Instance-Guided Context Rendering for Cross-Domain Person Re-Identification

Yanbei Chen[1] Xiatian Zhu[2] Shaogang Gong[1]
yanbei.chen@qmul.ac.uk eddy.zhuxt@gmail.com s.gong@qmul.ac.uk
[1]Queen Mary University of London [2]Vision Semantics Ltd., London, UK

ICCV 2019 Seoul, Korea

Queen Mary University of London

Introduction

Problem

- ❑ Person re-identification aims at searching person identities across cameras distributed over open surveillance space.

Motivation

- ❑ There exists inevitable domain gaps between datasets collected from different surveillance camera networks
- ❑ Our approach aims to hallucinate the same persons in diverse surveillance contexts, as if they were captured from different places and times in the target domain.

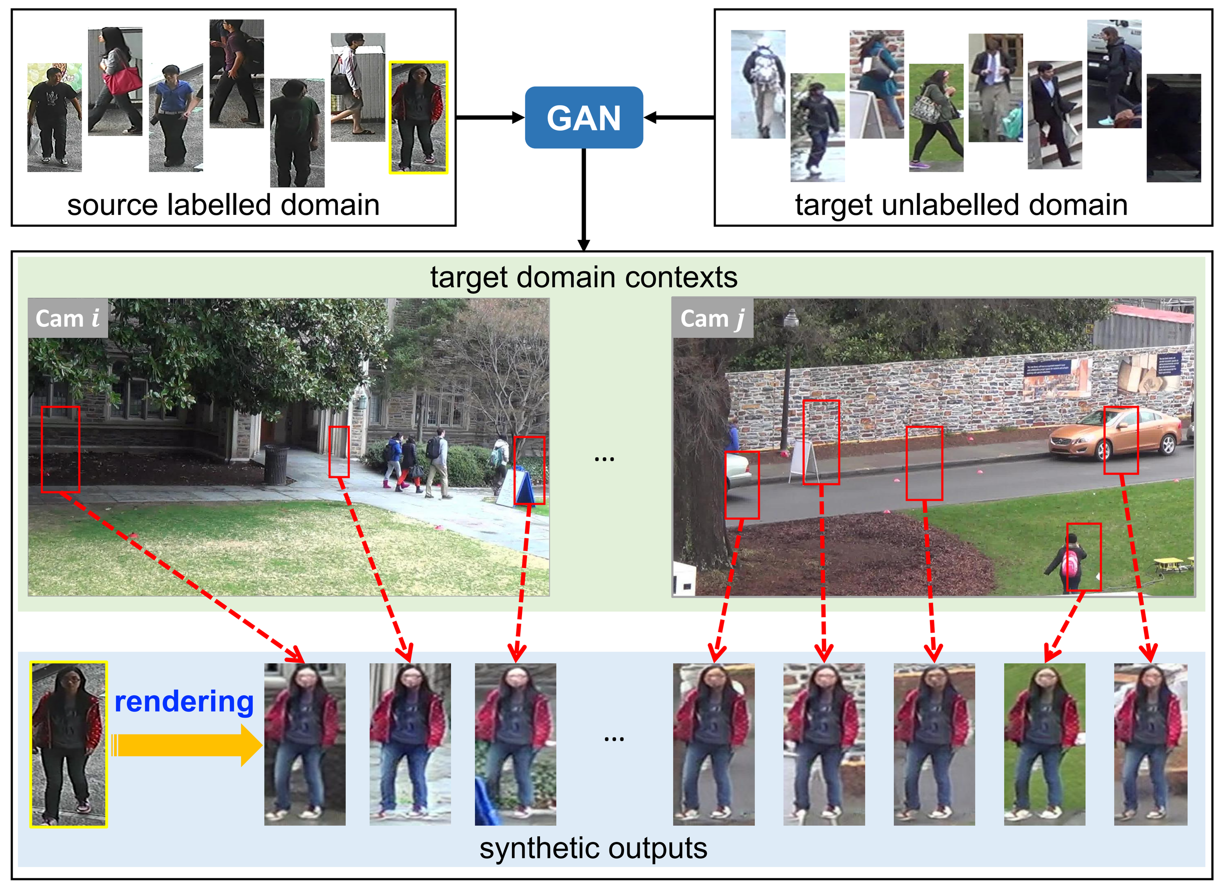


Figure 1. Motivation illustration: In open surveillance spaces, the *contextual variations* are quite diverse, due to ***wide-of-the-field imagery*** and ***varying times of the day.***

Main Idea

I. render the same person identities into diverse domain contexts to produce abundant synthetic training data

II. deploy the synthetic data for CNN training

A naïve alternative in image synthesis

- ❑ produce abundant synthetic training data under varying background contexts by "*cut and paste*"

- ❑ It is unsatisfactory because

I. various artifacts are introduced (e.g. missing identity related cues due to an incomplete person segmentation mask)

II. cannot capture domain drift in *colour tones* and *illuminations*

Methodology

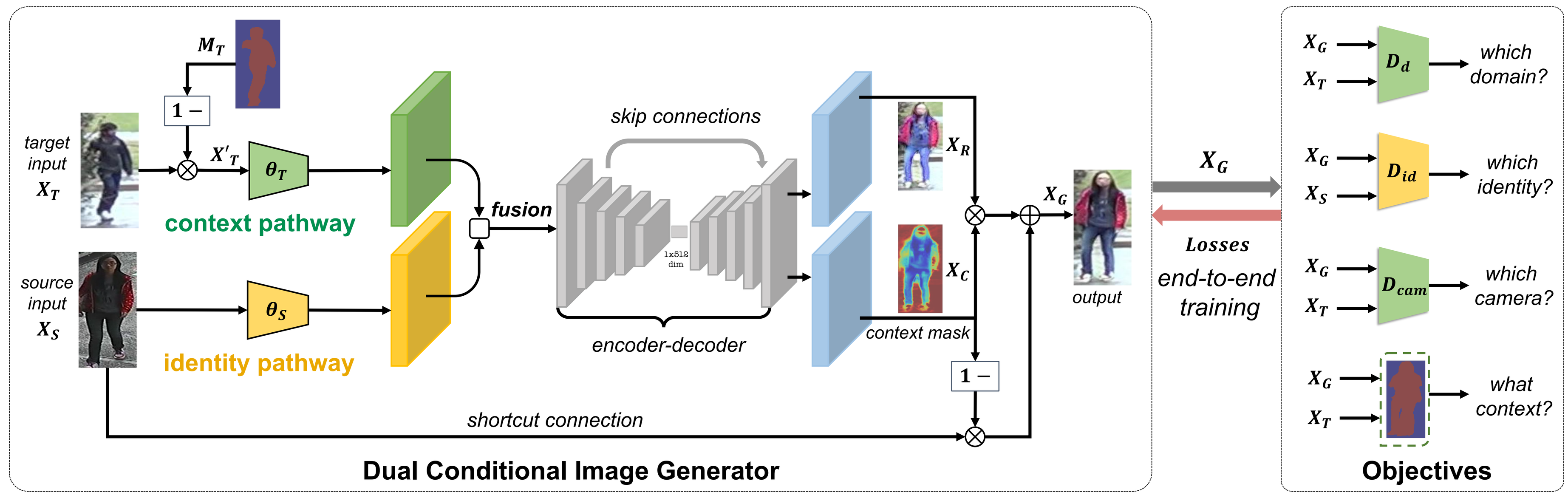


Figure 2. Model overview: We tackle the image-level domain drift by learning to render the source person image X_S into diverse domain contexts guided by arbitrary instances X_T from target domain.

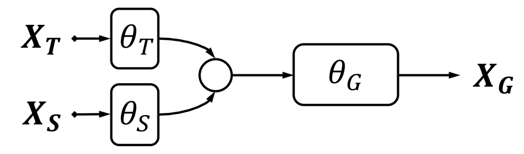
Figure 3. Deployment overview: The generator produces abundant synthetic images for CNN training.

Context Rendering Generative Adversarial Networks (CR-GAN)

❑ Dual Conditional Image Generator

- **Dual-Path Encoding:**
- A pair of source and target images as dual conditional inputs

$$X_G = G(X_S, X_T)$$



- Image Generator = U-Net + skip connection + residual blocks

❑ Learning Objectives

I. which domain? – Adversarial Loss.

$$\mathcal{L}_{\text{adv}} = \min_{G} \max_{D_d} \log D_d(X_T) + \log(1 - D_d(G(X_S, X_T)))$$

II. which identity? – Identity Loss.

$$\mathcal{L}_{\text{id}} = -\log(p(y_j | X_G))$$

III. which camera? – Camera Loss.

$$\mathcal{L}_{\text{cam}} = -\log(p(y_c | X_G))$$

IV. what context? – Context Loss.

$$\mathcal{L}_{\text{con}} = ||(X_G - X_S) \circ M_F||_2 + ||(X_G - X_T) \circ M_B||_2$$

CR-GAN Deployment for Cross-Domain Re-id Model Learning

- CR-GAN: generate abundant synthetic training data on-the-fly
- CNN (re-id model): train upon the synthetic data

Experiments & Ablation Studies

Experiments on re-id benchmark datasets

❑ Qualitative evaluation:

- CR-GAN augments the same person with diverse contexts explicitly guided by instances from the target domain.

Figure 4. Qualitative results I.

- CR-GAN renders the same person into different styles, i.e. varying *background clutters, colour tones* and *lighting conditions*.

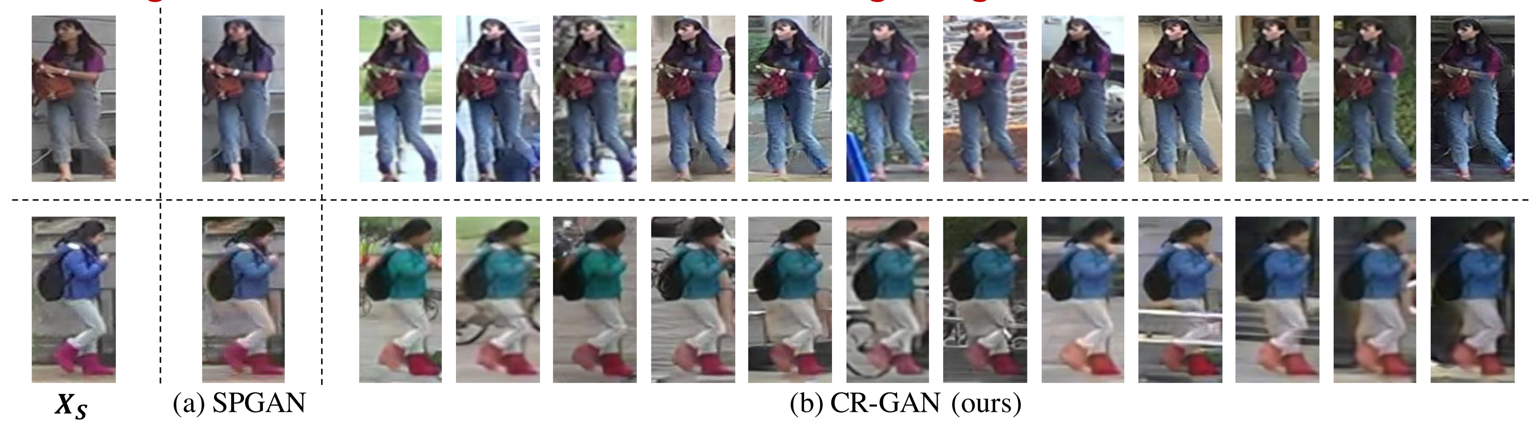


Figure 5. Qualitative results II.

❑ Quantitative evaluation:

- **Visual quality:** higher diversity and better fidelity (LPIPS↑, FID↓)
- **Re-id performance:** synthetic data boosts cross-domain re-id results.

S → T	Market→Duke		Duke→Market	
Metrics	LPIPS	FID	LPIPS	FID
Source-Target data	0.458	0.330	0.458	0.330
SPGAN [10]	0.099	0.171	0.099	0.115
CR-GAN	**0.281**	**0.058**	**0.269**	**0.096**

Table 1. Evaluation on visual quality.

Types	Source → Target	Market1501 → DukeMTMCreID				DukeMTMCreID → Market1501			
	Metrics (%)	R1	R5	R10	mAP	R1	R5	R10	mAP
Shallow	LOMO [31]	12.3	21.3	26.6	4.8	27.2	41.6	49.1	8.0
	BoW [62]	17.1	28.8	34.9	8.3	35.8	52.4	60.3	14.8
	UMDL [39]	18.5	31.4	37.6	7.3	34.5	52.6	59.6	12.4
Image	PTGAN [56]	27.4	-	50.7	-	38.6	-	66.1	-
	SPGAN+LMP [10]	46.4	62.3	68.0	26.2	57.7	75.8	82.4	26.7
	M2M-GAN+LMP [30]	54.4	-	-	31.6	63.1	-	-	30.9
	CR-GAN+LMP	**56.0**	**70.5**	**74.6**	**33.3**	**64.5**	**79.8**	**85.0**	**33.2**
Feature	PUL* [12]	30.0	43.4	48.5	16.4	45.5	60.7	66.7	20.5
	TJ-AIDL† [54]	44.3	59.6	65.0	23.0	58.2	74.8	81.1	26.5
	MMFA† [32]	45.3	59.8	66.3	24.7	56.7	75.0	81.8	27.4
	BUC* [33]	47.4	62.6	68.4	27.5	**66.2**	79.6	84.5	38.3
	TAUDL* [27]	**61.7**	-	-	**43.5**	63.7	-	-	41.2
Hybrid	HHL [64]	46.9	61.0	66.7	27.2	62.2	78.8	84.0	31.4
	SPGAN+TAUDL	66.1	80.0	83.2	47.2	66.5	81.8	86.6	38.5
	CR-GAN+TAUDL	**68.9**	**80.2**	**84.7**	**48.6**	**77.7**	**89.7**	**92.7**	**54.0**

Table 2. Evaluation on cross-domain re-id.

Reference

[1] Cut, paste and learn: Surprisingly easy synthesis for instance detection. ICCV2017
[2] Image-image domain adaptation with preserved self- similarity and domain-dissimilarity for person reidentification. CVPR2018
[3] CyCADA: Cycle-consistent adversarial domain adaptation. ICML2018
[4] Multimodal unsupervised image-to-image translation. ECCV2018