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. (2015) 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 research is to explore the synthesis of artwork with a target emotional profile. Primarily leveraging work by Tan et al. (2017) and Kim et al. (2018) to produce a generative architecture whose output not only has desirable abstract characteristics, but shows emotive capabilities. The proposed system would generate an image that satisfies a set of emotions provided. This will investigate both the efficacy with which a generative system can create emotive images, as well as give insight into the properties attributed to various emotions portrayed in image form.

In order to test, and verify the output of such a system, generated images will be exhibited to explore their emotional effect on humans, and any dissonance between the intended, and resulting emotional profile. This will further verify the accuracy with which an emotional profile can be synthesized into affective artwork with such an architecture.

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.

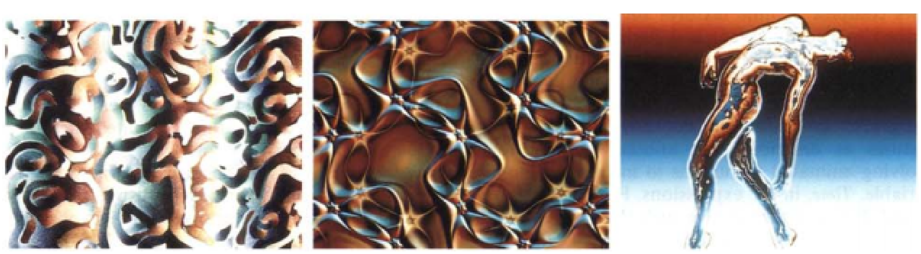

Figure 1: Images generated through the process of interactive evolution introduced by Sims (1993)

The area of evolutionary art and image generation has been explored for many years. Some of the first *human-in-the-loop* systems such as *NEvAr* (Machado and Cardoso, 2000) produced greatly impressive images leveraging 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

(Machado and Cardoso, 2000; Sims, 1993; den Heijer and Eiben, 2011, 2013; Ross et al., 2006).

Sims (1993) and Machado and Cardoso (2000) were able to produce images with visually striking characteristics, despite the slow nature of the interactive process. Ross et al. (2006) investigated the use of measures of aesthetics for fitness evaluation in artificially evolving images. This research primarily used the observation of 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 art 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).

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

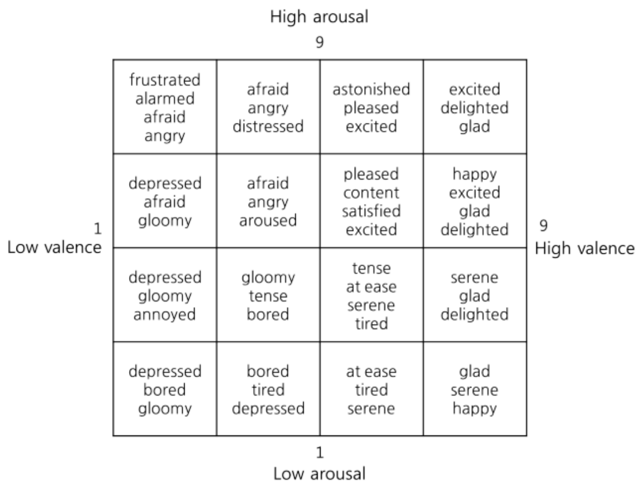

Figure 2: Distribution of emotions associated with levels of valence and arousal determined by DNN classifier produced by Kim et al. (2018)

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.

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.

Nguyen, A. M., Yosinski, J., and Clune, J. (2015). 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.

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.