Efficient Indexing of Regional Maximum Activations of Convolutions using Full-Text Search Engines

Giuseppe Amato, Fabio Carrara, Fabrizio Falchi, Claudio Gennaro

fabio.carrara@isti.cnr.it

Contribution

We exploit Deep Permutations [1] and Surrogate Text Representation [2] techniques to efficiently and effectively **index RMAC features using Lucene**, a Full-Text search engine based on inverted indexes.

- INRIA Holidays + MIRFlickr1M (K=127):
  - mAP**: ~0.70**
  - Time**: <0.4 s**
  - HW: Intel(R) Core(TM) i7-6800K CPU @ 3.40GHz x 8

RMAC features

- aggregated multi-region multi-resolution 2048-dimensional L2-normalized image descriptor
- have recently showed outstanding performance in visual instance retrieval - INRIA Holidays: 86.7 mAP, Oxford 5k: 83.1 mAP [3]
- **dense**, direct use of inverted indexes is difficult (they rely on sparsity of data)
- dot-product as similarity measure

Deep Permutations

Given a D-dimensional feature vector $x = [0.1, 0.3, 0.4, -0.15, 0.2]$

we represent it with a **permutation** (ranking -> component, obtained by arg-sorting in descending order) and its **inverse permutation** (component -> position):

$$\Pi_x = [3, 2, 5, 1, 2] \qquad \Pi_x^{-1} = [4, 2, 1, 5, 3]$$

The distance between two feature vectors x and y is defined as the **Spearman Rho** distance between their permutations:

$$d(x,y) = S_\rho(\Pi_x, \Pi_y) = \left\|\Pi_x^{-1} - \Pi_y^{-1}\right\|_2$$

Scoring Function

Given that $\|A - B\|^2 = \|A\|^2 + \|B\|^2 - 2A \cdot B$, we can rewrite:

$$d(x,y) = \underbrace{\left\|\Pi_x^{-1}\right\|_2^2 + \left\|\Pi_y^{-1}\right\|_2^2}_{constant} - 2 \cdot \Pi_x^{-1} \cdot \Pi_y^{-1}$$

Omitting constants, we define a scoring (similarity) function as follows:

$$s(x,y) = \Pi_x^{-1} \cdot \Pi_y^{-1}$$

Surrogate Text Representation

We use a **surrogate text representation** [2] of the inverse permutations to index them with Lucene:

1. First we **complement** the permutation to its size (D) plus one:
   $$\bar{\Pi}_x^{-1} = D + 1 - \Pi_x^{-1}$$
2. each component $\bar{\Pi}_x^{-1}(i)$ is associated with a codeword τ_i
3. each codeword τ_i is repeated $\bar{\Pi}_x^{-1}(i)$ times

E.g:

$$\begin{aligned} \Pi_x^{-1} &= [4, 1, 2, 5, 3] \\ \bar{\Pi}_x^{-1} &= [2, 5, 4, 1, 3] \\ \text{text} &= \tau_1\,\tau_1\,\tau_2\,\tau_2\,\tau_2\,\tau_2\,\tau_2\,\tau_3\,\tau_3\,\tau_3\,\tau_3\,\tau_4\,\tau_5\,\tau_5\,\tau_5 \end{aligned}$$

Approximation

We **truncate** the complemented permutations considering only the top-K rankings to obtain an approximate sparse representation:

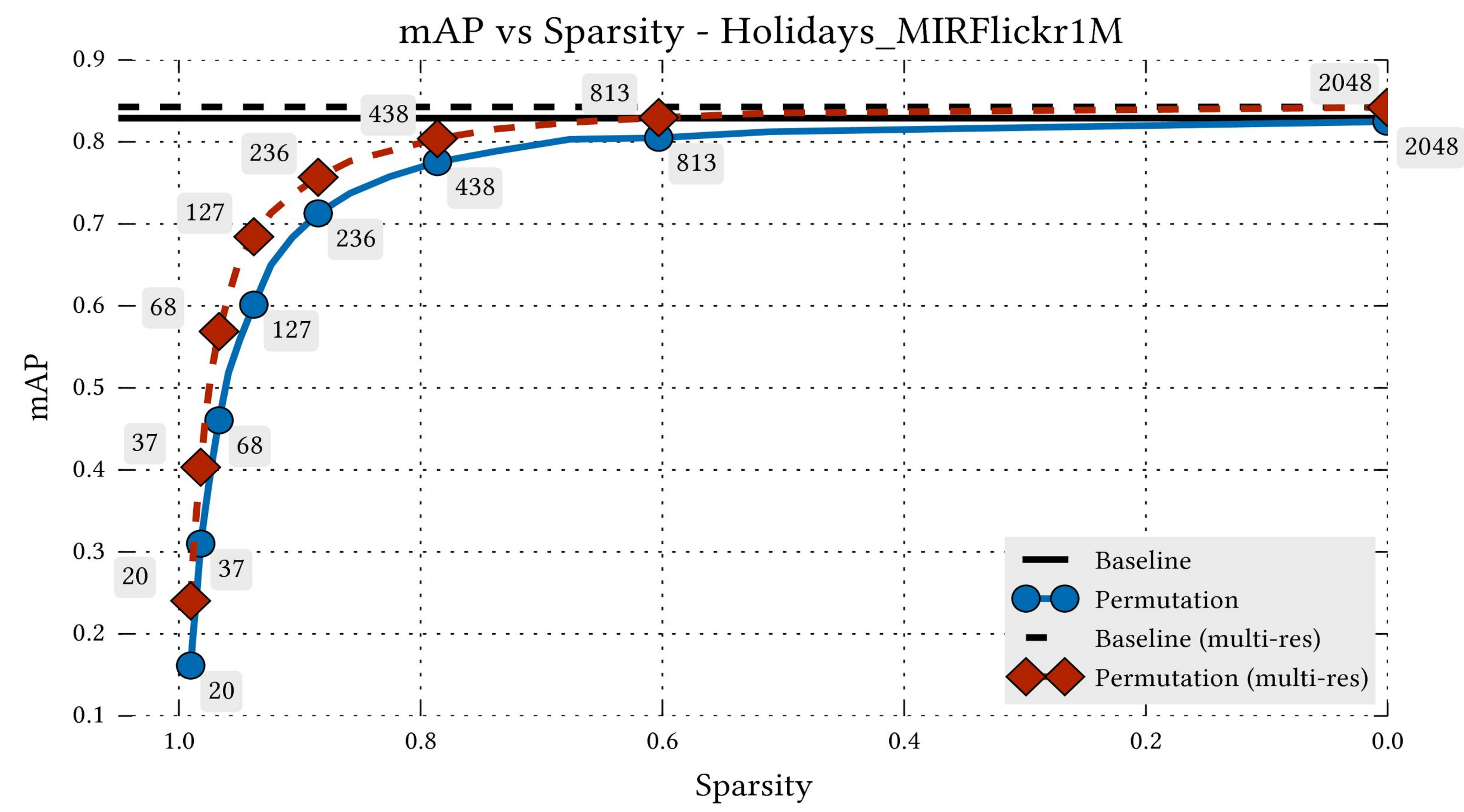
$$\tilde{\Pi}_x^{-1} = \max(\bar{\Pi}_x^{-1} - (K+1), 0)$$

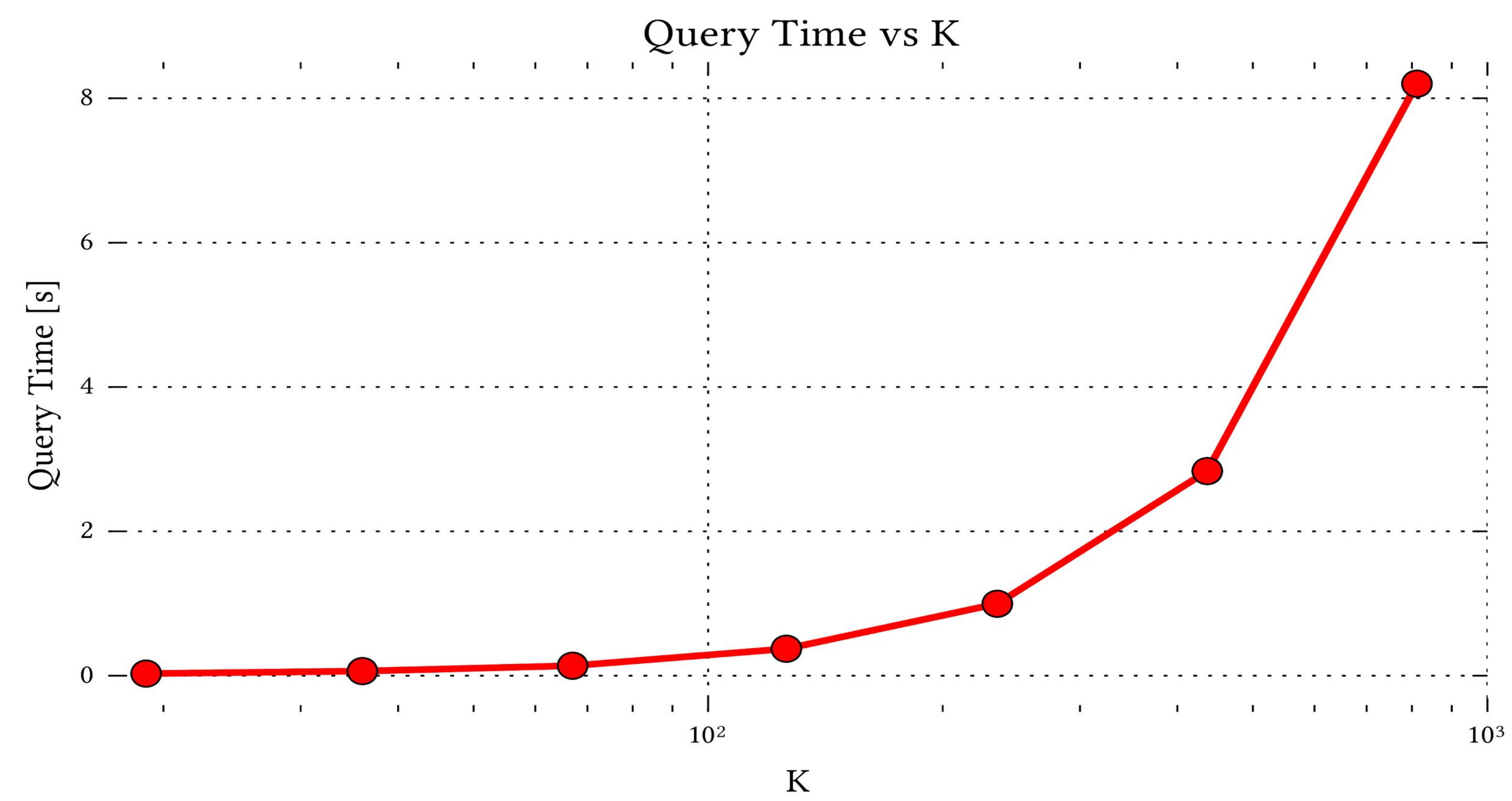
E.g:

$$\begin{aligned} \bar{\Pi}_x^{-1} &= [2, 5, 4, 1, 3] \\ \tilde{\Pi}_x^{-1} &= [0, 2, 1, 0, 0] \quad \text{with} \quad K = 2 \\ text &= \tau_2\,\tau_2\,\tau_3 \end{aligned}$$

The amount of sparsity is controlled by K:

$$\text{Sparsity} = 1 - \frac{K}{2048}$$





Example (D=5, K=2)

	q	a	b
x	$[-0.4, 0.2, 0.7, -0.1, 0.5]$	$[0.1, 0.3, 0.4, -0.15, 0.2]$	$[0.0, -0.8, 0.7, 0.9, 1.2]$
Π_x	$[3,\ 5,\ 2,\ 4,\ 1]$	$[3,\ 2,\ 5,\ 1,\ 4]$	$[3,\ 5,\ 2,\ 4,\ 1]$
Π_x^{-1}	$[5,\ 3,\ 1,\ 4,\ 2]$	$[4,\ 2,\ 1,\ 5,\ 3]$	$[4,\ 5,\ 3,\ 2,\ 1]$
$\bar{\Pi}_x^{-1}$	$[1,\ 3,\ 5,\ 2,\ 4]$	$[2,\ 4,\ 5,\ 1,\ 3]$	$[2,\ 1,\ 3,\ 4,\ 5]$
$\tilde{\Pi}_x^{-1}$	$[0,\ 0,\ 2,\ 0,\ 1]$	$[0,\ 1,\ 2,\ 0,\ 0]$	$[0,\ 0,\ 0,\ 1,\ 2]$
text$_x$	$\tau_3\ \tau_3\ \tau_5$	$\tau_2\ \tau_3\ \tau_3$	$\tau_4\ \tau_5\ \tau_5$

x	$\left\|\Pi_q^{-1} - \Pi_x^{-1}\right\|_2$	$\Pi_q^{-1} \cdot \Pi_x^{-1}$	$\bar{\Pi}_q^{-1} \cdot \bar{\Pi}_x^{-1}$	$\tilde{\Pi}_q^{-1} \cdot \tilde{\Pi}_x^{-1}$
a	2.0	53	53	4
b	3.74	48	48	2

1. Amato, Giuseppe, et al. "Deep Permutations: Deep Convolutional Neural Networks and Permutation-Based Indexing." International Conference on Similarity Search and Applications. Springer International Publishing, 2016.
2. Gennaro, Claudio, et al. "An approach to content-based image retrieval based on the Lucene search engine library." International Conference on Theory and Practice of Digital Libraries. Springer Berlin Heidelberg, 2010.
3. Gordo, Albert, et al. "Deep image retrieval: Learning global representations for image search." European Conference on Computer Vision. Springer International Publishing, 2016.