Adding Unlabeled Samples to Categories by Learned Attributes

CVPR 2013 Portland, Oregon June 23-28

Jonghyun Choi   Mohammad Rastegari   Ali Farhadi[W]   Larry S. Davis
http://umiacs.umd.edu/~jhchoi/addingbyattr
UMIACS · University of Maryland · University of Washington

THIS RESEARCH IS PARTLY SUPPORTED BY ONR MURI GRANT NO. N00014-10-1-0934
https://wiki.cs.umd.edu/cvss/

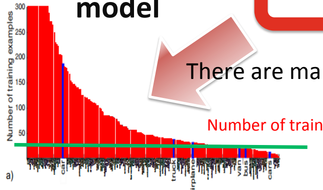
Problem

To obtain a better classifier

Better classification model

More training data

There are many categories that have a few training samples



Number of training samples in SUN 09 Dataset[1]

How to Add Samples

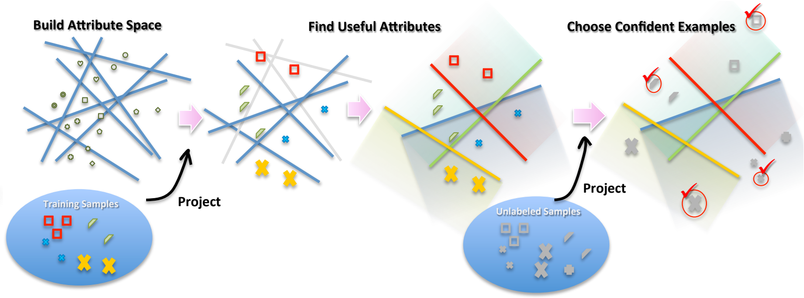
- Active learning
  - Require **human in the loop**
- Semi-supervised learning
  - Unlabeled samples are assumed to **follow the same distribution of labeled samples**

Our Solution

By learning data driven attributes based on [2]

- **No human required**
- **No underlying distribution assumed**

Approach Overview



Formulation

Joint optimization for discovering discriminative attributes and unlabeled samples

$$\min_{I_c \in I, w_c^v, w_c^a} \sum_c \left(\alpha J_c^v(I_c, w_c^v) + \beta J_c^a(I_c, w_c^a) \right) + M(I)$$

subject to

Max-margin Classifier on visual feature space

$$J_c^v(I_c, w_c^v) = \|w_c^v\|_2^2 + \lambda_v \sum_{i=1}^{n} \xi_{c,i}$$

$$I_{c,i} \cdot y_{c,i}(w_c^v x_i) \geq 1 - \xi_{c,i}, \quad \forall i \in \{1, \dots, n\}$$

Attribute Mapper

$$J_c^a(I_c, w_c^a) = \|w_c^a\|_2^2 + \lambda_a \sum_{j=1}^{n} \zeta_{c,j} - \sum_{k=l+1}^{n} I_{c,k} \left(w_c^a \phi(x_k) \right)$$

$$I_{c,j} \cdot y_{c,j}(w_c^a \phi(x_j)) \geq 1 - \zeta_{c,j}, \quad \forall j \in \{1, \dots, n\}$$

Max-margin Classifier on attribute space

$$\sum_{k=l+1}^{n} I_{c,k} \leq \gamma, \quad I_{c,k} = 1, \ \forall k \in \{1, \dots, l\}$$

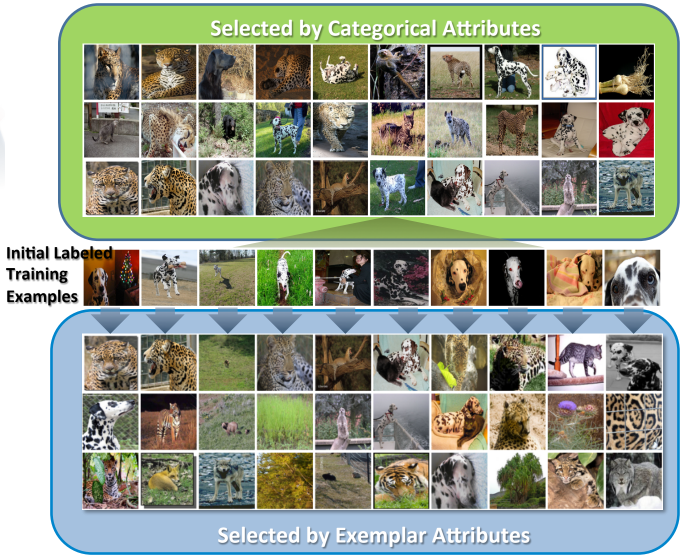
Top-Lambda Selector

$$M(I) = \sum_{c1 \neq c2} \sum I_{c1} \cdot I_{c2}.$$

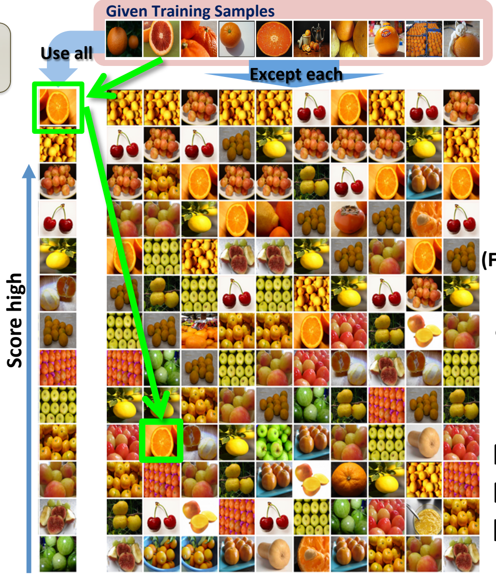
Mutual Exclusion

- Non-convex: I is discrete, w's are continuous
- Solve by block coordinate descent

Adding by Two Kinds of Attributes



How to Select by Exemplar-Attributes



Retrieval rank change by the absence of a training sample

(Full-set vs Leave-one-out Set)

$$e_j(x_i) = \frac{\mu}{r_g(x_i)} - \frac{\nu}{r_j(x_i)}$$

More stable than Exemplar-SVM w/o large negative set

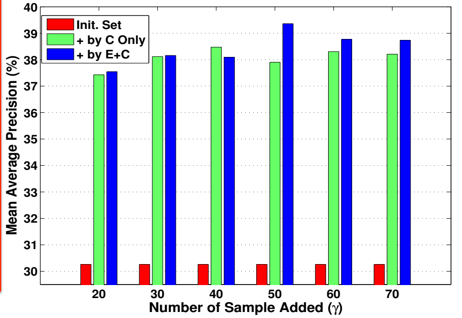
Experimental Results

Comparison with other methods

Category Name	Init.	NN	ALC	Cat.	E+C
Mashed Potato	45.03	34.02	51.15	61.39	63.92
Orange	29.84	16.29	26.97	40.61	41.05
Lemon	32.21	27.58	32.43	35.37	34.23
Green Onion	25.06	16.50	19.66	38.57	40.20
Acorn	13.09	11.05	15.41	19.35	20.10
Coffee bean	58.29	43.89	56.62	64.65	66.54
Golden Retriever	14.54	15.57	12.61	17.54	18.61
Yorkshire Terrier	29.62	13.62	27.63	41.41	45.65
Greyhound	15.24	15.73	15.64	14.75	15.22
Dalmatian	43.84	27.97	37.91	54.42	57.23
Miniature Poodle	26.10	12.50	21.16	28.87	30.21
Average	30.26	21.34	28.84	37.90	39.36

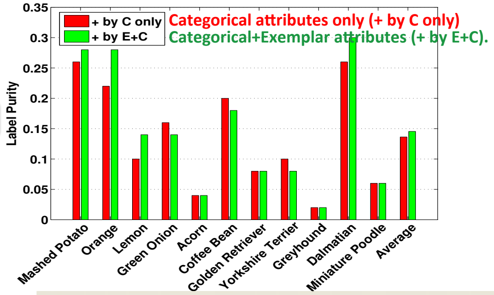
- 'Init.': initial labeled training set.
- 'NN': addition by 'nearest neighbor' in visual feature space
- 'ALC': addition by 'active learning criteria (ALC)' that finds the examples close to the current decision hyper-planes
- 'Cat.': our method of select examples using categorical attributes only.
- 'E+C': addition using both categorical and exemplar attributes.

Varying number of added images

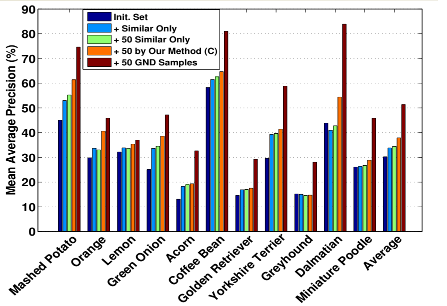


- Red: Using the initial labeled set (Init. Set)
- Green: The augmented set by our method using categorical attributes only (+ by C only)
- Blue: Categorical+Exemplar attributes respectively. (+ by E+C)

Purity of added examples

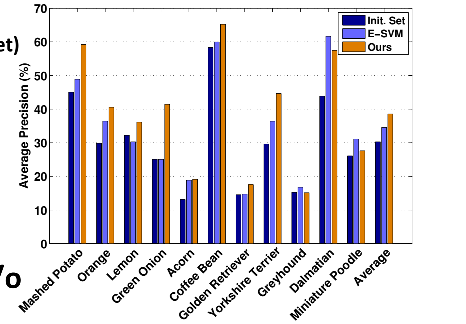


Low Purity still Improves Accuracy



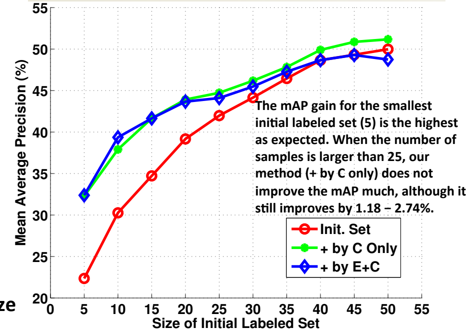
- Navy: using the initial labeled set (baseline).
- Blue: using only similar examples among the selected 50 examples.
- Green: using 50 similar examples to compare with the result of our selected 50 examples
- Orange: including both similar and exact examples.
- Red: using a set of 50 ground truth images, which is the best achievable accuracy (upper bound).

Our exemplar attribute discovery method (Sec. 4.2) to Exemplar SVM[3]



Our method outperforms the exemplar SVM in terms of category recognition accuracy by APs without the extra large negative example set (size = 50,000).

Size of Initial Labeled Set



[1] Salakhutdinov, Torralba, Tenenbaum, "Learning to Share Visual Appearance for Multiclass Object Detection", CVPR 2011 | [2] Rastegari *et al.*, "Attribute Discovery via Predictable Discriminative Binary Codes", ECCV 2012 | [3] Malisiewicz, Gupta, Efros, "Ensemble of Exemplar-SVMs for Object Detection and Beyond", ICCV 2011