Proceedings of the 4th International Conference on Computing and Data Science (CONF-CDS 2022)
DOI: 10.54254/2755-2721/2/20220541

Ground Based-cloud Classification Based on Comparing Different Classification Models

Tao Zhang[1,5], Yantian Ding[2], Wangjiang Gong[3] and Zeyu Hou[4]

[1]Department of Computer Science, Purdue University, West Lafayette, 47906, United States
[2]School of Information Science and Engineering, Lanzhou University, Lanzhou, 730000, China
[3]Reading Academy, Nanjing University of Information Science & Technology, Nanjing, 211800, China
[4]Beijing No. 80 High School, Beijing, 100102, China

[5]zhan3263@purdue.edu

Abstract. Clouds are one of the most common things in our daily life, and different types of clouds will foreshadow different weather conditions. Therefore, accurately identifying and classifying them are important for human forecasting the local weather. However, artificially classifying clouds could possibly cause a minor error. With the helping of machine learning classification model could avoid this situation as much as possible. In this study, our goal is to get sufficiently high accuracy of based-cloud classification. We used LBP as a preprocessing approach, and used KSVM, MLP, Custom Vision, and Resnet to compare each result. The dataset contains around 15 thousand cloud images in 5 types in the same format and pixels. With the implementation of four classification models, experiments showed that ResNet34 came out a great result with a test accuracy 94%. Therefore, this study demonstrates ResNet34 is a good image classification model in based-cloud classification field.

Keywords: LBP, MLP, Based-Cloud Classification, ResNet, Custom Vision.

1. Introduction

Clouds are an important meteorological phenomenon. When the sun shines on the Earth's surface, water evaporates to form water vapor. As the water vapor rises and begins to condense as the temperature drops, the water molecules gather around the condensation nuclei in the air and form shapes visible to the naked eye which creates clouds.

Clouds play a key role in weather forecasting. They help drive the water cycle bringing precipitation [1]. In addition, clouds are variable in time and space and promotion in the energy balance of the atmosphere because of their ability to absorb and reflect long-wave and short-wave radiation [2]. For the balance of radiation, cloud morphology are indispensable decision factors, such as vertical and horizontal thickness, cloud droplet size and concentration, and cloud droplet size and concentration [3]. Different clouds have different characteristics, Higher correct cloud classification rate is favorable study the radiation balance and weather forecasting.

According to the cloud classification from World Meteorological Organization [1], Clouds are di-

DOI: 10.54254/2755-2721/2/20220541

vided into many categories depending on their different heights and shapes in the sky. In terms of height, clouds can be divided into low (cloud base is usually about 2000 meters), medium (Cloud base are usually between 2000 and 7000 meters), and high cloud groups (Cloud base above about 5000 meters). There are also more types of clouds according to their shapes. Low clouds: Cumulus, Cumulonimbus, Stratocumulus, Stratus. Medium clouds: Altostratus, Altocumulus. High clouds: Cirrus clouds, Cirrostratus, Cirrocumulus. Most relevant studies in the field of earth sciences, especially atmospheric sciences, require cloud observations, such as the type and number of clouds [4]. These parameters have traditionally been obtained by human observers. However, human observations are not only time consuming but also somewhat inconsistent and subjective even if the observer is highly trained and experienced [3]. Therefore, it is urgent to be able to classify clouds efficiently and accurately.

Most of the sources of cloud maps come from two parts, one is satellite cloud maps. The other part is ground-based cloud pictures. Among them, ground-based cloud pictures are the most widely used, so our work is also based on ground-based cloud pictures for the development of new classification algorithms. In recent years, there are several sky imagers such as the whole-sky imager (WSI) [5], infrared cloud imager (ICI) [6], total-sky imager (TSL) ([4,7]), and whole-sky infrared cloud-measuring system (WSIRCMS) [8], all-sky imager (ASI) [9]. These instruments provide us with a large amount of cloud pictures.

This passage will study on accurately classifying five types of ground-based cloud. The dataset section will describe detailly information of our dataset. Next section – related works – will introduce several similar works. The methodology section demonstrates the classification models we used. The last section will conclude this experiment and propose the future works.

2. Dataset

This paper uses three different data sets, a total of 15542 cloud photos. Because there are different formats of images in different datasets, which will have adverse effects on classification, we change the image format to JPG with a resolution of 125 × 125, with a black background box. Considering that there are different classifications in the three datasets, we finally decided to divide the clouds into five categories: stratiform clouds (quantity: 1202), cumulus clouds (quantity: 2619), cirrus clouds (quantity: 2323), mixed clouds (quantity: 2020) and clear sky clouds (quantity: 2338).

2.1. TJNU multimodal ground-based cloud database (MGCD)

Resolution of cloud picture is 1024×1024. In order to make the different datasets in a uniform format, we reduced the resolution of the cloud map to 125×125 and also blackened the frame (unification treatment), and figure 1 shows several samples. The datasets are divided into 7 categories: cumulus, altocumulus, cirrus, clear sky, stratocumulus, cumulonimbus, mixed.

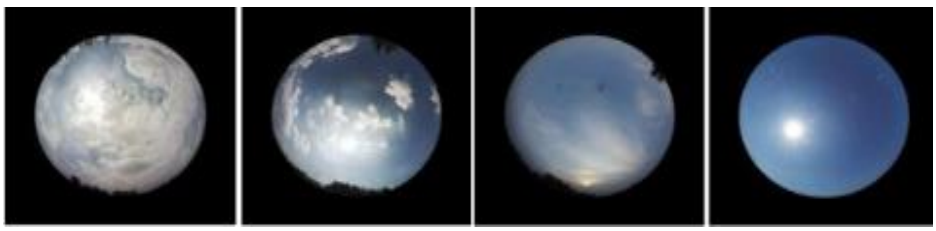

Figure 1. shows MGCD dataset after unification

2.2. Cirrus cumulus stratus nimbus (CCSN) data set

The data set contains 2,543 unique ground-based cloud images with resolution 256x256 pixels in JPG format, labels are the result of several rounds of discrimination by meteorologists with relevant experience [10]. The dataset classifies clouds into 11 categories, namely Ac, As, Cb, Cc, Ci, Cs, Ct, Cu, Ns, Sc, St. We had done a unification preprocess on this data set. Figure 2 shows CCSN dataset after unification treatment. From left to right is stratus, cumulus, cirrus.

DOI: 10.54254/2755-2721/2/20220541

Figure 2. shows CCSN dataset after unification treatment.

2.3. Ground-based all-sky cloud observation system and all-sky cloud map database.

The ground-based Total-sky Cloud Imager (TCI) is an automated ground-based cloud observation system. The camera unit uses a combination of an industrial camera with high stability and a fisheye lens with a viewing angle of not less than 180°. Unlike other visible band ground-based cloud observation equipment, TCI adopts a new protection scheme of full shading during idle time and full opening during observation, which reduces the complexity of the mechanical structure of the equipment significantly and is suitable for promotion and application. The public all-sky cloud map database contains 5000 all-sky images, divided into five categories: clear sky, cirrus clouds, stratus clouds, cumulus clouds and mixed sky, each containing 1000 images. These images were acquired by the TCI equipment between 2012 and 2014 in the Tibetan Plateau region of China. The resolution of the original images is 1392×1024. We had done a unification preprocess on this data set shown in figure 3.

Figure 3. shows ground-based all-sky cloud observation system and all-sky cloud map database.

3. Related works

A project made by Li et al. is about using deep fusion network to do cloud classification. This paper mainly illustrates an efficient and accurate approach to classify clouds with the utilization of a deep network. The method the authors use is a deep tensor fusion network (DTFN), and the purpose of applying it is to preserve the spatial information of the cloud images when uploading them to the dividing programs. Multimodal tensor subnetwork (MTN) and visual tensor subnetwork (VTN) are components of DTFN. By using VTN, the team can transform cloud images of the dataset into cloud visual tensors; by using MTN, the team can successfully enable the multimodal tensor to be compatible with the cloud visualization tensor mathematically. Moreover, they assert a tensor fusion layer to find the relationship between cloud visual tensor and multimodal tensor [11]. This paper explains a new way to help enhance the accuracy of classifying clouds, which stimulates us to consider and explore advanced methods of cloud classification. Yet the goal of our team is to create a program that contains a much larger dataset and a more accurate division with the photos taken by users from the ground. In short, this paper concentrates to improve the veracity by increasing new factors ------ it only focuses on the classifying process, but our paper focuses more on providing a whole application that can help the users classify clouds after they upload photos taken by their phones.

A project made by Wan et al. which is about using images taken from ground to classify clouds. A brand-new way which unite multiple aspects and random forest classifier that can be utilized to classify images of clouds taken on the ground. Just as what the authors define, the paper introduces a solution to a problem, cloud, which can affect the production of solar energy. The team first assembles

DOI: 10.54254/2755-2721/2/20220541

some basic aspects of the cloud, such as light extent, grain, and color, into a set. Then they utilize the method called Random Forest to classify different kinds of clouds and sky. At last, after combines the methods and results, the team divides the clouds into several kinds of groups. The data can be used to calculate the generation rate of solar energy. Multi-feature categorization accuracy is higher than single-feature and two-feature categorization accuracy [12]. After reading the paper, our group members share our new comprehensions and ideas of our topic, analyze the difference, and specify our goal. In the conclusion part, the authors present a limitation of their program, which brings us new ideas. Moreover, we read the codes it used, classify all the segments, and reconstruct ours. This paper concentrates on the classifying part and focuses on solar energy, yet our paper not only contains a classifying program, but also a function which can provide the results to the users. Reading this paper benefits us through helping us comprehend our goal and specify the necessary segments of the program.

A project made by Zhang et al. which is about determine the category of clouds by using neural network that is deep and convolutional. This paper introduces a new cloud classification method called CloudNet. The purpose of applying it is to find the effect of different cloud radiation effects. The team first set up datasets called Cirrus Cumulus, Stratus, Nimbus, including 11 classes under Meteorological Standard. The team also enlarge the scale of their previous dataset, specifically 3 times. With the help of a Deep Convolutional Neural Network, CloudNet has a high accuracy rate. Numerous test results indicate that CloudNet has a good efficiency with the swims - cat database and has a good performance with the complicate CCSN database in classifying ground-based cloud [13]. From this paper, we indicate that a larger dataset is likely to bring a more accurate result after our program study and be trained with it. Moreover, we also send an email to obtain the dataset this paper applies since it is mature and classified carefully. Besides, we also download more than 5,000 pictures from datasets of other websites. Despite the help, this paper only pays attention to the classifying method ------ there is a limited description about the use of the program. They just provide a successful method, yet our paper will not only explain the method we find and perfect it, but also illustrate some functions which our program can contribute to.

A project made by Hasenbalg et al. which is about 6 methods that can be used to classify clouds. This paper mainly talks about 6 different methods and the results after examining them. The first is a color-channel threshold-based algorithm, the second is a Clear. Sky. Library (CSL) based approach, the third is a Region. Growing. Algorithm, and the fourth is the Hybrid. Thresholding. Algorithm (HYTA), the fifth is a novel, HYTA-based. development named HYTA+, and the sixth is a deep. Convolutional. Neural. Network (FCN) is adapted via transfer learning to this problem. The team tests all the methods with a dataset that contains 829 images divided into 16 categories and figures out that CSL, HYTA+, and FCN are the most effective classifying methods. Moreover, FCN performs the best, specifically 97% accuracy, when tested with 160 images [14]. This paper brings us the idea of searching different methods and compare their results after uploading our dataset. Then we can find the most accurate method among them, improve it, and thus generate a new method with high accuracy in classifying cloud images. The greatest difference is that this paper only provides the information about the accuracy-test among the 6 methods, but ours also presents some improvements and a potential utilization channel. Our group also applies a much larger dataset, which may further improve the accuracy.

4. Methodology

Generally, to reach out sufficiently high accuracy rate for image classification, preprocessing methods and multiple classification models' comparison are required to do so. This section would primarily introduce preprocess method LBP and several image classification models used in our project, such as KSVM, MLP, Microsoft Custom Vision, and Resnet.

4.1. Domain LBP

The cloud classification is one kind of the Edge detection and recognition; therefore we used the domain LBP as the main preprocess function.

DOI: 10.54254/2755-2721/2/20220541

The original LBP operator is defined in the neighborhood of 3x3 pixels. In the neighborhood of the center pixel threshold value, the pixel value of eight neighboring pixels with gray value of the center neighborhood is compared. Shown in figure 4, If the center pixel value is greater than the surrounding pixels, then the pixel positions labeled as 1, otherwise 0 [15].

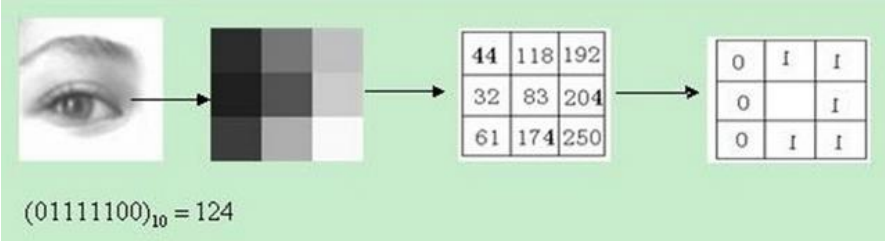

Figure 4. LBP Explanation [15]

Characterization of the spatial structure of the original partial image texture LBP operator is given by:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^P, \text{ where } s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (1)$$

LBP Operator Function (1)

As can be seen from the definition of LBP, LBP operator is the same gray scale, but not rotation invariant. Image rotation will result in different values LBP [16].

However, the LBP motivated, having the spatial structure for the two features of the image and contrast of texture, spatial pattern affected by the rotation, contrast, gradation affected. In order to obtain reliable method in gray scale and rotation changes, LBP achieve gray and rotational invariance, taking into account the differences in values and signs of gray pixels in the local community center rather than the exact value, minimum is found in all possible operator's value, called circular LBP shown in figure 5.

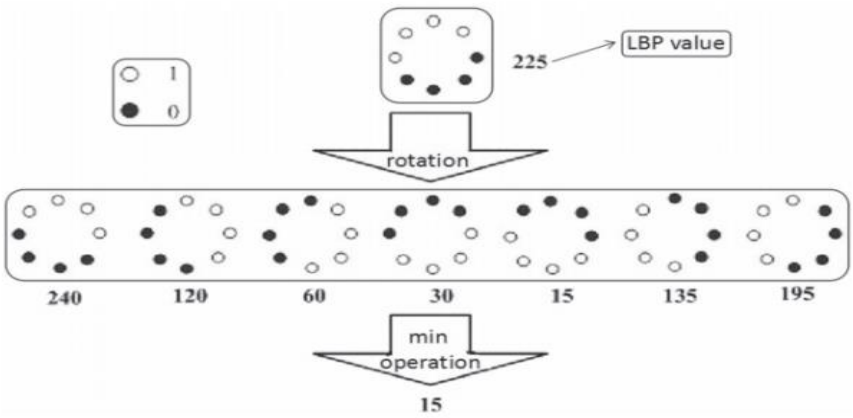

Figure 5. Approach of finding the minimum value among all possible values of operator

The circular LBP values is given by Circular LBP Values Function (2)

$$LBP_{P,R}^{ri} = \min_{0 \leq i \leq P-1} \{\sum_{P=0}^{P-1} s(g_p - g_c)2^{[(p+i) \bmod P]}\} \quad (2)$$

Where, in principle, (P, R) could choose any combination of P and R, however, in practice, they always take (8,1), (16,2) and (24,3). [16]

4.2. Kernel-SVM

Support vector machine (often abbreviated as SVM, also known as support vector network) is a supervised learning model for analyzing data in classification and regression analysis with associated learning algorithms [17].

In our project, we used the support vector machines in scikit-learn package. As scikit-learn provide

DOI: 10.54254/2755-2721/2/20220541

multiple different kernel functions, we were able to compare various situations to get the highest accuracy [17]. There are four kernel functions:

1.Linear: $< x, x^{'} >$

2.Polynomial: $\left(\gamma < x, x^{'} > + r\right)^{d}$, where d is specified by parameter degree, r by a coefficient.

3.RBF: known as radial basis function kernel, $K\left(x, x^{'}\right) = \exp\left(-\frac{\left||x - x^{'}|\right|^{2}}{2\sigma^{2}}\right)$

4.Sigmoid: $\tanh\left(\gamma < x, x^{'} > + r\right)$, where r is specified by a coefficient.

4.3. MLP

Neural networks are commonly used in the current machine learning field, for example, they can be used for image recognition, speech recognition, etc. It is a highly parallel information processing system with strong adaptive learning capability. Moreover, it has a good robustness to changes in the system parameters of the controlled object and external disturbances, which could handle complex multi-input, multi-output nonlinear systems, and the basic problem to be solved by neural networks is classification.

Multilayer Perceptron (MLP), also called Artificial Neural Network (ANN, Artificial Neural Network), can have multiple hidden layers in the middle of it, except for the input and output layers, and the simplest MLP contains only one hidden layer, i.e., a three-layer structure, shown in figure 6 [18].

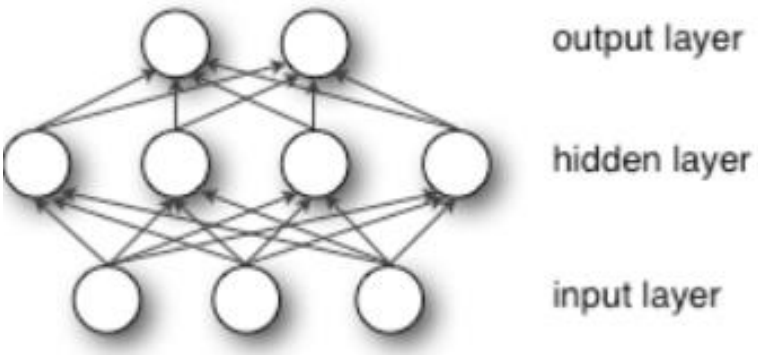

Figure 6. MLP Layer [18]

As can be seen from the above figure, the multilayer perceptron is fully connected between layers. The bottom layer of the multi-layer perceptron is the input layer, the middle layer is the hidden layer, and finally the output layer [17]:

1.Input Layer: Input data

2.Hidden Layer: $f(wx + b)$, with w is the weight (connection coefficient), b is the bias. We implement scikit-learn, and there are four activation functions for the hidden layer:

a)identity: no-op activation, useful to implement linear bottleneck, returns $f(x) = x$

b) logistic: $f(x) = 1 / (1 + exp(-x))$.

c)tanh: $f(x) = \tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$

d) relu: $f(x) = \max(0, x)$

3.Output Layer: the hidden layer to the output layer can be viewed as a multiclass logistic regression, such as softmax regression:

a)$softmax(w_2 x_1 + b_2)$, which $x_1$is the output of the hidden layer.

4.4. Resnet

In the problem of image classification, convolutional neural network has a good performance. Compared with the traditional fully connected neural network, the convolutional neural network adds a

DOI: 10.54254/2755-2721/2/20220541

convolutional layer and a pooling layer. The convolution layer uses a convolution kernel to extract the local details of the image through discrete convolution. The pooling layer is a down-sampling operation that maps an area of the image to a point in the next layer. Each convolutional layer in the convolutional neural network extracts local details from the image and gradually extracts higher-level and more abstract features. Therefore, convolutional neural network has a very good ability of feature recognition. In general, the more neural network layers are, the better the performance of the network will be.

However, in fact, If the network is too deep, problems such as gradient disappearance may occur, which makes training difficult. Maybe we can use some conventional methods to solve this problem, like data preprocessing or add the normalized layer. But all these methods are just cure the symptoms, not the disease. Shown in figure 7, as the depth of the network increases, there will be still a degradation problem: the accuracy of training will gradually reach a limit value and then decrease.

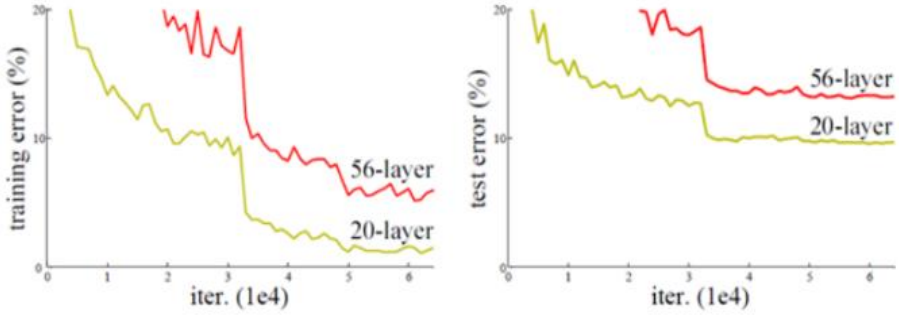

Figure 7. accuracy of training.

To solve the disease, people used the Residual Network. This model not only can increase the depth of the network, but also has a fast-training speed and a stronger performance than ordinary CNN. The implementation of these features mainly relies on residual blocks.

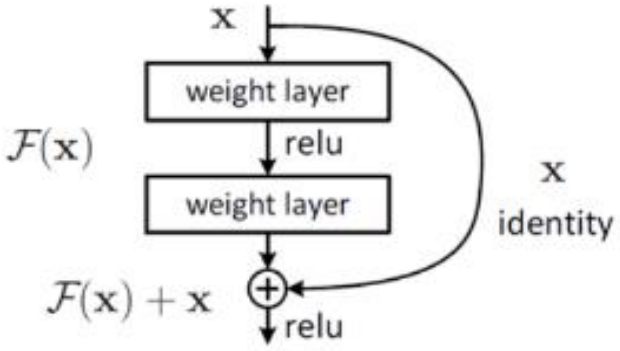

Figure 8. Residual Block.

Figure 8 is the Residual Block. The Residual Block not only had sequential convolution layers but also skipped some convolution layers with a Shortcut Connection side by side with the convolution layers. The shortcut connection can pass the data of the input residual block directly to the output, and add the output F(x) through the convolution layer. The output of the residual block is F(x) + x. In this way, in the training process, the error can be lossless back through the shortcut connection, and the problem of gradient disappearance is solved.

DOI: 10.54254/2755-2721/2/20220541

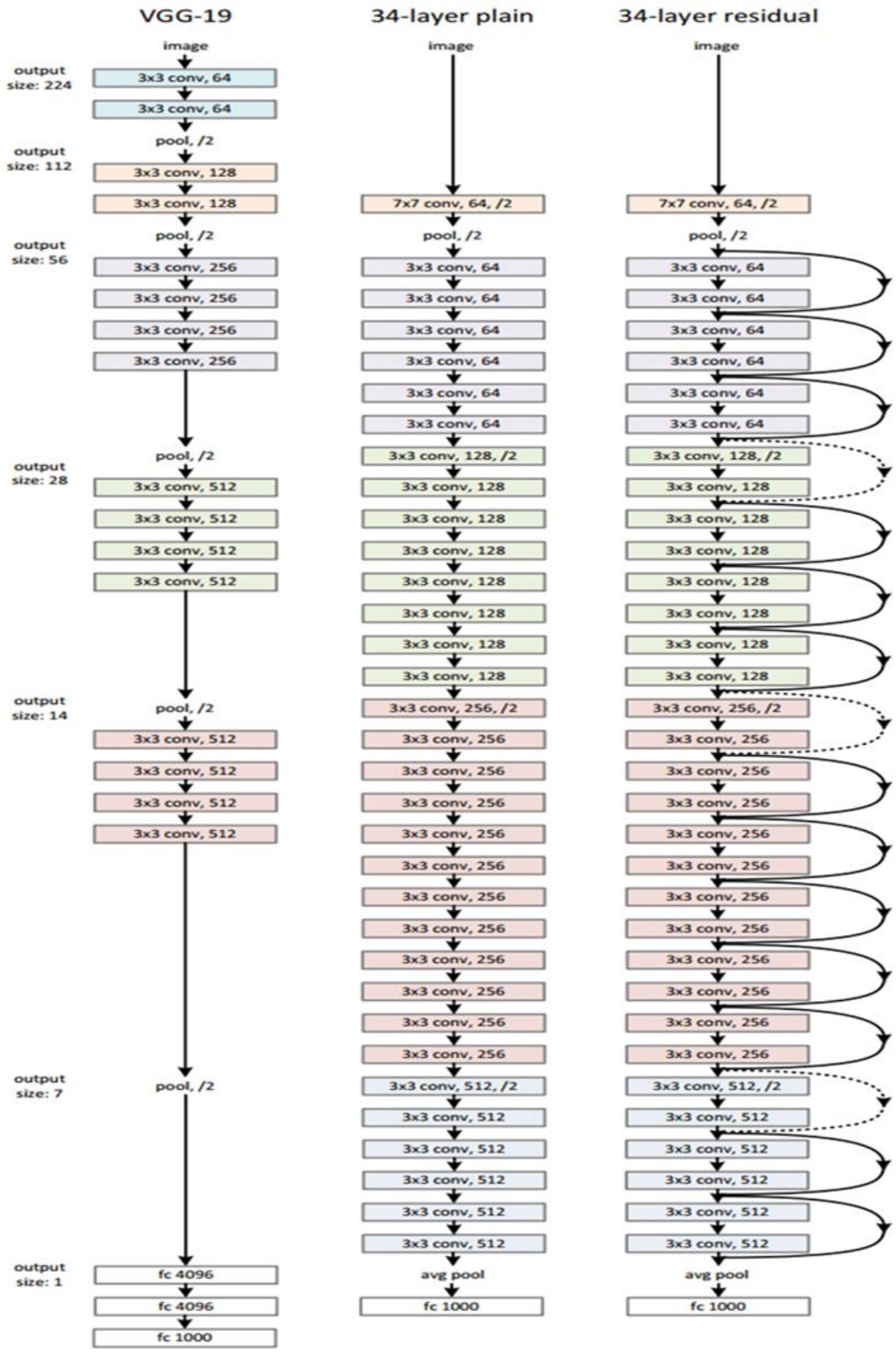

Figure 9. VGG19 network.

DOI: 10.54254/2755-2721/2/20220541

ResNet is based on and modified by VGG19. As the figure show below, it add the residual blocks between layers. This change mainly in RESNET convolution directly stride = 2 sub-samples, replace the full connection layer global average cell layer. RESNET an important design principle is that, when cut in half the size of the feature map, the number of feature map will be doubled, which maintains the complexity of the network layer. As it can be seen from the figure, compared with the conventional network, increasing the short-circuit between every two RESNET mechanism, a residual study, a broken line represents the number of feature map changes. The following figure shows ResNet 34 layers, which may also be constructed deeper network, shown below. As can be seen from the figure 9, for the ResNet18 and ResNet34, for a residual study between the two. When the network is deep, the residuals are learned among the three layers. Three convolution kernels are 1x1, 3x3 and 1x1. And it is 1/4 of the output feature map [19].

4.5. Microsoft custom vision

Azure Custom Vision is an image recognition service that can be used to build, deploy, and improve your own image identifiers. Image identifiers apply labels (representing classes or objects) to images based on their visual characteristics. Different with computer vision services, Custom Vision allows developers to specify labels and train custom models to detect them with high accuracy. Custom visual functions can be divided into two functions [20]:

1.Image Classification: adding one or more labels on one image.

2.Object Detection: similar with image classification, but it also returns the coordinates in the image, and the applied label can be found in the coordinates.

In our study, we mainly use the image classification feature in the Custom Vision as one of classification approach.

5. Experiment

5.1. KSVM

When we use the LBP value for every pixel and use the KSVM to classified the dataset with and without LBP value, we find the result as Table 1, Table 2, Table 3, and Table 4:

Table 1. Result without LBP

	Precision	Recall	F1-score	Support
A-sky	0.68	0.69	0.69	52
B-pattern	0.67	0.17	0.28	23
C-think-dark	0.33	0.37	0.35	59
D-think-white	0.35	0.39	0.37	38
E-veil	0.30	0.33	0.31	24
Accuracy	/	/	0.43	196
Macro avg	0.46	0.39	0.40	196
Weighted avg	0.46	0.43	0.43	196

kernel='poly', gamma=0.0001, C=1

DOI: 10.54254/2755-2721/2/20220541

Table 2. Result with LBP (8, 1)

	Precision	Recall	F1-score	Support
A-sky	0.56	0.69	0.62	51
B-pattern	0.96	1.00	0.98	25
C-think-dark	0.42	0.42	0.42	59
D-think-white	0.71	0.91	0.80	35
E-veil	1.00	0.15	0.27	26
Accuracy	/	/	0.62	196
Macro avg	0.73	0.64	0.62	196
Weighted avg	0.66	0.62	0.59	196

kernel='poly', gamma=0.0001,C=1

Table 3. Result with LBP (16, 2)

	Precision	Recall	F1-score	Support
A-sky	0.68	0.69	0.69	52
B-pattern	0.67	0.17	0.28	23
C-thick-dark	0.33	0.37	0.35	59
D-think-white	0.35	0.39	0.37	38
E-veil	0.30	0.33	0.31	24
Accuracy	/	/	0.43	196
Macro avg	0.46	0.39	0.40	196
Weighted avg	0.46	0.43	0.43	196

kernel='poly', gamma=0.0001, C=1

Table 4. Result with LBP (24, 3)

	Precision	Recall	F1-score	Support
A-sky	0.00	0.00	0.00	60
B-pattern	0.00	0.00	0.00	16
C-thick-dark	0.00	0.00	0.00	69
D-thick-white	0.00	0.00	0.00	34
E-veil	0.09	1.00	0.16	17
Accuracy	/	/	0.09	196
Macro avg	0.02	0.20	0.03	196
Weighted avg	0.01	0.09	0.01	196

kernel='poly', gamma=0.0001, C=1

With the KSVM report, we can find that even if LBP can improve the accuracy. Of course, LBP

DOI: 10.54254/2755-2721/2/20220541

can only have a positive effect on classification if it is in the right range. If the number of bits of LBP value is too large, it will cause less commonality between images under the same category. For example, in Table 4, when we extend the number of bits of LBP to 24, the classifier already approaches to distinguish the clouds of different categories. Also, there are still some patterns will do a negative effect. Also, the graph visualized by PCA show the same view. In the PCA graph, figure 10 and 11, we can see that the point cluster are more dis-perse. According to the paper [17], we choose the first 284 patterns as the feature.

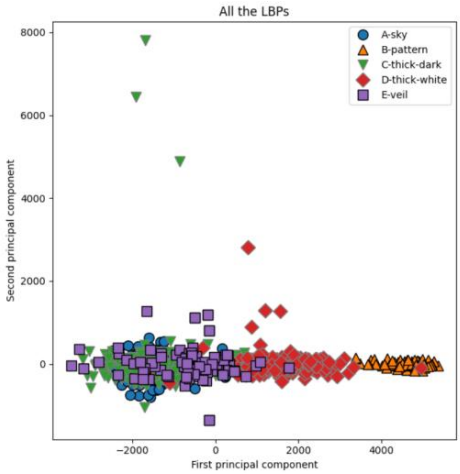

Figure 10. PCA with all LBPs.

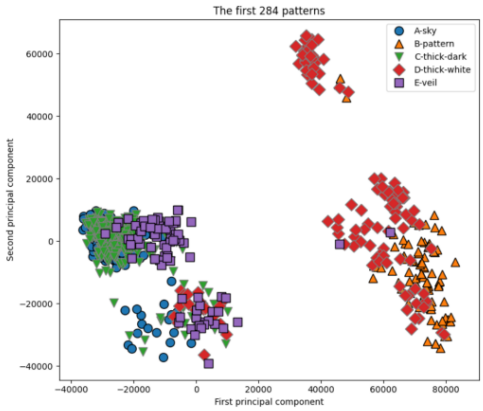

Figure 11. PCA with the first 284 patterns.

Also, in the KSVM, the report by using the first 284 patterns as features is as Table 5:

Table 5. Result with LBP (16, 2) – the first 284 patterns

	Precision	Recall	F1-score	Support
A-sky	0.56	0.63	0.59	51
B-pattern	1.00	0.84	0.91	25
C-thick-dark	0.54	0.58	0.56	59
D-thick-white	0.89	0.94	0.92	35
E-veil	0.83	0.58	0.68	26
Accuracy	/	/	0.69	196
Macro avg	0.77	0.71	0.73	196
Weighted avg	0.71	0.69	0.69	196

kernel='poly', gamma=0.0001,C=1

On this report we can find that the “B-pattern”, “C-thick-dark” and “D-thick-white” have been well classified. And the “A-sky” and “C-thick-dark” are still mixed together. As the graph by using PCA, the blue circle points and the green inverted triangle points are mixed, but the orange triangle, red rhombus and purple square points are been separated well.

DOI: 10.54254/2755-2721/2/20220541

5.2. Resnet

Table 6. Structure **of ResNet**

Layer name	Out-put size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2.x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3\times3, & 64 \\ 3\times3, & 64 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, & 64 \\ 3\times3, & 64 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, & 64 \\ 3\times3, & 64 \\ 1\times1, & 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, & 64 \\ 3\times3, & 64 \\ 1\times1, & 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, & 64 \\ 3\times3, & 64 \\ 1\times1, & 256 \end{bmatrix}\times3$
conv3.x	28×28	$\begin{bmatrix} 3\times3, & 128 \\ 3\times3, & 128 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, & 128 \\ 3\times3, & 128 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, & 128 \\ 3\times3, & 128 \\ 1\times1, & 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, & 128 \\ 3\times3, & 128 \\ 1\times1, & 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, & 128 \\ 3\times3, & 128 \\ 1\times1, & 512 \end{bmatrix}\times8$
conv4.x	14×14	$\begin{bmatrix} 3\times3, & 256 \\ 3\times3, & 256 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, & 256 \\ 3\times3, & 256 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, & 256 \\ 3\times3, & 256 \\ 1\times1, & 1024 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, & 256 \\ 3\times3, & 256 \\ 1\times1, & 1024 \end{bmatrix}\times23$	$\begin{bmatrix} 1\times1, & 256 \\ 3\times3, & 256 \\ 1\times1, & 1024 \end{bmatrix}\times36$
conv5.x	7×7	$\begin{bmatrix} 3\times3, & 512 \\ 3\times3, & 512 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, & 512 \\ 3\times3, & 512 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, & 512 \\ 3\times3, & 512 \\ 1\times1, & 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, & 512 \\ 3\times3, & 512 \\ 1\times1, & 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, & 512 \\ 3\times3, & 512 \\ 1\times1, & 2048 \end{bmatrix}\times3$
	1×1	Average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

There are many types of the ResNet shown in table 6. According to the depth, there are ResNet18, ResNet34, ResNet50, ResNet101, ResNet152 and so on. Because the dimension of cloud pictures is small, and have been preprocessed by the LBP, so the data information in these pictures is small. If we used to deep model, maybe cause overfitting. ResNet34 will done it well [21]. The table below shows the structure of ResNet we used in the paper.

We the optimizer we used is the SGD, and the parameter is learning rate=0.001, momentum=0.9, weight decay=5e-4. Also, the criterion we used is the Cross Entropy Loss. The structure of the ResNet is shown as Table 7.

DOI: 10.54254/2755-2721/2/20220541

Table 7. The structure of ResNet

Name of Layer	Output size	Parameters
Input Layer	125*125*3	The picture deal with LBP
Convolution layer 1	64*64*64	Convolution kernel=7*7, stride=64
Max pooling layer	32*32*64	Pooling kernel=3*3, step_size=2
Convolution group 2	32*32*128	$\begin{bmatrix} 3\times 3, & 64 \\ 3\times 3, & 64 \end{bmatrix} \times 2$
Convolution group 3	16*16*128	$\begin{bmatrix} 3\times 3, & 128 \\ 3\times 3, & 128 \end{bmatrix} \times 2$
Convolution group 4	8*8*256	$\begin{bmatrix} 3\times 3, & 256 \\ 3\times 3, & 256 \end{bmatrix} \times 2$
Convolution group 5	4*4*512	$\begin{bmatrix} 3\times 3, & 512 \\ 3\times 3, & 512 \end{bmatrix} \times 2$
Averaged pooling layer	1*1*512	Adaptive pooled kernel
Connection layer 1	10	512->10，Activation=Relu
Connection layer 2	5	10 -> 5, Output classification results

In 5 times epoch, the best report is shown in Table 8:

Table 8. The result of the ResNet

	Precision	Recall	F1-score	Support
A-sky	0.93	0.91	0.92	45
B-pattern	0.90	1.00	0.95	18
C-think-dark	0.92	0.92	0.92	51
D-thick-white	1.00	0.93	0.96	27
E-veil	0.94	1.00	0.97	17
Accuracy	/	/	0.94	158
Macro avg	0.94	0.95	0.94	158
Weight avg	0.94	0.94	0.94	158

Epoch = 5, batch size = 10.

5.3. Custom vision

From the custom vision, we can get multiple outputs:

1) Probability Threshold: this is the level of confidence that a prediction needs to have to be considered correct,
2) Precision: indicates the fraction of identified classifications that were correct.
3) Recall: indicates the fraction of actual classifications that were correctly identified.
4) AP: average precision

DOI: 10.54254/2755-2721/2/20220541

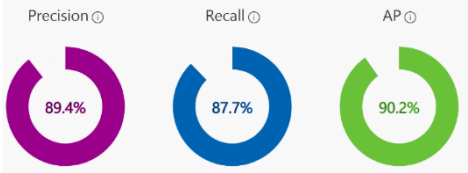

Figure 12. Azure Custom Vision Output

As figure 12 shows, after six-hour training in the custom vision, with the probability threshold in around 62%, we could reach the almost highest precision and recall in the custom vision, and the average precision is around 90.2%.

6. Conclusion

Various advanced classification models have been implemented today to do the image classification jobs in high precision field, including based-cloud classification. Increasing the accuracy of based-cloud classification not only could help scientists to minimize the error rate while classifying artificially, but also the public could receive benefits from it while they can scan a cloud-photo and know the type of the cloud on the sky. With the help of LBP preprocessing the dataset, this project had investigated and compared several different popular classification models, such as KSVM, MLP, Custom Vision, and Resnet. From the result of ResNet34, we successfully reach out a high accuracy around 94%.

In the future, since we had only investigated and used one preprocessing approach LBP, trying more preprocessing approaches will possibly make the accuracy higher. Moreover, we will use larger dataset that we introduced in the beginning to run over the classification model comparisons, because we had only implemented a small dataset SWIMMCAT with around 780 based-cloud images. Larger dataset may boost the final accuracy score. Furthermore, even though ResNet34 already provided a sufficiently high accuracy, using several more classification models may possibly also rise the accuracy.

7. Acknowledgement

Tao Zhang, Yantian Ding, these two authors contributed equally to this work and should be considered co-first authors.

References

[1] WMO. (2017) Understanding clouds. public.wmo.int/zh-hans/%E4%B8%96%E7%95%8C%E6%B0%94%E8%B1%A1%E6%97%A52017/%E4%BA%91%E7%9A%84%E5%88%86%E7%B1%BB.

[2] Chen, T., Rossow, W. B., and Zhang, Y.. (2000) Radiative effects of cloud-type variations. Journal of Climate, 13: 264–286.

[3] Ghonima, M. S., Urquhart, B., Chow C.W., Shields J. E., Cazorla A., and Kleissl J. (2012) A method for cloud detection and opacity classification based on ground based sky imagery. Atmospheric Measurement Techniques, 5: 2881–2892.

[4] Long, C. N.,Sabburg, J. M., Calbó, J., and Pagès, D.. (2006) Retrieving cloud characteristics from ground-based daytime color all-sky images. Journal of Atmospheric and Oceanic Technology, 23: 633–652.

[5] Shields, J. E., Karr, M. E., Tooman, T. P., Sowle, D. H., and Moore, S. T.. (1998) The whole sky imager – a year of progress. Proceedings of Eighth Atmospheric Radiation Measurement (ARM) Science Team Meeting: 677-685.

[6] Shaw, J. A., and Thurairajah, B.. (2003) Short-term Arctic cloud statistics at NSA from the infrared cloud image. Proceedings of the Thirteenth Atmospheric Radiation Measurement (ARM) Science Team Meeting: 1-7.

DOI: 10.54254/2755-2721/2/20220541

[7] Long, C., and Slater, D., Tooman, T.. (2001) Total Sky Imager Model 880 Status and Testing Results. DOE Office of Science Atmospheric Radiation Measurement (ARM) Program.
[8] Sun, X. J., Gao, T. C., Zhai, D. L., and Zhao, S. J.. (2008) Whole sky infrared cloud measuring system based on the uncooled infrared focal plane array (in Chinese). Infrared Laser Eng., 37: 761–764.
[9] Huo, J., and Lu, D..(2002) A primary study on cloud-cover using all-sky digital camera (in Chinese). Journal of Nanjing Institute of Meteorology, 25: pp. 242–246.
[10] Zhang, J. L., Liu, P., Zhang, F., and Song, Q.. (2018). CloudNet: Ground-based cloud classification with deep convolutional neural network. Geophysical Research Letters, 45: 8665–8672.
[11] Li, M., Liu S., and Zhang Z.. (2020) Deep tensor fusion network for multimodal ground-based cloud classification in weather station networks. Ad Hoc Networks, 96: 101991.
[12] Wan, X., and Du, J.. (2020). Cloud Classification For Ground-Based Sky Image Using Random Forest. ISPRS, XLIII-B3-2020: 835-842.
[13] Zhang, J., Liu, P., Zhang, F., and Song, Q.. (2018) CloudNet: Ground-based cloud classification with deep convolutional neural network. ResearchGate, 45: 8665-8672.
[14] Hasenbalg, M.,Kuhn, P., Wilbert, S., Nouri, B., Kazantzidis, A.. (2020) Benchmarking of six cloud segmentation algorithms for ground-based all-sky imagers. Solar Energy, 201: 596-614.
[15] Xi, D. Z. M.. (2014) Analysis of LBP principle and source code. blog.csdn.net/xidianzhimeng/article/details/19634.
[16] Wang, Y., Shi, C., Wang, C., and Xiao, B.. (2018) Ground-based cloud classification by learning stable local binary patterns. Atmospheric Research, Volume 207: 74-89.
[17] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011) Scikit-learn: Machine Learning in {P}ython. Journal of Machine Learning Research, Volume12: 2825-2830.
[18] Berseth, G.. (2019) Scalable deep reinforcement learning for physics-based motion control. https://www.semanticscholar.org/paper/Scalable-deep-reinforcement-learning-for-motion-Berseth/4c5dca886f1d5cc4213ece9cfb7895c7f48fd1be
[19] Tsang, S.. (2018) Review: ResNet — Winner of ILSVRC 2015 (Image Classification, Localization, Detection). https://towardsdatascience.com/review-resnet-winner-of-ilsvrc-2015-image-classification-localization-detection-e39402bfa5d8
[20] Patrick, F., and Coulter, D.. (2021) What is custom vision? – Azure cognitive services. docs.microsoft.com/en-us/azure/cognitive-services/Custom-Vision-Service/overview.
[21] Yang, Y., Wen, Z., and Xia, J.. (2021) Star-galaxy separation using ResNet[J]. Journal of Beijing Normal University (Natural Science).