Chosen Model: NLP For Image Captioning

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 the task of generating a natural language description for an image. This problem lies at the intersection of computer vision and natural language processing (NLP), and has become an active research area in recent years. One of the main approaches to solve this problem is to use NLP techniques [1] [2] to generate captions based on the features extracted from the images. This involves training a model to recognize the objects and context in an image, and then generating a descriptive caption in natural language. One popular dataset used for this task is the Flickr8K dataset provided by Illinois university [3] [4], which contains 8,000 images each paired with five different captions, making it a valuable resource for training and evaluating image captioning models.

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.