Improving Urban ReID with Data Augmentation and Advanced Loss Functions

Anonymous ICIP submission

Paper ID ***

Abstract

This paper explores the application of various techniques to improve urban object re-identification (ReID) using a baseline Part-Aware Transformer (PAT) model. We investigate the impact of dataset expansion, data augmentation, and alternative loss functions on model performance. Specifically, we focus on the inclusion of additional data from the Universidad Autónoma de Madrid (UAM) and the use of several data augmentation techniques, such as color jitter, random erasing, and AugMix. Furthermore, we examine the potential of the Adaptive Sparse Pairwise Loss (ASPL) as an alternative to the commonly used triplet loss function. Our experiments show that dataset expansion combined with random erasing led to the best performance, achieving a mean Average Precision (mAP) of 0.240. However, the inclusion of ASPL resulted in lower performance, with a mAP of 0.225. These findings highlight the importance of selecting the right loss function and augmentation techniques for improving the robustness and accuracy of ReID models.

1. Introduction

Urban object re-identification (ReID) [6] is a challenging task in computer vision, where the goal is to match objects across different camera views in an urban environment. This task is particularly difficult due to various factors such as appearance variation, occlusion, lighting conditions, and background clutter. The advent of deep learning models [4], especially those leveraging part-aware networks, has significantly improved the performance of ReID systems. However, these models often rely on large and diverse datasets to generalize effectively across different urban environments.

In this study, we explore several techniques aimed at improving the performance of a baseline Part-Aware Transformer (PAT) [7] [3] model for urban object ReID. First, we examine the effect of dataset expansion by incorporating the UrbAM-ReID Dataset [6], which introduces new variations in the data. Next, we evaluate the impact of various data augmentation strategies, including color jitter [2], random erasing [9], and AugMix [5], on model robustness. Finally, we investigate the effectiveness of the Adaptive Sparse Pairwise Loss (ASPL) [10] as a replacement for the widely used triplet loss function, which is known for its simplicity but can sometimes lead to inefficiencies in learning from large, imbalanced datasets.

Through a series of experiments, we aim to identify the most effective combination of dataset expansion, augmentation techniques, and loss functions that lead to improved accuracy in urban object ReID tasks.

2. Method

This study investigates several methods for improving the baseline performance in urban re-identification (ReID). The baseline model uses a Part-Aware Transformer (PAT) for object re-identification. To enhance its accuracy, we applied two main strategies: dataset expansion and advanced data augmentation techniques. The following sections detail the methods employed.

2.1. Dataset Expansion

The first technique utilized to improve the baseline model was expanding the dataset by incorporating additional data collected from the Universidad Autónoma de Madrid (UAM). The baseline dataset and the UrbAM-ReID Dataset [6], although similar, provided additional variations that allowed the model to generalize better in urban environments. This dataset expansion helped increase the diversity of training samples, making the model more robust to the challenges present in urban re-identification tasks.

2.2. Data Augmentation

Data augmentation was applied to further enhance the model's robustness [5]. Augmentation techniques are especially important in ReID tasks as they simulate various real-world conditions such as lighting changes, occlusions, and perspective shifts. Despite the two datasets being similar, the augmentation strategies introduced significant diversity in the data, helping the model generalize more effectively.

Each model was trained for 5 epochs initially, and then the most promising techniques were selected. After this selection, the model was retrained for an additional 60 epochs using these augmentation techniques.

The specific augmentation techniques applied include the following:

2.2.1. Color Jitter

Color jitter [2] is a technique that modifies the brightness, contrast, saturation, and hue of input images, creating variations that simulate different lighting conditions. We implemented two configurations of color jitter: *soft* and *strong*. In the soft configuration, with a probability of 0.5, the brightness, contrast, and saturation were adjusted by 15%, and the hue by 10%. The strong configuration also applied a 0.5 probability, but with more significant adjustments: brightness and contrast by 30%, saturation by 20%, and hue by 20%. These transformations helped improve the model's resilience to lighting and color inconsistencies, simulating real-world variations in environmental conditions.

2.2.2. Random Patch

Random patch augmentation [8] involves randomly cropping and inserting patches from different regions of the image. This technique allows the model to focus on diverse features and is especially helpful when dealing with partially occluded or cropped views of objects. In our implementation, we set the probability of applying random patch augmentation to 0.5, ensuring that the model learned to handle different spatial perspectives, making it more adaptable to situations where parts of an object may be obscured.

2.2.3. Random Erasing

Random erasing [9] involves randomly masking out portions of the image, forcing the model to make decisions based on incomplete visual information. This technique simulates real-world occlusions, such as object being partially blocked by other objects or individuals in an urban setting. We applied random erasing with a probability of 0.5, helping the model learn to re-identify objects even when critical parts of the image are missing.

2.2.4. AugMix

AugMix [5] is an advanced augmentation technique that combines multiple transformations to create mixed, augmented images. This technique generates images by mixing augmented versions of the original image, with the goal of enhancing model robustness to varied conditions. The AugMix method was implemented with the following parameters:

- **Augmentation Probability Coefficients**: A Dirichlet distribution with a coefficient of 1 was used to sample the mixture weights.
- **Mixture Width**: The number of augmented chains mixed for each image was set to 3.
- **Mixture Depth**: A random depth between 1 and 3 was selected for each augmentation chain.
- **Severity**: A severity level of 1 was used for the underlying augmentation operators, controlling the intensity of the transformations applied to each image.

By mixing augmented images with varying degrees of transformation, AugMix provides a diverse set of training samples, enhancing the model's ability to generalize and improving its robustness to variations in the input images.

2.3. Adaptive Sparse Pairwise Loss

In addition to the previously described methods, we also consider the potential benefits of switching from the traditional triplet loss to the **Adaptive Sparse Pairwise Loss** (ASPL) [10] in improving model performance. Triplet loss has been widely used in ReID tasks, aiming to minimize the distance between embeddings of the same identity while maximizing the distance between embeddings of different identities. However, triplet loss has limitations, particularly when it comes to handling large and imbalanced datasets, where the number of negative samples can overwhelm the model's learning process.

The **Adaptive Sparse Pairwise Loss** (ASPL) addresses some of these challenges by dynamically adjusting the weight of each pair during training, making it more adaptive to difficult cases. ASPL focuses on selecting informative negative pairs, which results in a more efficient use of the training data. By reducing the influence of easy negative samples (which are often too easy to distinguish), ASPL encourages the model to focus on harder examples, leading to better performance in terms of generalization and robustness.

Moreover, ASPL reduces the computational burden, as it does not require sampling all possible negative examples, unlike the triplet loss, which can be computationally expensive when dealing with large datasets. By promoting a sparse and adaptive selection of negative pairs, ASPL makes the model more effective in distinguishing fine-grained differences, even in highly challenging environments like urban re-identification.

In summary, the shift to **Adaptive Sparse Pairwise Loss** could provide more robust performance by focusing on harder and more informative pairs, which may lead to better generalization across varying urban environments, compared to the conventional triplet loss that often struggles with imbalanced negative samples in large-scale datasets.

3. Data

The dataset used in this study is the one provided by the **Urban Elements ReID Challenge** [1] on Kaggle. This dataset consists of images captured in urban environments, specifically designed for object re-identification (ReID) tasks. The challenge provides labeled images from multiple camera views, with the goal of matching objects across different urban locations.

To improve the performance of the baseline model, we expanded the dataset by incorporating the UrbAM-ReID Dataset [6], captured at the **Universidad Autónoma de Madrid (UAM)**. These additional images provided more diverse views of objects in urban environments, which helped the model generalize better across varying conditions such as lighting, occlusions, and camera angles. The UrbAM-ReID Dataset [6], while similar to the Urban Elements ReID dataset [1] in terms of the urban context, introduced new variations in the appearance of objects and environmental settings, offering a more comprehensive range of samples for training.

By combining the Urban Elements ReID Challenge dataset with the external UAM data, we were able to increase the diversity and size of the training set, providing the model with a richer set of examples to learn from and potentially improving its ability to re-identify objects across different urban scenarios.

4. Results and Ananysis

This section presents the results of the various augmentation techniques applied to the baseline model, as well as an analysis of their impact on model performance.

4.1. Dataset Expansion

By expanding the dataset with the UrbAM-ReID Dataset, we observed an improvement in the model's performance. Both the baseline model and the model using the UAM dataset were trained for 60 epochs (Figure 1). The baseline model achieved a mean Average Precision (mAP) of 0.201. After incorporating the new dataset, the mAP increased to 0.212. This improvement, although modest, highlights the benefit of dataset expansion in enhancing the model's ability to generalize across different urban environments.

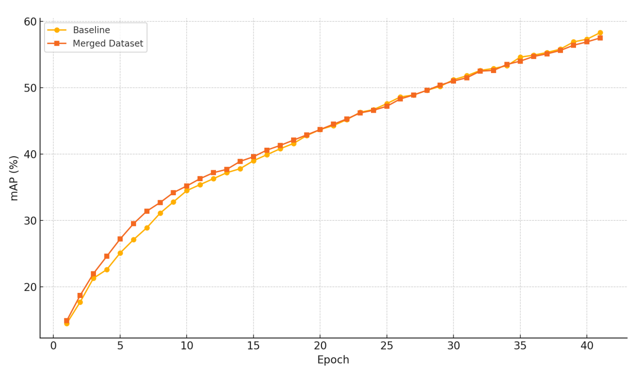

Figure 1. Comparison of mAP between merged dataset and baseline over the first 45 epochs.

4.2. Data Augmentation Results

After training each model for 5 epochs, the most promising augmentation techniques were selected for further analysis (Figure 2, 3). These included *Color Jitter* and *Random Erasing*, both of which were retrained for an additional 60 epochs. The models trained in this phase were: *REA* (Random Erasing Augmentation), *RPA* (Random Patch Augmentation), *CJ* (Color Jitter), and *AugMix*.

During the first 5 epochs, the baseline model outperformed all other models in terms of accuracy (Figure 2). However, this result is not significant as only a few epochs had passed, and the models had not yet fully adapted to the augmentation techniques. Additionally, it's important to note that the data was evaluated on a test set not provided by the challenge, which could have influenced the comparison.

The results of these augmentations are discussed below.

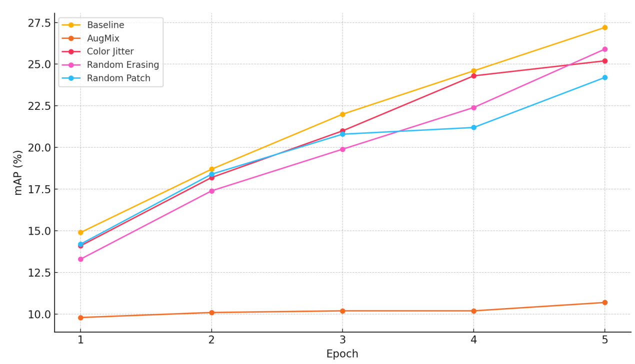

Figure 2. Mean Average Precision (mAP) over epochs for different augmentation techniques. In the first 5 epochs, the baseline model outperformed all other models, but this was not a significant result as the models had not yet fully adapted to the augmentations.

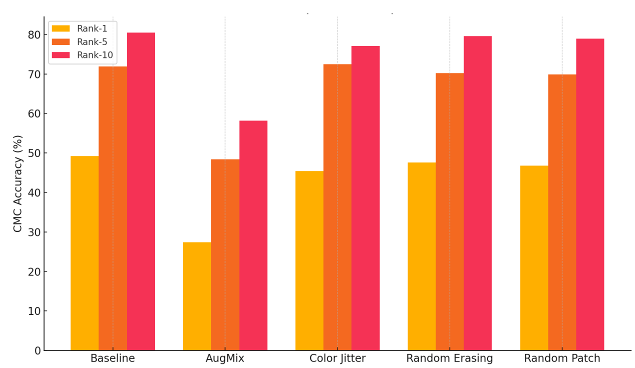

Figure 3. Cumulative Matching Curve (CMC) at epoch 5 for different models. The results show the relative performance of the models after the initial training period.

4.2.1. Color Jitter

The soft configuration of color jitter showed a minimal decrease in performance, with the mAP dropping slightly from 0.212 to 0.209. This result indicates that the soft color jitter did not negatively impact the model's performance, maintaining nearly identical accuracy to the baseline.

However, the strong configuration of color jitter had a more pronounced negative effect, with the mAP dropping significantly to 0.182. This suggests that more aggressive color jittering may have introduced too much noise, making it harder for the model to maintain accurate re-identification.

4.2.2. Random Erasing

The most promising results came from the *Random Erasing* augmentation. When combined with the dataset expansion, Random Erasing led to a substantial improvement in model performance, resulting in a mAP of 0.240. This combination of techniques allowed the model to better handle occlusions and partial views, improving its ability to re-identify objects even when parts of their images were missing.

In summary, the Random Erasing technique, especially when combined with dataset expansion, outperformed other augmentation strategies and provided the best results. The results indicate that Random Erasing is highly effective in enhancing the model's robustness in challenging urban re-identification scenarios.

4.3. Adaptive Sparse Pairwise Loss Results

We also evaluated the performance of the model using the **Adaptive Sparse Pairwise Loss** (ASPL) in combination with the previously applied techniques of dataset expansion and *Random Erasing*. The goal was to assess whether this alternative loss function could outperform the conventional triplet loss in terms of accuracy and robustness in the urban re-identification task.

Despite its potential advantages in dynamic pair selection and computational efficiency, the results obtained with ASPL were not as promising as expected. When using the combination of dataset expansion, *Random Erasing*, and ASPL, the model's performance showed a notable decrease. The mAP dropped to 0.225, which is significantly lower than the 0.240 achieved with the triplet loss combined with the same dataset expansion and augmentation techniques. This indicates that while ASPL might offer advantages in certain scenarios, it does not necessarily lead to better performance in the context of this specific re-identification task.

The drop in performance could be attributed to the adaptive nature of ASPL, which might not be fully exploiting the underlying structure of the dataset in this particular case. While ASPL has been shown to improve learning by focusing on informative pairs of negative samples, the results here suggest that the triplet loss may still be more effective in handling the complexities of urban re-identification tasks with the given data and augmentations.

In summary, although ASPL demonstrated potential in terms of efficiency and adaptive learning, it did not outperform the conventional triplet loss in this experiment. The combination of dataset expansion, *Random Erasing*, and triplet loss remains the most effective approach, yielding a higher mAP of 0.240.

5. Conclusions

In this paper, we have explored several methods to improve the performance of a Part-Aware Transformer (PAT) model for urban re-identification. Our experiments showed that dataset expansion, combined with random erasing, yielded the best results, achieving a mean Average Precision (mAP) of 0.240. This demonstrates that increasing dataset diversity and simulating occlusions through random erasing can significantly enhance model performance in urban re-identification tasks.

On the other hand, the Adaptive Sparse Pairwise Loss (ASPL), while promising in terms of adaptive learning and computational efficiency, did not outperform the traditional triplet loss. The model trained with ASPL, combined with dataset expansion and random erasing, achieved a lower mAP of 0.225, suggesting that ASPL may not be as effective in handling the complexities of urban re-identification in this case.

In summary, while ASPL showed potential in certain scenarios, the triplet loss, combined with dataset expansion and random erasing, remains the most effective approach for improving the model's robustness and accuracy. Future work may involve further fine-tuning of the ASPL or exploring other loss functions to improve its performance in ReID tasks.

Method	**mAP**
Baseline (no augmentation)	0.201
Dataset Expansion	0.212
Color Jitter (Soft)	0.209
Color Jitter (Strong)	0.182
Random Erasing	0.240
Random Erasing + ASPL	0.225

Table 1. Performance of various methods in terms of mAP.

References

[1] Urban reid challenge – dataset. https://www.kaggle.com/competitions/urban-reid-challenge/data, n.d. Accessed: 2025-06-08. 2, 3

[2] Yunpeng Gong, Liqing Huang, and Lifei Chen. Person re-identification method based on color attack and joint de-

fence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 4313–4322, 2022. 1, 2

[3] Shuting He, Hao Luo, Pichao Wang, Fan Wang, Hao Li, and Wei Jiang. Transreid: Transformer-based object re-identification. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 15013–15022, 2021. 1

[4] Xinchen Liu, Wu Liu, Tao Mei, and Huadong Ma. A deep learning-based approach to progressive vehicle re-identification for urban surveillance. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, pages 869–884. Springer, 2016. 1

[5] Timur Mamedov, Denis Kuplyakov, and Anton Konushin. Approaches to improve the quality of person re-identification for practical use. *Sensors*, 23(17), 2023. 1, 2

[6] Paula Moral, Álvaro García-Martín, and José M. Martínez. Long-term geo-positioned re-identification dataset of urban elements. In *2024 IEEE International Conference on Image Processing (ICIP)*, pages 124–130, 2024. 1, 3

[7] Hao Ni, Yuke Li, Lianli Gao, Heng Tao Shen, and Jingkuan Song. Part-aware transformer for generalizable person re-identification. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 11280–11289, 2023. 1

[8] Guiwei Zhang, Yongfei Zhang, Tianyu Zhang, Bo Li, and Shiliang Pu. Pha: Patch-wise high-frequency augmentation for transformer-based person re-identification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14133–14142, 2023. 2

[9] Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. Random erasing data augmentation. In *Proceedings of the AAAI conference on artificial intelligence*, pages 13001–13008, 2020. 1, 2

[10] Xiao Zhou, Yujie Zhong, Zhen Cheng, Fan Liang, and Lin Ma. Adaptive sparse pairwise loss for object re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 19691–19701, 2023. 1, 2