The
BRITISH
UNIVERSITY
IN EGYPT

Chosen Model: NLP For Image Captioning

2nd Phase Proposal

ID	Student Name	E-mail
196280	Ashraf Adel	ashraf196280@bue.edu.eg
194233	Farah Aymen	farah194233@bue.edu.eg
206562	Jacinta Samir	jacinta206562@bue.edu.eg
206069	Mohamed Negm	Mohamed206069@bue.edu.eg

Image captioning is a task that involves generating a description in natural language for an image. This problem is at the intersection of computer vision and natural language processing and has gained significant attention in recent years. One of the primary approaches to tackle this problem is to use NLP techniques [1] [2] to produce captions based on image features. This method involves training a model to recognize the objects and context in an image, and then generating a descriptive caption in natural language. From the previous phase, we have used the Flickr8K dataset provided by Illinois University [3] [4], which contains 8,000 images, each with five different captions, and we will continue to use it in this phase too.

Regarding phase 1, we attempted to use CBOW and skip grams as word embedding methods but found that both yielded unsatisfying results (loss of 83 and 5 respectively). Therefore, for phase 2, we have decided to embed the words using a pre-trained Glove Embedding, which is specifically designed for the English language. We will feed this word embedding to variations of LSTM, to evaluate its ability to generate captions accurately based on the image.

References

[1] M. Stefanini, M. Cornia, L. Baraldi, S. Cascianelli, G. Fiameni, and R. Cucchiara, “From Show to Tell: A Survey on Deep Learning-based Image Captioning.” arXiv, Nov. 30, 2021. doi: 10.48550/arXiv.2107.06912.

[2] Y. Feng, L. Ma, W. Liu, and J. Luo, “Unsupervised Image Captioning.” arXiv, Apr. 06, 2019. doi: 10.48550/arXiv.1811.10787.

[3] “Flickr 8k Data.” https://forms.illinois.edu/sec/1713398 (accessed Feb. 22, 2023).

[4] M. Hodosh, P. Young, and J. Hockenmaier, “Framing Image Description as a Ranking Task: Data, Models and Evaluation Metrics,” *Journal of Artificial Intelligence Research*, vol. 47, pp. 853–899, Aug. 2013, doi: 10.1613/jair.3994.