CRIPAC
智能感知与计算研究中心
Center for Research on Intelligent Perception and Computing

模式识别国家重点实验室
National Lab of Pattern Recognition

中国科学院自动化研究所
Institute of Automation Chinese Academy of Sciences

Dual Variational Generation for Low Shot Heterogeneous Face Recognition

Chaoyou Fu, Xiang Wu
Yibo Hu, Huaibo Huang and Ran He

NLPR & CRIPAC, CASIA
University of Chinese Academy of Sciences
Center for Excellence in Brain Science and Intelligence Technology, CAS

Heterogeneous Face Recognition

- Diverse modalities

NIR Thermal Sketch ID Card Video Profile

- Broad applications

Mobile Phone

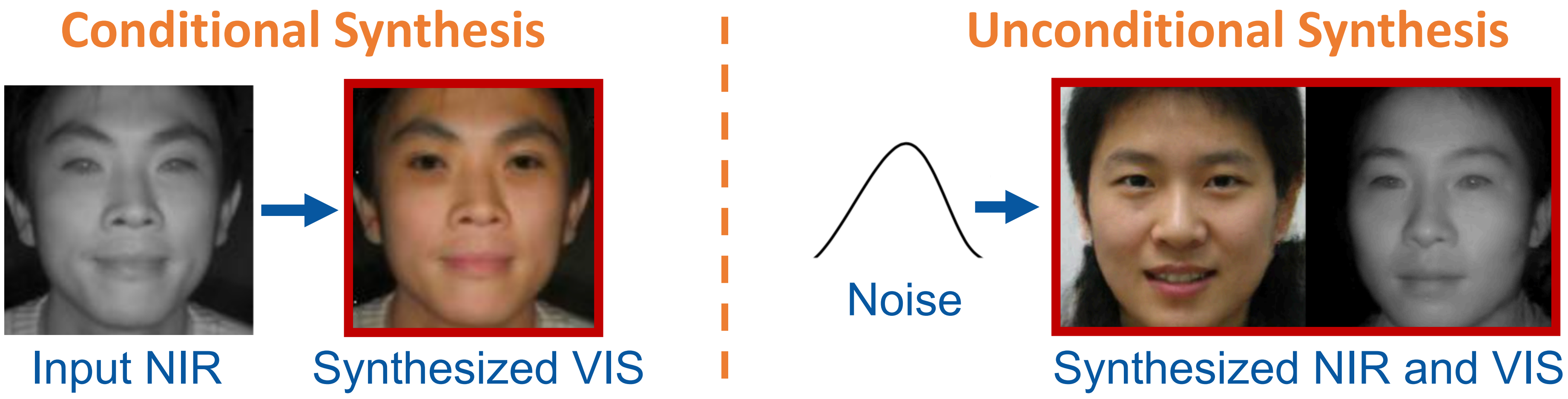
Criminology

Surveillance

Gate

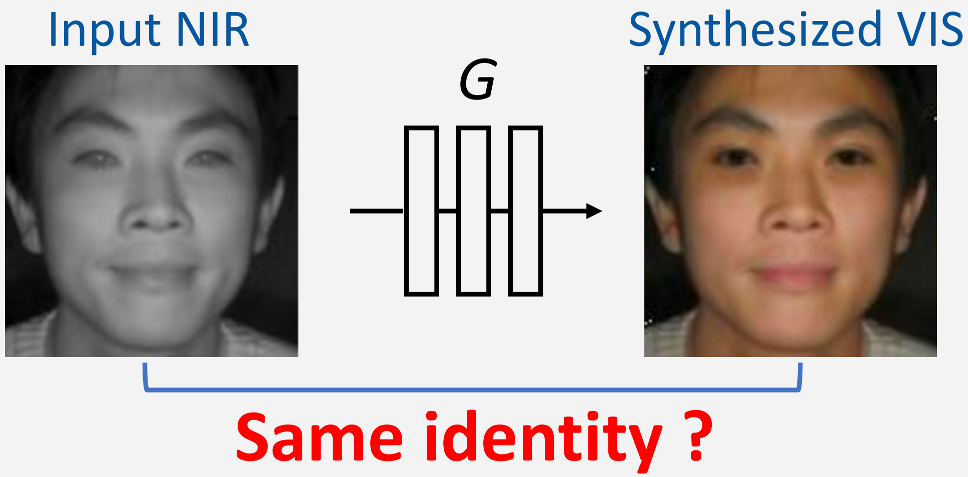
Heterogeneous Face Recognition

- Challenges in HFR
  - Large domain gap between heterogeneous data
  - The lack of large-scale databases
- Generative model for HFR
  - Conditional image synthesis - translate NIR to VIS to reduce domain gap
  - **Unconditional image synthesis** - generate images from noise



Conditional Image Synthesis

- Two challenges of such image-to-image translation methods
  - **Diversity:**
    Limited number of images and intra-class diversity
  - **Consistency:**
    Difficulty in preserving identity



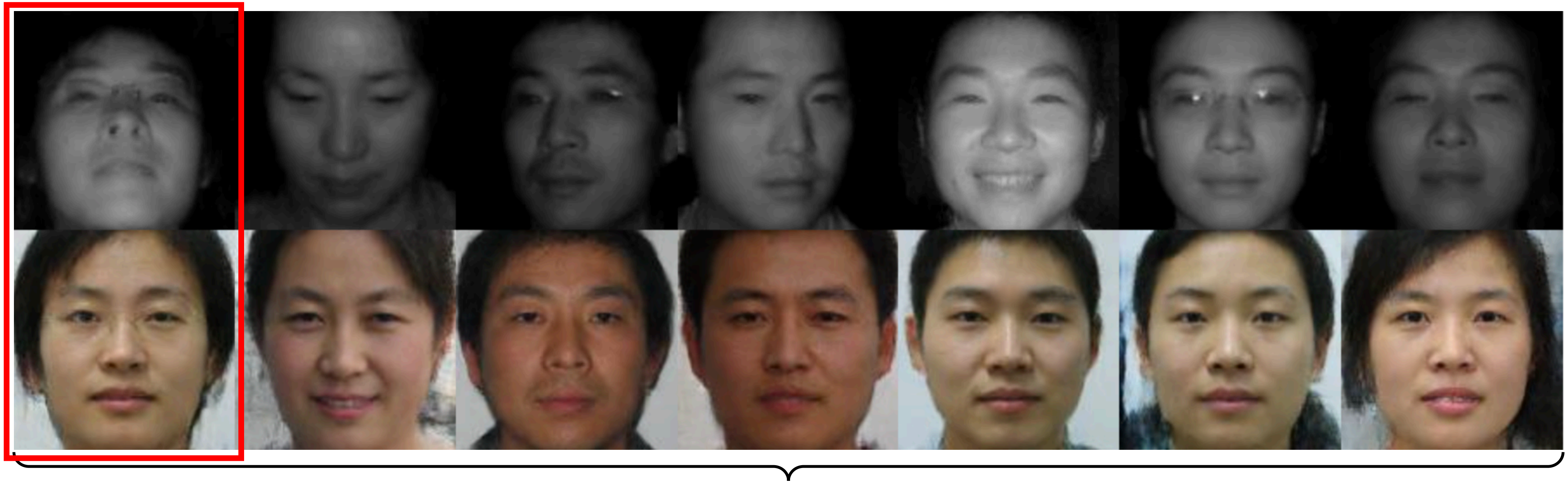
Only synthesize **one** new image of the target domain with **same attributes**

It is challenging to guarantee the **identity consistency**

Dual Variational Generation

- Generate **paired** new heterogeneous data from **noise**
  - ➢ Sample large-scale new images with abundant intra-class diversity
  - ➢ Ensure the identity consistency of the generated paired images

Same identity **Abundant intra-class diversity**



Large-scale new images

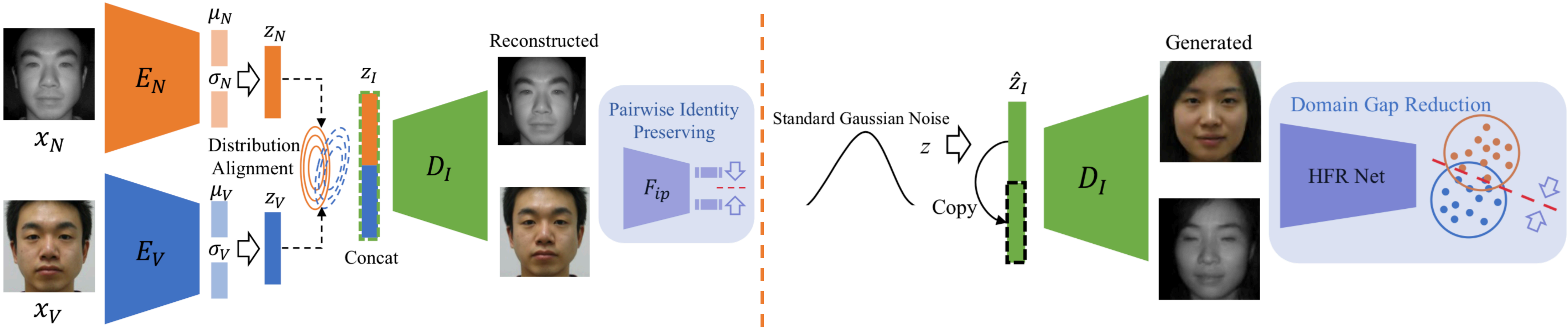
Dual Variational Generation

- Training method

➢ Learn the joint distribution of paired data

➢ Align the distributions via Wasserstein distance

➢ Preserve pairwise identity via F_{ip}



Training Framework

Testing Framework

Experiments

NIR-VIS

- CASIA NIR-VIS 2.0 database

Baseline: VR@FAR=0.1% = 97.4%

DVG: VR@FAR=0.1% = 99.8%

Improving 2.4%

- BUAA-VisNir database

Baseline: VR@FAR=0.1% = 89.4%

DVG: VR@FAR=0.1% = 97.3%

Improving 7.9%

- Oulu-CASIA NIR-VIS database

Baseline: VR@FAR=0.1% = 68.3%

DVG: VR@FAR=0.1% = 92.9%

Improving 24.6%

Experiments

Thermal-VIS

➢ Tufts Face database

Baseline: Rank-1 = 37.5%

DVG: Rank-1 = 53%

Improving 15.5%

Sketch-Photo

➢ IIIT-D Viewed Sketch database

Baseline: VR@FAR=1% = 81.04%

DVG: VR@FAR=1% = 97.86%

Improving 16.82%

Profile-Frontal Face

➢ Multi-PIE database

Baseline: Rank-1 = 65.4%

DVG: Rank-1 = 83.9%

Improving 18.5%

Poster: 05:30 -- 07:30 PM @ East Exhibition Hall B + C #66

Code is released: https://github.com/BradyFU/DVG

Dual Variational Generation for Low Shot Heterogeneous Face Recognition

Chaoyou Fu[1,2,3], Xiang Wu[1,2], Yibo Hu[1,2], Huaibo Huang[1,2], Ran He[1,2,3,4]

[1] Center for Research on Intelligent Perception and Computing, CASIA [2] National Laboratory of Pattern Recognition, CASIA
[3] University of Chinese Academy of Sciences [4] Center for Excellence in Brain Science and Intelligence Technology, CAS

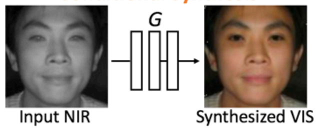
Scan to get code

CRIPAC 智能感知与计算研究中心 Center for Research on Intelligent Perception and Computing

Background

- Heterogeneous Face Recognition is a challenging issue because of the large domain discrepancy and a lack of heterogeneous data
- Previous image-to-image translation based methods face two challenges
  - **Diversity**
    Given one image, a generator only synthesizes one new image of the target domain, resulting in **limited number of images**. Moreover, two images before and after translation have same attributes except for the spectral information, leading to **limited intra-class diversity**
  - **Consistency**
    When generating large-scale samples, it is challenging to guarantee that the synthesized face images belong to the **same identity** of the input images

Conditional Synthesis



Input NIR Synthesized VIS

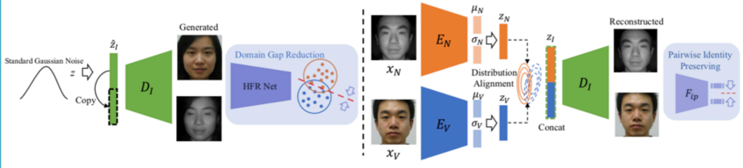
Dual Variational Generation

- Generate **paired** new heterogeneous data from **noise**
  - Sample large-scale new images with abundant intra-class diversity
  - Ensure the identity consistency of the generated paired images

Same identity Abundant intra-class diversity

Large-scale new images

Framework



The purpose (left part) and training model (right part). It generates large-scale new paired heterogeneous images with the same identity from standard Gaussian noise, aiming at reducing the domain discrepancy for HFR. A **distribution alignment** in the latent space and a **pairwise identity preserving** in the image space are imposed to guarantee the identity consistency of the generated paired images

Visual Results

CASIS NIR-VIS 2.0 Oulu BUAA

Quantitative Results

Method	CASIA NIR-VIS 2.0		Oulu-CASIA NIR-VIS			BUAA-VisNir		
	Rank-1	FAR=0.1%	Rank-1	FAR=1%	FAR=0.1%	Rank-1	FAR=1%	FAR=0.1%
IDNet [29]	87.1 ± 0.9	74.5	-	-	-	-	-	-
HFR-CNN [30]	85.9 ± 0.9	78.0	-	-	-	-	-	-
Hallucination [23]	89.6 ± 0.9	-	-	-	-	-	-	-
DLFace [28]	98.68	-	-	-	-	-	-	-
TRIVET [26]	95.7 ± 0.5	91.0 ± 1.3	92.2	67.9	33.6	93.9	93.0	80.9
IDR [10]	97.3 ± 0.4	95.7 ± 0.7	94.3	73.4	46.2	94.3	93.4	84.7
W-CNN [11]	98.7 ± 0.3	98.4 ± 0.4	98.0	81.5	54.6	97.4	96.0	91.9
DVR [35]	99.7 ± 0.1	99.6 ± 0.3	100.0	97.2	84.9	99.2	**98.5**	96.9
RCN [4]	99.3 ± 0.2	98.7 ± 0.2	-	-	-	-	-	-
MC-CNN [3]	99.4 ± 0.1	99.3 ± 0.1	-	-	-	-	-	-
LightCNN-9	97.1 ± 0.7	93.7 ± 0.8	93.8	80.4	43.8	94.8	94.3	83.5
LightCNN-9 + DVG	99.2 ± 0.3	98.8 ± 0.3	100.0	97.6	89.5	98.0	97.1	93.1
LightCNN-29	98.1 ± 0.4	97.4 ± 0.5	99.0	93.1	68.3	96.8	97.0	89.4
LightCNN-29 + DVG	**99.8 ± 0.1**	**99.8 ± 0.1**	**100.0**	**98.5**	**92.9**	**99.3**	**98.5**	**97.3**

Objective

- Lean the joint distribution

$$\mathcal{L}_{\text{rec}} = -\mathbb{E}_{q_{\phi_N}(z_N|x_N)\cup q_{\phi_V}(z_V|x_V)} \log p_\theta(x_N, x_V|z_I)$$

$$\mathcal{L}_{\text{kl}} = D_{\text{KL}}(q_{\phi_N}(z_N|x_N)||p(z_N)) + D_{\text{KL}}(q_{\phi_V}(z_V|x_V)||p(z_V))$$

- Align the distributions

$$\mathcal{L}_{\text{dist}} = \frac{1}{2}\left[||u_N^{(i)} - u_V^{(i)}||_2^2 + ||\sigma_N^{(i)} - \sigma_V^{(i)}||_2^2\right]$$

- Pairwise Identity Preserving

$$\mathcal{L}_{\text{ip-pair}} = ||F_{ip}(\hat{x}_N) - F_{ip}(\hat{x}_V)||_2^2$$

$$\mathcal{L}_{\text{ip-rec}} = ||F_{ip}(\hat{x}_N) - F_{ip}(x_N)||_2^2 + ||F_{ip}(\hat{x}_V) - F_{ip}(x_V)||_2^2$$

More Experiments

Thermal-VIS Sketch-Photo Profile-Frontal Face

- Tufts Face database
  Baseline: Rank-1 = 37.5%
  DVG: Rank-1 = 53%
  Improving 15.5%
- IIIT-D Viewed Sketch database
  Baseline: VR@FAR=1% = 81.04%
  DVG: VR@FAR=1% = 97.86%
  Improving 16.82%
- Multi-PIE database
  Baseline: Rank-1 = 65.4%
  DVG: Rank-1 = 83.9%
  Improving 18.5%

Contributions

- We provide a new insight into the problems of HFR. That is, we consider HFR as a dual generation problem, and propose a novel dual variational generation framework. This framework generates new paired heterogeneous images with abundant intra-class diversity
- We can sample large-scale diverse paired heterogeneous images from noise. By constraining the pairwise feature distances of the generated paired images in the HFR network, the domain discrepancy is effectively reduced