Paper link:

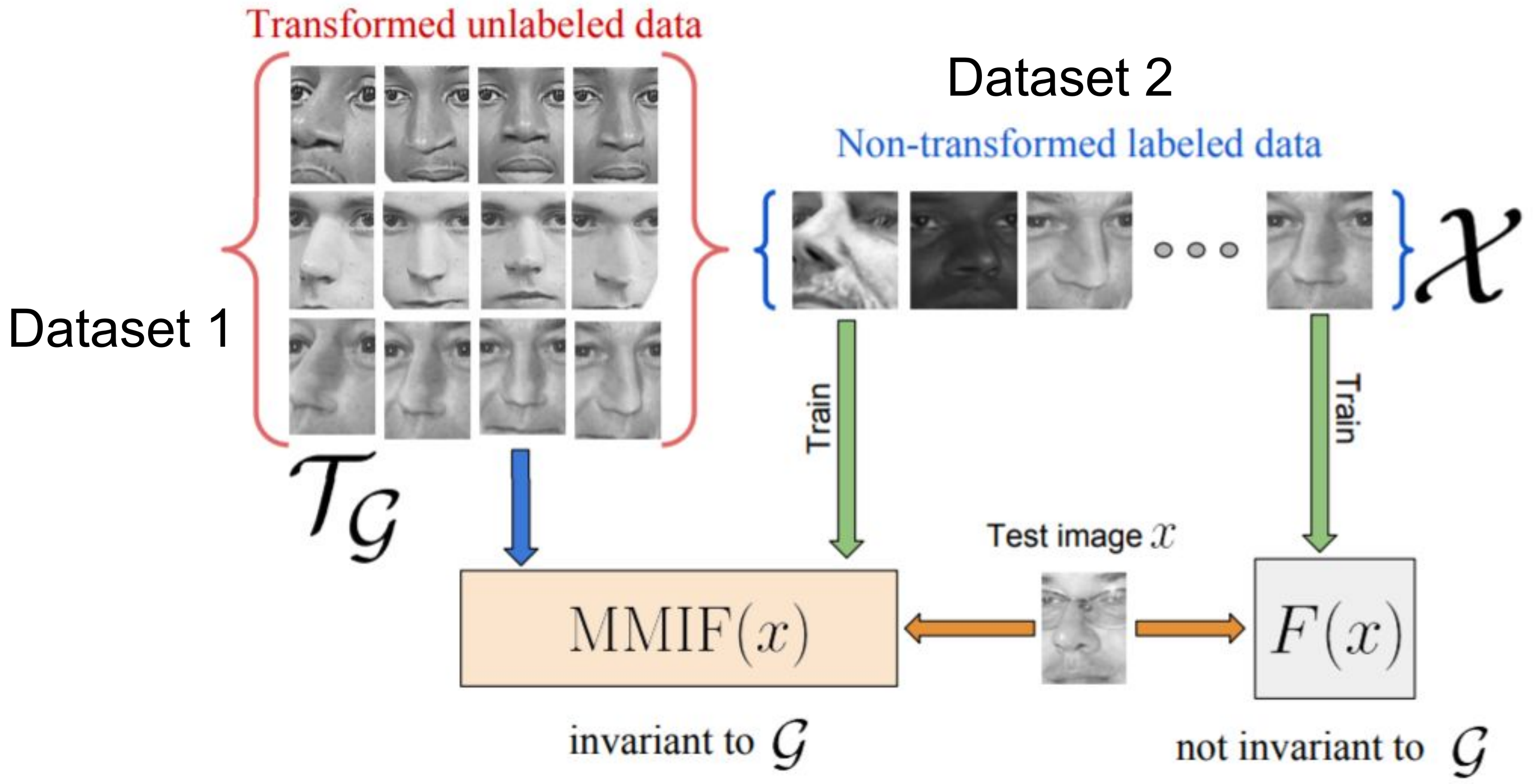
Max-Margin Invariant Features From Transformed Unlabeled Data

Dipan K. Pal, Ashwin A. Kannan*, Gautam Arakalgud* and Marios Savvides

{dipanp,aalapakk,garakalgud,marioss}@cmu.edu

The Problem

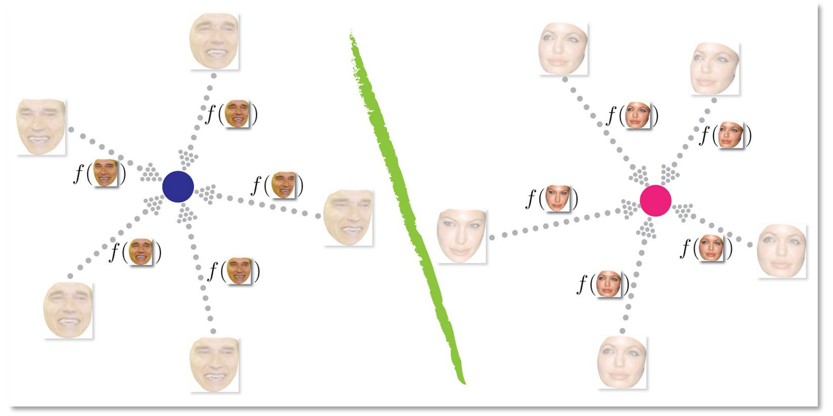
The Unlabeled Transformation Problem: Having access to transformed versions of the training unlabeled data (but not of labelled data), how do we learn a discriminative model of the labelled data, while being invariant to transformations present in the unlabeled data ?.

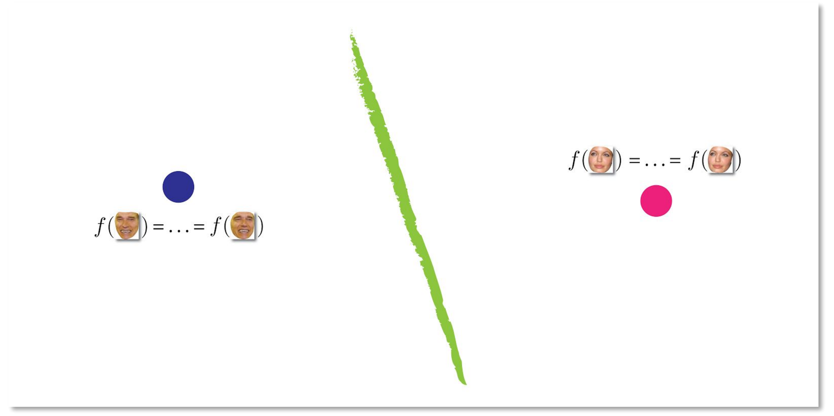


Key Idea

Unitary Group Invariant Kernels: We model transformations in data as unitary and develop unitary group invariant kernels. These invariant kernels allow us to build a set of group invariant SVMs called MMIF.

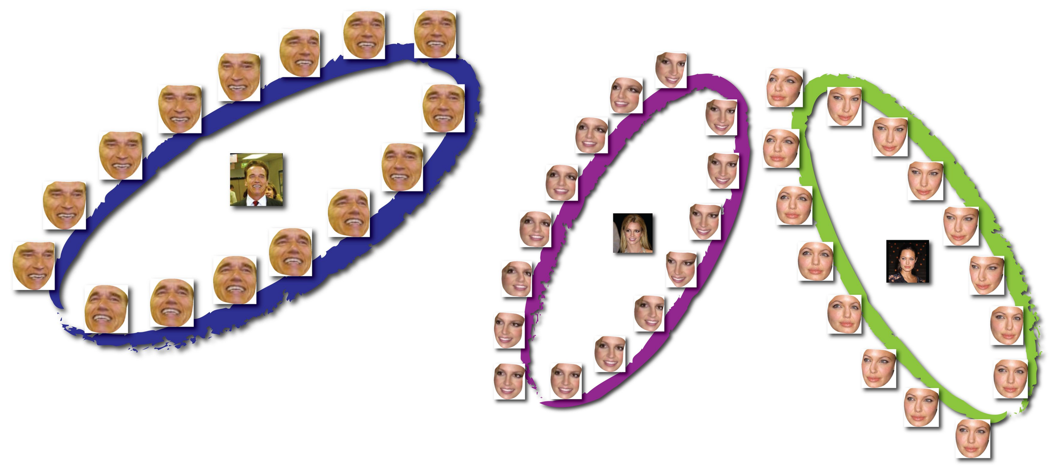
Invariance to Transformations: Nuisance transformations groups such as the translation, rotation group, increase complexity of the learning problem. Invariance to such transformations can drastically reduce complexity.



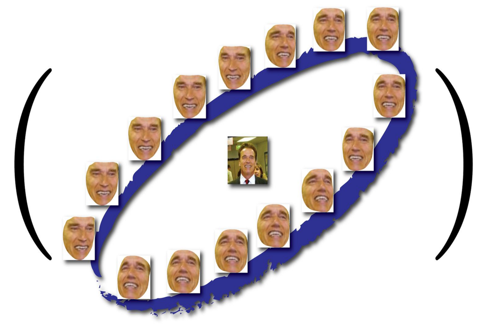


The Approach

When a group of transformations act on an object, they create an orbit.

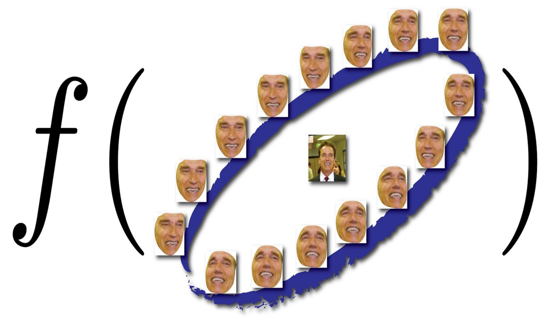


The orbit is unique to the object, and is an invariant to the transformation group



is invariant

A set of group integrated functions of the orbit is an invariant implicitly discriminative feature [1].

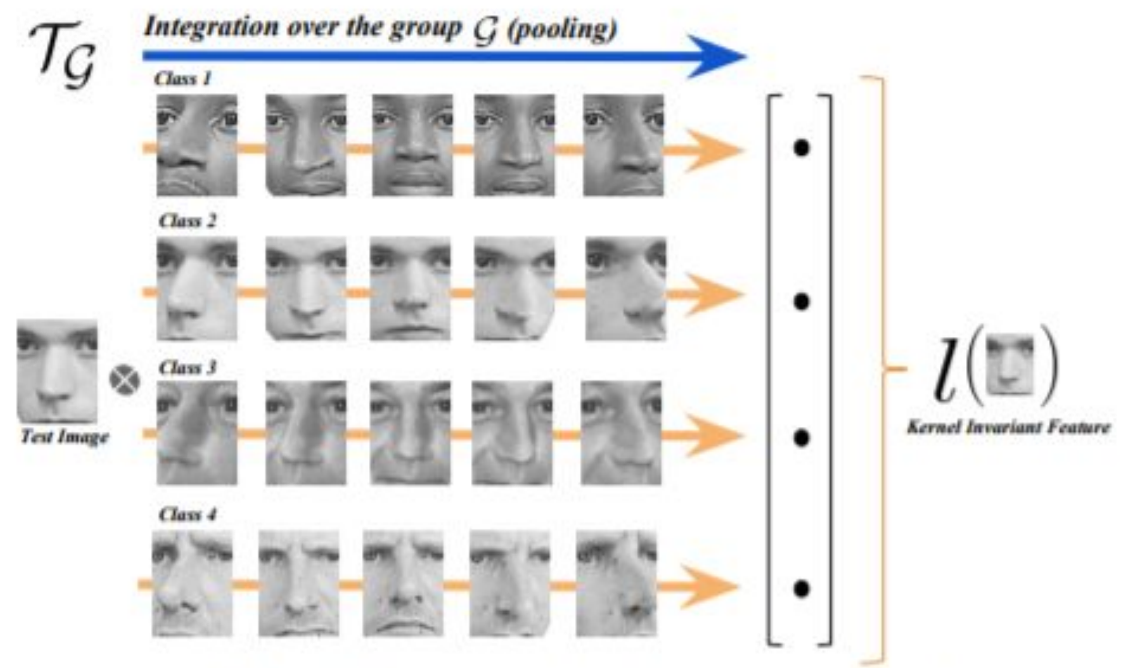


is invariant

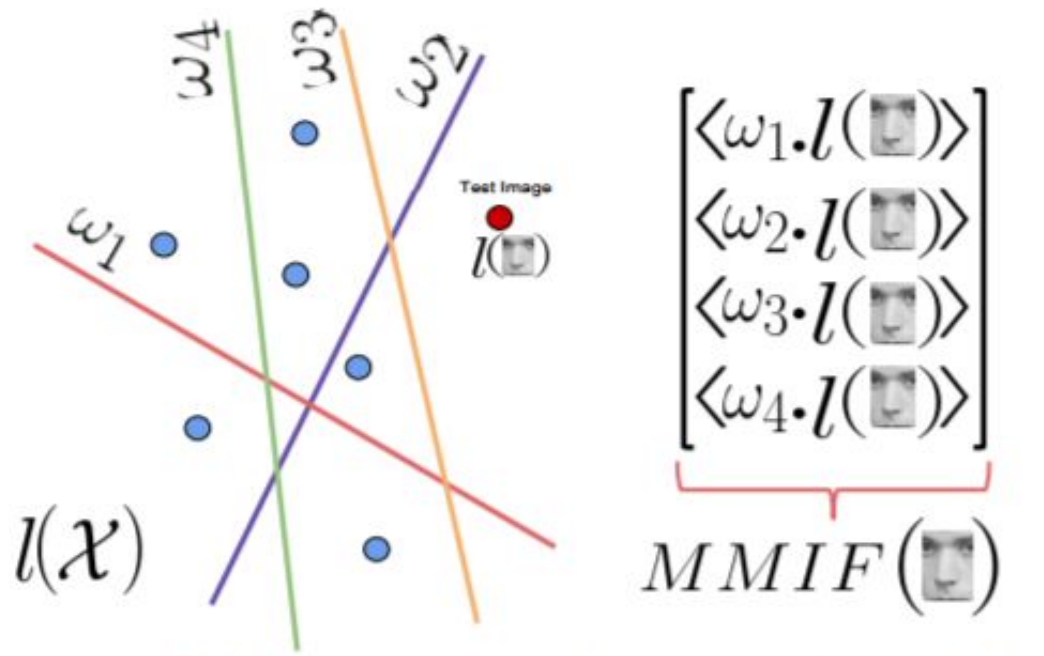
Max-Margin Discriminative Invariance: We extract group invariant features of the labeled data using kernel products and pooling. We then learn SVMs on random subsets of these features. The set of these SVMs is highly discriminative between the labeled data but invariant to transformations in the unlabeled data.

Our invariant kernel formulation through group integration allows for a SVM which is theoretically guaranteed to be invariant.

The learnt SVMs provide max-margin properties to the group invariant kernel features.



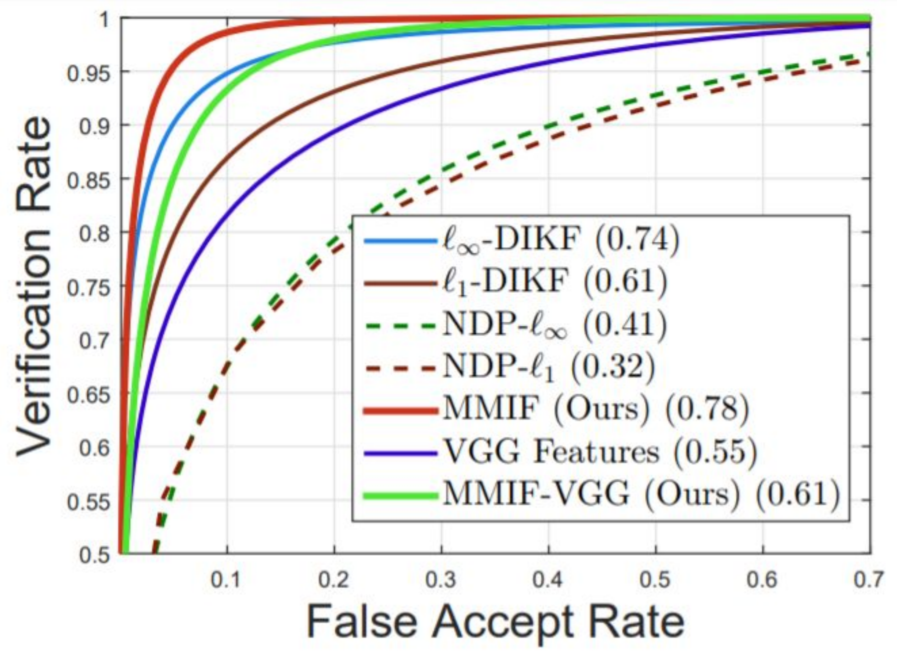
(a) Invariant kernel feature extraction

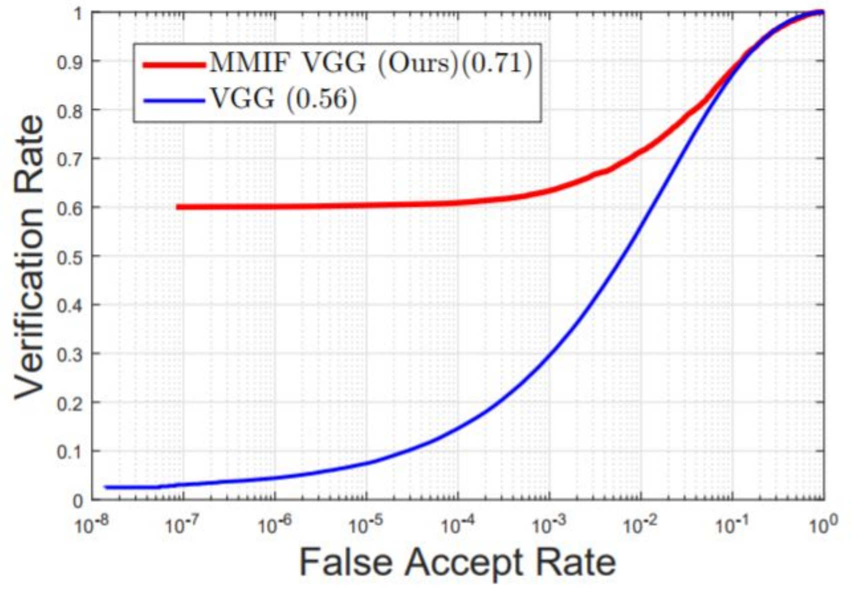


(b) SVM feature extraction leading to MMIF features

The Experiments

(1) ***Face recognition (153,000 semi-synthetic image dataset)***: 1000 subjects with 153 poses each. Images rendered from a 3D model with real texture. We compare MMIF against DIKF [2], sampled templates (NDP) [1] and a deep face recognition model (VGG-Face). We train on 38,000 images and test on 112,000 images (~13 billion matches).





(2) ***Face recognition (LFW)***: Max-pooled MMIF (in red) augments VGG features to be more discriminative within Dataset 2 while being invariant to nuisance transformations in Dataset 1. We train on 6300 images and test on 7000 images (49 million matches).

Paper link:

Carnegie Mellon University
CyLab Biometrics Center

[1] F. Anselmi, J. Z. Leibo, L. Rosasco, J. Mutch, A. Tacchetti, and T. Poggio. Magic materials: a theory of deep hierarchical architectures for learning sensory representations. MIT, CBCL paper, 2013.
[2] D. K. Pal, F. Juefei-Xu, and M. Savvides. Discriminative invariant kernel features: a bells-and-whistles-free approach to unsupervised face recognition and pose estimation. CVPR 2016