International Journal of Innovative Technology and Interdisciplinary Sciences
https://journals.tultech.eu/index.php/ijitis
ISSN: 2613-7305
Volume 9, Issue 2
DOI: https://doi.org/10.15157/IJITIS.2026.9.2.1311-1343
Received: 12.04.2026; Revised: 01.06.2026; Accepted: 08.07.2026

Research Article

Deep Multipath Network Architecture for Recognizing Visual Objects in Document Images

Ali J. Abboud* , **Maather Alshaibi** , **Saad Albawi**

Department of Computer Engineering, University of Diyala, Baquba, Iraq
***ali.j.abboud@uodiyala.edu.iq**

Abstract

Automated document processing systems are crucial tools for processing large volumes of document images. The primary visual elements present in these documents include stamps, logos, and signatures, which serve various purposes. This research proposes a novel multipath architectural network for simultaneous and accurate recognition of visual objects. This architecture uses deep learning and image patches as core components through three basic paths. The image characteristics extracted from these paths are combined to achieve the final feature vector, which is subsequently processed to determine the class of each image patch. The proposed architecture is validated through experiments conducted across four scenarios using the SPODS and Tobacco 800 datasets. The comprehensive evaluations demonstrate remarkable outcomes, achieving precision, recall, and accuracy rates of (99.54%). Furthermore, the results demonstrate that visual object recognition tasks, including the identification of signatures (100%), stamps (99.10%), and logos (99.30%), are executed with exceptional accuracy. In addition, the non-visual object recognition of text objects achieves a perfect score of (100%). To sum up, the findings of these evaluations showed significant improvements in accuracy compared with competitors.

Keywords: Multipath Network; Visual Objects Recognition; Image Patches; Deep Learning, Documents.

INTRODUCTION

Written documents serve as the most effective means of conveying intended information among humans. For example, clay tablets were used in ancient times as a means of communication, whereas paper documents were used in recent times to exchange thoughts more simply [1]. The emergence of computing devices in recent decades marks the starting point for using digital documents in our daily lives. The digital revolution has proven extremely powerful in acquiring, storing, processing, and transmitting various types of digital documents [2]. The combination of recurrent neural networks (RNNs), convolutional neural networks (CNNs), and deep learning artificial intelligence (AI) algorithms into document processing systems aims to enhance their efficiency,

performance, security, and automation [3-5]. Text document reduction, language translation, summarization, generative text creation, search capabilities, change tracking, intelligent automation, and detection of misinformation are some applications of AI for analysing and comprehending document content in practical settings such as legal documents and digital archives [6-8]. Many establishments, despite these technological advancements, continue to rely heavily on large volumes of paper documents, which must be scanned, transmitted, and stored on large storage devices. These documents include letters, contracts, memos, tenders, orders, and more [9]. In the current era of AI, the need for automated archiving and retrieval systems to manage vast volumes of documents has become imperative due to the widespread use of the internet and the emergence of 6G networks [10]. In addition, locating documents in such extremely large volumes of data is a tedious task, necessitating the reduction of search time as much as possible. To expedite the identification of targeted documents, text and visual elements in document images play significant roles in the search process [11].

Main components such as stamps, logos, and handwritten signatures significantly aid in matching documents. A logo provides a unique identity to the source of a document, whereas a stamp confirms its authenticity, and handwritten signatures serve as behavioural biometrics to verify individual identity [12,13]. Ultimately, signatures, logos, and stamps are the primary visual tools for executing crucial document-processing tasks such as identification, retrieval, verification, indexing, detection, and authentication [14]. The summary of the contributions in this manuscript is:

- A *task-specific hybrid multipath feature fusion framework* (DMPN-VOR) has been proposed and designed from primary concepts to tackle the unique characteristics of the aimed challenging problem. It is used to recognize various visual objects in the complex document images. These visual elements encompass handwritten signatures, logos, and stamps.
- Four unique deep learning models based on CNNs are developed from the ground up, with each model reflecting a distinct experimental scenario (with a different number of paths). Particularly, the developed architectures do not depend on any pre-trained backbone models such as ResNet, VGG or EfficientNet. Instead, a specific CNN branch was trained and created from scratch to acquire unique representations directly from the input image data with complementary description fusion.
- An extensive series of experiments is conducted on deep learning, patch-based techniques under four distinct experimental scenarios to identify the optimal scenario for effective visual object recognition in the difficult document images. The comparison with various approaches is presented, along with an ablation study of the proposed architecture.
- The SPODS dataset is labelled by tagging each object (visual and non-visual) in the document image.

The remainder of this paper is organized as follows: the section on related work provides an overview of the visual object recognition literature. The outline of the proposed methodology and materials is described in the next section. Details of the experimental results and their discussion across four scenarios is presented in the following section. Comparative analysis results and ablation study analysis are also presented in the next sections. The conclusions and the outline of potential future research directions related to the proposed method are presented at the end of the paper at last section.

RELATED WORKS

In document processing systems, the main tasks of detection, recognition, and retrieval of data are typically automated [15]. Visual and textual objects constitute the primary components used by these systems. The commonly present visual elements in document images are logos, stamps, and signatures [16]. To start, it is essential to explain the approaches employed to verify and detect signatures in document images, which poses a significant challenge in document image processing. This difficulty is because of their broad spectrum of structures in the same class and fragmented parts scattered across different curves in the image, necessitating highly precise detection algorithms to identify them within complex backgrounds of document images [17, 18]. In addition, the manual inspection of signature samples by officials is a process that is prone to errors, is time-consuming, and requires significant effort [19, 20]. In [19] used signatures of users, recurrent neural networks (RNN) and deep learning to authenticate the identity of users. The basic shortcoming of this method is its assumption that signatures can be used directly without segmenting from image. Authors at [12] improved the recognition accuracy of a handwritten signature system using deep CNN models and incorporated augmentation methods. This approach suffers in case the signatures overlap with other components in the image. Authors in [21] introduced a novel approach for handwritten signature identification and forgery verification, which utilizes explainable deep CNNs. The limitation of this approach is that accuracy will decrease noticeably with scarcity of balanced data. Authors at [22] proposed segmentation and recognition of offline signatures using CNNs and the performance of this approach relies on the quantity of training data. Researchers at [23] suggested real-time signature recognition by employing convolutional autoencoder (CAE) and deep learning and achieved good verification results even with insufficient signature data. Authors at [17] suggested an innovative method for detecting and identifying signatures using a multiscale saliency metric. This method's effectiveness relies on the effective recognition of salient textures in document images. Authors in [24] employed image patches for the detection and verification of signature forgery. The lack of this approach is based on one class classification idea for forgeries. Authors in [25] introduced a creative approach to find visual and textual objects in document images using Adaboost classifiers and well-known feature detectors. This approach suffers from low verification accuracy for handwritten signatures and acceptable

detection accuracy for other graphical objects. Researchers at [26] proposed a new scheme to detect handwritten signatures using refinement layers and CNNs. Authors in [27] developed a method using full CNNs to detect and segment stroke-based signatures. The last two methods need large amounts of data to train CNN models. Researchers in [28] developed a creative two-phase approach to detect various visual and textual objects in document images using deep learning. To sum up, these schemes demonstrated sufficiently accurate identification and extraction of signatures in cases where there is no overlap with other elements within the document. However, their performance decreased significantly when there are overlaps with other visual and textual objects in the document.

The logo is the second important visual object primarily employed in the tasks of retrieving documents. Authors at [29] suggested using logos structures and their geometrical details in developing an innovative scheme to recognize them. The limitation of this scheme is the inability to recognize well the logos whenever overlapped with other document elements. Researchers in [30] proposed a creative approach to recognize and detect logos using local binary pattern (LBP) apparatus however its performance will suffer significantly in case of bad logo segmentation. Authors in [31] developed a new scheme to localize and identify logos in the document images using gray level co-occurrence matrix (GLCM) and curvelet transform. The limitation of this technique is relying on the accuracy of segmentation algorithms. Researchers in [29] proposed a novel architecture using optimization algorithms and geometrical reconstruction to verify logos. In [32] it has been proposed an innovative method based on the document image contours to detect and recognize logos. However, the lack of this scheme is dependence on the performance of segmentation algorithms. Authors in [33] created a novel method using speeded up robust features (SURF) algorithm to discover and identify logos in the images of documents. Additionally, authors in [34] devised a creative logo localizing and recognizing scheme by utilizing texture data and FCM clustering.

Stamp is the third significant visual component within document image. Authors at [35] proposed a new approach to verify document stamps on restricted power devices as mobile phones. This method is not tested on real mobile phone datasets hence still its performance not well verified. Researchers in [36] employed Viola-Jones object detection algorithm to create an approach to distinguish stamps in the images. The main shortcoming of this approach is that it requires to select accurately the parameters to train classifiers. Authors at [37] developed an innovative approach using part-based features to segment and identify stamps. The stamp in this scheme is recognized regardless of geometry, colour and shape of stamps however the drawback of this method is that its performance relies on the results of segmentation. In [38] it has been developed a creative scheme by utilizing a group of intriguing properties for lines, shapes and regions in the document image to detect and classify stamps. Authors in [39] proposed a new method to localize and identify stamps in the ancient depositories. It employed a generative adversarial network (GAN) to select significant features of stamps from ancient records and then recognize them using their diverse properties. The disadvantage of these last two

approaches is that their performance depends on the quality of the extracted features from the document image.

The writings or publications on the simultaneous localization and recognition of several visual objects are scarce on a par with single visual object identification. For example, Authors at [40] utilized spectral filtering schemes to design a novel approach to simultaneously verify and locate stamps and logos in the document images. Researchers in [41] suggested a creative approach using deep learning to retrieve documents based on identifying their logos and signatures. In [42] it has been employed spectral filtering to develop new scheme to discriminate graphics and text in aim to recognize logos and stamps objects. Authors in [43] devised an innovative approach using outlier contents to localize and recognize stamps and logos. In [44] it has been designed a new office dataset and proposed a novel method to localize and identify logos, stamps and signatures in the document image using this dataset. Authors at [16] devised a creative two-stages scheme to localize and recognize visual components in the documents. In same vein, authors at [15] developed a new deep learning-based method to localize logos and seals in the ancient documents.

The investigation and analysis of up-to-date approaches for visual objects recognition, it is obvious that these techniques are created mainly to identify single objects. Hence, they may not perform as effectively in identifying multiple objects simultaneously in document images. In addition, these methods are constrained by specific shapes, colours, or training datasets, whereas visual objects themselves can present a wide range of shapes and colours. By leveraging deep learning techniques and image patches, our proposed multipath architecture effectively mitigates the aforementioned challenges by eliminating the need for preprocessing steps such as binarization and layout analysis in the system pipeline. Most of the previous methods focus on detecting and recognizing individual visual objects in documents. In addition, the major limitation of these approaches is that they rely solely on global structural and geometrical relationships to localize and verify visual objects. Our research aims to address the limitations by accurately and effectively recognizing visual objects within documents by introducing special framework that integrating CNN learned features representations by two supplementary feature streams to catch domain-relevant attributes not obviously formed by convolutional filters alone (i.e. slope-based descriptors and histogram-based descriptors).

THEORETICAL FRAMEWORK, RESEARCH HYPOTHESES AND GAPS

Let set a theoretical formalization for the proposed **DMPN-VOR** as follows:

- **Input**: patch x
- **Feature streams**:
  - $fcnn(x)$
  - $fslope(x)$
  - $fhist(x)$

- **Fusion**:

$$z = \phi([\text{fcnn}(x), \text{fslope}(x), \text{fhist}(x)])$$

- **Classification**:

$$\hat{y} = \text{softmax}(Wz + b)$$

The suggested multipath architecture is created on the presumption that various feature extraction algorithms catch complementary issues of visual elements in document images. Deep CNN are highly effective at learning hierarchical and discriminative feature representations directly from image data (i.e. shape, local texture and spatial patterns). However, additional statistical and structural information can be provided by handcrafted descriptors that may not be fully extracted by learned representations alone. Primarily, and structural and directional variations are characterized by slope descriptors, whereas the statistical distribution of visual content is captured by histogram descriptors. Hence, combining these heterogeneous feature representations is supposed to produce a more informative feature scope, leading to enhanced class separation and recognition accuracy. Based on this theoretical justification, the following research hypotheses are proposed:

H1: CNN features catch discriminative local patterns that are not fully extracted by handcrafted extractors.

H2: inserting histogram-based features marginally improves classification accuracy over CNN-only baseline.

H3: Slope-based features improve classification for text and handwritten categories but not for logos/stamps visual objects.

H4: The multipath framework boosts robustness under overlapping objects in relative to single-path CNN.

These hypotheses are assessed experimentally by ablation and comparative studies, where the performance of individual feature paths is contrasted against the proposed fused representation. Despite the important advancements reached in visual object recognition and document image analysis, several challenges (or research gaps) still insufficiently approached in existing literature:

- Adequately recognizing multiple visual objects in the hard document images.
- Minimizing reliance on preprocessing functions such as layout analysis and binarization.
- Utilizing complemental statistical, local and structural information concurrently.
- Keeping robustness under hard conditions including clutter, image degradation, occlusion and overlap.

METHODS AND MATERIALS

This section elucidates the proposed novel method of visual object recognition in the document images. Figure 1 illustrates the proposed multipath network architecture, which comprises three primary paths for recognizing individual image patches and classifying

them into one of four categories: signature, logo, stamp, or text. The first path uses a deep convolutional neural network to identify salient textures (features) in image patches and classify them. In the second path, the slopes of lines in the image patches serve as supplementary information to determine whether or not a patch corresponds to a handwritten signature, or text. The third path utilizes histogram of patch information to differentiate logos and stamps from other visual objects. These paths are subsequently concatenated to recognize the visual objects (logo, stamp, or/and signature) of a labelled document (or documents) with ID(s) of (1…n) in a document image database. The subsequent sections explain this architecture in detail. Additionally, Figure 2 in the subsection Path 2 (Slope Path) illustrates thoroughly the slope localization process of this architecture.

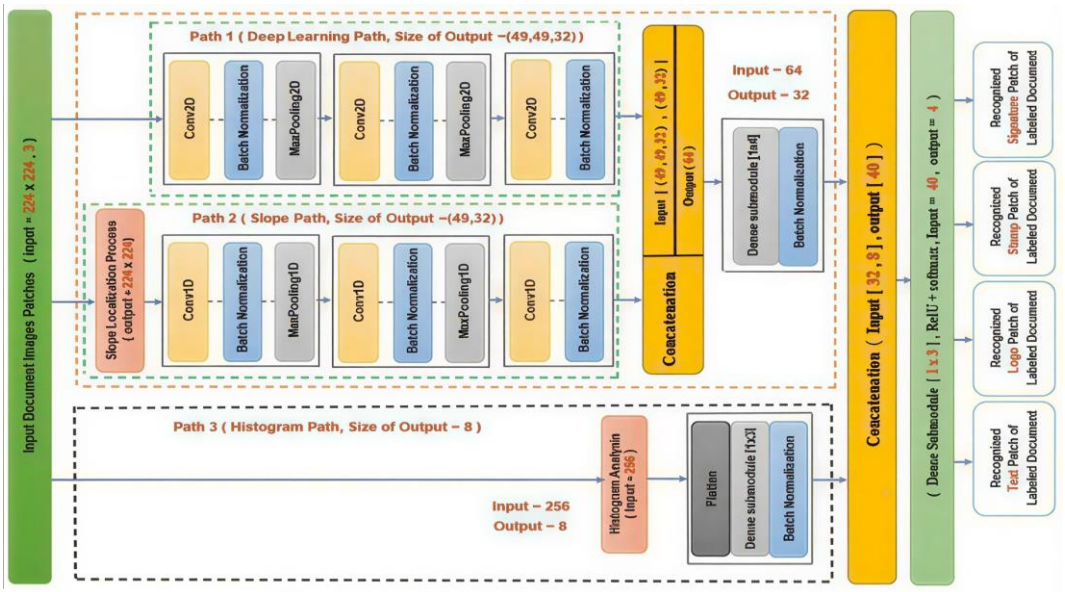

Figure 1. Structure of the deep multipath network for recognizing visual objects (**DMPN-VOR**) in the labelled document images

Path 1 (Deep Learning Path)

In the proposed architecture, deep convolutional neural networks serve as the foundational framework for identifying meaningful features in image patches and subsequently classifying them into one of the four categories. The deep convolutional neural network topology comprises multiple layers, each assigned a specific function that is crucial to the network's overall performance, as illustrated in Figure 1. An image patch of size (224 × 224 × 3) is fed to Path 1, and the output from this path is the feature vector of size (49 × 49 × 32).

Path 2 (Slope Path)

The second path in the proposed deep multipath architecture is designed to obtain the slopes of lines within visual objects. The slopes in the geometry and structure of handwritten signatures, for example, can be subsequently used for their recognition. In

addition, these slopes can be used to identify other visual objects such as stamps and logos. Although the slopes of lines are certainly beneficial for identifying textual objects, these lines can sometimes appear similar but with differing angles of rotation. Hence, recognizing such lines in image patches using only a deep learning path is challenging, and an additional slope-localization process is required to address object similarity. An image patch of size (224 × 224 × 3) is fed into the slope localization process shown in Figure 2 and then converted to grayscale (224 × 224).

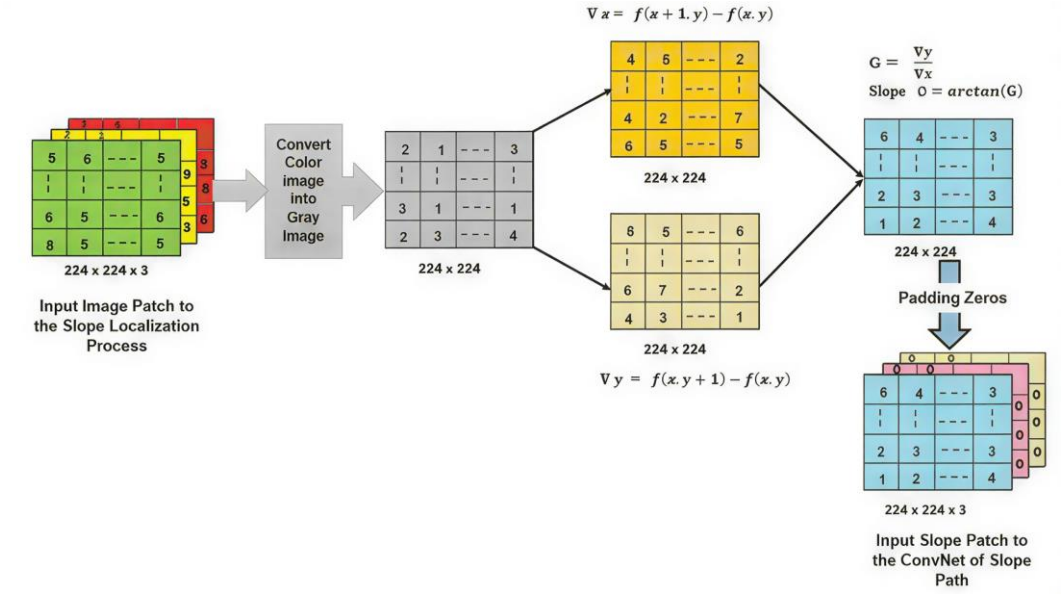

Figure 2. The slope localization process in the deep multipath network architecture (DMPN-VOR)

Afterword it is processed by gradient operators in both X and Y directions using equations (1) and (2) to get all changes in the lines within the image patch. Next, these gradient images are used to compute a gradient-magnitude image patch using equations (3), thereby obtaining the net change in lines within the patch. Then the resultant image is employed to calculate the slope image of size (224 x 224) using equation (4). This final array of angles in the range [-π, π] represents the slope of all lines within image path measured by radians. Additionally, the slope image is of size (224 x 224x3) processed through the slope path to obtain the desired features.

Gradient in X-direction $\nabla x = f(x+1, y) - f(x, y)$ (1)

Gradient in Y-direction $\nabla y = f(x, y+1) - f(x, y)$ (2)

Gradient Magnitude $G = \frac{\nabla y}{\nabla x}$ (3)

Slope $\emptyset = arctan(G)$ (4)

Path 3 (Histogram Path)

The image histogram serves as a significant tool for illustrating key characteristics within a document image. As certain visual segments are challenging to differentiate, the existing deep and slope features lack the necessary strength for effective discrimination. To overcome this limitation, a third histogram analysis path is incorporated into the proposed architecture. This enhancement is crucial as it encompasses distinctive patch features that can be utilized for the identification of specific categories of document image patches. The probabilities of gray levels are combined into a single feature vector of size (256) for a particular patch in the image document, which is subsequently fed into the multipath architecture to yield an output feature vector of size (8) from this path. The patch histogram is computed using equation (5).

$$p_r(r_k) = \frac{\mathrm{n_k}}{\mathrm{n}} \qquad \text{where } k = 0,1,2 \dots \dots, L-1 \tag{5}$$

p_r: gray level (r) probability of the in the image patch.

r_k: gray level (k).

n_k: number of the gray level (k) occurrences of in the image patch.

n: total count of pixels of the image patch.

L: count of distinct gray levels of the image patch.

A detailed summary of the proposed Deep Multipath Network architecture (DMPN-VOR) and its workflow for recognizing visual objects in labelled document images is presented as follows:

Step 1. Input document image patches (Ip) of size (224x224,3) of documents (1…...n).

Step 2. Do slope localization process on (Ip) to find slopes of lines within visual and text objects in the document image where and output slope patches (S1) of size = (224x224x3) as below:

Step 2.1. Convert colour image into grayscale image.

Step 2.2. Compute gradient in X-direction $\nabla x = f(x+1,y)-f(x,y)$.

Step 2.3. Compute gradient in Y-direction $\nabla y = f(x,y+1)-f(x,y)$.

Step 2.4. Compute gradient magnitude is $G = \nabla y/\nabla x$

Step 2.5. Compute slope $\emptyset = \arctan(G)$.

Step 3. Do histogram analysis process to find the frequency of distinct pixels within input image patches and output histograms (H1) of size = (256).

Step 4. Send simultaneously image patches (Ip, S1 and H1) to Path 1 (Deep Learning, Ip), Path 2 (Slope, S1) and Path 3 (Histogram, H1) respectively.

- Use (Ip) by deep learning path (Path1) to derive vector of features of size (49x49,32).
- Use (S1) by slope path (Path2) to generate vector of features of size (49x49,32).
- Use (H1) by histogram analysis process in the path (Path 3) to produce vector of features of size (8).

Step 5. Concatenate feature vectors produced by Path 1 and Path 2 to get concatenated feature vector (C1) of size (98x49,32).

Step 6. Reduce the size of concatenated feature vector (C1) from (98x49,32) to (32) to get reduced feature vector (R1).

Step 7. Concatenate feature vector R1 of size (32) with the output feature vector from Path3 of size (8) to obtain the final feature vector (C2) of size (40).

Step 8. Do classification process by utilizing C2 to obtain the type of image patches (Logo, Stamps, Signatures, Text) within labeled document images.

RESULTS AND DISCUSSION

The effectiveness and reliability of the proposed architecture (DMPN-VOR) for the simultaneous recognition of handwritten signatures, logos, stamps, and text were evaluated through series of experiments conducted on two widely used datasets. The subsequent sections elaborate on the experimental procedures, datasets, and evaluation of the experimental outcomes. Note that for the experiments, no preprocessing operations were accomplished on the images from the discussd datasets below, and the introduced deep learning-based models were designed from scratch using the two datasets.

Spods Dataset

The scanned pseudo-official data set (SPODS) is a complex collection of various document classes [44], freely and publicly available to the research community. It has been designed to evaluate the simultaneous identification of visual objects (stamps, logos, handwritten signatures) and text in the official document images. It is composed of 1088 images of different types of documents (12 documents, 32 signatures, 32 logos, and 32 stamps) as shown in Table 1. Additionally, there are 11 font types in the SPODS dataset, with various font sizes (10, 11, 11.5, 12, 14). All document images were created using Microsoft Office. Two ground truth sets (G1 and G2) were used as references to evaluate performance. G1 contains binary document images used to calculate detection performance metrics, whereas G2 contains metadata information in an Excel file for logos, signatures, stamps, and text. Figure 3 shows some examples of visual objects from the images in this dataset.

Figure 3. Visual object examples of the SPODS dataset

Table 1. Kinds of documents in the SPODS dataset

Document Type	Definition	Count of
1	Invitation for tender pre-qualification	32
2	Call for bid	32
3	Contract	32
4	Tender	32
5	Employment advertisement	32
6	memo	32
7	Letter (copy to many correspondence)	128
8	Letter (one to one correspondence)	288
9	Circular	224
10	Notice	96
11	Order	64
12	Table	96
Total		1088

Tobacco 800 Dataset

The Tobacco 800 dataset created by the Illinois Institute of Technology is a collection of complex document images [45, 46] that is freely and publicly available to the research community at Kaggle. It consists of 1290 images representing various types of documents, designed for the academic research of document analysis and processing. These images were obtained from 42 million document pages from tobacco companies using different scanning devices, and they are used as benchmark images. Each document image contains three categories of visual objects (signatures, logos, and stamps) of varying sizes and shapes, as well as text. The dataset includes multiple page images that are sequentially identified, making it a powerful tool for evaluating various document processing tasks. The images in this subset have a resolution ranging from 150 to 300 DPI, with dimensions of (1200 × 1600) pixels (height) and (2500 × 3200) pixels (width). A labeled ground-truth version of this dataset, created by humans, is also available to the public. Figure 4 provides examples of visual objects within the document images of this dataset.

Figure 4. Visual object examples of the Tobacco 800 dataset

Experimental Setup and Protocol

The experiments were conducted using a laptop with the characteristics: OS (Microsoft Windows 11 Home), Microprocessor (13th Gen Intel(R) Core (TM) i9-13900HX), Graphics Card (NVIDIA GeForce RTX 4070 Laptop GPU (8.0 GB) / Intel(R) UHD Graphics (1.0 GB)), and memory (8 GB and 32 GB). The deep learning libraries employed include TensorFlow, Keras, and scikit-learn. Our experimental protocol is based on dividing the used datasets into training (80%) and testing (20%) subsets. Good results are obtained using this protocol for both SPODS and Tobacco 800 datasets. Hence, the initial values of hyperparameters of the deep learning models in (Table 2) are used in our experiments.

Table 2. Hyperparameter settings in the deep learning model in the proposed architecture

	Hyperparameter	Setting
1	Learning Rate	0.001
2	Optimizer	Adam
3	Epochs	30
4	Batch Size	32
5	Patch Size	224 x 224
6	Loss Function	categorical_crossentropy

Performance Measures

The effectiveness of a visual document object recognition system is assessed using various established metrics, including the amounts of false positives (FP), false negatives (FN), true positives (TP), actual values (y_i), predicted values (y^) and sample size = N. All these metrics are important and each represents one aspect of the proposed model performance. Hence, one metric is insufficient to give us clear understanding of the model classification performance and always there is a need to use multiple metrics.

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$F1\ score = 2\ x\ \frac{(Precision\ x\ Recall)}{(Precision + Recall)} \tag{8}$$

$$Accuracy = \frac{(number\ of\ correct\ predictions)}{(total\ number\ of\ predictions)}\ x\ 100\% \tag{9}$$

$$Mean\ Absolute\ Error\ (MAE) = \frac{1}{N}\sum_{i=1}^{N}|y_i\ -\ \widehat{y_i}| \tag{10}$$

$$Mean\ Squared\ Error\ (MSE) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \widehat{y_i})^2 \quad (11)$$

Experimentations were tested to demonstrate the effectiveness of the introduced architecture (DMPN-VOR) for simultaneous recognition of all visual objects. Four distinct scenarios were applied to identify the most effective scenario for visual object recognition. A comprehensive discussion of these cases is provided in the subsections that follow.

First Scenario (Using Only Patches and Deep Learning)

In this series of experiments, only patches and deep learning techniques were employed in the proposed method. The results of recall, accuracy, precision, loss, MAE, MSE, and confusion matrix are illustrated in Figure 5 and Tables 3 and 4. The findings indicate high accuracy in recognizing all objects present in the document images. Furthermore, the error rates, as measured by MAE and MSE, were notably low, demonstrating the system's ability to identify signatures, stamps, logos, and text. Nevertheless, Table 3 suggests that the detection rates for logos and stamps were inferior to those for handwritten signatures and text, with a precision of (91.17) for logos and a recall of (90.27) for stamps. These results imply the need for further enhancement of object recognition accuracy in document images. For the Tobacco 800 dataset, the recall for signatures was (96.72), and the precision for logos was (97.14), with MSE and MAE values of (01.43) and (01.09), respectively. These findings indicate that the proposed method achieved a satisfactory level of accuracy in recognizing signatures and logos; however, there remains substantial potential for improvement.

Table 3. Total recognition performance of Loss (L), Accuracy (A), MAE (MA), MSE (MS), Precision (P) and Recall(R) metrics in S1

Dataset	L (↓)	A (↑)	MA (↓)	MS (↓)	P (↑)	R (↑)
SPODS	07.82	97.11	01.08	01.83	97.22	97.11
Tobacco	07.34	98.45	01.09	01.42	98.45	98.45

Table 4. Visual object recognition performance of Recall (R), Precision (P) and F1-Score (F1) measures in S1.

Dataset	Visual Object	P (↑)	R (↑)	F1 (↑)
SPODS	Signature	99.57	99.57	99.57
	Stamp	98.48	90.27	94.20
	Logo	91.17	100.00	95.38
	Text	100.00	98.46	99.22
Tobacco	Signature	100.00	96.72	98.33
	Logo	97.14	100.00	98.55

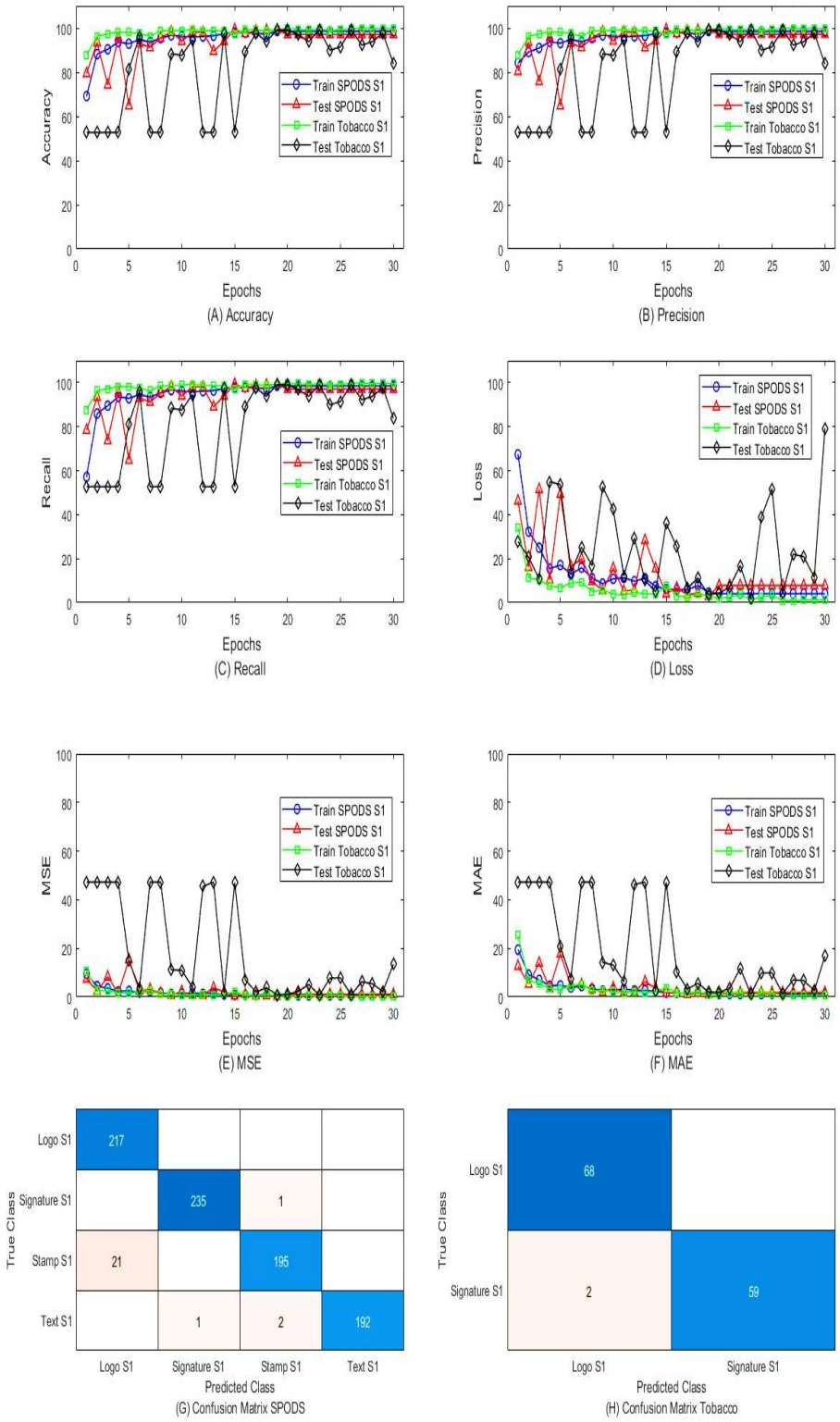

Figure 5. The results of the First Scenario S1. (A) accuracy, (B) precision, (C) recall, (D) loss, (E) MSE, (F) MAE, (G) confusion matrix SPODS and (H) confusion matrix Tobacco.

Second Scenario (Using Patches, Deep Learning and Histograms)

A group of experiments is executed in this scenario to test the accuracy of the developed approach. The deep learning algorithms, image patches and the histogram of patches are integrated to form the approach in this scenario. Figure 6 and Tables 5 and 6 show the results of experiments in terms of precision, accuracy, loss, MAE, MSE and confusion matrix. The findings denote that the recognition accuracy for logos, stamps, signatures and text existing in the document image is prominently high. Moreover, the noticed error rates are the smallest permissible. Such an impressive accomplishment is attributed to the discriminative information contained in the histograms of image patches.

These histograms can be a useful tool for adequately discriminating between different objects in the documents. Hence, these results reveal the superiority of combining patch histograms within the multi-path architecture, leading to a significant improvement in visual object recognition in documents. We can notice from the results of the Tobacco 800 dataset that the recall measure for the signature is (98.36), whereas the precision measure for the logo is (98.55). Furthermore, the overall accuracy, loss, MSE and MAE values are (99.23) and (02.55) and (01.23) and (00.79) respectively.

These outcomes of the Tobacco 800 dataset exhibit the strength and adequacy of the introduced architecture in recognizing logos and signatures.

Table 5. Total recognition performance expressed by Loss (L), Accuracy (A), MAE (MA), MSE (MS), Precision (P) and Recall(R) metrics in S2

Dataset	L (↓)	A (↑)	MA (↓)	MS (↓)	P (↑)	R (↑)
SPODS	01.75	99.54	00.23	00.66	99.54	99.54
Tobacco	02.55	99.22	00.79	01.22	99.22	99.22

Table 6. Visual object recognition performance expressed by Recall (R), Precision (P) and F1-Score (F1) measures in S2

Dataset	Visual Object	P (↑)	R (↑)	F1 (↑)
SPODS	Signature	100.00	100.00	100.00
	Stamp	100.00	90.27	94.20
	Logo	98.19	100.00	99.09
	Text	100.00	100.00	100.00
Tobacco	Signature	100.00	98.36	99.17
	Logo	98.55	100.00	99.27

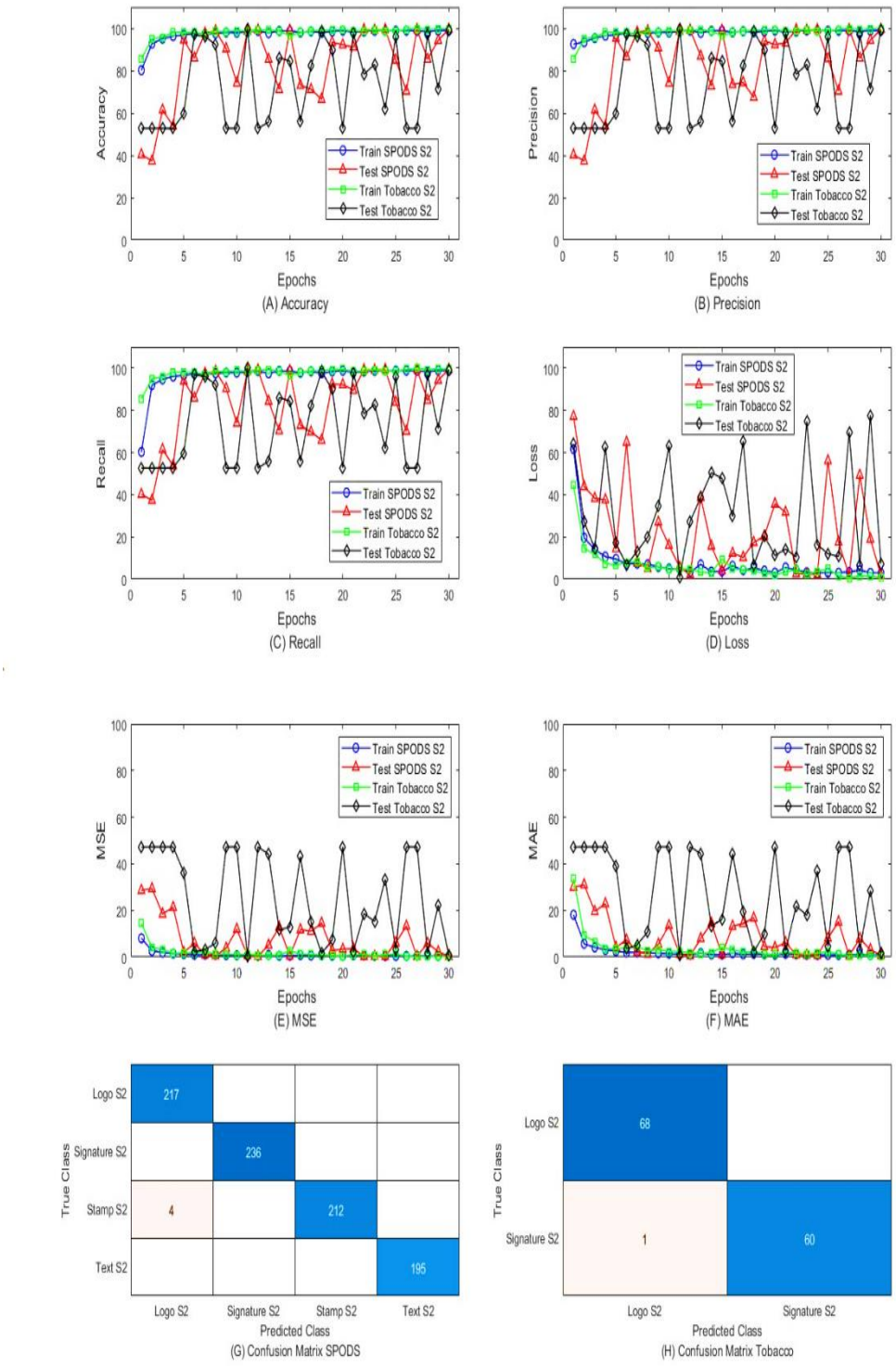

Figure 6. The results of the Second Scenario **S2**. (A) accuracy, (B) precision, (C) recall, (D) loss, (E) MSE, (F) MAE, (G) confusion matrix SPODS and (H) confusion matrix Tobacco.

Third Scenario (Using Patches, Deep Learning, And Slopes)

In the third set of experiments, image patches, deep learning techniques, and the slopes of lines within these patches were utilized to recognize visual objects in document images., Recall, loss, accuracy, precision ,MAE, MSE metrics, and the confusion matrix are used to illustrate the results in this scenario, as shown in Figure 7, as well as Tables 7 and 8. The findings reveal that the detection accuracy for all objects within the document image including signatures, logos, stamps, and text was inferior to that indicated in the last two scenarios. Furthermore, the error rates were the highest among all scenarios. However, the system was still able to recognize all signatures, stamps, logos, and text, albeit with lower accuracy than in the earlier cases. This diminished performance can be attributed to the limited capability of line slopes within the image patches to distinguish between the various objects in the document images. The precision for logos was (83.46) and the recall for stamps was (81.48). These results confirm that incorporating slope features into the proposed methodology does not improve object recognition accuracy. For the Tobacco 800 dataset, the recall for signatures was (88.52), while the precision for logos was (90.66). The MSE and MAE were (06.13) and (04.98), and the overall accuracy and loss were (94.57) and (30.00), respectively.

These results suggest that the introduced method can robustly recognize signatures and logos with satisfactory accuracy, although there remains significant potential for improvement.

Table 7. Total recognition performance in terms of Loss (L), Accuracy (A), MAE (MA), MSE (MS), Precision (P) and Recall(R) metrics in S3

Dataset	L (↓)	A (↑)	MA (↓)	MS (↓)	P (↑)	R (↑)
SPODS	23.80	94.91	02.27	02.66	94.90	94.79
Tobacco	30.00	94.57	04.98	06.14	94.57	94.57

Table 8. Visual object recognition performance expressed by Recall (R), F1-Score (F1) and Precision (P) measures in S3

Dataset	Visual Object	P (↑)	R (↑)	F1 (↑)
SPODS	Signature	100.00	98.31	99.15
	Stamp	100.00	81.48	89.79
	Logo	83.46	100.00	90.98
	Text	99.49	100.00	99.74
Tobacco	Signature	100.00	88.53	93.91
	Logo	90.66	100.00	95.11

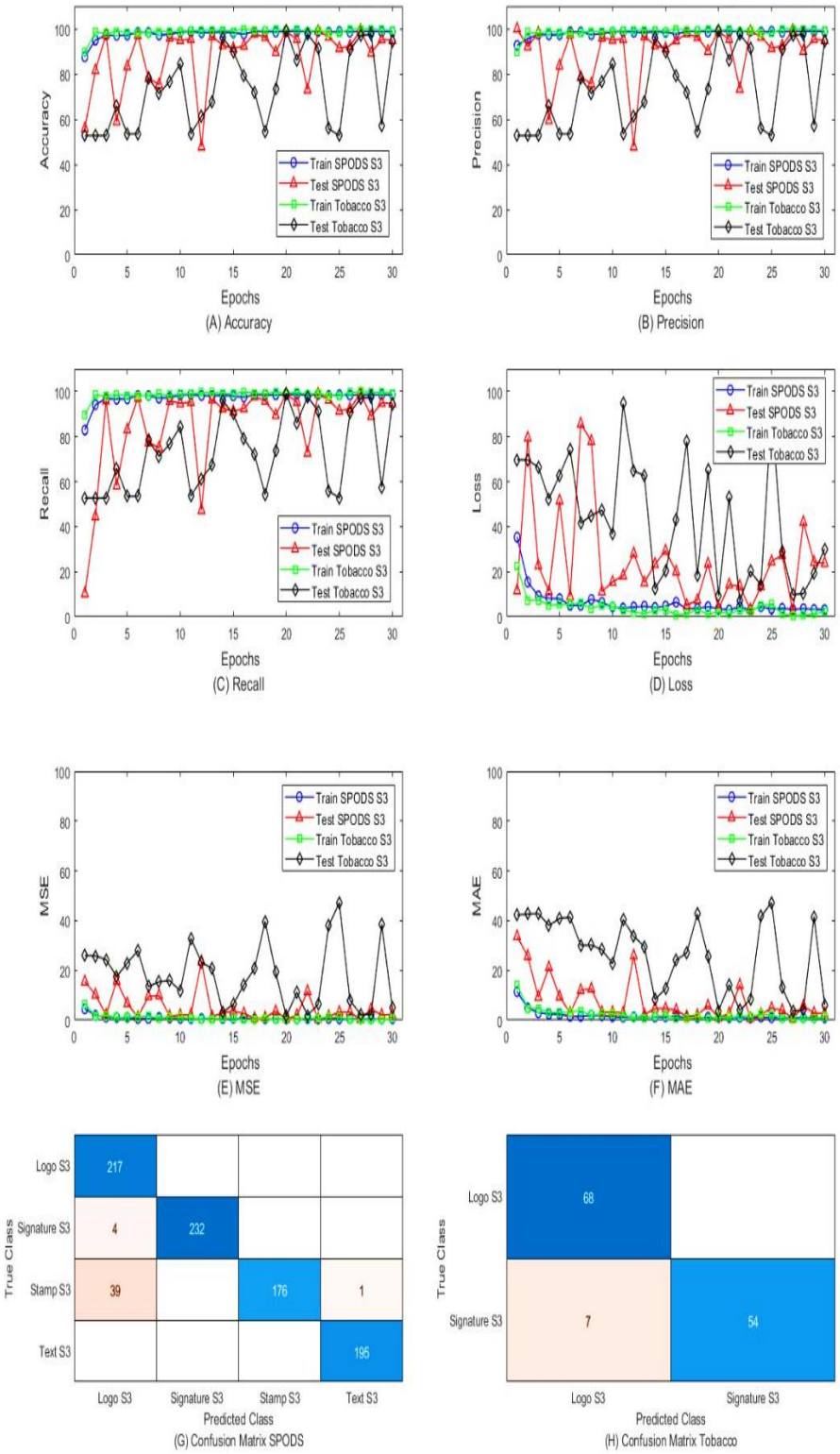

Figure 7. The results of the Third Scenario **S3**. (A) accuracy, (B) precision, (C) recall, (D) loss, (E) MSE, (F) MAE, (G) confusion matrix SPODS and (H) confusion matrix Tobacco.

Fourth Scenario (Using Patches, Deep Learning, And Slopes)

In the fourth set of experiments, patches, line slopes within image patches, histograms of image patches, and deep learning methodologies were employed to identify different visual objects in document images. The recall, loss , accuracy, precision, , MAE, MSE, and confusion matrix are used to illustrate results as shown in Figure 8 and Table 9 and 10. The findings indicate a notably effective recognition of handwritten signatures and logos; however, the recognition rates of text and stamps were comparatively lower, with precision for text recorded at (84.78) and recall for stamps at (81.94). The error rates were higher compared to the previous scenarios, with MAE at (02.06) and MSE at (02.83). This decline in overall detection performance can be attributed to the additional information from line slopes, which may have contributed to increased error rates. Thus, the inclusion of these features is indicated to adversely affect the proposed method's effectiveness in recognizing stamps and text. For the Tobacco 800 dataset, the recall for signatures was (95.08), while the precision for logos was (95.71), with MSE and MAE values being (04.52) and (02.21), respectively. The overall accuracy and loss for this scenario were (96.89) and (06.80), respectively.

These results suggest that although the proposed method demonstrates a satisfactory ability to recognize signatures and logos, there remains significant potential for enhancement.

Table 9. Total recognition performance in terms of Loss (L), Accuracy (A), MAE (MA), MSE (MS), Precision (P) and Recall(R) metrics in S4

Dataset	L (↓)	A (↑)	MA (↓)	MS (↓)	P (↑)	R (↑)
SPODS	23.79	95.25	02.07	02.83	95.25	95.14
Tobacco	06.80	96.89	02.21	04.51	96.89	96.89

Table 10. Visual object recognition performance delivered by F1-Score (**F1**), Precision (**P**), and Recall (**R**) measures in **S4**

Dataset	Visual Object	P (↑)	R (↑)	F1 (↑)
SPODS	Signature	99.57	99.57	99.57
	Stamp	99.44	81.94	89.85
	Logo	98.18	99.53	98.86
	Text	84.74	100.00	91.76
Tobacco	Signature	98.31	95.08	96.66
	Logo	95.71	98.53	97.10

A summary of the above results for the previously explained scenarios of the proposed multipath architecture is shown in Tables 11 and 12 and Figures 9 and 10 below to aid understanding of our experimental results.

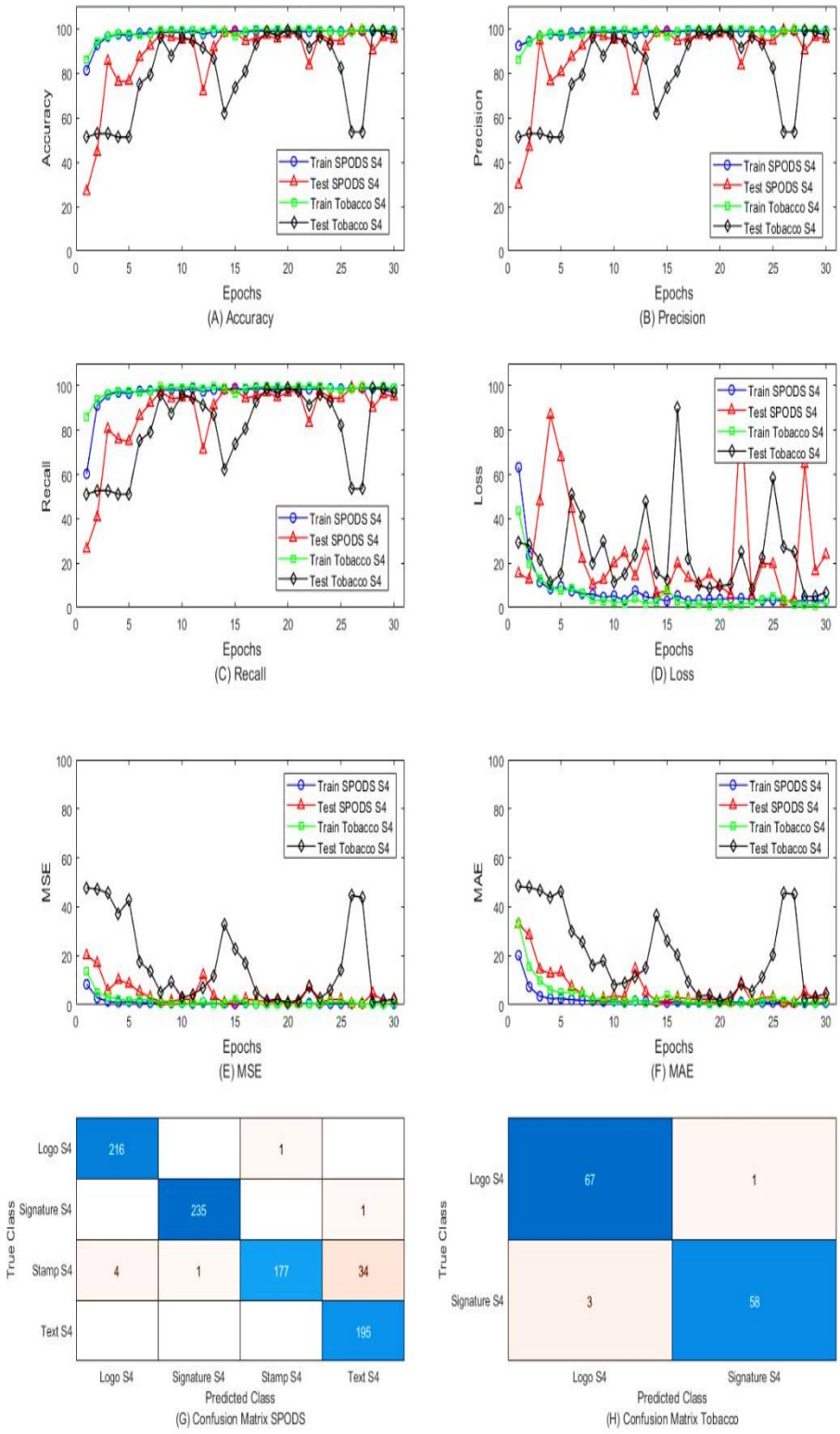

Figure 8. The results of the Fourth Scenario S4. (A) accuracy, (B) precision, (C) recall, (D) loss, (E) MSE, (F) MAE, (G) confusion matrix SPODS and (H) confusion matrix Tobacco.

Table 11. Summary table comparing the recognition performance in the Four scenarios in terms of (Loss, Accuracy, MAE, MSE, Precision and Recall) metrics

Dataset	Scenario	L (↓)	A (↑)	MA (↓)	MS (↓)	P (↑)	R (↑)
SPODS	S1	07.82	97.11	01.08	01.83	97.22	97.11
	S2	01.75	99.54	00.23	00.66	99.54	99.54
	S3	23.80	94.91	02.27	02.66	94.90	94.79
	S4	23.79	95.25	02.07	02.83	95.25	95.14
Tobacco	S1	07.34	98.45	01.09	01.42	98.45	98.45
	S2	02.55	99.22	0.79	01.22	99.22	99.22
	S3	30.00	94.57	04.98	06.14	94.57	94.57
	S4	06.80	96.89	02.21	04.51	96.89	96.89

Table 12. Summary table comparing visual objects recognition performance in the Four scenarios shown by Recall, F1-score and Precision metrics

Dataset	Visual Object	S1	S2	S3	S4
		Precision (↑)			
SPODS	Signature	99.57	100.00	100.00	99.57
	Stamp	98.48	100.00	100.00	99.44
	Logo	91.17	98.19	83.46	98.18
	Text	100.00	100.00	99.49	84.78
Tobacco	Signature	100.00	100.00	100.00	98.31
	Logo	97.14	98.55	90.66	95.71
		Recall (↑)			
SPODS	Signature	99.57	100.00	98.31	99.57
	Stamp	90.27	98.15	81.48	81.94
	Logo	100.00	100.00	100.00	99.53
	Text	98.46	100.00	100.00	100.00
Tobacco	Signature	96.72	98.36	88.53	95.08
	Logo	100.00	100.00	100.00	98.53
		F1 Score (↑)			
SPODS	Signature	99.57	100.00	99.15	99.57
	Stamp	94.20	99.07	89.79	89.85
	Logo	95.38	99.09	90.98	98.86
	Text	99.22	100.00	99.74	91.76
Tobacco	Signature	98.33	99.17	93.91	96.66
	Logo	98.55	99.27	95.11	97.10

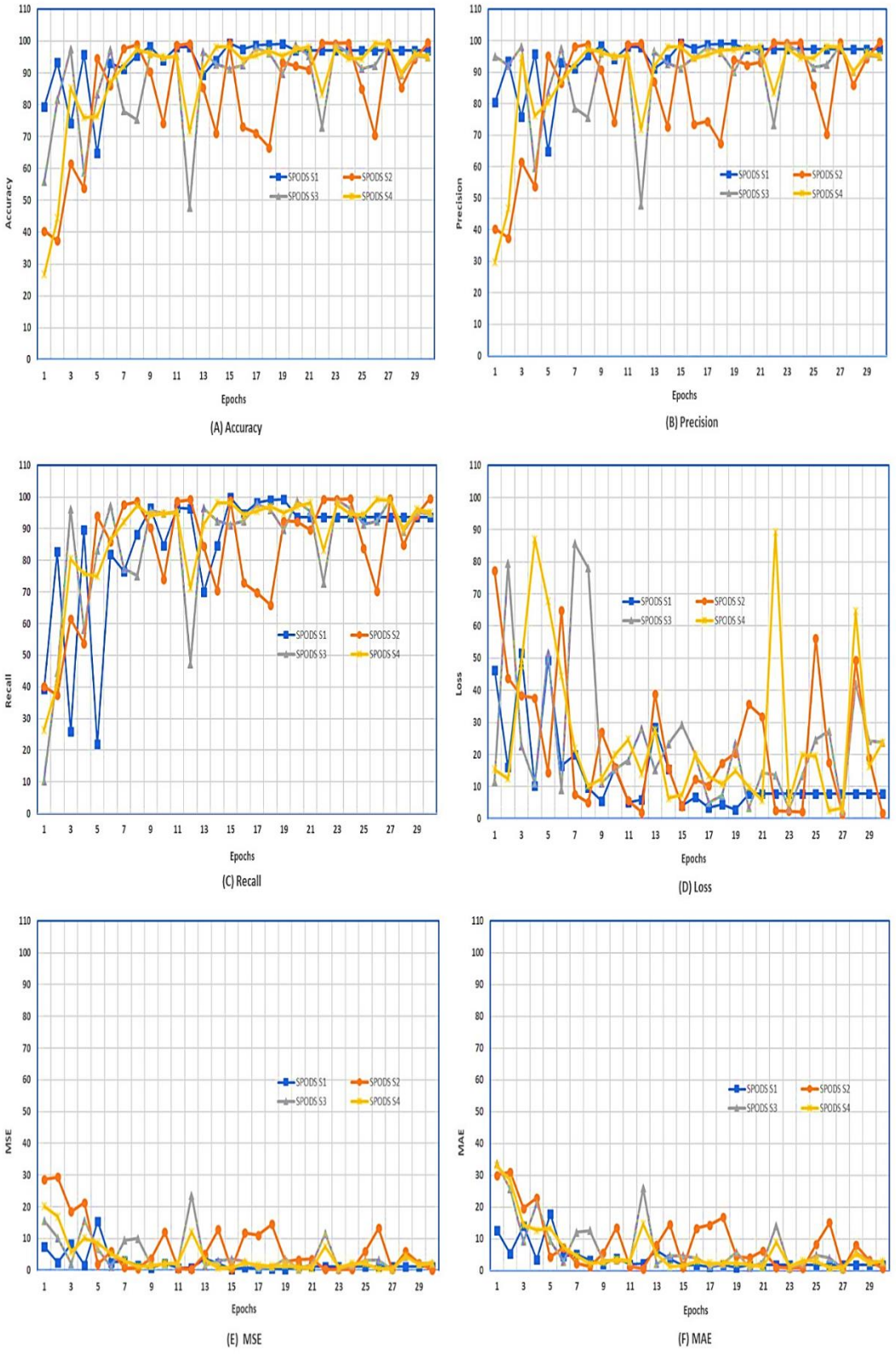

Figure 9. The charts comparing the performance of Four scenarios for SPODS dataset, using (A) accuracy, (B) precision, (C) recall, (D) loss, (E) MSE, (F) MAE

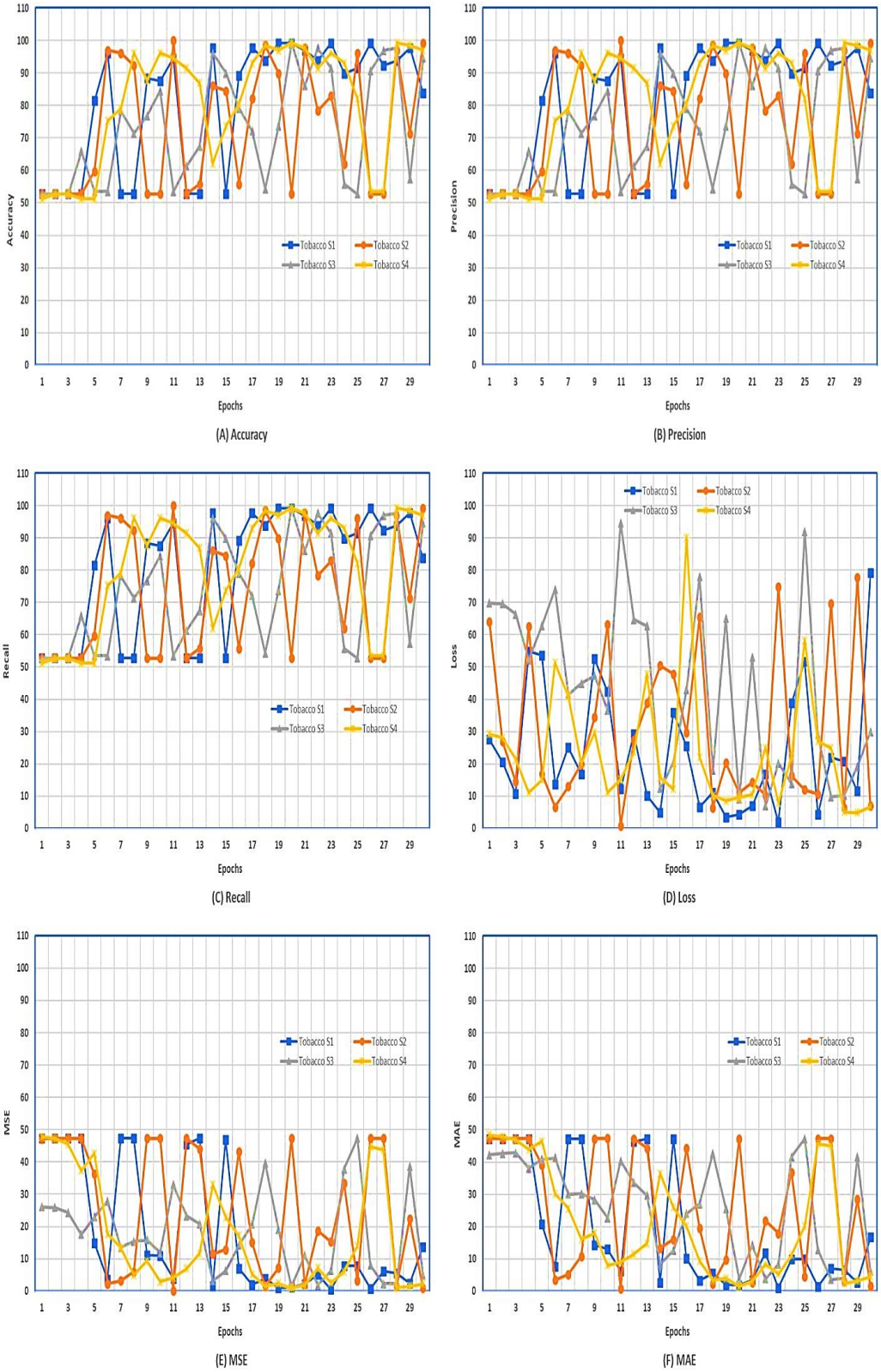

Figure 10. The charts comparing the performance of Four scenarios for Tobacco dataset, using (A) accuracy, (B) precision, (C) recall, (D) loss, (E) MSE, (F) MAE

The experimental outcomes demonstrate that histogram descriptors exceed slope descriptors whenever utilized separately. Such behaviour is attributed to the different types of information extracted by the two feature representations. Histogram descriptors distinguish the statistical dispersion of visual information within an image patch, forming them relatively insensitive to local, geometric distortions, partial occlusion, variations in object shape and noise. In contrary, slope descriptors depend on structural and directional continuity, which can be essentially influenced by overlapping objects, cluttered backgrounds cluttered backgrounds, irregular object boundaries and degraded document quality.

Consequently, histogram representations provide a more distinctive and stable representation for challenging document images.

COMPARATIVE EXAMINATION

To assess the importance of the suggested approach for recognizing objects of varying sizes, a comparative analysis was conducted against leading contemporary techniques. Tables 13 and 14 are given below for the SPODS and Tobacco 800 datasets. These tables show the results of comparing our proposed methodology with diverse approaches to visual object recognition.

The results in these tables show that our proposed multi-path architecture methodology outperforms the SOTA methods. The results reveal that our introduced model, with specific tasks, achieved the best performance among the compared approaches.

Table 13. The comparative analysis outcomes of SPODS dataset in respect to the F1 score (F1), Precision (P), Accuracy (A) and Recall(R) metrics

Approach	Visual Object	A (↑)	P (↑)	R (↑)	F1 (↑)
[49]	Signature	-	82.00	84.00	85.00
[37]	Stamp	-	66.00	64.00	67.00
[42]	Logo	-	99.00	98.00	99.00
	Stamp	-	69.00	91.00	76.00
[44]	Logo	-	99.00	98.00	99.00
	Stamp	-	89.00	90.00	87.00
	Signature	-	93.00	86.00	90.00
[36]	Stamp	-	98.50	62.85	-
[35]	Stamp	-	90.60	90.70	90.60
	Logo	99.54	98.19	100.00	99.09
Proposed	Stamp	99.54	100.00	98.15	99.07
	Signature	99.54	100.00	100.00	100.00

Table 14. The comparative analysis outcomes of the Tobacco 800 dataset in relation to Recall(R), Accuracy (A), F1 score (F1), and Precision (P) metrics

Approach	Visual Object	A (↑)	P (↑)	R (↑)	F1 (↑)
[17]	Signature	92.80	-	-	-
[18]	Signature	92.80	-	-	-
[47]	Signature	91.20	-	-	-
[48]	Logo	92.90	-	-	-
[29]	Logo	86.50	99.40	86.50	-
[32]	Logo	90.05	92.98	-	-
[33]	Logo	-	87.00	-	-
[50]	Signature	95.58	-	-	-
[51]	Signature	95.58	-	-	-
[52]	Logo	91.50	75.25	-	-
[53]	Logo	97.22	-	-	-
[54]	Logo	88.78	91.15	-	
[55]	Signature	-	**100.00**	84.00	-
[56]	Logo	89.52	-	-	-
[31]	Logo	91.47	98.10	-	-
[57]	Signature	-	72.00	-	-
[41]	Signature	89.60	-	-	-
	Logo	89.20	-	-	-
[30]	Logo	-	87.80	82.20	88.00
[9]	Signature	99.68	-	-	-
	Logo	99.50	-	-	-
[34]	Logo	96.86	98.93	-	-
[58]	Signature	94.04	-	-	-
[59]	Logo	95.00	-	-	-
[60]	Logo	-	98.90	89.50	-
[61]	Signature	94.40	-	-	-
[62]	Signature	78.00	-	-	-
Proposed	Signature	99.22	100.00	98.36	99.17
	Logo	99.22	98.55	100.00	99.27

ABLATION STUDY ANALYSIS

The significance of each element in the proposed multipath architecture is quantified through an ablation analysis. As shown in Table 15, the second scenario (with deep learning, patches, and histograms) achieved the highest recognition performance among all scenarios. This remarkable accomplishment (i.e., high accuracy) is due to the efficacy of histogram features derived from image patches, which facilitate the identification of various visual objects within the document image. The results presented in Table 15 show that including image patch slopes negatively affects recognition performance. This effect

is due to the fact that while handwritten signatures and text are primary objects that exhibit slopes, other visual elements rarely incorporate slopes in their structural design. A series of experiments is conducted on both the SPODS and Tobacco 800 datasets to determine whether adding two additional zero channels affect the performance of the proposed multipath network topology. The results in Tables 16, 17, 18, and 19 reveal that adding two additional zero channels to the slope process of the slope path increases the performance of the first (S1), third (S3), and fourth (S4) scenarios. However, the performance of the second scenario (S2) is still better with the addition of two zero-padding channels. In addition to experimental evidence, the study in [60] supports our claim that adding padding channels improves performance. Hence, we can say that the second scenario (S2), which adds two zero channels, is the best option if we want the best recognition performance.

Otherwise, if speed (saving memory and GPU resources) is the objective, the third scenario (S3), without adding zero channels, is the best option, at the expense of recognition performance. We note that if two zero channels are not added to the slope path, some changes are made to this path by fine-tuning the model to handle a one-channel grayscale input image patch.

Table 15. Evaluation of the proposed architecture in recognizing various elements of document images by, Precision (P), Recall(R) and Accuracy (A) metrics.

Case	Dataset	A (↑)	P (↑)	R (↑)
Path 1 (Deep Learning)	SPODS	97.10	97.22	97.11
	Tobacco	98.45	98.45	98.45
Path 1 +Path 2 (Deep Learning + Slopes)	SPODS	94.91	94.91	94.80
	Tobacco	94.57	94.57	94.57
Path 1 +Path 3 (Deep Learning + Histograms)	SPODS	99.54	99.54	99.54
	Tobacco	99.23	99.23	99.23
Path 1 + Path 2+ Path 3 (Deep Learning + Slopes + histograms)	SPODS	95.25	95.25	95.14
	Tobacco	96.89	96.89	96.89

Table 16. Overall performance of SPODS dataset in case of zero padding.

Metric/Scenario	S1	S2	S3	S4
Loss ↓	07.82	01.75	23.80	23.79
Accuracy ↑	97.11	99.54	94.91	95.25
MAE ↓	01.08	00.23	02.27	02.07
MSE ↓	01.83	00.66	02.66	02.83
Precision ↑	97.22	99.54	94.90	95.25
Recall ↑	97.11	99.54	94.79	95.14

Table 17. Overall performance of SPODS dataset in case of without zero padding.

Metric/Scenario	S1	S2	S3	S4
Loss ↓	10.76	06.16	06.20	07.32
Accuracy ↑	99.07	98.62	99.07	98.61
MAE ↓	01.04	01.12	00.56	01.13
MSE ↓	00.44	00.54	00.46	00.65
Precision ↑	99.07	98.61	99.07	98.60
Recall ↑	99.07	98.61	99.07	98.45

Table 18. Overall performance of Tobacco 800 dataset in case of zero padding.

Metric/Scenario	S1	S2	S3	S4
Loss ↓	07.34	02.55	30.00	06.80
Accuracy ↑	98.45	99.22	94.57	96.89
MAE ↓	01.09	00.79	04.98	02.21
MSE ↓	01.42	01.22	06.14	04.51
Precision ↑	98.45	99.22	94.57	96.89
Recall ↑	98.45	99.22	94.57	96.89

Table 19. Overall performance of Tobacco 800 dataset in case of without zero padding.

Metric/Scenario	S1	S2	S3	S4
Loss ↓	10.76	06.16	11.52	09.99
Accuracy ↑	99.07	98.62	99.22	98.44
MAE ↓	01.04	01.12	00.93	01.53
MSE ↓	00.44	00.54	00.78	01.47
Precision ↑	99.07	98.61	99.22	98.44
Recall ↑	99.07	98.61	99.22	98.44

CONCLUSION

An innovative multi-path architecture framework is created in this research to recognize various objects in the documents. Deep learning models and image patches are pivotal components of this framework for detecting, identifying, and distinguishing variable-sized visual objects. Deep learning, slope localization, and histogram analysis are the main three paths in this architecture, each performing a distinct job. The second experimental scenario, which utilizes patch histograms and deep learning techniques, exhibited a remarkable recognition performance, achieving approximately 100% recognition rates on the Tobacco 800 and SPODS datasets. However, the proposed framework architecture has a limitation: it cannot operate in real time on energy-constrained devices such as tablets and mobile phones. In future work, we will optimize the proposed multipath architecture to recognize visual objects on power-constrained devices and to extract handwritten and printed text patches across different document types. One limitation of this approach is real-time deployment on edge devices. Also, there

is a need to further investigate the proposed framework's ability to recognize various document layouts under challenging quality conditions. Such an intelligent investigation will require more training data and changing the architecture of the proposed framework.

In addition, the architecture and intelligence of DMPN-VOR will be further developed to analyse and understand the contents of difficult document images, especially those used in digital archives and legal documents. More experiments, statistical testing, quantitative/qualitative analysis, and comparisons with recent approaches for document analysis will be considered in future research. To sum up, the proposed method serves as an effective alternative tool to recognize visual objects of various sizes, colours, shapes, and orientations within document images.

NOMENCLATURE

RNNs	Recurrent Neural Networks
CNNs	Convolutional Neural Networks
AI	Artificial Intelligence
DMPN-VOR	Deep Multipath Network-Visual Object Recognition
SPODS	Scanned Pseudo-Official Data Set
CAE	Convolutional Autoencoder
LBP	Local Binary Pattern
GLCM	Gray Level Co-occurrence Matrix
SURF	Speeded up Robust Features
FCM	Fuzzy C-Means
GAN	Generative Adversarial Network
DPI	Dot Per Inch
MSE (MS)	Mean Square Error
MAE (MA)	Mean Absolute Error
ID	Identification
n	Number of labelled documents
X and Y	Directions
∇ x	Gradient in X-direction
∇ y	Gradient in Y-direction
G	Gradient Magnitude
∅	Slope
p_r	gray level (r) probability of the in the image patch
r_k	gray level (k)
n_k	number of the gray level (k) occurrences of in the patch

n	total count of pixels of the image patch
L	count of distinct gray levels of the image patch
Ip	document image patches
S1	output slope patches
H1	output histograms
C1	concatenated feature vector
R1	reduced feature vector
C2	Final feature vector
G1 and G2	Two ground truth sets
TP	true positives
FP	false positives
FN	false negatives
y_i	actual values
yˆ	predicted values
N	sample size
P	Precision
R	Recall
S1	First Scenario
S2	Second Scenario
S3	Third Scenario
S4	Fourth Scenario
L	Loss
A	Accuracy
F1	F1-Score

AUTHORS CONTRIBUTIONS

Conceptualization, A.J.A.; Methodology, A.J.A.; Software, M.A., and S.A.; Computationalization, A.J.A.; Modelling, A.J.A.; Validation, A.J.A., M.A., and S.A.; Formal Analysis A.J.A.; Investigation, A.J.A., and S.A.; Resources, A.J.A.; Data Curation, A.J.A., and M.A.; Writing Original Draft Preparation, A.J.A., and M.A.; Writing – Review & Editing, A.J.A.; Visualization, S.A.; Supervision, S.A.

ACKNOWLEDGMENT

We extend our appreciation to the College of Engineering at the University of Diyala for its invaluable support in making this research a reality. Furthermore, we express our

sincere gratitude to Abdallah, a skilled Python technician, for his support in developing and implementing the deep learning algorithms.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interest regarding the publication of this paper.

REFERENCES

1. Doermann, D. The indexing and retrieval of document images: A survey ", Computer Vision and Image Understanding, *Computer Vision and Image Understanding*, **2014**, *70*(3), 287-298.
2. Ferilli S. Automatic digital document processing and management: Problems, algorithms and techniques. *Advances in Computer Vision and Pattern Recognition Book Series, Springer*, **2011**.
3. Kukkar, A., Kaur. G. AEC: A novel adaptive ensemble classifier with LIME and SHAP-Based interpretability for fake news detection, *Expert Systems with Applications*, **2025**, *281*, 127751.
4. Momin, M.S., Sufian, A., Barman, D., Leo, M., Distante, C. and Damer, N. Explainable deepfake detection across different modalities: An overview of methods and challenges. *Image and Vision Computing*, **2025**, *163*, 105738.
5. El Bahi, H., Zatni, A., "Text recognition in document images obtained by a smartphone based on deep convolutional and recurrent neural network", *Multimedia Tools and Applications*, **2019**, *78*(18), 26453-81.
6. Mahadevkar, S.V., Patil, S., Kotecha, K., Soong, L.W., Choudhury, T. Exploring AI-driven approaches for unstructured document analysis and future horizons, *Journal of Big Data*, **2024**, *11*(1), 92.
7. Khan, M. S., Hongsong, C. Hybrid transformer deep neural architectures for enhanced misinformation detection on social media, *Expert Systems with Applications*, **2025**, *300*, 130470.
8. Liu, C., Yin, K., Cao, H., Jiang, X., Li, X., Liu, Y., Jiang, D., Sun, X. and Xu, L. Hrvda: High-resolution visual document assistant, *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, Washington, USA, **2024**, pp. 15534-15545.
9. Mandal, R., Roy, P.P., Pal, U., Blumenstein, M. Bag-of-visual-words for signature-based multi-script document retrieval, *Neural Computing and Applications*, **2019**, 31, 6223-6247.
10. Li, P., Fan, J., Wu, J. Exploring the key technologies and applications of 6G wireless communication network, *iScience*, **2025**, *28*(5), 112281.
11. Gauthier, I., Tarr, M. J. Visual object recognition: Do we (finally) know more now than we did?, *Annual review of vision science*, **2016**, *2*(1), 377-396.
12. Yapıcı, M.M., Tekerek, A., Topaloğl, N. Deep learning-based data augmentation method and signature verification system for offline handwritten signature, *Pattern Analysis and Applications*, **2021**, *24*(1), 165-179.
13. Sumalatha, U., Prakasha, K.K., Prabhu, S., Nayak, V.C. A comprehensive review of unimodal and multimodal fingerprint biometric authentication systems: Fusion, attacks, and template protection, *IEEE Access*; **2024**, 12, 64300-64334.
14. Zong, Y., Liu, S., Liu, X., Gao, S., Dai, X., Gao, Z. Robust synchronized data acquisition for biometric authentication, *IEEE Transactions on Industrial Informatics*, **2022**, *18*(12), 9072-9082.

15. Chanda, S., Prasad, P. K., Hast, A., Brun, A., Martensson, L., Pal, U. Finding Logo & Seal in Historical Document Images-An Object Detection Based Approach, *In Pattern Recognition.: 5th Asian Conference*, ACPR **2019**, pp. 821-834.
16. Forczmański, P., Markiewicz, A. Two-stage approach to extracting visual objects from paper documents, *Machine Vision and Applications*, **2016**, *27*(8), 1243-1257.
17. Zhu, G., Zheng, Y., Doermann, D., Jaeger, S. Multi-scale structural saliency for signature detection, *In 2007 IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN, USA, **2007**, pp. 1-8.
18. Zhu, G., Zheng, Y., Doermann, D., Jaeger, S. Signature detection and matching for document image retrieval, *IEEE Transactions on Pattern Analysis and Machine Intelligence Journal*, **2008**, *31*(11), 2015-2031.
19. Ghosh, R. A Recurrent Neural Network based deep learning model for offline signature verification and recognition system, *Expert Systems with Applications*, **2021**, *168*, 114249.
20. Lien, C. W., Vhaduri, S. Challenges and opportunities of biometric user authentication in the age of IoT: A survey, *ACM Computing Surveys*, **2023**, *56*(1), 1-37.
21. Kao, H.H., Wen, C.Y. An offline signature verification and forgery detection method based on a single known sample and an explainable deep learning approach, *Applied Sciences*; **2020,** *10*(11), 3716.
22. Ghazal, T.M. Convolutional neural network based intelligent handwritten document recognition, *Computer, Materials & Continua*, **2022**, *70*(3), 4563-81.
23. Vorugunti, C. S., Pulabaigari, V., Gorthi, R. K. S. S., Mukherjee, P. OSVFuseNet: Online Signature Verification by feature fusion and depth-wise separable convolution based deep learning, *Neurocomputing*; **2020**, *409*, 157-172.
24. Shariatmadari, S., Emadi, S. and Akbari, Y. Patch-based offline signature verification using one-class hierarchical deep learning, *International Journal on Document Analysis and Recognition*, **2019**, *22*(4), 375-385.
25. Markiewicz, A., Forczmański, P. Detection and classification of interesting parts in scanned documents by means of adaBoost classification and low-level features verification, *In Computing Analysis of Image and Pattern: 16th Int. Conference*, CAIP 2015, Valletta, Malta, September 2-4, 2015, Proceedings, Part II 16, **2015**, pp. 529-540.
26. Junior, C.A., da Silva, M.H.M., Bezerra, B.L.D., Fernandes, B.J.T., Impedovo, D. Fcn+ rl: A fully convolutional network followed by refinement layers to offline handwritten signature segmentation, In *IEEE 2020 International Joint Conference on Neural Networks (IJCNN)*. **2020**, pp. 1-7.
27. Melo, V.K.S.L., Dantas, B.B.L. A fully convolutional network for signature segmentation from document images, In *IEEE 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*. **2018**, pp. 540-545.
28. Forczmański, P., Smoliński, A., Nowosielski, A. and Małecki, K. Segmentation of scanned documents using deep-learning approach, In *International Conference on Computer Recognition Systems*; **2020**, *11*, 141-152.
29. Li, Z., Schulte-Austumand, M. and Neschen, M. Fast logo detection and recognition in document images, In *IEEE 2010 20th International Conference on Pattern Recognition*, Istanbul, Turkey, **2010**, pp. 2716-2719.

30. Veershetty, C., Hangarge, M. Logo retrieval and document classification based on LBP features, *In Data Analysis and learning: Proceeding of DAL 2018, Singapore*, **2019**, pp. 131-141.
31. Naganjaneyulu, G.V.S.S.K.R., Sai Krishna, C., Narasimhadhan, A.V. A novel method for logo detection based on curvelet transform using GLCM features, *In Proceeding of 2nd International Conference on Computer Vision and Image Processing: CVIP 2017*, **2018**, pp. 1-12.
32. Pham, T.A., Delalandre, M. and Barrat, S. A contour-based method for logo detection, *In IEEE 2011 International Conference on Document Analysis and Recognition*, Beijing, China, **2011**, pp. 718-722.
33. Jain, R., Doermann, D. Logo retrieval in document images, *In IEEE 2012 10th IAPR International Workshop on Document Analysis System*, **2012**, pp.135-139.
34. Zaaboub, W., Tlig, L., Sayadi, M. and Solaiman, B. Logo detection based on fcm clustering algorithm and texture features, *In International Conference on Image and Signal Processing*, Springer, Cham, **2020**, pp. 326-336.
35. Gayer, A., Ershova, D. and Arlazarov, V. Fast and accurate deep learning model for stamps detection for embedded devices, *Pattern Recognition and Image Analysis*, **2020**, *32*(4), 772-779.
36. Matalov, D.P., Usilin, S.A. and Arlazarov, V.V. About Viola-Jones image classifier structure in the problem of stamp detection in document images, *In SPIE Thirteenth International Conference on Machine Vision, Rome*, Italy, **2021**, pp. 241-248.
37. Ahmed, S., Shafait, F., Liwicki, M., Dengel, A. A generic method for stamp segmentation using part-based features, *In IEEE 13 12th International Conference on Document Analysis and Recognition*, **2013**, pp. 708-712.
38. Forczmański, P. and Markiewicz, A. Stamps detection and classification using simple features ensemble, *Mathematical Problems in Engineering*, **2015**, *2015*(1), 367879.
39. Jin, X., Mu, Q., Chen, X., Liu, Q., Xiao, C. Digital Archive Stamp Detection and Extraction, In *International Symposium on Artificial Intelligence and Robotics*, **2023**, pp. 165-174.
40. Nandedkar, A. V., Mukherjee, J., Sural, S. A spectral filtering based deep learning for detection of logo and stamp, In *IEEE 2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics*, Patna, India, **2015,** pp. 1-4.
41. Sharma, N., Mandal, R., Sharma, R., Pal, U., Blumenstein. Signature and logo detection using deep CNN for document image retrieval, In *IEEE 2018 16th international Conference on frontiers in handwriting Recognition*, **2018**, pp. 416-422.
42. Nandedkar, A.V., Mukhopadhyay, J., Sural, S. Text-graphics separation to detect logo and stamp from colour document images: A spectral approach, In *IEEE 13th International Conference on Document Analysis and Recognition*, **2015,** pp. 571-575.
43. Dey, S., Mukherjee, J., Sural, S. Stamp and logo detection from document images by finding outliers, In *IEEE 2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics*, **2015**, pp. 1-4.
44. Nandedkar, A. V., Mukherjee, J., Sural, S. SPODS: A dataset of colour-official documents and detection of logo, stamp, and signature, In *Computer Vision, Graphics and Image Processing: ICVGIP 2016 Satellite Workshops*, WCVA, DAR, and MedImage, Guwahati, India, December 19, **2017**; pp. 219-230.
45. Lewis, D., Agam, G., Argamon, S., Frieder, O., Grossman, D., Heard, J. Building a test collection for complex document information processing. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information Retrieval*, **2006**; pp. 665-666.

46. University of Maryland, Laboratory for Language, and Media Processing (LAMP), "Tobacco-800 signatures and logos dataset", **2006**.
47. Srinivasan, H., Srihari, S. Signature-based retrieval of scanned documents using conditional random fields, In *Computational Methods for Counterterrorism*, **2009**, pp. 17-32.
48. Wang, H. Document logo detection and recognition using Bayesian model, In *IEEE 2010 20th International Conference on Pattern Recognition* ,**2010**, pp. 1961-1964.
49. Ahmed, S., Malik, M. I., Liwicki, M., Dengel, A. Signature segmentation from document images, In *IEEE International Conference on Frontiers in Handwriting Recognition*, **2012**, pp. 425-429.
50. Mandal, R., Roy, P. P., Pal, U. Signature segmentation from machine printed documents using contextual information, *International Journal of Pattern Recognition and Artificial Intelligence*, **2012**, *26*(7), 1253003.
51. Mandal, R., Roy, P. P., Pal, U., Blumenstein, M. Signature segmentation and recognition from scanned documents, In *IEEE 2013 13th International Conference on Intelligent Systems Design and Applications*, **2013**, pp. 80-85.
52. Alaei, A., Delalandre, M., Girard, N. Logo detection using painting-based representation and probability features, In *IEEE 2013 12th International Conference on Document Analysis and Recognition*, **2013**, pp. 1235-1239.
53. Alaei, A., Delalandre, M. A complete logo detection/recognition system for document images, In *2014 11th IAPR IEEE International Workshop on Document Analysis Systems*, **2014**, pp. 324-328.
54. Le, V. P., Nayef, N., Visani, M., Ogier, J. M., De Tran, C. Tran. Document retrieval based on logo spotting using key-point matching, In *IEEE 2014 22nd international Conference on Pattern Recognition*, **2014**, pp. 3056-3061.
55. Butt, U. M., Ahmad, S., Shafait, F., Nansen, C., Mian, A. S., & Malik, M. I. Automatic signature segmentation using hyper-spectral imaging, In *IEEE 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, **2016**, pp. 19-24.
56. Dixit, U.D., Shirdhonkar, M.S. Automatic logo detection and extraction using singular value decomposition, In *IEEE 2016 International Conference on Communication and Signal Processing (ICCSP)*, **2016**, 0787-0790.
57. Wiggers, K.L., Britto, A.S., Heutte, L., Koerich, A.L., Oliveira, L.E.S. Document image retrieval using deep features, In *IEEE 2018 International Joint Conference on Neural Networks (IJCNN)*, **2018**, pp. 1-8.
58. Kanchi, S., Pagani, A., Mokayed, H., Liwicki, M., Stricker, D., Afzal, M. Z. EmmDocClassifier: Efficient multimodal document image classifier for scarce data, *Applied Sciences*, **2022**, 12(3), 1457.
59. Dixit, U.D., Shirdhonkar, M.S., Sinha, G.R. Automatic logo detection from document image using HOG features, *Multimedia Tools and Applications*, **2023**, *82*(1), 863-878.
60. Guijarro, M., Bayon, J., Martín-Carabias, D., Recas, J. A Multi-Stage Method for Logo Detection in Scanned Official Documents Based on Image Processing, *Algorithms*, **2024**, *17*(4), 170.
61. Thejashwini, B. L., Nagendraswamy, H. S., Supritha, B., Somanna, M. Automated Template Based Signature Detection and Extraction from Multilingual Documents, In *2024 IEEE International Conference on Computer Vision and Machine Intelligence (CVMI)*, **2024**, pp. 1-6.
62. Dinçer, E., Kalaycı, S., Yurdakul, E., Köroğlu, B. Removing Background from Noisy Handwritten Signatures on Banking Documents Using GANs, In *2024 9th IEEE International Conference on Computer Science and Engineering (UBMK)*, **2024**, pp. 846-850.