

# Abstract

A statistical language model assigns the probability of a sequence of words based on the textual information from a training corpus. Given an intuition where people usually imagine something when they are talking about a particular topic, we introduce a complementary information to a language model from visual modal system which constructs the method of image-based language modeling. The image-based language model (ILM) is basically a statistical language model which contains visual information from a parallel *image+caption* corpus. This visual information is called as a set of abstracted-image topics. There are many words which are visually associated. We aim to use these topics to improve a statistical language model in assigning probability.

Several experiments were conducted on various test sets. Based on the result, the ILM works interestingly well in reducing perplexity of the test set in this study. Other than that, the ILM works very well in finding associated words of any particular word which is visually grounded or related. We compared the list of associated words with the result of the word co-occurrence language model. According to a small test set of associated words, the experiment results show that the ILM at its best can increase the F-score almost three times higher than the word co-occurrence language model when the minimum frequency of word agreement is  $> 1$ . Furthermore, the experiment results might support an intuition where visual context holds a significant role in sentence generation since the abstracted-image topics makes the ILM works significantly better than a unigram language model.

**Keywords:** Language Modeling, Visual Information, Associated Words