Hindawi
BioMed Research International
Volume 2022, Article ID 4067581, 17 pages
https://doi.org/10.1155/2022/4067581

Review Article

Pooling *Operations in Deep* Learning: From “Invariable” to “Variable”

Zhou Tao,[1,2] Chang XiaoYu,[1] Lu HuiLing,[3] Ye XinYu,[1] Liu YunCan,[1] and Zheng XiaoMin[4]

[1] *School of Computer Science and Engineering, North Minzu University, Yinchuan 750021, China*
[2] *Key Laboratory of Image and Graphics Intelligent Processing of State Ethnic Affairs Commission, North Minzu University, Yinchuan 750021, China*
[3] *School of Science, Ningxia Medical University, Yinchuan 750004, China*
[4] *Research Institute for Reproductive Medicine and Genetic Diseases, Wuxi Maternity and Child Health Hospital, Wuxi 214002, China*

Correspondence should be addressed to Chang XiaoYu; yu0410yu@163.com and Lu HuiLing; lu_huiling@163.com

Received 6 May 2022; Accepted 27 May 2022; Published 20 June 2022

Academic Editor: Chen Li

Deep learning has become a research hotspot in multimedia, especially in the field of image processing. Pooling operation is an important operation in deep learning. Pooling operation can reduce the feature dimension, the number of parameters, the complexity of computation, and the complexity of time. With the development of deep learning models, pooling operation has made great progress. The main contributions of this paper on pooling operation are as follows: firstly, the steps of the pooling operation are summarized as the pooling domain, pooling kernel, step size, activation value, and response value. Secondly, the expression form of pooling operation is standardized. From the perspective of “invariable” to “variable,” this paper analyzes the pooling domain and pooling kernel in the pooling operation. Pooling operation can be classified into four categories: invariable of pooling domain, variable of pooling domain, variable of pooling kernel, and the pooling of invariable “+” variable. Finally, the four types of pooling operation are summarized and discussed with their advantages and disadvantages. There is great significance to the research of pooling operations and the iterative updating of deep learning models.

1. Introduction

Deep learning was proposed by Hinton [1] in 2006. A Convolutional Neural Network (CNN) was proposed by Le et al. [2], which is the first real multilayer structure learning algorithm. Deep learning has been widely used in various fields of multimedia such as text, image, and video, and it has become a research hotspot in the field of images. Convolution operation is the main step of the CNN, but it makes the whole network have a large number of calculations [3]. And it takes a long time for network training. Therefore, the pooling operation is an essential step in CNN.

Pooling operation is a common step in CNN and is also called subsampling operation or downsampling operation. When CNN was constructed, the pooling operation is usually placed behind the convolution layer. From the perspective of lightweight, pooling operation can reduce feature dimensionality, computation complexes, parameter number, and memory consumption and suppress noise. From the perspective of improving image quality, pooling operation can improve the horizontal movement invariance, scale invariance, and rotation invariance of the model. The pooling operation improves the robustness to image bias and rotation. From the perspective of the deep learning model, pooling operation can prevent overfitting. Pooling operation can improve the generalization ability of the model and expand the receptive field. Hence, it is very interesting to discuss pooling operation.

2. The “Variable” and “Invariable” of Pooling Operation

There are five parts in pooling operation, such as pooling domain, pooling kernel, step length, activation value, and response value [4]. The process of the pooling operation is shown in Figure 1. The “invariable” of pooling operation means that there is no change in the shape and size of the pooling domain and the shape, size, and parameter number of the pooling kernel. The “variable” of pooling operation means that there are two types: first is pooling domain changing and second is pooling kernel changing. About pooling domain changing, the shape and size are changed in the pooling domain. About pooling kernel changing, the shape, size, and parameters are changed in the pooling kernel [5].

Let the target image be $F(m, n)$, m is the width of the image, and n is the height of the image. The pooling domain is $f(i, j)$, where $i = 1, 2, \cdots, n$ and $j = 1, 2, \cdots, m$, which is the size of the receptive field in the pooling process. The pooling kernel is $w(i, j)$, where $i = 1, 2, \cdots, n$ and $j = 1, 2, \cdots, m$, which is the filter of the pooling domain. The pooling step length is the step length of the pooling nucleus every time it moves. The activation value $x_{i,j}$ is the feature value of the i-th row and j-th column in the feature map. The pooling feature value of the i-th row and j-th column is obtained by pooling operation.

In the pooling domain of size $i \times j$, the value in the pooling domain is determined by the feature map. The pooling domain $f(i, j)$ is determined by the pooling kernel $w(i, j)$, and the pooling kernel is a matrix of size $i \times j$. Generally, the pooling kernel has no parameters. The pooling domain $f(i, j)$ can be represented as

$$f(i, j) = \begin{bmatrix} x_{p,q} & x_{p,q+1} & \cdots & x_{p,q+j} \\ x_{p+1,q} & x_{p+1,q+1} & \cdots & x_{p+1,q+j} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p+i,q} & x_{p+i,q+1} & \cdots & x_{p+i,q+j} \end{bmatrix}_{i\times j}. \quad (1)$$

It can be concluded that the response feature map $g(i, j)$ is determined by the pooling kernel $w(i, j)$ in the feature map F of size $m \times n$. It is expressed as

$$g(i, j) = \begin{bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,j} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,j} \\ \vdots & \vdots & \ddots & \vdots \\ y_{i,1} & y_{i,2} & \cdots & y_{i,j} \end{bmatrix}_{i\times j}, \quad (2)$$

$$y_{i,j} = \beta \mathrm{down}(x(i, j)) + b,$$

where down(.) represents the downsampling operation, which is the pooling kernel $w(i, j)$ for the pooling operation of values in the pooling domain $f(i, j)$. β is the pooling coefficient. $y_{i,j}$ is the response value of the pooling domain. The response value is the value of the response feature map in the i-th row and j-th column. And b is the bias value of pooling operation.

The development of pooling operation is a process from “invariable” to “variable,” as shown in Figure 2. The most commonly used pooling operation is max pooling operation and average pooling operation. Max pooling operation and average pooling operation are calculated simply. They have few parameters and work well on most tasks. So far, these pooling operations are still the most commonly used pooling methods. In addition, max pooling operation and average pooling operation cannot achieve good results on specific tasks. Many scholars have further studied the pooling operation for different problems. Firstly, change the value of pooling selection. Secondly, change the pooling domain, which is including the change of shape and size. Thirdly, change the pooling kernel, which is including the change in shape, size, and parameters. Finally, there are various pooling methods that are combined, which have parallel methods and serial methods. The parallel method is the most commonly used.

In this paper, the “variable” and “invariable” of the pooling operation are divided into four categories: invariable of pooling domain, variable of pooling domain, variable of pooling kernel, and the invariable “+” variable of pooling. The overall framework is shown in Figure 3. The invariable of pooling domain means that there is no change in the shape and size of the pooling domain. The main methods are max pooling operation, maximum two mean pooling operation, K-max pooling operation, average pooling operation, median pooling operation, sum pooling operation, overlapping pooling operation, matrix 2-norm pooling operation, moment pooling operation, covariance pooling operation, dynamic adaptive pooling operation, and nested invariance pooling operation. The variable of pooling domain means that there is change in the shape and size of the pooling domain. The main methods are region of interest pooling method, fractional max pooling method, strip pooling method, chunk-max pooling method, and multiscale orderless pooling method. The variable of pooling kernel means that there is change in the parameters, selection method, and size of the pooling kernel. The main methods are generalized max pooling operation, parameter pooling operation, LP pooling operation, spatial attention pooling operation, stochastic pooling operation, spatial pyramid pooling operation, and rank-based pooling (rank-based average pooling operation, rank-based weighted pooling operation, and rank-based stochastic pooling operation) operation. The invariable “+” variable of pooling means that there are two or more pooling methods in pooling methods, including variable pooling operation and invariable pooling operation. The main methods are intermediate value pooling operation, mixed pooling operation, spatial pyramid and global average hybrid pooling operation, multiscale pooling operation, and scalable overlapping slide pooling operation.

3. Invariable of Pooling Domain

Invariable of pooling domain means that there is no change in the shape and size of the pooling domain. And there is

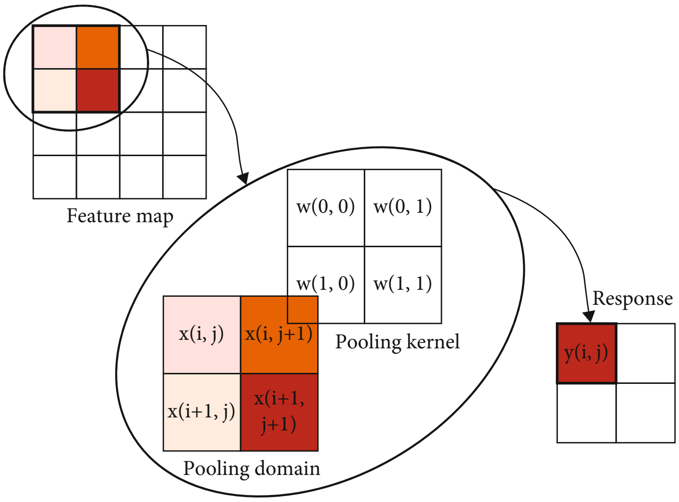


Figure 1: The process of pooling operation.

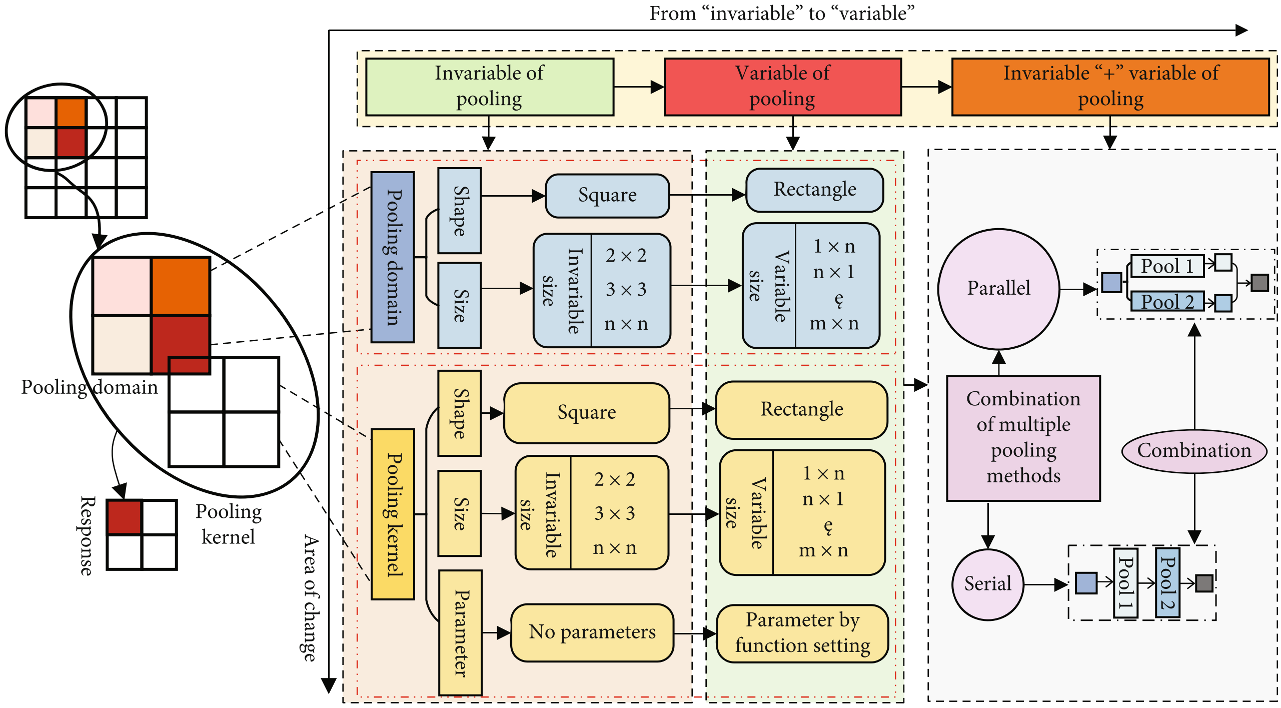


Figure 2: "Invariable" to "variable" of pooling operation.

change in the shape, size, and parameters of the pooling kernel. In this paper, the variable of pooling domain is divided into two methods: basic pooling operation and improved pooling operation. The basic pooling operation has the advantages of simple calculation process and wide application range. But the utilization rate is low in the feature map. The improved pooling operation uses different mapping rules to select the eigenvalues in the pooling domain. So the improved pooling operation can improve the utilization rate of the feature maps. But the calculation process is more complicated. The development process of invariable of pooling domain is shown in Figure 4.

3.1. Basic Pooling Operation. The basic pooling operation during the pooling operation is that the shape and size of the pooling domain are unchanged, and the shape, size, and parameters of the pooling kernel are unchanged. Select the eigenvalues with different numbers in the pooling domain, including max pooling operation, maximum two-mean pooling operation, *K*-max pooling operation, average pooling operation, median pooling operation, sum pooling operation, and overlapping pooling operation.

Firstly, max pooling operation selects the largest activation value in the pooling domain as the output response value. Other activation values are discarded at the same time.

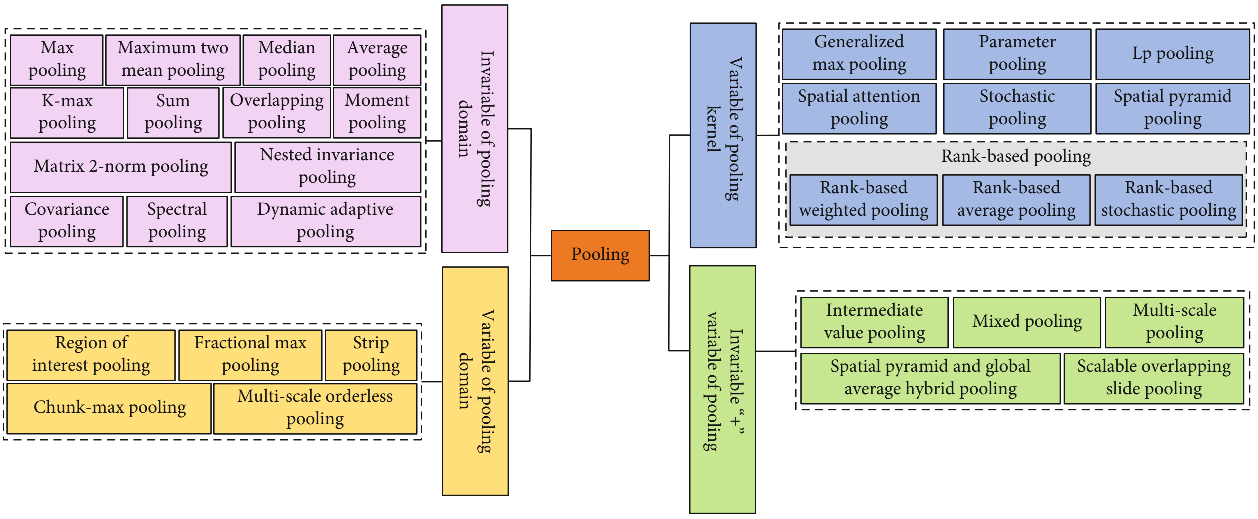

Max pooling
Maximum two mean pooling
Median pooling
Average pooling
K-max pooling
Sum pooling
Overlapping pooling
Moment pooling
Matrix 2-norm pooling
Nested invariance pooling
Covariance pooling
Spectral pooling
Dynamic adaptive pooling
Invariable of pooling domain
Region of interest pooling
Fractional max pooling
Strip pooling
Chunk-max pooling
Multi-scale orderless pooling
Variable of pooling domain
Pooling
Variable of pooling kernel
Generalized max pooling
Parameter pooling
Lp pooling
Spatial attention pooling
Stochastic pooling
Spatial pyramid pooling
Rank-based pooling
Rank-based weighted pooling
Rank-based average pooling
Rank-based stochastic pooling
Invariable "+" variable of pooling
Intermediate value pooling
Mixed pooling
Multi-scale pooling
Spatial pyramid and global average hybrid pooling
Scalable overlapping slide pooling

FIGURE 3: Summary of pooling methods.

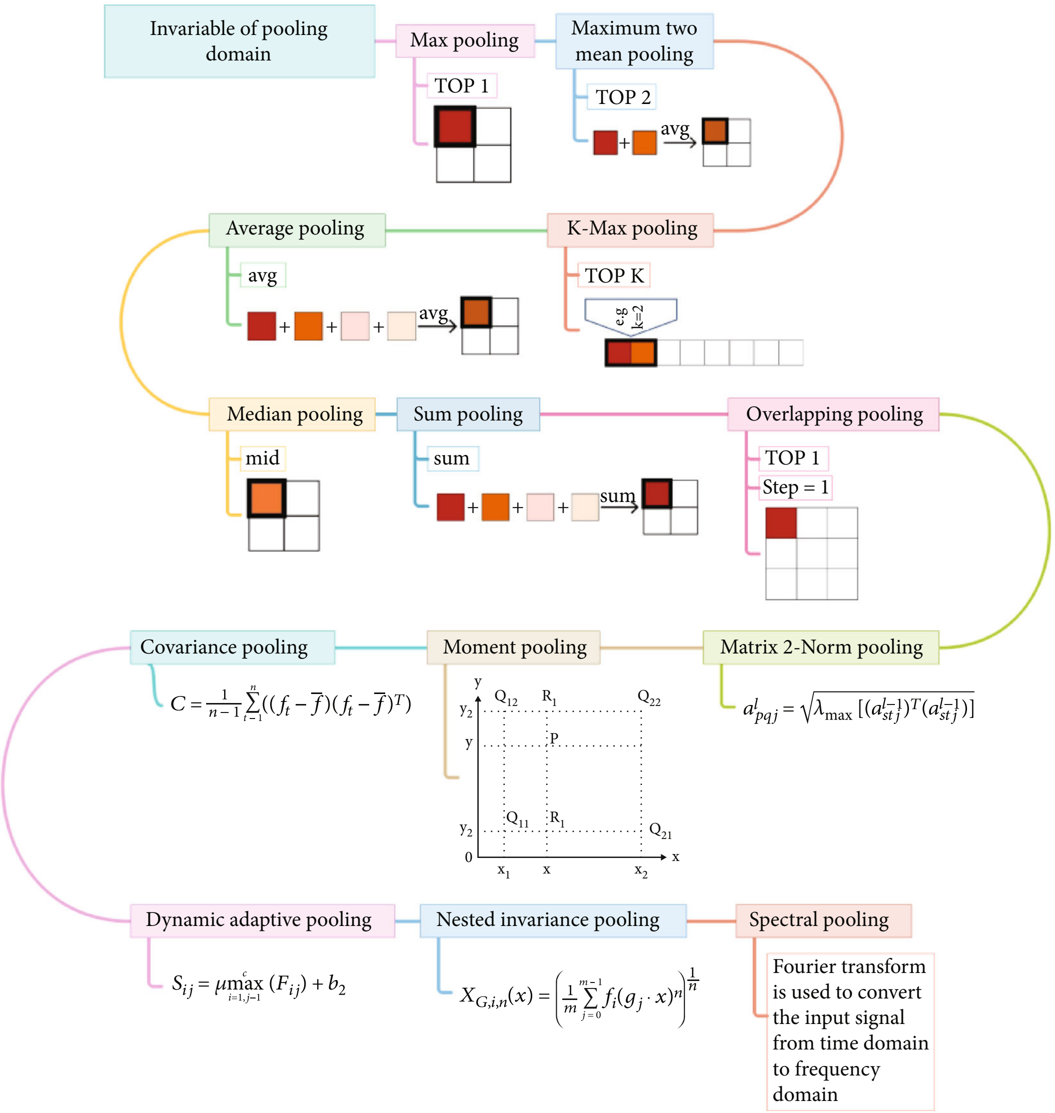

Invariable of pooling domain
Max pooling
TOP 1
Maximum two mean pooling
TOP 2
+
avg
Average pooling
avg
+
+
+
avg
K-Max pooling
TOP K
e.g k=2
Median pooling
mid
Sum pooling
sum
+
+
+
sum
Overlapping pooling
TOP 1
Step = 1
Covariance pooling
$C = \frac{1}{n-1}\sum_{t-1}^{n}((f_t - \overline{f})(f_t - \overline{f})^T)$
Moment pooling
y
Q_{12}
R_1
Q_{22}
y_2
y
P
Q_{11}
R_1
y_2
Q_{21}
0
x
x_1
x
x_2
Matrix 2-Norm pooling
$a^l_{pqj} = \sqrt{\lambda_{\max}[(a^{l-1}_{stj})^T(a^{l-1}_{stj})]}$
Dynamic adaptive pooling
$S_{ij} = \mu \max_{i=1,j-1}^{c} (F_{ij}) + b_2$
Nested invariance pooling
$X_{G,i,n}(x) = \left(\frac{1}{m}\sum_{j=0}^{m-1} f_i(g_j \cdot x)^n\right)^{\frac{1}{n}}$
Spectral pooling
Fourier transform is used to convert the input signal from time domain to frequency domain

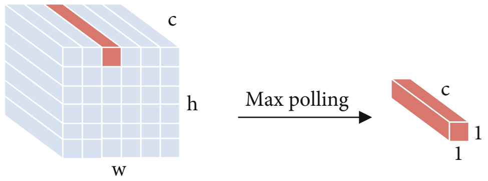
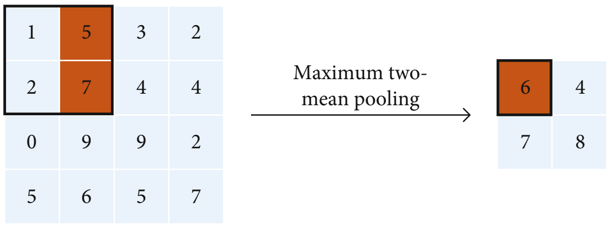
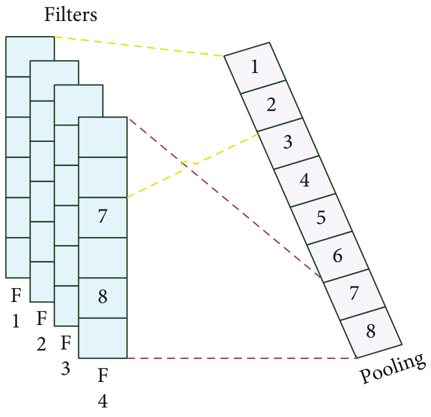
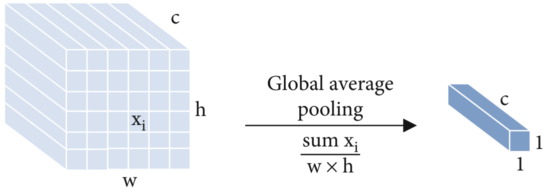
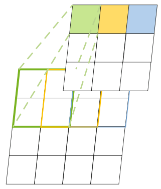
FIGURE 4: Invariable of pooling domain.

The advantage of max pooling operation is that salient features can be extracted, and the contour information of the image can be learned. However, max pooling operation loses too much features and information. The global max pooling operation is a case that max pooling operation is generalized to the global. The difference is that the size of the pooling domain is the same as the size of the entire feature map. Secondly, maximum two-mean pooling operation selects the two largest activation values in the pooling domain and averages them. Use the averaged value as the response value. The maximum two-mean pooling operation improves the disadvantage of ignoring features with larger influencing factors of max pooling. But there is some information missing. Thirdly, K-max pooling operation extracts the top K values in the pooling domain as the response value and preserves the order of these values. K-max pooling operation retains more information than max pooling, which can express the strength of a certain type of feature. K-max pooling operation retains position information, but it is the relative order between features, not absolute position information. Fourthly, average pooling operation means that the average of activation values in the pooling domain and the average value are used as the response value. Average pooling considers the global information, and there are no information loss and reducing overfitting. But it makes the features smooth and cannot extract the information of prominent features. The global average pooling operation is a case that average pooling operation is generalized to the global. The difference is that the size of the global average pooling domain is the same as the size of the entire feature map. Fifthly, median pooling operation takes the median of all activation values in the pooling domain as the response value. Median pooling operation has the characteristics of learning edge and texture structures and has a strong ability to resist noise, which can improve performance. Sixthly, sum pooling operation takes the sum value of all activation values in the pooling domain as the response value. Sum pooling operation can be applied to any encoding, and the impact of frequent descriptors is greater than that of rare descriptors. Seventhly, overlapping pooling [6] operation selects the largest activation value in the pooling domain as the output response value. The pooling domain of overlapping pooling operation has overlapping parts because its stride is 1. The overlapping pooling operation augments the output into multiple levels of smaller features. Overlapping pooling operation uses sparse coding to fuse multilevel features, which reduces the feature dimension of the output and improves the network performance. Table 1 shows the summary of basic pooling operation.

3.2. Improved Pooling Operation. The improved pooling operation means that there are no change in the shape and size of the pooling domain and change in the shape, size, and parameters of the pooling kernel. But the difference is the way to get the response value of eigenvalues in the pooling domain using the mapping rule. Improved pooling operation includes matrix 2-norm pooling operation, moment pooling operation, covariance pooling operation, dynamic adaptive pooling operation, nested invariance pooling operation, and spectral pooling operation.

Firstly, matrix2-norm pooling [7] operation makes the eigenvalues of feature map as the coefficient matrix of the singular equation system. Take the square root of the largest singular value obtained from this matrix as the response value of the pooling domain. When the image is geometrically distorted such as rotation, translation, and scaling, there are little changes in the singular value of the image matrix. This means that the energy information of the image has high stability, so the singular value of the matrix can make the geometric distortion of the image highly invariant. Secondly, moment pooling [8] operation calculates the central moment $c(i, j)$ of the pooling domain using the gray moment. Calculates the four neighborhoods of the central moment c. Calculate the probability of each neighborhood using the distances between the four neighborhoods $x(i, j)$, $x(i+1, j)$, $x(i, j+1)$, $x(i+1, j+1)$, and center $c(i, j)$. The closer the distance is, the greater the probability of being selected, which is conducive to the probability of selecting the response value. Because the randomness of moment pooling operation makes each selection different, it can effectively prevent oversuppression. Moment pooling operation randomly generates 2 probabilities. The computational complexity of the moment pooling operation is much lower than bilinear interpolation and close to the nearest neighbor interpolation. Thirdly, covariance pooling [9] operation calculates the covariance matrix using the eigenvalue matrix of the feature map and then adjusts the matrix to what you need. For example, the matrix is normalized to get the response values. Covariance pooling operation [10] belongs to the second-order pooling method, which can capture more information about the feature map than most first-order pooling methods. Fourthly, dynamic adaptive pooling operation [11] optimizes max pooling operation with a pooling factor μ. Dynamic adaptive pooling operation builds a mathematical model for function simulation, and other parameters are the same as max pooling operation. The pooling factor will not lose accuracy when it deals with pooling domains with obvious max characteristics. It can weaken the influence of max pooling operation when it deals with other pooling domains. Therefore, the CNN can extract more accurate features under different iterations. Fifthly, nested invariance pooling [12] operation changes the output of feature maps in the pooling domain using invariant transformation, which is the response feature map. It can be extended to any transform set. Lou et al. [13] obtain global descriptors by concatenating all feature maps. They construct nested invariance pooling operation from existing local translation invariant feature maps. Zhang et al. [14] showed that various types of transformation invariants can be linked. Sixthly, spectral pooling [15, 16] operation is performed for dimensionality reduction by projecting onto the frequency basis set and then truncating the representation. It is a pooling method based on fast Fourier transform, which keeps the same number of parameters and retains more information. Spectral pooling operation can be regarded as a denoising method that allows specifying any arbitrary output feature dimension. The truncation of frequency representation

Table 1: Summary of basic pooling operation.

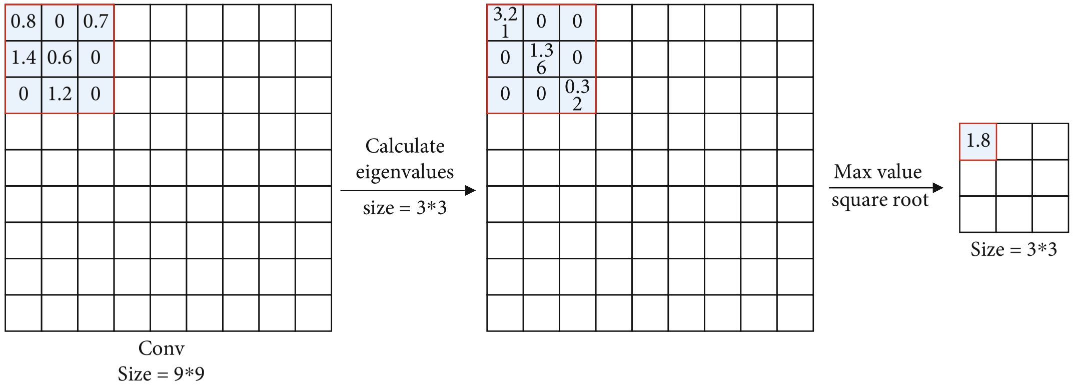
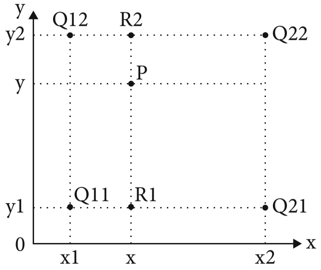
Pooling method	Characteristic	Sketch map
Max pooling	The max pooling operation is simple, but the features with strong influencing factors are ignored and lost information	
Maximum two-mean pooling	Maximum two-mean pooling operation improves the disadvantage of max pooling to a certain extent, which is ignoring features with larger influencing factors. But there is some problem of loss information	 1 5 3 2 2 7 4 4 0 9 9 2 5 6 5 7 Maximum two-mean pooling 6 4 7 8
K-max pooling	*K*-max pooling operation preserves more information than max pooling and preserves the relative order of eigenvalues, but there is no absolute position information	 Filters 7 8 F 1 F 2 F 3 F 4 1 2 3 4 5 6 7 8 Pooling
Average pooling	Average pooling operation takes global information into account, no information is lost, and overfitting is reduced. However, the features tend to be smooth, and the information of prominent features cannot be extracted	 c x_i h w Global average pooling $\frac{\text{sum } x_i}{w \times h}$ c 1 1
Median pooling	Median pooling operation can learn the characteristics of edge and texture structure and has a strong antinoise ability	$Y(F) = \text{Median}(S_1, S_2, \cdots, S_N)$ $S_i = \text{Median}\left(f_{i1}, f_{i1}, \cdots, f_{im_i}\right)$ $i = 1, 2, \cdots N$
Sum pooling	Sum pooling operation considers global information. There is no information loss, but it is vulnerable to extreme information	—
Overlapping pooling [6]	Overlapping pooling operation reduces the characteristic dimension of the output and reduces overfitting, but there are information that will be lost	

corresponds to the reduction in resolution. So spectral pooling operation can be supplemented with random regularization in the form of random resolution. Table 2 shows the summary of improved pooling operation.

4. Variable of Pooling Domain

Variable of pooling domain means that there is change in the shape and size of the pooling kernel. And there is no

Table 2: Summary of improved pooling operation.

Pooling method	Characteristic	Sketch map
Matrix2-norm pooling [7]	Matrix 2-norm pooling operation takes the energy of the image as the information transmitted to the next layer network to make geometric distortion of the image is highly invariable	 0.8 0 0.7 1.4 0.6 0 0 1.2 0 Conv Size = 9*9 Calculate eigenvalues size = 3*3 3.21 0 0 0 1.36 0 0 0 0.32 Max value square root 1.8 Size = 3*3
Moment pooling [8]	The randomness of moment pooling operation makes each choice different, to prevent over inhibition	 y y2 y y1 0 Q12 R2 Q22 P Q11 R1 Q21 x1 x x2 x
Covariance pooling [9]	Covariance pooling operation can capture more information on the feature map	$C=(1/n-1)\sum_{i=1}^{n}\left(\left(f_i-f_-\right)\left(f_i\bar{f}_-\right)^{T}\right)$
Dynamic adaptive pooling [11]	Dynamic adaptive pooling operation can adaptively adjust the weight of pooling and extract more accurate features	$y_{ij}=\mu\max_{i=1,j=1}^{c}\left(x_{ij}\right)+b_2$ $\mu=\rho\left(\lambda\left(x_{\max}-\lambda\right)/x_{\max}{}^{2}\right)+\theta\rho=c/1+\left(n_{epo}-1\right)c^{n_{epo}^{2}+1}$
Nested invariance pooling [12]	Nested invariance pooling operation can be extended to any arbitrary transformation set, allowing any arbitrary output feature dimension to be specified	$Y_{G,i,n}(x)=\left((1/m)\sum_{j=0}^{m-1}f_i\left(g_j.x\right)^{n}\right)^{1/n}$
Spectral pooling [15]	Spectral pooling operation has higher efficiency and lower cost	—

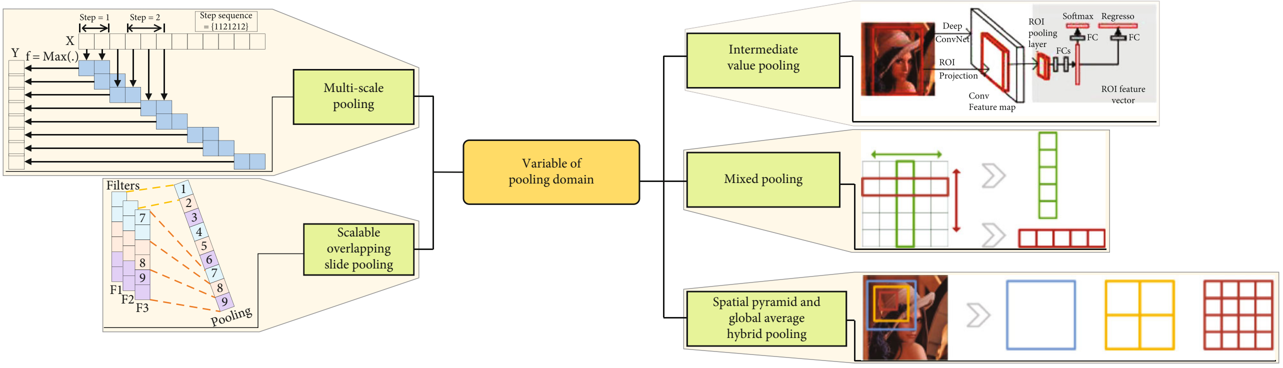


Figure 5: Variable of pooling domain.

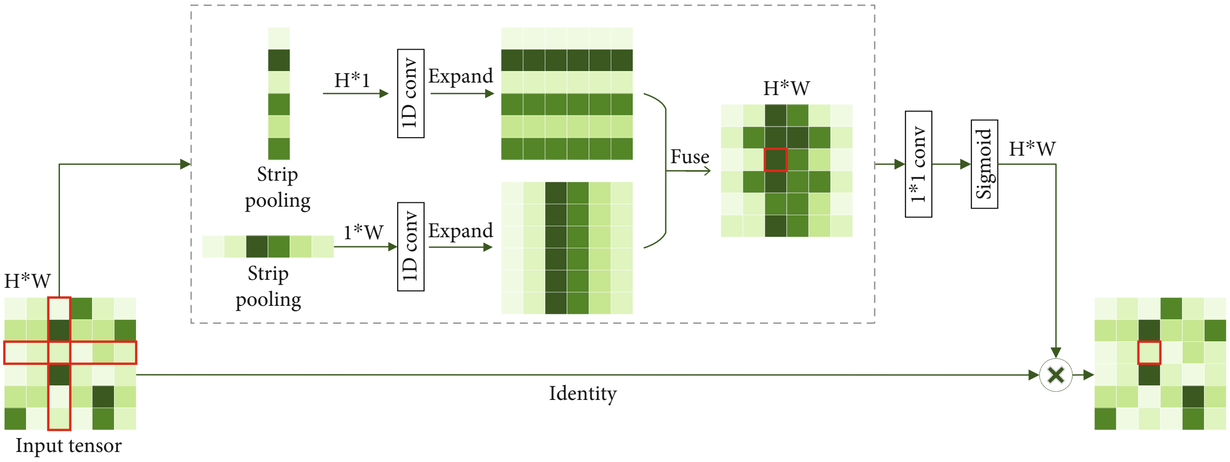


Figure 6: Schematic diagram of strip pooling [21].

change in the parameters of the pooling kernel. The shape and size of the pooling kernel are determined by the pooling domain. Variable of pooling domain can modify the size and shape of the pooling domain as needed, which is more flexible. But it leads to susceptibility to more irrelevant pixels when the pooling domain changes. Variable of pooling domain includes region of interest pooling operation, fractional max pooling operation, strip pooling operation, chunk-max pooling operation, and multiscale orderless pooling operation, as shown in Figure 5.

Firstly, region of interest pooling (ROI pooling) [17] is pooling just for the region of interest. There is no fix in the shape and size of the pooling domain which is depending on the region of interest. The pooling domain is the region of interest in which a max pooling operation in the pooling domain is done. ROI pooling operation was used in fast RCNN [17, 18] first. Its advantage is that the location area and shape size of the pooling domain can be determined due to the problem, and the size of the output feature map can be determined. And just the region of interest in the feature map needs to be pooled, which greatly reduces the amount of computation. In the calculation process, the position of the region of interest is easily selected due to the quantization operation. Secondly, fractional max pooling [19] operation divides the feature map of input size $b \times b$ into $a \times a$ blocks. These feature blocks of the $a \times a$ are $a \times a$ pooling domains, and then, select the max value of each pooling domain to obtain the response value. Fractional max pooling operation allows the pooling domain to be noninteger values. It avoids the problem of limited performance due to disjoint pooling kernels when the pooling domain is integer. Yue et al. [20] make random number sequences or pseudorandom sequences as the size of each pooling domain, and the integer sequence generates 1 or 2. It reduces the overfitting of the dataset due to the randomness of the sequence. Thirdly, strip pooling [21] operation performs pooling operations along the horizontal or vertical dimension using a strip pooling domain. The pooling domain of strip pooling operation is an $h \times 1$ or $1 \times w$ strip domain, and average values in the pooling domain rowwise or columnwise are shown in Figure 6. Strip pooling alleviates the problem of merging irrelevant regions in the process of processing objects with irregular shapes about average pooling. It can be easily embedded into any building block without training from scratch because the strip pooling module is lightweight. Strip pooling has long and narrow pooling field shapes. It is easy to establish long-range dependencies between discretely distributed regions and encode the strip-shaped regions. Fourthly, chunk-max pooling [22] operation segments the feature vector on the feature map as pooling domain. Chunk-max pooling operation performs max pooling operation for each pooling domain, obtaining multiple response values. This method of segmenting eigenvectors can be set in advance or dynamically divided according to eigenvalues. Static segmentation means that the number of paragraphs can be set in advance. Dynamic segmentation means that the boundary positions between chunks can be dynamically divided according to different inputs. Chunk-max pooling operation retains relatively coarse-grained fuzzy location information. Feature intensity can be captured if a strong feature occurs multiple times. Fifthly, multiscale orderless pooling [23] operation extracts deep activation features through three pooling domains of different sizes. Then, the local responses are aggregated on a finer scale by local aggregation vector (VLAD) coding. Finally, the original global deep activations are connected with the VLAD feature to form response value. The features obtained after activation of global CNNs lack geometric invariance, which includes rotation invariance and translation invariance. Multiscale orderless pooling operation is proposed to solve this problem. It limits their robustness to the classification and matching of variable scenes. Table 3 shows the summary of variable of pooling domain.

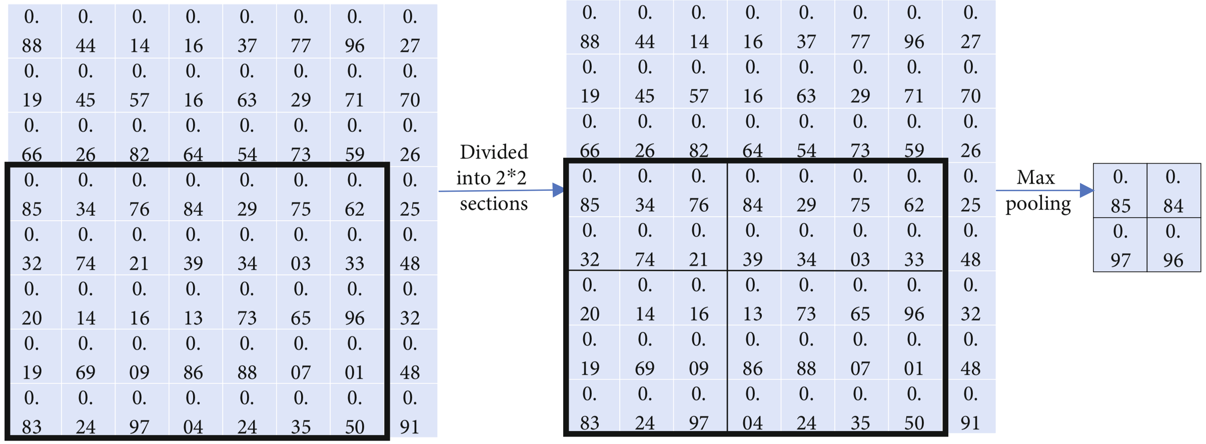
5. Variable of Pooling Kernel

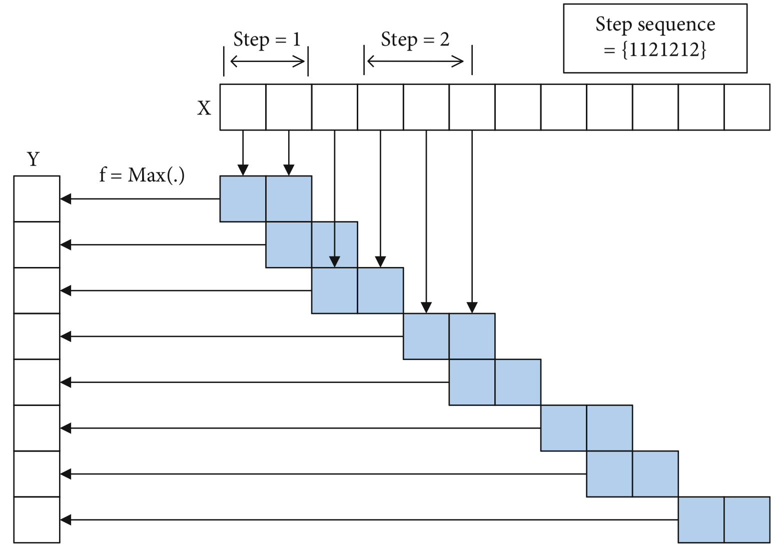
Variable of pooling kernel means that there is change in the shape, size, and parameters of the pooling kernel. Variable of pooling kernel can get a better solution for the problem to be solved, and the pooling changes the pooling kernel. In this paper, the variable of pooling kernel is divided into four categories: pooling kernel parameter invariable, random selection, rank-based pooling, and pooling kernel parameter variable, as shown in Figure 7. The pooling kernel parameter invariable is to set the parameters of the pooling kernel. The response value is obtained by multiplying the parameters in the pooling kernel with the activation value in the pooling domain. Random selection randomly selects values in the pooling domain as the output response value. Rank-based pooling sorts all the values in the pooling domain and then performs averages, weights, or random pooling on the ordered activation values. The pooling kernel parameter variable is the improvement of the pooling operation in front of the fully connected layer. The pooling kernel parameter variable uses three pooling kernels of different sizes to generate a fixed-length feature vector for the max pooling operation. The obtained result is passed to the fully connected layer.

Firstly, generalized max pooling [24] operation is a process of obtaining weights by solving the optimization problem. The weights are obtained according to the similarity between each element and other elements. Rare pixels with lower similarity will have larger weights. Finally, each element will be weighted and summed as the output. Generalized max pooling operation achieves the same effect as max pooling operation. It is also suitable for places other than the visual word package (BOV), especially for the Fisher vector. Secondly, parameter pooling [25] operation uses the operation rules of the pooling kernel parameters to obtain the parameters of the pooling kernel. The parameters of the pooling kernel are matrix multiplied by the activation value to obtain the response value. Generally, the pooling kernel has no parameters. Parameter pooling operation allows the model to choose the pooling method autonomously because it has a few parameters. Parameter pooling operation preserves feature information as much as possible. It reduces network errors, and the network performance is improved. Thirdly, LP pooling [26] operation uses a Gaussian kernel to weight the pixel values of the input feature map. At this time, the Gaussian kernel is the pooling kernel of the pooling operation, and the weighted value is the response value. The optimal pooling method for most problems is neither average pooling operation nor max pooling operation but a certain type between them. When $p = 2$, it is L2 pooling operation. There are good results that can be achieved in most image classification problems. Fourthly, spatial attention pooling [27] operation generates an attention parameter matrix for the entire feature map. The parameter matrix acts as a pooling kernel to perform a global pooling operation on the feature map. The value on the attention parameter matrix is multiplied by the values of the feature map to obtain a new matrix as the response result of spatial attention pooling operation. Spatial attention pooling operation is able to highlight regions that require attention. Fifthly, rank-based pooling [28] operation sorts the eigenvalues in the pooling domain from large to small. It performs average pooling operation, weighted pooling operation, and stochastic pooling operation on the sorted values, as shown in Figure 8. The traditional pooling methods operate directly on the values in the feature map. There are many possible values in a pooling domain, but the ranking of eigenvalues is relatively stable. Therefore, rank-based pooling operation can avoid the influence caused by the change of the value. Rank-based average pooling [29] (RAP) operation sorts the values of the feature maps from large to small and selects the t with the largest value to average the response value. RAP can alleviate the problem of information loss from max pooling operation and discriminative information loss from average pooling operation. Rank-based weighted pooling [30] (RWP) operation assigns weight p_r according to the size of the value. The value as the response value is obtained by multiplying the feature map matrix and the weight matrix. It can be assigned the larger weights for the higher activation value. The value ordering method of RWP can be a linear function or a nonlinear function. RWP assigns a reasonable weight value to each activation value in the pooling domain, which can improve the performance significantly. Rank-based stochastic pooling [31] (RSP) computes the probability p_r using exponential ranking on the resulting ranking values. It selects activation values by sampling from a multinomial distribution based on p_r. It is proposed to solve the problem that random pooling is just suitable for nonnegative values, and feature probability calculated by standardized eigenvalues is out of control, which is easy to lead to overfitting.

Sixthly, stochastic pooling [29] operation obtains the probability according to the value. The probability is used to select randomly, and the selected activation value is output. The higher the probability value, the greater the

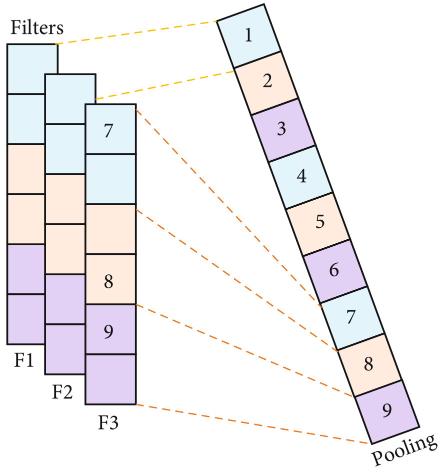
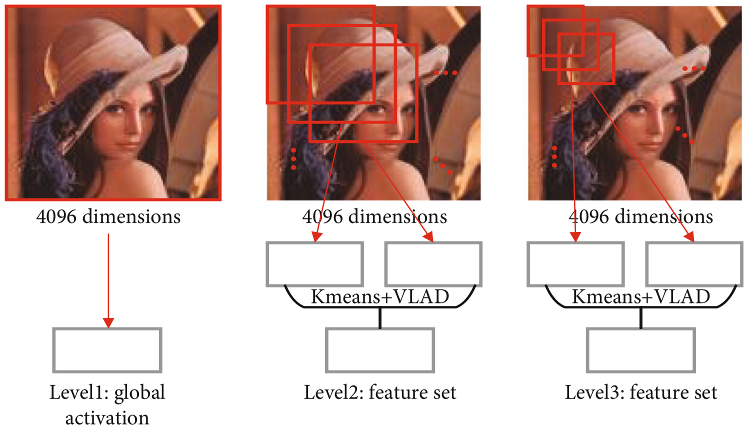
Table 3: Summary of variable of pooling domain.

Pooling method	Characteristic	Sketch map
Region of interest pooling [17]	ROI pooling operation converts feature maps within a region of interest of any size into feature maps of fixed size	 0. 88 0. 44 0. 14 0. 16 0. 37 0. 77 0. 96 0. 27 0. 19 0. 45 0. 57 0. 16 0. 63 0. 29 0. 71 0. 70 0. 66 0. 26 0. 82 0. 64 0. 54 0. 73 0. 59 0. 26 0. 85 0. 34 0. 76 0. 84 0. 29 0. 75 0. 62 0. 25 0. 32 0. 74 0. 21 0. 39 0. 34 0. 03 0. 33 0. 48 0. 20 0. 14 0. 16 0. 13 0. 73 0. 65 0. 96 0. 32 0. 19 0. 69 0. 09 0. 86 0. 88 0. 07 0. 01 0. 48 0. 83 0. 24 0. 97 0. 04 0. 24 0. 35 0. 50 0. 91 Divided into 2*2 sections Max pooling 0. 85 0. 84 0. 97 0. 96

Fractional max pooling [19]	Fractional max pooling operation allows the pooling domain to be noninteger values and reduces overfitting	 Step = 1
		Step = 2
		Step sequence = {1121212}
		X
		Y
		f = Max(.)

Strip pooling [21]	Strip pooling operation is easy to establish remote dependencies between discrete distributed regions. It can capture local details and can be easily embedded into any building block	$y_i^h = 1/w \sum_{0 \le j \le W} x_{i,j}, y_j^v = 1/H \sum_{0 \le i \le H} x_{i,j}$

Table 3: Continued.

Pooling method	Characteristic	Sketch map
Chunk-max pooling [22]	Chunk-max pooling operation retains the relative order information of multiple local max eigenvalues but does not retain the absolute position information	 Filters 7 8 9 F1 F2 F3 1 2 3 4 5 6 7 8 9 Pooling
Multiscale orderless pooling [23]	Multiscale orderless pooling operation improves the invariance of neural network activations without reducing its resolution	 4096 dimensions Level1: global activation 4096 dimensions Kmeans+VLAD Level2: feature set 4096 dimensions Kmeans+VLAD Level3: feature set

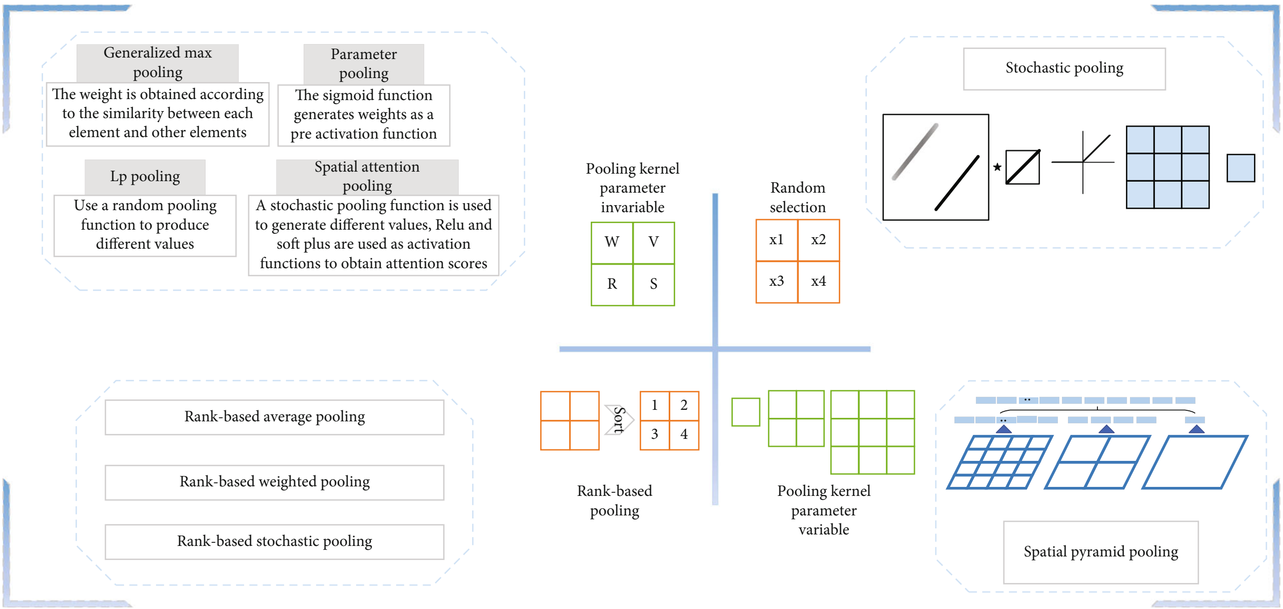

Generalized max pooling
The weight is obtained according to the similarity between each element and other elements
Parameter pooling
The sigmoid function generates weights as a pre activation function
Lp pooling
Use a random pooling function to produce different values
Spatial attention pooling
A stochastic pooling function is used to generate different values, Relu and soft plus are used as activation functions to obtain attention scores
Pooling kernel parameter invariable
W
V
R
S
Random selection
x1
x2
x3
x4
Stochastic pooling
Rank-based average pooling
Rank-based weighted pooling
Rank-based stochastic pooling
Sort
1
2
3
4
Rank-based pooling
Pooling kernel parameter variable
Spatial pyramid pooling

Figure 7: Variable of pooling kernel.

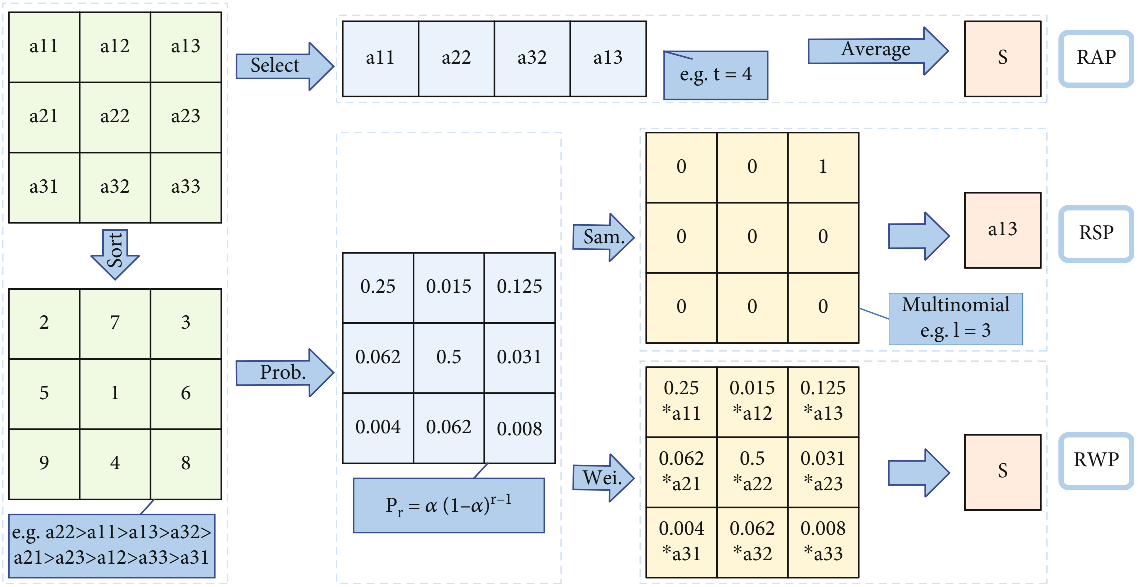


Figure 8: Rank-based pooling [28].

probability of being selected. Stochastic pooling operation is the process of random selection. When the selected pooling value is the max value in the pooling domain, the calculation process is equivalent to max pooling operation. Seventh, spatial pyramid pooling [33] operation performs max pooling operations on feature maps at three different scales. It replaces the last pooling layer with a spatial pyramid pooling layer, which can process information of different scales in the image to generate the fixed length feature information. Spatial pyramid pooling avoids the fact that most CNNs need to input fixed size images artificially, thus solving the limitation of image recognition accuracy due to fixed size. No matter how the feature map size changes, as long as the output length is fixed, the final number of features will be fixed. Images of different scales can be processed by Spatial Pyramid Pooling operation, so it is very flexible and can effectively prevent overfitting. However, some problems cannot be calculated in spatial pyramid pooling operation. When setting the output length, it is necessary to take into account the limitations of the actual situation. Table 4 shows the summary of variable of the pooling kernel.

6. The Invariable "+" Variable of Pooling

The invariable "+" variable of pooling means that it is a pooling method that uses different pooling methods for pooling separately and then fuses multiple pooling results.

Table 4: Summary of variable of the pooling kernel.

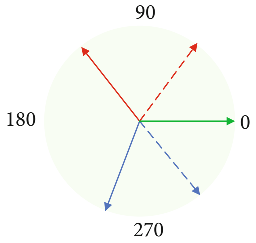
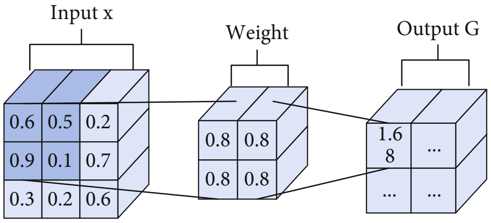
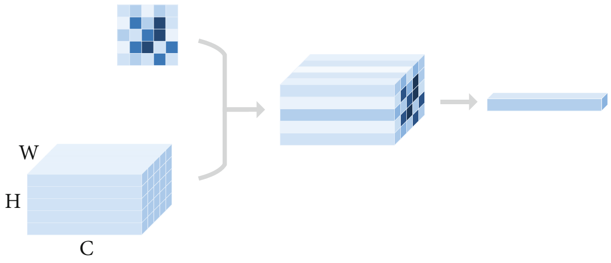
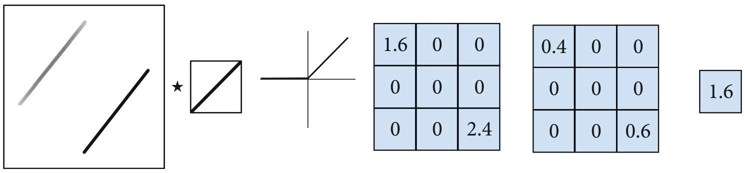
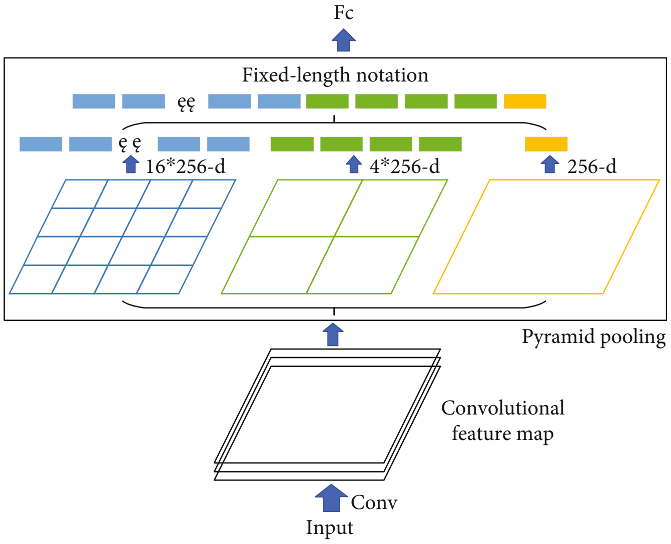
Pooling method		Characteristic	Sketch map
Generalized max pooling [24]		Generalized max pooling operation balances the influence of frequent pixels and rare data and improves the ability to extract fine-grained data	
Parameter pooling [25]		Parameter pooling operation converts the correlation operation into an interpretable pooling operation, which retains information and reduces errors	 Input x Weight Output G 0.6 0.5 0.2 0.9 0.1 0.7 0.3 0.2 0.6 0.8 0.8 0.8 0.8 1.6 8 ... ... $Y_c(i,j)=\sum_{m=0}^{p}\sum_{n=0}^{q}F_c(i+m,j+n)\cdot\beta(W_c(m,n))$
LP pooling [26]		LP pooling operation balances the effects of max pooling and average pooling.	$Y=\left(\sum\sum F(i,j)^P\times G(i,j)\right)^{1/P}$
Spatial attention pooling [27]		Spatial attention pooling operation mitigates the effects of distracting factors and focuses on meaningful parts of the image	
Rank-based pooling [28]	Rank-based average pooling	Alleviate the problem of loss of information in max pooling and loss of discriminative information in average pooling	$y_j=(1/t)\sum_{i\in f_j,r_i\le t}x_i$
	Rank-based weighted pooling	Assigning a weight value to each pixel in the pooling domain can improve performance	$p_r=\alpha(1-\alpha)^{r-1}$ $y_j=\sum_{i\in f_j}p_ix_i$
	Rank-based stochastic pooling	Alleviates the problem that random pooling is limited to nonnegative values and reduces overfitting	$p_r=\alpha(1-\alpha)^{r-1}$ $y_j=x_l$
Stochastic pooling [32]		Stochastic pooling operation is simple and has a strong generalization ability	 ⋆ 1.6 0 0 0 0 0 0 0 2.4 0.4 0 0 0 0 0 0 0 0.6 1.6 $p_i=x_i/\sum_{k\in f_j}x_k$ $y_j=x_l$, where $l\sim P\left(p_1,\cdots,p_{\lvert R_j\rvert}\right)$.

Table 4: Continued.

Pooling method	Characteristic	Sketch map
Spatial pyramid pooling [33]	Spatial pyramid pooling operation can handle images of different scales, is very flexible to use, and can effectively prevent overfitting. However, there are also practical constraints to consider	 Fc Fixed-length notation 16*256-d 4*256-d 256-d Pyramid pooling Convolutional feature map Conv Input

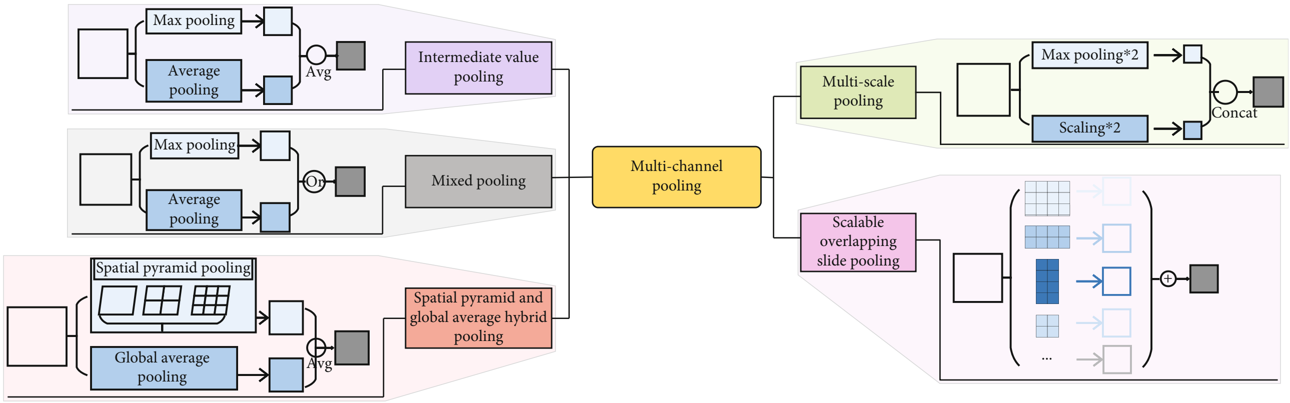


Figure 9: The pooling of invariable "+" variable.

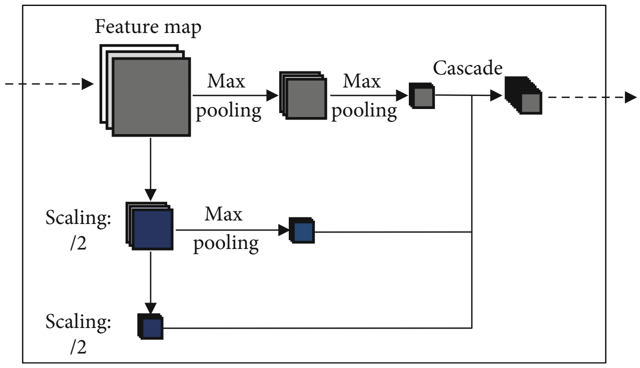
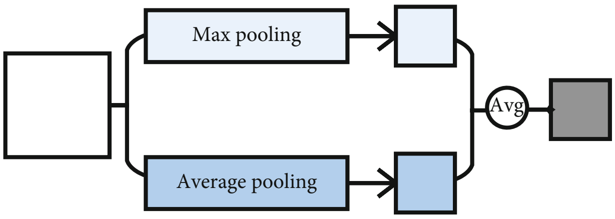
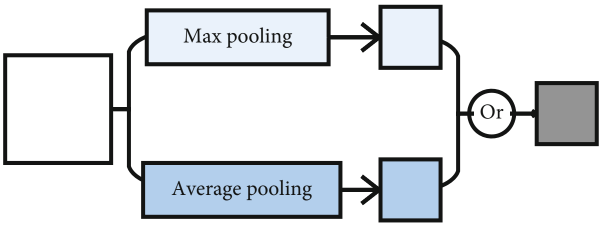
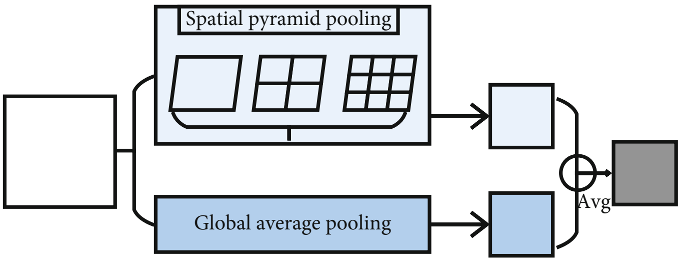
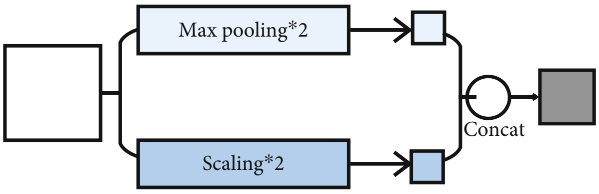


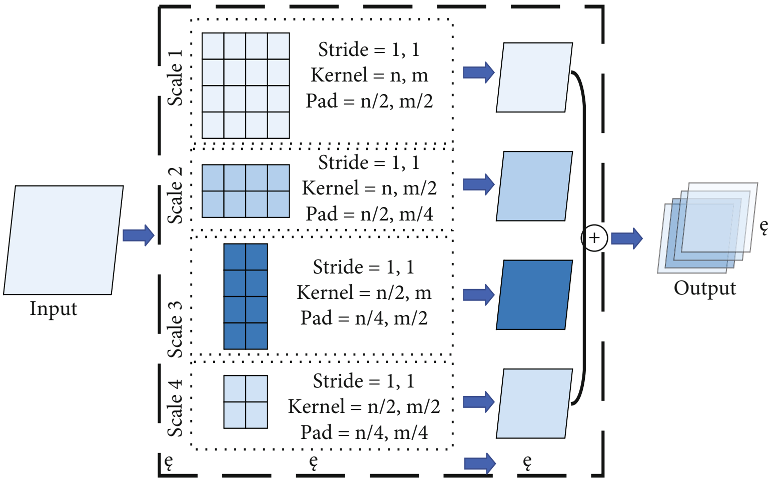
Figure 10: Multiscale pooling.

It can also be called multichannel pooling. Multichannel pooling operation means that there are two or more pooling methods in pooling methods to compensate for the defects of a single pooling operation. However, at the same time, it will increase the amount of computation. Multichannel pooling includes intermediate value pooling operation, mixed pooling operation, spatial pyramid and global average hybrid pooling operation, multiscale pooling operation, and scalable overlapping slide pooling operation, as shown in Figure 9.

Firstly, intermediate value pooling [34] operation is used to perform max pooling operation and average pooling operation on the same feature map in parallel. The response values of the two pooling methods are averaged to obtain the response feature map. It combines the characteristics of both and highlights the most discriminative features while preserving the information integrity of the feature map. It is generally suitable for general images, because the error of the model is smaller and the stability is higher. Secondly, mixed pooling [35] operation is random using max pooling operation or average pooling operation. Mixed pooling operation utilizes random sampling instead of deterministic pooling operations. To some extent, mixed pooling operation

Table 5: Summary of multichannel pooling.

Pooling method	Characteristic	Sketch map
Intermediate value pooling [34]	It takes into account the average pooling and the max pooling, so that it has smaller model error and higher stability	 $Y = 1/2 \times (\mathrm{AvgPool}(F) + \mathrm{MaxPool}(F))$
Mixed pooling [35]	It solves the problem of which to choose between max pooling and average pooling, but only using one of these pooling methods still has the disadvantage of max pooling or average pooling	 $y_{kij} = \lambda \max\left(x_{kqp}\right) + (1-\lambda)\left(1/\left\lvert f_{ij}\right\rvert\right) \cdot \sum x_{kqp}\,(p, q) \in f_{ij}$
Spatial pyramid and global average hybrid pooling [36]	It can extract local and global information, respectively, which learns more information	
Multiscale pooling	It is more flexible and can be used multiple times at the beginning, middle, or end of the network	

Scalable overlapping slide pooling	It considers the saliency of features at different scales and the relationship between adjacent feature elements, so that coarse-grained, medium-grained, and fine-grained multiple features can be extracted	 Input
		Scale 1
		Stride = 1, 1
		Kernel = n, m
		Pad = n/2, m/2
		Scale 2
		Stride = 1, 1
		Kernel = n, m/2
		Pad = n/2, m/4
		Scale 3
		Stride = 1, 1
		Kernel = n/2, m
		Pad = n/4, m/2
		Scale 4
		Stride = 1, 1
		Kernel = n/2, m/2
		Pad = n/4, m/4
		+
		Output

solves the problem of which to choose between max pooling operation and average pooling operation. It can greatly improve the overall performance of the model when combined with other forms of regularization methods. Thirdly, spatial pyramid and global average hybrid pooling [36] operation uses spatial pyramid pooling operation and global average pooling operation for pooling feature maps in parallel. There are two pooling methods fused to get response values.

Before fully connecting, global average pooling operation can reduce parameters, dimensionality, and overfitting without destroying the spatial characteristics of feature maps. Spatial pyramid pooling operation shows the information of local characteristics. Global average pooling shows the information of global characteristics. There are two pooling methods fused that can learn more comprehensive information. Fourthly, multiscale pooling operation performs max pooling operation twice serially. Then, perform scale operation twice serially. Next, perform max pooling operation and scale operation serially. Finally, concatenate the feature maps obtained by three pooling methods as the response value, as shown in Figure 10. Multiscale pooling operation replaces the max pooling operation in ResNet, which is used to improve the recognition effect of images collected at different resolutions. Compared with the spatial pyramid model, the advantage of multiscale pooling operation is that its substitution position is more flexible and it can be used many times at the beginning, middle, or end of a network. Fifthly, scalable overlapping slide pooling [37] operation pools the feature maps in parallel through pooling kernels of multiple scales. The overlapping slide pooling operation obtains coarse-grained, medium-grained, and fine-grained feature maps. Finally, the multiscale pooling operation response feature maps of multiscale pooling are spliced as the response value. Single-scale pooling operation does not consider the saliency of features in different scales and the relationship between adjacent feature elements, which leads to decreased inaccuracy. Multiscale pooling operation solved this problem. More comprehensive and accurate granular features can be obtained through multiscale overlapping sliding pooling operation. The detection ability of the network for small targets and the overall detection accuracy can be improved. Table 5 shows the summary of multichannel pooling.

7. Conclusions

With the continuous development of multimedia technology, deep learning is widely used in text, speech, image, video, etc. Among them, pooling operation in deep learning has been deeply studied. And pooling operation is iterated from “invariable” to “variable,” with different problems and tasks. These methods are constantly optimized due to different problems and tasks. Now, the most common methods of pooling are max pooling operation and average pooling operation. The newly proposed pooling method can improve the efficiency, improve the accuracy, and reduce the degree of overfitting.

In this paper, pooling operation in deep learning is comprehensively summarized. And introduce the pooling steps and their characteristics in detail and standardize the expression of pooling operation. The pooling operation is divided into the pooling domain, pooling kernel, step size, activation value, and response values. The improvement of pooling domain and pooling kernel in pooling operation is summarized as “variable” and “invariable” of pooling operation. Pooling operation can be classified into four categories: invariable of pooling domain, variable of pooling domain, variable of pooling kernel, and the pooling of invariable “+” variable. Invariable of pooling domain refers to the shape of the pooling domain being invariable in pooling operation. There are two types of invariable in pooling domain, including basic pooling methods and improved pooling methods. Variable of pooling domain means that shape and size of pooling domain are changed. Variable of pooling kernel means that the parameters, selection methods, and size of pooling kernel are changed. Pooling of invariable “+” variable means that there are two or more pooling methods in pooling methods. Although pooling operation is constantly being researched, deep learning is developing increasingly better, which has a better role for further research in deep learning.

Data Availability

The data that support the findings of this study are available from the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was financially supported by the National Natural Science Foundation of China (grant no. 62062003), Key Research and Development Project of Ningxia (grant no. 2020BEB04022), Natural Science Foundation of Ningxia (grant no. 2022AAC03149), and North Minzu University Research Project of Talent Introduction (grant no. 2020KYQD08).

References

[1] G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” *Science*, vol. 313, pp. 504–507, 2006.

[2] C. Le, Y. Boser, and B. Denker, “Backpropagation applied to handwritten zip code recognition,” *Neural Computing*, vol. 1, no. 4, pp. 541–551, 1989.

[3] T. Zhou, X. Ye, H. Lu, X. Zheng, S. Qiu, and Y. Liu, “Dense convolutional network and its application in medical image analysis,” *BioMed Research International*, vol. 2022, Article ID 2384830, 22 pages, 2022.

[4] T. Zhou, H. Lu, Z. Yang, S. Qiu, B. Huo, and Y. Dong, “The ensemble deep learning model for novel COVID-19 on CT images,” *Applied Soft Computing*, vol. 98, article 106885, 2021.

[5] Z. Tao, H. Bing-Qiang, L. Huiling, S. Hongbin, Y. Pengfei, and D. Hongsheng, “18F-FDG PETCT whole body imaging lung tumor diagnostic model: an ensemble E-Res Net-NRC with divided sample space,” *BioMed Research International*, vol. 2021, 13 pages, 2021.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[7] T. Li, Z. Meng, B. Ni, J. Shen, and M. Wang, “Robust geometric ℓp-norm feature pooling for image classification and action recognition,” *Image and Vision Computing*, vol. 55, no. 2, pp. 64–76, 2016.

[8] T. Otsuzuki, H. Hayashi, Y. Zheng, and S. Uchida, "Regularized pooling," in *Artificial Neural Networks and Machine Learning-ICANN 2020*, vol. 12397 of Lecture Notes in Computer Science, pp. 241–254, Springer, Bratislava, Slovakia, 2020.

[9] P. Li, J. Xie, Q. Wang, and Z. Gao, "Towards faster training of global covariance pooling networks by iterative matrix square root normalization," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 947–955, Salt Lake City, USA, 2018.

[10] D. Acharya, Z. Huang, D. P. Paudel, and L. Van Gool, "Covariance pooling for facial expression recognition," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 480–4807, Salt Lake City, UT, USA, 2018.

[11] L. Wanjun, X. Liang, and Q. Haicheng, "Learning performance of CNNs with different pooling models," *Journal of Image and Graphics*, vol. 21, no. 9, pp. 1178–1190, 2016.

[12] O. Morère, J. Lin, A. Veillard, L. Y. Duan, V. Chandrasekhar, and T. Poggio, "Nested invariance pooling and RBM hashing for image instance retrieval," in *The 2017 ACM on International Conference on Multimedia Retrieval*, pp. 260–268, Bucharest, Romania, 2016.

[13] Y. Lou, Y. Bai, J. Lin et al., "Compact deep invariable descriptors for video retrieval," in *Data Compression Conference*, p. 2375, Snowbird, USA, 2017.

[14] C. Zhang, G. Evangelopoulos, S. C. Voinea, L. Rosasco, and T. Poggio, "A deep representation for invariance and music classification," in *Proceedings Icassp IEEE International Conference on Acoustics Speech & Signal Processing*, pp. 6984–6988, Florence, Italy, 2014.

[15] O. Rippel, J. Snoek, and R. P. Adams, "Spectral representations for CNNs, International Conference on Neural Information Processing Systems," 2015, https://arxiv.org/abs/1506.03767.

[16] Z. Hao and J. Ma, "Hartley spectral pooling for deep learning," *CSIAM Transactions on Applied Mathematics*, vol. 1, no. 3, pp. 518–529, 2020.

[17] Y. Sun, C. Sun, D. Wang, Y. He, and H. Lu, "ROI pooled correlation filters for visual tracking," in *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5776–5784, Long Beach, CA, USA, 2019.

[18] Y. Zhu, T. Xu, L. Peng et al., "Faster-RCNN based intelligent detection and localization of dental caries," *Displays*, vol. 74, 2022.

[19] F. E. Raymond and B. W. Raymond, "Voluntarily relinquishing private property rights: the existence of risk-pooling wquilibria when facing environmental uncertainty," *Theoretical Economics Letters*, vol. 5, no. 2, pp. 225–237, 2015.

[20] Y. Kaiyu, F. Xu, and J. Yu, "Shallow and wide fractional max-pooling network for image classification," *Neural Computing and Applications*, vol. 31, no. 2, pp. 409–419, 2019.

[21] Q. Hou, L. Zhang, M. M. Cheng, and J. Feng, "Strip pooling: rethinking spatial pooling for scene parsing," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4002–4011, Seattle, WA, USA, 2020.

[22] J. Lin, L. Ma, and C. Jingxia, "A frequency-domain CNN architecture based on the frequency-domain randomized offset rectified linear init and frequency-domain chunk max pooling method," *IEEE Access*, vol. 8, pp. 98126–98155, 2020.

[23] Y. Gong, L. Wang, R. Guo, and S. Lazebnik, "Multi-scale orderless pooling of deep convolutional activation features," in *The European Conference on Computer Vision (ECCV)*, vol. 8695, pp. 392–407, Zurich, Switzerlan, 2014.

[24] V. Christlein, L. Spranger, M. Seuret, A. Nicolaou, P. Král, and A. Maier, "Deep generalized max pooling," in *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pp. 1090–1096, Sydney, Australia, 2019.

[25] J. Zetao, Q. Jiaqi, and Z. Shaoqin, "Parameterized pooling convolution neural network for image classification," *Acta Electronica Sinica*, vol. 48, no. 9, pp. 0372–2112, 2020.

[26] P. Sermanet, S. Chintala, and Y. Lecun, "CNNs applied to house numbers digit classification," in *International Conference on Pattern Recoginition*, pp. 3288–3291, Tsukuba Science City, Japan, 2012.

[27] J. Ma and G. Xiaodong, "Scene image retrieval with siamese spatial attention pooling," *Neurocomputing*, vol. 412, no. 2, pp. 252–261, 2020.

[28] S. Zenglin, Y. Yangdong, and Y. Wu, "Rank-based pooling for deep convolutional neural networks," *Neural Networks*, vol. 83, pp. 21–31, 2016.

[29] S. H. Wang, V. Govindaraj, J. M. Gorriz, X. Zhang, and Y. D. Zhang, "Explainable diagnosis of secondary pulmonary tuberculosis by graph rank-based average pooling neural network," *Ambient Intelligence and Humanized Computing*, 2021.

[30] Y. D. Zhang, S. C. Satapathy, D. Wu, D. S. Guttery, J. M. Górriz, and S. H. Wang, "Improving ductal carcinoma in situ classification by convolutional neural network with exponential linear unit and rank-based weighted pooling," *Complex & Intelligent Systems*, vol. 7, pp. 1295–1310, 2021.

[31] Y. D. Zhang, C. Pan, X. Chen, and F. Wang, "Abnormal breast identification by nine-layer convolutional neural network with parametric rectified linear unit and rank-based stochastic pooling," *Journal of Computational Science*, vol. 27, pp. 57–68, 2018.

[32] Y. D. Zhang, X. Jiang, and S. H. Wang, "Fingerspelling recognition by 12-layer CNN with stochastic pooling," *Mobile Networks and Applications*, 2022.

[33] Z. Ke, C. Le, and Y. Yao, "A multivariate grey incidence model for different scale data based on spatial pyramid pooling," *Journal of Systems Engineering and Electronics*, vol. 31, no. 4, pp. 770–779, 2020.

[34] J. Chang, L. Zhang, N. Gu et al., "A mix-pooling CNN architecture with FCRF for brain tumor segmentation," *Journal of Visual Communication and Image Representation*, vol. 58, pp. 316–322, 2019.

[35] B. A. Skourt, A. E. Hassani, and A. Majda, "Mixed pooling dropout for CNN regularization," *Journal of King Saud University Computer and Information Sciences*, 2021.

[36] Y. Wang, S. Xiaoning, and X. Wu, "Two-stream face spoofing detection network combined with hybrid pooling," *Journal of Image and Graphics*, vol. 25, no. 7, pp. 1408–1420, 2020.

[37] B. Qi, Z. Han, and Q. Kun, "Multi-scale stacking attention pooling for remote sensing scene classification," *Neurocomputing*, vol. 436, no. 12, pp. 147–161, 2021.