

Title	顔表情画像解析による人の複雑な感情推定に関する研究 －Emotion GANs (EmoGANs)による混合表情画像の合成による解析－
Author(s)	WIN, SHWE SIN KHINE
Citation	
Issue Date	2023-09
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/18774
Rights	
Description	Supervisor:小谷 一孔, 先端科学技術研究科, 博士

Facial expressions are the most interpretable visual signals used to perceive emotions in human social communication. Each emotion corresponds to distinct facial expressions. For instance, happiness is often conveyed through a smile, while surprise is expressed by the wide opening of the eyes and mouth. With advancements in technology, emotion recognition extends not only to humans but also to computers through the development of systems that can recognize emotions from signals like facial expressions.

Consequently, in literature, facial expressions are often accompanied by their corresponding emotion labels for analysis. However, the current state of the art in emotion research primarily focuses on basic emotions, namely anger, disgust, fear, happiness, sadness, and surprise. It is important to acknowledge that human emotions are complex and not solely limited to these basic emotions. Human emotions can be a mixture of multiple basic emotions. For example, graduating from a school can evoke a combination of happiness and sadness, often referred to as 'bittersweet' emotion. This illustrates that human emotions can be diverse and heterogeneous. However, the existing literature on computer vision has limited knowledge of such mixed emotions that are composed of multiple basic emotions. Besides, deep learning models are data demanding to generalize well and work efficiently.

Therefore, in the current study, we aim to estimate the emotions of mixed facial expression images using the Analysis-by-Synthesis (AbS) approach. By following the research objectives, we expect two research milestones: synthesized mixed facial expression images and mixed emotion labels. To accomplish the first milestone, we propose a generative model called Emotion Generative Adversarial Networks (EmoGANs) to synthesize mixed facial expression images that represent mixed emotions. The EmoGANs models operate by modeling the data from the existing prior distribution, also known as the latent space, and perform mapping of samples from the latent space onto the image space for image generation. Every sample from the latent space can be mapped onto the image space by the trained generator and creates a new image. As a result, it provides a rich environment for image formation. Besides, mixed facial expressions and subject identity information can be controlled by the model parameters, which are not available in the existing research in mixed emotion literature.

In the current study, we propose four generation models named EmoGANs, EmoGANs1, EmoGANs2, and EmoGANs3, based on the stability of their adversarial training, generated image quality, and spatial resolution. The generated images by all EmoGANs models are evaluated from various perspectives, including image quality, data diversity in image generation, involvement of mixed facial expressions, ability to control the image generation process, and disentanglement property.

For the second milestone, we approach the estimation of mixed emotions as a multi-label formulation. Each generated image contains two emotion labels out of the six basic classes, such as happiness and sadness in the example case. The estimation model is compared with the other state of art models. The result reported that our model obtained higher recognition performance than the compared models. Besides, we apply the model to estimate the emotions from the basic facial expression images as well. According to the result, the model also obtained high recognition performance on basic emotions. Moreover, we tested the model on unseen real images, and it can select good features on unseen data.

Keywords: Mixed Emotions, Analysis-by-Synthesis, EmoGANs, Facial Expressions Analysis, Mixed Emotions Estimation.