

---

# Automatic Caption Generation of Images

---

Supervisor: Oier Lopez de Lacalle & Eneko Agirre

## 1 Description

In this project you will build a deep learning model that generates descriptions of given photograph. Project requires to combined different deep learning architectures such as CNNs and LSTMs. Implemented model will be evaluated according to the standard metrics used in the field.

## 2 Objectives

The objectives are flexible depending on the desired level of difficulty:

- D1: Implement from scratch a caption generation model that uses a CNN to condition a LSTM based language model.
- D2: Extend the basic caption generation system that incorporates an attention mechanism to the model.

Objectives will be adjusted with the supervisor, and the final mark will depend on what is agreed.

## 3 Materials

We will provide some pointers to be able to code parts of the caption generation model. We will provide training and testing datasets, as well. Please contact Oier to check the details of the objectives, helper codes, as well as the datasets (oier.lopezdelacalle at ehu.eus).

## References

- [1] Vinyals, Oriol, Alexander Toshev, Samy Bengio, and Dumitru Erhan. "Show and tell: A neural image caption generator." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3156-3164. 2015.
- [2] Xu, Kelvin, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. "Show, attend and tell: Neural image caption generation with visual attention." In International conference on machine learning, pp. 2048-2057. 2015.