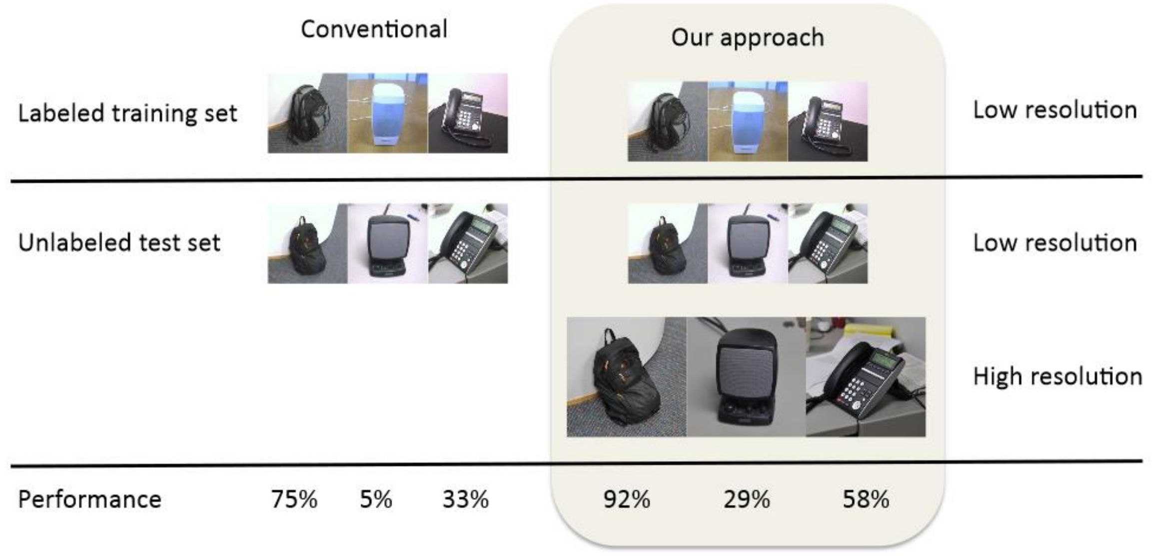
Learning to Recognize Objects from Unseen Modalities

C. Mario Christoudias, Raquel Urtasun, Mathieu Salzmann and Trevor Darrell
Presented by Deepak Kumar and Shishir Mathur for Machine Learning for Computer Vision
Guided by Dr. Vinay P. Namboodiri
Department of Computer Science and Engineering, Indian Institute of Technology Kanpur
Email deepakr@iitk.ac.in, mshishir@iitk.ac.in

Introduction

- The aim of the paper is to exploit multiple sources for object recognition when additional modalities are not present in the labelled training data.



- Classifier improves over the original one by effectively making use of all the available modalities while avoiding the burden of manually labelling new examples.

Hallucination

Inferring missing modalities in the labelled dataset to use them in conjunction with the old modalities in a probabilistic multiple kernel learning framework to improve the classifier's performance.

"Hallucinating" the missing modalities

- $X = [X^1 \ldots \ldots X^D]$ be the set of training inputs for the D modalities.
- $\bar{X} = [\bar{X}^1 \ldots \ldots \bar{X}^D]$ be the set of test inputs for the D modalities.
- $\bar{Z} = [\bar{Z}^1 \ldots \ldots \bar{Z}^M]$ be the set of M new modalities in test input.

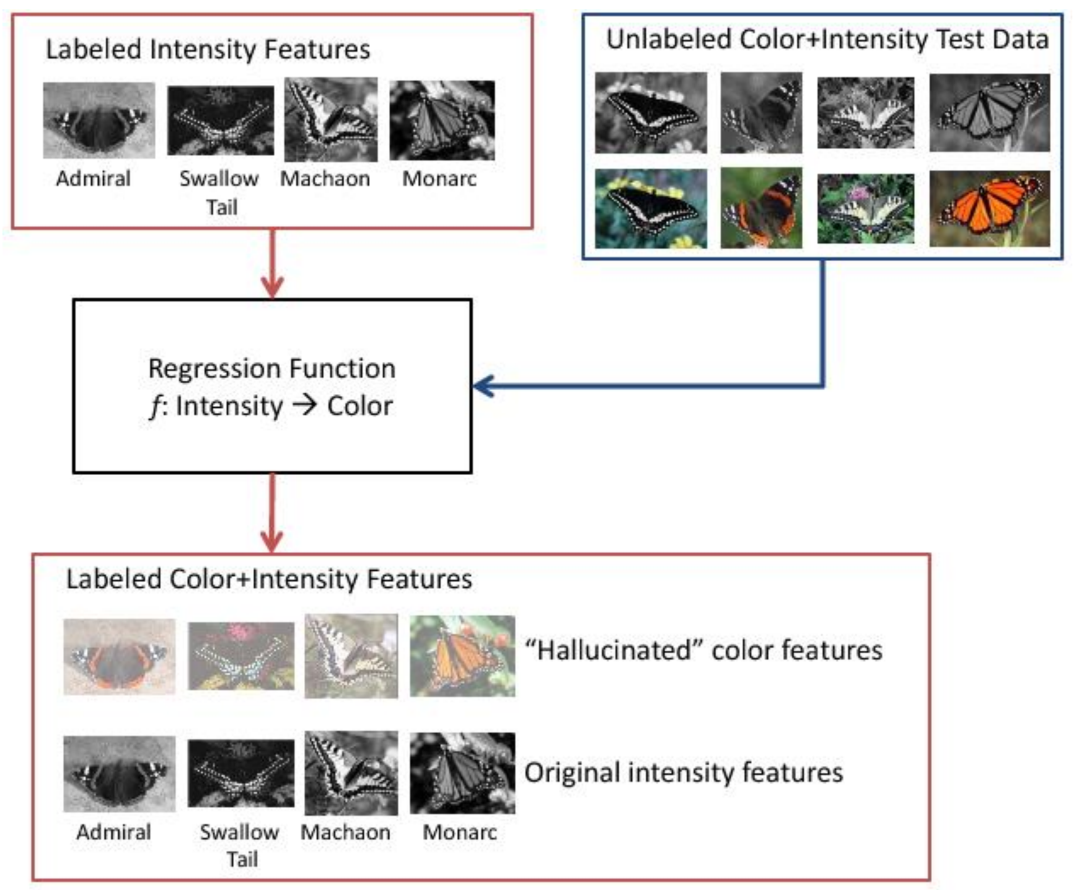
$$p(\bar{\mathbf{Z}}|\bar{\mathbf{X}}) = \prod_{m=1}^{M} \prod_{i=1}^{S_m} p(\bar{\mathbf{Z}}_{:,i}^{(m)}|\bar{\mathbf{X}})$$

$$= \prod_{m=1}^{M} \prod_{i=1}^{S_m} \mathcal{N}(\bar{\mathbf{Z}}_{:,i}^{(m)}; 0, \mathbf{K}^x)$$

where S_m is the dimensionality of the m^{th} new modality and

$$\mathbf{K}_{i,j}^x = \sum_{m=1}^{D} \alpha_m k^x(\bar{\mathbf{x}}_i^{(m)}, \bar{\mathbf{x}}_j^{(m)})$$

Where α_m is the hyper-parameter of the model.



Given the known modalities x_i for a labeled example, the predictive distribution under the Gaussian process is also Gaussian so,

$$p(\mathbf{z}_i|\mathbf{x}_i, \bar{\mathbf{X}}, \bar{\mathbf{Z}}) = \mathcal{N}(\mu(\mathbf{x}_i), \sigma(\mathbf{x}_i))$$

$$\mu(\mathbf{x}_i) = \mathbf{k}_i^x(\mathbf{K}^x)^{-1}\bar{\mathbf{Z}}$$

$$\sigma(\mathbf{x}_i) = k^x(\mathbf{x}_i, \mathbf{x}_i) - \mathbf{k}_i^x(\mathbf{K}^x)^{-1}\mathbf{k}_i^{xT}$$

Where k_i^x is the vector obtained from kernel function evaluation b/w $\bar{X}$ and x_i

Probabilistic framework based on old and new modalities

Probabilistic multiple kernel learning

To exploit all available sources of information for classification, we combine the hallucinated modalities and the old ones within a probabilistic multiple kernel learning framework.

$$\mathbf{K}_{i,j} = \mathbf{K}_{i,j}^x + \sum_{m=1}^{M} \beta_m k^z(\mathbf{z}_i^{(m)}, \mathbf{z}_j^{(m)})$$

Representation of the novel modalities

To insure that the MKL algorithm is able to exploit complex non linear kernels and to reduce the dimensionality of the feature map we rely on Kernel Principal Component Analysis which computes a low-dimensional representation of the possibly infinite dimensional.

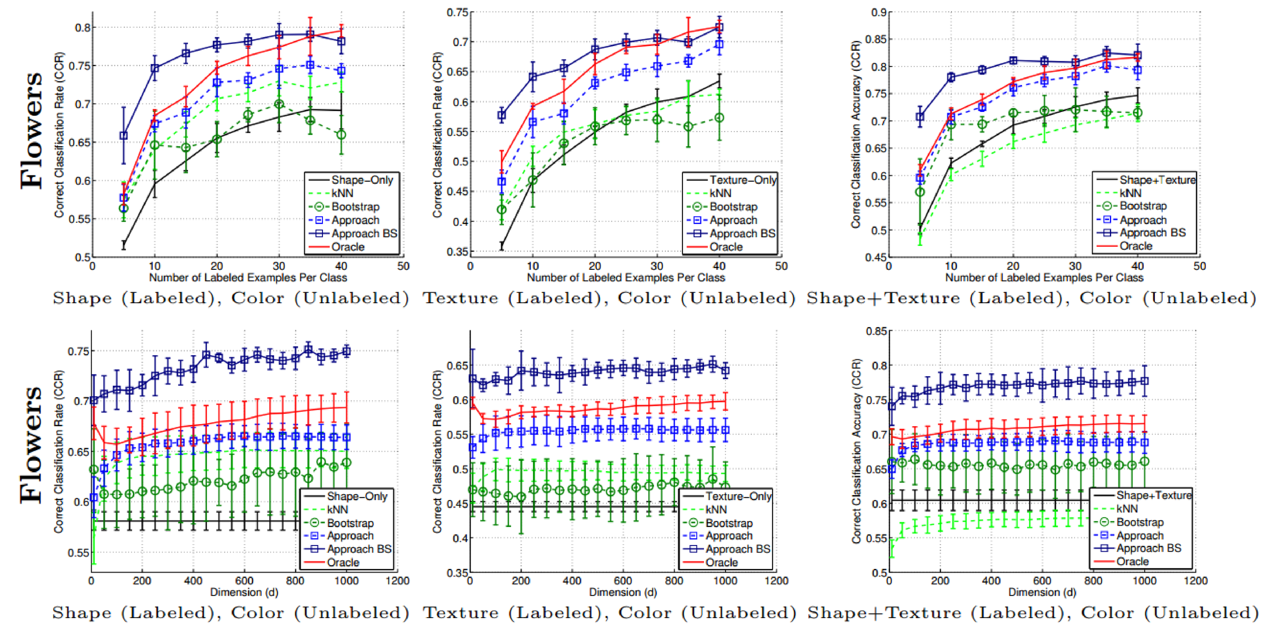
Bootstrapping

Bootstrapping strategies are employed to reduce the error of over fitting due to the unlabeled examples.

Test Case Examples

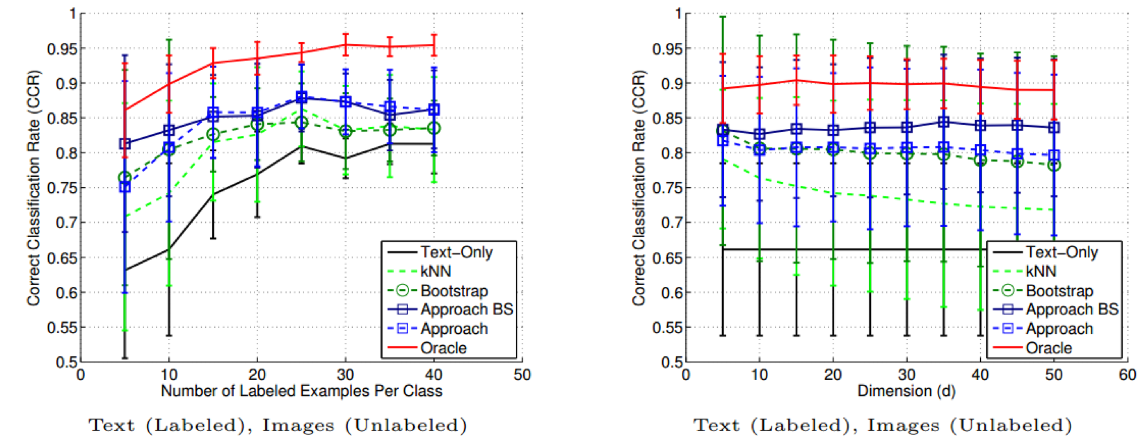
Unlabeled color images for greyscale object recognition

- Dataset: Oxford Flower Data
- Given: χ2 distance matrices over visual features- Color, Texture, Shape



Unlabeled images for sense disambiguation in text (labelled)

- Mouse- Animal or Computer Accessory?



- Image represented by dense SIFT features and text by word histogram.