

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

氏名	WIN, Shwe Sin Khine
学位の種類	博士 (情報科学)
学位記番号	博情第 509 号
学位授与年月日	令和 5 年 9 月 22 日
論文題目	Estimation of Human Complex Emotions by Analysis of Facial Expression Images —Analysis-by-Synthesis of Mixed Facial Expressions Image by Emotion GANs (EmoGANs)—
論文審査委員	小谷 一孔 北陸先端科学技術大学院大学 教授 池田 心 同 教授 長谷川 忍 同 教授 緒方 和博 同 教授 阿部 亨 東北大学 准教授

論文の内容の要旨

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.

論文審査の結果の要旨

本論文は、人の多様な心の感情が現れる複雑な顔表情を画像解析と機械学習により推定するモデルを与えた。

これまでの表情認識・感情推定の研究は主に「喜び」「悲しみ」などの基本6感情（クラス）について検討している。我々の日常ではこれら基本6表情以外に「うれし泣き」や「嘲笑」のように基本6感情と顔表情とが単純に対応しない場合をよく経験する。このため従来研究では人と機械（コンピュータ）との深いレベルでの感情コミュニケーションの実現は困難である。

本論文は上記の問題に対して「合成による解析」手法を与えた。その特徴は1) 機械学習を用いて複数の感情が混在して表出する顔表情画像を生成するモデル（EmoGAN, Emotion GAN）を提案、2) モデルはパラメータの選択により任意の複数の感情が混在した顔表情画像を生成でき、そのパラメータは機械学習におけるタグ（正解値）として用いることができる、3) パラメータ選択と識別器とを組み合わせることにより、複雑な顔表情を識別し、これに対応する感情を推定する手法を与えた。

近年、深層学習議などのAI技術の発達がめざましく、人の能力に迫るあるいは越える特性が実現されて驚くことが多い。しかしながら深層学習機は数万～数十万個もの学習データを必要とする。本論文は敵対的画像生成手法（GAN）の画像生成モデルに人の多様な感情情報をパラメータ化して特徴ベクトルを構成し、任意の複数の感情が混在する顔表情画画像を生成可能にして機械学習用のデータを生成可能にしたところが画期的である。シミュレーション実験によれば、

EmoGAN と深層学習機とを組み合わせた感情推定器は 90%程度の高い感情推定精度が得られている。また、これら研究成果は権威ある国際学術論文に採録され、高い評価を得ている。

以上、本論文は人の多様な感情が混在する顔表情認識・感情推定を高精度に実現する手法を提案し、学術的に貢献するところが大きい。本手法により、人と機械（コンピュータ）との深い感情コミュニケーションを実現でき、AI 技術として社会に与えるインパクトは大きい。よって、博士（情報科学）の学位論文として充分価値あるものと認めた。