

A technique to estimate emotional valence and arousal by analyzing images of human faces

January 26 2021, by Ingrid Fadelli



Credit: CC0 Public Domain

In recent years, countless computer scientists worldwide have been developing deep neural network-based models that can predict people's

emotions based on their facial expressions. Most of the models developed so far, however, merely detect primary emotional states such as anger, happiness and sadness, rather than more subtle aspects of human emotion.

Past psychology research, on the other hand, has delineated numerous dimensions of emotion, for instance, introducing measures such as valence (i.e., how positive an emotional display is) and arousal (i.e., how calm or excited someone is while expressing an emotion). While estimating valence and arousal simply by looking at people's faces is easy for most humans, it can be challenging for machines.

Researchers at Samsung AI and Imperial College London have recently developed a deep-neural-network-based system that can estimate emotional valence and arousal with high levels of accuracy simply by analyzing images of human faces taken in everyday settings. This model, presented in a paper published in *Nature Machine Intelligence*, can make predictions fairly quickly, which means that it could be used to detect subtle qualities of emotion in real time (e.g., from snapshots of CCTV cameras).

"Having long been working on the problem of affect estimation, it became clear to us that in general, discrete classes of emotional affect are too limited to represent the range of affect displayed by humans on a daily basis," the researchers who carried out the study told TechXplore via email. "As a result, we shifted our focus to more general dimensional measures of affect, namely valence and arousal."

Aside from highly performing hardware, building machine learning systems requires two fundamental ingredients: suitable datasets and algorithms. In their past studies, the team of researchers at Samsung AI and Imperial College thus compiled datasets that could be used to train [deep neural networks](#) for emotion recognition, including the AFEW-VA

and SEWA datasets.

"While creating the AFEW-VA [dataset](#), we showed that to obtain a method that works in naturalistic, as opposed to controlled laboratory conditions, the data on which that method is trained should also be collected in the wild," the researchers said. "Similarly, culture plays a critical role, as we showed in the SEWA project."

After they compiled datasets containing images of human faces shot in real-world settings, the researchers developed a model that merges traditional emotion recognition approaches with other emotion-related theories. The deep learning architecture they created can estimate valence and arousal with high levels of accuracy simply by processing images of human faces. Moreover, it performs well both when these images are taken in the lab and when they are taken in real-world settings.

"The main goal of our method is, given an image of a person's face, to estimate continuous valence (how positive or negative the state of mind) and arousal (how calming or exciting the experience) levels, reliably and in real-time," the researchers said.

The new system was trained on annotated images containing information about valence and arousal. In addition, it analyzed facial expressions using specific "landmarks," such as the location of a person's lips, nose and eyes, as a reference. This allows it to focus on areas of the face that are most relevant for estimating valence and arousal levels.

"We also used available labels for discrete emotion categories as an auxiliary task to provide additional supervision and obtain better performance on the main task of valence and arousal estimation," the researchers explained. "To prevent the network overfitting to any one of the tasks, we combine them using a randomized process, shake-shake

regularization."

In initial evaluations, the deep learning technique was able to estimate both valence and arousal from images of faces taken in naturalistic conditions with unprecedented levels of accuracy. Remarkably, when tested on the the AffectNet and SEWA datasets, the system performed as well as expert human annotators.

"Our network outperforms the agreement between expert human annotators on two datasets," the researchers said. "In practice, this means that if the network was considered as another annotator for these datasets, its average agreement with human annotators would be at least as good as the one between other human annotators, which is quite remarkable."

In addition to performing well, the deep learning method is non-intrusive and easy to implement, as it bases its predictions on simple images taken by regular cameras. This makes it ideal for a wide range of applications. For instance, it could be used to carry out market analyses or to create social robots that are better at understanding what humans are feeling and respond accordingly.

So far, the deep-neural-network-based system has only been trained to analyze static images. Although it could theoretically also be applied to [video footage](#), to perform equally well on videos it should also take temporal correlations into account. In their future work, the researchers thus plan to develop their system further, so that it can be used to estimate emotional valence and arousal both from static images and videos.

"The paper we presented at CVPR 2020, "Factorized Higher-Order CNNs with an Application to Spatio-Temporal Emotion Estimation," is a first step toward improving our network's performance on videos," the

researchers said. "In particular, we devised a novel method to train a neural network on static images first and then generalize to spatio-temporal data. This has the advantage of making the training of spatio-temporal networks faster while requiring less data."

More information: Estimation of continuous valence and arousal levels from faces in naturalistic conditions. *Nature Machine Intelligence*(2021). [DOI: 10.1038/s42256-020-00280-0](https://doi.org/10.1038/s42256-020-00280-0).

© 2021 Science X Network

Citation: A technique to estimate emotional valence and arousal by analyzing images of human faces (2021, January 26) retrieved 1 May 2024 from <https://techxplore.com/news/2021-01-technique-emotional-valence-arousal-images.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.