

Work completed last week :

- The papers related to CVPR/ICCV/ECCV person re-identification in 2020-2022 are summarized. It is found that the starting point of the current paper lies in the following areas:

Paper	Noise-refine	Feature-fusion	Label-refine	Description
Implicit Sample Extension for Unsupervised Person Re-Identification (CVPR 2022 )	☑			Progressive linear inter clustering problems: s
Joint Noise-Tolerant Learning and Meta Camera Shift Adaptation for Unsupervised Person Re-Identification(CVPR 2021)	☑			Focus on noisy labels feature variations caus
Learning with Twin Noisy Labels for Visible-Infrared Person Re-Identification(CVPR 2022)	☑			Focus on two noisy lat correspondence
FMCNet: Feature-Level Modality Compensation for Visible-Infrared Person Re-Identification (CVPR 2022 )		☑		Compensate the missi information in the feat image level
Channel Augmented Joint Learning for Visible-Infrared Recognition(ICCV 2021)		☑		Homogenously genera randomly exchanging
Discover Cross-Modality Nuances for Visible-Infrared Person Re-Identification (CVPR 2021)		☑		Discover cross-modal shoes, and the length explored, especially in different patterns for v
Farewell to Mutual Information: Variational Distillation for Cross-Modal Person Re-Identification(CVPR 2022 as Oral presentation)		☑		Variational Self-Distill scalable, flexible and a fitting the mutual infor estimating it.
Neural Feature Search for RGB-Infrared Person Re-Identification (CVPR 2021)		☑		Automate the process identity-related cues i and fine-grained spati
Optimal Transport for Label-Efficient Visible-Infrared Person Re-Identification(ECCV 2022)			☑	UDA for labeled RGB - Optimal-transport stra from visible to infrared
Part-based Pseudo Label Refinement for Unsupervised Person Re-identification(CVPR 2022)			☑	Employing the comple global and part feature

- Two research directions are proposed: 1) Nighttime Visible-Infrared Input person Re-ID: This is because the current papers say that the introduction of infrared image is: to deal with nighttime surveillance and low-light environments. However, the currently widely used dataset-SYSUMM01 does not have nighttime RGB images, so the RGB images used in the model training and testing stages are all daytime. So I ask a question: Under the conditions of satisfying the real situation, the visible and infrared images of the suspect will also be captured at night. If I input the nighttime visible into the existing model, the large domain gap will inevitably make the model in the RGB feature extractor stage will degrade performance. So is it possible to consider mixing daytime and nighttime RGB images during training and testing, or adding nighttime RGB images (day-day, night-night matched) when inputting nighttime infrared images, so as to add to test-find daytime (good illumination) RGB image. 2) Visible-Infrared Cloth-changing Person Re-ID: At present, RGB (single-modality) cloth-changing has attracted widespread attention, and there is currently no introduction of infrared images to body shape features (in single-modality-- 'pose-specific multi-branch network'), so I think cross-modality will achieve better performance in dealing with cloth-changing Re-ID.

• Notes taken briefly using Zotero while reading papers

总结 2020-2022 C...tml27.8 KB

Weekly report -- 0320-0326

Todo List

- Re-understand 'Optimal Transport for Label-Efficient Visible-Infrared Person Re-Identification (ECCV 2022)' and other UDA/Unsupervised methods to figure out our technical contributions.
- Think about whether there is a new novel classification/solution for noise on clustering besides using Lipschitz Continuity reduce cross- modality discrepancy.
- In the GANs we designed, since it is cross-modalities, the intra-modality is a zero-sum game, and in the inter-modality we can set other games (Stackelberg game!)
- \* Start reading three papers a day (cross-research fields), and systematically sort out good papers (brief introduction, advantage/disadvantage, what we can learn)

For the two cross-modality research directions mentioned last week, based on the survey of existing articles, the following ppt was made:

Cloth- changing reid

Existing works

Fine-Grained Shape-Appearance Mutual Learning for Cloth-Changing Person Re-Identification(CVPR 2021)

✓ They learns fine-grained discriminative body shape knowledge in a shape stream and transfers it to an appearance stream to complement the cloth-unrelated knowledge in the appearance features.

✓ Extracts fine-grained body shape features through pose-specific multi-branch network

✓ It also takes into account noises due to similar shape, pose variation and missing parts.

Disadvantage:

1. Body shape features obtained with a pose-specific multi-branch network may be affected by loose/tight clothing.

2. Under similar shape and same cloth, the situation of different identities is not discussed (or the degree of hard negative sample is not enough) .

3. The model relies heavily on the accuracy of the pose estimation or human parsing models

Cloth- changing reid

Cloth-Changing Person Re-identification with Self-Attention(WACV 2022)

• Propose vision-transformer based framework to address the inherent issues in CNN-based model (inductive biases)

• They pair the VIT baseline(establish long-range dependencies, exploit global structural patterns, and also retain fine-grained features) with a human body shape-motion modeling framework such as VIBE [13] that will provide the soft-bio metric features such as gait for handling CC-ReID problem.

Disadvantage:

1. Although long-range dependencies are established, noises brought by similar shapes or missing parts are not considered

2. Since camera shift is considered, the results show that the same clothes or different clothes under different viewing angles have not been recognized.

Cloth- changing reid

Motivation

1. RGB-T cloth-changing person re-identification is more realistic. Because most of the existing methods are based on the assumption that the same person wears the same clothes in short term, but they do not actually meet the real-world scene.

2. RGB-T cloth-changing is a more challenging task

3. We hope to use cross- modality to learn robust and unique feature representations that are invariant to clothing changes through shared and specific features between two modalities.

Challenge

1. There is currently no corresponding cloth-changing rgb-t dataset, rgb has 3 cloth-changing reid datasets: PRCC, LTCC and VCclothes (but if create a dataset corresponding to RGB-T, it will be time-consuming and lack label annotation)

2. Large domain gap led to the difficulty in cross-modal matching (due to different clothes during clustering, the same identity is divided into different classes according to the category of clothes)

3. The domain shift is different between modalities due to cloth-changing (Different clothes of the same identity in RGB have a greater impact on clustering and model judgment than Thermal.)

Cloth- changing reid

So what we focus on

□ Compensate the advantage of thermal on body shape feature to RGB, thus increasing the recognition confidence

□ By building a network of GANs, more confusing hard negative samples can be generated to increase the robustness of the model

□ Reduce the intra-class domain gap and increase the inter-class domain gap by introducing thermal images

On this cloth-changing reid, I think it is also possible to combine the two GANs proposed by us to conduct a Stackelberg game to get a more realistic hard negative sample and get a classifier with better performance.

- However, due to the lack of corresponding infrared data sets, nighttime and cloth-changing can be considered after creating a database corresponding to RGB-T-Event.
- Read the Steinberg game to get the idea, and try to apply the Stackelberg game to two GANs, first of all, it is a zero-sum game inside the two GANs, which refers to first fixing the Generator to optimize the Discriminator, and then fixing the Discriminator to optimize Generator and generates hard negative sample.
  - I mistakenly thought that two Discriminators can play a master-slave game, input a picture at the same time, and then the two Discriminators compete with each other to judge whether it is a synthetic image, so as to get a better Discriminator and then use self-guided adversarial learning to improve the performance of CODING-GANs, but later thought of the reason for cross-modality, the Discriminator of different modalities will not have good performance when input different modal data, so there are loopholes in the idea, discard.
  - Ask, Visible's anchor and hard negative sample + Infrared's anchor and hard negative sample can play a Stackelberg game, or change the game method to realize the inter-GAN game (continue to investigate later)

Summarize similar papers:

Optimal Transport for Label-Efficient Visible-Infrared Person Re-Identification (ECCV 2022)

• What they focus ?

Can they learn a cross-modality model only with one modal supervision or even without supervision?

• Methodology

1. Use UDA to generate pseudo labels for unlabeled visible data through visible labeled dataset

2. Use optimal transport to assign pseudo labels from visible to infrared modality.

3. Introduce uniform sample-wise and label-wise prior to match visible and infrared samples and set up prediction alignment loss to reduce the influence of incorrect label.

• Advantages

1. It can be achieved with only a part of the labeled visible dataset, less laboring efforts.

2. Improve accuracy by introducing additional technical contribution when matching visible and infrared sample (pseudo label assignment).

3. The model is experimentally proven to have little computational cost.

• Disadvantages

1. The impact of hard negative samples is not considered.

2. Vulnerable to UDA model performance

Self-Guided Hard Negative Generation for Unsupervised Person Re-Identification (IJCAI-22)

• What they focus ?

How to alleviate the misclustering problem due to hard negative sample?

• Methodology

1. Propose hard negative generation network (HNGN) to generate hard negative sample.

2. Use adversarial manner to let Generator and Discriminator to improve each other to get the hardest negative sample.

3. The re-ID network guides HNGN to generate challenging data, and HNGN in turn enforces the re-ID network to enhance discrimination ability.

• Advantages

1. The most confusing hard negative sample was obtained through the adversarial manner.

$$\mathcal{L}_{hng} = \frac{1}{N} \sum_{i=1}^N -\log \frac{\sum_{p \in \mathcal{P}^i} \exp(\text{sim}(\mathbf{v}^i, \mathbf{v}_p)/\tau)}{\sum_{q \in \mathcal{P}^i \cup \mathcal{N}^i} \exp(\text{sim}(\mathbf{v}^i, \mathbf{v}_q)/\tau)}, \quad (4)$$

$$\mathcal{L}_{con} = \frac{1}{N} \sum_{i=1}^N -\log \frac{\sum_{p \in \mathcal{P}_{con}^i} \exp(\text{sim}(\mathbf{v}^i, \mathbf{v}_p)/\tau)}{\sum_{q \in \mathcal{P}_{con}^i \cup \mathcal{N}_{con}^i} \exp(\text{sim}(\mathbf{v}^i, \mathbf{v}_q)/\tau)}, \quad (7)$$

where  $\mathcal{P}_{con}^i$  and  $\mathcal{N}_{con}^i$  are the positive set and negative set of image  $x_{anc}^i$ . They are defined as follows.

- 2. The mutual supervision of Re-ID network and HNGN improves the performance of re-id (so can we set how to supervise (self- guided) between two GANs and re-id network/or two GANs to serve as teachers to compete for re- id suggestion?)
  - 3. The part-level feature fusion strategy is realized through cosine similarity, which can be used as a naive method to obtain hard negative samples.
- Disadvantages
1. How to improve if the same identity is divided into two or mixed in the clustering stage.
2. In the sample stage, the negative image is only randomly selected, and the nearest one can be selected to improve the performance of the hard negative sample.
3. Has a high computation cost

In addition, the abstract has been modified to improve the topic introduction of the hard negative sample, increase the coherence of reading, and integrate the challenges into two points. And established a group weekly report Evernote.

Plan for next week:

1. Think about how to conduct inter-GANs game to increase the technical contribution of this paper
2. Solve how to better define and generate the cross- modality hard negative sample
3. Consider the relationship between two GANs and our re-id network.