Local Feature Matching

2016-10-27

Presented by: Cheng-Hao Tu

Outline

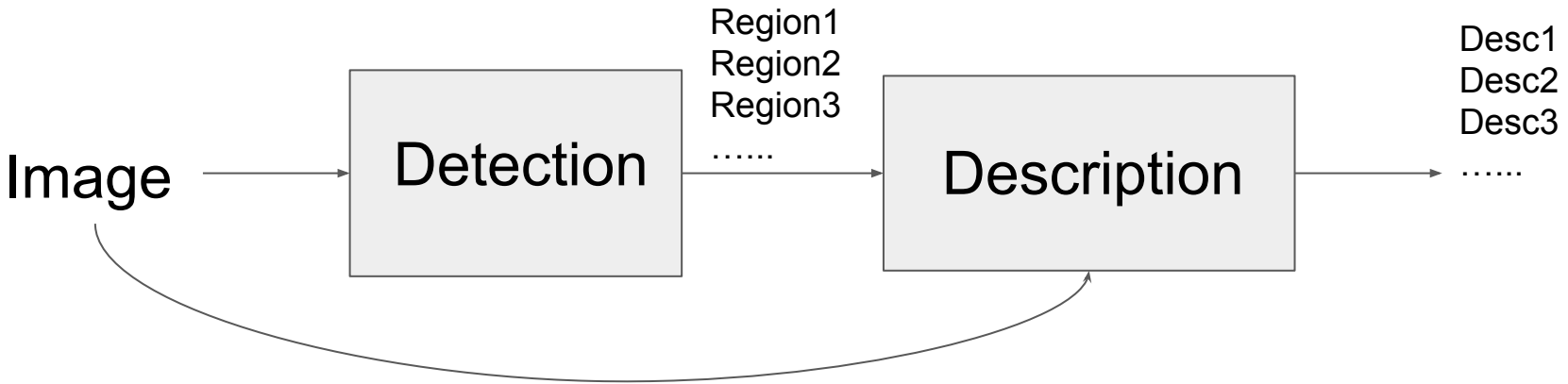
- Background
- Basic Pipeline
- Method
- Experiment
- Summary and Future

Background

- Panorama
  - Given multiple images, try to reconstruct the whole scene
- Popular method
  - SIFT

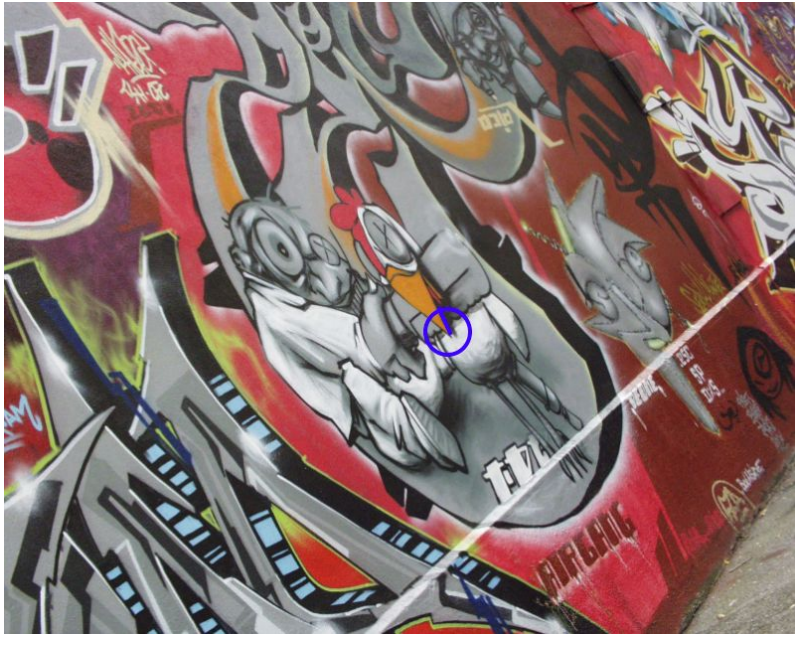
Basic pipeline

- Detector + Descriptor



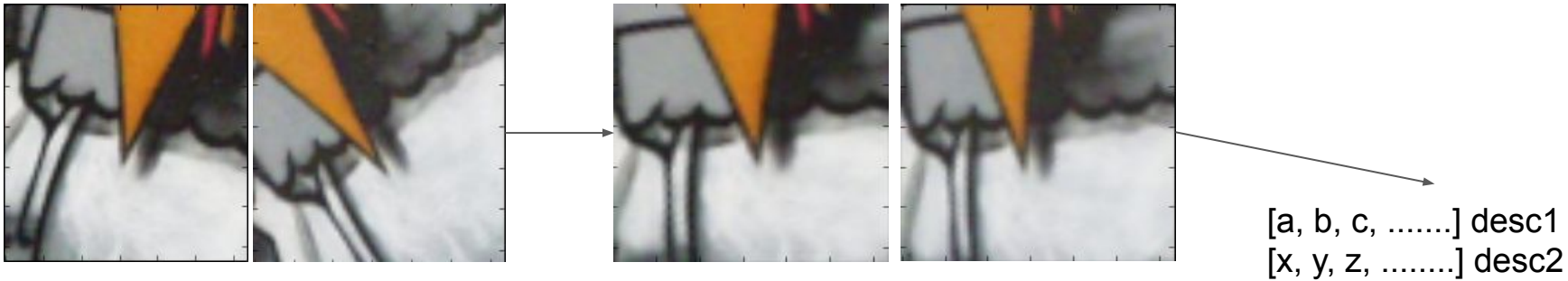
Basic pipeline

- Detector
  - Propose some key points on image
  - Estimate scale and rotation for each point



Basic pipeline (cont’d)

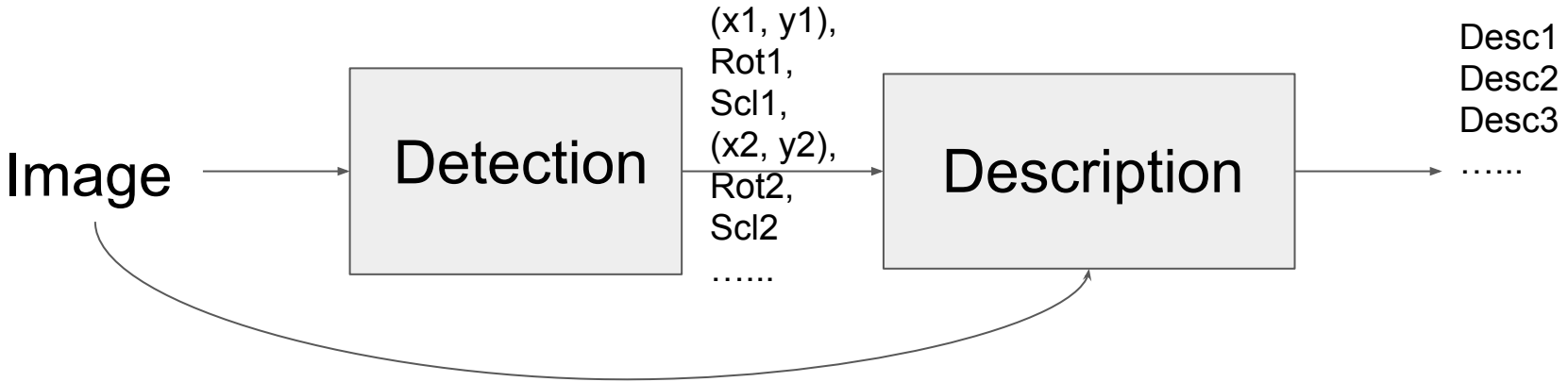
- Descriptor
  - Extract and normalize the the regions
  - Obtain a description for each region



- Match regions according to their descriptors
  - Calculate Distance(desc1, desc2) to determine whether they are correspondences

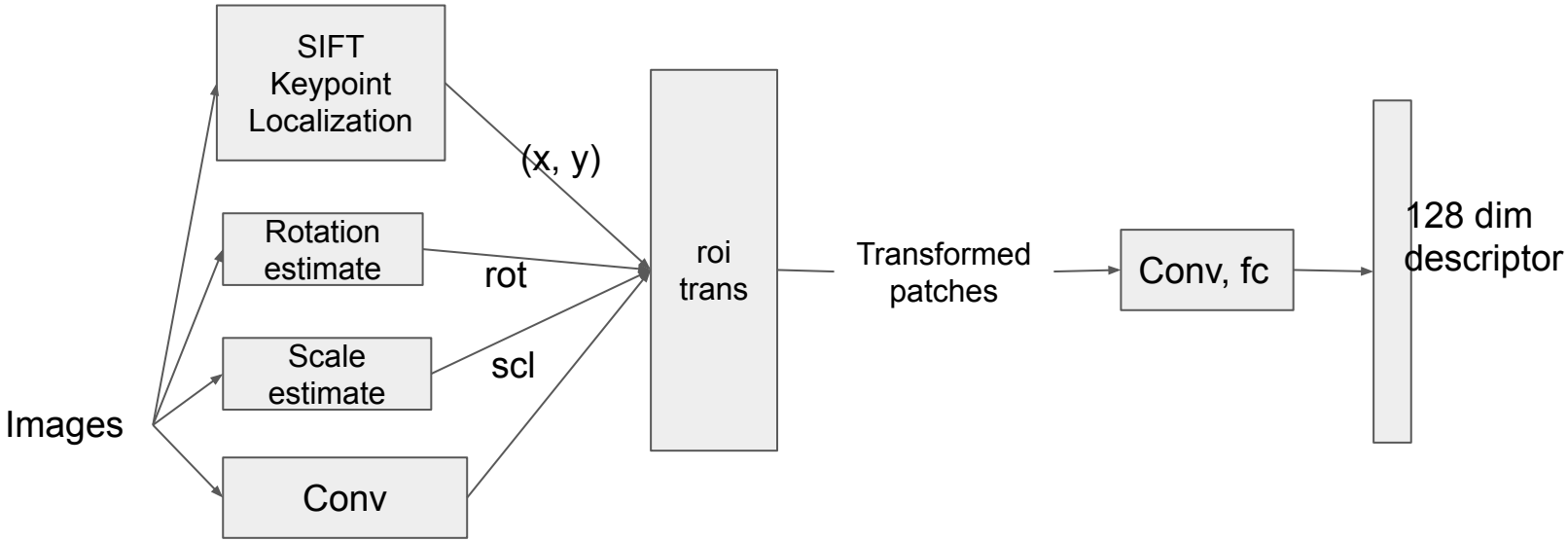
Basic pipeline (Summary)

- Detector + Descriptor
  - Location, rotation, scale of key points
  - Extracted regions
  - Obtain descriptors for regions to determine the correspondences



Method

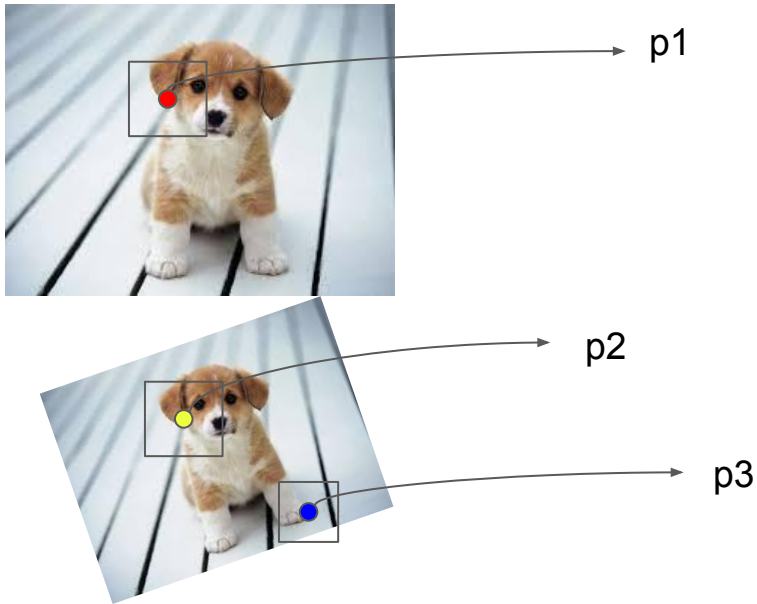
- In fact, SIFT detector can be simulated by CNN
- Testing Architecture



Jaderberg, Max, Karen Simonyan, and Andrew Zisserman. "Spatial transformer networks." *Advances in Neural Information Processing Systems*. 2015.

Method (cont'd)

- Training data
  - Unsupervised learning
  - Using augmented data for training
    - rotate, scale, lighting

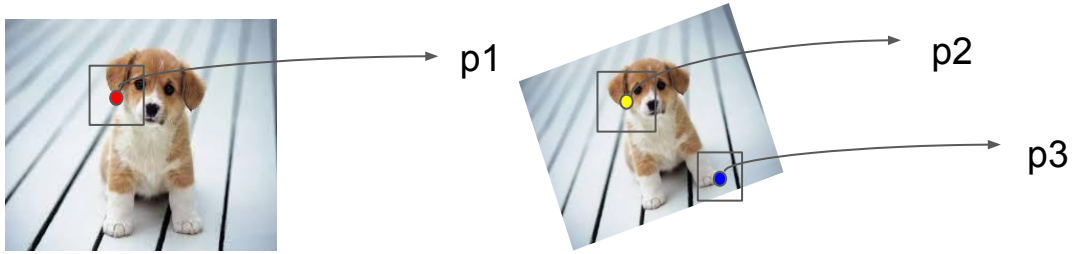


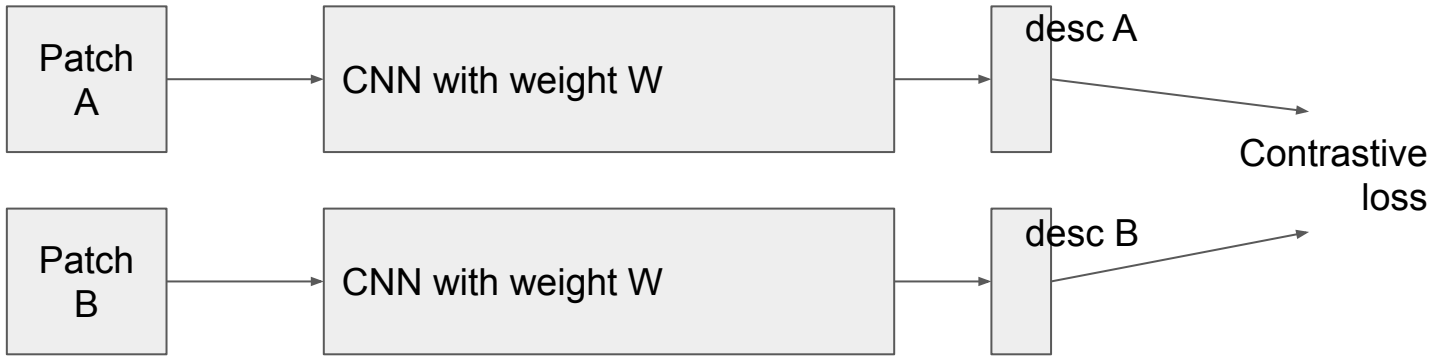
distance(desc(p1), desc(p2)) the smaller the better

distance(desc(p1), desc(p2)) the bigger the better

Method (cont'd)

- Training architecture
  - Siamese Network



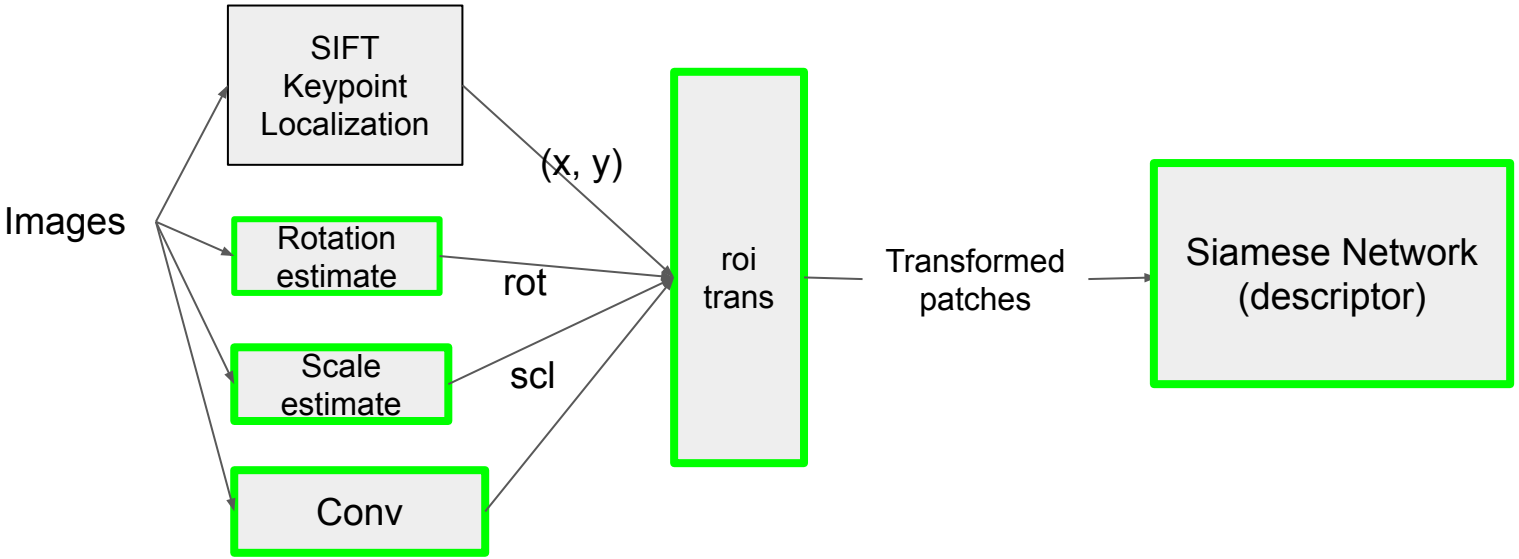


$$L = \begin{cases} distance(A,B)^2 & if\ A\ and\ B\ are\ the\ same \\ max(0, m - distance(A,B))^2 & if\ A\ and\ B\ are\ different \end{cases}$$

Chopra, Sumit, Raia Hadsell, and Yann LeCun. "Learning a similarity metric discriminatively, with application to face verification." *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. Vol. 1. IEEE, 2005.

Method (cont'd)

- Training architecture
  - Green: can do backpropagation

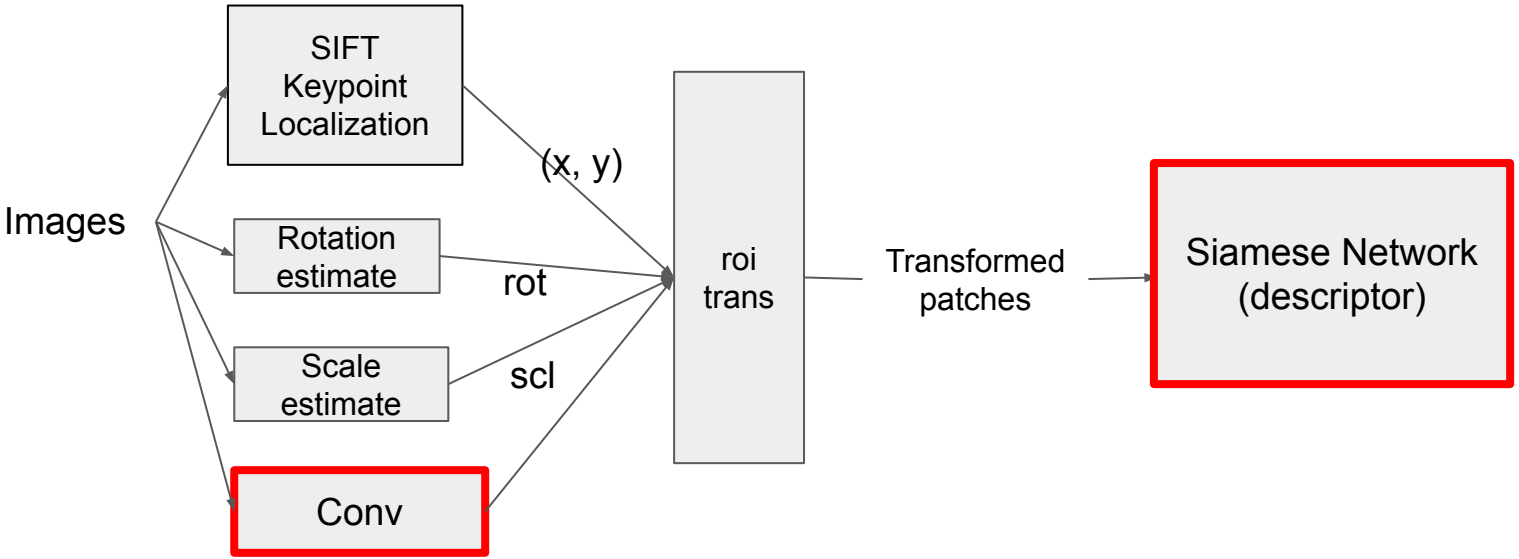


Experiment

- Train on about 8000 images Flicker images (about 48000 different keypoints , 480000 different patches, total 240000(pos)+480000(neg, sampled) pairs per epoch) and Validate on 100 images in MS COCO (500 keypoints for each image)

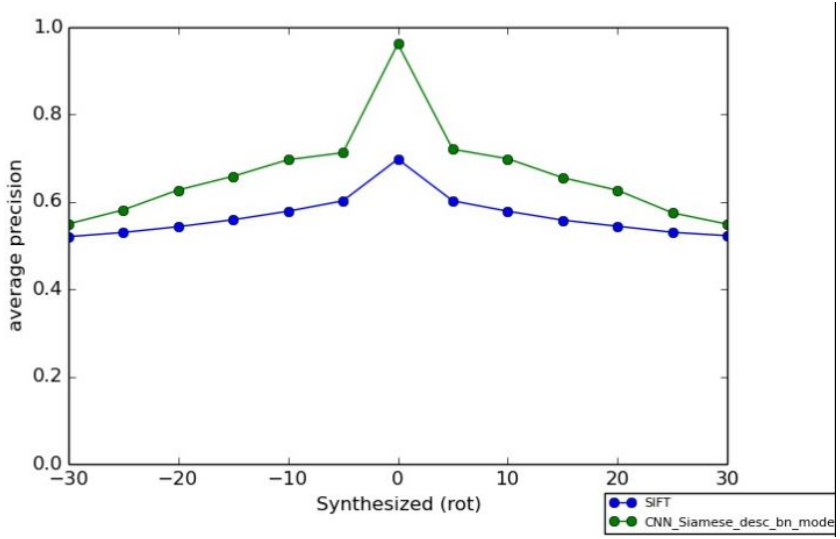
Experiment (cont'd)

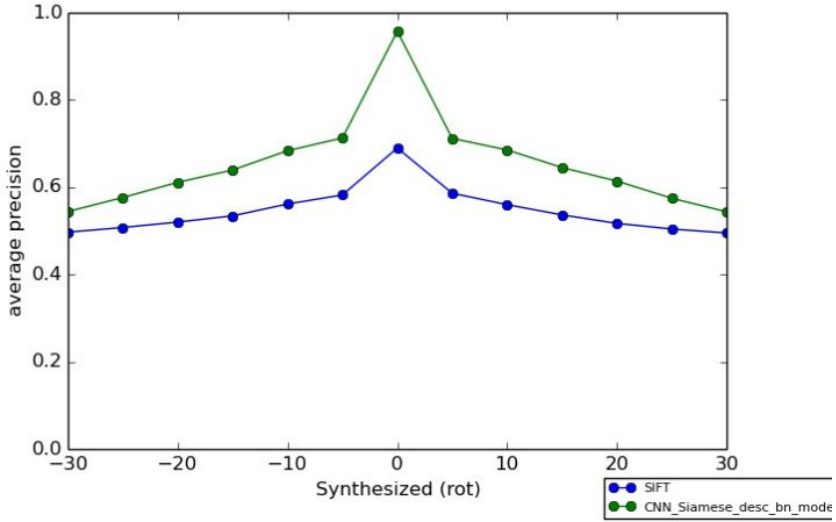
- Training Scheme
  - First, fix rot, scl, location (initialized as an approximation to sift) and update descriptor



Experiment (cont'd)

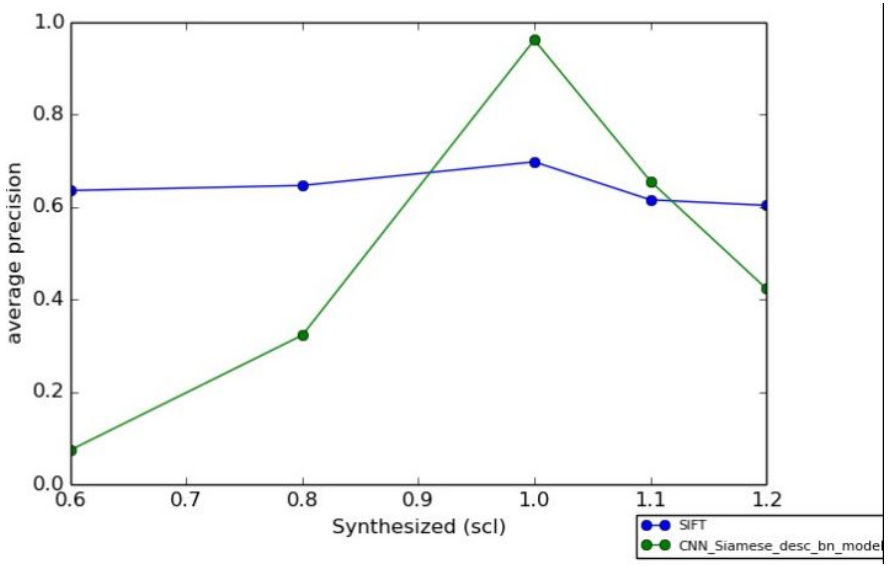
- For rotation transformation
  - Left: COCO IMG (Validation), Right: Flicker IMG (Train)

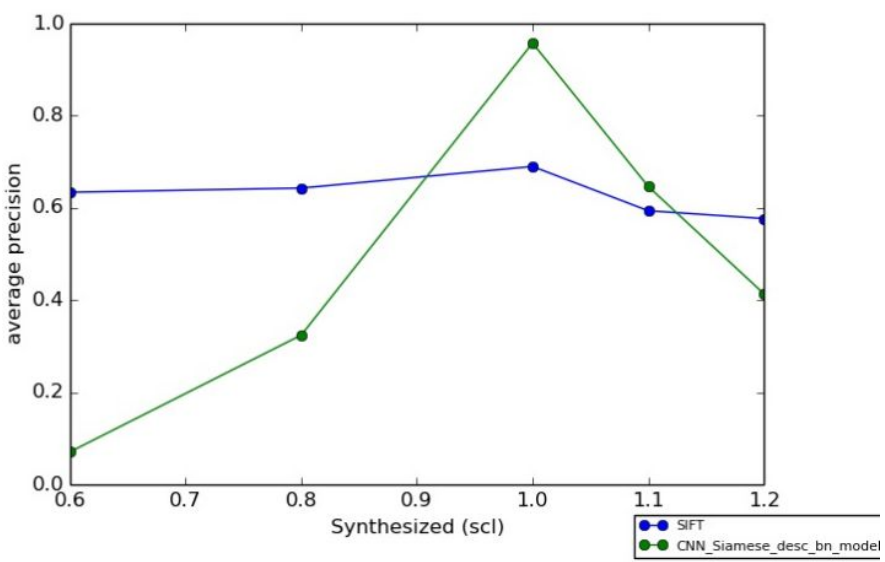




Experiment (cont'd)

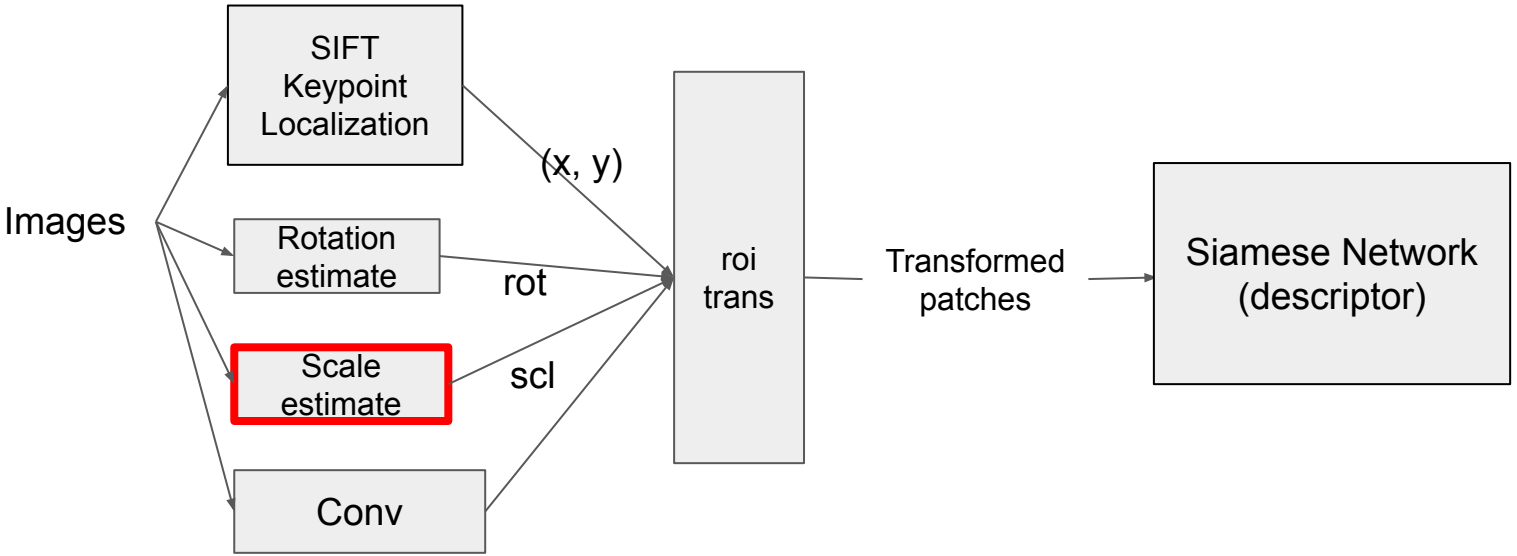
- For scale transformation
  - Left: COCO IMG (Validation), Right: Flicker IMG (Train)





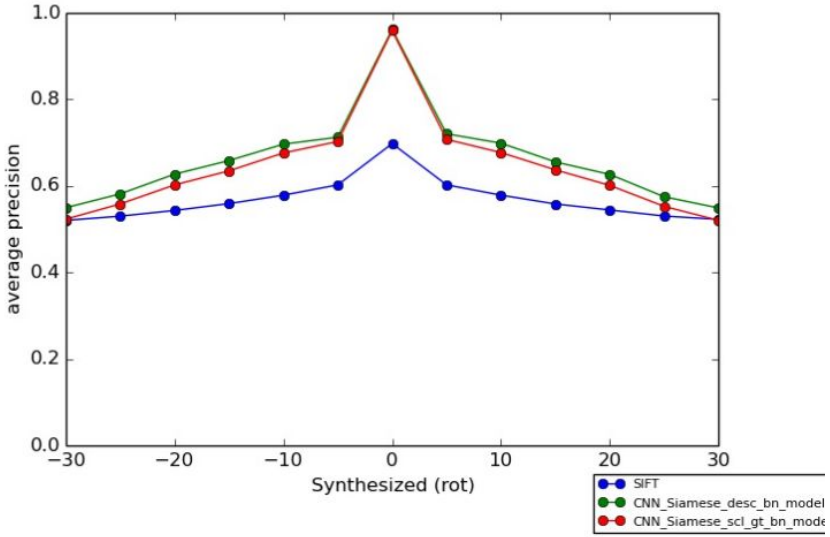
Experiment (cont'd)

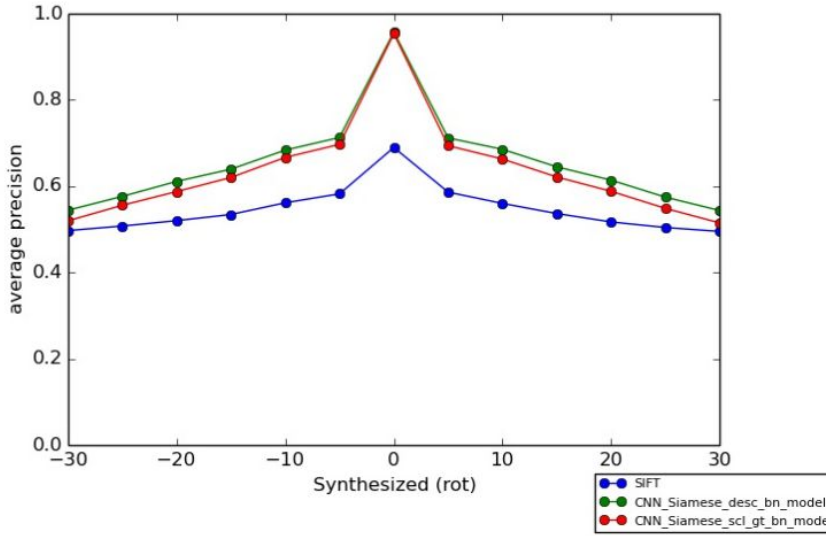
- Training Scheme
  - Second, fixed other components and update scale (with scale supervision generated by data augmentation)



Experiment (cont'd)

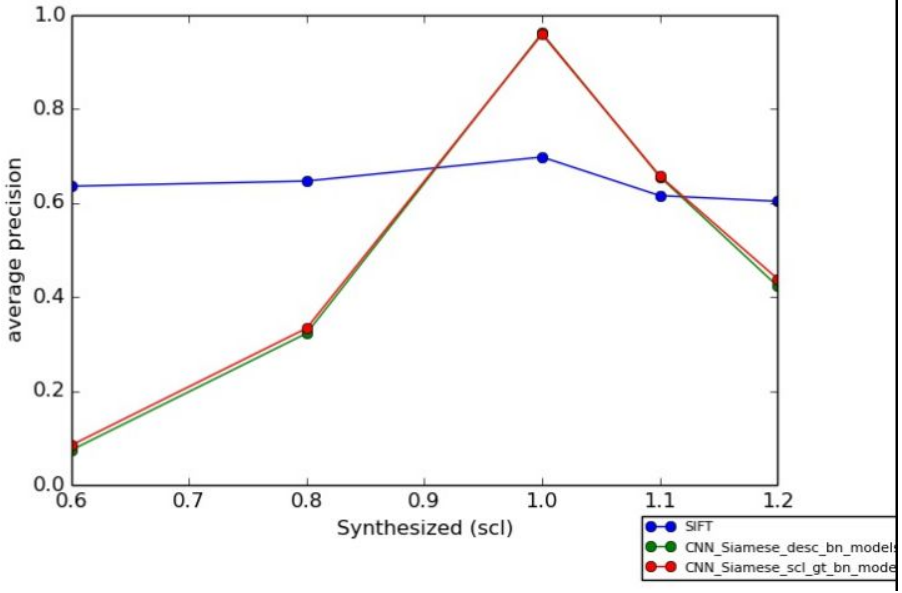
- For rotation transformation
  - Left: COCO IMG (Validation), Right: Flicker IMG (Train)

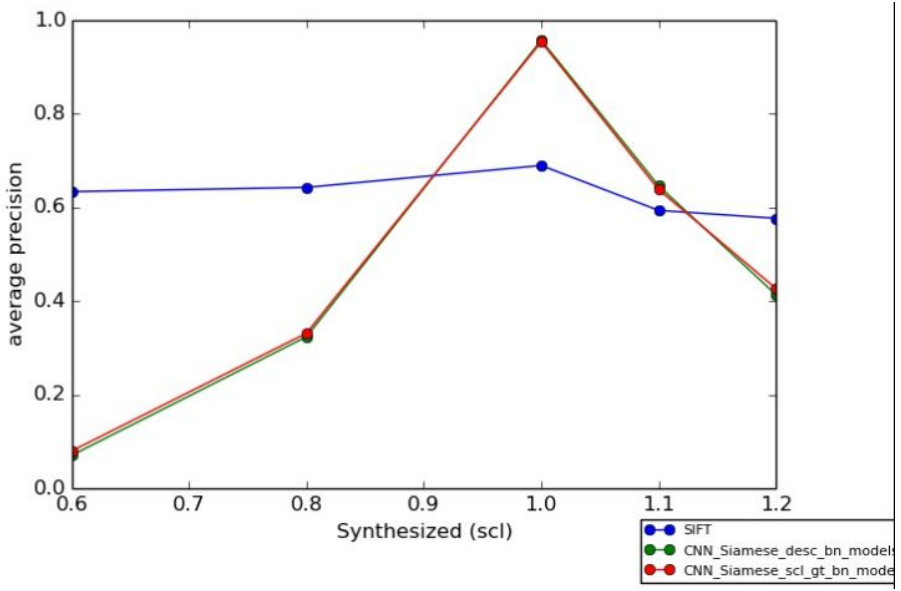




Experiment (cont'd)

- For scale transformation
  - Left: COCO IMG (Validation), Right: Flicker IMG (Train)



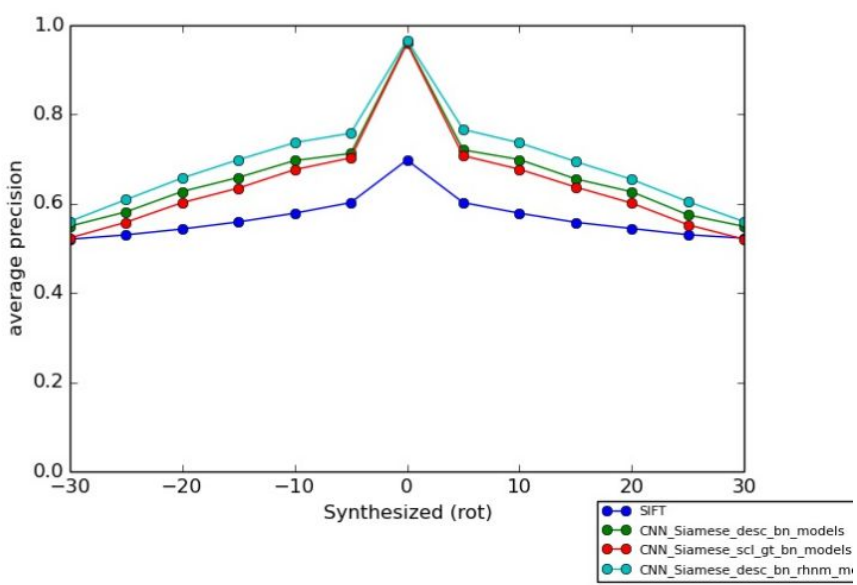


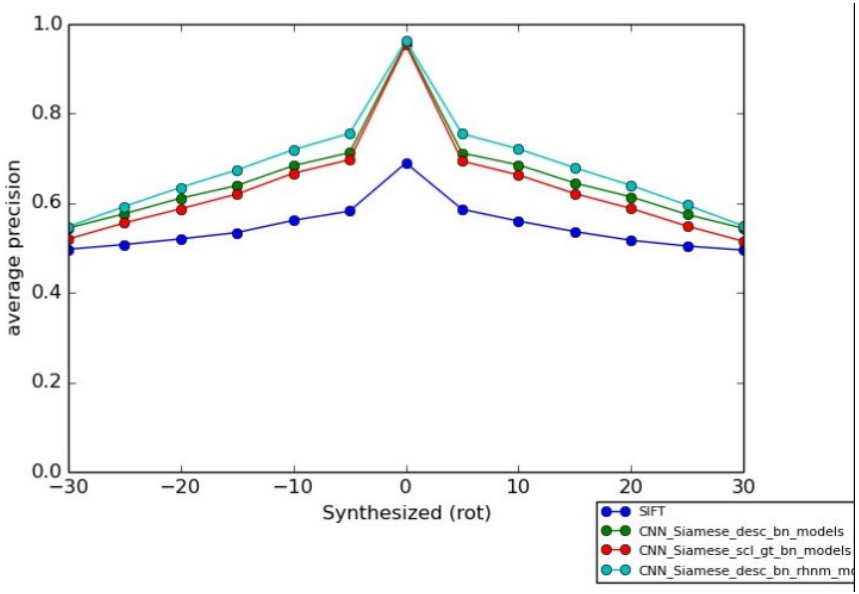
Experiment (cont’d)

- It seems that some training patches are underfit especially the scale transformation
- So we can try to backpropagate the hard samples with higher probability
  - Only update descriptor network currently
  - Epoch 0 ~ epoch 4, sampled negative pairs randomly
  - Epoch 5 ~ epoch 9, 50% of negative pairs are hard samples (with high loss), and 50% of negative pairs are randomly sampled
  - Epoch 10 ~ epoch 14, all pairs of negative examples are hard samples

Experiment (cont'd)

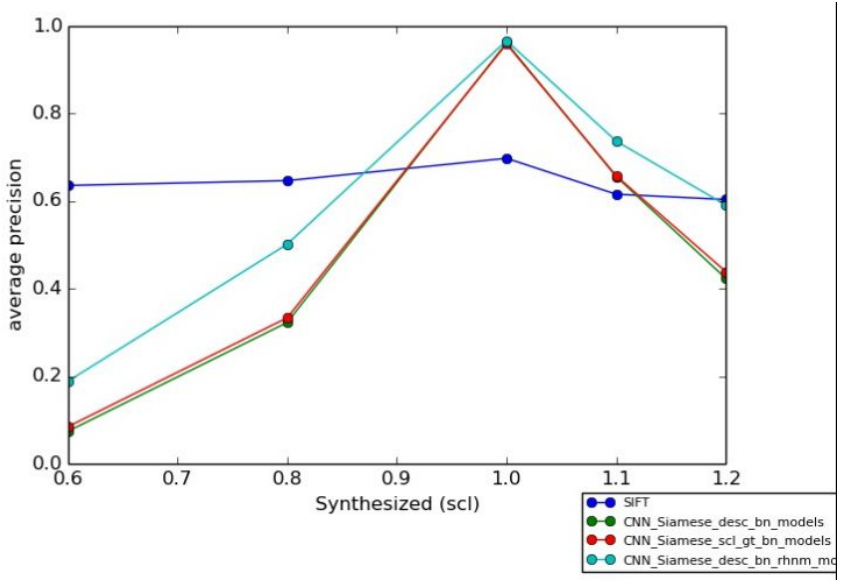
- For rotation transformation
  - Left: COCO IMG (Validation), Right: Flicker IMG (Train)

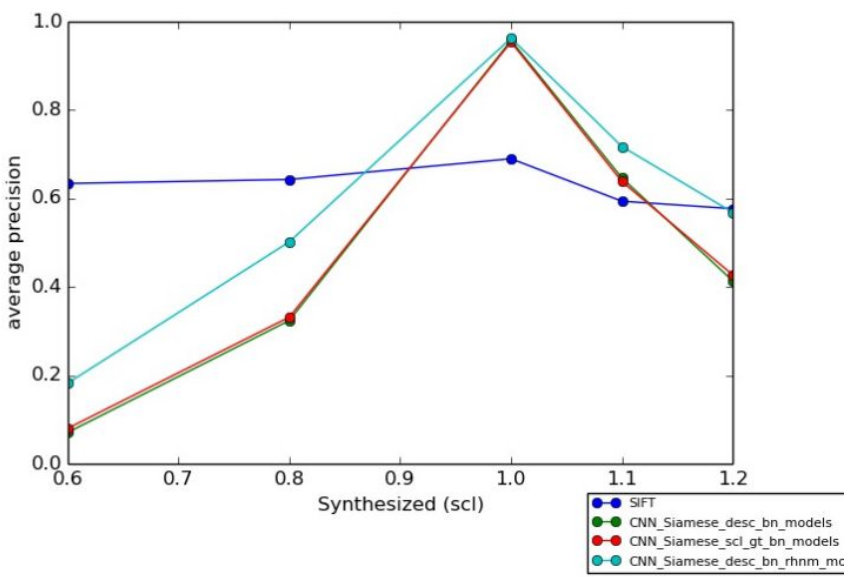




Experiment (cont'd)

- For scale transformation
  - Left: COCO IMG (Validation), Right: Flicker IMG (Train)





Summary

- Brief introduction to local descriptor matching

- Learning scale of local feature is much harder than rotation