Localizing and Orienting Street Views using Overhead Imagery
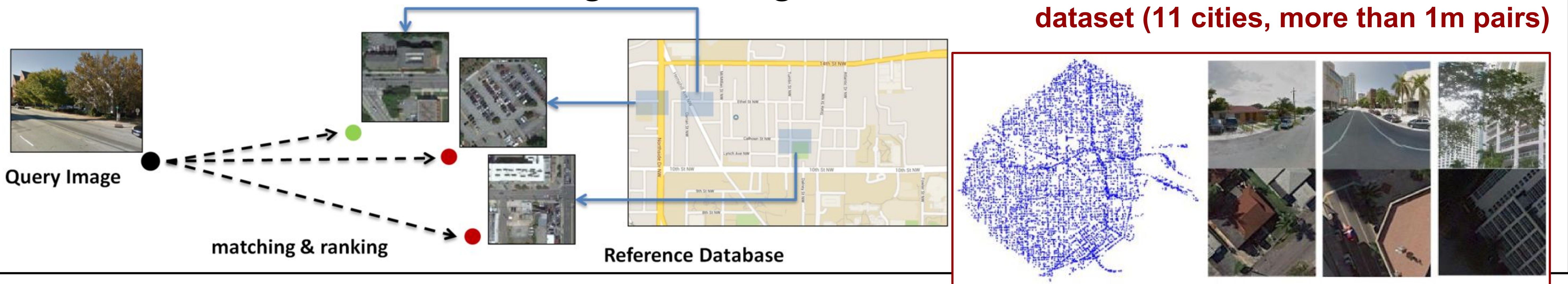
Nam Vo and James Hays, Georgia Tech

1. Problem formulation: image ranking
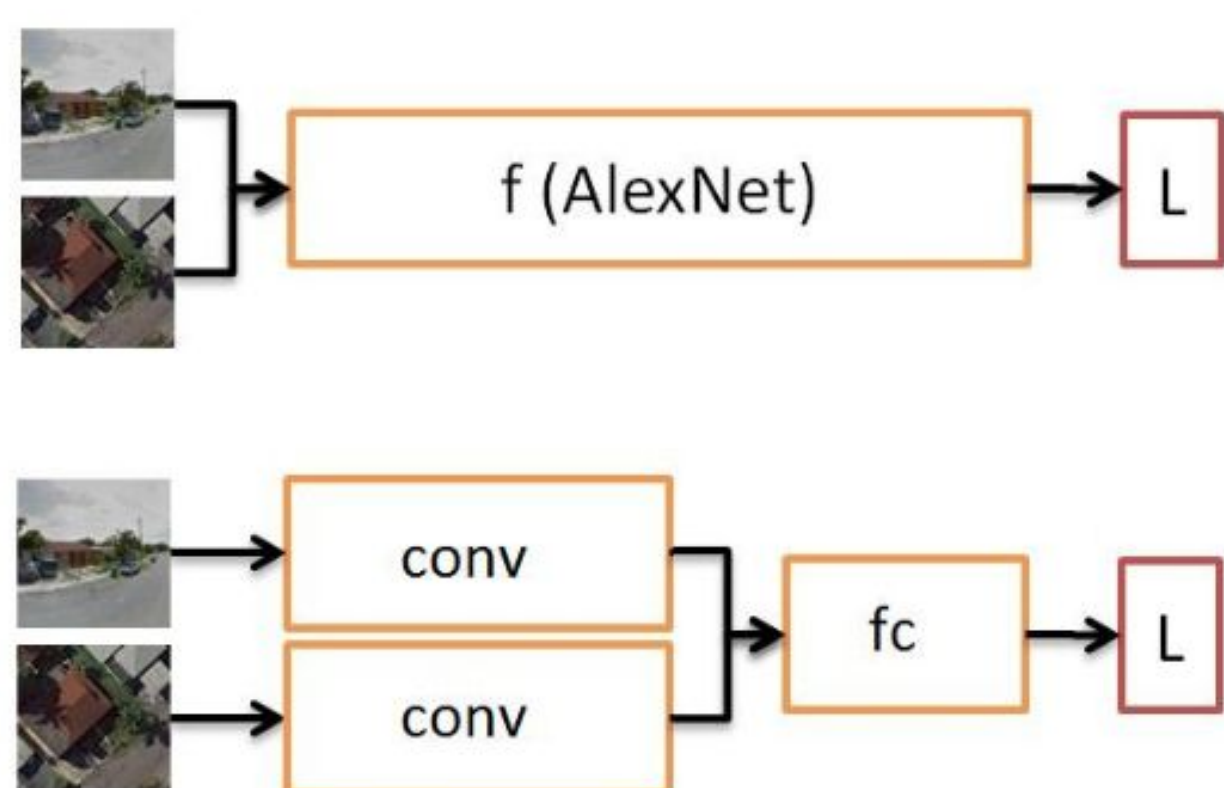

Query Image

matching & ranking

Reference Database

GTCrossView: a new large scale cross view dataset (11 cities, more than 1m pairs)

2. Deep learning approach

Distance based logistic (DBL) layer for Pair and Triplet
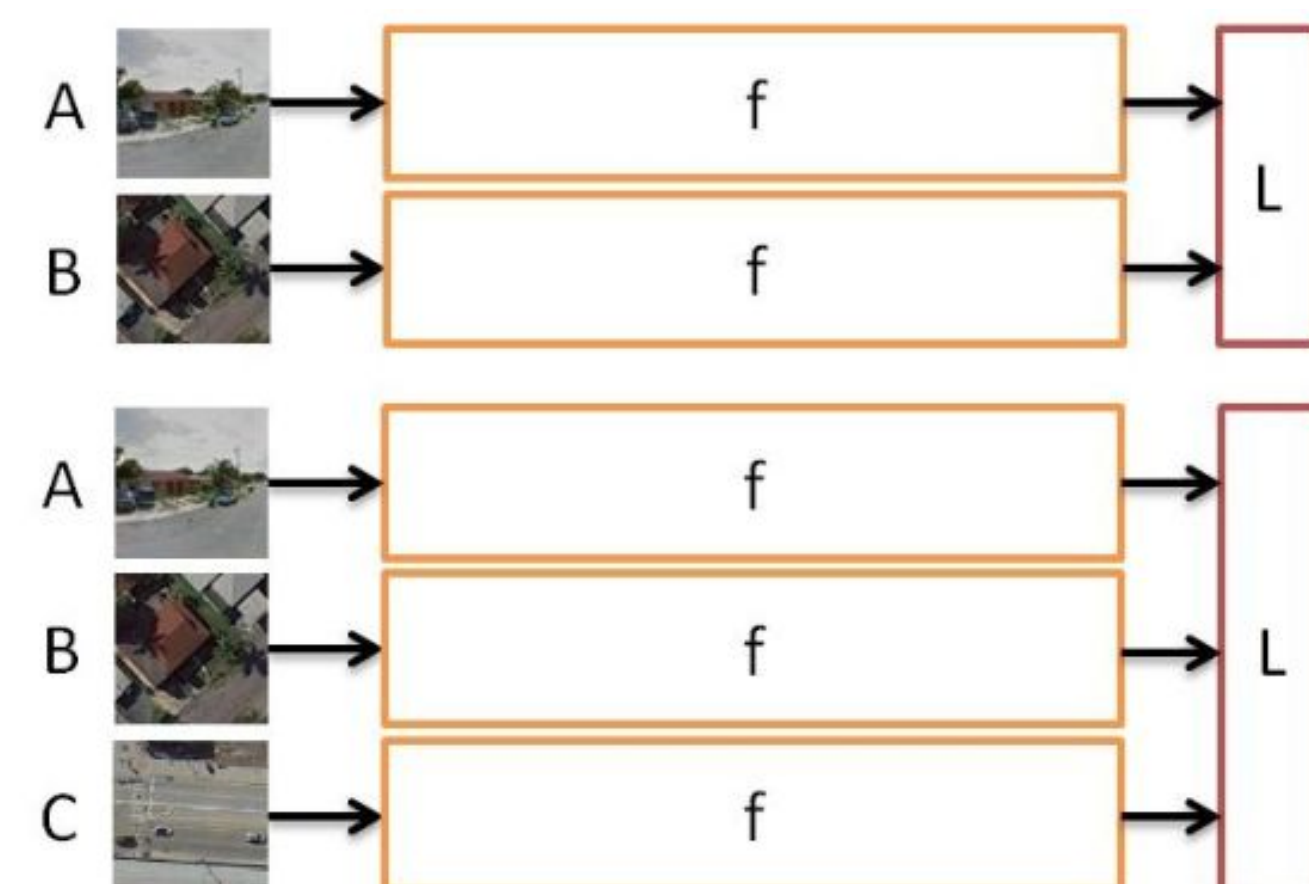
Classification
input both images and classify whether they are a match or not
- **Classification Alexnet**
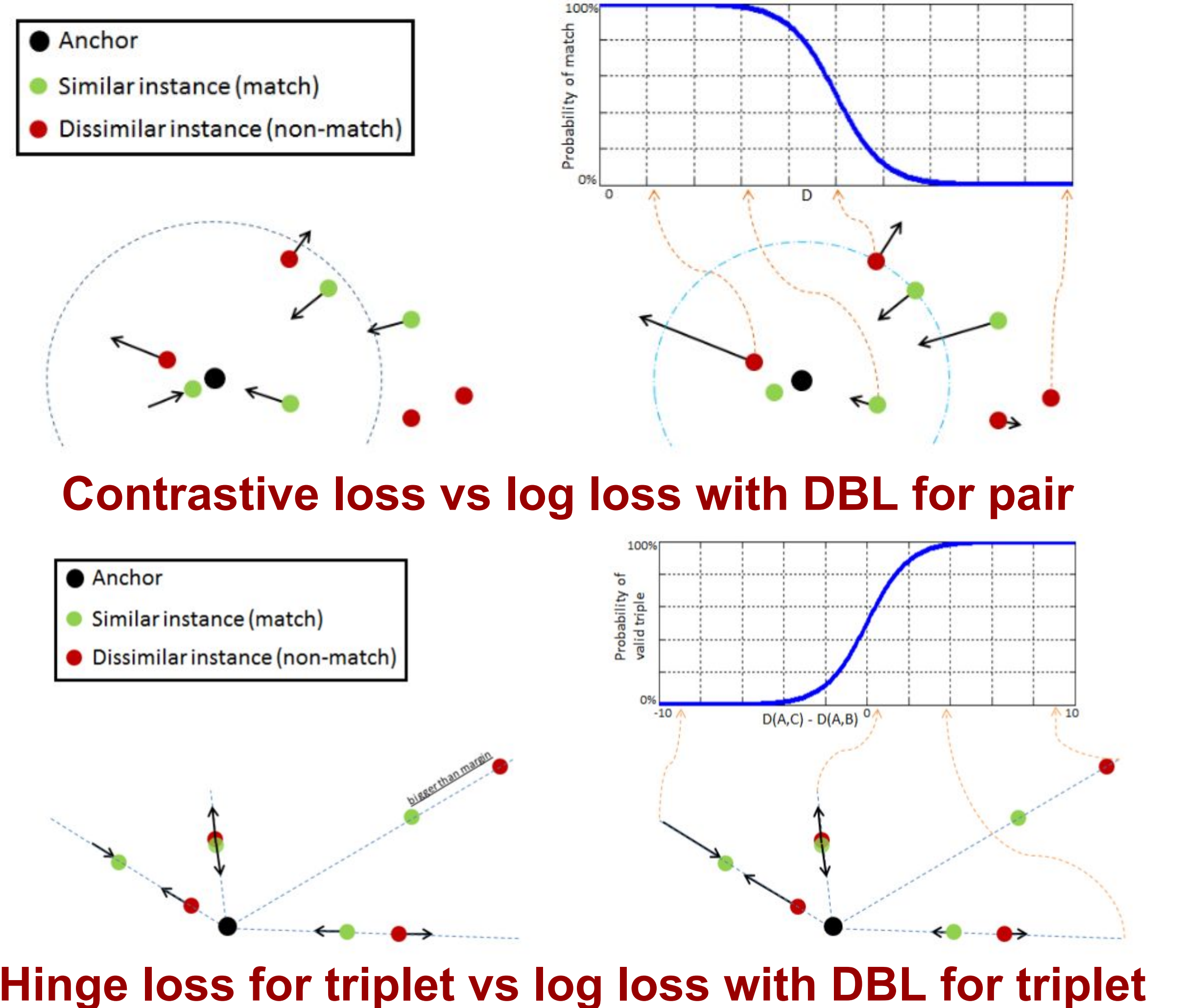- **Siamese-classification hybrid network**

Representation learning
embed both images in the same feature space and compare with Euclidean distance.
- **Siamese network**
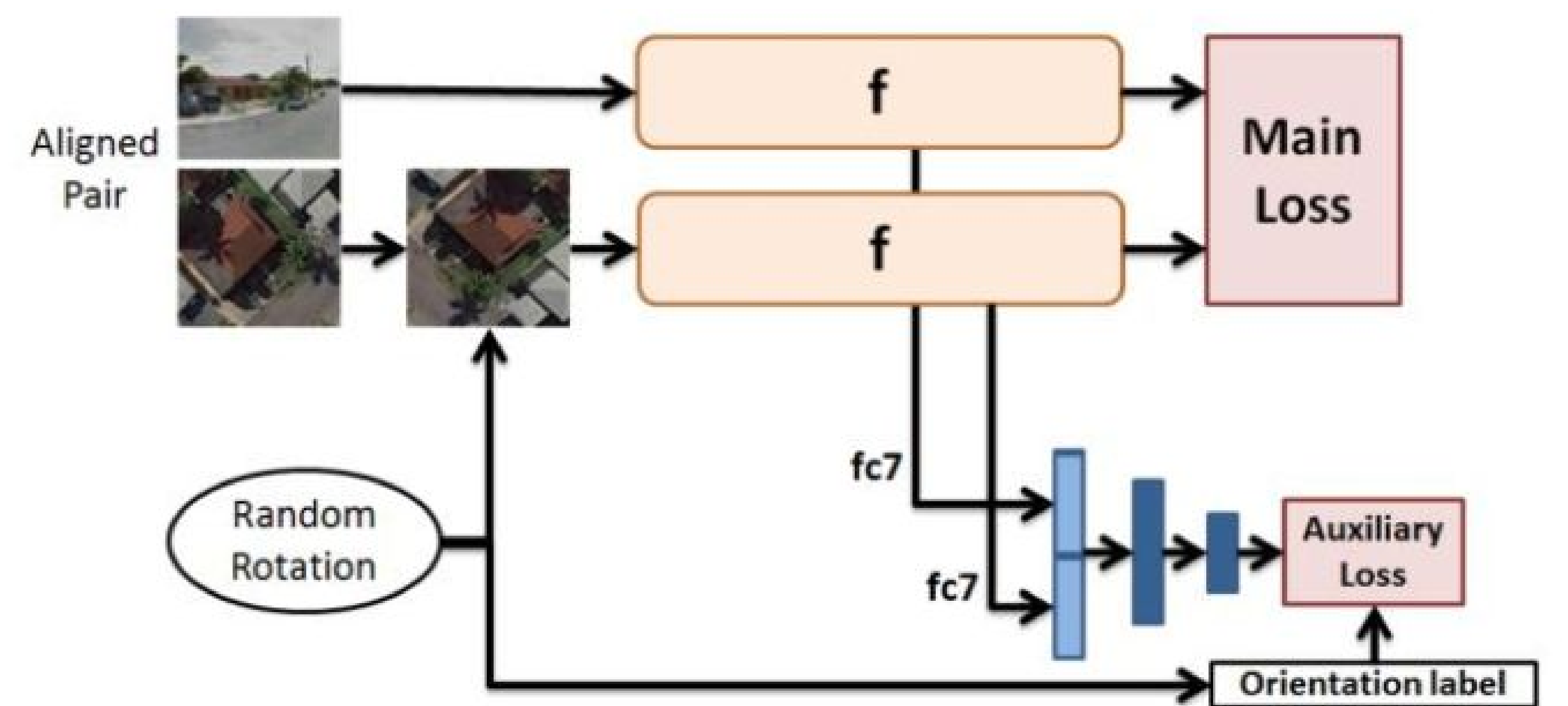- **Triplet (ranking) network**

- Anchor
- Similar instance (match)
- Dissimilar instance (non-match)

Contrastive loss vs log loss with DBL for pair

Hinge loss for triplet vs log loss with DBL for triplet
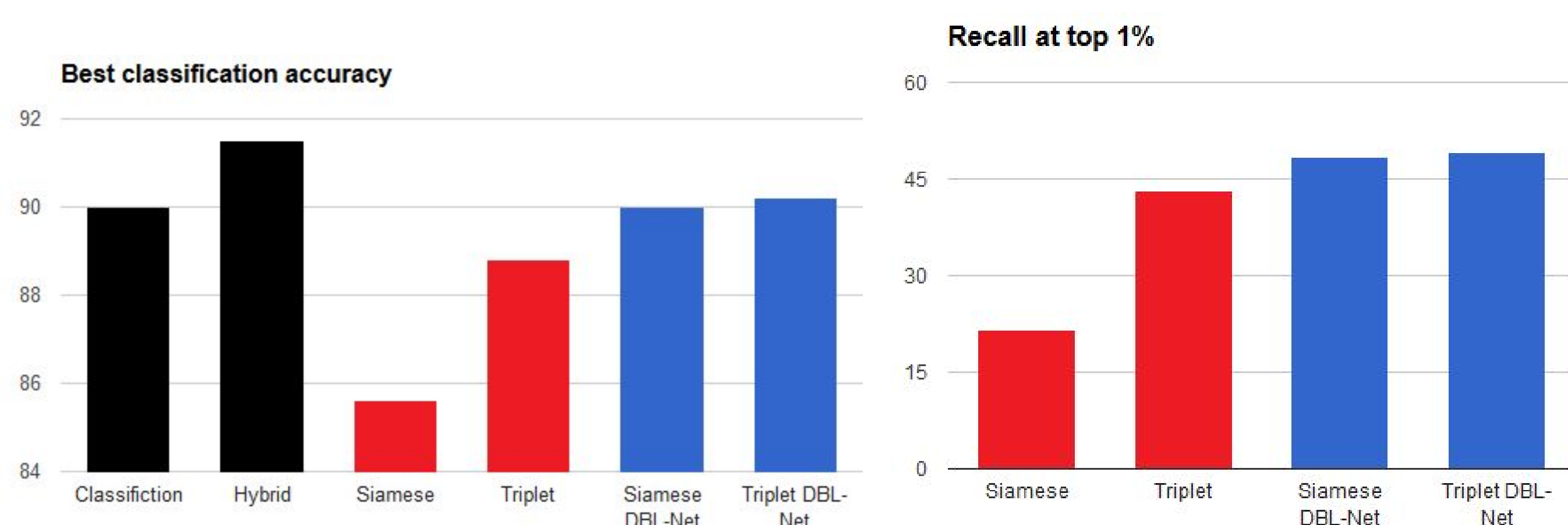
3. Rotational invariant matching

As orientation alignment is not available during testing, the desired representation should be invariant to rotation of overhead image, which is achieved by:

- Data augmentation by random rotation
- Better matching with multi-rotation testing and feature averaging
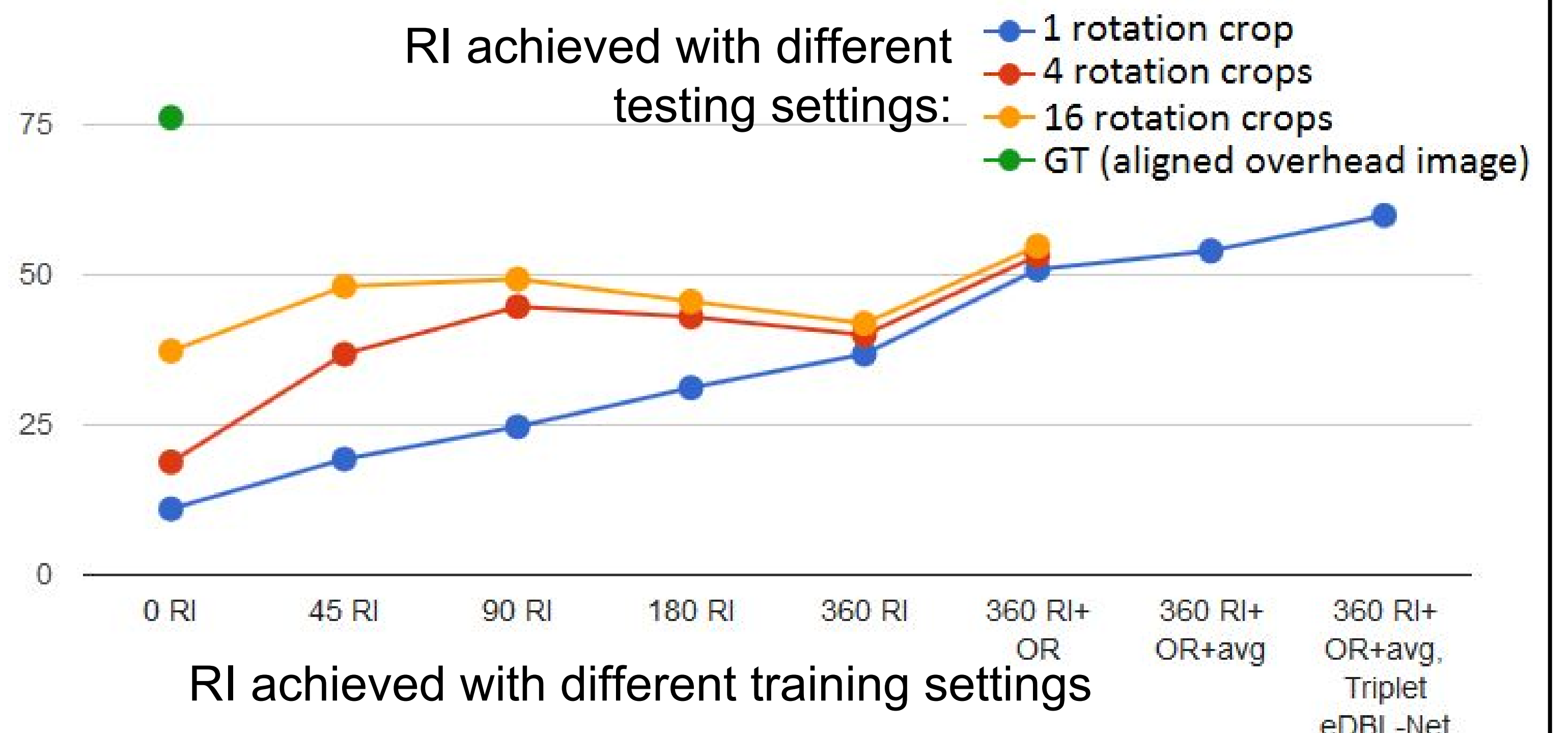- Learning better representation with orientation regression

Aligned Pair — f — Main Loss

Random Rotation

fc7 — fc7 — Auxiliary Loss

Orientation label

4. Experiments

Architecture comparison:

Best classification accuracy

Recall at top 1%

- Classification networks: loose upper bound
- Triplet: significantly better than Siamese
- **DBL-Net: additional boost in performance**

Ranking performance with Rotational Invariance

RI achieved with different testing settings:
- 1 rotation crop
- 4 rotation crops
- 16 rotation crops
- GT (aligned overhead image)

RI achieved with different training settings

0 RI, 45 RI, 90 RI, 180 RI, 360 RI, 360 RI+OR, 360 RI+OR+avg, 360 RI+OR+avg, Triplet eDBL-Net

Exhausting mini-batch sampling: improve performance & convergence rate