Feature selection methods are efficient in modern computer vision applications to reduce the computational cost and the chance of over-fitting. Recently, a novel selectable factor extraction (SFE) framework is proposed to simultaneously perform feature selection and extraction, and is theoretically and practically proved to be effective for high-dimensional data. Although it is advantageous in several aspects, SFE is only designed for either supervised or unsupervised learning, and is not suitable when there are limited labeled samples and a large number of unlabeled samples. To tackle this problem, we propose a novel manifold regularized SFE (MRSFE) framework for semi-supervised image classification.

We use a low rank penalized regression model to explore the label information. A low rank matrix of the regression coefficients, together with the ℓ 1 penalty is learned for joint feature selection and extraction. In addition, all the labeled and unlabeled samples are utilized in MRSFE to construct the data adjacency graph to approximate the subspace after dimension reduction. So we can implement feature selection and extraction simultaneously.

To tackle this problem, we propose a novel manifold regularized SFE (MRSFE) framework for semi-supervised image classification.

We use alternating optimization method to solve model (4) and (5), together with some thresholding operations which are at low cost. We demonstrate the superiority of the proposed method.

![Figure 1: Prediction accuracy vs. the number of selected features (left) using linear SVM (right) using nonlinear SVM](image)

significant features from X_L, while the orthogonal matrix V determines the subspace after dimension reduction. So we can implement feature selection and extraction simultaneously.

Considering the ℓ 1 penalty cannot handle the collinearity and may lead to inconsistent and biased estimation(4), similar to ℓ 2,1, we advocate to use non-convex constraint such as ℓ 2,0 instead of widely-used ℓ 1 penalty for S in (4)

\[
\min_{X \in \mathbb{R}^{d \times c}} \|Y_L - X_L S V^T\|_F^2 + \beta \text{tr}(V^T L_X L_U V^T S V^T), \text{s.t. } \|S\|_{0,2} \leq q_0
\]

where q_0 is a parameter to control the number of selected features. Using the constraint form instead of the penalty is intuitive, because we can directly control the number of features we need. Although the ℓ 2,0 penalty is nonconvex and hard to optimization in classical methods, it is doable in our algorithm. Once S is obtained from (4) or (5), we can select the significant features according to top-k index of the row-norms in descending order or nonzero rows of S. We call both of the models (4) and (5) manifold-regularized semi-supervised selectable factor extraction method (MRSFE).

We use alternating optimization method to solve model (4) and (5), an efficient and easy-to-implement algorithm is designed to find the solutions. Our algorithm only consists of SVD decomposition of small-scale W, together with some thresholding operations which are at low cost. We evaluate the effectiveness of our MRSFE by applying it to a challenge web image dataset, NUS-WIDE-OBJECT. Experiments on this dataset demonstrate the superiority of the proposed method.