Learning Discriminative and Transformation Covariant Local Feature Detectors

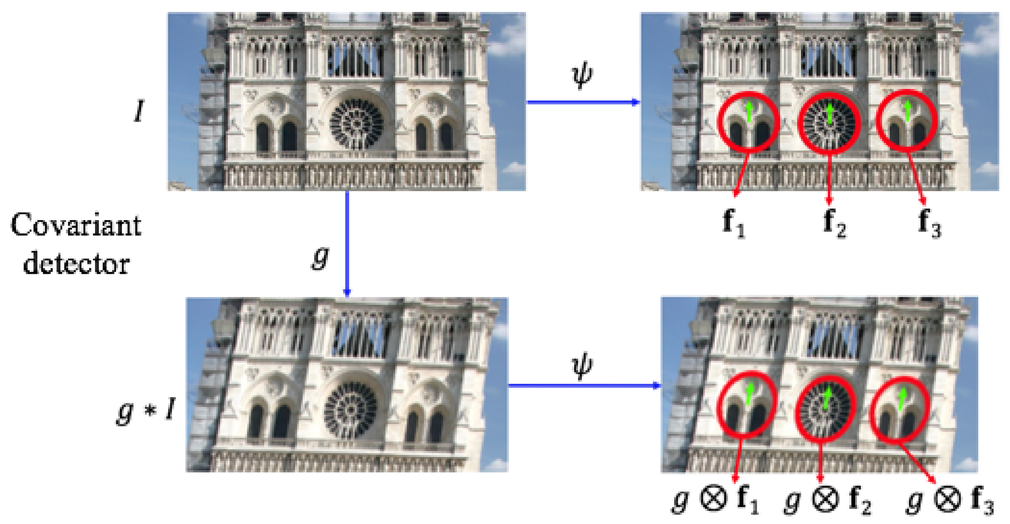
Xu Zhang[1], Felix X. Yu[2], Svebor Karaman[1], Shih-Fu Chang[1] [1]Columbia University, [2] Google Research

Covariant Detector

Discriminative: It can discover local discriminative information in the image.

Covariant: It can repeatedly detect consistent patterns when the scene undergoes diverse transformations.

Operator: Apply transform to image ($*$) and feature ($\otimes$). $\circ$: composition of transform.

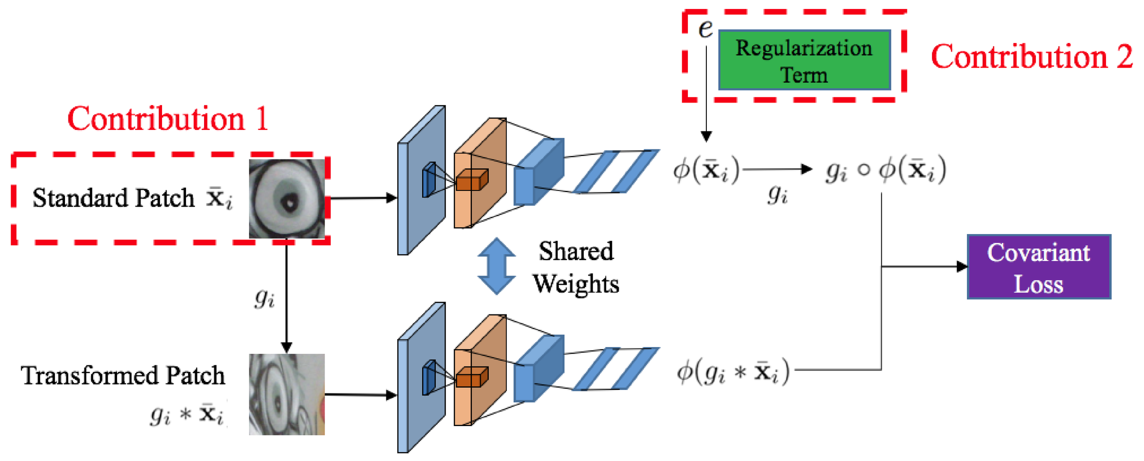


Contributions

Extend the covariant constraint proposed by Lenc and Vedaldi [1] by defining the concepts of "standard patch" and "canonical feature".

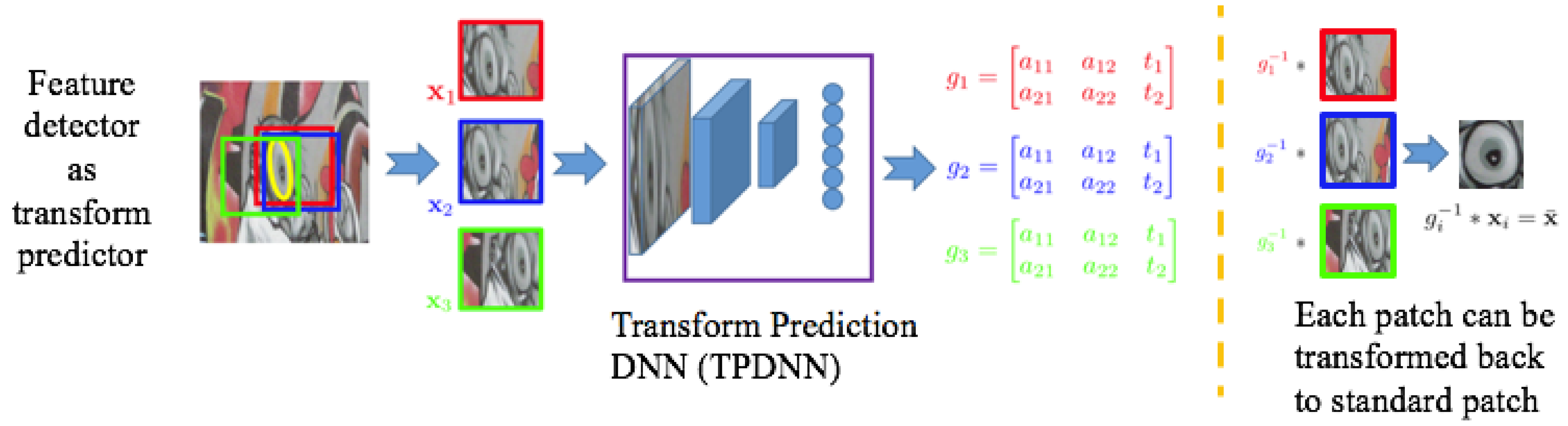
Show that the introduction of these concepts greatly simplifies the learning process, and also makes the detector more robust.

Covariant constraint: $\psi(g * \mathbf{x}) = g \otimes \psi(\mathbf{x})$.



Training: To preserve the input relations

Method



Transform Predictor (TPDNN): Predicts transform associated with a local patch.

Covariant constraint in transform space, $\phi(g * \mathbf{x}) = g \circ \phi(\mathbf{x})$, $\phi(\cdot)$ is the transform predictor.

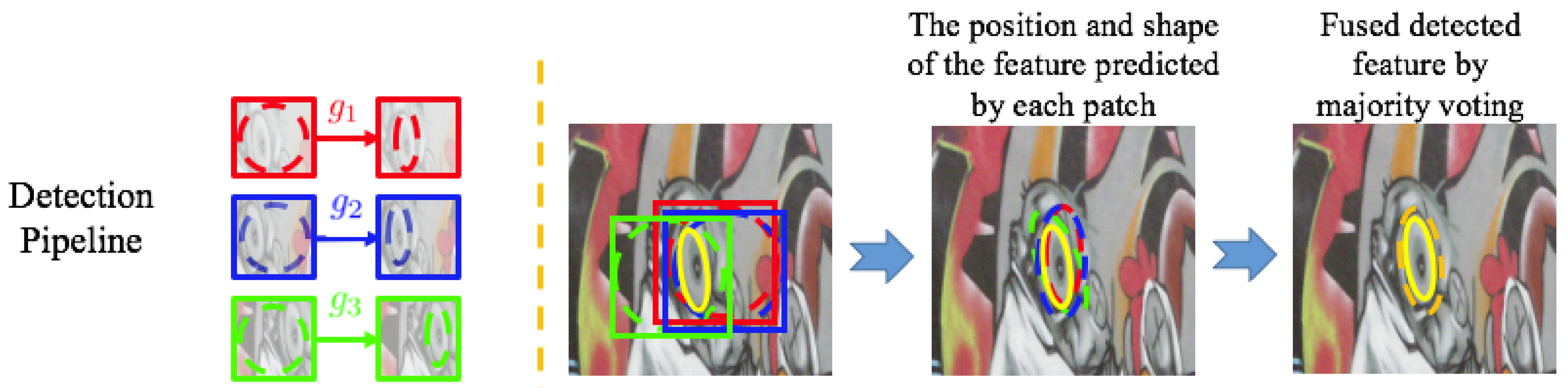
Covariant Loss: $\sum_{i=1}^{n} \| \phi(g_i * \mathbf{x}_i) - g_i \circ \phi(\mathbf{x}_i) \|_F^2$

Identity Loss: $\sum_{j=1}^{m} \| \phi(\bar{\mathbf{x}}_j) - e \|_F^2$

Training: Using synthesized patch from standard patch with the gt transform to train the siamese networks to preserve the input relation.

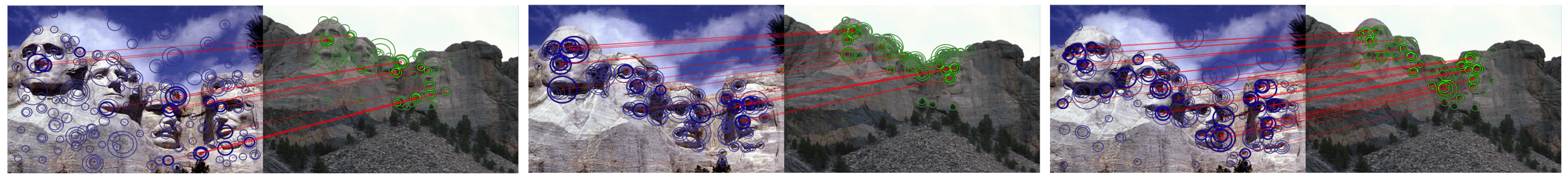
The Estimated Transform: The correct predicted transform is the one whose inverse can transform a patch ($\mathbf{x}_i$) back to a standard patch ($\bar{\mathbf{x}}_i$).

Detection: Multiple patches covering the same local feature are fused to refine the estimated location of the local feature.



$\mathbf{f}_0 \quad g_i \otimes \mathbf{f}_0 = \mathbf{f}_i \qquad g_i \otimes \mathbf{f}_0 = \mathbf{f}_i$

Results



T-P24 [2]: #Feat. 92, #Rep. 19 CovDet [1]: #Feat. 67, #Rep. 16 Ours: #Feat. 69, #Rep. 33

#Rep.: Number of correspondences (features that can be detected in both images).

Repeatability of different detectors

Method	Webcam #Feature 1000	Webcam #Feature 200	EF #Feature 1000	EF #Feature 200	VGG #Feature 1000	VGG #Feature 200
SIFT	29.5	19.1	20.8	10.9	47.1	41.7
SURF	46.0	33.4	39.7	23.4	61.2	58.3
MSER	45.1	29.4	37.1	18.9	54.1	38.4
HesLap	51.1	37.2	38.8	28.0	66.7	60.0
HesAff	42.5	34.5	26.6	21.8	66.4	59.6
T-P	35.4	29.0	26.3	16.3	54.6	46.1
T-P24	61.7	45.1	45.4	32.3	64.4	57.6
T-CNN	51.4	36.7	38.0	21.8	50.7	40.6
CovDet	49.9	32.2	42.7	23.8	62.0	48.0
Ours	**68.4**	**52.6**	**46.6**	**36.3**	**70.2**	**61.2**

Repeatability:: $\frac{\#\text{correspondences}}{\text{smaller number of features}}$

Matching Score: $\frac{\#\text{correct matches}}{\text{smaller number of features}}$. A correct match is a pair of correspondence that are also nearest neighbor in descriptor space.

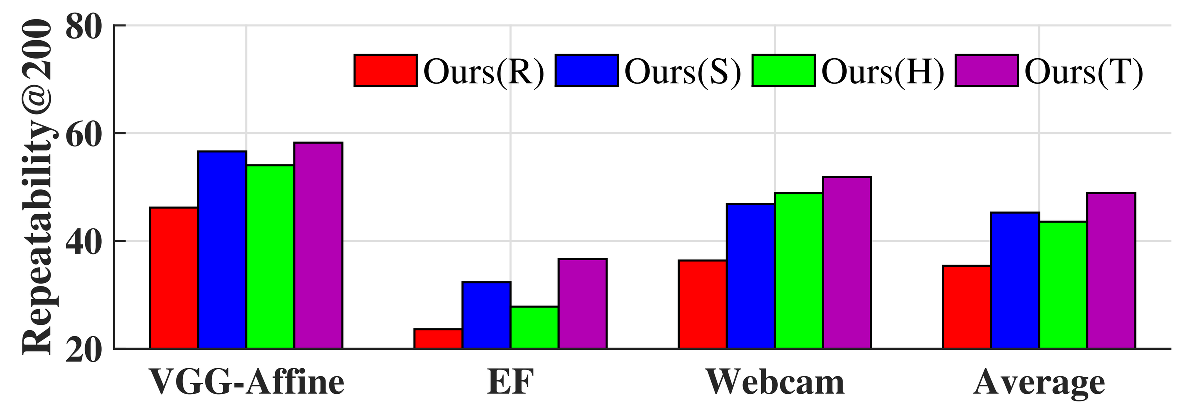
The result confirms the importance of

1. Defining discriminative patches;
2. The covariant constraint.

Matching score of different detectors

Dataset	Detector SIFT	Detector T-P24	Detector CovDet	Detector Ours
Webcam	12.9	13.4	12.0	**19.4**
VGG	42.8	44.5	43.1	**50.7**
EF	**10.2**	5.2	4.8	6.2

Standard Patches



Using different standard patches for training:

1. Randomly sampled patches (Ours(R))
2. SIFT detector (Ours(S))
3. Hessian Affine detector (Ours(H))
4. T-P24 [2] detector (Ours(T), best result)

Reference

[1] K. Lenc, and A. Vedaldi. Learning Covariant Feature Detectors In *ECCVW 2016*

[2] Y. Verdie, K. M. Yi, P. Fua, and V. Lepetit. TILDE: A Temporally Invariant Learned DEtector. In *CVPR 2015*

Source Code

```
https://github.com/
ColumbiaDVMM/Transform_
Covariant_Detector
```

Acknowledgement

This material is based upon work supported by the United States Air Force Research Laboratory (AFRL) and the Defense Advanced Research Projects Agency (DARPA) under Contract No. FA8750-16-C-0166. Any opinions, findings and conclusions or recommendations expressed in this material are solely the responsibility of the authors and does not necessarily represent the official views of AFRL, DARPA, or the U.S. Government.