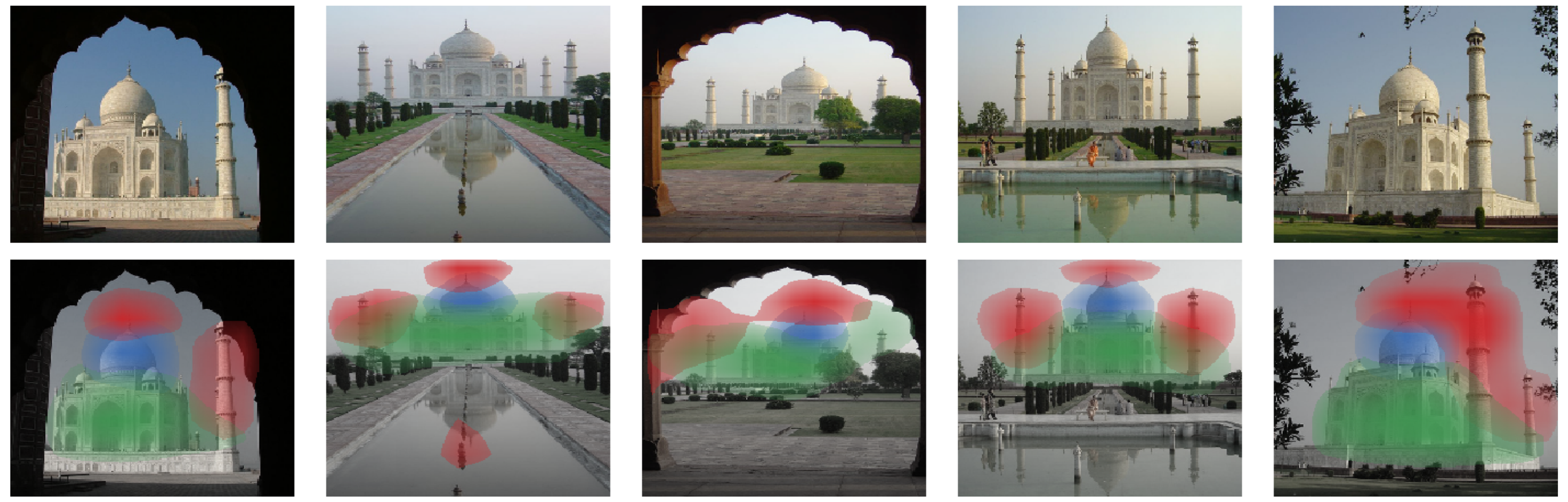
DEEP FEATURE FACTORIZATION FOR CONTENT-BASED IMAGE RETRIEVAL AND LOCALIZATION

Edo Collins, Sabine Süsstrunk, IVRL IC EPFL

DEEP FEATURE FACTORIZATION

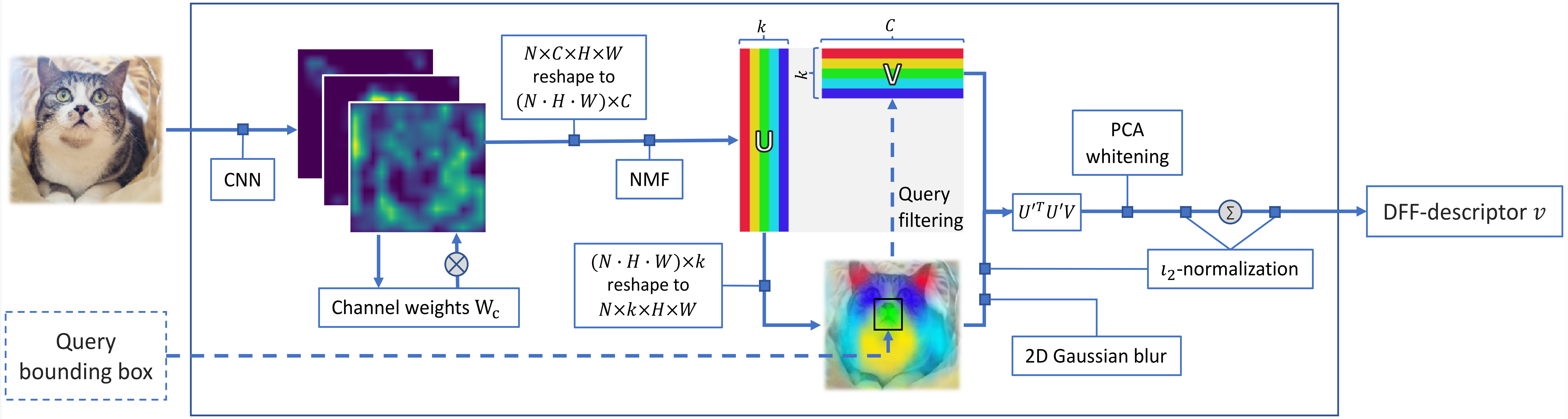
Activations of a neural network layer can be factorized $\boldsymbol{A} \approx \boldsymbol{UV}$ under a rank constraint k. The factors $\boldsymbol{U}$ and $\boldsymbol{V}$ decompose the image set as "what" ($\mathbf{V}$) in the image is located "where" ($\boldsymbol{U}$). For ReLU networks, which are non-negative, a natural factorization alg. is **non-negative matrix factorization** (NMF).

Each column of $\boldsymbol{U}$ induces a set of heatmaps, which form the basis of our localization technique. Taj Mahal example: shown below are $k = 3$ heatmaps, each with a different color.



The rows of $\boldsymbol{V}$ capture the most salient features in the image and form the basis of our descriptor for image retrieval. Taj Mahal example: $\boldsymbol{V}$ can be interpreted using gradient ascent visualization, as shown below.

IMAGE DESCRIPTOR



RESULTS

We evaluate retrieval on the Oxford- and Paris Buildings datasets (bottom).

After retrieval, we apply DFF to the query and top scoring images, and infer a bounding box for each non-query image (top).

Localization IoU	Oxford5k	Paris6k
AML (Tolias 2015)	51.3	51.4
Exhaustive R-MAC (Tolias 2015)	52.6	52.9
DFF VGG-16	49.3	67.6
DFF ResNet-50	**53.2**	**68.7**

Retrieval mAP	Oxford5k				Paris6k			
	VGG		ResNet		VGG		ResNet	
	-	QE	-	QE	-	QE	-	QE
MAC (Azizpour 2015)	55.7	60.2	57.2	66.5	68	77.6	69.9	81.2
R-MAC (Tolias 2015)	67.8	72.1	71	77.5	77.4	82.4	81.4	85.5
CroW (Kalantidis 2016)	65.4	68.5	63.3	68	74.3	77.1	71.7	76.4
CAM (Jimenez 2017)	71.2	73	69.9	-	80.5	83.6	80.4	80.4
DFF	**73.4**	**74.2**	**74.8**	**78.8**	**83**	**83.8**	**83.2**	**86.2**