Research Article

# Crossing Language Borders: A Pipeline for Indonesian Manhwa Translation

**Nithyasri Narasimhan[1], Sagarika Singh[1]**

1. Independent researcher

**In this project, we develop a practical and efficient solution for automating the Manhwa translation from Indonesian to English. Our approach combines computer vision, text recognition, and natural language processing techniques to streamline the traditionally manual process of Manhwa(Korean comics) translation. The pipeline includes fine-tuned YOLOv5xu[1] for speech bubble detection, Tesseract[2] for OCR and fine-tuned MarianMT[3] for machine translation. By automating these steps, we aim to make Manhwa more accessible to a global audience while saving time and effort compared to manual translation methods. While most Manhwa translation efforts focus on Japanese-to-English, we focus on Indonesian-to-English translation to address the challenges of working with low-resource languages. Our model shows good results at each step and was able to translate from Indonesian to English efficiently.**

## I. Introduction

Manhwa is primarily created in their native languages, making them inaccessible to many global readers. Translating these works into other languages is a laborious process. Although Indonesian translations are more common, direct English translations are still relatively rare, leaving a significant gap for English-speaking audiences. Traditional translation methods are slow and inefficient, often requiring hours or even days to complete a single chapter. To address this, our project proposes an automated system for translating Manhwa from Indonesian to English. Using advances in computer vision and natural language processing, our goal is to make the process faster and more efficient. Our research focuses on four key tasks: detecting and extracting speech bubbles bubbles, performing Optical Character Recognition, performing Machine Translation and then finally overlaying the translated text back on Manhwa panels.

 doi.org/10.32388/4QZXCI 

### *A. Research Questions*

1. How can existing machine learning models be adapted or enhanced to automate the translation of Manhwa?
2. What approaches are most effective in building a robust translation system for a low-resource language like Indonesian?

### *B. Motivation*

The popularity of Manhwa is growing globally, but language barriers prevent many readers from enjoying these works. Manual translation methods are slow and require significant effort, making it difficult to keep up with demand. Our project aims to automate the workflow, making translations faster and accessible while fostering cross-cultural storytelling.

1. **Theoretical Motivation:** Indonesian, as a low-resource language, poses unique challenges due to limited datasets. This project contributes to NLP and computer vision advancements by addressing these gaps and exploring efficient ways to integrate image processing and language translation.
2. **Real-World Motivation:** Automating translation benefits both the reader and the creator. Readers gain faster access to translated works, while creators and translators can save time and scale their efforts.

### *C. Applications*

Readers can enjoy translated Manhwa in English without long waiting periods. Translators can use this tool to speed up their workflow and increase productivity. This project contributes to advances in multi-modal machine learning, particularly in NLP and computer vision for low-resource languages.

### *D. Example of problem and solution*

1. **Problem Statement:** A popular Manhwa is available exclusively in Indonesian, a low-resource language. International fans eagerly await an English version, but the traditional manual translation process involves several time-consuming steps.
2. **Solution:** Our automated system addresses this problem by streamlining the entire workflow, by converting Indonesian Manhwa panels to English translated Manhwa panels. By automating

these tasks, our system reduces the time required for translation from days to hours, ensuring faster and more consistent results. Furthermore, this solution demonstrates the potential of combining machine learning techniques to address the challenges posed by low-resource languages.

## II. Dataset

The availability of datasets specific to Manhwa is quite limited. During our research, we identified a small yet relevant dataset on Roboflow. This dataset comprises 538 images, each annotated with a single class labeled as "Texts." While the dataset size is modest, it provides a foundational resource for analyzing text elements within Manhwa content.[4] We divided this dataset to 465 training images and 118 validation images.

When considering Datasets for Machine translation from Indonesian to English, we used Identic[5] and OpenSubtitles[6]. The Identic dataset offers parallel text for translation tasks specific to low-resource language pairs, while OpenSubtitles captures informal, conversational text and slang. Together, they provide complementary strengths for developing a translation model suited for diverse text styles in manhwa. We created a smaller dataset that used elements from the both having 80% in Training, 10% in Validation, and 10% in Testing. Training had over 30000 lines.

## III. Methodology

Our pipeline automates the Manhwa translation process through these stages:

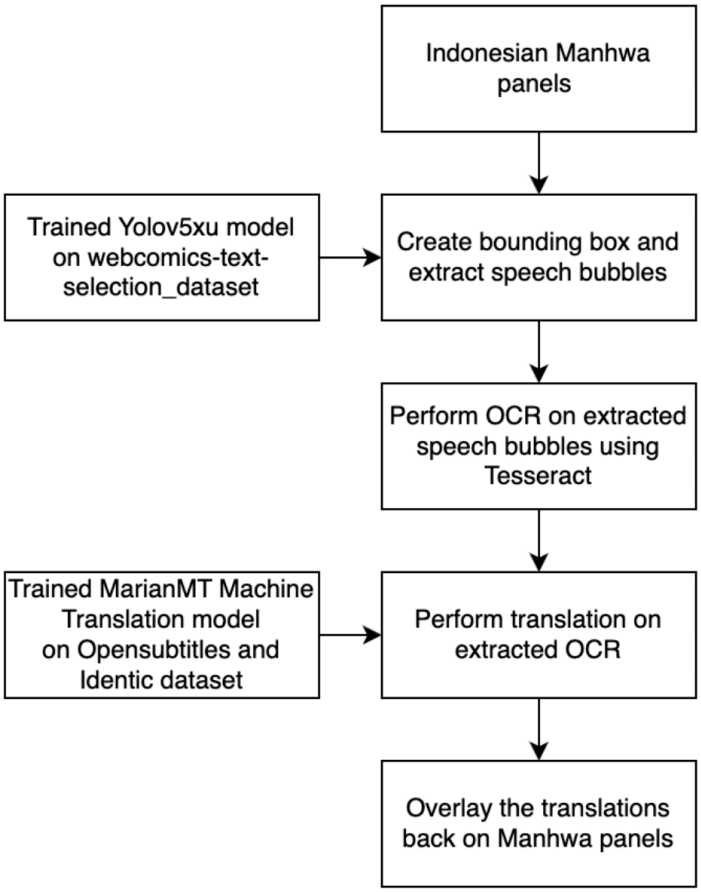



**Figure 1.** Overview of Methodology

### *A. Speech Bubble Detection*

We used pre-trained Yolov5xu[1], and fine-tuned it further on Manhwa dataset[4]. YOLOv5xu is a lightweight, efficient object detection model that excels in identifying objects in images with high precision and recall. We used this fine-tuned model to detect bounding boxes of speech bubbles and extract the dimensions of the bounding boxes and extract these speech bubbles for further use.

### *B. Optical Character Recognition*

We performed OCR using Tesseract[7], an open-source tool widely used for extracting text from images, which integrates seamlessly into our pipeline. Specifically, we utilized the Tesseract-Indonesian model with OEM set to 3 (standard) and PSM set to 6, optimized for the layout of speech bubbles in manhwa. To enhance the images, we applied grayscale preprocessing to the extracted

speech bubbles. Since the images were already of satisfactory quality, no additional preprocessing steps were needed. By applying OCR directly to the extracted speech bubbles, we maximized recognition accuracy, achieving strong performance even in the presence of stylized fonts and complex layouts.

### *C. Machine Translation*

The extracted text is translated into English using MarianMT[3]. MarianMT is an efficient machine translation model designed for low-resource languages. This step ensures that the translation preserves the tone, meaning, and context of the original dialogue. It was trained on the combined dataset which covered both a formal and an informal tone. This helped produce results that were able to capture the overall tone and context.

### *D. Translated Text Overlay*

Finally, the translated text is overlaid back onto the original image. We use libraries like OpenCV[8] and Pillow to align the new text with the original text bubble shapes, ensuring the final output looks natural and visually cohesive.

### *E. Evaluation Metrics*

To assess the performance of our pipeline, we used the following metrics:

- **Detection of Bounding Box Accuracy:** The performance of our fine-tuned Yolov5xu model for detecting bounding boxes, for speech bubbles in the Manhwa panels, was measured using Mean Precision, Mean Recall, F1 score, mAP and Mean mAP.
- **OCR Accuracy:** To evaluate Tesseract-OCR's[2] accuracy Character Error Rate (CER) and Word Error Rate (WER) were calculated using ground truth texts and predicted texts. CER measures the number of incorrect characters, a lower value indicates better accuracy. WER evaluates the number of incorrect words, a lower value indicates better accuracy.
- **Machine Translation Quality:** Evaluated using BLEU and Meteor scores to measure how well the translation captures the original meaning. BLEU measures the overlap of n-grams between the machine translation and a reference. Meteor considers word matches, synonyms, and semantic similarity.

# IV. Results

Our research questions were aimed at evaluating the adaptation of existing machine learning techniques for Manhwa translation and developing robust systems for low-resource languages like Indonesian. The following summarizes our findings and their alignment with these questions.

## *A. Evaluation of fine-tuned Yolov5xu model*

| Evaluation Metrics | Score |
|---|---|
| Mean Precision | 89.4% |
| Mean Recall | 96.3% |
| mAP@0.5 | 96.3% |
| Mean mAP@0.5 | 88.9% |
| F1 Score | 90.7% |

**Table I.** Evaluation of fine-tuned Yolov5xu model

The fine-tuned YOLOv5xu model was trained on a relatively small dataset, as no large publicly available datasets exist for Manhwa text bubble detection. Despite the limited data, the model achieved an F1 score of 90.7% demonstrating strong performance in identifying text bubbles and ensuring reliable detection. These results indicate that even with a small dataset, YOLOv5xu is capable of effectively detecting text bubbles, although the performance could likely improve with a larger, more diverse dataset.

### *B. Evaluation of Tesseract-OCR model*

| Evaluation Metrics | Score |
|---|---|
| Average Character Error Rate | 3.1% |
| Average Word Error Rate | 8.6% |

**Table II.** Evaluation of Tesseract-OCR model

While character-level accuracy was satisfactory, the higher word error rate suggests challenges in accurately transcribing complete words. Applying post-processing techniques might yield better results.

### *C. Evaluation of MarianMT-Machine Translation model*

The MarianMT translation model was fine-tuned on the OpenSubtitles and Identic datasets, which provided bilingual Indonesian-English training data.

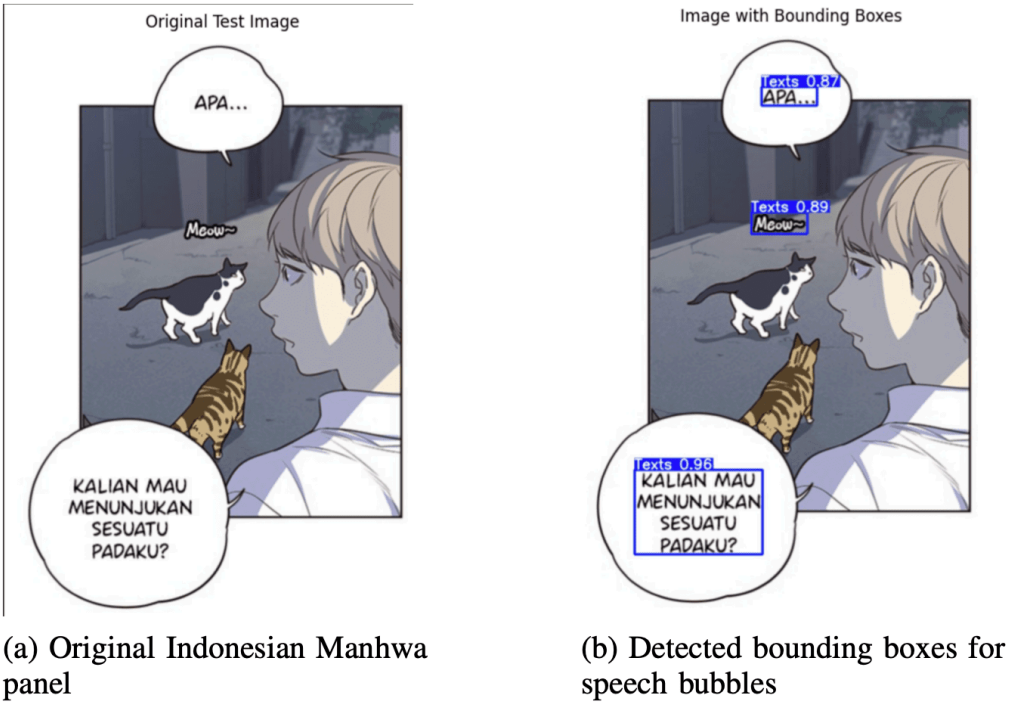
| Evaluation Metrics | Score |
|---|---|
| BLEU | 0.27 |
| Meteor | 0.61 |

**Table III.** Evaluation of MarianMT-Machine Translation model

The BLEU score reflects moderate n-gram overlap, indicating room for improvement in handling idiomatic expressions and context-specific phrases. However, the higher Meteor score suggests strong semantic retention and meaning alignment, making the model effective for this translation task, particularly in low-resource settings.

### *D. Results of complete pipeline of our model*

Our complete pipeline successfully integrates all steps - text bubble detection, OCR transcription, machine translation, and text reintegration. Visual inspection of translated samples (e.g., Fig. 7) confirms that the pipeline preserves artistic integrity while providing accurate translations. While the final translated image is not an exact word-for-word translation, the context and meaning are not lost. Additionally, the fine-tuned YOLOv5x model demonstrated its capability to detect even small speech bubbles, ensuring comprehensive text extraction from complex layouts. Furthermore, Tesseract OCR delivered strong character-level recognition, effectively transcribing text from speech bubbles despite challenges with stylized fonts and noisy backgrounds. This demonstrates the feasibility of using our approach for automating manhwa translation tasks, even with the challenges posed by a low-resource language like Indonesian.




(a) Original Indonesian Manhwa panel

(b) Detected bounding boxes for speech bubbles

**Figure 2.** Side-by-side comparison of the original panel and bounding box detection.


doi.org/10.32388/4QZXCI


(a) Extracted enhanced speech bubbles


(b) Final translated Manhwa panel

```
OCR Text:
Bubble 0: KALIAN MAU
MENUNJUKAN
SESUATU
PADAKU?
Bubble 1: NE
Bubble 2: APA...
```

(c) Results of OCR on extracted speech bubbles

```
Translations:
Bubble 0: Would you like to offer something to me?
Bubble 1: NE
Bubble 2: What...
```

(d) Translated text from OCR

**Figure 3.** Pipeline steps: (a) Enhanced speech bubbles, (b) Final translated panel, (c) OCR results, and (d) Translated text.

## V. Comparative Results

### *A. Object Detection Models Comparison*

CO-DETR achieved a box mAP of 66% on the COCO test-dev dataset for general object detection tasks[9]. YOLOX-L attained a COCO-style AP of 70.2% on the Manga109-s dataset[10]. Our fine-tuned

YOLOv5xu model achieved an mAP@0.5:0.95 of 88.9% using the Webcomice dataset[4] for text bubble detection. This result highlights the advantage of domain-specific fine-tuning and the effectiveness of using a curated dataset tailored for Manhwa.

### *B. OCR Models Comparison*

A segmentation-free OCR method on comic books, achieved a CER of 22.78% and a WER of 39.30%[11]. Using Tesseract OCR, our pipeline achieved a CER of 3.1% and a WER of 8.6%, significantly outperforming the segmentation-free approach. The smaller dataset of high-quality images in our study and effective pre-processing likely contributed to these superior results.

### *C. Machine Translation Models Comparison*

Indonesian-English Neural Translation on IWSLT datasets gave BLEU = 23%, and Meteor = 57%[12]. The MarianMT model in our pipeline achieved a BLEU score of 27% and a Meteor score of 61%. These scores reflect better semantic retention and formal translation quality, likely due to fine-tuning on OpenSubtitles and Identic datasets tailored for conversational text.

Our pipeline consistently demonstrates superior performance across all key components when compared to existing models and methodologies. The results underscore the importance of domain-specific datasets, effective pre-processing, and targeted fine-tuning. While our smaller dataset yields strong results, expanding and diversifying the data will further enhance the pipeline's robustness and scalability for broader applications.

## VI. Analysis

The results of our pipeline demonstrate its effectiveness in automating Manhwa translation while addressing key research questions. By integrating text detection, transcription, translation, and reintegration, the pipeline significantly streamlines traditionally manual workflows. The YOLOv5xu model achieves high recall and precision, ensuring reliable detection of even small speech bubbles in complex layouts. Combined with Tesseract OCR and a fine-tuned MarianMT model, the system retains semantic accuracy and adeptly handles contextual nuances, highlighting its domain-specific focus and superior performance compared to general-purpose models.

While limitations persist, such as the lack of a dedicated Manhwa dataset, challenges with informal language, and computational constraints, the results validate the pipeline's potential to enhance

accessibility and efficiency in Manhwa translation. This work demonstrates the power of adapting machine learning techniques to low-resource tasks, addressing both linguistic and artistic complexities, and establishing a robust foundation for future advancements in automated Manhwa translation.

## VII. Conclusion

We presented an automated system for translating Manhwa from Indonesian to English, addressing the challenges of low-resource language processing. Our pipeline effectively combines computer vision, OCR, and machine translation to streamline the traditionally manual process. Evaluation results demonstrate the strong performance of each component, with YOLOv5x achieving high detection accuracy, Tesseract handling transcription reliably, and MarianMT providing meaningful translations. The system shows potential for broader applications in translating image-based content across underrepresented language pairs.

## VIII. Future Work

Future work for this project includes several directions to enhance its capabilities and impact. Expanding and annotating a larger, more diverse dataset for Manhwa is essential to improving model performance. Incorporating post-processing techniques can improve both text bubble detection and artistic consistency in text reintegration. Extending the pipeline to support additional low-resource language pairs and enabling the system to process entire chapters or multiple panels simultaneously, rather than single panels, will broaden its applicability and scalability, making it a versatile tool for global audiences.

## Acknowledgments

I would like to express my gratitude to my professor, Dr. Cece Alm, for her invaluable guidance and encouragement throughout the course of this project as part of the PSYC:681-Natural Language Processing course at Rochester Institute of Technology.

## References


1. a, b *Jocher G. Ultralytics YOLOv5 [software]. 2020. Available from: https://github.com/ultralytics/yolov5. doi:10.5281/zenodo.3908559.*
2. a, b *Smith R (2007). "Tesseract OCR Engine". https://web.archive.org/web/20160819190257/tesseract-ocr.googlecode.com/files/TesseractOSCON.pdf.*
3. a, b *Junczys-Dowmunt M, Grundkiewicz R, Dwojak T, Hoang H, Heafield K, Neckermann T, Seide F, Germann U, Fikri Aji A, Bogoychev N, et al. "Marian: Fast Neural Machine Translation in C++." In: Proceedings of ACL: System Demonstrations; 2018. p. 116–121.*
4. a, b, c *Nunes LB. "Webcomics Text Selection Dataset". Roboflow Universe. Roboflow; 2024 Dec. Available from: https://universe.roboflow.com/luciano-bastos-nunes/webcomics-text-selection. Visited on 2024-12-10.*
5. ^ *Rosa RM, Warsono I, Murni PE, Ariyanti W, Ika SK, Adriani M (2010). "Identic: The Indonesian-English Parallel Corpus." In: Proceedings of the 8th Workshop on Asian Language Resources (ALR 2010). 2010. p. 71–74.*
6. ^ *Lison P, Tiedemann J (2016). "OpenSubtitles2016: Extracting Large Parallel Corpora from Movie and TV Subtitles." In: Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016). pp. 923--929.*
7. ^ *Zhang G, Yan X, Iwamura M, Matsuo Y (2021). "Automatic Manga Text Detection and Recognition." arXiv preprint arXiv:2012.14271v2.*
8. ^ *Baek Y, Lee B, Han D, Yun S, Lee H (2021). "Efficient and Accurate Scene Text Detector". arXiv preprint arXiv:2103.14027v3.*
9. ^ *Zong Z, Song G, Liu Y (2023). "DETRs with Collaborative Hybrid Assignments Training." In: Proceedings of the International Conference on Computer Vision. pp. 6725-6735. doi:10.1109/ICCV51070.2023.00621.*
10. ^ *Shinya Y. "USB: Universal-Scale Object Detection Benchmark." 2021 Mar. doi:10.48550/arXiv.2103.14027.*
11. ^ *Rigaud C, Burie JC, Ogier JM (2017). "Segmentation-Free Speech Text Recognition for Comic Books." In: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), vol. 03, pp. 29-34. doi:10.1109/ICDAR.2017.288.*

12. ^*Dwiastuti M. "English-Indonesian Neural Machine Translation for Spoken Language Domains." In: Alva-Manchego F, Choi E, Khashabi D, editors. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop. Florence, Italy: Association for Computational Linguistics; 2019. p. 309-314. Available from: https://aclanthology.org/P19-2043. doi:10.18653/v1/P19-2043.*

## Declarations

**Funding:** No specific funding was received for this work.

**Potential competing interests:** No potential competing interests to declare.