# Machine Translation of Cooking Videos Using Descriptions of the Images by Chain-of-Thought Augmentation

**Anonymous ACL submission**

## Abstract

English cooking videos often contain polysemous words and omitted expressions, making accurate translation challenging. This study aims to improve English-Japanese machine translation of cooking videos by utilizing images extracted from the video. We adopt a Chain-of-Thought Augmentation (CoTA) approach, where the model generates descriptions of images and utilizes them as auxiliary information for the translation task. In our experiments, we selected sentences from an English-Japanese cooking video corpus that were difficult to translate due to polysemous words. We evaluated the performance using GPT-4o and Qwen2-VL with COMET and BLEU scores. The results demonstrate that incorporating images improves translation accuracy, with a particularly strong tendency for CoTA applied to GPT-4o to produce more accurate translations.

## 1 Introduction

In recent years, with the proliferation of video sharing platforms, cooking videos in different languages have been shared on these platforms, increasing the demand for their translation. However, research specifically focusing on cooking video translation is scarce.

In translating cooking videos, accurately rendering polysemous words poses a challenge. English cooking videos are particularly difficult to translate due to the large number of polysemous words for English words. For example, the word "pepper" has several meanings, such as green pepper, paprika, chilli pepper and black pepper, making it difficult to select the appropriate meaning from the information contained in the text alone.

Multimodal machine translation (Delbrouck and Dupont, 2017; Calixto et al., 2017; Huang et al., 2016; Zhang et al., 2020; Sulubacak et al., 2020; Wu et al., 2021), which refers not only to text but also to speech and images, has the potential to address this a issue. Furthermore, the Chain-of-Thought (CoT) (Wei et al., 2022) approach, which processes complex problems in a step-by-step manner, is also considered effective. Inspired by the success of CoT, various Chain-of-X (CoX) (Xia et al., 2025) paradigms have been developed to address challenges across diverse domains and tasks. CoX is a generalization of CoT, constructing a continuous process with various components beyond reasoning thoughts, such as Chain-of-Feedback (Xu et al., 2025), Chain-of-Instructions (Hayati et al., 2025), and Chain-of-Histories (Xia et al., 2024). Notably, Chain-of-Thought Augmentation (CoTA) (Shim et al., 2024), which extends the chain with external knowledge when LLMs have limited information for specific tasks or domains, has emerged as an effective approach.

In this study, to improve the performance of multimodal machine translation for cooking videos, we propose a method that not only takes images as input but also the CoTA concept, which utilizes external information to expand the model's knowledge. In addition, we will construct a dataset to evaluate the effectiveness of multimodal machine translation. In the experiments, we primarily compare three methods: a baseline machine translation method that uses only text as input, a multimodal machine translation method that inputs images along with text, and our proposed methods that utilize CoTA concept.

## 2 Related Work

Chain-of-Thought(CoT) (Wei et al., 2022) is a method for solving complex reasoning tasks by decomposing them into steps, and has attracted much attention in recent years. The method divides a problem into a series of steps that are processed sequentially, allowing the model to produce more logical and explicable reasoning. Chain-of-Thought Augmentation (CoTA) (Shim et al., 2024) is one of

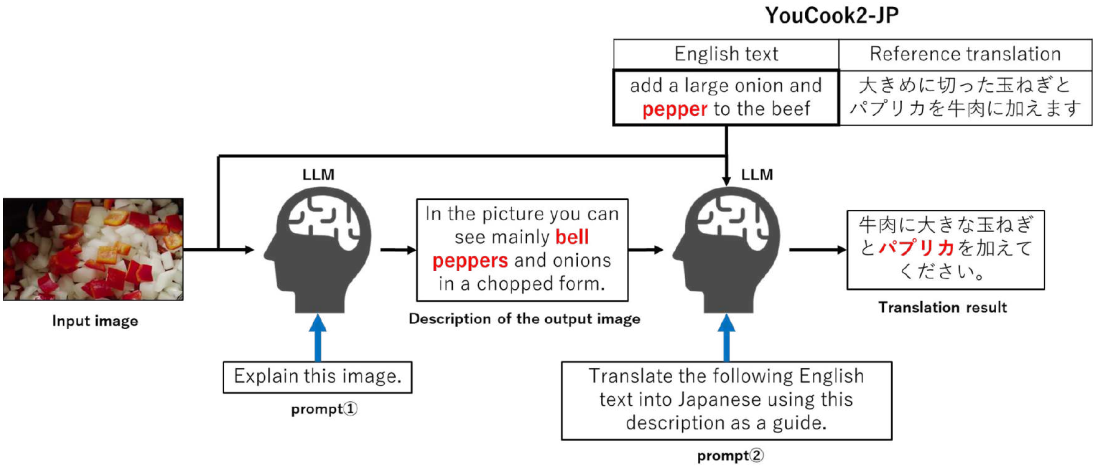


Figure 1: Overview of the proposed method

the Chain-of-X (CoX) (Xia et al., 2025) paradigms developed, inspired by the success of CoT. CoTA aims to extend the capabilities of LLMs by incorporating external additional information when they have limited knowledge for specific tasks or domains. A characteristic feature of this method is that the chain is constructed from various types of augmented data. For instance, some approaches integrate explicit retrieval steps into the reasoning chain, enabling LLMs to acquire external knowledge that is not present in pre-trained data or is outdated. Examples include ReAct (Yao et al., 2023) and Chain-of-Knowledge (Li et al., 2024), which enhance the LLM's reasoning process based on external information.

## 3 Proposed Method

### 3.1 Multimodal Machine Translation Using Chain-of-Thought Augmentation

In this proposed method (COTA), we aim to improve translation accuracy by incorporating image information into a stepwise inference process using the CoTA (Shim et al., 2024) concept, thereby enabling the model to gain a deeper understanding of the information obtained from the image. Specifically, as shown in Figure 1, the model is instructed to describe the content of an image acquired from a cooking video, and this description is then added as external knowledge to the translation process. Furthermore, we will also evaluate an applied version of this method (COTA+IMAGE), which re-inputs the image along with its description. Table 1 shows the Japanese prompts actually used. The English translations are provided in parentheses. Prompts for other methods are provided in Appendix B.

### 3.2 Construction of Evaluation Data

This section describes the evaluation dataset we have developed. To evaluate the effectiveness of multimodal machine translation, an evaluation set comprising text that is difficult to translate and requires image information for proper translation is considered useful. Therefore, we created a new dataset to confirm the effectiveness of multimodal machine translation utilizing images.

Since cooking videos are video data, they cannot be directly used for image-based multimodal machine translation. Thus, we constructed a dataset suitable for this purpose. First, from YouCook2-JP[1], a multimodal bilingual corpus with videos, we extracted a total of 150 English sentences that were manually judged to be difficult to translate using only textual information, such as those containing polysemous words like "pepper," and used them as evaluation data. Next, the following procedure was employed to obtain appropriate images for each English sentence. As the data contained timestamps indicating when each English sentence appeared in the video, we used this information to extract 10 images per English sentence using OpenCV[2]. Subsequently, the similarity between the English sentence and the images was measured using CLIP[3], and the image with the highest similarity was acquired from the 10 images. The details of this construction steps are illustrated in Appendix C.

[1] https://github.com/nlab-mpg/YouCook2-JP
[2] https://opencv.org/
[3] Phttps://github.com/openai/CLIP

| Method | Prompt |
|---|---|
| CoTA | この画像について説明してください． (Please describe this image.)<br>この文章を参考にして，次の英語の文章を日本語訳してください．<br>(Referring to this description, please translate the following English sentence into Japanese.) |
| CoTA+IMAGE | この画像について説明してください． (Please describe this image.)<br>この画像と画像に対する説明を参考にして，次の英語の文章を日本語訳してください．<br>(Referring to this image and the description of the image, please translate the following English sentence into Japanese.) |

Table 1: Prompts used in our CoTA methods. (English translations are in parentheses.)

## 4 Experimental Evaluation

### 4.1 Experimental Setup

In this study, gpt-4o-2024-08-06 (gpt, 2023) (GPT-4o) and Qwen2-VL-72B-Instruct[4] (Wang et al., 2024) (Qwen2-VL) were used as multimodal large language models. GPT-4o was accessed via API. We downloaded the Qwen2-VL model from Hugging Face[5] and ran it on our local machine.

We compared and evaluated five cases, with the English-only input serving as the baseline (TEXT-ONLY); inputting both English text and images (TEXT+IMAGE); using the CoT (Wei et al., 2022) approach (COT); using the CoTA (Shim et al., 2024) approach (COTA); and re-inputting images along with the generated descriptions in the CoTA approach (COTA+IMAGE). A simplified illustration outlining the differences between each method is provided in Appendix A

The evaluation dataset consists of 150 sentence pairs and images, and was constructed from YouCook2-JP, which is an English-Japanese translation of the first 600 videos from the English cooking video dataset, YouCook2[6], as described in Section 3.2. The COMET (Rei et al., 2020) and BLEU (Papineni et al., 2002) were used to evaluate the performance of the translation.

### 4.2 Experimental Results

The experimental results using GPT-4o are presented in Table 2, and those using Qwen2-VL are shown in Table 3, respectively. In the experiments utilizing GPT-4o, the CoTA approach, which incorporates image descriptions and images as input, achieved the highest scores in both BLEU and COMET. Furthermore, the BLEU score of TEXT+IMAGE improved by 2.1 points compared to TEXT-ONLY, and that of COTA+IMAGE improved by 1.71 points compared to COTA. This confirms

| | COMET | BLEU(%) |
|---|---|---|
| TEXT-ONLY | 0.8675 | 33.20 |
| TEXT+IMAGE | 0.8581 | 35.10 |
| COT | 0.8670 | 32.76 |
| COTA | 0.8788$^{*\dagger}$ | 35.75 |
| COTA+IMAGE | **0.8816**$^{*\dagger}$ | **37.46**$^{*\dagger}$ |

Table 2: Experimental results using GPT-4o

$^{*}$ $p < 0.05$ against TEXT-ONLY (paired t-test)
$^{\dagger}$ $p < 0.05$ against TEXT+IMAGE (paired t-test)

| | COMET | BLEU(%) |
|---|---|---|
| TEXT-ONLY | 0.8500 | 27.77 |
| TEXT+IMAGE | 0.8533 | **28.63** |
| COT | 0.8482 | 23.62 |
| COTA | **0.8559** | 26.28 |
| COTA+IMAGE | 0.8546 | 27.17 |

Table 3: Experimental results using Qwen2-VL

that for GPT-4o, incorporating image information into the input tends to significantly contribute to an increase in BLEU scores. Conversely, when the CoT approach was applied, the results indicated that neither BLEU nor COMET scores showed a statistically significant difference compared to the baselines, or were slightly lower. This suggests that the CoT approach may not directly contribute to translation performance improvement in this task.

In experiments employing the Qwen2-VL model, in contrast to the GPT-4o case, none of the proposed methods yielded results that statistically significantly outperformed the baselines.

Based on these results, it is inferred that the improvement in translation performance achieved by the CoTA approach is highly dependent on the performance of the model used. While the image-leveraging CoTA approach proves effective in enhancing translation performance with highly capa-

[4] https://huggingface.co/Qwen/Qwen2-VL-72B
[5] https://huggingface.co/
[6] http://youcook2.eecs.umich.edu/

| | GPT-4o | Qwen2-VL |
|---|---|---|
| TEXT-ONLY | 41.3% (62) | 17.6% (25) |
| TEXT+IMAGE | 44.7% (67) | 24.0% (36) |
| COT | 48.7% (73) | 25.3% (38) |
| COTA | 54.7% (82) | 21.3% (32) |
| COTA+IMAGE | **59.3% (89)** | **26.7% (40)** |

Table 4: Comparison of accuracy (Number of correct words in parentheses)

| | |
|---|---|
| **Image** | |
| **Image Description** | 画像では、木製のまな板の上でしいたけ(**shiitake mushroom**)をスライスしている場面が写っています。 |
| **English Sentence** | cut the **mushroom** into thin slices |
| **CoAT Output** | しいたけ(**shiitake mushroom**)を薄切りにする |
| **Reference** | しいたけ(**shiitake mushroom**)を薄切りにします |

Table 5: An example where image description was successfully reflected in the output (English translations are in parentheses)

ble multimodal models like GPT-4o, its efficacy might be limited with other models.

### 4.3 Analysis

We focused on specific words within each English sentence and individually assessed the accuracy of their translations. Our analysis targeted polysemous words like pepper, as well as context-dependent words such as sheet when translated as "nori" (seaweed) in the context of making sushi rolls. For each method, we counted the number of sentences out of the total 150 where the target word was correctly translated. Table 4 shows the result of the analysis. It shows that in many cases, the proposed methods outperformed the baselines. This suggests that these proposed methods are effective, particularly in contexts where translation is difficult due to literal interpretations. Furthermore, in experiments using GPT-4o, the application of the CoTA approach led to particularly notable performance improvements.

When utilizing CoTA, we observed cases where the model correctly grasped the image content, yet this information was not reflected in the translation. Table 5 is a successful example in COTA (GPT-4o). In this example, the content of the image description was successfully reflected in the translation, correctly translating "mushroom" (referring to mushrooms in general) as "shiitake." On the other hand, Table 6 is a failure example in COTA (GPT-4o). In this example, despite the image description stating "red bell pepper," the translation result rendered "pepper" as "kosho" (black pepper). Such cases were particularly frequently observed in the results with Qwen2-VL. This suggests that even if the model correctly understands the image content through CoTA, it may not always effectively reflect that information in the translation results. To solve this problem, we believe that it is necessary to design prompts and improve methods to more effectively reflect information obtained from images in the translation.

| | |
|---|---|
| **Image** | |
| **Image Description** | ボウルの中には、赤い野菜（おそらく赤ピーマン(**bell pepper**)）、グリーンの葉物野菜が見えます。 |
| **English Sentence** | put the chicken black beans green onions cilantro and **pepper** in the rice |
| **CoAT Output** | 鶏肉、黒豆、青ネギ、パクチー、そしてコショウ(**black pepper**)を、ご飯の中に入れてください。 |
| **Reference** | 鶏肉、ブラックビーンズ、青ねぎ、コリアンダー，パプリカ(**bell pepper**)）をご飯に混ぜます |

Table 6: An example where image description was not reflected in the output (English translations are in parentheses)

## 5 Conclusion

This study proposed multimodal machine translation methods leveraging image information to enhance the translation accuracy of polysemous and context-dependent expressions in cooking videos. We introduced the Chain-of-Thought Augmentation (CoTA) concept, which utilizes image descriptions as auxiliary information, and constructed a dataset to evaluate the effectiveness of multimodal machine translation. Experiments with GPT-4o demonstrated that COTA, particularly COTA+IMAGE achieved the highest scores in both BLEU and COMET, leading to a statistically significant improvement in translation quality compared to the baselines.

## Limitations

The evaluation data used in the experiments conducted in this study were small. An evaluation using more data is preferable.

## References


2023. GPT-4V(ision) System Card.

Iacer Calixto, Qun Liu, and Nick Campbell. 2017. Doubly-attentive decoder for multi-modal neural machine translation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1913–1924, Vancouver, Canada. Association for Computational Linguistics.

Jean-Benoit Delbrouck and Stéphane Dupont. 2017. Modulating and attending the source image during encoding improves multimodal translation. *Preprint*, arXiv:1712.03449.

Shirley Anugrah Hayati, Taehee Jung, Tristan Bodding-Long, Sudipta Kar, Abhinav Sethy, Joo-Kyung Kim, and Dongyeop Kang. 2025. Chain-of-instructions: Compositional instruction tuning on large language models. *Preprint*, arXiv:2402.11532.

Po-Yao Huang, Frederick Liu, Sz-Rung Shiang, Jean Oh, and Chris Dyer. 2016. Attention-based multimodal neural machine translation. In *Proceedings of the First Conference on Machine Translation: Volume 2, Shared Task Papers*, pages 639–645, Berlin, Germany. Association for Computational Linguistics.

Xingxuan Li, Ruochen Zhao, Yew Ken Chia, Bosheng Ding, Shafiq Joty, Soujanya Poria, and Lidong Bing. 2024. Chain-of-knowledge: Grounding large language models via dynamic knowledge adapting over heterogeneous sources. *Preprint*, arXiv:2305.13269.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: A Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.

Ricardo Rei, Craig Stewart, Ana C Farinha, and Alon Lavie. 2020. COMET: A Neural Framework for MT Evaluation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*.

Jay Shim, Grant Kruttschnitt, Alyssa Ma, Daniel Kim, Benjamin Chek, Athul Anand, Kevin Zhu, and Sean O'Brien. 2024. Chain-of-thought augmentation with logit contrast for enhanced reasoning in language models. *Preprint*, arXiv:2407.03600.

Umut Sulubacak, Ozan Caglayan, Stig-Arne Grönroos, Aku Rouhe, Desmond Elliott, Lucia Specia, and Jörg Tiedemann. 2020. Multimodal Machine Translation through Visuals and Speech. *Machine Translation*, pages 97–147.

Peng Wang, Shuai Bai, Sinan Tan, Shijie Wang, Zhihao Fan, Jinze Bai, Ke-Yang Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Yang Fan, Kai Dang, Mengfei Du, Xuancheng Ren, Rui Men, Dayiheng Liu, Chang Zhou, Jingren Zhou, and Junyang Lin. 2024. Qwen2-VL: Enhancing Vision-Language Model's Perception of the World at Any Resolution. *ArXiv*.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Ed H. Chi, Quoc Le, and Denny Zhou. 2022. Chain of Thought Prompting Elicits Reasoning in Large Language Models. *CoRR Journal*.

Zhiyong Wu, Lingpeng Kong, Wei Bi, Xiang Li, and Ben Kao. 2021. Good for misconceived reasons: An empirical revisiting on the need for visual context in multimodal machine translation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 6153–6166, Online. Association for Computational Linguistics.

Yu Xia, Rui Wang, Xu Liu, Mingyan Li, Tong Yu, Xiang Chen, Julian McAuley, and Shuai Li. 2025. Beyond chain-of-thought: A survey of chain-of-X paradigms for LLMs. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 10795–10809, Abu Dhabi, UAE. Association for Computational Linguistics.

Yuwei Xia, Ding Wang, Qiang Liu, Liang Wang, Shu Wu, and Xiao-Yu Zhang. 2024. Chain-of-history reasoning for temporal knowledge graph forecasting. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 16144–16159, Bangkok, Thailand. Association for Computational Linguistics.

Qingting Xu, Kaisong Song, Chaoqun Liu, Yangyang Kang, Xiabing Zhou, Jun Lin, and Yu Hong. 2025. COF: Adaptive chain of feedback for comparative opinion quintuple extraction. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 3236–3247, Abu Dhabi, UAE. Association for Computational Linguistics.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. 2023. React: Synergizing reasoning and acting in language models. *Preprint*, arXiv:2210.03629.

Zhuosheng Zhang, Kehai Chen, Rui Wang, Masao Utiyama, Eiichiro Sumita, Zuchao Li, and Hai Zhao. 2020. Neural machine translation with universal visual representation. In *International Conference on Learning Representations*.


## Appendix

### A Additional Experiments: COT+COTA

We also conducted additional experiments with COT+COTA, a method combining the comparative method COT and the proposed method COTA. The experimental results are shown in Tables 7 to 9.

| | COMET | BLEU(%) |
|---|---|---|
| TEXT-ONLY | 0.8675 | 33.20 |
| TEXT+IMAGE | 0.8581 | 35.10 |
| COT | 0.8670 | 32.76 |
| COTA | 0.8788*† | 35.75 |
| COTA+IMAGE | **0.8816***† | **37.46***† |
| **COT+COTA** | 0.8695 | 33.22 |

Table 7: Experimental results using GPT-4o

* $p < 0.05$ against TEXT-ONLY (paired t-test)
† $p < 0.05$ against TEXT+IMAGE (paired t-test)

This method generates a step-by-step thought process based on the image description. A simplified illustration of its differences from other methods is presented in Figure 2.

The results indicate some improvement in the translation of polysemous words, but the BLEU and COMET scores did not show significant improvement.

## B Prompts

The specific prompts for the five methods are shown in Table 10, and their English translations in Table 11.

For COT, controlling the output was challenging because we needed to generate not only the final Japanese translation but also the thought process. Therefore, the prompts were designed to clearly indicate and separate the output sections for the thought process and the Japanese translation using markers.

## C Data Set Construction Process

Figure 3 shows the procedure for constructing evaluation data.

## D AI Use Statement

We used AI assistants in coding and draft refinement, e.g., translation, grammar check, and rewriting.

| | COMET | BLEU(%) |
|---|---|---|
| TEXT-ONLY | 0.8500 | 27.77 |
| TEXT+IMAGE | 0.8533 | **28.63** |
| COT | 0.8482 | 23.62 |
| COTA | **0.8559** | 26.28 |
| COTA+IMAGE | 0.8546 | 27.17 |
| **COT+COTA** | 0.8502 | 26.47 |

Table 8: Experimental results using Qwen2-VL

| | GPT-4o | Qwen2-VL |
|---|---|---|
| TEXT-ONLY | 41.3% (62) | 17.6% (25) |
| TEXT+IMAGE | 44.7% (67) | 24.0% (36) |
| COT | 48.7% (73) | 25.3% (38) |
| COTA | 54.7% (82) | 21.3% (32) |
| COTA+IMAGE | **59.3% (89)** | 26.7% (40) |
| **COT+COTA** | 48.7% (73) | **27.3% (41)** |

Table 9: Comparison of accuracy (Number of correct words in parentheses)

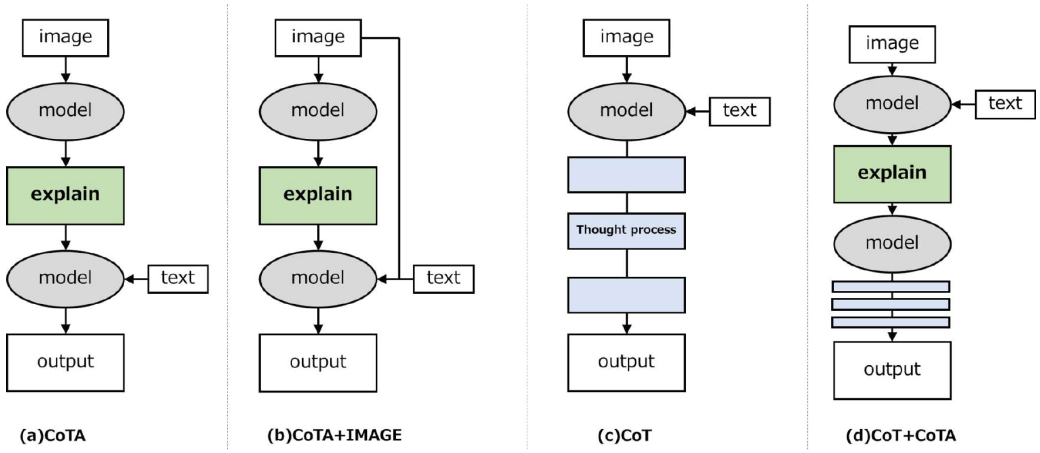


Figure 2: Multimodal translation methods used in this paper

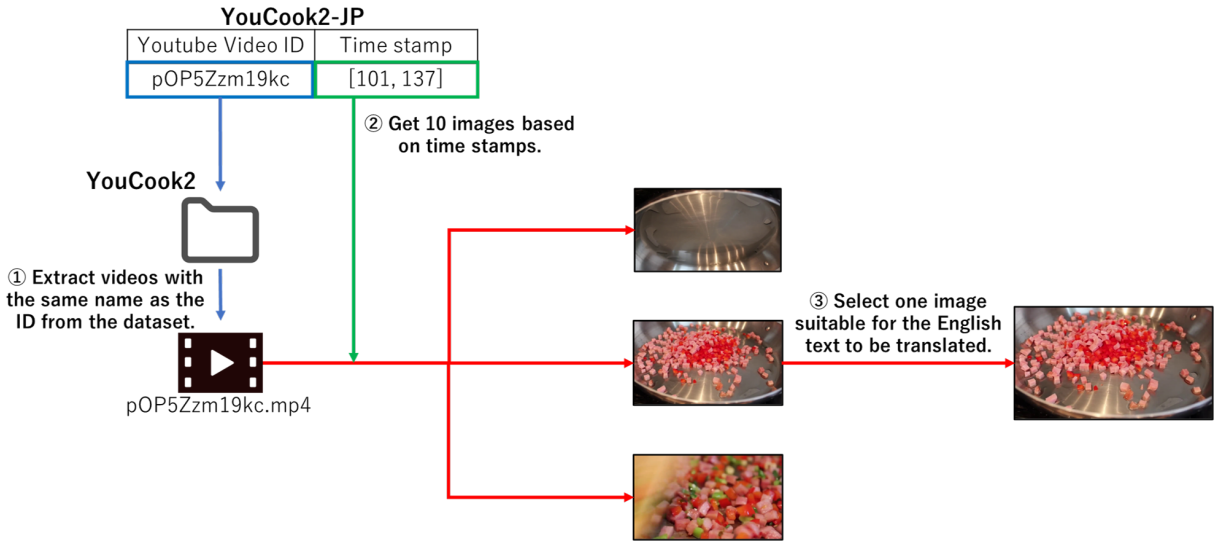


Figure 3: Procedure of constructing the evaluation data

| Method | Prompt |
|---|---|
| TEXT-ONLY | 次の英語文を日本語に翻訳し，翻訳結果のみを提供してください． |
| TEXT+IMAGE | 次の英語文を，次に示す画像の内容を考慮しながら翻訳し，翻訳結果のみを<br>提供してください．「申し訳ありませんが～」などの前置きを出力してはいけません． |
| COT | 以下の英文を日本語に翻訳してください．入力された画像から得られる情報を参考に<br>して，ステップバイステップで思考プロセスを出力してください．<br>**出力形式の指示:**<br>思考プロセスと最終的な日本語訳を，以下の指定されたマーカーで厳密に区切って<br>出力してください．他の説明や前置きは一切含めないでください．<br>翻訳できない場合でも，マーカーと空のセクションを出力してください．<br>—思考プロセス開始—<br>翻訳に至るまでの思考プロセスを詳細に記述してください．<br>—思考プロセス終了—<br>—最終的な日本語訳開始—<br>上記の思考プロセスを踏まえた，この英文に対する最終的な日本語訳のみをここに<br>記述してください．—<br>—最終的な日本語訳終了— |
| COTA | この画像について説明してください．<br>この文章を参考にして，次の英語の文章を日本語訳してください． |
| COTA+IMAGE | この画像について説明してください．<br>この画像と画像に対する説明を参考にして，次の英語の文章を日本語訳してください． |
| COT+COTA | 以下の英文を日本語に翻訳してください．まず入力された画像を説明し、その説明内容<br>を基にして，ステップバイステップで思考プロセスを出力してください．<br>**出力形式の指示:**<br>画像説明，思考プロセス，最終的な日本語訳を，以下の指定されたマーカーで厳密に<br>区切って出力してください．他の説明や前置きは一切含めないでください．<br>翻訳できない場合でも，マーカーと空のセクションを出力してください．<br>—画像説明開始—<br>入力された画像の内容を詳細に記述してください．<br>—画像説明終了—<br>—思考プロセス開始—<br>翻訳に至るまでの思考プロセスを詳細に記述してください．<br>—思考プロセス終了—<br>—最終的な日本語訳開始—<br>上記の思考プロセスを踏まえた，この英文に対する最終的な日本語訳のみをここに<br>記述してください．—<br>—最終的な日本語訳終了— |

Table 10: Prompts used in the six methods

| Method | Prompt |
|---|---|
| TEXT-ONLY | Please translate the following English sentence into Japanese and provide only the translation result. |
| TEXT+IMAGE | Please translate the following English sentence, considering the content of the image shown next, and provide only the translation result. Do not output any introductory remarks such as "I apologize, but...". |
| COT | Please translate the following English sentence into Japanese. Refer to the information obtained from the input image and output the thought process step-by-step.<br>Output Format Instructions:<br>Strictly separate the thought process and the final Japanese translation using the specified markers below. Do not include any other explanations or introductory remarks.<br>Even if translation is not possible, please output the markers and empty sections.<br>—Thought Process Start—<br>Describe in detail the thought process leading to the translation.<br>—Thought Process End—<br>—Final Japanese Translation Start—<br>Based on the thought process above, write only the final Japanese translation for this English sentence here.<br>—Final Japanese Translation End— |
| COTA | Please describe this image.<br>Referring to this description, please translate the following English sentence into Japanese. |
| COTA+IMAGE | Please describe this image.<br>Referring to this image and the description of the image, please translate the following English sentence into Japanese. |
| COT+COTA | Please translate the following English sentence into Japanese. First, describe the input image, and then, based on that description, output the step-by-step thought process.<br>Output Format Instructions:<br>Please output the image description, thought process, and final Japanese translation strictly separated by the specified markers<br>separated by the specified markers below. Do not include any other explanations or introductory remarks.<br>Even if translation is not possible, please output the markers and empty sections.<br>—Image Description Start—<br>Please describe the content of the input image in detail.<br>—Image Description End—<br>—Thought Process Start—<br>Describe in detail the thought process leading to the translation.<br>—Thought Process End—<br>—Final Japanese Translation Start—<br>Based on the thought process above, write only the final Japanese translation for this English sentence here.<br>—Final Japanese Translation End— |

Table 11: English translations of the prompts used in the six methods