

## Article

# Template-Driven Knowledge Distillation for Compact and Accurate Periocular Biometrics Deep-Learning Models

Fadi Boutros <sup>1,2,\*</sup> , Naser Damer <sup>1,2,\*</sup> , Kiran Raja <sup>3</sup> , Florian Kirchbuchner <sup>1</sup>  and Arjan Kuijper <sup>1,2</sup> 

<sup>1</sup> Fraunhofer Institute for Computer Graphics Research IGD, 64283 Darmstadt, Germany; florian.kirchbuchner@igd.fraunhofer.de (F.K.); arjan.kuijper@igd.fraunhofer.de (A.K.)

<sup>2</sup> Mathematical and Applied Visual Computing, Technical University of Darmstadt (TU Darmstadt), 64289 Darmstadt, Germany

<sup>3</sup> The Norwegian Colour and Visual Computing Laboratory, Norwegian University of Science and Technology (NTNU), 2815 Gjøvik, Norway; kiran.raja@ntnu.no

\* Correspondence: fadi.boutros@igd.fraunhofer.de (F.B.); naser.damer@igd.fraunhofer.de (N.D.); Tel.: +49-16094877953 (F.B.)

**Abstract:** This work addresses the challenge of building an accurate and generalizable periocular recognition model with a small number of learnable parameters. Deeper (larger) models are typically more capable of learning complex information. For this reason, knowledge distillation (kd) was previously proposed to carry this knowledge from a large model (teacher) into a small model (student). Conventional KD optimizes the student output to be similar to the teacher output (commonly classification output). In biometrics, comparison (verification) and storage operations are conducted on biometric templates, extracted from pre-classification layers. In this work, we propose a novel template-driven KD approach that optimizes the distillation process so that the student model learns to produce templates similar to those produced by the teacher model. We demonstrate our approach on intra- and cross-device periocular verification. Our results demonstrate the superiority of our proposed approach over a network trained without KD and networks trained with conventional (vanilla) KD. For example, the targeted small model achieved an equal error rate (EER) value of 22.2% on cross-device verification without KD. The same model achieved an EER of 21.9% with the conventional KD, and only 14.7% EER when using our proposed template-driven KD.

**Keywords:** biometrics; knowledge distillation; periocular verification



**Citation:** Boutros, F.; Damer, N.; Raja, K.; Kirchbuchner, F.; Kuijper, A. Template-Driven Knowledge Distillation for Compact and Accurate Periocular Biometrics Deep-Learning Models. *Sensors* **2022**, *22*, 1921. <https://doi.org/10.3390/s22051921>

Academic Editors: Abdeldjalil Ouahabi, Sébastien Jacques and Amir Benzaoui

Received: 26 January 2022

Accepted: 25 February 2022

Published: 1 March 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Biometric deployment on smartphones and different embedded devices enables novel applications for identity verification or service personalization. Facial characteristics are one of the commonly used biometric traits for identity recognition. This is driven by the high user acceptance [1] and the non-invasive touchless capture process [2–4]. Moreover, the high performance achieved by the recent approaches proposed in the literature [5–7] led to face recognition gaining larger deployment ground. However, face occlusions due to face masks, especially after the recent COVID-19 pandemic, present a challenge for face recognition systems [8]. Recently, several studies [9–12] evaluated the effect of wearing a face mask on face recognition performance. These studies [9–11] concluded that the verification performance of face-recognition solutions significantly degraded when the subject was wearing a face mask, compared to the case where their face was unmasked. This was followed by several efforts to enhance masked face recognition [13–15], including competitions where some of the submitted solutions proposed the use of the periocular recognition [16,17]. Periocular biometrics have a distinct advantage over facial biometrics when the face is largely occluded or when capturing a full face is less convenient than capturing the periocular region (e.g., a selfie on a smartphone or masked face recognition [11]), while maintaining the touchless nature of face capture. It also carries other benefits; for

example, perspective distortion affects the periocular region to a lower degree because the depth variation of the area is smaller than the that of the complete face [18]. Periocular recognition on smartphones and wearable devices has gained growing research attention, and with recent deep learning methods achieving superior accuracy and with applications in various domains have surfaced [19–22]. However, deploying large models based on deep learning on resource-critical consumer devices presents two main challenges. The first relates to the highly constrained computational requirements of these devices, requiring the models to be compact. Given the increased number of smart applications running on these devices, keeping the number of trainable parameters of any deep learning model to a minimum is essential for acceptance by customers against having applications that require gigabytes of memory. The second challenge is related to cross-device/cross-camera biometric performance, which often results in data with high variations. A small deep learning model for periocular recognition should be able to compensate for such variations in capture conditions. Learning complex image information such as identity is more realizable in relatively deeper models [23]. However, such models contain a large number of trainable parameters, as mentioned earlier. The post-training minimization of these complex models has been addressed mainly by parameter pruning, quantization, and knowledge distillation (KD). While the first two change the trained network structure, KD transfers its knowledge into a smaller (student) network. Conventional KD deployments optimize the distillation process so that the student network achieves similar output as the teacher (larger and trained) network.

Many biometric applications resort to deep learning models that are trained as identity classifiers. However, these models are used as template (feature) extractors by sampling pre-classification network layers [24,25]. This renders the recent proposals of deploying conventional KD on biometric networks [26] theoretically sub-optimal, as it teaches the less-complex student network to classify the identity correctly, rather than to extract an optimal (with respect to the teacher network) template.

The present work addresses this issue by proposing its main contribution, a novel template-driven KD for periocular recognition that transfers the optimal template extraction knowledge from a deeper network into a network with a significantly lower number of trainable parameters. The proposed approach is designed so that it can be adapted to any suitable student/teacher architecture combinations. We extensively demonstrate the applicability of our proposed approach for periocular verification, both for intra-device and cross-device scenarios, to show the effectiveness of our proposed template-driven KD. To demonstrate the generalizability of the proposed concept, we train our template-driven KD network with two different loss functions, mean squared error (MSE) loss and Cosine loss. Further, we compare the performance of the resulting student network with the same network when trained from scratch, as well as two different larger teacher networks to validate our assertion. Our results are further benchmarked against the conventional KD optimized for classification, consistently showing the benefits of our proposed solution. These achievements are demonstrated on a periocular dataset of 152 unique identities consisting of 6682 images captured with 2 different smartphones (iPhone 5S and Nokia Lumia 1020) to demonstrate generalizability across capture devices. We demonstrate that our template-driven KD solution successfully and consistently reduced the verification error rates under cross-device experimental settings. Our template-driven KD solution successfully reduced the cross-device false non-match rate (FNMR) at a false match rate (FMR) of 10% of the same efficient network structure from 39.45% when trained from scratch and 23.2% when conventional KD is applied, to 20.1%.

In the remainder of this article, we present a set of related works in Section 2 and the proposed approach along with baseline approaches in Section 3. We provide the details on the conducted experiments in Section 4 and the obtained results in Section 5. Finally, we present our conclusions in Section 6.

## 2. Related Work

Periocular recognition is preferred when capturing the full face consistently is a challenge [8,11,15], for instance when faces are occluded due to masks or clothing preferences. Park et al. [27] presented one of the first works proposing the use of the periocular region as a biometric trait captured under a controlled environment. Based on the approach of Park et al. [27], several subsequent works were proposed: Juefei-Xu et al. [28] utilized LBP to encode discrete transforms enabling a translation-robust descriptor. Mahalingam and Ricanek Jr. [29] proposed the use of multi-scale patch-based LBP feature descriptors. Ross et al. [30] presented a fusion-based schema to handle the variability in periocular input images. Recently, many approaches using deep learning have been proposed for periocular recognition [31–33]. Proenca et al. [31] trained convolutional neural networks (CNNs) to implicitly learn the region of interest (periocular area) without localizing the iris. Zhao and Kumar [33] proposed a semantics-assisted CNN (SCNN) model consisting of multiple CNN models for periocular recognition. Rattani et al. [32] evaluated off-the-shelf deep CNNs for periocular recognition and illustrated the benefit of fine-tuning CNNs.

As noted by Rattani et al. [32], the high variability between probe and gallery images, produced when the images are acquired using different devices, challenges the biometric performance. A generalized solution to such challenging data is needed to mitigate biometric performance loss, and some works have partially addressed this issue in recent years [20,34–37]. While previous deep-learning-based solutions addressed the problem to a certain degree, the computational complexity of the models is rather high, with many trainable parameters [34,37], and this specifically is not preferred in smartphone applications due to limited computation resources. Both models presented in [34,37] contain more than 12 million trainable parameters.

In order to address the constraints of compact model requirements, earlier works have minimized deep-learning-model size using mainly one of three methods, parametric pruning [38], quantization [39], and KD [40]. While the first two methods process a relatively large network into a smaller one, KD transfers the acquired knowledge learned by a larger network (teacher network) to a smaller one (student network). Common KD deployments optimize the distillation (teaching) process on the same network output that the teacher network used to optimize its own training, for example, classification accuracy [41].

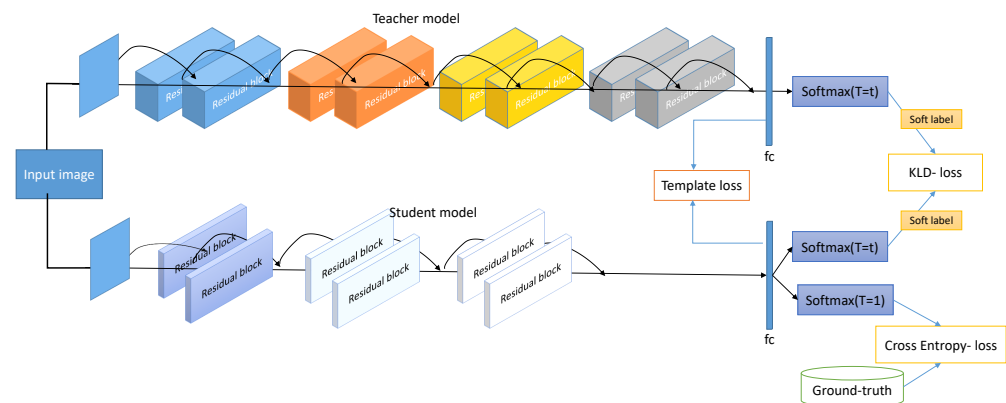
While this seems a feasible solution, biometric applications need to extract identity information (e.g., feature representation or templates) in order to achieve better performance [24,42]. Performing conventional KD, optimized on the classification output, has recently been applied in the biometric domain [26,43]. This makes the distillation of the classification knowledge sub-optimal for a student network that will be deployed as a template extractor, as will be shown later. In this work, although the teacher network is trained to optimally classify identities, we distill the knowledge to achieve optimized template generation by the student network that, by itself, is too shallow to learn templates which are as optimal as those of its teacher.

## 3. Baseline Models and Proposed Approach

In this work we present a template-driven KD framework to enhance the accuracy and generalizability of a compact CNN model for periocular verification under a smartphone verification scenario. Figure 1 illustrates the proposed template-driven KD framework. In this work, we propose to train the student model to learn the logits and the template embedding from the teacher model. We achieved this by introducing an additional loss function to the original KD loss operated on the feature extraction layer, as presented in Figure 1. In this section, we present the baseline model, the common KD method, and the proposed template-driven KD approach.

### 3.1. Baseline Models

We employed the widely used ResNet model [44] as a baseline model to evaluate our approach. The network architecture is based on an identity shortcut connection (residual connection) that skips one or more layers. The skip connection adds the input of a residual block to its output and passes it to the following residual block. ResNet won the ILSVRC 2015 challenge on the ImageNet database [45], with a top-5 error rate of 3.57%. The powerful representation ability of ResNet has motivated several computer vision tasks other than classification, such as object detection and face recognition [5]. In this work, we first evaluated the performance of three ResNet models—ResNet-34, ResNet-18, and ResNet-110 (without KD)—for periocular verification; all models are described in [44]. The first layer of ResNet-18 is a  $7 \times 7$  convolutional layer followed by a  $3 \times 3$  max-pooling layer. Then it stacks four convolutional groups. Each group consists of two residual blocks. Each block consists of two  $3 \times 3$  convolutional layers with 64, 128, 256, and 512 filters for layers in groups 1, 2, 3, and 4, respectively. The last layer of the network is the average pooling layer (produced 512 features vector), followed by the classification layer. ResNet-34 has the same form as ResNet-18, with 3, 4, 6, and 3 residual blocks in groups 1, 2, 3, and 4, respectively. The number of trainable parameters in ResNet-34 and ResNet-18 is 21.3 million (m) and 11.8 m, respectively. Additionally, ResNet proposed the use of a small filter size of (32, 16, 8) with three groups of residual blocks to create a compact version of the ResNet model (ResNet-110) designed for the CIFAR-10 database with fewer trainable parameters. The used ResNet-110 in this work contains 1.8 m trainable parameters.



**Figure 1.** Overview of the proposed template-driven KD approach for periocular verification based on the ResNet architecture. Note that both the template loss and KD output loss contribute to the distillation process.

In this work, we deployed ResNet-34 and ResNet-18 as teacher networks and the compact ResNet-110 as a student, as well as a stand-alone baseline. In order to adopt these models for periocular verification, we modified the number of classes in the classification layer to match the number of identities in our training dataset. We applied transfer learning on the models ResNet-34 and ResNet-18 pretrained on the ImageNet dataset [45] by fine-tuning the last group of residual blocks on images from our training dataset with the softmax classifier. During the test phase, the softmax classifier was removed, and the features  $f$  were extracted from the last layer, with dimension  $512 \times 1 \times 1$ .

The number features in the last layer of the ResNet-110 model is  $64 \times 1 \times 1$ . In order to enable KD on the template level, we introduced an additional fully connected layer (before the classification layer) of size 512 to the ResNet-110 model to match the size of the embedding template of the teacher model. We trained this model from scratch on images from our training dataset. The model was trained first using the softmax classifier for baseline evaluation. During the test phase, the features  $f$  were extracted from the last fully connected layer. Using the exact training setups, we trained and evaluated the ResNet-110

model with the conventional (vanilla) KD approach. Finally, we trained and evaluated the ResNet-110 model with our template-driven KD.

### 3.2. Knowledge Distillation

KD is a technique to improve the performance and generalizability of smaller models by transferring the knowledge learned by a cumbersome model (teacher) to a single small model (student). The key idea is to guide the student model to learn the relationship between different classes discovered by the teacher model that contains more information beyond the ground-truth labels [40]. Suppose we have teacher model  $T$ , student model  $S$ , and training dataset  $X, Y \in D$ , where  $X$  is the training images and  $Y$  is their class labels. The output of the teacher model for any input  $x_i \in X$  is a vector of class probabilities  $P^T$  computed for each class using the softmax function by converting the logits,  $z^T$ , into probabilities that sum to one  $P^T(x) = \text{softmax}(z^T)$ . Specifically, the probability  $p_i$  of class  $i$  is computed by comparing  $z_i$  with other logits as given:

$$p_i = \frac{\exp(z_i)}{\sum_{j=1}^N \exp(z_j)}. \quad (1)$$

This probability distribution will have a high probability value of  $p_i$  for the correct class  $y_i \in Y$ , with all other class probabilities close to zero. Thus, it does not provide more valuable information than ground-truth labels. Therefore, Hinton et al. [40] proposed to scale the logits using a temperature parameter  $t > 1$  before applying the softmax function. Thus, the teacher model can produce a softer distribution of the class probabilities, which provides more valuable information about classes similar to the predicted class. In this case, the output of the teacher model is  $P_s^T(x) = \text{softmax}(z^T/t)$  and the probability  $p_i$  of class  $i$  is given as:

$$p_i = \frac{\exp(z_i/t)}{\sum_{j=1}^N \exp(z_j/t)}. \quad (2)$$

Similarly, the student model  $S$  can produce a soft class probability distribution using the temperature parameter  $t$ ,  $P_s^S(x) = \text{softmax}(z^S/t)$ . The final loss for the student model is a weighted sum of two loss functions, cross-entropy loss  $L_{ce}$  and Kullback–Leibler divergence (KLD) loss  $L_{kld}$ , as follows:

$$L_{KD} = \alpha * L_{ce}(Y, P^S(x)) + \beta * t^2 * L_{kld}(P_s^S(x), P_s^T(x)), \quad (3)$$

where  $Y$  is the ground-truth label,  $P^S(x)$  is the standard softmax output produced by the student,  $P_s^S(x)$  is the parameterized softmax output produced by the student,  $P_s^T(x)$  is the parameterized softmax output produced by teacher, and  $\alpha$  and  $\beta$  are the loss weight hyper-parameters. Since the gradients of the  $L_{ce}$  loss are smaller than the gradients of the  $L_{kld}$  where the logits used for  $L_{kld}$  are divided by  $t$ ,  $L_{kld}$  is multiplied by  $t^2$  as suggested by Hinton et al. [40].

### 3.3. Template-Driven Knowledge Distillation

Recently, a number of approaches have been proposed to improve the conventional vanilla KD method by either introducing an additional loss function to the KD loss [26] or by forcing the student to learn from multilayers instead of learning only from the logits of the last layer [46]. One of the drawbacks of softmax loss used in KD is that it does not explicitly optimize the embedded feature representation (template) needed for biometric verification. The template produced by the student model is less informative than the embedded template produced by the teacher model due to the shallow architecture of the student model. Therefore, in this work we propose to enhance the performance of the student model in the KD framework by driving it to learn the embedding template from the teacher model. In order to achieve this, we introduce an additional loss function to the original KD loss operated on the feature extraction layer. We evaluated the performance

of the distilled model using two different loss functions: MSE and cosine embedding loss. Formally, the MSE is defined as follows:

$$l_{mse} = 1 - \frac{1}{n} \sum_{i=1}^n \left( \Phi_i^S(x) - \Phi_i^T(x) \right)^2, \quad (4)$$

where  $\Phi_i^S$  and  $\Phi_i^T$  are the templates obtained from the last fully connected layer of student and teacher models, respectively, and  $n$  is the size of the template. The final template KD loss, in this case, for the student model can be defined as:

$$L_{KD} = \alpha * L_{ce}(Y, P^S(x)) + \beta * t^2 * L_{kld}(P_s^S(x), P_s^T(x)) + \beta * l_{mse}(\Phi_i^S(x), \Phi_i^T(x)), \quad (5)$$

where  $L_{ce}$  and  $L_{kld}$  are the vanilla KD loss defined in Equation (3). In addition to MSE loss, we propose to evaluate the template distillation using cosine embedding loss defined as follows:

$$loss_{cosine} = 1 - \frac{1}{n} \sum_{i=1}^n \cos(\Phi_i^S(x), \Phi_i^T(x)). \quad (6)$$

The student loss using cosine distance for template KD can be defined as follows:

$$L_{KD} = \alpha * L_{ce}(Y, P^S(x)) + \beta * t^2 * L_{kld}(P_s^S(x), P_s^T(x)) + \beta * l_{cosine}(\Phi_i^S(x), \Phi_i^T(x)). \quad (7)$$

### 3.4. Training Paradigm

The training paradigm of the proposed template-driven KD approach can be described in a step-wise algorithmic manner as follows: (1) Given a high-performing teacher model  $T$  and compact student model  $S$ , a batch of  $n$  training samples  $(x_i, y_i)$  is passed into the teacher and then the student model.  $x_i \in X$  are periocular images of size  $224 \times 224$  pixels and  $y_i \in Y$  are their corresponding identity labels. (2) The outputs of the teacher model, that is, the predicted identity label  $P_s^T(x)$  and the feature embedding  $\Phi_i^T(x)$ , are used to calculate the loss function defined in Equation (7) for the student model. (3) The gradient descent method, stochastic gradient descent (SGD), is utilized to calculate the derivatives of the loss function with respect to the student model weights. (4) These weights are updated with the backpropagation. (5) Steps 1 to 4 are repeated until the student model is converged. It should be noted that the backpropagation is used only to update the weight of the student model. The teacher model is pretrained, and all its weights are frozen during the KD training paradigm.

## 4. Experimental Setup

This section presents the implementation details and evaluation settings used in this work.

### 4.1. Periocular Database for Evaluation

We evaluated our proposed approach on a semi-public internal dataset of periocular images captured using two different smartphones—iPhone 5S and Nokia Lumia 1020. We chose this database as it provides a realistic cross-device evaluation protocol that reflects the real use of such technology where the sensor that captures the reference might be different from the one capturing the probe. The database contains 152 unique periocular identities captured under mixed illumination settings with external illumination from sunlight and illumination from artificial room light in day-light. The periocular images were captured using the rear camera of the smartphones in a semi-cooperative manner (assisted capture setting). For each identity, a number of multiple samples were captured, resulting in a total set of 6642 images ( $3341 \times 2$  smartphones).

The periocular images in the database were captured to mimic everyday appearance variations that include make-up (e.g., the presence of mascara) and non-uniform illumi-

nation. The images also present a set of degradation due to motion blur and eye blinking, as encountered in everyday situations. Given that the images were captured in cross-illumination setting and with different cameras (iPhone 5S with 12 Mp and Nokia Lumia 1020 with 41 Mp), we first localized the periocular region using a Haar-cascade-based eye region detector. Further, the data were manually curated to eliminate any inconsistent segmentation in order to have only the periocular region. Given the nature of the images as illustrated in Figure 2, the challenging nature of periocular images could be deduced in cross-device verification settings. A careful observation Figure 2 reveals the variation in terms of appearance in different smartphones and the factors of degradation that were not constant across the phones/subjects.



**Figure 2.** Sample images from the periocular database employed.

For the sake of the experimental evaluation, we further divided the whole dataset of 152 unique periocular identities as a disjoint set of training and training with no identities in common. We chose the first 100 identities for the training dataset and the rest of 52 unique identities as a testing set to report our experimental results in the rest of this work.

#### 4.2. Baseline and KD Training

The presented models in this work use ResNet [44] as baseline architecture. Our choice of ResNet as the backbone was based on the accuracy achieved by the recent biometric solutions utilizing ResNet as a feature extraction model [5,6,47]. ResNet-34 and ResNet-18 were trained with a batch size of 128, and the ResNet-110 model was trained with a batch size of 16. All models were trained using an SGD optimizer with Nesterov momentum 0.9. The initial learning rates were set to  $\gamma = 0.001$  and  $\gamma = 0.1$  for teacher and student models, respectively, and it was dropped by a factor of 0.1 after epoch 10 for the ResNet-110 model. We resized the training and testing images to  $224 \times 224$  pixels to match the ResNet model input layer size. During the training, we augmented the training data by applying horizontal and vertical random shifting by up to 20% of the image width and/or height. ResNet-34 and ResNet-18 were fine-tuned for 5 epochs and ResNet-110 was trained for 25 epochs. ResNet-110 in conventional and template-driven KD approaches was trained using the exact ResNet-110 (without KD) training setup.

We followed the common choice for the knowledge distillation hyper-parameters [40,48,49] with temperature  $t \geq 4$ ,  $\alpha = 0.9$ , and  $\beta = 0.1$ .

#### 4.3. Evaluation Protocols and Metrics

We evaluated the verification performance with cosine distance for comparison. The result is reported first for the models ResNet-34, ResNet-18, and ResNet-110, without applying KD. Additionally, we report the result of the conventional KD approach with KLD loss using the ResNet-110 model as the student with either ResNet-34 or ResNet-18 as the teacher, denoted as ResNet-110\_KD34 and ResNet-110\_KD18 respectively. Moreover, we demonstrate the performance of the proposed template-driven KD by reporting the results

of using the ResNet-110 model as the student with either ResNet-34 or ResNet-18 as the teacher. These are denoted as ResNet-110\_KD34MSE and ResNet-110\_KD18MSE when the MSE loss is used, and as ResNet-110\_KD34COS and ResNet-110\_KD18COS when the cosine embedding loss is used.

For each of the settings, we investigated the verification performance under three different evaluation scenarios, defined as follows:

- iPhone verification scenario: The reference and the probe images were acquired using the camera of the iPhone smartphone.
- Nokia verification scenario: Similar to the previous scenario, the reference and the probe images were acquired using a Nokia smartphone.
- Cross-smartphone verification scenario: The reference and probe images were captured using two different smartphone cameras, where the reference images were captured using an iPhone camera and the probe images were captured using a Nokia camera.

The verification performance is reported using receiver operating characteristic (ROC) curves, area under the curve (AUC), FNMR at fixed FMR (FMR10: the lowest FNMR for  $FMR \leq 10\%$ , and FMR100: the lowest FNMR for  $FMR \leq 1\%$ ), and equal error rate (EER).

#### 4.4. Computational Efficiency

Table 1 presents the detailed analysis of the computational efficiency for each of the employed ResNet models. The computational efficiency of the deep learning approach depends on the number of trainable parameters and the inference latency. All evaluations were performed using PyTorch framework (Version 1.4) running on Linux OS with an Intel(R) Xeon(R) Gold 6130 CPU 2.10 GHz processor. ResNet-110 was the smallest model, with 1.8 m trainable parameters. The teacher models, ResNet-34 and ResNet-18, contained 21.3 m and 11.2 m trainable parameters. Regarding the inference time, ResNet-18 had the lowest latency of 0.007 ms. Note from Table 1 that the compact model (ResNet-110) had a higher latency than both ResNet-18 and ResNet-34. This is mainly because of the large number of convolutional layers in ResNet-110 (110 convolutional layers), as we discussed in Section 3, in comparison to 18 layers in ResNet-18 and 34 layers in ResNet-34. As the size of the model makes up the largest deployment challenge, the focus of this work was to enhance the accuracy of the small ResNet-110 model using the proposed template-driven KD.

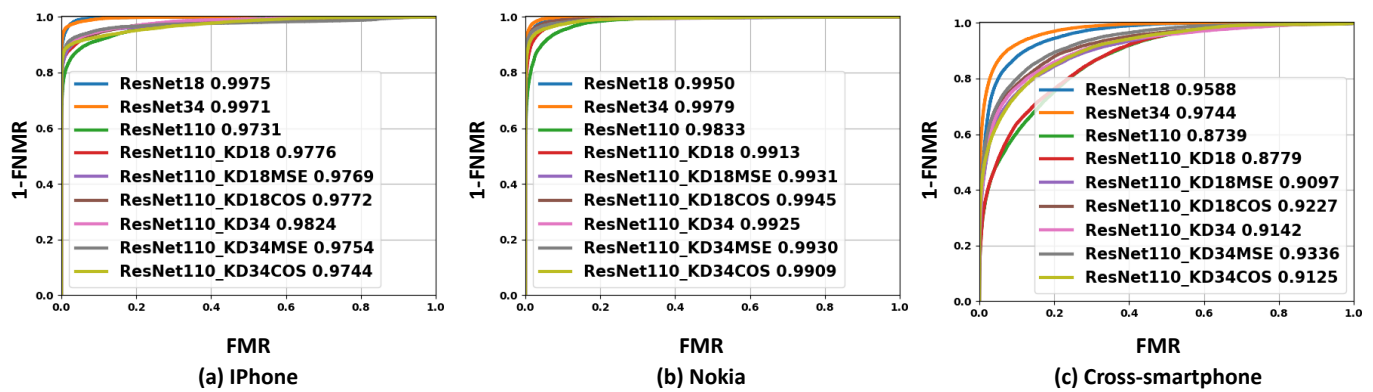
**Table 1.** The inference time (in milliseconds) and the number of trainable parameters (in millions (m)) for each of the evaluated models.

Model	Inference Time	No. Trainable Parameters
ResNet-110	0.015 ms	1.8 m
ResNet-18	0.007 ms	11.2 m
ResNet-34	0.008 ms	21.3 m

## 5. Results

The verification performances of the different experimental settings are presented in Figure 3 along with the EER, FMR10, and FMR100 values in Table 2. Figure 3a–c shows the achieved ROC curves of the iPhone verification scenario (a), the Nokia verification scenario (b), and the cross-smartphone verification scenario (c).





**Figure 3.** The achieved ROC curves for different experimental settings. For each experimental setting, the number next to the model label is the achieved AUC. Note the improvement in the ResNet-110 verification performance using our proposed template-driven KD approach.

**Table 2.** Performance obtained for different experimental settings. The first three rows of the table present the achieved result for the three evaluated models (without using KD), where the smallest model (ResNet-100) performed the worst. The next three rows of the table present the achieved verification performance by including KD in the training process using ResNet-18 as a teacher with conventional KD loss, and both the template-driven KD with MSE and cosine (COS) embedding loss. The last three rows present the achieved KD verification performance using ResNet-34 as a teacher with conventional KD, and both the template-driven KD with MSE and cosine embedding loss. The enhanced performance by the proposed method in comparison to the conventional KD is illustrated over all experimental setups. The lowest EER, FMR10 and FMR100 for each of the verification scenarios (iPhone, Nokia and Cross-Smartphone) achieved by ResNet-100 are in bold.

Model	Teacher	iPhone			Nokia			Cross-Smartphone		
		EER	FMR10	FMR100	EER	FMR10	FMR100	EER	FMR10	FMR100
ResNet-18	-	0.0251	0.0032	0.0557	0.0304	0.0108	0.0776	0.1104	0.1216	0.5171
ResNet-34	-	0.0273	0.0071	0.0431	0.0218	0.0033	0.0434	0.0871	0.0767	0.3267
ResNet-110	-	0.0883	0.0839	0.1984	0.0694	0.0473	0.2298	0.2216	0.3945	0.6932
ResNet-110_KD18	ResNet-18	0.0713	0.0547	0.1401	0.0471	0.0198	0.1284	0.2187	0.3619	0.6872
ResNet-110_KD18MSE	ResNet-18	0.0632	0.0500	0.1401	0.0353	0.0114	0.0763	0.1753	0.2539	0.5752
ResNet-110_KD18COS	ResNet-18	0.0702	0.0571	0.1171	<b>0.0330</b>	<b>0.0107</b>	0.0724	0.1587	0.2215	0.5051
ResNet-110_KD34	ResNet-34	0.0651	0.0530	0.1153	0.0413	0.0227	0.0844	0.1661	0.2321	0.5494
ResNet-110_KD34MSE	ResNet-34	<b>0.0614</b>	<b>0.0470</b>	<b>0.1136</b>	0.0375	0.0193	<b>0.0692</b>	<b>0.1469</b>	<b>0.2010</b>	<b>0.4984</b>
ResNet-110_KD34COS	ResNet-34	0.0775	0.0711	0.1161	0.0457	0.0265	0.0962	0.1725	0.2580	0.5368

### 5.1. Baseline

The first three rows of Table 2 present the achieved verification performance by ResNet-34, ResNet-18 and ResNet-110 (without KD). As expected, ResNet-34 and ResNet-18 achieved higher verification performances than the compact model ResNet-110 for different verification scenarios. In the cross-smartphone verification scenario, the best achieved EER was 8.17% by the ResNet-34 model, which indicates the high generalizability of this model. In this case, ResNet-18 achieved a competitive verification performance to ResNet-34 with 11.04% EER. ResNet-110 achieved the lowest verification performance, with 22.16% EER. For the iPhone verification scenario, the best achieved EER was 2.51% by ResNet-18 followed very closely by ResNet-34 with 2.73% EER, then ResNet-110 with 8.83% EER. The best-achieved EER for the Nokia verification scenario was 2.18% by ResNet-34, followed by ResNet-18 with 3.04% EER and ResNet-110 with 6.94% EER. Further, it can be noticed that the verification performances dropped for all evaluated models when the

reference and probe images were captured from different smartphones in comparison to the case where the probe and the reference images were captured from the same smartphone, as shown in Table 2. However, this degradation in the performance is a common issue for cross-smartphone verification scenarios, as reported in previous works [50].

### 5.2. Knowledge Distillation

The achieved results by the conventional KD approach are presented in Table 2 and Figure 1. It can be observed from the Table 2 that introducing conventional KD to the ResNet-110 training process (ResNet-110\_KD18 and ResNet-110\_KD34) highly improved the verification performances for different verification scenarios. For the cross-smartphone verification scenario, the EER was reduced from 22.16% to 21.87% and 16.61% when the knowledge was distilled from teacher ResNet-18 and ResNet-34, respectively. Similar improvements in verification performance can also be seen for Nokia and iPhone verification scenarios.

### 5.3. Template-Driven Knowledge Distillation

The results presented in Table 2 and Figure 3 illustrate that the verification performances of the ResNet-110 model were significantly improved by applying our proposed template-driven KD approach on the ResNet-110 training process in comparison to the case when ResNet-110 was trained with the conventional KD approach. For all verification scenarios, it can be observed from Table 2 that ResNet-110 trained with either cosine (ResNet-110\_KD18COS and ResNet-110\_KD34COS) or MSE embedding loss (ResNet-110\_KD18MSE and ResNet-110\_KD34MSE) outperformed the same model trained without KD and with the conventional KD approach. For the cross-smartphone verification scenario, the EER was improved from 21.87% (ResNet-110\_KD18) to 17.53% (ResNet-110\_KD18MSE) using MSE embedding loss and to 15.87% (ResNet-110\_KD18COS) using cosine embedding loss. A similar improvement in the verification performance was achieved using ResNet-34 as teacher with template-driven KD, where the best achieved EER, in this case, for the cross-smartphone scenario, was 14.69% (ResNet-110\_KD34MSE). For both Nokia and iPhone verification scenarios, the achieved verification performance by ResNet-110 using our proposed template-driven KD was significantly improved and even very close to the deeper models ResNet-34 and ResNet-18, as shown in Table 2 and Figure 3.

In comparison to the recent works that focused on providing compact periocular recognition models, our proposed model (1.8 m parameters) outperformed the DenseNet-20 (1.1 m parameters) proposed by [43] on the most challenging settings, cross-smartphone. The proposed DenseNet-20 by [43] was trained with conventional KD, where the compact model was trained to learn a similar classification decision to the teacher model. Different from conventional KD approaches [43], our proposed template-driven KD optimizes the student model so that it learns to output similar embeddings (template) to the teacher model, which is the requirement for a high-performing biometric template. The conventional classification KD in [43] is conceptually similar to our baseline ResNet-100\_KD18 and ResNet-110\_KD34. In this case, the best reported EER by [43] in cross-device verification was 15.54%, and the best reported one by our approach was 14.69%. Towards the goal of efficient periocular recognition, Almadan et al. [51] utilized conventional KD to train MobileNet-V2 [52] (3.5 m parameters), MobileNet-V3 [53] (2.5 m parameters), ResNet-20 [44] (1.3 m parameters), and ShuffleNetV2-50 [54] (1.4 m parameters) with ResNet50 [44] as a teacher model. These models were trained and evaluated on VISOB [55] and UFPR datasets [56]. Among the evaluated models, MobileNet-V2 (3.5 m parameters) achieved the lowest EER: 5.21% on VISOB and 5.38% on the UFPR dataset. Meanwhile, the ShuffleNetV2-50 model in [51], with a comparable size (1.4 m parameters) to our light model, scored an EER of 23.41% on VISOB and 22.52% on the UFPR datasets, although the results in [51] cannot be compared directly to the results in this work as they were evaluated on a different benchmark.

## 6. Conclusions

This work was motivated by the need for compact biometric deep learning models for deployment on resource-critical devices. Such models aim to have the capability to extract highly distinctive templates, as the larger models (more learned parameters) do. KD has been used to map such knowledge by teaching a smaller student model to produce a similar output as a larger teacher model. However, given that KD is commonly optimized on the classification output of such models, the knowledge of extracting biometric templates from pre-classification layers might not be optimally transferred to the student model. Therefore, we proposed a novel template-driven KD approach that aims at teaching the template extraction knowledge to the student model. The proposed approach was evaluated on smartphone periocular verification in intra- and cross-device settings. The achieved results showed that when the targeted small model was trained with our template-driven KD approach, it consistently outperformed similar models trained without KD or with the conventional KD approach.

**Author Contributions:** All the authors of this paper contributed to the work. The detailed contributions are as follows: Conceptualization, F.B. and N.D.; methodology, F.B., K.R. and N.D.; software, F.B.; validation, F.B. and N.D.; formal analysis, N.D.; investigation, K.R.; resources, K.R. and F.B.; data curation, K.R.; writing—original draft preparation, F.B.; writing—review and editing, F.B., N.D., K.R., F.K. and A.K.; visualization, F.B.; supervision, N.D. and A.K.; project administration, F.K.; funding acquisition, F.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research work has been funded by the German Federal Ministry of Education and Research and the Hessen State Ministry for Higher Education, Research and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

KD	knowledge distillation
MSE	mean squared error
FNMR	false non-match rate
FMR	false match rate
EER	equal error rate

## References

1. Krupp, A.; Rathgeb, C.; Busch, C. Social Acceptance of Biometric Technologies in Germany: A Survey. In Proceedings of the 2013 International Conference of the BIOSIG Special Interest Group (BIOSIG), Darmstadt, Germany, 4–6 September 2013; Volume P-212, pp. 193–200.
2. Gomez-Barrero, M.; Drozdowski, P.; Rathgeb, C.; Patino, J.; Todisco, M.; Nautsch, A.; Damer, N.; Priesnitz, J.; Evans, N.W.D.; Busch, C. Biometrics in the Era of COVID-19: Challenges and Opportunities. *arXiv* **2021**, arXiv:2102.09258.
3. Adjabi, I.; Ouahabi, A.; Benzaoui, A.; Taleb-Ahmed, A. Past, Present, and Future of Face Recognition: A Review. *Electronics* **2020**, *9*, 1188. [[CrossRef](#)]
4. Adjabi, I.; Ouahabi, A.; Benzaoui, A.; Jacques, S. Multi-Block Color-Binarized Statistical Images for Single-Sample Face Recognition. *Sensors* **2021**, *21*, 728. [[CrossRef](#)]
5. Deng, J.; Guo, J.; Xue, N.; Zafeiriou, S. Arcface: Additive angular margin loss for deep face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 4690–4699.
6. Boutros, F.; Damer, N.; Kirchbuchner, F.; Kuijper, A. ElasticFace: Elastic Margin Loss for Deep Face Recognition. *arXiv* **2021**, arXiv:2109.09416.

7. Boutros, F.; Damer, N.; Fang, M.; Kirchbuchner, F.; Kuijper, A. MixFaceNets: Extremely Efficient Face Recognition Networks. In Proceedings of the International IEEE Joint Conference on Biometrics, IJCB 2021, Shenzhen, China, 4–7 August 2021; pp. 1–8. [[CrossRef](#)]
8. Damer, N.; Boutros, F.; Süßmilch, M.; Kirchbuchner, F.; Kuijper, A. Extended evaluation of the effect of real and simulated masks on face recognition performance. *IET Biom.* **2021**, *10*, 548–561. [[CrossRef](#)] [[PubMed](#)]
9. Ngan, M.; Grother, P.; Hanaoka, K. *Ongoing Face Recognition Vendor Test (FRVT) Part 6A: Face Recognition Accuracy with Masks Using Pre-COVID-19 Algorithms*; Technical Report; NIST Interagency/Internal Report (NISTIR); National Institute of Standards and Technology: Gaithersburg, MD, USA, 2020. [[CrossRef](#)]
10. Department of Homeland Security. *Biometric Technology Rally at MDTF*; Technical Report. Available online: <https://www.dhs.gov/science-and-technology/news/2021/01/04/news-release-airport-screening-while-wearing-masks-test> (accessed on 4 January 2021).
11. Damer, N.; Grebe, J.H.; Chen, C.; Boutros, F.; Kirchbuchner, F.; Kuijper, A. The Effect of Wearing a Mask on Face Recognition Performance: An Exploratory Study. In Proceedings of the 2020 International Conference of the Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany, 16–18 September 2020; Gesellschaft für Informatik e.V.: Bonn, Germany, 2020; Volume P-306, pp. 1–10.
12. Damer, N.; Boutros, F.; Süßmilch, M.; Fang, M.; Kirchbuchner, F.; Kuijper, A. Masked Face Recognition: Human vs. Machine. *arXiv* **2021**, arXiv:2103.01924.
13. Huber, M.; Boutros, F.; Kirchbuchner, F.; Damer, N. Mask-invariant Face Recognition through Template-level Knowledge Distillation. In Proceedings of the 16th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2021, Jodhpur, India, 15–18 December 2021; pp. 1–8. [[CrossRef](#)]
14. Neto, P.C.; Boutros, F.; Pinto, J.R.; Damer, N.; Sequeira, A.F.; Cardoso, J.S. FocusFace: Multi-task Contrastive Learning for Masked Face Recognition. In Proceedings of the 16th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2021, Jodhpur, India, 15–18 December 2021; pp. 1–8. [[CrossRef](#)]
15. Boutros, F.; Damer, N.; Kirchbuchner, F.; Kuijper, A. Self-restrained triplet loss for accurate masked face recognition. *Pattern Recognit.* **2022**, *124*, 108473. [[CrossRef](#)]
16. Boutros, F.; Damer, N.; Kolf, J.N.; Raja, K.B.; Kirchbuchner, F.; Ramachandra, R.; Kuijper, A.; Fang, P.; Zhang, C.; Wang, F.; et al. MFR 2021: Masked Face Recognition Competition. In Proceedings of the International IEEE Joint Conference on Biometrics, IJCB 2021, Shenzhen, China, 4–7 August 2021; pp. 1–10. [[CrossRef](#)]
17. Deng, J.; Guo, J.; An, X.; Zhu, Z.; Zafeiriou, S. Masked Face Recognition Challenge: The InsightFace Track Report. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, ICCVW 2021, Montreal, BC, Canada, 11–17 October 2021; pp. 1437–1444. [[CrossRef](#)]
18. Damer, N.; Wainakh, Y.; Henniger, O.; Croll, C.; Berthe, B.; Braun, A.; Kuijper, A. Deep Learning-based Face Recognition and the Robustness to Perspective Distortion. In Proceedings of the 24th International Conference on Pattern Recognition, ICPR 2018, Beijing, China, 20–24 August 2018; IEEE Computer Society: Washington, DC, USA, 2018; pp. 3445–3450. [[CrossRef](#)]
19. RankOne. RankOne Inc.—Periocular Recognition Algorithm. 2020. Available online: <https://blog.rankone.io/2020/05/06/rank-ones-next-generation-periocular-recognition-algorithm/> (accessed on 6 May 2020).
20. Boutros, F.; Damer, N.; Raja, K.B.; Ramachandra, R.; Kirchbuchner, F.; Kuijper, A. Iris and periocular biometrics for head mounted displays: Segmentation, recognition, and synthetic data generation. *Image Vis. Comput.* **2020**, *104*, 104007. [[CrossRef](#)]
21. Boutros, F.; Damer, N.; Raja, K.B.; Ramachandra, R.; Kirchbuchner, F.; Kuijper, A. Periocular Biometrics in Head-Mounted Displays: A Sample Selection Approach for Better Recognition. In Proceedings of the 8th International Workshop on Biometrics and Forensics, IWBF 2020, Porto, Portugal, 29–30 April 2020; pp. 1–6. [[CrossRef](#)]
22. Raja, K.B.; Damer, N.; Ramachandra, R.; Boutros, F.; Busch, C. Cross-Spectral Periocular Recognition by Cascaded Spectral Image Transformation. In Proceedings of the 2019 IEEE International Conference on Imaging Systems and Techniques, IST 2019, Abu Dhabi, United Arab Emirates, 9–10 December 2019; pp. 1–7. [[CrossRef](#)]
23. Bianchini, M.; Scarselli, F. On the Complexity of Neural Network Classifiers: A Comparison Between Shallow and Deep Architectures. *IEEE Trans. Neural Netw. Learn. Syst.* **2014**, *25*, 1553–1565. [[CrossRef](#)]
24. Wu, X.; He, R.; Sun, Z.; Tan, T. A Light CNN for Deep Face Representation With Noisy Labels. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 2884–2896. [[CrossRef](#)]
25. Khaldi, Y.; Benzaoui, A.; Ouahabi, A.; Jacques, S.; Taleb-Ahmed, A. Ear Recognition Based on Deep Unsupervised Active Learning. *IEEE Sens. J.* **2021**, *21*, 20704–20713. [[CrossRef](#)]
26. Ge, S.; Zhao, S.; Li, C.; Li, J. Low-Resolution Face Recognition in the Wild via Selective Knowledge Distillation. *IEEE Trans. Image Process.* **2019**, *28*, 2051–2062. [[CrossRef](#)]
27. Park, U.; Ross, A.; Jain, A.K. Periocular biometrics in the visible spectrum: A feasibility study. In Proceedings of the 2009 IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems, Washington, DC, USA, 28–30 September 2009; pp. 1–6.
28. Xu, J.F.; Savvides, M. Subspace-based discrete transform encoded local binary patterns representations for robust periocular matching on NIST’s face recognition grand challenge. *IEEE Trans. Image Process.* **2014**, *23*, 3490–3505.
29. Mahalingam, G.; Ricanek, K. LBP-based periocular recognition on challenging face datasets. *EURASIP J. Image Video Process.* **2013**, *2013*, 36. [[CrossRef](#)]

30. Ross, A.; Jillela, R.; Smereka, J.M.; Boddeti, V.N.; Kumar, B.V.; Barnard, R.; Hu, X.; Pauca, P.; Plemmons, R. Matching highly non-ideal ocular images: An information fusion approach. In Proceedings of the 2012 5th IAPR International Conference on Biometrics (ICB), New Delhi, India, 29 March–1 April 2012; pp. 446–453.
31. Proença, H.; Neves, J.C. Deep-prwis: Periocular recognition without the iris and sclera using deep learning frameworks. *IEEE Trans. Inf. Forensics Secur.* **2017**, *13*, 888–896. [[CrossRef](#)]
32. Rattani, A.; Derakhshani, R. On fine-tuning convolutional neural networks for smartphone based ocular recognition. In Proceedings of the 2017 IEEE International Joint Conference on Biometrics (IJCB), Denver, CO, USA, 1–4 October 2017; pp. 762–767.
33. Zhao, Z.; Kumar, A. Accurate periocular recognition under less constrained environment using semantics-assisted convolutional neural network. *IEEE Trans. Inf. Forensics Secur.* **2016**, *12*, 1017–1030. [[CrossRef](#)]
34. Garg, R.; Baweja, Y.; Ghosh, S.; Singh, R.; Vatsa, M.; Ratha, N.K. Heterogeneity Aware Deep Embedding for Mobile Periocular Recognition. In Proceedings of the 9th IEEE International Conference on Biometrics Theory, Applications and Systems, BTAS 2018, Redondo Beach, CA, USA, 22–25 October 2018; pp. 1–7.
35. Alonso-Fernandez, F.; Raja, K.B.; Raghavendra, R.; Busch, C.; Bigün, J.; Vera-Rodríguez, R.; Fierrez, J. Cross-Sensor Periocular Biometrics: A Comparative Benchmark including Smartphone Authentication. *arXiv* **2019**, arXiv:1902.08123.
36. Ramachandra, R.; Stokkenes, M.; Mohammadi, A.; Venkatesh, S.; Raja, K.B.; Wasnik, P.; Poirer, E.; Marcel, S.; Busch, C. Smartphone Multi-modal Biometric Authentication: Database and Evaluation. *arXiv* **2019**, arXiv:1912.02487.
37. Ahuja, K.; Islam, R.; Barbhuiya, F.A.; Dey, K. Convolutional neural networks for ocular smartphone-based biometrics. *Pattern Recognit. Lett.* **2017**, *91*, 17–26. [[CrossRef](#)]
38. Zhu, M.; Gupta, S. To Prune, or Not to Prune: Exploring the Efficacy of Pruning for Model Compression. In Proceedings of the 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, 30 April–3 May 2018.
39. Jacob, B.; Kligys, S.; Chen, B.; Zhu, M.; Tang, M.; Howard, A.G.; Adam, H.; Kalenichenko, D. Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, 18–22 June 2018; IEEE Computer Society: Washington, DC, USA, 2018; pp. 2704–2713.
40. Hinton, G.E.; Vinyals, O.; Dean, J. Distilling the Knowledge in a Neural Network. *arXiv* **2015**, arXiv:1503.02531.
41. Chen, G.; Choi, W.; Yu, X.; Han, T.X.; Chandraker, M. Learning Efficient Object Detection Models with Knowledge Distillation. In Proceedings of the Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, Long Beach, CA, USA, 4–9 December 2017; pp. 742–751.
42. Cao, Q.; Shen, L.; Xie, W.; Parkhi, O.M.; Zisserman, A. VGGFace2: A Dataset for Recognising Faces across Pose and Age. In Proceedings of the 13th IEEE International Conference on Automatic Face & Gesture Recognition, FG 2018, Xi’an, China, 15–19 May 2018; IEEE Computer Society: Washington, DC, USA, 2018; pp. 67–74. [[CrossRef](#)]
43. Boutros, F.; Damer, N.; Fang, M.; Raja, K.B.; Kirchbuchner, F.; Kuijper, A. Compact Models for Periocular Verification Through Knowledge Distillation. In Proceedings of the 2020 International Conference of the Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany, 16–18 September 2020; Brömme, A., Busch, C., Dantcheva, A., Raja, K.B., Rathgeb, C., Uhl, A., Eds.; Gesellschaft für Informatik e.V.: Bonn, Germany, 2020; Volume P-306, pp. 291–298.
44. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
45. Deng, J.; Dong, W.; Socher, R.; Li, L.; Li, K.; Li, F. ImageNet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; IEEE Computer Society: Washington, DC, USA, 2009; pp. 248–255. [[CrossRef](#)]
46. Sun, S.; Cheng, Y.; Gan, Z.; Liu, J. Patient Knowledge Distillation for BERT Model Compression. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, 3–7 November 2019; Inui, K., Jiang, J., Ng, V., Wan, X., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 4322–4331. [[CrossRef](#)]
47. Liu, W.; Wen, Y.; Yu, Z.; Li, M.; Raj, B.; Song, L. SphereFace: Deep Hypersphere Embedding for Face Recognition. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, 21–26 July 2017; pp. 6738–6746. [[CrossRef](#)]
48. Cho, J.H.; Hariharan, B. On the efficacy of knowledge distillation. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019; pp. 4794–4802.
49. Furlanello, T.; Lipton, Z.C.; Tschannen, M.; Itti, L.; Anandkumar, A. Born-Again Neural Networks. In Proceedings of the 35th International Conference on Machine Learning Research, ICML 2018, Stockholm, Sweden, 10–15 July 2018; Volume 80, pp. 1602–1611.
50. Alonso-Fernandez, F.; Raja, K.B.; Busch, C.; Bigun, J. Log-likelihood score level fusion for improved cross-sensor smartphone periocular recognition. In Proceedings of the 2017 25th European Signal Processing Conference (EUSIPCO), Kos, Greece, 28 August–2 September 2017; pp. 271–275.
51. Almadan, A.; Rattani, A. Compact CNN Models for On-device Ocular-based User Recognition in Mobile Devices. In Proceedings of the IEEE Symposium Series on Computational Intelligence, SSCI 2021, Orlando, FL, USA, 5–7 December 2021; pp. 1–7. [[CrossRef](#)]

52. Reddy, N.; Rattani, A.; Derakhshani, R. Generalizable deep features for ocular biometrics. *Image Vis. Comput.* **2020**, *103*, 103996. [[CrossRef](#)]
53. Howard, A.; Pang, R.; Adam, H.; Le, Q.V.; Sandler, M.; Chen, B.; Wang, W.; Chen, L.; Tan, M.; Chu, G.; et al. Searching for MobileNetV3. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea, 27 October–2 November 2019; pp. 1314–1324. [[CrossRef](#)]
54. Ma, N.; Zhang, X.; Zheng, H.; Sun, J. ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design. In *Lecture Notes in Computer Science, Proceedings of the Computer Vision—ECCV 2018—15th European Conference, Munich, Germany, 8–14 September 2018*; Part XIV; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Springer: Berlin/Heidelberg, Germany, 2018; Volume 11218, pp. 122–138. [[CrossRef](#)]
55. Nguyen, H.M.; Reddy, N.; Rattani, A.; Derakhshani, R. VISOB 2.0—The Second International Competition on Mobile Ocular Biometric Recognition. In *Lecture Notes in Computer Science, Proceedings of the Pattern Recognition. ICPR International Workshops and Challenges, Virtual Event, 10–15 January 2021*; Part VIII; Bimbo, A.D., Cucchiara, R., Sclaroff, S., Farinella, G.M., Mei, T., Bertini, M., Escalante, H.J., Vezzani, R., Eds.; Springer: Berlin/Heidelberg, Germany, 2020; Volume 12668, pp. 200–208. [[CrossRef](#)]
56. Zanolensi, L.A.; Laroca, R.; Lucio, D.R.; Santos, L.R.; Britto, A.S., Jr.; Menotti, D. UFPR-Periocular: A Periocular Dataset Collected by Mobile Devices in Unconstrained Scenarios. *arXiv* **2020**, arXiv:2011.12427.