Technology for Translating, Visualizing, and Generating Recipes

Multimodal Machine Translation of Chinese into English

Chan Sin-wai

First published 2026

ISBN: 9781032308425 (hbk)
ISBN: 9781032308432 (pbk)
ISBN: 9781003306948 (ebk)

8 *VisualRecipe* for Visualizing Recipes through Image Technology

DOI: 10.4324/9781003306948-10

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

8 *VisualRecipe* for Visualizing Recipes through Image Technology

Introduction

Despite the fact that the quality of translation for general texts has improved enormously with the introduction of neural machine translation in 2013, domain-specific systems fare better in terms of accuracy and overall performance. That accounts for the development of a translation system for food and drinks, and a system for recipe translation *TransRecipe* in 2000 by the author. *TransRecipe* is a fully-automatic translation system for translating Chinese cookbooks into English based on the transfer model which combines the corpus-based, example-based, pattern-based, and rule-based approaches. With a lexical database of around 20,000 entries relating to Chinese food and drinks, a general dictionary of around 2,000 frequently used expressions, a database of 200 examples, and a total of 700 global parsing rules, *TransRecipe* can automatically translate a recipe from Chinese into English within seconds without any pre-processing or post-editing.

The success of *TransRecipe* leads to the creation of *VisualRecipe*, which is an online visual translation system for Chinese cookbooks.

VisualRecipe is an online translation system with three generation modes: textual translation, visual translation, and textual-visual translation.

Textual Translation

The interface of textual translation consists of three parts: the upper part is the title of the system, which is "VisualRecipe System", the left side of the middle part is the Text-in Box for inputting the source text, and the right side is the Text-out Box for displaying the target text output. The input text could be copied and pasted onto the Text-in Box for processing, or it could be retrieved from the recipe database stored in the computer or cloud. When the button of TT (abbreviation for Textual Translation) is clicked and "Translate" activated, the target text will be generated and displayed in the Text-out Box.

DOI: 10.4324/9781003306948-10

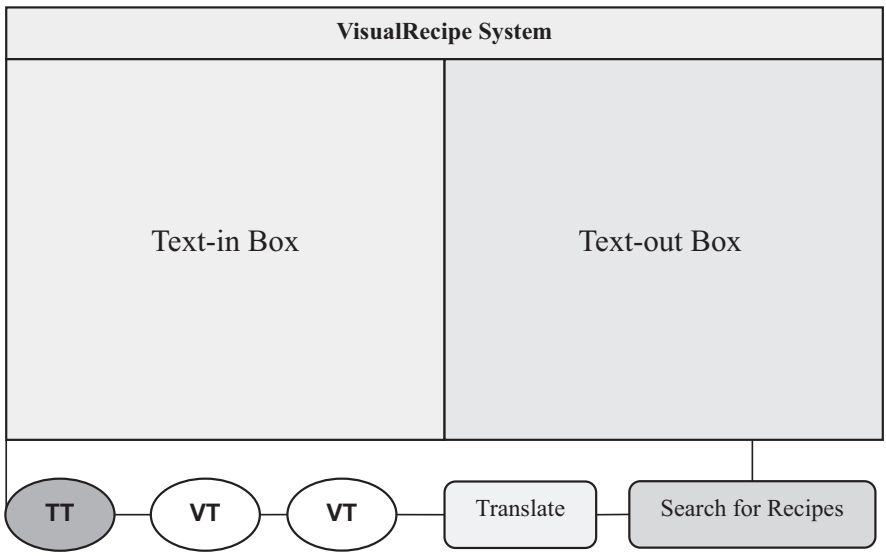

Figure 8.1 A recipe translation system.

Flowchart of Textual Translation

Machine translation systems work on sentences. When a source sentence in Chinese is put into the textual translation mode of the *VisualRecipe* system, it will be pre-edited automatically by a sentence-splitter to generate a new source sentence, in line with the short-sentence approach propagated by me. This new sentence will go through an example database to find out if an example sentence is stored in it. If it does, then the sentence will be translated as a target sentence. If it does not, then the pre-edited sentence will go through the general dictionary database, which has all its terms grammatically tagged, and the specialized dictionary database, with all the terms on food and cooking ingredients tagged according to their categories. Then the source sentence will be analysed according to its syntax, transferred to the target language according to the rules stored in the rule database, and the target sentence will be duly produced.

The process of textual translation is shown in the following flowchart.

Visual Translation

The interface of visual translation is the same as that of textual translation except that the Text-out Box will be for displaying pictures / images rather

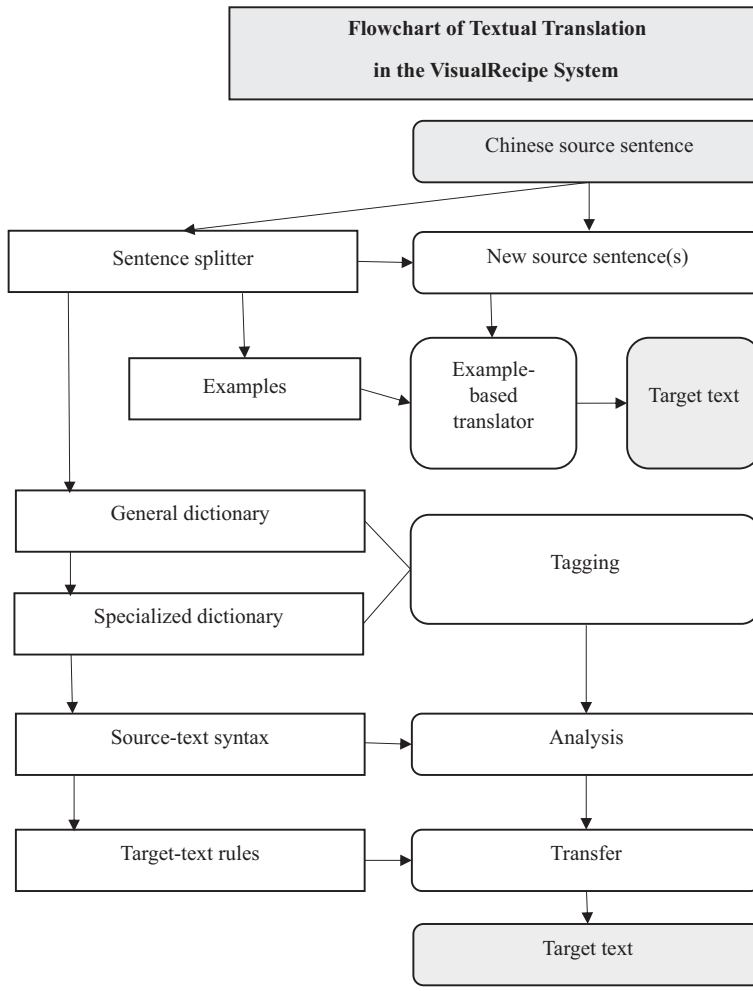

Figure 8.2 Flowchart of textual translation in the *VisualRecipe* system.

than the target text. The input text could be copied and pasted onto the box for processing, or it could be retrieved from the recipe database stored in the computer or cloud. When the button of VT (abbreviation for Visual Translation) is clicked and "Translate" activated, the target images will be displayed in the Picture-out Box.

Visual translation in the context of machine translation is about translating a text by pictures or images in general. When a source sentence in Chinese is put into visual translation, it will be translated into a picture. Theoretically, a recipe of a certain number of sentences will be translated into an equivalent

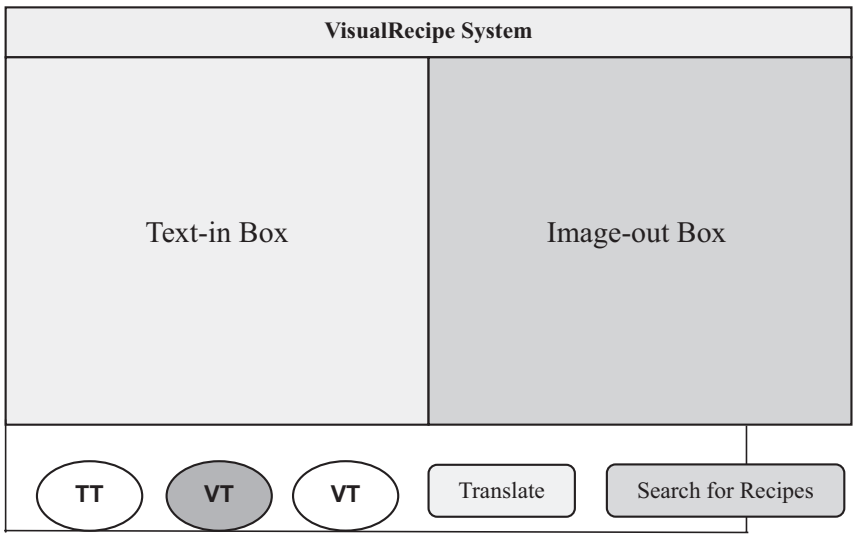

Figure 8.3 A proposed *VisualRecipe* system.

number of pictures. But in some cases, several sentences can be expressed through a single picture.

The process of visual translation is shown in the following flowchart.

Textual-visual Translation

Textual-visual translation is a combination of textual translation and visual translation, with an interface displaying the textual and visual translations of the original text. The generation processes of the textual and visual translations are the same as described above, the only difference lies with the way the translations are presented, as shown below.

This innovative textual-visual translation serves to cater for a large number of people who would like to prepare dishes according to the English translations of the Chinese recipes. What is most important here is that since pictures are language-independent, anyone can prepare Chinese dishes by following the order of pictures, which is a description of the process to cook and prepare a certain dish.

Mouse-over Visual Presentation

Another method to show visually the preparation of a dish is by mouse-over. Mouse-over, also known as hover, is a user interface interaction where an action is initiated by moving a cursor or pointer over a specific graphical

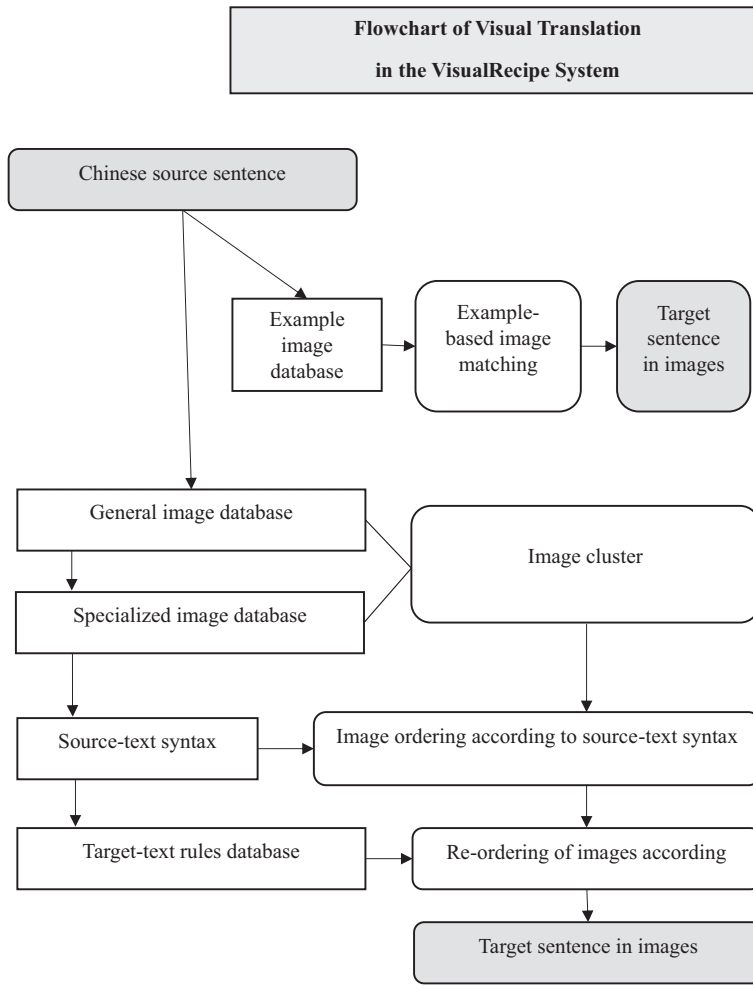

Figure 8.4 Flowchart of visual translation in the *VisualRecipe* system.

element on a screen without clicking it. This interaction enhances user experience by providing immediate feedback or additional details about items before any clicking is required, facilitating a more intuitive and informative navigation process.

In the case of preparing a dish from a recipe, mouse-over is a direct and straightforward way to show how a dish can be produced step by step, without referring to the text of a recipe. The following is a demonstration of a mouse-over procedure by moving the cursor to a sentence and a window showing the related action pops up on the screen.

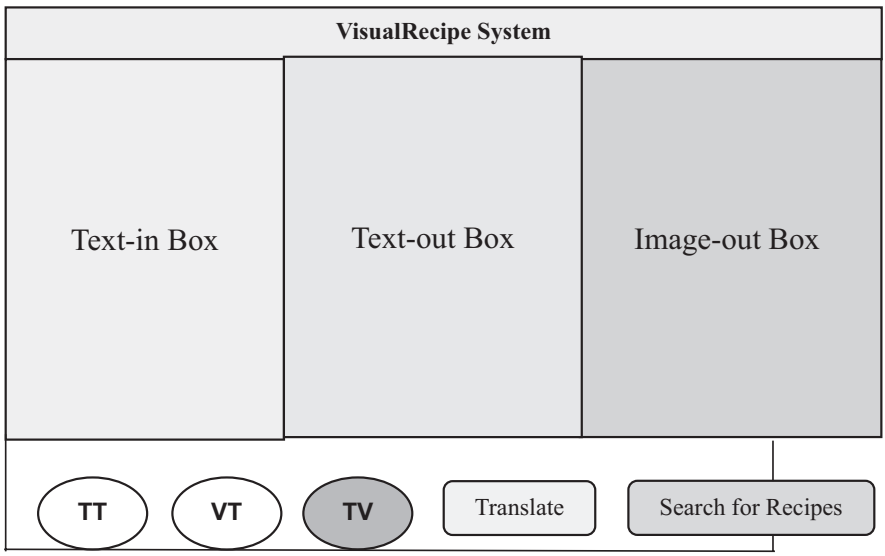

Figure 8.5 A proposed translation and *VisualRecipe* system.

〈乾炒牛河〉
將疊在一起的河粉分開。韭黃和葱，切段。
牛肉切片，加入醃料，醃約15分鐘。
雞蛋拂勻，把蛋漿淋在河粉上備用。
燒熱鑊下油，大火炒芽菜、韭黃及葱後盛起，再以大火煎牛肉至7成熟，盛起備用。
鑊再下油，轉中大火，以筷子炒河粉至微微焦香，加入生抽與老抽炒勻。
下芽菜、韭黃、葱與牛肉同炒，上碟後灑上芝麻即成。
"Stir-fried sliced beef with rice noodles in brown sauce"
Separate the rice noodles. Cut the garlic chives and green onions into sections. Slice the beef, add the marinade, and marinate for about 15 minutes.
Mix the egg white and yolk together, then pour them over the rice noodles and set aside.
Heat the wok and add oil. Stir-fry the bean sprouts, garlic chives, and green onions over high heat, then scoop them up. Next, fry the beef over high heat until medium well, then set it aside.

Add more oil to the wok, turn to medium-high heat, and use chopsticks to stir-fry the rice noodles until they are slightly crispy and fragrant. Then, add light soy sauce and dark soy sauce, and mix well.

Add the bean sprouts, garlic chives, green onions, and beef, and stir-fry together. Once plated, sprinkle with sesame seeds, and it's ready to serve.

Figure 8.6 Separating the rice noodles.

Figure 8.7 Slice the beef and add marinade.

Mouse-over Presentation of the *VisualRecipe* Output

將疊在一起的河粉分開。韭黃和葱，切段。
Separate the rice noodles. Cut the garlic chives and green onions into sections.
牛肉切片，加入醃料，醃約15分鐘。
Slice the beef, add the marinade, and marinate for about 15 minutes.
雞蛋拂勻，把蛋漿淋在河粉上備用。
Mix the egg white and yolk together, then pour them over the rice noodles and set aside.
燒熱鑊下油，大火炒芽菜、韭黃及葱後盛起，再以大火煎牛肉至7成熟，盛起備用。

Figure 8.8 Mix the egg and pour it over the rice noodles.

Figure 8.9 Stir-fry the ingredients and beef.

Heat the wok and add oil. Stir-fry the bean sprouts, garlic chives, and green onions over high heat, then scoop them up. Next, fry the beef over high heat until medium well done, then set it aside.
鑊再下油，轉中大火，以筷子炒河粉至微微焦香，加入生抽與老抽炒勻。
Add more oil to the wok, turn to medium-high heat, and use chopsticks to stir-fry the rice noodles until they are slightly crispy and fragrant. Then, add light soy sauce and dark soy sauce, and mix well.
下芽菜、韭黃、葱與牛肉同炒，上碟後灑上芝麻即成。

Figure 8.10 Stir-fry noodles and add soy sauce.

Figure 8.11 Stir-fry ingredients and beef and serve.

Add the bean sprouts, garlic chives, green onions, and beef, and stir-fry together. Once plated, sprinkle with sesame seeds, and it's ready to serve.

Concluding Remarks

In this chapter, we have proposed a textual-visual hybrid system in which Chinese recipes can be translated in words and presented in images, including photos and videos, to facilitate cooking a specific dish. Food photography itself has evolved into a significant aspect of food culture. The aesthetics

of food presentation have become as important as the taste and flavour. Professional food photographers use advanced techniques and equipment to capture the essence and appeal of the dish, making it look as inviting as possible. This not only influences public perception but also sets trends in the culinary world.

More important, nevertheless, is to notice the increasing use of image technology in recent years. Image recognition technology, for example, has started to change how people interact with recipes. Augmented reality (AR) and virtual reality (VR) are emerging technologies that take visualizing recipes to the next level. AR apps can project virtual images of recipes into the real world, providing users with an interactive experience. VR, on the other hand, can transport users to virtual cooking classes where they can learn cooking skills from chefs around the world as if they were in the same room.

The use of image technology in visualizing recipes has significant implications for accessibility. People with disabilities, such as those with visual impairments or learning difficulties, can benefit from tailored visual content that addresses their specific needs, making cooking more inclusive.

To conclude, the use of image technology in visualizing recipes is a dynamic and evolving field that has significantly influenced how recipes are shared, learned, and perfected. From enhancing the visual appeal of dishes and making cooking more accessible to fostering a global community of food enthusiasts, image technology continues to shape the culinary landscape in profound ways. As technology advances, it will undoubtedly continue to offer new and exciting ways to explore the art of cooking.

References

Batra, Devansh, Nirav Diwan, Utkarsh Upadhyay, Jushaan Singh Kalra, Tript Sharma, Aman Kumar Sharma, Dheeraj Khanna, Jaspreet Singh Marwah, Srilakshmi Kalathil, Navjot Singh *et al.* (2020) "Reciped: A Resource for Exploring Recipes", *Database*, 2020.

Bentivogli, Luisa, Arianna Bisazza, Mauro Cettolo, and Marcello Federico (2016) "Neural versus Phrase-based Machine Translation Quality: A Case Study", *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 257–267.

Bień, Michał, Michał Gilski, Martyna Maciejewska, Wojciech Taisner, Dawid Wisniewski, and Agnieszka Lawrynowicz (2020) "RecipeNLG: A Cooking Recipes Dataset for Semi-structured Text Generation", *Proceedings of the 13th International Conference on Natural Language Generation*, 22–28.

Borji, Ali (2022) "Generated Faces in the Wild: Quantitative Comparison of Stable Diffusion, Midjourney and Dall-e 2", *arXiv preprint arXiv:2210.00586*.

Bosch, Marc, Fengqing Zhu, Nitin Khanna, Carol J. Boushey, and Edward J. Delp (2011) "Combining Global and Local Features for Food Identification in Dietary Assessment", *2011 18th IEEE International Conference on Image Processing*, 1789–1792.

Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell *et al.* (2020) "Language Models are Few-shot Learners", *Advances in Neural Information Processing Systems*, 33: 1877–1901.

Catford, J.C. (1965) *A Linguistic Theory of Translation: An Essay in Applied Linguistics*, London: Oxford University Press.

Chan, Sin-wai (2002) "The Making of *TransRecipe*: A Translational Approach to the Machine Translation of Chinese Cookbooks", in Chan Sin-wai (ed.), *Translation and Information Technology*, Hong Kong: The Chinese University Press, 3–22.

Chan, Sin-wai (ed.) (2002) *Translation and Information Technology*, Hong Kong: The Chinese University Press.

Chandu, Khyathi, Eric Nyberg, and Alan W. Black (2019) "Storyboarding of Recipes: Grounded Contextual Generation", *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 6040–6046.

Chang, Minsuk, Léonore V. Guillain, Hyeungshik Jung, Vivian M. Hare, Juho Kim, and Maneesh Agrawala (2018) "Recipescape: An Interactive Tool for Analyzing

Cooking Instructions at Scale", *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–12.

Chen, Jing-jing and Chong-Wah Ngo (2016) "Deep-based Ingredient Recognition for Cooking Recipe Retrieval", *Proceedings of the 24th ACM International Conference on Multimedia*, 32–41.

Chen, Jing-Jing, Chong-Wah Ngo, Fu-Li Feng, and Tat-Seng Chua (2018) "Deep understanding of cooking procedure for cross-modal recipe retrieval" *Proceedings of the 26th ACM International Conference on Multimedia*, 1020–1028.

Chen, Jing-jing, Chong-Wah Ngo, and Tat-Seng Chua (2017) "Cross-modal Recipe Retrieval with Rich Food Attributes", *Proceedings of the 25th ACM International Conference on Multimedia*, 1771–1779.

Chen, Mei, Kapil Dhingra, Wen Wu, Lei Yang, Rahul Sukthankar, and Jie Yang (2009) "PFID: Pittsburgh Fast-food Image Dataset", *2009 16th IEEE International Conference on Image Processing (ICIP)*, 289–292.

Chen, Mei-Yun, Yung-Hsiang Yang, Chia-Ju Ho, Shih-Han Wang, Shane-Ming Liu, Eugene Chang, Che-Hua Yeh, and Ming Ouhyoung (2012) "Automatic Chinese Food Identification and Quantity Estimation", *SIGGRAPH Asia 2012 Technical Briefs*, 1–4.

Chen, Nicholas, Yun Young Lee, Maurice Rabb, and Bruce Schatz (2010) "Toward Dietary Assessment via Mobile Phone Video Cameras", *AMIA Annual Symposium Proceedings*, American Medical Informatics Association, 2010, 106.

Cho, Jaemin, Abhay Zala, and Mohit Bansal (2022) "DALL-Eval: Probing the Reasoning Skills and Social Biases of Text-to-image Generative Transformers", *arXiv preprint arXiv:2202.04053*.

Cho, Kyunghyun, Bart Van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio (2014) "Learning Phrase Representations Using RNN Encoder–decoder for Statistical Machine Translation", *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1724–1734.

Cho, Kyunghyun, Bart Van Merrienboer, Dzmitry Bahdanau, and Yoshua Bengio (2014) "On the Properties of Neural Machine Translation: Encoder-decoder Approaches", *arXiv preprint arXiv:1409.1259*.

Christiano, Paul F., Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei (2017) "Deep Reinforcement Learning from Human Preferences", *Advances in Neural Information Processing Systems*, 30.

Church, Kenneth W. and William A. Gale (1991) "Concordances for Parallel Texts", *Using Corpora: Proceedings of the 7th Annual Conference of the UW for the New OED and Text Research*, Oxford: Oxford University Press, 40–62.

Deeney, John J. (1995) "Transcription, Romanization, Transliteration", in Chan Sin-wai and David E. Pollard (eds.), *An Encyclopaedia of Translation: Chinese-English. English-Chinese*, Hong Kong: The Chinese University Press, 1085–1097.

DeFrancis, John (1996) *ABC Chinese-English Dictionary*, Hong Kong: The Chinese University Press.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2018) "Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding", *arXiv preprint arXiv:1810.04805*.

Fatemi, Bahare, Quentin Duval, Rohit Girdhar, Michal Drozdzal, and Adriana Romero-Soriano (2023) "Learning to Substitute Ingredients in Recipes", *arXiv preprint arXiv:2302.07960*.

Fei, Hongliang, Tan Yu, and Ping Li (2021) "Cross-lingual Cross-modal Pretraining for Multimodal Retrieval", *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 3644–3650.

Frome, Andrea, Greg S. Corrado, Jonathon Shlens, Samy Bengio, Jeffrey Dean, Marc'Aurelio Ranzato, and Tomas Mikolov (2013) "DeViSE: A Deep Visualsemantic Embedding Model", *Proceedings of the 26th International Conference on Neural Information Processing Systems-Volume 2*, 2121–2129.

Fu, Han, Rui Wu, Chenghao Liu, and Jianling Sun (2020) "Mcen: Bridging cross-modal gap between cooking recipes and dish images with latent variable model" *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,* 14570–14580.

Fujii, Tatsuki, Yuichi Sei, Yasuyuki Tahara, Ryohei Orihara, and Akihiko Ohsuga (2019) "'Never Fry Carrots without Chopping' Generating Cooking Recipes from Cooking Videos Using Deep Learning Considering Previous Process", *International Journal of Networked and Distributed Computing*, 7(3): 107–112.

Fung, Kam Ling 馮金陵, Lee Ngan Woon 李銀煥, and Hui Choi Yip 許彩葉 (1994) 《清爽涼拌》(*Cold Dishes*), Hong Kong: Food Paradise Publishing Co. 飲食天地出版社.

Gabeur, Valentin, Arsha Nagrani, Chen Sun, Karteek Alahari, and Cordelia Schmid (2022) "Masking Modalities for Cross-modal Video Retrieval", *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1766–1775.

Galatolo, Federico A., Mario G.C.A. Cimino, and Edoardo Cogotti (2022) "TeTImEval: A Novel Curated Evaluation Data Set for Comparing Text-to-image Models", *arXiv preprint arXiv:2212.07839*.

Gatt, Albert and Emiel Krahmer (2018) "Survey of the State of the Art in Natural Language Generation: Core Tasks, Applications and Evaluation", *Journal of Artificial Intelligence Research*, 61: 65–170.

Ghoddoosian, Reza, Saif Sayed, and Vassilis Athitsos (2022) "Hierarchical Modeling for Task Recognition and Action Segmentation in Weakly-labeled Instructional Videos", *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1922–1932.

Goel, Mansi, Pallab Chakraborty, Vijay Ponnaganti, Minnet Khan, Sritanaya Tatipamala, Aakanksha Saini, and Ganesh Bagler (2022) "Ratatouille: A Tool for Novel Recipe Generation", *2022 IEEE 38th International Conference on Data Engineering Workshops (ICDEW)*, 107–110.

Goodfellow, Ian J., Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David WardeFarley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio (2014) "Generative Adversarial Networks", *arXiv preprint arXiv:1406.2661*.

Gove, Philip Babcock (comp.) (1986) *Webster's Third New International Dictionary*, Springfield, Mass.: Merriam-Webster Inc.

Guerrero, Ricardo, Hai X. Pham, and Vladimir Pavlovic (2021) "Cross-modal Retrieval and Synthesis (X-MRS): Closing the Modality Gap in Shared Subspace Learning", *Proceedings of the 29th ACM International Conference on Multimedia*, 3192–3201.

Guo, Zhao, Lianli Gao, Jingkuan Song, Xing Xu, Jie Shao, and Heng Tao Shen (2016) "Attention-based LSTM with Semantic Consistency for Videos Captioning", *Proceedings of the 24th ACM International Conference on Multimedia*, 357–361.

Han, Fangda, Guoyao Hao, Ricardo Guerrero, and Vladimir Pavlovic (2021) "Multi-attribute Pizza Generator: Cross-domain Attribute Control with Conditional StyleGAN", *arXiv preprint arXiv:2110.11830*.

Harashima, Jun, Yuichiro Someya, and Yohei Kikuta (2017) "Cookpad Image Dataset: An Image Collection as Infrastructure for Food Research", *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1229–1232.

Hasyim, Muhammad, Firman Saleh, Rudy Yusuf, and Asriani Abbas (2021) "Artificial Intelligence: Machine Translation Accuracy in Translating FrenchIndonesian Culinary Texts", *International Journal of Advanced Computer Science and Applications*, 12 (3).

Haynes, Colin (1998) *Breaking down the Language Barriers*, London: Aslib.

Hearst, Marti A., Susan T. Dumais, Edgar Osuna, John Platt, and Bernhard Scholkopf (1998) "Support Vector Machines", *IEEE Intelligent Systems and Their Applications*, 13(4): 18–28.

Hertz, Amir, Ron Mokady, Jay Tenenbaum, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or (2022) "Prompt-to-prompt Image Editing with Cross Attention Control", *arXiv preprint arXiv:2208.01626*.

Heusel, Martin, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter (2017) "GANs Trained by a Two Time-scale Update Rule Converge to a Local Nash Equilibrium", *Advances in Neural Information Processing Systems*, 30.

Ho, Jonathan, Ajay Jain, and Pieter Abbeel (2020) "Denoising Diffusion Probabilistic Models", *Advances in Neural Information Processing Systems*, 33: 6840–6851.

Hochreiter, Sepp (1998) "The Vanishing Gradient Problem during Learning Recurrent Neural Nets and Problem Solutions", *International Journal of Uncertainty, Fuzziness and Knowledge-based Systems*, 6(2): 107–116.

Hochreiter, Sepp and Jürgen Schmidhuber (1997) "Long Short-term Memory", *Neural Computation*, 9(8): 1735–1780.

Horita, Daichi, Ryosuke Tanno, Wataru Shimoda, and Keiji Yanai (2018) "Food Category Transfer with Conditional Cyclegan and a Large-scale Food Image Dataset", *Proceedings of the Joint Workshop on Multimedia for Cooking and Eating Activities and Multimedia Assisted Dietary Management*, 67–70.

Hu, Fu 胡附(1984)《數詞和量詞》(*Numerals and Measure-words*), Shanghai: Shanghai Education Press 上海教育出版社.

Hu, Yuheng, Lydia Manikonda, and Subbarao Kambhampati (2014) "What We Instagram: A First Analysis of Instagram Photo Content and User Types", *Proceedings of the International AAAI Conference on Web and Social Media*, 8: 595–598.

Huang, Po-Yao, Xiaojun Chang, Alexander Hauptmann, and Eduard Hovy (2020) "Forward and Backward Multimodal NMT for Improved Monolingual and Multilingual Cross-modal Retrieval", *Proceedings of the 2020 International Conference on Multimedia Retrieval*, 53–62.

Ifrah, Georges (2000) *The Universal History of Numbers: From Prehistory to the Invention of the Computer*, tr. David Bellos, E.F. Harding, Sophie Wood, and Ian Monk, New York: Wiley.

Isola, Phillip, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros (2017) "Image-to-image Translation with Conditional Adversarial Networks", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1125–1134.

Ji, Lei, Chenfei Wu, Daisy Zhou, Kun Yan, Edward Cui, Xilin Chen, and Nan Duan (2022) "Learning Temporal Video Procedure Segmentation from an Automatically

Collected Large Dataset", *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1506–1515.

Kagaya, Hokuto, Kiyoharu Aizawa, and Makoto Ogawa (2014) "Food Detection and Recognition Using Convolutional Neural Network", *Proceedings of the 22nd ACM International Conference on Multimedia*, 1085–1088.

Kawano, Yoshiyuki and Keiji Yanai (2014) "Foodcam: A Real-time Mobile Food Recognition System Employing Fisher Vector", *MultiMedia Modeling: 20th Anniversary International Conference, MMM 2014, Dublin, Ireland, January 6-10, 2014, Proceedings,* Part II 20, 369–373.

Kawar, Bahjat, Shiran Zada, Oran Lang, Omer Tov, Huiwen Chang, Tali Dekel, Inbar Mosseri, and Michal Irani (2022) "Imagic: Text-based Real Image Editing with Diffusion Models", *arXiv preprint arXiv:2210.09276*.

Khan, Rijwan, Santosh Kumar, Niharika Dhingra, and Neha Bhati (2021) "The Use of Different Image Recognition Techniques in Food Safety: A Study", *Journal of Food Quality*, 1–10.

Kiddon, Chloé, Luke Zettlemoyer, and Yejin Choi (2016) "Globally Coherent Text Generation with Neural Checklist Models", *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 329–339.

Kitamura, Keigo, Toshihiko Yamasaki, and Kiyoharu Aizawa (2008) "Food Log by Analyzing Food Images", *Proceedings of the 16th ACM International Conference on Multimedia*, 999–1000.

Kitamura, Keigo, Toshihiko Yamasaki, and Kiyoharu Aizawa (2009) "Foodlog: Capture, Analysis and Retrieval of Personal Food Images via Web", *Proceedings of the ACM Multimedia 2009 Workshop on Multimedia for Cooking and Eating Activities*, 23–30.

Koehn, Philipp and Rebecca Knowles (2017) "Six Challenges for Neural Machine Translation", *Proceedings of the First Workshop on Neural Machine Translation*, 28–39.

Koehn, Philipp, Franz Josef Och, and Daniel Marcu (2003) "Statistical Phrase-based Translation", *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, 127–133.

Lee, Helena, Ke Shu, Palakorn Achananuparp, Philips Kokoh Prasetyo, Yue Liu, Ee-Peng Lim, and Lav R. Varshney (2020) "RecipeGPT: Generative Pretraining Based Cooking Recipe Generation and Evaluation System", *Companion Proceedings of the Web Conference 2020*, 181–184.

Lee, Ngan Woon 李銀煥 (1996) 《家常老幼靚湯》(*Soup for the Whole Family*), Hong Kong: Sea Shore Book Company 海濱圖書公司.

Li, Guang, Shubo Ma, and Yahong Han (2015) "Summarization-based Video Caption via Deep Neural Networks", *Proceedings of the 23rd ACM International Conference on Multimedia*, 1191–1194.

Li, Jiatong, Fangda Han, Ricardo Guerrero, and Vladimir Pavlovic (2021) "Picture-to-amount (pita): Predicting Relative Ingredient Amounts from Food Images", *2020 25th International Conference on Pattern Recognition (ICPR)*, 10343–10350.

Liu, Chang, Yu Cao, Yan Luo, Guanling Chen, Vinod Vokkarane, and Yunsheng Ma (2016) "Deepfood: Deep Learning-based Food Image Recognition for Computer-aided Dietary Assessment", *Inclusive Smart Cities and Digital Health: 14th International Conference on Smart Homes and Health Telematics, ICOST 2016, Wuhan, China, May 25-27, 2016. Proceedings 14*, 37–48.

Liu, Jiatong (2021) "Multimodal Machine Translation", *IEEE Access*.

Liu, Vivian and Lydia B. Chilton (2022) "Design Guidelines for Prompt Engineering Text-to-image Generative Models", *CHI Conference on Human Factors in Computing Systems*, 1–23.

Liu, Xiao, Yansong Feng, Jizhi Tang, Chengang Hu, and Dongyan Zhao (2022) "Counterfactual Recipe Generation: Exploring Compositional Generalization in a Realistic Scenario", *arXiv preprint arXiv:2210.11431*.

Liu, Zhiming, Kai Niu, and Zhiqiang He (2023) "ML-CookGAN: Multi-label Generative Adversarial Network for Food Image Generation", *ACM Transactions on Multimedia Computing, Communications and Applications*, 19(2s): 1–21.

Majumder, Bodhisattwa Prasad, Shuyang Li, Jianmo Ni, and Julian McAuley (2019) "Generating Personalized Recipes from Historical User Preferences", *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 5976–5982.

Malmaud, Jonathan, Jonathan Huang, Vivek Rathod, Nicholas Johnston, Andrew Rabinovich, and Kevin Murphy (2015) "'What's Cooking'? Interpreting Cooking Videos using Text, Speech, and Vision" *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 143–152.

Marin, Javier, Aritro Biswas, Ferda Ofli, Nicholas Hynes, Amaia Salvador, Yusuf Aytar, Ingmar Weber, and Antonio Torralba (2019) "Recipe1m+: A Dataset for Learning Cross-modal Embeddings for Cooking Recipes and Food Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(1): 187–203.

Metz, Cade and Priya Krishna (2022) "Can AI Write Recipes Better Than Humans?" *The Seattle Times* www.seattletimes.com/business/can-ai-write-recipes-better-than-humans/, retrieved on 17 April 2023.

Min, Weiqing, Shuqiang Jiang, Jitao Sang, Huayang Wang, Xinda Liu, and Luis Herranz (2016) "Being a Supercook: Joint Food Attributes and Multimodal Content Modeling for Recipe Retrieval and Exploration", *IEEE Transactions on Multimedia*, 19(5): 1100–1113.

Mirza, Mehdi and Simon Osindero (2014) "Conditional Generative Adversarial Nets", *arXiv preprint arXiv:1411.1784*.

Mohammadshahi, Alireza, Rémi Lebret, and Karl Aberer (2019) "Aligning Multilingual Word Embeddings for Cross-modal Retrieval Task", *Proceedings of the Beyond Vision and LANguage: inTEgrating Real-world kNowledge (LANTERN)*, 11–17.

Mokady, Ron, Amir Hertz, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or (2022) "Null-text Inversion for Editing Real Images Using Guided Diffusion Models", *arXiv preprint arXiv:2211.09794*.

Mori, Shinsuke, Hirokuni Maeta, Tetsuro Sasada, Koichiro Yoshino, Atsushi Hashimoto, Takuya Funatomi, and Yoko Yamakata (2014) "Flowgraph2text: Automatic Sentence Skeleton Compilation for Procedural Text Generation", *Proceedings of the 8th International Natural Language Generation Conference (INLG)*, 118–122.

Mori, Shinsuke, Hirokuni Maeta, Yoko Yamakata, and Tetsuro Sasada (2014) "Flow Graph Corpus from Recipe Texts", *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, 2370–2377.

Nagao, Makoto (1984) "A Framework of a Mechanical Translation between Japanese and English by Analogy", in Alick Elithorn and Ranan Barneiji (eds.), *Artificial and Human Intelligence*, Amsterdam: North-Holland Publishing Company, 73–80.

Newmark, Peter (1981) *Approaches to Translation*, Oxford: Pergamon Press.

Newmark, Peter (1988) *A Textbook of Translation*, Hertfordshire: Prentice-Hall.

Nichol, Alexander Quinn, Prafulla Dhariwal, Aditya Ramesh, Pranav Shyam, Pamela Mishkin, Bob Mcgrew, Ilya Sutskever, and Mark Chen (2022) "GLIDE: Towards

Photorealistic Image Generation and Editing with TextGuided Diffusion Models", *International Conference on Machine Learning*, 16784–16804.

Nishimura, Taichi, Atsushi Hashimoto, and Shinsuke Mori (2019) "Procedural Text Generation from a Photo Sequence", *Proceedings of the 12th International Conference on Natural Language Generation*, 409–414.

OpenAI (2022) "Introducing ChatGPT", accessed on 26 April 2023, available at https://openai.com/blog/chatgpt.

Pan, Liang-Ming, Jingjing Chen, Jianlong Wu, Shaoteng Liu, Chong-Wah Ngo, Min-Yen Kan, Yugang Jiang, and Tat-Seng Chua (2020) "Multi-modal Cooking Workflow Construction for Food Recipes", *Proceedings of the 28th ACM International Conference on Multimedia*, 1132–1141.

Papadopoulos, Dim P., Youssef Tamaazousti, Ferda Ofli, Ingmar Weber, and Antonio Torralba (2019) "How to Make a Pizza: Learning a Compositional Layer-based Gan Model", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8002–8011.

Papineni, Kishore, Salim Roukos, Todd Ward, and Wei-Jing Zhu (2002) "Bleu: A Method for Automatic Evaluation of Machine Translation", *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 311–318.

Pavlichenko, Nikita, Fedor Zhdanov, and Dmitry Ustalov (2022) "Best Prompts for Text-to-image Models and How to Find Them", *arXiv preprint arXiv:2209.11711.*

Petsiuk, Vitali, Alexander E. Siemenn, Saisamrit Surbehera, Zad Chin, Keith Tyser, Gregory Hunter, Arvind Raghavan, Yann Hicke, Bryan A. Plummer, Ori Kerret *et al.* (2022) "Human Evaluation of Text-to-image Models on a Multi-task Benchmark", *arXiv preprint arXiv:2211.12112.*

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik (1985) *A Comprehensive Grammar of the English Language*, New York: Longman Group Ltd.

Radford, Alec, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever *et al.* (2019) "Language Models Are Unsupervised Multitask Learners", *OpenAI Blog*, 1(8): 9.

Radford, Alec, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark *et al.* (2021) "Learning Transferable Visual Models from Natural Language Supervision", *International Conference on Machine Learning*, 8748–8763.

Ramesh, Aditya, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen (2022) "Hierarchical Text-conditional Image Generation with Clip Latents", *arXiv preprint arXiv:2204.06125.*

Rombach, Robin, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer (2022) "High-resolution Image Synthesis with Latent Diffusion Models", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10684–10695.

Sagart, Laurent (1999) *The Roots of Old Chinese*, Amsterdam and Philadelphia: John Benjamins Publishing Company.

Sahoo, Doyen, Wang Hao, Shu Ke, Wu Xiongwei, Hung Le, Palakorn Achananuparp, Ee-Peng Lim, and Steven C.H. Hoi (2019) "FoodAI: Food Image Recognition via Deep Learning for Smart Food Logging", *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2260–2268.

Salvador, Amaia, Michal Drozdzal, Xavier Giro-i Nieto, and Adriana Romero (2019) "Inverse Cooking: Recipe Generation from Food Images", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10453–10462.

Salvador, Amaia, Erhan Gundogdu, Loris Bazzani, and Michael Donoser (2021) "Revamping Cross-modal Recipe Retrieval with Hierarchical Transformers and Self-supervised Learning", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 15475–15484.

Salvador, Amaia, Nicholas Hynes, Yusuf Aytar, Javier Marin, Ferda Ofli, Ingmar Weber, and Antonio Torralba (2017) "Learning Cross-modal Embeddings for Cooking Recipes and Food Images", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3020–3028.

Sato, Takayuki, Jun Harashima, and Mamoru Komachi (2016) "Japanese-English Machine Translation of Recipe Texts", *Proceedings of the 3rd Workshop on Asian Translation (WAT2016)*, 58–67.

Schäffner, Christina and Helen Kelly-Holmes (1995) *Cultural Functions of Translation*, Clevedon: Multilingual Matters.

Sennrich, Rico, Barry Haddow, and Alexandra Birch (2016) "Improving Neural Machine Translation Models with Monolingual Data", *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 86–96.

Seo, Paul Hongsuck, Arsha Nagrani, Anurag Arnab, and Cordelia Schmid (2022) "End-to-end Generative Pretraining for Multimodal Video Captioning", *arXiv preprint arXiv:2201.08264*.

Simoons, Frederick J. (1991) *Food in China: A Cultural and Historical Inquiry*, Boca Raton: CRC Press.

Sohl-Dickstein, Jascha, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli (2015) "Deep Unsupervised Learning Using Nonequilibrium Thermodynamics", *International Conference on Machine Learning*, 2256–2265.

Song, Yang, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole (2020) "Score-based Generative Modeling through Stochastic Differential Equations", *arXiv preprint arXiv:2011.13456*.

Su, Han, Ting-Wei Lin, Cheng-Te Li, Man-Kwan Shan, and Janet Chang (2014) "Automatic Recipe Cuisine Classification by Ingredients", *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, 565–570.

Sugiyama, Yu and Keiji Yanai (2021) "Cross-modal Recipe Embeddings by Disentangling Recipe Contents and Dish Styles", *Proceedings of the 29th ACM International Conference on Multimedia*, 2501–2509.

Sulubacak, Umut, Ozan Caglayan, Stig-Arne Grönroos, Aku Rouhe, Desmond Elliott, Lucia Specia, and Jörg Tiedemann (2020) "Multimodal Machine Translation through Visuals and Speech", *Machine Translation*, 34: 97–147.

Sun, Chen, Austin Myers, Carl Vondrick, Kevin Murphy, and Cordelia Schmid (2019) "VideoBERT: A Joint Model for Video and Language Representation Learning", *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 7464–7473.

Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le (2014) "Sequence to Sequence Learning with Neural Networks", *Advances in Neural Information Processing Systems*, 27: 3104–3112.

Tai, Kuo-Chung (1979) "The Tree-to-tree Correction Problem", *Journal of the ACM (JACM)*, 26(3): 422–433.

Tam, Ng Wai Fung 譚吳威鳳 (1997) 《超值營養食譜》(*Healthy Money-saving Meals*), Hong Kong: Wan Li Book Co. Ltd 萬里機構and Food Paradise Publishing Co. 飲食天地出版社.

Tanno, Ryosuke, Daichi Horita, Wataru Shimoda, and Keiji Yanai (2018) "Magical Rice Bowl: A Real-time Food Category Changer", *Proceedings of the 26th ACM International Conference on Multimedia*, 1244–1246.

Teng, Chun-Yuen, Yu-Ru Lin, and Lada A. Adamic (2012) "Recipe Recommendation Using Ingredient Networks", *Proceedings of the 4th Annual ACM Web Science Conference*, 298–307.

Treffry, Diana (ed.) (1998) Collins English Dictionary, 4th edition. Glasgow: HarperCollins Publishers.

Ueda, Mayumi, Mari Takahata, and Shinsuke Nakajima (2011) "Recipe Recommendation Method Based on User's Food Preferences", *Proceedings of the IADIS International Conference on E-Society*, 591–594.

Van Asbroeck, Stephanie and Christophe Matthys (2020) "Use of Different Food Image Recognition Platforms in Dietary Assessment: Comparison Study", *JMIR Formative Research*, 4(12): e15602.

Varshney, Lav R., Florian Pinel, Kush R. Varshney, Debarun Bhattacharjya, Angela Schörgendorfer, and Y-M Chee (2019) "A Big Data Approach to Computational Creativity: The Curious Case of Chef Watson", *IBM Journal of Research and Development*, 63(1): 7–11.

Vasconcellos, Muriel (1996) *Recent Trends in Machine Translation*, London: Aslib.

Venuti, Lawrence (1995) *The Translator's Invisibility: A History of Translation*, London and New York: Routledge.

Vinay, Jean-Paul and Jean Darbelnet (1954, 1995) *Comparative Stylistics of French and English: A Methodology for Translation*, Juan C. Sager and M.-J. Hamel (tr. and ed.), Amsterdam and Philadelphia: John Benjamins Publishing Company.

Wang, Hao, Doyen Sahoo, Chenghao Liu, Ee-peng Lim, and Steven C.H. Hoi (2019) "Learning Cross-modal Embeddings with Adversarial Networks for Cooking Recipes and Food Images", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11572–11581.

Wang, Liping, Qing Li, Na Li, Guozhu Dong, and Yu Yang (2008) "Substructure Similarity Measurement in Chinese Recipes", *Proceedings of the 17th International Conference on World Wide Web*, 979–988.

Wang, Su, Honghao Gao, Yonghua Zhu, Weilin Zhang, and Yihai Chen (2019) "A Food Dish Image Generation Framework Based on Progressive Growing GANs", *Collaborative Computing: Networking, Applications and Worksharing: 15th EAI International Conference, CollaborateCom 2019, London, UK, August 19-22, 2019, Proceedings 15*, 323–333.

Wang, Wenjie, Ling-Yu Duan, Hao Jiang, Peiguang Jing, Xuemeng Song, and Liqiang Nie (2021) "Market2Dish: Health-aware Food Recommendation", *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 17(1): 1–19.

Weellisch, Hans (1975) *Transcription and Transliteration*, Silver Spring, MD: Institute of Modern Language.

Wikitionary (CC BY-SA 3.0) https://en.wiktionary.org/wiki/recipe accessed: 2023-04-27

Xie, Haoran, Lijuan Yu, and Qing Li (2010) "A Hybrid Semantic Item Model for Recipe Search by Example", *2010 IEEE International Symposium on Multimedia*, 254–259.

Xie, Zhongwei, Ling Liu, Lin Li, and Luo Zhong (2021) "Learning Joint Embedding with Modality Alignments for Cross-modal Retrieval of Recipes and Food Images", *Proceedings of the 30th ACM International Conference on Information and Knowledge Management*, 2221–2230.

Xu, Frank F., Lei Ji, Botian Shi, Junyi Du, Graham Neubig, Yonatan Bisk, and Nan Duan (2020) "A Benchmark for Structured Procedural Knowledge Extraction from Cooking Videos", *Proceedings of the First International Workshop on Natural Language Processing Beyond Text*, 30–40.

Yam, Lisa方任利莎 (1997)《方太食譜之魚蝦蟹》(*Lisa Yam's Cook Book: Seafood*), Hong Kong: Ming Pao Press Ltd. 明窗出版社.

Yamakata, Yoko, Shinji Imahori, Hirokuni Maeta, and Shinsuke Mori (2016) "A Method for Extracting Major Workflow Composed of Ingredients, Tools, and Actions from Cooking Procedural Text", *2016 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, 1–6.

Yamamoto, Kohei and Keiji Yanai (2022) "Text-based Image Editing for Food Images with CLIP", *Proceedings of the 7th International Workshop on Multimedia Assisted Dietary Management*, 29–37.

Yanai, Keiji, Kaimu Okamoto, Tetsuya Nagano, and Daichi Horita (2019) "Large-scale Twitter Food Photo Mining and its Applications", *2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM)*, 77–85.

Yang, Shulin, Mei Chen, Dean Pomerleau, and Rahul Sukthankar (2010) "Food Recognition Using Statistics of Pairwise Local Features", *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2249–2256.

Yu, Zhiwei, Hongyu Zang, and Xiaojun Wan (2020) "Routing Enforced Generative Model for Recipe Generation", *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 3797–3806.

Zhang, Deqin 張德鑫 (1999) 《數裏乾坤》(*The Mysterious World of Numerals*), Beijing: Peking University Press 北京大學出版社.

Zhang, Yixin, Yoko Yamakata, and Keishi Tajima (2019) "Categorization of Cooking Actions Based on Textual/Visual Similarity", *Proceedings of the 5th International Workshop on Multimedia Assisted Dietary Management*, 42–49.

Zhou, Kaiyang, Jingkang Yang, Chen Change Loy, and Ziwei Liu (2022) "Learning to Prompt for Vision-language Models", *International Journal of Computer Vision*, 130(9): 2337–2348.

Zhou, Luowei, Chenliang Xu, and Jason J. Corso (2018) "Towards Automatic Learning of Procedures from Web Instructional Videos", *Thirty-second AAAI Conference on Artificial Intelligence*.

Zhu, Bin, Chong-Wah Ngo, and Wing-Kwong Chan (2021) "Learning from Web Recipe-image Pairs for Food Recognition: Problem, Baselines and Performance", *IEEE Transactions on Multimedia*, 24: 1175–1185.

Zhu, Bin and Chong-Wah Ngo (2020) "CookGAN: Causality based text-to-image synthesis", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5519–5527.

Zhu, Bin, Chong-Wah Ngo, Jingjing Chen, and Yanbin Hao (2019) "R2gan: Cross-modal Recipe Retrieval with Generative Adversarial Network", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11477–11486.

Zhu, Jun-Yan, Taesung Park, Phillip Isola, and Alexei A Efros (2017) "Unpaired Image-to-image Translation Using Cycle-consistent Adversarial Networks", *Proceedings of the IEEE International Conference on Computer Vision*, 2223–2232.