International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

IJISAE

ISSN:2147-6799 www.ijisae.org Original Research Paper

Ensemble Model for 3D Face Reconstruction using Dual UNET along with PNCC and Depth Filtering Fine-Tuning for Forensic Investigation

Sincy John*[1], Ajit Danti[2]

Submitted: 29/12/2023 **Revised**: 05/02/2024 **Accepted**: 13/02/2024

Abstract: In the realm of facial recognition, accurate reconstruction of three-dimensional (3D) facial structures plays a vital role in various applications such as verification, biometrics, and forensic investigation. Existing models depend on the availability of labeled data for effectively reconstructing facial images. However, it is challenging to obtain labeled datasets which provide a diverse set of facial images with 3D face geometry. As a result, most of the research works develop synthetic data using morphable facial images. This research paper presents a novel approach to enhance 3D facial reconstruction by implementing Dual UNet architecture with Projected Normalized Coordinate Code (PNCC) and Depth Filtering Fine-Tuning (DFFT). The proposed methodology leverages the semantic segmentation and feature extraction abilities of Dual UNet framework in order to obtain color and depth information. This information is used to understand facial geometry and texture which significantly helps in the reconstruction of facial images. The PNCC used in this research enhances the capacity of the Dual UNet model to represent the nonlinear relationships within facial features. In addition, the PNCC with DFFT helps in modelling complex facial expressions, and thereby improves the reliability of 3D facial reconstruction. Experimental results demonstrate that the reconstruction of 3D face shapes with geometry details from only a single input image can efficiently be performed using the proposed approach.

Keywords: *3D Facial Reconstruction, Plane Normalized Coordinate Cross-Correlation, Depth Filtering Fine-Tuning, Dual UNet, Face Shapes*

1. Introduction

The field of computer vision has witnessed remarkable advancements over the years. One such advancement is facial reconstruction which plays a pivotal role in various applications spanning from animation and virtual reality (Thies et al., 2016a) [1] (Chen et al., 2018) [2] (You et al., 2021) [3] to biometric security and medical imaging (Wang & Zhang, 2023) [4]. Human faces exhibit numerous nuanced features and expressions that act as a medium for interpersonal communication. In the digital era, there is a great demand for constructing three-dimensional (3D) facial models with high accuracy and precision (Zollhöfer et al., 2018) [5]. 3D facial reconstruction is one of the prominent tools to precisely analyze the facial attributes which effectively helps the research community (Tarassoli et al., 2020) [6]. Various techniques have been introduced to interpret facial attributes such as face animation (Ichim et al., 2015) [7] (Thies et al., 2016b) [8] and morphable models (Tran et al., 2019) [9] (Tewari et al., 2021) [10]. However, these techniques are affected by different factors such as varying facial expressions, poor illumination, and position of the facial images. These factors affect the reliability of the reconstruction process. Hence, developing an automatic model for 3D face reconstruction is still a challenging task which needs to be addressed. Recently various research works have attempted to address the challenges related to facial reconstruction using a single image (Deng et al., 2019) [11] (Jiang et al., 2018) [12]. Most of the works have used model-based 3D face encoding (Tewari et al., 2018) [13] for different models such as 3D morphable model (3DMM) (Booth et al., 2018) [14] and Face Warehouse (Cao et al., 2014) [15]. However, the works that employ these models use shade in most of the image data and highlights the data to stabilize the 3DMM models. On the other hand, certain works have also used shape from shading (SFS) technique (Abada & Aouat, 2016) [16] which increases the problems related to the recovery of 3D shapes from shading variety. Fundamentally, SFS based techniques use the similarities observed in frivolous 3D facial models as a reference which are not effective in reconstructing 3D facial images. A convolutional neural network (CNN) is employed in (Richardson et al., 2017) [17] for reconstructing 3D facial images from a single image. The CNN model was used to reconstruct facial images by considering all details present within a single image. Although neural network models are effective in reconstructing 3D faces from a single image, there are certain drawbacks. Firstly, model-based techniques rely on the pre-defined data and hence are not suitable if the facial images are characterized with wider margins wherein the wrinkles and folds in the images are not properly encoded due to low dimensions. Shading based approaches can recreate facial points based on the shading cues. But these

[1] Department of Computer Science and Engineering, Christ University, Bengaluru, India
[2] Department of Computer Science and Engineering, Christ University, Bengaluru, India
** Corresponding Author Email: sincyjohn2@gmail.com*

approaches require a larger number of labeled data which are not easily available.

This research intends to address these drawbacks by developing a novel approach which can reconstruct 3D facial images from a single image. The main aim of this work is to assist in the forensic investigation process which majorly depends on facial images as evidence. The main contributions of this research are outlined in the points below.

●A novel Dual Net architecture is presented in this research for an effective reconstruction of 3D facial images. The proposed framework can capture the nonlinear relationships within different facial features which is critical in the reconstruction process.

●The Dual Net model is combined with PNCC technique for extracting complex facial expressions and enhancing the performance of the facial reconstruction model. In addition, the DFFT normalizes the depth of the channel and the filtering techniques enhance the quality of the depth map thereby improving the accuracy of the reconstructed 3D face.

●The contrast of the images is enhanced using a CLAHE technique which improves the quality by preventing over-amplification of noise in different image areas.

●The performance of the facial reconstruction model is quantitatively determined using evaluation metrics such as MAE, MSE, RMSE, Z Score, depth resolution, Skewness etc.

The remaining sections of the paper are structured as follows: Section 2 reviews various related works done for facial image reconstruction. Section 3 briefs the design and implementation of the Dual UNet model for 3D facial reconstruction and Section 4 discusses the results of the experimental analysis. Section 5 outlines the conclusion of the paper with key findings and future scope.

2. Related Works

Reconstruction of 3D facial images from single images is the most appealing topic of research in recent times. A large number of techniques have been introduced to solve the issues related to facial reconstruction. Initially, parametric techniques were used to represent the shape of the faces which are directly obtained from SKSS14, AMN19, AMAC17 (Rotger Moll et al., 2019) [18] or other existing datasets. For reconstructing faces using RGB images few works have employed specific templates. But the effectiveness of these works are affected due to the variations in the shape and size of the face images. In addition, human faces change with age and factors such as wrinkles, fine lines etc increase the complications and conventional techniques fail to capture these intricate details (Zhao & Qi, 2022) [19]. To address this drawback, (Yang et al., 2021) [20] proposed a photo-to-sketch based approach for obtaining training data to train the model for reconstruction from images and sketches. Furthermore, a unique loss function is incorporated in this study for refining the characteristics of the images. In comparison to other models, the proposed framework performs well in terms of reconstructing the images from the sketches.

The authors in (Tu et al., 2020) [21] worked to address the challenges related to the adoption of 3DMM by proposing an advanced 2D-based self-supervised learning (2DASL) method which uses two dimensional images with landmark data for improving the learning ability of the 2DASL for reconstruction. It was observed from the experimental outcome that this model achieved excellent performance in terms of 3D reconstruction and facial alignment. The emergence of deep learning (DL) has opened up a lot of new opportunities for the researchers to explore the facial reconstruction process. The application of deep neural networks (DNN) is discussed in (Dou et al., 2017) [22]. The DNN framework is used for reconstructing facial images from end to end using a single image. The architecture of DNN is combined with CNN for improving the reconstruction of facial expressions. Results validate the superiority of the DNN model. A voxel-based approach for facial reconstruction is presented in (Sharma & Kumar, 2020) [23] using DL. This framework utilized the variational auto encoders and BiLSTM for training the model for accurately reconstructing 3D facial images. A neural network model is proposed in (Chen et al., 2023) [24] along with an optimization algorithm known as simulated annealing (SA). The SA algorithm is used for extracting relevant features from the data along with labeling facial features and reconstruction of 3D images. The inclusion of SA significantly improved the performance of the DL model and it can be inferred that DL models provide highly accurate results and can effectively address the drawbacks of conventional techniques (Sharma & Kumar, 2022) [25]. Motivated by this aspect, this research employs a Dual UNet architecture for achieving accurate 3D facial reconstruction.

3. Proposed Research Methodology

The preliminary aim of this research is to understand the facial geometry from the 2D image and reconstruct 3D facial images. For this, this research deploys a Dual UNet architecture which accurately represents the facial features by concatenating all the coordinates of the 3D faces. The stages involved in the proposed approach are shown in figure 1 and are discussed in the below subsections:

3.1. Data Preparation

The facial images collected from the input are prepared for reconstruction by converting the color space of the image from BGR (Blue, Green, Red) to RGB (Red, Green, Blue). Further, the RGB color images are converted into grayscale wherein all images are in different shades of grey. Face detection is performed on the grayscale images since it is effective and exhibits better results. In this research, a pre-trained Haar Cascade classifier is used for detecting frontal faces. The Haar cascade classifier incorporates the ability of machine learning which is trained to detect faces from the images. The classifier uses a multiscale function for detecting faces from the grayscale images. Since the facial images obtained from the datasets are in varying sizes and orientations, they are resized or scaled using a scaling factor. The scaling factor plays a crucial role in the facial reconstruction process since it allows for a finer-grained scale detection. In this work, a scaling factor of 1.3 is used for face detection. The faces are identified by drawing a rectangle around the detected face.

3.2. Facial key-point detection using fast feature technique

A fast feature technique is employed for accurately identifying different facial key points such as eyes, nose, etc from the images. This technique leverages the advantage of image processing techniques such as quality enhancement, image smoothening, filtering, and background removal.

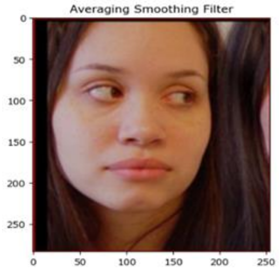


Fig 2: Average smoothened image

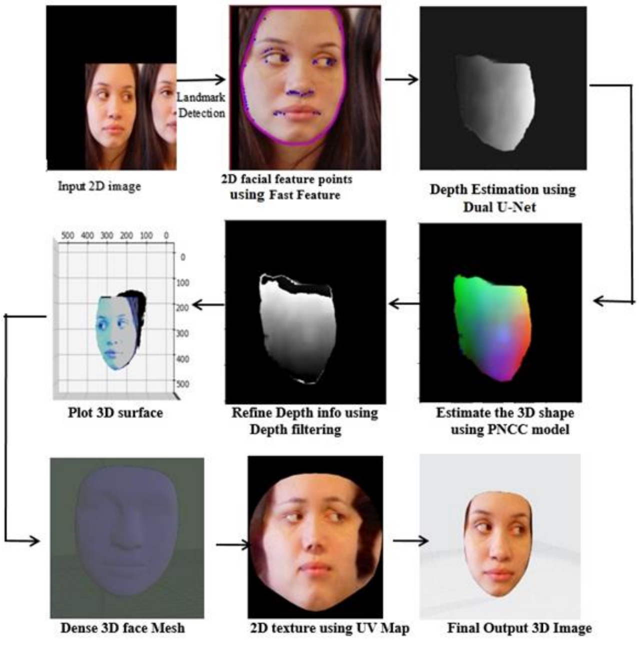

Landmark Detection
Input 2D image
2D facial feature points using Fast Feature
Depth Estimation using Dual U-Net
Plot 3D surface
Refine Depth info using Depth filtering
Estimate the 3D shape using PNCC model
Dense 3D face Mesh
2D texture using UV Map
Final Output 3D Image

Fig 1: Flow of the proposed approach

Initially, a 2D filter is applied to the images which performs 2D convolution operation on the facial images using kernels. The convolution operation applies a sliding kernel matrix to the input image and multiplies the values of overlapping pixels and aggregates the results. The convolution operation is also used in performing different actions such as blurring, sharpening, and edge detection. Further, a median blur operation is applied to smoothen the images using a median filter. In this process, the pixel values are replaced with the median value of the pixels. In this way, the noise in the image is reduced while preserving the edges. The smoothened image is shown in figure 2.

In the smoothening process, the BGR images are converted into LAB color space which distinguishes luminance (L), chrominance A (a), and chrominance B (b) components. The quality of the images are enhanced using a CLAHE (Contrast Limited Adaptive Histogram Equalization) technique. CLAHE technique enhances the contrast and improves the visualization of the image features especially when the images have non-uniform illumination. This technique applies a histogram equalization process to the image areas where some regions are too bright or dark and thereby limits the noise amplification and prevents over saturation of bright areas.

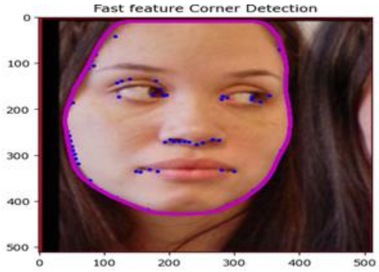


Fig 3: Facial Key point detection using Fast Feature Technique

Further, the images are subjected for background removal wherein the background from the facial images are removed in order to obtain a clearer representation of the face. A Canny Edge Detection (CED) technique is applied to the pre-processed image to identify the boundaries of the faces from the images and the features are detected using a corner detection approach. The corner coordinates of the images are identified and an overall 68 coordinates are identified in this work. The obtained coordinates for all images are stored in the local directory and are used to represent the original image using a Matplotlib function which visualizes the image with all detected corners and contours as shown in figure 3.

3.3. 3D Volumetric Face Reconstruction using Dual UNET Model

The Dual UNet is a neural network architecture which consists of two UNets into a single model for performing complex tasks. The UNet is a type of CNN with “U” shaped architecture that consists of an encoding path and a decoding path. In this architecture, the encoding path captures the information about facial features through a series of convolutional and pooling layers. This information is used by the decoding path to generate the output. The two UNets incorporated in the architecture are trained to work simultaneously for improving the quality and accuracy of the 3D facial reconstruction process. Dual UNet is mainly selected because of its ability to enhance the quality of images even during occlusions and presence of noise and to extract complex facial features from the image.

In this research, convolutional and deconvolutional blocks are used for constructing the U-Net architecture. These blocks perform convolutional operations using a ReLU activation function and normalization using a batch normalization technique. Further, a transposed convolution and deconvolution for up sampling and down sampling purposes.

The convolutional block function uses the Leaky ReLU activation function followed by a 2D convolutional layer and BatchNorm2d layer. On the other hand, the deconvolutional block function consists of a ReLU activation followed by a 2D transposed convolutional layer and BatchNorm2d layer in the U-Net architecture. For down sampling an encoder is used that consists of a series of convolutional blocks in order to reduce the spatial dimensions of the input. Each block reduces the spatial resolution through convolutional operations and increases the number of channels. For up sampling, a decoder is used that consists of a series of deconvolutional blocks to up sample the low-resolution feature maps to preserve prominent spatial information. The encoder and decoder are connected through skip connections which also helps in concatenating feature maps and thereby preserving the intricate and fine details during up sampling. The U-Net architecture uses a Dual U-Net class which combines two U-Nets for converting 2D images to 3D images. The constructor in the architecture initializes the Dual U-Net model which defines the 2D U-Net (Encoder) and the 2D U-Net (Decoder) using convolutional and deconvolutional blocks. The convolutional blocks (conv_down1 to conv_down8) are defined in the Encoder to down sample the input image, while the deconvolutional blocks (conv_up1 to conv_up11) are defined in the Decoder to up sample the features back to the original size. The convolutional blocks (conv_9 to conv_11) perform final 2D convolutions to generate the final output feature maps.

Mathematically x1 and x2 are defined as Input images (2D face images) and correspondingly y1 and y2 are the down-sampled features of x1 and x2, z1 and z2- Up-sampled features of y1 and y2 are the concatenated output also defined as final concatenated output of the Dual UNet. A forward method is employed to define the forward pass of the Dual U-Net model which takes two inputs simultaneously. The 2D U-Net (Encoder) processes the input 1 through the convolutional blocks, and the 2D U-Net (Decoder) up samples the features using deconvolutional blocks. The features obtained from both encoder and decoder are concatenated to generate final output feature maps. The input 2 is also processed independently using a different set of convolutional and deconvolutional blocks. The final output consists of the output feature maps from both the 2D and dual input branches, along with the concatenated output feature maps. Finally, the Dual U-Net architecture takes a 2D input image and its corresponding dual input image as inputs and performs both down sampling and up sampling operations using convolutional and deconvolutional blocks. The operation of the down sampling and up sampling process are mathematically defined as follows:

Downsampling operations

$$\mathrm{y}_1 = \mathrm{CB}\ (\mathrm{x}_1)\mathrm{and}\ \mathrm{y}_2 = \mathrm{CB}\ (\mathrm{x}_2)\ldots(1)$$

$$\mathrm{y}_1 = \mathrm{L\_ReLU}\ (\mathrm{WConv}_1 * \mathrm{x}_1 + \mathrm{bConv}_1)\ldots(2)$$

$$\mathrm{y}_2 = \mathrm{L\_ReLU}\ (\mathrm{WConv}_1 * \mathrm{x}_2 + \mathrm{bConv}_1)\ldots(3)$$

where, CB is the convolutional block, L_ReLU is the Leaky rectified linear unit activation function, Wconv1 is the convolutional weight, bconv1 is the convolutional bias which also performs convolution.

Up sampling operations

The upsampling operations are performed using the parameters of the deconvolution block using Input: y1 and y2 and Output: z1 and z2 which are mathematically represented as follows:

$$\mathrm{y}_2 = \mathrm{L\_ReLU}\ (\mathrm{WConv}_1 * \mathrm{x}_2 + \mathrm{bConv}_1)\ \ldots (4)$$

$$\mathrm{Z}_1 = \mathrm{ReLU}\ (\mathrm{Wdeconv}_1 * \mathrm{U}\ (\mathrm{y}_1) + \mathrm{bdeconv}_1)\ \ldots (5)$$

$$\mathrm{Z}_2 = \mathrm{ReLU}\ (\mathrm{Wdeconv}_1 * \mathrm{U}\ (\mathrm{y}_2) + \mathrm{bdeconv}_1)\ldots (6)$$

Where, deconv1 is the deconvolutional weight, b deconv1 is the deconvolutional bias, '∗' denotes deconvolution, U is the up sampling operation and is the rectified activation function. The outputs Z1 and Z2 are concatenated to generate the combined output as shown in equation 7.

$$\mathrm{CO} = \mathrm{Concatenate}\ (\mathrm{Z}_1, \mathrm{Z}_2)\ldots.\ (7)$$

Where, CO is the concatenated output.

Few convolution blocks are added in addition to further process the concatenated features and for all the features, a forward pass (FP) is applied which is given as follows:

FP (x_1,x_2)=(CO y_1,y_2,z_1,z_2) …. (8)

The output of the Dual U-Net model consists of the 2D and 3D feature maps, which can be further used for facial reconstruction tasks. The architecture is highly appropriate for the applications which require two input images and produce a final combined output. The final images are adjusted and cropped using a crop and adjust function which also helps in adjusting the size of the input image based on the specific bounding box. The bounding box can also be adjusted and crops the images according to the bounding box, increasing or decreasing the size based on the specific size.

3.4. Estimate the 3D shape and texture of the face using a PNCC model

PNCC is calculated using the Z-Buffer function, which takes as input the 3D vertex positions and the NCC texture (NCC). This operation effectively maps the NCC texture onto the 3D face model. For each visible point in the reconstructed 3D model of an object, the PNCC encodes its position as a 2D coordinate on the image plane. These 2D coordinates represent where each vertex would appear in the rendered image, allowing for easy visualization and analysis of the 3D structure in a two-dimensional context.This process enables the creation of a visually realistic representation of the 3D face model, with color information derived from the NCC texture and parameterized adjustments applied to account for facial identity and expression variations.

3.5. Depth Filtering:

A depth filtering technique is applied for normalizing the depth channels and a binary mask is created. This technique creates a normalized grid based on network results, and adjusts depth values, while handling invalid values in the depth data. The normalization and scaling are expressed as follows:

$$\mathrm{im}_{\mathrm{depth}} = \mathrm{im}_{\mathrm{depth}}.\mathrm{astype}(\mathrm{np.float64}), \mathrm{net}_{\mathrm{X}} \\ = \mathrm{im}_{\mathrm{depth}[:,:,0]}.(1.3674)/255 \ldots (10)$$

The depth values obtained from equation 10 are converted to float values and the X components are coordinates of the grid obtained in terms of depth and shape.

$$\mathrm{X} \\ = \mathrm{np.tile}(\mathrm{np.linspace}\left(-1,1, \mathrm{im}_{\mathrm{depth}_{\mathrm{shape}}[1]}\right), (\mathrm{im}_{\mathrm{depth}_{\mathrm{shape}}[0]}, 1)$$

…(11)

The grids are normalized based on the values obtained from equation 12 and the depth surface is calculated as shown in equation 12.

$$\mathrm{Z}_{\mathrm{surface}} = \mathrm{Z\,f}, \mathrm{Z}_{\mathrm{surface}}[\mathrm{mask} == \mathrm{False}] \\ = \mathrm{np.na} \ldots\ldots (12)$$

Where NaN values are applied to mask the surface. Here, the threshold values are set to -10 and 10. The line filters the im_depth array by setting values to 0 where the depth values are less than -10 or greater than 10. The thresholding operation preferably removes small depth values that are within the specified range. Further the depth information and depth filtering.

4. Results and Discussion

The performance of the proposed approach is evaluated using a combined approach of DualUNet, PNCC, and depth filtering wherein the strengths of each technique is leveraged. A Matplotlib function is used to visualize the depth map in 3D wherein the input is an image and the surface is the depth map. This function uses shading to enhance the visualization, and the resulting 3D facial image is displayed as output as shown in figure 4. The PNCC visualized and depth filtered image are shown in figure 5.

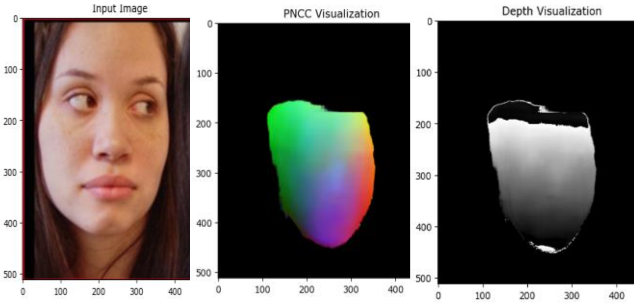


Fig.4. Enhanced PNCC and depth visualization techniques

The appearance of the image can be controlled by calibrating the parameters such as elevation, azimuth, and stride.

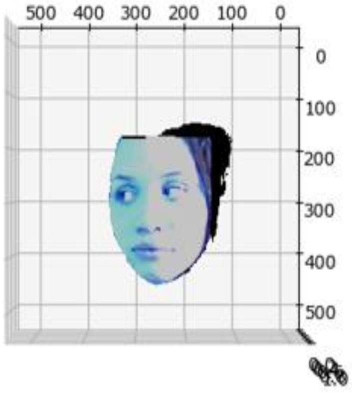


Fig.5. Resulting 3D facial image

The performance of the proposed approach is also quantitatively analysed with respect to different statistical parameters as shown in table 1, 2, and 3. The performance of the model is tested for the 3DMM dataset, and AFLW2000 - 3D dataset [39]. Figure 6 and figure 7 represent the qualitative analysis of the proposed work. Figure 7 depicts the sample set of images and the 3D reconstructed with texture details.

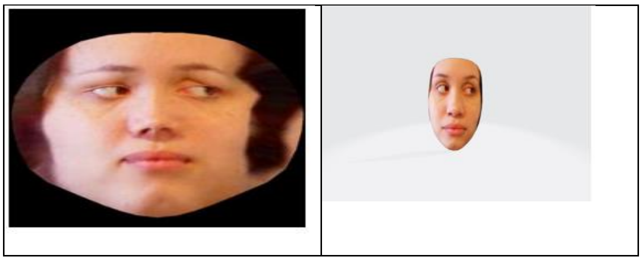


Fig 6: Representation of 3D image in 2D format

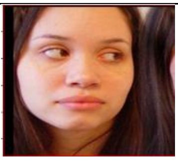
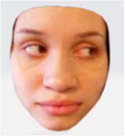
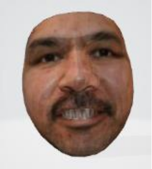
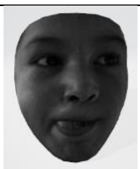
Sample Input Images	3D reconstruction with texture details
	
	
	

Fig 7: Input and corresponding 3D reconstruction with texture details

Table 1. Quantitative performance of the Dual U-Net model for the 3DMM and AFLW2000-3D dataset

Dataset	MAE	MSE	Depth Resolution	Z Score	Skewness
3DMM	3.4560	16.2555	0.3477	1.4810e-19	1.12633
AFLW2000-3D	3.4560	16.2371	0.1737	1.3511e-19	0.9538

In addition, the performance is compared in terms of the statistical parameters which are illustrated in table 1,2 and 3

In table 1 indicates that both datasets show relatively low MAE and MSE values, indicating good accuracy in depth prediction. The AFLW2000-3D dataset appears to have slightly better performance in terms of MAE and MSE compared to the 3DMM dataset. Depth resolution is higher for the 3DMM dataset, suggesting it provides more detailed depth predictions. Z scores are extremely low for both datasets, indicating excellent performance in matching the ground truth. Skewness values are close to 1 for both datasets, suggesting a slightly right-skewed distribution of errors, but overall, the distribution seems relatively symmetric.

In table 2 the proposed Dual UNET model outperforms all other methods listed in terms of RMSE, with the lowest value of 1.61. SynergyNet[29] has the second-best performance with an RMSE of 1.87. 3DDFA-V2 [28] and 2DASL [27] perform slightly worse with RMSE values of 2.04 and 2.05, respectively. PRNet [26] has the highest RMSE among the listed methods with a value of 2.25. The analysis suggests that the proposed Dual UNET model exhibits superior performance compared to the other methods listed in terms of RMSE.

Table 2. Quantitative performance of the Dual U-Net model for the 3DMM and AFLW2000-3D dataset

Methods	RMSE
PRNet [26]	2.25
2DASL [27]	2.05
3DDFA-V2 [28]	2.04
SynergyNet [29]	1.87
Proposed Dual UNET	1.61

Root Mean Squared Error (RMSE), a lower value indicates that the predictions of the model are nearer to the ground truth compared to models with higher RMSE values. In 3D face reconstruction, lower RMSE scores indicate better accuracy in reconstructing the three-dimensional structure of a face. Figure 8 depicts the RMSE for existing methods and proposed methods.

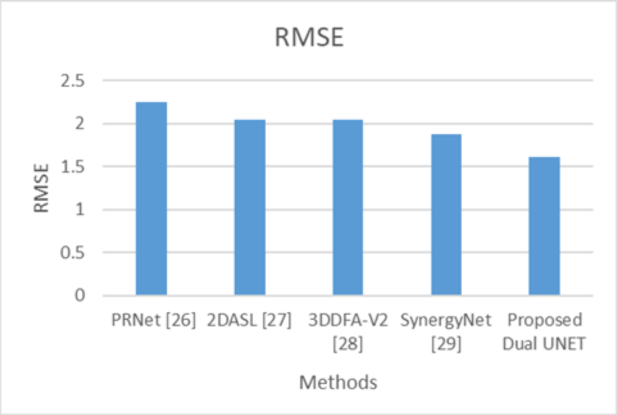


Fig 8: Comparative analysis with other existing works in terms of RMSE

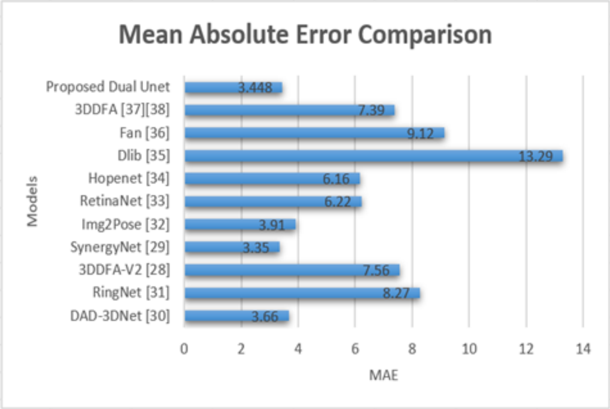


Fig.9. Comparison of the proposed approach in terms of MAE

The proposed Dual Unet model has the lowest MAE value (3.448), indicating better performance in 3D face reconstruction compared to the other listed models. A lower MAE suggests that the predicted 3D face shape aligns more closely with the ground truth. Figure 9 portrays the MAE for existing methods and proposed methods.

Table 3. Comparison based on Mean Absolute Error

Methods	MAE
DAD-3DNet [30]	3.66
RingNet [31]	8.27
3DDFA-V2 [28]	7.56
SynergyNet [29]	3.35
Img2Pose [32]	3.91
RetinaNet [33]	6.22
Hopenet [34]	6.16
Dlib [35]	13.29
Fan [36]	9.12
3DDFA [37][38]	7.39
Proposed Dual Unet	3.448

5. Conclusion

A novel approach combining a Dual U-Net architecture with PNCC and depth filtering technique is presented in this paper for reconstructing 3D facial images from a single input image. The quality of the images was enhanced using the PNCC technique and helped the U-Net architecture to understand the complex facial features. The depth filtering technique along with PNCC significantly improved the facial reconstruction process by reducing the skewness, error and improving the depth resolution. The proposed approach was quantitatively analyzed with respect to different parameters and results show that the combination of the Dual U-Net architecture with PNCC and depth filtering yielded a robust and accurate 3D facial reconstruction model which can effectively handle feature extraction, enhance perceptual features and refines the depth information which is crucial for obtaining a clear 3D face representation. The proposed model exhibits a minimum skewness of 0.9538 for AFLW2000-3D datasets.

5.1. Acknowledgment

None of the funding sources for this study were public, private, or nonprofit. We sincerely thank the professors of Christ University's department of computer science and engineering, Bangalore, India.

Conflict of Interest

The authors declare no conflict of interest

Reference

[1] Thies, J., Zollhöfer, M., Stamminger, M., Theobalt, C., & Nießner, M. (2016a). Facevr: Real-time facial reenactment and eye gaze control in virtual reality. arXiv preprint arXiv:1610.03151.

[2] Chen, S. Y., Gao, L., Lai, Y. K., Rosin, P. L., & Xia, S. (2018, March). Real-time 3d face reconstruction and gaze tracking for virtual reality. In 2018 IEEE Conference on virtual reality and 3d user interfaces (VR) (pp. 525-526). IEEE.

[3] You, Y., Song, L., & Yang, Y. (2021). High-Quality Facial Expression Animation Synthesis System Based on Virtual Reality. In VR/AR and 3D Displays: First International Conference, ICVRD 2020, Hangzhou, China, December 20, 2020, Revised Selected Papers 1 (pp. 21-32). Springer Singapore.

[4] Wang, L., & Zhang, R. (2023). Framework for facial recognition and reconstruction for enhanced security and surveillance monitoring using 3D computer vision. Journal of Electronic Imaging, 32(4), 042108-042108.

[5] Zollhöfer, M., Thies, J., Garrido, P., Bradley, D., Beeler, T., Pérez, P., ... & Theobalt, C. (2018, May). State of the art on monocular 3D face reconstruction, tracking, and applications. In Computer graphics forum (Vol. 37, No. 2, pp. 523-550).

[6] Tarassoli, S. P., Shield, M. E., Allen, R. S., Jessop, Z. M., Dobbs, T. D., & Whitaker, I. S. (2020). Facial reconstruction: a systematic review of current image acquisition and processing techniques. Frontiers in Surgery, 7, 537616.

[7] Ichim, A. E., Bouaziz, S., & Pauly, M. (2015). Dynamic 3D avatar creation from hand-held video input. ACM Transactions on Graphics (ToG), 34(4), 1-14.

[8] Thies, J., Zollhofer, M., Stamminger, M., Theobalt, C., & Nießner, M. (2016b). Face2face: Real-time face capture and reenactment of rgb videos. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2387-2395).

[9] Tran, L., Liu, F., & Liu, X. (2019). Towards high-fidelity nonlinear 3D face morphable model. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 1126-1135).

[10] Tewari, A., Seidel, H. P., Elgharib, M., & Theobalt, C. (2021). Learning complete 3d morphable face models from images and videos. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 3361-3371).

[11] Deng, Y., Yang, J., Xu, S., Chen, D., Jia, Y., & Tong, X. (2019). Accurate 3d face reconstruction with weakly-supervised learning: From single image to image set. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops (pp. 0-0).

[12] Jiang, L., Zhang, J., Deng, B., Li, H., & Liu, L. (2018). 3D face reconstruction with geometry details from a single image. IEEE Transactions on Image Processing, 27(10), 4756-4770.

[13] Tewari, A., Zollhoefer, M., Bernard, F., Garrido, P., Kim, H., Perez, P., & Theobalt, C. (2018). High-fidelity monocular face reconstruction based on an unsupervised model-based face autoencoder. IEEE transactions on pattern analysis and machine intelligence, 42(2), 357-370.

[14] Booth, J., Roussos, A., Ponniah, A., Dunaway, D., & Zafeiriou, S. (2018). Large scale 3D morphable models. International Journal of Computer Vision, 126(2), 233-254.

[15] Cao, C., Weng, Y., Zhou, S., Tong, Y., & Zhou, K. (2014). Facewarehouse: A 3d facial expression database for visual computing. IEEE Transactions on Visualization and Computer Graphics, 20(3), 413-425.

[16] Abada, L., & Aouat, S. (2016). Facial shape-from-shading using features detection method. International Journal of Advanced Intelligence Paradigms, 8(1), 3-19.

[17] Richardson, E., Sela, M., Or-El, R., & Kimmel, R. (2017). Learning detailed face reconstruction from a single image. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1259-1268).

[18] Rotger Moll, G., Moreno-Noguer, F., Lumbreras, F., & Agudo Martínez, A. (2019). Detailed 3D face reconstruction from a single RGB image. Journal of WSCG (Plzen, Print), 27(2), 103-112.

[19] Zhao, D., & Qi, Y. (2022, September). 3D Face Reconstruction with Geometry Details from a Single Color Image Under Occluded Scenes. In International Conference on Artificial Neural Networks (pp. 332-344). Cham: Springer Nature Switzerland.

[20] Yang, L., Wu, J., Huo, J., Lai, Y. K., & Gao, Y. (2021). Learning 3D face reconstruction from a single sketch. Graphical Models, 115, 101102.

[21] Tu, X., Zhao, J., Xie, M., Jiang, Z., Balamurugan, A., Luo, Y., ... & Feng, J. (2020). 3D face reconstruction from a single image assisted by 2D face images in the wild. IEEE Transactions on Multimedia, 23, 1160-1172.

[22] Dou, P., Shah, S. K., & Kakadiaris, I. A. (2017). End-to-end 3D face reconstruction with deep neural networks. In proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5908-5917).

[23] Sharma, S., & Kumar, V. (2020). Voxel-based 3D face reconstruction and its application to face recognition using sequential deep learning. Multimedia tools and applications, 79, 17303-17330.

[24] Chen, F. F., Guan, B., Kim, S., & Choi, J. (2023, August). 3-D Face Reconstruction Method Using Deep Learning Based Simulated Annealing. In International Conference on Intelligent and Fuzzy Systems (pp. 215-221). Cham: Springer Nature Switzerland.

[25] Sharma, S., & Kumar, V. (2022). 3D face reconstruction in deep learning era: A survey. Archives of Computational Methods in Engineering, 29(5), 3475-3507.

[26] Yao Feng, Fan Wu, Xiaohu Shao, Yanfeng Wang, and Xi Zhou. Joint 3d face reconstruction and dense alignment with position map regression network. In ECCV, pages 534–551,2018.

[27] Xiaoguang Tu, Jian Zhao, Mei Xie, Zihang Jiang, Akshaya Balamurugan, Yao Luo, Yang Zhao, Lingxiao He, Zheng Ma, and Jiashi Feng. 3d face reconstruction from a single image assisted by 2d face images in the wild. IEEE Trans_x0002_actions on Multimedia (TMM), 2020.

[28] Jianzhu Guo, Xiangyu Zhu, Yang Yang, Fan Yang, Zhen Lei,and Stan Z Li. Towards fast, accurate and stable 3d dense face alignment. In ECCV, 2020.

[29] Wu, Cho-Ying, Qiangeng Xu, and Ulrich Neumann. "Synergy between 3dmm and 3d landmarks for accurate 3d facial geometry." In 2021 International

Conference on 3D Vision (3DV), pp. 453-463. IEEE, 2021.

[30] Martyniuk, T., Kupyn, O., Kurlyak, Y., Krashenyi, I., Matas, J. and Sharmanska, V., 2022. Dad-3dheads: A large-scale dense, accurate and diverse dataset for 3d head alignment from a single image. In Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition (pp. 20942-20952).

[31] Soubhik Sanyal, Timo Bolkart, Haiwen Feng, and Michael J Black. Learning to regress 3d face shape and expression from an image without 3d supervision. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 7763–7772, 2019.

[32] Vitor Albiero, Xingyu Chen, Xi Yin, Guan Pang, and Tal Hassner. img2pose: Face alignment and detection via 6dof, face pose estimation. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 7617–7627, 2021.

[33] Jiankang Deng, Jia Guo, Evangelos Ververas, Irene Kotsia, and Stefanos Zafeiriou. Retinaface: Single-shot multi-level face localisation in the wild. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2020.

[34] Bardia Doosti, Shujon Naha, Majid Mirbagheri, and David J. Crandall. Hope-net: A graph-based model for hand-object pose estimation. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2020.

[35] Davis E. King. Dlib-ml: A machine learning toolkit. Journal of Machine Learning Research, 10:1755–1758, 2009.

[36] Feng-Ju Chang, Anh Tuan Tran, Tal Hassner, Iacopo Masi, Ram Nevatia, and Gerard Medioni. Faceposenet: Making a case for landmark-free face alignment. In International Conference on Computer Vision (ICCV) Workshops, Oct 2017.

[37] Jianzhu Guo, Xiangyu Zhu, and Zhen Lei. 3ddfa. https: //github.com/cleardusk/3DDFA, 2018.

[38] Xiangyu Zhu, Xiaoming Liu, Zhen Lei, and Stan Z Li. Face alignment in full pose range: A 3d total solution. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(1):78–92, 2017.

[39] Xiangyu Zhu, Zhen Lei, Xiaoming Liu, Hailin Shi, and Stan Z Li. Face alignment across large poses: A 3d solution. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 146–155, 2016.