Thermal Face Contrastive GAN (TFC-GAN): A Framework For Visible-to-Thermal Face Translation

Catherine Ordun [1 2] **Edward Raff** [3 2] **Sanjay Purushotham** [1]

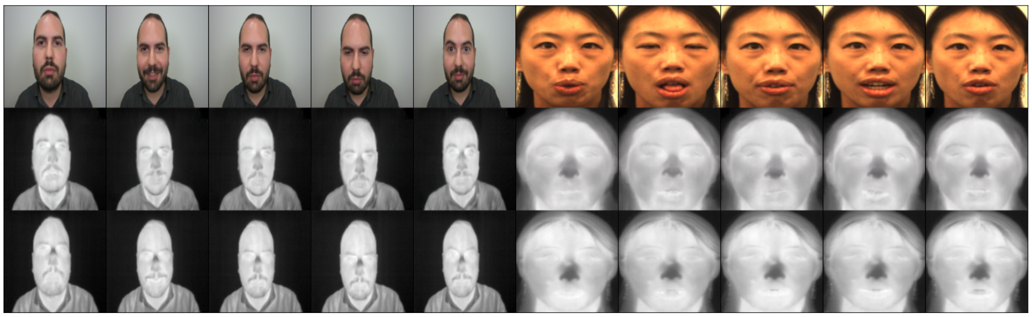

Figure 1. Generated thermal samples. Top and bottom rows are ground-truth visible and thermal, respectively. Middle row shows generated thermal faces from our approach, TFC-GAN.

Abstract

In our current work-in-progress, we explore the feasibility of generating high quality thermal faces from visible faces. This is an understudied task called Visible-to-Thermal (VT) face translation, where the majority of related works span defense and security applications for the opposite direction of Thermal-to-Visible (TV) for person re-identification. VT face translation holds promise for telemedicine whereby serving as a potential AI proxy for expensive thermal sensors, which are unavailable on smartphones and laptops. We show results from our approach called Thermal Face Contrastive GAN (TFC-GAN) that estimates a temperature vector and also learns a patch-based loss against regions of the face. Compared to seven well-known baselines used in VT face translation, and existing VT works, our method achieves a significant improvement in image quality at 41.75% and 81.15% FID improvement on the Eurecom and Devcom datasets, respectively.

[1]Dept. Information Systems, University of Maryland Baltimore County, Baltimore,MD, USA [2]Booz Allen Hamilton, Washington,DC, USA [3]Dept. Computer Science, University of Maryland Baltimore County. Correspondence to: Catherine Ordun <cordun1@umbc.edu>.

1. Introduction

The field of thermal physiology has a long history investigating thermal facial images called "thermograms" for detection of emotional states, empathy, distress, fear, cognitive stress, and correlation with gold standard vital measures like Galvanic Skin Response (Puri et al., 2005; Pavlidis et al., 2007; 2000; Merla, 2014; Ioannou et al., 2014; Rodríguez Medina et al., 2018; Ordun et al., 2020). Unlike the visible spectra, the Long-Wave Infrared (LWIR) spectra ($8um$ - 15 um) can detect radiated heat from the surface of the face due to principles of black body radiation (Bagavathiappan et al., 2009). Based on innervation of blood vessels at the surface of the facial skin due to stimulation from hormonal responses, thermograms visualize heat indicating various emotional responses that cannot be seen on visible spectra images (Pavlidis et al., 2000; Ioannou et al., 2014). However, thermal cameras are required for this task, but are not available on smartphones and computers commonly used in telemedicine sessions. As a result, we explore the feasibility of generating high quality thermograms using deep learning methods as a proxy or augmentation for thermal sensors when not available.

In particular, we introduce a paired conditional GAN (cGAN) translation framework called Thermal Face Contrastive GAN (TFC-GAN). We believe this work improves on the limited number of existing VT face translation works (Chen & Koltun, 2017; Mallat & Dugelay, 2020; Hermosilla et al., 2021; Mallat & Dugelay, 2021; Ordun et al., 2021; Pavez et al., 2022; Cao et al., 2022; Poster et al., 2021a) by

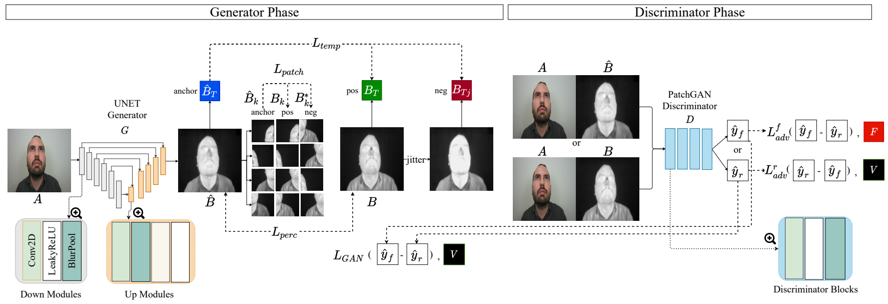

Figure 2. Overview of the proposed TFC-GAN. UNET (Ronneberger et al., 2015) Generator (G) and PatchGAN (Isola et al., 2017) Discriminator (D). G produces a fake thermal facial image, $\hat{B}$ from an input visible facial image, A: $\hat{B} = G(A; \theta^G)$. During the generator phase, four losses are calculated: triplet patch loss (L_{patch}), triplet temperature loss (L_{temp}), LPIPS perceptual similarity loss (L_{Perc}), and relativistic adversarial loss (L_{GAN}). D accepts the pair $(A, G(B))$ or (A, B), and outputs probability values across 16 x 16 patches for fakeness ($\hat{y}_f$) and realness ($\hat{y}_r$). D outputs a fake (L^f_{adv}) and real relativistic adversarial loss L^r_{adv}. Legend: A-Real Visible; B-Real Thermal; $\hat{B}$-Fake Thermal; $\hat{B}_T, B_T, B_{TJ}$ - Fake, real, jitter temperature vectors; $\hat{B}_k, B_k, B^*_k$ - Anchor, pos, neg patches; $\hat{y}_f, \hat{y}_r$ - fake and real probability values; V, F - valid and fake labels; $L_{patch}, L_{temp}, L_{perc}$ - Patch, Temperature, Perceptual Losses.

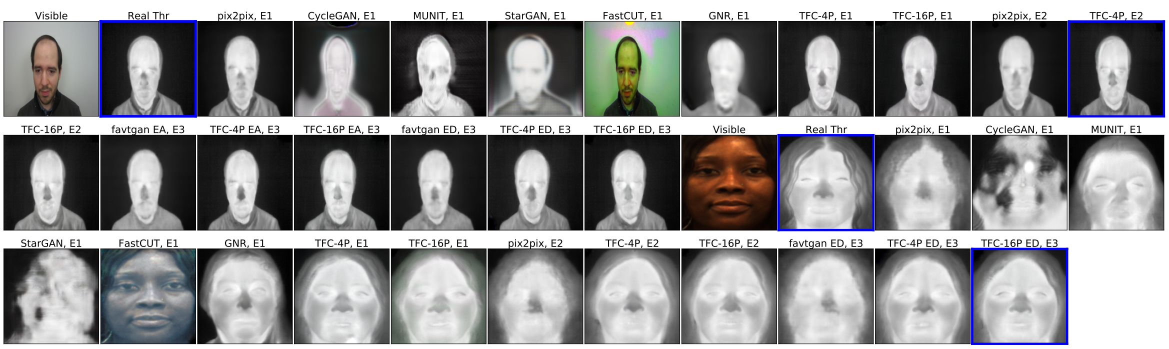

Figure 3. Samples of Eurecom and Devcom generated thermograms across all experiments. Blue frames indicate the real thermal image and generated thermal face from the highest scoring model by FID score per Tables 1 and 2.

estimating temperature vectors which are not commonly provided in datasets, as well as learning a series of contrastive losses to learn nose, eyes, mouth, and cheek regions of the face that carry important thermal physiological information. This brief report provides an overview of the key experimental settings, quantitative and qualitative results. Comprehensive material on the architecture, losses, experimental protocol, ablation studies, and additional contributions to include transfer effects from combining datasets across different thermal domains, evaluation using magnitude spectra, and extensive quantitative metrics using FID, LPIPS, SSIM, and PSNR, can be provided as a supplement.

2. Experiments

We use two paired facial datasets which are the Eurecom (Mallat et al., 2018) and the Devcom ARL Polarimetric Thermal Face Dataset (Poster et al., 2021b) datasets, and the paired FLIR Advanced Driver Assistance Systems (ADAS) dataset (FLIR, 2019). Our seven baselines are 1) pix2pix (Isola et al., 2017), 2) CycleGAN (Zhu et al., 2017), 3) StarGAN (Choi et al., 2018), 4) MUNIT (Huang et al., 2018a), 5) Fast Contrastive Unpaired Translation (FastCUT) (Park et al., 2020), 6) GansNRoses (Chong & Forsyth, 2021), and 7) favtgan (Ordun et al., 2020), designed for combination datasets. Although our datasets are paired, both paired and unpaired translation GANs are commonly used for VT translation. The ones we selected have been applied in previous studies (Anghelone et al., 2021; Bin et al.; Kniaz et al., 2018; Lu & Lu, 2021; Mallat et al., 2018; Ordun et al., 2021; Sasaki et al., 2021; Wang et al., 2020; Wang et al.; Zhang et al., 2018; Wang et al., 2019; Liu et al., 2020; Di & Patel, 2021; Li et al., 2021; Peri et al., 2021). Experiment 1 trains all baselines with a reduced Devcom dataset due to limited compute (2810 train pairs, 500 epochs). Experiment 2 trains the best performing baseline, (pix2pix) on the entire Devcom dataset (56,206 train pairs, 1000 epochs). Experiment 3 combines datasets from different domains, such as face-and-face or face-and-cityscape to understand transfer effects of similar thermal features. All experiments use the full Eurecom dataset (945 train pairs).

Table 1. Quantitative Metrics for Eurecom Generated Thermal Faces. **Bold** indicate best scores per experiment; **Bold*** indicate best scores overall. F-LPIPS (Face LPIPS values), P-LPIPS (Patch LPIPS values). For all experiments, we train the TFC-GAN model using four ("TFC-GAN-4P") and 16 ("TFC-GAN-16P") patches.

Dataset: Eurecom			
Experiment 1	**FID↓**	**F-LPIPS↓**	**P-LPIPS↓**
pix2pix (Isola et al., 2017)	63.694	0.149	0.127
CycleGAN (Zhu et al., 2017)	339.463	0.450	0.389
StarGAN (Choi et al., 2018)	290.331	0.517	0.455
MUNIT (Huang et al., 2018b)	220.863	0.294	0.242
FastCUT (Park et al., 2020)	204.355	0.758	0.736
GansNRoses (Chong & Forsyth, 2021)	221.877	0.431	0.374
TFC-GAN-4P	64.049	**0.117**	**0.095**
TFC-GAN-16P	**62.745**	0.124	0.100
Experiment 2	**FID↓**	**F-LPIPS↓**	**P-LPIPS↓**
pix2pix (Isola et al., 2017)	106.180	0.198	0.169
TFC-GAN-4P	**61.850***	0.126	0.103
TFC-GAN-16P	63.601	**0.121**	**0.096**
Dataset: Eurecom + ADAS			
Experiment 3	**FID↓**	**F-LPIPS↓**	**P-LPIPS↓**
favtgan (Ordun et al., 2020)	124.608	0.232	0.194
TFC-GAN-4P	**62.622**	**0.117***	**0.094***
TFC-GAN-16P	67.734	0.120	0.096

Table 2. Quantitative Metrics for Devcom Generated Thermal Faces. Same legend as Table 1.

Dataset: Devcom			
Experiment 1	**FID↓**	**F-LPIPS↓**	**P-LPIPS↓**
pix2pix (Isola et al., 2017)	198.105	**0.185***	**0.138***
CycleGAN (Zhu et al., 2017)	308.121	0.483	0.418
StarGAN (Choi et al., 2018)	261.635	0.397	0.336
MUNIT (Huang et al., 2018b)	87.726	0.361	0.309
FastCUT (Park et al., 2020)	130.818	0.603	0.543
GansNRoses (Chong & Forsyth, 2021)	86.270	0.397	0.348
TFC-GAN-4P	**48.820**	0.269	0.229
TFC-GAN-16P	55.901	0.276	0.237
Experiment 2	**FID↓**	**F-LPIPS↓**	**P-LPIPS↓**
pix2pix (Isola et al., 2017)	154.114	0.192	0.146
TFC-GAN-4P	**38.269**	0.266	0.224
TFC-GAN-16P	39.053	**0.262**	**0.224**
Dataset: Eurecom + Devcom			
Experiment 3	**FID↓**	**F-LPIPS↓**	**P-LPIPS↓**
favtgan (Ordun et al., 2021)	268.000	0.373	0.293
TFC-GAN-4P	36.406	**0.264**	**0.225**
TFC-GAN-16P	**34.439***	0.266	0.227

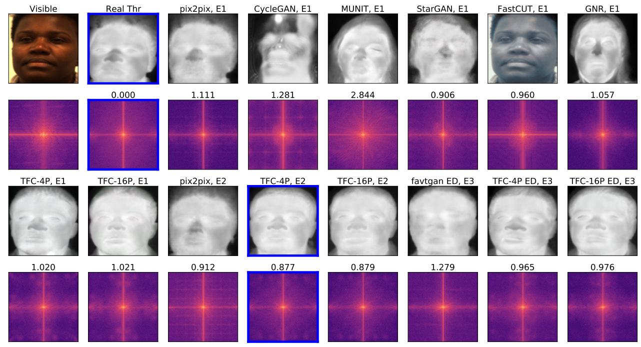

Figure 4. Sample magnitude spectra for Devcom. By decomposing the Fourier Transform into its magnitude component, image frequencies can be visualized (Mitra & Kuo, 2006). Our model "TFC-GAN-4P,E2" achieves the lowest MSE=0.877 which offers a different viewpoint compared to our best performing model "TFC-GAN-16P, ED" by FID score which scored MSE=0.976. Notice how the FastCUT spectrum resembles the real visible spectrum, and how the distorted results of CycleGAN, MUNIT, and StarGAN display geometric patterns inconsistent with the real thermal spectrum. Blue frames indicate the real thermal image and experiment with the lowest MSE.

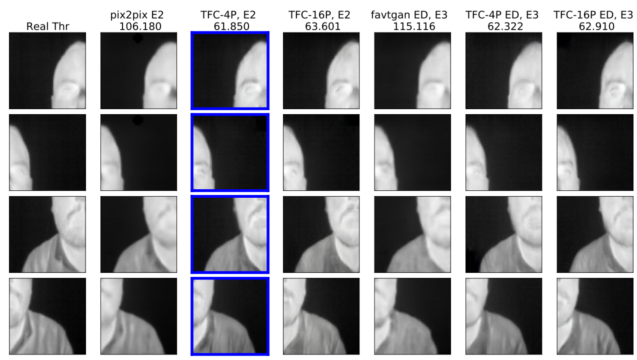

Figure 5. Patch samples from a Eurecom generated thermogram. Blue indicates the best scoring model by FID ("TFC-GAN-4P,E2"). FID scores shown.

3. Results

Across all experiments, TFC-GAN achieves the lowest Fréchet Inception Distance (FID) (Heusel et al., 2017) scores as shown in Table 1 and Table 2. For Eurecom, the "TFC-GAN-4P" in Experiment 2 trained only on the Eurecom datasets (FID=61.850) shows the best overall FID score. This is a 41.75% reduction in FID score compared to the best baseline, "pix2pix" (106.180). Likewise, the best FID score for Devcom results from "TFC-GAN-16P" trained on the combination Eurecom + Devcom dataset (FID=34.439). This is a 81.75% reduction in FID compared to the "favtgan ED" (268.000). When comparing our results to existing facial VT works, our FID scores show improvement over (Pavez et al., 2022) (reported FID 65.71) and (Cao et al., 2022) (reported FID 57.3). We show samples of generated thermograms across all experiments in Figure 3 and a Eurecom patch sample in Figure 5. The Eurecom "TFC-4P,E2" articulates the eyes and mouth, while maintaining overall temperature patterns consistent with the real thermal image. Although the pix2pix and favtgan baselines maintain similar temperature locations such as the nose, they are blurry with less detail. Distortion and/or loss of subject identity can be seen with "FastCUT,E1", "StarGAN,E1", "CycleGAN,E1", "MUNIT,E1", and "GNR,E1". Not shown here, we tried Pix2Pix-HD (Wang et al., 2018) and U-GAT-IT (Kim et al., 2019), but they did not make it into the baselines due to poor image quality and loss of subject identity. We also provide an alternative assessment by calculating Mean Squared Error between magnitude spectra from Fast Fourier Transform (FFT), of generated and real thermal samples shown in Figure 4.

References

Anghelone, D., Chen, C., Faure, P., Ross, A., and Dantcheva, A. Explainable thermal to visible face recognition using latent-guided generative adversarial network. In *IEEE International Conference on Automatic Face and Gesture Recognition*, 2021.

Bagavathiappan, S., Saravanan, T., Philip, J., Jayakumar, T., Raj, B., Karunanithi, R., Panicker, T., Korath, M. P., and Jagadeesan, K. Infrared thermal imaging for detection of peripheral vascular disorders. *Journal of medical physics/Association of Medical Physicists of India*, 34(1):43, 2009.

Bin, J., Zhang, R., Zhang, C., Liu, X., Du, S., Blasch, E., and Liu, Z. Improved object detection in thermal imaging through context enhancement and information fusion: A case study in autonomous driving.

Cao, X., Lai, K., Hsu, G.-S. J., Smith, M., and Yanushkevich, S. Cross-spectrum thermal face pattern generator. *IEEE Access*, 2022.

Chen, Q. and Koltun, V. Photographic image synthesis with cascaded refinement networks. In *CVPR*, 2017.

Choi, Y. et al. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *CVPR*, 2018.

Chong, M. J. and Forsyth, D. Gans n'roses: Stable, controllable, diverse image to image translation (works for videos too!). *arXiv*, 2021.

Di, X. and Patel, V. M. Multimodal face synthesis from visual attributes. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2021.

FLIR. Thermal dataset for algorithm training, 2019. URL https://www.flir.com/oem/adas/adas-dataset-form/.

Hermosilla, G., Tapia, D.-I. H., Allende-Cid, H., Castro, G. F., and Vera, E. Thermal face generation using stylegan. *IEEE Access*, 2021.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *NeurIPS*, 2017.

Huang, X., Liu, M.-Y., Belongie, S., and Kautz, J. Multimodal unsupervised image-to-image translation. In *ECCV*, 2018a.

Huang, X., Liu, M.-Y., Belongie, S., and Kautz, J. Multimodal unsupervised image-to-image translation. In *ECCV*, 2018b.

Ioannou, S., Gallese, V., and Merla, A. Thermal infrared imaging in psychophysiology: potentialities and limits. *Psychophysiology*, 2014.

Isola, P. et al. Image-to-image translation with conditional adversarial networks. In *CVPR*, 2017.

Kim, J., Kim, M., Kang, H., and Lee, K. U-gat-it: Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation. *arXiv preprint arXiv:1907.10830*, 2019.

Kniaz, V. V. et al. ThermalGAN: Multimodal color-to-thermal image translation for person re-identification in multispectral dataset. In *ECCV*, 2018.

Li, S., Han, B., Yu, Z., Liu, C. H., Chen, K., and Wang, S. I2v-gan: Unpaired infrared-to-visible video translation. In *ACM International Conference on Multimedia*, 2021.

Liu, S., Gao, M., John, V., Liu, Z., and Blasch, E. Deep learning thermal image translation for night vision perception. *ACM Transactions on Intelligent Systems and Technology*, 2020.

Lu, Y. and Lu, G. Bridging the invisible and visible world: Translation between rgb and ir images through contour cycle gan. In *IEEE International Conference on Advanced Video and Signal Based Surveillance*, 2021.

Mallat, K. and Dugelay, J.-L. Facial landmark detection on thermal data via fully annotated visible-to-thermal data synthesis. In *2020 IEEE International Joint Conference on Biometrics (IJCB)*, 2020.

Mallat, K. and Dugelay, J.-L. Indirect synthetic attack on thermal face biometric systems via visible-to-thermal spectrum conversion. In *CVPR*, 2021.

Mallat, K. et al. A benchmark database of visible and thermal paired face images across multiple variations. In *BIOSIG*, 2018.

Merla, A. Revealing psychophysiology and emotions through thermal infrared imaging. In *PhyCS*, 2014.

Mitra, S. K. and Kuo, Y. *Digital signal processing: a computer-based approach*. 2006.

Ordun, C., Raff, E., and Purushotham, S. Generating thermal human faces for physiological assessment using thermal sensor auxiliary labels. *arXiv*, 2021.

Ordun, C. et al. The use of AI for thermal emotion recognition: A review of problems and limitations in standard design and data. *AAAI*, 2020.

Park, T., Efros, A. A., Zhang, R., and Zhu, J.-Y. Contrastive learning for unpaired image-to-image translation. In *ECCV*, 2020.

Pavez, V., Hermosilla, G., Pizarro, F., Fingerhuth, S., and Yunge, D. Thermal image generation for robust face recognition. *Applied Sciences*, 2022.

Pavlidis, I., Dowdall, J., Sun, N., Puri, C., Fei, J., and Garbey, M. Interacting with human physiology. *Computer Vision and Image Understanding*, 2007.

Pavlidis, I. et al. The imaging issue in an automatic face/disguise detection system. In *IEEE Computer Vision Beyond the Visible Spectrum*, 2000.

Peri, N., Gleason, J., Castillo, C. D., Bourlai, T., Patel, V. M., and Chellappa, R. A synthesis-based approach for thermal-to-visible face verification. In *International Conference on Automatic Face and Gesture Recognition*, 2021.

Poster, D., Thielke, M., Nguyen, R., Rajaraman, S., Di, X., Fondje, C. N., Patel, V. M., Short, N. J., Riggan, B. S., Nasrabadi, N. M., et al. A large-scale, time-synchronized visible and thermal face dataset. In *WACV*, 2021a.

Poster, D. D., Hu, S., Short, N. J., Riggan, B. S., and Nasrabadi, N. M. Visible-to-thermal transfer learning for facial landmark detection. *IEEE Access*, 2021b.

Puri, C. et al. Stresscam: non-contact measurement of users' emotional states through thermal imaging. In *CHI'05*, 2005.

Rodríguez Medina, D. A., Domínguez Trejo, B., Cortés Esteban, P., Cruz Albarrán, I. A., Morales Hernández, L. A., and Leija Alva, G. Biopsychosocial assessment of pain with thermal imaging of emotional facial expression in breast cancer survivors. *Medicines*, 2018.

Ronneberger, O. et al. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, 2015.

Sasaki, H., Willcocks, C. G., and Breckon, T. P. Unit-ddpm: Unpaired image translation with denoising diffusion probabilistic models. *arXiv*, 2021.

Wang, G.-A., Zhang, T., Yang, Y., Cheng, J., Chang, J., Liang, X., and Hou, Z.-G. Cross-modality paired-images generation for rgb-infrared person re-identification. In *AAAI*, 2020.

Wang, T., Zhang, T., Liu, L., Wiliem, A., and Lovell, B. Cannygan: Edge-preserving image translation with disentangled.

Wang, T., Zhang, T., Liu, L., Wiliem, A., and Lovell, B. Cannygan: Edge-preserving image translation with disentangled features. In *ICIP*, 2019.

Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J., and Catanzaro, B. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8798–8807, 2018.

Zhang, T. et al. TV-gan: Generative adversarial network based thermal to visible face recognition. In *ICB*, 2018.

Zhu, J.-Y. et al. Toward multimodal image-to-image translation. In *NeurIPS*, 2017.