Abstract: FindingEmo – An Image Dataset for Emotion Recognition in the Wild

Laurent Mertens[1,2], Elahe' Yargholi[3], Jan Van den Stock[4,5], Hans Op de Beeck[3], and Joost Vennekens[6]

[1] KU Leuven, De Nayer Campus, Dept. of Computer Science J.-P. De Nayerlaan 5, 2860 Sint-Katelijne-Waver, Belgium
[2] Leuven.AI - KU Leuven Institute for AI, 3000 Leuven, Belgium
[3] Department of Brain and Cognition, Leuven Brain Institute, Faculty of Psychology & Educational Sciences, KU Leuven, 3000 Leuven, Belgium
[4] Neuropsychiatry, Leuven Brain Institute, KU Leuven, 3000 Leuven, Belgium
[5] Geriatric Psychiatry, University Psychiatric Center KU Leuven, 3000 Leuven, BE
[6] Vrije Universiteit Brussel, Brussels, Belgium
laurent.mertens@kuleuven.be

This is an extended abstract of a paper published at the 38th Annual Conference on Neural Information Processing Systems [4].

We present FindingEmo, a new image dataset containing annotations for 25,869 images, specifically tailored to Emotion Recognition. Contrary to existing datasets, it focuses on complex scenes depicting multiple people in various naturalistic, social settings, with images being annotated as a whole, thereby going beyond the traditional focus on faces or single individuals. Annotated dimensions include Russell's continuous Valence and Arousal dimensions [7] with integer scales $[-3, -2, \ldots, 3]$ for Valence and $[0, 1, \ldots, 6]$ for Arousal, and Emotion, as defined by Plutchik's discrete Wheel of Emotions (PWoE) [6]. PWoE defines 24 primary emotions, depicted as a flower organized in 8 leaves and 3 concentric rings. Each emotion leaf represents a group of 3, with opposite leaves representing opposite emotions. The rings represent the intensity levels, from most intense at the center to least intense at the outside.

Together with the annotations, we release the list of URLs pointing to the original images, as well as all associated source code on our dedicated repository [3], including the custom developed scraper and annotation interface. To mitigate the issue of broken URLs, we provide multiple URLs for a same image whenever possible. For copyright reasons, we do not share the images themselves.

The creation of the dataset was split into two phases. The first phase focused on gathering a large set of images, *prioritizing quantity over quality*, using a custom built image scraper that generates random search queries. Each query consists of three terms selected from predefined lists of, respectively, emotions, social settings/environments, and age groups of people (e.g., 'adults', 'seniors', etc.).The second phase consisted of collecting the annotations using a custom web interface. Annotators were recruited through the Prolific[7] platform.User

[7] https://www.prolific.com/

selection criteria were: fluent English speaker, (self-reported) neurotypical, and a 50/50 split male/female. In total, 655 participants contributed annotations.

Each image in the dataset has been annotated by one annotator. To assess the reliability of annotators, we used a set of 5 fixed images chosen specifically for being unambiguous. For each image, a default annotation was defined. Annotators' submissions for these images were compared to the reference, and scored. Users scoring below a fixed threshold were rejected. We also randomly presented previously annotated images to users in a controlled way, allowing to collect multiple annotations for 1,525 images (currently kept private). For 26.2% of the images, all annotators agreed on the emotion leaf, while for 46.6% of the images two labels were given. Out of these two-label annotations, 42.8% refer to adjacent emotion leafs. Annotators agree less on Arousal (average min-max difference of 2.7 ± 1.4) than on Valence (average min-max difference of 1.8 ± 1.2). Importantly, average Valence disagreement plateaus close to 2 with increasing number of annotations per image, while a linearely increasing trend is apparent for Arousal.

Keeping with the 8 leaves of PWoE, the distribution of annotations per leaf shows a clear imbalance. In particular "joy" and "anticipation" are overrepresented, and "surprise" and "disgust" heavily underrepresented, despite an added balancing mechanism to our system. A similar imbalance is found in popular facial expression datasets, such as FER2013 [1] and AffectNet [5]. Although EMOTIC [2] uses custom emotion labels, making a one-to-one comparison more difficult, it is also heavily skewed towards positive labels (top 3: "engagement", "happiness" and "anticipation"; bottom 3: "aversion", "pain" and "embarassement"). Compared to these other datasets, ours exhibits less imbalance.

Concerning the annotation values, as expected, perceived "negative" emotions ("fear", "sadness", "disgust" and "anger") have a negative average Valence, with the inverse being true for "positive" emotions ("joy", "trust"). Somewhat undecided are "surprise" and "anticipation", which can go either way. The highest Arousal values are reserved for "anger, "sadness" and "fear".Further analysis on the full emotion set verifies that also at this more fine-grained level, annotations conform to expectations, with Arousal levels increasing along with the intensity level of the PWoE ring, and Valence levels analogously increasing for "positive" and decreasing for "negative" emotions. Moreover, the association between Arousal and Valence annotations shows an expected collinearity between higher Arousal values and the extremes of the Valence range.

Baseline results for Emotion classification, and Arousal and Valence regression models obtained through transfer learning applied to multiple ImageNet-trained CNN architectures (AlexNet, VGG16, ResNet18/50/101 and DenseNet161), as well as transformer-based CLIP and DINOv2 models, show the dataset to be complex, and the tasks hard, with even modern models like CLIP and DINOv2 struggling. This suggests that in order to solve these tasks, novel Machine Learning roads might need to be explored.

References

1. Goodfellow, I.J., Erhan, D., Luc Carrier, P., Courville, A., Mirza, M., Hamner, B., Cukierski, W., Tang, Y., Thaler, D., Lee, D.H., Zhou, Y., Ramaiah, C., Feng, F., Li, R., Wang, X., Athanasakis, D., Shawe-Taylor, J., Milakov, M., Park, J., Ionescu, R., Popescu, M., Grozea, C., Bergstra, J., Xie, J., Romaszko, L., Xu, B., Chuang, Z., Bengio, Y.: Challenges in representation learning: A report on three machine learning contests. Neural networks **64**, 59–63 (2015)
2. Kosti, R., Alvarez, J.M., Recasens, A., Lapedriza, A.: Context based emotion recognition using emotic dataset. IEEE Transactions on Pattern Analysis and Machine Intelligence (2019), arXiv:2003.13401 [cs]
3. Mertens, L.: Gitlab repository containing the code and additional material for this paper. https://gitlab.com/EAVISE/lme/findingemo
4. Mertens, L., Yargholi, E., de Beeck, H.O., Van den Stock, J., Vennekens, J.: Findingemo: An image dataset for emotion recognition in the wild. In: Advances in Neural Information Processing Systems. vol. 37, pp. 4956–4996. Curran Associates, Inc. (2024)
5. Mollahosseini, A., Hasani, B., Mahoor, M.H.: Affectnet: A database for facial expression, valence, and arousal computing in the wild. IEEE transactions on affective computing **10**(1), 18–31 (2019)
6. Plutchik, R.: A general psychoevolutionary theory of emotion (1980)
7. Russell, J.A.: A circumplex model of affect. Journal of personality and social psychology **39**(6), 1161–1178 (1980)