Preprints.org

Article

Not peer-reviewed version

An Intelligent-Fusion Method of Product Core Feature Recognition Towards Design Patent Infringement Judgment

Siping Zeng , Lunjie Xiong , Wenguang Lin * , Renbin Xiao

Posted Date: 27 March 2026

doi: 10.20944/preprints202603.2216.v1

Keywords: core features; design patent; infringement judgment; attention mechanism; image similarity

Article

An Intelligent-Fusion Method of Product Core Feature Recognition Towards Design Patent Infringement Judgment

Siping Zeng [1], Lunjie Xiong [2], Wenguang Lin [2,*] and Renbin Xiao [3]

[1] School of Economics and Management, Xiamen University of Technology, Xiamen 361024, China
[2] School of Mechanical and Automotive Engineering, Xiamen University of Technology, Xiamen 361024, China
[3] School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China
* Correspondence: linwg@xmut.edu.cn

Abstract

The identification of core features plays an important role of design patent infringement judgment (DPIJ). The traditional method usually depends on the experience and knowledge of on-site engineers and the will of judges. There are obvious uncertainty and subjectivity, which is prone to occur miscarriage of justice and then hinders technological innovation. For overcoming the problem, this paper first proposes a fusion method for intelligent recognition of product core features, which combines resnet50 model with Convolutional Block Attention Module (CBAM) attention mechanism and cosine similarity algorithm. First, design patents of target products are downloaded and then the images of multi-views are extracted to build a database. Following this, Canny operator is used to extract the contour features of images for further fileting data. Second, cbam-resnet50 model is trained and optimized by transfer learning approach to realize the quantitative recognition of target products' salient feature regions. In addition, vgg16 model and cosine algorithm are integrated to calculate the similarity of contour features between target product and patent in database. Based on the two values, the target products' core features are determined by max score algorithm. The bathroom shower is taken as example to verify the effectiveness of proposed method. The results show that cbam-resnet50 model is superior to other models in terms of Intersection over Union (IOU) and Dice Similarity Coefficient (DSC) indexes, and the vgg16 model plus cosine similarity method has the highest discrimination for the contour feature. Moreover, the accuracy of the proposed method is higher than traditional method and even LLMs.

Keywords: core features; design patent; infringement judgment; attention mechanism; image similarity

---

1. Introduction

As market competition intensifies, enterprises increasingly prioritize appearance design to boost product appeal and capture consumer attention, often creating "hit products" [1,2]. However, this strategy also elevates the risk of imitation and intellectual property infringement. In judicial practice, design patent infringement disputes are the most prevalent type of adjudication. In China, for instance, such cases account for over 60% of total intellectual property disputes, with a significant portion hinging on the determination of similarity—specifically, whether "substantial differences" exist between designs[1].

Despite its importance, the standard for "substantial difference" has long been criticized for its subjectivity, as it relies heavily on judicial discretion. The lack of objective analytical tools often leads to inconsistent outcomes; statistics show that over 33.2% of design patent cases require re-judgment[1].

Thus, there is an urgent need to transition from qualitative discretion to a quantitative, data-driven auxiliary system. For the strong relationship between patent and R&D [3], quantitative patent analysis can be an effective means to minimize human bias. Building on this pursuit of objectivity, this study aims to transform the abstract legal concept of "substantial differences" into a specific, measurable task of core feature recognition.

According to the interpretations of the Supreme People's Court of China, "core features" of a design patent are defined by two dimensions: (1) parts easily viewed during normal use that impact the overall visual effect (high attention value), and (2) features distinct from existing designs (novelty)[2]. Although legal discourse emphasizes the necessity of distinguishing these aesthetic features [4,5], existing research rarely addresses how to identify and calculate them quantitatively. As artificial intelligence increasingly assists the legal profession [6], this paper proposes a core feature calculation framework that integrates deep learning and attention mechanisms with novelty evaluation to provide an objective auxiliary tool for design patent infringement judgment (DPIJ).

The main contributions of this study can be summarized as follows:

(1) *Problem-oriented contribution*: This work addresses the strong subjectivity inherent in existing DPIJ by providing an objective and effective identification tool for both judicial decision-making and enterprise R&D activities. The proposed approach helps reduce the risk of judicial misjudgment while simultaneously supporting appearance design avoidance strategies in product development.

(2) *Methodological contribution*: Focusing on the identification of core features in design patents, this study first conducts a systematic analysis and formal definition of core features based on relevant legal provisions. An attention mechanism and a novelty evaluation strategy are jointly introduced to construct a quantitative core feature computation framework, enabling integrated assessment of saliency and novelty.

(3) *Application-oriented contribution*: The research targets large-scale design patent data. Unlike conventional image datasets, design patent images emphasize contour information and multi-view representations, requiring feature identification from multiple perspectives. The proposed method is specifically designed to address these characteristics and is well suited for design patent analysis tasks.

The rest of the paper is organized as follows. In Section 2, the research framework is proposed. In Section 3, the research methods are introduced in detail. In Section 4, the methods are verified by shower design patent data. In Section 5, the content is discussed by judge case. In last section, the contribution, application and limitation are summarized.

2. Related Works

2.1. Research on Core Features of Design Patent

Different with utility patent, design patent is conceived to show the aesthetics rather than technical feature of design with combination, always including more than 6 images from different perspectives. So, the research on core features of design patent is a challenge task for a long time. Currently, extant literature on design patents is limited, primarily focusing on legal discussions regarding core features. For example, some studies emphasized the difference between design patents and inventions, which is not only reflected in ornamentally pleasing, but also reflected in distinctive [7]. In addition, the identification of appearance functionality and aesthetic characteristics of intelligent design works has been discussed from a judicial perspective [8]. Furthermore, analyses indicate that the novelty of existing design patents is essential to securing appearance licensing [9]. Moreover, it is widely held that design patents may be considered invalid if they lack novelty [4,5]. Recently, the Japanese Intellectual Property Office has allowed applicants to apply for design protection for works with unique shapes, patterns, or colors—or any combination thereof—that generate distinct visual artistic effects [10]. Obviously, those works mainly focused on the literal discussion of novelty, and rarely on the attention of design, and don't refer to the quantitative identification moreover.

2.2. Research on the Mechanism of Image Attention

The attention mechanism is a mechanism that imitates the human eye, selectively focusing on part of the given information and ignoring other information. The concept was first proposed for machine-translation tasks, allowing the model to dynamically focus on different parts of the input sentence and assign different weights, rather than relying solely on the last hidden state [11]. Then, an image attention model is proposed to learn and describe image content automatically [12]. Subsequently, the transformer model was proposed by Google, which is entirely based on the attention mechanism rather than RNN or CNN neural network architectures [13]. In addition, a Residual Attention Network was conceived by integrating attention mechanisms into convolutional neural networks [14].

Obviously, with the development of deep learning technology, it has become a common practice to integrate convolutional neural networks with attention mechanisms. This fusion strategy not only improves the sensitivity of the model to the key features of input data, but also enhances the learning ability of small sample data, and has been widely used in many industries. For instance, CNN was integrated with an attention mechanism, incorporating lateral inhibition and contrast suppression strategies to improve the detection efficiency of X-ray images in security screening [15]. Moreover, these two algorithms were combined to accurately distinguish between blurred and irregular land–sea boundaries [16]. In the same year, a Transformer model based on multi-head attention—which computes feature maps in parallel—alongside a convolutional neural network was employed for plant leaf disease recognition [17]. A semantic segmentation network built upon an existing CNN architecture—enhanced with a dual-attention module was proposed to achieve high-precision parsing of urban street scene semantics [18]. The lightweight CNN architecture ShuffleNetV2 was integrated into the YOLOv5 framework and incorporated a CBAM, significantly improving the detection capability for casting surface defects [19]. Channel attention and spatial attention mechanisms were combined with CNN algorithms to accomplish land cover feature recognition in satellite remote sensing images [20]. A hybrid CNN-integrated framework including cross-attention vision Transformer was constructed for accurately identifying and highlighting subtle differences in signatures [21]. A precise medical image segmentation is achieved by combining a pyramid pooling module with an attention mechanism [22]. In addition, a spatiotemporal convolutional neural network is fused with a Transformer-based multi-head attention mechanism to enhance the accuracy of fall risk prediction in elderly individuals [23].

In summary, current attention mechanism modules integrated with CNN are primarily applied in fields such as industry, agriculture, and healthcare, with relatively little involvement in patent image analysis. Unlike images from those fields, patent appearance images typically consist of geometric shapes and lines, and require comparison across multiple perspectives of the same design to identify and evaluate feature differences. Although a cross-attention mechanisms combined with CNN-LSTM algorithms was utilized for patent classification [24], the patent is treated as a whole, failing to adequately account for the distinctiveness of different design features within patent images.

2.3. Research on Patent Image Similarity

Image similarity is the key of patent multimodal research, which has received great attention from academic. For example, traditional geometric feature extraction method was applied on patent retrieval [25]. A patent flowchart recognition approach was proposed to enable structured visualization of flowchart images and support semantic queries [26]. Local features were extracted from patent drawings through superconducting wire curve recognition [27]. Patent textual descriptions, such as reference numbers, was used to identify features in patent drawings and applied the method within the USPTO system [28]. An improved cosine similarity measure combined with patent ownership probability is employed to perform patent classification [29]. Contextual local primitives were utilized for binary patent image retrieval [30]. A transfer learning approach utilizing a pre-trained VGG19 model was introduced for digital intellectual property protection [31]. Design feature vectors of patent images using convolutional neural networks was derived for visual material-type prediction and international patent classification section-label predictions [32]. Features were extracted from patent drawings using deep convolutional neural networks, incorporating a region proposal network (RPN) to improve recognition accuracy and reduce feature omission [33]. The YOLOv4 model was employed to crop images and subsequently applied a triplet convolutional neural network for trademark similarity analysis [34]. A multi-fusion deep learning framework capable of learning bimodal features from both text and images was proposed for patent retrieval [35]. Similarly, a patent knowledge graph (PKG)–driven patent similarity computation model that integrates graph similarity (GS) and image similarity (IS) was introduced [36]. A novel multimodal patent similarity evaluation framework that integrates artificial intelligence techniques with conceptual analysis of patent drawings was presented for patent classification [37]. A multimodal feature fusion approach was proposed to extract patent text and image features, and was integrated with an attention mechanism for design patent classification [38]. A hierarchical positive contrastive loss combined with the LIC classification method was introduced for patent image retrieval [39]. Structural similarity index (SSIM) was utilized to measure the similarity of U.S. design patent images [40]. Similarly, a density-based patent image feature extraction algorithm was proposed [41].

Although existing studies have made notable progress in patent image feature extraction and similarity analysis, several limitations remain. First, many approaches rely on geometric descriptors or binary image representations, which perform well for specific image types such as flowcharts or structural diagrams but exhibit limited capability in representing complex product appearances, making it difficult to accurately capture fine-grained design differences. Second, despite the substantial performance gains achieved by deep learning–based methods in recent years, most studies focus on single-view or whole-image feature representations, insufficiently accounting for multi-view shape variations of products and underestimating the critical role of local design elements in similarity assessment. Moreover, although some works enhance retrieval efficiency by integrating textual and visual information within multimodal frameworks [42], their primary emphasis lies in improving overall patent retrieval or classification performance, rather than directly addressing the specific requirements of core feature identification in design patent infringement determination.

Based on the above literature review, two key research gaps can be identified. First, there is a lack of systematic and quantitative analytical frameworks specifically designed for design patent infringement decision-making. Second, existing image recognition algorithms—regardless of whether they are based on attention mechanisms or similarity computation—remain difficult to directly adapt to the requirements of design patent analysis, where legally relevant visual features and fine-grained design distinctions must be explicitly captured and interpreted.

Against this background, this study addresses the intelligent decision-making requirements of design patent infringement analysis by proposing a core feature identification approach that simultaneously accounts for visual attention and design novelty, grounded in a systematic definition of core design features. Patent contour views are adopted as the primary research object, where contour extraction is used to suppress the influence of color, material, and scale variations, thereby enabling visual representations to focus more explicitly on the underlying shape and structural characteristics that are central to design patent infringement assessment. To overcome the information insufficiency inherent in single-view analysis for complex product designs, similarity is computed independently across multiple contour views and subsequently fused to achieve a more comprehensive similarity evaluation. At the feature level, deep learning models are employed to extract high-level contour features, which are combined with cosine similarity–based local feature comparison to enhance sensitivity to fine-grained design differences. This framework provides a more stable and interpretable quantitative basis for subsequent core feature aggregation and infringement decision support.

3. Research Framework

In this paper, the identification framework of the core characteristics of appearance products for intelligent judge is proposed, which integrates attention mechanism and patent similarity calculation. As shown in Figure 1, the approach comprises four main stages: Design patent data collection and processing, image model training and attention score calculation, product image feature similarity calculation and Product core feature calculation and identification.

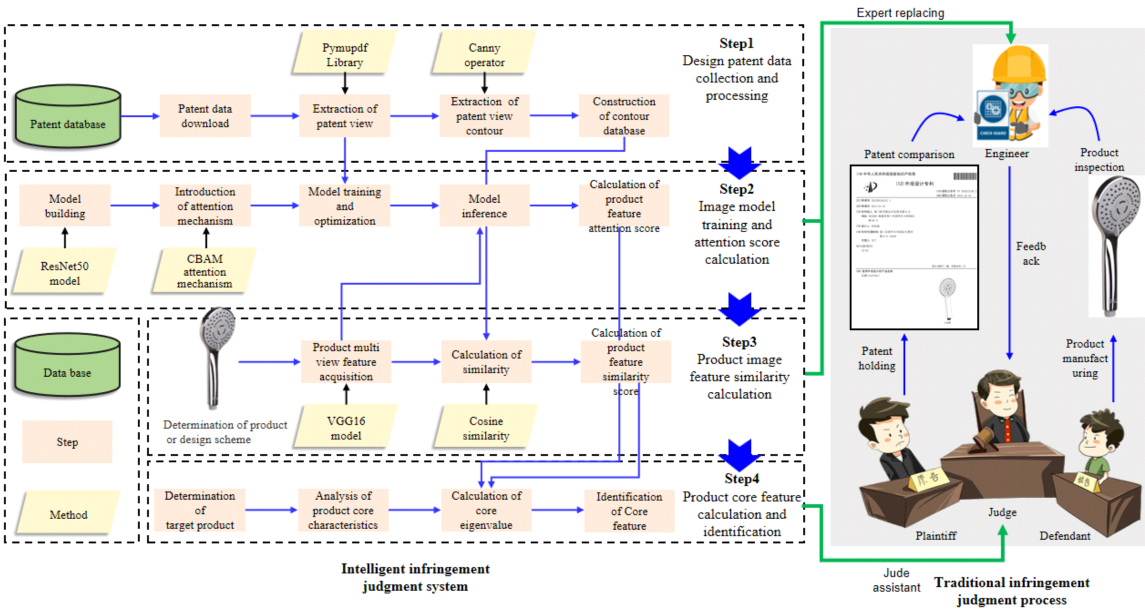


Figure 1. Research framework.

The specific implementation contents are introduced as follows:

Step 1*: Design patent data collection and processing.* To address the innovation requirements of a specific product category, the retrieval scope of relevant design patents is first defined. Design patent data are then collected from patent retrieval platforms. Based on this, multi-view design patent drawings are programmatically extracted using the PyMuPDF library to construct a structured image dataset. Subsequently, the Canny edge detection algorithm is applied to the patent drawings to extract contour information, which serves as a foundational representation for subsequent attention-based feature visualization and similarity computation.

Step 2*: Image model training and attention score calculation.* A ResNet50 architecture integrated with a CBAM is constructed and fine-tuned using a transfer learning strategy. The dataset extracted in Step 1 is divided into training and validation sets, which are used for parameter optimization and performance monitoring, respectively. During model training, cross-entropy loss is adopted as the objective function, and the Adam optimizer is employed to accelerate convergence. Dynamic learning rate scheduling and weight decay are further introduced to mitigate overfitting. After the model achieves the desired performance, multi-view images of the target product are fed into the trained network to obtain the corresponding design feature weight distributions, from which the salience scores of individual design features are calculated.

Step 3*: Product image feature similarity calculation.* For product feature similarity computation, a vector-space–based approach integrating deep feature extraction and distance measurement is developed. The design features of multi-view contour images of the target product are fed into a pre-trained VGG16 model, and the output of the penultimate fully connected (FC) layer is extracted as the feature vector representation. Subsequently, the cosine similarity metric is introduced. By traversing the database of product design patent contour images, the cosine of the angle between the feature vector of the target product and those of the database samples is computed. This value serves as the similarity measure for the corresponding feature, enabling the identification of the most similar sample within the database.

Step 4*: Product core feature calculation and identification.* The core feature score, denoted as Core, is defined as the maximum value of the ratio between the attention score and the maximum similarity score of a given product feature across different views. Based on the feature attention scores obtained in Step 2 and the feature similarity scores derived in Step 3, the Core value of each feature is calculated under multiple views of the target product. Ultimately, for a specific product view, the feature with the highest Core value is identified, and this feature is determined as the core feature of the product.

4. Method

4.1. Product Design Patent Image Extraction Based on PyMuPDF Library

Design patent downloaded from the patent office website, such as CNIPA, USPTO, and EPO, etc., is usually stored in PDF format, which not only contains a small amount of text descriptions, but also embeds many visual elements such as design drawings and schematic diagrams. In order to analyze the design patent, it is necessary to extract the image content from the PDF document. Compared with other extractors, PyMuPDF library has the obvious advantages of high speed, high coordinate accuracy and strong editing ability [43], so is introduced to mining the image. The pseudo code of extracting process is shown as Table 1.

Table 1. The pseudo code of product appearance design image extraction.

```
Input: PDF folder P, output folder O, target page index p
Output: Extracted images saved in O
For each pdf_i in P do
     Open pdf_i as document D
// fitz.open()
     If p ≥ number of pages in D then
        Continue to next pdf_i
     End If
     Load page p from D as page_p
// load_page()
     Obtain all images I = {I_1, I_2, ..., I_n} from page_p
// get_images(full=True)
    If I is empty then
        Close D
        Continue to next pdf_i
    End If
    Create subfolder O_i according to pdf_i name
    For each image I_j in I do
        Obtain image reference xref_j
        Extract image bytes b_j from D using xref_j
// extract_image()
         Decode b_j to image matrix M_j
         Save M_j to O_i with filename "page(p+1)_image(j+1).png"
// cv2.imdecode()
```

```
        End For
        Close D
    End For
    Return extracted image set
```

The process is including four steps:

Step 1. *Opening PDF document:* fitz.open() tool in PyMuPDF is used to load the PDF file, then the document object is obtained, and the number of pages in the document is checked to ensure whether the target page is existing.

Step 2. *Extracting images from pages:* The approach traverses all pages, and uses load_page() tool to load content of page, then calls get_images(full=True) to obtain all embedded image information in the page, including the image's xref identifier, size, color space, etc.

Step 3. *Extracting data from image:* For each image, the extract_image () tool is utilized to extract the data, including byte stream and other related information stored in the PDF image through xref label.

Step 4. *Image decoding and saving:* The extracted byte data is converted into an array via NumPy, then OpenCV's cv2.imdecode() function is used to decode array into an image matrix, which are saved in PNG format for subsequent operations such as contour extraction and model training.

4.2. Contour Extraction Algorithm

To enable subsequent feature region visualization and similarity computation, contour extraction is first applied to the design patent images. Commonly used contour extraction algorithms include Sobel operator, Laplacian operator, Roberts operator, and Canny edge detector [44].

With respect to the Sobel operator, although it is computationally efficient, its reliance on first-order derivatives often leads to fragmented edge extraction results, limiting contour continuity [45]. The Laplacian operator, as a second-order derivative method, enhances fine details but is excessively sensitive to image noise, making it unsuitable for applications requiring high contour integrity [44]. The Roberts operator, constrained by a small local neighborhood, exhibits weak noise robustness and limited capability in capturing continuous contours of complex structures [46]. In contrast, the Canny edge detector effectively suppresses noise through Gaussian filtering, accurately preserves fine details via non-maximum suppression, and maintains contour completeness using dual-threshold detection. As a result, it is widely recognized as one of the most effective edge detection methods [47].

This paper adopts the Canny operator for product appearance contour extraction. The progress is divided into the following 4 steps:

Step 1. *Grayscale conversion:* The input image is converted into a grayscale representation to produce a black-and-white appearance, thereby simplifying subsequent processing steps. The grayscale transformation is defined as follows:

$$Y = 0.299 \times R + 0.587 \times G + 0.114 \times B \tag{1}$$

where R, G, and B denote the pixel values of the red, green, and blue color channels, respectively, and Y represents the resulting grayscale intensity. The weighting coefficients are adopted from the ITU-R BT.601 standard [48], which accounts for the differences in human visual sensitivity to various color components.

Step 2. *Noise reduction:* Gaussian filtering is applied to smooth the image and reduce the influence of noise. The filtering operation is calculated as follows:

$$G(x,y) = \frac{1}{2\pi\sigma^2} exp\,(-\frac{x^2+y^2}{2\sigma^2}) \tag{2}$$

where x and y denote the spatial coordinates, σ represents the standard deviation of the Gaussian distribution, and $exp(\cdot)$ denotes the exponential function.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 March 2026 doi:10.20944/preprints202603.2216.v1

Step 3. *Edge detection:* The rate and direction of intensity variation in the image are first measured using gradient operators. The horizontal gradient is defined as $G_x = \partial I / \partial x$, and the vertical gradient is defined as $G_y = \partial I / \partial y$. Based on these gradients, the gradient magnitude is calculated as follows:

$$G = \sqrt{G_X^2 + G_y^2} \tag{3}$$

Meanwhile, the gradient direction is calculated as follows:

$$\theta = arctan\,(\frac{G_y}{G_x}) \tag{4}$$

Furthermore, non-maximum suppression is introduced to remove pixels that do not belong to edge regions, thereby retaining only thin and precise edge contours. Since computing the gradient direction of all pixels alone is insufficient to fully determine edge information, non-maximum values along the local gradient direction must be suppressed, and only pixels with locally maximal gradient responses are preserved.

Step 4. *Double threshold detection and edge connection:* To achieve precise edge selection, two thresholds are defined: a high threshold and a low threshold. Pixels are classified based on these thresholds into strong edges, weak edges, and non-edges: strong edges correspond to pixels with values above the high threshold, weak edges correspond to pixels between the low and high thresholds, and non-edges correspond to pixels below the low threshold. Strong edges are then connected to adjacent weak edges, and both are retained to generate the final edge detection result.

Lastly, based on the above description, the detailed procedural steps of the algorithm can be summarized in the following pseudocode.

Table 2. The pseudo code of canny operator contour detection.

```
Input: Image folder F, output folder O
Output: Edge-detected images saved in O
If O does not exist then
    Create folder O
End If
Obtain image set I = {I_1, I_2, ..., I_n} from F
// Images with extensions {png, jpg, jpeg}
For each image I_k in I do
    Read I_k as image M_k
    If M_k is invalid then
        Continue to next image
    End If
    Convert M_k to grayscale image G_k
    Apply Gaussian filter on G_k to obtain B_k
    // Gaussian kernel size is (5,5)
    Apply Canny edge detector on B_k to obtain edge map E_k
    // Low threshold and high threshold are determined
    Save E_k to O with filename "edges_I_k"
End For
Return edge-detected image set
```

4.3. Convolutional Neural Network Model with Attention Mechanism

4.3.1. Backbone Network: RESNET Model

Traditional CNN extract features by stacking multiple convolutional and down-sampling layers. However, as network depth increases, gradients can vanish or explode during backpropagation, leading to performance degradation. To address this issue, the ResNet model introduces batch normalization (BN) layers and employs a residual learning strategy [49]. By directly connecting a portion of the input to subsequent layers, the network learns the residual mapping between input and output rather than the full mapping function. This modification reduces the learning difficulty and effectively mitigates the degradation problem commonly observed in deep networks.

Considering that ResNet50 employs bottleneck residual blocks to significantly reduce parameter size and computational cost while preserving strong representational capability, it is adopted as the backbone network in this study. The overall architecture of the model is illustrated in Figure 2.

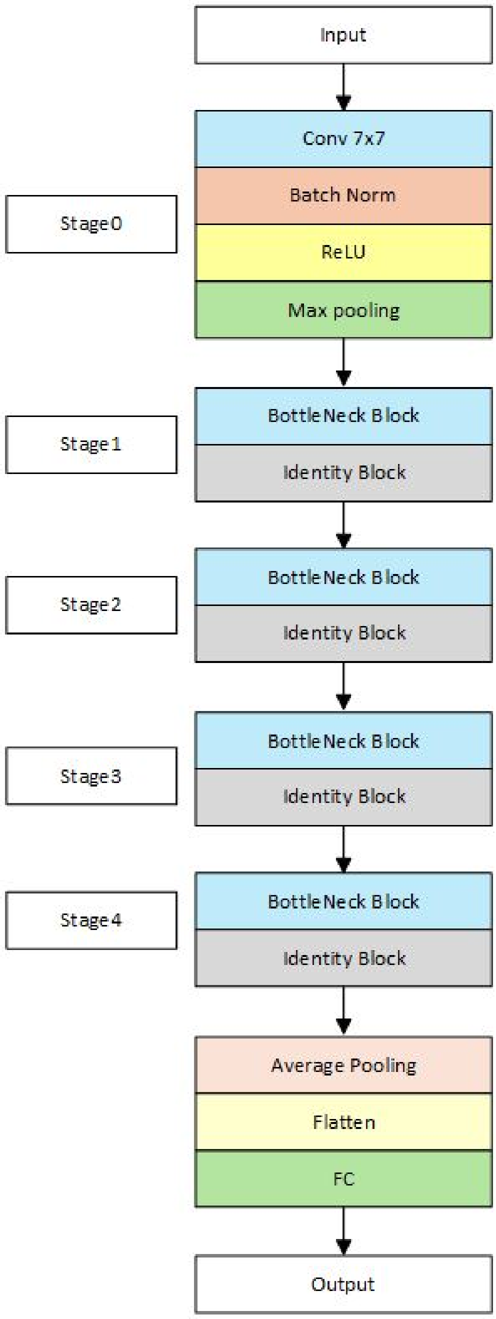


Figure 2. ResNet50 model architecture.

doi:10.20944/preprints202603.2216.v1

In the initial perception stage—Stage 0, a 7×7 convolution is applied in conjunction with BN and Rectified Linear Unit (ReLU) activation to capture low-level geometric features and ensure stable signal propagation in deeper layers. This is followed by a max pooling operation, which performs spatial down-sampling while retaining salient features.

In the feature mapping stage, deep representations are learned through four residual groups—Stage 1–Stage 4. Bottleneck residual blocks are responsible for flexible adjustment of feature map dimensions and down-sampling, whereas identity blocks propagate information through skip connections, allowing the network to focus on residual learning and effectively alleviating the degradation problem inherent in deep networks.

Finally, global average pooling aggregates channel-wise information, a flatten layer converts the multidimensional feature tensor into a one-dimensional high-dimensional vector, and a FC layer performs feature integration and mapping.

The core building block of ResNet50 is the bottleneck residual block, as illustrated in Figure 3. When processing large-scale design patent image datasets, this structure exhibits notable performance advantages. Specifically, a 1×1 convolution is first applied to reduce the dimensionality of the input feature map in stride 1, followed by a 3×3 convolution for feature extraction in stride 2, and finally another 1×1 convolution is used to restore the feature dimension in stride 3. This design preserves the capability to represent complex product details while substantially reducing the number of parameters and computational cost. Moreover, when the spatial resolution or channel dimensionality of the input and output does not match, the shortcut branch employs a 1×1 convolution for down-sampling and alignment, ensuring smooth propagation of residual information. Compared with basic residual structures, the bottleneck design in ResNet50 significantly improves training speed and inference efficiency on large-scale patent image datasets, providing a solid foundation for subsequent extraction of salient design features and enabling more efficient identification of visually critical product characteristics.

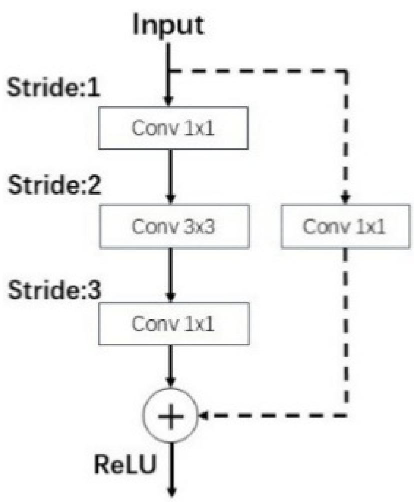


Figure 3. Residual module of ResNet50.

4.3.2. CBAM Attention Mechanism

In convolutional neural network (CNN) models, hierarchical feature extraction is achieved through successive convolution and pooling operations, capturing both low-level edge information and high-level semantic representations from patent drawings. However, conventional convolution operators rely on fixed-weight summation, which limits their ability to adaptively focus on regions that reflect visually salient product features, particularly in the presence of complex product. Given that different design features contribute unevenly to visual saliency across large-scale design patent datasets, the introduction of attention mechanisms is of significant importance. The core principle of attention mechanisms is to enable the network to automatically learn feature importance, enhancing informative components while suppressing irrelevant ones. In the computer vision domain, commonly used attention mechanisms include channel attention, spatial attention, temporal attention, and hybrid attention [50]. In this study, the CBAM is adopted, which is a lightweight hybrid attention mechanism composed of channel attention and spatial attention module [51]. Channel attention and spatial attention operate along the channel and spatial dimensions of the feature map, respectively, guiding the model to focus on salient features within target regions. The overall architecture of the CBAM attention mechanism is illustrated in Figure 4.

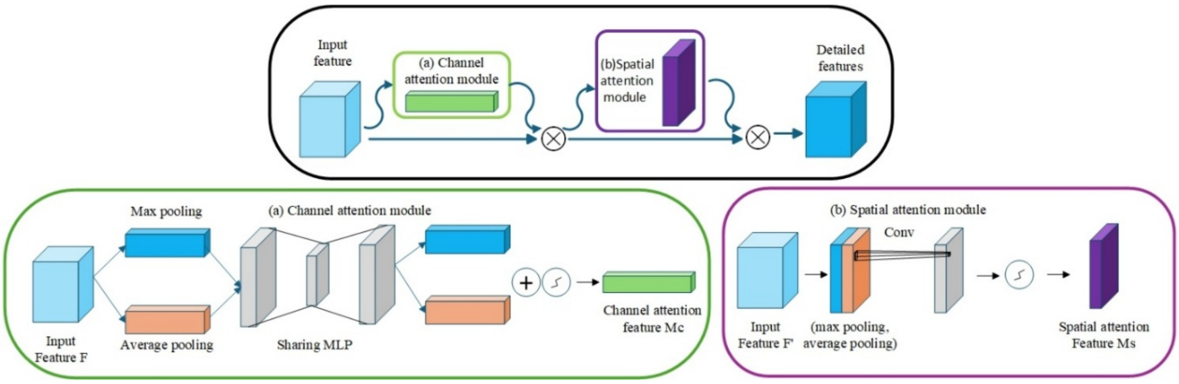


Figure 4. CBAM architecture.

As illustrated in Figure 4(a), the function of channel attention module is to adaptively assign different weights to each channel of the input characteristic graph, so as to enhance the response of the key characteristic channel and suppress the secondary channel. The principle is to perform global average pooling and global maximum pooling on the input characteristic graph, and generate two channel descriptors with the shape of $1 \times 1 \times C$. Then they are processed by a multi-layer perceptron (MLP), activated by the Sigmoid function, and finally fused to generate the channel attention weight matrix M_c. It is calculated as follows:

$$F_{avg} = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F(i,j) \tag{5}$$

$$F_{max} = max\, F(i,j) \tag{6}$$

$$M_c(F) = \sigma(W_2 ReLU(W_1 F_{avg}) + W_2 ReLU(W_1 F_{max})) \tag{7}$$

where F_{avg} and F_{max} denote the features obtained via global average pooling and global max pooling, respectively. Both descriptors are fed into a shared two- MLP. The first layer performs dimensionality reduction with C/r neurons, where r is the reduction ratio, followed by a ReLU activation function, while the second layer restores the feature dimension to the original number of channels. The outputs of the two branches are then summed and passed through a sigmoid activation function to generate the channel attention weight matrix M_c. This process enables the network to learn the relative importance of different channels. Finally, the input feature map F is multiplied by the attention weights to obtain the enhanced feature representation F'.

As illustrated in Figure 4(b), the spatial attention module further emphasizes informative spatial locations based on the channel-refined feature map F'. Specifically, global average pooling and global max pooling are applied to F' along the channel dimension, generating two two-dimensional feature maps that capture spatial saliency information. These two maps are then concatenated along the channel axis and fed into a standard convolutional layer followed by a sigmoid activation function. Through this process, the spatial attention weight matrix $M_s(F')$ is obtained, which highlights discriminative regions while suppressing less relevant spatial information. It is calculated as follows:

$$M_s(F') = \sigma\left(f^{7*7}([F'_{avg}, F'_{max}])\right) \tag{8}$$

where F' denotes the feature map refined by the channel attention module. F'_{avg} and F'_{max} represent the features obtained by applying global average pooling and global max pooling to F', respectively. $[F'_{\mathrm{avg}}, F'_{\mathrm{max}}]$ indicates the concatenation operation along the channel dimension. $f^{7\times7}$ denotes a convolutional layer with a kernel size of 7×7, and $\sigma(\cdot)$ denotes the sigmoid activation function. Through this process, the spatial attention weight map $M_s(F')$ is generated, which enhances informative regions while suppressing less relevant spatial areas.

In summary, the CBAM attention mechanism enables the network to adaptively learn both which features are important and where these important features are located by jointly modeling channel attention and spatial attention. Along the channel dimension, differentiated weights are assigned to individual feature channels, thereby enhancing the representation of critical semantic information. Along the spatial dimension, the model is further guided to focus on locally discriminative regions within the image. In addition, visualization of the learned attention weights provides an intuitive interpretation of the salient regions and structural characteristics emphasized during feature extraction, offering interpretable evidence for the identification and analysis of core design features in design patents.

4.3.3. RESNET Model Framework and Residual Module Integrated with CBAM

ResNet50 is adopted as the baseline network in this study due to its strong capability in extracting product appearance features while effectively mitigating gradient vanishing and exploding problems. On this basis, the CBAM module is integrated to further enhance feature representation through a dual-attention mechanism. In the channel attention branch, global average pooling and global max pooling are employed to capture channel-wise statistical information, and a multilayer perceptron is used to learn adaptive channel weights, thereby emphasizing informative features. In the spatial attention branch, channel-refined features are compressed and convolved to accurately localize visually salient regions in product drawings, enabling the model to focus on discriminative appearance characteristics.

In this study, a CBAM attention mechanism is integrated into the ResNet50 network to enhance product image feature learning. By incorporating attention-weighted feature maps and visualizing the learned attention responses, the model enables an intuitive analysis of the image regions that contribute most to its decision-making process. The overall architecture of the ResNet50 model with the embedded CBAM module is illustrated in Figure 5, while the modified residual block after integrating the CBAM attention mechanism is shown in Figure 6 [52].

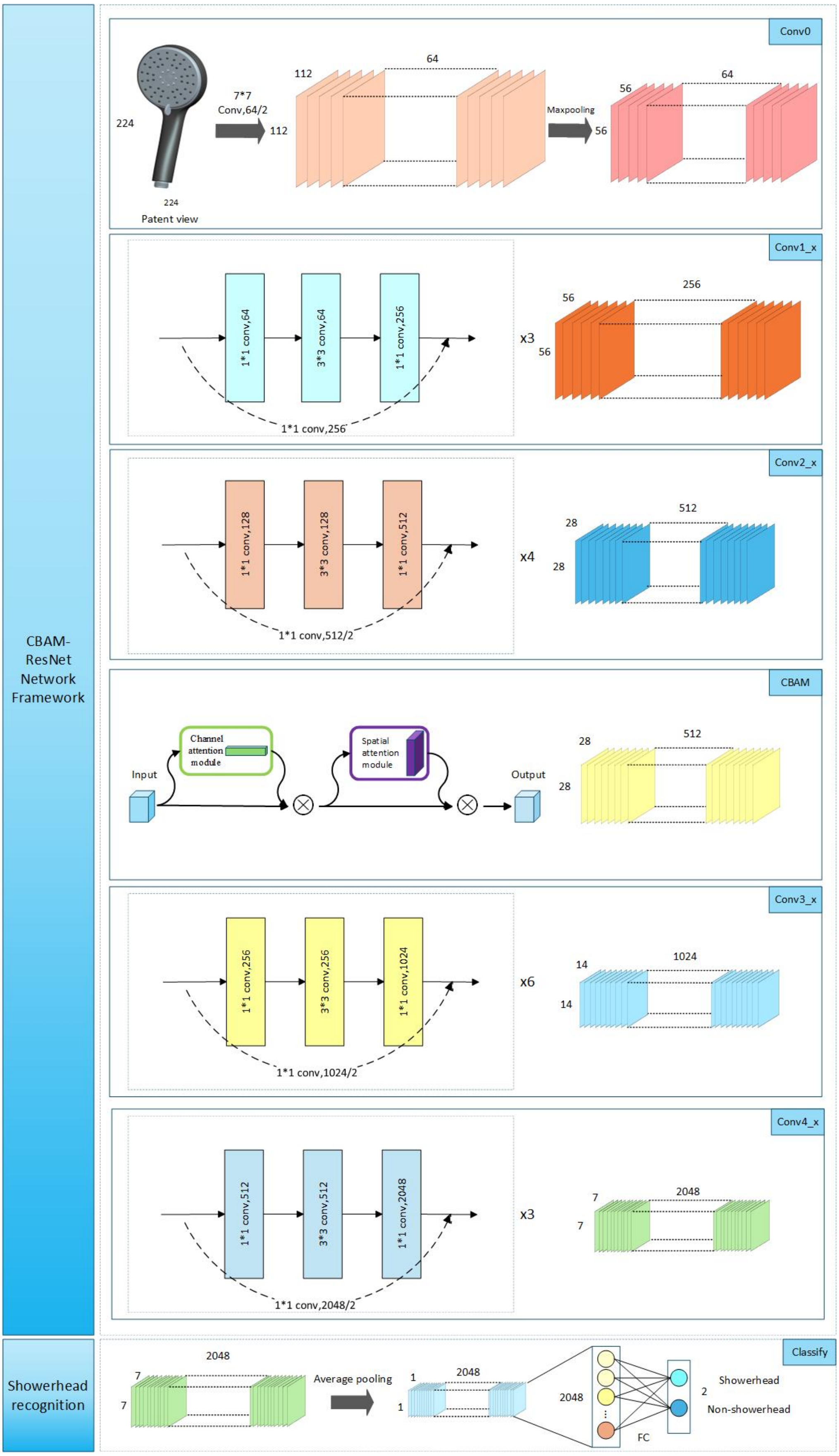


Figure 5. CBAM-ResNet model framework.

 doi:10.20944/preprints202603.2216.v1

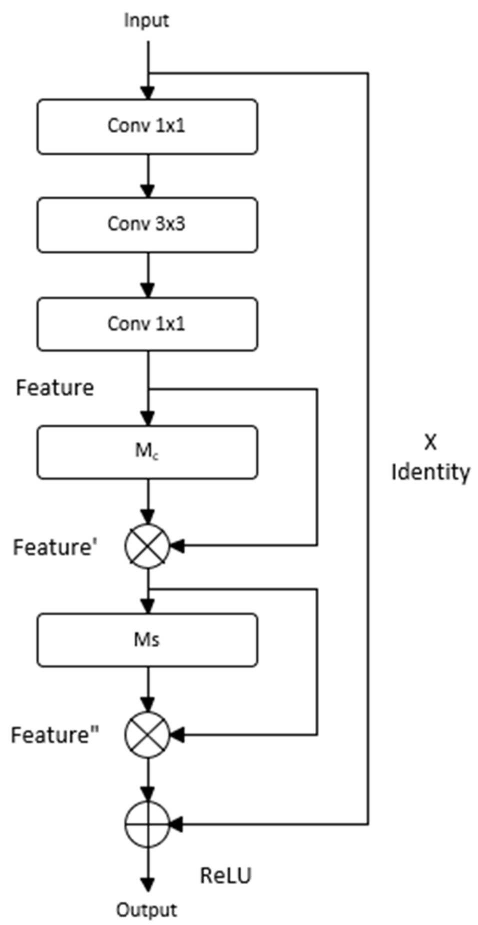


Figure 6. Residual module with CBAM attention mechanism inserted.

4.3.4. Attention Feature Visualization and Score Calculation

To improve the interpretability and intuitiveness of the experimental results, the attention maps produced by the model are visualized. By visualizing these attention maps, salient regions within the product images can be explicitly identified, enabling an assessment of whether the model effectively focuses on visually significant design features. The overall visualization procedure can be summarized in the following steps:

Step 1. *Preprocessing of input images:* Since product images vary in size, they are first resized to 224×224 for meeting the input requirements of the CBAM-ResNet50 model.

Step 2. *Extraction of attention information:* The preprocessed images are fed into the trained model to extract attention information. The feature maps output by the CBAM module—already weighted by both channel and spatial attention—are selected. These feature maps are then averaged along the channel dimension to generate the attention weight maps.

Step 3. *Resizing and normalization of attention maps:* The extracted attention maps are resized to match the dimensions of the contours detected by the Canny operator, normalized to the range [0,1], and stored as NumPy arrays for subsequent visualization.

Step 4. *Visualization on Canny edges:* The attention weights are overlaid on the Canny-detected edges, highlighting the salient regions for intuitive visual interpretation of the model's focus areas.

Considering that the model's attention may not be uniformly distributed across a given feature, the mean value of the attention weights for a specific feature is defined as the feature's attention score. This metric reflects the degree to which the model focuses on each individual product feature and provides a quantitative basis for the subsequent definition of core features. The computation can be expressed as follows:

$$F^{'} = M_c(F) \odot F \tag{9}$$

$$M(i,j) = M_s(F^{'}) \odot F^{'} \tag{10}$$

$$Attention\ Score = \frac{\sum_{i=1}^{H}\sum_{j=1}^{W} M(i,j) \cdot Mask(i,j)}{\sum_{i=1}^{H}\sum_{j=1}^{W} Mask(i,j) + \varepsilon} \tag{11}$$

where $\odot$ denotes the Hadamard product, representing element-wise multiplication of matrices; F denotes the input feature map to the CBAM module; F' denotes the feature map after applying channel attention weights; $M(i,j)$ represents the attention weight at spatial location (i,j); $\text{Mask}(i,j)$ denotes a binary indicator function for the target feature region, taking a value of 1 if pixel (i,j)belongs to the selected feature region, and 0 otherwise; $\sum_{i=1}^{H}\sum_{j=1}^{W}\text{Mask}(i,j)$denotes the total number of pixels within the masked region; $H$and $W$are the height and width of the image, respectively; and ε denotes a small constant added to prevent division by zero.

4.4. Similarity Algorithm

Common approaches for measuring image similarity include color histogram-based methods, structural similarity, feature extraction and matching, and deep learning-based methods. Color histogram-based methods evaluate similarity by statistically analyzing the color distribution of images. However, in the context of product design patents, color is typically not protected, which limits the applicability of this approach. Structural similarity algorithms (SSIM) assess image similarity across luminance, contrast, and structural dimensions, but in the case of contour images, significant variations in brightness can substantially interfere with matching accuracy. Methods based on feature extraction and matching rely on detecting and matching key points; however, for contour images, feature point extraction is often unstable, resulting in large fluctuations in matching performance [53].

The feature representation of product design patent contour images primarily relies on the geometric relationships of lines and lacks visual cues such as texture or color. Conventional methods struggle to ensure stable feature extraction for such data. In contrast, deep learning-based approaches leverage their powerful feature extraction capabilities to capture the morphological characteristics of contour images, while also providing scale invariance, effectively addressing the limitations of traditional algorithms in processing design patent data.

In this study, the pretrained VGG16 model is used to extract feature vectors from contour images [54], with the output of the second-to-last FC layer serving as the feature representation, and architecture is shown in Figure 7. Subsequently, cosine similarity is employed to measure the similarity between feature vectors. The core idea of cosine similarity is to evaluate the cosine of the angle between two vectors, under the assumption that the closer the directions of two vectors, the more similar the image content they represent. Assuming there are two vectors $\vec{A} = (A_1, A_2, \dots\dots, A_n)$ and $\vec{B} = (B_1, B_2, \dots\dots, B_n)$, the cosine similarity calculation formula is computed as follows:

$$cos(\theta) = \frac{\vec{A} \cdot \vec{B}}{\|A\| \cdot \|B\|} = \frac{\sum_{i=1}^{n} A_i \cdot B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \cdot \sqrt{\sum_{i=1}^{n} B_i^2}} \tag{12}$$

where $\vec{A} \cdot \vec{B}$ denotes the dot product of the two vectors, and $\|\mathbf{A}\|$ and $\|\mathbf{B}\|$ represent the magnitudes of the vectors, respectively. The resulting cosine similarity score ranges between -1 and 1. A value closer to 1indicates higher similarity, whereas a value closer to -1 indicates that the vectors are more dissimilar.

 doi:10.20944/preprints202603.2216.v1

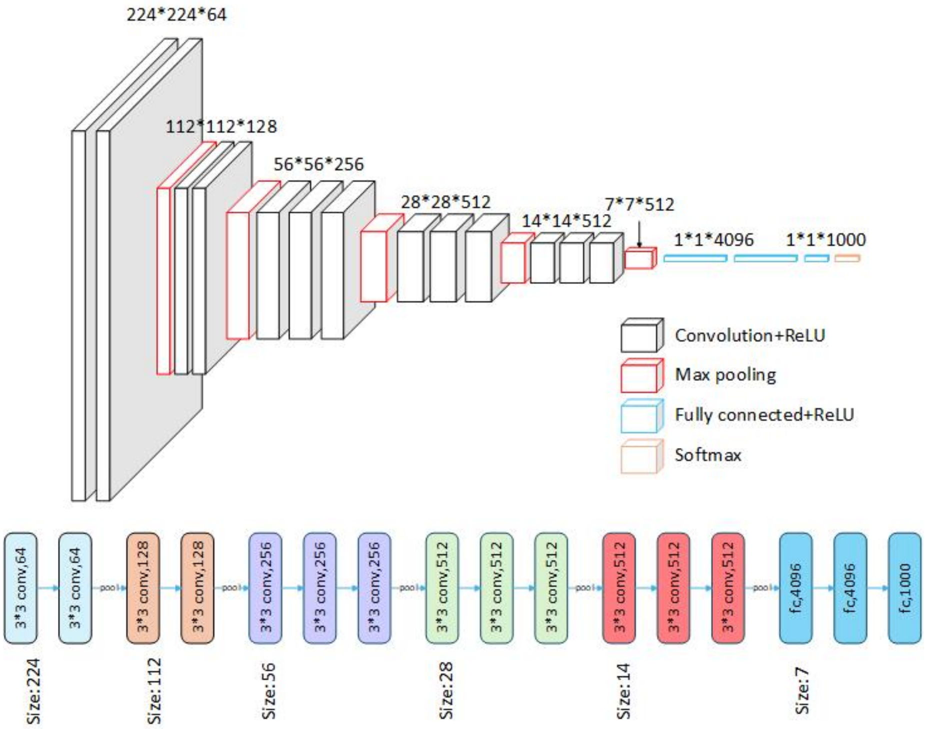


Figure 7. VGG16 model architecture.

4.5. Core Feature Identification for Product Appearance

It is well recognized that, although some design patents may include axial views, a product's design is typically represented across six standard perspectives: front, back, left, right, top, and bottom. In practice, the novelty and attention scores of the same feature can vary across different views. For instance, the faceplate of a shower may constitute a core feature in the front view but not in the side view, a situation that is also observed in other products. Accordingly, in the context of patent infringement evaluation, the core features of an allegedly infringing product should be assessed against the patent's multiple views, i.e., the target feature should be identified with reference to all six views.

Based on the definition of core features, and by leveraging the attention maps and scores obtained in Sections 4.3 and 4.4 along with the similarity maps and values, a method for identifying the core appearance features of a product is developed. To this end, a core feature score algorithm is introduced, with the computation formalized as follows:

$$Core = \max_{v \in V} \frac{Attention\ score[v(v_1 、 v_2 \ldots \ldots v_m)]}{most\ similar\ value[D(D_1 、 D_2 \ldots \ldots D_n)]} \tag{13}$$

where $Core$ denotes the importance of different product features, a higher value indicates higher feature importance; v denotes a specific view of the product; V represents the set of product views; $v_1, v_2 \ldots \ldots v_m$ denote the m features in product view v; D denotes the product design patent database; D_1、 $D_2 \ldots \ldots D_n$ denote the patent views of products in the product appearance database; $\max_{v \in V}$ denotes the maximum importance value of the same feature across different views.

It is evident that the *Core* score is positively correlated with the attention value and inversely correlated with feature similarity. Based on this formulation, judges and engineering professionals can perform an intelligent, quantitative identification of core design features for design patent infringement evaluation, rather than relying solely on traditional experience-based methods.

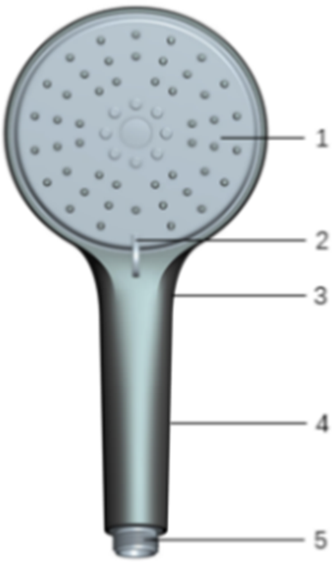
 doi:10.20944/preprints202603.2216.v1

5. Case Study

5.1. Case Overview

Showerheads are well-known export products in China. According to statistics, over 50% of shower products globally are manufactured in China, forming a huge industry that attracts numerous enterprises to participate. Meanwhile, showerheads, especially handle-shower are typical industrial products that contain both technological and appearance design elements. To gain higher market share and save manufacturing costs, enterprises often tend to attract user attention through appearance modifications and protect them with design patents. In China alone, there have been 6167 shower design patents in the past 15 years. These patents together form a tight patent protection network, increasing the risk of patent infringement for competitors. According to statistics, since 2014, there have been as many as 425 shower cases among Chinese design patent infringement disputes, accounting for 1.6% of the total number of infringements, ranking second in infringement cases among all products, and showing an increasing trend year by year[3]. In fact, shower products are composed of multiple mechanical and electronic components, belonging to representative products in the traditional electromechanical field. Moreover, shower design patents involve a large number of complex detailed features. In traditional DPIJ, relying on manual determination is often time-consuming and prone to deviations, with significant judgment errors. Moreover, using showerheads as a research object can also provide reference for studies of other similar products, such as toy, furniture, cloth, electrical products, etc. Therefore, this paper selects shower products as an application example, combining both practical significance and research value.

Therefore, this case demonstrates how existing shower design patent data can be effectively utilized for model training, enabling subsequent extraction and identification of core design features for other shower products. This approach shortens the evaluation time and improves the accuracy of infringement assessment. An example of shower is presented below to illustrate its appearance features, as shown in Figure 8.



1. Shower panel; 2. Regulator; 3. Outer contour; 4. Handle; 5. Thread

Figure 8. Schematic diagram of shower appearance features.

5.2. Collection and Processing of Shower Design Patent Data

In this study, shower design patents are selected as the database. According to retrieval query—combination of keywords and Locarno classification, shower design patents are collected from the

 doi:10.20944/preprints202603.2216.v1

Patsnap[4], a famous patent search engine. The corresponding search strategy for shower design patents is listed in Table 3. After data cleaning and filtering, a total of 3,000 valid shower design patents is obtained.

Table 3. Shower design patent search formula and results.

Patent retrieval	Search results	Filtering results	Pictures number
[TI = (*Shower)] AND [LOC = (23-06)] AND [AD = (2010-01-01 TO 2024-12-31)]	6167	2995	24493

Shower design patents mainly consist of three parts: authorization information, appearance design images, and brief description. For shower design patents, the most important part is the views section of the appearance design. Since design patent files are in PDF format, preprocessing is required to extract the required shower appearance views.

In this study, Python is adopted as the processing language, and the PyMuPDF library is utilized for PDF parsing to extract multiple views of showerheads from design patent documents, thereby constructing the image database. After further screening, a total of 13,518 shower appearance images is obtained, covering multiple views of each product. Representative examples are shown in Figure 9, where the blue boxes indicate different views corresponding to a single patent. As expected, each design patent consistently includes the six standard views (front, back, left, right, top, and bottom). These images provide a sufficient and diverse sample set for subsequent model training.

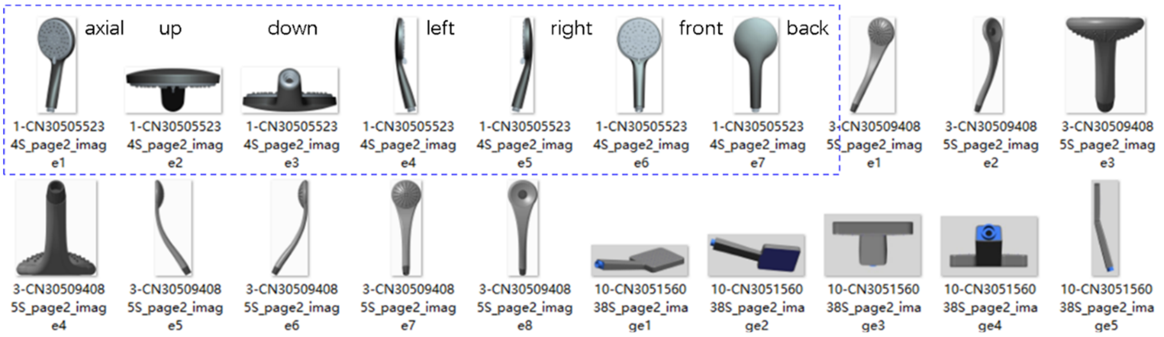


Figure 9. Partial images of the dataset.

5.3. Contour Extraction

Considering the characteristics of product design patents, the scope of protection primarily focuses on design elements such as shape, lines, and contour structures, rather than non-structural attributes such as color or material. In addition, design patent images may be affected by factors such as lighting conditions, imaging devices, and post-processing procedures during data acquisition, which can introduce noise and consequently reduce the accuracy of subsequent similarity computations.

To eliminate the above interference factors, this paper adopts a contour extraction algorithm. The specific steps are as follows: First, convert the design patent image into a grayscale image to remove the influence of color information; secondly, Gaussian filtering algorithm is utilized to smooth the image and reduce noise point interference; finally, the contour extraction algorithm is applied to extract the image contour.

This paper compares four common contour extraction algorithms and statistically analyzes the results from three dimensions: edge continuity, detail preservation, and computational complexity. Edge continuity reflects the algorithm's ability to preserve the overall contour structure of the product, while detail preservation evaluates the completeness of extracting key design details. Computational complexity measures the time and resource consumption of each algorithm in large-scale image processing. The comparative results are summarized in Table 4.

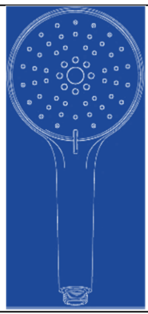
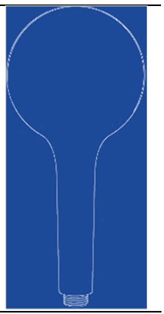
Table 4. Contour detection algorithm results.

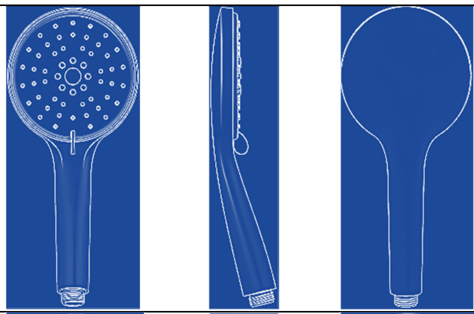
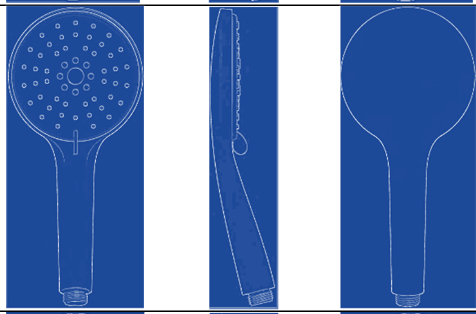
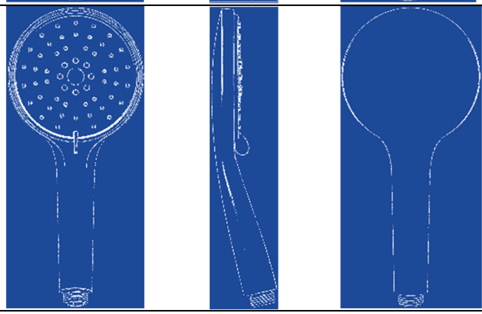
Type	Edge continuity	Detail preservation	Computational complexity
Canny operator	Good	Complete	High
Sobel operator	Average	Partial loss	Low
Laplacian operator	Poor	Noise interference	Medium
Roberts operator	Poor	Significant loss	Low

From the results in Table 4, it can be seen that the Canny operator performs best in edge continuity and detail preservation, being able to better retain the structural features of design patent images; the Sobel operator, although fast in computation, lacks sufficient detail preservation; the Laplacian operator is prone to introducing noise in detail extraction; the Roberts operator, although simple to implement, is too sensitive to image noise, resulting in many edge breaks in detection results, making it unsuitable for product appearance images with complex details. These results validate the selection of the Canny operator for contour extraction in this study, as it provides the most favorable overall performance for design patent images.

To provide a more intuitive comparison of the results produced by different algorithms, three representative views of a randomly selected design patent are chosen for contour extraction. The contour detection results obtained using different algorithms are presented in Table 5. Obviously, Canny operator clearly outperforms the other algorithms in both line detection and noise suppression.

Table 5. Contour detection results.

View / Method			
Grayscale and Gaussian filtering results			
Canny operator			

Sobel operator	
Laplacian operator	
Roberts operator	

5.4. Experimental Environment Configuration and Model Training

To address the limitation of insufficient training data in practical scenarios, this study adopts a transfer learning strategy to enhance model performance [55]. Specifically, a ResNet50 model pretrained on the ImageNet dataset is employed as the backbone network, which is loaded from the official PyTorch model repository [56]. Having learned rich and general-purpose visual representations, the pretrained model is further fine-tuned for the target task, thereby improving generalization capability and accelerating training convergence.

Due to variations in image size, image preprocessing is applied to resize all shower appearance images to 224×224 pixels. The constructed dataset contains a total of 13,518 handle-shower images, of which 70% are used for training and the remaining 30% for validation, as summarized in Table 6. In addition, data augmentation techniques—including random horizontal flipping, rotation, and brightness and contrast adjustment—are employed to expand the training dataset and mitigate overfitting.

Table 6. Shower dataset.

Type	Training set	Validation set	Total
Appearance Images	9463	4055	13518

The experiments are conducted using Python 3.12 with PyTorch 2.5.0 and CUDA 12.4. The hardware platform consists of an NVIDIA RTX 4060 GPU, an Intel i7-12700F CPU, and 32 GB of RAM. The parameter settings for all models are configured as follows: the initial learning rate, batch size, and number of training epochs are set to 0.0001, 32, and 30, respectively. The Adam optimizer is employed for model optimization, and ReduceLROnPlateau learning rate scheduler is used to automatically decrease the learning rate when the validation accuracy shows no improvement for an

 doi:10.20944/preprints202603.2216.v1

extended period, with a reduction factor of 0.5 and a patience value of 5. Cross-entropy loss is adopted as the objective function for training.

The training loss and accuracy curves of the CBAM-ResNet 50model over successive epochs are shown in Figure 10 and Figure 11, respectively. As the number of epochs increases, the training loss gradually decreases while the accuracy steadily improves. Moreover, the model also exhibits enhanced performance on the test set, confirming its reliable generalization capability.

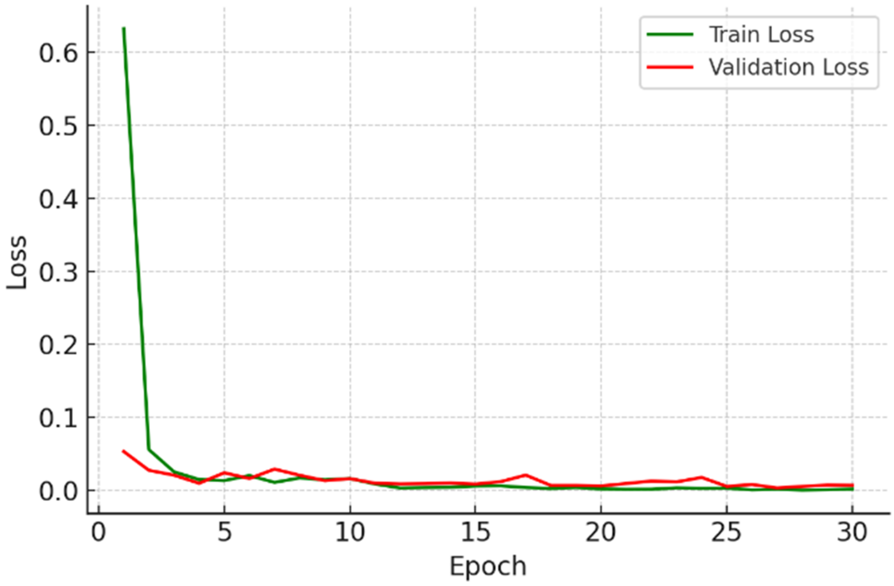


Figure 10. Loss function curve of the CBAM-ResNet50 model training process.

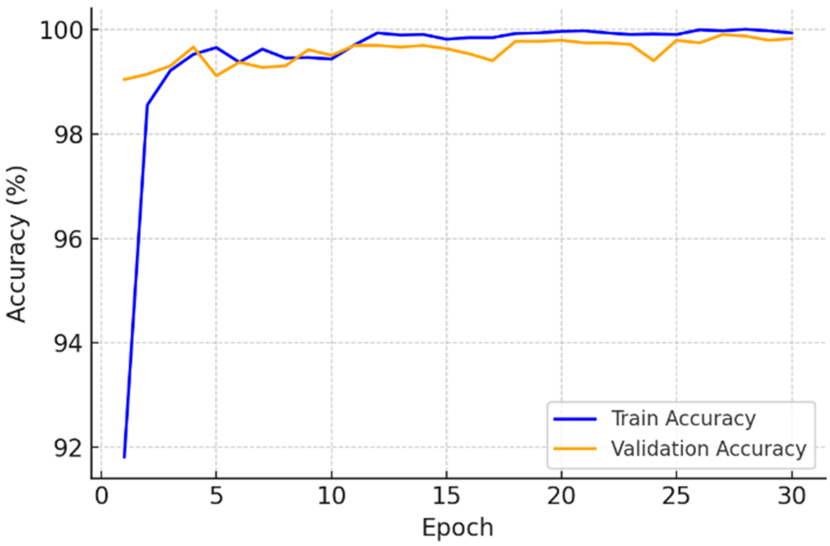


Figure 11. Accuracy curve of the CBAM-ResNet50 model training process.

5.5. Model Comparison and Ablation Experiments

To verify the effectiveness of the proposed CBAM-ResNet50 model, this paper designs three sets of comparative experiments, analyzing from three perspectives: different backbone networks, backbone network depth, and ablation experiments. All models are trained and tested on the same sprinkler product database, and the optimizer and learning rate are consistent to ensure the fairness of the comparison.

Considering the large size of the shower design patent database, 400 shower design patents are randomly selected as experimental samples. To ensure the reliability and rationality of the attention-

 doi:10.20944/preprints202603.2216.v1

based feature annotations, 20 patent engineers with extensive experience in design patent examination are invited to participate in the annotation process on line[5]. All experts are affiliated with Xiamen Zhihuichengrui Patent Agency, a firm that has long been engaged in patent applications for sanitary ware products since 2014. The list of participating experts is provided in Table 7. Obviously, most of the experts possess substantial professional experience in handling showerhead-related design patents.

Table 7. Expert group list.

Name	Seniority	Education	Number of shower patents represented
Yongjun Zheng	10+	master	23
Zhenrong Wu	10+	bachelor	3
Minghui Qiu	10+	master	5
Fuli Guo	10+	bachelor	51
Sifan Wei	10+	master	20
Suwei Lin	5	bachelor	5
Xiande Lin	7	bachelor	30
Weijun Xiao	4	bachelor	5
Xiaosi Chen	8	bachelor	26
Jinsheng Zheng	8	bachelor	11
Guineng Yang	8	bachelor	4
Huaixuan Chen	8	bachelor	24
Yufang Yang	10+	bachelor	38
Ping Luo	8	bachelor	8
Shuping Li	5	bachelor	14
Weiting Wang	5	bachelor	16
Jiede Lin	4	bachelor	13
Weifu Zhang	8	bachelor	1
Xiaoqin Wang	8	bachelor	10
Yufang Yang	10	bachelor	38
Wei Yang	8	bachelor	82

This paper adopts commonly used metrics for machine learning tasks, including accuracy, precision, recall, F1 score, IoU and DSC [57]. The calculation formulas are as follows.

$$Precision = \frac{TP}{TP + FP} \tag{14}$$

$$Recall = \frac{TP}{TP + FN} \tag{15}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{16}$$

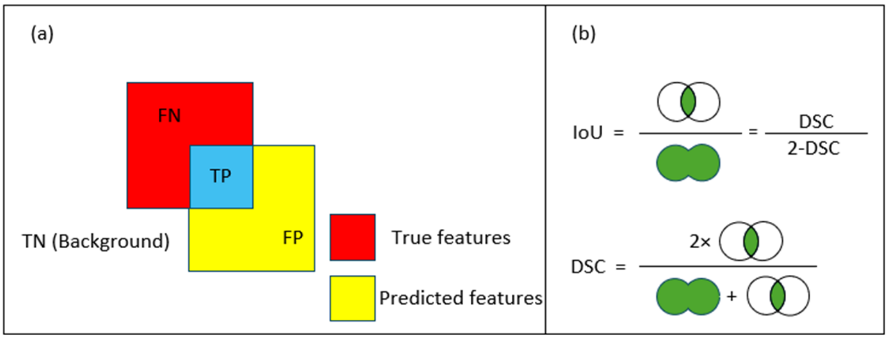
$$Accurary = \frac{TP + TN}{TP + TN + FP + FN} \tag{17}$$

where True Positives (TP) represents the number of mask pixels that the model correctly notices as product features; False Positive (FP) represents the number of mask pixels incorrectly identified as product features; True Negatives (TN) represents the number of correctly identified background pixels; False Negatives (FN) represents the number of incorrectly identified background pixels; Precision represents the proportion of features noticed by the model to the actual product features; Recall represents the proportion of actual product features successfully focused on by the model. F1 balances Precision and Recall to provide an overall evaluation of the model's attention-based feature recognition capability. Accuracy quantifies the overall pixel-wise classification correctness across the entire image.

IoU measures the degree of overlap between the predicted attention region and the ground-truth annotated region, while the DSC emphasizes the overall similarity between the two regions. Their

 doi:10.20944/preprints202603.2216.v1

corresponding formulations are given as follows, and the underlying computation principles are illustrated in Figure 12.



（a）True positive, false positive, false negative, and true negative (b) IoU and DSC.

Figure 12. Explanation of metrics.

$$IoU = \frac{TP}{TP + FP + FN} \tag{18}$$

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{19}$$

From the relationship between IoU and DSC, it can be derived that:

$$IoU = \frac{DSC}{2 - DSC} \tag{20}$$

5.5.1. Comparative Experimental Analysis of Backbone Networks

To evaluate the impact of different backbone networks on the model performance, several mainstream architectures are compared, including ResNet50, VGG16, InceptionV3, DenseNet121, and MobileNetV2 [58]. The 400 design patent images described in Section 5.5 are fed into each model to generate attention regions. These predicted attention regions are then compared with the ground-truth salient feature annotations in the dataset. The average values of the evaluation metrics are calculated, and the comparative results are summarized in Table 8.

Table 8. Different backbone network metric results.

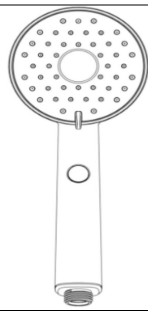
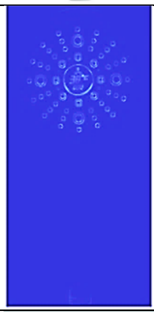
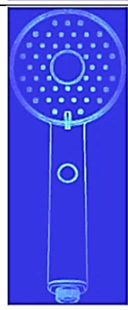
Metric / Model	Average（%）		
	Accuracy	IoU	DSC/F1
ResNet50	99.4	74.5	85.5
VGG16	99.0	60.7	75.4
InceptionV3	98.4	44.8	61.8
DenseNet121	98.1	42.1	59.3
MobileNetV2	98.7	52.1	69.4

From the Accuracy metric in Table 8, the performance of each model in overall pixel classification is relatively high, indicating that the models have strong stability in background region recognition. However, because this metric is sensitive to background pixels, it cannot fully reflect the model’s identification ability in salient feature regions.

In contrast, the IoU and F1 metrics more effectively reflect the models’ capability to accurately identify target feature regions. The results show that the ResNet50 model achieves IoU and F1 values of 74.5% and 85.5%, respectively, significantly outperforming the other architectures. This indicates that its attention regions more comprehensively cover the annotated core features while exhibiting less background interference. VGG16 ranks second, as it is able to capture the primary features but still misidentifies some non-feature regions. MobileNetV2 demonstrates moderate performance, whereas InceptionV3 and DenseNet121 yield relatively lower IoU and F1 values, suggesting more dispersed attention regions and insufficient focus on salient features.

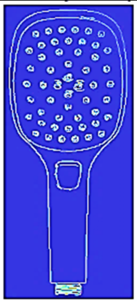
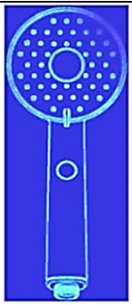
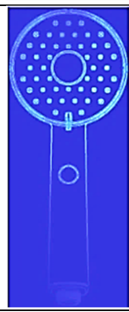
In summary, ResNet50 achieves the best overall performance in this task, providing comprehensive coverage of the core feature regions of shower designs. VGG16 ranks second, while MobileNetV2 and InceptionV3 exhibit moderate performance. In contrast, DenseNet121 produces more dispersed attention regions, making it difficult to accurately identify salient features. To provide a more intuitive comparison of the experimental results, three design patents are randomly selected from the 400 samples for visualization. As the principal view most comprehensively reflects product design characteristics, the main view of each product is selected for illustration. The corresponding results are presented in Table 9.

Table 9. Comparison of different backbone network results.

View / Model			
ResNet50			
VGG16			
InceptionV3			

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 March 2026 doi:10.20944/preprints202603.2216.v1

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 March 2026 doi:10.20944/preprints202603.2216.v1

DenseNet121			
MobileNetV2			

The results presented in Table 9 reveal noticeable differences among the models in terms of their attention focus regions. Although VGG16 is able to capture the overall structural features, it shows limitations in representing fine-grained details. Owing to its multi-scale architecture, InceptionV3 demonstrates improved attention to local details in certain images; however, its overall performance remains inferior to that of ResNet50. DenseNet121 is capable of focusing on the main structure of the showerhead, but it is less effective for identifying core design features. As a lightweight architecture, MobileNetV2 produces relatively reasonable attention regions; nevertheless, it still lags behind ResNet50 in terms of attention concentration and fine-grained feature recognition.

5.5.2. ResNet Network Depth Impact Comparison

To investigate the impact of network depth on model performance, several ResNet variants with different depths—namely ResNet18, ResNet34, ResNet50, and ResNet152—are selected for comparative experiments [59]. These models differ in their fundamental architectural designs: ResNet50 and ResNet152 adopt bottleneck structures incorporating 1×1 convolutions for dimensionality reduction and expansion, whereas ResNet18 and ResNet34 employ basic 3×3convolutional blocks. Under identical experimental settings, the effects of network depth and structural differences on model capability are systematically compared, and the result is shown in Table 10.

Table 10. ResNet network depth metric results.

Model \ Metric	Average（%）		
	Accuracy	IoU	DSC/F1
ResNet18	99.1	71.4	83.3
ResNet34	99.0	68.5	81.4
ResNet50	99.4	74.5	85.5
ResNet152	97.9	43.4	60.6

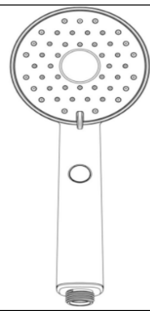
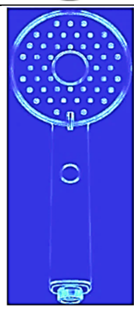
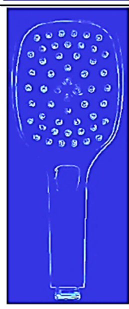
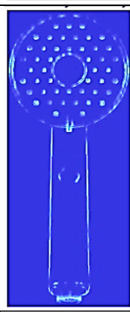
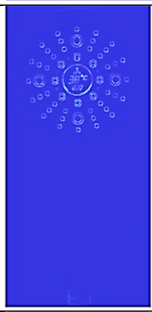
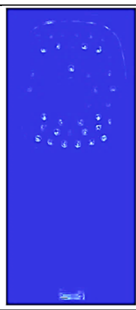
As indicated by the metrics reported in Table 10, increasing the depth of the ResNet architecture does not lead to a monotonic improvement in performance. ResNet50 achieves the best overall results across all evaluation metrics, demonstrating its ability to accurately and comprehensively focus on salient feature regions while effectively suppressing background interference. ResNet18 and ResNet34 exhibit slightly lower performance, suggesting that although they can capture the primary features, they remain limited in modeling fine-grained details and boundary information. In contrast,

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 March 2026 doi:10.20944/preprints202603.2216.v1

ResNet152 shows a notable decline in IoU and DSC/F1, indicating that excessive network depth weakens the model's ability to concentrate on core features.

In summary, a moderate network depth enables the model to enhance attention concentration while preserving effective feature extraction capability, and ResNet50 achieves the best overall performance in this task. To further illustrate the impact of different ResNet depths, the front views of three shower design patents selected in Section 5.5.1 are visualized for comparison. This qualitative analysis provides a more intuitive demonstration of the performance differences among ResNet architectures with varying depths. The result is shown in Table 11.

Table 11. Comparison of different ResNet network depth results.

View / Model			
ResNet18			
ResNet34			
ResNet50			
ResNet152			

 doi:10.20944/preprints202603.2216.v1

Evidently, although ResNet18 and ResNet34 offer higher computational efficiency, their relatively shallow architectures lead to extracted features that contain excessive fine-grained and irregular details, which is unfavorable for subsequent core feature identification. In contrast, while ResNet152 possesses stronger semantic abstraction capability due to its greater depth, the excessive down-sampling results in the loss of substantial low-level detail information, thereby degrading its ability to accurately characterize core product features. By comparison, ResNet50 achieves a better balance between feature concentration and detail preservation, yielding more reliable performance for core feature analysis.

5.5.3. Ablation Experiments

To verify the effectiveness of the proposed network and the rationality of the CBAM module placement, ablation experiments are conducted. The ResNet50 architecture consists of four convolutional stages. The proposed CBAM-ResNet50-C2 model integrates the CBAM attention mechanism after the Conv2_x stage of ResNet50. To evaluate the impact of different CBAM insertion positions, three comparable variants are further constructed, denoted as CBAM-ResNet50-C1, CBAM-ResNet50-C3, and CBAM-ResNet50-C4. Specifically, CBAM-ResNet50-C1 inserts the CBAM module after the Conv1_x stage, CBAM-ResNet50-C3 integrates CBAM after the Conv3_x stage, and CBAM-ResNet50-C4 places the CBAM module after the Conv4_x stage.

In addition, to evaluate the individual contributions of different attention components, the channel attention module (CAM) and the spatial attention module (SAM) of CBAM are separately integrated into the ResNet50 network for comparison. Furthermore, to demonstrate the advantages of CBAM over other attention mechanisms, the squeeze-and-excitation (SE) module and a self-attention module are also incorporated into the model for comparative analysis. To ensure a fair comparison, all attention mechanisms are inserted at the same position—after the Conv2_x stage of the ResNet50 network. The corresponding network architectures are illustrated in Figure 13, and the experimental results are reported in Table 12.

Table 12. Ablation experiment metric results.

Model \ Metric	Average（%）		
	Accuracy	IoU	DSC/F1
CBAM-ResNet50-C1	98.1	44.0	61.1
CBAM-ResNet50-C3	98.3	45.3	62.4
CBAM-ResNet50-C4	98.7	44.6	61.7
CAM-ResNet50	98.2	42.5	60.0
SAM-ResNet50	98.8	72.7	84.2
SE-ResNet50	98.0	42.6	59.7
Self-attention-ResNet50	96.0	38.8	55.9
CBAM-ResNet50-C2	99.4	74.5	85.5

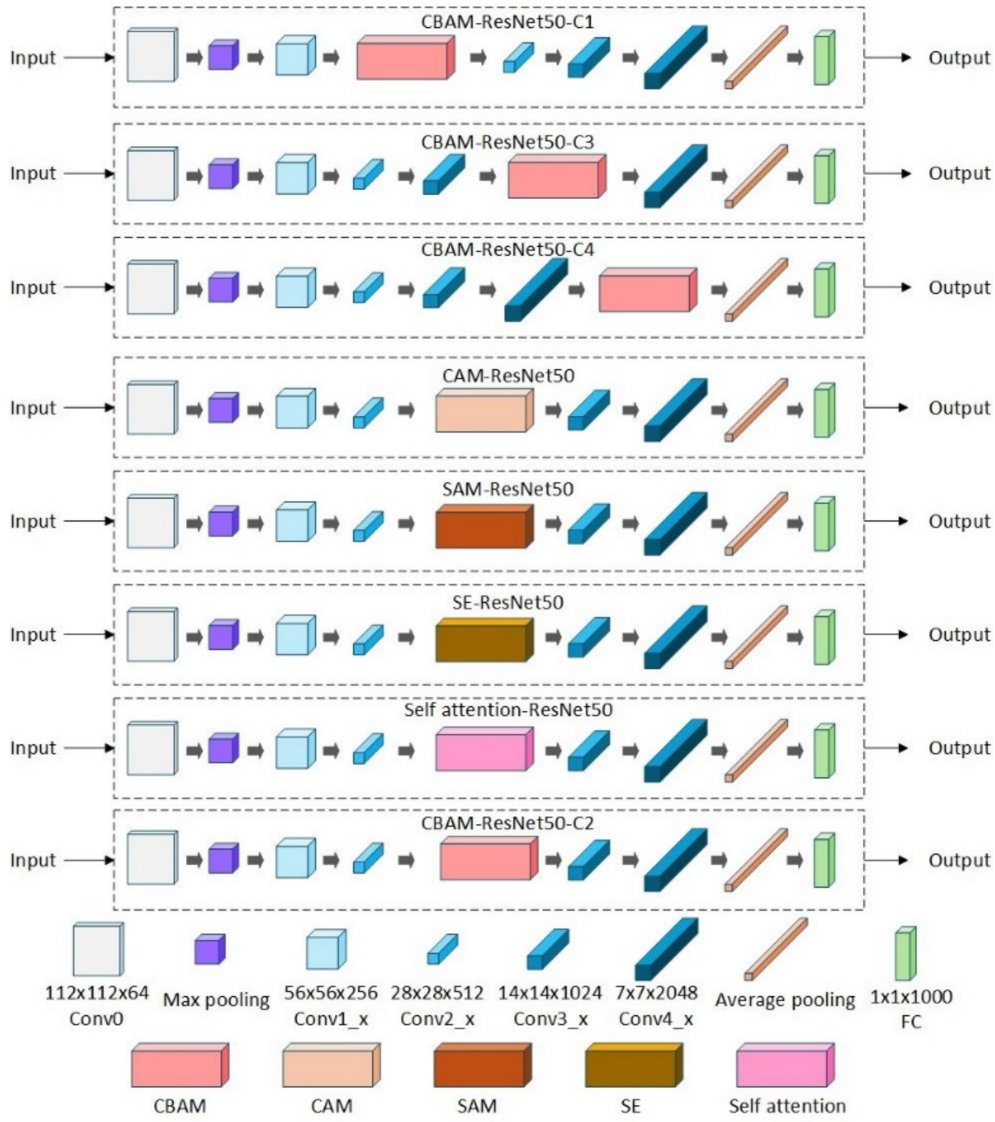
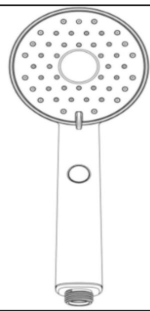
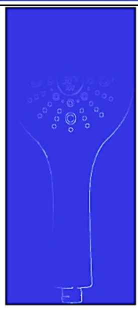
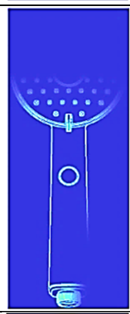
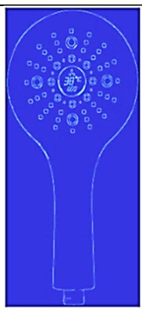
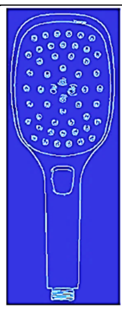
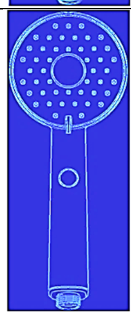
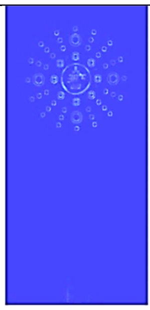


Figure 13. Network structure diagrams at different positions and with different attention mechanisms.

As shown in Table 12, the performance of CBAM-ResNet50-C1, CBAM-ResNet50-C3, and CBAM-ResNet50-C4 varies due to different CBAM insertion positions. All three variants exhibit noticeably lower IoU and F1 values than the proposed model—CBAM-ResNet50-C2, indicating insufficient responses to key feature regions. The CAM-ResNet50 model introduces channel-wise weighting; however, its lack of spatial selectivity prevents improvements in localization accuracy. Although the SAM-ResNet50 model demonstrates relatively strong feature localization capability, its attention distribution remains overly diffuse. The SE-ResNet50 model shows a decline in IoU and DSC/F1, suggesting that channel-only attention has inherent limitations in feature recognition and fails to adequately capture spatial details. In addition, the Self-attention-ResNet50 model exhibits excessive responses to non-feature regions, leading to reduced discrimination of salient features.

Overall, the results demonstrate that the proposed CBAM-ResNet50-C2 model achieves the best attention performance for shower feature recognition, and that the selected insertion position of the attention mechanism is also optimal. Through a comparative analysis of different attention mechanisms, the effectiveness and rationality of the proposed model and attention design for shower feature identification are further validated. To provide a more intuitive illustration, the attention results of the aforementioned three design patents are presented as Table 13.

Table 13. Comparison of ablation experiment results.

View / Model			
			
CBAM-ResNet50-C1			
CBAM-ResNet50-C3			
CBAM-ResNet50-C4			
CAM-ResNet50			
SAM-ResNet50			

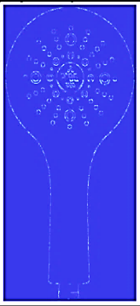
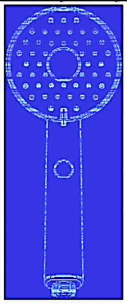
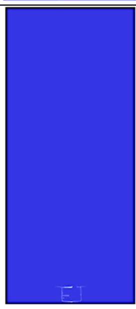
SE-ResNet50			
Self-attention-ResNet50			
CBAM-ResNet50-C2			

Table 13 presents the results of the ablation experiments. Due to different insertion positions of the attention modules, the first three models tend to focus on relatively abstract features, with attention predominantly concentrated on image edges or local structures, which is unfavorable for salient feature discrimination. CAM and SAM correspond to the channel attention and spatial attention components of the CBAM mechanism, respectively. The results show that CAM preserves an excessive amount of feature information, which hampers salient feature identification, whereas SAM suffers from insufficient feature completeness. In comparison, the CBAM mechanism adopted in this study effectively compensates for the limitations of both CAM and SAM by jointly modeling channel-wise and spatial attention, thereby achieving superior performance. Other attention mechanisms, such as SE and self-attention, either fail to capture meaningful salient features or attend to non-salient regions, resulting in inferior performance compared with the proposed method

5.6. Product Feature Attention Score Calculation

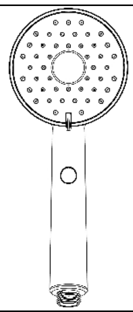
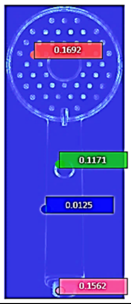
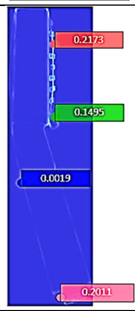
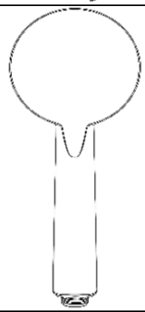
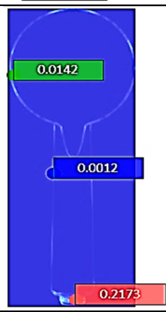
To further evaluate the model's focus on the core structural regions of shower designs, a quantitative analysis is conducted based on the attention maps extracted from the CBAM module. These attention maps incorporate both channel attention and spatial attention, which enhance feature representation from the perspectives of feature dimensions and spatial locations, respectively. Through their combined effect, the model is able to more accurately concentrate on the key regions of shower images.

First, the input feature maps are processed by global average pooling and global max pooling to extract complementary information from different perspectives. The pooled features are then fused through convolution operations and activation functions to generate the final attention weight map. Based on this attention map, normalization is applied, and the attention weights within the shower feature regions are averaged to obtain quantitative scores for evaluation. Meanwhile, these scores are overlaid on the attention maps, resulting in scores for different features of the shower across multiple views.

 doi:10.20944/preprints202603.2216.v1

Through this approach, complex attention distributions are transformed into intuitive numerical indicators. By processing shower images from multiple views, the attention maps generated by the model and their corresponding quantitative scores are jointly stored as input data for subsequent core feature identification. Taking a single shower product as an example, the attention scores of its features in the front view, side view, and rear view are reported in Table 14.

Table 14. Shower product original image, contour image, and attention image.

View	Original image	Contour image	Attention image
Front			
Side			
Rear			

5.7. Product Feature Similarity Calculation

To determine the core features of shower designs, relying solely on convolutional neural networks and attention mechanisms to extract salient regions is insufficient. Core design patent features of showerheads should exhibit both saliency—i.e., visually distinctive elements that characterize the product design—and novelty, meaning discernible differences compared with existing patented designs. Therefore, after obtaining the saliency results of shower features, image similarity computation is further employed to compare different shower samples and assess whether the identified features possess novelty relative to prior designs.

Considering that design patents protect contour features rather than color attributes, the contour images of the 400 shower design patents described above are selected as the test set. To ensure the objectivity of similarity annotation, 20 experts listed in Table 5. conducted manual evaluations of these samples. Based on the consensus of the expert panel, 200 pairs of samples with visually similar appearance features and 200 pairs of samples with significantly different appearance features are selected to construct the evaluation dataset.

Since contour images in design patents are typically black-and-white, similarity computation based on color histograms is not appropriate. Therefore, the dataset is evaluated using pixel-structure–based similarity computation methods for comparative analysis. These methods include

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 March 2026 doi:10.20944/preprints202603.2216.v1

the Structural Similarity Index (SSIM), feature matching–based approaches (e.g., SIFT and ORB), and cosine similarity based on deep learning feature representations (e.g., ResNet50+Cos, InceptionV3+Cos, DenseNet121+Cos, MobileNetV2+Cos, and VGG16+Cos). Each method outputs a corresponding similarity score for evaluation.

Since different algorithms may compute similarity scores using different criteria and scales, similarity cannot be determined solely based on absolute score values. Therefore, the ROC curve and the AUC metric are adopted in this study [60]. The ROC curve is not derived from a single fixed threshold; instead, it is generated by traversing all possible decision thresholds that distinguish similar and dissimilar samples. By plotting the ROC curve and computing the AUC value, the discriminative capability of each algorithm can be systematically evaluated.

The AUC represents the area under the ROC curve and serves as an overall performance indicator, with values ranging from 0 to 1. In the ROC curve, the horizontal axis corresponds to the False Positive Rate (FPR), and the vertical axis corresponds to the True Positive Rate (TPR), which are defined as follows:

$$TPR = \frac{TP}{TP + FN} \tag{21}$$

$$FPR = \frac{FP}{FP + TN} \tag{22}$$

Based on the similarity scores obtained for the similar and dissimilar sample pairs by each algorithm, the statistical results of the similarity scores are summarized in Table 15.

Table 15. Score statistical results of each algorithm.

Models	Similar group score range	Dissimilar Group score range	Average	
			Similar	Dissimilar
SSIM	0.6045-0.8621	0.4874-0.8472	0.7681	0.6879
SIFT	0.1301-0.6937	0.0214-0.8659	0.4770	0.4094
ORB	0.6909-0.8641	0.6463-0.7382	0.7318	0.7007
ResNet50+Cos	0.8938-0.9643	0.7886-0.9461	0.9348	0.8606
InceptionV3+Cos	0.8153-0.9390	0.6837-0.9081	0.8924	0.8044
DenseNet121+Cos	0.8216-0.9466	0.6635-0.8740	0.8910	0.7523
MobileNetV2+Cos	0.7977-0.9243	0.6219-0.8943	0.8767	0.7193
VGG16+Cos	0.8997-0.9317	0.5861-0.6762	0.8997	0.6277

Subsequently, the corresponding ROC curves and AUC values are plotted, as illustrated in Figure 14.

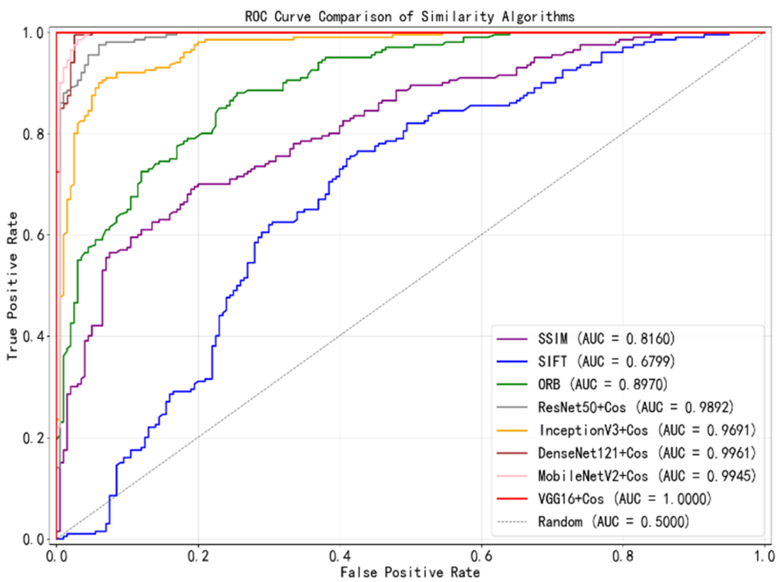


Figure 14. ROC curves and AUC indicators for each algorithm.

Specifically, cosine similarity methods based on deep learning feature representations demonstrate excellent performance in similarity measurement. Among them, the VGG16+Cos approach yields similarity scores concentrated in a high range (0.8997–0.9317) for the similar sample group, with an average value of 0.8997, while the scores for the dissimilar group drop significantly to 0.6277, resulting in a clear separation between the two groups. This pronounced distribution difference leads to a high AUC value of 1.0000, indicating that the method possesses outstanding discriminative capability for design patent contour image similarity analysis. In addition, DenseNet121+Cos and MobileNetV2+Cos also exhibit strong performance, achieving AUC values of 0.9961 and 0.9945, respectively, further confirming the superiority of combining deep learning–based feature extraction with cosine similarity measurement.

In contrast, traditional algorithms exhibit notable limitations in this task. Although the SSIM algorithm is able to distinguish similar and dissimilar image pairs to some extent, the similarity scores of the dissimilar group span a wide range (0.4874–0.8472), with an average value of 0.6879. This indicates that relatively high similarity scores are assigned to certain dissimilar image pairs, which may lead to misclassification. The ORB algorithm relies on local key point matching and shows limited robustness to global structural variations. Although it achieves a relatively high AUC value, the score distributions of the similar and dissimilar groups exhibit substantial overlap, resulting in limited discriminative power. The SIFT algorithm performs the weakest among the compared methods, with severe overlap between the two score distributions and an AUC of only 0.6799, making it ineffective for reliably distinguishing between similar and dissimilar contour images.

Considering both the AUC values and the distribution characteristics of similarity scores across different algorithms, cosine similarity methods based on deep learning feature representations demonstrate superior accuracy and robustness for similarity computation on design patent contour images. Among them, the VGG16 combined with cosine similarity achieves the best overall performance on the contour dataset, as it more effectively captures the global morphological characteristics and subtle differences of product appearance contours. Consequently, this approach is particularly well suited for similarity analysis tasks involving design patent contour images.

5.8. Core Calculation and Core Feature Determination

5.8.1. Comparison with Other Models

In this section, the samples of 400 shower design patents are selected as the study objects again. Their patent drawings are fed into the proposed model to obtain attention scores for individual features. Subsequently, a similarity algorithm is employed to compute feature-wise similarity scores between the contour images of these samples and those in the patent database. Finally, the importance of different features within each sample is calculated according to Eq. (11), which serves as the basis for determining the core design features.

To further assess the accuracy and rationality of the proposed approach, this section again invites the 20 experts listed in Table 5 to participate in the evaluation process. It should be noted that manual assessment is inherently subjective, and discrepancies may arise among experts when identifying core appearance features of products. To reduce the impact of subjective bias, an expert voting strategy is adopted. In cases where disagreement occurs regarding the identification of core features, the decision supported by the majority of experts (i.e., at least 70%) is accepted.

The evaluation procedure is conducted as follows. First, the core feature identification results of 400 randomly sampled shower design patents are obtained using the proposed method, as summarized in Table 16. These results are then submitted to the experts for online evaluation[4], where the experts assess the correctness of the identified core features. Subsequently, the expert evaluation results are aggregated. In cases where discrepancies occur among expert opinions, the majority decision is adopted as the final outcome.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 March 2026 doi:10.20944/preprints202603.2216.v1

Table 16. Sample results of product core features.

No.	Patent number	View	Core feature
1	CN305237649S		Shower Panel
2	CN305333393		Shower Panel
…	…	…	…
400	CN305466729S		Outer Contour

In addition, the accuracy of different methods is compared, and the results are illustrated in Figure15, 16, and 17.

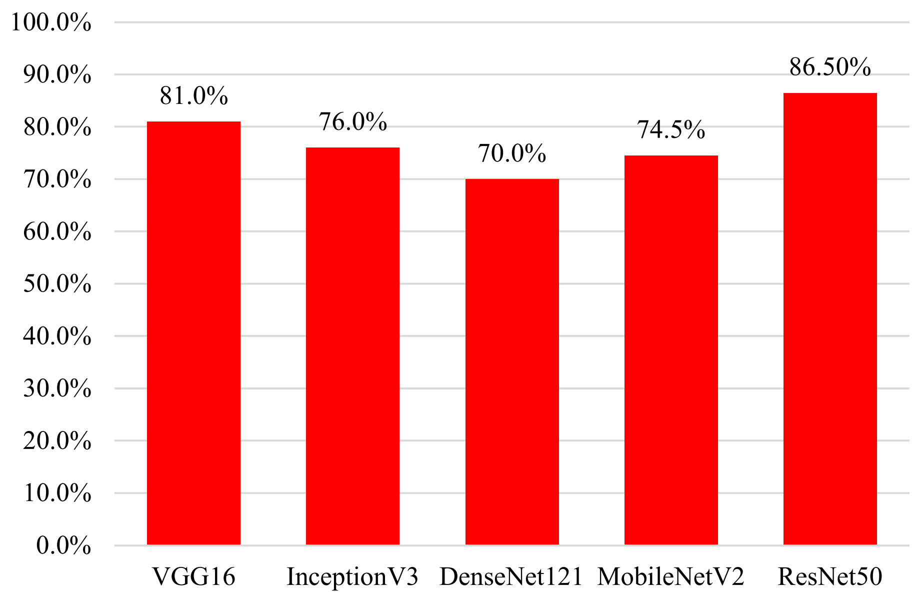


Figure 15. Accuracy results of different backbone networks.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 March 2026 doi:10.20944/preprints202603.2216.v1

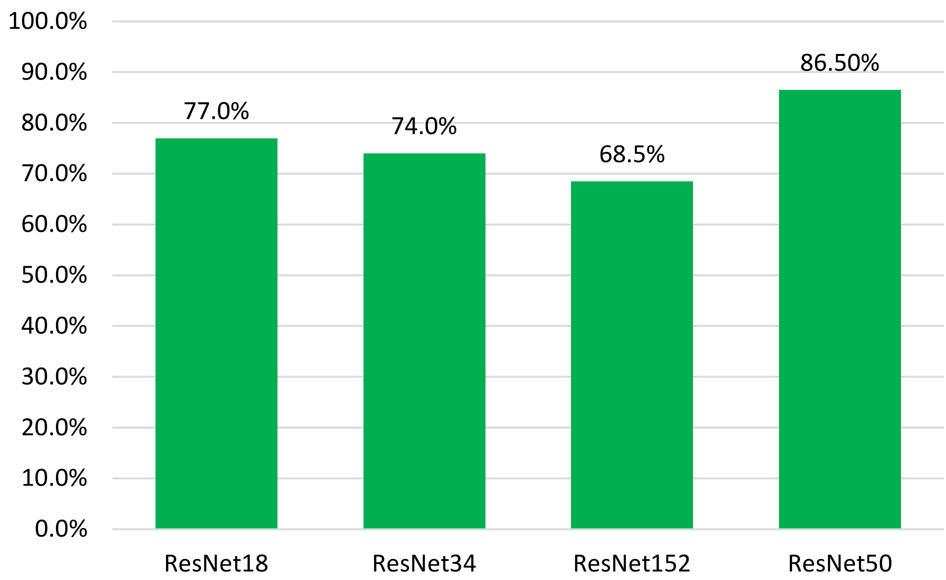


Figure 16. Accuracy results of ResNet with different network depths.

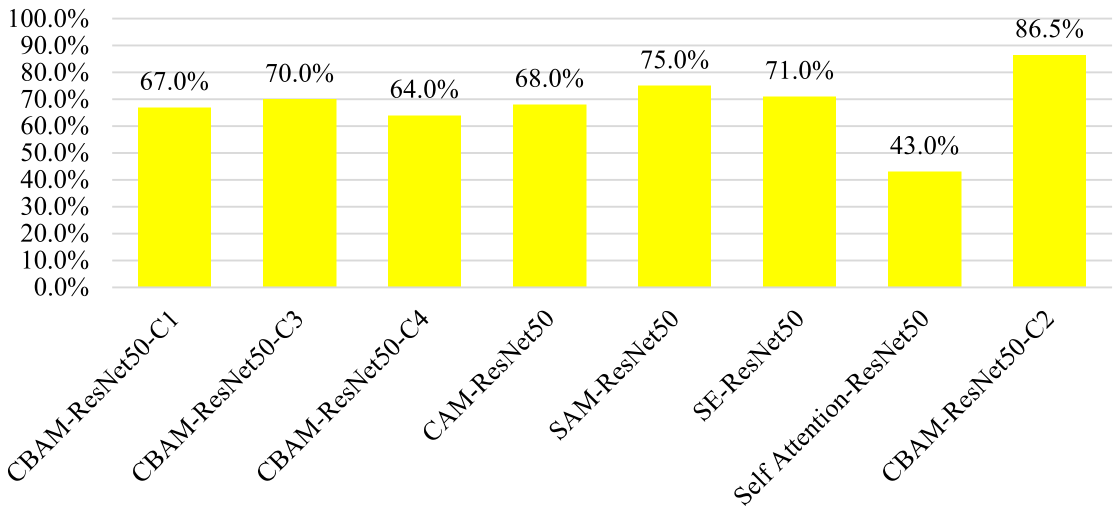


Figure 17. Accuracy results of ablation experiments.

This experiment first compares the results obtained using different backbone network architectures described in Section 5.5.1. By combining attention score computation and similarity analysis, the core feature identification results produced by each model are extracted. These results are then validated by patent domain experts, and the corresponding accuracy is calculated. The results show that, under identical experimental conditions, different backbone networks have a substantial impact on performance, with accuracy ranging from 70% to 86.5%. Evidently, the proposed model—CBAM-ResNet50 achieves the highest accuracy among all compared models.

Subsequently, the accuracy of ResNet models with different depths described in Section 5.5.2 is evaluated. The results show that the medium-depth ResNet50 model adopted in this study achieves the highest accuracy, whereas the deeper ResNet152 model exhibits a performance decline. This indicates that increasing network depth does not necessarily lead to improved accuracy; excessively deep architectures may cause feature overfitting or attention diffusion, thereby reducing the model's ability to focus on key feature regions.

Finally, the accuracy results of the ablation experiments described in Section 5.5.3 indicate that the CBAM attention mechanism adopted in this study, together with its selected insertion position,

 doi:10.20944/preprints202603.2216.v1

achieves higher accuracy than other attention mechanisms and alternative insertion configurations. These results validate that the CBAM-ResNet50-C2 module is more effective in focusing on core product features compared with other attention mechanisms, and that the insertion position is also optimal for enhancing feature identification performance.

In summary, the results of the three experiments demonstrate that different backbone networks exhibit varying performance in core feature identification, with the ResNet50 backbone outperforming the other models. Moreover, a network with moderate depth is more conducive to effective attention concentration. Finally, incorporating the CBAM module achieves superior performance compared with other attention mechanisms. Overall, the model integrating the CBAM-ResNet50 architecture—particularly the CBAM-ResNet50-C2 variant—achieves a significantly higher expert-evaluated accuracy, thereby fully validating the effectiveness of the proposed method for product core feature identification tasks.

5.8.2. Comparison with LLMs

To further validate the rationality of the proposed method and to explore the applicability of large-scale models to the task of core feature identification in design patents, multiple multimodal large models are introduced as comparative baselines. Core feature identification experiments are conducted on the same set of 400 shower design patent samples for comparison.

Specifically, under the same sample set and evaluation criteria, the views of each shower design patent are input into the large-scale model, which is guided by a unified prompt to identify the core design features from a visual appearance perspective. To ensure a fair comparison, the large model performs the assessment solely based on visual appearance information, without incorporating patent text, functional descriptions, or any manually annotated information.

Following the approach in [61], several mainstream multimodal large-scale models are employed for comparison, including Gemini-3, ChatGPT-5, Copilot, Qwen3-max, and LLaMA 3.2. The core feature identification results produced by these large models are subsequently submitted to the same group of experts for validation, using the same accuracy evaluation criteria as in the preceding experiments. When a model's output is consistent with the DPIJ of the majority of experts (≥70%), it is regarded as a correct identification, and the overall accuracy is computed accordingly. Taking ChatGPT-5 as an example, the corresponding dialogue content is presented in Table 17.

Table 17. The case of ChatGPT-5.

Query
 This is the three views of a shower product, including the main view, side view and rear view. Now you are an expert in industrial design and design patent examination. Now the core features of the appearance design are defined as follows: (1) visually salient regions that are easy to attract attention and have a significant impact on the overall impression of product. (2) design features that differ from existing designs and demonstrate novelty. Please conduct a comprehensive analysis of the appearance of the product based on the above definition to identify its core design features. At the same time, please mark the area determined as

doi:10.20944/preprints202603.2216.v1

the core feature on the original image, and give the judgment reasons based on the visual significance of the feature in the three views and its difference from the common existing designs.

Answer

Next, I will use the standard expression of "industrial design and design patent examination expert" to determine the core design features of the shower product, and make a systematic analysis combined with the three views you give.

1. Overview of the overall appearance of the product

From the front view, side view and rear view, it can be seen that the product is a hand-held shower. The classic structure of circular spray panel and slender handle is adopted as a whole. The modeling style is simple and modern, which is in line with the mainstream design trend of current sanitary products.

Main view: the circular nozzle panel is the visual center, and the spray holes are distributed radially;

Side view: the thickness of the nozzle is thin, and the transition between the handle and the nozzle is natural;

Rear view: the back shape is simple without obvious decorative ups and downs.

In terms of overall structure, the product continues the basic configuration of common sprinklers, but introduces innovative elements with significant recognition in local design.

2. Identification and annotation of core design features

The part judged as the core design feature: black circular area in the center of the nozzle has been marked with a red box in the original main view.

3. Reasons for core feature determination

(1) **Position centralization:** the circular display module is located in the geometric center of the nozzle; When the user observes the product at the first sight, the line of sight naturally focuses on this area, and the attention value is significantly higher than that of the nozzle and handle.

(2) **Strong contrast between color and material:** the black display area and the overall light translucent fuselage form a clear contrast between brightness and material; Digital displays (such as temperature values) further enhance the visual appeal, making them the visual focus in the main view.

(3) **The three views are consistent and perceptible:** Although the module is most obvious in the main view; However, its thickness and structure can still be identified in the side view, which are the constituent elements that cannot be ignored in the overall design.

Conclusion: this feature meets the requirements of "visually easy to attract attention and has a significant impact on the overall effect".

4. Comprehensive judgment conclusion

Comprehensive three view analysis shows that the circular display module in the center of the nozzle has high visual significance in position, shape and color contrast, which can significantly affect

the overall visual effect of the product. At the same time, this feature is not common in the traditional sprinkler design. By taking the visual display structure as an important part of the appearance design, the integration of function and shape is realized, which reflects the innovation different from the existing design. Therefore, the central display module can be identified as the core design feature of the shower appearance design.

*Visiting data is January 26th, 2026.

The comparison of the accuracy of several large models is shown in the Figure 18.

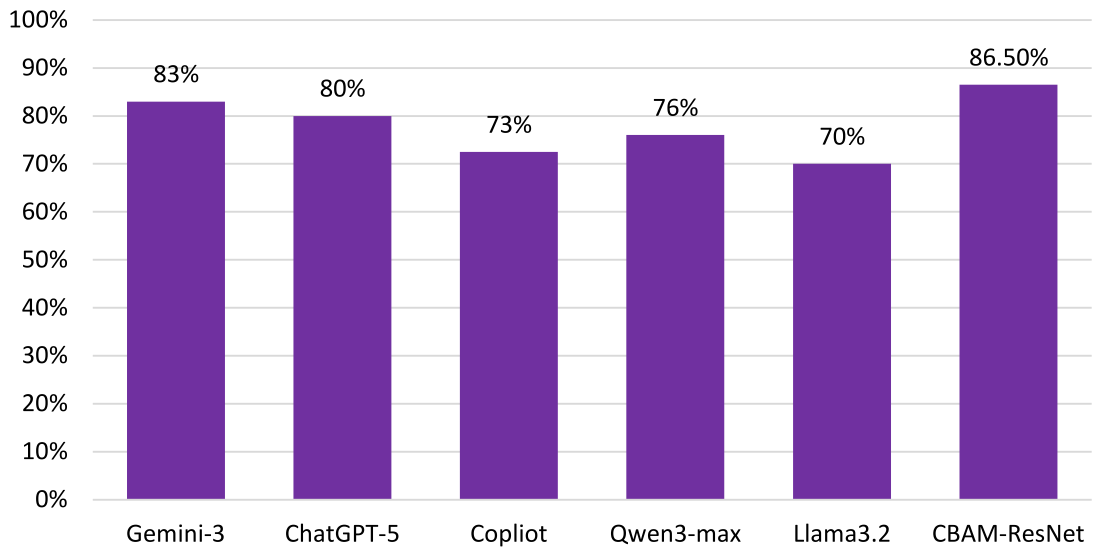


Figure 18. Accuracy of general LLM.

The experimental results indicate that general-purpose large models exhibit a certain degree of feasibility in identifying core features of product appearance designs; however, their overall accuracy remains lower than that of the proposed method. On the one hand, large models tend to make judgments based on holistic visual impressions, making their attention susceptible to factors such as color contrast, surface material appearance, and the presence of brand logos or textual elements in the images. As a result, some non-morphological elements may be mistakenly identified as core design features. On the other hand, when performing similarity analysis, general-purpose large models often rely on coarse-grained shape impressions—such as circular, rectangular, or elongated forms—to make decisions. This limits their ability to effectively distinguish designs that share similar overall contours but differ in fine-grained local geometric details, thereby constraining the accuracy of similarity-based judgments.

By contrast, the proposed method is more consistent with the fundamental principles of novelty assessment in design patent examination, which emphasize the substantial differences reflected in the overall visual effect of a design relative to prior art, rather than similarity judgments at an abstract form or general category level. Through the incorporation of attention constraints, the proposed approach is able to more accurately focus on key geometric regions that have a decisive impact on the overall appearance. This results in improved interpretability and provides further evidence of the effectiveness and robustness of the method for novelty analysis in design patent applications.

5.8.3. Illustration of the Computational Procedure

To make the computation process more intuitive, a shower product from Section 5.5 is selected as an example to illustrate the core feature calculation procedure in detail. This shower product consists of three primary appearance features, namely the shower faceplate, the outer contour of the

doi:10.20944/preprints202603.2216.v1

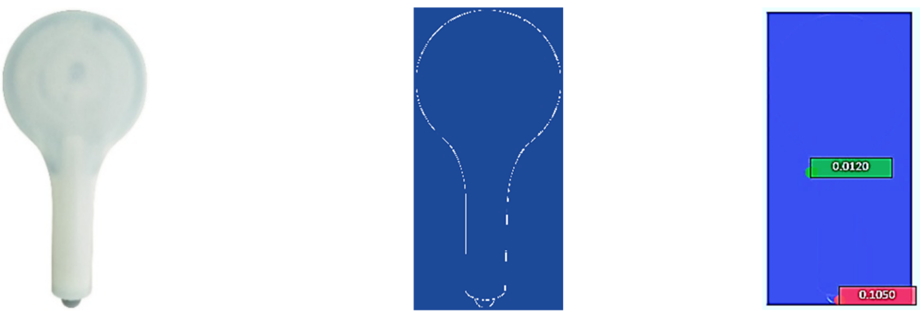
showerhead, and the threaded section. The calculation results for each feature are presented in Figure 19, 20, 21 and Table 18.

(a)Product front view (b) Contour image (c) Attention image

Figure 19. Product front view results.

(a) Product side view (b) Contour image (c) Attention image

Figure 20. Product side view results.



(a) Product rear view (b) Contour image (c) Attention image

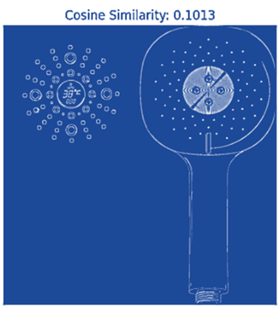
Figure 21. Product rear view results.

Table 18. Attention score results for each feature.

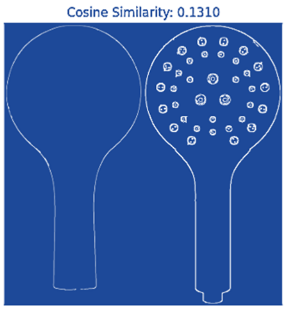
Feature / View	Shower Panel (A_1)	Outer Contour (A_2)	Thread (A_3)
Front View V_1	0.2250	0.0130	0.0950
Side View V_2	0.2060	0.0120	0.1057
Rear View V_3	0	0.0120	0.1050

The similarity of each feature is computed across different views of the product. Since the threaded section is a functional feature and does not fall within the scope of design patent protection, it is excluded from further consideration. The shower faceplate is denoted as feature A_1, and the outer contour is denoted as feature A_2. The corresponding computation results are shown in Figure 22 and Figure 23.

doi:10.20944/preprints202603.2216.v1

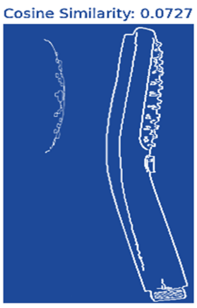


(a) Feature A_1

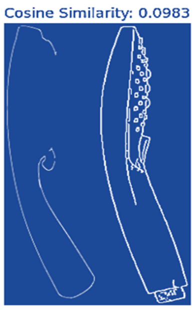


(b) Feature A_2

Figure 22. Front view feature similarity results.



(a) Feature A_1



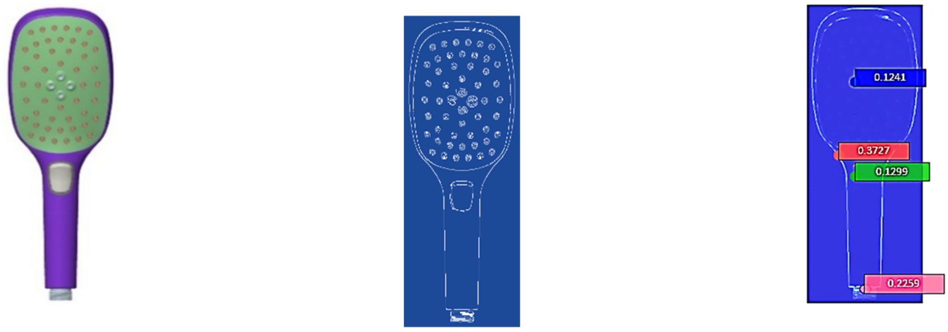
(b) Feature A_2

Figure 23. Side view feature similarity results.

Based on the attention results and similarity measurements, the core feature values of the shower can be computed for each view. Since the rear view of this product does not exhibit additional novel features and its primary characteristic is the outer contour, the outer contour in the main view is selected as the representative core feature. According to Eq. (13), the core feature values of each feature are $Core(V_1, A_1) = 2.22$ and $Core(V_1, A_2) = 0.10$ .And The calculation results of the core feature value for each feature in the side are $Core(V_2, A_1) = 2.83$ and $Core(V_2, A_2) = 0.12$.

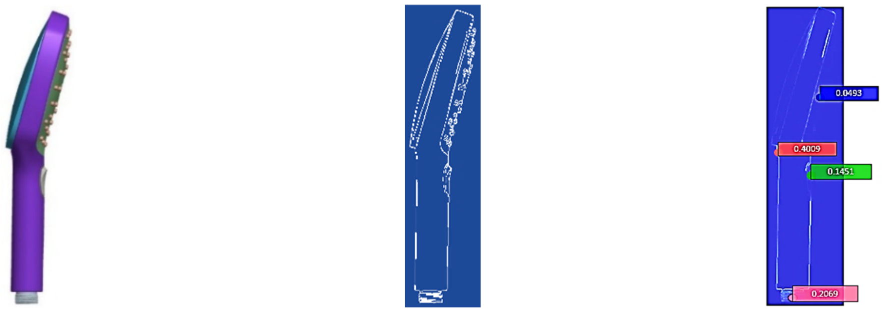
According to the computation results, the shower faceplate feature B_1achieves its maximum core feature value in the side view V_2, while the outer contour feature B_2attains its maximum value in the main view V_1. These two values are therefore selected as the representative core feature scores for the corresponding features. Since $\mathrm{Core}(V_2, A_1) > \mathrm{Core}(V_1, A_2)$, it can be concluded that the shower faceplate constitutes the core design feature of this shower product.

Similarly, take the calculation process of another shower product as an example. This shower mainly consists of 4 parts: shower panel, button, outer contour, and threaded part. Import the product views into the already trained model, output its views, contour images, and attention images. At the same time, mark the attention scores of each part feature of this shower on the views. The results are shown in the Figure 24, 25, and 26, and the attention scores for each feature of this product in different views are obtained as shown in Table 19.



（a） Front view (b) Contour image (c) Attention image

Figure 24. Product front view results.



（a） Side view (b) Contour image (c) Attention image

Figure 25. Product side view results.

（a） Rrear view (b) Contour image (c) Attention image

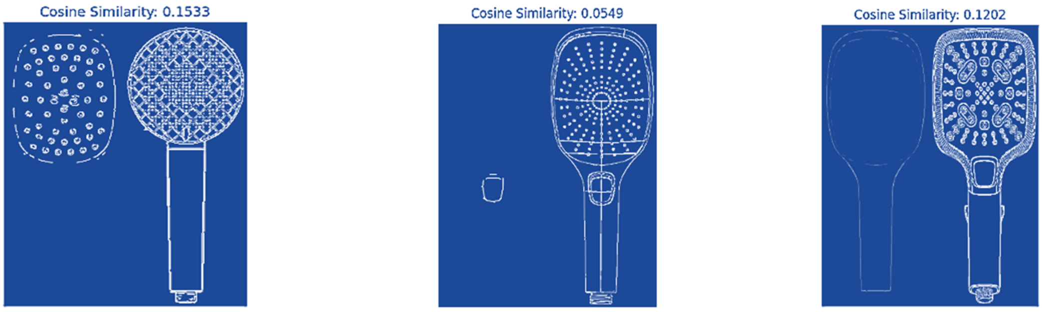
Figure 26. Product rear view results.

Table 19. Attention score results for each feature.

Feature / View	Shower Panel (B_1)	Button (B_2)	Outer Contour (B_3)	Thread (B_4)
Front View V_1	0.1241	0.1299	0.3727	0.2259
Side View V_2	0.0493	0.1451	0.4009	0.2069
Rear View V_3	0	0	0.4407	0.1827

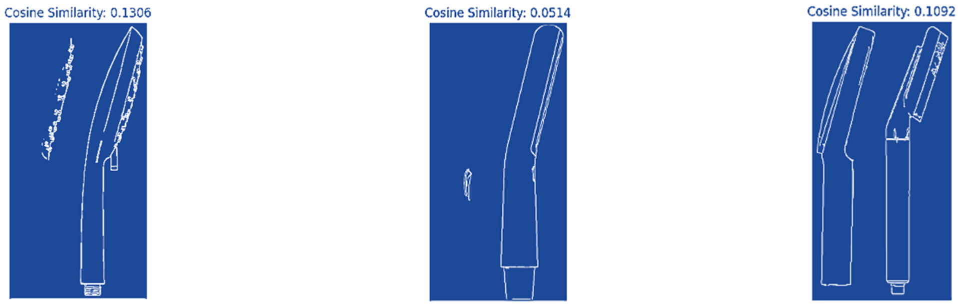
Obviously, the attention regions are primarily concentrated on the outer contour, followed by the threaded section, and subsequently on other features such as the shower faceplate and buttons. Since functional features are not considered within the scope of design patent protection, the threaded section is excluded from further computation. Accordingly, the shower faceplate is denoted as B_1, the button as B_2, the outer contour as B_3, and the threaded as B_4.

In addition, comparisons with other patent, design patent is required to satisfy the novelty criterion. The attention maps of the target shower features are therefore compared with those in the shower design patent database using similarity analysis. Taking the above product as an example, all shower views in the database are traversed to identify the most similar instance, and the corresponding similarity score is reported. The results are presented below.



（a）Feature B_1 (b) Feature B_2 (c) Feature B_3

Figure 27. Front view feature similarity result.



（a）Feature B_1 (b) Feature B_2 (c) Feature B_3

Figure 28. Left view feature similarity results.

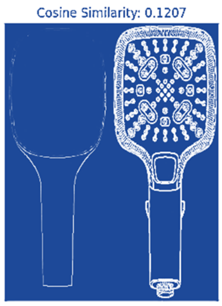


Figure 29. Rear view feature similarity result.

Based on the obtained feature attention scores and similarity results, the core feature values of each product feature under different views are computed according to the defined formulation. For the main view V_1, the core feature values are $\mathrm{Core}(V_1, B_1) = 0.81$, $\mathrm{Core}(V_1, B_2) = 2.37$, and $\mathrm{Core}(V_1, B_3) = 3.10$. For the side view V_2, the corresponding core feature values are $\mathrm{Core}(V_2, B_1) = 0.38$, $\mathrm{Core}(V_2, B_2) = 2.82$, and $\mathrm{Core}(V_2, B_3) = 3.67$. In addition, since the rear view contains only the outer contour feature, only the corresponding core feature value is computed, yielding $\mathrm{Core}(V_3, B_3) = 3.65$.

According to the computation results, $\mathrm{Core}(V_1, B_1)$, $\mathrm{Core}(V_2, B_2)$, and $\mathrm{Core}(V_2, B_3)$ are selected as the representative core feature values for the shower faceplate, button, and outer contour, respectively. Among the three views, the outer contour feature in the side view exhibits the highest core feature value. Therefore, it can be concluded that the outer contour constitutes the core design feature of this product.

 doi:10.20944/preprints202603.2216.v1

6. Discussion

To further validate the effectiveness of the proposed method, it is necessary to compare its results with those of existing real-world legal judgments. Accordingly, one representative shower design patent dispute cases are introduced for analysis and discussion.

6.1. Combined Determination with DPIJ Document

The representative case is the design patent infringement dispute between Grohe AG and China Jianlong Sanitary Ware Co., Ltd. (Case No.: (2015) Min Ti Zi No. 23) (民提字第23号). The views of the design patent (Application No. ZL200630113512.5) at issue are shown in Figure 30, and the appearance of the accused infringing product is illustrated in Figure 31.

（a）Stereogram

（b）Front view

（c）Side view

(d) Rear view

Figure 30. Patented product view.

Figure 31. The accused product alleged to infringe the design patent.

The Supreme People's Court of the People's Republic of China held that the shape of the spray surface constitutes one of the design features of the authorized design patent at issue. With respect to the shower submitted of Jianlong Corporation., the main view and the reference view of the

 doi:10.20944/preprints202603.2216.v1

product in use show that both ends of the spray surface are rectangular rather than arcuate, indicating that the spray surface is not racetrack-shaped.

Regarding the parts of the authorized design patent that are easily observable during normal use, the Court determined that the spray head, the handle, and the connection between them are all parts that can be readily observed by consumers. Furthermore, the designer of the authorized design patent deliberately adopted a racetrack-shaped button at the handle position in order to visually coordinate with the racetrack-shaped spray surface, thereby enhancing the overall aesthetic appearance of the product.

Finally, with respect to whether the appearance design of the accused product is identical or similar to the authorized design patent, the Supreme People's Court concluded that although both designs feature a highly similar racetrack-shaped spray surface, there are significant differences in the design characteristics related to the shape transitions of the spray head and its various surfaces, resulting in clearly distinct overall design styles. In addition, from the left-side view, the accused product exhibits a relatively large oblique angle at the connection between the spray head and the handle, which is neither identical nor similar to that of the authorized design. Consequently, the appearance design of the accused product does not fall within the scope of protection of the authorized design patent.

From the above judgment of the Court, it can be observed that the judicial assessment focuses on three core elements: the layout of the spray surface, the button, and the overall contour. The Court's ruling indicates that the spray surface design, particularly the racetrack-shaped configuration, can be regarded as a core design feature that determines the product's overall visual style. With respect to the push-button feature, the Court held that it is not merely functional but also possesses aesthetic attributes that contribute to the visual effect of the product. In addition, the contour characteristics at the connection between the handle and the spray head were also recognized as having a non-negligible visual impact.

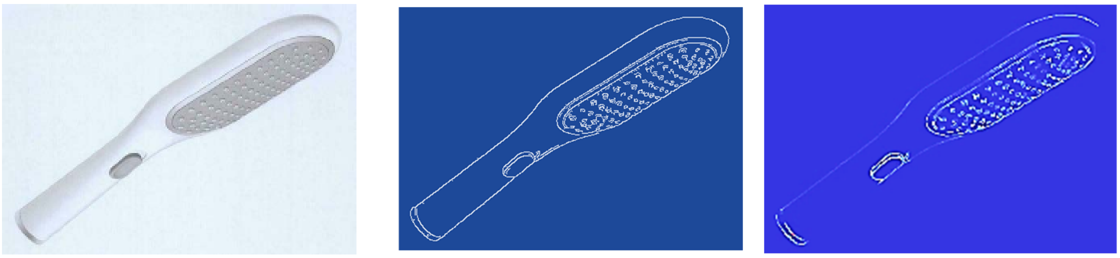
Accordingly, it can be concluded that the most salient feature of the product is the racetrack-shaped spray surface, followed by the visually influential push-button design and certain contour features. The views of this product are then input into the proposed model. Since the three-dimensional view provides a more comprehensive representation of the design features, it is selected for analysis. The resulting attention visualizations are shown in the following figure, and the corresponding quantitative evaluation metrics are summarized in the accompanying Table 20.

Table 20. Calculation results of product feature.

View	Feature	Attention score	Similar value	Core
Front view	Shower Panel	0.5853	0.2036	2.87
	Button	0.5163	0.1831	2.81
	Outer Contour	0.4516	0.1889	2.39
Side view	**Shower Panel**	**0.6442**	**0.1504**	**4.28**
	Button	0.5287	0.1305	4.05
	Outer Contour	0.4259	0.1404	3.03
Rear view	Shower Panel	0	0	0
	Button	0	0	0
	Outer Contour	0.4576	0.1889	2.42

As shown in Figure 32, the model's attention is primarily concentrated on the spray surface, the push button, and the contour regions around the connection between the shower handle and the head. These regions correspond closely to the features identified by the court as having a significant impact on the overall visual effect. In terms of attention weights, the model focuses predominantly on the spray surface and the push-button feature. The quantitative results indicate that the spray surface design is identified as the core feature, which is consistent with the court's conclusion that the racetrack-shaped spray surface constitutes a decisive design feature determining the product's visual style and should be given particular emphasis in similarity assessment.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 March 2026 doi:10.20944/preprints202603.2216.v1



（a）Product image （b）Contour image （c）Attention image

Figure 32. Product stereogram attention area results.

It should also be noted that the Court recognized the push-button feature as having visual influence in addition to its functional role. Using the proposed method, this feature also receives relatively high attention scores in both the main and side views. This further demonstrates that the proposed approach is applicable to judicial decision-making scenarios and can provide accurate, interpretable support for design patent infringement analysis.

6.2. Applications and Limitation

6.2.1. Applications

Beyond design patents, core feature identification can also be applied to other forms of intellectual property infringement analysis involving visual data, such as utility patents, image copyright, trademark rights, and integrated circuit layout design rights. For example, trademark infringement disputes frequently involve cases of similar or imitated trademarks, where effective adjudication likewise requires intelligent identification of core visual features. In this regard, the proposed core feature–based approach provides a generalizable solution for a wide range of image-based intellectual property protection scenarios, representing another important application of this study.

In addition, with respect to contour extraction and novelty evaluation, the proposed algorithms are also applicable to general 2D image processing tasks, such as product engineering drawings. For instance, existing technical drawings can be processed to extract 2D contour information and further converted into vector representations, which not only facilitates efficient storage but also provides a convenient basis for subsequent processing and analysis.

Regarding the attention-based approach, the proposed method can be extended to machine vision applications, where core object features are identified from a contour-based perspective. This enables faster recognition while simultaneously improving recognition accuracy, thereby demonstrating the broader applicability of the method beyond design patent analysis.

6.2.2. Limitations of the Research

This paper is applicable to the identification of core features of design patents that only protect shape, and does not involve design patents that simultaneously protect color, such as pattern design patents. The proposed algorithm is easily affected by color factors, which is also a problem that needs to be solved in the future.

At the same time, accuracy still needs further improvement, mainly concentrated in similarity, which is easily affected by noise. This is also an area that needs improvement in the future.

Another focus of design patent infringement is functionality and non-functionality, which is also not covered in this research and is a research problem that the team needs to solve in the future.

7. Conclusion

To overcome the problems of strong subjectivity and low efficiency of traditional judgment based on manual experience, and to achieve accurate identification of product core features, this

 doi:10.20944/preprints202603.2216.v1

paper proposes for the first time a method that integrates convolutional neural networks with attention mechanisms and combines similarity algorithms for product core feature identification. In addition, according to the characteristics of design patent documents, a method for extracting views from PDF files is proposed. Using the extracted data, a CBAM-Resnet50 model is constructed and trained. By outputting model attention maps and conducting similarity comparisons, the computer determines which feature is the core feature. Finally, professionals judge whether it is accurate and calculate the accuracy rate, thereby obtaining the conclusion that this method can be used for auxiliary judgment, and the effect is good.

Author Contributions: Conceptualization, S.Z., L.X., and R.X.; methodology, W.L.; software, W.L. and LX.; validation, W.L.; formal analysis, S.Z. and W.L.; investigation, S.Z. and W.L.; resources, W.L.; data curation, L.X.; writing—original draft preparation, S.Z., W.L., and L.X.; writing—review and editing, W.L. and S.Z.; visualization, W.L. and L.X.; supervision, S.Z., W.L., and W.L.; project administration, S.Z., W.L., and R.X.; funding acquisition, S.Z., W.L., and R.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China under grant No. 52275249, and the Social Science Foundation of Fujian Province under grant No. FJ2023B077.

Data Availability Statement: Data are contained within the article and supplementary materials.

Conflicts of Interest: The authors declare no conflicts of interest.

Note

1. https://wenshu.court.gov.cn (accessed on 23 March 2026)
2. http://gongbao.court.gov.cn/Details/4b8770fa8e9ced993f7a4b2edffeb8.html (accessed on 23 March 2026)
3. https://www.faxin.cn/ (accessed on 23 March 2026)
4. www.Patsnap.com(accessed on 23 March 2026)
5. https://www.wjx.top/vm/PpzuWzq.aspx#(accessed on 23 March 2026)
6. https://www.wjx.top/vm/PpzuWzq.aspx# (accessed on 23 March 2026)

References

1. Creusen, M.E.H.; Schoormans, J.P.L.. The different roles of product appearance in consumer choice. *J. Prod. Innov. Manag.* **2005**, *22*, 63–81.
2. Hu, H.; Liu, Y.; Lu, W.F.; Guo, X.. A quantitative aesthetic measurement method for product appearance design. *Adv. Eng. Inform.* **2022**, *53*, 101644.
3. Zeng, S.; Wang, T.; Lin, W.; Chen, Z.; Xiao, R. A Patent Mining Approach to Accurately Identifying Innovative Industrial Clusters Based on the Multivariate DBSCAN Algorithm. *Systems* **2024**, 12, 321.
4. Bartholomew, M.. Nonobvious design. *SSRN Electronic Journal* **2022**, 601–649.
5. Fackrell, S.. Uncreative designs. *Duke Law Journal* **2024**, *73*, 1437.
6. Villata, S.; Araszkiewicz, M.; Ashley, K.; Bench-Capon, T.; Branting, L.K.; Conrad, J.G.; Wyner, A. Thirty years of artificial intelligence and law: the third decade. *Artif. Intell. Law* **2022**, *30*, 561–591.
7. Buccafusco, C.; Lemley, M.A.. Screening functionality. *Virginia Law Review* **2016**, *103*, 1293–1378.
8. Buccafusco, C.; Lemley, M.A.; Masur, J.S.. Intelligent design. *Duke Law Journal* **2017**, *68*, 75.
9. Kielar, S.E.. Loosening the grasp of restriction by allowing multiple claims in design patents. *Buffalo Law Review* **2020**, *68*, 1169–1213.
10. Merrill, J.D.. Leveraging design patents and the Japanese Design Act to expand a global IP portfolio. *Harter Secrest & Emery LLP* **2024**.
11. Bahdanau, D.; Cho, K.; Bengio, Y.. Neural machine translation by jointly learning to align and translate. *arXiv* **2014**, arXiv:1409.0473.

12. Xu, K.; Ba, J.; Kiros, R.; Cho, K.; Courville, A.; Salakhudinov, R.; Zemel, R.; Bengio, Y.. Show, attend and tell: Neural image caption generation with visual attention. *In Proceedings of the 32nd International Conference on Machine Learning* **2015**, 2048–2057.
13. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I.. Attention is all you need. *Advances in Neural Information Processing Systems 30* **2017**.
14. Wang, F.; Jiang, M.; Qian, C.; Yang, S.; Li, C.; Zhang, H.; Wang, X.; Tang, X.. Residual attention network for image classification. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* **2017**, 3156–3164.
15. Zhang, H.; Zhao, Z.; Yang, J.. Attention-based prohibited item detection in X-ray images during security checking. *IET Image Processing* **2024**, *18*, 1119–1131.
16. Ji, X.; Tang, L.; Chen, L.; Hao, L.Y.; Guo, H.. Toward efficient and lightweight sea–land segmentation for remote sensing images. *Engineering Applications of Artificial Intelligence* **2024**, *135*, 108782.
17. Ye, Y.; Zhou, H.; Yu, H.; Hu, H.; Zhang, G.; Hu, J.; He, T.. Application of Tswin-F network based on multi-scale feature fusion in tomato leaf lesion recognition. *Pattern Recognition* **2024**, *156*, 110775.
18. Chen, W.; Miao, Z.; Qu, Y.; Shi, G.. HRDLNet: A semantic segmentation network with high-resolution representation for urban street view images. *Complex & Intelligent Systems* **2024**, *10*, 7825–7844.
19. Wu, K.; Sun, S.; Sun, Y.; Wang, C.; Wei, Y.. RBS-YOLO: A lightweight YOLOv5-based surface defect detection model for castings. *IET Image Processing* **2025**, 19, e70018.
20. Fayaz, M.; Dang, L.M.; Moon, H.. DLAN: A dual attention network for effective land cover classification in remote sensing. *Knowledge-Based Systems* **2025**, 113620.
21. Li, W.; Muhammat, M.; Xu, X.; Aysa, A.; Ubul, K.. Multi-scale CNN-CrossViT network for offline handwritten signature recognition and verification. *Complex & Intelligent Systems* **2025**, *11*, 400.
22. Shi, Y.; Li, F.; Zhao, S.; Yu, H.; Chen, X.; Liu, Q.. IAP-TransUNet: Integration of the attention mechanism and pyramid pooling for medical image segmentation. *Frontiers in Neurorobotics* **2025**, *19*, 1706626.
23. Zhao, Q.; Wu, R.; Chen, M.; Tsui, K.-L.; Zhao, Y.. MIEF-Net: Multimodal image-enhanced fusion network for intelligent fall risk prediction. *Neural Networks* **2025**, 108260.
24. Su, J.; Liang, J.; Zhu, J.; Li, Y.J.S.. HCAM-CL: A novel method integrating a hierarchical cross-attention mechanism with CNN-LSTM for hierarchical image classification. *Symmetry* **2024**, *16*, 1231.
25. Bhatti, N.; Hanbury, A.. Image search in patents: A review. *International Journal on Document Analysis and Recognition* **2013**, *16*, 309–329.
26. Rusiñol, M.; de las Heras, L.P.; Terrades, O.R.. Flowchart recognition for non-textual information retrieval in patent search. *Inf. Retr.* **2014**, *17*, 545–562.
27. Chen, J.; Li, M.; Jin, Q.; Bao, S.; Su, Z.; Yu, Y.. Lead curve detection in drawings with complex cross-points. *Neurocomputing* **2015**, *168*, 35–46.
28. Riedl, C.; Zanibbi, R.; Hearst, M.A.; Zhu, S.; Menietti, M.; Crusan, J.; Metelsky, I.; Lakhani, K.R.. Detecting figures and part labels in patents. *International Journal on Document Analysis and Recognition* **2016**, *19*, 155–172.
29. Lin, Z.C.; Wu, D.W.; Hong, G.E.. Combination of improved cosine similarity and patent attribution probability method. *Adv. Eng. Inform.* **2016**, *30*, 26–38.
30. Bhatti, N.; Hanbury, A.; Stottinger, J.. Contextual local primitives for binary patent image retrieval. *Multimed. Tools Appl.* **2018**, *77*, 9111–9151.
31. Trappey, A.J.C.; Trappey, C.V.; Shih, S.. An intelligent content-based image retrieval methodology using transfer learning for digital IP protection. *Adv. Eng. Inform.* **2021**, *48*, 101291.
32. Jiang, S.; Luo, J.; Ruiz-Pava, G.; Hu, J.; Magee, C.L.. Deriving design feature vectors for patent images using convolutional neural networks. *Journal of Mechanical Design* **2021**, *143*, 061405.
33. Li, M.; Li, H.. Application of deep convolutional neural network under region proposal network in patent graphic recognition and retrieval. *IEEE Access* **2021**, *10*, 37829–37838.
34. Trappey, A.J.C.; Trappey, C.V.; Lin, E.. Intelligent trademark recognition and similarity analysis using a two-stage transfer learning approach. *Adv. Eng. Inform.* **2022**, *52*, 101567.
35. Zhang, L.; Gao, Q.; Liu, M.; Gu, Z.; Lang, B.. Measuring similarity of dual-modal academic data based on multi-fusion representation learning. *IEEE Access* **2024**, 10.1109/ACCESS.2024.3427731.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 March 2026 doi:10.20944/preprints202603.2216.v1

36. Jing, L.; Zhou, C.; Feng, D.; Dou, Y.; Fan, X.; Jiang, S.. A patent infringement analysis approach based on patent knowledge graph driven fusion of graph and image similarity. *IEEE Access* **2025**, 10.1109/ACCESS.2025.3541874.
37. Kim, H.; Gim, G.. Enhancing patent document similarity evaluation and classification precision through a multimodal AI approach. *Appl. Sci.* **2025**, *15*, 9254.
38. Xie, X.; Wu, J.; Xiang, M.; Tang, J.; Sheng, Y.. Enhancing the efficiency of patent classification: A multimodal classification approach for design patents. *Journal of King Saud University – Computer and Information Sciences* **2025**, *37*, 183.
39. Kavimandan, K.; Nalmpantis, A.; Beauxis-Aussalet, E.; Sips, R.-J.. Hierarchical multi-positive contrastive learning for patent image retrieval. *arXiv* **2025**, arXiv:2506.13496.
40. Amoncio, E.; Chan, T.; Storz, C.. Using computer vision to measure design similarity: An application to design rights. *Research Policy* **2025**, *54*, 105309.
41. Lin, Y.H.; Hung, M.C.; Lee, C.F.. Density-refine: Patent image retrieval by density-based region extraction and feature fusion. *Journal of Mechanical Design* **2025**, *147*, 081703.
42. Lin, W.; Yu, W.; Xiao, R.. Measuring patent similarity based on text mining and image recognition. *Systems* **2023**, *11*, 294.
43. Adhikari, N.S.; Agarwal, S.. A comparative study of PDF parsing tools across diverse document categories. *arXiv* **2024**, arXiv:2410.09871.
44. Jing, J.; Liu, S.; Wang, G.; Zhang, W.; Sun, C.. Recent advances on image edge detection: A comprehensive review. *Neurocomputing* **2022**, *503*, 259–271.
45. Senel, H.G.. Gradient estimation using wide support operators. *IEEE Trans. Image Process.* **2009**, *18*, 867–878.
46. Muntarina, K.; Mostafiz, R.; Khanom, F.; Shorif, S.B.; Uddin, M.S.. MultiResEdge: A deep learning-based edge detection approach. *Intelligent Systems with Applications* **2023**, *20*, 200274.
47. Xu, Q.; Varadarajan, S.; Chakrabarti, C.; Karam, L.J.. A distributed Canny edge detector: Algorithm and FPGA implementation. *IEEE Trans. Image Process.* **2014**, *23*, 2944–2960.
48. Saravanan, C.. Color image to grayscale image conversion. *Proceedings of the Second International Conference on Computer Engineering and Applications* **2010**, 196–199.
49. He, K.; Zhang, X.; Ren, S.; Sun, J.. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* **2016**, 770–778.
50. Guo, M.H.; Xu, T.X.; Liu, J.J.; Liu, Z.N.; Jiang, P.T.; Mu, T.J.; Zhang, S.H.; Martin, R.R.; Cheng, M.M.; Hu, S.M.. Attention mechanisms in computer vision: A survey. *Computational Visual Media* **2022**, *8*, 331–368.
51. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S.. CBAM: Convolutional block attention module. *Proceedings of the European Conference on Computer Vision* **2018**, 3–19.
52. Cheng, Y.; Guan, R.; Zhou, S.; Zhou, X.; Xue, J.; Zhai, W.. Research on tool wear and breakage state recognition based on ResNet-CBAM. *Measurement* **2025**, *242*, 116105.
53. Ma, J.; Jiang, X.; Fan, A.; Jiang, J.; Yan, J.. Image matching from handcrafted to deep features: A survey. *International Journal of Computer Vision* **2021**, *129*, 23–79.
54. Simonyan, K.; Zisserman, A.. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
55. Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q.. A comprehensive survey on transfer learning. *Proceedings of the IEEE* **2020**, *109*, 43–76.
56. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L.. ImageNet: A large-scale hierarchical image database. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* **2009**, 248–255.
57. Wang, X.; Yue, Q.; Liu, X.. SBDNet: A deep learning-based method for the segmentation and quantification of fatigue cracks in steel bridges. *Adv. Eng. Inform.* **2025**, *65*, 103186.
58. Shiri, F.M.; Perumal, T.; Mustapha, N.; Mohamed, R.. A comprehensive overview and comparative analysis on deep learning models: CNN, RNN, LSTM, GRU. *arXiv* **2025**, arXiv:2305.17473.
59. Bhattacharjee, A.; Murugan, R.; Goel, T.; Mirjalili, S.. Pulmonary nodule segmentation framework based on fine-tuned and pretrained deep neural networks using CT images. *IEEE Transactions on Radiation and Plasma Medical Sciences* **2023**, *7*, 394–409.

 doi:10.20944/preprints202603.2216.v1

60. Huang, J.; Ling, C.X.. Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on Knowledge and Data Engineering* **2005**, *17*, 299–310.
61. Deroy, A.; Ghosh, K.; Raj, S.; Sarda, J.; Shrivastava, R.; Shrivastava, S.; Goyal, P. Investigating legal question generation using large language models. *Artif. Intell. Law* **2025**, doi:10.1007/s10506-025-09452-y.