

German International University of Applied Sciences
Informatics and Computer Science
Dr. Caroline Sabty
TA. Merna Said
TA. Ayman Iskandar

Advanced Machine Learning , Spring 2025
Assignment 2
Due date is April 26, 2025 at 11:59 PM

Project Title: Automatic Image Captioning using CNN-RNN on Flickr8k Dataset

Project Description: This project focuses on building an automatic image captioning system that generates natural language descriptions for given images. The model uses a deep learning pipeline combining **Convolutional Neural Networks (CNNs)** for visual feature extraction and **Recurrent Neural Networks (RNNs)** for caption generation.

The system also incorporates an image classification step to improve contextual understanding, which enhances the quality and relevance of the generated captions.

Dataset: Flickr8k

The project uses the **Flickr8k dataset**

This dataset is a processed version of the original Flickr8k image captioning dataset, containing 8,091 images, each annotated with five human-written captions in natural language. It has been split into training, validation, and test sets for use in this competition.

Dataset Structure

Split — Images — Captions — Description

train — 6,472 — 32,360 — Used to train your model

val — 810 — 4,050 — Optional validation/tuning set

test — 809 — 4,045 — Used for final evaluation & scoring

Data Preprocessing Steps:

- a) **Caption Cleaning & Tokenization**
- b) **Vocabulary Construction**
- c) **Sequence Preparation**
- d) **Image Feature Extraction (CNN):**
- e) **Train-Test Split:**
 - Split dataset into 60% training, 20% validation, 20% testing (based on image IDs)
 - Ensure captions are linked correctly to the corresponding image

Model Architecture & Implementation Steps:

- a) **CNN for Image Features**
- b) **Embedding Layer**

- c) **RNN Decoder**
- d) **Final Dense Layer**

Deliverables:

- a) you can join kaggle competition using this link : <https://www.kaggle.com/t/51485f5ab2a74ea49526ef993fb4e438>
- b) Your code needs to be submitted on the Google form (make sure the code runs and no errors in it).
- c) Google form link: <https://forms.gle/EbkCCgym85xV7ZgQ6>
- d) Make sure the code runs without any errors.
- e) You can also use any Different RNN or CNN architectures that are not discussed in the course, but you need to fully understand and describe them very carefully in the submitted notebook.
- f) Avoid writing all your code in a single cell; organize it logically into multiple cells. accompanied by inline comments to explain each step.
- g) Each team should submit only one file with the names and IDs of the other team members (file format .ipynb).

PLAGIARISM IS NOT TOLERATED AND COPIED WORK WILL BE AWARDED 0 POINTS FOR BOTH PERSONS INVOLVED!