Optical Characters Recognition

Classic computer vision approaches

The common pipeline looks as follows:

1. filtering – to suppress noise and highlight the foreground (characters),
2. contour detection - to recognize the characters,
3. image classification - to identify the characters

Advantages:

- use standard and well-known methods (noise-suppressing filters, edge detection filters, etc.),
- is not an ANN.

Drawbacks:

- hard to generalize,
- needs a lot of image preprocessing,
- fails in a more challenging tasks.

ANN approaches

The common pipeline looks as follows:

- a convolutional neural network (CNN) - to extract image features,
- a recurrent network (RNN):
    - to predict the position of letters in the image,
    - to classify/recognize each of the letter.

In other words, the OCR consists of two stages:

- text localization,
- text recognition.

Methods:

- DNN (Deep Neural Network)[6],
- Single-shot detection techniques (YOLO)[5],
- RCNN (Recurrent-Convolutional Neural Network)[1],
- RNN (Recurrent Neural Network)[8],
- CNN (Convolutional Neural Network)[10],

- Reinforcement learning approaches: RAM – (Recurrent Attention Model) and DRAM (Deep Recurrent Attention Model)[7].

Table recognition

- Semantic segmentation + (Fast) RCNN [9],
- RCNN (Recurrent-Convolutional Neural Network)[9],
- GNN (Graph Neural Network) [11, 12],
- DGCNN (Dynamic Graph Convolutional Neural Network)[13].