Efficient Large Scale Image Classification via Prediction Score Decomposition

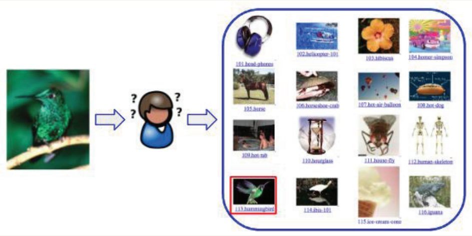
Duy-Dinh Le[1,2], Tien-Dung Mai[2], Shin'ichi Satoh[1], Thanh Duc Ngo[2], Duc Anh Duong[2]

(1)National Institute of Informatics (NII), Japan

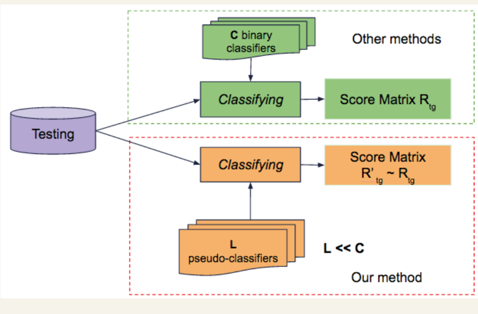
(2) University of Information Technology (UIT), Vietnam

ECCV'16 EUROPEAN CONFERENCE ON COMPUTER VISION

Introduction



- **Task:** Multi-class Classification → Given an image, predict which class that it belongs to.
- **Goal:** Reducing the test time complexity → necessary when the number of classes is large.
- **Challenges:**
  - The complexity in test time grows linearly with the number of classes when using the standard OvA approach.
  - The label tree approach usually suffers the well-known error propagation problem and it is difficult for parallelization for further speedup.
- **Proposal:**
  - The key idea is to use a smaller number of classifier evaluations to reduce the testing cost.
  - It is similar to the label tree approach, but hierarchical structure is not used.
  - It is done through prediction score matrix decomposition.



Method Overview

Let's revisit SVD.

$$R = U\Sigma V^T \quad (2)$$
$$RR^T = U\Sigma V^T V\Sigma U^T = U\Sigma^2 U^T \quad (3)$$
$$RR^T U = U\Sigma^2. \quad (4)$$

Instead of obtaining U directly by singular value decomposition, we take into account that U is the result of performing regression on the feature vectors of the images. To do so, we will pose the original problem as an eigenvalue problem,

$$RR^T u = \lambda u \quad (5)$$
$$u^T RR^T u = u^T \lambda u = \lambda \quad (6)$$

where u is an eigenvector and λ is the corresponding eigenvalue.

Now we consider u as regression result, namely, $u_{(i,k)} \approx g_k(v_i)$. We further assume linear regression:

$$g_k(v_i) = < w_k, v_i > + b_k \quad (7)$$
$$= [w_k\, b_k][v_i\, 1]^T \stackrel{\text{def}}{=} \tilde{w_k}\tilde{v_i}. \quad (8)$$

By defining the matrix of features:

$$\tilde{S} = \begin{bmatrix} v_1 & v_2 & \cdots & v_N \\ 1 & 1 & \cdots & 1 \end{bmatrix} \quad (9)$$

we want $u \approx \tilde{S}^T \tilde{w}$. Substituting this into (6), the problem becomes:

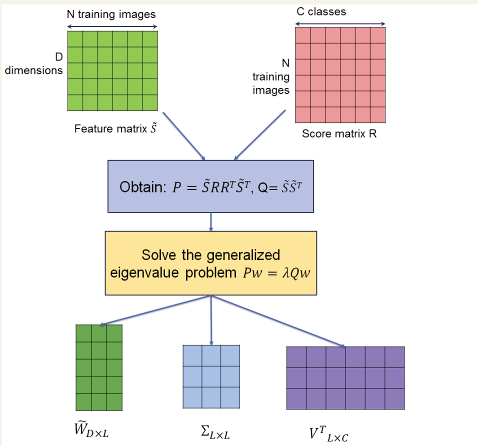
$$\text{maximize } \tilde{w}^T \tilde{S} RR^T \tilde{S}^T \tilde{w} \quad (10)$$
$$\text{such that } \tilde{w}^T \tilde{S}\tilde{S}^T \tilde{w} = 1 \quad (11)$$

We can use the Lagrange multipliers method to solve it.

$$J = \tilde{w}^T \tilde{S} RR^T \tilde{S}^T \tilde{w} - \lambda(\tilde{w}^T \tilde{S}\tilde{S}^T \tilde{w} - 1) \quad (12)$$
$$\frac{\partial L}{\partial \tilde{w}} = 2\tilde{S}RR^T\tilde{S}^T\tilde{w} - 2\lambda\tilde{S}\tilde{S}^T\tilde{w} = 0 \quad (13)$$
$$\tilde{S}RR^T\tilde{S}^T\tilde{w} = \lambda\tilde{S}\tilde{S}^T\tilde{w}. \quad (14)$$

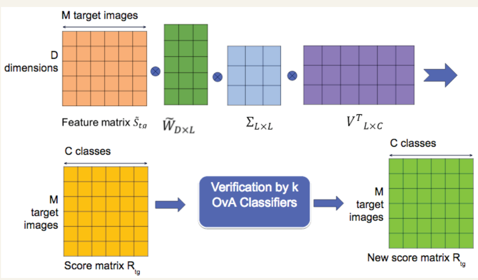
The above can be regarded as a generalized eigenvalue problem $Pw = \lambda Qw$, where $P = \tilde{S}RR^T\tilde{S}^T$ and $Q = \tilde{S}\tilde{S}^T$.

Training Stage



$\tilde{W}$: the eigenvectors corresponding to the L largest eigenvalues.
Σ : the eigenvalues corresponding to the L largest eigenvalues.
V^T: the right singular vectors corresponding to the L largest eigenvalues.

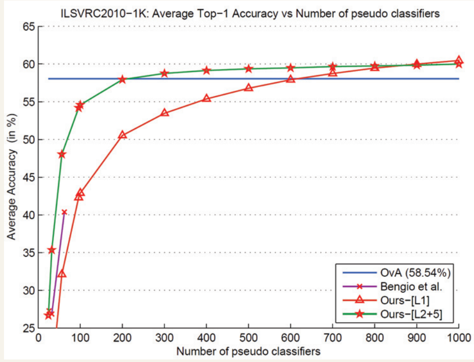
Testing Stage



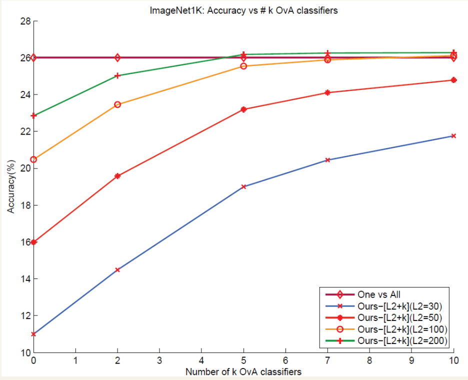
Accuracy and Test Speedup

Method	Flat		$T_{32,2}$ (L=96)		$T_{10,3}$ (L=56)		$T_{6,4}$ (L=32)	
	Acc(%)	S_{te}	Acc(%)	S_{te}	Acc(%)	S_{te}	Acc(%)	S_{te}
Fast-Balanced Tree[10]			11.9	10.3	8.92	18.20	5.62	31.3
Probabilistic Tree [24]			21.38	10.42	20.54	17.85	17.02	31.25
RDOC [2]			3.35	10.42	2.15	17.86	1.28	31.25
RSOC [2]			3.48	10.42	2.18	17.86	1.34	31.25
Spectral ECOC [39]			7.18	10.42	5.57	17.86	4.25	31.25
Attribute-based learning [21]			12.13	10.42	8.07	17.86	5.81	31.25
Ours-[L_1]			20.31	10.42	16.96	17.86	11.67	31.25
Ours-[$L_2 + k$]			**25.38**	10.42	**23.32**	17.86	**17.88**	31.25
1K-OvA(LIBLINEAR)	26.01	1.0						

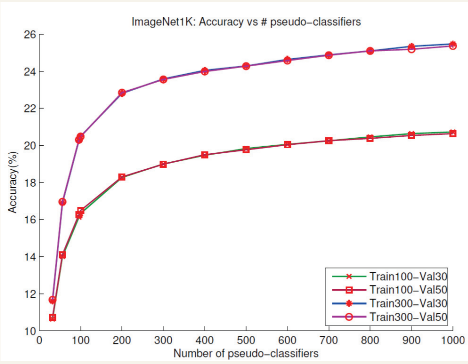
Results



Effect of the Number of Pseudo-Classifiers



Effect of the Verification Stage



Effect of Quality of the Score Matrix R (training)

Conclusion

- Propose an efficient method for large scale image classification .
- The method is based on joint optimization for prediction score matrix decomposition.
- Comprehensive evaluations on large datasets such as ImageNet-1K, ImageNet-10K show superiority over SoA methods such as OvA and label tree methods.