

MASKDUL: DATA UNCERTAINTY LEARNING IN MASKED FACE RECOGNITION

Libo Zhang, Weiming Xiong, Ku Zhao, Kehan Chen, Mingyang Zhong*

College of Artificial Intelligence, Southwest University, Chongqing, China

ABSTRACT

Since mask occlusion causes plentiful loss of facial feature, Masked Face Recognition (MFR) is a challenging image processing task, and the recognition results are susceptible to noise. However, existing MFR methods are mostly deterministic point embedding models, which are limited in representing noise images. Moreover, Data Uncertainty Learning (DUL) fails to achieve reasonable performance in MFR. Therefore, we propose a novel two-stream convolutional network, masked face data uncertainty learning (MaskDUL), that solves the problems by sampling uncertainty and intra-class distribution learning in MFR. Specifically, a Hard Kullback-Leibler Divergence (H-KLD) method is proposed to serve as an adaptive variance regularizer and a magnitude-based module is adopted to adaptively adjust the angular margin of different samples. Finally, insightful evaluation demonstrates the effectiveness and robustness of our MaskDUL.

Index Terms—Masked Face Recognition, Data Uncertainty Learning, Intra-class Distribution Learning

I. INTRODUCTION

Masked face recognition (MFR) has gradually aroused great attention because of COVID-19, and it also has a wide range of applications in academia and industry [1]. Due to the mask occlusion, extensive facial features are lost, as shown in Fig. 1(a). The masked face is difficult to recognize, and the result is also susceptible to noise [2]–[4]. In response to the aforementioned difficulties, many MFR methods based on Deterministic Point Embedding (DPE) have been proposed, such as Deep Cascaded Regression (DCR) [5], Upper Patch Attention (UPA) [6], and MFR algorithm based on large margin Cosine loss (MFCosface) [7]. However, it is difficult to estimate a precise point embedding for noise images in the latent space and the optimization direction is fixed towards the class center, which is prone to overfitting noise, as shown in Fig. 1(b). Therefore, MFR is a challenging and important image processing task [8].

As noise in images is inevitable, modeling the uncertainty is critical to the image processing task [9]. Data Uncer-

tainty Learning (DUL) is an ingenious uncertainty method that estimates a Gaussian distribution for each sample in latent space [10]. The classification model DUL_{cls} enhances the learning of low-uncertainty (high-quality) samples and weakening the learning of high-uncertainty noise, which is superior to other uncertainty methods in image processing [9], [11]. Unfortunately, due to the mask occlusion, DUL tends to regard the masked face as noise, thus weakening the optimization to it, as shown in Fig. 1(c). Therefore, DUL fail to achieve reasonable performance in MFR tasks (See Section II for details).

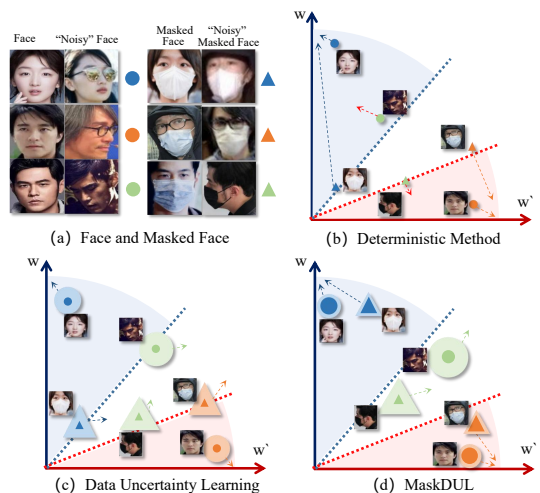


Fig. 1. (a) Samples of non-masked and masked faces, represented by circle and triangle respectively. (b) DPE method in MFR. w and w' are the class centers of different identities. The arrow indicates the optimization direction and the wrong direction is marked in red. (c) DUL in MFR. The inner circle represents the mean (feature) and the outer circle represents the variance (uncertainty). (d) MaskDUL in MFR. The high-quality faces are pulled to the class center, while the low-quality faces are pushed far away.

To address the aforementioned problems, we propose a novel two-stream convolutional network called Masked face Data Uncertainty Learning (MaskDUL). Specifically, we construct an adaptive variance regularizer for masked faces called Hard Kullback-Leibler Divergence (H-KLD) based on the DUL method and introduce a magnitude-based method to improve intra-class distribution. The effect is shown in Fig. 1(d). Finally, experiments prove the advancement of our

This work is supported by the National Nature Science Foundation of China (No. 62106205), the Oversea Study and Innovation Foundation of Chongqing (No. cx2021105), and the Natural Science Foundation of Chongqing (No. cstc2021jcyj-msxmX0824), and in part by the Chongqing Municipal Training Program of Innovation and Entrepreneurship for Undergraduate under Grant S202210635201. * Corresponding author.

MaskDUL. The contributions of this work are as follows:

(1) To the best of our knowledge, this is the first work that applies uncertainty learning to MFR tasks.

(2) The constructed two-stream convolutional network solves the problems of treating masked faces as noise in the traditional DUL methods.

(3) Our MaskDUL obtains high-quality sample distributions, in which high-quality masked and non-masked faces from the same class are close to the class center, while the real noise samples are far away from the center.

II. THE LIMITATIONS OF DUL

The softmax loss of DUL_{cls} is calculated as follows:

$$\mathcal{L}'_{cls} = \frac{1}{N} \sum_i^N -\log \frac{e^{\mathbf{w}_{y_i} \mathbf{s}_i}}{\sum_c^C e^{\mathbf{w}_c \mathbf{s}_i}} \quad (1)$$

where \mathbf{s}_i is an equivalent sampling representation consisting of identity feature of the face $\boldsymbol{\mu}_i$ and its uncertainty $\boldsymbol{\sigma}_i$, and ϵ is a standard Gaussian random noise independent of model. To avoid degenerating into a deterministic model [12], a regularization term is adopted to explicitly constrain $\mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\sigma}_i)$ to approach a standard normal distribution, $\mathcal{N}(\mathbf{0}, \mathbf{I})$, measured by Kullback-Leibler Divergence (KLD):

$$\begin{aligned} \mathcal{L}_{kl} &= KL [N(\mathbf{z}_i | \boldsymbol{\mu}_i, \boldsymbol{\sigma}_i^2) \| N(\epsilon | \mathbf{0}, \mathbf{I})] \\ &= -\frac{1}{2} (1 + \log \boldsymbol{\sigma}^2 - \boldsymbol{\mu}^2 - \boldsymbol{\sigma}^2) \end{aligned} \quad (2)$$

\mathcal{L}_{kl} assigns small variance to the high-quality sample to enhance the learning to it, and assign large variance to noise to avoid overlearning. The loss of facial features caused by mask occlusion will aggravate the uncertainty of identity recognition of masked faces. DUL_{cls} encounters some new challenges in the MFR tasks for two reasons:

(1) during the training period, masked faces are prone to be misjudged as noise, thus weakening the optimization;

(2) the model devotes to estimating uncertainty, but neglects its further use in intra-class distribution learning, which lead to overly scattered intra-class distribution.

LFW is a non-masked face dataset, while M-LFW is the corresponding synthetic masked dataset. The inconsistent sample distribution learned by DUL is shown in Fig. 2 (a).

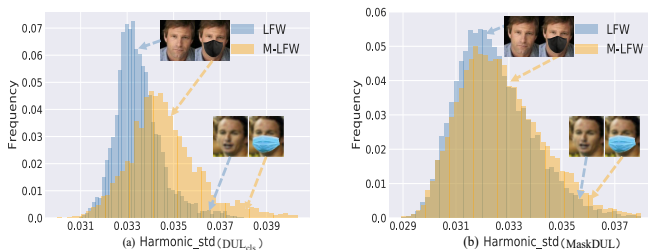


Fig. 2. The distribution learned by DUL_{cls} and MaskDUL.

III. PROPOSED METHOD

III-A. Overall Architecture

MaskDUL consists of a two-stream network with face and masked face branches, and an optimization mechanism of data uncertain learning based on hard variance constraint. The training pipeline of MaskDUL is illustrated in Fig. 3.

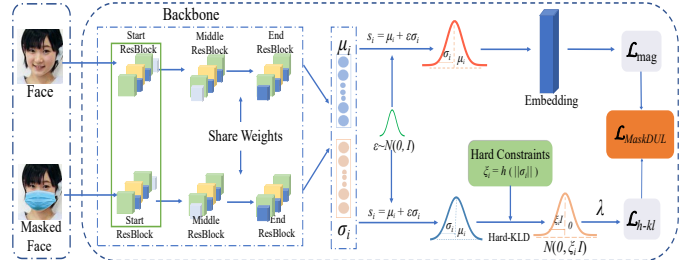


Fig. 3. Training pipeline of MaskDUL. Thereinto, M-VGGFace2 is adopted as the training set, which is a large-scale synthetic masked face dataset with 50% masked face proportion. Then we choose IResNet-50 as the backbone to extract two 512-D vectors, namely $\boldsymbol{\mu}_i$ and $\boldsymbol{\sigma}_i$.

Two branches have independent Start ResBlock and output different mean $\boldsymbol{\mu}_i$ and variance $\boldsymbol{\sigma}_i$, which are further reparameterized to obtain the equivalent sampling representation \mathbf{s}_i to calculate \mathcal{L}_{mag} . Weight sharing strategy enables the model to learn shared cross-domain feature between face and masked face. Then H-KLD module is adopted to impose a hard constraint on variance uncertainty $\|\boldsymbol{\sigma}_i\|$ to calculate \mathcal{L}_{h-kl} . MaskDUL can dynamically adjust the learning preference of face and masked face according to above two loss, and gain the optimal learning effect through the joint optimization of $\mathcal{L}_{MaskDUL}$ as follows:

$$\mathcal{L}_{MaskDUL} = \mathcal{L}_{mag} + \lambda \mathcal{L}_{h-kl}, \quad (3)$$

where λ is a trade-off hyperparameter for H-KLD.

III-B. Hard Kullback-Leibler Divergence

To diminish the variance of between the masked faces and non-masked faces, we adopt a standard normal distribution with smaller variance, $\mathcal{N}(\mathbf{0}, (\boldsymbol{\xi}\mathbf{I})^2)$, to constrain the estimated distribution of masked face, $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$. Thereinto, $\boldsymbol{\xi}$ indicates the constraint factor, which is used to control the constraint strength of masked face. Then, to achieve more accurate constraints, we define a linear decreasing function $h(\|\boldsymbol{\sigma}\|)$ on $[l_\sigma, u_\sigma]$ to determine the constraint factor $\boldsymbol{\xi} = h(\|\boldsymbol{\sigma}\|)$ according to the current variance uncertainty $\|\boldsymbol{\sigma}\|$ of masked face sample. For masked face sample:

$$h(\|\boldsymbol{\sigma}\|) = \begin{cases} \frac{u_\xi - l_\xi}{u_\sigma - l_\sigma} (u_\sigma - \|\boldsymbol{\sigma}\|) + l_\xi, & \|\boldsymbol{\sigma}\| \in [l_\sigma, u_\sigma] \\ u_\xi, & \|\boldsymbol{\sigma}\| \in [u_\sigma, +\infty) \end{cases} \quad (4)$$

where l_σ, u_σ represents the lower and upper bound of variance uncertainty, and l_ξ, u_ξ represents the lower and upper bound of constraint factors. For the sample, $\|\boldsymbol{\sigma}\| \in [l_\sigma, u_\sigma]$, we

put a monotonically decreasing constraint factor on it, while the sample beyond the upper bound, $\|\sigma\| \in [u_\sigma, +\infty)$, we keep the original constraint to avoid overfitting. Next, we also adopt KLD to measure the distance between above two distributions to construct a Hard KLD (H-KLD) with hard constraints, thus obtaining the expression of \mathcal{L}_{h-kl} as follows:

$$\begin{aligned} \mathcal{L}_{h-kl} &= KL [N(\mathbf{z}_i | \boldsymbol{\mu}_i, \boldsymbol{\sigma}_i^2) \| N(\boldsymbol{\epsilon} | \mathbf{0}, (\xi \mathbf{I})^2)] \\ &= -\frac{1}{2} \left(1 + \log \frac{\boldsymbol{\sigma}^2}{\xi^2} - \frac{\boldsymbol{\mu}^2}{\xi^2} - \frac{\boldsymbol{\sigma}^2}{\xi^2} \right). \end{aligned} \quad (5)$$

For face samples, \mathcal{L}_{h-kl} is equal to \mathcal{L}_{kl} , while for masked face samples, \mathcal{L}_{h-kl} imposes strong constraints on high-quality masked face samples and avoids overfitting noisy masked face samples simultaneously.

III-C. Magnitude-based module

We introduced the magnitude-based module to encode quality metrics into face representation, which optimize the model according to the magnitude $a_i = \|f_i\|$. Further, it designed a magnitude-aware angular margin $m(a_i)$ to impose a larger angular margin constraint on the high-recognizable samples to approach the class center, while a smaller constraint on the low-recognizable samples to approach the origin. In addition, to stabilize the intra-class structure, it designed the regularizer $g(a_i)$ to push each sample towards the boundary of the feasible region, and reward sample with large magnitude.

The loss function is calculated as follows:

$$\mathcal{L}_{Mag} = \frac{1}{N} \sum_{i=1}^N \mathcal{L}'_i, \quad (6)$$

where \mathcal{L}'_i is presented as follows:

$$\mathcal{L}'_i = -\log \frac{e^{s \cos(\theta_{y_i} + m(a_i))}}{e^{s \cos(\theta_{y_i} + m(a_i))} + \sum_{j \neq y_i} e^{s \cos \theta_j}} + \lambda'_g g(a_i), \quad (7)$$

where s is the scaling parameter and λ'_g is a trade-off hyperparameter between loss items. Finally, through theoretical derivation and experimental analysis, we determine the boundary of magnitude, construct a strictly increasing convex function $m(a_i)$ and a strictly decreasing convex function $g(a_i)$ and prove the convergence and monotonicity of \mathcal{L}_{Mag} .

IV. EXPERIMENTS

IV-A. Datasets and Implementation Details

The datasets used in our experiments are shown in Table I. MaskTheFace [13] tool is adopted to synthesize training set M-VGGFace2 with 5 mask styles and 50% masked face proportion. Then all the images are aligned to 112×112 by following the settings in ArcFace [3].

IResNet-50 [14] is adopted as backbone and the head of the baseline model is: Backbone – Flatten – FC – BN, while the Data Uncertainty Learning (DUL) methods have an additional head branch. All the models are trained for

30 epochs using a SGD optimizer [15] with a momentum of 0.9, weight decay of $5e-4$, dropout probability of 0.4 and batch size of 128. The initial learning rate is 0.05 and divided by 10 at 10, 18, 25 epochs. RandomCrop and RandomHorizontalFlip are used for data augmentation. For MaskDUL, l_ξ is 0.5 and u_ξ is 1, l_σ is 10 and u_σ is 25, and λ is set to 0.01. Other settings are consistent with those provided in literature [3]–[7], [9]–[12]. See the website (<https://github.com/MySky37/MaskDUL>) for details.

Table I. Description of the datasets.

Task	Datasets	#Identity	#Image	#Pair	#Variation
Train	M-VGGFace2 [16]	9.1K	6.6M	—	Full
	RMFRD* [17]	426	4,015	7178	Unconstrained
MFV	MFR2* [13]	53	269	1200	Limited
	M-LFW [7]	5,749	13,233	12,000	Limited
	M-CALFW [18]	5,749	12,174	12,000	Large-age
	M-AgeDB-30 [17]	568	16,488	12,000	Large-age
	M-CFP-FP [17]	500	7,000	14,000	Large-pose
	M-CASIA-FaceV5 [7]	500	2,500	10,000	Limited

* indicates real-world test set that is suitable for MFI task.

IV-B. Overall Comparison Results

We evaluate the proposed MaskDUL from the tasks of MFV and MFI with the benchmarks, baselines and results presented in Tables II and III. Overall, for MFV task, MaskDUL outperforms the baselines on most benchmarks, and achieves an average accuracy of 91.85, with a 1.11% to 12.97% performance improvement over the other methods. And for MFI task, MaskDUL outperforms the baselines in all benchmarks, especially on *RMFRD*, with a 1.11% to 19.02% performance improvement of Rank-1 over the other methods.

Then we analyze 5 categories of models as follows:

(1) For FR models, ArcFace [3] has the worst performance on both MFV and MFI tasks. Once again, the inapplicability of applying FR model directly to MFR task is proved though the performance of MagFace [4] improves to some extent.

(2) For Data Uncertainty Learning (DUL) models, such as PFE [11] and DUL_{cls} [12], which proposed to use probability embedding to represent noise samples more accurately. However, because mask occlusion will mislead the models assigning large variance to masked face, the DUL models only achieve limited performance improvement.

(3) OFR [19] models achieve good results in FR task with random and small-area occlusion, but they are not effective enough on MFR task.

(4) PFR [20] models adopt partial face as input, ignoring the overall visual appearance, which causes their performance improvement is limited.

(5) The other three MFR methods, such as the second place method, UPA [6] with dual-branch training strategy is effective to some extent. But it is still deterministic point embedding model with a limited performance.

(6) We proposes the hard regularization term, H-KLD, and magnitude-based module, which enables the model to learn

Table II. Masked face verification (MFV) accuracy (%) of various methods.

Category	Methods	Real-world		Synthetic					Avg
		Limited	Unconstrained	Limited		Large-age	Large-pose		
		MFR2	RMFRD	M-LFW	M-CASIA-FaceV5	M-CALFW	M-AgeDB_30	M-CFP_FP	
FR	ArcFace	89.37	64.81	86.63	88.36	76.36	75.11	71.52	78.88
	MagFace	93.64	70.97	91.02	92.78	83.41	81.50	80.41	84.82
DUL	PFE	91.13	68.03	88.69	90.85	79.78	77.93	74.93	81.62
	ProbFace	91.37	68.56	89.02	90.12	80.01	77.22	75.28	81.65
	DUL-clS	92.32	69.43	89.64	91.38	81.35	79.66	78.31	83.16
	DUL-GM	92.97	68.98	89.61	92.03	81.26	78.99	77.69	83.08
OFR	PDSN	95.76	74.24	94.91	94.57	87.84	85.19	84.33	88.12
PFR	DFM	96.61	76.89	94.52	95.43	86.33	86.43	84.25	88.64
MFR	DCR	97.27	77.87	94.98	97.28	89.86	88.02	85.47	90.11
	MFCosFace	97.35	78.09	95.17	96.03	90.02	89.22	88.93	90.69
	UPA	98.22	79.55	96.12	97.43	89.17	88.37	86.29	90.74
	MaskDUL	99.01	80.23	96.32	97.40	91.03	90.02	88.91	91.85

The best result is shown in bold (similarly hereinafter).

more accurate uncertainty representations and construct a more compact intra-class distribution in MFR, thus obtaining remarkable performance improvements.

Table III. Masked face identification (MFI) accuracy (%) of various methods.

Category	Methods	MFR2		RMFRD	
		Rank-1	Rank-5	Rank-1	Rank-5
FR	ArcFace	86.37	88.65	53.12	62.28
	MagFace	90.62	92.36	63.86	72.19
DUL	PFE	87.37	89.73	57.93	66.12
	ProbFace	87.01	89.02	58.65	66.93
	DUL-clS	88.11	90.78	60.68	68.65
	DUL-GM	88.93	90.85	61.54	70.23
OFR	PDSN	93.64	95.03	67.43	76.77
PFR	DFM	93.33	95.51	67.99	76.21
MFR	DCR	95.78	96.97	68.62	77.02
	MFCosFace	95.94	97.98	70.31	79.54
	UPA	96.72	98.73	71.03	80.11
	MaskDUL	97.15	99.05	72.14	80.73

IV-C. Ablation Study and Robustness Analysis

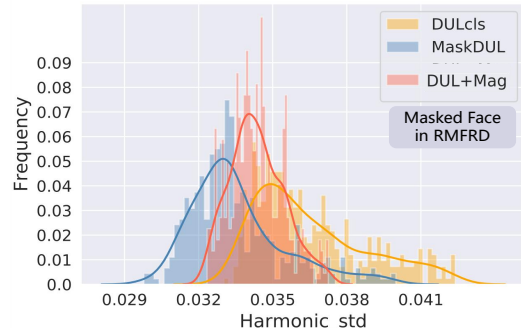
In Table IV, DUL-clS and MagFace are adopted as baselines to testify the improved performance brought by introducing MagFace into DUL-clS and H-KLD into MaskDUL. The performance improvement (1.85%~3.85%) achieved by DUL+Mag indicates that the advantages of the two models can complement each other and achieve better performance. The noise experiment is conducted on Gaussian blurred *M-VGGFace2*, and the results fully prove the effectiveness and strong anti-noise ability of H-KLD.

In Fig. 4, three models are used to visualize the uncertainty distributions on *RMFRD* to evaluate the learning ability to clean and noisy samples. Specifically, DUL-clS tends to assign large variance to masked faces. DUL+Mag makes the overall variance of masked face smaller and more concentrated.

Table IV. Performance and Robustness in Ablation Study.

Methods	MFV		MFI	Noise	
	RMFRD	M-CALFW	MFR2	RMFRD	M-CFP_FP
DUL-clS	69.43	79.66	90.78	51.51	62.42
MagFace	70.97	81.50	92.36	52.78	63.68
DUL+Mag	72.93	83.51	94.21	54.83	66.23
MaskDUL	80.23	90.02	99.05	62.36	73.15

But the optimization of the intra-class distribution is still necessary. MaskDUL further limits the variance of masked face to a low level, but also retains some true noises with large variance, thus distinguishing the high-recognizable masked face from the true noises in the intra-class distribution.

**Fig. 4.** Modules effectiveness in ablation study.

V. CONCLUSION

In this work, we propose a novel two-stream convolutional network with a Hard Kullback-Leibler Divergence (H-KLD) and magnitude-based module, called Masked Face Data Uncertainty Learning (MaskDUL). MaskDUL explores sample uncertainty and intra-class distribution learning in MFR, which enables the model to learn more accurate uncertainty representations and construct a more compact intra-class distribution in MFR. Finally, comprehensive experiments on MFR tasks prove the advancement of our MaskDUL.

VI. REFERENCES

- [1] A. T. Alzu'bi Ahmad, Albalas Firas, "Masked face recognition using deep learning: A review," *Electronics*, no. 10, 2021.
- [2] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815–823.
- [3] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2019, pp. 4690–4699.
- [4] Q. Meng, S. Zhao, Z. Huang, and F. Zhou, "Magface: A universal representation for face recognition and quality assessment," in *Computer Vision and Pattern Recognition*, 2021.
- [5] M. Geng, P. Peng, Y. Huang, and Y. Tian, "Masked face recognition with generative data augmentation and domain constrained ranking," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 2246–2254.
- [6] Y. Zhang, X. Wang, M. S. Shakeel, H. Wan, and W. Kang, "Learning upper patch attention using dual-branch training strategy for masked face recognition," *Pattern Recognition*, vol. 126, p. 108522, 2022.
- [7] H. Deng, Z. Feng, G. Qian, X. Lv, H. Li, and G. Li, "Mfcosface: a masked-face recognition algorithm based on large margin cosine loss," *Applied Sciences*, vol. 11, no. 16, p. 7310, 2021.
- [8] A. Alzu'bi, F. Albalas, T. Al-Hadhrami, L. B. Younis, and A. Bashayreh, "Masked face recognition using deep learning: A review," *Electronics*, vol. 10, no. 21, p. 2666, 2021.
- [9] K. Chen, Q. Lv, and T. Yi, "Fast and reliable probabilistic face embeddings in the wild," *arXiv: Computer Vision and Pattern Recognition*, 2021.
- [10] H. Ai, Q. Liao, Y. Chen, and J. Qian, "Gaussian mixture distribution makes data uncertainty learning better," in *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*. IEEE, 2021, pp. 01–08.
- [11] Y. Shi and A. K. Jain, "Probabilistic face embeddings," in *International Conference on Computer Vision*, 2019.
- [12] J. Chang, Z. Lan, C. Cheng, and Y. Wei, "Data uncertainty learning in face recognition," in *Computer Vision and Pattern Recognition*, 2020.
- [13] A. Anwar and A. Raychowdhury, "Masked face recognition for secure authentication," *arXiv preprint arXiv:2008.11104*, 2020.
- [14] I. C. Duta, L. Liu, F. Zhu, and L. Shao, "Improved residual networks for image and video recognition," in *2020 25th International Conference on Pattern Recognition*, 2021, pp. 9415–9422.
- [15] W. Ilboudo, T. Kobayashi, and K. Sugimoto, "Adaterm: Adaptive t-distribution estimated robust moments towards noise-robust stochastic gradient optimizer," *arXiv e-prints*, 2022.
- [16] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," in *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*. IEEE, 2018, pp. 67–74.
- [17] Z. Wang, G. Wang, B. Huang, Z. Xiong, Q. Hong, H. Wu, P. Yi, K. Jiang, N. Wang, Y. Pei *et al.*, "Masked face recognition dataset and application," *arXiv preprint arXiv:2003.09093*, 2020.
- [18] C. Wang, H. Fang, Y. Zhong, and W. Deng, "Mlfw: A database for face recognition on masked faces," *arXiv preprint arXiv:2109.05804*, 2021.
- [19] L. Song, D. Gong, Z. Li, C. Liu, and W. Liu, "Occlusion robust face recognition based on mask learning with pairwise differential siamese network," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 773–782.
- [20] L. He, H. Li, Q. Zhang, and Z. Sun, "Dynamic feature learning for partial face recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7054–7063.