Synthesizing emotive art

Konrad Cybulski

March 2019

1 Introduction

In the domain of creative artificial intelligence and evolutionary art, the desire exists for processes that produce imagery that is not only visually appealing, but images that exhibit abstract and emotive characteristics. There exists extensive research on the production of realistic, and target label accurate images. The use of quality-diverse (QD) algorithms in combination with deep neural network (DNN) image classifiers were used by Nguyen et al. (2015b) to generate images with high classification accuracy from pre-trained DNN. Bao et al. (2017) exmplifies recent use of variational generative adversarial network (GAN) architectures for the generation of realistic images from fine-grained target labels. GANs have shown their use in text-to-image synthesis (Reed et al., 2016; Zhang et al., 2017), producing realistic images reflecting the detailed description from which they are generated. With regard to creative image generation, Tan et al. (2017) explored techniques for synthesizing artwork with more abstract characteristics. Through the use of a target artist or genre, the generative network produced images that were highly abstract and stylistically accurate.

Little exploration however has been done on incorporating emotion into the process of art and image generation. Ali and Ali (2017) explored the idea of *emotion transfer*, using techniques such as image emotion assignment, and color/style transfer with the aim of altering image composition to reach a target emotion. Examples given use a target profile, with varying levels of emotions such as joy, anger, and fear, to alter the image's color composition. The classification of an image's affective emotion, the emotion with which a viewer classifies the image, has been explored in various works (Machajdik and Hanbury, 2010; Chen et al., 2015; Kim et al., 2018). Kim et al. (2018) produced a classifier for recognizing the emotion attributed to an image. This was done through the application of a DNN to decompose an image to a two-dimensional feature vector (valence and arousal) representing the image's emotion mapped to a continuous plane (see Figure 2).

2 Aims

The aim of this resarch is to better understand the visual patterns associated with various emotions conveyed in images. This will primarily leverage image emotion recognition architectures explored by Kim et al. (2018), in combination with the dataset produced by Zhao et al. (2016) containing over 1.4 million images with assigned valence, arousal, and dominance levels derived from their descriptions. Producing an architecture with which valence-arousal (VA) values can be assigned to images forms the basis for this research. Such a platform allows further exploration into the way in which it assigns VA values to more complex, multi-faceted and multi-layered images, and the efficacy with which this is done. Furthermore this process, and the patterns learned by it can be better understood through the use of generative processes which maximise given target features in a quality-diverse way following the methods introduced by Nguyen et al. (2015b) and Nguyen et al. (2015a). In the context of this research, such features include valence, arousal, dominance, happiness, sadness, etc. This will allow a great understanding of the visual patterns that such an architecture learns, and any psychological or artistic parallels that can be drawn.

3 Background

3.1 Unsupervised image synthesis

- □ Image generation methods explored: evolutionary algorithms, neural networks, line drawing, etc.
- □ Measuring *goodness* of a generated image: realism, abstractness, aesthetic appeal, etc.

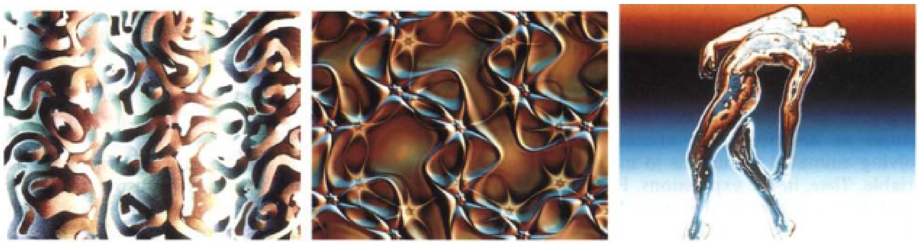


Figure 1: Images generated through the process of interactive evolution introduced by Sims (1993)

The area of image and art generation has been explored through various avenues. Some of the first *human-in-the-loop* systems such as *NEvAr* (Machado and Cardoso, 2000) produced greatly impressive images through evolutionary techniques. Evolutionary art leveraged methods introduced and exemplified by Sims (1993) such as those shown in Figure 1. Sims (1993) proposed using *Lisp* expressions for genotype definitions, which accepted a coordinate (x, y) which could be evaluated into a grayscale or RGB value, thus producing images. This genotype expression has been used in numerous further research into the process of both supervised and unsupervised image synthesis through evolutionary techniques (Machado and Cardoso, 2000; Sims, 1993; den Heijer and Eiben, 2011, 2013; Ross et al., 2006).

Despite the slow nature of the interactive process, Sims (1993) and Machado and Cardoso (2000) were able to produce images with visually striking characteristics. Ross et al. (2006) investigated measures of aesthetics for fitness evaluation in artificially evolving images. This research primarily used observations by Ralph (2006), that the distribution of colour gradients in fine art tend towards normal. While the images produced through this method did not meet the level of intricacy and detail as the results of Sims (1993) or Machado and Cardoso (2000), it represented a self-contained system able to generate appealing imagery without human interaction.

Introduction of the generative adversarial network architecture (GAN) by Goodfellow et al. (2014) allowed the process of image generation to be completely unsupervised. Common GAN application has involved the generation of realistic images, such as has been done by Bao et al. (2017), where images have been synthesized to fine-detailed target labels such as bird species' and actors. Zhang et al. (2017) and Reed et al. (2016) have recently explored text to image synthesis, in which detailed descriptions of birds and flowers have been converted into photo-realistic images using the GAN model. Tan et al. (2017) has explored the generation of art according to target genre and artist.

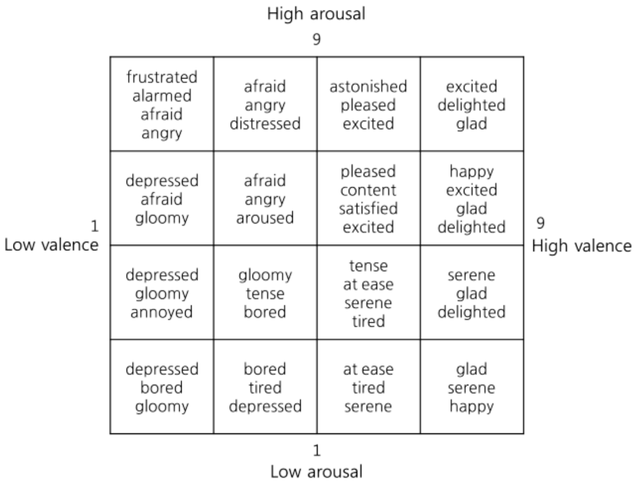
3.2 Generative adversarial networks

- □ What is the generative adversarial network, and differences in common architectures
- □ Why are GANs advantageous over the use of target feature analysis (aesthetics, etc.)
- □ Text-to-image
- □ Style transfer & image-to-image

Deep neural networks (DNN) have grown tremendously in popularity within the domain of image generation, and classification.

3.3 Image emotion recognition

- □ Sentiment classification: text & images.
- □ Emotion classification in images: facial expression, general imagery.
- □ Methods of classifying emotion in images: single target emotion, discrete categorical likelihood, decomposition into continuous vector (valence-arousal).



High arousal

9

1

Low valence

9

High valence

frustrated alarmed afraid angry	afraid angry distressed	astonished pleased excited	excited delighted glad
depressed afraid gloomy	afraid angry aroused	pleased content satisfied excited	happy excited glad delighted
depressed gloomy annoyed	gloomy tense bored	tense at ease serene tired	serene glad delighted
depressed bored gloomy	bored tired depressed	at ease tired serene	glad serene happy

1

Low arousal

Figure 2: Distribution of emotions associated with levels of valence and arousal determined by DNN classifier produced by Kim et al. (2018)

The area of image emotion and sentiment classification has been explored in a number of ways, primarily through image feature analysis derived from art and psychological factors (Machajdik and Hanbury, 2010); and more recently using techniques such as deep neural networks (Chen et al., 2015; Kim et al., 2018). Feature extraction and analysis has been used for various applications such as measuring aesthetic appeal (den Heijer and Eiben, 2010a,b, 2011) and as an emotional feature vector for sentiment classification (Machajdik and Hanbury, 2010). Due to the artistic and psychological underpinnings used by Machajdik and Hanbury (2010), the low-level features extracted from images can be understood at a high level. The relationship between an image's emotion and its core artistic components such as balance, harmony, and variety was further explored by Zhao et al. (2014), which uses a comparably small feature vector to Machajdik and Hanbury (2010), resulting however in a 5% classification increase to state-of-the-art approaches at the time.

Deep neural networks in this domain provide less transparency to the process with which emotions and sentiment are classified compared to feature analysis. The emotional content of an image can be decomposed in various ways. Image databases with singular emotion labels, and adjective-noun pairs (ANP) have been used for the training of deep neural network classifiers (Chen et al., 2014; Yang et al., 2018) with up to 200% performance gains over support vector machine classifiers.

4 Methodology

- □ Emotion profile representation
  - □ Discrete feature vector
  - □ Continuous valence-arousal space
- □ System architecture (training/evaluation method)
  - □ Generator: input type (related to emotion profile representation)
  - □ Generator: base architecture e.g. use ArtGAN (Tan et al., 2017)
  - □ Discriminator: use output image's emotion classification and error from target emotional profile as error function (autoencoder method)
  - □ Discriminator: use image label as target emotional profile and use standard logistic discrimination.
- □ Any interaction made between human and generator e.g. text-to-image synthesis using text emotion classification fed through to the generative model.

4.1 Datasets

There exist a few datasets in which non-facial images have labels of their affective emotion on the viewer. One particularly of interest to this research is the *WikiArt Emotions* dataset (Mohammad and Kiritchenko, 2018) in which images of artwork were labelled with an emotional profile. Each of the more than 4000 images in the dataset has an associated vector representing the proportion of viewers assigning a given emotion to the artwork. Emotions such as gratitude, happiness, anger, and arrogance are represented among the twenty emotions assigned to the images. Along with their respective emotional profile, each image is classified according to its artistic category (Impressionism, Baroque, etc.) and other desirable measures such as viewer rating, artwork title, artist name, and year of creation. Due to the level of detail relating to each artwork's affective emotional profile, and its classified style and category, this dataset will be investigated initially for creating the generative system detailed below.

4.2 Emotion profile representation

The method with which an emotional classification can be represented has been investigated as mentioned in the background section. With options ranging from a single target label (happy, sad, etc.), to the continuous two-dimensional circumplex model (valence-arousal) representation first introduced by Russell (1980). A model for representing emotion commonly used in classification tasks is that of a single label target, due its simplicity. However due to the subjectivity involved with the emotional classification of an image, a floating-point vector models the relative proportions with which an image's emotion is classified.

Such a model is used explicitly by Ali and Ali (2017) as a target emotion profile, and is the representation used by (Mohammad and Kiritchenko, 2018) due to statistical methods used in gathering data. The dataset created by Mohammad and Kiritchenko (2018) uses this representation of emotion to label the artwork available on WikiArt, and given it's relevance to both the domain of art, and emotion, this will be the representation first investigated.

The circumplex model of emotion will be investigated further with both the WikiArt Emotion dataset, which will use the valence-arousal (VA) decomposition presented by Kim et al. (2018) to evaluate the VA equivalence for each artwork. The process to convert the existing dataset's floating-point vector labels for emotion, to their respective VA values will involve applying the VA decomposition network to each of the dataset's images. Having both original vector and VA labels allows a direct comparison of the system's performance with both representations. A dataset has been created by Zhao et al. (2016) in which over 1,400,000 images are labelled according to the valence, arousal, and dominance values using image description text analysis. This dataset can be used in training an predictor architecture according to Kim et al. (2018) that can determine the VA values of a given image. This can then be applied to both the WikiArt dataset,and extended to a larger number of images available on WikiArt.

4.3 Generative architecture

The architecture through which images will be generated represents a core component of the entire creative system. Tan et al. (2017) specifies a GAN architecture which allows the propagation of a target style vector, and a process through which the assigned discriminator style label can be backpropagated to the generator network. Using such an architecture further allows the incorporation of emotion into the generator input.

5 Expected Outcomes & Contributions

The outcomes of this project will include a generative system with which art can be synthesized according to a target emotional profile. This system will be the combination of methods for the representation of emotion for use in a generative model, and an architecture with which such a model can be trained. Due to the exploratory nature of the project with respect to both system architecture and emotional profile representation, this research will have tested and analyzed various options and any comparative differences.

The proposed generative system will be used to create a collection of art, categorised by the target emotional profile with which they were seeded. The verification proposed involves the public exhibition of produced images, providing feedback to the generative process and pairing the generated images with a human assigned emotion label for use in any further system training.

References

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

Bao, J., Chen, D., Wen, F., Li, H., and Hua, G. (2017). Cvae-gan: fine-grained image generation through asymmetric training. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2745–2754.

Chen, M., Zhang, L., and Allebach, J. P. (2015). Learning deep features for image emotion classification. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 4491–4495. IEEE.

Chen, T., Borth, D., Darrell, T., and Chang, S.-F. (2014). Deepsentibank: Visual sentiment concept classification with deep convolutional neural networks. *arXiv preprint arXiv:1410.8586*.

den Heijer, E. and Eiben, A. (2010a). Using aesthetic measures to evolve art. In *IEEE Congress on Evolutionary Computation*, pages 1–8. IEEE.

den Heijer, E. and Eiben, A. (2011). Evolving art using multiple aesthetic measures. In *European Conference on the Applications of Evolutionary Computation*, pages 234–243. Springer.

den Heijer, E. and Eiben, A. (2013). Maintaining population diversity in evolutionary art using structured populations. In *2013 IEEE Congress on Evolutionary Computation*, pages 529–536. IEEE.

den Heijer, E. and Eiben, A. E. (2010b). Comparing aesthetic measures for evolutionary art. In *European Conference on the Applications of Evolutionary Computation*, pages 311–320. Springer.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680.

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.

Machado, P. and Cardoso, A. (2000). Nevar–the assessment of an evolutionary art tool. In *Proceedings of the AISB00 Symposium on Creative & Cultural Aspects and Applications of AI & Cognitive Science, Birmingham, UK*, volume 456.

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.

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).*

Nguyen, A., Yosinski, J., and Clune, J. (2015a). Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 427–436.

Nguyen, A. M., Yosinski, J., and Clune, J. (2015b). Innovation engines: Automated creativity and improved stochastic optimization via deep learning. In *Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation*, pages 959–966. ACM.

Ralph, W. (2006). Painting the bell curve: The occurrence of the normal distribution in fine art.

Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., and Lee, H. (2016). Generative adversarial text to image synthesis. *arXiv preprint arXiv:1605.05396.*

Ross, B. J., Ralph, W., and Zong, H. (2006). Evolutionary image synthesis using a model of aesthetics. In *2006 IEEE International Conference on Evolutionary Computation*, pages 1087–1094. IEEE.

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

Sims, K. (1993). Interactive evolution of equations for procedural models. *The Visual Computer*, 9(8):466–476.

Tan, W. R., Chan, C. S., Aguirre, H. E., and Tanaka, K. (2017). Artgan: Artwork synthesis with conditional categorical gans. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 3760–3764. IEEE.

Yang, J., She, D., Sun, M., Cheng, M.-M., Rosin, P. L., and Wang, L. (2018). Visual sentiment prediction based on automatic discovery of affective regions. *IEEE Transactions on Multimedia*, 20(9):2513–2525.

Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., and Metaxas, D. N. (2017). Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5907–5915.

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.

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.