Supervised Descriptor Learning for Multi-Output Regression

Xiantong Zhen[1], Zhijie Wang[2], Mengyang Yu[3], Shuo Li[2],

[1]Department of Medical Biophysics, Western University, London, ON, Canada. [2]GE Healthcare, London, ON, Canada. [3]Northumbria University, Newcastle, United Kingdom.

Descriptor learning has recently drawn increasing attention in computer vision, Existing algorithms are mainly developed for classification rather than for regression which however has recently emerged as a powerful tool to solve a broad range of problems, e.g., head pose estimation. In this paper, we propose a novel supervised descriptor learning (SDL) algorithm to establish a discriminative and compact feature representation for multi-output regression. By formulating as generalized low-rank approximations of matrices with a supervised manifold regularization (SMR), the SDL removes irrelevant and redundant information from raw features by transforming into a low-dimensional space under the supervision of multivariate targets. The obtained discriminative while compact descriptor largely reduces the variability and ambiguity in multi-output regression, and therefore enables more accurate and efficient multivariate estimation. We demonstrate the effectiveness of the proposed SDL algorithm on a representative multi-output regression task: head pose estimation using the benchmark Pointing'04 dataset. Experimental results show that the SDL can achieve high pose estimation accuracy and significantly outperforms state-of-the-art algorithms by an error reduction up to 27.5% . The proposed SDL algorithm provides a general descriptor learning framework in a supervised way for multi-output regression which can largely boost the performance of existing multi-output regression tasks.

Given a set of annotated training data $\{X_1, \ldots, X_L\}$ and the corresponding multivariate targets $\{Y_1, \ldots, Y_L\}$, where L is the number of training samples and $Y_i \in \mathbb{R}^d$, our task is to learn discriminative and compact representations of matrices. Instead of using a vectorized input space, we consider matrix representations, *i.e.*, $X_i \in \mathbb{R}^{M \times N}$, which could be any matrix representations of images, *e.g.*, raw pixel intensities.

We propose using GLRAM due to its efficient computation of dimension reduction of matrices [3]. This is to find two-side transformations: $W \in \mathbb{R}^{M \times m}$ and $V \in \mathbb{R}^{N \times n}$ with $m \ll M$ and $n \ll N$, and L matrices $D_i \in \mathbb{R}^{m \times n}$ such that WD_iV^{T} is an appropriate approximation of each X_i, $i = 1, \ldots, L$. We solve the following optimization problem of minimizing the reconstruction errors:

$$\underset{\substack{W,V,D_1,\ldots,D_L \\ W^{\mathrm{T}}W=I_m, V^{\mathrm{T}}V=I_n}}{\arg\min} \frac{1}{L}\sum_{i=1}^{L} \|X_i - WD_iV^{\mathrm{T}}\|_F^2 \tag{1}$$

where $\|\cdot\|_F$ is the Frobenius norm of a matrix, I_m is an identity matrix of size $m \times m$ and the constraints $W^{\mathrm{T}}W = I_m$ and $V^{\mathrm{T}}V = I_n$ ensure that W and V have orthogonal columns to avoid redundancy in the approximations.

In order to to achieve discriminative representations, we propose a supervised manifold regularization (SMR) to explore the multivariate target space for supervised descriptor learning. We impose discrimination on the low-rank representation $\{D_i\}_{i=1}^{L}$ by integrating the proposed SMR into (1). To this end, we first construct a weighted graph $G = (V, E)$ using the ε-neighborhood method [2], where V and E respectively represent L vertices and edges between vertices. The graph is built on the multivariate targets $(Y_1, \ldots, Y_L)$ rather than on inputs in conventional manifold regularization [1, 4], which naturally induces the supervision.

$$\sum_{i,j} \|D_i - D_j\|_F^2 S_{ij}. \tag{2}$$

Combining (2) and (1), we obtain the compact objective function of GLRAM with the SMR as follows:

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

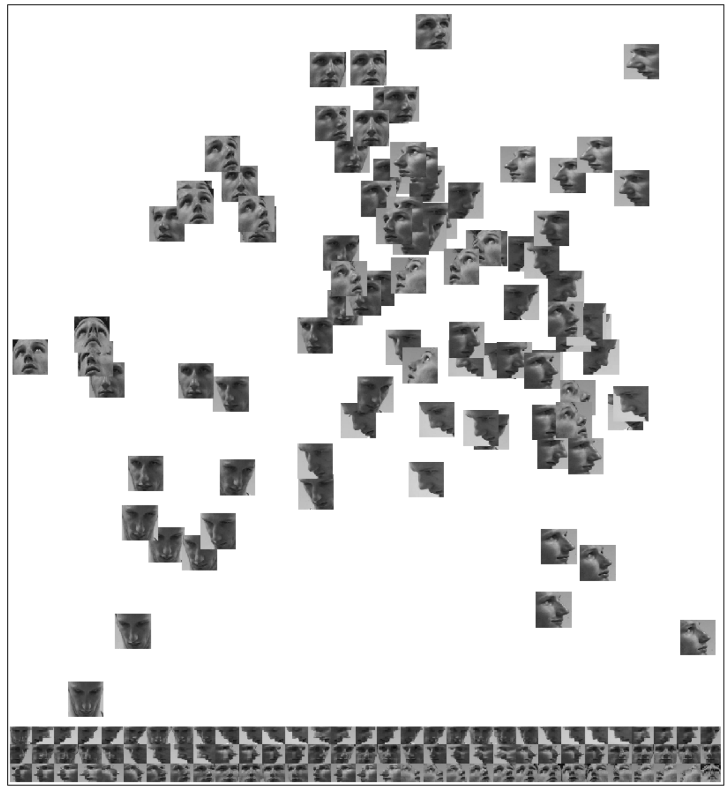

Figure 1: Illustration of head pose images in a two-dimensional space with $m = 1$ and $n = 2$. Head poses with similar orientations tend to be clustered while these with distinctive orientations are scattered away. The bottom are the images with all 93 different pose orientations.

$$\underset{\substack{W,V,D_1,\ldots,D_L \\ W^{\mathrm{T}}W=I_m, V^{\mathrm{T}}V=I_n}}{\arg\min} \underbrace{\frac{1}{L}\sum_{i=1}^{L} \|X_i - WD_iV^{\mathrm{T}}\|_F^2}_{\text{Low-rank approximation errors}} + \underbrace{\beta \sum_{i,j} \|D_i - D_j\|_F^2 S_{ij}}_{\text{Supervised manifold regularization}} \tag{3}$$

where $\beta \in (0, \infty)$ is a tuning parameter to balance the tradeoff between reconstruction errors and discrimination of the low-rank approximations, which also serves to keep the flexibility of the model.

In the objective function of (3), the first term guarantees the reconstruction fidelity in the low-rank approximation while the second SMR term introduces the discrimination to learned new representations. The objective function in (3) can not be solved straightforwardly using existing methods. We seek an alternative objective function which can be efficiently solved by an iterative algorithm via an alternate optimization.

[1] Mikhail Belkin, Partha Niyogi, and Vikas Sindhwani. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *JMLR*, 7:2399–2434, 2006.

[2] Xiaofei He and Partha Niyogi. Locality preserving projections. In *NIPS*, volume 16, page 153, 2004.

[3] Jieping Ye. Generalized low rank approximations of matrices. *Machine Learning*, 61(1-3):167–191, 2005.

[4] Zhenyue Zhang and Keke Zhao. Low-rank matrix approximation with manifold regularization. *IEEE TPAMI*, 35(7):1717–1729, 2013.