IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 15, 2021, accepted December 31, 2021, date of publication January 18, 2022, date of current version January 27, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3144308

Cross-Spectrum Thermal Face Pattern Generator

XINGDONG CAO[1], KENNETH LAI[1], (Member, IEEE), GEE-SERN JISON HSU[2], (Senior Member, IEEE), MICHAEL SMITH[1], (Life Senior Member, IEEE), AND SVETLANA N. YANUSHKEVICH[1], (Senior Member, IEEE)

[1]Department of Electrical and Software Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada
[2]Artificial Vision Laboratory, Department of Mechanical Engineering, National Taiwan University of Science and Technology, Taipei City 10607, Taiwan

Corresponding author: Svetlana N. Yanushkevich (syanshk@ucalgary.ca)

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) through the Grant "Biometric-Enabled Identity Management and Risk Assessment for Smart Cities," in Collaboration with National Taiwan University of Science and Technology (NTUST), Taiwan; and in part by the Mitacs Globalink Graduate Fellowship, Canada.

ABSTRACT Conversion of a visible face image into a thermal face image (V2T), or one thermal face image into another one given a different target temperature (T2T), is required in applications such as thermography, human body thermal pattern analysis, and surveillance using cross-spectral imaging. In this work, we propose to use conditional generative adversarial networks (cGAN) with cGAN loss, perceptual loss, and temperature loss to solve the conversion tasks. In our experiment, we used Carl and SpeakingFaces Databases. Frèchet Inception Distance (FID) is used to evaluate the generated images. As well, face recognition was applied to assess the performance of our models. For the V2T task, the FID of the generated thermal images reached a low value of 57.3. For the T2T task, we achieved a rank-1 face recognition rate of 91.0% which indicates that the generated thermal images preserve the majority of the identity information.

INDEX TERMS Generative adversarial networks, image-to-image translation, thermal pattern generation, face recognition, biometrics.

I. INTRODUCTION

Both visible and thermal spectra provide useful biometric information on human subjects. Most biometric tasks, such as face detection and recognition, focus on the visible spectrum. Thermal cameras, unlike visible spectrum ones, allow for capturing the low-light scene. However, in most cases, thermal images cannot be used for face recognition given that legacy databases of faces contain only visible spectra images. In thermography for healthcare applications, the opposite image-to-image translation may be needed.

This paper focuses on two tasks. The first is dedicated to answering the question: *Is it possible to convert a face image taken in visible spectrum into an image taken in thermal spectrum*? The second task poses the question: *How to convert a thermal face image taken at a certain measured body temperature into a thermal image given a different temperature*?

The associate editor coordinating the review of this manuscript and approving it for publication was Zahid Akhtar.

In this paper, we build a deep learning model to solve these two image-to-image translation tasks. We use a conditional generative adversarial network (cGAN) [1], conditioned on input visible or thermal images, to generate the output images. Our cGAN consists of a 'U-Net' generator (G) and a PatchGAN discriminator (D). We use the Carl Database [2] and the SpeakingFaces Database [3] containing visible and thermal images to train our model. For converting a visible face image into a thermal face image (V2T), we use the SpeakingFaces Database which contains paired visible face images and thermal images to train our model, with a cGAN and Mean Absolute Error (MAE) loss. For the second task, converting a thermal face image into another thermal face image with the target temperature (T2T), we modify our model to include temperature information. We use the Carl Database containing thermal faces with different temperatures to train our model. The model is trained using a combination of cGAN loss, perceptual loss [4], and temperature loss. We extend our previous work [5], by applying the

Frèchet Inception Distance (FID) [6] and face recognition techniques to measure the performance of the proposed cGAN. Applying face recognition techniques on the resulting generated thermal images shows an acceptable performance.

The main contributions of our work are listed below:

1) We propose to use cGAN to solve two cross-spectrum image-to-image translation tasks, V2T and T2T. To solve these tasks, we modified the structure of cGAN as $cGAN_{V2T}$ and $cGAN_{T2T}$. To the best of our knowledge, it is the first time that the T2T task is approached this way.
2) Two different databases are used to train $cGAN_{V2T}$ and $cGAN_{T2T}$, with different loss function combinations.
3) We evaluate the performance of our two cGAN models using FID and face recognition techniques.

The paper is organized as follows: Section II provides a literature review related to GAN and image to image translation tasks. Section III describes the architecture of the proposed cGAN. Section IV provides information on the databases used for training and describes the pre-processing steps applied to the images. Section V summarizes the results of the proposed cGAN application. Section VI concludes the paper.

II. RELATED WORK

A. GENERATIVE ADVERSARIAL NETWORKS

GANs were proposed in 2014 [7], followed by multiple successful applications in various fields [8]. A GAN is comprised of a generator (G) and a discriminator (D). D is trained on real samples from a collected database and fake samples generated by G and is responsible for classifying samples as real or fake. G is to generate fake samples that are real enough to fool D. The adversarial process is formulated as a minimax game [7]:

$$\min_G \max_D \mathbb{E}_{(x\sim p_{data})}[\log D(x)] + \mathbb{E}_{(z\sim p_z)}[\log(1 - D(G(z)))] \tag{1}$$

where the distribution of the real sample x is denoted by p_{data}, p_z represents the distribution of the noise input z, and $\mathbb{E}$ is to calculate the expectation of the following expression. By training G and D together, they compete with each other and achieve an equilibrium where G can implicitly learn the distribution of the collected samples, without the need of any crafted loss functions.

However, GANs suffer from adversarial training instability and mode collapse problems [9], [10]. An improved architecture called DCGAN was proposed in [9] to resolve the training instability problem, by using CNN and batch normalization [11]. Most recently, Wasserstein GAN [12], [13] and LSGAN [14] were proposed to improve the adversarial training stability and alleviate the mode collapse problem.

In our work, we adopt cGAN [1], with G aiming to generate thermal face images, conditioned on input visible or thermal images. Many image-to-image translation tasks, such as converting semantic labels to real city photos or converting architectural labels to real structural images, were approached using the cGAN to generate target images [15]. For the V2T task considered in our paper, the condition is the input visible image, and the proposed cGAN is built to generate a corresponding thermal image. This task is similar to the one described in [16] and [17]. K. Lai *et al.* [16] used cycle-GAN to convert images between visible and thermal domains. K. Landry *et al.* [17] uses GAN to convert images from thermal spectrum into visible. In our experiment, we focus on converting images from visible spectrum into thermal. For the T2T task, the condition includes both the input thermal image and a target temperature, and the proposed cGAN is constructed to generate a thermal image given a target temperature.

B. FACIAL ATTRIBUTE EDITING

Facial attribute editing aims to manipulate single or multiple attributes of a given face image, while preserving its identity details [18]. In contrast, our approach to T2T aims to edit the temperature attribute for a given thermal face. Many methods have been proposed recently to solve the facial attribute editing problem, and most of them have only been applied to visible face images. Li *et al.* [19] trained a deep identity-consistent attribute transfer model to add/remove an attribute to/from a visible face image. To do so, they employed an adversarial attribute loss and a deep identity feature loss. Shen and Liu [20] adopted the dual residual learning strategy to simultaneously train two networks by adding and removing a specific attribute. StarGAN [21] performed the facial attribute editing based on the attribute labels by introducing an attribute classification loss [22] and a cycle consistency loss [23], which can perform multiple attribute editing simultaneously using only a single model.

Wang *et al.* [24] proposed an IPCGAN with a cGAN module, an identity-preserved module, and an age classifier to address the face aging problem by editing the age attribute on a given face. Unlike this approach, in our paper, facial attribute editing is applied to the thermal spectrum, i.e., manipulating the temperature attribute of thermal face images. In our T2T task, we use thermal images from the Carl Database [2] to train our cGAN, with an adversarial, perceptual, and temperature loss function.

III. PROPOSED METHOD

In our work, we use cGAN to solve both of the image-to-image translation tasks. Our implementation of cGAN consists of a generator (G) and a discriminator (D). The minor differences in the structures of the G and D for V2T and T2T will be described in detail in Subsection III-A. For V2T conversion, the input of G_{V2T} is a visible image, and the expected output is a thermal image. The input of D_{V2T} includes both a visible image and a thermal image, and the output indicates whether the input thermal image is the real paired image of the input visible image. For T2T conversion, the input of G_{T2T} is a thermal image and a target temperature, and the output is the generated image with a thermal pattern

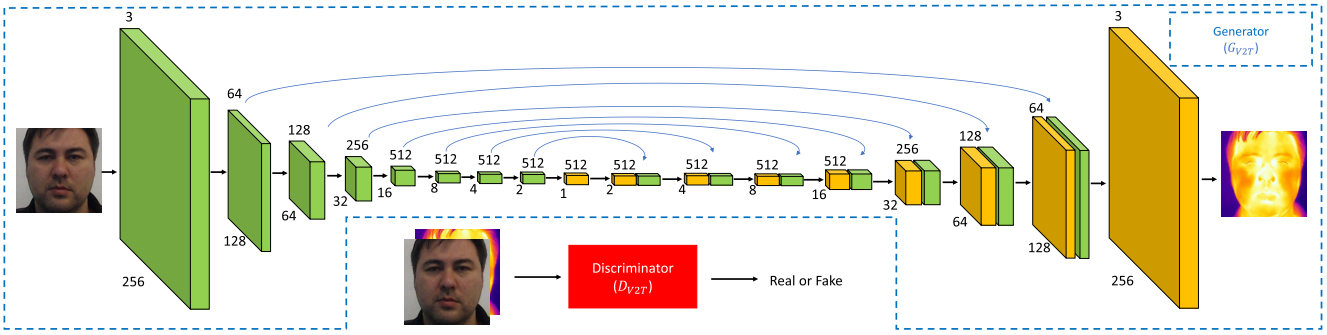


FIGURE 1. cGAN network for the V2T conversion. The generator (G_{V2T}) converts visible images into thermal images. G_{V2T} consists of an input block (the left-most green cuboid), 7 encoding blocks (the following 7 green cuboids), a bottleneck block (the left-most yellow cuboid), 7 decoding blocks (the following 7 yellow cuboids) and an output block (the right-most yellow cuboid). There is a direct skip (blue curve arrow) connecting each encoding-decoding pair. The Discriminator (D_{V2T}) is to determine whether the input visible and thermal images are paired. The structure of D_{V2T} is similar to the D_{T2T} presented in Fig. 2, and the minor difference will be described in Section III-A.

corresponding to the target temperature. The input of D_{T2T} is a thermal image, and the output indicates whether the input thermal image is real or generated. By training each G and D together, they should reach an equilibrium such that G generates a thermal image that is real enough to fool D.

A. PROPOSED FRAMEWORKS

The following is an overview of the cGAN used, similar to the cGAN described in [15], with a discussion of necessary changes. We use 'U-Net' [25] as our generator (G) and a 6-layer PatchGAN as our discriminator (D) for both conversions. Both possess the convolution-BatchNorm-LeakyRelu [9] blocks.

Fig. 1 shows the network structure of our G_{V2T}. Similar to our previous design [26], G_{V2T} consists of an input block, 7 encoding blocks, a bottleneck, 7 decoding blocks, and an output block. Each encoding block down-samples the previous block by 1/4 (1/2 of length and 1/2 of width) with *strides* $= 2$, and each decoding block up-samples the previous block by 4 times. For the i^{th} decoding block, we concatenate it with the last i^{th} encoding block in the channel dimension, before applying the LeakyRelu activation. The filter size is set to $4*4$ for all blocks. The filter number is set to 64 for the first encoding block, and doubles for each of the following blocks until it reaches 512; after that, it remains at 512. The filter number for each decoding block is the same as the encoding block connected to it. For the bottleneck block, the filter number is set to be 512, and the activation function is ReLU. For the output block, the filter number is set to 1, and the activation function is Sigmoid.

Fig. 2 shows the PatchGAN structure of D_{T2T}. It is a 6-layer CNN, with the number of filters for the layers set to 64, 128, 256, 512, 512, and 1, respectively. For the first four layers, the stride is set to 2, and for the last two layers, the stride is set to 1. The output is a $16 * 16$ matrix, and each value is to map a $70 * 70$ receptive field in input as real or generated.

The input layer of G_{T2T} and D_{T2T} are slightly different from G_{V2T} and D_{V2T}. For G_{T2T}, we add one more channel to incorporate the target temperature information so that the input layer has 4 channels, while for G_{V2T} it has 3 channels. For D_{T2T}, we only input the thermal images into the networks so that the input block has 3 channels, therefore, D_{V2T} has 6 channels with the concatenated visible and thermal images as input.

B. OBJECTIVE FUNCTION FOR V2T

The objective function of $cGAN_{V2T}$ is defined as follows:

$$\begin{aligned}\mathcal{L}^{cGAN}_{G_{V2T}}(G_{V2T}, D_{V2T}) &= \mathbb{E}_{(x)} \log(1 - D_{V2T}(x, G_{V2T}(x))) \\ \mathcal{L}^{cGAN}_{D_{V2T}}(G_{V2T}, D_{V2T}) &= -\mathbb{E}_{(x,y)} \log D_{V2T}(x, y) \\ &\quad -\mathbb{E}_{(x)} \log(1 - D_{V2T}(x, G_{V2T}(x)))\end{aligned} \tag{2}$$

where x and y are paired visible-thermal images. G_{V2T} tries to minimize $\mathcal{L}_{G_{V2T}}$, and D_{V2T} aims to minimize $\mathcal{L}_{D_{V2T}}$. Additionally, we appended the Mean Absolute Error (MAE) loss to help the generator converge faster, and also preserve the identity information of the ground-truth thermal images in the pixel-order. The MAE loss can be calculated by averaging the absolute difference between the pixel values of the same coordinate:

$$\mathcal{L}^{MAE}(G_{V2T}) = \mathbb{E}_{(x,y)} \frac{1}{V} [||y - G_{V2T}(x)||_1] \tag{3}$$

where V stands for the number of pixels in the input or output images. Combining these losses together, our final objective is expressed as follows:

$$\begin{aligned}\mathcal{L}^{cGAN}_{G_{V2T}} &= \mathcal{L}_{G_{V2T}} + \alpha \mathcal{L}^{MAE} \\ \mathcal{L}^{cGAN}_{D_{V2T}} &= \mathcal{L}_{D_{V2T}}\end{aligned} \tag{4}$$

where α controls the weight of $\mathcal{L}^{MAE}$ with respect to $\mathcal{L}_{G_{V2T}}$. Here, we set $\alpha = 100$, determined by the best visual effect of generated thermal images on the training set.

C. OBJECTIVE FUNCTION FOR T2T

The objective function of $cGAN_{T2T}$ conversion can be written as follows:

$$\mathcal{L}^{cGAN}_{G_{T2T}}(G_{T2T}, D_{T2T}) = \mathbb{E}_{(x)} \log(1 - D_{T2T}(G_{T2T}(x, t)))$$

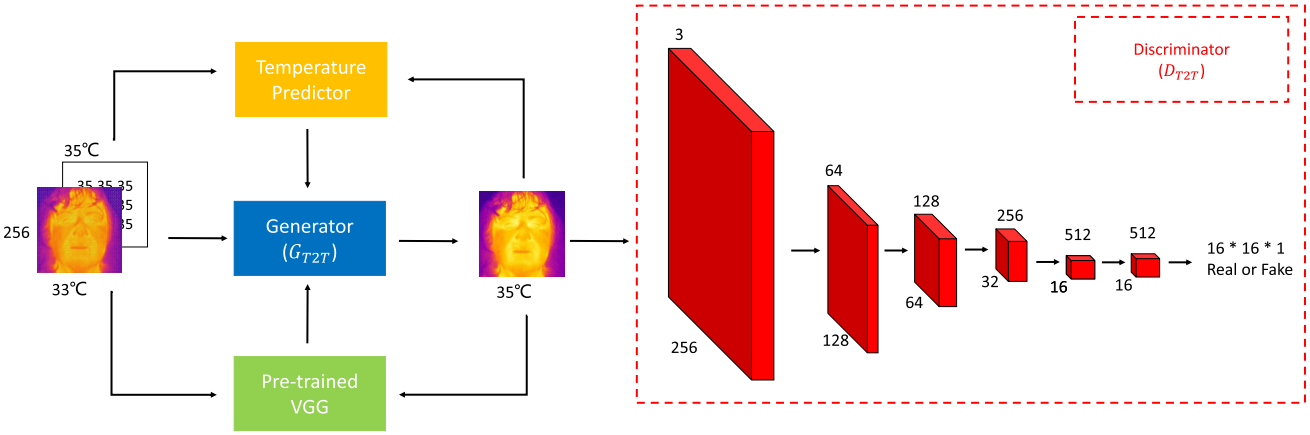


FIGURE 2. cGAN network for the T2T conversion. G_{T2T} converts a thermal image into another thermal image with target temperature, and we add one more channel to represent the temperature information in the input block. The temperature predictor is pre-trained to provide the temperature loss. VGG is to provide perceptual loss. D_{T2T} is a 6-layer PatchGAN that determines whether the input thermal image is real or generated.

$$\mathcal{L}_{D_{T2T}}^{cGAN}(G_{T2T}, D_{T2T}) = -\mathbb{E}_{(x)} \log D_{T2T}(x) \\ -\mathbb{E}_{(x)} \log(1 - D_{T2T}(G_{T2T}(x,t))) \tag{5}$$

where x is the input thermal image and t is the target temperature. Additionally, we use the perceptual loss $\mathcal{L}^{prp}$ as proposed by Johnson *et al.* [4]. It is formed using the features extracted from selected layers of the pre-trained VGG network [27] given by:

$$\mathcal{L}^{prp}(G_{T2T}) = \mathbb{E}_{(x)} \sum_{i=1}^{N} \frac{1}{V_i} [||F^{(i)}(x) - F^{(i)}(G_{T2T}(x,t))||_1] \tag{6}$$

where $F^{(i)}$ represents the i^{th} layer with V_i activations of the VGG network, and N is the number of selected layers in the VGG model. In our work, we empirically choose 4 layers of the VGG network as F, to calculate $\mathcal{L}^{prp}$. With $\mathcal{L}^{prp}$, we preserve the similarity of the high-level features of the input and output thermal images.

In addition, we use a temperature loss to enforce the generated thermal image fit into the target temperature. To get the thermal pattern corresponding to the target temperature, we pre-train a temperature predictor. We label each thermal image by the value of temperature, based on the temperature matrix provided by [2]. Our temperature predictor has the same architecture as Alexnet [28], except that the last fully-connected layer has only one unit with *tanh* activation. The temperature loss is defined as follows:

$$\mathcal{L}^{temp}(G_{T2T}) = \mathbb{E}_{(x)} \sigma\left(TP(G_{T2T}(x,t)), t\right) \tag{7}$$

In the above equation, $\sigma()$ corresponds to an MAE loss, and $TP()$ is our trained temperature predictor. During back-propagation, $\mathcal{L}^{temp}$ forces the parameters of G_{T2T} to change and generate faces that represent the target temperature. Combining these losses together, our final objective is expressed as follows:

$$\mathcal{L}_{G_{T2T}} = \mathcal{L}_{G_{T2T}}^{cGAN} + \alpha L^{prp} + \beta L^{temp} \\ \mathcal{L}_{D_{T2T}} = \mathcal{L}_{D_{T2T}}^{cGAN} \tag{8}$$

where α and β are set at 100 and 500, determined by the criterion of the best visual effect of the generated thermal images from the training set.

The reason that we use $\mathcal{L}^{prp}$ rather than $\mathcal{L}^{MAE}$ for the T2T task is that increasing or decreasing the temperature of the input thermal face image inevitably changes the pixel values of the image, thus causing a large $\mathcal{L}^{MAE}$. To some extent, $\mathcal{L}^{MAE}$ and $\mathcal{L}^{temp}$ are contradictory in the T2T task. However, one of our goals for the T2T task is to keep the subject identity information, so we use $\mathcal{L}^{prp}$ to preserve the high-level features of the input thermal faces.

IV. EXPERIMENT SETUP

A. DATABASES

In this paper, we use two databases. The first database is the Carl Database [2], in which visible and gray-scale thermal images, containing human faces, are collected simultaneously, using a TESTO 880-3. The database contains 41 subjects. For each subject, four image acquisition sessions were performed within two months, each with three different lighting settings (natural, infrared, and artificial), and five images for each lighting setting. This generates $41 * 4 * 3 * 5 = 2,460$ visible-thermal image pairs. However, the visible-thermal image pairs are not aligned, and, thus, cannot be used to train the $cGAN_{V2T}$ directly. We will provide details of our alignment process in Section IV-B. In addition, this database contains the temperature matrix corresponding to the thermal images, which provides the possibility to generate a thermal image with a different target temperature. In our work, to keep it consistent with another database, we used the Testo software [29] to convert the raw data into thermal images using an 'iron' palette. We used these images instead

of the provided gray-scale thermal images, to train and test the proposed model.

The second data set is the SpeakingFaces Database [3], which provides much more well-aligned visible-thermal face image pairs, compared to the Carl Database. In the SpeakingFaces Database, subjects are required to read commands, and their voice, thermal faces, and visible faces are recorded by microphones, FLIR T540 thermal camera, and Logitech C920 Pro HD web-camera simultaneously. The database contains 142 subjects. In our work, images of subjects $1-21$ are not used due to the incompleteness of these samples. For each subject, the image acquisition is conducted using two trials, each trial involves 8,100 frames. To record the images, the cameras are placed in 9 different positions in order to acquire the face images from 9 different angles, and 900 frames each. In our work, only the face images taken in the front angle from SpeakingFaces Database are used. Due to the high volume of the database and the high similarity between the images taken subsequently, we sampled only 5% of the frontal face images. Altogether, we use $122 * 2 * 900/20 = 10,980$ thermal-visible image pairs. The thermal images and visible images are aligned. However, unlike the Carl Database, the temperature information of the thermal images was not available in this database.

B. ALIGNMENT, FACE EXTRACTION AND RESIZING

The following is an overview of our approach to pre-processing of the two databases. We consider the resolution of $256 * 256$ of the face images used to train our two cGAN models. In the Carl Database, the visible and thermal images were not aligned and, thus, could not be used for training or testing directly. We applied the facial landmarks annotated manually by Alperen [30], and coordinated the mapping to extract the faces from visible and thermal images. We then resized the extracted faces to achieve the resolution of $256 * 256$. Fig. 3 shows an example of an original visible-thermal face image pair and the same image pair after the alignment and resizing. Based on the 6 facial-landmark-position pairs (blue points) provided by Alperen, we learn the coordinate mapping between the visible image and the thermal image. Details of the coordinate mapping will be described in the next paragraph. Next, we applied the pre-trained face detector tool `dlib` [31] to extract the face from the visible image and expand it by a factor of 1.3, in order for the whole face (solid green box) to be extracted. The learned coordinate mapping is used to map the solid green box within the visible image into the thermal image; this results in extracting the thermal face (dashed green box). At last, we resize the two extracted faces to achieve a resolution of $256 * 256$ and get the aligned visible-thermal face image pair.

Below we explain the coordinate mapping process. The Cartesian coordinates of the six points in the visible image are annotated as follows: $\{x_{vn}, y_{vn}\}$, $1 \leq n \leq 6$. The same coordinates in the thermal image are denoted as $\{x_{tn}, y_{tn}\}$, $1 \leq n \leq 6$. We map the points in the visible images into the points

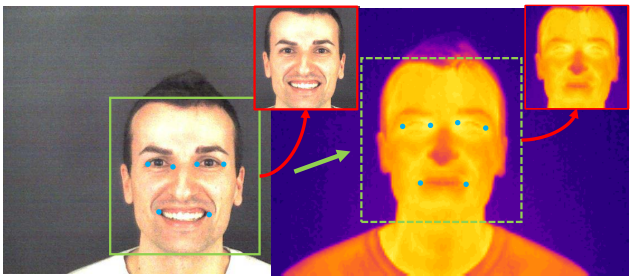


FIGURE 3. Alignment and resizing operation of both the visible and thermal images. Six blue points in each image are manually annotated [30]. The solid green box in the visible image was detected by the 'dlib' face detector. By learning the coordinate transformation through six-point pairs, we map the solid green box within a visible image into the dashed green box within a thermal one. Next, the two green boxes are resized to reach the $256 * 256$ resolution (red boxes).

in the thermal images through the linear transformation as described below:

$$\begin{aligned} a_x \times x_{vn} + b_x &= x_{tn} \\ a_y \times y_{vn} + b_y &= y_{tn} \end{aligned} \quad 1 \leq n \leq 6 \tag{9}$$

where a_x, b_x, a_y, b_y are the coefficients to control the linear mapping. We calculate the coefficients by minimizing the squared error:

$$\begin{aligned} a_x, b_x &= \arg\min_{a_x, b_x} \sum_{n=1}^{6} (a_x \times x_{vn} + b_x - x_{tn})^2 \\ a_y, b_y &= \arg\min_{a_y, b_y} \sum_{n=1}^{6} (a_y \times y_{vn} + b_y - y_{tn})^2 \end{aligned} \tag{10}$$

Since the visible and thermal images in the SpeakingFaces Database are already aligned, we directly use the `dlib` to extract the face from the visible images, and then expand it by a factor. We use the same coordinates to extract the thermal face. We then resize the extracted faces to achieve a resolution of $256 * 256$ and form the training set for the two cGAN models.

C. TRAINING AND TESTING FOR V2T

To train our V2T model, we use the visible-thermal image pairs of subjects $21-142$ from the SpeakingFaces Database. We use images collected from subjects $41-142$, all together $102 * 2 * 45 = 9,180$ image pairs, to train the model. Images collected from subjects $21-40$, provide $20 * 2 * 45 = 1,800$ images pairs, to test the model. The FID metric [6] is applied in our case to evaluate the similarity between the generated thermal images and the ground-truth thermal images. The FID is computed by passing the generated and real images into the pre-trained InceptionV3 model [32] and using the difference from the last pooling layer:

$$\text{FID} = ||\mu_r - \mu_g||^2 + tr(\Sigma_r + \Sigma_g - 2\sqrt{\Sigma_r \Sigma_g}) \tag{11}$$

where μ represents the mean for the real (r) and generated (g) images, Σ represents the covariance for the real (r) and generated (g) images, and tr is the trace linear function. The

TABLE 1. The forehead temperature range and distribution for the thermal images in Carl Database.

Temperature Range (°C)	# images (percent)
≤ 32.5	140 (5.8 %)
$32.5 - 33.0$	186 (7.8 %)
$33.0 - 33.5$	483 (20.2 %)
$33.5 - 34.0$	784 (32.7 %)
$34.0 - 34.5$	577 (20.1 %)
≥ 34.5	226 (9.4 %)

TABLE 2. Average identification performance (%) and standard deviation across two-fold cross-validation for V2T conversion. MobileNet performs the best (bolded) in three models.

	Models	True Acceptance Rate @ 1% FAR	True Acceptance Rate @ 0.1% FAR	Rank-1
Thermal Test	InceptionV3	38.5 ± 0.9	16.4 ± 0.3	31.1 ± 0.7
	Xception	43.6 ± 0.2	19.6 ± 0.4	35.6 ± 0.4
	MobileNet	$\mathbf{62.8 \pm 2.1}$	$\mathbf{41.0 \pm 0.3}$	$\mathbf{52.0 \pm 0.8}$
Thermal Valid	InceptionV3	55.0 ± 1.6	29.7 ± 0.7	45.1 ± 0.7
	Xception	62.2 ± 1.5	36.4 ± 0.3	51.4 ± 1.1
	MobileNet	$\mathbf{85.0 \pm 1.3}$	$\mathbf{67.2 \pm 1.5}$	$\mathbf{73.4 \pm 1.6}$
Visible Valid	InceptionV3	69.1 ± 0.9	47.2 ± 0.5	58.1 ± 0.0
	Xception	75.8 ± 2.5	55.1 ± 3.3	64.8 ± 2.3
	MobileNet	$\mathbf{86.1 \pm 0.2}$	$\mathbf{71.7 \pm 0.7}$	$\mathbf{76.2 \pm 0.1}$

lower FID score means that the quality of images generated by the generator is similar to the real ones; for example, the FID of 0 indicates that the generated and the ground-truth images are identical. Note that FID can vary between 0 and 600 in some cases.

D. TRAINING AND TESTING FOR T2T

In our experiment, we use images collected from subjects $1-30$, all together $30*4*3*5 = 1,800$ images, to train the model. The images collected from subjects $31 - 41$, totally 660 images, are used to test the model. The temperature of the centre of the subject's forehead on the thermal face image is considered to be the ground truth temperature. The temperature distribution of the thermal images in the different ranges is given in Table 1, which indicates that this is an unbalanced database. We used 80% of the images to train our temperature predictor, and use the remaining 20% to test the predictor. We linearly scale the temperature into [-1, 1] so that it falls into the range of *tanh* output. We use MAE to evaluate our temperature predictor. Note that the MAE is 0.27 °C in the testing database.

E. FACE RECOGNITION IN THERMAL SPECTRUM

In order to verify the similarity of the original and the generated face in both T2T and V2T, we apply face recognition in both thermal and visible spectra. The following methodology was used to evaluate the performance of such face recognition.

Face recognition is applied in this work to find the real identity of a test sample among all subjects in the database. We apply transfer learning [33] on three pre-trained models: InceptionV3 [32], Xception [34], and MobileNet [35].

The face recognition in T2T was performed using four-fold cross-validation, with each fold representing a different acquisition session. For V2T conversion, we use two-fold cross-validation. The number of folds is determined by the session numbers available in each database.

We train each CNN model with the images from three (for T2T) or one (for V2T) session(s) and validate it with the images from the remaining session. After fine-tuning the optimal parameters for the particular validation set, we test the final performance of the CNN with the synthesized thermal images. This way the parameters are adjusted using the criterion of the performance of the face generation. We compare the performance of the real and generated thermal images. We also perform face recognition in the visible spectrum for comparison.

Our approach employs Transfer Learning. It involves loading the pre-trained weights optimized for the ImageNet challenge first. The last fully-connected layer and classification layer from each model are removed, and an average pooling layer, two fully-connected layers with 512 units, and a classification layer are added to each model. Two training processes are applied sequentially. Initially, the model is trained with the other layers frozen, allowing only the newly added layers to update, with a higher learning rate for 50 epochs. Next, the entire model is trained with a lower learning rate for 50 epochs with a fine-tuning purpose [16]. Due to the high volume of the SpeakingFaces Database, for the face recognition in the V2T task, we resize the images to achieve the $128*128$ resolution. For the T2T task, we keep the resolution of $256 * 256$.

V. RESULTS

A. V2T GENERATED IMAGES

For V2T conversion, we use three instances of visible images: ground-truth thermal images using the 'iron' palette, generated 'iron' palette thermal images, and the corresponding FID values, as shown in Fig. 4. The generated thermal images reach satisfactory visual effects. For some parts of the generated images, our model does not perform well, such as a thermal image in the first column that seems to have an artifact in the hair part. For the second subject, the generated thermal image cannot reproduce the curly hair in the forehead. We use FID to measure the similarity between the generated thermal images and the ground-truth thermal images. The FID of all images in the test set achieved the value of 57.3 which is reasonably low compared to the maximum FID values that can reach 600 or higher.

1) V2T FACE RECOGNITION

The performance of face recognition (using $1 : N$ comparison) is evaluated by the rank performance and the True Acceptance Rate (*TAR*) at the targeted 1% and 0.1% False Acceptance Rate (*FAR*). For each subject, we calculate

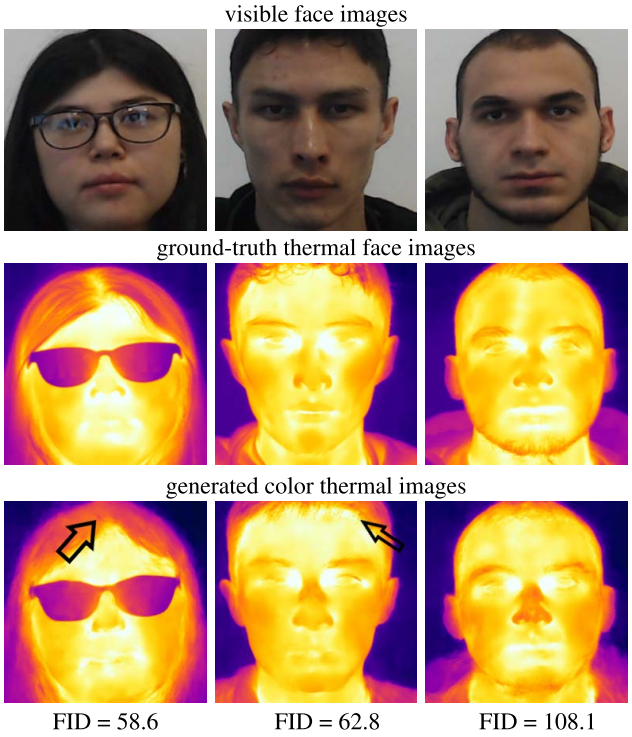


FIGURE 4. Three instances of visible images, ground-truth 'iron' thermal images, generated 'iron' thermal images, and the corresponding FID values. The black arrows show the parts that are not preserved well.

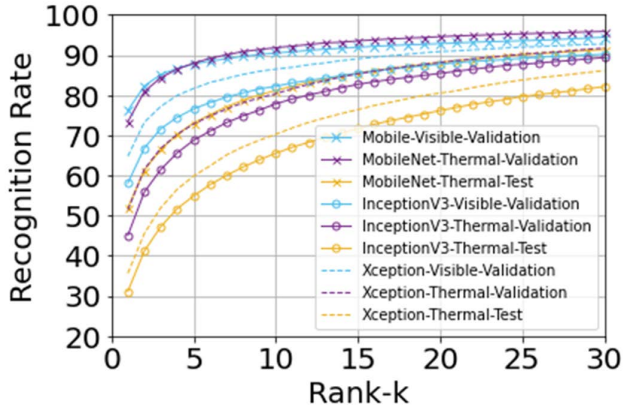


FIGURE 5. CMC curve for InceptionV3, Xception, and MobileNet using the SpeakingFaces database. Test case includes the synthesized thermal images, the validation case is composed of real thermal images, and the visible validation consists of visible images. MobileNet performs the best than the other two models. Performance on validation (Valid) is better than on testing (Test).

$TAR = TP/(TP + FN)$, where TP and FN represent the number of true positive and false negative, respectively. The overall TAR was taken as the average TARs for each subject. For evaluation, the targeted 1% or 0.1% FAR corresponds to a specific acceptance threshold. The testing set was formed based on the synthesized thermal images using the $cGAN_{V2T}$, and the validation set was formed using original thermal images.

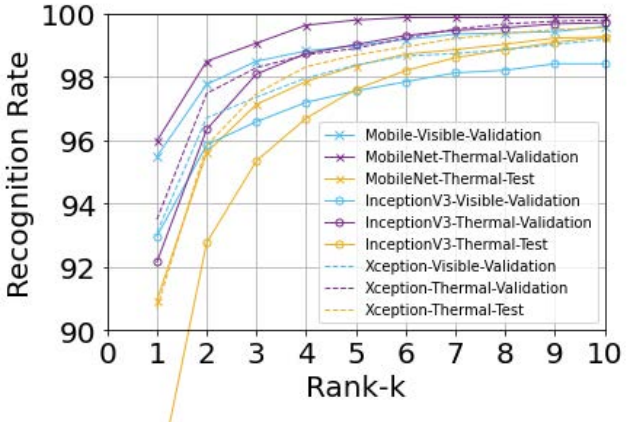


FIGURE 6. CMC curve for InceptionV3, Xception, MobileNet using the Carl database. The test case consists of the synthesized thermal images, validation case includes real thermal images, and visible validation consists of visible images. MobilNet performs the best than other two models. Performance on validation (Valid) is slightly better than on testing (Test).

Table 2 illustrates the performance for the validation set (Valid) and test set (Test) of the three different networks. Valid shows higher *TAR* because the validation images have the same spectrum as the training images. Test produces lower *TAR* since Test are formed from generated images. For testing, the MobileNet network demonstrates the best performance in terms of *TAR* at the targeted *FAR* and rank-1, and the Xception model reports the worst result.

Fig. 5 shows the Cumulative Matching Characteristic (CMC) for the face recognition rate of the three network models when accepting identities at ranks between 1 to 30. All the models perform better on Valid compared to Test, which indicates the loss of identity information following the V2T conversion. MobileNet performs the best, while InceptionV3 performs the worst.

We used $cGAN_{V2T}$ to convert visible face images into thermal face images, and FID and face recognition techniques to evaluate the generated thermal images. These images keep similarity to the ground-truth thermal images, and the overall FID reached 57.3. The rank-1 face recognition rate of Test is lower than Valid (14.0% for InceptionV3 and 15.8% for Xception). MobileNet has a better performance than the other two networks, with regard to TAR and rank-1 accuracy, on all of the sets.

B. T2T GENERATED IMAGES

For T2T conversion, we list four instances of the input and generated thermal images with different target temperatures as shown in Fig. 7. The four input thermal images have four different original temperatures, from lowest to highest. As seen in Fig. 7, from left to right, thermal images become lighter due to their higher facial temperature. We also used the trained temperature predictor to predict the temperatures of the generated thermal images, and list the MAE in Table 3.

TABLE 3. Performance summary on generated thermal images with different target temperatures.

target T(°C)	True Acceptance Rate @ 1% FAR	@ 0.1% FAR	Rank-1	MAE(°C)
33.0	92.8 ± 3.3	79.5 ± 8.8	87.1 ± 4.9	0.15
33.5	96.9 ± 3.1	88.4 ± 9.0	91.7 ± 4.7	0.15
34.0	97.3 ± 2.5	89.4 ± 7.8	92.5 ± 3.5	0.18
34.5	97.3 ± 2.2	90.3 ± 8.0	92.4 ± 4.8	0.20

TABLE 4. Average performance (%) and standard deviation across four-fold cross-validation for T2T conversion.

	Models	True Acceptance Rate @ 1% FAR	@ 0.1% FAR	Rank-1
Thermal Test	InceptionV3	92.2 ± 3.1	74.5 ± 7.6	85.1 ± 3.4
	Xception	**96.4 ± 3.4**	85.1 ± 9.3	90.7 ± 4.1
	MobileNet	96.2 ± 3.0	**87.0 ± 8.5**	**91.0 ± 4.3**
Thermal Valid	InceptionV3	97.8 ± 1.5	88.8 ± 4.7	92.2 ± 2.5
	Xception	97.8 ± 3.1	90.8 ± 10.1	93.5 ± 4.6
	MobileNet	**99.2 ± 1.0**	**95.2 ± 5.0**	**96.0 ± 2.8**
Visible Valid	InceptionV3	96.5 ± 3.8	90.5 ± 7.2	93.0 ± 4.0
	Xception	97.1 ± 3.2	89.8 ± 8.9	93.1 ± 4.7
	MobileNet	**98.5 ± 1.9**	**95.1 ± 4.9**	**95.5 ± 3.5**

C. T2T FACE RECOGNITION

Table 4 reports the performance for the Valid and Test, for the three network models. Valid shows higher performance because the validation images come from the remaining session of the database. Test produces lower performance, due to the fact that Test is generated by the $cGAN_{T2T}$. For testing, MobileNet and Xception show similar and good performance while InceptionV3 shows the worst.

Fig. 6 illustrates the performance of different network models when accepting identities from rank 1 to 10. For testing, the CMC curve shows that Xception and MobileNet models are similar in their performance, while the InceptionV3 model performs the worst. All the models perform slightly worse at the Test than Valid sets, which indicates that our $cGAN_{T2T}$ model preserves most of the identity information. We compared our results with Athira. S *et al.*'s work [36], which utilizes local binary pattern [37], pyramid histogram of oriented gradients [38], k-nearest neighbors [39], and support vector machine [40] for thermal face recognition with different schemes, with the highest being 91.0%. Our approach shows a higher recognition rate over most of the approaches.

Table 3 reports the performance of face recognition using MobileNet model at different target temperatures in Test. Generated thermal images with target temperatures of 33.5, 34.0 and 34.5 °C have similar performance in terms of TAR and rank-1 face recognition rate. The generated thermal images with a target temperature of 33.0 °C have slightly worse performance because of the insufficient amount of low-temperature samples in the database.

Table 5 reports the performance of face recognition using MobileNet model of each subject in both the Test and Valid set. A small fraction of subjects (6, 7, 8, 9, 11, 12, 20, 39) has an obvious decline of face recognition on Test compared with Valid set. Most of the subjects have a similar performance in terms of face recognition.

We use $cGAN_{T2T}$ to convert thermal images into thermal images with a given target temperature. Our generated thermal images show high structural similarity to the ground-truth thermal images. Similarly, the face recognition performance for each subject and each target temperature on Test with Valid sets indicates that our approach preserves the structure corresponding to the target temperature with high accuracy.

TABLE 5. Face recognition performance summary on generated thermal images for each subject.

	Subject	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Test	TAR @ 1 % FAR (%)	98.8	95.4	86.7	97.9	100.0	75.0	77.5	87.1	85.8	89.6	95.0	75.0	90.8	99.2
	TAR @ 0.1 % FAR (%)	85.4	77.1	70.8	88.8	99.6	56.2	55.4	73.3	74.6	79.2	79.6	62.5	80.0	94.6
	Rank-1 (%)	94.6	91.7	83.8	95.4	99.6	75.4	72.9	83.3	83.3	88.3	88.3	72.9	86.7	96.2
Valid	TAR @ 1 % FAR (%)	100.0	100.0	88.3	93.3	100.0	90.0	98.3	95.0	100.0	96.7	98.3	96.7	96.7	100.0
	TAR @ 0.1 % FAR (%)	93.3	91.7	76.7	75.0	100.0	76.7	98.3	91.7	95.0	80.0	95.0	85.0	86.7	100.0
	Rank-1 (%)	96.7	95.0	86.7	93.3	100.0	83.3	98.3	93.3	96.7	93.3	98.3	95.0	93.3	100.0
	Subject	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Test	TAR @ 1 % FAR (%)	97.9	98.8	95.0	100.0	98.8	84.6	92.9	100.0	97.3	87.9	100.0	99.2	97.1	100.0
	TAR @ 0.1 % FAR (%)	88.8	84.2	70.8	100.0	97.9	70.0	74.6	98.8	90.2	73.8	99.2	98.8	92.1	100.0
	Rank-1 (%)	92.5	93.8	89.6	100.0	97.9	80.8	87.1	99.2	94.7	86.2	98.8	98.3	95.4	100.0
Valid	TAR @ 1 % FAR (%)	100.0	96.7	100.0	100.0	100.0	100.0	98.3	100.0	100.0	91.7	100.0	100.0	100.0	100.0
	TAR @ 0.1 % FAR (%)	98.3	90.0	98.3	100.0	98.3	88.3	85.0	100.0	100.0	78.3	100.0	100.0	96.7	100.0
	Rank-1 (%)	98.3	93.3	98.3	100.0	98.3	96.7	96.7	100.0	100.0	91.7	100.0	100.0	96.7	100.0
	Subject	29	30	31	32	33	34	35	36	37	38	39	40	41	
Test	TAR @ 1 % FAR (%)	89.6	96.2	97.9	95.0	93.8	99.2	100.0	100.0	97.9	97.9	84.2	97.5	95.0	
	TAR @ 0.1 % FAR (%)	77.9	85.0	95.0	86.7	82.5	95.4	93.3	97.9	93.3	92.1	72.1	92.1	81.7	
	Rank-1 (%)	87.9	93.8	96.2	89.2	88.8	95.0	94.6	97.5	94.6	96.7	81.7	94.2	91.7	
Valid	TAR @ 1 % FAR (%)	98.3	95.0	98.3	100.0	100.0	100.0	100.0	100.0	98.3	100.0	98.3	100.0	75.0	
	TAR @ 0.1 % FAR (%)	81.7	88.3	96.7	98.3	88.3	96.7	95.0	95.0	96.7	100.0	90.0	95.0	58.3	
	Rank-1 (%)	93.3	95.0	96.7	98.3	96.7	96.7	98.3	96.7	96.7	100.0	96.7	96.7	80.0	

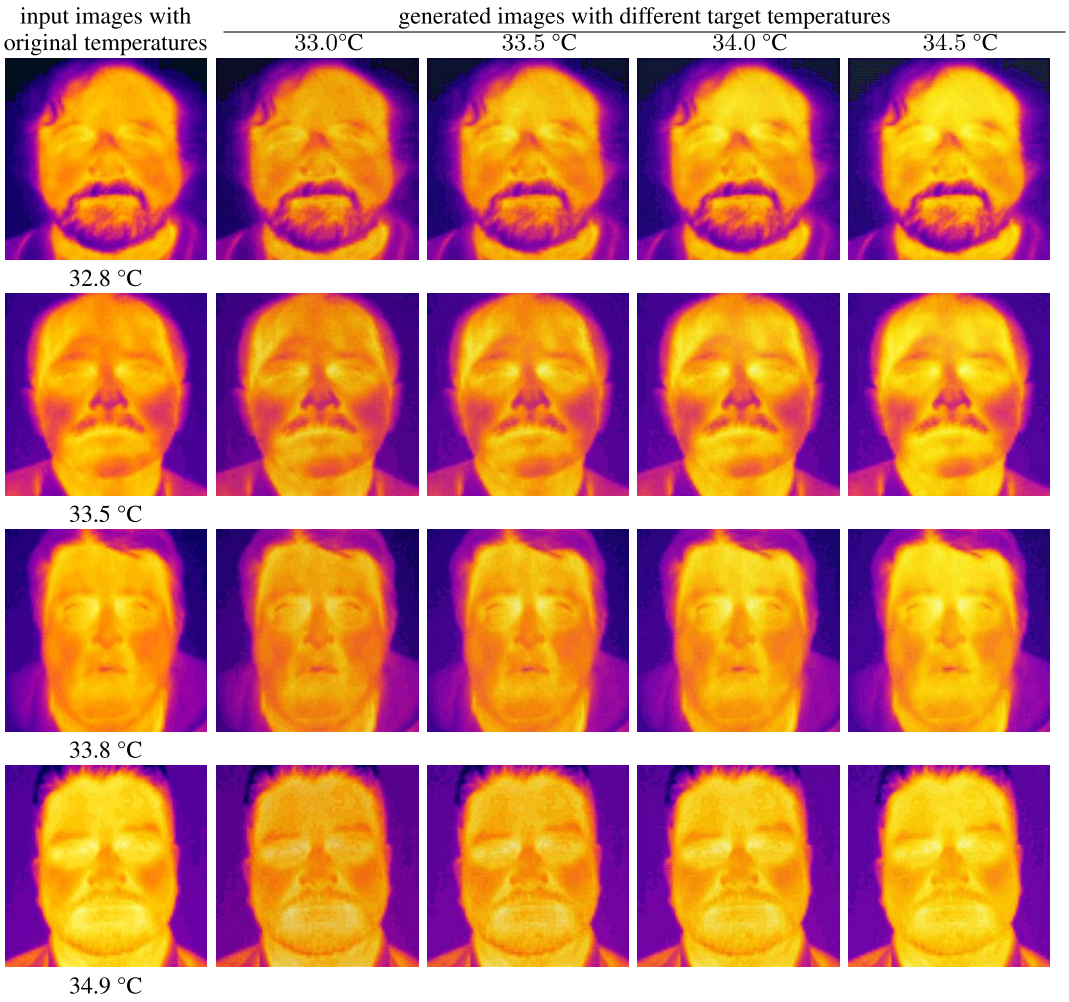


FIGURE 7. Four instances of input thermal image (left-most column) with original temperatures and the corresponding thermal images (other four columns) were generated given different target temperatures. The original temperatures of the input thermal images change from low (top) to high (bottom), and the target temperature varies from low (left) to high (right).

The rank-1 accuracy of Test has a slight decline compared to Valid: 2.8% lower for Xception, 5.0% lower for MobileNet, and 7.1% lower for InceptionV3. This means that the generated thermal images almost completely retain the identity information.

VI. CONCLUSION

In this paper, we describe an approach to the solution of two image-to-image translation tasks using cGAN: V2T and T2T. To convert visible to thermal images, we train our $cGAN_{V2T}$ using the criteria such as cGAN loss and MAE loss. The FID of the generated thermal images reached 57.3. To convert thermal to thermal images with different target temperatures, we train our $cGAN_{T2T}$ using the criteria of cGAN loss, perceptual loss, and temperature loss.

We also used face recognition techniques to evaluate the generated images. For T2T conversion, we reached a high face recognition rate, which means that the generated thermal images preserve the subjects' identity. The main outcome of this paper is proof of the feasibility of the proposed technique to generate thermal images given a target temperature, once either visible or thermal images of the subject are provided.

The proposed solutions address multiple applications from various fields. For example, in healthcare and medical sciences, one may need to model the thermal pattern distribution on subjects' faces given a target temperature. In surveillance, there might be a need to convert visible spectrum image stored in a legacy database to thermal, in order to compare with a probe image, acquired using thermal surveillance cameras, to determine the identity of a person.

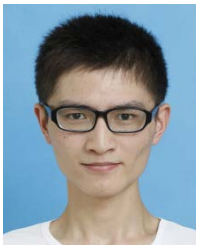
REFERENCES

[1] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*.

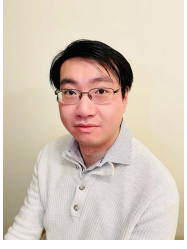
[2] V. Espinosa-Duró, M. Faundez-Zanuy, and J. Mekyska, "A new face database simultaneously acquired in visible, near-infrared and thermal spectrums," *Cognit. Comput.*, vol. 5, no. 1, pp. 119–135, Mar. 2013.

[3] M. Abdrakhmanova, A. Kuzdeuov, S. Jarju, Y. Khassanov, M. Lewis, and H. Atakan Varol, "SpeakingFaces: A large-scale multimodal dataset of voice commands with visual and thermal video streams," 2020, *arXiv:2012.02961*.

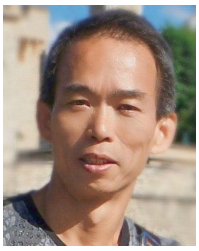
[4] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.* New York, NY, USA: Springer, 2016, pp. 694–711.
[5] X. Cao, K. Lai, S. Yanushkevich, and M. Smith, "Thermal face image generator," in *Proc. IEEE Int. Conf. Auto. Syst. (ICAS)*, Aug. 2021, pp. 1–5.
[6] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 6626–6637.
[7] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
[8] J. Gui, Z. Sun, Y. Wen, D. Tao, and J. Ye, "A review on generative adversarial networks: Algorithms, theory, and applications," *IEEE Trans. Knowl. Data Eng.*, early access, Nov. 23, 2022, doi: 10.1109/TKDE.2021.3130191.
[9] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*.
[10] S. Pei, R. Yi Da Xu, S. Xiang, and G. Meng, "Alleviating mode collapse in GAN via diversity penalty module," 2021, *arXiv:2108.02353*.
[11] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
[12] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 214–223.
[13] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of Wasserstein GANs," 2017, *arXiv:1704.00028*.
[14] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2794–2802.
[15] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1125–1134.
[16] K. Lai and S. N. Yanushkevich, "Multi-metric evaluation of thermal-to-visual face recognition," in *Proc. 8th Int. Conf. Emerg. Secur. Technol. (EST)*, Jul. 2019, pp. 1–6.
[17] L. Kezebou, V. Oludare, K. Panetta, and S. S. Agaian, "TR-GAN: Thermal to RGB face synthesis with generative adversarial network for cross-modal face recognition," in *Proc. Mobile Multimedia/Image Process., Secur., Appl.*, Apr. 2020, Art. no. 113990.
[18] Z. He, W. Zuo, M. Kan, S. Shan, and X. Chen, "AttGAN: Facial attribute editing by only changing what you want," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5464–5478, Nov. 2019.
[19] M. Li, W. Zuo, and D. Zhang, "Deep identity-aware transfer of facial attributes," 2016, *arXiv:1610.05586*.
[20] W. Shen and R. Liu, "Learning residual images for face attribute manipulation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4030–4038.
[21] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8789–8797.
[22] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier GANs," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 2642–2651.
[23] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Oct. 2017, pp. 2223–2232.
[24] X. Tang, Z. Wang, W. Luo, and S. Gao, "Face aging with identity-preserved conditional generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7939–7947.
[25] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput.-Assist. Intervent.* New York, NY, USA: Springer, 2015, pp. 234–241.
[26] X. Cao, K. Lai, S. N. Yanushkevich, and M. R. Smith, "Adversarial and adaptive tone mapping operator for high dynamic range images," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Dec. 2020, pp. 1814–1821.
[27] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
[28] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[29] Testo SE Co. KGaA. *TESTCO Software*. Accessed: Oct. 30, 2021. [Online]. Available: https://www.testo.com/en-U.S./products/thermography-irsoft
[30] A. Kantarcı and H. K. Ekenel, "Thermal to visible face recognition using deep autoencoders," in *Proc. Int. Conf. Biometrics Special Interest Group*, 2019, pp. 1–5.
[31] D. E. King, "Dlib-ml: A machine learning toolkit," *J. Mach. Learn. Res.*, vol. 10, pp. 1755–1758, Jul. 2009.
[32] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2818–2826.
[33] W. Dai, Q. Yang, G.-R. Xue, and Y. Yu, "Boosting for transfer learning," in *Proc. Int. Conf. Mach. Learn.*, Jun. 2007, pp. 193–200.
[34] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1251–1258.
[35] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.
[36] S. Athira and O. R. Murthy, "Face authentication using thermal imaging," in *Proc. Comput. Vis. Bio Inspired Comput.* New York, NY, USA: Springer, 2018, pp. 1006–1014.
[37] Z. Xie and Z. Wang, "Infrared face recognition based on personalized features selection of LBP," in *Proc. 7th Int. Conf. Intell. Hum.-Mach. Syst. Cybern.*, Aug. 2015, pp. 228–231.
[38] Y. Bai, L. Guo, L. Jin, and Q. Huang, "A novel feature extraction method using pyramid histogram of orientation gradients for smile recognition," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 3305–3308.
[39] B. Ameur, S. Masmoudi, A. G. Derbel, and A. Ben Hamida, "Fusing Gabor and LBP feature sets for KNN and SRC-based face recognition," in *Proc. 2nd Int. Conf. Adv. Technol. Signal Image Process. (ATSIP)*, Mar. 2016, pp. 453–458.
[40] J. A. K. Suykens and J. Vandewalle, "Recurrent least squares support vector machines," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 47, no. 7, pp. 1109–1114, Jul. 2000.



XINGDONG CAO received the B.Sc. degree in electrical engineering from Zhejiang University, Zhejiang, China, in 2019. He is currently pursuing the M.Sc. degree with the Biometric Technologies Laboratory, Department of Electrical and Software Engineering, University of Calgary, Calgary, AB, Canada, under the supervision of Professor Svetlana Yanushkevich. His research interest includes applying machine learning technologies to the biometrics field.



KENNETH LAI (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees from the Department of Electrical and Software Engineering, University of Calgary, Calgary, AB, Canada, in 2012, 2015, and 2022, respectively. He has published papers on using deep learning techniques for facial recognition, gesture recognition, and stress detection. His research interests include pattern recognition, machine learning, and biometrics and its application to security and health care systems.



GEE-SERN JISON HSU (Senior Member, IEEE) received the dual M.S. degree in electrical and mechanical engineering and the Ph.D. degree in mechanical engineering from the University of Michigan, Ann Arbor, in 1993 and 1995, respectively. From 1995 to 1996, he was a Postdoctoral Fellow with the University of Michigan. From 1997 to 2000, he was a Senior Research Staff with the National University of Singapore, Singapore. In 2001, he joined PenPower Technology, where he led research on face recognition and intelligent video surveillance. In 2007, he joined the Department of Mechanical Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan. His research interests include computer vision, deep learning, and particularly face recognition and license plate recognition. He received best/outstanding paper awards from ICMT 2011, CVGIP 2013, and CVPRW 2014. He and his team at PenPower Technology were recipients of the Best Innovation Award at the SecuTech Expo for three consecutive years, from 2005 to 2007. He is also a member of ACM, IAPR, and IEICE.

MICHAEL SMITH (Life Senior Member, IEEE) received the B.Sc. degree in special physics from the University of Hull, U.K., the B.Ed. degree in secondary science education, and the Ph.D. degree in experimental physics from the University of Alberta, Canada. Following a P.D.F. in physics and pathology with the University of Calgary, Canada, he spent time as a Junior and Senior High School Science and Mathematics Teacher.

He is currently a Professor Emeritus with the Department of Electrical and Software Engineering, University of Calgary, Canada. His signal and image processing research interests are varied—ranging from determination of brain visual network transfer functions, ultra-sound brain-barrier investigations, improved resolution approaches for TSAR and MRI image reconstruction and investigations of urban noise nuisance issues. On the software engineering side, he is exploring test driven development (TDD) approaches applied to biomedical engineering tools, embedded systems and FPGA development. In 2017, he was nominated for the Natural Sciences and Engineering Research Council of Canada's Science Promotion Award and received the Provincial Alberta Science Network Science Promotion Award, in 2019.

SVETLANA N. YANUSHKEVICH (Senior Member, IEEE) received the Dr.Sci. (Habilitation) degree from the Technical University of Warsaw, in 1999. She is currently a Professor with the Department of Electrical and Software Engineering (ESE), Schulich School of Engineering, University of Calgary. She directs the Biometric Technologies Laboratory, University of Calgary, the only research facility dedicated to biometric systems design in Canada. She was with the West-Pomeranian University of Technology, Szczecin, Poland, prior to joining the ESE Department, University of Calgary, in 2001. She has been contributed to the area of artificial intelligence for digital design and biometrics, since 1996. Most recently, she and her team have developed novel risk, trust and bias assessment strategies based on machine reasoning, with applications in biometric-enabled border control, forensics, and healthcare.