Spatio-Temporal Pain Recognition in CNN-based Super-Resolved Facial Images

Bellantonio, Marco; Haque, Mohammad Ahsanul; Rodriguez, Pau; Nasrollahi, Kamal; Telve, Taisi; Guerrero, Sergio Escalera; Gonzàlez, Jordi; Moeslund, Thomas B.; Rasti, Pejman; Anbarjafari, Gholamreza

Published in:
Video Analytics

DOI (link to publication from Publisher):
10.1007/978-3-319-56687-0_13

Publication date:
2017

Document Version
Accepted author manuscript, peer reviewed version

Spatio-Temporal Pain Recognition in CNN-based Super-Resolved Facial Images

Marco Bellantonio[1], Mohammad A. Haque[2], Pau Rodriguez[1], Kamal Nasrollahi[2], Taisi Telve[3], Sergio Escarela[1], Jordi Gonzalez[1], Thomas B. Moeslund[2], Pejman Rasti[3], and Gholamreza Anbarjafari[3]

[1] Computer Vision Center (UAB), University of Barcelona, Spain
marco.bellantonio@est.fib.upc.edu,sergio@maia.ub.es,{prodriguez,poal}@cvc.uab.es
[2] Visual Analysis of People (VAP) Laboratory, Aalborg University, Denmark
{mah,kn,tbm}@create.aau.dk
[3] iCV Research Group, Institute of Technology, University of Tartu, Tartu, Estonia
{tt,pejman,shb}@icv.tuit.ut.ee

Abstract Automatic pain detection is a long expected solution to a prevalent medical problem of pain management. This is more relevant when the subject of pain is young children or patients with limited ability to communicate about their pain experience. Computer vision-based analysis of facial pain expression provides a way of efficient pain detection. When deep machine learning methods came into the scene, automatic pain detection exhibited even better performance. In this paper, we figured out three important factors to exploit in automatic pain detection: spatial information available regarding to pain in each of the facial video frames, temporal axis information regarding to pain expression pattern in a subject video sequence, and variation of face resolution. We employed a combination of convolutional neural network and recurrent neural network to setup a deep hybrid pain detection framework that is able to exploit both spatial and temporal pain information from facial video. In order to analyze the effect of different facial resolutions, we introduce a super-resolution algorithm to generate facial video frames with different resolution setups. We investigated the performance on the publicly available UNBC-McMaster Shoulder Pain database. As a contribution, the paper provides novel and important information regarding to the performance of a hybrid deep learning framework for pain detection in facial images of different resolution.

Keywords: Super-Resolution, Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Pain detection

1 INTRODUCTION

Pain is a prevalent medical problem that reveals as an unpleasant experience and needs to be managed effectively as a moral and professional responsibility [5]. Traditionally, pain is measured by 'self-report'. However, self-reported pain level

assessment requires cognitive, linguistic and social competencies of the affected person. These aspects make self-report unfeasible to use for young children and patients with limited ability to communicate [38]. Thus, the notion of computer vision-based automatic pain level assessment was introduced [32,33].

Facial pain expression can be considered as a subset of facial expression and expresses emotion valley regarding to experiencing pain [2]. It can also provide information about the severity of pain that can be assessed by using the Facial Action Coding System (FACS) coding from [6,54]. For a long time the FACS has been used to measure facial expression appearance and intensity. Thus, vision-based approaches came into the scene to measure pain by using features from facial appearance change. Prkachin first reported the consistency of facial pain expressions for different pain modalities in [47] and then together with Solomon developed a pain metric called Prkachin and Solomon Pain Intensity (PSPI) scale based on FACS in [49].

The task of assessing the pain level from facial image or video is rather challenging. A substantial body of literature has been produced in the recent years to address the challenges [3,10,30,48,50]. A glimpse of the reason why pain level detection is difficult can be found in Figure 1 [14]. From the facial images in the figure, we can see that the pain and non-pain frames may not present enough visual difference; however, the self-report tells a different story about having pain and non-pain status. The challenges also increase in the presence of external factors like 'smiling in pain' phenomenon and gender difference (male's vs female's way of experiencing) to pain [29,55,31]. This in turns result to a non-linearly wrapped facial emotion levels in a high dimensional space [53].

Recent advances in facial video analysis using deep learning frameworks such as Convolutional Neural Networks (CNN) or Deep Belief Networks (DBN) provide the notion of realizing non-linear high dimensional compositions [51]. Deep learning architectures have been widely used in face recognition [44,36,19,57], facial expression recognition [58,25], emotion detection [51,22,24]. Pain level estimation using a deep learning framework was also proposed [59]. Employing deep learning framework for pain level assessment from facial video entails two kinds of information processing from facial video sequences: i) spatial information, ii) temporal information. Spatial information provides pain related information in the facial expressions of a single video frame. On the other hand, temporal information exhibits the relationship between pain expressions revealed in consecutive video frames.

While exploring spatial and temporal information from facial images, face quality (e.g. low face resolution) can also play important role as studied in [13,12,11]. The first limitation of the image resolution is created by the imaging acquisition devices or the imaging sensors [42]. The spatial resolution of the image capture is determined by the sensor size or the number of sensor elements. So, for increasing the spatial resolution of an imaging system, one of the easy ways is to increase the sensor density by reducing the sensor size. However, as the sensor size decreases, the amount of light incident on each sensor also decreases, causing the shot noise [42]. Also, the hardware cost of a sensor increases by making sensor

Figure 1: Pain and non-pain facial expression is sometimes very difficult to distinguish visually. Examples from the UNBC-McMaster shoulder pain database [39]. The pain frames are at the left and the non-pain frames are at the right.

density greater or corresponding image pixel density. Applying various signal processing tools is the other approach for enhancing the face resolution. One of the famous techniques is Super Resolution (SR). The basic idea behind SR methods is to obtain high resolution (HR) image from low resolution (LR) image or images [56,43]. Huang and Tsai [18] as pioneers of SR proposed a method in order to improve spatial resolution of satellite images of earth, where a large set of translated images of the same scene are available. They showed that the better restoration can be achieved rather than spline interpolation by using multiple offset images of the same scene and a proper registration. Since then, SR methods become common practice for many applications in different fields such as remote sensing [35], surveillance video [4,37], medical imaging such as ultrasound, magnetic resonance imaging (MRI), and computerized tomography (CT) scan[23,40,41,46].

The desire for HR stems from two principal application areas:

– Improvement of resolution for human interpretation: in these applications, human is ultimate goal for system. SR methods improve resolution and visual quality in captured image. For example, a doctor can diagnose or treat with image capture from outside and inside the patient's body.

– Helping representation for automatic machine perception: SR methods are used to improve the resolution and image quality, for facilitating the machine processing. SR methods are used in various problems such as optical character recognition (OCR) problem or machine face recognition [1,15,52,9].

In this paper, we investigate the plausibility of using a Recurrent Neural Network (RNN) [59] to exploit the temporal axis information from facial video using Long Short-Term Memory (LSTM) [16,8] to estimate pain level expression in the face. The RNN is fed with the features extracted by a CNN that explores spatial information. We employ a SR technique to generate super-resolved high-resolution images from low resolution faces and we employ the CNN+RNN based deep learning framework to observe the performance. We report our results through the publicly available challenging database called UNBC-McMaster Shoulder Pain database [39]. The major contribution of the paper are as follows:

– Analyzing the pain detection performance fluctuation due to facial image resolution.
– Determining the impact of employing SR techniques in pain expression detection.
– Employing a hybrid deep learning framework by combining CNN and RNN to exploit spatio-temporal information of pain in video sequences.

The rest of the paper is organized as follows. Section 2 describes the proposed methodology for pain level assessment. Section 3 presents the experimental environment and the obtained results. Section 4 contains the conclusions.

2 The Proposed Pain Detection Framework

In this section we first describe the facial pain-expression database to be used in our investigation. We then describe the procedure of generating facial images with different resolutions and, finally, the deep learning-based classification framework for the experiment.

2.1 The database

We use the UNBC-McMaster Shoulder Pain database collected by the researchers at McMaster University and University of Northern British Columbia [39]. The database contains facial video sequences of participants who had been suffering from shoulder pain and were performing a series of active and passive range of motion tests to their affected and unaffected limbs on multiple occasions. The database also contains FACS information of the video frames, self-reported pain scores in sequence level and facial landmark points obtained by an appearance model. The database was originally created by capturing facial videos from 129 participants (63 males and 66 females). The participant had a wide variety of occupations and ages. During data capturing the participants underwent eight standard range-of-motion tests: abduction, flexion, and internal and external

rotation of each arm separately. Participants' self-reported pain score along with offline independent observers rated pain intensity were recorded. At present, the UNBC-McMaster database contains 200 video sequences with 48398 FACS coded frames of 25 subjects.

2.2 Obtaining Pain-Expression Data with Varying Face Resolution

We created multiple datasets by obtaining the original images from the UNBC-McMaster database and then varying the resolutions by down-up sampling or SR algorithms. The down-up sampling was accomplished by simply down-sampling the original images and then up-sampling the down-sampled images to the same resolution of the original images by employing a cubic-interpolation.

In order to generate SR images, a state-of-the-art technique, namely example-based learning [27] is adopted. The work in [27] is an extension of [26] which uses kernel ridge regression in order to estimate the high-frequency details of the underlying HR image. Also a combination of gradient descent and kernel matching pursuit is considered and allows time-complexity to be kept to a moderate level. Actually the proposed method improves the SR method presented in [7]. In this algorithm, For a given set of training data points $(x_1, y_1), ..., (x_l, y_l) \subset \Re^M \times \Re^N$, the following regularized cost functional is minimized.

$$O\left(\left\{f^1, ..., f^N\right\}\right) = \sum_{i=1,...,N} \left(\frac{1}{2} \sum_{j=1,...,N} \left(f^i\left(x_j\right) - y_j^i\right)^2 + \frac{1}{2}\lambda \left\|f^i\right\|_H^2 \right) \quad (1)$$

where $y_j = \left[y_j^1, ..., y_j^N\right]$ and H is a reproducing kernel Hilbert space. Due to the reproducing property, the minimizer of equation 1 is expanded in kernel functions:

$$f^i\left(\cdot\right) = \sum_{j=1,...,l} a_j^i k\left(x_j, \cdot\right), for \quad i = 1, ..., N \quad (2)$$

where k is the generating kernel for H which, is choosen as a Gaussian kernel $\left(k\left(x, y\right) = exp\left(-\left\|x - y\right\|^2 / \sigma_k\right)\right)$. Equation 1 is the sum of individual convex cost functionals for each scalar-valued regressor and can be minimized separately. The final estimation of pixel value for an image location (x, y) is then obtained as the convex combination of candidates given in the form of a softmax :

$$Y\left(x, y\right) = \sum_{i=1,..N} w_i\left(x, y\right) Z\left(x, y, i\right) \quad (3)$$

where $w_i\left(x, y\right) = exp\left(-\frac{|d_i(x,y)|}{\sigma_C}\right) / \left[\sum_{j=1,...,N} exp\left(-\frac{|d_j(x,y)|}{\sigma_C}\right)\right]$ and Z is the initial SR image that is generated by a bicubic interpolation.
We use the down-sampled images as input to the SR algorithm and obtain the super-resolved images.

2.3 Deep Hybrid Classification Framework

We use a combination of CNN and RNN based hybrid framework to exploit both spatial and temporal information of facial pain expressions for pain detection. The hybrid pain detection framework is depicted in Figure 2. In order to extract discriminative facial features, we fine-tune `VGG_Faces` [45], a 16-layer pre-trained CNN with 2.6M facial images of 2.6K people. Concretely, we replace the last layer of the CNN by a randomly initialized fully-connected layer with the three pain levels to recognize, and set its learning rate as ten times the learning rate of the rest of the CNN.

Once, fine-tuned, we extract the features of the `fc7` layer of the fine-tuned model and use them as input to a Long-Short Term Memory (LSTM) Recurrent Neural Network (RNN) [17]. LSTMs are particular implementations of RNN that make use of the forget (f), input (i), and output (o) gates so as to solve the vanishing or exploding gradient problems, making them suitable for learning long-term time dependencies. These gates control the flow of information through the model by using point-wise multiplications and sigmoid functions σ, which bound the information flow between zero and one:

$$i(t) = \sigma(W_{(x\to i)}x(t) + W_{(h\to i)}h(t-1) + b_{(1\to i)}) \tag{4}$$

$$f(t) = \sigma(W_{(x\to f)}x(t) + W(h\to f)h(t-1) + b_{(1\to f)}) \tag{5}$$

$$z(t) = tanh(W_{(x\to c)}x(t)) + W_{(h\to c)}h(t-1) + b_{(1\to c)}) \tag{6}$$

$$c(t) = f(t)c(t-1) + i(t)z(t), \tag{7}$$

$$o(t) = \sigma(W_{(x\to o)}x(t) + W_{(h\to o)}h(t-1) + b_{(1\to o)}) \tag{8}$$

$$h(t) = o(t)tanh(c(t)), \tag{9}$$

where $z(t)$ is the input to the cell at time t, c is the cell, and h is the output. $W_{(x\to y)}$ are the weights from x to y. More detail can be found in the original implementation [34].

Labels are predicted sequence-wise, *i.e.* given a sequence of n frames $f_i \in \{f_1, ..., f_n\}$, the target prediction is the pain level of the f_n frame. Thus, training is set so that the information contained in the past frames is used in order to predict the current pain level. We optimize the LSTM with *Adam* [28] with an initial learning rate of 0.001 so as to alleviate the hyper-parameter tuning problem.

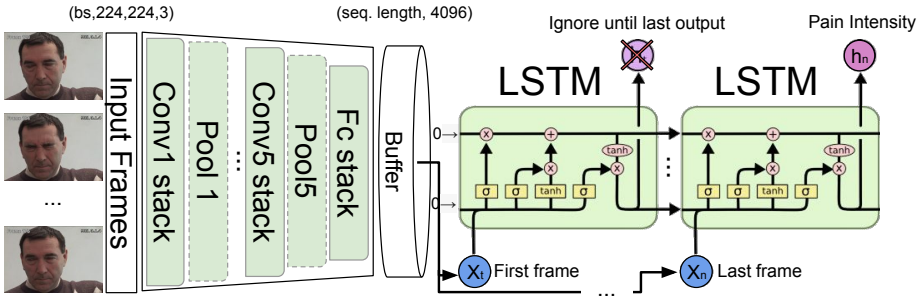

Figure 2: The block diagram of the deep hybrid classification framework based on a combination of CNN and RNN

3 Experimental Results and Discussions

3.1 Experimental Environment

As stated in the previous section, we evaluated the performance of pain detection in varying face resolution by employing the hybrid deep learning framework on the UNBC-McMaster Shoulder Pain database [39]. The video frames of the database showed patients who were suffering from shoulder pain while they were performing a series of active and passive range-of-motion tests. The pain indexes were computed by following Prkachin and Solomon Pain Intensity (PSPI) scale from [49] and the pain levels vary in the interval 0-16 based on the FACS codes. Following [21], we classified each pain index into three categories of no pain (pain index lower than 1), weak pain (pain index between 2 and 6) and strong pain (pain index greater than 6). The three categories have been balanced by dropping consecutive no-pain frames at the beginning and at the end of each video, or by discarding entire video sequences which do not contain pain.

We applied the down-up sampling and SR algorithm described in Section 2.2 to generate three experimental datasets. The first dataset was created by using the original images from the UNBC-McMaster database (also used by [20]). The second and third datasets were denoted by 'SR1/4' and 'SR1/2', and were created by employing down-up sampling with the values 1/4 and 1/2, respectively, on the first dataset. The fourth dataset were denoted by 'SR2', and was created by employing the SR algorithm from Section 2.2 on the down-sampled images with factor 1/2. The LSTM network was configured with 3 hidden-layers of 64 hidden-units each and a temporal window of 16 consecutive video frames. For the purpose of comparison, the experimental setup of the LSTM was kept fixed for all the experiments against the three datasets. The performances was estimated with leave-one-subject-out cross-validation protocol.

3.2 The Obtained Results

Table 1 shows the results of the proposed system against the three sets. Here we report the accuracy in percentage for each of the three categories, namely "No

pain", "Weak Pain" and "Strong Pain". From the experiments we can claim that the proposed method applied to super resolved images is crucial since it reaches better performance than using the plain down-sampled versions. The latter is denoted by the amount of improvement appearing in the pain detection rate using the super-resolved images as the subjects, while being compared against that of the LR ones. In other words, when recognizing the pains using the super-resolved images, a more powerful SR method leads to recognition rates closer to the case of considering the original ones. From the results we can see that pain detection

Table 1: Pain detection results for the four experimental datasets created from the UNBC-McMaster [39] database

Semantic Ground Truth	Pain Index	SR1/4	SR1/2	SR2
No Pain	0, 1	55.3%	62.22%	55.78%
Weak Pain	2, 3, 4, 5	73.1%	67.7%	75.94%
Strong Pain	>6	18.36%	5.86%	39.45%
F1-score		0.67	0.66	0.69
Total	0-16	62.43%	62.64%	65.34%

is much better in super-resolved images compared to down-sampled ones by a large margin in case of strong pain, while for the other two levels, namely no-pain and weak pain, the performances are slightly better. This is due to the fact that stronger pain (compared to weak or no pain case) imposes more changes on the face and these changes are more pronounced on super-resolved images hence the detection accuracy improves by far in the strong pain class compared to the other classes. In order to see how temporal information affects the final results, we provide the SR2 accuracy when using a linear classifier on the plain CNN features against the LSTM predictions, which aggregates the temporal information in Table 2. Here the results are reported for each subject in the considered data set. As it can be seen, temporal information improves the predictions for a large margin, a 16% in average, meaning that spatial features are not enough for determining the pain level on facial images. Thus, the temporal variation of the frames allows for finding higher level facial features, like FACS, which are central for predicting the PSPI pain score [50].

From Table 2 we notice that in two cases, specifically subject number 7 and subject number 8, the LSTM failed to improve the accuracy of the CNN. After a detailed study of the dataset, we notice that sometimes, for both subjects, the pain index changes very rapidly among consecutive frames. The same pattern occurs (in a lighter form) also for subject 6, which improvement in the accuracy is not as good as for the other subjects. In addition, subject 7 is the only one that contains only one video for the validation set, while subject 8 contains three videos, among which one very noisy video with only 20 frames. We think that the aforementioned differences could be the key problems which leads to such a different performance for different subjects.

Table 2: Comparison between CNN and LSTM performances on SR2 dataset (in accuracy %). The CNN relies on the information of a single frame, while the LSTM takes into account variations on the images in the temporal axis. As it can be seen, the LSTM enhances the accuracy prediction for all subjects, reaching a 16% in average.

Subj.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	**AVG**
CNN	40.9	50.3	52.0	50.7	50.3	42.5	**29.3**	**47.3**	50.4	44.0	50.0	51.3	51.9	30.2	45.8
LSTM	**58.0**	**61.5**	**63.0**	**65.6**	**82.5**	**48.0**	28.0	40.0	**81.0**	**65.4**	**82.0**	**66.0**	**65.5**	**60.5**	**61.9**

4 Conclusions

We investigated the performance of a recurrent deep learning framework trained against super-resolved high-resolution images for pain level classification. The system is a combination of CNN and a LSTM used to exploit both spatial and temporal information in videos. We evaluated our proposed method on UNBC_McMaster database by down sampling by different factors and by applying a super-resolution algorithm. From the experimental results of the pain detection performances we concluded that super-resolution and temporal information are key for obtaining good recognition results. Our experiments also showed that including deep temporal information within the model increases the generalization capabilities in discriminating among different levels of pain. Employing super-resolution techniques lead to an improvement of the performances in our pain detector. Down-sampling, on the other hand, worsen the system capabilities.

References

1. Capel, D., Zisserman, A.: Super-resolution enhancement of text image sequences. In: Pattern Recognition, 2000. Proceedings. 15th International Conference on. vol. 1, pp. 600–605. IEEE (2000)
2. Craig, K.D., Prkachin, K.M., Grunau, R.E.: The facial expression of pain. In: Handbook of Pain Assessment. Guilford Press (2011)
3. Craig, K.D., Hyde, S.A., Patrick, C.J.: Genuine, suppressed and faked facial behavior during exacerbation of chronic low back pain. Pain 46(2), 161 – 171 (1991)
4. Cristani, M., Cheng, D.S., Murino, V., Pannullo, D.: Distilling information with super-resolution for video surveillance. In: Proceedings of the ACM 2nd international workshop on Video surveillance & sensor networks. pp. 2–11. ACM (2004)
5. Debono, D.J., Hoeksema, L.J., Hobbs, R.D.: Caring for patients with chronic pain: Pearls and pitfalls. The Journal of the American Osteopathic Association 113(8), 620–627 (2013), +http://dx.doi.org/10.7556/jaoa.2013.023
6. Ekman, P., Friesen, W.: Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press (1978)
7. Freeman, W.T., Jones, T.R., Pasztor, E.C.: Example-based super-resolution. Computer Graphics and Applications, IEEE 22(2), 56–65 (2002)
8. Gers, F.A., Schmidhuber, J.A., Cummins, F.A.: Learning to forget: Continual prediction with lstm. Neural Comput. 12(10), 2451–2471 (Oct 2000)

9. Gunturk, B.K., Batur, A.U., Altunbasak, Y., Hayes, M.H., Mersereau, R.M.: Eigenface-domain super-resolution for face recognition. Image Processing, IEEE Transactions on 12(5), 597–606 (2003)
10. Hadjistavropoulos, T., LaChapelle, D.L., MacLeod, F.K., Snider, B., Craig, K.D.: Measuring movement-exacerbated pain in cognitively impaired frail elders. 16, 54–63 (2000)
11. Haque, M.A., Nasrollahi, K., Moeslund, T.B.: Real-time acquisition of high quality face sequences from an active pan-tilt-zoom camera. In: 2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance. pp. 443–448 (Aug 2013)
12. Haque, M.A., Nasrollahi, K., Moeslund, T.B.: Constructing facial expression log from video sequences using face quality assessment. In: 2014 International Conference on Computer Vision Theory and Applications (VISAPP). vol. 2, pp. 517–525 (Jan 2014)
13. Haque, M.A., Nasrollahi, K., Moeslund, T.B.: Quality-aware estimation of facial landmarks in video sequences. In: 2015 IEEE Winter Conference on Applications of Computer Vision. pp. 678–685 (Jan 2015)
14. Haque, M.A., Nasrollahi, K., Moeslund, T.B.: (submitted)pain expression as a biometric: Why patients' self-reported pain doesn't match with the objectively measured pain? In: 2017 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA) (Feb 2017)
15. Hennings-Yeomans, P.H., Baker, S., Kumar, B.V.: Recognition of low-resolution faces using multiple still images and multiple cameras. In: Biometrics: Theory, Applications and Systems, 2008. BTAS 2008. 2nd IEEE International Conference on. pp. 1–6. IEEE (2008)
16. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. 9(8), 1735–1780 (Nov 1997)
17. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation 9(8), 1735–1780 (1997)
18. Huang, T.S., Tsay, R.Y.: Multiple frame image restoration and registration. In: Advances in Computer Vision and Image Processing. pp. 317–339 (1984)
19. Huang, Z., Wang, R., Shan, S., Chen, X.: Face recognition on large-scale video in the wild with hybrid euclidean-and-riemannian metric learning. Pattern Recognition 48(10), 3113–3124 (2015)
20. Irani, R., Nasrollahi, K., Moeslund, T.B.: Pain recognition using spatiotemporal oriented energy of facial muscles. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 80–87 (June 2015)
21. Irani, R., Nasrollahi, K., Simon, M.O., Corneanu, C.A., Escalera, S., Bahnsen, C., Lundtoft, D.H., Moeslund, T.B., Pedersen, T.L., Klitgaard, M.L., Petrini, L.: Spatiotemporal analysis of rgb-d-t facial images for multimodal pain level recognition. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (June 2015)
22. Kahou, S.E., Bouthillier, X., Lamblin, P., Gulcehre, C., Michalski, V., Konda, K., Jean, S., Froumenty, P., Dauphin, Y., Boulanger-Lewandowski, N., Chandias Ferrari, R., Mirza, M., Warde-Farley, D., Courville, A., Vincent, P., Memisevic, R., Pal, C., Bengio, Y.: Emonets: Multimodal deep learning approaches for emotion recognition in video. Journal on Multimodal User Interfaces 10(2), 99–111 (2016)
23. Kennedy, J.A., Israel, O., Frenkel, A., Bar-Shalom, R., Azhari, H.: Super-resolution in pet imaging. Medical Imaging, IEEE Transactions on 25(2), 137–147 (2006)

24. Khorrami, P., Paine, T.L., Brady, K., Dagli, C., Huang, T.S.: How deep neural networks can improve emotion recognition on video data. In: 2016 IEEE International Conference on Image Processing (ICIP). pp. 619–623 (Sept 2016)
25. Kim, B.K., Roh, J., Dong, S.Y., Lee, S.Y.: Hierarchical committee of deep convolutional neural networks for robust facial expression recognition. Journal on Multimodal User Interfaces 10(2), 173–189 (2016)
26. Kim, K.I., Kim, D., Kim, J.H.: Example-based learning for image super-resolution (2004)
27. Kim, K.I., Kwon, Y.: Example-based learning for single-image super-resolution. In: Pattern Recognition, pp. 456–465. Springer (2008)
28. Kingma, D., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
29. Kunz, M., Gruber, A., Lautenbacher, S.: Sex differences in facial encoding of pain. The Journal of Pain 7(12), 915 – 928 (2006)
30. Kunz, M., Mylius, V., Schepelmann, K., Lautenbacher, S.: On the relationship between self-report and facial expression of pain. The Journal of Pain 5(7), 368 – 376 (2004)
31. Kunz, M., Prkachin, K., Lautenbacher, S.: Smiling in Pain: Explorations of Its Social Motives. Pain Research and Treatment 2013, e128093 (Aug 2013)
32. Kunz, M., Scharmann, S., Hemmeter, U., Schepelmann, K., Lautenbacher, S.: The facial expression of pain in patients with dementia. Pain 133(1-3), 221–228 (December 2007)
33. Lautenbacher, S., Niewelt, B.G., Kunz, M.: Decoding pain from the facial display of patients with dementia: A comparison of professional and nonprofessional observers. Pain Medicine 14(4), 469–477 (2013), http://painmedicine.oxfordjournals.org/content/14/4/469
34. Léonard, N., Waghmare, S., Wang, Y.: Rnn: Recurrent library for torch. arXiv preprint arXiv:1511.07889 (2015)
35. Li, F., Jia, X., Fraser, D.: Universal hmt based super resolution for remote sensing images. In: Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on. pp. 333–336. IEEE (2008)
36. Li, H., Hua, G.: Hierarchical-pep model for real-world face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4055–4064. CVPR '15 (2015)
37. Lin, F.C., Fookes, C.B., Chandran, V., Sridharan, S.: Investigation into optical flow super-resolution for surveillance applications (2005)
38. Lucey, P., Cohn, J.F., Matthews, I., Lucey, S., Sridharan, S., Howlett, J., Prkachin, K.M.: Automatically detecting pain in video through facial action units. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 41(3), 664–674 (June 2011)
39. Lucey, P., Cohn, J.F., Prkachin, K.M., Solomon, P.E., Matthews, I.: Painful data: The UNBC-McMaster shoulder pain expression archive database. In: 2011 IEEE International Conference on Automatic Face Gesture Recognition and Workshops (FG 2011). pp. 57–64 (Mar 2011)
40. Maintz, J.A., Viergever, M.A.: A survey of medical image registration. Medical image analysis 2(1), 1–36 (1998)
41. Malczewski, K., Stasinski, R.: Toeplitz-based iterative image fusion scheme for mri. In: Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on. pp. 341–344. IEEE (2008)
42. Milanfar, P.: Super-resolution imaging. CRC Press (2010)

43. Nasrollahi, K., Moeslund, T.B.: Super-resolution: a comprehensive survey. Machine Vision and Applications 25(6), 1423–1468 (2014)
44. Parkhi, O.M., Vedaldi, A., Zisserman, A.: Deep face recognition. In: British Machine Vision Conference. vol. 1, p. 6 (2015)
45. Parkhi, O.M., Vedaldi, A., Zisserman, A.: Deep face recognition. In: British Machine Vision Conference. vol. 1, p. 6 (2015)
46. Peled, S., Yeshurun, Y.: Superresolution in mri: application to human white matter fiber tract visualization by diffusion tensor imaging. Magnetic resonance in medicine 45(1), 29–35 (2001)
47. Prkachin: The consistency of facial expressions of pain: a comparison across modalities. 51, 297–306 (1992)
48. Prkachin, K.M., Berzins, S., Mercer, S.R.: Encoding and decoding of pain expressions: a judgement study. Pain 58(2), 253 – 259 (1994)
49. Prkachin, K.M., Solomon, P.E.: The structure, reliability and validity of pain expression: Evidence from patients with shoulder pain. 139, 267–274 (2008)
50. Prkachin, K., Schultz, I., Berkowitz, J., Hughes, E., Hunt, D.: Assessing pain behaviour of low-back pain patients in real time: concurrent validity and examiner sensitivity. Behaviour Research and Therapy 40(5), 595 – 607 (2002)
51. Ranganathan, H., Chakraborty, S., Panchanathan, S.: Multimodal emotion recognition using deep learning architectures. Institute of Electrical and Electronics Engineers Inc., United States (5 2016)
52. Sezer, O.G., Altunbasak, Y., Ercil, A.: Face recognition with independent component-based super-resolution. In: Electronic Imaging 2006. pp. 607705–607705. International Society for Optics and Photonics (2006)
53. Sikdar, A., Behera, S.K., Dogra, D.P.: Computer vision guided human pulse rate estimation: A review. IEEE Reviews in Biomedical Engineering PP(99), 1–1 (2016)
54. Sikka, K., Ahmed, A.A., Diaz, D., Goodwin, M.S., Craig, K.D., Bartlett, M.S., Huang, J.S.: Automated assessment of children's postoperative pain using computer vision. Pediatrics 136(1), 124–131 (2015)
55. Vallerand, A.H., Polomano, R.C.: The relationship of gender to pain. Pain Management Nursing 1(3, Supplement 1), 8–15 (Sep 2000)
56. Yang, J., Huang, T.: Image super-resolution: Historical overview and future challenges. Super-resolution imaging pp. 20–34 (2010)
57. Yang, J., Ren, P., Chen, D., Wen, F., Li, H., Hua, G.: Neural aggregation network for video face recognition. arXiv preprint arXiv:1603.05474 (2016)
58. Yu, Z., Zhang, C.: Image based static facial expression recognition with multiple deep network learning. In: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. pp. 435–442. ICMI '15, ACM, New York, NY, USA (2015)
59. Zhou, J., Hong, X., Su, F., Zhao, G.: Recurrent convolutional neural network regression for continuous pain intensity estimation in video. arXiv preprint arXiv:1605.00894 (2016)