HETEROGENEOUS FACE RECOGNITION USING DOMAIN INVARIANT UNITS

Anjith George[1] and Sébastien Marcel[1,2]

[1]Idiap Research Institute, Martigny, Switzerland
[2]University of Lausanne, Switzerland

{anjith.george,sebastien.marcel}@idiap.ch

ABSTRACT

Heterogeneous Face Recognition (HFR) aims to expand the applicability of Face Recognition (FR) systems to challenging scenarios, enabling the matching of face images across different domains, such as matching thermal images to visible spectra. However, the development of HFR systems is challenging because of the significant domain gap between modalities and the lack of availability of large-scale paired multi-channel data. In this work, we leverage a pretrained face recognition model as a teacher network to learn domain-invariant network layers called Domain-Invariant Units (DIU) to reduce the domain gap. The proposed DIU can be trained effectively even with a limited amount of paired training data, in a contrastive distillation framework. This proposed approach has the potential to enhance pretrained models, making them more adaptable to a wider range of variations in data. We extensively evaluate our approach on multiple challenging benchmarks, demonstrating superior performance compared to state-of-the-art methods.

Index Terms— Face Recognition, Heterogeneous Face Recognition, Domain Invariant Units, Biometrics.

1. INTRODUCTION

Face recognition (FR) has gained popularity as an access control tool due to its effectiveness and user-friendly nature. State-of-the-art FR methods often demonstrate outstanding performance in real-world conditions [1]. Nevertheless, these methods fail when applied in challenging situations such as low light conditions or over long distances. Typically, FR systems operate in a homogeneous domain, enrolling and matching individuals using the same modality—usually employing facial images captured through an RGB camera. This
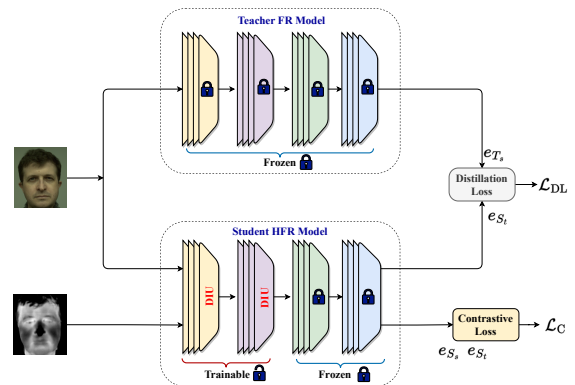
Fig. 1. The proposed Domain Invariant Unit (DIU) framework. The lower layers of the student model are trained in a contrastive framework to learn invariant features, while supervision from the distillation loss prevents overfitting.

approach proves inadequate when facing scenarios where enrollment and the probe images exhibit heterogeneity.

There are several instances where face matching in heterogeneous environments can prove beneficial. Consider a scenario in which the images used for enrollment originate from a controlled setting, while the target images come from a CCTV camera utilizing the Near Infrared (NIR) spectrum. NIR images can be used for both day and night regardless of illumination conditions. Similarly, the concept of matching images across different domains, such as visible and thermal images, proves beneficial in cases where the use of active illumination is not feasible. It's worth noting that although the term "heterogeneity" often refers to various modalities in the literature, discrepancies such as alterations in image quality (as seen in long-distance recognition) can also introduce heterogeneity. Heterogeneous Face Recognition (HFR) systems aim to facilitate cross-domain matching, allowing RGB images used for enrollment to be compared with images from any diverse modality. This obviates the necessity for separate modalities to undergo enrollment [2]. HFR is especially advantageous in scenarios where capturing good-quality visible images is challenging.

While being highly beneficial, achieving *HFR* using FR models trained using large-scale RGB datasets is challenging

due to the domain gap [3]. Moreover, the limited availability of paired datasets with the target modalities makes training invariant models from scratch challenging. Gathering large data for these new modalities can be cost-prohibitive. As a result, it's important to devise a framework that requires only a minimal set of paired data samples.

Several approaches have been proposed for heterogeneous face recognition (*HFR*) to extract invariant features that can match cross-modal images. Feature-based methods such as Difference of Gaussian (DoG) [4], and scale-invariant feature transform (SIFT) [5] have been proposed for *HFR* to reduce the domain gap. Another set of methods called common-space projection methods aims to learn a mapping to project different face modalities into a common shared subspace to reduce the domain gap [6, 7]. Recent HFR methods [8, 9], primarily use GAN-based synthesis due to its high-quality image generation. While using pre-trained FR models in synthesis-based HFR reduces the data needed for a new FR model's training, the synthesis increases computational costs, limiting real-world applicability. In [10], a new HFR method called Domain-Specific Units (DSU) was introduced. The authors argue that high-level features of CNNs trained on the visible spectrum can encode images from other modalities, indicating these features are domain-independent. Prepended Domain Transformers (PDT) [11] adds a separate module to a pretrained FR network for the target modality, converting it to an HFR network. In [12], authors treat different modalities as different styles and try to match them by modulating feature maps of the target modality using a conditional adaptive instance modulation (CAIM). In this work, in contrast to domain specific units (DSU), we propose to learn domain invariant units (DIU). We achieve this in a teacher-student distillation learning setting, adapting the lower network layers with a minimal set of paired samples. The main contributions of this work are listed below:

- We formulate the Heterogeneous Face Recognition (*HFR*) problem in a teacher-student distillation framework, leveraging a pretrained FR system as the teacher.

- The proposed approach can be used to improve pretrained models to more variability given a minimal amount of paired data.

- We have evaluated the proposed approach in several standard benchmarks to show the effectiveness of our approach.

Finally, the protocols and source codes will be made available publicly [1].

2. PROPOSED METHOD

Several recent works [10, 11] have shown that conditionally adapting low-level features can improve HFR performance.

For instance, Domain Specific Units (DSU) [10] proposes to create a copy of the initial layers from a Face Recognition (FR) network and adapt them specifically for the target modality. Prepended Domain Transfomers (PDT) on the other hand add a prepended module to the FR network specifically for the target modality. While these approaches perform well, they introduce an asymmetry in the data flow. The path for the source modality is always fixed (a pretrained FR network trained on a large face dataset), meaning the HFR training process has to always learn to map to this predefined space. This paradigm implies that the HFR training regime is invariably directed towards this pre-defined space. Such a constraint potentially curtails the training framework's capability to learn optimal invariant features. While training a single model for both source and target modality is possible, this will likely result in overfitting, especially given the limited volume of data typically found in heterogeneous datasets.

We propose to address this challenge by fine-tuning components of a pretrained face recognition network in a domain-invariant manner. More specifically, we aim to make the framework more flexible so that the source and target modality can align well in the embedding space by learning invariant layers. To counteract overfitting, we adopt a dual-strategy approach: 1) Restricting the training process to a specified subset of the lower layers, and 2) Incorporating supervisory signals from a teacher network, thereby safeguarding against catastrophic forgetting.

Let T denote a pretrained face recognition network with parameters given by Θ_{FR}, trained on a large-scale dataset. Let heterogeneous image pairs from the source and target modalities be represented by X_{s_i} and X_{t_i} respectively. The label y_i signifies if the pairs correspond to the same identity (1) or not (0). Given this, we define the embeddings from the model T for the source and target modalities as $e_{T_{s_i}}$ and $e_{T_{t_i}}$, respectively. In our formulation, the objective is to learn a *student* network S parameterized by Θ_{HFR}, such that the embeddings obtained for X_{s_i} and X_{t_i} belonging to the same identity ($y_i = 1$), $e_{S_{s_i}}$ and $e_{S_{t_i}}$ align in the embedding space.

Note that the values of Θ_{HFR} is initialized with Θ_{FR}. Now, we represent the weights of Θ_{HFR} into two categories.

$$\Theta_{HFR} = \{\Theta_{DIU_{i=1,2,..,K}}, \Theta_{Frozen}\} \quad (1)$$

Let Θ_{Frozen} represent the layers that remain unchanged during training, and Θ_{DIU_i} represent the ith DIU block out of a total of K blocks that have been trained to learn an invariant representation for both the source and target modalities.

To align the representation we use a cosine contrastive loss function, denoted by $\mathcal{L}_C$:

$$\mathcal{L}_C(e_{S_{s_i}}, e_{S_{t_i}}, y_i) = (1 - y_i) \cdot \max\left(0, \frac{e_{S_{s_i}} \cdot e_{S_{t_i}}}{\|e_{S_{s_i}}\|_2 \|e_{S_{t_i}}\|_2} - m\right)$$
$$+ y_i \cdot \left(1 - \frac{e_{S_{s_i}} \cdot e_{S_{t_i}}}{\|e_{S_{s_i}}\|_2 \|e_{S_{t_i}}\|_2}\right)$$
$$(2)$$

---

[1] https://gitlab.idiap.ch/bob/bob.paper.icassp2024_diu_hfr

Where m denotes the margin.

To ensure the stability of network training and prevent overfitting to the limited *HFR* data, we enforce a criterion such that the embeddings generated by the *student* network for the source modality images should match those produced by a teacher network. This approach resembles a distillation framework [13], wherein the student network is guided to mimic the representation learned by the teacher [14, 15]. In this context, it can be seen as a form of self-distillation; however, it is applied exclusively to images from the source modality and is focused solely on adapting the DIU layers.

Here the distillation loss has the following form $\mathcal{L}_{DL}$:

$$\mathcal{L}_{DL}(e_{T_{s_i}}, e_{S_{s_i}}) = \|e_{T_{s_i}} - e_{S_{s_i}}\|_2 \qquad (3)$$

Now the combined loss function to optimize can be written as:

$$\mathcal{L}(e_{S_{s_i}}, e_{S_{t_i}}, e_{T_{s_i}}, y_i) = (1 - \gamma) \cdot \mathcal{L}_C(e_{S_{s_i}}, e_{S_{t_i}}, y_i) \\ + \gamma \cdot \mathcal{L}_{DL}(e_{T_{s_i}}, e_{S_{s_i}}) \qquad (4)$$

where γ is a hyper-parameter to determine the contribution of each component in the total loss. The value of γ is empirically selected as 0.75, and the value of m is set as zero in all the experiments.

Pre-trained FR backbone: We employed the pre-trained *Iresnet101* face recognition model from AdaFace [16]. Specifically, the model we used was trained using the WebFace12M dataset [17], which contains over 12M images representing more than 600K identities. The model processes three-channel images with a resolution of 112×112 pixels. Before inputting into the Face Recognition network, the faces are aligned and cropped to align the eye centers with specific coordinates. For single-channel thermal images, the data is duplicated thrice during preprocessing.

Implementation details: The proposed Domain Invariant Units (DIU) is trained using a teacher-student distillation framework, complemented by a contrastive learning component. This framework was implemented using PyTorch and integrated with the Bob library [18, 19] [2]. We employed the Adam Optimizer, with a learning rate of 0.0001, and trained the framework for 50 epochs with a batch size of 48. The margin parameter m at zero and hyperparameter γ to 0.75 in all the experiments. We first initialize the student HFR models using the weights from the teacher. Only a certain number of lower layers of the student model are set as trainable, keeping the remaining ones frozen. The source channel image embeddings, obtained from the frozen teacher network, along with the contrastive loss acting on the student HFR network force the trainable layers to learn invariant representations for both modalities. During inference, solely the student network is required for performing the HFR task.

---

[2]https://www.idiap.ch/software/bob/

3. EXPERIMENTS

Databases and Protocols: The Polathermal dataset [20], offers polarimetric LWIR imagery alongside color images for a total of 60 participants. We employed the five-fold protocols suggested in [10] and presented the average Rank-1 identification rate across the five folds. The Tufts Face Database [21], consists of face images captured through different modalities, aimed for the *HFR* task. For assessing VIS-Thermal *HFR* performance, we utilized the thermal images from this dataset, covering 113 identities. We follow the standard protocols in [8] for reporting results. The SCFace dataset [22], comprises high-quality images suitable for face recognition during enrollment, whereas the probe samples are of lower quality, originating from surveillance cameras at different distances (heterogeneity in terms of image quality). This dataset contains four protocols, categorized according to the probe sample quality and distance: close, medium, combined, and far, with the "far" protocol representing the most challenging scenario.

Performance Metrics: We evaluate the models based on several performance metrics frequently used in recent literature, such as Area Under the Curve (AUC), Equal Error Rate (EER), Rank-1 identification rate, and Verification Rate at specific false acceptance rates (0.01%, 0.1%, 1%, and 5%).

3.1. Experimental results

The experiments conducted on various datasets and their results are elaborated in this section. For benchmarking, we contrast our results with those of the state-of-the-art, specifically compared with the methods presented in [11, 12].

Experiments with Polathermal dataset: Experiments were performed in the 'thermal-to-visible' recognition scenario in the Polathermal dataset, and the results are shown in Table 1. This table illustrates the mean Rank-1 identification rate across the five protocols of the Polathermal 'thermal to visible protocols' (following the reproducible protocols outlined in [10]). The introduced DIU method attains a mean Rank-1 accuracy of 97.8% with a standard deviation of 1.28%, achieving the state-of-the-art performance.

Table 1. Pola Thermal - Average Rank-1 recognition rate

Method	Mean (Std. Dev.)
DPM in [20]	75.31 % (-)
CpNN in [20]	78.72 % (-)
PLS in [20]	53.05% (-)
LBPs + DoG in [4]	36.8% (3.5)
ISV in [23]	23.5% (1.1)
DSU(Best Result) [10]	76.3% (2.1)
DSU-Iresnet100 [11]	88.2% (5.8)
PDT [11]	97.1% (1.3)
CAIM [12]	95.0% (1.63)
DIU (Proposed)	**97.8%(1.28)**

Experiments with Tufts face dataset: Table 2 showcases the performance of the DIU method in comparison to other

state-of-the-art methods in the VIS-Thermal protocol of the Tufts face dataset. This dataset poses significant challenges due to variations in pose and additional factors. Extreme yaw angles in the dataset lead to a performance dip for even the visible spectrum face recognition systems, paralleled by a decline in *HFR* performance. Nevertheless, the DIU method achieves the highest verification rate (85.9% at 1% FAR) and a Rank-1 accuracy of 82.94%. These results demonstrate the robustness of the proposed approach.

Table 2. Experimental results on VIS-Thermal protocol of the Tufts Face dataset.

Method	Rank-1	VR@FAR=1%	VR@FAR=0.1%
LightCNN [24]	29.4	23.0	5.3
DVG [25]	56.1	44.3	17.1
DVG-Face [8]	75.7	68.5	36.5
DSU-Iresnet100 [11]	49.7	49.8	28.3
PDT [11]	65.71	69.39	45.45
CAIM [12]	73.07	76.81	46.94
DIU (Proposed)	**82.94**	**85.9**	**74.95**

Experiments with SCFace dataset: We performed experiments on the SCFace dataset in the visible images protocol. This dataset poses a challenge due to the quality difference between gallery (high-res mugshots) and probe (low-res surveillance camera) images. We report the results on the most challenging "far" protocol with very low-quality probe images. From Table 3, it can be seen that the DIU approach achieves the highest Rank-1 accuracy at 94.55%. This shows that the proposed approach can address heterogeneity in terms of image quality as well, meaning it can even be used to improve the performance of pretrained models.

Table 3. Performance of the proposed approach in the SCFace dataset.

Protocol	Method	AUC	EER	Rank-1	VR@FAR=0.1%
Far	DSU-Iresnet100 [11]	97.18	8.37	79.53	58.26
	PDT [11]	98.31	6.98	84.19	60.00
	CAIM [12]	98.81	5.09	86.05	61.86
	DIU (Proposed)	**99.65**	**2.73**	**94.55**	**82.73**

Influence of the Number of DIU Layers: To understand the impact of varying the number of adaptable DIU layers, we performed an experiment on the Polathermal dataset, with different number of DIU layers. The results are shown in Tab. 4. Initially, it can be seen that adapting a greater number of layers enhances performance. However, this improvement plateaus at a certain point, after which performance begins to decline. Among the 48 layers, the best results were achieved when unfreezing up to the 24th layer.

Impact of γ: The hyper-parameter γ integrates the role of contrastive learning in aligning embeddings with the guidance of the teacher network during the distillation. To understand the influence of this parameter, we conducted experiments on the Polathermal dataset, examining how varying γ values affect the resulting HFR performance. As evidenced by Tab.

Table 4. Performance with different number of DIU blocks. 1-6 indicates the blocks from first to sixth layers are adapted. Experiment performed in Polathermal dataset.

Layers	AUC	EER	Rank-1	VR (0.1% FAR)	VR (1% FAR)
1	94.42±1.41	13.27±2.13	70.26±3.60	36.30±4.98	58.80±4.28
1-6	98.64±0.56	6.00±1.30	88.29±1.78	60.57±3.26	79.64±5.29
1-12	99.56±0.10	3.37±0.43	95.79±1.42	73.06±6.36	91.84±1.07
1-18	99.77±0.14	2.22±0.94	**97.33±1.08**	85.75±7.21	95.92±2.44
1-24	**99.84±0.11**	**2.06±0.87**	96.88±2.11	88.47±2.56	**95.97±2.20**
1-30	99.75±0.18	2.12±1.19	96.74±2.43	**89.07±7.15**	95.10±4.37
1-36	99.68±0.23	2.48±1.11	96.04±1.95	77.39±9.28	93.98±3.92
1-42	99.71±0.29	2.33±1.69	96.77±2.83	81.01±12.41	94.98±5.18
1-48	99.64±0.17	2.80±0.69	96.01±2.47	70.12±14.50	89.30±6.27

5, optimal performance is reached at $\gamma = 0.75$. Owing to the adaptation of only a select number of lower layers, DIU still achieves reasonable performance at $\gamma = 0$, which corresponds to supervision solely via contrastive loss. Incorporating the teacher's supervision further improves the results, achieving a peak AUC of 99.80%.

Table 5. Performance with different values of γ. Experiment performed in Polathermal dataset.

γ	AUC	EER	Rank-1	VR (0.1% FAR)	VR (1% FAR)
0.00	99.75±0.10	2.07±0.57	97.13±0.66	66.97±22.62	94.86±2.23
0.25	99.72±0.23	2.17±1.58	96.82±3.01	80.91±13.14	93.92±6.35
0.50	99.75±0.16	2.25±1.27	96.27±2.37	85.49±7.64	94.15±4.60
0.75	**99.80±0.16**	**2.02±1.06**	**97.82±1.28**	**87.50±6.85**	**96.34±2.53**
1.00	92.45±1.41	16.88±2.13	69.23±2.71	33.10±1.51	50.22±3.70

Effect of FR architecture: To compare the effectiveness of another architecture we considered the Iresnet50 model which is much smaller than the Iresnet100 model. As shown in Table 6, the performance is better for the bigger model for the HFR task (as well as for FR), however the small model still achieved good HFR performance showing the approach can be extended to both small and large FR models.

Table 6. Comparison of FR architectures on Tufts Face dataset.

Method	Rank-1	VR@FAR=1%	VR@FAR=0.1%
DIU (Iresnet100)	**82.94**	**85.9**	**74.95**
DIU(Iresnet50)	73.79	75.88	48.24

4. CONCLUSIONS

In this work, we introduce a novel approach for learning domain-invariant layers called domain invariant units (DIU) for the challenging task of heterogeneous face recognition. Leveraging guidance from a pretrained face recognition network, our method successfully extracts domain-invariant features while optimizing a contrastive learning objective. Our extensive experiments have showcased state-of-the-art performance achieved by our proposed method across a range of challenging benchmark datasets, demonstrating its effectiveness and robustness. The source codes and protocols will be made publicly available to facilitate the extension of our work.

5. REFERENCES

[1] Erik Learned-Miller, Gary B Huang, Aruni RoyChowdhury, Haoxiang Li, and Gang Hua, "Labeled faces in the wild: A survey," *Advances in face detection and facial image analysis*, vol. 1, pp. 189–248, 2016.

[2] Brendan F Klare and Anil K Jain, "Heterogeneous face recognition using kernel prototype similarities," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 6, pp. 1410–1422, 2012.

[3] Ran He, Xiang Wu, Zhenan Sun, and Tieniu Tan, "Wasserstein CNN: Learning invariant features for Nir-Vis face recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 7, pp. 1761–1773, 2018.

[4] Shengcai Liao, Dong Yi, Zhen Lei, Rui Qin, and Stan Z Li, "Heterogeneous face recognition from local structures of normalized appearance," in *International Conference on Biometrics*. Springer, 2009, pp. 209–218.

[5] Brendan Klare, Zhifeng Li, and Anil K Jain, "Matching forensic sketches to mug shot photos," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 3, pp. 639–646, 2010.

[6] Meina Kan, Shiguang Shan, Haihong Zhang, Shihong Lao, and Xilin Chen, "Multi-view discriminant analysis," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 1, pp. 188–194, 2015.

[7] Ran He, Xiang Wu, Zhenan Sun, and Tieniu Tan, "Learning invariant deep representation for Nir-Vis face recognition," in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[8] Chaoyou Fu, Xiang Wu, Yibo Hu, Huaibo Huang, and Ran He, "DVG-face: Dual variational generation for heterogeneous face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[9] He Zhang, Vishal M Patel, Benjamin S Riggan, and Shuowen Hu, "Generative adversarial network-based synthesis of visible faces from polarimetric thermal faces," in *2017 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 2017, pp. 100–107.

[10] Tiago de Freitas Pereira, André Anjos, and Sébastien Marcel, "Heterogeneous face recognition using domain specific units," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 7, pp. 1803–1816, 2018.

[11] Anjith George, Amir Mohammadi, and Sebastien Marcel, "Prepended domain transformer: Heterogeneous face recognition without bells and whistles," *IEEE Transactions on Information Forensics and Security*, 2022.

[12] Anjith George and Sebastien Marcel, "Bridging the Gap: Heterogeneous face recognition with conditional adaptive instance modulation," in *2023 International Joint Conference on Biometrics (IJCB)*. IEEE, 2023.

[13] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531*, 2015.

[14] Adriana Romero, Nicolas Ballas, Samira Ebrahimi Kahou, Antoine Chassang, Carlo Gatta, and Yoshua Bengio, "Fitnets: Hints for thin deep nets," *arXiv preprint arXiv:1412.6550*, 2014.

[15] Yonglong Tian, Dilip Krishnan, and Phillip Isola, "Contrastive representation distillation," *arXiv preprint arXiv:1910.10699*, 2019.

[16] Minchul Kim, Anil K Jain, and Xiaoming Liu, "Adaface: Quality adaptive margin for face recognition," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 18750–18759.

[17] Zheng Zhu, Guan Huang, Jiankang Deng, Yun Ye, Junjie Huang, Xinze Chen, Jiagang Zhu, Tian Yang, Jiwen Lu, Dalong Du, et al., "Webface260m: A benchmark unveiling the power of million-scale deep face recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 10492–10502.

[18] A. Anjos, M. Günther, T. de Freitas Pereira, P. Korshunov, A. Mohammadi, and S. Marcel, "Continuously reproducing toolchains in pattern recognition and machine learning experiments," in *International Conference on Machine Learning (ICML)*, Aug. 2017.

[19] A. Anjos, L. El Shafey, R. Wallace, M. Günther, C. McCool, and S. Marcel, "Bob: a free signal processing and machine learning toolbox for researchers," in *20th ACM Conference on Multimedia Systems (ACMMM), Nara, Japan*, Oct. 2012.

[20] Shuowen Hu, Nathaniel J Short, Benjamin S Riggan, Christopher Gordon, Kristan P Gurton, Matthew Thielke, Prudhvi Gurram, and Alex L Chan, "A polarimetric thermal database for face recognition research," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2016, pp. 119–126.

[21] Karen Panetta, Qianwen Wan, Sos Agaian, Srijith Rajeev, Shreyas Kamath, Rahul Rajendran, Shishir Paramathma Rao, Aleksandra Kaszowska, Holly A Taylor, Arash Samani, et al., "A comprehensive database for benchmarking imaging systems," *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 3, pp. 509–520, 2018.

[22] Mislav Grgic, Kresimir Delac, and Sonja Grgic, "SCface–surveillance cameras face database," *Multimedia tools and applications*, vol. 51, no. 3, pp. 863–879, 2011.

[23] Tiago de Freitas Pereira and Sébastien Marcel, "Heterogeneous face recognition using inter-session variability modelling," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2016, pp. 111–118.

[24] Xiang Wu, Ran He, Zhenan Sun, and Tieniu Tan, "A light CNN for deep face representation with noisy labels," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 11, pp. 2884–2896, 2018.

[25] Chaoyou Fu, Xiang Wu, Yibo Hu, Huaibo Huang, and Ran He, "Dual variational generation for low shot heterogeneous face recognition," in *Advances in Neural Information Processing Systems*, 2019.