



# Text Line Detection and Recognition of Greek Polytonic Documents

Panagiotis Kaddas<sup>1,2(✉)</sup>, Basilis Gatos<sup>1</sup>, Konstantinos Palaiologos<sup>1,3</sup>,  
Katerina Christopoulou<sup>1,4</sup>, and Konstantinos Kritsis<sup>5</sup>

<sup>1</sup> Computational Intelligence Laboratory, Institute of Informatics and Telecommunications, National Center for Scientific Research Demokritos, GR-153 10, Agia Paraskevi, Athens, Greece

{pkaddas,bgat,k.palaiologos,achristopoulou}@iit.demokritos.gr

<sup>2</sup> Department of Informatics and Telecommunications, University of Athens, 157 84 Athens, Greece

<sup>3</sup> Hellenic Institute, Royal Holloway, University of London, Egham Hill, Egham, Surrey TW20 0EX, UK

<sup>4</sup> Harokopio University, School of Environment, Geography and Applied Economics, Department of Economics and Sustainable Development, 17676 Athens, Greece

<sup>5</sup> Department of English Studies, University of Cyprus, 1678 Nicosia, Cyprus  
kritsis.konstantinos@ucy.ac.cy

**Abstract.** In this work we highlight the significance of Text Line Detection in documents. By utilizing a well-known Deep Neural Network and by proposing some simple but efficient modifications applied during training of such models, we can achieve very accurate results in different datasets of high diversity. Moreover, such models can be robust even when trained with few data. Our focus is on Greek polytonic documents (typewritten and handwritten) and we provide a new dataset to the public (GTLD-small) for text line detection. We evaluate our method through scenarios applied to the detection and recognition tasks, while demonstrating promising results when compared to popular commercial or open-source systems.

**Keywords:** Deep Neural Networks · Object Detection · Text Line Detection · Text Line Recognition · YOLOv5 · Circular Smooth Label

## 1 Introduction

Over the last few decades, there has been a growing interest in studying and examining images of historical and modern documents, largely driven and motivated by the need to investigate material stored in libraries and archives [4]. There are millions of pages that have been digitized and are accessible as images, but in order to facilitate efficient work with such documents by humanists, historians and the general public, there is ongoing research and academic discourse aimed at understanding and converting the content of these documents [17] in a digital form. This is equal to recognize and comprehend the textual and non-textual content through an automated procedure.

Recently, the research of document image analysis has investigated utilizing deep learning for the segmentation and recognition of typewritten and handwritten text. At this point, it is worth noting that images of historical documents are distinct from images of natural scenes, where deep learning techniques are already applied widely [8]. As deep learning techniques advance, the automated processing and understanding of documents has become increasingly efficient, concerning text recognition, while highlighting two key steps as crucial towards success [4]: Text line detection and text line recognition. Text line detection is a technique used in document processing and computer vision to detect and extract text lines from a scanned document [21]. It is often being applied as a pre-processing step for Optical Character Recognition (OCR) and Layout Analysis, while having great impact on the accuracy of each task [6].

In this work, by utilizing a variation of the well-known YOLOv5 ( [13]) Deep Neural Network model (YOLOv5-OBB<sup>1</sup>) and by proposing some simple but efficient modifications applied during training, we can achieve very accurate text line detection results on different datasets of high diversity. Moreover, we demonstrate the efficiency of such workflow on the text recognition task.

The contributions of this paper are as follows: a) We introduce a new dataset, named GTLD dataset, with annotated text line quadrilateral polygons and a smaller subset of 1.642 documents (including annotations on the Tobacco-3482 dataset) is publicly available<sup>2</sup>. b) We show in the experiments that the introduced oriented quadrilateral polygons can improve accuracy on the text line detection task, especially when text line boxes overlap. c) We provide promising results on text line recognition using an end-to-end workflow.

The rest of the paper is organized as follows. Section 2 presents related works, Sect. 3 introduces the proposed method, Sect. 4 demonstrates experimental scenarios and results and Sect. 5 presents the conclusion of this work.

## 2 Related Work

**Text Line Detection.** Text line detection methods have been proposed for many years [3, 16, 18, 22], where most techniques focus on historical documents because of their diverse nature. In addition, segmentation of unconstrained handwritten manuscripts is usually the most challenging task to comprehend [7]. In such cases, text line detection is sometimes equivalent to the detection of baselines.

In [10], a two-stage workflow is proposed to detect baselines, where the first stage is a pixel labeling (or goal-oriented binarization) deep hierarchical neural network (ARU-Net), which detects foreground baseline pixels and separators. Finally, second stage clusters extracted superpixels in order to build baselines. In [2], dhSegment is proposed, where a generic neural network (U-Net [20] based on ResNet50 [12]) enables multi-task detection for page segmentation, layout analysis, line detection and ornament extraction. A similar technique has been

<sup>1</sup> [https://github.com/hukaixuan19970627/yolov5\\_obb](https://github.com/hukaixuan19970627/yolov5_obb).

<sup>2</sup> <https://doi.org/10.5281/zenodo.8020403>.

proposed in [5], where Doc-UFCN utilizes a light U-shaped Fully Convolutional Network (FCN) without any dense layers for page segmentation and text line detection. In [1], text line detection is achieved when document image binarization is applied as a pre-processing step and connected component analysis enables the localization of text blocks and lines. Finally, in [8] the authors propose a Mask-RCNN [11] based architecture for detecting text lines in historical Arabic handwritten documents, where they group labeled results after detecting text lines on image patches.

**Text Line Recognition.** Recognizing the text in documents has been frequently considered as a subsequent step to the text line detection task. This paper focuses on recognition of Greek polytonic documents, where variations in writing styles and accents is a very challenging task [9]. In [23], the authors propose a workflow for the recognition of Greek printed polytonic scripts, using a character segmentation step combined with intensity-based feature extractor and  $k$ -nearest neighbors (KNN) classification. Following the neural network logic and by using only Long Short-Term Memory (LSTM) Networks, in [24], the authors demonstrate some interesting results for Greek polytonic scripts.

Over the years, many methods have been proposed for the Handwriting Text Recognition (HTR) problem, where most approaches focus on Deep Learning techniques [25]. In [17], an attention-based sequence-to-sequence (Seq2Seq) model was proposed, where a Convolutional Neural Network (CNN) with Bidirectional LSTM (BLSTM) networks encode the image into a feature representation. Then, a decoder that utilizes a recurrent layer interprets this information and an attention mechanism enables focusing on the most relevant encoded features.

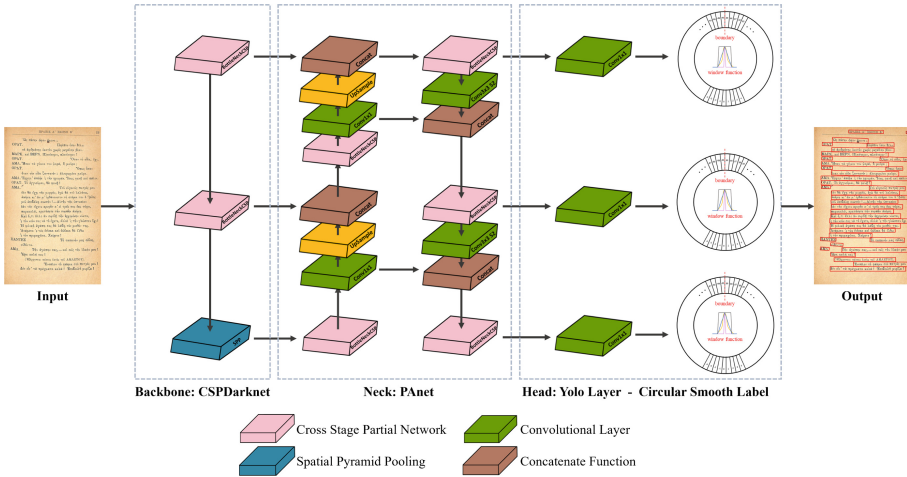
## 3 Proposed Method

### 3.1 Object Detection Model

The overall proposed architecture is provided in Fig. 1. Inspired by the success of the well-known YOLO [19] object detectors, we use a variation of the YOLOv5 model [13], utilized for the detection of oriented quadrilateral polygons [26], (named YOLOv5-OBb for convenience).

As shown in Fig. 1, the CSP-Darknet53 is the Darknet53 backbone architecture, combined the Cross Stage Partial (CSP) network strategy. CSP is applied in order to overcome the problem of redundant gradients by truncating the gradient flow, while preserving the advantage of residual and dense blocks used in the backbone in order to flow information to the deepest layers and to overcome the vanishing gradient problem. Using CSP, inference speed is increased (less FLOPS) while the number of parameters is being reduced. YOLOv5 has a major change in the Neck of the architecture, when compared to its predecessors. A variant of Spatial Pyramid Pooling layer feeds a Path Aggregation Network (PANet) in which the CSP strategy has been incorporated via the

BottleNeckCSP layer. Finally, the Head of the network is composed from 3 convolution layers for predicting the location of bounding boxes, as long as the object classes with their scores and the angle estimation.



**Fig. 1. Architecture of the proposed YOLOv5-OBB model:** The CSP-Darknet53 is the backbone feature extractor and a variant of Spatial Pyramid Pooling (SPP) layer feeds a Path Aggregation Network (PANet). The Head of the network is composed from 3 convolution layers combined with the Circular Smooth Label technique in order to classify the quadrilateral’s angle and to accurately detect the text lines in the document. In BottleNeckCSP convolutional layers, the Cross Stage Partial (CSP) technique has been incorporated in order to overcome the problem of redundant gradients.

### 3.2 Circular Smooth Label for Angle Estimation

In addition to the YOLOv5 model in order to localize text lines with orientation, this work follows the Circular Smooth Label (CSL) technique proposed in [26] in order to classify the rotation angle of the object’s bounding polygon. The use of regression methods can cause issues as ideal predictions fall outside the defined range. The result of a prediction can be controlled more effectively if the object angle is approached as a classification problem, where the angle is treated as a category label, and the number of categories is determined by the angle range. However, changing the method from regression to one-hot encoded classification can result in a slight decrease of accuracy.

With the CSL method, the angles are encoded in a circular/repeating pattern (Fig. 1), while windowing functions (pulse, rectangular, triangle and Gaussian) can be used to exterminate boundary problems in order to obtain better prediction of the angle’s class.

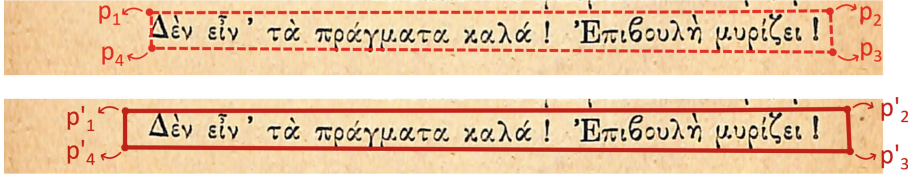
### 3.3 Extended Polygonal Labeling

In many cases, ground-truth text lines are strictly annotated and the polygonal edges almost overlap with the characters. Even when this is not the case, we observe that LSTM layers in a text line recognition system handle edge characters better, as demonstrated in the experiments Section. So, by loosening the enclosing polygon, both line detection and recognition systems improve. For this reason and during training of the YOLOv5-OBB model, we shift every point  $p_i = (x_i, y_i), i \in 1, 2, 3, 4$  of the  $n$ -th text line quadrilateral of width  $w_n$  over the horizontal axis, resulting to the new point  $p_i' = (x_i', y_i')$ , following Eqs. 1, 2:

$$x_i' = x_i + a \times k \times w_n \quad (1)$$

$$y_i' = \frac{(y_j - y_i)}{(x_j - x_i)} \times (x_i' - x_i) + y_i \quad (2)$$

where  $j = (i + 1) \bmod 4$ ,  $a = -1$  for the two leftmost points and  $a = +1$  for the two rightmost points. We empirically use  $k = 0.03$  as a shift factor relative to the width of the quadrilateral, since we observe that resulting polygons do not overlap, while starting and ending characters are recognized properly from the text line recognition system. A visualization is given in Fig. 2.



**Fig. 2. Extended Polygonal Labeling:** Ground-truth text lines are sometimes annotated strictly and the polygonal edges almost overlap with the characters (upper image). By loosening the enclosing polygon over the horizontal axis, both line detection and recognition systems improve (lower image).

## 4 Experiments

### 4.1 Datasets

Since the detection of text lines can be considered in general as a language-independent segmentation task, we combine two datasets with a totaling number of 17.133 images. We perform our experiments on the Tobacco-3482 dataset [14] along with a new Greek polytonic dataset named GTLD. The latter is a combination of two heterogeneous collections, with the first subset containing also handwritten documents. The overview of the datasets is given in Table 1. Smaller

portions of each dataset (denoted with a “*small*” suffix in Table 1) with a total number of 1.642 images are available along with the corresponding annotations at a text line level (See footnote 2).

**Tobacco-3482 Dataset.** The Tobacco-3482 [14] dataset consists of document images belonging to 10 classes such as letter, form, email, resume, memo, etc. The dataset has 3482 images and we provide annotations at a text line level.

**Greek Text Line Detection Dataset (GTLD).** We introduce a new dataset, named GTLD (See footnote 2), which contains a total of 13.651 images acquired from two collections: The ShakeIT collection contains images of Greek polytonic documents that originate from almost 140 books of Greek translations of Shakespeare’s plays published between 1842 and 1950. For this work, we use 28 books with a total of 3955 randomly selected pages. The PIOP collection consists of almost 100.000 digital documents from the Historical Archives of the Piraeus Bank Group Cultural Foundation (PIOP)<sup>3</sup>. This archive includes a significant number of collections starting from the early 20th century. In this paper, we use a total of 9696 randomly selected pages. Both collections contain mostly Greek polytonic characters, with rare cases of English/German. Exemplar images from both collections are shown in Fig. 3.

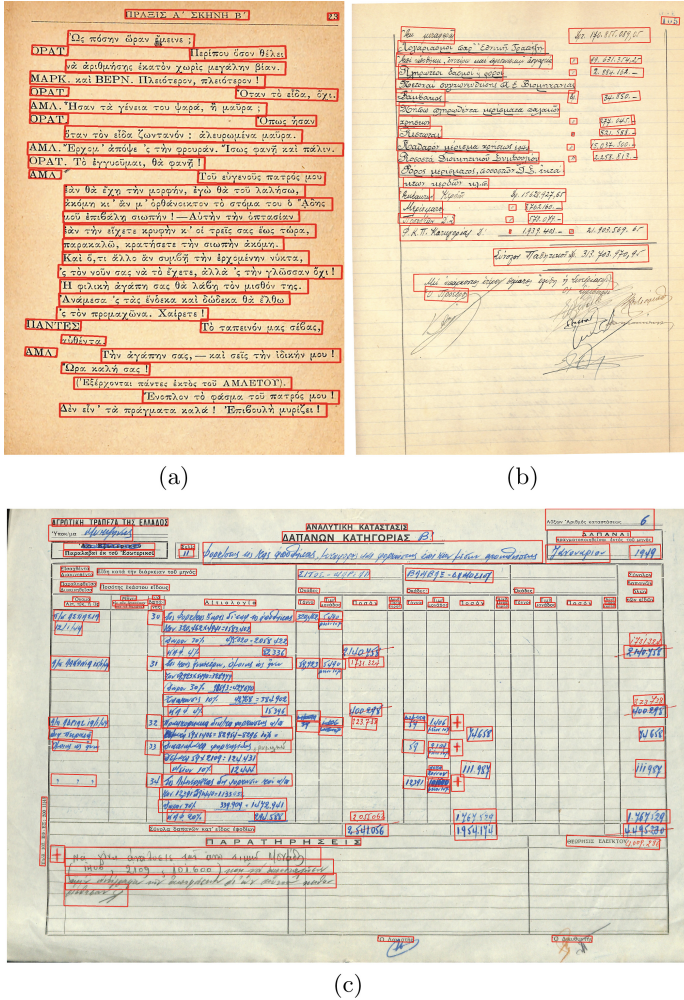
**Table 1.** Overview of the datasets included in this work and the number of images used for training, validation and testing.

Dataset	Collection	#Total	#train	#val	#test
GTLD	PIOP	9.696	6.782	969	1945
	ShakeIT	3.955	2.754	394	807
GTLD-small	PIOP-small	950	672	90	188
	ShakeIT-small	357	264	27	66
Tobacco-3482	Tobacco-3482	3.482	2.433	345	704
	Tobacco-3482-small	335	240	30	65

## 4.2 Experimental Results

**Text Line Detection:** For the text line detection task, we train all models shown in the following for 80–150 epochs (more epochs when using smaller subsets) using SDG algorithm with a starting learning rate of 0.001 and batch size equal to 4. We also apply common augmentation techniques during training (random affine transformations, color jitter, flipping, etc.) over the input images. For the Circular Smooth Label, we use Gaussian kernel with a radius of 4. As a

<sup>3</sup> <https://www.piop.gr/en/istoriko-arxeio.aspx>.



**Fig. 3.** Exemplar documents of the GTLD dataset: a): typewritten text (ShakeIT), b) handwritten text (PIOP), c) “Hybrid” text (PIOP).

YOLOv5 backbone, we use the YOLOv5-large<sup>4</sup> (YOLOv5l) model, pre-trained on MSCOCO dataset [15]. We keep the instance of the model that performed best on the validation set. As evaluation metrics, we use recall (R), precision (P), f1-score (F1) and mean Average Precision (mAP), evaluated at mAP@.5 and mAP@.5:.95, following the best evaluation practices in the domain of object detection. Table 2 shows the different architectures used in the following experiments, where “Rect” means rectangular polygons, “Quad” means quadrilateral

<sup>4</sup> <https://github.com/ultralytics/yolov5>.

polygons, “Extended” means extended quadrilateral or rectangular polygons and “CSL” means Circular Smooth Label

**Table 2.** Different architectures for the text line detection task.

Name	Rect	Quad	Extended	CSL
YOLOv5-rect	✓			
YOLOv5-rect-ext	✓		✓	
Yolov5-OBb		✓		✓
Yolov5-OBb-ext		✓	✓	✓

Table 3 presents the evaluation of all trained models for each collection. It is noted that each model has been trained either on all the “*small*” collections or the “*large*” collections and we do not train a separate model for each collection. At this point, it is important to highlight the following: a) For all small collections, it is expected to have lower metrics, because we use fewer training samples in contrast with the larger ones. b) For all setups, YOLOv5-rect is inferior to the variations proposed by this work. c) YOLOv5-rect-ext is superior to the YOLOv5-rect version, which proves the contribution of considering extended bounding polygons (for this case, bounding boxes). d) Application of the CSL technique (YOLOv5-OBb) significantly improves detection results (+7.5% on **mAP@.5:.95**) when compared to YOLOv5-rect. e) Extended Polygonal Labeling is a further improvement to the CSL technique for all the larger collections (+0.1%, +0.3%, +0.4% on mAP@.5:.95 respectively). For the smaller ones it is useful when there is need for more refined results (higher mAP@.5:.95 on PIOP-small and ShakeIT-small with Greek polytonic scripts), but if detail is not important (Tobacco-3482), extended quadrilaterals may be omitted.

In order to investigate the competitiveness of our proposed method, we evaluate two popular systems on our datasets. The first one, Google Cloud Vision API<sup>5</sup>, is commercial, while the second one is the well-known open-source OCR engine Tesseract<sup>6</sup> (version 5.0.0). In both cases we evaluate the systems using each one’s default setup. The evaluation results are given in Table 4. It is observed that the Tesseract engine fails to process all collections. One reason is that Tesseract provides only bounding rectangles instead of quadrilaterals and predictions suffer from the problems mentioned above. Google Cloud Vision API seems to perform better and gives more refined results on ShakeIT-small, were the layout of the documents is simpler. On the other hand, Our method is superior to all over the more diverse collections of PIOP-small and Tobacco-3482-small.

<sup>5</sup> <https://cloud.google.com/vision>.

<sup>6</sup> <https://github.com/tesseract-ocr/tesseract>.



**Table 3.** Evaluation results (%) for the GTLD (PIOP, ShakeIT) and Tobacco-3482 datasets

Collection	Model	P	R	mAP@.5	mAP@.5:.95
PIOP-small	YOLOv5-rect	<b>85.1</b>	59.4	68.9	34.5
	YOLOv5-rect-ext	82.7	58.0	69.8	37.2
	YOLOv5-OBB	84.7	<b>65.2</b>	<b>74.0</b>	39.9
	YOLOv5-OBB-ext	84.5	64.8	73.5	<b>41.4</b>
ShakeIT-small	YOLOv5-rect	95.5	95.4	95.5	53.8
	YOLOv5-rect-ext	95.7	95.3	95.6	52.5
	YOLOv5-OBB	<b>99.2</b>	<b>99.5</b>	<b>99.4</b>	65.4
	YOLOv5-OBB-ext	99.0	98.8	99.2	<b>68.1</b>
Tobacco-3482-small	YOLOv5-rect	87.6	79.7	83.9	42.9
	YOLOv5-rect-ext	85.8	81.0	84.5	42.3
	YOLOv5-OBB	88.5	<b>86.1</b>	<b>88.3</b>	<b>50.3</b>
	YOLOv5-OBB-ext	<b>89.3</b>	85.6	86.4	47.1
PIOP	YOLOv5-rect	87.0	74.3	81.1	47.6
	YOLOv5-rect-ext	<b>89.2</b>	78.3	85.7	50.4
	YOLOv5-OBB	88.4	79.6	86.5	54.4
	YOLOv5-OBB-ext	88.7	<b>80.9</b>	<b>87.1</b>	<b>54.5</b>
ShakeIT	YOLOv5-rect	95.2	94.9	96.6	65.8
	YOLOv5-rect-ext	99.3	99.2	99.4	68.0
	YOLOv5-OBB	99.3	<b>99.2</b>	<b>99.4</b>	71.9
	YOLOv5-OBB-ext	<b>99.4</b>	99.1	<b>99.4</b>	<b>72.2</b>
Tobacco-3482	YOLOv5-rect	85.5	82.7	85.4	50.9
	YOLOv5-rect-ext	90.5	86.0	89.5	57.6
	YOLOv5-OBB	<b>91.1</b>	87.3	90.6	58.8
	YOLOv5-OBB-ext	91.0	<b>88.8</b>	<b>91.3</b>	<b>59.2</b>

**Text Line Recognition:** Finally, order to signify the the robustness of the proposed method, we evaluate our systems indirectly over the subsequent task of text line recognition on Greek polytonic documents. For this reason, we perform the following steps:

**Table 4.** Comparison (mAP %) of the proposed method against other popular commercial/open-source systems.

Collection	Model	mAP@.5	mAP@.5:.95
PIOP-small	Google Cloud Vision API	52.0	31.9
	Tesseract 5.0.0	23.3	9.9
	<b>YOLOv5-OBB-ext</b>	<b>73.5</b>	<b>41.4</b>
ShakeIT-small	<b>Google Cloud Vision API</b>	95.2	<b>75.4</b>
	Tesseract 5.0.0	23.3	9.93
	<b>YOLOv5-OBB-ext</b>	<b>99.2</b>	68.1
Tobacco-3482-small	Google Cloud Vision API	67.6	31.3
	Tesseract 5.0.0	28.1	10.1
	<b>YOLOv5-OBB-ext</b>	<b>86.4</b>	<b>47.1</b>

- We train a text line recognition module, by utilizing the Calamari-OCR<sup>7</sup> engine. We use about 100.000 annotated text lines (OCR level) from the PIOP and ShakeIT collections. We choose the “*htr+*” architecture ([17]).
- We apply text line detection using four models (YOLOv5-rect, YOLOv5-rect-ext, YOLOv5-OBB and YOLOv5-OBB-ext) trained on the smaller collections.
- In order to construct the full page OCR from the predictions, we sort the detected text lines using Density-based spatial clustering (DBSCAN) and the coordinates of the extracted clusters.
- We consider the full page OCR of the 66 ShakeIT-small test images for evaluation. We choose this collection because it has simpler layout in order to minimize reading order errors from the previous step.
- As evaluation metrics, we consider the well-known Character Error Rate (CER) and Word Error Rate (WER).

From the results of Table 5 it is demonstrated again that the CSL method with extended quadrilaterals yields the best results and assists significantly to the text line recognition task. CER and WER results also validate the robustness of our text line recognition module. Another observation is that the text line recognition workflow that has been proposed above can be considered as an end-to-end workflow without any prior processing steps except for the line detection module combined with an grouping algorithm like DBSCAN for line ordering.

<sup>7</sup> <https://github.com/Calamari-OCR/calamari>.

**Table 5.** Evaluation of page OCR results using text line detection models on the ShakeIT-small test images.

Model	WER(%)	CER(%)
YOLOv5-rect	19.1	6.6
YOLOv5-rect-ext	17.2	5.3
YOLOv5-OBB	12.4	3.5
<b>YOLOv5-OBB-ext</b>	<b>9.8</b>	<b>3.2</b>

## 5 Conclusion

In this work, text line detection and recognition have been considered, with a focus on Greek polytonic documents. It was proposed that by utilizing YOLOv5 with some easy and simple techniques like the CSL method and extended polygonal labeling, there is a significant improvement on the text line detection task and accuracy results outperform popular commercial and open-source systems. Moreover, by applying this logic to the subsequent task of performing OCR even at a page level, similar improvements are noted and text line detection (with a line ordering technique) can be the only step before OCR in an end-to-end workflow. Finally, a new dataset was introduced (GTLTD), with a smaller subset being publicly available for future works.

**Acknowledgment.** This research has been partially co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH-CREATE-INNOVATE, project Culdile (Cultural Dimensions of Deep Learning, project code: T1EDK-03785), the Operational Program Attica 2014-2020, under the call RESEARCH AND INNOVATION PARTNERSHIPS IN THE REGION OF ATTICA, project reBook (Digital platform for re-publishing Historical Greek Books, project code: ATTP4-0331172) and the project “Corpus-assisted drama translation research: Shakespeare In Translation - ShakeIT”, under the call of internal research project funding of University of Cyprus (UCY <https://www.ucy.ac.cy/directory/en/>).

## References

1. Ahn, B., Ryu, J., Koo, H.I., Cho, N.I.: Textline detection in degraded historical document images. *EURASIP J. Image Video Process.* **2017**(1), 82 (2017)
2. Ares Oliveira, S., Seguin, B., Kaplan, F.: dhSegment: a generic deep-learning approach for document segmentation. In: 16th International Conference on Frontiers in Handwriting Recognition (ICFHR), 2018, pp. 7–12. IEEE (2018)
3. Basu, S., Chaudhuri, C., Kundu, M., Nasipuri, M., Basu, D.: Text line extraction from multi-skewed handwritten documents. *Pattern Recogn.* **40**(6), 1825–1839 (2007)
4. Boillet, M., Kermorvant, C., Paquet, T.: Robust text line detection in historical documents: learning and evaluation methods. *Int. J. Doc. Anal. Recog. (IJ DAR)* **25**(2), 95–114 (2022)

5. Boillet, M., Kermorvant, C., Paquet, T.: Multiple document datasets pre-training improves text line detection with deep neural networks. In: 2020 25th International Conference on Pattern Recognition (ICPR), pp. 2134–2141 (2021)
6. Diem, M., Kleber, F., Sablatnig, R.: Text line detection for heterogeneous documents. In: 2013 12th International Conference on Document Analysis and Recognition (ICDAR), pp. 743–747 (2013)
7. Diem, M., Kleber, F., Sablatnig, R., Gatos, B.: cBAD: ICDAR 2019 competition on baseline detection. In: 2019 International Conference on Document Analysis and Recognition (ICDAR), pp. 1494–1498 (2019)
8. Droby, A., Kurar Barakat, B., Alaasam, R., Madi, B., Rabaev, I., El-Sana, J.: Text line extraction in historical documents using mask r-CNN. *Signals* **3**(3), 535–549 (2022). <https://doi.org/10.3390/signals3030032>
9. Gatos, B., et al.: GRPOLY-DB: an old Greek polytonic document image database. In: 2015 13th International Conference on Document Analysis and Recognition (ICDAR), pp. 646–650 (2015)
10. Grüning, T., Leifert, G., Strauß, T., Michael, J., Labahn, R.: A two-stage method for text line detection in historical documents. *Int. J. Docu. Anal. Recogn. (IJ DAR)* **22**(3), 285–302 (2019)
11. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2980–2988 (2017)
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016)
13. Jocher, G., et al.: ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation (2022). <https://doi.org/10.5281/zenodo.7347926>
14. Kumar, J., Ye, P., Doermann, D.: Structural similarity for document image classification and retrieval. *Pattern Recogn. Lett.* **43**, 119–126 (2014). iCPR2012 Awarded Papers
15. Lin, T.-Y., et al.: Microsoft COCO: common objects in context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8693, pp. 740–755. Springer, Cham (2014). [https://doi.org/10.1007/978-3-319-10602-1\\_48](https://doi.org/10.1007/978-3-319-10602-1_48)
16. Louloudis, G., Gatos, B., Pratikakis, I., Halatsis, C.: Text line detection in handwritten documents. *Pattern Recogn.* **41**(12), 3758–3772 (2008)
17. Michael, J., Labahn, R., Grüning, T., Zöllner, J.: Evaluating sequence-to-sequence models for handwritten text recognition. In: 2019 International Conference on Document Analysis and Recognition (ICDAR), pp. 1286–1293 (2019)
18. Nicolas, S., Paquet, T., Heutte, L.: Text line segmentation in handwritten document using a production system. In: 9th International Workshop on Frontiers in Handwriting Recognition, pp. 245–250 (2004)
19. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779–788 (2016)
20. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
21. Sahare, P., Dhok, S.B.: Review of text extraction algorithms for scene-text and document images. *IETE Tech. Rev.* **34**(2), 144–164 (2017). <https://doi.org/10.1080/02564602.2016.1160805>

22. Shi, Z., Govindaraju, V.: Line separation for complex document images using fuzzy runlength. In: Proceedings of the First International Workshop on Document Image Analysis for Libraries (DIAL 2004), p. 306. DIAL 2004, IEEE Computer Society, USA (2004)
23. Sichani, A.M., Kaddas, P., Mikros, G.K., Gatos, B.: OCR for Greek polytonic (multi accent) historical printed documents: development, optimization and quality control. In: Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage, pp. 9–13. DATECH2019, Association for Computing Machinery, New York, NY, USA (2019)
24. Simistira, F., Ul-Hassan, A., Papavassiliou, V., Gatos, B., Katsouros, V., Liwicki, M.: Recognition of historical Greek polytonic scripts using LSTM networks. In: 2015 13th International Conference on Document Analysis and Recognition (ICDAR), pp. 766–770 (2015)
25. Teslya, N., Mohammed, S.: Deep learning for handwriting text recognition: existing approaches and challenges. In: 2022 31st Conference of Open Innovations Association (FRUCT), pp. 339–346 (2022)
26. Yang, X., Yan, J.: Arbitrary-oriented object detection with circular smooth label. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12353, pp. 677–694. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-58598-3\\_40](https://doi.org/10.1007/978-3-030-58598-3_40)