Generating Thermal Human Faces for Physiological Assessment Using Thermal Sensor Auxiliary Labels

Catherine Ordun[*], Edward Raff[*^], Sanjay Purushotham

[*]UMBC Dept. Information Systems, [*^]UMBC Dept. Computer Science, [*^]Booz Allen Hamilton

UMBC

Introduction and Motivation

- Thermal imagery, specifically long-wave IR (LWIR) has been studied for decades rooted in the intersection of physiological research and affective computing.
- Used for Facial Emotion Recognition (FER), and Facial Recognition (FR) person re-identification on thermal imagery.
- Medical applications for telemedicine.

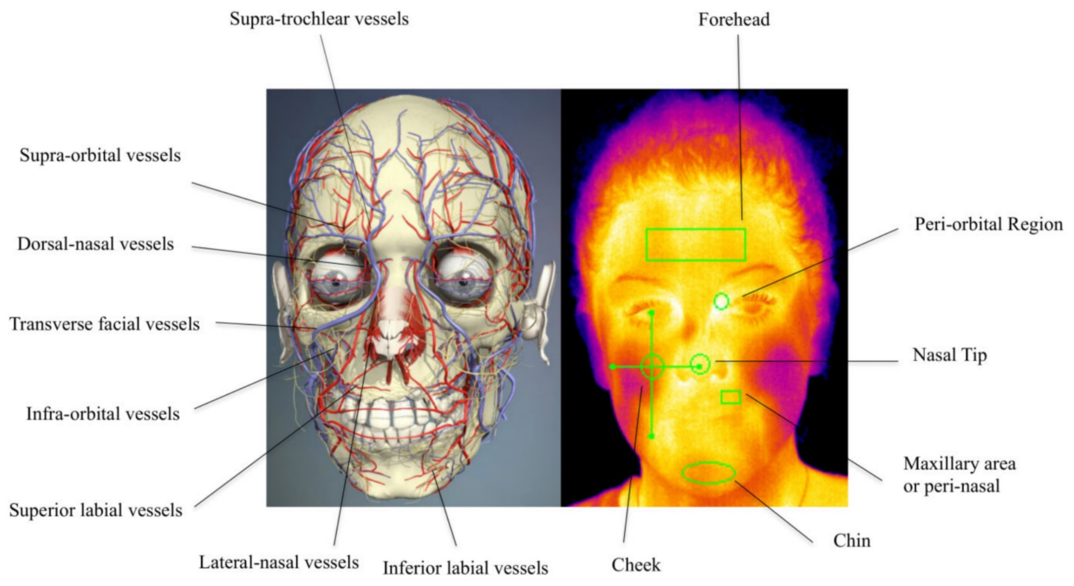


Figure 1. Thermal representation for extraction of ROIs along with a vascular representation of the major vessels affecting the subcutaneous temperature of the face (Berkovitz, Kirsch, Moxham, Alusi, & Cheesman, 2013).

Is invariant to lighting conditions unlike RGB, allowing the detection of physiological response (heat) to occur in low light or total darkness.

Facial temperature offers a reliable and accurate correlation to standard physiological measures like respiration and heart rate.

Is non-invasive making it convenient and non-intrusive and potentially relevant for non communicative persons.

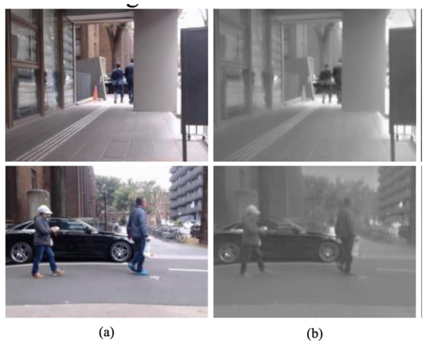
Resistant to intentional deceit since physiological responses cannot be faked, whereas visible facial expressions can be controlled.

Reveals facial disguises (i.e. wigs, masks) since these materials have high reflectivity.

Offers physiological signals of social interactions from person to person.

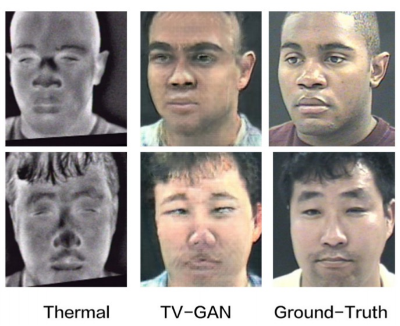
Related Work

- Our work focuses on the task of Image-to-Image translation using conditional Generative Adversarial Networks (GAN), translating one image mode to another mode.
- Successful $thermal \ \rightarrow visible\ (TV)$ works exist for law enforcement use cases.
- Downstream tasks for person re-identification, thermal face recognition, or object detection.
- Such models use only one dataset (see Ordun et al. for complete literature review of thermal FER datasets)

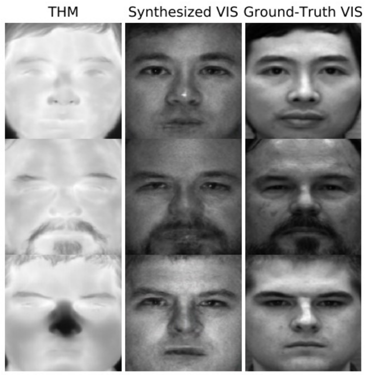


Object detection (Li, et al.)

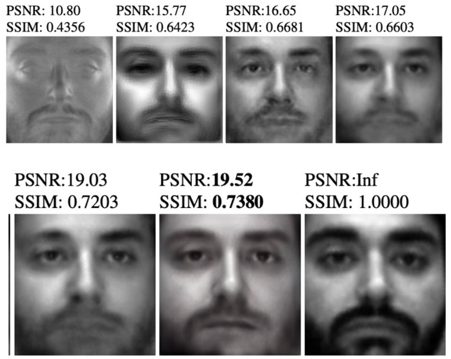
ThermalGAN (Kniaz, et al.)



TV-GAN (Zhang, et al.)



Semantic-Guided GAN (Chen, et al.)



Polarimetric Faces (Zhang, et al.)

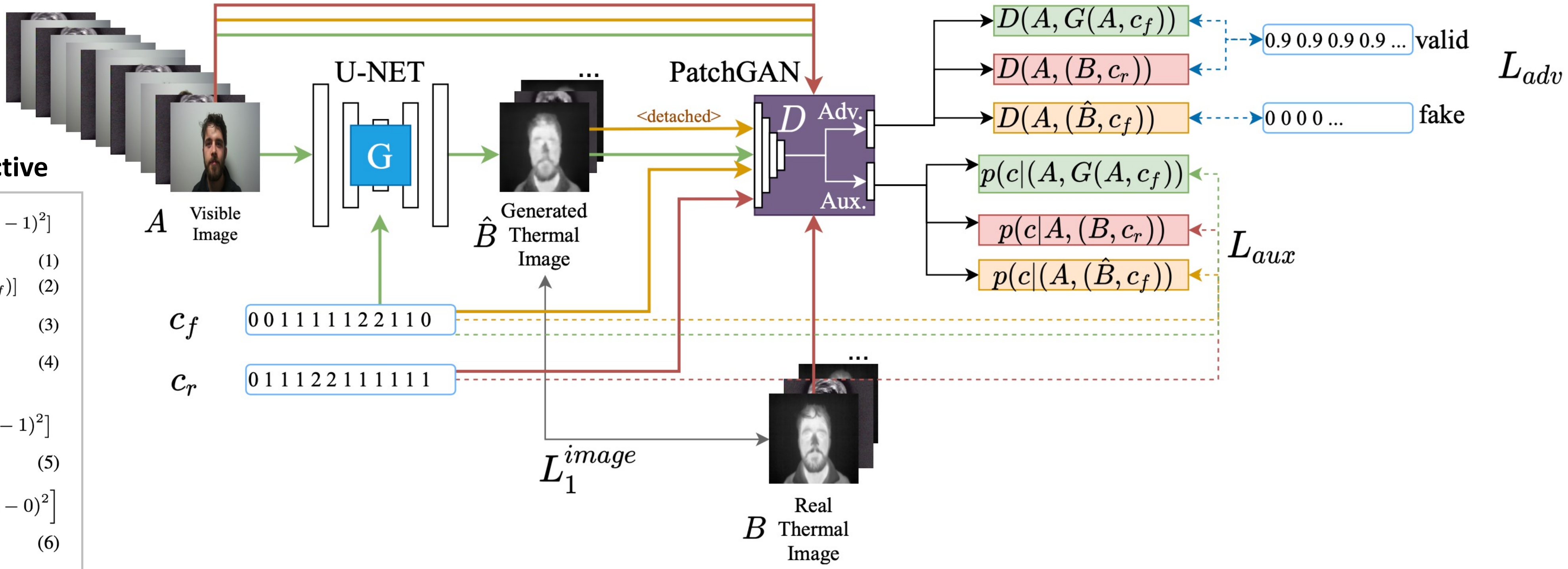
Contributions

- The first work to study VT translation of human faces, by developing a pix2pix-based favtGAN model.
- We study the image quality of generated thermal face images which is important for medical applications.
- We bootstrap training of image translation with additional data from different domains but similar thermal sensors to improve thermal image generation.

favtGAN

Notation:
A – Real visible image
B – Real thermal image
G – Generator (U-NET 256)
D - PatchGAN (16 x 16 patches)
$\hat{B}$ - Generated thermal image

c_f - Noisy sensor labels $\sim U[0,1]$
c_r - Real sensor class labels
L_{adv} - Adversarial loss (MSE)
L_{aux} - Auxiliary loss (CE)
L_1^{image} - Reconstruction loss (L1)



Equations for Generator and Discriminator and Training Objective

$$L_{Adv}(G) = \frac{1}{2} \mathop{\mathbb{E}}_{A\sim p_{vis}, c_f \sim U\{0,1\}, \hat{B}\sim p_G} [(D(A, \hat{B}, c_f) - 1)^2] \quad (1)$$

$$L_{aux}(G) = \mathop{\mathbb{E}}_{A\sim p_{vis}, c_f \sim U\{0,1\}, \hat{B}\sim p_G} [\log C(A, \hat{B}, c_f)] \quad (2)$$

$$L_1^{image}(G) = \mathop{\mathbb{E}}_{B\sim p_{thr}, \hat{B}\sim p_G} \|B - \hat{B}\|_1 \quad (3)$$

$$L_G = L_{Adv}(G) + L_{aux}(G) + \lambda L_1^{image}(G) \quad (4)$$

$$L_{Adv_{D_{real}}} = \frac{1}{2} \mathop{\mathbb{E}}_{A\sim p_{vis}, B\sim p_{thr}, c_r \sim p_{thr}^l} [(D(A, B, c_r) - 1)^2] \quad (5)$$

$$L_{Adv_{D_{fake}}} = \frac{1}{2} \mathop{\mathbb{E}}_{A\sim p_{vis}, c_f \sim U\{0,1\}, \hat{B}\sim p_G} \left[(D(A, \hat{B}, c_f) - 0)^2\right] \quad (6)$$

$$L_{aux_{D_{real}}} = \mathop{\mathbb{E}}_{A\sim p_{vis}, B\sim p_{thr}, c_r \sim p_{thr}^l} [\log C(A, B, c_r)] \quad (7)$$

$$L_{aux_{D_{fake}}} = \mathop{\mathbb{E}}_{A\sim p_{vis}, c_f \sim U\{0,1\}, \hat{B}\sim p_G} \left[\log C(A, \hat{B}, c_f)\right] \quad (8)$$

$$L_D = \frac{1}{2}[(L_{Adv_{D_{real}}} + L_{aux_{D_{real}}}) + (L_{Adv_{D_{fake}}} + L_{aux_{D_{fake}}})] \quad (9)$$

$$G^* = \min_G L_G + \min_D L_D \quad (10)$$

Experimental Methods – 4 paired datasets

Eurecom and ADAS – both microbolometers	Iris and OSU – both BST sensors



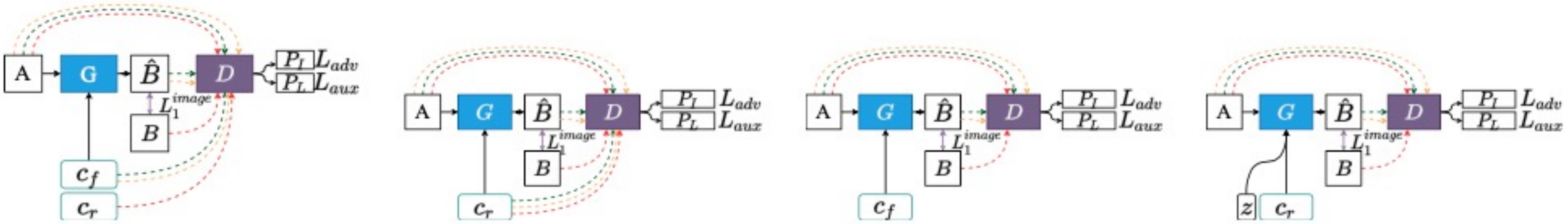

Eurecom Dataset	FLIR ADAS Dataset	Iris Dataset	Oklahoma State University (OSU) Dataset
Domain: Face Thermal Sensor: VOx Microbolometer	Domain: Cityscapes Thermal Sensor: VOx Microbolometer	Domain: Face Thermal Sensor: BST Ferroelectric	Domain: Cityscapes Thermal Sensor: BST Ferroelectric

			
FLIR Duo Pro – VOx Microbolometer <50 mK sensitivity LWIR spectral range: 7.5 - 13.5 μm Thermal res: 336 x 256 or 640 x 512	**FLIR Tau2 – VOx Microbolometer** <30 mK sensitivity LWIR spectral range: 7.5 - 13.5 μm Thermal res: 336 x 256 or 640 x 512	**Raytheon Palm-IR Pro – BST Ferroelectric** <100 mK sensitivity LWIR spectral range: 7.0 – 14.0 μm Thermal res: 320 x 240	**Raytheon 300D Thermal – BST Ferroelectric** (Approximate info) <100 mK sensitivity LWIR spectral range: 7.0 – 14.0 μm Thermal res: 320 x 240

Data	Num Sensors	Train Pairs	Train Subjects	Test Pairs	Test Subjects	Total Subjects	Total Pairs	Eurecom Test IDs	Iris Test IDs
Eurecom	**1**	**945**	**45**	**105**	**5**	**50**	**1050**	**1, 2, 21, 31, 36**	**n/a**
Iris	1	846	26	98	3	29	944	n/a	['Vijay', 'Meng', 'Vicky']
Adas	1	842	n/a	98	n/a	n/a	940	n/a	n/a
OSU	1	843	n/a	211	n/a	n/a	1054	n/a	n/a
EA	2	1787	45	203	5	50	1990	1, 2, 21, 31, 36	n/a
EI	2	1791	71	203	8	79	1994	1, 2, 21, 31, 36	['Vijay', 'Meng', 'Vicky']
IO	2	1689	26	309	3	29	1998	n/a	['Vijay', 'Meng', 'Vicky']

Experimental methods – favtGAN variations

- Train each variant on:
- Only Eurecom (face domain)
- Only Iris (face domain)
- Eurecom + Iris (only face domain)
- Eurecom + ADAS (similar VOx Microbolometer sensor, but different domains)
- Iris + OSU (similar BST Ferroelectric sensor, but different domains)
- Compare against unmodified pix2pix architecture



(a) favtGAN (b) No Noise (c) Noisy Labels (d) Gaussian Noise

Quantitative Results

Table 1: Image Quality Metrics using Mean SSIM and Mean PSNR for Generated Thermal Images, Translated from the Visible Test Set. SSIM % and PSNR % show the relative change compared to pix2pix trained only a single face dataset. EI: Eurecom + Iris dataset, EA: ADAS + Eurecom dataset, IO: OSU + Iris dataset, FG: favtGAN

			Eurecom		
Dataset	**Experiment**	**SSIM**	**PSNR**	**SSIM %**	**PSNR %**
Eurecom	pix2pix	0.906	32.048	-	-
EI	pix2pix	0.924	32.133	1.98%	0.26%
EI	FG-Baseline	0.925	32.366	2.09%	0.98%
EI	FG-No Noise	0.914	29.230	0.85%	-9.64%
EI	FG-Noisy Labels	0.925	31.835	2.02%	-0.67%
EI	FG-Gauss. Noise	0.909	28.242	0.36%	-13.48%
EA	FG-Baseline	0.931	33.104	2.69%	3.19%
EA	**FG-Noisy Labels**	**0.931**	**33.139**	**2.69%**	**3.29%**
			Iris		
Dataset	**Experiment**	**SSIM**	**PSNR**	**SSIM %**	**PSNR %**
Iris	pix2pix	0.685	24.158	-	-
EI	pix2pix	0.681	23.946	-0.54%	-0.89%
EI	FG-Baseline	0.682	24.060	-0.37%	-0.41%
EI	FG-No Noise	0.653	22.000	-4.91%	-9.81%
EI	FG-Noisy Labels	0.682	23.990	-0.42%	-0.70%
EI	FG-Gauss. Noise	0.652	22.083	-5.07%	-9.40%
IO	FG-Baseline	0.686	24.474	0.15%	1.29%
IO	**FG-Noisy Labels**	**0.690**	**24.797**	**0.72%**	**2.58%**

Qualitative Results

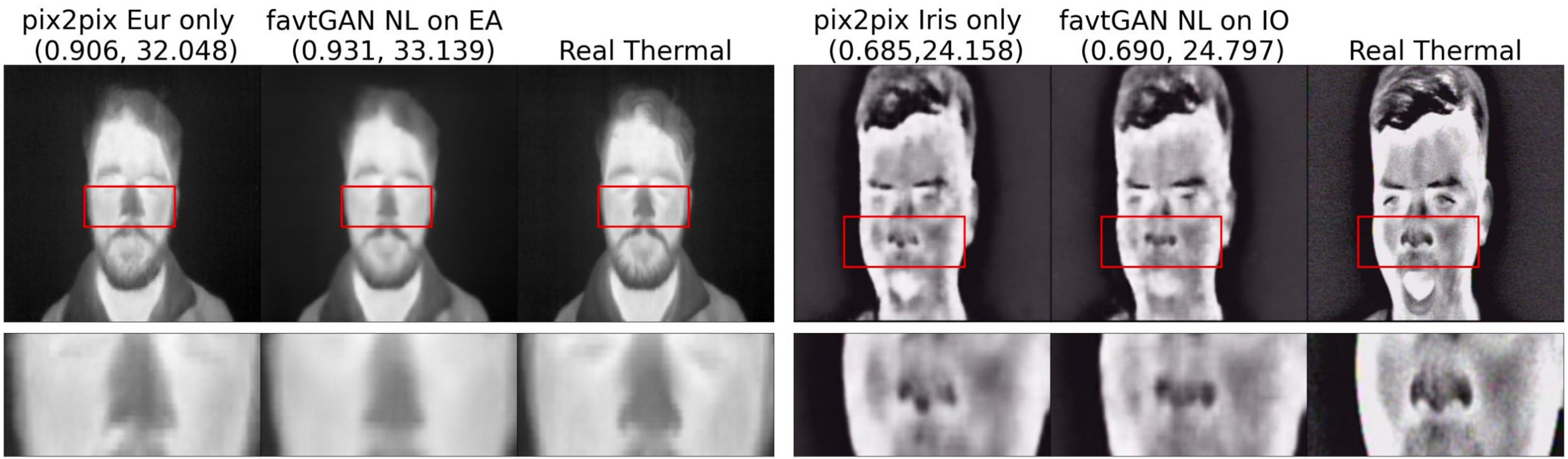


Fig. 2: **Generated Thermal Images Translated from the Visible Test Set.** Samples are shown from the best performing favtGAN experiment trained on combined face and cityscape datasets, compared to training pix2pix on a single dataset. Average SSIM and PSNR scores are provided. Red boxes show regions of interest, which are magnified in the second row. “NL”: Fig.1c Noisy Labels variation, “EA”: ADAS + Eurecom dataset, “IO”: OSU + Iris dataset.

Thanks

cordun1@umbc.edu
raff_edward@bah.com
psanjay@umbc.edu