

## *Retraction*

# **Retracted: A Hybrid Model for Commercial Brand Marketing Prediction Based on Multiple Features with Image Processing**

## **Security and Communication Networks**

Received 26 December 2023; Accepted 26 December 2023; Published 29 December 2023

Copyright © 2023 Security and Communication Networks. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.


The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

## **References**

- [1] F. Wang and L. Zhao, "A Hybrid Model for Commercial Brand Marketing Prediction Based on Multiple Features with Image Processing," *Security and Communication Networks*, vol. 2022, Article ID 5455745, 10 pages, 2022.

## Research Article

# A Hybrid Model for Commercial Brand Marketing Prediction Based on Multiple Features with Image Processing

Furong Wang<sup>1</sup> and Li Zhao<sup>2</sup> 

<sup>1</sup>Business School, Shanxi Technical College of Finance and Economics, Taiyuan 712000, China

<sup>2</sup>School of Management Engineering and Business, Hebei University of Engineering, Handan 056038, China

Correspondence should be addressed to Li Zhao; zhaoli@hebeu.edu.cn

Received 4 July 2022; Revised 9 August 2022; Accepted 18 August 2022; Published 15 September 2022

Academic Editor: Hangjun Che

Copyright © 2022 Furong Wang and Li Zhao. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Recently, deep learning has been employed in automatic feature extraction and has made remarkable achievements in the fields of computer vision, speech recognition, natural language processing, and artificial intelligence. Compared with the traditional shallow model, deep learning can automatically extract more complex features from simple features, which reduces the intervention of artificial feature engineering to a certain extent. With the development of the Internet and e-commerce, picture advertising, as an important form of display advertising, has the characteristics of high visibility, strong readability, and easy-to-obtain user recognition. An increasing number of Internet companies are paying attention to what kind of advertising pictures can attract more clicks. Based on deep learning technology, this paper studies the prediction model of click-through rate (CTR) for advertising and proposes an end-to-end CTR prediction depth model for display advertising, which integrates the feature extraction of display advertising and CTR prediction to directly predict the probability of an advertisement image being clicked by users. This paper studies the deep-seated nonlinear characteristics through the multilayer network structure of the deep network and carries out several groups of experiments on the private display advertising data set of a commercial advertising platform. The results show that the model proposed in this paper can effectively improve the prediction accuracy of CTR compared with other benchmark models and predict whether an advertisement is clicked or not by given advertisement information and user information. By establishing a reasonable advertising click-through rate prediction model, it can help the platform estimate future revenue so as to make cooperative decisions with advertisers. For advertisers, it is necessary to evaluate the price by predicting the click-through rate and estimate the bidding price of their own advertisements.

## 1. Introduction

**1.1. Research Background and Significance.** Online advertising, also known as online marketing, Internet advertising, or web advertising, is a form of advertising marketing that uses the Internet to deliver marketing information to consumers, including e-mail marketing, search engine marketing (SEM), social media marketing, and various forms of display advertising (such as web banner advertising) and mobile advertising. Compared with traditional TV, radio, magazines and newspapers, and other types of media advertising, Internet advertising has many natural advantages, mainly reflected in coverage, accurate positioning, and wide audience and targeting clear, fast, real-

time, update, flexible, and open interaction, the cost is more economic, and so on, thus widely favored by the industry and become an important part of modern marketing media strategy for enterprises. Online advertising is a multibillion-dollar business that generates huge revenues for many Internet companies, including Alibaba, Baidu, and Google.

Since its birth, online advertising has not stopped the pace of formal innovation. From the initial agreement-based advertising, which mainly focuses on display, it has continuously enriched the display content and methods of advertising with the progress of technology [1]. Online advertising has a variety of forms. According to the characteristics of application scenarios, it can be divided into sponsored search advertising (SSA) and contextual

advertising (CA) and display advertising [2]. SSA refers to the fact that advertisers determine keywords, titles, product descriptions, and other related attributes for their products according to the characteristics and incontinence of their products and conduct independent bidding for the advertising keywords [3]. CA refers to the commercial text advertising or display advertising related to web content automatically displayed at a certain position of the web page when Internet users browse the web page [4]. Display advertisement refers to the online advertisement directly displayed in the form of text, picture, and video when users browse the web page [5]. As an important form of display advertising, display advertising has the characteristics of high visibility, strong readability, and easy-to-obtain user recognition, and its application is increasing widely.

At present, most related work is mainly based on text and picture features and uses the features calculated by statistical learning methods to estimate the CTR. In this case, people usually extract colors (RGB, LAB, or HSV), brightness, saturation, texture, histogram of oriented gradient (HOG) [6], and scale-invariant feature transform (SIFT) [7] and apply them to advertising hit rate prediction tasks.

These picture features have weak generalization ability and are only effective for specific picture advertisements. They cannot be adjusted accordingly according to the changes in application scenarios. Effective picture features need to be screened manually. In addition, each feature is independent of each other, and the correlation between features is not fully reflected in the CTR prediction process. Constructing feature combinations is one of the important ways to mine the associated information between features. The traditional methods of constructing feature combination mainly rely on manual and prior knowledge and cannot construct higher dimensional feature combination. Therefore, how to automatically extract the visual features of pictures and fully mine the correlation between features with less manual intervention is of great research significance for the prediction of display advertising CTR.

Deep learning has become one of the hottest directions in the field of machine learning in recent years and has made amazing achievements in many application fields, such as speech recognition [8], image recognition [9], and business [10]. One of the core problems solved by deep learning is to automatically extract more complex features from simple features, so it is very suitable for data expression and feature extraction in the task of display advertising CTR prediction. In this paper, deep learning technology is applied to the prediction of display advertising CTR, which automatically completes the learning and combination of features, reduces the labor consumption of feature engineering, and improves the accuracy of advertising CTR prediction. Specifically, this paper divides the original features of display advertising into two parts, picture visual features and other basic features, and the Convolutional Neural Network (CNN) [11] in deep learning is used to extract the high-level semantic features of advertising pictures and enhance the representation ability of advertising picture features; then, the deep-level internal relationship between more features is mined

through deep neural network, so as to more effectively improve the effect of CTR prediction.

*1.2. Related Work and Analysis.* The accurate prediction of advertising CTR not only helps to improve the user experience but also is one of the important revenue sources of global Internet companies. It has important commercial value and academic research value. It has become an important research field in industrial and academic circles in recent years. On the whole, feature learning and feature combination are the key factors in improving the accuracy of display advertising CTR prediction. This paper will introduce the related research work of CTR prediction from these two angles.

In the aspect of feature learning, traditional research lacks effective methods to extract the high-level semantic features of advertising pictures. The traditional picture features are manual design features, which often focus on one aspect of the image, with limited representation ability and unable to capture the key high-level semantic features of the picture [12, 13]. For example, the hog feature focuses on the edge information of the picture, the SIFT feature focuses on the interest points of the local appearance of the picture, and the local binary pattern (LBP) [14] focuses on the texture of the picture. These picture features are only effective for specific tasks and cannot be adjusted according to different application scenarios to attract users' clicks. It is difficult to design and include all visual feature items highly related to CTR. The breakthrough development of deep learning provides a new idea to solve this problem and has been successfully applied to many image recognition problems [9, 15, 16]. Among them, CNN has made amazing achievements in the image field. It directly uses the original pixels of the picture as the input and retains all the information of the input picture to the greatest extent. The image features extracted by convolution and pooling operations of convolution neural network have good generalization. It not only achieves good performance in image classification tasks but also can be well generalized to other computer vision tasks, such as object detection, semantic segmentation, and video tracking. At present, there have been many researches on how to predict the CTR of advertising by using the high-level semantic features extracted by CNN. Feature learning, an image advertising feature learning architecture based on CNN [17], uses CNN to directly learn the high-level semantic features of images with recognition from the original pixels and user click feedback and then combine the basic features from advertising information. The CTR of each advertisement is predicted by logistic regression (LR) [18]. Feature learning solves the problem of data sparsity and cold start in advertising CTR prediction by adding additional picture high-level semantic features. It is one of the earlier works to apply picture high-level semantic features to advertising CTR prediction task. The deep CTR model proposed by Chen et al. [19] combines the picture high-level semantic features extracted by the convolution neural network with the advertising context features extracted by full connection layer; after batch normalization (BN) [20], it is

fed into the fully connected network together to integrate feature extraction and CTR prediction, to realize end-to-end training. DICM [21] (Deep Image CTR Model) model uses CNN to extract high-level semantic features from candidate advertising pictures and pictures clicked by users and jointly predict the probability that an advertisement is clicked by users by combining advertising features and user features.

In the aspect of constructing feature combination, most studies solve the problem of advertising CTR prediction based on traditional models. LR model has the characteristics of simple implementation, strong interpretability, and easy parallel. It is widely used in industry. However, in the LR model, the meaning of each dimension of features is fixed and isolated, which cannot fully mine the nonlinear association between features. Therefore, the LR model needs to manually construct combined features to mine the association between features, which has many problems, such as low efficiency and unable to migrate domain knowledge. Factorization machine (FM) uses matrix decomposition, regards the weight matrix as the inner product of two identical hidden vectors, and constructs the second-order combination between any two features [22], which effectively solves the problem of feature combination under large-scale sparse data. It is widely used in recommendation system, advertising CTR prediction, and other applications. Juan et al. [23] draw on the concept of the field. Jahrer et al. [24] add domain information to features on the basis of FM and propose a domain perception decomposition machine (field aware factorization machine, FFM) model. FFM model combines features with the same or similar properties into the same domain, so the implicit vector of features is not only related to the feature itself but also related to its corresponding domain. Compared with the FM model, the FFM model learns one more layer of domain information and better excavates the implicit information in features. FM model and FFM model can be constructed with any height in theory. However, due to the computational complexity, it can only express the pairwise combination relationship between features. In addition, the FM model constructs feature combination based on the linear method, which cannot fully mine the highly nonlinear association between features. Deep learning is famous for its ability to learn deep-seated nonlinear features, which reduces the intervention of artificial feature engineering to a certain extent. Most research methods use the corresponding network structure to automatically learn the high-order combination between features. The factor decomposition machine-supported neural network (FNN) proposed by Zhang et al. [25] adds a neural network on the basis of the FM model to better mine the internal relationship between features. Qu et al. [26] proposed a product-based neural networks (PNN) model based on vector product, which uses the second-order vector integration (pair wisely connected product layer) to carry out pairwise vector product on the embedded vector of FM model and then inputs it into a fully connected neural network to effectively construct high-order feature combination. Aiming at the defect that the FM model gives the same weight to all combined features, Xiao et al. [27] proposed a factor decomposition machine model based on

the attention mechanism (Attentional Factorization Machine, AFM), introducing the popular attention mechanism [28], automatically learning different weights of second-order combined features, paying attention to feature combinations with high influence, reducing the impact of invalid or even interfering information, better mining the internal correlation between features, and playing a certain role in improving the prediction effect. The classical width and depth model is proposed by Cheng et al. [29] (Wide & Deep, WDL model), the LR model in the width part and the neural network in the depth part are jointly trained, and the final fusion model effectively constructs the low-order combination and high-order combination of features, which has been successfully used in app ranking recommended by the app store. In order to extend the second-order combination of the FM model to a high-order combination, the depth and cross-network proposed by Wang et al. [30] (Deep & CrossNetwork, DCN) uses polynomial multiplication to directly extract by depth network to comprehensively mine the internal correlation between features.

To sum up, in view of the outstanding performance of CNN in the image field, this paper uses CNN to extract high-level semantic features of images with strong generalization ability. However, the current CNN is mainly used in image classification and object recognition tasks with category labeling information training and cannot be applied to advertising image feature extraction tasks for image advertising CTR prediction. Therefore, this paper needs to improve the existing convolution neural network structure to make it suitable for the feature extraction task of advertising pictures. Learning the high-level features of advertising pictures and fully mining the correlation between features by constructing combined features is the research content of this paper. It can be seen from the above that the deep learning neural network has a high ability for feature combination and construction, which has a certain reference significance for mining the correlation between features. This paper uses its multilayer network structure to mine the nonlinear correlation between features so as to more effectively improve the accuracy of hit rate prediction.

## 2. Method

In display advertising, most of the related work is mainly based on text and picture features, and the statistical learning method is used to estimate the CTR. In this case, most of the traditional methods of extracting advertising picture features are manual extraction based on special purposes, with weak generalization ability and only effective for specific tasks, requiring manual screening of effective visual features. In addition, each feature is independent of each other, and the correlation between features is not fully reflected in the CTR prediction process. It is necessary to rely on manual experience to construct combined features, which has many problems, such as low efficiency and unable to transfer domain knowledge. Aiming at the problem that traditional methods cannot filter features quickly and effectively and fully mine the correlation between features, this paper

proposes an end-to-end CTR prediction depth model for display advertising by using deep learning technology.

*2.1. Symbol Definition and Problem Description.* In order to explain the depth prediction model of display advertising CTR proposed in this paper, the important symbols and definitions in this paper are summarized in Table 1.

*2.2. Problem Description.* In order to publish activities in the display advertising system, advertisers upload their advertising pictures and specify the targeted target of the product (user division, time, region, etc.) and the advertising budget during the event. At the same time, advertisers will also allocate a small amount of budget to purchase statistics or institutional data, learn the user's feedback mode, and carry out effective audience orientation. When users initiate a web page request, advertisers use advertising pictures and corresponding historical CTR data to estimate the CTR of this display and purchase the opportunity to present the advertisement to the current user from the advertiser according to the estimated value. In this paper,  $x$  is used to represent an advertising picture. In advertising pictures, the features are divided into two parts: basic features and picture features. The basic features are generally advertiser ID, advertiser name, advertising space ID, advertising space name, advertising category ID, creative image width, creative image height, and so on. In this paper,  $p = [p_1, p_2, \dots, p_M]$  represents the basic features of advertising picture  $x$ , where  $M$  represents the number of basic features. The visual feature of the advertising picture is the advertising creative picture of  $224 \times 224 \times 3$ , where  $224 \times 224$  represents the size of the creative picture and  $\times 3$  represents that the creative picture is a three-channel color picture; that is, there are three channels under the RGB color model in the three primary color light mode, namely, red, yellow, and blue. In this paper, the creative image is marked as the three-dimensional matrix  $G \in R^{224 \times 224 \times 3}$  of  $224 \times 224 \times 3$ , and each element value in the matrix represents the RGB three-channel color value of the pixel in the creative image. In this paper, each advertising image  $x$  is defined as a binary composed of the basic feature  $p$  and the pixel matrix  $G$  of the creative image, that is,  $X = (p, G)$ .

Suppose that the training data set  $D$  is a data set containing  $N$  advertising pictures and corresponding CTRs.  $X = \{x_1, x_2, \dots, x_N\}$  is defined as the advertising picture data set, and the corresponding CTR is  $Y = \{y_1, y_2, \dots, y_N\}$ , where  $y_i \in [0, 1]$  indicates that the real CTR of advertising picture  $x_i$  is  $y_i$ . Given  $n$  advertising pictures  $X$  and corresponding CTR  $Y$ , the specific definition of training data set  $D$  is shown in

$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}. \quad (1)$$

Among them, each sample  $(x_i, y_i) \in D_i$  indicates that the real CTR of advertising picture  $x_i$  is  $y_i$ .

Given a display advertising training data set  $D$ , the problem of estimating the CTR of display advertising can be formally defined as follows: based on the training data set  $D$ , learn a model  $f(\cdot)$  to estimate the probability  $\hat{y}$  of users

clicking on the picture advertisement on the page after opening the page; see

$$\hat{y} = f(x; \theta), \quad (2)$$

where  $f(\cdot)$  represents the CTR prediction model,  $\hat{y} \in (0, 1)$  represents the click-through probability value calculated by the prediction model  $f(\cdot)$  for a given advertising picture  $x$ , and  $\theta$  represents the relevant parameters of model  $f(\cdot)$ .

Based on the definition of the above concepts and problems, the problem definition of picture advertisement CTR prediction based on deep learning studied in this paper is given as follows: given a picture advertisement training data set  $D$ , the task of the picture advertisement CTR prediction depth model is to automatically learn a prediction model  $f^*$  with the smallest error from the training data set  $D$ , so given a picture advertisement training data set  $D$ , the task of the picture advertisement CTR prediction depth model is to automatically learn a prediction model  $f^*$  with the smallest error from the training data set  $D$ , so as to minimize the error between the predicted value  $\hat{y}$  and the real value  $y$  given by the model as shown in

$$f^* = \arg \min (\text{error}(\hat{y}, y)). \quad (3)$$

*2.3. Display Advertising CTR Prediction and Optimization Objectives.* The CTR prediction of display advertising uses the basic characteristics, including advertisers, advertising spaces, advertising categories, creative image attributes, and other pieces of information, as well as the visual characteristics of advertising pictures, to predict the probability that an advertising picture is clicked by users. Therefore, this paper defines it as a regression problem to predict the specific CTR value. The most used performance measure for regression tasks is square loss, also known as the mean square error (MSE). It uses Euclidean distance as the measurement error to represent the difference between the predicted value and the real value. The smaller the loss function is, the closer the predicted value of the model is to the real value [31]. Therefore, this paper uses the square loss to measure the error of the model. The error of a single sample is defined in

$$\text{error}(\hat{y}, y) = (\hat{y} - y)^2. \quad (4)$$

Then, the loss function of the whole data set  $D$  can be defined as

$$L(\theta) = \frac{1}{N} \sum_1^N (\hat{y}_i - y_i)^2 + \lambda \theta_2^2. \quad (5)$$

Among them,  $\theta$  represents the relevant parameters of the prediction model,  $\theta_2^2$  represents the  $L_2$  norm regular term, which is used to prevent overfitting problems in the process of parameter optimization,  $N$  represents the number of samples, and  $y_i$  represents the real value of the  $i$ th sample.

Based on the definition of (5), the overall optimization objective of the display advertising CTR prediction depth model can be given, as shown in

TABLE 1: Important symbols and meanings.

Symbol	Meaning description
$x$	Represents an advertising picture
$p$	Represents the basic characteristics of an $x$ , represented by a multidimensional vector
$G$	Represents a picture feature of $x$ , corresponding to a three-dimensional matrix $G \in R^{224 \times 224 \times 3}$ representing the pixel information of the advertising creative image. Each element value in the matrix represents the RGB three-channel color value of the pixel in the creative image
$y$	Represents the real CTR of $x$
$X$	Advertising picture dataset $X = \{x_1, x_2, \dots, x_N\}$
$Y$	Real hit rate set $Y = \{y_1, y_2, \dots, y_N\}$
$D$	Training data set $D$
$N$	Number of training set samples
$f(x; \theta)$	Advertising CTR prediction depth model $\theta$ represents the correlation of the prediction model
$\hat{y}$	Click probability predicted by the model
$L(f(x; \theta))$	Loss function of display advertising CTR prediction depth model

$$\theta^* = \arg \min \left( \frac{1}{N} \sum_1^N (\hat{y}_i - y_i)^2 + \lambda \theta_2^2 \right). \quad (6)$$

As can be seen from (6), the goal of the picture advertisement CTR prediction depth model is to solve a set of parameters  $\theta^*$  of the model, which can minimize the square loss  $L$  based on the training data set  $D$ , so that the predicted value  $\hat{y}$  calculated by the model is as close as possible to the real value  $y$ . Recognition of handwritten character strings is a process of analyzing and processing handwritten note images and segmenting and recognizing characters to obtain electronic texts. This process needs to collect a large number of data images to improve the accuracy of codes.

#### 2.4. Depth Prediction Model of Display Advertising CTR.

This section first gives the overall network structure of the display advertising CTR prediction depth model proposed in this paper. Then, the basic principles of different components in the network are described in detail. Finally, the algorithm flow of the display advertising CTR prediction depth model proposed in this paper is completely displayed. The residual network still satisfies the nonlinear layer and then directly introduces a short connection from the input to the output of the nonlinear layer so that the whole mapping becomes a complete mapping. This is the core formula of the residual network; in other words, the residual is an operational construction of the network, and any network that uses such an operation can be called a residual network.

**2.4.1. Model Overview.** Figure 1 shows the overall network structure of the display advertising CTR prediction depth model. As can be seen from Figure 1, the model proposed in this paper is an end-to-end prediction model integrating feature extraction of display advertising and CTR prediction. The network structure of the model can be divided into three main parts from low level to high level: (1) residual network for extracting high-level semantic features of creative image; (2) embedding layer for transforming basic features into low-dimensional real number vectors; (3) mining the fully connected network between features and the output layer to output the hit rate prediction results. The remaining sections

of this paper will introduce the specific structure and basic principles of these three parts in detail. This shows the specific operation process of ctr, which is different from the continuous and dense data in the fields of image and speech, and the local correlation in space and time is good.

**2.4.2. Residual Network.** In this paper, the residual network is used to extract the high-level semantic features of advertising creativity image. The overall network structure is shown in Figure 2. The residual network further satisfies the nonlinear layer and then directly short-circuits the output of the nonlinear layer to the output so that the entire mapping becomes the core formula of the residual network. That is, the residual of any network that uses this function can be called a residual network.

**2.4.3. Basic Feature Embedding.** In display advertising, the basic features are usually composed of a large number of discrete features and a few continuous numerical features. The embedding layer transforms a high-dimensional sparse vector into a low-dimensional real vector, which can effectively reduce the dimension of features. In this paper, the embedding layer is used to image the basic features to several low-dimensional real number vectors to alleviate the sparsity of the basic features in display advertising. The following describes the basic characteristics of the single heat coding type and the numerical coding type, respectively.

The embedding process of unique heat coding type features is shown in Figure 3, and its input is a unique heat coding vector. If the  $i$ th basic feature  $p_i$  is a unique hot coding type feature, its embedding process is shown in

$$e_i = V_i p_i, \quad (7)$$

where  $e_i \in R^K$  represents the embedding representation of the unique heat coding feature  $p_i$  learned through the embedding layer,  $K$  represents the vector dimension after embedding,  $V_i \in R^{n_i \times K}$  represents the embedding matrix, and  $n_i$  represents the dimension of the binary vector  $p_i$ .

This is a summary of the entire process of feature selection. The so-called embedded feature selection is to fit the data through some special models, then use some attributes

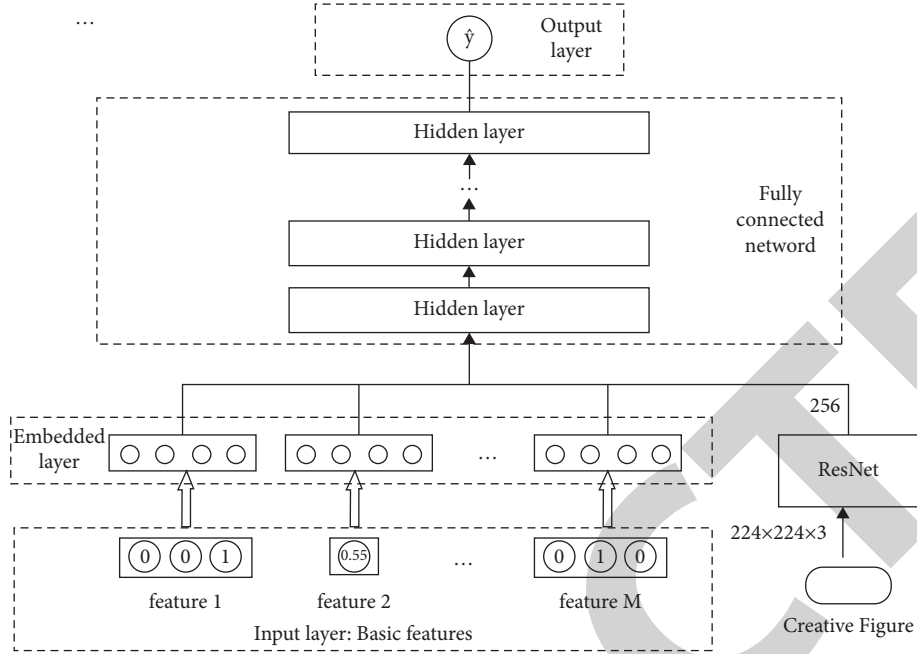


FIGURE 1: Overall network structure of display advertising CTR prediction depth model.

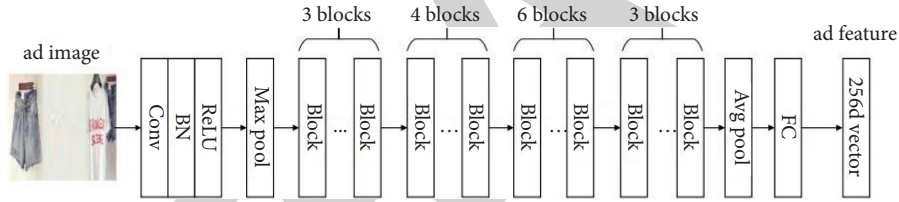


FIGURE 2: The overall network structure.

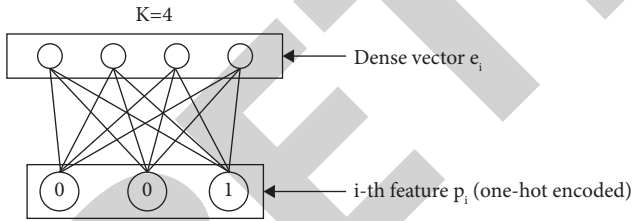


FIGURE 3: Schematic diagram of embedding layer of unique heat coding type features.

of the model itself to evaluate the features as evaluation indicators, and finally a selection is made using a packaging feature selection method. Of course, in many cases, we still stay at the stage of calculating the evaluation index, because the biggest problem of packaging feature selection is that the amount of calculation and time are the largest of the three. Assuming that the  $j$ th basic feature  $p_j$  is a numerical feature, the embedding process is shown in

$$e_j = V_j P_j, \quad (8)$$

where  $e_j \in R^K$  represents the numerical feature,  $p_j$  represents the embedded representation learned through the embedded layer,  $K$  represents the vector dimension after

embedding, and  $V_j \in R^K$  represents the embedded vector corresponding to  $p_j$ . Figure 4 shows the embedded layer of numerical features.

Connect the features of an advertising picture to generate a more comprehensive and effective feature expression  $e$  for an advertising picture  $x$ , as shown in

$$e = [e_1, e_2, \dots, e_M, e_G]. \quad (9)$$

**2.4.4. Fully Connected Network.** In this paper, a fully multilayer connected neural network is used to fully mine the nonlinear relationship between features, to improve the prediction results of CTR more effectively. There is no interconnection between neurons in the same layer. Each neuron is only connected to all neurons in the previous layer, receives the output of the previous layer, and inputs it to the next layer. The first layer is called the input layer. In this paper, the advertising picture feature  $e$  is used as the input  $a^{(0)}$  of the fully connected neural network, as shown in

$$a^{(0)} = [e_1, e_2, \dots, e_M, e_G]. \quad (10)$$

The fully connected neural network contains one or more hidden layers, and each hidden layer performs the calculation of (11) and (12):



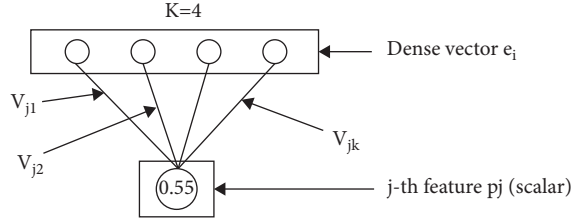


FIGURE 4: Schematic diagram of embedded layer of numerical features.

$$z^{(l+1)} = W^{(l)} a^{(l)} + b^{(l)}, \quad (11)$$

$$a^{(l+1)} = \sigma^{(l)}(z^{(l+1)}), \quad (12)$$

where  $l$  represents the depth of the hidden layer,  $\sigma^{(l)}$  represents the activation function of layer  $l$ ,  $m^{(l)}$  represents the

number of neurons in layer  $l$ ,  $z^{(l+1)} \in R^{m^{(l+1)} \times m^{(l)}}$  represents the net input of neurons in layer  $l+1$ ,  $a^{(l)} \in R^{m^{(l)}}$  represents the output of neurons in layer  $l$ , and  $W^{(l)} \in R^{m^{(l+1)} \times m^{(l)}}$  represents the weight matrix from layer  $l$  to layer  $l+1$ . Combine (11) and (12) to obtain

$$a^{(l+1)} = \sigma^{(l)}(W^{(l)} a^{(l)} + b^{(l)}), \quad (13)$$

In this way, the fully connected neural network can use its multilayer network structure to automatically learn the nonlinear correlation between features and obtain relatively high-order combined features. Among them, different hidden layers in multilayer networks are different potential representations of the input layer. They form more abstract high-level features by combining low-level features, and they can fully mine the association between features. Finally, the output definition of the fully connected network is shown in

$$h_L = \sigma^{(L)}(W^{(L)}(\sigma^{(L-1)}(W^{(L-1)} \dots \sigma^{(1)}(W^{(1)} a^{(0)} + b^{(1)})) + b^{(L-1)} + b^{(L)}), \quad (14)$$

where  $L$  represents the number of hidden layers in the fully connected network.

Finally,  $h$  is passed through the regression layer to obtain the predicted value  $\hat{y}$  of CTR, as shown in

$$\hat{y} = Wh_L + b, \quad (15)$$

where  $W$  represents the weight matrix of the regression layer and  $b$  represents the offset vector.

**2.4.5. Optimization Algorithm.** In order to better learn the depth prediction model of display advertising CTR proposed in this paper, we use the adaptive motion estimation (Adam) algorithm with an adaptive learning rate to optimize the objective function  $L$  defined in formula (5) [32]. In this paper, the first-order moment estimation and second-order moment estimation of the gradient are calculated by the Adam optimizer, independent adaptive learning rates are designed for different parameters, and only a few parameters are needed. Firstly, the algorithm initializes the relevant parameters  $\theta$  of the prediction model, the first-order moment variable  $s$ , the second-order moment variable  $R$ , and the time step  $t$  of the Adam optimizer; Then, taking  $m$  as a minibatch, a group of samples  $\{x^{(1)}, \dots, x^{(m)}\}$  are taken, where  $x^{(i)}$  corresponds to  $y^{(i)}$ , and the sample prediction is completed according to the given CTR prediction depth model  $f(x, \theta)$ . Finally, the objective function  $L$  is calculated, and the corresponding model parameter value  $\theta$  is updated based on the gap between the predicted value and the real value so that the prediction result of the model on this minibatch of data is closer to the real CTR.

### 3. Experimental Design and Result Analysis

#### 3.1. Experimental Dataset

**3.1.1. Data Description.** The data set used in this experiment comes from the real data set of display advertising privately owned by a commercial advertising platform. During the experiment, we collected all data from November 21 to 23, 2017, including 100000 samples involving 55725 creative images. In this paper, all fields of the display advertising data set are collected to form Table 2. In the display advertising dataset of this paper, each sample contains three kinds of information: (1) the advertiser, advertising space, advertising category, and corresponding creative attributes of an advertising picture, in which the creative attributes describe the relevant information of the corresponding advertising creative image from the aspects of format, name, width, height, and URL; (2) the creative image corresponding to the advertising picture; (3) the actual CTR of the advertising picture.

Data format description and sample display based on Table 2.

**3.2. Data Preprocessing.** The purpose of data preprocessing is to convert display advertising data into recognizable input during model operation. This section introduces the data preprocessing methods of the experiment from two aspects: creative diagram and basic features.

**3.2.1. Image Feature Preprocessing.** The image advertising data set in this paper contains a variety of creative drawings of different sizes, such as  $160 \times 200$ ,  $240 \times 164$ ,  $320 \times 175$ ,



TABLE 2: Detailed data format of display advertising data set.

Serial number	Attribute field	Sample	Description
1	advertiserID	1046141401	Advertiser ID
2	advertiserName	Pink-girl European and American women's wear	Advertiser name
3	adzoneID	34492608	Adzone ID
4	adzoneName	Wireless_flow packet_ online shopping_ hand amoy app_ hand panning focus image	Adzone name
5	maincateID	1	Maincate ID
6	maincateName	Women's wear	Maincate name
7	Date	2017-11-21	Date
8	ctr	5	CTR (%)
9	adboardID	842306140001	Adboard ID
10	adboardFormat	1	Adboard format
11	adboardName	Good looking series	Adboard name
12	adboardWidth	640	Adboard width
13	adboardHeight	200	Adboard height
14	imageUrl	TB1CfLzpL5TBuNjSspmSuuDRVXa	Image URL
15	templateID	2000000000	Template D

520 × 280, and 640 × 200. In this experiment, the bilinear interpolation method is used to uniformly adjust the original creative image of the data set to 224 × 224 size and make the target image retain all the information on the original creative image as much as possible.

**3.2.2. Basic Feature Preprocessing.** The discrete features in the basic features are processed as follows: the first step is to map some data in order to reduce the data storage space and facilitate feature extraction and data operation. In this paper, the discrete features of string identification are used in statistical samples, and then the character identification is mapped to the range of natural numbers starting from zero. The second step is to reprocess the discrete features in the whole data set. When a feature string is encountered, the natural number corresponding to the string is used to replace the feature.

### 3.3. Benchmark Model and Evaluation Index

**3.3.1. Benchmark Model.** In the experiment, the display advertising CTR prediction depth model proposed in this paper is compared with six other different models. In this paper, these models are divided into two categories, traditional advertising CTR prediction model and deep learning model, and two different traditional prediction models are compared. They are as follows: (1) LM: linear regression is used to estimate the CTR, and only the basic characteristics of advertising pictures are considered in the prediction process [33]; (2) LM\_HOG: linear regression is used to estimate the CTR, and both the basic characteristics of advertising pictures and HOG characteristics are considered in the prediction process [6]. In this paper, five different depth learning models are compared in the experiment, which are as follows:

- (1) DNN\_basic: use the fully connected network to directly predict the CTR of advertising pictures and only consider the basic characteristics of advertising pictures in the prediction process.

- (2) DL\_basic: the depth prediction model of display advertising CTR proposed in Section 3 of this paper only considers the basic characteristics of advertising pictures in the prediction process. DL\_basic model can be regarded as the CTR prediction model after removing the residual network.
- (3) Feature learning: consider both the basic features of advertising images and the high-level features of images. In the data set of this paper, due to the lack of label data of click events corresponding to advertising images, this paper uses advertising images and corresponding CTRs to pretrain its convolution neural network. Then, the trained network is used to extract the high-level semantic features of creative images. Finally, combining basic features with high-level visual features and the probability of an advertisement image being clicked by users is predicted by linear regression.
- (4) DeepCTR: the CTR prediction model of display advertising based on a deep network considers the basic characteristics of advertising pictures and high-level characteristics of pictures at the same time.
- (5) DL: the depth prediction model of display advertising CTR proposed in this paper considers the basic characteristics of advertising pictures and high-level characteristics of pictures at the same time.

For the DNN\_basic model, this paper conducted many comparative experiments and finally chose to set the number of fully connected network layers as 3 and the number of neurons at each layer as 1024. For the DL\_basic model, different combinations of fully connected layers and neurons are compared when the dimension of embedded representation is 5, 10, 15, 20, 25, and 30, respectively. The experimental results show that the DL\_basic model performs best in the case of a 15-dimensional embedding representation and two fully connected layers containing 1024 neurons. For the DL model, this paper uses the Keras application module to provide the ResNet50 network with pretraining weights and fine-tune network and feature

TABLE 3: Performance comparison of different CTR prediction models on test sets.

	Model	RMSE	MAPE (%)
Traditional prediction model	LM	0.0499	85.44
	LM_HOG	0.0455	74.55
Deep learning mode	DNN_basic	0.0365	63.20
	DL_basic	0.0254	44.48
	Feature_learning	0.0375	68.25
	DeepCTR	0.0263	44.80
	DL	0.0207	35.54

extraction on this basis. Specifically, this paper uses the ResNet50 network for image classification, pretrained on the ImageNet dataset, to remove the top full connection layer and add a full connection layer containing 256 neurons for fine tuning.

During the experiment, all models were trained on the training set, and the hyperparameters of the model were adjusted through the verification set. Finally, the performance of the model was tested on the test set. In this paper, the training cycle is set to 50, the batch size is set to 120, and the learning rate was set to 0.001 and reduced to 0.1 of the original learning rate after every 10 cycles. CTR refers to the ordering of the refined layer. Therefore, the candidate ranking set of the CTR model is generally thousands of orders of magnitude. CTR is generally divided into two layers: recall and sorting. Recall is responsible for roughly selecting thousands of items from millions of items. Common algorithms include collaborative filtering and user portraits, which are sometimes called rough sorting layers; sorting is responsible for fine sorting of thousands of items recalled by the recall layer, also called the refinement layer; it calculates the underlying general technology of advertising. Under the CPC/OCPC marketing model, the estimation accuracy plays a very important role in the advertiser's traffic purchase cost and platform monetization efficiency.

*3.4. Performance Comparison Experiment.* Table 3 completely shows the experimental results of performance comparison between the DL model and the other six benchmark models in the image advertising data set of this paper.

From the experimental results, the performance of the LM model based on a single feature is the worst, indicating that the LM model has limited performance and cannot effectively learn the internal association between features. The RMSE and MAPE performance of the DNN\_basic model was significantly better than the LM model, with a relative improvement of 26.9% and 26.0%, respectively. This shows that the multilayer network structure of deep learning can effectively mine the internal correlation between features and provide great help in performance improvement. DL\_basic introduces an embedding layer to transform basic features into low-dimensional real number vectors, which has a relative improvement of 30.4% and 29.6% compared with RMSE and MAPE indexes of the DNN\_basic model.

This shows that the embedded representations learned by the embedding layer can well represent the basic features. After these embedded representations are spliced into vector inputs to the neural network, the internal associations between features can be mined more effectively.

## 4. Summary

With the development of the Internet and e-commerce, display advertising, as an important form of display advertising, has the characteristics of high visibility, readability, and easy-to-obtain user recognition; an increasing number of Internet companies are paying attention to display what kind of advertising pictures can attract more clicks.

In view of the problem that traditional methods cannot quickly and effectively screen image visual features, this paper uses CNN to extract high-level image features quickly and effectively from advertising images and images with large-scale sparse features into low-dimensional dense real vectors through the embedding layer. Thus, the visual features of images are extracted explicitly, and the sparsity of image advertising data sets is solved, which is the basis of the subsequent research.

In addition, this paper uses the multilayer network structure of a deep network to learn the deep-seated non-linear features, to solve the problem that the traditional model only considers each feature independently and does not excavate the information defects hidden between features, thus improving the accuracy of CTR estimation more effectively.

Furthermore, in order to verify the correctness and effectiveness of the model proposed in this paper, we conducted several experiments on the private image advertising data set of a commercial advertising platform. In the performance comparison experiment, the results show that the proposed model can effectively improve the prediction accuracy of advertising CTR compared with other benchmark models. The neural network layers inside DNN can be divided into three categories, input layer, hidden layer, and output layer; as shown in the following figure, the first layer is the input layer, the last layer is the output layer, and the middle layers are the hidden floor. The number of layers of a neural network is calculated in this way. The input layer is not counted. From the hidden layer to the output layer, a total of several layers represents a neural network with several layers. This kind of hierarchical network learning is a problem-solving idea, which corresponds to multistep problem solving; that is, a problem is divided into multiple steps to solve step by step, and the end-to-end data is directly obtained from the input data result. That is, without pre-processing and feature extraction, throw the original data directly into the result. Feature extraction is contained within the neural network, so the neural network is an end-to-end network.

## Data Availability

Experimental data on the results of this study can be obtained from the corresponding author upon request.

## Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

## References

- [1] M. Sanghavi, "Midflight online advertisement campaign optimizer," 2011, <http://ip.com/patapp/US20110112900>.
- [2] M. S. Peteranetz, A. E. Flanigan, and D. F. Shell, "Computational creativity exercises: an avenue for promoting learning in computer science," *IEEE Transactions on Education*, vol. 60, no. 4, pp. 1-9, 2017.
- [3] A. Ghose and S. Yang, "An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets," *Management Science*, vol. 55, 2009.
- [4] A. Broder, M. Fontoura, and V. Josifovski, "A semantic approach to contextual advertising," in *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 559-566, ACM, Amsterdam, Netherland, July 2007.
- [5] "Display Advertising," *International Journal of Gynecology & Obstetrics*, vol. 83, no. 1, 2003.
- [6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," vol. 1, pp. 886-893, in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, IEEE, San Diego, CA, USA, June 2005.
- [7] D. Lowe, "Distinctive image features from scale-invariant key points," *International Journal of Computer Vision*, vol. 20, pp. 91-110, 2003.
- [8] K. Rao, H. Sak, and R. Prabhavalkar, "Exploring architectures, data and units for streaming end-to-end Speech\n, recognition with RNN-transducer," in *Proceedings of the 2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, IEEE, Mountain View, CA, USA, December 2018.
- [9] J. Si, H. Zhang, and C. G. Li, "Dual attention matching network for context-aware feature sequence-based person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5363-5372, Salt Lake City, UT, USA, June 2018.
- [10] Z. Bao and C. Wang, "A multi-agent knowledge integration process for enterprise management innovation from the perspective of neural network," *Information Processing & Management*, vol. 59, no. 2, Article ID 102873, 2022.
- [11] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, no. 2, 2012.
- [12] J. Azimi, R. Zhang, and Y. Zhou, "The impact of visual appearance on user response in online display advertising," in *Proceedings of the 21st International Conference on World Wide Web*, pp. 457-458, ACM, Lyon, France, April 2012.
- [13] H. Cheng, R. Zwol, and J. Azimi, "Multimedia features for click prediction of new ads in display advertising," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 777-785, ACM, Beijing, China, August 2012.
- [14] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971-987, 2002.
- [15] A. Taha, Y. T. Chen, and T. Misu, "Defense of the Triplet Loss for Visual Recognition," 2019, <https://arxiv.org/abs/1703.07737>.
- [16] L. H. Hughes, M. Schmitt, and L. Mou, "Identifying corresponding patches in SAR and optical images with a pseudo-siamese CNN," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, pp. 1-5, 2018.
- [17] K. Mo, B. Liu, and L. Xiao, "Image Feature Learning for Cold Start Problem in Display Advertising," in *Proceedings of the 24th International Conference on Artificial Intelligence*, pp. 3728-3734, Buenos Aires, Argentina, July 2015.
- [18] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic regression*, John Wiley & Sons, Hoboken, NJ, USA, 2013.
- [19] J. Chen, B. Sun, and H. Li, "Deep ctr prediction in display advertising," in *Proceedings of the 2016 ACM on Multimedia Conference*, pp. 811-820, ACM, Amsterdam, Netherland, October 2016.
- [20] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," 2015, <https://arxiv.org/abs/1502.03167>.
- [21] T. Ge, L. Zhao, and G. Zhou, "Image Matters: Jointly Train Advertising CTR Model with Image Representation of Ad and User Behavior," 2017, <https://www.arxiv-vanity.com/papers/1711.06505/>.
- [22] S. Rendle, "Factorization machines," in *Proceedings of the ICDM 2010, the 10th IEEE International Conference on Data Mining*, pp. 14-17, IEEE, Sydney, NSW, Australia, December 2010.
- [23] Y. Juan, Y. Zhuang, and W. S. Chin, "Field-aware Factorization Machines for CTR Prediction," in *Proceedings of the 10th ACM Conference on Recommender Systems*, September 2016.
- [24] M. Jahrer, A. Toscher, and J. Y. Lee, "Ensemble of collaborative filtering and featureengineered models for click through rate prediction," in *Proceedings of the KDDCup Workshop*, Washington DC, USA, 2012.
- [25] W. Zhang, T. Du, and J. Wang, "Deep Learning over Multi-Field Categorical Data," *European Conference on Information Retrieval*, Springer Cham, New York, NY, USA, 2016.
- [26] Y. Qu, C. Han, and R. Kan, "Product-based neural networks for user response prediction," in *Proceedings of the 2016 IEEE 16th International Conference on Data Mining (ICDM)*, pp. 1149-1154, IEEE, Barcelona, Spain, December 2016.
- [27] J. Xiao, H. Ye, and X. He, "Attentional Factorization Machines: Learning the Weight of Feature Interactions via Attention Networks," 2017, <https://arxiv.org/abs/1708.04617>.
- [28] K. Xu, J. Ba, and R. Kiros, "Show, Attend and Tell: Neural Image Caption Generation with Visual Attention," *Computer Science*, vol. 32, pp. 2048-2057, 2015.
- [29] H. T. Cheng, L. Koc, and J. Harmsen, "Wide & Deep Learning for Recommender Systems," in *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*, September 2016.
- [30] R. Wang, B. Fu, and G. Fu, "Deep & Cross Network for Ad Click Predictions," in *Proceedings of the ADKDD'17*, August 2017.
- [31] E. L. Lehmann and G. Casella, *Theory of point Estimation*, Wiley, Hoboken, NJ, USA, 1983.
- [32] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," 2014, <https://arxiv.org/abs/1412.6980>.
- [33] G. A. F. Seber and A. J. Lee, *Linear Regression analysis*, John Wiley & Sons, Hoboken, NJ, USA, 2012.