OrthoGAN: Visualizing teeth after Orthodontics treatment using Generative Models

Arbish Akram
PUCIT
Lahore, Pakistan
arbishakram@pucit.edu.pk

Nazar Khan
PUCIT
Lahore, Pakistan
nazarkhan@pucit.edu.pk

Abstract

In this paper, we propose OrthoGAN, conditional GANs based model to aid orthodontic patients and orthodontists to visualize synthetic post-treatment teeth images. Our proposed model learns a mapping between pre and post treatment clinical images. We employed generative models to facilitate orthodontists and orthodontic patients. In the orthodontics treatment, orthodontist takes X-rays, impression and images to diagnose and finalize treatment plan. Orthodontist and orthodontic patient uses these clinical images to identify and evaluate pre and post treatment effects on teeth. To our knowledge, this is the first work in the field of orthodontics. Experiments results demonstrate that our proposed OrthoGAN perform better as compared to the state-of-the-art models.

1 Introduction

With the development of Generative Adversarial Networks (GANs), computer vision has advanced significantly in medical applications. Recently, GANs have been used to develop many medical imaging applications [1, 2, 3, 4, 5].

Orthodontics is the field of dentistry that deal with overbites, cross bites, under bites, spacing between teeth, overcrowding of teeth and temporomandibular disorders (TMD) and to align misaligned teeth. Orthodontists most commonly use braces to fix these issues. We have employed GANs structure to facilitate orthodontists and orthodontic patients. Before treatment, the orthodontist maintains a complete set of clinical images that consist of intra-oral and extra-oral images. Orthodontist maintains these clinical images to diagnose, determine the best possible treatment and compare pre-treatment and post-treatment results [6, 7, 8, 9]. Extra-oral images contain information about a patient's facial features and smile aesthetics. It consist of these four images: frontal face with lips relaxed, frontal face with smile, right side profile and 45° profile. Intra-oral images provide in-depth view of teeth images from different angles to aid the orthodontist. It consist of these five images: Frontal in occlusion, left buccal in occlusion, right buccal in occlusion, upper occlusal using mirrors, and lower occlusal using mirrors.

We propose OrthoGAN, conditional generative adversarial network that takes pre-treatment teeth image as input and generates post-treatment teeth image. To the best of our knowledge, we are the first ones to apply GANs in the field of orthodontics. Some patients suffer from fear of orthodontic treatment [10]. These patients avoid or some times delay their appointments which leads to poor oral health. Orthodontists can use the proposed OrthoGAN to help such patients by showing them the synthetic post-treatment results as demonstrated in Figure 1.

We formulate OrthoGAN using conditional generative adversarial network. We verify correctness of our proposed method OrthoGAN, by using qualitative and quantitative evaluation. OrthoGAN generates photo-realistic synthetic post-treatment teeth image.

32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.

Figure 1: Pre-treatment image, Synthetic post-treatment image and real post-treatment image

2 Related Work

Generative Models: With the advent of Generative Adversarial Networks (GANs) [11] and deep convolutional GANs (DCGANs) [12], generative modeling have gained popularity in photo realistic image generation tasks. GANs consist of two adversaries: a generator (G) and a discriminator (D). The generator G tries to capture the data distribution while a discriminator D estimates the probability of a sample being real or fake using training data. Mathematically it can be written as:

$$\min_G \max_D L_{GAN}(D,G) = E_{x\sim p_{data}(x)}[\log D(x)] + E_{z\sim p_z(z)}[\log((1 - D(G(z)))] \tag{1}$$

Originally, GANs were designed to generate images from random noise [11, 12]. However, GANs variants guided by input image [13, 14, 15, 16, 17] or text have also become popular [18] and these are called conditional GANs (cGANs).

GANs extension to generate images on some condition called conditional Generative Adversarial Networks (cGANs). Conditional GANs receive conditioning information y on both G and D. The learning formulation now becomes:

$$\min_G \max_D L_{cGAN}(D,G) = E_{x,y\sim p_{data}(x,y)}[\log D(x,y)] + E_{x\sim p_{data}(x),z\sim p_z(z)}[\log(1 - D(x,G(x,z)))] \tag{2}$$

Image-to-Image translation: From many years, different image-to-image translation systems have been proposed in computer vision, image processing and graphics research [13, 19, 20, 21]. Recently, with the advent of GANs, photo-realistic results have been achieved in the image-to-image translation tasks [13, 16, 17, 14].

Isola et al. [13] pix2pix framework uses conditional adversarial network to learn mapping from source to target domain using paired data. They used ℓ_1 loss with adversarial loss in generator to generate images that are close to ground truth images as possible.

Sangkloy et al. [22] used similar idea to generate images from sketches. GANs have been used to reconstruct face images conditioned on different attributes (e.g. to add glasses, to change hair color, to smile or even change the gender of person) [23, 15, 24, 25, 16, 17]. Conditional GANs have been used for image manipulation guided by user [26] and style transfer [27]. GANs have been used by medical researchers also to develop medical imaging applications [1, 2, 5, 28]. Dai et al. [1] proposed structure correcting adversarial network (SCAN) to segment the lung fields and the heart from Chest X-ray images. Nie et al. [2] trained a patch-based GANs to translate brain MRI image to Computed Tomography (CT) image. Costa et al. [5] proposed a framework to learn the retinal vessel segmentation and then they learned mapping from binary vessel tree to new retinal image. In [29] synthetic focal liver lesions images are generated from CT and then they used CNN to classify liver lesion images. Xue et al. [30] proposed SegAN with a multiscale ℓ_1 loss function. Yang et al. [31] used GANs to introduce new CT image denoising method. They have used Wasserstein distance to improve performance of GANs and perceptual loss to suppress the noise. Hwang et al. [28] applied GANs in dental restorations by learning the mapping between input dental image and target image after placement of technician designed crown.

To the best of our knowledge this is the first model which attempt to aid orthodontic patients to visualize how their teeth will look after treatment

3 Method

Conditional generative adversarial network objective function can be defined as:

$$\min_G \max_D L_{cGAN}(D,G) = E_{x,y\sim p_{data}(x,y)}[\log D(x,y)] + E_{x\sim p_{data}(x)}[\log(1 - D(x, G(x)))] \tag{3}$$

Where D tries to maximize this objective against adversary G that tries to minimize this objective. We use ℓ_1 loss in generator with adversarial loss to make generator training more supervised.

$$L_{\ell_1}(G) = E_{x\sim p_{data}(x)}[||y - G(x)||_1] \tag{4}$$

Conditional generative adversarial network alone produces noise and artifacts on images that make generated images unrealistic. To avoid these issues and improve the quality of the generated images, we used feature reconstruction loss [32]. Let y and $\hat{y}$ be target and output images, we extract features from y and $\hat{y}$ images by passing them through feature reconstruction network ϕ (VGG16 [33] pretrained [34] model). Let ϕ^l(y) and ϕ^l(ŷ) denote features of target and output images extracted from layer l of the ϕ network, where l = $\{convk\_k\}_{k=1}^{3}$.

$$L_{FR}(G) = ||\phi^l(y) - \phi^l(\hat{y})|| \tag{5}$$

Our final objective is as follows:

$$G^* = \arg\min_G \max_D L_{cGAN}(D,G) + \lambda L_{\ell_1}(G) + L_{FR}(G) \tag{6}$$

4 Experiments

To show effectiveness of our proposed orthoGAN model, we make comparison with pix2pix [13]. The performance of models are evaluated quantitatively and qualitatively.

Dataset: We need dataset with set of image pairs, where each image pair contains pre treatment and post treatment teeth image. Since there no such dataset is available, We crawled intra-oral frontal occlusion pre and post treatment images from Google Images. In total, we collected 401 pre and post treatment teeth image pairs. We cropped and resized the mouth region to 64×128 and then aligned them manually. Some examples are shown in Figure 2. We randomly select 90 percent of data for training and 10 percent for testing.

Figure 2: Pre and post treatment images from our collected dataset

Network Architecture: OrthoGAN consists of generator and discriminator and a loss network. We have used the UNet [35] encoder decoder architecture in our generator. UNet introduced skip connections between encoding and decoding layers to allow sharing of low-level information like edges of objects. UNet with conditional GANs produces more realistic images as compared to previous approaches that produce blurry images. However, images produced by conditional generative adversarial networks still contain some noise and artifacts due to instability of their

training. Generator architecture consist of: Encoder (Conv64-Conv128-Conv256-Conv512) and Decoder (Deconv512-Deconv256-Deconv128-Deconv64). The feature maps at the end of decoder are mapped onto three color channels by applying convolution after last layer in decoder. BatchNorm is applied after each convolutional layer except first convolutional layer. The encoder uses Leaky ReLUs with slope 0.2, while the decoder uses ReLUs. For discriminator architecture, we have used the discriminator from PatchGAN [27] to classify each $N \times N$ patch of image and average all outputs to get final output. Since ℓ_1 loss encourages modelling of low level details, high level details are captures by giving attention to local patches. This is why we used PatchGAN discriminator architecture. Discriminator architecture consist of Encoder (Conv64-Conv128-Conv256-Conv512). To map one dimensional output we apply convolution after the last layer followed by sigmoid function. BatchNorm is applied after each convolutional layer except first convolutional layer. All ReLUs are leaky in discriminator with slope 0.2.

Training Procedure: We follow the training procedure of pix2pix [13], alternating between one gradient descent step on G and one step on D. We use Adam [36] with β_1 = 0.5 , β_2 and learning rate 0.0002. We use batch size 1 for all experiments. Each experiment is trained for 200 epochs.

Analysis of Results: To show effectiveness of our proposed OrthoGAN model, we conduct comparison with pix2pix [13] model. In Figure 3, we present qualitative comparison results. We observed in Figure 3, when pix2pix produced more blurry results Our proposed method OrthoGAN results are not so blurry. With our proposed OrthoGAN results are more realistic, post treatment images preserve shapes of teeth and color details. To evaluate our results quantitatively, we used Structure Similarity (SSIM) [37] and Mean Squared Error (MSE) to measure similarity and difference between synthetic and original post-treatment images. As shown in Table 1, compared to pix2pix [13], the MSE for OrthoGAN is lower while SSIM [37] is higher.

Table 1: Quantitative evaluation results

	pix2pix [13]	**OrthoGAN**
MSE_test	918.89	**851.97**
SSIM_test	0.5542	**0.5651**

5 Conclusion and Future Work

We have employed adversarial learning to generate synthetic post treatment teeth images to facilitate orthodontist and orthodontic patients. Experiment results demonstrate that our proposed OrthoGAN generates realistic synthetic post treatment intra-oral images as compared to state-of-the-art image to image translation model. In future, we will explore use of OrthoGAN with different datasets.

Figure 3: Teeth image synthesis on test dataset. First column shows pre treatment teeth image, second, third column shows pix2pix and OrthoGAN results, and right most column shows original post treatment teeth image

References

[1] Wei Dai, Nanqing Dong, Zeya Wang, Xiaodan Liang, Hao Zhang, and Eric P Xing. Scan: Structure correcting adversarial network for organ segmentation in chest x-rays. 2018.

[2] Dong Nie, Roger Trullo, Jun Lian, Caroline Petitjean, Su Ruan, Qian Wang, and Dinggang Shen. Medical image synthesis with context-aware generative adversarial networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 417–425. Springer, 2017.

[3] John T Guibas, Tejpal S Virdi, and Peter S Li. Synthetic medical images from dual generative adversarial networks. *arXiv preprint arXiv:1709.01872*, 2017.

[4] Maayan Frid-Adar, Idit Diamant, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. *arXiv preprint arXiv:1803.01229*, 2018.

[5] Pedro Costa, Adrian Galdran, Maria Ines Meyer, Meindert Niemeijer, Michael Abràmoff, Ana Maria Mendonça, and Aurélio Campilho. End-to-end adversarial retinal image synthesis. *IEEE transactions on medical imaging*, 37(3):781–791, 2018.

[6] UK Derby. Digital photography in orthodontics. *Journal of orthodontics*, 28:197–201, 2001.

[7] Uday Nandkishorji Soni, Shyama Dash, Dr Sagar Kausal, Mayuresh Baheti, Nilesh Mote, and Shubhangi Mani. Orthodontic photography–a clinical aspect.

[8] Jonathan Sandler and Alison Murray. Clinical photography in an orthodontic practice environment part 1. *Orthodontic Update*, 3(3):70–75, 2010.

[9] Carlos Henrique Theodoro Batista, Patrícia Rocha Coelho, Karla Daniela Malta Ferreira, and Maria das Graças Afonso Miranda Chaves. Digital photography in dentistry: Techniques and clinical importance.

[10] Surabhi R Jain and Sarvana Pandian. Prevalence of dental fear and anxiety among orthodontic patients (a survey). *Journal of Pharmaceutical Sciences and Research*, 8(9):1091, 2016.

[11] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[12] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[13] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. *arXiv preprint arXiv:1611.07004*, 2016.

[14] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint arXiv:1703.10593*, 2017.

[15] Guim Perarnau, Joost van de Weijer, Bogdan Raducanu, and Jose M Álvarez. Invertible conditional gans for image editing. *arXiv preprint arXiv:1611.06355*, 2016.

[16] Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jungkwon Lee, and Jiwon Kim. Learning to discover cross-domain relations with generative adversarial networks. *arXiv preprint arXiv:1703.05192*, 2017.

[17] Ming-Yu Liu, Thomas Breuel, and Jan Kautz. Unsupervised image-to-image translation networks. *arXiv preprint arXiv:1703.00848*, 2017.

[18] Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaolei Huang, Xiaogang Wang, and Dimitris Metaxas. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. In *IEEE Int. Conf. Comput. Vision (ICCV)*, pages 5907–5915, 2017.

[19] Aaron Hertzmann, Charles E Jacobs, Nuria Oliver, Brian Curless, and David H Salesin. Image analogies. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 327–340. ACM, 2001.

[20] David Eigen and Rob Fergus. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2650–2658, 2015.

[21] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440, 2015.

[22] Patsorn Sangkloy, Jingwan Lu, Chen Fang, Fisher Yu, and James Hays. Scribbler: Controlling deep image synthesis with sketch and color. *arXiv preprint arXiv:1612.00835*, 2016.

[23] Anders Boesen Lindbo Larsen, Søren Kaae Sønderby, Hugo Larochelle, and Ole Winther. Autoencoding beyond pixels using a learned similarity metric. *arXiv preprint arXiv:1512.09300*, 2015.

[24] Wei Shen and Rujie Liu. Learning residual images for face attribute manipulation. *arXiv preprint arXiv:1612.05363*, 2016.

[25] Shuchang Zhou, Taihong Xiao, Yi Yang, Dieqiao Feng, Qinyao He, and Weiran He. Genegan: Learning object transfiguration and attribute subspace from unpaired data. *arXiv preprint arXiv:1705.04932*, 2017.

[26] Jun-Yan Zhu, Philipp Krähenbühl, Eli Shechtman, and Alexei A Efros. Generative visual manipulation on the natural image manifold. In *European Conference on Computer Vision*, pages 597–613. Springer, 2016.

[27] Chuan Li and Michael Wand. Precomputed real-time texture synthesis with markovian generative adversarial networks. In *European Conference on Computer Vision*, pages 702–716. Springer, 2016.

[28] Jyh-Jing Hwang, Sergei Azernikov, Alexei A Efros, and Stella X Yu. Learning beyond human expertise with generative models for dental restorations. *arXiv preprint arXiv:1804.00064*, 2018.

[29] Maayan Frid-Adar, Idit Diamant, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. *arXiv preprint arXiv:1803.01229*, 2018.

[30] Yuan Xue, Tao Xu, Han Zhang, Rodney Long, and Xiaolei Huang. Segan: Adversarial network with multi-scale l_1 loss for medical image segmentation. *arXiv preprint arXiv:1706.01805*, 2017.

[31] Qingsong Yang, Pingkun Yan, Yanbo Zhang, Hengyong Yu, Yongyi Shi, Xuanqin Mou, Mannudeep K Kalra, Yi Zhang, Ling Sun, and Ge Wang. Low dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE Transactions on Medical Imaging*, 2018.

[32] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision*, pages 694–711. Springer, 2016.

[33] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[34] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015.

[35] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[36] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[37] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.