Metric Learning and Siamese Network

Computer Vision

Session Agenda

- Metric Learning
- Examples of Metric Learning
- Siamese Network
- Applications of Siamese Networks
- How to train Siamese Networks

Image Classification Task

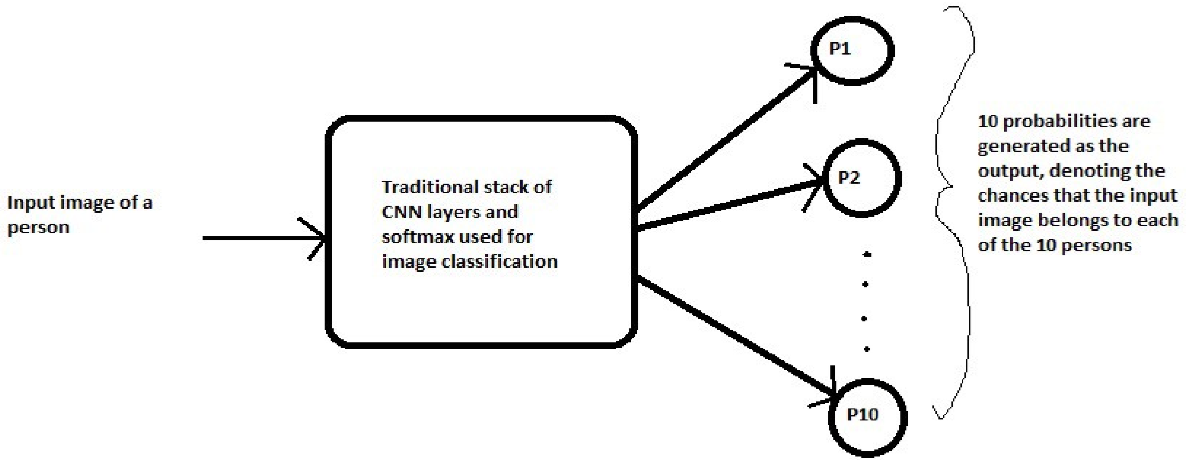
- Deep CNNs have become best performing methods for image classification tasks.
- In such Image related tasks using CNNs require a lots of labelled data which is one of the biggest limitations of using these algorithms.
- In many real- life applications, (eg: Building a face recognition model), collecting this much data is very difficult or not feasible.

Traditional Classification Task

Traditional Classification task - the input image is fed into a series of layers, and finally a probability distribution over all the classes (typically using a Softmax activation function).

- Two important points to be noted here -
  - Require large number of images for each class
  - If the network is trained only on, let's say, 3 classes of images, then we cannot expect to test it on any other class.
- If we want our model to classify the images of other class as well, then we need to first get a lot of images for that particular class and then we must re-train the model again.

Traditional Classification using CNN



Challenges in real world

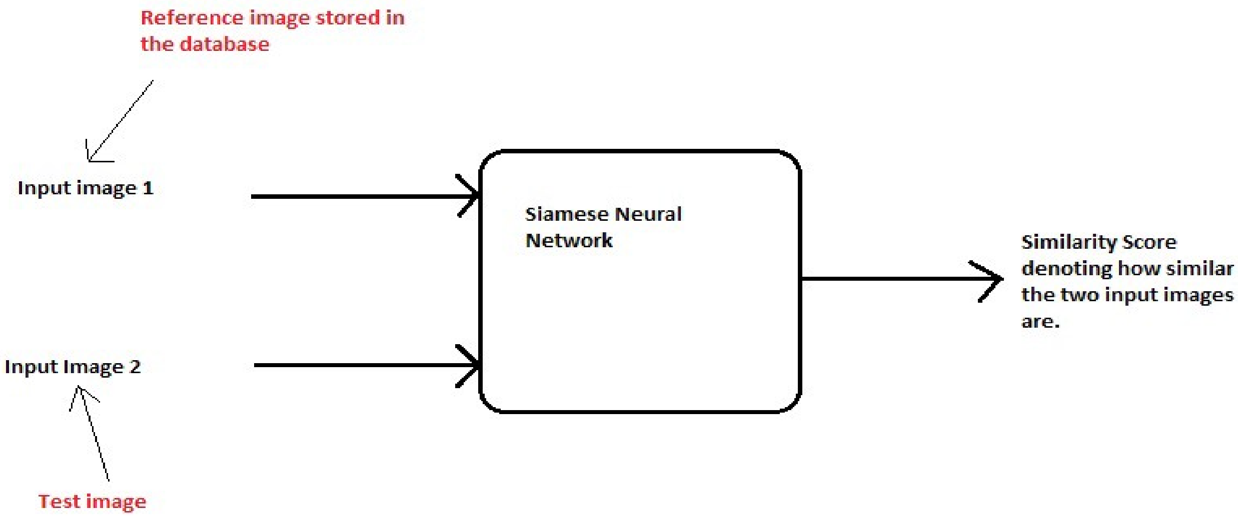
For some applications in real world, we neither have large enough data for each class and the total number classes is huge and it keeps on changing.

The cost and effort of data collection and periodical re -training is too high.

ONE SHOT LEARNING TO THE RESCUE!!

One-Shot Learning

One shot learning - requires only one training example for each class.



Metric Learning

Metric Learning

Metric is like a distance. It follows the following properties -

- Inverse of similarity
- It is symmetric
- It follows triangle inequality

Metric learning is the task of learning a distance function over objects.

Examples of Metric

If distance is considered, the objective is to **minimize** the distance measure.

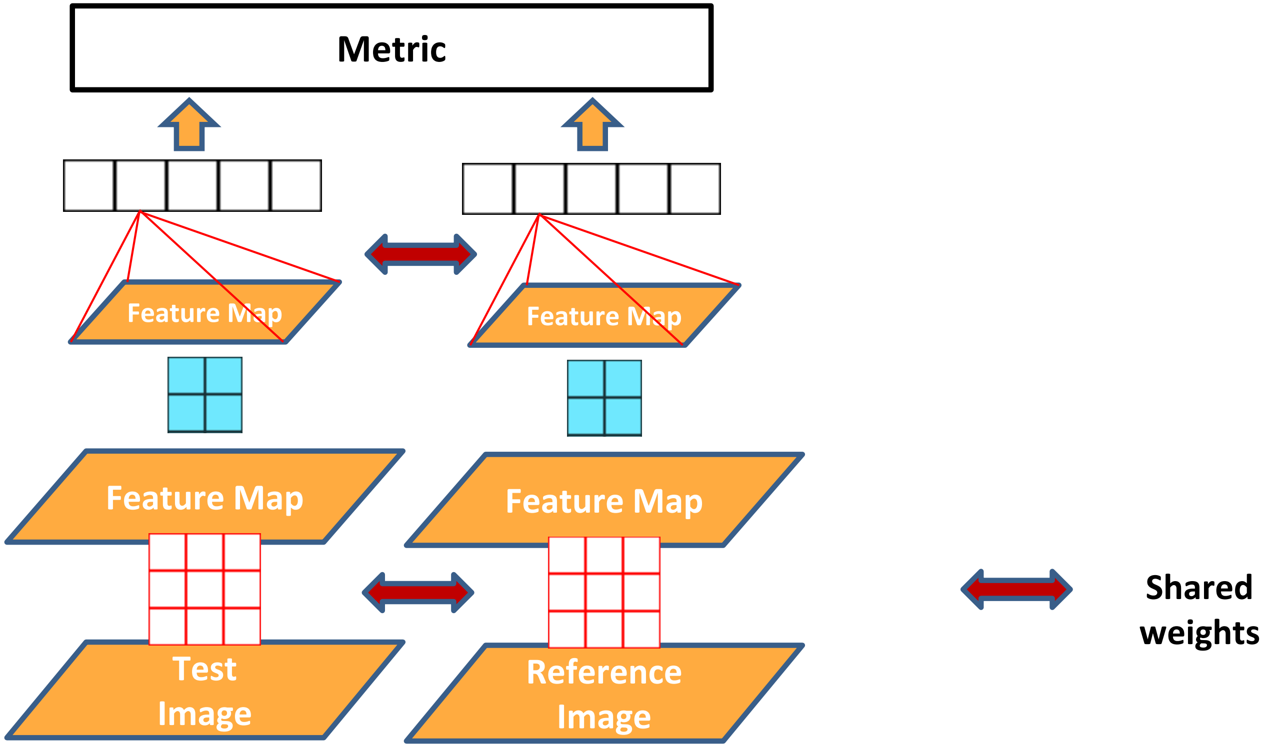
Euclidean and Manhattan distance.

If considering similarity, the objective is to **maximize** the similarity measure.

Dot product, RBF.

Siamese Network

Siamese network as metric learning



Siamese Network

- Siamese network is used to find we want to compare how similar two things are. Some examples of such cases are Verification of signature, face recognition
- Any Siamese network has two identical subnetworks, which share common parameters and weights.
- Siamese neural networks has a unique structure to naturally rank similarity between inputs.

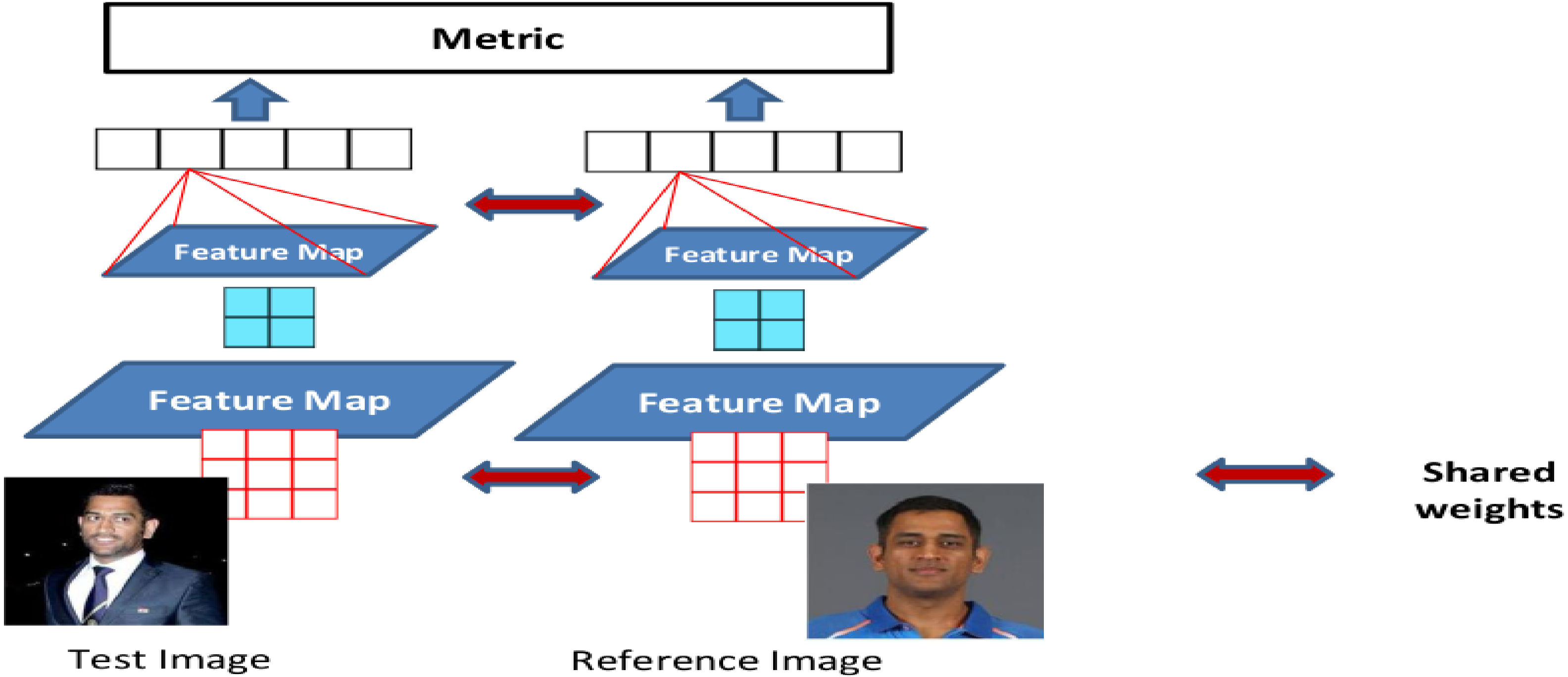
Applications of Siamese Network

- Signature verification
- Face verification
- Paraphrase scoring

And many more...

Generally, in such tasks, two identical subnetworks are used to process the two inputs, and another module will take their outputs and produce the final output.

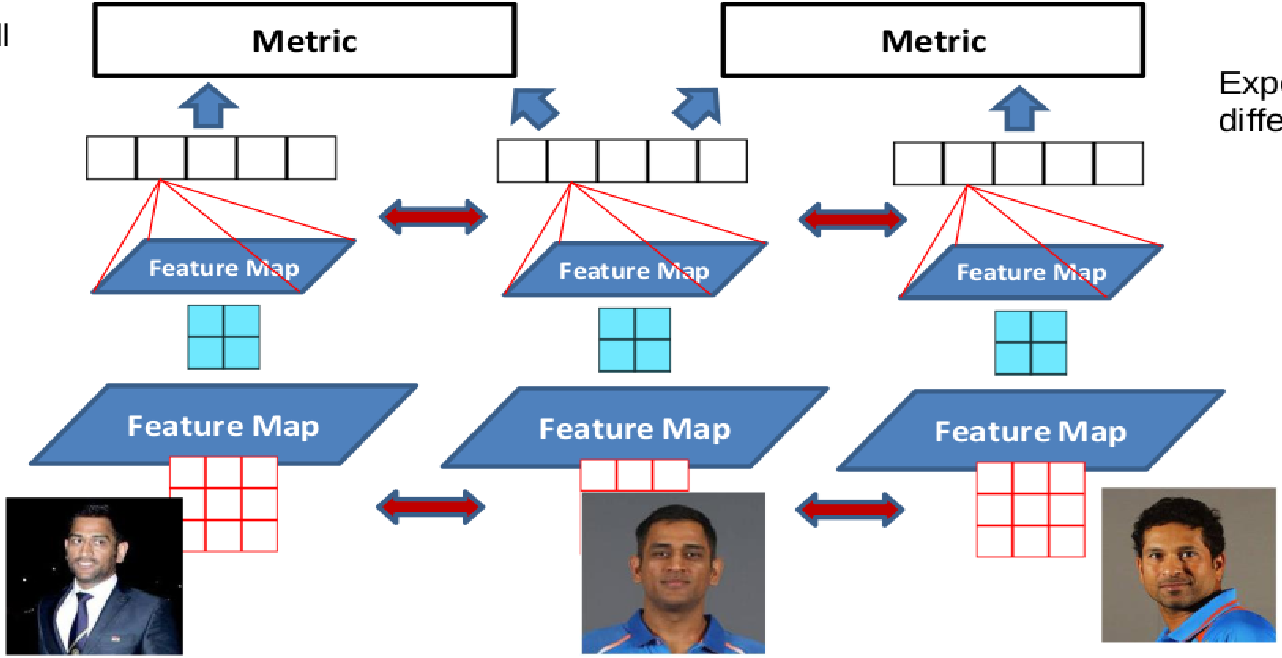
Example - Face Verification



Target values differ for similar and dissimilar pairs

Expect Small difference

Expect Large difference



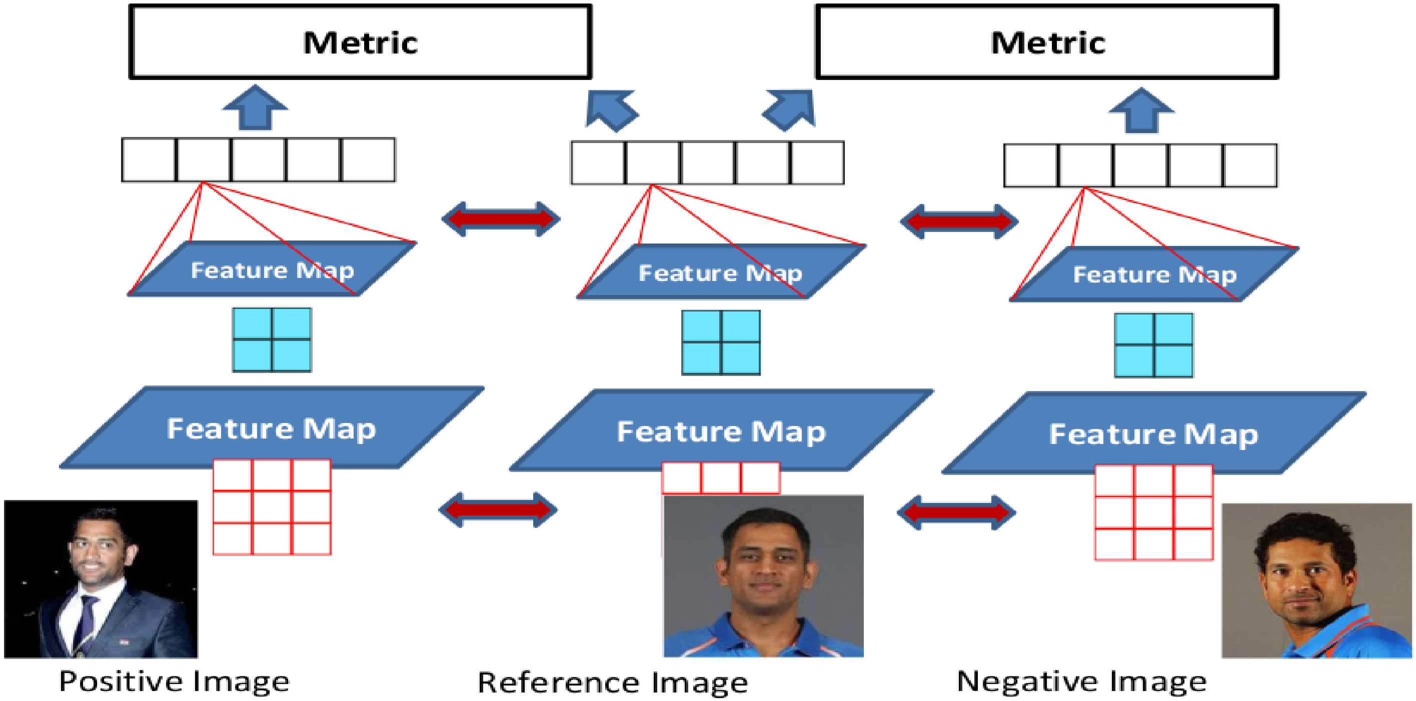
Q. Can we take pixel wise difference and say?

Positive Image

Reference Image

Negative Image

Or, the relative values are different



Two ways of viewing a metric

- Absolute terms (Regular Siamese training)
  - Distance (x_{ref}, x_+) = Low; Distance (x_{ref}, x_-) = High
  - Similarity (x_{ref}, x_+) = High; Similarity (x_{ref}, x_-) = Low
- Relative terms (Triplet Siamese training)
  - Distance (x_{ref}, x_-) – Distance (x_{ref}, x_+) > Margin
  - Similarity (x_{ref}, x_+) – Similarity (x_{ref}, x_-) > Margin
- Class probability was based on a single input
  - ClassProb (x,c) = High when $x \in c$; otherwise low

Some distance and similarity measures

- Distances examples
  - L2 norm of difference (Euclidean distance)
  - L1 norm of difference (City-block/Manhattan dist.)

- Similarity examples
  - Dot product
  - Arc cosine
  - Radial basis function (RBF)

Some distance and similarity measures

- Distances examples
  - $||(f(x_i) - f(x_j)||_2^2$
  - $|(f(x_i) - f(x_j)|_1$

- Similarity examples
  - $f(x_i)^T f(x_j)$ or $f(x_i) \cdot f(x_j)$
  - $f(x_i) \cdot f(x_j)$ / $(||f(x_i)|| \, ||f(x_j)||)$
  - $\exp(- ||x_i - x_j||^2/\sigma^2)$

Triplet Loss

- You can train the network by taking an anchor image and comparing it with both a positive sample and a negative sample.

- The dissimilarity between the anchor image and positive image must low and the dissimilarity between the anchor image and the negative image must be high.

Triplet Loss Function

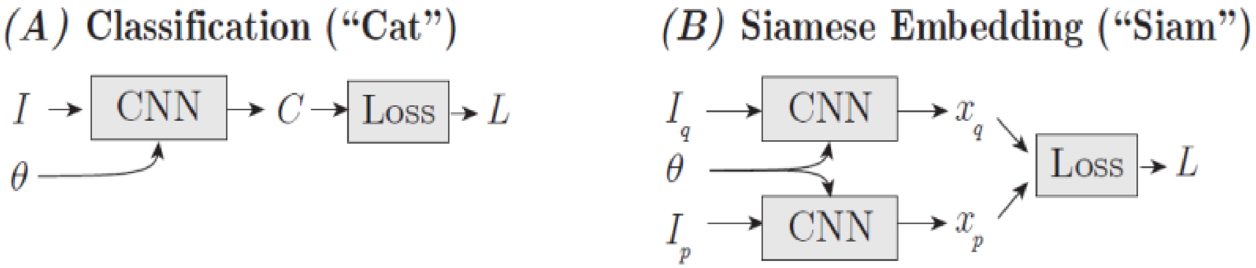
$$\mathcal{L} = max(d(a,p) - d(a,n) + margin, 0)$$

“a” - represents the anchor image
“p” - represents a positive image
“n” - represents a negative image
margin - is a hyperparameter. It defines how far away the dissimilarities should be.

Triplet Loss Function

greatlearning
Learning for Life

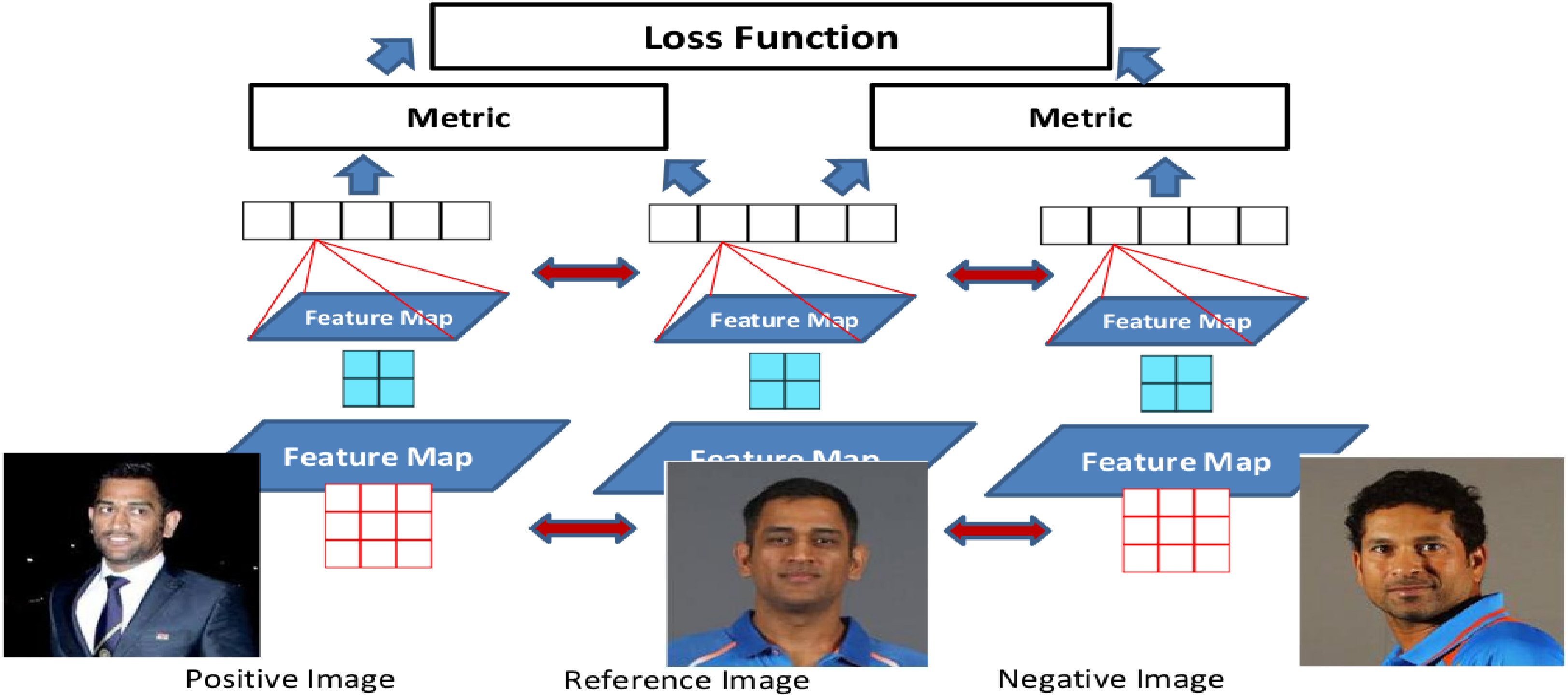
By using this loss function we calculate the gradients and with the help of the gradients, we update the weights and biases of the Siamese network.



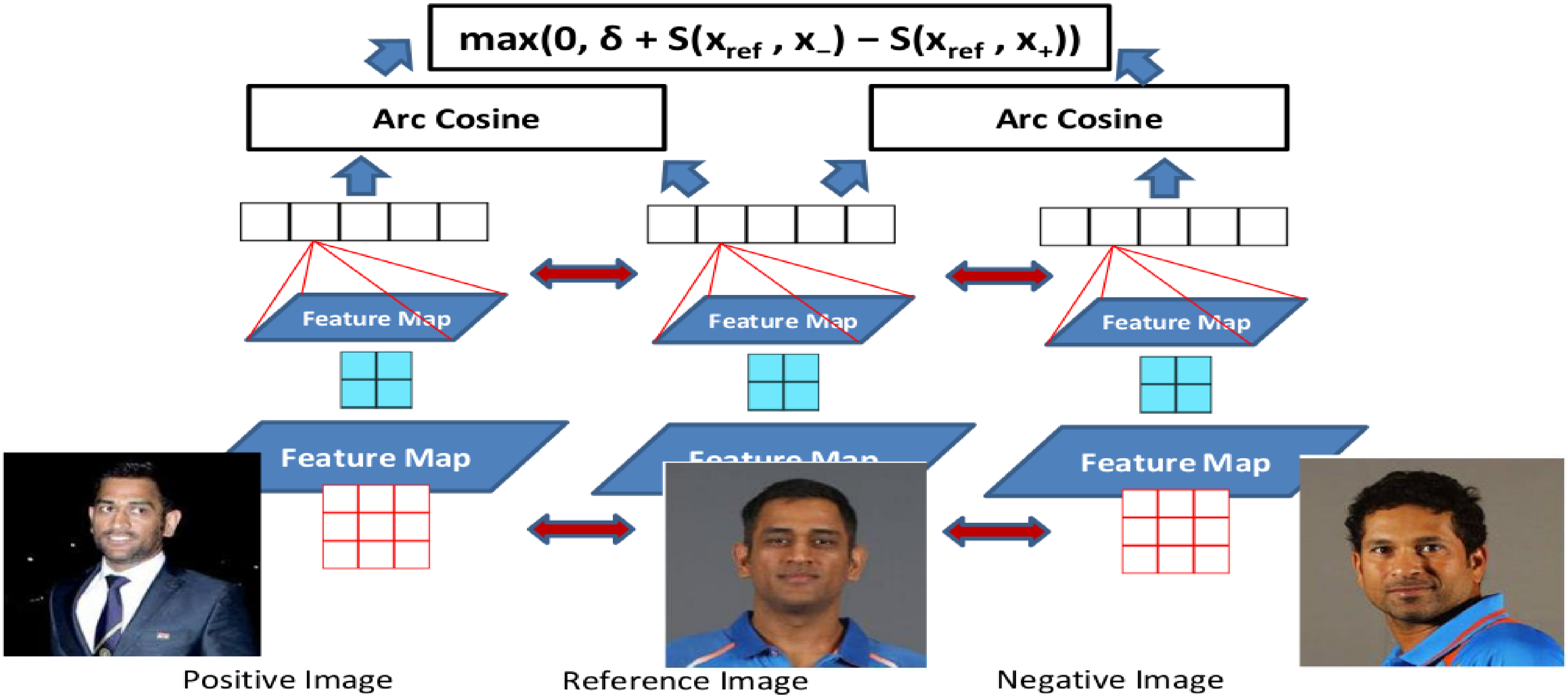
Paper o

Interesting Read - https://bamos.github.io/2016/01/19/openface-0.2.0/

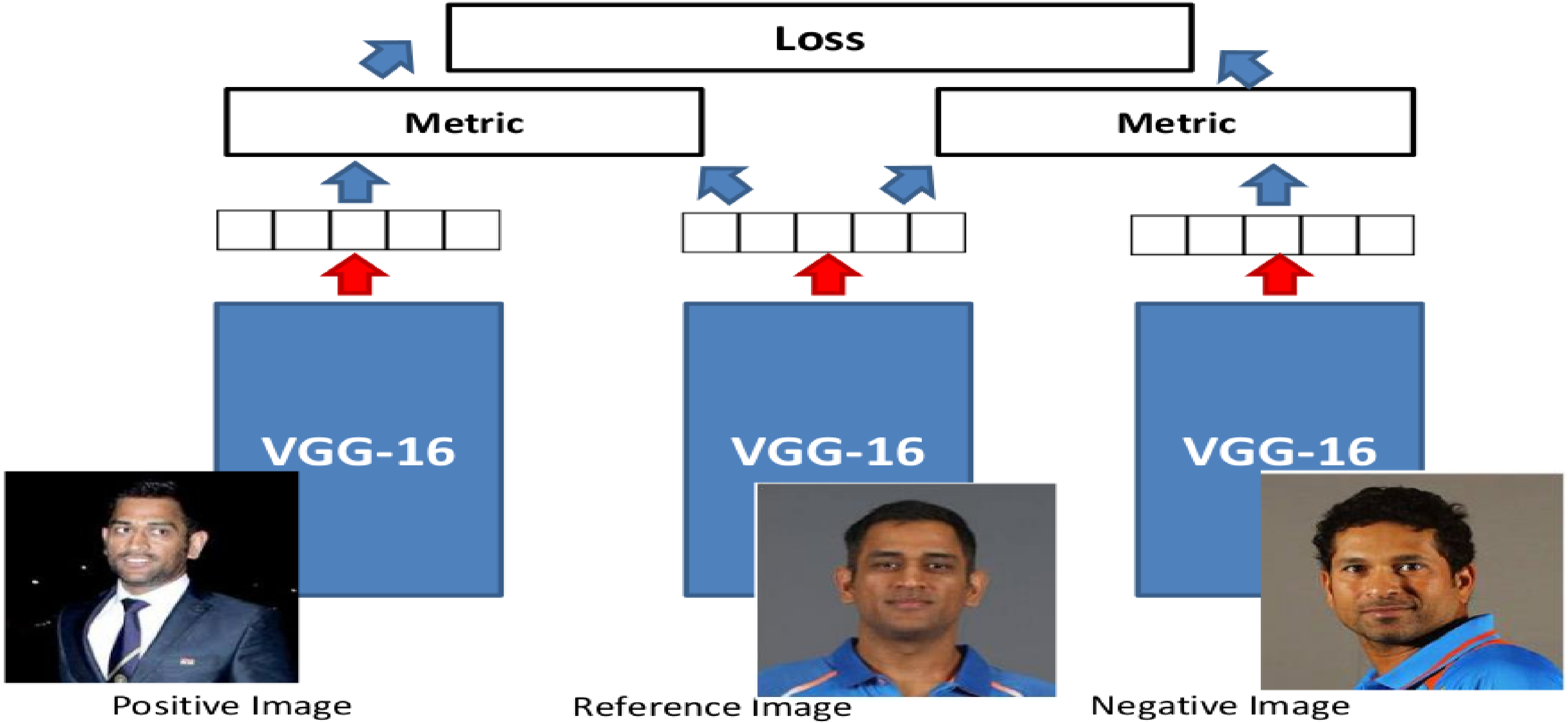
Loss gradient is propagated back



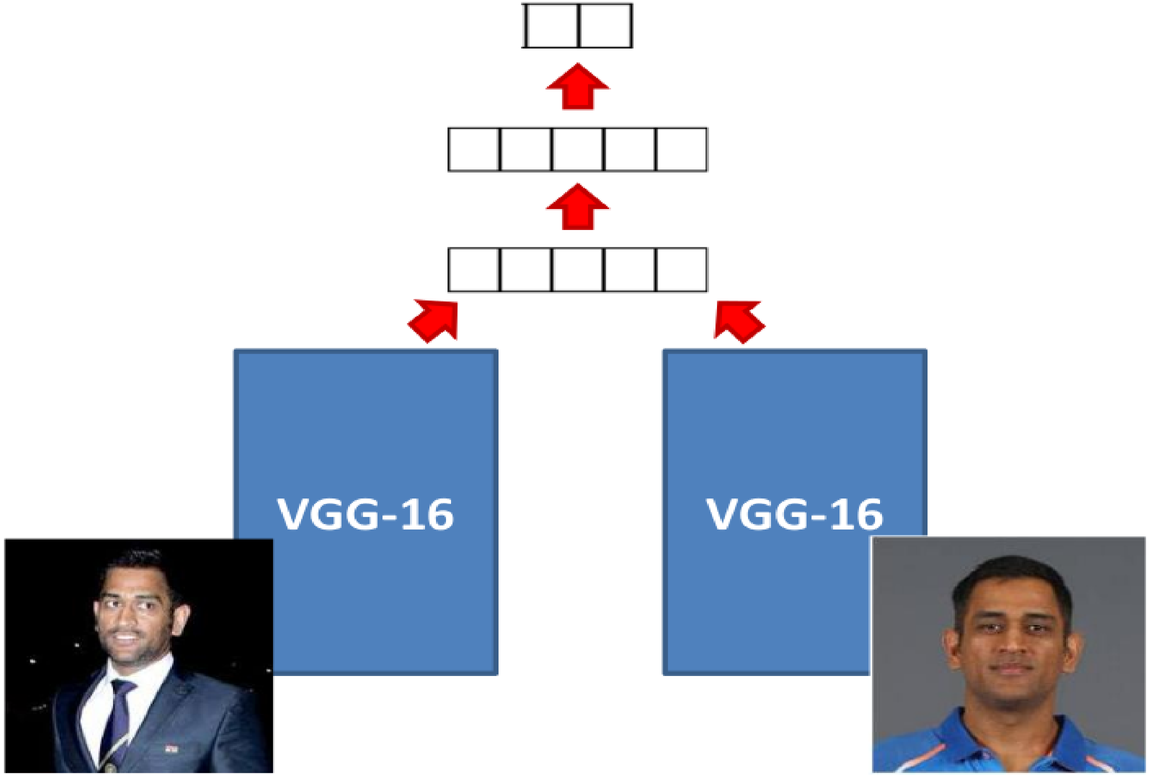
Loss gradient is propagated back



Pre-trained networks can be used



Some joint layers can also be added



Test Image

Reference Image

Face Recognition
Face Verification
Script Recognition
Signature verification

VGG or VGG-Face

Q. Can joint layers approach be symmetric n positive semi-definite?

Summary

We have learnt about,

- Metric Learning
- Siamese Network
- Applications of Siamese Networks

Thank you!

Happy Learning :)