# TR-GAN: thermal to RGB face synthesis with generative adversarial network for cross-modal face recognition

Kezebou, Landry, Oludare, Victor, Panetta, Karen, Agaian, Sos

Landry Kezebou, Victor Oludare, Karen Panetta, Sos Agaian, "TR-GAN: thermal to RGB face synthesis with generative adversarial network for cross-modal face recognition," Proc. SPIE 11399, Mobile Multimedia/Image Processing, Security, and Applications 2020, 113990P (21 April 2020); doi: 10.1117/12.2558166

Event: SPIE Defense + Commercial Sensing, 2020, Online Only

# TR-GAN: Thermal to RGB face synthesis with Generative Adversarial Network for Cross-Modal Face Recognition

Landry Kezebou[1], Victor Oludare[1], Karen Panetta[1], Sos Agaian[2]
[1]Tufts University, Medford, MA, 02155
[2]City University of New York, New York, NY 100161

## ABSTRACT

Unlike RBG cameras, thermal cameras perform well under very low lighting conditions and can capture information beyond the human visible spectrum. This provides many advantages for security and surveillance applications. However, performing face recognition tasks in the thermal domain is very challenging given the limited visual information embedded in thermal images and the inherent similarities among facial heat maps. Attempting to perform recognition across modalities, such as recognizing a face captured in the thermal domain given the corresponding visible light domain ground truth database or vice versa is also a challenge. In this paper, a Thermal to RGB Generative Adversarial Network (TR-GAN) to automatically synthesize face images captured in the thermal domain, to their RBG counterparts, with a goal of reducing current inter-domain gaps and significantly improving cross-modal facial recognition capabilities is proposed. Experimental results on the TUFTS Face Dataset using the VGG-Face recognition model without retraining, demonstrates that performing image translation with the proposed TR-GAN model almost doubles the cross-modal recognition accuracy and also performs better than other state-of-the-art GAN models on the same task. The generator in our network uses a U-NET like architecture with cascaded-in-cascaded blocks to reuse features from earlier convolutions, which helps generate high quality images. To further guide the generator to synthesize images with fine details, we optimize a training loss as the weighted sum of the perceptual, adversarial, and cycle-consistent loss. Simulation results demonstrate that the proposed model generates more realistic and more visually appealing images, with finer details and better reconstruction of intricate details such sunglasses and facial emotions, than similar GAN models.

***Keywords***—*Thermal spectrum, face recognition, Generative Adversarial Network, Thermal-RBG*

## 1. INTRODUCTION

Face recognition is a relatively solved problem in the visible spectrum [1]–[7] but remains an open challenge in non-ideal situations such as low illumination, total darkness, and thick smoke, haze or fog. In such non-ideal environmental conditions, information is often better captured in another domain such as the thermal or near-infrared domain. In thermal-infrared images faces are distinguished from their backgrounds based on the radiation difference, which makes them more suitable for low lighting conditions. However, directly using state-of-the-art visible face recognition methods on the thermal imagery data does not produce a satisfactory performance. This problem is relatively due to several factors including: (1) drastically smaller amount of thermal imagery data collected compared to the RGB data; (2) thermal images typically have low resolution, poor texture, and higher image noise; (3) face recognition in the thermal infrared domain has not received as much attention, in contrast with recognition in the visible-spectrum imagery; and (4) the inherent similarities among facial heat maps and skin color. Therefore, it is desirable to address these challenges by reducing these domain gaps and facilitate cross-modal recognition. The goal of this research is to effectively synthesize thermal face images to their visible RGB domain counterparts to help boost cross-modal recognition accuracy between thermal and visible spectrum domain.

Recent advancements in Deep Convolutional Neural Networks have help enable sophisticated facial detection and recognition systems, which prove valuable in surveillance and security systems applications. Existing state-of-the-art facial recognition systems have demonstrated high-performance accuracy for automatic face detection and identification/recognition tasks [1]–[7]. These models have become so reliable and efficient that they can run on small devices such as mobile phones, for self-identification. Availability of large-scale face datasets such CelebA [8] and IMDB-WIKI [9], helped foster the tremendous success of facial recognition systems such as VGG-Face [7] and other modern

Mobile Multimedia/Image Processing, Security, and Applications 2020, edited by Sos S. Agaian, Vijayan K. Asari, Stephen P. DelMarco, Proc. of SPIE Vol. 11399, 113990P · © 2020 SPIE
CCC code: 0277-786X/20/$21 · doi: 10.1117/12.2558166

recognition systems that we know today. However, such datasets extensively focused on face data in the visible spectrum, hence models trained on such data significantly degrade in performance when tested on images in other modalities such as thermal domain. Unlike RBG cameras, thermal cameras can capture information in the dark by reading heat maps and as such facial features are encoded differently making recognition in such domain very hard.

While many thermal face datasets are proposed [10]–[12], it is very difficult to train a face detection and recognition system from scratch to achieve comparable performance as existing systems on visible spectrum data. This is because it entails collecting and manually annotating millions of images and several days/ weeks of training, not to mention adequate time needed to fine-tune network architecture and corresponding hyperparameters.

Generative Adversarial Networks-(GAN) on the other hand have gained attention in the computer vision research community since its introduction in 2014 by Ian Goodfellow et al [13], and is considered the most innovative idea in Deep Learning in recent decades [14]. GANs have greatly evolved over the years and have found applications in numerous domains spanning, image-to-image translation, super-resolutions, voice synthesis, video synthesis, classifications tasks, security applications, image generation, photo inpainting, video predictions text-to-image translation face aging, and much more [15].

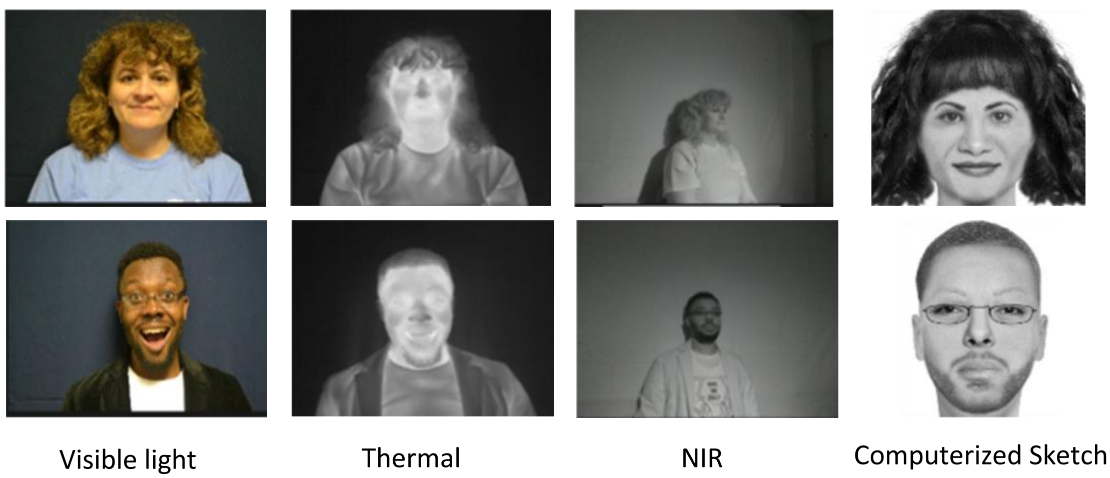



*Figure 1: Samples images from 4 domains taken from the Tufts Face Database* [11]

This work proposes a novel GAN model for automatically synthesizing face images in the thermal domain into their corresponding photorealistic images to address the above mentioned challenges. This unique GAN model will help achieve comparable performance result of existing face recognition systems without having to retrain the model on thermal images, hence addressing the current problem of efficiently recognizing faces captured in the thermal domain.

The contributions of this work are as follows:

- **a)** Developing an improved GAN model suitable for translating face images from the thermal domain to their RGB equivalent.
- **b)** Improving the accuracy of face recognition in the thermal domain by using existing pretrained models without having retraining from scratch on thermal image data.
- **c)** Providing a normalized paired dataset with proper alignment and uniform sizing to facilitate training of similar intermodal image-to-image translation models. The proposed improved dataset is derived from the Tufts Face Dataset [11] and will be made available as contribution on the author's Kaggle page.

The rest of the paper is organized as follows: Section 2 discusses related work and the shortcomings of existing approaches; Section 3 presents the proposed TR-GAN network architecture and associated loss functions; Section 4 presents the experimental setups and discusses the results; and finally conclusion and future work are presented in Section 5.

## 2. RELATED WORKS

Cross-modality face recognition aims to identify faces through distinct modalities such as thermal, near-infrared, and computerized sketches to mention a few. Cross-domain matching is still a challenging problem because of: (1) the modality gap caused by texture, resolution, and illumination, (2) the lack of coupled cross-modality databases, and lack of efficient learning tools; (2) difficulties in collecting faces of the same person with various cross-modality and attributes, which means limited inter-domain face images are available for training compared to the visible images.

While there has been tremendous success in facial detection and recognition systems in the visible light domain [1]–[9] much less work has been done to achieve such performance in other domains such as the thermal domain. State-of-the-art facial recognition models perform poorly when attempting to recognize faces captured in the thermal domain. This is due to the existing domain gap between images captured in the visible and thermal domain. A few methods have been proposed to address this domain gap problem.

Commonly used cross-modal face recognition methods utilize hand-crafted feature descriptors, which require prior knowledge to engineer relevant features, and which cannot exploit data-adaptive characteristics in feature extraction, hence resulting in unsatisfactory performance. For example, J. Choi et al. proposed using partial least squares-discriminant analysis (PLS-DA) framework to formulate facial recognition as a multimodal problem [16]. However, the reported recognition accuracy remained well below 50%. Antonio J, et al. [17] proposed a framework for automatically detecting faces in the thermal domain using haar cascade in combination with face contours, template matching and chamfer matching feature extractors. Felix J. et al. [18] proposed a facial recognition framework based on cross-spectral joint dictionary learning in combination with image reconstruction technique to achieve good recognition accuracy on face images across the near-infrared and visible light modality. Although this work achieved good accuracy, it required training the model from scratch on images from both domains and focused on frontal faces whereas our proposed approach has a design goal to be invariant to facial expressions, head rotation and partial occlusions with sun-glasses.

H. Zhang et al. [19] proposed a generative adversarial network to synthesize polarimetric thermal images into visible light domain. However, the target visible domain is the gray which is much closer to the polymetric thermal images hence relaxing the optimization constraint of the loss function. The model is limited for reconstructing colored images from the thermal counterpart and the synthesized images can exhibit blurriness and lack of finer details. Xin D. et al [20] also recently proposed a more stable GAN network for synthesizing visible face from polarimetric thermal counterpart via attribute preserved synthesis. The Generator network is a U-NET like architecture [21] with a multimodal compact bilinear pooling, and effectively fusing the attribute vector and input visible image features in the latent space to help guide the generator. Furthermore, the combination of perceptual and identify losses in the generator optimization function helps eliminate blurriness and generate sharper image resulting in better recognition accuracy than similar state-of-the-art GAN models [22]–[24]. However, these works do not extend to the visible RGB domain and focus only on frontal faces with no emotions, and no occlusions.

The most closely related work that attempts to synthesize faces in the visible RGB domain from their counterpart in the thermal-infrared is the TV-GAN proposed by Teng et al [25] which uses the base pix2pix [26] model with incorporated identity loss in the generator. However, this model generates some non-realistic colored images along with blurry and noisy patches, in the output image. Furthermore, faces generated by this model have limited sharpness and often misrepresent color information.

This work proposes to develop a more robust and more dynamic GAN network to synthesize photorealistic RGB visible faces from thermal-infrared face images. The proposed model uses a modified U-NET like architecture in the generator network with a cascaded-in-cascaded dense blocks. In addition to the adversarial and cycle-consistent losses, we incorporate the perceptual loss [27] to guide the synthesis of more realistic images and overcome the challenges such as blurriness, lack of attention and sharpness, present in other architectures.

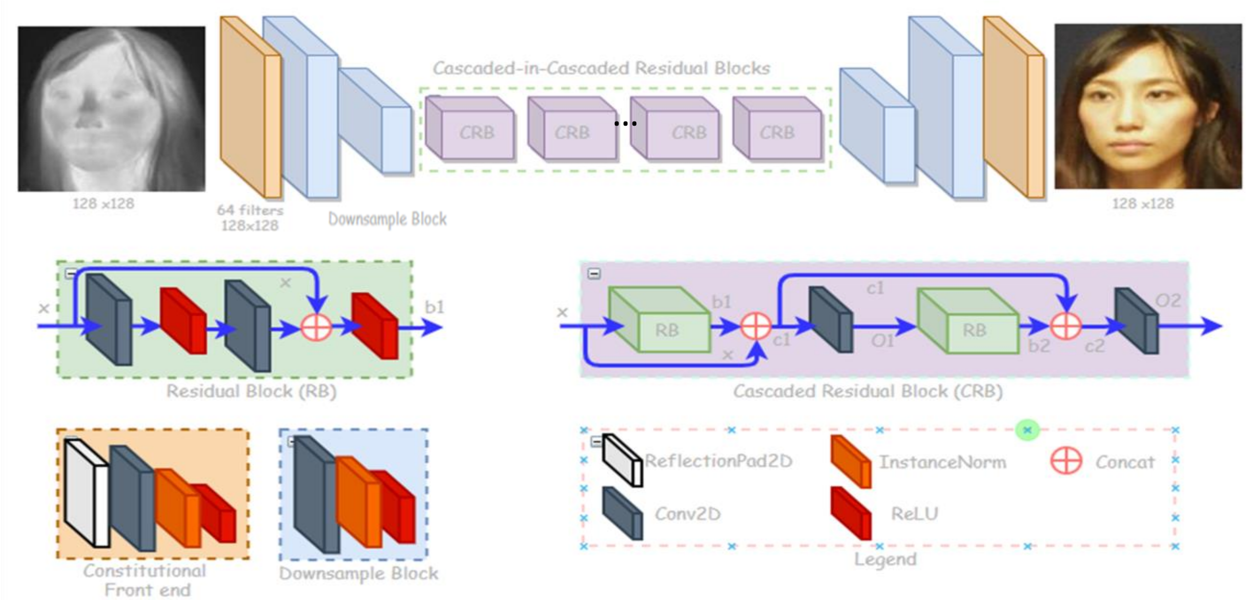



*Figure 2: TR-GAN generator network architecture.* ***Conv2D*** *denotes 2D convolution layer,* ***ReLU*** *denotes Rectified Linear Unit activation function,* ***Concat*** *indicates concatenation, InstanceNorm is for instance normalization. Blocks are color coded accordingly.*

## 3. PROPOSED METHOD

The goal of the proposed method is to learn a mapping function $G_{T\rightarrow R}$ that maps facial images in the thermal domain to their counterparts in the visible light RGB domain using paired samples from each domain. This is to boost cross-modal recognition accuracy of pretrained state-of-the-art face recognition models on thermal images by effectively eliminating the domain gap between thermal and RGB facial images. The proposed framework is inspired by CycleGAN [28], Pix2PixHD [29], and Residual Dense Nets [30] architectures and optimization functions.

### 3.1 Architecture

The TR-GAN framework employs a U-Net-like [21] architecture with cascade residual blocks in the generator network as shown in **Figure 2**. Unlike Pix2PixHD, TR-GAN uses only a global generator and eliminates the need for a local enhancer by incorporating cascaded residual blocks. The generator network consists of a convolutional front-end, two down sampling layers followed by a set of cascaded residual block, and finally two up-sampling layers and a transpose convolution layer. As a result, replacing the "*Resnet Blocks*" in the CycleGAN architecture with the *"Cascaded-in-Cascaded Residual Blocks"* shown in **Figure 2**, helps the generator to synthesize images with more consistent local and global structural information. This is particularly useful because the training dataset not only contains frontal face images as other datasets, but also captures four different emotions poses (neutral, surprised, sleepy, and smiling), side views (left and, right side view), and partial occlusion (with sun glasses) for each subject in the dataset, which renders training more complex. Furthermore, in addition to CycleGAN losses, the perceptual loss the generator objective function to help maintain realism and properly synthesize fine details is added. By virtue of cycle-consistency of the optimization function, the model simultaneously learns the inverse mapping function $F_{R\rightarrow T}$ during training, which can be used to generate paired thermal images for larger face datasets such as CelebA [8].

The TR-GAN model adapts the discriminator network architecture of CyleGAN [28] without any changes. This consists of two adversarial discriminators $D_T$ and $D_R$ ,where $D_T$ aims to discriminate between the original thermal images $\{t\}$ and the synthesized thermal image using the inverse mapping function $\{F(r)\}$; whereas $D_R$ aims to discriminate between the ground truth RGB images $\{r\}$ and reconstructed RGB images by the forward mapping function $\{G(t)\}$.

## 3.2 Objective Function

The objective function for the proposed GAN model aggregates the benefits of adversarial losses, cycle-consistent losses and perceptual loss respectively.

***Adversarial losses:*** The adversarial GAN loss function [13] enables modelling of the conditional statistical distribution of real RGB visible light images to help guide synthesis of realistic RGB images from given thermal input images. The adversarial loss consists of two parts, one for forward mapping function $G_{T\to R}$ and another for the inverse mapping function $F_{R\to T}$ as follows:

$$\boldsymbol{L_{GAN}(G,\ D_R\ ,T,R) = \min_{G}\max_{D_R}(\ E_{r\sim P_{data}(r)}\left[log\left(D_R(r)\right)\right]\ +\ E_{t\sim P_{data}(t)}\left[log\left(1-\ D_R(G(t))\right)\right])} \quad \textbf{(1)}$$

$$\boldsymbol{L_{GAN}(F,\ D_T\ ,R,T) = \min_{F}\max_{D_T}(\ E_{t\sim P_{data}(t)}\left[log\left(D_T(t)\right)\right]\ +\ E_{r\sim P_{data}(r)}\left[log\left(1-\ D_T(F(r))\right)\right])} \quad \textbf{(2)}$$

$$\boldsymbol{L_{ADV} = L_{GAN}(G,\ D_R\ ,T,R) +\ L_{GAN}(G,\ D_r\ ,R,T)} \quad \textbf{(3)}$$

Where $\boldsymbol{T}$ represents the thermal domain and $\boldsymbol{R}$ represents the visible light RGB domain; $\boldsymbol{t \in T}$ and $\boldsymbol{r \in R}$. $\boldsymbol{t\sim P_{data}(t)}$ and $\boldsymbol{r\sim P_{data}(r)}$ represents the data distribution in the thermal and visible RGB domains respectively.

***Cycle-Consistent losses****:* The cycle-consistent losses help maintain training consistency by effectively preventing the learned forward mapping function $\boldsymbol{G}$ from contradicting the learned inverse mapping function $\boldsymbol{F}$ [28] . This is done to ensure that for any given thermal input image $\boldsymbol{t_i}$ and corresponding RGB synthesized image $\boldsymbol{r'_i}$ using the forward mapping function $\boldsymbol{G}$, we can reconstruct the original thermal input image $\boldsymbol{t'_i}$ using the inverse mapping function $\boldsymbol{F}$, such that $\boldsymbol{t_i} \approx \boldsymbol{t'_i}$ , and vice versa.

$$\boldsymbol{t_i \to G(t_i) \to F(G(t_i)) \approx t_i} \text{ and } \boldsymbol{r_i \to F(r_i) \to G(F(r_i)) \approx r_i} \quad \textbf{(4)}$$

The cycle-consistent loss is expressed as follows:

$$\boldsymbol{L_{CYC}\ = E_{t\sim P_{data}(t)}[||F\left(G(t)\right) -\ t\ ||_1]\ +\ E_{r\sim P_{data}(r)}[||G\left(F(r)\right) -\ r\ ||_1]} \quad \textbf{(5)}$$

***Perceptual loss****:* Pix2PixHD [29] experiments found that adding perceptual loss to the objective function can significantly improve performance. We introduce the VGG loss [27] into our objective function to help the generator evaluate the perceptual feature difference between the original RGB image and the synthesized image. Essentially, we reuse the pretrained features of the VGG19 network for optimum performance.

$$\boldsymbol{L_{VGG} = \sum_{i=1}^{N}\frac{1}{M_i}\ [||F^{(i)}(r) -\ \ F^{(i)}(G(t))\ ||_1]} \quad \textbf{(6)}$$

Where $F^{(i)}\ and\ M_i$ respectively denote the $i^{th}$ layer of VGG19 network, and the $i^{th}$ element of that layer.

The complete objective function for the proposed TR-GAN architecture is defined as follows:

$$\boldsymbol{L_{TR-GAN} = arg\min_{G,F}\max_{D_T,D_R}(L_{ADV} + \lambda_1 L_{CYC} +\ \ \lambda_2 L_{VGG}\ )} \quad \textbf{(7)}$$

Where $\boldsymbol{\lambda_1 = 10.0}$ *and* $\boldsymbol{\lambda_2 = 2.0}$ respectively represent the weight for Cycle-consistent loss and VGG loss and control their relative importance in the overall optimization function.

## 4. EXPERIMENTS AND RESULTS

### 4.1 Dataset and Training

The proposed TR-GAN is trained on the Tufts Face Dataset [11] which contains ground truth images across seven different modalities including visible light, near-infrared, thermal, computerized sketch, video, LYTRO and 3D images. Corresponding image pairs in the Tufts Face Dataset are currently not properly aligned and capture part of the upper body as seen in **Figure 1**.

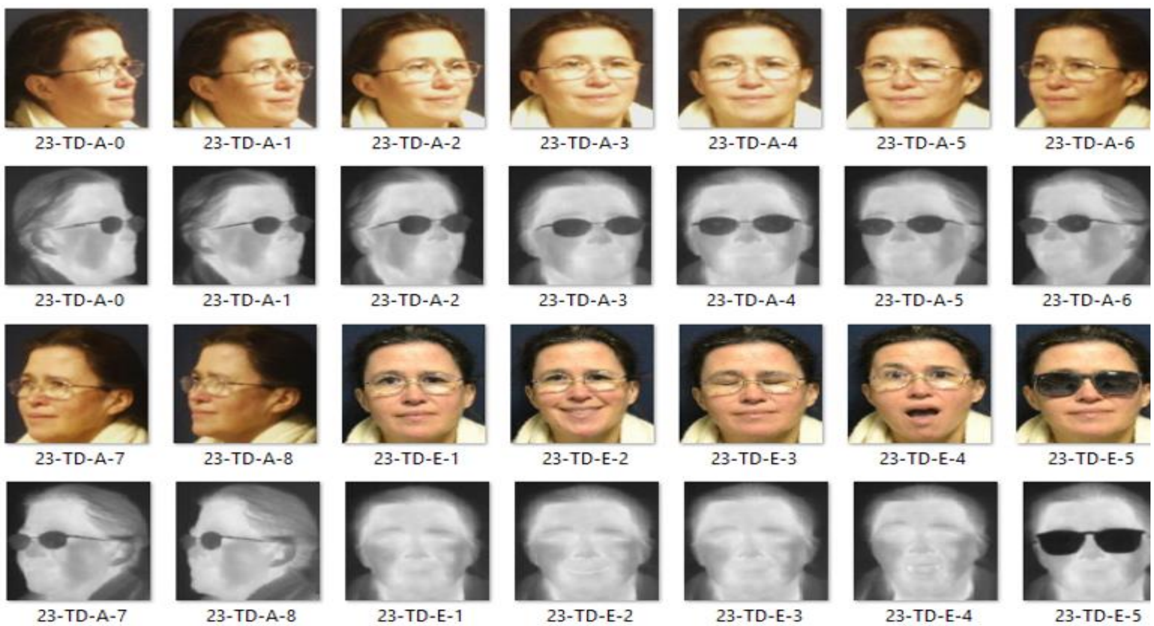


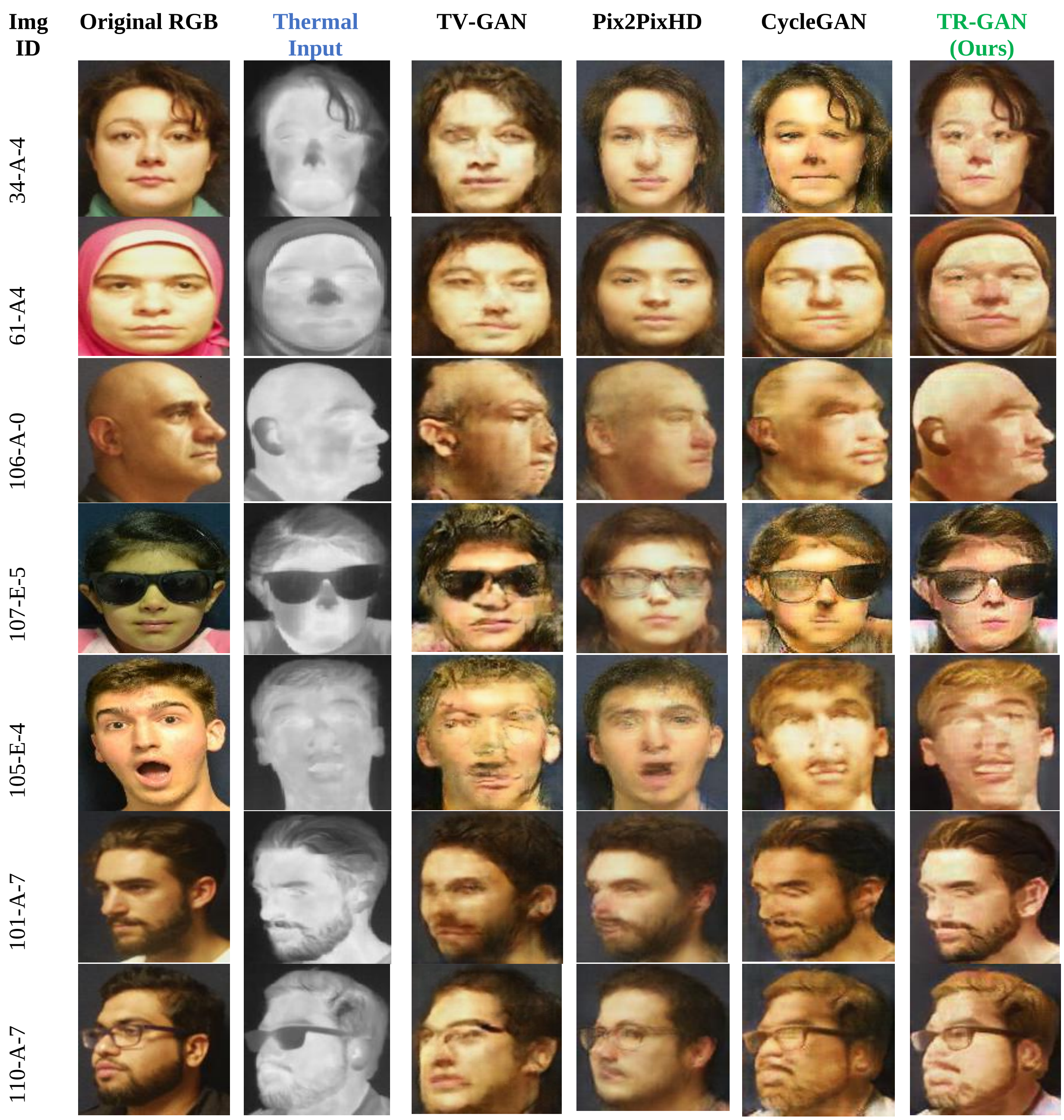

*Figure 3: Sample data of the normalized thermal-RGB dataset from Tufts Face Dataset. #23 indicate the subject ID number, $A_i$ denotes "Around" at ith angle, and $E_i$ denotes ith emotion*

Using the dataset as is would therefore cause the generator to generate images with misplaced eyes, nose or ears, and result in distorted reconstruction. To address this challenge, we derive a more face-focused and aligned dataset better suited for training GAN models for intermodal face synthesis. First, we extract matching pairs of thermal and RGB ground truth images and manually crop out the faces in each image, approximating the top of the hair, side of the ears and the bottom of the chin. The cropped faces are then resized to 128x128 pixels to normalize the dataset. The dataset contains 112 participants including 74 females and 38 males, from 15 countries. Furthermore, the dataset captures emotions such as neutral, smile, sleepy, surprised; occlusion with sunglasses; and side views at various angles. The proposed aligned dataset will be added as contribution to the Tufts Face Dataset currently available on Kaggle. **Figure 3** shows sample images of *subject #23* from the proposed normalized training set.

We randomly select 17 subjects for testing and use the rest for training. The training set contains a total of 1294 thermal images and corresponding 1294 RGB images, while the test set contains 238 thermal images and corresponding ground truth RGB. Training is conducted with a batch size of 4 and uses the Adam optimizer with an initial learning rate of 0.0002.

## 4.2 Results

Performance of the proposed model is compared against other state-of-the-art GAN algorithms such as TV-GAN [25], Pix2PixHD[29] and CycleGAN[28]. Synthesized images by each model as shown in **Figure 4** demonstrate that our model generates more realistic looking images and does a much better reconstruction job while preserving important and fine details.



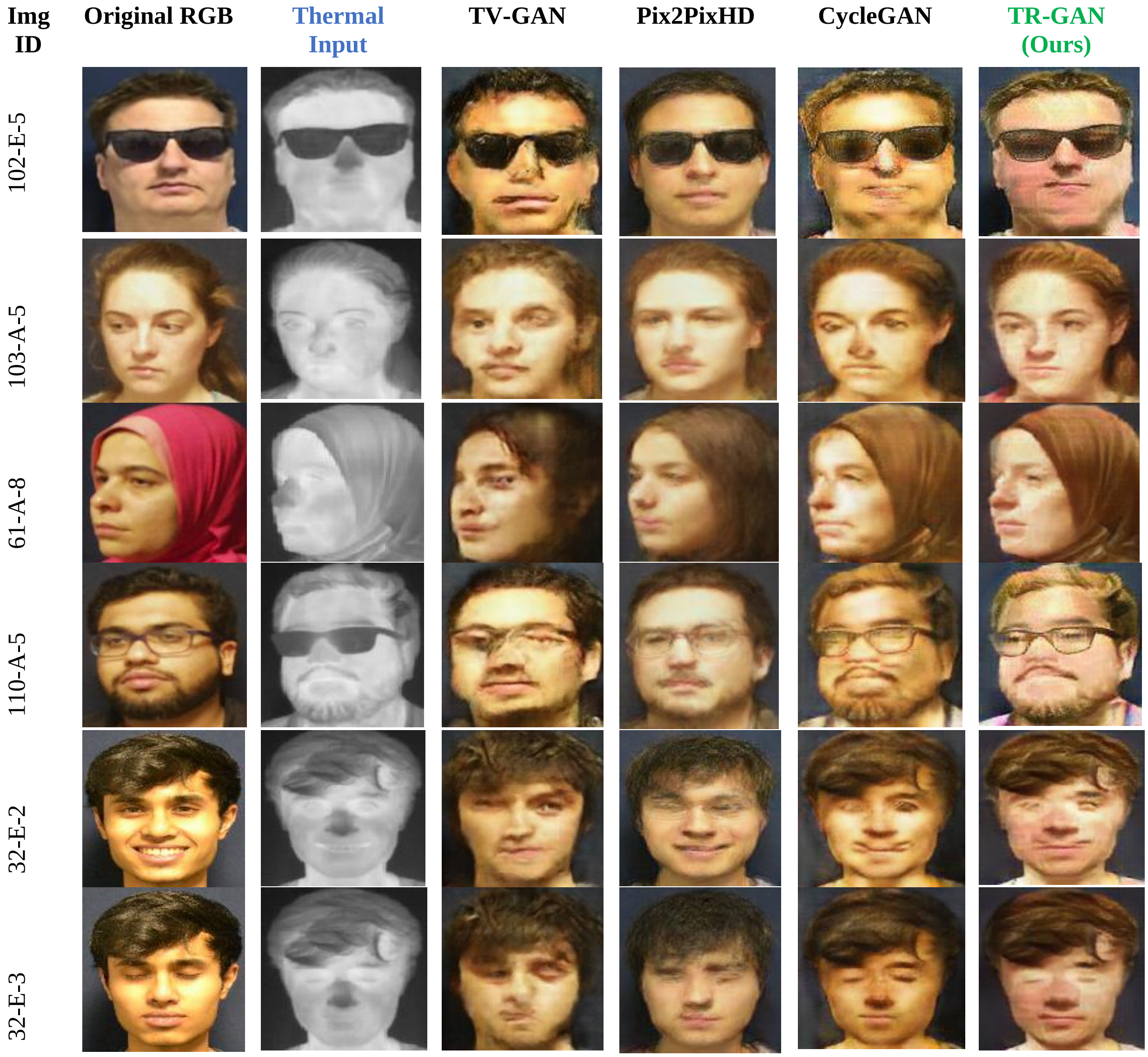



*Figure 4: Sample results of generated images by each model from the input thermal images. These images are from the test set and were not seen during training. Images in the first column are ground truth RGB images that models are trying to replicate, while images in second column are input thermal images to the model. Models' outputs are presented in each column.*

The TR-GAN model generates much sharper images with better matching features to the ground truth RGB image than other state-of-the-art models. Furthermore, TR-GAN can better synthesize intricate details such as facial emotions (surprised, smile, sleepy, neutral), sunglasses and corrective eyeglasses occlusions, and facial hair. Images generated by our model are therefore more visually appealing than other models' outputs, and much more similar to the ground truth RGB images as evidenced in sample results shown in **Figure 4**.

For the face recognition experiment, we use the VGG-Face recognition system [7], [31], [32], particularly the Keras implementation [32]. Ground truth RGB images from the test set consisting of 238 images from 17 subjects, are used to learn "*known embeddings*" against which generated images will be matched. Essentially, we reuse the VGG-Face model without modification nor retraining. Feature embeddings of synthesized RGB images by each model are extracted and compared against "*known embeddings"* to find potential matches. Recognition accuracies are as reported in the table below and is calculated as the ratio of the number of positive matches to the total number of test images.

*Table 1: Comparison of VGG-Face recognition accuracy on synthesized test images by various GAN models*

| | **Original RGB** | **Thermal** | **TV-GAN** | **Pix2PixHD** | **CycleGAN** | **TR-GAN (Ours)** |
|---|---|---|---|---|---|---|
| **Resnet50** accuracy (%) | 100.0 | 47.0 | 42.85 | 44.95 | 75.21 | **80.7** |
| **VGG16** accuracy (%) | 100.0 | 30.67 | 53.78 | 61.74 | 79.41 | **88.65** |

As evidenced in **Table 1**, the pretrained VGG-Face recognition model yields a **47%** accuracy on thermal images but up to **88.65%** recognition accuracy on the TR-GAN synthesized images. This demonstrates that bridging the domain gap by translating thermal images to their visible RGB counterpart using the proposed TR-GAN can almost double the face recognition accuracy. It is worth noting that the VGG-Face recognition accuracy on the ground truth RGB image is 100% for both "*resnet50*" and "*vgg16*" network models, which shows that VGG-Face is the adequate tool for properly evaluating impact of TR-GAN on cross-modal face recognition task.

## 5. CONCLUSION AND FUTURE WORK

In this paper, we introduced a sophisticated TR-GAN model capable of translating images from the thermal domain to the visible light RGB domain. The developed model (1) reduces the domain gap between the thermal and visible RGB domains, (2) synthesizes more realistic looking and more visually appealing images with finer details and better sharpness than existing state-of-the-art image-to-image translation GAN models, and (3) helps boost cross-modal recognition accuracy by using VGG-Face shelf recognition systems without the need to retrain the network. Experimental results show that the proposed TR-GAN model performs well reconstructing intricate details such as sunglasses, eyeglasses, facial emotions, and facial hair.

Furthermore, we introduced a normalized couple cross-modal (thermal-RGB) dataset as part of the Tufts Face Dataset to provide more suitable cross-modal data for training similar GAN models. The improved dataset eliminates the misalignment problem between paired data in the Tufts Face Dataset by manually cropping randomly sized paired images and adequately aligning and uniformly resizing them to 128x128 pixels. Moreover, by virtue its cycle consistency, the inverse mapping function of the proposed TR-GAN model can be used to create more accurate paired thermal samples from larger face datasets such as CelebA. Finally, VGG-Face recognition is performed on synthesized images to evaluate the benefits of the proposed model on cross-modal recognition task. Experimental results report **47%** cross-modal recognition accuracy on original thermal images given corresponding known ground truth RGB images, and up to **88.65%** on counterpart TR-GAN synthesized images. Therefore, translating thermal input images to their RGB counterparts using the proposed TR-GAN almost doubles the cross-modal recognition accuracy, which is also higher than the accuracy reported for other GAN models on the same test set.

For future work, we plan to integrate additional losses and incorporate attention mechanism to the generator and discriminator network to improve sharpness and color replication in synthesized output images, and further eliminate the crystallization effect. We also plan on analyzing facial expression recognition of visible, thermal, and fused imagery in an indoor and outdoor environment for human-computer interaction related applications.

## REFERENCES


[1] L. Wiskott, J. M. Fellous, N. Krüger, and C. der von Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775–779, 1997.

[2] P. Jonathon Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.

[3] P. J. Phillips *et al.*, "Overview of the face recognition grand challenge," in *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, 2005, vol. I, pp. 947–954.

[4] V. Bruce and A. Young, "Understanding face recognition," *British Journal of Psychology*, vol. 77, no. 3, pp. 305–327, Aug. 1986.

[5] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.

[6] X. He, S. Yan, Y. Hu, P. Niyogi, and H. J. Zhang, "Face recognition using Laplacianfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328–340, Mar. 2005.

[7] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face Recognition," British Machine Vision Association, 2015.

[8] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep Learning Face Attributes in the Wild *," 2015.

[9] "IMDB-WIKI - 500k+ face images with age and gender labels." [Online]. Available: https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/. [Accessed: 11-Mar-2020].

[10] S. Hu *et al.*, "A Polarimetric Thermal Database for Face Recognition Research," 2016.

[11] K. Panetta *et al.*, "A comprehensive database for benchmarking imaging systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Mar. 2018.

[12] S. Kamath K. M., R. Rajendran, Q. Wan, K. Panetta, and S. S. Agaian, "TERNet: A deep learning approach for thermal face emotion recognition," in *Mobile Multimedia/Image Processing, Security, and Applications 2019*, 2019, vol. 10993, p. 10.

[13] I. J. Goodfellow *et al.*, "Generative Adversarial Nets," 2014.

[14] "Introduction to generative adversarial network | Opensource.com." [Online]. Available: https://opensource.com/article/19/4/introduction-generative-adversarial-networks. [Accessed: 01-Mar-2020].

[15] "18 Impressive Applications of Generative Adversarial Networks (GANs)." [Online]. Available: https://machinelearningmastery.com/impressive-applications-of-generative-adversarial-networks/. [Accessed: 01-Mar-2020].

[16] J. Choi, S. Hu, S. S. Young, and L. S. Davis, "Thermal to visible face recognition," in *Sensing Technologies for Global Health, Military Medicine, Disaster Response, and Environmental Monitoring II; and Biometric Technology for Human Identification IX*, 2012, vol. 8371, pp. 83711L-83711L–10.

[17] "(PDF) Face Detection on Infrared Thermal Image." [Online]. Available: https://www.researchgate.net/publication/317013979_Face_Detection_on_Infrared_Thermal_Image. [Accessed: 10-Mar-2020].

[18] F. Juefei-Xu, D. K. Pal, and M. Savvides, "NIR-VIS Heterogeneous Face Recognition via Cross-Spectral Joint Dictionary Learning and Reconstruction." pp. 141–150, 2015.

[19] H. Zhang, V. M. Patel, B. S. Riggan, and S. Hu, “Generative adversarial network-based synthesis of visible faces from polarimetrie thermal faces,” in *IEEE International Joint Conference on Biometrics, IJCB 2017*, 2018, vol. 2018-January, pp. 100–107.

[20] X. Di, H. Zhang, and V. M. Patel, “Polarimetric thermal to visible face verification via attribute preserved synthesis,” in *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems, BTAS 2018*, 2018.

[21] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015, vol. 9351, pp. 234–241.

[22] B. S. Riggan, N. J. Short, and S. Hu, “Thermal to Visible Synthesis of Face Images Using Multiple Regions,” in *Proceedings - 2018 IEEE Winter Conference on Applications of Computer Vision, WACV 2018*, 2018, vol. 2018-January, pp. 30–38.

[23] B. S. Riggan, N. J. Short, S. Hu, and H. Kwon, “Estimation of visible spectrum faces from polarimetric thermal faces,” in *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems, BTAS 2016*, 2016.

[24] A. Mahendran and A. Vedaldi, “Understanding Deep Image Representations by Inverting Them,” 2015.

[25] T. Zhang, A. Wiliem, S. Yang, and B. Lovell, “TV-GAN: Generative adversarial network based thermal to visible face recognition,” in *Proceedings - 2018 International Conference on Biometrics, ICB 2018*, 2018, pp. 174–181.

[26] P. Isola, J.-Y. Zhu, T. Zhou, A. A. Efros, and B. A. Research, “Image-to-Image Translation with Conditional Adversarial Networks,” 2017.

[27] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, vol. 9906 LNCS, pp. 694–711.

[28] J.-Y. Zhu, T. Park, P. Isola, A. A. Efros, and B. A. Research, “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks Monet Photos,” 2017.

[29] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, “High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs,” 2018.

[30] Y. Zhang, Y. Tian, Y. Ukong, B. Zhong, and Y. Fu, “Residual Dense Network for Image Super-Resolution,” 2018.

[31] “Visual Geometry Group - University of Oxford.” [Online]. Available: http://www.robots.ox.ac.uk/~vgg/software/vgg_face/. [Accessed: 21-Mar-2020].

[32] “VGG-Face model for keras · GitHub.” [Online]. Available: https://gist.github.com/EncodeTS/6bbe8cb8bebad7a672f0d872561782d9. [Accessed: 21-Mar-2020].