EasyChair Preprint

№ 6909

Off-TANet:a Lightweight Neural Micro-Expression Recognizer with Optical Flow Features and Integrated Attention Mechanism

Jiahao Zhang, Feng Liu and Aimin Zhou

October 20, 2021

Off-TANet:A Lightweight Neural Micro-expression Recognizer with Optical Flow Features and Integrated Attention Mechanism *

Jiahao Zhang[1,2], Feng Liu[1,2,3], and Aimin Zhou[1,2,3]

[1] School of Computer Science and Technology, East China Normal University, China
zjh20000218@163.com
[2] Shanghai Institute for AI Education, East China Normal University, China
[3] Shanghai Key Laboratory of Mental Health and Psychological Crisis Intervention, China

Abstract. Micro-expression recognition is a video sentiment classification task with extremely small sample size. The transience and spatial locality of micro-expressions bring difficulties to constructing large micro-expression databases and designing micro-expression recognition algorithms. To reach the balance between classification accuracy and model complexity in this domain, we propose a lightweight neural micro-expression recognizer, Off-TANet, which is based on apex-onset optical flow features. The neural network contains a simple yet powerful triplet attention mechanism, and the powerfulness of this design could be interpreted in 2 aspects, FACS AU and matrix sparseness. The model evaluation is conducted with a LOSO cross-validation strategy on a combined database including 3 mainstream micro-expression databases. With obviously fewer total parameters (59,403) , the results of the experiment indicate that the model achieves an average recall of 0.7315 and an average F1-score of 0.7242, exceeding other major architectures in this domain. A series of ablation experiments are also conducted to ensure the validity of our model design.

Keywords: Micro-expression recognition · Attention module · Self-Attention mechanism · Optical flow features · Convolutional neural networks · Computational affection

1 Introduction

Micro-expression is a very brief and rapid facial motion that is provoked involuntarily , which could reveal an individual's true emotions even when true feelings are deliberately concealed. Due to the affinity between micro-expression

* This study supported by The Research Project of Shanghai Science and Technology Commission (20dz2260300) and The Fundamental Research Funds for the Central Universities. also supported by the Science and Technology Commission of Shanghai Municipality (No. 19511120601).
Corresponding author: Feng Liu, Aimin Zhou.

and true emotions, micro-expression has a wide range of applications in mental disorder treatment, such as emotion recognition ability recovery for Schizophrenia patients[20,21] and Alexithymia diagnosis[24]. Compared with regular facial expressions (macro-expression), micro-expression is more subtle both temporally and spatially. To be more specific, the duration of a micro-expression is rather low (between $1/25$s and $1/5$ s) [33], and a micro-expression only occurs in limited facial regions [13]. The nature of micro-expression not only brings challenges to automated micro-expression recognition but also causes data creation difficulties including human data labeling, sample video capturing, and micro-expression induction. As a consequence, the process of constructing large micro-expression datasets is severely delayed, and micro-expression recognition is still a small sample size problem even to this day.

In recent years, the MEGCs (Micro-Expression Grand Challenge) [22,32] accelerates the development of this domain. In MEGC 2019 [22], lightweight neural micro-expression classification approaches started to completely supersede handcrafted feature (LBP-TOP [8], etc.) based approaches with the help of the widely-used 'Less is more' onset-apex optical flow method [13]. In 2020, deep learning-based algorithms with more advanced techniques, such as graph neural networks and dilated convolution, are led into this domain [16,10]. These models, though drastically inflating the parameter scale, show better recognition accuracy than proposed methods in 2019. Nevertheless, neural micro-expression recognition models in these years still reflect some disadvantages, and the main demerit is that the balance between parameter scale and classification performance is fairly unsatisfying for those models.

To solve the parameter-accuracy balance problem mentioned above, we propose a novel optical flow-based neural network architecture called Off-TANet (Optical flow feature-Triplet Attention Net) for micro-expression sentiment classification. We design a triplet attention module including spatial attention, channel attention, and self-attention, and applied this attention module on a minimalist residual network. We summarize our main contributions as follows:

- We design a powerful triplet attention mechanism and find an interpretation for the powerfulness of the novel attention module based on FACS AU[3] and matrix sparsity.
- This paper proposes a simplified neural network architecture, Off-TANet, with the triplet attention mechanism. The architecture could prevent overfitting and greatly reduce the number of parameters.
- In a combined micro-expression database including $CASME$ [31], $CASME\ II$ [30] and $CAS(ME)^2$ [19], two evaluation metrics, UAR and UF1, are verified in experiments. In comparison to the listed mainstream models and ablation study results, our network, with an extremely low number of parameters, could reach the state-of-the-art.

2 Related Work

Owning to the difficulty to construct a micro-expression dataset, research of micro-expressions are based on some public databases with a limited number of samples. The mainstream databases are SAMM [1], SMIC [11], $CASME$ [31], $CASME\ II$ [30] and $CAS(ME)^2$ [19]. The total number of micro-expression image sequences in all the datasets mentioned above is extremely low (less than 1000). In those datasets, despite being unitary in ethnics, the three $CASME$ datasets with higher image resolution and better preprocessing, including face region segmentation and apex frame detection, are more capable of training neural micro-expression recognizers.

Micro-expression recognition is a new domain in computer vision. Automated micro-expression recognition firstly appears in 2009 [18], which is much later than the burgeon of macro-expression recognition algorithms since the 1990s. After the early explorations of handcrafted feature-based micro-expression recognition, the innovative 'Less is more' apex-onset optical flow method is proposed[13], and neural micro-expression classifiers began to emerge in 2019. ResNet-18 with adversarial training and expression magnification and reduction[14] shows a relatively satisfying performance in MEGC 2019. Several novel CNN structures are also mentioned in MEGC 2019, such as Off-ApexNet[4], Dual-Inception [34], and STSTNet [12]. In 2020, Lo et al. proposed a graph convolution network-based model called MER-GCN [16], which applies a GCN on top of a 3D convolution network to explore the dependencies among different FACS AUs. A real-time micro-expression recognizer, MACNN [10], with residual blocks and atrous convolutions is proposed by Lai et al.to categorize a micro-expression in a low response time. Wang et al. try to improve the performance of ResNet [5] in micro-expression recognition by adding micro-attention modules.

Attention mechanism plays a significant role in human perception. Computer vision researchers have made several previous attempts of leading attention mechanisms into convolutional networks to improve the performance of feature extraction. The spatial transformer is a typical form of spatial attention mechanism, which means the mechanism applies the same warping to each channel [9]. This structure could rotate and scale the feature map and focus on the regions with important features. By contrast, channel attention mechanism allocates the weight of every channel instead of calculating the importance of every pixel in each feature map. An instance of channel attention mechanism is the "Squeeze-and-Excitation" (SE) block, which adaptively recalibrates channel-wise feature responses by explicitly modelling interdependencies between channels [7]. The fusion of spatial and channel attention also shows satisfying perfomance [29,27]. Self-attention, as the essential operator of Transformers[25], was first widely used in the NLP (Natural Language Processing) tasks [25]. The early application of self-attention blocks in computer vision is the non-local neural network [28]. These days, the success of visual Transformers, including ViT[2], BoTNet[23], and Swin Transformer[15], also proves the effectiveness of this simple but powerful design.

3 Proposed Method

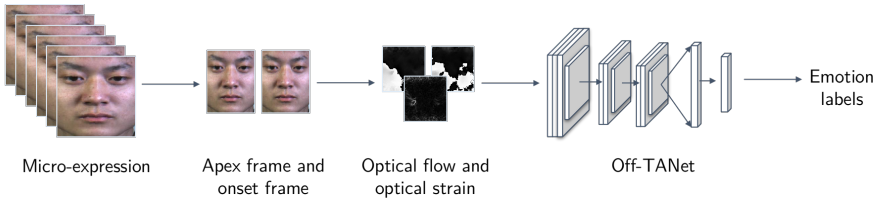


Fig. 1. The whole picture of our two-stage proposed method.

3.1 Optical Flow Feature Extraction

Considering that the facial movements in a micro-expression are extremely subtle, the difference between every two consecutive frames is inconspicuous. Instead of taking all frames as an input of the neural network, our micro-expression recognition pipeline (**Figure** 1) includes two main steps: onset-apex optical flow feature extraction and neural representation learning. This optical flow method firstly mentioned in [13] could memorably reduce the dimension of input features.

Let u and v denote the horizontal and vertical components of the optical flow vector field. In our feature extraction pipeline, the image partial derivatives are calculated by the Sobel operator, and u and v are solved by the TV-L1 optical flow algorithm [17].

Another optical flow-based feature called optical strain is also used in our work. It is capable of approximating the intensity of facial deformation [12], and can be defined as:

$$u = [u, v]^T \tag{1}$$

$$\epsilon(x, y) = \frac{1}{2}[\nabla u + (\nabla u)^T] \tag{2}$$

$$= \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{1}{2}(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}) \\ \frac{1}{2}(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}) & \frac{\partial v}{\partial y} \end{bmatrix} \tag{3}$$

The magnitude of optical strain is:

$$|\epsilon(x, y)| = \sqrt{(\frac{\partial u}{\partial x})^2 + (\frac{\partial v}{\partial y})^2 + \frac{1}{2}(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x})^2} \tag{4}$$

The optical flow features $\{u, v, |\epsilon|\}$ could be seen as a 3 channel image, and our neural network will take that 'image' as an input.

3.2 The Attention mechanism

Spatial and channel attention module. Experiments show that the convolutional block attention module (CBAM) is expected to boost the accuracy of lightweight networks [29]. We applied this argument-saving yet powerful structure to enhance the process of high-level feature extraction.

The CBAM includes two separated mechanisms: spatial attention and channel attention. In the spatial attention operator, the features are aggregated between different channels by both average and max pooling operations. Then the two spatial context descriptors are concatenated and convolved. In the channel attention operator, the features in each channel are aggregated by the two pooling operations, and then transformed by a shared MLP and merged by an element-wise summation. The spatial and channel attention map $SpA(\cdot)$ and $CA(\cdot)$ in our proposed method can be summarized as:

$$SpA(x) = Sigmoid(Conv([AvgPool(x); MaxPool(x)])) \tag{5}$$

$$CA(x) = Sigmoid(MLP(AvgPool(x)) + MLP(MaxPool(x))) \tag{6}$$

where $Conv$ denotes a convolution operator with a 3×3 kernel and MLP denotes a shared one-hidden layer perceptron. After the attention maps are calculated, the attention map could be applied by an element-wise multiplication on the input tensor.

Multi-head self-attention module. The burgeoning of visual Transformers showed the potential of the self-attention mechanism in computer vision tasks. Compared with spatial attention, which applies an identical map on each channel, the self-attention map on each channel differs. We applied this module after the CBAM module to construct a more powerful triplet attention mechanism.

The self-attention module in our network is similar to the self-attention block in NLP tasks. Specifically, the input tensor is transformed to three different representations query q, key k, and value v by three linear transformation matrices W_q, W_k, and W_v. Then we can calculate the output of the self-attention module as follows:

$$Attention(q, k, v, r) = Softmax(qk^T + qr^T) * v \tag{7}$$

$$r_{x,y} = PE(x, featureMapSize) + PE(y, featureMapSize) \tag{8}$$

$$PE(2i, d) = sin(1/10000^{2i/d}) \tag{9}$$

$$PE(2i + 1, d) = cos(1/10000^{2i/d}) \tag{10}$$

where r denotes the positional code, x and y represent pixel positions. The 2-D image positional code is constructed by adding the results of two 1-D sinusoidal positional code PE[25]. The attention map is then calculated by the element-wise summation of query-key matrix product and query-positional code matrix product. To obtain better performance, a multi-head self-attention mechanism is also applied by concatenating the output from self-attention blocks with unequal weights.

3.3 Network architecture

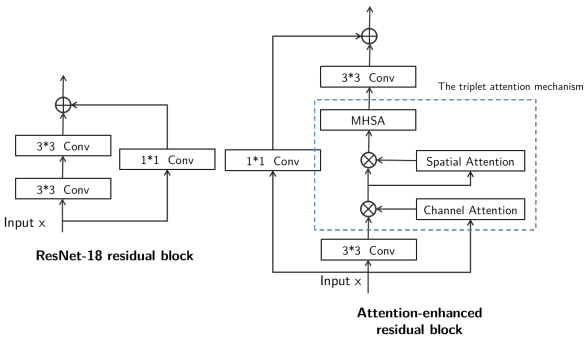


Fig. 2. The attention-enhanced residual block of our network, based on a ResNet-18 residual block. The triplet attention mechanism is added between the two convolution layers in a ResNet-18 residual block.

According to the experiments in BoTNet [23], apply an attention mechanism in the last residual block could improve the performance of ResNets. We replaced the last residual block in a minimalist residual convolutional network with a novel attention-enhanced residual block (**Figure** 2). Compared with the CBAM[29] attention module, our triplet attention module is expected to focus more on the important facial regions, as the CBAM channel and spatial attention extract features in a larger granularity, while the self-attention map differs both between different spatial positions and different channels.

The input apex images and onset images are firstly normalized to 112×112 with a cubic interpolation algorithm, then the $3 \times 112 \times 112$ optical flow features are extracted and sent to the network. The network architecture is shown in **Figure** 3. The number of channels in this architecture is under strict control to reduce the scale of parameters. The low-level features are extracted by a 7×7 convolution layer, and then the tensor is sent to two ResNet-18 style residual blocks. The attention-enhanced residual block with a triplet attention mechanism extracts the high-level representations. The output channels in the last residual block are reduced to shrink the FC layer, which could cause overfitting problems. More details about the network can be found in the published source code.

4 Experiments and Analysis

To validate the validity of our approach on micro-expression sentiment classification, we conduct experiments on a combined database including $CASME$ [31], $CASME\ II$ [30] and, $CAS(ME)^2$ [19]. Considering that mainstream models are tested on dissimilar benchmarks and datasets, all the models mentioned

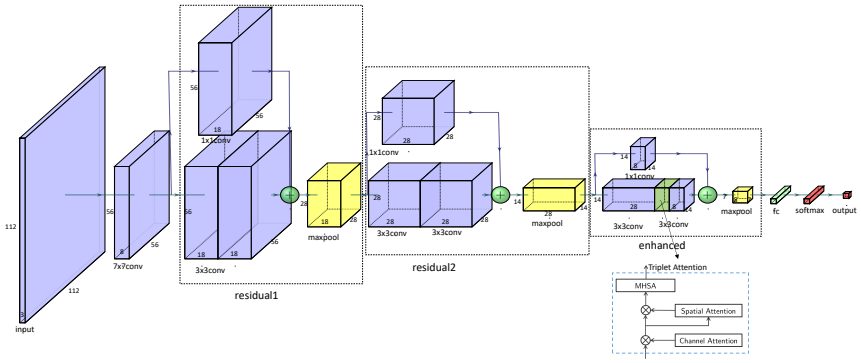


Fig. 3. The overall architecture of our network. The operator type and the output tensor shape in each layer are shown in this picture.

in **Table 2** are re-implemented and tested on this novel combined dataset with a 'MEGC 2019-like' benchmark.

Source code in Python, .csv format combined dataset (without images) and our running environments are available on
`https://github.com/ECNU-Cross-Innovation-Lab/PRICAI2021-Off-TANet`.

4.1 Data Preparation

The datasets used in our work are $CASME$ [31], $CASME\ II$ [30], and $CAS(ME)^2$ [19] respectively. The sentiment label, apex frame, onset frame of each micro-expression image sequence is provided by those datasets. Images are also properly cropped to get rid of the interference from pixels containing non-facial information. The input image will be normalized to different sizes with inter-cubic interpolation to adapt the input layer of each model, and RGB images are turned black and white before optical flow feature extraction.

To avoid confusing the learning process, We apply two 'MEGC 2019-like' data preparation methods. Among all the datasets, only main sentiment categories, which contain abundant micro-expression samples, are selected to form the combined dataset. Apart from that, macro-expression samples in $CAS(ME)^2$ [19] and 'Others' samples in $CASME\ II$ [30] are not used. To weaken the classification bias between datasets, we map those original labels to only three classes. Categorization information about the database can be seen in Table 1.

4.2 Algorithm Comparison

In this paper, all the methods mentioned in **Table** 2 are tested with a LOSO (Leave One Subject Out) protocol, and our metrics, UAR (Unweighted Average Recall) and UF1 (Unweighted F1-Score) are the same as metrics used in MEGC 2019 [22].

Table 1. Categorization information about the databases.

Label	Total samples	Number of samples			Original label
		$CASME$ [31]	$CASME\ II$ [30]	$CAS(ME)^2$[19]	
Positive	48	10	32	6	happiness
Negative	194	85	88	21	disgust repression sadness
Surprise	52	19	25	8	surprise

The LOSO protocol is a cross-validation strategy that repeats evaluation for 49 times by splitting out samples in each subject group in the 49-subject combined database. This widely used protocol effectively mimics realistic scenarios and ensures subject-independent evaluation.

The combined dataset is obviously imbalanced in category distribution, so class-balanced metrics are used in our experiments. The computation methods of UAR and UF1, which are also called balanced accuracy and macro-averaged F1-score, are as follows:

$$UAR = \frac{\sum_{c \in C} \frac{TP_c}{n_c}}{|C|} \tag{11}$$

$$UF1 = \frac{1}{|C|} \sum_{c \in C} \frac{2TP_c}{2TP_c + FP_c + FN_c} \tag{12}$$

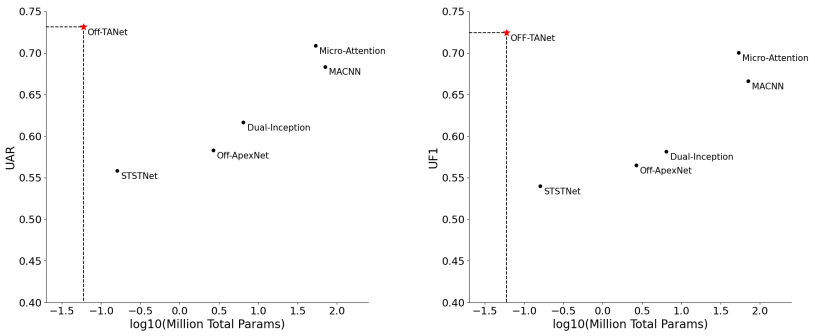
In these two formulas, C is the set of sentiment classes, n_c represents number of samples in class c, and TP_c, FP_c, FN_c means True Positive, False Positive, and False Negative. We assume that each left-out subject is of the same importance, so the final UAR and UF1 is the unweighted mean of all UARs and UF1s calculated in the 49-fold cross-validation.

Table 2. Results of Mainstream Approaches (sorted by UAR).

Model	**UAR**	**UF1**	**Params**	**Flops**	**MemR+W**
Off-ApexNet(2019)[4]	0.5832	0.5650	2.66M	3.87M	10.35MB
STSTNet(2019)[12]	0.5584	0.5399	162,051	526.98K	0.76MB
Dual-Inception(2019)[34]	0.6167	0.5814	6.45M	12.64M	25.67MB
MACNN(2020)[10]	0.6835	0.6660	70.57M	793.67M	1140.00MB
Micro-Attention(2020)[26]	0.7086	0.7003	53.38M	1.0G	237.97MB
Off-TANet(Ours)	**0.7315**	**0.7242**	**59,403**	**30.08M**	**5.64MB**

The results of mainstream approaches are illustrated in **Table** 2. **UAR**, **UF1**, the number of total parameters (**Params**), the total floating-point operation (**FLOPs**), and the total memory read/write (**MemR+W**) are listed in the

table. Parameter numbers and memory usage are measured by the **torchstat** Python package.



(a) TotalParams-UAR scatter diagram (b) TotalParams-UF1 scatter diagram

Fig. 4. The scatter plot of the test results. Our model reaches the highest UAR and UF1 with the lowest parameter number.

The hyperparameters of optimizers and the training process can be found in the code of this paper. The train epochs of re-implemented models are selected by a grid search to accommodate our new dataset, and other hyperparameters are determined according to the original papers. The architectures of MACNN [10] and Micro-Attention [26] are slightly adjusted to speed up the training process, avoid over-fitting and save GPU memory.

Table 2 and its corresponding scatter diagram **Figure** 4 directly shows the advantages of our architecture. With the lowest **Total Params (59,403)** and an obviously low **Total MemR+W (5.64MB)**, Off-TANet reached the highest **UAR (0.7315)** and **UF1 (0.7242)**.

4.3 Ablation Study

To ensure the validity of our model design, we have carried out a series of ablation experiment.

First of all, the effectiveness of the optical strain feature is examined. In the no optical strain experiment, we change the number of input channels of the first convolution layer and remove the optical strain feature from the input. The result also indicates the superiority of optical flow features compared with the end-to-end approach, which takes the raw onset and apex image as the input.

Despite the experiments in BotNet [23] show that the attention-enhancement should only be applied on the last residual block, we still compare the performance between Off-TANet and a network with identical output tensor shape in

Table 3. Results of ablation experiments on the input feature and the number of enhanced blocks.

Model	UAR	UF1
Off-TANet + no optical flow features	0.6631	0.6440
Off-TANet + no optical strain	0.7180	0.7078
Off-TANet + all residual block attention-enhanced	0.6895	0.6814
Off-TANet(Ours)	**0.7315**	**0.7242**

each block which change all the residual blocks with the triplet attention residual block. The accuracy of Off-TANet exceeds the accuracy of its counterpart which could cause overfitting problems.

Table 4. Results of ablation expreriments on the design of the last attention-enhanced residual block.

Model	UAR	UF1
No enhancement	0.6804	0.6745
+ CBAM[29]	0.6903	0.6744
+ multi-head self-attention	0.6943	0.6857
+ bottleneck Transformer[23]	0.7045	0.6937
Off-TANet(Ours)	**0.7315**	**0.7242**

The design of the attention-enhanced block is also discussed in this paper. On the original ResNet-18 residual block, we applied three different attention-enhancement approaches including only self-attention, only CBAM and bottleneck transformer. As visual Transformers are more data-hungry, the bottleneck transformer shows a less satisfying performance on micro-expression recognition – a small sample size problem. The self-attention enhancement and the CBAM approach also show their weakness compared with our design.

4.4 Off-TANet Attention Mechanism Analysis

Compared with raw CNN features, the attention features are more interpretable when visualized. To explore the intrinsic mechanism of our integrated attention block, we visualized the attention maps of our model (**Figure** 5) when inferencing sample EP01-5 from $CASME$ Subject 1. The attention maps are firstly resized to 112×112. The model is pre-trained on the training set of the first cross-validation fold (leaving $CASME$ Subject 1 out).

Sample EP01-5 contains a positive micro-expression with human-annotated FACS Action Unit[3] 'AU12', which means the lip corner puller and its related facial regions are colored red on the AU picture. The validity of the integration of CBAM[29] and self-attention can be seen from the figure, as the spatial attention

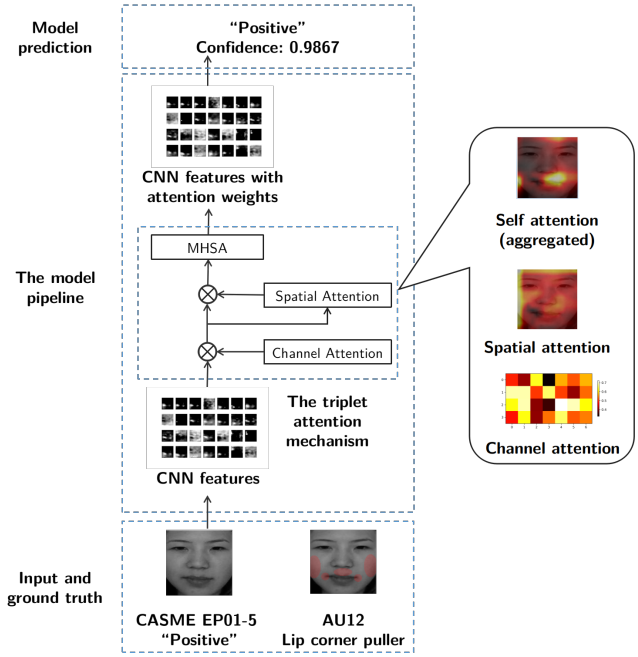


Fig. 5. The attention maps of Off-TANet when inferencing sample EP01-5. The CNN features are extracted by layers before the triplet attention mechanism. The spatial and self-attention maps are visualized by two heat maps. All the self-attention maps are aggregated by a channel-wise maximization to get exactly one attention map, and then the attention map is multiplicated on the spatial-attention map and visualized. Other details about the visualization process can be found in the source code.

map is general and vague, while the self-attention map only focuses on the most important facial areas. In addition, the self-attention map and the FACS AU facial regions correspond on the nasolabial fold, the right mouth corner, and the left cheekbone, and this shows the affinity between the Off-TANet triplet attention mechanism and human perception.

Despite the effectiveness of the fusion of channel attention and spatial attention is ensured by the experiments in [29], the lead-in of the self-attention module still needs more support phenomenons besides merely accuracy numbers. A possible explanation is that the self-attention enhanced CBAM[29] could further discriminate the spatial regions with significant features in comparison with the original CBAM[29]. This could mean that compared with the CBAM[29] spatial attention map, the attention map after the self-attention enhancement has greater matrix sparseness, as the attention weights for unimportant spatial positions are suppressed to zero. We applied two sparseness evaluation metrics in our experiment, soft 0-norm (the number of elements smaller than the threshold)

and Hoyer sparseness[6]. Their definitions are as follows:

$$Soft_t(x) = |\{i|x_i < t\}| \tag{13}$$

$$Hoyer(x) = \frac{\sqrt{n} - (\Sigma|x_i|)/\sqrt{\Sigma x_i^2}}{\sqrt{n} - 1} \tag{14}$$

where x denotes the matrix, x_i denotes its element and t denotes a threshold. We validate the mean value of these two indicators in multiple cross-validation folds (**Table** 5), and the sparseness assumption is confirmed.

Table 5. The sparseness of the self-attention map. Only folds with large test set sizes are contained.

Cross-val fold	**Test set size**	**Spatial attention**		**Self-attention**	
		$Soft_{0.01}$	$Hoyer$	$Soft_{0.01}$	$Hoyer$
1	22	0	0.0088	1470	0.1195
4	13	0	0.0154	1161	0.1768
6	23	0	0.0172	845	0.1196
41	31	0	0.0161	1036	0.1313

5 Conclusion

The integration of innovative neural network architectures and micro-expression recognition is an attractive topic. In our paper, we proposed a novel neural optical flow processor called Off-TANet for micro-expression sentiment classification. In this architecture based on a minimalist ResNet, a triplet attention mechanism is used to improve its classification performance. We test our model on a combined dataset with a LOSO protocol and showed that the UAR and UF1 of our design exceed the counterparts of other mainstream approaches. We also conduct a series of ablation experiments to ensure the validity of our design. We also give an possible interpretation for the intrinsic mechanism of the triplet attention module. Though this paper only explored the application of the triplet attention mechanism in micro-expression recognition, this design could be a general design for lightweight neural networks and we hope to release the potential of this simple yet powerful architecture on other tasks in our future work.

References

1. Davison, A.K., Lansley, C., Costen, N., Tan, K., Yap, M.H.: Samm: A spontaneous micro-facial movement dataset. IEEE transactions on affective computing 9(1), 116–129 (2016)
2. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
3. Ekman, P., Friesen, W.V.: Nonverbal leakage and clues to deception. Psychiatry 32(1), 88–106 (1969)
4. Gan, Y., Liong, S.T., Yau, W.C., Huang, Y.C., Tan, L.K.: Off-apexnet on micro-expression recognition system. Signal Processing: Image Communication 74, 129–139 (2019)
5. 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. pp. 770–778 (2016)
6. Hoyer, P.O.: Non-negative matrix factorization with sparseness constraints. Journal of machine learning research 5(9) (2004)
7. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7132–7141 (2018)
8. Huang, X., Zhao, G., Hong, X., Zheng, W., Pietikäinen, M.: Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns. Neurocomputing 175, 564–578 (2016)
9. Jaderberg, M., Simonyan, K., Zisserman, A., Kavukcuoglu, K.: Spatial transformer networks. arXiv preprint arXiv:1506.02025 (2015)
10. Lai, Z., Chen, R., Jia, J., Qian, Y.: Real-time micro-expression recognition based on resnet and atrous convolutions. Journal of Ambient Intelligence and Humanized Computing pp. 1–12 (2020)
11. Li, X., Pfister, T., Huang, X., Zhao, G., Pietikäinen, M.: A spontaneous micro-expression database: Inducement, collection and baseline. In: 2013 10th IEEE International Conference and Workshops on Automatic face and gesture recognition (fg). pp. 1–6. IEEE (2013)
12. Liong, S.T., Gan, Y., See, J., Khor, H.Q., Huang, Y.C.: Shallow triple stream three-dimensional cnn (ststnet) for micro-expression recognition. In: 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019). pp. 1–5. IEEE (2019)
13. Liong, S.T., See, J., Wong, K., Phan, R.C.W.: Less is more: Micro-expression recognition from video using apex frame. Signal Processing: Image Communication 62, 82–92 (2018)
14. Liu, Y., Du, H., Zheng, L., Gedeon, T.: A neural micro-expression recognizer. In: 2019 14th IEEE international conference on automatic face & gesture recognition (FG 2019). pp. 1–4. IEEE (2019)
15. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. arXiv preprint arXiv:2103.14030 (2021)
16. Lo, L., Xie, H.X., Shuai, H.H., Cheng, W.H.: Mer-gcn: Micro-expression recognition based on relation modeling with graph convolutional networks. In: 2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). pp. 79–84. IEEE (2020)

17. Pérez, J.S., Meinhardt-Llopis, E., Facciolo, G.: Tv-l1 optical flow estimation. Image Processing On Line 2013, 137–150 (2013)
18. Polikovsky, S., Kameda, Y., Ohta, Y.: Facial micro-expressions recognition using high speed camera and 3d-gradient descriptor (2009)
19. Qu, F., Wang, S.J., Yan, W.J., Li, H., Wu, S., Fu, X.: Cas (me) 2: A database for spontaneous macro-expression and micro-expression spotting and recognition. IEEE Transactions on Affective Computing 9(4), 424–436 (2017)
20. Russell, T.A., Chu, E., Phillips, M.L.: A pilot study to investigate the effectiveness of emotion recognition remediation in schizophrenia using the micro-expression training tool. British journal of clinical psychology 45(4), 579–583 (2006)
21. Russell, T.A., Green, M.J., Simpson, I., Coltheart, M.: Remediation of facial emotion perception in schizophrenia: concomitant changes in visual attention. Schizophrenia research 103(1-3), 248–256 (2008)
22. See, J., Yap, M.H., Li, J., Hong, X., Wang, S.J.: Megc 2019–the second facial micro-expressions grand challenge. In: 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019). pp. 1–5. IEEE (2019)
23. Srinivas, A., Lin, T.Y., Parmar, N., Shlens, J., Abbeel, P., Vaswani, A.: Bottleneck transformers for visual recognition. arXiv preprint arXiv:2101.11605 (2021)
24. Swart, M., Kortekaas, R., Aleman, A.: Dealing with feelings: characterization of trait alexithymia on emotion regulation strategies and cognitive-emotional processing. PloS one 4(6), e5751 (2009)
25. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. arXiv preprint arXiv:1706.03762 (2017)
26. Wang, C., Peng, M., Bi, T., Chen, T.: Micro-attention for micro-expression recognition. Neurocomputing 410, 354–362 (2020)
27. Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., Wang, X., Tang, X.: Residual attention network for image classification. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3156–3164 (2017)
28. Wang, X., Girshick, R., Gupta, A., He, K.: Non-local neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7794–7803 (2018)
29. Woo, S., Park, J., Lee, J.Y., Kweon, I.S.: Cbam: Convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV). pp. 3–19 (2018)
30. 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 9(1), e86041 (2014)
31. Yan, W.J., Wu, Q., Liu, Y.J., Wang, S.J., Fu, X.: Casme database: a dataset of spontaneous micro-expressions collected from neutralized faces. In: 2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG). pp. 1–7. IEEE (2013)
32. Yap, M.H., See, J., Hong, X., Wang, S.J.: Facial micro-expressions grand challenge 2018 summary. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). pp. 675–678. IEEE (2018)
33. Zhang, M., Fu, Q., Chen, Y.H., Fu, X.: Emotional context influences micro-expression recognition. PloS one 9(4), e95018 (2014)
34. Zhou, L., Mao, Q., Xue, L.: Dual-inception network for cross-database micro-expression recognition. In: 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019). pp. 1–5. IEEE (2019)