Low-Cost Image-based Diagnosis System for Cleft Palate Reconstruction using 3D-Neural Modelling.

Swaraj Shaw

Dublin Business School
10592394@mydbs.ie

Abstract: Palatoplasty, or the surgical treatment of children born with cleft palate, is continuously evolving, and there are several preoperative issues that a cleft surgeon needs to consider before embarking on a palatal surgery. Hence, cleft palate surgery remains an enigma for most cleft surgeons. The proposal aims to cater to millions of underprivileged infants and adults by lowering the treatment cost in the diagnosis and pre-surgery therapy. This work proposes a 3D imaging technology based on deep learning that provides a comprehensive recording of the facial morphology that lends itself to the objective and subjective assessment of cleft palate repair surgery.

Index Terms/Keywords: Cleft, Cleft Palate, Palatoplasty, 3D Modeling, Image Processing, Image Reconstruction, Image Analysis, Deep Learning, Gap Analysis, Presurgical Therapy, Neural Radiance Fields, NeRF, Generative Adversarial Networks.

1 Introduction.

A Cleft Palate is a congenital disability where a child is born with an opening in the roof of the mouth (the palate) where a defect exists in the hard or soft palate, which leaves a hole between the nose and the mouth. Cleft is one of the most commonly occurring congenital disabilities globally, affecting about 1 in every 500-750 live births. Peterson-Falzone, Sally J., et al. [1] Cleft is usually diagnosed by directly examining the inside of the baby's mouth, which is time-consuming, requires the cooperation of the patient (both the newborn the parents), possesses the risk of further injury, and only identifies the most apparent disproportions of the face due to limited visual perspective provided by the traditional imaging tools.

Recent developments in AI has opened doors to new possibilities in the healthcare sector with the emerge of computer vision techniques many research labs are working on to perform supervised and unsupervised methods for building solutions such as, "Identifying Dental Caries using edge diagnostic tools implementing convolution neural networks and self back propagation techniques in Dentistry." Fadhillah, Excel Daris, et al. [2] Similar applications were based on an exercise where, "The depth in a 2D image was analyzed using a Siamese Network to learn facial key points delivering high resolution depth maps for facial feature retention in 3D structure". Pini, Stefano, et al. [3]

Due to the therapeutic difficulties in this area, numerous alternatives have been put up. Bone substitutes should ideally be bio-compatible, undergo remodeling, have a low fibrotic response, and support bone neoformation. Technically speaking, artificial bone substitutes should be just as strong as the cortical and cancellous bone they are meant to replace. In order to overcome these limitations recent works in developing bio-materials suggests promising results in tissue engineering and cleft palate reconstruction. Materials such as biomimetics (synthesis of materials which mimic biological processes) made using bioceramics, polymers and composites are favorable in reconstruction and treatment. Martín-del-Campo, et al. [4] Brézulier, Damien, et al. [5] The creation of three-dimensionally customized bone replacement grafts is still in its infancy. To develop a novel approach we will need to explore more on what has already been done and devise a deep learning algorithm which is reliable, cost effective and high precision.

2 State Of the Art.

Cleft palates have a significant impact on children's and teenagers' quality of life, despite the fact that they cannot yet be totally prevented. Deep learning has been extensively used in many medical fields, including robotic surgery assistance, semantic segmentation of medical images, and assistant diagnosis of various diseases. For example, CNN has been used to identify postures and instruments in surgical scenes and to assess surgical techniques. In this proposal we will be focusing on how to approach constructing a cast for cleft using a 3D spatial model and then printing the 3D structure on a 3D printer using Deep learning application.

At the moment there are a few machine learning solutions present which aim to restore oral functionality at the cost of precision, normally training a machine learning model is easier done on a smaller dataset as training a model on a bigger dataset the time taken to execute is usually lengthy and the amount of learning rate is comparatively lower than state of the art Deep Learning models like CNNs (convolutional neural networks) which are capable of delivering high performance accuracy and learns information node by node, layer by layer, pixel by pixel of the image mimicking a human brain end to end. Identifying facial keypoints are the first step in detecting the region of interest of the affected facial area. Hu, James B., et al. [6]

AI-based models are dominating in assisted diagnosis, simplifying pre-surgical planning, and prognosis prediction. Understanding visual scenes from different angles becomes quite important for such models so that it is able to understand the geometry and stitch a mesh 3D model. A well studied deep architecture was unveiled by Meta at International Conference on Computer Vision (ICCV) addressing how leveraging a Mask R-CNN framework on a 2D object infers a high resolution 3D structure. The model was trained on 10,000 pairs of images on Pix3D dataset. Gkioxari et al. [7]

3 Methodology.

The proposed solution simplifies the construction of a cleft palate casting by analyzing the depth in sample 2D images further converting the samples into a 3D spatial model.

Fig 1.

3.1 Pipeline.

1. Capture Image of the region of interests using a smartphone.
2. Feed the image into a mobile application.
3. Image Synthesis to 3D model (Outlines a 3D spatial map analyzing key points and the depth map).
4. Print a cheap 3D palate plate for the infant.

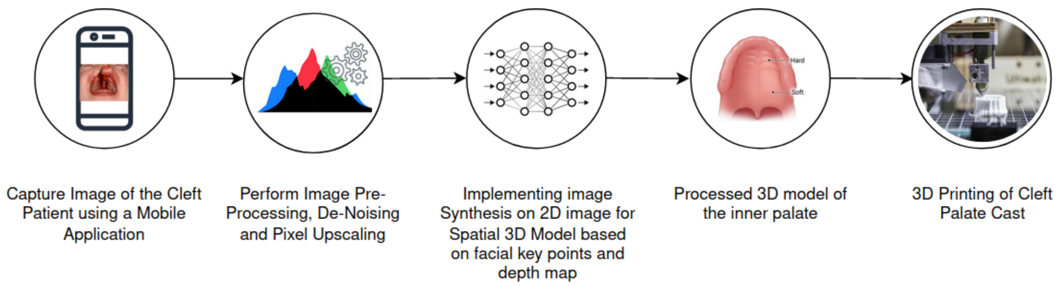

Fig 2.

3.2 Data Collection.

Our goal is to concentrate on the region of interest, the input image will be collected across multiple hospitals from a range of patients having cleft. The ideal image should be captured for a wide range of angles to stitch a 3D mesh using any smartphone. These images will be uploaded through a mobile application and further be stored in a block storage for further pre-processing and pixel up-scaling before training of the neural network.

3.3 Image Processing.

All the captured images will need to go through a tightly cropped reshaping technique of 128 x 128 pixels to maintain consistency among all images. All coloured images need to be gray-scaled to determine contours, edges and localized regions of interest will be processed for noise removal.

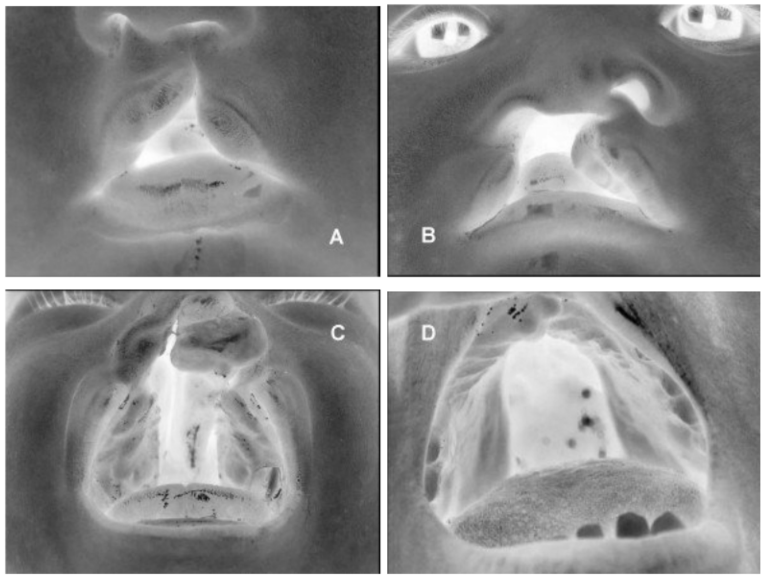

Fig 3. Cleft Sample Images. Stoll, Christian, et al. [8]

In the state where the sample of an image quality is blurred a up-scaling sampling technique will be applied to minimize the mean squared error(MSE) incurred based on a perceptual loss function using a deep residual network. In this step we train a generator model to obtain the best performing super resolution loss function. Here, the super-resolved image is ISR and the low resolution image is ILR and high resolution image IHR respectively. Goodfellow et al. [9] assuming D is the discriminator function where G is the generator.

$$\min_{\theta_G} \max_{\theta_D} \mathbb{E}_{I^{HR} \sim p_{\text{train}}(I^{HR})}[\log D_{\theta_D}(I^{HR})] + \mathbb{E}_{I^{LR} \sim p_G(I^{LR})}[\log(1 - D_{\theta_D}(G_{\theta_G}(I^{LR}))]$$

Fig 4. CVPR 2017 - SRGAN [10]

The processed output will go pre modeling stage where we perform a series of data augmentation techniques to generate new training sample (with existing limited samples) by rotating, flipping adding jitters to data, shearing and random shuffling so while training our model, it has a wide variety of perspective to learn from like real world scenarios. The preprocessed output is transformed to a batch of 2D images ready to be fed to our neural network.

3.4 Model Architecture.

Below is the proposed model architecture based on a neural rendering model (NeRF) that constructs a 3D scene on encoding a 2D image. We will be using a transfer learning technique to improvise the baseline model using ResNet-50 pretrained on ImageNet.

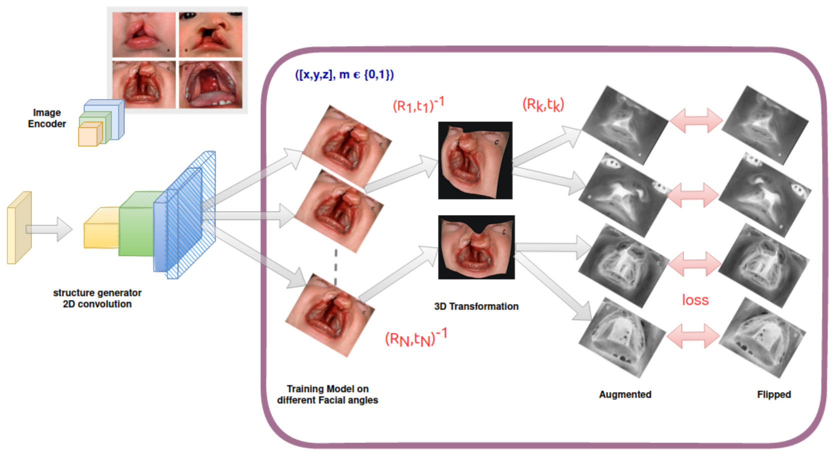

Fig 5. Encoding 2D image to 3D structure using neural modelling.

3.5 NeRF - Neural Radiance Fields.

Generating photo realistic models in 3D space is the motive behind neural rendering. It works by processing 2D images and letting the model learn on itself pixel by pixel how the image should look like in 3D space. It's a challenge for neural networks to understand the depth where each pixel intensity is computed in the range of 3 channels(RGB) of pixel values 0-255. Blending images with proper lighting conditions is necessary for any such models to look realistic and sharp. Neural Radiance fields as covered in the paper NeROIC by Zhengfei Kuang et al. 2022.
NERF is a neural network that tries to infer the color, opacity and radiance of each pixel using the images as inputs and guess the missing pixels for the small part of the objects that aren't present in the images. Legacy models usually need consistent lighting conditions and multiple samples increasing the size of the dataset available. The neural model takes in the image, segmentation masks it and makes assumptions based on positional and directional arguments and builds a radiance field to find the first guess of the density and color of each pixel. The rendering is separated from the lighting environment which improves the overall efficiency to adapt with varying lighting conditions. Using a 3D convolution with sobel filters the model analyses the edges on the 3D rendering. Spherical harmonics (group of basis functions defined on the sphere surface) representation of radiance is used in the lighting model optimizing the coefficients during training. This technique is prevalent in calculating lighting in a 3D model. Kuang et al. [11] The output is a realistic 3D model with a magnified amount of information which is what we are looking for in our 3D construction of facial cleft palate. The improvement on baseline NeRF model is expected to surpass with the inclusion of pretrained ResNet-50 as it will allow a broad perspective of different parameters.

4 Expected Outcome.

The 3D output model can then be extracted and used to print a plaster cast for palatal plates. The overall solution potentially can deliver quick 3D modeling under minutes, which drastically reduces the time taken in surgical procedures. The slice thickness of the 3D cleft palate should range from 0.5 to 1.5mm for best results. The inclusion of biomaterials and BPA free plastic reduces the risk of treatment. Given the feasibility the proposed solution can be deployed to larger markets with low income settings. The low cost prototype solution can also be deployed to regions with difficult to get medical facilities. With the production of 3D palate plates they can be implanted in the infants' mouths suffering from cleft can help in easier food intake.

5 Expected Timeline.

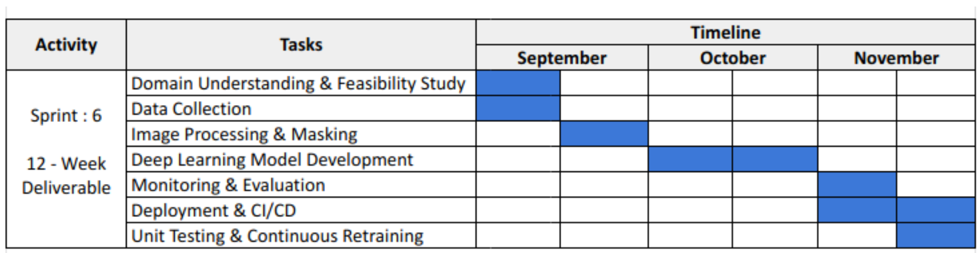

Fig 6. Expected Ball Park Estimate

6 Conclusion.

The overall solution proposes a brief on simplifying the production of cleft palates using deep learning techniques like neural modeling using radiance fields. In a way filling the gap of technology in the health care sector. The Proposed solution has sufficient validity for pre surgical therapeutic use for reconstruction of cleft palate.

References

[1] Sally J Peterson-Falzone, Mary A Hardin-Jones, Michael P Karnell, and Betty Jane McWilliams. *Cleft palate speech.* Mosby St. Louis, 2001.

[2] Excel Daris Fadhillah, Pravasta Caraka Bramastagiri, Riyanto Sigit, Sritrusta Sukaridhoto, Arya Brahmanta, and Bima Sena Bayu Dewantara. Smart odontogram: Dental diagnosis of patients using deep learning. In *2021 International Electronics Symposium (IES)*, pages 532–537. IEEE, 2021.

[3] Stefano Pini, Filippo Grazioli, Guido Borghi, Roberto Vezzani, and Rita Cucchiara. Learning to generate facial depth maps. In *2018 International Conference on 3D Vision (3DV)*, pages 634–642. IEEE, 2018.

[4] Marcela Martín-del Campo, Raúl Rosales-Ibañez, and Luis Rojo. Biomaterials for cleft lip and palate regeneration. *International Journal of Molecular Sciences*, 20(9):2176, 2019.

[5] Damien Brézulier, Louis Chaigneau, Sylvie Jeanne, and Ronan Lebullenger. The challenge of 3d bioprinting of composite natural polymers pla/bioglass: Trends and benefits in cleft palate surgery. *Biomedicines*, 9(11):1553, 2021.

[6] James B Hu, Andrew Guan, Qilong Zhangli, Lohrasb R Sayadi, Usama S Hamdan, and Raj M Vyas. Harnessing machine-learning to personalize cleft lip markings. *Plastic and Reconstructive Surgery Global Open*, 8(9 Suppl), 2020.

[7] Georgia Gkioxari, Jitendra Malik, and Justin Johnson. Mesh r-cnn. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9785–9795, 2019.

[8] Christian Stoll, Senait Mengsteab, Doris Stoll, Dieter Riediger, Axel M Gressner, and Ralf Weiskirchen. Analysis of polymorphic tgfb1codons 10, 25, and 263 in a german patient group with non-syndromic cleft lip, alveolus, and palate compared with healthy adults. *BMC Medical Genetics*, 5(1):1–9, 2004.

[9] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[10] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017.

[11] Zhengfei Kuang, Kyle Olszewski, Menglei Chai, Zeng Huang, Panos Achlioptas, and Sergey Tulyakov. Neroic: Neural rendering of objects from online image collections. *arXiv preprint arXiv:2201.02533*, 2022.