Pain Intensity Estimation using Spatiotemporal Facial Features

Anonymous Author(s)
Affiliation
Address
email

Abstract

Pain assessment is a hard, subjective, and critical problem in many medical situations. Thus, many computational approaches have been developed for pain detection and estimation using different types of data [1, 2, 3, 4, 5, 6]. We propose a guided-learning by warping the appearance surround the facial action units of pain (AUs). Sequences are processed to extract the temporal correspondence of facial features. Each sequence is generated from the original videos and must represent a single-stimulus effect in a short period, so we develop generation policies. Experimental results on the publicly available UNBC-McMaster database have demonstrated that our approach overcomes the-state-of-the-art.

1 Methodology

1.1 Dataset description

The UNBC-McMaster dataset [7] records 129 patients self-identified with chronic shoulder pain. In each video, the patient moves one arm, then, each frame is evaluated by the PSPI metric. Additionally, the dataset includes 66 facial landmarks per frame computed by an Active Appearance Model (AAM). Each sequence presents several disturbances in its pain level even though there is just one pain stimulus — higher the pain intensity, shorter its duration due to the patient's reflexes.

1.2 Pre-Processing

We apply a three-step pipeline: masking, frontalization, and resize. First, we calculate the facial landmarks to apply the convex hull algorithm for masking. The frontalization aims to solve the camera perspective error by estimating the projection matrix [8]. We use frontal-view markers from a pre-trained 3D-model [9] as a reference to calculate the camera matrix and the projection matrix. The projection matrix maps the original image to get the canonical normalized appearance, which undergoes a smooth symmetry process. We select the landmarks which surround the facial action units of pain to preserve the original features after frontalization.

1.3 Data Balancing

The unbalance affects every distribution, including samples per patient and pain levels. Sequence balancing has two stages; in the first one, the sequences split to generate new sub-sequences. In the second stage, we design downsampling and data augmentation policies considering that the label of a sub-sequence is its last frame's label.

Sub-sequence generation: The sequences split into a-length subsequences, considering a single-stimulus response cycle and painless segments. At the beginning of a sub-sequence, there may be b

Submitted to 33rd Conference on Neural Information Processing Systems (NeurIPS 2019). Do not distribute.

painless frames. The sub-sequences generation steps are: (1) the sequences split into single-stimulus response cycles and painless sub-sequences; (2) in case that a new cycle is a replica then it is attached to the previous period, however, if gradients change abruptly, then it is considered as a new cycle; (3) eliminate the sub-sequences with sizes smaller than a; and, (4) multiple sub-sequences are made using a displacement window over the previous fragments.

Balancing policies Let $M_{[p,l]}$ a the data distribution matrix over the patients and pain levels, being $\mathbb{P}$ and $\mathbb{L}$ the set of patients and levels, respectively. $T_{[p,l]}$ is the scaling matrix that balance $M_{[p,l]}$, being τ_{max} its fixed maximum value. When τ is less than 1, then downsampling is done with a probability of τ; else, augmentation policies run with same odds, except for the facial affine deformations. The policies include rotations ($3°, 6°$ and $9°$), vertical flipping and facial affine deformations. Facial deformations strongly affect the feature vector depending on its magnitude and location. By orienting the distortions, we seek to reinforce the pain features representation. We use Delaunay triangulation and piecewise-affine warping for the facial deformations. Deformations intensity depends on the landmark's relationship with the facial action units of pain proportionally from 0.2% to 10%.

1.4 Proposed Architecture

Figure 1 illustrates the spatiotemporal architecture. We employ the `fc6` layer outputs of a fine-tuned `VGG_faces` [10] as feature vectors due its temporal invariability. The fine-tuning handles an SGD optimizer with $\beta = 0.9$, $\alpha = 1e-3$ and a mini-batch size of 100. A two-layer GRU architecture correlates the feature vectors of the sequence. GRU units face the gradient vanishing problem in short sequences and reduce training time. Then, the data split randomly into 80% as the training set and 20% as the testing set using an RMSprop optimizer with $\alpha = 1e-4$.

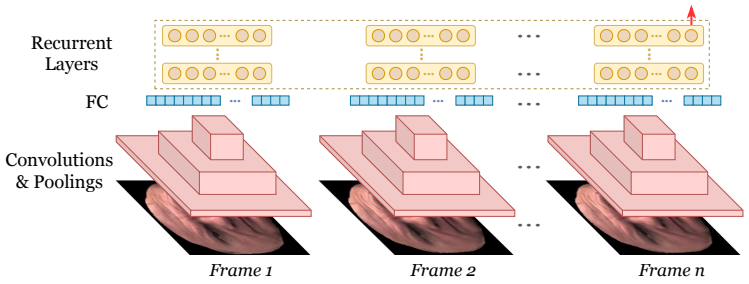

Figure 1: An overview of the proposed spatiotemporal architecture. Red: A pre-trained CNN perform the spatial extraction of features vectors (blue). Yellow: a two-layer GRU network explorers the temporal correspondence among the features vectors.

2 Results and Comments

Table 1 shows the comparison of the-state-of-the-art results for both analyses. Our results overcome previous works almost in every metric, except for the geometric approach raised by Rathee et al. [11]. Data distribution has a high-impact over the features representation; hence, data balancing acquires quite an importance. Our proposal faces the unbalance by structured augmentation policies, alongside a fine-tuning step for the spatial analysis and a pre-processing stage to ease the spatial analysis.

	MSE	PCC	ICC	ACC
Florea et al. [12]	1.18	0.55	-	-
Kaltwang et al. [13]	1.39	0.59	0.50	-
Rathee et al. [11]	-	-	-	**0.96**
Our	**0.634**	**0.692**	**0.529**	0.834

	MSE	PCC	ICC	ACC
Zhou et al.[1]	1.54	0.65	-	-
Nasrollahi et al.[2]	-	-	-	0.619
Rodriguez et al.[3]	0.74	**0.78**	0.45	-
Our	**0.622**	0.687	**0.615**	**0.854**

Table 1: Comparison of the-state-of-the-art metrics at frames-level (left) and sequences-level (right).

In the future, we are planning to untangle high-level expression and identity features from low-level pain features before the temporal analysis.

References

[1] Jing Zhou, Xiaopeng Hong, Fei Su, and Guoying Zhao. Recurrent convolutional neural network regression for continuous pain intensity estimation in video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 84–92, 2016.

[2] Kamal Nasrollahi, Taisi Telve, Sergio Escalera, Jordi Gonzalez, Thomas B Moeslund, Pejman Rasti, and Gholamreza Anbarjafari. Spatio-temporal pain recognition in cnn-based super-resolved facial images. In *Video Analytics. Face and Facial Expression Recognition and Audience Measurement: Third International Workshop, VAAM 2016, and Second International Workshop, FFER 2016, Cancun, Mexico, December 4, 2016, Revised Selected Papers*, volume 10165, page 151. Springer, 2017.

[3] Pau Rodriguez, Guillem Cucurull, Jordi Gonzalez, Josep M Gonfaus, Kamal Nasrollahi, Thomas B Moeslund, and F Xavier Roca. Deep pain: Exploiting long short-term memory networks for facial expression classification. *IEEE transactions on cybernetics*, (99):1–11, 2017.

[4] Daniel Lopez Martinez, Ognjen Rudovic, and Rosalind Picard. Personalized automatic estimation of self-reported pain intensity from facial expressions. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on*, pages 2318–2327. IEEE, 2017.

[5] Mohammad A Haque, Ruben B Bautista, Fatemeh Noroozi, Kaustubh Kulkarni, Christian B Laursen, Ramin Irani, Marco Bellantonio, Sergio Escalera, Golamreza Anbarjafari, Kamal Nasrollahi, et al. Deep multimodal pain recognition: A database and comparison of spatio-temporal visual modalities. In *Automatic Face & Gesture Recognition (FG 2018), 2018 13th IEEE International Conference on*, pages 250–257. IEEE, 2018.

[6] Dennis C Turk and Ronald Melzack. *Handbook of pain assessment*. Guilford Press, 2011.

[7] Patrick Lucey, Jeffrey F Cohn, Kenneth M Prkachin, Patricia E Solomon, and Iain Matthews. Painful data: The unbc-mcmaster shoulder pain expression archive database. In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 57–64. IEEE, 2011.

[8] Sandipan Banerjee, Joel Brogan, Janez Krizaj, Aparna Bharati, Brandon Richard Webster, Vitomir Struc, Patrick J Flynn, and Walter J Scheirer. To frontalize or not to frontalize: Do we really need elaborate pre-processing to improve face recognition? In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 20–29. IEEE, 2018.

[9] Tal Hassner, Shai Harel, Eran Paz, and Roee Enbar. Effective face frontalization in unconstrained images. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[10] Omkar M Parkhi, Andrea Vedaldi, Andrew Zisserman, et al. Deep face recognition. In *BMVC*, volume 1, page 6, 2015.

[11] Neeru Rathee and Dinesh Ganotra. A novel approach for pain intensity detection based on facial feature deformations. *Journal of Visual Communication and Image Representation*, 33:247–254, 2015.

[12] Corneliu Florea, Laura Florea, and Constantin Vertan. Learning pain from emotion: transferred hot data representation for pain intensity estimation. In *European Conference on Computer Vision*, pages 778–790. Springer, 2014.

[13] Sebastian Kaltwang, Ognjen Rudovic, and Maja Pantic. Continuous pain intensity estimation from facial expressions. In *International Symposium on Visual Computing*, pages 368–377. Springer, 2012.