**Description for AIS data and code**

**A novel ship movement classification from Automatic Identification System (AIS) data based on convolutional neural network**

Xiang Chena, Yuanchang Liub,\*, Kamalasudhan Achuthana,\* and Xinyu Zhangc

a Department of Civil, Environmental and Geomatic Engineering, University College London, Chadwick Building, London, WC1E 6BT, UK

b Department of Mechanical Engine>ering, University College London, Torrington Place, London, WC1E 7JE, UK

c Key laboratory of Maritime Dynamic Simulation and Control of Ministry of Transportation, Dalian Maritime University, Dalian 116026, China

Email Addresses: [xiang.chen.17@ucl.ac.uk](mailto:xiang.chen.17@ucl.ac.uk) [Yuanchang.liu@ucl.ac.uk](mailto:Yuanchang.liu@ucl.ac.uk) [k.achuthan@ucl.ac.uk](mailto:k.achuthan@ucl.ac.uk)

* **All codes in this paper have been implemented in Python. The essential libraries are: sklearn, tensorflow, numpy, opencv-python, math, matplotlib and pandas.**
* **Please read the following comments if you want to use the code and data.**

1. The Automatic Identification System (AIS) is an automatic tracking system used on ships and by Vessel Traffic Services (VTS). Information provided by AIS equipment, such as unique identification, position, course, and speed, can be displayed on a screen or an ECDIS. AIS is intended widely to assist a vessel's tracking, watchkeeping and allow maritime authorities to track and monitor vessel movement. The AIS data is derived from the shore-based station and satellite-AIS. These data formats is commonly used the NMEA coding format. There are many open sources for decoding the AIS data. And this project uses one of decoding programming to decode the data to .csv format. After that, we transfer these data into MySQL to do the preliminary process. When the data are selected in the MySQL platform, it can be easily converted into .csv format. The .csv format is compatible to process in the Pandas or Numpy packages under the Python development environment.
2. In the subfolder named ‘1.Process\_ais\_data’ of the folder ‘1. Ship\_movement\_images\_generation\_labeling’ contains the steps of processing the raw AIS data. The data for this paper is saved in the following link: <https://drive.google.com/open?id=1IW4D4ISWlH-b4zq9_nOcqtGhnDCK5ncA> . Before processing, please download the main dataset from the above link. The raw AIS data name is ‘raw\_ais\_data.csv’.
3. After processing the raw AIS data, the next step is to generate the images and related labels. The subfolder ‘2. Generate\_image\_labeling’ is the python file for generating the movement images and labelling. The data has been saved in the folder ‘movement\_images’ and ‘label.csv’ in the above link.
4. The folder ‘2. Training\_and\_Testing\_CNN’ is the function for the convolutional neural network for ship movement classification. The file ‘preprocess\_data.py’ is the preliminary step before sending the images into the convolutional neural network. It contains the decoding the images and resizing them. And, the data augmentation processes are random flip and standardisation. Please refer the code.
5. The file ‘train\_scratch.py’ is the framework of the CNN and it contains the different convolutional layers, batch normalisation, max pooling, dropout, flatten and dese operations. The different layers’ settings are explained in the table 2 of the paper.
6. The file ‘classical\_ML.py’ is the other classical machine learning algorithms (K-nearest neighbour network, support vector machine and decision tree).

Please contact with us through [xiang.chen.17@ucl.ac.uk](mailto:xiang.chen.17@ucl.ac.uk) if you have any question regarding the data and codes.

Thanks,

Xiang Chen