**Image Captioning with English and Turkish Data Sets with Deep Learning**

***Semiye DEMİRCAN1,\*[C:\Users\Abdullah\AppData\Local\Microsoft\Windows\INetCache\Content.Word\ORCID-iD_icon-16x16.gif](https://orcid.org/0000-0001-8570-2412), Şeyma ŞEKER2[C:\Users\Abdullah\AppData\Local\Microsoft\Windows\INetCache\Content.Word\ORCID-iD_icon-16x16.gif](https://orcid.org/0000-0002-8344-3688)Özlem KILIÇ3[C:\Users\Abdullah\AppData\Local\Microsoft\Windows\INetCache\Content.Word\ORCID-iD_icon-16x16.gif](https://orcid.org/0000-0003-3118-0668),***

*1Computer Engineering, Faculty of Engineering and Natural Science, Konya Technical University, Konya, Türkiye*

*2Computer Engineering, Faculty of Engineering and Natural Science, Konya Technical University, Konya, Türkiye*

*3Computer Engineering, Faculty of Engineering and Natural Science, Konya Technical University, Konya, Türkiye*

|  |
| --- |
| **Abstract**  In recent years, the increase in image data has led to the need to make sense of images. Therefore, the work on image captioning is gain importance. Image captioning means generating a description of an image. In this study, Flickr 8k data set [[1](#_ENREF_1)] was used as the data set to make image captioning. Long Short-Term Memory (LSTM) from Iterative Neural Network (RNN-Recurrent Neural Network) algorithms was used on this data set. First, the pictures were categorized using a Convolutional Neural Network (CNN). As a result of the experiments, it was observed that the highest success was in the VGG16 and Inception V3 architectures, and work was carried out on both of these architectures. The best results were obtained in the VGG16 architecture. The extracted images and the two caption descriptions found for each image were given to the LSTM algorithm and an English description was generated for a new image. In this application, the BLEU (Bilingual Evaluation Understudy) score was used as the Evaluation metric. The BLEU score is a measure of the difference between automatic translation and human translation. The BLEU-1score was found to be 0.51%. The “TasvirEt“ data set [[2](#_ENREF_2)] was also used to obtain the same data set with Turkish subtitles. This data set is a data set consisting of subtitles of the Flickr8k data set translated into Turkish. The extracted images and this Turkish subtitle were given to the LSTM algorithm with the TasviEt data set. The BLEU-1 score of the study with Turkish subtitles was found to be 0.31%. While the BLEU score was among the acceptable values for the English data set, it remained at a low value for the Turkish data set. |
| Keywords: Image Captioning, CNN, LSTM,Deep Learning |

**References**

[1] *Flickr8k Dataset for image captioning*. Available: https://[www.kaggle.com/datasets/adityajn105/flickr8k](http://www.kaggle.com/datasets/adityajn105/flickr8k)

[2] M. E. Unal, B. Citamak, S. Yagcioglu, A. Erdem, E. Erdem, N. I. Cinbis*, et al.*, " TasvirEt: Görüntülerden Otomatik Türkçe Açıklama Oluşturma İçin Bir Denektaşı Veri Kümesi (TasvirEt: A Benchmark Dataset for Automatic Turkish Description Generation from Images)." presented at the 24. IEEE Sinyal İşleme ve İletişim Uygulamaları Kurultayı (SIU 2016), Zonguldak, 2016.