Literature Review: Synthesising emotion-driven images

Konrad Cybulski

May 2019

1 Aims and Scope

The primary focus of the proposed research is to produce an image emotion classifier, and to further leverage it in the process of emotion-driven image synthesis. The two core themes explored are image generation techniques, and processes for classifying and generating emotional content. As a result this literature review aims to explore and understand the progression of knowledge in the fields of emotion representation and classification, computational image synthesis, and affective content synthesis. All of which represent the multiple facets of the proposed body of research with which an investigation into the use of image emotion classifiers for emotion-driven image generation will be explored.

2 Emotion representation

The field of emotion classification has surged in recent years given a popularity and rising interest in facial emotion recognition. Other research has focused on exploring ways in which emotion can be recognised in images, text, and more abstract content. Underlying these two areas of emotion recognition is the methods for quantitatively representing emotion in both a meaningful, and accurate way. This section will discuss the evolution of these aspects of emotion classification, in addition to their commonalities, and differences.

The computational recognition of emotion has been explored in countless studies and projects, in both the context of image emotion recognition [8, 15, 5] and facial emotion classification [11]. However a core component and key difference between a large number of such bodies of research is the method by which emotion is represented. Two of the most common representations represent a discrete, and continuous approach. Discrete emotion representations generally involve the categorisation of emotion to a series of labels. Discrete approaches represent the method used in a large number of papers [8, 1, 13, 10] however the number of emotion labels explored varies greatly. An example of the size of emotion label subsets used in studies of image emotion classification

using such a discrete model are 7 [1], 8 [8], 11 [13], and even 19 [10]. Due to this lack of consistency in emotion classification targets, such studies often resort to performing similar data gathering and classifier training methods. Despite the consequences of inconsistency in the emotion labels chosen, due to the simplicity of discrete categorical emotion assignment, data gathering can be performed with ease compared to continuous methods of emotion representation. In order to create large image datasets of labelled images can be completed with greater ease when the number of such labels is reduced. However there remains difficulty in labelling images with respect to emotion given the inherent variation in the emotion felt by someone when both viewing an abstract image, or accurately determining the emotion expressed by a facial expression.

The aforementioned difficulty associated with labelling images according to their respective emotional content is accentuated with the introduction of a dimensionally continuous representation of emotion. While proposed representations vary, the most recognised basis for many continuous models is the circumplex model of emotion [12]. The circumplex model of emotion introduced by Russell [12] asserts that emotion can be measured in terms of two continuous variables: valence, and arousal. Valence measures the positivity or negativity associated with an emotion; and arousal measures the excitement associated with it. For example, the discrete emotion label of **happiness** could be represented in the continuous valence-arousal (VA) space with high-valence, medium-arousal; while the label **depressed** would translate to low-valence, low-arousal; and **relaxed** being medium-valence, low-arousal. While this continuous dimensional model allows for greater continuity in measuring emotions, it is not without fault [7]. The primary limitations of such a model involve it's likelihood of misinterpretation particularly when considering the labels of each dimension, and any information loss arising from reducing more complex emotions to a two-dimensional space. One way in which the circumplex model of affect introduced by Russell [12] has been extended to address its weaknesses is through the introduction of a third-dimension: Dominance, the amount of control associated with an emotion. Despite its limitations, the circumplex model has been used extensively in the field of psychology [2, 14], and in computational emotion classification to a limited extent [16].

While both models for the representation of emotion have trade-offs, both suffer from issues related to cross-cultural differences in emotion expression and recognition. Cultures inherently differ with respect to how emotions are both felt and expressed [9]; this is becomes increasingly evident when attempting to classify the affective emotion embodied and expressed by more abstract content such as imagery and sound. The popular image-emotion dataset used for classification known as the International Affective Picture System (IAPS) was shown to have a significantly different valence-arousal assignment for up to 31.74% of images between Chinese and American young adults [4]. Valence-arousal values assigned to images often vary when considering a group of people. To address this potential for error in reporting the subjective emotion imparted by content, the self-assessment manikin (SAM) was introduced [6]. SAM, as can be seen in Figure 1, was developed to aid in the evaluation of emotion, particularly its

2

translation into the commonly used three dimensions of the circumplex model of affect: valence, arousal, dominance. It has proved itself as more accurate and effective than other methods of emotion self-assessment while being less complex [2].
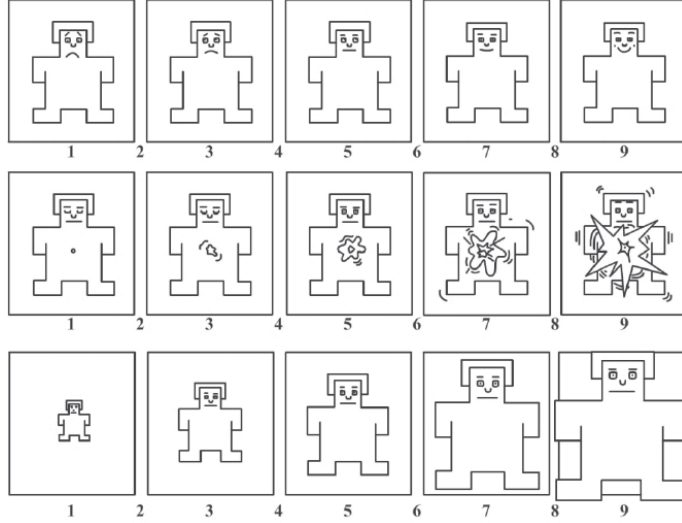

Figure 1: The self-assessment manikin: a guide for reporting individual emotional affect in the three dimensions valence (top), arousal (middle), and dominance (bottom) as introduced by Lang [6]

While both discrete and continuous methods of representing emotion have their respective benefits and trade-offs, the choice to use one over another often depends on the way in which the data gathering process for content emotion classification is performed. These two representations of emotion can be translated between, since all categorical emotions can be converted to the continuous circumplex model, and the inverse. However due to the higher complexity nature of the continuous model, even with the use of SAM, dimensional representations are more difficult to gather on a large scale in comparison to categorical emotions.

3 Emotion classification

Emotion classification has been researched in depth particularly with respect to two sub-domains: facial emotion recognition, and content emotion classification (image, text, sound). Facial emotion recognition has been more heavily researched in this field when compared to content emotion classification.

3.1 Facial emotion recognition

Humans are known to be more accurate in classifying facial emotion of another when they are culturally similar due to their in-group advantage. This was also found to be the case with computational facial emotion classifiers when trained more heavily on one culture over another [3].

3.2 Image emotion classification

4 Computational image synthesis

5 Affective content synthesis

6 Conclusion

References

[1] Ali, A. R. and Ali, M. (2017). Emotional filters: Automatic image transformation for inducing affect. *arXiv preprint arXiv:1707.08148*.

[2] Bradley, M. M. and Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59.

[3] Dailey, M. N., Joyce, C., Lyons, M. J., Kamachi, M., Ishi, H., Gyoba, J., and Cottrell, G. W. (2010). Evidence and a computational explanation of cultural differences in facial expression recognition. *Emotion*, 10(6):874.

[4] Huang, J., Xu, D., Peterson, B. S., Hu, J., Cao, L., Wei, N., Zhang, Y., Xu, W., Xu, Y., and Hu, S. (2015). Affective reactions differ between chinese and american healthy young adults: a cross-cultural study using the international affective picture system. *BMC psychiatry*, 15(1):60.

[5] Kim, H.-R., Kim, Y.-S., Kim, S. J., and Lee, I.-K. (2018). Building emotional machines: Recognizing image emotions through deep neural networks. *IEEE Transactions on Multimedia*, 20(11):2980–2992.

[6] Lang, P. J. (1980). Behavioral treatment and bio-behavioral assessment: Computer applications.

[7] Larsen, R. J. and Diener, E. (1992). Promises and problems with the circumplex model of emotion.

[8] Machajdik, J. and Hanbury, A. (2010). Affective image classification using features inspired by psychology and art theory. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 83–92. ACM.

[9] Markus, H. R. and Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological review*, 98(2):224.

[10] Mohammad, S. and Kiritchenko, S. (2018). Wikiart emotions: An annotated dataset of emotions evoked by art. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*.

[11] Mollahosseini, A., Chan, D., and Mahoor, M. H. (2016). Going deeper in facial expression recognition using deep neural networks. In *2016 IEEE Winter conference on applications of computer vision (WACV)*, pages 1–10. IEEE.

[12] Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161.

[13] Wang, Y. and Lewis, M. (2017). Arttalk: Labeling images with thematic and emotional content.

[14] Warriner, A. B., Kuperman, V., and Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behavior research methods*, 45(4):1191–1207.

[15] Zhao, S., Gao, Y., Jiang, X., Yao, H., Chua, T.-S., and Sun, X. (2014). Exploring principles-of-art features for image emotion recognition. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 47–56. ACM.

[16] Zhao, S., Yao, H., Gao, Y., Ji, R., Xie, W., Jiang, X., and Chua, T.-S. (2016). Predicting personalized emotion perceptions of social images. In *Proceedings of the 24th ACM international conference on Multimedia*, pages 1385–1394. ACM.