

# Study on Predicting Sentiment from Images using Categorical and Sentimental Keyword-based Image Retrieval

Sanghyun Seo · Dongwann Kang

Received: date / Accepted: date

**Abstract** Visual stimuli are the most sensitive stimulus to affect human sentiments. Many researches have attempted to find the relationship between visual elements in images and sentimental elements using statistical approaches. In many cases, the range of sentiment that affects humans varies with image categories, such as landscapes, portraits, sports, still-life, etc. Therefore, to enhance the performance of sentiment prediction, an individual prediction model must be established for each image category. However, collecting much ground truth sentiment data is one of the obstacles encountered by studies on this field. In this paper, we propose an approach that acquires a training data set for category classification and predicting sentiments from images. Using this approach, we collect a training data set and establish a predictor for sentiments from images. First, we estimate the image category from a given image, and then we predict the sentiment as coordinates on the Arousal-Valence (AV) space using the predictor of an estimated category. We show that the performance of our approach approximates performance using ground truth data. Based on our experiments, we argue that our approach, which utilizes big data on the web as the training set for predicting content sentiment, is useful for practical purposes.

**Keywords** Sentiment of Image · Sentiment Prediction · Image Classification · Image Retrieval

---

Sanghyun Seo

Smart Game Platform Section, Electronics and Telecommunications Research Institute (ETRI), 218 Gajeong-ro, Yuseong-gu, Daejeon, 305-700, South Korea.

E-mail: shseo@etri.re.kr

Dongwann Kang (Corresponding Author)

Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul, 156-756, South Korea.

E-mail: dongwann@cglab.cau.ac.kr

## 1 Introduction

Humans feel diverse sentiments through the effect of various stimuli, such as visual, acoustic, tactile, and olfactory [Reiman et al(1997)Reiman, Lane, Ahern, Schwartz, Davidson, Friston, Yun, and Chen]. Among these, visual stimuli are the most sensitive stimulus to affect human sentiments [Mayer et al(1990)Mayer, DiPaolo, and Salovey]. To feel sad or pleasant sentiments while watching a movie is one of the common experiences. On the one hand, these affections to sentiment are caused by context, although sometimes, they can also be caused only through visual stimulus without context. Feeling an sentiment from artwork we cannot understand is an example of this.

Several studies [Nicolaou et al(2012)Nicolaou, Gunes, and Pantic, Datta et al(2008)Datta, Li, and Wang] on finding a relationship between visual stimuli from images and human sentiments have been conducted for a long time in the affective computing field [Picard(1995),Picard(2003),Picard et al(2001)Picard, Vyzas, and Healey]. Many studies attempted to find the relationship between visual elements in images and sentimental elements using statistical approaches. For this, such studies extracted various visual elements from images by employing image processing and computer graphics techniques, quantified sentiments from images on a normalized sentimental space, and established a model for predicting sentiment from images by finding the mathematical relationship between them.

However, in many cases, the range of sentiments that affect humans varies with image categories, such as landscapes, portraits, sports, still-life, etc. Therefore, to enhance the performance of sentiment prediction, an individual prediction model must be established for each image category. At this time, many data that consist of image-sentiment pairs are required to establish a reliable prediction model. However, to establish an individual prediction model per image category, we need bigger data than the common model for all images. To collect much ground truth sentiment data is one of the obstacles in the studies on this field, and thus it is not easy to predict sentiments from images according to their category.

Human sentiment is subjective, and thus surveys on sentiments from images must be conducted with several people in order to acquire reliable ground truth sentiment data. To do this effectively, crowdsourcing has been employed for many studies. However, asking people to quantize sentiment for each image is tedious work for the survey respondents. Moreover, the quantized sentiment is not normalized generally. Therefore, many studies that employ crowdsourcing attempt to eliminate noise from survey responses and normalize such responses.

In this study, in order to overcome these limitations, we take an approach that acquires sentiment data easily by employing keyword-based image retrieval, and predicts sentiments from images according to category. On the Internet, there are numerous images combined with context. For example, some images are tagged with metadata. For other images, additional informa-

tion is included in the posts. Such context can be utilized for retrieving images for certain category or sentiment.

We assume that sentimental keywords in the context of images are highly related to sentiments from images. From this assumption, we use sentimental keywords as the ground truth sentiment from an sentiment by retrieving images that contain the sentimental keywords. In our sentimental keyword dictionary, each keyword has quantized coordinates on a normalized sentimental space. Therefore, we obtain the sentimental data of images by matching keyword coordinates in the dictionary and keywords in the context of an image. This method can be automated easily, and thus this method is more effective than user surveys that require much time and a high budget.

We collect images of a certain category and sentiment using categorical and sentimental keywords in image retrieval. We utilize images collected according to the category as the training set of machine learning for classifying image category. Then we extract several features from images, and establish an sentiment prediction model by finding the relationship between the features and sentimental coordinates of the images through machine learning. At this time, we conduct learning per category, so that we obtain the models for each category.

The contribution of our work is as follows. First, we propose a method that utilizes a number of unstructured data widely spread in the web as the machine learning training data for classifying images and predicting sentiments from images. This approach has the advantage of automatically acquiring and updating data. Moreover, this is a practical example where big data can be employed for sentimental research. Second, our method for predicting sentiments from images employs image categories so that our predictor has better prediction performance.

The remainder of this paper is organized as follows. In Ssection 2, we provide an overview of related works on sentiments from the imagesentiment of image. Next, we describe our method for collecting training data sets for image classification and sentiment prediction in Section 3, and define the features for being used fro for machine learning in Section 4. In Section 5, we present the results of our method. Finally, in Section 6, we conclude with a summary of our method, a discussion of its limitations, and the scope for future development.

## 2 Related work

One of the main methods for acquiring sentimental data for images is label analysis. In [Schmidt and Stock(2009)], users were asked to tag basic sentiments by applying scroll bars. By taking the mean value, they obtained the sentimental data of images. This is a simple and direct method. However, to ask persons about the sentiments of many images requires much time and human resources.

Another method is facial expression analysis. Human facial expressions reflect not only sentiments, but also mental activities, social interaction, and

physiological signals. Therefore, many studies that find sentiments from images utilize facial expressions [Fasel and Luetttin(2003), Yun et al(2013)Yun, Kim, Park, and Kim]. However, the main limitation of this approach is that humans and their faces must be present in the images. In our study, many image categories do not contain humans. Therefore, this method cannot be employed for general images.

The other method for obtaining sentiments from images is visual content analysis [Colombo et al(1999)Colombo, Del Bimbo, and Pala, Machajdik and Hanbury(2010)]. This method assumes that there is no explicit information for images. The visual image features are extracted based on low-level image properties, such as color and texture. Extracting features from images in our approach is essentially based on this method.

One of the studies on finding the relationship between the sentimental space and images is the Color Image Scale [Kobayashi(1981)], which is literature studied by Shigenobu Kobayashi and the Nippon Color & Design Research Institute. In their psychophysical research, they presented over 1,000 color combinations to express any mood, taste, or lifestyle that belongs to 174 semantic keywords on the mood perceived from colors. They labeled each color combination with one of 174 keywords. In addition, they devised a two-dimensional space for representing human sentiment, mood and adjective words, and the color image scale that consists of two axes that correspond to the scales hard-soft and cool-warm. On the color image scale, the researchers located every keyword according to its two scales measured by their numerous studies. In the color image scale, they also defined 15 categories, so that each keyword belongs to one of the categories.

A general method for measuring sentiment continuously is known as the Arousal-Valence (AV) space [Russell(1989), Schubert(1999)]. In this model, each sentiment is represented using two factors, such as arousal and valence. This approach has the advantage where all sentiments can be represented not discretely, but continuously, and thus many studies that measure sentiment employ this approach [Yang et al(2008)Yang, Lin, Su, and Chen, Lang et al(1999)Lang, Bradley, and Cuthbert, Dan-Glauser and Scherer(2011)]. In this study, we also use this approach to measure sentiment.

### 3 Image Acquisition with Ground Truth sentiment Data

#### 3.1 Keywords collection

Our approach for acquiring images is based on the keywords with ground truth sentiment coordinates on the AV space. In [Warriner et al(2013)Warriner, Kuperman, and Brysbaert], arousal and valence values for 13,915 English lemmas are surveyed. The lemmas include nouns, verbs, and adjectives. From these, we select adjectives only. The total number of adjective keywords is over 3,000. However, some keywords are for words not commonly used, and thus influ-

ential words must be selected. We select only 300 keywords ordered by the number of search results in Google search.

Although these keywords are the most influential, they are not guaranteed to cover the most range of sentiments. To select the keywords distributed on the AV space, we use the Voronoi diagram [Aurenhammer(1991)]. First, we generate the Voronoi diagram using the AV coordinates of 300 keywords as the Voronoi sites. Next, we find Voronoi cells whose area is bigger than a given threshold value. For each of those cells, we find an unselected keyword that divides the cell evenly into two cells by inserting it as a new site. We select a new keyword that minimizes the following equation:

$$size(a)^2 + size(b)^2 \quad (1)$$

Here,  $a$  is the keyword cell to be divided, and  $b$  is the new keyword cell. Through this method, we finally select 331 adjective keywords.

### 3.2 Image collection

For selected keywords, we collect images by retrieving keywords using Google image search. At this time, we use categorical keywords together (Figure 1). Portrait, landscape, sports, flower, animal, and architecture are employed for the categorical keywords in this study. For each adjective keyword, we search images using six categorical keywords. We collect the top 100 images from the search results. At the same time, we record the rank of each image search.

In many cases, duplicate images might be retrieved, and thus they must be removed. However, finding the exact same images from those collected requires too heavy computation. Instead, we prevent a heavy load in the duplicated reduction process by finding duplicate images in low resolution. For this, we employ the Antipole tree [Cantone et al(2005)Cantone, Ferro, Pulvirenti, Recupero, and Shasha], and use a thumbnail image resized to 16 16 as an index for the image. By inserting and searching thumbnail images, we find the same images rapidly. If a duplicate image is found for the same keyword, we remove the duplicate image. Otherwise, if a duplicate is found for other keywords, we simply add the keyword to the image.



**Fig. 1** Image Acquisition Process with Ground Truth Sentimental Data.

### 3.3 sentimental data collection

As indicated in Section 3.1, each keyword has its own AV coordinates surveyed through a user study. We assign the keyword coordinates to the images collected by retrieving the keyword. At this time, because of the duplication in image retrieving, many images have several keywords. For these images, we must assign single coordinates as the sentimental data. We calculate single AV coordinates by interpolating the keyword coordinates of the image. However, the influence of the keywords differs from each other for an image. In order to reflect the keyword influence, we utilize the image rank recorded during retrieval using equation 2.

$$c_I = \sum_{k \in k_I} w_{k_I} \times c_k \quad (2)$$

$$w_{k_I} = 1 - \frac{r_I}{N_k} \quad (3)$$

Here,  $c_I$  is the interpolated sentiment coordinates for image  $I$ ,  $k_I$  is the keyword for image  $I$ ,  $c_k$  is the coordinates of a keyword  $k$ ,  $r_I$  is the rank of image  $I$  from the retrieved results, and  $N_k$  is the number of images retrieved by keyword  $k$ .

## 4 Feature Extraction

For the collected images, we extract several visual features to conduct machine learning. For the aspects of color, texture, and composition, we define the features as explained in the following subsections.

### 4.1 Color

In general, the most significant visual features that relate to color are global hue and tone, contrast hue and tone, and color relationship [Kobayashi(1981)]. To calculate hue-related features, we convert all pixel colors into the *Lab* color space, and obtain  $a$  and  $b$  values. For the global hue, we calculate the average value of  $a$  and  $b$ . For hue contrast, we calculate the standard deviation of  $a$  and  $b$ . Contrary to hue, tone consists of two elements, brightness and saturation. For the tone features, we calculate the  $L$  value of all pixels in the *Lab* color space as brightness, and the  $s$  value in the *HSV* color space as the saturation. Similarly to hue, we compute the average and standard deviation of tone values. Although these features relate to global color and color contrast, they cannot reflect the global color relationship. Therefore, we classify the color relationship of images using a color harmony template [Moon and Spencer(1944), Tokumaru et al(2002) Tokumaru, Muranaka, and Imanishi]. We compute an images hue histogram, and find the best matching color harmony template by calculating the distance between the hue histogram and the templates [Cohen-Or et al(2006)Cohen-Or, Sorkine, Gal, Levvand, and Xu].

## 4.2 Texture

Texture is one of the important visual features for images. In many studies that consider texture features, Haralick features [Haralick et al(1973)Haralick, Shanmugam, and Dinstein,Haralick(1979)] are widely used. In this study, we also employ Haralick features as one of the texture features. In addition, in order to consider granularity and its direction, we define several features using the gradient value. For the degree of granularity, we compute the average and standard deviation of the gradient magnitudes. For the granularity direction, we calculate the average and standard deviation of the gradient vectors.

## 4.3 Composition

Image composition is one of the high-level visual features that strongly relate to image classification and sentiments from images. However, image composition is complex, and it is not easy to extract quantized features from image compositions. In this study, we consider only the position of important subjects in an image. To determine which objects are important, we employ visual saliency [Itti et al(1998)Itti, Koch, and Niebur]. First, we calculate the saliency of all pixels in an image. Next, we divide the image into several grids, e.g.,  $7 \times 7$  grids. For each grid, we add the saliency values of all pixels in the grid. Finally, we obtain a vector that consists of the added saliency values of the grids, e.g., 49 added saliency values. In this paper, we employ this vector as the feature for image composition.

## 5 Experimental Results and Discussion

To generate the category classifier and sentiment predictor, we employ machine learning techniques. First, for the images that we collected, we establish a category classifier using the features defined in Section 4 and a support vector machine (SVM) [Suykens and Vandewalle(1999)]. Each category has images collected by adjective keyword retrieval, so that there are over 30,000 images for each category. This is a statistically sufficient number of data for learning. Next, for each category, we conduct linear regression using the features we defined and the sentimental coordinates of the images. Consequently, we obtain an individual sentiment predictor for each category. Therefore, before predicting sentiments from images, classifying the category is required. Then, we perform our prediction using the model for classified category.

In our experiments, we used Google image search as the search engine. However, in Google's search application program interface (API), there is a limit to the number of search queries and retrieval results. Therefore, we developed a crawling tool using Google image search, and utilized it for the purpose of training only. To extract the features we defined, we used MATLAB. For machine learning, Weka [Hall et al(2009)Hall, Frank, Holmes, Pfahringer,

Reutemann, and Witten] was employed. The range of sentimental coordinates we used is [1 : 9].

We evaluate our classification performance using ten-fold cross validation. In our experiments, the performance of the category classification was recorded at approximately 87.2%. However, the category data used for our experiments are not the ground truth. Because we collected images using image retrieval, category data tagged by humans are required. For 200 images sampled randomly from our image collection, we surveyed a category employing a user study. We asked ten volunteers about the category of 200 images, and obtained a frequently answered category for each image. For those images, we evaluated the classification performance, which was recorded at approximately 81.5%.

Next, we evaluate our prediction performance using ten-fold cross validation. In our experiments, the performance of the sentiment prediction was recorded at approximately 83.1%. Similarly to the classification experiment, there are no ground truth sentiment data in the collected images. For this reason, we also conducted a user study to obtain the ground truth sentiment data. For 200 images sampled randomly from our image collection, we asked ten volunteers about the sentiments from images with two factors, arousal and valence. Then we calculated the average value of the two factors, and assigned it to each image as the ground truth data. For those images, the performance was recorded at approximately 74.0%. Figure 2 shows several results whose categories and sentiments are successfully predicted.

Figure 3 shows several failure results of our experiments. In some cases, a categorical keyword does not matched with ground truth category that is obtained by user survey (Figure 3(a)). For portrait images, facial expression occasionally affects the sentiment of image (Figure 3(b)). However, our method does not deal with the facial expression, so that this decreases the performance of our predictor. Figure 3 (c)-(d) show cases that wrong sentimental keywords are used.

Although the number of samples used for measuring performance was smaller than the number of images used for learning, our experiments show that the classification and prediction performance of our models approximate the performance measured using ground truth data. Therefore, we argue that the classifier and predictor generated by our approach are useful for practical purposes.

If a new image category is required, our approach can easily update the training data set by simply using image retrieval. The same approach can be employed for the sentiments from images. In general, sentiment varies according to culture, age, era, etc. Therefore, training sets for each culture, age, era, etc. are required for better performance. In this case, our approach offers a very simple and easy solution. Collecting images with specific search options that include language, region, etc. is one of the available solutions. In addition, the most important benefit of our approach is that collecting images, categories, and sentiment data does not require much time and human resources. Using our approach, establishing new data sets and models instantly is possible.



(a) gloomy landscape (error = 3.4%)



(b) lovely animal (error = 2.6%)



(c) lively sports (error = 5.5%)



(d) magnificent architecture (error = 3.9%)

**Fig. 2** Success cases of our prediction experiments. Captions denote the keywords used for image retrieval. Error means the distance between the coordinates of ground truth and prediction result on the AV space.

## 6 Conclusion

In this paper, we proposed an approach for acquiring a training data set for category classification and the prediction of sentiments from images. Using this approach, we collected a training data set, and established a predictor for the sentiments from images. First, we estimated the image category from a given image, and then we predicted the sentiments as coordinates on the AV space using the predictor of estimated categories. We showed that the performance of our approach approximated that of using ground truth data. Based on our experiments, we argue that our approach for using big data on the web as the training set for predicting sentiments from content is useful for practical purposes.

For future work, we will extend our work to various backgrounds, such as culture, age, and gender. As mentioned in Section 5, we believe that our approach can be extended easily by refining the search options. If a prediction model is established according to culture, age, and gender, the prediction performance will be enhanced dramatically.

For this paper, the evaluation was conducted using small samples with ground truth data. We will survey the ground truth category and sentiments from all the images employed in our experiments using physiology-based sen-



**Fig. 3** Failure cases of our experiments. Captions denote the keywords used for image retrieval. Error means the distance between the coordinates of ground truth and prediction result on the AV space.

timent recognition techniques [Park et al(2013)Park, Jang, Chung, and Kim] or crowdsourcing services, such as MTurk [Ipeirotis(2010)]. This will assist in obtaining more statistically reliable evaluation results.

### Acknowledgment

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(NRF-2013R1A1A2061611).

### References

- [Aurenhammer(1991)] Aurenhammer F (1991) Voronoi diagrams: a survey of a fundamental geometric data structure. *ACM Computing Surveys (CSUR)* 23(3):345–405
- [Cantone et al(2005)Cantone, Ferro, Pulvirenti, Recupero, and Shasha] Cantone D, Ferro A, Pulvirenti A, Recupero D, Shasha D (2005) Antipole tree indexing to support range search and k-nearest neighbor search in metric spaces. *Knowledge and Data Engineering, IEEE Transactions on* 17(4):535–550, DOI 10.1109/TKDE.2005.53

- [Cohen-Or et al(2006)Cohen-Or, Sorkine, Gal, Leyvand, and Xu] Cohen-Or D, Sorkine O, Gal R, Leyvand T, Xu YQ (2006) Color harmonization. In: *ACM Transactions on Graphics (TOG)*, ACM, vol 25, pp 624–630
- [Colombo et al(1999)Colombo, Del Bimbo, and Pala] Colombo C, Del Bimbo A, Pala P (1999) Semantics in visual information retrieval. *IEEE Multimedia* 6(3):38–53
- [Dan-Glauser and Scherer(2011)] Dan-Glauser E, Scherer K (2011) The geneva affective picture database (gaped): a new 730-picture database focusing on valence and normative significance. *Behavior Research Methods* 43(2):468–477, DOI 10.3758/s13428-011-0064-1, URL <http://dx.doi.org/10.3758/s13428-011-0064-1>
- [Datta et al(2008)Datta, Li, and Wang] Datta R, Li J, Wang JZ (2008) Algorithmic inferring of aesthetics and emotion in natural images: An exposition. In: *Image Processing, 2008. ICIIP 2008. 15th IEEE International Conference on*, IEEE, pp 105–108
- [Fasel and Luettin(2003)] Fasel B, Luettin J (2003) Automatic facial expression analysis: a survey. *Pattern recognition* 36(1):259–275
- [Hall et al(2009)Hall, Frank, Holmes, Pfahringer, Reutemann, and Witten] Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The weka data mining software: An update. *SIGKDD Explor Newsl* 11(1):10–18, DOI 10.1145/1656274.1656278, URL <http://doi.acm.org/10.1145/1656274.1656278>
- [Haralick(1979)] Haralick RM (1979) Statistical and structural approaches to texture. *Proceedings of the IEEE* 67(5):786–804
- [Haralick et al(1973)Haralick, Shanmugam, and Dinstein] Haralick RM, Shanmugam K, Dinstein IH (1973) Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on* (6):610–621
- [Ipeirotis(2010)] Ipeirotis PG (2010) Analyzing the amazon mechanical turk marketplace. *XRDS* 17(2):16–21, DOI 10.1145/1869086.1869094, URL <http://doi.acm.org/10.1145/1869086.1869094>
- [Itti et al(1998)Itti, Koch, and Niebur] Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence* 20(11):1254–1259
- [Kobayashi(1981)] Kobayashi S (1981) The aim and method of the color image scale. *Color research & application* 6(2):93–107
- [Lang et al(1999)Lang, Bradley, and Cuthbert] Lang PJ, Bradley MM, Cuthbert BN (1999) International affective picture system (iaps): Technical manual and affective ratings
- [Machajdik and Hanbury(2010)] Machajdik J, Hanbury A (2010) Affective image classification using features inspired by psychology and art theory. In: *Proceedings of the International Conference on Multimedia*, ACM, New York, NY, USA, MM '10, pp 83–92, DOI 10.1145/1873951.1873965, URL <http://doi.acm.org/10.1145/1873951.1873965>
- [Mayer et al(1990)Mayer, DiPaolo, and Salovey] Mayer JD, DiPaolo M, Salovey P (1990) Perceiving affective content in ambiguous visual stimuli: A component of emotional intelligence. *Journal of personality assessment* 54(3-4):772–781
- [Moon and Spencer(1944)] Moon P, Spencer DE (1944) Geometric formulation of classical color harmony. *JOSA* 34(1):46–50
- [Nicolaou et al(2012)Nicolaou, Gunes, and Pantic] Nicolaou MA, Gunes H, Pantic M (2012) Output-associative rvm regression for dimensional and continuous emotion prediction. *Image and Vision Computing* 30(3):186–196
- [Park et al(2013)Park, Jang, Chung, and Kim] Park BJ, Jang EH, Chung MA, Kim SH (2013) Design of prototype-based emotion recognizer using physiological signals. *ETRI Journal* 35(5):869–879
- [Picard(1995)] Picard RW (1995) Affective computing
- [Picard(2003)] Picard RW (2003) Affective computing: challenges. *International Journal of Human-Computer Studies* 59(1):55–64
- [Picard et al(2001)Picard, Vyzas, and Healey] Picard RW, Vyzas E, Healey J (2001) Toward machine emotional intelligence: Analysis of affective physiological state. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 23(10):1175–1191
- [Reiman et al(1997)Reiman, Lane, Ahern, Schwartz, Davidson, Friston, Yun, and Chen] Reiman EM, Lane RD, Ahern GL, Schwartz GE, Davidson RJ, Friston KJ, Yun LS, Chen K (1997) Neuroanatomical correlates of externally and internally generated human emotion. *American Journal of Psychiatry* 154(7):918–925

- 
- [Russell(1989)] Russell JA (1989) Measures of emotion. Academic Press
- [Schmidt and Stock(2009)] Schmidt S, Stock WG (2009) Collective indexing of emotions in images. a study in emotional information retrieval. *Journal of the American Society for Information Science and Technology* 60(5):863–876
- [Schubert(1999)] Schubert E (1999) Measuring emotion continuously: Validity and reliability of the two-dimensional emotion-space. *Australian Journal of Psychology* 51(3):154–165
- [Suykens and Vandewalle(1999)] Suykens JA, Vandewalle J (1999) Least squares support vector machine classifiers. *Neural processing letters* 9(3):293–300
- [Tokumaru et al(2002)Tokumaru, Muranaka, and Imanishi] Tokumaru M, Muranaka N, Imanishi S (2002) Color design support system considering color harmony. In: *Fuzzy Systems, 2002. FUZZ-IEEE'02. Proceedings of the 2002 IEEE International Conference on, IEEE*, vol 1, pp 378–383
- [Warriner et al(2013)Warriner, Kuperman, and Brysbaert] Warriner AB, Kuperman V, Brysbaert M (2013) Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behavior research methods* 45(4):1191–1207
- [Yang et al(2008)Yang, Lin, Su, and Chen] Yang YH, Lin YC, Su YF, Chen HH (2008) A regression approach to music emotion recognition. *Audio, Speech, and Language Processing, IEEE Transactions on* 16(2):448–457
- [Yun et al(2013)Yun, Kim, Park, and Kim] Yun Wh, Kim D, Park C, Kim J (2013) Hybrid facial representations for emotion recognition. *ETRI Journal* 35(6):1021–1028