Regularization by Integrating Co-Occurrence Domain Knowledge for Affect Recognition

Ines Rieger[1 2], Jaspar Pahl[1 2], Bettina Finzel[2], Ute Schmid[2]

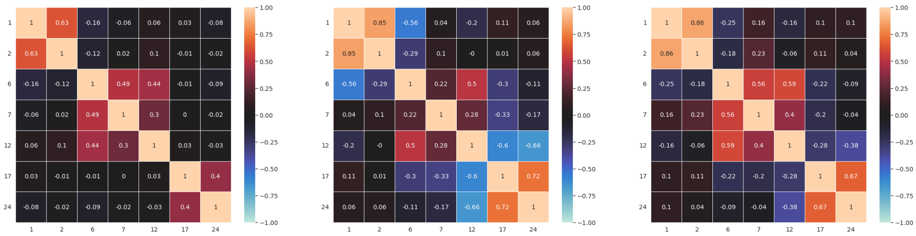
Abstract: While neural networks are widely used for affect recognition, they are often not sufficiently robust. To enhance robustness, we propose to use domain knowledge based on co-occurring facial movements. We integrate this knowledge as a regularization term into the loss function, which serves as a constraint during training. Our experiments show that our approach can enhance the generalizability of neural networks and can lower their risk of overfitting. Furthermore, our method can improve the calibration of neural networks on new facial expressions. This paper is an extended abstract of the work [Ri22], accepted at the 26th International Conference on Pattern Recognition.

Keywords: Neural Networks; Domain Knowledge; Regularization.

1 Introduction

A purely data-driven approach for training neural networks may reach its limits, for example, when there is training data of low quality or when there are constraints the model must satisfy such as natural laws or other regulations [Ru21]. Integration of domain knowledge can tackle these disadvantages by forcing the neural network to adhere to constraints.
In our approach, we propose to integrate domain knowledge on co-occuring target classes directly in the loss function to enhance affect recognition models. For our experiments, we concentrate on classifying facial movements called Action Units (AUs). AUs are a psychological framework to describe distinct, objective facial muscle movements such as lowering the brow, or raising the cheek in a modular way [EFH02].
One disadvantage of affective computing and especially AU datasets are the varying properties regarding their recording conditions, i.e. *in-the-lab* vs. *in-the-wild* or *acted* vs. *natural*. Training on datasets with very specific properties leads to models which suffer from bad generalizability and therefore do not evaluate well on datasets with different properties [Er19] in a cross-dataset setting. Domain knowledge can tackle this disadvantage, since it is to a certain degree disentangled from the dataset properties (e.g. recording setting or subject metadata) and therefore provides general information about the task.
For AUs, domain knowledge in the form of co-occurrences exist due to the fact that facial expressions such as emotions, pain or stress activate several AUs at the same time [DTM14]. Furthermore, because of the anatomically predetermined dependence of movements in the face, the contraction of muscles can also lead to the activation of several AUs at once. This

[1] Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany
[2] University of Bamberg, Cognitive Systems Group, Bamberg, Germany, firstname.lastname@uni-bamberg.de



(a) **Ground truth correlations** of facial AUs for happy faces: There is eye brow movement (AU 1-2), cheek raising and lid tightening (AU 6-7), and mouth movements that extend also to the cheek and chin (AU 12-24).

(b) Test data predictions, when trained with **binary crossentropy**. The neural network learns the correlations automatically to a certain degree.

(c) Test data predictions, when trained with **binary crossentropy (BCE) and CorrLoss**. This matrix is more similar to the ground truth correlation compared to when only trained with BCE.

Fig. 1: Intuition on how well the model learns ground truth correlations for happy faces with our CorrLoss regularization term. CorrLoss forces the NN to learn the true correlations between AUs. The example stems from experiments where we finetune models on facial expressions.

means that it is likely for many AUs to appear together in a facial expression, which can be expressed by the correlation strength between AUs. Since the patterns for the same facial expression are similar across subjects, we propose to use this co-occurrence information to enhance the model's generalization ability and to calibrate models on distinct facial expressions.

More specifically, we formulate the co-occurrence information as a weighted regularization term (**CorrLoss**) to optimize positive and negative AU correlations. During training, the CorrLoss minimizes the distance between the correlation matrix of the ground truth and the predicted labels. The total loss for each batch consists of the weighted CorrLoss and binary crossentropy loss (BCE). In contrast to other approaches that model the co-occurrence information in a hypothesis space (see [Cu20, So21]), we find that formulating this constraint as a regularization term a lightweight solution, which is furthermore flexible to steer as the domain knowledge does not need to be modeled first. However, the domain knowledge must be differentiable for training neural networks which is not always easy to realize.

For highlighting the interpretability aspect, we provide visualizations of the ground truth and learned co-occurrences that can be inspected with respect to plausibility (Fig. 1). To the best of our knowledge, we are the first to formalize a co-occurrence constraint directly in the loss function and to conduct a comprehensive cross-dataset evaluation when using co-occurrence knowledge.

Tab. 1: Cross-Dataset Evaluation. Results are in macro F1-score. Higher mean is bold and lower variance is underlined. Variable ρ weighs the CorrLoss and the BCE against each other.

		not balanced		balanced	
Train Data	Test Data	ρ=0	ρ>0	ρ=0	ρ>0
GFT (ρ=0.6)	Actor Study	0.289 ± 0.013	**0.301** ± <u>0.011</u>	0.304 ± <u>0.006</u>	0.309 ± 0.009
	AffWild2	0.150 ± <u>0.010</u>	**0.169** ± 0.014	0.199 ±0.016	**0.228** ± <u>0.011</u>
	BP4D	0.431 ± <u>0.020</u>	0.416 ± 0.026	0.450 ± 0.034	**0.481** ± <u>0.005</u>
	CK+	0.303 ± 0.026	**0.324** ± <u>0.019</u>	0.336 ± 0.023	**0.359** ± <u>0.014</u>
	EmotioNet	0.355 ± 0.017	**0.375** ± <u>0.016</u>	0.419 ± 0.019	**0.438** ± <u>0.007</u>

2 Main Results

To evaluate our approach we use several AU benchmark datasets: BP4D [Zh13], CK+ [KCT00], and GFT [Gi17], Actor Study [Se19], AffWild2 [Ko17], and EmotioNet [FBQSM16]. Our key findings are: (1) When evaluating the within dataset performance, using CorrLoss decreases the variance over different data folds, but does not significantly increase the mean results. The lower variance over several different data folds can indicate enhanced robustness. We can also observe a decreased risk of overfitting in the training. (2) When evaluating CorrLoss in a cross-dataset setting, the mean performance increases and variance decreases for most datasets compared to our baseline (see Table 1 as example for training dataset GFT). This means that CorrLoss can increase the robustness and generalizability of the model. This is also reflected in our state-of-the-art comparison, as our model outperforms state-of-the-art models that do not use co-occurrence information in a cross-dataset evaluation. (3) We can see a performance gain when we calibrate our trained models with CorrLoss on specific facial expression tasks like happiness or pain.

3 Discussion

In the following, we would like to bring up some discussion points related to our described approach.

1) Why did we chose to model our domain knowledge about co-occurring classes as correlation information instead of logical rules? We think that using correlation information in the loss function can be a lightweight and flexible solution, since finding explicit logical rules to describe this task can be difficult.

2) In what further applications could CorrLoss be used? In general, CorrLoss could be applied whenever predicted classes relate to each other–e.g., when sub-concepts (e.g., eye) relate to higher-order concepts (e.g., human and animal). In our opinion, non-causal and approximate relationships between classes, as they often occur in human behavior, could be an overall suitable application.

3) By regularizing the neural network with domain knowledge, one also introduces bias. To a certain extent, introducing bias when constraining a neural network is the goal–but the question arises, when is the regularization too much and turns into unwanted bias?

Acknowledgements This work was funded by grant 01IS18056 A/B of BMBF ML-3 (TraMeExCo), 405630557 of DFG (PainFaceReader), and 16SV7945K of BMBF (ERIK).

Bibliography

[Cu20] Cui, Zijun; Song, Tengfei; Wang, Yuru; Ji, Qiang: Knowledge Augmented Deep Neural Networks for Joint Facial Expression and Action Unit Recognition. Advances in Neural Information Processing Systems, 33, 2020.

[DTM14] Du, Shichuan; Tao, Yong; Martinez, Aleix M: Compound Facial Expressions of Emotion. Proceedings of the National Academy of Sciences, 111(15):E1454–E1462, 2014.

[EFH02] Ekman, Paul; Friesen, Wallace V; Hager, Joseph C: Facial Action Coding System: The Manual on CD ROM. A Human Face, Salt Lake City, pp. 77–254, 2002.

[Er19] Ertugrul, Itir Onal; Cohn, Jeffrey F.; Jeni, László A.; Zhang, Zheng; Yin, Lijun; Ji, Qiang: Cross-domain AU Detection: Domains, Learning Approaches, and Measures. In: 14th IEEE International Conference on Automatic Face & Gesture Recognition. pp. 1–8, 2019.

[FBQSM16] Fabian Benitez-Quiroz, C; Srinivasan, Ramprakash; Martinez, Aleix M: EmotioNet: An Accurate, Real-Time Algorithm for the Automatic Annotation of a Million Facial Expressions in the Wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5562–5570, 2016.

[Gi17] Girard, Jeffrey M; Chu, Wen-Sheng; Jeni, László A; Cohn, Jeffrey F: Sayette Group Formation Task (GFT) Spontaneous Facial Expression Database. In: 12th IEEE International Conference on Automatic Face & Gesture Recognition. IEEE, pp. 581–588, 2017.

[KCT00] Kanade, Takeo; Cohn, Jeffrey F; Tian, Yingli: Comprehensive Database for Facial Expression Analysis. In: Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580). IEEE, pp. 46–53, 2000.

[Ko17] Kollias, Dimitrios; Nicolaou, Mihalis; Kotsia, Irene; Zhao, Guoying; Zafeiriou, Stefanos: Recognition of Affect in the Wild Using Deep Neural Networks. In: Computer Vision and Pattern Recognition Workshops. pp. 1972–1979, 07 2017.

[Ri22] Rieger, Ines; Pahl, Jaspar; Finzel, Bettina; Schmid, Ute: CorrLoss: Integrating Co-Occurrence Domain Knowledge for Affect Recognition. 26th International Conference on Pattern Recognition (ICPR), 2022.

[Ru21] von Rueden, Laura; Mayer, Sebastian; Beckh, Katharina; Georgiev, Bogdan; Giesselbach, Sven; Heese, Raoul; Kirsch, Birgit; Walczak, Michal; Pfrommer, Julius; Pick, Annika; Ramamurthy, Rajkumar; Garcke, Jochen; Bauckhage, Christian; Schuecker, Jannis: Informed Machine Learning - A Taxonomy and Survey of Integrating Prior Knowledge into Learning Systems. IEEE Transactions on Knowledge and Data Engineering, 2021.

[Se19] Seuss, Dominik; Dieckmann, Anja; Hassan, Teena; Garbas, Jens-Uwe; Ellgring, Johann Heinrich; Mortillaro, Marcello; Scherer, Klaus: Emotion Expression from Different Angles: A Video Database for Facial Expressions of Actors Shot by a Camera Array. In: 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII). IEEE, pp. 35–41, 2019.

[So21] Song, Tengfei; Cui, Zijun; Zheng, Wenming; Ji, Qiang: Hybrid message passing with performance-driven structures for facial action unit detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6267–6276, 2021.

[Zh13] Zhang, Xing; Yin, Lijun; Cohn, Jeffrey F; Canavan, Shaun; Reale, Michael; Horowitz, Andy; Liu, Peng: A High-Resolution Spontaneous 3D Dynamic Facial Expression Database. In: 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). IEEE, pp. 1–6, 2013.