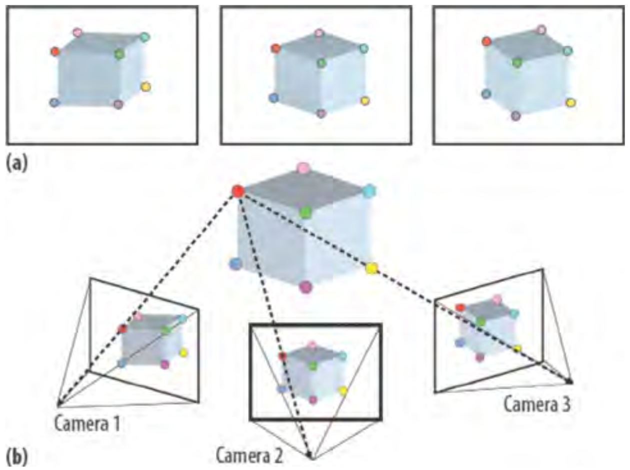
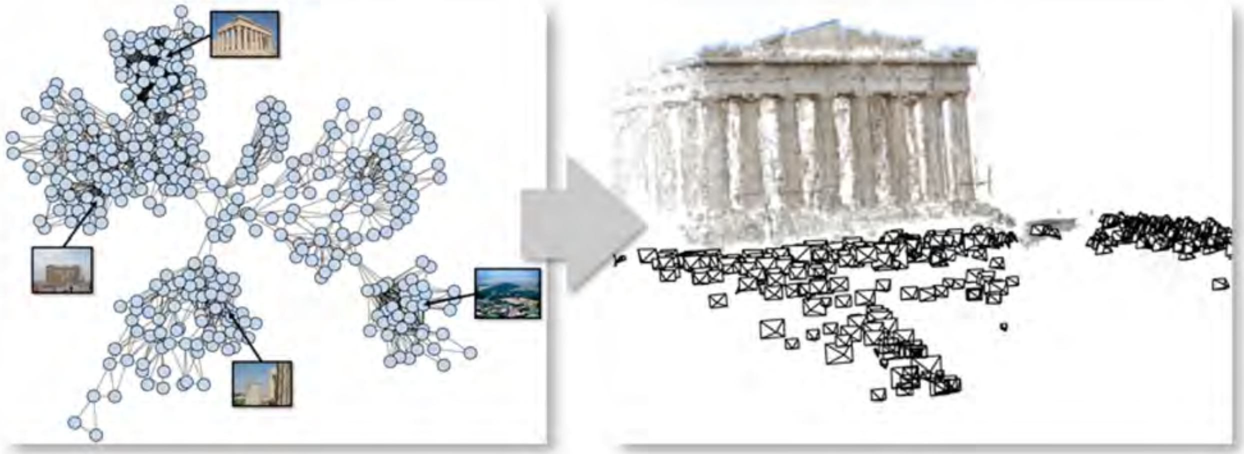
GIFT: Learning Transformation-Invariant Dense Visual Descriptors via Group CNNs

Yuan Liu, Zehong Shen, Zhixuan Lin, Sida Peng, Hujun Bao, Xiaowei Zhou, NeurIPS 2019
@ZJU-3DV

Background

- Keypoint Matching
  - Fundamental to many downstream tasks like *SfM*, *Visual Localization*, *SLAM*.



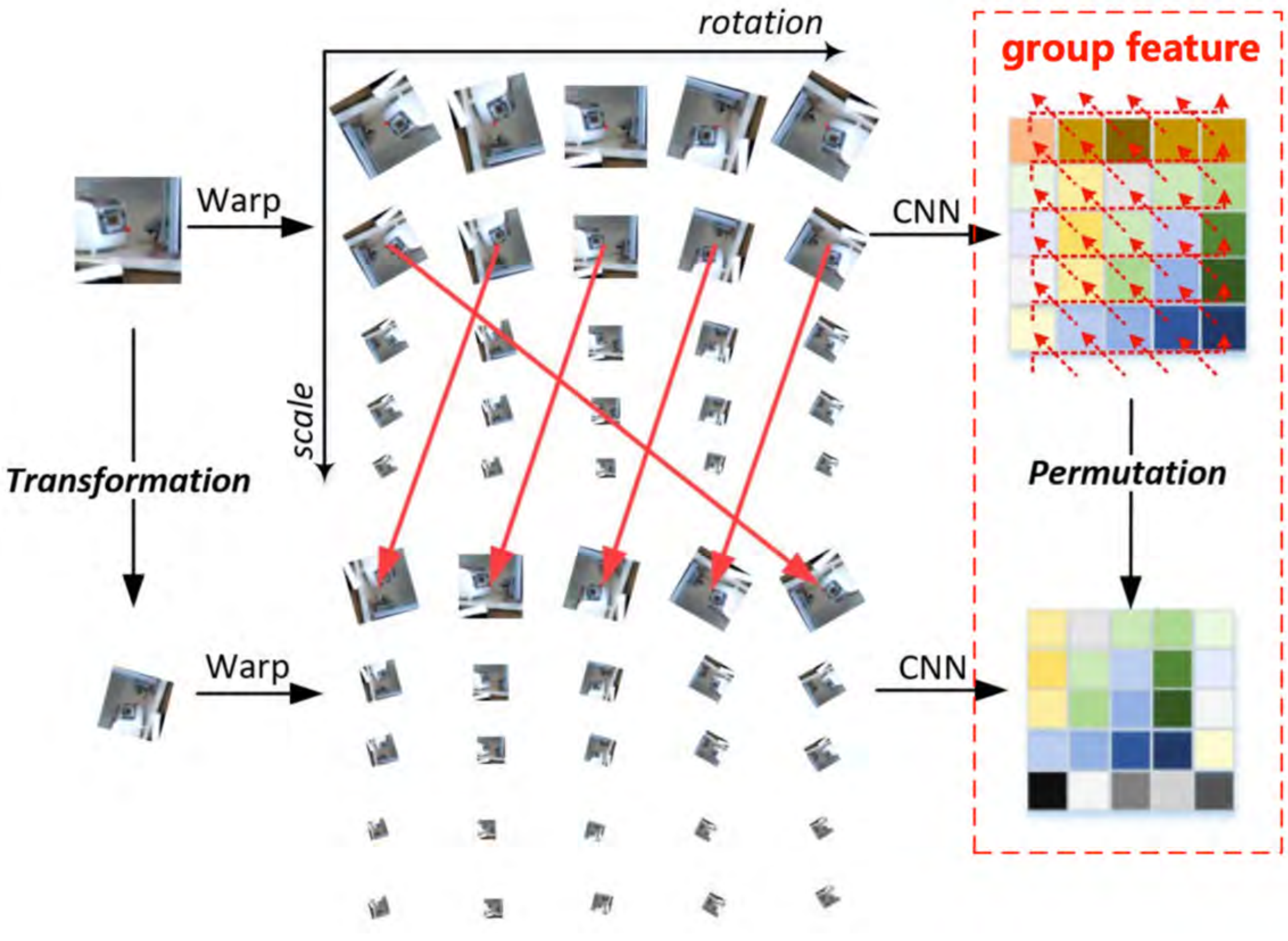


Background

- Failure cases under large transformation
  - Traditional methods heavily rely on detector
  - Convolution operator are only equivariant to translation of the input

- Can we learn a **invariant descriptor with theoretical guarantees**?
  - Rotation and Scaling

Extract Equivariance (Group) Feature



- Equivariance Definition

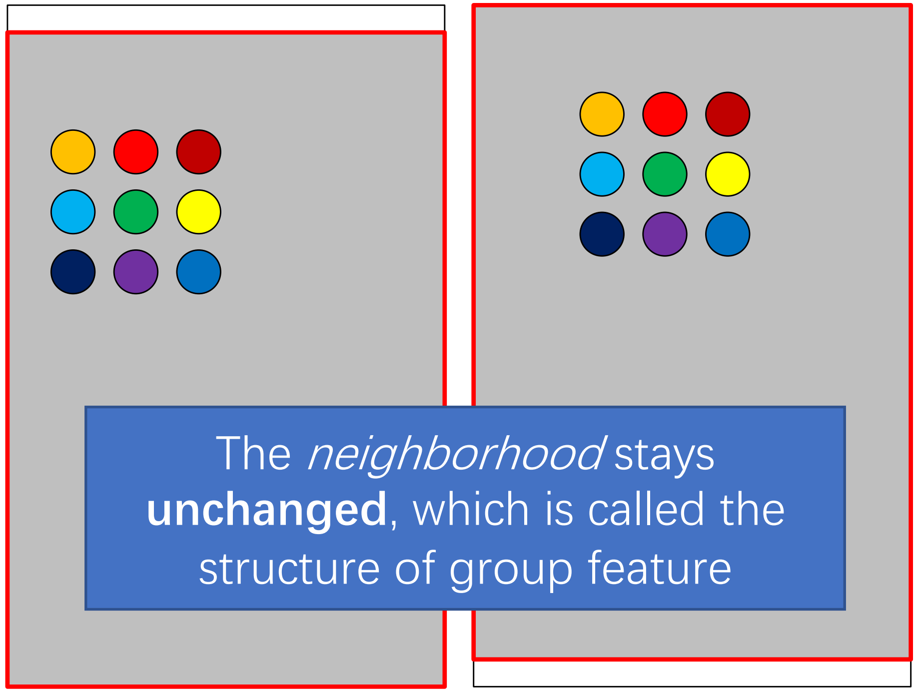
$$\begin{tikzcd} I \arrow[r, "f"] \arrow[d, "T_g"'] & f(I) \arrow[d, "T'_g"] \\ T_g \circ I \arrow[r, "f"] & f(T_g \circ I) \end{tikzcd}$$

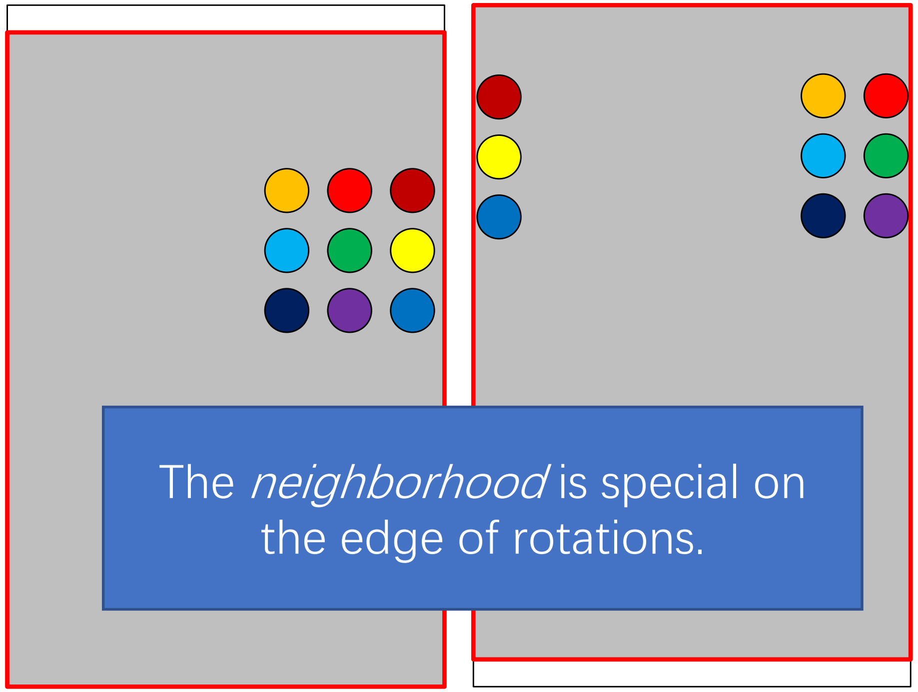
$$f(T_g \circ I) = T'_g \circ f(I)$$

- Here
  - Geometric transformation results in permutation

Structure of Group Feature

- Invariance of Local Structure
  - If we sample lots of scales and rotations, we will get two feature maps which are almost the same except for the edge of scales.





Group-Conv to Extract Information

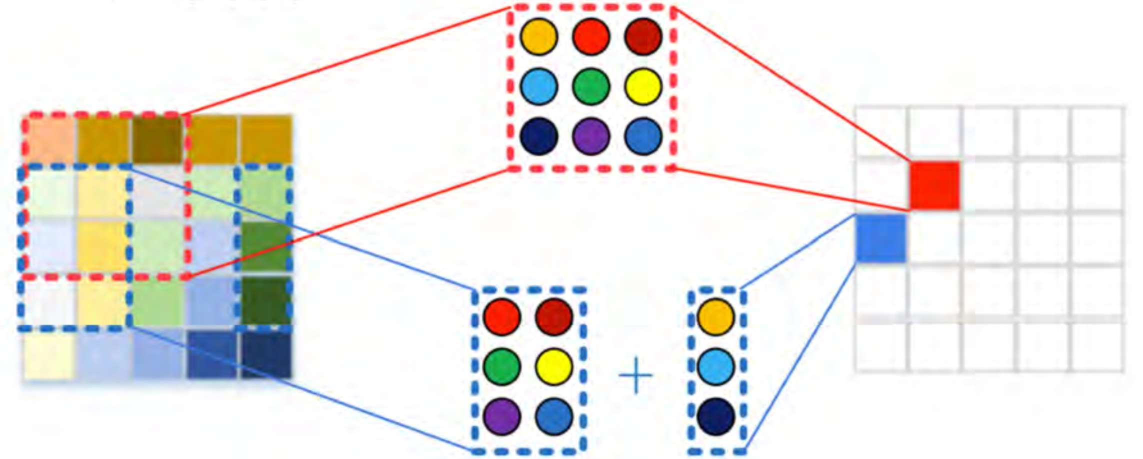
- Vanilla Convolution

$$f^{(l+1)}(x) = \sum_{h \in H} f^{(l)}(x+h)W(h)$$

- Group Convolution

$$f^{(l+1)}(g) = \sum_{h \in H} f^{(l)}(gh)W(h)$$

$$H = \{r, r^{-1}, s, s^{-1}, rs, rs^{-1}, r^{-1}s, r^{-1}s^{-1}, e\}$$

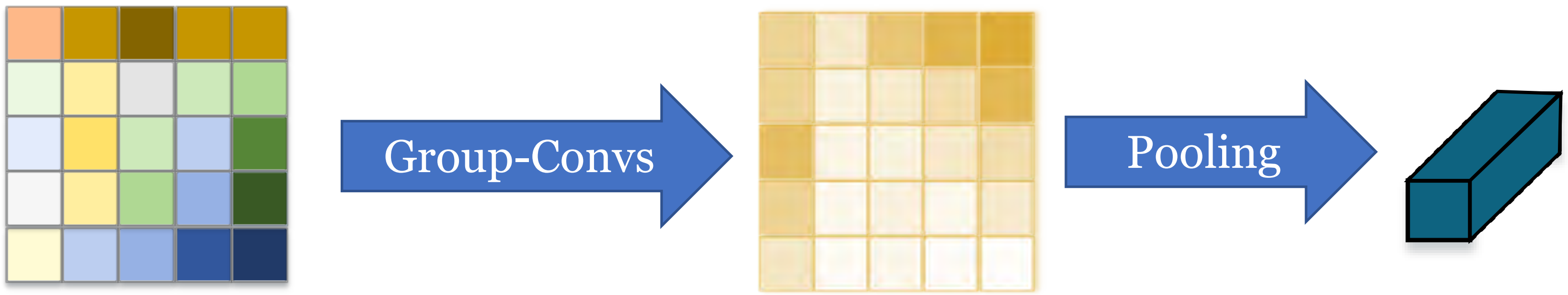


- *In our case, the scaling and rotation group, with regards to one point, looks like a cylinder*

- Vanilla convolution is also a group convolution which is defined on the translation group

After Multiple Group-Conv Layers

- We still get **a feature map**, but our goal is to get **a single invariant feature vector**

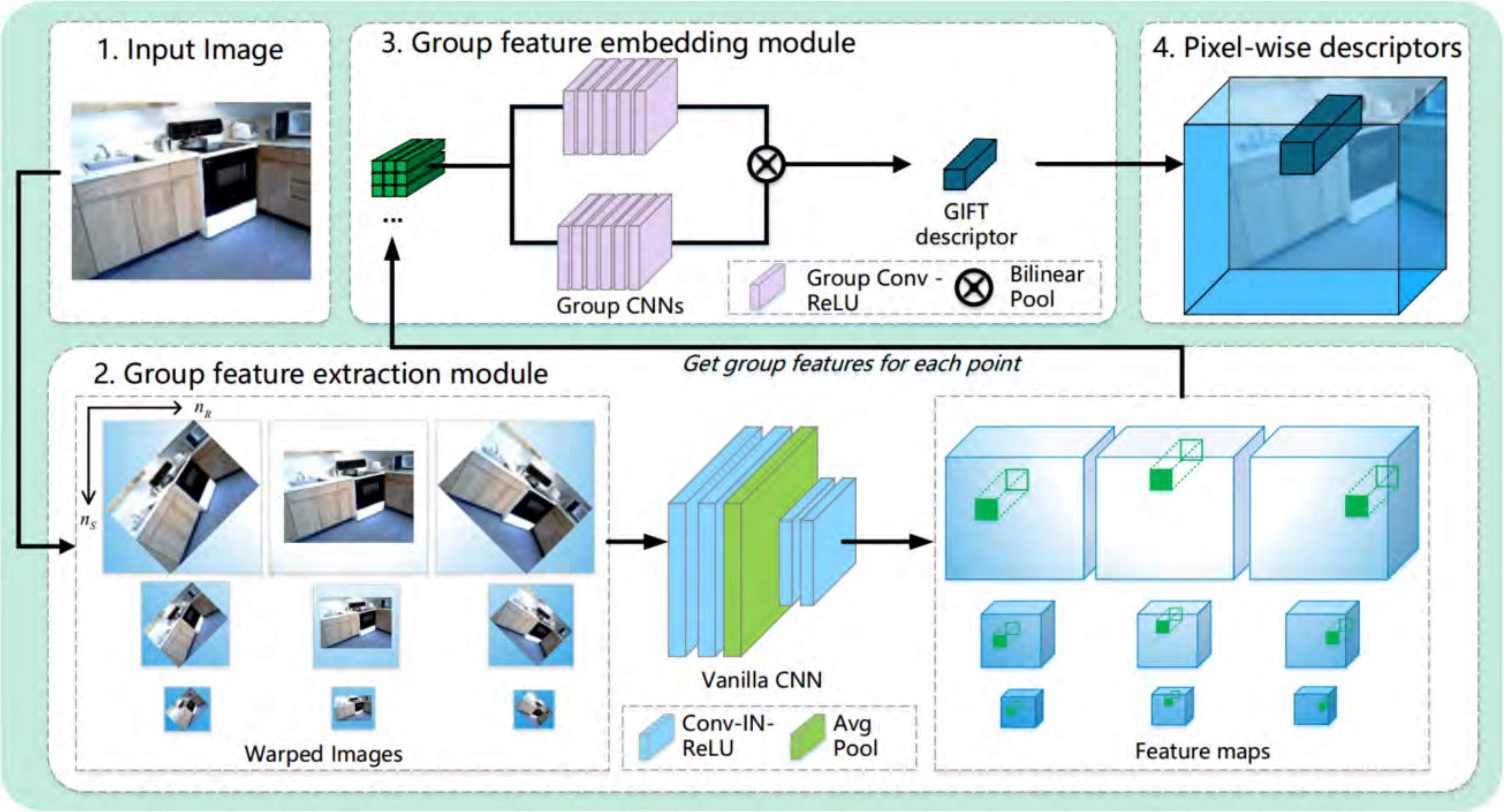


- Which pooling? Max/Average?
- Bilinear pooling
    - Second order statistics are more informative
    - Generalized form of previous descriptors

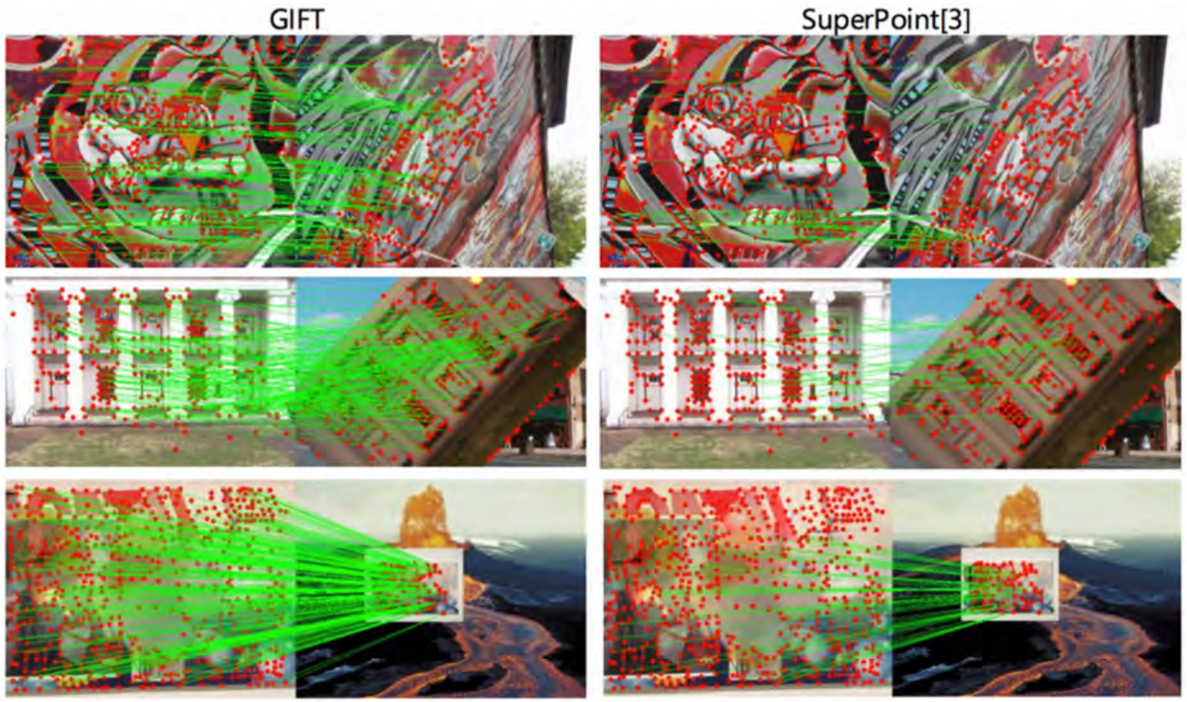
Refer to paper for more details!

Lin, Tsung-Yu, et. al. "Bilinear cnn models for fine-grained visual recognition." ICCV 2015.

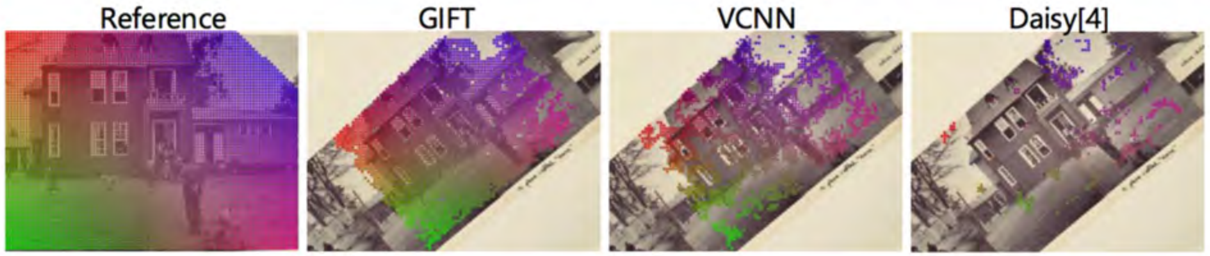
Proposed Method



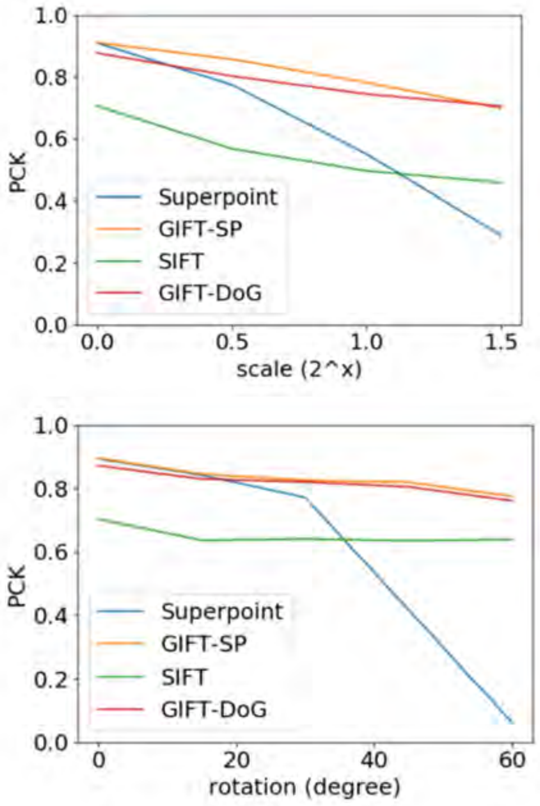
Results



a) Sparse Correspondence



b) Dense Correspondence



c) Systematical Analysis

code is available!

Thank You !

Zehong Shen @ ZJU-3DV, zhshen0917@gmail.com
12/17/2019