Body Part Detection from Neonatal Thermal Images Using Deep Learning

Fumika Beppu[1(✉)], Hiroki Yoshikawa[1], Akira Uchiyama[1], Teruo Higashino[1], Keisuke Hamada[2,3], and Eiji Hirakawa[4]

[1] Osaka University, Suita, Japan
f-beppu@ist.osaka-u.ac.jp
[2] Nagasaki Harbor Medical Center, Nagasaki, Japan
[3] Nagasaki University, Nagasaki, Japan
[4] Kagoshima City Hospital, Kagoshima, Japan

Abstract. Controlling thermal environment in incubators is essential for premature infants because of the immaturity of neonatal thermoregulation. Currently, medical staff manually adjust the temperature in the incubator based on the neonatal skin temperature measured by a probe. However, the measurement by the probe is unreliable because the probe easily peels off owing to immature skin of the premature infant. To solve this problem, recent advances in infrared sensing enables us to measure the skin temperature without discomfort or stress to the premature infant by using a thermal camera. The key challenge is how to extract skin temperatures of different body parts such as left/right arms, body, head, etc. from the thermal images. In this paper, we propose a method to detect the body parts from the neonatal thermal image by using deep learning. We train YOLOv5 to detect six body parts from thermal images. Since YOLOv5 does not consider relative positions of the body parts, we leverage the decision tree to check consistency among the detected body parts. For evaluation, we collected 4820 thermal images from 26 premature infants. The result shows that our method achieves precision and recall of 94.8% and 77.5%, respectively. Also, we found that the correlation coefficient between the extracted neck temperature and the esophagus temperature is 0.82, which is promising for non-invasive and reliable temperature monitoring for premature infants.

Keywords: Premature infant · Thermal image · Body part detection · Deep learning

1 Introduction

Premature infants require strict body temperature management because of the immaturity of neonatal thermoregulation to control their body temperature. Therefore, it is important to adjust the temperature in an incubator appropriately [3,5]. Currently, medical staff manually adjust the incubator temperature based on the skin temperature of a premature infant measured by a probe.

T. Hara and H. Yamaguchi (Eds.): MobiQuitous 2021, LNICST 419, pp. 438–450, 2022.
https://doi.org/10.1007/978-3-030-94822-1_24

However, the attachment of the probe to the neonatal skin is difficult owing to the premature neonatal skin, which often leads to inaccurate skin temperature measurement.

For non-invasive and reliable skin temperature measurement, thermal cameras attract attention recently. It enables measurement without giving premature infants discomfort or stress. However, medical staff need to manually specify regions in the thermal images to obtain the skin temperatures of body parts of interest such as face, neck and arms. This is a barrier to continuous measurement of neonatal skin temperature which enables appropriate control of incubator temperature.

Use of other non-invasive sensors such as a camera can help to detect body parts. However, we should avoid system complexity due to additional deployment and maintenance cost. Also, privacy issue may arise even for neonates. Therefore, we need a method to detect the body parts directly from thermal images. Although many methods such as OpenPose [1] for pose estimation have been proposed for visible cameras, thermal images are greatly different from visible images, which requires custom designs. Therefore, similarly to other research fields, deep learning has been applied for thermal images recently. Many face recognition methods for thermal images have been proposed for the purpose of nighttime surveillance [4]. Also, ThermalPose [2] achieves pose estimation for adults by re-training OpenPose for thermal images with ground truth obtained by a visible camera. However, the data collection of premature infants by both a thermal camera and a visible camera is difficult due to deployment cost and privacy concerns.

In this paper, we propose a method to detect six body parts from a thermal image of a premature infant in an incubator by using deep learning. We train YOLOv5 [6][1] to detect six body parts (i.e. head, torso, left/right arms, left/right legs) from thermal images. We choose the six body parts because they are key parts which enable further analysis for finding detailed parts. To enhance the accuracy, we apply the knowledge about the relative positions of the body parts. Since YOLOv5 does not consider relative positions of the body parts, we leverage the decision tree to check consistency among the detected body parts.

For evaluation, we collected 4820 thermal images from 26 premature infants in a hospital. The results show that our method achieves precision and recall of 94.8% and 77.5%, respectively. Also, to show the feasibility of the thermal images for the neonatal incubator control, we extract the skin temperature around the neck which is known to be close to the core body temperature. The neck region is defined based on the detected head and torso regions. The result demonstrates the correlation coefficient between the neck skin temperature and the esophageal temperature is 0.82, which is promising for non-invasive and reliable temperature monitoring for premature infants.

Our contributions are summarized below.

– To the best of the authors' knowledge, we are the first to propose body part detection from thermal images of premature infants using deep learning.

[1] https://github.com/ultralytics/yolov5.

Fig. 1. Data collection environment.

- Our method leverages the relative positions among the body parts to achieve accurate detection.
- We demonstrate the effectiveness of the skin temperature measurement by a thermal camera through the neck skin temperature extraction based on the detected body parts.

2 Data Collection

Figure 1 shows the data collection environment. A thermal camera is attached to the top of an incubator such that the whole body of a neonate fits within the thermal image. We captured an image every 20 min from the video taken for 72 h for each premature infant. In total, we obtained and labeled 4820 thermal images with various postures of 26 infants. The thermal image size is 320 × 256 pixels.

Some of the collected thermal images were recorded during intervention by medical staff. In addition, some others are not appropriate for temperature extraction since the target body parts are not visible, e.g., when a neonate lies on their side. Skin temperature extraction may be difficult and unreliable if we use such images. Therefore, we classify the input images into valid or invalid by using Convolutional Neural Network (CNN). For this purpose, we manually checked all the images and labeled them as valid if the neonate lies on their back without intervention and all of the six body parts are visible. Finally, 3868 images were labeled as invalid while the remaining 952 images were labeled as valid.

3 Proposed Method

3.1 Overview

Figure 2 illustrates the overview of the proposed method. First, we perform binary classification by CNN with thermal images as an input to remove invalid

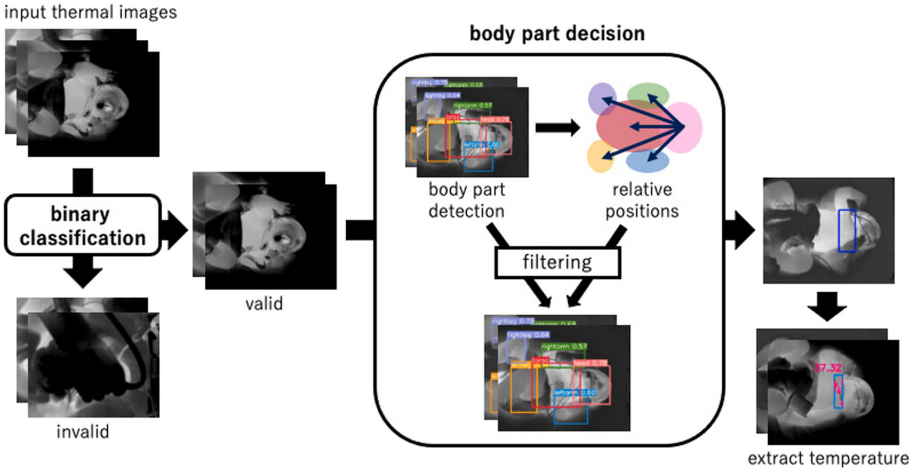

Fig. 2. Overview of the proposed method.

images as mentioned in Sect. 2. Second, we perform body part detection based on YOLOv5 trained by valid thermal images. Our target body parts are six, i.e. the head, torso, left/right arms, and left/right legs. Third, for the detected body parts, we extract features including relative positions to the detected head. The features of a detected body part are input to a decision tree model which classifies the input into one of the six body parts. If the classification result is consistent with the detection class by YOLOv5, we accept it. Otherwise, we reject the detection result as wrong detection. Finally, we extract the skin temperature of interest parts such as a neck based on the detected body parts. In this paper, we describe a method to extract the neck skin temperature based on the detected head and torso.

3.2 Classification of Valid Thermal Images

To filter the images which are not suitable for the body parts detection, we perform binary classification by CNN. The CNN architecture is shown in Fig. 3. We can distinguish between valid and invalid images by a simple model because there are clear differences in temperature distribution in addition to the shape and size. The effect of unnecessary parameters is also reduced by making the model smaller, which also avoids overfitting.

We used the LR range test [8] as a method to determine the initial learning rate. The LR range test gradually increases the learning rate over a certain range and adopts the learning rate of 0.00001 when the loss is the lowest. We built a model in which the convolution process by the 3×3 filter is performed 16 times in the first layer and 32 times in the second layer. We set the learning rate, the batch size, and the epoch to 0.00001, 16, and 100, respectively. The ReLU function is used as the activation function.

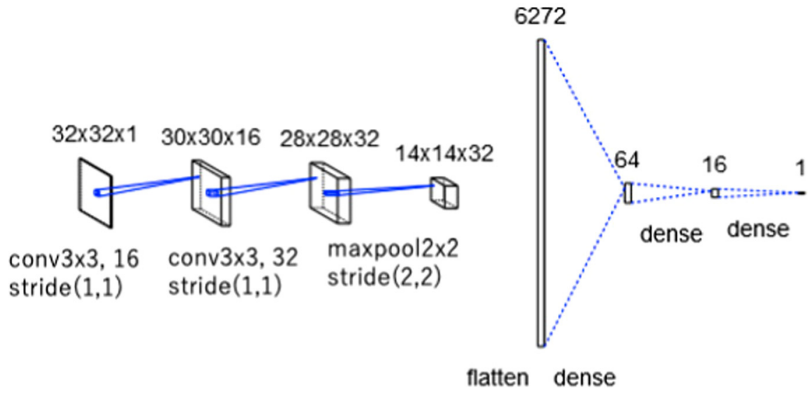

Fig. 3. CNN architecture for valid image classification.

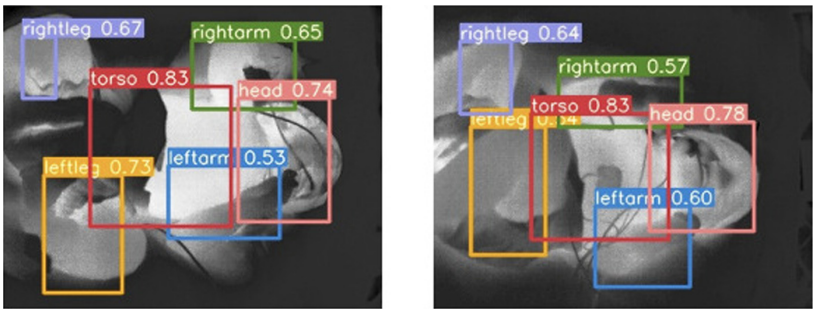

Fig. 4. Example of parts detection result.

3.3 Body Part Detection

We use the body part detection model learned by YOLOv5 [6]. The input is thermal images classified as valid by CNN. The output is the thermal images with the detected bounding boxes with their classes (i.e. body parts). We selected YOLOv5s model with 7 million parameters which is the smallest network among variations of YOLOv5. The batch size is set to 16 and the epoch is set to 1500 for learning.

In the proposed method, we only use at most one detection result with the highest confidence score for each body part because there is only one neonate in the incubator. The confidence score is given by YOLOv5 with the detection result. Figure 4 shows an example of the detection result.

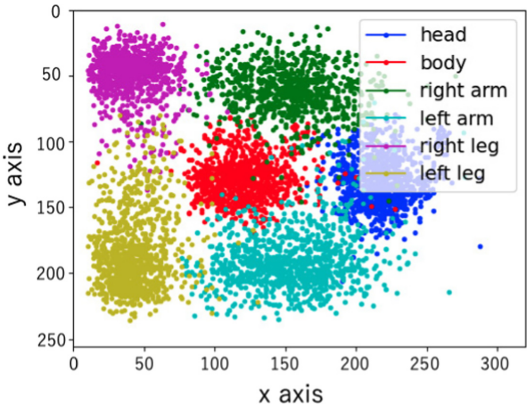

Fig. 5. Position distribution of each part.

Table 1. List of features.

Type	Features
Absolute position	x-coordinate, y-coordinate
Size	Bounding box area size
Relative position	Distance from head, x/y-component of directional vector from head

3.4 Filtering by Relative Positions

The input of the decision tree is the features calculated from the output of the body part detection model. The decision tree outputs one of the six body parts based on the input features.

We let B_p denote the bounding box of body part p output by the part detection model. We also let $D(B)$ be the output body part by the decision tree given the bounding box B as the input. We check the consistency between p and $D(B_p)$, and accept B_p if and only if $p = D(B_p)$. Otherwise, we reject B_p.

To design the features for the decision tree, we analyzed the distribution of the center coordinates of each part in the valid images as shown in Fig. 5. There are some trends in the relative positions among the parts. From this observation, we designed six features for the decision tree input as listed in Table 1. We note that the detected head is used as the reference of the relative positions. This is because the body part detection performance of the head detection is the highest, indicating stable and reliable results as shown in Sect. 4.3 later.

We discuss the effectiveness of each feature in Sect. 4.4. We use DecisionTreeClassifier implemented in scikit-learn.[2]

[2] https://scikit-learn.org/.

3.5 Extraction of Neck Temperature

To see the effectiveness of the skin temperature measurement by using the proposed method, we design a simple method to extract the neck skin temperature. Since the position of the thermal camera is fixed, the head of the neonate is always on the right side of the thermal images. Therefore, we assume that the neck area is between the head and the torso.

Specifically, we define the neck area as below. We let l_h and r_t denote the left side of the head bounding box and the right side of the torso bounding box, respectively. The neck area is a bounding box having the center coordinate which is the midpoint of the midpoints of l_h and r_t. The height and width of the bounding box are 60 and 20 pixels, respectively.

From the neck bounding box, we further extract the neck skin temperature. Since the bounding box may still contain other areas including backgrounds, we take the average of the top 25% in the temperature distribution of the neck bounding box. This process successfully extracts the neck skin temperature because it is known that the neck skin temperature is relatively high compared to other areas.

4 Evaluation

4.1 Evaluation Settings

For evaluation, we used the dataset collected as mentioned in Sect. 2. The dataset consists of 3868 invalid and 952 valid images. To avoid self-test, we randomly selected 762 invalid and 762 valid images for training of CNN. The same 762 valid images were also used for training of the body part detection model and decision tree. The remaining 3106 invalid and 190 valid images were used for the test. Since the ground truth of the body parts is labeled for the valid images, we used only the valid images for evaluation of the body part detection and decision tree. Meanwhile, we used the 3106 invalid and 190 valid images for evaluation of the extraction of the neck skin temperature.

Similarly to other classification problems, we use well-known metrics, i.e. precision and recall. Furthermore, object detection requires evaluation of the correctness of the detected area in addition to the classification. For this purpose, IoU (Intersection over Union) [7] is often used. IoU is defined as below.

$$\mathrm{IoU} = \frac{|B_p \cap \hat{B}|}{|B_p \cup \hat{B}|},$$

where B_p is the estimated bounding box labeled as body part p and $\hat{B}$ is the ground truth. The pair of the bounding boxes is chosen such that IoU is the highest. We note that IoU is regarded as zero if there is a mismatch in the class. In the following evaluation, we set the IoU threshold to 0.3 unless otherwise specified, i.e. a detected bounding box is regarded as correct if its IoU is 0.3 or more. We discuss the IoU threshold in Sect. 4.3.

Table 2. Confusion matrix of binary classification.

		Predicted class		
		Valid	Invalid	Recall(%)
Actual class	Valid	176	14	92.6
	Invalid	273	2833	91.2
	Precision(%)	39.2	99.5	

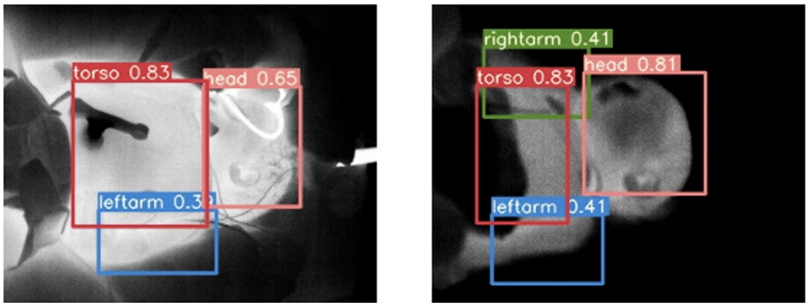

Fig. 6. An example of the body part detection for an invalid image classified as valid.

4.2 Binary Classification Performance

Table 2 shows the confusion matrix of the binary classification. The average precision and recall were 69.4% and 91.9%, respectively. Precision of the valid class is especially low because of the data imbalance between the classes (3106 vs. 190). Nevertheless, recall of both classes exceeds 91%.

We note that incorrect invalid classification means the loss of valid images while incorrect valid classification may lead to wrong body parts detection. Actually, we found that the images wrongly classified as valid were very similar to the valid images. The only difference was the lack of some body parts, e.g. both legs are out of the image. Even for such images, our method can still detects body parts that appear in the images. Figure 6 shows an example of the body part detection result for an invalid image classified as valid. Therefore, the performance of the binary classification is high enough to remove totally invalid images which lead to unreliable skin temperature extraction.

4.3 Body Part Detection Performance

To confirm the convergence in the training of the body part detection model, we show the mAP (mean Average Precision) over the epoch in Fig. 7. The result indicates that the training converges around 1000 epochs. We use this trained model for evaluation.

We evaluated precision and recall of the body part detection model. Figure 8 shows the results for the different IoU thresholds. We can see that decreasing

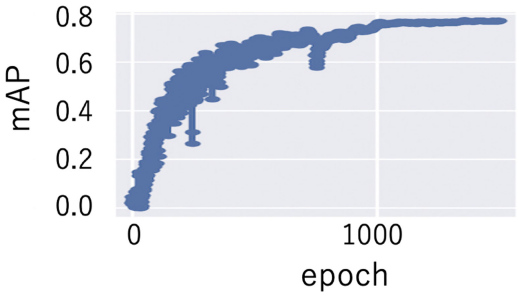

Fig. 7. mAP showing learning convergence of body part detection model.

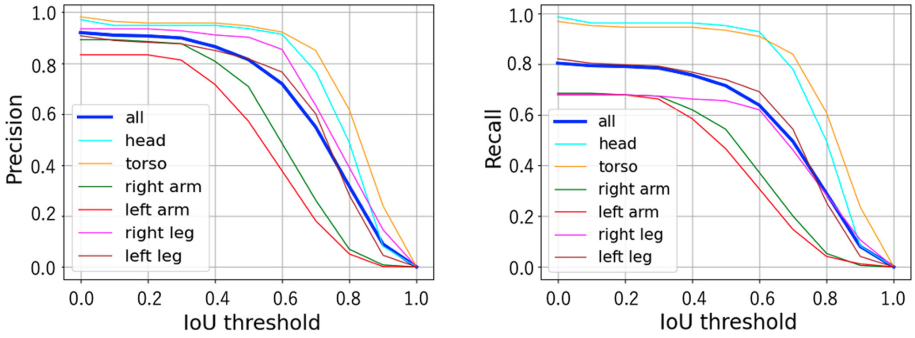

Fig. 8. Precision and recall for different body parts.

the IoU threshold improves both precision and recall. This is because, when the IoU threshold is low, the detected bounding box is regarded as a true positive even if the overlap with the ground truth is small. The IoU threshold less than 0.3 does not greatly improve the results. We also confirmed that the detected bounding boxes tend to be smaller than the ground truth.

However, if we successfully detect the rough positions of the body parts of interest, further processing based on the temperature distribution can extract the skin temperature of interest. Therefore, we set the IoU threshold to 0.3 in the following evaluation.

As for the difference between the body parts, the head and torso achieve high performance compared to the others (i.e. limbs). This is natural because the limbs are small and often appear in different positions owing to frequent movement.

4.4 Filtering Effect by Decision Tree

Figure 9 shows the decision tree model we trained. The results are also shown in Table 3. Precision and recall after filtering were 94.8% and 77.5%, respectively. Precision improved by 4.9% after filtering, which achieves approximately

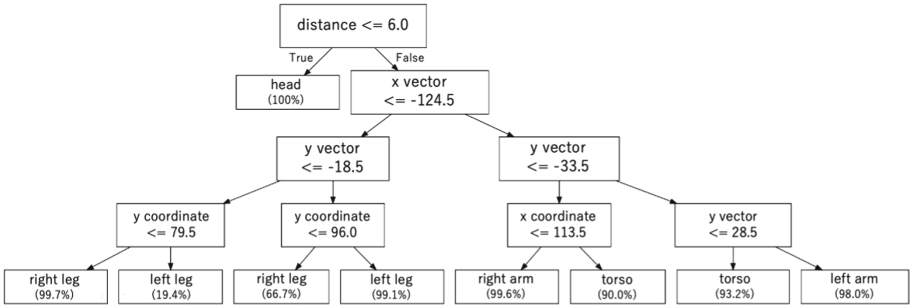

Fig. 9. Trained decision tree model.

Table 3. Part detection results before/after filtering.

	Precision	Precision (filtered)	Recall	Recall (filtered)
Head	94.8	94.8	96.4	96.4
Torso	95.8	**96.4**	94.7	94.7
Right arm	87.7	**94.9**	**67.5**	63.7
Left arm	81.2	**92.0**	**66.3**	65.9
Right leg	92.7	**95.8**	**67.5**	66.9
Left leg	87.6	**93.5**	**79.3**	77.4
Total	89.9	**94.8**	**78.6**	77.5

95% in the detection of the head and torso. In addition, precision of limbs also exceeded 90%. On the other hand, recall decreased by 1.1% after filtering. This result indicates some of the correct results were wrongly excluded due to filtering. Nevertheless, precision is the most important metric considering medical applications. Even if recall is low, we may be able to notify the fact that the target body parts are not detected. From the above results, we confirmed that filtering based on relative positions is effective to improve the precision while suppressing the decrease of recall.

Figure 10 shows the importance of each feature. We can see that the most important feature is the y-component (i.e. the vertical direction in the image) of the directional vector from the head position. We also see that the x-component of the directional vector and the distance from the head are equally important. These three features account for 98.2%, indicating the importance of the relative position from the head. This is because premature infants in incubators usually lie on their backs. This means the relative positions among the body parts are almost uniquely characterized by the direction and distance from the head.

4.5 Usefulness of Extracted Skin Temperature

Figure 11 shows examples of the extracted neck areas and neck skin temperature. We can see that most of the top 25% pixels are concentrated around the neck.

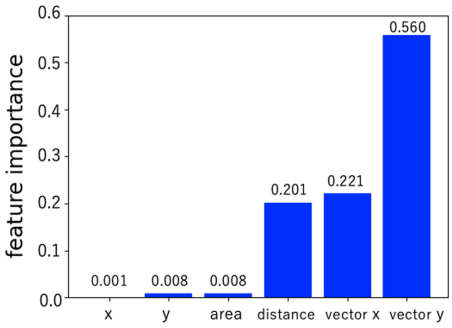

Fig. 10. Feature importance of the decision tree.

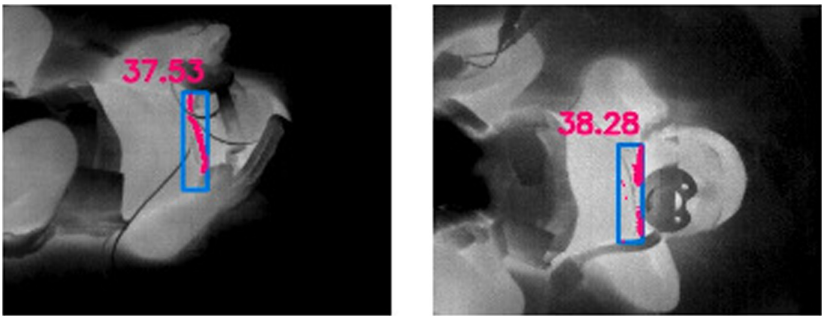

Fig. 11. Examples of extracted neck areas (blue boxes). The pink dots indicate the top 25% in the temperature distribution in the neck area. (Color figure online)

To see the usefulness of the extracted neck skin temperature, we see the correlation between the extracted neck skin temperature and the esophageal temperature measured by a probe. This is because the esophageal temperature is often used as the core body temperature, which is a key factor for medical care.

As we mentioned earlier, we used both invalid and valid images in this evaluation. This means the neck skin temperature may be extracted from some invalid images which are wrongly classified as valid. Even for invalid images, our method can extract the neck skin temperature as long as the head and torso are detected. Consequently, we extracted the neck skin temperature from 228 thermal images of which 170 are the valid images. Finally, we used 26 images to see the correlation between the esophageal temperature and the neck skin temperature since the esophageal temperature was measured for a part of the premature infants.

Figure 12 shows the correlation between the neck skin temperature and the esophageal temperature. The result shows that the correlation coefficient between the neck skin temperature and the esophageal temperature is 0.82, indicating strong correlation. Therefore, this result indicates the extracted skin temperature can be used as a reference to the core body temperature, highlighting the usefulness of our method.

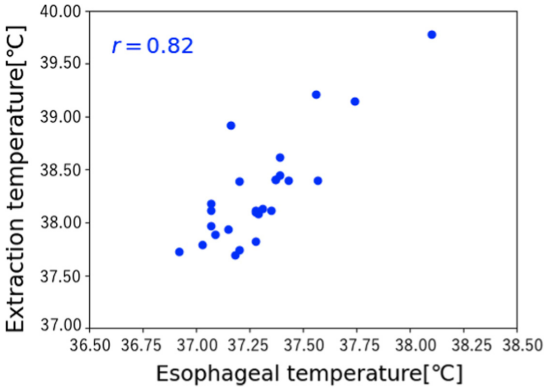

Fig. 12. Correlation between the esophageal temperature and the neck skin temperature extracted by the proposed method.

5 Conclusion

In this paper, we proposed body parts detection from neonatal thermal images using deep learning for the non-invasive skin temperature measurement. Our method combines CNN, YOLOv5, and the decision tree to improve the performance. The evaluation through 4820 thermal images collected from 26 premature infants shows that our method achieves precision and recall of 94.8% and 77.5%, respectively. Furthermore, we demonstrated that the neck skin temperature can be extracted based on the positions of the detected body parts. The correlation coefficient between the extracted neck temperature and the esophageal temperature was 0.82, which is promising for non-invasive and reliable temperature monitoring for premature infants.

For future work, we will further collect the data for analysis on the skin temperature and core body temperature. In addition, we will extract not only the neck temperature but also the temperature of other areas by using the detected body parts. Our future work also includes analysis on the appropriate temperature control of the neonatal incubators based on the extracted skin temperature.

References

1. Cao, Z., Hidalgo Martinez, G., Simon, T., Wei, S., Sheikh, Y.A.: OpenPose: realtime multi-person 2D pose estimation using part affinity fields. IEEE Trans. Pattern Anal. Mach. Intell. **43**(1), 172–186 (2019)
2. Chen, I.C., Wang, C.J., Wen, C.K., Tzou, S.J.: Multi-person pose estimation using thermal images. IEEE Access **8**, 174964–174971 (2020)
3. Knobel, R.B.: Thermal stability of the premature infant in neonatal intensive care. Newborn Infant Nurs. Rev. **14**(2), 72–76 (2014)
4. Krišto, M., Ivašić-Kos, M.: An overview of thermal face recognition methods. In: Proceedings of International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 1098–1103 (2018)

5. Ali, M., Abdelwahab, M., Awadekreim, S., Abdalla, S.: Development of a monitoring and control system of infant incubator. In: Proceedings of International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE), pp. 1–4 (2018)
6. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779–788 (2016)
7. Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., Savarese, S.: Generalized intersection over union. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 658–666 (2019)
8. Smith, L.N.: Cyclical learning rates for training neural networks. In: Proceedings of IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 464–472 (2017)