Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 3029866, 6 pages
https://doi.org/10.1155/2022/3029866

Research Article

A Semantic Image Retrieval Method Based on Interest Selection

Wenting Hu,[1,2] **Yin Sheng**,[3] **and Xianjun Zhu**[4]

[1]*Business College, Jiangsu Open University, Nanjing 210036, China*
[2]*Business College, Nanjing University, Nanjing 210093, China*
[3]*College of Internet of Things Engineering, Hohai University, Changzhou 213022, China*
[4]*School of Software Engineering, Jinling Institute of Technology, Nanjing 211169, China*

Correspondence should be addressed to Xianjun Zhu; mymailzxj@126.com

Received 4 November 2021; Revised 8 January 2022; Accepted 22 January 2022; Published 27 February 2022

Academic Editor: Narasimhan Venkateswaran

There is a semantic gap between people's understanding of images and the underlying visual features of images, which makes it difficult for image retrieval results to meet the needs of individual interests. To overcome the semantic gap in image retrieval, this paper proposes a semantic image retrieval method based on interest selection. This method analyses the interest points of individual selections and gives the weight of the interest selection in different regions of an image. By extracting the underlying visual features of different regions, this paper constructs two feature vector methods after users' interest point weighting. The two methods are called interest weighted summation and interest weighting. Finally, this paper compares the accuracy of different image classification methods using a support vector machine classification algorithm. The experimental results show that the target classification accuracy of the classification algorithm based on interest weighted summation is higher than that of the traditional and interest weighted methods. The classification algorithm based on interest weighted summation has the highest overall effect on target object classification in the four experimental scenarios. Therefore, the interest point selection method can effectively improve the overall satisfaction of image recommendation and can be used as a novel solution to overcome the semantic gap.

1. Introduction

With the continuous evolution of artificial intelligence technologies, such as computer vision, speech semantics, and machine learning, society is entering a new era of intelligence. This will cause the transformation of image retrieval modes and reshape the process experience of information retrieval to promote the intelligent upgrading and functional reconstruction of traditional information retrieval.

Images are the most important and effective method for human beings to obtain information because images are intuitive, comprehensible, and informative. In fact, an individual's understanding of an image not only is based on the visual similarity but also requires the semantic similarity of the image. Image processing algorithms are often used to extract underlying visual features, which cannot be used to fully describe the semantic information of an image [1]. People not only understand images through their accumulated experience, knowledge, and personal preferences from daily life but also understand images through a cognitive mode of thinking from a semantic perspective. This easily leads to a semantic gap between the image semantics and the underlying visual features. Relevant studies try to mimic the human visual attention mechanism to fundamentally solve the semantic gap problem [2–4]. The purpose of filtering redundant information from a large amount of visual information is to find useful information and obtain the high-level semantics of the image. Most of the methods were designed to model visual attention and have been evaluated by their congruence with fixation data obtained from experiments with eye gaze trackers. On the one hand, progress has been made in the construction of visual computing models to simulate the human visual attention mechanism, but these models are still in the simulation stage of human viewing scenes [5–7].

On the other hand, researchers use eye-tracking technology to obtain eye movement behaviour to depict the human visual attention mechanism [8–11]. It is difficult to collect individual eye movement information in large quantities because of the high cost and weak popularity of eye trackers.

In addition to the traditional visual attention calculation model and novel eye-tracking technology, scholars have tried to use other means to reflect and measure human visual attention phenomena, such as motion trajectory models and the click behaviour of a mouse cursor. Related research shows that users' attention behaviour is related to the mouse cursor trajectory. Users' interest selections are highly correlated with visual attention, and interest selection can be used to predict the location of the fixation point more accurately than the visual calculation model [12, 13]. Therefore, this paper takes users' interest points as feedback information to study the problem of interest point weighting in image semantics. Finally, the algorithm is used to study the accuracy of the image retrieval results fused with interest selection.

2. Classification Method Based on Interest Selection

2.1. Feature Vector Based on Interest Point Weighting

2.1.1. Weight Matrix. This paper collects interest data with a click experiment to obtain interest weight given by the grid sampling object region. We suppose that $x_i \in R^n (i = 1, 2, \ldots, n)$ is a set of interest data collected from the click experiment. $\Lambda = \{C_1, C_2, \ldots, C_T\}$ represents the object area after the object of the experimental scene was sampled by the uniform grid, and $t = 1, 2, \ldots, T$. $T = |\Lambda|$ represents the number of grid sampling objects in the experimental scene. The expression of the users' interests in the grid objects of the experimental scene C_t is the sum of the weights of the points of interest:

$$P_t = \sum_{i=1}^{m} k_i A_{x_i \longrightarrow C_t}, \tag{1}$$

where m represents the number of interest points that are divided into object C_t in the experimental scene. $A_{x_i \longrightarrow C_t}$ represents the value of interest point x_i in object C_t in the experimental scene, where the value of $A_{x_i \longrightarrow C_t}$ is 1. k_i represents the weight value given to $A_{x_i \longrightarrow C_t}$ according to different interest point weights, and $0 \le k_i \le 1$.

The one-dimensional matrix of the user's interest in the experimental scene grid object is obtained and expressed as $P = [P_1, P_2, \ldots, P_T]$. In this paper, the weight value given by the experimental scene grid object is described by the interest degree of the experimental scene grid object, and the weight value ω_t corresponding to object C_t in the experimental scene can be expressed as

$$\omega_t = \frac{p_t}{\sum_{t=1}^{T} P_t}, \sum_{t=1}^{T} \omega_t = 1, \text{and } 0 \le \omega_t \le 1. \tag{2}$$

Therefore, the one-dimensional weight matrix $[\omega_1, \omega_2, \ldots, \omega_T]$ is obtained to provide weight to the eigenvector of the grid object C_t.

2.1.2. Weighted Eigenvector. In this paper, the set of image eigenvectors extracted from grid objects in the experimental scene is expressed as $Y = \{y_t | t = 1, 2, \ldots, T\}$, where y_t is the underlying visual feature of the grid object C_t. On this basis, HSV (hue, saturation, and value) and LBP (local binary pattern) are combined to describe the underlying visual features of the grid objects, and a 1×131 dimensional visual eigenvector is obtained. This paper uses two methods to express the weighted eigenvector as follows.

The formula of the IWS (interest weighted summary) method is expressed as

$$u_{(Y)} = \sum_{t=1}^{T} \omega_t y_t. \tag{3}$$

The formula for the IW (interest weighting) method is expressed as

$$u_{(Y)}' = \{\omega_t y_t\}, \quad t = 1, 2, \ldots, T. \tag{4}$$

The formulas of $u_{(Y)}$ and $u_{(Y)}'$ express the two weighted underlying visual eigenvectors and allow them to be studied with the classification algorithm to explore the impact of interest decisions on the accuracy of image classification.

2.2. Classification Algorithm. The SVM (support vector machine) is a supervised learning method that can be widely used in statistical classification and regression analysis [14–16]. The initial appearance of SVM comes from a linear classifier. Suppose there is a two-class classification problem, and the data points are set as n dimensional vectors x of categories y, where the value of y is +1 or −1. If $f(x) = \omega^T x + b$ exists, y equals +1 and −1, which are separated on both sides of $f(x)$. It can be assumed that the y value corresponding to x in $f(x) < 0$ equals −1, while the y value corresponding to x in $f(x) > 0$ is +1. In most cases, the data are not always linearly separable, so we need to consider how to solve this problem. The SVM method maps the vector to a higher dimensional space, in which a maximum interval hyperplane is established. On both sides of the hyperplane, which separates data, there are two parallel hyperplanes. The optimization goal of separating hyperplanes is to maximize the distance between the two parallel hyperplanes.

To maximize the distance, the hyperplane can be vertically projected onto the corresponding point x_0 on the hyperplane for a point x. Then, the distance between x and x_0 can be defined as

$$x = x_0 + \gamma \frac{\omega}{\|\omega\|}. \tag{5}$$

Since x_0 is a point on the hyperplane that satisfies $f(x_0) = 0$, it can be obtained by substituting x_0 into the hyperplane equation $f(x) = \omega^T x + b$:

$$\gamma = \frac{\omega^T x + b}{\|\omega\|} = \frac{f(x)}{\|\omega\|}. \tag{6}$$

Since γ is signed at this time, the absolute value is required. It needs to be multiplied by the category of y:

$$\tilde{\gamma} = \gamma y. \tag{7}$$

Therefore, the objective function required by SVM is $\max(\tilde{\gamma})$, according to the constraint of $y_i(\omega^T x_i + b) = \hat{\gamma}_i \geq \hat{\gamma}$. $\hat{\gamma} = 1$ can be fixed to find the maximum value. The objective function becomes

$$\max \frac{1}{\|\omega\|}, \text{s.t.}, y_i\left(\omega^T x_i + b\right) \geq 1, \quad i = 1, \ldots, n. \tag{8}$$

We make a slight adjustment to the above formula, and then the function becomes

$$\min \frac{1}{2}\|\omega\|^2, s.t., y_i\left(\omega^T x_i + b\right) \geq 1, \quad i = 1, \ldots, n. \tag{9}$$

For the above problem, the Lagrange multiplier is used to calculate the extremum. The Lagrange equation is

$$L(\omega, b, \lambda) = \frac{1}{2}\|\omega\|^2 - \sum_{i=1}^{n} \lambda_i\left(y_i\left(\omega^T x_i + b\right) - 1\right). \tag{10}$$

Then, after the partial derivations of ω, b, λ, we can obtain

$$\begin{aligned} \frac{\partial L}{\partial \omega} = 0 &\longrightarrow \omega = \sum_{i=1}^{n} \lambda_i y_i x_i, \\ \frac{\partial L}{\partial b} = 0 &\longrightarrow \sum_{i=1}^{n} \lambda_i y_i = 0. \end{aligned} \tag{11}$$

Next, L is substituted; then, the problem becomes

$$\begin{aligned} &\max \sum_{i=1}^{n} \lambda_i - \frac{1}{2} \sum_{i,j=1}^{n} \lambda_i \lambda_j y_i y_j x_i^T x_j, \\ &\text{s.t.}, \lambda \geq 0, i = 1, ..., n, \sum_{i=1}^{n} \lambda_i y_i = 0. \end{aligned} \tag{12}$$

Finally, $\omega = \sum_{i=1}^{n} \lambda_i y_i x_i$ is substituted into $f(x)$; then, we can obtain

$$f(x) = \left(\sum_{i=1}^{n} \lambda_i y_i x_i\right)^T x + b = \sum_{i=1}^{n} \lambda_i y_i \langle x_i, x \rangle + b. \tag{13}$$

The above formula is the application of SVM in image classification by integrating the choice of interest.

3. Point of Interest Experiment

3.1. Purpose. According to the experimental requirements, the subjects needed to click on five points in the experimental picture that they were interested in, and the clicked positions represented the points of interest selected by the subjects. Based on this, the weighted eigenvector is brought into the classification algorithm to study whether the eigenvector based on points of interest has an impact on the accuracy of image classification.

3.2. Subjects. A total of 42 subjects (30 males and 12 females), aged 21 to 25, with an average age of 23, were invited. They had never participated in similar experiments before. They are right-handed and had a strong understanding of the experiment and completed the experiment well according to the requirements of the experimenter.

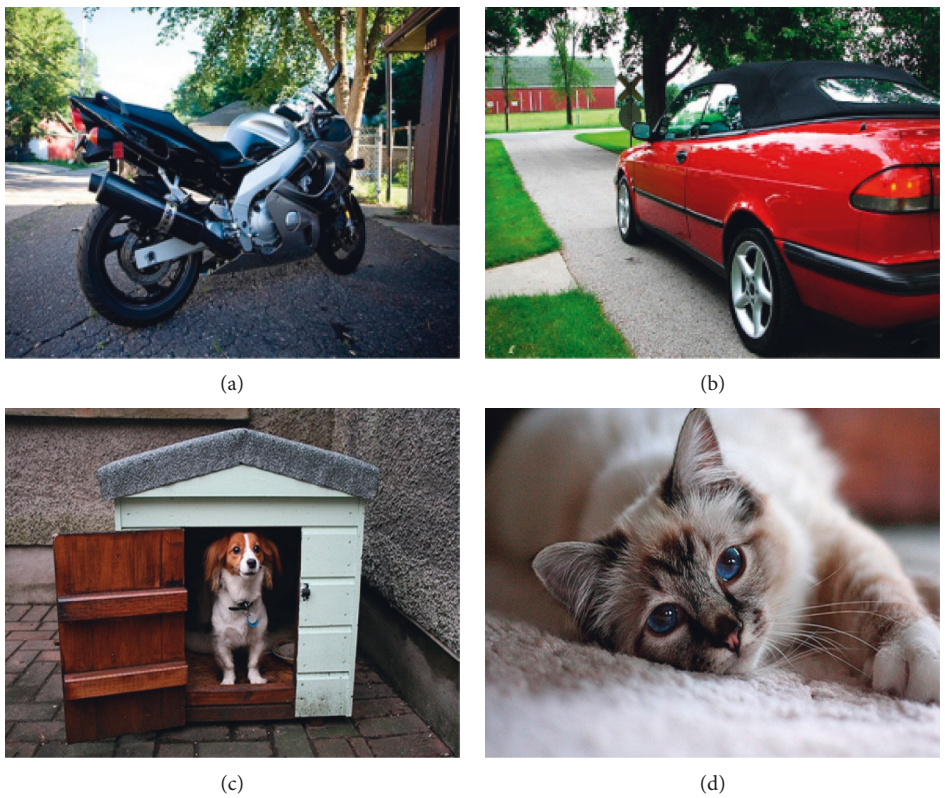
3.3. Experimental Equipment. The experiment used a desktop computer with a Dell OptiPlex 790 and a CPU frequency of 31 GHz. The capacity of the hard disk is 500 GB, and it runs on a Windows XP operating system.

3.4. Experimental Materials. In the experiment, four kinds of pictures, including kitten, puppy, motorcycle, and car pictures, in the PASCAL VOC2007 database were selected. Fifteen pictures of each kind were randomly selected to form a material library of 60 pictures, and the pictures were numbered from 01 to 60 consecutively. According to the specific requirements and specifications of the experiment, the experimental pictures were uniformly adjusted to 500×375 pixels. Figure 1 shows examples of the pictures used in the experiments.

3.5. Experimental Results and Analysis

3.5.1. Data Screening and Description. Each subject completed an experimental task with 30 pictures by selecting 5 points of interest from each picture; eventually, an average number of 105 interest points for each picture were obtained. According to the inquiry and investigation after the experiment, interest points that were incorrectly chosen by subjects were eliminated from the experiment to ensure the objectivity and accuracy of the experiment. After removing the abnormal data, the interest point coordinates of each experimental image were derived from the screen coordinates and transformed into 500×375 pixel coordinates by a mathematical transformation. Therefore, we ensured that the image coordinates were within the pixel range, in which the value of the X axis ranged from 0 to 500 and that of the Y axis ranged from 0 to 375.

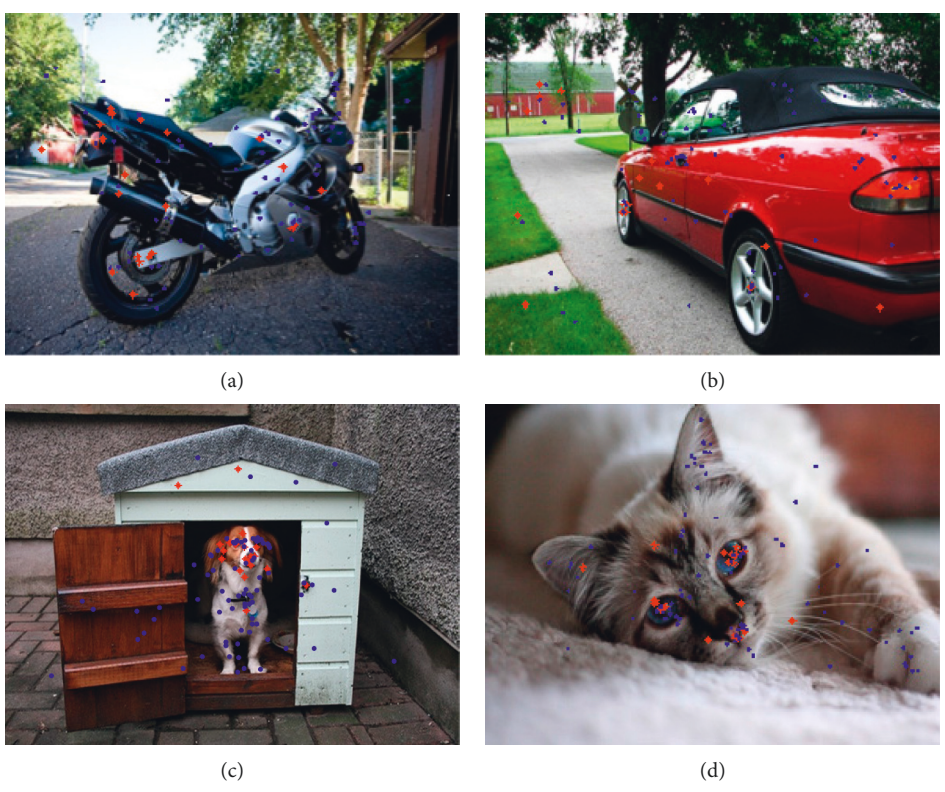
Figure 2 describes the effective distribution of the interest points of all subjects on the example pictures of motorcycles, cars, puppies, and kittens, in which the first points of interest selected by the subjects are marked with red asterisks, and the second to fifth points of interest are marked with blue dots. Figure 2 shows that the interest points are mainly distributed on the target object or foreground object rather than the picture background. The distributions of the interest points on the motorcycle and car target objects are relatively uniform, and the interest points on the dog and cat target objects are relatively concentrated. In particular, the first interest points on the puppy and cat pictures are mainly distributed on the faces.



(a) (b)

(c) (d)

Figure 1: Examples of experimental pictures. (a) Picture of a motorcycle. (b) Picture of a car. (c) Picture of a puppy. (d) Picture of a kitten.



(a) (b)

(c) (d)

Figure 2: Distributions of points of interest on experimental pictures. (a) Interest points on a motorcycle. (b) Interest points on a car. (c) Interest points on a puppy. (d) Interest points on a kitten.

Table 1: Accuracy of target object classification.

Target objects	SVM	IW-SVM	IWS-SVM
Kitten	0.80	0.74	0.85
Puppy	0.85	0.78	0.90
Motorcycle	0.70	0.72	0.84
Car	0.75	0.76	0.81
Average	0.78	0.75	0.85

3.5.2. Data Results and Analysis. First, object classification research is carried out according to the given standards for the four kinds of pictures of kittens, puppies, motorcycles, and cars. When studying the classification of a specific target, this paper takes one kind of picture as the target object and the other three kinds of pictures as interference objects. On this basis, the two classifications of sample objects are realized by combining the non-points of interest method, the IW method, and the IWS method with the SVM algorithm, and these methods are named SVM, IW-SVM, and IWS-SVM, respectively.

Second, in the case of different numbers of uniform grid object segmentations, the IW-SVM and IWS-SVM methods are used to study the average accuracy of target object classification. With the increasing number of regions obtained by uniform grid segmentation from 2×2 to 8×8, the average classification accuracies of the target objects of the two methods show a decreasing trend, indicating that the improvement in classification accuracy of the target objects is not consistent with the increase in the number of segmented regions. The study found that the average accuracy, which reached 0.8, was the highest in the case of a 3×3 segmented mesh.

Finally, according to the classification experimental results of the four kinds of target objects, this paper selects the 3×3 segmentation grid to explore the accuracy of target object classification in the four types of experimental scenes. We then randomly generate weights from a group of subject interest points. For example, the weights of the first to fifth interest points are [0.2, 0.2, 0.2, 0.2, 0.2]. This study finds that the weights of the first to the fifth interest points are [0.4, 0.3, 0.1, 0.1, 0.1], and the average accuracy of the target object classification result is the highest. It shows that the first and second interest points of the subjects better reflect the individuals' intentions and needs, indicating that the first and second interest points selected by the subjects have high research significance and value for target object classification. Table 1 describes accuracies of classification algorithms obtained by the SVM, IW-SVM, and IWS-SVM methods, which are 0.78, 0.75, and 0.85, respectively. Among them, the IWS-SVM method has the highest average accuracy. This shows that the addition of the IWS method can provide high-level semantics to the target object and improve the accuracy of target classification. In this study, we introduce the interest point selection method which can improve the overall satisfaction which can be found in Table 1.

4. Conclusion

To overcome the research disadvantages of visual attention computational models and eye-tracking technology in optimizing image recommendations, this paper proposes an image classification method based on interest selection. We focus on explaining the eigenvector of weighted interest points and complete relevant click experiments to realize the classification of experimental scene objects. The experimental results show that (1) the subjects' first and second choices of interest have a great impact on the target classification in the experimental scenes; (2) the IWS-SVM method has the best overall effect on the target object classification in the four kinds of experimental scenes; (3) the accuracy of target classification combined with the IWS classification algorithm is higher than that of the traditional methods and the IW method; and (4) the interest point method can effectively improve image information retrieval. Our results have shown that the interest point selection can be used as a novel solution to overcome the semantic gap. Therefore, future work could use other information (e.g., eye movements and electroencephalogram) to improve the overall satisfaction of image recommendation.

Data Availability

The images used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (no. 61903346), Postdoctoral Research Project in Jiangsu Province (nos. 2020Z034 and 2019K086), Natural Science Research Project of Colleges and Universities in Jiangsu Province (no. 18KJB520007), China Postdoctoral Science Foundation (no. 2020T130129ZX), and Research Project of Philosophy and Social Sciences at Jiangsu Universities (no. 2020SJA0767).

References

[1] Z. Yang, L. Yang, W. Huang, L. Sun, and J. Long, "Enhanced Deep Discrete Hashing with semantic-visual similarity for image retrieval," *Information Processing & Management*, vol. 58, no. 5, Article ID 102648, 2021.

[2] M. Q. Zeng, W. Y. Ma, and L. Li, "Efficient image retrieval scheme based on mixed similarity," *Computer Engineering*, vol. 45, no. 11, pp. 262–268, 2019.

[3] R. Lou and L. H. Jiang, "Research on semantic gap bridging of virtualization system based on feature selection," *Application Research of Computers*, vol. 36, no. 5, pp. 259–265, 2019.

[4] X. Wang, Y. Pang, and X. Ma, "Real distorted images quality assessment based on multi-layer visual perception mechanism and high-level semantics," *Multimedia Tools and Applications*, vol. 79, no. 35, pp. 25905–25920, 2020.

[5] H. Quan, S. Feng, C. Lang, and B. Chen, "Improving person re-identification via attribute-identity representation and visual attention mechanism," *Multimedia Tools and Applications*, vol. 79, no. 11, pp. 7259–7278, 2020.

[6] J. Y. Jiang, F. Guo, J. H. Chen, X.-H. Tian, and W. Lv, "Applying eye-tracking technology to measure interactive experience toward the navigation interface of mobile games considering different visual attention mechanisms," *Applied Sciences*, vol. 9, no. 16, p. 3242, 2019.

[7] A. Cyr and F. Thériault, "Bio-inspired visual attention process using spiking neural networks controlling a camera," *International Journal of Computational Vision and Robotics*, vol. 9, no. 1, pp. 39–55, 2019.

[8] C.-C. Wang, J. C. Hung, S.-N. Chen, and H.-P. Chang, "Tracking students' visual attention on manga-based interactive e-book while reading: an eye-movement approach," *Multimedia Tools and Applications*, vol. 78, no. 4, pp. 4813–4834, 2019.

[9] L. Wang, Y. K. Yang, J. H. Zheng, and X. Y. Wang, "Predicting consumer behaviors from the perspective of consumer neuroscience: current situation, challenges and future," *Journal of Industrial Engineering and Engineering Management*, vol. 34, no. 6, pp. 6–17, 2020.

[10] C. S. Wang, S. F. Chen, and S. L. Zheng, "User interest analysis method of web map interface based on eye movement data," *Geography and Geo-Information Science*, vol. 33, no. 2, pp. 57–62, 2017.

[11] L. J. Dun, Y. Xiong, S. X. Yang, C. Y. Ye, and R. Zhang, "Influence of product types and manifestations of eye movement technology on purchase decision," *Packaging Engineering*, vol. 40, no. 18, pp. 214–219, 2019.

[12] H. L. Li, Q. Xie, L. L. Tang, and Y. J. Liu, "POI recommendation algorithm integrating spatio-temporal information and the importance of interest points," *Computer Applications*, vol. 40, no. 9, pp. 2600–2605, 2020.

[13] C. M. Masciocchi and J. D. Still, "Alternatives to eye tracking for predicting stimulus-driven attentional selection within interfaces," *Human-Computer Interaction*, vol. 28, no. 5, pp. 417–441, 2013.

[14] H. Zhao, Y. Gao, H. Liu, and L. Li, "Fault diagnosis of wind turbine bearing based on stochastic subspace identification and multi-kernel support vector machine," *Journal of Modern Power Systems and Clean Energy*, vol. 7, no. 2, pp. 350–356, 2019.

[15] S. L. Karri, L. C. De Silva, D. T. C. Lai, and S. Y. Yong, "Identification and classification of driving behaviour at signalized intersections using support vector machine," *International Journal of Automation and Computing*, vol. 18, no. 3, pp. 480–491, 2021.

[16] A. Onan, S. Korukoğlu, and H. Bulut, "Ensemble of keyword extraction methods and classifiers in text classification," *Expert Systems with Applications*, vol. 57, no. 15, pp. 232–247, 2016.