



A Comparison between Average and Max-Pooling in Convolutional Neural Network for Scoliosis Classification

Nurbaiti Sabri¹, Haza Nuzly Abdull Hamed²,
Zaidah Ibrahim³, Kamalnizat Ibrahim⁴

¹Faculty of Computer and Mathematical Sciences,

Universiti Teknologi MARA Cawangan Melaka Kampus Jasin, 77300 Merlimau, Melaka, Malaysia.

^{1,2}School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia

³Faculty of Computer and Mathematical Sciences, University Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

⁴Fakulti Perubatan & Kesihatan Pertahanan, Universiti Pertahanan Nasional Malaysia (UPNM), Sungai Besi, 57000 Kuala Lumpur

ABSTRACT

The present study carried out a comparison between average and max-pooling in Convolutional Neural Network for scoliosis classification. In the past, around 2 to 4 per cent of adolescence has been reported to suffer with scoliosis. Currently, radiographic is the clinical approach used in identifying the Cobb angle to determine the suitable treatment for this category of patients. However, over exposure to radiographic have been seen to what is leading to the risk of cancer. As such, the present study proposed the used of photogrammetric approach to overcome the radiographic side effect. The photogrammetric of human's back is acquired to classify the scoliosis into Lenke Type 1 or Non-Type 1. Due to limited dataset, rotation, x-transition and y-transition of data augmentation was carried out. These data are classified using convolutional neural network. The convolutional neural network (CNN) consist of convolve layer, pooling layer, fully connected layer and softmax layer. Selection of the best pooling layer is important to increase the accuracy of classification. As mentioned earlier, the present study compares between average and max-pooling layer to classify the Lenke classification system. The result shows that the use of max-pooling can achieve a higher accuracy which is 84.6% compared to average pooling. Future studies are encouraged to collect more data to further prove the effectiveness of max-pooling layer.

Key words: Scoliosis, Lenke, Convolutional Neural Network, Max-pooling, Average pooling.

1. INTRODUCTION

Scoliosis is a structural three-dimensional deformity of the spine determine by more than 10 degrees of lateral curvature

[1]. The work to assess and monitor early-onset scoliosis or also known as spinal deformity in patients has been studied for less than 10 years [2]. It was reported that around 2 to 4 percent of adolescent idiopathic scoliosis (AIS) suffer from this condition and between 70 and 80 percent does not know the cause of scoliosis. It is important to identify scoliosis early by classify the different spine deformities, to assign appropriate treatment for the patient. The treatment for each angle is different according to the degree of angle and some of the cases do need a surgery. However, radiographic is the widely used approach to determine the scoliosis curve. This method consists of entire spine, thoracic (upper back) and lumbar (lower back). Cobb angle is the formal standard to classify the scoliosis curve by their location along the spine and degree of the curve [3].

Conversely, the use of radiographic image might increase risk of cancer associated with the high number of radiographs taken to measure scoliosis progression [4]. According to [5], patient that is diagnosed with mild scoliosis are ten (10) times more common compared to the serious cases of scoliosis. Also, there exist the subjects of unnecessary high exposure to radiographic. In which many alternative solutions have been introduced to overcome this problem. One of the first of such approach is moiré contour which is to replace radiographic image to identify Cobb angle from back surface [6]. This is followed by rasterstereography [7], the integrated shape imaging system (ISIS) [8] and the Quantec scanner [9]. Nevertheless, these approaches are still behind expectations because of the limitation of available technology.

On the other hand, photogrammetry is a comfort technique of image acquisition where it avoids a physical contact between patient and devices. It provides fast data capture compared to existing technique, such as radiographic or CAT scanning [10]. The image of back body surface is taken and processed using computer vision to start gaining attention in order to overcome the disadvantage of common radiographic approach for scoliosis angle detection.

Also, deep learning has been a very well know machine learning technique where numbers of higher accuracy have been reported on many domains, such as image analysis, recognition and more [11] [12][13]. Among all the techniques in deep learning, Convolutional Neural Network (CNN) is the most popular that produces a promising result with a simple and straight forward architecture compared to other deep learning techniques. The CNN architecture consists of pooling layer where it converts the joint feature representation by keeping the important information and remove the unwanted details [14]. The most common pooling are max and average pooling in which both are working in different ways.

Studies also show that, max-pooling has been used to classify ImageNet dataset, in which this pooling layer has achieved a high accuracy [15]. However, it only identifies the maximum element and ignores others pooling region. Meanwhile, average pooling proves an excellent image classification in Caltech101 dataset [16]. But the low magnitudes are also taken into consideration and contract reduction of the new feature map is produced after pooling. The most significant lost if the element is zero, is that the characteristic of the feature map will generally be reduced. In all, both pooling works differently where sparse coding and max-pooling method is better than the average pooling method, but not for the resolution in visual tracking [17]. Besides, vary result appeared according to different problems [14].

Based on this, the current study carried out a comparison between average and max-pooling in convolutional neural network for scoliosis classification using photogrammetric images of human’s back to identify the Lenke Type 1 and Non-Lenke type 1.

2. RESEARCH METHODOLOGY

In the present research, experiment on max-pooling and average pooling is carried out on photogrammetric image of the human’s back. Figure 1 shows the framework for reading photogrammetric image of the human’s back in the current experiment. This involves evaluating forty-five (45) images of the human’s back of AIS. Data are gathered from Pusat Perubatan University Malaya, Selangor, which consisted of Lenke type 1 and other images. Thereafter, the data augmentation process was executed due to the limited number of data and the requirement of convolutional neural network algorithm to train a large number of datasets. Also, Figure 2 shows the sample of scoliosis image of Lenke type 1. Meanwhile, Figure 3 shows the sample for Non-Lenke Type 1. It can be seen that there are large differences in photogrammetric image of both lenke and Non-Lenke Type 1 proven on the radiographic image, where for lenke type 1, it only consists of 1 major curve with more than 10 degrees. Meanwhile, for non-lenke type 1, it consists of more than 2

major curves in one radiographic image. As a result of this huge differences, coupled with the fact that the most common case in Malaysia is Lenke type 1, the present research focused on identifying lenke type 1 from the photogrammetric images of human’s back.

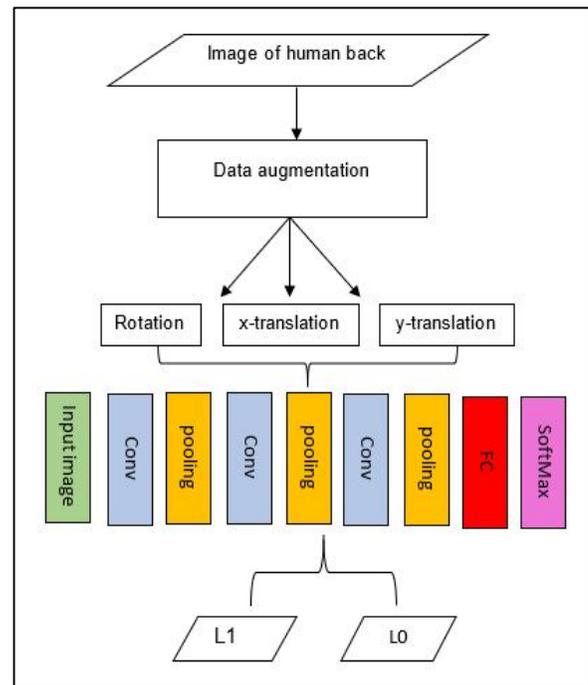


Figure 1: The framework for Lenke Type 1 classification

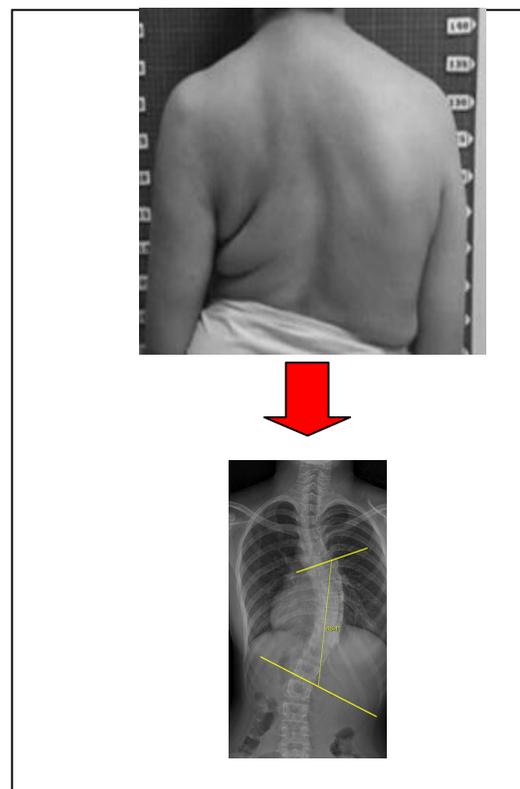


Figure 2: Sample photogrammetric image of Lenke type 1 and the radiographic image of Lenke type 1.

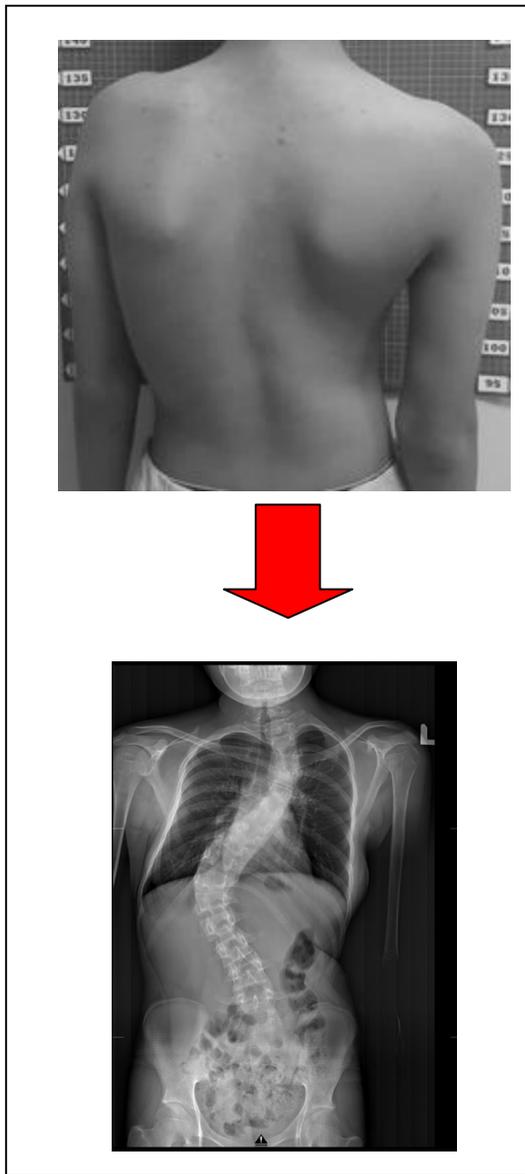


Figure 3: Sample photogrammetric image of Non-Lenke Type 1 and the radiographic image of Non-Lenke Type 1

2.1 Lenke Classification systems

A Lenke classification system was introduced by Lawrence Lenke in 2001 to define the curve types of scoliosis patients. In order to identify the curve type using Lenke classification system, the curve type (Cobb angle), lumbar modifier and sagittal profile need to be identified. There are six (6) different curve types in Lenke classification as listed in Table 1.

Table 1: Description of Lenke classification system.

Lenke Type	Structural	Non-structural
Type 1	main thoracic (major curve)	proximal thoracic lumbar or thoracolumbar
Type 2	Proximal thoracic (minor curve) main thoracic (major curve)	lumbar or thoracolumbar
Type 3	main thoracic (major curve) lumbar or thoracolumbar (minor curve)	proximal thoracic
Type 4	main thoracic (minor curve) proximal thoracic lumbar or thoracolumbar	
Type 5	lumbar or thoracolumbar (major curve)	main thoracic proximal thoracic
Type 6	main thoracic (minor curve) lumbar or thoracolumbar (major curve)	proximal thoracic

According to Table 1 above, the largest Cobb measurement is known as major curve, which are Cobb angle 45 degree and above. Meanwhile, the minor curve is defined according to Cobb angle measurement between 25 degree and more. Figure 4 shows the Lenke classification system on spine representation.

Curve type (1-6)						
Lumbar deviation (A-C)	Type 1 Single thoracic	Type 2 Double thoracic	Type 3 Double major	Type 4 Triple curve	Type 5 Thor.-lumbar or lumbar	Type 6 Thor.-lumbar or lumbar
A minimal						
B moderate						
C severe						
Sagittal plane	Normal	Cerv.-thor. kyphosis >20°	Thor.-lumb. kyphosis >20°	Cerv.-thor+ thor.-lumb. kyphosis >20°		

Figure 4: The curve type of human spine according to Lenke classification system [1]

Three (3) types of augmentation are implemented which are rotation, xtranslation and ytranslation [18]. Figure 5 shows results after augmentation process. The augmentation image produces 177 images consist of sixty-eight (68) Lenke Type 1 and one hundred and nine (109) for Non Lenke Type 1. This is followed by the implementation of CNN which is carried out from input image, convolve layer, pooling layer, fully connected layer and softmax layer. After the convolutional process, the softmax layer is responsible to classify the input images into two classes namely L1 (Lenke Type 1) or L0 (Non Lenke Type 1). In this stage, comparison between average and max-pooling layer was carried out.

2.1 Data Augmentation

It is difficult to build a large medical image, especially on scoliosis where the health screening is not yet established in Malaysia. Due to a smaller number of dataset available, data augmentation was suggested to overcome issues in medical [19]. In the present research, rotation, x-translation and y-translation are implemented to support the limited number

of datasets. The total data of 177 from 45 original images were produced using this approach.

A. Rotation

In this process equation (1), an original image of human's back θ are rotated into (-20°) and $(+20^\circ)$ to generate new image samples. The 4 by 4 matrix of $\cos \theta$ and $\sin \theta$ are multiply with x and y axis. The x and y represent the location of pixel where x is horizontal and y is vertical, where it maps each pixel (x, y) of an image to (x', y') [20].

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad (1)$$

B. X-Translation and Y-Translation

In order to increase the number of datasets, the x and y translation are implemented. The x and y-translation in equation (2) are done using (x_0, y_0) , to be the movement of the two images f_1 and f_2 . Equation 2 shows the x and y-translation [19]. The images were shifted (-5) , which is left and $(+5)$ which is right up in horizontal for x position and y vertical position.

$$f_2(x, y) = f_1(x - x_0, y - y_0) \quad (2)$$

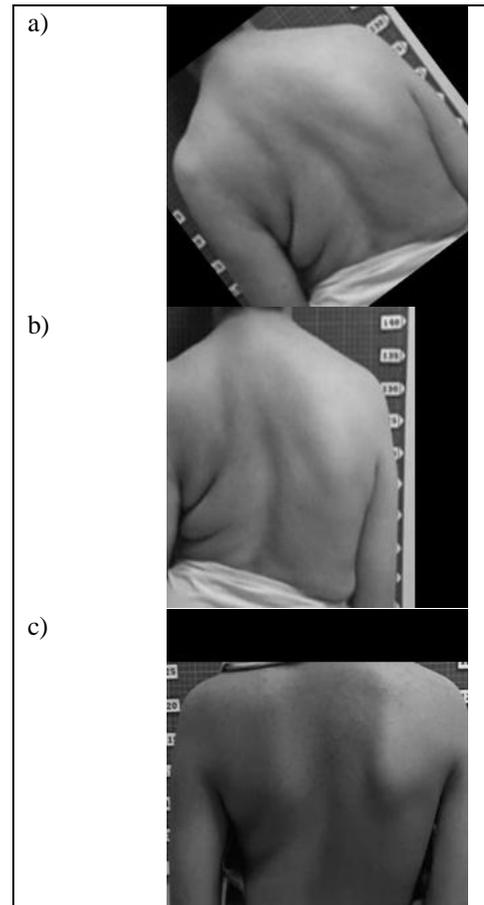


Figure 5: image after rotation, x-translation and y-translation

2.2 Convolutional Neural Network (CNN)

CNN is one of the deep learning architectures and has been implemented in many domains, such as in agriculture, signal processing and more. The basic architecture of CNN consists of 3 convolutional layers (1 fully connected layer and 1 softmax layer). The CNN input any color model and the present research focuses on 3 color channels which is Red, Green and Blue (RGB) images as an input. Experiment on three images sized 227 x 227, 180 x 180 and 160 x 160 were conducted [21]. It consists of three (3) convolve layers consisting of 20 filters sized 3x3. The padding consists of the computational of the size of the padding at training time. Therefore, the output will produce an equal size as the input when the stride equals 1. All neurons used a Rectified Linear Unit (ReLU) [22] with the weights initially being initialized from a Gaussian distribution with a 0 mean and a standard deviation of 0.01. The average pooling and max-pooling function of the window size two (2) and stride two (2) were applied, which further reduced the feature map. This process was executed repeatedly for three (3) times. Also, in the present research, three (3) convolve layers with the filter size of 3 by 3 were implemented. The deeper layer of convolutional affects the accuracy of the classification result. At the end, the final layer of max or average pooling was fed into a fully connected layer of size 2 which was connected to a softmax layer. The fully connected layer represents the number of classes to be classified which in this experiment is Lenke Type 1 or Non Lenke Type 1. Figure 6 shows the basic CNN architecture consisting of input layer, hidden layer and output layer.

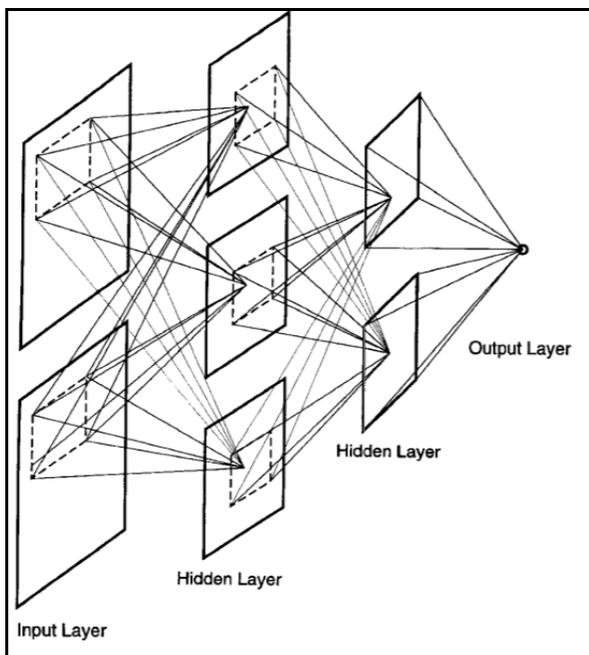


Figure 6: Basic CNN architecture [13]

2.3 Max-pooling Layer

In CNN, there are two conventional pooling methods, namely the max and average pooling. The collection of the largest element in each pooling region is done by max-pooling layer [18]. It translates invariance on each single cell and produces fully connected layers that require objects to be aligned. Max-pooling equation below shows an algorithm (3) where y_{kij} is the output of the pooling operator related to the k th feature map, and x_{kpq} is the element at (p, q) of R_{ij} which is pooling region that represents a local neighborhood around the position (i, j) [23] [24].

$$y_{kij} = \max_{(p, q) \in R_{ij}} x_{kpq} \quad (3)$$

2.4 Average Pooling Layer

The average pooling method takes the arithmetic mean of the elements in each pooling region as [18] where $|R_{ij}|$ stands for the size of the pooling region R_{ij} . This Equation (4) works by calculating the total pixel from each layer size and divides the total pixel from the same layer. The averages on each pixel are produced at the end of this pooling layer.

$$y_{kij} = \frac{1}{|R_{ij}|} \sum_{(p, q) \in R_{ij}} x_{kpq} \quad (4)$$

Therefore, a comparison to determine the best pooling layer between max-pooling layer and average pooling layer has been done. The result evaluated is based on the accuracy of classification on Lenke Type 1 and Non-Type 1.

3. RESULT AND ANALYSIS

The classification of Lenke Type 1 with non-Type 1 was done using seventy-seven (77) photogrammetric images of human's back after the augmentation process on original images. CNN algorithm was implemented to classify those images on three types of images sizes (180x180, 160x160 and 227x227). Additionally, the three (3) types of pooling layer (max-pooling, average pooling and max-average pooling) were experimented using three convolutional layers. Also, the pooling layer was tested on two (2) stages which were 2 max-pooling, 1 average pooling layer and 1 max-pooling and 2 average pooling. The results of the experiments are shown in Table 2.

Table 2: Result for 1 max-pooling

Max-pooling	
Image Size 180 x 180	84.6
Image Size 227 x 227	61.5
Image Size 160 x 160	61.5

Table 3: Result for 1 average pooling

Average Pooling	
Image Size 227 x 227	69.2
Image Size 160 x 160	76.9
Image Size 180 x 180	76.9

Table 4: Result for 2 max-pooling and 1 average pooling

2 Max-pooling and 1 Average Pooling	
Image Size 180 x 180	84.6
Image Size 227 x 227	53.8
Image Size 160 x 160	61.5

Table 5: Result for 1 max-pooling and 2 average pooling

1 Max-pooling and 2 Average Pooling	
Image Size 180 x 180	69.2
Image Size 227 x 227	76.9
Image Size 160 x 160	76.9

Table 2 until Table 5 shows the accuracy results of CNN with max-pooling layer, average pooling layer and max-average pooling layer while Figure 7 shows the comparison accuracy between all the combinations and different sizes of images used for testing purposes. It shows that the highest accuracy was achieved by using CNN with max-pooling which is

84.6% while, average pooling achieved 76.9% of accuracy. Furthermore, an experiment on max-average pooling shows a promising result of 84.6% using 2 max-pooling and 1 average pooling. It also shows that the use of a single max-pooling layer would result in the same finding as the combination of max-average pooling.

Meanwhile, the same result of single average pooling was achieved using 1 max-pooling and 2 average pooling. This is due to the image acquired using a photogrammetric image (the human's back) that is brighter with a darker background. This pooling layer collected all brighter pixels and ignores the darker pixels. However, lower result was obtained using average pooling due to the ability of this pooling to smooth the image, in which the sharp features were ignored to affect the decrement of the classification accuracy.

4. CONCLUSION

The present study carried out a comparison between average, max-pooling and max-average pooling in CNN for scoliosis classification in selecting the best pooling layer to classify Lenke Type 1 and Non-Lenke Type 1 of photogrammetric images of human's back. Findings show that by considering the maximum element of pixels, max-pooling layer was good enough to obtain a high accuracy. It also shows that the use of a suitable pooling method was important to increase the accuracy of classification. However, the implementation of average and max-pooling will be different according to the type of data and problem. Thus, more data is needed to be used for future study to experiment whether the number of data is crucial to produce a high accuracy classification model in CNN.

ACKNOWLEDGMENTS

This work is supported by the Ministry of Education Malaysia and Universiti Teknologi Malaysia through Research University Grant Scheme (Q.J130000.2651.16J63)

REFERENCES

1. S. Haleem, & C. Nnadi. **Scoliosis: a review**, *Symposium: surgery and orthopaedics*, 2018.
2. C. K. Hardesty, R. P. Huang, R. El-Hawary, A. Samdani, P. B. Hermida, T. Bas. **Early-Onset Scoliosis: Updated Treatment Techniques and Results**, *Spine Deformity* 6, pp. 467-472, 2018.
3. M. M. Adankon, J. Dansereau, H. Labelle and F. Cherieta. **Non Invasive Classification System of scoliosis curve types using Least-squares support vector machine**, *Artificial Intelligence in Medicine*, 2012.
4. I. Stokes and M. Moreland. **Concordance of back surface asymmetry and spine shape in idiopathic scoliosis**, *Spine*, vol. 14, pp. 73-78, 1989.

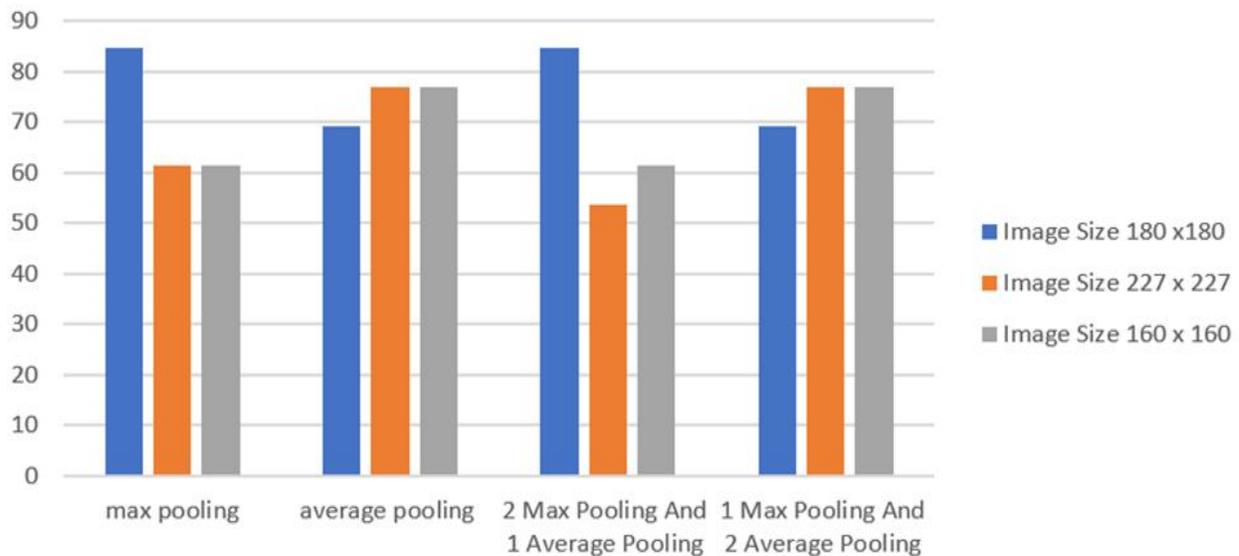


Figure 7: Comparison between max pooling, average pooling and image size.

5. A. Nachemson. **A long term follow-up study of non-treated scoliosis**, *Acta Orthopaedica Scandinavica*, vol. 39, pp. 466-476, 1968.
6. M. Batouche, R. Benlamri and M. Kholadi. **A computer vision system for diagnosing scoliosis using moirés images**, *Computers in Biology and Medicine*, vol. 26, pp. 339-353, 1996. [https://doi.org/10.1016/0010-4825\(96\)00014-5](https://doi.org/10.1016/0010-4825(96)00014-5)
7. J. A. Zubairi. **Application of computer-aided rasterstereography in spinal deformity detection**, *Image and Computer Vision Computing*, pp. 319-324, 2002.
8. A. Turner-Smith, J. Harris., G. Houghton, and R. Jefferson. **A method for analysis of back shape in scoliosis**, *Journal of biomechanics*, vol. 21, pp.497-509, 1988.
9. X. Liu , J. Thometz , R. Lyon , & J. Klein. **Functional classification of patients with idiopathic scoliosis assessed by the quantec system: a discriminant functional analysis to determine patient curve magnitude**, *Spine*, pp.1274–8, 2001.
10. H. Mitchell. **Application of digital photogrammetry to medical investigations**, *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 50, pp. 27-36, 1995.
11. W. Lumchanow, & S. Udomsiri. **Image classification of malaria using hybrid algorithms: convolutional neural network and method to find appropriate K for K-nearest neighbor**, *Indonesian Journal of Electrical Engineering and Computer Science*, 2019.
12. H. Sofian, J. Ming, S. Muhammad, & N. Mohd Noor. **Calcification detection using convolutional neural network architectures in Intravascular ultrasound images**, *Indonesian Journal of Electrical Engineering and Computer Science*, 2019 .
13. M. Zulqarnain, R. Ghazali, Y. Mohamad Hassim and M. Rehan. **A comparative review on deep learning models for text classification**, *Indonesian Journal of Electrical Engineering and Computer Science*, 2019. <https://doi.org/10.11591/ijeecs.v19.i1.pp325-335>
14. D. Yu, H. Wang, P. Chen, & Z. Wei. **Mixed Pooling for Convolutional Neural Networks**, *International Conference on Rough Sets and Knowledge Technology*, pp. 364-375, 2014.
15. A. Krizhevsky, I. Sutskever and G. E. Hinton. **ImageNet classification with deep convolutional neural networks**, *Advances in neural information processing systems*, pp.1097-1105, 2012.
16. K. Jarrett , K. Kavukcuoglu , M. Ranzato and Y. LeCun. **What is the best multi-stage architecture for object recognition?** , *IEEE 12th international conference on computer vision*, pp. 2146-2153, 2009.
17. Y.L.Boureau, J. Ponce and Y. LeCun. **A theoretical analysis of feature pooling in visual recognition**, *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010.
18. F. Calimeri, A. Marzullo, C. Stamile and G. Terracina. **Biomedical Data Augmentation using Generative Adversarial Neural Networks**, *International conference on artificial neural networks*, pp. 626-634, 2017.
19. C. Shorten and T. M. Khoshgoftaar. **A survey on Image Data Augmentation for Deep Learning**, *Journal of Big Data*, vol. 6, 2019.
20. S. Dielema, K. W. Willett and J. Dambre. **Rotation-invariant convolutional neural networks for galaxy morphology prediction**, *Monthly Notices of the Royal Astronomical Society*, Vol. 450, pp. 1441–1459, 2015.

21. K. Yanai, R. Tanno and K. Okamoto. **Efficient Mobile Implementation of A CNN-based Object Recognition System**, *Proceedings of the 24th ACM international conference on Multimedia*, pp. 362-366, 2016.
22. R. H. Hahnloser, R. Sarpeshkar, M. A. Mahowald, R. J. Douglas and S. H. Seung. **Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit**, *Nature*, 2000.
23. D. Scherer, A. Muller and S. Behnke. **Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition**, *20th International Conference on Artificial Neural Networks (ICANN)*, 2010.
24. S. Shindo, T. Goto, T. Kirishima and K. Tsuchida. **An optimization of facial feature point detection program by using several types of convolutional neural network**, *Indonesian Journal of Electrical Engineering and Computer Science*, pp. 827-834, 2019. <https://doi.org/10.11591/ijeecs.v16.i2.pp827-834>