Hindawi
Scientific Programming
Volume 2022, Article ID 8329745, 10 pages
https://doi.org/10.1155/2022/8329745

Research Article

Research on Modern Book Packaging Design Based on Aesthetic Evaluation Based on a Deep Learning Model

Wen Wen

Journalism & Communication Department, Anhui News and Publishing Vocational College, Hefei, Anhui 230601, China

Correspondence should be addressed to Wen Wen; 2007020061@ahcbxy.edu.cn

Received 6 August 2022; Revised 30 August 2022; Accepted 2 September 2022; Published 15 September 2022

Academic Editor: Lianhui Li

Through the analysis of the application and development of deep learning in the field of book design and publishing, the article expounds on the positive impact of deep learning on book design and publishing, discusses the shortcomings of deep learning in creative ability, aesthetic ability, emotion, etc., and then discusses the design and publishing of books. The future development direction of intelligent aided design and intelligent personalized design is proposed to provide a reference for researchers in deep learning and book design and publication.

1. Introduction

Deep learning was first proposed at the Summer Symposium on Deep Learning at Dartmouth College in Hanover, the USA, in 1956 [1]. Due to the wide range of research fields, the concept of deep learning is also divided. At present, the more recognized definition in the academic circle comes from the book "Deep Learning: A Modern Approach" by Stuart Russell and Peter Norvig: "deep learning" is the research and design of "intelligent agents" [2], and "an intelligent agent refers to a system that can observe the surrounding environment and take actions to achieve the goal." Today, deep learning has been widely used in speech recognition, machine vision, data mining, etc., and the cross-border between deep learning and publishing and design has gradually emerged [3]. Content is the starting point of book publishing, and content is often summarized and compiled by writers and scholars through learning, reading, investigation, and research, and this process requires a lot of time and energy to complete. In ancient times, people recorded text and image content by handwriting, and it was not until the advent of printing that books published close to modern times were widely disseminated. In recent years, we can even use speech recognition technology to allow computers to quickly convert language into text through "dictation," which is just the tip of the iceberg for deep learning [4]. In September 2015, Tencent developed a manuscript writing robot, Dreamwriter, which can generate manuscripts in a very short time. After more than two years of development, Dreamwriter has been able to generate templates through automatic learning, which has expanded from the initial financial field to movies, cars, games, and many other fields. In addition, Xinhuanet, Yicai, and other media also put the writing robot into use, and the content is mainly based on event description and analysis data. Poems and novels that require more complex rhetoric and grammar can also be completed by robots. For example, the poems created by Xiaobing, a poetry-writing robot developed by Microsoft (Asia) Internet Engineering Institute, have been submitted to newspapers and periodicals under multiple pseudonyms, and all of them have been received to publish invitation [5].

Since the end of 2013, a senior researcher from Microsoft Research Asia and an information design expert from the Academy of Arts and Design of Tsinghua University have been working on research in the field of the automatic layout. This research combines aesthetic principles in design with computable image features to creatively propose a computable prototype of an automatic typesetting framework. The prototype optimizes a series of key issues, such as the visual weight of text and pictures, the weight of visual space, the color harmony factor in psychology, and the importance of information in visual cognition and semantic understanding. The prior knowledge of experts in the fields of text semantics, design principles, and cognitive understanding is integrated into the same multimedia computing framework, creating the research direction of automatic visual text layout design [6]. At the 2017 Yunqi Conference Shanghai Summit, Alibaba iDST algorithm experts shared research titled "Visual Design in the Era of Deep Learning," proposing that automated and controllable visual content can be generated through deep learning. The process of intelligent design includes the spatial layout of design elements, color matching, background adaptation, font synthesis, style recommendation, intelligent interaction, etc. Among them, "automation" can automatically adapt to various sizes, automatically learn various styles, and automatically adapt to the number of elements; "controllable" means predictable and modifiable results. The introduction of deep learning into other fields seems unstoppable. In fact, most of the time, deep learning cannot be presented as an independent subject, and deep learning is more about transforming the process and mode of work in other fields [7].

Jane Hazus, the chief economist at Goldman Sachs Group, said: "In general, AI seems more likely to capture more valuable things in statistics than the last wave of innovation, and deep learning can reduce costs, reduce labor input for high value-added production types." China's modern book publishing industry is still dominated by paper media, and a series of work in publishing activities from topic selection, drafting, editing, and reviewing, to design publishing, and distribution is required. High value-added artificial brain power is to participate in the completion. Deep learning can improve the efficiency of book design and publication in the following three aspects [8].

The first is intelligent topic selection. The terminal of book publishing and distribution must be readers. The pain point of the traditional publishing topic selection process is that the distance between publishers and readers is too far. Book publishing forms a one-way process, and readers can only passively accept it. Or the publishing house obtains the data required by readers through traditional research and consultation methods and then selects a topic for publication and distribution, which must go through a long period of time. For some books whose market demand changes rapidly traditional data acquisition, the way will bring hysteresis [9]. It can be seen that data will be the source of future productivity. Based on deep learning and big data, readers' purchase and reading behaviors can be effectively recorded, the needs of the audience can be outlined through data, and the direction of topic selection will be more accurate.

The second is smart editing. The traditional editing work mainly includes drafting, revision, and proofreading. It should be said that each part requires a lot of time, and the error rate that occurs manually cannot be ruled out. Deep learning can be said to be the core technology of the current deep learning development. After the speech recognition and natural language processing technology has developed to a certain level, the language and grammar in the work can be corrected, and then, it can undertake a lot of tedious and time-consuming work. These problems can be solved by deep learning technology based on massive data support and with deep learning capabilities. However, modern editors can shift the focus of their work to the core values of judgment and decision making that cannot be completed by deep learning for the time being [10].

The third is intelligent design [11]. As a very important work in book publishing, book design affects the sales of books to a certain extent, and people are not satisfied with absorbing text content when reading. The graphic arrangement part of book design is the part of the largest workload after the overall creative design positioning of the book is completed in the early stage, and now, it is mainly completed by designers using arrangement software. According to the application research content mentioned above, the designer can analyze the reading audience according to the deep learning, propose a highly targeted book planning and design plan, and use the automatic layout software to arrange and adjust the content, effectively improving the work efficiency, and reduce the work intensity of design practitioners.

2. Knowledge about Deep Learning

2.1. Basics of Convolutional Neural Networks. Convolutional neural network (CNN) is a type of artificial neural network. It draws on the sparse response characteristics of biological neural networks and replaces the original fully connected layer with local connections to avoid overfitting in the training process due to too many model parameter problems [12]. The weight-sharing network structure of a convolutional neural network significantly reduces the complexity of the network, reduces the number of weights, and reduces the demand for training data. It is a research hotspot in the fields of speech analysis, image recognition, and target detection [13].

The general working principle of the convolutional neural network is as follows: first, the entire image is input into the convolutional neural network, and the network starts from the bottom pixel to learn the filters. These filters are used to extract the local edge and texture features of the image and then the middle layer filters. We learn the feature map processed by the upper-level edge filter and then extract the features that can describe different types of targets, and then learn those global features that describe the entire target by the high-level filter, and finally realize the target in the image through the nonlinear fitting of the activation function [14]. In the whole image recognition process, the network automatically learns the parameters of various types of filters from the image data. Its rich feature expression ability realizes the target recognition in the image and solves the problem of the traditional image recognition algorithm. Perform manual feature extraction and data reconstruction on image data. Figure 1 shows the development of convolutional neural networks.

The starting point of the development of convolutional neural networks is the neurocognitive machine model. LeCun, the originator of deep learning, proposed the first convolutional neural network model LeCun in 1989. Since

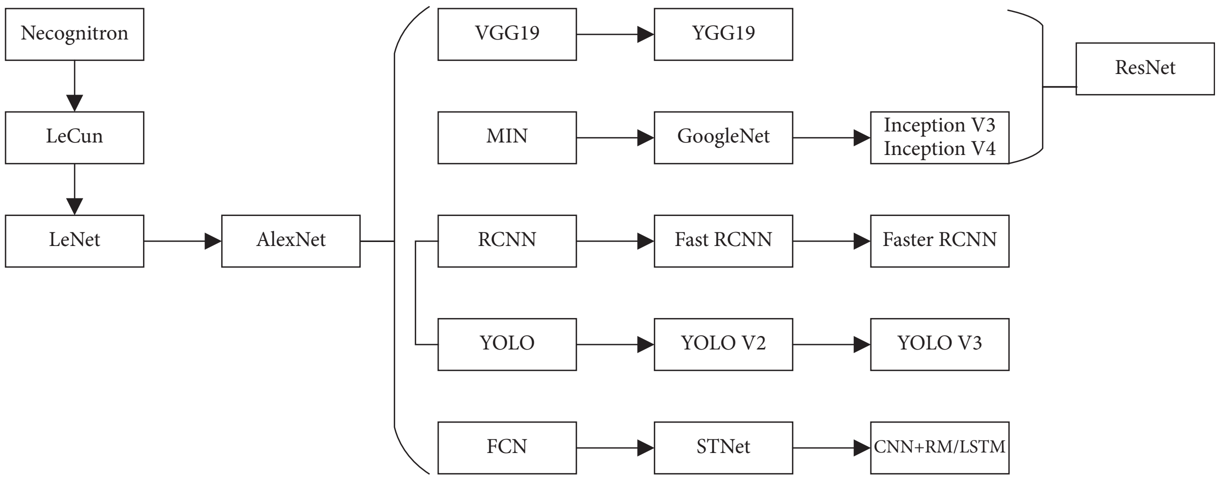

Figure 1: Development of convolutional neural networks.

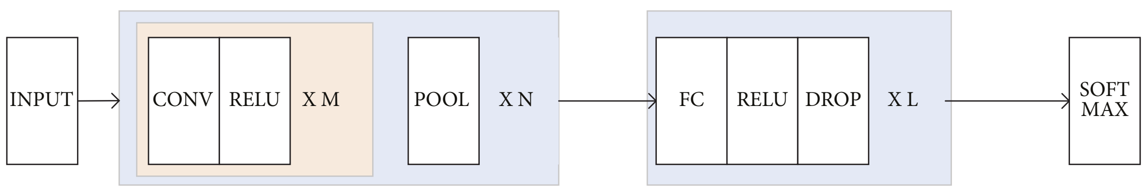

Figure 2: Classic convolutional neural network framework.

then, LeCun proposed the LeNet neural network model in 1998, but at that time, due to the superiority of hand-designed SVM and other classifiers, the convolutional neural network did not attract public attention along with the proposal of methods such as ReLU and Dropout, as well as the historical opportunities brought by GPU and big data, and the proposal of AlexNet in 2012 ushered in a historic breakthrough in convolutional neural networks. The evolution process of the convolutional neural network after AlexNet mainly consists of four directions: increasing the number of network layers and deepening the depth; enhancing the function of the convolutional layer from classification tasks to detection tasks; and adding new functional modules [15].

2.2. Convolutional Neural Network Architecture. A classic convolutional neural network for image classification mainly consists of five parts: input layer, convolutional layer, pooling layer, fully connected layer, and softmax layer. Usually, the network has only one input layer and one softmax output layer, and the convolution layer in the network can appear multiple times. The pooling layer is often located between the convolutional layers for data dimensionality reduction. The convolutional layer and the pooling layer are often connected in several adjacent convolution and pooling layers like this, which constitute a feature extraction layer of the network [16]. After multiple feature extraction and compression, the image data are input to the fully connected layer, and finally, the probabilistic classification result is output through the softmax layer. Figure 2 shows a classic convolutional neural network structure, in which M, N, and L are integers greater than zero to indicate the number of repetitions of the unit where it is located.

2.2.1. Convolutional Layer. The convolutional layer is the core part of the convolutional neural network. Its function is to extract the features of the input image. The convolutional layer completes the feature extraction through the convolution kernel. Each convolution kernel contains parameters such as size, stride, and edge padding. In shallow networks, convolution kernels extract low-level features, such as edges and corners, and in high-level networks, convolution kernels extract high-level features, such as faces, dogs, and cars [17]. Figure 3 shows the convolution operation process in the convolution layer.

In Figure 3, the input picture is a two-dimensional activation map K obtained after the convolution layer operation of the convolution kernel K, and its size is obtained by the following formulas:

$$h_{out} = \left(\frac{h_{in} - h_f + 2p}{s_h}\right) + 1, \tag{1}$$

$$w_{out} = \left(\frac{w_{in} - w_f + 2p}{s_w}\right) + 1, \tag{2}$$

$$d_{out} = k. \tag{3}$$

In formulas (1)–(3), h_{out} and w_{out} are the height and width of the output image,h_{in} and w_{in} are the height and width of the input image, as well as the height and width of the h_f convolution kernel, w_f is the width of the s_h

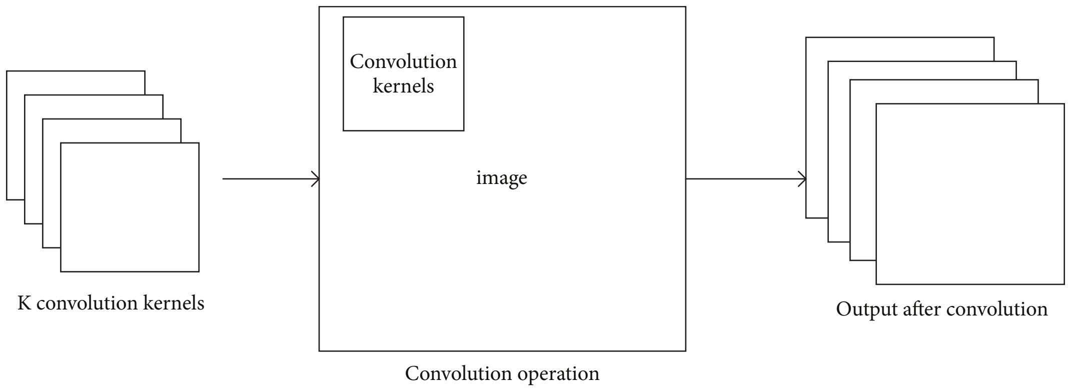

FIGURE 3: Convolution operation process in the convolution layer.

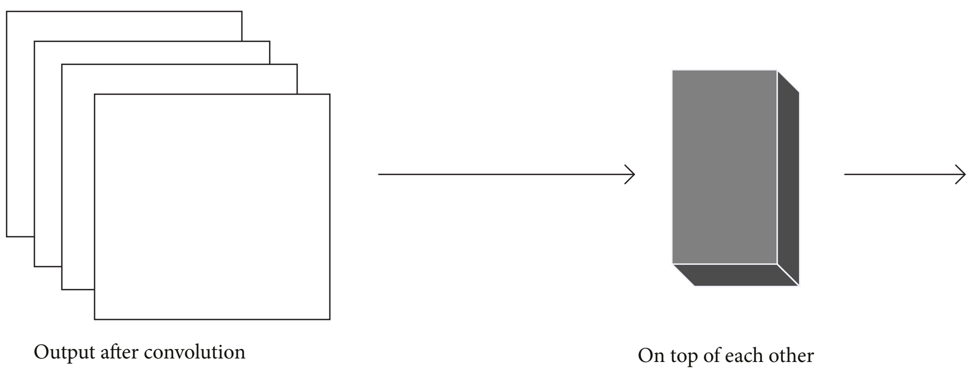

FIGURE 4: Activation maps are stacked and fed into the next convolutional layer.

convolution kernel, the sliding step size of the convolution kernel in the vertical direction and the s_w convolution kernel. The sliding step size in the horizontal direction p is the number of pixels to be supplemented by the edge, in which d_{out} is the dimension of the output.

Figure 4 shows that the K activation maps obtained after the convolution operation are stacked as the input to the next convolutional layer.

2.2.2. Pooling Layer. The function of the pooling layer is mainly to reduce the size of the feature map, and it is often used in the middle of two convolutional layers to reduce network parameters and reduce the overfitting of the model. Common types of pooling operations are max pooling and average pooling [18]. The maximum pooling operation is usually used in the middle of the convolutional network to reduce the size of the feature map, and the average pooling operation is generally used at the end of the network to replace the fully connected layer and reduce network parameters. Common pooling sizes are 2×2 and 3×3, and strides are usually 1×1 and 2×2.

If the size of the input feature map is $w_{in}\times h_{in}\times d_{in}$, the size after the pooling operation is

$$w_{out} = \left(\frac{w_{in} - w_f}{s_w}\right) + 1, \quad (4)$$

$$h_{out} = \left(\frac{h_{in} - h_f}{s_h}\right) + 1, \quad (5)$$

$$d_{out} = d_m. \quad (6)$$

In equations (4)–(6), w_{out} and h_{out} are the width and height of the output feature map, w_f and h_f are the width and height of the pooling window size, s_w and s_h are the horizontal and vertical strides, and d_{out} are the output dimensions. After the pooling operation, the size of the feature map is reduced to varying degrees according to the pooling window size and stride size [19].

2.2.3. Activation Layer. The main function of the activation layer is to introduce a nonlinear activation function, thereby increasing the nonlinearity of the network. The commonly used nonlinear activation functions are Sigmoid, Tanh, ReLU, ELU, Leaky ReLU, etc. Usually, an activation layer immediately follows each convolutional layer. For the input $w_{in}\times h_{in}\times d_{in}$ feature map, the output size after the activation layer is

$$w_{out} = w_{in}, \tag{7}$$

$$h_{out} = h_{in}, \tag{8}$$

$$d_{out} = d_{in}. \tag{9}$$

From equations (7)–(9), the activation layer usually does not change the size of the input feature map.

2.2.4. Fully Connected Layer. The convolution layer and pooling layer are mainly to complete the extraction of image features, while the fully connected layer is mainly to complete the classification task [20]. The fully connected layer is located at the end of the convolutional neural network, and the fully connected layer is usually followed by a softmax layer to calculate the final output of the network for each classification probability.

2.3. Data Augmentation. In deep learning, the training of the model often relies on a large amount of data to learn the parameters of the network, especially the deep network. However, in reality, it is often difficult to obtain enough data due to the limitation of practical conditions, and sometimes, a lot of manpower and material resources are wasted [21]. Data enhancement is to make some changes to the original data, but for the network model, it is "new" datum, thus easing the data requirements for deep learning model training. Data augmentation can improve the generalization ability of the model and improve the robustness of the model. Common data augmentation methods are flipping, rotating, scaling, cropping, translation, adding noise, etc. Compare model performance on the test set with and without data augmentation. It can be seen from the experiments that for the same model and the same dataset, when data augmentation is used, the accuracy and recall rate of the model on the test set are much better than when data augmentation is not used. It can be seen that, without adding any additional investment, the performance of the model can be significantly improved only through data augmentation operations.

2.4. Transfer Learning. Transfer learning refers to the simple adjustment of a model trained on one problem to make it suitable for a new problem. There are two common types of transfer learning. One is to use models trained on other datasets such as ImageNet datasets such as VGG, ResNet, and Inception as feature extractors, remove the final classification layer of the model, and replace it with new ones. This kind of transfer learning is suitable for the classification problem with a small dataset and is similar to the classification problem of the original model [22]; the other is fine-tuning that refers to replacing the last classification layer of the trained model with the classification layer of the new problem, initializing the classification layer, keeping the parameters of other layers unchanged, and then training the new classification layer separately, after a few rounds of iteration (warm-up) "Unfreeze" other layers to continue training, fine-tuning the parameters of the entire model. This method is suitable for a relatively large dataset of new problems and is similar to the problem of the original model. Since the first layer of the convolutional neural network generally extracts low-level features such as texture, corners, and colors, the features extracted by the closer convolutional layers are more advanced, abstract, and task-oriented, so during training, you can "Unfreeze" the later convolutional layers of the original model, keeping the initial convolutional layer parameters unchanged [23].

3. Modern Book Packaging Design Based on BBE Network Aesthetic Evaluation

3.1. Object Detection Task Overview. Image classification, object detection, and image segmentation are the three major tasks of deep learning applied to the field of image processing. Image classification means that when an input image is given, the deep learning algorithm needs to analyze and identify what all the objects in the image are, that is, the category they belong to; target detection means that when an input image is given, the deep learning algorithm needs to detect the specific positions of all objects in the image, and also be able to identify the category to which they belong; the image segmentation task is aimed at pixels in the image, which means that the deep learning algorithm needs to distinguish all pixels in the input image, that is, determine which pixels in the image belong to which targets [24].

3.2. Book Packaging Design Based on Object Detection. Object detection tasks in deep learning are classified into two-stage algorithms and one-stage algorithms.

3.2.1. Two-Stage Algorithm. The two-stage algorithm is characterized by high detection accuracy and slow detection speed. With the development of the two-stage algorithm, the tasks of each stage of target detection are integrated into a deep neural network [25]. Two-stage algorithms include RCNN, SPPNet, Fast-RCNN, Faster-RNN, and Mask-RCNN. The first two-stage algorithm was the R-CNN algorithm, followed by Fast R-CNN and Faster R-CNN gradually enabling object detection to be trained end-to-end. The principle of the two-stage algorithm is multistep. First, a large number of candidate frames are generated in the image, and then, the features of the selected regions of the candidate frames are extracted. Finally, according to the results of the feature extraction, more refined classification and localization operations are performed in the high-level network. It is precise because of the step-by-step detection process of the two-stage algorithm that the detection accuracy of the algorithm is high, and the detection speed is relatively slow.

The Faster R-CNN model is shown in Figure 5, and the algorithm completes the target detection step by step. First, we complete the work of extracting features from the input image. Then, the extracted feature maps are filtered through the region generation network. Finally, using the filtered feature map, the

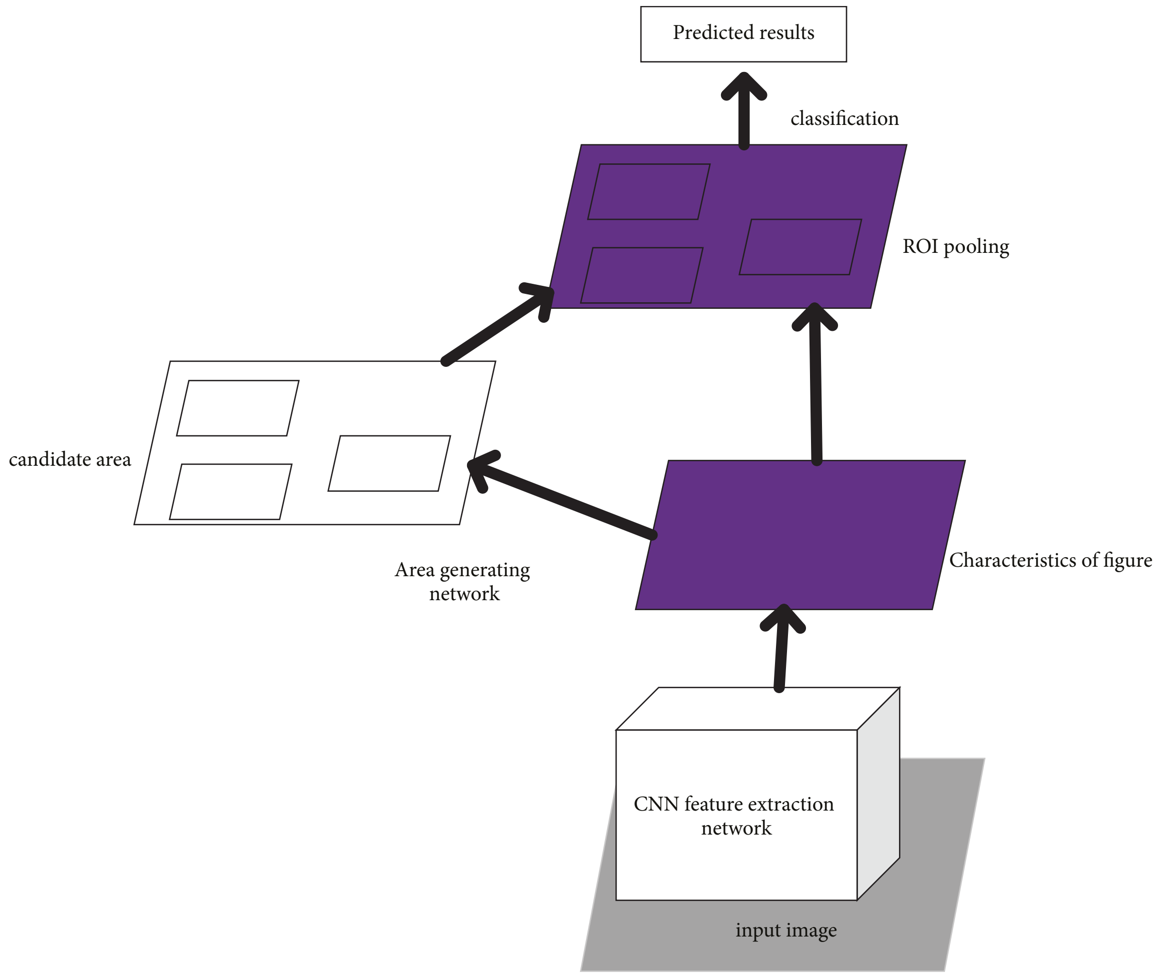

Figure 5: Faster R-CNN network model.

classification and localization of the target object are completed. Faster R-CNN has a unique area generation network, which enables the algorithm to perform more refined detection and identification. At the same time, because of the more complex network structure, the detection speed and training speed of the algorithm are relatively slow, and the computing performance of the hardware device is required high, and it is generally difficult to achieve real-time detection [26].

3.2.2. One-Stage Algorithm. The classic one-stage algorithm can be mainly divided into two series, namely, YOLO and SSD. The one-stage algorithm integrates feature extraction, classification, and regression into a deep learning network framework, and the detection speed is very fast.

YOLO is a representative of a one-stage target detection algorithm. The principle of the algorithm is to generate a large number of a priori frames on the input image, and then directly classify and locate the area selected by the a priori frame, that is, directly output the target object in the input image. Specifically, the image is first divided into grids, and then, a large number of prediction boxes are generated for each grid, and finally, the final prediction box is obtained through operations such as nonmaximum suppression and threshold analysis. Figure 6 shows the model architecture of the YOLO network, in which the input image is divided into 7×7 grids, and each grid has 30 data, including the coordinate offset of 2 bounding boxes, the target confidence, and the probability over a class. The YOLO algorithm does not perform well in detecting objects that are close to each other and small objects, because each grid of the algorithm only predicts two bounding boxes that belong to only one class, and when target objects with different aspect ratios appear in the image, the generalization ability of the algorithm is weak.

The difference between the SSD algorithm and YOLO is that the SSD algorithm directly uses CNN for detection. The SSD network has the following characteristics. First, SSD uses a multiscale strategy to detect target objects of different sizes in the image by extracting feature maps of different scales in the image, which helps to improve the detection accuracy of the network. Second, in order to simulate the size of different target objects in the image, the SSD network uses a priori frames with different aspect ratios and different sizes, which helps to improve the detection effect of the network on small target objects. Positioning will be more accurate. In general, the SSD algorithm not only takes the advantages of the YOLO algorithm but also draws on the candidate region generation network of the two-stage

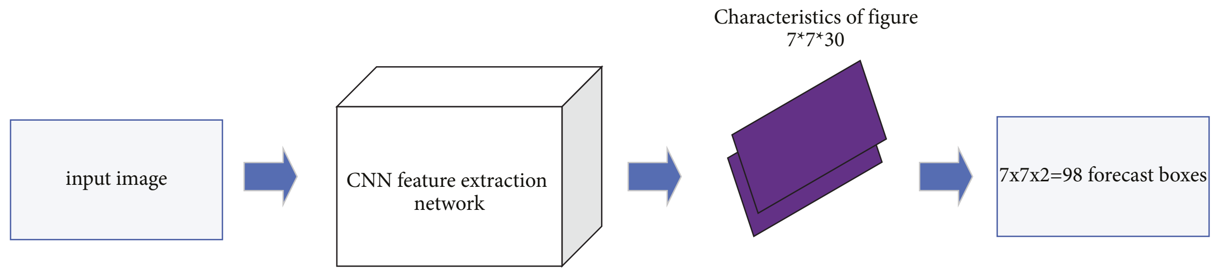

Figure 6: YOLO network model.

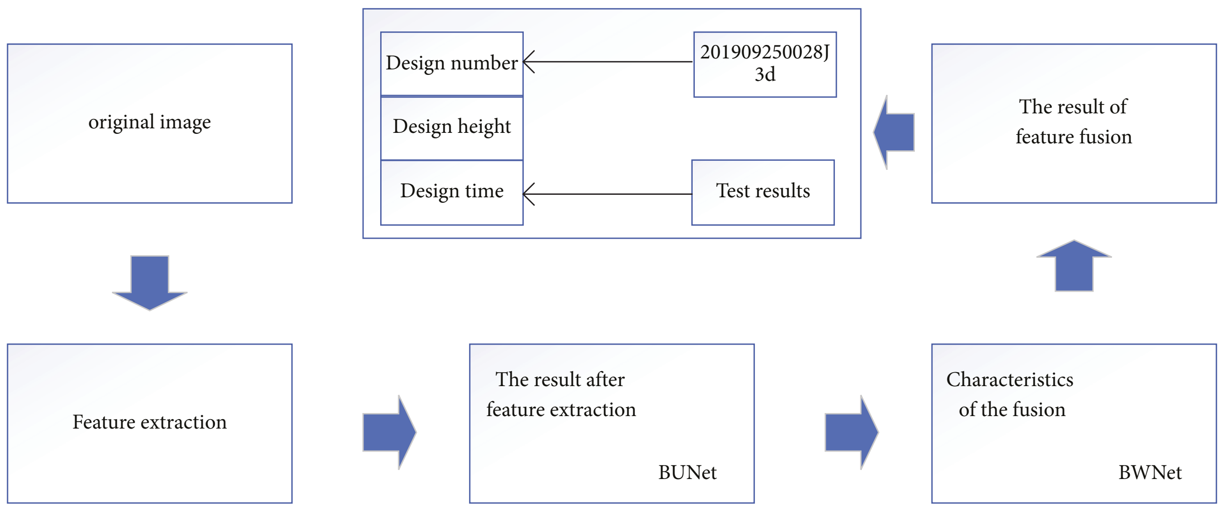

Figure 7: Overall design.

algorithm to achieve the speed of the one-stage algorithm and the accuracy of the two-stage algorithm, which is a relatively balanced accuracy and speed algorithm [27].

3.3. Overall Scheme and Algorithm Framework Design. The detection method of target detection can realize the end-to-end detection of inkjet character defects. However, due to the diversification of book design types, a large number of training sample datasets are required, and the detection accuracy is also easily interfered with by various factors. It is difficult to meet the requirements of industrial applications. The idea of the detection algorithm in this paper is to first obtain the category and position of each book through target detection. The overall design scheme is shown in Figure 7.

The BBE target detection network based on convolutional neural network has powerful feature extraction ability and has high detection accuracy and speed, which is very suitable for practical industrial detection. First, the collected images are input into the feature extraction network BUNet designed based on the EfficientNet core module for processing, and then, the extracted effective features are input into the designed feature fusion network BWNet for feature refinement and abstraction, and then, the fused features are classified and positioned through the classification network and the regression network, respectively. Finally, the quality inspection standards (number of code, height, and time) are set. We compare them to obtain the final overall detection result.

The structural framework of the BBE network, the feature extraction network, the feature fusion network, and the classification and regression network constitute the main frame structure of the algorithm network. The feature extraction network BUNet achieves good recognition with a small amount of parameters. The backbone of the feature extraction network is a general convolutional layer, which is used to perform convolution processing on the input image, and then continuously extract the depth features of the inkjet characters by connecting 7 basic modules with a total of 23 basic units (basic unit). It is based on the improvement of the core structure of EfficientNet. Based on the feature map pyramid network (FPN), the feature fusion network BWNet adds many connections to the network, fuses the feature maps of multiple intermediate layers, and continuously performs up and down weighted sampling and fusion features. It is a fast normalized multiscale weighting feature map pyramid network. The classification and regression networks are separate, the classification network classifies the target in the prior box, and the regression network adjusts the size and position of the prior box until the final prediction box is obtained. Finally, the redundant prediction frame is removed by the operation of nonmaximum suppression, and the category and position information of the inkjet characters are obtained.t

3.4. Feature Extraction Network Design

3.4.1. EfficientNet Algorithm. EfficientNet network was proposed by Google in 2019. On the basis of other networks, it improves the detection accuracy while greatly reducing the network parameter calculation, and has high detection accuracy and speed. To improve network performance, the following points should be paid attention to: (1) first, the network must be able to converge and be able to be trained; (2) the amount of parameters of the network should be minimized to ensure the high precision and speed of the network, and the model should be easy to train; and (3) improve the network structure, enabling it to learn useful deep features. The EfficientNet network does the above points well, and the network model uses a small amount of parameters to obtain good accuracy.

The construction of CNN often has the following characteristics: (1) increase the depth through the residual structure to improve the expressive ability of the network; (2) increase the number of feature layers extracted by each layer of the network, realize the extraction of multiscale features, and improve the width of the network and learn more features; and (3) increase the resolution of the input image, enrich the feature information that the network can learn, and improve the accuracy of the network. The EfficientNet network combines the above characteristics and adjusts the depth, height, and resolution of the input image to obtain a series of lightweight networks with balanced speed and accuracy.

3.4.2. Feature Extraction Network: BUNet. The feature extraction network in this paper is based on the core module of EfficientNet, which is designed, and the algorithm is based on a modular design, and the feature extraction capability of the network can be changed according to the needs of the task. The backbone of the network is a stack of 7 basic modules (23 basic units) with powerful feature extraction capabilities. At the end of the network are two max pooling layers of size 3×3, which are mainly used to downsample the extracted feature maps to make the obtained features more refined. Finally, 5 effective feature layers are selected from the network as output features. Among them, (BUConvk, $3 \times 3) \times n$ is the basic unit based on the inverted bottleneck structure, BU is the basic unit, k is the number of convolution kernels, n is the number of basic units, and the size of the convolution kernel is 3×3. Since the feature extraction network in this paper is continuously stacked by the same basic unit (basic unit), it is named BUNet.

3.5. Basic Unit. The basic unit of this paper is an inverted bottleneck structure as a whole, and many optimization strategies are added. After each calculation by a downsampling module, the resolution of the feature map is reduced to 1/2 of the original. First, the input channel is subjected to 1×1 convolution, BatchNorm normalization, and Swish activation operations, and then calculated by an improved depthwise separable convolution (xDepthwise Conv2d), which also performs normalization and activation operations, and then increases. An attention mechanism on channels is finally reduced by 1×1 convolution and normalized, and then connected to the large residual edge. The two most important operations in the base unit are the depthwise separable convolution and the inversion bottleneck structure.

3.5.1. Depthwise Separable Convolution. Depthwise separable convolution is composed of depthwise convolution and point-by-point convolution. Different from ordinary convolution operations, depthwise separable convolution reduces a lot of convolution calculations, which is widely used in most lightweight detection networks. For a three-channel input image, a depthwise convolution kernel is used to process one of the channels, and then, the number of channels of the output feature map is adjusted by point-by-point convolution. Point-by-point convolution is actually a 1×1 convolution, which can be used to adjust the number of output channels of the network, which plays a role in feature fusion to a certain extent and ensures the information exchange between each channel in the input feature map.

3.5.2. Inverted Bottleneck Structure. Residual structure and bottleneck structure are proposed in the ResNet network, which can solve the problem of gradient disappearance and gradient explosion caused by the deepening of the network, thus solving the problem that deeper networks are difficult to train. The residual structure has a bypass branch to connect the input directly to the output, so that the subsequent network layers can directly learn the residual between the input and the nonlinear convolution output, which not only protects the integrity of the input information but also simplifies the network. *Learning Goals and Difficulty*. The bottleneck structure first uses a 1×1 convolution to reduce the number of input channels. After the convolution calculation is completed, a 1×1 convolution is used to restore the number of output channels. This structure can not only improve the expressiveness of the network but also the computational complexity of the entire network can be reduced.

The inverted bottleneck structure first uses a 1×1 convolution to increase the dimension of the input channel. After the convolution calculation is completed, a 1×1 convolution is used to restore the number of output channels; that is, channel expansion is performed first, and then, channel compression is performed. The inverted bottleneck structure can learn more deep features in the middle convolutional layer of the network by first expanding the input feature map, and finally summarize and filter out useful features; that is, the inverted bottleneck structure can learn more about the input channel. Useful features have stronger feature extraction capabilities. The depthwise separable convolution (DWConv) is also applied in the inverted bottleneck structure. The operation of dilating and then compressing the feature channel does not increase the amount of computation, and the memory efficiency of the inverted design is much higher. The experimental effect is also better.

3.6. Network Optimization. The optimization strategy in the basic unit is elaborated below.

3.6.1. Improved Depthwise Separable Convolution. This section improves the depthwise separable convolution. Specifically, the number of input feature channels is adjusted by point-by-point convolution, and then, the feature extraction operation is completed by depthwise convolution. Among them, the depth convolution of 3×3 is split into two asymmetric convolutions of 1×3 and 3×1, which will not affect the function of the convolution, but can greatly reduce the calculation amount of the convolution. This can not only speed up the calculation speed of the network but also deepen the depth of the network and improve the nonlinear expression ability of the network.

3.6.2. Channel Attention Mechanism. After completing the convolution calculation for feature extraction, an attention mechanism is applied to the feature channels, which allows the network to learn and pay more attention to the channel where the effective features are located. First, perform global average pooling on the features extracted by the convolution operation, and reshape the feature channel into a dimension that can be convolved, then compress and expand the feature channel through 1×1 convolution, and use the Sigmoid function to obtain a ratio between 0 and 1. The probability value between them is the attention level of the channel. Finally, different feature channels are multiplied by their attention levels to obtain different levels of depth features.

3.6.3. Linear Activation Dimensionality Reduction and Element-Level Reduction Operations. The nonlinear activation function in the network can enhance the nonlinear expression ability of the network, and at the same time, it will also cause the model to lose part of the feature information, and the downsampling operation itself will also discard part of the feature information. If the feature channels are downsampled and processed with nonlinear activation functions at the same time, the expressive ability of the network will be reduced and the performance of the model will be reduced. In this paper, when 1×1 convolution is used to reduce the number of feature channels, the Swish function is not used as the activation function, and it is directly activated linearly to retain more feature information, ensure the expressive ability of the network, and thus ensure the performance of the model. The operation of using linear activation to retain more feature information is also mentioned in the MobileNet-v2 network model.

Element-level operations will not bring too much extra computation, but too many element-level operations will increase the memory consumption of the computer, reduce the calculation speed, and affect the performance of the model. Finally, the output channel after dimensionality reduction and the large residual edge (input) are stitched and fused by point convolution.

4. Conclusion

With the advancement of science and technology, digital technology has developed rapidly, and technological innovation has made people's cognition of the world more intuitive and vivid. The original perceptual cognition method is reorganized with a rigorous mathematical model, and various design elements are presented rationally and digitally. This new expression method will definitely bring a new design thinking and artistic presentation. The evolution of art design is rooted in society, and so it is the rise of digital media art. With the continuous development of digital art, the form and function of art design are constantly enriched and improved. As a discipline, we should also cater to the diversified development trend of design, and build a new theoretical system and thinking mode for the cultivation of artistic talents in line with social development. We should not only pay attention to the inheritance of traditional culture and the learning of advanced design theory but also keep up with the pace of the times, quickly master digital technology, continuously expand creative thinking, and comprehensively improve our operational capabilities. Only in this way can a scientific and sustainable design teaching system be built.

Data Availability

The dataset that supports the findings of the study can be accessed upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

[1] P. Aditya, M. Doron, and C. Caicedo Juan, "Image-based cell phenotyping with deep learning," *Current Opinion in Chemical Biology*, vol. 65, 2021.

[2] Y. Ding, J. Chen, and J. Shen, "Prediction of spectral accelerations of aftershock ground motion with deep learning method," *Soil Dynamics and Earthquake Engineering*, vol. 150, 2021.

[3] M. W. Wang, Y. Y. Zhang, and J. Xu, "A deep learning method based on multi-modality EEG for automatic depression screening," *International Journal of Psychophysiology*, vol. 168, no. S, pp. S205–S206, 2021.

[4] C. Wang, C. Wang, W. Li, and H. Wang, "A brief survey on RGB-D semantic segmentation using deep learning," *Displays*, vol. 70, Article ID 102080, 2021.

[5] M. Li and L. Jiang, "Deep learning nonlinear multiscale dynamic problems using Koopman operator," *Journal of Computational Physics*, vol. 446, Article ID 110660, 2021.

[6] O. F. Ertuğrul and M. F. Akıl, "Detecting hemorrhage types and bounding box of hemorrhage by deep learning," *Biomedical Signal Processing and Control*, vol. 71, 2022.

[7] J. He, J. Zhou, J. Dong, Z. Su, and H. Lu, "Revealing the effects of microwell sizes on the crystal growth kinetics of active pharmaceutical ingredients by deep learning," *Chemical Engineering Journal*, vol. 428, Article ID 131986, 2022.

[8] D. J Manuel and O. Lukumon, "Deep learning with small datasets: using autoencoders to address limited datasets in

construction management," *Applied Soft Computing Journal*, vol. 112, Article ID 107836, 2021.

[9] G. Lu, Lu Pan, and Y. Ren, "A Deep Learning Approach for Imbalanced Crash Data in Predicting Highway-Rail Grade Crossings accidents," *Reliability Engineering and System Safety*, vol. 216, Article ID 108019, 2021.

[10] H. Choi, C. Y. Um, K. Kang, H. Kim, and T. Kim, "Application of Vision-Based Occupancy Counting Method Using Deep Learning and Performance analysis," *Energy & Buildings*, vol. 252, Article ID 111389, 2021.

[11] N. M. Esrafilian and H. Fariborz, "Occupancy-based HVAC Control Using Deep Learning Algorithms for Estimating Online Preconditioning Time in Residential buildings," *Energy & Buildings*, vol. 252, Article ID 111377, 2021.

[12] K. Dutta Subrat, S. Sudarshan, B. Abhilasa et al., "Study on enhanced deep learning approaches for value-added identification and segmentation of striation marks in bullets for precise firearm classification," *Applied Soft Computing Journal*, vol. 112, Article ID 107789, 2021.

[13] N. Wang, H. Xu, F. Xu, and L. Cheng, "The algorithmic composition for music copyright protection under deep learning and blockchain," *Applied Soft Computing Journal*, vol. 112, Article ID 107763, 2021.

[14] D. Tian, Q. Ying, X. Jia, R. Ma, C. Hu, and W. Liu, "MDCHD: A Novel Malware Detection Method in Cloud Using Hardware Trace and Deep learning," *Computer Networks*, vol. 198, Article ID 108394, 2021.

[15] R. Xie and A. G. Dempster, "An on-line deep learning framework for low-thrust trajectory optimisation," *Aerospace Science and Technology*, vol. 118, Article ID 107002, 2021.

[16] W. Chao, J. Zhang, and X. Yuan, "DeepTIS: improved translation initiation site prediction in genomic sequence via a two-stage deep learning model," *Digital Signal Processing*, vol. 117, Article ID 103202, 2021.

[17] I. Hussain, S. Tan, B. Li, X. Qin, D. Hussain, and J. Huang, "A novel deep learning framework for double JPEG compression detection of small size blocks," *Journal of Visual Communication and Image Representation*, vol. 80, Article ID 103269, 2021.

[18] R. Liu, R. Yassine, and El Bagdouri Mohammed, "Multispectral background subtraction with deep learning," *Journal of Visual Communication and Image Representation*, vol. 80, Article ID 103267, 2021.

[19] N. Gentner, C. Mattia, K. Andreas, S. Gian Antonio, and Y. Yao, "DBAM: making Virtual Metrology/Soft sensing with time series data scalable through Deep Learning," *Control Engineering Practice*, vol. 116, Article ID 104914, 2021.

[20] J. Guo, Q. Wang, and Y. Li, "Evaluation-oriented façade defects detection using rule-based deep learning method," *Automation in Construction*, vol. 131, Article ID 103910, 2021.

[21] A. Akshay and F. You, "Quantum computing based hybrid deep learning for fault diagnosis in electrical power systems," *Applied Energy*, vol. 303, Article ID 117628, 2021.

[22] L. Gao, T. Liu, T. Cao, Y. Hwang, and R. Reinhard, "Comparing deep learning models for multi energy vectors prediction on multiple types of building," *Applied Energy*, vol. 301, 2021.

[23] X. Li, Y. Yang, Y. Ye, S. Ma, and T. Hu, "An online visual measurement method for workpiece dimension based on deep learning," *Measurement*, vol. 185, Article ID 110032, 2021.

[24] W. Chen, A. N. Wu, and F. Biljecki, "Classification of urban morphology with deep learning: application on urban vitality," *Computers, Environment and Urban Systems*, vol. 90, Article ID 101706, 2021.

[25] S. A. Azimi, A. Hossein, J. S. Chai, K. Martin, G. Abdelhakim, and R. Hofman, "Classification of radioxenon spectra with deep learning algorithm," *Journal of Environmental Radioactivity*, vol. 237, Article ID 106718, 2021.

[26] C. Neha, G. Gupta, V. Vallurupalli, and B. Indranil, "On the platform but will they buy? Predicting customers' purchase behavior using deep learning[J]," *Decision Support Systems*, p. 149, 2021.

[27] M. Huang, B. Zhang, W. Lou, and K. Ahsan, "A deep learning augmented vision-based method for measuring dynamic displacements of structures in harsh environments," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 217, Article ID 104758, 2021.