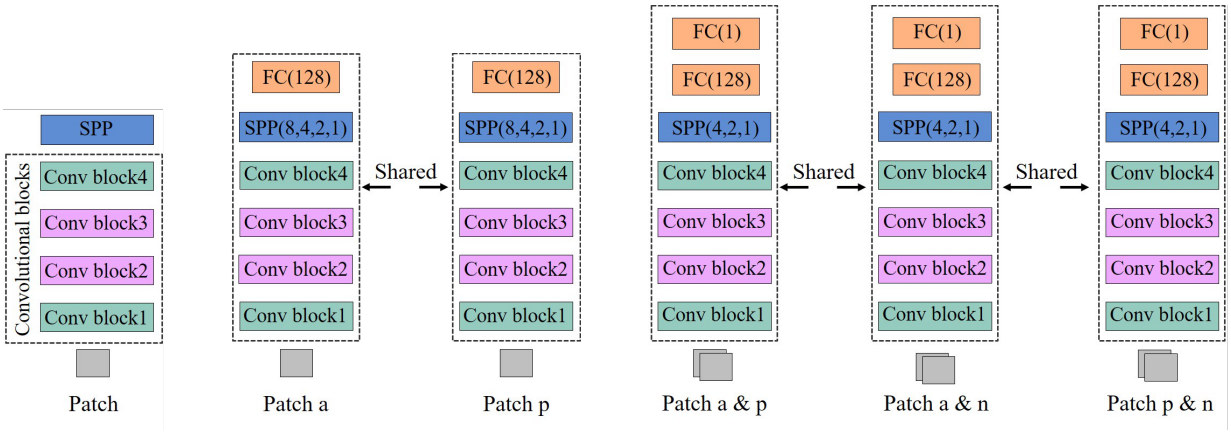
Supplementary material for the main article: Better and Faster: Exponential Loss for Image Patch Matching

Shuang Wang[1] Yanfeng Li[1] Xuefeng Liang[1,2] Dou Quan[1] Bowu Yang[1]
Shaowei Wei[1] Licheng Jiao[1]
[1]School of Artificial Intelligence, Xidian University,Shaanxi,China [2]Kyoto University, Kyoto, Japan
xliang@xidian.edu.cn

A. Model architectures



(a) Feature network (b) Siamese network (c) Triplet network

Figure 1: The networks used in our paper. (a) is the shared feature network. (b) is the Siamese network used in descriptor learning. (c) is the triplet network used in metric learning.

Figure 1 shows the network architectures used in this work. Figure 1 (a) is the shared feature network. The model contains four convolutional blocks and one SPP operator. Conv block1: C(32,3,1,1)-BN-ReLU-C(32,3,1,1)-BN-ReLU. Conv block2: C(64,3,1,2)-BN-ReLU-C(64,3,1,1)-BN-ReLU. Conv block3: C(128,3,1,2)-BN-ReLU-C(128,3,1,1)-BN-ReLU. Conv blcok4: C(128,3,1,1)-BN-ReLU-C(128,3,1,1)-BN. Figure 1 (b) is the Siamese network used in descriptor learning. The model has two branches that share weights. Two branches have the same settings as in the shared feature network. The output layer is a fully connected layer (FC) with 128 output units to obtain the feature descriptor. Figure 1 (c) is the triplet network used in metric learning. The model has three branches that share weights. For training efficiency, we use 3-level pyramid pooling and remove the last convolutional layer. What's more, we remove all BN layers. The model architecture is as following. Conv block1: C(32,3,1,1)-ReLU-C(32,3,1,1)-ReLU. Conv block2: C(64,3,1,2)-ReLU-C(64,3,1,1)-ReLU. Conv block3: C(128,3,1,2)-ReLU-C(128,3,1,1)-ReLU. Conv block4: Dropout(0.3)-C(128,3,1,1). To output a pairwise similarity, we add two fully connected linear layers: FC(128)-ReLU-FC(1)-Sigmoid, where C(n,k,s,d) is a convolutional layer with n filters of kernel size $k \times k$ applied with stride s and dilation d; BN represents Batch Normalization; SPP(a,b,c,d) means that 4-level pyramid pooling $a \times a, b \times b, c \times c, d \times d$ are included in a SPP operator.

B. Stability of exponential loss

To demostrate the stability of our proposed Exp-SLoss and Exp-TLoss, we conduct experiments on UBC and RGB-NIR benchmarks using our best training settings as discussed in the main paper. In particular, we utilize Exp-SLoss, Exp-TLoss,

l_2 loss and l_2^2 loss to optimize the network, and then plot the loss against training epochs in Figure 2 and Figure 3. As can be seen that the l_2 loss decreases slowly because it treats all training samples linearly, and can't effectively use hard samples to speed up training. For UBC benchmark, the l_2^2 loss value does not decrease in initial a few training epochs. For RGB-NIR benchmark, the network using l_2^2 loss dose not converge at all. By contrast, we train the network using the l_2 loss in the first epoch, and then replace it by our Exp-SLoss and Exp-TLoss in subsequent training. The experiment results show that our training strategy is more stable and performs better.

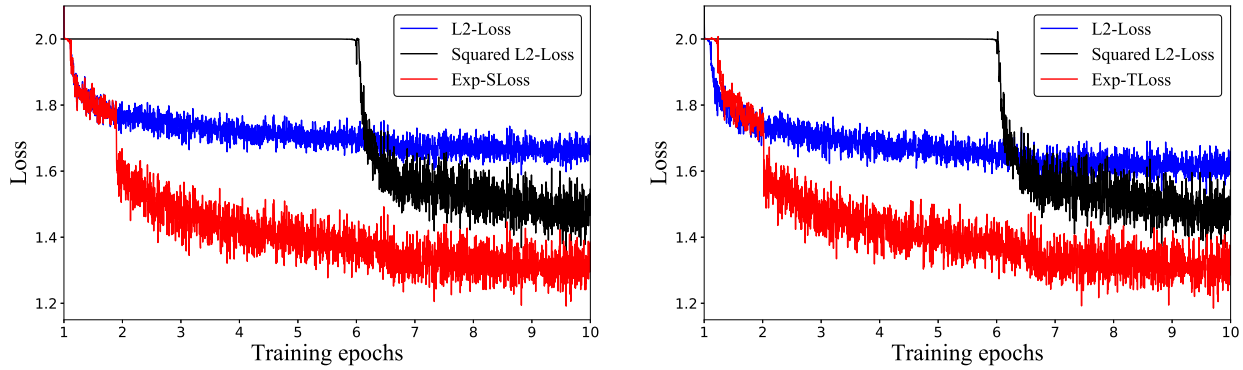


Figure 2: Training process using different loss functions on the UBC benchmark. The left and right figure show the comparisons for Exp-SLoss and Exp-TLoss respectively.

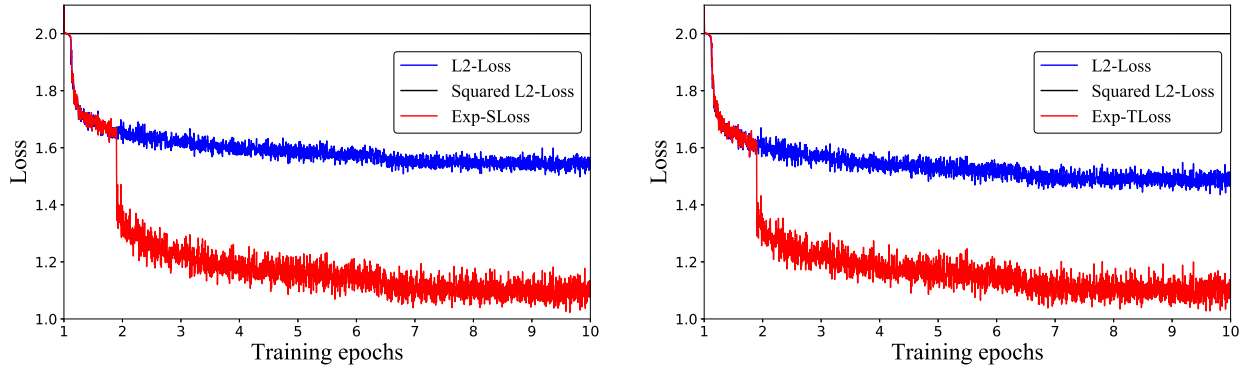


Figure 3: Training process using different loss functions on the RGB-NIR benchmark. The left and right figure show the comparisons for Exp-SLoss and Exp-TLoss respectively.

C. Stability of exponential loss

To study the influence of the hard sampling (of both positive and negative examples) more extensively, we do experiments using four sampling strategies: no hard sample mining (None), hard positive mining (HardP), hard negative mining (HardN) and hard positive and negative mining (Both). Table 1 below shows that both hard positive and negative mining performs the best.

Methods	None	HardP	HardN	Both
Metric learning	4.15/24.27	—	0.76/3.92	—
Desriptor learning	16.02/0.78	15.11/0.71	1.07/1.07	1.03/0.8

Table 1: Performance (FPR95/training time (hour)) using different sample mining strategies.