A Study on the Correlation Analysis between Product Cognitive Features and Visual Features

Zhuen Guo and Li Lin(✉)

School of Mechanical Engineering, Guizhou University, Guiyang, Guizhou 550025, China

715842042@qq.com

Abstract. Obtaining key feature elements based on design requirements is a fundamental task in the product design process. To explore the role of feature recognition in product design in algorithmic classification. This study explores the feasibility of using algorithmically derived product visual features to replace traditional methods for acquiring key product features by analyzing the correlation between user cognitive data obtained from eye tracking tests and visual data derived from category mapping activations based on algorithms. To begin, eye tracking tests are conducted to understand users' cognitive processes when viewing product images under different emotional objectives, and the findings are visualized through heatmaps. Subsequently, the ResNet34 classification algorithm is applied for image classification training, incorporating the Grad CAM visualization layer to obtain the visual feature data of product images. An association analysis is performed on the significant features obtained through both approaches. Finally, a conclusion is drawn regarding the feasibility of obtaining key product feature data based on the algorithmic model.

Keywords: Design Features, Eye-Tracking Tests, Grad-CAM

1 Introduction

Product innovation encompasses not only functional improvements but also innovations in form and other aspects [1]. In the consumer product market, where product functionalities tend to converge and product lifecycles become increasingly similar due to improved design and manufacturing capabilities, numerous products can meet users' basic functional and quality needs [2]. Consequently, major manufacturing enterprises are now focusing on enhancing product competitiveness beyond these basic functionalities. In the realm of product innovation design, meeting users' personalized emotional and psychological demands for homogeneous and functionally similar products can provide a competitive advantage in the target market and quickly gain consumer support [3]. Thus, the identification and localization of key product features that satisfy users' emotional needs become crucial in this process.

With the rapid advancement of intelligent technology, particularly deep learning, scholars have integrated this technology into various fields. Notably, computer vision research has become a prominent topic, with one direction focusing on using image

R. Appleby et al. (eds.), *Proceedings of the 2nd International Conference on Intelligent Design and Innovative Technology (ICIDIT 2023)*, Atlantis Highlights in Intelligent Systems 10,
https://doi.org/10.2991/978-94-6463-266-8_26

classification to overcome the inefficiencies of traditional manual classification [4]. In this context, combining image classification research with emotional classification and categorizing product images offers a means to incorporate intelligent technology into the design process. While intelligent technology effectively handles product image classification, it lacks insights into the algorithm's classification process compared to traditional methods like eye-tracking tests. However, the image classification process can be enhanced by visualizing the primary features of the classification process through the incorporation of the Class Activation Mapping (CAM) visualization layer. Consequently, our investigation aims to explore the correlation between visual feature data obtained through algorithm-based product image classification and cognitive feature data obtained from eye-tracking tests to improve emotion-based classification and rapidly acquire key product features [5-6].

The subsequent sections of this paper are structured as follows: Part 2 introduces the cognitive feature extraction process based on eye-tracking tests, Part 3 presents the classification algorithm and visual feature extraction process, Part 4 discusses the correlation analysis between cognitive and visual features, and finally, the study is summarized.

2 Cognitive Feature Extraction Based on Eye-Tracking Tests

The concept of "salience" is commonly categorized into cognitive salience and visual perceptual salience. Cognitive salience refers to the top-down process of human perception, involving the differentiation of objects under the strong influence of an observer's experiences, interests, and expectations [7-8]. On the other hand, visual perceptual salience refers to the bottom-up perceptual process in which objects attract attention based on their physical characteristics and degree of integration into the environment. In the design field, the implication of salience features is primarily studied in key design aspects, such as object recognizability-guided features, features influenced by product satisfaction, image-guided design features, and user demand-guided design features [9]. Building on these concepts, we designed a cognitive eye-tracking experiment using a dataset of automotive emotional samples as stimuli, with the emotion terms "stable" and "lively" as the classification categories [10]. The dataset of automotive emotional samples is shown in Figure 1.

Fig. 1. Automotive emotional samples

The cognitive eye-tracking experiment was conducted using the psychological operation software EPRIME-2. We employed an SMI-RED desktop remote eye tracker to measure users' eye movement data. The cognitive eye-tracking experiment took place in a controlled environment, featuring closed, quiet, softly lit, and noise-free laboratory settings. Prior to the experiment, participants' eye trackers were calibrated. At the onset of the experiment, instructions and operational guidelines were displayed on the screen, requiring participants to fully comprehend the experiment's objectives before proceeding. In the pre-experimental stage, participants familiarized themselves with the specific operational procedures and experiment flow to ensure smooth execution during the formal experiment. The sample images used in the pre-experiment phase comprised a fixed set of 30 images, selected and evaluated for emotional semantics by a focus group. Participants' choices were compared to the initial information, and an accuracy exceeding 90% indicated their understanding of the experiment's purpose, enabling them to proceed to the next stage of the formal experiment. Instances where accuracy was below 90% required retraining. The formal experiment followed the same procedures as the pre-experiment, with specific steps including the presentation of a "+" sign on the screen for 1 second to guide participants' attention. Subsequently, the stimulus sample image appeared, and participants judged the imagery words, clicking the left mouse button for "stable" semantics and the right mouse button for "lively" semantics. The data collected in this study consisted of heatmap data, representing the heatmaps created by all participants for each stimulus.

3 Deep Learning-Based Visual Feature Extraction

3.1 Design of Emotion Classification Recognition Model Network

Deep convolutional neural networks can extract higher-level image features. As automotive product images often possess complex geometric and topological characteristics, this section uses deeper convolutional neural networks as the foundation for constructing the emotion image recognition classification model. However, as the network depth increases, the problem of vanishing gradients becomes more severe, making network optimization increasingly challenging. The design of residual blocks in the ResNet network partially addresses the issue of vanishing gradients, enabling the network model to achieve good generalization performance while deepening the network. Therefore, a ResNet structure was adopted to construct the automotive emotion image recognition classification model. By comparing the performance of different-depth network models, the optimal emotion classification deep learning model was obtained. Taking ResNet-34 as an example, its network structure is illustrated in Figure 2.

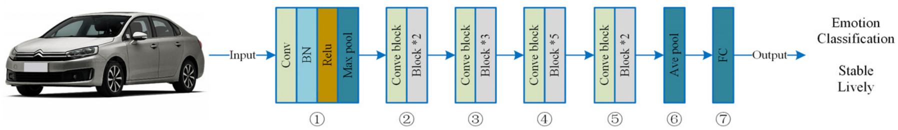

Fig. 2. ResNet-34 Network Architecture

3.2 Validation of Emotion Classification Model Performance

After setting up the network architecture and training parameters, the automotive front image emotion classification dataset was divided into training, validation, and testing sets in a ratio of 7:2:1. The training set was used to train the model, the validation set was utilized to fine-tune the model's parameters, and the evaluation on the validation set helped in finding the best hyperparameter combination to improve the model's generalization ability. The testing set was used to assess the model's generalization performance. The training, validation, and testing sets were randomly sampled from the annotated automotive front image emotion classification dataset.

The research used the server cluster of the Guizhou University National Key Laboratory of Public Big Data for training. The training environment was Ubuntu 18.04, and TensorFlow framework was used with a single NVIDIA A100 GPU for computation.

3.3 Visual Feature Extraction

Class Activation Mapping (CAM) is a commonly used feature visualization technique that highlights the image regions most important for the classification task. Gradient-weighted Class Activation Mapping (Grad-CAM) is an improved version of CAM that uses gradient information to weight the contribution of each pixel in CAM. By using gradient information to weight the contribution of each pixel in CAM, Grad-CAM can better capture the nonlinear behaviors of deep neural networks and provide more accurate visualization results. This helps to understand how the top convolutional layer focuses on different regions in the image and generates the classification result.

4 Analysis of Key Feature Elements

Based on the aforementioned Grad-CAM algorithm, a network structure for recognizing salient features in emotion imagery was constructed. The ResNet-34 model was used to batch classify automotive sample images and identify the visualized salient features, i.e., the large-scale emotion imagery key features. Several heatmaps and Grad-CAM visualizations were randomly selected, and key feature elements were extracted, as shown in the example Figure 3. By comparing the heatmaps and Grad-CAM visualizations of the samples, it can be observed that their primary feature elements exhibit consistency. Grad-CAM visualizations can reveal the key feature elements of the product, infer features under different emotional classifications, and provide insights for optimization and innovation.

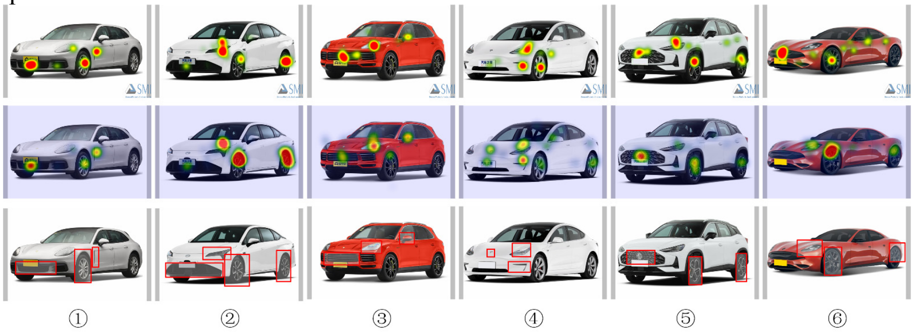

Fig. 3. Key Feature Elements Comparison

5 Conclusion

In order to explore whether emotion classification algorithms can represent key feature elements in the image classification process and provide visual explanations of the main product features in emotion classification, this study conducted two sets of experiments. The cognitive eye-tracking experiment was carried out to capture users' attention to product images under different emotional states, resulting in corresponding cognitive feature heatmaps. Additionally, based on the ResNet-34 classification algorithm and incorporating the Grad-CAM visualization layer, the main feature elements influencing the image classification process were obtained. By comparing the feature elements obtained from the two sets of experiments and analyzing their correlation, it was found that the visual features derived from the classification algorithm can represent the key feature elements of the product. The key feature elements obtained from emotion elicitation can provide new insights for design innovation.

Acknowledgments

This study was supported by the Guizhou Science and Technology Plan Project (no. ZK [2021]055, and the Guizhou University Cultivation Project (no. [2019]06). The corresponding author is Li Lin(715842042@qq.com).

Reference

1. Nguyen Ngoc H, Lasa G, Iriarte I (2022) Human-centred design in industry 4.0: case study review and opportunities for future research. Journal of Intelligent Manufacturing 33:35–76.
2. Zhang H, Han X, Li R, Qin S, Ding G, Yan K (2016) A new conceptual design method to support rapid and effective mapping from product design specification to concept design. Int J Adv Manuf Technol 87:2375–2389. doi: 10.1007/s00170-016-8576-6.
3. Guo F, Qu Q-X, Nagamachi M, Duffy VG (2020) A proposal of the event-related potential method to effectively identify kansei words for assessing product design features in kansei engineering research. International Journal of Industrial Ergonomics 76:102940. doi: 10.1016/j.ergon.2020.102940.
4. Wang J, Xu C, Zhang J, Zhong R (2022) Big data analytics for intelligent manufacturing systems: A review. Journal of Manufacturing Systems 62:738–752. doi: 10.1016/j.jmsy.2021.03.005.
5. Jiang P-T, Zhang C-B, Hou Q, Cheng M-M, Wei Y (2021) Layercam: Exploring hierarchical class activation maps for localization. IEEE Transactions on Image Processing 30:5875–5888.
6. Selvaraju RR, Das A, Vedantam R, Cogswell M, Parikh D, Batra D (2016) Grad-CAM: Why did you say that? arXiv preprint arXiv:161107450.
7. Takano H, Nagashima T, Nakamura K (2021) Characteristics of Visual Saliency Caused by Character Feature for Reconstruction of Saliency Map Model. Vision (Basel) 5. doi: 10.3390/vision5040049.
8. Kataoka N, Siritanawan P, Kotani K (2020) Integrating Human Visual Perception in Scene Understanding by Saliency Visual SLAM. In: 2020 59TH ANNUAL CONFERENCE OF THE SOCIETY OF INSTRUMENT AND CONTROL ENGINEERS OF JAPAN (SICE). IEEE, New York, pp 916–923.
9. Toet A (2011) Computational versus Psychophysical Bottom-Up Image Saliency: A Comparative Evaluation Study. IEEE Trans Pattern Anal Mach Intell 33:2131–2146. doi: 10.1109/TPAMI.2011.53.
10. Guo Z, Lin L, Liu B, Zhang L, Chang K, Li L, Jiang Z, Wu J (2023) Car Emotion Labeling Based on Color-SSL Semi-Supervised Learning Algorithm by Color Augmentation. Int J Intell Syst 2023:4331838. doi: 10.1155/2023/4331838.