# Deep Compact Person Re-IDentification with Distractor Synthesis via DC-GANs

**Víctor Ponce-López*, Tilo Burdghart, Yue Sun, Sion Hannuna, Dima Damen, Majid Mirmehdi**

*Dept. of Computer Science, Electric & Electronic Engineering, University of Bristol.*

vponcelop@gmail.com ; v.poncelopez@bristol.ac.uk

VILab
Visual Information Laboratory

University of BRISTOL

IAPR

## ABSTRACT

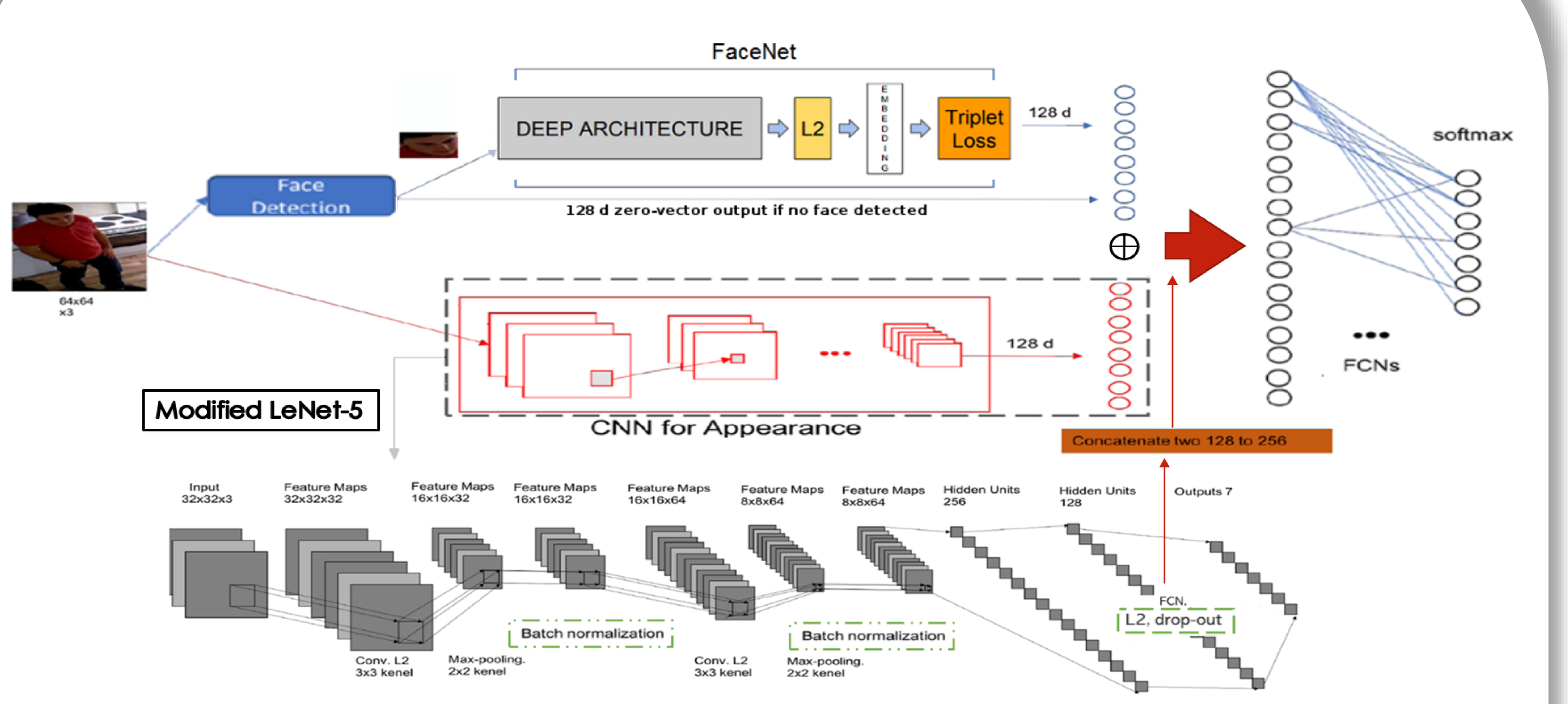
- Dual-Stream CNN learns both appearance and facial features in tandem from still images and infers person IDs.
- We leverage an alternative lightweight ID-CondenseNet architecture that integrates a face-guided DC-GAN to generate distractor person images for enhanced training.
- Both architectures are tested on FLIMA, a new extension of an existing person Re-ID dataset with added frame-by-frame annotations of face presence. We outperform the largest existing Re-ID dataset, MSMT17.

## 1. Introduction

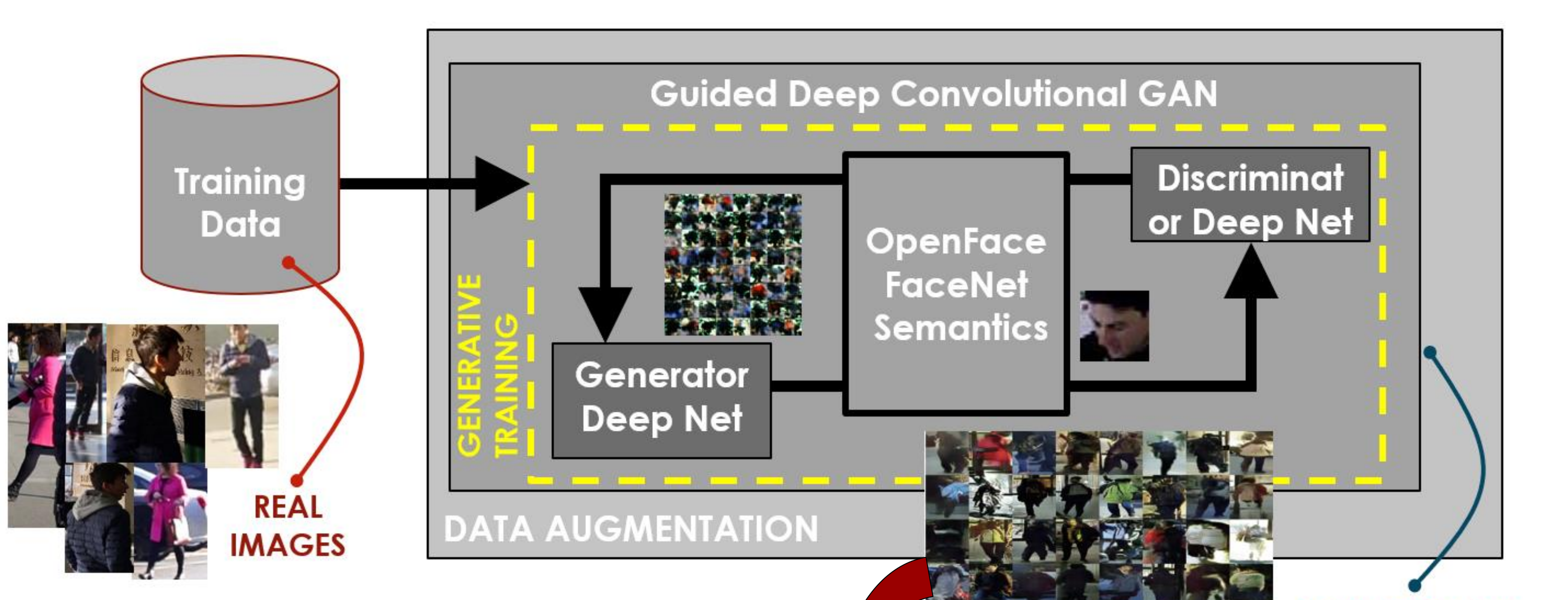
- Visual Person Re-ID links people across disjoint views.
- Challenging sub-domain in Computer Vision:
  - **Inherent viewpoint** and **illumination** changes, **partial occlusions**, **limitations** on **resolution**, significant **appearance alterations** (e.g. **changes** in **clothing**).
  - **Unimodal approaches**, such as **face recognition** systems, are *on their own* inadequate.
  - **Computational demand** for **network inference**.
- Methods evaluated on novel released datasets.

## 2. Proposed Method

### Dual-Stream Baseline




### Guided DC-GAN Compact Architecture




GAN and DC-GAN adversarial loss:

$$min_G \max_D V(D,G) = \mathbb{E}_{\mathbf{x}\sim p_{data}(\mathbf{x})}[\log(D(\mathbf{x}))] + \mathbb{E}_{\mathbf{z}\sim p_{\mathbf{z}}(\mathbf{z})}[\log(1-D(G(\mathbf{z})))]$$

Discriminators for real (**x**) and generated (**z**) images:

$$\mathcal{L}_{D_\mathbf{x}} = \frac{1}{m}\sum_{i=1}^{m}\left[\log\left(D\left(\mathbf{x}^{(\mathbf{i})}\right)\right)\right], \qquad \mathcal{L}_{D_\mathbf{z}} = \frac{1}{m}\sum_{i=1}^{m}\left[\log\left(1-D\left(G\left(\mathbf{z}^{(i)}\right)\right)\right)\right]$$

**Proposal**: Semantically modified discriminator losses for real and generated images:

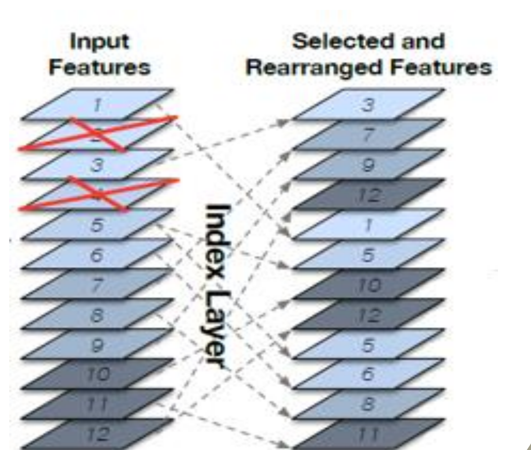
$$\mathcal{L}'_{D_\mathbf{x}} = \frac{1}{m}\sum_{i=1}^{m}\left[\log\left(D\left(\mathbf{x}^{(\mathbf{i})}\right)\right)\boxed{+\lambda_1\left(\Delta\left(\mathbf{x}^{(\mathbf{i})}\right)\right)}\right],$$

$$\mathcal{L}'_{D_\mathbf{z}} = \frac{1}{m}\sum_{i=1}^{m}\left[\log\left(1-D\left(G\left(\mathbf{z}^{(i)}\right)\right)\right)\boxed{-\lambda_2\left(\Delta\left(G\left(\mathbf{z}^{(\mathbf{i})}\right)\right)\right)}\right]$$

$\lambda_1, \lambda_2$ are penalisation factors.

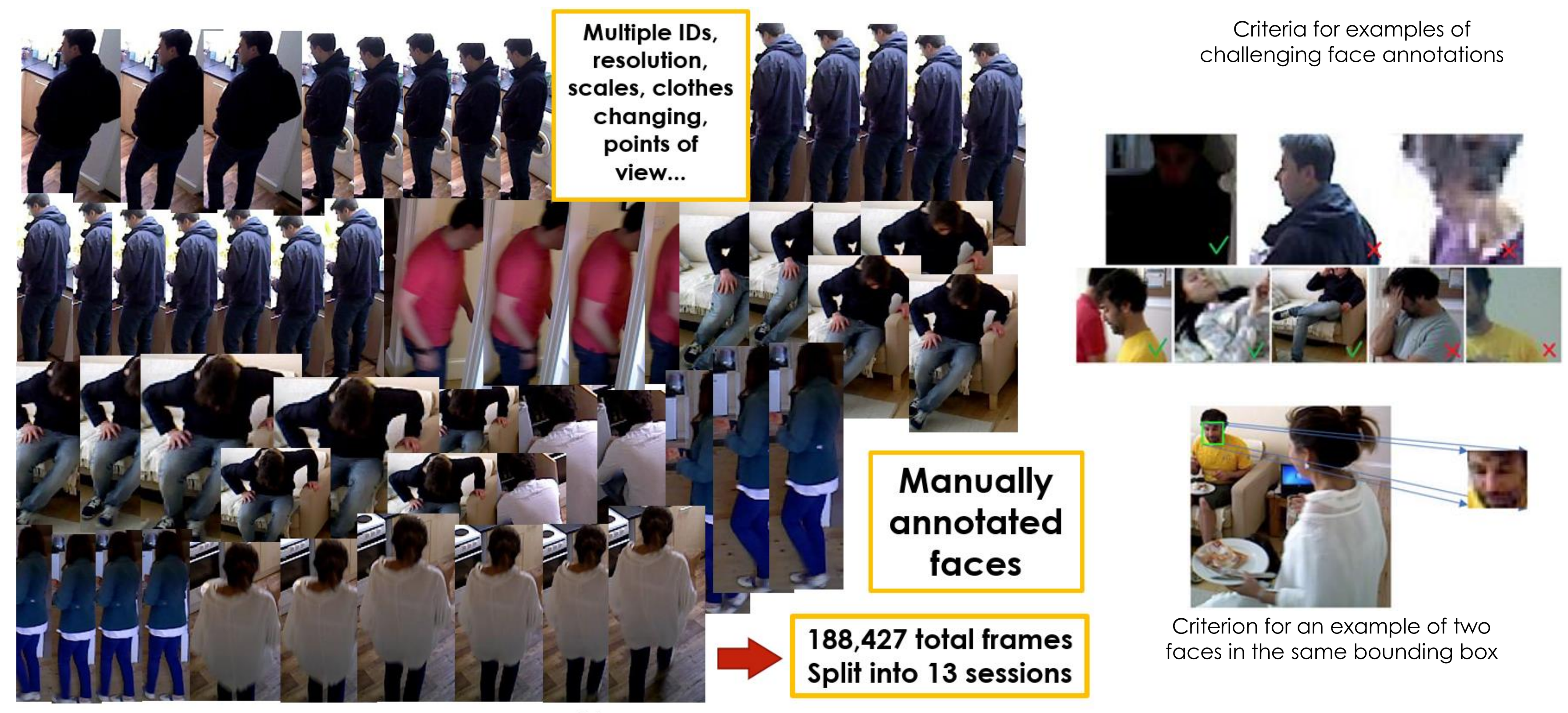
$\Delta(\cdot) = 1$ when there is no face detectable;
$\Delta(\cdot) = 0$ otherwise.

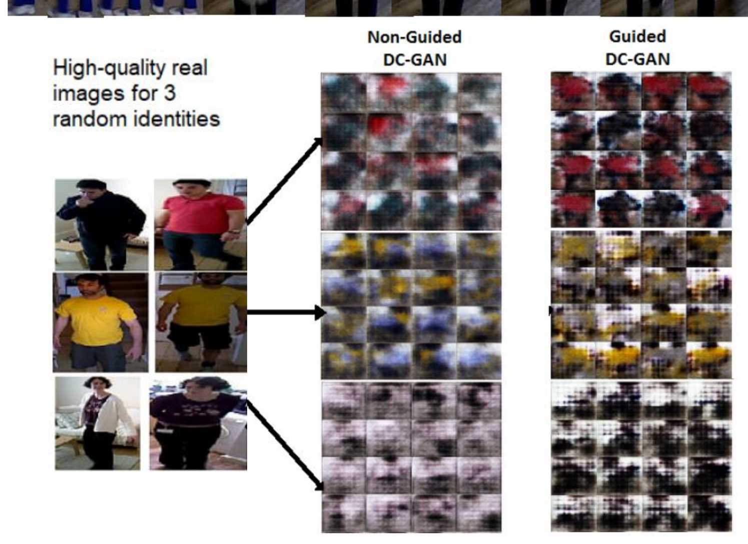
Discard – Forget unused weights




## 3. Experiments on FLIMA & MSMT17

### FACIAL Long-term Identity-aware Multi-target multi-camerA




Initial generated samples for 3 random identities – note cases of clothes changing.

Table 1. Recognition performance on FLIMA

| Method | # Test Images | Precision | Recall | F1-score |
|---|---|---|---|---|
| RCNN and RBF-SVM-Eigenface (Faces only) | 4,531 | 0.56 | 0.52 | 0.47 |
| Selective Augmentation Approach [19] | 14,494 | 0.75 | 0.74 | 0.74 |
| Our Guided DC-GAN trained CondenseNet | 14,494 | 0.85 | 0.85 | 0.85 |
| Our Appearance-Stream only | 14,494 | 0.92 | 0.81 | 0.86 |
| Our Full Dual-Stream | 14,494 | **0.93** | **0.90** | **0.91** |

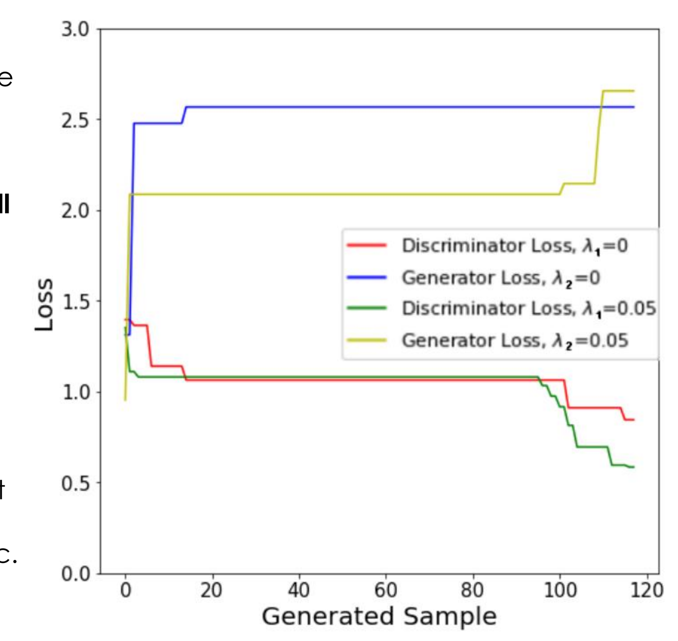
### Multi-Scene Multi Time 2017

Considerably **increased performance** for Recall with the **Dual-Stream** architecture.

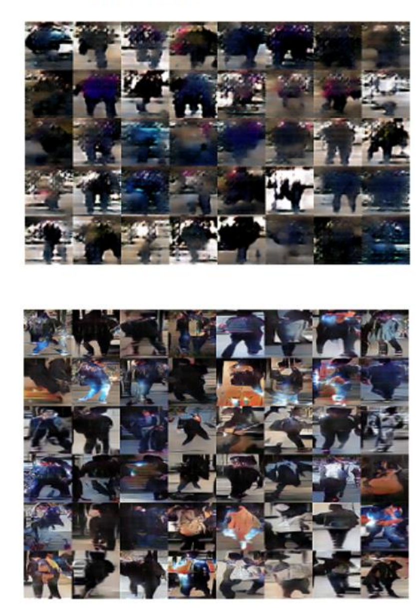
The Dual-Stream approach is advantageous for a **small number of individuals** and good **facial resolution**.

**DC-GAN trained CondenseNet** approach provides a significant improvement in **Rank@1 performance**.

4% performance increase above the next best performing method and 27% over GoogLeNet, without using expensive and time-consuming training of very-deep multi-stream networks that benefit the mAP metric.




Non-Guided : $\lambda_1, \lambda_2 = 0$



Guided : $\lambda_1, \lambda_2 = 0.05$

Table 2. Person Re-ID performance on MSMT17 dataset for single queries.

| Method | Rank@1 | mAP |
|---|---|---|
| Dual-Stream Architecture | 4.89 | 5.91 |
| GoogLeNet [25] | 47.6 | 23.0 |
| PDC [24] | 58.0 | 29.7 |
| GLAD [31] | 61.4 | **34.0** |
| Selective Augmentation Approach [19] | 61.5 | 15.01 |
| Our Guided DC-GAN ($\lambda_1, \lambda_2 = 0.05, 0.025$) trained CondenseNet | **63.85** | 16.64 |
| Our Guided DC-GAN ($\lambda_1, \lambda_2 = 0.05, 0$) trained CondenseNet | **65.51** | 18.57 |

Showing samples of **global distractors** for different values of $\lambda_1, \lambda_2 > 0$.

DC-GAN **Training on all identity samples** are used for the **generation** of distractors.

Note **improvements** when activating the **guidance**.

## 4. Conclusion

- Potential approaches for **person Re-ID** are based on the exploitation of **facial** and **person appearance** representations.
- **Guided DC-GAN** integrates the face detector, leveraged from the **face** stream of our **Dual-Stream CNN** architecture. It is used to **generate** person images for enhanced training.
- **Distractor augmentation** and **network compression** have a role to play for larger scale applications.

## References

[1] Z. Zheng, L. Zheng, Y. Yang. **Unlabeled samples generated by GAN improve the person re-identification baseline *in vitro***. International Conference on Computer Vision (ICCV), pages 3754–3762, 2017.
[2] G. Huang, S. Liu, L.-v.-d. Maaten, K,-Q. Weinberger. **CondenseNet: An Efficient DenseNet using Learned Group Convolutions**. arXiv preprint arXiv:1711.09224, 2017.
[3] A. Radford, L. Metz, S. Chintala. **Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks**. International Conference on Learning Representations (ICLR), 2015.
[19] V. Ponce-López, T. Burdghart, S. Hannuna, D. Damen, A. Masullo, M. Mirmehdi. **Semantically Selective Augmentation for Deep Contact Person Re-Identification**. ECCV Workshops, LNCS, vol. 11130, 2019.
[24] C. Su, J. Li, S. Zhang, J. Xing, W. Gao, Q. Tian. **Pose-driven Deep Convolutional Model for Person Re-identification**. International Conference on Computer Vision (ICCV), pp. 3960-3969, 2017.
[25] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich. **Going deeper with convolutions**. Computer Vision and Pattern Recognition (CVPR), pp. 1-9, 2015.
[31] L. Wei, S. Zhang, H. Yao, W. Gao, Q. Tian. **GLAD: Global–Local-Alignment Descriptor for Scalable Person Re-Identification.** IEEE Transactions on Multimedia, vol. 21(4), pp. 986-999, 2019.

EPSRC
Engineering and Physical Sciences Research Council

sphere