Deep Person Re-Identification with Camera-Label Based Batch Hard Mining

20170610 정찬이

Direction of Improvements

Adapt the loss function

A. Involve verification loss

Verification loss is often combined with the identity loss to improve the performance. (D. Chen et al., R. R. Varior et al.)

✓ B. Apply informative triplet mining

selected!

- Triplet loss determined by two randomly sampled person IDs
  ⟶ The large proportion of easy triplets dominate the training process resulting in limited discriminability
- Select informative triplets for computing triplet loss (H. Shi et al., A. Hermans et al.)

Topic

Improving the person re-identification baseline model introduced in

"A Strong Baseline and Batch Normalization Neck for Deep Person Re-identification"
(Luo et al.,2019)

by adding new triplet mining technique

that uses camera-ID label of the training data.

Baseline Model

"A Strong Baseline and Batch Normalization Neck for Deep Person Re-identification", Hau Luo et al.

- Closed-world, supervised learning
- Use global features of ResNet50
- Combine effective training tricks from literatures
- Propose a novel neck structure named batch normalization neck (BNNeck)
- Introduce center loss

Loss Function Design

"A Strong Baseline and Batch Normalization Neck for Deep Person Re-identification", Hau Luo et al.

- Identity loss

$$\mathcal{L}_{id} = -\frac{1}{n}\sum_{i=1}^{n}\log(p(y_i|x_i))$$

- Triplet loss
  - batch hard mining

$$\mathcal{L}_{tri}(i,j,k) = \max(\rho + d_{ij} - d_{ik}, 0)$$

- Center loss

$$\mathcal{L}_C = \frac{1}{2}\sum_{j=1}^{B}\left\|\boldsymbol{f}_{t_j} - \boldsymbol{c}_{y_j}\right\|_2^2$$

Triplet Loss

"A Strong Baseline and Batch Normalization Neck for Deep Person Re-identification", Hau Luo et al.

- $L_{Tri} = [d_p - d_n + \alpha]_+$

$d_p (d_n)$: distance between anchor sample and positive (negative) sample

α : margin parameter

$[z]_+$: max(z, 0)

- The anchor should be closer to the positive sample than to the negative sample by a pre-defined margin.
- However, the large proportion of easy triplets will dominate the training process if we directly optimize above loss function, resulting in limited discriminability.

→ **Triplet mining** (select informative triplets)

Batch Hard

"In defense of the triplet loss for person re-identification", A. Hermans, L. Beyer, and B. Leibe

- Triplet mining method
- For each sample in the batch, select the hardest positive and the hardest negative sample within the batch when forming the triplets for computing the loss
  - hardest positive(negative)
    : image sample that belongs to the same(different) person and hardest to correctly classify

$$\mathcal{L}_{\mathrm{BH}}(\theta; X) = \overbrace{\sum_{i=1}^{P} \sum_{a=1}^{K}}^{\text{all anchors}} \Big[m + \overbrace{\max_{p=1\ldots K} D\left(f_\theta(x_a^i), f_\theta(x_p^i)\right)}^{\text{hardest positive}} \qquad (5)$$

$$\underbrace{- \min_{\substack{j=1\ldots P \\ n=1\ldots K \\ j\neq i}} D\left(f_\theta(x_a^i), f_\theta(x_n^j)\right)}_{\text{hardest negative}} \Big]_+ ,$$

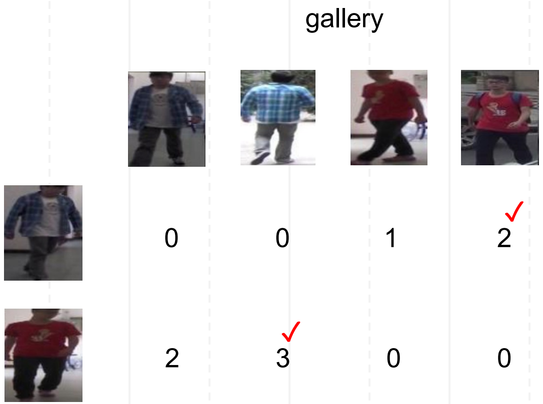
Batch Hard

"In defense of the triplet loss for person re-identification", A. Hermans, L. Beyer, and B. Leibe

- Select hardest sample within a small subset of data (not within the entire dataset)

  → Moderate, not too hard

- Distance matrix
  - Distance between the most recent embeddings of pairs of images in the batch
  - The image of the maximum(minimum) score is selected as the hardest positive(negative)



New : Camera-label Based Bath Hard

- Batch hard (Hermans et al.) is applied in the baseline.
- An improvement on batch hard
  - At initial stages, the embeddings of images are not accurate.
  - → Selecting hardest sample based on the embedding would not work so well.

Could we help selecting hard samples at initial stages with camera labels of the training images?

New : Camera-label Based Bath Hard

1. Add bias to the distance matrix depending on whether or not the cam-ID of a gallery image is same with that of the query image

- positive samples

  Those with the **different** cam-ID are harder to classify due to the **different** angle, light, background, etc.

- negative samples

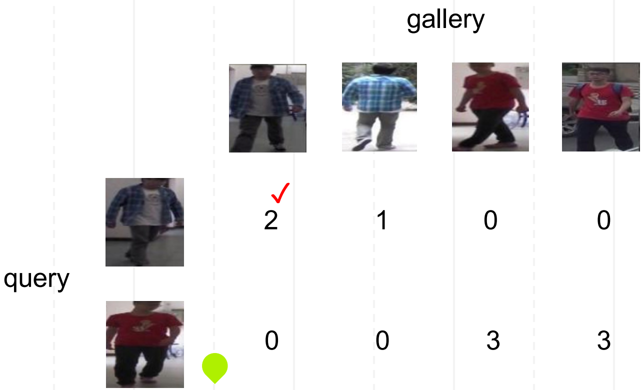
  Those with the **same** cam-ID are harder to classify due to the **similar** angle, light, background, etc.

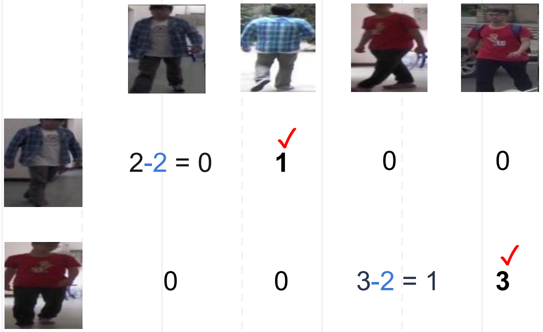
New : Camera-label Based Bath Hard

Positive-masked distance matrix

- Distance matrix in which the entries between different person identities are set zero.
- The image of the maximum score is selected as the hardest positive.

→ Decrease the distance between the images of same camera.



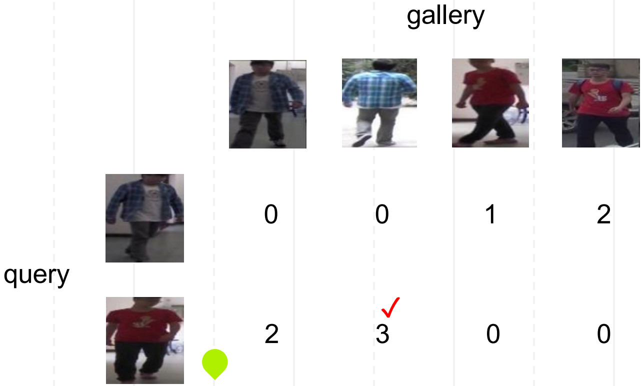


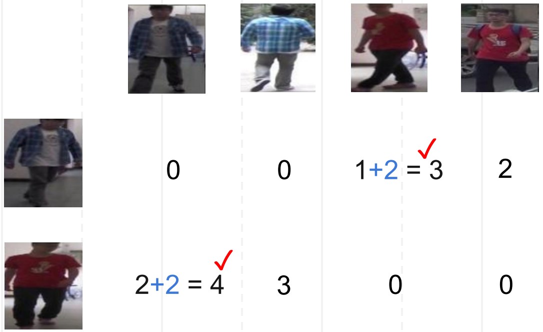
New : Camera-label Based Bath Hard

Negative-masked distance matrix

- Distance matrix in which the entries between same person identities are set zero.
- The image of the minimum score is selected as the hardest positive.

→ Increase the distance between the images of same camera.





New : Camera-label Based Bath Hard

2. As iteration grows, the image embeddings will get accurate so that it can select hard samples properly by itself.

→ Decay the bias as iteration grows

New : Camera-label Based Bath Hard

Bias formula

$$bias = k * std(D)/(1 + mt)$$

k : hyperparameter deciding the size of the bias

D : positive/negative-masked distance matrix

m : hyperparameter deciding the decaying speed of bias

t : iteration number

Experiment Setting

Pre-trained model	Resnet50
Pre-trained dataset	ImageNet
Training Dataset	Market1501
Training Batch size	16
Training epochs	120

Result

- Validation accuracy with different hyperparameters

Baseline : 85.8%(mAP), 94.5%(Rank-1)

k \ m	0.1	0.4	0.7
1	85.9, 94.2	85.7, 93.8	85.6, 93.6
5	**85.9, 94.6**	85.7, 94.1	85.6, 94.1
9	85.6, 94.2	85.9, 93.7	85.8, 94.0

Result
- Validation accuracy with different hyperparameters

Baseline : 85.8%(mAP), 94.5%(Rank-1)

More experiments with k=5, m close to 0.1

k \ m	0.05	0.1	0.2
5	85.6, 93.8	**85.9, 94.6**	85.6, 94.1

∴ k=5, m=0.1 gives the best performance.

Result – test accuracy

	mAP	Rank-1
Baseline	**94.1%**	**95.7%**
Ours (k=5, m=0.1)	94.0%	95.2%

The baseline generalizes better than our model.

Discussion

- Camera-label based batch hard didn't always result in better performance.
- However, there was a hyperparameter set which resulted in increased validation mAP and rank-1 accuracy.
- This provides the potential basis for further experiments of hyperparameter tuning and improvement of the method to achieve a better performance.

Limitation

- Plotting the loss curve of my model could give a better insight of its effectiveness compared with that of baseline.
  - Failed to find the way to use TensorBoard with the old pytorch-ignite version used in this implementation.

- More experiments could be done with more diverse settings.
  - Adding bias only in the early epochs
  - Non-decaying bias