HW2

New Attempt

Due May 7 by 10am **Points** 100 **Submitting** a file upload

HW2 is about image retrieval by matching keypoints.

Given a query image, q, and a large dataset of images, $D = \{r_i\}$, find in D a subset of K most similar images to q , denoted as $S_q(K) \subset D$. Similarity between q and $r \in D$, denoted as $s(q, r)$, is computed by matching their keypoints. Let f_k^q denote a deep feature of k-th keypoint in q, and f_l^r denote a deep feature of l-th keypoint in r. Then, we define $s(q, r)$ as a maximum total similarity of their matched pairs of keypoints:

$s(q, r) = \sum_{(f_k^q,\, f_l^r) \in M} s_{kl}$, where $s_{kl} = \frac{1}{2}\left(1 + \frac{(f_k^q)^\top f_l^r}{\|f_k^q\|_2 \|f_l^r\|_2}\right) \in [0, 1]$

In $S_q(K)$, there could be true and false positives. We say that an image $r \in S_q(K)$ is a true positive if r shows the same object as the query image q. Otherwise, r is a false positive. We will evaluate keypoint matching by estimating precision and recall of image retrieval, defined as

Precision(q, K) = $P(q, K)$ = [number of true positives in $S_q(K)$] / K

Recall(q, K) = $R(q, K)$ = [number of true positives in $S_q(K)$] / (number of true positives in D)

▸ ks:

- Download the set D of 136 images, and the set Q of 34 query images from **image_retrieval.zip.** (https://canvas.oregonstate.edu/courses/1811612/files/85435560/download?download_frd=1)
- For every image in D and Q, use your code from HW1 to:
  - Detect 20 strongest SIFT keypoints in the image;
  - Extract 20 image patches of size 32x32 pixels, centered at the detected SIFT keypoints, from the image;
- For every image in D and Q, use:
  - Your best CNN from HW1
  - our **CNN** (https://canvas.oregonstate.edu/courses/1811612/files/86781386/download?download_frd=1) with these **parameters** (https://canvas.oregonstate.edu/courses/1811612/files/85435522/download?download_frd=1)

learned on the image patches of HW1 to compute 20 deep features f of 20 32x32 image patches extracted in the previous step.

- For the two sets of deep features, and for every query image $q \in Q$
  - Match q to every image $r \in D$ by matching the deep features of their keypoints, using:
    1. **Many-to-many** matching:
       - Compute the similarity vector, $\mathbf{s} = [s_{kl}]$, of matching feature pairs (f_k^q, f_l^r) from the two images;
       - Find the binary indicator of matches, $\mathbf{x}^*$, from the following optimization problem specified for a continuous vector $\mathbf{x}$: maximize $\mathbf{s}^\top \mathbf{x}$, s.t. $\|\mathbf{x}\|_2 = 1$, and $\mathbf{x} \geq \mathbf{0}$
       - Compute the total similarity of matching: $s^{\text{many2many}}(q, r) = \mathbf{s}^\top \mathbf{x}^*$
    2. **One-to-one** matching:
       - Compute the cost matrix, $C = [c_{kl}]$, of matching feature pairs (f_k^q, f_l^r) from the two images, where $c_{kl} = 1 - s_{kl}$
       - Find $X^* = [x_{kl}^*]$ that minimizes the following constrained objective: minimize trace($C^\top X$), s.t. $\forall l : \sum_k x_{kl} = 1$ and $\forall k : \sum_l x_{kl} = 1$ and every $x_{kl} \in \{0, 1\}$
       - Compute the total similarity of matching: $s^{\text{one2one}}(q, r) = \sum_{k,l} s_{kl} x_{kl}^*$
  - For K=1, 2, 3, 4
    - Identify $S_q^{\text{one2one}}(K)$ and $S_q^{\text{many2many}}(K)$, i.e., the two sets of K most similar images from D to q with respect to the two matching criteria
    - Use the ground truth, given in **image_retrieval.zip** (https://canvas.oregonstate.edu/courses/1811612/files/85435560/download?download_frd=1) , to estimate:
      - Precision: $P^{\text{one2one}}(q, K)$ and $P^{\text{many2many}}(q, K)$
      - Recall: $R^{\text{one2one}}(q, K)$ and $R^{\text{many2many}}(q, K)$
- Use the ground truth, given in **image_retrieval.zip** (https://canvas.oregonstate.edu/courses/1811612/files/85435560/download?download_frd=1) , to compute your average precision and recall:
  - Precision: $P^{\text{one2one}}(K) = \frac{1}{|Q|} \sum_{q \in Q} P^{\text{one2one}}(q, K)$ and $P^{\text{many2many}}(K) = \frac{1}{|Q|} \sum_{q \in Q} P^{\text{many2many}}(q, K)$
  - Recall: $R^{\text{one2one}}(K) = \frac{1}{|Q|} \sum_{q \in Q} R^{\text{one2one}}(q, K)$ and $R^{\text{many2many}}(K) = \frac{1}{|Q|} \sum_{q \in Q} R^{\text{many2many}}(q, K)$

What to Turn in?

A zipped folder that contains the following files:

1. (20 points) "**precision_recall.pdf**" document with two figures, where one figure should show the two many-to-many precision-recall curves $\{(R^{\text{many2many}}(K), P^{\text{many2many}}(K)) : K = 1, 2, 3, 4\}$ for the two sets of deep features, and the other figure should show the two one-to-one precision-recall curves $\{(R^{\text{one2one}}(K), P^{\text{one2one}}(K)) : K = 1, 2, 3, 4\}$ for the two sets of deep features. **Plot the curves such that the precision values are on the y-axis and the recall values are on the x-axis.**
2. (20) "**retrieval.pth**" Python file that has a 4 x 34 x 4 tensor consisting of the four matrices of your image retrieval. In the tensor, each retrieval matrix has size 34 x 4, where the 34 rows correspond to the ID numbers of the query images, and the 4 columns represent the ID numbers of the top K=4 most similar images retrieved from the dataset D. In the tensor, organize the matrices as follows. The matrix at index=1 is obtained with many-to-many matching and your CNN; at index=2 is obtained with one-to-one matching and your CNN; at index=3 is obtained with many-to-many matching and our CNN; and at index=4 is obtained with one-to-one matching and our CNN.
3. (20) "**features1.pth**" and "**features2.pth**" two python files storing tensors of deep features, where "**features1.pth**" has features obtained with your best CNN from HW1, and "**features2.pth**" has features obtained with our **CNN** (https://canvas.oregonstate.edu/courses/1811612/files/86781386/download?download_frd=1) with these **parameters** (https://canvas.oregonstate.edu/courses/1811612/files/85435522/download?download_frd=1) . The tensor in each file has size 170 x 20 x 128, for a total of 170 images where the first 34 are the query images and the remaining 136 are images from D, and for the 20 keypoints detected in each image, and for the 128-dimensional deep features that you computed.

Submit your compressed folder on Canvas before 10am on May 7, 2021.

Grading

- 30 points = If you submit the four files as described above.
- +15 points = If your recall and precision are accurately computed from the two 34 x 4 matrices of your image retrieval in "**retrieval.pth**".
- +15 points = If the four 34 x 4 matrices in "**retrieval.pth**" are accurately computed from your deep features in the 170 x 20 x 128 tensor stored in "**features1.pth**" and "**features2.pth**".
- +10 points = If your recall and precision for every K and for the features computed with our CNN (https://canvas.oregonstate.edu/courses/1811612/files/86781386/download?download_frd=1) with these parameters (https://canvas.oregonstate.edu/courses/1811612/files/85435522/download?download_frd=1) are not worse more than 5% from ours.
- +20 points = If your recall and precision for every K and for the features computed with our CNN (https://canvas.oregonstate.edu/courses/1811612/files/86781386/download?download_frd=1) with these parameters (https://canvas.oregonstate.edu/courses/1811612/files/85435522/download?download_frd=1) are not worse more than 2% from ours.
- +10 points = If your recall and precision for every K and for the features computed with our CNN (https://canvas.oregonstate.edu/courses/1811612/files/86781386/download?download_frd=1) with these parameters (https://canvas.oregonstate.edu/courses/1811612/files/85435522/download?download_frd=1) are not worse more than 0.5% from ours.

▶