# Generative Adversarial Networks for Data Augmentation in Person Re-Identification

Laura Álvarez-González[1], Víctor Uc-Cetina[1*†], Anabel Martin-González[1†]

[1*]Facultad de Matemáticas, Universidad Autónoma de Yucatán, Periférico Norte, Mérida, 97000, Yucatán, Mexico.

*Corresponding author(s). E-mail(s): uccetina@correo.uady.mx; Contributing authors: laura.alvargonza@gmail.com; amarting@correo.uady.mx;

†These authors contributed equally to this work.

**Abstract**

People re-identification systems based on deep neural networks require as many training examples as possible. The possibility of automatically identifying a person in different images is highly relevant in various security surveillance applications. However, getting enough data to train these networks is sometimes problematic or not possible at all, which limits the quality of the re-identification model. One strategy to cope with this scarcity of data is through the generation of synthetic examples that can increase the size of our training dataset. In this article we propose a data augmentation methodology for re-identification systems, using generative adversarial networks. We provide empirical evidence showing that a StyleGAN model can be used to generate artificial but useful images, when they are used to further train a re-identification network. Specifically, with the manipulation of the latent space of a generative adversarial network, we successfully generate multiple images portraying an artificial person in various poses.

**Keywords:** person re-identification, generative adversarial models, data augmentation

## 1 Introduction

People re-identification is a technique used in the field of artificial intelligence and machine learning to recognize a person in different images or videos, even if they are in different angles, lighting or clothing. This technique is used in various applications, such as security surveillance, people identification in digital images, and behavior analysis in videos. It is based on the use of machine learning models that learn to recognize the characteristics that identify a person under different circumstances. These models are trained on a image and video data sets of people, along with information about the characteristics that identify each person. The re-identification of people can be challenging due to the variability of the characteristics that identify a person, such as clothing, hairstyle and other aspects that can change their appearance. It can also be challenging if you have a limited amount of training data. To overcome these challenges, image and video pre-processing techniques can be used, as well as deep learning techniques that allow the model to adapt to variations in a person's appearance and recognize relevant features in low-quality images or videos.

Currently, the most prominent training datasets for person re-identification exhibit a notable limitation in their scale, as they encompass only a modest number of images. To illustrate, the Market1501 dataset comprises a mere 1,501 individuals captured across 6 distinct camera viewpoints, whereas the DukeMTMC-reID dataset encompasses 702 individuals observed from 8 discrete camera perspectives. It is worth noting that among the widely utilized models, Market1501 and DukeMTMC-reID stand out. However, it is essential to acknowledge that DukeMTMC-reID's current usability is constrained by an impediment. Specifically, this dataset has been retracted from circulation and is advised not to be employed for further research purposes.

There are various challenges for the re-identification of people in images, such as low image resolution, variations in lighting and contrast, as well as other factors that complicate the task such as changes in clothing, presence of objects such as backpacks or sweaters, etc. Moreover, the presence of obstacles or people in the background limit the visibility of the person of interest in an open space.

This study investigates the use of a generative adversarial network, together with data augmentation techniques, to train a person re-identification model. The generative adversarial network is a type of machine learning model that is used to generate synthetic images that can be used as additional training data. Various techniques for extending training data, such as image generation and feature expansion, are discussed, and their effectiveness in training a person re-identification model using an adversarial generative network is evaluated.

Furthermore, it's important to note the basis for incorporating this augmentation approach from the perspective of generative adversarial networks. As demonstrated in the 2018 study by Antoniou et al. [1], they examined the enhancement of image classification metrics through artificial expansion of the training dataset using GAN architecture. This helps in understanding the significance of adopting this approach in the context of re-identifying individuals through generative adversarial networks.

# 2 Related work

Generative adversarial networks, originally proposed in 2014 by Ian Goodfellow et al. [2], are capable of artificially generating images with great diversity. Over time, new architectures have been proposed that improve the quality of the data generated, such as the CycleGan architecture, proposed in 2017 by Zhu et al. [3]. Apart from improving the quality of the generated images, it manages to transfer the style or domain of a group of images to another group, using two generative adversarial networks.

StyleGAN is a Generative Adversarial Network (GAN) architecture developed by NVIDIA in the year 2018 [4]. This architecture has been trained to be able to generate high quality images of non-existent people's faces. In this case, it was trained with the FFHQ database. which consists of images of faces of people from the Flickr social network.

One particularly significant property of StyleGAN is its capability to address the issue known as entanglement or "disentangle" that often plagues generative adversarial networks. This issue arises when the generated images become intertwined, leading to a mixture or confusion of different aspects of the image, like pose, gender, facial expression, etc. For instance, envision having the latent vectors of two facial images, the first depicting a girl's face and the second an adult woman's face. In models afflicted by the entanglement issue, interpolating between these images might result in entirely random intermediate images. In contrast, the StyleGAN model ensures coherent and smooth interpolation, allowing for a more controlled manipulation of the latent space to generate images more precisely.

Additionally, another remarkable feature of StyleGAN is its use of a generative network structure based on layers of styles. This unique architecture enables independent control over various aspects of the generated image, such as pose, facial expression, gender, and more. These independently modifiable style layers provide a high degree of flexibility and granularity in image synthesis, contributing to the model's ability to produce diverse and detailed visual outputs.

Due to these two aforementioned characteristics of StyleGAN3, it has been concluded that the model is exceptionally well-suited for the generation of multiple images depicting an artificial person. The disentanglement feature ensures that the different attributes of the generated images remain distinct and coherent, while the layered style architecture empowers precise control over each aspect, enabling the creation of a wide variety of images capturing various facets of the artificial persona.

Algorithms for re-identification of people have been proposed since 1996 [5]. Currently, the most widespread method is the use of a neural network as feature extractor. In recent years, there has been a great deal of interest in the use of adversarial generative networks in people re-identification models. In 2019, Hamed Alqahtani et al. [6] provided a detailed introduction to the state of the art in this field, describing different types of adversarial generative networks and 11 different architectures used for style transfer, labeled LSRO and the globalization of the model. Furthermore, Zhiyuan Luo et al. [7] focused exclusively on architectures that generate artificial images by switching styles between different cameras or databases. Finally, Yiqi Jiang et al. [8] conducted a detailed study on the quality of images generated by different adversarial generative network architectures in re-identification models, analyzing the details of these artificial images.

An important increment in the study of GANs for data augmentation in the training of re-identification models can be noted since 2018. The most relevant approaches can been grouped into three categories, corresponding to different methods used to generate new artificial images.

1. Style transfer [3, 9–13]. New images are artificially generated from an input image, using different styles, known also as domains. The styles are imposed on the new images at the moment they are generated by neural networks previously trained for that purpose. In the new generated images, you can see modifications with respect to the input image, such as color, tone, and lighting.
2. Pose transfer [14–18]. In this approach the inputs are one image of a real person and the target posture that we want to impose on that person. The posture can be specified whether as a heat map or by the joints that correspond to the skeleton of the desired posture. The model is capable to generate the image of the input person with a determined posture.
3. Random generation [19–24]. Methods in this category are less constrained and they focus on randomly generation of synthetic images with the only condition that the generated images should have similar characteristics of those images in the dataset that we want to augment.

# 3 Methods

In general, the operation of the re-identification model is simple. First, a database with tagged images of people is obtained through different security cameras. Second, the model is trained in a supervised manner so that it can classify a certain finite number of people from the training set. For example, using the Market 1501 database, we can train a model that is capable of identifying 751 people. Once the model has been trained, the following steps are followed:

1. Each one of the images is introduced into the model in order to obtain a vector of size 751. This vector represents the proportion of each of the 751 people that the input person contains.
2. A vector representing the image of the person we want to search for is generated. We measure the distance with each of the other vectors that represent the other images of people. Within the literature, the use of the cosine distance is a common practice.
3. A classification by shortest distance is made among all the images in the dataset. The least distance means that the image is more similar to the original and it is probably the same person.

In this article we propose a methodology that consists of four phases. Each of these phases is described next (see Fig. **??**).

a) Re-training the adversarial generative network StyleGAN3 with the Market-1501 database.
b) Generation of multiple images of artificial people in different postures is facilitated through the manipulation of latent vectors within the StyleGAN generator. Commencing with a seed image extracted from a random vector, we harness the potential of the StyleGAN generator to produce a series of artificial images portraying the

same individual in varying postures. This manipulation of latent vectors allows us to achieve controlled transformations and diverse poses in the generated images. Image filtering through automatic elimination of generated images that present noise or have been generated incorrectly.

c) Training of the re-identification model.

d) Inference of the re-identification model.

For the generation of artificial images, the architecture of the adversarial generative network we used StyleGAN3 [24]. It is a generative image model developed by the Nvidia research team in 2021. It is an improved version of the StyleGAN2 model, which is characterized by its ability to generate high-quality and realistic images in a wide variety of image categories. StyleGAN3 uses a deep learning approach based on generators and discriminators. The generator is a neural network that is trained to generate images that are as realistic as possible. To do this, you are shown a set of real images and asked to generate images that resemble them. As it trains, the generator learns to extract relevant features from actual images and use them to generate images that are as realistic as possible.

The discriminator is a neural network that is trained to distinguish between real and generated images. You are shown both real and generated images and are asked to determine which are real and which are generated. As it is trained, the discriminator learns to identify the features that differentiate real images from generated ones, and is used to guide training of the generator towards generating more realistic images.

StyleGAN is pre-trained with 25 million face images of which 70 thousand are real from the high quality FFHQ database of resolution 1024 × 1024 pixels and the rest were generated by the discriminator. Currently, StyleGAN3 works as a high-quality people face image generator. StyleGAN3 has been trained to generate faces and we will apply the transferred learning process to retrain it. For the re-training of StyleGAN3, we used the Market-1501 database, which consists of 51,247 images of 1,501 different people captured by six different cameras.

## 3.1 Evaluation of StyleGAN

To evaluate the performance of StyleGAN, we use the metric Fréchet inception distance (FID), proposed by P. Dimitrakopoulos et al. [25]. This distance metric is used to measure the similarity between two image distributions. The FID metric is based on the idea that the distance between two image distributions is the same as the distance between image features extracted from a deep neural network. To calculate the FID distance between two image distributions, features are first extracted from each distribution using a deep neural network, and then the distance between those features is calculated using the Fréchet distance. Mathematically, the FID distance between two image distributions can be calculated as follows:

$$\texttt{FID} = |\mu - \mu_w|^2 + \texttt{tr}(\sigma + \sigma_w - 2(\sigma\sigma_w)^{1/2}). \tag{1}$$

We compare the mean and the convariance matrix of the real and fictitious images, obtaining the data from one of the deepest layers of the neural network. It aims to mimic human perception to identify the similarity between two images using the

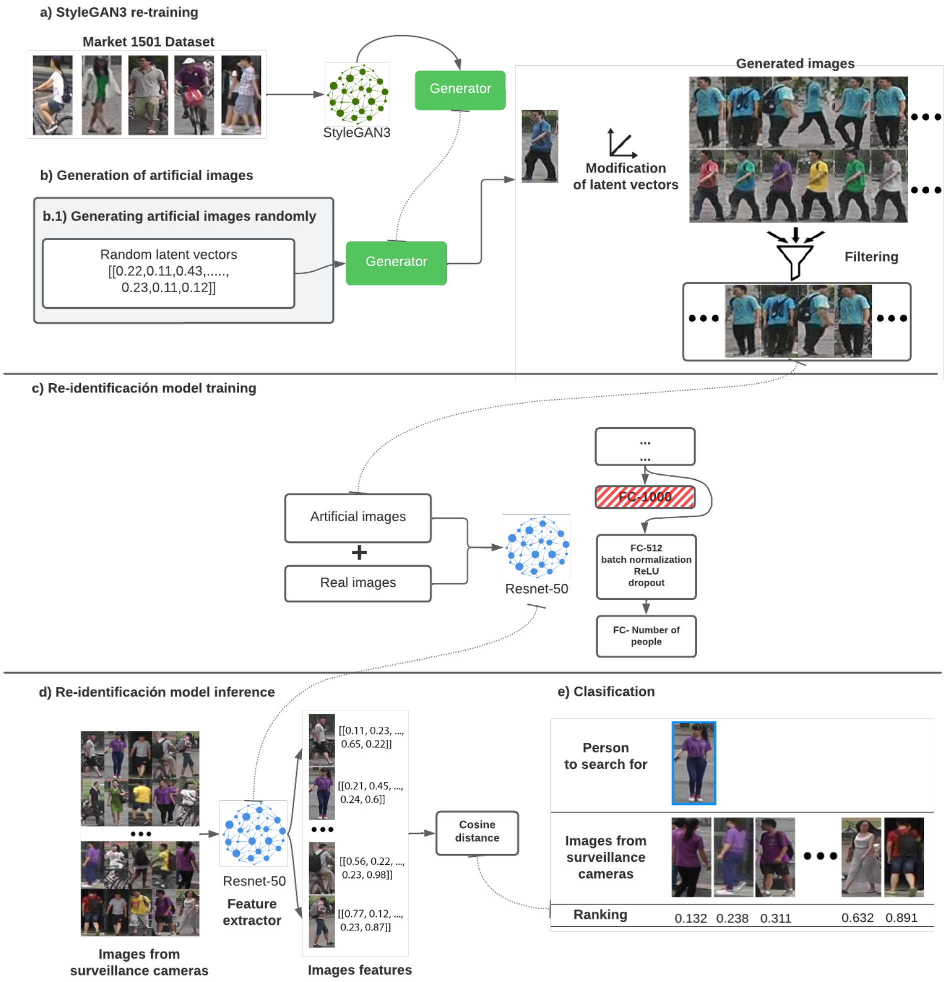


**Fig. 1** **a)** Retraining StyleGAN3 with images from the Market 1501 dataset. **b)** Generator Inference and Data Augmentation: Artificial images are generated using the generator by sampling from a random vector. This latent vector is traversed within the generator's latent space to create diverse images related to the original root image. Subsequently, noise-filtering filters are applied to all generated images to eliminate noisy instances. **c)** Re-Identification Model Training: The re-identification model is trained by utilizing ResNet-50. The final layer is removed and adjusted based on the number of distinct individuals within the training batch. **d)** The previously trained model is employed as a feature extractor. All images from the surveillance cameras are fed into the model, and a feature vector is extracted from each image. Subsequently, each feature vector is compared with the feature vector of the target person's image using cosine similarity, generating a ranking. The most similar images are those ranked highest.

discriminator as a feature extractor. If the returned value is zero, it indicates that the generated data and the actual data are identical, which means that the lower the returned value, the greater the similarity between the generated images and the actual images.

We have devised a methodology to generate a multitude of artificial images employing a stochastic process. In this approach, entirely randomized depictions of synthetic individuals are engendered. Leveraging a randomly chosen numerical value, often referred to as a seed, the generator allocates a latent vector that corresponds to an image. To procure diversifications of the initial image, an alternative random latent vector can be employed, facilitated either by a distinct seed or by means of interpolation techniques.

Within the various AdaIN (Adaptive Instance Normalization) layers of the model, akin to a fusion of artistic styles, the latent vector undergoes modifications, yielding a spectrum of deviations from the original image. This spectrum ranges from complete structural transformations to more nuanced alterations encompassing shifts in tonality, illumination, color palettes, saturation levels, and various other attributes. Another avenue for generating diversifications of the source image arises through manipulation of the primal latent vector via interpolation techniques.

However, due to the sheer volume of generated images, it becomes imperative to implement an automated filtering mechanism. This mechanism is designed to discern and discard images tainted by noise or those that have been inadequately generated. This automated filtration ensures that only high-quality synthetic images are integrated into the training dataset, thus augmenting the efficacy of subsequent re-identification model training processes."

## 3.2 YoloV4 tiny

The YOLOv4 tiny model, as proposed by Z. Jiang et al. [26], presents a streamlined variant of the YOLOv4 framework designed explicitly for object detection within images and videos. YOLO (You Only Look Once) distinguishes itself in the realm of object detection for its exceptional balance between swiftness and precision. The model has been meticulously fine-tuned for the accurate and rapid identification of pedestrians, achieving performance on par with more extensive models, all while significantly alleviating the resource burden. The diminutive iteration of the YOLOv4 architecture proves particularly advantageous for resource-constrained environments, as it boasts reduced computational demands, rendering it suitable for deployment on mobile devices and lower-performance computing platforms.

In essence, YOLOv4 tiny harnesses the power of a convolutional neural network to extract pertinent features from input images. Subsequently, a fusion of machine learning techniques is employed to execute the task of object detection.

## 3.3 SSIM

The Structural SIMilarity (SSIM) metric is a fundamental measure employed to quantify the structural resemblance between two images. This metric finds extensive application in evaluating the fidelity of a processed image relative to its original counterpart. The SSIM score is calculated by analyzing and contrasting the inherent structural characteristics of the images in question. What sets SSIM apart is its grounding in the understanding that human perception of image quality hinges not solely on pixel-level disparities, but crucially on the preservation of structural content.

In this context, the SSIM metric proves invaluable as it facilitates a nuanced assessment of the perceptual quality of processed images. However, to leverage this metric effectively, a novel filtering approach has been devised. This filtering mechanism conducts a comparative analysis between the initial seed artificial image, from which the image generation process commences, and the array of subsequently generated images representing the same person. By scrutinizing this spectrum of images, the filtering algorithm identifies and eliminates outliers that exhibit substantial deviations from the original seed image. This filtering process has the capacity to identify images marred by artifacts such as excessive noise or those that manifest drastic structural inconsistencies.

To carry out the calculation of the SSIM metric, three structural characteristics of two images are compared: their mean intensity, their intensity variance and their intensity covariance. The SSIM is obtained from the product of these three characteristics, and two images are considered to have high SSIM if they have similar mean intensity, similar intensity variance, and similar intensity covariance. The result is obtained from the product of these three characteristics, and is denoted as:

$$SSIM(x, y) = l(x, y) \cdot c(x, y) \cdot s(x, y), \tag{2}$$

where $x$ and $y$ are the two images being compared, $l(x, y)$ is the similarity in mean intensity, $c(x, y)$ is the similarity in intensity covariance and $s(x, y)$ is the similarity in intensity variance.

The similarity in mean intensity is calculated as:

$$l(x, y) = \frac{2 \cdot \mu_x \cdot \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \tag{3}$$

where $\mu x$ and $\mu y$ are the mean intensities of the images $x$ and $y$, respectively, and $C_1$ is a constant which is used to avoid division by zero.

The similarity in intensity covariance is calculated as:

$$c(x, y) = \frac{2 \cdot \sigma xy + C_2}{\sigma x^2 + \sigma y^2 + C_2}, \tag{4}$$

where $\sigma xy$ is the intensity covariance between the images $x$ and $y$, $\sigma x$ and $\sigma y$ are the intensity variances of the images $x$ and $y$, respectively, and $C_2$ is a constant used to avoid division by zero.

The similarity in intensity variance is calculated as:

$$s(x, y) = \frac{2 \cdot \sigma x \cdot \sigma y + C_3}{\sigma x^2 + \sigma y^2 + C_3}, \tag{5}$$

where $\sigma_{xy}$ is the intensity covariance between images $x$ and $y$, $\sigma_x$ and $\sigma_y$ are the intensity variances of images $x$ and $y$, respectively , and $C_3$ is a constant used to avoid division by zero.

In summary, the SSIM metric is calculated by comparing the mean intensities, intensity covariances, and intensity variances of two images. The SSIM is obtained as the product of the similarity in each of these characteristics, and is used to evaluate the quality of a processed image in comparison with an original image. It is based on the comparison of the structural characteristics of both images. The higher the result, the more variation there will be in the generated images.

$$\texttt{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma xy + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma x^2 + \sigma y^2 + c_2)}. \tag{6}$$

## 3.4 Re-identification model

A re-identification model is an algorithm used in image processing and artificial intelligence that makes it possible to identify and track objects or people in a sequence of images. These models are based on the comparison of visual characteristics between different images to determine if it is the same object or person. Mathematically, a re-identification model uses a similarity function to calculate the similarity between two images. This function takes two visual feature vectors (one from the reference image and the other from the image to be compared) and returns a value indicating the similarity between the two images. If the value returned by the similarity function exceeds a certain threshold, it is determined that the images correspond to the same object or person. To compute the visual feature vectors, the model uses a neural network that has been pre-trained on a dataset of labeled images. The neural network extracts relevant features from the images and groups them into a feature vector. These vectors are then used in the similarity function to determine the similarity between the images.

We used Resnet50 convolutional neural network, in which the last layer is modified so that the output adapts to the number of people with whom the model will be trained. During training, the cross-entropy loss is used as the loss function. This function is defined as:

$$L(y, \hat{y}) = -\frac{1}{N}\sum_{i=1}^{N} y_i \log \hat{y_i}.$$

In this equation, $y$ represents the actual label or desired value of the output, $\hat{y}$ represents the output predicted by the model, and $N$ is the number of examples in the data set.

Cross entropy is a loss function commonly used in classification problems, where the model output is interpreted as the probability that an instance belongs to each class. The idea behind cross-entropy is that if the model output is a good approximation of the true probability distribution, then the cross-entropy loss function will have a low value. In contrast, if the model output is very different from the true probability distribution, then the cross-entropy loss function will have a high value. It is used to measure how well the model is making predictions about the actual probability distribution of the classes. During model training, the loss function is optimized to improve its prediction accuracy. Once the training is finished, the model works as a feature extractor and the images are classified.

We introduce all the images to be evaluated one by one into the model to obtain their respective feature vectors and to classify which images are of the same person, each of the image vectors is compared with the vector of the original image using the cosine distance. It is a measure of similarity between two vectors in a vector space. This measure is calculated using the cosine of the angle between the two vectors and can be interpreted as the projection of the shorter vector onto the longer vector. The cosine distance between two vectors $\mathbf{a}$ and $\mathbf{b}$ is calculated as follows:

$$d_c(\mathbf{a}, \mathbf{b}) = 1 - \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}||\mathbf{b}|}.$$

In this equation, $\mathbf{a} \cdot \mathbf{b}$ is the dot product of the vectors $\mathbf{a}$ and $\mathbf{b}$, and $|\mathbf{a}|$ and $|\mathbf{b}|$ are the norms of the vectors $\mathbf{a}$ and $\mathbf{b}$, respectively. The cosine distance has a value between 0 and 1, where a value closer to 1 indicates a greater similarity between the vectors $\mathbf{a}$ and $\mathbf{b}$, and a value closer to 0 indicates a greater similarity. less similarity between them. After having obtained the cosine distance of all the images, they are ordered and those with the smallest cosine distance will be the closest to the original image, that is, they will be those that have been detected as images of the same person.

# 4 Results

The adversarial generative network StyleGAN3 [24] has been used for the generation of artificial images. The model is pre-trained with 25 million face images of which 70 thousand are real from the Flickr-Faces-HQ Dataset (FFHQ) of high quality resolution 512×512 pixels and the rest were generated using the discriminator. Training StyleGAN3 consisted required transfer learning and subsequent re-training with 51,247 images from the Market-1501 database (see Table 1).

| **Model** | **Training Images** | **Testing Images** | **Training Duration** | **Hardware** |
|---|---|---|---|---|
| StyleGAN3 | 51247 | N/A | 2 days 08 hours 24 minutes | Titan RTX |
| **Hyperparameters** | | | | |
| **Configuration** | **GPUs** | **Batch Size** | **Gamma** | **Training Images** |
| stylegan3-r | 1 | 16 | 2 | 5000 |

**Table 1** Technical details and hyperparameters of the StyleGAN3 training process for artificial image generation.

To measure training performance, the Fréchet Inception Distance [27] (FID) metric was used. This metric was applied to both generated and real images, with a smaller FID value indicating a better quality of generated images. In general, the performance of StyleGAN3 surpassed that of other generative adversarial networks trained with the same database (see Table 2).

During the experimentation, images of 401 artificial persons were generated in a completely random manner, and by modifying their latent vectors, 51 images per person were generated in different poses, resulting in a total of 20,451 images (see Fig. 2).

Once the artificial images were generated, two filters were applied to discard images that may have been generated incorrectly or contain noise.

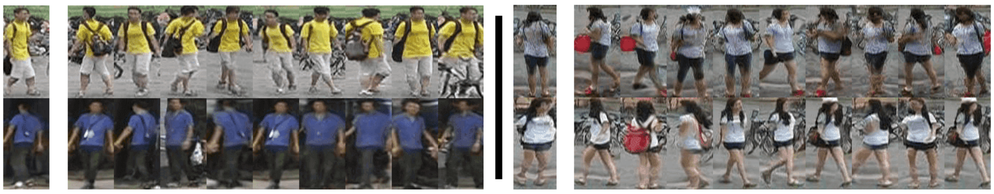


**Fig. 2** Seed Image (Left) - This refers to the initial image randomly generated, the latent vectors of which will be subsequently adjusted to alter its inherent features. Generated Images (Right) - These are the images produced by making modifications to the latent vectors of the seed image.

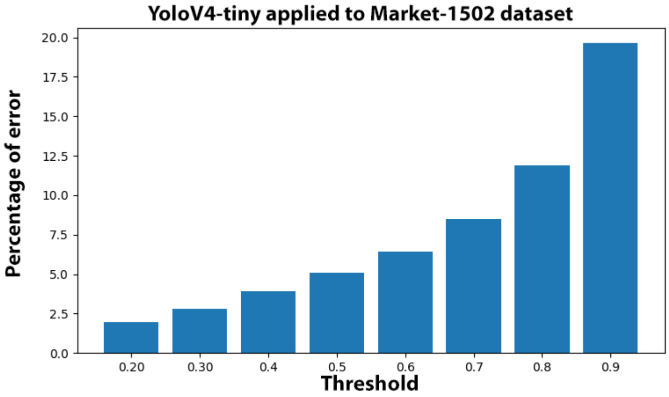


**Fig. 3** Application of YOLOv4 Tiny on images from the Market-1501 database using various thresholds reveals instances of misclassification, where objects are incorrectly labeled as non-pedestrians. This misclassification leads to false negatives, indicating the failure to detect actual pedestrians. Interestingly, the error percentage, which represents the proportion of images where these incorrect identifications occurred, starts to noticeably ascend when utilizing a threshold of 0.6.

- Filtered YoloV4 tiny The trained model YoloV4 tiny [26] was used for the detection of pedestrians in the generated images. All images whose classification was below the threshold of 0.6 were discarded, which was determined by analyzing a histogram created with different threshold Fig. 4, generating different percentages of images classified as non-pedestrians, on the actual images from the Market-1501 database. The Yolo V4 filter was applied to the generated images for pedestrian detection, eliminating 3,419 images that represent 16.7% of the total (see Table 4). Different examples of filtered images are shown in Fig. 5.
- SSIM Filtering The Structural Similarity Metric (SSIM) [28] was used to assess the similarity between two images. The methodology used to apply this metric consisted of selecting an image of a person and comparing it with the rest of the images of that same person in different postures. If the similarity value was equal to one, it was considered to be the same image. This metric was applied to the real images from the Market-1501 database and a histogram was generated. Through analysis of such histogram, it was determined that those images whose SSIM value was less than 0.75 would be discarded. Then the SSIM filter was applied, and 386 images, or

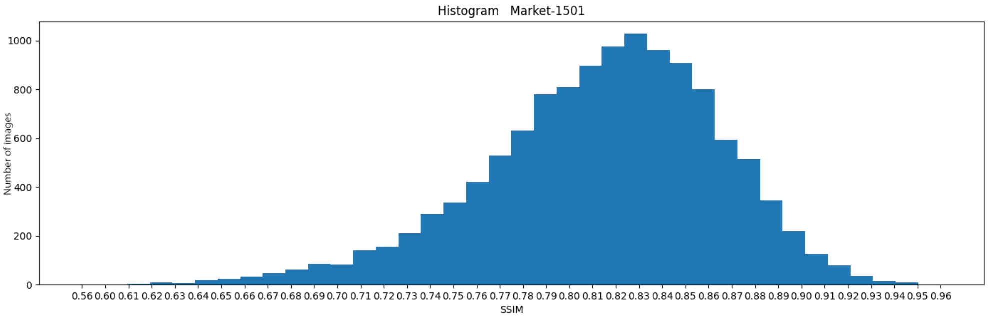


**Fig. 4** Histogram of the SSIM [28] metric on the Market-1501 dataset. Number of images that obtained the same SSIM value. One image per person was selected and compared to the rest of the images of that same person. As can be seen, most of the images are around the threshold of 0.75 and above.

| **Método** | **Market-1501 FID** |
|---|---|
| **Real** [29] | **7.22** |
| IS-GAN [29] | 281.63 |
| FD-GAN [23] | 257.00 |
| PG-GAN [12] | 151.16 |
| DCGAN [23] | 136.26 |
| LSGAN [12] | 136.26 |
| PN-GAN [12] | 54.23 |
| GCL [29] | 53.07 |
| DG-Net [29] | 18.24 |
| DG-GAN [23] | 18.24 |
| **StyleGAN3** | **9.29** |

**Table 2** Table comparing different generative adversarial networks (GANs) using Fréchet Inception Distance (FID)[27] as the performance metric. All models were trained on the Market-1501 dataset[30]. The first row represents the FID value obtained by applying the metric to real images from the dataset.

2.3% of the total, were discarded (see Table 4). Some examples of images discarded by this method can be seen in Fig. 6.

After applying the filters, a total of 3,815 images were discarded (see Table 4).

During the experimentation phase, the model underwent testing with varying numbers of additional individuals, wherein increments of ten persons were introduced sequentially until reaching a total of three hundred and twenty individuals. Figure 7 illustrates that the performance of the fundamental re-identification model remains consistently stable or exhibits slight declines initially, followed by subsequent performance improvements. This observed pattern may be attributed to the concurrent increase in the population of individuals, leading to a corresponding expansion in the number of distinct classes. Consequently, certain classes might exhibit disparities in image quantities, resulting in varying degrees of significance. This scenario can result in varying levels of informativeness among different classes.

In the context of benchmarking against state-of-the-art methods, our approach demonstrates substantial progress, as evidenced by the comparative results presented

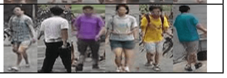
| Features | AdaIN Layers | Example |
|---|---|---|
| Fine | (12, 13, 14, 15) | |
| Medium | (5, 6, 7, 8, 9, 10, 11) | |
| Coarse | (0, 1, 2, 3, 4, 5) |  |

**Table 3** Layers modified to generate new images of the same person based on the modification of their fine, medium and coarse characteristics.

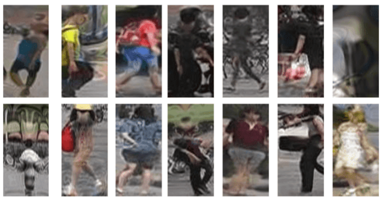


**Fig. 5** Example of some discarded images using the Yolo V4 tiny model for pedestrian detection.

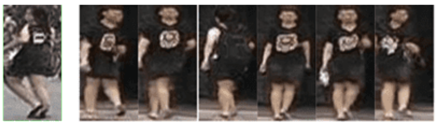


**Fig. 6** Example of some images discarded (right images) using the SSIM metric. Starting from one of the images of a person (left image), it is compared with the rest of the generated images of that same person.

| Method | Images Discarded | % |
|---|---|---|
| Yolov4-tiny [26] | 3.419 | 16.7 |
| SSIM [28] | 396 | 2.3 |
| **TOTAL** | **3.815** | **18.6** |

**Table 4** Number of discarded images during the application of different filters.

in Table 5. A noteworthy observation is the incremental enhancement achieved across both evaluation metrics, R1 and mAP. Particularly, our baseline implementation showcases competitive performance, exhibiting an R1 score of 89.3 and an mAP of 74.2. This establishes a robust foundation for our methodology. Furthermore, the introduction of an additional 280 instances, as indicated by 'Ours (280 added),' yields notable gains, resulting in an R1 score of 90.3 and an mAP of 76.8. This upwards trajectory in performance reaffirms the efficacy of our approach, indicating its potential utility in real-world applications.

## 5 Conclusions

The use of adversarial generative networks to augment data is a promising technique to improve performance in re-identification models. The results obtained in our experiments demonstrate that this technique is effective in generating high-quality data and the versatility of generating modifications thereof.

As a result of the experimentation, it was possible to observe that adding totally artificial people to the re-identification model could improve its performance by 1%.

| Method | R1 | mAP |
|---|---|---|
| CTGAN | 56.7 | 23.6 |
| LSRO | 83.9 | 66.0 |
| Multi-pseudo | 87.9 | 81.1 |
| PN-GAN | 89.4 | 72.5 |
| CamStyle | 89.4 | 71.5 |
| FD-GAN | 90.5 | 77.7 |
| SLSR | 91.8 | 86.0 |
| **Ours (baseline)** | 89.3 | 74.2 |
| **Ours (280 added)** | 90.3 | 76.8 |

**Table 5** Comparative Analysis of Our Approach Against Select Prominent State-of-the-Art Methods [23] on the Market-1501 Dataset.

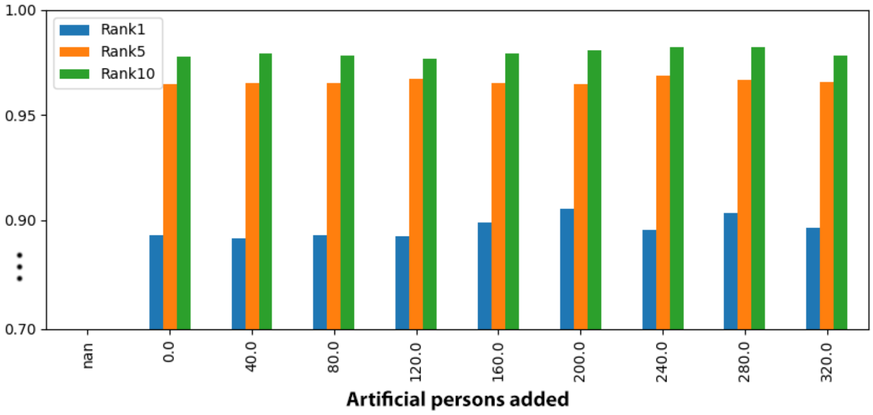


**Fig. 7** Performance of some models trained with different numbers of artificially generated persons (see Table 6). Adding 0, 40, 80, 120, 160, 200, 240, 280, and 320 persons.

The use of these generative adversarial networks allows the use of less original data in the training, which can reduce the resource and time requirements in the model training process. In summary, the use of adversarial generative networks in the realm of re-identification is a valuable technique that can bring a significant improvement in the performance of re-identification models. It allows the adaptation to different data sets and situations. This makes it a valuable tool not only for the field of re-identification, but also for other fields where data generation and improvement is required.

## 6 Competing Interests

The authors have no competing interests to declare that are relevant to the content of this article.

## 7 Authors contribution statement

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Laura Alvarez-González and Víctor

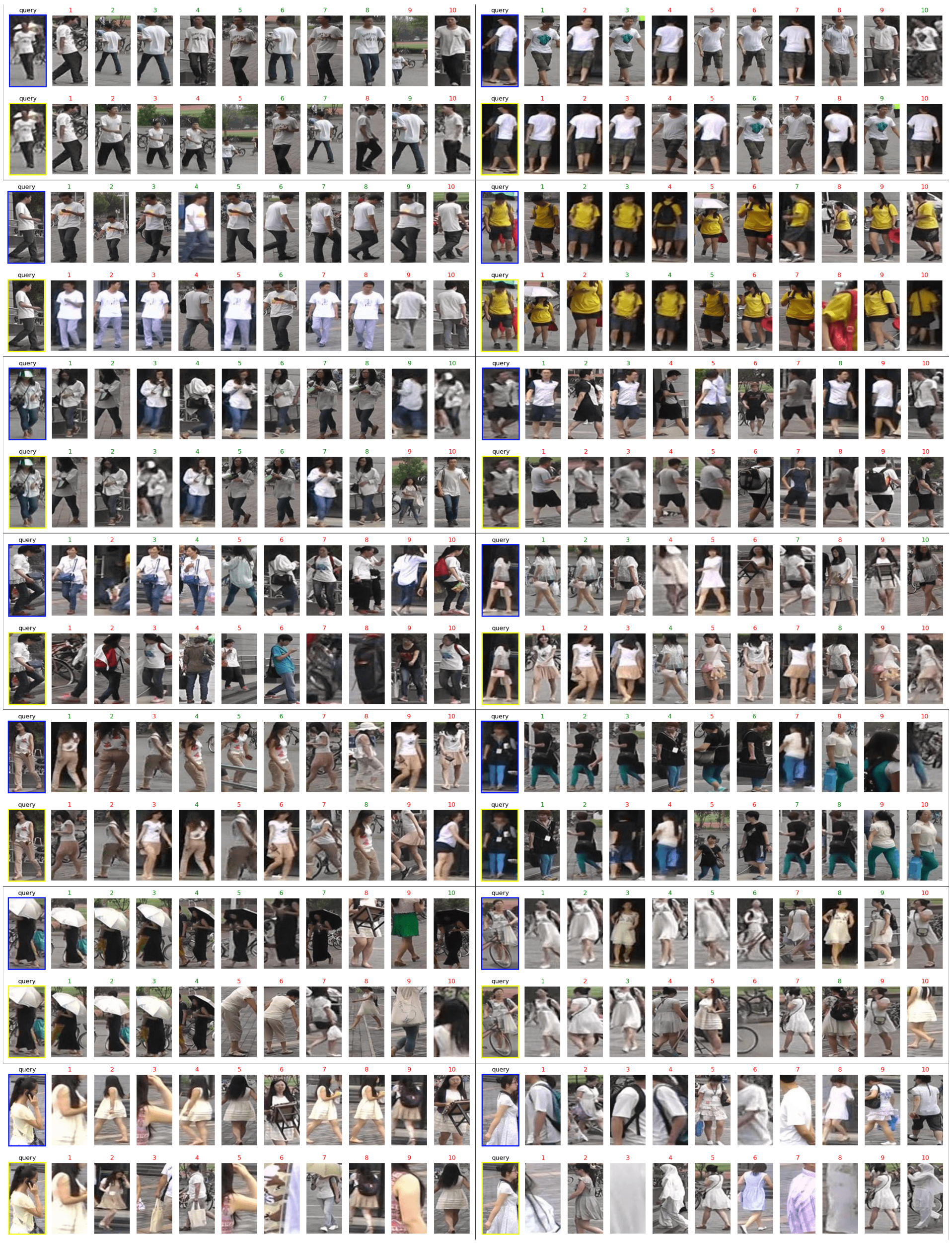


**Fig. 8** Results of re-identification model. Comparison between the results of the base model and the model trained after adding 280 artificial persons. The query is the image of the person to be searched for, and the following images represent the model's output, with green indicating a correct match and red indicating an error.

| **Pers.** | **Img.** | **Rank** |
|---|---|---|
| **0** | **0** | **Rank@1:0.893112 Rank@5:0.964964 Rank@10:0.977732 mAP:0.742632** |
| 10 | 473 | Rank@1:0.888955 Rank@5:0.964074 Rank@10:0.980404 mAP:0.743795 |
| 20 | 874 | Rank@1:0.892221 Rank@5:0.961105 Rank@10:0.974169 mAP:0.745116 |
| 30 | 1252 | Rank@1:0.891627 Rank@5:0.965558 Rank@10:0.979513 mAP:0.741414 |
| 40 | 1.682 | Rank@1:0.890143 Rank@5:0.964371 Rank@10:0.979513 mAP:0.748799 |
| **50** | **2.046** | Rank@1:0.894299 Rank@5:0.962589 Rank@10:0.978028 mAP:0.746853 |
| **60** | **2.427** | Rank@1:0.901128 Rank@5:0.967637 Rank@10:0.980404 mAP:0.751229 |
| 70 | 2.859 | Rank@1:0.892815 Rank@5:0.965261 Rank@10:0.978325 mAP:0.748843 |
| 80 | 3.214 | Rank@1:0.892518 Rank@5:0.965261 Rank@10:0.977732 mAP:0.748257 |
| **90** | **3647** | Rank@1:0.899347 Rank@5:0.964964 Rank@10:0.979216 mAP:0.758927 |
| **100** | **4058** | Rank@1:0.898159 Rank@5:0.964667 Rank@10:0.980998 mAP:0.755366 |
| **150** | **6.167** | Rank@1:0.898753 Rank@5:0.965261 Rank@10:0.979513 mAP:0.761433 |
| **200** | **8.223** | Rank@1:0.896378 Rank@5:0.967340 Rank@10:0.980701 mAP:0.763768 |
| **250** | **10.307** | Rank@1:0.893705 Rank@5:0.964964 Rank@10:0.980107 mAP:0.762484 |
| *280* | *11.244* | **Rank@1:0.903504 Rank@5:0.966746 Rank@10:0.982185 mAP:0.767955** |
| **300** | **12.320** | Rank@1:0.896081 Rank@5:0.963777 Rank@10:0.978919 mAP:0.768727 |
| **320** | **14.371** | Rank@1:0.896675 Rank@5:0.965855 Rank@10:0.978622 mAP:0.769509 |

**Table 6** Results of training with a varying number of synthetic personas. The initial row represents the baseline, with no supplementary images. Bold indicates improvements in results, and italics indicate instances of the best results achieved.

Uc-Cetina. Edition of the text was performed by Anabel Martin-González and Víctor Uc-Cetina.

# 8 Ethical and informed consent for data used

None ethical and informed consent was needed in order to use the Market1501 dataset, which is a third party dataset publicly available at https://zheng-lab.cecs.anu.edu.au/Project/project_reid.html

# 9 Data availability and access

The codes and datasets generated during the current study are available in the person-reidentification GitHub repository at https://github.com/uselessai/person-reidentification

# References

[1] Lata, K., Dave, M., K.N., N.: Data Augmentation Using Generative Adversarial Network. SSRN Electronic Journal, 1–14 (2019) https://doi.org/10.2139/ssrn.3349576 arXiv:arXiv:1711.04340v3

[2] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K.Q. (eds.) Advances in Neural Information Processing Systems, vol. 27. Curran Associates, Inc., ??? (2014). https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf

[3] Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. Proceedings of the IEEE International Conference on Computer Vision **2017-Octob**, 2242–2251 (2017) https://doi.org/10.1109/ICCV.2017.244 arXiv:1703.10593

[4] Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition **2019-June**, 4396–4405 (2019) https://doi.org/10.1109/CVPR.2019.00453 arXiv:1812.04948

[5] Cai, Q., Aggarwal, J.K.: Tracking Human Motion Using Multiple Cameras (1996)

[6] Alqahtani, H., Kavakli-Thorne, M., Liu, C.Z.: An introduction to person re-identification with generative adversarial networks. arXiv, 1–15 (2019)

[7] Luo, Z.: Review of GAN-Based Person Re-Identification (2021) https://doi.org/10.32604/jnm.2021.018027

[8] Jiang, Y., Chen, W., Sun, X., Shi, X., Wang, F., Li, H.: Exploring the Quality of GAN Generated Images for Person Re-Identification vol. 1, pp. 4146–4155. Association for Computing Machinery, ??? (2021). https://doi.org/10.1145/3474085.3475547

[9] Zhong, Z., Zheng, L., Zheng, Z., Li, S., Yang, Y.: CamStyle: A Novel Data Augmentation Method for Person Re-Identification. IEEE Transactions on Image Processing **28**(3), 1176–1190 (2019) https://doi.org/10.1109/TIP.2018.2874313

[10] Dai, P., Ji, R., Wang, H., Wu, Q., Huang, Y.: Cross-modality person re-identification with generative adversarial training. IJCAI International Joint Conference on Artificial Intelligence **2018-July**, 677–683 (2018) https://doi.org/10.24963/ijcai.2018/94

[11] Liang, W., Wang, G., Lai, J., Zhu, J.: M2M-GAN: Many-to-Many Generative Adversarial Transfer Learning for Person Re-Identification (2018) arXiv:1811.03768

[12] Zheng, Z., Yang, X., Yu, Z., Zheng, L., Yang, Y., Kautz, J.: Joint discriminative and generative learning for person re-identification. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition **2019-June**, 2133–2142 (2019) https://doi.org/10.1109/CVPR.2019.00224 arXiv:1904.07223

[13] Pang, Z., Guo, J., Sun, W., Xiao, Y., Yu, M.: Cross-domain person re-identification by hybrid supervised and unsupervised learning. Applied Intelligence (2021) https://doi.org/10.1007/s10489-021-02551-8

[14] Qian, X., Fu, Y., Xiang, T., Wang, W., Qiu, J., Wu, Y., Jiang, Y.-G., Xue, X.:

Pose-Normalized Image Generation for Person Re-identification. The European Conference on Computer Vision (ECCV), 650–667 (2018)

[15] Cao, Z., Simon, T., Wei, S.E., Sheikh, Y.: Realtime multi-person 2D pose estimation using part affinity fields. Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017 **2017-Janua**, 1302–1310 (2017) https://doi.org/10.1109/CVPR.2017.143 arXiv:1611.08050

[16] Borgia, A., Hua, Y., Kodirov, E., Robertson, N.M.: GAN-based Pose-aware Regulation for Video-based Person Re-identification (2019)

[17] Zhang, C., Zhu, L., Zhang, S.C., Yu, W.: PAC-GAN: An effective pose augmentation scheme for unsupervised cross-view person re-identification. Neurocomputing **387**, 22–39 (2020) https://doi.org/10.1016/j.neucom.2019.12.094 arXiv:1906.01792

[18] Ni, Z., Pei, J., Zhao, Y.: Affine transform for skew correction based on generative adversarial network method for multi-camera person re-identification. ACM International Conference Proceeding Series **PartF16898**, 89–95 (2021) https://doi.org/10.1145/3449365.3449380

[19] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the Inception Architecture for Computer Vision. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition **2016-Decem**, 2818–2826 (2016) https://doi.org/10.1109/CVPR.2016.308 arXiv:1512.00567

[20] Zheng, Z., Zheng, L., Yang, Y.: Unlabeled Samples Generated by GAN Improve the Person Re-identification Baseline in Vitro. Proceedings of the IEEE International Conference on Computer Vision **2017-Octob**, 3774–3782 (2017) https://doi.org/10.1109/ICCV.2017.405 arXiv:1701.07717

[21] Ainam, J.P., Qin, K., Liu, G., Luo, G.: Sparse Label Smoothing Regularization for Person Re-Identification. IEEE Access **7**, 27899–27910 (2019) https://doi.org/10.1109/ACCESS.2019.2901599 arXiv:1809.04976

[22] Eom, C., Ham, B.: Learning disentangled representation for robust person re-identification. Advances in Neural Information Processing Systems **32** (2019) arXiv:1910.12003

[23] Hussin, S.H.S., Yildirim, R.: StyleGAN-LSRO Method for Person Re-identification. IEEE Access, 13857–13869 (2021) https://doi.org/10.1109/ACCESS.2021.3051723

[24] Karras, T., Aittala, M., Laine, S., Härkönen, E., Hellsten, J., Lehtinen, J., Aila, T.: Alias-Free Generative Adversarial Networks (NeurIPS) (2021) arXiv:2106.12423

[25] DImitrakopoulos, P., Sfikas, G., Nikou, C.: Wind: Wasserstein Inception Distance for Evaluating Generative Adversarial Network Performance. ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings **2020-May**, 3182–3186 (2020) https://doi.org/10.1109/ICASSP40776.2020.9053325

[26] Jiang, Z., Zhao, L., Shuaiyang, L.I., Yanfei, J.I.A.: Real-time object detection method for embedded devices. arXiv **3**(October), 1–11 (2020)

[27] Yu, Y., Zhang Weibin, D., Yun: Frechet Inception Distance ( FID ) for Evaluating GANs (September), 0–7 (2021)

[28] Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (2004) https://doi.org/10.1109/TIP.2003.819861

[29] Chen, H., Wang, Y., Lagadec, B., Dantcheva, A., Bremond, F.: Joint Generative and Contrastive Learning for Unsupervised Person Re-identification, 2004–2013 (2020) https://doi.org/10.1109/cvpr46437.2021.00204 arXiv:2012.09071

[30] Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable Person Re-identification : A Benchmark University of Texas at San Antonio. Iccv, 1116–1124 (2015)