**Selected Architecture Notes:**

Data inspection:

-Image rotation is less than +/- 3 degrees

-Image scaling is less than +/- 3% on both axes

-Image capture offsets are less than +/- 60 px (off the image average) on both axes

Data processing and notes:

Of note, some of this is not super necessary since the data set is small and optimization on compute time is not really critical. I just included it here so you know I’m aware these things are possible in case we had compute limitations. I used AWS sagemaker so the whole thing didn’t take much more than a few minutes to train.

-1ch grey, 640x640 or smaller (no need for RGB)

-can employ image dynamic range normalization if needed

Initial Architecture Notes:

a.     Images are pre cropped to 640x640, utilizing random offset data augmentation, and are normalized to max dynamic histogram range [0-255] with 2% of original histogram edge clipped to reduce noise.

b.     7x7 filters is selected based on the relatively low spatial frequency of the points of interest (large object corners).

c.      The usage of 2 convolution layers are estimated based on the relatively simple image features to track.

d.     hyper parameters - architecture:

- filters size, stride step.

- adding dropout layer after convolution 1 if MSE error results suggests overfitting

e.     hyper parameters training (AWS sagemaker has great tuner algorithms):

- image data augmentation count [1500 planned].

- mini batch size = 16

- Adjusted learning rate (start with 0.01, divide by 10 if CV results start to flatten, with patience of 5-10 steps.

- gradient descend optimizer (adaptive momentum)

1.     1 Input layer:

640x640 x 8 bit (gray)

2.     C.1 Convolution layer:

Convolution filter size 7x7, stride=2, padding = 0, activation ReLu.

Number of filters: 8

Feature map size: [(640-7)/2]+1= 317

C.1: 317x317x 8

3.     P.1: Max pooling layer (2x2,16)

P.1: 159 x159 x16

4.     C.2 Convolution layer:

Convolution filter size 5x5, stride=2, padding = 0 , activation ReLu.

Number of filters: 16

Feature map size: [(159-5)/2]+1=  78x78

C.2: 78x78x16

5.     P.2: Max pooling layer (2x2,32)

P2: 39 x39 x32

6.     Fully connected layer feature vector: 1x 32, activation ReLu.

7.     Drop out some data randomly during training to minimize overfitting

Dropout layer (0.5)

8.     Regression prediction:  a vector of 6 numbers for points P1, P2, P3 respectively: (x1,y1,x2,y2,x3,y3)

Fully connected layer 1x 6         # is a regression head.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | bl\_x | bl\_y | ul\_x | ul\_y | ur\_x | ur\_y | x\_axis\_rot | SegX | SegY |  |  | d12 |  |  | d13 |  |  | d23 |
| Max rot (Deg) |  |  |  |  |  |  |  | 2.7 |  |  |  |  |  |  |  |  |  |  |  |
| Max rot (deg) |  |  |  |  |  |  |  | 1.5 |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| min scale: |  |  |  |  |  |  |  |  | 96% | 97% |  |  | 97% |  |  | 98% |  |  | 96% |
| max scale: |  |  |  |  |  |  |  |  | 102% | 102% |  |  | 102% |  |  | 103% |  |  | 102% |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Offset rangeY pix: | 114 |  | 101 |  | 114 |  | 113 | 8 | 10 | 16.5 |  |  |  |  |  |  |  |  |  |
| Offset rangeX pix : | 98 | 82 |  | 95 |  | 98 |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Min |  | 633 | 185 | 702 | 475 | 870 | 478 | 0 | 162 | 291.5 |  |  | 300 |  |  | 378 |  |  | 162 |
| Max |  | 715 | 286 | 797 | 589 | 968 | 591 | 8 | 172 | 308 |  |  | 315 |  |  | 398 |  |  | 172 |
| Average |  | 679.3 | 231.7 | 750.0 | 531.6 | 918.0 | 536.0 | 4.4 | 168.0 | 302.0 |  |  | 308 |  |  | 387 |  |  | 168 |
| Count |  | 150 | 150 | 150 | 150 | 150 | 150 | 150 | 150 | 150 |  |  | 150 |  |  | 150 |  |  | 150 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |