Class-Weighted Convolutional Features for Image Retrieval

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

DATA 61 CSIRO

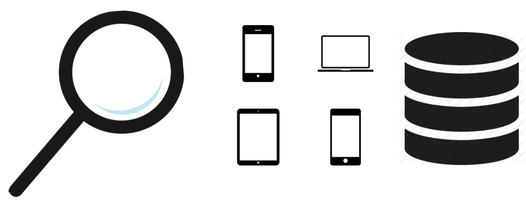

Albert Jiménez


José M. Álvarez

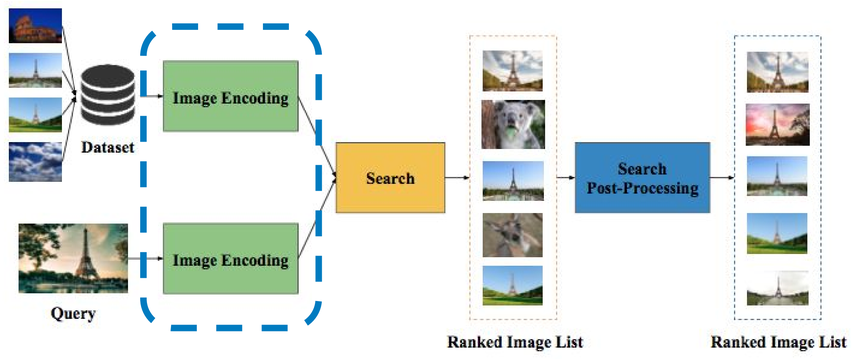

Xavier Giró-i-Nieto

Motivation

Search by Visual Similarity on Large Scale Databases.



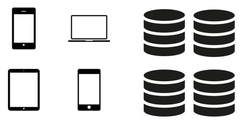
General Pipeline



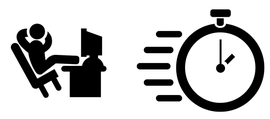
Desired Properties for Image Encoding

Invariant to scale, illumination & translation

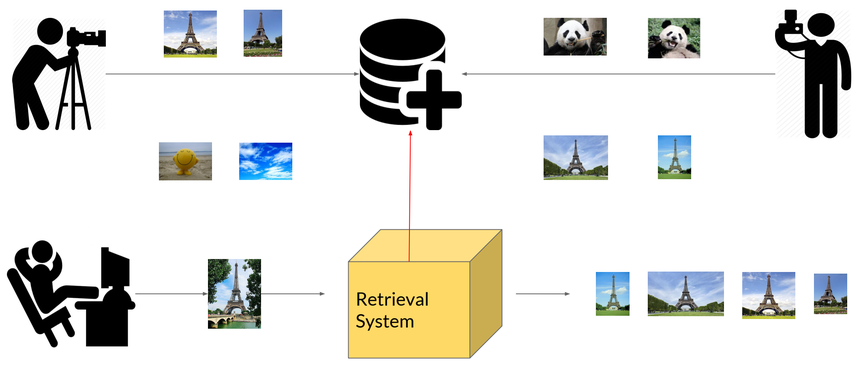
Memory Efficient



Be able to provide a **fast search**

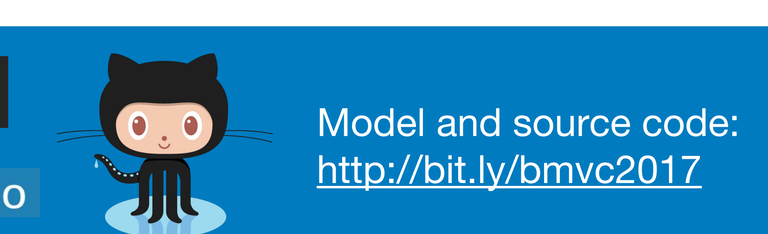


Challenge in Retrieval: Dynamic Datasets of Unlabeled Images



From Classification to Retrieval

Class Activation Maps (CAMs) [1] weight the Convolutional Features according to the predicted semantic classes.



$$CAM_i = w_{1,i} * conv_1 + w_{2,i} * conv_2 + ... + w_{K,i} * conv_K$$

[1] B. Zhou, A. Khosla, Lapedriza. A., A. Oliva, and A. Torralba. Learning Deep Features for Discriminative Localization. CVPR (2016).

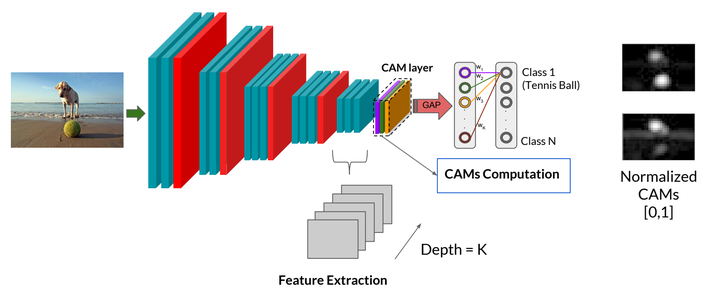
We have observed an important variations in the CAMs generated by different CNN architectures.

Top-1 predicted class

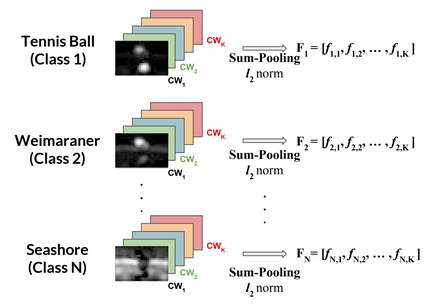
Top-1, Top-2 & Top-3 predicted classes

Proposed Image Encoding

1) Extraction of Convolutional Features & CAMs



2) Feature Weighting and Pooling

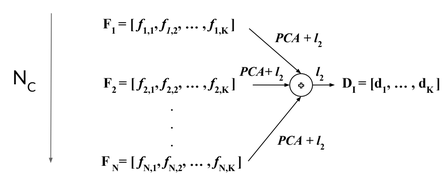


One vector per class:

$$F^c = [f_1^c, f_2^c, ..., f_K^c]$$

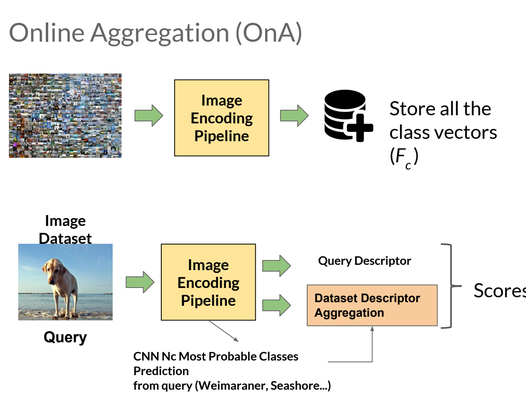
$$f_k^{(c)} = CW_k \sum_{i=1}^{W} \sum_{j=1}^{H} \chi_{i,j}^{(k)} \; CAM_{i,j}^{(c)}$$

3) Feature Aggregation

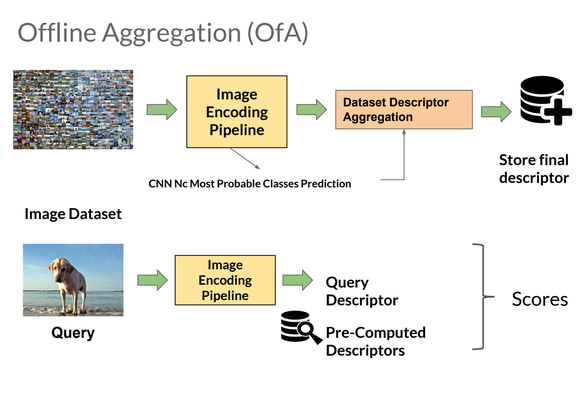


- ➔ Which classes do we aggregate?
- ➔ What is the optimal number of classes (N_C) ?
- ➔ What is the optimal number of classes (N_{PCA}) to compute the PCA ?

Online Aggregation (OnA)



Offline Aggregation (OfA)



4) Region Proposals for Reranking (R) and Query Expansion (QE)

Bounding boxes are fit to the largest connected components of the CAMs after a binarization at different thresholds.

Green - Ground Truth
Orange → Red - Different Thresholds

Experiments

Baseline Results

Descriptor Aggregation	Oxford5k
Raw Features	0.396
Raw + Crow (channel)	0.420
Raw Features + PCA	0.589
Raw + Crow(channel) + PCA	0.607

Computational Burden

Descriptor Aggregation	Time (s)	mAP
Raw + PCA	0.49	0.589
1 CAM	0.5	0.667
8 CAMs	0.6	0.709
32 CAMs	0.9	0.711
64 CAMs	1.5	0.712

Networks Comparison

Network	Oxf5k	Paris6k
VGG-16 (Raw)	0.396	0.526
VGG-16 (64CAMs)	0.712	0.805
Resnet-50 (Raw)	0.389	0.508
Resnet-50 (64CAMs)	0.699	0.804
Densenet-161 (Raw)	0.339	0.495
Densenet-161 (64CAMs)	0.695	0.799

Comparison with the State-of-the-Art for off-the-shelf features

Method	Dim	Oxf5k	Par6k	Oxf105k	Par106k
SPoC[3]	256	0.531	-	0.501	-
uCroW[14]	256	0.666	0.767	0.629	0,695
CroW[14]	512	0.682	0.796	0.632	0.710
R-MAC[31]	512	0.669	0.830	0.616	**0.757**
BoW[16]	25k	0.738	0.820	0.593	0.648
Razavian [22]	32k	**0.843**	0.853	-	-
Ours(OnA)	512	0.736	**0.855**	-	-
Ours(OfA)	512	0.712	0.805	**0.672**	0.733

$N_c = 64$, $N_{pca} = 1$

Method	Dim	R	QE	Oxf5k	Par6k	Oxf105k	Par106k
CroW	512	-	10	0.722	0.855	0.678	0.797
Ours(OnA)	512	-	10	0.760	0.873	-	-
Ours(OfA)	512	-	10	0.730	0.836	0.712	0.791
BoW	25k	100	10	0.788	0.848	0.651	0.641
Ours(OnA)	512	100	10	0.780	0.874	-	-
Ours(OfA)	512	100	10	0.773	0.838	0.750	0.780
RMAC	512	1000	5	0.770	**0.877**	0.726	**0.817**
Ours(OnA)	512	1000	5	**0.811**	0.874	-	-
Ours(OfA)	512	1000	5	0.801	0.855	**0.769**	0.800

$N_c = 64$, $N_{pca} = 1$, $N_{re\text{-}ranking} = 6$

PYTORCH

K theano

Model and source code: http://bit.ly/bmvc2017

Generalitat de Catalunya

GOBIERNO DE ESPAÑA MINISTERIO DE ECONOMÍA, INDUSTRIA Y COMPETITIVIDAD AGENCIA ESTATAL DE INVESTIGACIÓN

NVIDIA