ISSN: 2723-9535

Available online at www.HighTechJournal.org

HighTech and Innovation Journal

Vol. 5, No. 1, March, 2024

Automatic Recognition Technology of Library Books Based on Convolutional Neural Network Model

Jianping Hu [1], Yongkang Yan [1], Zhengguang Xie [2]*

[1] *Nantong University Library, Nantong, Jiangsu, China.*

[2] *School of Information Science and Technology, Nantong University, Nantong, 226019, Jiangsu, China.*

Received 08 November 2023; Revised 13 February 2024; Accepted 21 February 2024; Published 01 March 2024

Abstract

Background: The development of technological devices has changed many facets of our lives, particularly the way we engage with information and learning. The advent of automated technology for identification has had a had a revolutionary effect on how we read and organize books within the context of books and data searches. It starts by solving the difficulties in analyzing photos of book pages by using methods such as distortion rectification and book separation. *Objective:* The research compares the effectiveness of the suggested method with traditional straight-line identification techniques using real-world testing. *Methodology:* The Skip-Gram model in Word2Vec is used to accurately represent spoken language, allowing word vectors to be generated and input data to be preprocessed for CNN. The results show that the methodology created regarding the present investigation works better than alternatives concerning accuracy and efficiency during line identification. *Result:* This work advances the field of book suggestion systems by presenting a strong and effective method that leverages CNNs. The findings demonstrate deep learning techniques may be used to optimize system recommendations and improve customer service and happiness in a variety of contexts. This technique creates a bridge between natural language processing and picture evaluation and opens up new possibilities for suggestion advancement along with user satisfaction.

Keywords: Automatic Book Recognition; Convolutional Neural Network; Image Correction; Image Retrieval.

1. Introduction

The library utilizes more and more innovative tools throughout the age of technology to improve customer service and expedite processes. The automated recognizing technique (ART) of books from libraries represents these devices, which have been increasing in popularity. The computerized means of recognizing and cataloging library objects by a variety of methods, including Optics Characters Recognizing (OCR), Radio Frequency Identification (RFID), barcode reading, and algorithms used for machine learning, is known as ART. Adopting ART in libraries has several advantages, such as greater availability, accuracy, and productivity of services provided by libraries. Libraries can speed up the handling of library resources, minimize mistakes, and decrease manual labor by automating the recognition and cataloging of volumes. This enables libraries to concentrate more on offering useful solutions for users, such as neighborhood outreach along with academic support [1, 2].

Additionally, ART makes the materials in libraries easier to find by giving consumers simpler, quicker ways of getting data. Automatic book recognition technologies make it easier for users to find and check out a resource, which improves their time in libraries generally. ART also makes it possible for libraries to keep more precise records of their

* Corresponding author: xiezg@ntu.edu.cn

http://dx.doi.org/10.28991/HIJ-2024-05-01-015

collections, which lowers the number of abandoned or missing components and enhances the management of collections. Moreover, ART makes it easier to integrate libraries using other types of digital media, such as smartphones and internet-based catalogs. The scope and effect of library offerings are increased by this effortless integration, which enables readers to browse for, reserve, and utilize materials at libraries at any time or from any location [3]. A significant breakthrough in contemporary library administration is the Automated Identification Technologies of Public Literature, which gives institutions the chance to streamline activities, enhance customer service, and more. An important development in contemporary library administration is the Automatic Recognition Technology of Library Textbooks, which gives institutions the chance to enhance customer service, streamline activities, and adjust to the changing demands of customers throughout the world of digital media [4].

The objective of the study is to analyze the viability and efficacy of automating the processing of books from libraries in contemporary library environments. Assess the degree to which library cataloging and book authentication processes may be made more accurate and efficient by using computerized identification technologies. Figure 1 depicts the flow of the suggested structure.

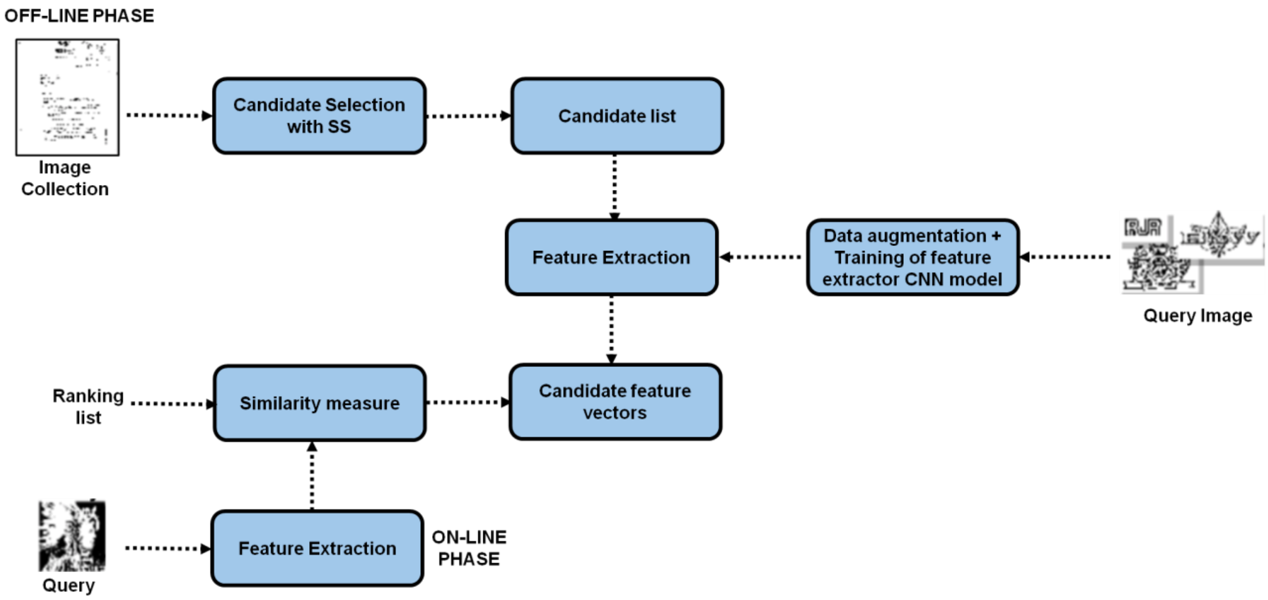

Figure 1. Framework of image retrieval page retrieval method based on convolutional neural network

The study by Anumula et al. [5] suggested that libraries increase the use of artificial intelligence (AI), which makes systems capable of activities that formerly required intellect from humans. In the end, creating computers with an intellect comparable to that of humans will influence the library profession. The research by Bairagi et al. [6] incorporates robotic intelligence; the library may transcend its material constraints and become intelligent and approachable. Libraries have to make hurried adjustments across their locations due to the present batch of AI technologies. The research by Mupaikwa [7] demonstrated that machine learning as well as intelligent technology are being applied in libraries to assist with a variety of offerings, including collection administration, retrieval of data, structuring and separating, recording, and librarianship. The study by González- Alcaide et al. [8] found that academic libraries' many technological platforms enhance their growth and efficiency and situate them within a larger framework in commercial. The research by Takher [9] explores how the use of near-field communication may assist India's agricultural sector and library to overcome their current problems. It concentrates especially on problems like scarce resources, antiquated facilities, and restricted availability.

Even though Automated Recognising Technologies (ART) with books from libraries have advanced, there are still a number of unanswered questions requiring additional study. Another is maximizing identification efficiency and precision within an array of library locations and groups, as present technologies might differ based on volume forms, illnesses, and available information, among many others. Others are investigating how ART affects customer interactions and procedures in libraries, evaluating its affordability, scaling, and accessibility, and looking at feedback and attitudes. Furthermore, since ART devices can analyze private data, there seems to be a dearth of studies related to the moral and security consequences of ART in library settings. Work ought to be directed towards creating accountable and open frameworks and resolving issues with permission, as well as data management, in order to guarantee the ethical implementation and uptake of ART through libraries.

2. Book Page Retrieval Method and Steps

As shown in Figure 1, the book page retrieval method based on a convolutional neural network consists of four parts: feature extraction in CNN, image segmentation, feature matching, image correction as well as retrieval. When offline, we input all standard images of candidate book pages saved in the database into the CNN and save them. During online retrieval, we first input the image of the book page to be inspected into the image segmentation module to remove the background area. Image distortion is corrected by an image correction module on an image-by-image basis [10]. Then we feed the corrected quasi-standard image into the feature extraction module. We extract feature codes through a convolutional neural network. Finally, we input the feature code of the image to be inspected into the feature matching module. We match it against standard image signatures saved in the database one by one. Finally, we select the top k standard images with the highest matching similarity as the retrieval result.

2.1. Image Segmentation

Because of the background interference and image distortion in the image of the book page to be inspected, it is still difficult to obtain the ideal retrieval accuracy by directly using the convolutional neural network to extract features. We propose a fully automatic, fast image segmentation algorithm to segment the book region. The purpose is to reduce the influence of background and image distortion on retrieval accuracy.

Existing interactive image segmentation algorithms usually require the user to enclose the target area with a rectangular frame. On the one hand, if the target bounding box specified by the user is not ideal, the image segmentation algorithm may give poor segmentation results. On the other hand, the interaction step will also reduce the user experience of the entire book page retrieval system [11]. Although some image segmentation algorithms can automatically initialize object bracketing according to visual saliency, the processing speed of such image segmentation algorithms is relatively slow. In this study, a coarse-to-fine automatic fast image segmentation algorithm is proposed to remove the background interference in the images of the book pages to be inspected. As shown in Figure 2, the proposed image segmentation method includes three steps: rough segmentation, object bounding box initialization, and fine segmentation. First, the input image is roughly segmented using a preset fixed target bounding box. Then a rectangle is fitted on the result of the rough segmentation as a new target bounding box. Finally, the input image is finely segmented again using the new target bounding box.

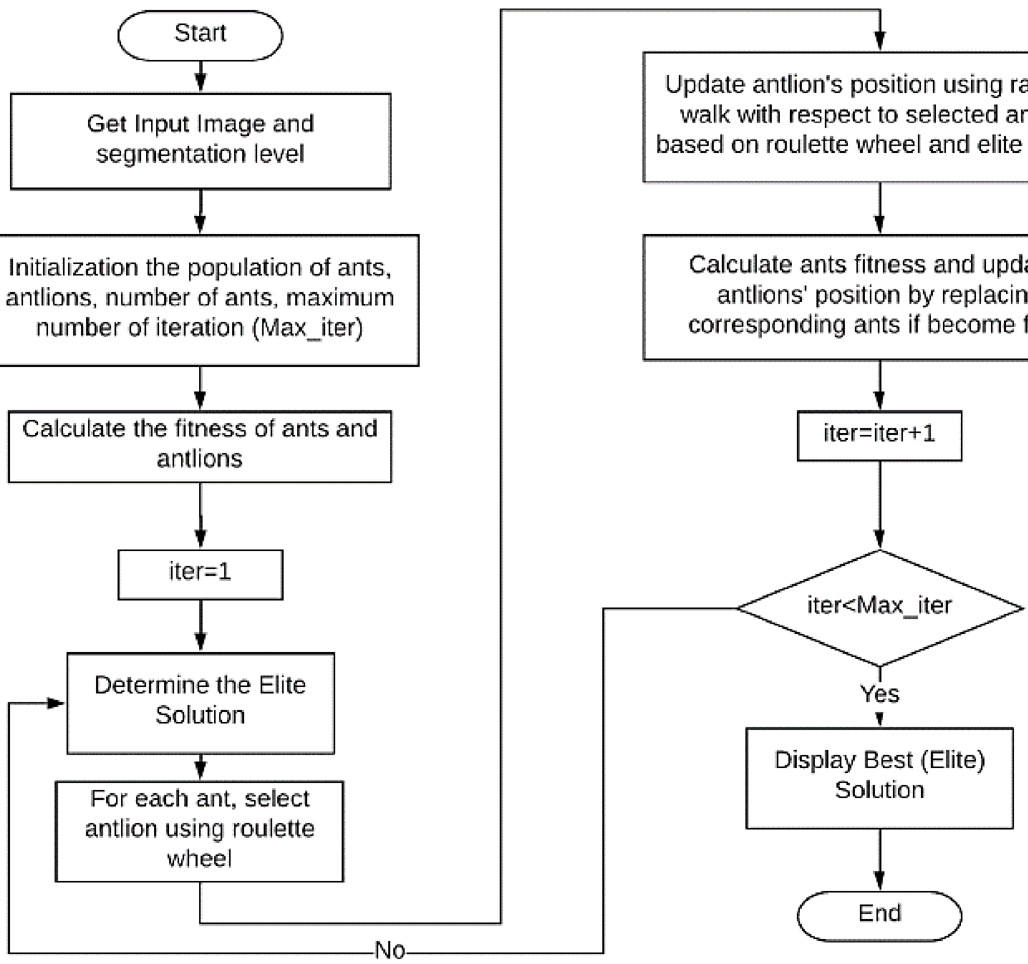

Figure 2. Automatic and fast book page image segmentation algorithm from coarse to fine

A Bayesian classifier is employed to roughly segment the images of the book pages for inspection. Let H_O denote the unnormalized color histogram counted from the bounding box O of the preset target image I of a book page, and B^{H_B} represent the unnormalized color histogram from the background region B. b_x denotes the corresponding column (bin) when the pixel at position x is projected to the histogram. Applying the Bayesian formula, the probability of pixels x is determined as shown in Equation 1.

$$p(x \in O|O, B, b_x) \approx \frac{p(b_x|x \in O)p(x \in O)}{\sum_{\Omega \in \{O,B\}} p(b_x|x \in \Omega)p(x \in \Omega)} \tag{1}$$

Where $p(b_x \mid x \in O)$ and $p(b_x \mid x \in B)$ are the likelihood probability values? It can be estimated from the color histogram:

$$p(b_x|x \in O) \approx H_O(b_x)/|O| \tag{2}$$

$$b(b_x|x \in B) \approx H_B(b_x)/|B| \tag{3}$$

Another way to estimate the previous likelihood is to use: $p(x \in O) \approx |O|/(|O| + |B|)$. Therefore, Equation 1 can be expressed in:

$$p(x \in O|O, B. b_x) \approx \frac{H_O(b_x)}{H_O(b_x) + H_B(b_x)} \tag{4}$$

If, $p(x \in O|O, B. b_x) > 0.5$ the pixel point x is classified as the target. The input image undergoes rough segmentation by a Bayesian classifier, followed by marking the connected components using the run-length encoding (RLE) connected component labeling algorithm. Subsequently, the bounding rectangle of the largest connected region is selected as the new target bounding box. Finally, the input image is finely re-segmented using the DenseCut algorithm. Because the rough segmentation method proposed in this paper has low time complexity. DenseCut is a real-time image segmentation algorithm with better segmentation results than GrabCut. Therefore, the automatic book page image segmentation algorithm has the advantages of good effect and fast speed [12]. When the initialization of the target bounding box is not ideal, image segmentation algorithms such as DenseCut generally find it difficult to give satisfactory book page segmentation results. The segmentation algorithm in this paper is tested on 2.5×104 images to be inspected. The results show that the algorithm can achieve satisfactory book page segmentation results.

2.2. Image Correction

Image distortion primarily occurs because users find it challenging to ensure that the camera head's plane is parallel to the book page during the shooting process. This distortion can be considered as a perspective distortion. Following image segmentation, perspective transformation is applied to correct the image distortion of the book pages being inspected. Let (u, v) be a point on the book page. (x, y) is the corresponding point on the image to be inspected after perspective mapping occurs. But:

$$(x', y', w') = (u, v, 1)\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \tag{5}$$

In the formula: $x = x'/w'; y = y'/w'$ to obtain the projection transformation parameters in Equation 5, we need to find at least 4 pairs of corresponding points from the map page and the image to be inspected.

Point pairs are identified from the result of image segmentation to solve the projective transformation parameters (Figure 3). Even with the perspective distortion, the book page is still a quadrilateral in the image. So we sample a series of discrete points on the edges of the image segmentation result. We fit a quadrilateral through a polygon approximation algorithm. The four vertices of the quadrilateral are denoted as $Q_0(x_0, y_0), Q_1(x_1, y_1), Q_2(x_2, y_2), Q_3(x_3, y_3)$. The corrected quasi-standard image of the book page is a square with a width of w, and the corresponding four vertices are: $P_0(u_0 = 0,\ v_0 = 0), P_1(u_1 = w,\ v_1 = 0), P_2(u_2 = w,\ v_2 = w), P_3(u_3 = 0,\ v_3 = w)$.

We can solve the projective transformation parameters by substituting the four-point pairs (Q_0, P_0), (Q_1, P_1), (Q_2, P_2), (Q_3, P_3) into Equation 5. Next, the corresponding coordinates on the image to be inspected are obtained through the perspective transformation of Equation 5 for each pixel coordinate on the standard image of the book page after correction. Subsequently, the pixel color value is taken from the corresponding coordinates of the image to be inspected and filled in the corrected image to obtain the quasi-standard image of the book page. Despite potential under-segmented or over-segmented outcomes, the distortion correction method can still achieve satisfactory results (Figure 3).

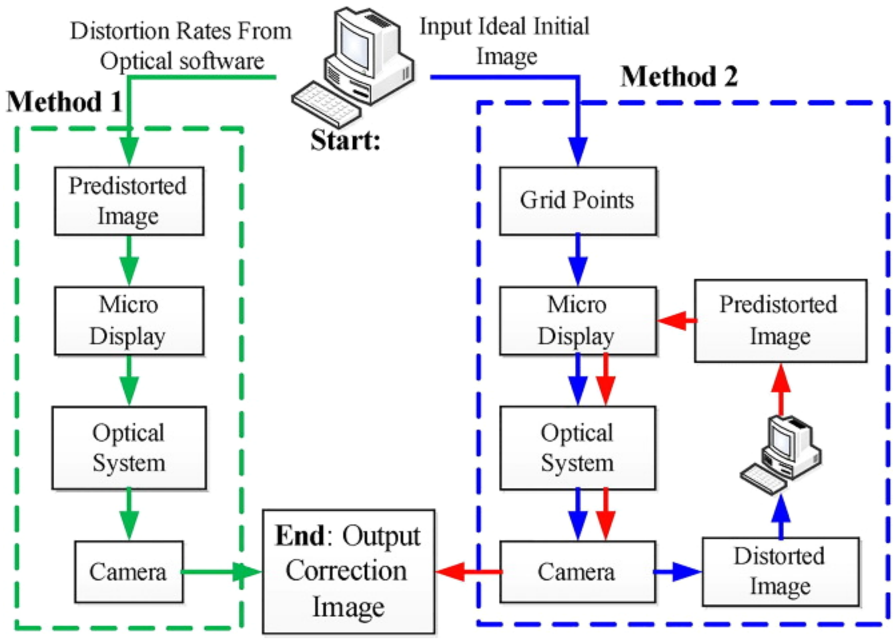

Figure 3. Image Distortion Correction

2.3. Feature Extraction

In recent years, convolutional neural networks have achieved remarkable results in the field of image retrieval. Although these methods can provide end-to-end image retrieval capabilities, they require millions, or even tens of millions, of target image data to train convolutional neural networks. Collecting such large-scale image data is a time-consuming and labor-intensive task.

A convolutional neural network text classification model is required. The model is trained on known book text data to accurately classify it. The loss function of the convolutional neural network model and the network model structure are redefined. Typical samples are extracted from previous models, and new sample data is added for training operations to obtain a new incremental network model. This allows for updates in model classification and predictions of new class samples. The incremental model is depicted in Figure 4.

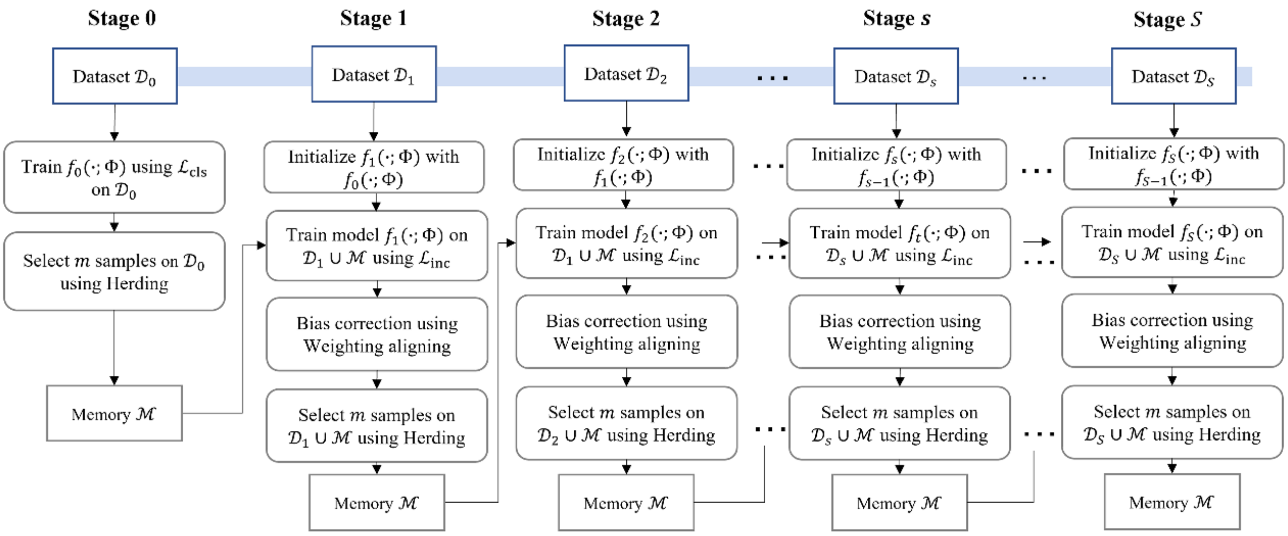

Figure 4. Incremental learning flow chart

When a new book text category sample arrives, the new sample needs to be merged with the old book text category sample to enable text expansion. We construct the final training set D of book texts. The D formula is shown Equation 6.

$$D \leftarrow \bigcup_{y=s,\cdots,t} \{(x,y): x \in X^y\} \cup \bigcup_{y=1,\cdots,s-1} \{(x,y): x \in P^y\} \tag{6}$$

$X^s, \ldots, X^t$ represents the sample set corresponding to each newly added book text category. P represents the current sample set. The algorithm needs to preprocess all the texts in the input book text sample set$P$ to be classified. After a convolutional neural network model, we extract text feature vectors. Each text corresponds to a feature vector. We add and average all the feature vectors in the text contained in each type of sample set obtained. This way we get the corresponding mean eigenvector $\mu_y \leftarrow \frac{1}{|P_y|}\sum_{p\in P_y}\phi(p)$. where$y = 1, \ldots, t$. When the CNN convolutional layers on which the classifier depends are changed, the classifier will be changed and modified adaptively.

In the second step, we prune the old class sample set and construct a new sample set. We compute the class-averaged vector $\mu \leftarrow \frac{1}{n}\sum_{x\in X}\phi(x)$of the$n$ feature vectors corresponding to the$n$ book texts in the current class X. Iterate over all $n$eigenvectors in turn and divide by the current number of samples to get the mean. We select the book text corresponding to the top $m$closest to the class average vectorμ features. We map it to the sample set of the corresponding class. Where $k = 1, \ldots, m$. After adding a new class, delete the samples with lower priority according to the priority list, that is, we reduce the samples of each classification.

In the third step, we use the newly constructed sample set. We take the gradient through the loss function and update the neural network parameters.

$$I(\theta) = I_{classification}(\theta) + I_{distillation}(\theta) \tag{7}$$

Where it contains the loss function; $I_{classification}(\theta) = -\sum_{(x_i,y_i)\in D}\sum_{y=s}\delta_{y=y_i}\log g_y(x_i) + \delta_{y\neq y_i}\log(1-g_y(x_i))$ and distillation loss function$I_{distillation}(\theta) = -\sum_{(x_i,y_i)\in D}\sum_{y=1}^{s-1} q_i^y \log g_y(x_i) + (1-q_i^y)\log(1-g_y(x_i))$.The classification loss function can make the data distinguish between the current category data and the data in the sample set. The distillation loss function can make the current model as close as possible to the response of the old model.

This study aims to train a convolutional neural network using an existing task-independent database and extract intermediate-layer features for book page retrieval. The ILSVRC2041 dataset was selected for training, which contains 1.2 × 106 educational graphics for one thousand different item types.

Numerous research studies have proposed convolutional neural networks with different structures. And they all successfully verified the ability of the method based on deep convolutional neural networks to learn image feature representation. In this study, the VGG convolutional neural network Fast version (VGG-F) with a depth of 8 layers is used to extract the image features of book pages. The VGG-F convolutional neural network includes an input layer, five convolutional layers (C1–C5), and three fully connected layers (F6–F8). The input layer receives a 3-channel 224 by 224 megapixel color picture. CNN layer C1 contains 64 convolutional huge kernels of 11 × 11 pixels. The strides are four pixels. A convolutional layer C2 contains 256 convolution kernels of size 5 × 5 pixels. Convolutional layers C3-C5 all contain 256 convolution kernels with a size of 3 × 3 pixels. Convolutional layers C1, C2, and C5 use pooling operations to decrease the characteristic's dimensions. F6 and F7 completely interconnected levels use the dropout operation to prevent overfitting. The output layer F8 uses softmax as the activation function.

The ILSVRC dataset is utilized to reduce the classification error for 1000 classes of objects, serving as the target for training VGG-F. Following the passage of data through the convolutional neural network, the output layer F8 is removed. From either the F6 or F7 layers, we can extract 4093-dimensional image features for the book page. The characteristics of the photograph taken using the F7 layer demonstrate superior retrieval performance [14]. Therefore, in this paper, the image after background segmentation and distortion correction is input into the trained VGG-F convolutional neural network, and the image feature codes of book pages are extracted from the F7 layer.

2.4. Feature matching and retrieval

The cosine distance of the included angle is used to measure the similarity between the image of the applicant's standards picture and the portion of the book that has been verified. Suppose X_i and X_i represent the graphic's attribute identifier extracted by the convolutional neural network, respectively. The similarity between the two is:

$$S_{i,j} = \frac{X_i X_j^T}{\sqrt{X_i X_i^T}\sqrt{X_j X_j^T}} \tag{8}$$

The formula $\sqrt{X_i X_i^T}$ does not affect the ranking of candidate standard images$j$. We can calculate it in an offline way. We rewrite the similarity calculation formula of Equation 6 as:

$$S_{i,j} = p_j(X_i X_j^T) \tag{9}$$

Where $p_j(X_j X_j^T)^{-1/2}$can be calculated offline? In this way, calculating the degree to which an applicant's standard picture, as well as the picture that has been verified, are comparable only consumes 40102 multiplications and 4093 additions during online retrieval. In this study, the exhaustive method is employed to directly retrieve standard images of book pages from the database. This involves using Equation 7 to calculate the similarity between each candidate standard image in the database and the image to be inspected individually. The K candidate standard images with the highest matching degree are then selected as the retrieval results [15]. The time complexity of the exhaustive retrieval algorithm is approximately O(dM), where d is the dimension of the feature code and M is the number of candidate standard images in the database. Using a single-threaded program on a laptop computer, we measured the time-consuming of the exhaustive retrieval algorithm when the dimension of the feature code is 4093 and the number of candidate standard images in the database is between 1×103 and 1×105. Figure 4 illustrates the relationship between the time-consuming t of the exhaustive retrieval algorithm in this paper and the number H of candidates in the database. When the number of candidate standard images in the database is 1×105, the time-consuming of the exhaustive retrieval algorithm in this paper is only 415.33 ms. This indicates that the exhaustive retrieval algorithm in this paper can fully meet the user's time-consuming requirements for book page retrieval in small and medium-sized databases. Large-scale and ultra-large-scale databases may use methods such as compressing signature dimensions or hash coding to reduce retrieval time, although this may reduce retrieval accuracy (Figure 5).

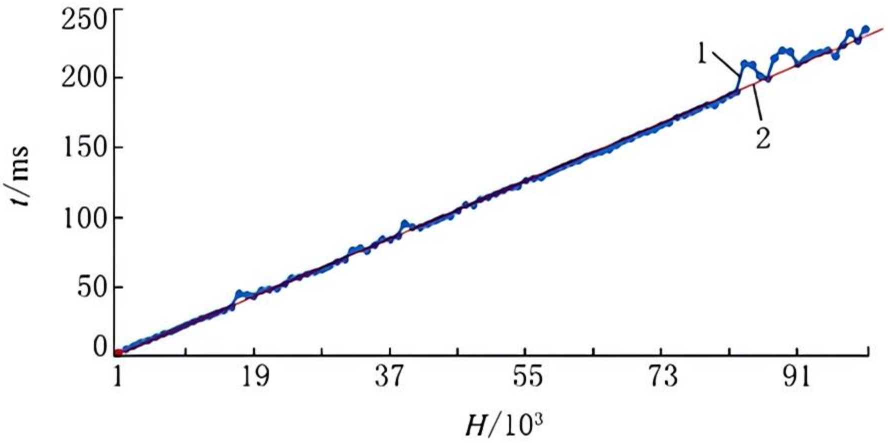

Figure 5. The relationship between the time-consuming of the exhaustive retrieval algorithm in this paper and the size of the database

3. Experimental Results and Analysis

3.1. Experimental Setup

The study collected 5×103 pages of books and created a test dataset to verify the proposed book page retrieval method. Each page of the book was scanned with a scanner to obtain candidate standard images, while five images of each page from different angles were captured using a smartphone as the images to be inspected. The test dataset thus comprised 5×103 candidate standard images and 2.5×104 images to be tested. Various harsh environments were simulated during the image capture process, including cluttered backgrounds, perspective changes, geometric distortion, scale changes, motion blur, illumination changes, and local highlights. We use the Top-k success rating was determined with the assessment which is a quantifiable index of the experimental results. $\gamma_k = N_k/N$ Determines the top-k hit ratings.

N represents the number of experiments. N_k Represents the quantity in successful retrievals (the first k images with the highest matching similarity given by the retrieval algorithm contain correct potential standards pictures).

3.2. Experimental outcome and investigation

Figure 6 illustrates the representative process in this paper, and each column in Figure 6 is the top 5 matching similarity from high to low from top to bottom. Correct results are marked with a green tick. The results in the first and second columns of Figure 5 show that the method in this paper can better distinguish book page images with high similarity. The retrieval results in the third column of Figure 6 demonstrate whether the suggested approach can get around the effects of image highlights and motion blur. Although the images in the fourth column of Figure 6 and the fifth column of Figure 6 are over-segmented, the method in this paper can still give correct retrieval results. One noteworthy finding is seen in Figure 6, specifically in the fifth column: the method effectively produces appropriate retrieval findings even if the rectified picture of the potential standard images within the information is rotated by 90 degrees. Furthermore, both of the initial corresponding resemblance pictures located in the seventh column of the figure show some lexical parallels to the photograph under examination, even if the subsequent match resemblance photograph is the right conclusion. This demonstrates how well the machine learning approach handles differences in library page photos, including crowded communities, distortion of images, local points of interest, blurred movement, along with additional frequent occurrences. As a result, the approach always produces the best recovery results, regardless of difficult circumstances. This highlights the efficiency and efficacy of the suggested method and its possibilities for use in practical situations where accurate and effective picture recovery is crucial. Consequently, the results show how well the technique can handle the complexity of imaging duties, offering a viable way to improve the effectiveness and precision of computerized processes across a range of industries.

Figure 6. Part of the retrieval results of the book page retrieval method in this article

The approach used in this work was contrasted with the approach across the datasets being tested, and the outcomes of the study are presented in Table 1 [16]. Reference is an end-to-end image retrieval algorithm based on a convolutional neural network. In the experiment, the convolutional neural networks of the two algorithms are trained using the ILSVRC dataset [17]. From the results in Table 1, it can be seen that the method γ1 of the literature is only 41.77%, and the method γ5 is only increased to 58.47%. This shows that the convolutional neural network trained on unrelated datasets is directly used for book page retrieval. Not ideal. The hit rate of the method in this paper is above 93%. γ5 also reaches 99.31%.

The image correction and image segmentation modules are gradually removed, and experiments are conducted on the test dataset to verify the method's rationality and the necessity of each module [18]. The experimental results in Table 2 indicate a sharp drop in the retrieval hit rate after removing the image segmentation and image correction modules [19], highlighting the necessity of these modules in the method [20]. The modular combination of the methods in this paper is deemed more reasonable.

Table 1. Comparison of hit rate % of the method and the literature paper

Method	γ1	γ2	γ3	γ4	γ5
Literature	41.77	41.24	36.13	49.84	58.47
This article	93.18	93.47	93.91	99.14	99.31

Table 2. Hit rate % after removing image segmentation and image correction modules

Method	γ1	γ2	γ3	γ4	γ5
Remove all modules only	93.18	93.47	93.91	99.14	99.31
Image Correction removes images	24.32	37.31	48.57	57.79	61.33
Segmentation, correction	17.64	24.73	37.53	47.62	56.69

The Euclidean distance and the cosine distance of the included angle shown in Equation 9 are used to measure the degree of similarity between the applicant's reference picture and the one being examined. Experiments are conducted on the test dataset [21]. The hit rates of the two feature similarity measurement methods are represented in Table 3. The outcomes of the test show Euclidean distance is 1.14%, 1.09%, 0.53%, 0.17%, and 0.03% lower than the $\gamma1 \sim \gamma5$ when the included angle cosine distance is adopted. In addition, when we use the cosine distance of the included angle shown in Equation 9, the average exhaustive retrieval time of each image to be inspected on the test data set is 12.13 ms. However, the average exhaustive retrieval time increases to 17.56 ms when using Euclidean distance.

Table 3. Comparison of hit rate % using different similarity measure methods

Method	γ1	γ2	γ3	γ4	γ5
Euclidean distance	95.04	95.38	93.45	93.102	102.28
Angle cosine distance	93.18	93.47	93.91	102.14	102.31

A thorough assessment of each module's efficiency was conducted on a laptop computer equipped with a Core i5-based 2.4GHz dual-core CPU and four gigabytes of RAM. Image segmentation was analyzed using pictures with a pixel dimension of 400×300, while distortion correction activities were performed on images with a pixel size of 224 × 300. Following convolutional neural networks' characteristic codes in the extraction procedure, the resulting feature codes were converted through a numbers format to expedite the pairing and retrieving procedures that followed. Carefully written in C++ employing OpenCV, the evaluation program ensures seamless implementation and effective operation despite requiring GPU or multitasking capabilities. After analyzing the running durations in the experiment's information set, significant numbers surfaced for every module that was studied. In particular, the deformity correcting unit showed a median time for processing of 5.04 ms, while the picture segmented unit showed an overall operating period of 28.102 ms. The characteristic coding collection procedure using convolutional neural network models required a mean of 51.58 ms, whereas the comparison and retrieving procedure took another 12.13 ms. The total time consumed by processing each of these steps added up to 103.42 milliseconds. Most importantly, our test results showed how amazing the technique is at providing almost immediate replies on desktop PCs. The above efficiency grade additionally meets, and sometimes exceeds, the reaction time constraints that are commonly anticipated in a variety of scenarios. The technique's exceptional effectiveness and productivity highlight its immediate relevance and adaptability, which makes it an excellent choice for implementation in a variety of contexts in which quick evaluation of images and recovery are essential elements of operations and decision-making processes [22].

3.3. Comparison of Existing with Proposed Method

In this paper, we have compared the proposed method of CNN with the existing methods such as Cascade Region Convolution Neural Network (Cascade R-CNN), You Only Look Once (YOLOv3), YOLOv4, and improved YOLOv4 [23], and precision, recall, F1 score, and detection time are parameters used to compare with the existing and proposed methods, as shown in Table 4.

Table 4. Numerical outcomes of proposed and existing method

Model	Precision	Recall	F1-Score	Detection time
Cascade R-CNN [23]	70.42	79.25	84.13	35.32
YOLOv3 [23]	73.98	88.01	80.38	16.52
YOLOv4	81.39	92.14	86.44	92.84
Improved YOLOv4	90.33	97.20	93.64	14.29
CNN (Proposed method)	95.37	98.43	96.23	13.91

Precision reflects the percentage of materials that are accurately recognized out of all the works of literature that the algorithm has deemed pertinent. Preciseness is essentially a measure of how well the method reduces errors in classification, or situations when an object is mistakenly classified as falling into a particular group of books. A high accuracy score means that the algorithm correctly classifies literature and successfully removes unnecessary ones. The value of precision in CNN obtained was 95.37%, which is higher than the existing methods such as Cascade R-CNN (70.42%), YOLOv3 (73.98%), YOLOv4 (81.39%), and improved YOLOv4 (90.33%), as shown in Figure 7.

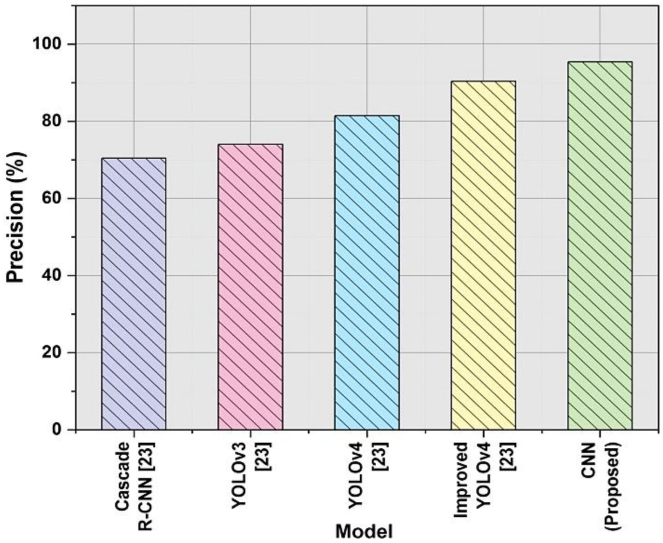

Figure 7. Graphical representation of precision

An important gauge of performance that's utilized to assess well-recognized by the system recognizes and locates books in the holdings of a library. Recall, often referred to as sensitivity, gauges how well the system for recognizing books recognizes each and every pertinent occurrence of a book within the collection of the library. Recall is essentially a measurement of how well able the system is to identify and collect each title in the library, with any being missed. It is determined by dividing the total amount of volumes in the library by the total amount of volumes that were accurately recognized. A strong recall score means that few errors or missed identities occur since the identification software correctly recovers a significant amount of the library's collection. Figure 8 depicts the value of recall in CNN (98.43%) and Cascade R-CNN (79.25%), YOLOv3 (88.01%), YOLOv4 (92.14%), and improved YOLOv4 (97.20%).

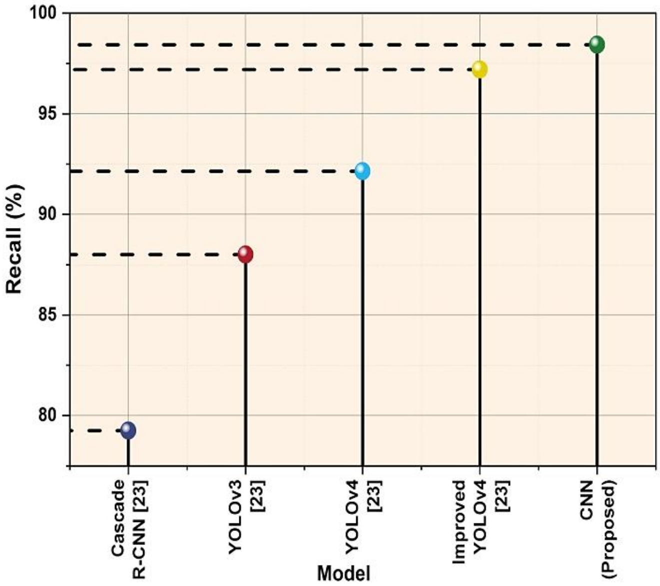

Figure 8. Comparison of recall

A comprehensive evaluation of the efficacy of a model is offered by the F1 score, which is an indicator of correctness that accounts for both precision and recall. Recall was the percentage of properly recognized occurrences over all truly good situations, whereas precise counts the percentage of properly determined occasions across every situation labeled as good. The F1 score provides a thorough assessment of the classification capability through taking into account both precision and recall, which makes such an invaluable tool to use in improving and perfecting recognition methods in library environments. F1 score ranges 96.23% in CNN, whereas existing methods like Cascade R-CNN at 84.13%, YOLOv3 at 80.38%, YOLOv4 at (86.44%), and improved YOLOv4 at (96.23%) as illustrated in Figure 9.

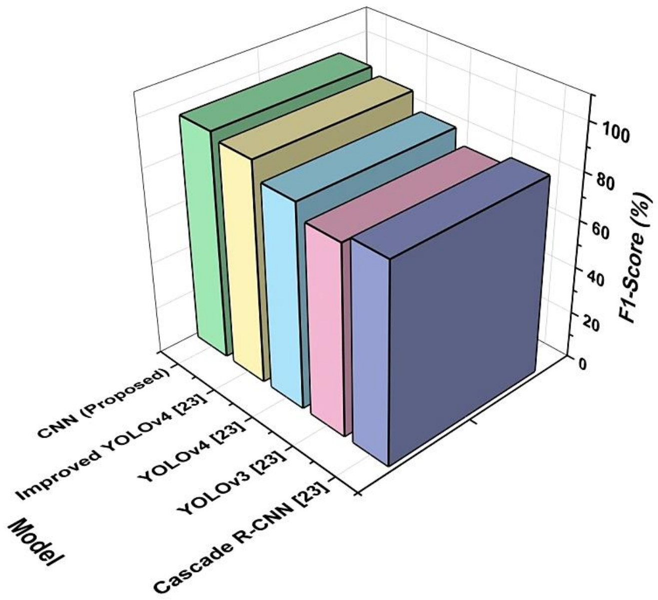

Figure 9. Comparison of F1 score

The length of time needed for the procedure to correctly assign books to a library collection. The concept includes the full procedure for taking pictures of covers or sections, analyzing them with recognition computations, and producing pertinent information for cataloging. ART methods' time for identification was a crucial component, especially in high-volume or actual-time library settings where quickness and effectiveness are crucial. Libraries may increase general effectiveness, optimize processes, and improve customer experience by identifying books more quickly when detection times are lower. CNN has obtained 13.91 ms in detection time, Cascade R-CNN (35.32 ms), YOLOv3 (16.52 ms), YOLOv4 (15.72 ms), and improved YOLOv4 (14.29 ms), as shown in Figure 10.

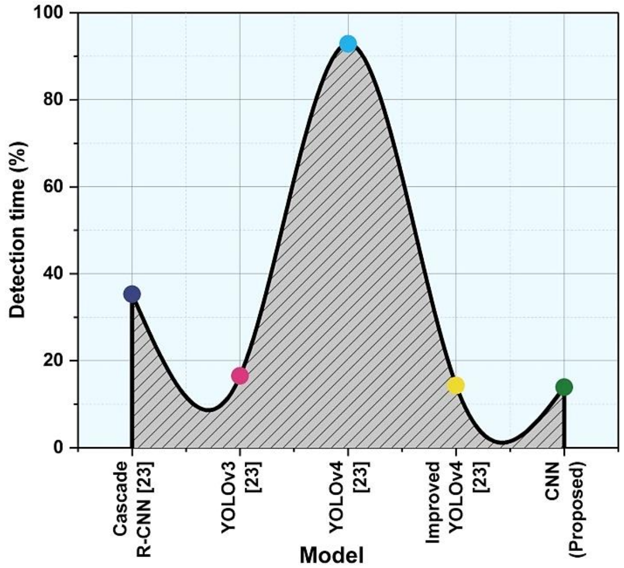

Figure 10. Graphical representation of Detection time

4. Conclusion

In this study, a unique approach for booking page recovery using convolutional neural networks (CNNs) was presented. The suggested method entails several crucial actions to improve the effectiveness and precision of retrieving. To enhance quality and uniformity, the acquired photos are first processed with foreground separation and distortion restoration procedures. Those previously treated photos are then sent through a CNN that has been constructed from task-independent information in order to acquire feature rules that operate on essentially distinct characteristics within each image. Using trigonometric proximity computations to compare the characteristic coding in the examined picture to the ones for prospective conventional pictures constitutes a few of the unique features of the suggested technique. The experiments conducted on the evaluation of a database, which showed outstanding retrieving precision, provided that this approach may yield accurate and dependable retrieving outcomes. The work brings out a fascinating limitation of end-to-end CNNs taught solely on task-independent information: they frequently score poorly when it comes to retrieving.

On the other hand, the suggested approach avoids the requirement requiring substantial gathering of information activities to construct enormous repositories of book page images. Rather than requiring job-specific data to be trained, it makes use of CNNs' strong visual feature characterization characteristics to attain higher retrieved reliability. When contrasted with conventional uniform recognition methods, the approach suggested in this investigation greatly increases the reliability and productivity of line recognition. Through the use of sophisticated computer modeling techniques, like CNNs, this investigation produces better evaluation and interpretation outcomes for book page pictures. To improve line detection reliability and system suggestions, the study presents a unique book system for suggestions that makes use of convolutional neural networks (CNN) and deep learning. It draws attention to how advanced machine learning techniques have revolutionized systems that provide recommendations. In the future, researchers hope to use hashing encoding and feature compression to significantly improve the approach's retrieving performance. With the help of such optimizations', the technique should be able to handle extremely massive book page retrieving workloads with effectiveness and stability. Through constant improvement and integration of novel approaches, the suggested strategy has the potential to progress in the area of book page recovery and tackle the difficulties related to managing massive picture databases. A groundbreaking development in library management software that provides availability, reliability, and economy motivates the scholarship. Libraries may concentrate on user interaction and collection growth since it spares both money and time. This equipment also creates new avenues for scholarly study, enabling students to work with librarians to improve identification systems and investigate how they affect customer service, availability, and data searching.

5. Declarations

5.1. Author Contributions

Conceptualization, J.H. and Y.Y.; methodology, Z.X.; software, J.X.; validation, J.H., Y.Y., and Z.X.; investigation, J.H.; resources, Z.X.; data curation, Y.Y.; writing—original draft preparation, J.H.; writing—review and editing, Z.X.; visualization, Z.X.; supervision, Z.X.; project administration, Z.X.; funding acquisition, J.H. All authors have read and agreed to the published version of the manuscript.

5.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

5.3. Funding

This research study is sponsored by these projects: project one: Industry-university-research collaborative education program of the Ministry of Education in China, the project number is 202102205018. Project two: universities philosophy and social science researches project in Jiangsu province, the project number is: 2019SJA1473.

5.4. Institutional Review Board Statement

Not applicable.

5.5. Informed Consent Statement

Not applicable.

5.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

6. References

[1] Shi, X., Tang, K., & Lu, H. (2021). Smart library book sorting application with intelligence computer vision technology. Library Hi Tech, 39(1), 220–232. doi:10.1108/LHT-10-2019-0211.

[2] Trinquet, C. (2023). Pepper as an assistant in the library: Identifying books using machine learning. Master’s Thesis in Applied Computer Science, June 2023, Norwegian University of Science and Technology, Trondheim, Norway.

[3] Martinez-Martin, E., Ferrer, E., Vasilev, I., & Del Pobil, A. P. (2021). The UJI aerial librarian robot: A quadcopter for visual library inventory and book localisation. Sensors (Switzerland), 21(4), 1–16. doi:10.3390/s21041079.

[4] Locher, R. (2022). NPOCR–Needle Printer Character Recognition: Deep learning-based image ID recognition. Master Degree Project (60 ECTS) in Informatics, University of Skövde, Skövde, Sweden.

[5] Anumula, V. R. B., Anumula, K. R. V. S. S., Anumula, M., & Damerla, S. (2024). Artificial Intelligence Using Library Information Science Professionals. International Journal of Progressive Research in Engineering Management and Science, 4(1), 591-594.

[6] Bairagi, M., & Lihitkar, S. Empowering Libraries: AI-Driven Tools and Techniques for Digital Transformation and Sustainable Innovation. In One Day National Conference on 'Preserving Indian Knowledge System and Envisioning the role of Libraries in NEP 2020 framework', 10(1), 388–396.

[7] Mupaikwa, E. (2025). The Application of Artificial Intelligence and Machine Learning in Academic Libraries. In Encyclopedia of Information Science and Technology, IGI Global, Sixth Edition, 1-18. doi:10.4018/978-1-6684-7366-5.ch041.

[8] González- Alcaide, G., Castelló- Cogollos, L., Navarro- Molina, C., Aleixandre- Benavent, R., & Valderrama- Zurián, J. C. (2008). Library and information science research areas: Analysis of journal articles in LISA. Journal of the American Society for Information Science and Technology, 59(1), 150-154. doi:10.1002/asi.20720.

[9] Takher, S. (2024). Revolutionizing Agriculture Libraries in India: A Comprehensive Study on Implementing Near Field Communication (NFC) Technology for Enhanced Access and Knowledge Sharing. Asian Journal of Agricultural Extension, Economics & Sociology, 42(1), 128–138. doi:10.9734/ajaees/2024/v42i12355.

[10] Sibiya, M., & Sumbwanyambe, M. (2019). A Computational Procedure for the Recognition and Classification of Maize Leaf Diseases Out of Healthy Leaves Using Convolutional Neural Networks. AgriEngineering, 1(1), 119–131. doi:10.3390/agriengineering1010000.

[11] Lin, J. D., Wu, X. Y., Chai, Y., & Yin, H. P. (2020). Structure Optimization of Convolutional Neural Networks: A Survey. Zidonghua Xuebao/Acta Automatica Sinica, 46(1), 24–37. doi:10.16383/j.aas.c180275.

[12] Zhou, D. X. (2020). Universality of deep convolutional neural networks. Applied and Computational Harmonic Analysis, 48(2), 787–794. doi:10.1016/j.acha.2019.06.004.

[13] Chouiekh, A., & El Haj, E. H. I. (2020). Deep convolutional neural networks for customer churn prediction analysis. International Journal of Cognitive Informatics and Natural Intelligence, 14(1), 1–16. doi:10.4018/IJCINI.2020010101.

[14] Abidi, S. F. H., Sumukhi, T., Kumar, V., & Santhosh, B. (2021). Lung Cancer Detection Using Deep Convolutional Neural Network. International Journal of Organizational and Collective Intelligence (IJOCI), 11(4), 13-20. doi:10.4018/IJOCI.2021100102.

[15] Yamada, A., Niikura, R., Otani, K., Aoki, T., & Koike, K. (2021). Automatic detection of colorectal neoplasia in wireless colon capsule endoscopic images using a deep convolutional neural network. Endoscopy, 53(8), 832–836. doi:10.1055/a-1266-1066.

[16] Sunaga, Y., Natsuaki, R., & Hirose, A. (2019). Land Form Classification and Similar Land-Shape Discovery by Using Complex-Valued Convolutional Neural Networks. IEEE Transactions on Geoscience and Remote Sensing, 57(10), 7907–7917. doi:10.1109/TGRS.2019.2917214.

[17] Yurong, G., & Ke, Z. (2019). End-to-end dual-channel feature recalibration for DenseNet image classification. Chinese Journal of Image Graphics, 11(08), 126–127.

[18] Jingdong, L., & Xinyi, W. (2019). Overview of Convolutional Neural Network Structural Optimization. Acta Automata, 14(09), 79-83.

[19] Surono, S., Rivaldi, M., Dewi, D. A., & Irsalinda, N. (2023). New Approach to Image Segmentation: U-Net Convolutional Network for Multiresolution CT Image Lung Segmentation. Emerging Science Journal, 7(2), 498-506. doi:10.28991/ESJ-2023-07-02-014.

[20] Surono, S., Afitian, M. Y. F., Setyawan, A., Arofah, D. K. E., & Thobirin, A. (2023). Comparison of CNN Classification Model using Machine Learning with Bayesian Optimizer. HighTech and Innovation Journal, 4(3), 531-542. doi:10.28991/HIJ-2023-04-03-05.

[21] Roy, A. M., Bose, R., & Bhaduri, J. (2022). A fast accurate fine-grain object detection model based on YOLOv4 deep neural network. Neural Computing and Applications, 34(5), 3895–3921. doi:10.1007/s00521-021-06651-x.

[22] Ji, S. J., Ling, Q. H., & Han, F. (2023). An improved algorithm for small object detection based on YOLO v4 and multi-scale contextual information. Computers and Electrical Engineering, 105, 108490. doi:10.1016/j.compeleceng.2022.108490.

[23] Pang, H., Zhang, Y., Cai, W., Li, B., & Song, R. (2022). A real-time object detection model for orchard pests based on improved YOLOv4 algorithm. Scientific Reports, 12(1), 13557. doi:10.1038/s41598-022-17826-4.