Cross-generalization: learning novel classes from a single example by feature replacement

Presented by: Goo Jun

CS395T Object Recognition
Spring 2007

Motivations

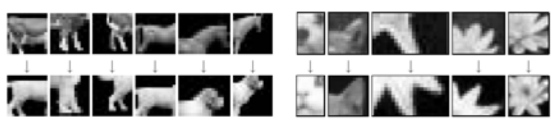
- Object classification methods generally require large number of training examples.
- Constructing training set is very time-consuming task, and it costs a lot. (time == money)
- Do we need many training examples for ALL classes? There exist a lot of classes in real world, and it is not possible to collect training images for all of them.
- Humans can distinguish >>10k classes

Overview

- Develop a feature-based object classification model that can learn a novel class from a single training example.
- Experience from already learned classes facilitate the learning and constrain overfitting.
- Features for a novel class are obtained by adapting the features from similar familiar (already learned) classes.
- Cross-generalized model outperforms the stand-alone algorithm on a large dataset with 107 classes of objects.

Introduction

- **Hypothesis :** A feature is likely to be useful for a novel class (e.g. dogs) if a similar feature proved effective for a similar class (e.g. cows).
- Assume a sufficient number of training examples are available for a set of object classes (say, cows, horses, and flowers) to extract suitable discriminating features.
- These classes are referred to as "known" or "familiar" classes.
- The objective is to learn a new class, say dog, from a single example.
    - Challenge : Obtain suitable features (restrict overfitting).
    - Proposed solution : Adapt the features from similar familiar classes.



Related Works

- Data Manufacturing
    - Pros : Can significantly improve classification when generative model is available.
    - Cons : Constructing generative models which reflect natural variations of visual objects is very difficult.
- Perona *et al.* proposed a parametric class model and obtains a prior for parameters for a novel class based on the examples of familiar class.
    - Pros : Avoid inaccurate parameter estimate and increases performance compared to no prior.
    - Cons : A single prior is used for all novel classes, hence biases novel class parameters towards frequently appearing familiar class.
- Freeman *et al.* proposed feature sharing between classes.
    - Pros : Reduces the total number of representative features.
    - Cons : Produces generic features (like edges) and requires simultaneous training of all classes.

Proposed algorithm obtains class-specific features and novel class features are adapted only from similar familiar class features.

Using Image Fragments as Features

- Feature Extraction
    - Extract sub-images of multiple sizes from multiple locations
    - With each fragment, its location in the original image is stored and used to determine relative locations of different fragments
    - Features are selected in a greedy manner that maximize the mutual information between the feature and the class it represents.
- For Classification
    - Set of fragments is searched for in the image, using the absolute value of normalized cross-correlation.
    - For each fragment F, the relevant locations in the image are determined by the location of F relative to the common reference frame.
    - Image patches at the relevant locations are compared with F, and the location with the highest correlation is selected. If this highest correlation exceeds a pre-determined threshold θ_F, the fragment is considered present, or active in the image

[2] E.Bart, E.Byvatov, and S. Ullman. "View-invariant recognition using corresponding object fragments." In ECCV, pages 113-127, 2002.
[17] S.Ullman, M.Vidal-Naquet, and E.Sali. "Visual features of intermediate complexity and their use in classification." Nature Neuroscience, 5(7):682-687, 202.

Classification by image fragments

- Let $f_i = 1$ or 0 indicate whether F_i exist or not in the test image.
- The classifier labels the image as belonging to the class C if

$$\sum_i w_i(f_i) > T \quad \text{where} \quad w_i(d) = \log \frac{P(f_i = d \mid C)}{P(f_i = d \mid \overline{C})}, \quad d \in \{0,1\}$$

- $P(f_i = d \mid C)$ represents the probability that fragment F_i is present in images belonging to class k.
- It is needed to invoke a new class if an image under testing does not get classified into one of the familiar classes.
- Should be able to estimate reliable features from only one example of the new class.

Cross-generalization

- A feature $\boldsymbol{F}$ is likely to be useful for class $\boldsymbol{C}$ if a similar feature $\boldsymbol{F'}$ proved effective for a similar class $\boldsymbol{C'}$.
- For an example image $\boldsymbol{E}$, each familiar fragment F_i^k is searched within $\boldsymbol{E}$ using Normalized cross-correlation.
- The location is selected with maximum cross-correlation S_i^k and a fragment F^{new} is extracted from the same location of image $\boldsymbol{E}$.
- The process continues to choose $\boldsymbol{J}$ new fragments from $\boldsymbol{E}$ that corresponds to $\boldsymbol{J}$ highest cross-correlation among all familiar fragments F_i^k. $(\boldsymbol{J} = 25)$
- The original fragments F_j^{new} extracted from $\boldsymbol{E}$ and their relative locations are used for the design of the new classifier.
- Note that features are chosen independently without incorporating any spatial correlation between them, although familiar classes offer this information.

Classification of Novel Class

- Each novel fragment F_j^{new} was nominated by some fragment F_i^k to which it was similar.
- Since F_j^{new} and F_i^k belong to different classes one needs to prevent F_j^{new} being detected in images that contain F_i^k, so a threshold higher that S_i^k is required.
- The authors set the threshold to $1.1 S_i^k$.
- A new classifier that uses the fragments F_j^{new} has the form

$$\sum_j w_j^{new}(f_j^{new}) > T^{new} \quad f_j^{new} \in \{0,1\}$$

$$\sum_i w_i^k(f_i^k) > T^k \quad \text{where} \quad w_i^k(d) = \log \frac{p(f_i^k = d \mid c = k)}{p(f_i^k = d \mid c = \overline{k})}, \quad d \in \{0,1\}$$

- The weights cannot be empirically estimates as shown before (one data point to count). The authors suggest

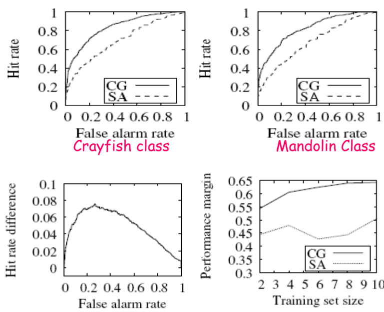
$$w_j^{new}(0) = w_i^k(0) \quad and \quad w_j^{new}(1) = w_i^k(1)$$

- Alternative : $w_j^{new}(0) = 0 \; and \; w_j^{new}(1) = 1$ offers poor performance!

Results (1)

- Classification results are compared against a stand-alone algorithm (SA). SA uses two examples for each class and two examples from a non-class for training.
- The tests are performed on a set of 107 object classes, with each class having 40 to 100 examples, along with a set of 400 non-class images.
- Leave-one-out scheme was used, where performance on each class as a novel class was tested with training done on the rest 106 classes. Test images consist of the novel class and a set of non-class images.

Results (2)



Crayfish class

Mandolin Class

Performance margin = 2(Area-0.5)

Hit rate difference = Hit rate$_{CG}$ – Hit rate$_{SA}$

Averaged over all examples

Conclusions

- A classification algorithm is proposed that is able to learn a novel class from a single example.
- New features are adapted from the similar features of familiar classes.
- The algorithm tries to mimic human cognition system. Will not work for a completely new class (having nothing common with familiar classes).
- No negative example is required to learn the new classifier (although negative examples are used to learn the familiar class discrimination).
- No spatial correlation is used to obtain features for the novel class. Scope for future research.

That's it!

Questions?