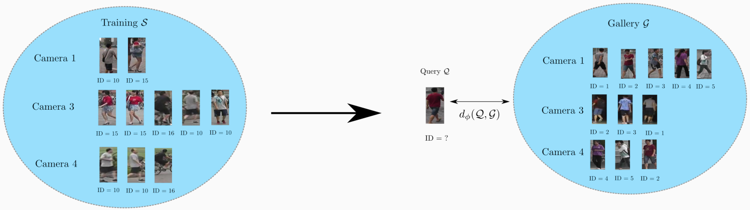
CANU-ReID: A Conditional Adversarial Network for Unsupervised person Re-IDentification

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

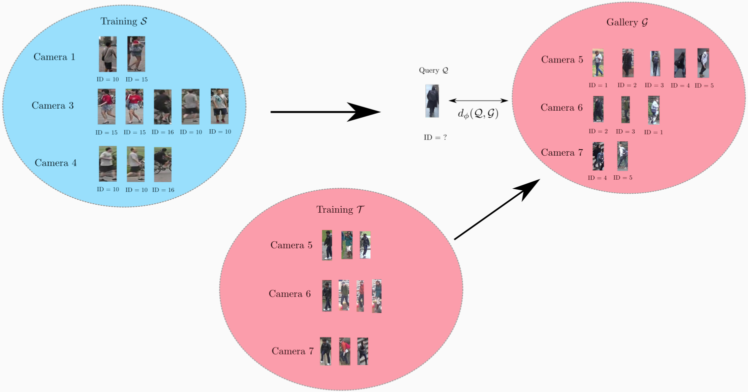
Person Re-Identification

Supervised/Unsupervised Person Re-Identification



- Supervised Re-ID: large annotated datasets and poor generalization.
- Unsupervised Re-ID: Labeled source $\mathcal{S}$, unlabeled target $\mathcal{T}$: optimizes re-ID performance on $\mathcal{T}$.

Supervised/Unsupervised Person Re-Identification



- Supervised Re-ID: large annotated datasets and poor generalization.
- Unsupervised Re-ID: Labeled source $\mathcal{S}$, unlabeled target $\mathcal{T}$: optimizes re-ID performance on $\mathcal{T}$.

Related Work

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 on source S in a supervised setting.
Alternates between:

1 - Clustering step runs clustering on $\mathcal{T} \rightarrow$ pseudo-ID labels $\tilde{\boldsymbol{p}}^{\mathcal{T}}$.

2 - Finetuning step ϕ finetuned using $\tilde{\boldsymbol{p}}^{\mathcal{T}}$.

---

[1] Yang Fu et al. "Self-similarity grouping: A simple unsupervised cross domain adapt. approach for person re-ID". In: *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 on source S in a supervised setting.

Alternates between:

1 - Clustering step runs clustering on $\mathcal{T} \rightarrow$ pseudo-ID labels $\tilde{\boldsymbol{p}}^{\mathcal{T}}$.

2 - Finetuning step ϕ finetuned using $\tilde{\boldsymbol{p}}^{\mathcal{T}}$.

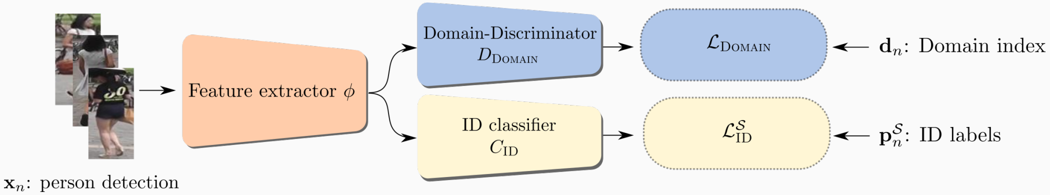
---

[1]Yang Fu et al. "Self-similarity grouping: A simple unsupervised cross domain adapt. approach for person re-ID". In: *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.

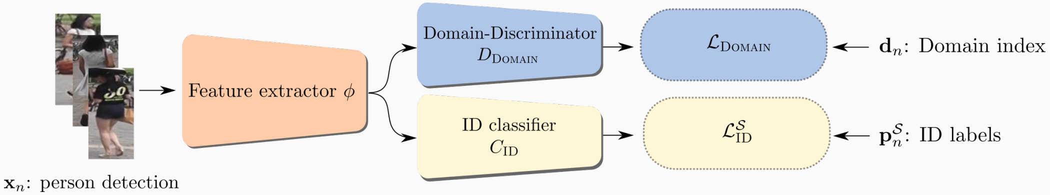


Domain Generalization[4]: Adversarial framework $\rightarrow$ *Negative Transfer*: discriminator learns **ID-related** instead of **domain-related features**. Happens when different ID prior distributions across domains.

[3] Yaroslav Ganin et al. "Domain-adversarial training of neural nets". In: *JMLR* (2016).
[4] Ya Li et al. "Deep Domain Generalization via Conditional Invariant Adversarial Networks". In: *ECCV*. 2018.

Adversarial Domain Adaptation

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



Domain Generalization[4]: Adversarial framework → *Negative Transfer*: discriminator learns **ID-related** instead of **domain-related features**. Happens when different ID prior distributions across domains.

[3] Yaroslav Ganin et al. "Domain-adversarial training of neural nets". In: *JMLR* (2016).
[4] Ya Li et al. "Deep Domain Generalization via Conditional Invariant Adversarial Networks". In: *ECCV*. 2018.

Contributions

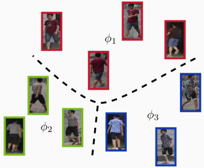
From this analysis we derive the following strategy:

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

Conditional Camera Adversarial Learning

Conditional camera adversarial training pipeline

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



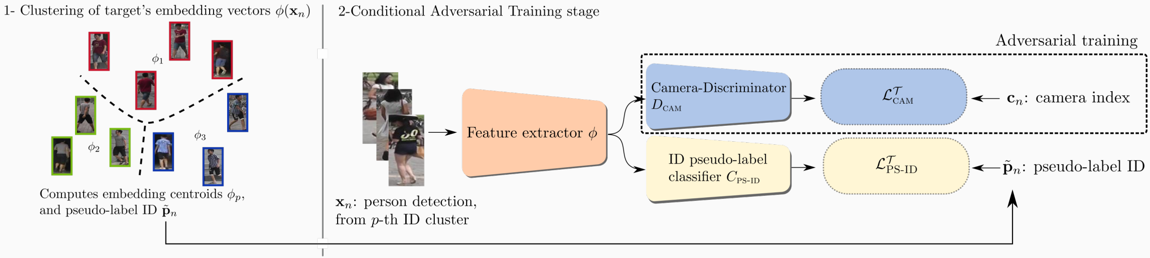
Computes embedding centroids ϕ_p,
and pseudo-label ID $\tilde{\mathbf{p}}_n$

Tackling Negative Transfer

$$\min_{\phi, C_{\text{PS-ID}}} \max_{D_{\text{C-CAM}}} \mathcal{L}^{\mathcal{T}}_{\text{PS-ID}}(\phi, C_{\text{PS-ID}}) - \mu \mathcal{L}^{\mathcal{T}}_{\text{C-CAM}}(\phi, D_{\text{C-CAM}}), \tag{1}$$

$$\mathcal{L}^{\mathcal{T}}_{\text{C-CAM}}(\phi, D_{\text{C-CAM}}) = -\mathbb{E}_{(\boldsymbol{x},\boldsymbol{p},\boldsymbol{c})\sim\mathcal{T}} \left\{ \log \left\langle D_{\text{C-CAM}}(\phi(\boldsymbol{x}), \boldsymbol{\phi_p}), \boldsymbol{c} \right\rangle \right\} \tag{2}$$

Conditional camera adversarial training pipeline

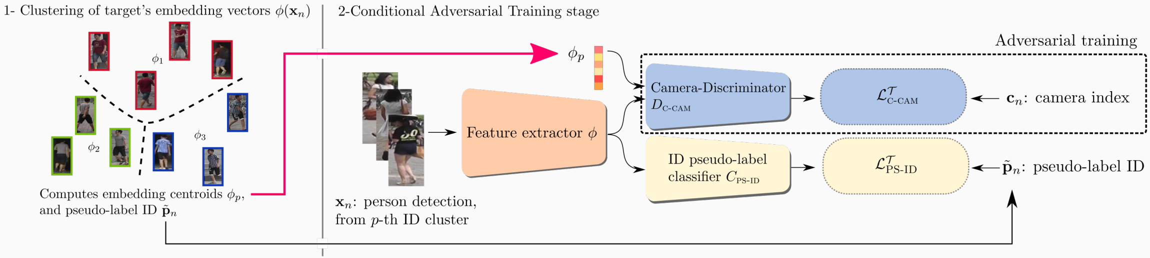


Tackling Negative Transfer

$$\min_{\phi, C_{\text{PS-ID}}} \max_{D_{\text{C-CAM}}} \mathcal{L}^{\mathcal{T}}_{\text{PS-ID}}(\phi, C_{\text{PS-ID}}) - \mu \mathcal{L}^{\mathcal{T}}_{\text{C-CAM}}(\phi, D_{\text{C-CAM}}), \qquad (1)$$

$$\mathcal{L}^{\mathcal{T}}_{\text{C-CAM}}(\phi, D_{\text{C-CAM}}) = -\mathbb{E}_{(x,p,c)\sim\mathcal{T}} \left\{ \log \left\langle D_{\text{C-CAM}}(\phi(x), \phi_p), c \right\rangle \right\} \qquad (2)$$

Conditional camera adversarial training pipeline



Tackling Negative Transfer

$$\min_{\phi, C_{\text{PS-ID}}} \max_{D_{\text{C-CAM}}} \mathcal{L}^{\mathcal{T}}_{\text{PS-ID}}(\phi, C_{\text{PS-ID}}) - \mu \mathcal{L}^{\mathcal{T}}_{\text{C-CAM}}(\phi, D_{\text{C-CAM}}), \tag{1}$$

$$\mathcal{L}^{\mathcal{T}}_{\text{C-CAM}}(\phi, D_{\text{C-CAM}}) = -\mathbb{E}_{(\boldsymbol{x},\boldsymbol{p},\boldsymbol{c})\sim\mathcal{T}} \left\{ \log \left\langle D_{\text{C-CAM}}(\phi(\boldsymbol{x}), \boldsymbol{\phi_p}), \boldsymbol{c} \right\rangle \right\} \tag{2}$$

Camera adversarial training

Advantages:

- Can be **plugged into any** clustering and finetuning strategy $\rightarrow$ CANU-MMT, CANU-SSG
- Explicitely **reduces errors** in pseudo-ID labels by making embedding space **invariant to camera information**

Camera adversarial training

Advantages:

- Can be **plugged into any** clustering and finetuning strategy $\rightarrow$ CANU-MMT, CANU-SSG
- Explicitely **reduces errors** in pseudo-ID labels by making embedding space **invariant to camera information**

Experimental Evaluation

Experimental Setup

- Evaluated using Market-1501 (Mkt), DukeMTMC-reID (Duke) and MSMT17 (MSMT) datasets.
- Standard Re-ID metrics (R1$\uparrow$ and mAP$\uparrow$) reported.

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
SSG [1]	73.0	53.4	80.0	58.3
CANU-SSG (ours)	**76.1**	**57.0**	**83.3**	**61.9**
MMT [2]	80.2	67.2	91.7	79.3
CANU-MMT (ours)	**83.3**	**70.3**	**94.2**	**83.0**

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

Method	Mkt ▶ MSMT		Duke ▶ MSMT	
	R1	mAP	R1	mAP
SSG [1]	31.6	13.2	32.2	13.3
CANU-SSG (ours)	**45.5**	**19.1**	**43.3**	**17.9**
MMT [2]	51.6	26.6	59.0	32.0
CANU-MMT (ours)	**61.7**	**34.6**	**66.9**	**38.3**

Conclusion

Conclusion

Merges **finetuning and clustering** with a **camera-based adversarial** strategy.

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

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

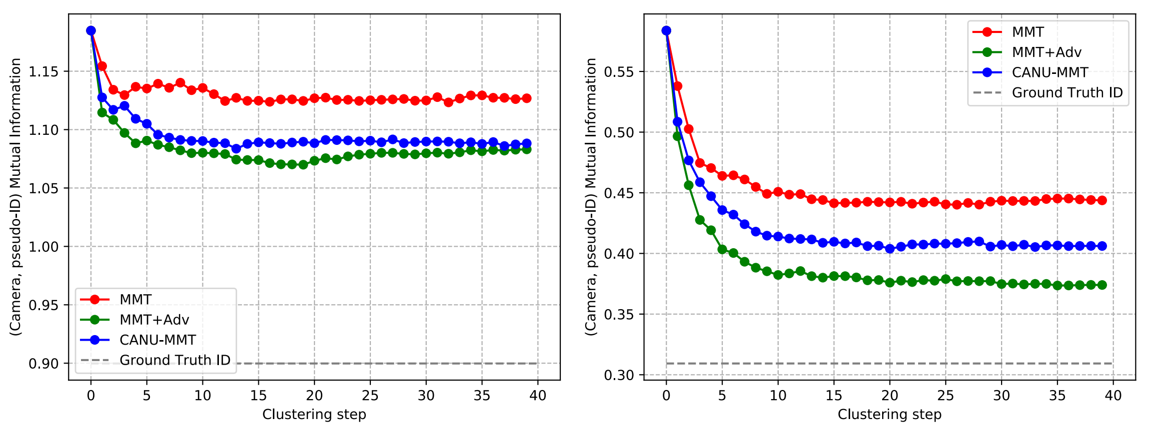
Supplementary Material

Camera adversarial vs Conditional camera adversarial

Table 3: Impact of the conditional strategy on baselines.

Method	Mkt ▶ Duke		Duke ▶ Mkt	
	R1	mAP	R1	mAP
SSG [1]	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 [2]	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

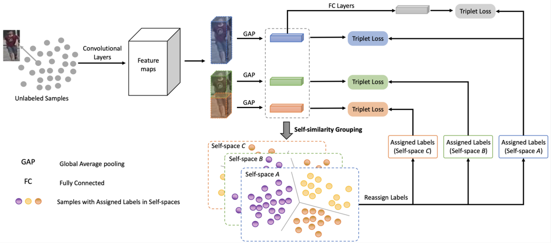


(a) Mkt ▶ Duke **(b)** Duke ▶ Mkt

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.

Clustering and Finetuning - examples

- **Self-similarity grouping** (SSG)[5] clusters on 3 visual subdomains (full body, upper/lower body),and rely on self-consistency to reduce clustering mistakes.
- **Mutual mean-teaching** (MMT)[6] uses teacher-student models, trained with hard pseudo-ID based loss and soft losses supervised by each other's predictions.

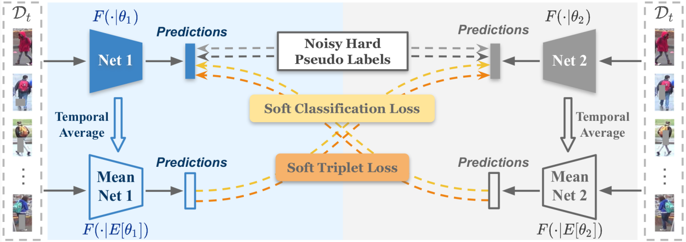


[5] Yang Fu et al. "Self-similarity grouping: A simple unsupervised cross domain adapt. approach for person re-ID". In: *ICCV*. 2019.

[6] 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)[5] clusters on 3 visual subdomains (full body, upper/lower body),and rely on self-consistency to reduce clustering mistakes.
- **Mutual mean-teaching** (MMT)[6] uses teacher-student models, trained with hard pseudo-ID based loss and soft losses supervised by each other's predictions.



[5]Yang Fu et al. "Self-similarity grouping: A simple unsupervised cross domain adapt. approach for person re-ID". In: *ICCV*. 2019.

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

Bibtex

@inproceedings{delorme:hal-02882285,
TITLE = {CANU-ReID: A Conditional Adversarial Network for Unsupervised person
Re-IDentification},
AUTHOR = {Delorme, Guillaume and Xu, Yihong and Lathuiliere, Stephane and
Horaud, Radu and Alameda-Pineda, Xavier},
URL = {https://hal.inria.fr/hal-02882285},
BOOKTITLE = {International Conference on Pattern Recognition},
ADDRESS = {Milano, Italy},
YEAR = {2021},
MONTH = Jan,
PDF = {https://hal.inria.fr/hal-02882285/file/delorme_icpr2020.pdf},
}