
Face2Emoji: Using Facial Emotional Expressions to Filter Emojis

Abdallah El Ali

University of Oldenburg,
Germany
abdallah.el.ali@uol.de

Wilko Heuten

OFFIS - Institute for IT
Oldenburg, Germany
wilko.heuten@offis.de

Torben Wallbaum

OFFIS - Institute for IT
Oldenburg, Germany
torben.wallbaum@offis.de

Susanne CJ Boll

University of Oldenburg,
Germany
susanne.boll@uol.de

Merlin Wasmann

OFFIS - Institute for IT
Oldenburg, Germany
merlin.wasmann@offis.de

Abstract

One way to indicate nonverbal cues is by sending emoji (e.g., 🤔), which requires users to make a selection from large lists. Given the growing number of emojis, this can incur user frustration, and instead we propose Face2Emoji, where we use a user's facial emotional expression to filter out the relevant set of emoji by emotion category. To validate our method, we crowdsourced 15,155 emoji to emotion labels across 308 website visitors, and found that our 202 tested emojis can indeed be classified into seven basic (including Neutral) emotion categories. To recognize facial emotional expressions, we use deep convolutional neural networks, where early experiments show an overall accuracy of 65% on the FER-2013 dataset. We discuss our future research on Face2Emoji, addressing how to improve our model performance, what type of usability test to run with users, and what measures best capture the usefulness and playfulness of our system.

Author Keywords

Face2Emoji; emoji; crowdsourcing; emotion recognition; facial expression; input; keyboard; text entry

ACM Classification Keywords

H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).
CHI'17 Extended Abstracts, May 06–11, 2017, Denver, CO, USA.
ACM 978-1-4503-4656-6/17/05.
<http://dx.doi.org/10.1145/3027063.3053086>

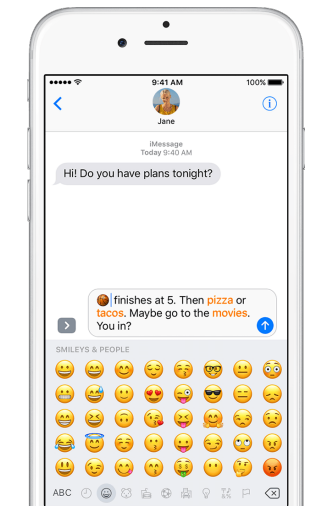


Figure 1: Apple® iOS 10 emoji keyboard within iMessage.

Introduction

Nonverbal behavior conveys affective and emotional information, to communicate ideas, manage interactions, and disambiguate meaning to improve the efficiency of conversations [14, 25]. One way to indicate nonverbal cues is by sending emoji, which are graphic icons (e.g., 🐱, 😊, 🌲) managed by the Unicode Consortium¹ that are identified by unicode characters and rendered according to a platform's font package.

Emojis enable people to express themselves richly, and while shown as screen graphics, they can be manipulated as text structures. Besides Pohl et al.'s EmojiZoom [22] who propose a zooming-based interface, entering emoji on smartphone keyboards currently requires users to make a selection from large lists (one list per category of emoji) (e.g., Apple® iOS 10 emoji keyboard² in Fig. 1). This makes emoji entry “a linear search task” [22], and given the growing number of emojis, we assume can incur user frustration. While no prior work explicitly addresses this, efforts such as Emojipedia³ highlight the need for better emoji search.

To address this, we propose Face2Emoji, a system and method to use users' facial emotional expressions as system input to filter emojis by emotional category. Despite that emojis can represent actions, objects, nature, and other symbols, the most commonly used emojis are faces which express emotion [3, 17, 24]. Moreover, previous work has shown that emojis can be ranked by sentiment (cf., Emoji Sentiment Ranking by Novak et al. [15]), textual notifications containing emojis exhibit differences in 3-valued sentiment across platforms [23], and for faces, emojis can be ranked by valence and arousal [24].

¹<http://unicode.org/emoji/> ; last retrieved: 14-02-2017

²Source: <https://support.apple.com/en-us/HT202332> ; last retrieved: 14-02-2017

³<http://emojipedia.org/> ; last retrieved: 14-02-2017

Motivation & Research Questions

Face2Emoji is motivated by two findings from the literature: that a primary function of emojis is to express emotion, and that most emojis used are face emojis. Cramer et al. [3] found that 60% (139/228) of their analyzed message by US participants were emoji used for expressing emotion. In an Instagram emoji study⁴, faces accounted for 6 of the top 10 emojis used, providing further evidence that people frequently use emoji to express emotion. Furthermore, according to a 2015 SwiftKey report⁵, faces accounted for close to 60 percent of emoji use in their analysis of billions of messages. Finally, in a qualitative study from Lee et al. [17] on emoticon sticker usage, they found that these stickers were used mainly for expressing emotions.

The study of nonverbal communication via emotions originated with Darwin's claim that emotion expressions evolved in humans from pre-human nonverbal displays [4]. Furthermore, according to Ekman [6, 7], there are six basic emotions which have acquired a special status among the scientific community: Anger, Disgust, Fear, Happiness, Sadness, and Surprise. Here, we draw on these six basic emotions, and additionally include the Neutral facial expression. By using computer vision and machine learning techniques for analyzing and recognizing emotional expressions, the user's face can be used as a natural interaction filter⁶. To test the validity of our proposed method, we used crowdsourcing to firstly identify whether a natural mapping be-

⁴<https://www.tumblr.com/dashboard/blog/instagram-engineering/117889701472> ; last retrieved: 14-02-2017

⁵<https://blog.swiftkey.com/americans-love-skulls-brazilians-love-cats-swiftkey-emoji-meanings-report/> ; last retrieved: 14-02-2017

⁶A filter according to Wikipedia ([https://en.wikipedia.org/wiki/Filter_\(higher-order_function\)](https://en.wikipedia.org/wiki/Filter_(higher-order_function))) is defined as “a higher-order function that processes a data structure (usually a list) in some order to produce a new data structure containing exactly those elements of the original data structure for which a given predicate returns the boolean value true.”

tween emojis and the seven facial expressions exists, and if so, what this mapping distribution looks like.

We address the following questions: Do the most frequently used emojis naturally map to the six basic (+ Neutral) facial emotional expressions? Can we achieve reasonable facial emotional expression recognition for these emotions using deep convolutional neural networks? The rest of the paper will address related work on natural, multimodal user interfaces and emoji communication, provide our crowdsourcing approach and results, our early emotion recognition experiments using deep convolutional neural networks, and sketch our future research steps and open questions.

Related Work

Multimodal User Interfaces and Emoji Entry

Related to our approach, Filho et al. [8] augmented text chatting in mobile phones by adding automatically detected facial expression reactions using computer vision techniques, resulting in an emotion enhanced mobile chat. For using the user's face as input, Anand et al. [1] explored a use-case of an eBook reader application wherein the user performs certain facial expressions naturally to control the device. With respect to emoji entry, Pohl et al. [22] proposed a new zooming keyboard for emoji entry, EmojiZoom, where users can see all emoji at once. Their technique, which was tested in a usability study against the Google keyboard, showed 18% faster emoji entry.

Emoji and Emotion Communication

The compactness of emojis reduces the effort of input to express not only emotions, but also serves to adjust message tone, increase message engagement, manage conversations and maintain social relationships [3]. Moreover, emojis do not have language barriers, making it possible for users across countries and cultural backgrounds to com-

municate [18]. In a study by Barbieri et al. [2], they found that the overall semantics of the subset of the emojis they studied is preserved across US English, UK English, Spanish, and Italian. As validation of the usefulness of mapping emojis to emotions, preliminary investigations reported by Jaeger et al. [13] suggest that emoji may have potential as a method for direct measurement of emotional associations to foods and beverages.

Emoji (Mis-)interpretation

Recently, Miller et al. [20] demonstrated how same emoji look differently across devices (iPhone, Android, Samsung) and is therefore differently interpreted across users. Even when participants were exposed to the same emoji rendering, they disagreed on whether the sentiment was positive, neutral, or negative around 25% of the time. In a related preliminary study, Tigwell et al. [24] found clear differences in emoji valence and arousal ratings between platform pairs due to differences in their design, as well as variations in ratings for emoji within a platform. In the context of our work, this highlights the need to account for multiple interpretations, where an emoji (as we show later) can be classified as belonging to one or more emotion categories.

Crowdsourcing Emoji to Emotion Mappings

Approach

To validate whether emojis, irrespective of function, can be categorized into one of the six basic emotional expressions (+ Neutral), and what such a mapping looks like, we adopted a crowdsourcing approach. Since currently as of Unicode 9.0, there are 1,394 emojis (not including modified emojis, or sequences)⁷, we decided to test only a subset. We selected emojis with greater than 100 occurrences from the Emoji Sentiment Ranking V1 [15] dataset, which resulted in a total of 202 emojis.

⁷<http://emojipedia.org/stats/> ; last retrieved: 14-02-2017

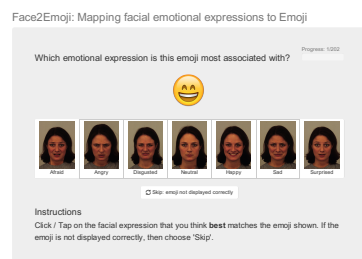


Figure 2: Snapshot of the Face2Emoji crowdsourcing website (showing female faces here).

Operating System	Labels
Win32	7113
MacIntel	3033
iPhone	2347
Android	1269
Linux	517
iPad	449
Win64	427

Table 1: OS's used to access the Face2Emoji website across all visitors (N=15,155)

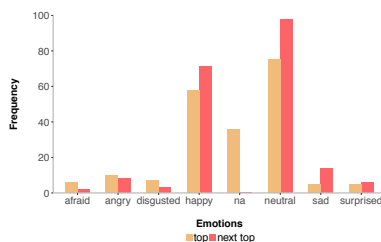


Figure 3: Distribution of top and next top crowdsource majority voting of 202 emojis across emotion categories, including NAs.

We built a website to crowdsource emoji to emotion labels (shown in Fig. 2). On the website, an emoji would be shown and a visitor has to choose one of seven emotion faces⁸: Afraid, Angry, Disgusted, Neutral, Happy, Sad, Surprised. Additionally, a 'Skip' option was provided in case the emoji was not displayed correctly. We tracked emojis and emotion labels using cookie session IDs, where the emoji index and associated unicode were used for all subsequent analysis. We additionally tracked a visitors' Operating System, however not the browser type (which can be a limitation). IP addresses were not tracked to avoid data privacy issues. Furthermore, we chose to render the unicode and not create images from them, in order to ensure users across platforms can provide emotion labels, irrespective of rendering. The website was distributed via online forums (e.g., Reddit) and the authors' social networks. Our datasets (raw and ranked) are publicly available⁹ for research purposes here: https://github.com/abdoelali/f2e_dataset

Descriptive Statistics

We collected a total of 15,155 labels, across 308 unique website visitors. Each emoji received an average of 75.0 labels ($Md=74.5$, $s=5.3$). From the total set, 1,621 (10.7 %) were 'skipped' (or labeled as NA's), where 10% of respondents who labeled NA made up 73.3% (1188/1621) of all NAs in our dataset. The distribution of operating systems used to access the website are shown in Table 1.

Annotation Quality

As a test for annotation quality, we independently (N=2) rated each emoji by classifying into one of the emotion categories, and computed unweighted Cohen's Kappa. Our ratings reached moderate agreement on classifying emojis

into emotions ($\kappa=0.55$, CI: [0.46,0.65]), where we agreed on 71.3% (144/202) of emojis. These joint labels were then compared with the top ranked (majority voted) emojis from the crowd, which gave an almost perfect agreement ($\kappa=0.85$, CI: [0.77,0.93]).

Classification Results

The distribution of the top most frequent (by majority vote) emotion labels, as well as the next top labels, across the 202 tested emojis are shown in Fig. 3. Interesting to observe here that for the majority of labels, none of the emojis tested were skipped due to unicode rendering. From our labeled data, it became clear that an emoji can be classified under two emojis (following a bimodal or at times multimodal distribution). For example, 🏆 was nearly equally labeled as Happy (N=32) (since a trophy, a sign of achievement can evoke happiness) and Neutral (N=34), since it is an object with no direct mapping to a facial expression. Therefore, to account for this variability, we classified whether an emoji belongs to an emotion label using our Emotion Class (EC) function:

$$EC = \frac{x_{ij}}{\max(x_i)} = \begin{cases} 1 & \text{if } EC > 0.5 \\ 0 & \text{if } EC \leq 0.5 \end{cases} \quad (1)$$

where: $x_i \in [1, 202]$, $x_j \in [1, 8]$. We chose a cutoff threshold of 0.5, where an emoji is classified as belonging to an emotion class if $EC > 0.5$. The result of applying our EC function to our data is shown in Table 2, where the emojis per emotion category are sorted by label count in ascending order.

⁸Female or male faces randomly chosen on page refresh.

⁹Our datasets contain no sensitive information and therefore comply with user privacy.

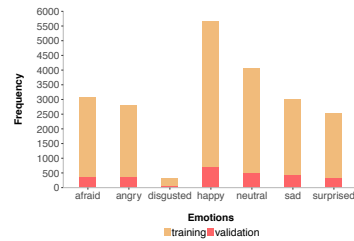


Figure 4: Training and validation data distribution across emotion categories.

Layer	Output Size
Input	48 x 48 x 1
Convolution	5 x 5 x 64 (activation = ReLU)
Max Pooling	3 x 3 (strides = 2)
Convolution	5 x 5 x 64 (activation = ReLU)
Max Pooling	3 x 3 (strides = 2)
Convolution	4 x 4 x 128 (activation = ReLU)
Dropout	value = 0.3
Fully Connected	3072
Softmax	7

Table 3: Our current CNN architecture.

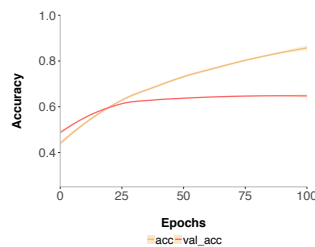


Figure 5: Accuracy and validation of our final network model across 100 epochs.

Neutral										Happy										Sad										Angry										Surprised										Disgusted										Afraid																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																															
🙄	🙄	🙄	🙄	🙄	🙄	🙄	🙄	🙄	🙄	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬	😬

Table 2: Resulting emojis per emotion class distribution after applying our EC function.

Deep CNN for Emotion Recognition

To build our emotion recognition module, we used deep Convolutional Neural Networks (CNNs). Deep Learning-based approaches, particularly those using CNNs, have been very successful at image-related tasks in recent years, due to their ability to extract good representations from data [12]. We chose to build our own recognition system instead of using available APIs (such as Microsoft's Emotion API¹⁰) because: (a) it allows us greater flexibility in inspecting the classification accuracies ourselves and determining why certain emotions are not correctly classified, (b) we can ensure user privacy by running all predictions directly on the device, and (c) it is free.

Dataset & Architecture

We used the FER-2013 facial expression dataset [9] for training and validation, which comprises 32,298 grayscale

48x48 pixel images of facial expressions, collected from the web using 184 emotion-related keywords. We implemented our network with TFLearn¹¹, a deep learning library featuring a higher-level API for TensorFlow¹². Our implementation and training procedure followed recent work by Gudi et al. [10] who used CNNs for emotion recognition. All faces were detected with OpenCV's Viola & Jones face detector (frontal) [26], resulting in a final training sample of 21,039 and validation sample of 1,546 images. The distribution across emotion labels is shown in Fig. 4, where it can be seen that most of the emotions are Happy, Neutral, Sad, and Surprised. After experimenting with different architecture and hyperparameters, our final network architecture is shown in Table 3, where training was done with a batch size of 32, using stochastic gradient descent with hyperparameters (momentum=0.9, learning rate=0.001, weight

¹⁰<https://www.microsoft.com/cognitive-services/en-us/emotion-api> ; last retrieved: 14-02-2017

¹¹<http://tflearn.org/> ; last retrieved: 14-02-2017

¹²<https://www.tensorflow.org/> ; last retrieved: 14-02-2017

neutral	0.03	0.01	0.02	0.04	0.10	0.02	0.78
surprised	0.01	0.01	0.08	0.03	0.03	0.81	0.04
sad	0.03	0.04	0.05	0.06	0.43	0.04	0.34
happy	0.01	0.01	0.00	0.86	0.03	0.03	0.07
afraid	0.11	0.04	0.43	0.01	0.12	0.12	0.17
disgusted	0.09	0.74	0.03	0.00	0.11	0.00	0.03
angry	0.46	0.05	0.08	0.03	0.16	0.03	0.20
	angry	disgusted	afraid	happy	sad	surprised	neutral

Figure 6: Performance matrix for our deep CNN model.

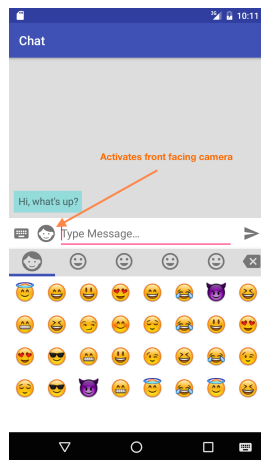


Figure 7: Early Android-based Face2Emoji prototype showing emojis (with Apple's® unicode rendering).

decay=0.0005) where loss was computed using categorical cross-entropy, and run on an NVIDIA GeForce GTX 970 GPU for 100 epochs.

Early Experiments & Results

Accuracy and validation accuracy plots across 100 epochs for our CNN model are shown in Fig. 5. Our network converged on a validation accuracy of 65%, which is comparable to human level performance on this dataset [9]. To evaluate our network performance, we tested our predictions on 1,523 FER-2013 test images. Additionally, we used the Radboud Faces Database (RaFD) [16], which consists of 8000 high resolution faces, as well as the Karolinska Directed Emotional Faces (KDEF) [19], which consists of 4900 pictures of human facial expressions of emotion.

The datasets differ on quantity, quality, and how much posing is involved. In this respect, the FER-2013 dataset shows emotions 'in the wild'. We took only the frontal images of the RaFD, and after face detection, we had a test set of 1,407 images. For the KDEF dataset, after preprocessing, we had 980 images. The performance of our network on the FER-2013 test set is 68%, 55% on the RaFD, and 46% on KDEF. We additionally experimented with a ResNet-32 [11] deep residual network (which have recently shown great promise on image recognition tasks), where we achieved up to 70% validation accuracy, however the network appeared to overfit and performed poorly on our FER-2013 test set (32.5%). For this reason, we left further experiments with this type of network for future work.

Our model prediction performance matrix is shown in Fig. 6. Best results were for Happy, Surprised, and Neutral emotions. This is in part due to the amount of training data used, but also to the difficulty in detecting certain emotions (e.g., surprise and fear are easily conflated due to similar coarse facial features). Given this, in our future work we in-

tend to test our Face2Emoji app with users for only those three emotions: Happy, Surprised, and Neutral.

Next Steps & Research Agenda

Our next steps are to experiment further with deep learning approaches for emotion recognition (e.g., using transfer-based learning to deal with our small dataset [21]). Furthermore, we are currently completing development of the Face2Emoji keyboard prototype for Android (shown in Fig. 7), and planning a usability test with users. In this usability test, we want to test Face2Emoji against a baseline custom keyboard we call EmoTabs, where the emojis are organized according to emotion labels (instead of the default categories). Our current plan is to evaluate Face2Emoji on selection time performance for individual emojis, but more importantly on whether emoji filtering using one's own emotional facial expressions is fun and natural.

For this work in progress, many open questions remain which steer our future work: how would users perceive such a system (i.e., does it raise social acceptability issues)? How to give feedback to the user on their current detected emotion? Should activating face recognition remain user-driven, or should it be system-driven (i.e., continuous recognition)? How would they react to such machine learning predictions, especially when they are incorrect or exhibit bias? How should misclassifications be explained and visualized to users? Since we have shown that emojis can be classified into emotion categories, can NLP methods be used to automate classification of new emojis? Finally, what other smartphone applications can benefit from facial emotional expression shortcuts? For distant future work, we intend on exploring a personalized form of Face2Emoji, where we would integrate contextual cues using word embedding models (including emoji2vec [5] emoji pre-trained embeddings) for personalized ranking and filtering.

References

- [1] B. Anand, B. B. Navathe, S. Velusamy, H. Kannan, A. Sharma, and V. Gopalakrishnan. 2012. Beyond touch: Natural interactions using facial expressions. In *proc. CCNC '12*. 255–259. DOI: <http://dx.doi.org/10.1109/CCNC.2012.6181097>
- [2] Francesco Barbieri, German Kruszewski, Francesco Ronzano, and Horacio Saggion. 2016. How Cosmopolitan Are Emojis?: Exploring Emojis Usage and Meaning over Different Languages with Distributional Semantics. In *Proc. MM '16*. ACM, New York, NY, USA, 531–535. DOI: <http://dx.doi.org/10.1145/2964284.2967278>
- [3] Henriette Cramer, Paloma de Juan, and Joel Tetreault. 2016. Sender-intended Functions of Emojis in US Messaging. In *Proc. MobileHCI '16*. ACM, New York, NY, USA, 504–509. DOI: <http://dx.doi.org/10.1145/2935334.2935370>
- [4] Charles Darwin. 1872/2009. *The Expression of the Emotions in Man and Animals* (anniversary ed.). Harper Perennial. <http://www.worldcat.org/isbn/0195158067>
- [5] Ben Eisner, Tim Rocktäschel, Isabelle Augenstein, Matko Bosnjak, and Sebastian Riedel. 2016. emoji2vec: Learning Emoji Representations from their Description. *CoRR* abs/1609.08359 (2016). <http://arxiv.org/abs/1609.08359>
- [6] Paul Ekman. 1992. An argument for basic emotions. *Cognition and Emotion* (1992), 169–200.
- [7] Paul Ekman and Wallace V. Friesen. 1971. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology* 17, 2 (1971), 124–129. <http://search.ebscohost.com/login.aspx?direct=true&db=pdh&AN=psp-17-2-124&site=ehost-live>
- [8] Jackson Feijó Filho, Thiago Valle, and Wilson Prata. 2014. Non-verbal Communications in Mobile Text Chat: Emotion-enhanced Mobile Chat. In *Proc. MobileHCI '14*. ACM, New York, NY, USA, 443–446. DOI: <http://dx.doi.org/10.1145/2628363.2633576>
- [9] Ian J. Goodfellow, Dumitru Erhan, Pierre Luc Carrier, Aaron C. Courville, Mehdi Mirza, Benjamin Hamner, William Cukierski, Yichuan Tang, David Thaler, Dong-Hyun Lee, Yingbo Zhou, Chetan Ramaiah, Fangxiang Feng, Ruifan Li, Xiaojie Wang, Dimitris Athanasakis, John Shawe-Taylor, Maxim Milakov, John Park, Radu-Tudor Ionescu, Marius Popescu, Cristian Grozea, James Bergstra, Jingjing Xie, Lukasz Romaszko, Bing Xu, Chuang Zhang, and Yoshua Bengio. 2013. Challenges in Representation Learning: A Report on Three Machine Learning Contests.. In *ICONIP (3)*, Minho Lee, Akira Hirose, Zeng-Guang Hou, and Rhee Man Kil (Eds.), Vol. 8228. Springer, 117–124.
- [10] Amogh Gudi. 2015. Recognizing Semantic Features in Faces using Deep Learning. *CoRR* abs/1512.00743 (2015). <http://arxiv.org/abs/1512.00743>
- [11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *Proc. CVPR '16*.
- [12] Yoshua Bengio Ian Goodfellow and Aaron Courville. 2016. Deep Learning. (2016). <http://goodfeli.github.io/dlbook/> Book in preparation for MIT Press.
- [13] Sara R. Jaeger, Leticia Vidal, Karrie Kam, and Gastón Ares. 2017. Can emoji be used as a direct method to measure emotional associations to food names? Preliminary investigations with consumers in {USA} and China. *Food Quality and Preference* 56, Part A (2017), 38 – 48. DOI: <http://dx.doi.org/10.1016/j.foodqual.2016.09.005>
- [14] Mark L. Knapp and Judith A. Hall. 2010. *Nonverbal Communication in Human Interaction* (7 ed.). Wadsworth: Cengage Learning, Bosten, USA.

- [15] Petra Kralj Novak, Jasmina Smailović, Borut Sluban, and Igor Mozetič. 2015. Sentiment of Emojis. *PLOS ONE* 10, 12 (12 2015), 1–22. DOI: <http://dx.doi.org/10.1371/journal.pone.0144296>
- [16] Oliver Langner, Ron Dotsch, Gijsbert Bijlstra, Daniel H. J. Wigboldus, Skyler T. Hawk, and Ad van Knippenberg. 2010. Presentation and validation of the Radboud Faces Database. *Cognition and Emotion* 24, 8 (2010), 1377–1388. DOI: <http://dx.doi.org/10.1080/02699930903485076>
- [17] Joon Young Lee, Nahi Hong, Soomin Kim, Jonghwan Oh, and Joonhwan Lee. 2016. Smiley Face: Why We Use Emoticon Stickers in Mobile Messaging. In *Proc. MobileHCI '16*. ACM, New York, NY, USA, 760–766. DOI: <http://dx.doi.org/10.1145/2957265.2961858>
- [18] Xuan Lu, Wei Ai, Xuanzhe Liu, Qian Li, Ning Wang, Gang Huang, and Qiaozhu Mei. 2016. Learning from the Ubiquitous Language: An Empirical Analysis of Emoji Usage of Smartphone Users. In *Proc. UbiComp '16*. ACM, New York, NY, USA, 770–780. DOI: <http://dx.doi.org/10.1145/2971648.2971724>
- [19] Daniel Lundqvist, Anders Flykt, and Arne Öhman. 1998. The Karolinska directed emotional faces (KDEF). *CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet* (1998), 91–630.
- [20] Hannah Miller, Jacob Thebault-Spieker, Shuo Chang, Isaac Johnson, Loren Terveen, and Brent Hecht. 2016. "blissfully happy" or "ready to fight": Varying interpretations of emoji. In *Proc. ICWSM 2016*. AAAI press, 259–268.
- [21] Hong-Wei Ng, Viet Dung Nguyen, Vassilios Vonikakis, and Stefan Winkler. 2015. Deep Learning for Emotion Recognition on Small Datasets Using Transfer Learning. In *Proc. ICMI '15*. ACM, New York, NY, USA, 443–449. DOI: <http://dx.doi.org/10.1145/2818346.2830593>
- [22] Henning Pohl, Dennis Stanke, and Michael Rohs. 2016. EmojiZoom: Emoji Entry via Large Overview Maps 😄🔍. In *Proc. MobileHCI '16*. ACM, New York, NY, USA, 510–517. DOI: <http://dx.doi.org/10.1145/2935334.2935382>
- [23] Channary Tauch and Eiman Kanjo. 2016. The Roles of Emojis in Mobile Phone Notifications. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct (UbiComp '16)*. ACM, New York, NY, USA, 1560–1565. DOI: <http://dx.doi.org/10.1145/2968219.2968549>
- [24] Garreth W. Tigwell and David R. Flatla. 2016. Oh That's What You Meant!: Reducing Emoji Misunderstanding. In *Proc. MobileHCI '16*. ACM, New York, NY, USA, 859–866. DOI: <http://dx.doi.org/10.1145/2957265.2961844>
- [25] Jessica L Tracy, Daniel Randles, and Conor M Steckler. 2015. The nonverbal communication of emotions. *Current Opinion in Behavioral Sciences* 3 (2015), 25 – 30. DOI: <http://dx.doi.org/10.1016/j.cobeha.2015.01.001>
- [26] Paul Viola and Michael J. Jones. 2004. Robust Real-Time Face Detection. *Int. J. Comput. Vision* 57, 2 (May 2004), 137–154. DOI: <http://dx.doi.org/10.1023/B:VISI.0000013087.49260.fb>