**A²-NET: Learning Attribute-Aware Hash Codes for Large-Scale Fine-Grained Image Retrieval**

(Supplementary Materials)

Xiu-Shen Wei¹,², Yang Shen¹, Xuhao Sun¹, Han-Jia Ye², Jian Yang¹

¹Nanjing University of Science and Technology
²State Key Lab. for Novel Software Technology, Nanjing University

In the supplementary materials, we present further information about the proposed A²-NET model, including: 1) Additional experimental results of other comparison methods, especially DSaH [4]; 2) More examples of retrieved results on other fine-grained benchmark datasets.

1 Additional experimental results of other comparison methods

Apart from ExchNet, DSaH [4] is another fine-grained hashing method which has achieved good retrieval accuracy. For fair comparisons, we strictly control empirical settings as the same as those of [4] and compare the results of our A²-NET with its results and three following methods, i.e., DPSH [6], DTQ [7] and HBMP [2]. Specifically, we follow the settings of DSaH [4] and conduct experiments on two fine-grained datasets, i.e., *Stanford Dogs* [5] and *CUB200-2011* [10]. In concretely, *Stanford Dogs* consists of 20,580 images in 120 classes while each class contains about 150 images. The dataset is divided into the train set (100 images per class) and the test set (totally 8,580 images for all categories). *CUB200-2011* contains 11,788 bird images from 200 bird species and is officially split into 5,994 images for training and 5,794 images for test. We use AlexNet as backbone and it is not fine-tuned on each dataset.

As shown in Table 1, our A²-NET significantly outperforms the other baseline methods on these two datasets by following the same settings of [4]. In particular, compared with DSaH [4], our A²-NET achieves 10% and 7% improvements on *Stanford Dogs* and *CUB200-2011* in average.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Stanford Dogs</th>
<th>CUB200-2011</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>12 bits</td>
<td>24 bits</td>
</tr>
<tr>
<td>DPSH [6]</td>
<td>17.7</td>
<td>22.1</td>
</tr>
<tr>
<td>DTQ [7]</td>
<td>18.5</td>
<td>18.7</td>
</tr>
<tr>
<td>HBMP [2]</td>
<td>19.0</td>
<td>23.8</td>
</tr>
<tr>
<td>DSaH [4]</td>
<td>24.4</td>
<td>28.7</td>
</tr>
<tr>
<td><strong>Ours</strong></td>
<td><strong>36.6</strong></td>
<td><strong>44.8</strong></td>
</tr>
</tbody>
</table>

2 More examples of retrieved results on other fine-grained datasets

We present more retrieval results on *Aircraft* [8], *Food101* [1], *NABirds* [9] and *VegFru* [3]. As shown in the following figures, our proposed A²-NET can retrieve well among multiple subordinate categories. There also exist several failure cases, where quite tiny differences (e.g., caused by different views) between the query image and the returned images are demanded by carefully observations.
Figure 1: Examples of top-10 retrieved images on Aircraft of 48-bit hash codes by our A²-NET.

Figure 2: Examples of top-10 retrieved images on Food101 of 48-bit hash codes by our A²-NET.

Figure 3: Examples of top-10 retrieved images on NABirds of 48-bit hash codes by our A²-NET.

Figure 4: Examples of top-10 retrieved images on VegFru of 48-bit hash codes by our A²-NET.
References


