID has some remarkable differences when compared to the classical ReID problems as person and vehicle. Vehicle ReID is a very challenging task as different cars may look very similar (inter-class similarity) when they share the same attributes as model and color. Similarly, person re-identification methods struggle to re-identify individuals that are using similar clothes (e.g. uniforms or sports jersey). Differently, as the motorcycle ReID uses information about the motorbike (e.g. model and color) and the motorcyclist (helmet, clothes and carrying objects), it reduces considerably the inter-class similarity.

Despite its importance, motorcycle ReID is a seldom addressed problem in the literature. As far as we know, Bike-Person Re-identification (BPReID) [42] is the unique work that considers the motorcycle ReID problem. In this work,

[1]The MoRe dataset and all the code can be found at http://smartsenselab.dcc.ufmg.br/dataset/more-a-large-scale-motorcycle-re-identification-dataset/ .


Figure 1: Some of the challenges present in the proposed MoRe dataset. Each column corresponds to the same individual captured in distinct camera views. The only exceptions are the last two columns that display distinct individuals.

the authors proposed a database for two-wheel vehicles (i.e. bicycles, motorcycles and electric-powered bikes) that was captured in a university campus. In truth, a campus setting does not resemble the challenges faced in a realistic urban surveillance system and more than 70% of the BPReID images correspond to bikes. Therefore, there is a demand for a large-scale motorcycle ReID dataset in the literature.

To address the above-mentioned issues and boost the research in the motorcycle ReID problem, we propose the **Mo**torcycle **Re**-Identification (MoRe) dataset, which is the first large-scale motorcycle ReID database captured by urban traffic cameras. Precisely, MoRe contains 3,827 distinct identities and 3,478 distractors captured by ten surveillance cameras in a total of 17,619 detected bounding-box images. The dataset presents important challenges to the research community due to the drastic changes of appearance that motorcycles endure between surveillance cameras as a result of different camera resolutions and views, aligned with the distinct illumination conditions, possible occlusions and the inter-class similarities (see Figure 1).

To promote research, we present the evaluation of a strong baseline model for motorcycle Re-ID in MoRe and BPReID databases. This strong baseline consists of a state-of-the-art model trained considering well-known tricks in the object re-identification literature. Experimental results in the BPReID database demonstrate the effectiveness of the strong baseline with an improvement of at least 19.27 percentage points in the *rank-1* when compared to the state-of-the-art in the motorcycle ReID problem. More importantly, we delve into the model using an explainable artificial intelligence (xAI) [1] method to highlight the regions of the images that are used to compute similarity between images. We believe that this analysis will provide valuable insights to the research community.

We also explore two crucial problems in the object re-identification community: the distractors [43], which are individuals included in the gallery without a corresponding probe image, and the domain shift [36, 28]. Experimental results demonstrate that the obtained *rank-1* diminishes as a consequence of adding distractors. Similarly, we show that models trained and tested in distinct datasets and even in different camera pairs observe a huge decline in performance due to the domain shift problem. For instance, the model trained on BPReID and tested on MoRe decreases the *rank-1* in more than 66 percentage points when compared to a model trained and tested on MoRe. We hope that these experiments highlight these important issues in the motorcycle ReID task.

## 2. Related Works

There are some works in the literature that consider problems as motorcycle detection and tracking, motorcycle helmet detection and motorcycle ReID using computer vision. In the following paragraphs, we present an overview of these works.

Motorcycle detection evolved from approaches based on background subtraction and handcrafted descriptors [31] to methods based on deep learning that work on dense [18] and occluded [13, 12, 10, 13] settings. A comprehensive survey of motorcycle detection and tracking is presented in [11].

Another topic that presented outstanding improvement recently is the motorcycle helmet detection. Initial works were based on constrained camera views, background subtraction and handcrafted descriptors [8]. Differently, recent methods used deep learning models and are trained on large scale databases [35, 4, 30]. As an example, Siebert and Lin [30] trained a deep learning model that detects the motorcycle, the riders position and predicts the helmet use surpassing human observers accuracy.

To the best of our knowledge, there is only one work that addressed the motorcycle ReID problem [42]. Yuan et *al.* [42] proposed the BPReID database and a method based on handcrafted appearance descriptors. BPReID contains 4,579 individuals between bikes, eletric-powered bikes and motorcycles in a total of 200,680 bounding box images captured by six cameras in a university campus. Nonetheless, only 940 of these individuals correspond to motorcycles. Differently, in this work, we propose the MoRe dataset, the first large-scale database entirely focused on motorcycles captured by urban traffic surveillance cameras. MoRe contains 3,827 individuals captured by ten urban traffic surveillance cameras. Furthermore, we present a strong baseline that is based on state-of-the-art deep learning models and training tricks.

# 3. MoRe Dataset

In this section, we present the proposed MoRe dataset and compare it with the only motorcycle ReID dataset in literature, the Motorcycles BPReID. Motorcycles BPReID is a subset of the BPReID that we construct containing only motorcycles and, therefore, is suitable to compare with the proposed MoRe dataset. Section 3.1 describes the dataset construction and its main characteristics. Then, Section 3.2 compares the main features of MoRe, BPReID and the Motorcycles BPReID datasets.

## 3.1. Construction

In this section, we present the dataset construction process. First, we selected six pairs of live feeds formed by devices broadcasting images of adjacent locations. This way, it could be expected that the same subject would be passing by both cameras in a short-time interval. Then, we use an instance of RetinaNet [23] model pre-trained in the MS-COCO [24] to detect both motorbikes and persons in images captured between 6 a.m and 6 p.m. We consider a true motorcycle (i.e., a motorbike with one or more riders) when both detections are above a threshold $\tau$ and the intersection-over-union (IoU) between motorbike and person is above a value $\gamma$. Finally, we merge person and motorbike bounding boxes, enlarge them 10% in each side and store them in the database with the corresponding timestamp and the camera capture information. We empirically determine the best values of $\tau$ and $\gamma$ as 0.3.

Once the data was captured, we annotated the image matches between nearby camera pairs. To lighten this overwhelming work, we used the timestamp to filter the potential matches. The result was 3,827 distinct identities in 14,141 detected bounding boxes. Notice that we have only few images per individual in each cameras as we employ real-world surveillance cameras with a low FPS. In addition to the correct matches, we included 3,478 distractors images to construct a more realistic dataset with 17,619 images. It is important to highlight that as the camera pairs are placed in distant regions, it is not possible that the same individual appears in more than a pair of cameras during the data capture.

Figure 2a presents the distribution of height and width of the detected bounding boxes, while Figure 2b shows a histogram of aspect ratios (i.e. the ratio between width and height). According to the figures, we notice that most of the detections are low-resolution and are taller than wide. As an example, 64% of the bounding boxes have resolution lower than 256x256 pixels and the mean aspect ratio is equal 0.69.

As the camera pairs corresponds to a different urban settings in Brazil, we notice that the traffic varies between cameras and, therefore, the number of detected motorcycles. Tab. 2 shows the distribution of images and examples of captures for each camera. For instance, we can observe that the $pair04$ corresponds to 42.7% of the total number of individuals. Differently, the $pair01$, $pair02$ and $pair03$ are placed in regions with small movement of vehicles, and when summed are equal only to 6.5% of the MoRe dataset. Similarly, Table 2 presents the image resolution and the number of individuals, bounding boxes and distractors annotated in each camera pair. We believe that the combination of the different characteristics of the cameras with the imbalanced number of samples captured by each camera make MoRe a very realistic dataset for motorcycle ReID.

## 3.2. Motorcycles Databases

In this section, we compare the proposed MoRe dataset with related datasets in the literature. Specifically, we consider the BPReID and the Motorcycles BPReID, which is a subset of the BPReID that contains only the motorcycle category. These comparisons are presented in Table 1.

To the best of our knowledge, BPReID is the only dataset that considers motorcycles in an object re-identification scenario. Despite being a large-scale dataset with 4,579 individuals, only a small segment of the dataset corresponds to motorcycles. As we focus on the motorcycles, we considered in this work the Motorcycles BPReID containing only 940 individuals. The proposed MoRe dataset presents advantages when compared to the Motorcycles BPReID such as the higher number of cameras and individuals (i.e. more than four times), the urban traffic scenario and detected bounding boxes. Nonetheless, we used real-world surveillance cameras with a low FPS and, therefore, we have the smallest number of bounding boxes and distractors images. In fact, the obtained experimental results demonstrate that this amount of distractor images is already enough to observe a deterioration in the performance of the model.

| Datasets | BPReID [42] | Motorcycles BPReid | MoRe |
|---|---|---|---|
| #Individuals | 4,579 | 940 | 3,827 |
| #BBoxes | 200,680 | 45,951 | 17,619 |
| #Cameras | 6 | 6 | 10 |
| #Distractors | 109,100 | 27,188 | 3,478 |
| Annotation | hand | hand | RetinaNet [23] |
| Environment | campus | campus | urban |

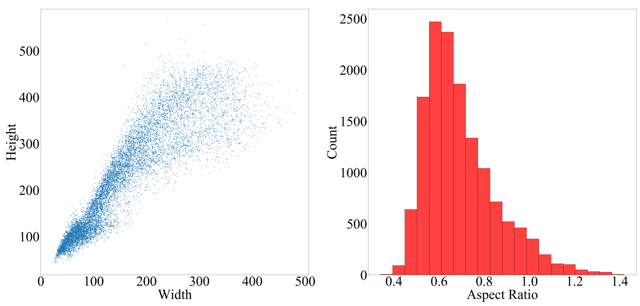
Table 1: Comparison between BPReID, Motorcycles BPReID and MoRe datasets. #Distractors are reported as the number of images.

# 4. Strong Baseline Model

In this section, we present the strong baseline model to tackle the motorcycle ReID problem. The strong baseline

| | pair01 | pair02 | pair03 | pair04 | pair05 | pair06 |
|---|---|---|---|---|---|---|
| #Individuals | 137 | 50 | 63 | 1,635 | 1,166 | 776 |
| #BBoxes | 524 | 217 | 315 | 4,767 | 5,382 | 2,936 |
| #Distractors | 178 | 178 | 402 | 925 | 1,127 | 846 |
| Cameras Resolution | 704x480 | 704x480 | 704x480 | 1920x1080 | 704x480 | 1920x1080 |

Table 2: Features of each camera pair in the proposed MoRe dataset.




(a) Resolutions. (b) Aspect Ratios.

Figure 2: Distribution of the detected bounding boxes resolutions and aspect ratios in the MoRe dataset.

Figure 3: From letf to right, examples of images in each camera pairs (columns) from *pair01* to *pair06*.

model is inspired in the model proposed by Luo et *al.* [26], where the authors evaluated a bag of training tricks for the person ReID problem. To accomplish that, we used a Resnet50 [16] pre-trained on ImageNet as the backbone model and included a fully connected layer mapping to the $N$ identities in the training set. In addition to the tricks proposed in [26], we included a comprehensive evaluation of three metric learning losses that can further improve the obtained experimental results. In the next sections, we describe both the tricks (Section 4.1) and metric learning losses (Section 4.2) considered in this work.

### 4.1. Tricks

In this section, we describe the tricks evaluated in the strong baseline model. In special, we focused on the following tricks: (1) Label Smoothing [34], (2) Warmup Learning Rate [14], (3) Last Stride [33], (4) BNNeck [26] and (5) Center Loss [37]. Despite being broadly used in the object ReID scenario, these tricks have not been graded in the motorcycle ReID problem.

Briefly, Label Smoothing is a widely used technique that avoids the overfitting in the training set by introducing a small perturbation on the training labels, and the Warmup Learning Rate improves the model convergence as it has a warmup strategy that increases the learning rate in the first training epochs. Similarly, the BNNeck improves the training process by including a Batch Normalization layer that leverages the goals of classification and metric learning losses, while the Last Stride trick adjusts the stride in the last down-sampling operation to obtain higher spatial resolution in the learned representation. Finally, the Center Loss reduces the intra-class variability by minimizing the distance of all samples of a specific class (i.e. multiple images of the same individual) from its centroid.

### 4.2. Metric Learning Losses

Different metric learning losses have been proposed in the past years. In this section, we present three widely accepted losses that are evaluated in the proposed strong baseline model. Specifically, these losses are the triplet loss [17], quadruplet loss [6] and margin sample mining loss (MSML) [38]. Despite the fact that all these losses reduce the distance between images of the same class (positive samples) and increase the distance between images of different classes (negative samples), they still have different working mechanisms. In fact, they employ different losses functions and the input samples are selected based on different strategies.

The triplet loss constructs triplets by fixing a sample as reference and, then, selecting a hard positive and negative image with respect to the reference. Differently, the quadruplet loss enhances the triplet loss by including in the loss a term that contemplates a negative pair of images with identities distinct from the reference image. Thus, the quadruplet is able to further reduce the inter-class similarities. Finally, the MSML constructs one hard positive and negative pairs per batch as the farthest positive samples and the closest negative samples, respectively.

## 5. Experiments

In this section, we first present the training and evaluation protocols. Section 5.1 evaluates the strong baseline model in the motorcycle ReID problem. Then, we present experiments evaluating two important issues in the object ReID literature: the distractors (Section 5.2) and the domain shift (Section 5.3). Finally, Section 5.4 evaluates the strong baseline in a dataset from literature. In the following experiments, we consider the BPReID, Motorcycle BPReID and the proposed MoRe datasets, described in Section 3.2.

**Training Protocol.** MoRe identities were divided in 1,913 for training and 1,914 for test. In the training stage, all images were resized to 256x256 pixels. As data augmentation techniques [29], we used random erasing, brightness transformation, horizontal flipping, rotation, translation and zooming, each of them with 50% of probability. Also, all images were converted to 32-bit floating point between [0,1] and normalized using ImageNet mean and standard deviation.

We first trained the feature extractor and classification layers for 120 epochs using only cross-entropy loss. Each training batch was built sampling two images for 16 random identities, in a total of 32 images per batch. For training, we started with the learning rate as $3.5 \times 10^{-4}$ and decreased by a factor of $10^{-1}$ at 40th and 70th epochs. The only exception occurred when evaluating the Warmup Learning Rate as described in [26]. Then, we added the metric learning loss in the learning process and trained the model for additional 120 epochs with the same learning rate scheme. When considering the metric learning, we constructed the batches by randomly sampling six identities, and for each identity sampling at most four images of that individual. Adam optimizer was used in all the training.

**Evaluation Protocol.** In the test set, for each pair of cameras, we used images from one camera as probe and from the other as gallery. We performed a multi-shot vs. multi-shot experiments and used the Euclidean distance between the learned feature representation. The only exception was when considering the BNNeck and the features after the Batch Normalization layer. In that case, we used the cosine distance due to its superior performance [26]. Experimental results are reported using the *rank-1* accuracy and mean Average Precision (mAP). For real applications, both metrics are important as the *rank-1* accounts for the capacity of the model to return the correct identity in the first position of the ranking of gallery images, while the mAP better measures the performance when considering all the correct gallery images (i.e. multiples ground truths).

### 5.1. Strong Baseline Evaluation

In this section, we evaluate the impact of the previously described tricks in the motorcycle ReID problem when considering the MoRe dataset. Besides, we assess the effect of distinct input shapes and aspect ratios. Finally, we investigate three widely used metric learning losses in the literature. Based on these results, we define a strong baseline for motorcycle ReID that researches can rely on to improve the state-of-the-art. More importantly, we present qualitative results showing the information learned by the strong baseline model to match the images.

**Tricks.** Table 3 presents the obtained experimental results when evaluating cumulatively the different tricks on the MoRe database. Based on these results, we notice that the direct application of the standard baseline from literature obtains suboptimal results. It is an issue as the baseline model struggles to highlight the still open problems in the motorcycle ReID. More importantly, we observed a boost in the both *rank-1* and mAP metrics when including the tricks. For instance, the *rank-1* and mAP increased 25.61 and 19.93 percentage points, respectively. These results demonstrates the importance of a strong baseline when proposing a novel database.

We credit the significant improvement in the model performance to some factors. The smoothing of the labels and the learning rate scheme reduced the overfitting and improved the convergence of the learned model. Differently, the last stride increased the spatial information and the center loss reduced the intra-class variations. Thus, they tackle problems as the low FPS and the background clutter present in the MoRe database.

| Model | r=1 | mAP |
|---|---|---|
| Standard Baseline | 57.80 | 64.45 |
| + Label Smoothing | 66.71 | 71.86 |
| + Warmup Learning Rate | 69.01 | 74.82 |
| + Last Stride + BNNeck | 72.15 | 76.72 |
| + Center Loss | **83.41** | **86.38** |

Table 3: The impact of adding tricks cumulatively to the Standard Baseline. The model is trained and tested on MoRe dataset.

**Input Shape.** Now, we evaluate the impact of the input shape on the obtained experimental results. Precisely, we consider the importance of the image resolution and the aspect ratio. To accomplish that, we consider the best model obtained in the previous section and the MoRe database.

Table 4 presents the achieved results when alternating the resolution with the aspect ratio fixed. These results show that reducing the image resolution to 128x128 pixels, we have a decrease in the performance as consequence of missing the fine-grained information. Similarly, the performance is worse when considering 320x320 pixels. It can be explained by the fact that most of the bounding box images present low resolution, as shown in Figure 2a.

| Input Shape | r=1 | mAP |
|---|---|---|
| 320x320 | 82.65 | 85.42 |
| 256x256 | **83.41** | **86.38** |
| 224x224 | 82.18 | 85.32 |
| 128x128 | 79.61 | 83.03 |

Table 4: The impact of the different image resolutions on the experimental results obtained in MoRe dataset.

Another important factor to consider when performing motorcycle ReID is the aspect ratio. Therefore, in Table 5, we evaluated different aspect ratios. In fact, we considered the image resolution of 256x256 as reference and increased (256x224), decreased (224x256) or selected a specific value of aspect ratio closer to the real distribution of the captured images (256x160), as shown in Figure 2b. Based on these results, we notice that the aspect ratio influences the performance of the method with worse results when using aspect ratio different from 1:1. In fact, these results corroborate with previous works in person ReID [26].

| Input Shape | r=1 | mAP |
|---|---|---|
| 256x256 (1.00) | **83.41** | **86.38** |
| 224x256 (1.14) | 81.91 | 85.33 |
| 256x224 (0.88) | 82.70 | 85.82 |
| 256x160 (0.63) | 81.44 | 84.93 |

Table 5: The influence of the different aspect ratios (between parenthesis) on the experimental results obtained in MoRe dataset.

**Metric Learning Losses.** Table 6 shows results regarding the different metric learning losses. Considering these results, we observed an improved performance of Quadruplet Loss when compared to the Triplet Loss and MSML losses. We connect these results to the advances of the Quadruplet loss when compared to the Triplet loss as additional inputs and terms in the loss function. More notably, as MSML obtains a single hard positive and negative pair per batch, it may be constrained to the same difficult pairs that may occur in realistic scenarios. Differently, the Quadruplet loss considers each image in a batch as reference image to obtain the hard pairs and, therefore, is more likely to see different pairs on the training.

**Explainable AI.** In the previous experiments, we presented an extensive evaluation of the different tricks that we employed to obtain a strong baseline for motorcycle ReID. Now, we go further and show the information that the model is learning to accomplish this task. We hope that these results provide insights to the research community in how to tackle the motorcycle ReID problem.

| Losses | r=1 | mAP |
|---|---|---|
| Triplet Loss | 81.09 | 84.19 |
| MSML | 82.73 | 85.92 |
| Quadruplet Loss | **83.41** | **86.38** |

Table 6: The influence of the different metric learning losses on the experimental results obtained in the MoRe dataset.


Figure 4: Ranking results for different probe images (i.e. each row) highlighted using xAI. The right matches are outlined in green.

We explored a visualization technique proposed by Stylianou *et al.* [32] that is designed for deep similarity networks. In short, this method is able to calculate the contribution of specific regions of the image (i.e. pixels) when matching two images. Likewise to the authors, we consider the layer immediately after the pooling operation (2048-D feature) to calculate the importance.

Figure 5 shows the obtained visualization results when considering pairs of probe and gallery that are correct (Figure 5a) and wrong (Figure 5b). According to these results, the model is using information from both the motorcycles and the motorcyclist to compute similarities. For instance, in the first and third columns of Figure 5a, the similarity is based mostly on the shorts and the motorcycle features as the trunk and front. Even though some parts of the images are not very discriminative (i.e. the motorcycle wheels), they allow the correct matching of probe and gallery images when combined with the information about the motorcyclists as illustrated in the second column of Figure 5a.

(a) Examples of *rank-1* identities that correspond to the probe. (b) Examples of *rank-1* identities that do not correspond to the probe.

Figure 5: Explainable AI (xAI) experiments obtained using MoRe dataset. The first row corresponds to the probe samples, while the second row shows the respective *rank-1* result. Regions are highlighted using the xAI method proposed in [32].

It is also interesting to observe some cases in which the model fails. For instance, in the second column of Figure 5b, we can observe that the model matches the car in the background with the gallery image. Differently, in the remaining columns, we observe that the wrong matches are very similar to the probe. Besides that, Figure 4 shows some ranking results highlighting the information employed by the model.

### 5.2. Distractors Experiments

In this section, we evaluated the robustness of the strong baseline model to the addition of distractors in the gallery set. These distractors correspond to a subset of individuals that were captured and that we did not have the correspondence in the probe set due to the low FPS of the surveillance cameras. Therefore, they share many characteristics with the gallery as the background and illumination conditions.

According to the results presented in the Figure 6, we notice that both the *rank-1* and mAP diminish as we include more distractors. For instance, the *rank-1* reduces more than 5 percentage points as we include 3,478 distractors in the gallery set.

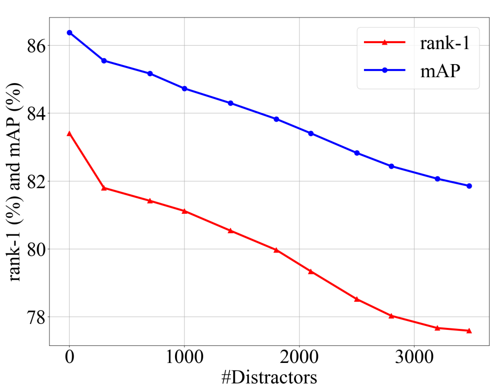



Figure 6: The impact of adding distractors in the gallery when evaluating the strong baseline in the MoRe data.

### 5.3. Domain Shift Experiments

A central problem when dealing with object re-identification is the poor performance of these models when the test set is different from the training ones (i.e. domain shift) [28, 36]. In this section, we evaluate the impact of the domain shift in the obtained experimental results in the MoRe dataset. In the following paragraphs, we provide experiments considering different camera pairs and different datasets as distinct domains, respectively.

**Cross-pairs Experiments.** Now, we provide an experiment to evaluate how the strong baseline model is able to generalize to unseen camera pairs. To accomplish that, we detach one camera pair (test set) for evaluation, while training the model on all the others camera pairs (training set).

Table 7 presents the achieved results for cross-pairs evaluation. Some pairs as the *pair01* and *pair03* present higher results due to the more similar camera viewpoints, as illustrated in Figure 3. Nonetheless, the model struggles to generalize to most of the unseen camera pairs. For instance, it reaches a *rank-1* of 40.33% when evaluated on *pair05*. We relate these poor results to a larger gallery set, the low-resolution images, the background clutter and the distinct viewpoint compared to the other pairs. Similarly, the *pair02* imposes some challenges to the model due to the low-resolution images. In fact, we hope that these results encourage researches to tackle the domain shift problem in the motorcycle ReID context.

**Cross-dataset Experiments.** In this section, we provide experimental results considering the domain shift between datasets captured in different scenarios (i.e., urban traffic vs campus). To accomplish that, we trained the strong baseline model in distinct datasets and evaluated on the test partition of the MoRe dataset. When training in the BPReID data, we followed the same protocol defined by the authors in [42].

The obtained experimental results are presented in Table 8. Based on these results, we can notice that the model

| Test Set | r=1 | mAP |
|---|---|---|
| pair01 | 74.18 | 81.40 |
| pair02 | 50.00 | 60.64 |
| pair03 | 85.38 | 84.84 |
| pair04 | 68.52 | 73.54 |
| pair05 | 40.33 | 47.27 |
| pair06 | 67.79 | 74.54 |

Table 7: Cross-pairs experiments in MoRe dataset.

trained on BPReID and evaluated in the MoRe dataset performs poorly. It can be related to the very different characteristics of images captured in urban traffic when compared to a university campus. Interestingly, even when training with both MoRe and BPReID datasets, we observed a drastic reduction in the performance. We attribute this result to the very different scenarios of MoRe and BPReID. More importantly, these result reiterate the importance of MoRe to the progress motorcycle ReID in urban traffic scenarios.

| Training Set | r=1 | mAP |
|---|---|---|
| BPReID | 16.94 | 23.08 |
| MoRe | **83.41** | **86.38** |
| BPReID+MoRe | 63.95 | 68.90 |

Table 8: Cross-dataset experiments. Different training sets are used to train the strong baseline model. Then, the model is evaluated in the test set of the MoRe dataset.

### 5.4. BPReid Experiments

In this section, we evaluate our strong baseline model in the Motorcycle BPReID database. To obtain a fairer comparison, we trained the strong baseline model using the BPReID training data (i.e. all categories) as defined by the authors [42]. Likewise, we consider the same evaluation protocol, which uses aggregation (max and mean pooling) for multi-shot evaluation and disregards the distractors.

As we focus on motorcycle ReID problem, we consider only the Motorcycle BPReID dataset when evaluating the methods. Table 9 presents the obtained experimental results for each camera pair of the BPReID dataset. Based on these results, we notice that the strong baseline surpasses all results reported in the state-of-the-art by a large margin. In fact, this boost in the performance links to the fact that the method proposed in [42] is based on handcrafted descriptors, while the strong baseline learns the descriptors from data using a deep learning model. More importantly, these results validate our model as a strong baseline for the motorcycle ReID task.

| Dataset | Yuan *et al*. [42] | Strong Baseline |
|---|---|---|
| cam1-2 | 50.80 | **91.99** |
| cam2-3 | 73.80 | **93.33** |
| cam3-5 | 73.70 | **97.61** |
| cam4-5 | 52.50 | **71.77** |
| cam5-6 | 29.50 | **77.66** |
| cam6-1 | 37.00 | **84.32** |

Table 9: Obtained mAP for the evaluated methods in each camera pair of the Motorcycles BPReID.

## 6. Conclusions

In this work, we proposed the MoRe dataset, the first large-scale dataset for motorcycle re-identification in urban traffic scenarios. In addition, we trained a deep learning model for the motorcycle ReID considering the best training practices from the object re-identification literature. The experimental results demonstrate that our strong baseline surpasses the state-of-the-art for a large margin. To further understand this strong baseline model, we highlighted the regions of the images used by the model to compute the similarity between images. Based on this analysis, we noticed that the model is able to incorporate information regarding the motorbikes, the motorcyclist vests and even additional clues as a backpack or a motorcycle trunk.

We also performed experiments that emphasized some open issues in the motorcycle ReID problem as the impact of distractors and the domain shift in the experimental results. Based on the results, we observed a reduction of more than 5 percentage points in the *rank-1* as we included 3,478 distractors in the gallery set. Furthermore, our domain shift results demonstrate a drastic reduction in the performance when training and evaluating the model in different environments (cross-dataset) and even in different pairs of cameras (cross-pairs).

Finally, we showed as xAI can be used as a tool to highlight where the models can be enhanced. For instance, we noticed that the learned model can be improved to better incorporate information about the motorcyclist helmet and the motorcycle model. We expect that the combination of a large-scale dataset, a strong baseline and a comprehensive experimental evaluation of open issues in the motorcycle ReID problem can boost the research in this important problem.

**Acknowledgement.** The authors would like to thank the National Council for Scientific and Technological Development – CNPq (Grants 438629/2018-3 and 309953/2019-7) and the Minas Gerais Research Foundation – FAPEMIG (Grants APQ-00567-14 and PPM-00540-17).

## References


[1] Amina Adadi and Mohammed Berrada. Peeking inside the black-box: A survey on explainable artificial intelligence (xai). *IEEE Access*, 6:52138–52160, 2018.

[2] Raja Muhammad Saad Bashir, Muhammad Shahzad, and MM Fraz. Vr-proud: Vehicle re-identification using progressive unsupervised deep architecture. *Pattern Recognition*, 90:52–65, 2019.

[3] Amran Bhuiyan, Yang Liu, Parthipan Siva, Mehrsan Javan, Ismail Ben Ayed, and Eric Granger. Pose guided gated fusion for person re-identification. In *The IEEE Winter Conference on Applications of Computer Vision*, pages 2675–2684, 2020.

[4] Aphinya Chairat, Matthew Dailey, Somphop Limsoonthrakul, Mongkol Ekpanyapong, and Dharma Raj KC. Low cost, high performance automatic motorcycle helmet violation detection. In *The IEEE Winter Conference on Applications of Computer Vision*, pages 3560–3568, 2020.

[5] Hao Chen, Benoit Lagadec, and Francois Bremond. Learning discriminative and generalizable representations by spatial-channel partition for person re-identification. In *The IEEE Winter Conference on Applications of Computer Vision*, pages 2483–2492, 2020.

[6] Weihua Chen, Xiaotang Chen, Jianguo Zhang, and Kaiqi Huang. Beyond triplet loss: A deep quadruplet network for person re-identification. 07 2017.

[7] Ruihang Chu, Yifan Sun, Yadong Li, Zheng Liu, Chi Zhang, and Yichen Wei. Vehicle re-identification with viewpoint-aware metric learning. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 8282–8291, 2019.

[8] Kunal Dahiya, Dinesh Singh, and C Krishna Mohan. Automatic detection of bike-riders without helmet using surveillance videos in real-time. In *2016 International Joint Conference on Neural Networks (IJCNN)*, pages 3046–3051. IEEE, 2016.

[9] Patricia C Dischinger, Gabriel E Ryb, Shiu M Ho, and Elisa R Braver. Injury patterns and severity among hospitalized motorcyclists: a comparison of younger and older riders. In *Annual Proceedings/Association for the Advancement of Automotive Medicine*, volume 50, page 237. Association for the Advancement of Automotive Medicine, 2006.

[10] Jorge E Espinosa, Sergio A Velastin, and John W Branch. Motorcycle detection and classification in urban scenarios using a model based on faster r-cnn. 2018.

[11] J. E. Espinosa, S. A. Velastín, and J. W. Branch. Detection of motorcycles in urban traffic using video analysis: A review. *IEEE Transactions on Intelligent Transportation Systems*, pages 1–16, 2020.

[12] Jorge Ernesto Espinosa Oviedo. Detection and tracking of motorcycles in urban environments by using video sequences with high level of oclussion. *Escuela de Sistemas*, 2019.

[13] Jorge Ernesto Espinosa-Oviedo, Sergio A Velastín, and John William Branch-Bedoya. Espinet v2: a region based deep learning model for detecting motorcycles in urban scenarios. *DYNA*, 86(211):317–326, 2019.

[14] Xing Fan, Wei Jiang, Hao Luo, and Mengjuan Fei. Sphereid: Deep hypersphere manifold embedding for person re-identification. *Journal of Visual Communication and Image Representation*, 60, 04 2019.

[15] Bing He, Jia Li, Yifan Zhao, and Yonghong Tian. Part-regularized near-duplicate vehicle re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3997–4005, 2019.

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[17] Alexander Hermans, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification. 03 2017.

[18] Chi-Kien Huynh, Thanh-Sach Le, and Kazuhiko Hamamoto. Convolutional neural network for motorbike detection in dense traffic. In *2016 IEEE Sixth International Conference on Communications and Electronics (ICCE)*, pages 369–374. IEEE, 2016.

[19] Sultan Daud Khan and Habib Ullah. A survey of advances in vision-based vehicle re-identification. *Computer Vision and Image Understanding*, 182:50–63, 2019.

[20] Amena Khatun, Simon Denman, Sridha Sridharan, and Clinton Fookes. Semantic consistency and identity mapping multi-component generative adversarial network for person re-identification. In *The IEEE Winter Conference on Applications of Computer Vision*, pages 2267–2276, 2020.

[21] Devinder Kumar, Parthipan Siva, Paul Marchwica, and Alexander Wong. Unsupervised domain adaptation in person re-id via k-reciprocal clustering and large-scale heterogeneous environment synthesis. In *The IEEE Winter Conference on Applications of Computer Vision*, pages 2645–2654, 2020.

[22] Mengliu Li, Han Xu, Jinjun Wang, Wenpeng Li, and Yongli Sun. Temporal aggregation with clip-level attention for video-based person re-identification. In *The IEEE Winter Conference on Applications of Computer Vision*, pages 3376–3384, 2020.

[23] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

[24] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.

[25] Yihang Lou, Yan Bai, Jun Liu, Shiqi Wang, and Ling-Yu Duan. Embedding adversarial learning for vehicle re-identification. *IEEE Transactions on Image Processing*, 28(8):3794–3807, 2019.

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

[27] Neeraj Matiyali and Gaurav Sharma. Video person re-identification using learned clip similarity aggregation. In *The IEEE Winter Conference on Applications of Computer Vision*, pages 2655–2664, 2020.

[28] Vishal M Patel, Raghuraman Gopalan, Ruonan Li, and Rama Chellappa. Visual domain adaptation: A survey of recent advances. *IEEE signal processing magazine*, 32(3):53–69, 2015.

[29] Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):60, 2019.

[30] Felix Wilhelm Siebert and Hanhe Lin. Detecting motorcycle helmet use with deep learning. *Accident Analysis & Prevention*, 134:105319, 2020.

[31] Romuere Silva, Kelson Aires, Rodrigo Veras, Thiago Santos, Kalyf Lima, and André Soares. Automatic motorcycle detection on public roads. *CLEI Electronic Journal*, 16(3):4–4, 2013.

[32] Abby Stylianou, Richard Souvenir, and Robert Pless. Visualizing deep similarity networks. In *2019 IEEE winter conference on applications of computer vision (WACV)*, pages 2029–2037. IEEE, 2019.

[33] 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 *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 480–496, 2018.

[34] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2818–2826, 2016.

[35] C Vishnu, Dinesh Singh, C Krishna Mohan, and Sobhan Babu. Detection of motorcyclists without helmet in videos using convolutional neural network. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 3036–3041. IEEE, 2017.

[36] Mei Wang and Weihong Deng. Deep visual domain adaptation: A survey. *Neurocomputing*, 312:135–153, 2018.

[37] Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao. A discriminative feature learning approach for deep face recognition. In *European Conference on Computer Vision*, pages 499–515. Springer, 2016.

[38] Qiqi Xiao, Hao Luo, and Chi Zhang. Margin sample mining loss: A deep learning based method for person re-identification. 10 2017.

[39] Jianhua Xu. Drive-away policing and situational crime prevention in china: an analysis of motorcycle ban (jinmo) policy in guangzhou. *International Journal of Offender Therapy and Comparative Criminology*, 56(2):239–264, 2012.

[40] 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.

[41] Ye Yuan, Wuyang Chen, Tianlong Chen, Yang Yang, Zhou Ren, Zhangyang Wang, and Gang Hua. Calibrated domain-invariant learning for highly generalizable large scale re-identification. In *The IEEE Winter Conference on Applications of Computer Vision*, pages 3589–3598, 2020.

[42] Yuan Yuan, Jianran Zhang, and Qi Wang. Bike-person re-identification: A benchmark and a comprehensive evaluation. *IEEE Access*, PP:1–1, 09 2018.

[43] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *Proceedings of the IEEE international conference on computer vision*, pages 1116–1124, 2015.

[44] Liang Zheng, Yi Yang, and Alexander G Hauptmann. Person re-identification: Past, present and future. *arXiv preprint arXiv:1610.02984*, 2016.

[45] Zhedong Zheng, Tao Ruan, Yunchao Wei, and Yezhou Yang. Vehiclenet: Learning robust feature representation for vehicle re-identification. In *CVPR Workshops*, volume 2, page 3, 2019.

[46] Yi Zhou and Ling Shao. Aware attentive multi-view inference for vehicle re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6489–6498, 2018.

[47] Yi Zhou and Ling Shao. Vehicle re-identification by adversarial bi-directional lstm network. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 653–662. IEEE, 2018.