Comparative Deep Learning of Hybrid Representations for Image Recommendations

Paper By: C. Lei, D. Liu, W. Li, ZJ Zha, H. Li

Presenter: Muhammad Ibtesam

Student ID: 2017208879

Contribution of the Paper

- Comparative Deep Learning method
  - Dual-net used for training of the deep network
  - Pair of images are used
  - Requires more training data than naïve deep learning
  - Achieves superior performance
- Dual-net deep network
  - Map the input of image and preference of user into same latent semantic space.
  - Distance between user and image is calculated

Problem Formulation of Comparative Learning

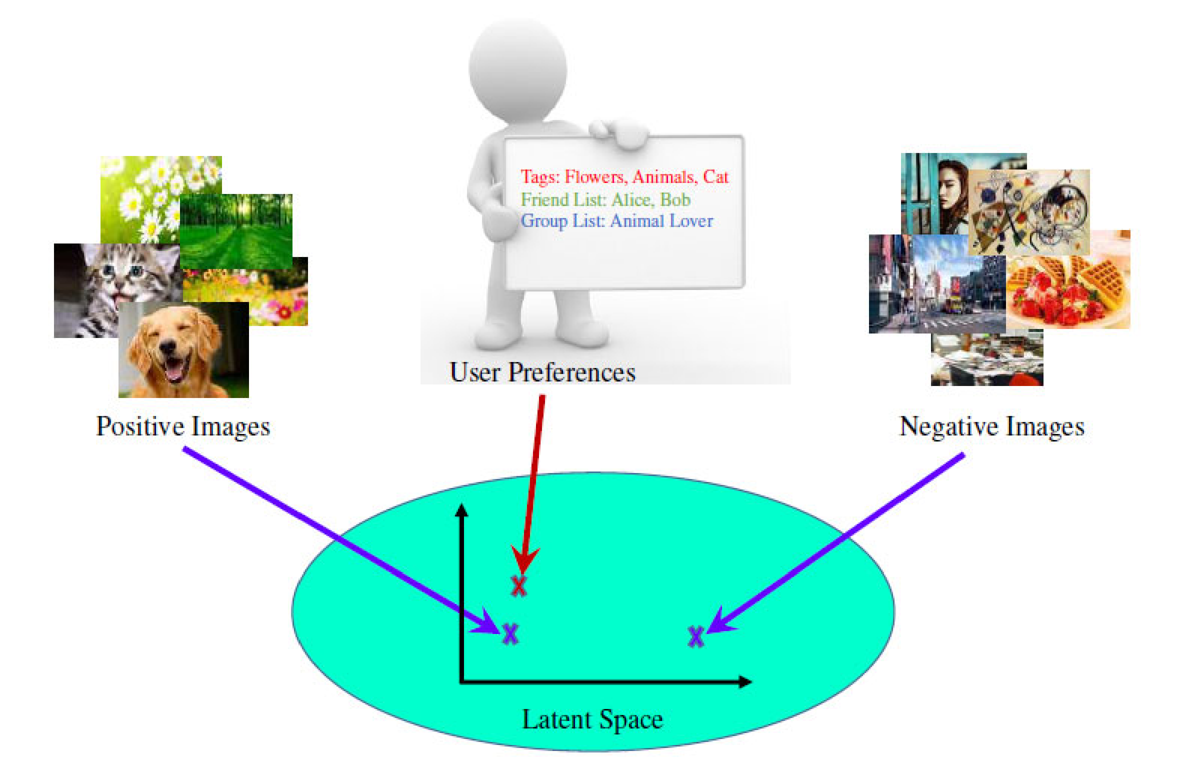
- Hybrid representation
  - Representation of both user and images
- Training Data
  - Triplets of user, positive image and negative image

$$\{\mathcal{T}_t = (U_t, I_t^+, I_t^-), t = 1, 2, ..., T\}.$$

$$D(\pi(U_t), \phi(I_t^+)) < D(\pi(U_t), \phi(I_t^-)), \forall t.$$

  - π(I) and φ(U) that map I and U to a same latent space and distance function D(·, ·)

Hybrid Representation



CONTD.

- Selection of Loss function
  - 0-1 loss function
  - Hinge loss function
  - Cross entropy loss function

$$P_{ij}^{U_t} = \frac{e^{o_{ij}^{U_t}}}{1 + e^{o_{ij}^{U_t}}},$$

$$\bar{P}_{ij}^{U_t} = \begin{cases} 0, & (i = I_t^+, j = I_t^-) \\ 1, & (i = I_t^-, j = I_t^+) \end{cases}$$

- Learning Objective

$$\min_{\pi,\phi,D} \mathcal{L}(\{\mathcal{T}_t\}) = \sum_t -\bar{P}_{ij}^{U_t} \log(P_{ij}^{U_t}) - (1 - \bar{P}_{ij}^{U_t}) \log(1 - P_{ij}^{U_t}).$$

Issues in Applying the CDL

- Preprocessing of user data as vectors
  - Word2vector technique is used.
  - Tags are converted to vectors then vectors are clustered by k means into 1024 semantic clusters
  - Tags are replaced by clusters and the bags-of-words
- Preparing triplets as training data
  - Positive images are handy
  - Negative images are not obvious.
- How to make recommendation?
  - Set of candidate images
  - Representations of these images
  - Distance is calculated
  - K nearest neighboring images as recommendations.

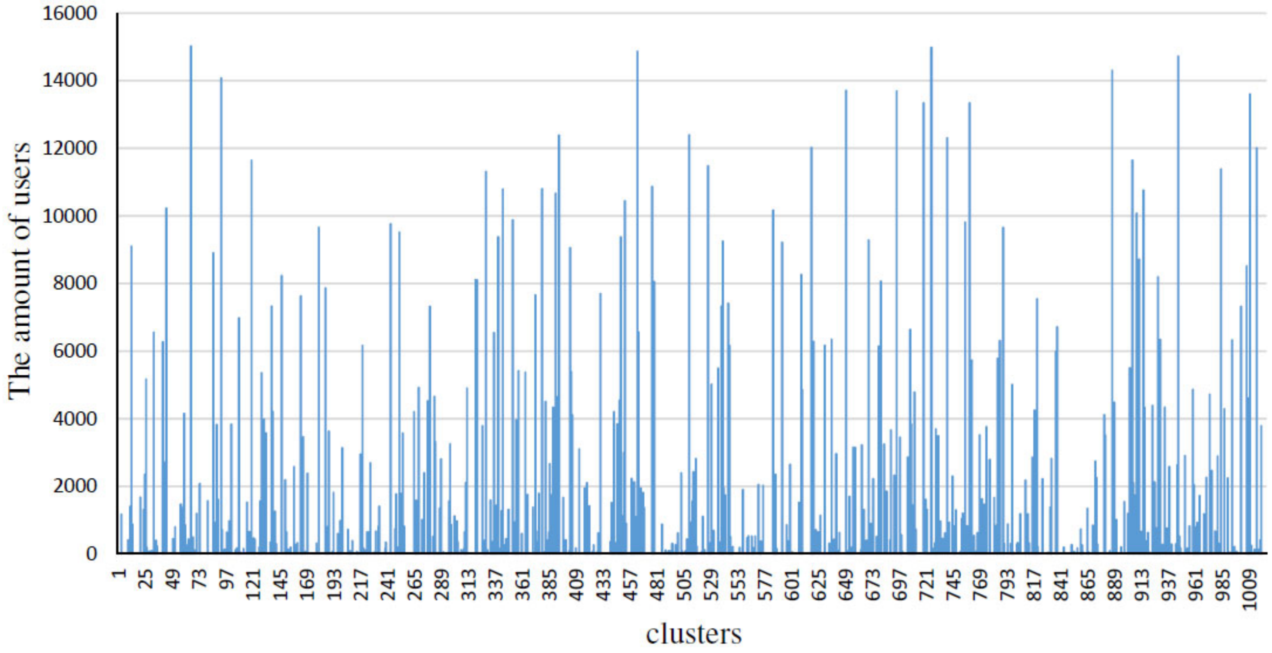
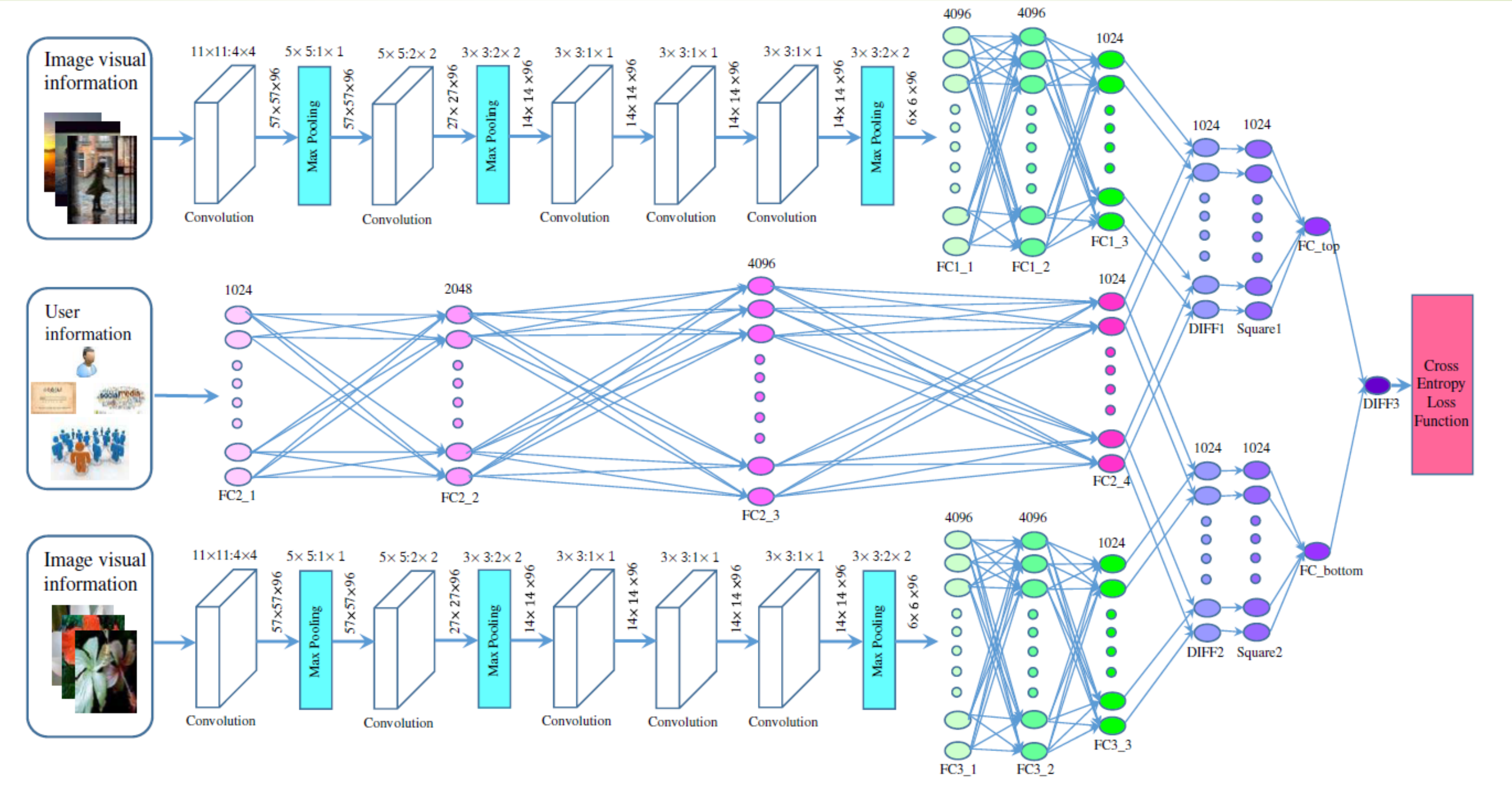


Figure 3. This figure shows the distribution of clusters. The x-axis displays 1024 clusters and the y-axis is the number of users having interests in this cluster. A user can be described by bag of words where words are indeed clusters.

Comparative Deep Learning

- Three sub-networks
  - Two convolutional neural network (CNN)
    - Capture image visual information
    - Identical configuration and shared parameters
    - One for each negative and positive image
  - Full connection neural network
    - For user information

Comparative Deep Learning



Experimental Settings

- Data Set:
  - Images from Flickr
  - 101,496 images, 54,173 users, 6439 groups and 35844 tags in this data set
  - Average 23.5 tags and 5.8 favorite images for each user.
  - filter out users that have less than 40 or more than 200 favorite images from test
  - filter out users that have interests in less than 80 or more than 280 clusters from training data
  - 8, 616 users for training and 15, 023 users for test.
  - For each user, 20 images are randomly selected from her favorite images and “concealed” for test

Compared Approaches

- Borda Count with SIDL
- Borda Count with BoW
- ImageNet
- LMNN
- Social + LMNN
- TwoNets

Implementation

- Open Source deep leaning software Caffe.
- Images resized to 256 x 256
- Dropout rate is set to 0.5
- Learning rate starts from 0.9 and momentum is 0.9.
- Mini-batch size of images is 128
- Weight decay parameter is 0.0005
- Single GeForce Tesla K20c GPU with 5GB graphical memory
- Took 4 days to finish training

Experimental results

- 20 images out of 100 candidates

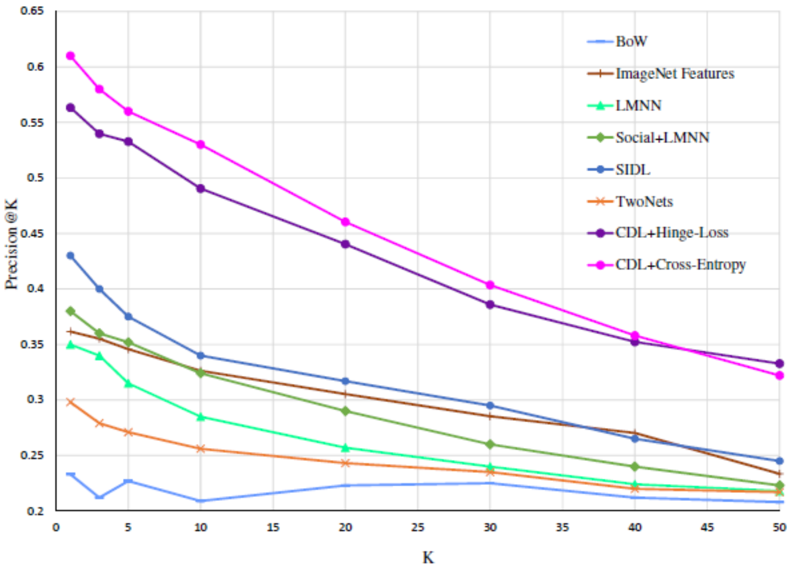


Figure 5. Precision@K for different K values of compared image recommendation methods.

CONTD.

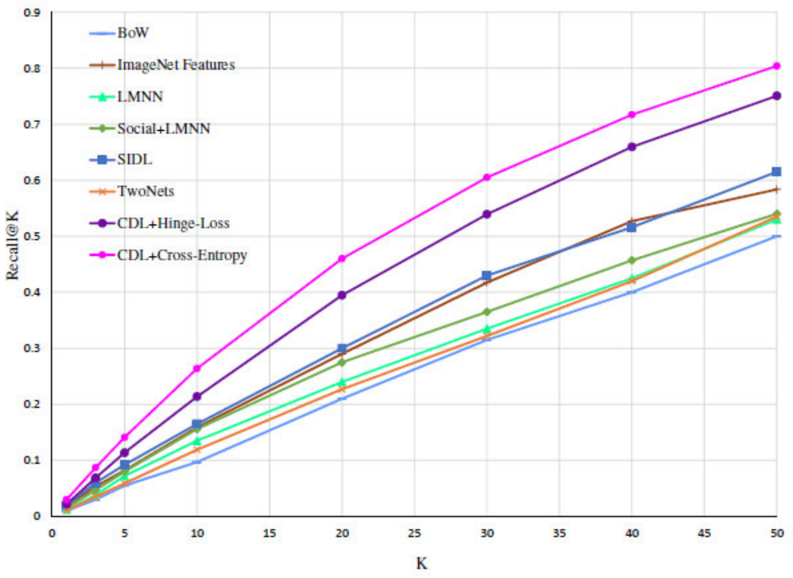


Figure 6. Recall@K for different K values of compared image recommendation methods.

Case Study

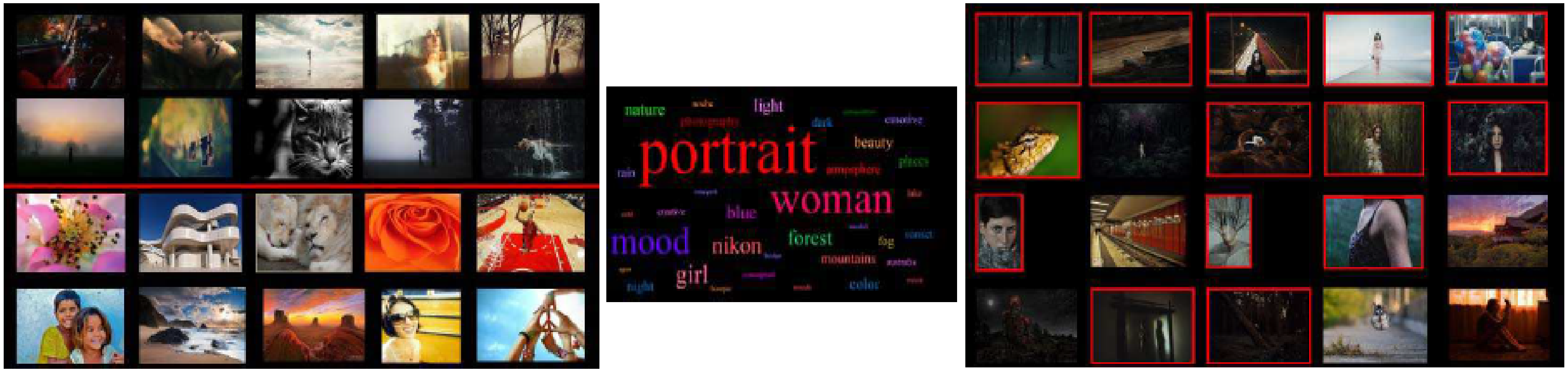


Figure 7. (Best view in color.) Case study of making recommendations to a selected user. Left: some samples of training images for this user, 10 positive and 10 negative, separated by the red line; unlike positive images that are indeed favorite images of this user, negative images are "assigned" by the process discussed in Section 5. Middle: the word cloud of this user's frequent tags retrieved from her tagging history and browsing history. Right: recommendation results sorted in relevance (ascending order of distance calculated by hybrid representations), where correct results are highlighted by red borders.

Result of User Representation

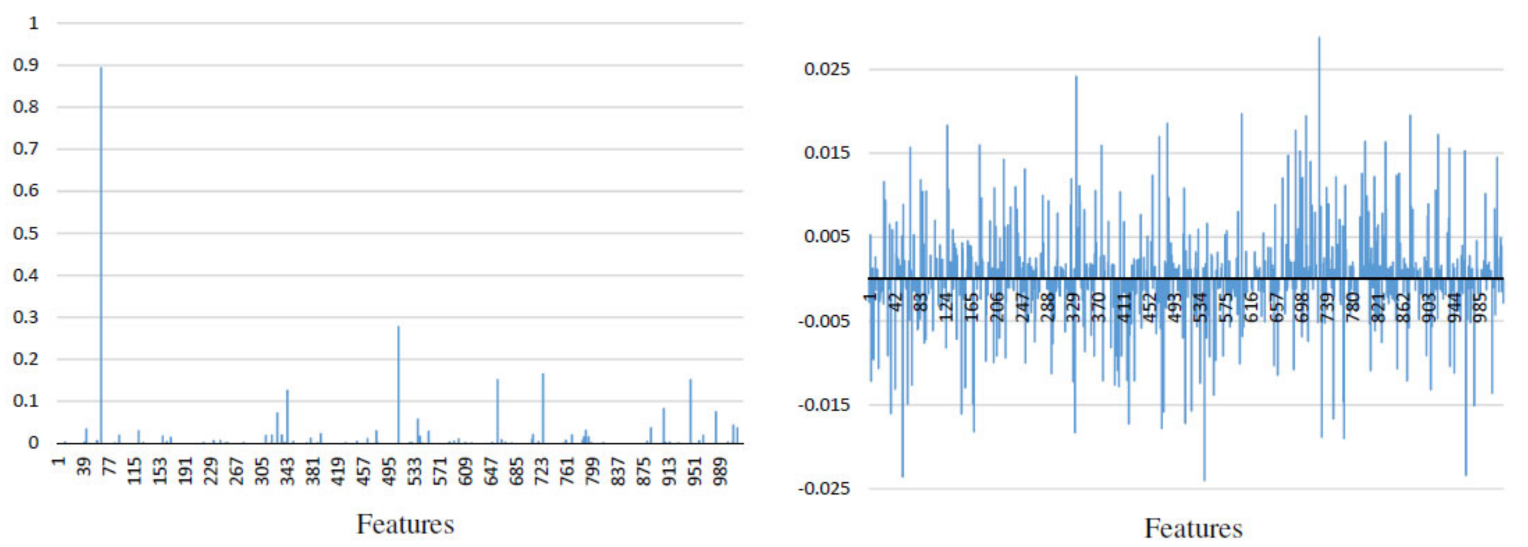


Figure 8. Exemplar input and output of the user sub-network in our designed dual-net deep network. Left: pre-processed user vector (input). Right: learnt user representation (output).

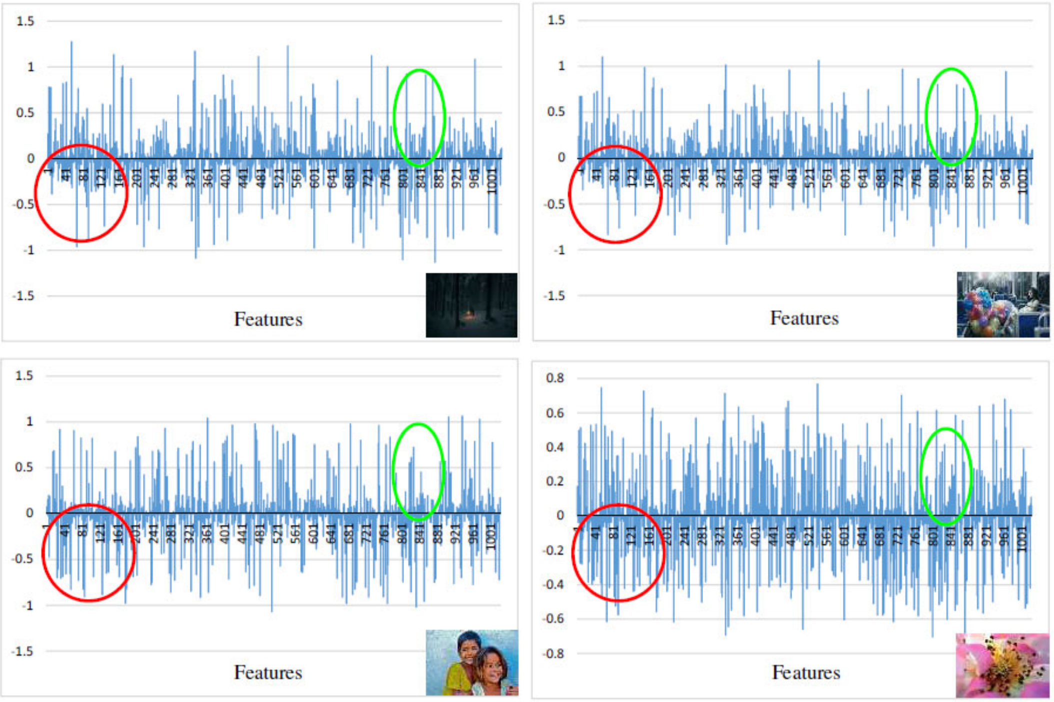


Figure 9. Exemplar learnt representations of positive images (top row) and negative images (bottom row). Note the similarity between positive images and dissimilarity between positive and negative images, especially in the circled areas.

ANY QUESTIONS?