Translating Manga with Multimodal Large Language Models: Challenges and Benchmarks

Tomasz Nitsch
tomasz.nitsch.stud@pw.edu.pl
Faculty of Electrical Engineering, Warsaw University of Technology

Introduction

- Japanese manga is globally popular, but translation is difficult.
- Japanese is a high-context language with ambiguous grammar and cultural nuances difficult to translate even for proffesional translators.
- Traditional translation is time-consuming and expensive.
- Can MMLLMs help reduce the workload for translators?

Challenges in Japanese Translation

- **Context-dependence:** Subjects often omitted.
- **SOV structure:** Unfinished sentences are common.
- **Cultural nuances:** No direct translations for many concepts.
- **Ambiguity:** Homophones and vague tenses.

Dataset

- Gathered **4345 manga pages** across genres
- **29631 line pairs** JP-EN
- Manga109 used for training page element detection
- Used previous works' OpenMantra dataset (1593 sentences) for testing

Model used for training

- Model: **Qwen2.5-VL-32B**
- Visual-language understanding
- Trained per-page with context (genre, image, bubble and text order)

Methodology

Preprocessing:

1. Panel and bubble detection (YOLOv5s)
2. Bubble text OCR + reading order

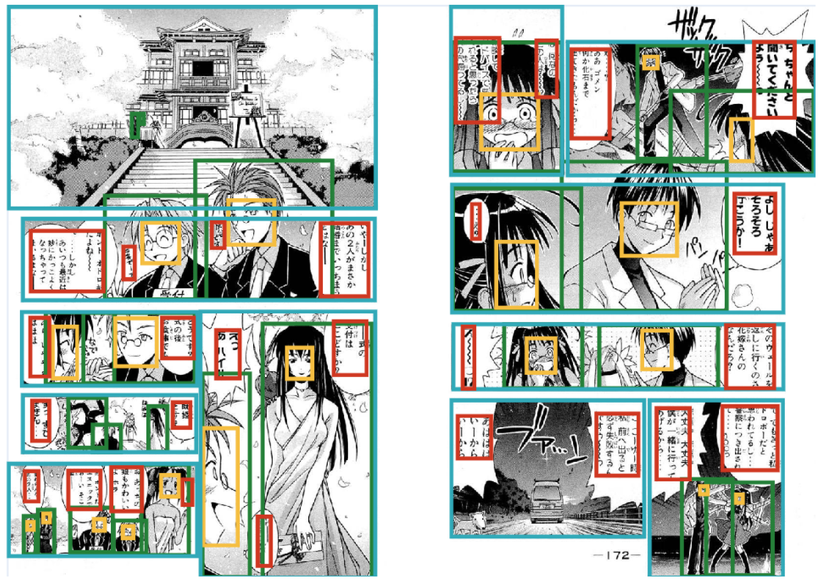

Figure 1: Page from Manga109 dataset used for training and testing, blue - panels, yellow - character's face, green - character's body, red - bubble text

Translation Methods:

- **PageByPage**
- **1+PageByPage+1** - has previous last and next first page's bubble
- **1+PageByPage+1 + synopsis**

Evaluation metrics:

Metric	Measures	Range	Good Score
Cosine	Angular similarity between sentence embeddings	-1 to 1	≥ 0.35 (decent) ≥ 0.5 (strong)
chrF	Character-level n-gram F-score (precision/recall)	0 to 100	≥ 30 (okay) ≥ 40 (good)
BLEU	Basic N-gram overlap	0 to 100	≥ 15 (passable) ≥ 25 (strong)
BLEURT	Captures fluency + adequacy	-1 to 1	≥ 0.5 (good) ≥ 0.6 (great)
xCMT	Source-target semantic alignment	0 to 1	0.7 (strong) ≥ 0.75 (very strong)

Table: Summary of translation evaluation metrics for Japanese-to-English LLM output

Translation Example

1. That's because...
   It's because my dream
2. ...I want to become an astronaut, see the world from space.
   is to become an astronaut and see the round earth.
3. This is my very first step towards my dream.
   This is my very first step towards my dream.
4. That's why I have to take these pictures myself.
   That's why I have to take the photo myself.
5. People only talk like that...
   Those who can speak of dreams like that
6. ...if they've never seen their dreams end
   Are only those who have never seen how dreams end.

Figure 2: Side-by-side translation: Model vs Human reference

Results

Performance Table (summary):

Method	Cosine	ChrF	BLEU	BLEURT	COMET
Qwen2.5-VL-32B	0.293	29.9	12.3	0.427	0.714
PageByPage	0.345	34.5	15.2	0.502	0.734
1+PBP+1	0.349	34.7	15.1	0.501	0.727
1+PBP+1+S	**0.389**	36.5	**15.6**	0.553	0.756
PBP-VIS-NUM	-	**36.8**	15.6	**0.582**	**0.776**

Table: Summary of evaluation results of models against OpenMantra dataset, PBP-VIS-NUM is previous work's best approach using gpt-4-turbo-2024-04-09

Conclusion

- MMLLMs show promising results in translating manga with context.
- While inferior to GPT-4 this significantly smaller model (1.4TB vs 32B parameters) can still offer a potential hybrid translation approach.
- Future work: Larger datasets, bigger/better models, hybrid pipelines.