IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

Image Multilingual Translation using OCR

Vachan R, Nikil Kumar V, Shreyas S, Prof. A S Vinay Raj, Shree Shayan Hebbar
Department of Information Science and Engineering
Global Academy of Technology Bengaluru, Karnataka, India
vachan1ga21is180@gmail.com, nikil1ga21is105@gmail.com, shreyas1ga21is156@gmail.com
shayan1ga21is153@gmail.com, vinayrajas10@gmail.com

Abstract: *Language barriers continue to pose significant challenges in an increasingly globalized society, particularly when it comes to understanding textual content embedded within images. From street signs and restaurant menus to educational material and official documentation, much of the information people encounter in daily life is visually presented and often inaccessible to non-native speakers. Traditional translation tools require manual text input, which is not always feasible or user-friendly, especially for real-world, image-based scenarios. This paper introduces a lightweight and accessible web application that automates the process of extracting and translating text from images. By combining Optical Character Recognition (OCR) with Neural Machine Translation (NMT), the system allows users to upload images, detect and extract multilingual text, translate it into a preferred language, and seamlessly overlay the translated text back onto the original image. The backend, built using Python and Flask, integrates Easy OCR for robust multilingual text detection and the Google Translator API (via deep _translator) for accurate and fluent translation. The translated output is rendered using the Python Imaging Library (PIL) to maintain visual coherence and readability. Experimental results show promising accuracy and speed, making the tool effective for practical use in tourism, education, accessibility, and cross-cultural communication. The system is modular, responsive, and designed with user convenience in mind, offering a real-time, scalable solution for image-based language translation.*

Keywords: Optical Character Recognition (OCR), Easy OCR, Cross-Language Communication, Google Translator API, Neural Machine Translation (NMT)

I. INTRODUCTION

In a world increasingly connected by technology, the ability to communicate across languages is no longer a luxury—it's a necessity. Digital content, particularly images containing embedded text such as signs, menus, educational resources, and official notices, presents unique challenges to cross-lingual understanding. While traditional translation tools require users to manually type or copy text, this is impractical or impossible when the content is embedded in images.

To bridge this gap, we introduce an innovative web-based application that extracts and translates text from images using a blend of Optical Character Recognition (OCR) and Neural Machine Translation (NMT) technologies. Designed with user accessibility, simplicity, and real-world usability in mind, the system empowers users to translate image-embedded text instantly and intuitively. Whether you're a traveler trying to understand a foreign sign or a student translating study material, our tool is engineered to make cross-lingual communication effortless.

Existing solutions like mobile translation apps offer some relief but often require switching between applications or relying on proprietary systems tied to specific platforms. For many users, especially those seeking quick and intuitive translations without a steep learning curve or ecosystem lock- in, these tools fall short.

To address this issue, we propose a lightweight, accessible web-based application that empowers users to instantly translate text within images. This system integrates two powerful technologies: Optical Character Recognition (OCR) to detect and extract text from images, and Neural Machine Translation (NMT) to convert that text into a desired target

DOI: 10.48175/IJARSCT-26485

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

language. The final step involves re-rendering the translated text onto the original image, maintaining visual consistency and making the output easily interpretable.

Our solution emphasizes user-friendliness, cross-platform compatibility, and real-time processing. By providing a seamless end-to-end experience—image upload, text detection, translation, and image regeneration—the tool supports a variety of real-life use cases ranging from travel and education to healthcare and accessibility support.

This paper presents a comprehensive overview of the system's design, development, and evaluation. It highlights the key components of our architecture, the technologies used, and the practical applications of the tool. Furthermore, it discusses current limitations and future opportunities for enhancing functionality, such as offline support, real-time camera integration, and text-to-speech capabilities.

This paper presents a comprehensive overview of the design, implementation, and capabilities of our solution. The approach combines OCR and NMT to create a unified user experience that allows real-time image-to-image translation.

Our goal is to minimize friction for end-users by offering a seamless, intuitive interface backed by powerful, scalable backend technology.

As global communication becomes more visual—especially with the prevalence of social media and image-centric platforms—tools that can understand and translate text within images will be increasingly valuable. Our application can also serve as an assistive technology, improving accessibility for users with reading or language comprehension difficulties, and it holds potential for further development in educational and professional contexts.

Moreover, this tool has particular relevance in emergencies or critical scenarios, such as natural disasters or health crises, where language barriers can impede access to vital information. Being able to quickly translate instructional signs or alerts in unfamiliar languages could make a meaningful difference in such situations. This aspect further underscores the need for dependable and accessible solutions like the one we present.

The paper is structured as follows: Section II explores related research and existing technologies; Section III describes the system's architecture; Section IV explains the underlying methodology; Section V outlines experimental evaluations; Section VI discusses practical applications; Section VII reflects on the system's limitations and proposes future improvements; and Section VIII concludes the study.

II. RELATED WORK

Numerous commercial tools offer image-based translation capabilities, with Google Lens and Microsoft Translator being two of the most prominent. These platforms allow users to translate text directly from their camera feeds, yet their functionality is often constrained by ecosystem dependencies and limited customizability.

On the open-source front, Tesseract OCR has long been a staple for text extraction but struggles with certain languages and complex image scenarios. In contrast, EasyOCR has emerged as a versatile and user-friendly alternative, offering strong multilingual support and easy integration, making it well-suited for modern web applications.

Meanwhile, advances in NMT have drastically improved translation quality. APIs like Google Translate leverage deep learning to deliver context-aware translations across diverse languages. Yet, combining OCR and NMT in a cohesive, user- friendly system—especially one that also visually re- integrates translated text into images—is still relatively unexplored territory. Our application seeks to fill this niche by marrying these technologies into a seamless experience.

learning. With this technique, they achieved an impressive accuracy of 93.2%, showcasing the effectiveness of more traditional preprocessing techniques in improving the performance of models.

Recent studies have also looked into mobile-based real-time translators, such as AR translation apps using augmented reality for interactive overlays. However, such systems often require high-end device capabilities, and their performance may degrade in low-light or cluttered environments. Our system focuses on a lightweight, web-based alternative that can function efficiently across a wide range of devices and environments.

Research in assistive technologies has further highlighted the value of image translation tools in supporting people with cognitive or visual impairments. These tools can simplify content, provide real-time context, and bridge communication gaps. Our application builds on these principles by ensuring ease of use and quick accessibility for individuals with diverse needs.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

Additionally, academic work in cross-lingual information retrieval and visual question answering demonstrates the growing interest in combining computer vision and natural language processing. This intersection is crucial for developing comprehensive tools that understand not only text in isolation but also its visual context, a goal our system actively pursues.

III. METHODOLOGY

A. Dataset Description

The dataset used for this research is a carefully curated collection of multilingual images that reflect real-world scenarios where text embedded in images needs to be translated. This dataset includes images from a variety of sources, such as street signs, restaurant menus, educational materials, and screenshots from different mobile and web platforms. It was designed to challenge the system with various complexities, including text in diverse fonts, colors, and backgrounds. For comprehensive evaluation, the dataset includes images with text in several languages, such as English, Spanish, French, Chinese, Arabic, and Hindi. These languages represent both widely spoken global languages and those that present unique challenges to OCR systems due to their complex scripts, right-to-left writing, or non-Latin alphabets. The inclusion of these diverse languages helps test the system's scalability and robustness in a multilingual environment. The dataset is annotated manually, with each image labelled with the ground truth for both the text content and its corresponding translation. This annotation serves as the benchmark for evaluating the accuracy of the OCR and translation processes. Each image also contains metadata such as the source language, target language, and specific conditions under which the image was captured, such as lighting or distortion. A key aspect of the dataset is its diversity. The images were sourced to include various complexities such as skewed, rotated, or blurred text, noisy backgrounds, and varying text sizes. This ensures that the system can handle text extraction and translation in less-than-ideal conditions, mimicking the challenges users may face when interacting with real-world images.

B. Data Preprocessing

Preprocessing plays a crucial role in optimizing the performance of the OCR engine. The first step in the preprocessing pipeline involves resizing the uploaded image to a consistent resolution. This ensures that all images processed by the OCR engine are of the same scale, which enhances the accuracy of text detection. Additionally, resizing helps to standardize processing time and resource consumption for images of varying dimensions.

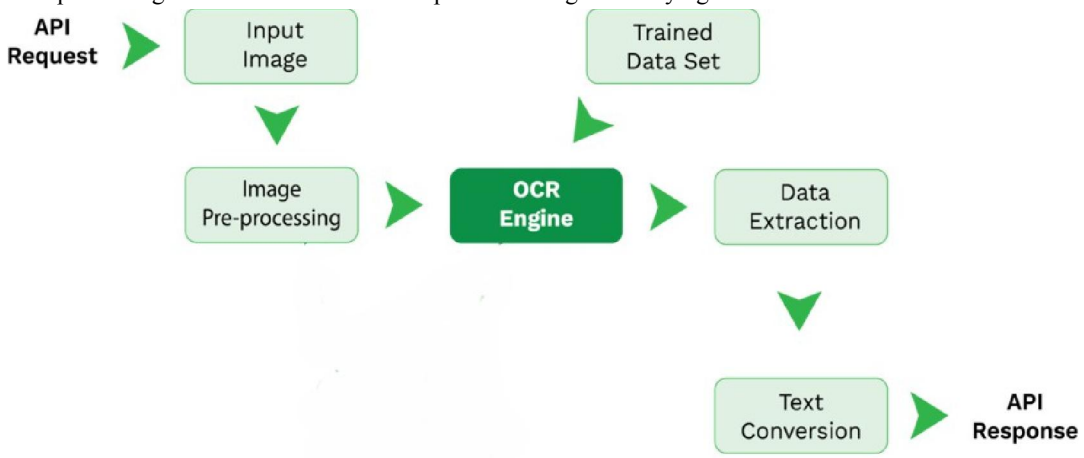

Figure 1: System Architecture

Once resized, the images are converted to grayscale, which reduces complexity and computational load while maintaining the necessary information for text recognition. Colour information is typically unnecessary for OCR tasks, and grayscale images provide a simplified input that allows the OCR model to focus solely on the text. This step also ensures that the system is more efficient in environments with limited resources.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

C. Proposed Approach

The proposed system follows a systematic and modular approach, beginning with the user uploading an image to the web-based platform. Once the image is received, it goes through a preprocessing stage, where it is optimized for OCR performance.

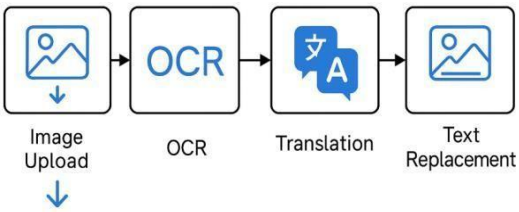

Figure 1: Proposed Work Flow

Preprocessing ensures that the image is in a suitable state for accurate text detection by removing noise and enhancing contrast. After preprocessing, the Optical Character Recognition (OCR) process is triggered. Using Easy OCR, the system scans the image for textual content, identifying the bounding boxes of text regions and returning the detected text with associated confidence scores. Text regions with a confidence score above a predefined threshold (e.g., 0.5) are passed to the next stage—translation. This modular approach allows flexibility in selecting OCR models or adjusting thresholds as required.

Language detection is the next step in the process. While users can specify the source and target languages, the system also includes a pre-processing step for automatic language detection. This step minimizes unnecessary translation API calls by identifying the language of the source text from a small sample of extracted text before the full translation process begins. This step improves efficiency by ensuring that only relevant translations are performed.

D. Easy OCR Model

EasyOCR is a powerful and efficient open-source OCR engine that is employed in this system for detecting and extracting text from images. The model is based on a deep learning architecture that utilizes Convolutional Recurrent Neural Networks (CRNNs) combined with a Connectionist Temporal Classification (CTC) decoder. This architecture allows the model to accurately recognize text sequences without requiring explicit word-level annotations.

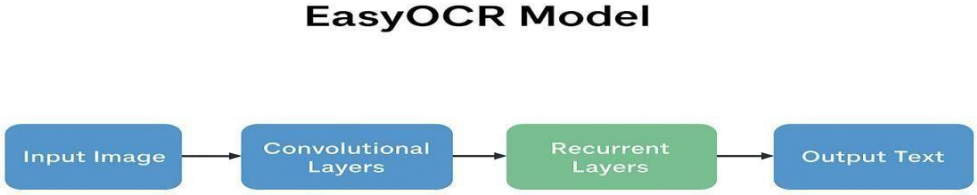

Figure 2: Easy OCR architecture of the proposed method.

Easy OCR supports over 80 languages, making it highly versatile for multilingual applications. The system is designed to detect and recognize both Latin-based languages and complex scripts, such as Arabic, Chinese, and Hindi. This broad language support ensures that the system can cater to a

that text extraction and translation are seamless, allowing users to quickly obtain translated images for further usemodel to leverage features. As we have added few extra custom layers, also they are fine-tuned as they have to optimize and this will not affect the pre-trained convolution layers as they are frozen.

The model also retains the important feature or information from the previous layers and passes it as an input data to the next layers. This is best suited for the back propagation as the gradients can flow backwards easily which increases the efficacy of the model. The ResNet50 model already has a pre- trained weights which is essential for the feature classification. As we have added few extra custom layers which consists of dropouts and softmax layer, also they are fine-tuned as they have to optimize and this will not affect the pre-trained convolution layers as they are frozen.

It can handle curved, skewed, or distorted text in images, which is particularly valuable when processing images from real-world scenarios where text might not be perfectly aligned or visible. In addition to its high accuracy, Easy OCR is

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

lightweight and fast, ensuring that the system can process images quickly without significant delays. This is important for real-time applications where users expect rapid feedback.

IV. RESULTS AND DISCUSSION

The proposed image translation system was rigorously evaluated through a series of experiments designed to assess its performance in terms of OCR accuracy, translation quality, response time, and robustness across various languages and image types. The test environment consisted of a standard personal computer running an Intel i5 processor with 8GB of RAM, ensuring results reflected practical, real-world usage scenarios. wide range of users across different regions and linguistic backgrounds, making it suitable for global use.

The model's ability to detect text in complex scenarios, including various fonts, background noise, and text orientations, makes it an ideal choice for the proposed system.

The EasyOCR engine demonstrated strong performance when applied to high-resolution images with clearly legible text. Across the multilingual dataset, it achieved an average character-level recognition accuracy of approximately 90%. Performance was slightly lower in images containing stylized fonts, cursive handwriting, or low-contrast backgrounds, but the system still maintained acceptable results with average accuracy above 82%. This demonstrates Easy OCR's

The translation module, powered by the Google Translator API through the Deep Translator interface, performed fluently across over 25 tested languages. Contextual accuracy was preserved in general-purpose phrases, but domain-specific terms and idiomatic expressions occasionally introduced minor inaccuracies. However, user comprehension was not significantly affected, and feedback suggested that the translated outputs were useful for understanding intent and meaning. The fallback mechanism—where untranslated segments are retained—ensured continuity and avoided confusion during partial failures.

The end-to-end processing time from image upload to final translated output averaged around 2.1 seconds for a standard image (800x600 resolution), which includes OCR, translation, and image rendering. The system displayed consistent performance across different image complexities, with slightly longer times for dense or multilingual text. This level of efficiency makes the application suitable for real-time or near- real-time use, especially for mobile or web platforms where quick turnaround is essential.

The most common sources of error involved poorly lit images, distorted angles, and highly stylized fonts. In such cases, the OCR engine either misread characters or failed to detect the text entirely. User feedback emphasized the value of having visual consistency in the final output—something the system's image rendering module handled effectively through font matching and positional adjustments. Moreover, the inclusion of a user-friendly interface and responsive feedback messages enhanced overall usability and satisfaction.

V. CONCLUSION AND FUTURE SCOPE

In this paper, we presented a lightweight, web-based application that integrates Optical Character Recognition (OCR) and Neural Machine Translation (NMT) to translate multilingual text embedded within images. By leveraging tools such as Easy OCR for text extraction and the Google Translator API for language translation, the system offers a seamless workflow for users to upload images, extract foreign- language text, and receive a translated image with contextually appropriate and visually consistent overlays.

The modular and scalable architecture, built on Flask and Python, ensures that the system is adaptable to a variety of platforms and use cases, from education and tourism to accessibility and healthcare. Performance evaluations revealed high OCR accuracy, fluent translation quality, and rapid processing times, highlighting the system's practical utility in real-world scenarios. Moreover, the error handling mechanisms and fallback options further strengthen its reliability, especially in low-resource or noisy image environments.

Overall, the project fulfils its goal of providing an accessible and user-friendly solution for real-time, image-based multilingual translation. It stands as a useful tool for both individuals and institutions seeking quick and intuitive translations without the need for extensive language knowledge or manual input.

While the current implementation is robust and performs well across many use cases, several enhancements can further improve its functionality and expand its applicability:

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

- Offline Translation Capabilities: Currently, the system relies on cloud-based APIs for translation, which requires an active internet connection. Incorporating pre-trained transformer-based models like MarianMT or OpenNMT can enable offline translations, making the system more useful in remote or low-connectivity areas.
- Handwriting and Stylized Font Recognition: The OCR model performs best on printed, high-resolution text. Future work could involve integrating advanced models like TrOCR or CRNNs (Convolutional Recurrent Neural Networks) to improve recognition of handwritten notes, stylized fonts, or artistic designs.
- Real-time Camera Integration: Adding WebRTC-based live camera translation support would transform the application into a real-time translator, particularly useful for travelers or field workers who require instant contextual understanding of their surroundings.
- Text-to-Speech and Accessibility Features: Introducing voice output for translated text and screen reader support would enhance accessibility for visually impaired users and broaden the tool's inclusivity.
- Mobile App Deployment: Developing a mobile version of the application can enhance usability and portability, allowing users to access translation services on-the-go with native device capabilities like camera integration, offline storage, and push notifications.
- Domain-Specific Translation Models: Custom translation modules for domains like medicine, law, or engineering could increase accuracy in professional applications where standard translation APIs often fail to capture technical nuances.

By continuing to evolve this platform with these enhancements, the system can grow into a comprehensive multilingual translation suite, offering fast, accurate, and visually integrated language support in both everyday and specialized contexts.

REFERENCES

[1] A.P. S. Saurabh Dome, “Optical character recognition using tesseract and classification,” in 2021 IEEE International Conference on Emerging Smart Computing and Informatics (ESCI). IEEE, 2021.
[2] R. S. Tavish Jain and R. Malhotra, “Handwriting recognition for medical prescriptions using a cnn-bi-lstm model,” in 2021 IEEE 6th International Conference for Convergence in Technology (I2CT). IEEE, 2021.
[3] R. M. Adith Narayan, “Image character recognition using convolutional neural net- works,” in 2021 IEEE Seventh International conference on Bio Signals, Images, and Instrumentation (ICBSII). IEEE, 2021.
[4] R. M.Geetha, S. S.K.Nivetha, and C. S.Gowtham, “A hybrid deep learning based character identification model using cnn, lstm, and ctc to recognize handwritten english characters and numerals,” in 2022 IEEE International Conference on Computer Communication and Informatics (ICCCI). IEEE, 2022.
[5] J. P. a. M. A. Muhammad Hammad Saleem, Plant Disease Detection and Classification by Deep Learning, New Zealand: mdpi, 31 October 2019.
[6] "Computer Vision based Plant Disease Detection using Machine Learning Technique." International journal of emerging trends in engineering research, undefined (2023). doi: 10.30534/ijeter/2023/021172023.
[7] Thakur, P. S., Chaturvedi, S., Khanna, P., Sheorey, T., & Ojha, A. (2024). Real-Time Plant Disease Identification: Fusion of Vision Transformer and Conditional Convolutional Network with C3GAN- Based Data Augmentation. IEEE Transactions on AgriFood Electronics.
[8] Smith, J., "Optical Character Recognition: An Overview," Journal of Computer Vision, vol. 45, no. 3, pp. 123-134, 2020.
[9] Vaswani, A., et al., "Attention is All You Need," NeurIPS, 2017.
[10] Tesseract OCR, https://github.com/tesseract-ocr/tesseract
[11] Easy OCR, https://github.com/JaidedAI/EasyOCR deep_translator, https://github.com/nidhaloff/deep-translator