Enhancing Neural Machine Translation with Image-based Paraphrase Augmentation

Johanes Effendi[1], Sakriani Sakti[1,2], Katsuhito Sudoh[1,2], Satoshi Nakamura[1,2]
{johanes.effendi.ix4,ssakti,sudoh,s-nakamura}@is.naist.jp

[1]Nara Institute of Science and Technology
[2]RIKEN, Center for Advanced Intelligence Project AIP

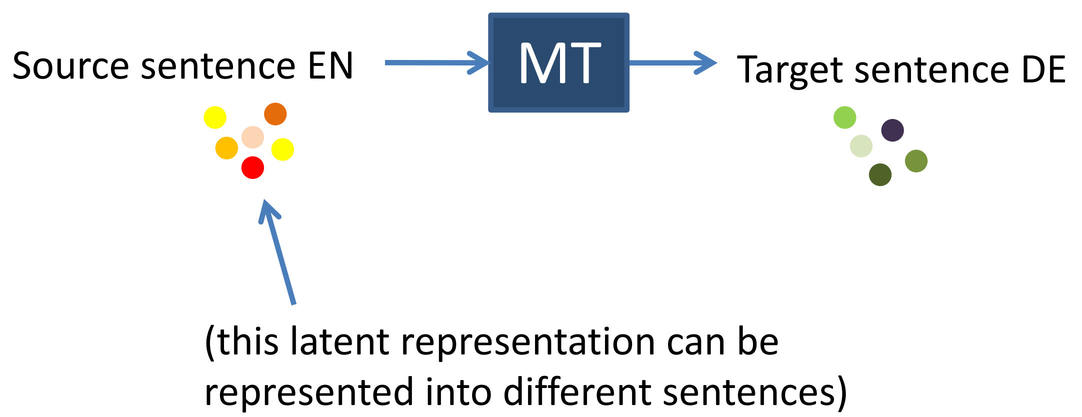
Outline

- Introduction
- Image-based Paraphrasing
- Proposed Idea
- Corpus Creation
- Experimental Settings
- Experiment Results
- Conclusion and Future Works

Introduction

Machine Translation

- ~~Text-to-text translation~~
- Parallel text dataset
- What about similar sentences?
- Concept-to-concept translation
  - Mapping latent representation into another latent representation



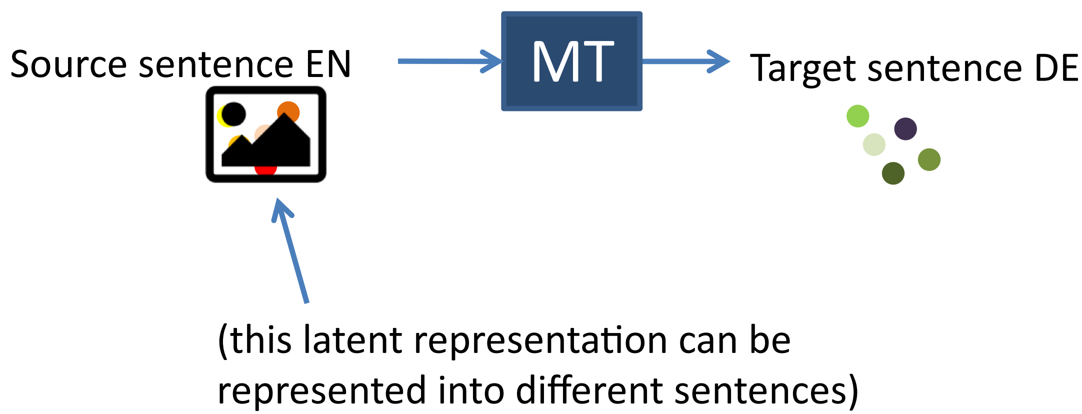
Multiple sources or references

- Multiple sources into one target

- Multiple references

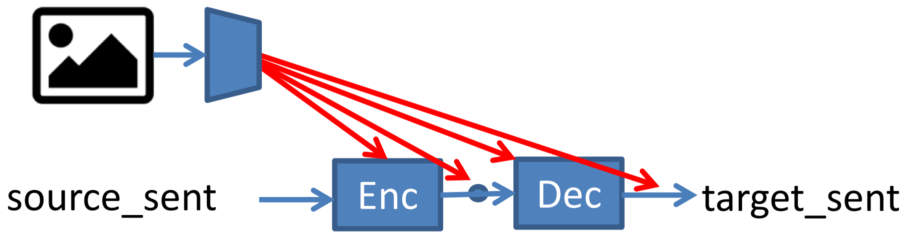
Multimodal NMT

- WMT17 Multimodal Translation Task
  - Translate a caption with the image provided
- Based on concept-to-concept idea:



Multimodal NMT (cont.)

- Common approach:
  - Incorporate latent image representation in various NMT components
    - Caglayan et al. (2016,2017), Calixto et al. (2017)



Multimodal NMT (cont.)

- Zhang et al. (2017) integrated similar image information as additional input

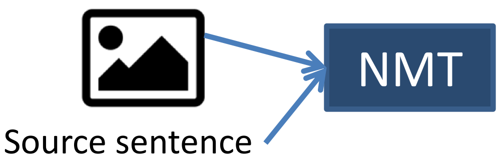
Difficulties with Multimodal NMT

- Powerful, but complicated
- The image encoder (VGG, ResNet) are resource intensive
- Difficulties combining latent spaces from different modalities
  - Not all information is useful for translation
- Improvement reached might not be as rewarding as the effort

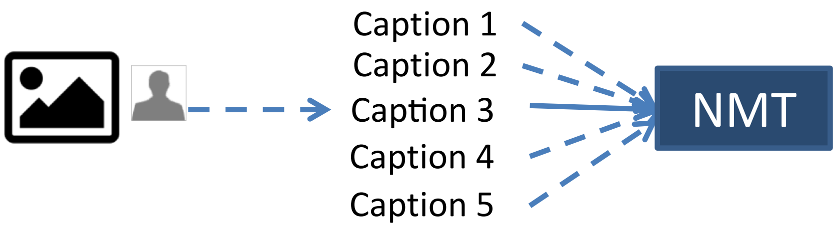
Image-based Paraphrasing

Image-based Paraphrasing

- Represent image as texts



Common approach



Our proposed approach

- **Image-based paraphrase**
- Rewrite source sentence with image as basis of paraphrasing
- Enable multi-source information in NMT

Difference with common MT paraphrases

- Paraphrasing to elaborate source language data
- Augment the dataset size in SMT
  - (Nichols et al., 2010, He et al., 2011)
- **Recent work:** only reordering and substitution are used

- **In this work:** with image as the basis of paraphrasing, deletion and insertion of information is possible

How to generate paraphrase from image?

- If random paraphrase is inputted, it might become noisy to each other
  - How many variations?
- Bhagat and Hovy (2013) studied on how many paraphrase operations language can possibly make
  - 25 classes of quasi-paraphrases
  - Survey the occurrence of each classes in Microsoft Research Paraphrase Corpus (MSR Corpus)

Classes of Quasi Paraphrases - Frequency

No	name	%Freq in MSR
1	Synonym substitution	19
2	Antonym substitution	0
3	Converse substitution	0
4	Change of voice	1
5	Change of person	1
6	Pronoun/Co-referent substitution	1
7	Repetition/Ellipsis	4
8	Function word variations	30
9	Actor/Action Substitution	0
10	Verb/Semantic-role noun substitution	0
11	Manipulator/Device substitution	0
12	General/Specific substitution	3
13	Metaphor substitution	1

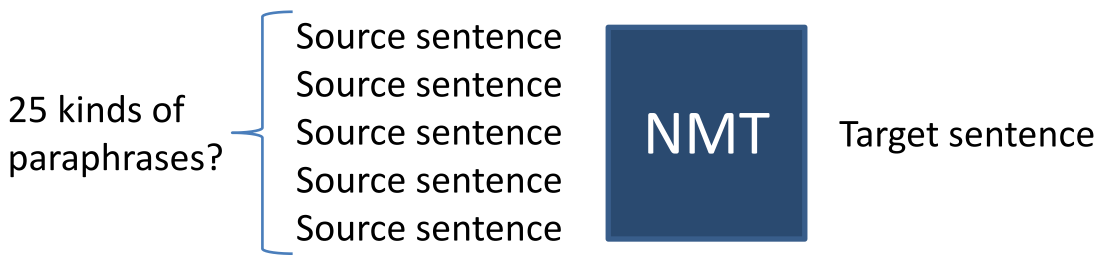
No	Name	%Freq in MSR
14	Part/Whole substitution	0
15	Verb/Noun conversion	3
16	Verb/Adjective conversion	0
17	Verb/Adverb conversion	0
18	Noun/Adjective conversion	0
19	Verb-preposition/Noun substitution	0
20	Change of tense	1
21	Change of aspect	0
22	Change of modality	0
23	Semantic implication	4
24	Approximate numerical equivalences	2
25	External knowledge	32

Some quasi-paraphrases have low frequency in MSR Corpus

Bhagat, R., & Hovy, E. (2013). What Is a Paraphrase? Computational Linguistics, 39(3), 463-472.

Simplify into four elements

- Some quasi-paraphrase classes:
  - have low frequencies
  - are too fine-grained
- Having 25 kinds of input sentences might be too difficult



Simplify into four elements (cont.)

- We grouped it into four elementary operations:
  - Deletion
  - Insertion
  - Reordering
  - Substitution

Simplify

Original
Deletion
Insertion
Reordering
Substitution

NMT

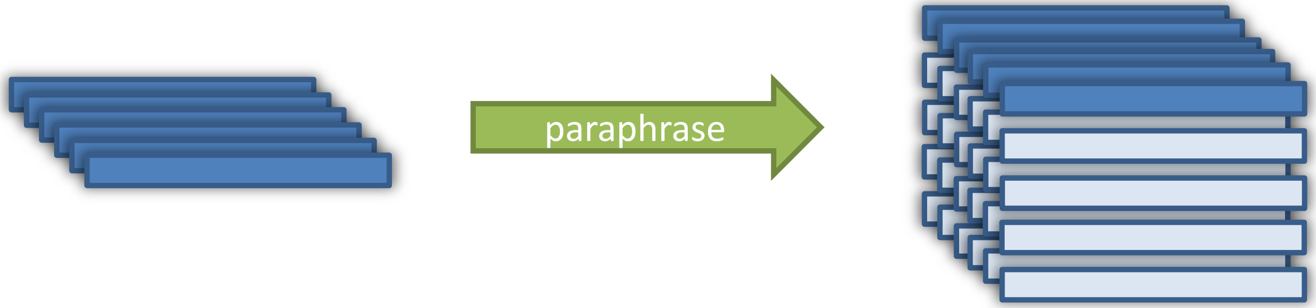
Target sentence

- Each source sentence now paraphrased into four paraphrase

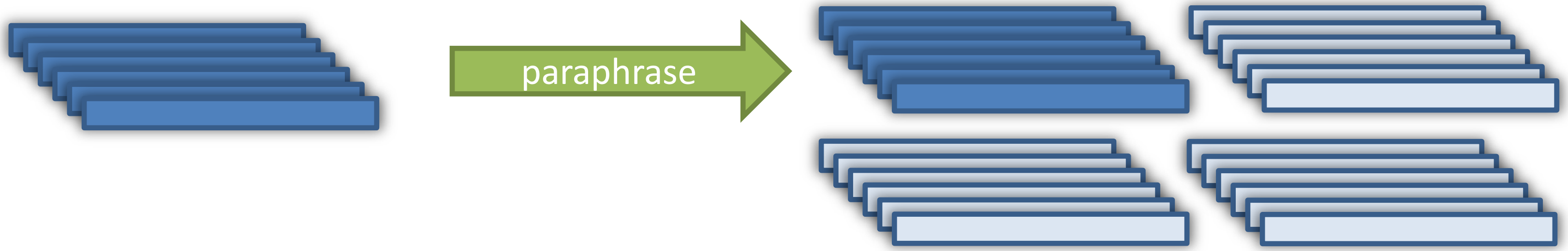
Proposed Idea

Two Possibilities on Data Usage

- Several paraphrase as input enables two scenario:
  - data augmentation
  - multi-source
- Simple data augmentation == combining all data

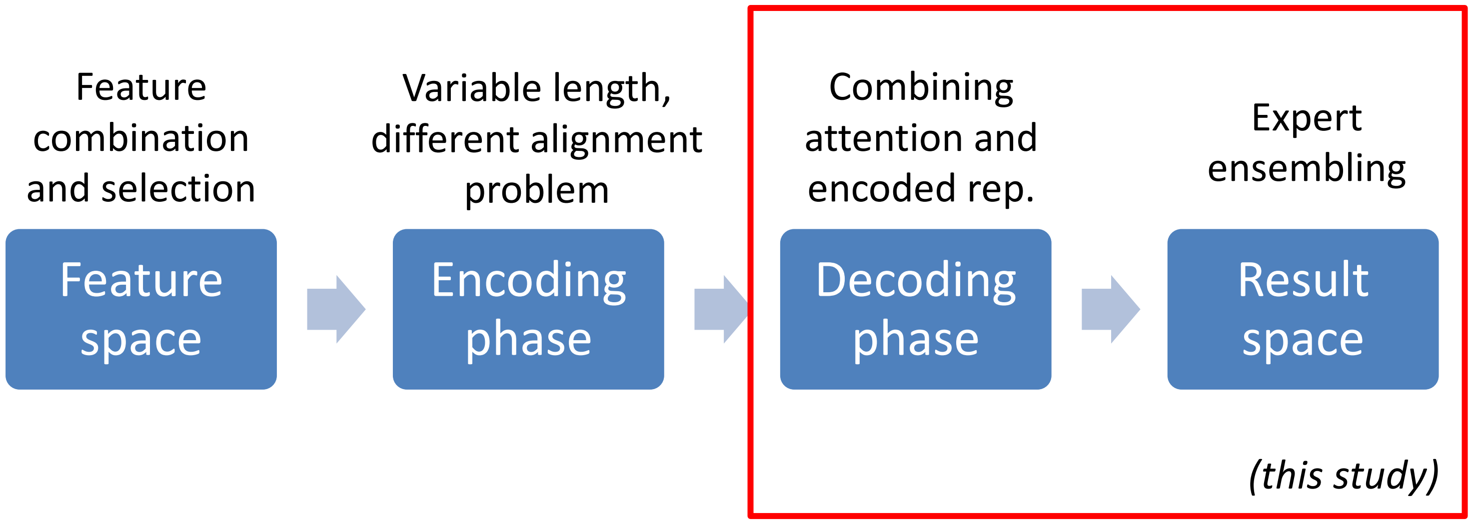


- Multi-source: separate dataset per paraphrase operation



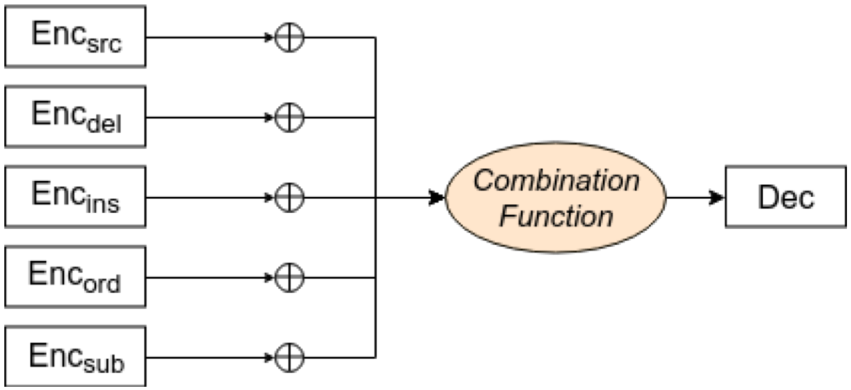
Determining Integration Point

- Multi-source combination:
  - preserves relation between paraphrases
  - on which NMT stage?
- Decoding phase and result space for this work
- Other phase is omitted for further study



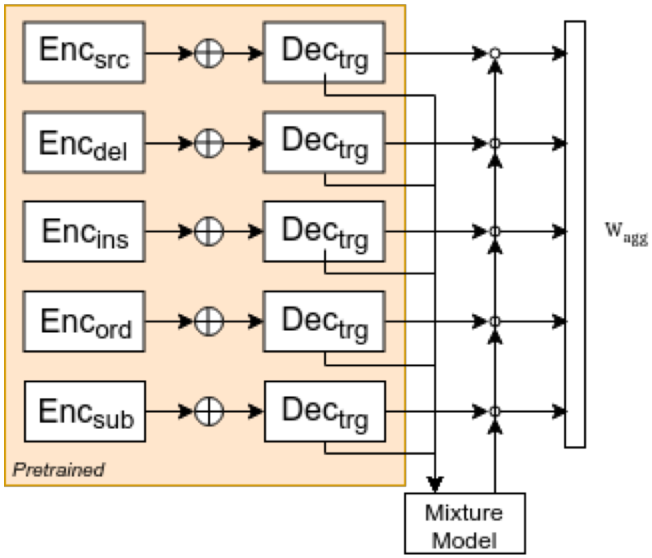
Multi-source and Multi-expert NMT

- Change input into paraphrases
- Multi-source NMT
  - Zoph and Knight (2016)



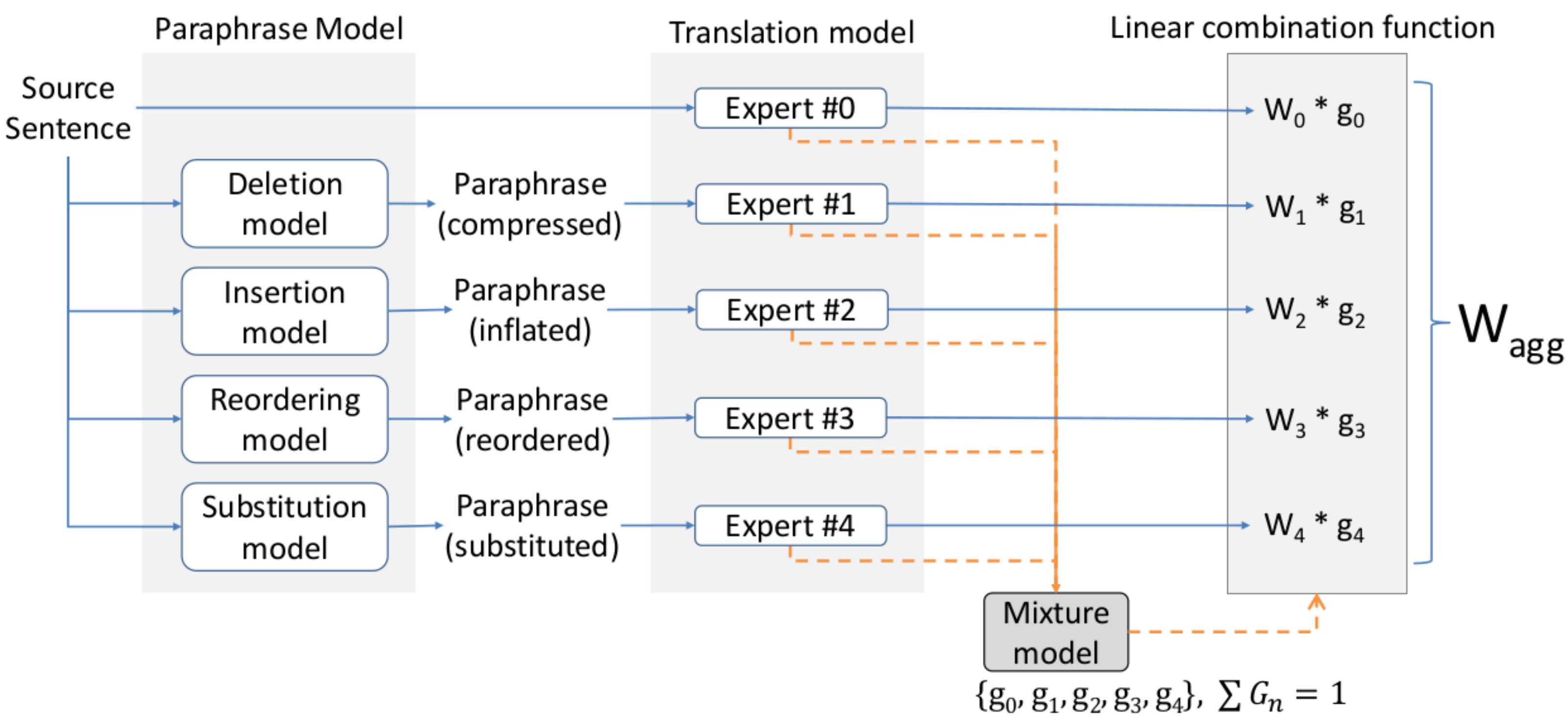
- Multi-expert NMT
  - Garmash and Mond (2016)
  - Final aggregated output weight is the linear combination:

$$W_{agg} = g_o W_0 + g_1 W_1 + \cdots + g_n W_n$$



Overall System: Paraphrase + Translation

- *For multi-expert NMT:*



Corpus Creation

Multi-paraphrase corpus creation

- Paraphrase WMT 2017 Multimodal Translation corpus
  - using crowdsourcing
- Using image as the basis of paraphrasing, the crowdworker paraphrase
  - Original -> {deletion, insertion, reordering, substitution}
- 3 months; 201 workers; 16 countries
  - English speaking countries, or at least English as second language
- Crowdsourced 10k of training data, dev, test

Caption	: A little gray dog jumps over a small hurdle.
Deletion	: A little gray dog jumps over a hurdle.
Insertion	: A little gray dog jumps over a small hurdle successfully.
Substitution	: A little gray dog pass over a small hurdle.
Reordering	: Over a small hurdle, a little gray dog jumps.

Generating the remaining paraphrases

- WMT dataset size is 29k pair
- Crowdsourcing successfully paraphrased 10k sentences

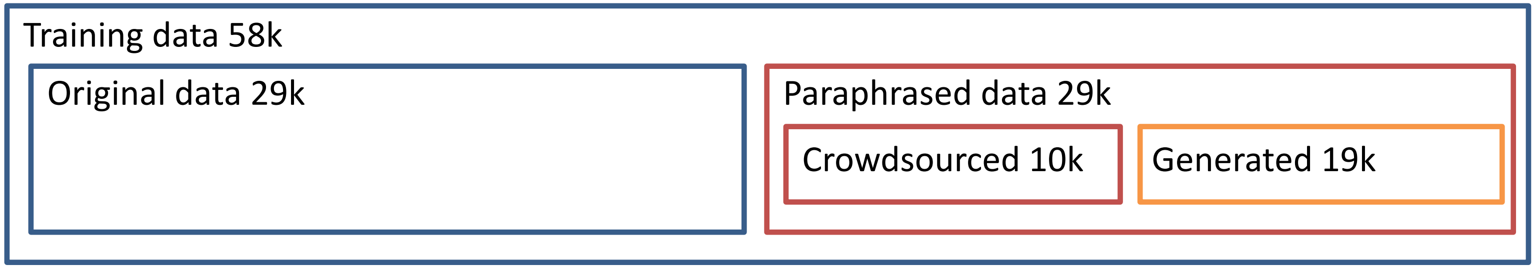
- Trained LSTM Encoder Decoder models for each paraphrase operation
  – Using 10k crowdsourced paraphrase
  – To generate remaining 19k paraphrase

Experimental Settings

Data Composition

- Combined paraphrased dataset with original dataset
  - Resulting 58k training data for each operation
  - The paraphrased data works as regularizer
- For dev and test dataset:
  - For paraphrasing : paraphrased dataset is used
  - For translation : original dataset is used

For each expert translation model:



Experiment Results

Experiment Result

Model Name	Test 2016		Test 2017		Test COCO 2017	
	BLEU	METEOR	BLEU	METEOR	BLEU	METEOR
Our NMT Baseline	37.7	55.6	30.1	49.7	25.0	44.6
Combine all data	36.7	53.9	29.6	47.7	25.1	43.7
Multi-source	37.0	55.0	30.8	49.6	25.0	44.3
Uniform weighted	39.6	56.9	31.4	50.7	26.7	46.0
Mixture of Expert	**40.5**	**57.6**	**32.5**	**51.3**	**28.0**	**46.8**

- Combining all data shows decrease in performance
- Mixture of Expert yields the best result
- Test COCO 2017 (ambiguous situation)

Result comparison with other models

Model Name	Type	Test 2016		Test 2017		Test COCO 2017	
		BLEU	MTR	BLEU	MTR	BLEU	MTR
Official WMT Baseline	Textual	32.5	52.5	19.3	41.9	18.7	37.6
Zhang et al. (2017)	Textual	-	-	31.9	53.9	28.1	48.5
Madhyastha et al. (2017)	Multimodal	-	-	25.0	44.5	21.4	40.7
Calixto et al. (2017)	Multimodal	41.3	59.2	29.8	50.5	26.4	45.8
Ma et al. (2017)	Multimodal	-	-	31.0	50.6	27.4	46.5
Helcl and Libovicky (2017)	Multimodal	36.8	53.1	31.1	51.0	26.6	46.0
Caglayan et al. (2017)	**Multimodal**	**41.0**	**60.4**	**33.4**	**54.0**	**28.5**	**48.8**
(Ours) Mixture-of-Expert	**Textual**	**40.5**	**57.6**	**32.5**	**51.3**	**28.0**	**46.8**

- Outperform almost all models, except one
- Works in par with other multimodal model
  - Only using textual information

Result Example - Unsuccessful

Type		Source Sentences	
(Data)	Original	a little girl climbing metal rope cables wearing a long pink skirt and black t-shirt .	
Translation Model	**Type**	**Target Sentences**	**BLEU+1**
Baseline /NMT	Original	ein kleines mädchen klettert metall an einem seil , das einen langen rosafarbenen rock und einem schwarzen t-shirt klettert .	**0.9**
(Data)	Target	ein kleines mädchen , das an metallseilen hochklettert und einen langen rosafarbenen rock und ein schwarzes t-shirt trägt .	-

Correct!

Correct!

Final hypothesis is quite different with target

Result Example - Successful

Type			
(Data)	Original	two motor	
Translation Model	**Type**		**BLEU+1**
Baseline /NMT	Original	zwei motorradfahrer fahren auf einer straße entlang .	0.75
(Data)	Target	**zwei motorräder fahren auf einer straße dem fluss entlang .**	-

The word "motorradfahrer" should be "motorräder fahren"

"dem fluss" is missing

Corrected in final result

Corrected in final result

Conclusions and Future Works

Conclusions

- A single caption cannot represents all the information of the image to which it refers to
- Generated multi-paraphrase of the WMT17 Multimodal Translation Task
  – Partially crowdsourcing with image as the basis of paraphrasing
  – Neural paraphrasing to complete the paraphrasing in semi-supervised way
- Proposed a textual model, in which the image information is not included in the model, but diffused in form of paraphrased caption
- +2.4 BLEU improvement over our NMT baseline

Future Works

- Try different combination strategies/integration point
- Investigate this proposed approach for another usage
  - Not limited for image caption translation
- Further investigate various methods of incorporating visual information

- Thanks for your attention!

References

- Rahul Bhagat and Eduard Hovy. What is a paraphrase? Computational Lin- guistics, 39(3):463–472, 2013.
- Philipp Koehn and Rebecca Knowles. Six challenges for neural machine trans- lation. In Proceedings of the First Workshop on Neural Machine Translation, pages 28–39. Association for Computational Linguistics, 2017.
- Yee Seng Chan, Hwee Tou Ng, and David Chiang. Word sense disambiguation improves statistical machine translation. In ACL, 2007.
- Marine Carpuat and Dekai Wu. Improving statistical machine translation using word sense disambiguation. In In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pages 61–72, 2007.
- Katharina Wäschle and Stefan Riezler. Integrating a large, monolingual corpus as translation memory into statistical machine translation. In Proceedings of the 18th Annual Conference of the European Association for Machine Translation, 2015.
- Çaglar Gülçehre, Orhan Firat, Kelvin Xu, Kyunghyun Cho, Loïc Barrault, Huei- Chi Lin, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. On using mono- lingual corpora in neural machine translation. CoRR, abs/1503.03535, 2015.
- Rico Sennrich, Barry Haddow, and Alexandra Birch. Improving neural machine translation models with monolingual data. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Pa- pers), pages 86–96. Association for Computational Linguistics, 2016.
  42Rajen Chatterjee, Matteo Negri, Marco Turchi, Marcello Federico, Lucia Specia, and Frédéric Blain. Guiding neural machine translation decoding with external knowledge, 01 2017.
- Liang Zhou, Chin-Yew Lin, and Eduard Hovy. Re-evaluating machine trans- lation results with paraphrase support. In Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 77–84, Stroudsburg, PA, USA, 2006.
- Santanu Pal, Pintu Lohar, and Sudip Kumar Naskar. Role of paraphrases in pb-smt. In Proceedings of the 15th International Conference on Computational Linguistics and Intelligent Text Processing - Volume 8404, CICLing 2014, pages 242–253, Berlin, Heidelberg, 2014. Springer-Verlag.
- Nitin Madnani and Bonnie J. Dorr. Generating targeted paraphrases for im- proved translation. ACM Trans. Intell. Syst. Technol., 4(3):40:1–40:25, July 2013.
- Eric Nichols, Francis Bond, D Scott Appling, and Yuji Matsumoto. Paraphrasing training data for statistical machine translation. Journal of Natural Language Processing, 17(3):3 101–3 122, 2010.
- Wei He, Shiqi Zhao, Haifeng Wang, and Ting Liu. Enriching smt training data via paraphrasing. In IJCNLP, 2011.
- Anabela Barreiro. SPIDER: A System for Paraphrasing in Document Editing and Revision — Applicability in Machine Translation Pre-editing, pages 365– 376. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.
- Chris Callison-Burch, Philipp Koehn, and Miles Osborne. Improved statistical machine translation using paraphrases. In Proceedings of the Main Conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics (HLT-NAACL ’06), pages 17–24, Stroudsburg, PA, USA, 2006.
  43Michel Simard, Cyril Goutte, and Pierre Isabelle. Statistical phrase-based post- editing. In Proceedings of the Main Conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics (HLT-NAACL ’07), pages 508–515, Rochester, NY, USA, 2007.
- Ekaterina Garmash and Christof Monz. Ensemble learning for multi-source neural machine translation. In COLING, 2016.
- Barret Zoph and Kevin Knight. Multi source neural translation. arXiv preprint arXiv:1601.00710, 2016.
- Ozan Caglayan, Walid Aransa, Adrien Bardet, Mercedes García-Martínez, Fethi Bougares, Loïc Barrault, Marc Masana, Luis Herranz, and Joost van de Wei- jer. LIUM-CVC submissions for WMT17 multimodal translation task. CoRR, abs/1707.04481, 2017.
- Jingyi Zhang, Masao Utiyama, Eiichiro Sumita, Graham Neubig, and Satoshi Nakamura. Nict-naist system for wmt17 multimodal translation task. In WMT, 2017.
- Jindrich Helcl and Jindrich Libovický. CUNI system for the WMT17 multimodal translation task. CoRR, abs/1707.04550, 2017.