Project-Out Cascaded Regression with an application to Face Alignment

Georgios Tzimiropoulos
School of Computer Science, University of Nottingham

Contributions. Cascaded regression approaches [2] have been recently shown to achieve state-of-the-art performance for many computer vision tasks. Beyond its connection to boosting, cascaded regression has been interpreted as a learning-based approach to iterative optimization methods like the Newton's method. However, in prior work [1],[4], the connection to optimization theory is limited only in learning a mapping from image features to problem parameters.

In this paper, we consider the problem of facial deformable model fitting using cascaded regression and make the following contributions: (a) We propose regression to learn a sequence of averaged Jacobian and Hessian matrices from data, and from them descent directions in a fashion inspired by Gauss-Newton optimization. (b) We show that the optimization problem in hand has structure and devise a learning strategy for a cascaded regression approach that takes the problem structure into account. By doing so, the proposed method learns and employs a sequence of averaged Jacobians and descent directions in a subspace orthogonal to the facial appearance variation; hence, we call it Project-Out Cascaded Regression (PO-CR). (c) Based on the principles of PO-CR, we built a face alignment system that produces remarkably accurate results on the challenging iBUG data set outperforming previously proposed systems by a large margin. Code for our system is available from http://www.cs.nott.ac.uk/~yzt/.

Shape and appearance models. We use parametric shape and appearance models. An instance of the shape model is given by $\mathbf{s}(\mathbf{p}) = \mathbf{s}_0 + \mathbf{S}\mathbf{p}$. An instance of the appearance model is given by $\mathbf{A}(\mathbf{c}) = \mathbf{A}_0 + \mathbf{A}\mathbf{c}$.

Face alignment via Gauss-Newton optimization. In this section, we formulate and solve the non-linear least squares optimization problem for face alignment using Gauss-Newton optimization. This will provide the basis for learning and fitting in PO-CR in the next section. In particular, to localize the landmarks in a new image, we would like to find $\mathbf{p}$ and $\mathbf{c}$ such that [3]

$$\arg\min_{\mathbf{p},\mathbf{c}} ||\mathbf{I}(\mathbf{s}(\mathbf{p})) - \mathbf{A}(\mathbf{c})||^2. \quad (1)$$

An update for $\mathbf{p}$ and $\mathbf{c}$ can be found by solving the following problem

$$\arg\min_{\Delta\mathbf{p},\Delta\mathbf{c}} ||\mathbf{I}(\mathbf{s}(\mathbf{p})) + \mathbf{J}_I\Delta\mathbf{p} - \mathbf{A}_0 - \mathbf{A}\mathbf{c} - \mathbf{A}\Delta\mathbf{c}||^2. \quad (2)$$

By exploiting the problem structure, the calculation for the optimal $\Delta\mathbf{c}$ at each iteration is not necessary. We end up with the following problem [3]

$$\arg\min_{\Delta\mathbf{p}} ||\mathbf{I}(\mathbf{s}(\mathbf{p})) + \mathbf{J}_I\Delta\mathbf{p} - \mathbf{A}_0||^2_{\mathbf{P}}, \quad (3)$$

where $\mathbf{P} = \mathbf{E} - \mathbf{A}\mathbf{A}^T$ is a projection operator that *projects out the facial appearance variation* from the image Jacobian $\mathbf{J}_I$. The solution to the above problem is readily given by

$$\Delta\mathbf{p} = -\mathbf{H}_P^{-1}\mathbf{J}_P^T(\mathbf{I}(\mathbf{s}(\mathbf{p})) - \mathbf{A}_0). \quad (4)$$

Face alignment via Project-Out Cascaded Regression. Based on Eqs. (3) and (4), the key idea in PO-CR is to compute from a set of training examples a sequence of averaged Jacobians $\widehat{\mathbf{J}}(k)$ from which the facial appearance variation is *projected-out* and from them and descent directions:

Step I. Starting from the ground truth shape parameters $\mathbf{p}_i^*$ for each training image $\mathbf{I}_i$, $i = 1,\dots,H$, we generate a set of K perturbed shape parameters for iteration 1 $\mathbf{p}_{i,j}(1)$, $j = 1,\dots,K$ that capture the statistics of the face detection initialization process. Using the set $\Delta\mathbf{p}_{i,j}(1) = \mathbf{p}_i^* - \mathbf{p}_{i,j}(1)$, PO-CR learns the averaged projected-out Jacobian $\widehat{\mathbf{J}}_P(1) = \mathbf{P}\widehat{\mathbf{J}}(1)$ for iteration 1 by solving the following weighted least squares problem

$$\arg\min_{\widehat{\mathbf{J}}_P(1)} \sum_{i=1}^{H}\sum_{j=1}^{K} ||\mathbf{I}(\mathbf{s}(\mathbf{p}_{i,j}(1))) + \mathbf{J}(1)\Delta\mathbf{p}_{i,j}(1) - \mathbf{A}_0||^2_{\mathbf{P}}, \quad (5)$$

Step II. Having computed $\widehat{\mathbf{J}}_P(1)$, we compute $\widehat{\mathbf{H}}_P(1) = \widehat{\mathbf{J}}_P(1)^T\widehat{\mathbf{J}}_P(1)$.
Step III. The descent directions $\mathbf{R}(1)$ for iteration 1 are given by

$$\mathbf{R}(1) = \widehat{\mathbf{H}}_P(1)^{-1}\widehat{\mathbf{J}}_P(1)^T. \quad (6)$$

Step IV. For each training sample, a new estimate for its shape parameters (to be used at the next iteration) is obtained from

$$\mathbf{p}_{i,j}(2) = \mathbf{p}_{i,j}(1) + \mathbf{R}(1)(\mathbf{I}(\mathbf{s}(\mathbf{p}_{i,j}(1))) - \mathbf{A}_0). \quad (7)$$

Finally, Steps I-IV are sequentially repeated until convergence and the whole process produces a set of L regressor matrices $\mathbf{R}(l)$, $l = 1,\dots,L$.

During testing,we extract image features $\mathbf{I}(\mathbf{s}(\mathbf{p}(k)))$ and then compute an update for the shape parameters from

$$\Delta\mathbf{p}(k) = \mathbf{R}(k)(\mathbf{I}(\mathbf{s}(\mathbf{p}(k))) - \mathbf{A}_0). \quad (8)$$

Results. We conducted a large number of experiments on LFPW, Helen, AFW and iBUG data sets. In the following figure, we show fiiting results from the challenging iBUG data set.

Figure 1: Application of PO-CR to the alignment of the iBUG data set.

[1] T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active appearance models. *TPAMI*, 23(6):681–685, 2001.

[2] Piotr Dollár, Peter Welinder, and Pietro Perona. Cascaded pose regression. In *CVPR*, 2010.

[3] Georgios Tzimiropoulos and Maja Pantic. Gauss-newton deformable part models for face alignment in-the-wild. In *CVPR*, 2014.

[4] Xuehan Xiong and Fernando De la Torre. Supervised descent method and its applications to face alignment. In *CVPR*, 2013.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.