

Article

Automatic Reproduction of Natural Head Position in Orthognathic Surgery Using a Geometric Deep Learning Network

Ji-Yong Yoo ^{1,†}, Su Yang ^{2,†}, Sang-Heon Lim ³, Ji Yong Han ³ , Jun-Min Kim ⁴, Jo-Eun Kim ⁵ , Kyung-Hoe Huh ⁵ , Sam-Sun Lee ⁵ , Min-Suk Heo ⁵, Hoon Joo Yang ^{6,*}  and Won-Jin Yi ^{2,3,5,*}

- ¹ Department of Biomedical Radiation Sciences, Graduate School of Convergence Science and Technology, Seoul National University, Seoul 08826, Republic of Korea; gd_yoo@snu.ac.kr
 - ² Department of Applied Bioengineering, Graduate School of Convergence Science and Technology, Seoul National University, Seoul 08826, Republic of Korea; s8431@snu.ac.kr
 - ³ Interdisciplinary Program in Bioengineering, Graduate School of Engineering, Seoul National University, Seoul 08826, Republic of Korea; smion123@snu.ac.kr (S.-H.L.); jiyong@snu.ac.kr (J.Y.H.)
 - ⁴ Department of Electronics and Information Engineering, Hansung University, Seoul 02876, Republic of Korea; jmkim@hansung.ac.kr
 - ⁵ Department of Oral and Maxillofacial Radiology, Dental Research Institute, School of Dentistry, Seoul National University, Seoul 03080, Republic of Korea; noel1st@snu.ac.kr (J.-E.K.); future3@snu.ac.kr (K.-H.H.); raylee@snu.ac.kr (S.-S.L.); hmslsh@snu.ac.kr (M.-S.H.)
 - ⁶ Department of Oral and Maxillofacial Surgery, Dental Research Institute, School of Dentistry, Seoul National University, Seoul 03080, Republic of Korea
- * Correspondence: yanghoonjoo@snu.ac.kr (H.J.Y.); wjyi@snu.ac.kr (W.-J.Y.)
- † These authors contributed equally to this work.



Academic Editors: Qi-Yong Hemis Ai, Kuo Feng Hung and Jinzhong Yang

Received: 12 November 2024

Revised: 10 December 2024

Accepted: 26 December 2024

Published: 27 December 2024

Citation: Yoo, J.-Y.; Yang, S.; Lim, S.-H.; Han, J.Y.; Kim, J.-M.; Kim, J.-E.; Huh, K.-H.; Lee, S.-S.; Heo, M.-S.; Yang, H.J.; et al. Automatic Reproduction of Natural Head Position in Orthognathic Surgery Using a Geometric Deep Learning Network. *Diagnostics* **2025**, *15*, 42. <https://doi.org/10.3390/diagnostics15010042>

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Background: Accurate determination of the natural head position (NHP) is essential in orthognathic surgery for optimal surgical planning and improved patient outcomes. However, traditional methods encounter reproducibility issues and rely on external devices or patient cooperation, potentially leading to inaccuracies in the surgical plan. **Methods:** To address these limitations, we developed a geometric deep learning network (NHP-Net) to automatically reproduce NHP from CT scans. A dataset of 150 orthognathic surgery patients was utilized. Three-dimensional skull meshes were converted into point clouds and normalized to fit within a unit sphere. NHP-Net was trained to predict a 3×3 rotation matrix to align the CT-acquired posture with the NHP. Experiments were conducted to determine optimal point cloud sizes and loss functions. Performance was evaluated using mean absolute error (MAE) for roll, pitch, and yaw angles, as well as a rotation error (RE) metric. **Results:** NHP-Net achieved the lowest RE of $1.918^\circ \pm 1.099^\circ$ and demonstrated significantly lower MAEs in roll and pitch angles compared to other deep learning models ($p < 0.05$). These findings indicate that NHP-Net can accurately align CT-acquired postures to the NHP, enhancing the precision of surgical planning. **Conclusions:** By effectively improving the accuracy and efficiency of NHP reproduction, NHP-Net reduces the workload of surgeons, supports more precise orthognathic surgical interventions, and ultimately contributes to better patient outcomes.

Keywords: orthognathic surgery; computed tomography; natural head position; head pose estimation; geometric deep learning

1. Introduction

Orthognathic surgery necessitates the accurate determination of the natural head position (NHP) as it is a fundamental reference posture extensively utilized in this field [1]. NHP

represents the patient's most natural and comfortable head position, providing a consistent baseline for assessing and diagnosing facial and cranial asymmetries [2–6]. This consistent baseline is essential for optimizing surgical outcomes by offering a standardized reference point, aiding in precise surgical planning, predicting post-operative results, and enhancing communication among dental professionals [7]. Consequently, the accurate determination of NHP is critical for achieving successful surgical interventions and improving patient outcomes in craniofacial analysis [6].

Misalignment in NHP can lead to suboptimal surgical outcomes and patient dissatisfaction. Achieving the correct head position involves defining NHP as the upright head position with eyes fixating on a distant point at eye level, serving as the ideal reference posture [5]. Figure 1 shows 3D skull meshes from CT scans before and after NHP reproduction for the patient with facial asymmetry, highlighting the importance of accurate NHP determination in surgical planning. NHP is particularly important for patients with facial asymmetries, as it ensures the surgical corrections are aligned with the patient's natural anatomical orientation. The importance of NHP extends beyond surgical procedures; it also plays a vital role in pre-operative planning and post-operative assessment, where even slight deviations can significantly impact the results and patient satisfaction [8].

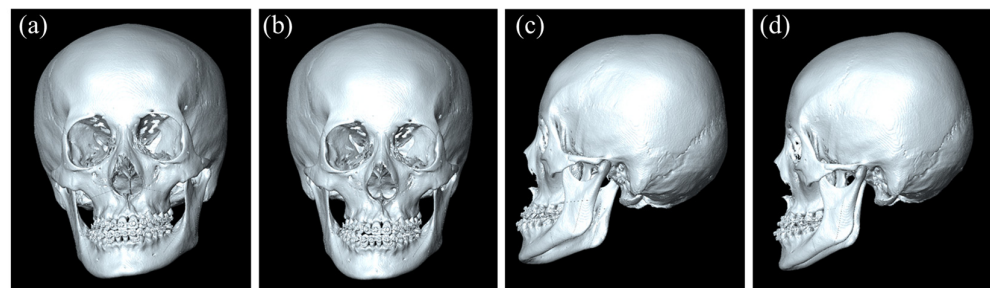


Figure 1. Three-dimensional skull mesh from CT scans before NHP reproduction (a) and after NHP reproduction (b) in the front view. Three-dimensional skull mesh from CT scans before NHP reproduction (c) and after NHP reproduction (d) in the side view.

However, the current methods for NHP reproduction have several limitations and inconveniences that impact their effectiveness and practicality. First, there is a lack of reproducibility due to individual variations and the difficulty for patients to maintain a consistent natural posture [9]. Patients may unintentionally alter their head position between measurements, leading to inconsistencies that can affect diagnostic accuracy and treatment outcomes. Second, traditional intracranial landmarks, such as the sella-nasion line and Frankfort horizontal plane [3,10,11], can be unreliable, especially in patients with significant craniofacial deformities [12–14], as these landmarks may not accurately reflect true anatomical orientation. Such discrepancies may lead to misalignment of surgical templates and inaccurate cephalometric measurements, which are essential for precise surgical planning. Deviations in NHP as small as 1–2° can result in cephalometric errors of 2–3 mm, leading to intraoperative adjustments, prolonged surgical time, and increased patient burden. This not only reduces surgical precision but also negatively impacts patient satisfaction [5,11,15–17]. To address these limitations, recent advancements in three-dimensional (3D) NHP reproduction methods have been introduced [14,18]. While these methods offer improvements, they often require additional scanning processes or equipment, which may not eliminate all practical challenges. Third, the use of external devices or markers, such as customized bite jigs, orientation sensors, radiopaque stickers, or pose estimation algorithms like a pose from orthography and scaling with iterations (POSIT), can be uncomfortable and may alter natural facial expressions [19–21]. These devices require additional installation procedures and are cumbersome and complex. This can lead to errors in surgical planning

and compromise patient comfort [22]. Finally, methods involving additional equipment or procedures, like 3D scanners or gyroscope sensors, increase time consumption, costs, and patient burden [23]. Extra scanning processes with additional surgical planning times require significant financial investment, which may not be feasible in all clinical settings. Therefore, an automatic and accurate method is required to reproduce NHP on CT scans from orthognathic patients.

Recent advances in deep-learning-based head pose estimation have significantly impacted fields such as computer vision and human–machine interaction, demonstrating precise head pose estimation techniques from images and videos for applications like gaze tracking, facial recognition, and virtual reality [24–26]. These advancements have shown that deep learning models can accurately predict head poses under various conditions, suggesting their potential applicability to medical fields [24,27,28]. As far as we know, however, no previous studies have reproduced NHP in CT scans to improve the accuracy of orthognathic surgery using deep learning.

The purpose of this study is to improve the accuracy of orthognathic surgery by automatically reproducing NHP in CT scans using a geometric deep learning network (NHP-Net). We developed NHP-Net that took a point cloud input from the skull mesh and regressed 3×3 rotation matrix elements to reproduce NHP directly. Our contributions include the following: (1) we developed a geometric deep learning network named NHP-Net that predicts rotation matrix elements of NHP in the CT scan of an orthognathic patient; (2) we conducted comprehensive experiments to assess the effectiveness of the size of points and loss functions.

2. Materials and Methods

2.1. Data Acquisition and Preparation

We included 150 patients (87 females and 63 males; mean age 28.42 ± 8.53 years) who underwent orthognathic surgeries at Seoul National University Dental Hospital from 2018 to 2023. The cohort was selected to represent a broad range of clinical presentations typical of orthognathic surgery cases. Specifically, patients with Class II and Class III malocclusions as well as various mandibular shapes were included to ensure diversity in anatomical features. However, patients with severe craniofacial anomalies beyond common skeletal discrepancies or those who had previously undergone extensive facial surgeries were excluded to reduce variability and maintain model robustness. This selection strategy aimed to create a dataset representative of most orthognathic surgery cases, ensuring that the model's performance would be applicable to typical clinical scenarios. CT scans from patients were obtained using a multi-detector computed tomography (SOMATOM Sensation 10, Siemens, Munich, Germany) operating at 120 kVp and 80 mA, with a slice thickness of 0.75 mm. All patient data were anonymized to protect privacy. The NHP for each patient was manually annotated by an oral and maxillofacial surgeon using commercial software (Mimics Research 19.0, Materialise, Leuven, Belgium). Maxillofacial bone areas were segmented using the global thresholding technique in CT scans, and the segmented maxillofacial bone areas were converted into skull mesh data (STL format). Specifically, global thresholding involves segmenting bone regions based on CT intensity values measured in Hounsfield units (HU). Let $I(x, y, z)$ represent the CT intensity at voxel (x, y, z) , and T (in HU) be the empirically determined threshold value. The segmented volume S is defined as

$$S(x, y, z) = \begin{cases} 1, & \text{if } I(x, y, z) \geq T \\ 0, & \text{if } I(x, y, z) < T \end{cases} \quad (1)$$

Only voxels with intensities above T are considered part of the bone. This thresholding ensures accurate extraction of bone structures, which are then converted into 3D skull

meshes for subsequent processing by NHP-Net. We uniformly sampled 2048 point clouds from each skull mesh data, and each point cloud input was normalized by shifting the centroid of the center points to the origin [29,30]. We chose to represent the skull geometry as a point cloud rather than using the raw mesh format. This approach eliminates the need to handle irregular mesh connectivity and ensures a uniform input structure. By sampling a fixed number of points, we maintain a consistent data representation, allowing NHP-Net to focus on orientation prediction without being influenced by varying mesh quality or resolution. The CT dataset was divided into training, validation, and test sets, consisting of 90, 30, and 30 patients, respectively.

2.2. Network Architecture of NHP-Net

In this study, we proposed a geometric deep learning network (NHP-Net) to automatically reproduce NHP from CT scans, as shown in Figure 2. NHP-Net took a point cloud input from the skull mesh and regressed 3×3 rotation matrix elements to reproduce NHP. Our NHP-Net has two set abstraction blocks (SAB), a global max-pooling (GMP), and fully connected (FC) layers, as shown in Figure 3. Each SAB progressively downsamples the point cloud while extracting local and global geometric features at multiple scales. To achieve this, each SAB performs three key operations: sampling, grouping, and convolution [29–31]. Sampling employs iterative farthest point sampling (FPS) to select centroids from the point cloud, ensuring uniform coverage of the 3D space. For a given input point set P , FPS selects a subset P' of centroids that are maximally distant from each other. This is achieved using the following selection criterion at iteration k :

$$p'_k = \underset{p \in P}{\operatorname{argmax}} d(p, P') \quad (2)$$

Grouping then identifies points within a radius r around each centroid p'_i , forming local point groups G_i as:

$$G_i = \{ p_j \in P \mid \|p_j - p'_i\| \leq r \} \quad (3)$$

These groups define the local regions from which geometric features are extracted.

Convolution is achieved using the PointConv operation, which generalizes convolution to point clouds. For each centroid p'_i and its local neighborhood G_i , the updated feature f'_i is calculated as:

$$f'_i = \sum_{p_j \in G_i} \frac{W(p_j - p'_i) \cdot f_j}{\rho(p_j)} \quad (4)$$

Here, $W(p_j - p'_i)$ is a learnable convolutional kernel defined by a multi-layer perceptron (MLP), and $\rho(p_j)$ adjusts for varying point densities. This approach captures local geometric relations and corrects for non-uniform distributions of points.

By employing these operations twice within the two SABs, NHP-Net progressively downsamples the point cloud, extracting hierarchical features at multiple levels. The final features are aggregated using a global max-pooling layer, and the fully connected (FC) layers predict the 3×3 rotation matrix for NHP reproduction. This hierarchical feature learning enables NHP-Net to capture both local and global geometric features, resulting in superior rotation matrix prediction accuracy. Overall, the model comprises approximately 1.46 million learnable parameters.

We employed the Wing loss (WL) function to handle both small and large errors effectively in network training. WL is particularly well-suited for angle regression tasks because it combines the advantages of both L1 and L2 loss functions, providing robustness to outliers while maintaining sensitivity to minor errors [32]. This characteristic is crucial

for accurately predicting roll, pitch, and yaw angles in natural head position estimation. WL is defined as

$$WL(x, y) = \begin{cases} \omega \ln(1 + |x - y|/\varepsilon) & \text{if } |x - y| < \omega \\ |x - y| - C & \text{otherwise} \end{cases} \quad (5)$$

where x and y are predicted 3×3 rotation matrix and ground truth of the 3×3 rotation matrix, respectively. ω , ε , and C are hyperparameters in WL, which control the smoothness of the loss curve, where C is defined as $\omega - \omega \ln(1 + \omega/\varepsilon)$. In this work, we set $\omega = 10$ and $\varepsilon = 3$ in WL, achieving the best performance.

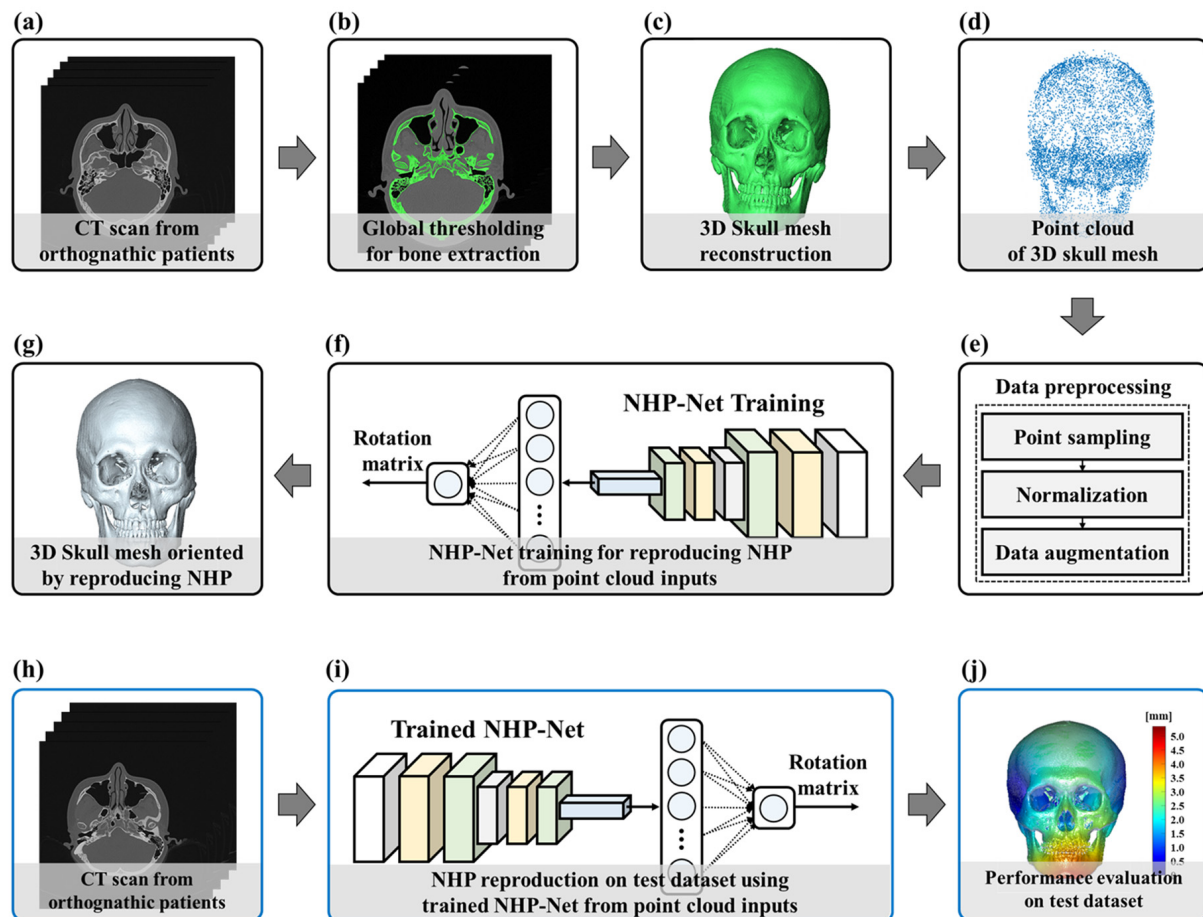


Figure 2. The schematic diagram of our proposed method. (a) CT scans from orthognathic patients on the training dataset. (b) Global thresholding for maxillofacial bone extraction. (c) Three-dimensional skull mesh reconstruction of maxillofacial bone area. (d) Point cloud extraction from 3D skull mesh. (e) Data preprocessing for deep learning. (f) NHP-Net training. (g) Three-dimensional skull mesh oriented by reproducing NHP from NHP-Net. (h) CT scans from orthognathic patients on the test dataset. (i) NHP reproduction on the test dataset using trained NHP-Net. (j) Evaluation process of the NHP-Net.

In the inference stage, the trained NHP-Net is used to estimate the reproducing NHP of CT scans from an orthognathic patient. We uniformly sampled point clouds on the input skull mesh and normalized them. Then, NHP-Net predicted the 3×3 rotation matrix for reproducing NHP. To ensure the orthogonality of the predicted 3×3 rotation matrix, we applied singular value decomposition (SVD) [33,34]. The output matrix R' was decomposed into $U\Sigma V^T$, and the refined rotation matrix R was obtained as $R = UV^T$.

Subsequently, the refined rotation matrix R was applied to the vertices of the skull mesh to align it to the NHP.

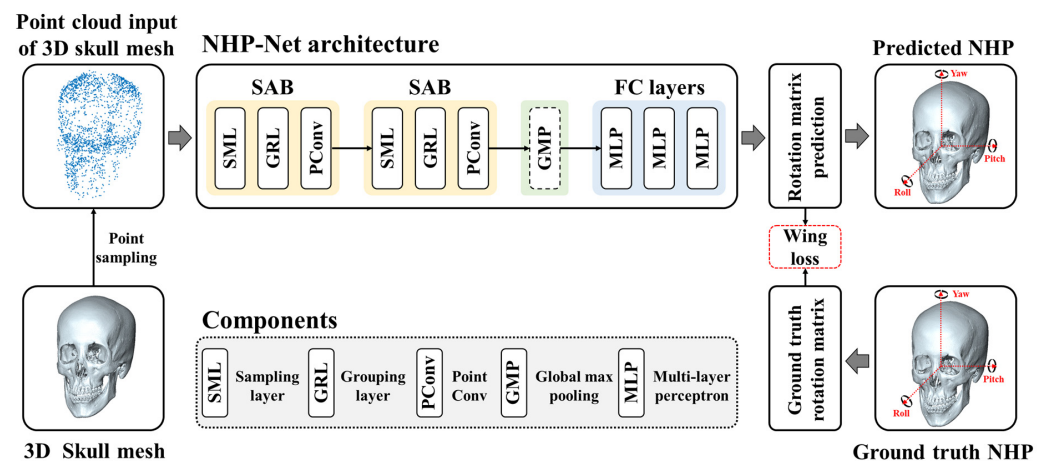


Figure 3. Illustration of our NHP-Net architecture. NHP-Net takes a point cloud input of the 3D skull mesh and regresses a 3×3 rotation matrix for reproducing NHP.

We used the Adam optimizer with an initial learning rate of 0.0001, where the learning rate was decreased to half of it every 50 epochs. The model was trained for 500 epochs with a batch size of 1 using a single NVIDIA Titan RTX GPU with 24 GB (NVIDIA Corporation, Santa Clara, CA, USA). We implemented NHP-Net architecture using Python3 and PyTorch 1.9.0. To enhance the robustness and generalizability of NHP-Net, we applied data augmentation techniques to the training dataset. Specifically, we increased the number of the training set by 30 times. Each augmented sample was created by randomly varying the original NHP roll, pitch, and yaw angles within $\pm 50\%$ of their original values. We chose this variation range to introduce significant diversity while ensuring the altered angles remained within physiologically plausible limits. By maintaining the average roll, pitch, and yaw angles of the data augmented dataset equal to those of the original data, we preserved the overall statistical properties, preventing any bias in the model training. This approach generated a total of 2700 training samples from the original 90 samples after data augmentation. Figure 4 illustrates augmented samples.

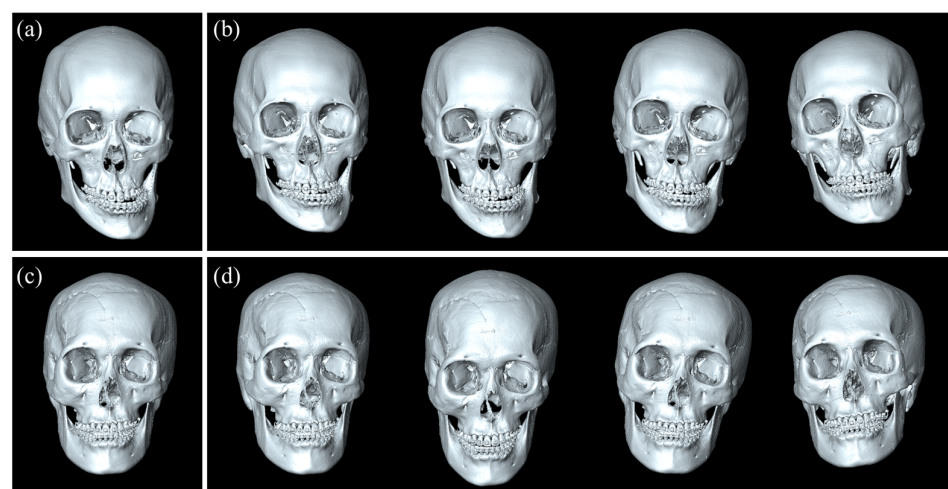


Figure 4. Illustrations of the data augmentation samples. The original 3D skull meshes (a,c) are augmented by randomly varying the roll, pitch, and yaw angles within $\pm 50\%$ of their original values, resulting in multiple augmented samples (b,d) used for training NHP-Net.

2.3. Evaluation Metrics

To evaluate the performance of deep learning models, we employed mean absolute error (MAE) and rotation error (RE), which were commonly used in regression tasks involving angular measurements. MAE assesses the average magnitude of the errors in predicting the individual roll, pitch, and yaw angles compared to the ground truth. For each angle, MAE is calculated as

$$\text{MAE}_{\text{angle}} = \frac{1}{n} \sum_{i=1}^n |\hat{\theta}_i - \theta_i| \quad (6)$$

where $\hat{\theta}_i$ is the predicted value of the angle (roll, pitch, or yaw) for the i -th sample, θ_i is the ground truth value of the same angle for the i -th sample, and n is the number of samples. The MAE is calculated separately for each angle. RE measures the angular difference between the predicted rotation matrix $\hat{R}$ and the ground truth rotation matrix R . It is calculated using the following formula:

$$\text{RE} = \cos^{-1} \left(\frac{\text{Tr}(\hat{R}R^T) - 1}{2} \right) \quad (7)$$

where $\text{Tr}(\cdot)$ denotes the trace of a matrix. To determine the statistical significance of our results, we performed paired t -tests comparing the MAE for each angle value of NHP-Net with those of other deep learning models, specifically PointNet, DGCNN, and Point Cloud Transformer. A p -value less than 0.05 was considered statistically significant. All statistical analyses were conducted using SPSS Statistics ver. 21 (SPSS Inc., Chicago, IL, USA).

3. Results

Table 1 shows the performance of NHP-Net using WL ($\omega = 10$ and $\varepsilon = 3$) with varying numbers of point cloud points (512, 1024, 2048, and 4096). As the number of points increased, the MAEs for roll, pitch, yaw, and RE generally decreased. Notably, at 2048 points, the RE was the lowest at $1.918^\circ \pm 1.099^\circ$, with MAE for roll at $0.665^\circ \pm 0.530^\circ$ and MAE for yaw at $0.810^\circ \pm 0.786^\circ$, being the lowest among all configurations. Although the MAE for pitch at 2048 points ($1.246^\circ \pm 1.155^\circ$) was slightly higher than at 1024 points ($1.142^\circ \pm 0.983^\circ$), the overall performance at 2048 points was optimal due to the lowest RE and MAEs for roll and yaw. However, at 4096 points, the RE slightly increased to $2.026^\circ \pm 0.943^\circ$, indicating a slight performance degradation. This suggests that increasing the number of points beyond 2048 does not lead to significant performance improvement and may even cause degradation in some angles. Therefore, 2048 points are considered the optimal number in terms of efficiency and overall performance.

Table 1. Quantitative performance from NHP-Net with different point cloud sizes. Mean absolute error is used to evaluate performance for roll, pitch, and yaw.

The Number of Points	Roll (°)	Pitch (°)	Yaw (°)	Rotation Error (°)
512	1.178 ± 0.905	1.306 ± 1.343	1.861 ± 1.322	3.007 ± 1.372
1024	0.959 ± 0.821	0.917 ± 0.936	1.212 ± 1.144	2.150 ± 1.212
2048	0.665 ± 0.530	1.246 ± 1.155	0.810 ± 0.786	1.918 ± 1.099
4096	0.764 ± 0.572	1.310 ± 0.897	0.940 ± 0.823	2.026 ± 0.943

Values are presented as mean $\pm$ standard deviation.

Table 2 presents the results of a comparison of various loss functions (WL, MAE loss, MSE loss, and Huber loss) in NHP-Net while fixing the number of point cloud points at 2048. When using WL ($\omega = 10$ and $\varepsilon = 3$), MAEs for roll, pitch, and yaw were $0.665^\circ \pm 0.530^\circ$, $1.246^\circ \pm 1.155^\circ$, and $0.810^\circ \pm 0.786^\circ$, respectively, with the lowest RE of $1.918^\circ \pm 1.099^\circ$. Notably, the absolute error for pitch was lowest when using MAE loss ($1.226^\circ \pm 0.962^\circ$), and the absolute error for yaw was lowest when using MSE loss ($0.749^\circ \pm 0.632^\circ$). The REs for MAE, MSE, and Huber losses were $2.033^\circ \pm 1.007^\circ$, $1.970^\circ \pm 0.922^\circ$, and $2.056^\circ \pm 1.023^\circ$, respectively. Although WL showed slightly better performance, the overall differences in performance across loss functions were insignificant.

Table 2. Quantitative performance from NHP-Net with different loss functions. Mean absolute error is used to evaluate performance for roll, pitch, and yaw.

	Roll ($^\circ$)	Pitch ($^\circ$)	Yaw ($^\circ$)	Rotation Error ($^\circ$)
WL ($\omega = 10$ and $\varepsilon = 3$)	0.665 ± 0.530	1.246 ± 1.155	0.810 ± 0.786	1.918 ± 1.099
MAE loss	0.865 ± 0.465	1.226 ± 0.962	0.894 ± 0.978	2.033 ± 1.007
MSE loss	0.796 ± 0.552	1.329 ± 1.033	0.749 ± 0.632	1.970 ± 0.922
Huber loss	0.823 ± 0.623	1.329 ± 1.100	0.786 ± 0.654	2.056 ± 1.023

Values are presented as mean $\pm$ standard deviation.

Table 3 displays the performance of NHP-Net according to various combinations of hyperparameters ω and ε in WL. The experimental results indicated that with the combination of $\omega = 10$ and $\varepsilon = 3.0$ in WL, the MAEs for roll, pitch, and yaw were $0.665^\circ \pm 0.530^\circ$, $1.246^\circ \pm 1.155^\circ$, and $0.810^\circ \pm 0.786^\circ$, respectively, and the RE was the lowest at $1.918^\circ \pm 1.099^\circ$. Overall, this combination provided the lowest RE and superior performance across all angles, leading us to select $\omega = 10$ and $\varepsilon = 3.0$ as the optimal hyperparameters for the WL.

Table 3. Quantitative performance from NHP-Net with different ω and ε hyperparameters in Wing loss. Mean absolute error is used to evaluate performance for roll, pitch, and yaw.

ω	ε	Roll ($^\circ$)	Pitch ($^\circ$)	Yaw ($^\circ$)	Rotation Error ($^\circ$)
5	0.5	0.788 ± 0.651	1.333 ± 1.020	0.782 ± 0.784	2.024 ± 1.118
	1.0	0.925 ± 0.664	1.292 ± 1.087	0.755 ± 0.714	2.038 ± 1.129
	2.0	0.870 ± 0.724	1.342 ± 1.143	0.719 ± 0.745	2.044 ± 1.146
	3.0	0.820 ± 0.600	1.289 ± 1.107	0.750 ± 0.697	1.994 ± 1.051
	4.0	0.855 ± 0.597	1.515 ± 1.015	0.882 ± 0.773	2.254 ± 1.077
10	0.5	0.817 ± 0.579	1.229 ± 0.992	0.836 ± 0.756	1.996 ± 1.047
	1.0	1.051 ± 0.804	1.323 ± 0.906	1.024 ± 1.014	2.328 ± 1.123
	2.0	0.858 ± 0.624	1.201 ± 0.878	0.820 ± 0.654	1.928 ± 0.851
	3.0	0.665 ± 0.530	1.246 ± 1.155	0.810 ± 0.786	1.918 ± 1.099
	4.0	0.808 ± 0.591	1.272 ± 1.040	0.802 ± 0.685	2.014 ± 1.032

Values are presented as mean $\pm$ standard deviation.

Table 4 shows the quantitative performance comparison results from different deep learning models. The number of point cloud points was fixed at 2048, and the loss function was set to WL ($\omega = 10$ and $\varepsilon = 3.0$). NHP-Net achieved MAEs of $0.665^\circ \pm 0.530^\circ$ for roll, $1.246^\circ \pm 1.155^\circ$ for pitch, and $0.810^\circ \pm 0.786^\circ$ for yaw, with the lowest RE of $1.918^\circ \pm 1.099^\circ$. The REs for other models (PointNet, DGCNN, and PTransformer) were $2.432^\circ \pm 1.305^\circ$, $2.295^\circ \pm 1.036^\circ$, and $2.238^\circ \pm 0.855^\circ$, respectively. Statistical significance tests showed that NHP-Net had significantly lower MAEs in roll, pitch, and RE compared to the other deep learning models ($p < 0.05$). However, for yaw, the difference was not statistically significant ($p > 0.05$). Additionally, comparisons of the surface distance errors in Figures 5 and 6 demonstrated that NHP-Net produced results most similar to the actual angles, suggesting that the model has high prediction accuracy. In Figure 7, we visually confirmed that NHP-

Net had a lower median error and a narrower distribution range, indicating stable and consistent performance.

Table 4. Quantitative performance from different deep learning models. Mean absolute error is used to evaluate performance for roll, pitch, and yaw.

	Roll (°)	Pitch (°)	Yaw (°)	Rotation Error (°)
NHP-Net	0.665 ± 0.530	1.246 ± 1.155	0.810 ± 0.786	1.918 ± 1.099
PointNet	0.825 ± 0.583 *	1.875 ± 1.480 *	0.715 ± 0.677	2.432 ± 1.305 *
DGCNN	0.934 ± 0.689 †	1.565 ± 1.081 †	0.805 ± 0.679	2.295 ± 1.036 †
PTransformer	0.854 ± 0.613 ‡	1.614 ± 0.793 ‡	0.867 ± 0.792	2.238 ± 0.855 ‡

Values are presented as mean ± standard deviation. * Significant difference between NHP-Net and PointNet ($p < 0.05$). † Significant difference between NHP-Net and DGCNN ($p < 0.05$). ‡ Significant difference between NHP-Net and PTransformer ($p < 0.05$).

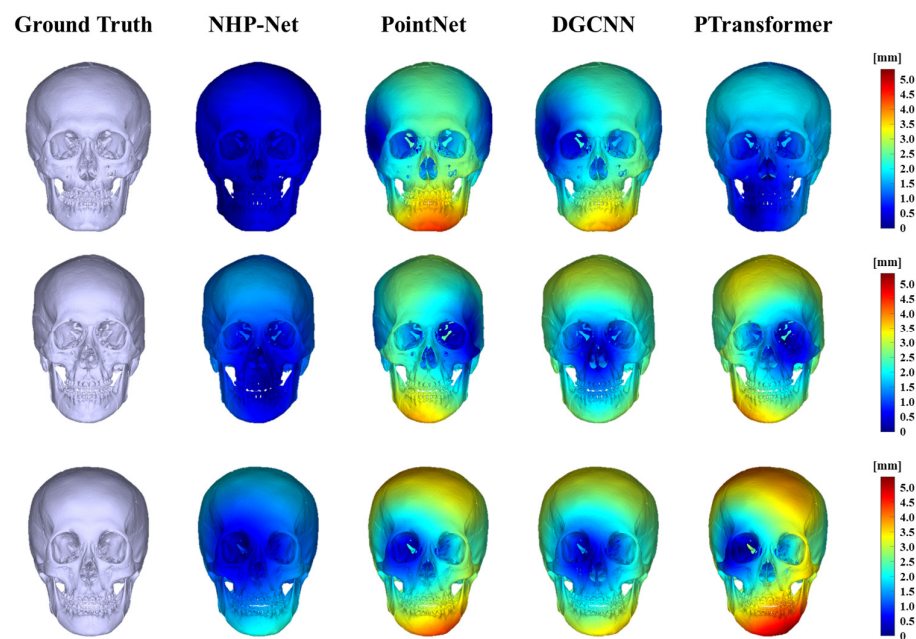


Figure 5. Surface distance errors between 3D skull meshes oriented by ground truth natural head position (NHP) and NHP reproduced by different deep learning models in the frontal view. Distributions of surface distance errors are visualized with a color map, where blue indicates low surface distance error and red indicates high surface distance. NHP-Net demonstrates the most accurate alignment with the reference, as shown by the lower surface distance errors than other deep learning models.

To evaluate potential model bias regarding gender, we performed a subgroup analysis. For male patients ($n = 9$), the average MAEs were $0.694^\circ \pm 0.422^\circ$ (roll), $1.093^\circ \pm 0.788^\circ$ (pitch), and $0.906^\circ \pm 0.854^\circ$ (yaw), with an RE of $2.083^\circ \pm 1.090^\circ$. For female patients ($n = 21$), the average MAEs were $0.653^\circ \pm 0.570^\circ$ (roll), $1.312^\circ \pm 1.275^\circ$ (pitch), and $0.769^\circ \pm 0.751^\circ$ (yaw), with an RE of $1.847^\circ \pm 1.095^\circ$. Independent t -tests showed no significant differences across genders ($p > 0.05$ for all metrics), indicating that the model's performance is consistent regardless of patient gender.

Figure 8 illustrates the progression of the Wing loss during training and validation, demonstrating a significant reduction in both losses over the epochs. The training loss decreased from 0.15925 to 0.01563, and the validation loss decreased from 0.31455 to 0.04625, indicating stable convergence and effective learning of the NHP-Net model.

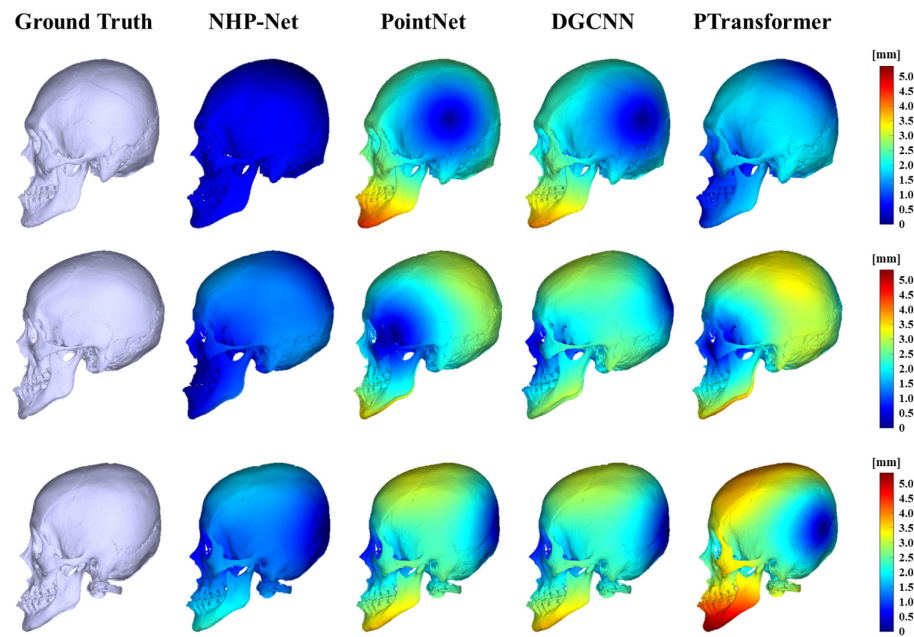


Figure 6. Surface distance errors between 3D skull meshes oriented by ground truth natural head position (NHP) and NHP reproduced by different deep learning models in the side view. Distributions of surface distance errors are visualized with a color map, where blue indicates low surface distance error and red indicates high surface distance. NHP-Net demonstrates the most accurate alignment with the reference, as shown by the lower surface distance errors than other deep learning models.

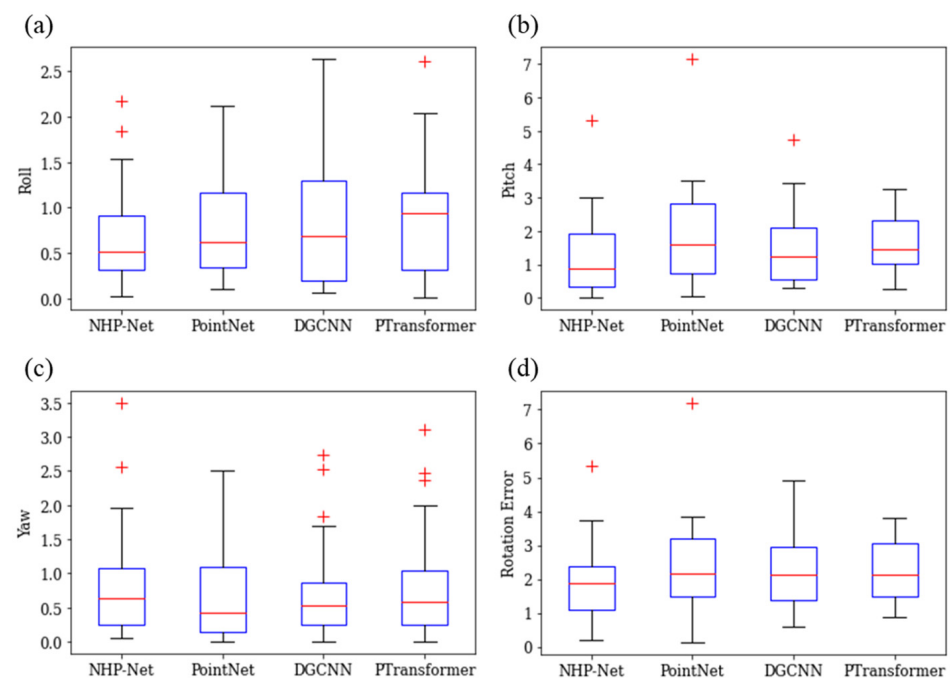


Figure 7. Boxplots of mean absolute error performance results for (a) roll, (b) pitch, (c) yaw, and (d) rotation error from different deep learning models. Each box represents the first and third quartiles of the data, with the median indicated by a red line inside the box. Whiskers extend above and below each box to ± 1.5 times the interquartile range (IQR), and outliers beyond 1.5 IQR are visualized as red '+' marks.

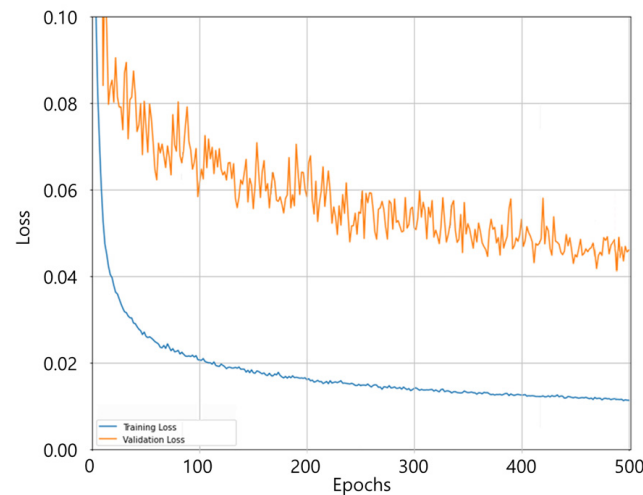


Figure 8. Wing loss during the training of the NHP-Net model with a batch size of 1. The training loss decreased from 0.15925 to 0.01563, and the validation loss decreased from 0.31455 to 0.04625 over the training epochs.

4. Discussion

Traditional techniques often suffer from reproducibility issues due to individual variations and the difficulty for patients to maintain a consistent natural posture [9]. Additionally, methods requiring external devices or markers can be uncomfortable and may alter natural facial expressions, leading to potential inaccuracies in surgical planning [19–21]. With the advent of deep learning and its success in head pose estimation in other fields, we introduced a deep learning approach utilizing geometric representations from point clouds of skull meshes to predict a rotation matrix of NHP from a CT scan. In this study, we aimed to develop a geometric deep learning network for reproducing NHP in CT scans from orthognathic patients. To the best of our knowledge, this is the first study to utilize deep learning techniques for reproducing NHP directly from CT scans. Our results demonstrate that NHP-Net outperformed other state-of-the-art deep learning models, such as PointNet, DGCNN, and Point Cloud Transformer, in terms of RE and MAEs for roll and pitch angles. Specifically, NHP-Net achieved the lowest RE of $1.918^\circ \pm 1.099^\circ$; significantly lower MAEs in roll and pitch; and a significantly lower RE compared to the other models ($p < 0.05$). The superior performance of NHP-Net over PointNet, DGCNN, and Point Cloud Transformer stems from key improvements in design and training. First, NHP-Net employs two set abstraction blocks (SABs) to downsample the point cloud while capturing hierarchical features. Unlike DGCNN's edge connections and PointNet's independent point processing, SABs enable NHP-Net to model both local and global geometric features crucial for NHP prediction. Compared to the self-attention of Point Cloud Transformer, SABs explicitly capture local geometric relationships, better representing the skull's detailed curvature. Second, the WL function enhances robustness by balancing sensitivity to small errors and resistance to large errors, outperforming MSE and MAE loss. This approach improves precision and robustness in NHP estimation, particularly in roll and pitch angles, which are critical for surgical planning. We employed two SAB blocks as this configuration balanced complexity and performance. Preliminary experiments indicated that increasing the number of SAB blocks did not yield meaningful improvements in accuracy, while potentially increasing computational cost and the risk of overfitting (Table 5).

Table 5. Performance and training time comparison using different numbers of SAB blocks. Performance is evaluated using mean absolute error (MAE) for roll, pitch, yaw, and rotation error. Training time is reported per epoch.

Number of SAB	Roll (°)	Pitch (°)	Yaw (°)	Rotation Error (°)	Training TIME (min/epoch)
1	0.712 ± 0.610	1.342 ± 1.200	0.845 ± 0.790	2.145 ± 1.130	5.2
2	0.665 ± 0.530	1.246 ± 1.155	0.810 ± 0.786	1.918 ± 1.099	6.0
3	0.672 ± 0.520	1.230 ± 1.140	0.806 ± 0.780	1.913 ± 1.095	8.5

Values are presented as mean ± standard deviation.

Despite experimenting with various loss functions, including MAE loss, MSE loss, Huber loss, and WL, our results showed that the overall performance differences among them were insignificant (Table 2). This minimal variance could be attributed to the robustness of the network architecture, which effectively captures the features necessary for NHP estimation regardless of the specific loss function used. The similarity in performance suggests that the model's capacity to learn the underlying data distribution and the rotational relationships is strong enough that the choice of loss function has a marginal impact on the final outcome. We chose WL as the loss function for several reasons. First, WL demonstrated the lowest RE among all tested loss functions, indicating a slight advantage in overall performance. Second, WL is specifically designed to handle both small and large errors effectively by combining the benefits of L1 and L2 losses [32]. This property is particularly beneficial in regression tasks involving angles, where outliers or extreme values can adversely affect model training. By using WL, the model becomes more robust to such outliers, potentially improving its generalization to unseen data. Moreover, the asymptotic nature of WL near zero error makes the model more sensitive to small deviations, which is crucial for precise NHP reproduction. In medical applications, even minor errors can have significant clinical implications. Therefore, opting for a loss function that prioritizes accuracy in small error regions aligns with the goal of achieving high precision in surgical planning. By employing 2048 point clouds and optimizing the hyperparameters of the WL function ($\omega = 10$ and $\varepsilon = 3.0$), we balanced computational efficiency with model accuracy. The superior performance of NHP-Net can also be attributed to its ability to effectively capture local and global features of point clouds of 3D skull meshes using geometric deep learning.

Compared to traditional methods for NHP reproduction, which often require additional equipment or manual intervention [21,23], our approach offers a fully automatic solution that operates solely on CT data. This reduces the time and cost associated with NHP determination and minimizes patient discomfort and potential errors arising from manual procedures. Unlike conventional methods requiring manual adjustments, fiducial markers, or specialized devices, NHP-Net operates solely on standard CT data, eliminating the need for external equipment. By automating NHP determination, NHP-Net reduces preoperative planning time and clinician intervention. For instance, conventional methods requiring fiducial markers or bite jigs can add 10–20 min per patient, whereas NHP-Net performs the process automatically after CT acquisition. This not only improves patient comfort but also minimizes procedural costs and operational complexity. Additionally, NHP-Net can be seamlessly integrated into existing hospital workflows, utilizing standard workstations with GPUs, thereby avoiding the need for substantial infra-structure changes. Our findings align with recent advancements in deep-learning-based head pose estimation in computer vision [24–26], extending their applicability to the medical field. We expect that NHP-Net will help to improve the workflow in orthognathic surgery.

Our study had several limitations. First, the CT dataset used for deep learning was relatively small, which may limit the generalizability of the model. To address this, future research will focus on assembling larger, more diverse datasets from multiple institutions, including hospitals from different regions and patient populations, to improve the model's robustness and applicability. Second, the ground truth NHP was determined by a single oral and maxillofacial surgeon. To reduce observer bias, future studies will incorporate annotations from multiple surgeons using consensus-based methods, such as inter-observer agreement analysis and majority voting. Beyond orthognathic surgery, NHP-Net has potential applications in other craniofacial procedures, such as maxillofacial trauma and cranial surgeries, where accurate orientation is crucial. Prospective clinical trials with heterogeneous datasets will further validate and refine the model, ultimately facilitating its adoption in diverse clinical environments. Third, our approach involves converting the 3D skull mesh into a point cloud. This conversion standardizes the input format and ensures consistent data dimensions but may result in the loss of some geometric details inherent in the raw mesh [35]. Future work could explore hybrid approaches that retain mesh connectivity information while leveraging the advantages of point cloud representations. Last, while NHP-Net showed significantly lower errors in roll, pitch, and RE compared to the other models, the improvement in yaw was not statistically significant ($p > 0.05$). NHP-Net may require further fine-tunings to enhance reproduction accuracy in yaw angles.

5. Conclusions

In this study, we introduced NHP-Net, a geometric deep learning network designed to automatically reproduce NHP in CT scans from orthognathic patients. By leveraging SAB, GMP, and FC layers and optimizing hyperparameters of the WL function, NHP-Net effectively captured both local and global geometric features of point clouds of 3D skull meshes. Our experimental results demonstrate that NHP-Net outperformed other state-of-the-art deep learning models, achieving the lowest RE of $1.918^\circ \pm 1.099^\circ$ and significantly lower MAEs in roll and pitch angles ($p < 0.05$). We expected that our proposed NHP-Net could be a significant advancement in automatic NHP reproduction from CT scans to reduce the workload of surgeons for orthognathic surgery planning and improve surgical precision and patient outcomes.

Author Contributions: Conceptualization, J.-Y.Y., S.Y., S.-H.L., J.Y.H., J.-E.K., K.-H.H., S.-S.L., M.-S.H., H.J.Y. and W.-J.Y.; methodology, J.-Y.Y., S.Y. and W.-J.Y.; software, J.-Y.Y., S.Y. and J.Y.H.; validation, J.-Y.Y., S.Y., H.J.Y. and W.-J.Y.; formal analysis, J.-Y.Y., S.Y., J.-M.K., H.J.Y. and W.-J.Y.; investigation, J.-Y.Y., S.Y., S.-H.L., J.Y.H. and H.J.Y.; resources, W.-J.Y.; data curation, J.-Y.Y., S.Y., S.-H.L., J.Y.H., J.-E.K., K.-H.H., S.-S.L., M.-S.H. and H.J.Y.; writing—original draft preparation, J.-Y.Y., S.Y., H.J.Y. and W.-J.Y.; writing—review and editing, J.-Y.Y., S.Y., J.-M.K. and W.-J.Y.; visualization, J.-Y.Y. and S.Y.; supervision, W.-J.Y.; project administration, W.-J.Y.; funding acquisition, W.-J.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government (MSIT) (no. 2023R1A2C200532611). This work was supported by a Korea Medical Device Development Fund Grant by the Korean government (Ministry of Science and ICT; Ministry of Trade, Industry, and Energy; Ministry of Health and Welfare; and Ministry of Food and Drug Safety) (Project Number: 1711194231, KMDF_PR_20200901_0011).

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (IRB) of Seoul National University Dental Hospital (CRI18001) on 30 May 2018.

Informed Consent Statement: Informed consent was waived under the approval of the Institutional Review Board (IRB) of Seoul National University Dental Hospital (CRI18001), in accordance with

Article 16, Paragraph 3 of the Bioethics and Safety Act of South Korea. The study met all required conditions, including the impracticality of obtaining consent and the minimal risk to research subjects.

Data Availability Statement: The datasets generated and/or analyzed during the current study are not publicly available due to the restriction by the Institutional Review Board (IRB) of Seoul National University Dental Hospital in order to protect patient's privacy but are available from the corresponding author on reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Xia, J.J.; McGrory, J.K.; Gateno, J.; Teichgraber, J.F.; Dawson, B.C.; Kennedy, K.A.; Lasky, R.E.; English, J.D.; Kau, C.H.; McGrory, K.R. A new method to orient 3-dimensional computed tomography models to the natural head position: A clinical feasibility study. *J. Oral Maxillofac. Surg.* **2011**, *69*, 584–591. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Moorrees, C.F.; Kean, M.R. Natural head position, a basic consideration in the interpretation of cephalometric radiographs. *Am. J. Phys. Anthropol.* **1958**, *16*, 213–234. [\[CrossRef\]](#)
3. Foster, T.; Howat, A.; Naish, P. Variation in cephalometric reference lines. *Br. J. Orthod.* **1981**, *8*, 183–187. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Arnett, G.W.; Bergman, R.T. Facial keys to orthodontic diagnosis and treatment planning. Part I. *Am. J. Orthod. Dentofac. Orthop.* **1993**, *103*, 299–312. [\[CrossRef\]](#)
5. Moorrees, C.F. Natural head position—A revival. *Am. J. Orthod. Dentofac. Orthop.* **1994**, *105*, 512–513. [\[CrossRef\]](#)
6. Lundström, A.; Lundström, F.; Lebre, L.; Moorrees, C. Natural head position and natural head orientation: Basic considerations in cephalometric analysis and research. *Eur. J. Orthod.* **1995**, *17*, 111–120. [\[CrossRef\]](#)
7. Dorafshar, A.H.; Brazio, P.S.; Munding, G.S.; Mohan, R.; Brown, E.N.; Rodriguez, E.D. Found in space: Computer-assisted orthognathic alignment of a total face allograft in six degrees of freedom. *J. Oral Maxillofac. Surg.* **2014**, *72*, 1788–1800. [\[CrossRef\]](#)
8. Welch, M.B.; Wahr, J.; Crowley, M.; Patient Positioning for Surgery and Anesthesia in Adults. UpToDate. 2018. Available online: <https://medilib.ir/upupdate/show/94593> (accessed on 5 November 2024).
9. Cooke, M.S.; Orth, D.; Wei, S.H. A summary five factor cephalometric analysis based on natural head posture and the true horizontal. *Am. J. Orthod. Dentofac. Orthop.* **1988**, *93*, 213–223. [\[CrossRef\]](#)
10. Solow, B.; Tallgren, A. Natural head position in standing subjects. *Acta Odontol. Scand.* **1971**, *29*, 591–607. [\[CrossRef\]](#)
11. Lundström, F.; Lundström, A. Natural head position as a basis for cephalometric analysis. *Am. J. Orthod. Dentofac. Orthop.* **1992**, *101*, 244–247. [\[CrossRef\]](#)
12. Üşümez, S.; Orhan, M. Inclinator method for recording and transferring natural head position in cephalometrics. *Am. J. Orthod. Dentofac. Orthop.* **2001**, *120*, 664–670. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Chen, C.-M.; Lai, S.; Tseng, Y.; Lee, K.-T. Simple technique to achieve a natural head position for cephalography. *Br. J. Oral Maxillofac. Surg.* **2008**, *46*, 677–678. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Weber, D.W.; Fallis, D.W.; Packer, M.D. Three-dimensional reproducibility of natural head position. *Am. J. Orthod. Dentofac. Orthop.* **2013**, *143*, 738–744. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Cooke, M.S.; Wei, S.H. Cephalometric errors: A comparison between repeat measurements and retaken radiographs. *Aust. Dent. J.* **1991**, *36*, 38–43. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Raju, D.S.; Naidu, D.L. Reliability and reproducibility of natural head position: A cephalometric study. *J. Indian Orthod. Soc.* **2012**, *46*, 340–347. [\[CrossRef\]](#)
17. Verma, S.K.; Maheshwari, S.; Gautam, S.N.; Prabhat, K.; Kumar, S. Natural head position: Key position for radiographic and photographic analysis and research of craniofacial complex. *J. Oral Biol. Craniofac. Res.* **2012**, *2*, 46–49. [\[CrossRef\]](#)
18. Serafin, M.; Baldini, B.; Cabitza, F.; Carrafiello, G.; Baselli, G.; Del Fabbro, M.; Sforza, C.; Caprioglio, A.; Tartaglia, G.M. Accuracy of automated 3D cephalometric landmarks by deep learning algorithms: Systematic review and meta-analysis. *Radiol. Med.* **2023**, *128*, 544–555. [\[CrossRef\]](#)
19. Schatz, E.C.; Xia, J.J.; Gateno, J.; English, J.D.; Teichgraber, J.F.; Garrett, F.A. Development of a technique for recording and transferring natural head position in 3 dimensions. *J. Craniofac. Surg.* **2010**, *21*, 1452–1455. [\[CrossRef\]](#)
20. Xia, J.J.; Gateno, J.; Teichgraber, J.F. Controversial Issues in Computer-Aided Surgical Planning for Craniomaxillofacial Surgery. In *Three-Dimensional Imaging for Orthodontics and Maxillofacial Surgery*; Blackwell Publishing Ltd.: Hoboken, NJ, USA, 2010; pp. 171–179.
21. Kim, D.-S.; Yang, H.J.; Huh, K.-H.; Lee, S.-S.; Heo, M.-S.; Choi, S.-C.; Hwang, S.J.; Yi, W.-J. Three-dimensional natural head position reproduction using a single facial photograph based on the POSIT method. *J. Cranio-Maxillofac. Surg.* **2014**, *42*, 1315–1321. [\[CrossRef\]](#)
22. Ariz, M.; Bengoechea, J.J.; Villanueva, A.; Cabeza, R. A novel 2D/3D database with automatic face annotation for head tracking and pose estimation. *Comput. Vis. Image Underst.* **2016**, *148*, 201–210. [\[CrossRef\]](#)

23. Choi, M.-H.; Lee, S.-J.; Yang, H.J.; Huh, K.-H.; Lee, S.-S.; Heo, M.-S.; Choi, S.-C.; Hwang, S.J.; Yi, W.-J. Automatic reproduction of natural head position using a portable 3D scanner based on immediate calibration. *Appl. Sci.* **2019**, *10*, 174. [\[CrossRef\]](#)
24. Asperti, A.; Filippini, D. Deep learning for head pose estimation: A survey. *SN Comput. Sci.* **2023**, *4*, 349. [\[CrossRef\]](#)
25. Ghosh, S.; Dhall, A.; Hayat, M.; Knibbe, J.; Ji, Q. Automatic gaze analysis: A survey of deep learning based approaches. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *46*, 61–84. [\[CrossRef\]](#)
26. Zheng, C.; Wu, W.; Chen, C.; Yang, T.; Zhu, S.; Shen, J.; Kehtarnavaz, N.; Shah, M. Deep Learning-based Human Pose Estimation: A Survey. *ACM Comput. Surv.* **2023**, *56*, 11. [\[CrossRef\]](#)
27. Patacchiola, M.; Cangelosi, A. Head pose estimation in the wild using convolutional neural networks and adaptive gradient methods. *Pattern Recognit.* **2017**, *71*, 132–143. [\[CrossRef\]](#)
28. Geng, X.; Qian, X.; Huo, Z.; Zhang, Y. Head pose estimation based on multivariate label distribution. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 1974–1991. [\[CrossRef\]](#)
29. Qi, C.R.; Su, H.; Mo, K.; Guibas, L.J. Pointnet: Deep learning on point sets for 3D classification and segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 652–660.
30. Qi, C.R.; Yi, L.; Su, H.; Guibas, L.J. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Adv. Neural Inf. Process. Syst.* **2017**, *30*. [\[CrossRef\]](#)
31. Wu, W.; Qi, Z.; Li, F.P. Deep convolutional networks on 3D point clouds. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 9613–9622.
32. Feng, Z.-H.; Kittler, J.; Awais, M.; Huber, P.; Wu, X.-J. Wing loss for robust facial landmark localisation with convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 2235–2245.
33. Arun, K.S.; Huang, T.S.; Blostein, S.D. Least-squares fitting of two 3-D point sets. *IEEE Trans. Pattern Anal. Mach. Intell.* **1987**, *PAMI-9*, 698–700. [\[CrossRef\]](#)
34. Umeyama, S. Least-squares estimation of transformation parameters between two point patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **1991**, *13*, 376–380. [\[CrossRef\]](#)
35. Feng, Y.; Feng, Y.; You, H.; Zhao, X.; Gao, Y. Meshnet: Mesh neural network for 3d shape representation. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 29 January–1 February 2019; pp. 8279–8286.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.