WILEY | Hindawi

Research Article

Car Emotion Labeling Based on Color-SSL Semi-Supervised Learning Algorithm by Color Augmentation

Zhuen Guo [ID],[1] **Li Lin** [ID],[1,2] **Baoqi Liu** [ID],[1] **Li Zhang** [ID],[1] **Kaixin Chang** [ID],[1] **Lingyun Li** [ID],[2] **Zuoya Jiang** [ID],[2] and **Jinmeng Wu** [ID][1]

[1]*School of Mechanical Engineering, Guizhou University, Guiyang, Guizhou 550025, China*
[2]*Key Laboratory of Modern Manufacturing Technology, Ministry of Education, Guizhou University, Guiyang, Guizhou 550025, China*

Correspondence should be addressed to Li Lin; linli_gzu@163.com

In the era of emotional consumption, it has become a hot topic that commodities meet consumers' emotional needs. As a necessity of life, the car also needs to meet the needs of consumers. To achieve that consumers can purchase cars according to their emotional needs, we need to label cars with emotional words. The car's appearance is the crucial medium of emotional information transmission, especially the car's color is an essential emotional factor. As the first impression of products, color affects people's emotional attitude. Therefore, introducing color features into the training process of sample marking is an excellent idea for intelligent labeling of a large number of product emotions. This paper proposes a semi-supervised learning method, Color-SSL, based on color data augmentation to realize the label of car emotion. Color-SSL takes FlexMatch as the framework of a semi-supervised learning model and augments data by extracting subject color. Compared with the baseline method, the accuracy of this method improved by 3.2%, 8.3%, 8.6%, and 1.4% with 10, 50, 100, and 200 training samples and 1000 test samples. The results show that Color-SSL obtains the best emotion-label result (94%). In addition, this study publishes pictures of emotional car datasets with high resolution, orthogonal perspective, and uniform background.

1. Introduction

With the rapid expansion of industrial globalization, people can experience and buy many brands and types of cars [1]. For ordinary users, cars have become a meaningful way to improve work efficiency and meet transportation needs. Consumers will choose cars according to their emotional needs, brand, price, fuel consumption, safety, and style in the car purchase process [2]. Also, emotional needs are considered the primary concern of consumers [3]. Every consumer will have different emotional needs, such as needing a car with streamlined, atmospheric, cute, and luxurious properties. Especially when the essential functions of cars are similar, car emotions are significant to consumers [4]. A car design that conforms to users' emotional preferences can induce consumers to buy. Designing cars that meet the

emotional needs of users is also the focus of major manufacturers [5].

In the context of economic globalization, consumers can experience different brands and types of cars locally [6]. Consumers have many options in buying cars, but it will also increase the difficulty of consumers when purchasing. There are many car brands. Each car brand has different types, and each type has a different emotional expression. If consumers check and understand each type, it will consume a lot of time and energy. Therefore, when buying a car, consumers often consider purchasing it according to their knowledge and experience, the recommendation of friends, their familiarity with the brand, and the introduction of the shopping guide [7]. However, this will cause consumers to have cognitive limitations when buying a car and cannot fully understand many car types, resulting in impulsive consumption [8].

Therefore, solving users' emotional needs and screening cars that meet consumers' emotional preferences at the beginning of car purchase is a critical way to reduce consumers' time and energy costs. It is also a meaningful way to reduce impulsive consumption. When consumers put forward their emotional needs, it is the key to solving the above problems to quickly select the car types with this emotional representation from many car types and then understand these screened car types. It is significant to build the emotional information of each car type, show the type of car with this emotion to consumers according to their emotional preferences, and improve the efficiency of consumers' choice of car buying.

However, there are more than 200 common car brands, such as Volkswagen, BMW, Mercedes Benz, Hyundai, and Toyota [9]. There are many types and car families under each brand. The construction of car emotional information requires many car types and emotional words, and the amount of data is vast. It is complex and colossal work to label each car type's emotional information manually. Furthermore, different people will have different attitudes towards products [10]. The attitude is uncertain, which challenges labeling products' emotional information. Therefore, how to quickly label the emotional information of each car type and solve the different emotional views of different groups on products is an urgent problem.

In recent years, the development of semi-supervised learning (SSL) algorithms provides a solution for the emotional information label of each car type [11]. SSL uses a large amount of unlabeled data and a small amount of labeled data for pattern recognition. Nowadays, SSL is widely used. Parthasarathy and Busso proposed a semi-supervised speech emotion recognition method using a ladder diagram network [12]. The proposed method creates a training framework with strong applicability for SSL and achieves better performance than fully supervised single task learning (STL) and MTL baseline. Xie et al. proposed unsupervised data augmentation to apply SSL [13]. On the IMDB dataset, through the UDA algorithm learning of only 20 labeled samples and more than 70000 unlabeled samples (after data augmentation), they finally achieved a better effect than the 2.5 W labeled dataset. Sohn et al. proposed the FixMatch method, which performs weak augmentation on the unlabeled image to generate false tags and strong augmentation on the unlabeled image to predict, including 94.93% accuracy on the CIFAR-10 with 250 labels [14]. Zhang et al. proposed the curriculum pseudo-labeling method, which can be applied to multiple semi-supervised methods without introducing new hyperparameters and additional computational overhead [15]. This method is applied to the FixMatch algorithm to achieve better performance than FixMatch in SSL. The above summarizes the recent popular SSL algorithms. Using these algorithms in SSL can effectively solve the prediction error of a small number of labels on unmarked data, reduce human work, and have high accuracy. SSL algorithm also provides a basis for this study to label huge cars' emotional information.

However, the traditional SSL algorithm is mainly used to label products, animals, expressions, texts, etc. The characteristic information of these data is obvious and easy to distinguish. For example, most SSL algorithms are mainly tested with datasets such as CIFAR-10, CIFAR100, STL-10, and SVHN. This study proposes to take cars as samples to label emotional information, which only contains one product type. The product differentiation is not obvious, which is easy to produce an overfitting phenomenon [16]. In addition, in the research of emotional labels, SSL is rarely used, which also brings excellent uncertainty for this study to try to use an SSL method for emotional labels.

Therefore, this study tries a new idea to verify the feasibility of a semi-supervised method to label pictures with emotion. Because the traditional emotional label is mainly manual, the emotional label of car pictures used in this study requires designers with professional skills to comprehensively judge each picture's color and shape [17]. For example, in the field of product design, it is generally believed that red products will produce a warm feeling, yellow products will have a warm feeling, and pink products will have a lovely sense. For the shape, the curve of the product will have a round feeling, while the straight line of the product is a strong feeling. Metal products often bring people a cold sense of materials, while wood products bring people a traditional sense. However, when labeling the emotional preferences of the car, the car material displayed in the picture has poor discrimination, and the picture can only stay in the same perspective, which cannot display the overall modeling information. Ding's research shows that product colors have different emotional perceptions in different distributions and combinations and affect users' emotional reactions to products [18]. Therefore, we only consider the influence of color in labeling emotional information on cars. We augmented the image data by adding color feature data and compared the label accuracy after data augmentation with that when the data were not augmented to verify our ideas when labeling emotional images in SSL. Also, we try to solve the inconsistency of different groups' emotional recognition of products through the method of the iconic samples proposed by our research group [19].

This paper makes the following research and analysis on labeling car emotion. (I) We construct the car emotion dataset and construct the test set through the iconic samples. (II) We propose the Color-SSL algorithm by improving the FlexMatch algorithm and enhancing the color content to obtain the training model. (III) We apply the trained model to the unlabeled car pictures to observe the accuracy of labeling and other indicators. (IV) The Color-SSL algorithm proposed in this paper is compared with the original FlexMatch algorithm, which proves the feasibility of introducing the product color data into the SSL process. The training process for emotional labels of car pictures is shown in Figure 1.

Our contributions in this article are as follows:

(1) We propose a Color-SSL method for car emotion labeling based on color augmented.

(2) Although Color-SSL is based on the existing Flex-Match model, the FlexMatch model used here is improved by color augmented. Color augmented is
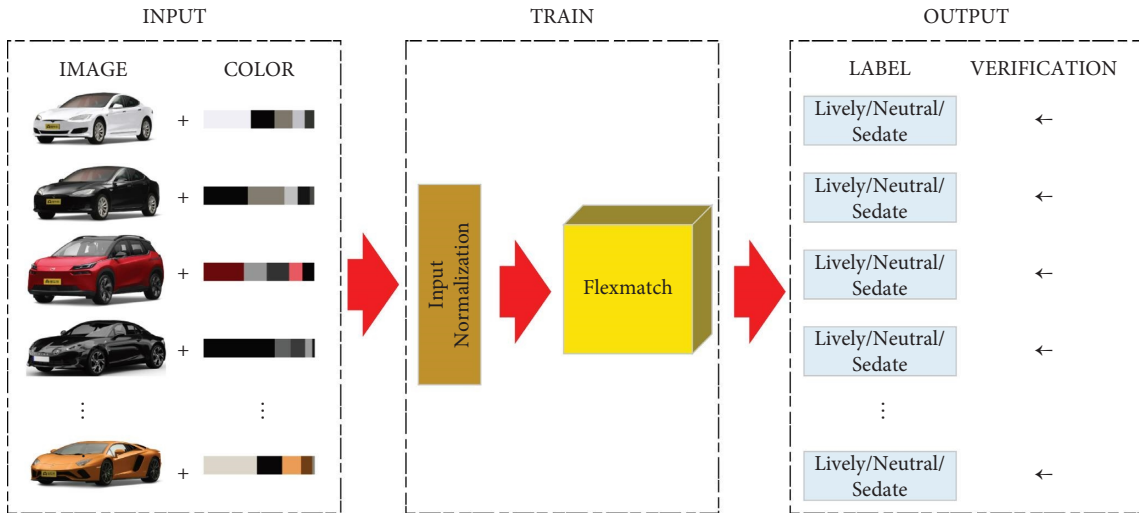
FIGURE 1: The training process of emotional labeling of car pictures. Each car image will be extracted with the five theme colors that account for the most significant proportion. The extracted color and picture are combined for data augmentation and then normalized. After data augmentation, the FlexMatch model is used for semi-supervised learning to obtain classification results.

applied to the semi-supervised learning training process to add color features as the key factor in decoding emotional image features. The results show that the accuracy of labeling can reach 94%, and the calculation efficiency is improved compared with the baseline method.

(3) We use the method of landmark samples to solve the quality of training samples. The quality of labeled information in training samples is an essential factor affecting the performance of semi-supervised learning. When labeling the emotional information of pictures, people's attitudes are inconsistent, which will lead to the uncertainty of the labeling information of the training set. Therefore, we use the method of the iconic samples proposed by our team to solve the problem of different groups' ambiguous emotional attitudes towards pictures.

(4) We publish emotional car datasets with high resolution, orthogonal perspective, and uniform background for the first time.

The rest of this paper is organized as follows. Section 2 introduces the dataset we made and reviews previous relevant research. Section 3 details the proposed approach. The experimental results are reported and analyzed in Section 4. Finally, Sections 5 and 6 give detailed discussion and conclusion.

2. Related Work

2.1. Datasets. We have collected the data published by existing research and learned that there are few public datasets of car pictures. Common automotive datasets are shown in Table 1. The public datasets of car pictures are all car pictures obtained on the actual road or actual scene, which have extensive background interference. They are used in traffic flow statistics, car detection, car classification,

and other applications and cannot meet the dataset requirements required by this study. Therefore, on this basis, this study needs to make an emotional car dataset suitable for product design and emotional.

To meet the requirements of this study, the image needs to meet the background which is a transparent or white background. In addition, to obtain the overall information of the car in a more extensive range, we need to adopt the orthogonal perspective. Therefore, based on the requirements of the above datasets, we have established a car emotion image dataset, as shown in Figure 2. Based on the types sold in the Chinese market, we use the two largest online car selection websites in China, Autohome (https://www.autohome.com.cn/) and Dongchedi (https://www.dongchedi.com/), to get the car pictures. Through screening, there are 25152 pictures. The dataset mainly includes brands, models, styles, and car types. The text information example is Mercedes Benz C-class 2022 C 260l sports green grandmother's green medium-sized car. These data will be made public.

2.2. Semi-Supervised Learning. Machine learning and deep learning algorithms have been widely used and have proved that fully supervised learning under a large number of labeled data is very effective [35]. In the actual scene, acquiring a large amount of data is very simple. However, it is challenging to obtain labeled data, and labeling a large amount of data requires a lot of time and energy. SSL is more suitable for practical applications [36]. SSL is the combination of supervised learning and unsupervised learning. SSL uses a small part of labeled data and a large amount of unlabeled data for training and prediction and finally realizes the label of a large amount of data. Nhi et al. enriched many image databases when building semantic-based image retrieval [37]. In order to solve the relationship between a large number of images and classes, they built an image generation framework through semi-supervised learning to

TABLE 1: Car public dataset.

Dataset	Traffic flow	Car detection	Car type	Car color	License plate	Car style	Reference
UA-DETRAC	√	√					[20]
BDD100 K	√	√					[21]
KITTI	√	√					[22]
BIT-Cardata			√				[23]
OpenData V11.0						√	[24]
CCPD					√		[25]
D²-City	√	√					[26]
Vehicle Color Recognition Dataset				√			[27]
VeRi dataset			√	√		√	[28]
Vehicle ID						√	[29]
VRAI			√	√			[30]
PKU-VD			√	√			[31]
CompCars						√	[32]
Stanford Cars Dataset						√	[33]
The Car Connection Dataset			√			√	[34]

FIGURE 2: Car emotion image dataset.

improve the efficiency of automatic retrieval. Chen et al. proposed a semi-supervised learning-based representation algorithm for repeated images when detecting image repeatability [38]. Their proposed method guarantees the description and semantic similarity and achieves excellent performance in image retrieval. Semi-supervised learning effectively solves the problem of less data and labeled information.

SSL algorithms include graph-based SSL (GSSL), pseudo-label, and hybrid methods [37]. The main idea of GSSL is to extract a graph from the original data. Each node represents a training sample, and the edge represents the similarity measurement of the sample pair. This graph contains labeled and unlabeled samples. The goal is to propagate labeled data from labeled nodes to unlabeled nodes [39]. The pseudo-label method works in two steps. The first step is to train the model on a limited labeled dataset. The second step is to use the same model to create a pseudo-label on unlabeled data and add the high-confidence pseudo-label as targets to the existing labeled dataset to make additional training data [40]. The primary process of SSL is shown in Figure 3.

Recently, the popular SSL algorithm mainly uses the pseudo-label and consistent regularization [41]. The pseudo-label is the target class after classifying unlabeled data. The pseudo-label can use them like an actual label during training. The model used when selecting pseudo-label is the class with the maximum prediction probability for each unlabeled sample. Consistency regularization disturbs the unlabeled data, the prediction results do not change significantly, and the output information is consistent. Through the combination of the two methods, the model of the SSL algorithm is not accessible to overfit and has a high generalization ability. SSL algorithms such as FixMatch, FlexMatch, UDA, and MixMatch have achieved perfect results in predicting unlabeled images.
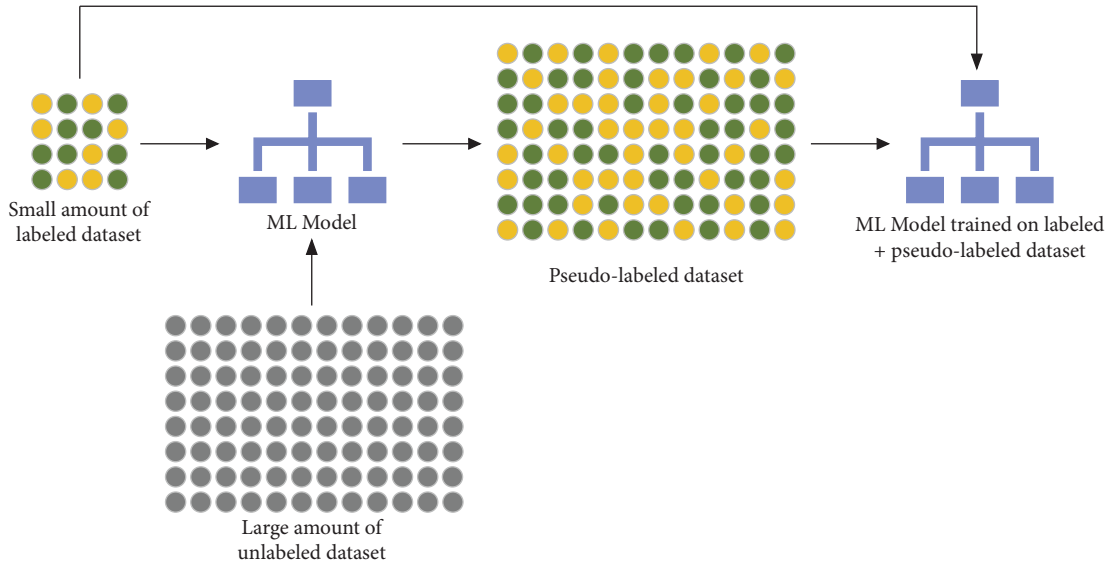
FIGURE 3: Main process of SSL. Semi-supervised learning (SSL) is the combination of supervised learning and unsupervised learning. SSL uses a small number of tagged examples and many unlabeled data, from which the model must learn and predict new examples. The basic process involves using existing tagged data to tag the remaining unlabeled data, thereby effectively helping to increase training data.

2.3. Data Augmentation. Data augmentation is to process more data from the original data without increasing the data. Data augmentation improves the quantity and quality of raw data to approach the value of more data [42]. In the data augmentation of pictures, the data augmentation of a single sample is mainly carried out through geometric operation, color transformation, random erasure, and adding noise [43]. Some recent studies have tried to use GAN for data augmentation and expanded the data augmentation methods. The problem of fewer data can be solved through data augmentation, and overfitting can also be prevented. Sedik et al. used the image rotation method to augment the data of X-ray chest images, achieving excellent performance and effectively solving the problem of deep learning in detecting and screening COVID-19 [44]. Srivastava et al. used three sampling methods to augment 3D data to solve the scarcity of 3D data in depth neural networks [45]. Rahman et al. used GAN to augment the COVID-19 dataset and verified the accuracy of the test with the deep learning model [46]. The above research has achieved excellent results in practical applications through data augmentation and improved the classifier's performance.

The above data augmentation method mainly adds graphics similar to the original data with the same structure. These methods have made outstanding contributions to data volume expansion, but the original data have not changed. People analyze emotional pictures and can focus on the main features of the picture, in which the theme color of the picture is one of the main features. With this idea, we will add the main features of the image to the original data, augment the original image data, and realize the rapid recognition of the depth neural network.

2.4. Iconic Samples. The iconic samples method is the method proposed by our team to screen samples when conducting psychological cognitive experiments [19]. The primary purpose of this method is to integrate the emotional evaluation of different groups and reduce the influence of subjective factors on emotional evaluation. Through emotional evaluation of products, we can get a preliminary emotional classification. Different groups will have different emotional evaluations when obtaining the primary emotional category of products. At this step, the Likert scale is mainly used to score, and then the score of each type of emotion is defined. Through the definition of scores, different product samples are preliminarily classified into three kinds of emotional sample sets that conform to the emotion, that do not conform to the emotion, and that are neutral, as shown in the following equation:

$$S = \begin{cases} S_1, & (V < k_1), \\ S_2, & (k_1 < V < k_2), \\ S_3, & (V > k_2), \end{cases} \tag{1}$$

where k_1 and k_2 are the defined values, V is the Likert scale value of the product sample, and S_1, S_2, and S_3 represent the sample set of preliminary emotion classification.

Then, we need to investigate the results of the preliminary classification again. The three-point semantic difference scale is used to survey at this stage. Each sample has three candidates, two relative emotional words and neutral. We make different groups judge the samples and obtain the classification results of each sample. We assume that the sample is iconic if all investigators have the same judgment results on a sample. The relationship between iconic and non-iconic samples is shown in Figure 4. Each sample in the iconic sample dataset has a clear emotional meaning (different groups have no deviation in the emotional evaluation of the sample). The corresponding experimental model is generated by experimenting with the iconic samples as
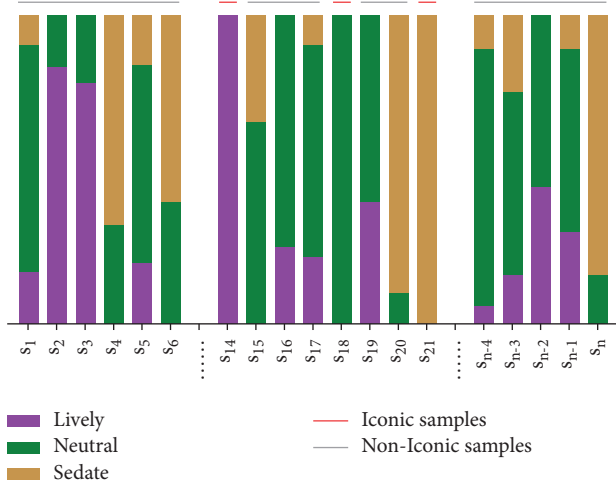
FIGURE 4: Schematic diagram of determining the iconic samples. When all people have the same evaluation results on the same sample, this sample is an iconic sample. The color of the histogram displayed in the figure represents different evaluation proportions. When the sample is an iconic sample, only one color is displayed.

testing samples, which can be applied to the verification process of other non-iconic samples. The iconic sample has also been proved feasible in many studies.

Semi-supervised learning predicts the information of many unlabeled samples through a small number of labeled samples. Therefore, the quality of labeled samples affects the prediction performance of semi-supervised learning. Our study needs to be labeled with emotional information. But the emotional reaction may be different for everyone. Only those pictures with obvious emotional significance will lead to consistent user responses. In order to improve the labeling quality of samples, it is necessary to clarify the emotional information of these few samples. The iconic samples method solves this problem. This paper will also reduce the impact of group emotional evaluation bias when labeling cars' emotional information through iconic samples.

3. Methods

In this section, we propose Color-SSL for the emotional labeling and classification of product pictures. We will take

FlexMatch as the main framework of SSL. We improved FlexMatch to use a small number of labeled and many unlabeled car images for training to realize the generation of a classification model.

3.1. FlexMatch. FlexMatch algorithm is an SSL model based on a pseudo-label algorithm [15]. For the current SSL algorithm based on the pseudo-label method, a relatively high and fixed threshold will be set, such as the threshold set in FixMatch which is 0.95. If the model's confidence for unlabeled samples exceeds the set threshold, the sample will be given a false label. This high and fixed threshold has some problems, leading to insufficient consideration of different training states and different types of training difficulties. Therefore, FlexMatch proposed curriculum pseudo-labeling (CPL). With the semi-supervised training process, the flexible threshold of each category is dynamically adjusted without introducing additional parameters and calculations. CPL assigns a pseudo-label to different classes at different time steps and adjusts the threshold in different training processes. The dynamic adjustment threshold is the main content of CPL. CPL adjusts the threshold by calculating the evaluation accuracy of each class, using the following equation:

$$\mathcal{T}_t(c) = a_t(c) \cdot \tau, \tag{2}$$

where $\mathcal{T}_t(c)$ is the flexible threshold of class c with time step t and $a_t(c)$ is the corresponding evaluation accuracy.

When evaluating the learning state, CPL proposes a new alternative method. The purpose of this method is not to introduce other reasoning processes and verification sets. When CPL assumes that the threshold is high, the learning effect of a class can be reflected by its prediction of the number of samples that belong to this class and are higher than the threshold. If the number of samples is small and the prediction confidence reaches the threshold, the class will be more challenging to learn. Its expression equation is as follows:

$$\sigma_t(c) = \sum_{n=1}^{N} l\big(\max\big(p_m, t\big(y|u_n|\big)\big) > \tau\big) \cdot l\big(\arg\max\big(p_m, t\big(y|u_n|\big)\big) = c\big), \tag{3}$$

where $\sigma_t(c)$ reflects the learning effect of class c at the time step t. $p_m, t(y|u_n|)$ is the prediction of unlabeled data by the model at the time step t, and N is the total number of unlabeled data.

Because this research is mainly about data augmentation, we specifically analyze FlexMatch data augmentation. This method is applied to two kinds of data augmentation: weak

augmentation and strong augmentation. Weak augmentation refers to flipping the picture with a probability of 0.5, cropping the picture at a random position, and normalizing the picture with the mean value. For strong augmentation, the RandAugment [47] strategy is mainly applied to select and combine image transformation randomly. The main algorithm flow is shown in Figure 5.
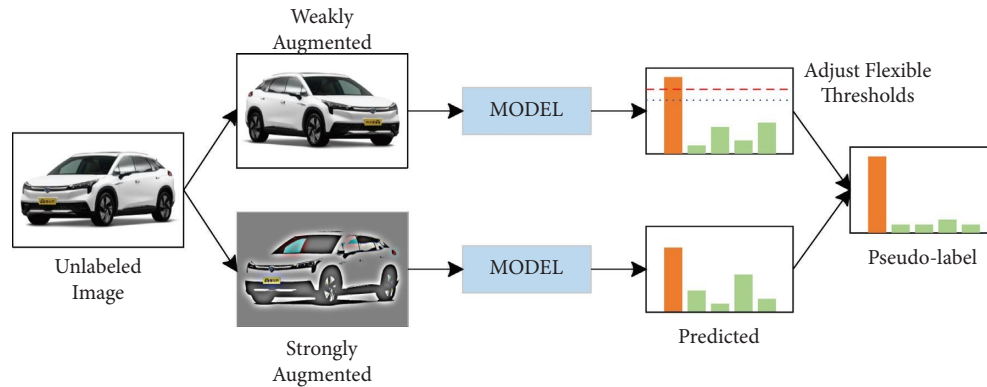
FIGURE 5: Main flowchart of FlexMatch algorithm. The flexible threshold of each class is adjusted to achieve the best unlabeled data effect.

3.2. Color-SSL for Emotional Product Image Labeling. In the SSL process, because there are fewer labeled data, it is necessary to expand the amount of data through data augmentation and other methods. Therefore, data augmentation is a vital link in SSL. As mentioned in the previous section, FlexMatch augments data through weak augmentation and strong augmentation to increase the number of data and reduce model overfitting. When labeling emotional information, especially when labeling products, we need to consider color and other information. Therefore, based on the FlexMatch algorithm framework, we propose a semi-supervised classification method based on color augmented. Its workflow is shown in Figure 6.

We combine prior knowledge and transform multiple samples for the data augmentation of multi-sample images. Researchers mainly use SMOTE [48], SamplePairing [49], Mixup [50], and other methods to construct the neighborhood values of known samples in the feature space. The color augmentation of samples proposed in this study mainly extracts the primary color from the original image and combines the main color features with the original image to obtain a new image based on the original image. The Color-SSL algorithm framework is shown in Figure 7.

The main steps are as follows:

Step 1. We make the original sample transparent to reduce the influence of background factors on the extracted main color. Background transparency is handled through the open-source PP-Matting model proposed by Baidu [51].

Step 2. We extract the main color of the transparent image. The main color can reflect the picture's main color characteristics and color composition information. Because the common features of cars, such as windows and wheels, have common colors and occupy a large proportion of the overall car color, multiple colors must be extracted to prevent the influence of common colors when extracting the main color. We use the Median Cut to obtain the five main colors with the highest proportion of the original image and divide them into color blocks of different sizes for combination according to the proportion of the main colors.
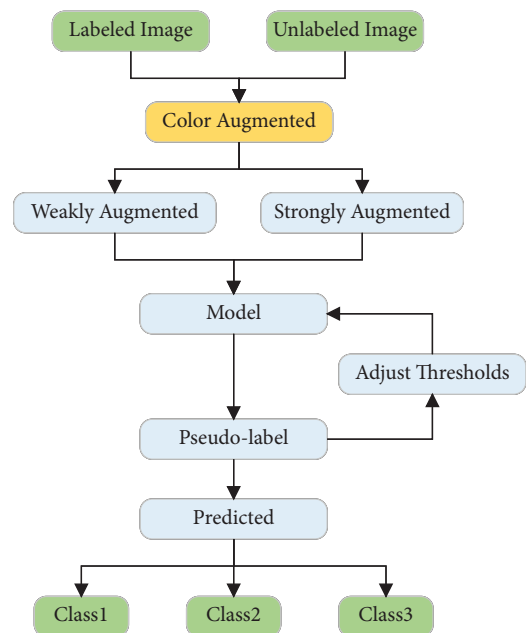

FIGURE 6: Flowchart of Color-SSL algorithm. Before weak augmentation and strong augmentation of data, we perform color augmentation. Then, we input the augmented data to model.

Step 3. We combine the extracted main color with the original image to generate an image based on the original image and color features.

Step 4. We input the color-augmentation image into the original FlexMatch model framework for training prediction to obtain the classification information of unlabeled samples.

4. Experiment and Result Analysis

4.1. Dataset Analysis. Our car emotion dataset contains 25152 pictures and covers 3910 different brands. Because we want to extract the emotional representation reflected in car pictures, 381 emotional words describing the appearance of cars are obtained from the Internet and related materials. The focus group screened these emotional words and eliminated some adjectives with obvious similarity and particularity. Then, the group combined the screened words
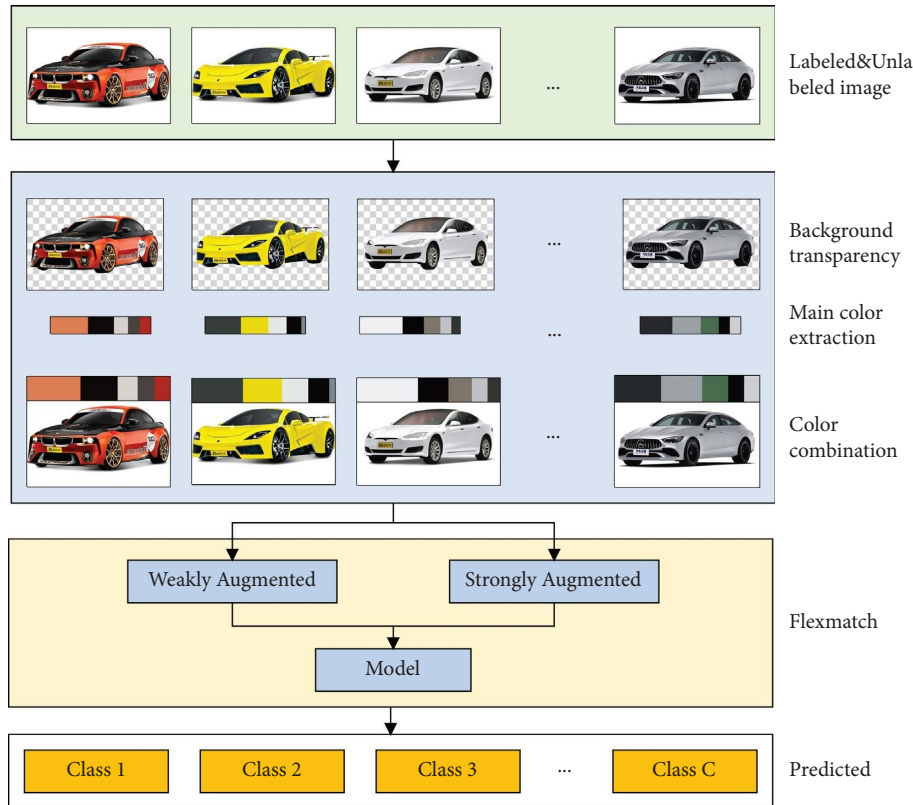
FIGURE 7: Color-SSL algorithm framework.

with adjectives with relative meanings to obtain 21 emotional word pairs. We evaluated the collected pairs of perceptual words in questionnaires to determine which three groups of words can better represent the needs of different groups for a car purchase. A total of 271 questionnaires were issued, and 269 questionnaires were recovered. The content of the questionnaire includes whether there is any recent idea about buying a car, age, gender, and other information. Through the questionnaire analysis, 128 people have the idea of purchasing a car soon. They are between 24 and 53 years old, and the ratio of males to females is 1.56 : 1. In all the questionnaires collected, the age is between 21 and 61. The male to female ratio is 1.26 : 1. According to the users' choice of perceptual words, among the total number of users and the users who have recently bought cars, the Lively-Stable option is the most, followed by the Offroad-Homely, as shown in Table 2. We then interviewed some users who chose Lively-Stable to understand the reasons for selecting this perceptual word pair. Most people say that Lively-Stable can reflect the purchase needs of people of different ages (young and middle-aged), gender groups (women and men), occupations (ordinary occupations and business people), and personality needs (personality and conservatism).

Since the Lively-Stable is selected for the most words, and the Lively-Stable can reflect the emotional needs of different groups. We take the Lively-Stable as the verification and introduce neutral as the transitional word between the two extreme adjectives to form the final emotional words.

4.2. Screening of Iconic Samples. We take Lively-Stable as the emotional words and conduct a questionnaire survey on the car samples using a five-point Likert scale. Because of the large number of samples, we randomly divided the picture samples into five groups and made them into questionnaires. We distributed 150 questionnaires. 147 returned completed questionnaires, 30 returned two-category questionnaires, and 29 returned three-category questionnaires. We analyzed the questionnaire and obtained the emotional score of each sample. According to the score value, the samples are preliminarily divided into three categories of emotional picture sets. The number of preliminary emotional picture sets is 9084 lively, 5611 neutral, and 10457 stable.

We subsequently recruited 5 car designers (3 men and 2 women) with over 5 years of experience, 5 car salespeople (1 man and 3 women), and 4 students in design (2 men and 2 women) to form the expert group. The expert group judged the preliminarily classified set of affective pictures on a three-point semantic scale. We obtained 7848 steady emotional pictures, with 2581 lively, 1753 neutral, and 3514 stable. The decision set relationships for the affective picture set are shown in Figure 8. We randomly selected 4500 iconic samples (1500 samples under each emotion vocabulary) from steady emotional pictures for the convenience of subsequent experiments. These samples will be trained and validated as labeled samples for the feasibility of the Color-SSL algorithm proposed in this study.

TABLE 2: Perceptual vocabulary selection.

Emotional words	Users with car purchase ideas	Total number of users
Luxurious-plain	11	31
Manliness-gentle	23	32
Classical-modern	8	15
Casual-business	45	105
Lively-stable	114	228
Cute-mature	6	18
Tough-soft	35	62
Common-stiff	68	79
Rough-delicate	12	31
Simple-complex	8	16
Flamboyant-connotative	32	95
Ordinary-personalized	21	89
Offroad-homely	89	216
Harmonious-contradictory	2	3
Male-female	72	127
Fashionable-traditional	21	89
Technological-old	32	48
Safe-dangerous	5	8
Beautiful-ugly	3	9
Ripe-young	21	29
Inconspicuous-conspicuous	12	15
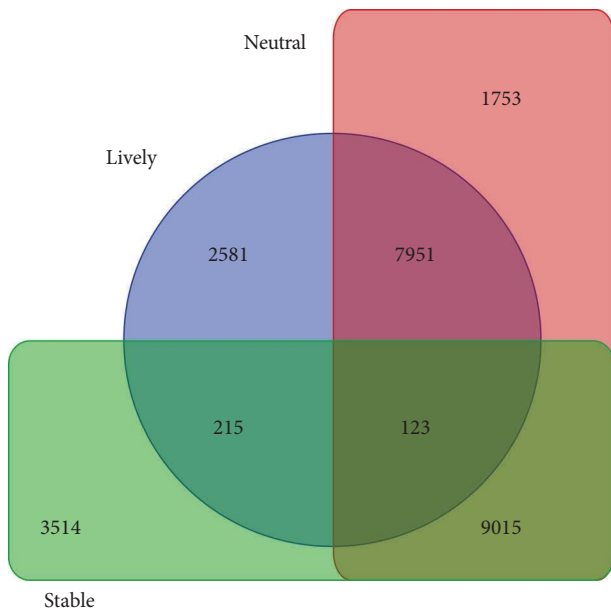
FIGURE 8: Venn diagram of judgment set relation of the emotional picture set.

TABLE 3: Hyperparameter setting information of Color-SSL algorithm.

Hyperparameters	Setting
epoch	1
num_train_iter	360000
num_eval_iter	10000
num_labels	10/50/100/200
batch_size	32
eval_batch_size	256
T	0.5
p_cutoff	0.7
ulb_loss_ratio	1
uratio	1
ema_m	0.999
optim	SGD
lr	0.03
momentum	0.9
weight_decay	0.0003
net	ResNet50
depth	0
widen_factor	0
leaky_slope	0.1
dropout	0

4.3. Parameter Optimization and Setting. Based on the SSL training and parameter optimization of the Color-SSL algorithm, we used a PC workstation equipped with NVIDIA GeForce GTX 3060ti GPU, Intel i5-12400 CPU, and 32 GB ram to conduct experiments. The whole algorithm is based on the TorchSSL. For a fair comparison, we use the same hyperparameter as FlexMatch. We used the ResNet50 for training and optimized it with random gradient descent with a momentum of 0.9. The detailed parameters are shown in Table 3.

4.4. Result Analysis. In order to obtain a more realistic classification prediction performance, we use 10, 50, 100, and 200 labeled samples as training sets and 1000 samples as test sets to verify. We compare the data-augmentation image with the non-data-augmentation image to predict accuracy. The comparison of different training samples is shown in Figure 9.

According to the prediction accuracy, the training results of the sample images augmented by color data are generally better than those not transparent. We will analyze the optimal value of the prediction results after the color
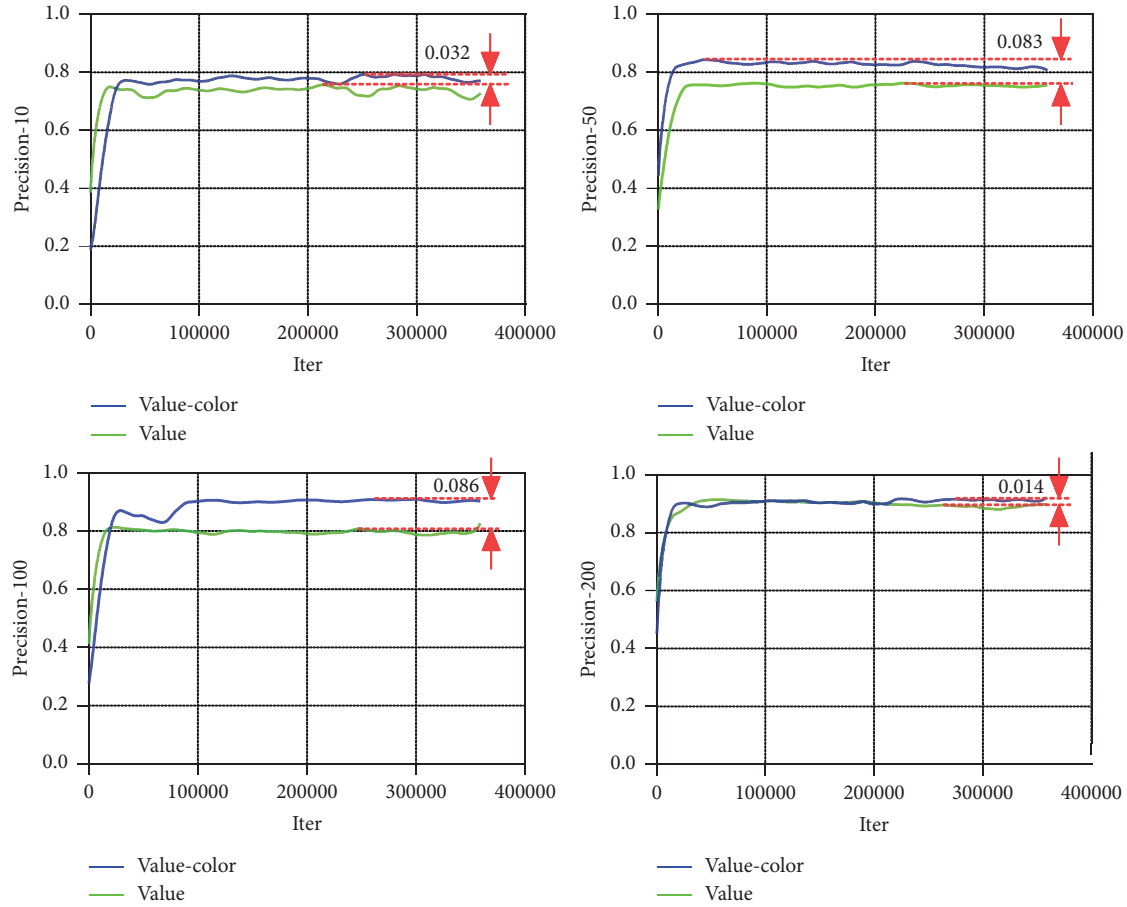
FIGURE 9: Comparison of accuracy of different training samples. The figures in the figure show the improvement in accuracy of our method compared with the baseline method.

TABLE 4: Comparison of accuracy between Color-SSL and FlexMatch.

Numbers of label sample	Method	Train and test				
		1	2	3	4	5
10	Color-SSL	77.52 (1.22)	76.26 (1.36)	75.89 (2.15)	77.34 (1.39)	75.95 (1.16)
	FlexMatch	72.75 (2.09)	73.69 (1.81)	71.05 (2.13)	72.69 (2.11)	71.36 (2.16)
50	Color-SSL	81.71 (0.89)	79.56 (1.12)	81.73 (0.95)	82.53 (0.68)	81.87 (1.69)
	FlexMatch	73.01 (2.19)	74.29 (1.25)	75.09 (0.92)	74.10 (1.28)	76.61 (0.93)
100	Color-SSL	91.15(0.34)	91.26 (1.01)	92.53 (0.65)	91.14 (0.65)	90.81 (0.85)
	FlexMatch	79.33 (2.35)	81.88 (1.65)	82.73 (1.32)	81.75 (1.16)	82.72 (0.95)
200	Color-SSL	94.77 (0.30)	91.47 (1.29)	93.26 (0.74)	93.65 (0.82)	92.34 (0.95)
	FlexMatch	87.91 (0.82)	90.03 (0.84)	92.31 (0.55)	93.15 (1.78)	91.51 (0.69)

augmentation and non-augmentation. When there are 50 and 100 sample pictures, the accuracy is greatly improved, which is 8.3% and 8.6%.

In addition, to verify the algorithm's stability in each training result, each experiment sets five different random numbers for training and obtains the accuracy of the prediction results. We recorded the accuracy of the last 20 checkpoints for each training and verification and obtained the average value and average deviation of the accuracy of the last 20 checkpoints. The specific values are shown in Table 4. The table shows that the average deviation of the accuracy obtained by Color-SSL is 2.15 at

most, and the average deviation of the accuracy obtained by FlexMatch is 2.35 at most. From the average deviation, we know that the method we proposed and the model obtained by FlexMatch training are both stable. Secondly, we found that in many experiments, the accuracy of our method has improved compared with the original Flex-Match. When there are 100 labeled samples, the predicted results have exceeded 92%. When there are 200 labeled samples, the accuracy exceeds 94%. It proves that the data augmentation of samples proposed by us is feasible in the SSL process and performs better than the original Flex-Match model.
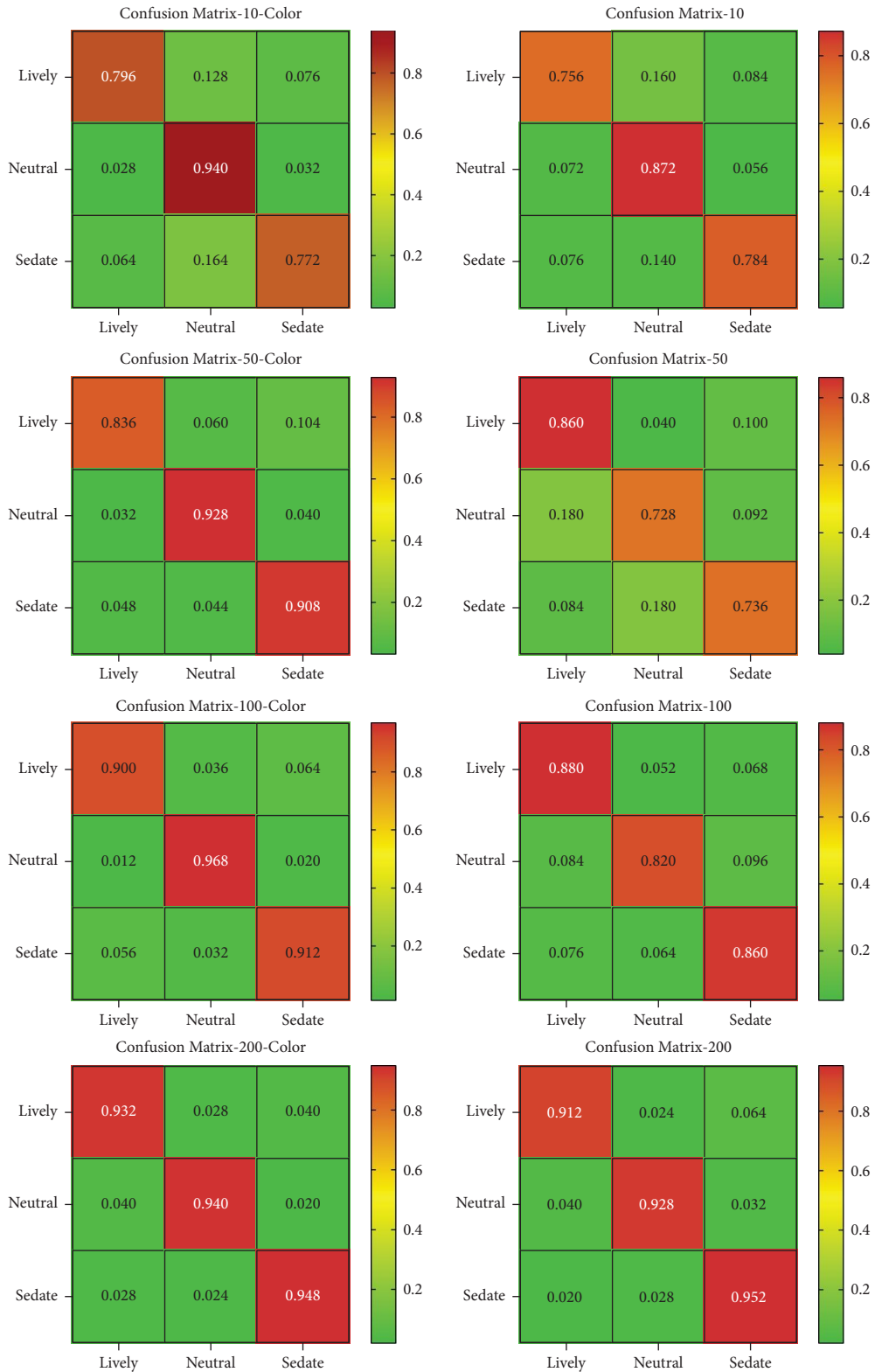
FIGURE 10: Confusion matrix of Color-SSL and FlexMatch under different training sample numbers. Using 1000 test samples to test the performance of the model, the prediction results of our method in the three classes are generally higher than those of the baseline method.

Figure 10 shows the confusion matrix of color and non-color augmentation under different training data. From the confusion matrix, it can be seen that there are obvious differences in the classification results between samples augmented based on color and samples not augmented. The classification results of samples based on color data augmentation are significantly better than those without data augmentation. In the sample training based on

augmentation, neutral samples' classification and label accuracy are higher when training with less or more data sample size. Through the correlation coefficient confusion matrix, the accuracy of training and labeling after color augmentation is significantly better than that without color augmentation.

5. Discussion

Through many experiments and different data, we verify the feasibility of the Color-SSL model based on color data augmentation in the emotional labeling of cars. We have improved the framework of FlexMatch by adding the steps of color data augmentation before image data augmentation. Compared with the original FlexMatch method, Section 4 shows that the SSL algorithm based on the Color-SSL model has achieved excellent results in labeling product. During the experiment, we also found interesting phenomena, including the relationship between the number of training samples and accuracy, complexity and overhead analysis, and the application of the model trained by iconic samples to non-iconic samples.

5.1. The Relationship between the Number of Training Samples and Accuracy. In Section 4, we can find that the Color-SSL proposed has better results than the original FlexMatch model. When the number of training samples is small or large, when the number of training samples is 10 or 200, improving the accuracy is not as obvious as when the number of training samples is 50 or 100. When the training sample set is 10 samples, the accuracy is improved by 3.2%. We assume that it may be related to the small number of samples trained. Color-SSL does not fully learn the characteristics of the samples, resulting in no obvious improvement compared with the original FlexMatch model. When the number of training samples is 200, the increase is minor, which is 1.4%. We assume that the number of samples is large. Color-SSL and FlexMatch models obtain more data features when learning image features, and the improvement effect is not apparent.

Because the SSL process belongs to the black box state, we do not know the SSL state and how to extract learning features in the training process. However, based on this fascinating phenomenon, we can boldly guess that the result of our Color-SSL is better than that of FlexMatch. The main reason is that when training is based on color data augmentation samples, the model can quickly identify the main features that can determine emotional information. This is also similar to human beings' rapid evaluation of products based on color data and empirical knowledge. Although this is a bold guess, we can provide an idea for optimizing the algorithm according to this guess. That is, the fusion of human concept information into the optimization process of the algorithm may have an unexpected result.

5.2. Complexity and Overhead Analysis. Color-SSL proposed in this paper is a semi-supervised learning algorithm based on FlexMatch improvement. Color-SSL is mainly used for emotional labeling of product pictures. In order to better demonstrate the practical application value of the method proposed in this paper, we conducted a complexity analysis of the method. The algorithm's complexity is mainly reflected in two aspects: time complexity and space complexity. Color-SSL and FlexMatch use the same network architecture and superparameters. However, during the training process, Color-SSL added the step of extracting the main color. We calculated that the average duration of each operation in extracting the main color is 0.7523 s, and the memory occupied is 66.056 Mib. Although Color-SSL proposed by us adds some computing processes compared with FlexMatch, compared with the time complexity and space complexity, the amount of computation and computing time generated can be ignored.

Our proposed Color-SSL also has a precise application scenario. In the process of sales and production of industrial products, many products of the same type will make it difficult for consumers to choose. The existing SSL methods are mainly used in object labeling and news labeling and are now gradually applied to new fields such as disease diagnosis. As far as we know, Color-SSL we proposed is the first emotional labeling method applied to consumer products. Color-SSL can be used as a relevant sales platform to quickly label the emotional labels corresponding to the products to facilitate consumers' purchases.

5.3. Application of Methods. Because this study is to label the emotional information of the same products, emotional labeling is different from that of products, and there is no direct or obvious distinguishing feature. Therefore, we applied the iconic samples method in this study, taking the steady emotional samples as training and prediction samples and obtaining relatively excellent results. However, this is only training and prediction based on iconic samples, so verifying whether the model trained by iconic samples can be applied to labeling non-iconic samples is necessary. The attitude of these non-iconic sample expert groups is not uniform, but we take the expert group's majority opinion as the sample's emotional information. The information is taken as the result of preliminary emotional classification.

We randomly selected 1500 pictures in the remaining non-iconic samples according to the results of the preliminary emotional classification, with 500 sample pictures in each category. We take these pictures as input and use the model trained by iconic samples to classify the output emotion. By comparing the prediction results with the results of the initial emotion classification, the prediction accuracy reaches 91.93%, and the prediction results of each category are shown in Figure 11. The model trained by iconic samples can be well transferred to the application of non-iconic samples.

Because the attitude towards emotional information is unclear, the expert group has multiple opinions on the same sample. Although we take the expert group's majority opinions as the sample's emotional classification results, we cannot ignore different opinions. We analyzed the samples with deviation in prediction, and the number of samples
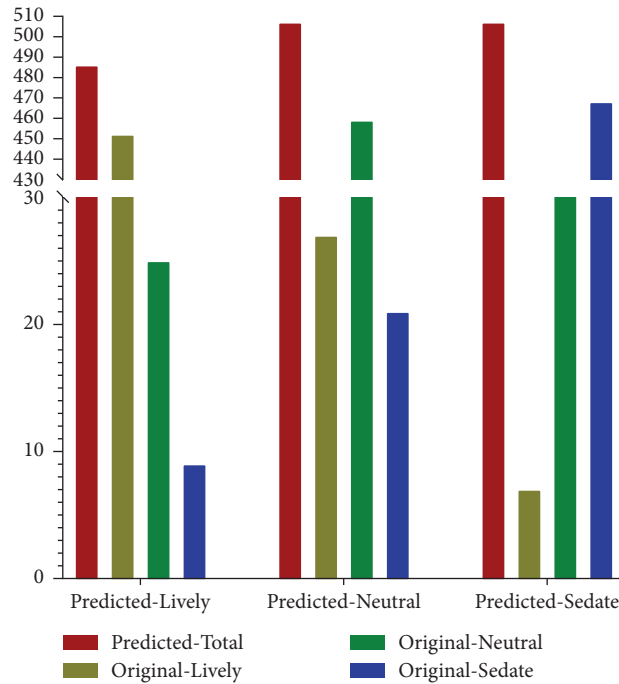
FIGURE 11: Prediction of non-iconic samples. Different colors represent sample information. Red represents the predicted results of our proposed method, and the remaining colors are actual results.
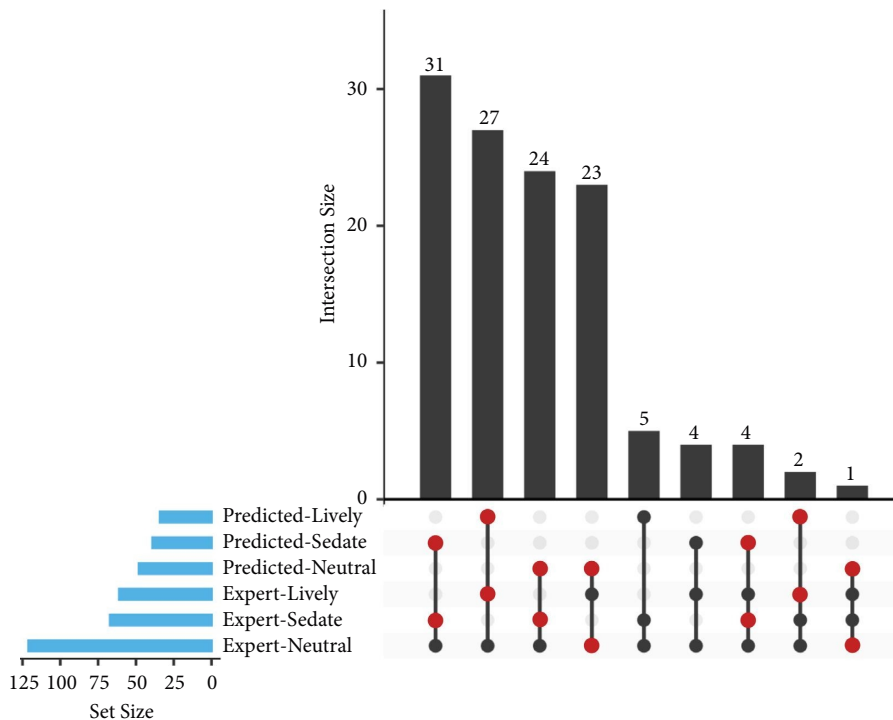

FIGURE 12: UpSet plot of relationship between prediction results and evaluation results of expert group. The prediction results are cross analyzed with the opinions of the expert group. The consistent samples are marked with red dots.

with deviation in prediction was 121. By comparing the different emotion judgments of these samples in the expert group, it is found that there is an intersection between the predicted results of 112 samples and the results determined by the expert group. That is, the predicted results are consistent with the evaluation results of the expert group, as shown in Figure 12, which shows the relationship between the predicted results and the evaluation results of the expert group. Although 121 results predicted by the model have deviations, compared with the evaluation of the expert

group, the consistent result has reached 92.56%, indicating that this model can meet the final labeling requirements in the prediction. It can be noted that because the samples we verified this time belong to non-iconic samples, the expert group has a variety of choices for the label of the emotional information of the samples. The predicted results are highly consistent with the preliminary emotional classification information obtained from the questionnaire, and the samples with prediction deviation are also consistent with the emotional information classification of the expert group. Through this verification, we know that in model training or related experiments, the establishment of models through landmark samples can be applied to the related process of non-iconic samples and has high consistency.

6. Conclusion

This study proposes a Color-SSL model based on color data augmentation. This model combines color data augmentation with the traditional FlexMatch model, making it more suitable for the scene of the emotional label. Specifically, this study innovatively introduces the human emotional evaluation of products based on color information into the SSL process. We compared the training results before and after color data augmentation and proved that our Color-SSL is better than FlexMatch. Secondly, SSL requires a small number of labeled samples, and the quality of label information of these labeled samples is essential. We used the method of iconic samples to screen the iconic samples as the training and testing sample set and obtained excellent results.

Moreover, through the application of non-iconic samples, the accuracy and coincidence of analysis have proved that our idea is correct, and the trained model can be well applied and extended. In addition, we have also made a dataset of car products for emotional research. This dataset collects large-scale pictures of car products with a unified background and orthogonal perspective and has each picture's brand, type, and other information. Although our Color-SSL has achieved some results and verified the feasibility of the iconic sample method proposed in our previous research, there are also some limitations. First, we need to obtain a certain number of iconic samples as the initial labeled samples which will consume some energy, but it takes much less time than labeling all samples. Second, we do not consider the factors that affect the shape and material of emotional information evaluation but only consider the factors of color, which also points out the direction for our future research. In addition, we can consider the combination of color data augmentation and more SSL models to develop a more practical and accurate model to provide ideas for the label of product emotional information.

Data Availability

The car emotion image dataset can be accessed from the following website: https://drive.google.com/drive/folders/1c2DBLNWMOhQHsukfeh8GyXqvb2bx6BlT?usp=sharing.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] D. Al Shishani, "Factors influence the buying behaviours of consumers," *International Journal of Recent Research in Social Sciences and Humanities (IJRRSSH)*, vol. 7, no. 3, pp. 14–21, 2020.

[2] A. N. E. Nassef and G. Sayed Abd El-Aziz, "Investigating the relationship between nostalgic advertisement, brand heritage, and automobile purchase intention (applying on mercedes–benz Egypt)," *The Academic Journal of Contemporary Commercial Research*, vol. 2, no. 1, pp. 16–34, 2022.

[3] C. Gillitzer and N. Prasad, "The effect of consumer sentiment on consumption: cross-sectional evidence from elections," *American Economic Journal: Macroeconomics*, vol. 10, no. 4, pp. 234–269, 2018.

[4] R. Dou, W. Li, G. Nan, X. Wang, and Y. Zhou, "How can manufacturers make decisions on product appearance design? A research on optimal design based on customers' emotional satisfaction," *Journal of Management Science and Engineering*, vol. 6, no. 2, pp. 177–196, 2021.

[5] T. Dhanabalan, K. Subha, R. Shanthi, and A. Sathish, "Factors influencing consumers' car purchasing decision in Indian automobile industry," *International Journal of Mechanical Engineering & Technology*, vol. 9, no. 10, pp. 53–63, 2018.

[6] S. Qazzafi, "Consumer buying decision process toward products," *International Journal of Scientific Research and Engineering Development*, vol. 2, no. 5, pp. 130–134, 2019.

[7] E. O'Neill, D. Moore, L. Kelleher, and F. Brereton, "Barriers to electric vehicle uptake in Ireland: perspectives of car-dealers and policy-makers," *Case Studies on Transport Policy*, vol. 7, no. 1, pp. 118–127, 2019.

[8] F. D. O. Santini, W. J. Ladeira, V. A. Vieira, C. F. Araujo, and C. H. Sampaio, "Antecedents and consequences of impulse buying: a meta-analytic study," *RAUSP Management Journal*, vol. 54, no. 2, pp. 178–204, 2019.

[9] M. Xiao, "Annual report on the trend of automobile consumption in China (2012)," in *Development of a Society on Wheels*, pp. 285–307, Springer, Berlin, Germany, 2019.

[10] E. Vogl, R. Pekrun, and K. Loderer, "Epistemic emotions and metacognitive feelings," in *Trends and Prospects in Metacognition Research across the Life Span*, pp. 41–58, Springer, Berlin, Germany, 2021.

[11] J. Liang, R. Li, and Q. Jin, "Semi-supervised multi-modal emotion recognition with cross-modal distribution matching," in *Proceedings of the 28th ACM International Conference on Multimedia*, pp. 2852–2861, New York, NY, USA, October 2020.

[12] S. Parthasarathy and C. Busso, "Semi-supervised speech emotion recognition with ladder networks," *IEEE/ACM*

Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 2697–2709, 2020.

[13] Q. Xie, Z. Dai, E. Hovy, T. Luong, and Q. Le, "Unsupervised data augmentation for consistency training," *Advances in Neural Information Processing Systems*, vol. 33, pp. 6256–6268, 2020.

[14] K. Sohn, D. Berthelot, N. Carlini et al., "FixMatch: simplifying semi-supervised learning with consistency and confidence," *Advances in Neural Information Processing Systems*, vol. 33, pp. 596–608, 2020.

[15] B. Zhang, Y. Wang, W. Hou et al., "FlexMatch: boosting semi-supervised learning with curriculum pseudo labeling," *Advances in Neural Information Processing Systems*, vol. 34, pp. 18408–18419, 2021.

[16] N. Lu, T. Zhang, G. Niu, and M. Sugiyama, "Mitigating overfitting in supervised classification from two unlabeled datasets: a consistent risk correction approach," in *Proceedings of the International Conference on Artificial Intelligence and Statistics*, pp. 1115–1125, PMLR, Sicily, Italy, August 2020.

[17] F. Tao, J. Cheng, Q. Qi, M. Zhang, H. Zhang, and F. Sui, "Digital twin-driven product design, manufacturing and service with big data," *International Journal of Advanced Manufacturing Technology*, vol. 94, no. 9-12, pp. 3563–3576, 2018.

[18] M. Ding, L. Zhao, H. Pei, and M. Song, "An XGBoost based evaluation methodology of product color emotion design," *Journal of Advanced Mechanical Design, Systems, and Manufacturing*, vol. 15, no. 6, Article ID 2021jamdsm0075, 2021.

[19] Z. Guo, L. Lin, M. Yang, and Y. Zhang, "Product image extraction model construction based on multi-modal implicit measurement of unconsciousness," *Jisuanji Jicheng Zhizao Xitong/Computer Integrated Manufacturing Systems, CIMS*, vol. 28, no. 4, pp. 1150–1163, 2022.

[20] S. Lyu, M. Chang, D. Du et al., "UA-DETRAC 2017: report of AVSS2017 & IWT4S challenge on advanced traffic monitoring," in *Proceedings of the 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1–7, IEEE, Lecce, Italy, September 2017.

[21] F. Yu, H. Chen, X. Wang et al., "Bdd100k: a diverse driving dataset for heterogeneous multitask learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2636–2645, Seattle, WA, USA, June 2020.

[22] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: the kitti dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.

[23] Z. Dong, Y. Wu, M. Pei, and Y. Jia, "Vehicle type classification using a semisupervised convolutional neural network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2247–2256, 2015.

[24] X. Li and Openits, *OpenData V11.0-Vehicle Recognition ID Dataset VRID-1*, 2021.

[25] Z. Xu, W. Yang, A. Meng et al., "Towards end-to-end license plate detection and recognition: a large dataset and baseline," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 255–271, Munich, Germany, September 2018.

[26] Z. Che, G. Li, T. Li et al., "D $\overset{*}{2}$-city: a large-scale dashcam video dataset of diverse traffic scenarios," 2019, https://arxiv.org/abs/1904.01975.

[27] P. Chen, X. Bai, and W. Liu, "Vehicle color recognition on urban road by feature context," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 2340–2346, 2014.

[28] X. Liu, W. Liu, H. Ma, and H. Fu, "Large-scale vehicle re-identification in urban surveillance videos," in *Proceedings of the 2016 IEEE international Conference on Multimedia and Expo (ICME)*, pp. 1–6, IEEE, Seattle, WA, USA, July 2016.

[29] H. Liu, Y. Tian, Y. Yang, L. Pang, and T. Huang, "Deep relative distance learning: tell the difference between similar vehicles," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2167–2175, Las Vegas, NV, USA, in.

[30] P. Wang, B. Jiao, L. Yang et al., "Vehicle re-identification in aerial imagery: dataset and approach," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 460–469, Seoul, South Korea, October 2019.

[31] K. Yan, Y. Tian, Y. Wang, W. Zeng, and T. Huang, "Exploiting multi-grain ranking constraints for precisely searching visually-similar vehicles," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 562–570, Venice, Italy, October 2017.

[32] L. Yang, P. Luo, C. Change Loy, and X. Tang, "A large-scale car dataset for fine-grained categorization and verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3973–3981, Boston, MA, USA, June 2015.

[33] J. Krause, M. Stark, J. Deng, and L. Fei-Fei, "3d object representations for fine-grained categorization," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 554–561, Sydney, Australia, December 2013.

[34] N. Gervais, *The Car Connection Picture Dataset*, 2020.

[35] Y. Ouali, C. Hudelot, and M. Tami, "An overview of deep semi-supervised learning," 2020, https://arxiv.org/abs/2006.05278.

[36] I. Z. Yalniz, H. Jégou, K. Chen, M. Paluri, and D. Mahajan, "Billion-scale semi-supervised learning for image classification," 2019, https://arxiv.org/abs/1905.00546.

[37] N. T. U. Nhi, T. M. Le, and T. T. Van, "A model of semantic-based image retrieval using C-tree and neighbor graph," *International Journal on Semantic Web and Information Systems*, vol. 18, no. 1, pp. 1–23, 2022.

[38] M. Chen, J. Yan, T. Gao, Y. Li, and H. Ma, "Duplicate image representation based on semi-supervised learning," *International Journal of Grid and High Performance Computing*, vol. 14, no. 1, pp. 1–13, 2022.

[39] Y. F. Li and D. M. Liang, "Safe semi-supervised learning: a brief introduction," *Frontiers of Computer Science*, vol. 13, no. 4, pp. 669–676, 2019.

[40] D. Lee, "Pseudo-label: the simple and efficient semi-supervised learning method for deep neural networks Workshop on challenges in representation learning," *ICML*, vol. 3, p. 896, 2013.

[41] J. Wang, C. Hq Ding, S. Chen, C. He, and B. Luo, "Semi-supervised remote sensing image semantic segmentation via consistency regularization and average update of pseudo-label," *Remote Sensing*, vol. 12, no. 21, p. 3603, 2020.

[42] X. Li, H. Yin, K. Zhou, and X. Zhou, "Semi-supervised clustering with deep metric learning and graph embedding," *World Wide Web*, vol. 23, no. 2, pp. 781–798, 2020.

[43] C. Tang, Q. Zhu, W. Wu, W. Huang, C. Hong, and X. Niu, "PLANET: improved convolutional neural networks with image enhancement for image classification," *Mathematical Problems in Engineering*, vol. 202010 pages, Article ID 1245924, 2020.

[44] A. Sedik, M. Hammad, F. E. Abd El-Samie, B. B. Gupta, and A. A. Abd El-Latif, "Efficient deep learning approach for augmented detection of Coronavirus disease," *Neural*

Computing and Applications, vol. 34, no. 14, pp. 11423–11440, 2022.

[45] A. M. Srivastava, P. A. Rotte, A. Jain, and S. Prakash, "Handling data scarcity through data augmentation in training of deep neural networks for 3D data processing," *International Journal on Semantic Web and Information Systems*, vol. 18, no. 1, pp. 1–16, 2022.

[46] M. A. Rahman, M. S. Hossain, N. A. Alrajeh, and B. B. Gupta, "A multimodal, multimedia point-of-care deep learning framework for COVID-19 diagnosis," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 17, no. 1s, pp. 1–24, 2021.

[47] E. D. Cubuk, B. Zoph, J. Shlens, and Q. V. Le, "Randaugment: practical automated data augmentation with a reduced search space," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 702-703, Seattle, WA, USA, June 2020.

[48] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[49] H. Inoue, "Data augmentation by pairing samples for images classification," 2018, https://arxiv.org/abs/1801.02929.

[50] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "Mixup: beyond empirical risk minimization," 2017, https://arxiv.org/abs/1710.09412.

[51] G. Chen, Y. Liu, J. Wang et al., "PP-matting: high-accuracy natural image matting," 2022, https://arxiv.org/abs/2204.09433.