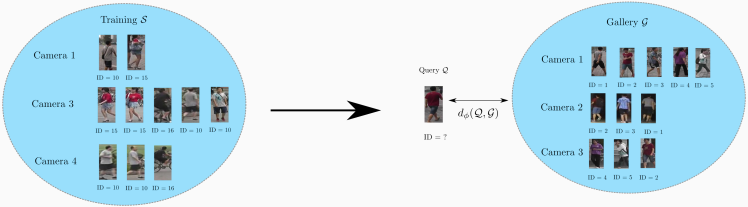
CANU-ReID: A Conditional Adversarial Network for Unsupervised person Re-IDentification

Guillaume Delorme[1], Yihong Xu[1], Stéphane Lathuilière[2], Radu Horaud[1], Xavier Alameda-Pineda[1]
[1]Inria, LJK, Univ. Grenoble Alpes, France [2] LTCI, Télécom Paris, IP Paris, France

12/01/2021, ICPR 2020, Milano, Italy

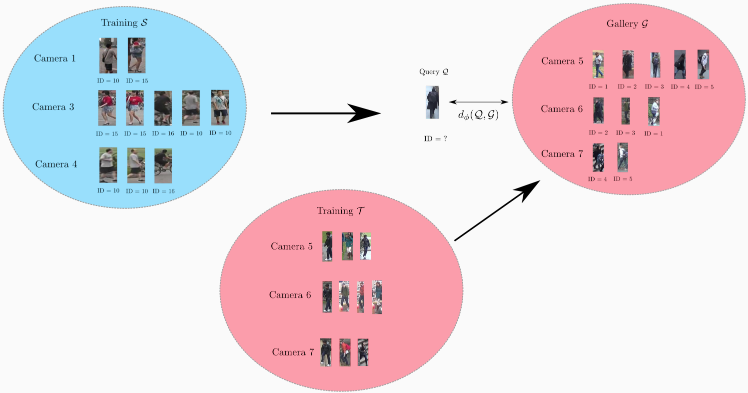
Unsupervised Person Re-Identification

Unsupervised Person Re-Identification



- supervised Re-ID: large annotated datasets $\rightarrow$ Unsupervised Person Re-ID.
- labeled source $\mathcal{S}$, unlabeled target $\mathcal{T}$: optimizes re-ID performance on $\mathcal{T}$.

Unsupervised Person Re-Identification



- supervised Re-ID: large annotated datasets $\rightarrow$ Unsupervised Person Re-ID.
- labeled source $\mathcal{S}$, unlabeled target $\mathcal{T}$: optimizes re-ID performance on $\mathcal{T}$.

Related Work

Overview

We need to review the following topics:

- Clustering and Finetuning
- Adversarial Domain Adaptation
- Negative Transfer

Clustering and Finetuning

Recent works in Unsupervised Person Re-ID are based on the *Clustering* and *Finetuning* framework (SSG[1], MMT[2]):

Source Pretraining ϕ pretrained with ID loss $\mathcal{L}_{\mathrm{ID}}^{\mathcal{S}}(\phi)$ on source S.

1 - Clustering step ϕ frozen, run clustering on $\mathcal{T}$ $\phi(\boldsymbol{x}_n^{\mathcal{T}}) \rightarrow$ pseudo-ID labels $\tilde{\boldsymbol{p}}_n^{\mathcal{T}}$.

2 - Finetuning step ϕ finetuned using $\tilde{\boldsymbol{p}}_n^{\mathcal{T}}$ with $\mathcal{L}_{\mathrm{PS\text{-}ID}}(\phi)$.

3 - Return to 1 until convergence.

[1] Yang Fu et al. "Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification". In: *IEEE ICCV*. 2019.

[2] Yixiao Ge, Dapeng Chen, and Hongsheng Li. "Mutual Mean-Teaching: Pseudo Label Refinery for Unsupervised Domain Adaptation on Person Re-identification". In: *ICLR* (2020).

Clustering and Finetuning

Recent works in Unsupervised Person Re-ID are based on the *Clustering* and *Finetuning* framework (SSG[1], MMT[2]):

Source Pretraining ϕ pretrained with ID loss $\mathcal{L}_{\text{ID}}^{\mathcal{S}}(\phi)$ on source S.

1 - Clustering step ϕ frozen, run clustering on $\mathcal{T}$ $\phi(\boldsymbol{x}_n^{\mathcal{T}}) \rightarrow$ pseudo-ID labels $\tilde{\boldsymbol{p}}_n^{\mathcal{T}}$.

2 - Finetuning step ϕ finetuned using $\tilde{\boldsymbol{p}}_n^{\mathcal{T}}$ with $\mathcal{L}_{\text{PS-ID}}(\phi)$.

3 - Return to 1 until convergence.

---

[1] Yang Fu et al. "Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification". In: *IEEE ICCV*. 2019.

[2] Yixiao Ge, Dapeng Chen, and Hongsheng Li. "Mutual Mean-Teaching: Pseudo Label Refinery for Unsupervised Domain Adaptation on Person Re-identification". In: *ICLR* (2020).

Clustering and Finetuning

Recent works in Unsupervised Person Re-ID are based on the *Clustering* and *Finetuning* framework (SSG[1], MMT[2]):

Source Pretraining ϕ pretrained with ID loss $\mathcal{L}_{\mathrm{ID}}^{\mathcal{S}}(\phi)$ on source S.

1 - Clustering step ϕ frozen, run clustering on $\mathcal{T}$ $\phi(\boldsymbol{x}_n^{\mathcal{T}}) \rightarrow$ pseudo-ID labels $\tilde{\boldsymbol{p}}_n^{\mathcal{T}}$.

2 - Finetuning step ϕ finetuned using $\tilde{\boldsymbol{p}}_n^{\mathcal{T}}$ with $\mathcal{L}_{\text{PS-ID}}(\phi)$.

3 - Return to 1 until convergence.

---

[1] Yang Fu et al. "Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification". In: *IEEE ICCV*. 2019.

[2] Yixiao Ge, Dapeng Chen, and Hongsheng Li. "Mutual Mean-Teaching: Pseudo Label Refinery for Unsupervised Domain Adaptation on Person Re-identification". In: *ICLR* (2020).

Clustering and Finetuning

Recent works in Unsupervised Person Re-ID are based on the *Clustering* and *Finetuning* framework (SSG[1], MMT[2]):

Source Pretraining ϕ pretrained with ID loss $\mathcal{L}_{\mathrm{ID}}^{\mathcal{S}}(\phi)$ on source S.

1 - Clustering step ϕ frozen, run clustering on $\mathcal{T}$ $\phi(\boldsymbol{x}_n^{\mathcal{T}}) \rightarrow$ pseudo-ID labels $\tilde{\boldsymbol{p}}_n^{\mathcal{T}}$.

2 - Finetuning step ϕ finetuned using $\tilde{\boldsymbol{p}}_n^{\mathcal{T}}$ with $\mathcal{L}_{\mathrm{PS\text{-}ID}}(\phi)$.

3 - Return to 1 until convergence.

---

[1] Yang Fu et al. "Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification". In: *IEEE ICCV*. 2019.

[2] Yixiao Ge, Dapeng Chen, and Hongsheng Li. "Mutual Mean-Teaching: Pseudo Label Refinery for Unsupervised Domain Adaptation on Person Re-identification". In: *ICLR* (2020).

Clustering and Finetuning

Recent works in Unsupervised Person Re-ID are based on the *Clustering* and *Finetuning* framework (SSG[1], MMT[2]):

Source Pretraining ϕ pretrained with ID loss $\mathcal{L}_{\mathrm{ID}}^{\mathcal{S}}(\phi)$ on source S.

1 - Clustering step ϕ frozen, run clustering on $\mathcal{T}$ $\phi(\boldsymbol{x}_n^{\mathcal{T}}) \rightarrow$ pseudo-ID labels $\tilde{\boldsymbol{p}}_n^{\mathcal{T}}$.

2 - Finetuning step ϕ finetuned using $\tilde{\boldsymbol{p}}_n^{\mathcal{T}}$ with $\mathcal{L}_{\mathrm{PS\text{-}ID}}(\phi)$.

3 - Return to 1 until convergence.

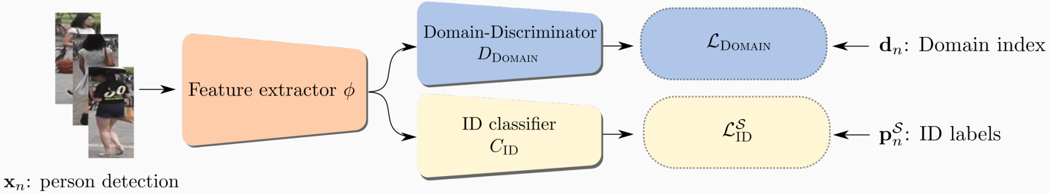
---

[1] Yang Fu et al. "Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification". In: *IEEE ICCV*. 2019.

[2] Yixiao Ge, Dapeng Chen, and Hongsheng Li. "Mutual Mean-Teaching: Pseudo Label Refinery for Unsupervised Domain Adaptation on Person Re-identification". In: *ICLR* (2020).

Adversarial Domain Adaptation

Adversarial Domain adaptation strategies[3] train a discriminator distinguishing target & source domain.



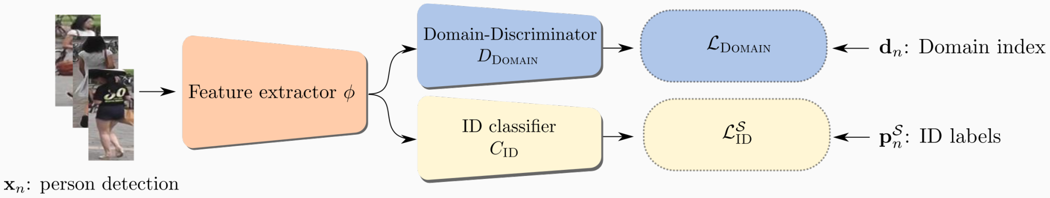
It is trained using the loss

$$\min_{\phi, C_{ID}} \max_{D_{\text{DOMAIN}}} \mathcal{L}^{\mathcal{S}}_{\text{ID}}(\phi, C_{\text{ID}}) - \mu \mathcal{L}_{\text{DOMAIN}}(D_{\text{DOMAIN}}) \quad (1)$$

[3] Yaroslav Ganin et al. "Domain-adversarial training of neural networks". In: *JMLR* (2016).

Adversarial Domain Adaptation

Adversarial Domain adaptation strategies[3] train a discriminator distinguishing target & source domain.



It is trained using the loss

$$min_{\phi, C_{ID}} \max_{D_{\text{DOMAIN}}} \mathcal{L}_{\text{ID}}^{\mathcal{S}}(\phi, C_{\text{ID}}) - \mu \mathcal{L}_{\text{DOMAIN}}(D_{\text{DOMAIN}}) \tag{1}$$

[3] Yaroslav Ganin et al. "Domain-adversarial training of neural networks". In: *JMLR* (2016).

Negative transfer

Domain Generalization[4] generalized this strategy to any number of domains.

Adversarial framework → *Negative Transfer*: discriminator learns **ID-related** instead of **domain-related features**.

Happens when **prior label distributions** are different accross domains.

[4] Ya Li et al. "Deep Domain Generalization via Conditional Invariant Adversarial Networks". In: *ECCV*. 2018.

Negative transfer

Domain Generalization[4] generalized this strategy to any number of domains.

Adversarial framework $\rightarrow$ *Negative Transfer*: discriminator learns **ID-related** instead of **domain-related features**.

Happens when **prior label distributions** are different accross domains.

[4] Ya Li et al. "Deep Domain Generalization via Conditional Invariant Adversarial Networks". In: *ECCV*. 2018.

Negative transfer

Domain Generalization[4] generalized this strategy to any number of domains.

Adversarial framework $\rightarrow$ *Negative Transfer*: discriminator learns **ID-related** instead of **domain-related features**.

Happens when **prior label distributions** are different accross domains.

[4] Ya Li et al. "Deep Domain Generalization via Conditional Invariant Adversarial Networks". In: *ECCV*. 2018.

Contributions

From this analysis derive the following strategies:

- **Camera adversarial-guided clustering**: in *Clustering* step, **viewpoint/camera variability** drives pseudo-label errors, and propose an adversarial strategy to reduce it.
- **Conditioned adversarial networks**: different ID prior distributions on different cameras lead to **negative transfer**.

Contributions

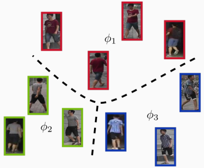
From this analysis derive the following strategies:

- **Camera adversarial-guided clustering**: in *Clustering* step, **viewpoint/camera variability** drives pseudo-label errors, and propose an adversarial strategy to reduce it.
- **Conditioned adversarial networks**: different ID prior distributions on different cameras lead to **negative transfer**.

Conditional Camera Adversarial Learning

Camera adversarial training pipeline

1- Clustering of target's embedding vectors $\phi(\mathbf{x}_n)$



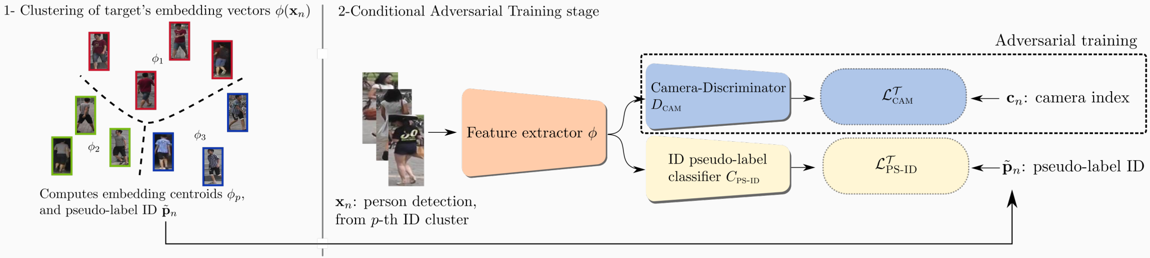
Computes embedding centroids ϕ_p, and pseudo-label ID $\tilde{\mathbf{p}}_n$

From Adversarial Domain Adaptation

$$\min_{\phi, C_{\text{PS-ID}}} \max_{D_{\text{CAM}}} \mathcal{L}^{\mathcal{T}}_{\text{PS-ID}}(\phi, C_{\text{PS-ID}}) - \mu \mathcal{L}^{\mathcal{T}}_{\text{CAM}}(\phi, D_{\text{CAM}}), \quad (2)$$

$$\mathcal{L}^{\mathcal{T}}_{\text{CAM}}(\phi, D_{\text{CAM}}) = -\mathbb{E}_{(\boldsymbol{x},\boldsymbol{c})\sim\mathcal{T}} \left\{ \log \langle D_{\text{CAM}}(\phi(\boldsymbol{x})), \boldsymbol{c} \rangle \right\} \quad (3)$$

Camera adversarial training pipeline

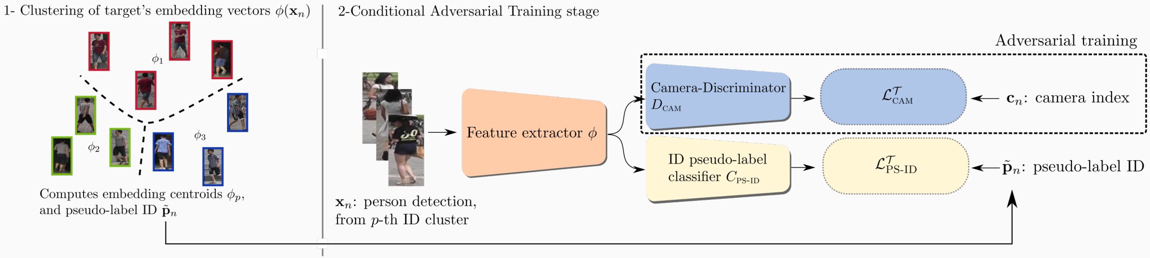


From Adversarial Domain Adaptation

$$\min_{\phi, C_{\text{PS-ID}}} \max_{D_{\text{CAM}}} \mathcal{L}^{\mathcal{T}}_{\text{PS-ID}}(\phi, C_{\text{PS-ID}}) - \mu \mathcal{L}^{\mathcal{T}}_{\text{CAM}}(\phi, D_{\text{CAM}}), \quad (2)$$

$$\mathcal{L}^{\mathcal{T}}_{\text{CAM}}(\phi, D_{\text{CAM}}) = -\mathbb{E}_{(\boldsymbol{x},\boldsymbol{c})\sim\mathcal{T}} \{\log \langle D_{\text{CAM}}(\phi(\boldsymbol{x})), \boldsymbol{c}\rangle\} \quad (3)$$

Camera adversarial training pipeline



From Adversarial Domain Adaptation

$$\min_{\phi, C_{\text{PS-ID}}} \max_{D_{\text{CAM}}} \mathcal{L}^{\mathcal{T}}_{\text{PS-ID}}(\phi, C_{\text{PS-ID}}) - \mu \mathcal{L}^{\mathcal{T}}_{\text{CAM}}(\phi, D_{\text{CAM}}), \tag{2}$$

$$\mathcal{L}^{\mathcal{T}}_{\text{CAM}}(\phi, D_{\text{CAM}}) = -\mathbb{E}_{(\boldsymbol{x},\boldsymbol{c})\sim\mathcal{T}} \left\{\log \langle D_{\text{CAM}}(\phi(\boldsymbol{x})), \boldsymbol{c}\rangle\right\} \tag{3}$$

Handling negative transfer

IDs are **unevenly distributed** across cameras $\rightarrow$ negative transfer:

- Can be solved by **adding the pseudo-ID label** information to the **discriminator input**.
- The **number of ID clusters** is big.
- Clustering algorithm **does not preserve** number of IDs and ordering.
- We use **centroids** ϕ_p provided by the clustering.

Handling negative transfer

IDs are **unevenly distributed** across cameras $\rightarrow$ negative transfer:

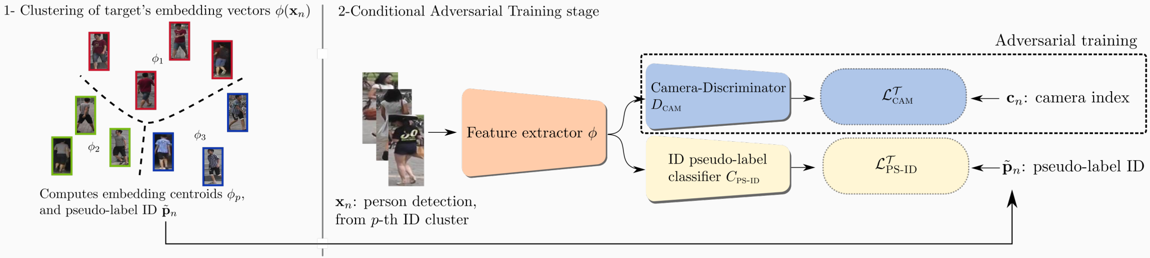
- Can be solved by **adding the pseudo-ID label** information to the **discriminator input**.
- The **number of ID clusters** is big.
- Clustering algorithm **does not preserve** number of IDs and ordering.
- We use **centroids** ϕ_p provided by the clustering.

Handling negative transfer

IDs are **unevenly distributed** across cameras $\rightarrow$ negative transfer:

- Can be solved by **adding the pseudo-ID label** information to the **discriminator input**.
- The **number of ID clusters** is big.
- Clustering algorithm **does not preserve** number of IDs and ordering.
- We use **centroids** ϕ_p provided by the clustering.

Camera adversarial training pipeline

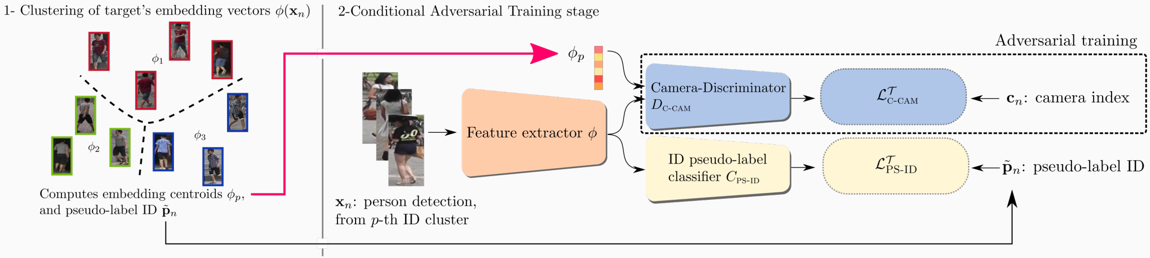


From Adversarial Domain Adaptation

$$\min_{\phi, C_{\text{PS-ID}}} \max_{D_{\text{CAM}}} \mathcal{L}^{\mathcal{T}}_{\text{PS-ID}}(\phi, C_{\text{PS-ID}}) - \mu \mathcal{L}^{\mathcal{T}}_{\text{CAM}}(\phi, D_{\text{CAM}}), \tag{4}$$

$$\mathcal{L}^{\mathcal{T}}_{\text{CAM}}(\phi, D_{\text{CAM}}) = -\mathbb{E}_{(\boldsymbol{x},\boldsymbol{c})\sim\mathcal{T}} \left\{ \log \left\langle D_{\text{CAM}}(\phi(\boldsymbol{x})), \boldsymbol{c} \right\rangle \right\} \tag{5}$$

Conditional camera adversarial training pipeline



Tackling Negative Transfer

$$\min_{\phi, C_{\text{PS-ID}}} \max_{D_{\text{C-CAM}}} \mathcal{L}^{\mathcal{T}}_{\text{PS-ID}}(\phi, C_{\text{PS-ID}}) - \mu \mathcal{L}^{\mathcal{T}}_{\text{C-CAM}}(\phi, D_{\text{C-CAM}}), \tag{6}$$

$$\mathcal{L}^{\mathcal{T}}_{\text{C-CAM}}(\phi, D_{\text{C-CAM}}) = -\mathbb{E}_{(\boldsymbol{x},\boldsymbol{p},\boldsymbol{c})\sim\mathcal{T}} \left\{ \log \left\langle D_{\text{C-CAM}}(\phi(\boldsymbol{x}), \boldsymbol{\phi_p}), \boldsymbol{c} \right\rangle \right\} \tag{7}$$

Camera adversarial training

Advantages:

- Can be **plugged into any** clustering and finetuning strategy: CANU-MMT, CANU-SSG
- Explicitely **reduce errors** in pseudo-ID labels,
- Make embedding space **invariant to camera information**, $\rightarrow$ better re-ID performance.

Camera adversarial training

Advantages:

- Can be **plugged into any** clustering and finetuning strategy: CANU-MMT, CANU-SSG
- Explicitely **reduce errors** in pseudo-ID labels,
- Make embedding space **invariant to camera information**, $\rightarrow$ better re-ID performance.

Camera adversarial training

Advantages:

- Can be **plugged into any** clustering and finetuning strategy: CANU-MMT, CANU-SSG
- Explicitely **reduce errors** in pseudo-ID labels,
- Make embedding space **invariant to camera information**, $\rightarrow$ better re-ID performance.

Experimental Evaluation

Experimental Setup

- CANU-SSG and CANU-MMT are evaluated.
- The clustering algorithm used is DBSCAN[5].
- The strategies are evaluated using Market-1501 (Mkt) [12], DukeMTMC-reID (Duke) [9] and MSMT17 (MSMT) [10] datasets with standard Re-ID metrics (R1 and mAP).

[5] Martin Ester et al. "A density-based algorithm for discovering clusters in large spatial databases with noise.". In: *Kdd*. 1996.

Experimental Setup

- CANU-SSG and CANU-MMT are evaluated.
- The clustering algorithm used is DBSCAN[5].
- The strategies are evaluated using Market-1501 (Mkt) [12], DukeMTMC-reID (Duke) [9] and MSMT17 (MSMT) [10] datasets with standard Re-ID metrics (R1 and mAP).

[5] Martin Ester et al. "A density-based algorithm for discovering clusters in large spatial databases with noise.". In: *Kdd*. 1996.

Experimental Setup

- CANU-SSG and CANU-MMT are evaluated.
- The clustering algorithm used is DBSCAN[5].
- The strategies are evaluated using Market-1501 (Mkt) [12], DukeMTMC-reID (Duke) [9] and MSMT17 (MSMT) [10] datasets with standard Re-ID metrics (R1 and mAP).

[5] Martin Ester et al. "A density-based algorithm for discovering clusters in large spatial databases with noise.". In: *Kdd*. 1996.

Comparison with State of the Art

Table 1: CANU on the Mkt ▶ Duke and Duke ▶ Mkt settings.

Method	Mkt ▶ Duke		Duke ▶ Mkt	
	R1	mAP	R1	mAP
PUL [3]	30.0	16.4	45.5	20.5
SPGAN [1]	41.1	22.3	51.5	22.8
Co-teaching [7]	77.6	61.7	87.8	71.7
SSG [4]	73.0	53.4	80.0	58.3
CANU-SSG (ours)	76.1 (+3.1)	57.0 (+3.6)	83.3 (+3.3)	61.9 (+3.6)
MMT [6]	80.2	67.2	91.7	79.3
CANU-MMT (ours)	**83.3** (+3.1)	**70.3** (+3.1)	**94.2** (+2.5)	**83.0** (+3.7)

Comparison with State of the Art

Table 2: CANU on the Mkt ▶ MSMT and Duke ▶ MSMT settings.

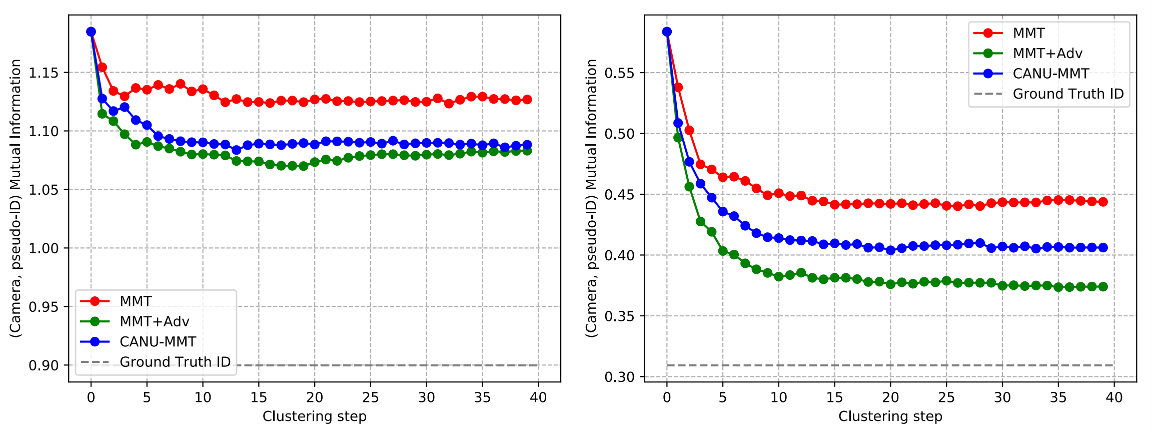
Method	Mkt ▶ MSMT		Duke ▶ MSMT	
	R1	mAP	R1	mAP
PTGAN [11]	10.2	2.9	11.8	3.3
ENC [13]	25.3	8.5	30.2	10.2
SSG [4]	31.6	13.2	32.2	13.3
CANU-SSG (ours)	45.5 (+13.9)	19.1 (+5.9)	43.3 (+11.1)	17.9 (+4.6)
MMT [6]	51.6	26.6	59.0	32.0
CANU-MMT (ours)	**61.7** (+10.1)	**34.6** (+8.0)	**66.9** (+7.9)	**38.3** (+6.3)

Camera adversarial vs Conditional camera adversarial

Table 3: Impact of the conditional strategy on baselines. When the mAP values are equal, we highlight the one corresponding to higher R1.

Method	Mkt ▶ Duke		Duke ▶ Mkt	
	R1	mAP	R1	mAP
SSG [4]	73.0	53.4	80.0	58.3
SSG+Adv.	75.4	56.4	**83.8**	**62.7**
CANU-SSG	**76.1**	**57.0**	83.3	61.9
MMT [6]	80.2	67.2	91.7	79.3
MMT+Adv.	82.6	70.3	93.6	82.2
CANU-MMT	**83.3**	**70.3**	**94.2**	**83.0**

Camera & Pseudo-ID dependancy analysis



(a) Mkt ▶ Duke   **(b)** Duke ▶ Mkt

Figure 1: Mutual information between pseudo labels and camera index evolution for the MMT baseline. Ground-truth ID comparison is displayed in dashed lines for both datasets.

Conclusion

Conclusion

Merge **finetuning and clustering** with a **camera-based adversarial** strategy, which can be plugged into **any unsupervised approach**.

Solve the **negative transfer** problem with a conditioned approach.

Demonstrate its performance on **two state of the art methods**.

Conclusion

Merge **finetuning and clustering** with a **camera-based adversarial** strategy, which can be plugged into **any unsupervised approach**.

Solve the **negative transfer** problem with a conditioned approach.

Demonstrate its performance on **two state of the art methods**.

Conclusion

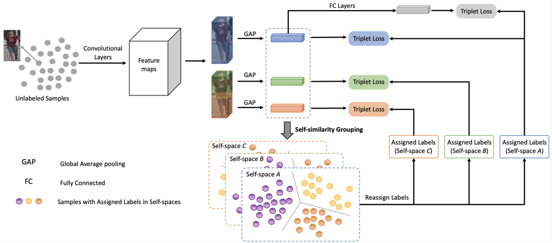
Merge **finetuning and clustering** with a **camera-based adversarial** strategy, which can be plugged into **any unsupervised approach**.

Solve the **negative transfer** problem with a conditioned approach.

Demonstrate its performance on **two state of the art methods**.

Clustering and Finetuning - examples

- **Self-similarity grouping** (SSG)[6] clusters on 3 visual subdomains (full body, upper/lower body),and rely on self-consistency to reduce clustering mistakes.
- **Mutual mean-teaching** (MMT)[7] uses teacher-student models, trained with hard pseudo-ID based loss and soft losses supervised by each other's predictions.

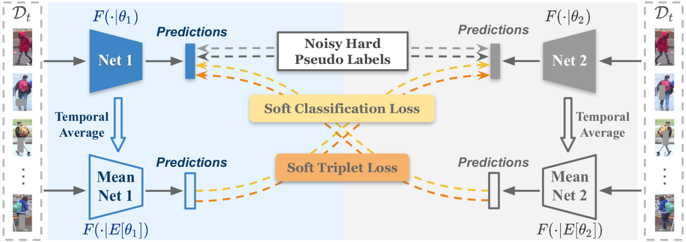


[6] Yang Fu et al. "Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification". In: *IEEE ICCV*. 2019.

[7] Yixiao Ge, Dapeng Chen, and Hongsheng Li. "Mutual Mean-Teaching: Pseudo Label Refinery for Unsupervised Domain Adaptation on Person Re-identification". In: *ICLR* (2020).

Clustering and Finetuning - examples

- **Self-similarity grouping** (SSG)[6] clusters on 3 visual subdomains (full body, upper/lower body),and rely on self-consistency to reduce clustering mistakes.
- **Mutual mean-teaching** (MMT)[7] uses teacher-student models, trained with hard pseudo-ID based loss and soft losses supervised by each other's predictions.



[6]Yang Fu et al. "Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification". In: *IEEE ICCV*. 2019.

[7]Yixiao Ge, Dapeng Chen, and Hongsheng Li. "Mutual Mean-Teaching: Pseudo Label Refinery for Unsupervised Domain Adaptation on Person Re-identification". In: *ICLR* (2020).