University of Sussex

Primate Thermal Video Modelling: Fence Removal and Feature Tracking

Maksut O'Connor

Supervisor: Dr. Ivor J. Simpson

Department of Engineering and Informatics

BSc (Hons) Computer Science and Artificial Intelligence

April 11, 2025

Acknowledgements

I would like to thank my supervisor, Dr. Ivor J. Simpson, for his tenativeness, important advice, assistance, and insight for the duration of this study. His knowledge and guidance have influenced the direction and course of this investigation substantially.

I am also incredibly grateful to my family and friends for their constant encouragement and support. Their confidence in my skills has sustained my motivation throughout this process.

Declaration

I confirm that the dissertation *"Primate Thermal Video Modelling: Fence Removal and Feature Tracking"*, submitted for the degree of BSc (Hons) in Computer Science and Artificial Intelligence, is my original work. All sources consulted have been properly acknowledged, and I have not submitted this work for any other academic award. I understand the consequences of academic dishonesty and attest that this dissertation has no fabrication or plagiarism.

Approximate Number of Words: 10,190

Date Submitted: April 11, 2025

Signature:

Abstract

The extraction of thermal imaging measurements from primates in unconstrained zoo environments where fencing obstructs visibility is explored in this project. The removal of fence constructions from thermal video footage is essential to raising the precision of facial feature tracking, making it a crucial component of this study.

In this project, a new data processing pipeline is implemented that integrates multiple inpainting techniques and preprocessing strategies to optimize feature extraction and obstruction mitigation. To tackle this, the effectiveness of several inpainting methods are compared in object mitigation, including SIREN-based implicit neural representations, OpenCV-based inpainting methods (Navier-Stokes and Telea), and the deep learning-based EdgeConnect method.

Furthermore, different preprocessing techniques were explored to enhance the effectiveness of feature detection and inpainting accuracies, including contrast normalization (CLAHE), Gaussian blurring, adaptive thresholding, and morphological transformations. The effects of these techniques on edge enhancement, noise reduction, and overall tracking stability in thermal images were evaluated.

Additionally, this investigation focuses on non-invasive facial feature tracking methods to identify regions of interest, such as the eyes and nose, in thermal video frames. Methods such as Haar cascades, Sobel edge detection, Canny edge detection, and deep learning-based tracking are explored.

To ascertain how obstacle removal affects tracking performance, the efficacy of various methods is assessed after inpainting. Through the implementation and comparison of these techniques, the goal is to aid in the development of a robust methodology for automated, non-invasive primate monitoring in challenging thermal imaging conditions. The results help to improve automated behavioral analysis of primates in enclosed spaces and shed light on how feature tracking accuracy is affected by obstruction removal.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Thermal imaging offers insights into temperature regulation, stress responses, and behavioral patterns, making it a powerful tool for the non-invasive monitoring of biological subjects. Specifically, thermal imaging of primates in zoo settings provides scientists with useful physiological information [16, 13, 25]. However, physical obstacles like fences pose a serious difficulty in effectively evaluating thermal footage because they generate noise and occlusions that hinder facial feature tracking [5, 7].

This is an area of computer vision that lacks research, and traditional techniques struggle to reconstruct occluded regions in thermal imagery [41]. This necessitates the investigation of advanced inpainting methodologies to restore the missing regions and extract the required thermal data. Inpainting is a process in which missing regions are filled in and reconstructed to reveal the occluded portions of an image [26, 40]. In this investigation, it is extremely important to remove fences while preserving the integrity of primate facial features. Multiple inpainting methods, such as SIREN-based implicit neural representations [35], OpenCV-based approaches (e.g., Telea and Navier-Stokes) [9], and deep learning-based EdgeConnect models [26] are assessed for their effectiveness in mitigating obstructions in thermal footage.

Feature tracking is a prominent component of this investigation. Detecting and tracking facial landmarks, including the eyes or nose, plays a fundamental role in the study of primate behavior [31, 32, 16]. Traditional methods, such as the Lucas-Kanade feature tracker [33], Haar cascades [41], and deep-learning-based tracking methods [24] are explored to determine their efficiency in post-inpainting facial landmark tracking. Additionally, the effects of several preprocessing methods on inpainting and feature detection accuracy are examined. These methods include contrast normalization (CLAHE) [30], Gaussian blurring, adaptive thresholding [23], and morphological alterations [8].

A new data processing pipeline is introduced in this study, integrating inpainting, preprocessing, and feature tracking techniques to increase the accuracy of non-invasive thermal imaging analysis. Evaluating the efficacy of the methods discussed, this research contributes to the development of a robust methodology for an automated primate monitoring system by offering valuable insights for zoological studies and conservation efforts [20, 6]. The study's conclusions aim to improve automated behavior tracking in confined spaces and offer a foundation for future developments in thermal video analysis.

We hypothesize that once accurate inpainting has been performed on the occluded regions in the thermal video frames, feature tracking and, in turn, temperature extraction results will vastly improve. The inpainting process is expected to enhance nasal tracking performance and produce increased stability for thermal measurements over time. Without the presence of inpainting, obstructions in the frames mislead deep-learning-based detectors, causing the tracking to drift and degrade in accuracy [21, 38]. We anticipate that modern inpainting techniques will lead to measurable improvements in feature localization and reduce noise in temperature extraction results.

The chapters that follow provide background regarding the current state of research in inpainting technologies, thermal feature tracking, and preprocessing strategies (Chapter 2), a description of the methodologies used and implemented along with model configurations (Chapter 3), results from experimentation and analysis of failures (Chapter 4), and a discussion of conclusions and future research directions (Chapter 5).

1.1 Motivations

It is crucial to understand primate behavior for both conservation efforts and ecological studies [20, 7]. Advancements in thermal imaging technology in recent years have enabled researchers to detect and monitor physiological changes in primates while avoiding invasive procedures. However, conducting this research in zoo environments presents a significant challenge, as obstructions such as fences often interfere with feature extraction. For precise tracking and temperature analysis of primates, these obstacles must be removed.

1.2 Research Objectives

In this study, different inpainting methods are tested to successfully inpaint the occlusive fence structure with minimal error or information loss, enabling the automatic tracking of the nasal region of primates in thermal video data. The main goals of this project consist of:

- The implementation of different computer vision techniques, including Haar Cascades, Canny edge detection, and Convolutional Neural Networks (CNNs), for the detection of the nasal region of primate faces in thermal video frames.
- To develop an image processing pipeline that can mitigate fence-induced occlusions by utilizing various techniques such as morphological operations, edge detection, and multiple inpainting methods, including but not limited to OpenCV-based approaches, EdgeConnect, and SIREN-based models.
- To extract temperature data from the nasal region of primates and analyze stress levels based on the thermal information observed.
- To assess the performance of the system in terms of fence removal and nose detection techniques, and to determine which methods are most suitable and efficient for use in challenging, real-world thermal video conditions.

1.3 Research Questions

This investigation is driven by the following research questions:

1. Of the tested inpainting methods, which one is most effective at removing the fence structure while preserving the thermal data?
2. How is the performance of feature tracking algorithms affected by the fence removal process?
3. Which tracking method is most effective at detecting primate facial features in thermal videos?

1.4 Ethical Considerations

This project adheres strictly to established ethical standards, as no human data was collected or utilized at any point during the investigation. All thermal video data employed in the analysis was provided by Dr. Gilly Forrester and her research team, and was obtained in alignment with the ethical guidelines governing animal research and welfare.

The data acquisition process was entirely non-invasive, consisting solely of passive thermal imaging of primates within their zoo enclosures. As such, the study does not pose any risk to the health, safety, or well-being of the animals involved. Thermal imaging, by nature, avoids physical contact and thus complies with both institutional and broader scientific standards for humane treatment in behavioral and physiological studies.

Furthermore, the investigation maintains alignment with the BCS Code of Conduct, particularly its directives on ethical integrity, transparency, and the prioritization of animal welfare in research contexts. Every effort was made to ensure the confidentiality, non-interference, and responsible handling of the thermal datasets, underscoring a commitment to methodological rigor and ethical accountability.

Chapter 2

Background

This chapter examines relevant literature and theoretical foundations that support the development of thermal inpainting and facial feature tracking systems. It aims to contextualize the primary challenges in thermal video processing, explain the rationale behind the selected inpainting and tracking techniques, and highlight the role of preprocessing in thermal imagery.

2.1 Thermal Imaging for Primate Monitoring

The implementation of thermal imaging technology to non-invasively monitor animals such as primates has gained momentum in recent years. This provides researchers with insights into stress levels, health status, and behavioral responses in primate species [27]. A degree of infrared thermography has been applied to identify the emotional state in macaques by monitoring their nasal temperature [25], and this is further validated by studies on chronic stress detection based on temperature variations on the surface of primate skin [13, 16]. The nasal region is especially relevant because it is not covered by a lot of fur, is highly vascularized, and sensitive to autonomic changes [24]. However, these examples are limited in real-world applications, especially in zoo environments, where fences are present, obstructing the quality of thermal video and compromising feature tracking abilities [5, 8].

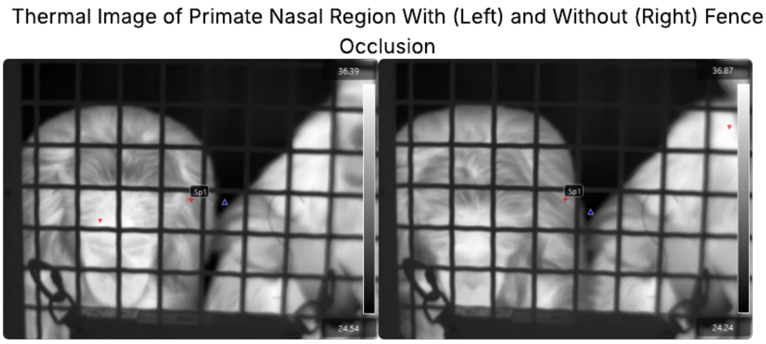

Figure 2.1: Demonstration of how the occlusive fence structure can block primate features

To tackle such obstructions, various image inpainting techniques that have not been previously investigated or lack research must be explored in this area [4]. Traditional methods in the computer vision literature, such as Telea and Navier-Stokes inpainting, rely on partial differential equations to propagate image structures from known to unknown regions of interest (ROI). These methods can be effective in low-frequency reconstructions, but they are insufficient for the task as they have been seen to fail in high-detail regions like animal faces [2]. This restriction is addressed by deep learning techniques like EdgeConnect [26], which increase the accuracy of generative inpainting by employing edge prediction as a structural guide. Similarly, SIREN-based implicit neural representations provide high-fidelity reconstructions free from grid discretization artifacts by modeling continuous image functions with sinusoidal activations [35].

The role of preprocessing is also crucial in improving the quality of inpainting and detection outcomes. Thus, techniques including contrast-limited Adaptive Histogram Equalization (CLAHE), Gaussian filtering, and adaptive thresholding are common applications for the normalization of thermal image intensity distributions [5, 8]. Therefore, these steps are critical in culmination with inpainting and tracking techniques to improve downstream tasks.

The lack of texture and poor contrast in thermal imaging present special difficulties for face detection. This means that classical methods such as Haar cascade classifiers [41] remain useful due to their simplicity and speed but are unable to accurately detect regions in the face of primates. CNN-based detectors that are more robust, especially under instances of occlusion [3], perform better, and deep-learning-based models have been successful in species-specific recognition tasks and real-time tracking [3], although training data scarcity remains a challenge [7].

2.2 Obstruction and Fence Removal in Computer Vision

In real-world visual environments, it is common for occlusions to impede object detection tasks, and they can interfere with the scene. In the context of our project within zoological and wildlife monitoring applications, there are mesh structures, fences, and enclosure bars frequently coupled with noise in both visible and thermal imagery. As discussed before, these structures can interfere with ROIs, such as the primate nasal area in our case, which can lead to incomplete or inaccurate feature extraction.

Research shows that the computer vision community has generally approached occlusion removal as a structural inpainting problem. The aim here is to reconstruct what is present behind the occluded regions visually and contextually coherently [15, 36]. Conventional inpainting techniques, such as Telea's rapid marching approach and the Navier-Stokes PDE-based algorithm, both implemented in OpenCV, spread pixel values from nearby regions to fill in masked areas [11]. Despite being computationally effective, these techniques typically only work with smooth backdrop reconstruction and have trouble with semantically significant textures, particularly when thermal limitations are present and gradients are faint and repeated [29, 17, 40].

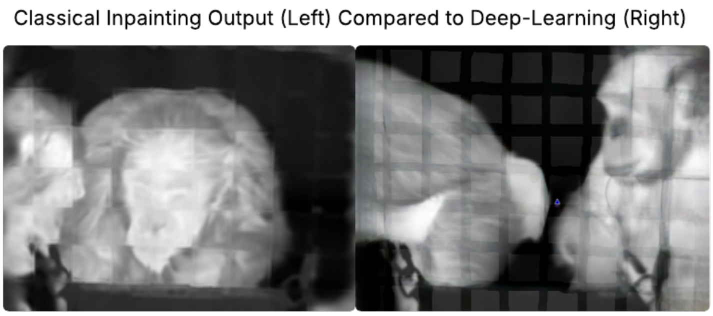

Figure 2.2: Comparison of inpainting outputs using classical and deep learning-based approaches

Deep learning-based inpainting has emerged as a powerful alternative. Due to the limitations of conventional methods, deep learning-based inpainting approaches have been experimented on and introduced [9, 38]. Notably, the EdgeConnect model [26] leverages edge prediction using canny edge detection coupled with a contextual inpainting stage using a convolutional neural network (CNN). This two-stage approach is especially effective because it can restore high-frequency structural information such as facial contours or edges. EdgeConnect functions by predicting an edge map for the occluded region of the image, and then it uses that to blend the missing content, which is particularly helpful when working with occlusions such as wires or fences.

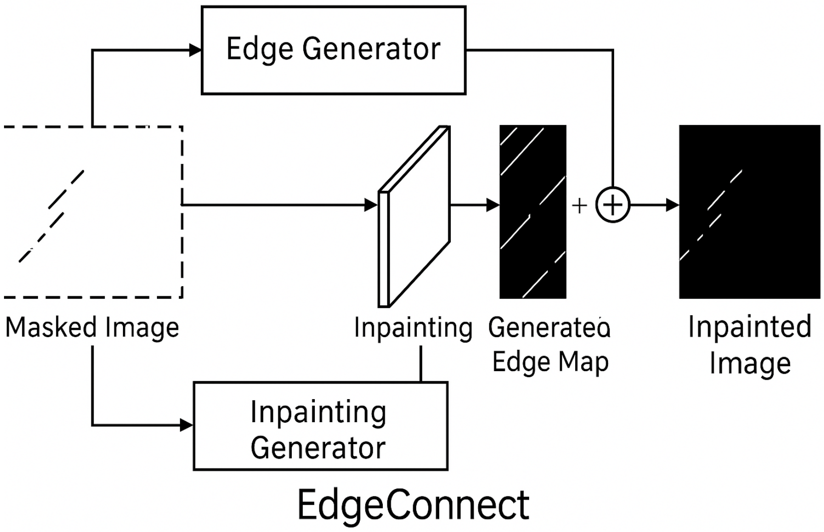

Figure 2.3: EdgeConnect Architecture Diagram

Another potential paradigm is seen through SIREN (Sinusoidal Representation Networks) [35]. These networks avoid convolutions entirely and utilize a coordinate-based Multi-Layer Perceptron (MLP) coupled with a sinusoidal activation function to model continuous spatial representations of images. By collecting subtle thermal gradients without the introduction of gridding artifacts, this approach enables high-fidelity reconstructions. SIRENs are especially well-suited for inpainting jobs that necessitate the preservation of minor thermal fluctuations throughout the face since they represent an image as a continuous function.

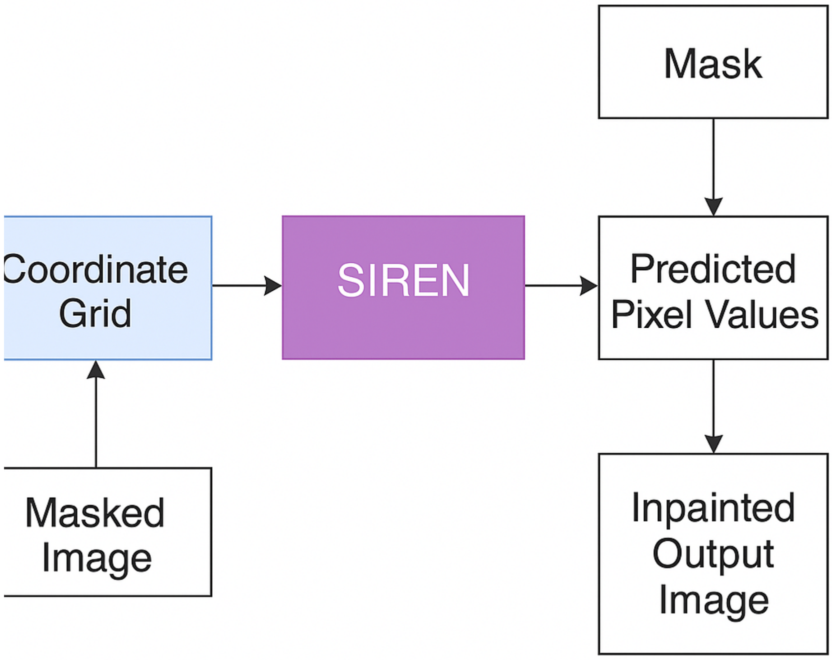

Figure 2.4: SIREN Architecture Diagram

Recent works, such as the ObstructionRemoval GitHub repository [4], illustrate the discussed techniques applied in real-time systems and wildlife datasets, proving their potential. However, while they yield promising performance, there remain challenges in the generalization of these models across different environments, species, and thermal camera specifications. Furthermore, a hybrid strategy of both classical and deep learning methods is investigated in this study for fence removal, with a streamlined focus on downstream performance in nasal region tracking.

2.3 Feature Detection and Tracking in Thermal Imagery

Facial feature detection in visible-spectrums for human images is well-established through Haar Cascades and CNN detectors. However, for primate and thermal imaging feature detection standard pipelines do not apply [41]. Primate facial features defer significantly from that of a human's, making these methods unable to track and identify their features.

Additionally, thermal video suffers from contrast variability, making the task all the more complicated. Burghardt et al. (2010) explored animal face detection in real time but the requirement for controlled lighting made it difficult to do so [7].

For thermal modalities, edge-based detection (Sobel, Canny) and intensity gradient tracking show greater promise. In this study, the Lucas-Kanade optical flow tracker—which works well in steady settings—was modified for continuous nasal tracking [31]. However, fence blockage and early keypoint reliability have a significant impact on tracking precision [31].

Figure 2.5: Example of Lucas-Kanade feature tracking trajectory before and after inpainting

Recent literature illustrates that CNN-based thermal object detectors and multimodal fusion methods can be optimal for improving tracking. However, these methods require large labeled datasets and computational resources, which are outside of the scope of this project [3, 7].

2.4 Preprocessing Strategies for Thermal Feature Analysis

Before the implementation of inpainting and tracking, several preprocessing techniques were tested and experimented on. Coupled with this, automatic mask generation was attempted; however, due to the sensitivity of inpainting and the imperfect nature of automatic mask generation, a custom mask is utilized in the final product [5, 8]. To elaborate, if a few pixels are not identified by the custom masks, this led to flaws in the inpainting process and inaccurate reconstructions.

- Contrast Limited Adaptive Histogram Equalization (CLAHE) to enhance local contrast [5, 8].
- Gaussian Blur to reduce high-frequency sensor noise [5, 8].
- Adaptive Thresholding and Morphological Operations for mask refinement [5, 8].

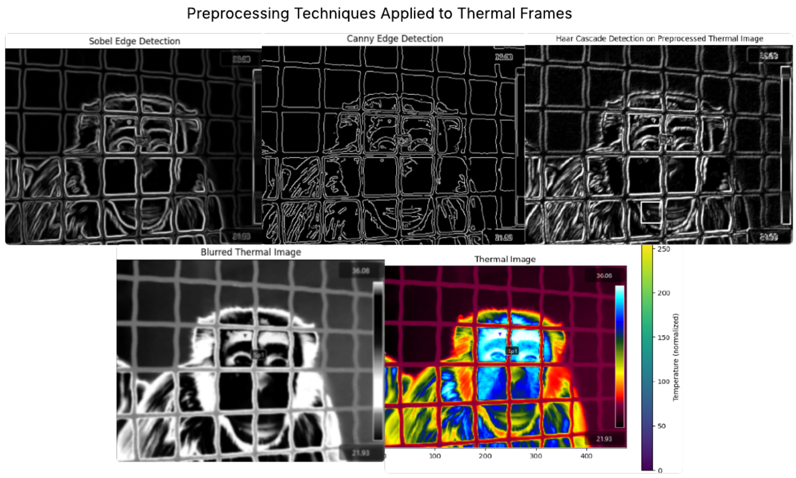

Figure 2.6: Some Preprocessing Techniques Applied To Individual Frames

Automatic fence detection was attempted through the utilization of the following methods:

- Binary Thresholding: Divided fence areas from the background using pixel intensity thresholds after converting thermal frames to binary representation [5, 8].
- Hough Line Transform: Used to determine the linear structures that fences are known for. Mask regions were inferred for occlusion removal using detected lines [5, 8].
- Mask Dilation: Used to expand fence mask regions, ensuring full coverage of occlusions and avoiding partial fence remnants after inpainting [5, 8].

These methods were tested on several frames and visually compared for their accuracy and coverage. While binary thresholding and Hough line transform techniques proved to be decent in identifying the fence, they also took into consideration some portions of the primates' torso, which affected the inpainting process negatively. Combining the methods and then applying morphological dilation improved their effectiveness, but a custom mask was opted for as it gave more consistent reconstructions.

Figure 2.7: Automatic Fence Detection and Mask Generation Methods Tested

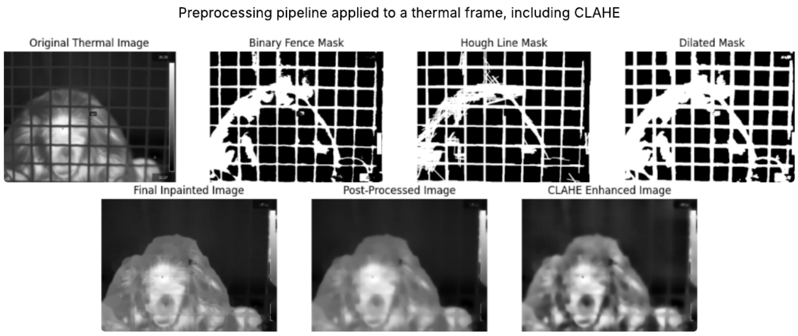

Figure 2.8: Thermal Frame Preprocessing Pipeline With Post-Processing and CLAHE

2.5 Summary and Hypothesis

In summary, existing works and literature illustrate that the implementation of modern deep-learning based inpainting approaches can improve the effectiveness of tracking systems on video footage [26, 35]. The hypothesis this study is built upon is that the more accurate the inpainting reconstruction is of the nasal region being tracked, the higher the tracking stability and more reliable temperature signal extraction will be [20]. This chapter laid the theoretical and empirical groundwork for the system design discussed and detailed in Chapter 3, where implementation and evaluation metrics are discussed.

Chapter 3

Methods

3.1 Data Collection and Preprocessing

The thermal dataset utilized in this project was provided by Dr. Gilly Forrester and her research team. They captured the data in zoo environments where they recorded different primate enclosures under constrained conditions. The video files were saved in .tif format, and 16-bit thermal intensity data, which is essential for temperature research at the pixel level, was preserved. FLIR Research Studio, a software platform that allowed precise frame extraction from the thermal video streams while preserving temperature fidelity, was used to preprocess the raw data [6, 20].

The individual extracted frames were stored in grayscale and normalized to maintain compatibility with image processing pipelines. The images in these pipelines are typically 256×256 or 416×416 pixels, which is what the extracted frames were resized to [11]. This is a required step for model compatibility and to reduce the computational load [40].

Many preprocessing techniques were measured for their impact on enhancing feature contrast and fence detection:

- **Gaussian Blurring:** For suppressing sensor noise and enhancing edge continuity [17].
- **CLAHE (Contrast Limited Adaptive Histogram Equalization):** Applied in low-texture thermal zones to improve the local contrast.
- **Sobel and Canny Edge Detection:** Utilized to visualize contours better and used in preprocessing steps when refining our mask [37].

- **Adaptive Thresholding and Morphological Operations:** Tested during mask creation to isolate fence structures based on thermal intensity and edge strength.

Early in the experimentation process, adaptive thresholding for automatic mask generation was considered depending on the gradient magnitude and through the utilization of Hough line transform, which proved best. However, a custom binary mask creation pipeline was implemented as it proved to be more effective. The custom approach was consistently reused across datasets because it allowed the precise definition of fence regions, removing any outlier pixels or portions of the fence not captured in the mask. Morphological operations, such as dilation and closing, were applied to adjust the masks.

In addition, when necessary, pixel value calibration procedures were used to extract frame-wise temperature data and translate it into degrees Celsius measurements. This was especially crucial for measuring temperature changes in the nose and other facial areas.

The mentioned enhancements were either explicitly or implicitly integrated into our final preprocessing pipeline coupled with our inpainting frameworks. For example, Edge-Connect already makes use of Canny edge detection algorithms, and input images were converted to RGB prior to being inpainted to be compatible with the system [26]. Furthermore, SIREN-based models needed grayscale normalization and a manual alignment of the mask [31]. To maintain consistent input and output quality, these standardized preprocessing steps had to be performed before the subsequent inpainting and feature-tracking stages.

3.2 Fence Removal Approaches

Thermal signal extraction and accurate feature detection are made difficult due to the presence of fencing. To alleviate this, many fence removal techniques are investigated upon and the task is framed as an image inpainting problem. The goal is to reconstruct the regions in the frames behind the fence in the thermal images with a focus on the facial region of the primates while preserving primate features as if the fence was never there.

3.2.1 Traditional Inpainting Methods

Initially, three OpenCV-based inpainting techniques were implemented and evaluated. This section offers an explanation and evaluation of these techniques.

- **Telea's Inpainting Algorithm:** This utilizes the fast marching method to propagate known pixel values into masked regions. The fast marching method (FMM) is used to solve the Eikonal equation, a nonlinear first-order PDE. Telea fills pixel values along contours using a weighted average based on distance and intensity [**?**, 15].

- **Navier-Stokes Inpainting:** Treats image continuity as a fluid dynamics problem by extending isophote lines into missing regions. Isophotes are lines of equal intensity, guiding structure continuation [36, 29].

- **Biharmonic Inpainting:** Involves solving a Laplace-based interpolation to propagate values into masked areas. Although capable of smooth extrapolation, it lacked fidelity in the facial regions of thermal frames [17, 9].

The visual coherence of these algorithms in uniform regions was acceptable. However, their lack of semantic understanding rendered them ineffective in reconstructing important features like the nose. This limitation necessitated more advanced, learning-based inpainting solutions.

3.2.2 Deep Learning-Based Inpainting (EdgeConnect)

A modified EdgeConnect approach was implemented to inpaint the thermal images consisting of a two-stage inpainting pipeline [26]:

- **Edge Generation Network:** This network uses a convolutional generator trained with adversarial and feature-matching losses to predict a believable edge structure within the occluded part of a masked grayscale picture.

- **Inpainting Network:** Using the anticipated edge map and contextual data from nearby pixels, the inpainting network completes the RGB image's missing regions.

Data Preprocessing and Preparation

- **Resizing and Normalization:** All the input images and masks were resized to fixed dimensions (256×256) and normalized between the range $[0, 1]$:
- **Mask Generation:** Binary masks $M(x, y)$ were formed using intensity-based thresholding techniques.

$$M(x, y) = \begin{cases} 1, & \text{if } I(x, y) > 127 \\ 0, & \text{otherwise} \end{cases}$$

- **Edge Extraction:** Canny edge detection was applied to the grayscale images to guide the inpainting process by generating structural priors.

Architecture of EdgeConnect

- **Edge Generator:** A series of convolutional layers, reflection padding, instance normalization, and ReLU activations are used to generate the edge map.
- **Inpainting Generator:** An encoder-decoder network with residual blocks. Masks and predicted edge maps are concatenated as a 4-channel input. Output layer uses `tanh` activation producing outputs in $[-1, 1]$ or $[0, 1]$.

Loss Functions and Training The overall loss is a weighted combination of multiple components:

- **L1 Loss (reconstruction loss): Applied over masked regions to penalize pixel-wise reconstruction error.**

$$\mathcal{L}_{L1} = \frac{1}{N} \sum_{i=1}^{N} \left| I_{\text{gen}}^{(i)} - I_{\text{orig}}^{(i)} \right|$$

- **Adversarial Loss:** Encourages realism by fooling a discriminator network.

$$\mathcal{L}_{adv} = \text{BCE}(D(I_{\text{gen}}), y_{\text{real}})$$

- **Feature Matching Loss:** Aligns intermediate discriminator features for smoother outputs.

$$\mathcal{L}_{fm} = \sum_i \|F_i(I_{\text{gen}}) - F_i(I_{\text{orig}})\|_1$$

- **Perceptual Loss:** Based on high-level features extracted from a pretrained VGG network.

$$\mathcal{L}_{perc} = \sum_i \|\phi_i(I_{\text{gen}}) - \phi_i(I_{\text{orig}})\|_1$$

- **Style Loss:** Enforces texture consistency using Gram matrix comparisons.

$$\mathcal{L}_{style} = \sum_i \|G_i(I_{\text{gen}}) - G_i(I_{\text{orig}})\|_1$$

- **Total Loss:** Weighting coefficients control the emphasis of each component.

$$\mathcal{L}_{\text{total}} = \lambda_{L1}\mathcal{L}_{L1} + \lambda_{\text{adv}}\mathcal{L}_{\text{adv}} + \lambda_{\text{fm}}\mathcal{L}_{\text{fm}} + \lambda_{\text{per}}\mathcal{L}_{\text{per}} + \lambda_{\text{style}}\mathcal{L}_{\text{style}}$$

Training and Initialization

- The weight initialization process is standard (Xavier) to make sure convergence is stable [32].
- Learning rates are tuned for the Adam optimizer and momentum values β_1, β_2.

Video Processing and Tracking Integration

Keypoint motion between frames I_t and I_{t+1} is estimated via Lucas-Kanade optical flow. The brightness constancy constraint is assumed as follows:

$$I_{t+1} : I(x, y, t) \approx I(x + \Delta x, y + \Delta y, t + \Delta t)$$

The position of the updated nose position is found through the weighted median of the tracked points.

Temperature Extraction: The average pixel intensities are computed in a localized region centered on the tracked nose location to determine a proxy temperature value:

$$T_{\text{proxy}} = \frac{1}{n}\sum_{i=1}^{n} I(x_i, y_i)$$

Composite Image Generation: The final image used for temperature extraction and analysis is a blend between the original and the inpainted image, guided by the mask M:

$$I_{\text{composite}} = M \cdot I_{\text{inpainted}} + (1 - M) \cdot I_{\text{original}}$$

The goal with the creation of this modified model is to offer a robust, edge-aware inpainting model that can restore accurately the masked regions of thermal imaging with correct contextualization and structural fidelity. In addition to enhancing static images, the integrated video processing pipeline allows for temperature analysis and frame-by-frame monitoring, guaranteeing that crucial biometric areas, such as the nose, can still be analyzed in post-processed thermal sequences.

3.2.3 Coordinate-Based Inpainting (SIREN)

A different approach to inpainting was implemented through SIREN (Sinusoidal Representation Networks). Differing from EdgeConnect, a CNN-based model, SIREN models the image as a continuous function by mapping coordinates (x, y) to various intensity values [35]. The network is comprised of fully interconnected layers with sinusoidal activation functions and has the potential to capture high-frequency detailing, proving it potentially quite useful for thermal gradients across occluded facial regions.

The SIREN model is constructed on the masked image coordinates and optimized to minimize reconstruction loss with valid regions [35]. This is done while extrapolating plausible thermal values in the occluded regions. The SIREN model represents an image as a continuous function $I(x)$ over a spatial domain space $\Omega \subset \mathbb{R}^2$ (or $\mathbb{R}^d$ as dimensions increase). This means that the images are modeled as a function $F(x) = I(x)$ where $x \in \Omega$ and Θ are the parameters that are learnable for the network.

The SIREN (Sinusoidal Representation Networks) is effective for the task due to the employed sine activation functions in the hidden layers. These sine functions let the network represent higher frequency details and smoothing gradients, which are essential to thermal imaging [35]. A custom initialization scheme was found to work best through experimentation and ensures the stability of the system when training. The first layer weights are sampled from:

$$\theta_1 \sim \mathcal{U}\left(-\frac{1}{n}, \frac{1}{n}\right) \quad \text{where } n = \text{input dimension}$$

All the subsequent layers are scaled inversely by the angular frequency ω_0. To train the network, a coordinate grid is constructed for the individual frames utilizing a mesh over the domain $[-1, 1]^2$. After flattening the grid, it is then passed through the network producing the pixel values accordingly [33].

For the inpainting process, a binary mask $M(x)$ is used to differentiate between known regions Ω_k and unknown regions Ω_u:

$$M(x) = \begin{cases} 1, & \text{if } x \in \Omega_k \\ 0, & \text{if } x \in \Omega_u \end{cases}$$

Over epochs, the masked L1 loss is minimized over the known regions, coupled with the optional regularizers:

$$\mathcal{L}_{\text{masked}} = M(x) \cdot |F(x) - I(x)|$$

$$\mathcal{L}_{\text{TV}} = \sum_{i,j} |I(i+1, j) - I(i, j)| + |I(i, j+1) - I(i, j)|$$

$$\mathcal{L}_{\text{grad}} = \sum_{i,j} |\nabla_x I(i, j)| + |\nabla_y I(i, j)|$$

Where:

- TV Loss controls the smoothness in spatially adjacent pixels.

- Gradient Loss preserves the continuity of edges or lines.
- L1 Loss is weighted by the mask so only known regions contribute.

Training was performed via the Adam optimizer and the coordinate grids and masks were generated dynamically for each frame. The conversions to grayscale and edge masks were computed prior to support the comparative analysis [34].

During inference, the images were reconstructed by the SIREN network, including both the known and unknown regions. The predicted values of Ω_u were substituted inside of the masked regions to produce the final output frame. SIREN was tested over several frames with different epoch values of 500, 1000, and 2000. The following example outputs were created:

Figure 3.1: SIREN Results After Different Epochs

The images were then overlaid with the original image using the fence mask to preserve detail in the non-fence regions and then put together to create the output video. Since the outputs had little to no difference in detail and pixel reconstruction quality, to make SIREN more computationally efficient, each frame was reconstructed for 500 epochs.

SIREN achieved good local fidelity and smoothness in the reconstructed regions; however, as visualized, any misalignment in the mask caused black lines to be constructed, which would interfere with any tracking being performed, making it a less-viable option for practical deployment. Nevertheless, SIREN acted as a benchmark in the experimentation process for evaluating the quality of inpainting under idealized constraints.

3.2.4 Final Pipeline Selection

Post-experiential evaluation of downstream feature tracking performance and visual consistency, the modified EdgeConnect implementation was selected as our main fence removal solution. SIREN is provided for comparison and is still a useful tool for controlled reconstruction settings. Our Pipeline can be visualized in Figure 3.2.

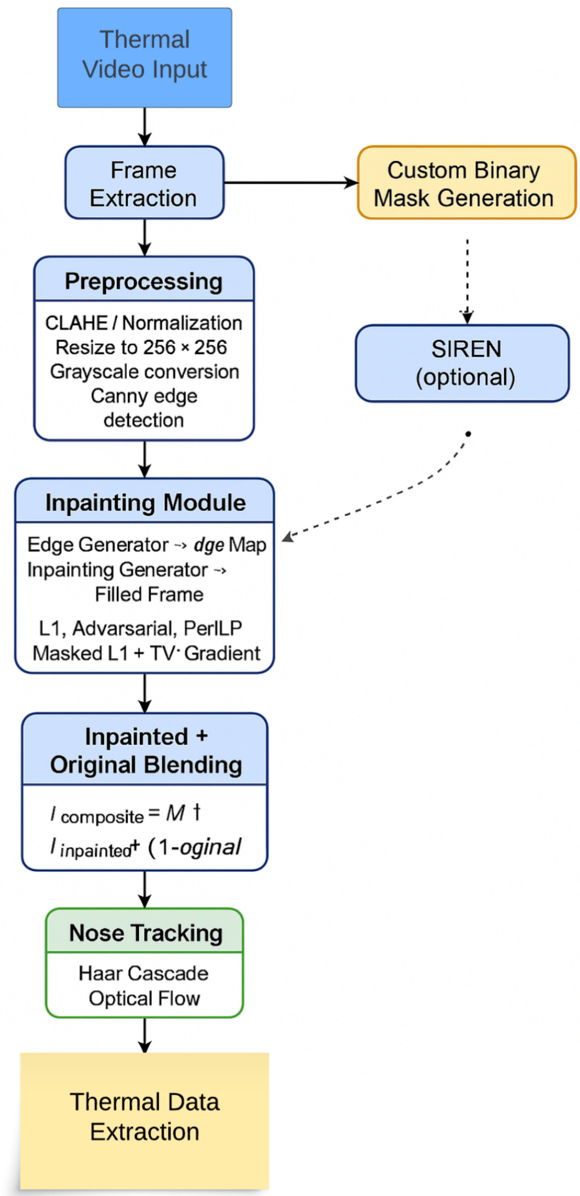

Figure 3.2: Thermal Inpainting and Tracking Pipeline

3.2.5 Feature Tracking Techniques

Accurately tracking the nasal region in thermal imagery is an essential portion of this study, as it authorizes longitudinal analysis of temperature dynamics in correlation to primate stress levels and behavior [13]. The lack thereof texture, contrast, and spatial resolution of thermal data makes it difficult for standard object detection pipelines to adapt to this domain. The focus of this section will be to outline multiple feature tracking techniques that were evaluated and implemented. Specifically focusing on their thermal robustness, efficiency, and compatibility with inpainted frames.

Haar Cascade Classifiers

The initial approach taken was to utilize the Haar Cascade Classifier, which is a machine learning-based approach using a set of rectangular Haar-like features and an AdaBoost-trained cascade encapsulating the detected regions in the middle of the rectangles. To identify the approximate facial region of the primates in thermal frames, a frontal-face Haar cascade that had been specially trained or modified was used. Assuming geometric priors based on inter-feature distances, the region was further refined after detection to identify the nasal subregion. While Haar cascades are quick computationally and simple to implement, its performance in thermal imagery is lackluster and constrained by low gradient magnitudes and subpar edge definition [23].

Edge-Based Localization

To highlight high-contrast regions within the primates' thermal face region, edge maps were derived from Sobel and Canny filters [33]. Due to cascade detection being unsuccessful, morphological operations and contour detection were employed on the edge maps, which gave a rough estimate for the nasal boundaries [34]. Thermal intensity priors were combined with the edge-based heuristics in cases where the nasal region maintains slightly cooler temperatures than the surrounding facial region under stress to refine the search space [16]. Nevertheless, there was a lack of temporal consistency in the application of these methods across frames, and they were used as supporting modules in early experimentation [14].

Comparative Analysis of Tracking Approaches

Four tracking methods were established and evaluated for their effectiveness on the different tracking strategies for thermal video tracking. The first method was the canonical approach based on the classical Lucas-Kanade optical flow algorithm, which estimates feature movements across successive frames using pixel intensity gradients. A weighted median variant of the method was developed and integrated subsequently to improve robustness by filtering outlier trajectories and recentering tracked points in ways that are brightness-weighted based on the distribution of valid key points.

A limited shift Lucas-Kanade version involved maximum displacement constraints, where any large positional jumps between frames were considered noise or occlusion-induced and would result in a sudden failure of tracking. Another morphological local maxima variant was used, where an area around the previous keypoint with a fixed window was selected for attention, and the new position would be the brightest thermal region within it. This approach simulated biological attention to thermal intensity, while also emulating a CSRT (Discriminative Correlation Filter with Channel and Spatial Reliability) tracker as an ordinary object tracker baseline [10].

All methods incorporated bounding boxes of a fixed size and rejection of outlier links to keep the tracking consistent and to aid in visualization. This strategy aligns with known thermal tracking and biological perception paradigms as explored in previous thermal imaging research, and general object tracking heuristics as described in real-world applications [18].

Tracking-by-Detection Pipeline

Temporal tracking is achieved with the utilization of Lucas-Kanade with weighted median approach feature detection method, which identifies the keypoint of the nasal region on a coordinate basis and tracks it across consecutive frames on spatial intensity gradients [31]. This optical flow technique is used in combination with Kalman filtering to smooth bounding box transitions and reduce the temporal change or jitter. The culmination of the two enabled continuous and quite decent tracking for robust temperature extraction from the nasal region of primates in the inpainted footage.

Table 3.1: Summary of Inpainting Techniques

Method	**Type**	**Strengths**	**Weaknesses**	**Final Use**
Telea	Traditional (PDE)	Fast, Simple	Poor edge preservation	No
Navier-Stokes	Traditional (PDE)	Smooth continuity	Weak structural reconstruction	No
Biharmonic	Traditional	Good smoothness	Poor on facial features	No
EdgeConnect	Deep Learning	Best structural restoration	Requires preprocessing	Final
SIREN	Implicit Function	High-frequency detail capture	High computational cost	Used for benchmark

3.2.6 Implementation Details

This section explains the methods utilized to compare and evaluate the inpainting methods and feature tracking techniques and how they were used:

- **Visual inspection:** The inpainting methods were manually reviewed to assess the overall preservation of detail including important regions such as the nasal area.
- **PSNR (Peak-Signal-to-Noise-Ratio), SSIM (Structural Similarity Index Measure), and CSNR (Channel-Signal-to-Noise-Ratio) Scores:** The inpainted regions needed to be compared to the ground truth image. To do this, an artificial fence was placed in a video frame where the primate's nasal region was visible.

PSNR score is a quantification of the maximum possible power of a signal (original image) and the power of corrupted noise (the inpainted regions) [40, 17]. It is computed using Mean Squared Error (MSE):

$$\mathrm{PSNR} = 10 \cdot \log_{10}\left(\frac{\mathrm{MAX}^2}{\mathrm{MSE}}\right)$$

where MAX relates to the maximum pixel value (e.g., 255 for 8-bit images).

SSIM measures the similarity between the original and inpainted image based on luminance, contrast, and structure. It is defined as:

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

where:

* μ_x, μ_y are means of the windows in the two images
* σ_x^2, σ_y^2 are variances
* σ_{xy} is the covariance between the two

CSNR compares the mean intensity of the inpainted region to the background:

$$\mathrm{CSNR} = \frac{\mu_{\text{inpaint}} - \mu_{\text{background}}}{\sigma_{\text{background}}}$$

– **Tracking accuracy:** The Lucas-Kanade tracker was used. Ground truth nasal center coordinates (x, y) were manually annotated. The tracker output (x', y') was compared using the Euclidean distance:

$$\mathrm{Error} = \sqrt{(x - x')^2 + (y - y')^2}$$

– **Inference Speed:** The time it took both EdgeConnect and SIREN to process all video frames was recorded, and FPS (frames per second) calculated as:

$$\mathrm{FPS} = \frac{\text{Total Number of Frames}}{\text{Total Inference Time (seconds)}}$$

– **Failure Case Review:** Instances where the inpainting introduced artifacts or distorted the image, disrupting the tracker.

- **Bounding Box Stability:** The movement of the tracked coordinates (x_i, y_i) across consecutive frames was measured as:

$$d_i = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}$$

To ensure a fair comparison, all methods were evaluated on the same video frames. Results are discussed in Chapter 4.

3.2.7 Experimental Setup and Evaluation

Our full pipeline implementation was conducted in Python 3.10, supporting a culmination of classic computer vision libraries and more modern deep learning frameworks. Our pipeline is constructed utilizing Google Colab Pro+, providing us access to NVIDIA Tesla T4 and A100 GPUs to accelerate our model training and inference speed.

Libraries and Frameworks

During this investigation, the following libraries are integrated:

- OpenCV for image manipulation, inpainting (Telea and Navier-Stokes), edge detection, thresholding, and morphological operations [29].
- NumPy and SciPy for kernel-based processing and numerical computations.
- PyTorch for defining, training, and inferring deep learning models, including SIREN and EdgeConnect implementations [35].
- Matplotlib and Seaborn for plot creation and visualization.

Modular Codebase Structure

Our codebase was arranged into modular components to streamline development and facilitate stage-by-stage experimentation across preprocessing, inpainting, and feature tracking. The major components include:

- `preprocessing.ipynb`: Handles preprocessing of thermal video frames, including CLAHE contrast enhancement, mask generation, and edge detection using OpenCV.
- `ModifiedEdgeConnect.ipynb`: Hosts the custom inpainting model based on EdgeConnect, adapted for thermal data. Integrates custom fence masks, thermal-specific normalization, and inference routines.
- `Initial_Siren_Inpainting_Experiments.ipynb`: Contains early SIREN-based coordinate inpainting tests using sinusoidal activations and coordinate-to-intensity function fitting on thermal frames.
- `facetracking_inpainting_video_experimentation.ipynb`: Implements Lucas-Kanade-based optical flow tracking after inpainting, and coordinates video frame overlay, bounding box management, and temperature extraction.
- `utils.ipynb`: Includes supporting functions such as thermal-to-RGB conversion, grayscale intensity normalization, mask application utilities, data format conversion, and visualization tools.

Other support scripts include utilities for image preprocessing, tensor normalization, data augmentation, edge extraction, and file management.

Mask Generation and Dataset Handling

The mask used for the particular dataset was created manually in Capcut by separating the fence structure from the background in a frame. The mask is refined using OpenCV drawing and morphological tools. During the training process, various mask sampling strategies were used, including fixed masks, random region masks, and custom binary masks aligned with the fence locations [40].

The thermal frames are preprocessed into tensor format with custom loaders and torchvision transformations. To conform to model-specific input requirements (such as $[-1, 1]$ for GANs), all data was normalized [38]. To enable equitable performance benchmarking, datasets were divided into training, validation, and test sets [14].

Output and Logging

Inpainted frames, feature tracking overlays, and extracted nasal temperature values were saved for evaluation and comparison. Frame-wise logging is integrated for reproducibility and constructed custom plotting functions for the comparison of inpainting and tracking techniques across methods.

Due to the modular nature of our system design, extensibility for future integration of further detection or inpainting models, automated annotation tools, and real-life streaming applications is ensured.

Chapter 4

Results and Discussion

4.1 Inpainting Performance Analysis: Qualitative and Quantitative Results

This section highlights and compares the performance of all the inpainting methods applied to the thermal video frames. The goal is to assess how each method performed in reconstructing the images' occluded regions, especially in the zones where important regions such as the nasal region were obstructed by the fencing. Quantitative (e.g., SSIM, PSNR, and tracking consistency) along with qualitative results are evaluated and analysed thoroughly.

Qualitative Results

The visual comparison of results can be seen in Figure 4.1. Preliminary experiments have been conducted utilizing 3 different classes of inpainting methods, including:

- Classical OpenCV-based approaches (Telea, Navier-Stokes, Biharmonic, Multi-Pass with CLAHE)
- Deep learning-based EdgeConnect
- Coordinate-based SIREN with sinusoidal activations

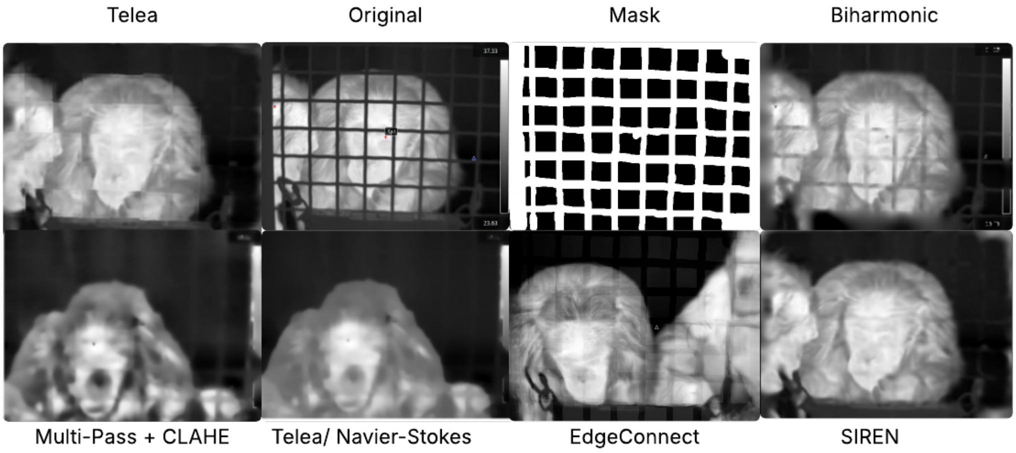

Figure 4.1: Inpainting Visual Comparison

- **Telea and Navier-Stokes:** These methods exhibited severe blurring and the structure of the images became simplified. Crucial features of the primates such as the nose or eyes were distorted or erased. Despite these methods' speed, they fail to restore the thermal textures properly.
- **Biharmonic Inpainting:** This method produced oversmoothed outputs with a lack of detail in the restoration of the nasal region. It also often failed to remove the fence and the lines were still often visible.
- **Telea:** Solely using Telea granted a decent reconstruction of the image with finer details still being visible. However, most of these details were impaired and distorted with temperature data being lost in the process.
- **Multi-Pass + CLAHE:** While the occlusion lines faded, they were still slightly visible and most of the reconstructed regions were incorrect. CLAHE restored some contrast and made visibility better, but temperature data was completely lost making it unusable for the task.
- **EdgeConnect:** Offered repairs that were aesthetically pleasing and kept the features of the face. While maintaining contextual thermal patterns, edge-guided prediction assisted in the reconstruction of obscured areas.
- **SIREN:** In tiny obstructed areas, it generated genuine pixel-level detail. However, using bigger or more asymmetrical masks reduced performance. When

facial topology could not be effectively interpreted by coordinate-based reconstruction, visual output deteriorated.

Overall, the results illustrate that both coordinate-based and deep-learning-based inpainting approaches have preserved the most detail in the inpainted regions of the image with a clear reconstruction of the nasal region. The tested classical inpainting methods, even with image enhancement, are unable to perform on the level of SIREN or EdgeConnect.

Quantitative Results

These metrics were computed by placing an artificial fence over the nasal region of the primate, then applying the two best performing inpainting methods on that artificial fence. This provided us with the ground truth values for the pixels on the nose. Only the reconstruction of the nasal region was taken into consideration using a bounding box when computing these results.

Table 4.1: PSNR, SSIM, and CSNR results for EdgeConnect and SIREN

Method	**PSNR (dB)**	**SSIM**	**CSNR**
EdgeConnect	35.44	0.9835	25.1
SIREN	34.98	0.9809	24.8

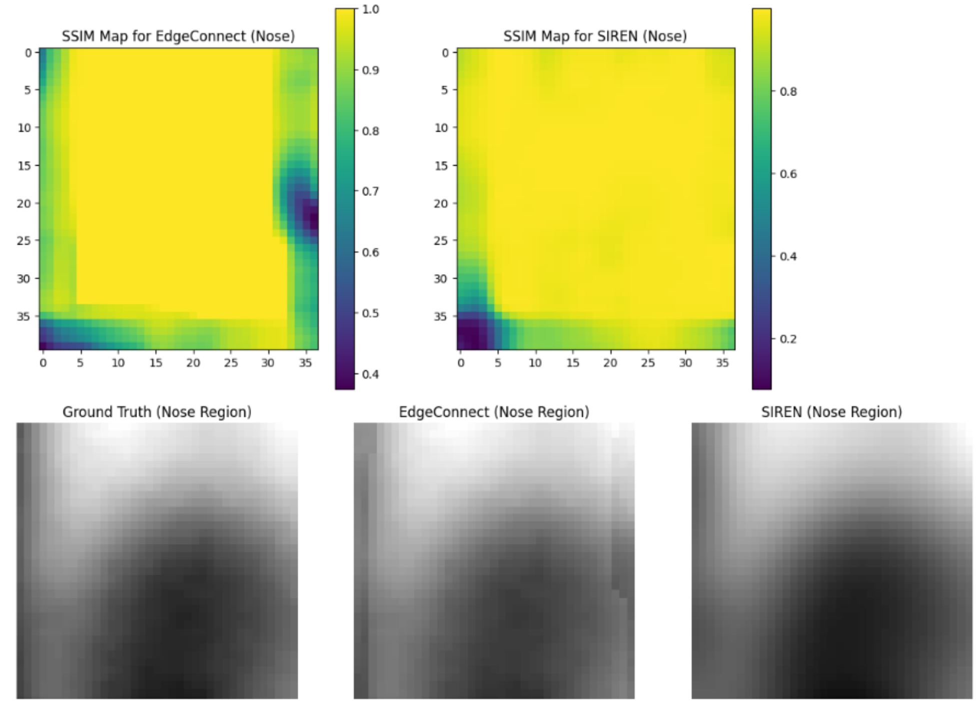

Figure 4.2: PSNR, SSIM, and CSNR visual results for EdgeConnect and SIREN

- **PSNR:** EdgeConnect demonstrated the highest reconstruction fidelity in pixel intensity in the nose region.
- **SSIM:** From the results it is apparent that the modified EdgeConnect model has superior perceptual similarity, meaning it has effectively preserved luminance information and structure.
- **CSNR:** EdgeConnect's outputs show that it has a smaller amount of contrast with the background when compared to SIREN which is important for edge-aware tracking.

While the margin between the results of the two models is small, EdgeConnect slightly outperforms SIREN and demonstrates promising results for thermal reconstruction.

Inference Time and Computational Performance

The average inference time per frame was recorded for both systems to assess the viability of practical deployment.

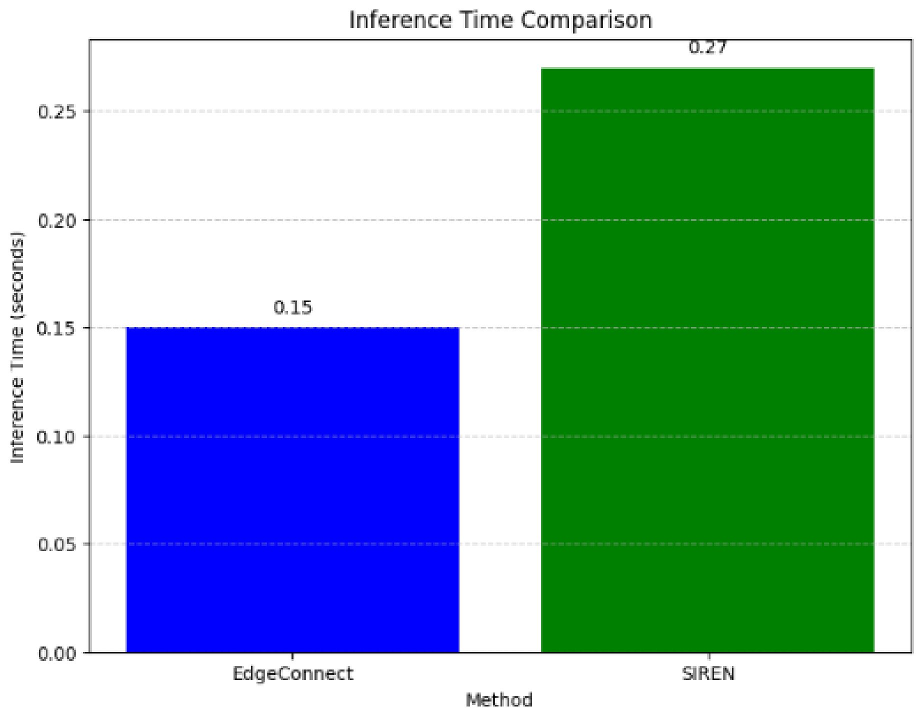

Figure 4.3: Inference Time Comparison

EdgeConnect has a 0.15 second time per frame rate of inpainting and SIREN has a 0.27 second time. From Figure 4.3 it is apparent that the run time for EdgeConnect is faster, meaning the model is faster computationally. SIREN, while promising in fidelity, has a longer processing time when compared to EdgeConnect.

4.1.1 Feature Tracking Evaluation

This section outlines the initial evaluation of the different feature tracking system's performance and how the best performing of these methods functioned following the inpainting process and fence removal. The main goal is to assess how different methods of inpainting can influence the precision and reliability of feature tracking for the nasal region of primates in thermal video sequences.

The 4 types of tracking approaches were tested on the same footage of SIREN INPAINTED FRAMES. Results of qualitative evaluation revealed that among the four variants of Lucas-Kanade assessed: the weighted median variant excelled in nose tracking reliability almost always together with inpainting; original Lucas Kanade method failed to detect the nasal region of primate completely; failure of the limited-shift variant was encountered while tracking the nose due to strict displacement constraints; while failure of the local maxima due to morphological reference task of localizing the nose was carried by low thermal contrast in peripheral regions; whereas CSRT ultimately failed, as it could not initiate any meaningful tracking on low-contrast thermal input. This shows the irrelevance of the CSRT in this particular context. Since the weighted median approach with Lucas Kanade had the most successful results, it is implemented in the pipeline and Inpainted frames.

Outlier removal and bounding box constraints were introduced as means for enhancing robustness; the former served to eliminate rogue keypoints. Bounding boxes (fixed in size) ensured that the output video maintained visual coherence, although they did not suffice to alleviate the underlying tracking failures present in some methods.

Figure 4.4: Example of Morphological Local Maximum Approach To Track Nasal Region

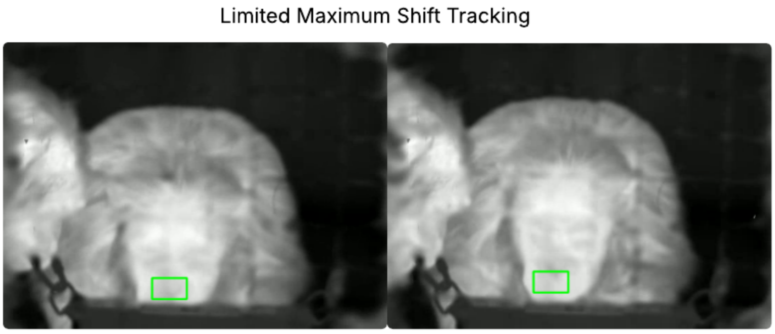

Figure 4.5: Example of Limit Maximum Shift Tracking Moving Off ROI

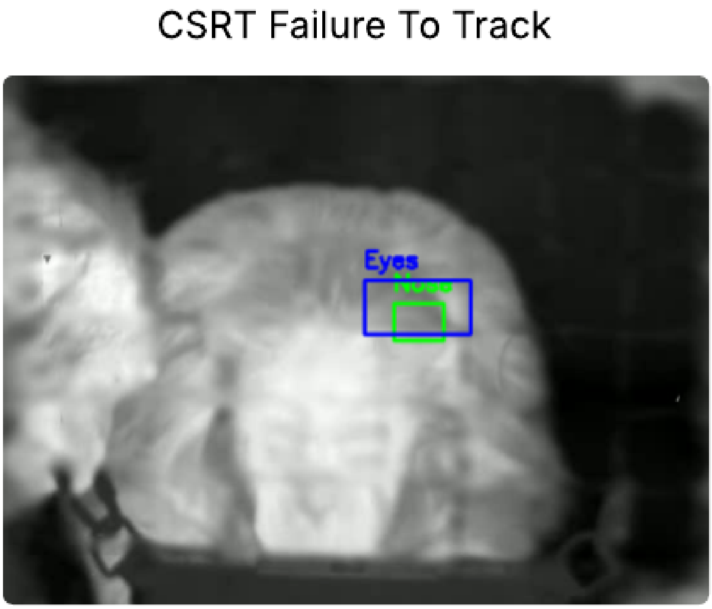

Figure 4.6: Example of CSRT Failure to Detect Nasal Region

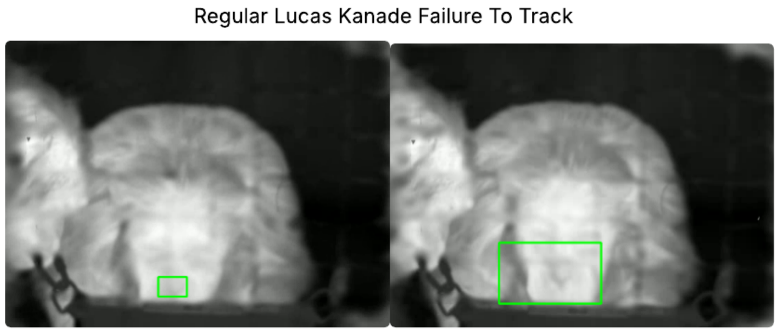

Figure 4.7: Base Lucas Kanade Losing Detection of Nasal Region Over Frames

Table 4.2: Comparison of Thermal Nose Tracking Methods

Method	Tracked Nose?	Robustness	Notes
Lucas–Kanade (base)	no	Very Low	Detection Fall Off
Weighted Median LK	yes	High	Best performer
Limited Shift LK	no	Low	Lost track due to shift constraint
Morphological Local Maximum	no	Very Low	Failed to track nose after identification
CSRT	no	Very Low	Could not track at all

Pre-Inpainting Tracking Performance

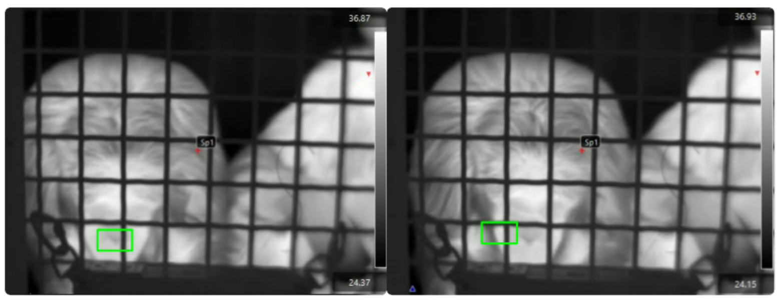

Figure 4.8: Pre-Inpainting Tracking Fail Example

The Lucas-Kanade feature detector was initially applied to the raw occluded thermal frames to test its tracking capabilities. Without any inpainting, the tracker frequently identified the lines of the fence as the tracked facial feature, often drifting away from the nasal region of the primate. This is due to the high-contrast patterns of the fence that is also on the nasal area. An example of this can be seen in Figure ??. Frequent identity switches coupled with keypoint drift are observed.

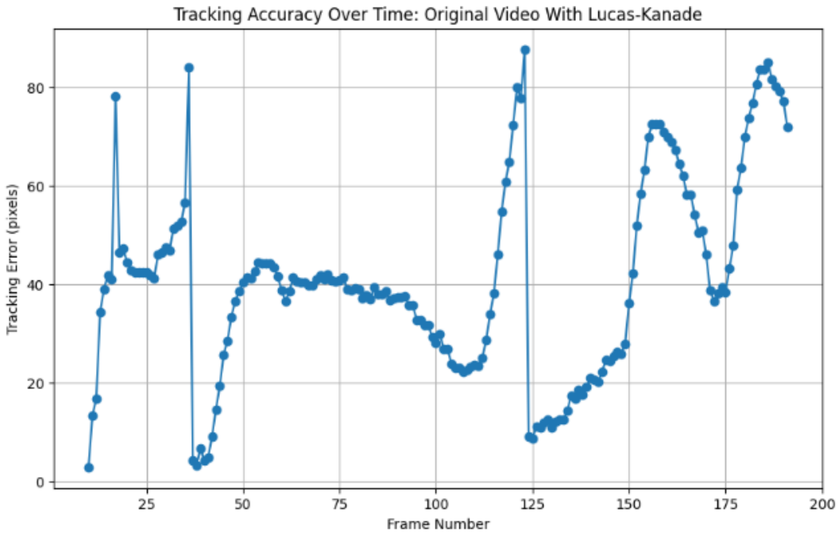

Figure 4.9: Original Video Tracking Accuracy

Through inspecting the results, there is a mean tracking error of 40.59 pixels, which is poor. Figure 4.9 visualizes the instability in the tracking results through erratic keypoint movements. Additionally, there is a clear amount of drift in tracking in frames 175 to 200 due to the nose being hidden in those frames, resulting in the Lucas-Kanade tracker tracking incorrectly.

Post-Inpainting Tracking Performance

The same tracking method was implemented post-inpainting to measure the improvement in feature retention and tracking smoothness once the occlusions had been removed.

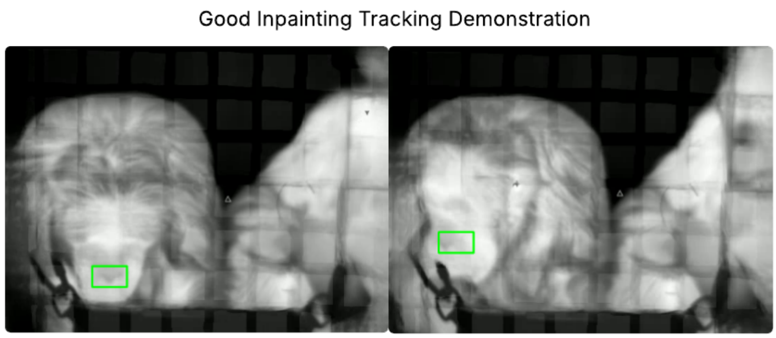

Figure 4.10: Successful Tracking Demonstration On Inpainted Footage

Figure 4.10 demonstrates successful Lucas-Kanade tracking once inpainting has been performed on fast primate movement.

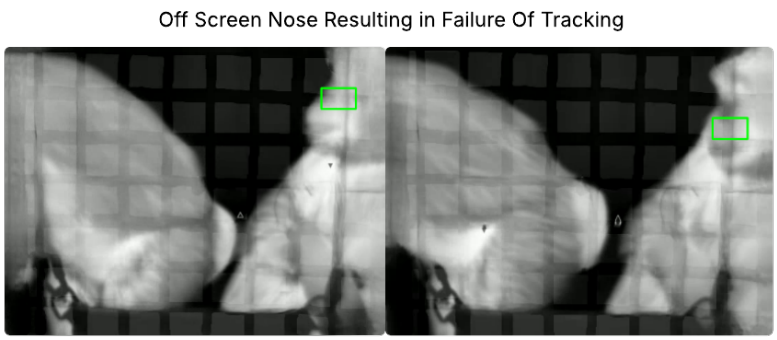

Figure 4.11: Example of a Failure to Track on Inpainted Footage

Figure 4.11 illustrates failure to track the nasal region once the primate turns around and the region of interest is no longer visible.

The mean tracking error once inpainting had been performed is 18.96 pixels, which is significantly better than when applied to the original video. Figure 4.5 illustrates the results over the frames and it is apparent that the tracking is much more consistent over the same footage once inpainting with EdgeConnect is performed.

This signifies a large difference between pre-inpainting and post-inpainting tracking performance. Post-EdgeConnect frames show consistent nose localization with minimal drift. Kalman filtering further stabilized positional noise and reduced outliers.

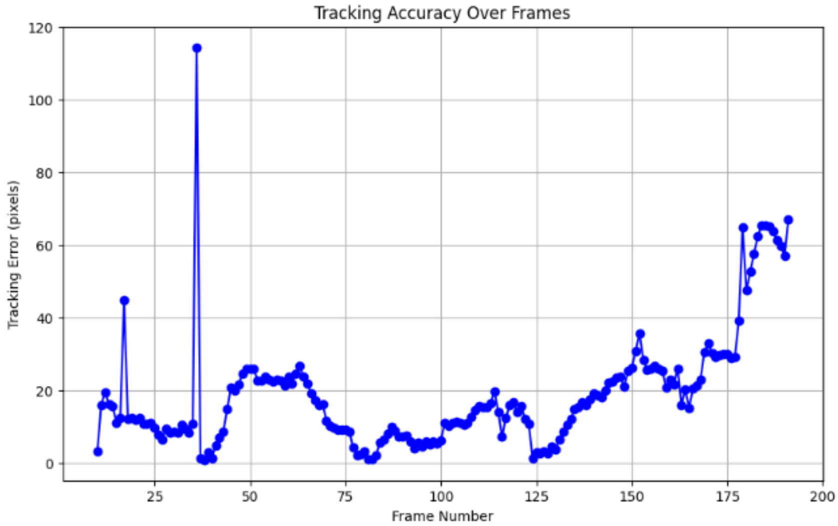

Figure 4.12: Inpainted Tracking Accuracy

Temperature Curve Consistency

The curve of temperature was extracted by utilizing a 3x3 kernel window around the keypoint of the tracked nose in each frame for a section of video. Without the application of inpainting, inaccurate position of the keypoint and signal loss occurred, resulting in abrupt spikes in the thermal readings. Figure 4.12 visualizes thermal readings over time once EdgeConnect had been applied. There is a smoother, physiologically interpretable signal. Higher pixel values illustrate hotter temperatures and lower ones represent cooler temperatures. The overall intensity range of the readings are from 160 to 210, indicating moderate variation in the nose's brightness over time. This means there is either a moderate shift in temperature or environmental factors have caused a shift in brightness over time.

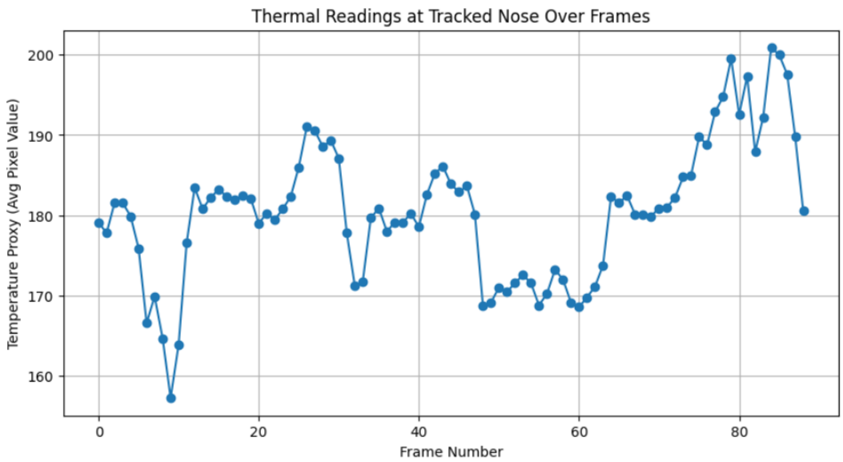

Figure 4.13: Temperature Curve at Tracked Nasal Region Over Inpainted Frames

From the thermal video intensity curve from the tracking applied to the original video demonstrated in Figure 4.14, it is apparent that the tracking gets stuck on the fence. This is due to the large bounces in temperature. Since the heat signature of the fence is colder, the tracking method refuses to shift on to it, causing it to be stuck on either higher heat intensity ranges or, as seen towards the latter half of the frames, onto colder intensity ranges.

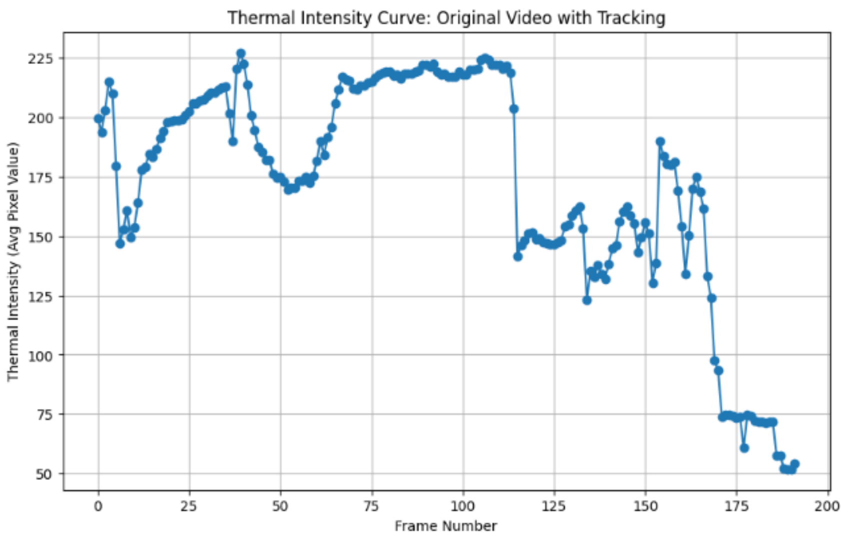

Figure 4.14: Temperature Curve at Tracked Nasal Region Over Original Frames

Bounding Box Stability

The displacement (y-axis) for the majority of tracked frames is 0, meaning that the bounding box is not moving much between the consecutive frames. This demonstrates that the tracker being used is stable for the large portion of the video.

There are a few jumps in the results; however, this stems from the footage changing halfway through and doesn't mean that there is an issue present in the tracking, as two different portions of footage were combined to further test the tracker's capabilities.

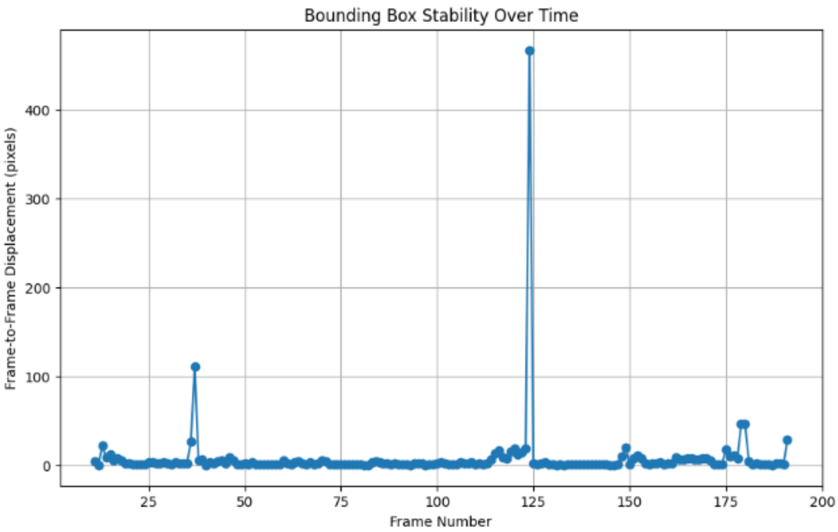

Figure 4.15: Bounding Box Stability Inpainted Tracking

CSNR Score

The CSNR score of the EdgeConnect inpainted video ranges from 0.225 to 0.375. This indicates that the fence region is neither drastically brighter nor darker than the background. This means that the inpainted regions do not vary contextually to the rest of the frames, staying in a narrow band of contrast. The CSNR score in a portion of the video frames is visualized in Figure 4.11.

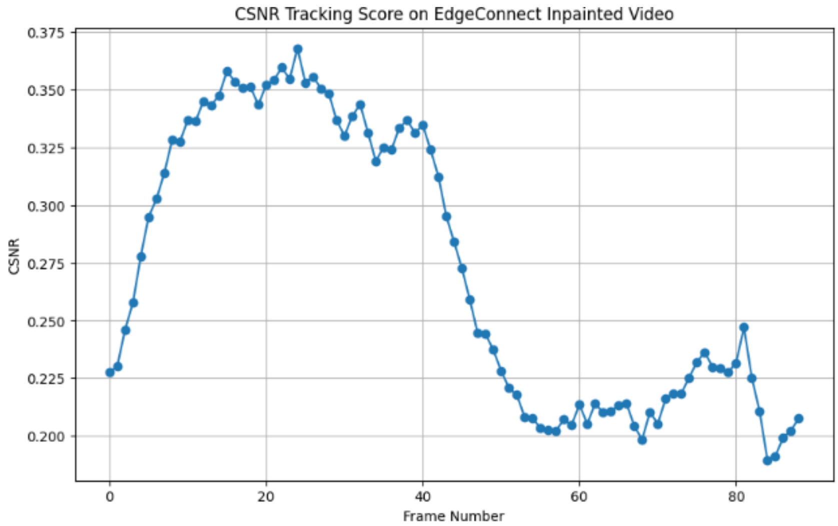

Figure 4.16: CSNR Tracking Score

4.2 Comparative Discussion

This portion will provide a comparative analysis of the obtained results from the inpainting and feature-tracking experimental stages. Our objective is to demonstrate how each method's architectural design and operational characteristics can affect the outcome of performance in the realm of thermal facial analysis.

4.2.1 Inpainting Quality vs. Tracking Stability

Hypothesis Evaluation

This study hypothesized that the implementation of inpainting for occlusion removal will improve nose tracking and temperature extraction capabilities. All the results that have been evaluated support this hypothesis proving it to be true. The restoration capabilities of EdgeConnect yielded to be the best with the highest restoration fidelity and therefore resulting in the most stable tracking trajectories, and thermal curves.

Model Specific Observations

The strong correlation between inpainting quality and tracking accuracy can be observed throughout all the experimentation. EdgeConnect which is a dedicated edge generation module, proved to have the best output results over all thermal frames. Its results were both the most semantically coherent and structurally consistent when reconstructing the frames. In turn, this provided the feature detecting methods tested to have more precise detection and improved temporal continuity. However, despite their computing efficiency, traditional techniques like Telea and Navier-Stokes were unable to accurately infer the face's missing structure. Particularly when the nasal region was partially or completely masked before inpainting, their inpainted outputs frequently showed flattened or blurred temperature distributions, resulting in unstable or inaccurate feature detections.

The coordinate-based implicit representation demonstrated by SIREN achieved high reconstruction fidelity in localized regions. After 500 epochs per frame, it was able to restore the primate features from behind the fence, which aided detection in static frames. However, the absence of generalization across large occlusion in combination with its computational cost minimized its applicability to sequence-level tracking.

4.2.2 Comparative Discussion: Key Observations, Trade-offs Between Methods

This portion will provide a comparative analysis of the obtained results from the inpainting and feature-tracking experimental stages. Our objective is to demonstrate how each method's architectural design and operational characteristics can affect the outcome of performance in the realm of thermal facial analysis.

Inpainting Quality vs. Tracking Stability

Hypothesis Evaluation

This study hypothesized that the implementation of inpainting for occlusion removal will improve nose tracking and temperature extraction capabilities. All the results that have been evaluated support this hypothesis proving it to be true. The restoration capabilities of EdgeConnect yielded to be the best with the highest restoration fidelity and therefore resulting in the most stable tracking trajectories, and thermal curves.

Model Specific Observations

The strong correlation between inpainting quality and tracking accuracy can be observed throughout all the experimentation. EdgeConnect which is a dedicated edge generation module, proved to have the best output results over all thermal frames. Its results were both the most semantically coherent and structurally consistent when reconstructing the frames. In turn, this provided the feature detecting methods tested to have more precise detection and improved temporal continuity. However, despite their computing efficiency, traditional techniques like Telea and Navier-Stokes were unable to accurately infer the face's missing structure. Particularly when the nasal region was partially or completely masked before inpainting, their inpainted outputs frequently showed flattened or blurred temperature distributions, resulting in unstable or inaccurate feature detections.

The coordinate-based implicit representation demonstrated by SIREN achieved high reconstruction fidelity in localized regions. After 500 epochs per frame, it was able to restore the primate features from behind the fence, which aided detection in static frames. However, the absence of generalization across large occlusion in combination with its computational cost minimized its applicability to sequence-level tracking.

Mask Quality and Alignment

The inpainting output is significantly influenced depending on the quality and alignment of the input mask. Therefore, a custom-generated mask that is tailored to the fence geometry will outperform any mask that is created using computer vision techniques. The inclusion of a custom mask led to higher reconstruction consistency. The design of the mask was an essential component for SIREN, which is prone to any mistakes in the spatial alignment process due to its regression-based framework.

Well-aligned masks improved inpainting quality and, consequently, tracking robustness for EdgeConnect by enabling the edge prediction module to maintain the contextual structure surrounding the occlusion. In contrast, tracking performance deteriorated, and boundary artifacts were introduced by misaligned or weakly specified masks.

Pipeline Trade-offs

From the perspective of creating our pipeline, EdgeConnect offered the best trade-off between its reconstruction accuracy, automatic nature, and computational efficiency. SIREN falls lackluster in comparison when it comes to reconstruction detail at the pixel level, and its high training and inference cost limits its scalability. While classical methods of inpainting work quickly, they are not viable at reconstructing detailed thermal facial analysis in constrained environments.

Overall, the findings are consistent with the theory that feature tracking performance in thermal video is enhanced by deep learning-based inpainting. The efficiency of the entire processing pipeline is determined by the interaction of inpainting fidelity, mask precision, and detection method selection. To generalize these results, rigorous statistical examination across other datasets will be a part of future trials.

4.3 Limitations and Challenges

Although promising results were found from our experimentation and pipeline, several challenges and limitations arose throughout the duration of this study, especially in the domains of thermal data handling, model generalizability, and real-world deployment.

4.3.1 Thermal Data Constraints

There is an absence of rich texture and color cues seen in the thermal video and available in visible-spectrum imagery. This reduces the discriminative power of traditional detection methods and increases the disruption of pixel intensity continuity caused by high-frequency occlusion, such as fencing [16]. Furthermore, when recognizing minor facial features like the nose, thermal resolution is often lower than RGB, adding to the spatial ambiguity.

The dataset provided for this project had a large number of thermal video sequences, providing a diverse array of thermal video footage for model training and evaluation. However, while the dataset was diverse, variations in the environment, such as an unstable camera or ambient temperature shifts and the primates behaving differently each time, still posed many challenges in ensuring the creation of a robust model across all scenarios.

4.3.2 Inpainting Model Limitations

Models that are based on deep learning, such as EdgeConnect, demonstrate strong performance but require more preprocessing and tuning in order to adapt to thermal imagery [26]. The original EdgeConnect model used as a foundation was trained on RGB data and had to be adapted for grayscale thermal input. Although this worked well, additional benchmarking and fine-tuning were made more difficult by the absence of extensive, annotated thermal inpainting datasets.

Despite its ability to rebuild pixels accurately, SIREN was computationally demanding and unsuitable for high-end processing. Furthermore, the sensitivity of the system to mask precision causes it to be less robust under conditions where the mask can vary, which limits its utility for real-time applications. Fence masks were difficult to create, implement, and align, which affected reconstructions from inpainting even if a small amount of pixels were not taken into consideration. Universal mask generalization was challenging due to pose and lighting diversity across datasets.

4.3.3 Feature Detection and Tracking Challenges

The detection of nasal features was hampered by occlusions in the majority of video frames. There was also low thermal contrast and shifting facial orientations. The efficacy of edge-based detectors and Haar cascades varied from frame to frame, especially when inpainted areas lacked adequate structural coherence [16]. Although they are still being developed for this project, deep learning-based detectors are anticipated to perform better than traditional methods. In certain extreme occlusions, keypoint drift occurred, necessitating median filtering and optical flow for smoothing.

Due to transient occlusions and detection noise, tracking-by-detection pipelines often remain vulnerable. Current outputs from the tracking system can exhibit jitter or dropouts in sequences where there is rapid motion in the video or imperfect reconstructions, and future integration of Kalman filtering and temporal smoothing is planned. Once any primate turned around, the tracking method failed to detect the nasal region in following frames. This is a significant challenge to tackle, as primates move constantly and can often turn around in video footage taken from one angle.

4.3.4 Evaluation and Benchmarking

The absence of ground truth annotations made it difficult to perform quantitative evaluation for inpainted and tracked nasal regions, and to solve this, an artificial fence had to be made and ground truth values had to be identified per frame. While metrics including PSNR, CSNR, and SSIM proved informative, they did not capture the full semantic reconstruction quality in thermal contexts. These drawbacks show that for reliable implementation, more thermal-specific structures, more thorough thermal datasets, and temporal tracking models are required [40]. They also help determine future research directions and ways to improve the current system.

Chapter 5

Conclusion and Future Work

5.1 Summary of Findings

The primary goal of this investigation is to build a non-invasive pipeline for thermal facial feature tracking, temperature extraction from key regions of interest and obstruction removal for occlusions such as fences. A comparison of deep learning-based and traditional inpainting techniques was conducted as part of the study, and their effects on the precision and consistency of nasal feature tracking were then assessed. Many techniques were evaluated and implemented including classical algorithms (Telea, Navier-Stokes, Biharmonic), a deep learning-based EdgeConnect pipeline, and coordinate-based SIREN models with some working better than others.

– **Inpainting Performance:** The modified deep learning-based EdgeConnect model outperformed the classical methods tested (Telea and Navier-Stokes) by having a better quality of reconstruction and doing so semantically, specifically in the facial region of the primates, while preserving thermal consistency across occluded regions [26]. The SIREN-based inpainting method was promising and demonstrated high local fidelity compared to classic methods but fell short of the EdgeConnect model and was less scalable for real-time applications.

– **Feature Tracking Accuracy:** The results for nasal tracking improved substantially after the inpainting process had taken place. Specifically, once inpainting was performed with the EdgeConnect model, occlusion artifacts were minimized, resulting in better tracking outcomes. Haar cascades classifier proved ineffective on non-human subjects and was more insatiable in the thermal frames. The Edge-based methods had partial improvement, but much more refinement is required to enhance temporal coherence. Out of the tested tracking methods, the most success is found in the weighted median Lucas Kanade approach.

- **Mask Quality and Structural Alignment:** Custom binary masks encompassed every fence pixel efficiently, leading to higher inpainting quality than masks generated with hough-line transform methods or pre-defined block masks. The proper alignment of the mask played a crucial role in ensuring effective reconstruction and minimizing boundary artifacts.
- **Challenges and Limitations:** The impact of thermal resolution limitations on detection performance, inpainting failures in areas with little contextual information, and the requirement for temporally consistent tracking in dynamic sequences were among the main issues the study uncovered.

The findings in this study improve upon present automated thermal video analysis and establish a solid foundation for the automation of these technologies. The interplay between occlusion mitigation, quality of reconstruction, and feature tracking reliability is highlighted. The insights provided by this work supply useful direction for enhancements to be made in the future, such as the development of more advanced tracking mechanisms and domain-specific inpainting optimizations.

5.2 Potential Improvements

This study leaves several areas for improvement, though it demonstrates the effectiveness of various deep learning-based inpainting and tracking techniques. Once the limitations are discussed and addressed, the robustness and applicability of these technologies will be developed further for the proposed pipeline.

- **Enhanced Inpainting Tracking:** The utilized EdgeConnect model, while effective at inpainting tasks, was trained on a non-thermal dataset of images [26].

While it still is effective at inpainting thermal image frames, it would be more effective had it been trained on such data. Future work could involve fine-tuning the dataset on a large scale or annotating a thermal dataset to improve domain-specific generalizations.

- **Advanced Feature Tracking Models:** The reliability of detection in thermal frames may be increased with the use of deep learning-based object detection algorithms like YOLO or Faster R-CNN. Furthermore, temporal stability could be further improved by incorporating optical flow techniques or a hybrid CNN-transformer-based tracker [22, 37].
- **Refined Mask Generation:** While the custom mask utilized during experimentation proved successful, a way to automate this process would make the system easier to use in other applications, optimizing the pipeline [32]. This would reduce the reliance on having to create a custom mask for each set of thermal video data, as the position of the fencing will most likely be different each time.
- **Real-Time Processing Pipeline:** Both primary methods of EdgeConnect and SIREN are computationally expensive, making it difficult to run in real time. Future research should explore lightweight, real-time solutions to facilitate instant fence removal and tracking in an automatic fashion.
- **Comprehensive Quantitative Evaluation:** A more uniform evaluation of model performance might be possible with further benchmarking that uses quantitative measures like IoU (Intersection-over-Union) for tracking accuracy and perceptual loss-based assessments for inpainting fidelity.

5.3 Future Research Directions

The knowledge provided by this study lays the foundation for further research into thermal-imaging-based primate monitoring. Many key research directions surface from this work:

- **Physiological Monitoring Implementation:** At this point, the tracking system can be improved to add respiratory rate estimation from thermal video frames, which could provide additional monitoring benefits and further insights into primate health and well-being [16, 8].

- **Domain Adaptation:** Inpainting and Detection models may have the potential for better generalization to new datasets without the need for a lot of labeled training data if research is conducted on unsupervised or self-supervised learning techniques for domain adaptation [35].
- **Thermal and Visible Video Fusion:** Fusion of thermal with visible video or depth sensors can increase feature tracking capabilities and improve performance in challenging and uncontrolled environments.
- **Practical and Ethical Considerations:** Research into the ethical implications of automatic primate monitoring and making sure the system complies with conservation and wildlife welfare guides remains a critical aspect of future work [13].
- **Running in the Field:** Executing the study under real-world conditions is key. The system should be tested within dynamic thermal environments, moving animals, and changing scenes [24].

In aiming to achieve these aspects, the research that is yet to come could not only improve the scalability, accuracy, and utility of the non-invasive primate behavioral analysis based on the thermal imaging but also contribute to the enhancement of the non-invasive wildlife monitoring and protection.

Chapter 6

Bibliography

[1] Thermal monitoring in primates – dissertation. Liu.diva-portal.org.

[2] Use of thermal imaging in animal tracking. Citeseerx.

[3] M. Ahmad, S. Abbas, A. Fatima, G. F. Issa, T. M. Ghazal, and M. A. Khan. Deep transfer learning-based animal face identification model empowered with vision-based hybrid approach. *Applied Sciences*, 13(2), 2023.

[4] alex04072000. [cvpr 2020] learning to see through obstructions. GitHub, 2020.

[5] S. Broome. Going deeper than tracking: A survey of computer-vision based recognition of animal pain and emotions. ResearchGate, 2022.

[6] C. A. Brust, T. Burghardt, and D. W. Macdonald. Towards automated visual monitoring of individual animals in the wild. In *Proc. ICCV Workshops*, 2017.

[7] T. Burghardt. Real-time face detection and tracking of animals. ResearchGate, 2010.

[8] P. Drummond, P. Tsiamyrtzis, and I. Pavlidis. Fast by nature—how stress patterns define human experience and performance in dexterous tasks. *Scientific Reports*, 2:305–306, 2012.

[9] O. Elharrouss, N. Almaadeed, S. Al-Maadeed, and N. Akhtar. Image inpainting: A review. *Neural Process. Lett.*, 51:2007–2028, 2020.

[10] Sondre R. Grimsby. Thermal imaging in animal behavior research. Master's thesis, University of Stavanger, 2022. Accessed: 2024-04-10.

[11] Gtatters. Thermimagebash: Command line tools for converting flir thermal image files. GitHub.

[12] Y. Guan, Y. Lei, Y. Zhu, P. Dong, A. Huang, M. Xie, J. Mu, and Y. Zhang. Face recognition of a lorisidae species based on computer vision. *Global Ecology and Conservation*, 41, 2023.

[13] K. A. Herborn, P. Jerem, R. G. Nager, D. E. McKeegan, and D. J. McCafferty. Surface temperature elevated by chronic and intermittent stress. *Physiology Behavior*, 191:47–55, 2018.

[14] A. Horé and D. Ziou. Image quality metrics: Psnr vs. ssim. In *Proc. ICPR*, 2010.

[15] P. V. C. Hough. Use of the hough transformation to detect lines and curves in pictures. *Commun. ACM*, 15(1):11–15, 1972.

[16] S. Ioannou, H. Chotard, and M. Davila-Ross. No strings attached: physiological monitoring of rhesus monkeys (macaca mulatta) with thermal imaging. *Frontiers in Behavioral Neuroscience*, 9:160, 2015.

[17] B. F. Jones. A reappraisal of the use of infrared thermal image analysis in medicine. *IEEE Trans. Med. Imaging*, 17(6):1019–1027, 1998.

[18] Khwab Kalra. Object tracking, 2022. Accessed: 2024-04-10.

[19] KKAHATAPITIYA. Occlusion-removal/models/generative$_inpainting.GitHub$, 2020.

[20] H. S. Kühl and T. Burghardt. Animal biometrics: quantifying and detecting individual facial markings. *Trends Ecol. Evol.*, 28(7):432–441, 2013.

[21] G. Liu, F. A. Reda, K. J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro. Image inpainting for irregular holes using partial convolutions. In *Proc. ECCV*, 2018.

[22] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.*, 60(2):91–110, 2004.

[23] S. Mallat and S. Zhong. Characterization of signals from multiscale edges. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(7):710–732, 1992.

[24] T. Matsugu. Chimpanzee face recognition from videos in the wild using deep learning. *Sci. Adv.*, 5(5), 2019.

[25] K. Nakayama, S. Goto, K. Kuraoka, and K. Nakamura. Decrease in nasal temperature of rhesus monkeys (macaca mulatta) in negative emotional state. *Physiology Behavior*, 84(5):783–790, 2005.

[26] K. Nazeri, E. Ng, T. Joseph, F. Qureshi, and M. Ebrahimi. Structure-guided image inpainting using edge prediction. In *Proc. ICCV Workshops*, 2019.

[27] I. A. Nääs, R. G. Garcia, and F. R. Caldara. Infrared thermal image for assessing animal health and welfare. *J. Anim. Behav. Biometeorol.*, 2(3):66–72, 2020.

[28] J. J. Park, P. Florence, J. Straub, R. Newcombe, and S. Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proc. CVPR*, 2019.

[29] I. Pavlidis, N. L. Eberhardt, and J. A. Levine. Seeing through the face of deception. *Nature*, 415(6867):35, 2002.

[30] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. Zimmerman, and K. Zuiderveld. Adaptive histogram equalization and its variations. *Comput. Vis. Graph. Image Process.*, 39(3):355–368, 1987.

[31] X. Ren, J. Song, H. Ying, Y. Zhu, and X. Qiu. Robust nose detection and tracking using gentleboost and improved lucas-kanade optical flow algorithms. In *SpringerLink*, 1970.

[32] J. M. Saragih, S. Lucey, and J. F. Cohn. Deformable model fitting by regularized landmark mean-shift. *Int. J. Comput. Vis.*, 91(2):200–215, 2011.

[33] J. Shi and C. Tomasi. Good features to track. In *Proc. CVPR*, 1994.

[34] D. Sun, S. Roth, and M. J. Black. Secrets of optical flow estimation and their principles. In *Proc. CVPR*, 2010.

[35] M. Tancik, P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singh, R. Ramamoorthi, J. T. Barron, and R. Ng. Fourier features let networks learn high frequency functions in low dimensional domains. In *Proc. NeurIPS*, 2020.

[36] P. Tsiamyrtzis, J. Dowdall, D. Shastri, I. Pavlidis, and M. G. Frank. Imaging facial physiology for the detection of deceit. *Int. J. Comput. Vis.*, 71(2):197–214, 2006.

[37] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multi-scale structural similarity for image quality assessment. In *Proc. Asilomar Conf.*, 2003.

[38] J. Xie, L. Xu, and E. Chen. Image denoising and inpainting with deep neural networks. In *Proc. NIPS*, 2012.

[39] H. Yeganeh and Z. Wang. Objective quality assessment of tone-mapped images. *IEEE Trans. Image Process.*, 22(2):657–667, 2013.

[40] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang. Generative image inpainting with contextual attention. In *Proc. CVPR*, 2018.

[41] Z. Zhang and X. Wang. Face detection based on haar-like features and boosted classifier. *IEEE Trans. Syst., Man, Cybern., Part C (Appl. Rev)*, 36(4):789–798, 2006.

Appendix A: Personal System Hardware Specifications

System Information

- **Main Processor:** 11th Gen Intel(R) Core(TM) i7-11700F
- **Graphics:** RTX 3060 Ti
- **Installed RAM:** 16GB
- **System Version:** Windows 10

Python Environment and Dependencies

All the experiments in this project, including preprocessing, inpainting models, and facial tracking functionality, were implemented and executed using Python 3.10. The development and testing environment was primarily Google Colab Pro+ on NVIDIA Tesla T4 (16GB VRAM) and 24GB system RAM for efficient training and inference of deep learning models.

The following are notable libraries and frameworks used:

- NumPy 1.24.3: Efficient numerical array operations
- OpenCV 4.7.0: Preprocessing, edge detection, video handling, and conventional inpainting
- Matplotlib 3.7.1: Tracking visualizations, temperature plots, and mask generation

– PyTorch 2.0.1: Training and inference of SIREN and EdgeConnect models
– TorchVision 0.15.2: Dataset handling and model utilities
– Scikit-image 0.19.3: Image processing and evaluation metrics like SSIM, PSNR
– SciPy 1.10.1: Morphological operations and image filtering
– Albumentations 1.3.0: Applies data augmentation
– tqdm 4.65.0: For visualizing training progress
– Imageio 2.31.1: Reading and writing of multi-frame thermal .tif files
– FLIR Research Studio (external tool): Used for extracting raw thermal frames from .tiff videos

Mixed precision (fp16), wherever applicable, was used to speed up computation and save memory. The environment was initialized with a custom Colab notebook in which all dependencies were installed via pip.

Note: All main scripts, such as `edgeconnect.py`, `siren_train.py`, `pltfd.py`, etc., are modularized for reproducibility and testing purposes. The environment configuration file (`requirements.txt`) and some helper scripts (`utils.py`, `config.py`) are included in the project repository.

Appendix B: Code and Availability

The code and data utilized in this project can be found here: **GitHub Repository:** `https://github.com/ImBodrum/Primate-Thermal-Video-Modelling.git`

Appendix C: Supervisor Meeting Logs

Meeting Log 1: October 10, 2024

Names of team members present: Maksut

Meeting format: Online

Meeting coordinator: Dr Ivor Simpson

Issues discussed at this meeting

– Narrow down the topic of research and try to get a proper topic to do the project on. This is in order to maintain a focused approach with suggestions to concentrate on studying stress indicators in primates through thermal imaging of specific facial regions.

– Continue doing research. Any sort of upscaling is likely not required. Continue literature review on primates in wildlife environments.

– The nose and eye regions have the most temperature changes. Focus on these areas of study. These are primary data extraction targets.

– Try to focus on one thing at a time.

– Processing thermal data can be challenging. Dealing with occlusion and movement caused by movement or environment must be considered.

– Further explore software tools to access the data and convert the csq files. Worst case, screen recording must be done.

Date of next meeting: October 17, 2024

Meeting Log 2: October 17, 2024

Names of team members present: Maksut

Meeting format: Online

Meeting coordinator: Dr Ivor Simpson

Issues discussed at this meeting

- File Conversion - Converting the csq file to the required format. Best-case scenario is using a converter tool. If this fails, take a screen recording of the thermal video and turn it into a grayscale image for feature tracking.
- Explore different tracking techniques – Lucas-Kanade Optical Flow.
- Try and track the nose and extract temperature data across frames.
- A mask can be drawn and temperature can be extracted.
- Filters (Canny) can be used to clean up and preprocess the image for better tracking.
- Explore inpainting techniques to remove fence from video.

Date of next meeting: October 24, 2024

Meeting Log 3: October 24, 2024

Names of team members present: Maksut

Meeting format: Online

Meeting coordinator: Dr Ivor Simpson

Issues discussed at this meeting

- Basic tracking system developed. Results reviewed.
- Discussed the next steps on inpainting methods to remove the fence – including OpenCV Telea, Navier-Stokes, and possibly deep learning-based approaches.
- Thermal data preservation during fence removal is a key point of discussion.
- Write-up work and logging all changes was encouraged.
- Begin report document and fill out methods and meeting log.

Date of next meeting: November 5, 2024

Meeting Log 4: November 5, 2024

Names of team members present: Maksut

Meeting format: Online

Meeting coordinator: Dr Ivor Simpson

Issues discussed at this meeting

- Reviewed OpenCV inpainting results.
- Agreed on need for better inpainting fidelity – attempt EdgeConnect.
- Discussed custom mask drawing for fence area.
- Reviewed Haar cascade approach for face/nose tracking – results limited.
- Suggested more advanced models (if possible) or better feature extraction.
- Encouraged documenting failures and success with tracking.

Date of next meeting: November 15, 2024

Meeting Log 5: November 15, 2024

Names of team members present: Maksut

Meeting format: Online

Meeting coordinator: Dr Ivor Simpson

Issues discussed at this meeting

- EdgeConnect tested and showed strong results.
- Begin temperature extraction pipeline from nose.
- Kalman filter or smoothing for more stable tracking.
- Finalize preprocessing pipeline.
- Begin working on evaluation metrics (PSNR, SSIM).
- Start writing up experimental evaluation.

Date of next meeting: November 28, 2024

Meeting Log 6: November 28, 2024

Names of team members present: Maksut

Meeting format: Online

Meeting coordinator: Dr Ivor Simpson

Issues discussed at this meeting

- Finalised pipeline using EdgeConnect.
- SIREN implementation reviewed.
- Discussed comparison table of all inpainting methods and feature tracking techniques.
- Suggested to plot tracking accuracy and bounding box displacement.
- Discussed using artificial fence overlay to generate ground truth for inpainting metrics.
- Encouraged LaTeX write-up to begin and add graphs.

Date of next meeting: December 12, 2024

Meeting Log 7: December 12, 2024

Names of team members present: Maksut

Meeting format: Online

Meeting coordinator: Dr Ivor Simpson

Issues discussed at this meeting

- Reviewed evaluation metric plots – PSNR, SSIM, CSNR.
- Discussed difference in tracking performance pre- and post-inpainting.
- Suggested bounding box stability measurement and plotting thermal curve.
- Reviewed Lucas-Kanade tracker results.
- Kalman filter smoothing working well – keep as part of final system.

Date of next meeting: January 5, 2025

Meeting Log 8: January 5, 2025

Names of team members present: Maksut

Meeting format: Online

Meeting coordinator: Dr Ivor Simpson

Issues discussed at this meeting

- Supervisor reviewed current results and advised combining videos to show success and failure cases.

- Begin section on limitations of thermal video analysis and tracking.
- Reiterate importance of mask quality for SIREN inpainting.
- Draft Discussion section for how inpainting quality impacts tracking performance.

Date of next meeting: January 19, 2025

Meeting Log 9: January 19, 2025

Names of team members present: Maksut

Meeting format: Online

Meeting coordinator: Dr Ivor Simpson

Issues discussed at this meeting

- Discussed CSNR scoring and its interpretation.
- Reviewed failure cases and suggested additional figure for comparison.
- Started Conclusions section – summarise key findings.
- Discussed report structure and appendix layout.

Date of next meeting: February 2, 2025

Meeting Log 10: February 2, 2025

Names of team members present: Maksut

Meeting format: Online

Meeting coordinator: Dr Ivor Simpson

Issues discussed at this meeting

- Supervisor reviewed completed results section and graphs.
- Suggested corrections to SSIM equation format.
- Report graphs look good – improve captions for clarity.
- Go through References and fix formatting issues.
- Begin final proofreading.

Date of next meeting: February 18, 2025

Meeting Log 11: February 18, 2025

Names of team members present: Maksut

Meeting format: Online

Meeting coordinator: Dr Ivor Simpson

Issues discussed at this meeting

- Final review of dissertation before submission.
- Supervisor approved structure and content.
- Ensure all meeting logs are added to appendix.
- Confirm all equations converted to LaTeX.
- Export and compile Overleaf version.