Anand Gupta, 2024, 12:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology

An Open Access Journal

Shape-Guided Conditional Image Synthesis for Tooth Alignment

Anand Gupta[1], Deep Mehta[2], Arpan Dholakiya[3], Soorajkumar Suresh[4], Siddhant Jadhav[5], Amit Yadav[6], Yaduvir Singh[7]

Toothlens Healthcare Private limited, Delaware, United States[1,2,3,4,5]
7Niet, Greater Noida, Uttar Pradesh, India[1,6,7]

Abstract- Orthodontics is a field of dentistry that focuses on correcting misaligned teeth. Orthodontic treatment often involves the use of braces, aligners, retainers, and other dental appliances to align teeth, correct bite issues, and improve overall dental health and aesthetics. But fixing them can be a long and complicated process. So, it's important to have a picture showing what your teeth will look like after treatment. This helps dentists explain things better and makes patients more likely to go through with the treatment. In this paper, we present an approach for generating the conditional image of a misaligned tooth to an aligned tooth from a diffusion model it takes 2D images as input (e.g., captured by the camera, or smartphones), and aligns the teeth part in the image within 2D image space to generate post-treatment like images, our method employs 3D STL data scanned by intra-oral scanner projecting pre- and post-treatment, diffusion model to learn the spatial movement of each tooth followed by post-processing to convert the teeth area more realistic textures. We validate our pipeline over various facial photographs, illustrating its outstanding effectiveness and broad utility within the field of Orthodontics.

Keywords- Orthodontic Treatment, Pre- and Post-treatment Images, DDPM (Denoising Diffusion Probabilistic Models), Facial Recognition, Image Segmentation, smile design, Contour Segmentation, Edge detection, Conditional Diffusion Model, Tooth Alignment, Unet, Resnet18, Facial landmarks, Neural Network, Deep Learning, Image Synthesis, Shape-guided image synthesis, Loss Function, Training Pipeline, Inference, MaskGIT(Masked Generative Image Transformer), Data Pipeline, Image Generation, Visualization, Multilayer Perceptron(MLP)

I. INTRODUCTION

In recent years, there have been significant advances in Generative AI, particularly in the field of Image Generation, which holds promising implications for medical images. These advancements have led to the development of sophisticated models capable of generating realistic and high-fidelity images, thereby opening up new avenues for enhancing the quality and diversity of medical images. As smile simulation gains increasing attention [3,4]. particularly in the realm of frontal face aesthetics, its potential to attract individuals to undergo orthodontic treatment becomes apparent. Patients opting for invisible teeth alignment procedures naturally desire insight into how their frontal facial appearance may evolve across various treatment stages. Orthodontic treatment offers an effective solution for correcting tooth misalignment and improving overall dental

alignment and aesthetics. while orthodontic intervention holds the potential to address malocclusion problems that affect over 90% of individuals to varying degrees, the benefits extend beyond mere physiological improvements. By enhancing oral health and function, orthodontic treatment plays a pivotal role in bolstering patients' confidence and overall psychological well-being. Nevertheless, the intricate and time-consuming nature of orthodontic procedures spanning months or even years may act as a deterrent for individuals considering treatments. Therefore, the creation and presentation of anticipated post-treatment facial images featuring aesthetically pleasing teeth assume paramount importance. These predictive visuals serve not only to captivate and inspire patients but also to facilitate more meaningful interactions between orthodontists and their patients.

While orthodontists may not be able to manually generate 2D teeth images based solely on their professional expertise and inform patients that these images represent the expected shapes of their teeth under the current treatment schedule, they can still offer reliable information. This includes providing 3D teeth models and their planning prescriptions, detailing how each tooth will move at every orthodontic stage. Therefore, the primary objective of our system is to reconstruct the teeth region of the facial images

Figure 1: we're showcasing before-and-after images of people's smiles. On the left we show the misaligned teeth and on the right post-treatment images of aligned teeth

In Traditional Clinical procedure orthodontists utilize a technique known as "Visual Treatment Objective" (VTO) to visualize patients' appearance after orthodontic treatment, this method typically involves manipulating X-ray images to adjust soft tissues and skeleton based on specific landmarks [21,25]

In the realm of computer vision, recent advancements in deep learning, particularly in generative networks, have shown remarkable progress. However, many of these models rely heavily on paired images, which poses a challenge for our specific task. Collecting paired pre- and post-orthodontic facial photographs is arduous due to the lengthy orthodontic process and the evolving facial appearance over time. Moreover, 2D photographs lack the depth information necessary to represent the structure of teeth accurately. Consequently, bridging the gap between clinical knowledge of tooth alignment, which was defined on 3D tooth models, and 2D photographs presents a significant challenge [1]. In this research, our goal is to take photographs and generate aligned teeth and brighter teeth with realistic textures, as shown in Figure 1. Each generated photograph followed the unique tooth alignment property.

In this paper, we propose a shape-guided conditional image synthesis for tooth alignment in photographs. In essence, our main objective is to acquire the clinical understanding of tooth alignment as delineated in the 3D intra-oral scanning models [17], we followed the learned property to guide the post-orthodontic image generation. we use a pre-trained model which is trained by collecting a set of paired pre- and post-orthodontic intra-oral scanning tooth models and applying them to the oral region of a 2D facial image. [1][22]. Post a Diffusion Model [11] approach is utilized to grasp information regarding tooth alignment. for generating tooth contours of the post-treatment image, threshold for Canny Edge (50,100) (as shown in Figure 2). Another diffusion model is integrated to generate realistic photographs with aligned teeth and incorporate teeth-lightening information. In this experiment, a substantial quantity of photographs is gathered from patients exhibiting varying degrees of

malocclusion issues. our method outperforms state-of-the-art techniques, including GANs [8].

II. LITERATURE REVIEW

Table 1: Summary of work done on generation of teeth alignment

Author Name	Feature Extraction	Result and finding	Remark
Yulong Dou1, Lanzhuju Mei1, Dinggang Shen, Zhiming Cui	Faical landmarks, mouth detection, oral segmentation,tooth contours, skin color	State-of-the-art result for generating post treatment images	Pioneering approach in computer vision and conditional diffusion model to convert pre-treatment image to post-treatment image
Richard P McLaughlin and John C Bennett[21]	Analyze orthodontic tooth movement, possibly introducing a new method called "dental VTO"	Dental VTO's accuracy in predicting tooth movements and its correlation with actual treatment outcomes	Practical Implications of the Dental VTO's and its potential impact on clinical decision
Shen Feihong, Liu jingjing, Li Haizhen[2]	The system utilizes image-based facial editing techniques to simulate orthodontic treatment effects accurately	Effectiveness of the proposed approach in predicting treatment outcomes, with OrthoGAN outperforming other models in terms of simulation authenticity	Deep learning-based smile simulation approach for virtual teeth alignment effects in frontal face images with high precision generate realistic output
Lingchen Yang, Zefeng Shi, Yiqian Wu, Xiang Li, Kun Zhou, Hongbo Fu, and Youyi Zheng [3]	Teeth geometry, enabling fine-scale geometry control and lightening condition.	iOrthoPredictor accurately predicts teeth alignment, demonstrating practical applicability in orthodontics through validation and user studies	iOrthoPredictor offers a significant advancement in predicting teeth alignment from TSynNet, TAligNet,TGeoNet
Beijia Chen, Hongbo Fu, Kun Zhou, and Youyi Zheng[4]	Teeth region of portrait photographs and embeds them into the latent space of StyleGAN, facilitating,geometrically	Demonstrates alignment prediction across diverse orthodontic cases and short video clips	StyleGAN and BlendingNet enhance alignment prediction and blending for versatile application.
Zhiming Cui, Changjian Li, and Wenping Wang [5]	Utilizing a two-stage network: edge extraction and 3D region proposal.	Represented 3D teeth models reconstructed from CBCT images, along with augmenting the base network	Tooth segmentation and identification from CBCT images
Zhiming Cui, Yu Fang, Lanzhuju Mei, Bojun Zhang, Bo Yu, Jiameng Liu, Caiwen Jiang, Yuhang Sun, Lei Ma, Jiawei Huang, et al[5]	ROI Extraction, Tooth ROI, Tooth centered, Tooth Skeleton, Tooth classification, Multi-task tooth segmentation	Using deep learning achieve high accuracy fully automatic tooth and alveolar bone segmentation	Advances in tooth and bone segmentation
W De Vos, Jan Casselman,Swennen [7]	Reviewed literature on CBCT imaging techniques	Identified clinical applications and highlighted the lack of evidence-based data on radiation dose.	Suggested guidelines for CBCT device parameters.
Ian Goodfellow, Jean Pouget-	The generative model	GAN training improves	Compatible with various

Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio[8]	captures data distribution and the discriminative model estimates the probability of training data	the model and discriminator (The adversarial process refines both models)	image generation frameworks for generative model
Andreas Wirtz, Florian Jung, Matthias Noll, Anqi Wang, and Stefan Wesarg[9]	Photo-specific 2D coupled shape models the model extracts tooth contours from each image	Promising 3D tooth reconstruction with 0.848mm average symmetric surface distance, the method achieves accurate reconstructions	Automatic 3D reconstruction from orthodontic photographs is feasible,This approach offers the potential for automating orthodontic analysis.
Chenglei Wu, Derek Bradley, Pablo Garrido, Michael Zollhöfer, Christian Theobalt, Markus H Gross, and Thabo Beele[10]	Parametric tooth row prior, the model uses a pre-learned database of tooth shapes to guide reconstruction	Non-invasive reconstruction of entire tooth rows from mouth photos, the method creates realistic tooth models from simple images	The first approach for photo-based tooth row reconstruction, applicable to various image sources, this innovative method offers a less invasive alternative to dental scanners for creating tooth models
Jonathan Ho, Ajay Jain, and Pieter Abbeel[11]	Appearance, geometry, high-quality image synthesis	State-of-the-art image synthesis on CIFAR10 and LSUN datasets, the model achieves high-quality image generation	Promising diffusion models with progressive decompression scheme
YANG Lingchen, SHI Zefeng, WU Yiqian, LI Xiang, ZHOU Kun, FU Hongbo, and Youyi Zhen[12]	Frontal face image, misalignment teeth extraction with teeth geometry	iOrthoPredictor accurately predicts teeth alignment in photos, Disentanglement of teeth geometry and in-mouth appearance is key	The network has the potential for patient consultations and treatment planning from the post-treatment image generation
Dor Bank, Noam Koenigstein, Raja Giryes[13]	Learn from edges, shapes, textures, colors, spatial relationships, anomalies	Effectively learn compressed and meaningful data representations	Challenges remain in setting hidden state size and distribution for generative models
Diederik P. Kingma, Max Welling[14]	Extract latent features from images	Efficient and scalable gradient-based variational posterior inference	VAE remains a widely used tool for probabilistic modeling
Guillaume Alain, Yoshua Bengio, Li Yao, Jason Yosinski, Eric Thibodeau-Laufer, Saizheng Zhang, and Pascal Vincent[15]	GSN captures underlying data structures, facilitating tasks like denoising and generative modeling	GSN simplifies generative model training, capturing data distribution via Markov chain. Training sufficient-capacity models to denoise and reconstruct suffices for distribution capture	GSN extends beyond auto-encoders, aids classification, and offers fast sampling in deep models

Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi[16]	Utilizes conditional diffusion models for image-to-image translation, surpassing baselines without task-specific tuning	Unified framework based on conditional diffusion models outperforms GAN and regression baselines on image-to-image translation tasks	Promotes standardized evaluation in image-to-image translation; generalist diffusion model matches specialist models' performance.
Hidemichi Kihara, Wataru Hatakeyama, Futoshi Komine, Kyoko Takafuji, Toshiyuki Takahashi, Jun Yokota, Kenta Oriso, and Hisatomo Kondo[17]	Evaluation of accuracy and practicality of intraoral scanners and verification methods.	Intraoral scanners' accuracy is influenced by lighting. Repeatable for partial prostheses. Challenges for full-arch prostheses.	Intraoral scanner accuracy depends on the lighting. Good for partial edentulous cases..
Vahid Kazemi and Josephine Sullivan[18]	Ensemble of regression trees estimates landmark positions from sparse pixel intensities for face alignment	Achieved super-realtime performance with high-quality predictions, handling missing or partially labeled data.	Utilizing gradient boosting framework with appropriate priors for efficient feature selection and combating overfitting.
Changqian Yu, Jingbo Wang, Chao Peng, Changxin Gao, Gang Yu, and Nong Sang[20]	Bilateral Segmentation Network balances spatial information and receptive field for semantic segmentation.	BiSeNet achieves 68.4% Mean IOU on Cityscapes with 105 FPS, balancing speed and performance	BiSeNet offers a promising solution to the trade-off between speed and segmentation accuracy
Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros[29]	Conditional adversarial networks learn mapping and loss functions, versatile for image-to-image translation tasks	Effective synthesis of photos from label maps, object reconstruction from edge maps, and image colorization	Shift from hand-engineered mapping and loss functions to automated learning in image translation
Vedant Singh, Surgan Jandial, Ayush Chopra, Siddharth Ramesh, Balaji Krishnamurthy, and Vineeth N Balasubramanian[37]	Conditional diffusion models for image editing, stock photo generation, and 3-D object creation.	Conditional noise input enhances image control, diverging from traditional Gaussian noise approaches, input improves semantic attribute conditioning in diffusion models for image generation	Promising approach enhances image control and conditioning in diffusion models for diverse image generation.
Olaf Ronneberger, Philipp Fischer, and Thomas Brox[51]	Network combines contracting and expanding paths for context and localization	Outperforms sliding-window network on ISBI challenge for neuronal structure segmentation	Fast segmentation, trained from few images, wins ISBI cell tracking challenge
Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Muller, Harry Saini Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, Dustin Podell, Tim Dockhorn Zion English, Kyle Lacey, Alex Goodwin, Yannik Marek, Robin Rombach[69]	Improved noise sampling techniques bias towards perceptually relevant scales in rectified flow models	Superior performance in high-resolution text-to-image synthesis compared to established diffusion formulations	Transformer-based architecture enables bidirectional flow between image and text tokens, improving synthesis and comprehension

In recent years, there has been a strong push towards digitizing the dental field. Intraoral scanners, which create impressions digitally, are seen as transformative for future dental treatments [17]. CBCT [7] provides dentists about information, the structure of patient's teeth. Various novel methods have emerged in associated domains, such as tooth segmentation, 3D tooth reconstruction [9],[10] and 3D tooth arrangement [35]. Lingchen et al [12] present a novel approach to visually predict teeth alignment in photographs along with a corresponding 3D teeth model as input and generate using the concept of StyleGAN. Feihong et al. [2] conclude their study with a high-precision visualization for orthodontic smile simulation, the system utilizes advanced image-based facial editing techniques to provide highly realistic simulation results. Experiments and user studies confirm its effectiveness in predicting treatment outcomes in digital orthodontics.

Image generation in computer vision involves creating new digital images using algorithms or models, such as GANs [8]. Bank et al. [13] Autoencoder design to compress input data and then reconstruct it to resemble the original input. VAE [14], Denoising Diffusion Probabilistic Models [11] Saharia et al. [16] present image-to-image translation based on conditional diffusion models.

III. METHODOLOGY

1. Overview

In this research paper, we embark on a comprehensive exploration of Teeth alignment reconstruction. Our goal is to develop a tooth alignment multilayer network that understands the spatial movement of each instance of teeth that guides the generation of post-treatment photographs.

We have pre and post-intra-oral scanning pre-trained models S = {S1,S2,...,SN} of the same patients. And facial photographs I = {I1,I2,...,IM} captured by smartphones. We used a module that is design for tooth alignment which seamlessly integrates 3D structural insights from intra-oral scanning models. This module dynamically selects pre- and post-orthodontic intra-oral scanning models (i.e., Sr ∈ S and Sˆ r ∈ Sˆ) for a given unpaired facial photograph (Ir ∈I). It then executes a coarse 2D-3D registration. Then the 3D structure is projected onto the 2D facial photograph to obtain the contours of pre- and post-orthodontic Cr ∈ R 128×256.

Post we used a pre-trained segmentation module to locate the mouth region and segment the binary mask and tooth contours from frontal face photographs. In addition, a generation model used which is from the conditional self-attention diffusion model along with Unet architecture to the realistic images After the initial processing, we apply a post-processing technique using an instance segmentation model to focus on specific instances and extract individual tooth masks from the segmented masks. Subsequently, we determine the shade of each tooth by identifying individual teeth and computing the mean saturation of the masked regions. We then adjust the H (Hue) and S (Saturation) channels to shift the tooth shade grade accordingly. The threshold list used for this adjustment includes values of 20, 40, 50, and 65.

Yulong Dou1 et al.

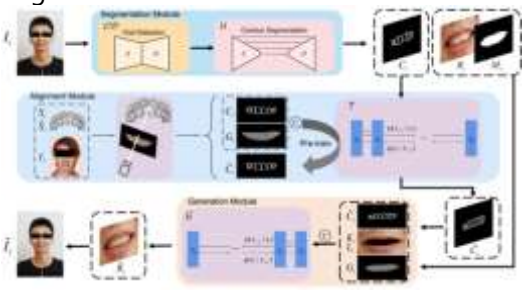

Figure 2: Flow of pipeline. When a 2D photograph is input into networks. It goes to the mouth detection module to obtain the mouth mask inside the lip area, and tooth contours. Then it passes through Align-Mod to predict we aligned tooth contours Lastly, the processed image undergoes the Generation Module (Gen-Mod), which generates a facial photograph displaying well-aligned teeth[1].

2. Image Segmentation

Detect face alignment (i.e position of the face) [18][19] and get the face size of Fi ∈ R 512×512

from input facial image Ii ∈ I ⊆ R. and locate the inside lip area from mouth segmentation model trained on ResNet18 backbone OD(.), YOLO v8. we propose a mouth detection model [20][23], segment the mask, and crop inside the lip area mask Ri ∈ R 128×256, for segmentation of the tooth we employ U-Net[51] Attention Refinement Module(ARMs), calculation of segmentation is as

$$Ri, Mi = OD(Ii),$$

$$Ci = U(Ri, Mi), \forall i = 1, 2,...,N, \quad (1)$$

Where OD(.) represents mouth detection in photographs Ii , U(•) represent U-Net network. we obtain the pre-orthodontic tooth contours (Ci), mouth region (Ri), and oral mask (Mi) using the mouth segmentation model.

We used pre-trained network training utilizing a combination of Dice Loss [22] and Weighted Cross-Entropy Loss [24] to handle class imbalance effectively. Dice Loss emphasizes foreground learning, ideal for imbalanced scenarios, while Weighted Cross-Entropy Loss adjusts weighted proportions to address imbalance concerns. The loss function:

$$L = w_{dice} \cdot L_{dice} + w_{ce} \cdot L_{ce}. \quad (2)$$

3. Orthodontic Alignment

We refer to the approach that innovatively integrates 3D intra-oral scan data into the Align-Mod model [1], crucial for orthodontic treatment planning. Unlike prior methods focused on reconstructing 3D dental models from tooth photographs, we project intra-oral scans onto 2D facial images using a 3D-to-2D rendering approach. Additionally, we've devised a conditional DDPM Model-based network to glean orthodontic insights from rendered tooth contours (Cr) achieved by applying edge detection that detects strong gradients with the threshold of 50-100 [9][27].

We apply coarse registration for establishing spatial correspondences between photographs based on the tooth landmarks [28]. Therefore, the standard contours Cr, Cˆ r ∈ R 128×256

$$\rho m = (KR^T \mid -KR^TC)\,(M\ 1), \quad (3)$$

where ρ is projection, K denotes the matrix of internal parameters, R denotes the rotation matrix indicating the camera's orientation, C is the camera spatial location M denotes the coordinates of the point in the world coordinate system, symbolized as M = (X,Y,Z)^ T and m denotes the coordinate of corresponding point in pixel coordinate system as m=(u, v,1)^T.

4. Image-to-Image Generation

We refer image to image DDPM model [16] for the generation of the mouth regions with Gaussian noise Gr ∈ R^ 128×256 and we fine-tune the parameter beta schedule the number of time steps is 150 in mask-to-mask generation, contour-to-contour generation combined with conditional diffusion model. During training post orthodontic tooth contours C^r = T (Cr c Gr), where c represents concatenation(channel-wise), T represents learn the clinical knowledge.

Contours derived from the Image segmentation model (Ci)with Gaussian noise (Gi) applied into a diffusion model anticipating a dependable prediction for straight well-aligned teeth contours (C^i).

$$Cˆ i = T(Ci\ c\ Gi), \forall i = 1,2,...,N.$$

To generate a complete mouth region, we concatenate conditional information skin color Ki ∈ R 128×256, straight tooth contour, and Gaussian noise Gi ∈ R 128×256 to feed in our generation model.

5. Enhance Teeth

We proposed a teeth color enhancement network to focus on the teeth area. trained instance segmentation (Is) model, segmentation results are post-processed to obtain masks for individual teeth. Cropped mask from generated image and calculated the mean saturation of the masked image to identify tooth shade. The formula for mean saturation is:

$$M^s = \Sigma(Y^s)/\ N \quad (4)$$

Where M^s represent mean saturation, Y^s is yellow saturation filtered out zero values from the saturation channel of the HSV representation of the input image, and N is the number of filtered saturation values.

We then adjusted the tooth shade based on a threshold. modify the Hue (H) and Saturation (S) channels of the HSV image to shift the tooth shade grade. If the pixel was part of the tooth (determined by the mask), the saturation was updated according to the required shade. Additionally, all non-yellow hues were converted to yellow hues to handle edge cases. The modified HSV image was then converted back to the BGR color space to obtain the final result.

IV. EXPERIMENTS

we use a pre-trained model for facial segmentation to the crop mouth area, for mouth segmentation model and generation model [1], and trained mouth detection and semantic segmentation model for segmenting teeth used in post-processing that generate more realistic textures. we set the batch size of 60.

V. RESULT

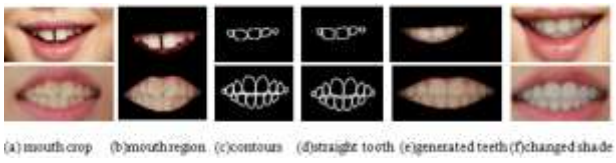

Figure3: Aligned teeth generation process, crop the mouth(a), segment mouth region(b), tooth edges(c), generated tooth edges (d), generated mouth region(e), adjusted shade of teeth

We were able to generate the post-treatment image from 2d image captured by smartphones from our experiment with conditional image-to-image generation method using a diffusion model. The results obtained from our experiment are not just promising, demonstrating the potential of our approach. This success paves the way for further advancements in this field.\

1. Comparison

We conducted a comparative analysis of our conditional image synthesis tooth alignment model with Pix2Pix GAN [8][29], StyleGAN [3], and 3D Structure-guided Network [1]. The results revealed that our conditional diffusion post-processing network produces more realistic outputs compared to Pix2Pix, StyleGAN, and the 3D Structure-guided Network.

2. Discussion

In this experiment, we propose a shape-guided conditional image synthesis pipeline to generate post-treatment images from pre-treatment 2d photographs. As per experimental results give clinical significance. Additionally, we introduce pre-processing and post-processing approach to feed in Diffusion Model in task of mouth region image generation.

We trained our own instance segmentation model for individual teeth detection to detect and classify individual teeth for generating more shiny images. Our approach preserves the potential of orthodontic knowledge and ensures ease of use in practice within the orthodontic industry.

The limitation of pre-trained models is evident in our approach, as it struggles with severely misaligned cases and contour creation. While it successfully resolves merging issues in (collision and occlusion) in some cases, it remains unresolved for others. However, our method effectively addresses the generation of realistic shiny teeth colors and ensures that gums are generated with natural skin tones.

There are some guidelines for input images due to our use of the dlib shape_predictor_68_face_landmarks detection pre-trained model, it will be unable to generate if the frontal face is missing and unable to generate from side views image. Additionally, if two or more consecutive teeth are missing, the model may not be able to generate these individual teeth accurately. These limitations highlight areas for future research and improvement

VI. FUTURE SCOPE

The future scope for the Shape-Guided Conditional Image Synthesis using the diffusion model is incredibly promising in the healthcare industry and spans various domains. Here are several domains where Shape-Guided Conditional Image Synthesis using a diffusion model is expected to have a significant impact.

Post-Treatment Reconstruction: Skin Diseases and Tumor Residuals: Shape-Guided Conditional Synthesis image generation has potential to reconstruct the pre-treatment image of Skin Diseases and Tumor Residuals.

Cosmetic Surgery: This approach could be used to simulate the result of cosmetic procedures that can help patients make informed decisions about their treatment.

Virtual Reality: Image synthesis generation can contribute to developing realistic virtual environment for enlarging the shape of objects.

Fashion Design: This can be used for simulating the mask of the clothes using shape and size and design. It could generate realistic images of how different clothing items or masks would fit.

VII. CONCLUSION

In conclusion, this project has presented comprehensive research and implementation of Shape Guided Image-Image to generation using conditional diffusion models. We started by exploring the fundamental concepts of smile design principles, and teeth alignment patterns through deep learning and moved to the generation from diffusion models. The process encompassed from data collection, preprocessing of data, facial landmarks detection, mouth detection, mouth segmentation, teeth instance segmentation, image-to-image generation, and postprocessing. our approach stands out from existing methods this ensures the reliability of our approach and its potential applicability in the healthcare sector.

Acknowledgements

This study was supported by the Noida Institute of Engineering and Technology (Department of Artificial Intelligence) and Toothlens Healthcare Private Limited.

REFERENCES

1. Yulong Dou1, Lanzhuju Mei1, Dinggang Shen, Zhiming Cui: 3D Structure-guided Network for Tooth Alignment in 2D Photograph,15(2), 2023
2. Shen Feihong, Liu jingjing, Li Haizhen: ORTHOGAN: HIGH-PRECISION IMAGE GENERATION FOR TEETH ORTHODONTIC VISUALIZATION,15(10),2022
3. Lingchen Yang, Zefeng Shi, Yiqian Wu, Xiang Li, Kun Zhou, Hongbo Fu, and Youyi Zheng. iorthopredictor: model-guided deep prediction of teeth alignment. ACM Transactions on Graphics (TOG), 39(6):1–15, 2020.
4. Beijia Chen, Hongbo Fu, Kun Zhou, and Youyi Zheng. Orthoaligner: Image-based teeth alignment prediction via latent style manipulation. IEEE Transactions on Visualization and Computer Graphics, 2022.
5. Zhiming Cui, Changjian Li, and Wenping Wang. Toothnet: automatic tooth instance segmentation and identification from cone beam ct images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 6368– 6377, 2019
6. Zhiming Cui, Yu Fang, Lanzhuju Mei, Bojun Zhang, Bo Yu, Jiameng Liu, Caiwen Jiang, Yuhang Sun, Lei Ma, Jiawei Huang, et al. A fully automatic ai system for tooth and alveolar bone segmentation from cone-beam ct images. Nature communications, 13(1):2096, 2022.
7. W De Vos, Jan Casselman, and GRJ19464146 Swennen. Cone-beam computerized tomography (cbct) imaging of the oral and maxillofacial region: a systematic review of the literature. International journal of oral and maxillofacial surgery, 38(6):609–625, 2009.
8. Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks.

Communications of the ACM, 63(11):139–144, 2020.

9. Andreas Wirtz, Florian Jung, Matthias Noll, Anqi Wang, and Stefan Wesarg. Automatic model-based 3-d reconstruction of the teeth from five photographs with predefined viewing directions. In Medical Imaging 2021: Image Processing, volume 11596, pages 198–212. SPIE, 2021.
10. Chenglei Wu, Derek Bradley, Pablo Garrido, Michael Zollhöfer, Christian Theobalt, Markus H Gross, and Thabo Beeler. Model-based teeth reconstruction. ACM Trans. Graph., 35(6):220–1, 2016.
11. Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems, 33:6840–6851, 2020.
12. YANG Lingchen, SHI Zefeng, WU Yiqian, LI Xiang, ZHOU Kun, FU Hongbo, and Youyi Zheng. iorthopredictor: model-guided deep prediction of teeth alignment. ACM Transactions on Graphics, 39(6):216, 2020.
13. Dor Bank, Noam Koenigstein, Raja Giryes: Autoencoders,2021.
14. Diederik P. Kingma, Max Welling: An Introduction to Variational Autoencoders 2019.
15. Guillaume Alain, Yoshua Bengio, Li Yao, Jason Yosinski, Eric Thibodeau-Laufer, Saizheng Zhang, and Pascal Vincent. GSNs: generative stochastic networks. Information and Inference: A Journal of the IMA, 5(2):210–249, 2016.
16. Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In ACM SIGGRAPH 2022 Conference Proceedings, pages 1–10, 2022.
17. Hidemichi Kihara, Wataru Hatakeyama, Futoshi Komine, Kyoko Takafuji, Toshiyuki Takahashi, Jun Yokota, Kenta Oriso, and Hisatomo Kondo. Accuracy and practicality of intraoral scanner in dentistry: A literature review. Journal of prosthodontic research, 64(2):109–113, 2020.
18. Vahid Kazemi and Josephine Sullivan. One millisecond face alignment with an ensemble of regression trees. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1867–1874, 2014.
19. Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 815–823, 2015.
20. Changqian Yu, Jingbo Wang, Chao Peng, Changxin Gao, Gang Yu, and Nong Sang. Bisenet: Bilateral segmentation network for real-time semantic segmentation. In Proceedings of the European conference on computer vision (ECCV), pages 325–341, 2018.
21. Richard P McLaughlin and John C Bennett. The dental vto: an analysis of orthodontic tooth movement. Journal of Clinical Orthodontics: JCO, 33(7):394–403, 1999.
22. Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In 2016 fourth international conference on 3D vision (3DV), pages 565–571. Ieee, 2016.
23. Cheng-Han Lee, Ziwei Liu, Lingyun Wu, and Ping Luo. Maskgan: Towards diverse and interactive facial image manipulation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 5549–5558, 2020.
24. Reuven Rubinstein. The cross-entropy method for combinatorial and continuous optimization. Methodology and computing in applied probability, 1:127–190, 1999.
25. G Power, J Breckon, M Sherriff, and F McDonald. Dolphin imaging software: an analysis of the accuracy of cephalometric digitization and orthognathic prediction. International journal of oral and maxillofacial surgery, 34(6):619–626, 2005.
26. Pascal Vincent. A connection between score matching and denoising autoencoders. Neural computation, 23(7):1661–1674, 2011
27. TX Zheng, Shuai Huang, YF Li, and MC Feng. Key techniques for vision based 3d reconstruction: a review. Zidonghua Xuebao, 46:631–652, 2020.
28. Bo D Tapley and JM Lewallen. Comparison of several numerical optimization methods.

Journal of Optimization Theory and Applications, 1:1–32, 1967.
29. Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1125–1134, 2017.
30. Amit Yadav, Anand Gupta, Ms. Aarushi Thusu: Recognition of Sentiment using Deep Neural Network. in International Journal of Trend in Scientific Research and Development (ijtsrd), 2456-6470, 2023.
31. Phillip Roe, Kitichai Rungcharassaeng, Joseph Y.K. Kan, Rishi D.Patel, Wayne V. Campagni, James S. Brudvik: The Influence of Upper Lip Length and Lip Mobility on Maxillary Incisal Exposure, 2012
32. Mohan Bhuvaneswaran: Principles of smile design, 0.4103/0972-0707.73387,2010
33. Karel Zuiderveld. Contrast limited adaptive histogram equalization. Graphics gems, pages 474–485, 1994
34. Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 586–595, 2018
35. Guodong Wei, Zhiming Cui, Yumeng Liu, Nenglun Chen, Runnan Chen, Guiqing Li, and Wenping Wang. Tanet: towards fully automatic tooth arrangement. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XV 16, pages 481–497. Springer, 2020.
36. Manoj Kumar Sundar and BDS Venkataraman Chelliah. Ten steps to create virtual smile design templates with adobe photoshop cs6. Compendium, 39(3), 2018.
37. Vedant Singh, Surgan Jandial, Ayush Chopra, Siddharth Ramesh, Balaji Krishnamurthy, and Vineeth N Balasubramanian. On conditioning the input noise for controlled image generation with diffusion models. arXiv preprint arXiv:2205.03859, 2022.
38. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and ComputerAssisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, pages 234–241. Springer, 2015.
39. Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14, pages 694–711. Springer, 2016.
40. Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, Sungroh Yoon: ILVR: Conditioning Method for Denoising Diffusion Probabilistic Models, 2021
41. Rameen Abdal, Yipeng Qin, and Peter Wonka. Image2stylegan: How to embed images into the stylegan latent space? In ICCV, 2019.
42. Rameen Abdal, Yipeng Qin, and Peter Wonka. Image2stylegan++: How to edit the embedded images? In CVPR, 2020.
43. Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale gan training for high fidelity natural image synthesis. arXiv preprint arXiv:1809.11096, 2018.
44. Jooyoung Choi, Jungbeom Lee, Yonghyun Jeong, and Sungroh Yoon. Toward spatially unbiased generative models. arXiv preprint arXiv:2108.01285, 2021.
45. Ming-Yu Liu, Xun Huang, Arun Mallya, Tero Karras, Timo Aila, Jaakko Lehtinen, and Jan Kautz. Few-shot unsupervised image-to-image translation. In ICCV, 2019
46. Guilin Liu, Fitsum A Reda, Kevin J Shih, Ting-Chun Wang, Andrew Tao, and Bryan Catanzaro. Image inpainting for irregular holes using partial convolutions. In ECCV, 2018.
47. Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In CVPR, 2020.
48. Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In CVPR, 2019.
49. Sachit Menon, Alexandru Damian, Shijia Hu, Nikhil Ravi, and Cynthia Rudin. Pulse: Self-

supervised photo upsampling via latent space exploration of generative models. In CVPR, 2020.
50. Taesung Park, Alexei A Efros, Richard Zhang, and JunYan Zhu. Contrastive learning for unpaired image-to-image translation. In ECCV, 2020.
51. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. Unet: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention, 2015.
52. Yujun Shen, Jinjin Gu, Xiaoou Tang, and Bolei Zhou. Interpreting the latent space of gans for semantic face editing. In CVPR, 2020.
53. Lynton Ardizzone, Carsten Lüth, Jakob Kruse, Carsten Rother, and Ullrich Köthe. 2019. Guided Image Generation with Conditional Invertible Neural Networks. In arXiv:1907.02392.
54. Connelly Barnes, Eli Shechtman, Adam Finkelstein, and Dan B Goldman. 2009. PatchMatch: A Randomized Correspondence Algorithm for Structural Image Editing. ACM Transactions on Graphics (Proc. SIGGRAPH) 28, 2009.
55. Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. 2017. Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587 (2017).
56. Philipp Kopp, Derek Bradley, Thabo Beeler, Markus Gross: Analysis and Improvement of Facial Landmark Detection, 10.13140/RG.2.2.10980.42886, 2019.
57. J. Yang, Q. Liu, and K. Zhang. Stacked Hourglass Network for Robust Facial Landmark Localisation. In 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pages 2025–2033, 2017
58. Amir Zadeh, Tadas BaltruÅ ¸aaitis, and Louis-Philippe Morency. Convolutional Experts Constrained Local Model for Facial Landmark Detection. arXiv:1611.08657 [cs], November 2016. arXiv: 1611.08657
59. Xiangyu Zhu, Zhen Lei, Xiaoming Liu, Hailin Shi, and Stan Z. Li. Face alignment across large poses: A 3d solution. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 146–155, 2016
60. Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks. IEEE Signal Processing Letters, 23(10):1499–1503, October 2016. arXiv: 1604.02878
61. Abdul Mueed Hafiz, Ghulam Mohiuddin Bhat: A Survey on Instance Segmentation: State of the art, arXiv:2007.00047, 2020.
62. Chien-Yao Wang, Alexey Bochkovskiy, Hong-Yuan Mark Liao: YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, arXiv:2207.02696, 2022.
63. Panos Achlioptas, Olga Diamanti, Ioannis Mitliagkas, and Leonidas Guibas. 2018. Learning representations and generative models for 3D point clouds. In International Conference on Machine Learning (ICML). 40–49
64. Yazeed Alharbi, Neil Smith, and Peter Wonka. 2019. Latent Filter Scaling for Multimodal Unsupervised Image-to-Image Translation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 1458–1466
65. Matthew Amodio and Smita Krishnaswamy. 2019. TraVeLGAN: Image-to-image Translation by Transformation Vector Learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 8983–8992
66. Moab Arar, Yiftach Ginger, Dov Danon, Ilya Leizerson, Amit Bermano, and Daniel Cohen-Or. 2020. Unsupervised Multi-Modal Image Registration via Geometry Preserving Image-to-Image Translation. arXiv preprint arXiv:2003.08073,2020.
67. Tero Karras, Samuli Laine, Timo Aila: A Style-Based Generator Architecture for Generative Adversarial Networks, arXiv:1812.04948, 2018.
68. Patrick Esser , Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Muller, Harry Saini Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, Dustin Podell, Tim Dockhorn Zion English, Kyle Lacey, Alex Goodwin, Yannik Marek, Robin Rombach: Scaling Rectified Flow

Transformers for High-Resolution Image Synthesis, arXiv:2403.03206, 2024
69. Blattmann, A., Dockhorn, T., Kulal, S., Mendelevitch, D., Kilian, M., Lorenz, D., Levi, Y., English, Z., Voleti, V., Letts, A., et al. Stable video diffusion: Scaling latent video diffusion models to large datasets. arXiv preprint arXiv:2311.15127, 2023
70. Brooks, T., Holynski, A., and Efros, A. A. Instructpix2pix: Learning to follow image editing instructions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 18392–18402, 2023.
71. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin: Attentio Is All You Need, arXiv:1706.03762, 2017
72. Huiwen Chang Han Zhang Lu Jiang Ce Liu˚ William T. Freeman: MaskGIT: Masked Generative Image Transformer, 2022.