

ANTHROPOMETRIC IDENTIFICATION SYSTEM USING CONVOLUTION NEURAL NETWORK BASED ON REGION PROPOSAL NETWORK

Ho Nguyen Anh Tuan*, Pham Dang Dieu*,
Nguyen Dao Xuan Hai**, Nguyen Truong Thinh**, Le Gia Vinh***

ABSTRACT.

The operating of the anthropometric identification supported by Convolution Neural Network system is to locate precious anthropology spots and certain distances between each feature area on a person's face. Identifies the anthropometric points from 2D captured pictures by 3 perspective views promised an implementation between medical diagnostics to solve the problem of data retrieval time and efficiency compared to other manual measures.

Key words: *Anthropometric, human face, identification system*

I. INTRODUCTION

In the 18th century and early 19th century, Johann Sigismund Elsholtz was the first man developed the technique of anthropometry [1-3]. Elsholtz's technique was widely used for human growth studies and classification of populations based on quantitative morphology. Especially, the development of anthropological instruments such as spreading and sliding calipers that allowed more comprehensive body measurements. Throughout the last century, anthropometry has witnessed as an extensive development,

computed aid with deep-learning for location and automatic detection landmarks on nasal anthropology are not known about this method in South East Asian countries.

Being dissatisfied with the limitations of visual examination, Farkas began exploring the application of classical anthropometric techniques for quantitative evaluation of faces before and after surgery. His collaboration with anthropologist Karel Hajnis paved the way to develop a facial measurement scheme for patients with congenital anomalies and traumatic facial deformities. Leslie Farkas can be considered to be the pioneer of modern craniofacial anthropometry having published 241 scientific works including four seminal books about anthropometry [4-6].

Generally, anthropometric measurements are made directly from the object by spreading and sliding calipers (**Fig. 1**). After measuring the necessary parameters; calculating the size, length and angle of each part on the human face will be computed based on those direct measurements. To mark the anthropometric points and measure them manually will take a lot of time, but the accuracy of the results is not really high, especially in the eye area and nasal cavity. In addition, the selection of each measuring equipment which fits for each area on the surface of the eyes, ears, nose is difficult and requires a certain level of expertise and experience. Moreover, using the same measuring equipment for many sensitive areas such as eyes, nose, mouth is a risk of

* *University of Medicine Pham Ngoc Thach, Ho Chi Minh City*

** *Ho Chi Minh City University of Technology and Education*

*** *Vietnam Medical Association*

Responsible person: Ho Nguyen Anh Tuan

Email: drhotuan.pnt@gmail.com

Date of receipt: 07/7/2021

Date of scientific judgment: 05/8/2021

Reviewed date: 25/8/2021

infection with unwanted pathogens, thus, each time the measurement is completed that have to conduct disinfection for the measuring equipment. This problem is wasting a lot of time and cost of hospitals or research institutes in over the world, especially in the developing countries. Therefore, the anthropometric identification system can overcome the disadvantages of manual methods, with increased accuracy, speeding up the data collection process and helping to reduce the infectiousness and number of tools used to direct contact with patient.

II. MATERIALS AND METHODS

The system's primary goal is to identify anthropometric points on the human face. As a result, our patient visual data collection system consists of the camera and the frame. Various frame designs have been suggested, but the authors have opted to use the version in which the patient sits constantly in one position and the camera rotates around the patient at a 180 degree angle to capture adequate visual like **Figure 1**

1 due to the priority provided to improve precision and minimize shooting time. The collection system then transmits the patient's images to the image processing center and the processor will begin to use the trained data in combination with the software to process the patient's photos and output the results. Data used in this Study consists of 3 images with three different shooting angles including main picture shooting directions, horizontal angle and basal image. The male and female sex ratios in the collected photographs are 32% and 68%, respectively. The sampled people's average age ranges from 19 to 57, and the total number of pictures after the sampling process is 2200. The next step is to label the data for machine learning after completing the data collection process from a variety of locations and ages. Since marked 27 anthropometric points would be augmented on people's faces in different angles and perspective like **Figure 1**. Although labeling photos took a long time, training for CNN for detecting facial landmarks could give the best visualization.

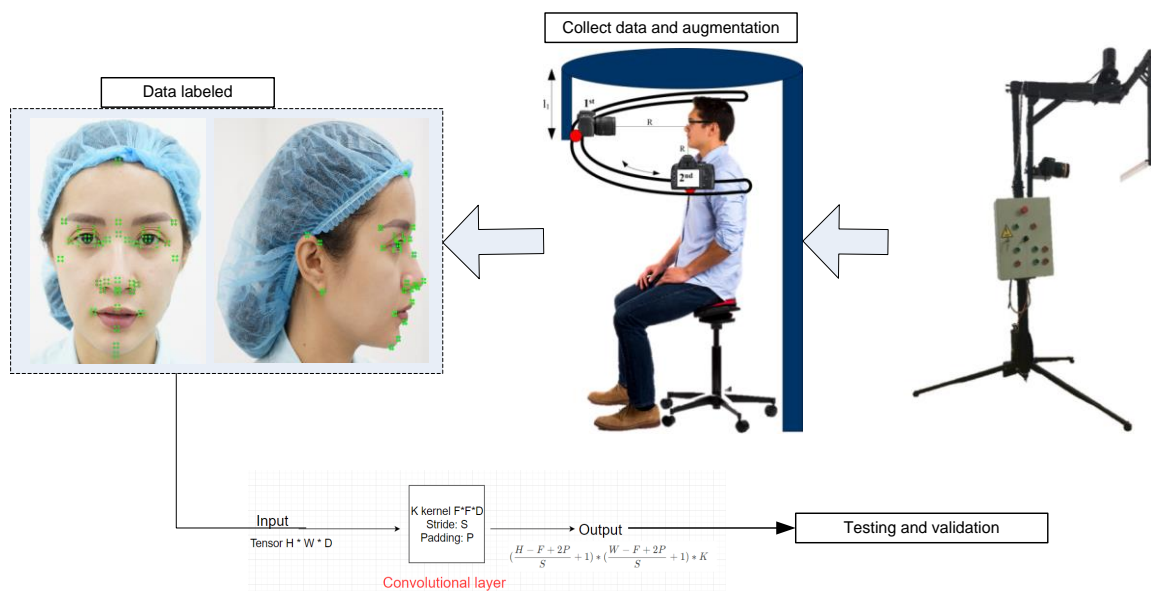


Figure 1. Anthropometric points were labeled in three perspective views of participant.

III. CONVOLUTIONAL NEURAL NETWORK

Convolutional Neural Network (CNN) is one of the models of deep learning. The effect of this algorithm is to create intelligent, responsive systems with high accuracy. The most basic application of this algorithm is layering, that is when the image is fed to the computer, the image will be two-dimensional pixels. The CNN algorithm in

the computer will still recognize the learned image even though the pixels change. The CNN input is divided into 3 dimensions: width, height, and depth. The neurons in the network are not completely linked to the whole of the neurons, but only to a small region. Finally, an output stage is reduced to a vector of the probability value shown as figure 2.

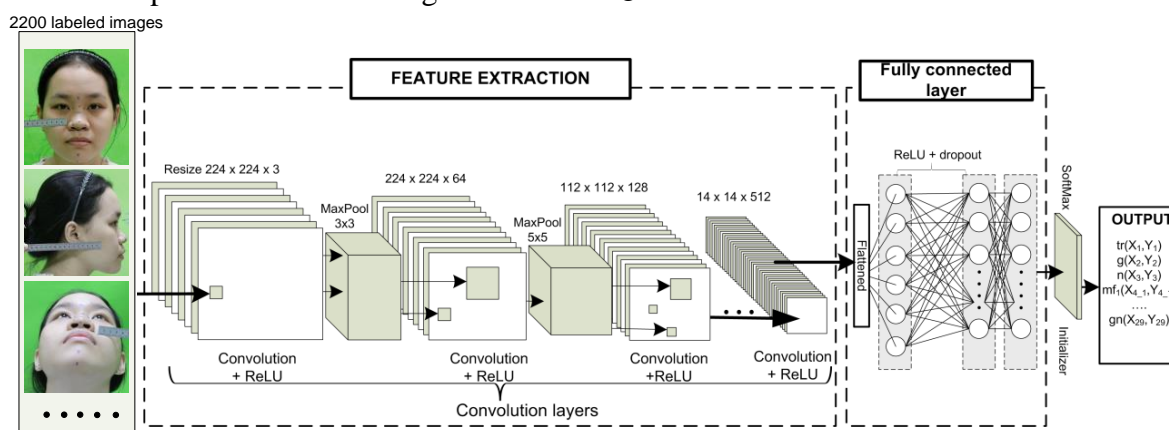


Figure 2. Proposed Convolutional Neural Network (CNN) model for anthropological localization

In this section, we use the YOLOv4 algorithm to perform the detection of human anthropometric features. Images has been provided, taken by an automated camera and processed by hand, which is shown in the section that demonstrates the YOLOv4 model training procedure on Google Colab. The test and evaluation part of the trained model. The YOLOv4 algorithm basically inherits the methods of other YOLO versions, but YOLOv4 applies fast object detection algorithms, optimizes parallel operations to speed up recognition and increase accuracy.

1. Structure of object recognition of YOLOv4

The object recognition structure of YOLOv4 has 3 parts:

- **Backbone:** Backbone is a pre-train model of another transfer learning model to learn features and locations of objects. Transfer learning models are usually VGG16, ResNet-50, ... The transfer learning model applied in YOLOv4 is CSP Darknet53.

- **Head:** The head is used to increase feature specialization to predict the class and bounding-box. In the head part, 1 floor or 2 floors can be applied:

- **Level 1:** Dense Prediction predicts on the whole picture with RPN, YOLO, SSD, ... models.

- **Level 2:** Sparse Prediction predicts with each array predicted objects with R-CNN series models,...

- **Neck:** In the middle of the Backbone and Head, there is usually a Neck section.

Necks are often used to enrich information by combining information between bottom-up and top-down (because some information is too small when going through bottom-up process is lost, so the process top-down cannot reproduce). The models used in the Neck process of YOLOv4 are SPP, PAN.

Anthropometric points were detected using the YOLO model, we have performed the following steps:

Step 1: Search and collect photos of objects that need to be identified, and label each picture.

Step 2: Divide the data into training and test sets.

Step 3: Train the model to use data in the training set.

Step 4: Check and evaluate the model using images captured by the automatic camera.

2. Dataset and Validation

Deep learning model is being trained in 2200 images dataset, which is a prerequisite for us to train the model. Images data is susceptible to external factors (at each time it is different, the image pixel size, image color will be different). So we need a large enough data set for the machine to be able to identify the objects we need to identify. Reality increases accuracy.

The data set is divided into 3 parts:

- Training part: This is the part where we let machine learn the properties of the object to be identified on the image, so this is the biggest part. To create a training set, we need to label the objects, the labeling helps us determine the coordinates of the location of the object to be identified, thereby calculating the loss function and the IoU ratio, updating the weights increase model accuracy.

- Validation part: Used to check the correctness of the model during training, often the validation set will take about 30% of data from the training set to evaluate the effectiveness of the model without participating in the training. weight update. Evaluating the effectiveness of the training session will tell us if we need to adjust the parameters again, whether the network layer is high-low bias, high-low variance or not.

- Test part: This is a very important episode used to evaluate the model after completing the training process. It will show the tolerance based on the actual results we want. For that reason, the test set cannot reuse the training or validation data to avoid overfit or underfit phenomena like as table 1.

Table 1. Model evaluation

Training part error	5%	20%	20%	1.5%
Validation part error	50%	18%	40%	2.25%
Model evaluation	Model has high variance (overfitting)	Model has high bias (underfitting)	Model has high bias and high variance	Model works stably

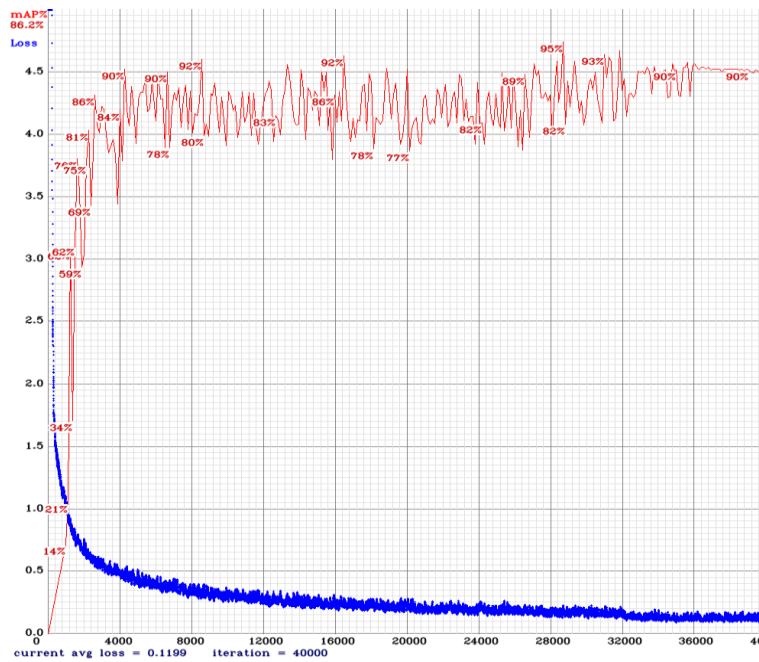


Figure 3. Number of iterations and the average error of the system.

At first, training can take a lot of time shown as figure 3, it had taken about 183 hours continuous training. However, facial landmarks can be easily detected very fast and precious accuracy which is shown in figure 4 and table 2. Detection of facial landmarks of 195 models with trained model have given the fastest as 0.07533s and lowest have been given was 0.25145s.

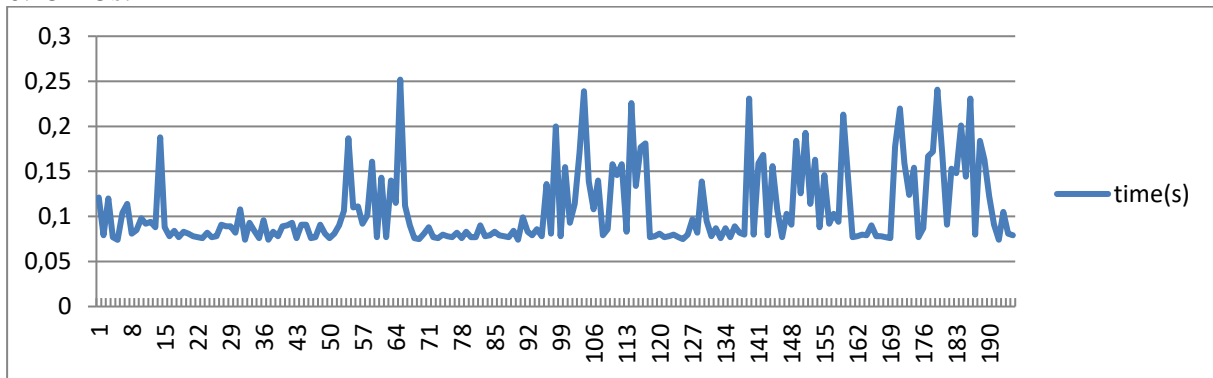


Figure 4. Time period testing with CNN recognition on 195 models

Table 2. Accuracy training for CNN model construction process

	Training	Validation	Testing
Accuracy	98.2258%	98.5326%	97.8693%
Mean absolute error		0.1199%	

When comparing over 182 persons utilizing photographs in three different positions identified by anatomy, with comparison facial landmarks after being included in the system's automatic anatomical landmark points detection, the results are pretty favorable, as seen below figure 5 to 7.



Figure 5. Lateral view image reached 97.8693% of precious accuracy



Figure 6. Frontal view image reached 93.5435% of precious accuracy

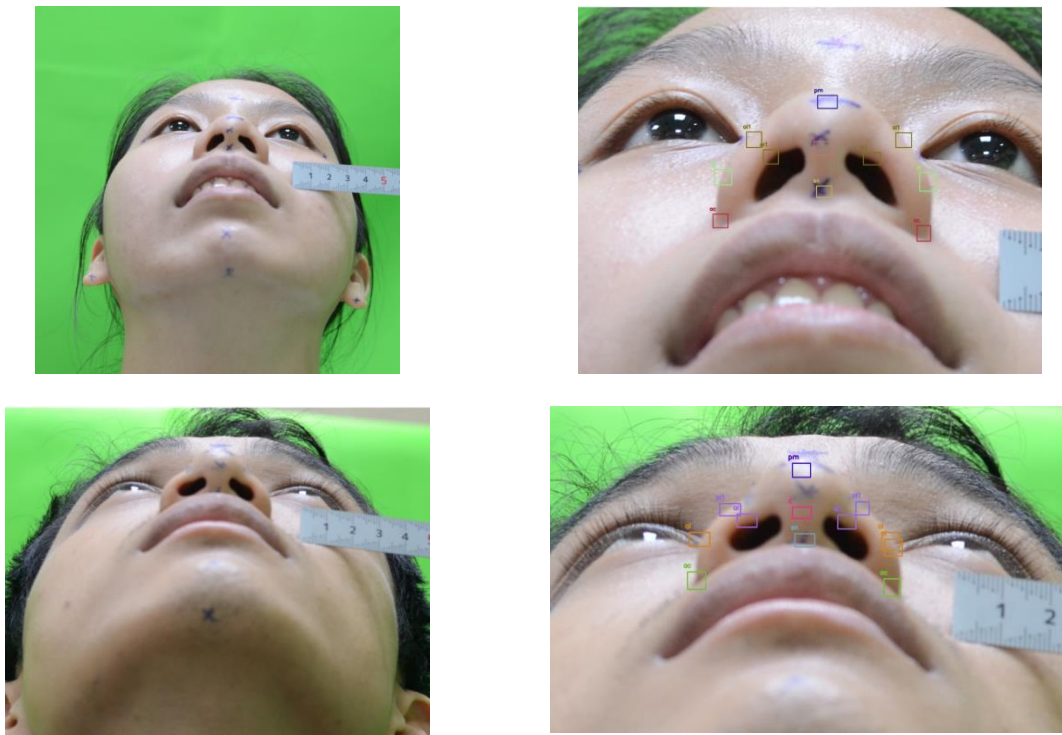


Figure 7. Basal view image reached 95.5645% of precious accuracy

IV. CONCLUSION

In the realm of medicine, a method of measuring based on the neural network model could help detect anthropological on facial landmark points. Not only reducing waste time by processing many pictures at the same time but also giving quick and accurate measurement results in diagnosing medical images according to standard posture and full algorithms. However, the lack of data is inevitable to reduce model errors, which is necessary to improve the accuracy of identification findings by providing more precise and some data is also private data. This will make a significant contribution to the fields of anthropology, aesthetics, and nutrition research.

REFERENCES

1. **S.-J. Ulijaszek and C. G. N. Mascie-Taylor**, *Anthropometry: The Individual and the Population*, Cambridge University Press, Cambridge, UK, 2005.
2. **A. Albrizio**, "Biometry and anthropometry: from Galton to constitutional medicine," *Journal of Anthropological Sciences*, vol. 85, pp. 101-123, 2007.
3. **T. Cuff**, "Biometric method, past, present, and future," in *Historical Anthropometrics*, J. O. Irwin, Ed., pp. 363-375, Ashgate, Aldershot, UK, 1998.
4. **Kolar J C, Salter E M.** *Springfield, IL: CC Thomas; 1997. Craniofacial anthropometry: Practical measurement of the head and face for clinical, surgical, and research use.*