Concept for Anonymous Re-Identification

1st Robert Kathrein
University of Applied Sciences Kufstein
Josef Ressel Zentrum - Vision2Move
University of Passau, Germany
Kufstein, Austria
robert.kathrein@fh-kufstein.ac.at

2nd Oliver Zeilerbauer, 3rd Johannes Larcher,
5th Mario Döller
University of Applied Sciences Kufstein
Josef Ressel Zentrum - Vision2Move
Kufstein, Austria
{oliver.zeilerbauer,johannes.larcher,mario.doeller}
@fh-kufstein.ac.at

4th Kris Raich
University of Passau, Germany
Kufstein, Austria
kris.raich@gmail.com

***Abstract*—This work presents a novel concept for object Re-Identification (ReID) based on multiple object detecting visual sensors with focus on data privacy and ethical usage of Artificial Intelligence (AI) enhanced deployments. By challenging the current state of the art solutions for ReID, the proposed concept is optimising individual privacy in terms of data protection and transportation. With attention to the laws, covered by the General Data Protection Regulation (GDPR) inside the European Union, AI driven applications especially with deployments in public places are regularly prohibited. Within this paper announced methods all personal data neither leaves the secured sensors, nor gets stored on any media at any point of time - which is a requirement for GDPR. Furthermore, most of the currently used methods are compared with temporal finite datasets in retrospect. Our vision is to design a GDPR compliant platform for near real time ReID that supports contemporaneous auditing and not after the fact.**

I. Introduction

In recent years, the field of computer vision has witnessed remarkable advancements, particularly in the area of object detection and tracking. One of the fundamental challenges in this domain is the task of re-identifying objects across multiple cameras or sensors, commonly known as the Re-Identification problem [1]. ReID plays a crucial role in various applications, such as traffic management, video surveillance or crowd monitoring in complex environments (e.g. events, fire service, etc.). The ReID problem arises when multiple cameras or sensors observe the same or partially overlapping scene, capturing objects from different viewpoints, angles, and under varying lighting conditions. These differences in viewpoints and environmental factors often lead to significant variations in appearance, making it difficult to accurately track and identify objects across different sensors. Moreover, variations in pose, clothing, occlusions, and partial appearances further exacerbate the challenge. In a multi-object multi-sensor tracking scenario, the goal is to develop robust algorithms and techniques that can associate objects across different sensors, enabling reliable tracking and identification. This scenario involves processing data from multiple cameras, depth sensors, LiDAR, or any other sensing modalities, with the aim of establishing correspondences between objects appearing in different sensor streams. To address the ReID problem in this complex setting, researchers and engineers have leveraged advancements in deep learning, computer vision, and pattern recognition. Deep neural network architectures, such as Siamese networks [2], [3], triplet networks [4], and generative adversarial networks (GANs) [5], have shown promising results in learning discriminative representations for individual identities. In most of these research initiatives, data protection, privacy concerns and ethical topics are not considered. However, especially in Europe the General Data Protection Regulation (GDPR) act plays an important role. Therefore, this paper is focused on privacy and data protection issues in a multi-object multi-sensor tracking system. In general, most research is exploring algorithms and methods where the visual data streams of all sensors are analyzed in one stack, neglecting privacy or data protection concerns. While this is acceptable for some use cases, especially the European law demands for active privacy and data protection support. Based on that, concepts and research are needed in order to tackle this important ethical area. This paper proposes an universal concept for anonymous ReID systems and highlights, the individual components and steps. The paper is structured as follows: Section II gives a broad overview of various methods and technologies used to address the occurring problems related to ReID. Afterwards in section III the proposed concept is discussed in terms of data privacy and is juxtaposed to actual procedures. Also, several theorems are suggested, a system has to fulfil to comply with GDPR within this part of the work. Finally conclusions and open research topics are discussed in section IV.

II. Related Work

Several very recent surveys on ReID [7] or the subproblem Multi Object Tracking (MOT) [8] have been proposed. In the following, a short overview of the main findings concerning anonymous ReID systems is given.

A. *Definition for Re-Identification*

The topic of ReID in recent research is mostly based on person ReID or vehicle ReID and accomplished by utilising single RGB image frames of pre-recorded video data. For the purpose of the proposed concept this section sets the scope and explicates those inconsistencies to archive a status quo. As for the definition of the term itself, most researchers agree that detected objects of different sensors are matched. However the community around this subject is divided on specific

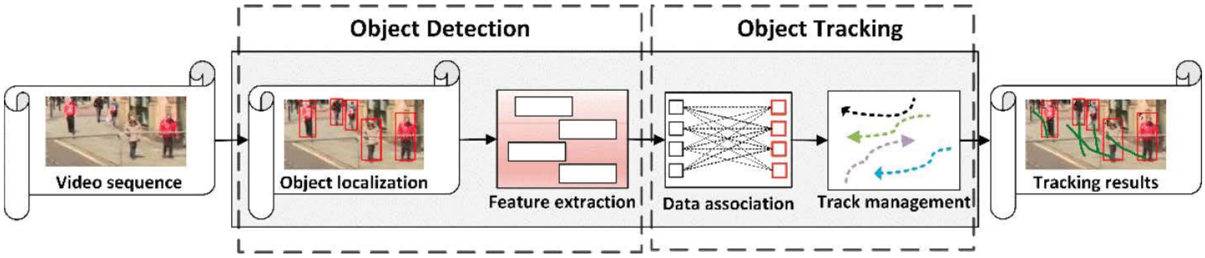

Fig. 1. Generic Steps to accomplish ReID in one sensor (see [6])

definitions. Some authors define ReID as a process assuming only non-overlapping field of views [9], [10]. Whereas other definitions incorporate overlapping field of views as well [11], [12].

B. Steps of a Re-ID System

Figure 1 presents the most important steps of a MOT system. In general, this result is the outcome of the processing pipeline within one camera sensor. This is followed by the interconnection of the individual outcomes of several camera processing pipelines where the same objects are identified and numbered (see Figure 2).

In the following, recent scientific results are introduced. Please note, we do not focus on object detection approaches such as DETR [13] or Swin Transformer [14] as we identify this as a separate research question to be solved. The starting point of the ReID process is a 100 percent object detection performance.

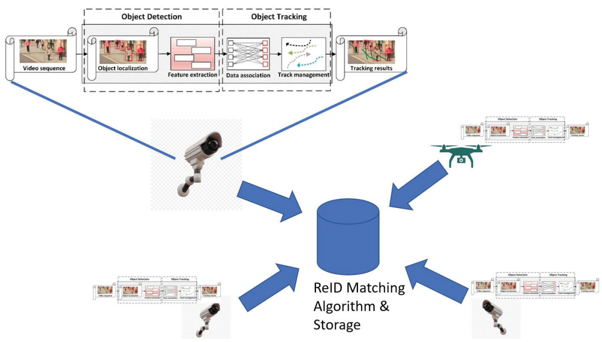

Fig. 2. Generic Steps to accomplish ReID over multiple sensors

C. Single-Sensor Multi-Object Tracking

Modern MOT methods can be separated into two main categories: tracking-by-detection [15], [16], [16]–[21] and joint-detection-tracking approaches [22]–[26]. Tracking-by-detection methods involve the first step of detecting the objects and then associating them based on cues of appearance and motion. Due to the advancements in object detection techniques, these methods have been successful in dominating the MOT task for an extended period. The SORT [21] algorithm has an implementation of a Kalman filter algorithm for motion based MOT. Its improvement, the DeepSORT [20] is a state-of-the-art algorithm which successfully incorporates appearance features into object association. In comparison Joint-detection-tracking methods, such as JDE [26] and FairMOT [23], incorporate appearance embedding or motion prediction into detection frameworks, thus achieving comparable performance with low computational costs. However, these joint trackers face the challenge of competition between the individual parts, which limits their tracking performance.

D. Re-identification

Re-identification is a crucial component in multi-camera traffic flow analysis, aiming to identify vehicles across different cameras. Convolutional Neuronal Network (CNN)-based ReID methods have shown strong feature representation, employing various loss functions (triplet loss [27], online instance matching loss [28], circle loss [29]) to learn discriminative features. A model trained with triplet loss is, based on a distance metric, advised to view pairs of the same samples closer to each other than different ones [27]. Online instance matching loss is used for joint-detection-tracking approaches to combine detection and re-identification by separating between three classes (labeled identities, unlabeled identities and backclutter) and calculating the cosine distance of the features of labeled and unlabeled identities in a circular queue [28]. Upon obtaining the extracted features, the final stage of the vehicle ReID process entails inter-camera association. This involves subjecting trajectories built from the acquired features to a matching algorithm resulting in a system-wide consistent ID. Recent works propose the usage of traffic rules, spatial-temporal constraints, to reduce search space as well as accelerating matching algorithms like greedy algorithms [30] and hierarchical clustering [31].

E. State-of-the-art performances

In the past decade, the Re-Identification task got more and more attention. With the advancement of Neural Networks and AI-based technologies the accuracy of such tasks improved from year to year. According to paperswithcode.com the most accurate Person-re-identification approaches, depending on mean Average Precision (mAP), peak at 98.3% [32], 97.1% [33] and 86.5% [34] mAP on the Market-1501 [35], DukeMTMC-reID [36] and MSMT17 [37] dataset respectively. However, the accuracy drops to 27% [38] mAP on the Market-1501-C dataset [39], which consists of algorithmically generated corruptions like Noise: Gaussian, speckle; Blur: defocus, frosted glass, motion, zoom; Weather: snow, fog, rain;

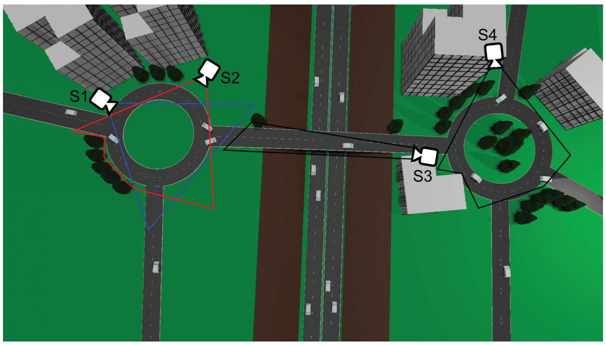

Fig. 3. Sample model of an common traffic situation, where two connected roundabouts are connected over an highway.

and Digital: contrast, pixel, etc. with different severity levels, reflecting the difficulty of real-life problems. Benchmarking those difficult settings Minghui Chen et. al [39] comes to the conclusion transformer based models are more robust than CNN ones.

In a similar field, vehicle-re-identification (VReID), which has specific problems due to small inter-variability (different car, but similar model and color) and high intra-variability (same car, front/back view difference) of the data, similar challenges were made. Although VReID got more attention in recent years there is less published and reproducible code out there than in person-re-identification (PReID). A well-known dataset for VReID is the Veri-776 dataset [40]. According to paperswithcode.com the best approaches reach 88.0%, 87.1% and 83.4% mAP respectively [41]. While there are public datasets for standard situations, there is no extensive research in artificially corrupted datasets, like the Market-1501-C dataset [39] from PReID, yet.

Another known challenge series is the AI-City-Challenge series. The last AI-City-Challenge with a mAP score including a ReID-problem was Track1 in the 2021 AI-City-Challenge. The best performing team achieved a mAP score of 74.45% [42]. To fully understand the complete scope of ReID additional metrics need to be accounted for in official competition. Despite high mAP-scores in some cases it is still a challenging task far from solved, especially in real-life applications.

To the authors best knowledge, contemporary ReID approaches do not prioritize feature selection for the purpose of guaranteeing anonymous ReID in the current scientific literature. Consequently, we intend to elaborate on this concept in the following chapter.

III. Conceptual Design of anonymous ReID systems

Figure 3 displays a common traffic situation captured by multiple cameras with partially to fully overlapping observing areas. In this example solving the ReID problem is a necessity for a traffic management system. To tackle those problems, our conceptual design for an anonymous ReID system is defined as follows.

Let $\mathbf{S} = (S_1, S_2, \ldots, S_n)$ be a n-tuple of sensors (e.g. RGB-cameras) and S_i the i-th sensor, with $i, n \in \mathbb{N}$ and $1 \leq i \leq n$. Without loss of generality (WLOG) one generic sensor pipeline and the connection from sensors to a collective database is described in the following.

From an arbitrary sensor S_i, we get the sensor output (e.g. a frame from a RGB-camera), denoted as O_i. On the given sensor output O_i, object detection models (e.g. a YOLO-Model) are applied to detect objects present in the current sensor output. From a sensor output O_i we extract $D_i = \{d_1, d_2, \ldots, d_k\}$, with $k \in \mathbb{N}$ where D_i represents the set of detected objects in sensor output O_i, and d_j represents the j-th detected object (e.g. an array with a bounding box, a confidence score and an object class), with $1 \leq j \leq k$. With the newly gained information D_i, advanced feature extraction is performed on the sensor output O_i to obtain m amount of feature descriptors for the set of detected objects D_i. For the set of detected objects D_i we extract $F_i \in \mathbb{R}^{k \times m}$, where F_i represents the feature matrix obtained from the sensor output O_i, after getting the information of the detected objects D_i, and $f_{j,l} \in F_i$ represents the feature descriptor associated with the j-th detected object and its l-th feature descriptor. So every column holds one specific feature descriptor (e.g. color) of every detected object and every row holds all the feature descriptors of one detected object d_j.

With the feature matrices of different sensor outputs, a matching algorithm then performs ReID. To improve reliability one matching algorithm can be performed before connecting the data of multiple sensors and another one thereafter. Therefore the feature matrices can get improved with a temporary ID on a single sensor to improve the ReID results of a multiple-sensor-system. As data protection is important to uphold, the first matching algorithm could be processed in one camera with more sensible data to improve results, while guaranteeing the raw data (e.g. RGB-frames) never leaving the camera. Another approach suggested by Dangwal et al. [43] mitigates reverse engineering attacks on local feature descriptors compromising accuracy while improving data anonymity. To preempt such reverse engineering attacks, it is crucial to select suitable features, to ensure even in the event of a successful attack, sensitive data remains protected. We expand on this part in section III-A. For further considerations on our concept we define the following terms as follows.

Definition 1 (personal data): In accordance with the provisions and guidelines stipulated by the General Data Protection Regulation GDPR [44] the following statements are delineated:

- "Personal data is any information that relates to an identified or identifiable living individual. Different pieces of information, which collected together can lead to the identification of a particular person, also constitute personal data."
- "Personal data that has been de-identified, encrypted or pseudonymised but can be used to re-identify a person remains personal data and falls within the scope of the GDPR."

- "Personal data that has been rendered anonymous in such a way that the individual is not or no longer identifiable is no longer considered personal data. For data to be truly anonymised, the anonymisation must be irreversible."

Definition 2 (named entity): Related to definition 1 we define a named entity as: an individual living which can be identified with an uniquely associated name. Additionally, it encompasses any object that, when combined with certain distinguishing features (e.g. licence plates, document numbers, registration numbers, etc.), facilitates the unique identification of such individual living.

Definition 3 (field of view): As discussed in the II-A subsection, the scientific community is divided into overlapping and non-overlapping ReID environments. Our concept supports both kind of views as a ReID system. This is especially important to cover larger areas such as whole cities or large malls.

Based on the formalized ReID scenario and definitions, we enumerate the following individual principles which define a GDPR compliant $\oplus ReID$ system:

Theorem 1: An *anonymous* ReID system $\oplus ReID$ is defined as a system where tracked objects can not be identified as named entities.

Theorem 2: An *anonymous* ReID system $\oplus ReID$ consists of in-sensor iS_i processing and out-sensor oS_i processing. Features $f_{j,l} \in F_i$ that leave the in-sensor processing are not capable to restore the whole scenery. Raw sensor data is not allowed to be transferred.

Theorem 3: The anonymity of named entities must not be violated even through enrichment with external interpolated features.

Theorem 4: An *anonymous* $\oplus ReID$ feature set is defined that their granularity is set that neither a single feature nor their cross section can pin point a named entity.

Theorem 5: An *anonymous* ReID system $\oplus ReID$ is defined as a system where positions and those calculations are accomplished by using an *anonymous* location reference system.

In the following subsections, the individual topics according to the stated theorems are discussed in more detail.

A. Feature Restoring

As specified beforehand, one main focal point of an anonymous ReID system is the set of features that are transferred between sensor and further processing engines. In this context, we can define two sets of features, namely anonymous feature vs. ReID optimized feature. Figure 4 demonstrates this issue. An anonymous feature is defined as a feature which does not allow to restore the original scenery. Whereas a ReID optimized feature has its own focus on optimising the matching algorithm of entities.

As an example, we can propose a simple color histogram of an object as an anonymous feature which does not allow to restore the original RGB image. However, the color histogram has a very poor ReID matching capability. In contrast to that, a licence plate detector has a very strong ReID matching capability for cars but a very poor anonymous capability. Of course, the optimal solution for an anonymous ReID system is the selection of features that are in the union of these two sets. A detailed research in this direction has been elaborated by [43], where features are evaluated according to its characteristic to restore RGB images.

B. Linkability as Anonymity Benchmark

Linkability pertains to the capacity to establish connections between anonymized data and external data sets or supplementary information in order to re-identify named entities. Evaluating linkability entails assessing the ease of degree with which the anonymized data can be linked to external sources. A lower level of linkablity signifies a heightened level of data anonymity, indicating the increased difficulty in associating the anonymized data with specific individuals through external data sources. This evaluation provides insights into the effectiveness of data anonymization techniques in preventing the ReID of individuals through linkage with external information.

To measure the anonymity of the linkage of internal features and external features, we introduce the metric "Level of Anonymity (A)". A takes the value of 1, if given a collection of features, where it is impossible to ascertain the object's identification as a named entity, when connected with external information. By assigning A the value of 1, it signifies a heightened level of anonymity, implying that no combination of features can yield sufficient evidence to associate the object with a named entity.

Let F be a set of features (extracted within a sensor such as the pose of a person, etc.), let F_e be a set of external features (e.g., social media data, etc.) which uniquely identify a named entity. The set including all such sets F_e, which uniquely identify a named entity, is denoted as EF. Let $A \in \{0, 1\}$ be a binary variable, denoting the level of anonymity associated with an entity. A = 1 if the following statement is true:

$$\forall F_e \in EF : F \cap F_e \neq F_e,$$

otherwise A = 0.

It must be noted that achieving an accurate implementation of the mathematical formula in practice is challenging. A suitable trade-off between anonymity and features needs to be identified. While this is theoretically correct, one has to consider the scope of the ReID system. The smaller the scope becomes, the more features can be taken into consideration to influence the level of anonymity.

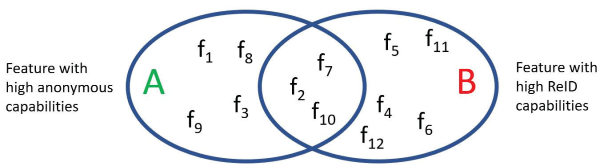

Fig. 4. Anonymous feature vs. ReID optimized feature

C. Process of ⊕ ReID system

To fix the issues discussed in section I and partially in section II the novel concept is introduced within this section. For the sake of explanation we use a hypothetical use case and sensor array - but the concept is not limited by the scenario and should be universally applicable. A representation for this hypothetical use case is shown in figure 3. During the research for ReID we could identify and differentiate between several sequential steps which are essential to perform ReID. To accomplish an anonymous approach, we suggest the following deviations to the procedure.

1) *Data gathering*: In the beginning data has to be gathered which lays within the scope of the regarded region. At this point the produced data of multiple sensors (e.g. S1, S2 of our example) can be ambiguous, in our hypothetical example we have multiple video streams of a generic traffic situation as data source. For the reason of data privacy this stream of data should not be sent to any external server or be exchanged between sensors. But processed directly at the recording sensor or a directly connected companion processing device respectively. We call this in-sensor (iS_i) processing.

2) *Object detection*: Depending on the data, a specialized algorithm or deep learning process is required for detecting the contained objects. Whereas this may consist of audio snippets, bounding boxes e.g., in our example we would use a YOLO object detection CNN to get the bounding boxes of vehicles. This detection should be conducted in-sensor (iS_i) where no public services should be used for detection.

3) *Object classification / metadata interpolation*: This step is required to enrich the detected objects with additional metadata. Frame by frame all detections are fed through multiple classification algorithms or neuronal networks in a parallel manner. These classifiers need to be constructed in a modular design for effortless exchanging depending on the situation and the sensors capabilities. In our example we would use analysing methods for location (camera based), vehicle type (CNN), vehicle color (histogram analysis). The metadata analysis should again be performed in-sensor without any external resources to comply with data privacy.

4) *Re-Identification algorithm*: As all previous steps are fulfilled in-sensor (iS_i) the ReID step itself is accomplished on a bigger centralised server (oS_i). Therefore, the data of the detected objects has to get transmitted to this service. We propose to utilise a kind of data broker service for data exchange. Needless to say, the transferred messages are exclusively composed by corresponding metadata of detected objects. A message (M) can be defined as the combination of the sensor's id (i) a timestamp (t) and finally the feature set of the corresponding timestamp ($F_i(t)$).

$$M = (i, t, F_i(t))$$

Depending on calculation and transmission speed, multiple combinations of timestamps and feature sets can be collected and send within one message.

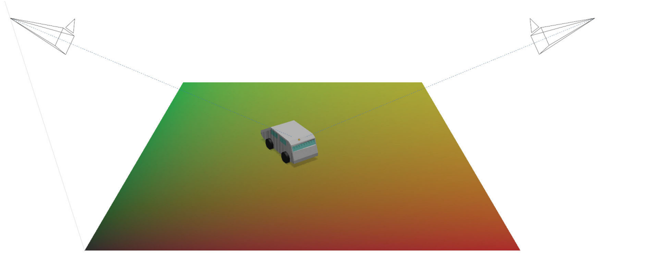

Fig. 5. Visualisation of the proposed camera based anonymous location reference system

D. Anonymous location reference system

The two most anonymous parameters in the case of named entities, but features with great impact on algorithms for matching objects are the objects exact position at what exact point of time. An exact localisation of an object by using only a camera sensor is a challenging task. For this problem we want to propose the following method as a possible solution. The cameras properties consist of the intrinsic sensor values like tilt angle, focal length, zoom level and so on. Combined with the global orientation, the sensors exact GPS location or the relative position to other sensors and the sensors mounting height; all information for an exact pixel wise localisation, is given. To keep privacy issues, we propose, to span a two dimensional color space gradient (i.e. red and green channel of RGB color space) extending the whole observed area as visualised in figure 5. For larger areas, a grid like system similar to raster map tile systems can be introduced to retain a fine granularity where the third color channel could be used as tile identifier. The sensor system has to render an image of the view ports color gradient only once and use it as a overlay mask. The great benefit of such a system is that regarding of viewing angle, field of view, and camera position, not a single calculation is required for localisation. In terms of data anonymity any subsequent calculations for distances, object height and positions can be based on this alternative location system. Therefore without explicit knowledge of the exact sensor positions, the real positions of detected objects can be used for calculations in an obfuscated environment.

IV. Conclusion and Open Research topics

In this paper the benchmarks of vehicle and person ReID approaches are presented and the privacy issues of the status quo are mentioned. Furthermore a conceptual design to improve the data anonymity of ReID systems in the future is outlined in various steps. The problem-structure gets defined and future goals of the practical realisation of the concept are articulated. Furthermore, the term Linkability is defined to recognize if acquired features are anonymous. Finally a method to guarantee anonymous location tracking is proposed. For future work, extensive research of anonymous features in combination with high performance is needed. Additionally the automatic generation of anonymous training data would improve the performance as the data gathering of real-life applications are difficult to acquire without violating the GDPR. Finally, the anonymous localisation data of objects without displaying GPS-coordinates has jet to be implemented.

Some open research issues of ReID are as follows:

- Classification of feature sets - anonymity: Future research in this domain necessitates conducting an extensive investigation into the nature of features, delving deeper into their implications concerning anonymity. Furthermore, thorough exploration of privacy-preserving algorithms is crucial, with a focus on comprehending the effects of diverse anonymisation techniques.
- Comparison of ReID Matching algorithms: In order to extend the current body of knowledge, it is essential to undertake a contemporary research building upon previous surveys [7], [45] that have compared re-identification matching algorithms. This endeavor should employ modern state-of-the-art benchmarks to assess the advancements made in recent years and determine the most efficacious approaches for ReID.
- Implementation of location reference system: As pointed out in section III-D, the objects location is a potent feature for ReID. The proposed method for a camera based reference system has yet to be implemented and validated against anonymity and performance.

ACKNOWLEDGEMENT

The marked authors are funded by Josef Ressel Center for multimedia analysis in the mobility domain (Vision2Move). The financial support by the Austrian Federal Ministry for Digital and Economic Affairs and the National Foundation for Research, Technology and Development and the Christian Doppler Research Association is gratefully acknowledged.

REFERENCES

[1] Vittorio Murino and M. Cristani. Re-identification of persons in videos and images. 2013.

[2] Meng Zheng, Srikrishna Karanam, Ziyan Wu, and Richard J. Radke. Re-identification with consistent attentive siamese networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5728–5737. IEEE, 2019. DOI: 10.1109/CVPR.2019.00588.

[3] Dahjung Chung, Khalid Tahboub, and Edward J. Delp. A two stream siamese convolutional neural network for person re-identification. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017. IEEE.

[4] Zengming Tang and Jun Huang. Harmonious multi-branch network for person re-identification with harder triplet loss. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 18(4):1–21, 2022.

[5] Hamed Alqahtani, Manolya Kavakli-Thorne, and Charles Z. Liu. An introduction to person re-identification with generative adversarial networks. *arXiv:1904.05992*, 2019.

[6] Anton Milan, Laura Leal-Taixe, Ian Reid, Stefan Roth, and Konrad Schindler. Mot16: A benchmark for multi-object tracking. *arXiv:1603.00831v2*, 2016.

[7] Mang Ye, Jianbing Shen, Gaojie Lin, T. Xiang, Ling Shao, and S. Hoi. Deep learning for person re-identification: A survey and outlook. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44:2872–2893, 2020.

[8] Shuman Guo, Shichang Wang, Zhenzhong Yang, Lijun Wang, Huawei Zhang, Pengyan Guo, Yuguo Gao, and Junkai Guo. A review of deep learning-based visual multi-object tracking algorithms for autonomous driving. *Applied Sciences*, null:null, 2022.

[9] Su Huynh, Nam-Hoang Nguyen, Ngoc-Thanh Nguyen, Vinh Nguyen, C. Huynh, and Chuong H. Nguyen. A strong baseline for vehicle re-identification. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, null:4142–4149, 2021.

[10] Tsung-Wei Huang, Jiarui Cai, Hao Yang, Hung-Min Hsu, and Jenq-Neng Hwang. Multi-view vehicle re-identification using temporal attention model and metadata re-ranking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019.

[11] Zhibin Zhao, Yinuo Yao, Huicheng Chen, Jun Wang, and Yongming Yan. Research on cross-camera person re-identification using overlapping field of view. pages 699–704, 2021.

[12] Yu-Jhe Li, Xinshuo Weng, Yan Xu, and Kris M. Kitani. Visio-temporal attention for multi-camera multi-target association. pages 9834–9844, October 2021.

[13] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. *ArXiv*, abs/2005.12872:null, 2020.

[14] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and B. Guo. Swin transformer: Hierarchical vision transformer using shifted windows. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, null:9992–10002, 2021.

[15] Yifu Zhang, Pei Sun, Yi Jiang, Dongdong Yu, Zehuan Yuan, Ping Luo, Wenyu Liu, and Xinggang Wang. Bytetrack: Multi-object tracking by associating every detection box. *Applied Sciences*, 2021.

[16] F. Yu, Wenbo Li, Quanquan Li, Yu Liu, Xiaohua Shi, and Junjie Yan. Poi: Multiple object tracking with high performance detection and appearance feature. *ArXiv*, abs/1610.06136:null, 2016.

[17] Jiangmiao Pang, Linlu Qiu, Xia Li, Haofeng Chen, Qi Li, Trevor Darrell, and F. Yu. Quasi-dense similarity learning for multiple object tracking. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, null:164–173, 2020.

[18] Yunhao Du, Yang Song, Bo Yang, and Yanyun Zhao. Strongsort: Make deepsort great again. *ArXiv*, abs/2202.13514:null, 2022.

[19] Nir Aharon, Roy Orfaig, and B. Bobrovsky. Bot-sort: Robust associations multi-pedestrian tracking. *ArXiv*, abs/2206.14651:null, 2022.

[20] Jinkun Cao, Xinshuo Weng, Rawal Khirodkar, Jiangmiao Pang, and Kris Kitani. Observation-centric sort: Rethinking sort for robust multi-object tracking. *ArXiv*, abs/2203.14360:null, 2022.

[21] A. Bewley, ZongYuan Ge, Lionel Ott, F. Ramos, and B. Upcroft. Simple online and realtime tracking. *2016 IEEE International Conference on Image Processing (ICIP)*, null:3464–3468, 2016.

[22] Xingyi Zhou, V. Koltun, and Philipp Krähenbühl. Tracking objects as points. *ArXiv*, abs/2004.01177:null, 2020.

[23] Yifu Zhang, Chunyu Wang, Xinggang Wang, Wenjun Zeng, and Wenyu Liu. Fairmot: On the fairness of detection and re-identification in multiple object tracking. *International Journal of Computer Vision*, 129:3069 – 3087, 2020.

[24] Chao Liang, Zhipeng Zhang, Yi Lu, Xue Zhou, Bing Li, Xiyong Ye, and Jianxiao Zou. Rethinking the competition between detection and reid in multiobject tracking. *IEEE Transactions on Image Processing*, 31:3182–3196, 2020.

[25] Bing Shuai, Andrew G. Berneshawi, Xinyu Li, Davide Modolo, and Joseph Tighe. Siammot: Siamese multi-object tracking. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, null:12367–12377, 2021.

[26] Zhongdao Wang, Liang Zheng, Yixuan Liu, and Shengjin Wang. Towards real-time multi-object tracking. *ArXiv*, abs/1909.12605:null, 2019.

[27] Alexander Hermans, L. Beyer, and B. Leibe. In defense of the triplet loss for person re-identification. *ArXiv*, abs/1703.07737:null, 2017.

[28] Tong Xiao, Shuang Li, Bochao Wang, Liang Lin, and Xiaogang Wang. Joint detection and identification feature learning for person search. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, null:3376–3385, 2016.

[29] 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 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, null:6397–6406, 2020.

[30] Xipeng Yang, Jinxing Ye, Jincheng Lu, Chenting Gong, Minyue Jiang, Xiangru Lin, Wei Zhang, Xiao Tan, Yingying Li, Xiaoqing Ye, and Errui Ding. Box-grained reranking matching for multi-camera multi-target tracking. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, null:3095–3105, 2022.

[31] Hung-Min Hsu, Jiarui Cai, Yizhou Wang, Jenq-Neng Hwang, and Kwang-Ju Kim. Multi-target multi-camera tracking of vehicles using metadata-aided re-id and trajectory-based camera link model. *IEEE Transactions on Image Processing*, 30:5198–5210, 2021.

[32] paperswithcode. Person Re-Identification on Market-1501. https://paperswithcode.com/sota/person-re-identification-on-market-1501. Accessed: 2023-06-20.

[33] paperswithcode. Person re-identification on dukemtmc-reid. https://paperswithcode.com/sota/person-re-identification-on-dukemtmc-reid, 2023. Accessed: 2023-06-20.

[34] paperswithcode. Person re-identification on msmt17. https://paperswithcode.com/sota/person-re-identification-on-msmt17, 2023. Accessed: 2023-06-20.

[35] Lin Wu, Yang Wang, Junbin Gao, and D. Tao. Deep co-attention based comparators for relative representation learning in person re-identification. *ArXiv*, abs/1804.11027:null, 2018.

[36] Ergys Ristani, Francesco Solera, Roger S. Zou, R. Cucchiara, and Carlo Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. *ArXiv*, 2016.

[37] Longhui Wei, Shiliang Zhang, Wen Gao, and Q. Tian. Person transfer gan to bridge domain gap for person re-identification. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, null:79–88, 2017.

[38] paperswithcode. Person re-identification on market-1501-c. https://paperswithcode.com/sota/person-re-identification-on-market-1501-c, 2023. Accessed: 2023-06-20.

[39] Minghui Chen, Zhiqiang Wang, and Feng Zheng. Benchmarks for corruption invariant person re-identification. *ArXiv*, abs/2111.00880:null, 2021.

[40] Zhedong Zheng, Tao Ruan, Yunchao Wei, Yi Yang, and Tao Mei. Vehiclenet: Learning robust visual representation for vehicle re-identification. *IEEE Transactions on Multimedia*, 23:2683–2693, 2020.

[41] paperswithcode. Vehicle re-identification on veri-776. https://paperswithcode.com/sota/vehicle-re-identification-on-veri-776, 2023. Accessed: 2023-06-20.

[42] AI-City-challenge2021. 2021 challenge winners. https://www.aicitychallenge.org/2021-challenge-winners, 2023. Accessed: 2023-06-20.

[43] Deeksha Dangwal, Vincent T. Lee, Hyo Jin Kim, Tianwei Shen, M. Cowan, R. Shah, Caroline Trippel, Brandon Reagen, Tim Sherwood, Vasileios Balntas, Armin Alaghi, and Eddy Ilg. Analysis and mitigations of reverse engineering attacks on local feature descriptors. *ArXiv*, abs/2105.03812:null, 2021.

[44] Verordnung (EU) 2016/679. Verordnung (eu) 2016/679 des europäischen parlaments und des rates vom 27. april 2016 zum schutz natürlicher personen bei der verarbeitung personenbezogener daten, zum freien datenverkehr und zur aufhebung der richtlinie 95/46/eg (datenschutz-grundverordnung) (text von bedeutung für den ewr), 2016.

[45] Hongbo Wang, Jiaying Hou, and Na Chen. A survey of vehicle re-identification based on deep learning. *IEEE Access*, 7:172443–172469, 2019.