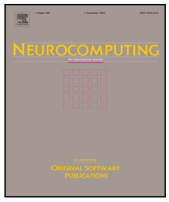
Neurocomputing 598 (2024) 128011

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom



Applying deep learning image enhancement methods to improve person re-identification

Oliverio J. Santana *, Javier Lorenzo-Navarro, David Freire-Obregón, Daniel Hernández-Sosa, Modesto Castrillón-Santana

Institute of Intelligent Systems and Numeric Applications in Engineering, University of Las Palmas de Gran Canaria, 35001 Las Palmas de Gran Canaria, Spain

ARTICLE INFO

Communicated by C. Yang

Keywords:
Artificial intelligence
Deep learning
Computer vision
Image processing
Biometrics
Person re-identification

ABSTRACT

Person re-identification has gained significant attention in recent years due to its numerous practical applications in video surveillance. However, while artificial intelligence and deep learning methods have enabled substantial progress in particular aspects of this domain, putting together those individual advances to generate practical systems remains a computer vision challenge. Existing methods are typically designed assuming the target person's images are captured under uniform, stable conditions with similar lighting levels, but this assumption may not hold in real-world scenarios, such as outdoor monitoring over 24 h, as image quality can vary considerably throughout day and night. In this paper, we propose a framework that incorporates image enhancement techniques to improve the performance of a person re-identification model. The proposed approach achieves a significant improvement in a demanding re-identification dataset, raising the mAP from 9.0% using a zero-shot baseline to 65.8% through the combined use of low-light image enhancement methods and noise reduction.

1. Introduction

In the past few years, artificial intelligence has significantly affected the evolution of surveillance and person identification strategies [1,2]. Traditional biometric features, such as fingerprints or iris patterns, can be used to identify any individual reliably, but to accomplish this requires a collaboration level that may be considered an abuse of privacy in certain environments [3]. In addition, using these features implies that the individual is made aware of being under surveillance, which may not be desirable. Instead, video monitoring systems can capture images of individuals in a non-intrusive way without them being aware of being recorded. The application of computer vision methods to these images makes it possible to extract biometric features related to the physical characteristics of the person's body shape, as well as behavioral characteristics like their pose and their movement patterns [4].

Among the many practical applications that can be developed with this technology, person re-identification (ReId) is one of the most challenging. Given a particular individual captured by a video camera, ReId systems seek to match individuals across the footage recorded by different cameras [5]. From an engineering point of view, implementing these systems faces two significant problems when dealing with real-world scenarios. On the one hand, images may be recorded at different locations, involving changes in the camera angle, the distance to the subject of interest, and the image's background [6]. On the other hand, the images may be recorded a long time apart, which implies that the appearance of the individuals may change [7]. As a result, relying solely on obvious clues like clothing becomes impractical for designing an effective ReId system.

These spatial and temporal changes also give rise to another related issue, which is the wide range of illumination conditions that ReId systems have to deal with. Specifically, this issue becomes pronounced in the context of nighttime or low-light images, where obscured figures and the diminished perceptibility of specific details hampers the performance of computer vision systems [8]. Addressing this problem with hardware solutions is a challenging task. Increasing the exposure time of cameras may allow more light to be captured, but it concurrently introduces undesired noise and motion blur artifacts into the resultant images [9]. Similarly, utilizing infrared technologies presents a potential remedy for low illumination scenarios; however, it entails sacrificing important visual information, including the discernment of color attributes associated with distinct elements [10].

This paper explores the application of deep learning techniques to improve the quality of low-light images and boost the performance of a ReId system working with a challenging dataset containing images of individuals taken in diverse locations under varying illumination conditions [11].

* Corresponding author.
E-mail address: oliverio.santana@ulpgc.es (O.J. Santana).

https://doi.org/10.1016/j.neucom.2024.128011
Received 16 November 2023; Received in revised form 10 May 2024; Accepted 7 June 2024
Available online 10 June 2024

In our experiments, we find that the zero-shot approach using models trained on widely-used ReId datasets yields unsatisfactory results, consistently achieving mAP values below 9.0%, unlike in other scenarios where better performance was observed [12]. Therefore, we train a specific model tailored to this task, resulting in significantly improved outcomes compared to the zero-shot approach, with a mAP of 62.5%.

To further enhance the aforementioned results, we introduce a low-light image enhancement pre-processing step to address night images. This leads to an increase in mAP from 62.5% to 64.9%. Finally, considering that low-light image enhancement methods introduce noise in the enhanced images, we add a final denoising step, resulting in a slight improvement, achieving a final mAP of 65.8%. These results show the great potential that deep learning methods have when combined to generate comprehensive frameworks to solve complex problems.

The remainder of this paper is organized as follows. Section 2 conducts a review of previous related work. Section 3 introduces the proposed ReId pipeline, integrating deep learning models for image enhancement. Section 4 describes the dataset used to evaluate the proposal. Section 5 reports our experiments and discusses our evaluation results. Finally, Section 6 presents our concluding remarks.

2. Related work

Our work explores how using low-light image enhancement methods influences the performance of ReId models in scenarios with challenging illumination conditions. For proper contextualization, reviewing the state of the art in both fields is necessary. We start discussing previous work on ReId, and then we review the deep learning architectures proposed in the literature for low-light image enhancement, as well as image deblurring and denoising.

2.1. Person re-identification

The ReId problem is understood today as matching the visual appearance of people captured with a recording device with images acquired in different places and at different times [5]. Typically, these recording devices are video cameras, which means that the ReId system counts on multiple images of each individual to be identified, something commonly referred to as the multi-shot approach [13,14].

To perform a ReId task, it is necessary to retrieve the distinctive features of the recorded subjects and build with them a representation of each individual that can be compared by the ReId system. It is not feasible to define a fixed set of rules to extract these features due to the significant variability that the images can present: camera position and angle, occlusions, changes in the background, changes in illumination, etc. For this reason, it has become popular to use deep learning methods capable of learning by themselves to extract these features [1,15].

ReId can be considered a classification problem, where each individual is a category, and the deep learning model is trained to distinguish between them [16]. More commonly, a convolutional neural network (CNN) is fed with the images of the individuals. However, instead of using its output, a feature vector is extracted from an intermediate layer [17] and used to compare the individuals and rank them.

Triplet loss [18] is the most frequently used strategy to train these networks. A triplet is a set of three images, two of the same individual (positive samples) and one of a different individual (negative sample). The objective of this training strategy is to achieve a transformation from an initial representation space of the person images to a new one, wherein images of the same individual are mapped closely together. In contrast, images of different individuals are mapped far apart.

An essential factor to consider is the individuals' pose in the images. One strategy to address this problem is to estimate the pose of each individual and realign the image to a standard pose [19]. However, it requires increasing the complexity of the ReId system with additional pre-processing steps and may result in deformed images that hinder ReId. A more straightforward approach is to divide the images into blocks, obtain the features of each block, and align the blocks of the different images to reconstruct the global feature vector of each image [20]. Since the images of individuals undergoing ReId are usually upright, taller than wide, a clever strategy to divide the images into blocks is to split them into horizontal bands [12].

Many works have been published in recent years [2], showing the research community's great interest in the subject. A variety of problems related to the captured individuals have been addressed, such as occlusions [21–23] and variations in viewpoint [24–26], pose [27–30] or scale [31–34]. Issues related to the captured images have also been addressed, like background clutter [35] and low image resolution [36,37]. However, the issue of poorly illuminated images has yet to receive detailed discussion within this context.

2.2. Low-light image enhancement

The subjective perception of the color of an object by a particular individual remains constant regardless of the lighting conditions. This effect, described by the Retinex Theory [38], is the foundation of many algorithms developed to enhance images captured in low light conditions. Retinex algorithms decompose color images into two components: reflectance and illumination. Reflectance describes the object's color, while illumination indicates the intensity of the light incident on the object.

Traditional low-light image enhancement algorithms are computationally expensive and time-consuming [39] due to their iterative nature. Different strategies have been proposed to reduce the computational cost, like focusing just on the illumination component [40] or changing the color space [41]. However, the processing times of these algorithms are still on the order of several minutes per image, making them impractical approaches to deal with large image datasets.

In contrast, deep learning presents a viable solution to this issue. Training a neural network for image enhancement can be accomplished within a reasonable time frame thanks to the computational power of GPUs facilitated by deep learning software libraries, taking into account that it is a once-only process. After the network has been trained, the inference time taken to enhance an image is just a few seconds.

One of the first approaches to improve low-light images using neural networks is the LLNET architecture proposed by Lore et al. [42]. Based on deep auto-encoders, this network can brighten and denoise natural low-light images. Subsequently, researchers have proposed novel architectures and specialized datasets to advance this domain further.

Wei et al. [43] generated a LOw-Light dataset (LOL) with pairs of low-light/normal-light images. This dataset was used to train Retinex-Net, an architecture that decomposes the images following the principles of the Retinex Theory, then adjusts and reconstructs them. Chen et al. [44] built the See-in-the-Dark (SID) dataset pairing short-exposure images with the same long-exposure images as reference. Using this dataset, they developed a low-light image processing pipeline based on a neural network that operates directly on the raw sensor data. Ren et al. [45] designed a residual encoder–decoder architecture [46] trained with the DPED dataset [47]. The edges of the resulting images are later refined using a spatially variant recurrent neural network to prevent fine structure details from being lost.

Some authors have presented alternatives to the standard pattern of direct conversion between low-light images and normal-light images. Cai et al. [48] built a neural network architecture trained with a paired image dataset of low-contrast images and their corresponding high-quality images. It focused on improving images with different exposure levels, both underexposed and overexposed. Wang et al. [49] trained a neural network not to generate improved images, but to generate an intermediate image-to-illumination mapping that can be used later to enhance the low-light images.

Zhang et al. [50] drew inspiration from the Retinex Theory to propose Kind++, an architecture that splits images into two components and processes them in separate branches. The illumination branch

adjusts the lighting, while the reflectance branch restores degrading effects caused by the low light during image capture, such as noise and color distortion. This architecture is an advanced version of the Kind network [51] that has been enhanced by adding a multi-scale illumination attention module to mitigate certain visual defects occasionally generated by the reflectance restoration branch.

Hao et al. [52] proposed a decoupled network that divides the image enhancement problem into two independent tasks. The first stage focuses on improving the illumination of the scene. In contrast, the second stage increases the appearance fidelity by suppressing other degenerative factors that degrade the image's visual quality, such as noise and color distortion. This division aims to prevent the neural network from reaching sub-optimal results while attempting to meet multiple objectives simultaneously.

All the architectures discussed in the preceding paragraphs are based on supervised learning, that is, the network must be instructed on the desirable outcome for each image in the training dataset. This is a limiting factor because it is not possible to have pairs of low-light/normal-light images taken from real or in-the-wild environments. Consequently, these networks are trained with artificial datasets generated under controlled conditions, which may adversely impact the enhancement process when applied to real images.

Yang et al. [53] developed a semi-supervised framework in two stages. The first stage, trained using pairs of low-light/normal-light images, obtains an enhanced image from a low-light input image. The second stage, trained using unpaired high-quality images, further enhances the resulting image to make it more visually appealing for the human eye.

Fully unsupervised learning approaches effectively overcome the need for paired data to train image enhancement models [54]. Some of these models are trained with well illuminated images and learn what the expected result is. Jiang et al. [55] proposed EnlightenGAN, an unsupervised generative adversarial network that enhances low-light images after performing unpaired training with self-regularization. Guo et al. [56] moved away from GAN architectures and presented the zero-reference deep curve estimation method (Zero-DCE), which uses a lightweight deep network to estimate pixel-wise and high-order curves, enabling dynamic range adjustments. This method implicitly measures image quality using non-reference loss functions, thus not needing a paired image dataset for the training process.

More recently, some unsupervised frameworks that execute their training in the opposite direction have been introduced, examining low-light images and learning how to improve them. Liu et al. [57] proposed the Retinex-inspired Unrolling with Architecture Search (RUAS), a framework capable of developing and optimizing low-light image enhancement deep learning networks following the basic image decomposition rule of the Retinex Theory. Adopting a similar approach, Ma et al. [58] proposed a Self-Calibrated Illumination (SCI) learning framework. This framework comprises an illumination estimation module and a self-calibrated module that reduces the computational cost and allows the models obtained using this framework to adapt to general real-world scenes efficiently. Wen et al. [59] take another approach, further developing the idea behind Zero-DCE to design the Self-Reference Deep Adaptive Curve Estimation (Self-DACE) method. This method is divided into two stages; the first one adjusts the luminosity of the scene with newly designed adaptive adjustment curves, more flexible than those used by Zero-DCE, while the second one aims to remove the noise from the image.

Regardless of the chosen model, it should be remembered that low-light image enhancement methods may have collateral consequences, such as increasing the noise level of the images or blurring the edges of the different elements in the scene. Deep learning approaches have also been used recently to address these issues. Kupyn et al. [60] proposed DeblurGAN, a relativistic conditional generative adversarial network designed for single image motion deblurring using a Feature Pyramid Network [61] generator. Chen et al. [62] simplified existing methods, removing the nonlinear activation functions without degrading performance, which led to a Nonlinear Activation Free Network (NAFNet) featuring CA/GELU and Simplified Channel Attention (SCA) along with SimpleGate.

3. Methodology

The primary objective of the ReId task is to establish correspondences between individuals captured by distinct non-overlapping cameras. However, a significant challenge arises when there are substantial variations in illumination conditions across these cameras, such as those encountered between nighttime and daytime acquisitions. Fig. 1 shows the general schema proposed in this work to improve ReId performance addressing this challenge.

3.1. Image acquisition and preparation

The images used in this work belong to the TGC20ReId dataset [11]. The images were captured at different times of the day and night, thus providing a varied collection of pictures acquired under good and poor illumination conditions.

The original footage consists of full HD 1920 × 1080 resolution images captured with Sony Alpha 6400 cameras at 50 frames per second. Bodies are detected in these images with Faster R-CNN [63] using Inception v2 as feature extractor [64], as proposed by Huang et al. [65]. This body detector was trained using the COCO dataset [66].

Only images with detected bodies were considered and, in particular, those with body dimensions (height and width) greater than 30% of the image dimensions. This resulted in many images; thus, one image per minute was subsampled. Finally, the dataset annotators identified the subjects of interest by hand and selected representative images for each individual.

3.2. ReId pipeline

In the context of this paper, we define the problem of ReId as taking the image of an individual (the probe) captured at a particular location and comparing it to a set of images of persons (the gallery) captured at different locations and times. The ReId process is successful when an image in the gallery corresponding to the probe is correctly identified.

The practical implementation of a ReId system requires the development of a methodology to represent each individual's distinctive features as numerical vectors that can be compared. According to the literature's terminology, we refer to these numerical vectors as embeddings. Our ReId pipeline aims to extract the embeddings corresponding to each of the subjects of interest.

Since the primary objective of this paper does not revolve around the computation of embeddings, we employ the AlignedReID++ architecture as an embedding extractor [12]. This architecture uses ResNet50 [46] as the backbone. It adopts a triplet function loss that leverages the Dynamically Matching Local Information (DMLI) technique for dynamically aligning local information without additional supervision. By mitigating the negative effects of inaccurate person detection boxes, DMLI effectively reduces the limitations associated with human pose misalignment.

In order to compute the embeddings, the images of the subjects are cropped to suppress the images of other individuals that may interfere with the ReId process. This cropping process uses the bounding box provided by the body detector as the cutting edge. The cropped bodies are then resized to dimensions of 256 × 128 and fed to the AlignedReID++ architecture. The embeddings are extracted from the last convolution layer of the ResNet50 backbone, resulting in numerical vectors with 2048 elements that comprehensively represent each individual.

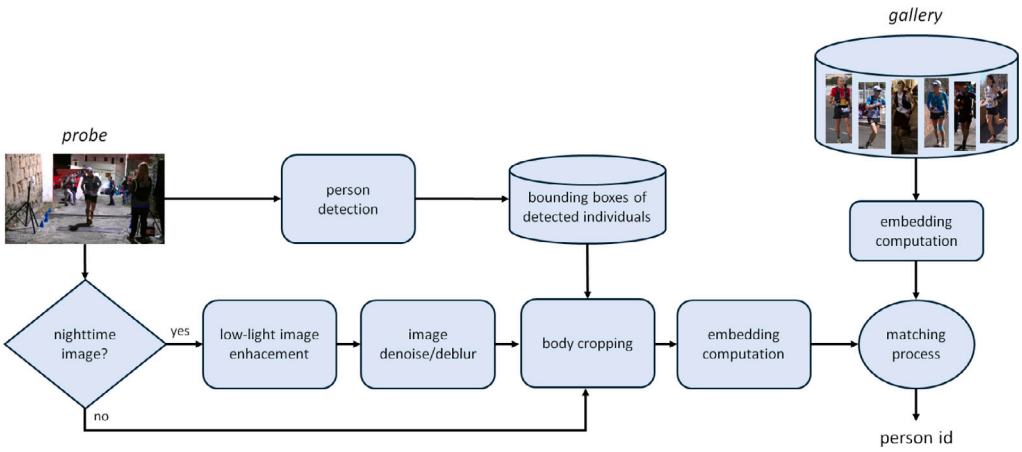


Fig. 1. Proposed schema for ReId with illumination correction.

3.3. Low-light image enhancement

For images captured under good illumination conditions, only the embeddings are computed. When images are captured under poor illumination conditions, two previous stages must be accomplished before the embeddings are computed. The initial stage involves illumination correction using a low-light image enhancement method, which may introduce some noise. Therefore, a post-process is applied to remove the introduced noise and blur.

As can be appreciated in Fig. 1, these two pre-processing stages are applied to the full images. The bodies of the subjects of interest are cropped after the images have been enhanced and then delivered to the feature extractor to perform the ReId process by computing the similarity between the reference image of the probe and all images in the gallery.

This framework represents a significant evolution from the naive approach we introduced in an earlier work [67]. Our previous version overlooked the noise introduced in the enhanced images, resulting in the omission of a crucial post-processing stage. Moreover, in our earlier design, the enhancement method was applied to all images, irrespective of their illumination level, leading to color artifacts in well-illuminated original images.

It is interesting to note that the level of daylight illumination in each image can be easily determined from the metadata of the original footage, which includes the time and date of the recording. This information makes it possible to implement a simple heuristic based on the daytime without needing any laborious image relabeling process.

4. Dataset

To the best of our knowledge, the TGC20ReId dataset [11] we use in this work is the only publicly available ReId benchmark dataset that includes both daytime and nighttime images. This dataset features images of runners participating in the Transgrancanaria ultra-trail race conducted in March 2020 on Gran Canaria island. Due to the extensive length of the race, the first runner completed it in 12 h, while the last runner took almost 30 h.

As the conditions of an ultra-trail race are extremely tough, the runners experience a wide variety of situations, resulting in many different poses. Hence, this challenging dataset is a representative example of the complex situations any real-world ReId system would have to face. Furthermore, since the race started at 11 pm, the dataset encompasses images captured under varying lighting conditions, with day and night footage.

We have considered images captured at four recording points (RPs) distributed along the race as indicated in Fig. 2. Due to the race's starting time, the runners were captured at RP1 and RP2 during the night, illuminated only by artificial light and occasionally by the light from their headlamps. The runners at RP3 and RP4 were captured in broad daylight, except for a few runners captured at dusk in RP4.

Table 1
Time of passage for the first and the last runner considered at each RP.

	First runner	Last runner	Enhanced images
RP1	0:14	1:00	All
RP2	1:19	2:46	All
RP3	10:30	17:34	None
RP4	11:44	19:33	Images captured after 19:00

Table 1 indicates the recording times of the first and last runners at each RP. As an illustrative example, Fig. 3 shows images of some runners at each RP. The points recorded during the day exhibit focused and well-lit images, while the points recorded during the night result in images with low light, noise issues, and even motion blur in some cases.

Since the place and time at which each image was captured is known, a simple heuristic can be implemented in our ReId pipeline to ensure that image enhancement is applied solely to those images that need it. As described in Table 1, all night images captured in RP1 and RP2 are enhanced. In the rest of the RPs, only the images captured at dusk in RP4 are processed. For more complex scenarios, where this information was not known, it would be possible to implement more intricate heuristics using image quality assessment metrics [68].

We generated our ReId test datasets by selecting a group of 109 runners detected at the four RPs from all the identities annotated in the original dataset. In this way, the ReId performance evaluation is conducted on a complete and closed set of identities. The resulting dataset comprises a total of 1269 images. The ReId training dataset consists of 3809 images encompassing 503 runners not included in the test set. These training images were acquired at multiple RPs, but none of the runners were captured at all four RPs. The distribution of the number of identities and images in each RP is shown in Table 2.

ReId training is conducted with images of runners that have been cropped from the footage captured in the different RPs. However, we must use complete images to train the various low-light image-enhancement methods evaluated, and thus we have generated a separate training dataset for this purpose, referred to as TGC20lime, excluding the images from which the runners used in the test dataset were cropped. This dataset contains 5113 complete images, as captured by the cameras, with no additional processing. Of these, 2392 are nighttime images and 2721 were recorded in broad daylight. Table 3 shows the distribution of these images in each RP.

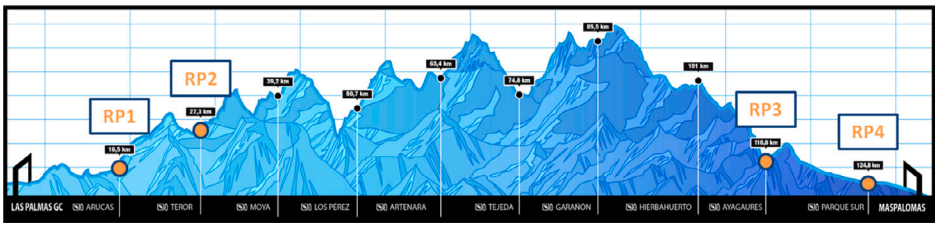


Fig. 2. Track profile of the Transgrancanaria Classic 2020 ultra-trail race (courtesy of Arista Eventos S.L.U.).

Fig. 3. Samples of cropped images captured at each RP.

Table 2
Distribution of identities and images in the ReId datasets.

	RP1	RP2	RP3	RP4
TGC20ReId test identities	109	109	109	109
TGC20ReId test images	290	188	209	245
TGC20ReId train identities	102	168	339	238
TGC20ReId train images	400	450	1560	1399

Table 3
Distribution of images in the low-light image enhancement datasets.

	RP1	RP2	RP3	RP4
TGC20lime night-light images	1109	1283		
TGC20lime day-light images			1545	1176

5. Experiments

To assess the performance of the proposed pipeline, we use the mean average precision (mAP), a widely used metric in ReId problems. Given a number n of probe individuals, mAP is calculated as:

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP_i$$

where AP_i denotes the area under the precision–recall curve for individual i. Hence, this metric considers the trade-off between precision and recall, maximizing both metrics' influence on the evaluation process. In addition, it considers all appearances of each individual in the gallery, recognizing that the same individual may have been captured previously on more than one occasion.

Since the dataset has been generated from a running competition, we preserve the chronological order of the RPs, comparing each probe RP only with the galleries corresponding to previous RPs. This evaluation strategy leads to six tests per experiment, as shown in Table 4.

For brevity, we only report the average and the standard deviation of the six tests for each experiment. This average is calculated as follows:

$$\overline{mAP} = \frac{1}{6}\sum_{i=2}^{4}\sum_{j=1}^{i-1} mAP(RP_i\ vs.\ RP_j)$$

Table 4
Tests performed in each experiment.

Probe	Galleries		
RP2	RP1		
RP3	RP1	RP2	
RP4	RP1	RP2	RP3

Table 5
ReId performance using different AlignedReID++ models and the original non-enhanced images. Each line of the table shows the average mAP value for the six tests conducted in the corresponding experiment and the standard deviation.

Model	mAP
CUHK03	$9.0 \pm 3.2\%$
Market1501	$8.8 \pm 3.0\%$
DukeMTMCReId	$7.5 \pm 2.0\%$
MSMT17	$8.9 \pm 3.0\%$
TGC20ReId	$62.5 \pm 13.1\%$

5.1. ReId models

Using off-the-shelf deep learning architectures to develop image processing pipelines is a widespread practice in the community. However, to make the most of these architectures, it is crucial to train them with the appropriate data for the problem they will be applied to.

Table 5 presents the performance evaluation on the TGC20ReId test dataset of five models trained on different datasets. The first four rows represent a cross-dataset scenario where the models are trained with the CUHK03 dataset [15], the Market1501 dataset [69], the DukeMTMCReId dataset [70], and the MSMT17 dataset [71].

The last row displays the results obtained when the model is trained using the TGC20ReId train dataset. Notably, the models trained with datasets captured under less challenging illumination conditions than the TGC20ReId dataset exhibit a substantially lower performance, highlighting the importance of training the ReId models with appropriate data for the specific scenario under evaluation.

The 62.5% average mAP provided by our model is an excellent result for the harsh conditions imposed by the selected dataset. It is important to keep in mind that not only there are significant variations in the illumination of the different RPs, affecting the perception of the colors of clothing and equipment, but also that this dataset is subject to moderate clothing changes, as runners may swap their outfits between two RPs, making the task even more challenging. Improving performance under these constraints is very difficult and every percentage point gained matters.

5.2. Low-light image enhancement

Table 6 shows the average mAP value achieved by our proposed pipeline applying various low-light image enhancement methods. We

Fig. 4. Original image of a group of runners at RP2 and the corresponding re-lighted versions generated with supervised low-light image enhancement methods.

Table 6
ReId performance using low-light image enhancement methods. The first column of results shows the average mAP value for the six tests conducted in the experiment when only the corresponding low-light image enhancement method is applied. The remaining columns show the mAP when the enhanced images are later post-processed with a deblurring or denoising method; performance improvements are shown in green, and performance degradations are shown in red. The best performance value is underlined.

Enhancement method	Train dataset	mAP	Deblur GANv2	NAFNet denoising	NAFNet deblurring
Retinex-Net (2018)	LOL	38.8	38.8	38.7	38.9
Self-DACE (2023)	SICE	50.7	50.7	50.9	51.0
RUAS (2021)	LOL	51.2	51.2	50.9	50.7
RUAS (2021)	DarkFace	51.3	51.3	51.3	48.9
Self-DACE (2023)	TGC20lime	52.0	52.1	52.6	52.8
EnlightenGAN (2021)	custom	53.6	53.5	53.1	53.8
Kind++ (2021)	LOL	54.3	54.2	53.9	54.1
Decoupled Net (2022)	LOL	58.9	58.8	59.3	59.0
Zero-DCE (2020)	SICE	58.9	58.9	59.3	59.6
RUAS (2021)	MIT5K	59.4	59.5	59.2	59.6
EnlightenGAN (2021)	TGC20lime	59.8	59.6	57.9	58.9
RUAS (2021)	TGC20lime	60.2	60.2	59.6	60.2
SCI (2022)	medium	60.2	60.3	60.8	61.0
SCI (2022)	difficult	61.5	61.5	61.3	62.4
SCI (2022)	easy	63.9	64.0	62.8	63.6
SCI (2022)	TGC20lime	64.7	64.7	62.6	63.8
Zero-DCE (2020)	TGC20lime	64.9	64.7	64.3	65.8

have evaluated three methods trained using pairs of low/normal-light images selected from the LOL dataset [43]. These methods are Retinex-Net [43], Kind++ [50], and the decoupled low-light image enhancement network [52]. Applying these methods fails to improve ReId performance and degrades it.

Fig. 4 shows a nighttime image of some runners who are barely distinguishable in the original footage, as well as the resulting images after processing the original footage with the three supervised methods. The low-light image enhancement methods increase the visibility of the runners, but only the decoupled network provides an acceptable result, albeit with saturated colors. The other two methods show overly saturated colors and considerable noise. Generating new models for adapting these methods to our ReId scenario is not feasible because we are working with outdoor images. In a real uncontrolled environment, obtaining two versions of each image, one with natural lighting conditions and another with desirable lighting, would be highly complicated.

The best strategy for improving ReId performance using enhanced images is to tailor the low-light image enhancement methods to the working scenario. This can only be accomplished with unsupervised learning methods that do not require paired data for training. We have evaluated three unsupervised methods that are trained on nighttime images: RUAS [57], SCI [58], and Self-DACE [59]. The authors of the RUAS framework used it to generate three models derived from the LOL dataset [43], the MIT-Adobe 5K dataset [72], and the DarkFace dataset [73] respectively. The authors of the SCI framework also generated three models, based this time on three levels of difficulty subjectively selected: the easy model was generated from the MIT-Adobe 5K dataset, the medium model was generated from the LOL dataset and the LSRW dataset [74], and the difficult model was generated from the DarkFace dataset. The authors of Self-DACE trained the method to generate a single model based on the SICE dataset [48]. We have adapted these methods to our working scenario by training them with the nighttime TGC20lime dataset.

Table 6 shows that most of the models trained by the original authors of these image enhancement methods result in a ReId performance degradation, falling below the average 62.5% mAP achieved using the original non-enhanced images. These results prove that merely applying an image enhancement method selected from the literature does not guarantee good outcomes. It is possible to achieve an improvement using a zero-shot model, as seen with the SCI method trained with the easy dataset. However, obtaining this improvement would require a trial-and-error approach, applying various methods until one delivers satisfactory results.

Fig. 5 shows that SCI provides strong output, while RUAS and Self-DACE tend to brighten the images excessively, providing overexposed results that alter the colors and hinder ReId performance. In contrast, models generated using the TGC20lime dataset better preserve the original colors. Table 6 shows that the TGC20lime model consistently performs best for each method. The best result is provided by the SCI TGC20lime model, achieving 64.7% average mAP.

We have evaluated two unsupervised methods that were originally trained with well-lit images, Zero-DCE and EnlightenGAN. Guo et al. [56] trained their Zero-DCE model using 3022 images with varying exposure levels selected from the SICE dataset [48]. As for Jiang et al. [55], they trained their EnlightenGAN model using 1930 images manually selected from the RAISE dataset [75] and the LOL dataset [43], as well as high dynamic range images from the SICE dataset [48] and the dataset generated by Kalantari and Ramamoorthi [76]. In order to adapt these methods to our particular scenario, we have generated new models training Zero-DCE and EnlightenGAN with the daytime TGC20lime dataset.

As shown in Table 6, the best mAP results are obtained by the Zero-DCE enhancement method trained using the TGC20lime dataset, a 64.9% mAP value, reasserting the importance of training specific models for the particular task at hand. Zero-DCE does not rely on reference images during training thanks to its carefully formulated non-reference loss functions that measure enhancement quality, making this method more versatile. However, the model trained by the authors tend to over-saturate the colors, resulting in visually appealing images at the cost of not preserving the original tones. Our Zero-DCE TGC20lime model may generate less eye-catching images, but it better preserves the colors, facilitating ReId with images from other locations.

Fig. 5. Image of a group of runners at RP2 re-lighted with unsupervised low-light image enhancement methods. The first row shows the results of unsupervised methods as trained by their authors; when multiple models were available, we selected the best-performing ones, i.e., MIT5K for RUAS and easy for SCI. The second row shows the results of unsupervised methods trained using the TGC20lime dataset.

The EnlightenGAN TGC20lime model also significantly improves over the model trained by the authors, but the resulting images show certain artifacts. As can be observed in Fig. 5, the EnlightenGAN TGC20lime model does not uniformly improve the lighting of the whole scene, rather seems to focus on the sky and the cobblestones, maintaining a dark section in the area of the runners with a distinguishable boundary at the top. The uneven quality of the image enhancement, with more pronounced improvements in the floor and sky compared to the runners, likely arises from the entanglement problem associated with GANs. These networks consist of a generator and a discriminator that work collaboratively to generate images. However, this collaborative nature can result in prioritizing particular image features over others, leading to a disparity in the extent of improvement across different parts of the image [77]. In the case of the image mentioned above, the GAN may have overemphasized enhancing some elements, such as the floor and sky, while not allocating sufficient resources to enhancing the runners, underscoring the challenge of achieving a balanced and uniform enhancement of all elements in the image. This effect ultimately prevents the EnlightenGAN TGC20lime model from reaching the performance achieved by the Zero-DCE TGC20lime model.

5.3. Image denoising and deblurring

Our Zero-DCE TGC20lime model provides the best ReId performance of all the state-of-the-art deep learning methods evaluated, but there is still room for further improvement. Images captured in low-light conditions and enhanced with software methods often exhibit issues such as noise or blurriness. To address these problems, additional enhancement methods can be applied, such as deblurring and denoising.

Firstly, we utilize DeblurGAN [60], whose integration of a Feature Pyramid Network within its generator enables the utilization of different backbones, striking a balance between performance and efficiency based on backbone complexity. The authors trained the DeblurGANv2 architecture with a dataset of approximately 10000 clear/blurry image pairs selected from the GoPro dataset [78], the DVD dataset [79], and the NFS dataset [80]. The blurry images were generated by interpolating the frames [81] of the original 240 fps videos to generate 3840 fps videos and then averaging consecutive frames to simulate the effects of high-exposure image recording.

Secondly, we employ NAFNet [62], which adopts a classical single-stage U-shaped architecture with skip-connections but without non-linear activation functions, which reduces complexity and improves computational efficiency. The NAFNet architecture can be used for various image enhancement tasks based on the dataset used to train it. On the one hand, the authors trained NAFNet with the GoPro dataset mentioned above [78] to generate an image deblurring model. On the other hand, they trained NAFNet with the SIDD dataset [82] to generate an image denoising model. In both cases, these are datasets with paired images, one blurred/noisy image and one clear image representing the desirable result.

The last three columns of Table 6 show the results obtained by applying these denoising/deblurring models after processing the images using the different low-light image enhancement models. We have not trained any denoising/deblurring models because we do not have high-quality clear images to pair with the noisy/blurry images found in the TGC20lime dataset.

In general, the effect is minimal, but a clear pattern in the performance variations can be observed. The DeblurGANv2 network and the NAFNet denoising models do not provide good results. They improve ReId performance for four and five low-light image enhancement methods respectively, but the former degrades ReId performance for five methods and the latter for eleven methods. On the other hand, the NAFNet deblurring model provides a performance improvement for ten of the evaluated methods, including our Zero-DCE TGC20lime model.

In order to provide more insight into this behavior, Fig. 6 shows the original image of a runner captured in RP2, the improved version after applying our Zero-DCE TGC20lime model, and the result of applying different denoising and deblurring methods. The only of the three methods evaluated that has an appreciable effect at first glance is the NAFNet deblurring model, making it the best candidate for integration in our ReId framework.

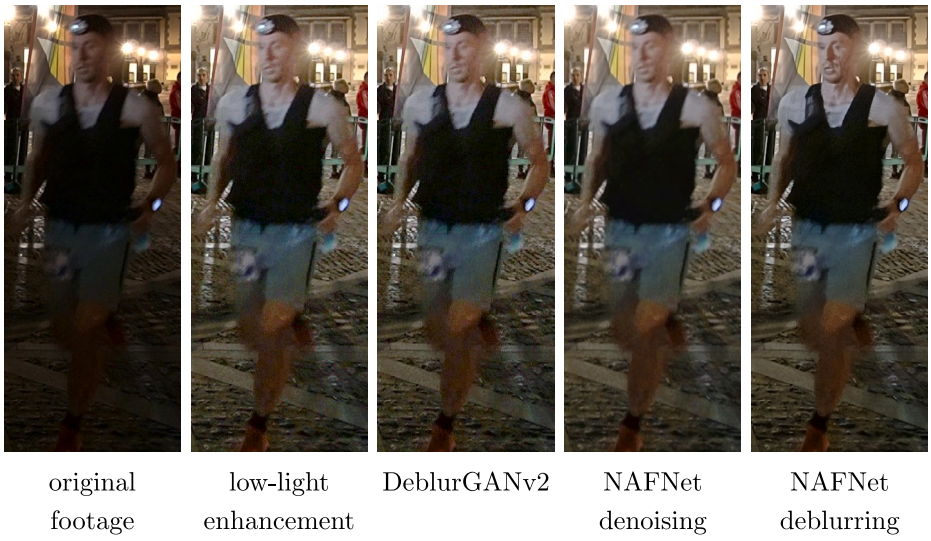


Fig. 6. Image of an individual captured at RP2 with various denoising/deblurring methods. The first image shows the original footage, and the second image shows the footage enhanced using the Zero-DCE architecture trained with the TGC20lime dataset. After low-light image enhancement, the remaining images show the results of applying different denoising/deblurring methods.

Table 7
Detailed mAP, original images.

		Gallery		
		RP1	RP2	RP3
Probe	RP2	65.0%	–	–
	RP3	53.8%	54.6%	–
	RP4	51.0%	58.8%	91.6%

Table 8
Detailed mAP, enhanced images.

		Gallery		
		RP1	RP2	RP3
Probe	RP2	66.0%	–	–
	RP3	56.4%	61.8%	–
	RP4	56.7%	62.0%	91.8%

Table 9
Detailed ranking (Rank-1/Rank-5/Rank-10), original images.

		Gallery		
		RP1	RP2	RP3
Probe	RP2	69.0/84.5/89.3	–	–
	RP3	59.1/72.1/78.4	48.1/74.0/81.7	–
	RP4	54.1/73.0/82.8	54.9/73.0/77.5	91.0/97.1/98.8

Table 10
Detailed ranking (Rank-1/Rank-5/Rank-10), enhanced images.

		Gallery		
		RP1	RP2	RP3
Probe	RP2	71.7/85.0/90.9	–	–
	RP3	60.1/77.9/82.7	60.6/76.9/82.7	–
	RP4	60.9/80.2/84.0	57.6/74.5/80.2	91.4/97.1/98.8

5.4. Discussion by RP

We have achieved the best results by enhancing low-light images using the Zero-DCE architecture trained with daytime images from the TGC20lime dataset and applying the NAFNet deblurring model. Tables 7 and 8 present detailed results per RP for the original non-enhanced and enhanced images, respectively.

As anticipated, the improvement is primarily observed in combinations involving nighttime RPs as probe or gallery, along with daytime RPs as probe or gallery. For instance, when comparing RP3 with RP2, the mAP increases from 54.6% to 61.8% after processing the images with the enhancement pipeline. Similarly, when comparing RP4 with RP1, the mAP improves from 51.0% to 56.7%.

Tables 9 and 10 present ranking metrics for the original non-enhanced images and the enhanced images. For each RP, the tables show the percentage of instances in which the correct individual was placed in the first position (Rank-1), among the first five positions (Rank-5), and among the first ten positions (Rank-10) of the match list. Improvements are observed not only in Rank-1, but also in Rank-5 and even Rank-10 for some combinations, showing the robustness of our system. For example, when comparing RP3 with RP1, Rank-5 rises from 72.1% to 77.9% and Rank-10 rises from 78.4% to 82.7%.

Regarding the comparison between the two RPs filmed during daytime, the observed improvement is slight because this combination presents the highest baseline value by far. Nevertheless, this improvement shows the model's adaptability to illumination variations in the input images. In the daytime RPs, image enhancement is only applied to the runners recorded in RP4 during dusk. The illumination level of the original images changes significantly during this period, as can be seen in Fig. 7, but the results provided by the enhancement process exhibit uniform illumination, which facilitates ReId. A variation in the background shade is observed in the later images, where artificial illumination has more influence, but the subjects are unaffected.

6. Conclusions

Indubitably, the ability to discern the identity of an individual across diverse temporal and spatial junctures is crucial for any video surveillance system in diverse contexts, making it one of the top priorities for academic and industrial research in computer vision. There is a plethora of works on this subject in the literature. However, they tend to present particular solutions to specific problems, which are evaluated using datasets that are not representative of the challenges any real-world video surveillance system would encounter.

This work combines different elements into a single framework designed to handle images captured under highly variable lighting conditions. Our framework provides significant ReId performance improvements thanks to new models trained with datasets composed of both daytime and nighttime images. Finally, the performance is

Fig. 7. Original images recorded at dusk in RP4 (upper row) and their corresponding enhanced versions pre-processed with our Zero-DCE TGC20lime model and the NAFNet deblurring model (lower row). Each column shows the approximate passing time of the depicted individual.

increased even more by integrating deep learning techniques in the ReId pipeline to enhance the quality of nighttime images. The achieved results evidence the significant potential of this strategy to improve the performance of current ReId systems.

CRediT authorship contribution statement

Oliverio J. Santana: Conceptualization, Data curation, Investigation, Methodology, Resources, Software, Visualization, Writing – original draft, Writing – review & editing. **Javier Lorenzo-Navarro:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **David Freire-Obregón:** Conceptualization, Data curation, Investigation, Methodology, Resources, Visualization, Writing – review & editing. **Daniel Hernández-Sosa:** Conceptualization, Data curation, Investigation, Methodology, Resources, Visualization, Writing – review & editing. **Modesto Castrillón-Santana:** Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Visualization, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work is partially funded by the Spanish Ministry of Science and Innovation under project PID2021-122402OB-C22 and by the ACIISI-Gobierno de Canarias and European FEDER funds under project ULPGC Facilities Net and Grant EIS 2021 04. We would like to thank Arista Eventos S.L.U. for allowing us to record images during Transgrancanaria 2020 and granting us their usage for research purposes.

References

[1] Z. Ming, M. Zhu, X. Wang, J. Zhu, J. Cheng, C. Gao, Y. Yang, X. Wei, Deep learning-based person re-identification methods: A survey and outlook of recent works, Image Vis. Comput. 119 (2022) 104394, http://dx.doi.org/10.1016/j.imavis.2022.104394.

[2] A. Zahra, N. Perwaiz, M. Shahzad, M.M. Fraz, Person re-identification: A retrospective on domain specific open challenges and future trends, Pattern Recognit. 142 (2023) 109669, http://dx.doi.org/10.1016/j.patcog.2023.109669.

[3] E.J. Kindt, Privacy and Data Protection Issues of Biometric Applications, Springer Dordrecht, 2013, http://dx.doi.org/10.1007/978-94-007-7522-0.

[4] P. Tome, J. Fierrez, R. Vera-Rodriguez, M.S. Nixon, Soft biometrics and their application in person recognition at a distance, IEEE Trans. Inf. Forensics Secur. 9 (3) (2014) 464–475, http://dx.doi.org/10.1109/TIFS.2014.2299975.

[5] N. Gheissari, T. Sebastian, R. Hartley, Person reidentification using spatiotemporal appearance, in: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR'06, Vol. 2, 2006, pp. 1528–1535, http://dx.doi.org/10.1109/CVPR.2006.223.

[6] A. Bialkowski, S. Denman, S. Sridharan, C. Fookes, P. Lucey, A database for person re-identification in multi-camera surveillance networks, in: 2012 International Conference on Digital Image Computing Techniques and Applications, DICTA, 2012, pp. 1–8, http://dx.doi.org/10.1109/DICTA.2012.6411689.

[7] Y. Huang, Q. Wu, J. Xu, Y. Zhong, Z. Zhang, Clothing status awareness for long-term person re-identification, in: 2021 IEEE/CVF International Conference on Computer Vision, ICCV, 2021, pp. 11875–11884, http://dx.doi.org/10.1109/ICCV48922.2021.01168.

[8] Z. Rahman, Y.-F. Pu, M. Aamir, S. Wali, Structure revealing of low-light images using wavelet transform based on fractional-order denoising and multiscale decomposition, Vis. Comput. 37 (2021) 865–880, http://dx.doi.org/10.1007/s00371-020-01838-0.

[9] P. Wieschollek, M. Hirsch, B. Schölkopf, H.P. Lensch, Learning blind motion deblurring, in: 2017 IEEE International Conference on Computer Vision, ICCV, 2017, pp. 231–240, http://dx.doi.org/10.1109/ICCV.2017.34.

[10] A. Wu, W.-S. Zheng, H.-X. Yu, S. Gong, J. Lai, RGB-infrared cross-modality person re-identification, in: 2017 IEEE International Conference on Computer Vision, ICCV, 2017, pp. 5390–5399, http://dx.doi.org/10.1109/ICCV.2017.575.

[11] A. Penate-Sanchez, D. Freire-Obregón, A. Lorenzo-Melián, J. Lorenzo-Navarro, M. Castrillón-Santana, TGC20ReId: A dataset for sport event re-identification in the wild, Pattern Recognit. Lett. 138 (2020) 355–361, http://dx.doi.org/10.1016/j.patrec.2020.08.003.

[12] H. Luo, W. Jiang, X. Zhang, X. Fan, J. Qian, C. Zhang, AlignedReID++: Dynamically matching local information for person re-identification, Pattern Recognit. 94 (2019) 53–61, http://dx.doi.org/10.1016/j.patcog.2019.05.028.

[13] M. Farenzena, L. Bazzani, A. Perina, V. Murino, M. Cristani, Person re-identification by symmetry-driven accumulation of local features, in: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010, pp. 2360–2367, http://dx.doi.org/10.1109/CVPR.2010.5539926.

[14] L. Bazzani, M. Cristani, A. Perina, M. Farenzena, V. Murino, Multiple-shot person re-identification by HPE signature, in: 2010 20th International Conference on Pattern Recognition, 2010, pp. 1413–1416, http://dx.doi.org/10.1109/ICPR.2010.349.

[15] W. Li, R. Zhao, T. Xiao, X. Wang, DeepReID: Deep filter pairing neural network for person re-identification, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2014, pp. 152–159, http://dx.doi.org/10.1109/CVPR.2014.27.

[16] L. Zheng, H. Zhang, S. Sun, M. Chandraker, Y. Yang, Q. Tian, Person re-identification in the wild, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2017, pp. 3346–3355, http://dx.doi.org/10.1109/CVPR.2017.357.

[17] T. Matsukawa, E. Suzuki, Person re-identification using CNN features learned from combination of attributes, in: 2016 23rd International Conference on Pattern Recognition, ICPR, 2016, pp. 2428–2433, http://dx.doi.org/10.1109/ICPR.2016.7900000.

[18] D. Cheng, Y. Gong, S. Zhou, J. Wang, N. Zheng, Person re-identification by multi-channel parts-based CNN with improved triplet loss function, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2016, pp. 1335–1344, http://dx.doi.org/10.1109/CVPR.2016.149.

[19] L. Zheng, Y. Huang, H. Lu, Y. Yang, Pose-invariant embedding for deep person re-identification, IEEE Trans. Image Process. 28 (9) (2019) 4500–4509, http://dx.doi.org/10.1109/TIP.2019.2910414.

[20] L. He, J. Liang, H. Li, Z. Sun, Deep spatial feature reconstruction for partial person re-identification: Alignment-free approach, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 7073–7082, http://dx.doi.org/10.1109/CVPR.2018.00739.

[21] A. Aich, M. Zheng, S. Karanam, T. Chen, A.K. Roy-Chowdhury, Z. Wu, Spatio-temporal representation factorization for video-based person re-identification, in: 2021 IEEE/CVF International Conference on Computer Vision, ICCV, 2021, pp. 152–162, http://dx.doi.org/10.1109/ICCV48922.2021.00022.

[22] Y. Wang, P. Zhang, S. Gao, X. Geng, H. Lu, D. Wang, Pyramid spatial-temporal aggregation for video-based person re-identification, in: 2021 IEEE/CVF International Conference on Computer Vision, ICCV, 2021, pp. 12006–12015, http://dx.doi.org/10.1109/ICCV48922.2021.01181.

[23] C. Eom, G. Lee, J. Lee, B. Ham, Video-based person re-identification with spatial and temporal memory networks, in: 2021 IEEE/CVF International Conference on Computer Vision, ICCV, 2021, pp. 12016–12025, http://dx.doi.org/10.1109/ICCV48922.2021.01182.

[24] H. Huang, W. Yang, J. Lin, G. Huang, J. Xu, G. Wang, X. Chen, K. Huang, Improve person re-identification with part awareness learning, IEEE Trans. Image Process. 29 (2020) 7468–7481, http://dx.doi.org/10.1109/TIP.2020.3003442.

[25] A. Porrello, L. Bergamini, S. Calderara, Robust re-identification by multiple views knowledge distillation, in: Computer Vision – ECCV 2020, Springer International Publishing, Cham, 2020, pp. 93–110, http://dx.doi.org/10.1007/978-3-030-58607-2_6.

[26] S. Lian, W. Jiang, H. Hu, Attention-aligned network for person re-identification, IEEE Trans. Circuits Syst. Video Technol. 31 (8) (2021) 3140–3153, http://dx.doi.org/10.1109/TCSVT.2020.3037179.

[27] Z. Zhang, C. Lan, W. Zeng, X. Jin, Z. Chen, Relation-aware global attention for person re-identification, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2020, pp. 3183–3192, http://dx.doi.org/10.1109/CVPR42600.2020.00325.

[28] P. Wang, Z. Zhao, F. Su, X. Zu, N.V. Boulgouris, HOReID: Deep high-order mapping enhances pose alignment for person re-identification, IEEE Trans. Image Process. 30 (2021) 2908–2922, http://dx.doi.org/10.1109/TIP.2021.3055952.

[29] T. He, X. Jin, X. Shen, J. Huang, Z. Chen, X.-S. Hua, Dense interaction learning for video-based person re-identification, in: 2021 IEEE/CVF International Conference on Computer Vision, ICCV, 2021, pp. 1470–1481, http://dx.doi.org/10.1109/ICCV48922.2021.00152.

[30] Z. Zhang, H. Zhang, S. Liu, Y. Xie, T.S. Durrani, Part-guided graph convolution networks for person re-identification, Pattern Recognit. 120 (2021) 108155, http://dx.doi.org/10.1016/j.patcog.2021.108155.

[31] X. Qian, Y. Fu, T. Xiang, Y.-G. Jiang, X. Xue, Leader-based multi-scale attention deep architecture for person re-identification, IEEE Trans. Pattern Anal. Mach. Intell. 42 (2) (2020) 371–385, http://dx.doi.org/10.1109/TPAMI.2019.2928294.

[32] N. Martinel, G.L. Foresti, C. Micheloni, Deep pyramidal pooling with attention for person re-identification, IEEE Trans. Image Process. 29 (2020) 7306–7316, http://dx.doi.org/10.1109/TIP.2020.3000904.

[33] Y. Huang, S. Lian, H. Hu, D. Chen, T. Su, Multiscale omnibearing attention networks for person re-identification, IEEE Trans. Circuits Syst. Video Technol. 31 (5) (2021) 1790–1803, http://dx.doi.org/10.1109/TCSVT.2020.3014167.

[34] G. Chen, T. Gu, J. Lu, J.-A. Bao, J. Zhou, Person re-identification via attention pyramid, IEEE Trans. Image Process. 30 (2021) 7663–7676, http://dx.doi.org/10.1109/TIP.2021.3107211.

[35] X. Ning, K. Gong, W. Li, L. Zhang, X. Bai, S. Tian, Feature refinement and filter network for person re-identification, IEEE Trans. Circuits Syst. Video Technol. 31 (9) (2021) 3391–3402, http://dx.doi.org/10.1109/TCSVT.2020.3043026.

[36] Z. Cheng, Q. Dong, S. Gong, X. Zhu, Inter-task association critic for cross-resolution person re-identification, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2020, pp. 2602–2612, http://dx.doi.org/10.1109/CVPR42600.2020.00268.

[37] Z. Feng, J. Lai, X. Xie, Resolution-aware knowledge distillation for efficient inference, IEEE Trans. Image Process. 30 (2021) 6985–6996, http://dx.doi.org/10.1109/TIP.2021.3101158.

[38] E.H. Land, The retinex theory of color vision, Sci. Am. 237 (6) (1977) 108–129, URL: http://www.jstor.org/stable/24953876.

[39] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, X. Ding, A weighted variational model for simultaneous reflectance and illumination estimation, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2016, pp. 2782–2790, http://dx.doi.org/10.1109/CVPR.2016.304.

[40] X. Guo, Y. Li, H. Ling, LIME: Low-light image enhancement via illumination map estimation, IEEE Trans. Image Process. 26 (2) (2017) 982–993, http://dx.doi.org/10.1109/TIP.2016.2639450.

[41] S. Liu, W. Long, Y. Li, H. Cheng, Low-light image enhancement based on membership function and gamma correction, Multimedia Tools Appl. 81 (2022) 22087–22109, http://dx.doi.org/10.1007/s11042-021-11505-8.

[42] K.G. Lore, A. Akintayo, S. Sarkar, LLNet: A deep autoencoder approach to natural low-light image enhancement, Pattern Recognit. 61 (2017) 650–662, http://dx.doi.org/10.1016/j.patcog.2016.06.008.

[43] C. Wei, W. Wang, W. Yang, J. Liu, Deep retinex decomposition for low-light enhancement, in: 2018 British Machine Vision Conference, BMVC, 2018, pp. 1–12.

[44] C. Chen, Q. Chen, J. Xu, V. Koltun, Learning to see in the dark, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, IEEE Computer Society, Los Alamitos, CA, USA, 2018, pp. 3291–3300, http://dx.doi.org/10.1109/CVPR.2018.00347.

[45] W. Ren, S. Liu, L. Ma, Q. Xu, X. Xu, X. Cao, J. Du, M.-H. Yang, Low-light image enhancement via a deep hybrid network, IEEE Trans. Image Process. 28 (9) (2019) 4364–4375, http://dx.doi.org/10.1109/TIP.2019.2910412.

[46] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2016, pp. 770–778, http://dx.doi.org/10.1109/CVPR.2016.90.

[47] A. Ignatov, N. Kobyshev, R. Timofte, K. Vanhoey, DSLR-Quality photos on mobile devices with deep convolutional networks, in: 2017 IEEE International Conference on Computer Vision, ICCV, 2017, pp. 3297–3305, http://dx.doi.org/10.1109/ICCV.2017.355.

[48] J. Cai, S. Gu, L. Zhang, Learning a deep single image contrast enhancer from multi-exposure images, IEEE Trans. Image Process. 27 (4) (2018) 2049–2062, http://dx.doi.org/10.1109/TIP.2018.2794218.

[49] R. Wang, Q. Zhang, C.-W. Fu, X. Shen, W.-S. Zheng, J. Jia, Underexposed photo enhancement using deep illumination estimation, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2019, pp. 6842–6850, http://dx.doi.org/10.1109/CVPR.2019.00701.

[50] Y. Zhang, X. Guo, J. Ma, W. Liu, J. Zhang, Beyond brightening low-light images, Int. J. Comput. Vis. 129 (2021) 1013–1037, http://dx.doi.org/10.1007/s11263-020-01407-x.

[51] Y. Zhang, J. Zhang, X. Guo, Kindling the darkness: A practical low-light image enhancer, in: Proceedings of the 27th ACM International Conference on Multimedia, 2019, pp. 1632–1640, http://dx.doi.org/10.1145/3343031.3350926.

[52] S. Hao, X. Han, Y. Guo, M. Wang, Decoupled low-light image enhancement, ACM Trans. Multimedia Comput. Commun. Appl. 18 (4) (2022) http://dx.doi.org/10.1145/3498341.

[53] W. Yang, S. Wang, Y. Fang, Y. Wang, J. Liu, Band representation-based semi-supervised low-light image enhancement: Bridging the gap between signal fidelity and perceptual quality, Trans. Img. Proc. 30 (2021) 3461–3473, http://dx.doi.org/10.1109/TIP.2021.3062184.

[54] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: 2017 IEEE International Conference on Computer Vision, ICCV, 2017, pp. 2242–2251, http://dx.doi.org/10.1109/ICCV.2017.244.

[55] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, Z. Wang, EnlightenGAN: Deep light enhancement without paired supervision, IEEE Trans. Image Process. 30 (2021) 2340–2349, http://dx.doi.org/10.1109/TIP.2021.3051462.

[56] C. Guo, C. Li, J. Guo, C.C. Loy, J. Hou, S. Kwong, R. Cong, Zero-reference deep curve estimation for low-light image enhancement, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2020, pp. 1777–1786, http://dx.doi.org/10.1109/CVPR42600.2020.00185.

[57] R. Liu, L. Ma, J. Zhang, X. Fan, Z. Luo, Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2021, pp. 10556–10565, http://dx.doi.org/10.1109/CVPR46437.2021.01042.

[58] L. Ma, T. Ma, R. Liu, X. Fan, Z. Luo, Toward fast, flexible, and robust low-light image enhancement, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2022, pp. 5627–5636, http://dx.doi.org/10.1109/CVPR52688.2022.00555.

[59] J. Wen, C. Wu, T. Zhang, Y. Yu, P. Swierczynski, Self-reference deep adaptive curve estimation for low-light image enhancement, 2023, http://dx.doi.org/10.48550/arXiv.2308.08197, arXiv preprint arXiv:2308.08197.

[60] O. Kupyn, T. Martyniuk, J. Wu, Z. Wang, DeblurGAN-v2: Deblurring (orders-of-magnitude) faster and better, in: 2019 IEEE/CVF International Conference on Computer Vision, ICCV, 2019, pp. 8877–8886, http://dx.doi.org/10.1109/ICCV.2019.00897.

[61] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2017, pp. 936–944, http://dx.doi.org/10.1109/CVPR.2017.106.

[62] L. Chen, X. Chu, X. Zhang, J. Sun, Simple baselines for image restoration, in: S. Avidan, G. Brostow, M. Cissé, G.M. Farinella, T. Hassner (Eds.), Computer Vision – ECCV 2022, Springer Nature Switzerland, Cham, 2022, pp. 17–33.

[63] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, IEEE Trans. Pattern Anal. Mach. Intell. 39 (06) (2017) 1137–1149, http://dx.doi.org/10.1109/TPAMI.2016.2577031.

[64] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, in: Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37, 2015, pp. 448–456, http://dx.doi.org/10.5555/3045118.3045167.

[65] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, K. Murphy, Speed/accuracy trade-offs for modern convolutional object detectors, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2017, pp. 3296–3297, http://dx.doi.org/10.1109/CVPR.2017.351.

[66] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C.L. Zitnick, Microsoft COCO: Common objects in context, in: Computer Vision – ECCV 2014, 2014, pp. 740–755, http://dx.doi.org/10.1007/978-3-319-10602-1_48.

[67] O.J. Santana, J. Lorenzo-Navarro, D. Freire-Obregón, D. Hernández-Sosa, M. Castrillón-Santana, Evaluating the impact of low-light image enhancement methods on runner re-identification in the wild, in: Proceedings of the 12th International Conference on Pattern Recognition Applications and Methods - ICPRAM, SciTePress, INSTICC, 2023, pp. 641–648, http://dx.doi.org/10.5220/0011652000003411.

[68] G. Zhai, W. Sun, X. Min, J. Zhou, Perceptual quality assessment of low-light image enhancement, ACM Trans. Multimedia Comput. Commun. Appl. 17 (4) (2021) http://dx.doi.org/10.1145/3457905.

[69] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, Q. Tian, Scalable person re-identification: A benchmark, in: 2015 IEEE International Conference on Computer Vision, ICCV, 2015, pp. 1116–1124, http://dx.doi.org/10.1109/ICCV.2015.133.

[70] E. Ristani, F. Solera, R. Zou, R. Cucchiara, C. Tomasi, Performance measures and a data set for multi-target, multi-camera tracking, in: Computer Vision – ECCV 2016 Workshops, Springer International Publishing, Cham, 2016, pp. 17–35, http://dx.doi.org/10.1007/978-3-319-48881-3_2.

[71] L. Wei, S. Zhang, W. Gao, Q. Tian, Person transfer GAN to bridge domain gap for person re-identification, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2018, pp. 79–88, http://dx.doi.org/10.1109/CVPR.2018.00016.

[72] V. Bychkovsky, S. Paris, E. Chan, F. Durand, Learning photographic global tonal adjustment with a database of input / output image pairs, in: 2011 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2011, pp. 97–104, http://dx.doi.org/10.1109/CVPR.2011.5995332.

[73] W. Yang, Y. Yuan, W. Ren, J. Liu, W.J. Scheirer, Z. Wang, T. Zhang, Q. Zhong, D. Xie, S. Pu, Y. Zheng, Y. Qu, Y. Xie, L. Chen, Z. Li, C. Hong, H. Jiang, S. Yang, Y. Liu, X. Qu, P. Wan, S. Zheng, M. Zhong, T. Su, L. He, Y. Guo, Y. Zhao, Z. Zhu, J. Liang, J. Wang, T. Chen, Y. Quan, Y. Xu, B. Liu, X. Liu, Q. Sun, T. Lin, X. Li, F. Lu, L. Gu, S. Zhou, C. Cao, S. Zhang, C. Chi, C. Zhuang, Z. Lei, S.Z. Li, S. Wang, R. Liu, D. Yi, Z. Zuo, J. Chi, H. Wang, K. Wang, Y. Liu, X. Gao, Z. Chen, C. Guo, Y. Li, H. Zhong, J. Huang, H. Guo, J. Yang, W. Liao, J. Yang, L. Zhou, M. Feng, L. Qin, Advancing image understanding in poor visibility environments: A collective benchmark study, IEEE Trans. Image Process. 29 (2020) 5737–5752, http://dx.doi.org/10.1109/TIP.2020.2981922.

[74] J. Hai, Z. Xuan, R. Yang, Y. Hao, F. Zou, F. Lin, S. Han, R2RNet: Low-light image enhancement via real-low to real-normal network, J. Vis. Commun. Image Represent. 90 (2023) 103712, http://dx.doi.org/10.1016/j.jvcir.2022.103712.

[75] D.-T. Dang-Nguyen, C. Pasquini, V. Conotter, G. Boato, RAISE: A raw images dataset for digital image forensics, in: Proceedings of the 6th ACM Multimedia Systems Conference, 2015, pp. 219–224, http://dx.doi.org/10.1145/2713168.2713194.

[76] N.K. Kalantari, R. Ramamoorthi, Deep high dynamic range imaging of dynamic scenes, ACM Trans. Graph. 36 (4) (2017) http://dx.doi.org/10.1145/3072959.3073609.

[77] Z. Chen, H. Zhao, P. Aarabi, R. Jiang, SC2gan: Rethinking entanglement by self-correcting correlated GAN space, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, 2023, pp. 4457–4466.

[78] S. Nah, T. Kim, K. Lee, Deep multi-scale convolutional neural network for dynamic scene deblurring, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2017, pp. 257–265, http://dx.doi.org/10.1109/CVPR.2017.35.

[79] S. Su, M. Delbracio, J. Wang, G. Sapiro, W. Heidrich, O. Wang, Deep video deblurring for hand-held cameras, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2017, pp. 237–246, http://dx.doi.org/10.1109/CVPR.2017.33.

[80] H.K. Galoogahi, A. Fagg, C. Huang, D. Ramanan, S. Lucey, Need for speed: A benchmark for higher frame rate object tracking, in: 2017 IEEE International Conference on Computer Vision, ICCV, 2017, pp. 1134–1143, http://dx.doi.org/10.1109/ICCV.2017.128.

[81] S. Niklaus, L. Mai, F. Liu, Video frame interpolation via adaptive separable convolution, in: 2017 IEEE International Conference on Computer Vision, ICCV, 2017, pp. 261–270, http://dx.doi.org/10.1109/ICCV.2017.37.

[82] A. Abdelhamed, S. Lin, M.S. Brown, A high-quality denoising dataset for smartphone cameras, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2018, pp. 1692–1700, http://dx.doi.org/10.1109/CVPR.2018.00182.

Oliverio J. Santana obtained his B.Sc. and M.Sc. in Computer Science from the University of Las Palmas de Gran Canaria (ULPGC) in 2000, and his Ph.D. from the Universitat Politècnica de Catalunya in 2005. With significant research experience in computer architecture, his current interests lie in the practical applications of artificial intelligence and deep learning across various fields such as oceanography, satellite data processing, facial expression analysis, and biometric-based person recognition. He currently serves as an Associate Professor of operating systems administration at ULPGC and conducts research at the Institute of Intelligent Systems and Numerical Applications in Engineering.

Javier Lorenzo-Navarro holds an M.Sc. in Computer Science from the University of Las Palmas de Gran Canaria (ULPGC, 1992) and a Ph.D. from ULPGC (2001). He is an Associate Professor in the Department of Computer Science and Systems and a researcher at the Institute of Intelligent Systems and Numerical Applications in Engineering (SIANI), ULPGC. With over 100 papers published in journals and presented at conferences, and having contributed to over 20 research projects. As an Associate Professor, he teaches subjects related to artificial intelligence alongside his research, focusing on computer vision for human interaction and biometrics, person re-identification, machine learning, and data mining.

David Freire-Obregón earned his M.Sc. and Ph.D. in Computer Science from the University of Las Palmas de Gran Canaria (ULPGC), completing his Master's in 2010 and his Ph.D. in 2014. His research primarily explores biometrics and digital forensics, encompassing various topics including machine learning, generative models, image processing, and computer graphics. He currently serves as an Associate Professor at ULPGC.

Daniel Hernández-Sosa is an Associate Professor at the Department of Computer Science and Systems of the University of Las Palmas de Gran Canaria. He received the B.Sc. and M.Sc. degrees in Computer Science in 1992, and the Ph.D. in 2003. He is a researcher in the Institute of Intelligent Systems and Numerical Applications in Engineering. His areas of interest include computer vision, evolutionary optimization, and artificial intelligence techniques applied to autonomous systems.

Modesto Castrillón-Santana received the M.Sc. and Ph.D. degrees in Computer Science from the University of Las Palmas de Gran Canaria University (ULPGC), in 1992 and 2003, respectively. He is currently a Full Professor with the Department of Computer Science and Systems at ULPGC. His main research activities focus particularly on automatic facial analysis, covering various topics related to image processing, perceptual interaction, human-machine interaction, biometrics, and computer graphics. He is a member of the AEPIA and AERFAI-IAPR, having coauthored more than 100 articles, including peer-reviewed international journals, book chapters, and conference proceedings. Additionally, he has acted as an external expert for the Chilean, Italian, Qatari, and Spanish Research Agencies. He serves the community through participation in various conference programs and technical committees. Currently, he co-chairs the International Conference on Pattern Recognition Applications and Methods and serves as an Associate Editor for Pattern Recognition Letters and Image and Vision Computing.