U.P.B. Sci. Bull., Series C, Vol. 86, Iss. 1, 2024 ISSN 2286-3540

AUGMENTED RANDOMIZATION INJECTION TRANSFER FRAMEWORK FOR FACE EXPRESSION RECOGNITION

Andrei Racoviteanu[1], Corneliu Florea[1], Constantin Vertan[2], Laura Florea[3]

In this paper we approach the theme of face recognition. Here difficulties arise due to the perceived subjectiveness of human observers, making the annotation process hard and costly. We propose a transfer based solution, in which the key element is the injection of a randomized perturbation within controlled amplitude for efficient regularization of the flow between two different domains, one with supervised data and one with unsupervised data. On the technical side, our method uses the self labeling paradigm and, as the images from the two cases, annotated and not annotated, may be drawn from biased distributions. To cope with the bias, a random perturbation is injected in the loss function while training. On the application side, to assess the efficiency of the proposed method we experiment two scenarios that have been rarely investigated before; these refer to the separability of anxiety-originated expressions in the wild and, respectively, to the study of face expression recognition in children.

Keywords: deep learning,transfer learning, random injection, facial expression recognition

1. Introduction

Face expression recognition is unique among computer vision and pattern recognition tasks, being also informative but difficult.. We refer the reader to the recent analyses by Sariyanidi et al. and, respectively, Corneanu et al. [1, 2] for a comprehensive summary of the numerous solution proposals and challenges in the respective field. in the respective field. In the aforementioned evaluations, the major applications, trends, and solutions for classification are organized and elaborated. The expressions ”neutral,” ”anger,” ”fear,” ”disgust,” ”happy,” ”sad,” and ”surprise,” and sometimes ”contempt,” are categorized into one of Ekman’s [3] six fundamental sets. This work also addresses the issue of recognizing ”anxiety” as a distinct category.

Particularly, human face expression labeling is difficult and expensive. On CIFAR 10 (general image classification), the average untrained user scored 94% accuracy for image classes [4]. Susskind et al. [5] found that psychology students, experienced observers, had 89.2% accuracy in a 6-expression face expression experiment.

[1]PhD Student, University ”Politehnica” of Bucharest, Romania, e-mail: andrei.racoviteanu@upb.ro

[2]Professor, University ”Politehnica” of Bucharest, Romania

[3]Professor, University ”Politehnica” of Bucharest, Romania

[4] Assoc. Professor, University ”Politehnica” of Bucharest, Romania

In addition, Bartlett et al. [6] and Ekman et al. [3] found that recognizing face movements with 70% accuracy (the minimum requirement for FACS certifica- tion) requires 100 hours of training. Domingos [7] argued that more data is better than complex algorithms in a learning article. Facial expression recognition is diffi- cult to meet this requirement.Thus, we argue that face expression analysis problems benefit from adding unlabeled data instead of annotations to improve performance.

Contribution. In this paper, we examine strategies that rely on additional un- labeled data to enhance deep learning baselines in several face expression recogni- tion problems. This paper is a continuation of our previous work [8], which defined the Annealed Label Transfer algorithm. In addition to providing more information and experimental results, we have made one significant improvement. Opposite to the Annealed Label Transfer, we discover that increasing the maximal amplitude as we iterate in training is a more effective strategy for randomization injection.

Overall, we contribute by: (1) a new domain transfer method based on aug- mented injection of randomized (AIR) perturbation. (2) We report a systematic analysis of the recognition of facial expressions in children and demonstrate the efficacy of transfer learning in boosting the performance; this application is partic- ularly essential given the notorious lack of data on this task. (3) We examine the performance of detecting the expression of anxiety in images in the wild, either in a purely supervised manner or in a transfer scenario; the latter aspect has not been extensively studied to our knowledge.

Paper structure. The remainder of the paper is organized as follows. In the Section 2, we present a summary of prior related work. The method is developed into Section 3. The implementation and the tests' results on the four mentioned scenarios are detailed in Section 4. The paper ends with a discussion on the achieved results and conclusions.

2. Related Work

Random strategies for backward update in deep learning. In the last pe- riod, a number of works utilizing a deep learning framework have explored various strategies involving the injection of randomization as a means of regularizing the flow of information in either forward propagation or backward weight adjustment.

The conventional strategies are dropout and shake–shake. Dropout was in- troduced by Srivastava et al. [9] and it presume to suspend a set of randomly se- lected weights after every iteration. In a multi-branch network (such as residual networks), shake-shake regularization [10] replaced the conventional aggregation of parallel branches with a stochastic (randomly chosen) affine combination. The method combines the outputs of numerous branches of a deep neural network dur- ing training in a random manner, as opposed to simply adding them together. Our solution can be used in conjunction with dropout layers, whereas the other two vari- ants, despite being designed for fully supervised learning, can be viewed as a more detailed and computationally intensive version of our regularization. Obviously, all the earlier mentioned techniques are useful to combat overfitting.

Face expression in children databases. The problem of recognizing facial expressions in adults dates back at least 20 years, whereas the recognition of facial expressions in children is relatively new. Several proposals have only been pub- lished in the recent past.

.

The biggest collection of data that is accessible is The Child Affective Facial Expression (CAFE) [11]; however, the data set was introduced in the psychology domain, and the computer vision community has conducted little research on it. In the sense of ethics, while images from the database are exclusively for research, they cannot be utilized in publications. Baker et. al. [12] used a combination of SVM and features to identify the CAFE child emotion on this database. Nojavansghari et. al. [13] presented a new multi-modal database and experimented with several feature+classifier variants.

As the problem of recognizing facial expressions in children attracts the atten- tion of multiple organizations, it is worth noting that Khan et.al. [14] recently intro- duced a database containing images of children and reported automatically obtained results. However, results on the largest and most widely used database, CAFE, have not yet been published using the deep learning technique. Taking into consideration that we achieve 100 % accuracy with our proposed method, we will also report it on the LIRIS database, which allows for comparison with prior deep learning methods.

Recognizing Anxiety and Worry.

Anxiety is sometimes considered a subcategory of fear in emotion analy- sis, but there are reasons to classify it separately. A psychological experiment by Perkins et al. [25] found that many observers can distinguish ”anxiety” and ”fear” expressions. The difference is that stress is a short-term experience, while anxiety is a chronic condition. Another term that should be explained is that of ”worry”. Compared to worry anxiety is more intense, includes mental and verbal imagery, and lasts longer.

Carneiro et.al. [15] explored multiple automated techniques for detecting anx- ious individuals in terms of automatic recognition, but the multi-modal data neces- sitate the temporal dimensions of videos. Giannakakis et al. [16] constructed a laboratory-induced set of images with the faces of stressed people and reported approx 88% accuracy. They used a processing chain that included face region de- scription with the location of keypoints given by Active Appearance Models, optical flow, and K-Nearest Neighbor. The latter is, to the best of our knowledge, the only study reporting automated recognition of stress expressions in images. However, the cited works did not include images captured in the wild, but only images acquired in a laboratory setting.

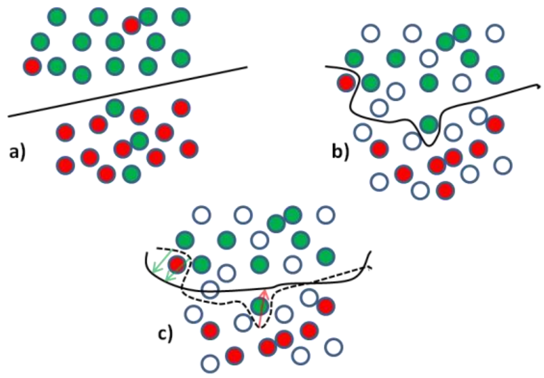

FIGURE 1. Separation boundary for a) Ideal separation -supervised b) Pseudolabel- semi-supervised c) AIR - Transfer learning

3. **Method**

The idea is based on 2 different data sets, supervised and unsupervised. After training a network on labeled data, the pseudolabel [17] algorithm transfers the information to the unsupervised domain. Next, the learner organizes the unsupervised domain to boost prediction confidence.The transfer is not straight forwards because the domains are not ideal and a randomization process i used to smooth the process. A regularization method allows unsupervised samples to change network weights if performance improvement is sufficient. Thus, distribution disparities mitigated negative effects.

The transfer smoothing process starts with the injection of an arbitrary quan- tity into the gradient with the help of a random variable function g, thereby regular- izing the flow between databases. Figure 1 presents a graphical illustration of the dual domain strategy. The line is the decision boundary for tagged (red and green dots) and untagged data (blue dots). Because of the random input, ALT attempts to change certain weaker boundaries in a random manner, while normal pseudo–labels preserve the current ones.

For the supervised batch, the net update can be defined as:

$$L_{sup} = L(y^v; \mathbf{d}^v, \Theta_n) \tag{1}$$

where y^v and d^v represent the data and the labels for a validation set, while Θ_n are the nework updated parameters for the current batch. The weights update Θ_n is done with SGD. Similarly it can be defined the unsupervised net adjustment:

$$L_{unsup} = L(y^u; \mathbf{d}^u, \Theta_{n2}) \tag{2}$$

where y^u and d^u represent the unlabeled data and the self-predicted proba- bility distributions of the netwotk. Providing more details, the total loss for the unsupervised data is:

$$L_{unsup} = L(y^u; \mathbf{d}^u, \Theta_{n2}) \quad n$$

where y_{pred} is the class probability distribution for the unlabeled data d^u predicted with the self-labeling process (pseudo-label) and y^u is an ideal probability distribution with value of ”1” on the $argmax(y_{pred})$ position and ”0” in rest. If the update is potentially positive (the loss decreases) $(L_{sup} - L_{unsup}) > \beta(n) > 0$ the updated parameters on the unsupervised batch Θ_{n2} are kept. $\beta(n)$ is chosen to be 0.2 of the main loss function.

Randomization is injected at a certain step by the random variable function g : $\{1, N_{epochs}\} \times [-1, 1] \rightarrow [-1, 1]$, where λ is a uniformly distributed radom variable in [-1,1]. Despite the fact that there are other options for the function, the favored variant in this study is increasing followed by cutting:

$$g(n, \lambda) = \begin{cases} \frac{\lambda n}{50}, & n < 50 \\ 0, & n \geq 50 \end{cases} \tag{4}$$

where n, here, indexed the training epoch and went up to $N_{epochs} = 150$.

4. Implementation and results

For the implementation, we relied on the standard architecture of the ResNet- 50 [18], but occasionally we also tested with AlexNet [19] and VGG-16 [20]; all included batch normalization. The training was done with SGD with learning rates preset to $10^{\{-3,-4-5\}}$ for 50 epochs to a total of 150. The implementation was done in Pytorch. The faces were pre–processed by cropping based on MTCNN [21] facial detector. During training, the databases were augmented using horizontal flip, random small angle rotation, and random slight contrast change.

4.1. Databases

RAF-DB [22] contains facial color images in the wild, which are, often, large enough such that cropped faces require downsizing to 224×224. The database is annotated by at least 40 trained annotators per image and divided into 12271 training images and 3078 testing images. It is labeled for seven basic emotions. On the RAF-DB database, the prior works reported standard accuracy (defined the number of correctly recognized cases normalized by the total number of cases) and the average of the main diagonal of confusion matrix, denoted as average accuracy. Because the set of worry/anxiety images did not contain enough samples, it were added to RAF-DB for the classification task.

The unlabeled data for recognizing anxiety is a subset of the **MegaFace** data- base [23], containing ≈ 311.000 images with faces randomly selected from the In- ternet. The MegaFace images contain faces in the wild that have an expression, but there is no information about it. Both datasets are public and images are exclusively for research. Examples from the two datasets are shown in Figure 2.

Children Expressions Databases. Khan et al. noted that CAFE is still the largest database of children's expressions when introducing LIRIS[1]. This database comprises pictures of infants aged two to eight years. he collection includes 90 fe- male children and 64 male children who posed for each of the seven standard facial expressions. The publicly accessible database has 1192 expert-annotated images be- cause not all children could pose for all facial expressions. The psychology-focused database has only recently been used by computer scientists.

For our experiments, the database was arbitrarily divided into 45% for train- ing, 10% for validation, and 45% for testing. This division is person wise, as each individual exists in only one of the three subsets. It should be noted that results remain consistent regardless of the division responsible for ensuring 30% of the database is included in the train set.

As unlabeled data, we selected 1389 images containing the faces of 12 chil- dren from the LIRIS database [14] and all images of children from the IMDB-WIKI database [24]. The IMDB-WIKI database was created to estimate ages. On the ba- sis of IMDB-WIKI annotations, there should have been approximately 3000 images of children aged 1 to 10 years; however, after manual validation, only 1154 were retained, with the remaining images having faulty annotations.

FIGURE 2. Left (a-c) - Face crop images from MegaFace, used as unlabeled source of information. Right (d-f) - Images form RAF-DB dataset.

4.2. **Recognition of Anxiety/Worry**

As it was mentioned in a previous section, anxiety is more intense and per- sists for a longer duration than worry Thus, we will not distinguish them only by appearance. Figure 3 shows famous actors' worried expressions.

Several works address depression recognition simultaneously [26]. Anxiety and depression are psychological disorders, but they have different expressions.

Anxiety causes fear, while depression causes anger. Anxiety, unlike depression which is associated with decreased head movement, was linked to eye darting and head rotation. According to Russel's circumflex model of emotion, anxiety is in the high-arousal, negative-valence quadrant, while depression is in the low-arousal quadrant. The two manifestations are distinct, conclusively. We will only concen- trate on ”anxiety” in our work.

FIGURE 3. Examples of expressions (genuine - surprised by pa- parazzi or posed - acting in films) from famous actors (Halle Berry, Alain Delon, Lucy Liu). One might note that the expression of “worry/anxiety” is distinct from other fundamental ones.

[1]LIRIS database can be obtained from `https://childrenfacialexpression.projet.liris.cnrs.fr/site/requestnew` and CAFE dataset from `https://www.childstudycenterrutgers.com/the-child-affective-facial-expression-se`. Images can not be published and can be used for research purposes only

In order to investigate the facial expression of anxiety, we created a collection of Internet-sourced images containing such expressions. The images have been validated as conveying ”worry/anxiety” by all of observers, which are experienced in face analysis. Perkins et al. [25] used a similar validation procedure to identify the expression of anxiety, and a comparable framework was used to develop the RAF-DB database. In our previous work, we reported the collection of 176 Internet images depicting anxiety and 43 Internet images depicting former soldiers with Post Traumatic Stress Disorder (PTSD). For this task, we have added 306 images to the database.

For the images gathered from the Internet, the procedure involved browsing Google for images using different keywords (”stress”, ”PTSD”, ”anxiety”, ”worry”) alone, or in conjunction with industries that have a high incidence of stress, such as ”(sport) coach”, ”solicitor”, ”farmer”, etc. In a subsequent stage, we expanded the search to include the names of specific notable individuals (such as politicians with a public role) in addition to the keywords. n particular, we have searched for state officials (president, prime minister, monarchical representative) along with the word ”worry”. In this way, we aimed to balance the database with the inclusion of various ethnic and racial groups.

Frequently, the manifestation of ”worry/anxiety” does not have a long dura- tion, and public figures in positions of authority attempt to conceal it. Thus, in order to have images of ”worry,” a photographer must have captured the precise moment of concern. o improve the number of images, the search was conducted in all of the main languages Google Translate supports. It resulted a collection of approximately 1000 images. According to RAF-DB, the images were evaluated by several expert viewers (included in the cropped form) and only those that passed were kept. 180 training, 36 validation, and 90 test examples were included. We intentionally included all 43 PTSD-diagnosed soldier images in the test set. The images are available at `http://imag.pub.ro/common/staff/cflorea/ Anxiety/anxiety.zip` and can be used for research purposes only.

Pekins et al. [25] conducted a study in which human participants were able to distinguish the ”stress” expression from the other fundamental facial expressions identified by Ekman’s research. To replicate this research, we have combined im- ages containing ”worry/anxiety” with the RAF-DB database (taken with its initial separation into train and test). Overall, the experiment has eight classes.

Unlabeled additional images (reference database) have been extracted from the MegaFace database. Despite the fact that this collection is vast and may con- tain any type of expression, we have not (manually) discovered any images that express tension or anxiety. Considering that the labels have distinct distributions, the scenario involves domain transfer. The confusion matrix for the version based on ResNet-50 with AIR is in Table 2, and the recognition rates, including multi- ple baselines, are in 1. Additional data improves stress detection in all scenarios, supporting the AIR method.

The confusion matrix for”anxiety” class is particularly interesting and sev- eral observations are possible. The errors are spread across Anger, Sadness, Sur- prise, and Neutral, and anxiety images are not perceived as fear. The RAF-DB does not have Contempt, so this confusion was not studied. The low-intensity ”worry/anxiety” expression is often misinterpreted as neutral, but this confusion is often done even by human observers. Secondly, no image annotated with standard expressions is misinterpreted as anxiety, arguing for the separation of the anxiety expression from the rest. In addition, the ”sad” image is arbitrarily assigned as ”anxious” in successive trials, while the overall result remains unchanged.

In Figure 4 it is presented the loss variation during 50 epochs for RAFDB facial expression recognition. The left figure expose the unsupervised loss (blue) variation in comparison with the supervised one (orange). As expected, the su- pervised loss decrease faster at the beginning, but after epoch 12 the unsupervised component tend to be lower than supervised one. In the right figure the important lines are green (the difference between supervised and unsupervised loss) and or- ange (the threshold used for updating the weights parameters). In can be observed that for the first 20 epochs the system is actually pure supervised because the cond- tion is not fulfilled. Once the training procedure advances the threshold is passed and the system begin to learn from unsupervised data batches.

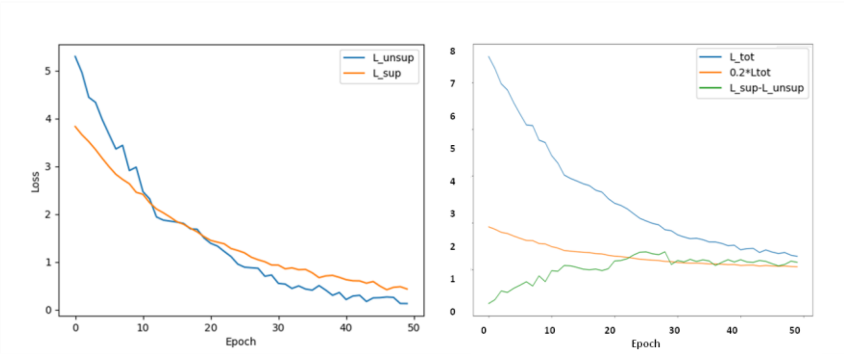

FIGURE 4. Left : Supervised and unsupervised loss variation dur- ing 50 epochs. Right : Total loss, difference between supervised and unsupervised loss and 20% of total loss which represented the β function used as threshold for parameter updating. The loss are reprted for RAFDB (supervised) dataset and Megaface (unsuper- vised)

4.3. Face Expression in children

As mentioned, adult automatic expression recognition research has many no- table proposals. Expression recognition in children has been studied less with lim- ited results. Annotated databases of children's faces are the main issue. This lack of information may be due to fragile ethics and our need to protect their image from malevolent intentions. However, after careful ethics management, several databases have been acquired, and two of them, CAFE and LIRIS, are used in our investigations.

TABLE 1. Performance (recognition rate) within 8-class problem on images of worry/anxiety/stress added to test set of RAF–DB.

Method	VGG-16 superv	VGG-16 + PL	VGG-16 + AIR	ResNet-50 superv	ResNet-50 + PL	ResNet-50 + AIR
Anxiety Recog.	60.0	52.22	58.89	58.89	63.33	67.53
Overall Recog.	77.25	83.21	85.11	75.87	83.21	**86.23**

TABLE 2. Confusion matrix obtained with the ResNet-50 + AIR for the 8 expressions problem on the (test set of RAF-DB + anxiety) database

Express.	**Sur**.	**Fear**.	**Disg**.	**Hap**.	**Sad**.	**Ang**.	**Neut**.	**Anx**
Sur.	**86.02**	1.52	1.82	2.13	1.82	2.13	4.56	0
Fear	14.86	**62.16**	1.35	9.46	5.41	2.7	4.05	0
Disg.	2.5	1.25	**56.88**	10.63	8.13	6.88	13.75	0
Hap.	0.76	0.08	0.59	**95.27**	0.51	0.17	2.62	0
Sad	0.63	0.42	3.56	5.23	**82.85**	0.84	6.28	0.21
Ang.	1.85	1.85	4.32	6.17	3.7	**77.16**	4.94	0
Neut.	2.50	0.15	1.76	3.24	5	0.59	**86.76**	0
Anx.	10.0	1.11	2.33	0	3.33	2.22	14.44	**67.78**

Experiments on the CAFE database. In this scenario, with respect to the nature of images (i.e. the probability density function describing the data), both CAFE and LIRIS databases have been acquired in laboratory with LIRIS containing slightly older children. Thus, similar probability density functions can be assumed for the data. However, the IMDB-WIKI subset contains Internet-sourced images, which are therefore in the wild. Consequently, there is a clear distinction between the domains of the labeled (CAFE) and unlabeled (LIRIS + IMDB-WIKI) data, transforming the framework to transfer learning.

One of the unique characteristics of databases with children is their restricted size. In this instance, there are too few individuals who are, however, sufficiently distinct. Direct training and testing on the database (for various train/test ratios, regularization, or data augmentation) resulted in blunt over-fitting: 100% accuracy on train and arbitrary chance on test. To enhance generalization, images from the RAF-DB database with annotations were included. Consequently, the training set will comprise CAFE and RAF-DB images.

The obtained results are presented in table 3. Before our previous work [8], only Baker et.al. [12] reported peer–reviewed, automatic results on this database using features and SVM. It trained only on CAFE images and reported results from 1000 random images without person separation. Deep learning solutions easily outperformed presented results. We trained AlexNet from scratch in our previous work [8]. We also report AlexNet's supervised training performance to establish a baseline. No matter the training strategy, ResNet-50 yielded 100% accuracy.

Overall, as the results were perfect and the strength of the residual connec- tion appeared satisfactory, we will conduct additional experiments on the LIRIS database.

Experiments on the LIRIS database In this case, the CAFE database is not utilized at all. The IMDB-WIKI database is used as a source of unlabeled data, while from LIRIS we retain the split: ”80 percent of frames for training and 10 percent for validation procedure”. In addition, we impose that a sequence be included in either the training set or the testing set. In this instance, no image from a different database is used.

In the table 4 the obtained results and comparisons to previous work can be found. Overall, our solution is 9 percent superior to the previous work. The version of transfer learning based on AIR regularization outperforms the baseline by 4%, a significant margin that demonstrates the effectiveness of the method.

TABLE 3. Performance (recognition rate) within 7-class problem on the test set of CAFE Database.

Method	Accuracy
SVM -based [12]	62.5
AlexNet - superv [8]	83.50
AlexNet + Pseudo–Labels [8]	90.29
AlexNet + ALT [8]	99.29
AlexNet + AIR	100
ResNet-50 - supervised	100
ResNet-50 + AIR	100

TABLE 4. Performance (recognition rate) within 5-class problem on the test set of LIRIS database containing expression of children.

Method	Accuracy
VGG-16 [14] - supervised	67.2
VGG-16 - +AIR	68.5
ResNet-50 - supervised	72.3
ResNet-50 + AIR	76.6

5. **Conclusions**

Automatic face expression recognition from images and videos has many practical applications, but manual annotating complexity and cost make it a prime candidate for transfer learning-based techniques like semi-supervised learning and domain adaptation. Having few annotated images and a lot of unlabeled data, methods aim to achieve the best recognition rates across realistic scenarios.

In the studied scenarios the bias must be counterbalanced by a domain adapta- tion technique that collaborates perfectly with inference over unlabeled data. In this paper, we proposed a method named Augmented Randomization Injection (**AIR**) that combines the Pseudo-Labels technique with random quantity injection into the loss function gradient. The solution performed better than the baseline in experiments.

We built on Perkins et al.'s findings that humans distinguish between faces of ”fear”/”anger” and those of ”worry/stress/anxiety.” We provided additional exam- ples, and our numerical simulation showed a good separation. The transfer strategy improved baseline and outcomes by incorporating new information through ran- domization regularization. We emphasize our community contribution by publicly compiling ”worry/anxiety” images.

In the LIRIS database, additional information improves children's expressions. Possible explanations include the lack of annotated data and the presence of expression at the apex in the child databases. Thus, children have many facial expressions, and additional information and regularization reduce bias once more. We also emphasize that recognizing children's facial expressions has received little attention and that our work establishes stronger baselines, requiring more images in the wild with annotations.

Ackknowledgement. The results presented in this article has been funded by the Ministry of Investments and European Projects through the Human Capital Sectoral Operational Program 2014-2020, Contract no. 62461/03.06.2022, SMIS code 153735.

REFERENCES

[1] *E. Sariyanidi, H. Gunes, A. Cavallaro*, Automatic analysis of facial affect: A survey of registration, representation, and recognition, IEEE Trans. on Pattern Analysis and Machine Inteligence, **37**(6) (2015) 1113 – 1133.

[2] *C. Corneanu, M. Sim on, J. Cohn, S. Escalera*, Survey on RGB, 3d, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications, IEEE Trans. on Pattern Analysis and Machine Inteligence, **38**(8) (2016) 1548 – 1568

[3] *P. Ekman, E. Rosenberg*, What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the FACS, Oxford Scholarship, 2005.

[4] *T. Ho-Phuoc*, CIFAR10 to compare visual recognition performance between deep neural networks and humans, CoRR abs/1811.07270

[5] *J. Susskind, G. Littlewort, M. Bartlett, J. Movellan, A. Anderson*, Human and computer recognition of facial expressions of emotion, Neuropsychologia, **45**(1) (2007) 152 – 162.

[6] *M. Bartlett, J. Hager, P. Ekman, T. Sejnowski*, Measuring facial expressions by computer image analysis, Psychophysiology,**36** (2) (1999) 253 – 263.

[7] *P. Domingos*, A few useful things to know about machine learning, Communications of the ACM, **55** (10) (2012) 78–87.

[8] *C. Florea, L. Florea, C. Vertan, M. Badea, A. Racoviteanu*, Annealed label transfer for face expression recognition, in: Proc. of British Machine Vision Conference, 2019, p. 12

[9] *N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov*, Dropout: a simple way to prevent neural networks from overfitting, The journal of machine learning research, **15** (1) (2014) 1929–1958.

[10] *X. Gastaldi*, Shake-shake regularization of 3-branch residual networks, in: Proc. of Int. Conf. on Learning RepresentationsWorkshops, 2017, p. 5 pg

[11] *V. LoBue, C. Thrasher*, The child affective facial expression (CAFE) set: Validity and reliability from untrained adults, Frontiers in Psychology, **5**(2015) 1532

[12] *L. Baker, V. LoBue, E. Bonawitz, P. Shafto*, Towards automated classification of emotional facial expressions., in: Proc. Annual Conf. of the Cognitive Science Society, 2017, pp. 1574 – 1579

[13] *B. Nojavanasghari, T. Baltrusaitis, C. Hughes, L. Morency*, Emoreact: a multimodal approach and dataset for recognizing emotional responses in children, in: Proc. of ACM Int. Conf on Multimodal Interaction, 2016, pp. 137 – 144

[14] *R. A. Khan, A. Crenn, A. Meyer, S. Bouakaz*, A novel database of children's spontaneous facial expressions (liris-cse), Imag. Vis. Computing **83–84** (2019) 61 – 69.

[15] *D. Carneiro, J. Carlos-Castillo, P. Novais, A. Fernandez-Caballero, J. Neves*, Multimodal behavioral analysis for non-invasive stress detection, Expert Systems with Applications, **39** (18) (2012) 13376 – 13389.

[16] *G. Giannakakis, M. Pediaditis, D. Manousos, E. Kazantzaki, F. Chiarugi, P. Simos, K. Marias, M. Tsiknakis*, Stress and anxiety detection using facial cues from videos, Biomedical Signal Processing and Control, **31** (2017) 89 – 101.

[17] *D. Lee*, Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks, in: Proc. of Int. Conf. on Machine Learning Workshops, 2013, p. 6 pg

[18] *K. He, X. Zhang, S. Ren, J. Sun*, Deep residual learning for image recognition, in: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, 2016, pp. 770–778

[19] *A. Krizhevsky*, One weird trick for parallelizing convolutional neural networks, arXiv preprint arXiv:1404.5997

[20] *K. Simonyan, A. Zisserman*, Very deep convolutional networks for large-scale image recogni- tion, in: Proc of Int. Conf. on Learning Representations, 2015, p. 14 pg

[21] *K. Zhang, Z. Zhang, Z. Li, Y. Qiao*, Joint face detection and alignment using multitask cascaded convolutional networks, IEEE Signal Proc. Letters, **23**(10) (2016) 1499 – 1503

[22] *W. Li, F. Abtahi, Z. Zhu*, Action unit detection with region adaptation, multi-labeling learn- ing and optimal temporal fusing, in: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, 2017, pp. 6766 – 6775

[23] *I. Kemelmacher-Shlizerman, S. Seitz, D. Miller, E. Brossard*, The megaface benchmark: 1 million faces for recognition at scale, in: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, 2016, pp. 4873 – 4882

[24] *R. Rothe, R. Timofte, L. V. Gool*, Deep expectation of real and apparent age from a single image without facial landmarks, Int. J. on Computer Vision (2016) 144 – 157

[25] *A. Perkins, S. Inchley-Mort, A. Pickering, P. Corr*, A. Burgess, A facial expression for anxiety., J. of Personality and Social Psychology, **102**(5) (2012) 910.

[26] *J. M. Girard, J. F. Cohn, M. H. Mahoor, S. M. Mavadati, Z. Hammal, D. P. Rosenwald*, Non-verbal social withdrawal in depression: Evidence from manual and automatic analyses, Image and vision computing, **32**(10) (2014) 641–647.