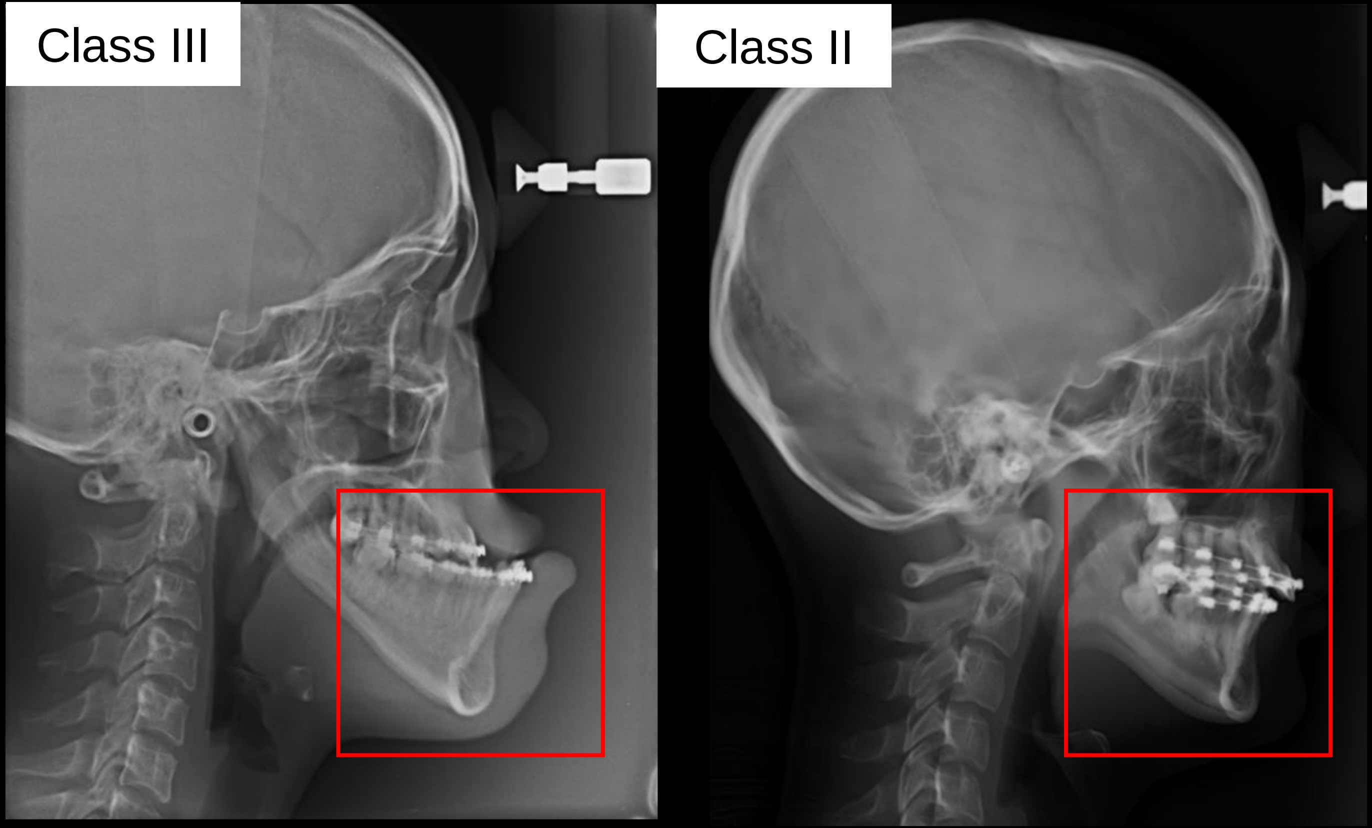
Asan Medical Center

Prediction on lateral cephalograms of post-orthognathic surgery using graph convolution neural network and diffusion model

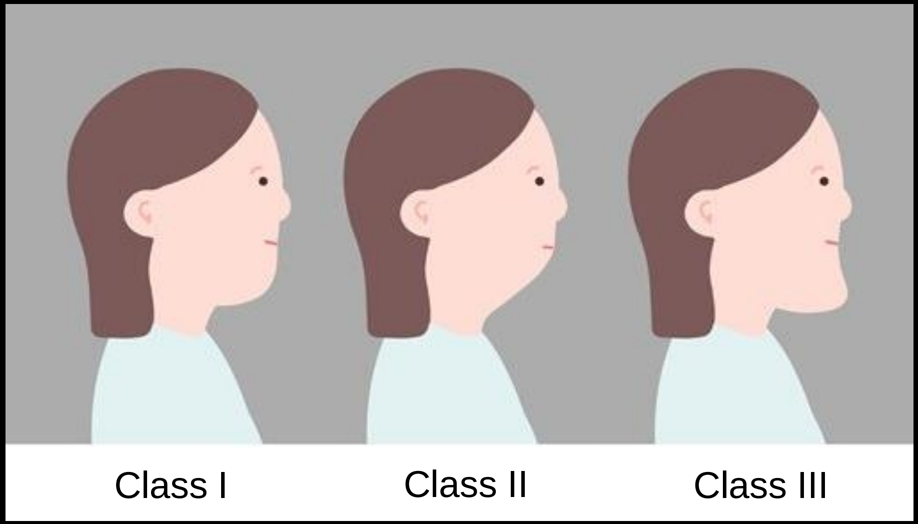
In-Hwan Kim*, Jiheon Jeong*, **Jun-Sik Kim***, Jin-Hyoung Cho, Mihee Hong, Kyung-Hwa Kang, Minji Kim, Su-Jung Kim, Yoon-Ji Kim, Sang-Jin Sung, Young Ho Kim, Sung-Hoon Lim, Seung-Hak Baek, Jae-Woo Park†, **Namkug Kim†**

Asan Medical Center, Univ. of Ulsan College of Medicine, South Korea

Orthognathic surgery (OGS)



Patients in need of Orthognathic Surgery (OGS)



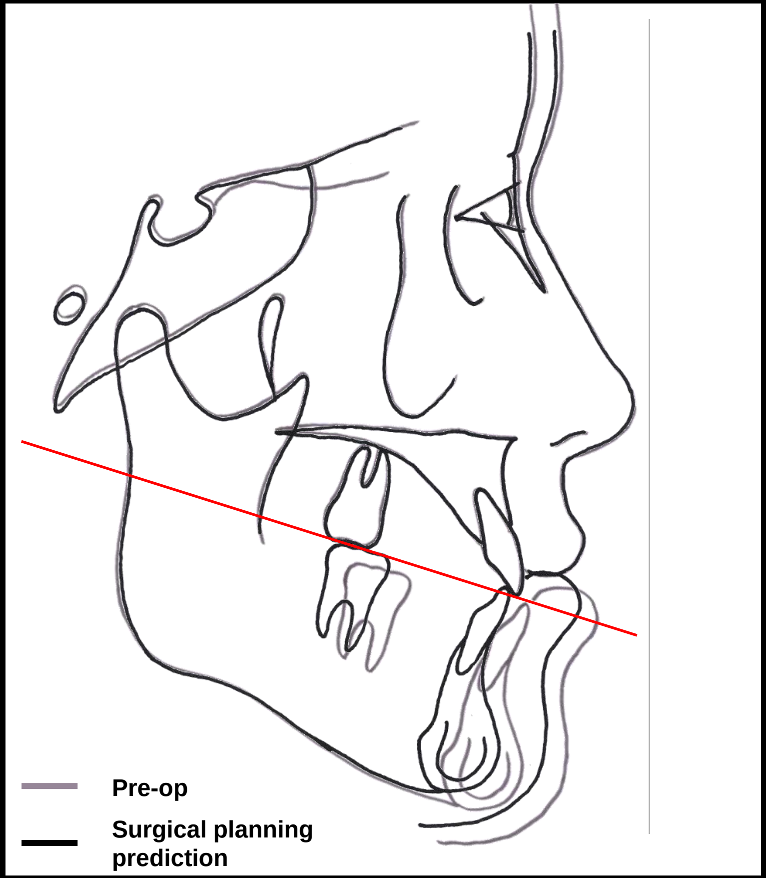
Category of skeletal deformity

Patients classified as Class II and Class III typically have skeletal deformities that cannot be corrected through conventional orthodontic treatments. They often require treatment through orthognathic surgery (OGS). For OGS, setting surgical treatment goals and predicting surgical outcomes are necessary to achieve a balance between aesthetics, functionality, and stability, ultimately increasing patient satisfaction.

Surgical planning for OGS

Asan Medical Center

MIRL



	Surgical Planning
U1 to MxOp (º)	52.0
U1 to PP (º)	111.5
L1 to MnOp (º)	73.0
IMPA (º)	96.0
AB to MxOp (º)	77.0
FH to MxOp (º)	16.5
Stm-Me'/Sn-Stm	2.04
Soft tss facial pl (º)	91.0
FH to A′B′ (º)	84.0
A′B′ to UL:LL:Pog′	6:6.5:2.5

Example of surgical planning

Therefore, Surgical Planning for Orthognathic Surgery (OGS) involves the following steps: 1. **Setting pre-surgical correction goals**, 2. **Developing precise surgical objectives** to achieve the best functional and aesthetic outcomes, and 3. **Generating facial profile goals** that can be used as visual references during consultations.

Low accuracy of prediction

Asan Medical Center

However, **the current predictions are less accurate** because 1. the prediction program sets the soft tissue landmarks to move in accordance with the underlying bony landmarks or 2. interpolates changes among adjacent landmarks with a moderate average, which compromises accuracy.

1. Both **Quick Ceph** and **V-Ceph** showed discrepancies between predicted values and actual measurements.

Myoung-Kyun Kim, Youn-Sung Choi, Song-Woo Chung, Young-Mi Jeon, & Jong-Ghee Kim (2005). Accuracy of soft tissue profile change prediction in mandibular set-back surgery patients: a comparison of Quick Ceph Image Pro™(ver 3.0) and V-Ceph™(ver 3.5). *THE KOREAN JOURNAL of ORTHODONTICS, 35(3), 216-226.*

2. When comparing the manual measurements with those predicted by Dolphin, discrepancies were observed in 7 out of the 16 measurements for the manual method and in 9 out of the 16 measurements for Dolphin.

Gossett, Christel Buck, et al. "Prediction accuracy of computer-assisted surgical visual treatment objectives as compared with conventional visual treatment objectives." *Journal of oral and maxillofacial surgery 63.5 (2005): 609-617.*

3. After reviewing nine research papers, it cannot be concluded that computer-based predictions are more accurate.

Kaipatur, Neelambar R., and Carlos Flores-Mir. "Accuracy of computer programs in predicting orthognathic surgery soft tissue response." *Journal of Oral and Maxillofacial Surgery 67.4 (2009): 751-759.*

Purpose

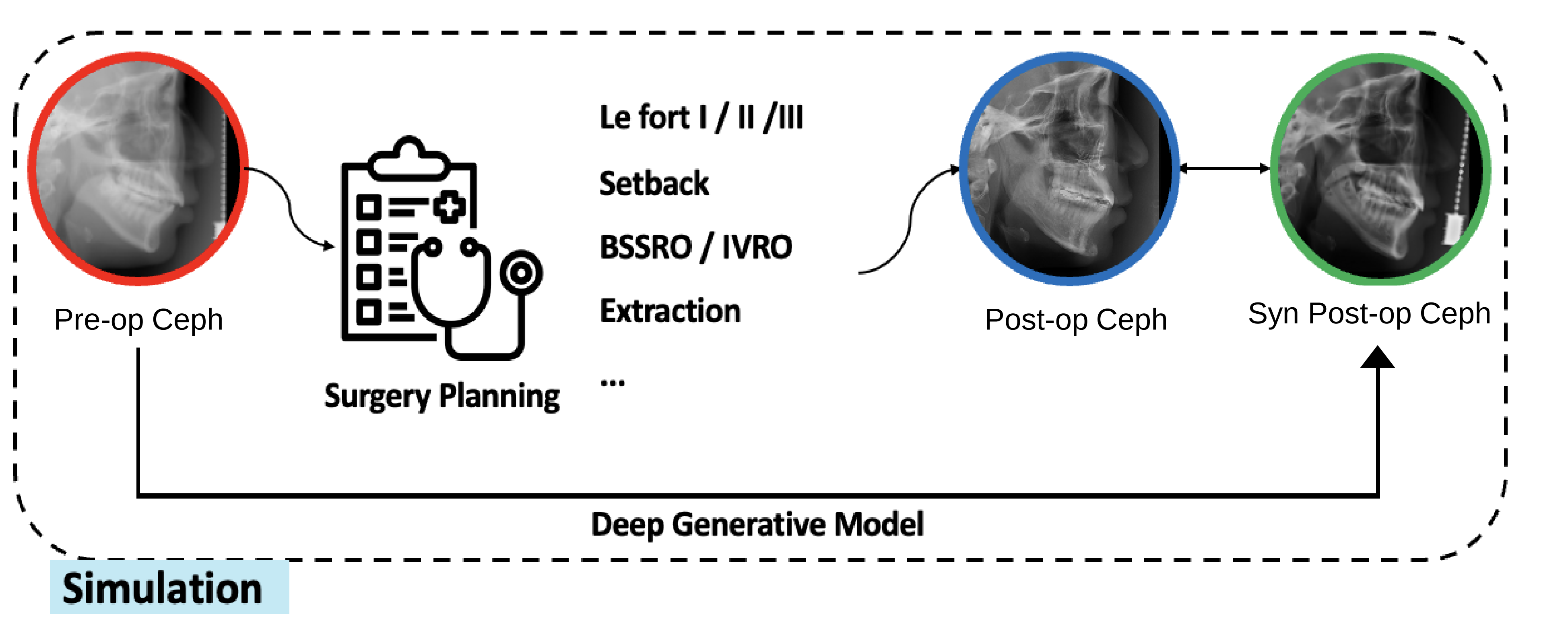
Therefore, our purpose is to develop a generative model that generates post-op cephalograms based on the surgical movement of pre-op cephalograms.

- To evaluate the synthesized post-op cephalograms quantitatively and qualitatively in collaboration with experts with over 20 years of experience.

- To evaluate the digital twin that controls the surgical movement of this model.

Virtual surgical planning using generative model

Asan Medical Center



Overall concept of our study

As a result, we aim to assist in surgical planning by using the generative model to replace inaccurate methods. Furthermore, our study will demonstrate its superiority through a total of three analyses.

1. Visual Turing test with 4 DDSs
2. Tracing & Superimposition with one DDS
3. Digital twin for surgical planning

Datasets and surgery type distribution

Asan Medical Center

Dataset collected by 9 university hospital and one dental hospital

A (SNUBH)	B (KHUDH)	C (KNUDH)	D (KOO)	E (WUDH)	F (KUDH)	G (EUMC)	H (CNUDH)	I (AUDH)	J (AMC)
330	160	30	130	9	8	9	5	7	19

Split the dataset into internal and external validation

Internal External

A (SNUBH)	B (KHUDH)	C (KNUDH)	D (KOO)	E (WUDH)	F (KUDH)	G (EUMC)	H (CNUDH)	I (AUDH)	J (AMC)
330	160	30	130	9	8	9	5	7	19

Eligible patients (n = 650)

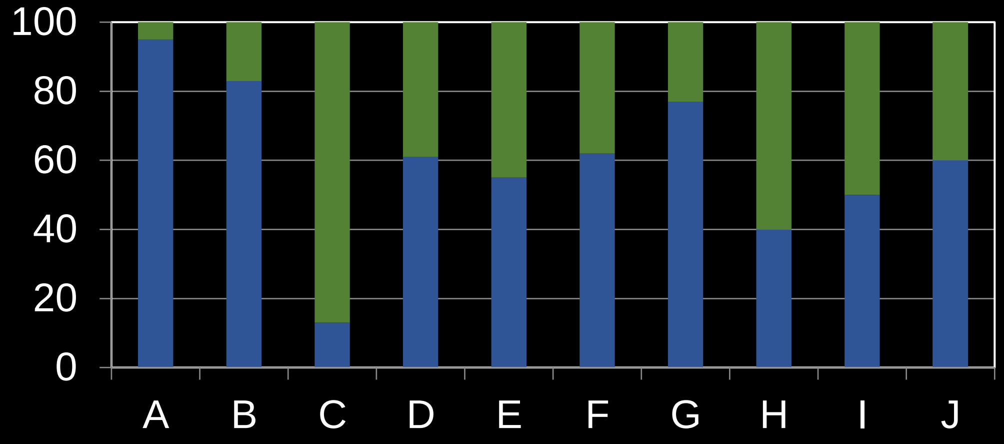
Eligible patients (n = 57)

Training (n = 550)

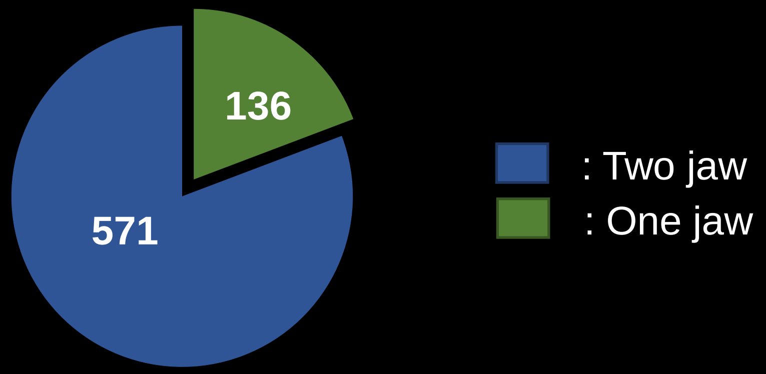
Validation (n = 50)

Internal Test (n = 50)

External Test (n = 57)



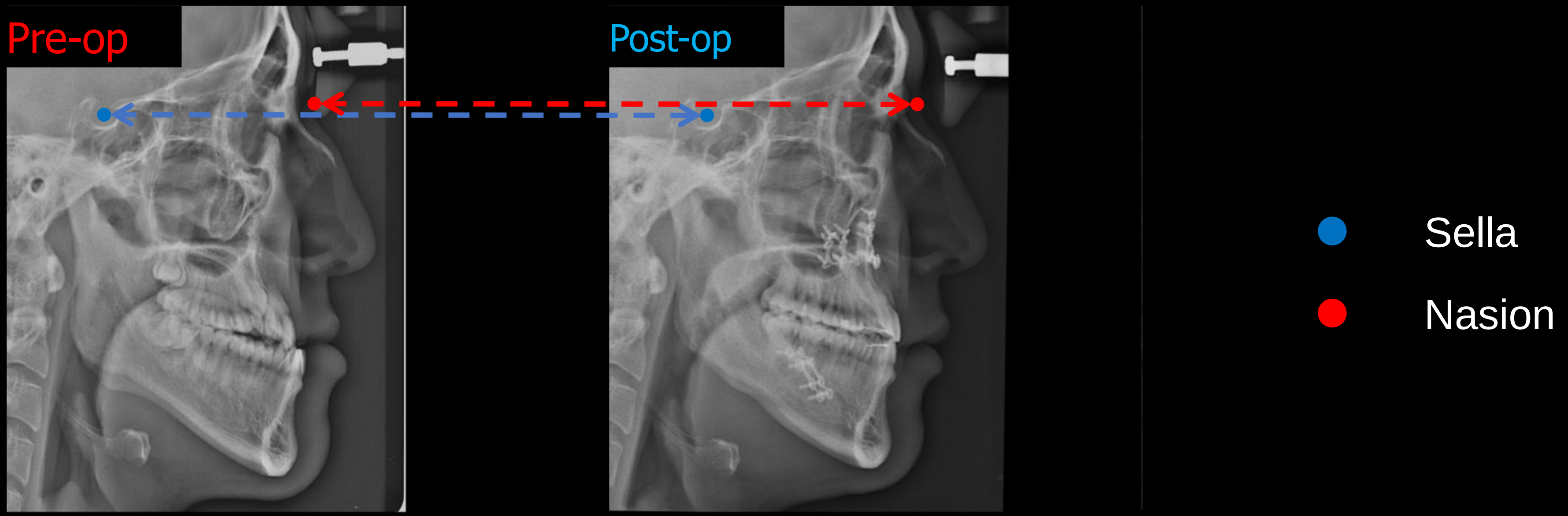
Surgery type by each hospital



Surgery type distribution

We constructed the dataset by considering the ratio to ensure that the dataset does not have a biased distribution of one and two jaw surgeries.

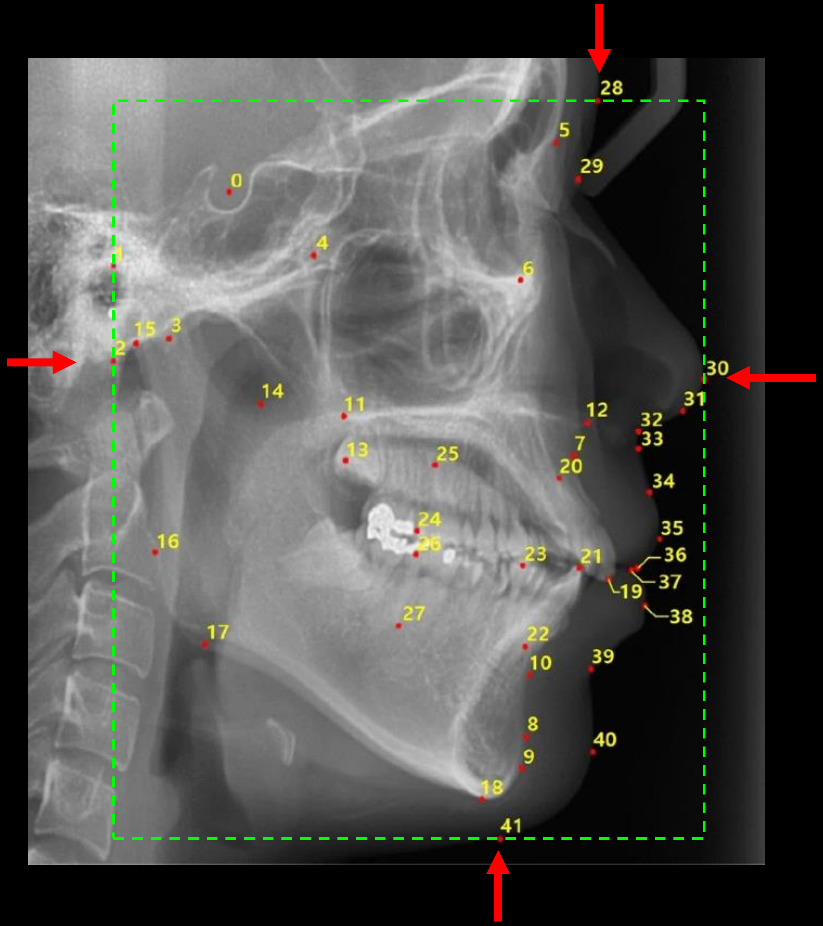
Registration between pre-op and post-op ceph



First, we aligned the points based on the Sella and Nasion and then rotated and resized them for alignment. Since Sella and Nasion are points that do not change due to surgery, we aligned the pre-op cephalogram and post-op cephalogram using these two reference points.

Cephalogram landmarks and crop protocol

Asan Medical Center

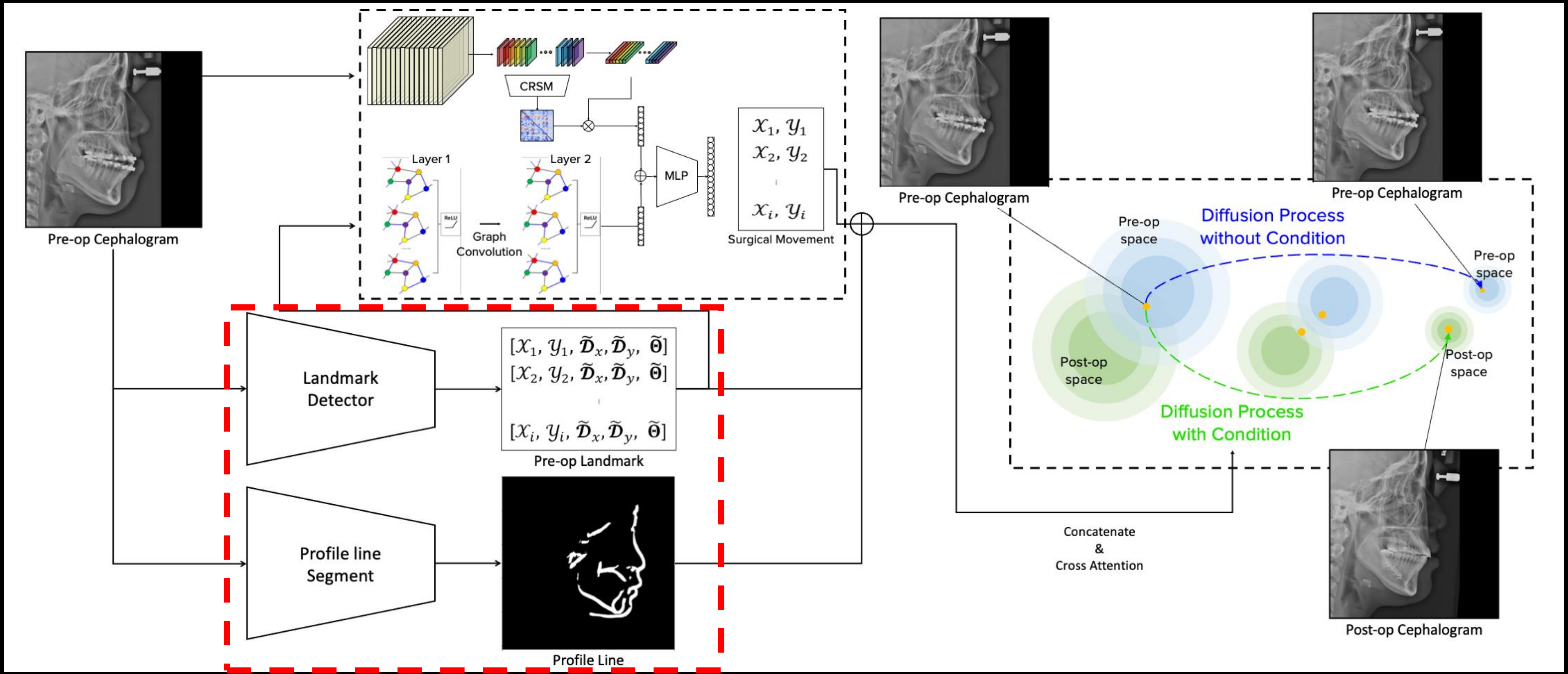


Index	Landmark	Index	Landmark	Index	Landmark	Index	Landmark
0	Sella	10	B-Point	21	Md 1Crown	31	Columella
1	Porion	11	PNS	22	Md 1 Root	32	Subnasale
2	Basion	12	ANS	23	Occlusal plane point	33	Soft tissue A
3	Hinge axis	13	R1	24	Mx 6 distal	34	Labrale Superius
4	Pterygoid	14	R3	25	Mx 6 root	35	Upper lip
5	Nasion	15	Articulare	26	Md 6 distal	36	Stomion superius
6	Orbitale	16	Ramus down	27	Md 6 root	37	Stomion inferius
7	A-Point	17	Corpus left	28	Glabella	38	Lower lip
8	Protuberanc e Menti	18	Menton	29	Soft tissue nasion	39	Soft tissue B
9	Pogonion	19	Mx 1 Crown	30	Pronasale	40	Soft tissue pogonion
		20	Mx 1 Root			41	Soft tissue menton

Next, we cropped the image by creating a square using the longest length among points <u>2(Basion)</u>, <u>41(Soft tissue menton)</u>, <u>30(Pronasale)</u>, and <u>28(Glabella)</u> as the reference, ensuring that all points are contained within the square

Landmark and profile line on pre-op. cephalogram

Asan Medical Center



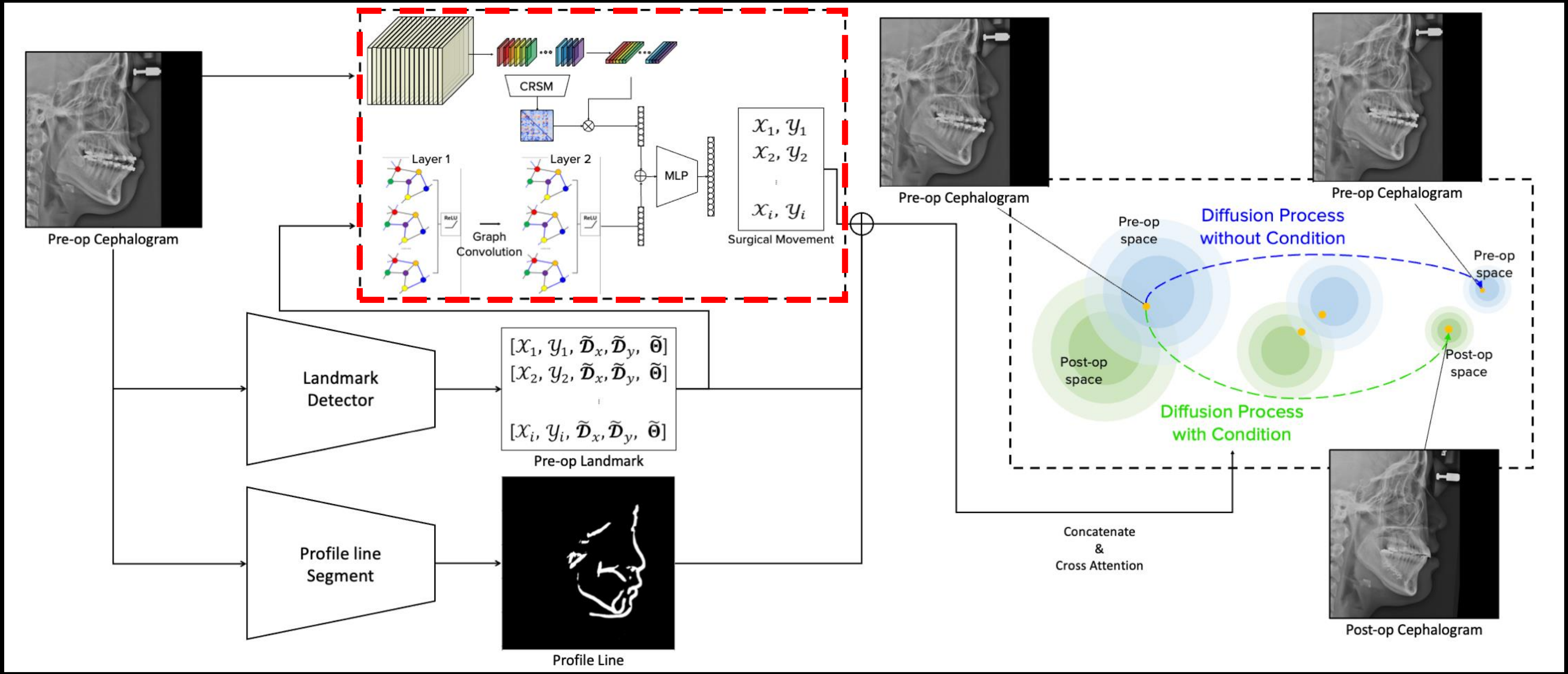
Generative Prediction for Orthognathic Surgery using Ceph (GPOSC Network)

When measuring 42 landmarks using the landmark detection model, the average error was found to be **0.79 ± 0.91 mm**. Considering the orthodontists' reproducibility error of **0.80 ± 0.79 mm**, it can be concluded that this is a sufficiently meaningful result. Furthermore, we trained a model using U-Net to predict profile lines in order to utilize additional information.

Kim, In-Hwan, et al. "Comparing intra-observer variation and external variations of a fully automated cephalometric analysis with a cascade convolutional neural net." *Scientific Reports* 11.1 (2021): 7925.

Post-op. landmark prediction module

Asan Medical Center



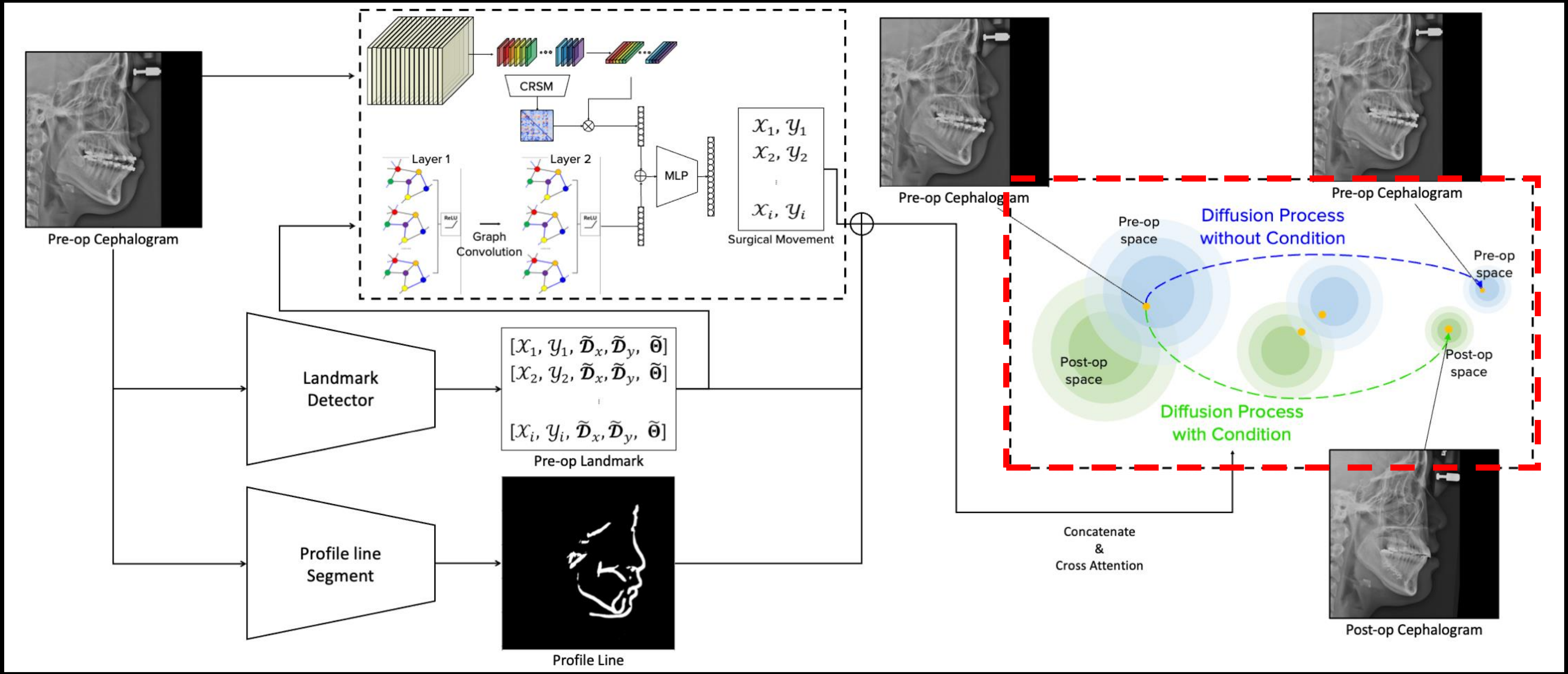
Generative Prediction for Orthognathic Surgery using Ceph (GPOSC Network)

We used **CNN-backbone** and **GCN-backbone** for predicting the surgical movement. For image embedding and topological structure representation, we adopted High-Resolution Network (HR-Net) and Graph Convolution Network (GCN). As a result, the surgical movement of all landmarks averaged 1.38 ± 0.97 mm.

Kim, In-Hwan, et al. "Orthognathic surgical planning using graph CNN with dual embedding module: External validations with multi-hospital datasets." *Computer Methods and Programs in Biomedicine* (2023): 107853.

Post-op. cephalogram generation module

Asan Medical Center



Generative Prediction for Orthognathic Surgery using Ceph (GPOSC Network)

We used diffusion-based model for generating synthetic cephalogram. When generating the synthetic post-op cephalogram, we used the condition as pre-op cephalogram, pre-op landmark, pre-op profile line and surgical movement predicted by surgical movement prediction model.

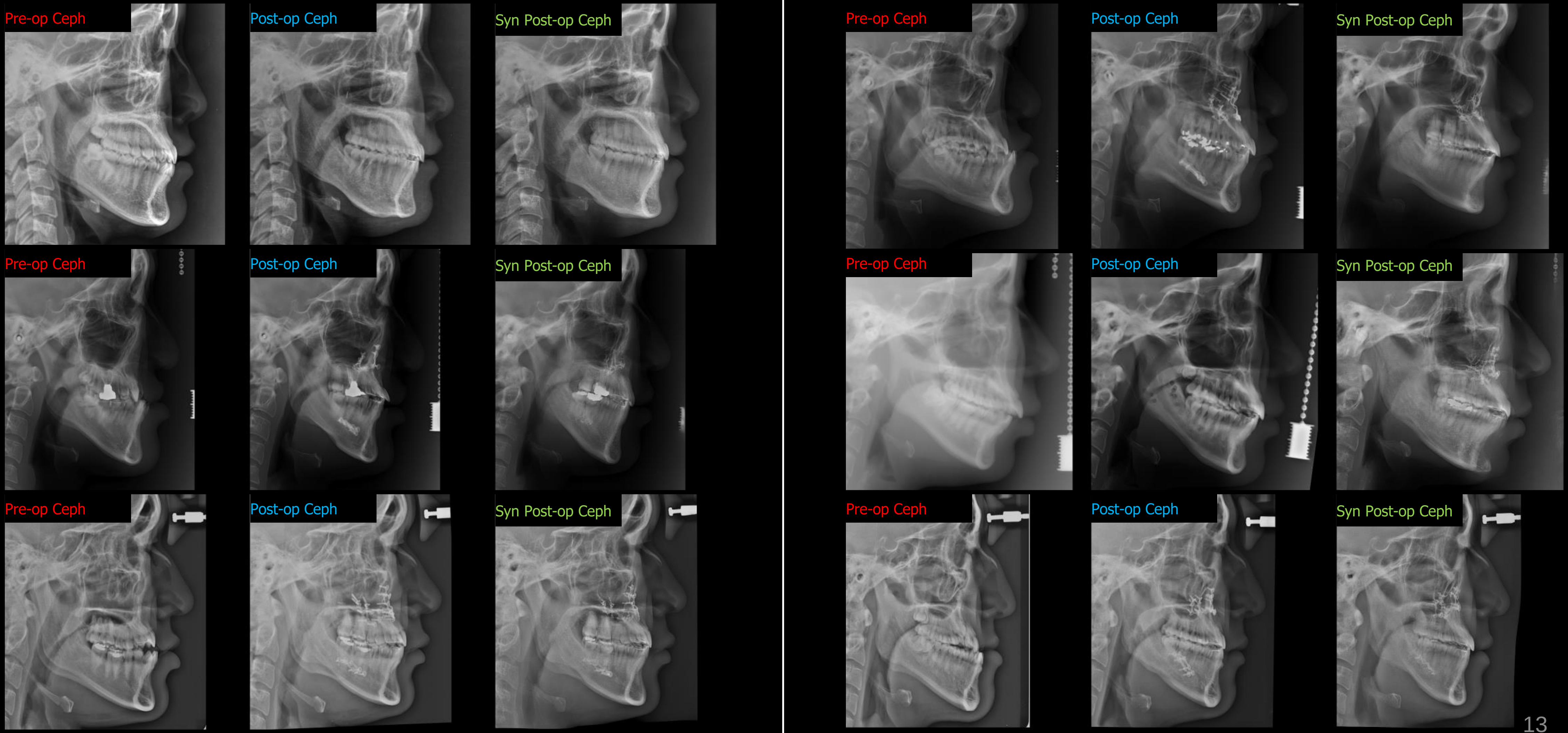
Rombach, Robin, et al. "High-resolution image synthesis with latent diffusion models." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022.

Synthetic post-op. ceph. generation by GPOSC-Net

Asan Medical Center

MIRL

Generation result of the synthetic post-op. images using GPOSC-Net



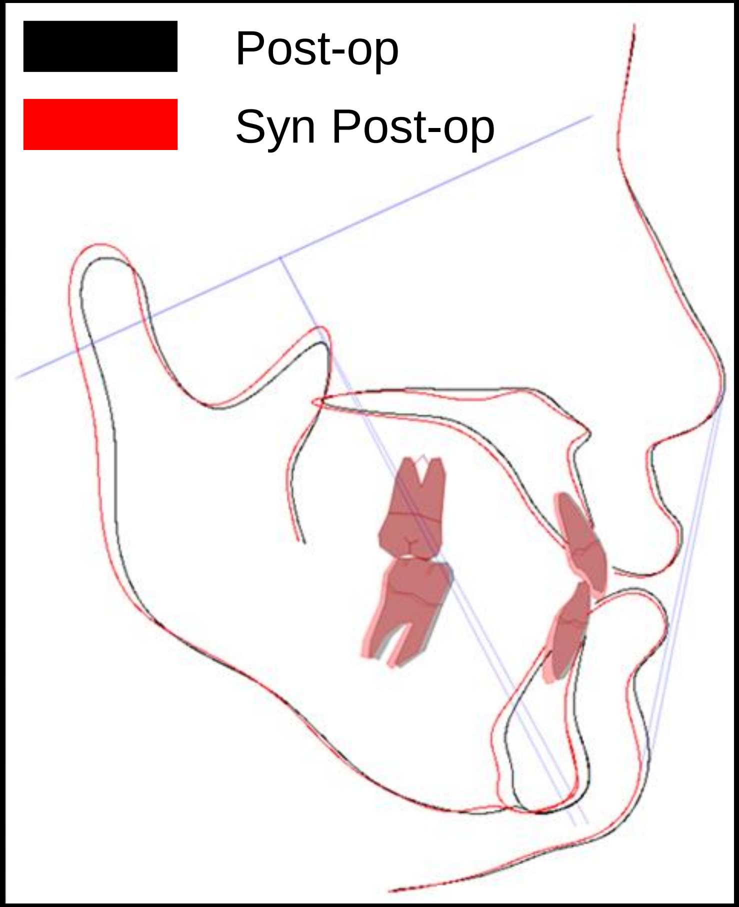
Tracing among reals and fakes

Asan Medical Center

Comparison of tracing between real post-op. ceph. and synthetic post-op. ceph.

Post-op.

Syn Post-op.

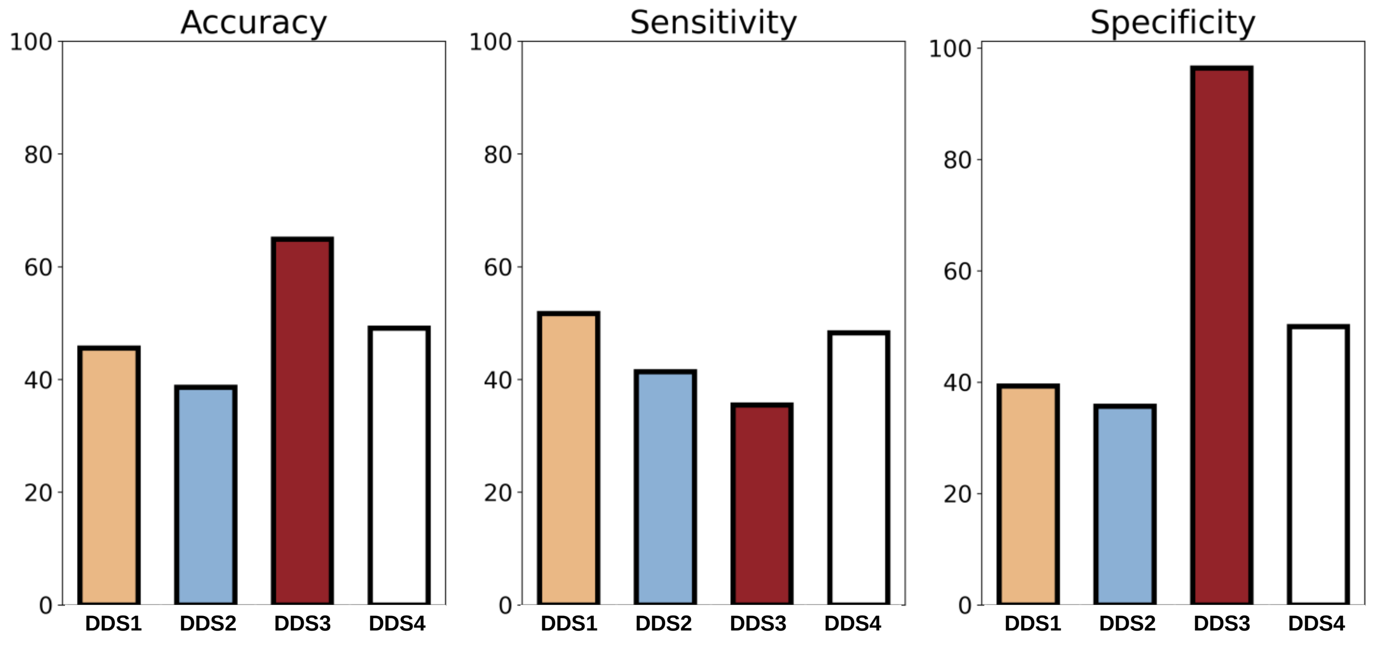


Tracing of orthodontist with 20 years experience

DDS: Doctor of Dental Surgery

Visual Turing test

To distinguish between real images and synthetic images

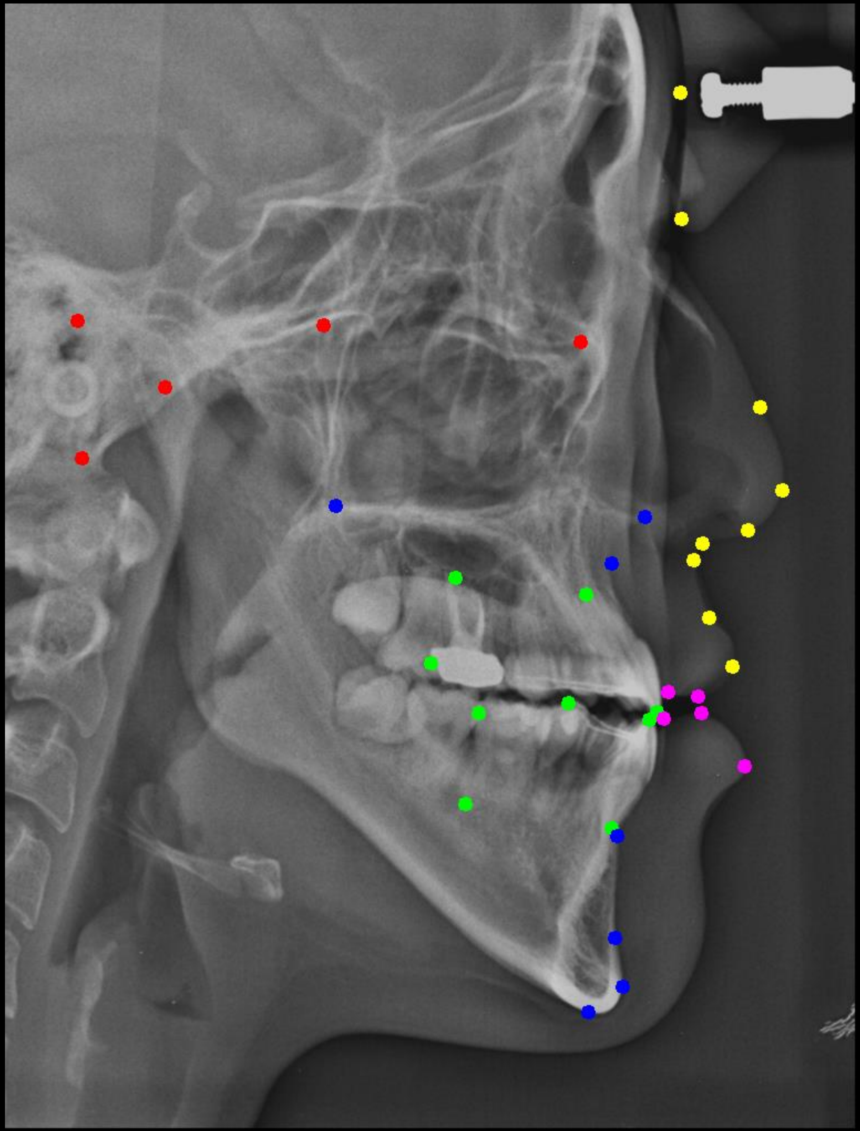


Visual Turing test of 4 Doctor of Dental Surgery (DDSs)

We requested a Visual Turing test from four DDSs, and the result showed an average accuracy of **46%.**
This means that it is **challenging** for the doctors **to distinguish between real post-op ceph and synthetic post-op. ceph**.

Landmark by anatomical group

Asan Medical Center

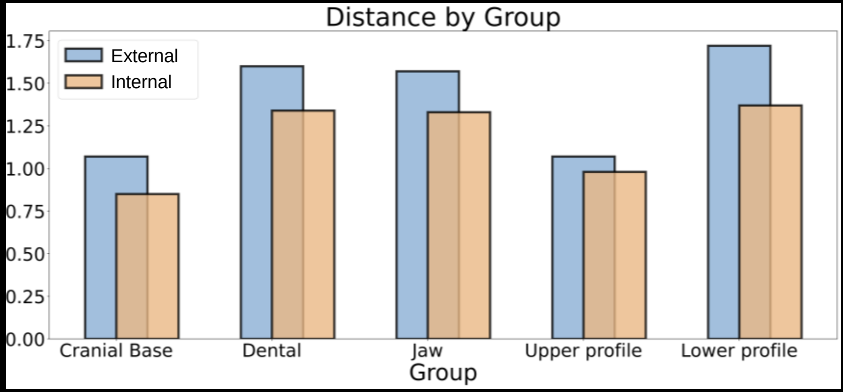


Cranial Base	Porion, Basion, Hinge axis, Pterygoid, Orbitale
Dental	Mx 1 Crown, Mx 1 Root, Md 1Crown, Md 1 Root, Occlusal plane point, Mx 6 distal, Mx 6 root, Md 6 distal, Md 6 root
Jaws	A-point, PM, Pogonion, B-point, PNS, ANS, Menton
Upper Profile	Glabella, Soft tissue nasion, Pronasale, Columella, Subnasale, Soft tissue A, Labrale Superius, Upper lip, Stomion superius
Lower Profile	Stomion inferius, Lower lip, Soft tissue B, Soft tissue pogonion, Soft tissue menton

Landmark errors by anatomical groups

Asan Medical Center

Comparison of landmark error selected from post-op cephalogram and synthetic post-op cephalogram



Landmark errors by anatomical groups (mm)

	Internal	External
Cranial Base	0.85±0.62 mm	1.07±0.79 mm
Dental	1.34±0.83 mm	1.60±1.08 mm
Jaws	1.33±0.86 mm	1.57±0.94 mm
Upper Profile	0.98±0.68 mm	1.07±0.76 mm
Lower Profile	1.37±0.89 mm	1.72±1.07 mm

Note: metric: Euclidean distance between post-op and syn post-op

The landmark error between **Post-op Ceph's landmarks** and **Synthetic Post-op Ceph's landmarks** measured by an orthodontist with 20 years experience.

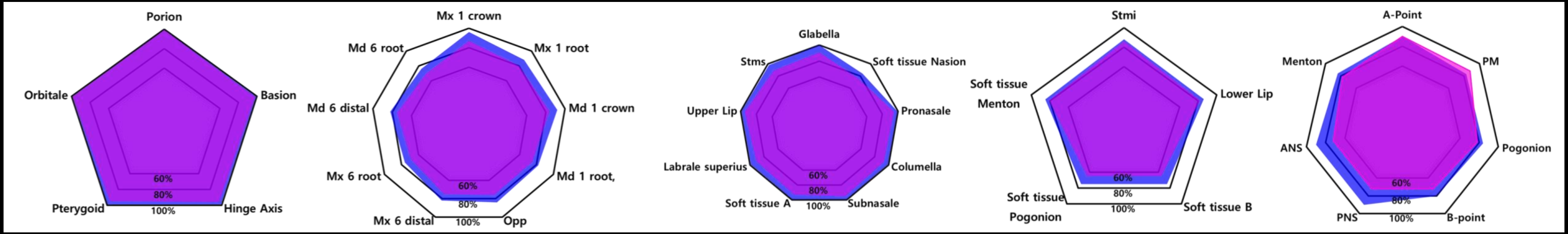
As a result, almost all landmark distance error came out within a medically significant range of 2mm, and furthermore, there is no significant difference between internal and external test set.

Successful prediction ratio (SPR)

Asan Medical Center

Successful prediction ratio (SPR) - each landmark was assessed according to errors < 2.0mm,
(as determined by an expert orthodontist with more than 20 years of experience.)

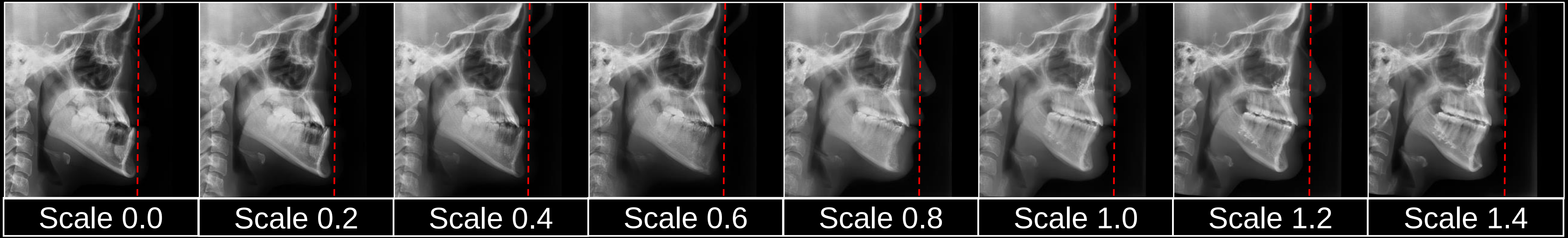
Cranial Base | Dental | Upper profile | Lower profile | Jaws



We calculated the SPR that determines whether landmarks of difference between post-op ceph and synthetic post-op ceph fall **within 2.0mm**. For all points belonging to the five groups in both internal and external categories, we achieved a hit ratio of over **80%.** This means that the landmarks are accurately located in precise positions.

Digital twin for surgical planning

Simulating with surgical movement by adjusting guidance scaling of diffusion model

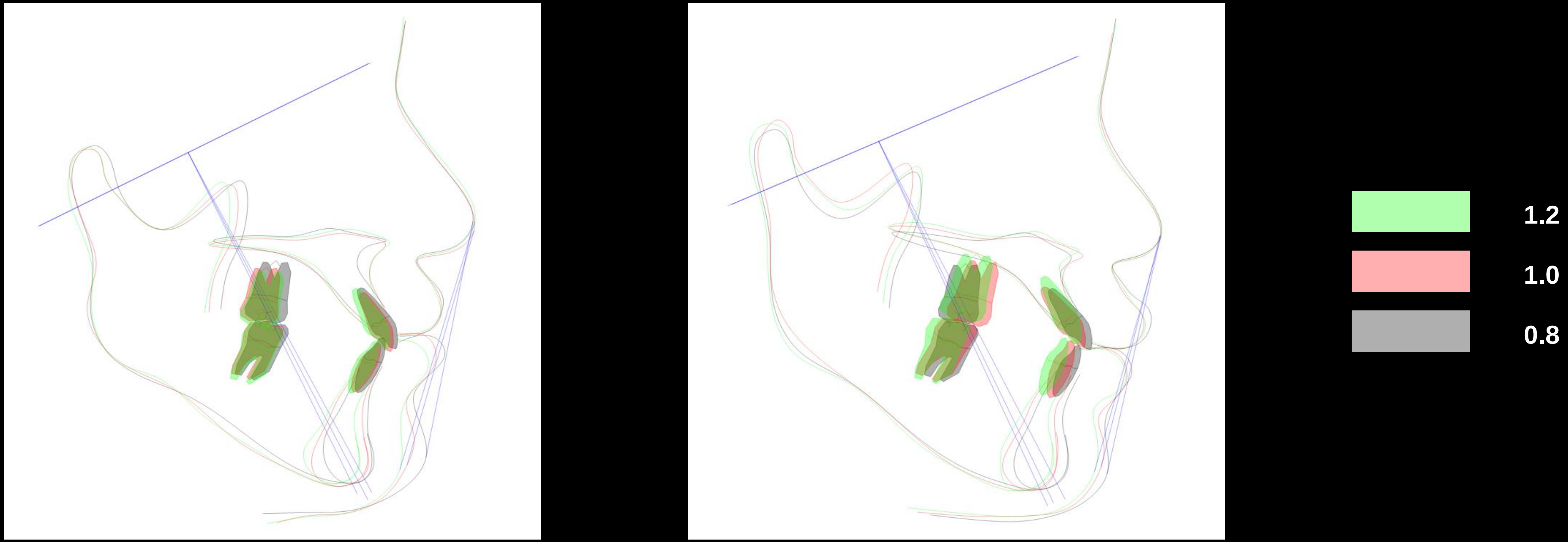


Digital twin in generation based on the guidance scale.

We have developed a digital twin that can predict surgical outcomes in various ways by adjusting the classifier-free guidance of the diffusion model. As the scale increases, it results in a greater amount of surgical movement. By adjusting the scale in the figure, it can be observed that the object moves further away from the red line.

Digital twin for surgical planning

Simulating with surgical movement by adjusting guidance scaling of diffusion model



Tracing of orthodontist with 20 years experience

To assess the changes in facial structure resulting from surgery, we adjusted the intensity of the condition in our generative model. The red corresponds to an embedding intensity of 1.0, the green to 1.2, and the gray to 0.8, representing the outcomes when generated under these conditions.

Conclusion and Discussion

Conclusion

1. To proposed an algorithm that uses a deep generative model to predict and generate post-op. lateral cephalograms.

2. The synthetic post-op. cephalograms were quantitatively and qualitatively evaluated by a panel of 2 to 4 doctors with over 15 years of experience, and the results were assessed as medically plausible.

3. To make a digital twin that allows the adjustment of surgical movement by controlling the guidance scale.

Limitation

1. We conducted the experiments using only one race from a single country.

2. The clinical utility evaluation is still lack.

Discussion

1. Since we used only one ethnicity in our study, we have the potential to conduct experiments on various ethnicities. Using a diverse range of ethnicities could make the experiment more meaningful.

2. Currently, our experiments have focused on surgical procedures, but we can also observe changes that may occur due to growth or other factors in the future.

3. While our experiments were conducted using cephalograms, we have the potential to expand to 3D modeling methods such as CT scans.

Q & A