Schneider Electric European Hackathon by Nuwe

Data Science Junior

Group: No Support from Vector Machine

María García Girón
Diego González López
Santiago Valbuena Rubio

19/11/2022

Previous steps followed

Firstly, a feature engineering process has been carried out. Our objective has been to obtain, through the longitude and latitude fields, an additional field that shows the country that each of the image paths represents.

Once it has been understood that there are only images corresponding to one country, Indonesia, we have proceeded to use additional fields that allow us to obtain more information such as the city, state, and country names.

The GeoPy [1] library has been used, which makes it easy for Python developers to locate the coordinates of addresses, cities, countries, and landmarks around the world using third-party geocoders and other data sources.

After performing this transformation, our dataset is ready for training.

Used tools

Next, the tools used in this project will be shown.

- Python [2] programming language (version 3.6) under a PyCharm environment.
- Keras [3] is a high-level library written in Python for the development of neural networks whose objective is to make the use of this type of system as easy as possible.
- Pandas [4] library is a python library specialized in data structure analysis.
- Matplotlib [5] library is a library used in visualization tasks.
- Numpy [6] python library for data analysis, especially for large volumes of data.
- Pip [7] tool used to install Python libraries in a more comfortable, fast and efficient way.
- Tensorflow [8] tool that allows compiling deep neural network code.

- Scikit-learn [9].
- OpenCV [10] for computer vision applications.
- GitHub repository to manage projects and control code versions.

Dataset used

For the challenge, the data sets provided by Schneider Electric European have been used.

- Train SE.csv
- Test.csv
- Train_test_data.zip: Containing the training image datasets and testing datasets.

Contemplated alternatives

Developing a convolutional network design from scratch is not a simple process. There are numerous parameters that must be decided and normally there is no basis behind all the decisions that must be made. For example, one would have to decide the number of layers, the type of layer, the size of the kernel or even the optimal number of trainable parameters.

Currently, there is no defined pattern or steps to follow this complex process. Then the concept of transfer learning arises. It is based on already trained models that have worked correctly for other types of problems and that have achieved an optimum image recognition hit rate.

The *EfficientNet* [11] network that has been tested for the problem of this project.

In the following figure, we can see that in addition to using the pretrained network, the information from the text dataframe has been added to the training.

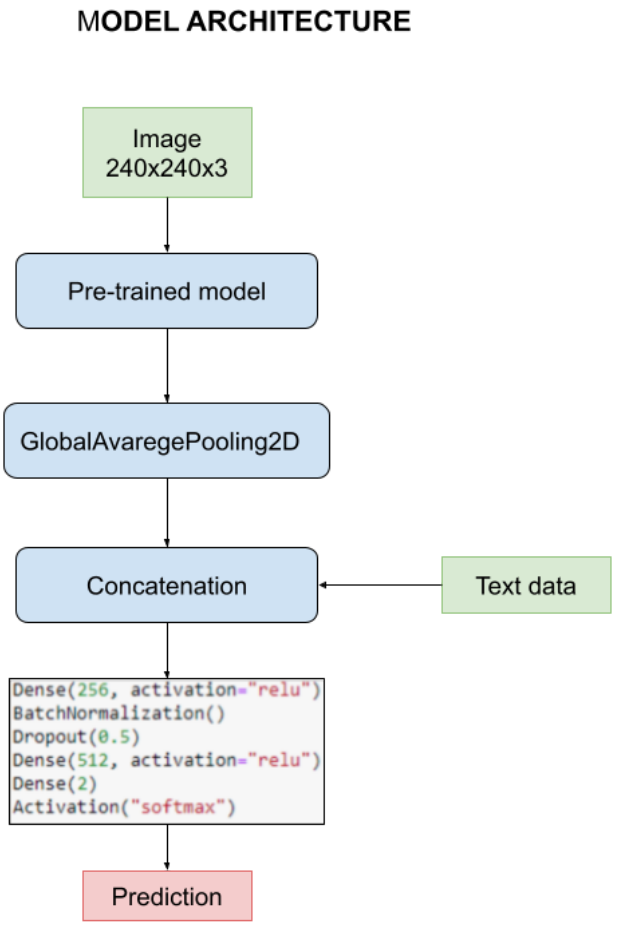

Fig [1] Architecture of the EfficientNet model used

Final score

In this section we show the final f1 score [12] result obtained is near 94% in training and 50% in validation.

```
Epoch 30/30
21/21 [==============================] - 97s 5s/step - loss: 0.1501 - accuracy: 0.9494 - f1_m: 0.9481 - val_loss: 13.9126 - val_accuracy: 0.5000 - val_f1_m: 0.5000
```

References

[1] Geopy library to locate the coordinates. https://geopy.readthedocs.io/en/stable/

[2] Python. February, 2019. https://www.python.org/

[3] Keras. February, 2019. https://keras.io/

[4] Pandas. https://pandas.pydata.org

[5] Matplotlib. https://matplotlib.org

[6] Numpy. https://numpy.org

[7] Pip. https://pypi.org/project/pip/

[8] Tensorflow. https://www.tensorflow.org/learn?hl=es-419

[9] Scikit-learn. https://scikit-learn.org/stable/

[10] OpenCV. https://opencv.org

[11] EfficientNet: Improving Accuracy and Efficiency through AutoML and Model Scaling by Mingxing Tan, Staff Software Engineer and Quoc V. Le, Principal Scientist, Google AI. https://ai.googleblog.com/2019/05/efficientnet-improving-accuracy-and.html

[12] F1 score. https://towardsdatascience.com/the-f1-score-bec2bbc38aa6