Milvus Paper sharing

Efficient Indexing of Billion-Scale datasets of deep descriptors

godchen fishpenguin

ZILLIZ

Content

1.Background

2.The NO-IMI structure

3.Indexing

4.Codebook learning

5.Query

6.Experiments

background

SIFT(Scale-invariant feature transform)

IVFADC and IMI have been performing very well, achieving state-of-the-art recall in several milliseconds.

DNN

SIFT-like descriptors, however, are quickly being replaced with descriptors based on deep neural networks (DNN) that provide better performance for many computer vision tasks.IMI is suitable for SIFT but not suitable for DNN because the structure is different.

The paper seeks to develop a good indexing structure for deep descriptors.

background

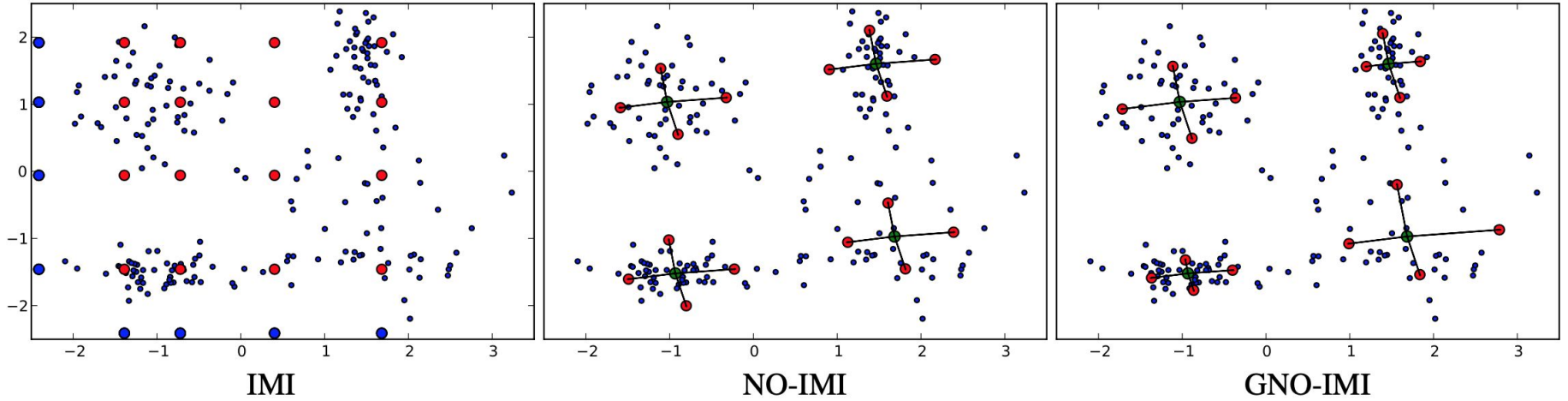


Figure 1. The cell centroids (red points) produced by different indexing structures for the same sample of two-dimensional data (blue points). For all structures parameter K was set to four, hence 16 centroids were produced. The left plot corresponds to the IMI structure and large blue points on the axes denote the codewords of the underlying PQ decomposition. The middle and right plots correspond to the NO-IMI and the GNO-IMI structures respectively. On the both plots green points correspond to the "first-order centroids" $S_1, \ldots, S_4$. The GNO-IMI centroids represent the actual data distribution more accurately than the other structures.

The NO-IMI structure

Let us assume that a database P = {p1,...,pN} of D-dimensional points is given.

The cells in the NO-IMI are constructed based on two codebooks S = {S1,...,SK} and T = {T1 , . . . , TK }

$$c_i^j = S_i + T_j, \qquad i, j = 1, \ldots, K \qquad C_i^j = \{x \in \mathbf{R}^D | i, j = \arg\min_{k,l} \|x - (S_k + T_l)\|^2\}$$

S1, . . . , SK can be interpreted as cluster centroids in the original data space and we refer to them as first-order centroids.

Similarly,the codewords T1,...,TK can be interpreted as centroids in the space of displacements of data points from the first-order centroids.

The generalized NO-IMI

an element α[i, j] is a factor for the j-th second- order centroid Tj in the i-th first-order cluster. After the incorporation of these factors, the expression for the correponding centroid becomes:

$$c_i^j = S_i + \alpha[i,j]T_j, \qquad i,j = 1,\ldots,K$$

Indexing

$$c_i^j = S_i + \alpha[i,j]T_j, \qquad i,j = 1,\ldots,K$$

Let us now assume that S, T , α are given and describe the process that distributes the dataset points p1 , . . . , pN assigning them to different (G)NO-IMI cells.

$$i,j = \arg\min_{k,l} \|p - (S_k + \alpha[k,l]T_l)\|^2$$

$$\|p - (S_k + \alpha[k,l]T_l)\|^2 = \|p - S_k\|^2 + \alpha[k,l]^2\|T_l\|^2$$
$$-2\alpha[k,l]\langle p, T_l\rangle + 2\alpha[k,l]\langle S_k, T_l\rangle \qquad (6)$$

codebook learning

$$\sum_{i=1}^{N} \left\| p_i - (S_{k^i} + \alpha[k^i, l^i] T_{l^i}) \right\|^2 \to \min_{\substack{S_k \in \mathbf{R}^D, |S| = K, \\ T_k \in \mathbf{R}^D, |T| = K, \\ \alpha \in \mathbf{R}^{D \times D} \\ k^i, l^i \in \{1, \ldots, K\}}}$$

$$\alpha[k, l] = \frac{\sum_{i: k_i = k, l_i = l} \langle p_i - S_k, T_l \rangle}{\sum_{i: k_i = k, l_i = l} \|T_l\|^2} \tag{10}$$

$$\sum_{i=1}^{N} \left\| p_i - (S_{k^i} + \alpha[k^i, l^i] T_{l^i}) \right\|^2 =$$

$$= \sum_{k=1}^{K} \sum_{l=1}^{K} \sum_{i: k_i = k, l_i = l} \|p_i - S_k - \alpha[k, l] T_l\|^2$$

$$\sum_{i: k_i = k, l_i = l} \|p_i - S_k - \alpha[k, l] T_l\|^2 \to \min_{\alpha[k,l]}$$

codebook learning

$$\sum_{i=1}^{N} \left\| p_i - (S_{k^i} + \alpha[k^i, l^i] T_{l^i}) \right\|^2 \to \min_{\substack{S_k \in \mathbf{R}^D, |S| = K, \\ T_k \in \mathbf{R}^D, |T| = K, \\ \alpha \in \mathbf{R}^{D \times D} \\ k^i, l^i \in \{1, \ldots, K\}}}$$

$$\alpha[k, l] = \frac{\sum_{i: k_i = k, l_i = l} \langle p_i - S_k, T_l \rangle}{\sum_{i: k_i = k, l_i = l} \|T_l\|^2} \tag{10}$$

$$\sum_{i=1}^{N} \left\| p_i - (S_{k^i} + \alpha[k^i, l^i] T_{l^i}) \right\|^2 =$$

$$= \sum_{k=1}^{K} \sum_{l=1}^{K} \sum_{i: k_i = k, l_i = l} \|p_i - S_k - \alpha[k, l] T_l\|^2$$

$$\sum_{i: k_i = k, l_i = l} \|p_i - S_k - \alpha[k, l] T_l\|^2 \to \min_{\alpha[k,l]}$$

codebook learning

similarly:

$$\alpha[k,l] = \frac{\sum\limits_{i:k_i=k,l_i=l} \langle p_i - S_k, T_l \rangle}{\sum\limits_{i:k_i=k,l_i=l} \|T_l\|^2} \tag{10}$$

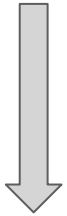
$$T_l = \frac{\sum\limits_{k=1}^{K} \alpha[k,l] \sum\limits_{i:k_i=k,l_i=l} (p_i - S_k)}{\sum\limits_{k=1}^{K} \sum\limits_{i:k_i=k,l_i=l} \alpha[k,l]^2}$$

$$S_k = \frac{\sum\limits_{l=1}^{K} \sum\limits_{i:k_i=k,l_i=l} (p_i - \alpha[k,l]T_l)}{\sum\limits_{l=1}^{K} \sum\limits_{i:k_i=k,l_i=l} 1}$$

Query

$$distances[k] = \|q - S_k\|^2, \quad k = 1, \ldots, K$$

$$k_1, \ldots, k_r = distances.argsort()[1:r]$$



$$cells = \{(k_i; l)\}$$

$$k_i \in \{k_1, \ldots, k_r\}, \quad l = 1, \ldots, K$$

for each $cell = (k_i; l)$ **in** $cells$**:**

$$cellDistances[(k_i; l)] = \|q - S_{k_i}\|^2 + \alpha[k_i, l]^2 \|T_l\|^2 - 2\alpha[k_i, l]\langle q, T_l\rangle + 2\alpha[k_i, l]\langle S_{k_i}; T_l\rangle$$

Experiments

1.SIFT1B dataset[13] that contains one billion of 128- dimensional SIFT descriptors along with precomputed groundtruth for 10, 000 queries. A hold-out learning set of 100 million descriptors is also provided.

2. DEEP1B dataset. Our DNN had the GoogLeNet architecture and was trained on the ImageNet dataset[1]. The outputs were then compressed by PCA to 96 dimensions and l2-normalized. We also prepared hold-out sets containing 350 millions of descriptors for learning and 10, 000 for querying (with known ground truth for nearest neighbors in the main set).

Experiments

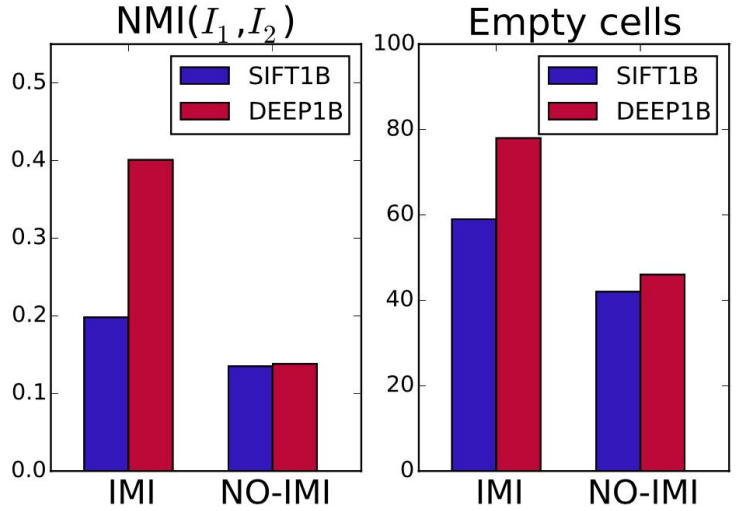


Figure 3. Normalized mutual information between the indices of the closest codewords in two codebooks (left) and the percent of empty index cells (right) for the IMI and NO-IMI systems. The high value of NMI in the case of DEEP1B and the IMI means that the implicit assumption of independent subspaces in the IMI does not hold for DEEP1B. The large percent of empty cells also indicates that cell centroids in the IMI represent DEEP1B data distribution poorly.

Experiments

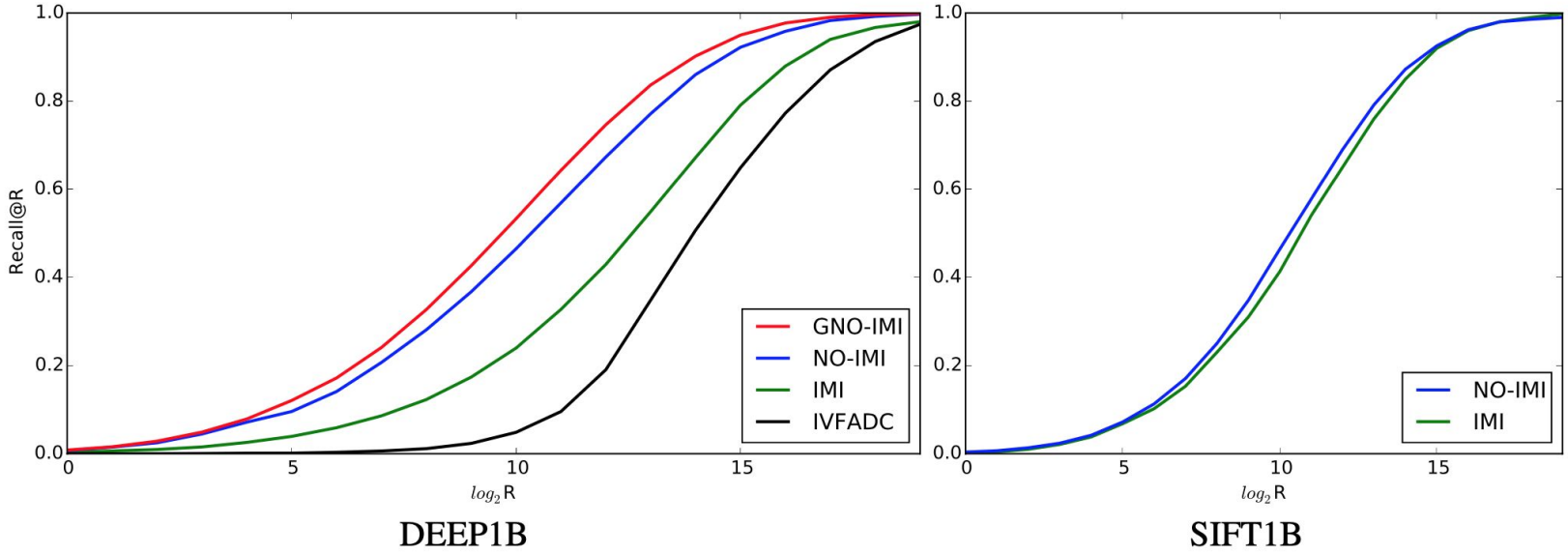


Figure 2. Recall as a function of the candidate list length. On DEEP1B (left plot) we compare four systems: IVFADC with 2^{17} codewords, the IMI with $K = 2^{14}$ and preliminary orthogonal transformation, the NO-IMI and the GNO-IMI with $K = 2^{14}$. For all recall levels the (G)NO-IMI provides much shorter candidate lists. For SIFT1B dataset (right plot) the advantage of the NO-IMI is negligible.

Experiments

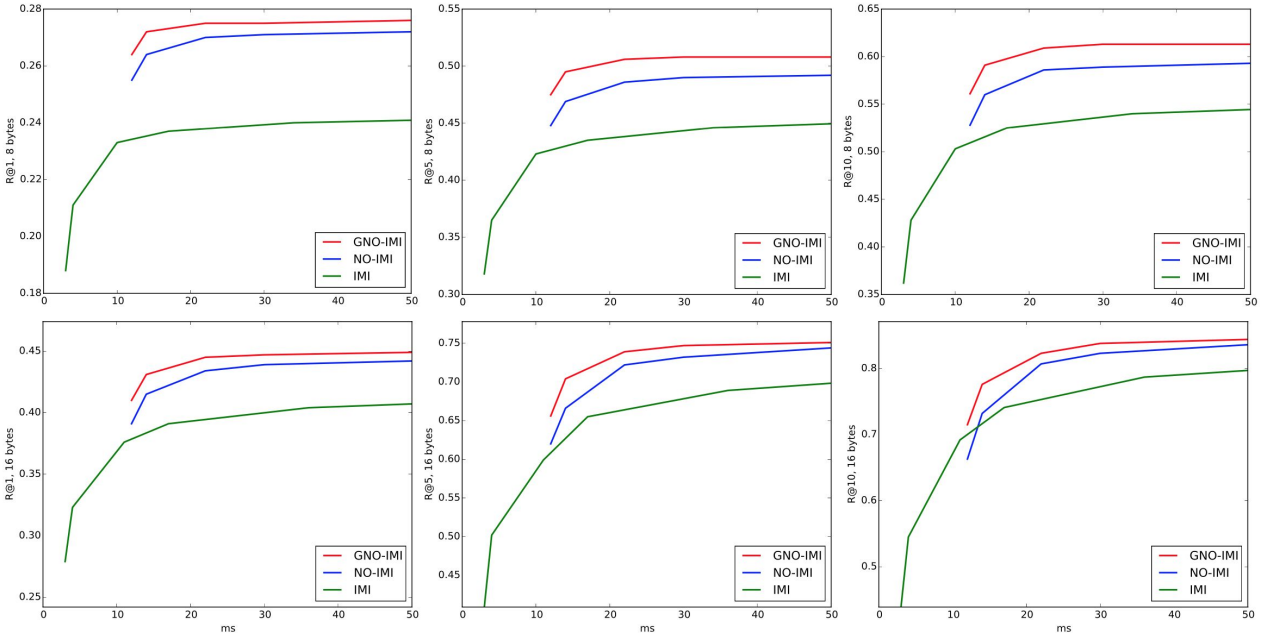


Figure 4. Comparison of the original IMI, NO-IMI and GNO-IMI in terms of recall after reranking, and runtime on the DEEP1B dataset. For all systems we used OPQ with local codebooks to compress database points. With few exceptions, for any given time budget above 11 ms the (G)NO-IMI provides considerably higher recall compared to the IMI-based scheme.

Thanks

ZILLIZ