



Multi30K

Multilingual German-English Image Descriptions

Desmond Elliott, Stella Frank,
Khalil Sima'an, Lucia Specia

Flickr30K



31,000

Professional
Translations

155,000

Crowdsourced
Descriptions

Translated Sentence



A brown dog is running
after the black dog.



Ein brauner Hund rennt
dem schwarzen Hund
hinterher

Independent Descriptions



A brown dog is running
after the black dog.

Ein schwarzer und ein
brauner Hund rennen auf
steinigem Boden
aufeinander zu

Multimodal Machine Translation



Source:
A brown dog is running after the black dog.

translate



Ein brauner Hund ...



evaluate

Gold Target:

Ein brauner Hund rennt dem schwarzen Hund hinterher.

Crosslingual Image Description



English Descriptions:

*A brown dog is running after the black dog.
Two dogs run towards each other on a ...
A brown dog is running after a black ...
Two dogs run across stones near a body ...
Two dogs playing on a beach.*

describe

Zwei Hunde ...

evaluate

Gold German Descriptions:

*Ein schwarzer und ein brauner Hund rennen ...
Zwei Hunde rennen über einen steinigen Platz.
Zwei Hunde spielen auf dem Strand.
Zwei Hunde rennen am Strand.
Zwei Hunde tollen in der Nähe des Meeres.*

Related Datasets

	Images	Sentences
COCO-FR/DE	1,000	5K De, Fr, and En
DeCOCO	1,000	1K De and En
TasvirET	8,000	24K Tr and En
Flickr8K-CN	8,000	40K & 40K Zh and En
Multi30K	31,000	31K and 155K De and En

How should we build massively multilingual multimodal corpora?

References

Lucia Specia, Stella Frank, Khalil Sima'an, and Desmond Elliott. A Shared Task on Multimodal Machine Translation and Crosslingual Image Description. Proceedings of the First Conference on Machine Translation.

Janarthanan Rajendran et al. Bridge Correlational Neural Networks for Multilingual Multimodal Representation Learning. Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies

Mesut Erhan Unal, Begum Citamak, Semih Yagcioglu, Aykut Erdem, Erkut Erdem, Nazli Ikizler Cinbis, Ruket Cakici. TasvirEt: A Benchmark Dataset for Automatic Turkish Description Generation from Images. IEEE Sinyal İşleme ve İletişim Uygulamaları Kurultayı

Xirong Li, Weiyu Lan, Jianfeng Dong, Hailong Liu. Adding Chinese Captions to Images. Proceedings of the International Conference on Multimodal Retrieval.

Julian Hitschler, Shigehiko Schamoni, Stefan Riezler. Multimodal Pivots for Image Caption Translation. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics.