F. Boutros, N. Damer, M. Fang, M. Gomez-Barrero, K. Raja,
C. Rathgeb, A. Sequeira, and M. Todisco (Eds.): BIOSIG 2024,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2024

NeutrEx-Light: Efficient Expression Neutrality Estimation for Facial Image Quality Assessment

Thanh Tung Linh Nguyen[1], Marcel Grimmer[2], Raymond Veldhuis[3], Christoph Busch[4]

Abstract: Face recognition systems have numerous practical applications, including device authentication, border control, and attendance monitoring. These systems must achieve high recognition performance while processing large volumes of data. To ensure that low-quality images do not degrade biometric performance, it is crucial to quantify the quality of biometric samples. The draft international standard ISO/IEC 29794-5 introduces the concept of component quality, which measures the quality of individual quality elements. In this work, we focus on NeutrEx, a recently proposed component quality measure that quantifies the expression neutrality of facial images. We optimize NeutrEx to improve its efficiency in terms of parameters, storage space, and inference time. We investigate the applicability of optimization techniques, including pruning and knowledge distillation, to enhance the throughput rate of NeutrEx to tailor it for high frequency real-world applications. All code and pre-trained experiments will be publicly available upon acceptance.

Keywords: Face Recognition, Quality Assessment, Pruning, Knowledge Distillation.

1 Introduction

The image-capturing process of biometric characteristics is crucial to the effectiveness of any biometric system, significantly impacting its overall recognition accuracy [AFFOG12, Os22]. In biometric quality assessment, *utility* refers to how a biometric sample contributes to the overall biometric performance. To this end, the draft international standard of ISO/IEC 29794-5 [IS24b] extends the concept of *utility* into *unified quality* and *component quality*. Unified quality assesses the overall image quality, while component quality evaluates how individual quality elements influence recognition performance.

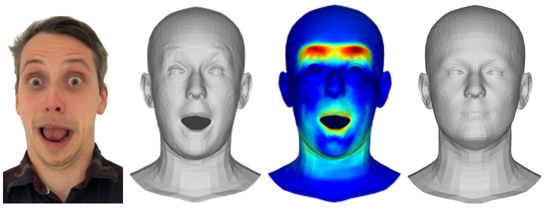
(a) Input (b) 3D Mesh (c) Residuals (d) Anchor

Fig. 1: Residual maps visualizing per-vertex Euclidean distances from the neutral anchor extracted from [Gr23].

Among these quality components, *expression neutrality* is a critical factor for the recognition capability of face recognition systems (FRS). However, quantifying expression neu-

[1] Hochschule Darmstadt, linhnt96@gmail.com
[2] Norwegian University of Science and Technology, marceg@ntnu.no
[3] University of Twente, r.n.j.veldhuis@utwente.nl
[4] Hochschule Darmstadt, christoph.busch@h-da.de

trality is challenging due to the high diversity of facial expressions, which must be disentangled from subject-specific attributes (*e.g.*, head shape) and demographic factors (*e.g.*, ethnicity or age). To address this, Grimmer et al. [Gr23] proposed a novel method - *NeutrEx*, which measures expression neutrality. NeutrEx shows promising results compared to state-of-the-art approaches in expression neutrality estimation. However, the authors also express concerns regarding the computational demand caused by the EMOCA [DBB22] and DECA [Fe21] encoders in the NeutrEx algorithm. This can be a potential bottleneck in systems with high throughput rates or mobile applications with limited resources.

In this work, we implement a more lightweight version of NeutrEx - *NeutrEx-Light* that focuses on efficiency by combining *pruning* and *knowledge distillation* optimization strategies with the main goal to reduce computational overhead while maintaining competitive performance with the original model. We benchmark NeutrEx-Light against NeutrEx extensively, including both efficiency and quality assessment metrics evaluated on two independent datasets (FEAFA+ [Ga22], Multi-PIE [Gr10]).

2 Background

2.1 NeutrEx

NeutrEx is developed in compliance with the current draft international standard of ISO/IEC 29794-5 [IS24a], focusing on assessing expression neutrality. NeutrEx measures how much a facial image deviates from a neutral expression anchor. Given a facial image, two monocular 3D face reconstruction encoders are used to infer the parameters required for the 3D face reconstruction. Both encoders use the ResNet-50 architecture, with approximately 50 million parameters each. The first encoder, DECA [Fe21], maps the image to a coarse shape representation, while the second encoder, EMOCA[DBB22], extracts emotion-sensitive information. These output parameters are translated into the FLAME [Li17] parameter space for 2D-to-3D reconstruction. To eliminate identity-specific or pose-specific biases, the 3D face model is transformed into a generic head shape with a frontalised pose.

To calculate the NeutrEx score, a dataset-specific neutral expression anchor is reconstructed in FLAME space composed of 5,023 vertices. The algorithm iterates over all vertices of the reconstructed 3D facial image, calculating the per-vertex Euclidean distance between the reconstruction and the anchor. The NeutrEx score is then mapped between 0-100 range using min-max scaling. NeutrEx-light aims to significantly reduce the parameters of the two encoders while maintaining competitive performance with the original model.

2.2 Pruning

Pruning refers to the process of systematically removing parameters from a neural network to reduce computational overhead and memory size [Bl20]. The goal of pruning is to derive a more lightweight model that retains similar performance to the original model. An extensive survey on pruning techniques [Bl20] concludes that pruning often leads to consistent reductions in computation and memory requirements.

In our study, we focus on structured pruning, which removes groups of parameters, such as channels [HZS17] or convolutional layers [CZM18]. This leads to changes in the model's structure. Li et al. [Li16] remove filters and their connecting filters in subsequent layers based on the L_1-norm. [He19] prunes filters based on the Geometric Median by assuming that the filter closest to others geometrically contains information that can be represented by those other filters and can thus be removed without significant performance loss.

2.3 Knowledge Distillation

Knowledge distillation, or teacher-student framework, is a form of transfer learning where learned parameters from a source model (teacher) are leveraged during the training of a smaller-sized model (student). The goal is for the student to mimic the behavior of the teacher, accepting minor performance degradation in favor of efficiency. Ideally, the student model can replicate the teacher model's predictions given the same input data.
Knowledge distillation can be applied during the training of the teacher model, updating both models simultaneously [Go21]. For example, [Bo22] distils knowledge from the teacher to the student at multiple points during training, allowing the student to continuously acquire knowledge and learn more complex patterns as both models converge. Alternatively, knowledge distillation can be applied after the teacher model is fully trained. This approach is particularly valuable when training a large neural network from scratch is infeasible due to hardware, dataset, or time constraints.

3 Experimental Setup

For brevity, we refer to our experiments using the following format: *method-dataset-pruner-encoder-sparse_ratio*, where *method* can be `pruning` or `finetuning`; *pruner* can be `l1` or `l2` for one of the pruners; *encoder* can be `cse` for the coarse shape encoder or `ee` for the expression encoder. The *sparse ratio* is chosen empirically between 0.1 to 0.7 with a 0.1 increment. The *dataset* tag is used only if a training dataset is used. Each component is further described in the next subsections.

3.1 Datasets

For the evaluation of NeutrEx-Light, we use the same method as the original NeutrEx, which is to measure predictive performance on two datasets Multi-PIE [Gr10] and FEAFA+ [Ga22]. Multi-PIE contains images taken in strictly controlled environment with variations in expression only, while FEAFA+ contains images that are varied in both expression, pose and lighting. The choices in the evaluation datasets show how the quality measure performs under different experimental setups. In addition to that, we also use a subset of AffectNet [MHM17] in our fine-tuning experiment, as it is originally used in training of the DECA [Fe21] and EMOCA [DBB22] encoders.

3.2 Pruning Setup

We utilize the `Neural Network Intelligence` (NNI) library[5] to facilitate pruning. Among the supported pruning algorithms, we select *L1 Norm pruner* and *L2 Norm pruner* as our algorithms. Both algorithms prune structurally, removing interconnected filters to enhance both storage efficiency and inference speed. During pruning, the most crucial hyperparameter is the *sparse ratio*, a value between 0 and 1 that specifies the proportion of encoder weights to remove. To maintain model consistency, we group pruning targets,

[5] `https://github.com/microsoft/nni`

such that each connected network component that shares the same *dependency group id* is removed during pruning. This can lead to additional pruning beyond the specified sparse ratio, potentially resulting in significant parameter reduction depending on the model's structure. We prune the encoder globally, i.e all convolutional layers and linear layers within the EMOCA and DECA ResNet-50-encoders are pruned, except for the output layer. Our experiments are influenced by the choice of the pruner and the sparse ratio.

3.3 Knowledge Distillation Setup

After pruning, the lightweight encoders are still generic and not tailored to a specific task. We implement a knowledge distillation setup illustrated in Figure 2, in which the original encoders act as teachers to fine-tune the pruned student encoders. By fine-tuning the pruned encoders, we can better adapt them to the expression neutrality measure and close the performance gap caused by the network size reduction. Each image in the training dataset is inferred twice using the pruned encoder and the original encoder. We calculate the *mean squared error* (MSE) between the inferred FLAME encodings, which acts as a loss function to optimize the parameters in the pruned encoder. We only apply fine-tuning to the pruned expression encoders, as its expression neutrality estimation performance significantly degrades after reducing the network size (see Section 4.2). We use a subset of 30k images from AffectNet [MHM17] for training, and another 3k images for validation.

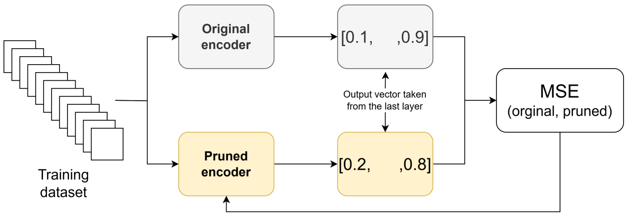

Fig. 2: Fine-tuning schema for pruned encoders. Values in the output vectors are examples.

4 Experimental Results

4.1 Pruning the Coarse Shape Encoder

In this section, we modify the coarse shape encoder, while keeping the expression encoder fixed. We calculate NeutrEx scores for images in both Multi-PIE and FEAFA+ and benchmark on each dataset separately in a leave-one-dataset-out approach, such that the neutral anchor is computed independently. All experiments effectively reduce the number of parameters of the encoders. Table 1 shows the pruning results with sparse ratios of 0.1 and 0.7 as example. The L1 Norm pruner consistently removes fewer parameters compared the L2 pruner at the same sparse ratio. At 0.7 ratio, L2 reduces the coarse shape encoder to 26k parameters, representing a removal of 99.9% of the original encoder weights.

Given the extensive number of experiments, we focus on the most insightful findings in subsequent discussions (Figure 3). We choose `pruning-l1-cse-0.1` and `pruning-l1--cse-0.7`, representing the least and most pruned encoder models while not having too many parameters removed. In terms of *Error-vs-Discard Characteristic* (EDC) curves [Sc22], the 0.1 sparse ratio experiments closely match NeutrEx's performance. Within the 0% to

Name	#Parameters	FLOPs	Size (MB)	pAUC (%) Multi-PIE / FEAFA+
`neutrex-cse`	25,848,108	4.09 billion	281.83	1.12 / 1.46
`pruning-l1-cse-0.1`	21,011,447	3.34 billion	245.18	1.14 / 1.57
`pruning-l1-cse-0.7`	2,404,745	401.50 million	64.15	1.25 / 1.31
`pruning-l2-cse-0.1`	21,011,447	3.34 billion	245.18	1.13 / 1.51
`pruning-l2-cse-0.7`	26,956	6.89 million	5.73	1.19 / 1.28

Tab. 1: Statistics after pruning the coarse shape encoder with selected pruning methods.

30% interval, defining the discard ratio at which facial images with large distances to the neutral anchor are removed, our experiments slightly underperform NeutrEx on Multi-PIE. `pruning-l1-cse-0.7` outperforms the original NeutrEx marginally on FEAFA+.

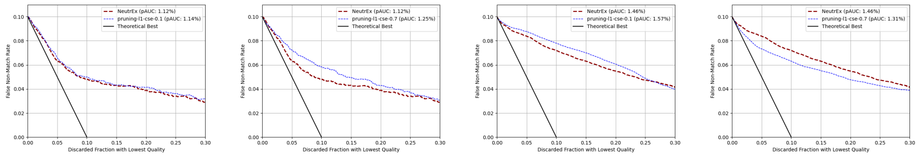

Fig. 3: EDC curves for `pruning-l1-cse-0.1` and `pruning-l1-cse-0.7`. Benchmarked with Multi-PIE (first two columns) and FEAFA+ (last two columns).

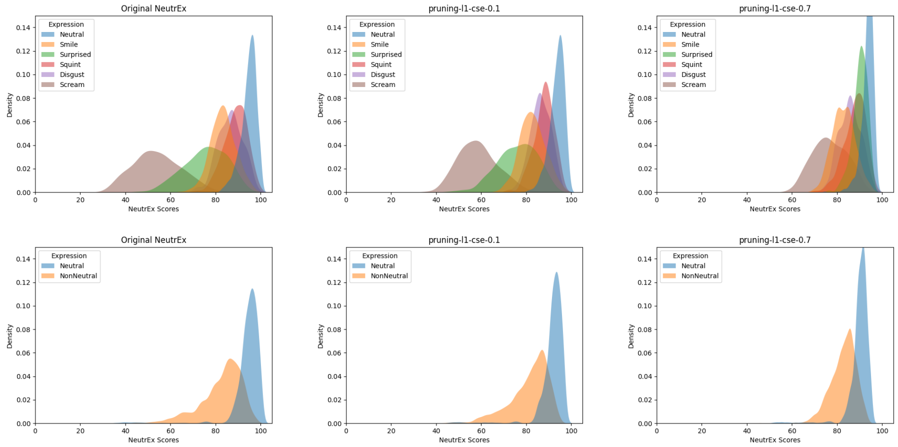

Fig. 4: CWD in Multi-PIE (top) and FEAFA+ (bottom). NeutrEx (left), `pruning-l1-cse-0.1` (middle), `pruning-l1-cse-0.7` (right) for comparison.

Figure 4 shows the *class-wise distributions* (CWD) on the Multi-PIE and FEAFA+ datasets. Our pruned encoder keeps the distribution characteristics for the majority of labels but the distribution of the 0.7 variant is heavily shifted to the right in both datasets, while the 0.1 variant maintains a similar distribution to NeutrEx. This indicates a lower performance in utility prediction tasks where we want to discard images with lower scores as they contribute less to the system. Our experiments indicate that pruning the coarse shape encoder is a viable method for parameter reduction in NeutrEx. The choice of pruners does not significantly impact the measured pAUCs. We observe diminishing performance differences at higher ratios, while experiments with lower ratios, such as 0.1, closely approximate NeutrEx. This demonstrates robust performance even with substantial parameter reduction.

4.2 Pruning the Expression Encoder

In line with our previous experiments, we pruned the expression encoder while fixing the coarse shape encoder from NeutrEx. We omit the L2 pruner due to its tendency to over-remove parameters. Table 2 summarizes the number of parameters retained in the expression encoder after pruning at sparse ratios of 0.1 and 0.7.

Name	**#Parameters**	**FLOPs**	**Size**	**pAUC (%)**
			(MB)	**Multi-PIE / FEAFA+**
`neutrex-ee`	25,657,458	4.09 billion	281.06	1.12 / 1.46
`pruning-l1-ee-0.1`	20,839,769	3.34 billion	244.50	1.15 / 1.53
`pruning-l1-ee-0.7`	21,748	6.89 million	5.71	1.17 / 1.74

Tab. 2: Statistics after pruning the expression encoder with our chosen pruner.

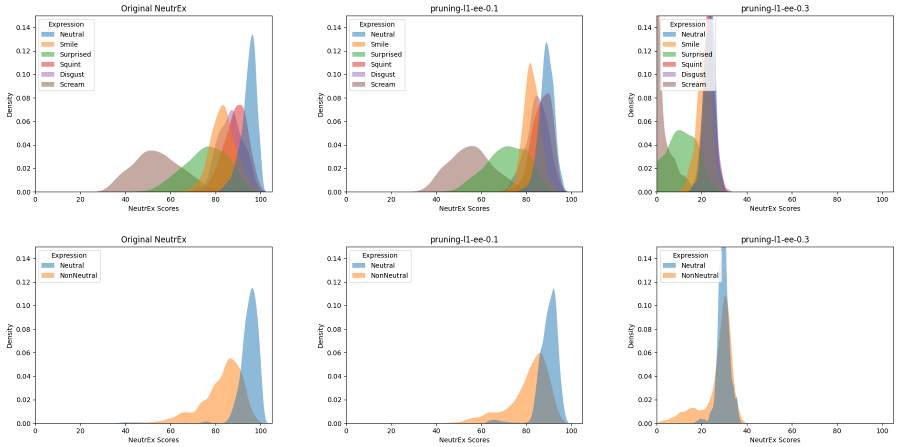

Fig. 5: CWD in Multi-PIE (top) and FEAFA+ (bottom). NeutrEx (1st column), `pruning-l1-ee-0.1` (2nd column) and `pruning-l1-ee-0.3` (3rd column).

Unlike previous experiments, pruning the expression encoder significantly degrades the expression neutrality measures, as shown in the compressed and overlapping CWD in Figure 5. Even at the lowest ratio of 0.1, all expression classes exhibit compressed and overlapping quality value distributions, thus making it more challenging to separate expression neutrality from deviating expression classes. Uniformly, we observe unacceptable performances when choosing sparse ratios beyond 0.1. At certain ratios, such as `pruning-l1-ee-0.3`, encoder functionality became compromised, highlighting the critical importance of choosing appropriate pruning strategies and ratios.

4.3 Knowledge Distillation Experiments

We selectively fine-tuned a subset of pruned expression encoders. We prioritize encoders with more remaining parameters to ensure convergence during fine-tuning. From the two pruners used for the expression encoder (see Section 4.2), we focus on the L1-norm pruning due to slightly more promising results compared to the alternative pruning techniques. We also present the experiment with faulty `pruning-l1-ee-0.3` to showcase the effectiveness of our fine-tuning schema. Figure 6 shows the EDC curves after fine-tuning with AffectNet. There are noticeable improvements when evaluated on the Multi-PIE dataset, as the fine-tuned model outperforms the original NeutrEx encoder. However, the FEAFA+ benchmark still indicates a performance degradation compared to NeutrEx's performance.

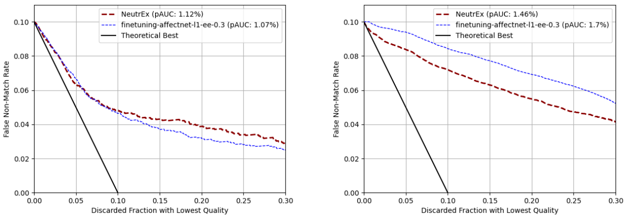

Fig. 6: EDC curves after fine-tuning (2nd column). Benchmarked with Multi-PIE (left) and FEAFA+ (right).

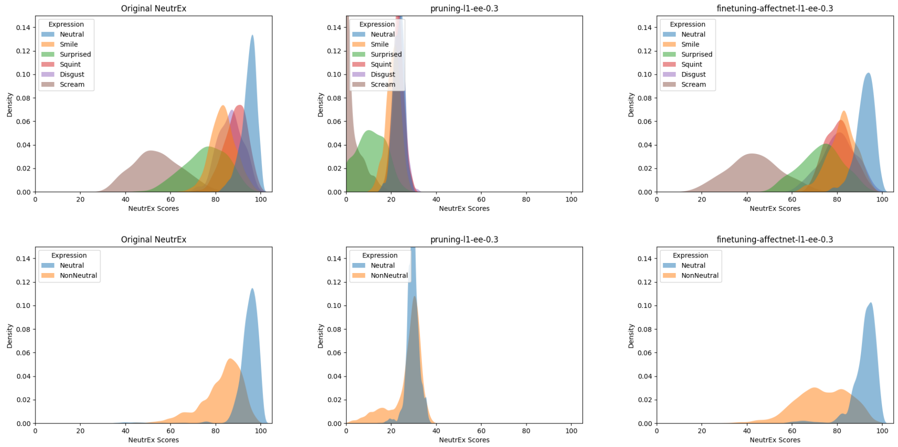

Fig. 7: CWD in Multi-PIE (top) and FEAFA+ (bottom). Original NeutrEx (1st column) for overall comparison. Pre-fine-tuning versions (2nd column and 4th column) and their respective fine-tuned ones (3rd and 5th column).

Figure 7 compares CWD on both datasets. Our experiments realign the quality measure distributions to closely resemble NeutrEx. For the *Neutral* expression class in both datasets, we maintain quality values within the [80, 100] range, crucial to discern it from other expression classes. While pruning focuses solely on parameter reduction, fine-tuning addresses task-specific optimization to align the pruned models to the original NeutrEx which proves to be crucial in mitigating severe performance degradation after pruning, as evidenced by `finetuning-affectnet-l1-ee-0.3`.

5 NeutrEx-Light

For NeutrEx-Light, we replace both the coarse shape encoder and the expression encoder in NeutrEx with our optimized versions. We use two encoders from previous experiments: a minimal coarse shape encoder (`pruning-l2-cse-0.7`), and a fine-tuned expression encoder (`finetuning-affectnet-l1-ee-0.3`) which comprises 12,670,175 parameters in total, representing a reduction of approximately 38 million parameters or almost 75% of the initial NeutrEx encoders.

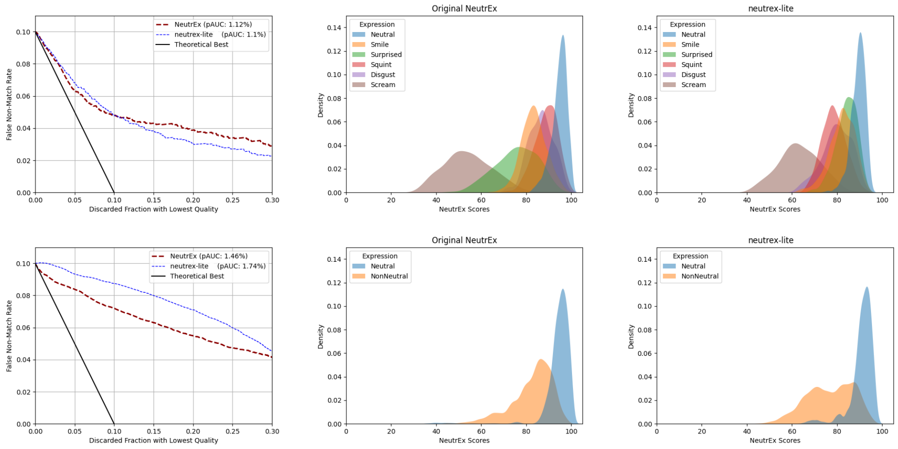

Fig. 8: EDC curves of `NeutrEx-Light` (left). CWD of NeutrEx (middle) and NeutrEx-Light (right). Benchmarked with Multi-PIE (top) and FEAFA+ (bottom).

We benchmark NeutrEx-Light against NeutrEx and report the results in Figure 8. NeutrEx-Light has a pAUC of 1.1% compared to 1.2% of NeutrEx when benchmarked with Multi-PIE, but 0.28% higher pAUC with FEAFA+. The CWD closely resembles those of NeutrEx in both datasets, matching our intended goal of mirroring the model's initial behavior. Our experiments reveal a fundamental trade-off between optimizing a single encoder to achieve both parameter reduction and performance improvement. While pruning the coarse shape encoder has shown little effect on the utility prediction performances, the expression encoder proved more volatile, requiring lower sparse ratios in combination with task-specific fine-tuning to maintain competitive performance to NeutrEx.

6 Conclusion

In this study, we introduce NeutrEx-Light, a more efficient variant of NeutrEx while maintaining its effectiveness as a facial expression quality measure in compliance to ISO/IEC 29794-5 [IS24b]. We demonstrate that pruning in combination with knowledge distillation effectively reduces computational complexity. With a combined 11 million parameters, our encoders reduce the number of parameters from the original NeutrEx model by almost 75% while still retaining the initial performance to a large extent, making NeutrEx-Light a viable option within face image quality assessment pipelines with high throughput rates.

Acknowledgment

This research work has been supported 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.

References

[AFFOG12] Alonso-Fernandez, Fernando; Fierrez, Julian; Ortega-Garcia, Javier: Quality Measures in Biometric Systems. IEEE Security & Privacy, 10(6):52–62, 2012.

[Bl20] Blalock, Davis; Gonzalez Ortiz, Jose Javier; Frankle, Jonathan; Guttag, John: What is the state of neural network pruning? Proceedings of machine learning and systems, 2:129–146, 2020.

[Bo22] Boutros, Fadi; Siebke, Patrick; Klemt, Marcel; Damer, Naser; Kirchbuchner, Florian; Kuijper, Arjan: Pocketnet: Extreme lightweight face recognition network using neural architecture search and multistep knowledge distillation. IEEE Access, 10:46823–46833, 2022.

[CZM18] Chin, Ting-Wu; Zhang, Cha; Marculescu, Diana: Layer-compensated pruning for resource-constrained convolutional neural networks. arXiv preprint arXiv:1810.00518, 2018.

[DBB22] Daněček, Radek; Black, Michael J; Bolkart, Timo: EMOCA: Emotion driven monocular face capture and animation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20311–20322, 2022.

[Fe21] Feng, Yao; Feng, Haiwen; Black, Michael J; Bolkart, Timo: Learning an animatable detailed 3D face model from in-the-wild images. ACM Transactions on Graphics (ToG), 40(4):1–13, 2021.

[Ga22] Gan, Wei; Xue, Jian; Lu, Ke; Yan, Yanfu; Gao, Pengcheng; Lyu, Jiayi: FEAFA+: an extended well-annotated dataset for facial expression analysis and 3D facial animation. In: Fourteenth International Conference on Digital Image Processing (ICDIP 2022). volume 12342. SPIE, pp. 307–316, 2022.

[Go21] Gou, Jianping; Yu, Baosheng; Maybank, Stephen J; Tao, Dacheng: Knowledge distillation: A survey. International Journal of Computer Vision, 129:1789–1819, 2021.

[Gr10] Gross, Ralph; Matthews, Iain; Cohn, Jeffrey; Kanade, Takeo; Baker, Simon: Multi-pie. Image and vision computing, 28(5):807–813, 2010.

[Gr23] Grimmer, Marcel; Rathgeb, Christian; Veldhuis, Raymond; Busch, Christoph: NeutrEx: A 3D Quality Component Measure on Facial Expression Neutrality. arXiv preprint arXiv:2308.09963, 2023.

[He19] He, Yang; Liu, Ping; Wang, Ziwei; Hu, Zhilan; Yang, Yi: Filter Pruning via Geometric Median for Deep Convolutional Neural Networks Acceleration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019.

[HZS17] He, Yihui; Zhang, Xiangyu; Sun, Jian: Channel pruning for accelerating very deep neural networks. In: Proceedings of the IEEE international conference on computer vision. pp. 1389–1397, 2017.

[IS24a] ISO/IEC JTC1 SC37 Biometrics: . ISO/IEC 29794-1 Information Technology - Biometric Sample Quality - Part 1: Framework. International Organization for Standardization, 2024.

[IS24b] ISO/IEC JTC1 SC37 Biometrics: . ISO/IEC DIS 29794-5 Information Technology - Biometric Sample Quality - Part 5: Face Image Data. Intl. Organization for Standardization, 2024.

[Li16] Li, Hao; Kadav, Asim; Durdanovic, Igor; Samet, Hanan; Graf, Hans Peter: Pruning filters for efficient convnets. arXiv preprint arXiv:1608.08710, 2016.

[Li17] Li, Tianye; Bolkart, Timo; Black, Michael J; Li, Hao; Romero, Javier: Learning a model of facial shape and expression from 4D scans. ACM Trans. Graph., 36(6):194–1, 2017.

[MHM17] Mollahosseini, Ali; Hasani, Behzad; Mahoor, Mohammad H: Affectnet: A database for facial expression, valence, and arousal computing in the wild. IEEE Transactions on Affective Computing, 10(1):18–31, 2017.

[Os22] Osorio-Roig, D.; Schlett, T.; Rathgeb, C.; Tapia, J.; Busch, C.: Exploring Quality Scores for Workload Reduction in Biometric Identification. In: 2022 International Workshop on Biometrics and Forensics (IWBF). pp. 1–6, 2022.

[Sc22] Schlett, Torsten; Rathgeb, Christian; Henniger, Olaf; Galbally, Javier; Fierrez, Julian; Busch, Christoph: Face image quality assessment: A literature survey. ACM Computing Surveys, 54(10s):1–49, 2022.