Three Master thesis Projects
- in AI and Machine Learning related to Text Recognition and Face recognition

Supervisor: Anders Hast

anders.hast@it.uu.se
https://andershast.com

I have more project ideas so please contact me if you want to discuss other projects than the ones listed below, or if you have similar ideas that you want to investigate more.

Learning Important Features for Classification

CNN’s and also Visual Transformers computes features of the whole image in order to make classification. This thesis would investigate how classical Computer Vision techniques for feature extraction can be used to learn from scale and rotation invariant features, such as SIFT. After classification of features, a bag of words-model can be used to tell what is visible in the image. The image can be handwritten text, or images of objects. The learning process should be much faster this way, since not everything in the image is processed, but rather the important parts. Moreover, since scale and orientation is know, a framework can be built, which is invariant to these factors, and hence less training data is necessary. An interesting extension/variant is described next.

Visual transformers have gained a lot of attention recently and has Achieved state of the art results for some applications. However, they are data hungry and therefore require extensive computing resources. The reason might be that they divide images into equal size patches that are flattened and fed into the transformers. This means that the networks needs images of objects in different sizes and orientations to be able to recognise them. Moreover, patches that contain no objects are also feed into the transformers. A possible remedy for this would be to use scale and rotation invariant key point detectors (e.g. SIFT), which find points of interest in objects together with their estimated sizes. By feeding the network with patches from the most prominent key points, the Transformer needs to handle much less data. Moreover, the patches will be scaled and rotated so that patches taken from the same objects will tend to have the same size, which will further reduce the need for training images. The positional embedding must be changed accordingly, which would use the key point position. This project aims at comparing the how learning time, performance and accuracy will change when using the above mentioned approach instead of the normal approach.

Age Estimation in Photographs

A persons age can be very hard to tell from face photos. However, deep learning approaches have the capacity to learn from huge amounts of data. Several such databases exist online. The purpose of this project is to learn from such datasets and make estimations of the age of persons, regardless of gender and ethnicity. The results will be very useful for the EB-CRIME project where age estimation is one important part. Especially it is interesting to tell whether a persons has reached the age of 15 or 18. It is also interesting to determine how well the algorithm works for different groups (age, gender and ethnicity). Generally it is easier to tell which age span a person belongs to, e.g. 11-20, 21-30 etc. By classifying different overlapping age spans, one should be able to narrow down to a more exact age. E.g. if one classifier tells that a person belongs to 21-30 and another tells 26-35, then one can conclude that a person more probably belongs to the group 26-30. However, the choice of group is not necessarily binary, but rather it is fuzzy. This can hopefully be used to give a more accurate analysis of a persons age, i.e. give a probability of what a persons age is (e.g. 26 = 5%, 27=42%, 28=82%, 29 14%, 30=1%) , rather than just one age. This project could be extended to also estimate gender and ethnicity.

Improved OCR

Optical Character Recognition (OCR) is generally much simpler than Handwritten Text Recognition (HTR). Since printed characters (using types molded in led in matrix typesetting, or a typewriter), usually have a fixed size and are rather similar for each character, it presents itself as an easier problem than HTR. However, types age and can be damaged and they also come out a bit differently when being molded. The ink can be abundant (or lacking), especially for typewriters, so a ‘c’ can like like an ‘e’. Moreover, combinations like ‘rn’ can look more like an ‘m’, etc. Hence, automatically OCR’d text has very varying quality and sometimes lots of errors can be found, which decreases the readability. Also, when a page contains several columns, titles and captions, there is a big risk that these are not properly identified and the text becomes unreadable, when these are accidentally mixed.

The idea in this work is to use existing software for document analysis and HTR and do transfer learning so that it works for OCR. This means that instead of identifying the extent of each character using the small space between them and then classify it (as is done in OCR), it takes on the HTR approach and tries to identify characters regardless character space. Hence, the character space problem is avoided, which might lead to better OCR! Comparisons can be done to classical OCR such as Tesseract and Kraken.