



SyPer: Synthetic periocular data for quantized light-weight recognition in the NIR and visible domains

Jan Niklas Kolf^{a,b,*}, Jurek Elliesen^a, Fadi Boutros^a, Hugo Proença^c, Naser Damer^{a,b}

^a Fraunhofer IGD, Fraunhoferstr. 5, Darmstadt 64283, Hessa, Germany

^b Technical University Darmstadt, Karolinenplatz 5, Darmstadt 64289, Hessa, Germany

^c University of Beira Interior, IT: Instituto de Telecomunicações, Portugal

ARTICLE INFO

Article history:

Received 15 February 2023

Received in revised form 15 April 2023

Accepted 28 April 2023

Available online 02 May 2023

Keywords:

Deep learning

Quantization

Synthetic data

Biometrics

Periocular

ABSTRACT

Deep-learning based periocular recognition systems typically use overparameterized deep neural networks associated with high computational costs and memory requirements. This is especially problematic for mobile and embedded devices in shared resource environments. To perform model quantization for lightweight periocular recognition in a privacy-aware manner, we propose and release SyPer, a synthetic dataset and generation model of periocular images. To enable this, we propose to perform the knowledge transfer in the quantization process on the embedding level and thus not identity-labeled data. This does not only allow the use of synthetic data for quantization, but it also successfully allows to perform the quantization on different domains to additionally boost the performance in new domains. In a variety of experiments on a diverse set of model backbones, we demonstrate the ability to build compact and accurate models through an embedding-level knowledge transfer using synthetic data. We also demonstrate very successfully the use of embedding-level knowledge transfer for near-infrared quantized models towards accurate and efficient periocular recognition on near-infrared images. The SyPer dataset, together with the evaluation protocol, the training code, and model checkpoints are made publicly available at <https://github.com/jankolf/SyPer>.

© 2023 Elsevier B.V. All rights reserved.

1. Introduction

Biometric systems use physical and behavioral characteristics of a person for recognition and are increasingly used in the everyday life [1,2]. Especially in smartphones, such systems are used to unlock the mobile phone or important data and functions. Modern smartphones are equipped with selfie cameras that allow the use of the face as a modality [2,3]. However, using the entire face as a biometric feature is not always possible, especially when wearing face masks as in the COVID-19 pandemic or when parts of the face are obscured by hands, device posture, or head position [4,5]. Instead, it is possible to use the periocular region for recognition in this use-case [6,7], as well as other use-cases restricted by the capturable region, e.g. head-mounted displays [8,9]. This region includes the eye, eyelids, iris, eyebrows and parts of the cheekbone [10]. It has sufficient textural and color information and contains enough uniqueness to allow for good recognition [3,11]. Depending on the application and purpose, either images of the face or the periocular region in the visible spectrum or in the near-infrared

(NIR) spectrum are used. NIR light is not visible or perceptible to humans. These images are used especially in the automotive sector [12]. Images captured in the NIR spectrum are much less susceptible to errors, as the user is not blinded by a light source.

Modern systems for face recognition (FR) and periocular recognition (PR) use deep neural networks (DNN) with a large number of high-precision parameters [13–15]. This causes a high memory footprint and high computational cost, limiting the applicability in devices with shared resources like smartphones [16–19]. An indication of the targeted memory footprint on embedded devices is the ICCV 2019 Workshop challenge on Eye Tracking organized by Facebook [20]. In the competition, the model size seen suitable for embedded and smart devices was set to under 1 MB. To enable the use of DNN on resource-restricted domains, various model architectures for DNN are presented that reduce the number of parameters and calculations to minimize the computational burden [21–24]. Especially for FR lightweight models are developed, such as Mixfacenets [25], ShuffleFaceNet [26] and MobileFaceNet [27]. To transfer knowledge from a larger and more representative model to a smaller and more compact model, a special training technique called knowledge distillation can be used [28,29]. Here, the smaller model is optimized to compute the same output for an input as the larger model, the teacher. In this way, a similar

* Corresponding author at: Fraunhofer IGD, Fraunhoferstr. 5, Darmstadt 64283, Hessa, Germany.

E-mail address: jan.niklas.kolf@igd.fraunhofer.de (J.N. Kolf).

performance can be achieved as with the teacher model. Although less computational effort can be achieved through more compact models, special hardware such as floating-point arithmetic units is still required for execution.

In order to circumvent the high computational effort that arises from floating point arithmetic and to reduce the memory footprint of a DNN, model quantization can be used [30]. In model quantization, the parameter space is reduced from floating-point to integer values that have a much shorter bit width than the floating point number [30–34]. Therefore the representation of these integer parameters requires much less memory, which reduces the overall memory space required for the model in both persistent and volatile memory. Also, depending on the underlying hardware architecture, integer calculations are faster to calculate than floating-point arithmetic, which overall leads to a faster inference time. Deep learning frameworks like Pytorch can run quantized models more efficiently, cutting computation time by a factor of 2 to 4 [35,36]. Typical quantization sizes for integers are 8, 6 and 4 bit [35,37–41]. Recent works [30] applied a pre-trained generative adversarial network (GAN) to synthesize images and use them for label-free quantization of FR models. While the method was used with great success, the GAN used is limited to a pre-trained generator model that synthesized face images.

In our previous work [42], we were the first to show that model quantization for PR can effectively reduce the model size and computational cost for PR with DNN, while keeping the recognition performance at a similar level. However, one of the major drawbacks of the model quantization method used is that the original training dataset is needed for achieving high recognition performance. This is problematic because the use of the original data is often not possible without restrictions due to data protection and privacy, which is of utmost relevance, especially for biometric data of the face and its parts. This also limits the quantization process to data of the same domain (e.g. capture spectrum) and to data with identity labels. The previously applied quantization approach performs the quantization on the classification level and can therefore not be performed with label-free data.

In this work, we present a set of novel contributions towards building lightweight PR models in a privacy-friendly framework. The same framework is also used to quantize PR models in a manner that adapts their performance to new domains, the NIR domain in our case. This work is also the first to target the synthetic generation of images of the wider periocular region and to use them in the quantization process. Towards these contributions, we adapt the capabilities of the GAN StyleGAN2-ADA [43], that is extensively used in research and showed the success of the architecture in multiple fields, and create a synthetic periocular image dataset and generation model, SyPer. With a specifically trained generator for PR we differ from the approach used in [30] that applied a pre-trained GAN. The created SyPer dataset allows us to train quantized models completely without access to privacy-sensitive authentic data.

To perform quantization without requiring training data with identity labels, we propose using embedding-level knowledge transfer. In our proposed approach, the expertise of a pre-trained, full-precision teacher network, that uses floating-point arithmetic, is transferred to quantized student network. As a teacher network, a DNN that was previously trained on authentic data with identity labels is used. The student is utilizing the teacher network with the trained weights, but the network weights are quantized to integers before knowledge transfer is applied. With this approach we overcome one major challenge of our previously applied quantization approach in [42]. We take advantage of this proposed method in two aspects. First, by enabling the quantization process using our label-free synthetic SyPer data in a privacy-aware manner (i.e. without using privacy-sensitive authentic data with identity labels) with a very minor drop in the performance in comparison to quantization based on authentic data, eliminating the other major challenge of our previous approach [42]. Second, by enhancing the performance

of the quantized model in a domain different than that of the full precision model. This is achieved by the embedding-level knowledge transfer with data from the NIR domain, where the teacher model trained on visible spectrum images (more widely available) is quantized using NIR images. With this, we successfully demonstrated the ability of NIR-quantization that can act as a domain adaption to the NIR domain with promising results. We successfully demonstrate the effectiveness of our method on three diverse model backbones on the evaluation protocols of the UFPR-Periocular and CASIA-Iris-M1-S3 databases. As a further contribution, we will release the training code as well as the created synthetic dataset and generation model (SyPer), the landmarks, and the cropping algorithm for all datasets used. The data, code, and pre-trained models are to be made available at <https://github.com/jankolf/SyPer>.

2. Related work

Previously works on efficient methods for biometric recognition, especially for mobile devices, were presented with different hand-crafted and deep learning-based approaches. Scale Invariant Feature Transformation (SIFT), among others, by Ahuja et al. [44], Alonso-Fernandez et al. [45], Ross et al. [46] or Raja et al. [47]. Park et al. [48] applied Local Binary Patterns (LBP) to extract several features of the periocular region. To generate a feature set of the periocular region Tan et al. [49] used Leung-Mallik filters. In [50] Alonso-Fernandez et al. apply successfully Symmetry Assessment by Feature Expansion descriptors for PR. With the advent of the new wave of deep learning, DNN were used more frequently and successfully. A convolutional version of Restricted Boltzman Machines was used by Nie et al. [51] in an unsupervised learning approach to extract features. With a focus on smartphones, Raja et al. [52] also used deep learning with deep sparse filtering to extract features from the periocular region. Zhang et al. [53] combined the iris and the periocular region and applied a convolutional neural network (CNN) to extract and fuse features for biometric recognition. Unsupervised learning with a convolutional autoencoder was applied by Reddy et al. [54], beating a supervised trained ResNet50 in their evaluation. To allow recognition in unconstrained scenarios with the periocular region Zanlorenzi et al. [55] reduced the within-class variability with a generative model. Alonso-Fernandez et al. [56] used several feature extractors to combine them into one. They also successfully conducted cross-spectrum experiments between visible and NIR images. Reddy et al. [57] proposed OcularNet which is using several small CNNs, extracting from six overlapping periocular patches individual features and fuse them together, beating a ResNet50 with a model that is 15.6x smaller. Pruning techniques were used by Almadan et al. [58], creating an efficient DNN for ocular recognition aimed towards mobile devices. Boutros et al. [28,59] used knowledge distillation to create a small student network that learns from a larger teacher model. This was applied either with a traditional setup [21,28] or on the template level [59]. Despite these efforts towards efficient PR, only the work that we extend in this paper [42] proposed model quantization, but it limits itself to quantizing models evaluated on images of the visible domain and that are quantized using privacy-sensitive authentic images with identity labels.

A main contribution of this work is to generate synthetic periocular images and use them for model quantization, eliminating the need of identity labels in the quantization process. Previous works touched on generating iris and ocular identity-specific images, however, given a set of authentic images of that identity and an arbitrary semantic segmentation [8,60]. The only closely related work on synthetic data generation is that by Tomasevic et al. [61]. They used a generative model to create synthetic ocular images in the visible and near-infrared domain with the main goal of creating segmentation training data for ocular regions. Simultaneously, a semantic segmentation mask is created as a segmentation label. However, this was not a generation for the whole wide periocular region,

making our work here the first to present the generation of synthetically generated data (both the model and data will be made public).

As PR techniques are strongly related to those of FR, we shortly go over related works that addressed FR. Qui et al. [62] use synthetization models trained on face images to generate artificial face images. In order to increase the performance and reduce the domain gap, among other things caused by low intra-class variations, an identity and domain mixup is used. A 3d morphable face model is used by Kortylewski et al. [63] to synthetically generate images in different poses and illuminations. This, together with realistic data, has allowed the authors to achieve an improvement, but the synthetic data itself is not sufficient to achieve good recognition performance. In a follow up paper [64] the authors have identified the dataset bias of synthetic data and show that they can be used to reduce the required authentic images, i.e. non-generated images. In [30], Boutros et al. used a pre-trained data generator for synthetic face images to train a quantized model from a full-precision and larger teacher model using knowledge transfer. In [65], the authors created synthetic facial images with identity information. This means that no privacy-sensitive authentic images are needed to train FR models. In another work, Boutros et al. [66] also used a generator system to generate synthetic images, but this time they used further augmentation methods and an unsupervised learning approach to train FR models. However, synthetic periocular images were never previously proposed to be generated, nor used to train or quantize a recognition model.

In this paper, we are the first to investigate and propose the combined use of synthetic generation to create a synthetic periocular image dataset together with model quantization and knowledge transfer. We overcome the challenge of our previous work that requires privacy-sensitive images with identity labels by applying model quantization on the embedding level using synthetic, label-free data. While efficient models and learning techniques such as knowledge transfer and model quantization have been explored for FR, as well as synthetic datasets, recent works [30] rely on pre-trained generators with a focus on face images [30] or privacy-sensitive data for periocular images [42]. Overall, the proposed harmonization between the used techniques and limitations imposed by the domain and data restrictions have not yet been applied to PR.

3. Methodology

The goal of representation learning is to create an embedding vector from an input image. In the case of biometrics, this vector encodes the identity of the person depicted in the image, and the system encoding the identity information is optimized to create embedding vectors of samples of an identity that are similar with respect to a similarity measure. Creating and learning such an embedding vector is strongly dependent on the loss function used. In this paper we use an angular margin penalty-based softmax loss, Arcface [15]. This loss function is used to train a full-precision network that uses floating-point parameters. To train a compact network from such a network, the network is quantized, i.e. all parameters are mapped from floating-point space towards integer space. This process is described in Section 3.1 and is also used later in the work to enhance PR in the NIR domain. A core component of this work, the generation of synthetic data for privacy-friendly model quantization, is explained in Section 3.2.

3.1. Model quantization for periocular models

3.1.1. Model quantization

The aim of model quantization is to change the parameters and the activation functions from floating-point, which are often represented with 32 bits, to integers with the bit width b . Integers with different bit widths such as 8, 6, or 4 bits can be used [37–39]. The bit width for parameters and activation functions can also be changed individually [39,67]. In order for the network to have

similar performance, the DNN must behave almost identically with integer parameters as with floating-point numbers, which is not the case in practice. Therefore, an extra fine tuning process called quantization-aware training (QAT) is needed [67]. This is described in more detail in Subsection 3.1.2. The fact that the parameters are only represented with b bits reduces the memory requirements of the model and speeds up the computation of a forward pass [35,36]. According to the IEEE standard [68] for floating-point numbers with 32 bits (FP32) their value range r is $r \in [-3.4 \times 10^{38}, 3.4 \times 10^{38}]$ [69]. Since regularization methods and certain activation functions are used in the training of DNN, the complete value range of FP32 is not utilized. In practice, the values of the weights and the activation function of the DNN are in a limited range of values. This range of values can now also be represented by fewer numbers. A signed integer r_Q with bit size b has 2^b different values and a value range of $r_Q \in [-2^{b-1}, 2^{b-1} - 1]$. If 256 different numbers are sufficient to represent the range of FP32 parameters and the transformations between individual latent spaces of the DNN, this value range can be represented by a $b = 8$ bit signed integer. Let $r \in [\alpha, \beta]$ [70] be the minimum and maximum occurring value within an FP32 value range. The representation is realized by a mapping from the FP32 space to the integer space. The minimum value α is mapped to the lower limit of the integer and β to the upper limit. The range of values in between can either be mapped uniformly to the integer space, where the individual values are equidistant from each other, or non-uniformly, where the values are spaced differently. In this work, uniform mapping is used. It is defined as [39,70].

$$Q(r) = \text{round}\left(\frac{r}{S}\right) - Z, \quad (1)$$

where $\text{round}(\cdot)$ rounds the floating-point number to the nearest integer. The constant Z specifies the zero-point shift and the mapping between the two value ranges. It is represented as a b bit signed integer. S is the constant scaling factor and divides the value range into equally sized partitions. The constants are calculated as [39].

$$S = \frac{\beta - \alpha}{2^b - 1}, \quad (2)$$

$$Z = \text{round}\left(\beta \cdot \frac{2^b - 1}{\beta - \alpha} - 2^{b-1}\right). \quad (3)$$

The dequantization operator reconstructs the original FP32 value from a quantized input [39,70]:

$$D(r_Q) = S \cdot (r_Q + Z). \quad (4)$$

This dequantization operation is associated with a loss of information since a target value can be occupied several times when projecting onto a range of values with fewer different values. The upper and lower bounds of the value range r , α , and β , can be calculated either dynamically or statically [70]. In the dynamic calculation, the upper and lower bounds are saved during the fine tuning of the quantized model and then fixed after a certain time. In this work, we use dynamic calculation because it provides better parameter mapping. Quantization granularity refers to the calculation of ranges for specific groups of parameters. In a DNN, the convolution filters and fully-connected layers can have different parameter ranges. Therefore, they can also be quantized differently. We follow our previous work and use channel-wise quantization, where the bounds are calculated separately for filters of a convolution [42,71,72].

3.1.2. Quantization aware training

When a DNN is quantized, the parameters are not trained to deal with the information loss due to the mapping. In order to achieve similar performance, QAT training is used. In this case, the mapping between

the value ranges is adjusted once again on a dataset [70]. In this way, the quantization error can be reduced, which leads to a better performance of the model. In practice, this is implemented by simulating the quantization operations in the network by mapping an input FP32 value to an integer value and then dequantizing it again. The backward pass through this operation is implemented by a straight-through estimator (STE) [73], where the derivative is implemented as a static 1 and thus the gradient is passed on directly [70]. The original dataset can be used as the training set, if available. If this is not available, e.g. due to privacy, data protection reasons, or the use of label-less synthetic data, another dataset can be used with the help of knowledge transfer, as we will discuss in the next section. In this case, the pre-trained FP32 model is used as the teacher and the quantized model as the student. The quantized model learns to predict the original output and can adjust the individual upper and lower bounds, which leads to better performance.

3.1.3. Embedding-level knowledge transfer

QAT typically requires the original labeled training dataset to fine-tune the quantized model [39]. This is often impossible due to privacy concerns or for the lack of labels [74]. Emerging works proposed to fine-tune a quantized model with generated images [39,75]. Unlike supervised learning for periocular images, where the identity labels are used to learn a good separation of the individual identities, embedding-level knowledge transfer (KT) does not use identity labels. Instead, a teacher and student model is used, with the aim of transferring the knowledge about encoding identities from an input image into a vector representation of the teacher model to the student. This is achieved by giving an image I as input to both networks and calculating the embedding vectors for both the teacher, e_t and the student, e_s . The loss function is mean squared error, where the distance of the two embedding vectors is calculated:

$$L = \frac{1}{d} \sum_{i \in N} (e_{s,i} - e_{t,i})^2$$

The student is optimized to calculate embeddings that are as identical as possible to those of the teacher. The original datasets are often not used for KT because they are no longer available or cannot be used for privacy reasons. Instead, synthetic data is used more frequently. The generation of synthetic periocular images is one of the core contributions of this work and is described methodically in Section 3.2.

3.2. SyPer - synthetic periocular image dataset

For QAT by KT, a dataset is needed that has, at best, very similar distribution as the original training data. In reality, this data may not be available, especially in biometrics, as it contains sensitive information and is subject to special regulations. Therefore, methods are needed to optimize and quantize models privacy-friendly. Generative Adversarial Networks (GAN) introduced by Goodfellow et al. [76] are a method established by a growing body of research to generate synthetic images. In this minimax-based learning method, a generator network \mathcal{G} and a discriminator network \mathcal{D} are used. The generator creates an image from a latent vector z . The discriminator receives the generated image and authentic images from a training set. The discriminator tries to decide which images examined by it are authentic or artificially generated. \mathcal{D} is optimized to discriminate artificially generated images from authentic images from the training set. \mathcal{G} is optimized to trick the discriminator into classifying the generated image as an authentic image. This creates a competing scheme by which \mathcal{G} is optimized to generate images as similar as possible to the training set. In this work, we use the established StyleGAN2-ADA [43] as the synthesis model, as it combines the properties of stable training with good synthesis ability. Preparation of the training data and the training process of the GAN are described in Section 4.3.

4. Experimental setup

In this section, the used and generated datasets are presented along with details about the implementation, training parameters, and evaluation procedure. All the training processes are performed on a GPU server with 2×16 core Intel Xeon Gold 6130 s, 256GB RAM, and 4 Nvidia GeForce RTX 2080 Ti 11GB GPUs is used as hardware for the model training.

4.1. Authentic visible periocular dataset

The UFPR-Periocular dataset by Zanlorensi et al. [80] is one of the most comprehensive and diverse periocular datasets. In total, using 196 different mobile devices, 33,600 images from 1,122 subjects were taken of both the left and right periocular regions. The images were taken in 3 different sessions per participant, with a minimum interval of 8 hours between sessions to allow for more variation in the dataset. There were also no restrictions on the recording, which took place in uncontrolled environments. This results in blur, occlusion, or poor lighting in the images, as one would expect in reality. In this way, a highly diverse dataset with a realistic scenario is achieved. The pre-processing including alignment and splitting of the images was performed by the authors of the dataset. The dataset consists of various training and evaluation protocols covering the individual scenarios of a biometric comparison. The closed world protocol covers the identification scenario. The goal is to find the identity from a given database with a $1 : N$ comparison. In this protocol, the identities are included in both the training and test sets. The verification scenario, where a $1 : 1$ comparison takes place with the aim of deciding whether the comparison is between the identical identity or different identities, is covered in the open world/closed validation and open world/open validation scenarios. In these protocols, the identities in the test set are not included in the training or validation set. The two protocols differ in the structure of the evaluation set. In the closed validation case, the identities in the training and evaluation set are the same, in the open validation the identities in all splits, training, validation, and test, are different. Each protocol is in three folds and is evaluated in cross-validation fashion. We follow the approach of Uzair et al. [81] and flip the images with label left so that they have the same orientation as the periocular images of the right side. The images are scaled to the common size 224×224 pixels and normalized with a mean of 0.5 and a standard deviation of 0.5. We follow the dataset authors and train and evaluate our FP32 models for both the closed world protocol for the identification task and open world/closed validation for the verification task. We also use the same evaluation metrics, which are described in more detail in the Subsection 4.6.

4.2. Authentic near-infrared periocular dataset

CASIA-Iris-M1-S3 [79] is used as the database in the NIR spectrum. The dataset contains 3600 authentic images of 360 subjects. The images were captured with a mobile NIR iris scanner. As the images in the dataset do not contain highly-detectable landmarks, they are annotated manually. Bounding boxes are placed around the eyes, the tip of the nose, and the corner of the mouth and the respective center point is determined. Not all landmarks are included in every image; regularly the corners of the mouth are not completely visible. With the visible landmarks, the image is aligned and the periocular regions are cropped. A detailed description of the alignment is in Section 4.3, and a graphical representation is shown in Fig. 2. The first 180 identities of the dataset are used as the training set, the remaining 180 as the test set, as defined in [79]. The sets are identical for the respective folds of the UFPR-Periocular models and are averaged over all 3 models used. The evaluation on the dataset takes place in verification mode, where each periocular image from the test set is compared with every other image from a different identity. Images of the same identity are compared if

they are from the same side of the eye. Overall 6,460,200 comparisons are made. All FP32 models trained on the identification and verification protocols of the UFPR-Periocular dataset are tested on the NIR dataset. The same metrics as for the visible spectrum dataset are used.

4.3. Our SyPer dataset

To allow QAT through KT to be applied on pre-trained full-precision PR models, realistic synthetic periocular images are needed. To generate such data, a GAN is trained. The Flickr-Faces-HQ (FFHQ) dataset by [77] is used as a foundation for the generator training. To extract the periocular region, for each image of the dataset, the landmarks for eye centers e_r and e_l , nose tip, and mouth corners are calculated using a pre-trained Multitask Cascaded Convolutional Networks (MTCNN) [78] implementation. Using these landmarks, the image is aligned to a normalized size of 1000×1000 pixels. A replicated border of 250 pixels is placed around the image. The horizontal distance between the two eye centers is calculated as $d = \|e_{l,x} - e_{r,x}\|$. The periocular region is considered to be a square with edge length $\frac{d}{2}$ centered on the eye landmark. This periocular region extraction follows the procedure defined in [80]. The image extracted from the region is scaled to the final size of 256×256 pixels. The periocular region extracted from the left side of the image is flipped horizontally so that the left and right periocular regions have identical alignment. In Fig. 1 the whole alignment process is illustrated. Since the landmarks can not be recognized in all input images, a total of 139,974 periocular images are cropped from the 70,000 images of the FFHQ dataset. As the underlying GAN network setup StyleGan2-ADA by [43] is used, with the Pytorch version published by the authors being used for training. We follow the training setup of the reference implementation but disable image flip augmentations to reduce alignment errors in the synthesis process. The StyleGan2-ADA is trained on 25,000,000 image iterations. From the trained model, 99,840 synthetic periocular images are extracted based on randomly samples latent vectors. Each generated image is scaled to the final resolution of 224×224 pixels. Example images cropped from the FFHQ dataset and synthesized from the trained GAN are shown in Fig. 4. Both the SyPer dataset and the SyPer trained model will be made publicly available.

4.4. Model architectures and full-precision training

The full-precision FP32 baseline models are each trained with the ArcFace [15] loss on the UFPR-Periocular dataset. Each FP32 model is trained once on the open world/closed validation protocol of the UFPR dataset for all three folds. These models are referred to as verification models in this paper. The models are also trained for all folds on the closed world protocol for the identification task and are referred to as identification models. We use the same parameters shown to be

optimal in our previous work and set $s = 64$, $m = 1.0$, with an embedding size of 512. Well established and performing architectures, ResNet18 [82], ResNet50 [82] and MobileFaceNet [27], are used as backbone. The variation in the backbone architecture and model size are aimed at proving the generalizability of the proposed procedure. The models are trained with a learning rate of 0.1 on 20 epochs. The model with the best performance on the validation set is selected. We also reduce the learning rate at a factor of 10 after the 8th and 15th epochs. Momentum of 0.9 together with weight decay of $5e - 4$ are used with the Stochastic Gradient Descent optimizer. A batch size of 16 was used for MobileFaceNet and ResNet18, and a batch size of 8 for ResNet50. Pytorch is used as the framework.

4.5. Quantization procedure

After pre-training the FP32 baseline models on authentic, visible spectrum UFPR images, all verification and identification models are quantized to 8 (W8A8), 6 (W6A6), and 4 (W4A4) bit integers. These models are referred to as quantized models. The QAT is performed using three different databases, depending on the targeted experiment.

The first database used is the training set of the authentic UFPR-Periocular database. The second database is our novel synthetic dataset for periocular images, SyPer. The training set of the NIR CASIA-Iris-M1-S3 dataset is used as the third database.

In the experiments using the first, authentic database, the QAT for the quantized models is performed using all three folds of the UFPR-Periocular training data using the respective verification (open world/closed validation) or identification (closed world) protocols. The quantization is performed on the classification level, requiring the identity labels of the respective training dataset. After QAT the models are tested on the respective testset of the used protocol. The models that are quantized using authentic data are labeled as "Auth."

In the case of the second database, the baseline FP32 models are quantized and QAT is performed with the synthetic images of SyPer, allowing the QAT to be performed in a privacy-friendly manner. This is achieved by utilizing knowledge transfer, where the QAT is performed on the embedding level. This significant contribution of the SyPer data generation is described in Section 4.3. After the QAT is finished, the models are tested on the respective authentic testset of the UFPR-Periocular protocols used for the FP32 model training. All models quantized on SyPer are labeled as "Synth." or "VIS / Synth.", as they are quantized on synthetic data in the visible spectrum.

The NIR CASIA-Iris-M1-S3 dataset, the third database used in our experiments, is applied for QAT on the baseline models as well. All FP32 models that were trained on the UFPR-Periocular identification and verification protocols are quantized and QAT is performed on the training set of CASIA-Iris-M1-S3. The applied procedure for QAT is knowledge transfer on the embedding level. After QAT all quantized

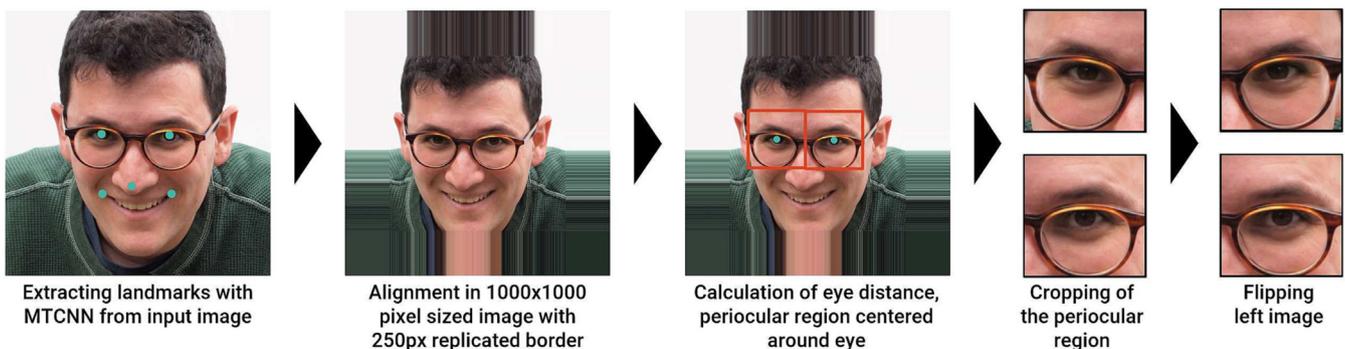


Fig. 1. The alignment process for the FFHQ dataset [77] differs from the NIR alignment process in that the landmarks are extracted automatically by MTCNN [78]. A detailed description of the alignment can be found in Section 4.3.

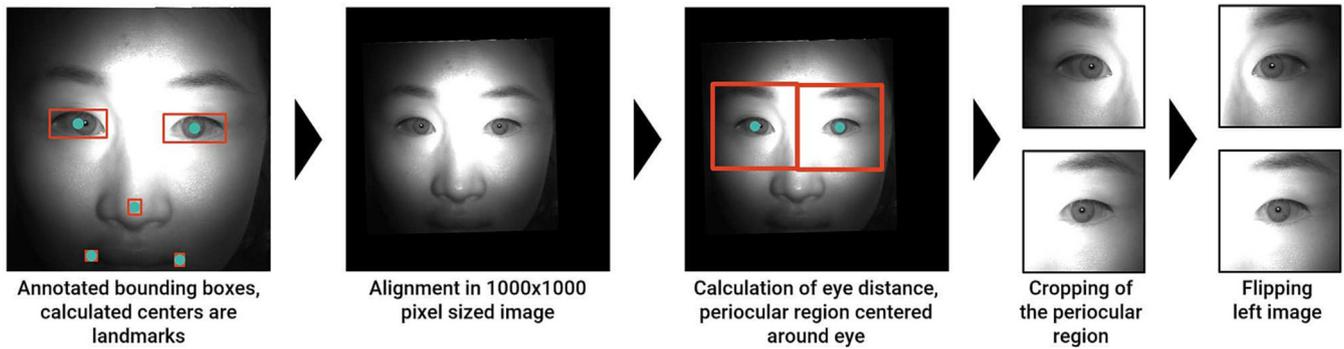


Fig. 2. The alignment process for the CASIA-Iris-M1-S3 dataset [79] differs from the visible spectrum alignment in that the landmarks for this dataset were annotated by hand. A detailed description of the alignment can be found in Section 4.3.

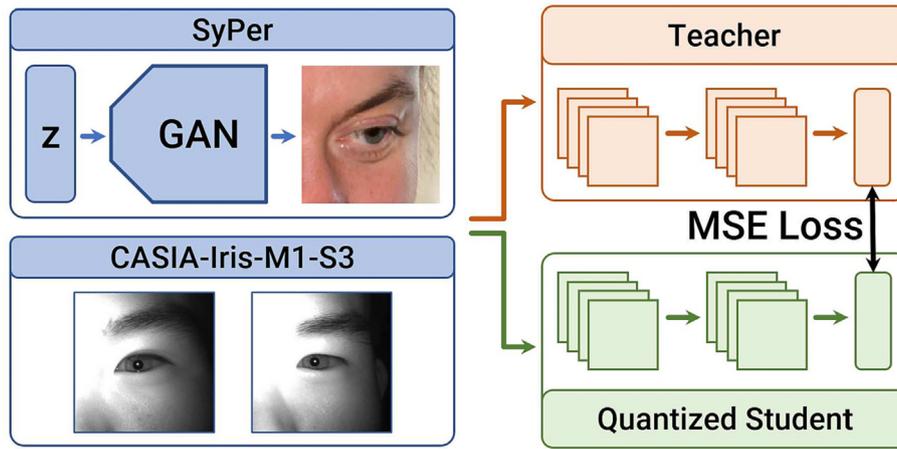


Fig. 3. The quantization is applied through embedding-level knowledge transfer. A dataset is used for quantization aware training, in this case the SyPer dataset with synthetic periocular images created in this work, or with the CASIA-Iris-M1-S3 dataset [79] to achieve better and more efficient NIR recognition performance. Both the teacher and the student model receive an image as input. The student is optimized to compute the same embedding vector as the teacher.



(a) Authentic periocular images cropped from the Flickr-Faces-HQ dataset.



(b) Samples of the SyPer data generated with our trained StyleGan2-ADA trained on periocular-cropped FFHQ dataset.

Fig. 4. Sample images from the cropped FFHQ dataset and synthesized images from the custom trained StyleGan2-ADA.

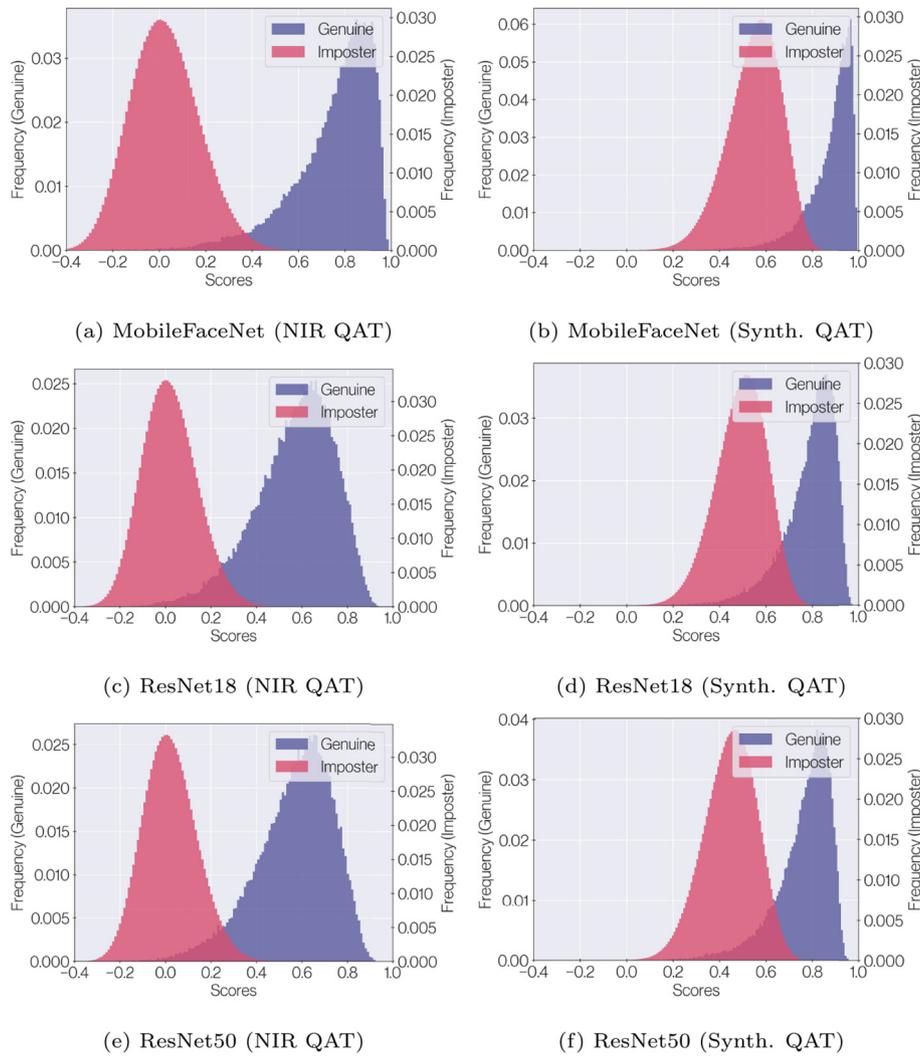


Fig. 5. Genuine and imposter score distributions calculated on the CASIA-Iris-M1-S3 test set. Distributions are generated by the respective models that were quantized with data from either CASIA-Iris-M1-S3 training set (NIR QAT) or with SyPer dataset (Synth. QAT).

Table 1

The results for all three verification backbones with their respective quantization level of 8 (W8A8), 6 (W6A6), and 4 (W4A4) bit. The QAT was performed on the UFPR [80] training set for open world/closed validation in the verification scenario, labeled "Auth.", as well as on the SyPer dataset presented in this paper, labeled "Synth.". All models were tested on the open world/closed validation test protocol of the UFPR dataset.

Model	Params	Bits	Size (MB)	QAT	Verification (1:1)	
					AUC (%)	EER (%)
ResNet18	62.560 M	FP32	250.24	-	98.51 ± 0.15	5.76 ± 0.38
		W8A8	62.56	Auth.	98.41 ± 0.16	5.99 ± 0.39
				Synth.	97.55 ± 0.24	7.64 ± 0.45
		W6A6	46.92	Auth.	98.35 ± 0.15	6.17 ± 0.40
				Synth.	96.72 ± 0.34	9.13 ± 0.61
		W4A4	31.28	Auth.	75.69 ± 3.17	29.18 ± 2.70
ResNet50	82.125 M	FP32	328.50	-	98.47 ± 0.17	5.88 ± 0.38
		W8A8	82.13	Auth.	98.39 ± 0.18	5.99 ± 0.41
				Synth.	97.47 ± 0.25	7.68 ± 0.51
		W6A6	61.59	Auth.	98.29 ± 0.18	6.28 ± 0.41
				Synth.	95.46 ± 0.51	11.18 ± 0.74
		W4A4	41.06	Auth.	66.02 ± 1.07	37.78 ± 0.94
MobileFaceNet	1.276 M	FP32	5.10	-	99.23 ± 0.05	3.86 ± 0.21
		W8A8	1.28	Auth.	99.11 ± 0.05	4.22 ± 0.09
				Synth.	98.51 ± 0.17	5.49 ± 0.45
		W6A6	0.96	Auth.	99.18 ± 0.06	4.02 ± 0.19
				Synth.	97.71 ± 0.18	7.20 ± 0.34
		W4A4	0.64	Auth.	55.01 ± 3.24	50.49 ± 8.39
		Synth.	61.52 ± 5.96	41.67 ± 4.45		

Table 2

Comparison of the verification ResNet18, ResNet50 and MobileFaceNet models quantized to 8 (W8A8), and 6 (W6A6) bit and a set of models reported in [80]. The training of our verification models was carried out on the synthetic dataset SyPer created in this work. All models are tested on the open world/closed validation test set of the UFPR dataset.

Model	Params	Size (MB)	Verification (1:1)	
			AUC (%)	EER (%)
VGG16 [80,83]	135.89 M	1088	97.38 ± 0.53	8.52 ± 0.92
VGG16-Face [80,84]	135.89 M	1088	97.70 ± 0.42	7.78 ± 0.75
InceptionResNet [80,85]	55.25 M	445	99.10 ± 0.24	4.61 ± 0.65
ResNet50V2 [80,86]	49.79 M	400	98.73 ± 0.28	5.69 ± 0.64
ResNet50 [80,82]	24.61 M	198	98.60 ± 0.28	5.98 ± 0.67
ResNet50-Face [14,80]	24.61 M	198	99.18 ± 0.16	4.38 ± 0.47
Xception [80,87]	21.91 M	176	98.93 ± 0.16	5.23 ± 0.42
DenseNet121 [80,88]	7.79 M	64	99.51 ± 0.12	3.39 ± 0.46
Multi-task [80]	4.49 M	37	99.67 ± 0.08	2.81 ± 0.39
MobileNetV2 [80,89]	3.13 M	26	99.56 ± 0.08	3.17 ± 0.33
Siamese [80]	2.55 M	21	97.27 ± 0.64	8.10 ± 1.01
Pairwise [80]	2.35 M	20	98.62 ± 0.72	5.77 ± 1.57
ResNet18 W8A8	62.56 M	62.56	97.55 ± 0.24	7.64 ± 0.45
ResNet18 W6A6	62.56 M	46.92	96.72 ± 0.34	9.13 ± 0.61
ResNet50 W8A8	82.13 M	82.13	97.47 ± 0.25	7.68 ± 0.51
ResNet50 W6A6	82.13 M	61.59	95.46 ± 0.51	11.18 ± 0.74
MobileFaceNet W8A8	1.28 M	1.28	98.51 ± 0.17	5.49 ± 0.45
MobileFaceNet W6A6	1.28 M	0.96	97.71 ± 0.18	7.20 ± 0.34

models are tested on the test split of the dataset. To investigate the impact of the QAT on NIR data, all models quantized using the second database SyPer are also tested on the testset of the NIR database. These models were not quantized using NIR data. As the UFPR protocols are utilizing three folds, the results of the NIR test set is the average over all three folds. Models quantized on NIR data of CASIA-Iris-M1-S3 are labeled as “NIR / Auth.”, as the data used consists of authentic NIR images.

The FP32 models trained on authentic UFPR-Periocular images serve as the teacher, while the quantized models are the students. A graphical overview of the proposed knowledge transfer quantization procedure on embedding level is shown in Fig. 3. In total, each quantized model is trained for 10 epochs, with the first 5 epochs serving as the calibration phase, where the upper (β) and lower (α) bounds are set. After the 5 epochs, these values are fixed and the parameters are adjusted to this mapping.

Table 3

The results for all three identification backbones with their respective quantization level of 8 (W8A8), 6 (W6A6), and 4 (W4A4) bit. The QAT was performed on the UFPR [80] training set for closed world in the identification scenario, labeled “Auth.”, as well as on the SyPer dataset presented in this paper, labeled “Synth.”. All models were tested on the closed world test protocol of the UFPR dataset.

Model	Params	Bits	Size (MB)	QAT	Identification (1:N)		Verification (1:1)	
					Rank 1 (%)	Rank 5 (%)	AUC (%)	EER (%)
ResNet18	62.560 M	FP32	250.24	–	99.61 ± 0.08	99.88 ± 0.03	99.77 ± 0.01	1.75 ± 0.05
		W8A8	62.56	Auth.	99.54 ± 0.11	99.85 ± 0.05	99.76 ± 0.01	1.75 ± 0.02
		W6A6	46.92	Synth.	99.37 ± 0.08	99.77 ± 0.05	99.51 ± 0.01	2.54 ± 0.07
				Auth.	99.48 ± 0.11	99.79 ± 0.05	99.73 ± 0.01	1.92 ± 0.04
		W4A4	31.28	Synth.	98.39 ± 0.18	99.42 ± 0.13	99.25 ± 0.01	3.44 ± 0.06
				Auth.	50.96 ± 12.54	62.01 ± 11.25	86.18 ± 3.96	19.17 ± 4.12
ResNet50	82.125 M	FP32	328.50	–	21.53 ± 4.22	41.77 ± 5.44	88.95 ± 1.17	18.75 ± 1.39
		W8A8	82.13	Auth.	99.54 ± 0.04	99.81 ± 0.04	99.76 ± 0.02	1.76 ± 0.04
		W6A6	61.59	Synth.	99.47 ± 0.02	99.82 ± 0.04	99.75 ± 0.01	1.74 ± 0.02
				Auth.	99.42 ± 0.03	99.72 ± 0.01	99.51 ± 0.02	2.51 ± 0.06
		W4A4	41.06	Synth.	99.34 ± 0.07	99.71 ± 0.08	99.73 ± 0.01	1.95 ± 0.09
				Auth.	96.60 ± 0.51	98.65 ± 0.16	98.95 ± 0.08	4.33 ± 0.23
MobileFaceNet	1.276 M	FP32	5.10	–	33.24 ± 3.82	46.29 ± 4.09	80.55 ± 2.03	26.75 ± 1.98
		W8A8	1.28	Auth.	1.92 ± 0.24	6.41 ± 1.28	75.22 ± 1.73	31.46 ± 1.35
		W6A6	0.96	Synth.	99.87 ± 0.06	99.92 ± 0.03	99.86 ± 0.01	1.26 ± 0.09
				Auth.	99.79 ± 0.05	99.91 ± 0.03	99.86 ± 0.01	1.18 ± 0.08
		W4A4	0.64	Synth.	99.77 ± 0.03	99.88 ± 0.04	99.66 ± 0.03	1.91 ± 0.15
				Auth.	99.80 ± 0.06	99.91 ± 0.03	99.86 ± 0.01	1.22 ± 0.13
Synth.	99.53 ± 0.07	99.79 ± 0.07	99.53 ± 0.03	2.45 ± 0.14				
Auth.	5.53 ± 1.13	11.19 ± 2.46	62.28 ± 3.30	42.50 ± 3.99				
Synth.	7.01 ± 3.29	14.67 ± 6.72	75.72 ± 7.04	30.63 ± 6.01				

4.6. Evaluation metrics

For all experiments, we report the Equal Error Rate (EER), which is defined as the false match rate (FMR) or the false non-match rate (FNMR) at the operation point where they are equal. This metric is based on the ISO/IEC 19795–1 standard [94]. The receiver operating characteristics (ROC) curve plots the $1 - \text{FNMR}$ at a given FMR. To quantify the ROC curve, we report area under curve characteristic (AUC), which is calculated from the ROC curve and specifies the area under it. An area of 1.0 is the optimum and means that the system can perfectly distinguish between the same identity, genuine, and different identities, imposter, in comparisons. For the UFPR dataset two evaluation scenarios are defined by the database authors [80], the verification protocol and the identification protocol explained in Section 4.1. For the verification protocol, we report the EER and AUC metrics calculated from the given 1 : 1 comparisons of the test set. No ranking metrics are reported as no 1 : N database is given. In the identification scenario the respective identification metrics, rank 1 and rank 5, are used following the protocol defined in [80]. Each sample of the test set is evaluated in 1 : N comparison against the given database of identities. A ranking of the N comparison scores is made with the most similar sample being ranked first and the most dissimilar sample being ranked last. Rank 1 specifies the percentage of cases in which the first place contains the identity that was queried. Rank 5 specifies in how many percent of the cases the identity was ranked in the first five places. In this work, the metric used is cosine similarity (as recommended in [15]), where the angle between two embedding vectors is calculated. In the identification protocol, we also apply 1 : 1 verification comparisons between all the samples in the test set, reporting the EER and AUC metric of this setup. For the NIR CASIA-Iris-M1-S3 [79] evaluation, we also follow the reported metrics and report the EER and AUC, all in percentage.

5. Results

In this section, we summarize and discuss the results of our experiments. First, we review the results for quantization on visible image data and show the differences between quantized models in comparison to full precision ones along with comparing quantization based on authentic and synthetic data. Next, the evaluation and quantization on NIR data are discussed.

Table 4

Comparison of the identification ResNet18, ResNet50 and MobileFaceNet models quantized to 8 (W8A8), and 6 (W6A6) bit and a set of models reported in [80]. The training of our identification models was carried out on the synthetic dataset SyPer created in this work. All models are tested on the closed world identification test set of the UFPD dataset.

Model	Params	Size (MB)	Identification (1:N)		Verification (1:1)	
			Rank 1 (%)	Rank 5 (%)	AUC (%)	EER (%)
VGG16 [80,83]	135.89 M	1088	50.56 ± 3.30	68.73 ± 3.01	99.41 ± 0.11	3.59 ± 0.32
VGG16-Face [80,84]	135.89 M	1088	56.29 ± 1.62	73.84 ± 1.48	99.43 ± 0.08	3.44 ± 0.28
InceptionResNet [80,85]	55.25 M	445	65.16 ± 2.45	81.53 ± 1.99	99.78 ± 0.15	1.85 ± 0.40
ResNet50V2 [80,86]	49.79 M	400	63.18 ± 2.14	77.79 ± 1.81	99.74 ± 0.04	2.24 ± 0.18
ResNet50 [80,82]	24.61 M	198	71.06 ± 1.14	85.22 ± 0.82	99.89 ± 0.02	1.41 ± 0.10
ResNet50-Face [14,80]	24.61 M	198	73.76 ± 1.43	86.86 ± 1.02	99.83 ± 0.03	1.74 ± 0.12
Xception [80,87]	21.91 M	176	57.43 ± 1.43	75.88 ± 1.52	99.77 ± 0.04	2.19 ± 0.18
DenseNet121 [80,88]	7.79 M	64	75.54 ± 1.36	88.53 ± 0.97	99.93 ± 0.02	1.11 ± 0.09
Multi-task [80]	4.49 M	37	84.32 ± 0.71	94.55 ± 0.58	99.96 ± 0.01	0.81 ± 0.06
MobileNetV2 [80,89]	3.13 M	26	77.98 ± 1.08	90.19 ± 0.79	99.93 ± 0.01	1.13 ± 0.07
Siamese [80]	2.55 M	21	–	–	98.94 ± 0.22	4.86 ± 0.44
Pairwise [80]	2.35 M	20	–	–	99.44 ± 0.66	3.06 ± 1.84
ResNet18 W8A8	62.56 M	62.56	99.37 ± 0.08	99.77 ± 0.05	99.51 ± 0.01	2.54 ± 0.07
ResNet18 W6A6	62.56 M	46.92	98.39 ± 0.18	99.42 ± 0.13	99.25 ± 0.01	3.44 ± 0.06
ResNet50 W8A8	82.13 M	82.13	99.42 ± 0.03	99.72 ± 0.01	99.51 ± 0.02	2.51 ± 0.06
ResNet50 W6A6	82.13 M	61.59	96.60 ± 0.51	98.65 ± 0.16	98.95 ± 0.08	4.33 ± 0.23
MobileFaceNet W8A8	1.28 M	1.28	99.77 ± 0.03	99.88 ± 0.04	99.66 ± 0.03	1.91 ± 0.15
MobileFaceNet W6A6	1.28 M	0.96	99.53 ± 0.07	99.79 ± 0.07	99.53 ± 0.03	2.45 ± 0.14

5.1. Quantization and visible periocular recognition

The results of the FP32 models, the quantization on 8, 6, and 4 bit with respectively authentic and synthetic data for the QAT of the visible identification task are shown in Table 3. For each of the models used, the number of parameters and the size of the model in MB are given. For mobile devices, it is particularly relevant that there is an acceptable trade-off between recognition performance and model size. The results for the verification task on visible images are shown in the same format in Table 1. Both tables show that the recognition rates only slightly drop in almost all cases when quantizing without class labels on the synthetic dataset SyPer instead of authentic data, however, comparable performance is still maintained.

At 8 bits, the EER for ResNet50 in the verification task increases from 5.99% with authentic data to 7.68% when synthetic data are used for quantization. Similar increases of 1 to 1.5 percentage points occur with both ResNet18 and MobileFaceNet. If the models are quantized

to 6 bits, the EER for the synthetic quantized models in the verification protocol increases by about 2 to 3 percentage points.

In the identification task, ResNet50 loses about 3 percentage points in the rank 1 metric. While ResNet18 and MobileFaceNet with synthetic quantization have a weaker EER than the quantization stage with authentic data, they hardly lose any performance in the rank 1 and rank 5 metrics. When the models are quantized to 4 bits, performance in verification and identification tasks radically decreases for both models, those quantized on authentic data and those quantized on synthetic data. Due to the reduced bit width, it is harder for the quantized model to encode the identity into a meaningful vector.

A comparison of the results of the SyPer quantized model with those of the established methods reported in [80], shown for verification in Table 2, for identification in Table 4, shows that similarly competitive results can be obtained with the synthetic quantization process when model size and EER are considered together. MobileFaceNet in particular

Table 5

The results for all three verification backbones with their respective quantization level of 8 (W8A8), 6 (W6A6), and 4 (W4A4) bit. The QAT was performed on NIR images of the CASIA-Iris-M1-S3 training set, labeled "NIR / Auth.", as well as on the synthetically created SyPer dataset with visible spectrum images, labeled "VIS / Synth.". All models were tested on test set of CASIA dataset.

Model	Params	Bits	Size (MB)	QAT	Verification (1:1)	
					AUC (%)	EER (%)
ResNet18	62.560 M	FP32	250.24	–	96.54 ± 0.19	8.62 ± 0.40
		W8A8	62.56	NIR / Auth.	98.96 ± 0.08	4.35 ± 0.21
				VIS / Synth.	96.09 ± 0.33	9.28 ± 0.62
		W6A6	46.92	NIR / Auth.	98.54 ± 0.09	5.51 ± 0.22
				VIS / Synth.	95.82 ± 0.32	9.88 ± 0.49
		W4A4	31.28	NIR / Auth.	81.3 ± 1.79	26.33 ± 1.64
ResNet50	82.125 M	FP32	328.50	–	96.66 ± 0.15	8.28 ± 0.25
		W8A8	82.13	NIR / Auth.	99.06 ± 0.10	4.12 ± 0.33
				VIS / Synth.	96.27 ± 0.19	8.87 ± 0.30
		W6A6	61.59	NIR / Auth.	98.1 ± 0.04	6.56 ± 0.14
				VIS / Synth.	95.08 ± 0.25	11.02 ± 0.32
		W4A4	41.06	NIR / Auth.	63.44 ± 0.41	40.36 ± 0.32
MobileFaceNet	1.276 M	FP32	5.10	–	60.42 ± 1.09	42.49 ± 0.80
		W8A8	1.28	NIR / Auth.	99.48 ± 0.05	2.7 ± 0.25
				VIS / Synth.	97.52 ± 0.25	6.63 ± 0.38
		W6A6	0.96	NIR / Auth.	99.16 ± 0.16	3.91 ± 0.42
				VIS / Synth.	96.87 ± 0.53	8.04 ± 0.93
		W4A4	0.64	NIR / Auth.	74.27 ± 5.58	32.01 ± 4.34
		VIS / Synth.	65.14 ± 5.01	39.16 ± 4.07		

Table 6

Comparison of the verification ResNet18, ResNet50 and MobileFaceNet models quantized to 8 (W8A8), and 6 (W6A6) bit and a set of models reported in [79], tested on CASIA-Iris-M1-S3 dataset [79]. Only through NIR-quantization are our models able to achieve similar performance values with regard to EER and model size.

Method	Params	Size (MB)	EER (%)
LBP [48,79]	-	-	29.4
Gabor [79,90]	-	-	17.6
SIFT [79,91]	-	-	7.03
TIFS17-CosineDistance [79,92]	-	-	12.6
AlexNet [79,93]	60 M	232	2.83
VGG-16 [83, 79]	138 M	528	2.56
Maxout CNNs [79]	4 M	16	1.89
ResNet18 W8A8	62.56 M	62.56	4.35 ± 0.21
ResNet18 W6A6	62.56 M	46.92	5.51 ± 0.22
ResNet50 W8A8	82.13 M	82.13	4.12 ± 0.33
ResNet50 W6A6	82.13 M	61.59	6.56 ± 0.14
MobileFaceNet W8A8	1.28 M	1.28	2.70 ± 0.25
MobileFaceNet W6A6	1.28 M	0.96	3.91 ± 0.42

is convincing at the 8 bit quantization level due to its small model size and performance, which is significantly better than models with substantially more parameters and larger memory footprints.

While the data distributions of FFHQ, and subsequently the SyPer (generator trained on FFHQ), and UFPR authentic images might be different, the similar performances of models quantized using the SyPer and the authentic UFPR data indicate that this possible domain bias does not significantly effect the model performance. Furthermore, the model performance on authentic data shows that QAT on synthetic SyPer data retains the ability of identity discrimination on authentic images.

In general, the quantization process (on authentic or synthetic data) did prove to reduce the model size by multiple folds while maintaining most of the accuracy, especially when the quantization is performed to 6 and 8 bits models.

5.2. Quantization and NIR periocular recognition

The pre-trained FP32 models were quantized to three bit levels using the training set of the NIR dataset. The results for the verification-based models are shown in Table 5, and the results for the

Table 7

The results for all three identification backbones with their respective quantization level of 8 (W8A8), 6 (W6A6), and 4 (W4A4) bit. The QAT was performed on NIR images of the CASIA-Iris-M1-S3 training dataset, labeled "NIR / Auth.", as well as on the synthetically created SyPer dataset with visible spectrum images, labeled "VIS / Synth.". All models were tested on test set of CASIA dataset.

Model	Params	Bits	Size (MB)	QAT	Verification (1:1)	
					AUC (%)	EER (%)
ResNet18	62.560 M	FP32	250.24	-	95.96 ± 0.26	9.57 ± 0.46
		W8A8	62.56	NIR / Auth.	98.68 ± 0.08	5.11 ± 0.06
				VIS / Synth.	95.41 ± 0.19	10.43 ± 0.36
		W6A6	46.92	NIR / Auth.	97.96 ± 0.03	6.87 ± 0.08
				VIS / Synth.	94.9 ± 0.23	11.26 ± 0.35
		W4A4	31.28	NIR / Auth.	79.17 ± 2.10	28.11 ± 1.87
ResNet50	82.125 M	FP32	328.50	-	96.0 ± 0.28	9.48 ± 0.44
		W8A8	82.13	NIR / Auth.	98.84 ± 0.05	4.81 ± 0.07
				VIS / Synth.	95.69 ± 0.39	9.94 ± 0.61
		W6A6	61.59	NIR / Auth.	97.49 ± 0.19	7.87 ± 0.29
				VIS / Synth.	94.63 ± 0.39	11.8 ± 0.58
		W4A4	41.06	NIR / Auth.	63.34 ± 1.09	40.43 ± 0.71
MobileFaceNet	1.276 M	FP32	5.10	-	60.26 ± 0.79	42.66 ± 0.56
		W8A8	1.28	NIR / Auth.	96.97 ± 0.41	7.63 ± 0.65
				VIS / Synth.	99.44 ± 0.02	2.94 ± 0.12
		W6A6	0.96	NIR / Auth.	96.35 ± 0.37	8.76 ± 0.63
				VIS / Synth.	99.26 ± 0.02	3.7 ± 0.07
		W4A4	0.64	NIR / Auth.	95.98 ± 0.21	9.57 ± 0.29
		VIS / Synth.	77.43 ± 4.35	29.72 ± 3.64		
				71.01 ± 5.14	34.61 ± 4.07	

identification-based models are listed in Table 7. For both verification and identification models, it is clear that the models quantized on NIR data show very good performance in comparison even when compared to the full precision model. This might be due to the fact that the quantization process can act as a domain adaption to the NIR domain, where the model is modified to expect a data distribution similar to that of the NIR. If the bit width is reduced, the performance for $b = 6$ bit decreases only slightly. It is also the case for these models that 4 bit are not sufficient to obtain a stable recognition result. The models quantized on SyPer and therefore without NIR-quantization perform significantly worse but only slightly worse than the full precision model. Their EER is approximately twice as high as that of the models quantized on NIR data. Comparing the NIR-quantized models with the models reported in [79], verification in Table 6, identification in Table 8, it becomes clear that even this short NIR-quantization is sufficient to obtain comparable results with small compact models. As in the visible spectrum, MobileFaceNet is a solid choice. With only a fraction of the parameters and model size of VGG-16 and AlexNet, it achieves on $b = 8$ bit quantization results that are very close in level.

The results of NIR-quantization clearly show the capabilities of quantizing models on data of other spectrums. To further investigate the recognition performance, the genuine and imposter distributions on the NIR CASIA-Iris-M1-S3 test dataset using models with quantization-aware training either on NIR data or synthetic SyPer data are shown in Fig. 5. The genuine and imposter distributions of the model quantized on authentic NIR images are stronger separated in comparison to the distributions of models quantized on synthetic SyPer images of the visible spectrum. In the NIR quantization approach the distributions utilize a larger value range while maintaining separability. This underlines the possible adjustment of pre-trained models to images of new domains through QAT. Although the FP32 teacher model itself was not trained on NIR data, the QAT of the parameter and quantization mappings between floating and integer space is sufficient to achieve good separation and discrimination. Although the distribution of the NIR images does not correspond to those of the visible spectrum, the network is still able to extract the identity information from the image. This is because the structure of the skin and the eye is also visible in the NIR spectrum and the FP32 model trained on visible data can also extract this information, as shown in the results. The proposed method of NIR-quantization is a potential candidate to boost the

Table 8

Comparison of the identification ResNet18, ResNet50 and MobileFaceNet models quantized to 8 (W8A8), and 6 (W6A6) bit and a set of models reported in [79], tested on CASIA-Iris-M1-S3 dataset [79]. Only through NIR-quantization are our models able to achieve similar performance values with regard to EER and model size.

Method	Params	Size (MB)	EER (%)
LBP [48,79]	–	–	29.4
Gabor [79,90]	–	–	17.6
SIFT [79,91]	–	–	7.03
TIFS17-CosineDistance [79,92]	–	–	12.6
AlexNet [79,93]	60 M	232	2.83
VGG-16 [83, 79]	138 M	528	2.56
Maxout CNNs [79]	4 M	16	1.89
ResNet18 W8A8	62.56 M	62.56	5.11 ± 0.06
ResNet18 W6A6	62.56 M	46.92	6.87 ± 0.08
ResNet50 W8A8	82.13 M	82.13	4.81 ± 0.07
ResNet50 W6A6	82.13 M	61.59	7.87 ± 0.29
MobileFaceNet W8A8	1.28 M	1.28	2.94 ± 0.12
MobileFaceNet W6A6	1.28 M	0.96	3.70 ± 0.07

performance in the NIR domain without the need of identity labels for the NIR domain training samples. This approach can be mapped to any image-level domain gap. Thus, a short NIR-quantization of visible-based models on NIR data can be sufficient to train a suitable NIR-compatible model. This is an advantage, as significantly more images are available in the visible spectrum than in the NIR spectrum and because it does not require identity labels of the NIR images.

6. Conclusion

This work presented a set of contributions towards PR solutions that are lightweight, adaptable to different domains, developed in a privacy-friendly framework, along with the generation of synthetic periocular images. This work is the first to apply generative models to periocular images, beyond the tight ocular region. A large synthetic dataset and generation model - SyPer - was created on which three popular neural architectures were trained by quantization-aware training through knowledge transfer on the embedding level. The evaluation and comparison with state-of-the-art models of the underlying benchmark have shown that SyPer can reduce the model size and computational complexity without largely compromising performance. Additionally, by our NIR-quantization that does not require identity labels, we were able to show that a student model can achieve very good recognition performances in the NIR domain, even if the teacher model was only trained on the visible domain.

Data availability

Datasets, Code and Model Checkpoints are available at Github: <https://github.com/jankolf/SyPer>

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.

Acknowledgement

This research work has been funded by the German Federal Ministry of Education and Research and the Hessian Ministry of Higher Education, Research, Science and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE. The contributions due to Hugo Proença in this work were funded by FCT/MCTES through national funds and co-funded EU funds under the project UIDB/EEA/50008/2020.

References

- [1] A.K. Jain, A. Ross, S. Prabhakar, An introduction to biometric recognition, *IEEE Trans. Circuit. Syst. Video Technol.* 14 (1) (2004) 4–20.
- [2] R. Spolaor, Q. Li, M. Monaro, M. Conti, L. Gamberini, G. Sartori, Biometric authentication methods on smartphones: a survey, *PsychNology J.* 14 (2) (2016).
- [3] R. Sharma, A. Ross, Periocular biometrics and its relevance to partially masked faces: a survey, *Comput. Vis. Image Underst.* 226 (2023), 103583, <https://doi.org/10.1016/j.cviu.2022.103583>.
- [4] N. Damer, F. Boutros, M. Süßmilch, F. Kirchbuchner, A. Kuijper, Extended evaluation of the effect of real and simulated masks on face recognition performance, *IET Biom.* 10 (5) (2021) 548–561, <https://doi.org/10.1049/bme2.12044>.
- [5] M. Fang, N. Damer, F. Kirchbuchner, A. Kuijper, Real masks and spoof faces: on the masked face presentation attack detection, *Pattern Recogn.* 123 (2022), 108398, <https://doi.org/10.1016/j.patcog.2021.108398>.
- [6] F. Boutros, N. Damer, J.N. Kolf, K.B. Raja, F. Kirchbuchner, R. Ramachandra, A. Kuijper, P. Fang, C. Zhang, F. Wang, D. Montero, N. Aginako, B. Sierra, M. Nieto, M.E. Erakın, U. Demir, H.K. Ekenel, A. Kataoka, K. Ichikawa, S. Kubo, J. Zhang, M. He, D. Han, S. Shan, K. Grm, V. Struc, S. Seneviratne, N. Kasthuriarachchi, S. Rasnayaka, P.C. Neto, A.F. Sequeira, J.R. Pinto, M. Saffari, J.S. Cardoso, MFR 2021: Masked face recognition competition, *International IEEE Joint Conference on Biometrics, IJCB 2021, Shenzhen, China, August 4–7, 2021, IEEE 2021*, pp. 1–10, <https://doi.org/10.1109/IJCB52358.2021.9484337>.
- [7] M. Fang, F. Boutros, A. Kuijper, N. Damer, Partial attack supervision and regional weighted inference for masked face presentation attack detection, *16th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2021, Jodhpur, India, December 15–18, 2021, IEEE 2021*, pp. 1–8, <https://doi.org/10.1109/FG52635.2021.9667051>.
- [8] F. Boutros, N. Damer, K.B. Raja, R. Ramachandra, F. Kirchbuchner, A. Kuijper, Iris and periocular biometrics for head mounted displays: segmentation, recognition, and synthetic data generation, *Image Vis. Comput.* 104 (2020), 104007, <https://doi.org/10.1016/j.imavis.2020.104007>.
- [9] F. Boutros, N. Damer, K.B. Raja, R. Ramachandra, F. Kirchbuchner, A. Kuijper, Periocular biometrics in head-mounted displays: a sample selection approach for better recognition, *8th International Workshop on Biometrics and Forensics, IWBF 2020, Porto, Portugal, April 29–30, 2020, IEEE 2020*, pp. 1–6, <https://doi.org/10.1109/IWBF49977.2020.9107939>.
- [10] F. Alonso-Fernandez, J. Bigün, J. Fierrez, N. Damer, H. Proença, A. Ross, Periocular biometrics: a modality for unconstrained scenarios, *CoRR* (2022), <https://doi.org/10.48550/arXiv.2212.13792> abs/2212.13792. arXiv:2212.13792.
- [11] P. Kumari, K. Seeja, Periocular biometrics: a survey, *J. King Saud Univ. Comp. Inform. Sci.* 34 (2019) 1086–1097.
- [12] M. Rezende, Evaluation of face recognition technologies for access authentication in automotive passive entry systems with near infrared camera, *Projetos e Dissertações em Sistemas de Informação e Gestão do Conhecimento*, 4, 2015.
- [13] F. Boutros, N. Damer, F. Kirchbuchner, A. Kuijper, Elasticface: Elastic margin loss for deep face recognition, *IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2022, Computer Vision Foundation / IEEE, 2022*.
- [14] Q. Cao, L. Shen, W. Xie, O.M. Parkhi, A. Zisserman, Vggface2: A dataset for recognising faces across pose and age, *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, IEEE 2018, pp. 67–74.
- [15] J. Deng, J. Guo, N. Xue, S. Zafeiriou, Arcface: Additive angular margin loss for deep face recognition, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2019*, pp. 4690–4699.
- [16] Y. Cai, Z. Yao, Z. Dong, A. Gholami, M.W. Mahoney, K. Keutzer, Zeroq: A novel zero shot quantization framework, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2020*, pp. 13169–13178.
- [17] J. Deng, J. Guo, D. Zhang, Y. Deng, X. Lu, S. Shi, Lightweight face recognition challenge, *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019*.
- [18] Y. Martinez-Daz, H. Méndez-Vázquez, L.S. Luevano, L. Chang, M. Gonzalez-Mendoza, Lightweight low-resolution face recognition for surveillance applications, *2020 25th International Conference on Pattern Recognition (ICPR)*, IEEE 2021, pp. 5421–5428.
- [19] Y. Martinez-Diaz, M. Nicolas-Diaz, H. Mendez-Vazquez, L.S. Luevano, L. Chang, M. Gonzalez-Mendoza, L.E. Sucar, Benchmarking lightweight face architectures on specific face recognition scenarios, *Artif. Intell. Rev.* 54 (8) (2021) 6201–6244.
- [20] 2019 IEEE/CVF International Conference on Computer Vision Workshops - The 2019 OpenEDS Workshop: Eye Tracking for VR and AR, *ICCV Workshops 2019, Seoul, Korea (South), October 27–28, IEEE, 2019*, URL <https://research.facebook.com/the-2019-openeds-workshop-eye-tracking-for-vr-and-ar/> URL.
- [21] F. Boutros, P. Siebke, M. Klemm, N. Damer, F. Kirchbuchner, A. Kuijper, Pocketnet: extreme lightweight face recognition network using neural architecture search and multi-step knowledge distillation, *IEEE Access* 10 (2022) 46823–46833, <https://doi.org/10.1109/ACCESS.2022.3170561>.
- [22] H.-C. Li, Z.-Y. Deng, H.-H. Chiang, Lightweight and resource-constrained learning network for face recognition with performance optimization, *Sensors* 20 (21) (2020) 6114.
- [23] X. Li, F. Wang, Q. Hu, C. Leng, Airface: Lightweight and efficient model for face recognition, *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019*.
- [24] P. Zhang, F. Zhao, P. Liu, M. Li, Efficient lightweight attention network for face recognition, *IEEE Access* 10 (2022) 31740–31750.
- [25] F. Boutros, N. Damer, M. Fang, F. Kirchbuchner, A. Kuijper, Mixfacenets: Extremely efficient face recognition networks, *2021 IEEE International Joint Conference on Biometrics (IJCB)*, IEEE 2021, pp. 1–8.

- [26] Y. Martinez-Diaz, L.S. Luevano, H. Mendez-Vazquez, M. Nicolas-Diaz, L. Chang, M. Gonzalez-Mendoza, Shufflefacenet: A lightweight face architecture for efficient and highly-accurate face recognition, Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019.
- [27] S. Chen, Y. Liu, X. Gao, Z. Han, Mobilefacenets: Efficient cnns for accurate real-time face verification on mobile devices, Chinese Conference on Biometric Recognition, Springer 2018, pp. 428–438.
- [28] F. Boutros, N. Damer, M. Fang, K. Raja, F. Kirchbuchner, A. Kuijper, Compact models for periocular verification through knowledge distillation, 2020 International Conference of the Biometrics Special Interest Group (BIOSIG), IEEE 2020, pp. 1–5.
- [29] J. Liu, H. Qin, Y. Wu, J. Guo, D. Liang, K. Xu, Coupleface: Relation matters for face recognition distillation, in: S. Avidan, G.J. Brostow, M. Cissé, G.M. Farinella, T. Hassner (Eds.), Computer Vision – ECCV 2022 – 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XII, Vol. 13672 of Lecture Notes in Computer Science, Springer 2022, pp. 683–700, https://doi.org/10.1007/978-3-031-19775-8_40.
- [30] F. Boutros, N. Damer, A. Kuijper, Quantface: Towards lightweight face recognition by synthetic data low-bit quantization, 26th International Conference on Pattern Recognition, ICPR 2022, Montreal, Quebec, August 21–25, 2021, IEEE, 2022.
- [31] R. Dong, Z. Tan, M. Wu, L. Zhang, K. Ma, Finding the task-optimal low-bit sub-distribution in deep neural networks, ICML, Vol. 162 of Proceedings of Machine Learning Research, PMLR 2022, pp. 5343–5359.
- [32] W. Fei, W. Dai, C. Li, J. Zou, H. Xiong, General bitwidth assignment for efficient deep convolutional neural network quantization, IEEE Trans. Neural Netw. Learn. Syst. 33 (2021) 5253–5267.
- [33] Y. Ma, T. Jin, X. Zheng, Y. Wang, H. Li, G. Jiang, W. Zhang, R. Ji, Ompq: Orthogonal mixed precision quantization, arXiv (2021) 1–13 preprint arXiv:2109.07865.
- [34] J. Wu, C. Leng, Y. Wang, Q. Hu, J. Cheng, Quantized convolutional neural networks for mobile devices, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016, pp. 4820–4828.
- [35] R. Krishnamoorthi, Quantizing deep convolutional networks for efficient inference: a whitepaper, arXiv (2018) 1–36 preprint arXiv:1806.08342.
- [36] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al., Pytorch: an imperative style, high-performance deep learning library, Adv. Neural Inf. Proces. Syst. 32 (2019).
- [37] R. Banner, Y. Nahshan, D. Soudry, Post training 4-bit quantization of convolutional networks for rapid-deployment, Adv. Neural Inf. Proces. Syst. 32 (2019).
- [38] J. Gong, H. Shen, G. Zhang, X. Liu, S. Li, G. Jin, N. Maheshwari, E. Fomenko, Highly efficient 8-bit low precision inference of convolutional neural networks with intelcffe, Proceedings of the 1st on Reproducible Quality-Efficient Systems Tournament on Co-designing Pareto-efficient Deep Learning 2018, p. 1.
- [39] B. Jacob, S. Kligys, B. Chen, M. Zhu, M. Tang, A.G. Howard, H. Adam, D. Kalenichenko, Quantization and training of neural networks for efficient integer-arithmetic-only inference, 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18–22, 2018, Computer Vision Foundation / IEEE Computer Society 2018, pp. 2704–2713, <https://doi.org/10.1109/CVPR.2018.00286>, URL http://openaccess.thecvf.com/content_cvpr_2018/html/Jacob_Quantization_and_Training_CVPR_2018_paper.html, URL.
- [40] Q. Jin, J. Ren, R. Zhuang, S. Hanumante, Z. Li, Z. Chen, Y. Wang, K. Yang, S. Tulyakov, F8net: Fixed-point 8-bit only multiplication for network quantization, in: International Conference on Learning Representations, 2022, URL <https://openreview.net/forum?id=CfpJazzXT2>.
- [41] B. Zhuang, C. Shen, M. Tan, L. Liu, I. Reid, Towards effective low-bitwidth convolutional neural networks, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2018, pp. 7920–7928.
- [42] J.N. Kolf, F. Boutros, F. Kirchbuchner, N. Damer, Lightweight periocular recognition through low-bit quantization, IEEE International Joint Conference on Biometrics, IJCB 2022, Abu Dhabi, United Arab Emirates, October 10–13, 2022, IEEE 2022, pp. 1–12, <https://doi.org/10.1109/IJCB54206.2022.10007980>.
- [43] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, T. Aila, Training generative adversarial networks with limited data, in: H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, H. Lin (Eds.), Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6–12, 2020, virtual, 2020.
- [44] K. Ahuja, A. Bose, S. Nagar, K. Dey, F. Barbhuiya, Isure: User authentication in mobile devices using ocular biometrics in visible spectrum, 2016 IEEE International Conference on Image Processing (ICIP), IEEE 2016, pp. 335–339.
- [45] F. Alonso-Fernandez, J. Bigun, Best regions for periocular recognition with nir and visible images, 2014 IEEE International Conference on Image Processing (ICIP), IEEE 2014, pp. 4987–4991.
- [46] A. Ross, R. Jillela, J.M. Smereka, V.N. Boddeti, B.V. Kumar, R. Barnard, X. Hu, P. Pauca, R. Plimmons, Matching highly non-ideal ocular images: An information fusion approach, 2012 5th IAPR International Conference on Biometrics (ICB), IEEE 2012, pp. 446–453.
- [47] K.B. Raja, R. Raghavendra, M. Stokkenes, C. Busch, Smartphone authentication system using periocular biometrics, 2014 International Conference of the Biometrics Special Interest Group (BIOSIG), IEEE 2014, pp. 1–8.
- [48] U. Park, A. Ross, A.K. Jain, Periocular biometrics in the visible spectrum: A feasibility study, 2009 IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems, IEEE 2009, pp. 1–6.
- [49] C.-W. Tan, A. Kumar, Human identification from at-a-distance images by simultaneously exploiting iris and periocular features, Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), IEEE 2012, pp. 553–556.
- [50] F. Alonso-Fernandez, A. Mikaelyan, J. Bigun, Compact multi-scale periocular recognition using SAFE features, 23rd International Conference on Pattern Recognition, ICPR 2016, Cancun, Mexico, December 4–8, 2016, IEEE 2016, pp. 1455–1460, <https://doi.org/10.1109/ICPR.2016.7899842>.
- [51] L. Nie, A. Kumar, S. Zhan, Periocular recognition using unsupervised convolutional rbm feature learning, 2014 22nd International Conference on Pattern Recognition, IEEE 2014, pp. 399–404.
- [52] K.B. Raja, R. Raghavendra, C. Busch, Collaborative representation of deep sparse filtered features for robust verification of smartphone periocular images, 2016 IEEE International Conference on Image Processing (ICIP), IEEE 2016, pp. 330–334.
- [53] Q. Zhang, H. Li, Z. Sun, T. Tan, Deep feature fusion for iris and periocular biometrics on mobile devices, IEEE Trans. Inform. Forens. Secur. 13 (11) (2018) 2897–2912.
- [54] N. Reddy, A. Rattani, R. Derakhshani, Robust subject-invariant feature learning for ocular biometrics in visible spectrum, 2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS), IEEE 2019, pp. 1–9.
- [55] L.A. Zanlorensi, H. Proença, D. Menotti, Unconstrained periocular recognition: Using generative deep learning frameworks for attribute normalization, 2020 IEEE International Conference on Image Processing (ICIP), IEEE 2020, pp. 1361–1365.
- [56] F. Alonso-Fernandez, K.B. Raja, R. Raghavendra, C. Busch, J. Bigun, R. Vera-Rodríguez, J. Fierrez, Cross-sensor periocular biometrics in a global pandemic: comparative benchmark and novel multialgorithmic approach, Inform. Fusion 83 (2022) 110–130.
- [57] N. Reddy, A. Rattani, R. Derakhshani, Ocularnet: deep patch-based ocular biometric recognition, 2018 IEEE International Symposium on Technologies for Homeland Security (HST), IEEE 2018, pp. 1–6.
- [58] A. Almadan, A. Rattani, Compact cnn models for on-device ocular-based user recognition in mobile devices, 2021 IEEE Symposium Series on Computational Intelligence (SSCI), IEEE 2021, pp. 1–7.
- [59] F. Boutros, N. Damer, K.B. Raja, F. Kirchbuchner, A. Kuijper, Template-driven knowledge distillation for compact and accurate periocular biometrics deep-learning models, Sensors 22 (5) (2022) 1921, <https://doi.org/10.3390/s22051921>.
- [60] N. Damer, F. Boutros, F. Kirchbuchner, A. Kuijper, D-id-net: Two-stage domain and identity learning for identity-preserving image generation from semantic segmentation, 2019 IEEE/CVF International Conference on Computer Vision Workshops, ICCV Workshops 2019, Seoul, Korea (South), October 27–28, 2019, IEEE 2019, pp. 3677–3682, <https://doi.org/10.1109/ICCVW.2019.00454>.
- [61] D. Tomasevic, P. Peer, V. Struc, Bioculargan: bimodal synthesis and annotation of ocular images, IEEE International Joint Conference on Biometrics, IJCB 2022, Abu Dhabi, United Arab Emirates, October 10–13, 2022, IEEE 2022, pp. 1–10, <https://doi.org/10.1109/IJCB54206.2022.10007982>.
- [62] H. Qiu, B. Yu, D. Gong, Z. Li, W. Liu, D. Tao, Synface: Face recognition with synthetic data, Proceedings of the IEEE/CVF International Conference on Computer Vision 2021, pp. 10880–10890.
- [63] A. Kortylewski, A. Schneider, T. Gerig, B. Egger, A. Morel-Förster, T. Vetter, Training deep face recognition systems with synthetic data, CoRR (2018) 1–8, URL <http://arxiv.org/abs/1802.05891>, arXiv:1802.05891.
- [64] A. Kortylewski, B. Egger, A. Schneider, T. Gerig, A. Morel-Förster, T. Vetter, Analyzing and reducing the damage of dataset bias to face recognition with synthetic data, IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2019, Long Beach, CA, USA, June 16–20, 2019, Computer Vision Foundation / IEEE 2019, pp. 2261–2268, <https://doi.org/10.1109/CVPRW.2019.00279>, URL http://openaccess.thecvf.com/content_CVPRW_2019/html/BEFA/Kortylewski_Analyzing_and_Reducing_the_Damage_of_Dataset_Bias_to_Face_CVPRW_2019_paper.html.
- [65] F. Boutros, M. Huber, P. Siebke, T. Rieber, N. Damer, Sface: Privacy-friendly and accurate face recognition using synthetic data, IEEE International Joint Conference on Biometrics, IJCB 2022, Abu Dhabi, United Arab Emirates, October 10–13, 2022, IEEE 2022, pp. 1–11, <https://doi.org/10.1109/IJCB54206.2022.10007961>.
- [66] F. Boutros, M. Klemm, M. Fang, A. Kuijper, N. Damer, Unsupervised Face Recognition Using Unlabeled Synthetic Data, FG, IEEE, 2023 1–8.
- [67] Y. Guo, A survey on methods and theories of quantized neural networks, arXiv (2018) 1–17 preprint arXiv:1808.04752.
- [68] D. Zuras, M. Cowlshaw, A. Aiken, M. Applegate, D. Bailey, S. Bass, D. Bhandarkar, M. Bhat, D. Bindel, S. Boldo, et al., IEEE standard for floating-point arithmetic, IEEE Std. 754 (2008) (2008) 1–70.
- [69] S. Van Der Walt, S.C. Colbert, G. Varoquaux, The numpy array: a structure for efficient numerical computation, Comp. Sci. Eng. 13 (2) (2011) 22–30.
- [70] A. Gholami, S. Kim, Z. Dong, Z. Yao, M.W. Mahoney, K. Keutzer, A survey of quantization methods for efficient neural network inference, arXiv (2021) 1–33 preprint arXiv:2103.13630.
- [71] Q. Huang, D. Wang, Z. Dong, Y. Gao, Y. Cai, T. Li, B. Wu, K. Keutzer, J. Wawrzyniak, Codenet: Efficient deployment of input-adaptive object detection on embedded fpgas, The 2021 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays 2021, pp. 206–216.
- [72] D. Zhang, J. Yang, D. Ye, G. Hua, Lq-nets: Learned quantization for highly accurate and compact deep neural networks, Proceedings of the European Conference on Computer Vision (ECCV) 2018, pp. 365–382.
- [73] Y. Bengio, N. Léonard, A. Courville, Estimating or propagating gradients through stochastic neurons for conditional computation, arXiv (2013) 1–12 preprint arXiv:1308.3432.
- [74] Y. Choi, J.P. Choi, M. El-Khamy, J. Lee, Data-free network quantization with adversarial knowledge distillation, 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR Workshops 2020, Seattle, WA, USA, June 14–19, 2020, Computer Vision Foundation / IEEE 2020, pp. 3047–3057, <https://doi.org/10.1109/CVPRW50498.2020.00363>, URL https://openaccess.thecvf.com/content_CVPRW_2020/html/w40/Choi_Data-Free_Network_Quantization_With_Adversarial_Knowledge_Distillation_CVPRW_2020_paper.html, URL.

- [75] S. Xu, H. Li, B. Zhuang, J. Liu, J. Cao, C. Liang, M. Tan, Generative low-bitwidth data free quantization, in: A. Vedaldi, H. Bischof, T. Brox, J. Frahm (Eds.), *Computer Vision - ECCV 2020 - 16th European Conference*, Glasgow, UK, August 23–28, 2020, Proceedings, Part XII, Vol. 12357 of Lecture Notes in Computer Science, Springer 2020, pp. 1–17, https://doi.org/10.1007/978-3-030-58610-2_1.
- [76] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A.C. Courville, Y. Bengio, Generative adversarial networks, *Commun. ACM* 63 (11) (2020) 139–144, <https://doi.org/10.1145/3422622>.
- [77] T. Karras, S. Laine, T. Aila, A style-based generator architecture for generative adversarial networks, *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019*, Long Beach, CA, USA, June 16–20, 2019, Computer Vision Foundation / IEEE 2019, pp. 4401–4410, <https://doi.org/10.1109/CVPR.2019.00453>.
- [78] K. Zhang, Z. Zhang, Z. Li, Y. Qiao, Joint face detection and alignment using multitask cascaded convolutional networks, *IEEE Sign. Process. Lett.* 23 (10) (2016) 1499–1503, <https://doi.org/10.1109/LSP.2016.2603342>.
- [79] Q. Zhang, H. Li, Z. Sun, T. Tan, Deep feature fusion for iris and periocular biometrics on mobile devices, *IEEE Trans. Inf. Forens. Secur.* 13 (11) (2018) 2897–2912, <https://doi.org/10.1109/TIFS.2018.2833033>.
- [80] L.A. Zanlorensi, R. Laroca, D.R. Lucio, L.R. Santos, A.S. Britto Jr., D. Menotti, A new periocular dataset collected by mobile devices in unconstrained scenarios, *Sci. Rep.* 12 (1) (2022) 17989.
- [81] M. Uzair, A. Mahmood, A. Mian, C. McDonald, Periocular biometric recognition using image sets, *2013 IEEE Workshop on Applications of Computer Vision (WACV)*, IEEE 2013, pp. 246–251.
- [82] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016*, pp. 770–778.
- [83] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: Y. Bengio, Y. LeCun (Eds.), *3rd International Conference on Learning Representations, ICLR 2015*, San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings, 2015, URL <http://arxiv.org/abs/1409.1556>, URL.
- [84] O.M. Parkhi, A. Vedaldi, A. Zisserman, Deep face recognition, in: X. Xie, M.W. Jones, G.K.L. Tam (Eds.), *Proceedings of the British Machine Vision Conference 2015, BMVC 2015*, Swansea, UK, September 7–10, 2015, BMVA Press 2015, pp. 41.1–41.12, <https://doi.org/10.5244/C.29.41>.
- [85] C. Szegedy, S. Ioffe, V. Vanhoucke, A.A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.
- [86] K. He, X. Zhang, S. Ren, J. Sun, Identity mappings in deep residual networks, *European Conference on Computer Vision*, Springer 2016, pp. 630–645.
- [87] F. Chollet, Xception: Deep learning with depthwise separable convolutions, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017*, pp. 1251–1258.
- [88] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017*, pp. 4700–4708.
- [89] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen, Mobilenetv2: Inverted residuals and linear bottlenecks, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2018*, pp. 4510–4520.
- [90] J. Daugman, High confidence visual recognition of persons by a test of statistical independence, *IEEE Trans. Pattern Anal. Mach. Intell.* 15 (11) (1993) 1148–1161, <https://doi.org/10.1109/34.244676>.
- [91] D.G. Lowe, Distinctive image features from scale-invariant keypoints, *Int. J. Comput. Vis.* 60 (2) (2004) 91–110, <https://doi.org/10.1023/B:VISI.0000029664.99615.94>.
- [92] Z. Zhao, A. Kumar, Accurate periocular recognition under less constrained environment using semantics-assisted convolutional neural network, *IEEE Trans. Inf. Forens. Secur.* 12 (5) (2017) 1017–1030, <https://doi.org/10.1109/TIFS.2016.2636093>.
- [93] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: P.L. Bartlett, F.C.N. Pereira, C.J.C. Burges, L. Bottou, K.Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a meeting held December 3–6, 2012, Lake Tahoe, Nevada, United States 2012*, pp. 1106–1114, URL <https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html>, URL.
- [94] A. Mansfield, Information technology–biometric performance testing and reporting–Part 1: Principles and framework, *ISO/IEC (2006) 19795–1*.