清華大學
Tsinghua University

i-VisionGroup

L2-Net: Deep Learning of Discriminative Patch Descriptor in Euclidean Space(CVPR2017)

Yurun Tian Bin Fan Fuchao Wu

National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China

Network architecture

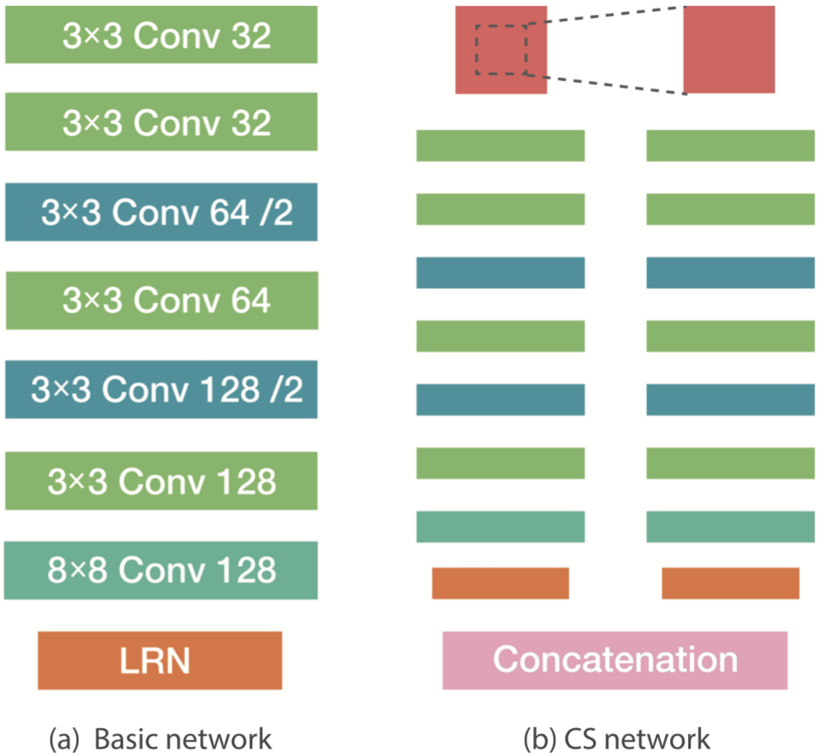

Figure 1. Network Architecture. 3×3 Conv = Convolution +Batch Normalization + Relu. 8×8 Conv = Convolution +Batch Normalization.

LRN: local response normalization
(这里用在全联接层之后，实际上就是L2归一化)

Dataset

- Brown dataset and Hpatches
- Each patch in the dataset has a unique 3D point index, patches with identical 3D point index are matching ones.
- For each 3D point, there are at least 2 matching patches
- All patches are down sampled to 32x32 for training.

Sampling strategy

- wholedataset: P points
- The sampling strategy:
  take p1 points sequentially, and then take an extra of p2 points fromthe rest (P − p1) points randomly. (a mixup of trained points and new points)

- A batch: $X = \{X_1^1, X_1^2, \ldots, X_i^1, X_i^2, \ldots X_p^1, X_p^2\}$
  Output: $Y = \{Y_1^1, Y_1^2, \ldots, Y_i^1, Y_i^2, \ldots Y_p^1, Y_p^2\}$
  distance matrix: $D = \sqrt{2(1 - Y_1^T Y_2)}$
  similarity matrix:

$$s_{ij}^c = \exp(2 - d_{ij}) / \sum_m \exp(2 - d_{mj})$$
$$s_{ij}^r = \exp(2 - d_{ij}) / \sum_n \exp(2 - d_{jn})$$

Loss functions

$$E_1 = -\frac{1}{2}\left(\sum_i \log s_{ii}^c + \sum_i \log s_{ii}^r\right)$$

Loss function for descriptors compactness:
目的是减少维度与维度之间的相关性

$$\text{denote } \mathbf{Y}_\mathbf{s}^\mathbf{T} \text{ as } \left[\mathbf{b}_1^s, \cdots, \mathbf{b}_i^s, \cdots, \mathbf{b}_q^s\right]$$

$$r_{ij}^s = \frac{\left(\mathbf{b}_i^s - \bar{b}_i^s\right)^T\left(\mathbf{b}_j^s - \bar{b}_j^s\right)}{\sqrt{\left(\mathbf{b}_i^s - \bar{b}_i^s\right)^T\left(\mathbf{b}_i^s - \bar{b}_i^s\right)}\sqrt{\left(\mathbf{b}_j^s - \bar{b}_j^s\right)^T\left(\mathbf{b}_j - \bar{b}_j^s\right)}}$$

$$E_2 = \frac{1}{2}\left(\sum_{i \neq j}\left(r_{ij}^1\right)^2 + \sum_{i \neq j}\left(r_{ij}^2\right)^2\right)$$

Loss function

- Immediate feature maps

$$E_3 = -\frac{1}{2}\left(\sum_i \log v_{ii}^c + \sum_i \log v_{ii}^r\right)$$

- Total loss function
  $E = E_1 + E_2 + E_3$

Hard-Net
Working hard to know your neighbor’s margins: Local descriptor learning loss(NIPS2017)

1 Szkocka Research Group, Ukraine
anastasiya.mishchuk@gmail.com
2 Visual Recognition Group, CTU in Prague

Loss function

- A batch: $X = \left\{X_1^1, X_1^2, \dots, X_i^1, X_i^2, \dots X_p^1, X_p^2\right\}$
  Output: $Y = \left\{Y_1^1, Y_1^2, \dots, Y_i^1, Y_i^2, \dots Y_p^1, Y_p^2\right\}$
  distance matrix: $D = \sqrt{2(1 - Y_1^T Y_2)}$

- Triple margin Loss function:

$$L = \frac{1}{n} \sum_{i=1,n} \max\left(0, 1 + d(a_i, p_i) - \min\left(d(a_i, p_{j_{min}}), d(a_{k_{min}}, p_i)\right)\right)$$

SOSnet:Second Order Similarity Regularization for Local Descriptor Learning

Yurun Tian∗,1,2 Xin Yu3 Bin Fan1 Fuchao Wu1
Huub Heijnen4 Vassileios Balntas4

1 National Laboratory of Pattern Recognition,
Institute of Automation,

Chinese Academy of Sciences, Beijing, China

SOSNet loss function

- Second order similarity

  difference between distance:

$$d^{(2)}(\boldsymbol{x}_i, \boldsymbol{x}_i^+) = \sqrt{\sum_{j \neq i}^{N} \left(d(\boldsymbol{x}_i, \boldsymbol{x}_j) - d(\boldsymbol{x}_i^+, \boldsymbol{x}_j^+)\right)^2},$$

$$\mathcal{R}_{\text{SOS}} = \frac{1}{N} \sum_{i=1}^{N} d^{(2)}(\boldsymbol{x}_i, \boldsymbol{x}_i^+).$$

$$\mathcal{L}_{\text{FOS}} = \frac{1}{N} \sum_{i=1}^{N} \max\left(0, t + d_i^{\text{pos}} - d_i^{\text{neg}}\right)^2,$$

$$d_i^{\text{pos}} = d(\boldsymbol{x}_i, \boldsymbol{x}_i^+),$$

$$d_i^{\text{neg}} = \min_{\forall j, j \neq i} \left(d(\boldsymbol{x}_i, \boldsymbol{x}_j), d(\boldsymbol{x}_i, \boldsymbol{x}_j^+), d(\boldsymbol{x}_i^+, \boldsymbol{x}_j), d(\boldsymbol{x}_i^+, \boldsymbol{x}_j^+)\right),$$

SOS

- Explanation
  满足first order similarity的描述子空间还是太大了(只需要matching points描述子距离小于non-matching points描述子距离) 添加second order similarity约束之后, 可能能够减过拟合