Product Design Sketch Decision Model Based on Improved Capsule Network

Zhou Qingyan[1], Zhang Jincheng[2], Wei Tangwei[3], Li Hao[4] ,Wang Jing[5] and Bao Xiaozhong[6]

[1,2,3,5]School of Information and Artificial Intelligence, Anhui Agricultural University,
[1]871066225@qq.com, [2]zhangjincheng50@foxmail.com, [3]1784725066@qq.com,
[5]2519621283@qq.com
[4] China Telecom Anhui Branch, 741123734@qq.com
[6] Keyi College of Zhejiang Sci-Tech University, bxz@ky.zstu.edu.cn

Corresponding author: Zhou Qingyan, 871066225@qq.com

Abstract: The outstanding capabilities of neural networks have led to their increasing use in combination with various other fields. However, their application in sketch design remains limited. To address the challenge of making decisions based on design sketches in the product design process, this paper proposes an innovative model called the Product Design Sketch Decision Network (PDSDNet). This model is based on an enhanced capsule network and introduces a reverse dot product attention routing mechanism. The ConvNeXt serves as the backbone network to improve the design sketch image's feature extraction ability, and the ADAMW optimizer further enhances the model's training speed and stability. To validate the accuracy of the model's decisions, a dataset of design sketches for table lamps was established. Experts evaluated the dataset and supplemented it with semantic tags and key region image segmentation annotations. PDSDNet was compared with AlexNet, ResNet50, and the classic capsule network. The experimental results revealed that in the decision-making process for design area sketch styling semantics, the decision accuracy of PDSDNet and ResNet50 was higher compared to the other two models. PDSDNet was more accurate in the semantic and feasibility decision-making of overall sketch modeling, with F1 scores of 0.81 and 0.96, respectively. These results demonstrate that the model focuses on abstract features and subtle semantic information.

Keywords: Product design; sketch; capsule network.
DOI: https://doi.org/10.14733/cadaps.2025.369-383

1 INTRODUCTION

Industrial design, as a crucial component of innovative design, aids in establishing a connection between a product's multifunctional attributes and the subjective perception of users, thereby facilitating the realization of value in innovative outcomes [14]. Despite the availability of many digital design methods, many designers still prefer to use sketches during the product design stage for analysis and communication with clients. With the widespread application of computer assistance in product design, numerous researchers have utilized relevant technologies to improve efficiency and accuracy in decision-making for product design sketches. Press and Cooper [9] detailed the decision-making process in product design, while Schmid, et al. [11] focused on analyzing and researching sketch strokes. Fonseca, et al. [5] achieved parameterization and recognition of sketch images through topological constraints. Yan, et al. [16] used generative adversarial networks to facilitate the computer-assisted transformation of sketches into final images using AI. These studies have somewhat enhanced the accuracy of sketch decision-making but often involve transforming sketches into another form of data, leaving a gap in decision-making model research focused directly on the sketches themselves.

Deep learning, which allows computational models with multiple processing layers to learn and represent data with multiple levels of abstraction, mimics how the brain perceives and understands multimodal information. This approach implicitly captures the complex structure of large-scale data, with computer vision being one of the most prominent use cases of deep learning [15]. Currently, computer vision methods like convolutional neural networks [3] and recurrent neural networks [18] have been successfully applied in the field of industrial design decision-making for product design images. However, these decision-making models require a large number of design images containing detailed design information and primarily utilize images of finished product designs, which limits their early-stage assistive role in product design.

The capsule network is a new concept in deep learning proposed by Sabour, et al. [10]. Unlike traditional Convolutional Neural Networks (CNNs) that only recognize local pixel features in images without understanding the hierarchical relationships of these features, capsule networks address this issue. Additionally, one of the significant benefits of capsule networks is that they require learning only a small portion of data compared to CNNs to achieve good recognition performance. Essentially, a capsule network is a collection of neurons, with its basic unit being a capsule containing a vector. Thus, capsules can indicate not only the probability of the existence of features but also the direction of different features.

Building on the traditional capsule network, Tsai, et al. [13] introduced a new inverted dot-product attention routing algorithm. In this algorithm, unlike the dynamic routing mechanism of traditional capsule networks where lower-level capsules compete for the attention of higher-level capsules, the routing decision depends on the consistency between the pose of higher-level capsules and the votes formed by lower-level capsules for these higher-level poses. This approach also transforms the sequential iterative routing in the original capsule network into concurrent iterative routing. Experiments have shown that these changes can effectively improve the performance of the capsule network while reducing some of the training parameters.

The inverted dot-product attention routing capsule network has already seen various applications. For instance, Paoletti, et al. [8] utilized this network to propose a new method for hyperspectral image classification, addressing the issue of capsule networks not being able to correctly model the hierarchical spectral relationships between different images. Dinani and Caragea [4] significantly improved the accuracy of classifying disaster images on social networks as informative or non-informative using the inverted dot-product attention routing capsule network. This enhancement aids in more accurately filtering out messages seeking help or reporting about disasters when they occur. Ng and Liu [7] applied the inverted dot-product attention routing along with the Tanimoto loss function to improve the accuracy of voice emotion recognition.

This paper addresses the aforementioned issues by proposing a product design sketch decision model based on an improved capsule network - PDSDNet (Product Design Sketch Decision Network). This model utilizes the ConvNeXt backbone network and inverted dot-product attention routing and employs the AdamW optimizer to enhance the capsule network. This approach overcomes the limitations of previous product design decision models, which struggled with making decisions based on the sketches themselves and were not applicable during the sketching stage of product design. The following sections summarize the parameters used in the various pre-set styles. You should never really have to deal with these parameters directly, but only select one of the pre-formatted styles. But for completeness of documentation, this information is given here so that future managers of this template may more easily make selective changes to some of the styles.

2 PRODUCT DESIGN SKETCH DECISION MODEL BASED ON PDSDNET

2.1 PDSDNet Network Model Structure

The basic structure of the PDSDNet network used in this model is illustrated in Figure 1. Images that have been preprocessed and semantically annotated are fed into the PDSDNet. Initially, the ConvNeXt backbone network is used to extract image features, which are then inputted into the precursor capsules of the capsule network. This part then serves as the main backbone and is frozen. Subsequently, the data passes through two convolutional capsule layers and a fully connected capsule layers. Finally, a softmax function is applied to the fully connected layers to obtain the classification results.

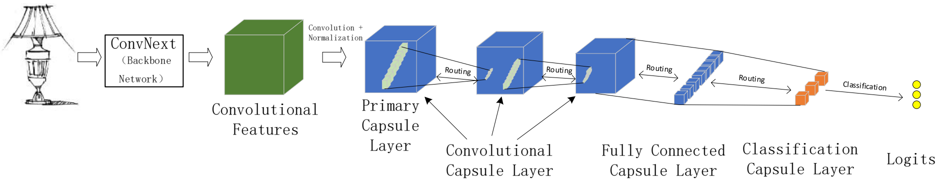


Figure 1: PDSDNet network architecture.

2.1.1 ConvNeXt Backbone Network

In the original capsule network architecture, multiple ResNet computation blocks or a single convolutional layer are used as the backbone network. However, considering that design sketches often contain less explicit information, a backbone network with stronger feature extraction capabilities and a more comprehensive architecture is needed. Therefore, in PDSDNet, we use ConvNeXt instead of the traditional ResNet as the model's backbone network, with a structure as Figure 2. Compared to the ResNet backbone, ConvNeXt also adopts a modular stacked design, but with a different stacking ratio in its four modules, diverging from ResNet's [3:4:6:3] ratio. ConvNeXt utilizes an optimized [1:1:9:1] network structure, allowing later modules to have more computational power [6]. In the design of each Block, unlike the bottleneck layer design of the ResNet Block, the ConvNeXt Block uses an inverted bottleneck layer design, adjusting the computational load in each layer of the Block, making it easier for the network to propagate gradients and reduce the problem of vanishing gradients that may occur. To compensate for the reduced model capacity that comes with using group convolution to decrease the model's computational load, the number of channels is increased from 64 in ResNet to 96 in ConvNeXt, achieving higher precision while reducing the overall computational load of the model. To further optimize the accuracy of the network, certain optimizations were also made to the ConvNeXt backbone network: less frequent use of activation functions, switching from ReLU to GeLU functions; using fewer normalization layers, adding separate downsampling layers, and replacing batch normalization with layer normalization to minimize errors that may arise from batch normalization when the batch size is small.

Considering the issue of limited information content in sketches, PDSDNet opts for the ConvNeXt-S scale of the ConvNeXt network as the backbone of the model. This choice ensures adequate semantic feature extraction while maintaining an appropriate network depth. By enhancing the feature extraction capability, the model's accuracy in making decisions about design sketches is improved.

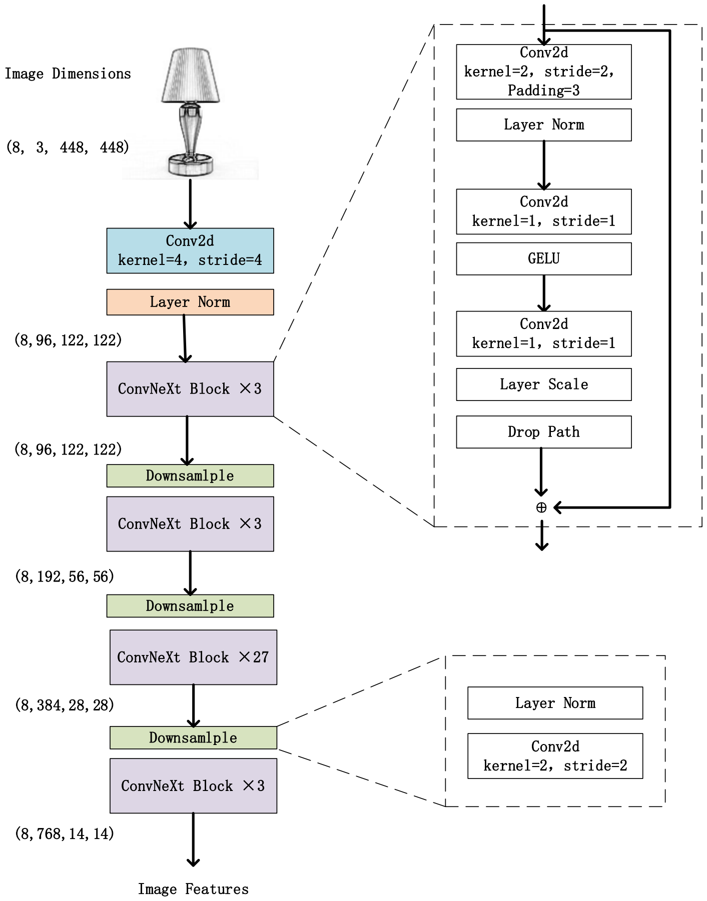


Figure 2: ConvNeXt backbone network structure.

2.1.2 Improved Capsule Network

In traditional capsule networks, the routing between parent and child capsule layers typically involves sequential iterative routing. In this process, child capsules pass input vectors to parent capsules, and the parent capsules dynamically adjust their weights and update their capsule states based on the received information. The structure of capsules, compared to traditional Convolutional Neural Networks (CNNs), handles ambiguity well and can reduce the data required for training to some extent. However, this structure can lead to training difficulties and hyperparameter sensitivity due to a large number of input parameters and redundant features.

Given that product design sketches are often composed of multiple independent design elements and contain less information, and the differences between sketches are not as pronounced as those in final product images, we have adopted an improved capsule network based on inverted dot-product attention routing. Contrary to sequential iterative routing, in this algorithm, parent capsules attempt to attract the attention of child capsules. The routing probability is derived from the pose state of the parent in the previous iteration, as well as the votes cast by the child capsules for the parent's pose state in the current iteration. The main steps of this routing method are as follows:

In the routing mechanism, l represents the layer index, and $p_i^l \in R^{s^l}$denotes the i^{th}capsule in the p^lmatrix of layer l. Inspired by the Expectation-Maximization algorithm, the pose vector defined by p_i^l is reshaped into a matrix array $R^{\sqrt{s^l}\times\sqrt{s^l}}$. Once the capsules are initialized, the routing mechanism is applied between the lower and higher levels. There are two main steps in the application of the routing mechanism: consistency calculation and pose update.

Consistency calculation involves transforming the matrix pose p_i^l into a voting $V_{i,j}^l$, to calculate the consistency between the higher-level p_j^{l+1} and the lower level p_i^l, as shown in Equation (1):

$$v_{i,j}^l = W_{ij}^l \times p_{i,j}^l \tag{1}$$

Where $W_{ij}^l \in \mathrm{R}^{\sqrt{s^{l+1}} \times \sqrt{s^l}}$is the learned transformation matrix. Once the votes for the pose p_j^{l+1} are obtained, the agreement a_{ij}^l is determined by Equation (2):

$$a_{ij}^l = (P_j^{l+1})^T \cdot v_{i,j}^l \tag{2}$$

Once the agreement is obtained, the pose at the higher level will be updated. From this perspective, the agreement coefficient is processed by the softmax function to determine the routing probabilities r_{ij}^l between the lower and higher levels.

$$r_{ij}^l = \frac{\exp\,(a_{ij}^l)}{\sum_Z \exp\,(a_{iZ}^l)} \tag{3}$$

Equation (3) defines the competition for the attention of higher-level capsules towards the lower-level capsules, where r_{ij}^l is the score obtained from the inverted dot-product attention. Once the higher-level capsules receive these scores, they update their poses using the information from the lower-level capsules, as shown in Equation (4):

$$P_j^{l+1} = Normalization(\textstyle\sum_i r_{ij}^l\, v_{i,j}^l) \tag{4}$$

In the original capsule network model, the learning rate is adjusted using the Stochastic Gradient Descent (SGD) optimizer. Although this optimizer is structurally simple and widely applicable, its convergence speed can be affected by the learning rate and may face convergence issues. Therefore, in PDSDNet, we use the AdamW optimizer instead of the traditional SGD for optimization. Compared to other optimizers, AdamW can automatically adjust the learning rate without the need for extensive parameter tuning, reducing redundancy. Additionally, it automatically adjusts the weight decay coefficient, making the decision model more stable and helping to avoid overfitting.

2.2 Design Sketch Decision Model Overall Structure

The model takes product design sketches as input, preprocesses the images, and performs semantic segmentation. The segmented key design parts are then semantically annotated to build a dataset. The annotated dataset is trained through the PDSDNet network to develop a product design sketch decision-making method. This method is used to evaluate the product design sketches in the test set and verify their consistency with human evaluative annotations, thereby assessing the performance of decision-making for product design sketches. The model mainly consists of two parts: training the model using training data and making decisions on test data based on the training results. The overall structure of the model is illustrated in Figure 3.

3 EXPERIMENTAL DESIGN

3.1 Construction of Experimental Dataset

3.1.1 Collection and Annotation of Product Design Sketches

Product designs often contain different stylistic semantics in various target areas, necessitating semantic segmentation of different styling areas for intelligent decision-making in the field of product design sketch form decisions. This requires detailed semantic segmentation of images according to the edge contours of different design areas.

Considering the difficulty in obtaining product design sketches, as well as to better reflect the differences in stylistic semantics between different design areas, this experiment uses household desk lamp design sketches as the dataset. Based on commonly used drawing techniques in product design, 102 hand-drawn sketches of desk lamp designs were collected from the works of industrial design students. These sketches were created using widely used techniques in industrial design, such as one-point and two-point perspectives. [12].

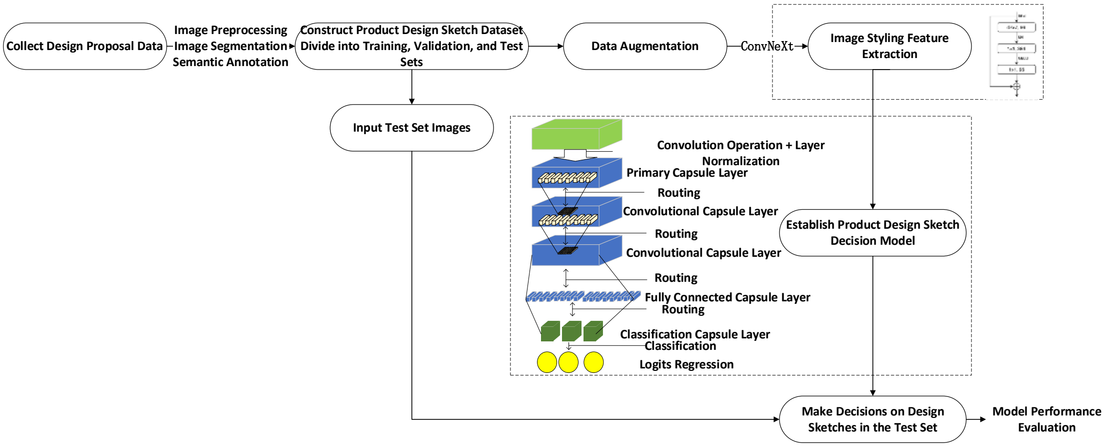

Collect Design Proposal Data
Image Preprocessing
Image Segmentation
Semantic Annotation
Construct Product Design Sketch Dataset
Divide into Training, Validation, and Test Sets
Data Augmentation
ConvNeXt
Image Styling Feature Extraction
Input Test Set Images
Convolution Operation + Layer Normalization
Primary Capsule Layer
Routing
Convolutional Capsule Layer
Routing
Convolutional Capsule Layer
Routing
Fully Connected Capsule Layer
Routing
Classification Capsule Layer
Classification
Logits Regression
Establish Product Design Sketch Decision Model
Make Decisions on Design Sketches in the Test Set
Model Performance Evaluation

Figure 3: Overall structure of the product design sketch decision model based on PDSDNet.

Given the limited volume of collected data, additional desk lamp design sketches meeting the criteria were gathered from the internet. After excluding invalid samples that did not meet the experimental requirements from this expanded dataset, a total of 500 design sketch proposals were selected. Each of these design sketch proposals is a one-point or two-point perspective grayscale sketch of a household desk lamp. Figure 4 displays a portion of the dataset.

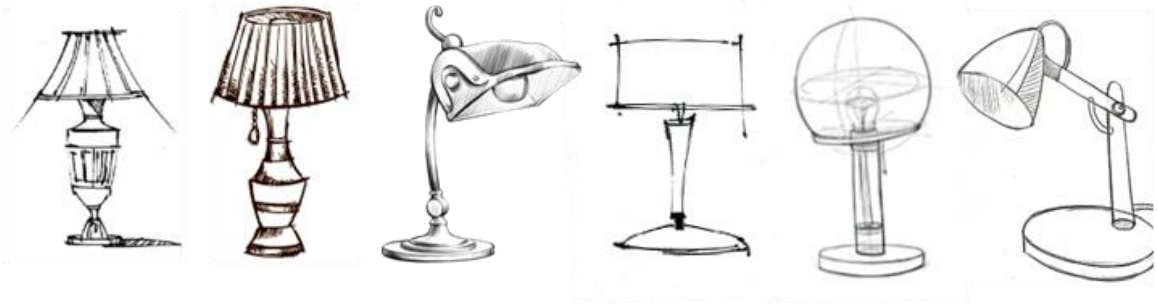


Figure 4: Partial dataset of table lamp design sketches.

After constructing the basic dataset, it is necessary to perform semantic segmentation and annotation of the images based on the edge contours of different design areas. In the case of desk lamp design proposals, both hand-drawn sketches by designers and sketches selected from the internet are generally clear, with a simple background and no obstructions. Therefore, in this experiment, the contours of the 102 design sketch proposals were initially drawn using LabelMe software. Subsequently, the efficient image segmentation network Deeplabv3+ [2] was utilized to segment and extract contours from other images, and the extracted results were manually corrected.

Based on the analysis of desk lamp-related design patents, the key parts of desk lamp design are mainly located in the lampshade, lamp pole, and base. Therefore, these three parts were annotated as Region 1 (blue), Region 2 (orange), and Region 3 (green), respectively. These three parts influence each other and collectively contribute to the overall design image of the desk lamp. The data marking areas for the key parts of the desk lamp after segmentation are illustrated in Figure 5.

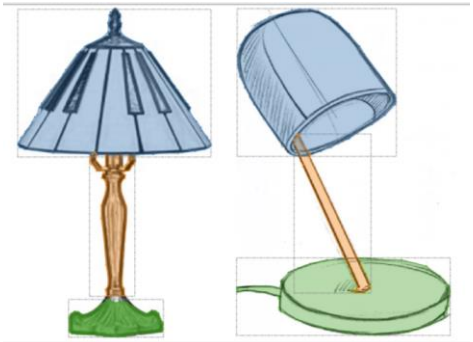


Figure 5: Key annotation areas after structural division of the desk lamp.

3.1.2 *Selection of Stylistic Semantics and Feasibility Evaluation*

Based on a broad collection of 200 image words related to lighting products from the internet and magazines, and building upon previous research [1],[17], eight pairs of image adjectives that designers focus on during desk lamp design were identified. These pairs are: simple-complex, popular-personalized, generous-elegant, geometric-streamlined, warm-cold, rugged-delicate, business-artistic, and plain-gorgeous. Given that some characteristics are not evident during the sketch design stage, certain word pairs were excluded, while new pairs such as 'classic-modern' and 'lively-quiet' were introduced. Additionally, the pair 'business-artistic' was modified to 'business-leisure,' resulting in a total of nine pairs of image adjectives.

In addition to the stylistic semantic labels for the three parts, overall styling semantic labels also need to be incorporated into the design proposals. Given that designers frequently overlook the manufacturing challenges of actual products during the sketching process, it is crucial to include feasibility labels in the design proposals, thereby offering a valuable reference for product design.

Based on the semantic segmentation of the overall design and individual components of the desk lamp sketches, four experienced designers with over seven years of experience and three teachers of sketch design courses were invited to evaluate the style of the entire dataset, including each key part. In addition to this, the experts also rated the feasibility of the sketches on a 10-point scale based on their implementability.

After the experts completed their evaluations and scoring, six style combinations were identified through cluster analysis for the overall semantic description. In the semantic description of the design area's styling, each styling area was divided into three different semantics. All semantic combinations are presented in Table 1.

Styling Area	*Semantic Description*

Overall Styling	Classic, Complex;
	Business, plain;
	Leisure, Lively;
	Warm, Quiet;
	Geometric, Cold;
	Personalized, Delicate.

Region 1 (Lampshade)	Gorgeous;
	Warm;
	Plain.

Region 2 (Lamp Pole)	Streamlined;
	Geometric;
	Complex.

Region 3 (Lamp Base)	Delicate;
	Simple;
	Rugged.

Table 1: Semantic descriptions of different styling areas of the desk lamp.

Subsequently, clustering methods labeled other semantics with similar meanings as one of the selected semantic descriptions. For feasibility, the scores from all expert reviewers were taken, and a quadratic averaging method was used to eliminate discrepancies between expert scores. The final scores were categorized as follows: 0-3.3 points defined as difficult to implement, 3.3-6.6 points as feasible, and 6.6-10 points as easy to implement.

Finally, the LabelMe software was used to complete the semantic annotation of the dataset for both types of evaluations. Table 2 presents an example of the annotated data.

Image Number	*Overall Styling Description*	*Area Styling Description*			*Feasibility Evaluation*
		Region 1 (Lampshade)	*Region 2 (Lamp Pole)*	*Region 3 (Lamp Base)*	
1	Classic, Complex	Gorgeous	Streamlined	Delicate	Difficult to Implement
2	Business, Plain	Plain	Geometric	Simple	Easy to Implement
3	Leisure, Lively	Warm	Streamlined	Simple	Easy to Implement
n	...	...	...	...	...

Table 2: Semantic annotations of desk lamp styling design sketch dataset.

3.2 Experimental Validation and Training Parameter Settings

The dataset of desk lamp design sketches comprises 500 images, each containing three-part styling semantic descriptions, one overall semantic description, and one feasibility evaluation. The dataset was randomly divided into training, validation, and test sets. The training set, consisting of 300 images, was used for training and establishing the decision model. The validation set included 100 images for adjusting the model's hyperparameters and for a preliminary assessment of the model's capabilities. The test set also contained 100 images and was used to evaluate the performance of the trained model.

Before training, all images were resized to 448×448 pixels for easy processing by the decision model. To further increase the diversity of the data and thereby enhance accuracy, the images were randomly flipped and pixel-wise normalized before training commenced. These data augmentation methods indirectly increased the diversity of the training data, improving the model's generalization ability and performance. Additionally, the ConvNeXt pre-trained model was integrated into our model. This transfer learning approach helps the model better capture high-level features of the data, improves generalization capability, and accelerates convergence speed in sketch decision tasks.

After processing, each batch of sketch images entering the PDSDNet product design sketch decision model has an input dimension of (8, 3, 448, 448), representing the batch size, color channels, and spatial dimensions. The images first pass through an initial convolution layer, where the dimension changes to (8, 96, 112, 112). Then, they proceed through the four stages of the ConvNeXt backbone network, ultimately reaching the dimensions of (8, 768, 14, 14).

In the capsule network section, the primary capsule layer further processes the images, with an output dimension of (8, 32, 14, 14, 16). This is followed by two main capsule layers, which convert the feature dimensions to (8, 32, 6, 6, 16) and (8, 32, 4, 4, 16), respectively. Finally, the classification capsule layer produces an output of (8, x, 16), where x corresponds to the number of possible styling semantic description categories.

This series of transformations demonstrates the network's ability to increase feature dimensions while reducing spatial size, reflecting the gradual abstraction of images from original pixels to high-level features. The primary capsule layer and the main capsule layer each contain 32 capsules, while the classification capsule layer includes capsules corresponding to the number of styling semantic description categories. Between capsule layers, the inverted dot-product attention routing method is used for two updates: the first using sequential routing and the second using concurrent routing. This ensures effective information extraction and processing at each network stage.

The entire experiment was conducted using Python 3.7 for programming, with the PyTorch deep learning framework. The training was performed using a 4-core CPU and an NVIDIA A4000 graphics card. The learning rate initialized with AdamW was set to 0.0005, with a total of 100 epochs of training and a batch size of 8.

3.3 Evaluation Metrics

To directly observe the performance of the decision model, we used precision, recall, accuracy, and F1 score for evaluation in the experiment. Precision represents the degree of accuracy in predicting positive sample results; recall is the probability of correctly predicting a positive sample from all positive samples. Accuracy indicates the model's predictive precision, but in cases of unbalanced sample sizes, it may not accurately reflect the model's performance. The F1 score is a crucial indicator, with higher scores indicating better overall performance of the model. The definitions of precision, recall, accuracy, and the F1 score are shown in Equation (5). To comprehensively evaluate model performance, the macro average method is used to calculate the average value of various metrics.

$$\begin{cases} \text{precision} = \frac{TP}{TP+FP} \\ \text{recall} = \frac{TP}{TP+FN} \\ \text{accuracy} = \frac{TP+FN}{TP+FP+FN+TN} \\ \mathrm{F}_1 = \frac{2TP}{2TP+FP+FN} \end{cases} \tag{5}$$

The explanations of TP, FP, FN, and TN in Equation (5) are shown in Table 3.

Confusion Matrix		*Actual Values*	
		Positive	*Negative*
Predicted Values	Positive	TP(True Positive)	FP(False Positive)
	Negative	FN(False Negative)	TN(True Negative)

Table 3: The explanations of TP, FP,FN and TN.

4 RESULTS AND DISCUSSION

To verify the decision-making performance of the PDSDNet product design sketch decision model, this experiment selected three networks for comparative analysis: the traditional capsule network, ResNet50, and AlexNet. Their decision-making performance was compared in terms of design area styling semantics, feasibility, and overall styling semantics evaluation.

4.1 Decision-Making in Design Area Styling Semantics

First, various models were employed to make decisions regarding the styling semantics of three design areas within the desk lamp design sketches. The outcomes of these decisions are presented in Table 4.

Design Area	*Decision Network Model*	*Precision*	*Recall*	*Accuracy*	*F1*
Region 1 (Lampshade)	AlexNet	0.6799	0.7083	0.6800	0.6938
	ResNet50	0.7778	0.7913	0.7800	0.7845
	CapsuleNet	0.7237	0.7476	0.7300	0.7326
	PDSDNet	0.7660	0.7820	0.7600	0.7739
Region 2 (Lamp Pole)	AlexNet	0.6087	0.5491	0.5900	0.5774
	ResNet50	0.7071	0.7157	0.7200	0.7114
	CapsuleNet	0.7753	0.6404	0.6900	0.7014
	PDSDNet	0.7563	0.7178	0.7300	0.7365
Region 3 (Lamp Base)	AlexNet	0.6486	0.6713	0.6300	0.6598
	ResNet50	0.6928	0.7130	0.6900	0.7028
	CapsuleNet	0.7237	0.7476	0.7300	0.7326
	PDSDNet	0.7387	0.7488	0.7400	0.7437

Table 4: Network decision results on different desk lamp styling design areas.

The experimental results in Figure 6 show that PDSDNet and ResNet50 achieved the highest decision accuracy among the four network models tested, with F1 scores surpassing the other two networks in most of the design areas. A comparison of decision results across different design areas reveals that all models performed better in design area 1 compared to the other two areas. This could be because designers often pay more attention to the design of the lampshade area during the design and sketching process, resulting in more design elements in this part of the image and making it easier for the model to train and make decisions. Additionally, the shaping of the lamp pole is often more ambiguous in the design process, and it occupies a smaller proportion of the overall design sketch, making it less distinctive compared to the other two design areas. Therefore, this aspect likely influences the decision-making model's accuracy in styling decisions for the lamp pole area.

4.2 Decision-Making on the Feasibility and Styling Semantics of Overall Design

In the experiment, the overall design sketches served as the subjects for decision-making. The model was initially trained and subsequently used to make decisions regarding the feasibility of the overall styling design. The results of these decisions are presented in Table 5. Different models have achieved satisfactory results in making decisions about the feasibility of the overall design, with PDSDNet slightly outperforming the other models in terms of decision accuracy. This is because the complexity of product design often has a significant relationship with its feasibility.

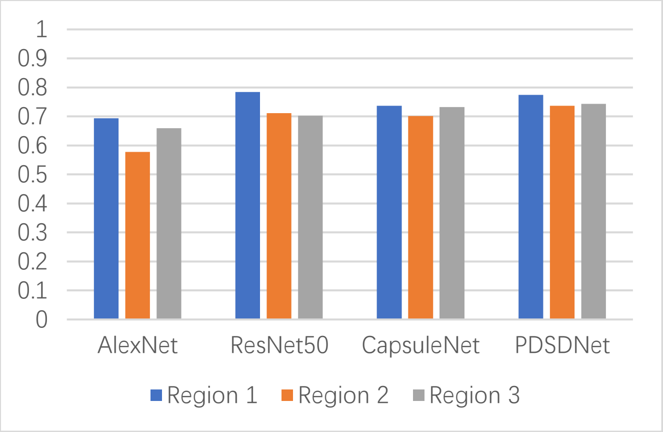


Figure 6: F1 score of the model for decision-making in different design areas.

Decision Network Model	*Precision*	*Recall*	*Accuracy*	*F1*
AlexNet	0.8200	0.8125	0.8200	0.8129
ResNet50	0.9073	0.8846	0.8900	0.8991
CapsuleNet	0.9101	0.8960	0.9000	0.9037
PDSDNet	0.9672	0.9531	0.9600	0.9654

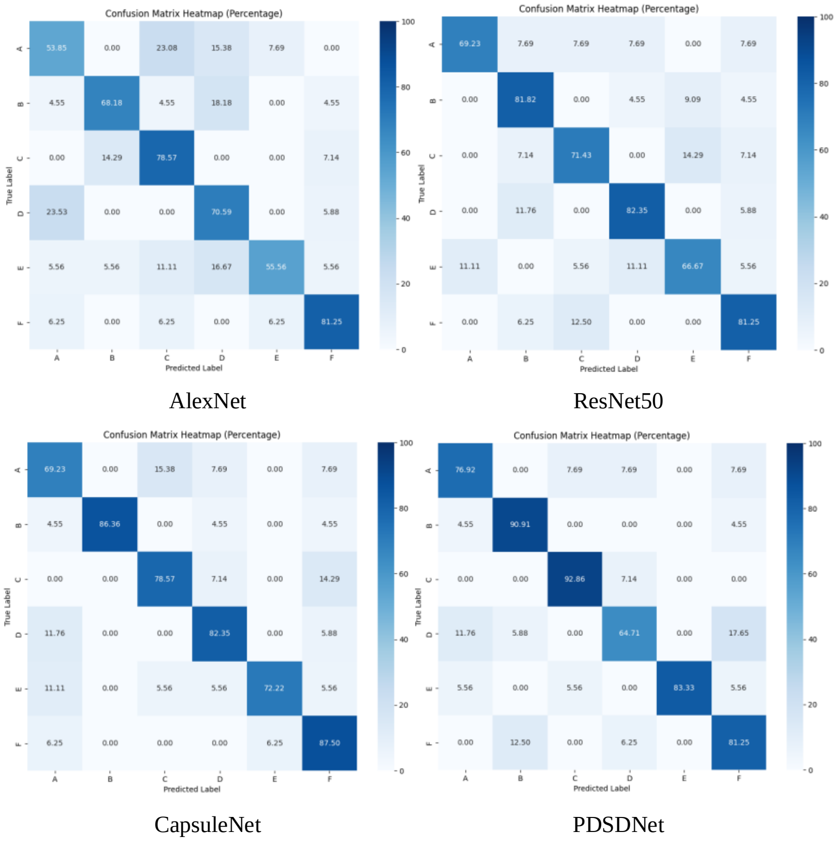
Table 5: Decision results on the feasibility of the overall design.

Design sketches of complex product drafts tend to receive lower scores in feasibility assessments. Therefore, the product design sketch decision model can implicitly learn the complexity of the sketches through the network, effectively realizing the feasibility of decision-making for product design sketches.

In the decision-making regarding the overall semantic styling, different models exhibited significant differences in performance, as shown in Table 6. Figure 7 displays the confusion matrices of the decision results of different networks on the test set.

Decision Network Model	*Precision*	*Recall*	*Accuracy*	*F1*
AlexNet	0.6857	0.6800	0.6800	0.6737
ResNet50	0.7608	0.7546	0.7600	0.7558
CapsuleNet	0.7987	0.7937	0.8000	0.7909
PDSDNet	0.8201	0.8166	0.8200	0.8146

Table 6: Network decision results on the overall styling of desk lamp design.



AlexNet

ResNet50

CapsuleNet

PDSDNet

Figure 7: Comparison of confusion matrices for decision results: (a) AlexNet, (b) ResNet50, (c) CapsuleNet, and (d) PDSDNet.

Figure 8 visually displays the decision-making results of models built using different networks on the test set. The decision models utilizing capsule networks exhibit notably higher accuracy in the decision-making process for the overall semantic styling of sketches compared to those using convolutional neural networks. This enhanced performance is due to the capsule network's more effective capture of the relationships between individual design styling areas within the sketches and the overall design proposal. By conducting a detailed analysis of poses in sketches, the capsule network can more comprehensively understand and express the interactions between different elements, thereby improving the accuracy of the decision model in terms of overall semantic styling. This improvement is derived from the capsule network's sensitivity to spatial relationships, enabling it to understand better and infer the relationships between design elements and their impact on overall styling semantics. Consequently, capsule networks demonstrate superior decision-making capabilities in tasks involving overall semantic styling.

In the experiment, our proposed PDSDNet, compared to the traditional CapsuleNet, achieved further improvements in accuracy for overall styling semantics decisions. This improvement is attributed to the stronger feature extraction capabilities of the ConvNeXt backbone network used in the PDSDNet model, which helps to capture abstract features and subtle semantic information in sketches more accurately. The use of the AdamW optimizer leads to more effective weight updates, assisting the model in converging faster and maintaining better generalization performance during training. These factors contributed to the significant improvement of PDSDNet in decision-making tasks for overall styling semantics. The capsule network employing Inverted dot-product attention routing not only speeds up computations but also addresses the issue of gradient vanishing commonly encountered in the training process of classic capsule networks.

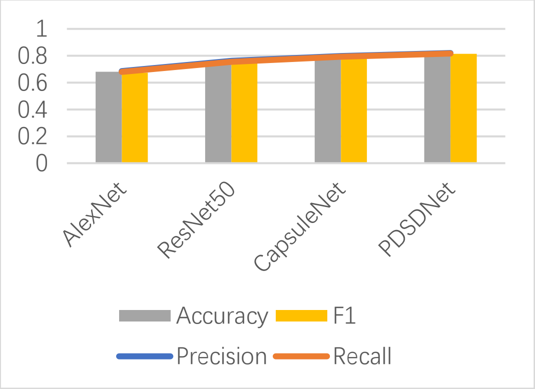


Figure 8: Results of network decision on the overall styling of desk lamp design.

Figure 9 shows the loss curves of CapsuleNet and PDSDNet during the training process. It can be observed that PDSDNet is more stable than CapsuleNet during training, and PDSDNet tends to stabilize at the 70th epoch, while CapsuleNet requires until the 85th epoch to reach stability.

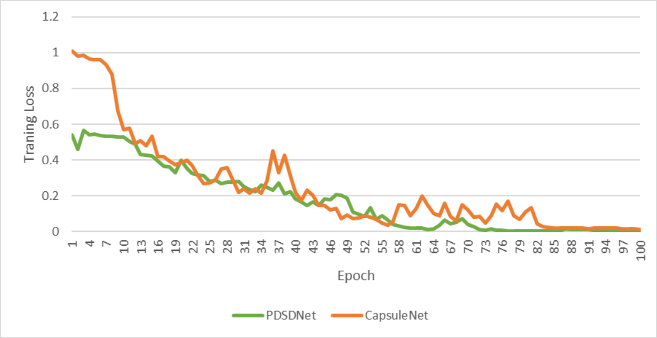


Figure 9: Training loss curves of PDSDNet and CapsuleNet.

5 CONCLUSIONS

To further enhance the efficiency of decision-making in product design sketches, this paper introduces the PDSDNet product design sketch decision model. This model utilizes an improved inverted dot-product attention routing capsule network to overcome the gradient vanishing issue common in classical capsule networks. ConvNeXt, as the backbone network, extracts image features from sketches, thereby increasing attention to the details in the sketch images. In PDSDNet, the AdamW optimizer is used instead of the traditional SGD, enhancing the training speed and stability of the capsule network and strengthening the model's learning and inference abilities regarding the overall styling semantics. The experimental results demonstrate that in the decision-making of overall styling semantics, PDSDNet achieves an F1 score of 0.8146, outperforming other models, thus providing valuable assistance to designers in the sketching phase of product design. In addition, in order to verify the generalization ability of the PDSDNet model, we apply it to the decision-making process of automobile sketch. Experimental results show that PDSDNet also performs well in the overall style semantic decision-making of automobile sketches, reaching 0.8 F1 score. This result shows that although the model was initially built for the lamp design sketch, it has a high applicability in dealing with specific fields such as car design sketches. We believe that this is due to the model's deep understanding of the details of the sketch and its powerful feature extraction capabilities, which enable the model to capture the key design elements and style features in the car sketch.

This paper selects semantic annotation of product design from a pre-defined semantic database without considering the relationship between a single word and a specific morphological feature region, the semantic description of key feature regions in images inevitably has some inaccuracies, which will be the focus of future research. Additionally, our model can only be applied to 2D grayscale sketches, not to color pictures or 3D data. we are considering integrating this model with technologies such as Magic3D to enable intelligent decision-making throughout the entire design process from sketch to final product, further facilitating designers to enhance work efficiency using AI technology throughout the entire design cycle.

6 ACKNOWLEDGEMENT

This work was supported by the Talent Research Program of Anhui Agricultural University (Grant No. rc482003); the University Natural Science Research Program of Anhui Province(Grant No. KJ2021A0181); Zhejiang Provincial Association for Higher Education 2024 Special Key Project on 'Research on the Application of Artificial Intelligence in Empowering Education and Teaching'(Grant No. KT2024436).

Zhou Qingyan, http://orcid.org/0000-0002-1057-5882
Zhang Jincheng, http://orcid.org/0009-0008-7054-0893
Wei Tangwei, http://orcid.org/0009-0006-5448-2895
Li Hao, http://orcid.org/0009-0007-5671-2022
Wang Jing, http://orcid.org/0009-0009-8413-7963
Bao Xiaozhong, http://orcid.org/0009-0000-8829-985X

REFERENCES

[1] Cao X.; Xie X.; Xiao W.; Zou N.; He X.: Koch Fractal-Based LED Lamp Appearance Design Method, Machine Learning and Intelligent Communications: First International Conference, 2016,2017,209-216. https://doi.org/10.1007/978-3-319-52730-7_21

[2] Chen L. C.; Zhu Y.; Papandreou G.; Schroff F.; Adam H.: Encoder-decoder with atrous separable convolution for semantic image segmentation, Proceedings of the European Conference on Computer Vision (ECCV),2018,801-818. https://doi.org/10.1007/978-3-030-01234-2_49

[3] Ding M.; Cheng Y.; Zhang J.; Du G.: Product color emotional design based on a convolutional neural network and search neural network, Color Research & Application, 46(6), 2021,1332-1346. https://doi.org/10.1002/col.22668

[4] Dinani S. T.; Caragea D.: Disaster Image Classification Using Capsule Networks, 2021 International Joint Conference on Neural Networks (IJCNN),2021,1-8. https://doi.org/10.1109/IJCNN52387.2021.9534448

[5] Fonseca M. J.; Ferreira A.; Jorge J. A.: Sketch-based retrieval of complex drawings using hierarchical topology and geometry, Computer-Aided Design, 41(12), 2009, 1067-1081. https://doi.org/10.1016/j.cad.2009.09.004

[6] Liu Z.; Mao H.; Wu C. Y.; Feichtenhofer C.; Darrell T.; Xie S.: A convnet for the 2020s, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022,11976-11986. https://doi.org/10.48550/arXiv.2201.03545

[7] Ng A. J. B.; Liu K. H.: The investigation of different loss functions with capsule networks for speech emotion recognition, Scientific Programming, 2021, 1-12. https://doi.org/10.1155/2021/9916915

[8] Paoletti M. E.; Tao X.; Han L.; Wu Z.; Moreno-Álvarez S.; Haut J. M.: Deep Attention-Driven HSI Scene Classification Based on Inverted Dot-Product, IGARSS IEEE International Geoscience and Remote Sensing Symposium, 2022, 1380-1383. https://doi.org/10.1109/IGARSS46834.2022.9883028

[9] Press M.; Cooper R.: The design experience: the role of design and designers in the twenty-first century. Routledge, London, 2003. https://doi.org/10.4324/9781315240329

[10] Sabour S.; Frosst N.; Hinton G. E.: Dynamic routing between capsules, Advances in neural information processing systems, 30, 2017, 3856. https://doi.org/10.48550/arXiv.1710.09829

[11] Schmid C.; Soatto S.; Tomasi C.: Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, San Diego, CA, 2005. https://doi.org/10.1109/CVPR.2005.277

[12] Shlahova A.: Problems in the Perception of Perspective in Drawing, Journal of Art & Design Education, 19(1), 2000, 102-109. https://doi.org/10.1111/1468-5949.00207

[13] Tsai Y.H.; Srivastava N.; Goh H.; Salakhutdinov R.: Capsules with Inverted Dot-Product Attention Routing, ArXiv, 2002.04764, 2020. https://doi.org/10.48550/arXiv.2002.04764

[14] Tovey M.: Styling and design: intuition and analysis in industrial design, Design Studies, 18(1), 1997, 5-31. https://doi.org/10.1016/S0142-694X(96)00006-3

[15] Voulodimos A.; Doulamis N.; Doulamis A.; Protopapadakis E.: Deep learning for computer vision: A brief review, Computational intelligence and neuroscience, 2018, 2018, 1-13. https://doi.org/10.1155/2018/7068349

[16] Yan H.; Zhang H.; Liu L.; Zhou D.; Xu X.; Zhang Z.; Yan S.: Toward Intelligent Design: An AI-Based Fashion Designer Using Generative Adversarial Networks Aided by Sketch and Rendering Generators, IEEE Transactions on Multimedia, 25, 2022, 2323-2338. https://doi.org/10.1109/TMM.2022.3146010

[17] Zhao Y.: Study on Lamp Product Modeling Based on Parametric Bionic Design, M.S. Thesis, Donghua University, Shanghai, China, 2022. https://doi.org/10.27012/d.cnki.gdhuu.2022.000810

[18] Zong L.; Wang N.: Research on the Decision Model of Product Design Based on a Deep Residual Network, Scientific Programming, 2022(1), 2022, 276-280. https://doi.org/10.1155/2022/8490683