How is object matching different from object classification?

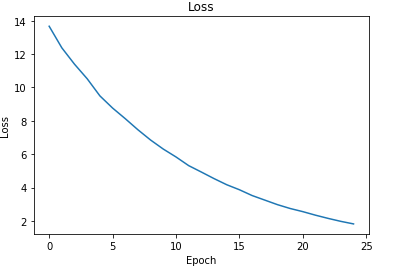
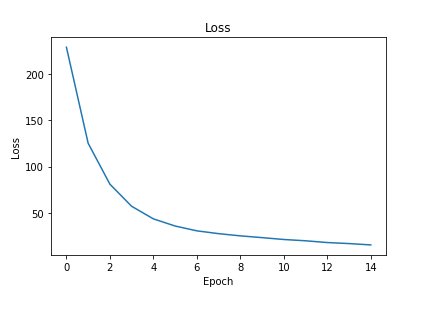
The major difference between object matching and object classification is in terms of dimensionality of search. For object matching there is one to one comparison. Thus, it is either if the object is present in second image or it is not present (thus more like binary class), whereas in the second case it is more about figuring out to which of the class the features mostly closely matches and then it will output a fixed class to which it closely matches.

One of the important constraints with object matching is the number of objects to be matched. For each pair of images there can be only a single object matched.

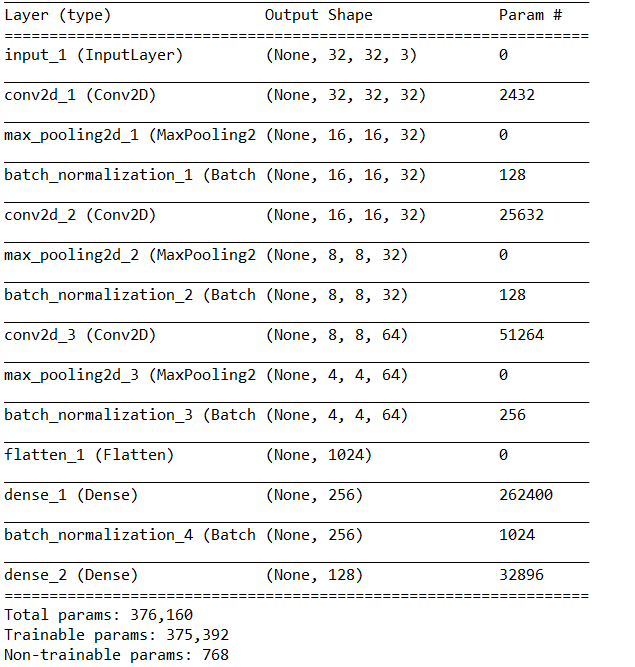
The objects to be matched may differ in terms of its rotation angle, scale or it may be translated across certain axis. Also, the illumination and brightness may vary for the two objects to be matched. Thus, while detecting the image, the features of the images used by model should have following properties:

* geometric variation
* photometric variation

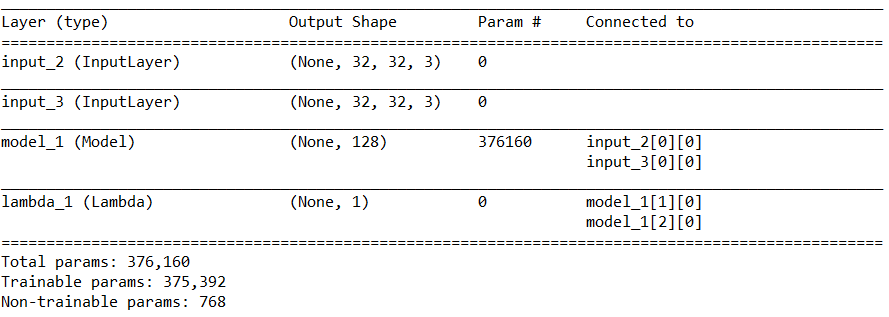
**Loss over epochs**



**Model Summary:**



**Similarity comparison layers**:



**Problems with loss and accuracy:**

The loss function requires to have difference between distance between anchor and positive and anchor and negative images. Later maximum of the difference and 0 should be taken. In order to do this, it is required to process anchor, positive and negative as a single example. To take difference between two. However this later requires that while inference phase also 3 requirements are passed, while inference requires to evaluate similarity between only two images.

Tried splitting it into anchor positive and anchor and negative and than later add the two. But slicing the tensor was difficult. Tried doing this using eval function. But the eval function threw dict error

Tried calculating loss directly by multiplying loss and labels and then calculating max value for each error. But that doesn’t take difference of a,p and a,n.

Working approach:

Used anchor, positive and negative as pair of input. Calculated distance between a and p, a and n. Calculated their difference between and returned that as output of Siamese network.

For calculating loss created a custom loss function in Keras called triplet loss. Which simply returns the mean of max(0,loss) for each example.

For calculating accuracy created a custom function which considers value as True if the diff between sim(a,p) and sim(a,n) is zero or less than zero else it is considered as false.

Finally to reload the model each time with custom objects we have to pass these function parameters as arguments(while loading the model)

**Issues:**

**Accuracy is decreasing over time**

When checked for predictions by performing inference on anchor, positive , anchor as pair and anchor negative, anchor as pair (since 3 inputs are req I have provided 3rd input as anchor so that the distance between them is zero and nothing is subtracted from distance between anchor and test image) it was found that the distance between a and p and distance between a and n were almost similar and most of them were close to margin.

One of the possible solutions observed is triplet selection:

It was hinted that triplet selection should be not completely random because if the distance between a and n is much larger than a and p than the network will not learn much effectively. Thus focus should be on picking hard triplets.

**Viewpoint :** Ourlosscontinuously decreased at steady rate which contradicts with the case of easy triplets in which loss doesn’t increase steadily

Also, one should make sure that triplets are not too difficult, else the difference between sim(a,p) and sim(a,n) would converge to margin

**Viewpoint:** Thisseems to be the case for us, but our selection process is completely random

Started working on creating contrastive loss model

Collect data for creating pairs

**Look at papers related to Metric learning**

**What are the similarity**

**Explainability using AI**

**Latest Vizviz paper**

**What kind of questions are asked by blind people**

**Applying to Clothing patterns using finger-mounted camera**