Comparison of Traditional Transformations for Data Augmentation in Deep Learning of Medical Thermography

Ahmet Haydar ORNEK, Murat CEYLAN
Department of Electrical & Electronics Engineering
Faculty of Engineering and Natural Science, Konya Technical University
Konya, Turkey
ahmethaydarornek@gmail.com
mceylan@ktun.edu.tr

***Abstract*—Convolutional neural networks (CNN) models demonstrate high performance in image-based applications such as image classification, segmentation, noise reduction and object recognition. However, balanced and sufficient data are required to effectively train a CNN model but this is not always possible. Since conditions of both hospitals and patients are not always appropriate for collecting data and patients with the same disease are not always available, the problem of collecting balanced and sufficient data are often occurred in medical fields. In this study, comparison of traditional data augmentation methods such as rotating, mirroring, zooming, shearing, histogram equalization, color changing, sharpening, blurring, brightness enhancement and contrast changing were performed by using neonatal thermal images. These images belonged to 19 unhealthy and 19 healthy neonates were obtained from Selcuk University, Faculty of Medicine, Neonatal Intensive Care Unit. A combination of three different augmentation methods were implemented to original images in each one of the 10 different comparisons accomplished and CNN was used to classify the these comparisons in the study. When contrast changing, sharpening and blurring methods were used, increased the sensitivity rate by 23.49%, the specificity rate by 29.09% and the accuracy rate by 26.29%. The obtained results show that simple and low–cost traditional data enhancement methods can improve the performance of classification.**

***Index Terms*—classification; convolutional neural networks; data augmentation; medical thermography; neonate**

I. Introduction

Thermal imaging technique is called thermography and obtained thermal images are called thermograms. All substances with temperature above 0 Kelvin radiate infrared thermal waves. These infrared thermal waves are caught by thermal cameras and then converted to thermograms [1]. Since thermograms contain both temperature and image information, applications are carried out without the need for external temperature measuring and imaging device.

Thermography which can be practiced day and night and is a non-contact, non-ionized and non-invasive method is used in military, environmental, industrial and medical fields.Thermoregulation [2], breast cancer [3], neonatal monitoring [7], [4], [10], [9], [8], [5], [6], [11], urology [12] and vascular diseases [13] are performed using thermography in medical field. When studies about neonatal monitoring are examined, it has been observed that machine learning algorithms and conventional methods such as template matching are used [**?**]. To the best of our knowledge, there had not been any study about neonatal imaging using deep learning methods.

Convolutional neural networks [14], which is one of the deep learning models, demonstrates high performance for applications such as image classification, segmentation, generation and noise reduction. But, there are problems of creating appropriate model and obtaining balanced and sufficient data. Since suitable environments for obtaining image data cannot always be available, the often faced problem in medical fields is collecting balanced and sufficient data. There are various data augmentation methods that can be used to overcome the problem of sufficient data [15]. These methods can be divided into two categories as traditional (changing rotation, contrast, brightness etc.) and advanced (texture and style transfer etc.). While advanced methods include different learning models and time consuming methods, traditional methods are low–cost and simple to apply.

In this study, the use of different traditional data augmentation methods (rotating, mirroring, zooming, shearing, histogram equalization, color changing, sharpening, blurring, brightness enhancement and contrast changing) in the classification of medical thermograms was compared by using 3800 thermograms belonged to 19 healthy and 19 unhealthy neonates taken at Selcuk University, Faculty of Medicine, Neonatal Intensive Care Unit. Although these methods are simple and low-cost, performance of classification increased the accuracy rate by 26.29%.

II. Material and Methods

A. Used Data

The thermograms used in this study were obtained from neonatal intensive care unit by using Variocam HD infrared

This study was supported by the Scientific and Technological Research Council of Turkey (TUBITAK, project number: 215E019).

thermal camera over a one year period. The resolution of the thermal camera is 640x480 and thermal sensitivity is approximately 0.01 centigrade. Thermograms were captured from a distance of 60-100 cm from neonates lying in the supine position. The imaging setup is shown in Fig. 1.

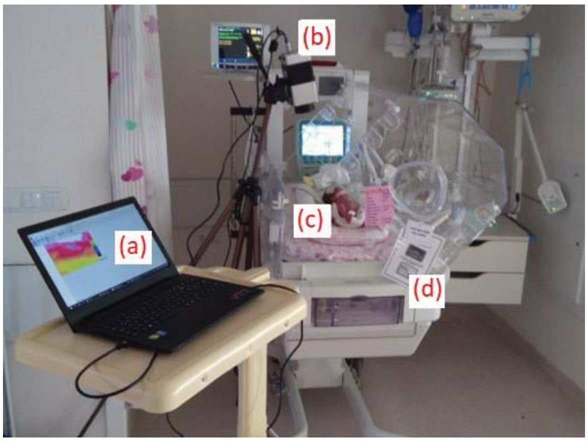

Fig. 1. Imaging setup (a) portable computer (b) thermal camera (c) neonate (d) incubator.

B. Data Augmentation

Data augmentation methods can be divided into two categories as traditional and advanced methods. While traditional methods are easy to apply and have a low complexity time, advanced methods such as generative adversarial networks are complicated [15]. In this study, comparison of traditional methods was evaluated. All parameters used for rotating, mirroring, zooming, shearing, histogram equalization, color changing, sharpening, blurring, brightness enhancement and contrast changing were selected empirically.

Rotation changing. The operation of 45 degrees rotation in a counterclockwise direction around a center point was applied to change the rotation of images. As shown in Fig. 2 (b), the portion except the 640 x 480 frame was cropped to avoid changing the size of the images.

Mirroring. Mirroring was performed by taking the symmetry of each pixel value according to Y axis in the image. According to this operation, a pixel's position was changed from (x, y) to (640-x, y). Obtained image is shown in Fig. 2 (c).

Zoom in. This operation contains two stages (1) 640 x 480 sized image was cropped from column 140 to column 570 and from row 140 to row 400 (2) as shown in Fig. 2 (d), cropped portion was rendered to 640 x 480 by using image resizing.

Shearing. 2D affine transforms were applied to images to obtain different positions of the thermograms. Occurred image is shown in Fig. 2 (e).

Histogram equalization. The method used to create a black-white balance in gray - level images is called histogram equalization. To apply histogram equalization, first of all, the images were divided into red, green and blue bands and then images were reassembled by applying the histogram equalization to every band. Obtained image is shown in Fig. 2 (f).

Color changing. To change the image's color, images were divided into red, green and blue bands once again. While images are being reassembled, red, green and green bands were used. Thus, as shown in Fig. 2 (g), images with different colors were obtained.

Contrast and brightness changing. Contrast of images were changed by multiplying all pixel values by 0.75 and brightness of images were changed by adding 0.2 to all pixel values. Occurred images are shown in Fig. 2 (h, i).

Blurring and sharping. Low-pass Gaussian filter with standard deviation of 2 was used to blur the images. Images were sharpened by subtracting a blurred version of the image from itself. Standard deviation of low-pass Gaussian filter was selected as 1.5. Obtained images are shown in Fig. 2 (j, k).

C. Convolutional Neural Networks

CNN is one of the deep learning models that had been demonstrating high performance in image-based applications such as image classification and segmentation over the past decade [14]. When classifying an image, either all pixels or features that are extracted from images are given as inputs to a classifier model. CNN is a learning model that gives features to the classifier model after extracting them from images. It contains two types of layers; convolution layer for feature extraction-reduction and fully-connected layer for classification. The CNN model used in this study is shown in Fig. 3.

Convolutional layer includes two operations as convolution and pooling. Convolution operation extracts features of images and pooling operation reduces the size of these features. Fully-connected layer classifies the flattened features and after that classification is realized.

D. Evaluation Metrics

Specificity (1), sensitivity (2) and accuracy (3) metrics [16] were calculated by using confusion matrix to evaluate the classification results. Confusion matrix is a table that allows visualization of the performance of a classification. As shown in Fig. 4, each row of the matrix depicts the actual class and each column depicts the predicted class. Specificity represents the rate of thermograms classified as healthy among healthy neonates while sensitivity represents the rate of thermograms classified as unhealthy among unhealthy neonates.

$$specificity = \frac{TN}{TN + FP} \tag{1}$$

$$sensitivity = \frac{TP}{TP + FN} \tag{2}$$

$$accuracy = \frac{TN + TP}{TN + TP + FN + FP} \tag{3}$$

Where TP, TN, FP and FN are the abbreviation of true positive, true negative, false positive and false negative, respectively. TP refers to the number of unhealthy neonates that were correctly labeled by the classifier, TN refers to the

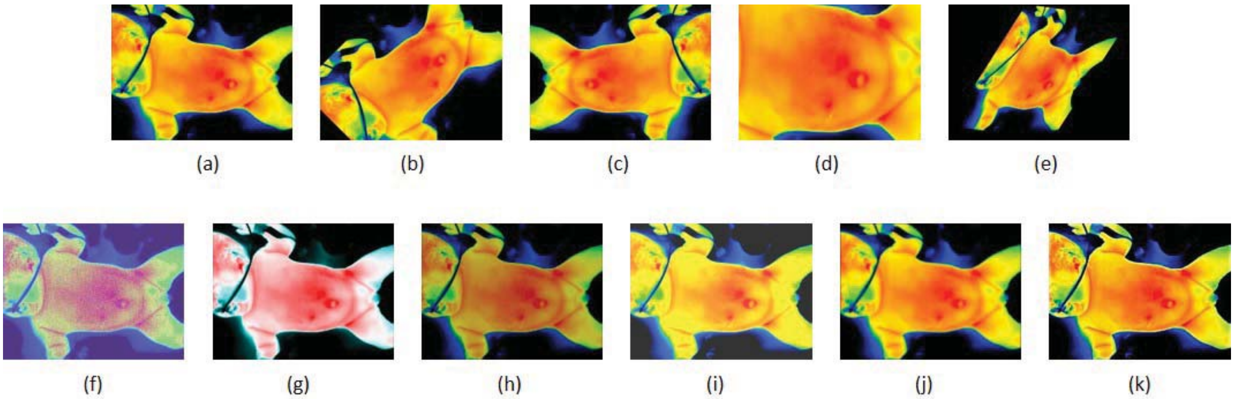

Fig. 2. Original thermal image and augmented thermal images (a) original (b) rotation (c) mirroring (d) zooming (e) sharing (f) histogram equalization (g) color changing (h) contrast changing (i) brightness changing (j) blurring (k) sharping.

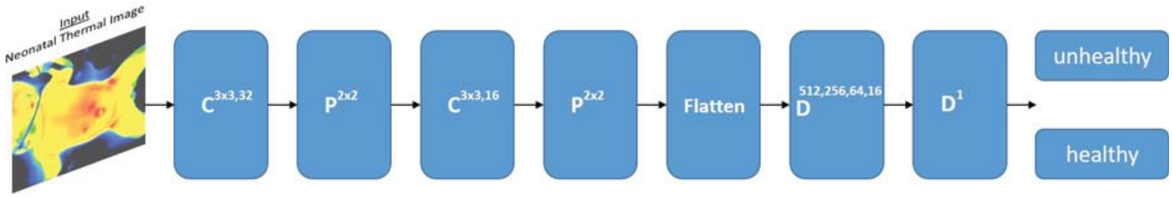

Fig. 3. Created convolutional neural networks model. C a x b, n describes convolutional layer, P a x b describes pooling layer, A and b are dimension of kernel and n is number of kernel. D k describes dense layer, k is number of neuron. This CNN model contain 2252017 trainable parameters. First convolutional layer contain 32 different 3 x 3 kernel and second 16 different 3x3. All pooling kernel was selected 2 x 2 as max pooling. Five dense layers was used to classify of obtained features. There are 512, 256, 64, 16, 1 neurons from one to five layer, respectively.

	Predicted Class	
Actual Class	TP	FN
	FP	TN

Fig. 4. A Confusion Matrix.

number of healthy neonates that were correctly labeled by the classifier. FN refers to the number of unhealthy neonates that were incorrectly labeled by the classifier, FN refers to the number of unhealthy neonates that were incorrectly labeled by the classifier.

E. 10-Fold Cross Validation

In classification tasks part of the data is defined as test and the rest of the data is defined as train part and evaluations are carried out according to the test part. But, this technique is inadequate due to the fact that it depends on a part of the data thus it gives misleading results. To avoid this inadequacy K-fold cross validation [17] is used instead of this technique. According to k-fold cross validation, data set is divided into K parts and while K part is used as the test set, K-1 parts are used as the train set for K times. Also confusion matrices are calculated for K times and all confusion matrices are summarized to calculate the evaluation metrics. K value was selected as 10 and 1520 images were used as the test set while 13680 images were used as the train set for ten times.

III. Experiments and Results

In this study, medical thermograms were increased by using ten different traditional data augmentation methods such as rotating, mirroring, zooming, shearing, histogram equalization, color changing, sharpening, blurring, brightness enhancement and contrast changing. Ten different combinations were created using these ten data augmentation methods. In each combination, in addition to the original images, image sets were created using three different augmented images. A total of 15200 images (3800 original, 11400 augmented) were separated according to 10-fold cross validation technique and classified with CNN. All of the obtained confusion matrices are shown in Fig. 5. All experiments and used methods are given in Table 2.

IV. Conclusion

With the development of deep learning models, new applications have been developing in different fields such as military, environmental, industrial and medical. Especially CNNs have been demonstrating high performance for image classification. However, CNNs have some limitations such as the amount of image data used as inputs. There are some data augmentation methods that can be used to overcome the lack of data problem. These methods can be divided into two categories as

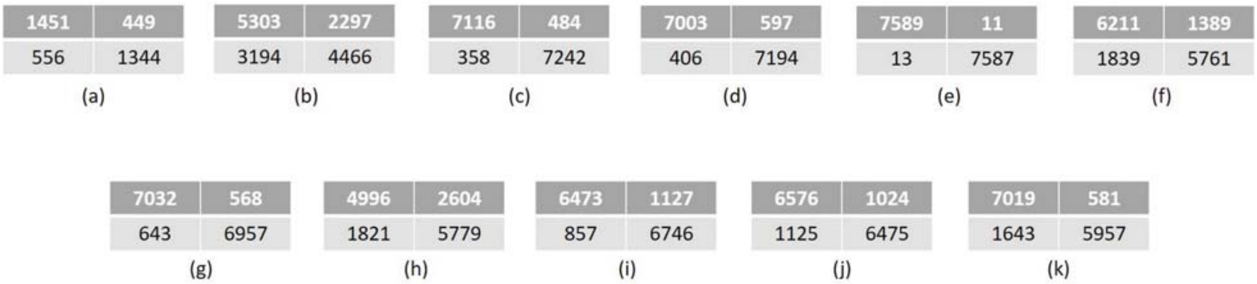

Fig. 5. Obtained confusion matrices from (a) original (b) rotation (c) mirroring (d) zooming (e) sharing (f) histogram equalization (g) color changing (h) contrast changing (i) brightness changing (j) blurring (k) sharping.

TABLE I
ALL EXPERIMENTS AND SELECTED METHODS

Trials	Used Methods			Sensitivity (%)	Specificity (%)	Accuracy (%)
A	Without Data Augmentation			76,36	70,73	73,55
B	Rotation changing	Mirroring	Zooming	69,77	58,76	64,26
C	Histogram equalization	Color changing	Rotation changing	93,63	95,28	94,46
D	Contrast changing	Brightness changing	Sharping	92,14	94,65	93,40
E	Contrast changing	Sharping	Blurring	99,85	99,82	99,84
F	Sharing	Blurring	Mirroring	81,72	75,80	78,76
G	Zooming	Color changing	Brightness changing	92,52	91,53	92,03
H	Rotation changing	Sharing	Zooming	65,73	76,03	70,88
I	Sharping	Sharing	Color changing	85,17	88,76	86,96
J	Rotation changing	Blurring	Brightness changing	86,52	85,21	85,86
K	Histogram equalization	Zooming	Brightness changing	92,35	78,38	85,36

traditional and advanced methods. While traditional methods simple and low-cost, new learning methods such as generative adversarial networks need to be implemented for advanced methods. In this study, thermal images were augmented by using traditional methods such as rotating, mirroring, zooming, shearing, histogram equalization, sharpening, blurring, brightness enhancement and contrast changing.

When the augmented images obtained using three different data augmentation methods were added to the original images, sensitivity, specificity and accuracy values were found to be over 99%. The results shows that when the correct methods are selected, the use of traditional methods instead of advanced data augmented methods increase the classification performance. In future studies, it is aimed to design an automatic system that determines which diseases the neonates have. First of all, the results will be evaluated by using traditional data augmented methods. If the desired results are not reached, advanced methods such as new image generation, style and texture transformation will be used.

REFERENCES

[1] Jagannathan G: A Case for Joint Development of IR Cameras in India, 2007

[2] Heimann K, Jergus K, Abbas AK, Heussen N, Leonhardt S, Orlikowsky T: Infrared thermography for detailed registration of thermoregulation in premature infants, 2013

[3] Kontos M, Wilson R, Fentiman I: Digital infrared thermal imaging (DITI) of breast lesions: sensitivity and specificity of detection of primary breast cancers. Clinical Radiology 66:536-539, 2011

[4] Knobel-Dail RB, Holditch-Davis D, Sloane R, Guenther BD, Katz LM: Body temperature in premature infants during the first week of life: Exploration using infrared thermal imaging. Journal of thermal biology 69:118-123, 2017

[5] Ornek AH, Savasci D, Ceylan M, Ervural S, Soylu H: Determination of correct approaches in evaluation of thermograms, 2018

[6] Savaşci D, Ceylan M: Thermal image analysis for neonatal intensive care units (First evaluation results). Proc. 2018 26th Signal Processing and Communications Applications Conference (SIU): City, 2-5 May 2018 Year

[7] Clark R, Stothers J: Neonatal skin temperature distribution using infrared colour thermography. The Journal of physiology 302:323-333, 1980

[8] Abbas AK, Heimann K, Jergus K, Orlikowsky T, Leonhardt S: Neonatal non-contact respiratory monitoring based on real-time infrared thermography. Biomed Eng Online 10, 2011

[9] Abbas AK, Heimann K, Blazek V, Orlikowsky T, Leonhardt S: Neonatal infrared thermography imaging: Analysis of heat flux during different clinical scenarios. Infrared Phys Techn 55:538-548, 2012

[10] Abbas AK, Leonhardt S: Neonatal IR-Thermography Pattern Clustering based on ICA Algorithm, 2014

[11] Nur R: Identification of thermal abnormalities by analysis of abdominal infrared thermal images of neonatal patients. Carleton University 2014)

[12] Kee Ng W, Ng E, Tan Y: Qualitative Study of Sexual Functioning in Couples with Erectile Dysfunction Prospective Evaluation of the Thermography Diagnostic System, 2009

[13] Bagavathiappan S, et al.: Infrared thermal imaging for detection of peripheral vascular disorders. Journal of medical physics 34:43-47, 2009

[14] Krizhevsky A, Sutskever I, Hinton GE: Imagenet classification with deep convolutional neural networks. Proc. Advances in neural information processing systems: City

[15] Mikołajczyk A, Grochowski M: Data augmentation for improving deep learning in image classification problem. Proc. 2018 International Interdisciplinary PhD Workshop (IIPhDW): City, 9-12 May 2018 Year

[16] Zhu W, Zeng N, Wang N: Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations. NESUG proceedings: health care and life sciences, Baltimore, Maryland 19:67, 2010

[17] Kohavi R: A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection, 2001