frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 30 June 2023
DOI 10.3389/fnins.2023.1230786



OPEN ACCESS

EDITED BY
Anguo Zhang,
University of Macau, China

REVIEWED BY
Xingshuo Han,
Nanyang Technological University, Singapore
Yongcheng Zhou,
Chongqing University, China

*CORRESPONDENCE
Qing Wang
✉ wangqingait@cau.edu.cn

RECEIVED 29 May 2023
ACCEPTED 15 June 2023
PUBLISHED 30 June 2023

CITATION
You J and Wang Q (2023) Sublinear information bottleneck based two-stage deep learning approach to genealogy layout recognition. *Front. Neurosci.* 17:1230786. doi: 10.3389/fnins.2023.1230786



# Sublinear information bottleneck based two-stage deep learning approach to genealogy layout recognition

Jianing You and Qing Wang*

College of Information and Electrical Engineering, China Agricultural University, Beijing, China

As an important part of human cultural heritage, the recognition of genealogy layout is of great significance for genealogy research and preservation. This paper proposes a novel method for genealogy layout recognition using our introduced sublinear information bottleneck (SIB) and two-stage deep learning approach. We first proposed an SIB for extracting relevant features from the input image, and then uses the deep learning classifier SIB-ResNet and object detector SIB-YOLOv5 to identify and localize different components of the genealogy layout. The proposed method is evaluated on a dataset of genealogy images and achieves promising results, outperforming existing state-of-the-art methods. This work demonstrates the potential of using information bottleneck and deep learning object detection for genealogy layout recognition, which can have applications in genealogy research and preservation.

KEYWORDS

**genealogy layout recognition, sublinear information bottleneck, YOLOv5 detector, ResNet, deep learning**

## 1. Introduction

A genealogy is a special document that records the family lineage and important figures in a family's history in the form of a chart (Wang and Zhang, 2008). It is a characteristic of Chinese civilization, and is a historical record of the bloodline of a kinship group, including the people and events related to the same ancestor. Genealogy is a valuable humanistic resource that plays an irreplaceable and unique role in in-depth research in fields such as history, folklore, demography, sociology, and economics. However, due to wars and social upheavals in history, the lineages and genealogies of many families have been destroyed or lost. Therefore, digital preservation of genealogy has become necessary. Through digital technology, genealogy can be digitally stored and disseminated, making it convenient for researchers and scholars to access and study, and protecting the cultural value and inheritance of genealogy (Chang, 2014).

Genealogy recognition technology is one of the important means of digital preservation of genealogy. By recognizing the ancient books of genealogy, the information in the genealogy can be automatically extracted and processed, thus realizing the digital storage and dissemination of genealogy (Fan, 2013). However, due to the complex and diverse layout structures of genealogy ancient books, the recognition difficulty is also high. Therefore, accurate layout detection and positioning technology is an important prerequisite for genealogy recognition.

Genealogy recognition relies heavily on document layout analysis technology. The most famous and widely used traditional document layout analysis algorithm is the Docstrum

algorithm proposed by L. (1993). It sequentially divides the black and white connected domains in the image into text, text lines, and text blocks from bottom to top, thus obtaining the layout. For table recognition, the table lines are obtained through erosion, dilation, and other operations, the row and column areas are divided, and then the cells are combined with text contents to reconstruct the table object. In response to some shortcomings of the algorithm, subsequent researchers have proposed corresponding optimization algorithms. For example, Wieser and Pinz (1994) proposed a method of combining bottom-up merging and top-down cutting for newspaper page segmentation. Watanabe et al. (1995) introduced a classification tree to manage different types of layout structures, and proposed a method for recognizing the layout structure of documents with multiple table formats. Liu-Gong et al. (1995) used a universal model to convert document images into layout structures. Lee et al. (2000) used geometric structure analysis to propose a knowledge-based method for analyzing complex geometric structures of journal pages. Lee and Ryu (2001) constructed a pyramid quadtree structure for multiscale analysis based on a parameter-free method, and proposed a periodicity measurement method to find the periodic properties of text regions. In addition, in the application of textual literature, Bukhari et al. (2011) proposed a layout distribution system for extracting text in reading order from scanned images of Arabic texts written in different languages and styles. However, traditional algorithms still face some technical challenges, mainly in (1) layout analysis and table structure extraction; (2) image processing methods relying on various threshold and parameter selections; (3) difficulty in ensuring generalization of document images in different scenarios.

In recent years, deep learning has shown great promise in improving the accuracy and efficiency of genealogy layout recognition. With its ability to automatically learn and extract features from large datasets, deep learning models can also adapt to different variations in genealogy images, such as variations in font styles, sizes, and orientations, without the need for manual feature engineering (Li et al., 2022). However, there are still some challenges that need to be addressed to improve the performance of deep learning models in genealogy recognition. One of the major challenges is dealing with images that are poorly scanned or contain noise or artifacts, which can affect the accuracy of recognition. Another challenge is handling complex layouts, which can make it difficult for deep learning models to distinguish between text and non-text regions. Moreover, the lack of large and diverse annotated datasets for genealogy recognition limits the performance of deep learning models. This is because deep learning models require large amounts of annotated data to effectively learn and generalize to new data. Therefore, efforts to create more annotated datasets for genealogy recognition are needed to improve the performance of deep learning models in this domain.

Therefore, in this paper, we present a two-stage deep learning approach for genealogy image layout recognition. Firstly, genealogy images are fed into the ResNet classifier model to identify whether the image contains a bordered or borderless image. Based on the classification results, the image is then directed to either a borderless or a bordered YOLOv5 object detection model. Both the deep learning classifier and object detection models are trained using large amounts of labeled data, which can lead to high accuracy in recognizing different image features and layouts, and can be easily scaled to recognize a large number of different image layouts, making them suitable for recognizing different genealogy image layouts. Further, the proposed method can also be relatively fast at recognizing image layouts, making them suitable for real-time or near real-time applications. Furthermore, we introduce the sublinear information bottleneck (SIB) algorithm to compress the intermediate feature representation of the network model as much as possible while ensuring the accuracy of the model's output, thus achieving high generalization and strong robustness in layout recognition. The main contributions of this paper can be summarized as follows:

(1) Proposing a two-stage deep learning approach that providing advantages of high accuracy, scalability, flexibility, and speed performance;

(2) Introducing the SIB compression technology to improve the network's generalization performance, making it more adaptable to genealogy images.

(3) Using rich collection resources, we scan and manually label common genealogy images to establish a genealogy image and layout positioning standard dataset for model training and experimental result testing.

The rest of this paper is organized as follows: Section 2 reviews related work on deep learning-based layout recognition and the basic concept of information bottleneck. In Section 3, we introduce our proposed layout recognition method in detail, including the SIB, SIB-ResNet classifier and SIB-YOLOv5 object detector. In Section 4, we construct a genealogy layout dataset that we designed and established independently and perform performance testing on our proposed method. Finally, Section 5 summarizes the whole paper.

## 2. Related works

### 2.1. Deep learning based genealogy layout recognition

In recent years, deep learning-based approaches have become increasingly popular for genealogy layout recognition due to their ability to automatically learn discriminative features from data. Borges Oliveira and Viana (2017) proposed a fast automatic document layout method based on convolutional neural networks (CNN), which greatly improved overall performance. Moreover, Kosaraju et al. (2019) proposed a texture-based convolutional neural network model called DoT-Net, which can effectively recognize document component blocks such as text, images, tables, mathematical expressions, and line graphs, solving the problems caused by location transformations, inter-class and intra-class variations, and background noise. Singh and Karayev (2021) unveil an architecture for a Handwritten Text Recognition (HTR) model based on Neural Networks, which is capable of recognizing complete pages of handwritten or printed text without the need for image segmentation. It is built on the Image to Sequence architecture, allowing it to accurately extract text from an image and sequence it correctly. Additionally, it can be trained to generate

auxiliary markup that pertains to formatting, layout, and content. Wu et al. (2023) proposed a genealogical knowledge graph model to implement the construction and applications of genealogical knowledge graphs. One of the challenges in genealogy layout recognition is the lack of large datasets, as well as the presence of various types of noise, such as text overlap, low contrast, and curved text. To address this, researchers have proposed different strategies, such as dataset collection and data augmentation.

In Singh and Karayev (2021), Sumeet et al. presented TableBank, a new image-based table detection and recognition datase with 417K high quality labeled tables, allowing building strong baselines of deep neural networks. Zhong et al. (2019) introduced the PubLayNet dataset for document layout analysis, the dataset is created by automatically associating the XML representations with the content of more than 1 million publicly available PDF articles on PubMed Central. Data augmentation is also a commonly used technique to increase the size and diversity of training data. To address the issue of data scarcity for rare family relationships, He et al. (2021) leveraged data augmentation technology to generate additional synthetic data. Subsequently, they developed a multitask-based artificial neural network model capable of simultaneously detecting names, extracting relationships between individuals, and assigning attributes such as birth and death dates, residence, age, and gender.

Deep learning-based approaches have shown promising results for genealogy layout recognition. However, there is still room for improvement, particularly in handling complex genealogy layouts with overlapping and curved text.

## 2.2. Information bottleneck

Information Bottleneck (IB) was first proposed by Tishby and Zaslavsky (1999) for traditional machine learning methods. In 2015, Tishby hypothesized in his paper that deep learning is an information bottlenecking procedure that compresses data noise as much as possible and keeps the information that the data wants to convey (Tishby and Zaslavsky, 2015). This suggests that neural networks are like squeezing information into a bottleneck, leaving only the features that are most relevant to the general concept and removing the large amount of irrelevant and noisy data. Later, it was used in Schwartz-Ziv and Tishby (2017) for the study of interpretability of deep learning, and realized the effective combination of information bottleneck theory and deep learning networks.

The Information Bottleneck is an information theory method used for tasks such as data compression and classification, which effectively extracts key information from data. The core idea is to minimize the uncertainty of the output information while retaining the maximum amount of input information.

Specifically, given input random variable $X$ and output random variable $Y$, the IB method finds an intermediate random variable $T$ to describe the relationship between input and output, which maximally preserves the information of $X$, while minimizing the information entropy between $T$ and $Y$, i.e.,

$$I(T;X) - \beta I(T;Y) \tag{1}$$

where $I(T;X)$ and $I(T;Y)$ are the mutual information between $T$ and $X$, and $T$ and $Y$, respectively, and $\beta$ is a tuning parameter that balances the information entropy between $X$ and $Y$, and that between $T$ and $Y$.

The advantage of the IB method is that it can automatically learn the most important features from data without requiring prior knowledge, thus extracting the most useful information. It has been widely used in natural language processing, image recognition, signal processing, and other fields.

Dong and He (2023) used the optimization objective proposed by the information bottleneck theory, added a loss function to the tensor input to the linear classification layer in the model, and aligned the clean samples with the high-level features obtained when the adversarial samples are input to the model by sample cross-training. Li and Liu (2019) employed IB theory to understand the dynamic behavior of convolutional neural networks (CNNs) and investigate how the fundamental features have impact on the performance of CNNs. To construct a classifier which is more robust to small perturbations in the input space, Pensia et al. (2020) propose a novel strategy for extracting features in supervised learning. The experimental results for synthetic and real data sets show that the proposed feature extraction methods indeed produce classifiers with increased robustness to perturbations. In Song et al. (2022), Song et al. investigated for the first time a novel and flexible multimodal representation learning method, multi-feature deep information bottleneck (MDIB), for breast cancer classification in CESM. Moreover, Amjad and Geiger (2020) investigate training deep neural networks (DNNs) for classification via minimizing the information bottleneck (IB) functional. The above studies show that information bottleneck theory has a positive impact on feature extraction, model optimization and performance improvement of deep learning models.

# 3. Proposed method

## 3.1. Sublinear information bottleneck

The SIB is a novel method for unsupervised feature selection and compression that extends the original information bottleneck (IB) method by incorporating second-order statistics of the input data (Kolchinsky et al., 2018).

The SIB method aims to find a compressed representation $T$ of the input data $X$ that preserves the most relevant information for a given task $Y$. Specifically, the method seeks to minimize the following objective function:

$$\mathcal{L} = I^2(X;T) + \beta H(Y|T) \tag{2}$$

where $I^2(X;T)$ is the second-order mutual information between $X$ and $T$, and $H(Y|T)$ is the conditional entropy of $Y$ given $T$.

The first term $I^2(X;T)$ measures the amount of second-order statistical dependence between $X$ and $T$, while the second term $H(Y|T)$ measures the amount of uncertainty in predicting $Y$ given $T$. By minimizing this objective function $\mathcal{L}$, the SIB method finds a compressed representation $T$ that preserves the most relevant information for predicting $Y$.

The optimization problem is typically solved using a Lagrangian relaxation approach, which leads to a set of non-linear equations that can be solved iteratively (Juttner et al., 2001). Compared to the original IB method (Owen, 2001), the SIB method takes into account the second-order statistics of the input data, which can capture higher-order dependencies and correlations between input variables. This can lead to better feature selection and compression performance, especially in complex datasets with non-linear dependencies.

The conventional problem in IB theory is to minimize the mutual information $I(X;T)$ with respect to the encoding mapping $p(t|x)$, given a fixed input distribution $p(x)$. This function is convex. On the other hand, maximizing the conditional entropy $H(Y|T) = H(Y) - I(T;Y)$ with respect to the decoding mapping $p(y|t)$, given a fixed joint distribution $p(x,y)$, is a concave function. The entropy $H(Y)$ is a constant for the dataset or assumed to have a small fluctuation for a patch of training data. Therefore, the SIB function $\mathcal{L}$ used to determine the global or local minimum is a concave function.

We define the optimal parameter set $\omega$ to achieve the best performance of the minimal value of the loss $\mathcal{L}$ and a randomly generated parameter set $\phi$. Let the representation predicted by $\phi$ be $\hat{Y}$, and $\hat{y}$ be an instance of $\hat{Y}$. Thus, we have

$$\begin{aligned} H\left(q_\phi(Y|T)\right) &\leq H\left(q_\phi(Y|T)\right) + D_{KL}\left(q_\omega(Y|T) \parallel q_\phi(Y|T)\right) \\ &= -\mathbf{E}_{q_{\omega(Y,T)}}\left[\log q_\phi(Y|T)\right] \\ &= -\mathbf{E}_{q_{\omega(Y,\hat{Y})}}\left(\mathbf{E}_{q_{\phi(\hat{Y},T)}}\left[\log q_\phi(Y|\hat{Y})\right]\right) \\ &= -\mathbf{E}_{q_{\omega(Y,\hat{Y})}}\left[\log q_\phi(Y|\hat{Y})\right] \\ &= \mathcal{C}\left(q_\phi(Y|\hat{Y})\right) \end{aligned} \tag{3}$$

where $\mathbf{E}[\cdot]$ denotes the expectation value, $\mathcal{C}(\cdot)$ is the cross-entropy, and (a) is due to

$$q_{Y|T}(y|t) = q_{Y|q(\hat{y}|t)}\left(y|q\left(\hat{y}|t\right)\right) = q_{Y|\hat{Y}}\left(y|\hat{y}\right) \tag{4}$$

The equality in (3) is achieved only when $q_\phi(y|t)$ is identical to the optimal mapping $q_\omega(y|t)$. Moreover, if the Kullback-Leibler (KL) divergence $D_{KL}\left(q_\omega(Y|T) \parallel q_\phi(Y|T)\right) \to 0$, then $H\left(q_\phi(Y|T)\right) \to \mathcal{C}\left(q_\phi(Y|\hat{Y})\right)$. This implies that minimizing the distance between the network parameter set $\phi$ and the optimal set $\omega$ leads to a smaller gap between $H\left(q_\phi(Y|T)\right)$ and its upper bound. The term $I(X;T)$ denotes the information that is compressed from the input signal $X$ to the intermediate activation $T$:

$$\begin{aligned} I(X;T) &= \sum_{x,t} q(x,t) \log\left(\frac{q(x,t)}{p(x)\,q(t)}\right) \\ &= \sum_{x,t} q(x,t) \log\left(\frac{q(t|x)}{q(t)}\right) \\ &= \sum_{x,t} q(x,t) \log q(t|x) - \sum_t q(t) \log q(t) \end{aligned} \tag{5}$$

Computing the marginal distribution of $T$, $q(t) = \sum_x q(t|x)p(x)$, may pose a challenge. Taking inspiration from VIB (Alemi et al., 2017), we employed the variational distribution $r(t)$ to approximate $q(t)$. As the KL divergence is non-negative by definition, we obtain:

$$\begin{aligned} D_{KL}\left(q(T) \parallel r(T)\right) &= \sum_t q(t) \log q(t) - \sum_t q(t) \log r(t) \\ &\geq 0 \end{aligned} \tag{6}$$

According to (5) and (6), we have

$$\begin{aligned} I(X;T) &\leq \sum_{x,t} q(x,t) \log q(t|x) - \sum_t q(t) \log r(t) \\ &= \sum_{x,t} p(x)q(t|x) \log q(t|x) - \sum_{x,t} p(x)q(t|x) \log r(t) \\ &= \frac{1}{N} \sum_{n=1}^{N} q(t|x_n) \log \frac{q(t|x_n)}{r(t)} \\ &= \frac{1}{N} \sum_{n=1}^{N} D_{KL}\left[q(T|x_n) \parallel r(T)\right] \end{aligned} \tag{7}$$

where $N$ is the number of data samples that has been defined before.

By combining (3) and (7) with the constraints $H(Y|T) \geq 0, I(X;T) \geq 0$, we established an upper bound for our proposed SIB, which is given by:

$$\mathcal{L} \leq \bar{\mathcal{L}} = \mathcal{C}\left(q_\phi(Y|\hat{Y})\right) + \beta\left[\frac{1}{N} \sum_{n=1}^{N} D_{KL}\left[q(T|x_n) \parallel r(T)\right]\right]^2 \tag{8}$$

The minimization of the loss function $\mathcal{L}$ can be transformed into the minimization of the upper bound $\bar{\mathcal{L}}$, thereby achieving the objective of reducing $\mathcal{L}$.

## 3.2. SIB-ResNet

In recent years, ResNet has become one of the most popular deep neural network architectures for image classification and other computer vision tasks (He et al., 2016). One of the keys to its success is the use of residual connections, which allow the network to effectively learn features at multiple scales while minimizing the vanishing gradient problem. However, ResNet still suffers from the curse of dimensionality, as the feature maps tend to become increasingly complex and high-dimensional as they pass through the network. To alleviate this issue, we propose to insert four proposed sublinear information bottleneck (SIB) layers into ResNet, which aims to extract the most relevant information from the input while discarding the redundant information. These layers perform dimensionality reduction by encoding the feature maps into a compressed representation, which can then be decoded back to the original dimensionality. By adding these SIB layers, we aim to improve the efficiency and scalability of ResNet, while maintaining its high accuracy. As shown in Figure 1, the constructed loss for SIB-ResNet is

$$\mathcal{L}_{SIB-ResNet} = I^2(X;T_1) + \sum_{i=1}^{3} \alpha_i I(T_i;T_{i+1}) + \beta H(Y|T_4) \tag{9}$$

where $\alpha_i > 0, (i = 1,2,3)$ and $\beta > 0$ are designed parameters to balance the weight of each term.

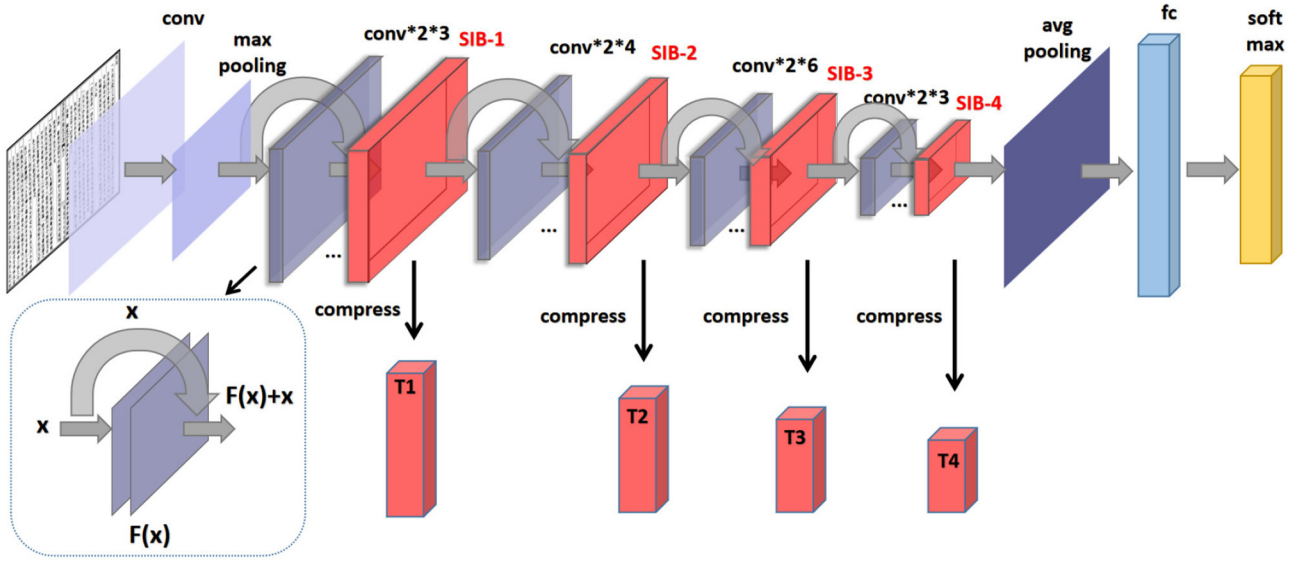


FIGURE 1
The proposed structure of SIB-ResNet.

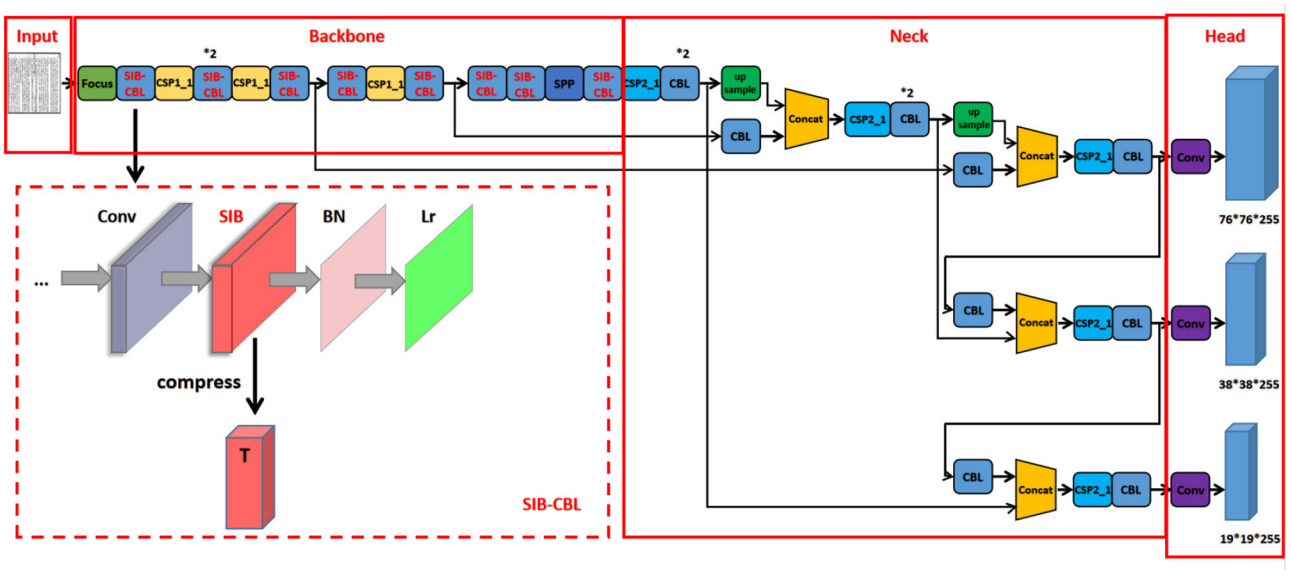


FIGURE 2
The proposed structure of SIB-YOLOv5. *Represents the number of times the module is repeated in the graph.

## 3.3. SIB-YOLOv5

The YOLOv5 model utilizes several loss functions to optimize its performance during training. These include:

(1) Objectness loss: This function aims to identify whether the object is present or not in a given anchor box. It penalizes false positives or negatives when detecting an object in the anchor box.

$$\mathcal{L}_{\text{obj}} = \lambda_{\text{obj}} \sum_{i \in \text{anchors}}^{S^2} \sum_{j=0}^{B} \mathbf{1}_{ij}^{obj} \left(\text{IOU}_{pre} - \text{IOU}_{true}\right)^2 \quad (10)$$

where $i$ denotes the index of anchor boxes, $j$ denotes the index of bounding boxes, $\mathbf{1}_{ij}^{obj}$ is an indicator function that is equal to 1 if the anchor box $i$ is assigned to the ground-truth box $j$, and 0 otherwise. $S$ is the grid size, and $B$ is the number of predicted bounding boxes each anchor predicts. $\text{IOU}_{pre}$ represents the intersection over union (IOU) between the predicted bounding box and its assigned ground truth box, while $\text{IOU}_{true}$ represents the true intersection over union (IOU) between the ground truth box and anchor box.

(2) Classification loss: This function helps to classify the object detected in the anchor box. It computes the cross-entropy between

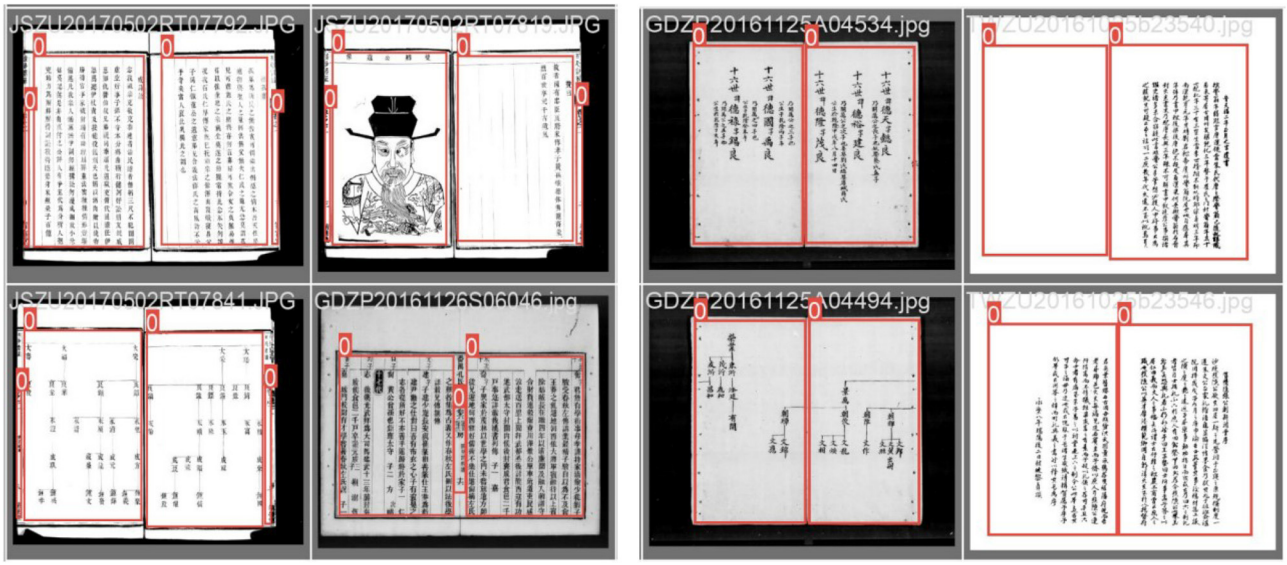


FIGURE 3
The two styles of genealogical picture and their corresponding tag boxes.

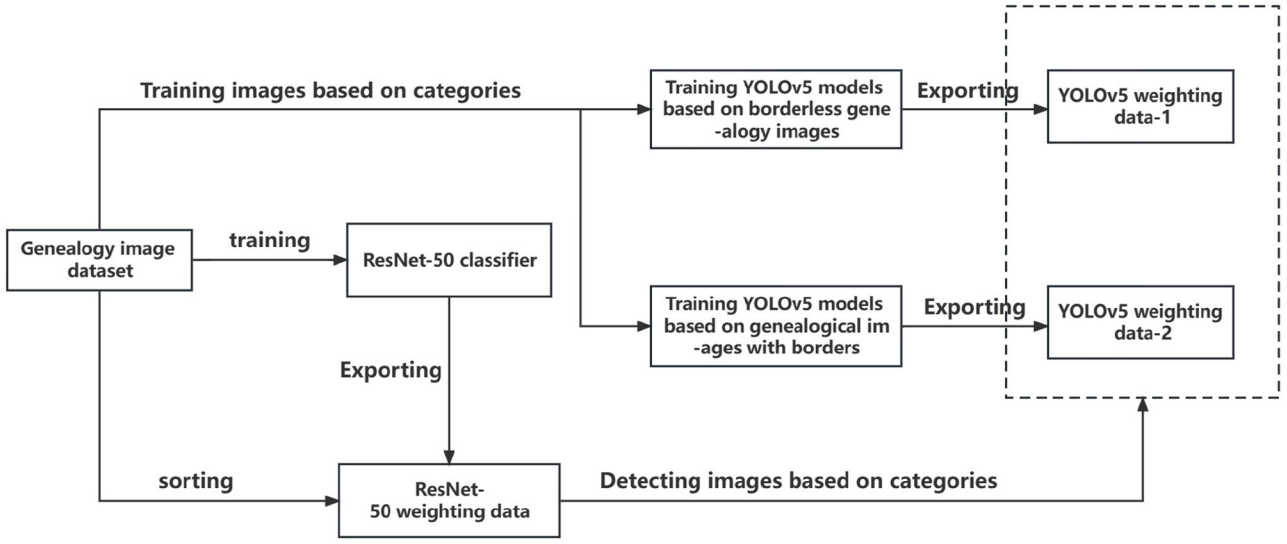


FIGURE 4
The implementation process of model.

the predicted class probabilities and the true class of the object.

$$\mathcal{L}_{\text{cls}} = \lambda_{\text{cls}} \sum_{i \in \text{anchors}}^{S^2} \sum_{j=0}^{B} \mathbf{1}_{ij}^{obj} \sum_{c \in C} \left[ p_{ij}^{c} \log\left(\hat{p}_{ij}^{c}\right) + \left(1 - p_{ij}^{c}\right) \log\left(1 - \hat{p}_{ij}^{c}\right) \right] \tag{11}$$

where $p_{ij}^{c}$ is the indicator function for the $j_{th}$ bounding box in the $i_{th}$ anchor box having class $c$ and $\hat{p}_{ij}^{c}$ is the predicted probability for the same class. $\mathbf{1}_{ij}^{obj}$ is the same as in the objectness loss.

(3) Localization loss: This function predicts the bounding box of the object in the image. It computes the mean squared error between the predicted box coordinates and the true box coordinates.

$$\mathcal{L}_{\text{loc}} = \lambda_{\text{loc}} \sum_{i \in \text{anchors}}^{S^2} \sum_{j=0}^{B} \mathbf{1}_{ij}^{obj} \sum_{m \in \{x,y,w,h\}} (\sigma_{ij}^{m})^2 \left(\hat{t}_{ij}^{m} - t_{ij}^{m}\right)^2 \tag{12}$$

where $\sigma_{ij}^{m}$ is the mask to select the predicted value of $m$, and $t_{ij}^{m}$ is the true value of $m$ for the $j_{th}$ bounding box in the $i_{th}$ anchor box. $\hat{t}_{ij}^{m}$ is the predicted value for the same parameter.

These loss functions work together to optimize the YOLOv5 model during training, allowing it to detect objects accurately and precisely in images.

TABLE 1 Comparison results of Experiment 1.

| Model | Weights [MB] | mAP [%] | Recall [%] | Speed [ms] |
|---|---|---|---|---|
| ResNet | 87.32 | 96.2 | 92.5 | 27 |
| GooLeNet | 26.33 | 90.8 | 88.9 | 19 |
| VGG | 474.66 | 97.1 | 94.8 | 39 |
| IB-ResNet | 74.49 | 92.3 | 88.3 | 22 |
| SIB-ResNet | 78.29 | 96.8 | 92.5 | 24 |

The CBL module is a key component of the YOLOv5 object detection model, it performs point-wise linear transformations coupled with pointwise activation functions on a subset of feature maps. This approach reduces the channel dimension input, greatly enhancing the non-linear characteristics of CNN features and improving their descriptive power. The resulting reduction of model complexity lowers computation costs while preserving significant spatial granularity. Compared to YOLOv4 (Bochkovskiy et al., 2020), in YOLOv5, the CBL module has been further improved by introducing a Swish activation function. This activation function is known to provide non-linearity superior to ReLU, making it a popular choice. The CBL module's mathematical formulation is then as follows:

$$y = \mathrm{CBL}(x) = \sigma(\gamma(x+\beta)) \times \mathrm{Swish}(x) \tag{13}$$

Here, $x$ is the input tensor, the bias shift $\beta$, and scaling parameter $\gamma$ are learnable parameters, while $\sigma$ and Swish denote the sigmoid and Swish activation functions, respectively. The YOLOv5 model also makes use of a similar but more complex CBL block, which leverages the power of skip connections to perform deep feature fusion, reducing computation costs and enhancing the model's accuracy simultaneously.

As shown in Figure 2, the SIB layer is embedded into CBL module to reduce the dimensions of the CBL input, forcing the module to learn a more concise and reduced representation of the input, thus improving its generalization abilities. Define the original input $X$ by $T_0$, the overall loss of our proposed SIB-YOLOv5 can be summarized as

$$\mathcal{L}_{SIB-YOLOv5} = \sum_{i=0}^{K} \alpha_i I^2 T_i;\, T_{i+1} + \lambda_1 \mathcal{L}_{\mathrm{obj}} + \lambda_2 \mathcal{L}_{\mathrm{cls}} + \lambda_3 \mathcal{L}_{\mathrm{loc}} \tag{14}$$

where $K$ is the number of CBL modules with SIB, $\alpha_i > 0, (i = 1, \ldots, K)$ is the ratios of each SIB, $\lambda_1, \lambda_2$ and $\lambda_3$ are the term weights of original YOLOv5 loss.

# 4. Experimental results

## 4.1. Genealogy dataset

The Chinese genealogy has undergone thousands of years of development, from the undefined format before the Pre-Qin period, to the simple graph format created by Sima Qian, to Ban Gu's four-generation and one-turn format, and then to the graphic transmission and separation format during the Northern and Southern Dynasties. It has gradually improved over time, and even today, new discoveries are still being made in the compilation methodology of Chinese genealogy. There are currently six common types of genealogy samples, and different family tree styles have different effects on the fitting of deep learning models. In this experiment, we mainly scanned and extracted two common styles (with and without borders on the inner pages) and manually labeled the positions of the tag boxes, recording the positions of the upper left and lower right vertices of the tag boxes relative to the images to achieve image labeling. The two styles of genealogical picture and their corresponding tag boxes are shown in Figure 3.

## 4.2. Experimental settings

The experiment was conducted on a server running Ubuntu 20.04 operating system with an Intel(R) Xeon(R) Platinum 8255C CPU and an RTX 3080 GPU with a memory size of 10GB. The training was accelerated using CUDA 11.3, and PyTorch 1.10.0 deep learning framework was used for training. The Resnet classification model and YOLOv5-6.0 prediction model were used, with an input image size of 640 × 640. The initial learning rate was set to 0.01 and the final learning rate was set to 0. The SGD optimizer had a momentum of 0.937, and the batch size for training was set to 16.

This layout detection method for Chinese genealogy image recognition and region localization, which is divided into two parts: classification and detection. The classification part employs an SIB-ResNet network optimized by the SIB theory for feature extraction, achieving high accuracy. The detection part uses an SIB-YOLOv5 model, where the SIB enhances the model's precision and computational speed compared to the original network. The specific process is shown in the Figure 4. To validate the proposed method, a set of experiments was designed for classification, detection, and overall performance testing.

Experiment 1 compares the classification performance of ResNet, VGGNet, IB-ResNet, and SIB-ResNet through a set of comparative trials. All models were trained on the same data and tested on the same test set to compare their detection speed and accuracy.

Experiment 2 focuses on the detection part and compares the computational efficiency, mAP value of the training set, and accuracy and recall of specific data sets between YOLOv5 and SIB-YOLOv5 models with the same data set and parameter settings during the training process.

Experiment 3 is a comprehensive test that utilizes a specific test set to evaluate the performance of SIB-YOLOv5 models trained on randomly scattered data sets and SIB-YOLOv5 models trained on data sets selected by the SIB-ResNet classifier. The extracted detection areas are compared with the selected positions marked in the local data set, and cosine similarity is calculated. If the similarity reaches the threshold, it is considered a successful recognition. The total number of actual images in the test set, the number of correctly classified images, the number of correct detections, the average completion time, and the overall performance are recorded and evaluated.

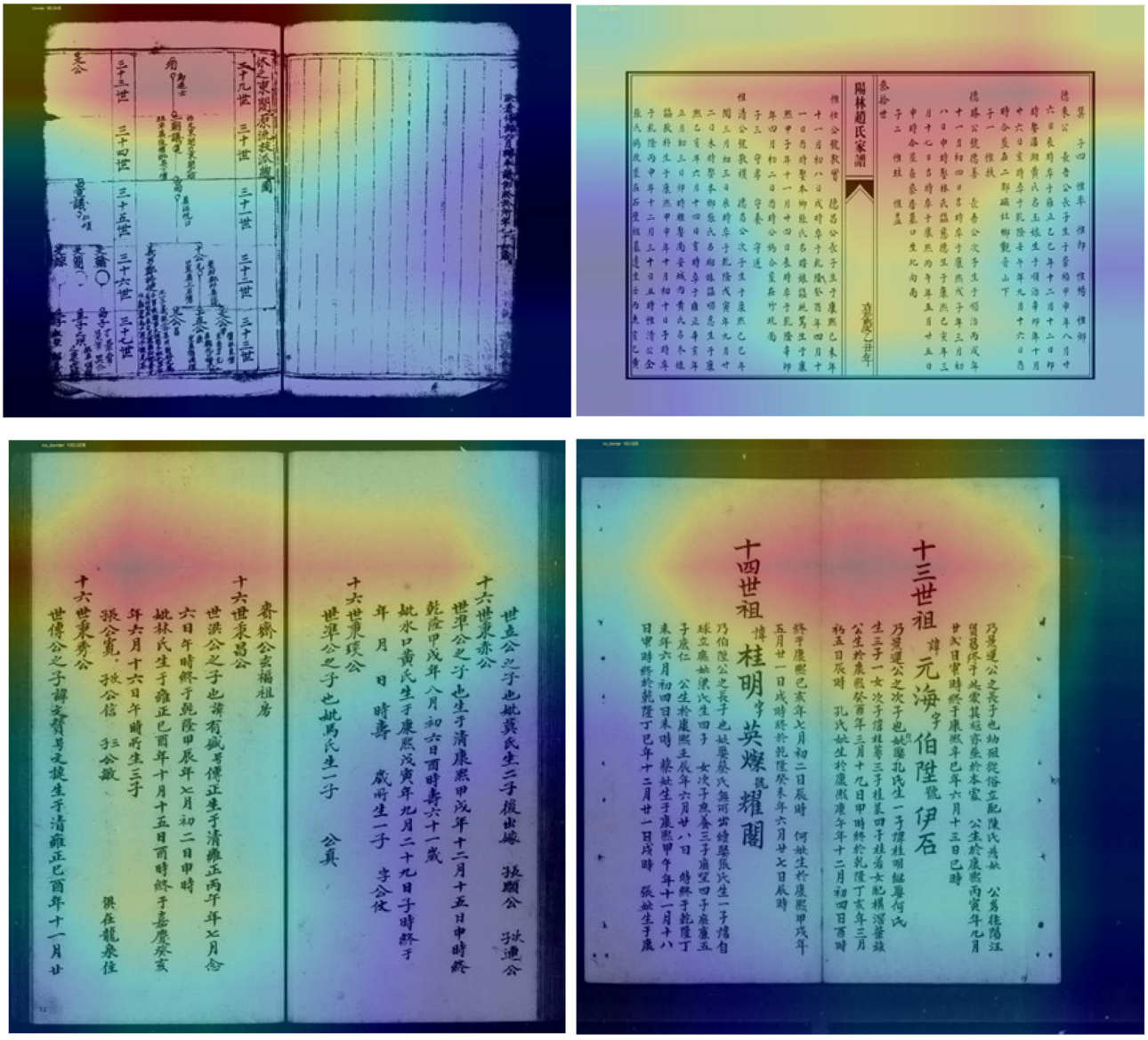

FIGURE 5
The feature heat map of the SIB-ResNet network.

## 4.3. Evaluation metrics

This paper evaluates the accuracy of the detection model using the metrics of recall, precision, average precision (AP), and mean average precision (mAP). Before introducing these metrics, the following concepts are defined: TP (true positives) refers to correctly assigned positive samples; TN (true negatives) are the correctly assigned negative samples; FP (false positives) are the incorrectly assigned positive samples; and FN (false negatives) are the incorrectly classified negative samples.

Precision: the proportion of correctly classified positive samples to all samples that the classifier identifies as positive:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{15}$$

TABLE 2 Comparison results of Experiment 2.

| Model | Weights [MB] | mAP [%] | Recall [%] | Speed [ms] |
|---|---|---|---|---|
| YOLOv5 | 41.9 | 90.6 | 93.7 | 33 |
| IB-YOLOv5 | 34.5 | 87.2 | 90.8 | 22 |
| SIB-YOLOv5 | 35.6 | 89.2 | 92.3 | 24 |

Recall: the proportion of correctly classified positive samples to all actual positive samples:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{16}$$

AP: the area under the curve formed by the precision-recall curve, where recall is on the x-axis and precision is on the y-axis,

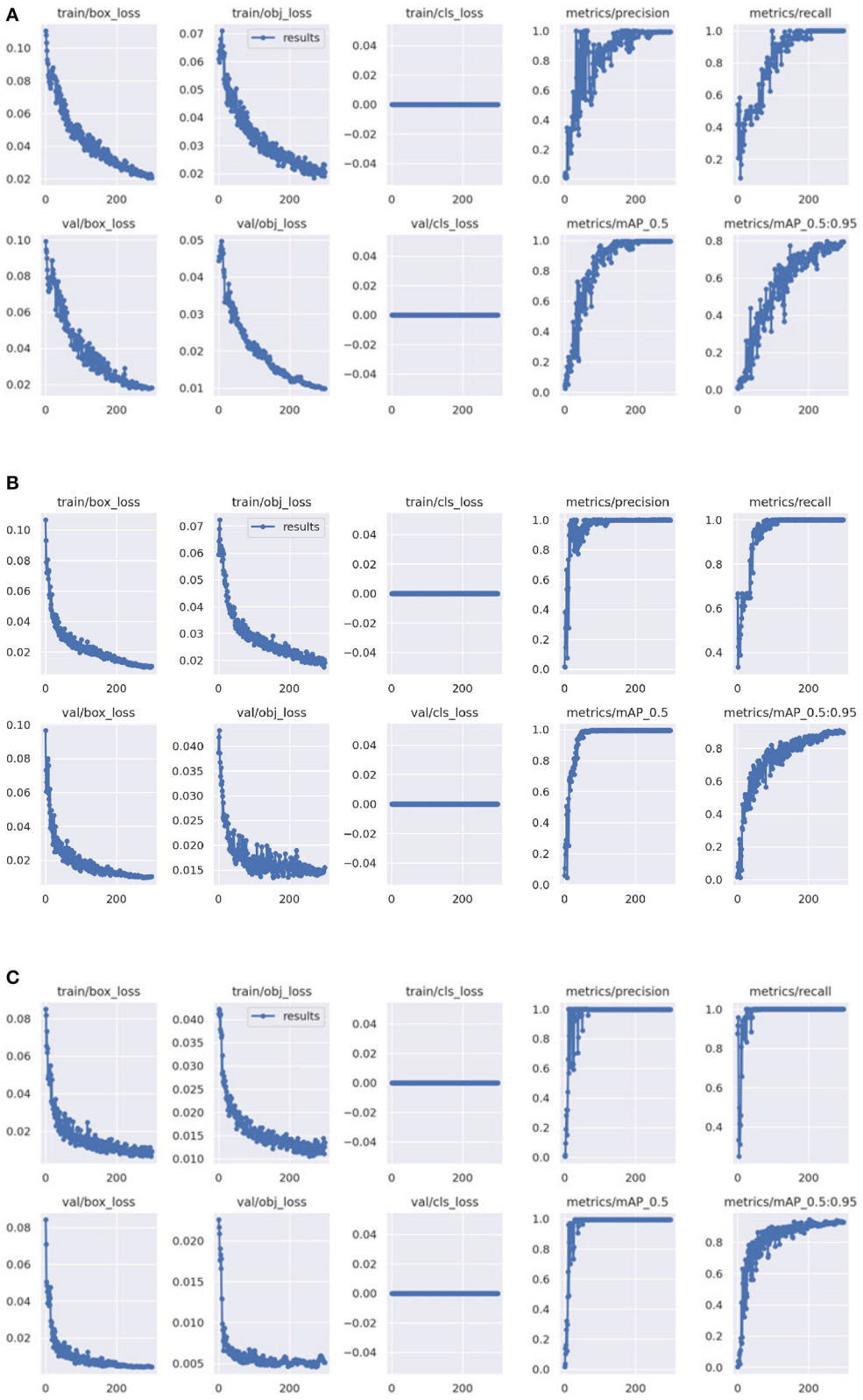


FIGURE 6
Comparison of the training loss. **(A)** The training loss of YOLOv5. **(B)** The training loss of IB-YOLOv5. **(C)** The training loss of SIB-YOLOv5.

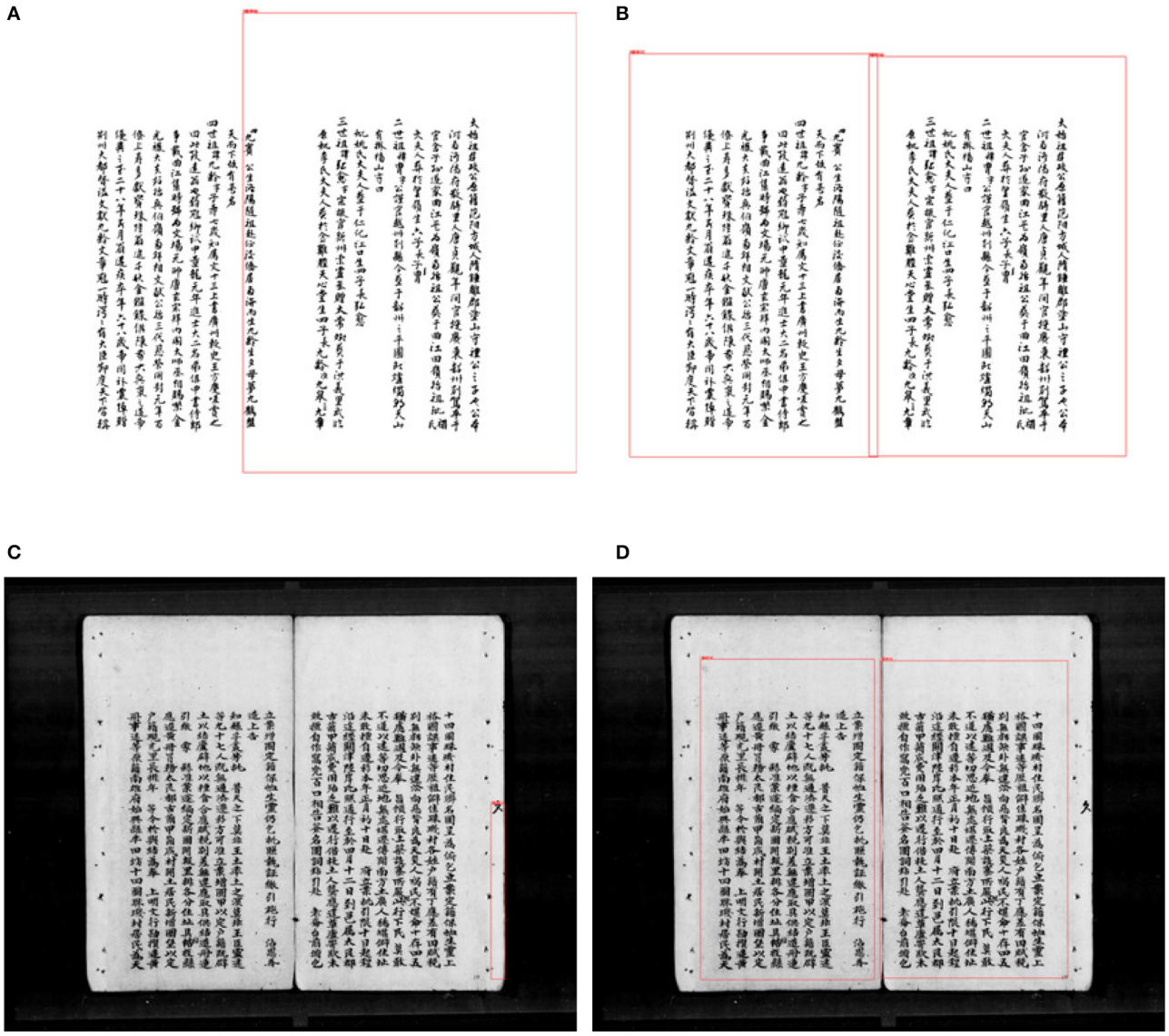


FIGURE 7
The detection results between the YOLOv5 model trained on the dataset after the SIB-ResNet classifier selection and the YOLOv5 model trained on the randomly selected dataset under different genealogy images. **(A)** Borderless pictures recognized by SIB-YOLOv5. **(B)** Borderless pictures recognized by SIB-YOLOv5 after the SIB-ResNet classifier selection. **(C)** Framed pictures recognized by SIB-YOLOv5. **(D)** Framed pictures recognized by SIB-YOLOv5 after the SIB-ResNet classifier selection.

as shown in (17):

$$\mathrm{AP} = \sum_{i=1}^{n-1} (r_{i+1} - r_i) p(r_{i+1}) \tag{17}$$

mAP: the mean average precision of all categories in the dataset, as shown in (18):

$$\mathrm{mAP} = \frac{1}{m} \sum_{i=1}^{m} (AP_i) \tag{18}$$

## 4.4. Experimental results

In Experiment 1, we compared our proposed SIB-ResNet with ResNet, GooLeNet, VGG, and conventional IB-ResNet, results can be found from Table 1. We can see that, the Resnet network model with the addition of the information bottleneck algorithm is lightweight. SIB-ResNet improves accuracy by nearly 5 percent over the equally lightweight GooLeNet. Compared with the traditional ResNet network, SIB-ResNet obtains faster recognition time while the precision and recall are not affected. SIB-ResNet model has only 4 MB more weights than the IB-ResNet model, and although it is 2 ms slower than IB-ResNet in terms of inference time, the average precision and recall improved by 4.5 and 4.2 percentage points, respectively, over IB-ResNet. VGG is slightly better than SIB-

ResNet in terms of average precision and recall, but its weights are about 6 times the weights of SIB-ResNet model, and its inference speed is slightly slower than that of SIB-ResNet.

We also presented the feature heat map of the SIB-ResNet network, as shown in Figure 5, through which we can observe which areas the network focuses on more for the purpose of correct classification. Based on the heat map, we can see that the high response areas are indeed concentrated in the border position area, which is the area that we have identified as the most helpful for making judgments, regardless of whether it is a border or borderless layout.

In Experiment 2, the same data sets and the same parameter settings were used in the training process of both YOLOv5 and SIB-YOLOv5 models to compare the computational efficiency, the mAP values of the training set, and the accuracy and recall of detection for a specific data set of the two different models, and the results were given in Table 2.

Figure 6 represents the comparison results of various evaluation metrics between YOLOv5 and SIB-YOLOv5. From the figure, it can be seen that the SIB-YOLOv5 model converges faster and has smaller loss values compared to the traditional YOLOv5 model, indicating that training the deep learning network using the introduced sublinear information bottleneck (SIB) improves the convergence ability of the network.

To better verify the feasibility of the model proposed in this paper, we designed experiment 3 and selected some images of different categories for testing, as shown in Figure 7 for the comparison of the detection results between the YOLOv5 model trained on the dataset after the SIB-ResNet classifier selection and the YOLOv5 model trained on the randomly selected dataset under different genealogy images. Figures 7A, B indicate the detection results for the borderless genealogy images, Figure 7A shows the detection results of the YOLOv5 model trained on the randomly selected dataset, and Figure 7B shows the detection results of the YOLOv5 model trained on the dataset after the SIB-ResNet classifier selection. In the prediction of borderless family tree, the classifier-optimized model is significantly better than the non-classifier-optimized model for the positioning of the borders, and the prediction frames are placed at reasonable locations. Figures 7C, D are the detection result of the genealogy image with border, it can be seen that the detection error of the model with borders mainly comes from the confusion between the borders and text, and the model trained by the classifier can solve this problem well.

Through a series of experiments, we found that the classification of genealogical images using the ResNet34 network has higher training efficiency and recognition accuracy compared to other deep learning networks, and the introduced information bottleneck theory approach enables a lighter and more adaptive model. In the target detection part of Experiment 2, we demonstrate that the SIB-YOLOv5 model shows better performance than the traditional YOLOv5 model, which has poor performance in complex and diverse family tree versions and low target localization accuracy compared to the SIB-YOLOv5 model. The results of Experiment 3 illustrate that the sublinear information bottleneck (SIB) and the two-stage deep learning approach proposed in this paper are robust to complex genealogical picture types, thus showing superior performance as well as more accurate localization accuracy. In the future, as the classifier classifies more and more categories, the model is continuously optimized to be able to perform comprehensive and accurate recognition of genealogical images.

## 5. Conclusion

In this paper, we present a novel sublinear information bottleneck (SIB) approach for genealogy layout recognition and apply it to the ResNet classifier and YOLOv5 object detection network, resulting in the SIB-ResNet and SIB-YOLOv5 models. Compared to traditional IB methods, our SIB is more effective in compressing various types of noise present in genealogy layout images, such as text overlap, low contrast, and stains, while adding minimal additional computational complexity. Our proposed method can simultaneously address the recognition of genealogy layout images with and without borders, demonstrating greater adaptability. Through a series of experimental results, we demonstrate the effectiveness of our approach, achieving excellent performance in both recognition accuracy and computational speed.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Author contributions

JY: conceptualization, methodology, validation, formal analysis, writing—original draft preparation, writing—review and editing, and visualization. QW: funding acquisition. All authors have read and agreed to the published version of the manuscript.

## Funding

This research was funded by National Key Research and Development Program of China under grant no. 2021YFE0101600.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

# References


Alemi, A., Fischer, I., Dillon, J. V., and Murphy, K. (2017). "Deep variational information bottleneck," in *International Conference on Representation Learning* (Toulon), 1–19. doi: 10.48550/arXiv.1612.00410

Amjad, R. A., and Geiger, B. C. (2020). Learning representations for neural network-based classification using the information bottleneck principle. *IEEE Trans. Pattern Anal. Mach. Intell.* 42, 2225–2239. doi: 10.1109/TPAMI.2019.2909031

Bochkovskiy, A., Wang, C. Y., and Liao, H. (2020). YOLOv4: optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*. doi: 10.48550/arXiv.2004.10934

Borges Oliveira, D. A., and Viana, M. P. (2017). "Fast CNN-based document layout analysis," in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)* (Venice), 1173–1180. doi: 10.1109/ICCVW.2017.142

Bukhari, S. S., Shafait, F., and Breuel, T. M. (2011). "High performance layout analysis of arabic and urdu document images," in *2011 International Conference on Document Analysis and Recognition* (Beijing), 1275–1279. doi: 10.1109/ICDAR.2011.257

Chang, J. (2014). The organization, research and digitization of chinese genealogical data. *J. Anhui Univ.* 38, 95–105.

Dong, Q., and He, J. (2023). Robustness enhancement method of deep learning model based on information bottleneck. *J. Electron. Inform. Technol.* 1–8.

Fan, J. (2013). The connotation of digital humanity and the deep development of digitization of ancient books. *Res. Library Sci.* 4–5.

He, K., Yao, L., Zhang, J., Li, Y., and Li, C. (2021). Construction of genealogical knowledge graphs from obituaries: multitask neural network extraction system. *J. Med. Internet Res.* 23, e25670. doi: 10.2196/25670

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*.

Juttner, A., Szviatovszki, B., Mecs, I., and Rajko, Z. (2001). "Lagrange relaxation based method for the QOS routing problem," in *IEEE INFOCOM 2001: The Conference on Computer Communications* (Anchorage), 859–868.

Kolchinsky, A., Tracey, B. D., and Kuyk, S. V. (2018). Caveats for information bottleneck in deterministic scenarios. *arXiv preprint arXiv:1808.07593*. doi: 10.48550/arXiv.1808.07593

Kosaraju, S. C., Masum, M., Tsaku, N. Z., Patel, P., Bayramoglu, T., Modgil, G., et al. (2019). "Dot-Net: document layout classification using texture-based CNN," in *2019 International Conference on Document Analysis and Recognition (ICDAR)* (Sydney), 1029–1034. doi: 10.1109/ICDAR.2019.00168

L, O. (1993). The document spectrum for page layout analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 15, 1162–1173.

Lee, K.-H., Choy, Y.-C., and Cho, S.-B. (2000). Geometric structure analysis of document images: a knowledge-based approach. *IEEE Trans. Pattern Anal. Mach. Intell.* 22, 1224–1240. doi: 10.1109/34.888708

Lee, S.-W., and Ryu, D.-S. (2001). Parameter-free geometric document layout analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 23, 1240–1256. doi: 10.1109/34.969115

Li, J., and Liu, D. (2019). Information bottleneck methods on convolutional neural networks. *arXiv preprint arXiv:1911.03722*. doi: 10.48550/arXiv.1911.03722

Li, K., Chen, Y., and Liu, J. (2022). Survey of deep learning-based object detection algorithms. *Comput. Eng.* 48, 1–12. doi: 10.19678/j.issn.1000-3428.0062725

Liu-Gong, Y., Dubuisson, B., and Pham, H. (1995). "A general analysis system for document's layout structure recognition," in *Proceedings of 3rd International Conference on Document Analysis and Recognition* (Montreal, QC), 597–600. doi: 10.1109/ICDAR.1995.601967

Owen, J. H. (2001). *Computational Combinatorial Optimization*. Springer.

Pensia, A., Jog, V., and Loh, P. (2020). Extracting robust and accurate features via a robust information bottleneck. *IEEE J. Select. Areas Inform. Theory* 1, 131–144. doi: 10.1109/JSAIT.2020.2991005

Schwartz-Ziv, R., and Tishby, N. (2017). Opening the black box of deep neural networks via information. *arXiv preprint arXiv:1703.00810*. doi: 10.48550/arXiv.1703.00810

Singh, S. S., and Karayev, S. (2021). "Full page handwriting recognition via image to sequence extraction," in *Document Analysis and Recognition – ICDAR 2021*, eds J. Lladós, D. Lopresti, and S. Uchida (Cham: Springer International Publishing), 55–69.

Song, J., Zheng, Y., Wang, J., Ullah, M. Z., Li, X., Zou, Z., and Ding, G. (2022). Multi-feature deep information bottleneck network for breast cancer classification in contrast enhanced spectral mammography. *Pattern Recogn.* 131, 108858. doi: 10.1016/j.patcog.2022.108858

Tishby, N., and Zaslavsky, N. (1999). The information bottleneck method. *arXiv preprint arXiv:physics/0004057*. doi: 10.1145/345508.345578

Tishby, N., and Zaslavsky, N. (2015). "Deep learning and the information bottleneck principle," in *Information Theory Workshop*. Jeju Island. doi: 10.1109/ITW.2015.7133169

Wang, R., and Zhang, X. (2008). From cemetery to genealogy to ancestral temple: the changing form of family bonds in Qixia, Shandong province during the Ming and Qing dynasties. *History Res.* 75–97.

Watanabe, T., Luo, Q., and Sugie, N. (1995). Layout recognition of multi-kinds of table-form documents. *IEEE Trans. Pattern Anal. Mach. Intell.* 17, 432–445.

Wieser, J., and Pinz, A. A. (1994). "Layout analysis finding text, titles, and photos in digital images of newspaper pages," in *Proceedings of the 17th Meeting of the Austrian Association for Pattern Recognition on Image Analysis and Synthesis* (Cambridge), 241–253.

Wu, X., Jiang, T., Zhu, Y., and Bu, C. (2023). Knowledge graph for China's genealogy. *IEEE Trans. Knowledge Data Eng.* 35, 634–646. doi: 10.1109/TKDE.2021.3073745

Zhong, X., Tang, J., and Jimeno Yepes, A. (2019). "PublayNet: largest dataset ever for document layout analysis," in *2019 International Conference on Document Analysis and Recognition (ICDAR)* (Sydney), 1015–1022.