The 34th Annual Conference of the Japanese Society for Artificial Intelligence, 2020

A New Convolution for a Deep Learning using Two-dimensional Geometric Distance

Michihiro Jinnai [*1] Kyle Armstrong [*2] Taisuke Sano [*3] Yasuhisa Uesugi [*4]

[*1] Nagoya Women's University [*2] University of Adelaide, Australia [*3] Quan Inc. [*4] ship shape LLC

For deep learning classification, a new convolution called the Geometric Distance, which numerically evaluates the degree of likeness between the input image and the filter on the convolution layer is proposed. Traditionally, the convolution known as the cosine similarity has been used widely to measure likeness. The traditional method does not perform well in the presence of noise or pattern distortions. In this paper, a new mathematical model for similarity is proposed that overcomes the limitations of the earlier model, and a new algorithm based on a one-to-many point mapping is proposed to realise the mathematical model. In the GD, when a "difference" occurs between peaks of the input pattern and the filter with a "wobble" due to noise, the "wobble" is absorbed and the distance metric increases monotonically according to the increase of the "difference". We performed numerical experiments and confirmed the effectiveness of the GD.

1. Introduction

In image recognition, deep learning methods such as convolutional neural network (CNN) has received much attention. In CNN, similarity is calculated on a convolution layer between the input image and the filter (Figure 1). Conventionally, cosine similarity has been used widely to measure the degree of likeness. Conventional cosine similarity compares the patterns using one-to-one mapping, but the resulting distance metric is highly sensitive to noise, and the distance metric changes in a staircase pattern when a difference occurs between peaks of the input image and the standard image (filter). As an improvement, we have developed a new convolution (similarity scale) called the "Geometric Distance (GD)" [1-7] as a superior alternative to cosine similarity. The GD is more accurate than the conventional cosine similarity in a noisy environment [8-13].

In sound recognition, a waveform of the microphone output can be converted into a sound spectrogram using the Linear Predictive Coefficient (LPC). This allows the waveform of the sound to be processed as a spectrogram image. In this paper, we propose a new convolution (similarity scale) for both image recognition and sound recognition in CNN.

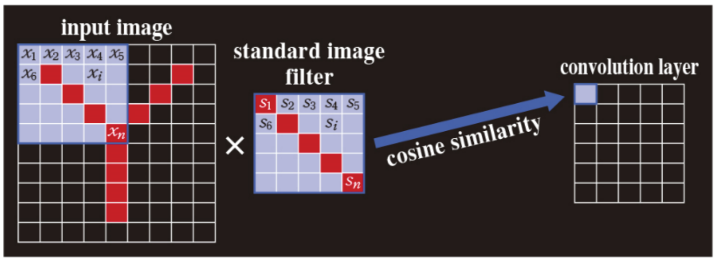


Figure 1. Convolution between input image and filter in CNN

Contact: Michihiro Jinnai, mjinnai@nagoya-wu.ac.jp

2. The LPC spectrogram of bat calls

Figure 2B shows the waveform of a bat echolocation call. Figure 2A shows the spectrogram (time-frequency-power) of the same signal. The waveform has been segmented with a 1.004 msec frame width and a 0.008 msec frame period, and the LPC spectrum has been calculated for each frame. Next, the spectrogram has been coloured according to the logarithmic power of the LPC spectrum. We have set the software analysis parameters for this example bat call recorded at 250 kHz sampling frequency, 16 bit quantization, to an LPC order of 11; restricted the spectral frequency range from 0 Hz to 125 kHz, with a 179.6 Hz frequency resolution; and set a dB threshold filter of 0 dB to -60 dB logarithmic power spectrum. If transient signals such as bat calls are analysed, a short frame width needs to be specified, as is shown at the bottom of Figure 2B. The LPC spectrum analysis is suitable for such transient signals.

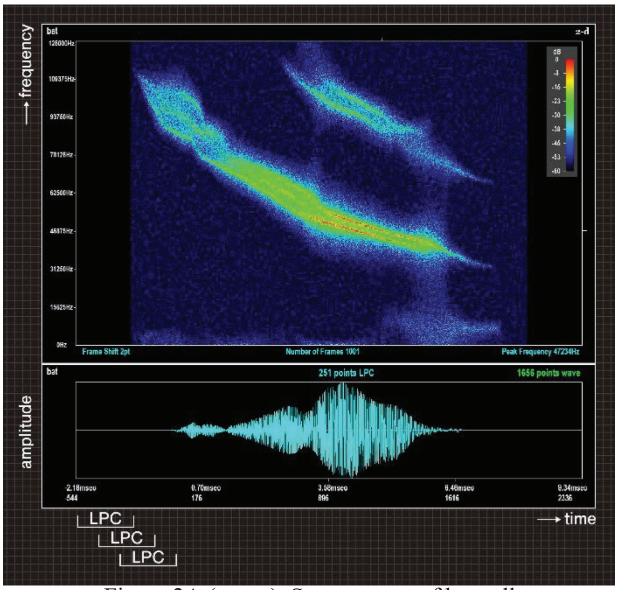


Figure 2A (upper). Spectrogram of bat call
Figure 2B (lower). Oscillogram of bat call

3. Conventional convolution

In CNN, similarity is calculated between the input pattern and the filter on a convolution layer for feature extraction. In conventional CNN, the convolution known as the cosine similarity has been used widely to measure likeness. The cosine similarity compares the input pattern with the standard pattern (filter in CNN) using one-to-one mapping. A possible result of one-to-one mapping is that input patterns with different shapes might have the same distance from the standard pattern when the patterns have the "difference" plus "wobble". This section describes the shortcomings that are found in the conventional cosine similarity using typical examples of spectrograms and images.

Here, we create a standard pattern vector $\boldsymbol{s}$ having s_i (i = 1, 2, …, n) components of the standard sound (or image), and an input pattern vector $\boldsymbol{x}$ having x_i (i = 1, 2, …, n) components of the input sound (or image), and represent them as follows.

$$\begin{aligned}\boldsymbol{s} &= (s_1, s_2, \cdots, s_i, \cdots, s_n)\\ \boldsymbol{x} &= (x_1, x_2, \cdots, x_i, \cdots, x_n)\end{aligned} \tag{1}$$

As shown in Figure 1, Equation (1) expresses the shapes of the standard sound (or image) and the input sound (or image) by the n pieces of component values of the power spectrum (or the image density), respectively. The cosine similarity is then calculated using the following Equation (2). Note that Figures 3B-6B show the angle θ, which is calculated using an arccosine in Equation (2).

$$\cos\theta = \frac{\sum_{i=1}^{n} s_i \cdot x_i}{\sqrt{\sum_{i=1}^{n} s_i^2}\sqrt{\sum_{i=1}^{n} x_i^2}} \tag{2}$$

In the spectrogram of the bat echolocation call, a peak of energy can be observed in one area (Figure 2A). In each call, the position of peak energy can vary according to both time and frequency, even amongst successive calls from the same individual bat.

● Figure 3A shows an example of the "difference" where the standard sound has two peaks in the spectrogram, and input sounds 1, 2 and 3 have a different position for the first peak. Note that both the standard and input sounds have the same volume. In Figure 3B, the bar graph on the left shows the cosine similarities θ_1, θ_2 and θ_3 between the standard sound and each of the input sounds 1, 2 and 3. Given that the cosine similarities have the relationship of θ_1=θ_2=θ_3, and therefore, the input sounds 1, 2 and 3 cannot be distinguished from one another.

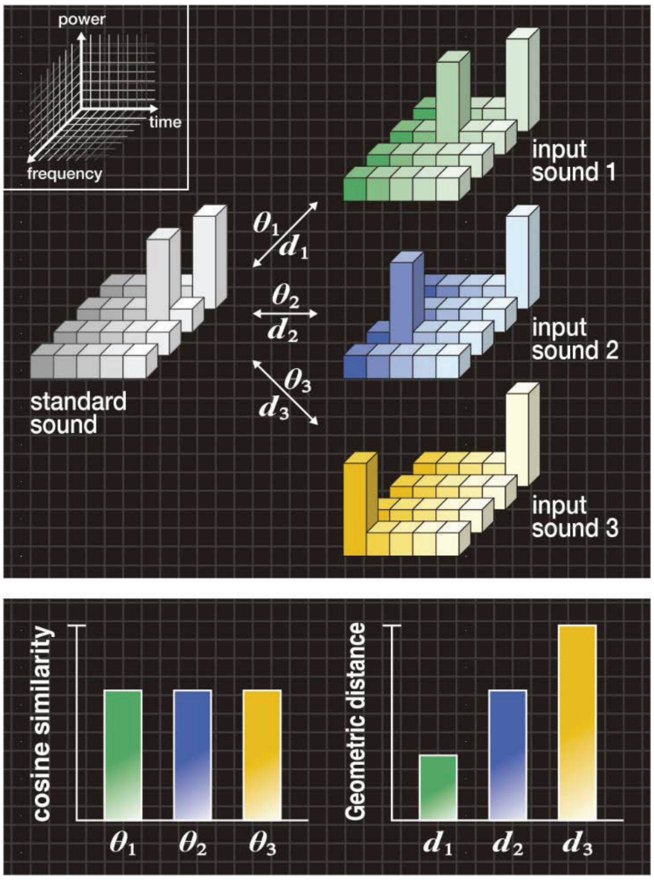


Figure 3A (upper). Typical example of "difference in peak"
Figure 3B (lower). Values of convolution

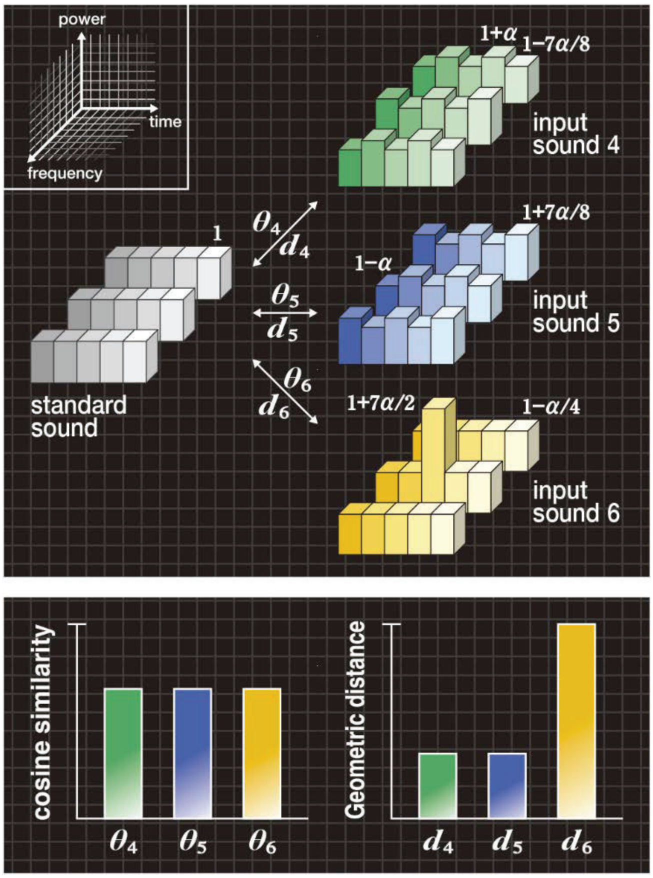


Figure 4A (upper). Typical example of "wobble"
Figure 4B (lower). Values of convolution

In the spectrogram of an example bat echolocation call (Figure 2A), a “spectrum intensity wobble” can be observed in a noisy environment. The “spectrum intensity wobble” occurs for every call, even amongst successive calls from the same individual bat.

● Figure 4A shows an example of the “wobble” where the standard sound has a flat spectrogram, input sounds 4 and 5 have the “wobble” on the flat spectrogram, and input sound 6 has a single peak. Each sound is assumed to have variable α, so the standard and input sounds always have the same volume. In Figure 4B, the bar graph on the left shows the cosine similarities θ_4, θ_5 and θ_6 between the standard sound and each of the input sounds 4, 5 and 6. The cosine similarities have the relationship of θ_4=θ_5=θ_6, and therefore, the input sounds 4, 5 and 6 cannot be distinguished from one another.

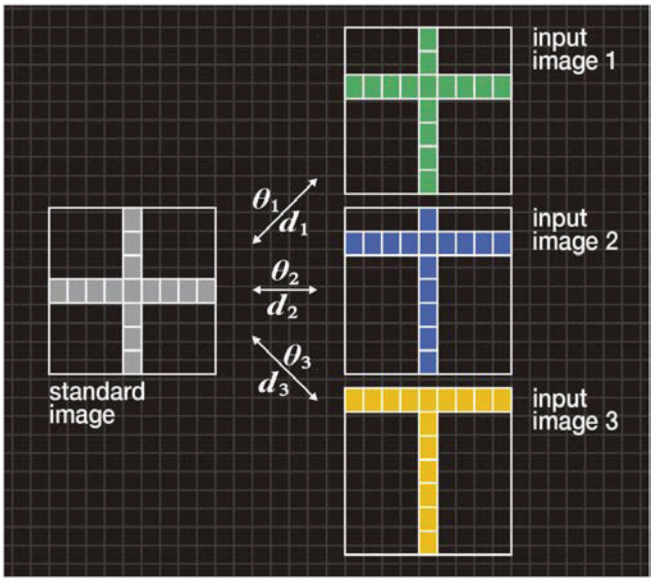
With respect to image recognition, a “difference in position” of line occurs in every handwritten character, even when the same character is written.

● Figure 5A shows an example of the “difference in position” where the standard image has a symbol “+”, and input images 1, 2 and 3 have a different position on the horizontal bar. In Figure 5B, the bar graph on the left shows the cosine similarities θ_1, θ_2 and θ_3 between the standard image and each of the input images 1, 2 and 3. The cosine similarities have the relationship of θ_1=θ_2=θ_3, and therefore, input images 1, 2 and 3 cannot be distinguished from one another.

“Character deformation” (changes in position and length of lines composing the character) occurs in every handwritten character, even when the same character is written.

● Figure 6A shows an example of “character deformation” where the standard image has the letter “E” and input images 4, 5 and 6 have the letters “E”, “F” and “G”, respectively. In Figure 6B, the bar graph on the left shows the cosine similarities θ_4, θ_5 and θ_6 between the standard image and each of the input images 4, 5 and 6. The cosine similarities have the relationship of θ_4>θ_5>θ_6, and therefore, the input letter “E” cannot be recognised correctly.

In order to overcome the above limitations of the conventional cosine similarity, we have developed a new convolution (a new similarity scale) called “Geometric Distance (GD)” [1-7]. With GD, when a “difference” occurs between peaks of the standard and input patterns with a “wobble” due to noise, the “wobble” is absorbed, and the distance metric increases monotonically according to the increase of the “difference”.



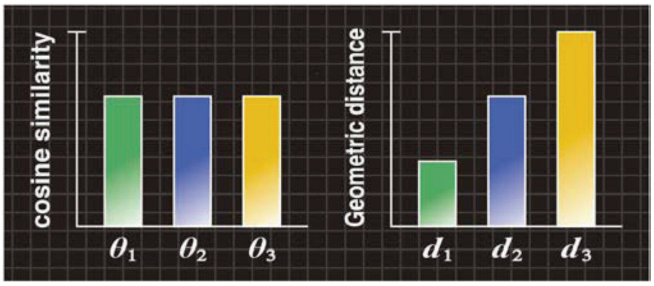
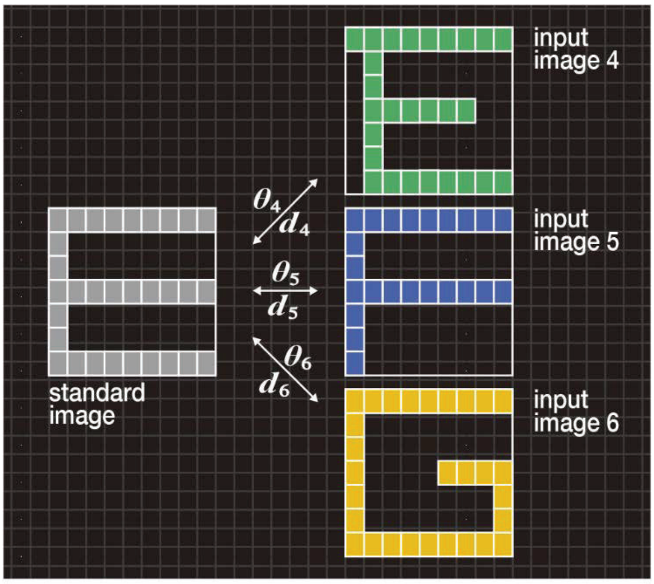


Figure 5A (upper). Typical example of “difference in position”
Figure 5B (lower). Values of convolution



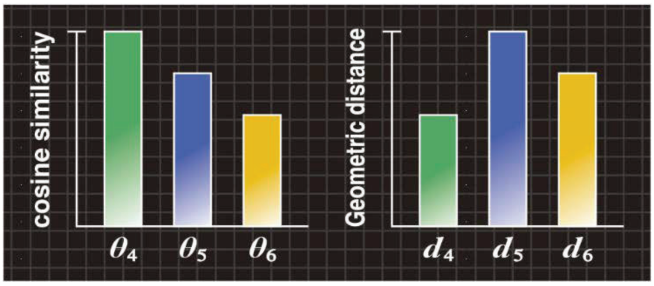


Figure 6A (upper). Typical example of “character deformation”
Figure 6B (lower). Values of convolution

4. Mathematical model for new convolution

A similarity scale is a concept that should concur intuitively with the human concept of similarity for hearing and sight. First we need to develop a mathematical model for the similarity scale so that we can perform numerical processing by computation. For GD, a mathematical model of similarity is proposed to improve the shortcomings that are found in the cosine similarity and other indices. For the calculation of GD, a mathematical model incorporating the following two characteristics is used:

<1> A distance metric that shows good immunity to noise.

<2> A distance metric that increases monotonically when a difference increases between peaks of the standard and input patterns.

Bar graphs on the right of Figures 3B, 4B and 5B express the characteristics <1> and <2> of the mathematical model.

5. New convolution algorithm

A new algorithm based on a one-to-many point mapping is proposed to realise the mathematical model. In statistical analysis, the normal distribution is often used for models exhibiting many phenomena. A "kurtosis" is used to verify whether the phenomenon obeys the normal distribution or not. In the GD, the difference in shapes between standard and input patterns is replaced by the shape change of the normal distribution, and the magnitude of this shape change is numerically evaluated as a variable of the kurtosis. Then, we obtain a new similarity scale (a new convolution) [1, 4].

Furthermore, in order to remove pseudo difference in shapes, we use a weighting vector that consists of a rate of change of the kurtosis, and create two weighted pattern vectors by performing the product-sum operation using the weighting vector and the standard pattern vector, and the product-sum operation using the weighting vector and the input pattern vector. Then, we use the angle between these weighted pattern vectors as the new Geometric Distance. This method reduces the processing overhead during an input pattern recognition process by separating the calculation of the shape change of the normal distribution into a standard pattern registration process and an input pattern recognition process [2, 6]. From the numerical experiments, we confirmed that the GD algorithm matches characteristics <1> and <2> of the mathematical model. Using GD, recognition experiments in bioacoustics were carried out in noisy environments [8, 10, 11, 13]. Furthermore, experiments in abnormal sound recognition of concrete structural integrity were conducted [12, 13]. In all cases, a significant improvement in recognition accuracy was demonstrated.

6. Conclusions and future work

We have described the shortcomings that are found in the conventional cosine similarity. We then introduced a mathematical model that overcomes the limitations of the earlier models. Of significance, we have proposed a new convolution algorithm based on a one-to-many point mapping to realise the mathematical model. In our future work, we will continue to improve the recognition software.

References

[1] M. Jinnai, S. Tsuge, S. Kuroiwa, F. Ren and M. Fukumi, "New Similarity Scale to Measure the Difference in Like Patterns with Noise", *International Journal of Advanced Intelligence*, Volume 1, Number 1, pp.59-88 (2009) http://aia-i.com/ijai/index.html

[2] M. Jinnai, S. Tsuge, S. Kuroiwa and M. Fukumi, "A New Geometric Distance Method to Remove Pseudo Difference in Shapes", *International Journal of Advanced Intelligence*, Volume 2, Number 1, pp.119-144 (2010) http://aia-i.com/ijai/index.html

[3] M. Jinnai, S. Tsuge, S. Kuroiwa, F. Ren and M. Fukumi, "A New Optimization Method of the Geometric Distance using Weighted Random Numbers", *International Journal of Advanced Intelligence*, Volume 4, Number 1, pp.133-154 (2012) http://aia-i.com/ijai/index.html

[4] M. Jinnai, H. Yamaguchi, Y. Ishihara, J. Ohshima and Y. Kidu, "Apparatus for detecting abnormal sound and method for judging wrong in machine", *Patent No.* US6170333(2001), JP3426905(2003)

[5] M. Jinnai and H. Yamaguchi, "Method for detecting similarity between standard information and input information", *Patent No.* US7006970(2006), JP3342864(2002), AU766266(2004), CA2357255(2010), EU1189207(2010), HK1045016(2010)

[6] M. Jinnai, "Method for detecting similarity between images, voices, oscillation waves, moving images and solids", *Patent No.* US7194133(2007), CA2422003(2008), JP3422787(2003), AU2003201379(2006)

[7] M. Jinnai, Y. Akashi, K. Hashimoto and S. Hayashi, "Method for detecting abnormal sound and method for judging abnormality in structure by use of detected value thereof", *Patent No.* US9552831(2017), AU2016200487(2017), CA2918533(2018), JP5956624(2016)

[8] M. Jinnai, N. Boucher, J. Robertson and S. Kleindorfer, "Design considerations in an automatic classification system for bird vocalisations using the Two-dimensional Geometric Distance and cluster analysis", *20th International Congress on Acoustics, ICA2010*, 130, (August 2010, Sydney) https://www.geometricdistance.com

[9] M. Jinnai, Y. Akashi, S. Nakaya, F. Ren and M. Fukumi, "Recognition of Abnormal Vibrational Responses of Signposts using the Two-dimensional Geometric Distance and Wilcoxon Test", *6th International Conference on Natural Language Processing and Knowledge Engineering, IEEE NLP-KE 2010*, 26, (August 2010, Beijing) https://www.geometricdistance.com

[10] M. Jinnai, N. Boucher, M. Fukumi and H. Taylor, "A new optimization method of the geometric distance in an automatic recognition system for bird vocalisations", *International Congress on Acoustics*, 105, (April 2012, Nantes, France) https://www.geometricdistance.com

[11] M. Jinnai, "A new segmentation method of bird vocalisations recorded in the distance", *2018 Autumn Meeting, Acoustical Society of Japan*, 3-1-5 https://www.geometricdistance.com

[12] M. Jinnai, K. Hashimoto, S. Hayashi and E. J. Pedersen, "Recognition of abnormal vibrational responses of concrete structures", *2019 Spring Meeting, Acoustical Society of Japan*, 1-6-15 https://www.geometricdistance.com

[13] M. Jinnai and E. J. Pedersen, "New similarity scale to recognize bird calls and abnormal sounds of concrete/machine", *The 33rd Annual Conference of the Japanese Society for Artificial Intelligence*, 2019, 3Rin2-26 https://www.geometricdistance.com