

DIGITAL AUTOMATION OF MARINE DATA

SHOM & IMT Atlantique

AUTHORS: RUNLU QU, XUE WANG, ZHANG HANG, ZHENG YUAN
 SUPERVISOR: CÉCILE BOTHOREL
 CLIENTS: THIERRY SCHMITT, ERIC LE GUEN, JULIAN LE DEUNF

• INTRODUCTION

As part of the digitization of its archives, the Shom has a large number of bathymetric maps on which are handwritten annotations. In order to exploit these documents, digitization of the maps is important. It is necessary to automate the recognition of the information they contain and in particular the bathymetry or depth of water information.

• OBJECTIVE

- Set up one or several algorithms for identifying and extracting handwritten annotations from digital maps in the form of an image.
- Have a tool for statistical analysis of classification performance.

• PROCESSUS



$$2 + 0. + 3 = 20.3$$

METHODE A

Prepare the training data set:

1. Get the images of digits from the original map (0-9, 0. - 9., .0 - .9)
2. Randomly combine the single number into strings of different lengths as the training set, such as 21.5, 16.7, etc.

Input (original map)

Remove noises

Object detection

$$3.6 = 3. + 6$$

METHODE B



- Two CNN networks:
 1. Determine string length
 2. Recognize strings

Output: the entire string

20.3

RESULTS & DISCUSSION

After testing these two methods:

- The performance of method A is not satisfied because of its massive classes (from 0.1 to 200) which is up to 2000 classes. It requires the complexity of the neural network if we want to get an accurate result.
- The accuracy of method B is depending on the accuracy of positioning accuracy (60% - 90%), detection of decimal point accuracy (80% - 95%), segmentation-recognition-combination accuracy (51% - 100%). The final accuracy is between 24% to 85%. The performance of our algorithm reaches to our expectation, but still could be improved.

Rotate the splitted image and do pre-processing

Split the positioned image into individual digits

apply CNN to recognize digits

Output : Combine the recognized digits into a string

3.6