EGU23-5746, updated on 15 Apr 2024
https://doi.org/10.5194/egusphere-egu23-5746
EGU General Assembly 2023

Confidence estimation of DNN predictions for on-board applications

Nicolas Dublé[1], François De Vieilleville[1], Adrien Lagrange[1], and Bertrand Le Saux[2]
[1]AGENIUM Space, TOULOUSE, France
[2]ESA, PHILAB, FRASCATI, Italy

Most of the DNNs are designed to predict a class, a segmentation map or detections, no matter it is interpolation or extrapolation. Then, a confidence score answers to the need of having interpretable outputs and it could help an AI4EO end-user to take a decision.

The first investigated use case was binary classification of small Sentinel-2 tiles containing ships or not (with 2 classes “tile containing ship” or “tile not containing ship”). The database gathered 16,947 small 140x140 tiles extracted from 37 Sentinel-2 products. The ground truth was generated using Danish AIS data and then checked by human-eye. It was divided into several datasets for training, validation, testing, and active learning.

The second investigated use case was the classification of 10 geophysical phenomena from Sentinel-1 wave mode [Wang et al, 2018]. The database gathered 30,032 images with a quite balanced repartition between the 10 classes.

Classification networks (VGG16) were trained on the training datasets of both use cases, reaching high performances (>95% accuracy). We added several Out Of Distribution (OOD) examples for the ship classification use case, and used the test database provided for the Ocean Features use cases. Models reach around 70% accuracy on these 2 harder datasets so that regressing confidence could have an interest, with many examples of wrong classifications.

The solution developed used the confidNet approach developed by Corbière et al. Without retraining the classification DNN, we added a second DNN, composed by several dense layers, taking latent space from the classification network as input, which objective was to estimate a confidence score, by trying to approach the True Class Probability. It proved to be easy to train when enough failure examples are available in the database.

The main objective of the confidNet is to find the “ID”/”OOD” boundary, qualifying which examples the classifier should be able to predict (interpolation), and those it should fail to predict (extrapolation). An important work was done to try to qualify the quality of the predictions of the confidNet (confidence score) to ensure that it didn’t just learn to map the subset of the dataset where the classifier fails, and the one where the classifier was right. It presented interesting properties of generalization and turned out to be less “dataset-dependent” than a classical DNN.

21 different network configurations were tested, making the size of the architecture vary from 4k

to 2.5M parameters. It showed that many of these configurations could reach similar results, and that the number of layers was more decisive than the number of parameters of the intermediate feature maps.

The main results obtained in this study are the relevance to utilize the confidNet approach in AI4EO scenarios, the possibility to reduce the network in an on-boarding interest, and a first warranty that the confidNet approach can learn in a different way from classification networks, with interesting properties of generalization. This study demonstrates the possibility to associate confidence scores to the predictions of a DNN in a satisfying way.