SEMI-SUPERVISED MULTI-OUTPUT IMAGE MANIFOLD REGRESSION

Hui Wu[⋆] *Scott Spurlock*[†] *Richard Souvenir*[‡]

[⋆] IBM Research
[†] Department of Computing Sciences, Elon University
[‡] Department of Computer and Information Sciences, Temple University

ABSTRACT

We present a data-driven method for semi-supervised multi-output regression on image manifolds, which simultaneously considers the manifold structure of the input data and complex output labels. Compared to related methods, our method achieves superior prediction accuracy on a variety of data sets, with as few as 5% of the input examples labeled. Also, with a few labeled examples and no domain-specific tuning, our method performs on par with specialized algorithms for tasks such as face landmark detection.

Index Terms— semi-supervised; image manifolds

1. INTRODUCTION

Contour-based segmentation and articulated pose estimation can be framed as multi-output regression (Figure 1), with image input and multi-dimensional structured output. Typical methods for acquiring and verifying labels (e.g., crowdsourcing) are often insufficient for multi-dimensional label vectors; often domain expertise is needed (e.g., left ventricle segmentation from noisy ultrasound imagery). A semi-supervised approach can balance the increasing availability of visual data with the expense of expert manual annotation.

We present a semi-supervised multi-output algorithm designed for image manifold regression. Compared to methods that apply regularization only on the data manifold, our approach considers the manifold structure of both the images and multi-dimensional labels and applies regularization in both spaces. Image manifolds provide a natural model for semi-supervised learning; we review methods for manifold-regularized, semi-supervised learning. There have been methods suited to classification tasks [1, 2]. For regression on image sets, supervised approaches, such as support vector regression (SVR) [3] and Kernel Supervised PCA (KSPCA) [4], have been considered, including an approach for supervised, multi-output regression [5]. For semi-supervised regression; one method [6] addresses the problem of scalar (single-output) regression on image manifolds.

Other multi-manifold problems include manifold alignment and canonical correlation analysis (CCA). In manifold alignment, a projection is learned from each manifold to a common low-dimensional space, which generally preserves inter- and intra-manifold correspondences [7, 8]. CCA finds the embedding that maximizes the consensus between two data sets, and recent work has explored both semi-supervised and kernelized variants [9, 10]. However, rather than learning a common embedding, we aim to leverage the correspondences between manifolds to learn the missing labels.

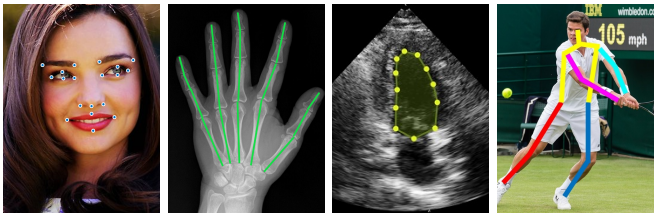


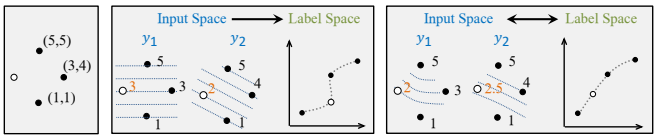
Fig. 1: Our algorithm provides a domain-agnostic approach for a variety of tasks, such as pose estimation.

In the broader area of structured output learning, there have been extensions to the semi-supervised setting [11, 12, 13]. The main drawback is that most of these approaches require task-specific models for the joint feature space of the input and labels. In some cases, modeling these joint kernels and defining efficient searches in these joint spaces is tantamount to designing a specialized application for the task.

Our method (depicted in Figure 2) leverages the underlying manifold structure of both the image and label spaces to learn smoothly-varying, multi-dimensional labels. The main contributions are: (1) extending image manifold regression to the multi-output setting; (2) an efficient, data-driven, dual regularization algorithm for semi-supervised, multi-output regression; and (3) applying multi-output image manifold regression to specialized tasks, e.g., facial landmark detection and left ventricle segmentation.

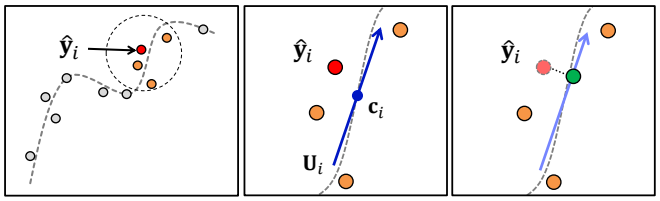
2. METHOD

Consider an image set, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]^\intercal$, where $\mathbf{x}_i \in \mathcal{R}^{D_X}$ corresponds to the D_X-dimensional image feature representation of image i, and corresponding multi-dimensional labels, $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N]^\intercal$, where $\mathbf{y}_i \in \{\mathcal{R}^{D_Y}, \varnothing\}$, D_Y



(a) Input (b) Manifold Regularization (c) Dual Regularization

Fig. 2: (a) This example with 2D labels, $\vec{y}$, shows one unlabeled and three labeled input points. (b) For traditional manifold regularization, the predictions (each dimension shown separately) for the unlabeled example vary smoothly only in the input space. (c) With our method, the predicted labels vary smoothly in *both* the input and label spaces.



(a) (b) (c)

Fig. 3: Each point represents a multi-dimensional label in label space. For a given label (denoted by the red point), the regularizer encourages the predicted label to be close to the local tangent space (denoted by the green point).

is the dimensionality of image labels, and $\mathbf{y}_i = \varnothing$ means image i is unlabeled. We assume that the images are samples from a manifold, $\mathcal{M}_X$, embedded in the feature space, and there exists a smooth function, $f : \mathcal{R}^{D_X} \rightarrow \mathcal{R}^{D_Y}$, which maps the input image features to the labels, and that the ideal labels, $\mathbf{y}_i^\star = f(\mathbf{x}_i)$, are samples from this output manifold, $\mathcal{M}_Y$. Our goal is to predict the optimal label set, $\hat{\mathbf{Y}} = [\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \ldots, \hat{\mathbf{y}}_N]^\intercal$, such that $\hat{\mathbf{y}}_i = \mathbf{y}_i^\star$, using:

$$\underset{\hat{\mathbf{Y}}}{\text{argmin}} \quad R(\mathbf{X}) + \lambda_S S(\hat{\mathbf{Y}}) + \lambda_L L(\mathbf{Y}; \hat{\mathbf{Y}}) \tag{1}$$

where R regularizes the function over the input image manifold, S regularizes the function over the output label manifold, and L is a loss function on the labeled examples.

2.1. Image Manifold Regularization

The Hessian regularizer has been previously applied to nonlinear dimensionality reduction [14] and scalar label prediction in the semi-supervised [6] and robust [15] settings. For a point on the manifold, the local Hessian functional is defined on its associated tangent space as the Frobenius norm of the Hessian matrix. The local measure is then averaged over the entire manifold to provide a global measurement, which is an extension of the average Frobenius norm of the Hessian of a function in the Euclidean space to manifolds.

Consider one dimension, $\hat{y}$, of the label, $\hat{\mathbf{y}}$. For an input point, $\mathbf{x}_i$, let $\mathcal{N}_i$ represent the neighborhood of K nearest neighbors and $\mathbf{z}_j^{(i)}$ represents the coordinates of $\mathbf{x}_j \in \mathcal{N}_i$ in the d-dimensional tangent space of $\mathbf{x}_i$, where $\mathbf{z}_i^{(i)}$ is defined as the origin. The local Hessian functional estimates a second-order polynomial, f, near $\mathbf{x}_i$ of the form:

$$f = \hat{y}_i + \mathbf{J}^{(i)}\mathbf{z}^{(i)} + \frac{1}{2}\mathbf{z}^{(i)\intercal}\mathbf{H}^{(i)}\mathbf{z}^{(i)} \tag{2}$$

where $\mathbf{J}^{(i)}$ and $\mathbf{H}^{(i)}$ are the local Jacobian and Hessian matrices, respectively, $\hat{y}_i$ is the predicted label, and $\mathbf{z}^{(i)}$ is the d-dimensional tangent space coordinate. Equation 2 is linear with respect to $\mathbf{J}^{(i)}$ and $\mathbf{H}^{(i)}$. This leads to the global regularizer, $\hat{\mathbf{y}}^\intercal\mathbf{B}\hat{\mathbf{y}}$. (Details for the construction of $\mathbf{B}$ can be found in [14].) Minimizing the global Hessian functional encourages the label function to be locally linear on the image manifold. Extending this for multi-dimensional output, gives:

$$\text{Tr}(\hat{\mathbf{Y}}^\intercal\mathbf{B}\hat{\mathbf{Y}}) \tag{3}$$

2.2. Label Space Regularization

We assume the multi-dimensional labels sparsely sample the D_Y-dimensional label space and lie on or near a d_Y-dimensional manifold, $\mathcal{M}_Y \in \mathcal{R}^{D_Y}$, providing an additional avenue for regularization: the predicted labels should be points drawn from (or near) a locally linear output manifold. Using the tangent space estimated by a small set of neighboring labels, we estimate the projection of a point on the output manifold, as shown in Figure 3.

Let $\mathcal{N}_i^Y$ be the set of K_Y nearest neighbors of label, $\hat{\mathbf{y}}_i \in \hat{\mathbf{Y}}$. The tangent space of $\hat{\mathbf{y}}_i$ is modeled using PCA to obtain the mean, $\mathbf{c}_i$, and the basis, $\mathbf{U}_i$, where $\mathbf{U}_i$ is a matrix of size $d_Y \times D_Y$. The projection of $\hat{\mathbf{y}}_i$ to the tangent space defined by $\mathbf{c}_i$ and $\mathbf{U}_i$ is $(\mathbf{U}_i)^\intercal\mathbf{U}_i(\hat{\mathbf{y}}_i - \mathbf{c}_i) + \mathbf{c}_i$. The output label manifold regularizer minimizes the difference between the predicted label and its reconstruction on the associated local tangent space:

$$\sum_{i=1}^{N} ||\hat{\mathbf{y}}_i - ((\mathbf{U}_i)^\intercal\mathbf{U}_i(\hat{\mathbf{y}}_i - \mathbf{c}_i) + \mathbf{c}_i)||_2^2 \tag{4}$$

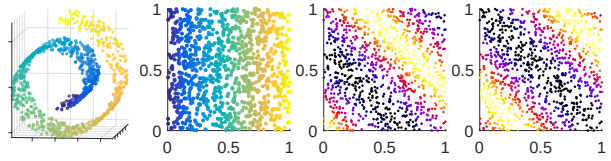
2.3. Algorithm

Even though the Hessian operator, $\mathbf{B}$, is positive semi-definite, substituting Equation 3, Equation 4, and $L2$ loss for the labeled examples into Equation 1 leads to a non-convex optimization due to the label regularization. However, fixing $\mathbf{c}$ and $\mathbf{U}$ leads to a quadratic function of $\hat{\mathbf{Y}}$, which can be solved efficiently. Our method, *Semi-Supervised Dual-Regularized Manifold Regression* (`SS-DRMR`), outlined in Algorithm 1, applies an alternating minimization approach, iterating between updating $\mathbf{c}$ and $\mathbf{U}$ and solving for $\hat{\mathbf{Y}}$.

Algorithm 1 SS-DRMR

Input: image features, $\mathbf{X}$; labels, $\mathbf{Y}$
Output: predicted labels, $\hat{\mathbf{Y}}$

Compute the global Hessian operator, $\mathbf{B}$
Estimate $\mathbf{c}$ and $\mathbf{U}$ using labeled examples
Solve for $\hat{\mathbf{Y}}_{(0)}$
$k \leftarrow 0$
repeat
  $k \leftarrow k + 1$
  With $\hat{\mathbf{Y}}_{(k-1)}$, estimate $\mathbf{c}$ and $\mathbf{U}$
  Solve for $\hat{\mathbf{Y}}_{(k)}$
until $||\hat{\mathbf{Y}}_{(k)} - \hat{\mathbf{Y}}_{(k-1)}||_F < \epsilon$ or $k = k_{max}$
$\hat{\mathbf{Y}} \leftarrow \hat{\mathbf{Y}}_{(k)}$



(a) Input (3D) (b) Input (2D) (c) Label (y_1) (d) Label (y_2)

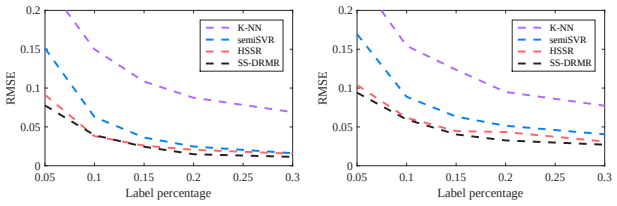
Fig. 4: For the 2D label, $\mathbf{y} = [y_1\ y_2]$, the value of each dimension is indicated by the color of the points. For clarity, (b)-(d) are plotted using 2D manifold coordinates.

3. EVALUATION

We compare SS−DRMR with: (1) $K-$NN: average of the K nearest labeled input examples; (2) Semi-supervised support vector regression (SemiSVR) [16] with RBF kernel; (3) Hessian semi-supervised regression (HSSR) [6]: scalar regression method for nonlinear manifolds; (4) Kernel Supervised PCA (KSPCA) [4]. The free parameters of each method are optimized using grid search and 5-fold cross-validation. KSPCA is fully-supervised, so trained using only the labeled examples. For the single-output methods, output dimensions were predicted independently. In §3.3, we apply SS−DRMR to computer vision tasks that can be viewed as instances of multi-output regression. For these problems, the experiments are not meant to be an exhaustive evaluation, but a comparison of SS−DRMR and representative methods.

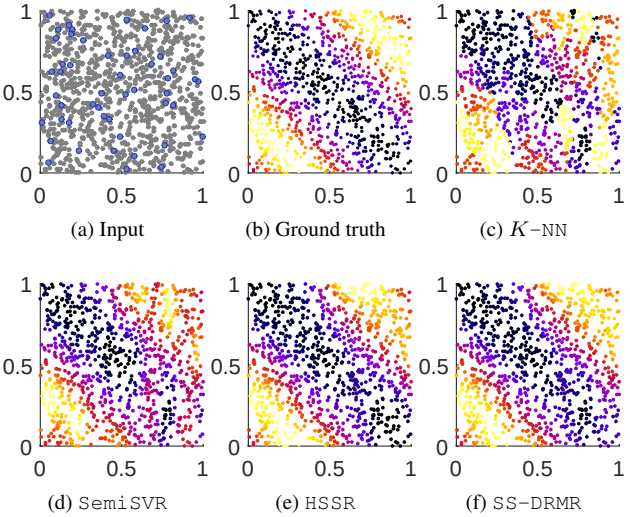
3.1. Implementation Details

For SS−DRMR, the free parameters can be specified using prior knowledge of the data or estimated from the input. To estimate the intrinsic manifold dimensionality, we apply PCA on a neighborhood of 20 points from a small set of randomly selected examples and use the value corresponding to the "elbow point" of the residual variance curve. We found that the algorithm was robust to a large range of neighborhood sizes and set $K = 0.05N$ for both the input and output manifolds. For the regularization parameters, λ_S and λ_L, we use 5-fold cross validation to select values in the range $[10^{-4}, 10^{4}]$. For the termination criteria, we observed that for a range of values of ϵ, most experiments converged within 20 iterations, so we set $\epsilon = 10^{-2}$ and $k_{max} = 20$. The values of each label dimension are scaled to the range $[0, 1]$. Results are reported as the root mean squared error (RMSE) of the predictions averaged over 10 trials.



(a) Clean (b) Noisy

Fig. 5: RMSE of 10 repeated trials on Swiss Roll data.



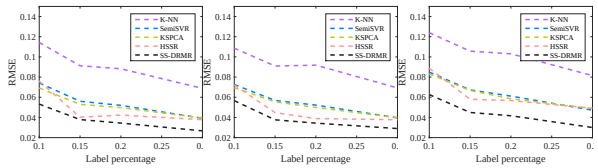
(a) Input (b) Ground truth (c) $K-$NN
(d) SemiSVR (e) HSSR (f) SS−DRMR

Fig. 6: Swiss Roll results with 5% labeled points. (a) shows the labeled input examples as blue circles. (b) shows the ground truth values for one dimension of the output labels. (c)-(f) show the output of each method.

3.2. Semi-Supervised Multi-Output Prediction

The **Swiss Roll** data (Figure 4) consists of 1,000 points sampled from a 2D manifold embedded in 3D. For each 3D input point, the 2D label is cosine and sine of the sum of the manifold coordinates, respectively. Though the input is 3D, for ease of visualization, the plots use the 2D manifold coordinates. In Figure 4, the value of each dimension of the label (Figure 4(c)) and (d)) is indicated by color.

We varied the percent of labeled input examples from 5% to 30% and added Gaussian noise ($\sigma = 0.05$) to the labels.



(a) Width (b) Height (c) Area

Fig. 7: Results from the Leaf Images experiment. Each plot shows the RMSE averaged over 10 trials for each dimension of the output label prediction task.

(a) (b) (c) (d) (e)

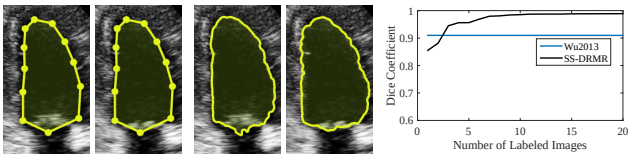
Fig. 8: Facial landmark detection output for SS−DRMR (with (a) 2, (b) 5, and (c) 10 labeled examples) and (d) Asthana2013. (e) RMSE (in pixel units) for facial landmarks as a function of the number of labeled examples.

Figure 5 shows the RMSE of the predicted label compared to the ground truth. With or without label noise, SS−DRMR has the lowest RMSE among all methods. HSSR and SS−DRMR, which both apply Hessian regularization, performed similarly well at this toy prediction task. Figure 6 shows the predictions for an experiment with 5% of the input labeled. K−NN shows local patches of errors, and SemiSVR shows smoothness, but global distortions compared to ground truth.

The **Leaf Images** dataset varies due to non-rigid leaf shape change and rotation. The labels for each image are three shape descriptors: height, width, and area. As before, we varied the percentage of labeled examples from 5% to 30% and added Gaussian noise ($\sigma = 0.05$) to each label dimension. Figure 7 shows the prediction error for the Leaf Image data, with SS−DRMR being the top performer. There is an increased margin between the previously-comparable HSSR and SS−DRMR methods. For more complex label spaces, the benefit of dual regularization becomes more evident as our approach, SS−DRMR, provides more accurate predictions than the single regularization approaches.

3.3. Application to Computer Vision Tasks

Facial landmarks are typically represented as a set of 2D points located near facial features (eyes, nose, mouth). We compare SS−DRMR to a recent facial landmark detection algorithm (Asthana2013 [17]). The input consists of low-resolution images (100×72) from YouTube Faces [18]. SS−DRMR uses a randomly-selected subset as labeled examples. Asthana2013 is pre-trained, so does not require



(a) SS−DRMR (b) Wu2013 (c)

Fig. 9: Representative segmentation from (a) SS−DRMR (with 5 labeled images) and (b) texttttWu2013. (c) Average Dice coefficient (higher is better) as a function of the number of labeled examples.

labeled input. Figure 8 shows example frames with landmark detection results and the RMSE (in pixel units). With as few as 5 labeled images (<4% of the data set), SS−DRMR performs on par with Asthana2013, and, with additional labeled input, shows decreasing prediction error.

For **Left Ventricle Segmentation**, the input is an ultrasound video consisting of 180 frames (∼5 cardiac cycles). The contour is represented by the 2D image locations of 11 control points. We compare SS−DRMR to a recent image segmentation method (Wu2013 [19]), based on an adaptive diffusion flow active-contour model. The free parameters of Wu2013 were optimized using grid search. Figure 9 shows representative results and reports the Dice coefficient, which measures the overlap between the predicted segmentation and ground truth. With as few as 3 labeled images (<2% of the data set), SS−DRMR performs as well as the active contour approach, Wu2013. With 10 labeled images (∼5% of the data set), SS−DRMR achieves nearly ideal segmentation. For this challenging segmentation task with low-gradient edges, the specialized method, Wu2013, tended to overestimate the area of the segmented region in the presence of low contrast structures. Additionally, Wu2013 is sensitive to its free parameters; using default settings, the average Dice coefficient much lower (0.72). Our approach requires no problem- or data-specific tuning.

4. CONCLUSIONS AND FUTURE WORK

We presented an algorithm for multi-output regression on high-dimensional manifolds. In addition to the contribution to image manifold learning, this approach provides an alternative to obtaining and tuning specialized code for new data. Despite the minimal supervision, our approach is trained directly on the data of interest, which is certainly advantageous compared to off-the-shelf approaches. This highlights an explicit trade-off between ease-of-use, manual effort, and expected accuracy, which the end-user must balance for a particular task. For future work, we plan to develop a formulation of label space regularization as a loss function for deep convolutional neural network (CNN) learning, to take advantage of the inherent scale and parallelization benefits.

5. REFERENCES

[1] Mikhail Belkin, Partha Niyogi, and Vikas Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *The Journal of Machine Learning Research*, vol. 7, pp. 2399–2434, 2006.

[2] Rob Fergus, Yair Weiss, and Antonio Torralba, "Semi-supervised learning in gigantic image collections," in *Advances in Neural Information Processing Systems 22*, Y. Bengio, D. Schuurmans, J.D. Lafferty, C.K.I. Williams, and A. Culotta, Eds., pp. 522–530. Curran Associates, Inc., 2009.

[3] Chih-Chung Chang and Chih-Jen Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 27, 2011.

[4] E. Barshan, A. Ghodsi, Z. Azimifar, and M. Zolghadri Jahromi, "Supervised principal component analysis: visualization, classification and regression on subspaces and submanifolds," *Pattern Recognition*, vol. 44, no. 7, pp. 1357–1371, 2011.

[5] Guangcan Liu, Zhouchen Lin, and Yong Yu, "Multi-output regression on the output manifold," *Pattern Recognition*, vol. 42, no. 11, pp. 2737–2743, 2009.

[6] Kwang I Kim, Florian Steinke, and Matthias Hein, "Semi-supervised regression using hessian energy with an application to semi-supervised dimensionality reduction," in *Advances in Neural Information Processing Systems*, 2009, pp. 979–987.

[7] Chang Wang and Sridhar Mahadevan, "Manifold alignment using procrustes analysis," in *Proceedings of international conference on Machine learning*. ACM, 2008, pp. 1120–1127.

[8] Chang Wang and Sridhar Mahadevan, "Manifold alignment preserving global geometry," in *Proceedings of international joint conference on Artificial Intelligence*, 2013, pp. 1743–1749.

[9] Chenping Hou, Changshui Zhang, Yi Wu, and Feiping Nie, "Multiple view semi-supervised dimensionality reduction," *Pattern Recognition*, vol. 43, no. 3, pp. 720–730, 2010.

[10] Matthew B Blaschko, Christoph H Lampert, and Arthur Gretton, "Semi-supervised laplacian regularization of kernel canonical correlation analysis," in *Machine Learning and Knowledge Discovery in Databases*, pp. 133–145. Springer, 2008.

[11] V. Badrinarayanan, I. Budvytis, and R. Cipolla, "Semi-supervised video segmentation using tree structured graphical models," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 11, pp. 2751–2764, Nov 2013.

[12] Yasemin Altun, Mikhail Belkin, and David A Mcallester, "Maximum margin semi-supervised learning for structured variables," in *Advances in neural information processing systems*, 2005, pp. 33–40.

[13] Ha Q Minh and Vikas Sindhwani, "Vector-valued manifold regularization," in *Proceedings of the International Conference on Machine Learning*, 2011, pp. 57–64.

[14] David L Donoho and Carrie Grimes, "Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data," *Proceedings of the National Academy of Sciences*, vol. 100, no. 10, pp. 5591–5596, 2003.

[15] Hui Wu and Richard Souvenir, "Robust regression on image manifolds for ordered label denoising," in *Computer Vision and Pattern Recognition, IEEE Conference on*, 2015.

[16] G. Camps-Valls, J. Munoz-Mari, L. Gomez-Chova, K. Richter, and J. Calpe-Maravilla, "Biophysical parameter estimation with a semisupervised support vector machine," *Geoscience and Remote Sensing Letters, IEEE*, vol. 6, no. 2, pp. 248–252, April 2009.

[17] Akshay Asthana, Stefanos Zafeiriou, Shiyang Cheng, and Maja Pantic, "Robust discriminative response map fitting with constrained local models," in *Computer Vision and Pattern Recognition, IEEE Conference on*, 2013, pp. 3444–3451.

[18] Lior Wolf, Tal Hassner, and Itay Maoz, "Face recognition in unconstrained videos with matched background similarity," in *Computer Vision and Pattern Recognition, IEEE Conference on*, 2011, pp. 529–534.

[19] Yuwei Wu, Yuanquan Wang, and Yunde Jia, "Adaptive diffusion flow active contours for image segmentation," *Computer Vision and Image Understanding*, vol. 117, no. 10, pp. 1421–1435, 2013.