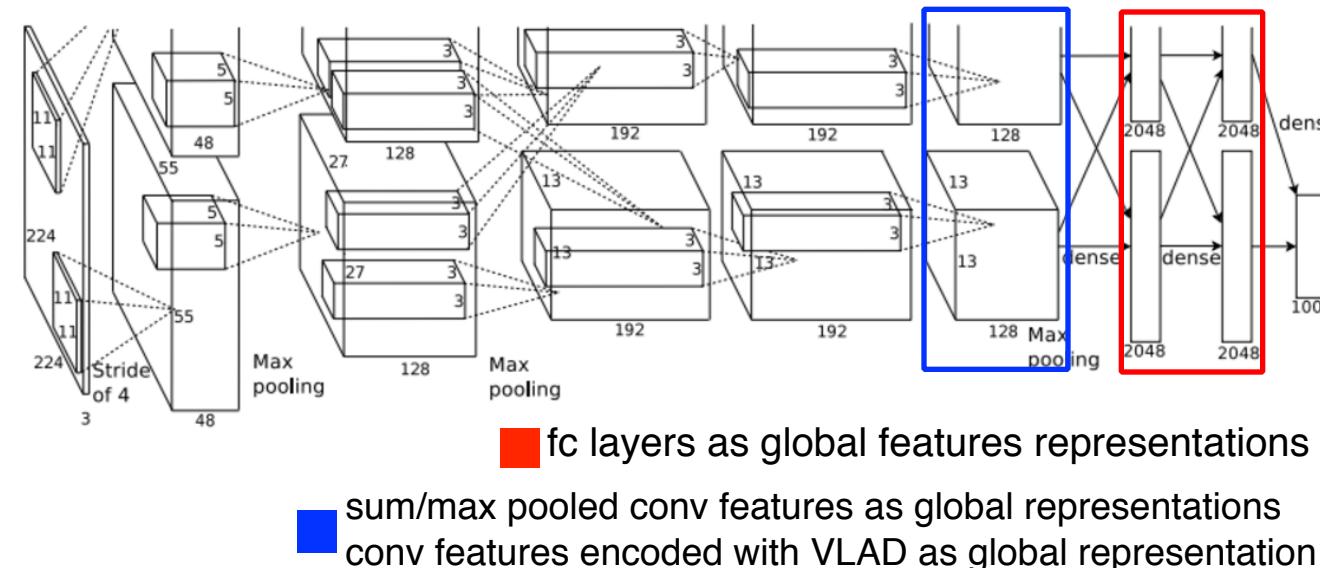


Motivation

CNN Features for Instance Search



Compact & Dense
(e.g. sum/max pooling conv feats, FC feats) → Capacity?

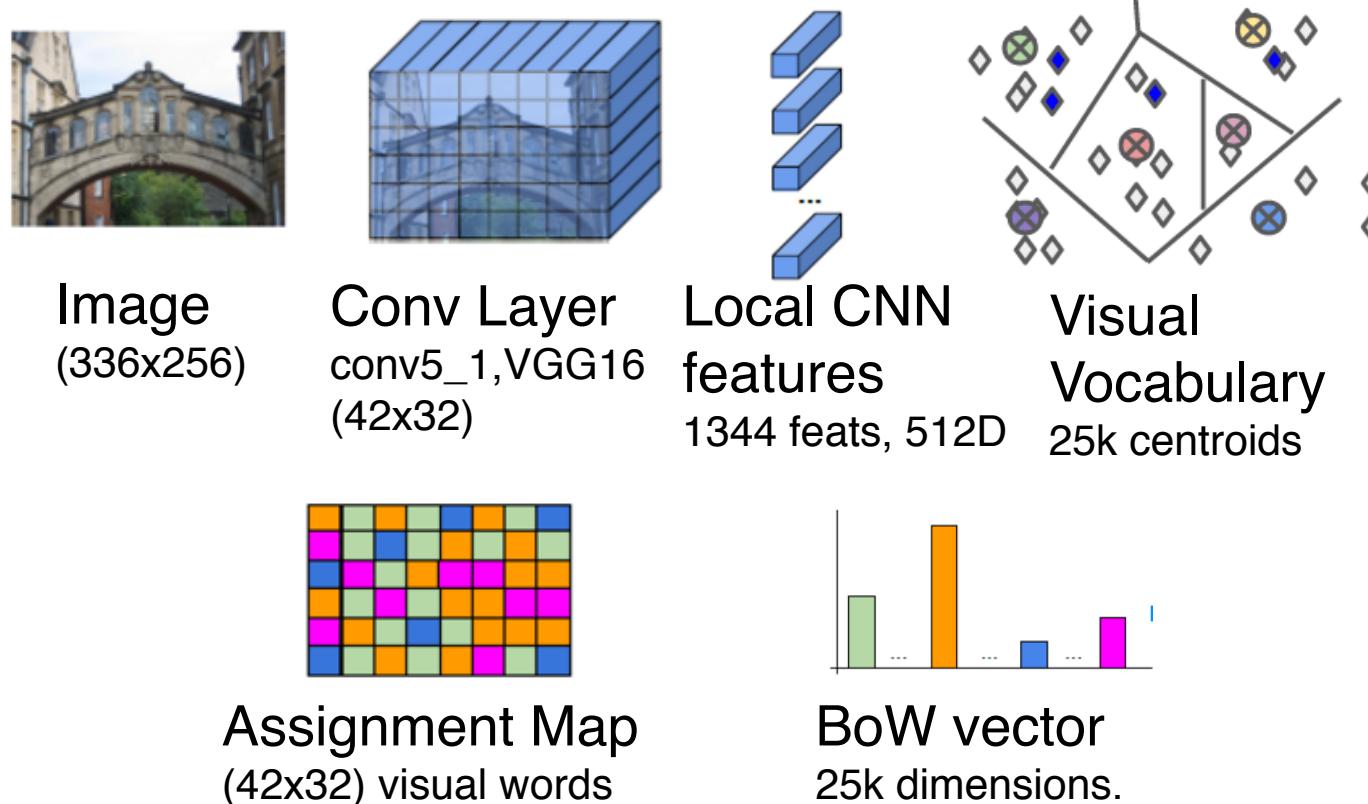
High-dimensional & Dense
(e.g. VLAD encoding) → Scalability?

High-dimensional & Sparse Bag of Visual Words

Contributions

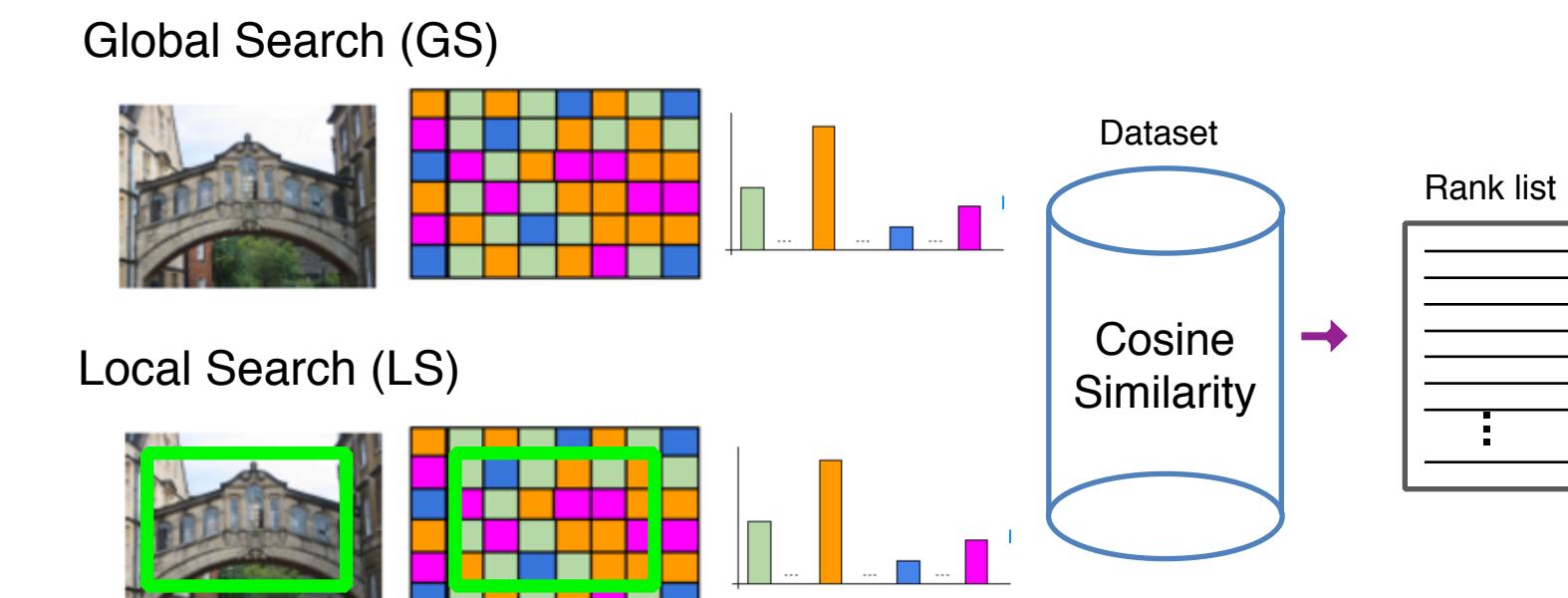
- Sparse visual representation based on a **Bags of Convolutional Features**, which allows fast retrieval by means of an inverted index.
- Assignment map** as a new compact representation of the image, allowing fast composition of BoW descriptor for any region of the image.
- Local analysis** of multiple image regions for reranking followed by query expansion using the obtained object locations.

Bag of Words framework

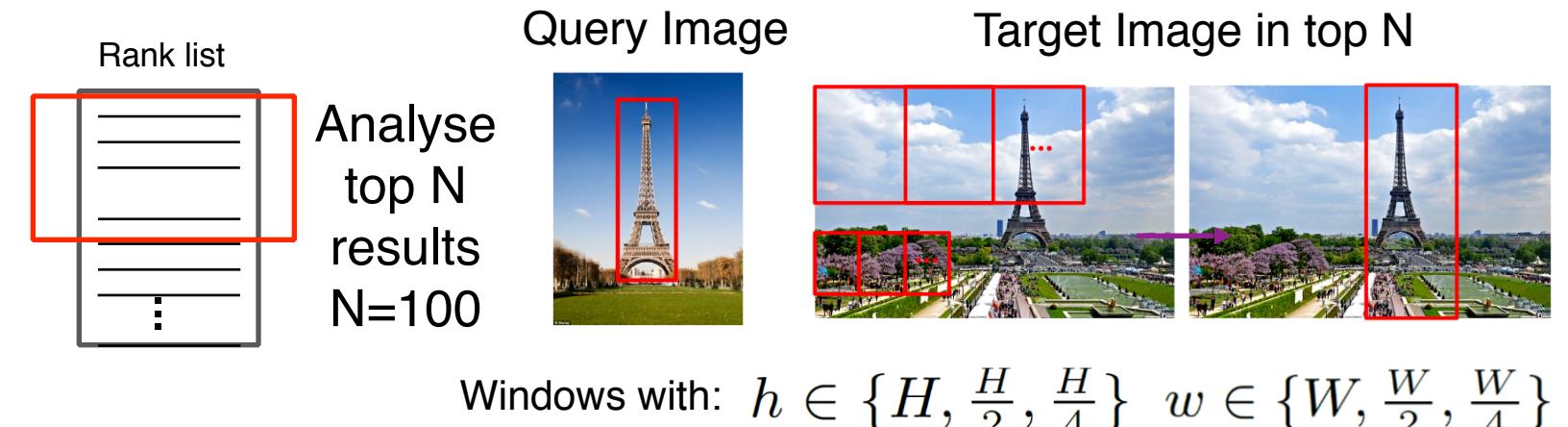


Instance Retrieval

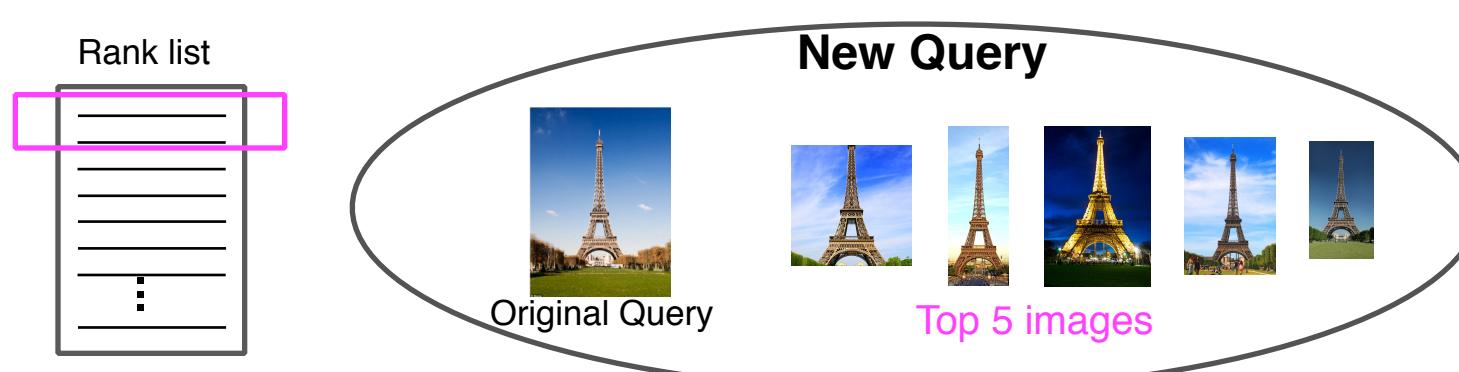
Query Representation



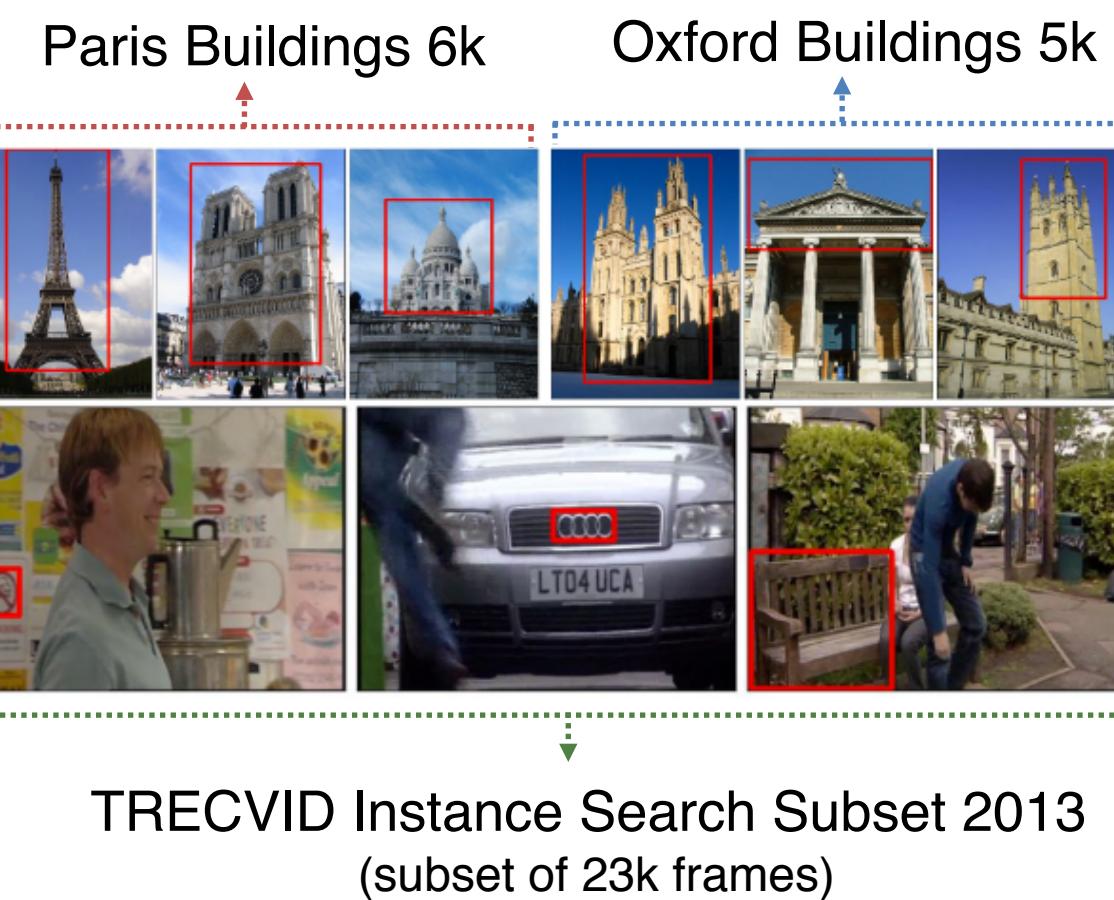
Spatial Reranking



Query Expansion



Datasets



Experiments

Comparison with state-of-the-art

	Oxford		Paris	
	5k	105k	6k	106k
Ng et al. [10]	0.649	-	0.694	-
Razavian et al. [16]	0.844	-	0.853	-
SPoC [2]	0.657	0.642	-	-
R-MAC [20]	0.668	0.616	0.830	0.757
CroW [7]	0.682	0.632	0.796	0.710
uCroW [7]	0.666	0.629	0.767	0.695
GS	0.652	0.510	0.698	0.421
LS	0.739	0.593	0.820	0.648
CroW + GQE [7]	0.722	0.678	0.855	0.797
R-MAC + R + GQE [20]	0.770	0.726	0.877	0.817
LS + GQE	0.773	0.602	0.814	0.632
LS + R + LQE	0.788	0.651	0.848	0.641

Results on TRECVID subset

	Oxford 5k	Paris 6k	INS 23k
BoW	GS	0.650	0.698
	LS	0.739	0.819
Sum pooling (as ours)	GS	0.606	0.712
	LS	0.583	0.742
Sum pooling (as in [7])	GS	0.672	0.774
	LS	0.683	0.763
			0.139
			0.120



Conclusions

- High-dimensional sparse representation suitable for fast retrieval
- Competitive results in two image retrieval benchmarks
- Well suited and more robust for scenarios where only small number of features in the target images are relevant to the query.