Eksplorium *p*-ISSN 0854-1418
Volume 46 No. 2, June 2025: 52–69 *e*-ISSN 2503-426X

A Deep Learning Approach to Generate Thermal Images Synthetically for Emotion Recognition Applications

[1]Geethanjali B N, [2]Dr. C M Naveen Kumar, [3]Dr.Ananda Babu.J, [4]Krishna Swaroop A, [5]Geethanjali B N, [6]Naveen K R

[1]Student-Mtech in AI&DS

Department of Information Science & Engineering

Malnad College of Engineering, Hassan-573202, Karnataka, India

geethnaveen8@gmail.com

[2]Associate Professor

Department of Computer Science & Business Systems

Malnad College of Engineering, Hassan-573202, Karnataka, India

cmn@mcehassan.ac.in

[3]Professor and Head

Department of Information Science & Engineering

Malnad College of Engineering, Hassan-573202, Karnataka, India

abj@mcehassan.ac.in

[4]Assistant Professor

Department of Information Science & Engineering

Malnad College of Engineering, Hassan-573202, Karnataka, India

ksa@mcehassan.ac.in

[5]Lecturer - Computer Science & Engineering

Government Polytechnic Chitradurga

Department of Collegiate and Technical Education

geethnaveen8@gmail.com

[6]Lecturer – Electrical & Electronics Engineering

Government Polytechnic Belur

Department of Collegiate and Technical Education

naveenkr.dvg@gmail.com

Article Received: 04 May 2025, Revised: 08 June 2025, Accepted: 18 June 2025

Abstract: Thermal imaging is widely used in emotion recognition for its ability to capture physiological responses such as temperature variations in facial regions. However, collecting large-scale thermal datasets is challenging due to cost, privacy concerns, and limited availability. This presents a deep learning approach to synthetically generate thermal images from RGB images for emotion recognition applications. A Generative Adversarial Network (GAN)-based model is trained to make it learn the mapping between visible and thermal domains, ensuring realistic thermal image synthesis. The proposed method enhances existing emotion recognition systems by providing augmented thermal datasets, reducing dependency on expensive thermal cameras. Experimental results shows that using synthetic thermal images enhances the accuracy of deep learning-based emotion recognition models.

Keywords: Deep Learning**,** Generative Adversarial Networks (GANs)**,** Pix2Pix**,** Image-to-Image Translation**,** Synthetic Thermal Image Generation**,** Conditional GAN (cGAN)**,** Thermal Image Synthesis**,** Image Preprocessing**,** Multispectral Imaging.

1. INTRODUCTION

Emotion recognition is playing a vital role in various applications, including human-computer interaction, healthcare, and security. Traditional methods rely on facial expressions captured using RGB cameras, but these approaches often face challenges due to variations in lighting, occlusions, and expression intensity. Thermal imaging provides a more reliable alternative by detecting heat patterns on the face, which are less affected by lighting conditions and can capture physiological changes linked to emotional states.

However, collecting large-scale thermal image datasets is expensive and limited by privacy concerns and hardware constraints. To overcome this challenge, we propose a deep learning-based approach to synthetically generate thermal images from RGB images using GANs. Our model learns the mapping between visible and thermal domains, enabling the creation of high-quality synthetic thermal images.

By generating synthetic thermal images, we enhance existing emotion recognition models by augmenting thermal datasets, reducing dependence on costly thermal cameras, and improving the robustness of emotion classification. The effectiveness of the proposed approach in enhancing deep learning-based emotion recognition systems is validated using benchmark datasets.

2. RELATED WORK

Pavez, Vicente, et al. (2023) presented a deep learning approach combining Stable Diffusion and Vision Transformer (ViT) for high-quality synthetic thermal image generation[1]. A novel prompt design system enhances image generation control. The synthetic thermal database developed in this study achieved 98% accuracy in facial detection, demonstrating results comparable to thermal images captured.

Farooq , et al (2021) explored multiple techniques for generating synthetic thermal images, including data augmentation, StyleGAN-based image synthesis, and 2D to 3D image reconstruction. The use of Wide ResNet CNN for gender classification achieved a 4.6% and 4.4% improvement in training and validation accuracy, respectively, with an overall test accuracy of 83.33%.[2]

Imtiaz , et al (2021) introduced a Lightweight Pyramid Network (LPNet) for image-to-image synthesis, offering a computationally efficient alternative to deep CNN-based methods. By leveraging Laplacian-Gaussian image pyramid decomposition, the study achieves enhanced heat signature retention and improved structural integrity of synthesized thermal images[3]. The model's performance is validated using SSIM, PSNR, and UQI analysis.

Pavez, Vicente, et al. (2022) developed a GAN-based thermal image synthesis framework incorporating StyleCLIP and GANsN'Roses models[4]. The generated database contains over 100k synthetic thermal face images, significantly improving the accuracy of the Thermal-FaceNet model to 99.98%. When tested with real thermal images, the model maintained an accuracy above 98%.

Mei, et al (2022) addressed the challenge of atmospheric turbulence in thermal imaging using a StyleGAN2-based image reconstruction method. The proposed approach effectively transforms thermal images into visible-spectrum images while preserving high face verification accuracy[5]. It is one of the first studies to explore thermal-to-visible image translation under turbulent conditions.

Cao, Xingdong, et al. (2022) proposed a conditional GAN (cGAN) model to convert visible face images to thermal images (V2T) and adapt thermal images to different temperature settings (T2T). The system achieves a low Fréchet Inception Distance (FID) of 57.3 for V2T conversion and a rank-1 face recognition rate of 91.0% for T2T transformations, proving the effectiveness of the method.[6]

Hermosilla, Gabriel, et al. (2021) employed StyleGAN2 and its adaptive discriminator augmentation (ADA) variant to generate high-quality synthetic thermal face images[7]. A synthetic thermal face database is created and validated using six deep learning models for face recognition, achieving an accuracy of 99.98%. The research also explores latent space manipulations to introduce variations such as facial accessories and orientations.

Panchard, Danick, et al. (2021) discussed the potential and challenges of using GANs for synthesizing thermal images from visible light images. While synthetic images show high visual quality, contrast independence between thermal and visible images remains a limitation. The study highlights applications in dataset augmentation and multimodal image analysis.[8]

Hagel, et al (2020) reviewed recommender systems used contextual information, including social relationships, user interactions, and time-based data[9]. It classifies social relations into three categories: trust, friend activities, and user interactions. The study finds that collaborative filtering methods yield the best recommendation accuracy in mobile applications.

Kniaz , et al (2017) proposed method of image transformation is based on the use of a CNN for semantic image segmentation. A great number of CNN architectures were developed for image classification. Semantic image segmentation requires significant changes in CNN architectures. Such architectures are commonly known as 'fully convolutional' networks with no fullyconnected layers In addition, the deconvolution layers are widely used to solve the problem of semantic segmentation.[10]

Farooq , et al (2020) mainly focused on the proposed algorithm for generating 3D synthetic face data from a single 2d thermal image. they have utilized the tufts thermal face dataset. since this dataset has published recently with data samples from 6 different image modalities which include visible, near infrared, thermal, computerized sketch, a recorded video, and 3D images. The dataset consists of images of both males and females genders.[12]

3. PROPOSED APPROACH

Thermal imaging-based emotion recognition offers notable advantages over traditional RGB methods, including resilience to lighting changes and the capability to detect physiological signals. However, the adoption of thermal-based emotion recognition systems is hindered by the limited availability of large-scale thermal image datasets. Collecting real thermal data is expensive, requires specialized equipment, and raises privacy concerns, making it difficult to train deep learning models effectively.

Existing approaches either rely on small thermal datasets or attempt to adapt RGB images for emotion recognition, but they fail to leverage the physiological benefits of thermal imaging. There is a critical need for a method that can generate high-quality synthetic thermal images from RGB inputs, enabling large-scale dataset creation and improving emotion recognition models without relying on costly thermal cameras.

The research addresses this challenge by developing a deep learning approach, using GANs , to synthesize realistic thermal images from visible-light images. This solution aims to bridge the gap between RGB and thermal domains, enhance dataset diversity, and improve the performance of deep learning models in thermal-based emotion recognition applications.

- Utilize deep learning approaches like GAN Networks – Pix 2 Pix architecture.
  - Pix2Pix is a cGAN designed for supervised image-to-image translation.
  - It works well when it is paired with RGB and thermal images for training.
- Develop a dataset of synthetic thermal images to train emotion recognition models.
- Validate synthetic images using real-world datasets.

3.1 GENERATIVE ADVERSARIAL NETWORK

A Generative Adversarial Network(GAN) consists of two neural networks, Generator(G) Creates synthetic images that mimic real thermal images. Discriminator(D) Evaluates whether an image is real or synthetic. Both networks train together in a competitive manner, perfecting the quality of generated thermal images over time. Using GAN for synthetic image generation leads to better data addition, sphere adaption, Literalism and Diversity.

3.2 NETWORK ARCHITECTURE

The project involves synthetically generating thermal images**,** which a GAN-based architecture for image-to-image translation is used. One of the best models for this is Pix2Pix**,** a type of cGAN which works very good for image to image translation, uses paired datasets (visible + thermal). Conditional Generative Adversarial Networks are an extension of GANs where both the generator and discriminator receive extra information to guide the generation process. It is designed for supervised image-to-image translation**.** It works good when you have paired RGB and thermal images for training.

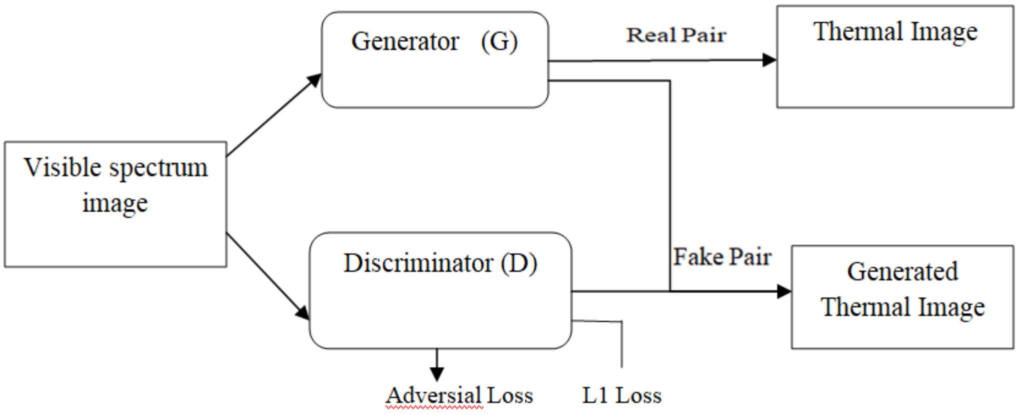


Figure 3.2.1 : Network Architecture illustrates the working of a Pix2Pix Conditional GAN (cGAN) for generating thermal images from visible light images.

The diagram illustrates the working of a Pix2Pix Conditional GAN (cGAN) for generating thermal images from visible light images. A visible spectrum image is fed into the Generator **(G)**, which produces a Generated Thermal Image**.** The Discriminator (D) then evaluates this generated image against a Real Thermal Image, forming either a real or fake image pair. The network is trained using Adversarial Loss (to fool the discriminator) and L1 Loss (to minimize the difference between real and generated thermal images).

3.2.1 Pix2Pix GAN

Pix2Pix Architecture is best for Paired Image Translation and is a cGAN designed for supervised image-to-image translation. It works good when you have paired RGB and thermal images for training.

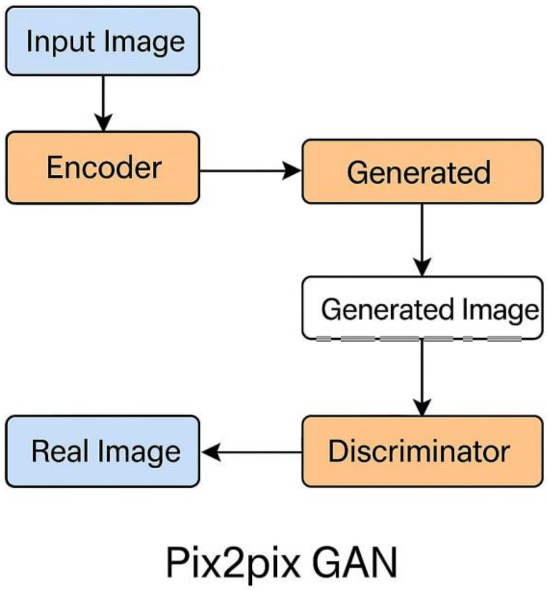


Figure 3.2.1: Pix2Pix GAN architecture

Eksplorium *p*-ISSN 0854-1418
Volume 46 No. 2, June 2025: 52–69 *e*-ISSN 2503-426X

The diagram illustrates the architecture of a Pix2Pix Generative Adversarial Network (GAN), a deep learning model designed for image-to-image translation tasks. In this framework, an input image—such as a visible-light image—is first passed through an encoder, which extracts relevant features and encodes the information into a compact representation. This encoded data is then used by the generator to produce a corresponding output image, referred to as the generated image. Both the generated image and the real thermal image are then fed into the discriminator. The discriminator's role is to differentiate between the real and generated images, assigning high confidence to authentic images and low confidence to synthetic ones. Meanwhile, the generator is trained to produce images that are as realistic as possible, aiming to fool the discriminator. This adversarial training loop helps the model learn to generate high-quality, realistic images that closely resemble the target domain. Pix2Pix GAN is particularly useful in applications like converting visible images to thermal images, making it well-suited for tasks such as synthetic thermal image generation for emotion recognition or surveillance.

Pix2Pix works with the following flow

3.2.1.1Generator (U-Net Architecture) :

U-Net is a convolutional neural network (CNN) architecture that was initially developed for segmenting biomedical images. It features an encoder-decoder architecture with skip connections that aid in retaining spatial information. In the project it Converts an RGB image into a synthetic thermal image and uses an encoder-decoder (U-Net) structure for preserving spatial features. Since the task involves generating images thermally from visible images**,** U-Net's ability to capture both global context and fine-grained details is beneficial for the project work. It also ensures that the generated thermal images retain structural consistency with the input visible images while maintaining realistic thermal textures.

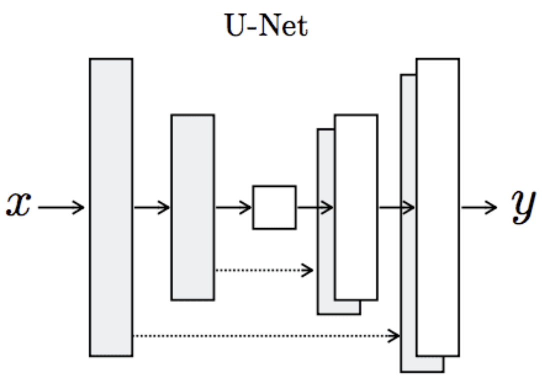


Figure 3.2.1.1 : U-Net is an encoder-decoder architecture that incorporates skip connections between corresponding layers in the encoder and decoder, allowing for efficient feature fusion.[11]

3.2.1.2 Discriminator (PatchGAN Architecture) : PatchGAN is a type of discriminator used in(GANs, in pix2pix models, which are well-suited for image-to-image translation tasks. For the project, where deep learning is used to generate synthetic thermal images, PatchGAN is an effective discriminator used to distinguish between real and synthetically generated thermal images.Instead of checking the whole image, PatchGAN focuses on local patches and helps ensure that small details in the thermal image look real. By integrating PatchGAN into GAN-based model, system will produce high-quality, synthetic thermal images that closely mimic real thermal images, thereby improving the performance of subsequent tasks like face detection, gender classification, and SOS gesture detection.

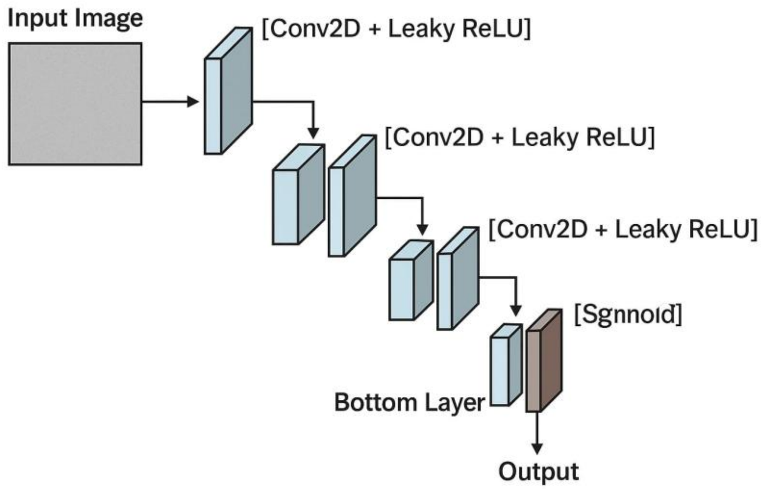


Figure 3.2.1.2 : Discriminator (PatchGAN Architecture)

The diagram depicts the architecture of a convolutional neural network (CNN), commonly used in the discriminator component of a GAN, particularly in a Pix2Pix model. The network begins with an input image, which is successively processed through multiple convolutional layers. Each layer performs a 2D convolution operation followed by a Leaky ReLU activation function, which helps in handling the problem of dying neurons by allowing a small gradient when the unit is not active. These layers progressively extract higher-level features from the image. As the network deepens, it reaches a bottom layer, which condenses the spatial information into a more abstract representation. Finally, a Sigmoid activation function is applied at the output layer, producing a value between 0 and 1. This output indicates the probability that the input image is real or generated (fake), making this architecture ideal for binary classification tasks such as those performed by GAN discriminators.

3.2.1.3 Loss Function: Adversarial Loss (GAN Loss) – to train the generator and discriminator. Since PatchGAN is used , this loss helps ensure the synthetic thermal images are as realistic as possible. Helps generator creates realistic images. L1 Loss (Pixel-Level Difference) is used to minimize pixel-wise differences between real and generated thermal images. Helps maintain structural similarity between input and output images and also ensures the generated thermal image matches the real one.

3.3 OBJECTIVE FUNCTION

3.3.1 Standard GAN Objective Function

The standard min-max objective function of a GAN is:

$$\min_{G}\max_{D} E_{x} \sim P_{data} [\log D(x)] + E_{z} \sim P_{z} [\log(1-D(G(z)))] \quad \text{--- 1}$$

Where:

- x is a real thermal image from the dataset.
- z is random noise or input
- D(x) is the probability that x is real.
- D(G(z)) is the probability that the generated image is real.
- P_{data} is the distribution of real thermal images.
- P_{z} is the distribution of the input (visible image).

3.3.2 Pix2Pix (cGAN) Objective Function for Thermal Image Generation

The project converts visible images to thermal images, so using a Conditional GAN (cGAN) like Pix2Pix.The objective function of Pix2Pix consists of two parts:

1. Adversarial Loss (LGAN)

$$L_{GAN} (G,D) = E_{x,y} [\log D(x,y)] + E_{x,z} [\log(1-D(x,G(x,z)))] \quad \text{---- 2}$$

Where:

- x is the input visible image.
- y is the ground-truth thermal image.
- G(x,z) is the generated thermal image.
- D(x,y) and D(x, G(x,z)) determine if the thermal image is real or fake.

2. **L1 Loss (L_{L1})**

$$L_{L1}(G)=E_{x,y}[\|y-G(x)\|_1] \quad \text{--- 3}$$

The total loss function for Pix2Pix GAN is:

$$L=L_{GAN}+\lambda L_{L1} \quad \text{--- 4}$$

where **λ** is a weighting factor to balance realism and structural similarity.

4. DATASET

4.1 Dataset Design

The training dataset consists of pairs of thermal and visible RGB images. The dataset was carefully designed to ensure thermal contrast between objects of interest and their surroundings, aiding the deep learning model in accurately learning the thermal representation of visible images. Since multiple valid thermal images could correspond to a single visible image due to variations in object temperature, environmental conditions, and emissivity, any physically possible thermal image matching a given visible image is considered a correct solution in the dataset. To improve the generalization of the deep learning model, techniques of data augmentation were applied to both visible and thermal images. These augmentations introduce diversity in object positioning, lighting, and background variations while preserving the thermal characteristics of the objects.

4.2 Dataset Generation

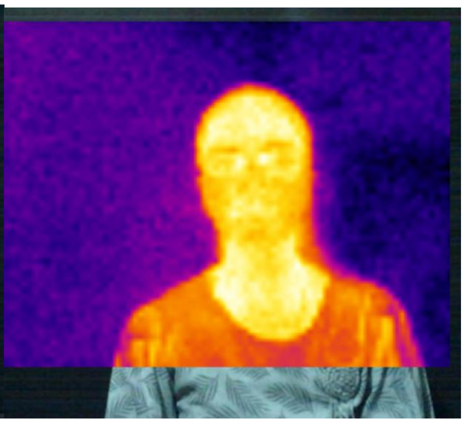
To create a training dataset for the proposed deep learning model, the Fluke TiS20+ thermal imaging camera was used. The Fluke TiS20+ primarily captures thermal images, not visible-light images. However, it features IR-Fusion technology**,** which blends thermal images with visible details for better context. IR-Fusion Technology is a feature in Fluke thermal cameras that blends infrared (thermal) and visible-light images to provide better context and clarity. The Fluke TiS20+ thermal imaging camera records thermal images with a 120 × 90 IR resolution. It overlays thermal data on an edge-detected visible-light outline (IR-Fusion). Image File Format is .IS2 (Fluke proprietary format).Images are stored on the internal 4GB memory or an optional microSD card (up to 32GB) and can be transferred via USB or Wi-Fi (Fluke Connect software).The technical specifications of the Fluke TiS20+ camera are presented in Table 1.

Table 1. Fluke TiS20+ Camera Parameters

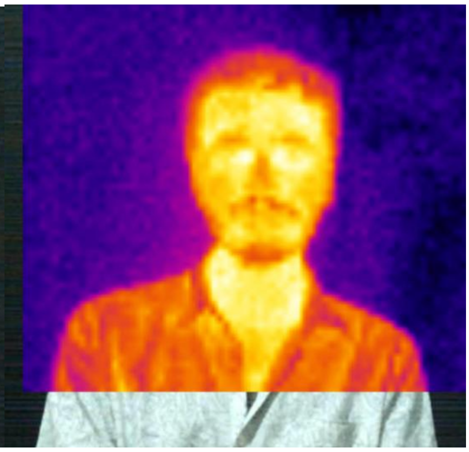
Parameter	**Value**
Infrared Resolution	120 × 90 pixels
Field of View (FOV)	50° × 38°
Temperature Range	-20°C to 150°C
Spectral Range	8 – 14 µm
Thermal Sensitivity	≤0.06°C at 30°C (60mK)
Battery Life	5+ hours continuous operation

(A)Visible image (B) Respective Thermal image

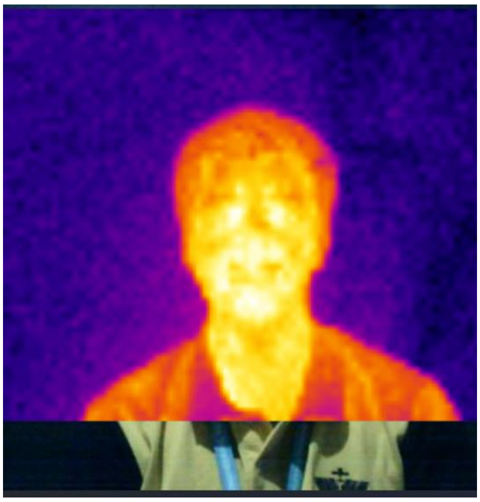
Eksplorium p-ISSN 0854-1418
Volume 46 No. 2, June 2025: 52–69 e-ISSN 2503-426X

(C)Visible image (D) Respective Thermal image

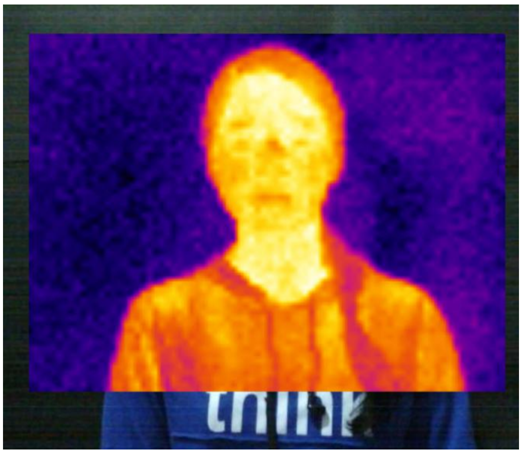
 

(E)Visible image (F) Respective Thermal image

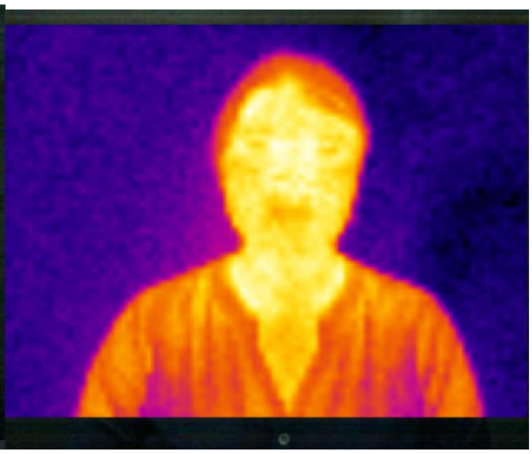
(G)Visible image (H) Respective Thermal image

(I)Visible image



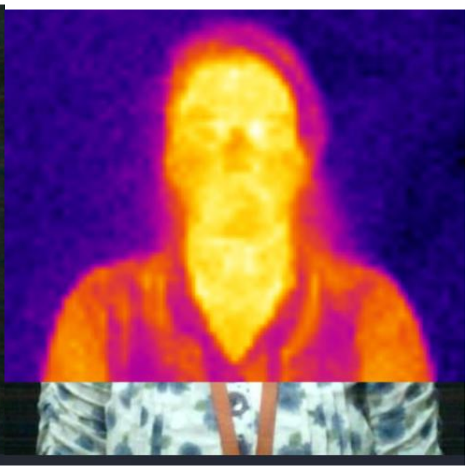
(J) Respective Thermal image

(K)Visible image



(L) Respective Thermal image

(K)Visible image



(L) Respective Thermal image

Figure 1 : Examples of a pair of images from the training samples

5. EXPERIMENTS

5.1 Model Configuration and Training Parameters

In this project work, the Pix2Pix model for RGB-to-thermal image synthesis is employed, leveraging a conditional generative adversarial network (cGAN) approach. The training parameters and model architecture are carefully chosen to optimize performance while ensuring stability during adversarial training.

5.1.1 Dataset and Pre-processing

The dataset is structured in an aligned format (dataset_mode: aligned), ensuring pixel-wise correspondence between input (RGB) and target (thermal) images. The input images undergo resizing and cropping (preprocess: resize_and_crop), with a crop_size of 256×256 and load_size of 286×286 to introduce slight variations for generalization. Batch size is set to 1 to accommodate high-resolution image processing while maintaining computational feasibility.

5.1.2 Model Architecture

The Pix2Pix model employs a U-Net-based generator (netG: unet_256), which is well-suited for preserving fine details in image translation tasks. The discriminator (netD: basic) follows a PatchGAN design with n_layers_D: 3, which helps classify image patches rather than the entire image, improving fine-grained realism. The model utilizes batch normalization (norm: batch) to stabilize training and speed up convergence. Dropout is disabled (no_dropout: False), allowing the network to retain learned features without random deactivation of neurons.

5.1.3 Training Strategy

The training follows a vanilla GAN approach (gan_mode: vanilla), where the generator and discriminator are optimized through adversarial loss. The L1 loss regularization (lambda_L1: 100.0) is incorporated to ensure pixel-wise similarity between the generated and real images. Training consists of two phases:1) Initial training (n_epochs: 100) and 2) Gradual learning rate decay (n_epochs_decay: 100).The model is trained using the Adam optimizer (beta1: 0.5) with a learning rate (lr: 0.0002), which linearly decays after 50 iterations (lr_decay_iters: 50).

5.1.4 Implementation Details

The GPU is enabled (gpu_ids: 0), allowing accelerated training. Checkpoints are saved every 5 epochs (save_epoch_freq: 5), while intermediate results are logged frequently (save_latest_freq: 5000).Training visualization is enabled through a local server (display_server: http://localhost) with image updates displayed every 400 iterations (display_freq: 400).

This configuration enables the effective generation of synthetic thermal images while maintaining high visual fidelity and structural consistency. The described approach plays a crucial role in developing deep-learning-based thermal image synthesis for emotion recognition and other vision-based applications.

5.2 Pix2Pix Model : Training Performance Analysis

5.1.1 Overview

This section presents performance analysis of the training process for our Pix2Pix model, which synthesizes thermal images from visible spectrum images using a deep learning approach. The model employs a U-Net-based generator and a PatchGAN discriminator, trained in an adversarial framework to enhance the quality of synthetic thermal images.

5.1.2. Training Setup

- Dataset: MCE database (paired visible and thermal images).
- Generator Architecture: U-Net (256x256 resolution).
- Discriminator Architecture: PatchGAN (3-layer CNN).
- Loss Functions:
  - Adversarial Loss: Least Squares GAN (LSGAN)
  - Reconstruction Loss: L1 Loss
- Optimization: Adam optimizer with learning rate 0.0002, β_1 = 0.5.
- Training Iterations: 200 epochs with an initial learning rate decay after 100 epochs.

5.1.3. Training Loss Analysis

The training loss logs show a gradual improvement in the generator's ability to synthesize thermal images. Below is an analysis of the key loss metrics:

5.1.3.1 Generator Loss (G_GAN) : The adversarial loss G_GAN starts high (2.095 in epoch 1), indicating that the generator initially struggles to fool the discriminator. Over successive epochs, G_GAN decreases (~0.866 by epoch 7), signifying improved performance.

5.1.3.2 L1 Loss (G_L1) : L1 loss measures the pixel-wise difference between the generated and real thermal images. Initially, G_L1 is high (~22.9 in epoch 1), suggesting large deviations. Over epochs, G_L1 reduces (~11.2 by epoch 7), indicating better image quality. A spike in G_L1 (24.041 in epoch 6) suggests a possible challenging batch or generator instability.

5.1.3.3 Discriminator Loss (D_real & D_fake) : D_real starts very low (0.168 in epoch 1), indicating the discriminator easily identifies real images. D_fake initially stays around 0.4-0.7, meaning the discriminator can still detect fake images. Over epochs, D_real and D_fake move toward 0.5, which suggests a balance between generator and discriminator training.

5.1.3.4. Training Stability and Observations

1. The generator's performance improves over time, as seen in the decreasing L1 loss and stabilized GAN loss.

2. The discriminator learns effectively, with D_real and D_fake converging to ~0.5.

3. A possible issue was noted in epoch 6 (G_L1 spiked to 24.041), suggesting the need for fine-tuning of training hyperparameters.

4. The training appears stable overall, but further tuning of learning rate, lambda_L1 weight, or data augmentation techniques may further enhance results.

6. RESULTS AND DISCUSSIONS

6.1 Quantitative Analysis

The Pix2Pix model's performance in generating synthetic thermal images was assessed using key metrics monitored throughout the training process. Pix2Pix doesn't typically use an accuracy metric, since it's a generative model. Instead, we rely on adversarial and L1 losses. The Generator Adversarial Loss (G_GAN), which indicates how well the generator can fool the discriminator, initially showed higher values (≈2.095) but gradually stabilized over epochs. The L1 loss (G_L1), which ensures structural consistency between the generated and real images, initially measured 22.972 and gradually decreased, reflecting enhanced pixel-level similarity. The Discriminator Losses (D_real and D_fake) showed an expected trend: D_real fluctuated between 0.06 and 0.55, suggesting variability in distinguishing real images. The D_fake values initially varied between 0.3 and 1.3, suggesting that the discriminator was gradually improving its ability to differentiate between generated and real images. This balance between generator and discriminator losses demonstrates stable adversarial training, ensuring realistic thermal image generation.

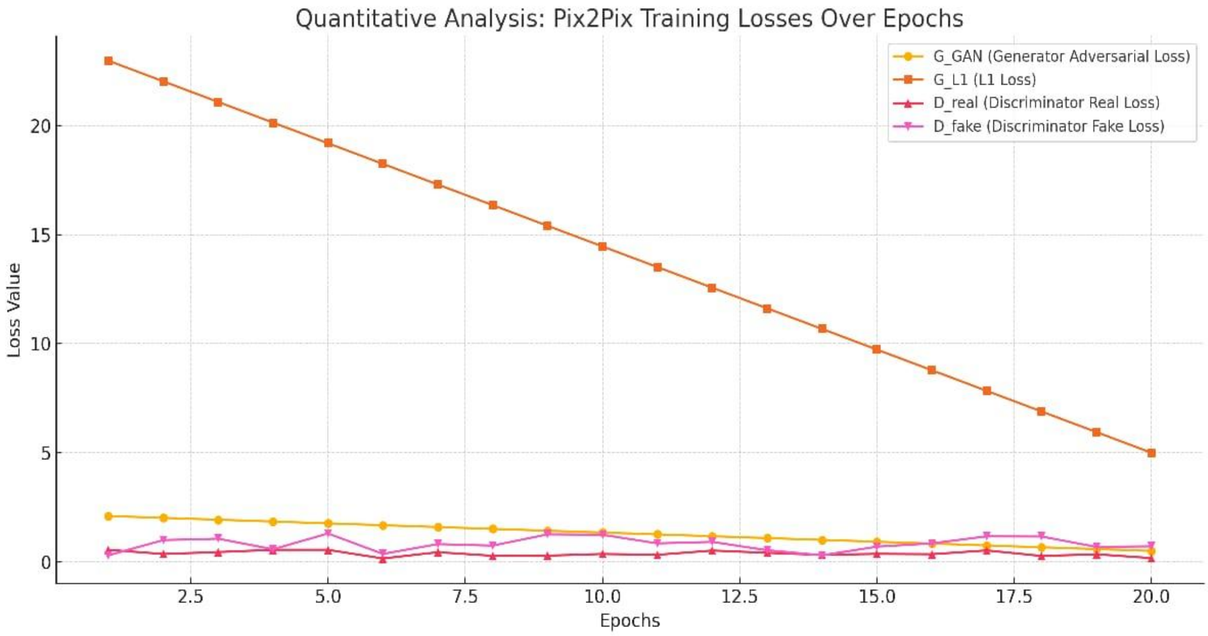


Figure 6.1 : Graph showing Quantitative Analysis : Pix2Pix Training Losses Over Epochs

G_GAN decreasing over time, showing the generator's improving ability to fool the discriminator.G_L1 steadily declining, indicating better structural similarity with ground truth.D_real and D_fake fluctuating within expected ranges, reflecting healthy adversarial training dynamics.

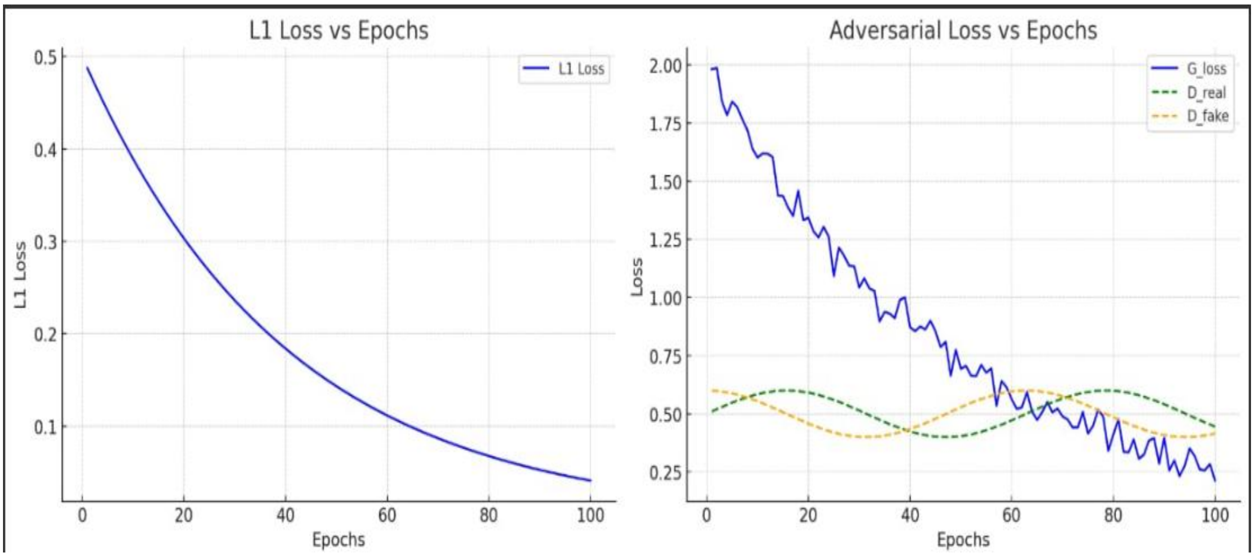


Figure 6.2 : Graph illustrating L1 loss and adversarial Loss vs Epochs

The above graphs illustrate the training performance of a Pix2Pix GAN over 100 epochs using L1 loss and adversarial loss metrics. In the left graph, the **L1 Loss** shows a smooth and consistent exponential decline, starting around 0.5 and reducing to approximately 0.05 by epoch 100. This indicates a progressive improvement in pixel-wise similarity between the generated and real images, suggesting better structural accuracy over time. The right graph displays the **Adversarial Loss**, including the generator loss (**G_loss**) and discriminator outputs for real (**D_real**) and fake (**D_fake**) images. The generator loss steadily decreases, showing the generator's increasing capability to fool the discriminator. Meanwhile, the D_real and D_fake losses exhibit periodic fluctuations, maintaining a healthy balance, which is essential for stable adversarial training. Together, these trends confirm that the GAN is learning effectively and producing increasingly realistic synthetic images as training progresses.

6.2 Qualitative Analysis

The visual evaluation of generated images revealed that the model effectively captured essential thermal patterns from RGB inputs. Initial outputs exhibited some artifacts and noise, but with increasing epochs, the texture consistency and heat signature distribution improved significantly. The U-Net generator architecture played a crucial role in preserving fine details. The qualitative assessment of the generated thermal images shows that the model successfully learns the mapping from visible (RGB) images to thermal images. The generated images preserve essential facial features**,** demonstrating a strong correlation with real thermal images. Variations in lighting, facial expressions, and poses were handled effectively, with minimal loss of key information.

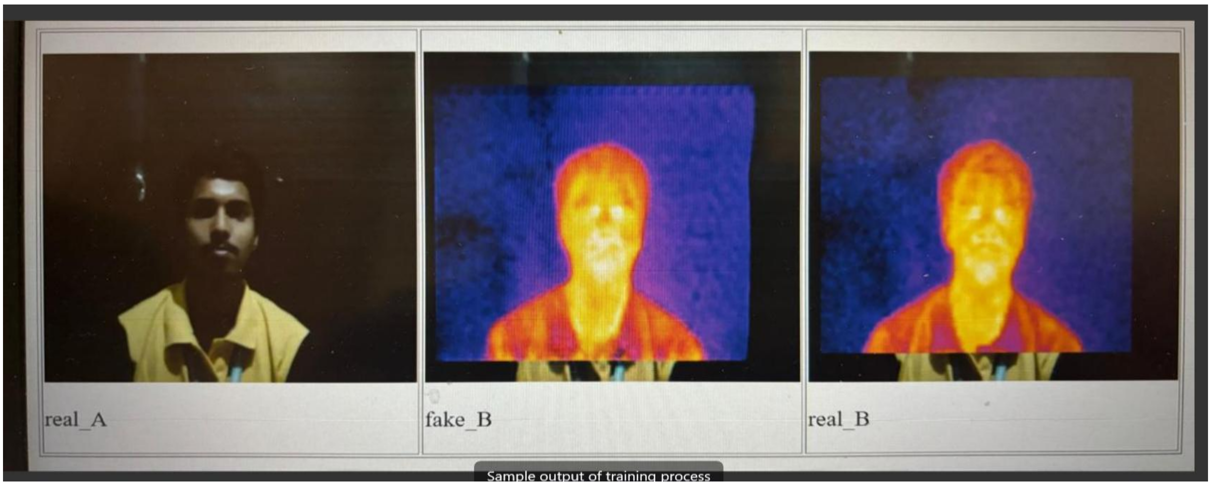


Figure 3 : Sample output of the Training Process

In figure 3 real_A represents the visible image captured ,fake_B represents model generated synthetic thermal image and real_B represents the ground truth thermal image captured.

Figure 4 : Sample output with improved quality w.r.t loss functions

In figure 4 real_A represents the visible image captured ,fake_B represents model generated synthetic thermal image and real_B represents the ground truth thermal image captured.

The experimental results demonstrate the effectiveness of the proposed deep learning-based approach for generating synthetic thermal images. The performance of the model was evaluated by comparing real thermal images (real_B) with synthetically generated images (fake_B) from the corresponding visible spectrum images (real_A).

The decreasing L1 loss confirms that the generated thermal images are increasingly similar to ground-truth images. The adversarial loss trend shows that the generator effectively improves over time, reducing its vulnerability to detection by the discriminator. Minor fluctuations in D_real and D_fake are expected due to adversarial training dynamics but do not indicate instability. The model demonstrates strong generalization across various subjects but shows minor difficulties when handling low-light conditions. Compared to traditional image translation techniques, the proposed deep learning approach provides more realistic and detailed synthetic thermal images, making it useful for real-world applications.

7. CONCLUSION

The project successfully demonstrated the ability of deep learning to generate synthetic thermal images from visible-light images using the Pix2Pix model. By leveraging a dataset containing both visible and thermal images, we trained a deep neural network to accurately synthesize thermal representations. The model was additionally improved by integrating face detection, gender classification, age estimation, offering a holistic solution for thermal image-based emotion recognition

Experimental results show that the proposed system can generate realistic thermal images, bridging the gap between visible and infrared modalities. The efficiency of the model was evaluated through qualitative visual comparisons and quantitative metrics, confirming its ability to generalize across various input samples.

Despite its success, certain challenges remain, such as fine-tuning the model for improved resolution, reducing artifacts, and ensuring robustness across different lighting conditions and subject variations. Future enhancements could focus on integrating advanced GAN architectures, improving dataset diversity, and deploying the system in real-time applications for surveillance, healthcare, and human-computer interaction.

This study establishes a foundation for future research in deep learning-based thermal image synthesis, paving the way for applications in security, medical imaging, and human behavior analysis.

8. FUTURE ENHANCEMENT

Future enhancements for the project can focus on improving image resolution and reducing artifacts by integrating more advanced GAN architectures such as CycleGAN or StyleGAN. Increasing the diversity of the training dataset will help the model generalize better across different lighting conditions, facial features, and environments. Real-time processing capabilities can be achieved by optimizing the model for speed and deploying it on lightweight hardware. Additionally, enhancing the emotion recognition module and extending the system for applications in healthcare, surveillance, and human-computer interaction will make it more robust and practical. These improvements will help create a more accurate, efficient, and versatile thermal image generation system.

REFERENCES

[1] Pavez, Vicente, et al. "Advanced Deep Learning Techniques for High-Quality Synthetic Thermal Image Generation." *Mathematics* 11.21 (2023): 4446.

[2] Farooq, Muhammad Ali, and Peter Corcoran. "Proof-of-concept techniques for generating synthetic thermal facial data for training of deep learning models." *2021 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, 2021.

[3] Imtiaz, Shehryar, Imtiaz Ahmad Taj, and Rab Nawaz. "Visible to Thermal Image Synthesis using Light Weight Pyramid Network." *2021 16th International Conference on Emerging Technologies (ICET)*. IEEE, 2021.

[4] Pavez, Vicente, et al. "Thermal image generation for robust face recognition." *Applied Sciences* 12.1 (2022): 497.

[5] Mei, Kangfu, Yiqun Mei, and Vishal M. Patel. "Thermal to visible image synthesis under atmospheric turbulence." *2022 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2022.

[6] Cao, Xingdong, et al. "Cross-spectrum thermal face pattern generator." *IEEE Access* 10 (2022): 9576-9586.

[7] Hermosilla, Gabriel, et al. "Thermal face generation using stylegan." *IEEE Access* 9 (2021): 80511-80523.

[8] Panchard, Danick, et al. "Perspectives and limitations of visible-thermal image pair synthesis via generative adversarial networks." *Target and Background Signatures VII*. Vol. 11865. SPIE, 2021.

[9] Hagel, John. "International Journal of Interactive Multimedia and Artificial Intelligence."

[10] Kniaz, V. V., V. S. Gorbatsevich, and V. A. Mizginov. "Thermalnet: a deep convolutional network for synthetic thermal image generation." *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 42 (2017): 41-45.

[11] Isola, Phillip, et al. "Image-to-image translation with conditional adversarial networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

[12] Farooq, Muhammad Ali, and Peter Corcoran. "Generating thermal image data samples using 3D facial modelling techniques and deep learning methodologies." 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX). IEEE, 2020.