NYU TANDON SCHOOL OF ENGINEERING

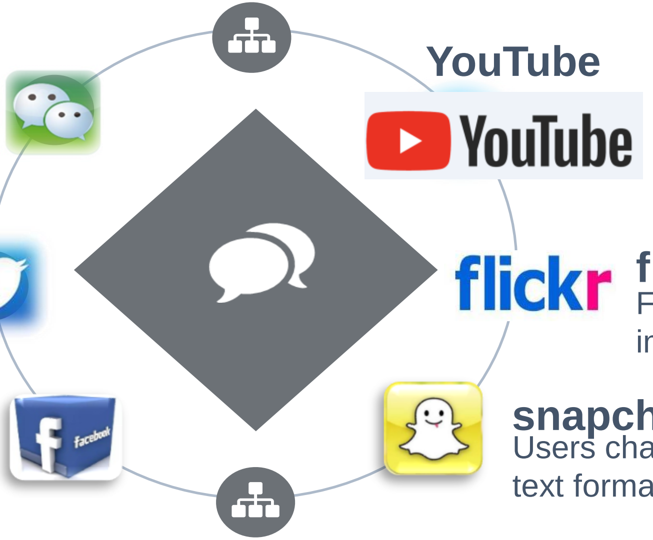
Cross-Model RETRIEVAL

EL-GY 6123

2019 fall final project

Guyu Liu
Han Jiang
Jiacheng Wang
Zhong Han

BACK GROUND



Wechat

Users share their daily life on the Wechat Moment, which include huge amount of images and texts.

Twitter

Billions of images, videos and texts are uploaded on the Twitter everyday

Facebook

Like Twitter, Billions of images, videos and texts are uploaded

YouTube

300 hours video were uploaded on the youtube every minutes

fliker

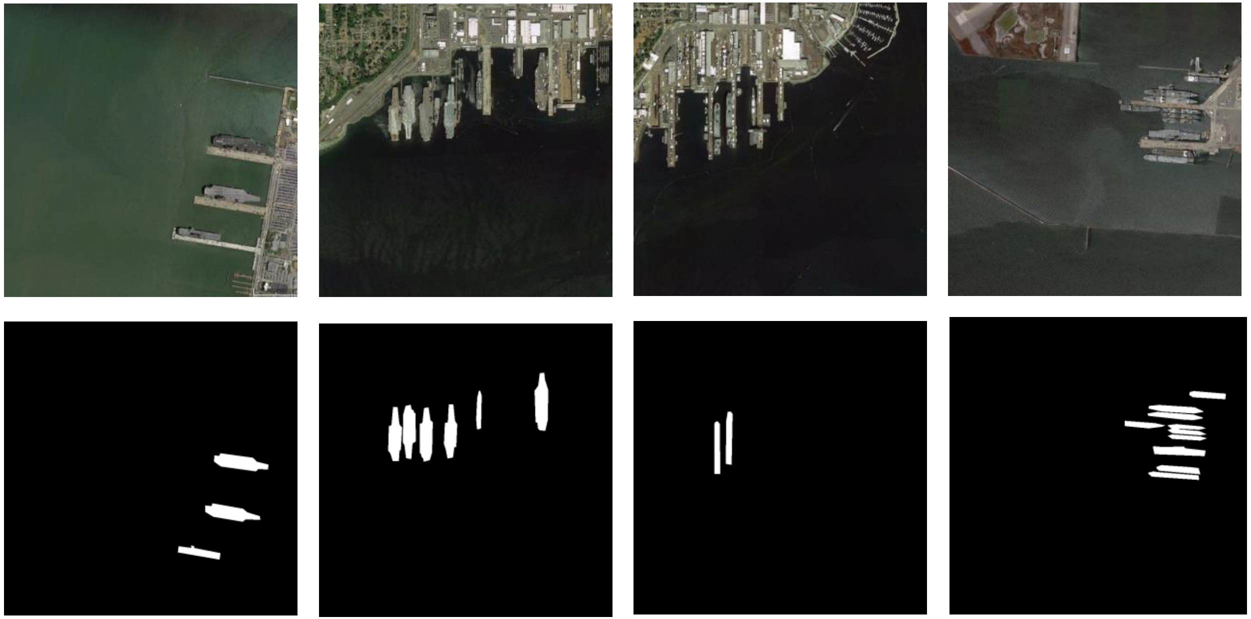
Fliker has billions of images

snapchat

Users chat in images or text format everyday

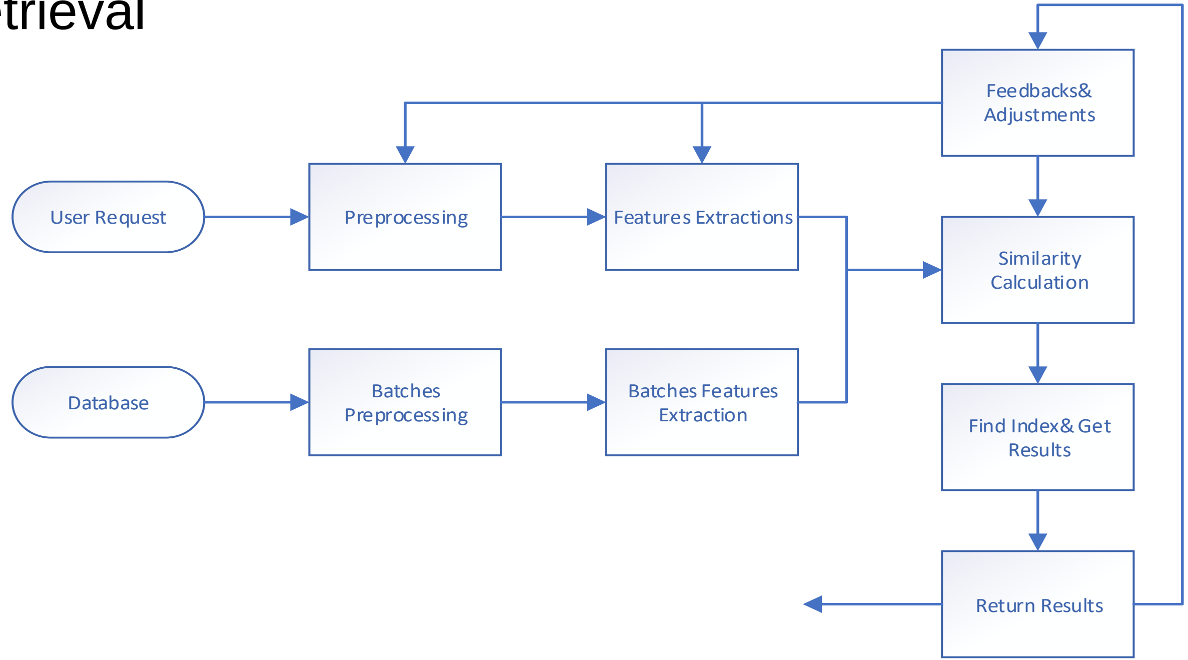
Image Retrieval

- Based on labels.
- Based on contents.



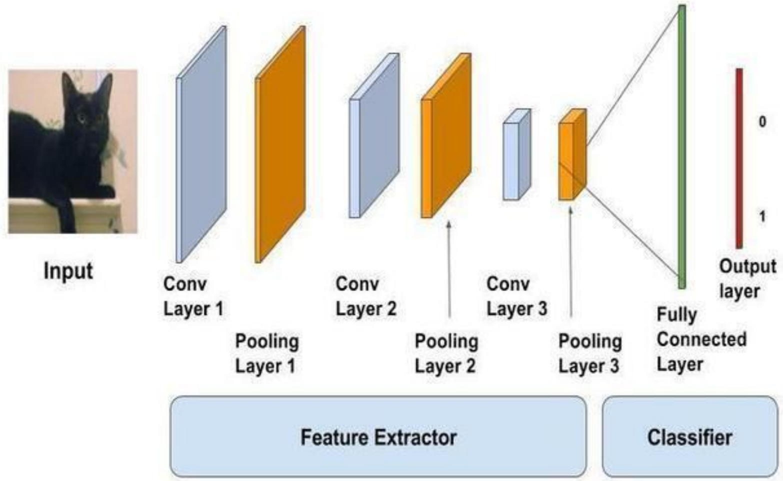
Binary Label

Image Retrieval



Retrieval progress

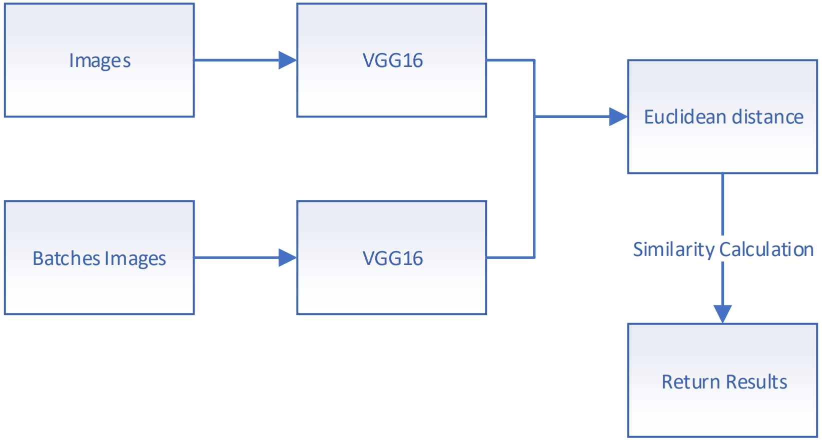
Image Retrieval



Feature clustering, t-SNE visualization

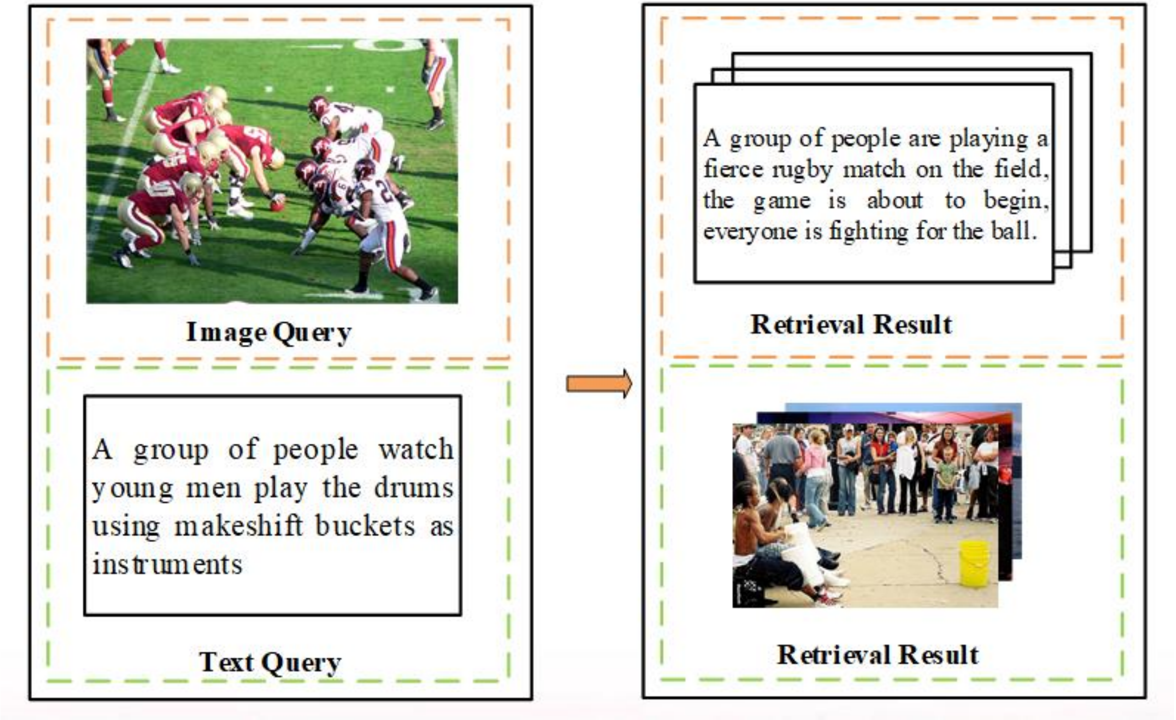
Image Retrieval

- Using VGG16 to extract features from images
- Computing feature similarity by using euclidean metric



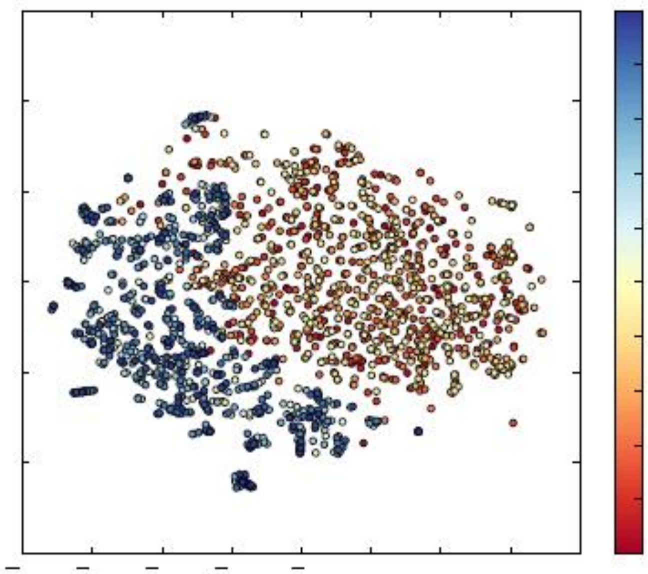

```
Matches are:
        Distance: 65.44046783447266, FileName: image_24.jpg
        Distance: 73.20362091064453, FileName: image_25.jpg
        Distance: 81.2601318359375, FileName: image_15.jpg
        Distance: 82.78900146484375, FileName: image_4.jpg
```

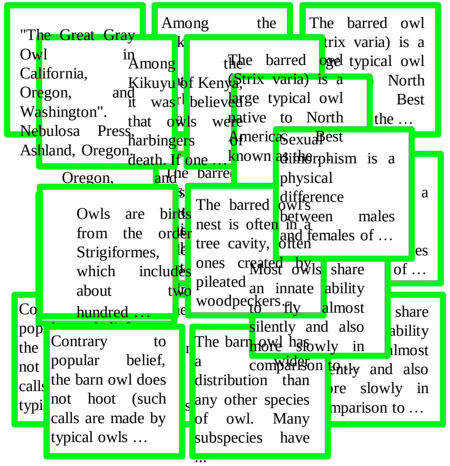
Cross-Modal Retrieval



Cross-Modal Retrieval

- Images and texts are heterogeneous
- Cannot estimate the distance directly

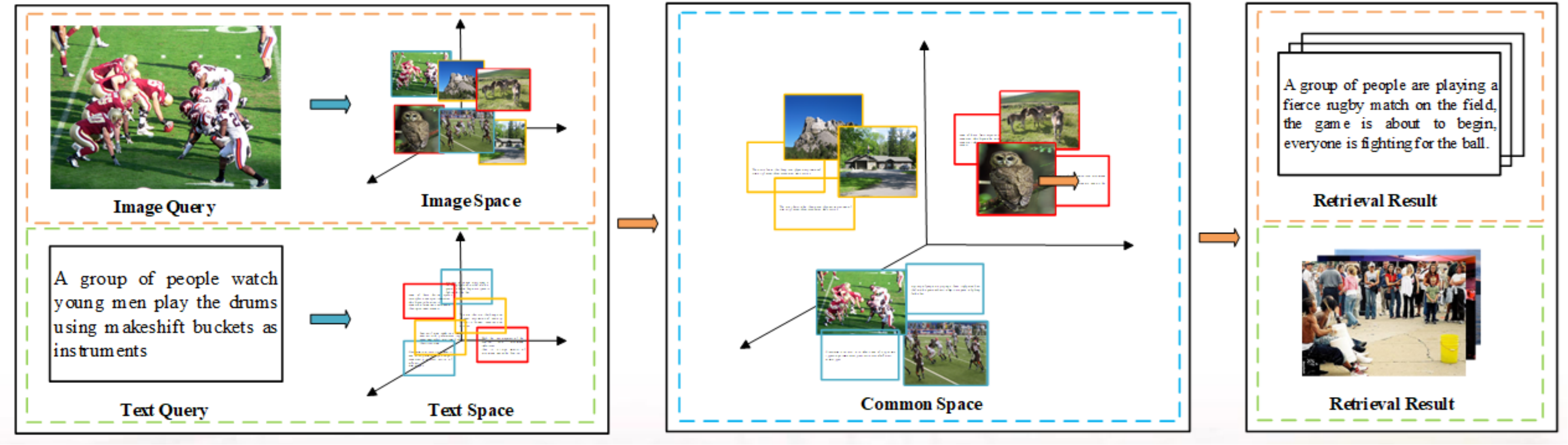




NYU TANDON SCHOOL OF ENGINEERING

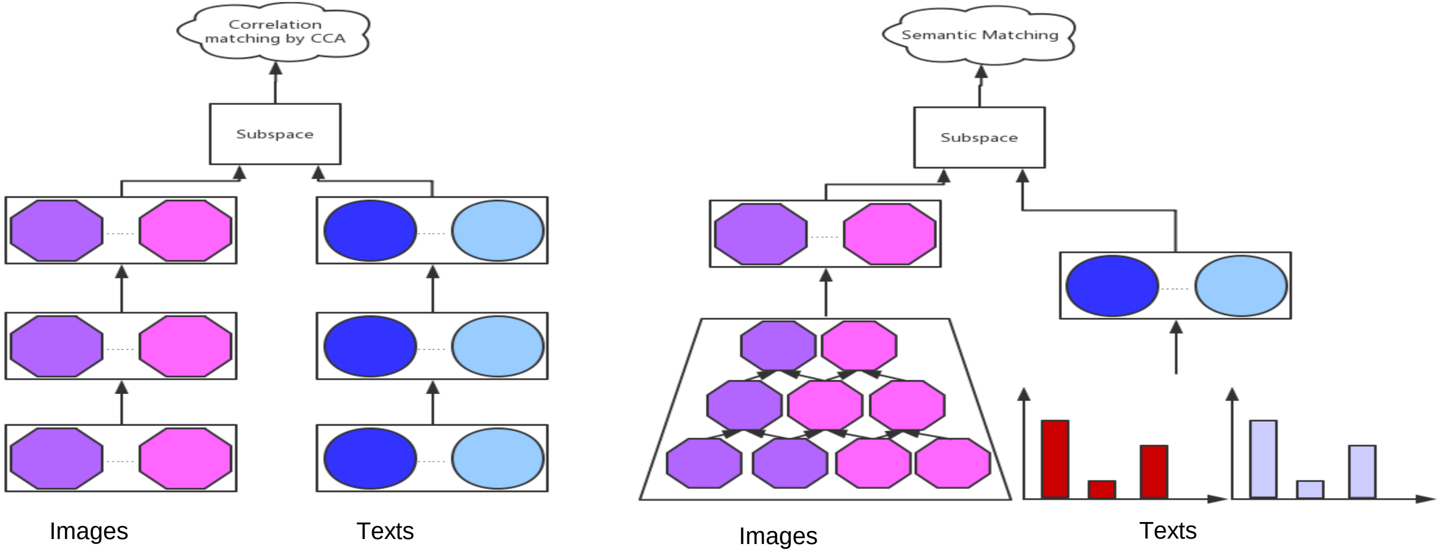
Cross-Modal Retrieval

- Find similarity between different information model.
- Common methods: Constructing common space
- Find the correlation between images and texts

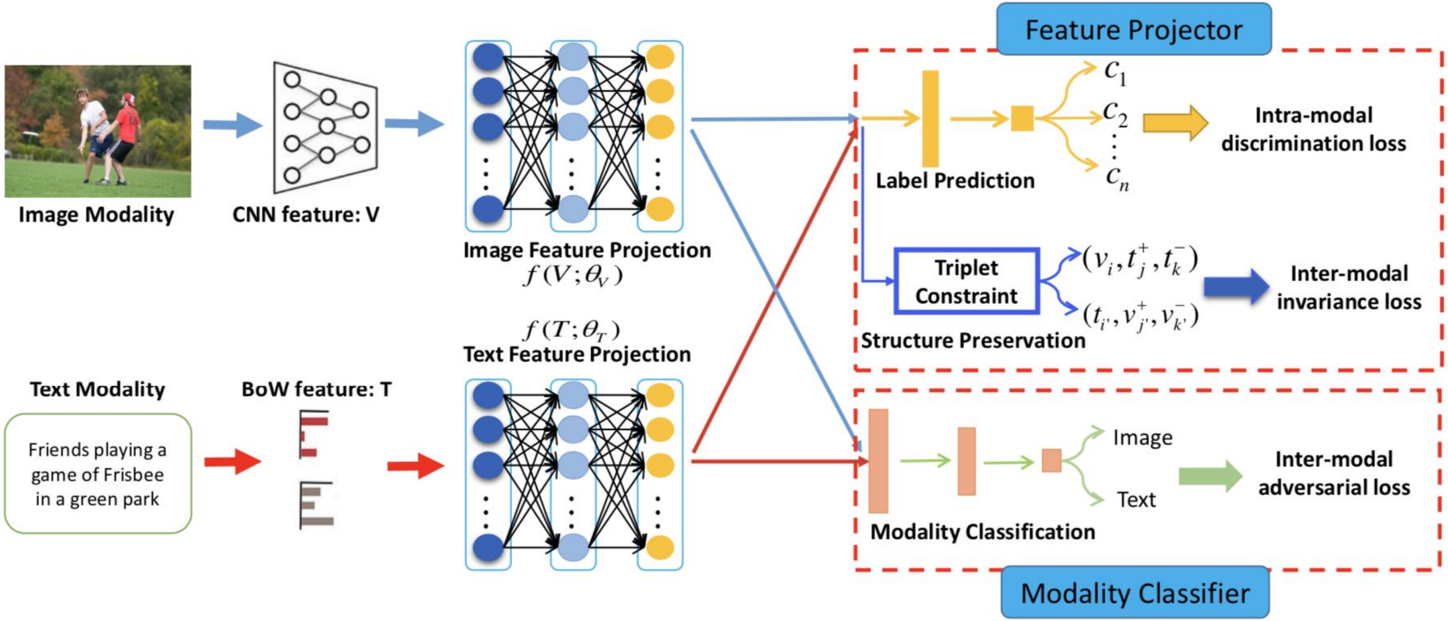


Cross-Modal Retrieval

- Traditional method : Canonical Correlation Analysis，CCA
- Deep learning method:



Cross-Modal Retrieval



Modality classifier distinguishing the items in terms of their modalities, and feature projector generating modality-invariant and discriminative representations and aiming to confuse the modality classifier.

I love artificial intelligence and machine learning

1st window

$X_{I,\ love}$ += 1
$X_{I,\ artificial}$ += 1

4th window

$X_{intelligence,\ love}$ += 1
$X_{intelligence,\ artificial}$ += 1
$X_{intelligence,\ and}$ += 1
$X_{intelligence,\ machine}$ += 1

7th window

$X_{learning,\ and}$ += 1
$X_{learning,\ machine}$ += 1

Glove model : $J = \sum_{i,j}^{N} f(X_{i,j})\,(v_i^T v_j + b_i + b_j - \log(X_{i,j}))^2$

Cross-Modal Retrieval

Loss functions:

- Classification loss:

$$L_{imd}\left(\theta_{imd}\right) = -\frac{1}{n}\sum_{i=1}^{n}\left(y_i \cdot \left(\log \hat{p}_i\left(v_i\right) + \log \hat{p}_i\left(t_i\right)\right)\right)$$

- Triplet Loss

$$L_2\left(v,t\right) = \left\| f_V\left(v;\theta_v\right) - f_T\left(t;\theta_T\right)\right\|_2$$

$$L_{imi,V}\left(\theta_V\right) = \sum_{i,j,k}\left(l_2\left(v_i,t_j^+\right) + \lambda \max\left(0, \mu - l_2\left(v_i,t_k^0\right)\right)\right)$$

$$L_{imi,\mathrm{T}}\left(\theta_\mathrm{T}\right) = \sum_{i,j,k}\left(l_2\left(t_i,v_j^+\right) + \lambda \max\left(0, \mu - l_2\left(t_i,v_k^0\right)\right)\right)$$

- adversarial and embedding losses

$$L_{emb}\left(\theta_V,\theta_\mathrm{T},\theta_{imd}\right) = \alpha L_{imi} + \beta L_{imd} + L_{reg}$$

Cont.

- MMD_loss:

$$MMD(X,\mathrm{Y}) = \left\| \frac{1}{n}\sum_{i=1}^{n}\Phi(x_i) - \frac{1}{m}\sum_{j=1}^{n}\Phi(y_j) \right\|_H^2$$

- Deep_CORAL_loss

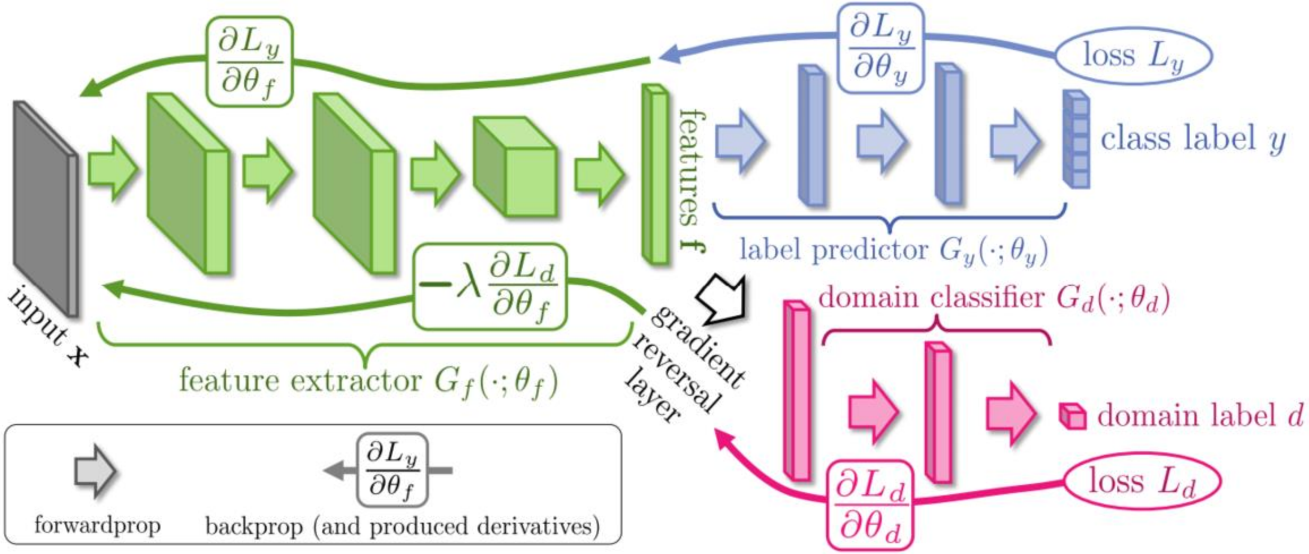
$$l_{CORAL} = \frac{\left\| C_S - C_T \right\|_F^2}{4d^2}$$

$$C_S = \frac{1}{n_S - 1}\left(D_S^T D_S - \frac{\left(1^T D_S\right)^T \left(1^T D_S\right)}{n_S} \right)$$

$$C_T = \frac{1}{n_T - 1}\left(D_T^T D_T - \frac{\left(1^T D_T\right)^T \left(1^T D_T\right)}{n_T} \right)$$

Cross-Modal Retrieval

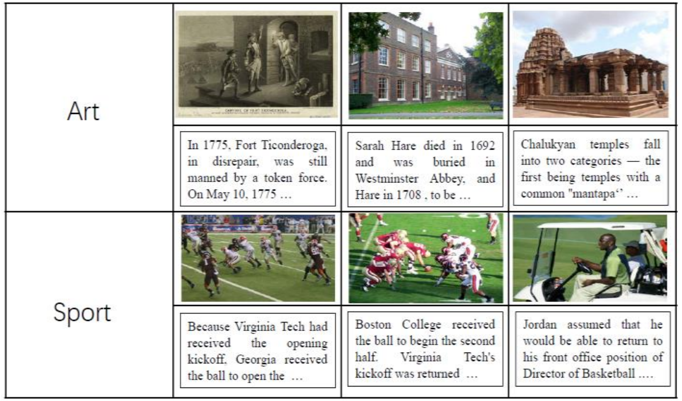
- Transfer learning



Cross-Modal Retrieval

- **Dataset**

Data set	Training data / test data	Number of class
Wikipedia	2173/693	10
Pascal Sentence	800/200	20

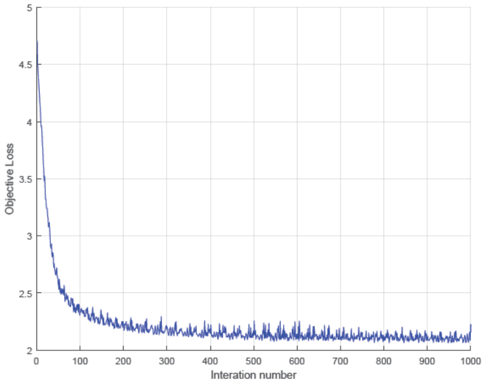


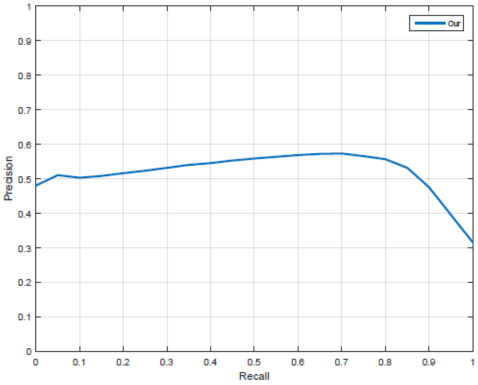
Example of Wikipedia dataset

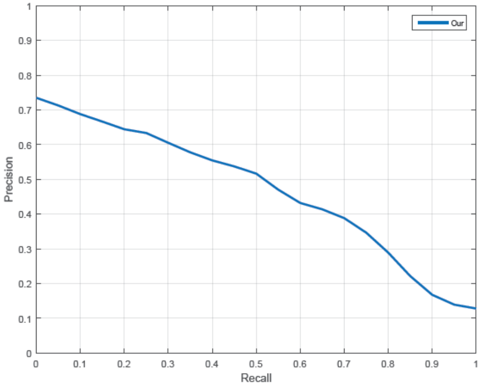
NYU TANDON SCHOOL OF ENGINEERING

MAP results

Data set	Retrieval the texts based on images(MAP)	Retrieval the images based on texts(MAP)	Average MAP
Wikipedia	0.53	0.46	0.495
Pascal Sentence	0.52	0.54	0.530







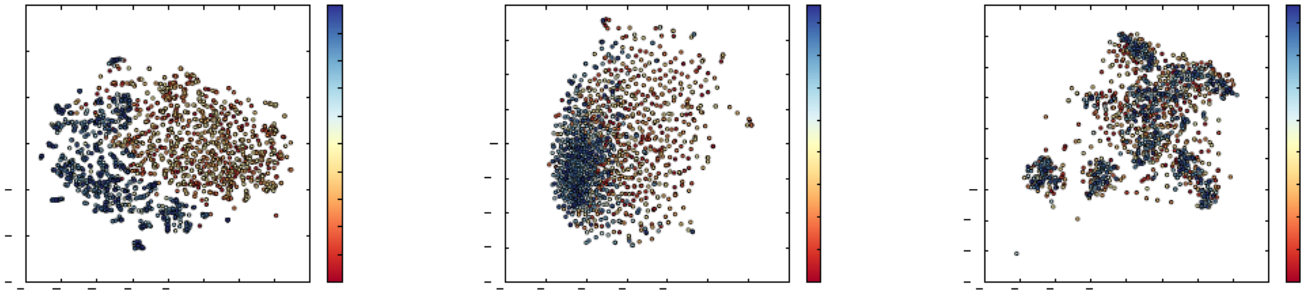
Cont.

Texts feature	Retrieval the texts based on images(MAP)	Retrieval the images based on texts(MAP)	Average MAP
bow	0.50	0.42	0.460
word2vec	0.52	0.44	0.480
glove	0.53	0.46	0.495

loss3	Retrieval the texts based on images(MAP)	Retrieval the images based on texts(MAP)	Average MAP
Adversarial loss	0.53	0.46	0.495
MMD_loss	0.51	0.47	0.490
Deep_CORAL_loss	0.51	0.44	0.475
Correlation_loss	0.47	0.40	0.435

Cont.

loss	Retrieval the texts based on images(MAP)	Retrieval the images based on texts(MAP)	Average MAP
Only loss1	0.35	0.40	0.375
Loss1+loss2	0.44	0.42	0.430
All loss	0.53	0.46	0.495



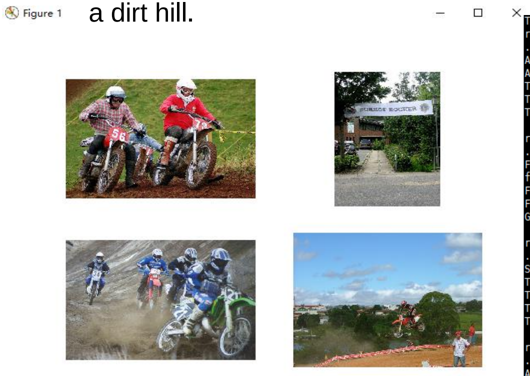
Tsne visualization

Training and testing results

```
Epoch: [ 0][   0/  33] time: 2.2069, emb_loss: 3067.68994141, domain_loss: 1.84295118, label_loss: 4.940
52410, triplet_loss: 2573.63745117 , 322.92285156 ,323.04046631 , 322.92288208 ,332.75079346
Epoch: [ 0][   1/  33] time: 2.4639, emb_loss: 2685.25878906, domain_loss: 1.58900285, label_loss: 4.982
82814, triplet_loss: 2186.97607422 , 276.64929199 ,278.58090210 , 276.64929199 ,301.13574219
Epoch: [ 0][   2/  33] time: 2.6943, emb_loss: 2383.67700195, domain_loss: 1.45281708, label_loss: 5.009
81522, triplet_loss: 1882.69555664 , 236.04240417 ,236.48657227 , 236.04241943 ,241.44207764
Epoch: [ 0][   3/  33] time: 2.9520, emb_loss: 2364.36547852, domain_loss: 1.53052711, label_loss: 4.993
94321, triplet_loss: 1864.97119141 , 234.26106262 ,233.35289001 , 234.26104736 ,244.48635864
Epoch: [ 0][   4/  33] time: 3.1950, emb_loss: 2228.92504883, domain_loss: 1.64426064, label_loss: 5.066
38718, triplet_loss: 1722.28625488 , 217.11094666 ,219.10539246 , 217.11096191 ,229.91780090
Epoch: [ 0][   5/  33] time: 3.4177, emb_loss: 2065.30249023, domain_loss: 1.80180085, label_loss: 4.928
36571, triplet_loss: 1572.46582031 , 199.23208618 ,199.86497498 , 199.23208618 ,220.18995667
Epoch: [ 0][   6/  33] time: 3.6639, emb_loss: 1999.90185547, domain_loss: 1.92574072, label_loss: 4.842
94176, triplet_loss: 1515.60766602 , 189.42655945 ,190.01152039 , 189.42654419 ,188.84634399
Epoch: [ 0][   7/  33] time: 3.9184, emb_loss: 1941.54296875, domain_loss: 1.80381215, label_loss: 4.922
92929, triplet_loss: 1449.25000000 , 183.50570679 ,184.12406921 , 183.50572205 ,201.88299561
Epoch: [ 0][   8/  33] time: 4.1573, emb_loss: 1837.40197754, domain_loss: 1.74798739, label_loss: 4.994
46821, triplet_loss: 1337.95520020 , 168.20713806 ,167.93162537 , 168.20712280 ,176.38444519
Epoch: [ 0][   9/  33] time: 4.4026, emb_loss: 1760.11706543, domain_loss: 1.82267380, label_loss: 4.964
19048, triplet_loss: 1263.69799805 , 158.25349426 ,158.58523560 , 158.25347900 ,160.45164490
Epoch: [ 0][  10/  33] time: 4.6373, emb_loss: 1696.84472656, domain_loss: 1.76925004, label_loss: 4.984
93910, triplet_loss: 1198.35083008 , 149.77940369 ,150.93249512 , 149.77938843 ,148.71063232
Epoch: [ 0][  11/  33] time: 4.8737, emb_loss: 1630.70397949, domain_loss: 1.81159520, label_loss: 4.836
00330, triplet_loss: 1147.10363770 , 145.24467468 ,147.06393433 , 145.24467468 ,158.47909546
Epoch: [ 0][  12/  33] time: 5.1453, emb_loss: 1550.39489746, domain_loss: 1.72135830, label_loss: 5.064
90755, triplet_loss: 1043.90417480 , 131.53250122 ,135.37445068 , 131.53248596 ,136.24621582
Epoch: [ 0][  13/  33] time: 5.4354, emb_loss: 1469.01892090, domain_loss: 1.50265455, label_loss: 5.053
89786, triplet_loss: 963.62915039 , 123.00194550 ,122.28170776 , 123.00194550 ,144.30845642
Epoch: [ 0][  14/  33] time: 5.7008, emb_loss: 1492.22583008, domain_loss: 1.70078802, label_loss: 5.096
18378, triplet_loss: 982.60742188 , 122.87506104 ,123.55076599 , 122.87506104 ,122.79236603
```

Traing results

Input：Four bikers are riding on a dirt hill.



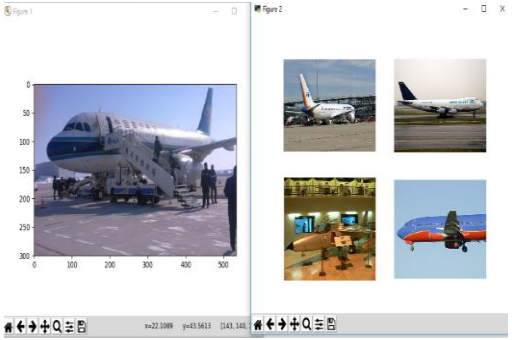

```
Text retrieval result
retrieval 634
./data/texts_pascal/motorbike/2008_008246.txt   motorbike
African man on blue bike posing with 2 friends.
A man sitting on a blue motorcycle with two men beside him.
Three men pose with a blue motorcycle.
Three men posing with a blue motorcycle.
Two guys standing next to a person on a blue motorcycle with a truck in the background.

retrieval 234
./data/texts_pascal/motorbike/2008_005213.txt   motorbike
Four bikers are riding on a dirt hill.
four motorbikes competing in the dirt
Four people race through the dirt on BMX bikes.
Four people racing motor bikes.
Group of four dirt bikers riding down dirt trail.

retrieval 761
./data/texts_pascal/motorbike/2008_007955.txt   motorbike
Several people riding dirt bikes with number plates on the bikes.
The elder rider looks to make his move on the leader.
Three men in helmets race motorcycles on a dirt track.
Two men ride motocross bikes
Two people wearing helmets ride dirt bikes.

retrieval 971
./data/texts_pascal/motorbike/2008_001203.txt   motorbike
A close-up of bright colored houses.
A line of closely packed buildings.
Alleyway with blue doors into building.
A view of buildings with various different colors alongside each other.
Brick, turquoise, and blue exteriors of buildings on a street.
```

Input an image:




```
Text retrieval result
retrieval 854
./data/texts_pascal/aeroplane/2008_008424.txt   aeroplane
Airplane on runway in front of buildings.
An aircraft sitting on the runway.
A white airplane is on a runway.
Front image of a parked silver airplane.
Plane on runway with buildings in the background

retrieval 604
./data/texts_pascal/aeroplane/2008_008146.txt   aeroplane
A jet flies high in the blue sky.
A plane flying at a distance.
A plane in the sky
A small jet airplane.
Fighter jet flying through the blue sky.

retrieval 240
./data/texts_pascal/aeroplane/2008_008044.txt   aeroplane
A gray jet parked near the building.
A U.S. military jet fighter on display.
A white airplane is parked on the cement.
A white plane on the runway
Large grey parked airplane.

retrieval 939
./data/texts_pascal/aeroplane/2008_008471.txt   aeroplane
An airplane is flying over a tree in the blue sky.
A plane is flying in the distance.
A small aircraft flies in the blue sky above the trees.
A small airplane flying above the trees.
The back of an airplane.
```