

Fourier-Based Strategies to Improve Ethnic Feature Generation during Visible-to-Thermal Facial Translation (A work in progress)

Catherine Ordun, Edward Raff, Sanjay Purushotham

Overview. The task of *visible* $\rightarrow$ *thermal* (VT) facial translation is an under-studied task where given a visible facial image, its thermal pair must be generated. The application is relevant for telemedicine, healthcare, and other fields where thermal physiology is important to detect signs of inflammation and stress. In our previous research such as development of the favtGAN [1] and AI risk assessment in thermal Facial Emotion Recognition [2], we find that most VT/TV face translation works focus largely on image quality. However, the loss of subject identity is an under-reported challenge that ties closely with ethnicity and race. In our current work that improves upon favtGAN called Thermal Face-Contrastive GAN (TFC-GAN), we observe that generating thermal face quality is straight-forward, but translating ethnic features of hair and eyes that largely comprise the subject’s identity is difficult. We show initial findings of a work currently in progress.

Motivation. This is an outstanding problem in existing VT/TV face datasets since the majority of subjects are Caucasian (such as the Eurecom [3] and Carl [4] datasets) as inventoried by [2]. Hence, racial identity under thermal spectra has not been fully investigated in image translation. One recent dataset called the Devcom Army Research Lab (ARL) dataset [5] provides a more diverse set of subjects. We attempted to manually label each of the 188 training subjects and 46 test subjects in the Devcom dataset, and found that only 25% of the training subjects are non-white (Asian, Black, other Non-White), whereas 75% are White.

Experiments. As a result, we use the Devcom dataset to explore strategies to improve the translation of visible-to-thermal faces for non-Caucasian features. We first develop the TFC-GAN to produce high quality thermal faces by adding anti-aliasing layers [6] to the UNET [7] generator and PatchGAN [8] discriminator in addition to spectral normalization [9]. Further, we compose our objective through a triplet loss that controls regional face patches of the thermal image, a triplet loss for temperature distribution, an LPIPS [10] perceptual loss, and a relativistic adversarial loss [11]. We find that calculating the Fourier amplitude and phase of different patches of the face, and learning an L1 loss per [12], improves the quality of the generated thermal face. An example of magnitude spectra from Fast Fourier Transform (FFT) of real and generated thermal face patches is shown in Figure 2, demonstrating the frequency contrast between the subject and its falsely generated identity.

We explore four strategies to generate more accurate representations of ethnic subject identity: 1) TFC-GAN without Fourier Loss (“FFT”), 2) TFC-GAN using KL-Divergence across Fourier signals of patches, hair, and eyes, 3) TFC-GAN using a mix of KL-Divergence and L1 Loss for Fourier

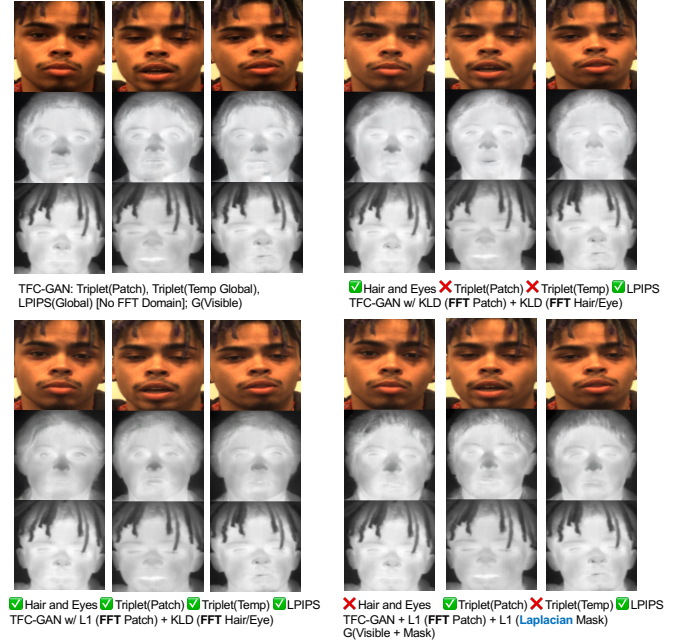


Fig. 1: Samples of generated thermal faces based on four sets of experiments. Top row - real visible, Middle row - generated thermal, Bottom row - real thermal

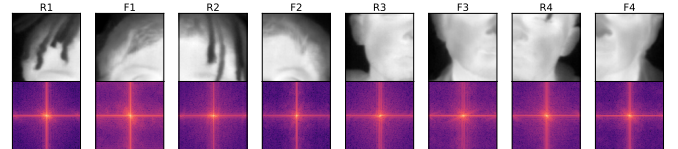


Fig. 2: Example of Real Thermal (“R”) and Fake Thermal (“F”) magnitude spectra across regions of the face.

signals of patches, hair, and eyes, and 4) TFC-GAN with a generator that accepts the visible image and its Laplacian mask, learning the L1 losses of Fourier signals of patches, but no hair and eyes. Our rationale is that since the Fourier domain consists of amplitude and phase, whereby they represent different distributions of signal frequencies, KL-Divergence might be applicable to these distributions to reconstruct high frequency hair features. The Laplacian masks are used to extract high frequency edges which we observe are more common in ethnic hair features such as dreadlocks as shown in Figure 1. We show qualitative results in Figure 1. While all four strategies struggle to translate the identity of the subject, the “Masked” experiment at the bottom right appears to show greater improvement for expression and hair frequencies and provides a promising direction for future study.

REFERENCES

- [1] C. Ordun, E. Raff, and S. Purushotham, “Generating thermal human faces for physiological assessment using thermal sensor auxiliary labels,” *arXiv*, 2021.
- [2] C. Ordun *et al.*, “The use of AI for thermal emotion recognition: A review of problems and limitations in standard design and data,” *AAAI*, 2020.
- [3] K. Mallat *et al.*, “A benchmark database of visible and thermal paired face images across multiple variations,” in *BIOSIG*, 2018.
- [4] V. Espinosa-Duró, M. Faundez-Zanuy, and J. Mekyska, “A new face database simultaneously acquired in visible, near-infrared and thermal spectrums,” *Cognitive Computation*, 2013.
- [5] D. Poster, M. Thielke, R. Nguyen, S. Rajaraman, X. Di, C. N. Fondje, V. M. Patel, N. J. Short, B. S. Riggan, N. M. Nasrabadi *et al.*, “A large-scale, time-synchronized visible and thermal face dataset,” in *WACV*, 2021.
- [6] R. Zhang, “Making convolutional networks shift-invariant again,” in *ICML*, 2019.
- [7] O. Ronneberger *et al.*, “U-net: Convolutional networks for biomedical image segmentation,” in *MICCAI*, 2015.
- [8] P. Isola *et al.*, “Image-to-image translation with conditional adversarial networks,” in *CVPR*, 2017.
- [9] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, “Spectral normalization for generative adversarial networks,” *arXiv*, 2018.
- [10] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, “The unreasonable effectiveness of deep features as a perceptual metric,” in *CVPR*, 2018.
- [11] A. Jolicœur-Martineau, “The relativistic discriminator: a key element missing from standard gan,” *arXiv*, 2018.
- [12] D. Fuoli, L. Van Gool, and R. Timofte, “Fourier space losses for efficient perceptual image super-resolution,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 2360–2369.