ICIP 2017 IEEE International Conference on Image Processing
September 17-20, 2017, Beijing, China

Part Based Fine-grained Bird Image Retrieval Respecting Species Correlation

Cheng Pang[1,2], Hongdong Li[2,3], Anoop Cherian[3], Hongxun Yao[1]
[1]School of Computer Science & Technology, Harbin Institute of Technology [2]College of Engineering & Computer Science, Australian National University
[3]Australian Centre for Robotic Vision, Australian National University

Motivations

Fine-grained objects usually present some fixed parts which can be used to distinguish between different species. Most fine-grained categorization and retrieval methods build their algorithms on this observation. However, the visually similarity of a part carries some correlations between different species. These correlations has helped biologists exploring the evolution of a part and cross-species behavioral similarities. As shown in this Figure, birds from different species with similar tails share interesting correlations: two of them are gulls and they are both sea birds.

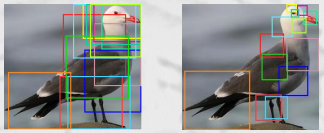
Based on the above observations, we present a novel task to search for instances with visually similar parts but from different species.

Geometry-constrained Part Pooling

In order to generate features for every part, we design a geometry-constrained part pooling layer. Not like some existing networks which take part features directly from their layers, our pooling layer take part from original images. And the taken parts have reasonable size of the receptive field (RF) which can be adjusted automatically and ensure there is less overlaps between different parts. As a comparison, a receptive field of a part in conv5 layer of Alexnet has a size of 163. Compared to its input size 224, the receptive field is too large and may lead to many overlaps, which is not suitable for our task.

In our method, the RF of each part is adjusted automatically. We first compute the distances between every part and find the k-nearest neighbor parts for every part. That is, if D denotes the average distance to all parts and its top-k neighbours, and if $\overline{D}$ denotes the average distance to all parts, then we scale the default receptive field size S_d as:

$$S=\left(1+\frac{D-\overline{D}}{\overline{D}}\right)\times S_d$$



RF of parts in some existing networks

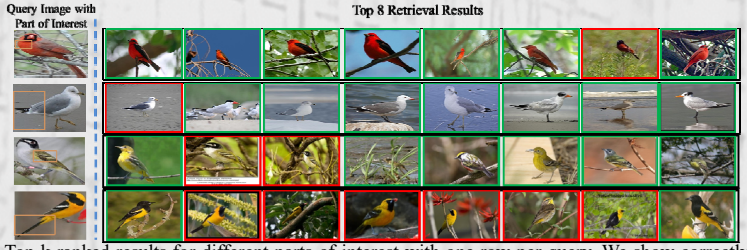
RF of parts in the proposed part pooling layer

Experiments

Dataset, benchmarks and settings

We test our method on CUB200-2011 bird dataset which is a challenging fine-grained dataset. To evaluate the performance of our scheme, we need to define a list of classes for each part based on the similarity between the corresponding parts. In this paper, we show evaluations on three parts: chest, tail, and wing. We manually generate 3 such class lists for these parts. We use the standard cumulative match characteristic (CMC) and Recall@K for evaluating the performance.

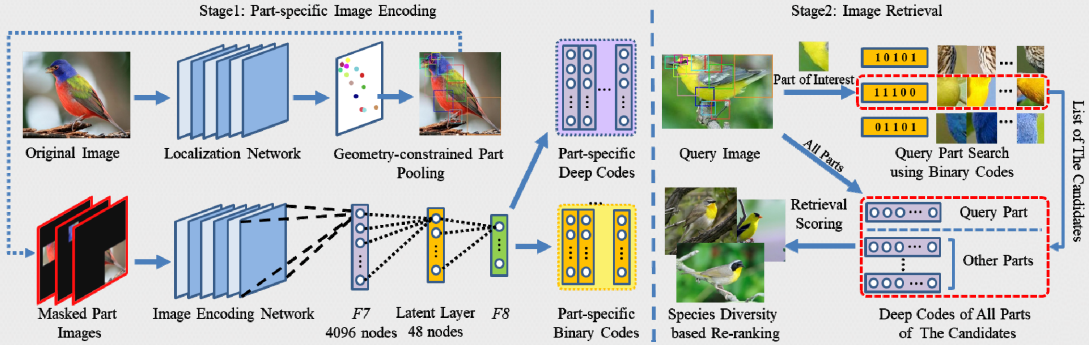
Qualitative Analysis



Top k ranked results for different parts-of-interest with one row per query. We show correctly retrieved instances in green and incorrect ones in red.

The Proposed Pipeline

The proposed method takes birds as examples, but not limited in birds. For example, one likes the collar and sleeves of a coat but not the other parts, which can be quite a common situation of online-shopping. Out pipeline consists of 2 parts: Part-specific image encoding and retrieval. Our pipeline consists of 2 steps: Part-specific image encoding and retrieval.

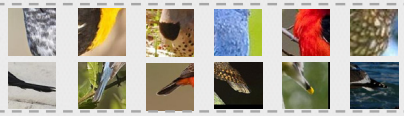


We use 2 networks to encode an image. A localization network trained using the bird dataset gives part localizations of query images. Then a novel geometry-constrained part pooling layer extract masked part images and feed them into the Image encoding network. Notice that the encoding network has a binary layer, generating a kind of binary codes together with deep features for all parts. This benefits the final retrieval which we will detailed later.

Part-based Retrieval

Give both the binary codes and deep codes generated by the encoding network, the proposed retrieval strategy can be done in 2 steps: First, given the part of interest which should be visually similar to the part of query image, we look for candidates with k nearest binary codes of the part of interest. Then we apply precise retrieval using deep codes of all parts of these candidates. The 2 step strategy balances the retrieval accuracy and computational costs.

Illustration of some parts
Row 1 – chest
Row 2 – tail



Our ranking function is defined as follows. Suppose $\chi = \{x\hat{}1, x\hat{}2, \cdots, x\hat{}m\}$, are deep codes for m parts and let "χ" $_q$ is such a set for the query image. Further let, $x\_q\hat{}t \in$ "χ" $_q$ is the deep code for part t which we use for the query. Then, the ranking function computes:

$$F\left(x_q^t, \chi_q, \chi\right) = \log P\left(x_q^t, x^t\right) + \frac{\omega}{m-1}\sum_{\substack{s=1 \\ s\neq t}}^{m} \log\left(1 - P\left(x_q^s, x^s\right)\right)$$

where $P\left(x_i^s, x_j^s\right) = \frac{1}{1+dist\left(x_i^s, x_j^s\right)}$.

Quantitative Analysis

The CMC curves: results using deep codes with novel species diversity ranking function, wins by a large margin over binary codes.

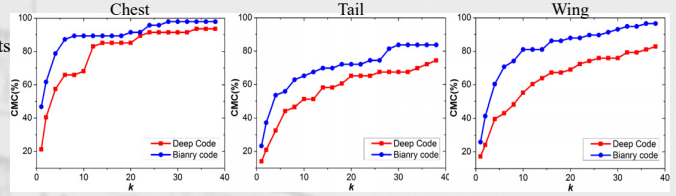


Table 1. Recall@K Scores on CUB200-2011.

part	Recall(%)@K	1	10	40	60	80	100	120	140	160	180	200	220	240	260	280	300
Chest	Retrieval w/o part pooling layer and with part diversity	0	0.2	4.8	8.4	12.8	16.3	21.2	26.0	28.2	30.4	32.6	36.6	39.2	40.5	43.2	46.7
	Retrieval with part pooling layer and w/o part diversity	0	1.3	4.9	10.6	13.2	18.1	21.2	26.0	29.5	32.6	35.7	40.5	43.6	47.1	49.3	51.1
	Retrieval with part pooling layer and with part diversity	0.4	3.1	7.5	12.3	16.7	20.7	25.1	29.1	32.6	34.8	37.4	42.7	46.3	49.3	52.0	53.7
Tail	Retrieval w/o part pooling layer and with part diversity	0	0.8	1.3	2.8	4.4	5.2	7.0	8.0	8.8	10.1	10.8	12.1	12.1	12.4	12.4	12.6
	Retrieval with part pooling layer and w/o part diversity	0	0.9	6.2	10.6	11.5	12.4	15.9	17.7	18.6	21.2	21.2	22.1	23.9	24.8	24.8	25.7
	Retrieval with part pooling layer and with part diversity	0	0.9	7.1	11.5	11.5	15.0	16.8	19.5	19.5	21.2	22.1	22.1	23.9	24.8	25.7	26.6
Wing	Retrieval w/o part pooling layer and with part diversity	0	1.8	5.5	10.9	15.5	15.5	16.4	17.3	17.3	17.3	21.8	25.5	27.3	29.1	29.1	30.9
	Retrieval with part pooling layer and w/o part diversity	0.9	4.6	10.9	14.5	18.2	20.9	20.9	21.8	24.6	25.5	26.4	26.4	27.3	33.6	33.6	33.6
	Retrieval with part pooling layer and with part diversity	0.9	5.5	10.9	16.4	18.2	22.7	24.6	24.6	26.4	26.4	27.3	27.3	28.2	33.6	33.6	33.6

In Table 1, we show results evaluating the advantages of our part-pooling layer in combination with the species diversity ranking function.

Conclusion

We proposed a novel fine-grained bird image retrieval task that searches for instances having similar body parts, but from different species. A novel two step strategy is designed to solve this task consisting of first generating a list of candidate retrievals using a network with a novel part-pooling layers, and then re-ranking these candidates using a function that promotes species diversity. Experiments have shown the effectiveness of the proposed method. Our method successfully discovered interesting correlations among distinct species. Although, in this paper we focused on bird images only, it can be straightforwardly extended to other fine-grained retrieval applications.