Document Sanctity Check Using Machine Learning

Thesis by
Muhammed Faris Mukthar M V

In Partial Fulfillment of the Requirements for the
Degree of
M.Sc. Computer Science With Specialization in Data Analytics

SCHOOL OF DIGITAL SCIENCES
KERALA UNIVERSITY OF DIGITAL SCIENCE, INNOVATION AND TECHNOLOGY
Trivandrum, Kerala

Supervisors: Dr.Anoop V.S

2024
Defended [Exact Date]

Muhammed Faris Mukthar M V
ORCID: [Author ORCID]

BONAFIDE CERTIFICATE

This is to certify that the project report entitled "Document Sanctity Check Using Machine Learning" submitted by **Muhammed Faris Mukthar M V (Reg. No: 223037)** in partial fulfilment of the requirements for the award of Master of Science in Computer Science with Specialization in Data Analytics is a bonafide record of the work carried out at School of Digital Sciences under our supervision.

Guide	**Internal Guide**
Dr. Anoop V.S Research Officer, School of Digital Sciences,	Dr. Anoop V.S Research Officer, School of Digital Sciences,

DECLARATIONS

I, Muhammed Faris Mukthar M V, student of Masters of Science in Computer Science with specialization in Data Analytics, hereby declare that this report is substantially the result of my own work, except where explicitly indicated in the text, and has been carried out during the period February 2024 to June 2024.

Place: Thiruvananthapuram

Date: June - 2024

ACKNOWLEDGEMENTS

I would like to offer my sincere gratitude and appreciation to everyone who has helped and encouraged me along the way. I could not have accomplished my goals and gotten to where I am now without their support, direction, and encouragement..

I want to start by gratefully thanking **Dr. Anoop V. S.**, my guide, for his continuous support and guidance during my research endeavour. His knowledge, counsel, and unwavering encouragement have been extremely helpful in directing my work and inspiring me to pursue greatness.

For providing excellent instruction and a supportive study atmosphere, **Prof.John Eric Stephen** and our institution, the institution of Digital Sciences, have my sincere gratitude. I am really grateful to **Dr. Saji Gopinath**, our Vice Chancellor, for providing me with the opportunity to grow in such a remarkable university as Digital University Kerala. Their commitment to promoting an intellectually curious society has had a major influence on both my academic and personal development.

I would be negligent if I did not acknowledge the help and comprehension of my friends and family. Throughout my journey, their steadfast support, words of wisdom, and trust in me have been a continual source of strength. I consider myself very lucky to have such a strong support network.

ABSTRACT

Technological advancements have created new opportunities for document analysis tasks, such as resume shortlisting, invoice extraction, and bill summarizing, by extracting content efficiently. This project focuses on verifying documents by assessing their integrity, authenticity, and validity using machine learning techniques. Our aim was to develop an automated machine learning model for ensuring document sanctity.

To achieve this, we utilized transformers and detection models. Initially, we collected various documents relevant to our problem, including those with headers, those without headers, and various structured forms containing seals and signatures. Extracting meaningful information from these documents required a model capable of understanding their structure and context. For this purpose, we employed LayoutLMv3, a state-of-the-art model that can be fine-tuned for such tasks. Through extensive research, we identified LayoutLMv3 as the most suitable model for our needs. Additionally, we used Faster R-CNN as our detection model.

We developed a fine-tuned LayoutLMv3 model for classifying documents based on their structure and content. Integrating the Faster R-CNN model enabled us to detect signs and seals within documents, thereby verifying their integrity. By combining these two models, we created a seamless pipeline that automates the document sanctity check process. This innovative approach is expected to streamline various processes and reduce bottlenecks significantly.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Number *Page*

LIST OF TABLES

Chapter 1

INTRODUCTION

The necessity of automated document verification is becoming increasingly clear in all industries in this new digital era. The most often used form of document verification at the moment is manual verification, which depends solely on human examination to confirm the authenticity and integrity of the papers. This labour- and time-intensive method causes processing delays and inefficiencies since it takes a lot of work. Moreover, human error is a common occurrence, leading to inconsistent precision and dependability. Thus, the advancement of computer vision and machine learning technologies has been driven by the need for efficient, accurate, and dependable verification systems. These days, all organisations rely primarily on digital documents and digital processes for essential operations; therefore, it is crucial to quickly and accurately check the document.

This project's main objective is to demonstrate and develop a simple system that can reliably identify and detect, classify certificates according to their type, identify validation certificates, and reject papers that aren't necessary. When the machine detects uncertainty, those instances will be marked for manual verification to ensure that any ambiguities are thoroughly examined.

There are numerous significant sections in this work. First, the generation of the dataset, which ensures diversity in the representation and structural arrangement of various documents by obtaining from the customer a varied set of documents containing photos. Titled, nontitled, and form-based papers are among the several kinds of documents. This is an important stage since it offers an extensive dataset that captures the different forms, styles, and complexity found in real-world situations. Then, in order to provide ground truth data necessary for training and assessing the detection model, these documents are annotated using Label-Studio to annotate the signs and seals for detection and key value pair annotation for structural analysis of the text.

A key component of our methodology is the LayoutLMv3 model, a specialised model created to comprehend and handle complicated document layouts by combining text, layout, and visual information. It is very useful for document processing tasks including form interpretation, table extraction, and document categorization since it improves upon earlier versions of LayoutLM in handling multimodal data. It uses layout embeddings to tokenize the text and capture spatial arrangements, and it further provides context with visual features from a pre-trained CNN. The model learns the links between tokens by considering their placements and contents through its transformer encoder with self-attention processes. As a result, LayoutLMv3 can precisely recognise entities and associate keys with the appropriate values.

For the detection we use the Faster R-CNN model, a specialised model built to understand and handle complicated object identification tasks by integrating region proposal networks with convolutional neural networks, is a crucial part of our methodology. It handles region-based detection better than previous versions and is very efficient for tasks like object detection, instance segmentation, and image classification. It generates potential object bounding boxes using a region proposal network (RPN) and uses a CNN that has already been trained to extract information from these regions. With the use of its fully linked layers and region of interest (RoI) pooling, the model learns the relationships between regions by taking into account their contextual and spatial information. Consequently, objects may be accurately recognised using Faster R-CNN, which can also connect bounding boxes with the relevant class labels.

By utilising these technologies, the workload necessary for manual inspection is greatly decreased in addition to automating document classification and verification. In high-volume situations when human inspectors are prone to errors and tiredness, this automation is quite helpful. Organisations can significantly increase operational dependability and efficiency by simplifying these processes.The advantages of such a system go beyond simple operational efficiency. Automating certificate verification has far-reaching implications for security and compliance.

Ensuring certificate validity is essential to upholding trust and integrity in industries like business, government, and education. In addition to lowering the possibility of fraud and improving overall document security, automated solutions can produce results consistently. Organisations can strengthen their document reliability and prevent fraud by putting these technologies into practice.

Achieving high levels of accuracy and dependability in certificate verification is the main focus of the project. Using cutting-edge technology techniques improves accuracy and efficiency while also making the process simpler and quicker. By automating certificate detection and authentication, organisations can anticipate notable gains in productivity and operational dependability.

In the end, this project represents a comprehensive attempt to address the difficulties associated with manual document verification by utilising cutting-edge technologies in the most efficient domains and techniques. To that end, we have developed and validated methodologies for dataset generation, model execution, and model assessment through test files in order to attain the best possible outcomes. This commitment to project advancement is focused on automating document verification systems.

Summarizing, our results show how well our approach works to extract data from particular kinds of documents. We were able to precisely parse and extract useful data from a test set of documents using our algorithm, and we also achieved excellent results on the trained dataset when we used Faster R-CNN to detect seals and signatures. These results demonstrate how these methods can be successfully implemented to change the way organisations authenticate documents, guaranteeing greater accuracy, efficiency, and security. This introduction lays the foundation for a thorough exploration of the project's methodologies, findings, and implications, giving a thorough overview of our approach to enhancing document verification through automation.

Chapter 2

BACKGROUND

The literature analysis focuses on important advances in document understanding, notably Visual Document Understanding (VDU) and Relation Extraction (RE) within Natural Language Processing (NLP). It addresses how classic RE jobs have shifted from sentence-level to document-level difficulties, needing models that can handle both textual information and visual characteristics in documents. Specifically, the integration of models such as LayoutLM, LayoutLMv2, LayoutLMv3, and BROS tackles the issues given by visually rich documents (VRDs), which require spatial arrangement and picture information. These models use techniques like positional encoding and include both absolute and relative spatial data to efficiently extract connections between things. Recent research emphasises the effectiveness of models such as LayoutLMv3 in improving document verification duties via automated systems, which may be combined with technologies such as Faster R-CNN for full document analysis. These innovations not only increase operational efficiency by automating verification procedures, but they also hold promise for broader applications across sectors, highlighting the continued development and potential impact of multi-modal methods to document comprehension and processing.

The research offers the SC-Faster R-CNN algorithm, a revolutionary method for increasing object identification performance, particularly when objects are obscured or distorted. By incorporating skip pooling and contextual information fusion into the Faster R-CNN architecture, the algorithm hopes to improve identification accuracy, particularly for tiny or partially occluded objects. Experimental assessments of PASCAL VOC datasets show that the approach outperforms previous techniques. The SC-Faster R-CNN method performs well in general object identification tasks, but it struggles with distorted, spinning, and camouflaged objects. Future research topics include looking into ways to increase detection efficiency for difficult samples and optimising the algorithm for real-time system performance. The proposed approach demonstrates the potential for improving item identification skills in complicated circumstances, emphasising the significance of future research to solve constraints and improve overall detection accuracy in a variety of real-world applications.

LayoutLMv3: Pre-training for Document AI with Unified Text and Image Masking" presents LayoutLMv3, a transformational multimodal pre-trained model designed specifically for Document AI workloads. Unlike typical techniques that use separate CNNs for picture embeddings, LayoutLMv3 combines text, image, and multimodal representations into a single Transformer architecture. This innovation not only simplifies model parameters but also eliminates the requirement for region annotations, increasing efficiency and making model deployment easier. LayoutLMv3 achieves world-class performance across a wide range of tasks, including form understanding, receipt analysis, document image classification, and layout analysis, by leveraging a comprehensive pre-training framework that includes Masked Language Modelling (MLM), Masked Image Modelling (MIM), and Word-Patch Alignment (WPA). Its adaptability and resilience in both text-centric and image-centric tasks demonstrate its suitability as a flexible tool for Document AI applications. Furthermore, LayoutLMv3's simplicity and reliability make it an invaluable tool for improving document comprehension and analysis in practical contexts. LayoutLMv3, as a pioneering model in multimodal pre-training for document processing, establishes a strong foundation for future research and development in document AI capabilities and industry-specific difficulties.

The study presents Fast R-CNN, an efficient and enhanced variant of R-CNN and SPPnet for object identification. It delivers cutting-edge data and conducts rigorous trials to provide fresh insights into object identification systems. The work emphasises the importance of sparse object suggestions in improving detector quality, an aspect that was previously difficult to investigate owing to time restrictions but is now achievable using Fast R-CNN. The approach includes extracting features for proposals using max-pooling on feature maps, with several output sizes pooled and concatenated in a manner similar to spatial pyramid pooling. Fast R-CNN dramatically increases detection during test time, with speeds ranging from 10 to 100 times faster than earlier approaches. Training time is also decreased thrice as a result of quicker proposal feature extraction. The research continues by mentioning the possibility of additional developments in object identification approaches that might improve the performance of dense boxes to match sparse suggestions, so adding to the speeding of object detection procedures.

[You can have chapters that were published as part of your thesis. The text style of the body should be single column, as it was submitted to the publisher, not formatted as the publisher did.]

Chapter 3

MATERIALS AND METHODS

3.1 LayoutLMv3

Developed to comprehend and handle documents that contain text and layout information, including visual information, LayoutLMv3 is a cutting-edge model. By taking use of the spatial organisation of the text within a document, LayoutLMv3 aims to improve document comprehension. Along with linguistic understanding of documents, it also merges picture and text modalities to boost performance in tasks like reading forms, interpreting receipts, and classifying documents.

EMBEDDINGS

Text/Word Embeddings

Text in documents is converted into numerical representations in LayoutLMv3 by text embedding. The model's ability to comprehend and digest text effectively depends on this procedure. For complete document understanding, LayoutLMv3 combines visual information with text embeddings created using sophisticated natural language processing algorithms.

Visual Embedding/ Image Embedding

Processing the document's structure and layout is a part of visual embedding. LayoutLMv3 captures spatial connections and visual elements in the document using visual transformers. These embeddings aid in the model's comprehension of the context that the layout provides, including the locations of text blocks, graphics, and other document components.

Process:

- Linear Embedding: After splitting up images into patches, each patch is given a linear embedding.
- CNN: Convolutional neural networks (CNNs) are used to analyse pictures and produce feature maps that resemble grids.
- Faster R-CNN: Specific regions of interest within the image are identified and embedded.

TRANSORMERS

Multimodal Transformer The multimodal transformer, which incorporates both text and visual embeddings, is the central component of LayoutLMv3. This transformer creates a comprehensive knowledge of the document by analysing the merged data. LayoutLMv3 is able to handle complicated document analysis tasks more correctly because to the multimodal transformer, which takes into account both textual content and layout. This allows the model to exploit both textual and visual information for full document comprehension.

Pre-training Objectives

LayoutLMv3 improves its performance by using many pre-training objectives:

- Masked Word Token Classification: enhances the model's capacity for language understanding by teaching it to anticipate words that are hidden inside the text.
- Masked Patch Token Classification: E improves the model's capacity to anticipate picture regions that are veiled, aiding in the comprehension of the image's structure. of any length.
- Origin Image Reconstruction: makes sure the model keeps all of the visual information by reassembling the original picture.
- Masked Region Feature Regression: optimises the model's capacity to recognise and anticipate particular areas inside pictures.

PROCESSING

Document Processing LayoutLMv3 excels in processing a variety of document types by leveraging its integrated text and visual understanding capabilities. It is particularly effective in tasks such as form understanding, receipt processing, and document classification. The model's ability to analyze both content and layout allows it to handle complex documents with varied structures efficiently.

3.2 Faster R-CNN

LAYERS

Region Proposal Network (RPN) To create a convolutional feature map, the input picture is processed using a deep convolutional neural network (ConvNet), usually a pre-trained model like VGG16 or ResNet. With the ability to capture important high-level visual aspects like edges, textures, and forms, this feature map provides a high-dimensional representation of the input image. These characteristics are essential for the creation of region proposals as well as item categorization.

RoI Pooling Layer

Extraction of fixed-size feature vectors from the suggested regions produced by the RPN is the responsibility of the Region of Interest (RoI) pooling layer. The RoI pooling layer collects features from each mapped region once it has been suggested and mapped into the convolutional feature map. The size of these features is normalised via ROI pooling, which fixes their size so that the ensuing fully linked layers can process them reliably. Because it enables the network to efficiently accommodate region proposals of different sizes, this normalisation is crucial.

Fully Connected Layers

A sequence of fully linked layers are next applied to the pooled features from the ROI pooling layer. The traits that were taken out of the suggested areas are further processed and refined in these levels. The high-level characteristics are consolidated by the fully connected layers, setting them up for the latter phases of bounding box regression and classification.

OUTPUT HEADS

Two parallel branches are involved in the Faster R-CNN final stage:

- Softmax Layer: Over a predetermined set of object classes, this branch generates a probability distribution. The odds that each suggested region belongs to a certain class are output by the softmax layer. This allows the model to correctly categorise the items that are observed.

- Bounding Box Regressor:Modifications to the bounding box coordinates are output by the second branch. By making these changes, the initial area recommendations are improved and the bounding boxes are made to suit the observed items precisely. By adjusting the locations and dimensions of the observed objects, the bounding box regressor increases the localization accuracy.

3.3 Label Studio

An open-source data labelling application called Label Studio offers a flexible platform for annotating many different kinds of data, such as text, photos, audio, and video. When producing the datasets needed to train machine learning models, it is very helpful. Label Studio is essential for annotating documents with key-value pairs and object identification labels in the context of Faster R-CNN and other document-related tasks.

Key-Value Pair Annotation

Key-value pair annotation is made easier with Label Studio, which is crucial for organising and retrieving particular information from documents. Form processing, and other document comprehension applications frequently employ this kind of annotation. Label Studio aids in the creation of a well-annotated dataset that can be used to train models for tasks like information extraction and form interpretation by labelling sections of text as keys and related values.

Object Detection for Faster R-CNN Labeling

Label Studio offers powerful tools for annotation of pictures with bounding boxes for object detection tasks, which are utilised in the training of Faster R-CNN models. Bounding boxes surrounding interesting things can be drawn by users, who can then identify the boxes. In order for Faster R-CNN to effectively recognise and categorise objects inside pictures, this procedure is essential for producing high-quality training data.

Chapter 4

EXPERIMENTAL ANALYSIS

The development of an automated system for document comprehension and sign/seal identification requires a number of intricate procedures.

1. Data Collection and Preprocessing

   Collected a set of documents from the client and conducted an initial review. Filtered out outliers, removed unwanted and duplicate files to ensure a clean dataset.Sorted documents into different categories based on document types.

2. Custom Dataset Creation for LayoutLMv3

   Utilized Label-Studio to create a custom dataset for LayoutLMv3 model training. Annotated documents using Label-Studio's OCR template, capturing key-value pairs for model comprehension.Exported annotated documents in json.min format, totaling approximately 700 files for further processing.

3. Fine-tuning LayoutLMv3 Model

   Fine-tuned LayoutLMv3 using the custom dataset with 10,000 epochs. Studied dataset performance and selected the best model based which is obtained after fine tuning on manual inference verification.

4. Dataset Creation for Faster R-CNN:

   Focused on sign and seal detection, created a dataset by labeling relevant items using Label-Studio.

5. Fine-tuning Faster R-CNN Model

   Fine-tuned Faster R-CNN using the COCO format with 350 epochs and 480 labeled images. Achieved improved model performance suitable for the specific project use case.

The work flow of the project is as shown below:

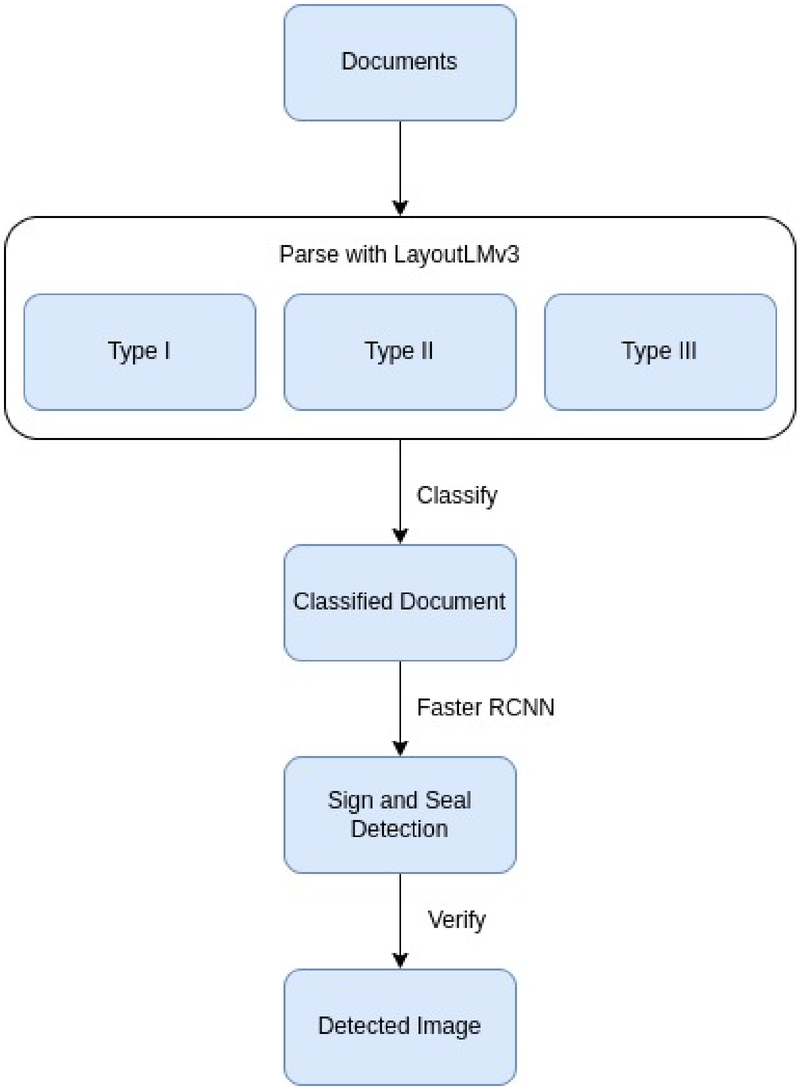

Figure 4.1: Work flow of the project

4.1 COMPARISON OVER OTHER METHODS

1. LayoutLM

   Document understanding problems were revolutionised by LayoutLM, which was first offered with a model architecture that included a BERT-based model with a Masked Region Proposal Network (MRPN). LayoutLM addressed the difficulties presented by documents with complicated structures, such as forms, bills, and receipts, by seamlessly merging text and layout information. Its capacity to effectively extract structured data from documents with a variety of layouts is its primary invention, which positions it as a fundamental tool in the fields of document processing and OCR (Optical Character Recognition).

2. LayoutLMv2

   With improved design and performance metrics, LayoutLMv2 expands upon the framework created by its predecessor. Precision, recall, and F1-score measures are all higher with the new model architecture than with LayoutLM, allowing it to manage even more complex document layouts with more efficiency. Additionally, LayoutLMv2 includes new training methodologies, namely in pre-training tactics and transfer learning. By addressing the scalability and performance limits seen in previous versions, these enhancements seek to increase the model's flexibility across a variety of document kinds and languages.

3. LayoutLMv3

   At the cutting edge of document understanding technology, LayoutLMv3 has cutting edge improvements in both its architecture and performance. With optimal inference speed, resource utilisation (GPU memory, for example), and overall inference time, this version excels in accuracy and efficiency. Furthermore, LayoutLMv3 may be precisely customised to certain domains or document formats, providing unmatched fine-tuning versatility. LayoutLMv3's enhanced customisation possibilities and smooth integration into pre-existing document processing workflows make it an excellent tool for practical applications.

Thus, I inferred that the transition from LayoutLM to LayoutLMv3 signifies a noteworthy progression in the field of text comprehension technology. Every iteration has improved upon the shortcomings and strengthened the areas of its predecessor. For applications requiring accurate data extraction and strong document layout analysis, LayoutLMv3 becomes the clear choice. It provides improved fine-tuning capabilities, improved model design, and greater performance metrics. I found out that LayoutLMv3 is the best model that can be used within these versions for my project.LayoutLMv3's developments establish it as a fundamental component of contemporary document processing systems, facilitating accurate and efficient document comprehension for a wide range of industries and use cases.

Chapter 5

RESULTS AND DISCUSSIONS

Dataset Preparation and Annotation

Preparing and annotating the dataset was the first significant step in this effort. From the customer, a wide range of documents, including form-based, titled, and untitled documents, were obtained. In order to get rid of any duplicates or unnecessary files, these papers were carefully sorted and cleaned. We labelled about 700 pages with Label-Studio, labelling signs and seals for detection and key-value pairs for structural analysis. The LayoutLMv3 and Faster R-CNN models were trained on a strong basis thanks to this annotated dataset.

LayoutLMv3 Fine-tuning

The custom dataset generated from the annotated documents was used to refine the LayoutLMv3 model. 10,000 epochs were used in the fine-tuning phase to help the model learn and comprehend the complex linkages and features found in the document layouts. When the optimised LayoutLMv3 model's performance was assessed using common metrics, the following outcomes were obtained:

Metric	**Value**
Eval Loss	0.077
Eval Precision	0.768
Eval Recall	0.763
Eval F1	0.765
Eval Accuracy	0.989

Table 5.1: Evaluation Metrics and Values for LayoutLMv3

These findings show that the LayoutLMv3 model achieved high precision, recall, and overall accuracy by successfully learning to identify and comprehend the texts' structure and content.

Faster R-CNN Fine-tuning

We employed the Faster R-CNN model for the detection of signs and seals. Between 380 and 400 papers' worth of signs and seals were labelled to provide a different dataset. The COCO format was used to export this dataset, which had JSON files and labelled photos. 480 labelled pictures and 350 epochs were used to refine the Faster R-CNN model. The table below provides an overview of the performance of the optimised Faster R-CNN model:

Metric	**SEALS**	**SIGNS**	**Overall**
Precision	0.87	0.82	0.845
Recall	0.83	0.79	0.81
F1 Score	0.85	0.80	0.825
mAP	0.81	0.78	0.795

Table 5.2: Performance Metrics of Faster R-CNN

These measurements show that the Faster R-CNN model is a dependable tool for document verification jobs since it can recognise and categorise signs and seals properly.

Manual Verification and Validation

After training both models, we manually verified the inferences to ensure their accuracy and reliability. The models demonstrated a high degree of precision in identifying key-value pairs, signs, and seals in the documents. This manual verification step was crucial in confirming the models' capability to perform in real-world scenarios.

Overall System Performance

An automated system for document verification has been created through the combination of LayoutLMv3 for document layout analysis and Faster R-CNN for sign and seal identification. This technology lowers the danger of human mistake, improves operational efficiency, and drastically decreases the labour needed for manual inspection. Now, businesses can rely on this system for reliable, accurate, and consistent document verification—a feature that is especially helpful in settings with large processing volumes. The project's outcomes demonstrate how well-suited sophisticated machine learning models like as LayoutLMv3 and Faster R-CNN are for automating activities related to document verification. Organisations may improve the security, efficiency, and accuracy of their document processing operations by

شركة لكنيت السعودية المحدودة
SAUDI TECHINT LTD.

Date: 31. Aug. 2009

TO WHOM IT MAY CONCERN

This is to certify that ***Mr. Renold Christopher Pereira,*** *Indian national holding passport # H-0465714 was employed in our organization as a* ***Project Technical Coordinator for CAD & TDC,*** *from 01st Aug 2006 to 31st Aug 2009 for Khursaniyah Pipeline Project.*

Mr. Pereira *was involved in preparing Piping, Civil construction, Structural & Electrical IFC construction & As-Built drawings, Furthermore he was responsible for Technical office that includes Engineering / Vendor Technical documentation based on the contract of construction of Upstream & Downstream Pipelines for Khursaniyah Project.*

During his stay with us, his performance was excellent with very good leadership qualities and management skills. His approach towards the job and conduct is excellent.

This Letter of Appreciation is issued for his efforts towards the completion of Project.

Saudi Techint Ltd. By

SAUDI TECHINT LTD.
C.R. 2051015148

Manuel Aguilar
Construction Manager

Issued this on 31th August 2009 at Techint – Dhahran Main Office, Kingdom of Saudi Arabia. By Saudi Techint Ltd.

Figure 5.1: Obtained output

Figure 5.2: Detected header using LayoutLMv3

Figure 5.3: Detected Seal using Faster R-CNN

Figure 5.4: Detected Sign using Faster R-CNN

utilising these technologies. These models have been successfully implemented and refined, indicating their potential to revolutionise manual verification procedures and open the door to more dependable and automated solutions across a range of sectors.

Chapter 6

LIMITATIONS

- In contrast to its name, faster R-CNN is not the quickest object identification model. In comparison to single-stage detectors, its two-stage detection process may be slower.
- The Faster R-CNN is a complex model that can be challenging to adjust, optimise, and put into practice. It consists of several components, including the Fast R-CNN detector and the Region Proposal Network (RPN), each of which needs to be fine-tuned for better performance.
- When using Faster R-CNN for the detection of seals on certificates, some emblems are mispredicted as seals, indicating difficulty in distinguishing between visually similar elements.
- LayoutLMv3 does a great job understanding structured documents, but it could have trouble with papers with very different or unconventional layouts. Documents with complex formatting, unusual font choices, and a variety of graphical features that differ from the training set are examples.

Chapter 7

CONCLUSION AND FUTURE WORK

7.1 Conclusion

This project successfully demonstrated the development and implementation of an automated document verification system capable of reliably identifying, detecting, and categorizing certificates based on their kind, validating certificates, and rejecting unneeded documents. The system employs a variety of cutting-edge technologies, including LayoutLMv3 for document layout analysis and Faster R-CNN for object detection. The results reveal that the system is capable of extracting data from a wide range of documents with great accuracy and efficiency.

The project's success has far-reaching consequences for businesses that rely significantly on manual document verification processes, which are frequently labor-intensive, time-consuming, and prone to human mistake. By automating this process, enterprises can reduce the burden necessary for manual screening, increase operational efficiency, and improve document security. The system's capacity to effectively detect and classify documents can help to reduce errors and improve document processing quality.

Furthermore, the system's versatility and scalability make it a viable option for a variety of industries, including finance, healthcare, education, and government. The system's flexibility to adapt to various document formats and layouts makes it an invaluable tool for organizations that handle a diverse range of papers. In conclusion, this project has shown how automated document verification systems may change document processing. We can improve the efficiency, accuracy, and security of document verification by incorporating cutting-edge technologies and advanced machine learning algorithms.

7.2 Future Work:

- Dataset Expansion

  Expanding the dataset is essential for improving model performance and generalizability. This can be accomplished by gathering a diversified set of papers from various sectors, domains, and format types. The extended dataset should comprise texts with diverse layouts, fonts, and graphical features, as well as documents of varying complexity and structure. Furthermore, the dataset should be labeled and annotated with high-quality information to guarantee that the model can effectively learn from it. Data augmentation techniques can be used to expand dataset size while reducing overfitting. Furthermore, active learning and transfer learning can be employed to lower annotation expenses while increasing annotation quality.

- Improved model architectures:

  Improved model architectures can also help to improve performance. Single-stage detectors like YOLO or SSD can be investigated as alternatives to Faster R-CNN, which may be faster and more accurate. Transfer learning can be used to extract features from pre-trained models or to fine-tune them for specific tasks. Hybrid models, which combine the strengths of multiple models, can also be investigated. Ensemble approaches, also known as stacking, can be used to increase accuracy by combining numerous models' predictions. Furthermore, domain adaptation methods can be utilized to modify the model to fit new domains or layouts.

- LayoutLMv3 Enhancements

  LayoutLMv3 is an effective model for comprehending structured texts; yet, it may struggle with uneven layouts or papers with sophisticated formatting. To address this issue, strategies such as layout-aware attention mechanisms can be created to focus on specific sections of the document based on layout data. Self-attention processes can be utilized to represent the links between different parts of the material. Furthermore, domain adaptation techniques can be utilized to adapt the model to new domains or layouts. Layout-aware post-processing algorithms can also be created to improve predictions using layout information.

- Error Correction Mechanisms

  Error correction strategies are crucial for reducing errors and increasing model accuracy. Active learning can be used to choose uncertain samples for human annotation while also incorporating human feedback into the training process. Manual review or re-training are examples of post-processing approaches that can be used to fix faults. Uncertainty estimation approaches can be used to estimate the uncertainty in model predictions and incorporate it into decision-making. Ensemble approaches, such as stacking, can be used to merge many models to enhance accuracy. By implementing these error correction strategies, the model will become more robust and accurate.

By addressing these areas, we can develop a more robust and accurate automated document verification system capable of effectively identifying, detecting, and classifying certificates with various layouts and structures.

Chapter 8

CONSENT FORM