report-part
PlayCDC

15.07.2018

Dataset creation

INTRO DEPENDING ON WHAT IS BEFORE
As there was no suitable dataset of cards to use, we decided to create it by our own. We did this in basically three main steps to obtain at the end **(Nr. images)**images, as well as the necessary Bounding Box information required for training our YOLO NN. First we had to create the data doing photos and arrange it in such a way that we can easily use it around our workflow. The next, we had to detect the convex hulls of the card numbers and suits and the last perform rotations/re scaling/... on the images to generate new data.

Data preparation
We decided to work with a deck of 52 cards, where each card contain two times the card logo. We took for each card 2 different photos **(TODO: check this there were not really 52)** and after this we extract the cards of the photos and re-scaled them to 600x900 pixels. We choose this resolution arbitrary, as we though that it was a nice size to work with. For this part, we used the selection/rotation/cropping/re-scale -tools provided by GIMP[1]. Concluding this, we proceed by detecting the convex hulls of the cards numbers and suits. We detect the convex hulls using SciPy libraries[2]in a semi-automatic way, verifying for each card manually that we got the desired result **[SEE FIGURE]**. At the last step, we saved the convex hulls as an NumPy array **(cite??)**

Data generation
The goal in this step was to generate a big amount of new data for each card destined for training or NN. We wanted to perform image transformations, like translations, rotations and re scaling, as well as blurring and sharping to generate new data. We decided to use the imgaug python library, because it provided a nice way to keep track of the convex hulls after doing transformations.

---

[1]GIMP 2.8.22 - GNU Image Manipulation Program
[2]SciPy: Open Source Scientific Tools for Python