mathematics

MDPI

Article

PRNet: A Priori Embedded Network for Real-World Blind Micro-Expression Recognition

Xin Liu [1,†], Fugang Wang [1,†], Hui Zeng [2], Yile Chen [1], Liang Zheng [1] and Junming Chen [1,*]

[1] Faculty of Humanities and Arts, Macau University of Science and Technology, Macau 999078, China; 3220001492@student.must.edu.mo (X.L.); fgwang@must.edu.mo (F.W.); chenyile@must.edu.mo (Y.C.); zhengliang@must.edu.mo (L.Z.)
[2] School of Design, Jiangnan University, Wuxi 214122, China; 7230306023@stu.jiangnan.edu.cn
* Correspondence: jmchen@must.edu.mo
† These authors contributed equally to this work.

Abstract: Micro-expressions, fleeting and often unnoticed facial cues, hold the key to uncovering concealed emotions, offering significant implications for understanding emotions, cognition, and psychological processes. However, micro-expression information capture presents challenges due to its instantaneous and subtle nature. Furthermore, it is affected by unpredictable degradation factors such as device performance and weather, and model degradation issues persist in real scenarios, and directly training deep networks or introducing image restoration networks yields unsatisfactory results, hindering the development of micro-expression recognition in real-world applications. This study aims to develop an advanced micro-expression recognition algorithm to promote the research of micro-expression applications in psychology. Firstly, Generative Adversarial Networks (GANs) are employed to build high-quality micro-expression generation models, which are then used as prior decoders to model micro-expression features. Subsequently, the GAN priors of deep neural networks are fine-tuned using low-quality facial micro-expression images. The designed micro-expression GAN module ensures that the generation of latent codes and noise inputs suitable for micro-expression GAN blocks from the deep and shallow features of deep neural networks. This approach controls the reconstruction of facial structure, local details, and accurate expressions to enhance the stability of subsequent recognition networks. Additionally, a Multi-Scale Dynamic Cross-Domain (MSCD) module is proposed to dynamically adjust the input of reconstructed features to different task representation layers. Doing so effectively integrates reconstructed features and improves the micro-expression recognition performance. Experimental results demonstrate that our method consistently achieves superior performance on multiple datasets, achieving particularly significant performance improvements in micro-expression recognition for severely degraded facial images in real scenarios.

Keywords: micro-expression recognition; image classification; transfer learning; image reconstruction; generative adversarial network

MSC: 68T07; 68T20; 68T30; 68T45

check for updates

Academic Editors: Jian Luo, Zheming Gao and Xin Yan

Received: 10 February 2025
Revised: 22 February 2025
Accepted: 24 February 2025
Published: 25 February 2025

Citation: Liu, X.; Wang, F.; Zeng, H.; Chen, Y.; Zheng, L.; Chen, J. PRNet: A Priori Embedded Network for Real-World Blind Micro-Expression Recognition. *Mathematics* **2025**, *13*, 749. https://doi.org/10.3390/math13050749

Copyright: © 2025 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

Facial expressions are highly efficient nonverbal communication tools that are crucial in interpersonal emotional exchanges [1], which typically encompass macro- and micro-expressions. Macro-expressions involve facial muscle movements with relatively longer

durations and greater magnitudes, and their prominence and perceptibility enable the clear conveyance of an individual's emotional state. Unlike typical macro-expressions, micro-expressions often emerge when an individual subjected to stimuli attempts to conceal genuine emotions. Micro-expressions are characterized by subtle facial changes, rapid onsets, and brief duration, rendering their identification significantly more challenging than macro-expressions. Non-professionals cannot recognize or classify micro-expressions without assistance, while the information they convey represents unconscious genuine responses to external stimuli. Given the substantial correlation between micro-expression manifestations and psychological states during deception, micro-expression recognition holds broad application prospects in corporate management, national security, clinical medicine, criminal investigation, and judicial interrogation [2,3].

Micro-expression recognition involves the simultaneous analysis of semantics and micro-movements. Often difficult to observe with the naked eye, micro-expression capture typically requires high-speed cameras. Current micro-expression feature extraction and recognition methods mainly fall into three categories: optical flow methods, improved local binary pattern methods, and convolutional neural networks (CNNs). Although CNNs have made significant progress in micro-expression recognition, they still exhibit certain limitations in real-world scenarios, including issues with motion blur and model degradation due to image quality compression [4]. These challenges hinder the development of micro-expression recognition for understanding human emotions, cognition, and psychological processes in real-world research contexts.

To the best of our knowledge, this is the first integration of image reconstruction into the micro-expression recognition network to advance the research on micro-expression recognition from real-world low-quality images. The high-frequency information in the images, including texture and structure, is restored first to produce more explicit micro-expression images. Subsequently, the newly generated micro-expression images are utilized for image recognition and micro-expression recognition. The proposed approach significantly enhances the model's adaptability to extreme scenarios and demonstrates a remarkable effectiveness across multiple datasets.

The primary contributions of this study are as follows:

1. A prior network framework is proposed for micro-expression generation to capture the feature distribution of micro-expression datasets. This framework fine-tunes facial micro-expression recognition networks by leveraging facial structure and expression information, thus overcoming facial micro-expression degradation in real-world scenarios. Moreover, its versatility enables applications across various micro-expression recognition networks.
2. A multi-scale dynamic cross-domain module, MSCD, is employed to efficiently reconstruct features for integration into micro-expression recognition tasks. It dynamically adjusts the features of the reconstruction task to ensure their effective transmission to the representation layer of subsequent recognition tasks.
3. In this study, we propose a novel method that contributes significantly to micro-expression recognition. It attains top-tier results on multiple standard benchmark datasets like CASME II, CASME3, SAMM, and SMIC, outperforming many existing approaches. Additionally, its remarkable effectiveness in complex degradation scenarios showcases its unique value. This validates the method's superiority in handling real-world challenges.

2. Related Works

2.1. Face Reconstruction

Facial image restoration is a crucial subfield in image restoration with a rich history of research. An early study by Zhang, H. et al. [5] introduced a method combining blind image restoration and recognition to address facial recognition issues in low-quality images through sparse representation. Meanwhile, Nishiyama et al. employed predefined blur kernels to restore blurry faces, thus enhancing its recognition performance. With the unprecedented success of Deep Neural Networks (DNNs) in various image restoration tasks such as denoising, deblurring, restoration, and image super-resolution [6–10], numerous DNN-based facial image restoration methods have emerged and significantly advanced the development of traditional approaches.

Li, Xuanchen et al. [11] proposed the first automatic dynamic facial geometry and texture reconstruction framework based on 3D Gaussian splatting. This framework is capable of directly reconstructing a topologically consistent geometric surface and a dynamic 8K texture map containing pore-level details from multi-view videos. Jia, Haozhe et al. [12] focused on exploring and learning the disentangled control of high-dimensional facial semantic information and explicit 3DMM (3D Morphable Model) parameters in diffusion-based generative facial image editing. Without the need for additional data, it can achieve explicit editing and driving of the pose, expression, and lighting of facial images, and effectively retain the original ID information of the person. However, current 3D methods lack datasets in the field of micro-expressions and cannot ensure real-time performance. At the same time, when the quality of micro-expression images is degraded, it further increases the difficulty of 3D reconstruction.

However, current facial recognition heavily relies on precise prior knowledge learned from stable scenes. Yet, the significant disparity between real-world scenarios and stable environments often leads to severe model performance degradation and erroneous predictions. Facial image restoration methods [13,14] can be expanded to micro-expression recognition by considering the specific structural characteristics of facial images. The current challenges in micro-expression recognition can be addressed by restoring severely degraded micro-expression images into clear and distinct ones without any knowledge of model degradation, which is a crucial and imperative issue to address.

2.2. Micro-Expression Recognition

Recent research on micro-expression detection has predominantly utilized methods that explicitly represent image motion. For instance, optical flow techniques [15–18] capture motion information in image sequences and feed it into deep CNNs and other classifiers for micro-expression recognition. This approach has yielded top-tier results in expression detection. However, the computational complexity of the above methods is relatively high due to the extensive calculations for feature extraction through optical flow methods and computation within deep CNN models. Ref. [19] proposed a novel solution by introducing a shallow optical flow CNN model to predict the likelihood scores of micro-expressions from optical flow to address these drawbacks. Their method reduced the time complexity of neural networks but failed to effectively reduce the time complexity of optical flow computation. In other related studies, scholars have employed motion amplification in preprocessing to enhance micro-expression saliency. For instance, ref. [20] initially applied Eulerian motion magnification (EMM) to amplify motion, followed by optical flow computation, which was input into a graph attention network. However, this approach exhibits redundancy as the motion information undergoes two extraction processes: first through EMM algorithm amplification and then through optical flow algorithm measurement.

The studies above focus on computational complexity and enhancing micro-expression saliency but overlook a crucial issue: compromised image quality scenarios due to motion blur or image compression [21,22]. In such cases, introducing motion information via optical flow reconstruction may induce additional noise that adversely affects the quality of features, consequently impacting task performance. Therefore, novel approaches are needed to recognize micro-expressions in complex real-world environments.

2.3. Generative Model

Researchers have proposed several generative model-based approaches for better object motion capture. For instance, Monkey-Net, FOMM, and MRAA [23–25] encode the motion information of key points or regions in videos through self-supervised learning, and their performance relies on sample diversity. Considering the subtle and challenging nature [26] of micro-expression motion and the inability to directly employ such modeling features for classification tasks, a prior knowledge module is imperative for reconstructing micro-expressions compatible with recognition tasks.

Researchers have proposed methods for transferring prior knowledge of Generative Adversarial Networks (GANs). For instance, [27] applied GANs to image generation using the domain adaptation techniques, achieving the knowledge transfer of the generative model by employing a knowledge mining network. Ref. [28] also introduced a novel approach based on the GAN architecture for transfer learning. These studies successfully transferred knowledge from the original domain to a closely related target domain.

This study embeds prior knowledge for micro-expression generation into the micro-expression reconstruction process adapted to real-world degraded images [9]. Subsequently, the embedded reconstruction features are jointly fine-tuned to facilitate the further integration of cross-task features. To our knowledge, this is the first framework to address micro-expression recognition in real-world degraded scenarios.

3. Materials and Methods

3.1. Materials

The continuous evolution of artificial intelligence has brought about significant breakthroughs in micro-expression recognition. However, micro-expression identification in images with a compromised quality remains challenging. A definitive solution is needed to address image degradation and loss as this task falls under a typical ill-posed inverse problem. Let (X) represent the space of degraded low-quality (LQ) facial images, and (Y) denote the space of original high-quality (HQ) facial images. Given an input LQ facial image $(x \in X)$, current approaches based on DNNs primarily focus on learning a mapping function (Φ) to directly predict micro-expressions, i.e., $(\Phi(x) \rightarrow y)$.

However, DNNs face challenges predicting micro-expressions in images lacking low-frequency information such as facial structures, textures, and identity cues. Although image super-resolution algorithms can reconstruct low-quality micro-expression images, they only yield marginal improvements in image quality. One contributing factor to this phenomenon is that super-resolution algorithms are utilized solely for recovering partial high-frequency information from low-quality images, failing to reconstruct low-frequency details. Another factor is that directly feeding output images from super-resolution algorithms into micro-expression recognition networks leads to the loss of information regarding feature space reconstruction. Consequently, recognition networks struggle to model and comprehend the images accurately. Therefore, a comprehensive approach is needed to effectively tackle image quality degradation, particularly in micro-expression recognition tasks.

Inspired by conditional image generation, this study proposes a Micro-Expression Generator Module (MEGAN) equipped with cross-task transfer fine-tuning instead of utilizing optical flow and leveraging additional information like previous approaches. MEGAN is a decoder for training micro-expression generation models, enabling the network to generate micro-expression images through ($G(z) \rightarrow y$)

$$Face = Generator(z) \tag{1}$$

Subsequently, the encoder of a CNN is integrated into the model to allow the decoder to accept image features. The encoder learns to map the inputted degraded image (x) to the expected latent encoding (z) in the latent representation space (Z) of micro-expressions.

$$Latent = Network(LQ) \tag{2}$$

Then, the micro-expression prior network reproduces the expected high-quality micro-expression images via ($G(z) \rightarrow y$). Finally, the reconstructed micro-expression images are fed into a classification network. To further integrate the restored high-frequency information and the features of the classification network, a Multi-scale Cross-Domain Block is proposed for fusing information from different task domains, thereby enhancing the fusion of facial high-frequency information extracted by the restoration model for subsequent classification network information.

$$Face_latent = Encoder(LQ) + Latent \tag{3}$$

$$CLS = CLS_Network(Decoder(Face_latent)) \tag{4}$$

CLS_Network refers to any classification network, and *CLS* stands for the classification task. Our method can be integrated into any classification network for learning, enabling the model to improve its performance.

3.2. Network Architecture

3.2.1. Micro-Expression Prior Generation Model

The numerous applications of U-Net in image restoration tasks demonstrate its effectiveness in preserving image details. Therefore, this study initially adopts the U-Net architecture to reconstruct the micro-expression space. Drawing inspiration from advanced StyleGAN and ConditionGAN, a micro-expression generation model combining latent encoding (z) with specified categories is constructed first, projecting them into a less entangled space ($w \in W$). This allows the proposed model to model arbitrary micro-expression feature spaces. Unlike traditional conditional GAN, this study encodes micro-expression information into the latent space in advance to enable subsequent generation networks to perceive category information and transfer it to other tasks, as illustrated in Figure 1a. Subsequently, the intermediate features characterizing the generation process are broadcasted to each MEGAN block. Meanwhile, the feature mapper of the proposed MEGAN blocks aggregates the generated features with reserved feature map regions, thereby allowing features of the generation task to propagate through skip connections.

To train our model, we adopt three loss functions: the feature matching loss L_A, the adversarial loss L_A, and the content loss L_C.

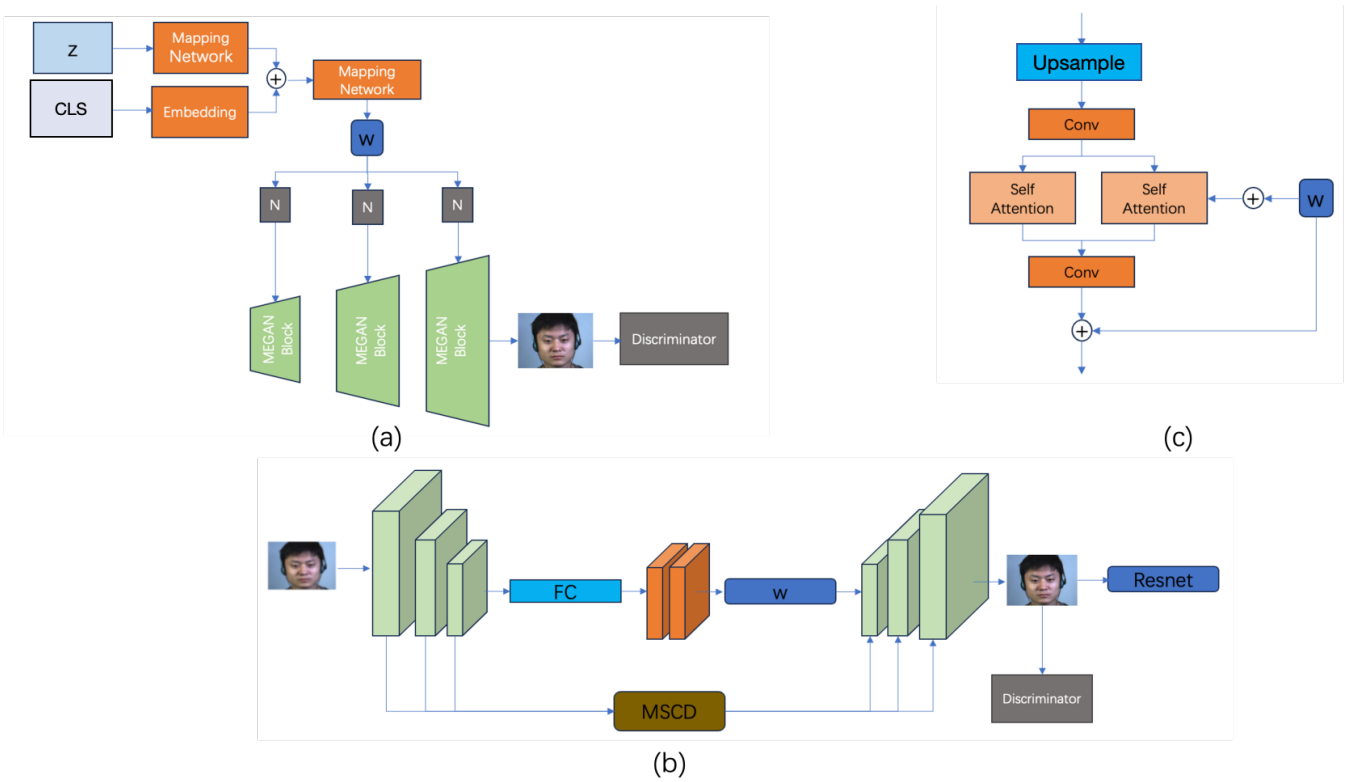

Figure 1. The PRnet architecture includes (**a**) the conditional micro-expression generation model, (**b**) the full network architecture of PRnet, and (**c**) the detailed structures of a MEGAN block. The proposed network first utilizes the conditional micro-expression generation model to generate accurate micro-expression images. Then, the MEGAN block, capable of modeling micro-expressions, is incorporated into the image restoration task. The restored micro-expression images are then fed into subsequent classification models.

L_F is similar to the perceptual loss but it is based on the discriminator rather than the pre-trained VGG network to fit our task. It is formulated as follows:

$$L_F = \min_G E_{(X)} \left(\sum_{i=0}^{T} \left\| D^i(X) - D^i(G(\tilde{X})) \right\|_2 \right) \tag{5}$$

where T is the total number of intermediate layers used for feature extraction. $D^i(X)$ is the extracted feature at the i-th layer of discriminator D.

L_A is inherited from the GAN prior network where X and $\tilde{X}$ denote the ground-truth HQ image and the degraded LQ one, G is the generator during training, and D is the discriminator. L_C is defined as the L1-norm distance between the final results of the generator and the corresponding ground-truth images.

$$L_A = \min_G \max_D E_{(X)} \log(1 + \exp(-D(G(\tilde{X})))) \tag{6}$$

The final loss L is as follows:

$$L = L_F + L_A + L_B \tag{7}$$

The content loss L_C enforces the fine features and preserves the original color information. By introducing the feature matching loss L_F on the discriminator, the adversarial loss L_A can be better balanced to recover more realistic face images with vivid details.

3.2.2. Multi-Scale Cross-Domain Information Bridging

Skip connections are widely employed to enhance the effective transfer of features in various applications. This approach establishes additional feature transmission pathways by connecting one layer's output to the subsequent layers' input. Such a design enables skipping specific layers in feature propagation, allowing the network to learn more discriminative features, thereby improving the model performance. However, using skip connections in current tasks faces two issues. Firstly, each reconstruction layer can only access single-scale encoded features. Secondly, the reconstructed features cannot effectively transfer to the classification task due to inter-task conflicts.

$$F_{ms} = \sum_{i=1}^{N} w * MLP(F) + (1 - w) * CBAM(F) \tag{8}$$

To address these issues, a novel Multi-Scale Cross-Domain Information Bridging (MCI-Bridge) mechanism is proposed. Specifically, the reconstructed features of different scales are first aggregated to a unified size. Subsequently, the features are mapped onto five scales utilizing a self-attention mechanism, dynamically allocating them for different tasks. This mechanism is illustrated in Figure 1b. This approach enhances the reconstruction quality while effectively integrating the reconstructed features into arbitrary classification networks.

3.2.3. Overview Pipeline

Furthermore, our model's training and inference are facilitated by Algorithm 1, a widely recognized facial micro-expression dataset, which was selected to train a pre-existing network for reconstructing facial micro-expressions using degradation techniques provided by Real-ESRGAN. As the decoder, this network is embedded into a U-Net DNN. Subsequently, the latent code and noise introduced by the GAN network are merged in the MSCD module and forwarded to the decoding layers. This process governs the global facial structure, local micro-expression details, and the global structural representation of background reconstruction. Finally, the pre-existing network and the above MCI-Bridge mechanism are input into an arbitrary classification network for fine-tuning. Doing so facilitates mutual adaptation learning between the encoding part of the reconstruction, the decoder, and micro-expression recognition tasks. The specific pipeline is illustrated in Figure 1b.

Algorithm 1 Algorithm of PRNet

Input: Image
Output: High-quality image
if Training **then**
  $LQ = ((Image \otimes k) \downarrow_s + n_\sigma)_{JPEG_q}$
  $W_{face} = FC(UNET_{Encoder}(LQ))$
  $MSCD_{feature} = UNET_{Encoder}(LQ)$
  $HQ = Decoder(W_{face}, MSCD_{feature})$
else
  $HQ = PRNet(Image)$
end if
return HQ

Our workflow is shown in Figure 2. Our algorithm is simplified into three steps. In the first step, the main goal is to enable the encoder of the generation model to fully learn the feature extraction ability of the micro-expressions. Subsequently, in the second step, we define a reconstruction model that allows the model to accept low-quality images. By using the

trained micro-expression feature extraction model and the decoder, it can output reconstructed and clear micro-expression images. In the third step, we concatenate an arbitrary classification model so that the classification model can easily learn from the micro-expression images instead of low-quality images, thus reducing the learning burden of the model.

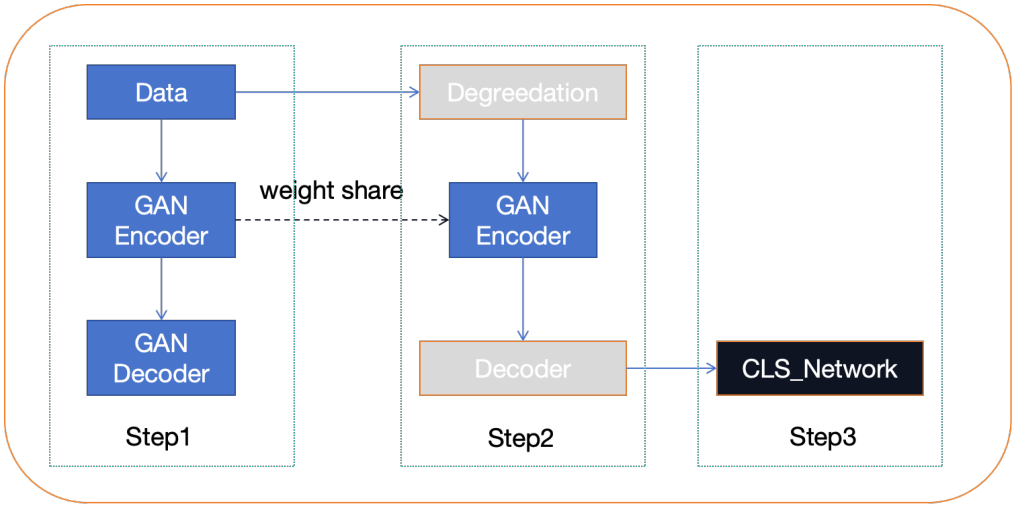

Figure 2. The workflow implemented by our algorithm.

4. Experiment and Results

4.1. Dataset

This study selects the SAMM [29], SMIC [30], CASME II [31], and CASME III [32] datasets for experimentation. SAMM, SMIC, and CASME II are merged into a comprehensive dataset with the same emotion labels for micro-expression recognition tasks to ensure consistency and comparability. Emotion categories are divided as follows: the "positive" emotion category includes "happiness", the "negative" emotion category includes "sadness", "disgust", "contempt", "fear", and "anger", while the "surprise" emotion category includes "surprise". Table 1 summarizes the attributes of these datasets.

Table 1. The experiments are implemented on SAMM [29], SMIC [30], CASME II [31], and CASME III [32] databases. SAMM, SMIC, and CASME II are merged into one composite dataset, and the same labels in these three datasets are adopted for micro-expression tasks.

Database	SAMM	CASME II	SMIC	CASME III
Subjects	28	26	16	100
Samples	133	145	164	943
Frame rate	200	200	100	30
Cropped image resolution	224 × 224	224 × 224	224 × 224	224 × 224
Negative	92	88	70	508
Positive	26	32	51	64
Surprise	15	25	43	201
Onset index	✓	✓	✓	✓
Offset index	✓	✓	✓	✓
Apex index	✓	✓		✓

The experiments in this study are primarily divided into three evaluation tasks. First, the performance of the proposed micro-expression reconstruction method is assessed on the original dataset, followed by its recognition performance in degraded scenarios. In order to demonstrate the effectiveness of our micro-expression reconstruction method, the

effectiveness of the proposed reconstruction model is evaluated by comparing it to the latest reconstruction models.

Further information on the pre-processed dataset: The SAMM dataset comprises 28 participants, with 133 micro-expressions and 147 long videos containing 343 macro-expressions. This dataset is rich in action unit encoding, providing comprehensive facial expression information. SAMM also offers the onset, apex, and offset indices of micro-expressions. The original samples in the dataset have a resolution of 2040 $\times$ 1088 pixels, with a frame rate set at 200 fps. Emotion categories of images in SAMM include "disgust", "fear", "contempt", "anger", "repression", "surprise", "happiness", and "others". After categorizing into the above three emotion categories, the quantities for "negative", "positive", and "surprise" are 92, 26, and 15, respectively.

The CASME II dataset comprises data from 24 subjects, totaling 145 samples corresponding to 145 emotional states. All its samples are captured using laboratory cameras, with a frame rate of 200 fps and a size of 640 $\times$ 480 pixels. Samples in CASME II are categorized into "happiness", "surprise", "disgust", "sadness", "fear", "repression", and "others". After merging into the above three emotion categories, the quantities for "negative", "positive", and "surprise" are 88, 32, and 25, respectively. Micro-expression onset, apex, and offset indices are annotated in CASME II.

CASME III represents the third generation of facial micro-expression databases, distinguished by its inclusion of depth information and high ecological validity, rendering it highly valuable for micro-expression recognition. The first part of CASME III comprises data from 100 subjects, totaling 943 samples corresponding to 943 emotional states. Samples are captured using laboratory cameras, with a frame rate of 30 fps and an original resolution of 1280 $\times$ 720 pixels. Samples in the first part of CASME III are categorized into "happiness", "anger", "fear", "disgust", "surprise", "others", and "sadness". The total counts for "negative", "positive", and "surprise" are 508, 64, and 201, respectively.

The SMIC (SMIC-HS) dataset comprises data from 16 subjects, totaling 164 samples corresponding to 164 emotional states. All samples are captured using laboratory cameras, with a frame rate of 100 fps and an original size of 640 $\times$ 480 pixels. SMIC samples are categorized as "negative", "surprise", and "positive". The quantities for "negative", "positive", and "surprise" are 70, 51, and 43, respectively. While onset and offset indices are provided in SMIC, peak indices are not included. Detailed information regarding these three datasets can be found in Table 2.

In this paper, we mainly discuss the impact of real-world degradation. Since micro-expression recognition is a complex task, and the loss of image quality will further exacerbate the loss of image information, making effective recognition impossible. To solve this problem, in the experimental design, first, to ensure fairness, we divided the dataset into an 8:2 ratio and guaranteed the balance of subjects and the data allocated for testing to avoid data leakage. In addition, we further expanded the test set, enhancing it in a 1:10 ratio using Formula (9) to ensure fairness. The specific categories of the test set are detailed in Table 2. Meanwhile, with a limited dataset, we can further expand the dataset by using the degradation function, which can ensure the diversity of data during training and enhance the robustness of the model.

In this paper, for the sake of fairness, we divided the dataset into a 8:2 ratio and ensured the balance of subjects and the data allocated for testing to avoid data leakage. Additionally, we further augmented the test set, enhancing it in a 1:10 ratio using Formula (1) to ensure fairness. Specific test set categories are detailed in Table 2.

Table 2. The unweighted F1-score (UF1) and unweighted average recall (UAR) performance of sota methods and our PRNet on SMIC [30], CASME II [31], CASME III [32], and SAMM [29]. Bold text indicates the best result.

Approaches	**SMIC**		**CASME II**		**CASME III**		**SAMM**	
	UF1	**UAR**	**UF1**	**UAR**	**UF1**	**UAR**	**UF1**	**UAR**
AlexNet [33]	0.6201	0.6373	0.7994	0.8312	0.2570	0.2634	0.6104	0.6642
GoogLeNet [34]	0.5123	0.5511	0.5989	0.6414	0.2658	0.2713	0.5124	0.5992
VGG16 [35]	0.5800	0.5964	0.8166	0.8202	0.3209	0.3400	0.4870	0.4793
Resnet50 [36]	0.7251	0.7615	0.8249	0.8556	0.3491	0.3651	0.7260	0.7435
STSTNet [37]	0.6801	0.7013	0.8382	0.8686	0.3795	0.3792	0.6588	0.6810
CapsuleNet [38]	0.5820	0.5877	0.7068	0.7018	0.2478	0.2516	0.6209	0.5989
Dual-Inception [39]	0.6645	0.6726	0.8621	0.8560	0.3844	0.4001	0.5868	0.5663
RCN [40]	0.6326	0.6441	0.8621	0.8512	0.3928	0.3893	0.7601	0.6715
FeatRef [41]	0.7011	0.7083	0.8915	0.8873	0.34938	0.3413	0.7372	0.7155
HTNet [42]	0.8049	0.7905	0.9532	0.9516	0.5767	0.5415	0.8131	0.8124
PR-AlexNet	0.6654	0.6691	0.8518	0.8487	0.3159	0.3209	0.6227	0.6542
PR-GoogLeNet	0.6078	0.6128	0.6499	0.7011	0.3278	0.3317	0.5745	0.6218
PR-VGG16	0.6476	0.6550	0.8731	0.8639	0.3815	0.3843	0.6854	0.7032
PR-Resnet50	0.8257	0.8082	0.9625	0.9516	0.5892	0.5760	0.8328	0.8345

4.2. Experiment Details

The experiments in this study utilize Dlib for facial keypoint detection and subsequent image cropping based on these keypoints. All cropped images across datasets are resized to 224 × 224 pixels. The method in Real-ESRGAN is adopted to degrade the images, and the corresponding data are collected. Equation (9) represents the input facial image, where k denotes the blur kernel, $n\sigma$ signifies Gaussian noise with a standard deviation of σ, and Id represents the degraded image. $\otimes$ denotes 2D convolution, $\downarrow s$ denotes standard s-fold downsampling, and JPEG_q represents the JPEG compression operator with a quality factor q.

$$I^d = ((I \otimes k) \downarrow_s + n_\sigma)_{JPEG_q} \tag{9}$$

Numerous existing studies have demonstrated the effectiveness of the above degradation models in simulating various real-world degradation scenarios [9]. In our implementation, the blur kernel (k) for each image is randomly selected from a set of blurring models, including Gaussian blur and motion blur with varying kernel sizes. The additive Gaussian noise (σ) is sampled channel-wise from a normal distribution, and (σ) is selected from 0 to 25. The value of (s) is randomly and uniformly sampled from 10 to 200 (i.e., up to 200 times downscaling), and (q) is randomly and uniformly sampled from 5 to 50 (i.e., up to 95% JPEG compression) per image. Then, the proposed model is divided into two stages. First, a micro-expression generation model is trained to generate data in the same distribution as the datasets, thus deepening our understanding of image representations. Afterwards, the pre-trained generation module is embedded into the reconstruction and classification models for further training. AdamW is employed during training to optimize the network with a learning rate initialized to 0.001. At every 30th, 60th, and 90th epoch, the learning rate is set to 0.0001, 0.00001, and 0.000001, respectively. The model is trained using the cross-entropy loss function.

5. Ablation Experiment

5.1. Standard Experimental Comparison

The comparison results in Table 2 show that the proposed method achieves the best results across various evaluation metrics. Thus, our method can better reconstruct micro-

expressions, thereby assisting in classification tasks, without relying on additional information or complex network designs.

5.2. Degraded Micro-Expression Recognition

This section discusses the micro-expression recognition task in images degraded to low quality based on the above equations. The ablation experiments are conducted in three parts.

According to Table 3, a significant amount of high-frequency information is lost when the image is degraded into a low-quality one, which affects the extraction of features such as micro-expressions and textures. Existing methods have been greatly affected by directly learning from low-quality images, while PRNet only incurs a minimal loss. These experimental data indicate that PRNet can maintain high competitiveness in addressing image quality loss in the real world. Meanwhile, the proposed algorithm applies to and improves the performance of any classification model.

Table 3. The unweighted F1-score (UF1) and unweighted average recall (UAR) performance of sota methods and our PRNet on the degraded SMIC [30], CASME II [31], CASME III [32], and SAMM [29]. Bold text indicates the best result.

Approaches	**SMIC**		**CASME II**		**CASME III**		**SAMM**	
	UF1	**UAR**	**UF1**	**UAR**	**UF1**	**UAR**	**UF1**	**UAR**
AlexNet [33]	0.2340	0.3059	0.4496	0.4787	0.0837	0.1373	0.2242	0.3585
GoogLeNet [34]	0.1962	0.3305	0.3594	0.3948	0.1049	0.1528	0.2150	0.3595
VGG16 [35]	0.2350	0.3582	0.4699	0.4821	0.1083	0.1940	0.1858	0.3876
Resnet50 [36]	0.3125	0.4568	0.4940	0.5314	0.1096	0.2190	0.2904	0.4457
STSTNet [37]	0.2900	0.4208	0.4171	0.5011	0.1493	0.1746	0.2799	0.4084
CapsuleNet [38]	0.2410	0.3585	0.3387	0.4307	0.0986	0.1506	0.2484	0.3395
Dual-Inception [39]	0.2823	0.4028	0.3552	0.5136	0.1434	0.2200	0.2347	0.2265
RCN [40]	0.2663	0.4021	0.3448	0.5101	0.1864	0.1946	0.3300	0.3358
FeatRef [41]	0.3006	0.4342	0.4457	0.4910	0.1247	0.1707	0.3189	0.3577
HTNet [42]	0.3524	0.4752	0.4766	0.5758	0.2883	0.2708	0.4065	0.4062
PR-AlexNet	0.6432	0.63781	0.8108	0.8211	0.2942	0.3001	0.6135	0.6158
PR-GoogLeNet	0.5902	0.6005	0.6172	0.6716	0.3018	0.3100	0.5471	0.5822
PR-VGG16	0.6342	0.6307	0.8555	0.8389	0.3643	0.3622	0.6611	0.6897
PR-Resnet50	0.8176	0.7937	0.9542	0.9479	0.5690	0.5689	0.8213	0.8235

In the ablation experiment of the MSCD module, our bridging mechanism can effectively transfer the reconstructed features to the subsequent classification model so that the model not only receives a single output of the reconstructed image (Table 2).

For micro-expression recognition, this study synthesizes LQ facial images on multiple datasets for evaluation using the degradation model in Equation (9) and the same dataset as in Section 4.2. The proposed PRNet is compared with several state-of-the-art reconstruction algorithms, and the PSNR and FID results are listed in Table 4. Compared with other competitive methods, the proposed method in this study achieved significantly better results in PSNR and performed well in terms of FID and LPIPS indicators. The proposal of this task also contributes to the advancement of micro-expression recognition towards realistic degradation.

Table 4. PSNR, FID, and LPIPS comparison of different micro-expression restoration methods on the degraded SMIC [30], CASME II [31], CASME III [32], and SAMM [29]. Bold text indicates the best result.

Approaches	SMIC		CASME II		CASME III		SAMM	
	$PSNR_{\uparrow}$	$FID_{\downarrow}$	$PSNR_{\uparrow}$	$FID_{\downarrow}$	$PSNR_{\uparrow}$	$FID_{\downarrow}$	$PSNR_{\uparrow}$	$FID_{\downarrow}$
Pix2PixHD [43]	19.61	73.97	20.09	80.22	20.18	80.20	20.65	79.40
Super-FAN [44]	20.67	132.11	21.21	142.27	21.10	140.26	22.34	133.92
HiFaceGAN [45]	20.51	54.45	21.01	59.19	20.68	59.21	22.18	60.67
PRNet (ours)	**23.12**	**29.72**	**23.28**	**30.34**	**23.62**	**30.02**	**23.94**	**31.34**

5.3. Real-World Face Recognition

Figure 3 demonstrates the superiority of our method. The first and third rows contain the low-quality images we collected, while the second and fourth rows show the reconstruction effects of the model. Our method can effectively restore details and add more natural colors, and the results are significantly better than the low-quality input images.



Figure 3. Comparison diagrams of real-world face reconstruction. The first and third rows represent the low-quality input images, while the second and fourth rows represent the reconstructed images.

6. Discussion and Limitations

This study proposes a prior-embedded network, PRNet, for real-world blind micro-expression recognition, which addresses the ineffective micro-expression recognition on low-quality images in the real world. This research is essential to better understand emotions, cognition, and psychological processes in the natural world.

The proposed network is mainly divided into three steps. First, a generative network is designed to generate micro-expression images using the micro-expression category control. Then, the trained generation module is embedded into the micro-expression restoration model. Finally, the restoration model is integrated into the classification model, enabling it to classify degraded micro-expression images in the real world. Additionally, this study proposes an MSCD for dynamically adjusting the features for different tasks in the reconstruction process to further enhance the utilization of restoration features.

Experimental results demonstrate that PRNet consistently outperforms the state-of-the-art methods on four micro-expression benchmark datasets, including SAMM, CASME II, SMIC, and CASME III. Moreover, the proposed method maintains good performance in classifying severely degraded images. Additionally, PRNet can be easily transferred to other methods, contributing to the advancement of micro-expression recognition and classification in the real world.

This paper mainly discusses the degradation issue of micro-expression recognition in the real world. Current challenges include defining micro-expressions and assessing their degradation degrees in different scenarios. This study primarily adopts recognized methods that simulate real-world degradation to address micro-expression degradation in real-world settings. Although the methods above have achieved particular effectiveness, they still cannot accurately capture micro-expression distribution in practical scenarios, such as in corporate micro-expression management or micro-expression emotional analysis based on interviews. Therefore, generating micro-expression images more inclined toward real and complex scenarios will be of significant research value in future studies to promote the development of these specific fields.

This study proposes a prior-embedded network, PRNet, for blind micro-expression recognition in the real world, aiming to address the issue of ineffective micro-expression recognition on low-quality images in reality. This research is of great significance for a better understanding of emotions, cognition, and psychological processes in the natural world. Specifically, the network proposed in this paper is mainly divided into three steps. Firstly, a generative task is utilized to obtain an encoder capable of extracting features. Secondly, an image reconstruction task is designed to reconstruct low-quality micro-expression images. Thirdly, the reconstructed images are fed into a classification network for learning. Furthermore, in order to promote the fusion of features between the reconstruction model and the classification model, this paper proposes a Multi-Scale Cross-Domain module (MSCD), which is used to dynamically adjust features for different tasks during the reconstruction process, so as to further improve the utilization rate of the restored features.

Experimental results demonstrate that PRNet consistently outperforms the current state-of-the-art methods on four micro-expression benchmark datasets, including SAMM, CASME II, SMIC, and CASME III. Moreover, this method can maintain good performance when classifying severely degraded images. Additionally, PRNet can be easily transferred to other methods, contributing to the advancement of micro-expression recognition and classification technologies in the real world.

Current research algorithms mainly focus on fixed datasets, while this paper is dedicated to simulating the degradation in the real world. This study primarily adopts the recognized methods for simulating real-world degradation to address the issue of micro-expression degradation in real-world scenarios. Although the above-mentioned methods have achieved certain effectiveness, they still cannot accurately capture the distribution of micro-expressions in practical situations, such as in corporate micro-expression management or micro-expression emotional analysis based on interviews. At the same time, existing methods still tend to use larger models to improve accuracy, which fails to ensure

real-time performance and further restricts the application of micro-expression recognition algorithms. In the future, this paper believes that, on the one hand, the use of generative models can help improve data quality and simulate the data distribution of various demand scenarios. Meanwhile, strengthening the research on lightweight models, such as methods like knowledge distillation, can further promote the implementation and practicality of this direction and drive the development of the field.

Author Contributions: Conceptualization, X.L. and J.C.; Methodology, J.C.; Software, J.C.; Validation, X.L. and J.C.; Formal analysis, J.C.; Investigation, J.C.; Resources, X.L.; Data curation, H.Z. and J.C.; Writing—original draft, X.L., F.W., H.Z., Y.C., L.Z. and J.C.; Writing—review and editing, X.L., F.W., H.Z., Y.C., L.Z. and J.C.; Visualization, J.C.; Supervision, X.L. and J.C.; Project administration, J.C.; Funding acquisition, J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The datasets presented in this article are not readily available because institutions providing datasets require users to submit an application for use using the email address of the academic institution, and the application must indicate the specific details of the intended use. Requests to access the datasets should be directed to dfuc.mmu@amail.com.

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

References

1. Zhang, F.; Zhang, T.; Mao, Q.; Xu, C. Joint pose and expression modeling for facial expression recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 3359–3368.
2. Ekman, P.; Friesen, W.V. Nonverbal leakage and clues to deception. *Psychiatry* **1969**, *32*, 88–106. [CrossRef]
3. Ekman, P. *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage (Revised Edition)*; WW Norton & Company: New York, NY, USA, 2009.
4. Li, G.; Shi, J.; Peng, J.; Zhao, G. Micro-expression recognition under low-resolution cases. In Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications-Volume 5: VISAPP, Prague, Czech Republic, 25–27 February 2019; Science and Technology Publications: Setúbal, Portugal, 2019; pp. 427–434.
5. Zhang, H.; Yang, J.; Zhang, Y.; Nasrabadi, N.M.; Huang, T.S. Close the loop: Joint blind image restoration and recognition with sparse representation prior. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 770–777.
6. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Change Loy, C. Esrgan: Enhanced super-resolution generative adversarial networks. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, 8–14 September 2018; pp. 63–79.
7. Ma, C.; Jiang, Z.; Rao, Y.; Lu, J.; Zhou, J. Deep face super-resolution with iterative collaboration between attentive recovery and landmark estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 5569–5578.
8. Chan, K.C.; Wang, X.; Xu, X.; Gu, J.; Loy, C.C. Glean: Generative latent bank for large-factor image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual Conference, 19–25 June 2021; pp. 14245–14254.
9. Wang, X.; Xie, L.; Dong, C.; Shan, Y. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Virtual Conference, 11–17 October 2021; pp. 1905–1914.
10. Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; Fu, Y. Residual dense network for image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 2472–2481.
11. Li, X.; Cheng, Y.; Ren, X.; Jia, H.; Xu, D.; Zhu, W.; Yan, Y. Topo4D: Topology-Preserving Gaussian Splatting for High-fidelity 4D Head Capture. In Proceedings of the European Conference on Computer Vision, Milan, Italy, 29 September–4 October 2024; Springer: Berlin/Heidelberg, Germany; pp. 128–145.
12. Jia, H.; Li, Y.; Cui, H.; Xu, D.; Wang, Y.; Yu, T. DisControlFace: Adding Disentangled Control to Diffusion Autoencoder for One-shot Explicit Facial Image Editing. *arXiv* **2023**, arXiv:2312.06193.
13. Tu, X.; Zhao, J.; Liu, Q.; Ai, W.; Guo, G.; Li, Z.; Liu, W.; Feng, J. Joint face image restoration and frontalization for recognition. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *32*, 1285–1298. [CrossRef]

14. Liao, Y.; Lin, X. Blind image restoration with eigen-face subspace. *IEEE Trans. Image Process.* **2005**, *14*, 1766–1772. [CrossRef] [PubMed]
15. Brox, T.; Malik, J. Large displacement optical flow: Descriptor matching in variational motion estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *33*, 500–513. [CrossRef]
16. Sun, S.; Kuang, Z.; Sheng, L.; Ouyang, W.; Zhang, W. Optical flow guided feature: A fast and robust motion representation for video action recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 1390–1399.
17. Xu, L.; Jia, J.; Matsushita, Y. Motion detail preserving optical flow estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *34*, 1744–1757.
18. Peng, M.; Wang, C.; Chen, T.; Liu, G.; Fu, X. Dual temporal scale convolutional neural network for micro-expression recognition. *Front. Psychol.* **2017**, *8*, 273835. [CrossRef]
19. Liong, G.B.; See, J.; Wong, L.K. Shallow optical flow three-stream CNN for macro-and micro-expression spotting from long videos. In Proceedings of the 2021 IEEE International Conference on Image Processing (ICIP), Anchorage, AK, USA, 19–22 September 2021; pp. 2643–2647.
20. Kumar, V.; Durst, F.; Ray, S. Modeling moving-boundary problems of solidification and melting adopting an arbitrary Lagrangian–Eulerian approach. *Numer. Heat Transf. Part B Fundam.* **2006**, *49*, 299–331. [CrossRef]
21. Kaneko, T.; Harada, T. Blur, noise, and compression robust generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual Conference, 19–25 June 2021; pp. 13579–13589.
22. Oh, T.; Lee, S. Blind sharpness prediction based on image-based motion blur analysis. *IEEE Trans. Broadcast.* **2015**, *61*, 1–15.
23. Siarohin, A.; Lathuilière, S.; Tulyakov, S.; Ricci, E.; Sebe, N. First order motion model for image animation. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 641.
24. Siarohin, A.; Lathuilière, S.; Tulyakov, S.; Ricci, E.; Sebe, N. Animating arbitrary objects via deep motion transfer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2377–2386.
25. Siarohin, A.; Woodford, O.J.; Ren, J.; Chai, M.; Tulyakov, S. Motion representations for articulated animation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual Conference, 19–25 June 2021; pp. 13653–13662.
26. Lu, H.; Kpalma, K.; Ronsin, J. Motion descriptors for micro-expression recognition. *Signal Process. Image Commun.* **2018**, *67*, 108–117. [CrossRef]
27. Wang, Y.; Wu, C.; Herranz, L.; Van de Weijer, J.; Gonzalez-Garcia, A.; Raducanu, B. Transferring gans: Generating images from limited data. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 218–234.
28. Frégier, Y.; Gouray, J.B. Mind2mind: Transfer learning for gans. In Proceedings of the Geometric Science of Information: 5th International Conference, GSI 2021, Paris, France, 21–23 July 2021; Springer: Berlin/Heidelberg, Germany, 2021; pp. 851–859.
29. Davison, A.K.; Lansley, C.; Costen, N.; Tan, K.; Yap, M.H. Samm: A spontaneous micro-facial movement dataset. *IEEE Trans. Affect. Comput.* **2016**, *9*, 116–129. [CrossRef]
30. Li, X.; Pfister, T.; Huang, X.; Zhao, G.; Pietikäinen, M. A spontaneous micro-expression database: Inducement, collection and baseline. In Proceedings of the 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (fg), Shanghai, China, 22–26 April 2013; pp. 1–6.
31. Yan, W.J.; Li, X.; Wang, S.J.; Zhao, G.; Liu, Y.J.; Chen, Y.H.; Fu, X. CASME II: An improved spontaneous micro-expression database and the baseline evaluation. *PLoS ONE* **2014**, *9*, e86041. [CrossRef] [PubMed]
32. Li, J.; Dong, Z.; Lu, S.; Wang, S.J.; Yan, W.J.; Ma, Y.; Liu, Y.; Huang, C.; Fu, X. CAS (ME) 3: A third generation facial spontaneous micro-expression database with depth information and high ecological validity. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 2782–2800. [CrossRef]
33. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [CrossRef]
34. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
35. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
36. 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, 26 June–1 July 2016; pp. 770–778.
37. Liong, S.T.; Gan, Y.S.; See, J.; Khor, H.Q.; Huang, Y.C. Shallow triple stream three-dimensional cnn (ststnet) for micro-expression recognition. In Proceedings of the 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019), Lille, France, 14–18 May 2019; pp. 1–5.

38. Van Quang, N.; Chun, J.; Tokuyama, T. CapsuleNet for micro-expression recognition. In Proceedings of the 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019), Lille, France, 14–18 May 2019; pp. 1–7.
39. Zhou, L.; Mao, Q.; Xue, L. Dual-inception network for cross-database micro-expression recognition. In Proceedings of the 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019), Lille, France, 14–18 May 2019; pp. 1–5.
40. Xia, Z.; Peng, W.; Khor, H.Q.; Feng, X.; Zhao, G. Revealing the invisible with model and data shrinking for composite-database micro-expression recognition. *IEEE Trans. Image Process.* **2020**, *29*, 8590–8605. [CrossRef]
41. Zhou, L.; Mao, Q.; Huang, X.; Zhang, F.; Zhang, Z. Feature refinement: An expression-specific feature learning and fusion method for micro-expression recognition. *Pattern Recognit.* **2022**, *122*, 108275. [CrossRef]
42. Wang, Z.; Zhang, K.; Luo, W.; Sankaranarayana, R. HTNet for micro-expression recognition. *arXiv* **2023**, arXiv:2307.14637. [CrossRef]
43. Park, T.; Liu, M.Y.; Wang, T.C.; Zhu, J.Y. Semantic image synthesis with spatially-adaptive normalization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2337–2346.
44. Bulat, A.; Tzimiropoulos, G. Super-fan: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with gans. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 109–117.
45. Yang, L.; Wang, S.; Ma, S.; Gao, W.; Liu, C.; Wang, P.; Ren, P. Hifacegan: Face renovation via collaborative suppression and replenishment. In Proceedings of the 28th ACM International Conference on Multimedia, Seattle, WA, USA, 12–16 October 2020; pp. 1551–1560.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.