* 1. **Problem statement:**

My idea is to count how many people in a picture or a video by a simple convolutional neural network. It is inspired when I watched some political events on the television. Those different channels would show you some videos or pictures with different numbers of people although it is from the same event. TV channels might want to be my clients and provide a more accurate number of people in their stories. They should be willing to show a really confident number especially when they are reporting particular events like the presidential election campaign. It is not only good in convincing their audience, but also in building up their reputation.

* 1. **Dataset:**

Non-facial images:

1. CIFAR10 (C10)
2. CalTech Cars Rear background (bg)
   1. <http://www.robots.ox.ac.uk/~vgg/data3.html>
3. CalTech Cars Rear (car)
   1. <http://www.robots.ox.ac.uk/~vgg/data3.html>
4. Place365 CNNs (ls)
   1. <http://places2.csail.mit.edu/download.html>

Facial images:

1. MS-Celeb-1M (<https://www.msceleb.org/>)
   1. Sample ImageThumbnails Data (full)
   2. Sample FaceCropped Data (crop)
   3. Sample FaceAligned Data (align)
2. Labeled faces in the wild (lfw)
   1. <http://vis-www.cs.umass.edu/lfw/>
3. CalTech Human Face Front (face)
   1. <http://www.robots.ox.ac.uk/~vgg/data3.html>
   2. Summary of findings

I trained CNNs with different combinations of facial images and non-facial images. Then, I tested the trained model with completely unrelated datasets. The results show that model 5 and 8 have greater scores than the rest and the difference of them is that model 5 has more capability to recognize the face with the body because it has higher accuracy scores in “full” and “face” testing datasets. In order to get a better classifier, I also tried Keras’s augmentation preprocessing in model4\_aug (32x32) and model5\_aug (32x32). Unfortunately, both of them did not give me a better accuracy score.

Model training:

* BGR images (32 x 32) if not specifically indicated.

Model structure



Model testing:

* aug: augmentation
* ∆: different model structures compared to the rest of models

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Training dataset | | | | Accuracy from other facial images  Red: not included in the training | | | | |
| Non-facial images | | Facial  images | | align | crop | full | face | lfw |
| model | C10 | 60,000 | align | 1,375 | 0.98 | 0.04 | 0.03 | 0.02 | 0.01 |
| model\_slice | C10 | 1,375 | align | 1,375 | 0.98 | 0.38 | 0.22 | 0.31 | 0.57 |
| model3 (gray scale) | C10 | 1,375 | align | 1,375 | 0.99 | 0.17 | 0.1 | 0.08 | 0.36 |
| model4 | ls | 1,375 | align | 1,375 | 0.99 | 0.08 | 0.06 | 0.04 | 0.23 |
| model5 | ls | 4,256 | align  full  crop | 4,256 | 0.98 | 0.97 | 0.88 | 0.74 | 0.86 |
| model6 | ls | 13,233 | lfw | 13,233 | 0.32 | 0.47 | 0.15 | 0.5 | 1 |
| model7 | ls | 10,000 | lfw | 10,000 | 0.28 | 0.58 | 0.23 | 0.52 | 1 |
| model8 | ls | 2,750 | align  crop | 2,750 | 0.99 | 0.96 | 0.52 | 0.51 | 0.91 |
| Model1\_aug (150 x 150) ∆ | bg | 1,370 | align | 1,375 | 0.96 | 0.74 | 0.66 | 0.48 | 0.85 |
| Model2\_aug (64x64) ∆ | bg | 1,370 | align | 1,375 | 0.98 | 0.65 | 0.55 | 0.42 | 0.83 |
| Model3\_aug (32x32) ∆ | bg | 1,370 | align | 1,375 | 0.97 | 0.48 | 0.43 | 0.57 | 0.76 |
| Model4\_aug (32x32)  \* based on model8 | ls | 2750 | align  crop | 2750 | 0.99 | 0.98 | 0.61 | 0.63 | 0.95 |
| Model5\_aug (32x32)  \*more augmentations | ls | 2750 | align  crop | 2750 | 0.8 | 0.65 | 0.39 | 0.44 | 0.56 |