

## Deep Learning Based Text to Image Synthesis for Accurate Facial Composites in Criminal Investigations

**Authors :** Zhao Gao, Eran Edirisinghe

**Abstract :** The production of an accurate sketch of a suspect based on a verbal description obtained from a witness is an essential task for most criminal investigations. The criminal investigation system employs specifically trained professional artists to manually draw a facial image of the suspect according to the descriptions of an eyewitness for subsequent identification. Within the advancement of Deep Learning, Recurrent Neural Networks (RNN) have shown great promise in Natural Language Processing (NLP) tasks. Additionally, Generative Adversarial Networks (GAN) have also proven to be very effective in image generation. In this study, a trained GAN conditioned on textual features such as keywords automatically encoded from a verbal description of a human face using an RNN is used to generate photo-realistic facial images for criminal investigations. The intention of the proposed system is to map corresponding features into text generated from verbal descriptions. With this, it becomes possible to generate many reasonably accurate alternatives to which the witness can use to hopefully identify a suspect from. This reduces subjectivity in decision making both by the eyewitness and the artist while giving an opportunity for the witness to evaluate and reconsider decisions. Furthermore, the proposed approach benefits law enforcement agencies by reducing the time taken to physically draw each potential sketch, thus increasing response times and mitigating potentially malicious human intervention. With publically available 'CelebFaces Attributes Dataset' (CelebA) and additionally providing verbal description as training data, the proposed architecture is able to effectively produce facial structures from given text. Word Embeddings are learnt by applying the RNN architecture in order to perform semantic parsing, the output of which is fed into the GAN for synthesizing photo-realistic images. Rather than the grid search method, a metaheuristic search based on genetic algorithms is applied to evolve the network with the intent of achieving optimal hyperparameters in a fraction the time of a typical brute force approach. With the exception of the 'CelebA' training database, further novel test cases are supplied to the network for evaluation. Witness reports detailing criminals from Interpol or other law enforcement agencies are sampled on the network. Using the descriptions provided, samples are generated and compared with the ground truth images of a criminal in order to calculate the similarities. Two factors are used for performance evaluation: The Structural Similarity Index (SSIM) and the Peak Signal-to-Noise Ratio (PSNR). A high percentile output from this performance matrix should attribute to demonstrating the accuracy, in hope of proving that the proposed approach can be an effective tool for law enforcement agencies. The proposed approach to criminal facial image generation has potential to increase the ratio of criminal cases that can be ultimately resolved using eyewitness information gathering.

**Keywords :** RNN, GAN, NLP, facial composition, criminal investigation

**Conference Title :** ICCGCV 2018 : International Conference on Computational Geometry and Computer Vision

**Conference Location :** Tokyo, Japan

**Conference Dates :** October 08-09, 2018