Available online at www.ijpe-online.com

International Journal of Performability Engineering

vol. 14, no. 3, March 2018, pp. 434-444
DOI: 10.23940/ijpe.18.03.p4.434444

CNN-based Flow Field Feature Visualization Method

Tang Bin[a,*] and Li Yi[b]

[a]*College of Shipbuilding Engineering, Harbin Engineering University, Harbin, 150001, China*
[b]*College of Computer Science & Technology, Harbin Engineering University, Harbin, 150001, China*

Abstract

The feature-based visualization method can separate important areas for users from flow field data, which can better highlight the feature structure. However, most of the current feature extraction methods are only applicable to single typical features, and they need complex mathematical analysis. Based on the above reasons, this paper proposes a universal feature visualization method, recognizes demand in the region of flow data, shows the characteristics of structure protruding from the global visual effect in the design of a multi-dimension parallel convolution kernel that contains the recognition model, and further puts forward the method of feature visualization based on a convolutional neural network. Compared with the classical three level BP neural network model, our model gets a high accuracy rate. We verify the effectiveness of the method and solve the problem of insufficient expansion of existing methods.

Keywords: flow field; CNN; feature-based visualization

(Submitted on December 6, 2017; Revised on January 24, 2018; Accepted on February 15, 2018)

1. Introduction

Researches have always established a similar physical experiment environment to analyze the fluid data of ships or aircrafts, which, as a kind of physical experiment, is unlikely to provide various fine quantitative data indicators. In order to meet the urgent demand of studying flow field features, the changing laws of using computational fluid mechanics, and the progress of Internet technology, scientific research projects such as Numerical Wind Tunnel and Numerical Tank were born, which can provide low-loss and high-precision flow field information test results through virtual flow field experiments and simplify the process of related data analysis by scholars [14]. In the data post-processing stage, the scientific visualization method can also work as a form of the flow field data, such as simulation test results that can directly display the flow field features with image and graphics, making it convenient for relevant research workers to study the law. In derivative areas of flow field visualization method, a key research topic is to locate flow field areas efficiently and highlight their structures.

As the neutral network has nonlinear mapping characteristics, it can not only recognize the typical characteristics of flow field in the flow field visualization area, but also recognize and classify the atypical characteristics that are needed by researchers. However, it can only be selected through complex reanalysis and methods after sample training and studying [3]. This happens to make up for the disadvantages of stronger monotony and insufficient universality of the current mainstream feature extraction methods. At the same time, compared with traditional methods that extract physical or topological features based on artificial settings, the neural network carries out less complex and uncomplicated mathematical analysis and processing. It can directly input the original data as a sample and conduct adaptive training and verification. Moreover, CNN can effectively resist recognition difficulties such as deformation, translation, scaling and weak rotation by virtue of its excellent performance of recognition. It is superior to the ordinary BP neural network in terms of actual recognition effects and accuracy.

Therefore, the paper tries to explore a more universal feature visualization solution by employing CNN to recognize and locate features and seek a more reasonable mapping solution. We highlight features and remove or weaken non-relevant parts in the subsequent visualization process, thus achieving flow field feature visualization and making it convenient for researchers to study the law of features.

* Corresponding author.
E-mail address: tangbin@hrbeu.edu.cn

2. Related work

The feature-based visualization method mines important areas of special significance to users from the data, reduces and weakens the visual impact generated by redundant data, and draws visual images by means of direct, geometric and texture methods to simply and efficiently highlight key features and alleviate the visualization problems caused by chaos or shadowing. Therefore, the flow field extraction method, as the most important link in feature visualization, has always been a hot area of research by experts and scholars at home and abroad.

Since the topological structure analysis method was proposed, the structure of multiple flow field features generated through classification of critical points has greatly increased the information amount in the flow field data. Afterwards, the methods of high order singular point and closed streamline were also put forward successively. Despite that, the vortex feature has not yet been strictly defined in the later research. Working out its extraction algorithm has gradually become a popular research direction of researchers due to its importance in the fluid structure. There are some typical methods based on identical physical mechanism and velocity gradient tensor, including Q, λ_2, Δ, and Γ_2, etc. In addition, Ebling et al. proposed a template matching feature detection method based on the Clifford algebra in Literature [1], which offers a detection method that replaces attributive analysis with known structures. In China, Xu L. et al. combined Shannon's information theory with the flow field data for measurement and put forward the concept of flow information entropy in Literature [12]. According to this method, areas with larger entropy values are used to reflect flow field features. Liu X et al. proposed the feature template matching method based on the flow field information entropy and the Clifford algebra [6]. According to the method, regions with larger information entropy values are first targeted, and then templates with user-defined features are used to recognize features based on the Clifford algebra. Xu was the first to propose and realize the detection and recognition of flow field features by use of one hidden layer BP neural network. According to the method, a BP network model that can recognize corresponding features after training is obtained through supervising and studying the interest feature samples marked by researchers, which is applied to feature recognition in visualization. Compared with template matching, this method can avoid the repeat of error recognized in the same area through a second training that marks the area where wrong recognition occurs.

As there have been no strict and mathematical definitions for many important features in the flow field to this day, different flow field feature extraction standards were proposed successively. However, most of these recognition methods are only specific to one particular feature structure and lack universal extraction and visualization means. Consequently, the combination with neural network recognition and the classification method offers an opportunity to address the problem.

3. CNN-based flow field feature visualization method

Two layers of definition are applicable to the flow field features: special structures of great significance in the flow field, such as vortex and shock wave, and special areas in data that interest researchers [10]. The latter may include typical features, undefined new features, and areas of research value in a particular environment. Most of the current feature extraction methods are mainly targeted at simple typical features and lack universality for complex new features or atypical features in a particular environment, which are needed by researchers. CNN is a supervised learning algorithm with excellent nonlinear classification capability, which recognizes flow field features by constructing a CNN model and selecting feature sample data sets to train the CNN model. Theoretically, effective recognition can be achieved for typical features or user-interested features in the flow field. Based on the above analysis, this paper attempts to propose a CNN-based flow field feature recognition method, according to which an applicable CNN model is designed and adopted to recognize features. Based on such methods, the CNN-based feature visualization method is finally proposed.

3.1. CNN feature recognition model

3.1.1. Network structure with multi-dimension convolution kernels

Compared with the common neural network that has a fully connected structure, the CNN model structure that contains convolution excels at recognizing features that are close to each other in space and is thus widely applied in the field of image recognition. Figure 1 shows the CNN model structure containing multi-dimension parallel convolution kernels that is finally determined through an experiment after referring to the classical CNN models of LeNet-5 and GoogLeNet. Among them, C1, C2, C3 are convolutional layers, S4 and S5 are sub sampling layers, and F6 is a fully connected layer. The final output is classified by Softmax.

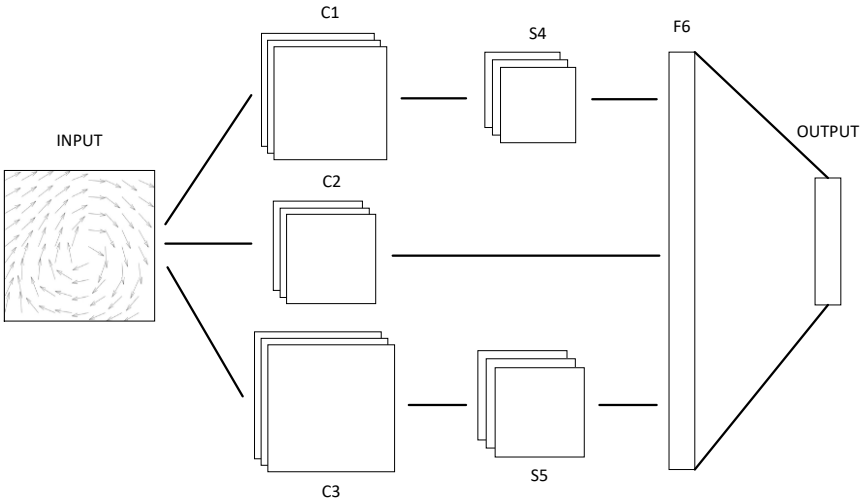


Figure 1. CNN feature recognition model structure

The C_1, C_2 and C_3 that are parallel in the structure use different sizes of convolution kernels to convolve the input two-channel matrix samples, respectively. Formal representations are as follows:

$$\begin{aligned} C_1 &= \sigma_1(\mathrm{cov}(W_1, x) + b_1) \\ C_2 &= \sigma_2(\mathrm{cov}(W_2, x) + b_2) \\ C_3 &= \sigma_3(\mathrm{cov}(W_3, x) + b_3) \end{aligned} \tag{1}$$

where cov represents convolution; W_1, W_2 and W_3 are corresponding convolution kernel parameters; b_1, b_2 and b_3 are corresponding bias terms; σ_1, σ_2 and σ_3 are corresponding activation functions. The convolution results of C_1 and C_3 shall be pooled according to S_4 and S_5, respectively. $\mathrm{max\_pool}$ represents the maximum pooling operation. The calculation of this method is as follows:

$$\begin{aligned} S_4 &= \max\_\mathrm{pool}(C_1) \\ S_5 &= \max\_\mathrm{pool}(C_3) \end{aligned} \tag{2}$$

Through the above calculation, the outputs of C_2, S_4 and S_5 are integrated and passed to the fully connected layer of F_6. The formal representation is as follows:

$$F_6 = \sigma_6(W_{F_6} \bullet \mathrm{concat}(C_2, S_4, S_5) + b_{F_6}) \tag{3}$$

where σ_6 refers to the activation function of the present layer, and concat represents integrated processing. The output of F_6 is finally passed to the output layer of the model for actual classification.

In the design of the neural network structure, the choice of activation functions of convolutional layers and fully connected layers are very important. The current mainstream activation functions are $\mathrm{sigmoid}$, $\tanh$ and ReLU [11]. As ReLU is more like the activation model of biological neurology, it takes less time for its gradient descent than that of the other two functions, which accelerates its convergence. Therefore, ReLU is applied the most widely. In this paper, the ReLU is also adopted as the activation function of convolutional layers and fully connected layers $\sigma_1, \sigma_2, \sigma_3$ and σ_6.

3.1.2. Layer parameter setting

The sample size in this paper is 9×9, a total of 81 nodes. But, since each node contains two attributes of μ and ν, it is actually a two-channel input and can be expressed as $9 \times 9 \times 2$. In the network structure designed in this paper, C_1, C_2 and C_3 receive sample inputs simultaneously, which are convolved with different sizes of convolution kernels respectively.

For the convolutional layer C_1, the vertical and horizontal shift step lengths of the convolution filter kernel are both 1, the size 4×4 and corresponding depth 2. It can thus be expressed as $4\times4\times2$. 16 types of convolution kernels are adopted at this layer. The narrow convolution VALID is used on the edge, i.e., there is no use of the edge zero-padding method [15]. The calculation algorithm is Formula (3-4), where H is the size after convolution, W the size before convolution, F the size of convolution kernel size, and S the shift step length. As a result, the size of the output feature matrix is 6×6, the quantity 16, and it can thus be expressed as $16\times[6\times6]$.

$$H=\lceil(W-F+1)/S\rceil \tag{4}$$

For the sub sampling layer S_4, the output of the C_1 layer matrix is 16 × [6 × 6]. The vertical and horizontal shift step lengths of the sampling are both 2, the size 2 × 2, and the quantity 16. The maximum sampling method is adopted, with the SAME method used on the edge. The size of output sampling matrix is 3 × 3, which can be expressed as 16 × [3 × 3].

For the convolutional layer C_2, no sub sampling layer follows subsequently. Instead, the method of increasing step length is adopted [9]. The vertical and horizontal shift step lengths of the convolution filter kernel are both 2, the size 3×3 and corresponding depth 2. It can thus be expressed as $3\times3\times2$. 16 types of convolution kernels are adopted for this layer. The narrow convolution VALID is used on the edge. As a result, the size of the output feature matrix is $\lceil(9-3+1)/2\rceil\times\lceil(9-3+1)/2\rceil$, i.e., 4×4 or quantity 16, and can thus be expressed as $16\times[4\times4]$. Despite that the size is aliquant when calculating, it can be known through simple calculation that there is no edge missing in the actual movement.

For the convolutional layer C_3, the vertical and horizontal shift step lengths of the convolution filter kernel are both 1, the size 2×2 and corresponding depth 2. It can thus be expressed as $2\times2\times2$. 16 types of convolution kernels are adopted for this layer. The narrow convolution VALID is used on the edge. As a result, the size of the output feature matrix is 8×8, the quantity 16, and it can thus be expressed as $16\times[8\times8]$.

For the sub sampling layer S_5, the output of the C_3 layer matrix is 16 × [8 × 8]. The vertical and horizontal shift step lengths of this layer are the same with that of S_4, and the quantity is the same with that of C_3, which is 16. The maximum sampling method is adopted with the SAME method used on the edge. The size of the output sampling matrix is 4×4, which can be expressed as $16\times[4\times4]$.

The fully connected layer F6 can be seen as a special convolutional layer, which converts multi-dimensional data to one-dimensional data. This layer integrates the outputs of C_2, S_4, and S_5 as inputs. The number of neurons in this layer shall be first calculated with an empirical formula and then adjusted and optimized. There are many empirical formulas for calculation of the quantity of neurons in the fully connected layer, most of which are related to the number of inputs and outputs of the layer. The empirical formula adopted in this paper is as follows:

$$L=\frac{3\sqrt{M\times N}}{2} \tag{5}$$

In the Formula (5), M and N represent the number of input nodes and output nodes for F6, respectively. The final L value calculated with the empirical formula is usually fluctuated based on the actual situation. In this paper, $M=16\times4\times4+16\times4\times4+16\times3\times3$. There are four categories of feature samples, i.e., clockwise vortex samples, anticlockwise vortex samples, saddle-shaped samples and normal samples, which are encoded (1,0,0,0), (0,1,0,0), (0,0,1,0) and (0,0,0,1), respectively in this paper, hence N=4. By formula (5), L can be finally adjusted to 80. For the flow field features of other data set, the L value may be optimized and modified by Formula (5) and the actual number of feature categories, while other parts of the CNN model need no further major modifications.

In the end, the output of the layer F6 shall be used to calculate the normalized data of the multi-classification results with Softmax for actual classification.

3.1.3. Recognition effects

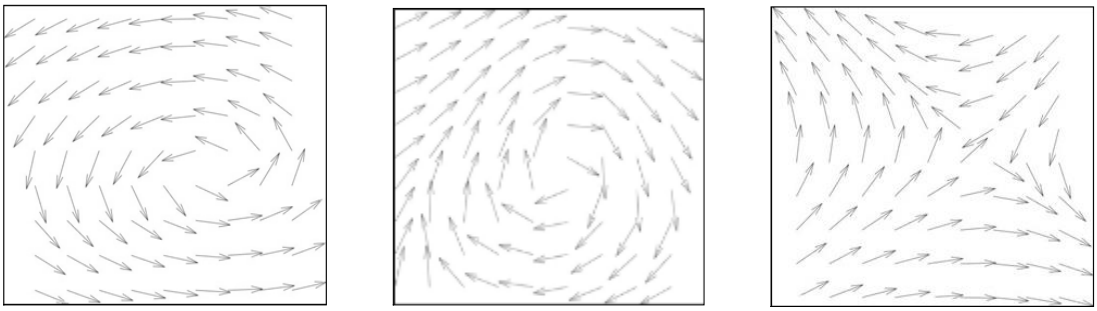
Under the premise of ensuring the diversity of data features, a set of two-dimensional vector feature sample data is labelled in this paper for the training and verification of the validity of the CNN flow field feature recognition method. The feature

categories selected for this sample data set include clockwise vortex, anticlockwise vortex and saddle-shaped areas. Along with the normal samples that are used as background data, there are four categories in total. Among them, clockwise vortex and anticlockwise vortex can be regarded as one same kind of vortex feature structure, which are of great significance to researchers under special circumstances, such as cyclones and anticyclones. Normal samples, however, refer to diversified background data other than those three selected feature samples. In practical application, researchers can choose a particular kind of data that interest them according to their needs for sample classification and annotation and further recognize such kinds of features.

The size of a single sample matrix used in this paper is 9×9, among which each node contains two attributes, i.e., μ and v. During the model training, μ_0 and v_0 that have gone through nodal velocity unitization are used as two-channel data inputs. The unitization method is shown as Formula (6).

$$\begin{aligned} u_0 &= u / sqrt(u^2 + v^2) \\ v_0 &= v / sqrt(u^2 + v^2) \end{aligned} \tag{6}$$

There are a total of 513 clockwise vortex samples, 503 anticlockwise vortex samples and 525 saddle-shaped samples labelled and selected in this paper. The rest are normal background samples. For each category, 150 samples are selected randomly as test sets, with a total number of 600. The rest are allocated as training set samples.



(a) (b) (c)

Figure 2. Visualization effects of feature sample point icons

The visualization effects of the unit point icons of feature samples are shown as Figure 2, where (a), (b) and (c) represent clockwise vortex samples, anticlockwise vortex samples and saddle-shaped samples, respectively. The single hidden layer BP neural network model and the CNN model designed in this paper are used to train, test and recognize the data set below.

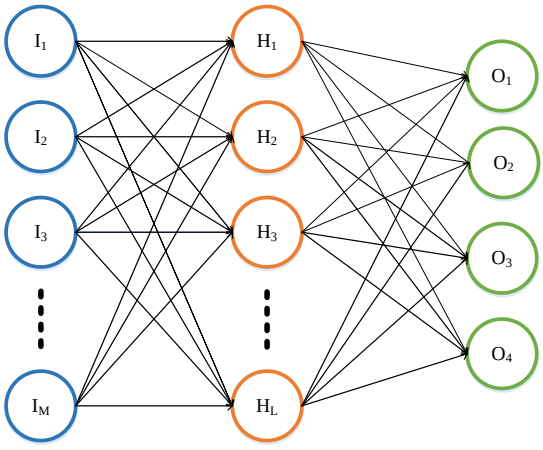


Figure 3. Structure of three-layer BP neural network

The BP neural network is a feedforward neural network model that has a fully connected structure, which is widely applied in the combination of neural network and the interdisciplinary field [2]. In this paper, the parameters of flow field feature data set are adjusted based on the typical three-layer BP neural network. As shown in Figure 3, the structure contains

two fully connected layers, which can also be divided into three layers, i.e., the input layer I, the hidden layer H, and the output layer O. For the input layer I, the received two-channel sample input is $9\times9\times2$ and the actual number of neuron nodes represented by M is 162; for the hidden layer H, the number of neuron nodes L is 25; for the output layer O, the number of nodes equals the actual number of categories, which is 4. The actual final recognition results are shown in Table 1 and Table 2.

Table 1. BP network sample recognition results

BP recognition	Actual classification of prediction results				
	Normal	Saddle-shaped	Clockwise vortex	Anticlockwise vortex	Total
Normal	137	8	4	10	159
Saddle-shaped	4	142	5	0	151
Clockwise vortex	4	0	141	0	145
Anticlockwise vortex	5	0	0	140	145
Total	150	150	150	150	600

Table 2. CNN sample recognition results

CNN recognition	Actual classification of prediction results				
	Normal	Saddle-shaped	Clockwise vortex	Anticlockwise vortex	Total
Normal	141	4	1	2	148
Saddle-shaped	2	146	4	0	152
Clockwise vortex	3	0	145	0	148
Anticlockwise vortex	4	0	0	148	152
Total	150	150	150	150	600

Table 3 shows the comparison of accuracy calculated according to the recognition results in Table 1 and Table 2, where the total accuracy relates to the number of correctly recognized normal samples while the feature accuracy is related to the number of correctly recognized feature samples only.

Table 3. Comparison of recognition accuracy

Accuracy	Total number of samples	Number of feature samples	Number of correct feature samples	Number of correct normal samples	Total accuracy	Accuracy of feature samples
BP recognition	600	450	423	137	93.33%	94.00%
CNN recognition	600	450	439	141	96.67%	97.56%

In terms of this test set, the overall accuracy of CNN recognition can reach 96.67%, while the accuracy for 450 feature sample is 97.56%. Both indicators are higher than the final results of BP recognition, and the overall recognition effects are also superior. Figure 4, Figure 5, and Figure 6 show the precision ratio, recall ratio and the harmonic mean F1 of those three feature samples calculated according to Table 1 and Table 2. For the precision ratio, the accuracy of the CNN model is always higher than 96%, which is superior to the BP network model in recognition effects all the time. For recall ratio, CNN BP network is far ahead of BP network in terms of recognition effects, with the former effectively reducing omission of features. As can be seen from the comparison of the harmonic mean F1, CNN is better than BP network in respect of the overall precision and recall for the data set in this paper.

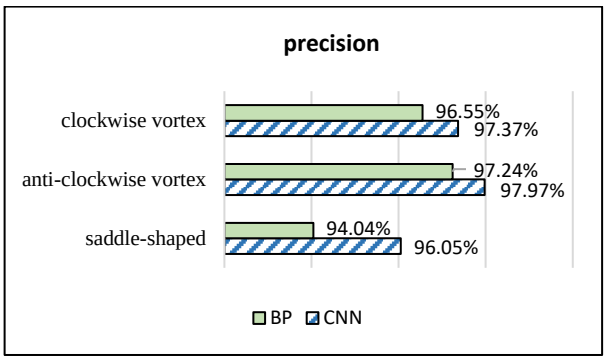


Figure 4. Precision ratio chart

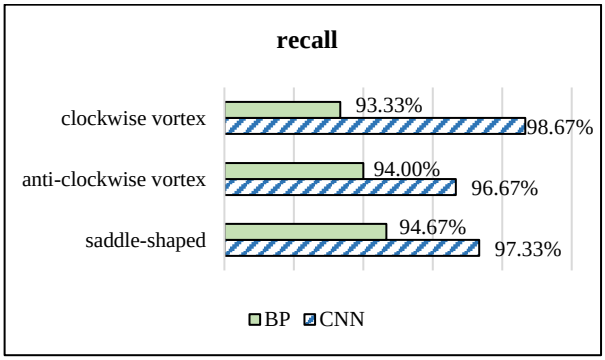


Figure 5. Recall ratio chart

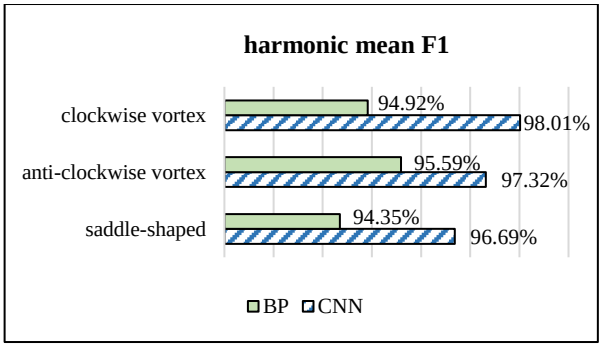


Figure 6. Harmonic mean F1 chart

3.2. CNN feature recognition visualization

In the flow field, the feature-based visualization method shifts the focus from the whole to the part through visually highlighting the feature parts in the data while filtering and weakening non-relevant data to distinguish them, which can avoid the independent feature structure recognition that is relied upon researchers after data visualization. Figure 7 is a framework diagram of the CNN feature visualization method designed in this paper. The method is mainly divided into three stages, i.e., data pre-processing, feature area detection and feature visualization, among which the feature detection stage shall cover the sample recognition process by use of the aforementioned CNN feature recognition method.

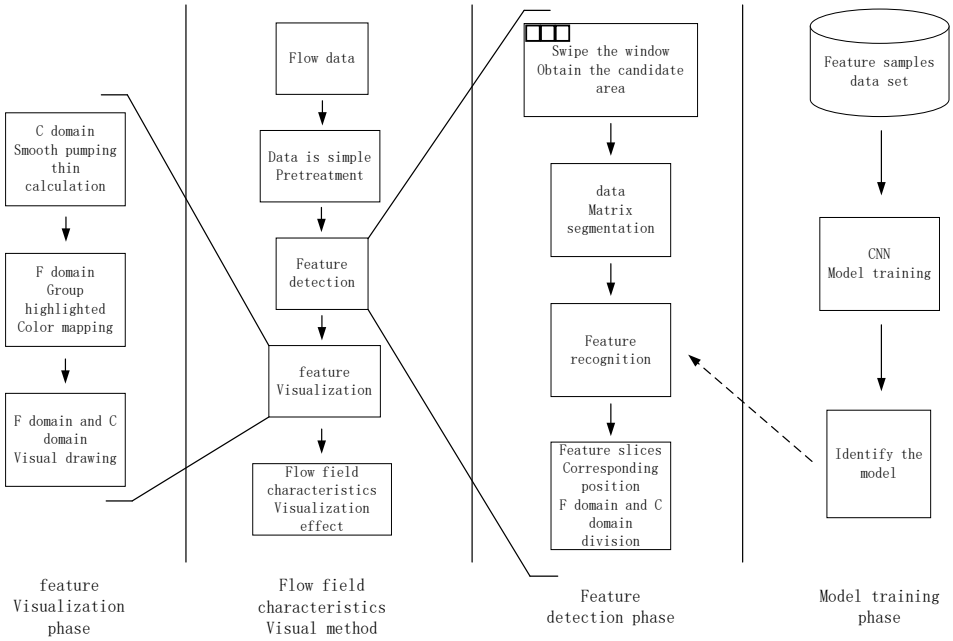


Figure 7. Framework diagram of CNN feature visualization method

In the data pre-processing stage, the main task is to filter and extract relevant data from the flow field data that needs to be visualized. In the feature detection stage, the main tasks include sliding and cutting the data field matrix that has gone pre-processing, testing and recognizing samples with the CNN model that is designed and trained in the above text, and exploring feature areas that correspond with the set feature categories and recording them according to their categories. In the feature visualization stage, the feature areas obtained through the feature detection stage shall be divided into Region C and Region F according to the F+C technical idea. Researchers then highlight the feature areas and weaken non-relevant parts through respective calculating and visualization, and thus complete the feature visualization.

3.2.1. Data pre-processing

This paper mainly studies the visualization of features from the vector data in two-dimensional flow field made of uniform grid. When the point icon method is adopted for vector data visualization, in order to achieve the best visual effects, unitized equal-length vector arrows are often used to map vector nodes, and the node size attribute is labelled by changing the color of node arrows. As the color of arrows are used to distinguish feature area categories according to the final visualization method in this paper, the grids where the nodes are located are thus used for color mapping as the underlying background of visualization instead of arrows. Based on the above analysis, in terms of the flow field data that is about to be visualized, the major experimental data related attributes that should be paid attention to in the data pre-processing stage shall be horizontal and vertical coordinates and their corresponding velocity components. In addition, the vector data shall be unitized according to Formula (6), and the size of vectors before unitization shall be recorded for color mapping.

3.2.2. Pre-selection of feature areas

In order to recognize the features in the data field, the first step is to segment the vector matrix in the data field by sliding the window matrix. Slide and segment the matrix in the horizontal and vertical directions. Slide back and forth to make sure boundaries are reached. Derive and recognize the sample matrixes in the sample data set that have the same size. They are then used as the input to pass through the well-trained CNN model for recognition and classification. Judge their feature and category to in order to seek the feature area that meets the requirement. This method is quite common and can also be applied to the recognition of typical features and atypical features that are needed by researchers. In the actual process of feature visualization, when the final demand features conform to particular requirements, the pre-selection method may be used to quickly filter the matrix data and conduct feature recognition in the filtered candidate areas. Among the three categories of feature samples in this paper, the clockwise vortex and anticlockwise vortex can be regarded as one same kind of vortex feature, which is identical to the saddle-shaped feature in their feature areas. Both contains the feature critical point, and their vector attributes are all 0 in all directions according to the vector topology analysis theory. Therefore, the pre-selection of candidate areas can be realized through a simple search for the feature points in the vector matrix.

An immediate existence of the ideal feature point is hardly impossible, which is generally located within the data grid. For a grid cell, when the vector components of the four vertices of the cell are converted in the horizontal and vertical directions positively and negatively each at least once, we shall determine that this cell contains a feature critical point [8]. Therefore, the matrix sample with such cell as its center can be listed as a candidate area for classification and recognition judgment. As shown in Figure 8, as the vector is converted in the horizontal and vertical direction positively and negatively in cell 1, it can be taken as a candidate area; while in cell 2, where the vector is not converted in in the horizontal direction positively and negatively, it cannot be taken as a candidate area. The list of candidate points can be obtained by scanning the whole data matrix and thereafter by recording possible locations of candidate feature points.

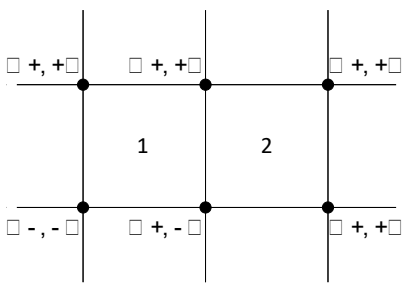


Figure 8. Sketch of feature point judgment

3.2.3. Region C and Region F division

The region division is based on the feature recognition results. It is possible to carry out recognition and classification after obtaining the list of feature candidate points. Take each node that corresponds with the location of each element in the list of

feature candidate points as center points while expanding them equidistantly to the size of feature samples. The matrix that is to be recognized will be obtained. Pass this matrix to the trained CNN feature recognition model. Based on the recognition results, record and store the area location that corresponds with the node coordinate in the list of its features. Through this step, different mapping operations for various features may be possible in the subsequent visualization process.

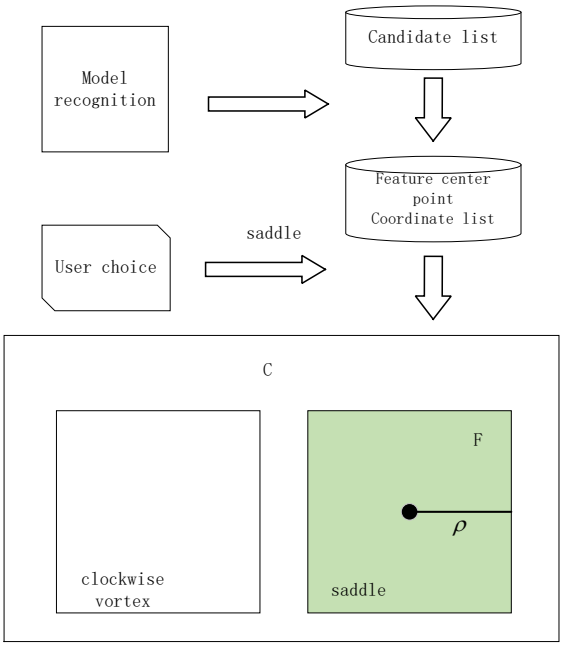


Figure 9. Sketch of F+C region division

According to the F+C technical idea, the important visual points of interest are regarded as the highlights of Region F, while other non-relevant areas are weakened in Region C. When this idea is applied to the visualization of the flow field, the tested and recognized feature area is often regarded as Region F. As the visualization methods are designed with interactivity in this paper, the actual demands of users shall be the main basis for the division of Region F and Region C. For example, there are three categories of features that interest users, but only one category can be highlighted in each actual visualization. Therefore, as shown in Figure 9, despite that the data vector matrix contains three categories and all of them have been recognized in the final actual visualization process, researchers need to weaken the Region C that contains the clockwise vortex, anticlockwise vortex and other background areas when visualizing the saddle-shaped features. At the same time, in order to ensure better visualization effects for features in Region F, the threshold of feature range is introduced when highlighting multiple features for differentiation among colors. Researchers may control the actual visualization range of Region F.

3.2.4. Smoothing and thinning calculation

According to the requirements of the fisheye view method of the F+C technology, Region C and Region F shall be processed differently subsequent to the region division in order to highlight Region F and the set feature structure in the vector matrix. Based on the regular fisheye view method, the data shall often be skewed to increase the level sense of visual points. As the real flow field data have some physical properties and are not suitable for distortion processing, the data in Region C are thinned in this paper in order to change local intensity and visual level of vector matrix. The thinning method used in this paper is based on the equal step length thinning method. It is a smoothing and thinning method that calculates the overall vector direction of the area that is represented by a vector node, which is defined as the trend point.

The data vector matrix of the size $m \times n$ is M, and the thinned matrix is M'. Each element in these two two-dimensional vector matrixes is represented by the horizontal and vertical vector component of (μ, v). In order to smooth and thin the matrix M with a thinning intensity of $\delta = k$, the first step is to divide M in the horizontal and vertical direction with a step length of *k* into *t* sub-matrixes, among which the value of *t* is $\lfloor m/k \rfloor \times \lfloor n/k \rfloor$. Then, mark the arbitrary submatrix of M as N. Each N is smoothed and thinned and converted for M' times. The thinning results are hence obtained. When dividing M into *t* sub-matrixes, the marginal part that is not fully divided as N shall be used to construct the matrix N by use of zeros padding method with the actual residual size. The trend point in the final N shall be converted for M' times. If

the coordinates exceed the size during conversion, this submatrix shall be abandoned. According to the smoothing and thinning method, the element in N is brought into Formula (7) to work out the means of horizontal and vertical vector components, i.e., μ' and v', respectively. If N is a regular submatrix, the value of p is k^2; if N is an edge completion submatrix, the value of p shall be the number of real surplus nodes. The obtained μ' and v' shall be unitized according to Formula (6) to derive the trend point, whose location is $(\lceil k/2 \rceil, \lceil k/2 \rceil)$ in N. At last, finish the thinning operation by converting this trend point to M'.

$$u' = \frac{\sum_{i=1}^{k}\sum_{j=1}^{k} u_{i,j}}{p}$$
$$v' = \frac{\sum_{i=1}^{k}\sum_{j=1}^{k} v_{i,j}}{p} \tag{7}$$

3.2.5. Data synthesis mapping

After region division and smoothing and thinning, the overall data shall be synthesized to derive the final data matrix used for mapping, and the colors shall be mapped according to data properties to achieve visualization.

For data synthesis, the first step is to smooth and thin the entire vector matrix to derive the thinned matrix M'. Then, take the elements in the list of feature center in the formerly divided Region F as central nodes while expanding them to the data node matrix set by the threshold of the feature range, with the actual size of $(2\rho+1)\times(2\rho+1)$. Redefine the color attributes of all nodes in the feature matrix with the color attribute type of this central node. Then, cover M' with the feature matrix according to the corresponding locations of nodes.

Figure 10 is the data synthesis sketch of Region F and Region C, where each grid part represents a vector node. The yellow nodes are thinned, while the red areas are directly covered. For the final mapping, we first draw a gradual background according to the vector size in the matrix, which then draws arrow icons with different colors and sparsity. In the end, the visualization effect with different levels of structure can be achieved.

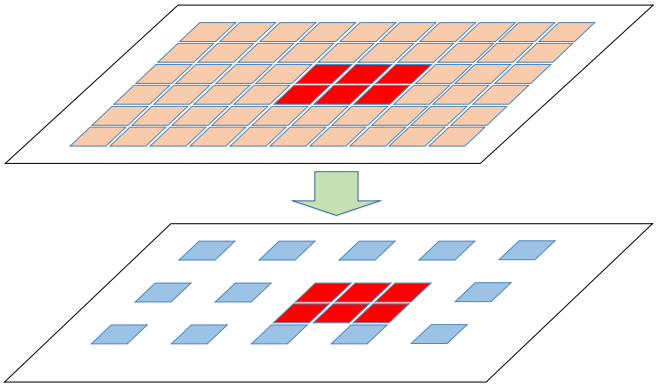


Figure 10. Sketch of vector data mapping

4. Application

Based on the above feature visualization process, Figure 11 shows the final effect of the CNN feature visualization method proposed in this paper, with the thinning intensity of $\delta = 2$ and feature range of $\rho = 5$. For the left picture, the user chooses to display the full set of features, where the bright red parts represent the clockwise vortex features, the bright yellow parts the anticlockwise vortex features, the bright white parts saddle-shaped features. For the right picture, the user chooses to display the clockwise vortex features only. In the above synthetic mapping process, the direction of vector arrows in the chosen Region F always corresponds with the actual direction of the original data when highlighting Region F, which can thus correctly reflect the real attributes and structure of the feature areas in the original data. The flow field attributes of the divided Region C can still show the overall moving direction after the visual effect is weakened through smoothing and thinning, making it convenient for researchers to make reference analysis.

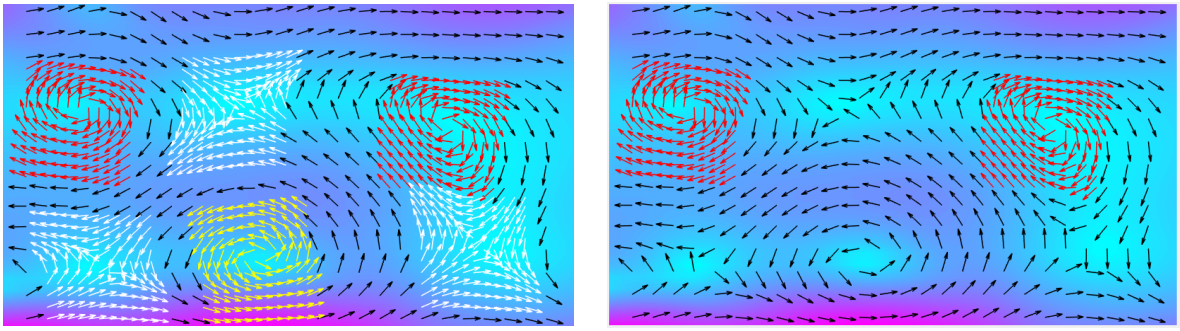


Figure 11. Data synthetic mapping

5. Conclusions

In this paper, we propose a CNN-based flow field feature visualization method, design and provide the CNN model structure that contains multi-dimension parallel convolution kernels, and compare it with the classical BP neural network through program implementation and experimental verification. Through analysis of experimental results, the CNN feature recognition method designed and implemented in this paper has a higher recognition accuracy. For the subsequently designed feature visualization method, we adopt arrow icons as mapping means, divide the feature areas according to the recognition and classification results and the categories that are chosen by users, and change the density and mapping means of the whole area through highlighting of feature areas and smoothing and thinning of non-relevant areas, making it easier for researchers to distinguish feature areas and achieve good visual effects.

References

1. J. Ebling, G. Scheuermann, "Clifford Fourier transform on vector fields," *IEEE Transactions on Visualization & Computer Graphics*, 2005, 11(4):469-479P.
2. S. C. Huang, X. Y. Fang, J. Zhou, "Image Local Fuzzy Measurement Based on BP Neural Network," *Journal of Image and Graphics*, 2015, 20(1):20-28P.
3. S. K. Li, X. Cai, W. K Wang, P. Wang and H. H. Wang, "Large-Scale Flow Field Scientific Visualization," National Defense Industry Press, 2013:158-162P.
4. Y. Lin, C. Wang, J. X. Wang, Z. Dou, "A Novel Dynamic Spectrum Access Framework Based on Reinforcement Learning for Cognitive Radio Sensor Networks," *Sensors*, 2016, 16(10): 1-22.
5. Y. Lin, X. Zhu and Z. Zheng, "The individual identification method of wireless device based on dimensionality reduction and machine learning," *The Journal of Supercomputing*, 2017: 1-18.
6. X. Liu, W. Zhang, N. Zheng, "Flow Feature Extraction Based on Entropy and Clifford Algebra," *Image and Graphics. Springer International Publishing*, 2015:292-300P.
7. H. Obermaier, R. Peikert, "Feature-Based Visualization of Multifields," *Scientific Visualization. Springer London*, 2014:189-196.
8. P. Skraba, B. Wang and G. Chen, "2D Vector Field Simplification Based on Robustness," *Visualization Symposium. IEEE*, 2014:49-56P.
9. J. T. Springenberg, A. Dosovitskiy and T. Brox, "Striving for Simplicity: The All Convolutional Net," *Eprint Arxiv*, 2014:1-14P.
10. M. S. Tang, "3D data field visualization," *Tsinghua University Press*, 1999:174-176P.
11. Y. Wu and W. G. Qiu, "Face Recognition Based on Improved Depth Convolution Neural Network," *Computer Engineering and Design*, 2017, 38(8):2246-2250 P.
12. L. Xu, T. Y. Lee and H. W. Shen, "An information-theoretic framework for flow visualization," *IEEE Transactions on Visualization & Computer Graphics*, 2010, 16(6):1216-1224P.
13. H. X. Xu, S. K. Li, L. Zeng, "A Novel Intelligent Feature Detection and Recognition Method of Fluid Fields," *Computer Engineering & Science*, 2009, 31(5):27-30.
14. F. Zhao, C. S. Wu, S. F. Huang and Z. R. Zhang, "Route map on virtual tank," *Journal of Ship Mechanics*, 2014, 18(8):924-932P.
15. Z. Y. Zheng and S. Y. Gu, "Tensorflow: Actual Google deep learning framework," *Electronic Industry Press*, 2017:144-146P.