UCF
CENTER FOR RESEARCH IN COMPUTER VISION

NMF-KNN: Image Annotation using Weighted Multi-view Non-negative Matrix Factorization

Mahdi M. Kalayeh, Haroon Idrees, Mubarak Shah
{mahdi| haroon| shah } @eecs.ucf.edu , *Center for Research in Computer Vision @ UCF*

IEEE 2014 Conference on Computer Vision and Pattern Recognition
CVPR 2014

Problem

- Assigning relevant tags to query images based on their visual content

Challenges

- Finding the most relevant tags among many possible ones.
- There are tags that do not occur frequently in the dataset.
- Images that share many tags may conceptually be very different.

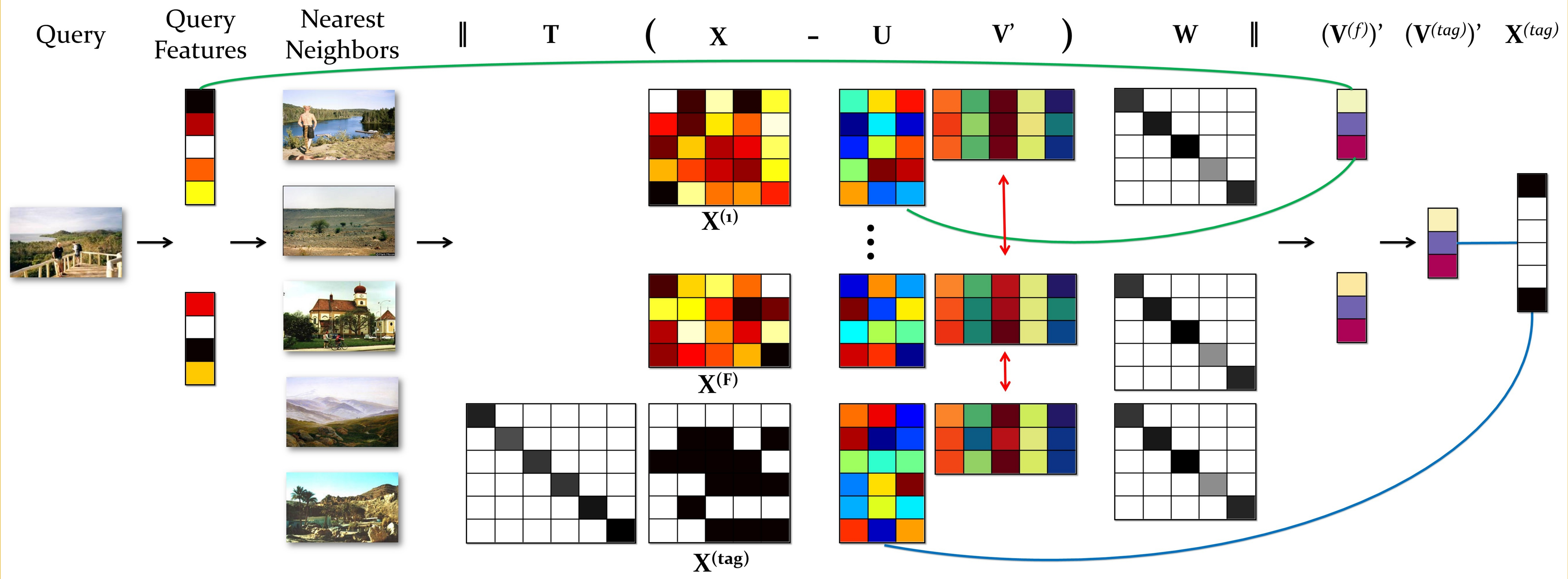
Drawbacks of Existing Methods

- Addition of images and tags requires retraining the models.
- ad-hoc feature fusion approaches are usually taken.

Our Contributions

- A query-specific model (no global training!)
- A natural solution to feature fusion
- Handling dataset imbalance through weighted NMF formulation
- $O(n)$ test-time complexity
- Straightforward extension for sub-linear test-time complexity

Proposed Approach



Query-specific Training

- Minimizing L via an iterative alternative approach (U and V are unknown)
- Training finds the optimum U that minimizes L.
- T penalizes inaccurate matrix factorization severely for rare tags.
- W is to bias the learning towards a more accurate factorization of images with rare tags.

$$L = \sum_{f=1}^{F+1} \left\| T\left(X^{(f)} - U^{(f)} V'^{(f)}\right) W \right\|_F^2 + \sum_{f=1}^{F+1} \lambda_f \left\| W'\left(V^{(f)} Q^{(f)} - V^*\right) \right\|_F^2 \quad s.t. \quad \forall\, 1 \le f \le F+1, \quad U^{(f)}, V^{(f)}, V^* \ge 0$$

$$Q^{(f)} = Diag\left(\sum_{m=1}^{M} U_{m,1}^{(f)}, \sum_{m=1}^{M} U_{m,2}^{(f)} \cdots \sum_{m=1}^{M} U_{m,K}^{(f)} \right)$$

Recovering Tags of Query (Testing)

1. Project query's feature vectors on corresponding basis matrices U
2. Approximate $V^{(tag)}$ of query by averaging over F different $V^{(visual\ features)}$
3. Predict score of different tags by computing $U^{(tag)} \times (V^{(tag)})'$
4. Select relevant tags with the highest scores

Experimental Results

- Datasets: Corel5K and ESP Game
- Evaluation metrics: Precision, Recall and N+

Method	P	R	F_1	$N+$
CRM[17]	16	19	17.3	107
InfNet[23]	17	24	19.9	112
NPDE[37]	18	21	19.3	114
SML[2]	23	29	25.6	137
MBRM[7]	24	25	24.4	122
TGLM[20]	25	29	26.8	131
JEC[22]	27	32	29.2	139
TagProp-ML[15]	31	37	33.7	146
TagProp-σML[15]	33	42	36.9	160
Group Sparsity[38]	30	33	31.4	146
FastTag[3]	32	43	36.7	166
NMF-KNN	38	56	45.2	150

Table 1. Performance evaluation on Corel5k dataset

Method	P	R	F_1	$N+$
MBRM[7]	18	19	18.4	209
JEC[22]	22	25	23.4	224
TagProp-σSD[15]	39	24	29.7	232
TagProp-ML[15]	49	20	28.4	213
TagProp-σML[15]	39	27	31.9	239
FastTag[3]	46	22	29.7	247
NMF-KNN	33	26	29.0	238

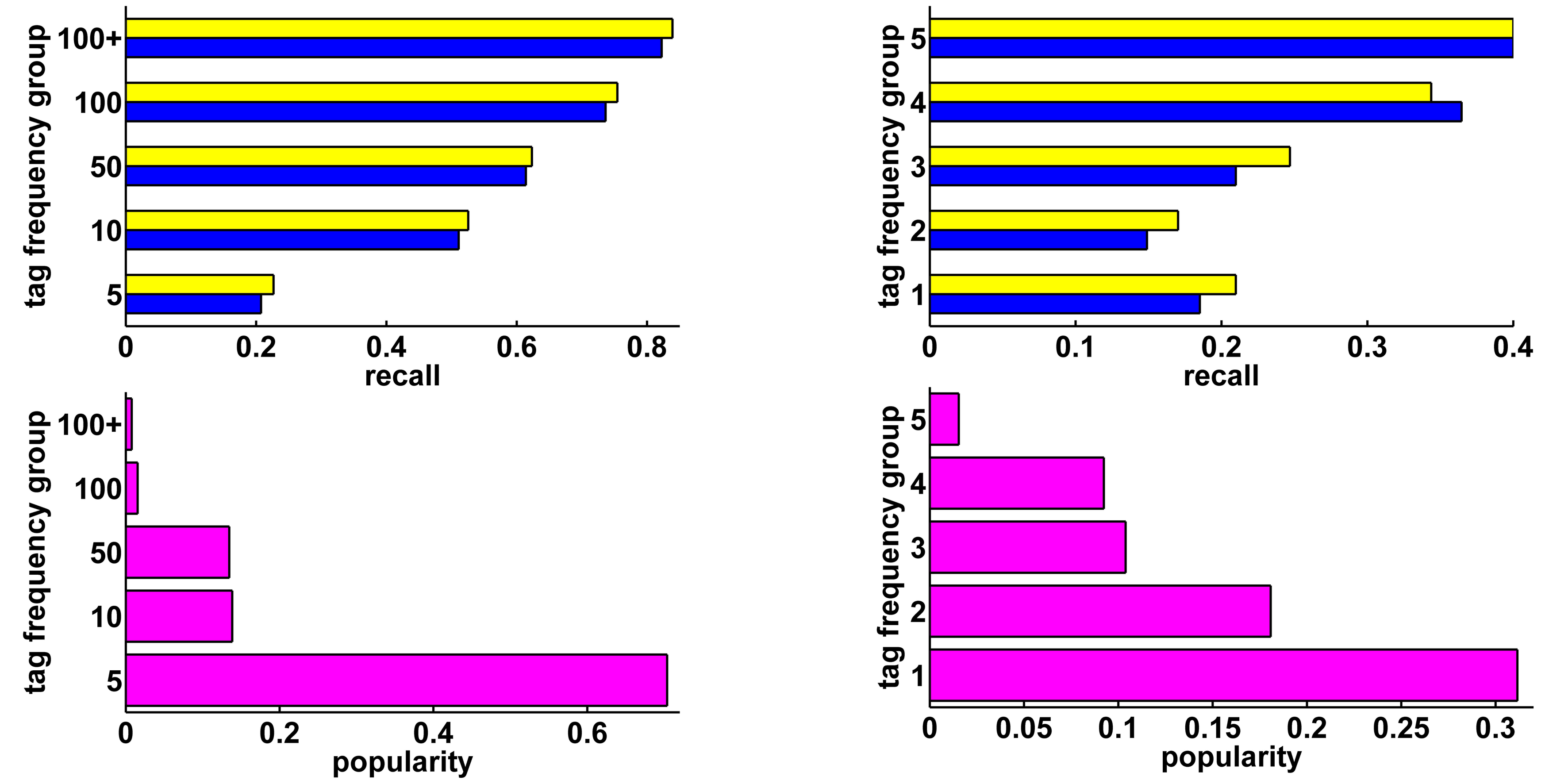
Table 2. Performance evaluation on ESP Game dataset

Qualitative Results

Predicted tags in **green** appear in the ground truth while **red** ones do not.

Effect of Weight Matrices (W and T)

with weights
without weights



Acknowledgment: This work was supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center contract numbers D11PC20066. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the U.S.Government.