2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

Microexpression to Macroexpression: Facial Expression Magnification by Single Input

Yaqi Song[1], Tong Chen[2], Shigang Li[3] and Jianfeng Li[4,*]

***Abstract*— Microexpressions are expressions that people inadvertently express, and therefore often represent a person's true emotion. However, because it has a low intensity and a short duration, it is hard to be recognized correctly. In this paper, we propose a deep learning magnification method to generate macroexpressions from a single microexpression image. In the first stage, we extract the expression information from a single microexpression image. Then, We combine the idea of cyclegan and optical flow consistency to model the extracted expression features as the optical flow field between the neutral face and microexpressions. To extract a reliable optical flow field from the expression information, we design an optical flow refiner. In the second stage, we adopt an encoder-decoder network and let it learn to magnify the optical flow. Finally, the magnified optical flow guided the microexpression images to generate macroexpression images. We compare our single input based network with current two-frames-input based networks. The results show that our method performs better, even in wild images. We fed our magnified images directly into a simple ResNet18 network for recognition, achieving a competitive score under the MEGC2019 standard, compared with recent complex recognition networks.**

I. Introduction

In daily life, in addition to macroexpressions(ordinary facial expressions), there is a special type of expression called "microexpressions" (MEs). In the physical sense, microexpressions have a movement trend corresponding to macroexpressions, but this trend is very weak. ME was first discovered by Haggard and IsaacsI [1] in 1966 during a psychotherapy session. Microexpressions usually unconsciously occur and only when an emotion is present, so it often is the most real emotions. Due to its reliability, the analysis of ME has broad application prospects [2], [3], [4] and can be applied in research, psychology, medical treatment, law enforcement, business negotiation and other fields. However, microexpressions are more difficult to recognize than macroexpressions. The duration of microexpressions is extremely short, and the amplitude is very low [5], [6]. In a psychological experiment, Frank et al. [7] reported that even after training, a person has only 45% accuracy in this 5-group classification recognition task. Therefore, a method for microexpression recognition based on computer vision and pattern analysis techniques [8] with high accuracy is urgently required.

[1]Yaqi Song, [2]Tong Chen and [4]Jianfeng Li are with College of Electronic and Information Engineering, Southwest University, China
[3]Shigang Li is with Graduate School of Information sciences, Hiroshima City University, Japan
[1]e-mail: songyq934050131@163.com
[*]e-mail: popqlee@swu.edu.cn, corresponding author

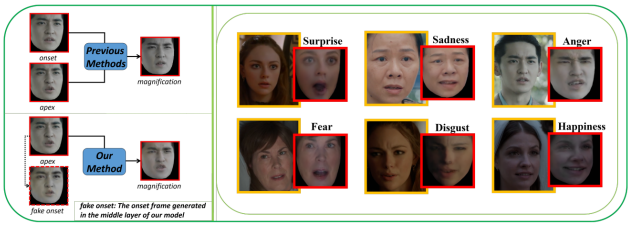

Fig. 1. The left top is the current two-frame-input based magnification method, which requires onset and apex frames. The left down is the proposed single input-based magnification method, which generates the onset in the network. The right is magnified samples in six basic facial expressions based on the proposed method. The image in orange border represents the input, the image in the red border represents the output.

To tackle the problem of low amplitude, a common approach is to enhance the facial features via a method [9], [10]. Tae-Hyun Oh et al. [11] proposed learning-based video motion magnification (LVMM), and Lei et al. [12], [13] proved that the method based on deep learning was conducive to recognition. However, this model relies on the setting of the magnification factor, it is susceptible to face distortion caused by overmagnification. In this case, Song et al. [14] proposed a method recognizing microexpressions as macroexpressions by the Teacher-student Framework Network (MEMM) to magnify the microexpressions. It was evaluated for the first time on wild samples in this field and achieved ideal results. Although MEMM no longer relies on factors, it still needs two aligned inputs: the onset frame and the apex frame. Because the onset frame and apex frame of the face are not easily aligned in the wild samples, the two-frame-input based method limits the application prospect. To address these issues, we propose the first deep learning method for microexpression magnification using a single image. Specific contributions are as follows:

- We propose the first deep learning method to generate macroexpressions from a single microexpression image. Inspired by cyclegan [15], we first extract a reliable facial optical flow field from a single image. Then, we design a magnification network to magnify the optical flow, which is used for macroexpression generation.
- To accurately extract the subtle facial features of microexpressions, we design an optical flow refiner.
- We compare the proposed single input based method with the previous method that takes two frames as input. The results show that our method can obtain an ideal magnification effect on both microexpression datasets and wild samples. According to the MEGC2019 [16] standard, we received a competitive score compared with recent complex recognition networks.

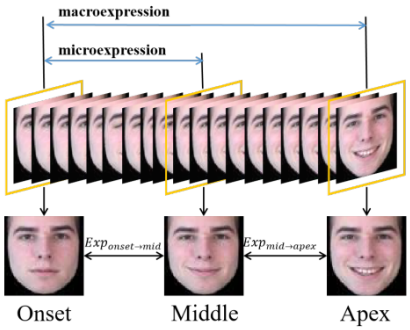

Fig. 2. The relationship between microexpression sequence and macroexpression sequence

II. Related work

A. Microexpression

Research on microexpression is urgent because it shows the true emotions of humans. To date, many recognition methods have been proposed. Zhou et al. [17] proposed a dual-inception model to tackle the Cross-DB challenge of the MEGC 2019 [16]. Wang et al. [18] proposed a novel attention mechanism called microattention to help the network focus on facial areas of interest to recognize microexpressions. A dual-stream deep learning network was designed by Nie et al. [19] to perform multiple tasks of gender recognition and microexpression recognition. Zhou et al. [20] proposes a novel Feature Refinement (FR) with expression-specific feature learning and fusion for microexpression recognition. Zhang el at. [21] propose a novel spatio-temporal transformer architecture for microexpression recognition. Zhai el at. [22] propose a novel framework Feature Representation Learning with adaptive Displacement Generation and Transformer fusion (FRL-DGT).

B. Microexpression Magnification

In addition to the research on recognition methods, microexpression magnification also improves the recognition rate. Park et al. [23] magnified the motion vector obtained based on the feature point tracking method and subsequently extracted features for recognition. Park et al. [24] proposed an adaptive magnification method based on time features. Many amplification tasks are improved based on the linear Euler-based video magnification algorithm (EVM) [9] and phase-based video magnification method [10]. Lei et al. [12] used magnification method LVMM [11] based on deep transfer learning for the first time, which has not specifically applied to microexpression magnification. Then, Song et al. [14] proposed a teacher-student framework for microexpression magnification, and the effect was achieved in the wild sample test.

III. Methods

A. Problem Formulation

In the previous magnification methods based on deep learning [11], [14], the features between two given frames are extracted; then, magnification is performed. Unlike their methods, our goal is to magnify the expression with only one microexpression image.

According to the theory in [14], microexpressions have a very similar movement of corresponding macroexpressions under the same emotion label. In other words, the microexpression variation can be seen as a part of macroexpression variation. As shown in Fig. 2, in macroexpression sequence, the sequence between the onset frame and the middle frame is taken as the microexpression sequence, and the middle frame is regarded as the apex frame of the microexpression sequence.

Refer to the above research, we also turn microexpression magnification task into generation task from the middle frame to the apex frame. But different from them, we get the neutral face (the onset frame in Fig. 2) by decreasing the expression information from the microexpression image (the middle frame in Fig. 2). As shown in (1):

$$Onset = Middle - Exp_{onset\rightarrow mid} \tag{1}$$

Further, if the microexpression image gets more expression information, it becomes a macroexpression image (apex frame in Fig. 2). As shown in (2):

$$Apex = Middle + Exp_{mid\rightarrow apex} \tag{2}$$

In (1), the expression information $Exp_{onset\rightarrow mid}$ is contained in the microexpression image $Middle$. Therefore, how to accurately extract reliable expression information from microexpression images and decrease these information is the first problem we want to solve.

$$Onset = Deexpression(Middle) \tag{3}$$

In (2), in order to get the macroexpression image $Apex$, we need to focus on the expression information $Exp_{mid\rightarrow apex}$. Since macroexpression is a continuous motion variation, $Exp_{onset\rightarrow mid}$ in (1) can be regarded as the previous part of $Exp_{mid\rightarrow apex}$, how to infer $Exp_{mid\rightarrow apex}$ from $Exp_{onset\rightarrow mid}$ is the second problem we need to solve:

$$Exp_{mid\rightarrow apex} = Mag(Exp_{onset\rightarrow mid}) \tag{4}$$

Based on (2) and (4), we get:

$$Apex = Middle + Mag(Exp_{onset\rightarrow mid}) \tag{5}$$

From (1), we know that $Exp_{onset\rightarrow mid} = Middle - Onset$. If we combine (3) and (5), we can generate the corresponding macroexpression image from a given single microexpression image:

$$\begin{aligned} &Apex = Middle + \\ &Mag(Middle - Deexpression(Middle)) \end{aligned} \tag{6}$$

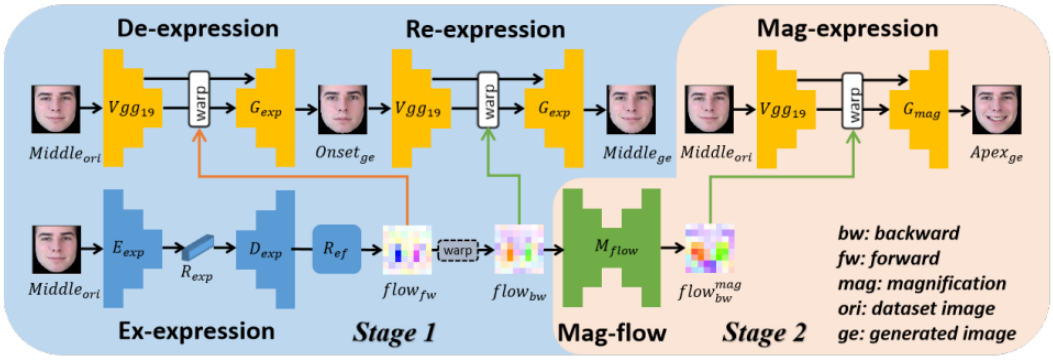

Fig. 3. General framework of the proposed approach. The details of module R_{ef} are shown in Fig. 3.

B. Network Framework

To solve the two main problems, we split our network into two stages, which are end-to-end. In stage 1 (stage 1 in Fig. 3), we employ encoder-decoder architectures E_{exp} and D_{exp} to extract expression information from a single microexpression image. This process is defined as Ex-expression:

Ex-expression: Ex-expression is defined as modeling the expression information R_{exp}, which is extracted from a single microexpression image $Input$ and expressed as optical flow $flow_{fw}$.

To ensure that the extracted optical flow is reliable and effective enough, inspired by cyclegan [15], we apply an encoder-decoder network loop including **De-expression** and **Re-expression** processes.

De-expression: De-expression is defined as removing expression information R_{exp} (R_{exp} is expressed as optical flow $flow_{fw}$) from the single microexpression image input $Input$. It outputs the corresponding neutral face image $Onset_{ge}$.

Re-expression: Re-expression is defined as adding the neutral face image $Onset_{ge}$ with expression information R_{exp} (R_{exp} is expressed as optical flow $flow_{bw}$). It outputs the corresponding microexpression image $Middle_{ge}$.

In stage 2 (stage 2 of Fig. 3), to obtain the magnified expression features, we feed the optical flow of stage 1 into an encoder-decoder network M_{flow}. This process is called Mag-flow:

Mag-flow: Mag-flow is defined as learning the magnified optical flow $flow_{bw}^{mag}$ using the extracted optical flow $flow_{bw}$.

Finally, to successfully obtain the macroexpression image from the magnified optical flow, we introduce an encoder-decoder network **Mag-expression**.

Mag-expression: Mag-expression is defined by magnifying the input $Input$ to generate the macroexpression image $Output(Apex_{ge})$ by fusing the magnified optical flow $flow_{bw}^{mag}$.

C. Extracting Reliable Expression Features from a Single Image

Based on the motivation in subsec:3.1, in this subsection, we will introduce how to extract reliable expression features from a single microexpression image. As shown in stage 1 in Fig. 3, we take the microexpression image as the input. We first employ an encoder network E_{exp} that can extract low-dimensional expression information from a single image R_{exp}:

$$R_{exp} = E_{exp}(Input) \tag{7}$$

Next, we adopt a decoder network D_{exp} to model the expression feature R_{exp} as differential features between microexpression face and neutral expression face.

$$Dif_{feat} = D_{exp}(R_{exp}) \tag{8}$$

(7) and (8) represent **Ex-expression**. Due to the large error of coordinate information contained in the feature map $Dif_{feat} \in R^{H\times W\times W}$, it cannot meet the requirements of our task. To further refine the differential features and generate the optical flow, we design a pipeline to integrate $Dif_{feat} \in R^{H\times W\times W}$ refer to [25]. This pipeline is shown in Fig. 4 and (9), mainly using soft-argmax. The visualization of refiner is shown in Fig. 5.

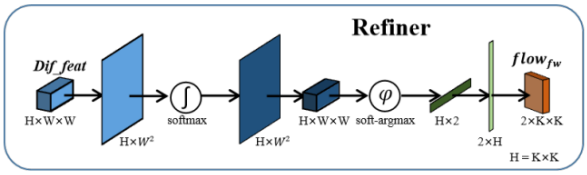

Fig. 4. Module R_{ef} in Fig. 3.

$$flow_{fw} = T(\varphi(T(exp(\frac{T(Dif_{feat})}{\sigma})))) \tag{9}$$

Here, $\varphi(\cdot)$ is $soft-argmax$, and $T(\cdot)$ is the transpose and reshape operation. Finally, we obtain the refined forward $(Middle \rightarrow Onset)$ optical flow $flow_{fw} \in R^{2\times k\times k}$. Based on the consistency of forward and backward optical flow [26], [27], which means that a pixel traversing flow vector forward and then backward should arrive at the same position [28], we can obtain the backward $(Onset \rightarrow Middle)$ optical flow $flow_{bw} \in R^{2\times k\times k}$ from the forward optical flow $flow_{fw}$ by a warping function.

$$flow_{bw} = -\omega(flow_{fw}, flow_{fw}) \tag{10}$$

We use bilinear interpolation to implement the warping operation $\omega(\cdot)$ as in [29].

Obviously, we cannot have a true optical flow for reliable training; we refer to the idea of cyclegan [15]. We design periodic loops from microexpression faces $Input$ to neutral faces $Onset_{ge}$ and subsequently to microexpression faces $Middle_{ge}$ by combining $flow_{fw}$ and $flow_{bw}$.

To generate the target face image $Onset_{ge}$ from $Input \in R^{3\times n\times n}$, we first obtain coarse-to-fine feature maps $Feat$ of the input face image at different scales from the $conv2\_1$ and $conv3\_1$ layers of the VGG19 network pretrained on ImageNet [30]. Based on the given forward optical flow $flow_{fw}$, we simply use the bilinear interpolation method $ds(\cdot)$ to obtain $ds(flow_{fw})$ which has the same size as $Feat$. Then, $ds(flow_{fw})$ is used to warp the corresponding feature map $Feat$ and decoder G_{exp} is used to generate the neutral face image $Onset_{ge}$. The **De-expression** can be implemented as follows:

$$Onset_{ge} = G_{exp}(\omega(Feat, ds(flow_{fw}))) \tag{11}$$

Additionally, we believe that by warping the corresponding feature maps $Feat$ using $ds(flow_{bw})$ and feeding the result into decoder G_{exp}, we can reconstruct the microexpression image $Middle_{ge}$ from the neutral expression image $Onset_{ge}$. This process is called **Re-expression**:

$$Middle_{ge} = G_{exp}(\omega(Feat, ds(flow_{bw}))) \tag{12}$$

D. Cyclic consistency of facial expressions

As mentioned above, to successfully extract expression information from a single image, we further model the extracted expression information as the optical flow field between neutral expressions and microexpressions. Then, based on the optical flow consistency [26], [27], we obtain $Input + flow_{fw} \rightarrow$ **De-expression** $\rightarrow Onset_{ge} + flow_{bw} \rightarrow$ **Re-expression** $\rightarrow Middle_{ge}$. In this step, the groundtruth can be used to train the optical flow, and we should have:

$$\begin{aligned} Middle &= Middle_{ge} \\ Onset &= Onset_{ge} \end{aligned} \tag{13}$$

To make the expression information extracted from a single image sufficiently reliable, we introduce a total of three losses to evaluate the reconstruction error. First, the first loss is the L1 loss. We will calculate the loss between the generated image and the true image:

$$\begin{aligned} L_1 = &\mid Middle - Middle_{ge} \mid + \\ &\mid Onset - Onset_{ge} \mid \end{aligned} \tag{14}$$

Then, we introduce perceptual loss [31] using the pretrained VGG19 network. As shown in (13):

$$\begin{aligned} L_2 = &\sum_{i=1}^{n} \mid N_i(Middle) - N_i(Middle_{ge}) \mid + \\ &\sum_{i=1}^{n} \mid N_i(Onset) - N_i(Onset_{ge}) \mid \end{aligned} \tag{15}$$

Among them, the channel features are extracted from the specific VGG-19 layer, where the $conv2\_1$, $conv3\_1$ and $conv4\_1$ layers are used here. To ensure the quality of the generated images, we also introduce SSIM loss [32]. It is represented by the weighting and product of three elements of luminance (l), contrast (c) and structure (s) of two images, which takes into account human visual perception:

$$\begin{aligned} SSIM(x,y) &= [l(x,y)^{\alpha} \cdot c(x,y)^{\beta} \cdot s(x,y)^{\gamma}] \\ &= \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \end{aligned} \tag{16}$$

μ is the mean, σ is the variance, σ_{xy} is the covariance, and α, β and γ are generally taken as 1. C_1 and C_2 are constants to avoid the instability problem when the denominator is 0. Based on (13) and (16), we have:

$$\begin{aligned} L_3 = &SSIM(Middle, Middle_{ge}) + \\ &SSIM(Onset, Onset_{ge}) \end{aligned} \tag{17}$$

E. Mag-expression

Based on the aforementioned facial motion cycle consistency, we extract reliable expression features from a single microexpression face image. To generate macroexpression images from a single microexpression image, we must magnify the extracted expression features. As shown in stage 2 in Fig. 3, the backward optical flow $flow_{bw}$ in stage 1 is fed into a network M_{flow} with an encoder-decoder structure to learn to magnify the backward $(Middle_* \rightarrow Apex_*)$ optical flow. This process is called **Mag-flow**:

$$flow_{bw}^{mag} = M_{flow}(flow_{bw}) \tag{18}$$

We obtain the macroexpression image $Apex_{ge}$ by using $flow_{bw}^{mag}$ to warp the pixels of $Input$ based on warping function $\omega(\cdot)$:

$$Apex_{ge} = G_{exp}(\omega(Feat, ds(flow_{bw}^{mag})) \tag{19}$$

The process of magnifying an expression from a microexpression image to a macroexpression image is called **Mag-expression**. Ideally, the resulting magnified image $Apex_{ge}$ should look identical to the apex image $Apex_{ori}$:

$$Apex = Apex_{ge} \tag{20}$$

We apply the three aforementioned losses to calculate the reconstruction error between ground truth image and generated image:

$$\begin{aligned} L = &a* \mid Apex - Apex_{ge} \mid + \\ &b * \sum_{i=1}^{n} \mid N_i(Apex) - N_i(Apex_{ge}) + \\ &c * SSIM(Apex, Apex_{ge}) \end{aligned} \tag{21}$$

IV. EXPERIMENTS

A. Train details and application

1) Train: In our experiment, we used databases MMI [33], [34] and CK+ [35] to train our model. There are a total of 766 sequence samples. In MMI, the expression sequence of each sample was from the neutral expression to the peak expression and subsequently to the neutral expression. We selected the middle three frames of each sample sequence

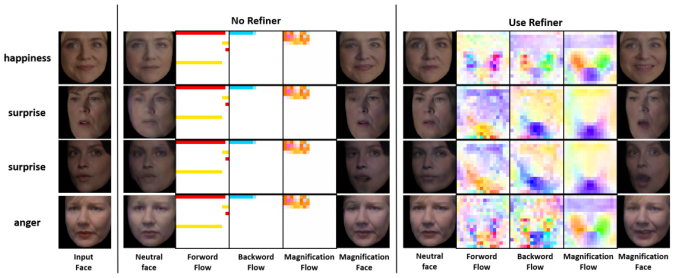

Fig. 5. Results of no refiner and using a refiner.

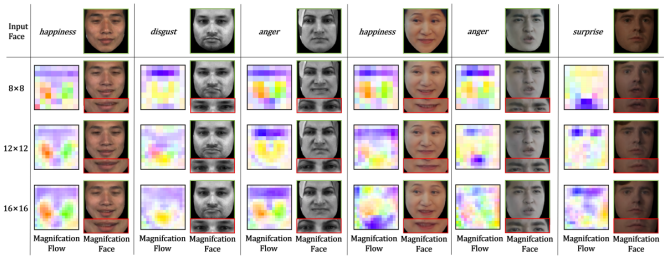

Fig. 6. Ablation experiments on the size of the optical flow.

as the peak expression and only reserved the part from the neutral expression to the peak expression. The sequence of samples in CK+ was the neutral expression to the peak expression, so no additional processing was required. For all macroexpression datasets, we aligned facial images in each sample and output 112×112 images.

TABLE I

COMPARISON OF $FID \downarrow$ ON DIFFERENT DATASETS.

	CASME II	SAMM	SMIC	Microexpression hybrid database	The wild samples
LVMM3	20.995	31.480	58.498	32.805	34.602
MEMM	19.986	**16.404**	16.866	14.081	32.342
Ours	**14.725**	19.911	**13.930**	**12.002**	**27.052**

Throughout the training process, we used the onset frame, middle frame, and apex frame of each sample sequence. When the sample sequence frame number was $k > 9$, the onset frame was the first frame, one random frame from the third to the penultimate was taken as the middle frame, and the last frame was the apex frame. When the number of sample sequence frames was $k \leq 9$, the onset frame was the first frame, the $k/2$ frame was the middle frame, and the last frame was the apex frame.

In stage 1 of training, we trained E_{exp}, D_{exp} and G_{exp} with epochs set to 15. In stage 2, we trained M_{flow} and G_{mag}, and the epoch was set to 40. We used the Adam [36] optimizer to train our network, with bachsize set to 16 and learning rate $1e-5$ for all training sessions.

We used CASME II [37], SAMM [38] and SMIC [39] microexpression datasets to apply microexpression magnification to our trained models. In total, 442 samples from three databases were included in the application. We used the apex frame of each sample sequence as the input and output a magnified image. We used the wild samples that were used by Song et al. [14] for testing. Again, we cropped and aligned all samples and output 112×112 images.

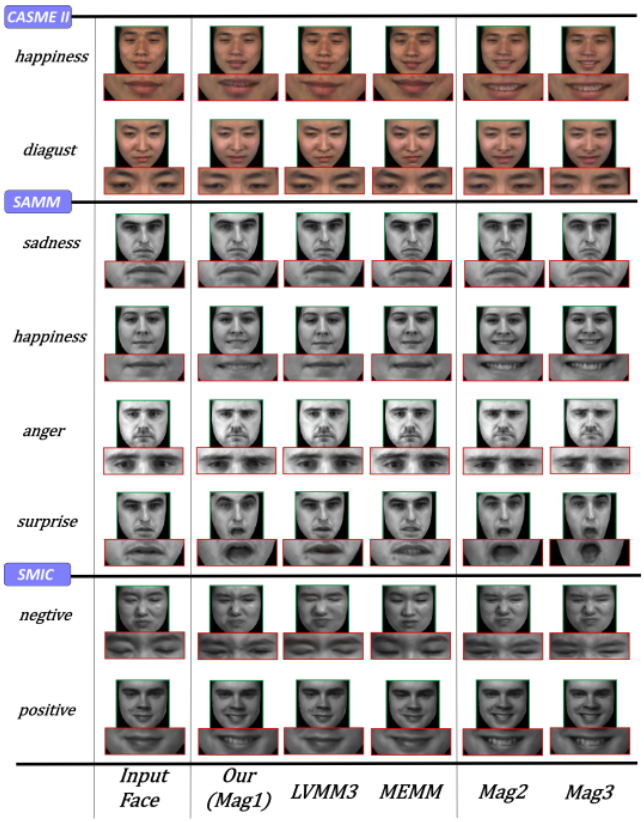

Fig. 7. Comparison with previous methods on the microexpression datasets. The first two samples are from CASME II, the middle fuel samples are from SAMM, and the last two samples are from SMIC.

TABLE II

COMPARING WITH RECENT COMPLEX RECOGNITION MODELS UNDER MEGC 2019 STANDARD.

Methods	*SAMM UF1*	*SAMM UAR*	*SMIC UF1*	*SMIC UAR*	*CASME II UF1*	*CASME II UAR*	*FULL UF1*	*FULL UAR*
LBP-TOP [40] (2007)	0.3954	0.4102	0.2000	0.5280	0.7026	0.7429	0.5882	0.5785
Bi-WOOF [41] (2018)	0.5211	0.5139	0.5727	0.5829	0.7805	0.8026	0.6296	0.6227
OFF-ApexNet [42] (2019)	0.5409	0.5392	0.6817	0.6695	0.8764	0.8681	0.7196	0.7096
CapsuleNet [43] (2019)	0.6209	0.5989	0.5820	0.5877	0.7068	0.7018	0.6520	0.6506
Dual-Inception [17] (2019)	0.5868	0.5663	0.6645	0.6726	0.8621	0.8560	0.7322	0.7278
STST-Net [44] (2019)	0.6588	0.6810	0.6801	0.7013	0.8382	0.8686	0.7353	0.7605
EMR [18] (2020)	0.7754	0.7152	**0.7461**	**0.7530**	0.8293	0.8209	0.7885	0.7824
LFM-based [45] (2020)	0.6700	0.6000	0.7200	0.7100	0.8700	0.8400	0.7700	0.7500
GEME [19] (2021)	0.6868	0.6541	0.6288	0.6570	0.8401	0.8508	0.7395	0.7500
FAUF [13] (2021)	0.7751	[0.7890]	0.7192	0.7215	0.8798	0.8710	0.7914	0.7933
FR [20] (2022)	0.7372	0.7155	0.7011	0.7083	0.8915	0.8873	0.7838	0.7832
SLSTT-LSTM [21] (2022)	0.7150	0.6430	0.7400	0.7200	0.9010	0.8850	[0.8160]	0.7900
ME-PLAN [46] (2022)	0.7358	0.7687	N/A	N/A	0.8941	0.8962	0.7979	0.8041
FRL-DGT [22] (2023)	0.7720	0.7580	[0.7430]	[0.7490]	[0.9190]	[0.9030]	0.8120	[0.8110]
LVMM3 [11] +ResNet50 (2018)	0.7130	0.6551	0.6735	0.6782	0.8241	0.8065	0.7493	0.7225
MEMN [47]+ResNet50 (2022)	N/A	N/A	N/A	N/A	N/A	N/A	0.7590	0.7450
MEMM4 [14] +ResNet50 (2022)	[0.8100]	0.7753	0.7326	0.7305	**0.9367**	**0.9279**	**0.8240**	0.8060
Ours+ResNet18	**0.8221**	**0.8146**	0.7349	0.7447	0.8610	0.8746	0.8045	**0.8153**

B. Ablative analysis

1) Optical flow refiner: In Fig. 5, we show the intermediate results (including the generated neutral face, forward optical flow, backward optical flow, and magnified optical flow) and the final magnified results of our model; the case of using the refiner is shown on the right, and the case without using the refiner is shown on the left. In Fig. 5, without the refiner, our method cannot extract an effective optical flow field from a single picture. In addition, the generated images will be distorted, and the magnification effect is not satisfactory. After using the refiner, the optical flow output by the model can effectively map the facial motion with coordinates. The generated neutral faces and magnified faces show no distortion, and the magnified effect is significantly

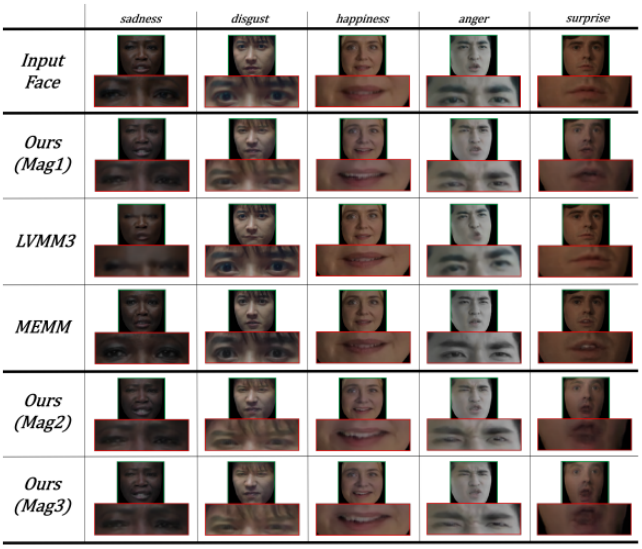

Fig. 8. Cascade magnification of the wild samples.

improved. Therefore, we can prove that our designed refiner is effective.

2) Size of the optical flow field: In the previous subsection, we set the optical flow field to a size of 16×16. Since facial expression is a very subtle feature, to find the best size of the optical flow field for our network, we conducted a series of ablation studies on the size of the optical flow field. We compared them in three sizes: 8×8, 12×12 and 16×16. Fig. 6 is a demonstration of some samples (they are from the microexpression database CASME II, SAMM, SMIC and wild samples). From the magnification effect, the optical flow at three scales can capture the characteristics of the key parts. Overall, we found that the network with a size of 8×8 had the best performance, and it could focus more features on inconspicuous areas such as the eyebrow in the second sample, the eyebrow in the sixth sample, and the mouth in the final sample. Therefore, we choose 8×8 as the size of our optical flow field.

C. Comparison with previous methods

1) Comparison of the microexpression datasets: We test our method against LVMM [11] and MEMM [14] on the CASME II [37], SAMM [38] and SMIC [39] microexpression databases. Both LVMM and MEMM take two frames as input, while LVMM musty set the magnification factor. LVMM has been effectively applied for preprocessing in recent microexpression recognition studies [12], [13]. They use LVMM to capture small movements between two frames and amplify them by a corresponding factor (magnification factor). We set the magnification factor to 3 according to the discussion of a previous paper [12], [13]. MEMM is the latest work and the first work that focuses on magnifying microexpressions; it can achieve a good result without factors. As shown in Fig. 7, the results show that our method can achieve the best results among all methods, even when we use a single frame input.

2) Comparison of wild samples: Datasets in laboratory environments tend to have faces directly facing the camera and constant lighting conditions. However, in wild samples, it is often impossible to ensure that the face is frontal in images, and there are various lighting environments. Therefore, the performance in wild samples is more challenging. To further demonstrate our method, we tested it using wild samples. In Fig. 8, the results show that our method can achieve a better magnification effect than the other two methods.

3) Comparison of image quality : In this section, we evaluated the quality of the generated images. We measured the distance between two data distributions using FID [48], which is commonly used to investigate whether a generated dataset is natural. As shown in Table I, our method performs the best.

D. Cascade testing

To examine our approach, we introduced the idea of cascading in the testing process. Fig. 7 and Fig. 8 display the effect of cascade testing of our method on the mixed dataset of microexpressions and samples in the wild. The magnification is enhanced when the number of cascades increases. Within three cascaded magnifications (Mag3), our method can magnify microexpression images to satisfactory results.

E. Evaluation criteria

To further demonstrate our approach, we input images into Resnet18 to perform microexpression recognition. As shown in Table II We conducted a convincing experimental evaluation using the composite database evaluation (CDE) of MEGC 2019 [16], where Unweighted F1-score (UF1) and unweighted average recall (UAR) were used to evaluate the results. We simply feed the enlarged image to the ResNet18 network, and the results show that our method exceeds the previous recognition work. The previous amplification methods used double frames as input, and our method obtained enough competitive scores even on the basis of single frame as input.

F. Conclusion

In this paper, we proposed a deep learning method to generate macroexpression images from a single microexpression image. We tested both microexpression datasets and wild samples and compared the test results with the current methods. Our method can generate sufficiently desirable macroexpression images even if we take a single image as the input. Moreover, our method greatly improves the application prospects of microexpression amplification tasks.

REFERENCES

[1] E. A. Haggard and K. S. Isaacs, "Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy," in *Methods of research in psychotherapy*. Springer, 1966, pp. 154–165.

[2] S.-J. Wang, Y. He, J. Li, and X. Fu, "Mesnet: A convolutional neural network for spotting multi-scale micro-expression intervals in long videos," *IEEE Transactions on Image Processing*, vol. 30, pp. 3956–3969, 2021.

[3] Z. Lautman and S. Lev-Ari, "The use of smart devices for mental health diagnosis and care," p. 5359, 2022.

[4] B. Allaert, I. M. Bilasco, and C. Djeraba, "Micro and macro facial expression recognition using advanced local motion patterns," *IEEE Transactions on Affective Computing*, vol. 13, no. 1, pp. 147–158, 2019.

[5] B. Bhushan, "Study of facial micro-expressions in psychology," *Understanding facial expressions in communication: Cross-cultural and multidisciplinary perspectives*, pp. 265–286, 2015.

[6] W.-J. Yan, Q. Wu, J. Liang, Y.-H. Chen, and X. Fu, "How fast are the leaked facial expressions: The duration of micro-expressions," *Journal of Nonverbal Behavior*, vol. 37, pp. 217–230, 2013.

[7] M. Frank, M. Herbasz, K. Sinuk, A. Keller, and C. Nolan, "I see how you feel: Training laypeople and professionals to recognize fleeting emotions," in *The Annual Meeting of the International Communication Association. Sheraton New York, New York City*, 2009, pp. 1–35.

[8] S. Nag, A. K. Bhunia, A. Konwer, and P. P. Roy, "Facial micro-expression spotting and recognition using time contrasted feature with visual memory," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 2022–2026.

[9] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman, "Eulerian video magnification for revealing subtle changes in the world," *ACM transactions on graphics (TOG)*, vol. 31, no. 4, pp. 1–8, 2012.

[10] N. Wadhwa, M. Rubinstein, F. Durand, and W. T. Freeman, "Phase-based video motion processing," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 4, pp. 1–10, 2013.

[11] T.-H. Oh, R. Jaroensri, C. Kim, M. Elgharib, F. Durand, W. T. Freeman, and W. Matusik, "Learning-based video motion magnification," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 633–648.

[12] L. Lei, J. Li, T. Chen, and S. Li, "A novel graph-tcn with a graph structured representation for micro-expression recognition," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 2237–2245.

[13] L. Lei, T. Chen, S. Li, and J. Li, "Micro-expression recognition based on facial graph representation learning and facial action unit fusion," in *Proceedings of the IEEE/CVF workshops on Computer Vision and Pattern Recognition*, 2021, pp. 1571–1580.

[14] Y. Song, W. Zhao, T. Chen, S. Li, and J. Li, "Recognizing microexpression as macroexpression by the teacher-student framework network," in *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 2022, pp. 548–553.

[15] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[16] J. See, M. H. Yap, J. Li, X. Hong, and S.-J. Wang, "Megc 2019–the second facial micro-expressions grand challenge," in *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*. IEEE, 2019, pp. 1–5.

[17] L. Zhou, Q. Mao, and L. Xue, "Dual-inception network for cross-database micro-expression recognition," in *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*. IEEE, 2019, pp. 1–5.

[18] C. Wang, M. Peng, T. Bi, and T. Chen, "Micro-attention for micro-expression recognition," *Neurocomputing*, vol. 410, pp. 354–362, 2020.

[19] X. Nie, M. A. Takalkar, M. Duan, H. Zhang, and M. Xu, "Geme: Dual-stream multi-task gender-based micro-expression recognition," *Neurocomputing*, vol. 427, pp. 13–28, 2021.

[20] L. Zhou, Q. Mao, X. Huang, F. Zhang, and Z. Zhang, "Feature refinement: An expression-specific feature learning and fusion method for micro-expression recognition," *Pattern Recognition*, vol. 122, p. 108275, 2022.

[21] L. Zhang, X. Hong, O. Arandjelović, and G. Zhao, "Short and long range relation based spatio-temporal transformer for micro-expression recognition," *IEEE Transactions on Affective Computing*, vol. 13, no. 4, pp. 1973–1985, 2022.

[22] Z. Zhai, J. Zhao, C. Long, W. Xu, S. He, and H. Zhao, "Feature representation learning with adaptive displacement generation and transformer fusion for micro-expression recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 22 086–22 095.

[23] S. Park and D. Kim, "Subtle facial expression recognition using motion magnification," *Pattern Recognition Letters*, vol. 30, no. 7, pp. 708–716, 2009.

[24] S. Y. Park, S. H. Lee, and Y. M. Ro, "Subtle facial expression recognition using adaptive magnification of discriminative facial motion," in *Proceedings of the 23rd ACM international conference on Multimedia*, 2015, pp. 911–914.

[25] A. Siarohin, S. Lathuilière, S. Tulyakov, E. Ricci, and N. Sebe, "First order motion model for image animation," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[26] M. J. Black and P. Anandan, "The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields," *Computer vision and image understanding*, vol. 63, no. 1, pp. 75–104, 1996.

[27] J. Hur and S. Roth, "Mirrorflow: Exploiting symmetries in joint optical flow and occlusion estimation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 312–321.

[28] Y. Zou, Z. Luo, and J.-B. Huang, "Df-net: Unsupervised joint learning of depth and flow using cross-task consistency," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 36–53.

[29] D. Sun, X. Yang, M.-Y. Liu, and J. Kautz, "Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8934–8943.

[30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[31] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European conference on computer vision*. Springer, 2016, pp. 694–711.

[32] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[33] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in *2005 IEEE international conference on multimedia and Expo*. IEEE, 2005, pp. 5–pp.

[34] M. Valstar, M. Pantic, *et al.*, "Induced disgust, happiness and surprise: an addition to the mmi facial expression database," in *Proc. 3rd Intern. Workshop on EMOTION (satellite of LREC): Corpora for Research on Emotion and Affect*. Paris, France., 2010, p. 65.

[35] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *2010 ieee computer society conference on computer vision and pattern recognition-workshops*. IEEE, 2010, pp. 94–101.

[36] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[37] W.-J. Yan, X. Li, S.-J. Wang, G. Zhao, Y.-J. Liu, Y.-H. Chen, and X. Fu, "Casme ii: An improved spontaneous micro-expression database and the baseline evaluation," *PloS one*, vol. 9, no. 1, p. e86041, 2014.

[38] A. K. Davison, C. Lansley, N. Costen, K. Tan, and M. H. Yap, "Samm: A spontaneous micro-facial movement dataset," *IEEE transactions on affective computing*, vol. 9, no. 1, pp. 116–129, 2016.

[39] X. Li, T. Pfister, X. Huang, G. Zhao, and M. Pietikäinen, "A spontaneous micro-expression database: Inducement, collection and baseline," in *2013 10th IEEE International Conference and Workshops on Automatic face and gesture recognition (fg)*. IEEE, 2013, pp. 1–6.

[40] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 6, pp. 915–928, 2007.

[41] S.-T. Liong, J. See, K. Wong, and R. C.-W. Phan, "Less is more: Micro-expression recognition from video using apex frame," *Signal Processing: Image Communication*, vol. 62, pp. 82–92, 2018.

[42] Y. S. Gan, S.-T. Liong, W.-C. Yau, Y.-C. Huang, and L.-K. Tan, "Off-apexnet on micro-expression recognition system," *Signal Processing: Image Communication*, vol. 74, pp. 129–139, 2019.

[43] N. Van Quang, J. Chun, and T. Tokuyama, "Capsulenet for micro-expression recognition," in *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*. IEEE, 2019, pp. 1–7.

[44] S.-T. Liong, Y. S. Gan, J. See, H.-Q. Khor, and Y.-C. Huang, "Shallow triple stream three-dimensional cnn (ststnet) for micro-expression recognition," in *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*. IEEE, 2019, pp. 1–5.

[45] D. Y. Choi and B. C. Song, "Facial micro-expression recognition using two-dimensional landmark feature maps," *IEEE Access*, vol. 8, pp. 121 549–121 563, 2020.

[46] S. Zhao, H. Tang, S. Liu, Y. Zhang, H. Wang, T. Xu, E. Chen, and C. Guan, "Me-plan: A deep prototypical learning with local attention network for dynamic micro-expression recognition," *Neural Networks*, vol. 153, pp. 427–443, 2022.

[47] W. Zhao, L. Lei, T. Chen, S. Li, and J. Li, "Micro-expression recognition by combining progressive-learning intensity magnification with self-attention-convolution classification," in *2022 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 2022, pp. 1–10.

[48] M. Lucic, K. Kurach, M. Michalski, S. Gelly, and O. Bousquet, "Are gans created equal? a large-scale study," *Advances in neural information processing systems*, vol. 31, 2018.