Image Recognition in New and Emerging Drugs (NEDs) Package with Convolutional Neural Network

Chih-Ping Yen

Department of Information Management, Central Police University,

Taoyuan 33304, Taiwan, ROC

ORCID: 0000-0002-1189-4922, peter@mail.cpu.edu.tw

Abstract

Today, many new and emerging drugs (NEDs) are packaged in a "foodized" way by mixing ingredients such as coffee, candy powder, jelly, etc. In an attempt to attract young people to take and avoid police investigations, which leads to the spread and abuse of drugs. Because people can't recognize the contents of the package as drugs, they are almost exposed to drugs without defense, and naïve teenagers are the most direct victims. In order to quickly identify whether it is an NED package that has been seized in the past by first-line staff. Based on deep learning, this study will propose a Multi-channel convolutional neural network (MCCNN) architecture to check the suspected packaging by taking photos with smart phones. We will use very few NEDs packaging images and applied 7 kinds of data augmentation methods, including brightness, scale, translation, shearing transformation, blur, rotation, and random crop to expand the training image data. Next, the proposed MCCNN method compares Uniform LBP (i.e. non-CNN) and other well-known CNN classification methods include AlexNet, VGG-16, VGG-19, GoogleNet, and ResNet. Finally, experiments have proven that the proposed MCCNN has the best accuracy compared with the non-CNN and state-of-the-art CNN methods, and reaches 98.56%.

Keywords: New and emerging drugs (NEDs), Deep learning, Multi-channel convolutional neural network (MCCNN), Data augmentation

1. Introduction

New and emerging drugs (NEDs), or new psychoactive substances (NPS), are also known by many other names, such as synthetic drugs, bath salts (monkey dust), herbal aromas, herbal highs, aphrodisiac teas, NBOM and legal higher drugs [1]. In addition,

NEDs may have cute names like Devil, Hello Kitty, Green Giant, Joker, N-bomb, or Flakka. Looking at the countries of the world, the proliferation of NEDs has become a global phenomenon. As of December 2018, the United Nations Office on Drugs and Crime (UNODC) reported a total of 888 NEDs reports [2]. In Taiwan, up to 132 new drugs have been seized as of 2018. These NEDs are often packaged in bright logos as shown in Figure 1 [3], and in "foodized" way by mixing ingredients such as coffee, candy powder, and jelly, in an attempt to attract young people to take. Coupled with the development of social media, NEDs are becoming more widespread. Then, according to the statistics on the types of drug abuse reported by medical institutions in Taiwan, the methamphetamine abuse ratios from suspected suspects' urine test between 2016 and 2018 was 29.4%, 34%, 45% respectively [3]. In this regard, the statistical data shows a trend of increasing significantly year by year.

For the above reasons, the motivation of this study is to propose a framework that assists the first-line staff to use the smartphone to quickly identify the NED package image. The query will quickly retrieve the input package image to identify whether it is an emerging drug packaging that has been seized in the past. This will increase the performance of investigating drug crimes in the future. In the identification of NED package images, the more similar research is logo recognition. Among the many logo recognition methods, the LBP-based approach is the most classic. In the past 3 years of logo-like recognition research, Shirazi et al. [4] directly applied the LBP method to establish the Persian logo recognition system; the results of the study showed that the recognition rate increased with the increase of the number of each class of training images. Wasim et al. [5] also believe that the LBP method is a very simple and reliable technology that can quickly and accurately identify objects of different shapes, so it is suitable for applications in the fields of surveillance, medicine, industry, and so forth. Yu et al. [6] proposed an OE-POEM based vehicle identification method, which is an improved version of the POEM descriptor; since LBP encodes rich edge information and the global structure of the object image, the POEM feature uses gradients in each direction and captures self-similarity between image regions by LBP. Sotoodeh et al. [7] also present two adaptive color descriptors for color image retrieval, named WCRMCLBP and WPDM based on LBP; which has the appropriate speed and the highest retrieval accuracy than the well-known methods by experiments.

In recent years, the introduction of deep learning has greatly improved the accuracy and efficiency of recognition. Therefore, more and more researchers have applied it to various fields, including computer vision, image recognition, image restoration, language processing, speech recognition, bioinformatics, visual art processing, robotics, etc. Thereby, deep learning has become the most popular method in machine learning [8] [9]. Based on deep learning, our work will propose a pattern recognition framework to assist first-line police officers in performing duties, such as pull over, spot check, and when checking for suspicious packaging, taking photos with smart phones to quickly identify whether it is a NED package that has been seized in the past, thereby increasing the police's ability for drug investigations.

The rest of the paper is organized as follows. In Section 2, we describe related work in pattern recognition methods, image processing, data augmentation, and Convolutional neural network (CNN). Section 3 proposes an architecture of Multi-channel CNN (MCCNN) and details each step. In Section 4, we describe the experimental NED package database, loss function, validation, and evaluated the MCCNN method we proposed. Finally, Section 6 concludes the whole paper.

Figure 1: NEDs are packaged in bright logos [3]

2. Related Work

2.1 Uniform local binary pattern (Uniform LBP)

LBP is the classic method, first proposed by Ojala and colleagues [10], which labels the pixels of an image by thresholding the 3×3 neighborhood of each pixel and considers the result as a binary number. In Figure 2, given an example of LBP code

which can be described using the following equation:

$$LBP_{R,P} = \sum_{p=0}^{P-1} s(g_p - g_c) \cdot 2^p \ , \quad (1)$$

$$s(g_p - g_c) = \begin{cases} 1 & g_p \geq g_c \\ 0 & g_p < g_c \end{cases} , \quad (2)$$

where,

g_c and g_p denote the gray values of the central pixel and its neighbor

p is the index number of the neighbor

R is the radius of the circular neighborhood and

P is the number of the neighbors.

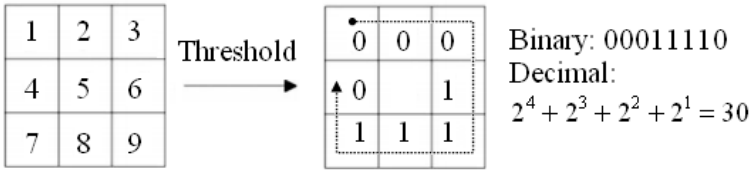

Figure 2: An example of the pattern and LBP

Many methods of local binary mode (LBP) [10] and its extension have also been widely used in many pattern fields, including face recognition [11], facial expression recognition [12], pedestrian detection [13], texture classification [14], and Content based image retrieval (CBIR) [15], satellite image detection [16], and medical image detection [17], etc. Since the LBP method combines structural and statistical features to improve the performance of texture analysis, many combining other feature or LBP-like methods have subsequently been proposed, such as Uniform LBP, LTP, LDiP, CLBP, VLBP, LGBP, DLBP, LDeP, LTrP, etc [18]. Among the foregoing methods, Uniform LBP [19] has a relatively small feature dimension and can be suitably applied to smartphone of which is limited in computing power.

A LBP is called uniform if the binary pattern contains at most two bitwise transitions from 0 to 1 or vice versa when the bit pattern is considered circular. For instance, the pattern 00001000 is uniform; while the pattern 00001001 is non-uniform considering contains 4 transitions. According to statistics, there are totally 58 different uniform patterns in neighborhood of eight sampling points, as shown in Figure 3 and occupying more than 90% of the ratio. The 58 uniform patterns are expressed in terms

of LBP values, which are 0, 1, 2, 3, 4, 6, 7, 8, 12, 14, 15, 16, 24, 28, 30, 31, 32, 48, 56 , 60, 62, 63, 64, 96, 112, 120, 124, 126, 127, 128, 129, 131, 135, 143, 159, 191, 192, 193, 195, 199, 207, 223, 224, 225 , 227, 231, 239, 240, 241, 243, 247, 248, 249, 251, 252, 253, 254, 255. The above 58 different uniform patterns are aggregated into 58-bins histograms, and all non-uniform patterns are assigned to the same bin, so the final dimension size of the Uniform LBP histogram is 59.

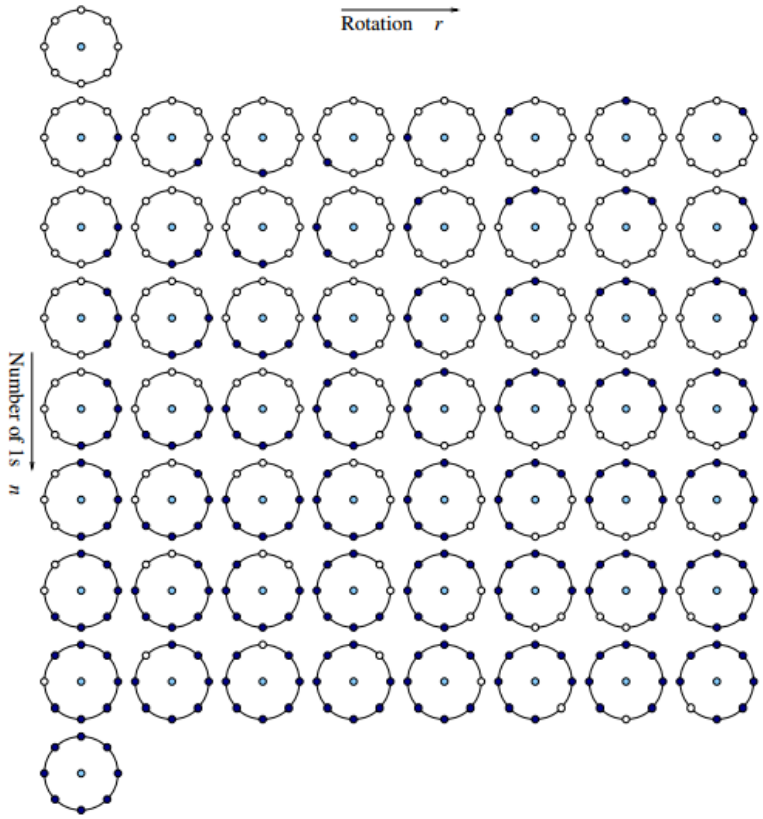

Figure 3: The 58 different uniform patterns [19]

The Uniform LBP is used in image classification by dividing the package image into multiple non-overlapping local regions, extracting local Uniform LBP histograms from them, and then joining them into a single spatial feature histogram, as shown in Figure 4. The resulting histogram has the advantage of encoding both the local texture and the global shape of the package image.

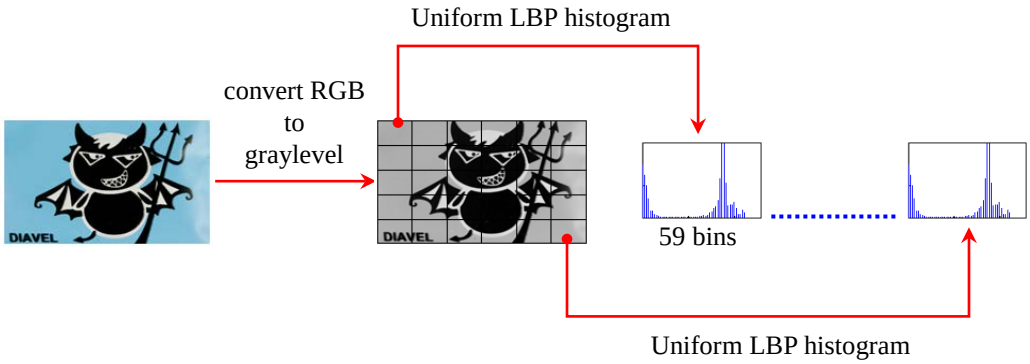

Figure 4: Uniform LBP based image description

In addition, another advantage of the uniform pattern is statistical robustness. Using a uniform pattern will be better recognized in many applications than using all possible patterns because they are more robust and less prone to noise.

2.2 Entropy of an Image

Information entropy (also called "Shannon entropy") was first introduced by Shannon [20] to measure the degree of uncertainty that exists in a system. For an event X with n possible outcomes and probabilities $p_1, \ldots, p_n$, The Shannon entropy of X is defined as

$$H(X) = H(p_1, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log_2(p_i), \quad (3)$$

where p_i is the probability of the event i and

$$\sum_{i=1}^{n} p_i = 1, \qquad 0 \le p_i \le 1. \quad (4)$$

According to Shannon's definition of entropy (3), Pun [21] first proposed the entropy of the gray image which grey level is $\{0, 1, \cdots, L-1\}$:

$$H = -\sum_{i=1}^{L-1} p_i \log_2(p_i); \qquad p_i = N_i/N, \quad (5)$$

where N is bin number, N_i represents the number of cases in each bin, and p_i is just the probability of each gray level.

For color images, convert to grayscale images and then use Equation (5) to evaluate entropy. In other words, the more complex the texture in the image, the higher the entropy value. Therefore, when doing random crop for data augmentation, the blocks with the higher entropy value are taken.

2.3 Convolutional neural network (CNN)

Deep learning is a subfield of machine learning (ML) in artificial intelligence (AI) that uses multi-layer artificial neural networks to provide state-of-the-art accuracy in tasks such as computer vision, image recognition, object detection, speech recognition, language processing, and so on. Convolutional neural network (CNN or ConvNet) is a class of deep neural network (DNN) and one of the well-known algorithms in deep learning, proposed by Lecun et al. [22] in 1998. They designed the classic model LeNet, which is simple, clear and easy to understand for the implement of CNN architecture, while effectively solves the recognition of handwritten and machine printed characters. As shown in Figure 5, the LeNet architecture has seven layers, consists of two sets of convolution and subsampling layers, followed by a flat convolutional layer, then two full connection layers and finally a softmax classifier. The detailed description of each layer is as follows:

Layer C1: The input to LeNet is a 32×32 grayscale image that passes through the first convolutional layer with six filters of size 5×5 and a stride of one pixel. The original image size changed from 32×32×1 to the feature map 28×28×6.

Layer S2: Subsampling the feature maps of the upper layer, that is, using a filter of size 2×2 and a stride of two pixels to perform average or maximum pooling. The resulting maps dimensions will be reduced to 14×14×6.

Layer C3: Feature maps of the previous layer that passes through the second convolutional layer with 16 filters of size 5×5 and a stride of one pixel. The maps size changed from 14×14×6 to 10×10×16.

Layer S4: Do the same subsampling as Layer S2, and the resulting map size will be reduced to 5×5×16.

Layer C5: This layer is a full connection with 120 feature maps each of size 1×1. Each of the 120 units is connected to all 400 nodes (=5×5×16) in the previous S4 layer. After adding 120 bias parameters, there are a total of 48,120 trainable parameters.

Layer F6: The sixth layer contains 84 units and is fully connected to layer C5. After adding 84 bias parameters, there are a total of 10,164 trainable parameters.

Output Layer: Finally, there is a fully connected softmax output layer with 10 possible values corresponding to numbers from 0 to 9.

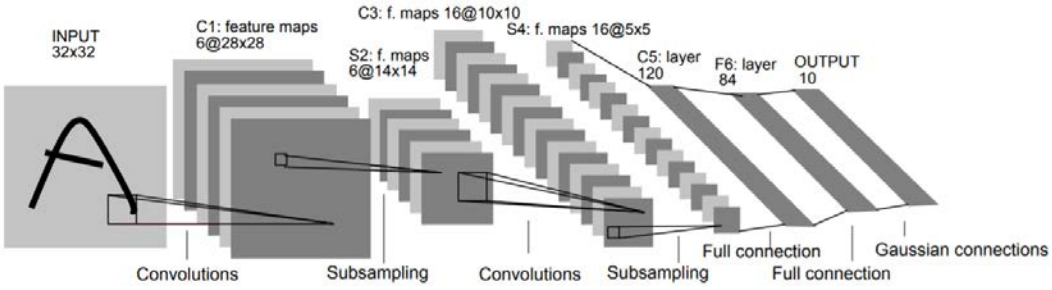

Figure 5: An illustration of the architecture of LeNet [22].

Until 2012, Krizhevsky et al. [23] proposed another CNN architecture, AlexNet shown in Figure 6, consisting of eight layers with millions of weighted parameters; which is divided into two layers and runs on two GPUs considering memory space. If there is enough memory space, we can also combine the architecture of the two paths into one path, also known as the CaffeNet architecture shown in Figure 7, and then execute on single GPU [24]. In the competition of ImageNet Large-Scale Visual Identity Challenge (ILSVRC), AlexNet won the championship and with a superior performance of 15.4% error rate, exceeding the gap of 10% in the second place. The detailed description of each layer is as follows:

Layer 1: The input to AlexNet is a 227×227×3 RGB image that passes through the first convolutional layer with 96 filters of size 11×11×3 and a stride of 4 pixels. The original image size changed from 227×227×3 to the feature map 55×55×96.

Layer 2: Feature maps of the previous layer that passes through the second convolutional layer with filter of size 5×5 and a stride of one pixel. Then, using a filter of size 3×3 and a stride of 2 pixels to perform max pooling. The feature maps size changed from 55×55×96 to 27×27×256.

Layer 3: This layer is same as the second layer except it has 384 feature maps. The resulting maps dimensions will be reduced to 13×13×384.

Layer 4: Repeat the same processing as the upper layer.

Layer 5: Repeat the same processing as the upper layer except it has 256 feature

maps. The resulting maps size is 13×13×256.

Layer 6: The previous layer of output is flattened by a fully connected layer with 4,096 feature maps, each of which has a size of 1×1.

Layer 7: This is again a fully connected layer with 4,096 feature maps each of size 1×1.

Layer 8: The final step is a layer of softmax with 1000 possible values.

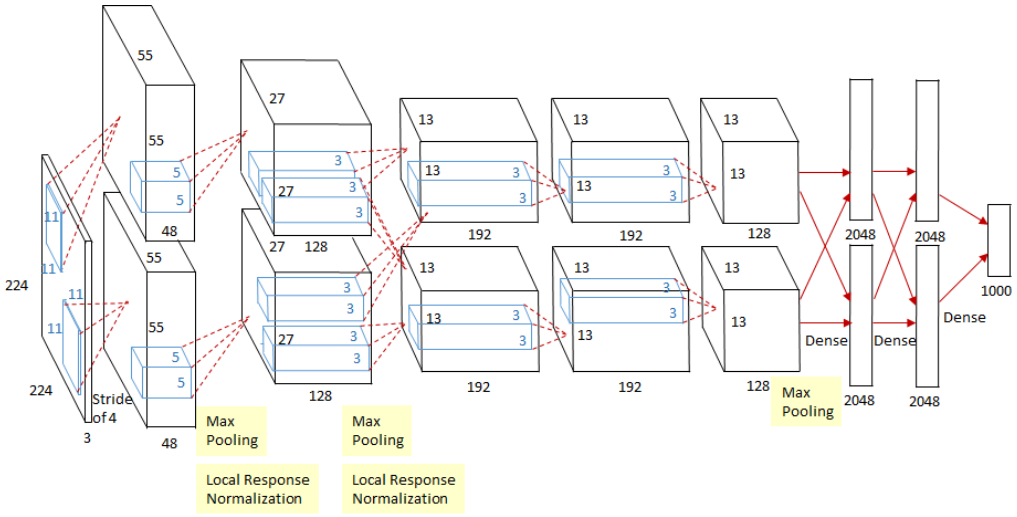

Figure 6: An illustration of the AlexNet architecture and running on two GPUs [23]

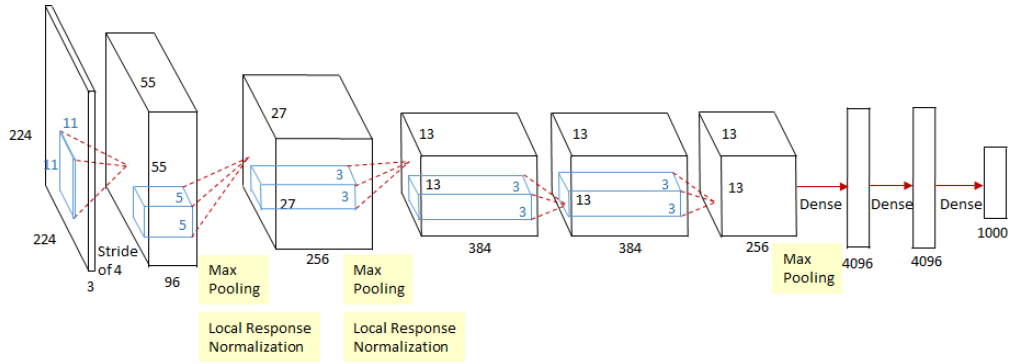

Figure 7: An illustration of the CaffeNet architecture and running on single GPU [24]

Later, some CNN designs such as ZFnet, GoogleNet, and ResNet also won the championships in the 2013-2015 competition with error rates of 11.7%, 6.7%, and 3.57%, respectively. Additionally, VGGNet, which won the second place in the ILSVRC competition in 2014, also received attention with an error rate of 7.3% [25]. Thereby, CNN is currently the most popular and leading algorithm.

2.4 Data augmentation

Some studies have pointed out that the use of data augmentation can improve the overfitting learning caused by deep learning [23] [26] [27]. Moreover, the amount of data is a key factor in whether the deep learning model can be successfully trained. Since the training data is too small, the accuracy will be reduced. Therefore, when the training data is limited, the data augmentation method becomes one of the options for expanding the data volume. In the AlexNet architecture, some data augmentation techniques are also used for training data, including image translations, horizontal reflections, and changing the intensity of the RGB channels [23].

NEDs packaging images are often the evidence in the process of investigation by smart phones or general digital cameras, and often taken in the natural environment or indoors, not in a professional photography environment. In addition, it is not easy to collect NEDs packaging images under different environmental conditions. Therefore, the collected NEDs packaging images conduct data augmentation, which commonly use include horizontal flip, vertical flip, brightness, noise, rotation, translation, zoom, stretch, and blur to enlarge database. Figure 8 is a partial schematic diagram of the NEDs package image after data augmentation, and the description is as follows:

Horizontal flip: Mirror the image along the center vertical line.

Vertical flip: Mirror the image along the center horizontal line. In fact, vertical flip is equivalent to rotating the image 180 degrees and then performing a horizontal flip.

Brightness: Adjust the degree to which the image becomes brighter and darker.

Scale: Adjust the size of the image outward or inward. When scaling outward, the image size will be larger than the original image size. Otherwise, it is smaller than the original image size.

Translation: Translation refers to moving an image in the X or Y direction or both, which can help CNN look everywhere in the image.

Noise: Over-fitting occurs when CNN learns high-frequency features that may be useless (a large number of patterns). This condition can be improved by adding an appropriate amount of noise. Commonly used are Gaussian noise or salt-and-pepper noise.

Shearing transformation: An affine transformation is a linear mapping method

that preserves points, lines, and planes. It can be used for scaling, translation, cropping, rotation, etc., where the shear transformation shifts each point horizontally or vertically by an amount proportional to its coordinates.

Blur: Image blurring is achieved by convolving the image with a low pass filter kernel that helps to eliminate noise. It actually removes high frequency content (e.g., noise, edge) from the image, causing the edges to blur when the filter is applied. Commonly used are Gaussian, average, median, etc.

Rotation: Rotate the image at different angles clockwise or counterclockwise.

Random crop: Randomly crop the specified range from the image, then resize this range to the original image size.

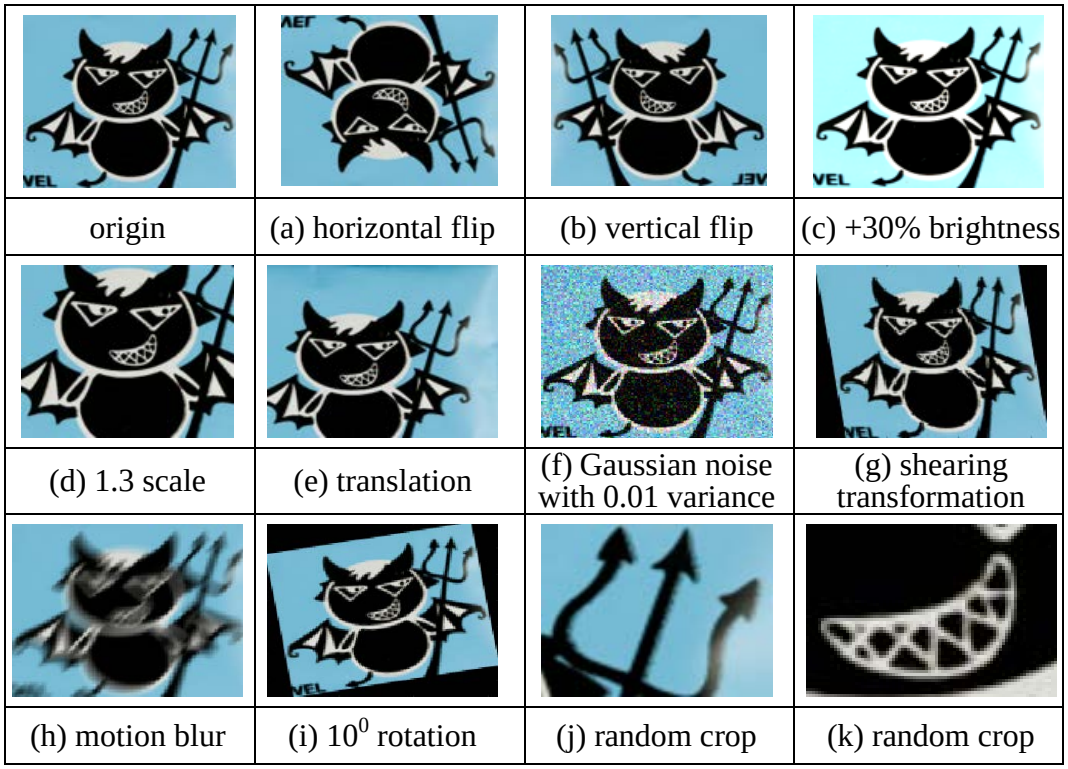

Figure 8: Data augmentation of NEDs packaging images

3. The Proposed Multi-channel CNN (MCCNN) Architecture

Since this study is going to propose a pattern recognition architecture to assist the first-line police officers in performing duties, such as pull over and spot check, use smart phones to take photos of suspicious NED packages to quickly identify whether they have been seized in the past. Thereby, it is necessary to consider that the first-line policeman is shooting a suspicious NED package in a natural environment

with a handheld smart phone. Some of these shots may have a slight translation, shearing transformation, rotation, blur, darker or lighter, larger or smaller. Therefore, based on CNN and reference to Multi-column DNN (MCDNN) [28], we design a Multi-channel CNN (MCCNN) architecture for recognition of NED package, as shown in Figure 9. In order to match the actual situation when using the mobile phone as much as possible, the training image must pass through 7 kinds of data augmentation. However, the testing image is not needed, in the case of limited computing resources of smart phones. Then, each column of CNN is used to process different images which are generated by pre-processing and Uniform LBP transformation. In more details, the MCCNN is described in the following subsections.

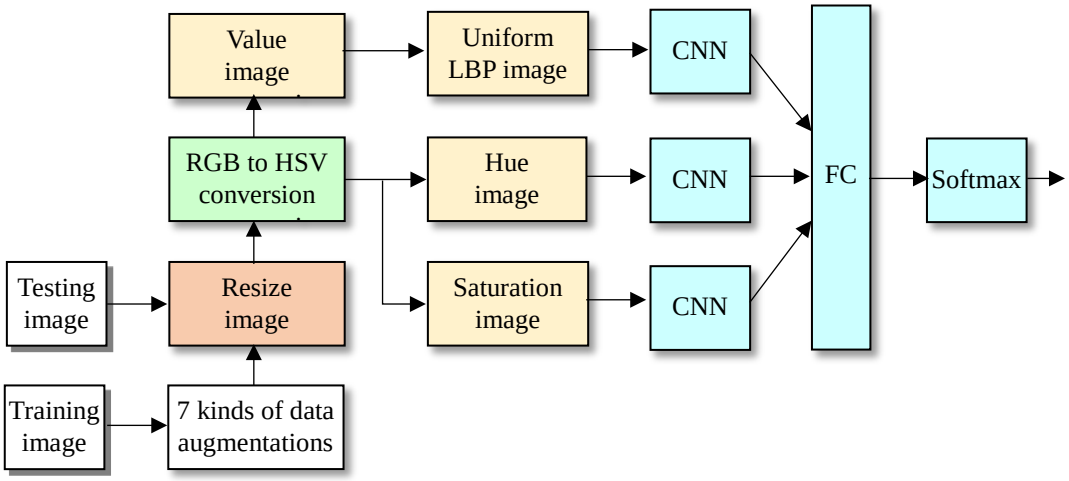

Figure 9: The architecture of the proposed Multi-channel CNN (MCCNN)

3.1 7 kinds of data augmentation

We considered that the first line of police officers used the smart phone to shoot the NED package on duty. These images will have 7 possible differences, including brightness, scale, translation, shearing transformation, blur, rotation, and random crop. Therefore, the experimental database will expand 8 images for each of the 7 possible conditions, that is, each original training image will be enlarged by 56. The 7 kinds of data augmentation will be explained in detail in Section 4 below.

3.2 Resize image

The NED package image taken by the smart phone has a large size, which is not favorable to the computing speed when recognizing. For example, the HTC 10 has a size of 4208×3120, 4208×2368, 2976×2976, and 1920×1080 pixels. In fact, the higher-order smart phone, the larger the size of the photos that can be taken. Therefore, both the training image and the testing image of this study need to be reduced in size to 224×224 pixels. In addition, the shrink algorithm uses bicubic interpolation, whose output pixel value is a weighted average of pixels in the nearest 4-by-4 neighborhood.

3.3 RGB to HSV conversion

HSV color model is commonly used in computer graphics, image processing, pattern recognition and other fields. In this space, the color is represented by three components based on cylinder coordinates: hue (H), saturation (S), and value (V). Hue is used to distinguish colors, saturation is the percentage of white light added to a pure color, and the value represents the brightness perception of a particular color. The advantage of HSV is that each of its properties directly corresponds to a human understanding of color, with the ability to separate color and achromatic components. Therefore, so as to identify NED package images with similar logos but different tones, as shown in Figure 10, the HSV color model is the best choice. So the converted H and S image will be fed into the next stage of CNN to achieve this.

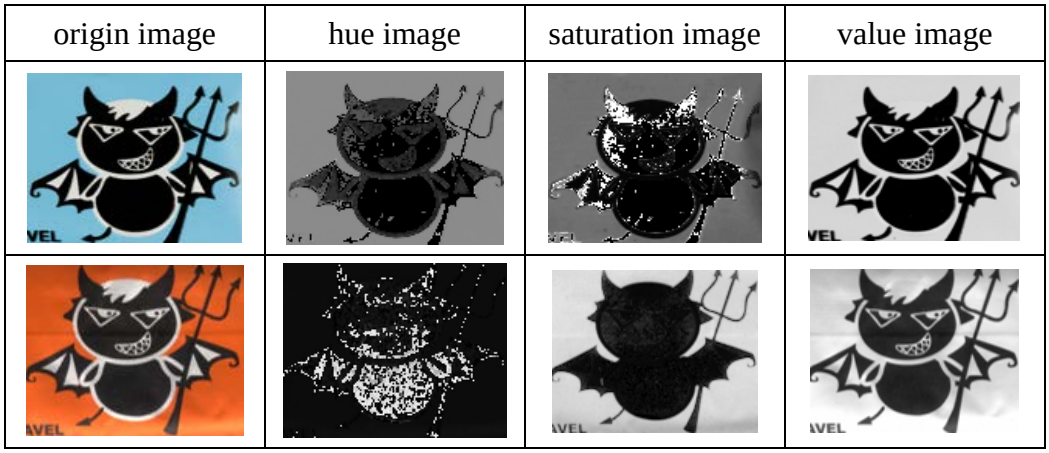

Figure 10: NED package images with similar logos but different tones

3.4 Uniform LBP image

According to Subsection 2.1, Local Binary Pattern (LBP) is a simple but powerful spatial feature descriptor that reduces computational effort and improves classification accuracy. Thus, the Uniform LBP further reduces the LBP dimension (or complexity) to suit the computing conditions of the smart phone. So the V image generated in the previous stage performs a Uniform LBP conversion as the CNN input for the next stage, as shown in Figure 11.

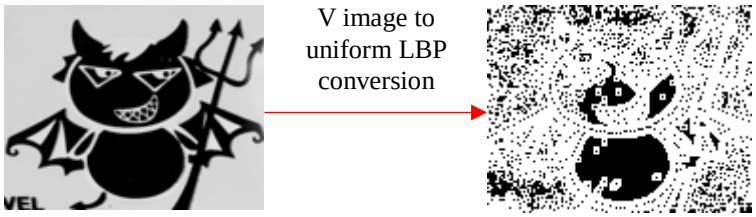

Figure 11: V image to Uniform LBP conversion

3.5 Convolutional neural network (CNN)

Convolutional neural networks (CNN) have powerful feature extraction capability, which have been widely used to extract features from images. In our MCCNN, we design three proposed CNNs to handle hue, saturation, and Uniform LBP image. Each CNN structure is based on CaffeNet, with five convolution layers and one fully connected layer, as illustrated in Figure 12. But in order to consider the efficiency of implementation, each convolution layer uses only one-eighth of the number of filters, while the fully connected layer works with 2,048 neural units for logo images.

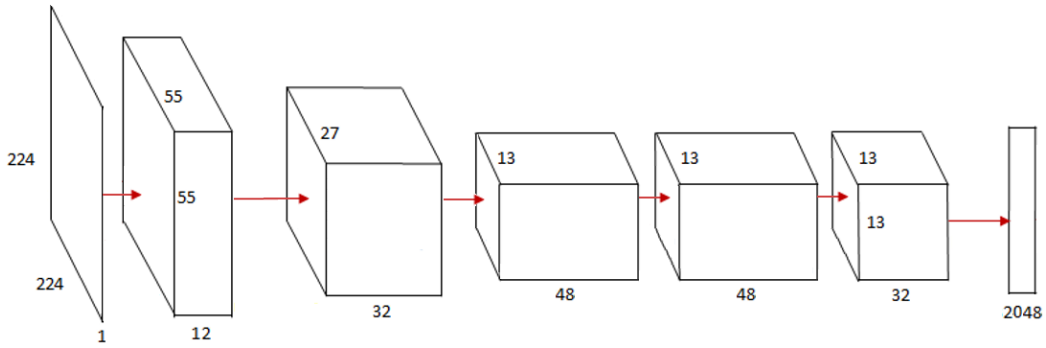

Figure 12: Overview of the proposed CNN

3.6 Fully connected (FC) layer

The proposed CNN in the previous layer can extract features. Then, this fully connected layer itself can classify these features. In addition, the more neural units in the fully connected layer, the more complex the objects that can be described. Sometimes only a fully connected layer does not solve the nonlinear problem. If there are two or more layers, the nonlinear problem can be solved well, that is, the nonlinear expression ability of the model is improved. Hence, add a fully connected (FC) layer at this stage and combine the outputs of the three proposed CNNs, as shown in Figure 13.

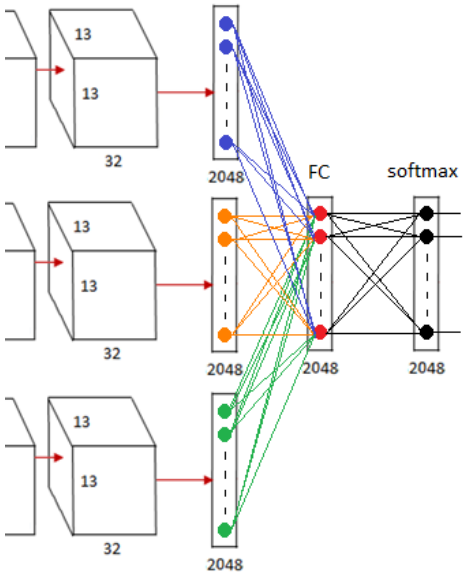

Figure 13: Overview of the fully connected (FC) layer

3.7 Softmax

Softmax is used for multi-classification problems, prior to applying softmax, some outputs may be negative, or greater than one, they may not sum to 1. Instead, the softmax layer is used to output the probability distribution as illustrated Figure 12, and then the output sum is 1 using the following equation:

$$P_i = e^{Z_i} / \sum_{j=1}^{n} e^{Z_j}, \quad (6)$$

where

P_i probability of the i-th output,

Z_i output score,

j the j-th class,

n number of classes, in this case 2,048.

Finally, we select the maximum probability node as our prediction target. This additional constraint of the softmax helps the training converge faster.

4. Experiments and Results

This section will describe the NED package and non-NED package image database we collected, and explain how to use 7 kinds of data augmentation methods to expand the image amount of the database. In addition, in order to make the correct prediction better under softmax output, we use the loss function of cross-entropy. And then the experiment will apply hold-out as the validation method, and 70% of the data is used for training, and the remaining 30% is used for testing. Finally, the experiment is implemented in Matlab programming and run on a PC-based machine with an Intel Core i5-6500 CPU, 3.2GHz, and 8G RAM.

4.1 NED package image database

The experimental database of this study collected 82 kinds of NED package images published on the Anti-Drug website [3], some of which are shown in Figure 14. These package images are seized by actual crimes, but data augmentation is required due to the number of restrictions. We judged that the first line of police officers used the smart phone to shoot the NED package on duty.

Figure 14: Partial images in the NED package database [3]

These images will have 7 possible differences, including brightness, scale, translation, shearing transformation, blur, rotation, and random crop. Therefore, the experimental database will expand 8 images for each of the 7 possible conditions, that is, each original package image will be enlarged by 56. At the same time, we collect logo images of the brands that are not NED packages on the Internet. A total of 1,966 images belong to 50 different categories, for example advertising, arts and design, business, education, food and drinks, game, media, music, shopping, sports, etc. Finally, this database contains 116,736 (57×2,048) images with original images.

In addition, since each of the seven types of data augmentation requires four parameters, we conducted experimental statistics with 20 people. The method is for each of 20 people, each taking 8 photos for 7 kinds of data augmentation. Then, summarized the appropriate 8 parameter values, shown as Table 1.

Table 1: List of data augmentation parameters

Data augmentation	Parameter description	The number images
brightness	brighten 10%, 20%, 30%, 40%, and darken 10%, 20%, 30%, 40%	8
scale	scale factor 0.6, 0.7, 0.8, 0.9, 1.1, 1.2, 1.3, 1.4	8
translation	10% and 20% displacement of image sizes in 4 directions up, down, left, and right	8
shearing transformation	horizontal and vertical shearing with shear factor m=-1, -0.5, 0.5, 1, respectively	8
blur	Gaussian filter with standard deviation σ = 0.5, 1, 2, 3; motion blur with angle θ=0^0, 10^0 and corresponding length L=5, 10 pixels	8
rotation	rotated by 30, 45, 90, 135, 150, 180, 225, 315 degrees	8
random crop	size n×n and a stride of x pixel, cut the image first, then take the top 8 with the highest image entropy value, by Equation (5)	8
total		56

4.2 Loss function

The loss function is used to optimize the parameters of the neural network. That is,

the goal of optimizing neural network weights is achieved by minimizing the loss function. The loss is calculated using the loss function by matching the actual value with the predicted value through the neural network. Then, the gradient descent method is used to optimize the weight of the network until the loss is minimized. In order to make the correct prediction probability as high as possible under softmax output, the following cross-entropy equation is used to make the convergence rate faster when it is close to accurate classification [29].

$$\text{Cross-entropy} = -\sum_{c=1}^{M} y_{i,c} \log(p_{i,c}), \tag{7}$$

where

M: number of classes,

$y_{i,c}$: a binary indicator (0 or 1) of whether class label c is the correct classification for observation i,

$p_{i,c}$: predicted probability observation i is of class c.

The cross entropy will calculate a value which summarizes the average difference between the actual and predicted probability distributions for predicting class 1. The value is minimized and the perfect cross entropy value is 0, but the experiment is set to be less than 10^{-4} or epoch up to 6 times to stop training.

4.3 Validation

The purpose of validation is to confirm the performance of the model (or system). The number of collected data in the database determines the validation method used. When the amount of data is large enough, usually using the hold-out method, as shown in Figure 15, it splits the data set into "training" and "testing" sets, such as 70% of the data for training, and the remaining 30% of the data for testing. The training set is what the model is trained on, and the testing set is used to see how well that model performs on unseen data.

Figure 15: Hold-out validation [30]

While the amount of data is insufficient, it will tend to use Cross-validation (also

known as k-fold cross-validation). This method randomly divides the data set into k groups. One of them is used as a testing set, and the other is used as a training set. Thereby, the model is trained on the training set and evaluated on the testing set. The process is then repeated until each group is used as a test set. Here, if the value of k is equal to the amount of data in the entire database, it is called leave-one-out cross validation (LOOCV). This applies when the collected data is really scarce.

Since the NED package image of this research has used data augmentation to increase the image amount of the database, the experiment will apply hold-out as the validation method, and 70% of the data is used for training, and the remaining 30% is used for testing.

4.4 Results and analysis

The proposed MCCNN method compares non-CNN and other well-known CNN classification methods. The former non-CNN method is Uniform LBP, and the latter CNN methods include AlexNet, VGG-16, VGG-19, GoogleNet, and ResNet. Accuracy is the score that matches the predicted label to the true label of the validation set. As shown in Table 2, CNN technology has better classification accuracy than non-CNN technology, and more than 10%, of which our proposed MCCNN has the highest accuracy of 98.56%. As further analysis, the main reason why the proposed MCCNN has high performance. When the foreground objects of the two logo images are similar, but the background colors are completely different, it can be correctly classified, so that the false positive rate is the lowest. Compared with the Uniform LBP method, the false positive rate is the highest.

Table 2: Results of recognition for NEDs package image

Method	# layers	Accuracy rate (%)
Uniform LBP	Non-CNN	82.89
AlexNet	8	92.59
VGG-16	16	95.67
VGG-19	19	95.78
GoogleNet	22	95.20
ResNet-50	50	98.32
Proposed MCCNN	9	98.56

5. Conclusion

We propose a novel deep learning architecture, named MCCNN, to assist the first-line police officers in performing duties, such as pull over and spot check, using smart phones to take photos of suspicious NED packages to quickly identify whether they have been seized in the past. In order to better evaluate the performance of the NED package image recognition method under actual conditions, we collected and labeled 82 NED package images and 1,966 non-NED package images, and then expanded them into an experimental database with 116,736 images using data augmentation. Experiments have proven that the proposed MCCNN has the best accuracy compared with the non-CNN and state-of-the-art CNN methods, and reaches 98.56%. In the future, the proposed architecture can be built using cloud computing technology for criminal investigation.

Reference

[1] Victoria State Government, “Synthetic drugs (new psychoactive substances),” https://www.betterhealth.vic.gov.au/health/HealthyLiving/synthetic-drugs, Access: 28 Jul. 2019.

[2] United Nations Office on Drugs and Crime (UNODC), “UNODC early warning advisory (EWA) on new psychoactive substances (NPS),” https://www.unodc.org/LSS/Page/NPS, Access: 20 May 2019.

[3] Anti-Drug, “New and emerging drugs (NEDs) information,” https://antidrug.moj.gov.tw/cp-1190-4840-2.html, Access: 20 May 2019.

[4] A. A. Shirazi, A. Dehghani, H. Farsi and M. Yazdi, “Persian logo recognition using local binary patterns” 2017 3rd International Conference on Pattern Recognition and Image Analysis (IPRIA), Apr. 2017.

[5] M. Wasim, A. Aziz and S. F. Ali, “Object‘s shape recognition using local binary patterns,” International Journal of Advanced Computer Science and Applications (IJACSA), vol. 8, no. 8, 2017, pp. 258-262.

[6] Y. Yu, J. Wang, J. Lu, Y. Xie and Z. Nie, “Vehicle logo recognition based on overlapping enhanced patterns of oriented edge magnitudes,” Computers & Electrical Engineering, vol. 71, Oct. 2018, pp. 273-283.

[7] M. Sotoodeh, M. R. Moosavi and R. Boostani, “A novel adaptive LBP-based descriptor for color image retrieval,” Expert Systems with Applications, vol. 127, Aug. 2019, pp. 342-352.

[8] R. Wason, “Deep learning: Evolution and expansion,” Cognitive Systems Research, vol. 52, Dec. 2018, pp. 701-708.

[9] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu and F. E. Alsaadi, “A survey of deep neural network architectures and their applications,” Neurocomputing, vol. 234, Apr. 2017, pp. 11-26.

[10] T. Ojala, M. Pietikäinen and T. Mäenpää, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, 2002, pp. 971–987.

[11] A. F. Zhilali'l, M. Nasrun and C. Setianingsih, “Face recognition using local binary pattern (LBP) and Local Enhancement (LE) Methods At Night Period,” 2018 International Conference on Industrial Enterprise and System Engineering (ICoIESE 2018), Mar. 2018.

[12] C. Shan, S. Gong and P. W. McOwan, “Facial expression recognition based on Local Binary Patterns: A comprehensive study,” Image and Vision Computing, vol. 27, no. 6, 2009, pp. 803–816.

[13] N. Jiang, J. Xu, W. Yu and S. Goto, “Gradient local binary patterns for human detection,” 2013 IEEE International Symposium on Circuits and Systems (ISCAS2013), May 2013.

[14] C. P. Yen, “Fusion of motif co-occurrence matrix and local binary pattern based on intuitionistic fuzzy set for texture classification,” Scientific Visualization, Accepted Jun. 2019.

[15] S. R. Dubey, S. K. Singh and R. K. Singh, “Multichannel decoded local binary patterns for content-based image retrieval,” IEEE Transactions on Image Processing, vol. 25, issue 9, Sep. 2016, pp. 4018-4032.

[16] F. Yang, Q. Xu and B. Li, “Ship detection from optical satellite images based on saliency segmentation and structure-LBP feature,” IEEE Geoscience and Remote Sensing Letters, vol. 14, issue 5, May 2017, pp. 602-606.

[17] J. S. Athertya, G. S. Kumar and J. Govindaraj, “Detection of Modic changes in MR images of spine using local binary patterns,” Biocybernetics and Biomedical Engineering, vol. 39, issue 1, 2019, pp. 17-29.

[18] M. Pietikäinen and G. Zhao, “Two decades of local binary patterns: A survey,” Advances in Independent Component Analysis and Learning Machines, 2015, pp. 175-210.

[19] T. Ojala, M. Pietikainen and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 24, no. 7, 2002, pp. 971-987.

[20] C. E. Shannon, “A mathematical theory of communication,” The Bell System Technical Journal, vol. 27, Jul. 1948, pp. 379-423.

[21] T. Pun, “A new method for gray-level picture thresholding using the entropy of the histogram,” Signal Processing, vol. 2, 1980, pp. 223-237.

[22] Y. LeCun, B. Boser and J. S. Denker et al., “Backpropagation applied to handwritten zip code recognition,” Neural Computation, vol. 1, issue 4, 1989, pp. 541-551.

[23] A. Krizhevsky, I. Sutskever and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” In Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, Dec. 2012, pp. 1097-1105.

[24] Y. Jia, E. Shelhamer, J. Donahue et al., “Caffe: Convolutional architecture for fast feature embedding,” Proceedings of the 22nd ACM international conference on Multimedia, Orlando, Florida, USA, Nov. 2014, pp. 675-678.

[25] IMAGENET, “Large Scale Visual Recognition Challenge (ILSVRC),” http://image-net.org/challenges/LSVRC/, Access: 6 Feb. 2019.

[26] Nguyen and Khanh, “Multiple kernel learning with data augmentation,” Asian Conference on Machine Learning, 2016, pp. 49-64.

[27] Tran and Toan, “A Bayesian data augmentation approach for learning deep models,” Advances in Neural Information Processing Systems, 2017, pp. 2794-2803.

[28] D. Ciresan, U. Meier and J. Schmidhuber, “Multi-column deep neural networks for image classification,” 2012 IEEE Conference on Computer Vision and Pattern Recognition, USA, Jun. 2012, pp. 3642-3649.

[29] P. Golik, P. Doetsch and H. Ney, “Cross-Entropy vs. Squared Error Training: a Theoretical and Experimental Comparison,” 14th Annual Conference of the International Speech Communication Association, Lyon, France, Aug. 2013, pp. 1756-1760.

[30] ebc, “Cross-Validation Strategies,” http://www.ebc.cat/2017/01/31/cross-validation-strategies/, Access: 28 Jul. 2019.