Faces in Places: Compound Query Retrieval

Yujie Zhong[1]
yujie@robots.ox.ac.uk
Relja Arandjelović[2]
relja.arandjelovi@inria.fr
Andrew Zisserman[1]
az@robots.ox.ac.uk

[1] Visual Geometry Group
Dept. of Engineering Science
University of Oxford, UK

[2] WILLOW project
INRIA, France

The goal of this work is to retrieve images containing both a target person and a target scene type (*e.g* Barack Obama on the beach) from a large dataset of images. At run time this compound query is handled using a face classifier trained for the person, and an image classifier trained for the scene type.

We make three contributions: **first** we propose a hybrid convolutional neural network architecture that produces place-descriptors that are aware of faces and their corresponding descriptors. We show that our jointly trained Place-CNN is able to ignore large faces in the images; and it produces descriptors that are amenable to the combination rule we choose - the relevance of a database image to the compound query is computed as the minimum of the query face and query place classification scores.

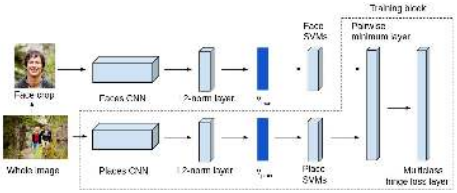

Figure 1: **Hybrid network architecture.** The network is used to generate face-descriptors and face-aware place-descriptors (v_{face} and v_{place}) for the face crop and the whole image respectively.

Second, we propose an image synthesis system to render high quality fully-labelled face-and-place images (as shown in Figure 2), and train the network only from these **synthetic** images. This synthesis system alleviates the difficulty of collecting training data and the problem of severe class-imbalance.

Some examples are shown in Figure 3. The automatic face search and replacement system successfully deals with different lighting condition and head poses, as well as accessories like hats and glasses.

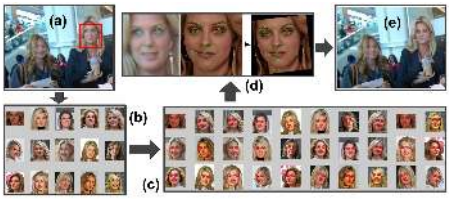

Figure 2: **Automatic synthetic rendering pipeline.** Replacing the face of the unknown person in the airport with the celebrity face (Gage Golightly).

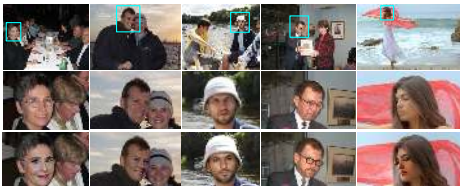

Figure 3: **Example images from the synthetic dataset.** The first row shows the synthetic images, and the second and third row show the close-up of the original and replaced face inside the blue box respectively.

Third, To test our method, and to facilitate research in compound query retrieval, we collect and annotate a 'Celebrity In Places' (CIP) Dataset of **real** images containing celebrities in different places. We test the compound query retrieval on the CIP dataset plus distractor images. As shown in the full paper, retrieval performance for compound queries is significantly improved using the face-aware place-descriptors. Figure 4 shows the top 2 retrieved images of various queries in a dataset of 73k images using our method.

Figure 4: **Examples of top ranking images returned by our retrieval system.**