International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 08 | Aug 2022 www.irjet.net p-ISSN: 2395-0072

Generative Network for 3D Face Modelling

Varanasi L V S K B Kasyap[1], Battula Karthikeya[2]

[1]*School of Computer Science and Engineering, VIT-AP University, India*

Abstract - *3D Face modelling is not same as 2D Face image generation using DeepFake. This paper suggests a model, in solving the problem of responsive 3D face generation using less training data. By using Deep Convolutional Neural Networks (CNNs), the loss function is defined on feature maps. Optimization problem is solved using Stochastic Gradient Descent (SGD). Generative Adversarial Networks (GANs) are used here to generate 3D Face Model from feature maps. Recurrent Neural Network (RNN) makes it to classify the image to be progressed or not. This model is evaluated against dataset generated with 30 people in laboratory and validates the acceptable performance and boosts up the Inception Score (IS) in 3D Face generation with contemplate limits 3D Face modelling is not same as 2D Face image generation using DeepFake. This paper suggests a model, in solving the problem of responsive 3D face generation using less training data. By using Deep Convolutional Neural Networks (CNNs), the loss function is defined on feature maps. Optimization problem is solved using Stochastic Gradient Descent (SGD). Generative Adversarial Networks (GANs) are used here to generate 3D Face Model from feature maps.*

Key Words: Face Modelling, GANs, Feature Modelling.

1. INTRODUCTION

Generating 3D Faces from images of 2D Faces is the predominant application of the recent Generative Neural Networks. Besides generating the virtual 3D Faces, features of the face to be generated are obtained using CNNs and by training them over RNN gives the better result. Generating faces using Conditional Generative Adversarial Networks(cGANs) [8], makes the face more realistic. So far, CNNs are used in semantic segmentation, 2D image generation [2], Mu Li et al. developed a method to embed the Human-Identity in CNN. However, the combination of GANs and CNNs could be able to generate 3D Faces with less training data (images, video clips). Yu Song et al. developed Face Recognition Algorithm using facial feature data extraction [1], follows the Euclidean distance measure, Angel Feature measure, Curvature Distance measure and Volume Feature measure. Cahit et al. work of designing a Face Recognition system [3] is used in this paper to recognize the primary facial features (eyes, nose, ear, mouth, forehead). Facial Features play key role in identifying a person. Human Brain also uses these features in recognizing human faces and this spatial data is stored in the synapses, the work of Y.

2. Area Measurement

The distance between the features of the face can be calculated using line Euclidean distance between eye-forehead, eye-eye, eye-nose, nose-mouth, eye-mouth, eye-ear. The face can be divided into smaller regions and calculating the area of these small regions and storing them in the database help to generate 3D Face model, however finding the regions of face basing on only one projection is not enough to make 3D Face model, since the other side of the facial data is lost. To overcome this problem, model at least needs three images of same person i.e., Front View, Left Side View, Right Side View and divided as ten regions, six regions, six regions simultaneously. The ten regions of the face for Front view Projection are FR1: the area of the triangle with left eye, glabella(lateral point on forehead)and left ear as vertices, FR2: the area of the triangle with left eye, right eye, glabella as vertices, FR3: the area of the triangle with right eye, glabella and right ear as vertices, FR4: the area of the triangle with left eye, nose and left ear as vertices, FR5: the area of the triangle with left eye, right eye and nose as vertices, FR6: the area of the triangle with right eye, nose and right ear as vertices, FR7: the area of the triangle with left ear, nose and left end point of mouth as vertices, FR8: the area of the quadrilateral with left end point of mouth , nose, right end point of mouth and upper end point of upper lip as vertices, FR9: the area of the triangle with right end point of mouth, nose and right ear as vertices, FR10: the area of the triangle with left end point of mouth ,upper end point of upper lip ,right end point of mouth and right end point of mouth as vertices. The six regions of the face for Left Side View Projection are LS1: the area of the triangle with left eye, lateral point of head and Imaginary Point1(shown in Figure3), LS2: the area of the quadrilateral with left eye, nose, Imaginary Point1,Imaginary Point2 (shown in Figure3) and nose, LS3: the area of the triangle with nose ,mouth and Imaginary Point2, LS4: the area of the triangle with left ear, lateral point of head and Imaginary Point1, LS5: the area of the triangle with left ear, Imaginary Point1,Imaginary Point2, LS6: the area of the triangle with left ear, mouth, Imaginary Point2.

3. Identity Loss

For transferring the attributes of 2D Faces into 3D model, CNN is used to develop a feature vector, the 2D Face embeddings are enumerated using encoders presented in [9]. Encoder maps images and features to parallel embedding space such that all the features essential for 3D Face model are mapped to feature vector with a high inner product. Since

2D Face images with kindred content should have kindred CNN features, the L2 loss defines the perpetual loss. As given in [2] Let the Ψ be the Face network, □(f) be the feature map of the fthCNN layer with reference to the input image Image1, Ri be the aspect ratio of the image feature map (Height (Hi) * Width (Wi)) The perceptual loss in between the two images Img1 and Img2 of the fth CNN layer is given as L2 loss between two feature map representations.

4. Model Design

The generator and discriminator in the GANs use convolutional architecture like Deep Convolutional Generative Adversarial Network(DC-GAN)[9]. A noise vector in the generator of 128 dimensions, sampled from N(0,1). The features of the face are passed to the function Φ and the output Φ(□□) = Ψ, is flattened to 128 dimensions via fully connected layer with leaky Rectified Linear Unit(leaky ReLU) activation function. The output is then chained with the noise vector and transformed to a linear projection and then deconvolution is done using leaky ReLU activation till the 64*64*64*3 dimension is achieved. In the discriminator, the output 3D face model of generator is passed as input model through series of the convolutional layers. The spatial dimension of the face becomes 4*4, the feature embeddings are compressed to a vector by fully connected layer. These compressed feature embeddings here are spatially recreated and merged in deeper layers of convolutional feature network. The focus of this paper is not much on the inner architecture of the discriminator and generator. Accordingly, to obtain better performance there is slight deviation from DC-GAN architecture and appended one more residual layer[9] in discriminator and two more residual layers in generator[9].

5. Result Analysis

The model is validated on 2D image data set generated with 30 people in the laboratory. For 2D image-based evaluation, the model is trained on Kaggle face data and achieved a performance of 97.83%. Model is also trained on the ResNet and showed the performance measure of 88.2% which is 9% less than Kaggle dataset performance. For video-based evaluation, the video clips are taken as small frames of images as unit and empirically tested, with the performance of 92.23%.Yet, the best performance can be expected with this model using video clips. The training accuracy and training loss of the model is shown in Figure 8.The Inception Score is also adopted to assess quality and divergence of the generated 3D Face model, the L2 construction error between the distance features and the 3D Face model demonstrates this model can better conserve the spatial data. This model implements passage□wise attention in both generator and discriminator, to better obtain effectiveness of this mechanism, intermediate results are visualized and corresponded to feature maps in different stages.

5. Conclusion

The proposed model of combined CNN and GAN, which can generate 3D Face model using the 2D Face image data is based on the regional area parameters and Euclidean distance measures between the key features on the face in 2D image. The novel component introduced in this model is the combination of CNN and GAN architecture that can visualize the 3D face better than CNN and RNN architecture. The endorsement of Inception score and perpetual loss is used in decreasing the randomness of the 3D Face model. Experimental results show the effectiveness and divergence in the 3D Face models generated using this model.

REFERENCES

[1] Y. Song, W. Wang, & Y. Chen, “Research on 3D Face Recognition Algorithm,” First International Workshop

[2] on Education Technology and Computer Science,” Convolutional Network for Attribute-driven and Identity-

[3] Preserving Human Face”,2009.

[4] 2. Mu Li, Wangmeng Zuo & David Zhang, “Convolutional Network for Attribute-driven and Identity-preserving

[5] Human Face Generation.” ArXiv abs/1608.06434 (2016).

[6] 3. Gürel Cahit & Erden Abdulkadir, “Design of a Face Recognition System”, “The 15th International Conference

[7] on Machine Design and Production, Denizli, Turkey”2012.

[8] 4. Li Yuezun & Lyu Siwei, “Exposing DeepFake Videos by Detecting Face Warping Artifacts”, “CVPR

[9] Workshop”, 2018.

[10] 5. Thanh Thi Nguyen, Cuong M. Nguyen, Dung Tien Nguyen, Duc Thanh Nguyen & Saeid Nahavandi, “Deep

[11] Learning for Deepfakes Creation and Detection: A Survey”, ArXiv abs/1909.11573 (2019).

[12] 6. Choi, Hyeong-Seok, Changdae Park, Kyogu Lee, “From Inference to Generation: End-to-end Fully Self-

[13] Supervised Generation of Human Face from Speech.” ArXiv abs/2004.05830 (2020).

[14] 7. Wei Wei, Jiayi Liu, Xianling Mao, Guibing Guo, Feida Zhu, Pan Zhou, Yuchong Hu, “Emotion-aware Chat

[15] Machine: Automatic Emotional Response Generation for Human-like Emotional Interaction",

[16] ArXiv abs/2106.03044 (2021).

[17] 8. Bowen Li, Xiaojuan Qi, Thomas Lukasiewicz, Philip H. S. Torr, "Controllable Text-to-Image Generation",

[18] ArXiv abs/1909.07083 (2019)

[19] 9. Bodnar Cristian, "Text to Image Synthesis Using Generative Adversarial Networks", ArXiv:1805.00676(2018)