1 Retrieval Measurement of Model Performance

The main goal of evaluating the performance of a CBIR system is to measure the quality of the image retrieval for a given query. After the training process, Let, $\mathbf{b}_g^G$ is the binary hash code of an arbitrary image $\mathbf{x}_g^G \in \mathbf{X}_G$ and $|\mathbf{X}_G| = U_2$. $d_H(\mathbf{b}_g^G, \mathbf{b}_q^Q)$ is the HD between hash pair $\mathbf{u}_g$ and $\mathbf{b}_q^Q$, where $\mathbf{b}_q^Q$ is the hash code of a query image $\mathbf{x}_q^Q \in \mathbf{X}_Q$. Consider, the normalized hash codes $\mathbf{u}_g = \frac{\mathbf{b}_g^G}{\|\mathbf{b}_g^G\|_2}$, and $\mathbf{u}_q = \frac{\mathbf{b}_q^Q}{\|\mathbf{b}_q^Q\|_2}$. $(\mathbf{b}_g^G)^t$ denotes the transpose of $\mathbf{b}_g^G$. Let, The Euclidean distance between two binary hash codes $\mathbf{u}_g$ and $\mathbf{u}_q$ of length K is denoted by $\|\mathbf{u}_g - \mathbf{u}_q\|_2$ and is defined by,

$$\begin{aligned}
\|\mathbf{u}_g - \mathbf{u}_q\|_2^2 &= (\mathbf{u}_g - \mathbf{u}_q)^t(\mathbf{u}_g - \mathbf{u}_q) \\
&= \|\mathbf{u}_g\|_2^2 + \|\mathbf{u}_q\|_2^2 - 2(\mathbf{u}_g^t\mathbf{u}_q) \\
&= 1 + 1 - 2\|\mathbf{u}_g\|\|\mathbf{u}_q\| * cos(\mathbf{u}_g, \mathbf{u}_q) \\
&= 2(1 - cos(\mathbf{u}_g, \mathbf{u}_q)) \\
&= 2(1 - cos(\mathbf{b}_g^G, \mathbf{b}_q^Q)) \\
&= 2(1 - \frac{(\mathbf{b}_g^G)^t\mathbf{b}_q^Q}{\|\mathbf{b}_g^G\|\|\mathbf{b}_q^Q\|}) \\
&= \frac{2}{K}(K - (\mathbf{b}_g^G)^t\mathbf{b}_q^Q) \\
\Rightarrow \left\| \frac{\mathbf{b}_g^G}{\|\mathbf{b}_g^G\|} - \frac{\mathbf{b}_q^Q}{\|\mathbf{b}_q^Q\|} \right\|_2^2 = 2(1 - cos(\mathbf{b}_g^G, \mathbf{b}_q^Q)) &= \frac{2}{K}(1 - (\mathbf{b}_g^G)^t\mathbf{b}_q^Q))
\end{aligned} \tag{1}$$

Since, $\|\mathbf{b}_g^G\| = \|\mathbf{b}_q^Q\| = \sqrt{K}$

The relationship between cosine similarity and normalized Euclidean distance between two binary hash codes $\mathbf{b}_i$ and $\mathbf{b}_q^Q$ of length K can be expressed as follows:

$$d_H(\mathbf{b}_g^G, \mathbf{b}_q^Q) = \frac{K}{4}\left\| \frac{\mathbf{b}_g^G}{\|\mathbf{b}_g^G\|} - \frac{\mathbf{b}_q^Q}{\|\mathbf{b}_q^Q\|} \right\|_2^2 = \frac{K}{2}(1 - cos(\mathbf{b}_g^G, \mathbf{b}_q^Q)) = \frac{1}{2}(K - (\mathbf{b}_g^G)^t\mathbf{b}_q^Q) \tag{2}$$

The range of $d_H(\mathbf{b}_g^G, \mathbf{b}_q^Q)$ lies on $[0, K]$. Retrieval performance is evaluated using two metrics: mAP and nDCG (?). The quality of image retrieval is measured using mAP, while nDCG assesses the rank quality of the retrieved images across all query images. We discuss the computation process for both these metrics as below in the following subsections.

1.0.1 Mean average precision

$mAP@p$ represents the mAP for the top-p retrieved images from $\mathbf{X}_G$. It is calculated by finding the average precision ($AP_q@p$) for each $\mathbf{x}_q^Q \in \mathbf{X}_Q$ based on the top-p retrieved images from $\mathbf{X}_G$. Let, $\mathbf{x}_r^G$ is the r-th ranked image from top-p retrieve images. Then $AP_q@p$ is defined as,

$$AP_q@p = \frac{\sum_{r=1}^{p} P_q(r)R_q^{mAP}(r)}{\sum_{r=1}^{p} R_r^{mAP}(r)} \tag{3}$$

where, $P_q(r)$ denotes the precision for the top-r retrieval of query image $\mathbf{x}_q^Q$ and is defined by,

$$P_q(r) = \frac{\sum_{r=1}^{r} R_q^{mAP}(r)}{r} \tag{4}$$

The relevance score $R_q^{mAP}(r)$ for image $\mathbf{x}_q^Q$ is discussed later.

Finally,

$$mAP@p = \frac{1}{U_3} \sum_{\mathbf{x}_q^Q \in \mathbf{X}_Q} AP_q@p \tag{5}$$

The range of $mAP@p$ lies on $[0, 1]$.

1.0.2 Normalized discounted cumulative gain

In order to compute $nDCG@p$ for top-p retrieval, first we need to calculate $DCG@p$ for top-p retrieval. The mathematical formulation for $DCG_q@p$ of query image $\mathbf{x}_q^Q$ is given by,

$$DCG_q@p = \sum_{r=1}^{p} \frac{2^{R_q^{nDCG}(r)} - 1}{log_2(r+1)} \tag{6}$$

We normalize this by dividing it with the maximally achievable value or Ideal DCG (iDCG). Finally to obtain,

$$nDCG_q@p = \frac{DCG_q@p}{iDCG_q@p} \tag{7}$$

where $iDCG_q@p = DCG_q@p$ of ideal ranking or best possible ranking. The relevance score $R_q^{nDCG}(r)$ for image $\mathbf{x}_q^Q$ is discussed later.

Finally,

$$nDCG@p = \frac{1}{U_3} \sum_{\mathbf{x}_q^Q \in \mathbf{X}_Q} nDCG_q@p \tag{8}$$

$iDCG@p$ is 0 indicates that the absence of similar images to the query in the retrieval list. $nDCG@p$ is applicable only when there is at least one similar image to the query image in the retrieval list. The range of $nDCG@p$ lies on $[0, 1]$.