

cases	doc_1		doc_2		decision	id
	authors	<ul style="list-style-type: none">Viktar AtlihaDmitrij Å eÅĵok	authors	<ul style="list-style-type: none">Atliha, ViktarÅ eÅĵok, Dmitrij	DUPLICATES	296
	title	Text Augmentation Using BERT for Image Captioning	title	Text Augmentation Using BERT for Image Captioning		
	publication_date	2020-08-28 00:00:00	publication_date	2020-01-01 00:00:00		
	source	SupportedSources.INTERNET_ARCHIVE	source	SupportedSources.CORE		
	journal	MDPI AG	journal	Applied Sciences		
	volume		volume			
	doi	10.3390/app10175978	doi	None		
	urls	<ul style="list-style-type: none">https://web.archive.org/web/20200830085240/https://res.mdpi.com/d_attachment/applsci/applsci-10-05978/article_deploy/applsci-10-05978.pdf	urls	<ul style="list-style-type: none">https://core.ac.uk/download/357745342.pdf		
	id	id2699235022872611662	id	id-6404761069829391659		
	abstract	Image captioning is an important task for improving human-computer interaction as well as for a deeper understanding of the mechanisms underlying the image description by human. In recent years, this research field has rapidly developed and a number of impressive results have been achieved. The typical models are based on a neural networks, including convolutional ones for encoding images and recurrent ones for decoding them into text. More than that, attention mechanism and transformers are actively used for boosting performance. However, even the best models have a limit in their quality with a lack of data. In order to generate a variety of descriptions of objects in different situations you need a large training set. The current commonly used datasets although rather large in terms of number of images are quite small in terms of the number of different captions per one image. We expanded the training dataset using text augmentation methods. Methods include augmentation with synonyms as a baseline and the state-of-the-art language model called Bidirectional Encoder Representations from Transformers (BERT). As a result, models that were trained on a datasets augmented show better results than that models trained on a dataset without augmentation.	abstract	Image captioning is an important task for improving human-computer interaction as well as for a deeper understanding of the mechanisms underlying the image description by human. In recent years, this research field has rapidly developed and a number of impressive results have been achieved. The typical models are based on a neural networks, including convolutional ones for encoding images and recurrent ones for decoding them into text. More than that, attention mechanism and transformers are actively used for boosting performance. However, even the best models have a limit in their quality with a lack of data. In order to generate a variety of descriptions of objects in different situations you need a large training set. The current commonly used datasets although rather large in terms of number of images are quite small in terms of the number of different captions per one image. We expanded the training dataset using text augmentation methods. Methods include augmentation with synonyms as a baseline and the state-of-the-art language model called Bidirectional Encoder Representations from Transformers (BERT). As a result, models that were trained on a datasets augmented show better results than that models trained on a dataset without augmentation.This research was funded by Google Cloud Platform Education Programs Grant		
	versions		versions			