engineering proceedings

MDPI

Proceeding Paper

A Stereo-Photogrammetry System for Multi-Plane Facial Image Acquisition and Landmark Identification Using Convolutional Neural Networks †

Nguyen Minh Trieu [1], Nguyen Truong Ngoc Lam [1], Nguyen Thien Bao [1], Le Trinh [2], Xuan Long Gia [3] and Nguyen Truong Thinh [1,*]

[1] Institute of Intelligent and Interactive Technologies, University of Economics Ho Chi Minh City (UEH), Ho Chi Minh City 700000, Vietnam
[2] Institute of Information Technology, Vietnam Academy of Science and Technology, Hanoi 100000, Vietnam
[3] Institute of Mechanics, Vietnam Academy of Science and Technology, Hanoi 100000, Vietnam
* Correspondence: thinhnt@ueh.edu.vn
† Presented at the 2026 International Conference on Energy, Infrastructure and Environmental Research (EIER 2026), Ho Chi Minh City, Vietnam, 15–17 July 2026.

Abstract

Accurate identification of craniofacial landmarks is essential in anthropometry and reconstructive surgery, particularly for nasal reconstruction and surgical planning. Manual measurements are time-consuming and prone to variability, while existing 3D imaging systems are often costly and complex. This study presents a mechanically stabilized stereo-photogrammetry system for standardized multi-plane facial image acquisition combined with a convolutional neural network (CNN) for facial landmark detection. The system captures high-resolution frontal, profile, and nasal base images using a fixed-baseline stereo configuration with controlled alignment and standardized inter-camera spacing. A dataset of 2803 Vietnamese-origin participants was collected, with 25–32 annotated anthropometric landmarks per subject. The system achieved a mean NME of 0.0990, with PCK@0.05 and PCK@0.10 of 69.82% and 86.51%, respectively. These results demonstrate reliable multi-plane landmark localization and strong generalization capability. The proposed integration of stereo-photogrammetry and deep learning provides a cost-effective and scalable solution for automated facial anthropometry and future 3D reconstruction applications.

Keywords: facial landmark; landmark detection; CNN; Stereo photogrammetry; Vietnam case

1. Introduction

Facial landmarks are anatomically defined points on the human face that describe craniofacial morphology and spatial relationships among facial structures [1]. In anthropology, these landmarks form the foundation of morphometric analysis, enabling quantitative comparison of facial proportions across populations, age groups, and genders [2,3]. Beyond anthropological research, precise landmark localization plays a crucial role in clinical practice, particularly in craniofacial diagnosis, treatment planning, and outcome assessment [3,4]. In reconstructive and aesthetic surgery, especially rhinoplasty and auricular reconstruction, accurate identification of nasal, orbital, and perioral reference points is essential for restoring anatomical symmetry and functional harmony. Nasal reconstruction presents a unique challenge because the nose is both structurally complex and centrally positioned on the face [5]. Its morphology is defined by a combination of osseous, cartilaginous, and soft tissue components. Small deviations in nasal bridge length, tip projection, alar width, or

Check for updates

Academic Editor: Le Kim Hung

Published: 25 November 2026

Copyright: © 2026 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.

base symmetry can significantly affect both aesthetics and respiratory function [6]. Similarly, ear reconstruction requires reliable spatial references relative to the facial midline and orbital landmarks to achieve natural positioning and proportion [7]. Consequently, objective and reproducible landmark detection is indispensable for surgical simulation, preoperative planning, postoperative evaluation, and long-term anthropometric research.

Traditionally, craniofacial measurements have been obtained through manual anthropometry using calipers, rulers, and angle-measuring devices [8]. Manual measurement is highly operator-dependent, and inter-observer variability often introduces significant error [9]. Many anatomical landmarks are located in sensitive or difficult-to-access regions, such as the nasal base or infraorbital area. Direct contact measurement in these areas may cause patient discomfort and compromise precision. Finally, manual anthropometry is time-consuming, particularly when large datasets are required for research or population studies [10]. Image-based anthropometry enables indirect measurement on high-resolution photographs, reducing physical contact and improving repeatability. More recently, convolutional neural networks (CNNs) have demonstrated strong performance in automated facial landmark identification [11].

Three-dimensional (3D) imaging systems based on electromagnetic radiation, such as computed tomography (CT) or other radiographic modalities, provide detailed structural information but are associated with high equipment costs and potential health concerns related to ionizing radiation exposure [12]. Non-radiative 3D surface scanners and structured-light systems offer safer alternatives; however, they remain relatively expensive and may suffer from sensitivity to lighting conditions and surface reflectivity [13]. Two-dimensional (2D) photographic methods are more affordable and widely accessible [14]. Nevertheless, single-plane imaging, typically limited to frontal or lateral views, cannot fully capture the complex three-dimensional structure of the nasal region. The nasal base, in particular, is inadequately represented in standard frontal and profile photographs. Accurate assessment of alar width, nostril symmetry, and columellar alignment requires an inferior or basal perspective. Without multi-plane acquisition, critical geometric information is lost, limiting the precision of landmark determination and subsequent morphometric analysis.

Another significant limitation of existing systems is the lack of integrated mechanical stabilization combined with automated detection algorithms. While deep learning models such as region-based convolutional neural networks and one-stage detectors have demonstrated strong performance in general object detection tasks, their accuracy is highly dependent on input image consistency [15]. Variations in head orientation, camera distance, and optical axis alignment can degrade model performance. Many studies focus exclusively on algorithmic optimization without addressing the mechanical and geometric conditions under which images are acquired. This disconnect between hardware design and software intelligence reduces overall system robustness and repeatability. Therefore, there is a need for a unified system that integrates mechanical alignment, standardized multi-plane stereo-photogrammetry, and advanced CNN-based landmark detection. Such integration would ensure geometric normalization at the acquisition stage while leveraging artificial intelligence for automated analysis.

The objective of this study is to design and evaluate a mechanically assisted stereo-photogrammetry system capable of capturing standardized frontal, profile, and nasal base images for automated facial landmark identification using convolutional neural networks. The proposed system aims to combine precise mechanical head positioning, controlled camera geometry, and high-resolution imaging to minimize pose-related distortion and improve measurement repeatability. By integrating multi-plane image acquisition with deep learning-based detection algorithms, the system seeks to provide accurate, efficient, and reproducible landmark localization suitable for anthropological research and reconstructive

surgical applications. Ultimately, this work endeavors to bridge mechanical engineering design and artificial intelligence-driven image analysis to establish a reliable framework for advanced craniofacial measurement and 3D reconstruction.

2. Methodology

2.1. Mechanical Design of the Capture System

The imaging module employs a rotational stereo camera configuration with a fixed baseline distance B. The baseline is defined as the Euclidean distance between the optical centers of the stereo camera pair:

$$B = \|C_L - C_R\| \quad (1)$$

where C_L and C_R represent the left and right camera centers, respectively.

A fixed baseline ensures consistent disparity measurements across subjects and is essential for reliable depth reconstruction in stereo-photogrammetry shown in Figure 1. The camera assembly is mounted on a semi-circular rotational frame, allowing controlled angular repositioning to acquire three standardized views: Frontal plane (0° relative to facial midline); Profile plane (90° lateral rotation); Nasal base plane (inferior orientation with downward optical axis). The rotational mechanism is supported by precision bearings to maintain a constant baseline length during angular displacement. Locking joints with angular scale markings enable repeatable positioning within ±0.5°.

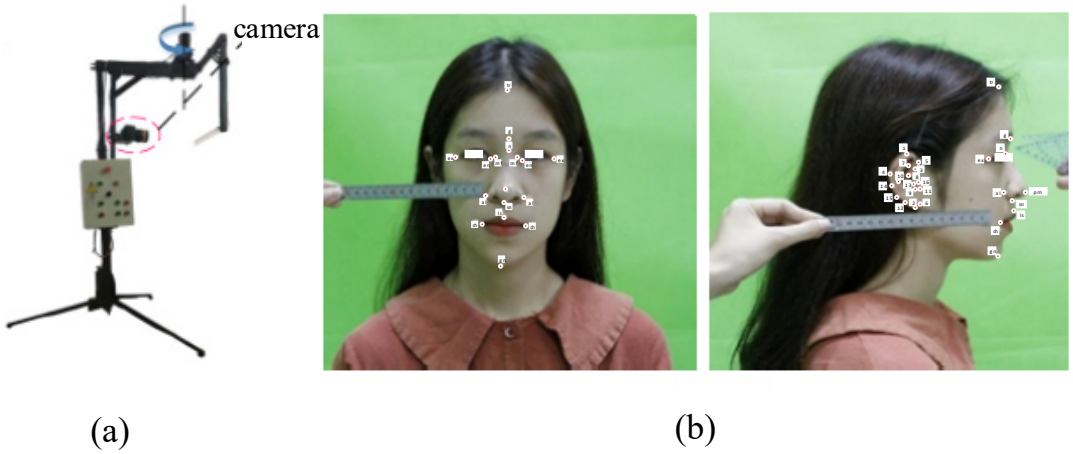


Figure 1. (**a**) Schematic diagram of the rotational stereo camera rig showing fixed baseline B, rotation axis, and three capture orientations. (**b**) Illustration of Frankfort plane alignment.

To eliminate variability due to body movement, participants are seated on a non-reclining chair with rigid back support. The chair height is adjustable vertically to align the participant's Frankfort horizontal plane with the optical axis of the camera. The head position is stabilized using a passive positioning system composed of lateral supports and a posterior headrest. No invasive fixation is used. Accurate craniofacial measurement requires standardization of head orientation. The Frankfort horizontal plane, defined by the line connecting the tragion and orbitale points, is used as the anatomical reference, as shown in Figure 1b. Proper alignment ensures that images are captured without rotational bias in pitch.

Camera-to-subject distance directly influences image scale and perspective distortion. Therefore, fixed distances are enforced for each capture orientation: $D_{FP} = 70$ cm (frontal and profile views); $D_{NB} = 45$ cm (nasal base view). These distances are measured from the optical center to the pronasale reference point. The shorter distance for nasal base

imaging increases spatial resolution for inferior nasal structures. Perspective projection can be modeled as

$$x = \frac{fX}{Z}, y = \frac{fY}{Z} \tag{2}$$

where f is the focal length, (X, Y, Z) are world coordinates, and (x, y) are image coordinates. Maintaining constant $Z \approx D$ across participants ensures consistent magnification:

$$M = \frac{f}{Z} \tag{3}$$

Thus, distance standardization preserves comparable scale factors across the dataset. Mechanical vibrations degrade stereo matching accuracy. The system frame is constructed from aluminum alloy with a high stiffness-to-weight ratio. Rubber isolation feet reduce floor-transmitted oscillations. Exposure time is minimized to further limit motion blur. Rigid mounting prevents micro-displacement between stereo cameras. The relative pose matrix $[R|t]$ must remain constant:

$$P = K[R|t] \tag{4}$$

Any unintended shift alters extrinsic parameters and introduces depth error. Stereo depth Z is estimated via disparity d as

$$Z = \frac{fB}{d} \tag{5}$$

A larger baseline increases depth sensitivity but may introduce occlusion. The selected baseline balances nasal depth resolution with frontal facial coverage. Uniform illumination minimizes shadow-induced feature distortion. A green background is employed to enhance foreground segmentation and improve CNN-based detection. Diffused LED panels positioned symmetrically reduce specular highlights. The system utilizes a 120-megapixel digital camera sensor to capture high-detail facial features. High spatial resolution is critical for identifying small anatomical landmarks such as the subnasale or alare.

The capture workflow follows a sequential, standardized procedure: Frontal stereo acquisition, profile stereo acquisition, and nasal base stereo acquisition. Between captures, only the camera module rotates; the participant remains stationary. This reduces inter-view variability. Images are stored in RAW format and later converted to high-quality JPEG for processing.

Stereo images undergo rectification to align epipolar lines horizontally. Let H_L and H_R be homography matrices applied to the left and right images:

$$I'_L = H_L I_L, I'_R = H_R I_R \tag{6}$$

Rectification ensures corresponding points share identical vertical coordinates, simplifying disparity computation. The complete architecture integrates mechanical alignment, standardized geometry, calibrated stereo imaging, and computational correction. By combining structural engineering principles with photogrammetric calibration, the system ensures consistent multi-plane facial acquisition suitable for CNN-based landmark detection and future 3D reconstruction. The synergy between mechanical stability and computational accuracy forms the foundation for reliable anthropometric analysis and robotic-assisted clinical applications.

2.2. Data Acquisition and Dataset Preparation

In this study, a population-specific dataset is constructed using participants of Vietnamese origin to ensure anthropometric homogeneity and reduce inter-ethnic morphological variability. A dataset of N = 2803 participants was enrolled under standardized acquisition conditions using the mechanically stabilized stereo-photogrammetry system

described previously. For each subject, three synchronized views were acquired, namely frontal, profile (lateral), and nasal base (inferior). Thus, the total number of raw captured images is 8409. In practice, additional augmented samples expanded the effective dataset size during training.

To ensure generalization capability, the dataset was divided into training and testing groups according to age distribution. Let the age variable be denoted by $A = 35.09 \pm 11.56$ years, with mean μA and standard deviation σ_A. The dataset was partitioned into a training set of 80% of subjects and a testing set of 20% of subjects (Figure 2). This stratification ensures that the trained model generalizes across age-related morphological differences such as nasal cartilage elasticity or soft tissue thickness. The probability density of the age distribution can be approximated by a Gaussian model:

$$f(A) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left(-\frac{(A-\mu)^2}{2\sigma^2}\right)} \tag{7}$$

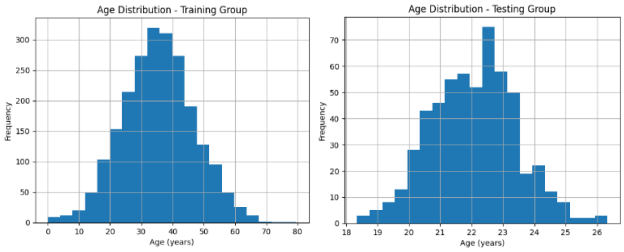


Figure 2. Histogram illustrating age distribution for training and testing groups.

Strict inclusion criteria were applied, such as neutral facial expression, absence of occluding accessories, and correct Frankfort plane alignment. Subjects with excessive head rotation or obscured anatomical regions were excluded. This controlled acquisition protocol minimizes systematic bias in landmark geometry and ensures reproducibility.

Accurate landmark definition is fundamental to anthropometric measurement. A total of 25–32 anatomical landmarks were identified, depending on the viewing perspective. These landmarks represent discrete, reproducible anatomical reference points. Let the landmark set be denoted by

$$L = \{L_1, L_2, \ldots, L_k\}, \quad 25 \leq k \leq 32 \tag{8}$$

For stereo reconstruction, a corresponding 3D coordinate can later be derived:

$$L_i^{3D} = (X_i, Y_i, Z_i) \tag{9}$$

Ground truth annotation was performed manually by trained experts using image visualization software, as shown in Figure 3. Each landmark was labeled independently by two observers using the list of facial landmarks in Table 1. Inter-observer reliability was quantified using the mean Euclidean distance:

$$E_{\text{inter}} = \frac{1}{k}\sum_{i=1}^{k} \sqrt{\left(x_1^{(1)} - x_i^{(i)}\right)^2 + \left(y_1^{(1)} - y_i^{(i)}\right)^2} \tag{10}$$

If E_{inter} exceeded a predefined tolerance threshold, re-evaluation was conducted. This ensures high annotation consistency before CNN training.

Deep neural networks require large and diverse datasets to generalize effectively. To mitigate overfitting and simulate real-world variability, multiple augmentation techniques were applied. Let the original image be $I(x,y)$. The augmented image is denoted by $I'(x,y)$.

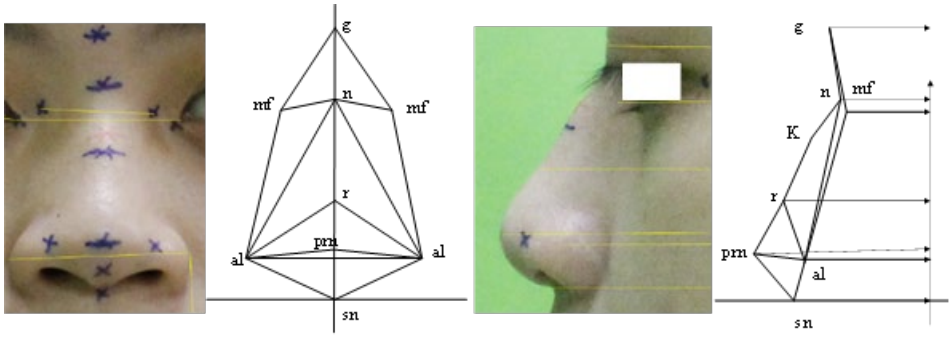


Figure 3. Annotated frontal and profile images showing landmark positions.

Table 1. List of facial landmarks.

Name	Abbreviation	Name	Abbreviation
Trichion	tr	Subnasale	sn
Endocanthion	en	Rhinion	r
Exocanthion	ex	Maxillofrontale	mf
Nasion	n	Cheilion	ch
Pronasale	prn	Pogonion	pg
Alare	al	Gnation	gn

2.3. CNN-Based Landmark Detection Framework

The automatic identification of craniofacial landmarks in multi-plane stereophotogrammetry images is formulated as a supervised object detection problem. Each anatomical landmark is treated as a localized object instance, and the task of the convolutional neural network (CNN) is to predict both the class label and spatial coordinates of bounding boxes corresponding to predefined anthropometric points. Given an input image $I \in \mathbf{R}^{H\times W\times 3}$, the objective of the network is to estimate the spatial coordinates of predefined landmark points belonging to the set L. Each landmark L_i is represented by its 2D image coordinate and optionally by a bounding region when treated under an object-detection paradigm. The CNN model learns a nonlinear mapping

$$f_\theta : I \rightarrow \hat{L} \tag{11}$$

where θ denotes the network parameters and $\hat{L}$ denotes the predicted landmark set.

Facial landmark localization can be approached in two principal ways: direct coordinate regression or object-based detection. In this framework, landmarks are modeled as small, localized regions and detected using a CNN-based object detection architecture. For each landmark class c, the network predicts the following: a confidence score $s_c \in [0, 1]$; bounding-box coordinates $\hat{b}_c = (x_c, y_c, w_c, h_c)$. The detection problem is thus defined as minimizing the discrepancy between predicted bounding boxes $\hat{b}_c$ and ground-truth boxes b_c^*, while simultaneously classifying the landmark type.

The Intersection over Union (IoU) metric is used to evaluate spatial overlap:

$$IOU\left(B_{pred}, B_{gt}\right) = \frac{\left|B_{pred} \cap B_{gt}\right|}{\left|B_{pred} \cup B_{gt}\right|} \tag{12}$$

A detection is considered correct if $IoU \geq \tau$, where τ is typically set to 0.5.

The proposed framework employs a deep CNN, as in Figure 4, composed of three major components: a backbone network for hierarchical feature extraction, feature aggregation layers for multi-scale representation, and a detection head for landmark localization and

classification. The backbone network transforms the input image into progressively abstract feature maps. Let the convolutional operation at layer l be expressed as

$$F_l = \sigma(W_l * F_{l-1} + b_l) \tag{13}$$

where $*$ denotes convolution, W_l are learnable kernels, b_l is bias, and σ is a nonlinear activation function. Pooling layers reduce spatial resolution while increasing receptive field size. The hierarchical representation enables detection of both global facial geometry and fine anatomical structures such as the subnasale or alare.

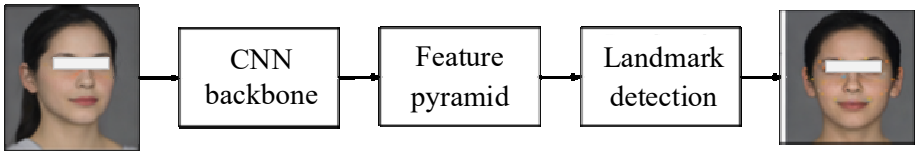


Figure 4. Diagram illustrating CNN backbone, feature pyramid, and landmark detection head.

Facial landmarks vary in scale across different planes (frontal, profile, and nasal base). Multi-scale detection layers allow the model to detect small landmarks (e.g., rhinon, subnasale) and larger ones (e.g., trichion) simultaneously. Feature Pyramid Networks (FPNs) combine low-level spatially precise features with high-level semantic features:

$$P_l = Conv(F_l + Up(P_{l+1})) \tag{14}$$

where $Up(\cdot)$ denotes upsampling.

The detection head outputs landmark predictions for each spatial location. For grid-based detectors, the image is divided into $S \times S$ cells, and each cell predicts B candidate boxes. Bounding-box parameterization follows:

$$\begin{array}{cc} x = \sigma(t_x) + c_x, & y = \sigma\left(t_y\right) + c_y \\ w = p_w e^{t_w}, & h = p_h e^{t_h} \end{array} \tag{15}$$

where (c_x, c_y) is the cell offset and (p_w, p_h) are anchor-box priors. The final landmark coordinate is derived as the center of the predicted bounding box

$$\hat{L}_i = (x_i, y_i) \tag{16}$$

Non-Maximum Suppression (NMS) removes redundant detections by suppressing boxes with high IoU overlap but lower confidence scores.

Training is performed by minimizing a joint loss function:

$$L_{total} = L_{cls} + \lambda L_{reg} \tag{17}$$

where L_{cls} is the classification loss, and L_{reg} is the bounding-box regression loss.

Binary cross-entropy is used for landmark presence prediction:

$$L_{cls} = -[y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})] \tag{18}$$

Here, y denotes the ground truth label and $\hat{y}$ denotes the predicted confidence. To improve localization precision, the IoU loss is incorporated:

$$L_{IoU} = 1 - IoU\left(B_{pred} - B_{gt}\right) \tag{19}$$

This directly optimizes spatial overlap rather than independent coordinate differences. Model performance is evaluated using Precision, Recall, and mean Average Precision (mAP) across all landmark classes. Since the system captures frontal, profile, and nasal base

views, predicted 2D landmarks from each plane are geometrically consistent through stereo triangulation. Given disparity d, focal length f, and baseline B, 3D depth is computed as

$$Z = \frac{fB}{d} \tag{20}$$

This enables the reconstruction of spatial landmark coordinates:

$$(X, Y, Z) = \left(\frac{xZ}{f}, \frac{yZ}{f}, Z\right) \tag{21}$$

Such integration enhances robustness by enforcing anatomical plausibility across views. As illustrated in Figure 5, a process for reconstructing 3D facial models from facial landmarks is proposed as a direction for future research. This approach involves using stereo imagery in conjunction with a convolutional neural network (CNN) to detect 2D landmarks, followed by the application of stereo triangulation to determine their 3D spatial coordinates. These results will serve as the foundation for reconstructing 3D facial geometry, thereby enhancing the capacity to analyze and simulate facial morphological features.

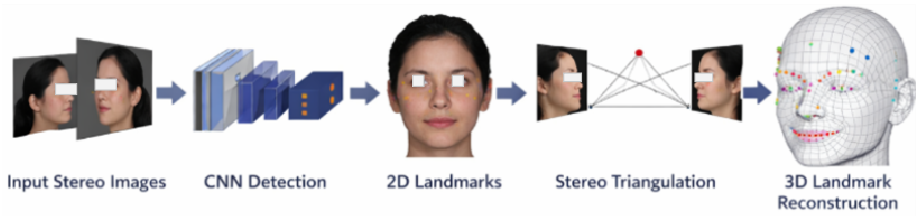


Figure 5. Flowchart of 3D craniofacial landmark reconstruction process.

3. Results and Discussions

The performance of the proposed CNN-based facial landmark detection framework was evaluated using the curated validation and strictly subject-disjoint test sets. All reported quantitative results correspond to the held-out test set, which was not used during network training, hyperparameter tuning, or model selection. This separation ensures an unbiased estimation of generalization capability under the same acquisition protocol and annotation standards.

Landmark localization accuracy was quantified using two complementary metrics: normalized mean error (NME) and percentage of correct keypoints (PCK). The normalized mean error was defined as the Euclidean distance between predicted and ground-truth landmark positions, averaged over all landmarks and test images, and normalized by the inter-alar distance in each image (Figure 6). For a single image with k landmarks, the normalized error is given by

$$NME = \frac{1}{k}\sum_{i=1}^{k} \frac{\sqrt{(\hat{x}_i - x_i)^2 + (\hat{y}_i - y_i)^2}}{d_{al-al}} \tag{22}$$

where $(\hat{x}_i, \hat{y}_i)$ and (x_i, y_i) denote predicted and reference coordinates of landmark i, and d_{al-al} represents the inter-alar distance used as a scale normalization factor. This normalization ensures scale invariance and allows for comparison across subjects with different facial sizes. The percentage of correct keypoints (PCK) measures the fraction of landmarks whose normalized localization error falls below a predefined tolerance threshold τ. Formally,

$$PCK(\tau) = \frac{1}{kN}\sum_{j=1}^{N}\sum_{i=1}^{k} \mathbf{1}\left(\frac{\left\|\hat{L}_{ij} - L_{ij}\right\|_2}{d_{al-al,j}} < \tau\right) \tag{23}$$

where N is the number of test images and $\mathbf{1}(\cdot)$ is the indicator function.

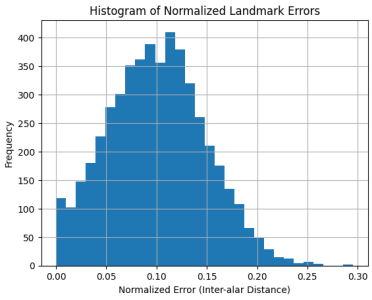
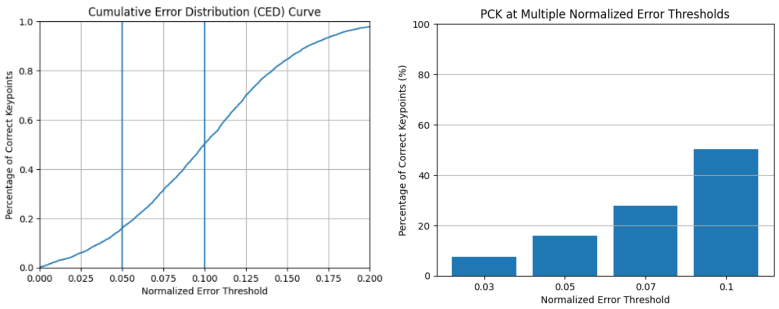


Figure 6. Histogram of normalized landmark errors with normalized error values and frequency.

Thresholds are expressed as fractions of the inter-alar distance to maintain clinical interpretability. These metrics provide complementary perspectives: NME reflects the average magnitude of localization error across all landmarks, while PCK captures the proportion of landmarks that meet clinically meaningful precision thresholds shown in Figure 7a. On the independent test set, the mean inter-alar-normalized NME across all landmarks and images was $NME_{test} = 0.0990 \pm 0.4254$. This result indicates that, on average, predicted landmark positions deviate from the reference positions by approximately 9.9% of the inter-alar distance. Given that the inter-alar distance typically represents a stable nasal width reference, this magnitude of error corresponds to clinically acceptable localization precision for many anthropometric measurements. In terms of threshold-based performance, the model achieved PCK@0.05 = 69.82% and PCK@0.10 = 86.51%, as shown in Figure 7b. These values indicate that nearly 70% of all landmarks were localized within 5% of the inter-alar distance from their ground-truth positions, and more than 86% fell within a 10% tolerance band. The improvement from PCK@0.05 to PCK@0.10 reflects the expected tradeoff between strict and relaxed localization criteria.



(**a**) (**b**)

Figure 7. (**a**) Cumulative error distribution (CED) curve for facial landmark benchmarking. (**b**) Chart of PCK at multiple thresholds (0.03, 0.05, 0.07, 0.10).

The NME suggests stable average performance across the full landmark constellation, including cranial, orbital, nasal, and chin landmarks. Given that the inter-alar distance is anatomically consistent and directly relevant to rhinoplasty and nasal reconstruction, normalization by this metric provides practical interpretability. The relatively high PCK@0.10 value (86.51%) demonstrates robust detection under clinically realistic tolerance levels. For surgical planning scenarios where sub-millimetric precision may not always be required, such performance supports the feasibility of automated anthropometric assistance. It is important to note that performance metrics were computed on a subject-disjoint test set. No identity overlap occurred between training and test subsets. Therefore, the reported results reflect the model's ability to generalize to previously unseen faces captured under the same standardized stereo-photogrammetry protocol. Although the overall NME is low, the reported standard deviation (± 0.4254) indicates variability across landmarks and

images. Errors are typically larger for nasal base landmarks (due to occlusion and inferior perspective), small anatomical structures (e.g., rhinon), and regions with soft tissue variability. Landmark detection and calculation were automatically performed on different views that is shown in Figure 8. Future work may incorporate multi-view geometric consistency constraints or heatmap-based regression to further reduce localization variance.

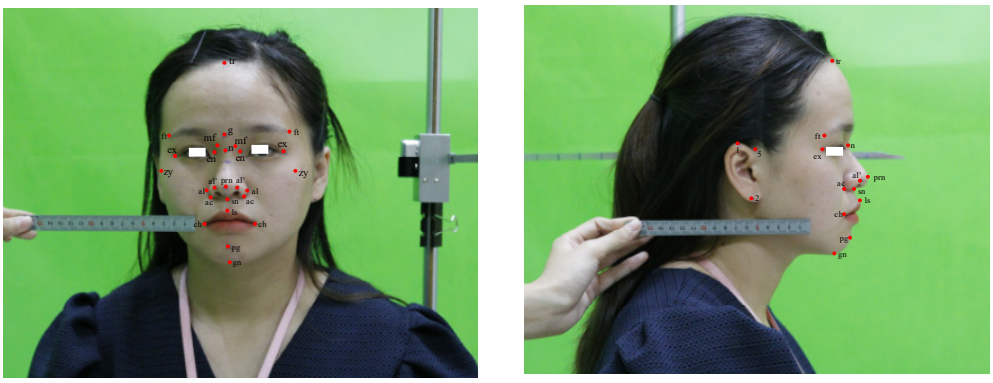


Figure 8. Landmark detection and calculation were automatically performed on different views.

4. Conclusions

This study presents a mechanically stabilized stereo-photogrammetry system for standardized multi-plane facial image acquisition integrated with a CNN model. The proposed system enables reliable identification of anthropometric landmarks across frontal, profile, and nasal base planes. A population-specific dataset of 2803 Vietnamese-origin participants was constructed under standardized conditions, with 25–32 anatomically defined landmarks annotated per subject. The CNN-based detection framework, evaluated using normalized mean error (NME) and percentage of correct keypoints (PCK), demonstrated robust localization accuracy on a strictly subject-disjoint test set. The achieved NME of 0.0990 and PCK@0.10 of 86.51% indicate that the majority of predicted landmarks fall within clinically acceptable tolerance levels when normalized by inter-alar distance. The results confirm that the integration of controlled stereo-photogrammetry acquisition and deep learning–based detection provides a cost-effective and scalable alternative to conventional high-cost 3D scanning systems.

Author Contributions: Conceptualization, N.T.T.; Methodology, N.M.T. and N.T.T.; Software, N.T.B. and N.M.T.; Validation, N.T.N.L. and N.M.T.; Formal analysis, N.T.B. and N.M.T.; Investigation, N.T.N.L. and N.M.T.; Resources, N.M.T.; Data curation, N.T.N.L. and X.L.G.; Writing—original draft, N.T.N.L. and N.M.T.; Writing—review and editing, N.T.T.; Visualization, L.T.; Supervision, N.T.T.; Project administration, N.T.T.; Funding acquisition, N.T.T. All authors have read and agreed to the published version of the manuscript.

Funding: This work belongs to the project "Modeling Artificial Material Implants in Cosmetic Rhinoplasty Based on Facial Landmarks", grant No: B2026-KSA-01 funded by the Vietnam Ministry of Education and Training and hosted by the University of Economics Ho Chi Minh City—UEH, Vietnam.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and was approved by the Biomedical Research Ethics Committee of Pham Ngoc Thach University of Medicine (Approval Code: 66/HĐĐĐ; Approval Date: 10 May 2019).

Informed Consent Statement: Written informed consent was obtained from all participants involved in the study. Written informed consent for the use and publication of identifiable facial images was also obtained from the participants.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: This research is funded by the University of Economics Ho Chi Minh City (UEH) under the Research Group number: TRL-039.

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

References

1. Katina, S.; McNeil, K.; Ayoub, A.; Guilfoyle, B.; Khambay, B.; Siebert, P.; Sukno, F.; Rojas, M.; Vittert, L.; Waddington, J.; et al. The definitions of three-dimensional landmarks on the human face: An interdisciplinary view. *J. Anat.* **2016**, *228*, 355–365. [CrossRef] [PubMed]
2. Jaiswal, S.; Rana, A. Exploring Morphometry of Human Faces: A review of Anthropometric parameters. *Int. J. Sci. Dev. Res.* **2024**, *9*, 288–298.
3. Mbonani, T.M.; L'Abbé, E.N.; Ridel, A.F. Analyzing 3D facial morphology: Insights from a comparative European and South African study on population affinity, sex, age, and allometry. *Forensic Sci. Int.* **2024**, *365*, 112282. [CrossRef] [PubMed]
4. Kim, S.-G. Evaluation and Treatment Plan for Orthognathic Surgery Patients. In *Handbook of Oral and Maxillofacial Surgery and Implantology*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 1–13.
5. Ruse, M.K.; Calhoun, M.; Davis, B.K. Prosthetic nasal reconstruction. *Surgery* **2024**, *1*, 8.
6. Raza, I.G.; Snelling, S.J.; Mimpen, J.Y. Defining the extracellular matrix in non-cartilage soft-tissues in osteoarthritis: A systematic review. *Bone Jt. Res.* **2024**, *13*, 703–715. [CrossRef] [PubMed]
7. Yen, F.-Y.; Lin, Y.-A.; Fang, Q. Improving neuroimaging headgear placement robustness using facial-landmark-guided augmented reality. *Neurophotonics* **2025**, *12*, 045005. [CrossRef] [PubMed]
8. Misăiloaie, A.; Tărăboanță, I.; Stan, C.I.; Budacu, C.C.; Misăiloaie, D.-M.; Sava, A. Comparison of craniofacial anthropometric measurement accuracy of manual technique vs. cone-beam CT scanning. *Diagnostics* **2024**, *14*, 2595. [CrossRef] [PubMed]
9. Mine, K.; Ono, K.; Tanpo, N. The reliability of palpatory examinations for pelvic landmarks to determine pelvic asymmetry: A systematic review and meta-analysis. *Phys. Ther. Rev.* **2022**, *27*, 181–190. [CrossRef]
10. Pal, A.; Patel, T.; Khro, K. A comparative study of the effectiveness of photogrammetric versus manual anthropometric measurements. *Work* **2024**, *78*, 735–746. [CrossRef] [PubMed]
11. Trieu, N.M.; Thinh, N.T. Development of grading system based on machine learning for dragon fruit. In *Regional Conference in Mechanical Manufacturing Engineering*; Springer: Berlin/Heidelberg, Germany, 2021.
12. He, Y.; Song, J.; Li, M.; Sakhatskyi, K.; Li, W.; Feng, X.; Yang, B.; Kovalenko, M.; Wei, H. Perovskite computed tomography imager and three-dimensional reconstruction. *Nat. Photonics* **2024**, *18*, 1052–1058. [CrossRef]
13. Chen, X.; Huang, X.; Zhang, G.; Yang, B.; He, Z.; Xu, W. Validation of single dynamic 3D structured light in precise pedicle puncture navigation. *BMC Surg.* **2025**, *26*, 49. [CrossRef] [PubMed]
14. Chandran, P.; Zoss, G.; Gotardo, P.; Bradley, D. Infinite 3d landmarks: Improving continuous 2d facial landmark detection. In *Computer Graphics Forum*; Wiley Online Library: Hoboken, NJ, USA, 2024.
15. Vu, N.H.; Trieu, N.M.; Tuan, H.N.A.; Khoa, T.D.; Thinh, N.T. Facial anthropometric, landmark extraction, and nasal reconstruction technology. *Appl. Sci.* **2022**, *12*, 9548. [CrossRef]

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.