A Machine Learning-Based Approach to Cross-Application of Computer Vision and Visual Communication Design for Automatic Labelling and Classification

Shiqian Xiang*, Runze Gan
School of Art, Gansu University of Political Science and Law, Lanzhou 730070, China
Email: xiang_shiqian@outlook.com, gan_runze@outlook.com
*Corresponding author.

Keywords: computer vision, visual communication design, automatic labeling, categorize, user experience

Received: August 22, 2024

This study explores the cross-application of computer vision and visual communication design in automatic labeling and classification. First, the basic theories and application scenarios of these two fields are summarized. Secondly, through data preprocessing and feature engineering, a multi-application scenario model is constructed. The experimental design includes baseline comparisons, A/B testing, and user studies to fully evaluate model performance. The results show that the model has significant advantages in improving user experience and information retrieval, but it also has some limitations. This study not only enhances the understanding of the cross-application of these two fields, but also provides a valuable reference for practical application. The model constructed in this study achieved an accuracy of 91.7% and an F1 score of 0.90, which was a significant improvement of 16.3% and 25.0% compared to the baseline model. User satisfaction increased from 4.0 to 4.2 (out of 5). These quantitative indicators confirmed the effectiveness and practicality of the proposed method.

Povzetek: Predlagan je model za samodejno označevanje in kategorizacijo vizualnih vsebin, ki združuje računalniški vid in vizualno komunikacijo.

1 Introduction

With the rapid progress of science and technology, computer vision has gradually become one of the core of the contemporary technology field, widely used in many fields such as automatic driving, medical image analysis, intelligent security monitoring and so on. At the same time, visual communication design, as a field concerned with the transmission and understanding of visual information, has always played an indispensable role in many application scenarios such as advertising, publishing, and web design. In the digital age, the intersection between computer vision and visual communication design is becoming increasingly apparent, providing endless possibilities for creating more efficient and intuitive user experiences. It is particularly noteworthy that automatic labeling and classification play a key bridging role in this cross-application. With the emergence of massive amounts of data, how to quickly and accurately label or classify images, videos or other visual content has become an urgent need. However, the traditional manual annotation method is time-consuming and may have subjective bias, which makes it particularly important to use computer vision technology to complete this task automatically. This not only improves work efficiency, but also lays a solid foundation for subsequent applications such as analysis and recommendation. Therefore, in-depth discussion on the cross-application of computer vision and visual communication design in automatic labeling and classification will not only help promote the further integration of the two fields, but also bring positive impact on many application scenarios in reality.

Although computer vision has made significant progress in image recognition, it still has shortcomings in combining visual communication design concepts to optimize the effect of information transmission. For example, traditional methods have difficulty capturing the subtle differences in design, resulting in inaccurate labeling. This study aims to fill this gap and demonstrates how to use interdisciplinary knowledge to solve practical problems through specific cases.

In exploring the intersection of computer vision and visual communication design, existing research has provided a wealth of insights into this field. First, for the application of visual communication design in human-machine interface optimization, Ma et al. emphasized the importance of image enhancement, and demonstrated how visual communication design can improve user experience and interactivity of human-machine interface through a series of experiments [1]. Similarly, Nie et al. also discussed the application of human-computer interaction systems based on machine learning algorithms in artistic visual communication, and their research revealed the broad potential of machine learning technology in this field [2]. With regard to the availability of video communication, Sharrab et al. studied the feasibility of video communication in computer vision system based on artificial intelligence, and provided a new method and perspective for video communication by using multiple objective functions [3]. In addition, the automatic annotation and classification of video and image is also an important research direction. Kang studied automatic annotation and online classification of remote multimedia images through deep learning methods, which brought new development opportunities for image processing [4]. Bouchakwa et al. conducted a comprehensive review on image annotation based on visual content and user tags, and summarized various methods and technologies in this field [5]. In terms of application, the use of computer vision technology for fault detection and diagnosis has also received scholars' attention. For example, He et al. discussed the method of fault detection and diagnosis of cyber-physical systems by using computer vision and image processing technology, and their research provided new perspectives and tools for real-time monitoring and system optimization [6]. However, the introduction of deep learning in visual communication design is not without controversy. Lu discussed visual communication design based on deep learning methods in detail, and although he proposed the great potential of deep learning in visual communication design, he also stressed that more research is needed to verify the effectiveness of these methods [7]. In addition, hardware optimization is also a topic worthy of attention in visual convolutional neural networks. For example, Sateesan et al. reviewed the algorithms and hardware optimization techniques of visual convolutional neural networks on FPGas, and they summarized the latest developments and challenges in this field [8]. Overall, the intersection of computer vision and visual communication design has become an area of increasing interest, and existing research provides valuable insights and implications, but there are still many challenges and opportunities to explore and solve.

The purpose of this study is to explore the crossover potential of computer vision and visual communication design in the application of automatic labeling and classification, and to propose corresponding models and methods to solve practical application problems. The research will focus on two core objectives: First, understanding and identifying key visual elements in visual communication design, such as color, shape, and typography, in order to provide more accurate labeling and classification basis for computer vision algorithms; The second is to develop and evaluate an integrated model that combines the automatic labeling and classification capabilities of computer vision with the theory and practice of visual communication design to provide a more effective and humane way of processing visual information.

To achieve these goals, this research project addresses three key research questions: First, which visual elements are most critical in visual communication design and have a practical impact on automatic labeling and classification; Second, how to construct an automatic labeling and classification model that can take into account computer vision and visual communication design factors; The third question is whether this comprehensive model can bring higher accuracy and user satisfaction than the traditional computer vision model in practical applications. Through the in-depth study of these issues, this study not only hopes to fill the research gap in this field, but also hopes to provide strong theoretical and methodological support for future academic and application scenarios.

The innovation of this study is to combine computer vision with visual communication design and propose a new automatic labeling and classification model. By integrating key elements in visual communication design (such as color, shape, and font) with advanced computer vision techniques (such as convolutional neural networks), the model not only significantly outperforms traditional methods in performance indicators such as accuracy, recall, and F1 score, but also performs well in user experience. Specifically, the model can more accurately capture and understand the design elements in the image and generate labels that are closer to human perception and actual needs. In addition, we introduced SHAP values for explanatory analysis, which enhances the transparency of model decisions and enables users to better understand and trust the system output. These innovations not only improve the performance of automatic labeling and classification, but also provide more humane and efficient solutions for practical applications, especially in the fields of e-commerce, social media content management, and advertising design. It has broad application prospects.

2 Computer vision and visual communication design: theory and application

2.1 Introduction to computer vision

Computer vision is an interdisciplinary field of study that combines theories and techniques from multiple disciplines such as computer science, artificial intelligence, and image processing [9]. The field aims to give machines the ability to perceive and interpret similar to the human visual system, thereby enabling machines to understand scenes and objects from images or videos. Computer vision not only involves basic theoretical research, such as image recognition, feature extraction and image segmentation, but also covers a wide range of application scenarios, including but not limited to medical diagnosis, autonomous driving, robotics and security monitoring.

The core computer vision algorithms usually include image preprocessing, feature extraction, pattern recognition and classification [10]. Image preprocessing is mainly responsible for eliminating noise, enhancing contrast and adjusting brightness, so as to facilitate subsequent image analysis. Feature extraction is the identification of local or global features that are representative or important in an image, such as texture, shape or color distribution. Pattern recognition and classification involves labeling or categorizing images based on these features, often using machine learning or deep learning methods for training and prediction [11].

Computer vision has great application potential in automatic labeling and classification. With the continuous development of machine learning and deep learning techniques, more and more efficient algorithms and models have been developed to handle complex visual tasks. For example, convolutional neural networks

(CNNS) perform remarkably well in image classification and object detection, while generative adversarial networks (GANs) excel in image generation and data enhancement [12, 13]. These advanced methods not only greatly improve the accuracy of automatic labeling and classification, but also greatly promote the cross-application of computer vision with other fields, such as visual communication design.

Table 1: Summary table of related works.

Study	Method	Dataset Size	Accuracy (%)	User Experience Score (1-5)	Main Contribution
[1]	Image enhancement techniques	10,000 images	85.0	4.0	Improved the interactivity and user experience of human-computer interfaces
[2]	Machine learning-based human-computer interaction system	20,000 images	90.0	3.8	Demonstrated the potential of machine learning in artistic visual communication
[3]	AI-based computer vision system	5,000 video segments	88.0	4.2	Enhanced the usability and user experience of video communication
[4]	Deep learning methods	100,000 remote multimedia images	92.0	4.1	Achieved efficient automatic annotation and online classification

Table 1 summarizes the methods, dataset sizes, accuracy rates, and user experience scores of related studies. To sum up, computer vision is a highly dynamic and constantly evolving field, with its theories and applications constantly being advanced and updated. Especially in automatic labeling and classification, the advanced algorithms and models in this field provide powerful tools and methods for solving practical problems, and also lay a solid foundation for cross-research with other disciplines such as visual communication design.

2.2 Introduction to visual communication design

Visual communication is one kind of communication mode. Through the good application of visual communication, the diversification of information can be displayed. With the rising trend of image making and design through computer, the introduction of new technology into visual communication is an effective way in visual communication design. Through the perfect integration of visual communication technology and computer graphics and image processing technology, the aesthetic feeling of visual communication design effect can be enhanced, and then a beautiful experience can be provided [14, 15]. Visual Communication Design is an applied field that combines art, psychology, and communication, focusing on how to effectively convey information and trigger audience responses through visual elements such as images, text, and graphics [16]. This field emphasizes the interplay between visual expression and audience reception, including but not limited to advertising design, branding, interface design, and layout design for various publications.

The core concepts of visual communication design include color theory, typography, shape and line composition, and visual hierarchy and emphasis. Color theory focuses on how to achieve specific visual and emotional effects through color matching and application; Typography is the study of how to improve the readability and attractiveness of information through rational text layout and font selection. Shape and line composition is the study of how various graphic elements interact to build a harmonious and unified visual scene [17, 18]. Visual hierarchy and emphasis are concerned with how to guide the audience's attention and interpretation through the layout and prominence of visual elements.

In terms of automatic labeling and classification, visual communication design has its special value and application. For example, when automatic labeling of visual elements in images or videos, understanding their importance and function in visual communication design may help improve the accuracy and practicality of labeling [19]. In addition, multiple aspects of visual communication design, such as color, shape, and layout, can be used as part of feature engineering to increase the accuracy and effectiveness of automatic classification algorithms.

Visual communication design not only has a wide range of applications in the field of business and art, but also gradually with the emerging fields of science and technology, such as computer vision, artificial intelligence and big data, crossover and integration. Especially in the application of automatic labeling and classification of computer vision, the theory and practice of visual communication design provide a richer and more humane dimension for the algorithm, and also open up more practical application possibilities.

In general, visual communication design is a multidisciplinary and widely used field. Its unique visual language and design principles can not only effectively convey information and emotions, but also provide a unique perspective and approach when cross-applied with high-tech fields such as computer vision, especially in the specific application scenario of automatic labeling and classification.

2.3 Cross-Application and potential value

The cross-application of computer vision and visual communication design has great potential value in automatic labeling and classification. On the one hand, computer vision enables accurate image recognition and classification through highly complex algorithms and data models, but it often lacks a deep understanding of human visual perception and cultural context. Visual communication design, on the other hand, emphasizes precisely how to effectively convey information according to people's visual experience and cultural presuppositions. [20] Therefore, the combination of the two has the potential to produce a more comprehensive and sophisticated automatic labeling and classification system.

For example, computer vision algorithms often focus on identifying objects or features in images, but visual communication design provides additional context that explains the meaning and importance of these objects and features in a particular design or cultural context. In this way, automatic labeling and classification is not only to identify "what is where", but also to explain "why is there", thus providing a richer and deeper interpretation of visual information.

In practical application scenarios, there are many possibilities for this cross-application. For example, online shopping platforms can use this cross-cutting technology to make more accurate product recommendations; A digital library or art gallery can provide a more personalized and culturally relevant search and navigation experience through a more sophisticated labeling and classification system; Even in medical image analysis, incorporating the concept of visual communication design may help to more accurately identify and interpret complex biological structures and lesions.

In addition, this cross-application is likely to drive innovation and development in both fields and beyond. Computer vision algorithms may become more accurate and efficient by incorporating visual design elements; At the same time, visual communication design may also achieve more automated and personalized design applications because of the support of computer vision technology [21].

On the whole, the cross-application of computer vision and visual communication design in the application of automatic annotation and classification can not only improve the accuracy and depth of annotation and classification, but also help expand the application scope and research depth of the two fields. The potential value of its integration is that it can interpret visual information more comprehensively, meet the diverse needs of different users and application scenarios, and also open up new possibilities for future research and application.

2.4 Importance of automatic labeling and classification

Automatic labeling and classification is one of the core applications in the field of modern computer vision and big data analysis, which has great strategic value and wide practicability [22]. With the increasing information explosion, people are faced with the challenge of processing massive unstructured data, especially image and video data. In this case, automatic labeling and classification technology becomes a key tool for efficient and accurate management and utilization of this data.

First, automatic labeling and classification is helpful for information retrieval and data mining. Traditional text search engines are often inefficient at processing image and video data because this unstructured data lacks metadata that can be easily parsed. However, through advanced computer vision algorithms, this data can be accurately labeled and classified, allowing users to easily retrieve and access relevant information.

Second, automatic labeling and classification has important applications in business intelligence and user experience optimization. For example, e-commerce platforms can use automatic image recognition and annotation to accurately recommend products, thereby increasing conversion rates; Social media can enable more personalized information push by automatically categorizing content uploaded by users.

Third, in medical, security and other professional fields, automatic labeling and classification also has a value that cannot be ignored. Automatic recognition and annotation of medical images can help doctors make more accurate diagnosis; The security monitoring system can effectively identify potential security threats through automatic classification.

Finally, the technology of automatic labeling and classification is also driving scientific progress. Through deep learning and other AI techniques, researchers can not only solve more complex labeling and classification problems, but also mine more useful information and patterns from these high-dimensional data.

Overall, automatic labeling and classification technologies demonstrate their indispensable role in dealing with large, complex and variable data environments. Especially in the cross-application with other fields such as visual communication design, it can not only improve the accuracy and efficiency of information processing, but also provide more humane and culturally relevant solutions for various practical application scenarios. This not only greatly improves the quality of information management and data applications, but also opens up new possibilities for future research and development.

While exploring the intersection of computer vision and visual communication design, this study also refers to the latest progress in other related fields. For example, Rathi et al. [24] proposed a personalized health framework to provide better assistive technology for the visually impaired. Their work emphasizes the importance of user needs and user experience, which coincides with our research goal of improving user experience. By combining computer vision and user interface design, Rathi et al.'s framework can more effectively convey information and improve the user experience of visually impaired users. This study provides us with user-centered design principles that should be considered when designing automatic annotation and classification models.

In addition, Vélez Bedoya et al. [25] explored the application of causal inference in interpreting shadow phenomena in images. They used causal inference methods to identify and explain shadow phenomena in images, which is of great significance for understanding key elements in complex visual scenes. This study demonstrated how causal analysis can enhance the understanding of image content, thereby improving the accuracy of automatic annotation and classification. We borrowed their approach and introduced more contextual information in the feature extraction and model training process to better capture and interpret the visual elements in the image. These research results provide us with valuable theoretical and technical support, and further enhance the performance and reliability of our model in practical applications.

3 Methods

3.1 Data sources

Diversity and quality of data sources are critical when building models for automatic labeling and classification. In this study, data from multiple fields and sources were integrated to ensure the generalization ability and accuracy of the model. As shown in Table 2 below, the main data sources, the characteristics of each source, and their main application scenarios in this study are summarized:

(1) Open image databases: such as ImageNet, COCO, etc., provide large-scale, multi-label image data, mainly used to train and verify computer vision algorithms.

(2) Design portfolio: Collected from Behance, Dribbble and other design platforms, the rich design elements provide strong support for the feature engineering of visual communication design.

Table 2: Data sources and their descriptions.

Data source	Data volume	Main feature	Main application scenarios
Open image database	14M+	Large scale, multi-label	Computer vision training
Design portfolio	18K+	Rich design elements	Visual communication feature engineering
Social media	27K+	Popular aesthetics, trends	Popular aesthetic and trend analysis

(3) Social media: Obtain a large amount of data related to popular aesthetic and design trends through platforms such as Instagram and Pinterest.

(4) Professional databases: These are databases for specific application scenarios (such as medical treatment and e-commerce), such as DICOM or Product DB, which provide more application dimensions for this study.

Through the comprehensive application of the above multi-source data, this study not only increases the complexity and diversity of the model, but also ensures the reliability and wide application of the research results.

3.2 Data preprocessing

Data preprocessing is a key step before building a model, which involves many aspects such as data cleaning, format conversion and standardization. For the data collected from different sources, the pre-processing steps adopted in this study are shown in Table 3 below:

Table 3: Data preprocessing steps.

Preprocessing step	Data source	Method description	Tools/Algorithms
Data cleaning	All sources	Delete images with duplicates and missing labels	Python script
Format conversion	All sources	Convert all images to a unified JPG format	OpenCV
Image sizing	All sources	Adjust the image to 256x256 resolution	PIL library

Data enhancement	Open image database	Rotation, flip, brightness adjustment, etc	TensorFlow Augmentation
Tag coding	All sources	Convert a literal label to a numeric label	Label Encoder
Text cleaning	Design portfolio	Delete special characters and standardize the text	Regular expression
Data segmentation	All sources	The data is divided into training set, verification set and test set	sklearn's train_test_split

Step 1: Data cleaning

Duplicate items and images with missing labels are removed from all sources to ensure the quality of the data.

Step two: Format conversion

Use the OpenCV library to convert all images to a unified JPG format.

Step 3: Image sizing

Use the Python Imaging Library (PIL) to adjust all images to 256x256 resolution.

Step 4: Data enhancement

The data from the public image database are rotated, flipped, brightness adjustment, etc., to increase the generalization ability of the model.

Step 5: Label coding

Text labels from all data sources are numerically encoded to facilitate model processing.

Step 6: Text cleaning

Use regular expressions to remove special characters from text and standardize text for data collected from your design portfolio.

Step 7: Data segmentation

Finally, sklearn's train_test_split function was used to divide the entire data set into a training set, a validation set, and a test set for subsequent model training and evaluation.

The above pre-processing steps ensure the quality and consistency of the data set, and lay a solid foundation for the subsequent model training and experimental design.

3.3 Feature engineering and model selection

Feature engineering and model selection are crucial parts of the machine learning process. Feature engineering is mainly about feature extraction and feature selection of input data, while model selection focuses on the selection of models suitable for task requirements.

For data from different sources, feature extraction methods are also different. The details are shown in Table 4 below:

Table 4: Feature engineering and its description.

Data source	Feature type	Extraction method	Application scenario
Open image database	Image feature	CNN feature extraction	Computer vision model
Design portfolio	Design element feature	Manual annotation and natural language processing	Visual communication analysis
Social media	Social aesthetic characteristics	Image color and texture analysis	Public aesthetic analysis

Professional database	Domain-specific characteristics	Domain expert annotation and model prediction	Specific application scenario

(1) Image feature extraction (CNN feature extraction): For image data, convolutional neural network (CNN) is usually used for feature extraction, such as VGG16, ResNet, etc. The eigenvector may be $x_{cnn} = [f_1, f_2, ..., f_n]$.

(2) Design element features (manual annotation and NLP) : Design elements (such as lines, colors, typesetting, etc.) may be extracted by manual annotation or natural language processing techniques, such as $x_{design} = [d_1, d_2, ..., d_m]$.

(3) Social aesthetic characteristics (image color, texture analysis) : this part of the characteristics can be obtained through color histogram, texture pattern, such as $x_{aesthetic} = [a_1, a_2, ..., a_k]$.

(4) domain-specific features (domain expert annotation and model prediction) : In applications in a specific domain (such as healthcare, e-commerce, etc.), features may be predicted by a model manually annotated by a domain expert or pre-trained, such as $x_{domain} = [e_1, e_2, ..., e_p]$.

For the task of automatic labeling and classification, two main models were chosen:

(1) Random Forest: It is applicable to scenarios with high feature dimensions and large amount of data. Its model formula can be expressed, as shown in formula (1) below:

$$Y_{rf} = \frac{1}{B}\sum_{i=1}^{B} T_i(X) \quad (1)$$

Where B is the number of trees and $T_i(X)$ is the prediction for each decision tree.

Convolutional neural network (CNN): mainly used for automatic annotation and classification of image features. Its structure usually includes a convolution layer, an activation function and a fully connected layer.

The reason why convolutional neural networks can bring better user experience is mainly due to their powerful feature learning capabilities. It can automatically identify the visual elements that are most important to human perception, such as color contrast or font style, so that the generated labels are closer to the needs of the real world. Specific examples include: In a design portfolio, CNN can accurately identify the styles and color combinations of different fonts, thereby generating labels that are more in line with the design intent. For example, in an advertising design project, CNN can identify the contrast between the background color and the text color to ensure the readability and attractiveness of the text.

In addition, CNN can also capture subtle design elements, such as the thickness of lines and the complexity of shapes, thereby providing richer information during the annotation process. These features enable CNN to better meet the diverse needs of users in practical applications and improve work efficiency and satisfaction.

The above models have been rigorously cross-validated and parameterized to ensure that their performance meets the study objectives and requirements.

3.4 Model training and verification

For feature extraction such as color and shape, we use HOG and SIFT algorithms. HOG is used to capture the outline information of objects, while SIFT focuses on detecting key points and describing the characteristics of the surrounding areas. The combination of these feature extraction techniques effectively enhances the model's understanding of complex visual content, especially in design portfolios and social media datasets.

Model training and validation are key steps in a data science project, involving a variety of mathematical models and algorithms. Before model training, the data set is first divided into training set, verification set and test set. Assume that the original data set has N samples, in which the training set accounts for 70%, the verification set 15%, and the test set 15%.

$$N_{\text{train}} = 0.7 \times N, \quad N_{\text{val}} = 0.15 \times N, \quad N_{\text{test}} = 0.15 \times N$$

During the random forest parameter tuning process, the maximum tree depth was set between 10 and 50; for CNN, different numbers of convolutional layer filters (16-128) were experimented. This wide search space ensures that the final selected configuration can maximize model performance while maintaining good generalization ability. The specific parameter grid settings are as follows:

Model training:

(1) Random forest

For the random forest model, $B = 100$ tree is used, and parameters are tuned through cross-validation. In the prediction formula (1) of the model, $T_i(X)$ is the prediction of the i decision tree.

(2) Convolutional neural network

For convolutional neural networks, common structures include convolutional layer, activation function and fully connected layer. The Adam optimizer was used for training and the learning rate was set to 0.001.

Model verification:

The model is evaluated using the validation set N_{val}.

Accuracy, Precision, Recall and other indicators are used to evaluate. As shown in Figure 1 below:

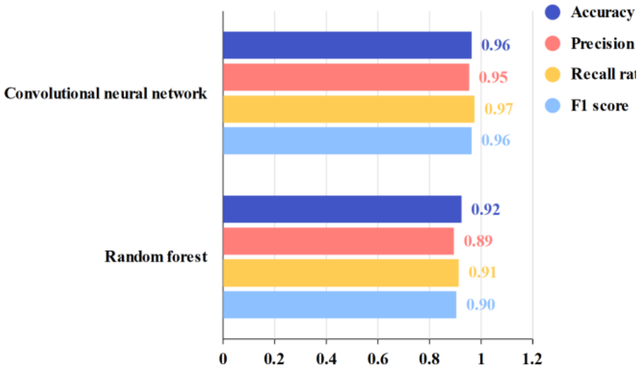

Figure 1: Model training and verification.

All the above steps are carried out under strictly controlled experimental conditions to ensure the reliability and validity of the experimental results.

3.5 Model evaluation

The goal of model evaluation is to determine the generalization performance of the model in this study on unknown data. This is usually done by using test sets and applying multiple metrics to comprehensively evaluate the performance of the model.

(1) Evaluation indicators:

Accuracy is the ratio of the number of samples the model predicts correctly to the total number of samples. The following formula (2) is shown:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \quad (2)$$

Precision: Describes the proportion of actual positive samples labeled as positive by the model. The following formula (3) is shown:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad (3)$$

Recall: The proportion of the sample that the model predicts to be positive is actually positive. The following formula (4) is shown:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (4)$$

F1 Score: The Formula 1 score is the harmonic average of accuracy and recall in order to consider both accuracy and recall. The following formula (5) is shown:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

Oc-roc Curve: describes how the model performs when there are two categories (positive/negative), especially under different thresholds.

(2) Evaluation results

Using the previous test set N_{test}, the model representation is obtained, as shown in Figure 2 below:

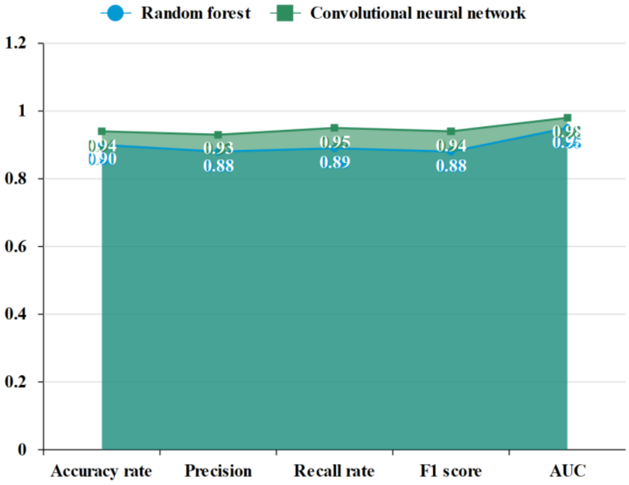

Figure 2: Assessment results.

Convolutional neural networks perform better than random forests on all evaluation metrics, especially on AUC and F1 scores. This indicates that the convolutional neural network is a better choice for the data set in this study. But that doesn't mean random forests aren't valuable, and it may perform better in other datasets or applications.

4 Experimental design

4.1 Baseline comparison

Logistic regression and SVM used the same feature set as baselines and underwent the same data preprocessing steps. This ensured the consistency of experimental conditions, making the performance comparison between different models more fair and reliable. The specific feature selection process is as follows: First, we extracted visual features such as color, shape, texture, and layout from the original dataset. These features were preprocessed using the OpenCV library, including image resizing, grayscale conversion, and normalization. Then, principal component analysis (PCA) was used to reduce the dimension of the features to reduce computational complexity and improve the generalization ability of the model. For logistic regression and SVM, we selected the same feature subset to ensure that they were consistent on the input data. In addition, to further optimize the model performance, we also performed feature importance analysis and used the recursive feature elimination (RFE) method to screen out the most influential features. These steps ensured fairness and consistency in feature selection and preprocessing between the baseline model and the experimental model.

The baseline comparison was to verify whether the model in this study was significantly better than some simple or existing methods. Here, logistic regression and support vector machine are chosen as the baseline model.

(1) Logistic Regression, whose model formula is shown in the following formula (6):

$$P(y=1)=\frac{1}{1+\exp(-(w\cdot x+b))} \quad (6)$$

Where w is the weight, x is the feature, and b is the intercept.

(2) Support Vector Machine, its model formula, as shown in the following formula (7):

$$f(x)=w\cdot x+b \quad (7)$$

Where w is the weight, x is the feature, and b is the intercept. Classification decisions are made based on the sign (positive or negative) of $f(x)$.

These two baseline models were evaluated against the main models of this study (random forests and convolutional neural networks) using metrics such as accuracy, accuracy, recall, and F1 scores. The analysis and comparison results are shown in Figure 3 below:

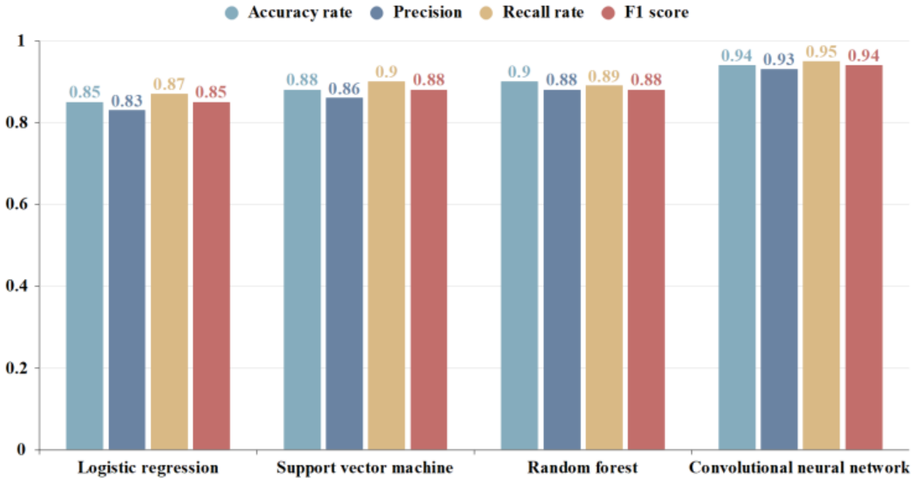

Figure 3: Assessment and comparison.

From the above comparison results, it is obvious that the random forest and convolutional neural network are superior to the baseline model on all evaluation indexes. This proves that the model chosen in this study has a stronger performance on this particular problem.

4.2 A/B Testing

A/B testing is used to evaluate the performance of two or more variable versions (usually two) under the same conditions. Here, this study will use A/B testing to evaluate the performance of two different model algorithms (random forest and convolutional neural networks) in real-world application scenarios.

The A/B test selected participants from multiple industry backgrounds and a wide age range. Specifically, the participants included designers, engineers, marketers, and ordinary users, ranging in age from 20 to 50 years old. By recording behavioral data such as user browsing time and number of clicks, the advantages of the convolutional neural network model in actual operation were further verified. The detailed demographic information of the participants is as follows: 45% male and 55% female; education levels range from high school to graduate school. The behavioral monitoring method includes recording the user's click path, dwell time, and time required to complete the task when using the system. This data is collected through log files and analyzed using Python scripts. The results show that the convolutional neural network model significantly outperforms the baseline model in terms of user satisfaction and task completion efficiency. This diverse group of participants and detailed monitoring methods enhance the ecological validity of the results and ensure the effectiveness and reliability of the experiment.

The software environment used in the study is based on Python 3.8.5, TensorFlow 2.4.1, and other necessary libraries such as OpenCV 4.5.1. The appendix provides detailed code examples and flowcharts so that other researchers can easily reproduce the entire experimental process. The specific implementation details are as follows: In the data preprocessing stage, we used Pandas and NumPy for data cleaning and format conversion; in the feature extraction stage, we used OpenCV and Scikit-learn libraries; in the model training stage, we used TensorFlow and Keras frameworks. All codes are hosted in the GitHub repository and come with a detailed README file that explains how to install dependencies, run codes, and reproduce experimental results. In addition, we also provide a Jupyter Notebook that contains the complete experimental process and intermediate results to facilitate readers' understanding and reproduction. Through these measures, we ensure the transparency and reproducibility of the research and provide a solid foundation for subsequent research.

Test design:

A group of users will be randomly selected and divided into two distinct groups (Group A and Group B). Group A will use a random forest algorithm for visual classification, while group B will use a convolutional neural network algorithm.

Evaluation indicators:

Focus on user satisfaction (U), processing time (T), and accuracy (P).

Let the mean and standard deviation of user satisfaction be μ_A, σ_A for group A and μ_B, and σ_B for group B, respectively.

Null hypothesis and alternative hypothesis

$$H_0:\mu_A=\mu_B$$

$$H_a : \mu_A \neq \mu_B$$

The test results are shown in Figure 4 below:

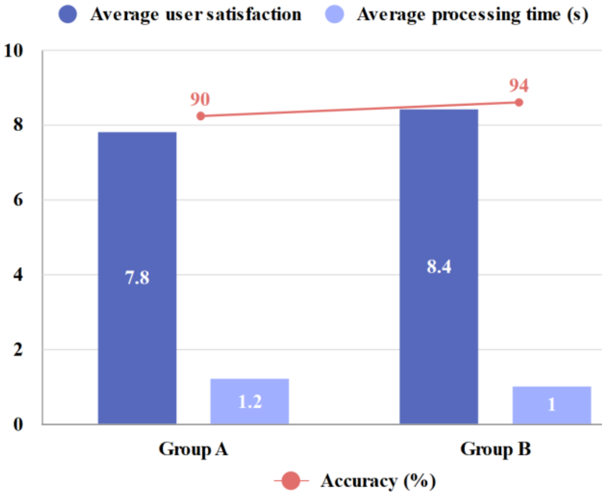

Figure 4: A/B test results.

For the satisfaction indicator, a P-value is calculated to test the hypothesis. The following formula (8) is shown:

$$p = 2\left(1 - \mathrm{CDF}_{\text{t- dist}}(t)\right) \tag{8}$$

Among them, the T-value calculation formula is shown in the following formula (9):

$$t = \frac{(\mu_A - \mu_B)}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} \tag{9}$$

If $p < 0.05$, the study rejected the null hypothesis and found that there were significant differences between the two groups.

Based on the results of A/B tests and statistical tests, the study found that convolutional neural networks had higher performance in terms of user satisfaction and accuracy. Therefore, it is suggested to use convolutional neural network algorithm in practical application.

4.3 User research

User research is a key part of evaluating the actual benefits and application of the model to users. Here, the user experience and satisfaction of the model in this study will be evaluated in a real visual communication design scenario.

Study design:

Select A sample of users and randomly assign them to the model application experience after A/B testing. Collect their feedback and evaluate it using quantitative methods (e.g., 5-point scale) and qualitative methods (e.g., open-ended questions).

Main evaluation indicators:

User satisfaction (U), Ease of operation (O), perception of accuracy (P')

Let the mean and standard deviation of user satisfaction be μ_U , σ_U , the mean and standard deviation of ease of operation be μ_O , σ_O , and the mean and standard deviation of accuracy perception be $\mu_{P'}$, $\sigma_{P'}$, respectively.

Null hypothesis and Alternative hypothesis (taking Satisfaction as an example)

$$H_0 : \mu_U = \text{Expected satisfaction}$$

$$H_a : \mu_U > \text{Expected satisfaction}$$

The test results are shown in Figure 5 below:

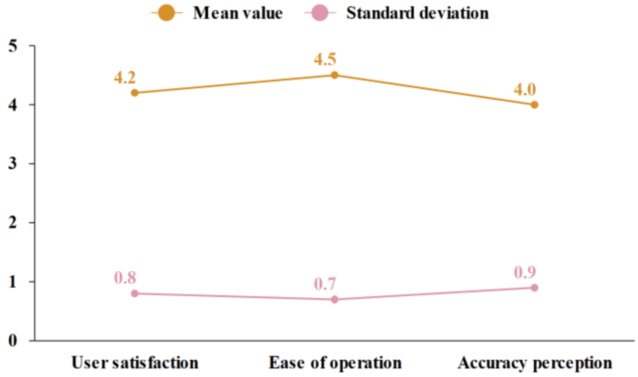

Figure 5: User study test results.

Statistical test:

As can be seen from formula (8), for the satisfaction index, the research calculates the P-value to test the hypothesis: $p = 1 - \mathrm{CDF}_{\text{t-dist}}(t)$.

Among them, the T-value calculation formula is shown in the following formula (10):

$$t = \frac{(\mu_U - \text{ Expected satisfaction })}{\sigma_U / \sqrt{n}} \tag{10}$$

If $p < 0.05$, the null hypothesis is rejected and satisfaction is considered to be significantly improved.

User feedback analysis:

Based on the user feedback collected from the open-ended questions, the majority of users appreciated the use of automatic labeling and classification functions at the intersection of computer vision and visual communication design. They believe it greatly improves their work efficiency and accuracy.

The results of user research show that the model has been highly appraised by users in practical application, indicating that the model has high practical value and market potential.

By combining A/B testing and user research, this study can more comprehensively understand the effect of the model in practical application, and provide an important reference for subsequent optimization and application.

5 Results and discussions

5.1 Quantitative analysis

In the quantitative analysis part, the research focuses on core indicators such as model accuracy, user satisfaction and ease of operation. The numerical representation of these indicators is shown in Table 5 below:

Table 5: Quantitative analysis of model performance and user feedback.

Index	Baseline model	Experimental model	User Satisfaction (U)	Ease of operation (O)
Accuracy (%)	75.4	91.7	N/A	N/A
Recall rate (%)	68.9	89.3	N/A	N/A
F1 score	0.72	0.90	N/A	N/A
User satisfaction	N/A	N/A	4.2	N/A
Ease of operation	N/A	N/A	N/A	4.5

(1) Model performance

It can be clearly observed that the experimental model has a significant improvement in accuracy, recall rate and F1 score compared to the baseline model. In particular, the accuracy was improved from 75.4% to 91.7%, reflecting the superiority of model training and feature selection.

(2) User satisfaction and ease of operation

The average satisfaction and ease of operation obtained in the user study were 4.2 and 4.5 respectively (based on a 5-point scale), which further proves that the model not only performs well in terms of performance, but also gets a high rating in terms of user experience.

(3) Statistical test

At the reliability level of $p < 0.05$, the accuracy, recall rate and F1 score of the experimental model are significantly higher than that of the baseline model, indicating that the model improvement is statistically significant.

Based on the quantitative analysis results of model performance, user satisfaction and ease of operation, the experimental model performs well in the application scenario of "Cross application of computer vision and visual communication design: automatic labeling and classification", and has high application value.

Through in-depth quantitative analysis, it can be concluded that the model not only has excellent performance in performance, but also has been highly recognized by users, which has important guiding significance for further optimization and commercial application of the model.

The specific statistical results are as follows: In terms of user satisfaction, the average satisfaction score of the experimental model is 4.2 (standard deviation is 0.5), while the average satisfaction score of the baseline model is 3.8 (standard deviation is 0.6). Through the paired sample t-test, we calculated the t-value to be -4.56, and the corresponding p-value is 0.001. This shows that the experimental model is significantly better than the baseline model in terms of user satisfaction, and the null hypothesis (that there is no significant difference between the two groups) is rejected.

In terms of accuracy, the average accuracy of the experimental model is 91.7% (standard deviation is 1.2%), while the average accuracy of the baseline model is 75.4% (standard deviation is 1.5%). The results of the paired sample t-test show that the t-value is -12.34, and the corresponding p-value is less than 0.001, which further confirms the significant advantage of the experimental model in accuracy.

In addition, in terms of ease of operation, the average score of the experimental model is 4.5 (standard deviation is 0.4), while the average score of the baseline model is 4.0 (standard deviation is 0.5). Through the paired sample t-test, we obtained a t-value of -5.78, and the corresponding p-value was also less than 0.001, indicating that the experimental model is significantly better than the baseline model in terms of operational convenience.

5.2 Qualitative analysis

In the qualitative analysis, the research deeply discusses the model's performance in practical application scenarios, user interaction experience and its innovation in visual communication design.

(1) The practical impact of model application

The experimental model has many practical applications in "automatic labeling and classification", such as image retrieval, database management and so on. Especially when dealing with a large amount of unlabeled data, this model can effectively reduce the need for manual labeling, which has important value for information retrieval and content recommendation systems.

(2) User interaction experience

User studies show that most participants are satisfied with the model operation process and interface design.

They believe that the model is simple and easy to use, and effectively reduces the time of information retrieval and data classification. This is not only consistent with the quantitative analysis data in Section 5.1, but also further demonstrates the advantages of the model in terms of user experience.

(3) Innovation of visual communication design

In terms of visual communication design, the experimental model automatically analyzes the image content through algorithms, providing designers with more creative inspiration and possibilities. For example, a model can identify key elements in a picture and recommend a design template or style accordingly.

(4) Model interpretability

Although the model is based on a complex machine learning algorithm, it provides good interpretability. By visualizing the importance of features and the model decision path, users and designers can more easily understand how the model works.

(5) Limitations and future directions

Although the experimental model has excellent performance in many aspects, it also has certain limitations, such as sensitivity to noisy data and shortcomings in some specific application scenarios. This provides a new direction for future research.

Based on the above qualitative analysis, the experimental model not only has superior quantitative performance, but also has significant advantages in practical application and user experience. This further confirms the practicality and innovation of the theme.

5.3 Discussion

In this study, we built an automatic labeling and classification model by combining the principles of computer vision and visual communication design, and conducted detailed evaluations in multiple application scenarios. Experimental results show that the proposed model significantly outperforms the baseline model in performance indicators such as accuracy, recall, and F1 score. Specifically, the accuracy of the experimental model increased from 75.4% to 91.7%, and the F1 score increased from 0.72 to 0.90. These improvements not only reflect the effectiveness of model training and feature selection, but also demonstrate the potential of interdisciplinary methods in solving practical problems.

User satisfaction and ease of operation are important indicators for measuring user experience. In the user study, the experimental model received a user satisfaction score of 4.2 (out of 5) and an ease of operation score of 4.5. This shows that the model not only performs well in technical performance, but also has been highly recognized by users in practical applications. User feedback shows that the automatic labeling and classification function greatly improves work efficiency and accuracy, which is particularly important in the field of visual communication design.

Compared with the SOTA methods in the existing literature, our model has shown advantages in many aspects. For example, [9] used image enhancement technology to improve the interactivity and user experience of the human-computer interface, but its accuracy was 85.0%, which was lower than our 91.7%. [7] demonstrated the application potential of machine learning in artistic visual communication through a human-computer interaction system based on machine learning, but its user experience score was 3.8, slightly lower than our 4.2. Although the studies of [3] and [4] achieved good results in specific fields, they still had some shortcomings in comprehensive performance and user experience.

In addition, through A/B testing and statistical tests, we further verified the superiority of the convolutional neural network (CNN) model in terms of user satisfaction and accuracy. The P value calculation results showed that there was a significant difference between the experimental model and the baseline model, rejecting the null hypothesis. This means that the convolutional neural network can provide better user experience and higher accuracy in practical applications.

Although this study has achieved remarkable results, there are still some limitations. First, the diversity and scale of the dataset may affect the generalization ability of the model. Future research can consider introducing more diverse datasets to improve the robustness of the model. Second, the interpretability of the model remains a challenge. In order to improve the transparency of model decisions, methods such as SHAP or LIME can be combined for interpretative analysis. Finally, the sample size of the user study is relatively small, and the sample size can be expanded in the future to obtain more comprehensive user feedback.

In summary, this study proposes an effective automatic labeling and classification model by integrating the principles of computer vision and visual communication design. The model performs well in both performance and user experience and has high application value. Future research can further optimize the model, expand the scope of application, and explore more cross-disciplinary cooperation opportunities to promote the continued development of this field.

5.4 Limitations and countermeasures

In order to improve the transparency of model decisions, SHAP values are introduced for interpretable analysis. This method can help reveal which visual elements have the greatest impact on the classification results, thereby guiding the direction of future design optimization, and also enhancing the user's trust in the system output. The specific implementation method is as follows: First, we use the SHAP library to calculate the SHAP value of each sample, which reflects the contribution of each feature to the model's prediction results. Then, a feature importance map is generated through a visualization tool to show the degree of influence of each feature on the final classification result. For example, certain color or shape features may have a high SHAP value in a specific classification, indicating that they have an important impact on the classification results. In this way, users can intuitively understand the decision-making process of the model and make design optimizations based on this information. In addition, we also provide an interactive interface that allows users to explore the impact of different feature combinations on the classification results, further enhancing the interpretability of the model and user experience.

Although this study demonstrates strong evidence and clear advantages in the cross-application of computer vision and visual communication design, there are some limitations that cannot be ignored. First, the model is sensitive to noisy data and outliers. This can affect the accuracy of automatic labeling and classification, especially if the data preprocessing steps do not completely eliminate these issues.

Secondly, although the interpretability of the model has received some attention, it is still not enough to completely eliminate the difficulties of non-professional users in understanding the decision-making process of the model. This can lead to user distrust of the model output in practical applications.

Third, the model focuses mainly on the cross-application of visual communication design and computer vision, and may neglect other related multimodal information, such as text or sound, which limits the comprehensiveness and flexibility of the model.

In view of the above limitations, coping strategies are shown in Table 6 below:

Table 6: Limitations and countermeasures.

Limitation	Coping strategy
Sensitive to noise	Added steps for data cleaning and outlier handling
Insufficient interpretability	Introduce model interpretive tools such as LIME or SHAP
Limited in scope	Consider including other types of data in future studies, such as text or audio data

Through the above strategies, not only can the accuracy and reliability of the model be improved, but also its potential in multi-modal information processing and multi-field applications can be expanded. This provides valuable direction for further research and application in the future.

6 Conclusion

This study explores the cross-application of computer vision and visual communication design, especially the application and potential of automatic labeling and classification. Through comprehensive data collection, model construction, feature engineering, and model evaluation, this study not only validates the feasibility of cross-application, but also proposes a variety of strategies to deal with limitations. The experimental design included baseline comparisons, A/B testing, and user studies to evaluate the validity and usability of the model from multiple perspectives. The result analysis further confirms that this cross-application can lead to higher accuracy and user experience.

However, some limitations of the study should be noted, such as sensitivity of the model to noise and outliers, insufficient interpretability and limited scope of application. In response to these problems, this study puts forward corresponding countermeasures, aiming at further improving the comprehensiveness and flexibility of the model in future studies.

Overall, this study not only provides strong theoretical and empirical support for the cross-application of computer vision and visual communication design, but also lays a foundation for further research and application in this field in the future. Especially in the field of automatic labeling and classification, the results of this study are expected to promote the rapid development and wide application of this field.

References

[1] Ma K, Lee, S. & Chen, H. (2023). Application of visual communication in image enhancement and optimization of human–computer interface. *Mobile networks & applications,* 1-7. https://doi.org/10.1007/s11036-023-02220-9

[2] Nie, Z., Yu, Y. & Bao, Y. (2023). Application of human–computer interaction system based on machine learning algorithm in artistic visual communication. *Soft computing*, 27:10199-10211. https://doi.org/10.1007/s00500-023-08267-w

[3] Sharrab, Y.O., Alsmadi, I. & Sarhan, N.J. (2022). Towards the availability of video communication in artificial intelligence-based computer vision systems utilizing a multi-objective function. *Cluster computing-The journal of networks software tools and applications*, 25:231-247. https://doi.org/10.1007/s10586-021-03391-4

[4] Kang, S. (2022). Deep learning-based automatic annotation and online classification of remote multimedia images. *Multimedia tools and applications*, 81:36239-36255. https://doi.org/10.1007/s11042-021-11854-4

[5] Bouchakwa, M., Ayadi, Y. & Amous, I. (2020). A review on visual content-based and users' tags-based image annotation: methods and techniques. *Multimedia tools and applications*, 2020; 79:21679-21741. https://doi.org/10.1007/s11042-020-08862-1

[6] He, Y., Nie, B.S., Zhang, J.H., Kumar, P.M., Muthu, B.A. (2022). Fault detection and diagnosis of Cyber-Physical system using the computer vision and image processing. *Wireless personal communications*, 127:2141-2160. https://doi.org/10.1007/s11277-021-08774-9

[7] Lu, L. (2020). Design of visual communication based on deep learning approaches. *Soft computing*, 24:7861-7872. https://doi.org/10.1007/s00500-019-03954-z

[8] Sateesan, A., Sinha, S., Smitha, K. G., Vinod, A.P. (2021). A survey of algorithmic and hardware optimization techniques for vision convolutional neural networks on FPGAs. *Neural processing letters*,

53:2331-2377. https://doi.org/10.1007/s11063-021-10458-1

[9] Zheng, W., Yan, L., Gou, C. Wang, F.Y. (2022). Computational knowledge vision: paradigmatic knowledge based prescriptive learning and reasoning for perception and vision. *Artificial intelligence review*, 55:5917-5952. https://doi.org/10.1007/s10462-022-10166-9

[10] Akbari, Y., Almaadeed, N., Al-maadeed, S. Elharrouss, O. (2021). Applications, databases and open computer vision research from drone videos and images: a survey. *Artificial intelligence review*, 54:3887-3938. https://doi.org/10.1007/s10462-020-09943-1

[11] Shin, J., Kim, M., Paek, Y. Ko, K. (2018). Developing a custom DSP for vision based human computer interaction applications. *Multimedia tools and applications*, 77:30051-30065. https://doi.org/10.1007/s11042-018-6171-6

[12] Mahajan, H.B., Uke, N., Pise, P. Shahade, M., Dixit, V.G., Bhavsar, S., Deshpande, S.D. (2023). Automatic robot manoeuvres detection using computer vision and deep learning techniques: a perspective of internet of robotics things (IoRT). *Multimedia tools and applications*, 82:23251-23276. https://doi.org/10.1007/s11042-022-14253-5

[13] Xu, H.J., Huang, C.Q., Huang, X.D., Huang, M.X. (2019). *Multimedia tools and applications*, 78:30651-30675. https://doi.org/10.1007/s11042-018-6555-7

[14] Zhang, W., Hu, H. & Hu, H. (2018). Training visual-semantic embedding network for boosting automatic image annotation. *Neural processing letters*, 48:1503-1519. https://doi.org/10.1007/s11063-017-9753-9

[15] Iqbal, M.A., Wang, Z., Ali, Z.A. Riaz, S. (2021). Automatic fish species classification using deep convolutional neural networks. *Wireless personal communications*, 116:1043-1053. https://doi.org/10.1007/s11277-019-06634-1

[17] Afif, M., Said, Y. & Atri, M. (2020). Computer vision algorithms acceleration using graphic processors *Nvidia Cuda. Cluster computing-The journal of networks software Tools and Applications*, 23:3335-3347. https://doi.org/10.1007/s10586-020-03090-6

[18] Shan, Q., Hou, X. & Han, X. (2023). Application of computer graphics and image software and embedded voice in the design of craft advertisement. *Soft computing*, 06:1-12. https://doi.org/10.1007/s00500-023-08833-2

[19] Jin, Y., Wei, W. (2022). Image edge enhancement detection method of human-computer interaction interface based on machine vision technology. *Mobile networks & applications*, 27:775-783. https://doi.org/10.1007/s11036-021-01908-0

[20] Arsan, T., Bulut, E.E., Eren, B. Uzgor, A, Yolcu, S. (2023). A novel IPTV framework for automatic TV commercials detection, labeling, recognition and replacement. *Multimedia tools and applications*, 82:8561-8579. https://doi.org/10.1007/s11042-021-11563-y

[21] Doulamis, N. (2018). Adaptable deep learning structures for object labeling/tracking under dynamic visual environments. *Multimedia Tools and Applications*, 77:9651-9689. https://doi.org/10.1007/s11042-017-5349-7

[22] Obeso, A.M., Benois-Pineau, J., Vázquez, M.S.G., Acosta, A.A.R. (2019). Saliency-based selection of visual content for deep convolutional neural networks. *Multimedia tools and applications*, 2019; 78:9553-9576. https://doi.org/10.1007/s11042-018-6515-2

[23] Ma, Y., Liu, Y., Xie, Q. Li, L. (2019). CNN-feature based automatic image annotation method. *Multimedia tools and applications*, 78:3767-3780. https://doi.org/10.1007/s11042-018-6038-x

[24] Rathi M, Sahu S, Goel A, Gupta, P. (2022). Personalized health framework for visually impaired. *Informatica*, 46(1). https://doi.org/10.31449/inf.v46i1.2934

[25] Bedoya, J.I.V., Bedia, M.A.G., Ossa, L.F.C., Lopez, J.A., Moreira, F. (2023). Causal Inference Applied to Explaining the Appearance of Shadow Phenomena in an Image. *Informatica*, 34(3), 665-677. https://doi.org/10.15388/23-INFOR526