一种大规模图像检索算法 ——DELF

Noh, H., Araujo, A., Sim, J., Weyand, T., & Han, B. (2017). Large-Scale Image Retrieval with Attentive Deep Local Features. *2017 IEEE International Conference on Computer Vision (ICCV)*, 3476-3485.

全局特征解决不了的问题

- 背景复杂
- 有遮挡
- 光照变化

Google-Landmarks Dataset

1. dense localized feature extraction
2. keypoint selection
3. dimensionality reduction
4. indexing and retrieval

Dense localized feature extraction

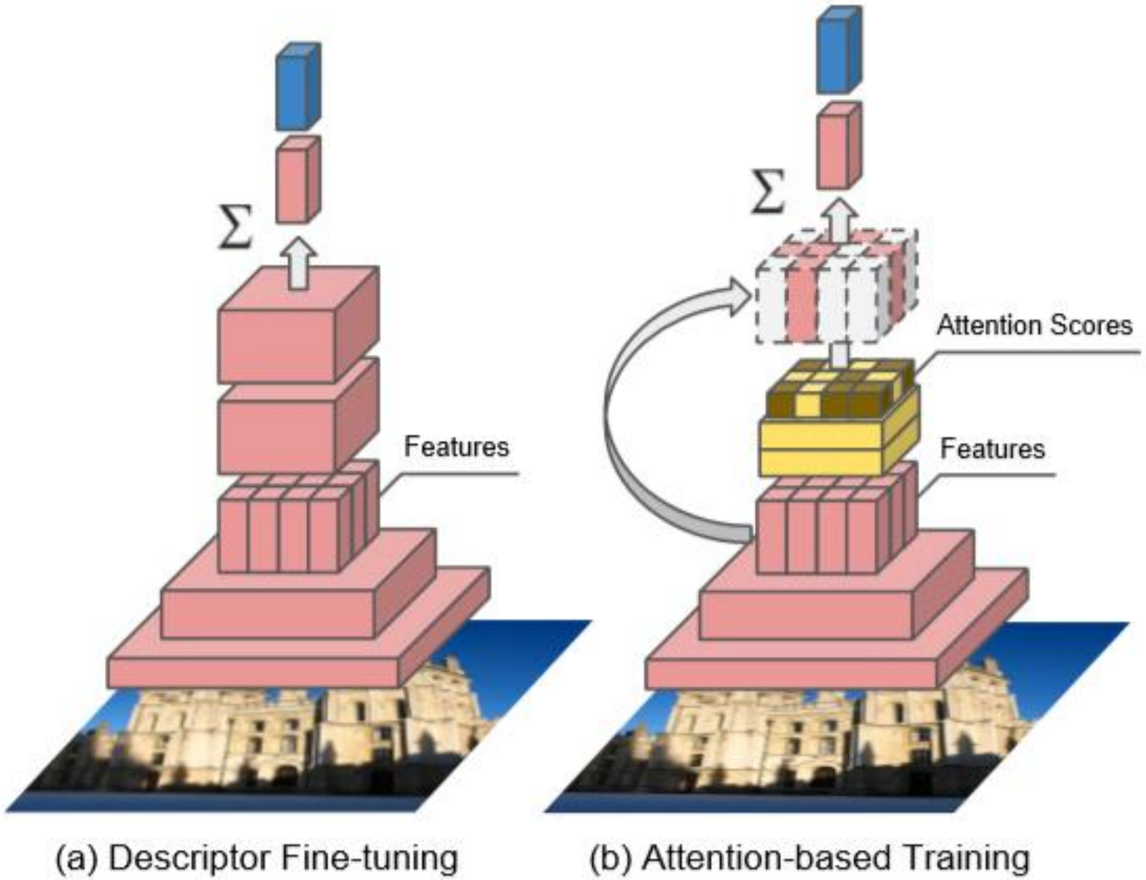
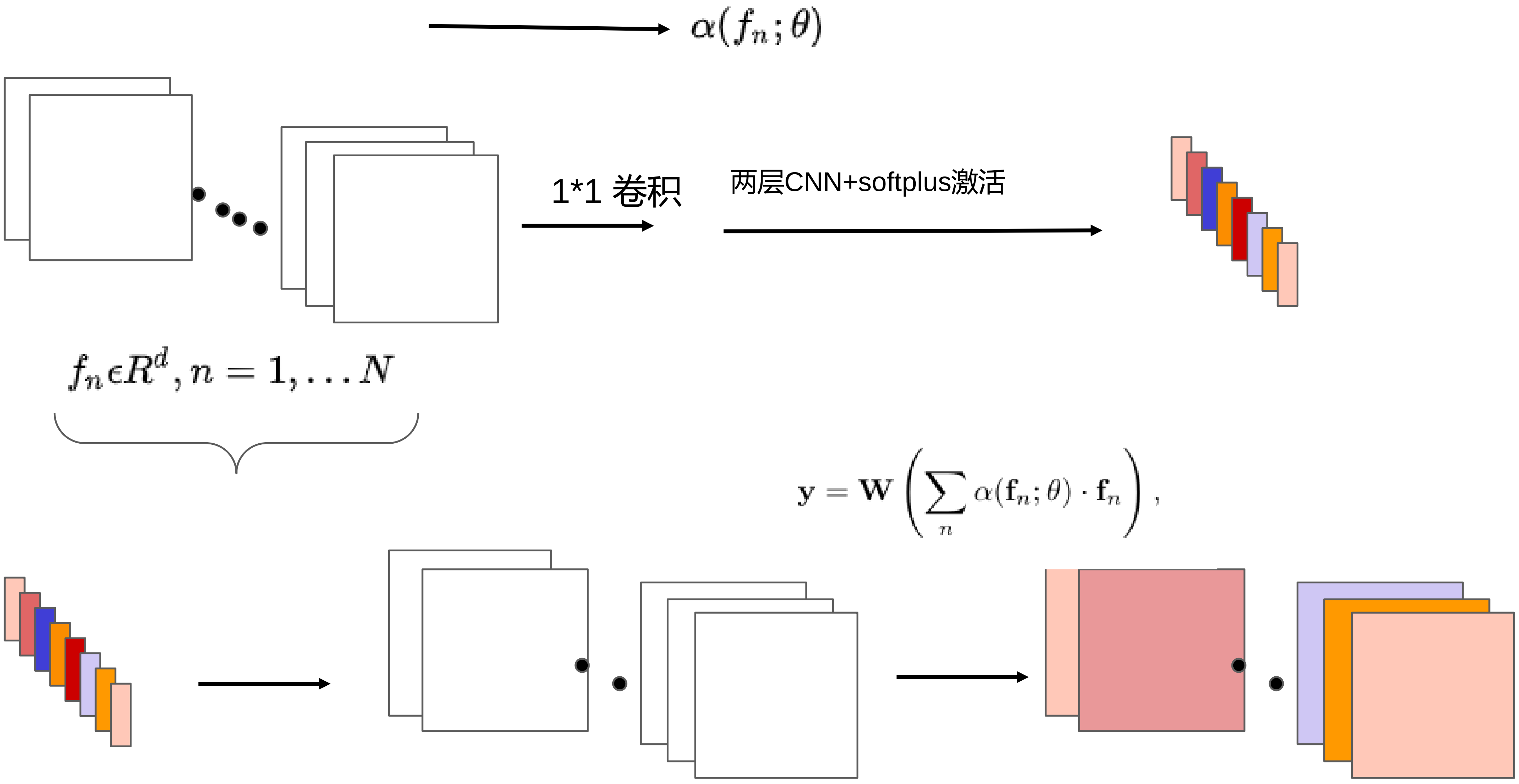


Figure 4: The network architectures used for training.

attention



$\alpha(f_n;\theta)$
1*1 卷积
两层CNN+softplus激活
$f_n \epsilon R^d, n = 1, \ldots N$
$\mathbf{y} = \mathbf{W}\left(\sum_n \alpha(\mathbf{f}_n;\theta)\cdot \mathbf{f}_n\right),$

计算分数

预测值: $$\mathbf{y} = \mathbf{W}\left(\sum_n \alpha(\mathbf{f}_n;\theta)\cdot \mathbf{f}_n\right),$$

真值: $\mathbf{y}^*$

损失函数: $$\mathcal{L} = -\mathbf{y}^* \cdot \log\left(\frac{\exp(\mathbf{y})}{\mathbf{1}^{\mathrm{T}}\exp(\mathbf{y})}\right)$$

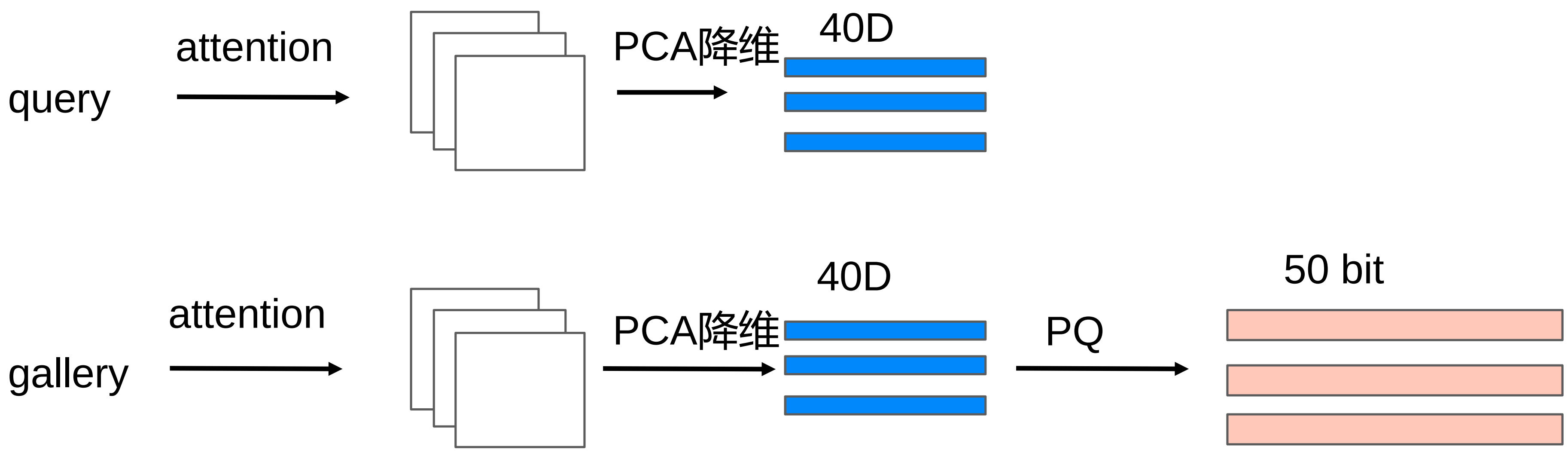
反向传播: $$\frac{\partial \mathcal{L}}{\partial \theta} = \frac{\partial \mathcal{L}}{\partial \mathbf{y}}\sum_n \frac{\partial \mathbf{y}}{\partial \alpha_n}\frac{\partial \alpha_n}{\partial \theta} = \frac{\partial \mathcal{L}}{\partial \mathbf{y}}\sum_n \mathbf{W}\,\mathbf{f}_n \frac{\partial \alpha_n}{\partial \theta}$$

获得局部特征的分数: $\alpha_n \equiv \alpha(\mathbf{f}_n;\theta)$

Dimensionality Reduction

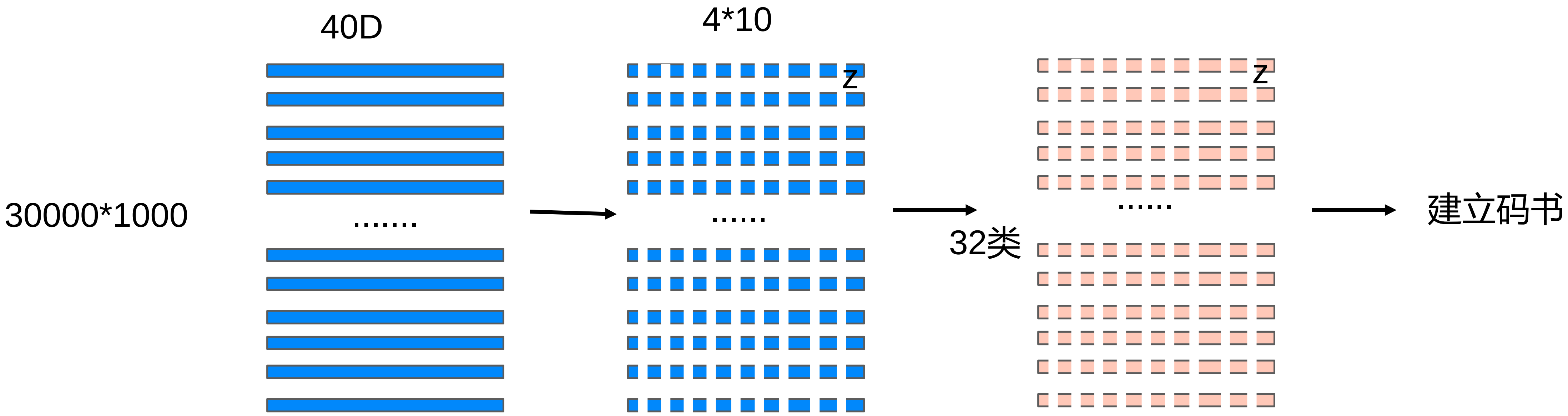
L2正则化+ pca降维 +L2正则化

Indexing

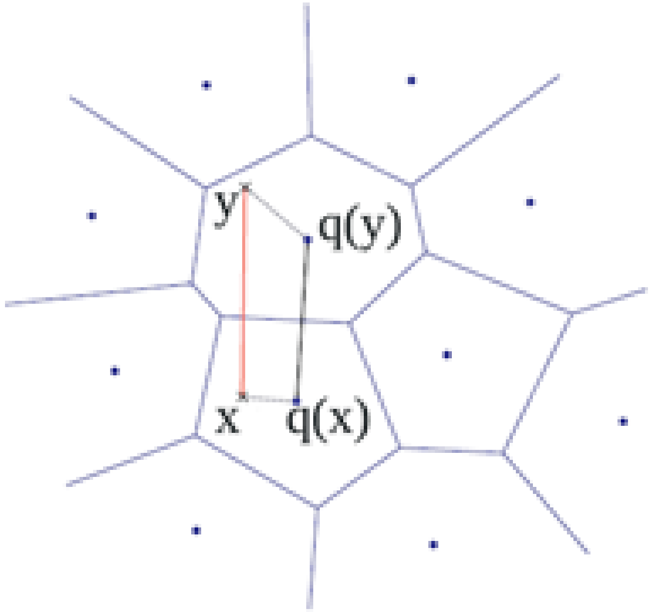


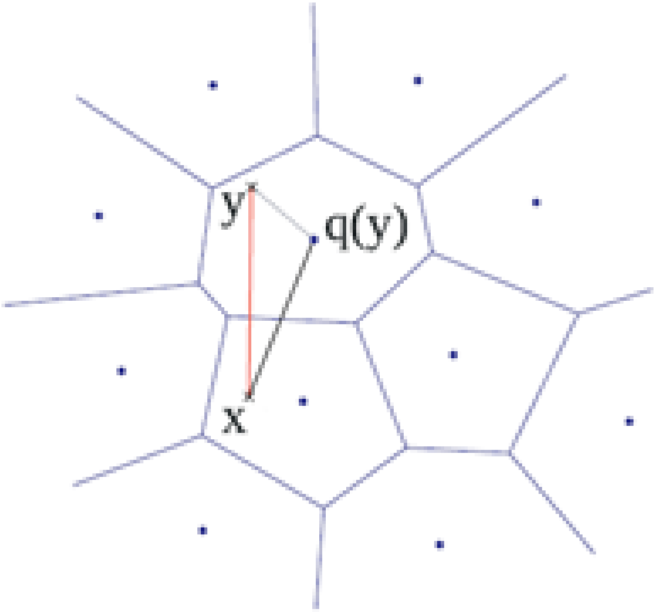
attention
PCA降维
40D
query
attention
PCA降维
40D
PQ
50 bit
gallery

PQ(Product Quantization)



非对称计算

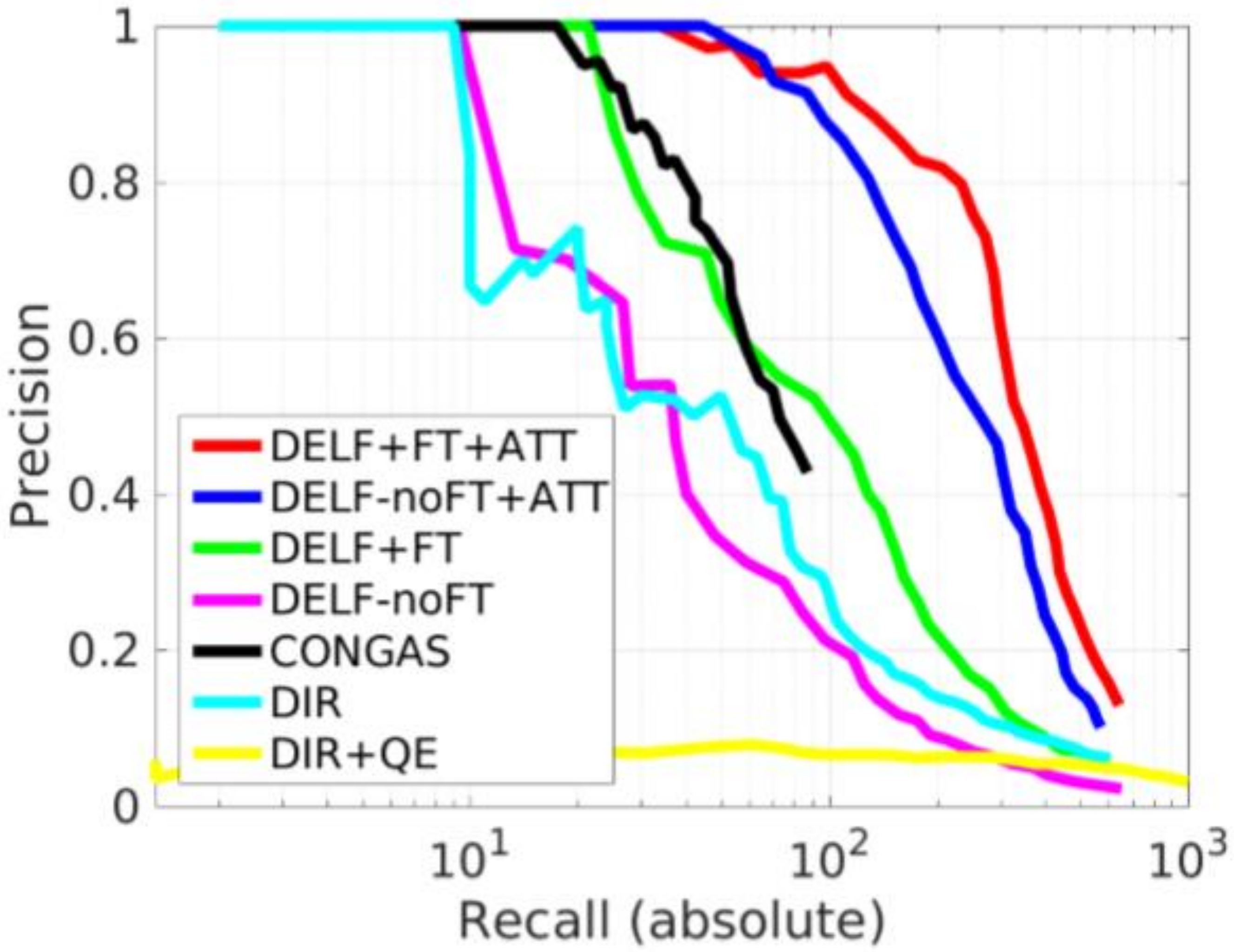




Retrieval

query $\xrightarrow{\text{最近邻检索}}$ 选取topk个局部特征 $\xrightarrow{\text{RANSAC}}$ 局内点

实验结果



Precision
Recall (absolute)
1
0.8
0.6
0.4
0.2
0
10^1
10^2
10^3
DELF+FT+ATT
DELF-noFT+ATT
DELF+FT
DELF-noFT
CONGAS
DIR
DIR+QE

(a) (b) (c)

Figure 6: Examples where DELF+FT+ATT outperforms DIR: (a) query image, (b) top-1 image of DELF+FT+ATT, (c) top-1 image of DIR. The green borders denote correct results while the red ones mean incorrect retrievals. Note that DELF deals with clutter in query and database images and small landmarks effectively.

(a) (b) (c)

Figure 7: Examples where DIR outperforms DELF+FT+ATT: (a) query image, (b) top-1 image of DELF+FT+ATT, (c) top-1 image of DIR. The green and red borders denotes correct and incorrect results, respectively.

Figure 9: Comparison of keypoint selection methods. (a) Input image (b) L_2 norm scores using the pretrained model (DELF-noFT) (c) L_2 norm scores using fine-tuned descriptors (DELF+FT) (d) Attention-based scores (DELF+FT+ATT). Our attention-based model effectively disregards clutter compared to other options.

Table 1: Performance evaluation on existing datasets in mAP (%). All results of existing methods are based on our reproduction using public source codes. We tested LIFT only on Oxf5k and Par6k due to its slow speed. (* denotes the results from the original papers.)

Dataset	Oxf5k	Oxf105k	Par6k	Par106k
DIR [11]	86.1	82.8	94.5	90.6
DIR+QE [11]	87.1	85.2	95.3	91.8
siaMAC [29]	77.1	69.5	83.9	76.3
siaMAC+QE [29]	81.7	76.6	86.2	79.8
CONGAS [8]	70.8	61.1	67.1	56.8
LIFT [40]	54.0	–	53.6	–
DIR+QE* [11]	89.0	87.8	93.8	90.5
siaMAC+QE* [29]	82.9	77.9	85.6	78.3
DELF+FT+ATT (ours)	83.8	82.6	85.0	81.7
DELF+FT+ATT+DIR+QE (ours)	**90.0**	**88.5**	**95.7**	**92.8**