RESEARCH ARTICLE

Detection of Newspaper Layouts Using YOLO12

Atul Kumar[1,*] **and Gurpreet Singh Lehal**[2]

[1] Department of Computer Science, R.G.M. Government College, Joginder Nagar 175015, India
[2] Department of Computer Science, Punjabi University, Patiala 147002, India

Abstract

This study presents a robust and scalable method for automatic layout detection in digitized newspapers to facilitate efficient knowledge extraction and information retrieval. A custom dataset comprising annotated newspaper images in English, Hindi, and other languages was developed, with layout regions categorized into five primary classes. An enhanced YOLOv12 object detection model was trained on this dataset and evaluated using the mean Average Precision (mAP) metric across various Intersection over Union (IoU) thresholds. The model achieved a mAP@50 of 0.88, demonstrating strong detection performance and outperforming several stateof-the-art object detection models in the same task. The findings validate the effectiveness of the proposed approach in handling multilingual, structurally diverse newspaper formats. This research provides a practical framework for integrating automated layout analysis into digital archiving systems, OCR pipelines, and media monitoring applications. It also supports broader efforts to digitize historical print media and improve accessibility to regional content, thereby enabling enhanced research, journalism, and public engagement.

Keywords: newspapers, YOLO, segmentation, layout analysis.

1 Introduction

In the digital age, automatic analysis and digitization of printed media, particularly newspapers, are crucial for preservation, content extraction, and information retrieval. Conventional Optical Character Recognition (OCR) algorithms often prove inadequate when applied directly to complex newspaper layouts, which usually include multi-column formats, integrated graphics, a range of fonts, and various article structures. Accurate layout detection is required for effective downstream tasks like article segmentation, reading order reconstruction, and content categorization.

Recent advances in computer vision, particularly deep learning-based object detection models, have considerably improved document layout analysis. Among them, the You Only Look Once (YOLO) model family has emerged as a leading strategy owing to its real-time performance and accuracy. The most recent generation, YOLOv12, improves architectural efficiency, spatial attention mechanisms, and multi-scale feature extraction, making it ideal for the complex layouts found in newspapers. The digitization of newspapers can be done with the help of OCR, a method of automatically identifying characters while optically scanning and digitizing text [1]. The analysis of newspaper layouts plays a pivotal role

Submitted: 10 November 2025
Revised: 14 December 2025
Accepted: 31 December 2025
Published: 09 February 2026

Vol. 2, **No.** 2, 2026.
10.62762/TMI.2025.846033

***Corresponding author:**
✉ Atul Kumar
atulkmr02@gmail.com

Citation
Kumar, A., & Lehal, G. S. (2026). Detection of Newspaper Layouts Using YOLO12. *ICCK Transactions on Machine Intelligence*, 2(2), 77–87.

© 2026 ICCK (Institute of Central Computation and Knowledge)

in understanding the structure and content of a newspaper page, allowing for a holistic examination of its design and informational elements. The layout also depends on many parameters like editor's style, column size, types of images, advertisements, inverted text etc. It is so difficult to analyze the layout of these newspapers [2].

Although researchers in this sector have recently established a number of methodologies to analyze newspaper layouts, gaps remain in handling multilingual and structurally diverse formats. The objective of this work is to perform a newspaper layout analysis, which has the following workflows:

1. Construction of a dataset specifically for newspaper layouts in COCO format, annotated with bounding boxes across five classes.
2. Training on the dataset using the YOLOv12 model [3] architecture.
3. Comparison with other state-of-the-art object detection models.

The remainder of this paper is organized as follows: related work, challenges in the layout analysis of newspapers, methodology, data preparation, experiments, and conclusion.

2 Related Work

Over the years, considerable research has been conducted in the field of computer vision and image processing, leading to the development of numerous methods for newspaper layout analysis and recognition. With the advent of deep learning techniques, significant advancements have been made in solving the problem of newspaper layout analysis. By employing neural networks, it has become possible to accurately identify and extract different components of a newspaper page. However, the diverse designs of newspapers pose significant challenges for layout analysis and comparison. To address these challenges, the layout and content of a newspaper, as well as its style and format, can all impact the variability. Newspaper layouts can be found in a variety of structures [4], including rectangular, Manhattan, non-Manhattan, and multi-column Manhattan formats. Ref. [5] comprehensively analyzed different layout types and discussed the framework of layout analysis. The phases of the layout analysis algorithm are derived from an extensive review of the literature and have been comprehensively discussed in this study. Classical top-down, bottom-up, and hybrid approaches based on the XY cut algorithm [6], whitespace-based segmentation using maximal empty rectangles [7], and the run length smoothing algorithm [8] have been proposed to segment blocks in documents. Deep learning techniques have shown promise for a range of applications, including headline recognition, text segmentation, and object recognition. Ref. [9] proposed a rule-based method for logical layout analysis of historical French newspapers and evaluated it against two machine learning models (RIPPER and Gradient Boosting), using a dataset of press documents digitized from the Bibliothèque Nationale de France archives. Ali et al. [10] demonstrated that document layout analysis serves as a critical step for providing article-level access to digitized historical newspaper collections, proposing a computer vision workflow in which page segmentation precedes OCR-based metadata enrichment and content retrieval, and Ref. [11] presented a comparative evaluation of YOLO-based architectures (YOLOv8 through YOLOv11) for layout analysis of historical newspaper images, introducing a new annotated dataset of historical Brazilian newspapers and benchmarking model performance using mAP across layout detection tasks. Ref. [12] introduced the Newspaper Navigator dataset by applying a deep learning pipeline to extract headlines, photographs, illustrations, maps, comics, editorial cartoons, and advertisements from over 16 million historic newspaper pages in Chronicling America. Ref. [13] proposed an enhanced YOLO-based model for document layout analysis, incorporating multi-convolutional deformable separation modules to handle the high variability in aspect ratios among document layout elements. Ref. [14] introduced DocLayNet, a large human-annotated document layout segmentation dataset in COCO format containing over 80,000 pages from diverse document sources—including financial reports, manuals, scientific literature, legal documents, patents, and government tenders—demonstrating that models trained on varied layouts achieve stronger generalization than those trained on single-domain corpora. Ref. [15] proposed an integrated layout analysis and OCR system for printed documents, employing YOLO-based deep learning models to identify the positions of titles, text paragraphs, tables, and images, followed by CNN-based text recognition to produce editable output. Ref. [16] proposed deep learning models for tasks such as text detection and structure identification, providing pre-trained neural network models and comprehensive tools

for effective data annotation on document images and model fine-tuning to accommodate various levels of customization. To enhance the performance of area recognition, YOLO paired with a Feature Pyramid Network [17] to integrate features at multiple levels to recognise smaller regions. You Only Look Once (YOLO), one of CNN's most iconic depictions, is an original and uncomplicated solution to the object detection challenge [18]. For OCR-based character recognition, the open-source Tesseract engine [19] provides a robust solution, employing a two-pass adaptive classification process and novel line-finding algorithms capable of handling skewed pages and complex document layouts. Ref. [20] presented a comprehensive review of YOLO architectures from YOLOv1 to YOLOv8 and YOLO-NAS, summarizing the key architectural innovations and training strategies that have driven real-time object detection performance improvements. Ref. [21] refined the multimodal model by developing a multimodal method that combines textual and visual clues for the semantic segmentation of historical newspapers. Addressing the complexity of magazine and newspaper pages, Ref. [22] proposed a language-agnostic framework using CNNs and transformers to detect newspaper page elements—including news articles, advertisements, and page headers—and to categorize newspaper pages based on their visual features. Ref. [23] suggested a successful hybrid bottom-up/top-down technique. Ref. [24] presented a comprehensive survey of deep learning methods applied to historical document analysis and recognition, categorizing approaches by task—including layout segmentation, text detection, and recognition—and demonstrating that recent research has shifted from adapting existing algorithms to addressing novel tasks and applications specific to historical documents. Ref. [25] proposed a lightweight network based on dilated convolutions for memory-efficient document layout analysis. Ref. [26] presented a comprehensive survey on the digitization of historical paper documents—covering the full pipeline from layout analysis and OCR to information retrieval and mining—and highlighted persistent challenges in processing structurally complex and multilingual newspaper collections.

3 Challenges in Newspaper Digitization

Numerous problems could arise when analyzing the newspapers for digitization. The typical ones include:

1. Preprocessing the image is crucial before moving on to other stages due to the inherent low quality of old newspaper editions and potential digitization errors (the page is not positioned correctly on the scanner, or the scanner automatically selects the incorrect binarization threshold).
2. To further divide the columns and lines in a newspaper story with photographs, the text must be separated from the images. Due to the fact that photographs can span two or more columns, there is no longer a requirement for continuous white lines to separate the columns.
3. There is no standard layout format for newspapers; instead, each publication may have a unique layout. Because of this, it could be challenging to develop an algorithm that can accurately evaluate and extract information from different newspaper layouts.
4. Because newspapers typically employ images and graphics to convey information, automated systems find it challenging to evaluate and extract information from them.
5. Text may overlap in certain newspaper structures, making it difficult for an algorithm to discriminate between different parts and efficiently extract information.
6. Different language scripts are complex (other than English) and can be difficult for computer programmers to read, especially if the font is not conventional. This may lead to mistakes in text recognition and layout analysis.
7. It may be difficult for algorithms to establish the proper orientation and correctly extract the material because it can be written either vertically or horizontally.

4 Methodology

This study employs a pipeline for detecting textual and layout-based components in newspaper images using a deep learning approach. The complete methodology is divided into six key stages, as illustrated in the workflow Figure 1.

4.1 Data Acquisition

High-resolution newspaper page images were collected from a diverse range of sources, including digital newspaper archives, scanned printed editions, and online repositories, as shown in Table 1.

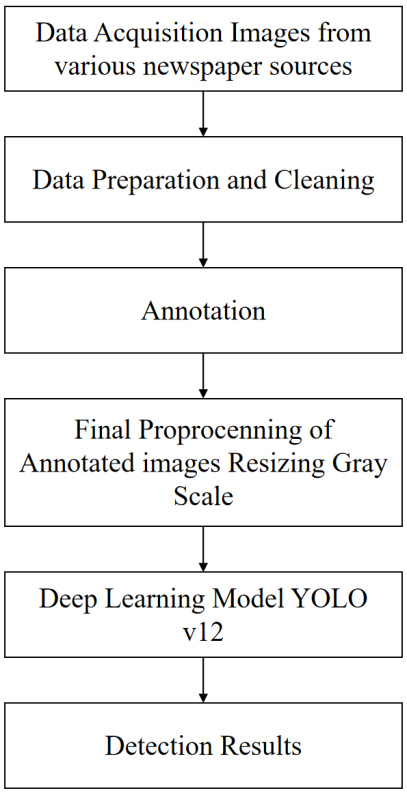


Figure 1. Flow diagram of newspaper recognition system.



Figure 2. Sample image of layout annotation using Roboflow.

The goal was to compile a comprehensive and representative dataset encompassing various layouts, fonts, languages, and noise levels.

Table 1. Collection of Newspapers from different resources.

Source	Number of Newspapers' images collected
Digitized archives	875
Local Newspapers from Libraries	450
Multilingual	175
Total	**1500**

Table 2. Number of annotations per Class.

Annotation Type	Count
Text	14352
Headline	8654
Image	9463
Advertisement	3212
Caption	6543
Total	**42224**

4.2 Data Preparation and Cleaning

The acquired raw images underwent preprocessing to ensure uniformity and reduce noise. This involved:

- Eliminating corrupted or low-quality images.
- Standardizing image formats and resolutions.
- Enhancing contrast and brightness where necessary.
- Eliminating irrelevant margins or marks using cropping techniques.

4.3 Annotation

Following the cleaning process, a manual and semi-automated annotation process was performed to label different newspaper elements such as Text, Headline, Image, advertisement, and caption. This annotation stage was crucial for supervised learning and was facilitated using tools like LabelImg or custom annotation platforms like Roboflow [27]. Figures 2, 3, and 4 show the annotation process using Roboflow. The number of annotations per class is shown in Figure 5. The detailed count of annotations per class is provided in Table 2.

5 Final Preprocessing

Annotated images were subjected to final preprocessing steps to optimize them for deep learning:

Resizing: All images were resized to a fixed dimension of 640x640 compatible with the YOLO V12 model, preserving aspect ratio as much as possible. To ensure robustness and generalizability, the dataset was curated to maintain a balanced distribution across categories and layout types, encompassing both single- and multi-column formats, image-heavy and text-dominant designs, as well as pages with varying advertisement densities. The annotation strategy focused on capturing the semantic and structural diversity of newspaper layouts, enabling the detection model to distinguish among visually similar but



Figure 3. Raw annotation data structure in JSON format showing bounding box coordinates and metadata for layout components.

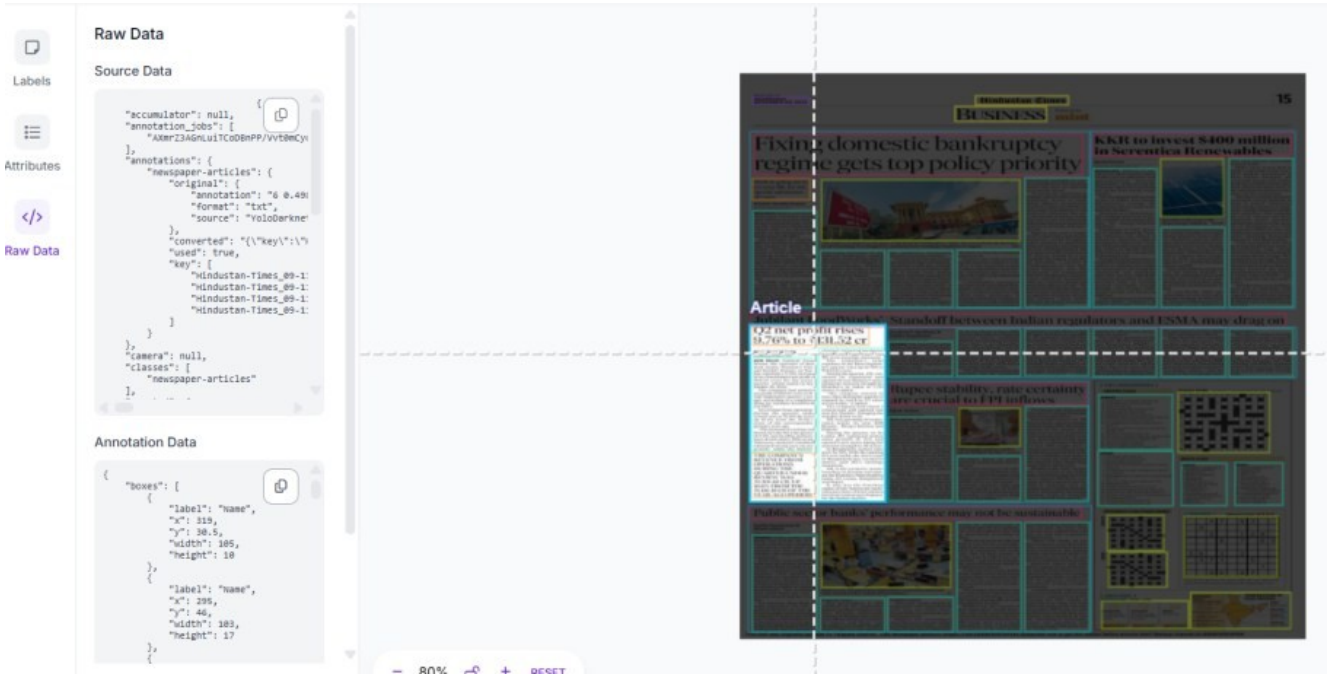


Figure 4. Annotation data example showing JSON structure alongside extracted newspaper headlines.

contextually different components. The choice of input resolution plays a crucial role in achieving an optimal trade-off between model accuracy and computational efficiency in layout detection tasks. The resolution(640x640) was selected based on the YOLO architecture's inherent requirements, which operate most effectively with standardized square input dimensions. The 640 × 640 resolution offers a well-balanced design that maintains a manageable computational footprint for quicker training and inference while preserving important layout elements like headlines, column boundaries, and advertisement blocks. Experiments showed that 640 × 640 provided adequate spatial fidelity for the detection of macro-structural components typical of newspaper layouts, without significantly degrading accuracy, even though higher resolutions (e.g., 1024 × 1024) could capture finer text-level features.

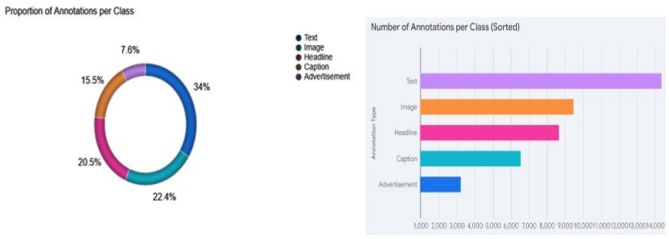


Figure 5. Annotation statistics for the newspaper layout dataset: class proportions and sorted counts.

Grayscale Conversion: Images were converted to grayscale to reduce computational complexity and emphasize textual and layout-related features over colour variations. The YOLOv12 model input layer was accordingly configured to accept single-channel images by setting the input channels parameter to 1, consistent with the grayscale preprocessing applied to the dataset.

5.1 YOLOv12 Architecture

YOLOv12 features an attention-centric architecture, designed to blend the speed of traditional CNN-based YOLO models with the performance benefits of attention mechanisms. Its core innovations include an optimized backbone leveraging Residual Efficient Layer Aggregation Networks (R-ELAN) for enhanced feature extraction and stable training, and a novel "area attention" module that strategically partitions feature maps to reduce computational overhead while maintaining a large effective receptive field. This architecture (Figure 6) also integrates Flash Attention for memory efficiency and employs 7x7 separable convolutions to further reduce computational burden, making it highly efficient for real-time object detection across various scales.

6 Experiments

6.1 Dataset Preparation

To train our YOLO model, we used a total dataset of 1,500 newspaper images. This dataset was partitioned into three subsets: 1,070 images (71.3%) for training, 300 images (20.0%) for validation, and 130 images (8.7%) for testing. The training set is used to teach the model, the validation set is used to tune hyperparameters and prevent overfitting, and the test set provides a final, unbiased evaluation of the model's performance. The complete distribution is detailed in Table 3 and Figure 7.

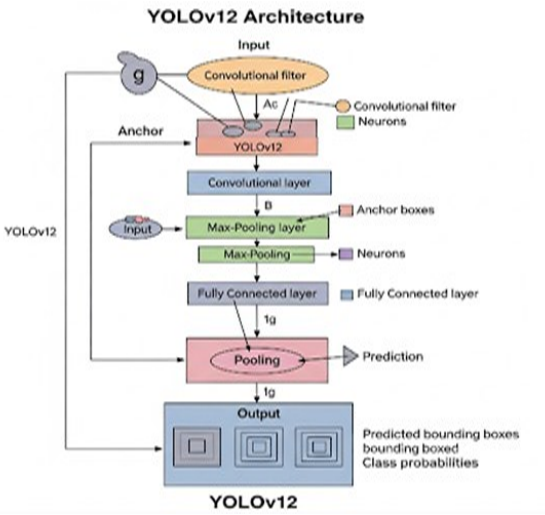


Figure 6. YOLOv12 architecture diagram showing the processing pipeline from input image to predicted bounding boxes and class probabilities.

Table 3. Dataset distribution for Model.

Type	Number of Images
Training Set	1070
Validation Set	300
Testing Set	130
Total	1500

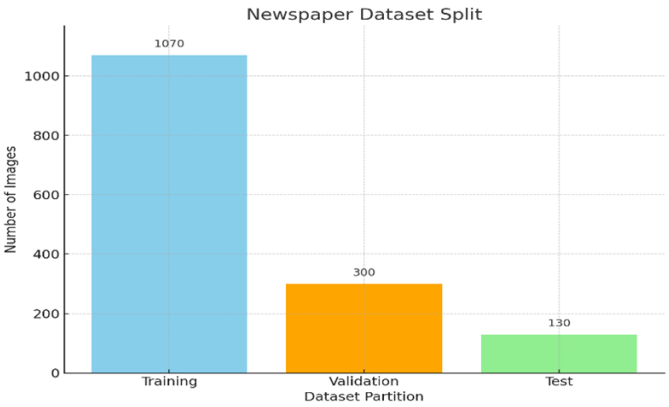


Figure 7. Dataset of Newspaper images with training, validation, and test.

6.2 Evaluation Metrics

The performance of the proposed model is assessed using a set of standard evaluation metrics, namely Precision (P), Recall (R), and Mean Average Precision (mAP), as shown in Table 4. The mathematical formulations used to compute each of these metrics are presented in Table 4, providing a thorough evaluation of both the model's accuracy and computational efficiency.

Table 4. Summary of key evaluation metrics.

Metric	Formula	Purpose
Precision	$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}}$	Measures the correctness of positive predictions
Recall	$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}}$	Measures coverage of actual positives
Average Precision	$\text{AP} = \int_0^1 P(R)\, dR$	Area under the precision-recall curve for a class
Mean Average Precision	$\text{mAP} = \frac{1}{n} \sum_{i=1}^{n} \text{AP}_i$	Mean of AP over all classes

6.3 Experimental Details

The experimental setup for this study is based on an Ubuntu 22.04 operating system, utilizing PyTorch version 2.6.0, Python 3.11.12, and CUDA 12.4. The hardware configuration comprised NVIDIA A100 GPUs. Each model was trained for 200 epochs using the Stochastic Gradient Descent (SGD) optimizer, with an initial learning rate of 0.01 and momentum 0.9. Given the high resolution of images in the dataset, using the original dimensions would significantly extend training time, while excessively reducing the input size could compromise accuracy. All input images were uniformly resized to 640 × 640 pixels to balance computational efficiency and model performance.

6.4 Training

The model was trained for 200 epochs. The plots indicate effective model training over 200 epochs in Figure 8. Training and validation losses (box, classification, and DFL) consistently decrease, demonstrating stable convergence and improved model fitting.

The model was evaluated on 130 newspaper images from the test set, as shown in Table 5.

Table 5. Detection of the Model on Different Categories

Category	Original category(Count)	Detected category(Count)	Accuracy (%)
Text	1563	1430	91.49
Headline	476	438	92.01
Image	520	504	96.92
Advertisement	76	69	90.78
Caption	274	252	91.97

To assess the effectiveness of the proposed method, the enhancement strategy was applied to various YOLOv12 model variants and compared against their respective baseline counterparts. The experimental outcomes are presented in Table 6, with the highest-performing results for each variant emphasized in bold. A detailed analysis of the results demonstrates that YOLOv12 consistently surpasses the baseline models in terms of both detection accuracy and model efficiency. For the comparative evaluation against state-of-the-art models in Table 7, the YOLOv12m variant was selected as the proposed model and further optimised, achieving a mAP@50 of 0.88, which represents an improvement over the variant-level results reported in Table 6.

Table 6. Performance comparison between YOLO12n, YOLO12s and YOLO12m.

Model	mAP@50 (%)	mAP@50–95 (%)	Params
YOLO12n	0.76	0.71	2.6
YOLO12s	**0.81**	0.74	9.3
YOLO12m	0.80	**0.78**	20.2

Table 6 compares three YOLOv12 model variants: YOLOv12n, YOLOv12s, and YOLOv12m based on their detection performance and model complexity. YOLOv12n, the most lightweight model with 2.6 million parameters, achieves a mAP@50 of 76% and mAP@50–95 of 71%, making it suitable for resource-constrained environments. YOLOv12s, with 9.3 million parameters, offers improved accuracy with a mAP@50 of 81% and mAP@50–95 of 74%, balancing performance and efficiency. YOLOv12m, the most complex model at 20.2 million parameters, attains the highest mAP@50 of 80% and mAP@50–95 of 78%, indicating strong and consistent detection performance across a range of IoU thresholds.

6.5 Comparative Analysis and Model Performance

YOLOv12's performance was benchmarked against several state-of-the-art models on the same dataset. To thoroughly assess the performance of YOLOv12, we compared this against several state-of-the-art models, including Faster R-CNN [28], Mask R-CNN [29], YOLOv5 [30], YOLOv8 [31], YOLOv10 [32], and YOLOv11 [33]. The comparison

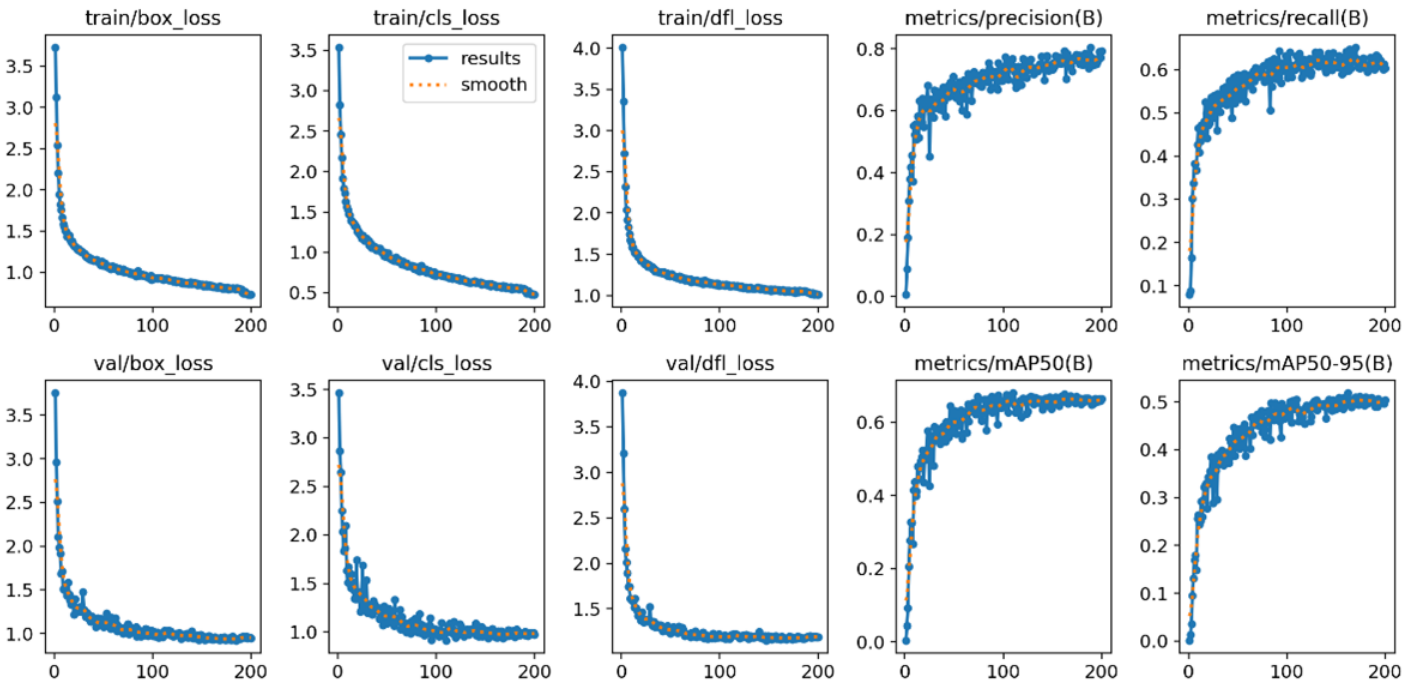


Figure 8. Training and validation performance metrics of the object detection model.

Table 7. Comparison of YoloV12 with Other State-of-the-art models.

Model	mAP@50 (%)	mAP@50–95 (%)	Params (M)
Faster R-CNN [28]	0.70	0.69	42
Mask R-CNN [29]	0.76	0.73	45
YOLOv5 [30]	0.80	0.79	7.2
YOLOv8 [31]	0.83	0.82	11.1
YOLOv10 [32]	**0.85**	0.73	14.3
YOLOv11 [33]	0.83	0.81	18.5
YOLOv12 (proposed)	**0.88**	**0.86**	**20.2**

Table 8. Detection Accuracy (%) for Each Category Across Tested Models.

Model	Text	Headline	Image	Advertisement	Caption	Average Accuracy (%)
Faster R-CNN [28]	86.42	87.30	90.15	84.67	85.92	86.89
Mask R-CNN [29]	88.73	88.54	91.23	86.10	87.48	88.42
YOLOv5	88.24	90.92	93.12	87.34	89.50	89.82
YOLOv8	89.35	91.87	92.21	87.42	90.76	90.32
YOLOv10 [32]	89.02	91.65	94.02	85.34	91.55	90.32
YOLOv11	90.02	91.56	94.45	86.11	92.02	90.83
YOLOv12 (Proposed)	**91.49**	**92.02**	**96.92**	**90.78**	**92.97**	**92.84**

results are summarized in Table 7, with the top two performers highlighted in red and blue on the similar dataset. Table 7 shows the performance of different YOLOv12 versions. YOLOv12 achieved the highest overall accuracy and efficiency due to its attention-based feature extraction and optimized multi-scale fusion. These results demonstrate a measurable improvement in both precision and recall compared to its predecessors. Figure 9 shows the detection results on various newspaper images.

The findings clearly show that YOLOv12 performs better than all previous iterations in terms of accuracy (mAP@50 = 0.88) and generalization (mAP@50–95 = 0.86) without requiring an excessive amount of computing power. Notably, the proposed YOLOv12 demonstrates superior stability across different IoU thresholds and delivers a 5–7% increase in mAP@50 when compared to YOLOv10 and YOLOv11. This demonstrates YOLOv12's improved adaptability to multilingual scripts, visually complex page structures,

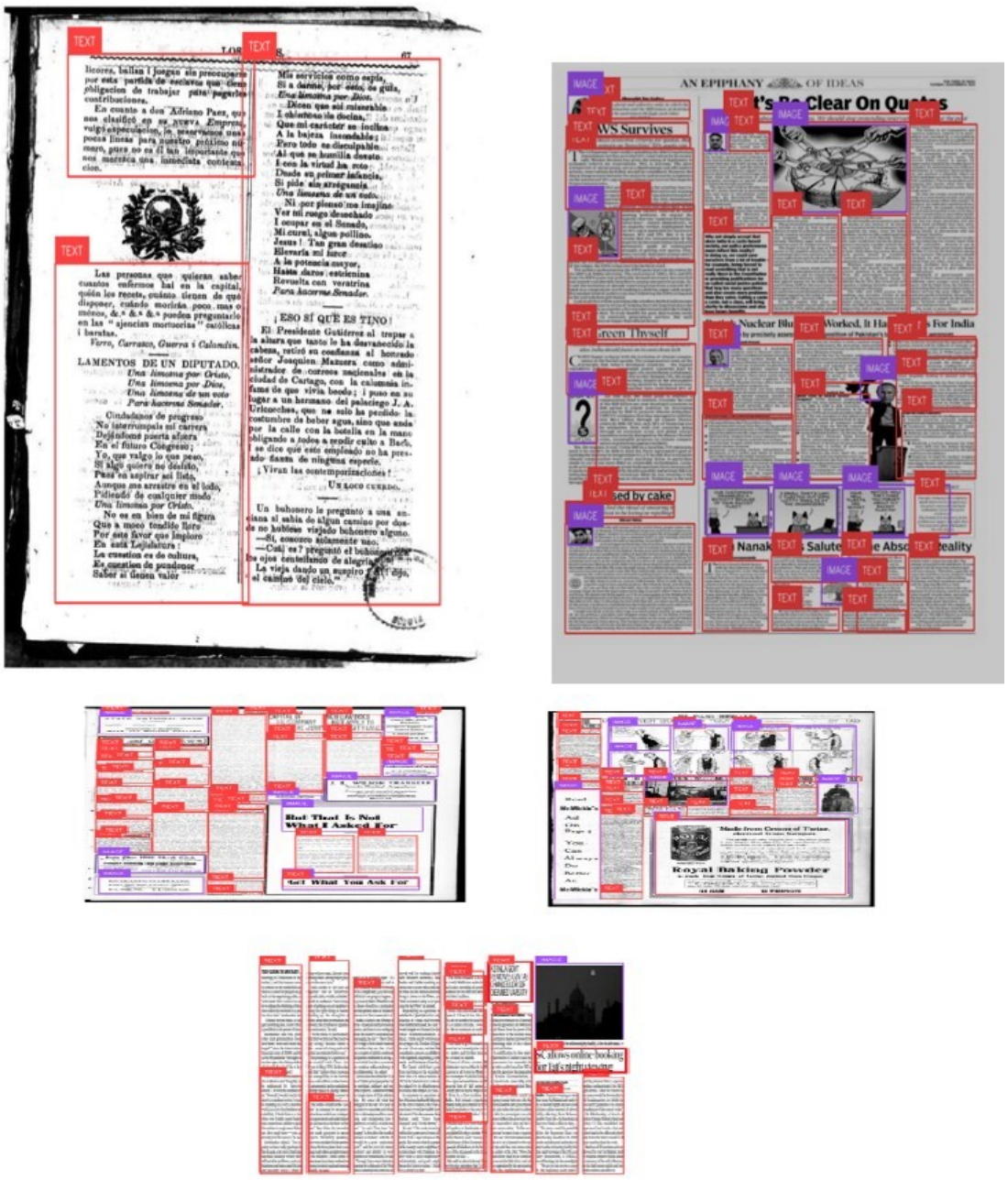


Figure 9. Layouts detection for different newspaper images.

and varied newspaper layouts.

The YOLO series outperforms region-based techniques like Faster R-CNN and Mask R-CNN, as Table 8 shows a steady increase in detection accuracy across models when evaluated on the same test dataset. With the greatest average accuracy (92.84%), the proposed YOLOv12 demonstrates exceptional capacity to recognize both textual and visual layout elements. Due to the variety of visual structures, the Image category performs the best (96.92%), whereas the Advertisement category exhibits somewhat lower accuracy (90.78%). Overall, the findings demonstrate YOLOv12's improved accuracy and robustness in complex newspaper layout recognition, suggesting that it is appropriate for tasks involving automated structural analysis and large-scale document digitization.

7 Conclusion

This paper presents an advanced and systematic framework for newspaper layout recognition employing the improved YOLOv12 model. Through the development of a custom multilingual dataset and the integration of attention-based architectural

enhancements, YOLOv12 demonstrates remarkable capability in accurately detecting and classifying heterogeneous newspaper components. The model achieved a mAP@50 of 0.88 and an average detection accuracy of 92.84%, outperforming established benchmarks such as Faster R-CNN, Mask RCNN, YOLOv10, and YOLOv11 in both precision and computational efficiency.

The results substantiate YOLOv12's robustness and generalization across diverse linguistic and structural formats, effectively addressing challenges inherent in multilingual scripts, irregular columnar layouts, and overlapping visual entities. Beyond achieving state-of-the-art detection accuracy, this study contributes a novel annotated dataset and a reproducible methodological framework for future research in document image analysis. Consequently, the proposed approach provides a substantial advancement toward intelligent document understanding, automated archival preservation, and scalable newspaper digitization within digital humanities and media informatics domains.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Breuel, T. M. (2008, January). The OCRopus open source OCR system. In *Document recognition and retrieval XV* (Vol. 6815, pp. 120-134). SPIE. [CrossRef]

[2] Namboodiri, A. M., & Jain, A. K. (2007). Document structure and layout analysis. In *Digital Document Processing: Major Directions and Recent Advances* (pp. 29-48). London: Springer London. [CrossRef]

[3] Ultralytics. (2025). *Ultralytics YOLO12: Attention-centric object detection*. Retrieved June 24, 2025, from https://github.com/ultralytics/yolo

[4] Kise, K. (2014). Page segmentation techniques in document analysis. In *Handbook of Document Image Processing and Recognition* (pp. 135-175). Springer, London.

[5] Binmahashen, G. M., & Mahmoud, S. A. (2019). Document layout analysis: A comprehensive survey. *ACM Computing Surveys, 52*(6), 1–36. [CrossRef]

[6] Sutheebanjard, P., & Premchaiswadi, W. (2010, April). A modified recursive xy cut algorithm for solving block ordering problems. In *2010 2nd International Conference on Computer Engineering and Technology* (Vol. 3, pp. V3-307). IEEE. [CrossRef]

[7] Breuel, T. M. (2002). Two geometric algorithms for layout analysis. In D. Lopresti, J. Hu, & R. Kashi (Eds.), *Document Analysis Systems V* (Lecture Notes in Computer Science, Vol. 2423, pp. 188-199). Springer, Berlin, Heidelberg. [CrossRef]

[8] Sun, H. M. (2006). Enhanced constrained run-length algorithm for complex layout document processing. *International Journal of Applied Science and Engineering, 4*(3), 297-309.

[9] Gutehrlé, N., & Atanassova, I. (2022). Processing the structure of documents: Logical layout analysis of historical newspapers in French. *Journal of Data Mining and Digital Humanities*. [CrossRef]

[10] Ali, D., Milleville, K., Verstockt, S., Van de Weghe, N., Chambers, S., & Birkholz, J. M. (2024). Computer vision and machine learning approaches for metadata enrichment to improve searchability of historical newspaper collections. *Journal of Documentation, 80*(5), 1031-1056. [CrossRef]

[11] Santos Júnior, E. S. D., Paixão, T., & Alvarez, A. B. (2025). Comparative performance of YOLOv8, YOLOv9, YOLOv10, and YOLOv11 for layout analysis of historical documents images. *Applied Sciences, 15*(6), 3164. [CrossRef]

[12] Lee, B. C. G., Mears, J., Jakeway, E., Ferriter, M., Adams, C., Yarasavage, N., ... & Weld, D. S. (2020, October). The newspaper navigator dataset: Extracting headlines and visual content from 16 million historic newspaper pages in chronicling america. In *Proceedings of the 29th ACM international conference on information & knowledge management* (pp. 3055-3062). [CrossRef]

[13] Deng, Q., Ibrayim, M., Hamdulla, A., & Zhang, C. (2024). The YOLO model that still excels in document layout analysis. *Signal, Image and Video Processing, 18*(2), 1539-1548. [CrossRef]

[14] Pfitzmann, B., Auer, C., Dolfi, M., Nassar, A. S., & Staar, P. (2022, August). Doclaynet: A large human-annotated dataset for document-layout segmentation. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining* (pp.

3743-3751). [CrossRef]

[15] Li, D. L., Lee, S. K., & Liu, Y. T. (2025). Printed document layout analysis and optical character recognition system based on deep learning. *Scientific Reports, 15*(1), 23761. [CrossRef]

[16] Shen, Z., Zhang, R., Dell, M., Lee, B. C. G., Carlson, J., & Li, W. (2021, September). Layoutparser: A unified toolkit for deep learning based document image analysis. In *International Conference on Document Analysis and Recognition* (pp. 131-146). Cham: Springer International Publishing. [CrossRef]

[17] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017, July). Feature Pyramid Networks for Object Detection. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 936-944). IEEE. [CrossRef]

[18] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788). [CrossRef]

[19] Smith, R. (2007, September). An overview of the Tesseract OCR engine. In *Ninth international conference on document analysis and recognition (ICDAR 2007)* (Vol. 2, pp. 629-633). IEEE. [CrossRef]

[20] Terven, J., Córdova-Esparza, D. M., & Romero-González, J. A. (2023). A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. *Machine learning and knowledge extraction, 5*(4), 1680-1716. [CrossRef]

[21] Barman, R., Ehrmann, M., Clematide, S., Oliveira, S. A., & Kaplan, F. (2021). Combining visual and textual features for semantic segmentation of historical newspapers. *Journal of Data Mining & Digital Humanities*, (HistoInformatics). [CrossRef]

[22] Almutairi, A. (2025). Newspaper elements detection and newspaper pages categorization using CNNs and transformers. *International Journal on Document Analysis and Recognition (IJDAR), 28*(2), 225-237. [CrossRef]

[23] Singh, V., & Kumar, B. (2014, January). Document layout analysis for Indian newspapers using contour based symbiotic approach. In *2014 International Conference on Computer Communication and Informatics* (pp. 1-4). IEEE. [CrossRef]

[24] Lombardi, F., & Marinai, S. (2020). Deep learning for historical document analysis and recognition—A survey. *Journal of Imaging, 6*(10), 110. [CrossRef]

[25] Zhao, H., Min, W., Wang, Q., & Wei, Z. (2023). Memory-efficient document layout analysis method using LD-net. *Multimedia Tools and Applications, 82*(3), 4371-4386. [CrossRef]

[26] Girdhar, N., Coustaty, M., & Doucet, A. (2024). Digitizing history: transitioning historical paper documents to digital content for information retrieval and mining—a comprehensive survey. *IEEE Transactions on Computational Social Systems, 11*(5), 6151-6180. [CrossRef]

[27] Roboflow. (n.d.). *Roboflow: Computer vision platform for image annotation and model training*. Retrieved June 24, 2025, from https://roboflow.com/

[28] Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(6), 1137–1149. [CrossRef]

[29] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2020). Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 42*(2), 386–397. [CrossRef]

[30] Jocher, G. (2020). *YOLOv5 by Ultralytics*. Retrieved June 24, 2025, from https://github.com/ultralytics/yolov5.

[31] Jocher, G., Chaurasia, A., & Qiu, J. (2023). *Ultralytics YOLO* (Version 8.0.0). Retrieved June 24, 2025, from https://github.com/ultralytics/ultralytics.

[32] Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., & Han, J. (2024). Yolov10: Real-time end-to-end object detection. *Advances in Neural Information Processing Systems, 37*, 107984-108011.

[33] Ultralytics. (2024). YOLOv11: The latest version of the YOLO series for object detection. Ultralytics Documentation. Retrieved June 24, 2025, from https://docs.ultralytics.com/models/yolo11/



Dr. Atul Kumar is an Assistant Professor in the Department of Computer Science at R.G.M. Government College, Joginder Nagar, District Mandi, India. He earned his Ph.D. from Punjabi University, Patiala, and has over 10 years of teaching experience. Dr. Kumar has authored 15 research papers in reputed journals and conferences. His research interests include Computer Science, Natural Language Processing (NLP), and Pattern Recognition, where his work has contributed significantly to advancing methods and applications in these domains. (Email: atulkmr02@gmail.com)



Dr. Gurpreet Singh Lehal is a Retired Professor from the Department of Computer Science, Punjabi University Patiala. Currently, he is working as a Senior Project Consultant at IIIT, Hyderabad. He has over three decades of contributions to computational linguistics, digital humanities, and Indic language technologies, particularly in Punjabi, Urdu and Hindi languages. His leadership spans major national and international projects in OCR, machine translation, transliteration, language modeling, and digital text processing, including large-scale government-funded initiatives. He has developed more than two dozen language software systems, supervised 23 PhD scholars, published over 200 indexed research papers, and served on influential committees such as ICANN panels and national standards bodies. His work has significantly advanced the technological ecosystem for Indian languages. (Email: gslehal@gmail.com)