2nd ADNE assignement

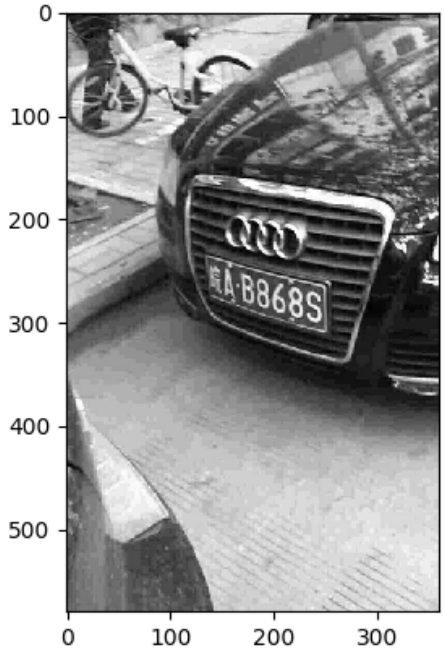
Contents

There are 3 tasks on this assignment. The weights of each of the tasks for grading purposes are:

task 1. 13/20

task 2. 3/20

task 3. 4/20



!h

Figure 1: Example of a task-1 dataset image

1 Task 1

You must create a convolutional neural network that identifies the license plate characteres in images of chinese cars like those in this task-1 dataset.

1.1 task 1 dataset

The data for this task is in the google drive folder:

`https://drive.google.com/drive/folders/1icyeu_vw4bMvdFewooEQ1sIW4h0jQsbl?usp=sharing`

It consists of 40 tensorflow **tfrecords**[1] containing more than 130000 grayscale images of cars with chinese license plates.

[1]see `https://www.tensorflow.org/guide/data` for the use of tfrecords in the tf.data.Dataset API.

The script 'inspectTFrecord.py' (included in the above drive folder) shows how to acess images and labels of the examples in the task-1 *tfrecords*.

1.2 Requirements for task 1:

1. Your code must be written in keras (tensorflow-2.xx). You must identify which tensorflow-2.xx you used to run your code because there may be some incompatibilities between different tensorflow-2.xx versions.

2. There must be a function, *train*, that trains your model from scratch, using a **tf.data.TFRecordDataset** that consumes tfrecords from task-1 dataset.

3. There must be a function, *predict*, that receives tensors with shape (?,580,360,1) and produces a tensor with shape (?,7) with the predictions of the license plate characteres for each (580,360,1) tensor slice.

4. There must be a function, *evaluate*, that evaluates any 'tf.data.dataset' with examples like those in task-1 dataset. Your code will be tested on images that do not belong to task-1 dataset but are similar, in particular they have the same shape.

5. You must deliver a report that contains:

   - A description of your train, validation and test datasets: which tfrecords you used for each set.
   - A description of the loss and accuracy functions you used.
   - The number of epoches you use for training and the results (loss, accuracy) on the train, validation and test datasets.

1.3 About the approach and dataset used in task 1

1.3.1

Our dataset is based on the CCPD (Chinese City Parking Dataset), [1]. We reduced the images resolution and use only grayscale instead of RGB.

The approach we use in this task is much simpler than the one proposed in [1] and can achieve similar results ($\approx$ 98% accuracy). One can simplify the approach in [1]: we adapt the detection module in order to use it for recognition, and not use at all their recognition module.

1.3.2 Labels description

License plates in our dataset have 7 characters and are identified by sequence of 7 integers:

One chinese character: it identifies the china province. There are 31 possible characteres. It corresponds to the first number in the label, between 0 and 30. In our dataset, in 96% of the examples, the label for the chinese province is 0.

A letter: there are 24 possible letters. It corresponds to the second number in the label, between 0 and 23. Two letters are absent 'i' and 'o'. In our dataset, in 91% of the examples, the label for the letter is 0.

A sequence of 5 characteres, letters or numbers: there are 34 (=24+10) possible values for the number that identifies each of the 5 characteres.

Because some labels are much more frequent than others, it is very easy for a model to get around 30% accuracy: it does not mean that the model has learned anything more than just taking the average.

2 Task 2

This task is of a more theoretical nature than the other two tasks. You should write a text (no more than 3 pages) where you compare the 4 different attention mechanism approaches described in [1], [2], [3] and [4]. Assume that we are interested on the application of those mechanisms to image recognition. Keep in mind that you should list the advantages and disadvantages of each approach against the others and not just make a summary of the attention mechanism used on each paper.

3 Task 3

You must create a model, that identifies the street house numbers in images like those in task-3 dataset, using the attention mechanism presented in *Attention-Based Models for Speech Recognition* [3].

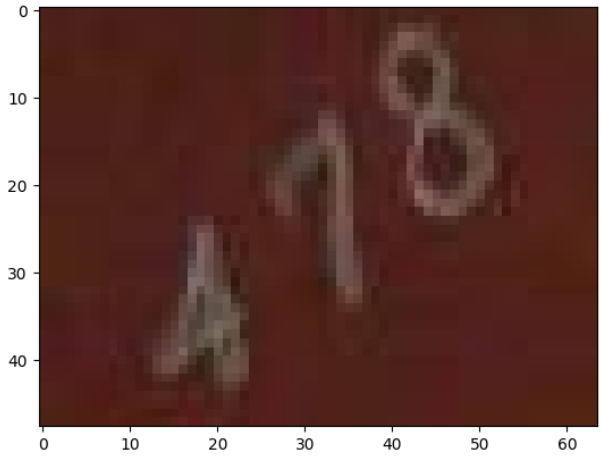


Figure 2: Example of a task 3 dataset image

3.1 task 3 dataset

Train, validation and test tfrecords are in the google drive folder `https://drive.google.com/drive/folders/1AxKQSJVBtfbLs-H1Nvrzlq1SKXDrblMB?usp=sharing`.

The script 'inspectTFrecord.py' (included in the above drive folder) shows how to acess images and labels of the examples in the task-3 *tfrecords*.

3.2 Requirements for task 3:

1. Your code must be written in tensorflow-2.xx. You must identify which tensorflow-2.xx you used to run your code because there may be some incompatibilities between different tensorflow-2.xx versions.

2. There must be a function 'train' that trains your model from scratch using task-3 dataset.

3. There must be a function *predict* that receives tensors with shape (?,48,64,3) and produces a tensor with shape (?,6) with the predictions of house numbers for each (48,64,3) tensor slice.

4. There must be a function 'evaluate' that evaluates any 'tf.data.dataset' with examples like those in task-3 dataset. Your code will be tested on images that or not in this task dataset but are similar, in particular they have the same shape.

5. You must deliver a report that contains:

   - A description of the loss and accuracy functions you used.
   - The number of epoches you use for training and the results (loss, accuracy) on the train, validation and test datasets.

3.3 About the approach and dataset used in task 3

Our dataset is based on *The Street View House Numbers (SVHN) Dataset*[5]. The numbers in our images may have from 1 to 6 digits, therefore, it makes sense to use a *generative* model as a decoder. Possibly there are simpler aproaches that achieve equal or better better results in this dataset. The appropriate dataset to use with this, or similar, attention mechanism would be *The French Street Name Signs Dataset* [6], but training oon that dataset could last several weeks and we don't have that time.

The labels we use are lists with 6 integers. If the image contains two digits, the corresponding label will have those two digits and the number 10 repeated 4 times: the number 10, in the label, means **no digit**.

4 Dates and rules for the 2nd assignement

1. The assignement can be done individually or in groups of at most 2 elements.

2. The answer to each task must be submited in the Moodle before the following dates:

   task 1 : May 18

   task 2 : May 25

   task 3 : May 30

   Warning: Task 1 should be straight forward, with some small difficulties related to the large dimension of the training data, so, if you plan to

solve any of the other two tasks, you should finish task 1 much earlier than the limit time because the other two tasks will require much more time than task 1.

3. There will be an oral discussion of the assigment with each group, between June 1 and June 5.

4. Before May 11 each student must answer a questionaire in the moodle, telling if he works alone or in a group and, in the last case, who is the other member of his group.

5. All the elements in the group are responsible for the totality of the group work and will have to answer for it in the oral discussion.

6. Different groups are not allowed to share code.

5 Bibliography

1. "Towards End-to-End License Plate Detection and Recognition: A Large Dataset and Baseline", Xu, Zhenbo and Yang, Wei and Meng, Ajin and Lu, Nanxue and Huang, Huan, Proceedings of the European Conference on Computer Vision (ECCV), pages 255–271, 2018 pdf.

2. "Attention-based Extraction of Structured Information from Street View Imagery",Zbigniew Wojna, Alex Gorban, Dar-Shyang Lee, Kevin Murphy, ICDAR (2017)pdf. Qian Yu, Yeqing Li, Julian Ibarz

3. "Attention-Based Models for Speech Recognition", Jan Chorowski, Dzmitry Bahdanau, Dmitriy Serdyuk, Kyunghyun Cho, Yoshua Bengio, NIPS Proceedings, 2015.pdf

4. "Recurrent Models of Visual Attention ", Volodymyr Mnih Nicolas Heess Alex Graves Koray Kavukcuoglu, NIPS Proceedings, 2014pdf.

5. "Reading Digits in Natural Images with Unsupervised Feature Learning", Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, Andrew Y. Ng, NIPS Workshop on Deep Learning and Unsupervised Feature Learning 2011

6. "End-to-End Interpretation of the French Street Name Signs Dataset", Smith, Raymond, et al., European Conference on Computer Vision. Springer International Publishing, 2016.