Automatic Layout Detection in Historical Civil Records Using Deep Object Detection

Anonymous Author(s)

Anonymous Institution(s)

Abstract. Accurate layout detection in historical documents is a critical step in enabling automated transcription and large-scale analysis of archival records. This paper presents a deep learning-based approach for automatic layout detection in historical civil birth records from the Belfort archives, covering demographic data over a century. The layout detection of historical civil birth records represents a substantial challenge due to varying handwriting styles, overlapping components, and skewed marginal annotations. Therefore, a custom dataset was created through manual transcription and structured tagging, complemented by an automated annotation using a dedicated XML-generating tool. The YOLOv8 object detection model was fine-tuned through an iterative training strategy that incorporated model predictions and manual verification using a custom-built post-processing interface. The model achieved strong detection performance, with a $mAP_{0.5}$ of 0.9945 and $mAP_{0.5:0.95}$ of 0.9152, and was successfully applied to segment over 214,000 layout components across the full archival collection. The overall results justify the application of YOLOv8 for large-scale layout detection of heterogeneous historical documents. This research represents an initial development toward more accurate handwritten text recognition and automated information extraction from historical archives.

Keywords: Document layout analysis · Historical documents · Deep learning · Object detection · YOLO

1 Introduction

Historical civil records are essential for analyzing the demographic development of urban populations [31]. Belfort stands out as a French city with a particularly unique demographic history in the nineteenth century. Its significance grew notably after France's defeat by Prussia in 1871 and the annexation of Alsace-Lorraine by Germany. In the aftermath, Belfort experienced a sharp population increase, driven by the arrival of Alsatians who chose to remain on French soil or relocated there to work in the expanding textile and mechanical industries established after 1871 [7, 13]. As a result, the city serves as a remarkable case study for examining three major transformations: urban expansion, migratory movements, and shifting social dynamics, particularly in relation to sexuality.

Additionally, an analysis of Belfort's birth records provides substantial historical insights into diverse social and cultural practices [12]. These include evolving

patterns of cohabitation, midwives' roles and responsibilities, the frequency of births outside of marriage, the age of parents at the birth of their first child, and trends in the legal recognition of children. The records also shed light on how witnesses were chosen for official declarations, common naming practices for newborns, and whether births took place at home or in medical institutions [15]. Figure 1 illustrates a typical page from these civil birth registers.

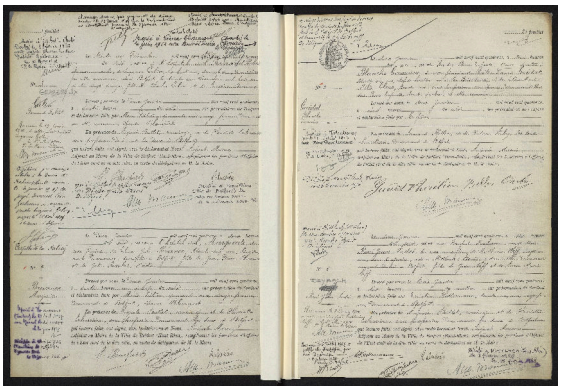


Fig. 1. Sample of Belfort civil registers of births.

To effectively investigate and address these historical concerns, it is crucial to create a comprehensive knowledge database [24]. This process involves transcribing archival content through two primary methods. The first method is manual transcription, which, although accurate, is costly and time-consuming. The second method is automatic transcription, which employs deep learning techniques [10].

Handwritten text recognition (HTR) is a process that automatically converts handwritten text from a digital image into a machine-readable format. The need for robust HTR methods is growing as it has become essential for transcribing many types of documents. These include historical archives, letters, general forms, and books [2]. The entire transcription process typically involves several stages, as shown in figure 2.

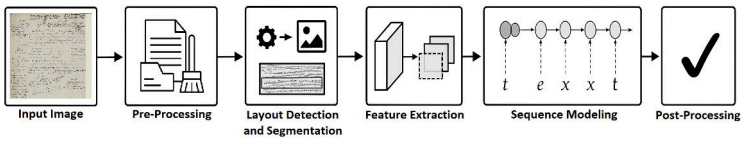


Fig. 2. The essential stages involved in the transcription of handwritten text images.

To identify and extract the structural components of historical archives. The YOLO (You Only Look Once) object detection algorithm offers a cutting-edge, real-time detection model known for its speed and accuracy [1]. Unlike traditional methods that rely on multiple stages, YOLO approaches object detection as a single regression task—predicting both bounding boxes and class labels in one go from the entire image. This streamlined process makes it especially well-suited for handling large volumes of high-resolution historical documents [21].

This study addresses the specific challenges encountered during the layout component detection phase, a critical step in facilitating the transcription of historical birth records from Belfort. The primary objective is to segment these document components automatically, which presents notable difficulties due to the presence of both handwritten and printed text, various writing styles, text overlapping, text skew, ink stains, and degradation. Applying deep learning models directly to full-page historical documents is particularly complex under these conditions [26]. Although Optical Character Recognition (OCR) technologies have seen substantial advancements in recent years, they continue to underperform when applied to such heterogeneous archival materials.

To mitigate these limitations, the YOLO object detection algorithm [14] was employed to identify and localize essential structural elements within the civil registers of Belfort. This structured layout detection enables a more systematic analysis of the documents and serves as a foundational step toward accurate transcription and automated information extraction.

The primary goal of this article is to propose and evaluate a robust, scalable approach for automatic layout detection in historical civil birth records, using deep object detection techniques. Rather than addressing handwritten text recognition directly, the focus is placed on accurately identifying and segmenting the main structural components of complex archival documents. By combining a custom-constructed historical dataset, an iterative YOLOv8 training strategy, and a dedicated post-processing tool for validation and correction, this work aims to demonstrate how layout detection can be reliably applied at the scale of an entire archival collection.

The remainder of this paper is organized as follows: Section 2 provides an overview of recent developments in the relevant literature. Section 3 discusses the specific characteristics and challenges of the Belfort civil birth registers and details the proposed methodology. The experimental results are summarized in Section 4. Finally, Section 5 concludes the paper and offers recommendations for future research avenues.

2 State-of-the-Art Overview

Building on prior work [2], this study addresses a significant need for a new deep learning method specifically designed to transcribe the Belfort civil registers of births (BCRB). Nowadays, systems possess the capability to analyze document layouts at various levels, such as text lines, paragraphs, and documents as a whole.

2.1 Historical Document Layout Analysis

Recent research has seen the emergence of diverse deep learning techniques tailored to historical document layout analysis. In [20, 8], authors conducted surveys, where they systematically reviewed contemporary methods of text line and baseline detection applied to historical documents, thereby highlighting both advances and persistent challenges. Other researchers [27] proposed an unsupervised method to create layout masks from unlabeled document images, particularly useful for under-annotated archival collections. While [5] demonstrated the

effectiveness of self-supervised learning (SSL) to successfully segment historical handwritten documents, achieving strong results with minimal annotated data. Their method highlights the potential of SSL to enhance the process of layout detection in archives. Authors of [11] contributed a new layout-classification benchmark for historical printed Hebrew books, facilitating model assessment in heritage contexts.

2.2 The YOLO Framework and Evolution

The You Only Look Once (YOLO) family of models has been one of the key factors in the advancement of real-time object detection since its initial release [22]. Unlike traditional multi-stage detectors like Regions with Convolutional Neural Networks (R-CNN) and Faster R-CNN, YOLO treats object detection as one single regression problem, where it predicts bounding boxes and class probabilities directly from full images in one evaluation step. Over time, the YOLO architecture has been subjected to numerous refinements, each iteration introduced improvements in accuracy, efficiency, and architectural design of the model [23, 6].

In 2023, Ultralytics introduced new features in YOLOv8 [14], which were the main selling points of the system in terms of performance, flexibility, and efficiency. These features include tasks such as detection, classification, segmentation, and pose estimation. Architecturally, YOLOv8 demonstrates higher performance across benchmarks such as COCO, while preserving real-time inference speeds. Moreover, its modular structure (defined via YAML configuration files) allows users to flexibly scale across different model variants (n, s, m, l, x), which trade accuracy for speed and parameter count. The architecture of the model can be described in three major components:

Backbone: A 3×3 stem convolution and stacked C2f modules (cross-stage partial concatenation). The bottleneck blocks use 3×3 convolutions, which have dual effect of better gradient flow and computational efficiency.

Neck: A hybrid Feature Pyramid Network (FPN) and Path Aggregation Network (PAN) architecture merges multi-scale feature maps, using lightweight concatenation operations that relax channel-size matching constraints, thereby reducing parameters while improving feature fusion.

Head: An anchor-free, decoupled detection head separates classification and bounding-box localization, eliminating anchor boxes. The regression of each detected object is predicted from its center point, making inference and Non-Maximum Suppression (NMS) more efficient.

In this work, we select YOLOv8l for its higher capacity, which is beneficial for complex handwritten layouts.

2.3 YOLO for Document Layout Detection

YOLO-based methods have also been widely adopted for general document layout detection across various modern domains. [29] applied YOLOv5 on images of documents to identify elements such as paragraphs, tables, and figures with high detection rates on a custom dataset. In one experiment [25], authors studied the efficiency of YOLOv5s and YOLOv5m in segmenting the layout of

the scans of books in Arabic and English. Moreover, DocLayout-YOLO framework [34] advances this approach by pre-training YOLOv8 on a synthetic dataset (DocSynth-300K), and recorded the highest speed and accuracy on the DocStructBench benchmark in comparison with state-of-the-art methods.

Recent YOLO Advances (YOLOv9–YOLOv11): Since the release of YOLOv8 in 2023, further versions have kept improving the framework. YOLOv10 -CBRC [33] achieved high-precision layout analysis of Tibetan newspaper images using the novel AbaTND dataset. [32] proposed SEMA-YOLO, a YOLOv11-based model that is capable of detecting small objects in remote sensing. It managed to achieve the best results of precision and recall on the DIOR dataset [16] by using shallow-layer enhancement and multi-scale adaptation, outperforming baseline YOLOv11 in lightweight deployment. These successive iterations demonstrate the rapid pace of development in the YOLO family, although YOLOv8 remains a widely adopted and stable version for document analysis research.

YOLOv8 for Historical Document Applications: Authors of [30] utilized YOLOv8 to identify characters in Greek papyri. Their approach combined YOLOv8 with DeiT (a transformer model) and SimCLR (self-supervised learning) to refine classification. In the ICDAR 2023 Competition on Greek papyri, their method outperforms other competitors at relaxed IoU thresholds. The work was extended to more than 4,500 images of the Oxyrhynchus Papyri, demonstrating the power of YOLOv8 in large-scale historical manuscript analysis. [18] proposed iForal, a system based on YOLOv8 for layout recognition, segmentation, and transcribing of medieval Portuguese manuscripts (67 royal charters). The system achieved high accuracy rates for layout detection and line segmentation.

Additionally, authors of [28] gathered a multimodal dataset of 72,081 wood-engraved illustrations from The Illustrated London News (1842-1890). They annotated 908 pages for a YOLOv8 model training and validation process. Subsequently, the model was able to pinpoint images on 56,699 pages with high accuracy, enabling large-scale multimodal retrieval through Open-CLIP embeddings combined with OCR for captions. This work is a demonstration of YOLOv8's capabilities for the extraction of high-precision visual content in historical periodicals. [26] conducted a comparative evaluation of YOLOv8, YOLOv9, YOLOv10, and YOLOv11 for historical document layout analysis. The study with archival record benchmark datasets concluded that YOLOv8 is still the most computationally efficient for large-scale deployment. Table 1 depicts the details and accuracy achieved by different YOLO-based approaches. Performance is reported in terms of standard object detection metrics, which are formally defined in Section 4.1.

3 Methodology

3.1 Belfort civil registers of births

The civil birth registers of the Belfort commune contain a total of 39,627 declarations, all written in French and digitized at a resolution of 300 dpi. These declarations cover births dated according to the Gregorian calendar from 1807 to 1919. Initially, all entries were fully handwritten; however, in 1885, a shift

Table 1. Summary of YOLO-based approaches for document layout analysis.

Author Year	YOLO Version	Dataset	Detected Components	P	R	mAP†	mAP‡
[29] (2023)	YOLOv5	153 pages (books, journals)	Title, Text, Image, Caption, Image_caption, Table, Table_caption	0.911	0.971	0.970	0.801
[25] (2024)	YOLOv5	950 scientific engineering book pages (Arabic & English)	Title, Subtitle, Text, Table, Table caption, Figure, Figure caption, List, Page number, Reference	0.927	0.920	0.945	–
[30] (2024)	YOLOv8	ICDAR 2023 Greek papyri Oxyrhynchus Papyri AL-PUB v2	Greek characters	0.932	0.986	0.932	0.514
[18] (2025)	YOLOv8	iForal (Portuguese medieval royal charters)	Text block, Image, Capital	0.980	0.950	0.980	–
[28] (2025)	YOLOv8	Illustrated London News (1842–1890)	Illustrations	–	–	0.964	0.920
[26] (2025)	YOLOv8	BHN, GBN, BlaLet GT	Text, Image, Graphic, Table	0.845	0.809	0.891	–
[26] (2025)	YOLOv9	BHN, GBN, BlaLet GT	Text, Image, Graphic, Table	0.754	0.712	0.757	–
[34] (2024)	YOLOv10	DocLayNet	Title, List-item, Table, Figure, Caption, Footnote, Formula, Reference, Header, Page-footer	–	–	0.797	0.934
[33] (2025)	YOLOv10	AbaTND	Text blocks, annotations, layout components	–	–	–	0.776
[26] (2025)	YOLOv10	BHN, GBN, BlaLet GT	Text, Image, Graphic, Table	0.949	0.495	0.692	–
[32] (2025)	YOLOv11	RS-STOD	Small vehicle, Large vehicle, Ship, Airplane, Oil tank	0.748	0.699	0.725	0.468
[26] (2025)	YOLOv11	BHN, GBN, BlaLet GT	Text, Image, Graphic, Table	0.854	0.801	0.871	–

P: Precision; **R**: Recall; **mAP**†: $\text{mAP}_{0.5}$; **mAP**‡: $\text{mAP}_{0.5:0.95}$.

occurred toward partially printed forms that included designated blank spaces for recording essential details about each birth.

This change towards partially printed declarations concerned nearly 57.5% of the total declarations. The registers up to 1902 are publicly available online via the municipal archives portal.[1]. Besides that, it was possible to grant access for complete registers, including records up to 1919, with formal authorization from the municipal archives of Belfort.

In the registers, each document includes mostly four declarations. However, the number of declarations may vary between three and six declarations per document within some registers. These birth declarations basically contain informa-

[1] `https://archives.belfort.fr/search/form/e5a0c07e-9607-42b0-9772-f19d7bfa180e` (accessed February 6, 2026)

tion such as registration numbers, names, dates, addresses, and other contextual information. An example of a declaration is shown in Figure 3, and Table 2 provides an overview of the structure and contents of a typical declaration in the archive.

Fig. 3. A sample birth declaration from the Belfort civil registers. (1) indicates the declaration number, which uniquely identifies the entry within the register. (2) marks the name of the individual at birth. (3) highlights the main paragraph of the text, which contains detailed information about the birth event. (4) refers to marginal annotations, typically used to record additional details such as marriage, death, or later amendments related to the individual.

Table 2. Structure and content of a birth declaration in the Belfort civil registers.

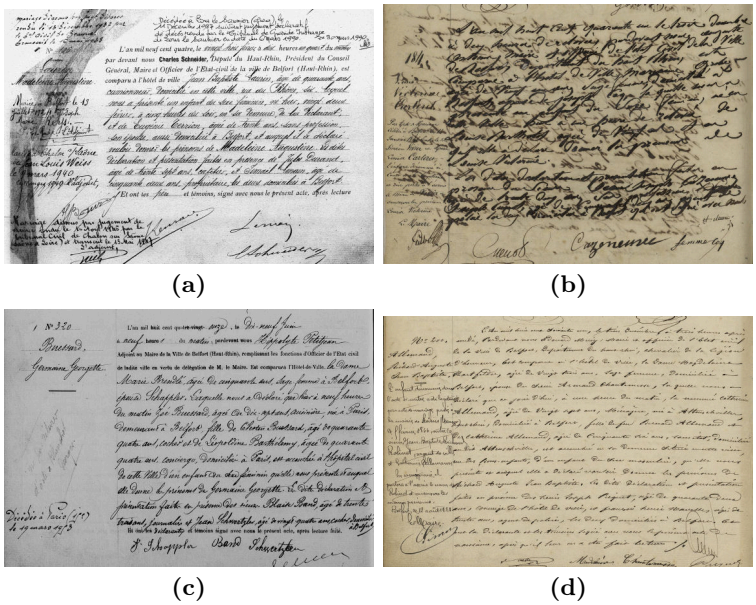
Structure	Content	Location
Number	Declaration registration number.	Left head margin
Individual names	first and last names of the person born.	Left head margin
Main body text	Time and date of declaration. Surname, first name, and position of the registering official. Surname, first name, age, profession, and address of the declarant. Surname(s), sex of the newborn, time and date of birth. Father's first name and surname. Mother's first name and surname, status (married or other), profession, and address. Witnesses' first names and surnames, ages, professions, and addresses. Mention of absence of signature or illiteracy of the declarant.	Center of the image
Marginal annotations	Official recognition of paternity/maternity: name and surname of the declarant, date of recognition. Marriage: date, wedding location, name and surname of spouse. Divorce: date, location. Death: date of death, place, date of the declaration of death.	Top-left to bottom-left, and bottom-left to bottom-right

Layout detection Challenges

The detection of components within the Belfort archive declarations presents several distinct challenges. Some documents consist of double-page spreads in which each page may contain either a single complete declaration or a varying number of declarations. Additionally, there are instances where a declaration begins on one page and continues onto the following page, further complicating the segmentation and analysis process.

The reading sequence of the components within a page is expected to follow a specific logical order: beginning with the declaration number, then the individual's names, followed by the main paragraph, and lastly, the marginal annotations. This sequence generally adheres to a left-to-right, top-to-bottom reading flow. However, the marginal annotations are typically added afterward to provide supplementary information. They are often positioned randomly in relation to the main paragraph of the declaration, and in some cases, they overlap with other components. Their variable placement presents additional challenges for consistent detection and interpretation.

Additionally, the documents exhibit a wide range of handwriting styles, including angular and spiky letters, as well as inconsistencies in character size. Consequently, this diversity often leads to text overlapping within the components of the document, complicating the layout detection and segmentation processes. Furthermore, many components display skew anomalies, including vertical aligned text that has been rotated by 90 degrees. The quality of the images is often compromised by the physical deterioration of the original documents, manifested through faded ink, smudging, ink blotches, and the yellowing of paper over time. Figure 4 illustrates examples of these challenges.



(a) **(b)** **(c)** **(d)**

Fig. 4. Examples illustrating challenges in Belfort birth declarations: (a) severe ink bleeding, water damage, overlapping strokes, and varying handwriting styles and ink intensities; (b) ink bleed-through and overlaid handwriting caused by reuse of paper or ink seeping from the reverse side; (c) faded marginal annotations and removed text; (d) marginal annotations overlapping with names and the main paragraph, disrupting structural boundaries between components.

Dataset

Following prior studies [4, 3], a dataset derived from the BCRB was constructed. The construction of this dataset involved two primary stages. In the first stage, selected declarations from the archival pages were manually transcribed. These declarations were carefully chosen to represent the full range of years covered by the archive, ensuring that all possible layout variations were accounted for during the model training phase. A structured tagging system was employed throughout the manual transcription process to accurately identify and organize the various components within each declaration. Table 3 provides an overview of the tag types used during this stage.

Table 3. Set of tags used during the manual transcription phase of the dataset (Source: [4, 3]).

Tag	Description
<begin>	Beginning of a declaration.
<text>...<\text>	Main paragraph of the declaration.
<margin>...<\margin>	Marginal text.
<ptext>...<\ptext>	Printed text.
<striped>...<\striped>	Struck-through (deleted) text.
<unreadable>...<\unreadable>	Unreadable text.
<added above>...<\added above>	Text added above the main text line.
<added below>...<\added below>	Text added below the main text line.
<page>	Start of a new page.

At present, a total of 319 transcription files in .txt format have been completed, encompassing 1,010 individual birth records. Each file contains up to four records, depending on the structure of the original document. In addition to the main record content, the transcriptions include 984 marginal annotations. Collectively, the dataset comprises 21,939 lines of text, which amount to 189,976 words and approximately 1,177,354 characters.

The second stage focused on the development of a dedicated tool designed to generate XML files, thereby facilitating the efficient preparation of transcribed declarations for input to the deep learning model. This tool systematically links each visual component within the images to its corresponding transcription at both the paragraph and line levels, using the predefined tags. Additionally, it records the precise coordinates of each component. The resulting XML files are structured to include the following attributes:

- Image Information: Includes metadata such as the image name, and page type (single or double page), as well as the height and width of the image.
- Reading Order: Assigns a unique index to each component to define the correct sequence for reading and processing.
- Declaration Number and Name: Contains the bounding box coordinates of these elements within the image, along with their corresponding transcribed text.
- Paragraphs and Marginal Annotations: Provides the coordinates and transcriptions of both the main body paragraphs and marginal annotations. This

section also includes detailed information about individual text lines within these components, such as unique identifiers, spatial coordinates, and the corresponding transcribed content.

The choice of XML as the annotation format was guided by several practical considerations. The historical documents handled in this study include multiple levels of organization, from pages and declarations down to layout regions and individual text lines, which XML proved convenient to represent. In addition, XML allows the reading order of components to be explicitly encoded alongside spatial information, which is important for complex historical layouts. While formats such as JSON could also be used, XML integrates more easily with the document analysis and digital humanities tools already employed during annotation and verification, without requiring additional adaptation. Table 4 offers a detailed overview of the dataset composition.

Table 4. Summary of generated XML files and associated image components.

Component type	Number of images
XML files generated	306
Number images	965
Name images	1,935
Primary paragraph images	994
Marginal annotation images	982

3.2 Layout Detection Model

In this study, we adopted the YOLOv8 object detection architecture. This version is stable and introduces several architectural enhancements over earlier YOLO models, optimizing it for high-resolution document image analysis as consistently demonstrated in the state of the art.

We utilized the YOLOv8l model and fine-tuned it on our custom dataset, which includes four target classes: Number, Name, main text Paragraph, and Marginal annotations. The dataset was organized in the standard YOLO format, with annotations specified in a data.yaml file that defined the number of classes, class labels, and paths to the training images.

Training was performed on CPU-based systems with an emphasis on memory efficiency and model generalization. Automatic mixed-precision (AMP) and caching mechanisms were enabled to enhance performance despite limited computational resources. The key parameters used during training are summarized in Table 5.

The final model configuration, outlined in Table 6, consists of 209 layers and approximately 43.6 million parameters, with a computational complexity of 165.4 GFLOPs. This high-capacity architecture is well-suited to capturing the intricate visual characteristics of handwritten historical documents.

3.3 Post Processing tool

To refine and validate the output of the YOLOv8-based object detection model, a custom post-processing tool was developed. This tool provides an interactive interface for manual review, editing, and reordering detected layout components

Table 5. YOLOv8l training configuration.

Category	Settings
Model	YOLOv8-Large (yolov8l.pt)
Training	100 epochs, batch size 4, image size 1024×1024
Optimizer	SGD, LR $0.003 \rightarrow 0.001 \rightarrow 0.0001$, momentum 0.937, weight decay 0.0005
Scheduler	Cosine LR, warmup 3 epochs
Loss weights	Box 7.5, Class 1.0, DFL 2.0
Thresholds	Confidence 0.75, IoU 0.75
Augmentation	HSV (0.015, 0.5, 0.3), translate 0.05, scale 0.2, shear 1.5, flip (H 0.2 / V 0.05)
Efficiency	AMP, cache enabled, 8 workers
Training setup	Backbone frozen (layer 0), validation enabled, resume from `last.pt`

Table 6. Summary of the YOLOv8 model architecture used for layout detection.

Model attribute	Value
Number of layers	209
Number of parameters	43,632,924
Number of gradients	0
FLOPs (GFLOPs)	165.4

extracted from the historical birth record images. The goal is to ensure both the spatial accuracy of bounding boxes and the logical reading order of components before segmentation.

The tool begins by processing the raw output of the YOLOv8l model. For each input image, the model produces a set of bounding boxes, each associated with a predicted class label (e.g., Number, Name, main text Paragraph, Marginal annotations) and a confidence score. To reduce noise, a maximum threshold is defined per class, limiting the number of retained detections. Detections are then sorted in descending order of confidence.

Moreover, the tool applies an automated region-based sorting strategy. This method divides each image into four quadrants using vertical and horizontal midlines and groups detections accordingly. Within each region, components are sorted top-to-bottom by their vertical positions, and within each class, they are ordered based on a fixed reading hierarchy: Number, Name, main text Paragraph, and Marginal annotations. This pre-sorting simulates the natural left-to-right, top-to-bottom reading flow.

Furthermore, a full-screen OpenCV-based interface enables users to interactively inspect and refine the detections, where users can add, edit, or delete bounding boxes. Also, assign labels via keyboard shortcuts. Additionally, define custom reading orders by clicking boxes in the desired sequence, and visualize region boundaries and current box indices in real time. The editor also offers an automatic reordering feature, which reapplies the region-based sorting if needed.

Upon confirmation, the tool saves each labeled component as an individual image crop, using a standardized filename format (<index>_<label>.jpg). It also generates a .txt file containing metadata for each component, including label, confidence score, and bounding box coordinates. This information is essential for further training data.

The tool supports batch processing of an entire folder of images. It is particularly useful for quality assurance, enabling manual correction of model predictions and preserving accurate structural layouts, which are critical for downstream transcription and data extraction tasks.

4 Results

4.1 Evaluation Metrics

To assess the performance of the model, we employed standard object detection metrics: Precision (P), Recall (R), Mean Average Precision at IoU 0.5 ($\text{mAP}_{0.5}$), and Mean Average Precision averaged over IoU thresholds from 0.5 to 0.95 ($\text{mAP}_{0.5:0.95}$). Precision and Recall are defined as [19]:

$$\text{P} = \frac{TP}{TP + FP}, \qquad \text{R} = \frac{TP}{TP + FN}, \tag{1}$$

where TP, FP, and FN denote true positives, false positives, and false negatives. We report $\text{mAP}_{0.5}$ (VOC, IoU = 0.5) [9] and $\text{mAP}_{0.5:0.95}$ (COCO, IoU $= 0.50 : 0.05 : 0.95$) [17]. Metrics were computed with Ultralytics YOLOv8 [14].

4.2 Validation Results

The proposed YOLOv8-based layout detection approach was applied to the historical Belfort civil birth dataset, focusing on four classes (entry numbers, names, main text paragraphs, and marginal annotations). The detection was carried out in successive stages, employing a fine-tuning strategy designed to incrementally improve model performance across a diverse and evolving dataset.

The initial training phase was based on the manually annotated subset of the dataset described in Section 3.1, which served as the foundation for the base model. Over the course of ten iterative stages, the model was applied to additional full-page records from the archive to predict layout components. These predictions were subsequently reviewed and corrected using the custom post-processing tool. The validated annotations were added to the training data for further fine-tuning. This iterative bootstrapping approach progressively expanded the dataset, enabling the model to adapt to diverse handwriting styles, complex layouts, and common degradation artifacts found in historical documents. The strategy proved effective in refining model performance with minimal manual annotation effort, leveraging model-assisted labeling to enhance both efficiency and accuracy.

Model performance was evaluated using the best-performing checkpoint. The model achieved a Precision of 0.9902, a Recall of 0.9904, an of $\text{mAP}_{0.5}$ of 0.9945, and an $\text{mAP}_{0.5:0.95}$ of 0.9152. Training results, presented in Table 7, demonstrate the model's strong capability to accurately detect all four layout components with a high degree of reliability.

Figure 5 presents further insight into the model's performance. These include F1-Confidence, Precision-Confidence, Precision-Recall, and Recall-Confidence curves. Finally, Figure 6 presents the final layout detection performance on the BCRB dataset and provides a contextual comparison with representative YOLOv8-based studies.

Table 7. Evaluation results for the YOLOv8l training runs.

Run	Precision	Recall	$mAP_{0.5}$	$mAP_{0.5:0.95}$
1	0.9359	0.9175	0.9377	0.6632
2	0.9225	0.9291	0.9483	0.6616
3	0.9361	0.9440	0.9544	0.6688
4	0.8909	0.9363	0.9463	0.7279
5	0.9504	0.9016	0.9229	0.7529
6	0.9617	0.8018	0.8749	0.7221
7	0.9238	0.9377	0.9504	0.7354
8	0.9818	0.9808	0.9907	0.8555
9	0.9702	0.9832	0.9904	0.8563
10	0.9902	0.9904	0.9945	0.9152

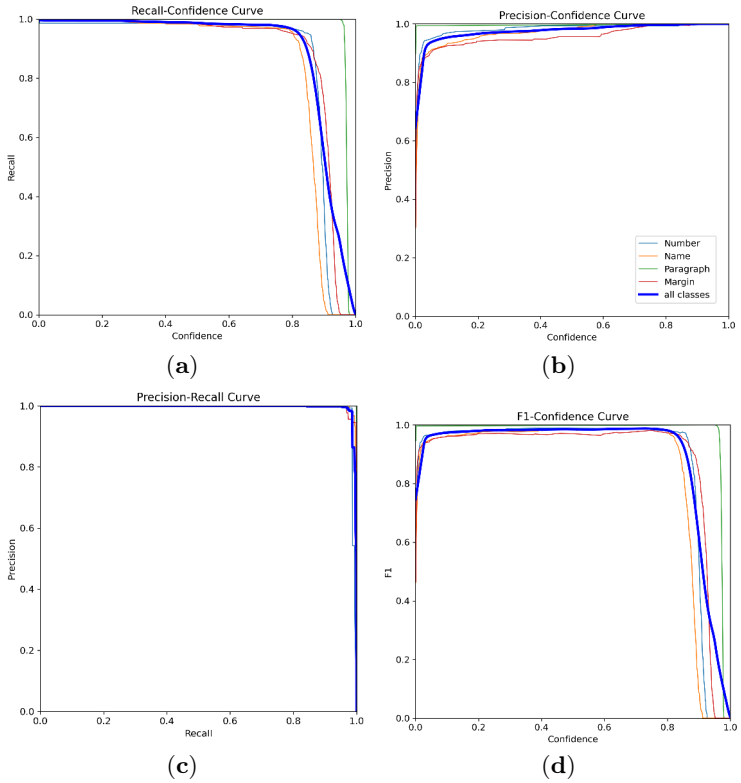


Fig. 5. Evaluation curves of the layout detection model. (a) Recall–confidence curve. (b) Precision–confidence curve. (c) Precision–recall curve. (d) F1–confidence curve.

4.3 Application to Full Archive

The trained model was applied to the full archive collection, comprising 134 archival volumes. In total, the system identified 40,218 numbers, 87,303 names, 45,075 main text paragraphs, and 41,608 marginal annotations, resulting in 214,204 segmented images. For each document, a corresponding .txt file was generated, containing detailed information about each segment, including its

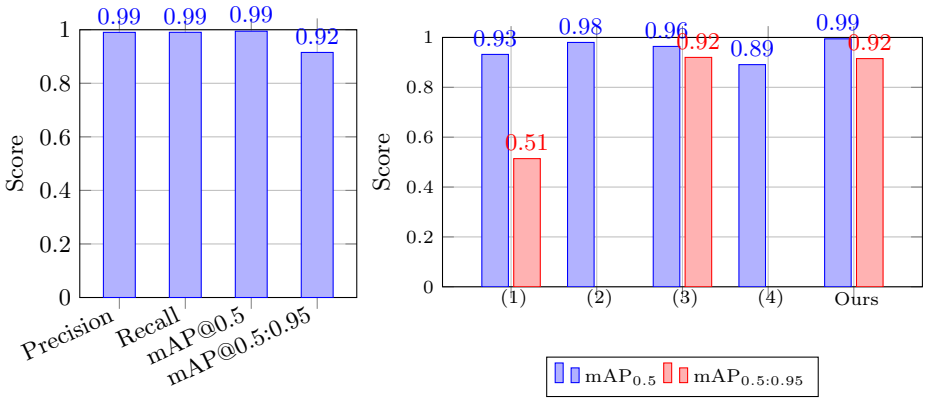


Fig. 6. (a) Layout detection performance of the proposed YOLOv8 model on the BCRB dataset. (b) Contextual comparison with representative YOLOv8-based studies: (1) Turnbull et al. [30], (2) Matos et al. [18], (3) Smits et al. [28], and (4) Santos et al. [26]. Reported values are reproduced from the respective publications and are provided for contextual reference only; differences in datasets, label definitions, and evaluation protocols prevent direct benchmarking.

position within the reading order. All segments and their respective reading sequences were manually verified using the post-processing tool described in Section 3.3.

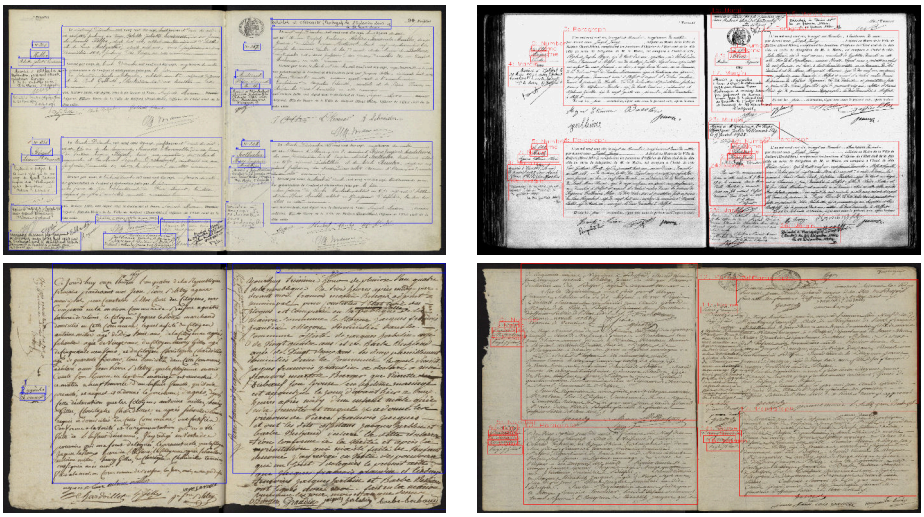


Fig. 7. Examples of layout detection results on archival birth records, including successful detections (top row) and representative failure cases (bottom row).

Although the model demonstrated strong overall performance, it occasionally failed to accurately identify relevant components in some cases. These errors are primarily associated with challenging input conditions, including heavily skewed marginal annotations, overlapping text, and degraded or low-contrast document

quality. Such challenges are likely attributable to the limited representation of these edge cases in the training dataset. Furthermore, issues were observed in identifying the reading order of components. All identified errors were subsequently addressed through manual verification using the post-processing tool described in Section 3.3. Figure 7 presents representative examples of successful layout detection results on the archive collection, along with illustrative failure cases encountered under challenging conditions.

5 Conclusions

In this paper, we presented an automated methodology for layout detection in historical civil birth records at the archives of Belfort through a YOLOv8-based deep learning model. The documents spanning over 100 years of officious records expose various structural, textual, and visual complications, skewed annotations, and overlapping text and components, which create challenges for the traditional layout detection methods.

To address those complex challenges, a unique dataset was developed through a manual transcription, structured tagging, and automated annotation using a dedicated XML generator. We fine-tuned the YOLOv8 model in a progressive training strategy that integrated model predictions with iterative human verification using a custom-built post-processing tool. This approach enabled efficient scaling of training data while maintaining annotation accuracy.

The final model demonstrated strong performance, indicating its robustness in detecting complex document layouts. Ultimately, the model and accompanying tools facilitated the complete segmentation of the Belfort birth registers. These structured outputs lay the foundation for more accurate handwritten text recognition research efforts and historical data extraction. Future work will focus on integrating the layout detection outputs with HTR systems and adapting the model to generalize to more historical records and other archival datasets.

References

1. Ali, M.L., Zhang, Z.: The yolo framework: A comprehensive review of evolution, applications, and benchmarks in object detection. Computers **13**(12), 336 (2024)
2. AlKendi, W., Gechter, F., Heyberger, L., Guyeux, C.: Advancements and challenges in handwritten text recognition: A comprehensive survey. Journal of Imaging **10**(1), 18 (2024)
3. Alkendi, W., Gechter, F., Heyberger, L., Guyeux, C.: Belfort birth records transcription: Preprocessing, and structured data generation. In: 4th International Conference on Image Processing and Vision Engineering. pp. 32–43. SCITEPRESS-Science and Technology Publications (2024)
4. AlKendi, W., Gechter, F., Heyberger, L., Guyeux, C.: Optimized data structuring and preprocessing techniques for belfort civil registers of birth transcription. SN Computer Science **6**(5), 550 (2025)
5. Baloun, J., Prantl, M., Lenc, L., Martínek, J., Král, P.: On self-supervision in historical handwritten document segmentation. International Journal on Document Analysis and Recognition (IJDAR) pp. 1–16 (2025)
6. Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M.: Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934 (2020)

7. Delsalle, P.: Histoires de familles: Les registres paroissiaux et d'état civil, du Moyen Âge à nos jours: Démographie et généalogie (Family History: Parish and Civil Status Registers, from the Middle Ages to the Present Day: Demography and Genealogy). Presses universitaires de Franche-Comté, Besançon (2009)
8. Ding, Y., Han, S.C., Lee, J., Hovy, E.: Deep learning based visually rich document content understanding: A survey. arXiv preprint arXiv:2408.01287 (2024)
9. Everingham, M., Eslami, S.A., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes challenge: A retrospective. International journal of computer vision **111**(1), 98–136 (2015)
10. Garrido-Munoz, C., Rios-Vila, A., Calvo-Zaragoza, J.: Handwritten text recognition: A survey. arXiv preprint arXiv:2502.08417 (2025)
11. Gogawale, S., Bambaci, L., Kurar-Barakat, B., Shapira, D.V., Ezra, D.S.B., Dershowitz, N.: Netlay: Layout classification dataset for enhancing layout analysis. Magazén: International Journal for Digital and Public Humanities (2024)
12. Gourdon, V.: L'histoire sociale de la famille en France à l'époque moderne et au XIXe siècle : traditions historiographiques et renouvellements thématiques (The Social History of the Family in France in the Modern Era and the 19th Century: Historiographical Traditions and Thematic Renewals). In: García González, F., Guzzi-Heeb, S. (eds.) Historia de la familia, historia social. Experiencias de investigación en España y en Europa (siglos XVI-XIX) (History of the Family, Social History. Research Experiences in Spain and Europe (16th-19th Centuries)), pp. 167–193. Ediciones TRea/Ediciones de la Universidad de Castilla-La Mancha, Gijón (2023)
13. Heyberger, L.: L'industrialisation de Belfort : une conséquence positive du siège de 1870-1871 ? Approche par l'histoire anthropométrique (The Industrialization of Belfort: A Positive Consequence of the Siege of 1870-1871? An Approach Through Anthropometric History). In: Belot, R. (ed.) 1870 de la guerre à la paix (1870 from War to Peace), pp. 207–217. Hermann, Strasbourg-Belfort, Paris (2013)
14. Jocher, G., Chaurasia, A., Qiu, J.: Ultralytics YOLOv8. `https://github.com/ultralytics/ultralytics` (2023), version 8.0.0. License: AGPL-3.0. Accessed: 2025-08-02
15. Laslett, P., Oosterveen, K., Smith, R.M.: Bastardy and Its Comparative History: Studies in the History of Illegitimacy and Marital Nonconformism in Britain, France, Germany, Sweden, North America, Jamaica, and Japan. Edward Arnold, London (1980)
16. Li, K., Wan, G., Cheng, G., Meng, L., Han, J.: Object detection in optical remote sensing images: A survey and a new benchmark. ISPRS journal of photogrammetry and remote sensing **159**, 296–307 (2020)
17. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: European conference on computer vision. pp. 740–755. Springer (2014)
18. Matos, A., Almeida, P., Correia, P.L., Pacheco, O.: iforal: Automated handwritten text transcription for historical medieval manuscripts. Journal of Imaging **11**(2), 36 (2025). https://doi.org/10.3390/jimaging11020036, `https://www.mdpi.com/2313-433X/11/2/36`
19. Powers, D.M.: Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. arXiv preprint arXiv:2010.16061 (2020)
20. Rabaev, I., Litvak, M.: Recent advances in text line segmentation and baseline detection in historical document images: a systematic review. International Journal on Document Analysis and Recognition (IJDAR) pp. 1–37 (2025)

21. Ramos, L.T., Sappa, A.D.: A decade of you only look once (yolo) for object detection. arXiv preprint arXiv:2504.18586 (2025)
22. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 779–788 (2016)
23. Redmon, J., Farhadi, A.: Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767 (2018)
24. Rosenzweig, R.: Clio Wired: The Future of the Past in the Digital Age. Columbia University Press, New York (2011)
25. Salim, H., Mustafa, F.S.: A comprehensive evaluation of yolov5s and yolov5m for document layout analysis. European Journal of Interdisciplinary Research and Development **23**, 21–33 (2024)
26. Santos Júnior, E.S.d., Paixão, T., Alvarez, A.B.: Comparative performance of yolov8, yolov9, yolov10, and yolov11 for layout analysis of historical documents images. Applied Sciences **15**(6), 3164 (2025). https://doi.org/10.3390/app15063164, https://www.mdpi.com/2076-3417/15/6/3164
27. Sheikh, T.U., Shehzadi, T., Hashmi, K.A., Stricker, D., Afzal, M.Z.: Unsupdla: Towards unsupervised document layout analysis. In: International Workshop on Document Analysis Systems. pp. 142–161. Springer (2024)
28. Smits, T., Warner, B., Fyfe, P., Lee, B.C.G.: A fully-searchable multimodal dataset of the illustrated london news, 1842–1890. Journal of Open Humanities Data **11**(1), 10 (2025). https://doi.org/10.5334/johd.284, https://openhumanitiesdata.metajnl.com/articles/10.5334/johd.284
29. Sugiharto, H., Silviana, Y., Nurpazrin, Y.S.: Unveiling document structures with yolov5 layout detection. arXiv preprint arXiv:2309.17033 (2023)
30. Turnbull, R., Mannix, E.: Detecting and recognizing characters in greek papyri with yolov8, deit and simclr. International Journal on Document Analysis and Recognition (IJDAR) **28**(2), 277–285 (2025). https://doi.org/10.1007/s10032-024-00504-8, https://doi.org/10.1007/s10032-024-00504-8
31. Vézina, H., St-Hilaire, M., Bournival, J.S., Bellavance, C.: The linkage of microcensus data and vital records: An assessment of results on quebec historical population data (1852–1911). Historical Methods: A Journal of Quantitative and Interdisciplinary History **51**(4), 230–245 (2018)
32. Wu, Z., Zhen, H., Zhang, X., Bai, X., Li, X.: Sema-yolo: Lightweight small object detection in remote sensing image via shallow-layer enhancement and multi-scale adaptation. Remote Sensing **17**(11), 1917 (2025)
33. Wu, Z., Wang, W., Li, H.: Yolov10-cbrc: A high-precision document image layout analysis model. Journal of King Saud University Computer and Information Sciences **37**(6), 1–21 (2025)
34. Zhao, Z., Kang, H., Wang, B., He, C.: Doclayout-yolo: Enhancing document layout analysis through diverse synthetic data and global-to-local adaptive perception. arXiv preprint arXiv:2410.12628 (2024)