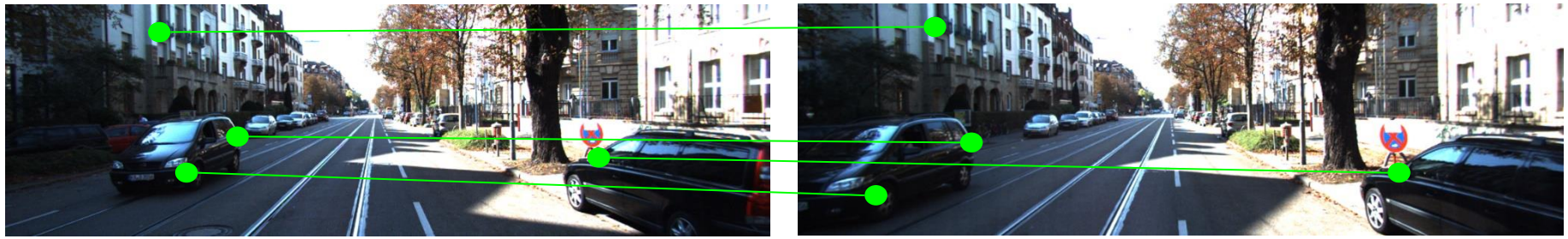
Feature Point Matching in Cross-Spectral Images with Cycle Consistency Learning

Ryosuke Furuta Naoaki Noguchi Xueting Wang Toshihiko Yamasaki
The University of Tokyo

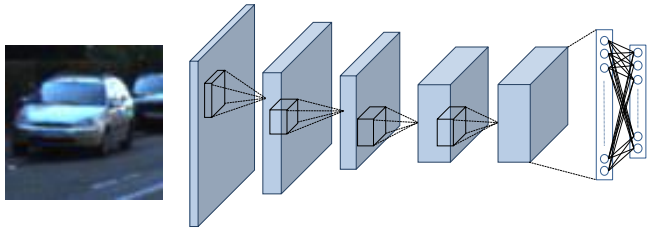
Introduction

Feature Point Matching



Applications
- Image stitching
- 3D reconstruction
- Camera calibration
- etc.

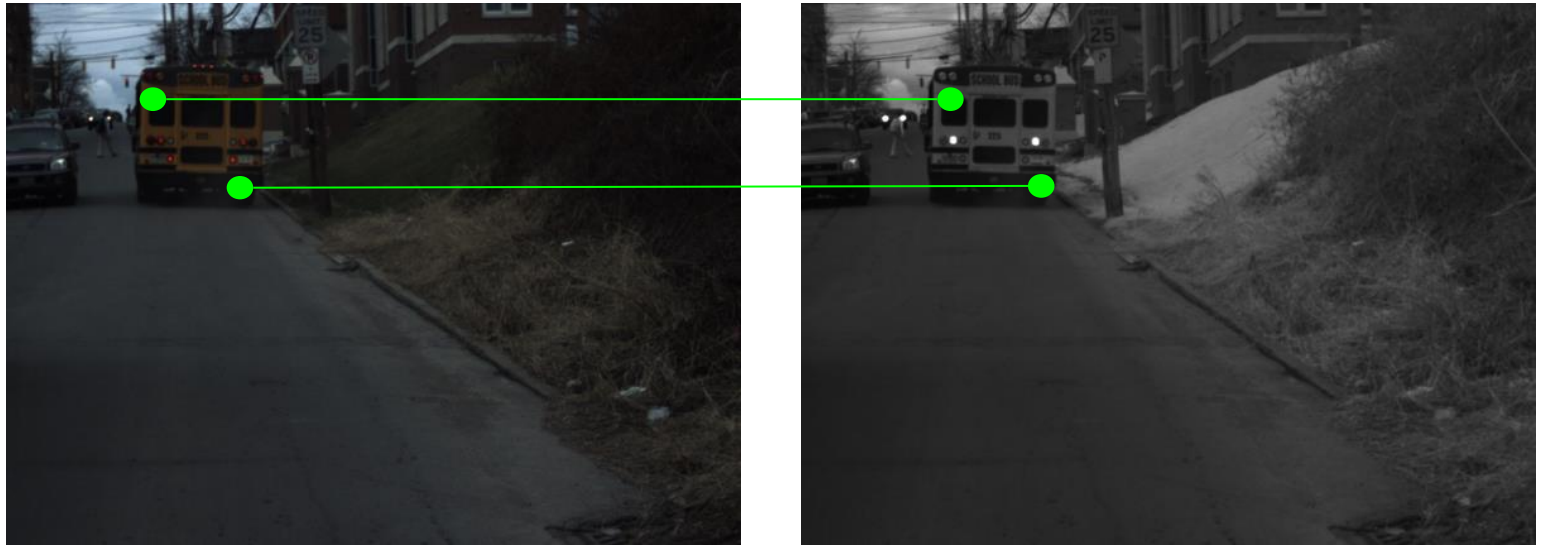
Deep learning-based approaches have been proposed as local descriptors
- HardNet [Mishchuk, NeurIPS17]
- AffNet [Mishkin, ECCV18]
- D2Net [Dusmanu, CVPR19]



feature vector: $\boldsymbol{f}$

Introduction

Feature Point Matching in **Cross-Spectral Images**



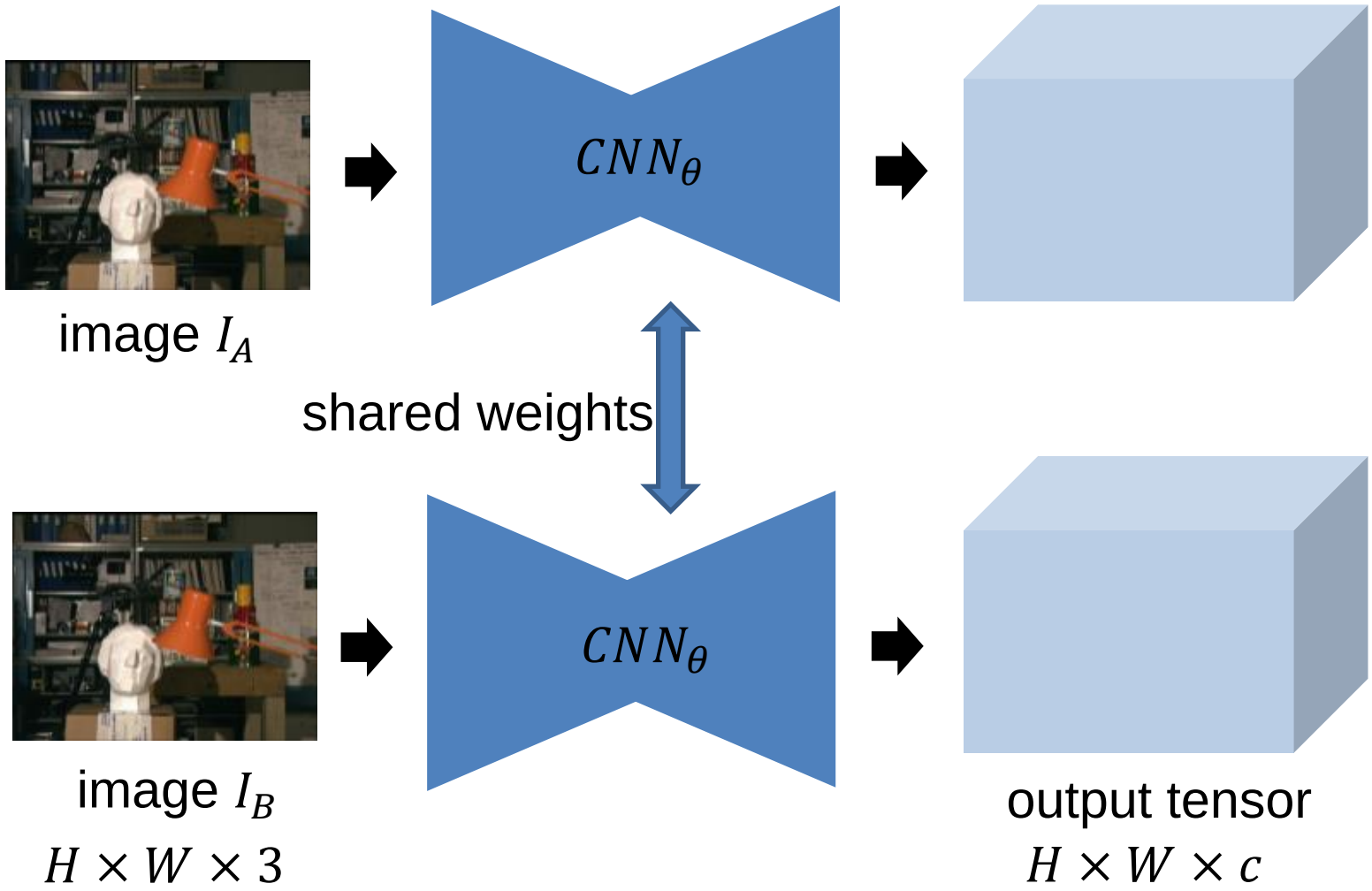
E.g., between RGB and Near-Infrared (NIR) images

Not easy to obtain ground-truth correspondences

We propose a **self-supervised learning method** to train feature extraction networks by utilizing the cycle consistency of the corresponding points.

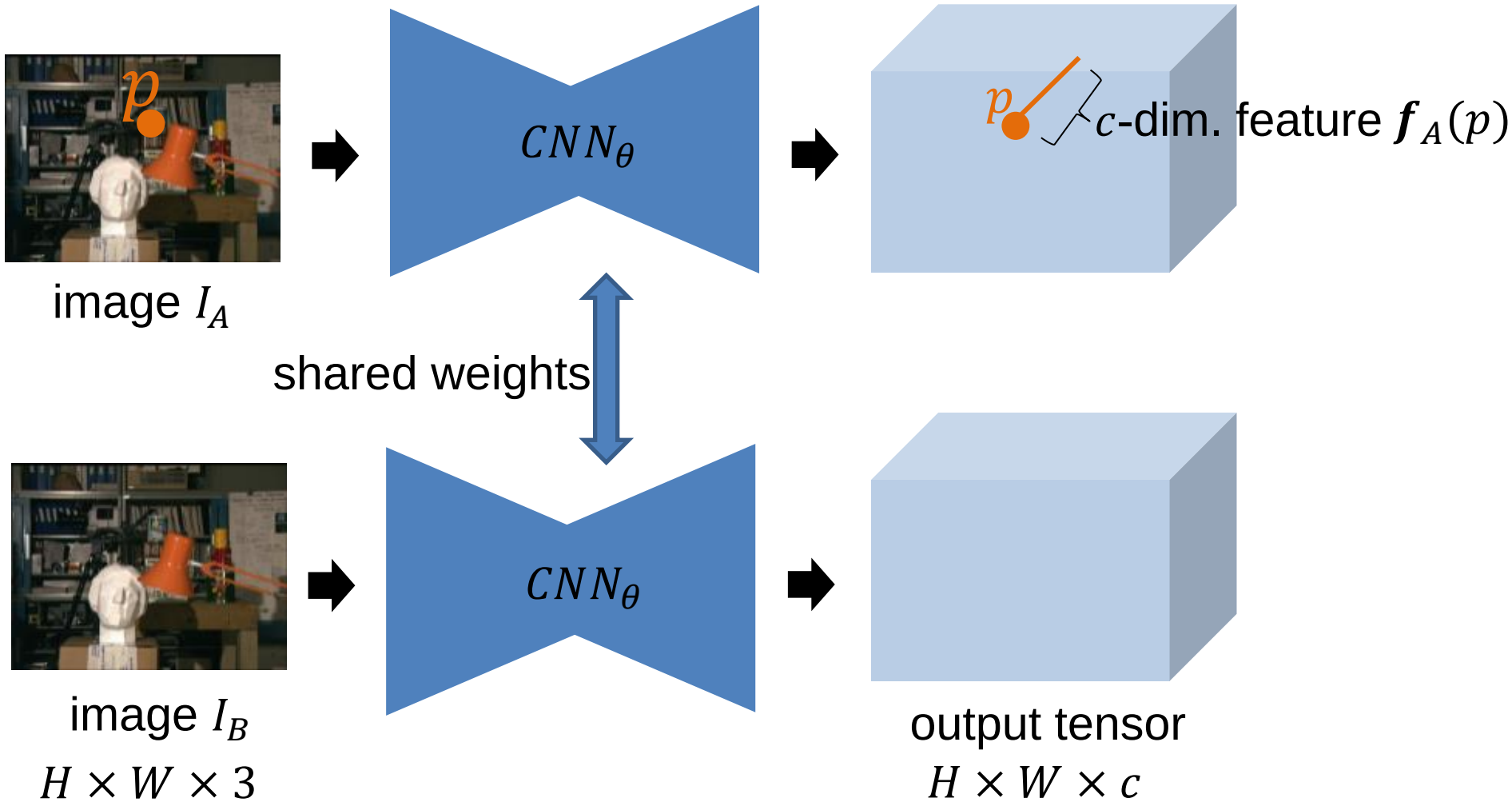
Proposed Method

Training phase



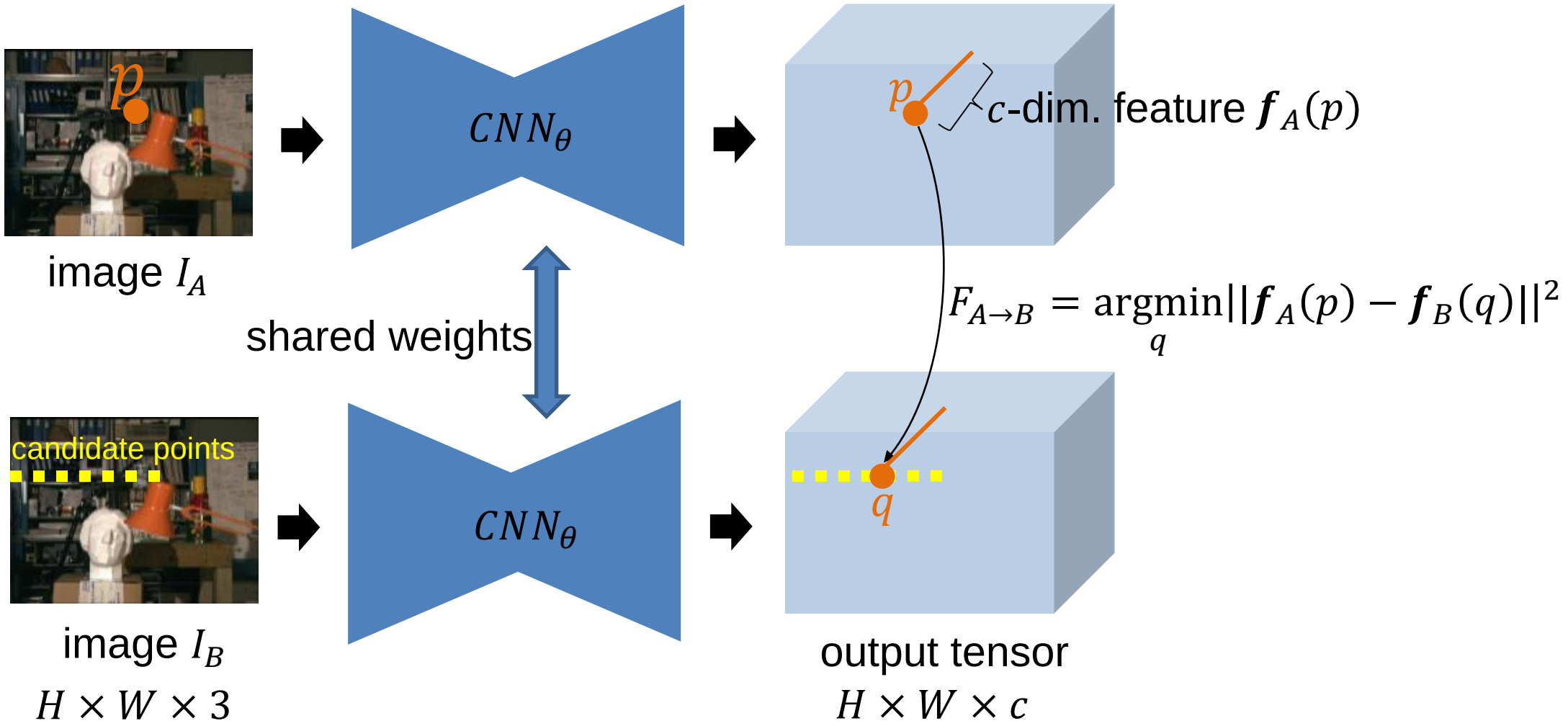
Proposed Method

Training phase



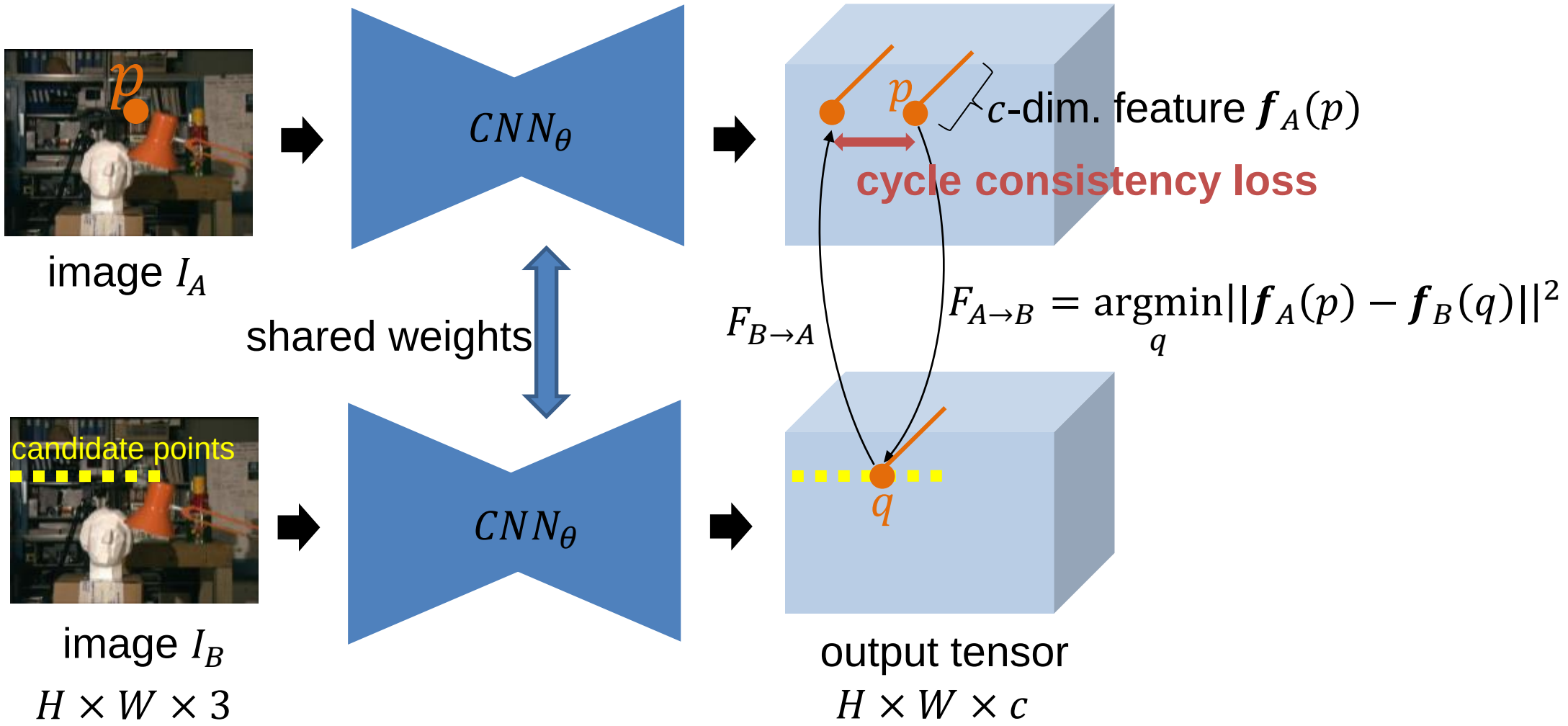
Proposed Method

Training phase



Proposed Method

Training phase



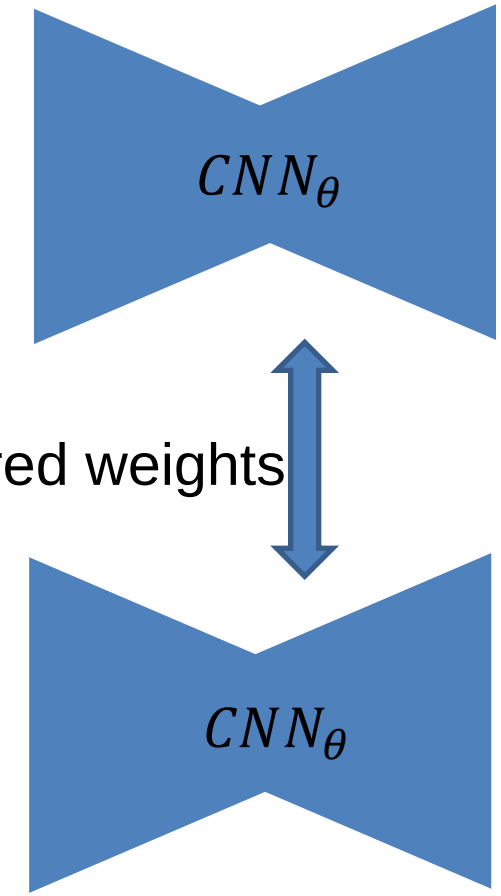
Proposed Method

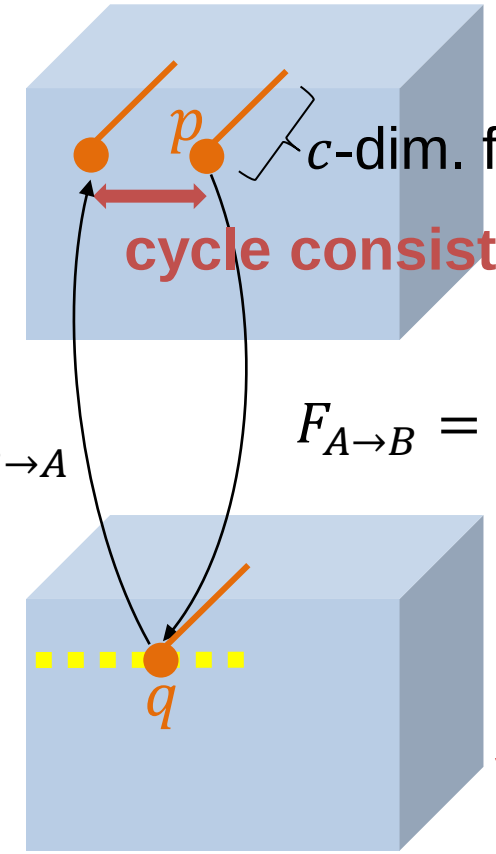
Training phase

image I_A

image I_B

$H \times W \times 3$





c-dim. feature $\boldsymbol{f}_A(p)$

cycle consistency loss

$$F_{A\to B} = \sum_q w_{pq} q$$

$$w_{pq} = \frac{\exp(-\|\boldsymbol{f}_A(p) - \boldsymbol{f}_B(q)\|^2)}{\sum_q \exp(-\|\boldsymbol{f}_A(p) - \boldsymbol{f}_B(q)\|^2)}$$

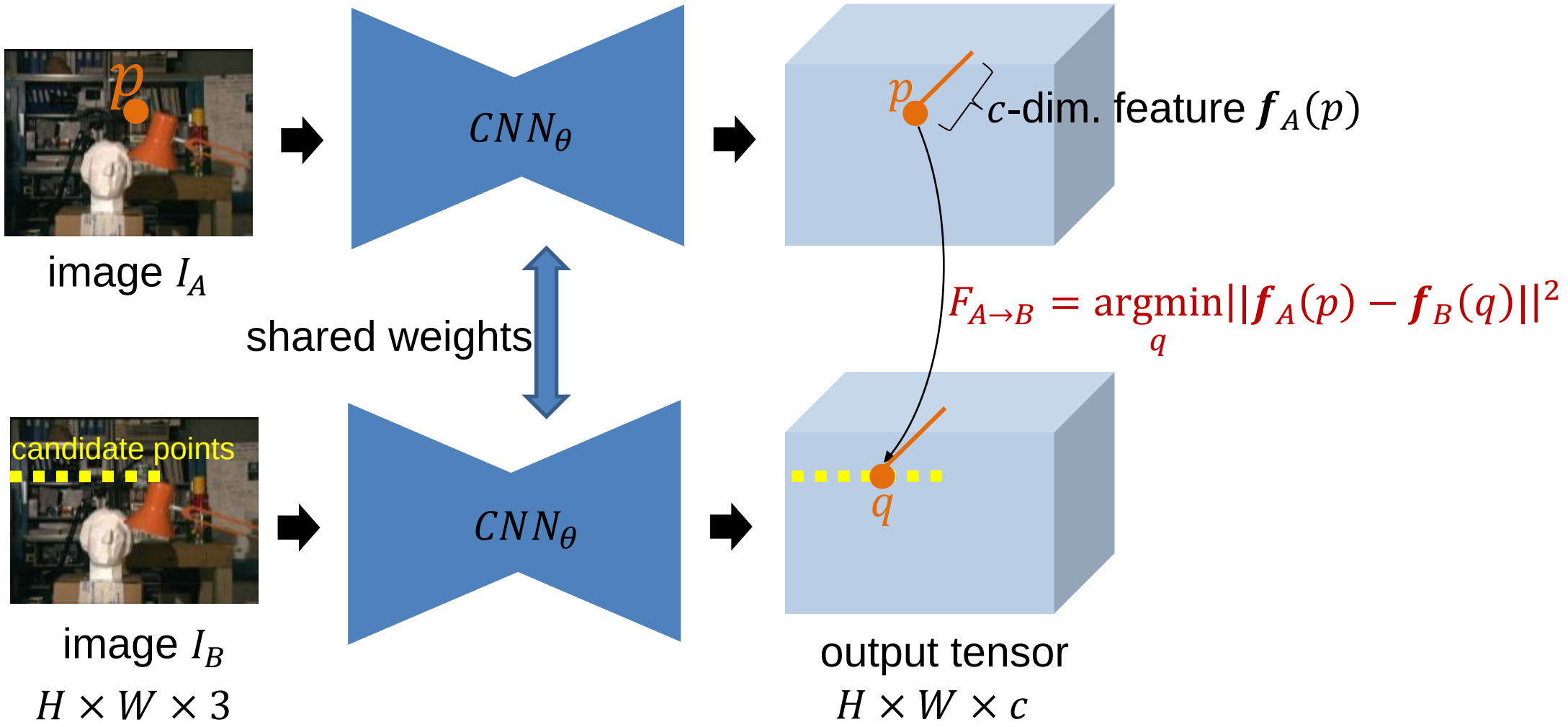
Soft nearest neighbor [Dwibedi, CVPR19]

output tensor

$H \times W \times c$

Proposed Method

Test phase



Experiments

Stereo matching on KITTI 2012 dataset

- 390 image pairs for training
- 194 image pairs for testing

RGB stereo

Simulated three types of cross-spectral settings

- RGB stereo
- RGB2gray
- anaglyph

RGB2gray

anaglyph

Compared methods

- Hand-crafted cost function + nearest-neighbor matching:
  - Baseline
- Hand-crafted cost function + smoothness regularization (guided filter) + post-processing:
  - Cost-volume filtering (CVF) [Hosni, TPAMI12]

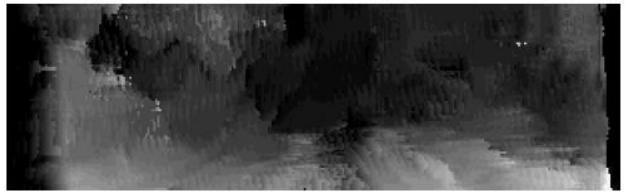
Experiments

	Error rate [%] ↓			Mean error [pix] ↓		
	RGB stereo	RGB2gray	anaglyph	RGB stereo	RGB2gray	anaglyph
Ours	**39.0**	**35.4**	**27.6**	**5.29**	**4.93**	**4.20**
Baseline	52.8	49.9	57.9	7.31	6.92	8.20
CVF [Hosni, TPAMI12]	43.9	43.7	34.0	5.65	5.59	4.71

Left image

Right image

Ground truth



Ours
(Learned feature + NN matching)

Baseline
(Handcrafted feature + NN matching)

CVF
(Handcrafted feature
+ Filtering + Post-processing)

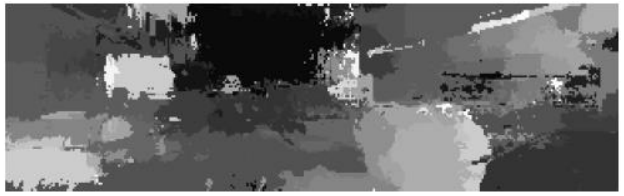
Left image

Right image

Ground truth

Ours
(Learned feature + NN matching)

Baseline
(Handcrafted feature + NN matching)



CVF
(Handcrafted feature
+ Filtering + Post-processing)

Conclusions

- General feature point matching including cross-spectral settings
- Proposed method:
  - Self-supervised method with cycle consistency learning
- Experimental results on cross-spectral stereo matching:
  - Better accuracy than hand-crafted methods on KITTI dataset
  - Not as accurate as the compared methods but much faster on PittsStereo dataset
- Future works:
  - Deal with occlusions for better accuracy
  - Apply to other feature point matching problems such as image stitching and optical flow estimation