JOIV : Int. J. Inform. Visualization, 7(4) - December 2023 2513-2522

INTERNATIONAL JOURNAL
ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

A Review of Neural Network Approach on Engineering Drawing Recognition and Future Directions

Muhammad Syukri Mohd Yazed [a], Ezak Fadzrin Ahmad Shaubari [a,*], Moi Hoon Yap [b]

[a] *Department of Multimedia, Universiti Tun Hussein Onn Malaysia (UTHM), Batu Pahat, Johor, 86400, Malaysia*
[b] *Manchester Metropolitan University, All Saints Building, Manchester, M15 6BH, United Kingdom*
*Corresponding author: *ezak@uthm.edu.my*

***Abstract*— Engineering Drawing (ED) digitization is a crucial aspect of modern industrial processes, enabling efficient data management and facilitating automation. However, the accurate detection and recognition of ED elements pose significant challenges. This paper presents a comprehensive review of existing research on ED element detection and recognition, focusing on the role of neural networks in improving the analysis process. The study evaluates the performance of the YOLOv7 model in detecting ED elements through rigorous experimentation. The results indicate promising precision and recall rates of up to 87.6% and 74.4%, respectively, with a mean average precision (mAP) of 61.1% at IoU threshold 0.5. Despite these advancements, achieving 100% accuracy remains elusive due to factors such as symbol and text overlapping, limited dataset sizes, and variations in ED formats. Overcoming these challenges is vital to ensuring the reliability and practical applicability of ED digitization solutions. By comparing the YOLOv7 results with previous research, the study underscores the efficacy of neural network-based approaches in handling ED element detection tasks. However, further investigation is necessary to address the challenges above effectively. Future research directions include exploring ensemble methods to improve detection accuracy, fine-tuning model parameters to enhance performance, and incorporating domain adaptation techniques to adapt models to specific ED formats and domains. To enhance the real-world viability of ED digitization solutions, this work highlights the importance of conducting testing on diverse datasets representing different industries and applications. Additionally, fostering collaborations between academia and industry will enable the development of tailored solutions that meet specific industrial needs. Overall, this research contributes to understanding the challenges in ED digitization and paves the way for future advancements in this critical field.**

***Keywords*— Engineering drawings; ED analysis; neural network; object detection and recognition; industrial practice.**

Manuscript received 26 Mar. 2023; revised 27 Jul. 2023; accepted 25 Aug. 2023. Date of publication 31 Dec. 2023.

I. INTRODUCTION

Engineering drawings (EDs), such as piping and instrumentation diagrams, electrical and electronic drawings, and mechanical drawings, are commonly archived in various formats, including computer-aided design (CAD) drawings, PDF, and image formats [1], [2], and [3]. Digitizing EDs is necessary for applying technology in the shipbuilding and plant engineering industries, which involves applying digital image processing and computer vision techniques for pre-processing, element detection, classification, and sometimes inferring relations between elements [2], [4]–[6]. Recent ED digitization trends have involved artificial intelligence, with many deep learning-bound neural network models proposed for object and element recognition [7]–[10]. Various methods have been proposed, including back propagation neural networks [11], semi-automated heuristic methods [9], and the use of regional proposal networks (RPN) and convolutional neural networks (CNN) [12], with a focus on finding and recognizing ED elements.

The digitization process for engineering drawings, particularly complex EDs such as process flow diagrams, piping, and instrumentation diagrams, requires accurate object and element detection and a certain degree of contextualization [1]. That is, the meaning and relevance of the digitized information have to be interpreted according to a set of rules for a specific application. Early efforts to digitize these complex EDs relied on relatively rudimentary software, which has since become obsolete due to its incompatibility with modern hardware and software requirements [1]. However, more recent techniques, such as convolutional neural networks, have been developed to cater to the increasing demands for ED digitization across various industries [2].

Convolutional neural networks (CNN) have shown significant promise for digitizing EDs, with a reported accuracy rate exceeding 98% for engineering drawing classification [2]. Using CNN for ED analysis also has the added speed advantage compared to traditional digitization algorithms. To maximize their efficiency, these CNN models often employ data enhancement techniques such as rotation transformation, random cutting, and applying salt and pepper noise to expand the dataset, thereby improving the model's training [2]. Despite these advancements, considerable challenges remain, such as achieving a high degree of automation in the digitization process, particularly for more complex engineering diagrams [1].

According to previous studies, engineering drawings consist of various elements such as text, symbols, characters, units of measurement, notation, visual projection style, and page layout [1]. This study focuses explicitly on extracting three elements, which are text, symbols of components, and characters [9], [13]–[17]. With the increasing volume of data in engineering and manufacturing businesses, it is necessary to digitize EDs to process, analyze, and utilize them efficiently [18]. Manual analysis of EDs is time-consuming, and researchers have proposed using neural networks for ED digitization and image processing techniques for image enhancement [1].

This study highlights the significance of neural networks in improving the ED analysis process in academia and industry. By using neural networks, specifically CNN, the study demonstrates that it is possible to develop advanced computer vision models capable of accurately detecting and recognizing various elements within EDs, such as symbols and text. These neural network-based models offer a robust and efficient approach to digitizing EDs, reducing the need for manual analysis and increasing productivity. The potential benefits of adopting neural network techniques in ED analysis include faster and more accurate results, improved object specification and assembly processes, and benefits in academic research and practical applications in industries such as shipbuilding and plant engineering.

The paper is organized into sections, starting with a review of related works and industrial practices of ED analysis in section II. It is followed by a proposed neural network approach in section III, which analyses the previous and recent methods. Section IV provides the results and discussion. Finally, section V provides the conclusion.

II. Material and Methods

Various digital image processing techniques are involved in digitizing EDs, which include pre-processing, symbol detection, classification, and contextualization. Binarization or image thresholding is a commonly used pre-processing method to remove noise and improve object localization while thinning or skeletonization is used to discard object volume [19]. Review papers have been written on the digitization of these drawings, with some focusing on the field of EDs such as musical notes [20], CAD file creation for 3D reconstruction from paper-based mechanical drawings [21], [22], and OCR [23], [24]. Symbol detection [25] and classification [26] have also been reviewed in different parts of the digitization process. For instance, [27] proposed methods for interpreting mechanical drawings.

Artificial neural networks, also known as neural networks, are inspired by how neurons in the brain process information and learn [28]. These networks are created for specific tasks, such as pattern recognition or data classification, and can learn from either examples or experience. One of the earliest and most established CNN architectures is the CNN architecture of Yann LeCun [29], which has greatly contributed to the development of deep learning. This architecture can learn weights and biases and uses an input image to learn output probabilities for each class. CNN consists of various layers, including convolution, ReLU, pooling, and fully connected layers. Yann LeCun's CNN architecture is considered one of the foundational models in deep learning [30], as shown in Fig. 1.

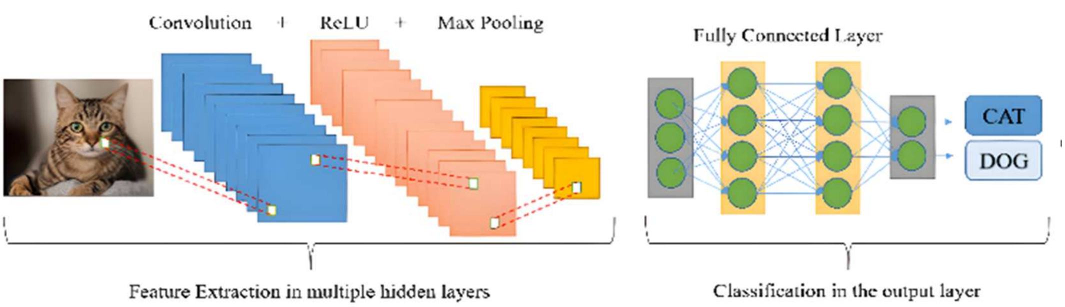

Fig. 1 Architecture of a simple CNN model by Yann LeCun [30]

The advancements in deep learning have led to the development of CNNs with specialized layers, which are used to automate the feature extraction process and improve the accuracy of object classification, localization, detection, identification, or segmentation tasks [35]. Furthermore, deep learning has opened up new possibilities for problem-solving, inspiring researchers to overcome previous limitations in the field [36]. Various machine learning methods, such as decision trees, support vector machines, and linear regression, also contribute to the learning process [31]–[33].

YOLOv4, the latest version of the CNN-based architecture, has been used to recognize symbols in electrical drawings with an accuracy rate of over 80% [34]. The use of CNN as the base of all detection and recognition models provides a better solution for the problem of digitizing and analyzing the elements in ED, and several deep learning-based neural networks have been developed for this purpose in recent years

[35]. Previous research has focused on symbol and text or character detection and recognition, which will be discussed in the following subsections [9], [10], and [35].

TABLE I
THE SUMMARY OF PREVIOUS RESEARCH WORK REGARDING SYMBOL AND TEXT OR CHARACTER

Ref.	Proposed Technique	Evaluation Metrics	Issues
[36]	Heuristic-based Image Processing	Accuracy (96.52%)	Symbol and text overlapping issue
[9]	Heuristic-based CNN	Accuracy (95.84%)	The limited size of the dataset
[2]	YOLO & GAN model	Accuracy (94%)	Focus on symbol detection
[37]	CNN	Precision (90%)	Classification only
[38]	CNN	Accuracy (95%)	Focus on symbol classification only
[6]	YOLO	Accuracy (80%)	Focus on symbol classification
[34]	YOLOv4	Accuracy (80%)	Focus on symbol detection
[4]	EAST & LSTM model	Accuracy (86%)	Used pretrained EAST model in detecting text
[39]	Hybrid of CNN-RNN-LSTM	Accuracy (95.2%)	Not tested in ED
[40]	Multi-Channel CNN (MCCNN)	Accuracy (93%)	Not tested in ED
[41]	Faster R-CNN with multiple RPN	Precision (91.81%),	Not tested in ED
[42]	CRNN	Precision (85.35%),	Test on the front view of railway CAD drawing
[16]	CNN (ResNET-50)	-	No information regarding the percentage of matrices.
[35]	CNN	Accuracy (98%)	Classification only

A. Symbols

In 2018, researchers presented semi-automatic and heuristic methods for recognizing and localizing symbols in P&ID drawings [9]. They used both supervised and unsupervised learning methods to enhance classification precision and achieved an accuracy of 95% when tested on a dataset of symbols representing the drawing standard. However, due to the small size of the dataset, consisting of only 37 symbol kinds in total, there is a lack of a consistent reference dataset [9].

In 2020, Mani et al. [37] proposed a pipeline for digitizing P&IDs automatically using computer vision techniques. This pipeline uses a trained CNN for symbol detection with an average precision of over 90% and a graph search strategy to find relationships between symbols through lines. The proposed method enables various applications, such as equipment-to-sensor mapping and asset hierarchy creation, but the author noted that overlapping elements remain a challenging task that requires further improvement.

The use of CNN for symbol detection in ED is becoming increasingly popular, as seen in studies by authors [16] and [35], who have claimed a high level of accuracy in their proposed methods. While the specific elements and types of drawings were disclosed in their research, it is believed that symbol detection is a crucial aspect of their studies.

Luo et al. [34] proposed an engineering drawing identification approach based on the YOLOv4 algorithm, which can identify component targets under circuit diagrams with accuracies ranging from 83% to 97% for 14 selected categories of electrical components. The YOLOv4 algorithm, unlike traditional manual methods, can learn about picture attributes more comprehensively and generate predictions based on information from the entire picture, making it a more efficient approach for drawing audits. While the accuracies for two components reached 97%, the rest were below that, leading the authors to propose increasing the robustness and precision of the remaining components by 97%.

As the use of YOLO for object detection and recognition continues to evolve, the latest versions, YOLOv5, YOLOv6, and YOLOv7, have made significant progress [43], [44]. However, accurate symbol detection in EDs still depends on having access to sufficient training datasets, which are currently not publicly available [2], [34]. As a result, the process of manually labeling or annotating symbols in EDs can be time-consuming. To fully digitalize EDs and transform them into CAD format, it is necessary to accurately recognize graphical elements [21], [26].

B. Text and Character

Despite the success of deep learning methods for detecting and recognizing text and characters in EDs, challenges still exist, especially regarding complex representations of text and its proximity to other drawing elements. Jamieson et al. [4] proposed a text detection and recognition method in EDs using EAST and LSTM, which achieved a 90% detection rate for text strings, including vertical text strings. However, certain non-text diagram elements were detected as text. To address this, text recognition was used to obtain text strings in 86% of the cases where text was detected. Nonetheless, the authors pointed out the need for further improvement in the detection of more complex representations of text, such as those located near other drawing elements [36], [45].

Geetha et al. [39] proposed a deep neural network model for hybrid handwritten text recognition (H2TR) using a sequence-to-sequence (Seq2Seq) approach in the following year. The model utilizes the features of CNN and recurrent neural networks (RNN) with the LSTM network. The CNN extracts the features from the handwritten image, which are then modeled using a sequenced approach and passed to the RNN-LSTM to encode the visual features and decode the available letter sequences in the handwritten image. The authors evaluated their model using integrated assessment modeling (IAM) [46] and Recognition and Indexing of Handwritten Documents and Faxes Handwritten Databases (RIMES) [47], which demonstrated competitive word accuracy results of 95.2% and 98.14%, respectively. Nonetheless, the author concluded that this method is only suitable for text in sequential order, whereas the text in ED is not sequential.

Chiney et al. [40] and Nagaoka et al. [41] proposed innovative automated handwritten text recognition methods. Chiney et al. utilized the Multi-Channel CNN (MCCNN) model and anchor-based object detection method, which were trained on three versions of hybrid datasets, to achieve a high classification accuracy above 93% for digit, alphabet, and digit and alphabet datasets. Meanwhile, Nagaoka et al. [41] proposed a multi-feature method on Faster R-CNN for robust text detection in natural scene images, achieving an 83.33% higher F-score than the baseline model, although the proposed method takes more time to execute than the baseline Faster R-CNN model.

Fang and Yin [42] developed a convolutional recurrent neural network (CRNN) for text recognition in train control system drawings that are difficult to automatically recognize due to the text's various changes, such as rotation, tilt, font changes, and proximity to lines. To address this issue, the authors trained their CRNN model using 6 million Chinese and 7 million English entry images, focusing on single-word rotation through iterative training with post-processing techniques such as tilting, noise addition, and blurring. As a result, their model achieved an accuracy of 98.36% in text recognition, demonstrating its ability to handle complex scenarios.

While OCR was traditionally utilized to digitize letters and numbers, with the aid of algorithms to differentiate between handwritten and machine-written numbers [48], the emergence of deep learning approaches such as those used by Geetha et al. [39], Abhinandan Chiney et al. [40], and Nagaoka et al. [41], as well as the detection and recognition of text in ED by Jamieson et al. [2], reveals a shared emphasis on the identification of textual or character information within images through various methods [49]. It is necessary to accurately recognize graphical elements [21], [26].

C. *Industrial Practice on ED Analysis*

The digital recognition of various elements within engineering drawings is an essential process. These elements are text, symbols, and characters, as illustrated in Fig. 2. Image conversion and pre-processing techniques are commonly used to prepare these elements for classification and detection. In recent years, CAD inspection software, also known as ballooning software, has gained popularity in analyzing EDs in industries. This software can perform extractions for dimensions, tolerance measurements, and notes from text-selectable AutoCAD or PDF files, as well as manage different markup edits and drawing balloons into existing structures. The software can also process PDF output and exports, as shown in the sample in Fig. 3. Overcoming the challenges related to ED file improvement is crucial for accurate detection and recognition of ED, which requires addressing issues such as congestion, overlap, tolerance dimensions, font styles, raster pictures, superfluous components, stamps, and shading.

The process of digitizing text and characters has become easier with the advent of Optical Character Recognition (OCR) platforms, which can electronically identify and convert text or print documents into digital text documents [50]–[52]. OCR has become the first layer in text or characters [53] and is commonly used by companies and industries to automate the processing of managing physical typewritten documents, such as creating digital copies of structured documents like invoices, receipts, bank statements, and other accounting documents that need to be managed. In the case of industries, OCR software is used to analyze ED.

OCR software is widely used to convert text or printed documents into machine-readable codes for data processing, with SimpleOCR and Tesseract being some of the most popular [50]. However, OCR is limited to the structure of typewritten and handwritten documents [53]. To improve OCR technology, machine learning and computer vision algorithms have been applied, such as Intelligent Character Recognition (ICR), which can examine a document's layout to determine what data to extract and can read both typed and handwritten text with no restrictions, making it the second layer of capabilities to recognize text or characters. Although the accuracy levels for reading handwritten text may not always be exceptionally high, ICR algorithms can still achieve 97–99 percent accuracy rates in structured forms when handling capital letters and simple-to-segment figures [17]. However, ICR fails when faced with more challenging conditions, such as unconstrained texts or non-separable (such as cursive) handwriting, leading to Intelligent Word Recognition (IWR).

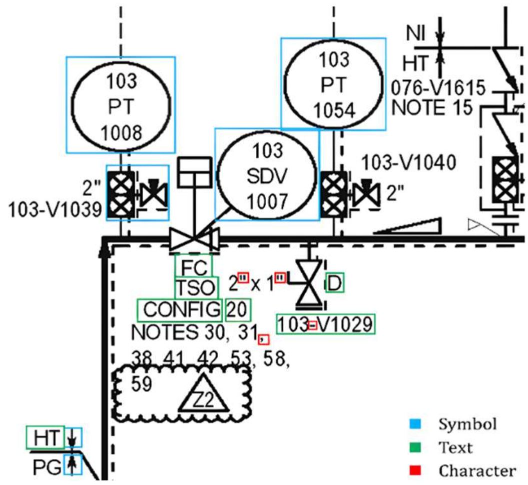

Fig. 2 Close-up view of P&ID text, symbol, and character [23]

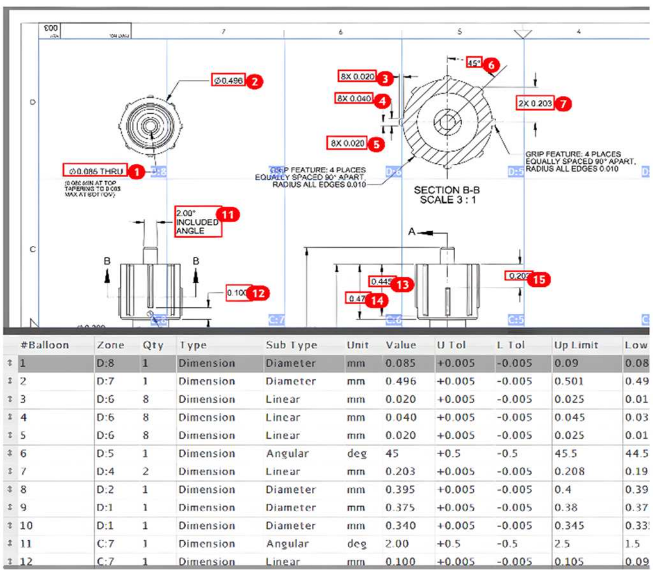

Fig. 3 Sample view of the ballooning software used for analyzing ED

Optimized for processing real-world documents that contain mostly free-form, hard-to-recognize data fields that are inherently unsuitable for ICR, Intelligent Word Recognition (IWR) is the third layer in text or character recognition that works with unstructured information (e.g., full words or phrases) from documents [53]. Although still a developing technology, some products such as Google Vision API and Amazon Textract have the ability to decode (scanned) printed or handwritten text, and a few prototypes are being tested and validated in pertinent environments, making IWR supposedly more advanced than handwritten ICR.

If the ED designs are not in compliance with the ISO's standard ED formatting structures, the accuracy of the built-in OCR in CAD inspection software for text and character recognition could drop by 80% below; however, using neural network models such as the EAST model and a LSTM model for text detection and recognition can increase accuracy to 86%, although these models cannot still identify overlapping text, symbols, and characters as highlighted in Fig. 4 and previously discussed by Cao and Tan [54] and Roy et al. [55] due to the arbitrary lengths and sizes of text strings that describe symbols [4], [56]–[59].

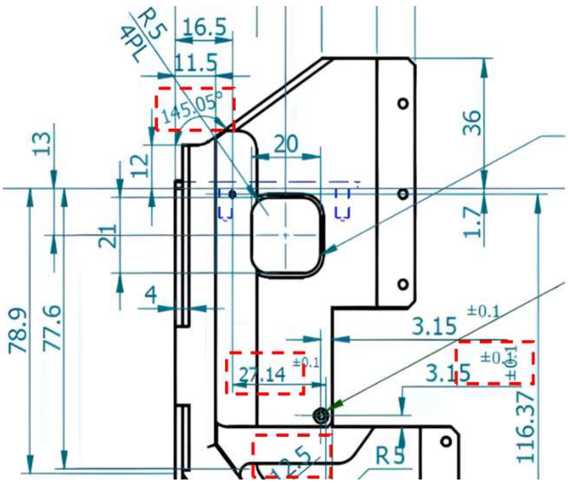

Fig. 4 Close-up view of overlapping text and symbol on ED

Although the built-in OCR in CAD inspection software can make inspecting ED documents or images for text and characters easier, misinterpretation can still occur due to error-prone text interpretation. This can make it difficult to associate corresponding text or characters with symbols, as demonstrated by the congested drawing in Fig. 5, where dimensions too close to each other can be read as one and different font styles used in the drawing result in incorrect ballooning, such as misidentification of dots as squares or inability to distinguish between a real dot and a dashed line. Despite this challenge, the use of pre-processing methods can improve the ballooning software's ability to recognize the different types of tolerances shown in Fig. 6, especially the non-formal ones, although capturing tolerance formats is still difficult. It is also crucial to convert the image before ballooning. Moreover, incorporating ICR and IWR techniques can greatly enhance text and character recognition in ED, which is essential for describing dimension and tolerance information in drawings [53].

In addition to the challenges of scaling, lighting, and pose variations in digitizing symbols, the compilation of a well-defined and clearly labeled dataset is a complex task due to the lack of a benchmark and publicly available dataset, which can make the classification of symbols problematic, as noted in the paragraph. While there have been significant efforts to recognize one or two elements in the ED using image processing, computer vision, and neural networks to achieve better solutions, the models are still under improvement as they can wrongly detect information or parts as symbols, as demonstrated in Fig. 7. In the image, the green boxes indicate the area of the symbol that the detection and recognition model should identify, but the symbol can be incorrectly detected and recognized, as shown in the red boxes.

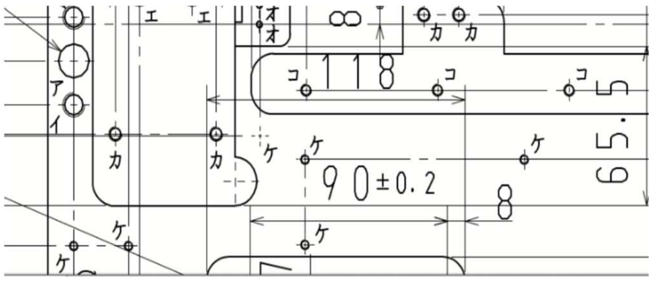

Fig. 5 Congested drawing with ISOCP font style

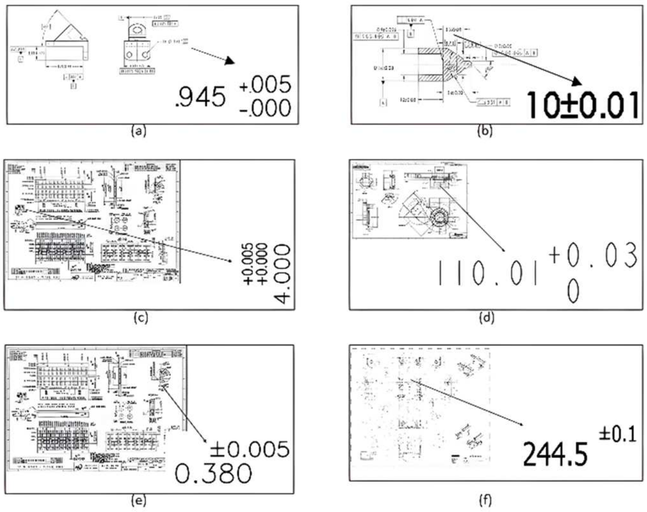

Fig. 6 Types of tolerance dimensions format in ED

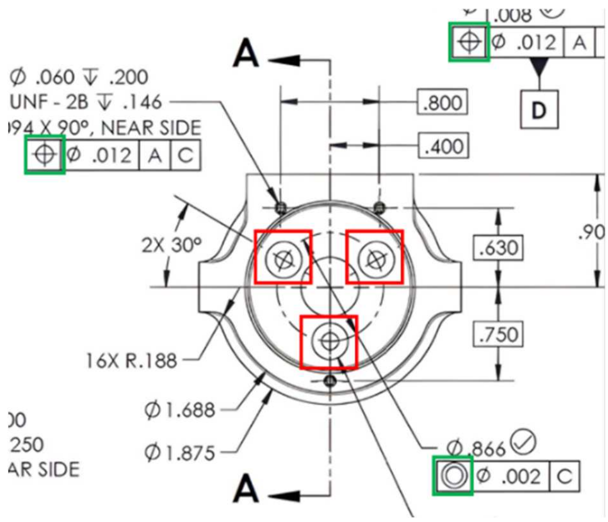

Fig. 7 Detect unnecessary parts as symbols

According to Moreno et al. [19], CNN is a recognition model that offers great affinity and functionality when used for computer vision tasks, given its capability to classify a wide pool of images with varying sizes and characteristics, as noted in the paragraph. Some researchers believe that complex ED digitization can be addressed through multiple CNN-based techniques [60]–[62], including AlexNet [63], VGGNet [64], ResNet [65], and Darknet [66]. The next section will discuss the current approaches that have been developed for the digitization, detection, and recognition of ED.

III. Result and Discussion

A. Proposed Approach

While machine learning techniques, such as neural networks, show promise as an end-to-end approach to detecting symbols, text, and characters in electronic documents (EDs), there are still challenges to overcome, such as the availability of annotated training data and the effort required for training [19]. Recent works in the past five years, as shown in Table 1, highlight the gaps in current approaches, indicating the need for further research to address these challenges. Although studies on ED digitization have been conducted for over three decades, significant obstacles still need to be addressed to improve the suitability of the existing models [3].

Despite the effectiveness of CNNs in reducing the number of parameters and producing high-quality models for detecting and recognizing any element in an ED [67], the model is still imperfect, as it struggles to classify images with a degree of tilt or rotation [68]. To overcome this limitation, researchers have developed several approaches, including a combination of algorithms, techniques, and methods, to improve the performance of the existing CNN model [67].

Several studies have attempted to improve the CNN model for detecting and localizing symbols in ED, such as the heuristic-based method proposed by [9] and the Faster R-CNN technique with multiple RPN to detect texts from different resolution feature maps by [41]. Hybrid models, utilizing existing CNN models, have also been experimented with by researchers, as seen in the works of [2] and [39]. Meanwhile, [40] improved a single model using two techniques. Despite these efforts, it can be observed that the recognition of ED symbols is still undergoing algorithmic enhancements within the existing model.

Deep learning has become increasingly popular for various purposes and applications, including computer vision, recognition, clustering, detection, segmentation, and classification. CNNs have been widely utilized due to their high performance in computer vision applications. In the context of ED digitization, CNN-based models have been developed to address specific tasks despite the usual limitations caused by rotation, translation, degradation, and overlapping, among others, which can affect symbol classification accuracy. However, the need to manually collect and correct large amounts of input data for training still poses a significant challenge. Several methods based on data augmentation have been proposed, including affine transformations of pre-existing data samples to generate more data for a specific class.

Many techniques can be used to optimize neural network models and reduce overfitting, such as transfer learning, which involves taking part in an existing pre-trained model and making minimal modifications to it to fit the current problem [64]. Additionally, fine-tuning the weights or layers can be done based on the expected outcome or domain of the model [64]. Another effective method is data augmentation, where pre-existing data samples are transformed using affine transformations to produce more data for specific classes [68]. Overfitting occurs when a model is too complicated for the problem it is trying to solve, leading to poor performance against new testing data [68]. Thus, it is crucial to employ techniques such as data augmentation, transfer learning, and model optimization to improve the accuracy and precision of domain prediction.

B. Analyzing Approach

This section discusses and evaluates prior research and industry practices related to ED. While progress has been made in detecting and recognizing ED elements, achieving 100% accuracy is still impossible due to factors that persist during the analysis process. To understand the novelty of this study, a review of prior research in the ED element detection field is presented. Despite notable progress, the ED analysis process remains highly challenging due to various factors. The nature of the ED is one of the key factors affecting its analysis process. Different kinds of ED exist, including mechanical, electrical, electronic, and civil drawings, which can be further categorized as systems, infrastructures, structures, or model drawings. The accuracy and precision of ED element detection and recognition largely depend on the specific type of drawing and the proposed method. Researchers such as [2], [4], [9], [36], and [37] utilize a mechanical drawing known as P&ID, while [38] uses an electronic drawing dataset to obtain high accuracy and precision ranging from 86% to 97%.

The second factor affecting ED detection and recognition of EDs is the file format used for testing. EDs can be found in various formats, including DXF, PDF, STEP, and image formats like TIFF and PNG. However, due to the prevalence of scanned images of older EDs, existing digitization methods often use raster images, which have low resolution. A pre-processing step is needed to enhance the quality of these images, which involves techniques such as thresholding to reduce noise, background subtraction, color conversion, pattern matching, and convolution-based approaches [9], [36]. Another approach to improving scanned images is to convert them from raster images to vector images, which is effective in previous research [69]. For example, Marek Ciezobka used a pixel vector field approach with a Hopefield neural network model to convert scanned ED images into CAD formats (DWG and DXF) [69]. Although no details are available on the evaluation metrics used, the author claims that scanned ED images can be recreated and suggests further development for recognition.

The third factor concerns the availability of an ED dataset. As with any field of deep learning, training requires a dataset to learn and improve. Unfortunately, some research papers utilize proprietary datasets that are not accessible to the public, while others employ existing datasets that can be found online. However, in the case of ED datasets, there is a lack of availability compared to other datasets like animals, groceries, human body parts, and stationery, which have multiple datasets accessible online, such as MS-COCO, ImageNet, and CIFAR-10. Currently, the only available ED dataset is a symbol dataset made public by the author [9] through GitHub.

The subsequent factor to consider is the viability of utilizing a deep learning approach that can recognize and detect all elements in ED. From 2017 to 2022, researchers attempted to use CNN for symbols, text, and characters as a basic deep-learning approach. However, the author [2] discovered a problem that needed to be solved to detect and recognize the elements in ED completely. The biggest challenge is related to overlapping elements and congested drawings, which have been discussed since 2002 by Lu [70], where graphics or symbols must be separated first. The author suggests that the Hough transform and a rule-based algorithm be used to segment the elements. In [2], [9], and [36], the authors use the Hough transform and SVM to separate text from symbols, while [37] uses a lighter CNN to segment text in ED. In [4], a segmentation method is also described. The fully end-to-end neural network is still under development, and the suitability of the existing deep learning approach is still being tested and not yet implemented in a real-world situation.

One of the most popular approaches is the use of YOLO-based methods. Many researchers, programmers, engineers, and technicians are currently testing YOLOv7. Wang et al. [44], the founders of YOLOv7, trained this model using the COCO dataset and obtained the highest accuracy of the real-time model (YOLOv7-E6E, 56.8% AP) [44]. In this research, YOLOv7 CNN architecture was employed to assess its effectiveness in symbol detection within engineering drawings. The results revealed a promising potential for accurate symbol detection.

However, achieving higher accuracy will require further optimization and fine-tuning of the model. Table II shows the results of the YOLOv7 model training at different epochs (1st to 500th) with a batch size of 16. The model was trained on 207 ED images and tested on 21 ED images. The mean average precision result obtained is 0.23 mAP, shown in red boxes. Fig. 8 displays the training result of ED using YOLOv7. The model's performance is evaluated based on several metrics: Precision (P), Recall (R), mean Average Precision at IoU threshold 0.5 (mAP_0.5), and mean Average Precision over IoU thresholds from 0.5 to 0.95 (mAP_0.5:0.95). Additionally, the time taken for each training epoch is recorded in minutes.

The YOLOv7 results from previous research work in Table II and Table I compare various techniques used for engineering drawing (ED) element detection and recognition. The previous research encompasses different methods, each with its respective evaluation metrics and limitations.

In [36], a heuristic-based image processing technique achieved an accuracy of 96.52%, but it faced challenges with symbol and text overlapping. [9] employed a heuristic-based CNN with an accuracy of 95.84%, but the limited size of the dataset was a concern. The YOLO & GAN model in [2] achieved an accuracy of 94% but primarily focused on symbol detection. Similarly, [37] and [38] used CNNs with precision and accuracy of 90% and 95%, respectively, but were specialized in symbol classification. The YOLO model in [6] had an accuracy of 80% and focused solely on symbol classification, while YOLOv4 in [34] also achieved 80% accuracy but concentrated on symbol detection.

Other techniques like EAST & LSTM model [4] attained an accuracy of 86% but used a pretrained EAST model for detecting text. The hybrid of CNN-RNN-LSTM in [39] achieved 95.2% accuracy but was not tested on ED. MCCNN in [40] achieved an accuracy of 93% but, similar to other approaches, was not evaluated on ED. Faster R-CNN with multiple RPN in [41] achieved a precision of 91.81%, but its evaluation did not include ED testing. CRNN in [42] reached a precision of 85.35% but was tested on front views of railway CAD drawings, not ED. [16] used CNN (ResNET-50) without specific metrics, and [35] achieved 98% accuracy but only focused on classification.

Comparing these prior techniques with the results of YOLOv7 in this paper, it is evident that YOLOv7 shows promise in ED detection and recognition. While some previous methods achieved high accuracy, they often had limitations such as focusing on specific tasks (e.g., symbol detection or classification only), using pre-trained models, or not being tested on ED datasets. YOLOv7, on the other hand, exhibits a balanced performance in terms of precision, recall, and mAP scores, and its training process shows continuous improvement over epochs. Therefore, the paper's contribution lies in providing insights into the challenging ED analysis process while demonstrating the potential of YOLOv7 in addressing these challenges and its suitability for practical implementation in the industry.

TABLE II
YOLOV7 PERFORMANCE METRICS DURING TRAINING

Epochs	P	R	mAP_0.5	mAP_0.5:0.95
50	0.823	0.545	0.597	0.415
100	0.558	0.737	0.72	0.525
150	0.668	0.739	0.764	0.559
200	0.761	0.775	0.808	0.602
250	0.857	0.752	0.819	0.611
300	0.876	0.744	0.823	0.611
400	0.851	0.754	0.815	0.606
500	0.864	0.758	0.803	0.604

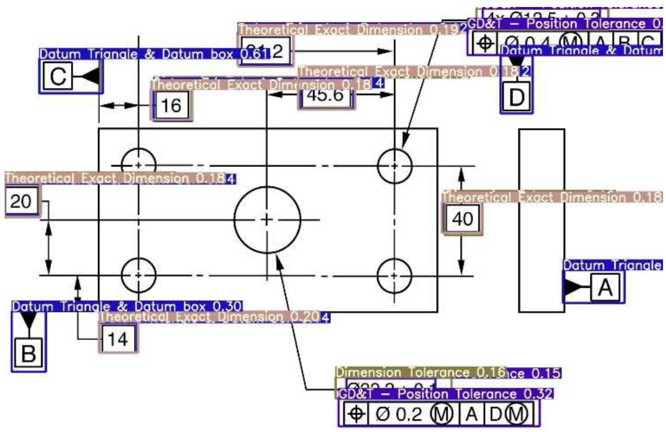

Fig. 8 YOLOv7 result on ED dataset.

Besides, the pre-processing procedures required to generate an improved ED image are often unavailable in the sample CAD inspection software. Wang et al. [71] have recently shown that a neural network model can be used to enhance the resolution of an image. The model, called Real-ESRGAN and developed by [71], can upscale raster images by utilizing training pairs and a more effective degradation process, aiding in image enhancement and restoration. The results of applying the Real-ESRGAN model to a mechanical ED image before and after enhancement are presented in Fig. 9(a) and 9(b), respectively.

The digitization of ED still requires a significant amount of research, although previous researchers have succeeded in their methods and aim to achieve 100% detection and recognition of the ED elements they focus. Based on the previous work discussed, an end-to-end solution is feasible.

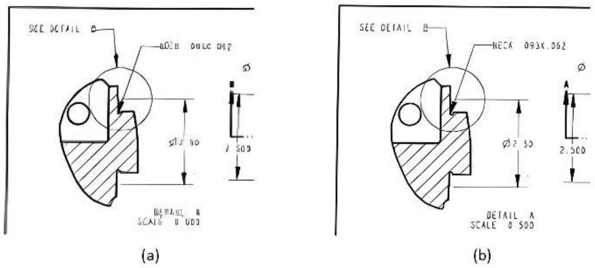

Fig. 9 The enhancement of ED image using the Real-ESRGAN model

IV. Conclusion

This research contributes to engineering drawing (ED) digitization by addressing the challenging ED element detection and recognition process. The study highlights the significance of neural networks, particularly YOLOv7, in improving the accuracy and efficiency of this critical process in both academic and industrial settings. A comprehensive review of previous works shows that achieving 100% accuracy in ED analysis remains elusive due to various persistent factors. However, the results obtained from the YOLOv7 model demonstrate promising performance, with progressively improving precision, recall, and mAP scores across training epochs.

By comparing the YOLOv7 results with prior research techniques, it becomes evident that YOLOv7 exhibits a balanced and versatile performance in ED detection and recognition. Unlike previous approaches focused solely on symbol detection or classification, YOLOv7 simultaneously demonstrates competence in both tasks. Furthermore, its reliance on neural networks enhances adaptability and potential applications in various industries.

This paper's findings underscore the importance of continued research and development in ED digitization, as it plays a pivotal role in transforming manual workflows into efficient and automated processes. The utilization of neural networks, particularly YOLOv7, promises to revolutionize ED analysis, paving the way for more accurate, reliable, and time-efficient practices in academic research and industrial applications.

This study opens new avenues for future research, such as exploring ensemble methods, fine-tuning network parameters, and conducting extensive real-world testing. We hope this research will inspire further investigations and collaborations to overcome the existing challenges in ED digitization, bringing us closer to achieving higher accuracy and efficiency in this critical domain. Integrating advanced technologies like YOLOv7 in ED analysis will contribute significantly to advancements in engineering, architecture, and various other industries that rely on accurate digital representations of complex drawings.

Acknowledgment

This research is supported by the Ministry of Higher Education (MOHE) under Fundamental Research Grant Scheme (FRGS/1/2020/ICT02/UTHM/03/4). The authors also thank the Faculty of Computer Science and Information Technology (FSKTM) for allowing us to use all of the facilities in the faculty for this research purpose.

References

[1] F. Chollet and others, *Deep learning with Python*, vol. 361. Manning New York, 2018.

[2] E. Elyan, L. Jamieson, and A. Ali-Gombe, "Deep learning for symbols detection and classification in engineering drawings," Neural Networks, vol. 129, pp. 91–102, Sep. 2020, doi: 10.1016/j.neunet.2020.05.025.

[3] B. Scheibel, J. Mangler, and S. Rinderle-Ma, "Extraction of dimension requirements from engineering drawings for supporting quality control in production processes," Computers in Industry, vol. 129, p. 103442, Aug. 2021, doi: 10.1016/j.compind.2021.103442.

[4] L. Jamieson, C. F. Moreno-Garcia, and E. Elyan, "Deep Learning for Text Detection and Recognition in Complex Engineering Diagrams," in *2020 International Joint Conference on Neural Networks (IJCNN)*, 2020, pp. 1–7.

[5] K. Mahasivabhattu, D. Bandi, S. K. Singh, P. Kumar, and others, "Engineering Data Management Using Artificial Intelligence," in *Offshore Technology Conference*, 2019.

[6] Y. Zhao, X. Deng, and H. Lai, "A Deep Learning-Based Method to Detect Components from Scanned Structural Drawings for Reconstructing 3D Models," Applied Sciences, vol. 10, no. 6, p. 2066, Mar. 2020, doi: 10.3390/app10062066.

[7] N. Kriegeskorte, "Deep neural networks: a new framework for modeling biological vision and brain information processing," *Annu Rev Vis Sci*, vol. 1, pp. 417–446, 2015.

[8] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Physica D*, vol. 404, p. 132306, 2020.

[9] E. Elyan, C. M. Garcia, and C. Jayne, "Symbols classification in engineering drawings," in *2018 International Joint Conference on Neural Networks (IJCNN)*, 2018, pp. 1–8.

[10] K. Seeliger *et al.*, "Convolutional neural network-based encoding and decoding of visual object recognition in space and time," *Neuroimage*, vol. 180, pp. 253–266, 2018.

[11] D. Wang, J. Fan, H. Fu, and B. Zhang, "Research on optimization of big data construction engineering quality management based on RNN-LSTM," *Complexity*, vol. 2018, 2018.

[12] S. Ren, K. He, R. Girshick, X. Zhang, and J. Sun, "Object Detection Networks on Convolutional Feature Maps," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 7, pp. 1476–1481, Jul. 2017, doi: 10.1109/tpami.2016.2601099.

[13] J. Dekhtiar, A. Durupt, M. Bricogne, B. Eynard, H. Rowson, and D. Kiritsis, "Deep learning for big data applications in CAD and PLM--Research review, opportunities and case study," *Comput Ind*, vol. 100, pp. 227–243, 2018.

[14] D. Benjamin, P. Forgues, E. Gulko, J. Massicotte, and C. Meubus, "The use of high-level knowledge for enhanced entry of engineering drawings," [1988 Proceedings] 9th International Conference on Pattern Recognition, doi: 10.1109/icpr.1988.28186.

[15] Chan Pyng Lai and R. Kasturi, "Detection of dimension sets in engineering drawings," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 16, no. 8, pp. 848–855, 1994, doi:10.1109/34.308483.

[16] D. Van Daele, N. Decleyre, H. Dubois, and W. Meert, "An Automated Engineering Assistant: Learning Parsers for Technical Drawings," Feb. 2019, [Online]. Available: http://arxiv.org/abs/1909.08552.

[17] R. Ptucha, F. Petroski Such, S. Pillai, F. Brockler, V. Singh, and P. Hutkowski, "Intelligent character recognition using fully

convolutional neural networks," Pattern Recognition, vol. 88, pp. 604–613, Apr. 2019, doi: 10.1016/j.patcog.2018.12.017.
[18] S. H. Joseph and T. P. Pridmore, "Knowledge-directed interpretation of mechanical engineering drawings," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, no. 9, pp. 928–940, 1992, doi: 10.1109/34.161351.
[19] C. F. Moreno-García, E. Elyan, and C. Jayne, "New trends on digitisation of complex engineering drawings," Neural Computing and Applications, vol. 31, no. 6, pp. 1695–1712, Jun. 2018, doi: 10.1007/s00521-018-3583-1.
[20] D. Blostein, "General diagram-recognition methodologies," Lecture Notes in Computer Science, pp. 106–122, 1996, doi: 10.1007/3-540-61226-2_10.
[21] K. Tombre, "Analysis of engineering drawings: State of the art and challenges," in *International Workshop on Graphics Recognition*, 1997, pp. 257–264.
[22] T. Kanungo, R. M. Haralick, and D. Dori, "Understanding Engineering Drawings: A Survey," in *Proceedings of First IARP Workshop on Graphics Recognition*, Citeseer, Ed., 1995, pp. 119–130.
[23] Y. Lu, "Machine printed character segmentation —; An overview," Pattern Recognition, vol. 28, no. 1, pp. 67–80, Jan. 1995, doi: 10.1016/0031-3203(94)00068-w.
[24] S. U. Ahmad, C. R. Kulkarni, and A. B. Barbadekar, "Text Detection and Recognition: A Review IRJET Journal Art ificial Neural Net works for Document Analysis and Recognit ion Text Detection and Recognition: A Review," *International Research Journal of Engineering and Technology (IRJET)*, vol. 4, no. 6, pp. 179–185, 2017, [Online]. Available: www.irjet.net.
[25] L. P. Cordella and M. Vento, "Symbol recognition in documents: a collection of techniques?," International Journal on Document Analysis and Recognition, vol. 3, no. 2, pp. 73–88, Dec. 2000, doi:10.1007/s100320000036.
[26] S V Ablameyko and S. Uchida, "Recognition of engineering drawing entities: review of approaches," *Int J Image Graph*, vol. 7, no. 04, pp. 709–733, 2007.
[27] V. Nagasamy and N. A. Langrana, "Engineering drawing processing and vectorization system," *Comput Vis Graph Image Process*, vol. 49, no. 3, pp. 379–397, 1990.
[28] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
[29] K. Sun, J. Zhang, C. Zhang, and J. Hu, "Generalized extreme learning machine autoencoder and a new deep neural network," *Neurocomputing*, vol. 230, pp. 374–381, 2017.
[30] Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio, "Object Recognition with Gradient-Based Learning," Lecture Notes in Computer Science, pp. 319–345, 1999, doi: 10.1007/3-540-46805-6_19.
[31] V. Podgorelec, P. Kokol, B. Stiglic, and I. Rozman, Journal of Medical Systems, vol. 26, no. 5, pp. 445–463, 2002, doi:10.1023/a:1016409317640.
[32] K. Florence, "Logistic regression and Kernelized SVM," pp. 1–5, 2014.
[33] W. D. Pan, Y. Dong, and D. Wu, "Classification of Malaria-Infected Cells Using Deep Convolutional Neural Networks," Machine Learning - Advanced Techniques and Emerging Applications, Sep. 2018, doi: 10.5772/intechopen.72426.
[34] Y. Luo, T. Yu, J. Zheng, and Y. Ding, "Design of engineering drawing recognition system based on Yolo V4," 2022 IEEE 6th Information Technology and Mechatronics Engineering Conference (ITOEC), Mar. 2022, doi: 10.1109/itoec53115.2022.9734453.
[35] L. li, C. Yuhui, and L. Xiaoting, "Engineering Drawing Recognition Model with Convolutional Neural Network," Proceedings of the 2019 International Conference on Robotics, Intelligent Control and Artificial Intelligence, Sep. 2019, doi: 10.1145/3366194.3366213.
[36] C. F. Moreno-García, E. Elyan, and C. Jayne, "Heuristics-Based Detection to Improve Text/Graphics Segmentation in Complex Engineering Drawings," Communications in Computer and Information Science, pp. 87–98, 2017, doi: 10.1007/978-3-319-65172-9_8.
[37] S. Mani, M. A. Haddad, D. Constantini, W. Douhard, Q. Li, and L. Poirier, "Automatic Digitization of Engineering Diagrams using Deep Learning and Graph Search," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 176–177.
[38] H. Wang, T. Pan, and M. K. Ahsan, "Hand-drawn electronic component recognition using deep learning algorithm," *International Journal of Computer Applications in Technology*, vol. 62, no. 1, pp. 13–19, 2020.
[39] R. Geetha, T. Thilagam, and T. Padmavathy, "Effective offline handwritten text recognition model based on a sequence-to-sequence approach with CNN–RNN networks," Neural Computing and Applications, vol. 33, no. 17, pp. 10923–10934, Jan. 2021, doi:10.1007/s00521-020-05556-5.
[40] A. Chiney et al., "Handwritten Data Digitization Using an Anchor based Multi-Channel CNN (MCCNN) Trained on a Hybrid Dataset (h-EH)," Procedia Computer Science, vol. 189, pp. 175–182, 2021, doi: 10.1016/j.procs.2021.05.095.
[41] Y. Nagaoka, T. Miyazaki, Y. Sugaya, and S. Omachi, "Text Detection Using Multi-Stage Region Proposal Network Sensitive to Text Scale," Sensors, vol. 21, no. 4, p. 1232, Feb. 2021, doi: 10.3390/s21041232.
[42] Y.-Y. Fang and Z.-H. Yin, "A Text Correction and Recognition for Intelligent Railway Drawing Detection," 2021 IEEE 16th Conference on Industrial Electronics and Applications (ICIEA), Aug. 2021, doi:10.1109/iciea51954.2021.9516318.
[43] Do Thuan, "Evolution of YOLO Algorithm and YOLOv5: The State-of-The-Art Object Detection Algorithm," Oulu University of Applied Sciences, 2021. Accessed: Mar. 22, 2022. [Online]. Available: https://www.theseus.fi/handle/10024/452552.
[44] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," *ArXiv*, Jul. 2022, [Online]. Available: http://arxiv.org/abs/2207.02696.
[45] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network," Physica D: Nonlinear Phenomena, vol. 404, p. 132306, Mar. 2020, doi:10.1016/j.physd.2019.132306.
[46] U.-V. Marti and H. Bunke, "A full English sentence database for off-line handwriting recognition," in *Proceedings of the Fifth International Conference on Document Analysis and Recognition. ICDAR'99 (Cat. No. PR00318)*, 1999, pp. 705–708.
[47] F. Menasri, J. Louradour, A.-L. Bianne-Bernard, and C. Kermorvant, "The A2iA French handwriting recognition system at the Rimes-ICDAR2011 competition," in *Document Recognition and Retrieval XIX*, 2012, p. 82970Y.
[48] P. Banerjee *et al.*, "Automatic Creation of Hyperlinks in AEC Documents by Extracting the Sheet Numbers Using LSTM Model," in *TENCON 2018-2018 IEEE Region 10 Conference*, 2018, pp. 1667–1672.
[49] M. Shaker and M. ElHelw, "Optical character recognition using deep recurrent attention model," in *Proceedings of the 2nd International Conference on Robotics, Control and Automation*, 2017, pp. 56–59.
[50] A. F. Biten, R. Tito, L. Gomez, E. Valveny, and D. Karatzas, "OCR-IDL: OCR Annotations for Industry Document Library Dataset," *ArXiv*, Feb. 2022, doi: 10.48550/arXiv.2202.12985.
[51] H. Takahashi, N. Itoh, T. Amano, and A. Yamashita, "A spelling correction method and its application to an OCR system," Pattern Recognition, vol. 23, no. 3–4, pp. 363–377, Jan. 1990, doi:10.1016/0031-3203(90)90023-e.
[52] N. H. Imam, V. G. Vassilakis, and D. Kolovos, "OCR post-correction for detecting adversarial text images," Journal of Information Security and Applications, vol. 66, p. 103170, May 2022, doi:10.1016/j.jisa.2022.103170.
[53] F. Martínez-Plumed, E. Gómez, and J. Hernández-Orallo, "Futures of artificial intelligence through technology readiness levels," Telematics and Informatics, vol. 58, p. 101525, May 2021, doi:10.1016/j.tele.2020.101525.
[54] R. Cao and C. L. Tan, "Text/Graphics Separation in Maps," Graphics Recognition Algorithms and Applications, pp. 167–177, 2002, doi:10.1007/3-540-45868-9_14.
[55] P. P. Roy, E. Vazquez, J. Lladós, R. Baldrich, and U. Pal, "A System to Segment Text and Symbols from Color Maps," Graphics Recognition. Recent Advances and New Opportunities, pp. 245–256, doi: 10.1007/978-3-540-88188-9_23.
[56] G. Henzold, *Geometrical dimensioning and tolerancing for design, manufacturing and inspection: a handbook for geometrical product specification using ISO and ASME standards*, 2nd ed., vol. 3. Oxford, UK: Elsevier, 2006.
[57] I. Cristofolini, "Datums Concepts by Asme and ISO Standards," AMST'02 Advanced Manufacturing Systems and Technology, pp. 641–648, 2002, doi: 10.1007/978-3-7091-2555-7_74.
[58] I. Popov and S. Onuh, "Critical notes and considerations on the use of ISO 286-1 for CAD modelling and rapid product development," International Journal of Agile Systems and Management, vol. 2, no. 2, p. 214, 2007, doi: 10.1504/ijasm.2007.015790.

[59] S. Tornincasa, “Geometrical Specification for Non-Rigid Parts,” Springer Tracts in Mechanical Engineering, pp. 283–290, Nov. 2020, doi: 10.1007/978-3-030-60854-5_12.

[60] A. Dhillon and G. K. Verma, “Convolutional neural network: a review of models, methodologies and applications to object detection,” Progress in Artificial Intelligence, vol. 9, no. 2, pp. 85–112, Dec. 2019, doi: 10.1007/s13748-019-00203-0.

[61] A. Ajit, K. Acharya, and A. Samanta, “A Review of Convolutional Neural Networks,” 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), Feb. 2020, doi: 10.1109/ic-etite47903.2020.049.

[62] A.-A. Tulbure, A.-A. Tulbure, and E.-H. Dulf, “A review on modern defect detection models using DCNNs – Deep convolutional neural networks,” Journal of Advanced Research, vol. 35, pp. 33–48, Jan. 2022, doi: 10.1016/j.jare.2021.03.015.

[63] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks.” [Online]. Available: http://code.google.com/p/cuda-convnet/.

[64] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, International Conference on Learning Representations, ICLR, 2015.

[65] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, “Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning.” [Online]. Available: www.aaai.org.

[66] J. Redmon and A. Farhadi, “Yolov3: An incremental improvement,” *arXiv preprint arXiv:1804.02767*, 2018.

[67] M. A. El-Sayed and M. A. Khafagy, “Automated Edge Detection Using Convolutional Neural Network,” 2013. [Online]. Available: www.ijacsa.thesai.org.

[68] N. Yao, G. Shan, and X. Zhu, “Substation Object Detection Based on Enhance RCNN Model,” 2021 6th Asia Conference on Power and Electrical Engineering (ACPEE), Apr. 2021, doi:10.1109/acpee51499.2021.9437086.

[69] M. Cieżobka, “Conversion of raster images into vector graphics,” *Czasopismo Techniczne. Mechanika*, vol. 105, pp. 3–10, 2008.

[70] Zhaoyang Lu, “Detection of text regions from digital engineering drawings,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 4, pp. 431–439, Apr. 1998, doi:10.1109/34.677283.

[71] X. Wang, L. Xie, C. Dong, and Y. Shan, “Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data,” *ArXiv*, Jul. 2021, [Online]. Available: http://arxiv.org/abs/2107.10833