Supplementary Material for DeepCD: Learning Deep Complementary Descriptors for Patch Representations

Anonymous ICCV submission

Paper ID 1557

In this supplementary material, we show some details and figures that were not able to be included in the paper due to the space constraint. We first display the ROC curves (true positive rate vs. false positive rate) for several methods on all six training-testing configurations of the Brown dataset. Next, we give details about the computation cost for different methods. In Section 4.5 of the paper, we discuss the matching performance and the computation cost for different methods. We used floating point operations per pair (FLOPP) as the computation cost. Here, we detail how FLOPPs were obtained for these methods.

1. ROC curves

Figure 2 shows the ROC curves of several methods on the six training/testing combinations for the Brown dataset which we showed their FPR95 values in Table 1 of the paper. On the top of each plot, we show the combination of the training subset and the testing subset. For example, **liberty_to_notredame** stands for the case in which liberty was used for training and notredame was used for testing. Our method outperforms all competing methods in all six combinations.

2. Computation costs

In Section 4.5 of the paper, we discussed the trade-offs between the matching performance and the computation cost for different methods. Here, we consider the cost for calculating the distance between a pair of patches using their descriptors. Since bit operations are negligible compared to floating point operations, FLOPP (Floating point operations per pair) was used as metric for the computation cost. The result was shown in Figure 5 of the paper and is reproduced here as Figure 1 for clarity.

L_2 distance and Hamming distance. The calculation of the Hamming distance requires no floating point operation. For two 128-d descriptors, the calculation of L_2 distance requires $128 * 2 + (128 - 1) = 383$ FLOPP.

Decision network. Here, we take the decision network of the 2ch-2stream model as an example since the model

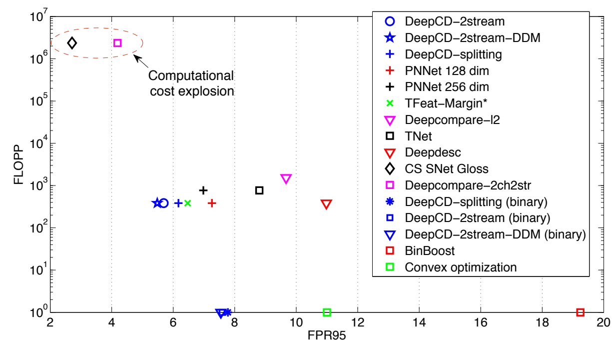

Figure 1. Performance-computation tradeoffs of different methods on the Brown dataset.

has the best performance in the Deepcompare and the Global loss paper. The structure of the decision network is $\{\text{FC}(1536 \rightarrow 768) + \text{ReLU} + \text{FC}(768 \rightarrow 1)\}$. We use the function $f\,(i \rightarrow t)$ to denote the total float point operations per image pair for a fully connected layer $y = Wx + b$ whose input vector $x_{(i\times 1)}$ contains i dimensions and the output vector $y_{(t\times 1)}$ contains t dimensions. The weight matrix $W_{(t\times i)}$ and the bias term $b_{(t\times 1)}$ are the parameters of the layer. We have $f\,(i \rightarrow t) = t * [i + (i-1)] + t = 2ti$. For the first fully connected layer, $f\,(1536 \rightarrow 768) = 2,359,296$. We ignore ReLU for simplicity. For the second fully connected layer, $f\,(768 \rightarrow 1) = 1,536$. In sum, it takes more than 2M FLOPP for calculating the distance between a patch pair if the 2ch-2stream decision network is used.

DeepCD. In DeepCD, in addition to the L_2 distance and the Hamming distance, product is required for late fusion. As an example, let's assume that we choose to use a 128-d real-valued leading descriptor and a 256-bit binary complementary descriptor. It would require $[128 * 2 + (128 - 1)] + 0 + 1 = 384$ FLOPP. If both the leading and the complementary descriptors are binary, it only requires 1 FLOPP.

Thus, there are roughly three classes of methods in Figure 1, the one with binary descriptors, the one involved with L_2 distance and the one with complicated decision networks.

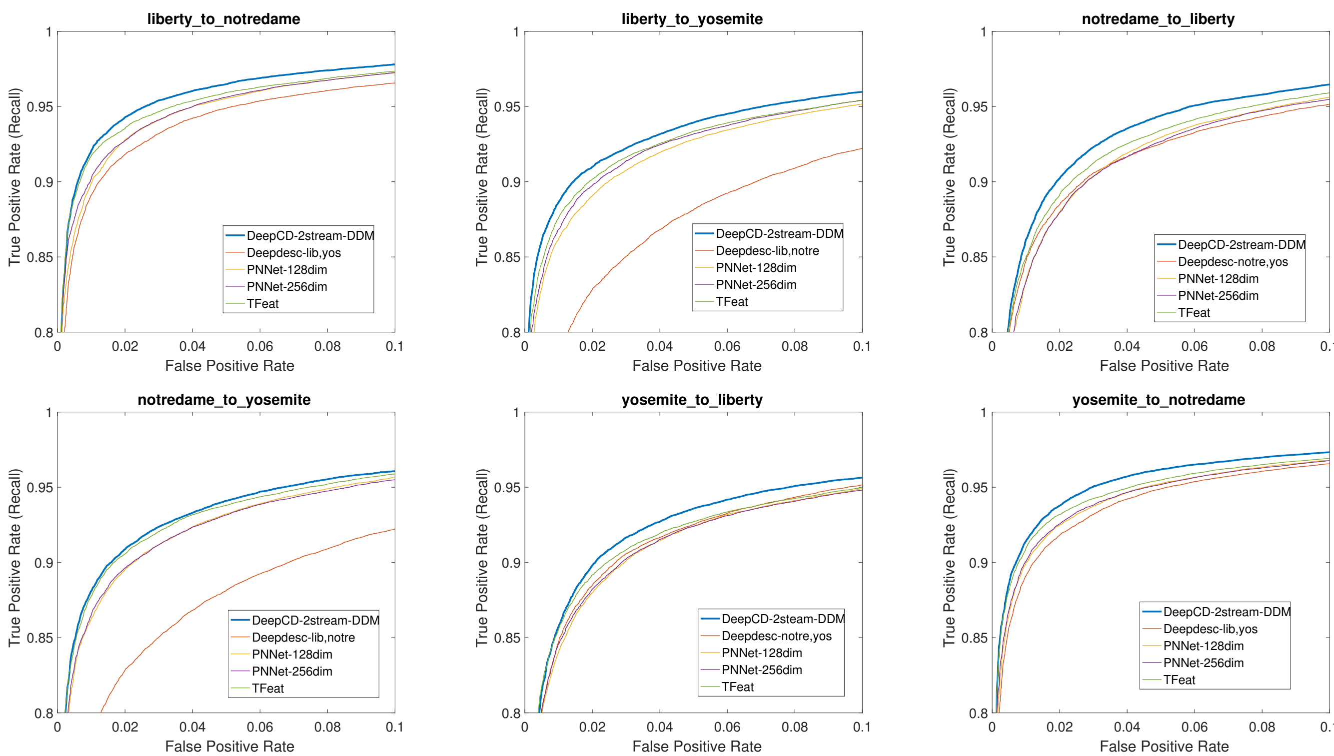

Figure 2. ROC curves of several methods on different training/testing combinations of the Brown dataset.