**Tutorial Overview**

This tutorial is divided into 6 parts; they are:

1. Photo and Caption Dataset
2. Prepare Photo Data
3. Prepare Text Data
4. Develop Deep Learning Model
5. Train With Progressive Loading (**NEW**)
6. Evaluate Model
7. Generate New Captions

**In this tutorial, you will discover how to develop a photo captioning deep learning model from scratch.**

After completing this tutorial, you will know:

* How to prepare photo and text data for training a deep learning model.
* How to design and train a deep learning caption generation model.
* How to evaluate a train caption generation model and use it to caption entirely new photographs.

## Photo and Caption Dataset

A good dataset to use when getting started with image captioning is the Flickr8K dataset.

The reason is because it is realistic and relatively small so that you can download it and build models on your workstation using a CPU.

The definitive description of the dataset is in the paper “[Framing Image Description as a Ranking Task: Data, Models and Evaluation Metrics](https://www.jair.org/media/3994/live-3994-7274-jair.pdf)” from 2013.

The dataset is available for free. You must complete a request form and the links to the dataset will be emailed to you. I would love to link to them for you, but the email address expressly requests: “*Please do not redistribute the dataset*“.

You can use the link below to request the dataset:

* [Dataset Request Form](https://illinois.edu/fb/sec/1713398)

Within a short time, you will receive an email that contains links to two files:

* **Flickr8k\_Dataset.zip** (1 Gigabyte) An archive of all photographs.
* **Flickr8k\_text.zip** (2.2 Megabytes) An archive of all text descriptions for photographs.

Download the datasets and unzip them into your current working directory. You will have two directories:

* **Flicker8k\_Dataset**: Contains 8092 photographs in JPEG format.
* **Flickr8k\_text**: Contains a number of files containing different sources of descriptions for the photographs.

The dataset has a pre-defined training dataset (6,000 images), development dataset (1,000 images), and test dataset (1,000 images).

One measure that can be used to evaluate the skill of the model are BLEU scores.