ICASSP2021
TORONTO
Canada
June 6-11, 2021
Metro Toronto Convention Centre

2021 IEEE International Conference on Acoustics, Speech and Signal Processing
6-11 June 2021 • Toronto, Ontario, Canada
Extracting Knowledge from Information

CANET: CONTEXT-AWARE LOSS FOR DESCRIPTOR LEARNING

Tianyou Chen[1], Xiaoguang Hu[1], Jin Xiao[1], Guofeng Zhang[1], and Hui Ruan[1]

[1]Beihang University Beijing 100191, China

Introduction

What is Local Feature Descriptor?

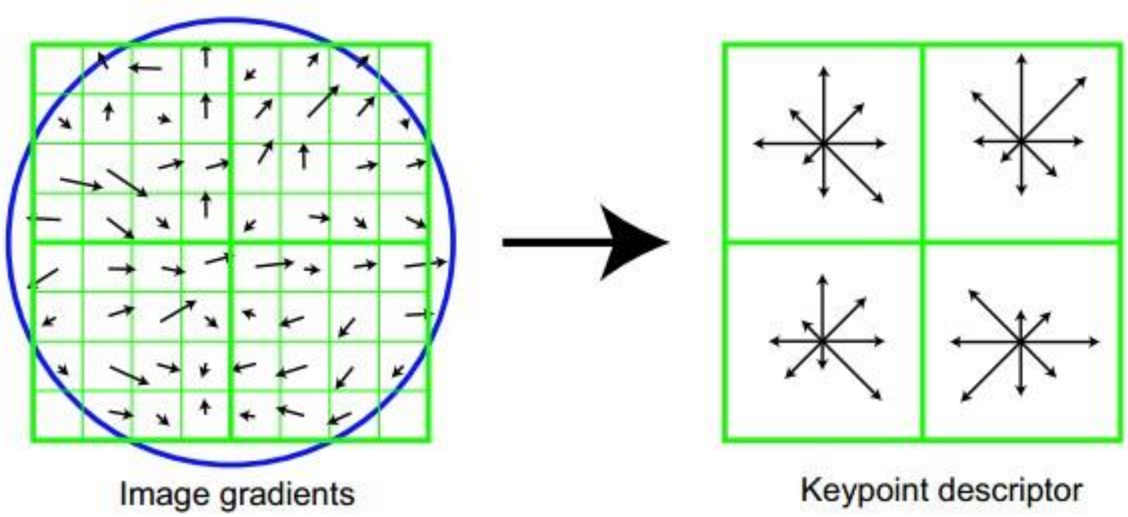
Encoding local images into representative vectors to compare local patches across images.

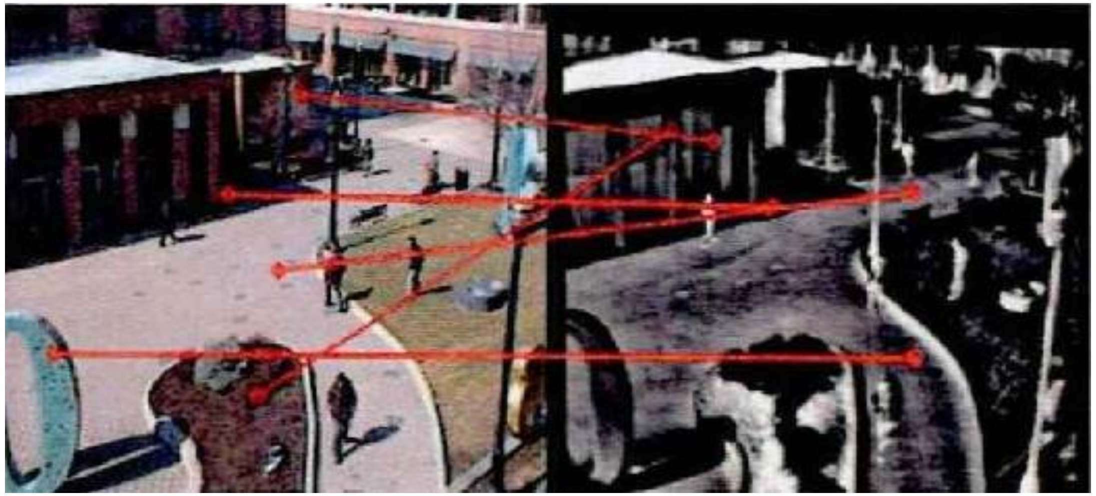
Application domains

- 3D reconstruction
- Wide-baseline matching
- Image retrieval

Existing works

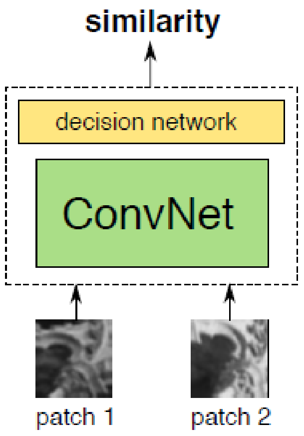
Traditional approaches



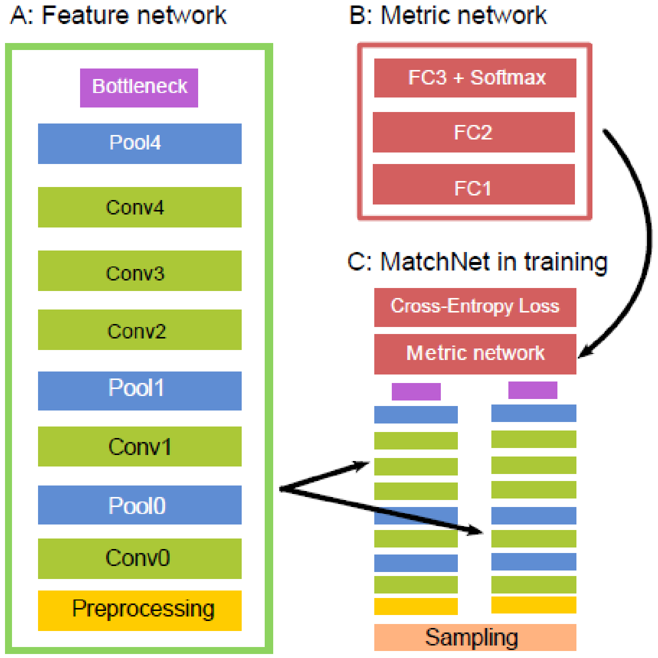


Existing works

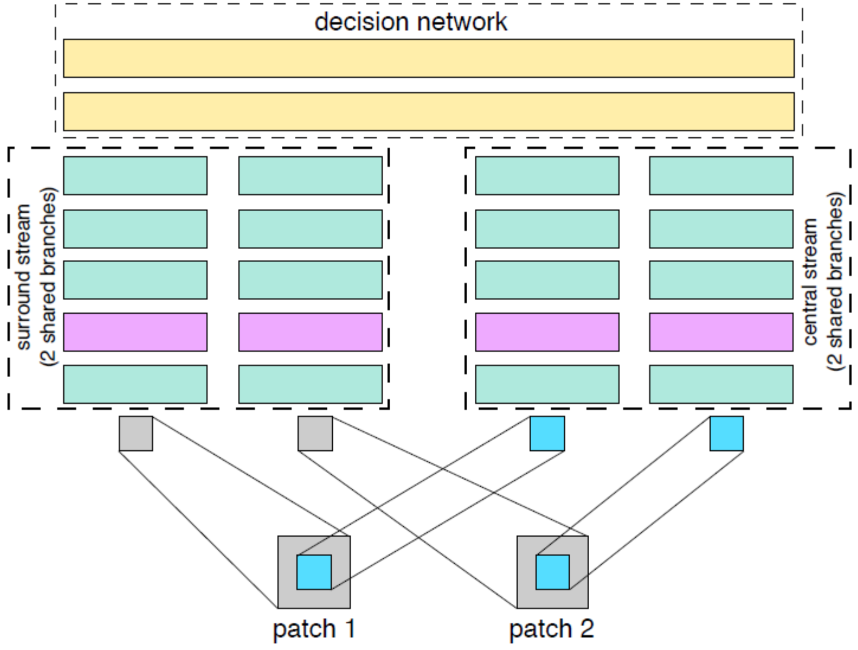
Deep learning based approaches (Siamese network)



MatchNet



DeepCompare



MatchNet: Han et al., MatchNet: Unifying feature and metric learning for patch-based matching. CVPR 2015.
DeepCompare: Zagoruyko et al., Learning to compare image patches via convolutional neural networks. CVPR 2015.

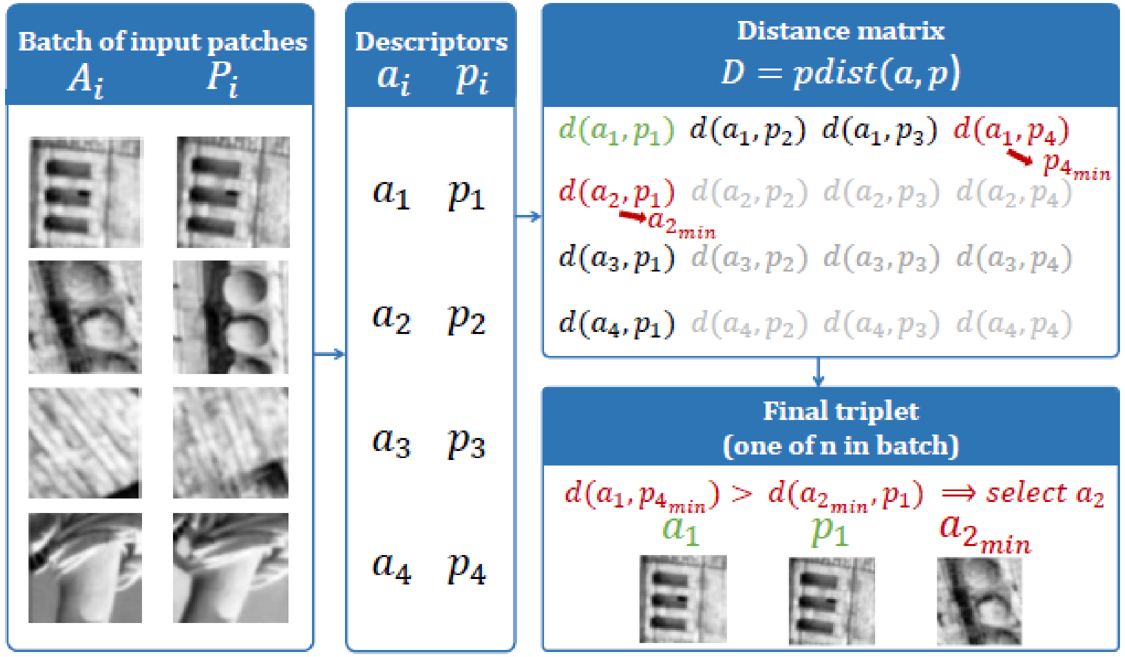
Introduction

Deep learning based approaches (Single network)

L2-Net

3×3 Conv 32
3×3 Conv 32
3×3 Conv 64 /2
3×3 Conv 64
3×3 Conv 128 /2
3×3 Conv 128
8×8 Conv 128
LRN

HardNet



L2-Net: Tian et al., L2-Net: Deep learning of discriminative patch descriptor in Euclidean space. CVPR 2017.
HardNet: Mishchuk et al., Working hard to know your neighbor's margins: Local descriptor learning loss. NIPS 2017.

Drawbacks

Siamese network

Networks with metric learning layers typically treat the matching of local patches as a binary classification task, so there does not exist the concept of descriptor.

Single network

Networks perform descriptor learning by combing hard mining sampling strategies with Siamese loss or triplet loss without fully utilizing context information.

Solution

- First order similarity loss

1. Generate a batch of matching patches $\mathrm{X} = \{(A_i, P_i), i = 1,2, \ldots, n\}$ and the corresponding descriptors $\chi = \{(a_i, p_i), i = 0,1, \ldots, n\}$
2. Compute the distance matrix $D = \{d(a_i, p_j) = \left\|a_i - p_j\right\|_2, i = 0,1, \ldots, n; j = 1,2, \ldots, n\}$
3. Find k closest non-matching descriptors $\mathrm{K}_i = \{q_{i,m}, m = 1,2, \ldots, k\}$ for each anchor patch descriptor a_i and ensure $d(a_i, q_{i,1}) \leq d(a_i, q_{i,2}) \leq \cdots \leq d(a_i, q_{i,k})$
4. Build a virtual descriptor v_i for a_i, $d(a_i, v_i) = \sum_{m=1}^{k} W_{i,m} d(a_i, q_{i,m})$, $w_{i,m} = \left(\frac{d(a_i, q_{i,k})}{d(a_i, q_{i,m})}\right)^2$, $W_{i,m} = \frac{w_{i,m}}{\sum_{l=1}^{k} w_{i,l}}$
5. Adopt the anchor swap strategy and create another virtual descriptor z_i for p_i

$$\mathcal{L}_1 = \frac{1}{n} \sum_{i=1}^{n} \max\left(0, 1 + d(a_i, p_i) - \min(d(a_i, v_i), d(p_i, z_i))\right)$$

Solution

- Second order similarity regularization

1. Compute two distance matrixes $D_a = \{d(a_i, a_j), i = 1,2, \ldots, n; j = 1,2, \ldots, n\}$ and $D_p = \{d(p_i, p_j), i = 1,2, \ldots, n; j = 1,2, \ldots, n\}$
2. Find t closest descriptors $S_i = \{(a_{s_{i,j}}), j = 1,2, \ldots, t\}$ for a_i and ensure $d(a_i, a_{s_{i,1}}) \leq d(a_i, a_{s_{i,2}}) \leq \cdots \leq d(a_i, a_{s_{i,t}})$
3. Find t closest descriptors $C_i = \{(p_{c_{i,j}}), j = 1,2, \ldots, t\}$ for p_i and $d(p_i, p_{c_{i,1}}) \leq d(p_i, p_{c_{i,2}}) \leq \cdots \leq d(p_i, p_{c_{i,t}})$.
4. Compute the regularization term $\mathcal{L}_2 = \mathcal{L}_{2a} + \mathcal{L}_{2p}$

$$\mathcal{L}_{2a} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{t} W_{i,j}^{A}\left|d\left(a_i, a_{s_{i,j}}\right) - d\left(p_i, p_{s_{i,j}}\right)\right|,$$

$$w_{i,j}^{a} = \left(\frac{d\left(a_i, a_{s_{i,t}}\right)}{d\left(a_i, a_{s_{i,j}}\right)}\right)^2,$$

$$W_{i,j}^{A} = \frac{w_{i,j}^{a}}{\sum_{l=1}^{t} w_{i,l}^{a}},$$

$$\mathcal{L}_{2p} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{t} W_{i,j}^{P}\left|d\left(a_i, a_{c_{i,j}}\right) - d\left(p_i, p_{c_{i,j}}\right)\right|,$$

$$w_{i,j}^{p} = \left(\frac{d\left(p_i, p_{c_{i,t}}\right)}{d\left(p_i, p_{c_{i,j}}\right)}\right)^2,$$

$$W_{i,j}^{P} = \frac{w_{i,j}^{p}}{\sum_{l=1}^{t} w_{i,l}^{p}}.$$

ICASSP 2021 TORONTO Canada June 6-11, 2021 Metro Toronto Convention Centre

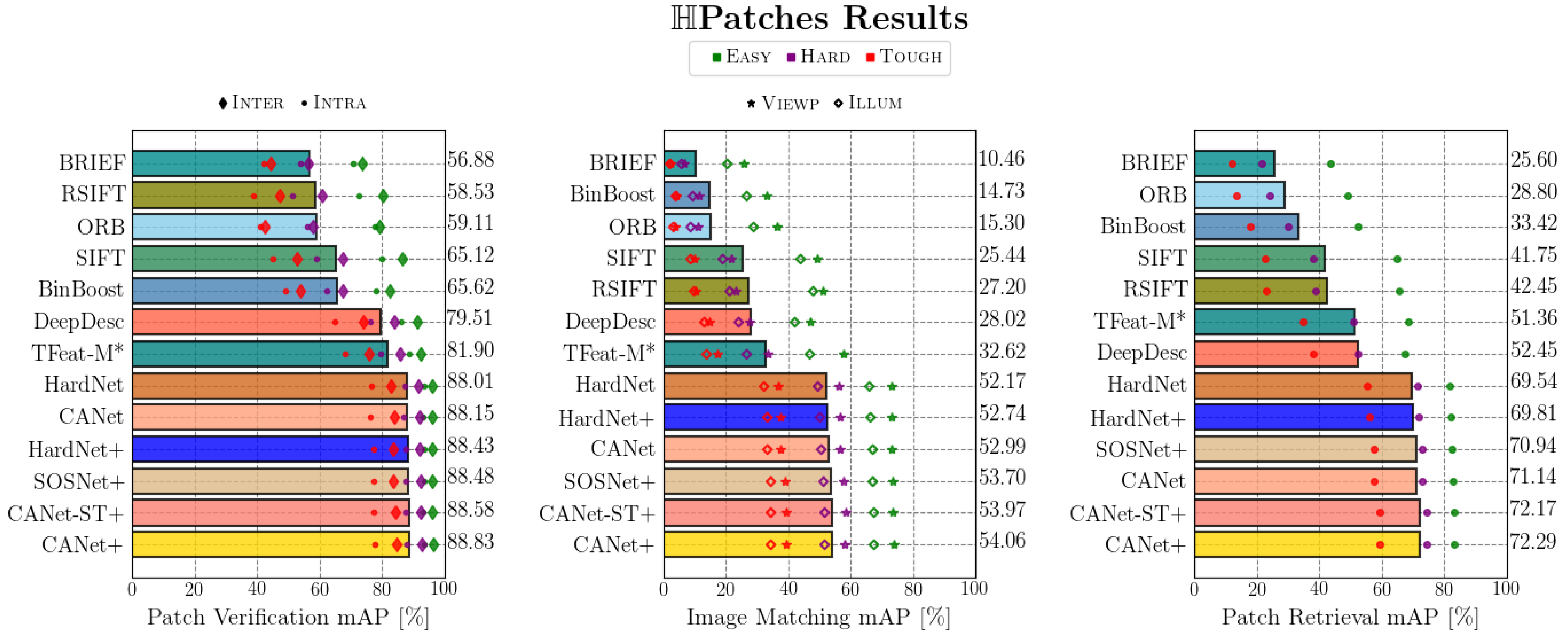
Results and comparisons

UBC Phototour dataset

Training	Notredame	Yosemite	Liberty	Yosemite	Liberty	Notredame	Mean
Test	Liberty		Notredame		Yosemite		
SIFT [4]	29.84		22.53		27.29		26.55
DeepDesc [14]	10.9		4.40		5.69		6.99
MatchNet [20]	7.04	11.47	3.82	5.65	11.6	8.7	8.05
TFeat-M [15]	7.39	10.31	3.06	3.8	8.06	7.24	6.64
L2Net [21]	3.64	5.29	1.15	1.62	4.43	3.30	3.24
CS L2Net [21]	2.55	4.24	0.87	1.39	3.81	2.84	2.61
HardNet [10]	1.47	2.67	0.62	0.88	2.14	1.65	1.57
Keller et. al. [16]	1.79	2.96	0.68	1.02	2.51	1.64	1.77
SOSNet [11]	1.25	2.84	0.58	0.87	1.95	1.25	1.46
CANet (Ours)	1.19	2.69	0.42	0.77	1.58	1.15	1.30
Data Augmentation							
L2Net+ [21]	2.36	4.7	0.72	1.29	2.57	1.71	2.23
CS L2Net+ [21]	1.71	3.87	0.56	1.09	2.07	1.3	1.76
HardNet+ [10]	1.49	2.51	0.53	0.78	1.96	1.84	1.51
DOAP+ [23]	1.54	2.62	0.43	0.87	2.00	1.21	1.45
DOAP-ST+ [23-24]	1.47	2.29	0.39	0.78	1.98	1.35	1.38
SOSNet+ [11]	1.08	2.12	0.35	0.67	1.03	0.95	1.03
CANet+ (Ours)	1.29	2.46	0.45	0.75	1.23	1.10	1.21
CANet-ST+ (Ours)	1.25	2.49	0.42	0.69	1.36	1.15	1.23

ICASSP 2021 TORONTO Canada
June 6-11, 2021
Metro Toronto Convention Centre

Results and comparisons



ICASSP2021
TORONTO
Canada
June 6-11, 2021
Metro Toronto Convention Centre

2021 IEEE International Conference on Acoustics, Speech and Signal Processing
6-11 June 2021 • Toronto, Ontario, Canada
Extracting Knowledge from Information