e-ISSN 2724-3923

magazén
Vol. 5 – Num. 2 – December 2024

NetLay: Layout Classification Dataset for Enhancing Layout Analysis

Sharva Gogawale
Tel Aviv University, Israel

Luigi Bambaci
École Pratique des Hautes Études, Paris, France

Berat Kurar-Barakat
Tel Aviv University, Israel

Daria Vasyutinsky Shapira
Tel Aviv University, Israel

Daniel Stökl Ben Ezra
École Pratique des Hautes Études, Paris, France

Nachum Dershowitz
Tel Aviv University, Israel

Abstract Within the domain of historical document image analysis, the process of identifying the spatial structure of a document image is an essential step in many document processing tasks, such as optical character recognition and information extraction. Advancements in layout analysis promise to enhance efficiency and accuracy using specialized models tailored to distinct layouts. We introduce NetLay, a new dataset for benchmarking layout classification algorithms for historical works. It consists of over 1,300 images of pages of printed Hebrew (or Hebrew-character) books in a variety of styles, categorized into four different classes based on their layout (the number of text columns and regions). Ground truth was crafted manually at the page level. Furthermore, we conduct an in-depth performance evaluation of various layout classification algorithms, which are based on deep-learning models that learn to extract spatial features from images. We evaluate our algorithms on NetLay and achieve state-of-the-art results on the task of layout classification for historical books.

Keywords Historical document analysis. Layout analysis. Layout classification. Multi-label classification. Convolutional neural networks. Deep learning.

Edizioni Ca'Foscari

Peer review

Submitted	2024-04-04
Accepted	2024-09-23
Published	2024-12-17

Open access

Citation Gogawale, S. et al. (2024). "NetLay: Layout Classification Dataset for Enhancing Layout Analysis". *magazén*, 5(2), 223-240.

DOI 10.30687/mag/2724-3923/2024/02/003

1 Introduction

Numerous institutions and libraries worldwide are digitizing their archives to democratize access and safeguard them from physical deterioration. This calls for an ability to perform primary processing of numerous texts automatically. In the field of document image processing,[1] benchmark datasets with corresponding ground truth are essential for evaluating, developing, and comparing algorithms, as they also drive the creation of new approaches to address emerging challenges. Recent advancements in image analysis and computer vision have automated most of the tasks in the pipeline for automatic document analysis. Document layout analysis acts as a crucial preliminary step for various document image analysis tasks. Advancements in this field hold immense potential for boosting efficiency and accuracy through the development of specialized models tailored to diverse document layouts. Document image processing encompasses classical machine learning techniques, requiring meticulous feature selection, and deep neural network-based approaches where features are inherently learned within the network. While both techniques play a role, recent breakthroughs in image classification have been primarily driven by deep-learning methods.

A key advantage deep learning offers over traditional approaches lies in its inherent ability to extract features directly from the data. This not only liberates paleographers from spending weeks or months on feature selection but also empowers neural networks to uncover novel and intricate features that might evade even the most discerning human expert. A critical aspect of this endeavour is addressing the challenges inherent in ancient and medieval handwriting studies, necessitating the training of specialized models tailored to distinct layouts. However, the scarcity of diverse stylistic representations poses challenges for developing multi-domain general layout analysis, compounded by the predominance of datasets containing Latin script.

Addressing these disparities is imperative for advancing historical document analysis research and development, particularly in historical document layout analysis. However, the current landscape of available datasets suffers from two major limitations that hinder progress in historical document analysis. Firstly, the lack of stylistic

This research was funded in part by the European Union (ERC, MiDRASH, Project No. 101071829). Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

1 We use the term 'document' in its general sense, ranging from literary works to personal notes, from full-length books to individual pages.

diversity can significantly hamper the development of general layout analysis methods capable of functioning effectively across multiple domains. Secondly, the vast majority of existing datasets primarily cater to documents in English, neglecting the inherent differences in text features present in other languages. This disparity can lead to problems when applying these methods to languages like Hebrew, highlighting a critical gap in resources dedicated to historical document layout analysis datasets. While significant strides have been made in the domain of modern documents, addressing this discrepancy is paramount to propelling research and development forward in the field of historical document analysis.

Long-standing efforts have been devoted to creating layout analysis datasets, with the huge dataset PubLayNet (Zhong, Tang, Jimeno Yepes 2019) for contemporary documents emerging recently. However, existing datasets tailored for historical documents remain limited in scope. The majority of openly available historical document layout datasets mostly address more popular scripts and languages. The Europeana Newspapers Project (ENP) (Clausner et al. 2015) contains common European languages like Dutch, English, German, etc., from the seventeenth century onward, and contains 500 page images. The PRImA Layout Analysis Dataset (Antonacopoulos et al. 2009) places emphasis on magazines and technical/scientific publications, the majority in Latin script. Addressing these disparities and incorporating the representation of less common and older languages – like Hebrew – in datasets are imperative for advancing historical document analysis research and development.

Before we address the more complicated question of Hebrew 'manuscript' layout, we must solve the problem of automatic layout classification for 'printed' Hebrew books. Hebrew books often have nonstandard layouts, multiple languages (Hebrew/Aramaic; Hebrew/Yiddish, etc.) per page written in the same script and alphabet, and different script type-modes per page (Ashkenazi square plus Oriental semi-cursive ["Rashi"]). Sometimes, different text fields are not clearly distinguishable.

To address these challenges, we present NetLay, a dataset containing 1352 pages, taken from books with diverse layouts sourced from the collection of the National Library of Israel (NLI). In addition, we propose several benchmark techniques to perform layout classification. We implement various deep-learning models. We also propose a multi-label encoding scheme based on the spatial and global interdependencies of distinct layout elements.

The remainder of this paper is organized as follows: Section 2 is a short survey of the related literature. Section 3 explains the properties of the dataset proposed. Section 4 describes various methods used for layout classification. In Section 5 we evaluate several deep-learning classifiers and present our results.

2 Related Work

Understanding the layout of a document serves as a preliminary step for various document image processing tasks. These tasks include information retrieval, page segmentation, word spotting, and optical character recognition (OCR), which aims to extract meaningful textual information from these images. Breuel (2003) proposed novel algorithms and statistical methods for flexible page layout analysis, combining globally optimal geometric algorithms with robust statistical models and meticulous engineering techniques. Page segmentation algorithms typically fall into two categories: bottom-up and top-down. Bottom-up algorithms work in a hierarchical manner to group elements such as pixels, patches, or connected components into progressively larger regions. In contrast, top-down algorithms divide the entire page into regions in a single step. Many of the early page layout analysis methods often relied on assumptions about document structure and employed a top-down approach, particularly for well-formatted, modern binary (black and white) documents. Typically, these methods rely on mathematical morphology and connected components (Alarcón Arenas, Yari, Meza-Lovon 2018), Voronoi diagrams (Kise, Sato, Iwata 1998), or run length smearing algorithms (Wong, Casey, Wahl 1982).

There are, however, also numerous other techniques that don't fit neatly into one of the above categories. These so-called mixed or hybrid approaches aim to merge the efficiency of top-down methods with the robustness of bottom-up ones. Corbelli et al. (2016) proposed a hybrid layout analysis pipeline, integrating both top-down and bottom-up approaches. They employ the X-Y cut algorithm and a support vector machine (SVM) classifier for illustration detection, coupled with a convolutional neural network (CNN) and random forest classifier for content classification identifying different classes of layout entities. Pixel classification approaches have also been explored for page segmentation. Wei et al. (2013) framed the problem as pixel classification, where each pixel is represented as a feature vector based on the image's color. They employed techniques like Gaussian mixture models (GMM), multi-layer perceptrons (MLP), and SVM to classify pixels into categories such as decoration, background, periphery, and text pixels. Chen et al. (2014) subsequently improved upon this work by incorporating more comprehensive features encompassing texture and colour properties like smoothness, Laplacian, Gabor dominant orientation histograms, local binary patterns, and colour variance.

With the onset of deep learning, many authors have addressed the problem of layout segmentation and analysis using different deep neural network configurations. Borges Oliveira and Viana (2017) introduced a novel one-dimensional CNN approach for rapid automatic

layout detection of structured documents. Barakat and El-Sana (2018) presented a binarization-free method for page layout analysis of historical Arabic manuscripts, training an FCN to predict the class of each pixel and segmenting main text and side text regions. Kosaraju et al. (2019) proposed DoT-Net, a texture-based CNN for document layout analysis that can capture textural variations among the multiclass regions of documents. Alaasam, Kurar and El-Sana (2019) proposed a Siamese network-based layout analysis method tailored for challenging historical Arabic manuscripts. Da et al. (2023) introduced a two-stream vision grid transformer for layout analysis, conducting visual pre-training in two stages utilizing 2D token-level and segment-level understanding.

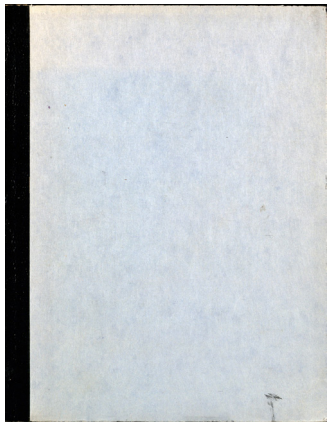
Although layout analysis and segmentation have been extensively explored, layout classification remains relatively understudied. This process involves categorizing documents based on their spatial arrangement, aiming to comprehend the overall layout of content within a document. This understanding serves as a cornerstone for the development of advanced algorithms for segmentation and OCR. Hu, Kashi and Wilfong (1999) introduced interval encoding, a novel feature set for capturing layout information. They utilize this encoding within an HMM framework for fast document image classification based solely on spatial layout.

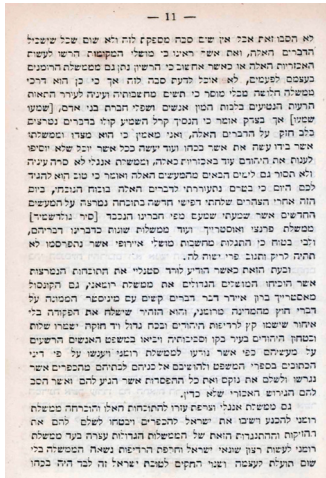
3 Dataset

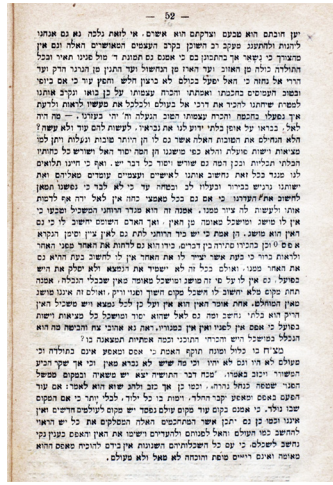
There is a critical necessity of implementing a layout classifier to augment the efficacy of dedicated models used in transcription systems like eScriptorium (Kiessling et al. 2019). To that end, we sourced images from the digital collections of NLI, tailored specifically for this task. High-resolution images of pages in the NetLay dataset were curated from a random selection of printed Hebrew books at NLI. From each book, one page image was carefully chosen for inclusion in the dataset. The dataset includes a total of 1352 images of single pages or facing pages. It is balanced and comprises the following classes: no text (“empty”), single column, two columns (occasionally on facing pages), and complex layout (three or more regions, or regions with insets), with 300, 442, 300, and 310 samples, respectively, for each class.

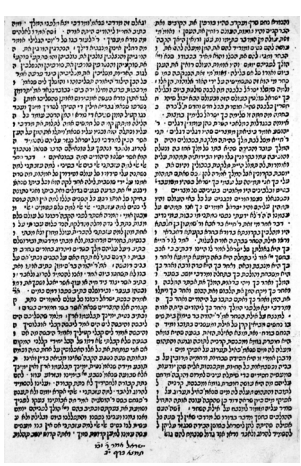
Facing pages, each containing one column, are usually one continuous work, but may also be two related works, one on even numbered pages and the other on odd ones. Two-column text may be read across both columns (as in poetry, for example), or column by column, or they may be two works side by side – in the same language or in two (perhaps a translation or commentary), in the same font or not. Complex layouts often contain separate, but related, works by different authors **[fig. 1]**.

e-ISSN 2724-3923









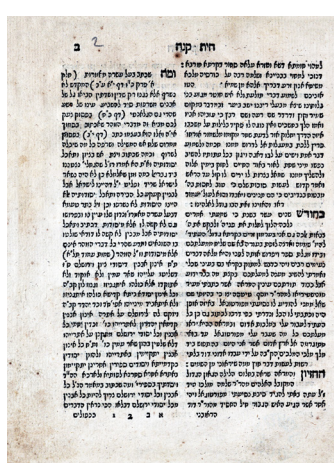


Figure 1 Document samples from NetLay: (a) no text; (b) single column; (c) two column; (d) complex layout. The figure contains illustrative examples of document images representing each class within the dataset. The dataset is publicly available at `https://github.com/TAU-CH/midrash_layout_classification_using_multilabel_vgg/tree/main/data`

4 Methods

The challenge of layout class identification presents itself as an image classification task, where the goal is to assign a specific class to a given document image. Given the complexity and variability of layouts, employing deep-learning models emerges as the most effective strategy for image classification tasks. Therefore, our approach uses deep-learning-based models to accurately categorize document images into distinct layout structures. In this section, we outline the experimental setup, including model architecture, training methodology, and evaluation procedures. We adopt state-of-the-art deep-learning models tailored for image classification tasks. To assess the performance of our approach, we conduct several benchmark experiments. These experiments aim to evaluate the efficacy of the proposed deep-learning models in accurately classifying layout structures. To ensure a robust evaluation, we divided our dataset into three distinct subsets: training (80%), validation (10%), and testing (10%). This split allows for effective model training, hyperparameter tuning, and unbiased performance evaluation. All the experiments for training the deep-learning models were conducted on a machine equipped with an NVIDIA Titan T4 GPU with 15 GB of memory.

Predictions are evaluated based on four standard performance metrics: accuracy, precision, recall, and F1-score.

We employ two methods for the task of document image layout classification.

1.1 Single Label Classification

Single-label classification involves assigning one class label to each instance from a predefined set of classes. In the context of document layout classification, our objective is to categorize layouts into four distinct classes: no text, single column, double column, or complex.

Below, we explore various architectures and propose methods employed for this task.

EfficientNetV2 We utilize EfficientNetV2 (Tan, Le 2021) for spatial feature extraction, pretrained on the ImageNet dataset. The core architecture employs the mobile inverted bottleneck (MBConv) (Sandler et al. 2018), with squeeze and excitation optimization.

In the EfficientNet family, comprising models from EfficientNet B0 to B7 (Tan, Le 2019) which employs mobile inverted bottleneck convolution (MBConv) with squeeze and excitation optimization. The variations can be seen in MBConv block count, width, depth, resolution, and overall size of the model. EfficientNetV2 introduces enhancements like fused-MBConv blocks alongside regular MBConv blocks, which lead to higher accuracies with fewer parameters. EfficientNetsV1s demonstrate adaptability through transfer learning, excelling when trained on diverse datasets. However, challenges such as slow training with large image sizes and inefficiencies in early layers due to depthwise convolutions are evident. Addressing these concerns, EfficientNetV2 introduced novel design elements and employs training-aware neural architecture search and scaling strategies to jointly optimise model accuracy, training speed, and parameter size.

Table 1 The multi-label encoding scheme

Class	Page width text line	Half page width text time	Page height vertical separator	Half page height vertical separator	Multiple fonts
Empty	0	0	0	0	0
Single column	1	0	0	0	0
Two columns	0	1	1	0	0
Complex layout	1	1	0	1	1

Table 2 Performance metrics (accuracy, precision, recall, and F1), for each class, using EfficientNetV2, ViT, and VGG16 with multi-label encoding

Method	Class	Accuracy	Precision	Recall	F1 Score
Efficient-Net	0 (Empty)	98.50%	0.94	1.00	0.97
	1 (Single column)	93.98%	0.97	0.84	0.90
	2 (Two columns)	95.49%	0.83	1.00	0.91
	3 (Complex layout)	93.98%	0.90	0.84	0.87
ViT	0 (Empty)	99.25%	1.00	0.97	0.98
	1 (Single column)	94.78%	0.89	0.95	0.92
	2 (Two columns)	98.51%	0.94	1.00	0.97
	3 (Complex layout)	94.03%	0.93	0.81	0.86
VGG16	0 (Empty)	99.26%	0.97	1.00	0.98
	1 (Single column)	99.26%	1.00	0.98	0.99
	2 (Two columns)	98.53%	1.00	0.93	0.96
	3 (Complex layout)	98.53%	0.95	1.00	0.97

Vision transformer We also experiment with the vision transformer (ViT) architecture (Dosovitskiy et al. 2021), which transforms image processing by dividing input images into fixed-sized patches, departing from the conventional pixel-based evaluation of CNNs. ViT encapsulates each patch into a latent representation while retaining positional information, forwarding them through a transformer encoder. The input image, denoted $x \in \mathbf{R}^{H \times W \times C}$, undergoes transformation into a sequence of flattened 2D patches $x_p \in \mathbf{R}^{N \times (P2 \cdot C)}$, where $N = W \cdot H/P^2$ signifies the resulting number of patches of size $P \times P$, and $H \times W$ is the resolution of the image. With C representing the channels, typically 3 for RGB images, our model embraces a patch size of 16 × 16 pixels. This architecture facilitates the breakdown of images into manageable patches, subsequently processed through transformer layers adept at capturing both local and global dependencies. Our methodology aligns with the ViT paradigm, expanding the adaptability of transformers to encompass image classification tasks.

4.1 Multi-Label Classification

Multi-label classification involves the assignment of multiple labels to each instance simultaneously. It involves predicting multiple categories or classes for a given input, making it a more complex problem compared to traditional single-label classification. To address potential overlap in class characteristics, we also employ a multi-label classification approach. Each of the four classes is encoded as a five-dimensional vector, allowing for shared attributes across classes **[tab. 1]**. This method offers distinct advantages, particularly in

handling overlapping attributes among certain classes. Furthermore, the extraction of spatial document image features for layout classification is facilitated through the utilization of a VGG16 (Simonyan, Zisserman 2015) based backbone.

Complex layout classification We delve deeper into understanding the complexities of layout structures. Figure 2 showcases various examples from the dataset, highlighting the variability in spatial arrangements of text columns within the complex layout structure. For instance, Figure 2(a) exhibits a C type structure, while Figure 2(b) displays an L type arrangement. Moreover, Figures 2(c) and (d) portray complex spatial configurations bearing resemblance to an O and a U, respectively **[fig. 2]**.

We identified seven distinct subcategories within the complex layout arrangement **[fig. 3]**. These subcategories are characterized by different spatial configurations of text columns, including variations such as C, L, U, and O shapes, along with their corresponding reflected counterparts – C2, L2, and U2. Each of these subcategories captures unique layout features, contributing to the complexity of the overall structure, and poses different challenges for accurate classification. Through training an end-to-end CNN-based classifier, we aimed to comprehend these features and effectively capture the nuanced spatial relationships within the complex layout structures. Our experiments yielded a classification accuracy of 60%, indicating the model's ability to distinguish these spatial features significantly better than random guessing.

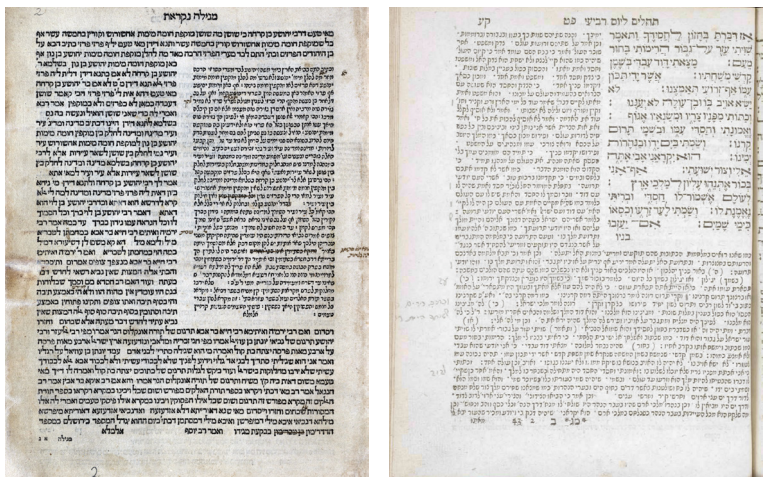


Figure 2 Examples of complex layouts

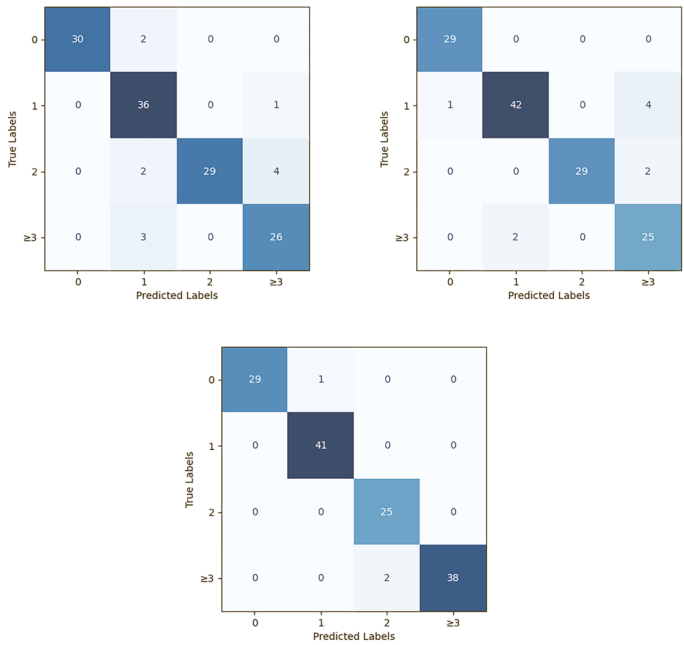
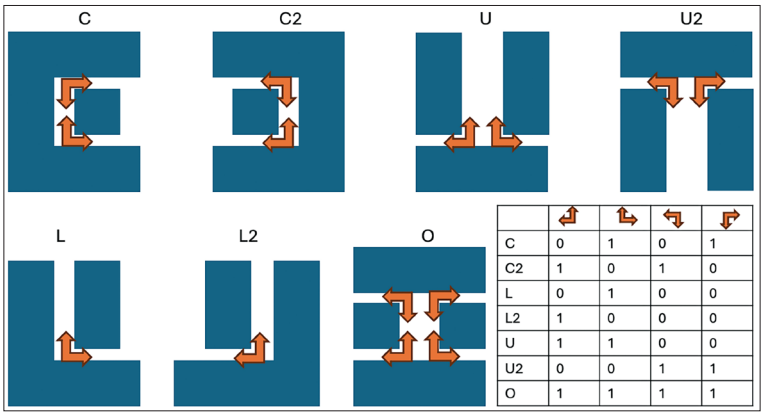


Figure 3 Confusion matrices for the different classifiers. (a) Confusion matrix for EfficientNetV2; (b) Confusion matrix for ViT; (c) Confusion matrix for multilabel encoding with VGG16

5 Results

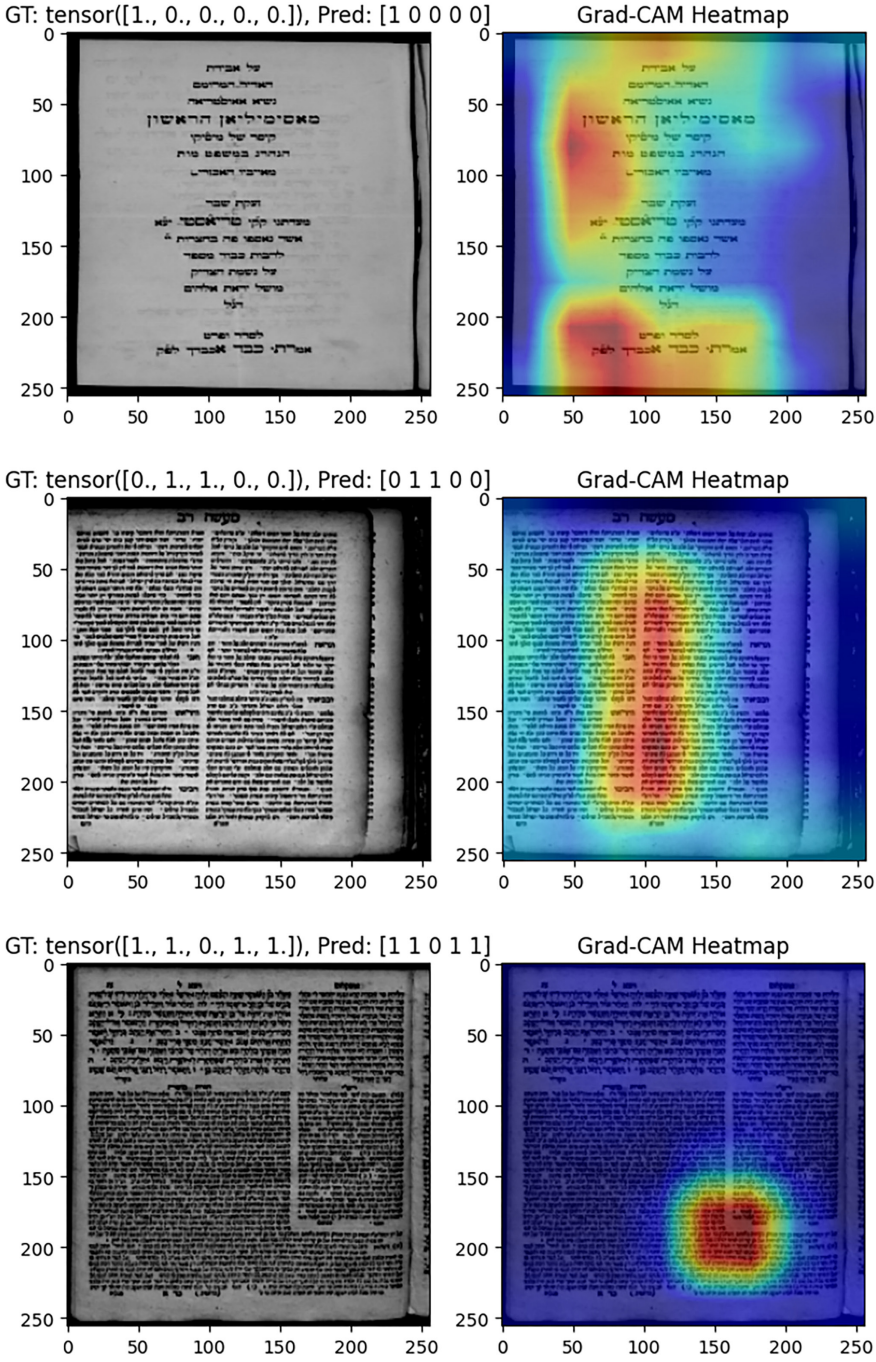
In this section, we present the outcomes obtained from various deep-learning classifiers, which serve as foundational benchmarks for future comparative analyses. The aim was to assess the effectiveness of the proposed features and methods introduced here for facilitating efficient document layout classification. We achieved competitive performance on the document classification task. Figures 4a-c showcase the confusion matrices corresponding to the trained models.



C	0	1	0	1
C2	1	0	1	0
L	0	1	0	0
L2	1	0	0	0
U	1	1	0	0
U2	0	0	1	1
O	1	1	1	1

Figures 4a-c Examples of complex layout structures with corresponding spatial arrangement features

The evaluation metrics, including accuracy, precision, recall, and F1, are utilized to assess the models' performance across different classes, as showcased in Table 2. Notably, employing EfficientNetV2 yielded an impressive overall accuracy of 90.98%, while the ViT model achieved an even higher accuracy of 93.28%. Furthermore, leveraging the multi-label encoding approach with VGG16 resulted in the highest accuracy of 97.79%. To elucidate the influential features guiding the model's final prediction, we employ the gradient-weighted class activation mapping (Grad-CAM) technique (Selvaraju et al. 2017). This approach leverages the gradients of a target class flowing into the underlying CNN architecture, specifically VGG16 in our study, to generate a coarse localization map, thereby accentuating pivotal regions crucial for predicting the target class. Figure 5 depicts the salient features relevant to the classification of layout structures **[fig. 5]**.


GT: tensor([1., 0., 0., 0., 0.]), Pred: [1 0 0 0 0]
Grad-CAM Heatmap
GT: tensor([0., 1., 1., 0., 0.]), Pred: [0 1 1 0 0]
Grad-CAM Heatmap
GT: tensor([1., 1., 0., 1., 1.]), Pred: [1 1 0 1 1]
Grad-CAM Heatmap

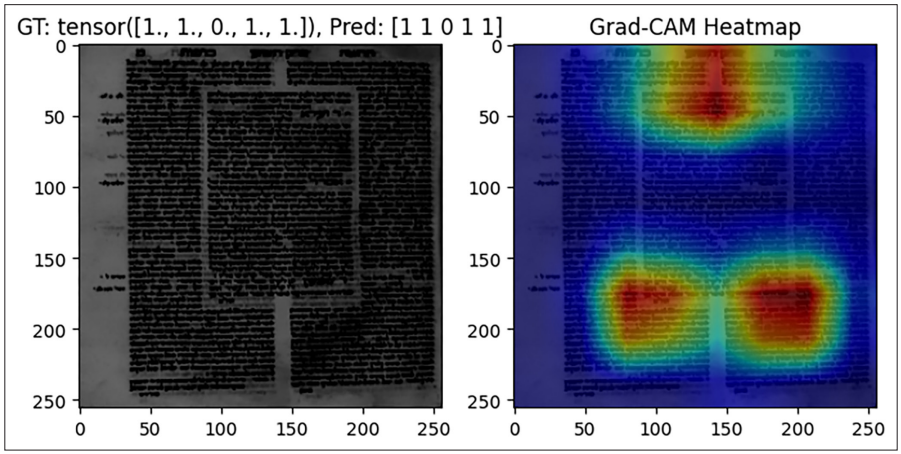


Figure 5 Visualization of important features for classification using Grad-CAM

6 Conclusions and Future Work

Conducting layout analysis on simple layouts, containing one or two columns of text, is relatively straightforward, but analysing complex layouts that feature text columns in structures diverging from the standard one or two columns, such as L, U, O, and C shapes, alongside other complexities, presents significant challenges. Therefore, layout classification is vital for distinguishing between simple and complex layouts. This distinction allows for the application of existing layout analysis algorithms on simple layout document images but specialized analysis methods for complex layout document images.

We have introduced a dataset designed for benchmarking layout classification methods, along with a single-label multi-classification algorithm and a multi-label multi-classification algorithm to address the layout classification challenge. Our findings indicate that multi-label encoding leads to a more separable feature space, thereby enhancing accuracy. The visualization of classifiers further supports this conclusion, revealing that the classifiers indeed focus on features employed to encode the multi-labels for each class.

Future work includes further improving results for complex layout classification in a variety of languages and scripts, considering pages with marginal and intertextual comments, considering books with changes of script size and/or language within paragraphs, and pages from incunabula and other early printed books with unusual nonstandard layouts. This will be combined with reading-direction recognition, language, and script detection to achieve complex page

analysis. These algorithms would serve as a solid base for efficient automatic processing of printed books. At the same time, the automatic classification of page layouts for printed books is an important preparatory step for the more challenging task of page layout analysis of handwritten manuscripts.

Bibliography

Alaasam, R.; Kurar, B.; El-Sana, J. (2019). "Layout Analysis on Challenging Historical Arabic Manuscripts Using Siamese Network". *2019 International Conference on Document Analysis and Recognition (ICDAR) = Conference Proceedings* (Sydney, NSW, Australia, 20-25 September 2019). New York: Institute of Electrical and Electronics Engineers (IEEE), 738-42.
https://doi.org/10.1109/ICDAR.2019.00123

Alarcón Arenas, S.W.; Yari, Y.; Meza-Lovon, G. (2018). "A Document Layout Analysis Method Based on Morphological Operators and Connected Components". *XLIV Latin American Computer Conference (CLEI) = Conference Proceedings* (São Paulo, Brazil, 1-5 October 2018). New York: Institute of Electrical and Electronics Engineers (IEEE), 622-31.
https://doi.org/10.1109/CLEI.2018.00080

Antonacopoulos, A.; Bridson, D.; Papadopoulos, C.; Pletschacher, S. (2009). "A Realistic Dataset for Performance Evaluation of Document Layout Analysis". *10th International Conference on Document Analysis and Recognition (ICDAR) = Conference Proceedings* (Barcelona, Spain, 26-29 July 2009). New York: Institute of Electrical and Electronics Engineers (IEEE), 296-300.
https://doi.org/10.1109/ICDAR.2009.271

Barakat, B.K.; El-Sana, J. (2018). "Binarization Free Layout Analysis for Arabic Historical Documents Using Fully Convolutional Networks". *IEEE 2nd International Workshop on Arabic and Derived Script Analysis and Recognition (ASAR) = Conference Proceedings* (London, UK, 12-14 May 2018). New York: Institute of Electrical and Electronics Engineers (IEEE), 151-5.
https://doi.org/10.1109/ASAR.2018.8480333

Borges Oliveira, D.A.; Palhares Viana, M. (2017). "Fast CNN-based Document Layout Analysis". *IEEE International Conference on Computer Vision Workshops (ICCVW) = Conference Proceedings* (Venice, Italy, 22-29 October 2017). New York: Institute of Electrical and Electronics Engineers (IEEE), 1173-80.
https://doi.org/10.1109/ICCVW.2017.142

Breuel, T. (2003). *High Performance Document Layout Analysis*. Technical report, May.
https://www.researchgate.net/publication/2564797_High_Performance_Document_Layout_Analysis

Chen, K. et al. (2014). "Page Segmentation for Historical Handwritten Document Images Using Color and Texture Features". *14th International Conference on Frontiers in Handwriting Recognition (ICFHR) = Conference Proceedings* (Hersonissos, Crete Island, Greece, 1-4 September 2014). New York: Institute of Electrical and Electronics Engineers (IEEE), 488-93.
https://doi.org/10.1109/ICFHR.2014.88

Clausner, C. et al. (2015). "The Enp Image and Ground Truth Dataset of Historical Newspapers". *13th International Conference on Document Analysis and Recognition*

(ICDAR) = Conference Proceedings (Tunis, Tunisia, 23-26 August 2015). New York: Institute of Electrical and Electronics Engineers (IEEE), 931-5.
https://doi.org/10.1109/ICDAR.2015.7333898

Corbelli, A. et al. (2016). "Historical Document Digitization Through Layout Analysis and Deep Content Classification". *23rd International Conference on Pattern Recognition (ICPR) = Conference Proceedings* (Cancún, Mexico, 4-8 December 2016). New York: Institute of Electrical and Electronics Engineers (IEEE), 4077-82.
https://doi.org/10.1109/ICPR.2016.7900272

Da, C. et al. (2023). "Vision Grid Transformer for Document Layout Analysis". *IEEE/CVF International Conference on Computer Vision (ICCV) = Conference Proceedings* (Paris, France, 1-6 October 2023). New York: Institute of Electrical and Electronics Engineers (IEEE), 19405-15.
https://doi.org/10.1109/ICCV51070.2023.01783

Dosovitskiy, A. et al. (2021). "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale". *International Conference on Learning Representations (ICLR) = Conference Proceedings* (Austria, 3-7 May 2021).
https://openreview.net/forum?id=YicbFdNTTy

Hu, J.; Kashi, R.S.; Wilfong, G. (1999). "Document Image Layout Comparison and Classification". *Proceedings of the Fifth International Conference on Document Analysis and Recognition (ICDAR)* (Bangalore, India, 20-22 September 1999). New York: Institute of Electrical and Electronics Engineers (IEEE), 285-8.
https://doi.org/10.1109/ICDAR.1999.791780

Kiessling, B. et al. (2019). "eScriptorium: An Open Source Platform for Historical Document Analysis". *International Conference on Document Analysis and Recognition Workshops (ICDARW) = Conference Proceedings* (Sydney, Australia, 22-25 September 2019). New York: Institute of Electrical and Electronics Engineers (IEEE), 19.
https://doi.org/10.1109/ICDARW.2019.10032

Kise, K.; Sato, A.; Iwata, M. (1998). "Segmentation of Page Images Using the Area Voronoi Diagram". *Computer Vision and Image Understanding*, 70(3), 370-82.
https://doi.org/https://doi.org/10.1006/cviu.1998.0684

Kosaraju, S.C. et al. (2019). "DoT-Net: Document Layout Classification Using Texture-based CNN". *International Conference on Document Analysis and Recognition (ICDAR) = Conference Proceedings* (Sydney, Australia, 20-25 September 2019). New York: Institute of Electrical and Electronics Engineers (IEEE), 1029-34.
https://doi.org/10.1109/ICDAR.2019.00168

Sandler, M. et al. (2018). "MobileNetV2: Inverted Residuals and Linear Bottlenecks". *IEEE/CVF Conference on Computer Vision and Pattern Recognition = Conference Proceedings* (Salt Lake City, UT, USA, 18-23 June 2018). New York: Institute of Electrical and Electronics Engineers (IEEE), 4510-20.
https://doi.org/10.1109/CVPR.2018.00474

Selvaraju, R.R. et al. (2017). "Grad-CAM: Visual Explanations from Deep Networks Via Gradient-based Localization". *IEEE International Conference on Computer Vision (ICCV) = Conference Proceedings* (Venice, Italy, 22-29 October 2017). New York: Institute of Electrical and Electronics Engineers (IEEE), 618-26.
https://doi.org/10.1109/ICCV.2017.74

Simonyan, K.; Zisserman, A. (2015). "Very Deep Convolutional Networks for Large-scale Image Recognition". Bengio, Y.; LeCun, Y. (eds), *3rd International Conference on Learning Representations (ICLR) = Conference Track Proceedings* (San Diego, 7-9 May 2015). San Diego, 1-14
http://arxiv.org/abs/1409.1556

Tan, M.; Le, Q. (2019). "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks". Chaudhuri, K.; Salakhutdinov, R. (eds), *Proceedings of Machine Learning Research (PMLR)*. Vol. 97, *Proceedings of the 36th International Conference on*

Machine Learning (Long Beach, California, USA, 9-15 June 2019). Long Beach (CA), 6105-14.
https://proceedings.mlr.press/v97/ tan19a.html

Tan, M.; Le, Q. (2021). "Efficientnetv2: Smaller Models and Faster Training". Meila, M.; Zhang, T. (eds), *Proceedings of Machine Learning Research (PMLR)*. Vol. 139, *Proceedings of the 38th International Conference on Machine Learning* (18-24 July 2021), 10096-106.
https://proceedings.mlr.press/v139/tan21a.html

Wei, H. et al. (2013). "Evaluation of SVM, MLP and GMM Classifiers for Layout Analysis of Historical Documents". *12th International Conference on Document Analysis and Recognition = Conference Proceedings* (Washington, DC, USA, 25-28 August 2013). New York: Institute of Electrical and Electronics Engineers (IEEE), 1220-4.
https://doi.org/10.1109/ICDAR.2013.247

Wong, K.Y.; Casey, R.G.; Wahl, F.M. (1982). "Document Analysis System". *IBM Journal of Research and Development*, 26(6), 647-56.
https://doi.org/10.1147/rd.266.0647

Zhong, X.; Tang, J.; Jimeno Yepes, A. (2019). *PubLayNet: Largest Dataset Ever for Document Layout Analysis*. September.