# Convolutional aggregation of local evidence for large pose face alignment

Adrian Bulat
adrian.bulat@nottingham.ac.uk
Georgios Tzimiropoulos
yorgos.tzimiropoulos@nottingham.ac.uk

Computer Vision Laboratory
University of Nottingham
Nottingham, UK

Methods for unconstrained face alignment must satisfy two requirements: (a) they must not rely on accurate initialization/face detection and (b) they should perform equally well for the whole spectrum of facial pose. To the best of our knowledge, there are no methods meeting these requirements to satisfactory extent, and in this paper, we propose Convolutional Aggregation of Local Evidence (CALE), a Convolutional Neural Network (CNN) architecture particularly designed for addressing both of them.

In particular, CALE by-passes the requirement for accurate face detection by firstly using a CNN detector to perform facial landmark detection, providing at the same time confidence scores for the location of each of the facial landmarks (local evidence). Next, our system aggregates the local evidence for each facial landmark through joint CNN regression of the confidence scores, in order to refine the landmarks' location. Besides playing the role of a graphical model, CNN regression is a key feature of our system, guiding the network to rely on context for predicting the location of occluded landmarks, typically encountered in very large poses. The proposed architecture (Fig. 1) is simple and can be trained end-to-end with intermediate supervision. We show that our system achieves large performance improvement on AFLW-PIFA[1], which is, to the best of our knowledge, by far the most difficult test set for face alignment to date.

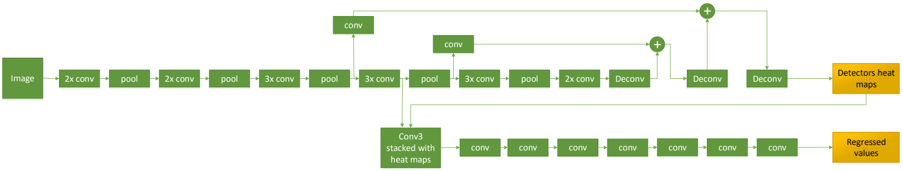

Figure 1: Proposed architecture for Convolutional Aggregation of Local Evidence (CALE).

Our second contribution in this paper is an investigation of CALE's alignment performance beyond human faces and, in particular, on animal faces. As animal faces exhibit a much larger degree of variability in shape and appearance as well as in pose and expression, animal face alignment is a much more difficult problem which, to the best of our knowledge, has never been systematically explored in the past by the Computer Vision community. Although drawing a direct comparison is not possible, our results, show that CALE's performance on animal faces is not far from that on human faces.

When applied to AFLW-PIFA[1], our method provides more than 50% absolute gain in localization accuracy when compared to other recently published methods [2, 3] for large pose face alignment. Note that prior work reports on visible points, only. To the best of our knowledge we are the first to report results on non-visible landmarks too. Remarkably, the performance of CALE when evaluated on all points - both visible and occluded surpasses the performance of all existing methods when these are evaluated on visible points only.

[1] Amin Jourabloo and Xiaoming Liu. Pose-invariant 3d face alignment. In *ICCV*, 2015.

[2] Amin Jourabloo and Xiaoming Liu. Large-pose face alignment via cnn-based dense 3d model fitting. In *CVPR*, 2016.

[3] Shizhan Zhu, Cheng Li, Chen Change Loy, and Xiaoou Tang. Unconstrained face alignment via cascaded compositional learning. In *CVPR*, 2016.