

图4 微表情的原始图及灰度图序列

本文用仿射变换进行数据增强. 所的方法包括平移、翻转和旋转 3 类. 具体地, 对于水平翻转、垂直翻转; 旋转角度 90° 、旋转角度 180° 、使用旋转角度 270° , 单种方式有 5 种, 同类仿射变换不混合的混合方式有 6 种, 水平翻转加垂直翻转本身及和旋转三个角度混合有 4 种, 共 15 种仿射变换.

3.2 光流特征提取

光流特征提取是对微表情视频进行光流估计, 提取微表情视频的低级特征. 对于微表情识别的视频序列, 用光流法计算面部区域的微小移动, 计算工具使用 OpenCV 库^[24], 得到与原图大小相等的双通道图像, 双通道分别表示强度和方向. 为了更直观地可视化光流, 可以使用 Munsell 颜色系统^[30] 将强度和方向矩阵转换为可视化图像, 使用该颜色系统的微表情光流分布如图 5 所示. 得到的两个通道的光流估计序列视为低级特征, 和一个通道的灰度图像序列合并为 3 个通道图像序列, 共同输入 C3D 的输入层中, C3D 将自动提取出时域和空域上的特征并进行最后的分类.

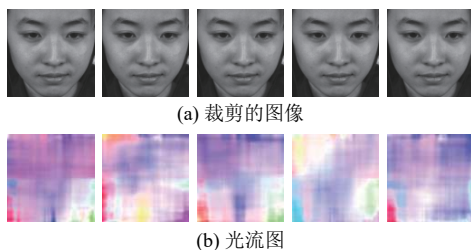


图5 使用 Munsell 颜色系统表示的微表情光流分布

3.3 C3D 网络结构设计

在深度学习中, 一个好的神经网络结构在训练过程中起着重要作用, 良好的结构模型可以有效地提取特征. 因此, C3D 结构上的参数需要适当地设置, 包括层数、每层的核种类等. 本文所提出的 C3D 结构如表 1 所示. 其中, 网络结构的卷积层核是 $(3 \times 3 \times 3)$. 表 1 中: (1) Conv, Pool 和 F 分别是卷积层, 最大池化层和完全连接层的缩写. (2) 第一层 Conv1 的输入大小为 $(96 \times$

$96 \times 95 \times 3)$, 其中 95 表示一个视频在时域上的大小, (96×96) 表示一个视频在空域上的大小, 3 表示输入视频的通道, 包括两个光流通道和一个上一帧灰度帧. (3) 最后一层 F2 用于分类, 本文中的微表情类数为 5, 因此输出大小为 (5×1) . 这个网络结构设计借鉴了经典的深度学习网络 VGG^[31] (Visual Geometry Group) 的优点, VGG 在图像处理方面表现出良好的性能. 它具有 3 个优点: (1) VGG 使用多个小卷积核而不使用过多大卷积核, 一方面可以在获得相同大小的特征映射的情况下捕获到更多的空间上下文信息, 但使用较小的卷积核时, 使用的参数和计算量较少. 另一方面, 因为使用更小的核, 意味着要使用更多的滤波器, 即能够使用更多的激活函数, 从而使卷积网络学习到更具区分性的映射函数. (2) 层组的设计. 两个卷积层后面紧接着一个池化层, 其中前两个卷积层更注重局部特征, 适合于需要提取微小局部信息的微表情识别. (3) 第 2 组层组比第 1 组层组多一个卷积层, 这意味着可以在第 1 组层组的基础上进一步细化特征的学习.

表 1 C3D 网络结构参数

网络层	输入大小	核(kernel)	stride	Pad	本层输出大小
Input	$96 \times 96 \times 95 \times 3$	-	-	-	$96 \times 96 \times 95 \times 3$
Conv1	$96 \times 96 \times 95 \times 3$	$3 \times 3 \times 3$	1	0	$94 \times 94 \times 94 \times 16$
Pool1	$94 \times 94 \times 94 \times 16$	$2 \times 2 \times 2$	2	0	$47 \times 47 \times 47 \times 16$
Conv2a	$47 \times 47 \times 47 \times 16$	$7 \times 7 \times 7$	1	0	$41 \times 41 \times 41 \times 32$
Conv2b	$41 \times 41 \times 41 \times 32$	$7 \times 7 \times 7$	1	0	$35 \times 35 \times 35 \times 32$
Pool2	$35 \times 35 \times 35 \times 32$	$2 \times 2 \times 2$	2	0	$17 \times 17 \times 17 \times 32$
Conv3a	$17 \times 17 \times 17 \times 32$	$3 \times 3 \times 3$	1	0	$15 \times 15 \times 15 \times 64$
Conv3b	$15 \times 15 \times 15 \times 64$	$3 \times 3 \times 3$	1	0	$13 \times 13 \times 13 \times 64$
Pool3	$13 \times 13 \times 13 \times 64$	$2 \times 2 \times 2$	2	0	$6 \times 6 \times 6 \times 64$
F1	$96 \times 96 \times 95 \times 3$	-	-	-	128×1
F2	$96 \times 96 \times 95 \times 3$	-	-	-	5×1

4 实验

4.1 实验参数说明

网络模型的学习过程由 Keras 编码实现. 参数是经反复试验决定的. 本文中, 初始学习率设置为 0.01. 对于 C3D, 训练 epoch 设定为 160. 对于损失函数, 使用均方误差 (Mean Square Error, MSE), 均方误差损失函数是使用最广泛的函数, 并且在大部分情况下, 均方误差有着不错的性能, 因此被用作损失函数的基本衡量指标. 实验的主要硬件设备是两块 NVIDIA Titan X GPU, 编程语言使用 Python.

4.2 实验结果和分析

在本节中, 我们通过评估本文提出的方法在 CASME

II数据集的分类准确率,并与其他现有方法进行了比较,包括现有的手工描述特征方法和深度学习方法。

4.2.1 和现有方法的对比

由于留一受试者交叉验证方法(LOSO)能防止学习过程中的主体偏差^[15];因此,我们的实验采用LOSO交叉验证法。在此情景下,我们将本文所提出方法和其他现有的方法进行比较,包括手工描述特征方法和深度学习方法。

所提出方法的识别准确率比较如表2所示。如表中所示,所提出的方法优于其他现有方法。与手工描述特征的方法相比,深度学习方法通过调整参数和权重,能够自动学习特征并在训练期间优化模型。深度学习方法尽管更依赖于训练样本的数量,但这个问题可以通过数据增强来解决,通过逐层学习样本,获取到深层次的特征。如表2所示的基于深度学习的方法总体比手工描述特征方法表现更好。特别地,本文所提出的方法结果比手工描述特征方法中的最佳方法高约4%,这表明本文所提出的方法作为一种深度学习方法,能够自动提取特征,省去了人工寻找特征的步骤,也提高了识别准确率。

表2 本文提出的方法与现有方法的微表情识别准确率比较

方法类别	方法名称	准确率 (%)
手工描述特征方法	LBP-TOP ^[22]	63.41
	STLBP-IP ^[7]	59.51
	STCLQ ^[6]	58.39
	MDMO ^[10]	52.12
	FDM ^[9]	41.96
	MMFL ^[8]	59.81
深度学习方法	Bi-WOOF ^[11]	59.26
	SDF ^[15]	47.30
	文献[16]提出的方法	60.98
	C3D(本文)	61.34
	本文所提出的方法	67.53

4.2.2 光流法对微表情识别的影响分析

本文所提出的方法是通过计算光流获取低级特征,在时间维度上提取相邻帧上的强度和方向的特征,以便捕获更多的动态信息,然后对C3D进行训练,提取高级特征,实现微表情的自动识别。

从表2可以看到,本文所提出的方法比C3D的识别准确率高了6.19%。即光流法贡献了6.19%的识别正确率。原因是微表情视频是一个动态时域上出现动作变化的视频,光流法能通过计算出微表情的微小运动的大小和方向,逐帧地提取微表情的动态特征,捕捉

到更多的动作信息,从而提高识别准确率。

5 结束语

本文提出了结合C3D与光流法的微表情自动识别方法,通过光流法逐帧提取微表情的动态信息,得到的光流序列和原始灰度序列输入C3D网络,通过C3D提取时域和空域上的特征,同时捕捉微表情的动态信息。实验中,为了满足大量的深度学习数据训练需要,采用数据增强策略,扩大了微表情数据规模,防止深度学习网络容易过拟合。在开放的微表情数据集CASME II上进行了模拟实验,实验表明所提出的方法提高了微表情识别准确率,准确率达到67.53%。

参考文献

- 1 Corneanu CA, Simón MO, Cohn JF, *et al.* Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 38(8): 1548–1568. [doi: 10.1109/TPAMI.2016.2515606]
- 2 Russell TA, Chu E, Phillips ML. A pilot study to investigate the effectiveness of emotion recognition remediation in schizophrenia using the micro-expression training tool. *British Journal of Clinical Psychology*, 2006, 45(4): 579–583. [doi: 10.1348/014466505X90866]
- 3 Pfister T, Li XB, Zhao GY, *et al.* Recognising spontaneous facial micro-expressions. *Proceedings of 2011 International Conference on Computer Vision*. Barcelona, Spain. 2011. 1449–1456.
- 4 Wang YD, See J, Phan RCW, *et al.* LBP with six intersection points: Reducing redundant information in LBP-TOP for micro-expression recognition. *Proceedings of the 12th Asian Conference on Computer Vision*. Singapore. 2015. 525–537.
- 5 Wang YD, See J, Phan RCW, *et al.* Efficient spatio-temporal local binary patterns for spontaneous facial micro-expression recognition. *PLoS One*, 2015, 10(5): e0124674. [doi: 10.1371/journal.pone.0124674]
- 6 Huang XH, Zhao GY, Hong XP, *et al.* Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns. *Neurocomputing*, 2016, 175: 564–578. [doi: 10.1016/j.neucom.2015.10.096]
- 7 Huang XH, Wang SJ, Zhao GY, *et al.* Facial micro-expression recognition using spatiotemporal local binary pattern with integral projection. *Proceedings of the IEEE International Conference on Computer Vision Workshop*. Santiago, Chile. 2015. 1–9.
- 8 He JC, Hu JF, Lu X, *et al.* Multi-task mid-level feature

- learning for micro-expression recognition. *Pattern Recognition*, 2017, 66: 44–52. [doi: [10.1016/j.patcog.2016.11.029](https://doi.org/10.1016/j.patcog.2016.11.029)]
- 9 Xu F, Zhang JP, Wang JZ. Microexpression identification and categorization using a facial dynamics map. *IEEE Transactions on Affective Computing*, 2017, 8(2): 254–267. [doi: [10.1109/TAFFC.2016.2518162](https://doi.org/10.1109/TAFFC.2016.2518162)]
- 10 Liu YJ, Zhang JK, Yan WJ, *et al.* A main directional mean optical flow feature for spontaneous micro-expression recognition. *IEEE Transactions on Affective Computing*, 2016, 7(4): 299–310. [doi: [10.1109/TAFFC.2015.2485205](https://doi.org/10.1109/TAFFC.2015.2485205)]
- 11 Liong ST, See J, Wong K, *et al.* Less is more: Micro-expression recognition from video using apex frame. *Signal Processing: Image Communication*, 2018, 62: 82–92. [doi: [10.1016/j.image.2017.11.006](https://doi.org/10.1016/j.image.2017.11.006)]
- 12 Huang WB, Fan LJ, Harandi M, *et al.* Toward efficient action recognition: Principal backpropagation for training two-stream networks. *IEEE Transactions on Image Processing*, 2019, 28(4): 1773–1782. [doi: [10.1109/TIP.2018.2877936](https://doi.org/10.1109/TIP.2018.2877936)]
- 13 Young T, Hazarika D, Poria S, *et al.* Recent trends in deep learning based natural language processing. *IEEE Computational Intelligence Magazine*, 2018, 13(3): 55–75. [doi: [10.1109/MCI.2018.2840738](https://doi.org/10.1109/MCI.2018.2840738)]
- 14 Xiong W, Wu L, Alleva F, *et al.* The Microsoft 2017 conversational speech recognition system. *Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing*. Calgary, AB, Canada. 2018. 5934–5938.
- 15 Patel D, Hong XP, Zhao GY. Selective deep features for micro-expression recognition. *Proceedings of 2016 23rd International Conference on Pattern Recognition*. Cancun, Mexico. 2016. 2258–2263.
- 16 Kim DH, Baddar WJ, Ro YM. Micro-expression recognition with expression-state constrained spatio-temporal feature representations. *Proceedings of the 24th ACM International Conference on Multimedia*. Amsterdam, the Netherlands. 2016. 382–386.
- 17 Peng M, Wang CY, Chen T, *et al.* Dual temporal scale convolutional neural network for micro-expression recognition. *Frontiers in Psychology*, 2017, 8: 1745. [doi: [10.3389/fpsyg.2017.01745](https://doi.org/10.3389/fpsyg.2017.01745)]
- 18 Khor HQ, See J, Phan RCW, *et al.* Enriched long-term recurrent convolutional network for facial micro-expression recognition. *Proceedings of 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition*. Xi'an, China. 2018. 667–674.
- 19 Tran D, Bourdev L, Fergus R, *et al.* Learning spatiotemporal features with 3D convolutional networks. *Proceedings of the IEEE International Conference on Computer Vision*. Santiago, Chile. 2015. 4489–4497.
- 20 Barron JL, Fleet DJ, Beauchemin SS. Performance of optical flow techniques. *International Journal of Computer Vision*, 1994, 12(1): 43–77. [doi: [10.1007/BF01420984](https://doi.org/10.1007/BF01420984)]
- 21 Negahdaripour S. Revised definition of optical flow: Integration of radiometric and geometric cues for dynamic scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, 20(9): 961–979. [doi: [10.1109/34.713362](https://doi.org/10.1109/34.713362)]
- 22 Yan WJ, Li XB, Wang SJ, *et al.* CASME II: An improved spontaneous micro-expression database and the baseline evaluation. *PLoS One*, 2014, 9(1): e86041. [doi: [10.1371/journal.pone.0086041](https://doi.org/10.1371/journal.pone.0086041)]
- 23 Ji SW, Xu W, Yang M, *et al.* 3D convolutional neural networks for human action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(1): 221–231. [doi: [10.1109/TPAMI.2012.59](https://doi.org/10.1109/TPAMI.2012.59)]
- 24 Farnebäck G. Two-frame motion estimation based on polynomial expansion. *Proceedings of the 13th Scandinavian Conference on Image Analysis*. Halmstad, Sweden. 2003. 363–370.
- 25 Li XB, Pfister T, Huang XH, *et al.* A spontaneous micro-expression database: Inducement, collection and baseline. *Proceedings of 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*. Shanghai, China. 2013. 1–6.
- 26 Yan WJ, Wu Q, Liu YJ, *et al.* CASME database: A dataset of spontaneous micro-expressions collected from neutralized faces. *Proceedings of 2013 10th IEEE International Conference and Workshops on Automatic face and Gesture Recognition (FG)*. Shanghai, China. 2013. 1–7.
- 27 Kazemi V, Sullivan J. One millisecond face alignment with an ensemble of regression trees. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, OH, USA. 2014. 1867–1874.
- 28 Zhou ZH, Zhao GY, Pietikäinen M. Towards a practical lipreading system. *Proceedings of CVPR 2011*. Providence, RI, USA. 2011. 137–144.
- 29 Zhong Z, Zheng L, Zheng ZD, *et al.* CamStyle: A novel data augmentation method for person re-identification. *IEEE Transactions on Image Processing*, 2019, 28(3): 1176–1190. [doi: [10.1109/TIP.2018.2874313](https://doi.org/10.1109/TIP.2018.2874313)]
- 30 Gargi U, Kasturi R, Strayer SH. Performance characterization of video-shot-change detection methods. *IEEE Transactions on Circuits and Systems for Video Technology*, 2000, 10(1): 1–13. [doi: [10.1109/76.825852](https://doi.org/10.1109/76.825852)]
- 31 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *The 3rd International Conference on Learning Representations*. San Diego, CA, USA. 2015. 1–5.