IEEE xplore

**Abstract:**

Facial expression generation involves creating facial images with specific expressions using computational methods and finds extensive applications in face editing, film production, and data augmentation. The advent of Generative Adversarial Networks (GANs) has led to significant advancements in facial expression generation. However, images generated by these methods often suffer from issues such as overlap and blurriness, resulting in a lack of realism. To address these challenges, this paper introduces a Multi-scale Mixed Attention Generative Adversarial Network (MMA-GAN) aimed at producing highquality facial expression images. The proposed MMA-GAN incorporates global residual connections at the beginning and end of the generator to preserve skin color and ignore irrelevant background content. Additionally, a multi-scale mixed attention module is integrated within the generator to adaptively learn features of key regions, thereby enhancing the learning of critical areas in the images. Experiments conducted on the publicly available AffectNet dataset validate the effectiveness of the MMAGAN model. Results indicate that MMA-GAN outperforms related methods in both qualitative assessments and quantitative analysis metrics.

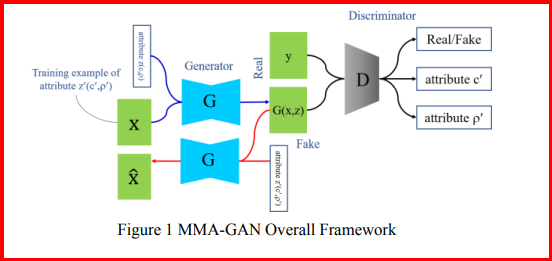
**Data Sources:** The images in AffectNet [2] were obtained through search engines using relevant keywords that correspond to the eight basic emotions (anger, disgust, fear, happiness, sadness, surprise, neutral, and contempt) as well as other complex emotional expressions derived from psychological research.

**Data preprocessing:** The dataset employed was meticulously divided into two main parts: the training set and the test set. This division was completed during the data preprocessing stage, ensuring the independence of the training and evaluation processes.

Furthermore, to accurately capture facial expressions, we utilized advanced automatic face detection algorithms to identify and crop the facial regions in each image. To ensure the consistency of the input data, all facial expression images were uniformly resized to a resolution of 128128 pixels and set to a 3-channel (RGB color mode). The primary purpose of the training set is for the training and validation processes of the network model, optimizing the model parameters and structure; meanwhile, the test set is used to evaluate the performance of the trained network model, which helps in verifying the model's generalization ability to unseen data.

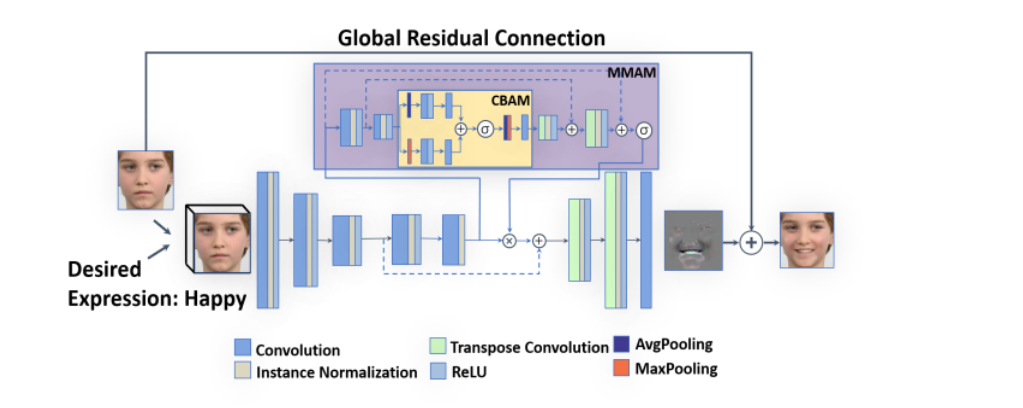
**FLOW CHART DIAGRAM OF MMA-GAN:**

we propose a novel Generative Adversarial Network architecture named Multi-Scale Mixed Attention Generative Adversarial Network (MMA-GAN), which is capable of generating images with both continuous and discrete attribute features. Specifically, the input image x is combined with continuous attribute values and discrete attribute labels c to generate images x with subtle attribute variations. is a real value in the interval [0,1], representing the change in attribute strength, while c indicates the attribute labels that need to remain unchanged. The attribute vector is encoded in the form of z=(c,), where serves as an adjustable continuous variable, which, when combined with c, forms the attribute encoding z. This combination is achieved through element-wise multiplication ⊙.

****

**Multi-scale Mixed Attention Generator**

To make the model focus only on the parts that need to be changed according to the text, without generating parts unrelated to the expression, this paper adds a Global Residual Connection (GRC) Module at the beginning and end of the model**.** The core of MMARM includes an encoder-decoder network, mixed attention module, and residual connections. The encoder extracts features at various scales, and the decoder uses these features to recreate the feature map, aiming for an output matching the original's dimensions.



**Discrimimator**

Discriminators typically map an input image to a single real number, representing the probability of the image being a real sample. In contrast, the approach of PatchGAN maps the input to a matrix, where each element of the matrix represents the probability of the corresponding image patch being a real sample.