Supplementary Materials of Image Retrieval with Well-Separated Semantic Hash Centers

Liangdao Wang[1,2], Yan Pan[1⋆], Hanjiang Lai[1], Jian Yin[1]

[1] School of Computer Science and Engineering, Sun Yat-sen University
[2] wangld5@mail2.sysu.edu.cn

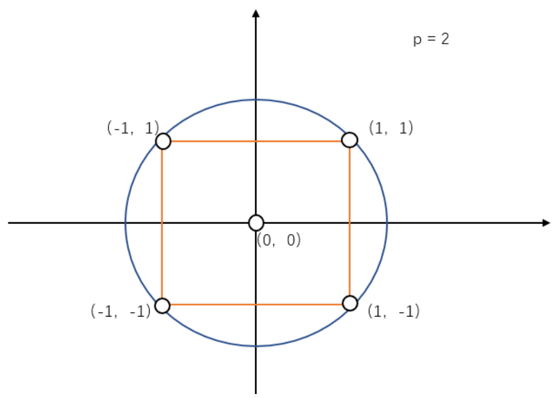

Fig. 1. Illustrative example of Eq.1 when $p = 2$ and $q = 2$

For ease of understanding, we show an example of the following equation with $p = 2$ and $q = 2$ in Fig.1 and then will prove it in terms of both sufficiency and necessity:

$$z \in \{-1, 1\}^q \Leftrightarrow z \in [-1, 1]^q \bigcap \{z : ||z||_p^p = q\} \tag{1}$$

where $p \in (0, +\infty)$ and q is the length of hash code

Proof. We first proof the sufficiency of Eq.1:

$$z \in \{-1, 1\}^q \Rightarrow z \in [-1, 1]^q \bigcap \{z : ||z||_p^p = q\}$$

As $\{-1, 1\}^q \subset [-1, 1]^q$, given $z \in \{-1, 1\}^q$, $z \in [-1, 1]^q$ must hold. As $\{-1, 1\}^q \subset \{z : ||z||_p^p = q\}$, given $z \in \{-1, 1\}^q$, $z \in \{z : ||z||_p^p = q\}$ must hold. Combining this two points, we obtain that given $z \in \{-1, 1\}^q$, $z \in [-1, 1]^q \bigcap \{z : ||z||_p^p = q\}$ must hold.

Then we proof the necessity of Eq.1:

$$z \in [-1, 1]^q \bigcap \{z : ||z||_p^p = q\} \Rightarrow z \in \{-1, 1\}^q$$

⋆ Corresponding Author

As $z \in [-1,1]^q$, then $\forall i$, $|z_i| \leq 1$, and the equation holds iff $z_i \in \{-1,1\}$. As $p \in (0,+\infty)$ and $|z_i| \leq 1$, we have $|z_i|^p \leq 1$, and the equation holds iff $z_i \in \{-1,1\}$. Then we obtain $||z||_p^p = \sum_i^q |z_i|^p \leq q$, and the equation holds iff $\forall i, z_i \in \{-1,1\}$. Thus we obtain that if $z \in [-1,1]^q \bigcap \{z : ||z||_p^p = q\}$, then $z \in \{-1,1\}^q$ must hold.

Next, to project the range of $w_i^T w_j$ from $[min(w_i^T w_{\sim i}), max(w_i^T w_{\sim i})]$ to $[-q,q]$, the following equation will be used in our work and proved here:

$$w_i^T w_j = -q + 2q\frac{w_i^T w_j - min(w_i^T w_{\sim i})}{max(w_i^T w_{\sim i}) - min(w_i^T w_{\sim i})} \tag{2}$$

where q is the length of hash code, w is category feature extracted from ResNet-50 model, and $w_{\sim i} = [w_1, ..., w_{i-1}, w_{i+1}, ..., w_m]$ represents the matrix that consists of $w_j (1 \leq j \leq m, j \neq i)$.

Proof. Firstly, we need to claim that to transform the variable $x \in [a,b]$ to variable $y \in [A,B]$, the linear extension equation is:

$$y = \frac{B-A}{b-a}(x-a) + A \tag{3}$$

therefore, to project $w_i^T w_j \in [min(w_i^T w_{\sim i}), max(w_i^T w_{\sim i})]$ to $w_i^T w_j \in [-q,q]$:

$$\begin{aligned} w_i^T w_j &= \frac{q-(-q)}{max(w_i^T w_{\sim i}) - min(w_i^T w_{\sim i})}(w_i^T w_j - min(w_i^T w_{\sim i})) + (-q) \\ &= -q + \frac{2q}{max(w_i^T w_{\sim i}) - min(w_i^T w_{\sim i})}(w_i^T w_j - min(w_i^T w_{\sim i})) \\ &= -q + 2q\frac{w_i^T w_j - min(w_i^T w_{\sim i})}{max(w_i^T w_{\sim i}) - min(w_i^T w_{\sim i})} \end{aligned} \tag{4}$$

Table 1. Comparison in mAP using category features generated from three pre-trained models

Method	ImageNet (mAP@1000)			Stanford Cars (mAP@All)			NABirds (mAP@All)		
	16 bits	32 bits	64 bits	16 bits	32 bits	64 bits	16 bits	32 bits	64 bits
With AlexNet	0.8511	0.8761	0.8982	0.7998	0.8483	0.8705	0.6489	0.7194	0.7533
With VGG16	0.8555	0.8806	0.8982	0.8111	0.8521	0.8745	0.6578	0.7250	0.7588
With ResNet50	**0.8616**	**0.8851**	**0.8982**	**0.8218**	**0.8569**	**0.8771**	**0.6693**	**0.7381**	**0.7599**

To explore the effects of hash centers generated using semantic information of different pre-trained models, we use the category features from VGG16, AlexNet and ResNet respectively, to generate hash centers, and compare their retrieval results. As shown in Tab.1, the method with hash centers generated by ResNet-50 performs better than other methods, which indicates that a better pre-trained model can obtain more accurate semantic information, and it is beneficial to improve the final retrieval performance.