西安交通大學
XI'AN JIAOTONG UNIVERSITY

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

Polynomial Universal Adversarial Perturbations for Person Re-Identification

Presenter：Wenjie Ding

Wenjie Ding[1], Xing Wei[1], Rongrong Ji[2], Xiaopeng Hong[1], Yihong Gong[1]

[1]Faculty of Electronic and Information Engineering, Xi'an Jiaotong University

[2]Department of Artificial Intelligence, School of Information, Xia'men University

Poster

Introduction

- **Universal Adversarial Perturbation (UAP)**
  - Single fixed perturbation map that perturbs the images of interest.
  - No online optimization, implementation convenient
- **Problems:**
  - Simple constant form limits the attack power.
  - UAP attack on open-set tasks is challenging.

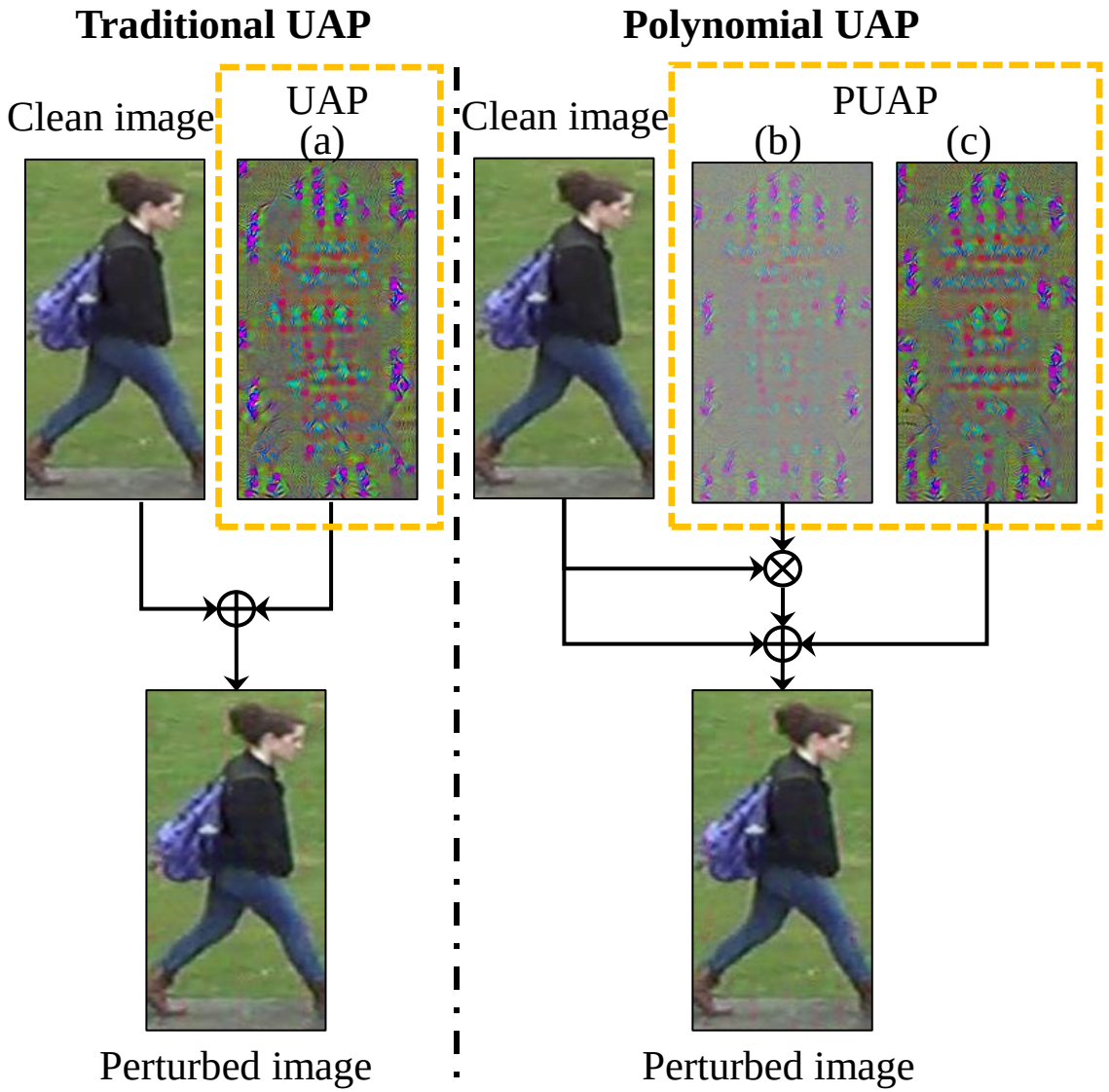
Main Contributions

a. Formulating the perturbation signal as a polynomial form for the first time

b. Provide a mechanism to control the energy of perturbation signal

c. Introduce a Pearson correlation coefficient (PCC) loss

Main Idea



Comparisons of traditional additive UAPs and the proposed polynomial UAPs.

⊕ Element-wise addition

⊗ Hadamard product

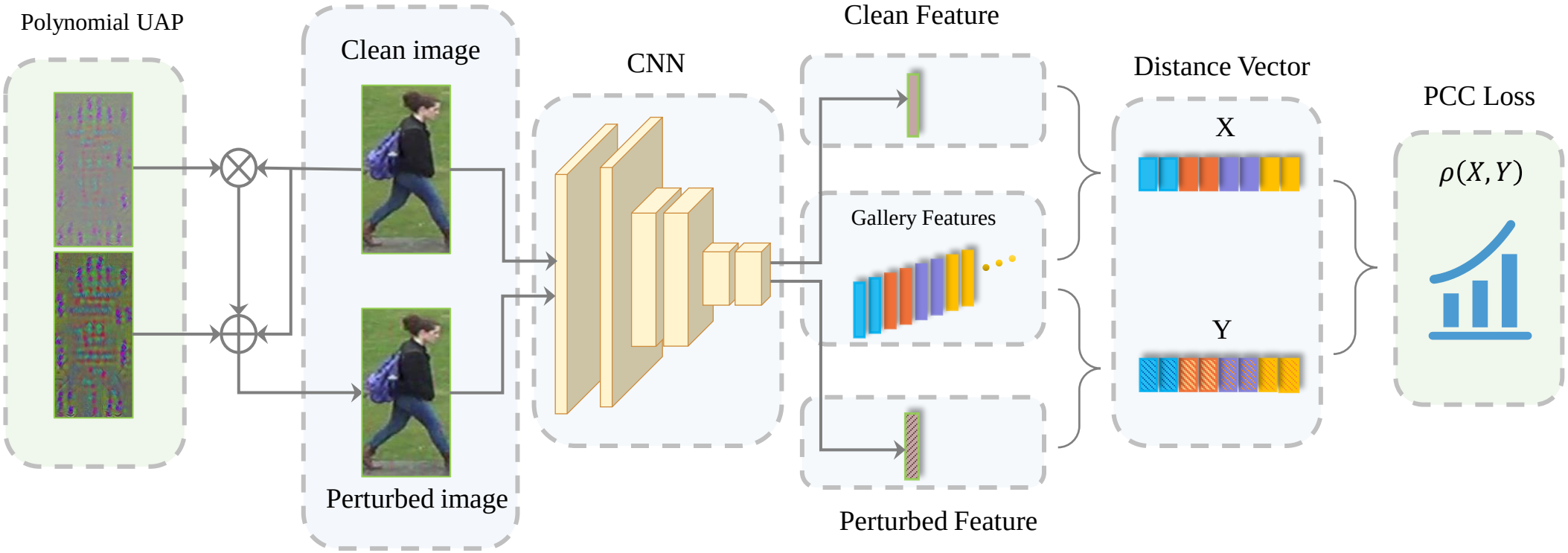
Main Idea

Traditional additive UAPs *Vs.* Polynomial UAPs:

$$\phi(x) \neq \phi(x+\sigma) \qquad \textit{Vs.} \qquad \phi(x) \neq \phi(\alpha x+\sigma)$$

ϕ represents the whole Re-ID system including the feature extraction and the sorting operation. σ represents the (traditional) additive perturbation and α represents the multiplicative factor.

Framework of the proposed method



1. Different colors in the features stand for different IDs.
2. Texture-rendered feature vector indicate the attack.
3. The model weights are fixed and only the perturbation signals are updated during training.

Mathematical form of the framework

Polynomial UAPs:

$$\phi(x) \neq \phi(\alpha x + \sigma)$$

ϕ represents the whole Re-ID system including the feature extraction and the sorting operation. $\boldsymbol{\sigma}$ represents the (traditional) additive perturbation and $\boldsymbol{\alpha}$ represents the multiplicative factor.

Pearson Correlation Coefficient Loss:

$$\mathcal{L}_{PCC} = \frac{\sum_{i=1}^{N}(a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_{i=1}^{N}(a_i - \bar{a})^2}\sqrt{\sum_{i=1}^{N}(b_i - \bar{b})^2}},$$

$$a_i = d(F_\theta(x^q), f_i^g), \quad b_i = d(F_\theta(x^{q\prime}), f_i^g), \quad x^{q\prime} = \alpha x^q + \sigma,$$

$\boldsymbol{a_i}$ and $\boldsymbol{b_i}$ are distance vectors. $\boldsymbol{x^q}$ and $\boldsymbol{x^{q\prime}}$ denotes the clean query image and the corresponding perturbed ones.

Mathematical form of the framework

Perturbation Energy Control:

$$\begin{cases} \alpha \leftarrow \arg\min_{\alpha'} \|\alpha - \alpha'\|_p, \ \ s.t. \| \dfrac{x^q}{\|x^q\|_p} \odot \alpha' \|_p \leq \epsilon\lambda, \\ \sigma \leftarrow \arg\min_{\sigma'} \|\sigma - \sigma'\|_p, \ \ s.t. \|\sigma'\|_p \leq \epsilon(1-\lambda), \end{cases}$$

$\odot$ represents Hadamard product. λ represents the percentage of the multiplicative term. ϵ denotes the total magnitude constrain of perturbations.

Experiments

Datasets: DukeMTMC-reID, Market-1501, MARS,

Evaluation Metrics : mean Drop Rate (mDR)

$$mDR = \frac{mAP(x^q) - mAP(x^{q'})}{mAP(x^q)}$$

Implementation Details:

- **Models:** ResNet50, DenseNet121, VGG16, SeNet154, ShuffleNet trained with BoT.
- **Comparative methods:** IR-UAP, GD-UAP, I-FGSM
- **Experimental setting:** We randomly choose 800 images for image-based datasets and 1000 tracklets for video-based dataset during perturbation training. λ is set to 0.1. During test, we project the perturbation on the corresponding l2 ball to satisfy the constrain of total magnitude.

Results

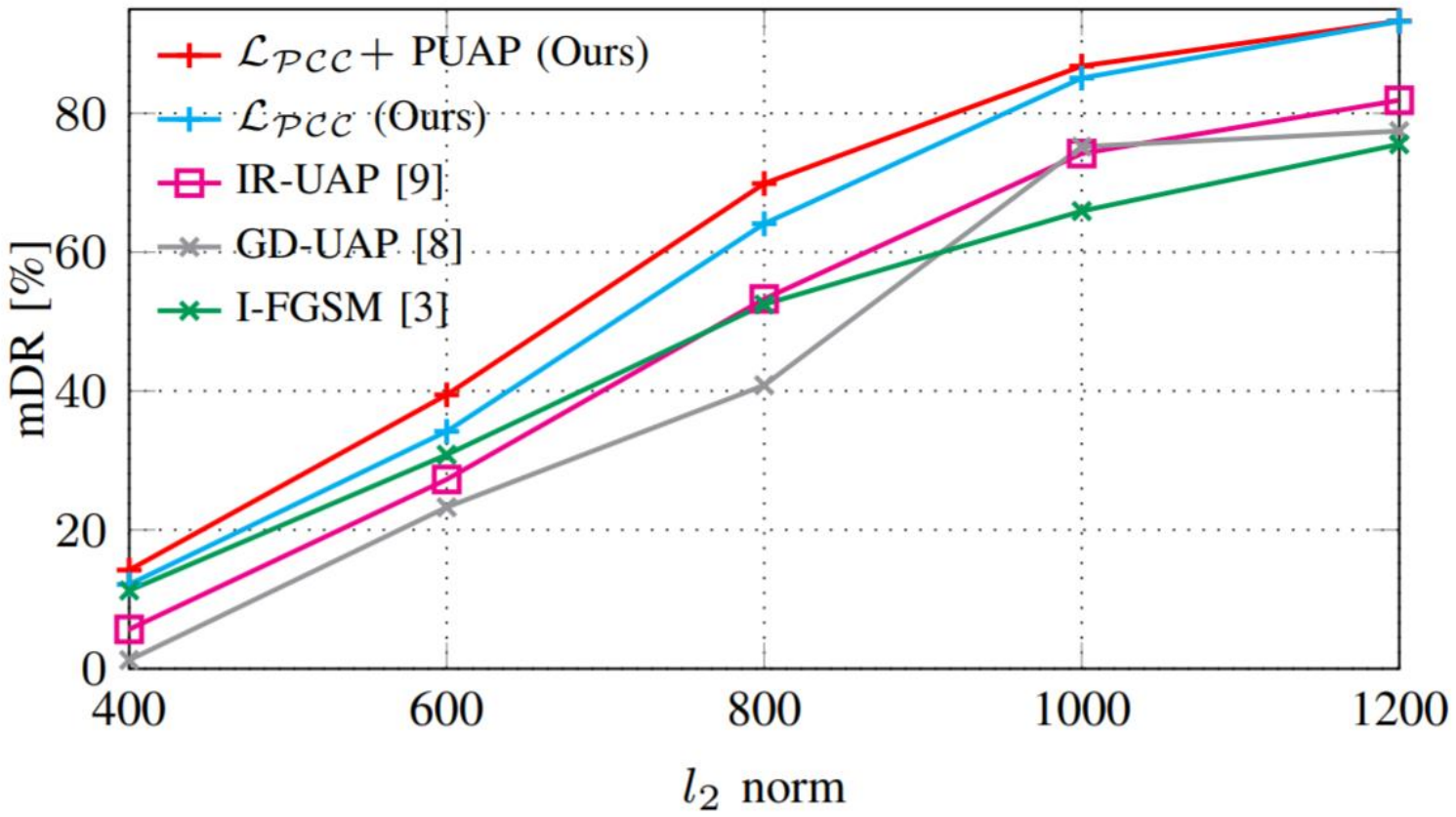


Figure: The attack performance comparison with state-of-the-art methods on DukeMTMC-reID. The red and cyan lines describe the performance of the proposed methods.

Results

mean Drop Rate (mDR)		Market-1501					DukeMTMC-ReID				
		400	600	800	1000	1200	400	600	800	1000	1200
ResNet50	$\mathcal{L}_{Base}$	11.47	27.81	48.90	65.58	79.18	14.66	35.82	52.47	65.92	75.54
	$\mathcal{L}_{PCC}$	11.66	37.49	66.04	83.88	92.55	**14.81**	34.18	64.10	85.07	93.24
	+PUAP	**15.07**	**44.18**	**72.46**	**87.65**	**95.31**	14.21	**39.45**	**69.88**	**86.80**	**93.29**
DenseNet121	$\mathcal{L}_{Base}$	9.79	28.28	44.49	58.38	71.76	10.66	32.77	53.36	67.52	73.72
	$\mathcal{L}_{PCC}$	12.45	42.83	68.10	85.63	92.41	10.98	33.92	66.01	83.60	93.32
	+PUAP	**17.27**	**48.53**	**71.13**	**88.78**	**94.71**	**17.02**	**41.70**	**68.94**	**86.19**	**93.74**
Vgg16	$\mathcal{L}_{Base}$	25.91	57.70	74.84	86.13	92.96	19.45	51.61	75.68	85.90	91.75
	$\mathcal{L}_{PCC}$	28.22	58.15	83.71	95.60	98.24	21.43	50.63	79.87	94.59	97.79
	+PUAP	**28.75**	**60.19**	**87.79**	**96.75**	**98.72**	**29.43**	**61.73**	**83.80**	**95.38**	**98.04**
SENet154	$\mathcal{L}_{Base}$	**13.45**	32.39	48.25	63.40	72.10	11.62	26.20	43.48	57.87	67.79
	$\mathcal{L}_{PCC}$	12.00	32.43	61.02	76.90	88.41	11.35	29.35	55.66	75.38	88.21
	+PUAP	13.02	**34.04**	**61.43**	**79.99**	**90.40**	**11.76**	**30.40**	**58.14**	**80.08**	**89.16**
ShuffleNet	$\mathcal{L}_{Base}$	38.75	60.93	81.34	88.33	92.84	35.98	54.05	70.51	84.84	91.82
	$\mathcal{L}_{PCC}$	50.21	88.97	97.30	99.15	99.46	42.55	74.62	91.04	96.32	98.28
	+PUAP	**56.57**	**91.23**	**98.03**	**99.23**	**99.55**	**45.70**	**79.11**	**92.59**	**96.57**	**98.36**

Table: Experimental results (%) on Market-1501 and DukeMTMC-reID dataset. As can be seen, the proposed PUAP consistently improve the attack performance

Take Home Message

Polynomial Universal Adversarial Perturbations (PUAP) for Person Re-ID

- The extension from the constant formulation of existing UAP methods to a more general polynomial case.
- The introduction of Pearson correlation coefficient loss to disrupt the entire similarity rank of Re-ID.

西安交通大學
XI'AN JIAOTONG UNIVERSITY

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

Thanks!

Email: dingding@stu.xjtu.edu.cn; hongxiaopeng@mail.xjtu.edu.cn.