Object Recognition

Computer Vision 2023 Autumn

Exam preparations

Exam topics

- Bag-of-words algorithm
- General concepts

Bag-of-words

- Explain the main steps for BOW
- Pros/Cons of BOW
- Interpretation of BOW representation examples
- K-means clustering
- HOG features

Bag-of-words

- Explain the main steps for BOW
- Pros/Cons of BOW
- Interpretation of BOW representation examples
- K-means clustering
- HOG features

Bag-of-Words

Images


Positive (1)


Negative (0)

Feature extraction

Codebook construction

Codebook / dictionary: visual words

Bag-of-words vector encoding

Train image classifier

Image classification

Binary label: 0/1 ?

Step 1: Dictionary learning

Step 2: Bag-of-words vector encoding

Step 3: Image classification

Bag-of-words

- Explain the main steps for BOW
  - Extract features (e.x., SIFT)
  - Cluster to construct visual word dictionary (k-means)
    - Cluster center: '**visual words**'
    - **Q:** Higher vocabulary size results in better classifier performance? **A:** False.
  - Vector encoding: assign each feature to the nearest visual word
  - Build histogram: count occurrences of each visual word
  - Test image classification:
    - Given a test image → bag-of-words histogram
    - Find its nearest neighbor training histogram
    - Predict: assign it the category of this nearest training image (0/1)

Bag-of-words

- Explain the main steps for BOW
- Pros/Cons of BOW
- Interpretation of BOW representation examples
- K-means clustering
- HOG features

Bag-of-words

- Pros:
  - Flexible to geometry / deformations / viewpoint
  - Compact summary
  - Vector representation for sets
- Cons:
  - Ignores geometry
  - Background and foreground mixed
  - Optimal vocabulary formation remains unclear

Bag-of-words

- Explain the main steps for BOW
- Pros/Cons of BOW
- Interpretation of BOW representation examples
- K-means clustering
- HOG features

Bag-of-words

- Interpretation of BOW representation examples
  - **Q:** We have a dictionary of 4 visual words, and they represent a hand, a foot, a leg, and a face, respectively. Write down the typical BoW histogram of an image of a person (with all body parts visible).
  - **A:** [2, 2, 2, 1]

  - **Q:** Assume a visual word codebook containing the following 3 words: ”O”, ”X”, ”+”. Now consider bag-of-word representations of the two samples A=”O X +” and B=”X + O”. What is the difference between the bag-of-words representation between sample A and B?
  - **A:** It will not differentiate. The BoW histograms for A and B will be identical (i.e. [1 1 1] in both cases).

Bag-of-words

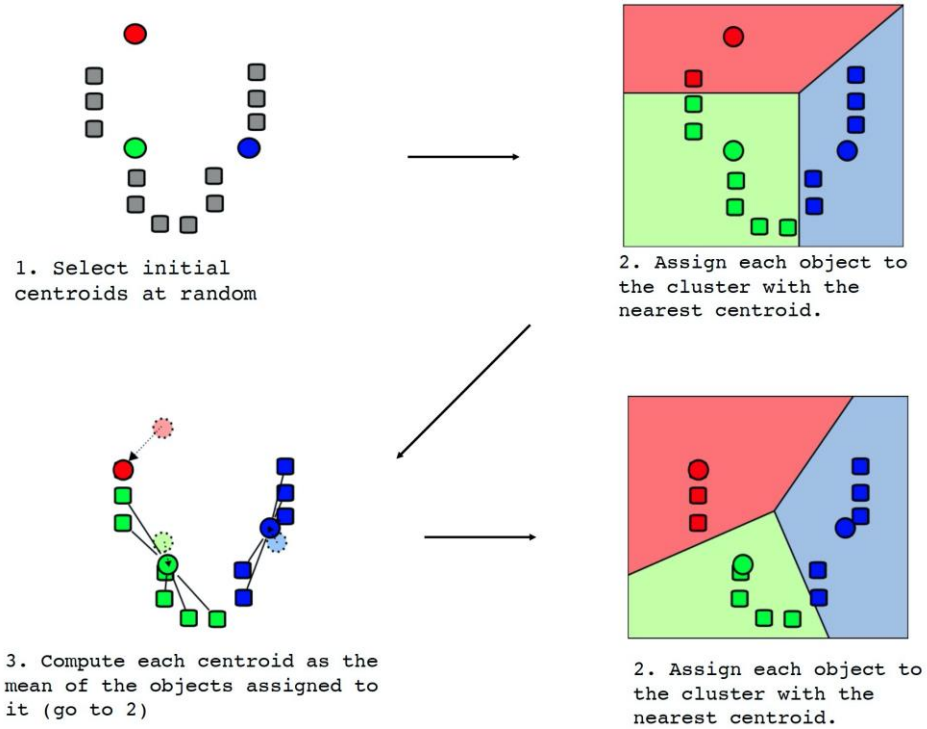
- Explain the main steps for BOW
- Pros/Cons of BOW
- Interpretation of BOW representation examples
- K-means clustering
- HOG features

Bag-of-words

- K-means clustering
  - Explain main steps
  - Note: Centers randomly initialized

  - **Q:** given a set of points, and initial centers, compute several k-means iterations

  - **Q:** How is it used in BoW image categorization?
  - **A:** K-means is used for learning the vocabulary. From all detected features in all training samples, we specify the vocabulary size K and run the K-means clustering to extract the most prominent features (visual words).



Bag-of-words

- Explain the main steps for BOW
- Pros/Cons of BOW
- Interpretation of BOW representation examples
- K-means clustering
- HOG features

Bag-of-words

- Histogram of Gradient (HOG) features
  - **Q:** If you use HOG to describe an image patch, the patch is split into h × w equally sized cells. How would you describe each cell?

    What will be the dimension of the vector describing a cell?

    What will be the dimension of the final descriptor?
  - **A:** Each cell is described by a *b*-bin (e.g. 8) histogram over the orientation of the gradients.

    Dimension is *b*, e.g. 8.

    The final dimension is *w x h x b*

General concepts

- **Q:** k-NN algorithm does more computation on test time rather than training time?
- **A:** Yes. The training phase consists only of storing the feature vectors and class labels of the training samples. During test, a test point is classified by assigning the label which are most frequent among the k training samples nearest to that query point.

- **Q:** The receptive fields get larger for deeper / further layers in AlexNet?
- **A:** True.