# Can Convolution Neural Network (CNN) Triumph in Ear Recognition of Uniform Illumination Invariant?

# Nursuriati Jamil<sup>1</sup>, Ali Abd Almisreb<sup>2</sup>, Syed Mohd Zahid Syed Zainal Ariffin<sup>3</sup>, N. Md Din<sup>4</sup>, Raseeda Hamzah<sup>5</sup>

<sup>1,3,5</sup>Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
<sup>2,4</sup>College of Graduate Studies, Universiti Tenaga Nasional, Jalan IKRAM-UNITEN, 43000 Kajang, Malaysia

## **Article Info**

# ABSTRACT

| Article history:     |  |  |  |
|----------------------|--|--|--|
| Received Mar 1, 2018 |  |  |  |
| Revised Apr 21, 2018 |  |  |  |
| Accepted May 1, 2018 |  |  |  |

## Keywords:

Convolution Neural Network Ear Recognition Uniform Illumination Invariant Lumens

Current deep convolution neural network (CNN) has shown to achieve superior performance on a number of computer vision tasks such as image recognition, classification and object detection. The deep network was also tested for view-invariance, robustness and illumination invariance. However, the CNN architecture has thus far only been tested on non-uniform illumination invariant. Can CNN perform equally well for very underexposed or overexposed images or known as uniform illumination invariant? This is the gap that we are addressing in this paper. In our work, we collected ear images under different uniform illumination conditions with lumens or lux values ranging from 2 lux to 10,700 lux. A total of 1,100 left and right ear images from 55 subjects are captured under natural illumination conditions. As CNN requires considerably large amount of data, the ear images are further rotated at every 5° angles to generate 25,300 images. For each subject, 50 images are used as validation/testing dataset, while the remaining images are used as training datasets. Our proposed CNN model is then trained from scratch and validation and testing results showed recognition accuracy of 97%. The results showed that 100% accuracy is achieved for images with lumens ranging above 30 but having problem with lumens less than 10 lux.

> Copyright © 2018 Institute of Advanced Engineering and Science. All rights reserved.

# Corresponding Author:

Nursuriati Jamil, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia. Email: lizajamil@computer.org

# 1. INTRODUCTION

Convolutional Neural Networks (CNN) has been successfully applied in many recognition problems. The conceptual architecture of CNN is inspired by [1]'s seminal work on the cat's striate cortex called receptive field. Later, Fukushima [2], explained the Neocognitron, which defines the layer wise structure of neural networks and explains the spatial invariance characteristic of simple cells and complex cells of visual primary cortex. LeCun introduced the structure of CNN for face and digit recognition [3-5] which demonstrated better recognition results than probability density function methodologies such as Gaussian Bayesian approaches and Gaussian Mixture models. They developed a multi-layer artificial neural network called LeNet-5 which could classify handwritten digits. Like other neural networks, LeNet-5 has multiple layers and can be trained with the backpropagation algorithm [6]. It can obtain effective representations of the original image, which makes it possible to recognize visual patterns directly from raw pixels with little-to-none preprocessing. Generally, CNNs are constructed by stacking interweaved layers of two types: convolutional layers and pooling (subsampling) layers [7]. The pooling layer plays an important role in CNNs since it is mainly responsible for the invariance to data variation and perturbation. The pooling

layer can be obtained through pooling operation which typically contains two steps. Firstly, pooling operator scans the feature map and aggregates information (responses) within each local region [8].

Deep convolution neural network has shown the trend of dominating computer vision's world by triumphing over traditional machine intelligence approaches. It has excelled in image classification [9, 10], character recognition [11] and many more. The robustness of CNN has also been shown in work by [12] where the layers of CNN are studied to better understand how viewpoint invariance is achieved. The datasets used are RGB-D dataset [13] and Pascal3D+ dataset [14] as both datasets contain multiview tabletop objects and images of object categories exhibiting high variability, captured in uncontrolled settings and under many different poses. When used on pre-trained CNNs, the network captures representations that highly preserve the manifold structure at most of the network layers, including the fully connected layers. It is also noted that the later layers such as Pool 5 shows better representation for the view-manifold than early layers like Pool 1. When the network is fine-tuned, very good pose estimation performance is achieved. In [15], CNN's tolerance to image variations such as translation, scale, pose and illumination are investigated using a largescale synthetic dataset. Their datasets consists of image objects of 16 categories, 8 rotation angles, 11 cameras on a semicircular arch for random viewpoints and distances, 5 lighting conditions, 3 focus levels, variety of backgrounds generating over 20 million images in total. A pre-trained network, namely AlexNet CNN is used to discover the expressive capabilities of the pool5 and fc7 layers for object and parameter predictions. Results showed that both fc7 and pool5 produced accurate results for classification implying that these layers have good discriminative power for object recognition. As for parameter predictions, the accuracies for lighting, rotation, and camera are 100%, 77%, and 62%, respectively. This implies that the layers have difficulties most in predicting camera view, while lighting variations pose very little prediction difficulty. As changing camera views alter produce geometric variants, prediction task become more complicated. This finding is further proven when an assessment was done on the network to learn the power of CNN in transferring the learned parameter over one object category to another. Lighting parameter is easily transferred from seen objects to unseen object, while variations in camera views degraded the performance of the network for unseen objects. As lighting variations do not transform the geometric shapes of the seen object, it has the simplest knowledge to be transferred on unseen categories.

Even though viewpoint or camera view invariance remains a challenge in any object recognition [17], we only consider illumination invariant in this paper. Despite the fact that CNNs have shown to handle illumination invariant for object recognition, we believe that further study need to be done for images of uniform illumination invariant as most work focus only on non-uniform invariant [18]. Figure 1 and 2 illustrate previous work [15, 16] on illumination invariant where the image datasets used are only non-uniform illumination invariant caused by changes in lighting intensity, direction and spectrum.



Figure 1. A boat under 5 different illuminations [15]

On the other hand, Vonikakis et al. [18] constructed a new benchmark dataset named Phos featuring a greater variety of imaging conditions, compared to existing databases; containing images captured both under uniform and non-uniform illumination. Phos comprises 15 scenes captured under different illumination conditions. Each scene contains 9 images taken under dissimilar uniform illuminations, and 6 images under different degrees of non-uniform illuminations. Figure 3 demonstrates an example of a scene in Phos database. Uniform illumination is attained by using several diffusive light sources evenly distributed around the objects. Using the shutter speed, four overexposed (+) and four underexposed (-) images are generated from the correct exposure image. A strong directional light source with different strengths is projected to the objects producing 6 non-uniform illumination images.

Can Convolution Neural Network (CNN) Triumph in Ear Recognition of Uniform... (Nursuriati Jamil)



Figure 2. RGB images taken at the same location every hour of a 24-hour period [16]



Figure 3. RGB images taken at the same location every hour of a 24-hour period [16].

Even though, Phos image datasets are valuable and is freely accessible for public use, the uniform illumination is in a controlled environment. Under and over-exposed pixels in RGB images are the major source of artefacts in the illumination invariant, particularly during daylight hours [16]. Therefore, our first contribution in this paper is the collection of uniform illumination images under natural source illuminations. Our uniform illumination images are ear images captured in uncontrolled indoor and outdoor environment. We choose ear biometrics domain as ear recognition using deep CNN is still scarce due to lack of large-scale datasets [19]. At the time of writing, only four published work [19]-[22] of employing deep learning for ear recognition are found, with two papers from the same team. Although they also utilized CNN architecture, but the datasets used are under controlled environment or ear images with non-uniform illumination captured in the wild. Consequently, this paper intends to investigate the accuracy of CNN for ear biometrics captured under uniform illumination invariant environment.

## 2. THE UNIFORM ILLUMINANT INVARIANT EAR IMAGE DATASET

Illumination is the amount of light present in an area [23] and is measured in lux (lx) or lumen (lm) per square meter. Lux, also known as luminance is a measurement of brightness measured using a lux meter. The uniqueness of our datasets is that the ear images are acquired in the indoor and outdoor environments, and illuminations are objectively quantified using lux meter at actual locations. The acquisition process started with the subject seated on a provided chair. The subject's ear is faced towards the camera, and the camera is set to be at the same level with the ears. Then, the illumination is measured using lux meter. The lux meter sensor is placed directly under the ear to ensure an accurate reading. Details of the image acquisition can be found in our previous work [24]. For the uniform illumination variance, we managed to collect ear images with illuminations ranging from 2 lux to 10700 lux from 55 subjects. Ten ear images are collected from each subject, with various lux values depending on the surrounding lighting. The final

**5**61

compilation of ear dataset consists of a total of 1,100 images comprising 550 left and 550 right images. Table 1 lists the collection of ear images in different lux for both left and right ear images. Meanwhile, Figure 4 shows some examples of the ear images in selected lux values. The resolution for the ear images are 2048 x 1056 pixels.

| Table 1. Lux variations of the ear image dataset |      |       |       |           |      |       |       |             |      |       |       |
|--|------|-------|-------|-----------|------|-------|-------|-------------|------|-------|-------|
| Lux  | Left | Right | Total | Lux       | Left | Right | Total | Lux         | Left | Right | Total |
| 1-10   | 130  | 106   | 236   | 101-200   | 52   | 50    | 102   | 2001-3000   | 16   | 14    | 30    |
| 11-20  | 54   | 80    | 134   | 201-300   | 34   | 32    | 66    | 3001-4000   | 4    | 6     | 10    |
| 21-30  | 32   | 28    | 60    | 301-400   | 48   | 50    | 98    | 4001-5000   | 8    | 8     | 16    |
| 31-40  | 6    | 6     | 12    | 401-500   | 2    | 8     | 10    | 5001-6000   | 6    | 6     | 12    |
| 41-50  | 30   | 36    | 66    | 501-600   | 4    | 6     | 10    | 6001-7000   | 0    | 2     | 2     |
| 51-60  | 14   | 8     | 22    | 601-700   | 6    | 4     | 10    | 7001-8000   | 6    | 2     | 8     |
| 61-70  | 8    | 4     | 12    | 701-800   | 4    | 4     | 8     | 8001-9000   | 2    | 2     | 4     |
| 71-80  | 10   | 18    | 28    | 801-900   | 6    | 4     | 10    | 9001-10000  | 0    | 0     | 0     |
| 81-90  | 22   | 16    | 38    | 901-1000  | 2    | 6     | 8     | 10001-11000 | 0    | 2     | 2     |
| 91-100   | 12   | 14    | 26    | 1001-2000 | 32   | 28    | 60    | Total       | 550  | 550   | 1100  |



Figure 4. Our uniform illumination-invariant ear image dataset

Data augmentation using rotation at every  $5^{\circ}$  is applied to the ear images to generate adequate data training and testing for the CNN model. After data augmentation, a total of 25,300 images are generated and for each subject, we chose approximately 88.3% (i.e. 410 images per subject) to be used as training dataset and the remaining 11.7% (50 images per subject) is used for validation and testing datasets.

## 3. PROPOSED CNN MODEL

In this paper, we present the use of deep convolutional neural networks (CNN) to identify a person based on ear biometrics. On the basis of an analysis of structure and parameters in CNN, the gradient-descent algorithm can be applied to train CNN. A total of 25000 ear images are resized to 125x125 to correlate with the size of the input layer, and then the resized images are used to train the CNN model. This CNNs model can improve the convergence speed while training the parameters in CNNs, and obtain higher recognition accuracy compared to conventional neural network or support vecto machine. The advantages of the proposed CNN are that images can be input directly to the model. Figure 5 demonstrates the proposed CNN model and the layers are subsequently described in the following subsections.

Can Convolution Neural Network (CNN) Triumph in Ear Recognition of Uniform... (Nursuriati Jamil)

# 562

## **3.1. Convolutional Layers**

First, an image input layer inputs images to a network and applies data normalization. Then, we can classify images from a reduced data set applying convolution and pooling [25]. The details of the first convolutional layer are presented in Table 2.



Figure 5. Our proposed CNN model

| Table 2. First convolutional layer of our proposed CNN |                     |  |  |  |
|--|---------------------|--|--|--|
| Layer Type   | Layer Detail        |  |  |  |
| Input image  | (125,125,1)         |  |  |  |
| Filters  | 16                  |  |  |  |
| Filter size  | 3*3                 |  |  |  |
| Number of weights per filter                           | 3*3*1=9             |  |  |  |
| Number of parameters in the layer                      | (9+1)*16=160        |  |  |  |
| Stride   | 1                   |  |  |  |
| Number of neurons in each feature map                  | (125-3+2*1)/1+1=125 |  |  |  |
| Total number of neurons in the layer                   | 125*125*16=250000   |  |  |  |

Secondly, a batch normalization layer is added after the convolutional layer to speed up training of convolutional neural networks and reduce the sensitivity to network initialization. Batch Normalization allows us to use much higher learning rates and be less careful about initialization. It also acts as a regularizer, in some cases eliminating the need for Dropout [26]. Thirdly, Rectified Linear Units (ReLU) is used as the activation function. The ReLU layer performs a threshold operation to each pixel element of the input as seen in Equation 1 [27], where f(x) is the image and x is the pixel element. If the pixel element, x is less than zero it will be reset to 0. Otherwise, the pixel element remains as is.

$$f(x) = \begin{cases} x, & x \ge 0\\ 0, & x < 0 \end{cases}$$
(1)

Fourthly, a max pooling layer performs down-sampling by dividing the input into rectangular pooling regions, and computing the maximum of each region. We configured this layer to use  $4 \times 4$  pooling region size and 2 for the stride. This structure is repeated 3 times to form 3 folds of layers. The max pooling layer is placed between each of the two folds as illustrated in Figure 5.

Table 3 shows the layer details of the second convolutional layer. It consists of 32 filters of size  $6 \times 6$  and padding of 1. The convolutional layer is further followed by a pooling layer comprising  $2 \times 2$  pooling region size and 2 for the stride. The third convolutional layer contains 64 filters and the size of each filter is  $16 \times 16$  as can be seen in Table 4.

Indonesian J Elec Eng & Comp Sci, Vol. 11, No. 2, August 2018: 558 - 566

| Layer Type                            | Layer Details       |
|---------------------------------------|---------------------|
| Filters                               | 32                  |
| Filter size                           | 6×6                 |
| Number of weights per filter          | 6×6×1=36            |
| Number of parameters in the layer     | (36+1)*32=1184      |
| Stride                                | 1                   |
| Number of neurons in each feature map | (125-32+2*1)/1+1=96 |
| Total number of neurons in the layer  | 96*96*32=294912     |
| Total number of neurons in the layer  | 96*96*32=294912     |
| Table 4. Third convolutional layer    | of our proposed CNN |
| Layer Type                            | Layer Details       |
|                                       |                     |

| Layer Type                            | Luyer Details       |
|---------------------------------------|---------------------|
| Filters                               | 64                  |
| Filter size                           | 16×16               |
| Number of weights per filter          | 16×16×1=256         |
| Number of parameters in the layer     | (256+1)*64=16448    |
| Stride                                | 1                   |
| Number of neurons in each feature map | (125-64+2*1)/1+1=64 |
| Total number of neurons in the layer  | 64*64*64=262144     |

#### 3.2. Fully-Connected Layer

In CNN, all the convolutional layers and pooling layers are followed by several fully-connected layers used for the determination in the output layer. The fully-connected layer can be considered as the hidden-layer in artificial neural network. All the neurons in the fully-connected layer are fully connected to the neurons of both previous and next layers. To reduce overfitting in the fully connected layer, a regularization method called "dropout" [28] is usually employed. Dropout technique enables the complex co-adaptations of neurons to reduce greatly. Therefore, it is forced to learn more robust features in an image and exhibits substantial overfitting. The output of the last fully connected layer is fed to a Softmax layer [29]. This layer has 55 outputs, which produce probability distributions calculated using Equation 2. The probability is used to label each class, that is the recognized person. Hence, probabilities vector of size  $1 \times 55$  where each vector element corresponds to a class of dataset is obtained.

$$p_j = \frac{e^{x_j}}{\sum_{i=1}^{k} e^{x_k}}, \text{ for } j = 1..k$$
<sup>(2)</sup>

where x is the net input. Output values of p are between 0 and 1 and their sum equals to 1.

#### 3.3. Classification Layer

This layer is designated to count the error (loss) using cross-entropy function for 1-of-k coding scheme as in Equation 3.

$$E(\theta) = -\sum_{i=1}^{n} \sum_{j=1}^{k} t_{ji} \ln y_j(x_i, \theta)$$
(3)

where  $\theta$  is the parameter vector,  $t_{ji}$  is the indicator that the *i*th sample belongs to the *j*th class, and  $y_j(x_i, \theta)$  is the output for the sample, *i*. The output  $y_j(x_i, \theta)$  can be interpreted as the probability that the network associates *i*th input with class, *j*, that is,  $P(t_j = 1|x_i)$ . The output unit activation function is the Softmax function.

### 4. **RESULTS AND DISCUSSIONS**

The proposed CNN model is trained using batches of images containing 60 images, with learning rate fixed at 0.001 to achieve a reliable training. We used Stochastic Gradient Descent with Momentum for training the CNN. The value of momentum is 0.9 inspired by [29]-[31]. The maximum number of epoch is set to 50 to ensure network convergence. The training and validation outcomes are presented in Figure 6. Since the CNN model is trained from scratch, the recognition accuracy started with a value close to 0 and gradually improves along with the number of iterations. After iteration 5,000, the accuracy stabilized at close

564 🗖

to value 100%. Therefore, the training is stopped at validation accuracy of 99.76% as further training is deemed unnecessary. The loss rate is illustrated in Figure 7. At the early stage of training, the loss rate is rather high at a value of close to 60. As the training proceeded, the loss rate gradually decreases eventually reaching to loss value close to 0.



Figure 6. Training and validation of the proposed CNN model



Figure 7. Loss rate of the proposed CNN model



Figure 8. Wrongly recognized subjects with its corresponding luminance value

Indonesian J Elec Eng & Comp Sci, Vol. 11, No. 2, August 2018: 558 - 566

For testing, a total of 2,750 ear images are used for recognition and the result shows that 2,742 images are correctly recognized. Figure 8 illustrates the wrongly recognized 8 subjects with its corresponding luminance value. As can be seen, with the exception of one image, all the wrongly recognized images have lux values of less than 10. In total, there are 236 ear images (Refer Table 1) with the lux values of less than 10, and our proposed CNN model is able to recognize 229 images (97%) correctly. Even though the ear in the image is not even visible by naked eyes, the proposed model is able to correctly recognized majority of the subjects. As for subject #38, the augmented left ear of subject #14 is wrongly recognized even though adequate luminance of 22 lux is recorded for the image. This is probably due to the rotated operation performed on the image.

# 5. CONCLUSION

Deep learning using convolution neural network (CNN) has shown to produce exceptional performance over traditional methods in many application domains [32]. In this paper, a CNN model is built and trained for ear biometrics in various uniform illuminations measured using lumens. As far as we know, this is the first work of testing the performance of CNN on very underexposed or overexposed images. Our results showed that for images uniformly illuminated with luminance of above 25 lux, 100% recognition accuracy is achieved. Even though the CNN model has problems recognizing few images of below 10 lux, the overall accuracy of 97% suggests that CNN architecture performs equally well for uniform illumination-invariant images. However, data augmentation for our dataset only involves rotations. More robust operations should be included in data augmentation to ultimately test the performance of CNN model for ear biometrics.

### ACKNOWLEDGEMENTS

Due acknowledgment is accorded to the Research Management Center and Faculty of Computer & Mathematical Sciences, Universiti Teknologi MARA, for the funding received through the BESTARI Grant, (600-IRMI/DANA 5/3/BESTARI (0002/2016) and Universitii Tenaga Nasional.

#### REFERENCES

- [1] Hubel DH, Wiesel TN. Receptive Fields of Single Neurones in the Cat's Striate Cortex. *The Journal of. Physiology*. 1959; 148: 574-591.
- [2] Fukushima K. Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position," *Biol. Cybern.* 1980; 36(4): 193–202.
- [3] LeCun Y, Bengio Y. Convolutional networks for images, speech, and time series. *The Handbook of Brain Theory and Neural Networks*. 1995 Apr; 3361(10):1995.
- [4] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-Based Learning Applied to Document Recognition. Proceedings of the IEEE. 1998 Nov; 86(11): 2278-324.
- [5] LeCun Y, Huang FJ, Bottou L. *Learning Methods for Generic Object Recognition with Invariance to Pose and Lighting*. Proceedings of the 2004 IEEE Computer Vision and Pattern Recognition, 2004; 2:II-104.
- [6] Hecht-Nielsen R. Theory of the Backpropagation Neural Network. In Neural Networks for Perception; 1992: 65-93.
- [7] Sun M, Song Z, Jiang X, Pan J, Pang Y. Learning Pooling for Convolutional Neural Network. *Neurocomputing*. 2017 Feb 8; 224: 96-104.
- [8] Scherer D, Müller A, Behnke S. Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition. International Conference on Artificial Neural Networks, Springer, Berlin, Heidelberg; 2010 Sep 15: 92-101.
- [9] Krizhevsky A, Sutskever I, Hinton GE. Imagenet Classification with Deep Convolutional Neural Networks. In Advances in Neural Information Processing Systems; 2012: 1097-1105.
- [10] Lee SH, Chan CS, Mayo SJ, Remagnino P. How Deep Learning Extracts and Learns Leaf Features for Plant Classification. *Pattern Recognition*. 2017 Nov 30; 71: 1-3.
- [11] Boufenar C, Kerboua A, Batouche M. Investigation on Deep Learning for Off-line Handwritten Arabic Character Recognition. Cognitive Systems Research. 2017 Nov 13.
- [12] Bakry A, Elhoseiny M, El-Gaaly T, Elgammal A. *Digging Deep into the Layers of CNNs: In Search of How CNN Achieve View Invariance*. Proceedings of the International Conference on Learning Representations (ICLR 2016) arXiv preprint arXiv:1508.01983. Cornell University Library. 2015 Aug 9.
- [13] Lai K, Bo L, Ren X, Fox D. A Large-Scale Hierarchical Multi-view RGB-D Object Dataset. Proceedings of the 2011 IEEE International Conference on Robotics and Automation (ICRA), 2011 May 9; 1817-1824.
- [14] Xiang Y, Mottaghi R, Savarese S. Beyond Pascal: A Benchmark for 3D Object Detection in the Wild. 2014 Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), 2014 Mar 24; 75-82.
- [15] Borji A, Izadi S, Itti L. What Can We Learn About CNNs From a Large Scale Controlled Object Dataset? arXiv preprint arXiv:1512.01320. Cornell University Library. 2015 Dec 4.

Can Convolution Neural Network (CNN) Triumph in Ear Recognition of Uniform... (Nursuriati Jamil)

- [16] Maddern W, Stewart A, McManus C, Upcroft B, Churchill W, Newman P. Illumination Invariant Imaging: Applications in Robust Vision-Based Localisation, Mapping and Classification for Autonomous Vehicles. In Proceedings of the IEEE International Conference on Robotics and Automation Visual Place Recognition in Changing Environments Workshop, (ICRA). Hong Kong, China. 2014, May; 2: 3.
- [17] Yong L, Yuanzhi C. A Local Feature Descriptor Invariant to Complex Illumination Changes. *Telkomnika* (Telecommunication Computing Electronics and Control). 2016 Jun 2; 14(2A): 289-97.
- [18] Vonikakis V, Chrysostomou D, Kouskouridas R, Gasteratos A. A Biologically Inspired Scale-Space for Illumination Invariant Feature Detection. *Measurement Science and Technology*. 2013 Jun 12; 24(7):074024.
- [19] Emeršič Ž, Štepec D, Štruc V, Peer P. Training Convolutional Neural Networks with Limited Training Data for Ear Recognition in the Wild. arXiv preprint arXiv:1711.09952. Cornell University Library. 2017 Nov 27.
- [20] Emeršič Ž, Štruc V, Peer P. Ear Recognition: More Than a Survey. *Neurocomputing*. 2017 Sep 13; 255: 26-39.
- [21] Zhang Y, Mu Z. Ear Detection Under Uncontrolled Conditions with Multiple Scale Faster Region-Based Convolutional Neural Networks. *Symmetry*. 2017 Apr 10; 9(4): 53.
- [22] Galdámez PL, Raveane W, Arrieta AG. A Brief Review of the Ear Recognition Process using Deep Neural Networks. *Journal of Applied Logic*. 2017 Nov 1; 24: 62-70.
- [23] Devlin K, Chalmers A, Reinhard E. Visual Calibration and Correction for Ambient Illumination. ACM Transactions on Applied Perception (TAP). 2006 Oct 1; 3(4): 429-452.
- [24] Ariffin SM, Jamil N, Rahman PN. DIAST Variability Illuminated Thermal and Visible Ear Images Datasets. In Proceedings of the IEEE Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), 2016 Sep 21; 191-195.
- [25] Krizhevsky A, Sutskever I, Hinton GE. Imagenet Classification with Deep Convolutional Neural Networks. In Advances in Neural Information Processing Systems 2012; 1097-1105.
- [26] Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In Proceedings of International conference on Machine Learning, 2015 Jun 1; 448-456.
- [27] Nair V, Hinton GE. Rectified Linear Units Improve Restricted Boltzmann Machines. In Proceedings of the 27th international Conference on Machine Learning (ICML-10), 2010; 807-814.
- [28] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *The Journal of Machine Learning Research*. 2014 Jan 1; 15(1): 1929-58.
- [29] Christopher MB. Pattern Recognition and Machine Learning. Springer-Verlag New York; 2016.
- [30] Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, Liu T, Wang X, Wang G, Cai J, Chen T. Recent Advances in Convolutional Neural Networks. *Pattern Recognition*. 2017 Oct 11.
- [31] Deng W, Fang Y, Xu Z, Hu J. Facial Landmark Localization by Enhanced Convolutional Neural Network. *Neurocomputing*. 2018 Jan 17; 273: 222-9.
- [32] Borodo SM, Shamsuddin SM, Hasan S. Big Data Platforms and Techniques. *Indonesian Journal of Electrical Engineering and Computer Science*. 2016 Jan 1; 1(1): 191-200.