Received: 23 June 2022 | Revised: 6 February 2023 | Accepted: 31 May 2023

DOI: 10.1002/rob.22219

RESEARCH ARTICLE

WILEY

Neural network-based optical flow versus traditional optical flow techniques with thermal aerial imaging in real-world settings

Tran Xuan Bach Nguyen[1] | **Kent Rosser**[1,2] | **Asanka Perera**[3] | **Philip Moss**[2] | **Javaan Chahl**[1]

[1]University of South Australia, Mawson Lake, South Australia, Australia

[2]Defence Science and Technology Group, Edinburgh, South Australia, Australia

[3]University of New South Wales, Canberra, Australian Capital Territory, Australia

Correspondence
Tran Xuan Bach Nguyen, University of South Australia, Mawson Lake, South Australia 5095, Australia.
Email: nguty219@mymail.unisa.edu.au

Funding information
Australian Government Research Training Program (RTP) Scholarship

Abstract

The study explores the feasibility of optical flow-based neural network from real-world thermal aerial imagery. While traditional optical flow techniques have shown adequate performance, sparse techniques do not work well during cold-soaked low-contrast conditions, and dense algorithms are more accurate in low-contrast conditions but suffer from the aperture problem in some scenes. On the other hand, optical flow from convolutional neural networks has demonstrated good performance with strong generalization from several synthetic public data set benchmarks. Ground truth was generated from real-world thermal data estimated with traditional dense optical flow techniques. The state-of-the-art Recurrent All-Pairs Field Transform for the Optical Flow model was trained with both color synthetic data and the captured real-world thermal data across various thermal contrast conditions. The results showed strong performance of the deep-learning network against established sparse and dense optical flow techniques in various environments and weather conditions, at the cost of higher computational demand.

KEYWORDS

deep learning, LWIR, navigation, optical flow, thermal imaging, UAVs

1 | INTRODUCTION

Autonomous navigation is a crucial requirement for unmanned aerial vehicles (UAVs) to be integrated more deeply into the economy and society. Currently, UAVs rely on Global Navigation Satellite Systems (GNSS) for many navigation applications. However, despite being cost effective and widely available, GNSS can be unreliable in built-up urban areas, or in environments with lots of vegetation, and it is not available underground. Furthermore, GNSS does not intrinsically provide any environment or obstacle-sensing capabilities, making the solution unreliable in dynamic environments.

Many researchers have attempted to solve this problem using vision-based systems with promising results (Chahl et al., 2004a; Conroy et al., 2009; Lu et al., 2018; Miller et al., 2018; Popov et al., 2016; Rosser & Chahl, 2019). Unlike GNSS which is based on off-board navigation signals, vision-based systems can provide information in real-time about obstacles in dynamic environments. Vision-based systems do not rely on artificial sources of information which makes them more resistant to jamming of navigation signals

(Nguyen et al., 2021). On the other hand, there are challenges to use a vision-based system on UAVs for guidance and control due to spatial and temporal limitations of sensors, such as motion blur caused by high angular rates of the platform and low rates of ground movement at high altitude (Gan & Sukkarieh, 2011; Ross et al., 2013; Tran et al., 2022, 2021; Zuo et al., 2022). Furthermore, due to the high number of DOF of UAVs, there are also problems in dealing with various viewing angles transformed in roll, pitch, and yaw which may affect captured images (Perera et al., 2018; Tran et al., 2020).

Since the introduction of AlexNet (Krizhevsky et al., 2012) in 2012, many researchers have applied neural networks to solve many computer vision tasks with great success (Ciregan et al., 2012). Some neural networks even surpass human's benchmark in the image classification ImageNet competition, such as Inception v4 (Szegedy et al., 2016) and ResNet (He et al., 2016). Some of the recent applications of neural networks in the literature include object detection (Liu, Ouyang, et al., 2020; Pathak et al., 2018; Wu et al., 2020; Zhao & Zheng, 2019), image segmentation (Ghosh et al., 2019; Goceri, 2019; Guo et al., 2019; Minaee et al., 2021), autonomous driving (Feng et al., 2018; Kocić et al., 2019; Wu et al., 2017; Zhang et al., 2022), recognition of human activities and emotions (Negi, Kumar, Chaudhari, et al., 2021; Sharma, Kumar, & Kumar, 2017; Sharma, Kumar, & Singh, 2017; Vijayvergia & Kumar, 2018), human's disease's prediction (Alok et al., 2021; Chowdhury & Iraj, 2020; Haq et al., 2018; Kumar et al., 2020; Reddy et al., 2021), diseases detection in agriculture (Jasim & Al-Tuwaijari, 2020; Lu et al., 2017; Negi & Kumar, 2021a; Negi, Kumar, and Chauhan, 2021), real-time face mask detection (Kodali & Dhanekula, 2021; Militante & Dionisio, 2020; Negi et al., 2020; Negi & Kumar, 2021b; Sethi et al., 2021).

Optical flow is defined as motions of brightness patterns across two consecutive frames, caused by movement of objects or the camera, or both (Horn & Schunck, 1981). Optical flow in flight control was inspired to some extent by observations of flying insects (Chahl & Mizutani, 2010). For example, honeybees rely on optical flow for grazing landings (Chahl et al., 2004b; Srinivasan et al., 2000) and obstacle sensing (Srinivasan, 2010). Furthermore, in the field of computer vision, optical flow can be used for tracking (Li et al., 2016; Rosser & Chahl, 2019), environmental reconstruction (Godard et al., 2015), autonomous driving (Shah & Xuezhi, 2021) or face recognition (Ranftl et al., 2017). Optical flow has been proposed to have roles in medical imagery, including medical image registration (Hermann & Werner, 2013), breast tumor analysis (Abdel-Nasser et al., 2017), and diagnosis of bladder cancer (Weibel et al., 2012).

Previously, the techniques to determine optical flow have been dominated by traditional methods of spatiotemporal image processing, such as the Horn and Shuck algorithm (Horn & Schunck, 1981) and the Farnedback technique (Farnebäck, 2003). Another class of optical flow techniques is typified by the gradient-based Lucas–Kanade (LK) techniques (Baker & Matthews, 2004) and their optimization (Bouguet, 2001; Sharmin & Brad, 2012), correlation and block matching (Dabov et al., 2006), and the image interpolation technique (Srinivasan, 1994).

With the ease of access to more powerful graphic processing units (GPUs), scientists have been experimenting with optical flow implementations based on deep learning concepts with great success. FlowNet (Dosovitskiy et al., 2015) was the first model but its performance was worse than traditional algorithms. FlowNet2 (Ilg et al., 2017) was an improved version by stacking multiple FlowNet layers, which greatly increased the performance and outperformed most traditional techniques, but is very expensive to run.

The next-generation models have better efficacy and smaller memory footprint by borrowing some of the concepts from traditional techniques. For example, SpyNet (Ranjan & Black, 2017) relies on a coarse-to-fine and spatial-pyramid to deal with small and larger movements. LiteFlowNet (Hui et al., 2018), LiteFlowNet2 (Hui et al., 2020), and LiteFlowNet3 (Hui & Loy, 2020) rely on an inference approach at each pyramid level through a lightweight cascaded network and a feature-driven local convolution, with later models are more refined and are smaller and have better efficacy than the former. Additionally, pyramid, warping, and cost (PWC)-Net (Sun et al., 2018) utilizes pyramidal processing, warping, and the use of a cost volume which results in 17 times smaller size compared with FlowNet2. Most recently, the Recurrent All-Pairs Field Transform (RAFT) (Teed & Deng, 2020b) and its lighter version, RAFT-s, were introduced that achieved state-of-the-art efficacy while also having one of the lowest memory requirements. The RAFT models were inspired by optimization-based approaches from traditional optical flow techniques.

In the field of autonomous navigation, the vast majority of attempts so far have only been done with optical visual reflected light sensors (Nguyen et al., 2021) during daytime with abundant natural lights, which leads to a substantial gap in knowledge about approaches for night operation. Given night and use of thermal payloads typically accounts for half of long missions, it is necessary to understand the performance and behavior of the night sensor suite. There are many reasons why it is more difficult to test a sensor suite designed for night, including higher cost for thermal sensors, the difficulty to operate the aircraft after dark due to regulation restriction and challenges of launching and recovering small aircraft at night.

The paper is organized into nine sections. Section 2 summarizes related works with thermal sensors in the field of robotic navigation. Section 3 outlines our motivations and the contributions made in this paper. Section 4 presents the theory behind optical flow including optical flow equation and algorithms to estimate the optical flow field. Section 5 outlines the construction of the payload including components, the configuration of long wave infrared (LWIR) thermal sensor, a rescaling technique that satisfies the constraints imposed when computing optical flow and the conditions of the environment where new data were collected. Section 6 outlines how the data were divided into training and evaluation sets, how the optical ground truth was generated, how the model was trained and our assessment methodology. Section 7 presents the results of this experiment. Section 8 presents the lessons learned and Section 9 concludes the paper and outlines future research.

2 | RELATED WORKS

Thermal imaging has certain advantages over visual spectrum reflected light in various applications such as to reveal hidden details that cause changes in surface temperature and aid navigation in low light conditions that are not visible to the naked eye. The fundamental concepts of thermal sensor, their advantages and disadvantages, as well as their usage are well documented in Nguyen et al. (2021).

Beside being utilized for navigation, thermal sensor has been used in agriculture to monitor crops (Speth et al., 2022), infrastructure monitoring (Chokkalingham et al., 2012; Fuentes et al., 2021; Stypułkowski et al., 2021; Wu et al., 2018), objects detection and tracking (Leira et al., 2021; Liu, Li, et al., 2020; Liu et al., 2017, 2022).

For navigation applications, some researchers have demonstrated early promise from combined LWIR thermal and optical light sensors to enhance and detect hidden features in low light conditions. Maddern and Vidas (2012) used LWIR thermal data to compensate for adverse effects such as sun-induced lens flare in RGB images during daylight, and to enhance hidden features after dark. The result showed that the extra data from the LWIR sensor improved the resilience of the tested system over longer operating periods. Brunner et al. (2013) made use of thermal data to more efficiently detect and reject bad features such as dust and reflective surfaces that appear in the visible spectrum.

Mouats et al. (2014) proposed multispectral stereo odometry for unmanned ground vehicle. Mouats et al. (2015) later developed a purely thermal approach to the problem. Another study (Poujol et al., 2016) showed that combining thermal and visible data, the performance of the system greatly improved in various lighting conditions.

While the above work has shown the potential of thermal sensor in the field, they struggled to some extent due to the lack of a truly radiometric LWIR sensor. Early sensors were built with Automatic Gain Control (AGC) systems that would maximize, and in so doing, substantially disrupt the contrast of the output when a relatively hot or cool object would enter or exit the scene. A radiometric sensor is calibrated against the standard black body (Nguyen et al., 2021) leading to greatly improved accuracy and consistency of thermal data. Furthermore, a radiometric sensor can output pre-AGC thermal data representing thermal emissivity in 14- or 16-bit depth that can be handled differently to maintain contrast between two frames.

Furthermore, with the introduction of modern small and low power consumption radiometric LWIR sensors such as those from FLIR Corporation, the sensor can now be embedded into small UAVs.

Saputra et al. (2020) proposed a thermal–inertial network hallucinating features from thermal data and fusing them with an inertial sensor for pose estimation. The results showed the proposed network outperform their visual counterparts under various lighting conditions. However, the network was limited to a frame rate of 4–5 fps, with accuracy severely degraded at higher or lower frame rates.

Khattak et al. (2019b) proposed a fusion of visual light, radiometric LWIR sensors with an inertial sensor to help small UAVs navigate in a dark tunnel filled with fog and dust. The drone carried a short-range light source to illuminate the environment for the optical sensor while the thermal sensor provided data beyond visible wavelengths. They later developed a direct thermal system without the optical sensor (Khattak et al., 2019a).

The same team (Khattak et al., 2018) also proposed the use of thermal fiducial markers as a low-cost and reliable solution for robot localization in known but visually degraded environments. More recently, Khattak, Papachristos et al. (2020) proposed a thermal inertial system that utilized full 14-bit radiometric data for a small UAV. The study indicated that using full unprocessed radiometric data can not only bypass the troublesome AGC, but can also be more resilient against loss of track of features over time, which results in more consistent performance. Most recently, Khattak, Nguyen et al. (2020) proposed the fusion of thermal and visual inertial with light detection and ranging (LiDAR) to improve reliability for pose estimation.

Some research groups have been attempting to enhance simultaneous localization and mapping (SLAM) vision-based systems with thermal data (Chen et al., 2017; De Pazzi et al., 2022; Saputra et al., 2021; Zhao et al., 2020). Shin and Kim (2019) proposed a direct thermal-infrared algorithm that utilized full 14-bit radiometric thermal data to measure six degrees of freedom (DOF) motion. They demonstrated that the system was more robust under various lighting conditions across both day and night. Chen et al. (2022) developed an edge-based infrared/LiDAR SLAM framework to reliably generate a dense depth network throughout the day.

3 | MOTIVATIONS AND CONTRIBUTION

The materials in this paper is a continuation of our program (Nguyen, Rosser, & Chahl, 2022; Nguyen, Rosser, Perera, et al., 2022; Rosser et al., 2021) to systematically explore the application of optical flow from thermal imaging, "thermal flow," for airborne navigation with an LWIR microsensor. Our first study in Rosser et al. (2021) explored the use of an LWIR optical flow sensor to reduce lateral drift in a closed-loop control system instead of an optical reflected light sensor, the PX4Flow (Honegger et al., 2013), for fixed-wing UAVs.

In our second study in Nguyen, Rosser, Perera et al. (2022), we investigated the performance of thermal flow over 24 h from one test site. We learned that thermal flow with the LK sparse technique, performs comparably to the PX4Flow during daylight, while suffering during cold-soaked conditions when there are few features in the thermal image.

In a recent study by Nguyen, Rosser, and Chahl (2022), a comparison between a dense optical flow technique, the Image Interpolation Algorithm (I^2A) (Srinivasan, 1994), to the sparse technique, the LK, showed that the I^2A performs well to the LK in normal conditions, better in cold-soaked condition while suffering from the aperture problem in some scenes.

In this study, we investigate the feasibility of deep-learning neural networks being used to compute thermal flow from real-world aerial data. The Recurrent All-Pairs Field Transform for the Optical Flow (RAFT-s) model (Teed & Deng, 2020b) was chosen due to its ability to achieve state-of-the-art accuracy on a standard synthetic data set, with substantially less of parameters at 990M with a small memory footprint of 4 MB.

Two synthetic data sets, the flying chairs (Dosovitskiy et al., 2015) and MPI-Sintel (Butler & Wulff, 2012), and our own real-world thermal data set including some that were captured from our previous studies (Nguyen, Rosser, & Chahl, 2022; Nguyen, Rosser, Perera, et al., 2022; Rosser et al., 2021) were used to train the RAFT-s network to answer these questions:

- Can a network designed for and trained on RGB images works on thermal data?
- Can this modern deep-learning network outperform traditional techniques in variable conditions, with high- and low-contrast thermal data?
- Does the network suffer from the aperture problem without being explicitly trained for it?

Our contribution is significant because:

- We will show that the modern neural network performs comparatively well, or even outperform the traditional techniques under several operating conditions.
- A transfer learning modern neural network from synthetic data, can be retained with real-world thermal data to detect very low-contrast thermal features in cold-soaked conditions.
- Optical flow ground truth can be generated with traditional techniques to deal with real-world problems.

4 | OPTICAL FLOW EQUATION

Most traditional optical flow techniques operate based on three assumptions (Horn & Schunck, 1981):

- *Brightness constancy*: Pixel brightness and contrast do not change between two consecutive frames.
- *Small displacement*: Pixel movements between two frames must not be too large.
- *Spatial coherence*: Neighboring pixels move together across two consecutive frames.

Considering the first assumption, with pixel $I(x(t), y, t)$ in the first frame displaced by (u, v) pixels in the next frame. We have

$$f(x) \equiv I(x, y) = I(x + u, y + v). \tag{1}$$

Since the pixel intensity does not change over time:

$$\frac{\partial f(x)}{\partial t} = 0. \tag{2}$$

The second assumption, small movements, which is the movement of pixels is small across two frames. Applying the Taylor series approximation to the right-hand side then simplify

$$\begin{aligned} I(x + u, y + v) &= I(x, y) + \frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + HigherOrderTerms \Rightarrow \\ &\approx I(x, y) + \frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v. \end{aligned} \tag{3}$$

From Equations (1) and (3), we get the *optical flow* equation:

$$I_x u + I_y v + I_t = 0, \tag{4}$$

where I_t is the time derivative of the image at pixel (x, y).

Equation (4) cannot be solved with one equation as there are two unknowns, u and v, per pixel. This is a geometrical constraint for drone imagery, where the speed of the aircraft and height above ground will define the amount of movement.

The optical flow equation has traditionally been solved as a mathematical energy minimization problem over the last decades. With the emergence of the deep-learning approach in the field, combined with some advanced data sets, optical flow estimation with neural network has demonstrated good results.

4.1 | Traditional methods

Traditional optical flow estimation techniques can be divided into three broad categories: pixel based (Chen & Koltun, 2016; Menze et al., 2015), feature based (Brox & Malik, 2010; Bruhn et al., 2005; Farnebäck, 2003), and energy based (Brox et al., 2004; Steinbrücker et al., 2009). In practice, there are further two variants which are dense and sparse optical flow techniques. The dense technique will compute optical flow for every pixel in a frame while a sparse technique only computes for selective pixels. The advantage of sparse techniques is that they are generally less computationally demanding and can work on embedded systems (Nguyen et al., 2021), where the entire optical flow field is not required. The Pyramidal LK spare implementation in OpenCV library (Bradski, 2000) is a very popular choice for real-time embedded systems.

Additionally, while the first condition: brightness constancy, must be met for traditional methods, it also applies on deep-learning-based technique due to all available data sets maintaining the same conditions across the images.

4.2 | Deep-learning-based methods

On the basis of the concept of deep-learning and convolutional neural networks (CNNs), these types of techniques learn to compute optical flow from sequences of images, and have been showing promising results. Some of the notable optical flow deep-learning network, such as FlowNet (Dosovitskiy et al., 2015), FlowNet2 (Ilg et al., 2017), LiteFlowNet (Hui et al., 2018), PWC-Net (Sun et al., 2018), and RAFT (Teed & Deng, 2020b).

Instead of relying on predefined features or energy minimization like traditional techniques, deep-learning methods rely on a large amount of quality labeled data for training. This is a major issue in real-world applications since it is extremely difficult to obtain accurate dense airborne optical flow ground truth in real scenes (Shah & Xuezhi, 2021). So far, researchers have relied on synthetic data such as Middlebury (Baker et al., 2011), KITTI (Geiger et al., 2012; Menze & Geiger, 2015), MPI-Sintel (Butler & Wulff, 2012), and flying chairs (Dosovitskiy et al., 2015) to train and evaluate neural networks. Although these synthetic data sets display movement, they do not reproduce realistic effects, such as noise, blur, fog, shadows, contrast disparity, and so forth. Many state-of-the-art models that were trained on synthetic data may experience challenges in real situations with more complex image sequences (Shah & Xuezhi, 2021). Hence, the lack of a quality labeled real-world data set currently remains a challenge for deep-learning optical flow approaches under field conditions.

4.2.1 | The RAFT and RAFT-s model

With limited size and weight options for UAVs, the overall techniques should have the least computational demand possible. The RAFT-s model was chosen for this study due to its low footprint of 4 MB and low parameter count of 990K, while still achieving state-of-the-art accuracy in several benchmark standard data sets (Teed & Deng, 2020b). Another reason is that the model is open-source and its implementation in Pytorch is available on Github (Teed & Deng, 2020a).

RAFT is a composition of CNN and recurrent neural network architectures, consisting of three main blocks: a feature/context encoder, a convolution layer, and a recurrent Gated Recurrent Unit (GRU)-based layer.

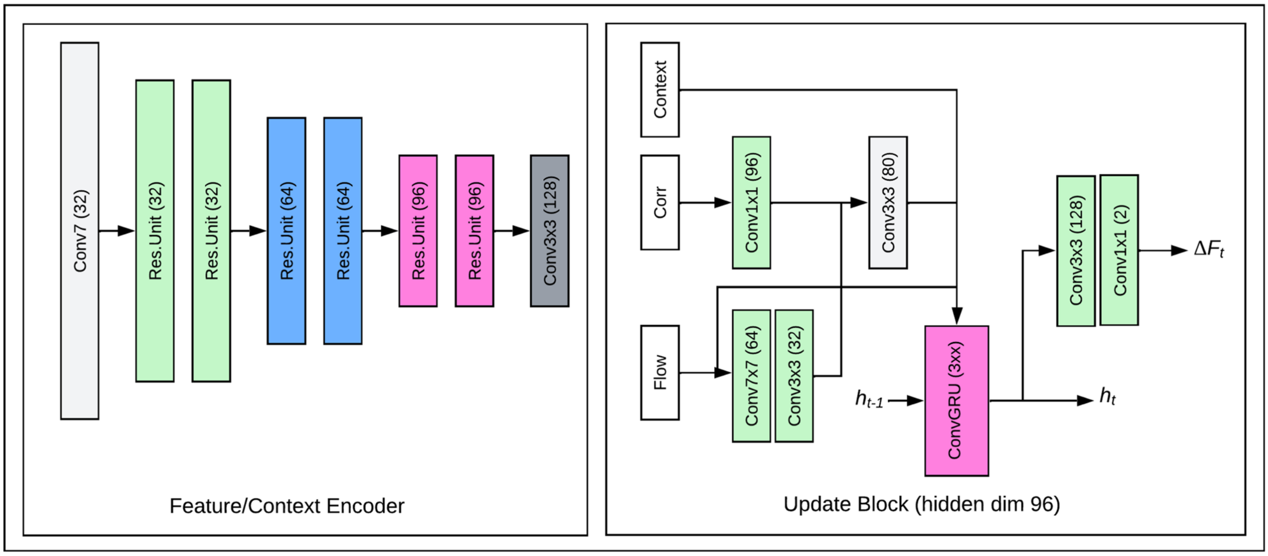

FIGURE 1 The model was used in this study (Teed & Deng, 2020b).

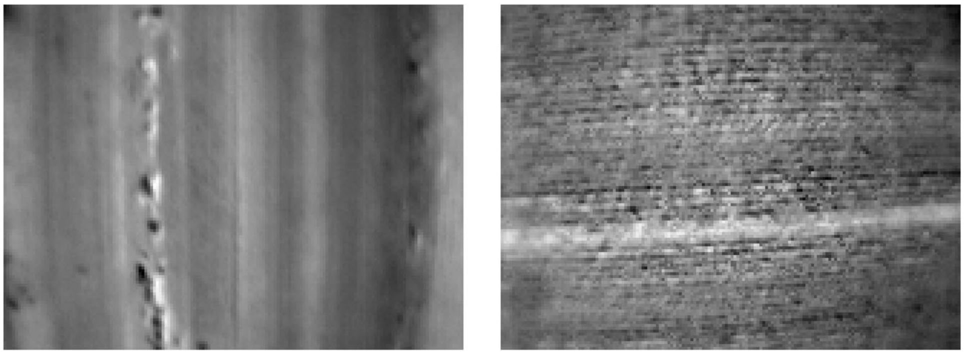

FIGURE 2 Sample thermal images show vertical and horizontal lines from data set (Nguyen, Rosser, Perera, et al., 2022).

In a feature extractor layer, the input consists of two consecutive frames, which are similar to the FlowNetCorr architecture, where features are extracted from two images separately. In this layer, per-pixel features are extracted both input frames and along side with one context encoder layer that only extracts features from the first frame. In a convolution layer, a four-dimensional (4D) correlation volume contains high values pixels and then processed by average pooling the last two dimensions with kernel sizes of 1, 2, 4 then 8.

The CNN architecture is inspired by ResNet, which consists of six residual layers with the resolution is reduced by haft while a number of channels increase. The only difference between RAFT and RAFT-s models is that the small model only uses a single GRU with 3 × 3 filter instead of two convolutional GRU

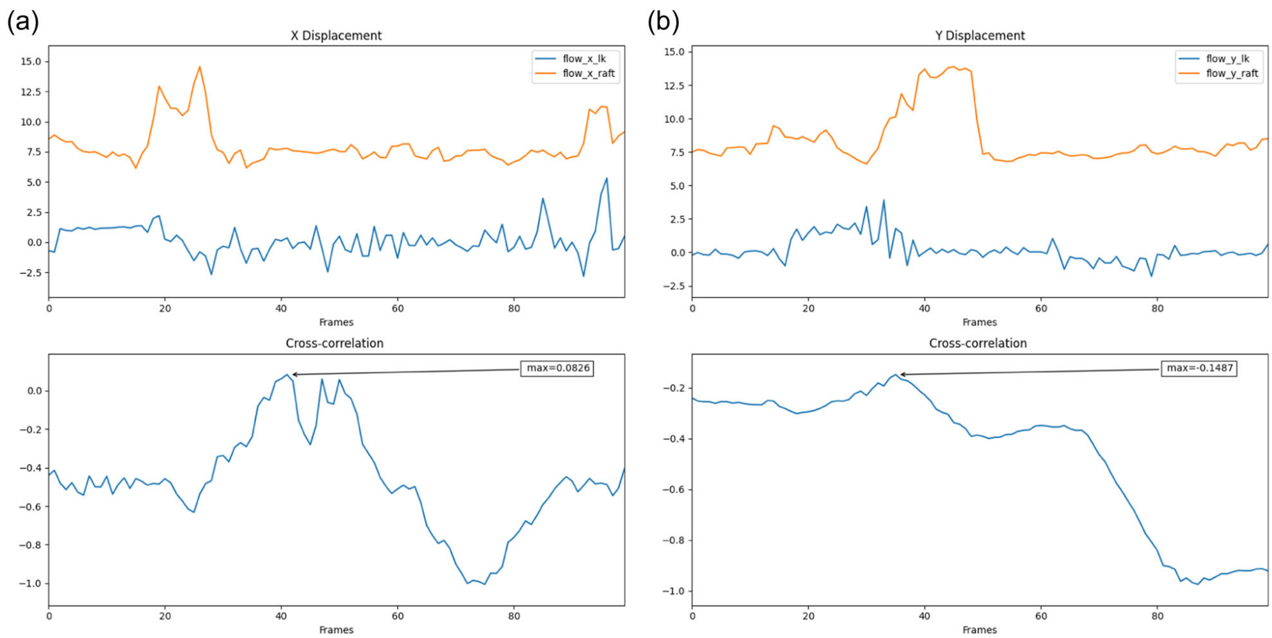

FIGURE 3 *X* and *Y* displacements with pretrained RAFT-s networks. Low normalized cross-correlation indicates the RAFT-s trained with RGB data set alone does not work well with thermal data. (a) *X* displacement and (b) *Y* displacement. RAFT-s, Recurrent All-Pairs Field Transform for the Optical Flow.

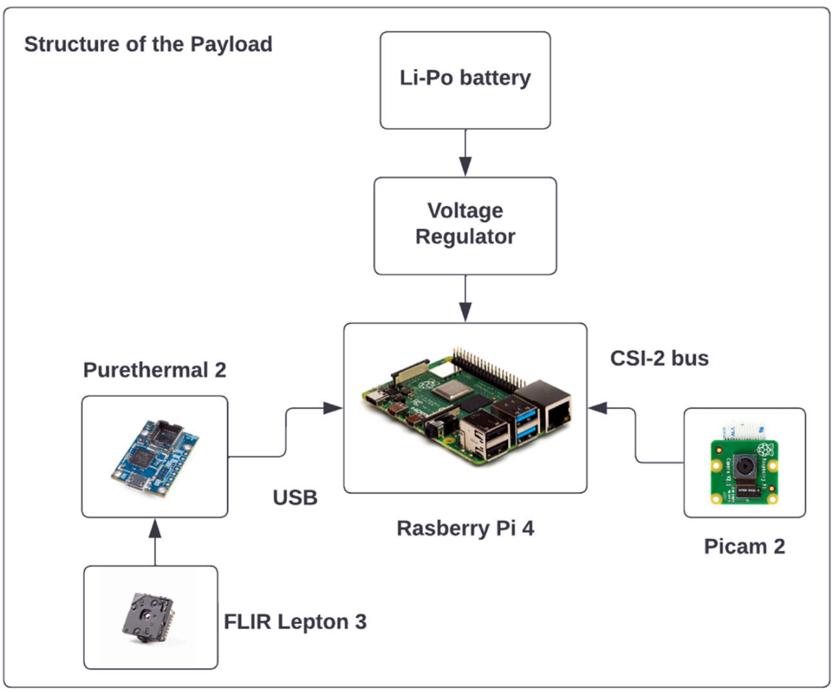

FIGURE 4 A block diagram of the payload showing the selected components. CSI, Camera Serial Interface; USB, Universal Serial Bus.

update blocks. Figure 1 shows the model adapted from Teed and Deng (2020b).

4.2.2 | Thermal frames with pretrained RAFT-s model

This section investigates how the RAFT-s network trained on RGB synthetic data performs with aerial thermal data. We used a pretrained model on the flying chairs three-dimensional (3D) and MPI-Sintel data set, and the test sequence was from Nguyen, Rosser, Perera et al. (2022) above a wheat field during a period with good thermal contrast. Some of the images are shown in Figure 2.

Figure 3 shows the overlaid signal of the RAFT-s and the LK technique. The signals of the model in both *X* and *Y* displacements show that the model does not provide a good optical flow field with thermal data despite achieving good results from the color synthetic data set.

Hence, it is necessary to retrain the RAFT-s network on a real-world thermal aerial data set. In this study, we use real aerial thermal data collected from our previous studies to train and test the RAFT-s model.

5 | SYSTEM IMPLEMENTATION

This section outlines our selected components, their configurations, thermal sensor, and AGC issue that affects the optical flow assumption.

The system block diagram is shown in Figures 4 and 5 illustrating inside and outside of the constructed payload. The system consists of a Raspberry Pi 4 (Pi 4) as the main processor, a radiometric FLIR Lepton 3 thermal sensor, and an RGB Picam 2. The payload is powered by the Li-Po battery through an Ultimate Battery Elimination Circuit voltage regulator at providing 3 A at 5 V. The FLIR Lepton 3 was mounted into the Purethermal 2 board which connects to the Raspberry Pi 4 via Universal Serial Bus connector. The Picam 2 connects to the Pi 4 via onboard CSI-2 bus. The purpose of the Picam 2 is to collect color images for terrain analysis and completeness of the data set.

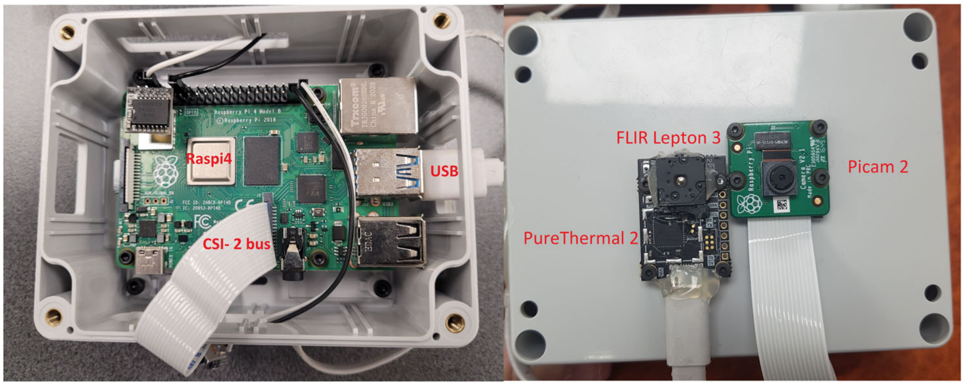

FIGURE 5 Inside (on the left) and outside (on the right) of the payload. The outside shows the Lepton 3 and the Picam, and the inside shows the Pi 4 with CSI and USB connections to the Picam and Lepton 3, respectively. CSI, Camera Serial Interface; USB, Universal Serial Bus.

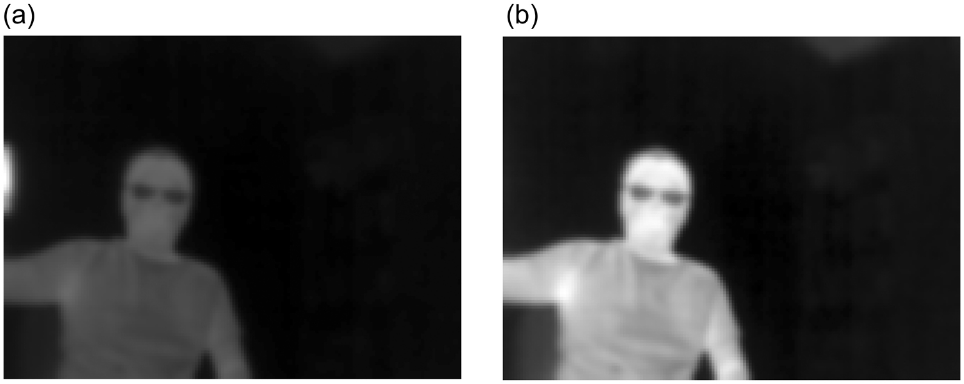

FIGURE 6 AGC changes the contrast in an image when a hot cup exits a scene: 1–2. (a) Frame 1 and (b) Frame 2. AGC, Automatic Gain Control.

5.1 | Thermal sensor configuration

The FLIR Lepton 3 (FLIR Corporation, 2018) is an LWIR-calibrated radiometric microthermal sensor. The sensor can output thermal image data at 160 × 120 resolution and 14-bit depth at 8.7 Hz. The sensor was shown to be sufficient for navigation application with small angular movement errors without further calibration (Rosser et al., 2021).

A Purethermal 2 board was used to interface with the Lepton 3, it provides raw radiometric data in 14-bit depth to the onboard computer. Its firmware was modified to be in HY16 mode and to disable the Flat Field Correction process since the sensor was mounted on a constantly moving platform (Khattak et al., 2019b; Nguyen, Rosser, Perera, et al., 2022) to ensure the output is not interrupted. The saved full radiometric 14-bit data will later be used for training and validation.

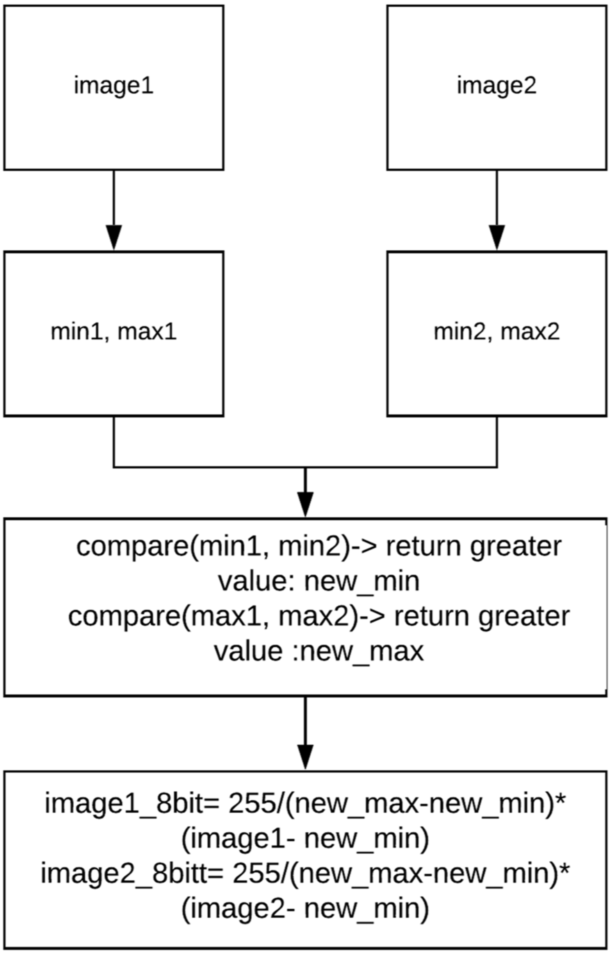

FIGURE 7 The 14–8-bit downsampling technique from Nguyen, Rosser, Perera et al. (2022).

5.2 | AGC with optical flow assumption

AGC is a closed-loop feedback system that is built into many modern thermal sensors (Nguyen et al., 2021), including the FLIR Lepton 3. The AGC is essential when converting 14-bit raw thermal data to 8-bit data. The purpose of the AGC is to enhance thermal contrast when there is a substantial change in the temperature of a pixel in the image, when a hottest or coolest object enters or exits the scene. This issue is particularly troublesome for many navigation and feature-based techniques that rely on pixel intensities matching between two frames. Figure 6 shows the change in contrast between a pair of thermal images when a hot cup exits the scene.

Due to the brightness being inconsistent between two frames, the optical flow condition of brightness constancy as described in Section 4 is violated. Due to the current implementation of the RAFT-s network designed to work with 8-bit data, this is a major issue.

5.3 | Rescaling technique

This section revisits our proposed technique in Nguyen, Rosser, Perera et al. (2022), to downsampling 14-bit to 8-bit data based on

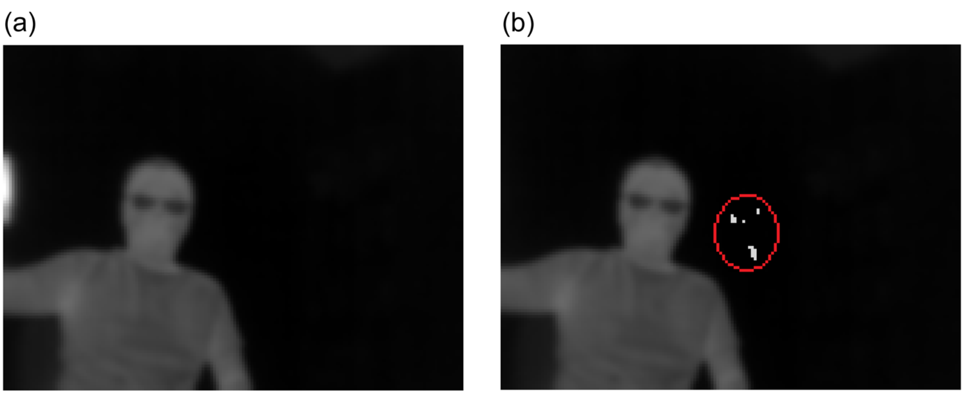

FIGURE 8 A pair from Figure 6 with our technique with small artifacts are circled in red. (a) Frame 1 and (b) Frame 2.

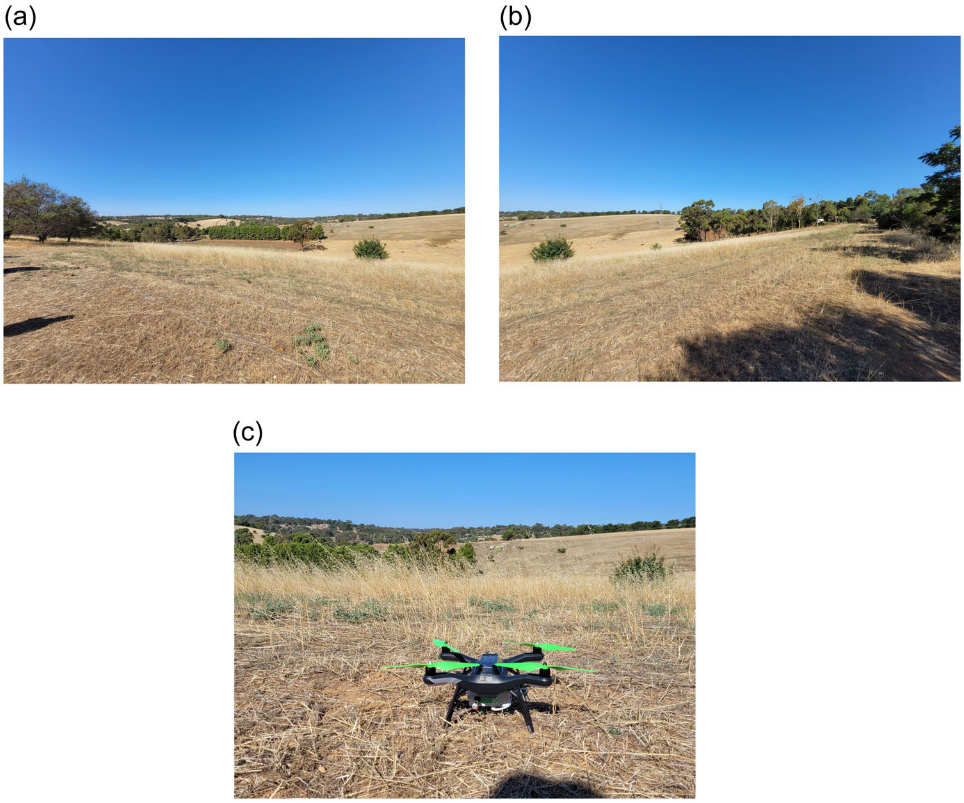

FIGURE 9 Color images of the testing site showing the left and right of the field as well as the image of the drone with the payload mounted underneath. (a) Left side of the field, (b) right side of the field, and (c) the SOLO on the ground.

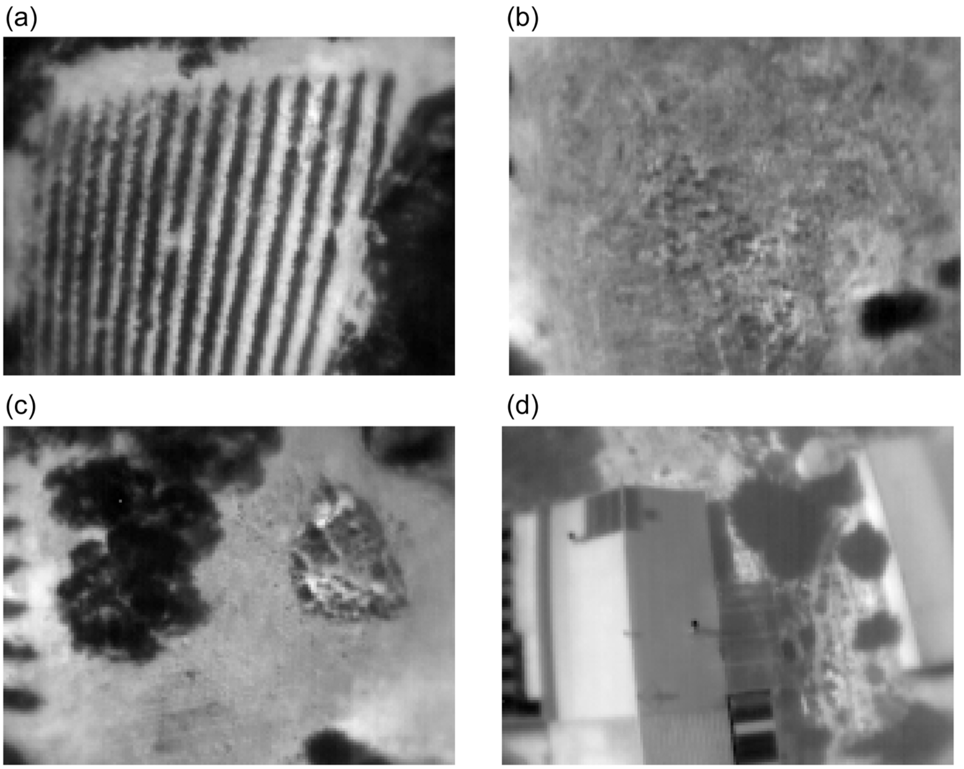

FIGURE 10 Some collected 8-bit thermal images of the site over a vineyard, an empty field, a large tree, and over a house with solar panels. (a) Frame 1: over a vineyard, (b) Frame 2: over an empty field, (c) Frame 3: over a large tree, and (d) Frame 4: over a house with solar panels.

the maximum and minimum pixel intensity of each pair. The technique is demonstrated in Figure 7.

Figure 8 shows the results with our technique from Figure 6. The results show no change in image contrast across the pair with minimal artifacts in this extreme case. The artifacts normally appear at otherwise undetected optical flow regions, so its impact is minimal.

5.4 | Flying platform

The payload as described in Section 5 was mounted underneath the 3DR SOLO quadrotor. The drone can take off with the maximum payload of up to 500 g, with flight endurance is up to 10 min. We did a total of two flights to collect our data, which is around 20 min in mid-air. The drone was flown randomly under manual control with different altitudes, heading, and velocities over a field.

5.4.1 | Field experiment

An empty field and a house were chosen as a site for this study. The site provides a clear view of the sky, flat terrain, thermal texture from the artificial objects, such as house, solar panels, car, and so forth. The flight was conducted in the afternoon at 4 p.m. for maximize thermal emissivity. The temperature at the site was 27°C (Bureau of Meteorology, 2022) with clear and sunny conditions. Figure 9 shows some of the images of the field on the day of the experiment, Figure 10 shows some collected 8-bit

TABLE 1 Information about the training and evaluation sets.

	Source	Training set	Evaluation set	Site condition	Total images
Data set 1	Rosser et al. (2021)	Yes: 10,894	Yes: 2000	High contrast	12,894
Data set 2	This study	Yes: 5792	No	High contrast	5792
Data set 3	Nguyen, Rosser, Perera et al. (2022)	No	Yes: 9892	High and low contrast	9892
Data set 4	Nguyen, Rosser, and Chahl (2022)	No	Yes: 2800	High and low contrast	2800
Total images		16,686	14,692		27,586

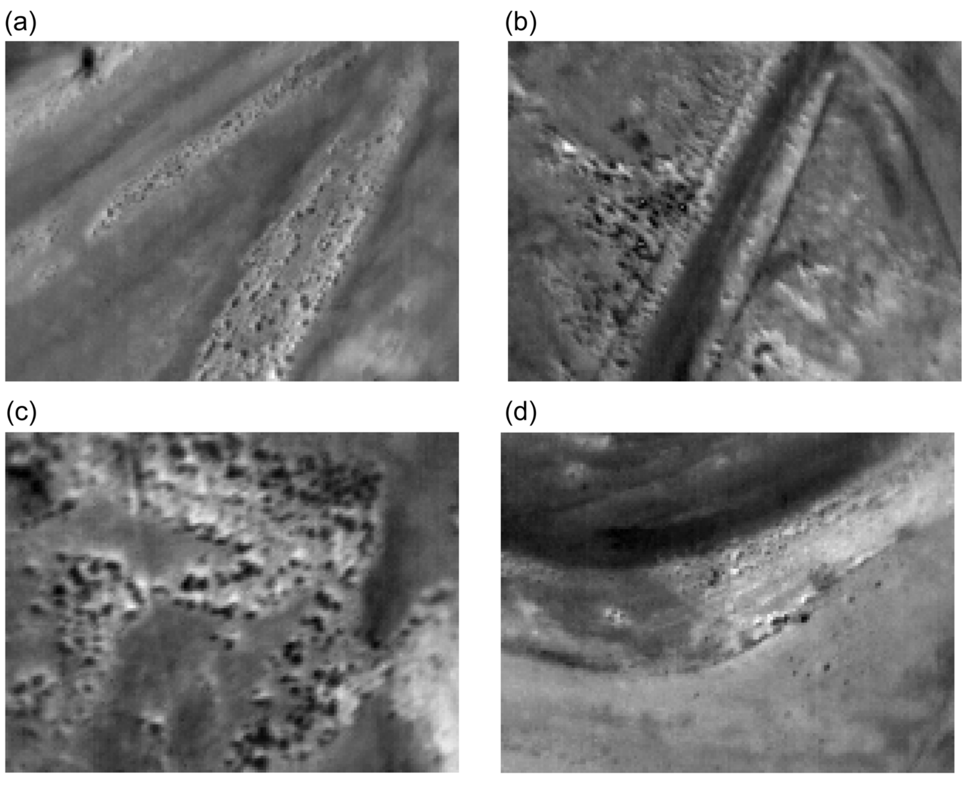

FIGURE 11 Some thermal frames from Data set 1, over some interesting features of the field includes dry creek beds, a runway, and a dirt road. (a) Frame 1: above dry creek beds, (b) Frame 2: over a runway, (c) Frame 3: above different dry creek beds at different angles, and (d) Frame 4: over a dirt road.

thermal frames of the test site. In total, 5792 images were collected during this flight.

6 | DATA AVAILABILITY

Our data set consists of 14-bit full radiometric data across three different locations: an empty wheat field, an arid desert, and an empty top hill. The data set contains images from our previous studies in Nguyen, Rosser, and Chahl (2022), Nguyen, Rosser, Perera et al. (2022), and Rosser et al. (2021) and our newly collected thermal data from Section 5.4.1. To the best of my knowledge, there is currently no public 14-bit aerial thermal data set available. There is one available data set provided by FLIR Corporation (2021) for object detection for night driving. However, the data set cannot be used for this study because the data set is not airborne, and there is no optical flow ground truth provided either.

Perfect ground truth from real-world data is difficult to obtain, yet it is essential to train the network. To overcome this issue, ground truths were generated from our data with traditional dense techniques, the implementation of the dense optical flow Farneback technique in OpenCV. The Farneback technique was chosen as it is one of the most reliable techniques for generating dense optical flow fields. Furthermore, only high-contrast sequences of our data set were used for training to achieve the best possible ground truth for the network. For validation, both the high- and low-contrast sequences will be used.

One limitation is that due to the ground truth being generated from traditional techniques, the trained model may not show its full potential, and may not be better than the traditional techniques. On the other hand, both the sparse and dense traditional techniques have proven to be superior to the PX4Flow (Nguyen, Rosser, & Chahl, 2022; Nguyen, Rosser, Perera, et al., 2022), hence the comparison is still valid.

6.1 | Data set division

In total, there were four different data sets collected over three sites at different times, varying from Summer through Autumn to Winter. We divided our data into two sets: the training and evaluation sets.

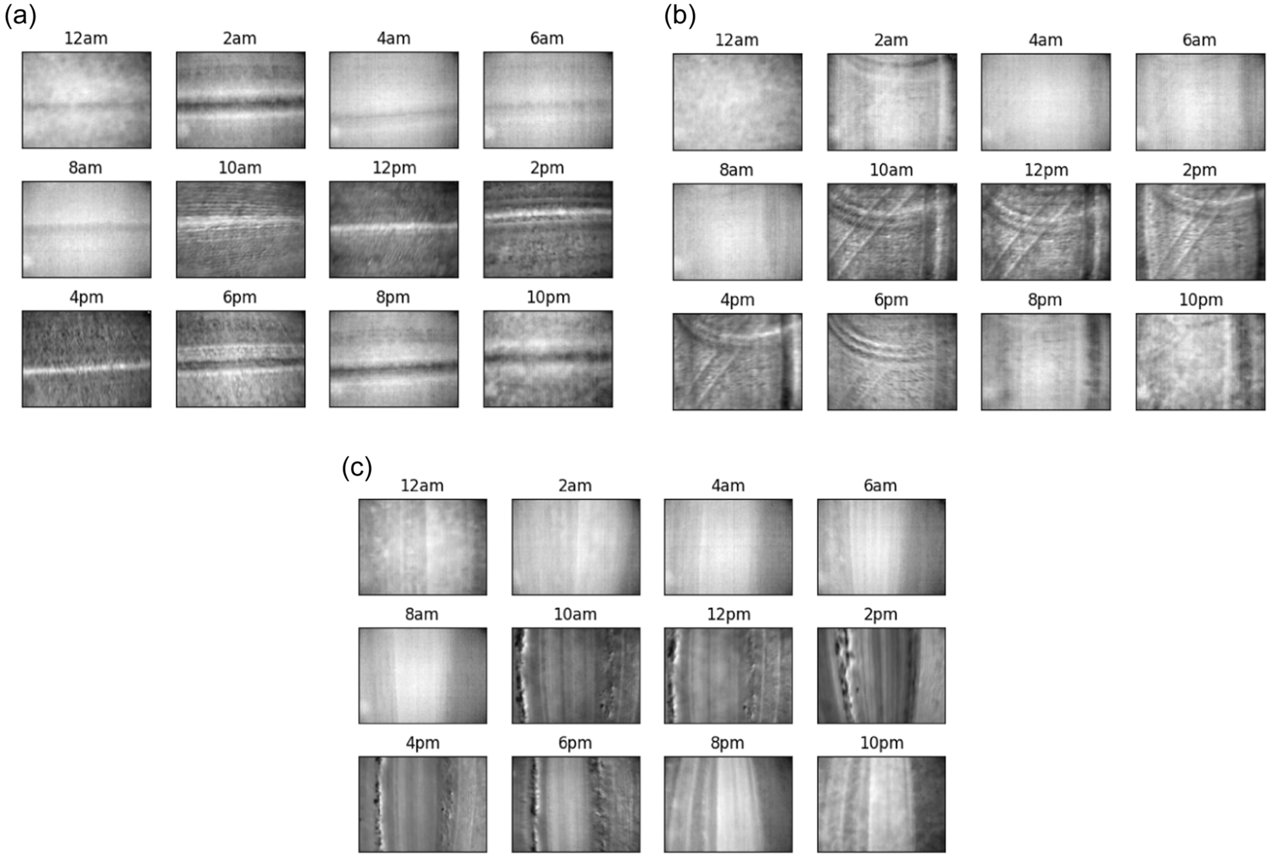

FIGURE 12 Features of the ground at some interesting locations including the wheat field, the top corner, and the dirt road over 24 h of the data set. Only sequences at 4 p.m., 2 a.m., 12 a.m., and 4 a.m. were chosen. (a) Features over the wheat field, (b) features over the top left corner, and (c) features over the main road.

Table 1 shows a summary of our data sets, including which data set and how many images were dedicated to train and evaluate the model and conditions of each site.

6.2 | Training set

The training set contains sequences captured during high thermal contrast condition to generate the best optical flow field possible. The set contains two data sets which we labeled "Data set 1" from Rosser et al. (2021) and "Data set 2" which was newly captured at the site under the conditions described in Section 5.4.1.

6.2.1 | Data set 1

Data set 1 includes 12,894 thermal frames, showing images of an arid expanse with sparse knee-high vegetation. The data were captured during late Summer under clear conditions and the temperature was 34°C (Bureau of Meteorology, 2020). Due to its remoteness, there are minimal artificial objects present on the site. Figure 11 shows some images from the data set.

We also reserved some sequences for the evaluation set: 10,894 images for training and 2000 for the evaluation set.

6.2.2 | Data set 2

Data set 2 was described in Section 5.4.1. The whole data set contains 5792 images that were all used for training.

6.3 | Evaluation set

The evaluation set consists of "Data set 3" from Nguyen, Rosser, Perera et al. (2022), "Data set 4" from Nguyen, Rosser, and Chahl (2022), and approximately 2000 images from "Data set 1."

6.3.1 | Data set 3

Data set 3 was collected over 24 h as described in Nguyen, Rosser, Perera et al. (2022). The data set exhibited strong parallel lines across the horizontal and vertical frames, the two-dimensional (2D) motion of which cannot be determined unambiguously due to the presence

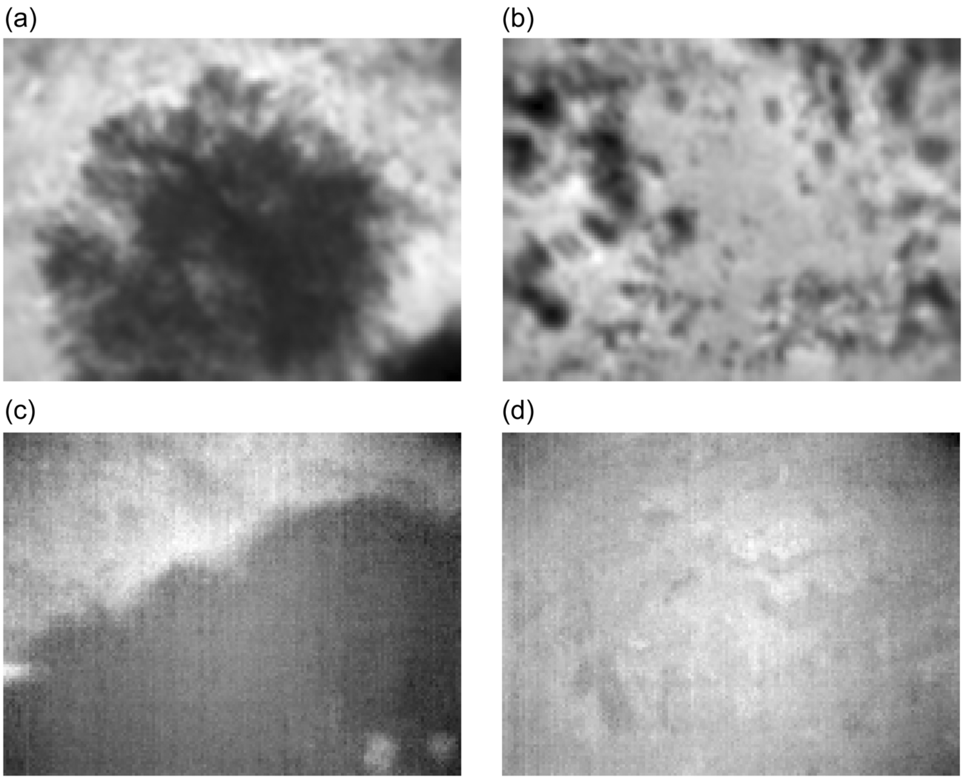

FIGURE 13 Some of the images from Data set 4. Frames 1 and 2 show the field during high-contrast condition, and frames 3 and 4 show thermal images at approximately at the same location but in low-contrast conditions. (a) Frame 1: above a big tree with high contrast, (b) Frame 2: over an empty field with high contrast, (c) Frame 3: over a big tree during low contrast, and (d) Frame 4: over an empty field during low contrast.

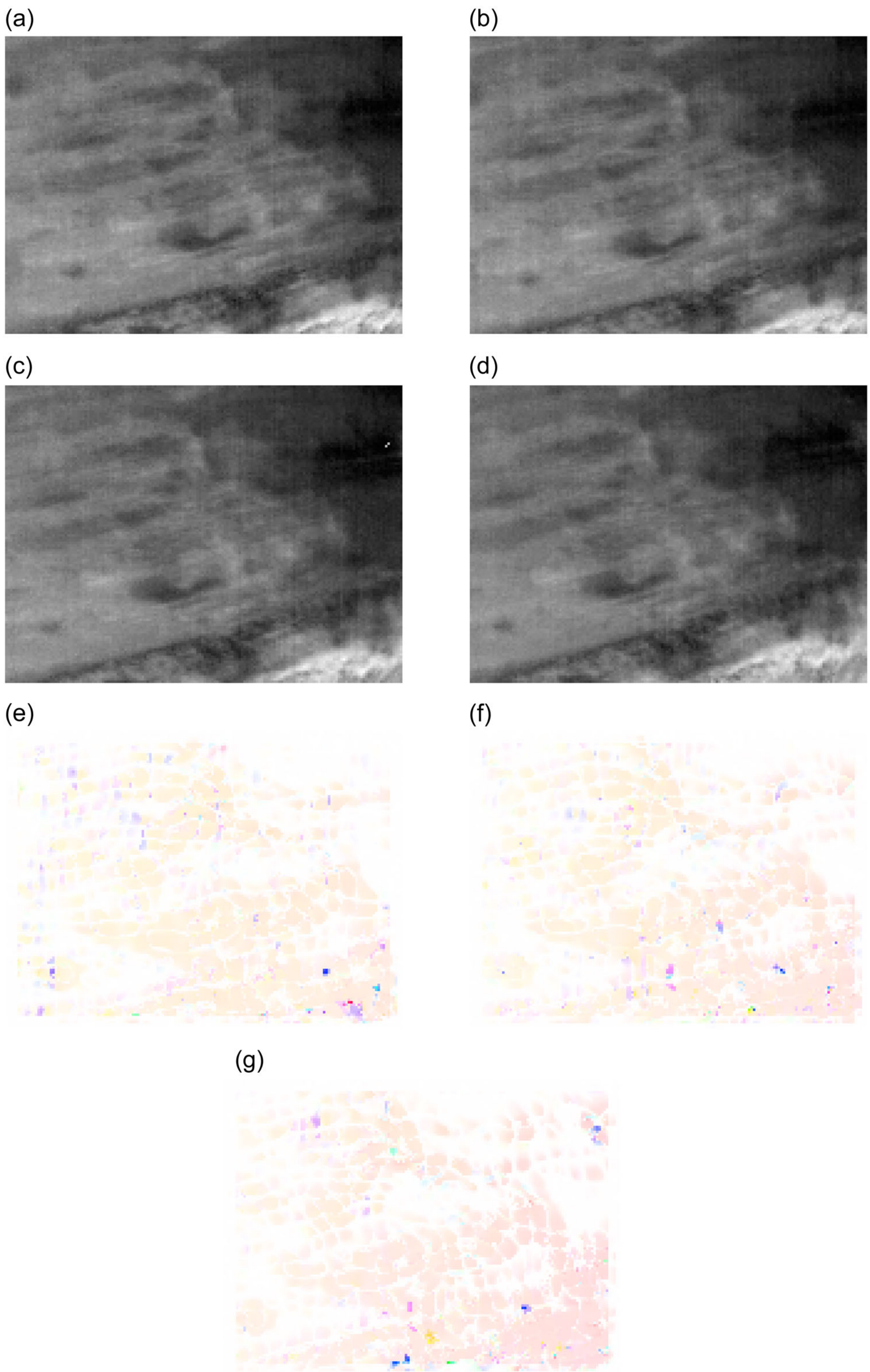

FIGURE 14 A sample sequence of thermal data and its generated ground truth from dense optical flow from the Farneback algorithm in OpenCV. (a) Frame 1, (b) Frame 2, (c) Frame 3, (d) Frame 4, (e) flow output from frames 1 and 2, (f) flow output from frames 2 and 3, and (g) flow output from frames 3 and 4.

of information on only one axis, which is commonly known as the "aperture problem" (Binder et al., 2009). Dense techniques suffer from this issue as shown in Nguyen, Rosser, and Chahl (2022), so it cannot be used to generate reliable ground truth. On the other hand, this data set can be used to evaluate the network to analyze if it suffers from the aperture problem over time.

Figure 12 shows thermal frames of the Data set 3 at some interesting locations over 24 h. We only selected some sequences that exhibit vastly different thermal contrast conditions at 4 p.m., 2 a.m., 12 a.m., and 4 a.m. to evaluate the performance of the neural network.

6.3.2 | Data set 4

Our data set from Nguyen, Rosser, and Chahl (2022) contains approximately 2800 usable frames at the same site as Data set 2 but not at the same time. Additionally, it was captured during Autumn and Winter which represent two distinct conditions: high and low contrast. Figure 13 shows some of the thermal images of the site in both conditions.

6.4 | Generated optical flow ground truth

Figure 14 shows one sequence from Data set 1 and its generated flow ground truth. The sensor was mounted underneath a moving aircraft and every pixel moved in the same direction and approximately at the same speed. However, there are "white" pixels in the image indicating "no movement," which were anomalies. Hence, the generated ground truth is not entirely perfect with missing movement in some pixels.

Additionally, our system mimics the output of the PX4Flow, which is outputting optical flow as a 2D vector, that contains: flow_x indicating movement in the *X*-direction, and flow_y indicating movement in the *Y*-direction. Because all of the pixels move with the aircraft, in theory, we only need one good pixel movement (x, y) to get good optical flow measurement. This is the case with our work in Nguyen, Rosser, Perera et al. (2022) using the sparse technique, LK in OpenCV to compute optical flow from a small number of pixels. Hence, missing a few pixels movement is not an issue when outputting optical flow as a 2-D vector.

6.5 | Training the model

We trained the model with both color and the thermal data set collected as described in Section 6.2. The model was trained using the original weights from the flying chairs and Sintel (Teed & Deng, 2020b). Batch sizes of 10, at 160,000 steps, with a learning rate of 0.0001 and weight decay of 0.0001.

The network was trained using an Intel Core i7-7700 CPU, 64 GB of RAM, and Nvidia GTX 1080 Ti GPU. The operating system was Ubuntu 20.04, other programs, including Pytorch version 1.6.0, torchvision 0.7.0, cudatoolkit 10.1, python 3.8, and OpenCV 4.5.5.

6.6 | Model evaluation methods

We systematically evaluated the performance of the model based on two conditions: The ideal conditions and cold-soaked conditions. In hot and high-contrast conditions, the traditional techniques worked very well, while in cold and low thermal contrast conditions, thermal flow worked between reasonably well to not working at all. High contrast indicates when strong thermal emission values were present in the scene which usually translates to good features that can be used for tracking. On the other hand, low contrast means there are fewer features in the scene for tracking.

Table 2 shows the data set divided into seven (7) test cases including the traditional technique used and its operation status and if the scene exhibits strong patterns that may cause the aperture problem.

TABLE 2 Our seven (7) test cases with their characteristics including the traditional technique used in various thermal conditions.

Case	From data set	Compared with	Operation status	Aperture problem
High contrast				
1	[1]	LK	Very well	No
2	[4]	I^2A	Very well	No
3	[3] at 4 p.m.	LK	Very well	Yes
Low contrast				
4	[4]	I^2A	Reasonably well	No
5	[3] at 2 a.m.	LK	Somewhat operational	Yes
6	[3] at 12 a.m.	LK	No	Yes
7	[3] at 4 a.m.	LK	No	Yes

Note: The operation status was determined by comparing the signals to the PX4Flow system.
Abbreviation: LK, Lucas–Kanade.

7 | RESULTS

This section outlines the results of thermal from the RAFT-s network and versus the LK and the I^2A. The normalized cross-correlation value between each signal in both X and Y axes, will indicate how closely matched the two signals are. The normalized cross-correlation range is between [−1; +1], where the closer the positive value near (+1), the more closely two signals correlate to each other and vice versa.

7.1 | Ideal conditions, high contrast

7.1.1 | Test case (1)

Figure 15 shows the result for test case (1). A high correlation of value 0.83 in both X and Y displacements shows the RAFT-s model works well compared with the sparse LK in this scenario.

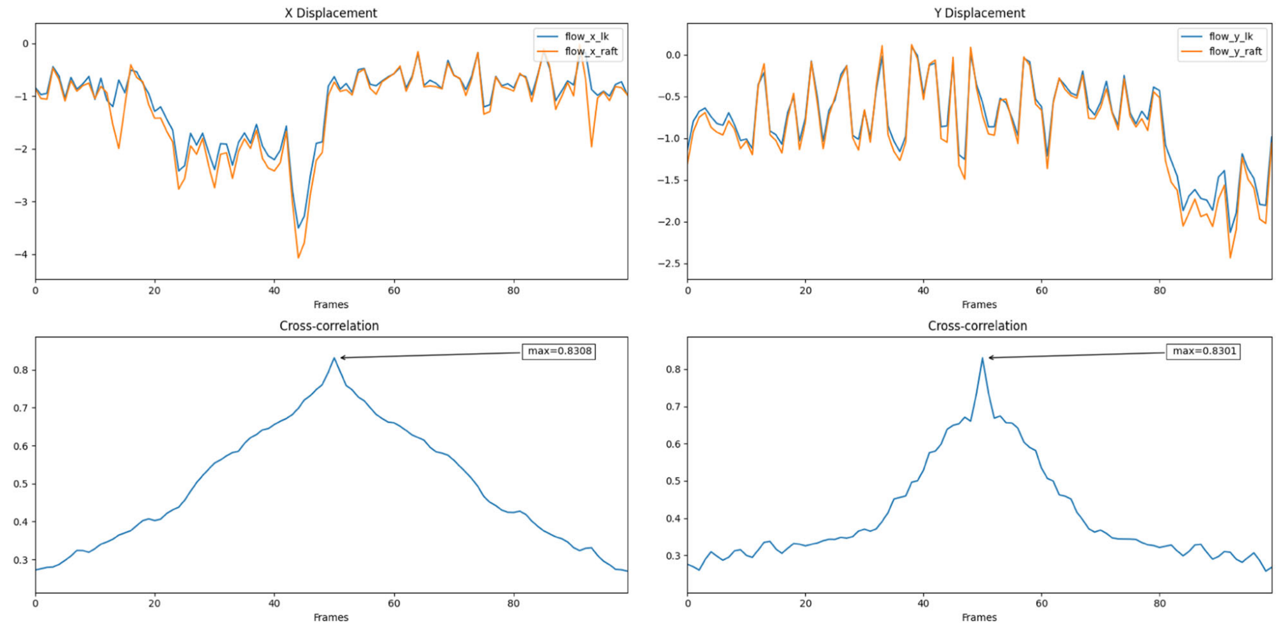

FIGURE 15 X and Y displacements for test case (1). The high normalized cross-correlation value indicates the two signals are similar.

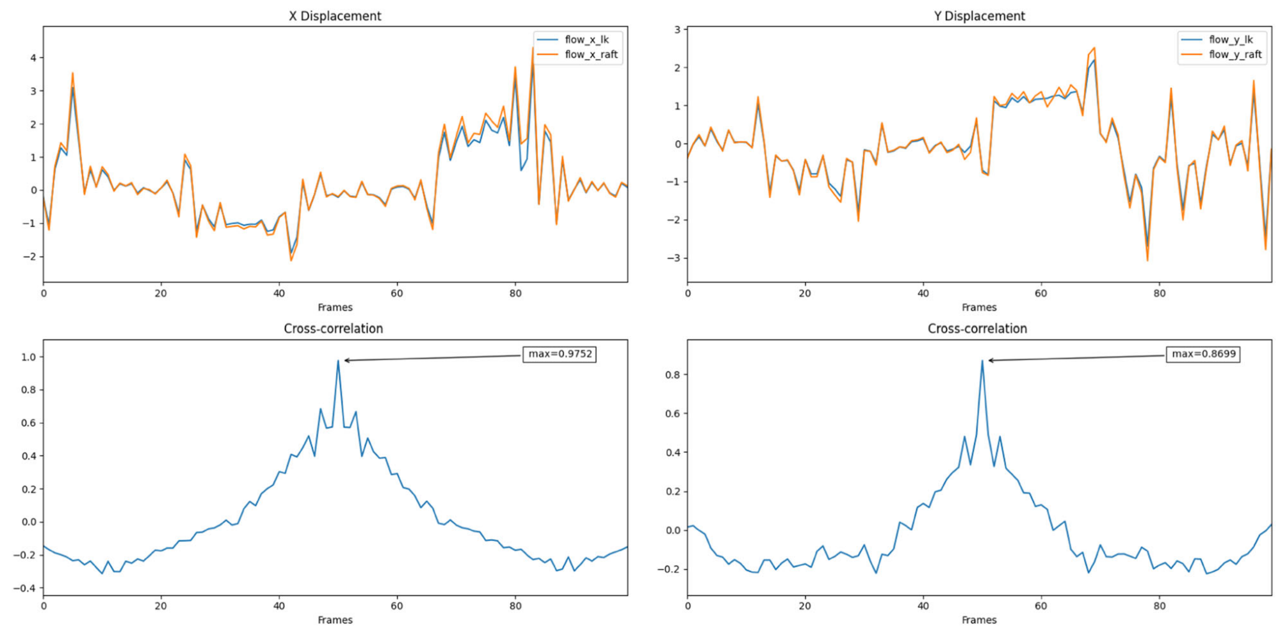

FIGURE 16 X and Y displacements for the test case (2). The high normalized cross-correlation value indicates the two signals are the same.

7.1.2 | Test case (2)

Figure 16 shows the result for test case (2). A high correlation of value 0.97 in X and 0.87 in Y displacements shows the RAFT-s model also performs well compared with the dense I^2A in this scenario.

7.1.3 | Test case (3)

Figure 17 shows the result for test case (3). A high correlation of 0.527 in X and 0.33 in Y displacements shows the model works comparatively to the LK in this case. Additionally, the results

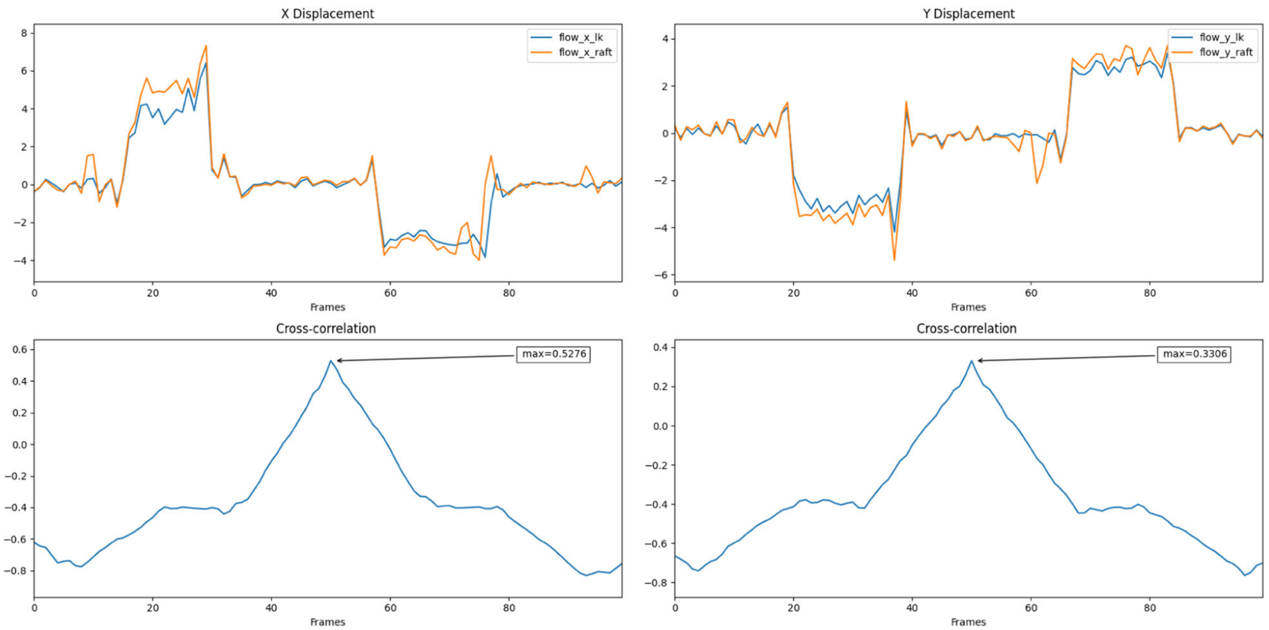

FIGURE 17 X and Y displacements for the test case (3). The high normalized cross-correlation value indicates the two signals are the closely correlated.

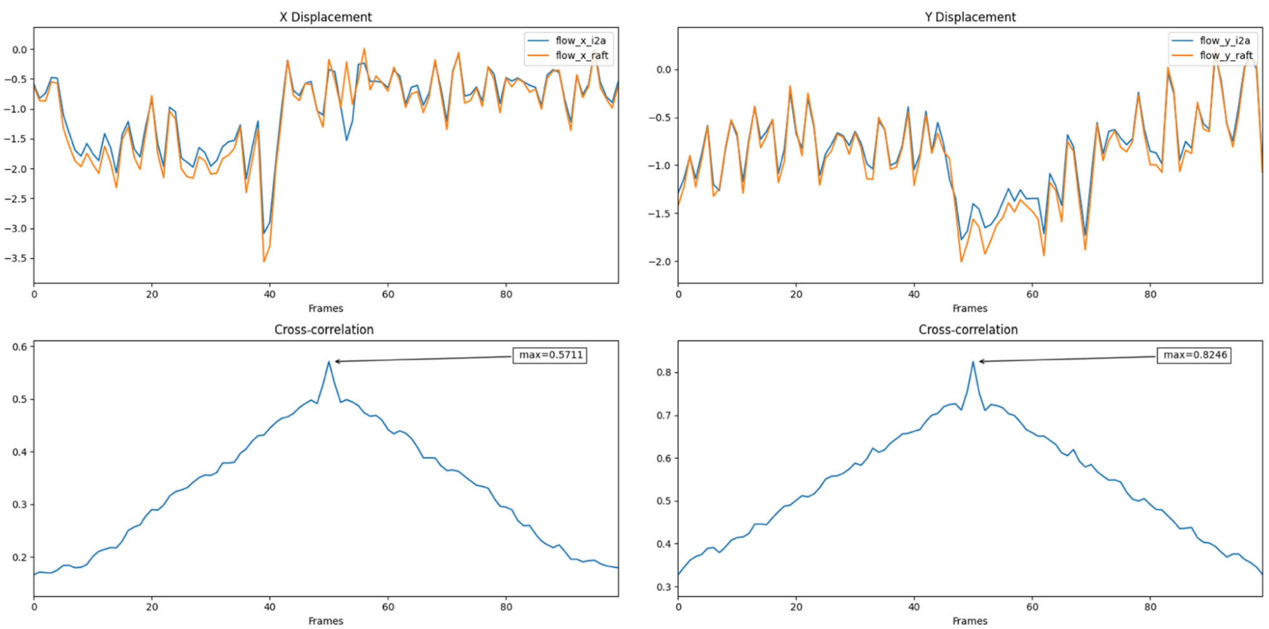

FIGURE 18 X and Y displacements for the test case (4). The high normalized cross-correlation value indicates the two signals are the same.

indicate that the RAFT-s model does not suffer from the aperture problem.

7.2 | Cold-soaked condition

7.2.1 | Test case (4)

Figure 18 shows the result for test case (4) in cold-soaked condition. A high correlation value of 0.571 in *X* and 0.8246 in *Y* displacements show the model works well compared with the I^2A in this case.

7.2.2 | Test case (5)

Figure 19 shows the result for test case (5). A good correlation value of 0.269 in *X* and 0.3891 in *Y* displacements shows the model works comparatively well compared with the LK in cold-soaked low-contrast conditions.

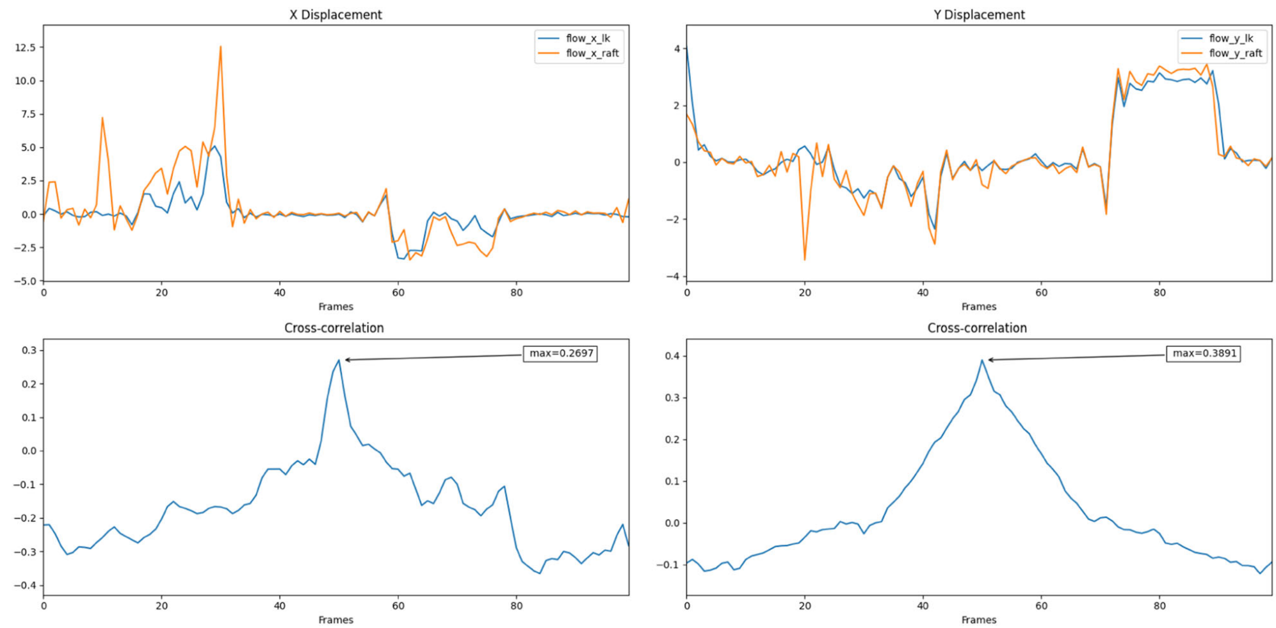

FIGURE 19 *X* and *Y* displacements for the test case (5). The high normalized cross-correlation value indicates the two signals are the same.

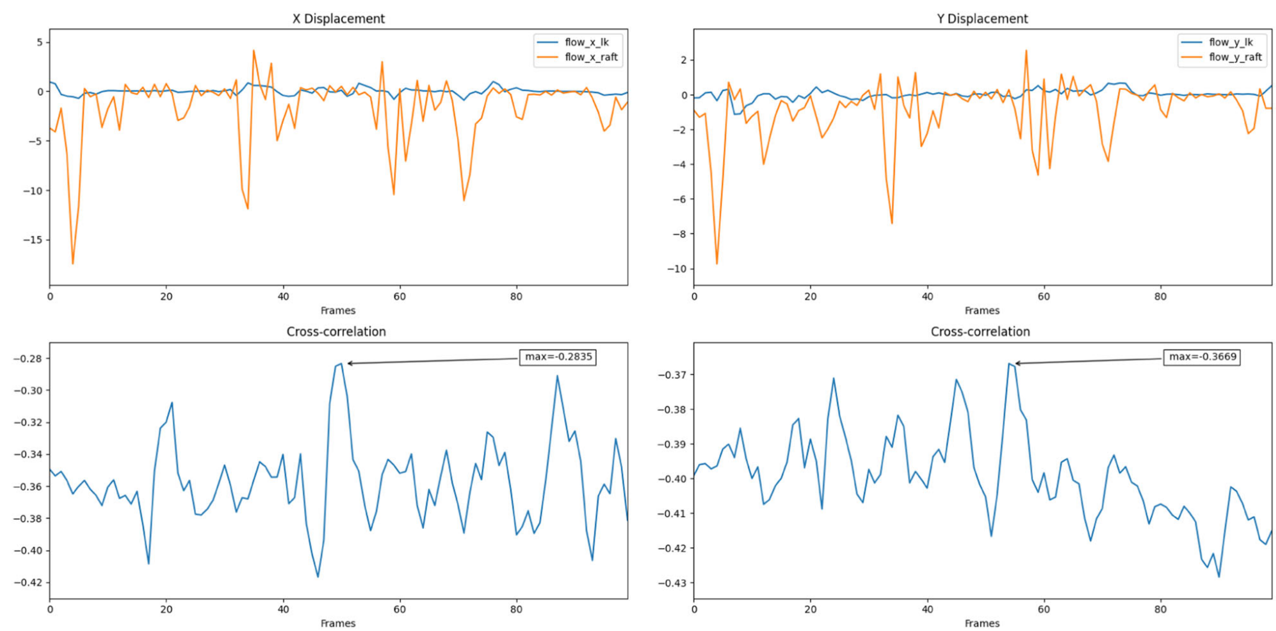

FIGURE 20 *X* and *Y* displacements for the test case (6). The results show that the model is still functional while the LK does not appear to be. LK, Lucas–Kanade.

7.2.3 | Test case (6)

Figure 20 shows the result for test case (6). The signals show that the model still works to some degree in this case while the LK does not. It demonstrates that the neural network was more reliable than the LK in this test.

7.2.4 | Test case (7)

Figure 21 shows the result for test case (7). The measured *X* and *Y* displacements are very low and fluctuate around 0.5 pixels which shows both the neural network and the traditional technique do not work in this case.

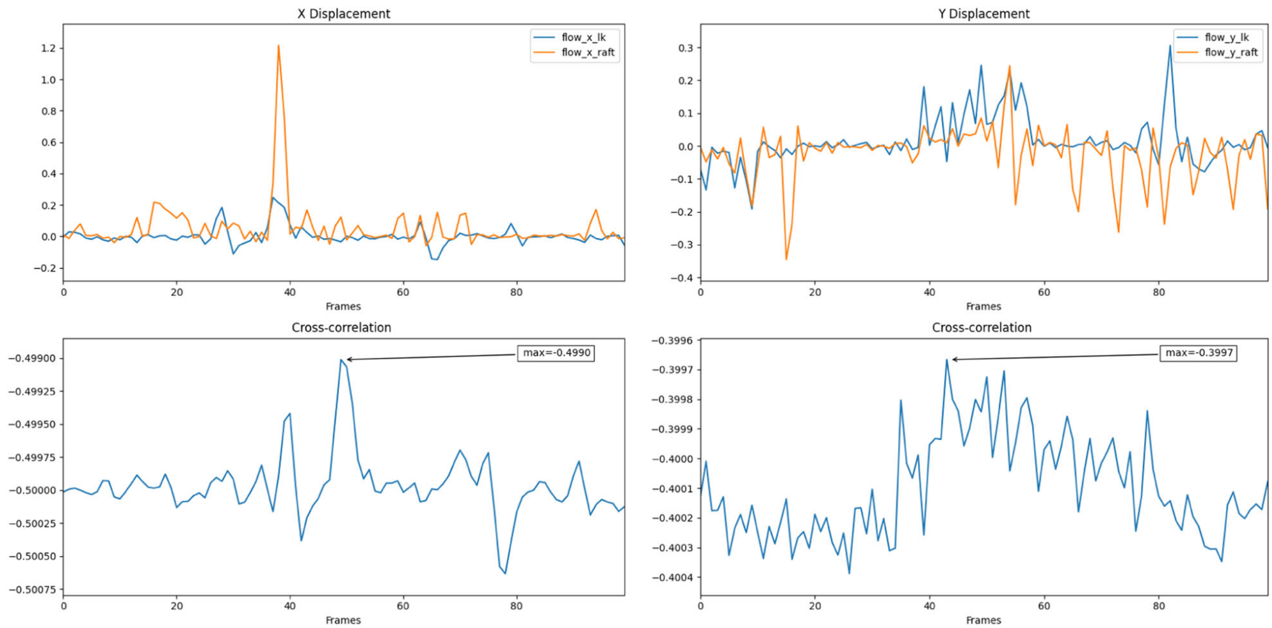

FIGURE 21 *X* and *Y* displacements for the test case (7). It shows that both the neural network and the LK do not work in this case. LK, Lucas–Kanade.

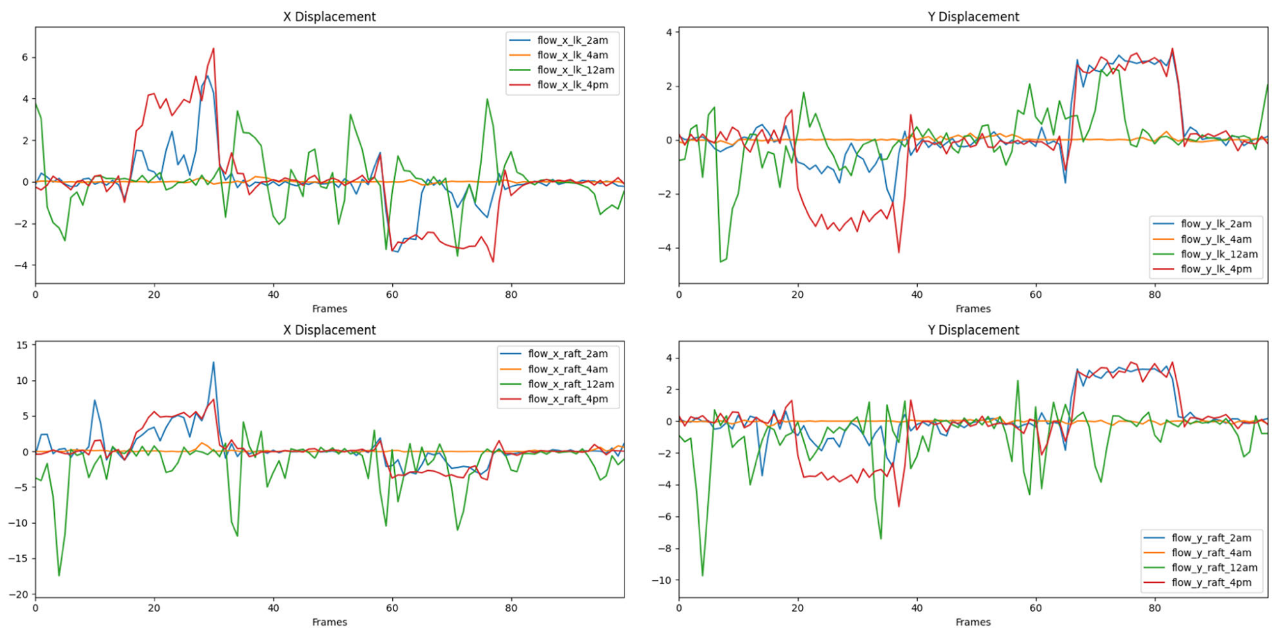

FIGURE 22 Signals from the model as contrast of the scene worsens through the night over time from the highest at 4 p.m.–2 a.m.–12 a.m. to the lowest at 4 a.m.

7.3 | RAFT-s signals from high- to low-contrast conditions

Figure 22 shows the result for overlaid signals from tests (3), (5), (6), and (7), which demonstrates how the model performs over time at the same site. We can see the model performs very well at 4 p.m. (test 3) then slowly decays in cold-soaked conditions at 2 a.m. (test 5), giving some signals at 12 a.m. (test 6), and completely stops functioning at 4 a.m. (test 7).

Table 3 shows a summary of maximum correlation values in *X* and *Y* displacements of the neural network over seven test cases.

Figure 23 shows the change over the ground across 24 h adapted from Nguyen, Rosser, Perera et al. (2022). There are few thermal features over the ground in test (6) when the LK does not function. The RAFT-s, on the other hand, appears to pick up high-level information that allows it to compute optical flow successfully in some instances in this case. Since the network was not trained with low-contrast thermal frames specifically, it can be argued that the color synthetic data set helps the network somewhat with strong generalization to determine optical flow across two very low-contrast scenes in test (6).

In test (7), however, both the LK and the network do not function in this extremely low thermal contrast condition.

TABLE 3 Maximum cross-correlation value of *X* and *Y* displacements of seven test cases with their thermal conditions and data set.

Case	Thermal condition	Data set	Maximum cross correlation of *X* displacement	Maximum cross correlation of *Y* displacement
1	High contrast	[1]	0.8308	0.8301
2	High contrast	[4]	0.9752	0.8699
3	High contrast	[3] at 4 p.m.	0.5276	0.3306
4	Low contrast	[4]	0.5711	0.8246
5	Low contrast	[3] at 2 a.m.	0.2697	0.3891
6	Low contrast	[3] at 12 a.m.	−0.2835	−0.3669
7	Low contrast	[3] at 4 a.m.	−0.4990	−0.3997

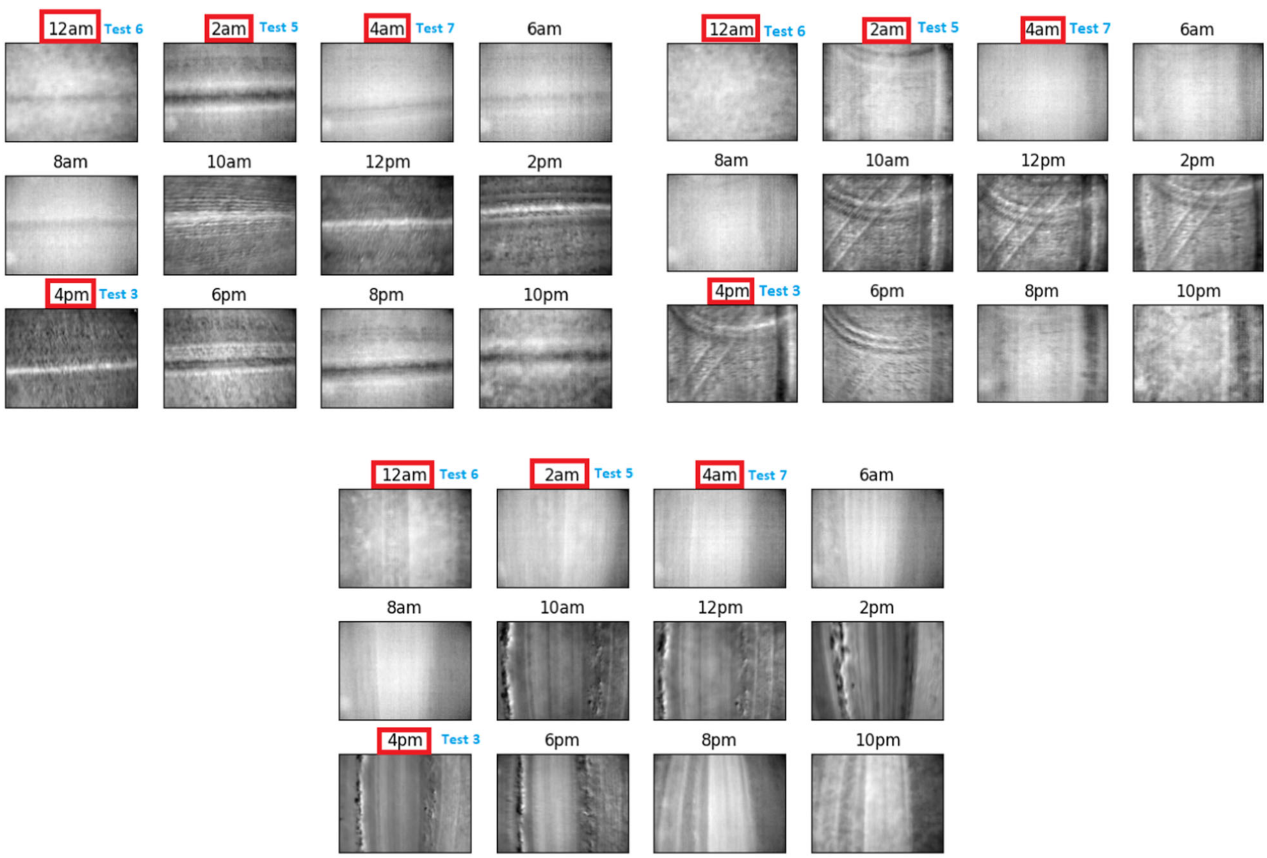

FIGURE 23 Features of the ground over 24 h of the field, with the used sequences are highlighted in red squares.

8 | DISCUSSION

The study shows that the RAFT-s model trained on both RGB synthetics with perfect ground truth, and trained on real-world thermal aerial frames with ground truth generated from traditional techniques, works comparatively well compared with both sparse and dense traditional techniques during hot, high thermal contrast conditions. Very high normalized cross-correlation values shown in Figures 15–17 indicate a very high relation between the two signals.

During the cold-soaked low-contrast conditions, the model works well compared with the I^2A as shown in Figures 18 and 19. Relatively strong normalized cross-correlation values indicate a strong relation between the two signals.

In test (6), the results in Figure 20 show that the model is still functional to some degree, while the LK does not work. In test (7), Figure 21 both the LK and the model do not work in this cold-soaked low thermal condition.

Importantly, the model does not suffer from the aperture problem without being specifically trained for that behavior. This can be explained as the model having strong generalization from the synthetic RGB data.

While the model shows strong performance against the test algorithms, it comes at much higher computational cost. Although the RAFT-s model has the fewest parameters and smallest memory footprint, it cannot be deployed on low-cost and lightweight embedded systems, such as the Pi 4, which limits its use on small UAVs.

It would be interesting to see how the model runs on specific embedded boards designed for deep-learning application.

Limitations are due to the ground truth which is learned from traditional techniques, it may not allow the full potential of the network to be expressed. With good labeled thermal data, the network may outperform traditional techniques in extremely low-contrast condition.

9 | CONCLUSION

This study compares the performance of a thermal flow-based neural network trained with synthetic color and our collected thermal data sets versus traditional optical flow techniques in real-world conditions. The results show that thermal flow from deep-learning network can work well in high-contrast conditions, and even outperforms existing techniques in low-contrast environments.

The pretrained neural network that was trained on the RGB data set must be retrained with thermal data to work well. Imperfect ground truth from traditional dense technique could be used to train the network, possibly with some consequence to ultimate performance.

The results also show that the neural network has strong generalization from synthetic RGB data sets to detect very low thermal contrast features where the traditional optical flow techniques fail to detect.

On the other hand, like other deep-learning models, the RAFT-s neural network requires a dedicated GPU to work well. This potentially limits its use with low-cost and lightweight systems such as the Raspberry Pi 4 or other systems suited to small and micro-UAVs.

Future studies will focus on implementing a much smaller model for thermal aerial application with the goal to further reduce computational requirements while maintaining accuracy.

ACKNOWLEDGMENTS

This research was supported by an Australian Government Research Training Program (RTP) Scholarship. The authors would like to thank Sam Teague from the University of South Australia for his assistance. Open access publishing facilitated by University of South Australia, as part of the Wiley - University of South Australia agreement via the Council of Australian University Librarians.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Tran Xuan Bach Nguyen http://orcid.org/0000-0002-9914-2265
Kent Rosser http://orcid.org/0000-0003-4017-6396

REFERENCES

Abdel-Nasser, M., Moreno, A., Rashwan, H.A. & Puig, D. (2017) Analyzing the evolution of breast tumors through flow fields and strain tensors. *Pattern Recognition Letters*, 93, 162–171.

Alok, N., Krishan, K. & Chauhan, P. (2021) Deep learning-based image classifier for malaria cell detection. *Machine Learning for Healthcare Applications*, 12, 187–197.

Baker, S. & Matthews, I. (2004) Lucas–Kanade 20 years on: a unifying framework. *International Journal of Computer Vision*, 56, 221–255.

Baker, S., Scharstein, D., Lewis, J., Roth, S., Black, M.J. & Szeliski, R. (2011) A database and evaluation methodology for optical flow. *International Journal of Computer Vision*, 92(1), 1–31.

Binder, M.D., Hirokawa, N. & Windhorst, U. (Eds.) (2009) *Aperture problem*. Berlin, Heidelberg: Springer, p. 159.

Bouguet, J.-Y. et al. (2001) Pyramidal implementation of the affine Lucas–Kanade feature tracker description of the algorithm. *Intel Corporation*, 5(1–10), 4.

Bradski, G. (2000) The OpenCV library. *Dr. Dobb's Journal: Software Tools for the Professional Programmer*, 25(11), 120–123.

Brox, T., Bruhn, A., Papenberg, N. & Weickert, J. (2004) High accuracy optical flow estimation based on a theory for warping. In: *European Conference on Computer Vision*. Prague, Czech Republic: Springer, pp. 25–36.

Brox, T. & Malik, J. (2010) Large displacement optical flow: descriptor matching in variational motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(3), 500–513.

Bruhn, A., Weickert, J. & Schnörr, C. (2005) Lucas/Kanade meets Horn/Schunck: combining local and global optic flow methods. *International Journal of Computer Vision*, 61(3), 211–231.

Brunner, C., Peynot, T., Vidal-Calleja, T. & Underwood, J. (2013) Selective combination of visual and thermal imaging for resilient localization in adverse conditions: day and night, smoke and fire. *Journal of Field Robotics*, 30(4), 641–666.

Bureau of Meteorology. (2020) *Wommera weather*.

Bureau of Meteorology. (2022) *On tree hill weather*.

Butler, D.J., Wulff, J., Stanley, G.B. & Black, M.J. (2012) A naturalistic open source movie for optical flow evaluation. In: *European Conference on Computer Vision*. Florence, Italy: Springer, pp. 611–625.

Chahl, J. & Mizutani, A. (2010) Biomimetic attitude and orientation sensors. *IEEE Sensors Journal*, 12(2), 289–297.

Chahl, J.S., Srinivasan, M.V. & Zhang, S.-W. (2004a) Landing strategies in honeybees and applications to uninhabited airborne vehicles. *The International Journal of Robotics Research*, 23(2), 101–110.

Chahl, J.S., Srinivasan, M.V. & Zhang, S.-W. (2004b) Landing strategies in honeybees and applications to uninhabited airborne vehicles. *The International Journal of Robotics Research*, 23(2), 101–110.

Chen, Q. & Koltun, V. (2016) Full flow: optical flow estimation by global optimization over regular grids. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 34, Las Vegas, NV, USA: pp. 4706–4714.

Chen, L., Sun, L., Yang, T., Fan, L., Huang, K. & Xuanyuan, Z. (2017) RGB-T SLAM: a flexible SLAM framework by combining appearance and thermal information. In: *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 5682–5687.

Chen, W., Wang, Y., Chen, H. & Liu, Y. (2022) EIL-SLAM: depth-enhanced edge-based infrared-lidar SLAM. *Journal of Field Robotics*, 39(2), 117–130.

Chokkalingham, S., Chandrasekhar, N. & Vasudevan, M. (2012) Predicting the depth of penetration and weld bead width from the infrared thermal image of the weld pool using artificial neural network modeling. *Journal of Intelligent Manufacturing*, 23, 1995–2001.

Chowdhury, K. & Iraj, M.A. (2020) Predicting autism spectrum disorder using machine learning classifiers. In: *2020 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT)*. Bangalore, India: IEEE, pp. 324–327.

Ciregan, D., Meier, U. & Schmidhuber, J. (2012) Multi-column deep neural networks for image classification. In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*. Providence, RI, USA: IEEE, pp. 3642–3649.

Conroy, J., Gremillion, G., Ranganathan, B. & Humbert, J.S. (2009) Implementation of wide-field integration of optic flow for autonomous quadrotor navigation. *Autonomous Robots*, 27(3), 189.

Dabov, K., Foi, A., Katkovnik, V. & Egiazarian, K. (2006) Image denoising with block-matching and 3D filtering. In: *Image processing: algorithms and systems, neural networks, and machine learning*, Vol. 6064. San Joe, Californita, USA: International Society for Optics and Photonics, p. 606414.

De Pazzi, D., Pertile, M. & Chiodini, S. (2022) 3D radiometric mapping by means of lidar SLAM and thermal camera data fusion. *Sensors*, 22(21), 8512.

Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., Van Der Smagt, P., Cremers, D. & Brox, T. (2015) FlowNet: learning optical flow with convolutional networks. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2758–2766.

Farnebäck, G. (2003) Two-frame motion estimation based on polynomial expansion. In: *Scandinavian Conference on Image Analysis*. Maui, Hi, USA: Springer, pp. 363–370.

Feng, D., Rosenbaum, L. & Dietmayer, K. (2018) Towards safe autonomous driving: capture uncertainty in the deep neural network for lidar 3D vehicle detection. In: *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. Maui, Hi, USA: IEEE, pp. 3266–3273.

FLIR Corporation. (2018) *FLIR Lepton engineering data sheet*. Wilsonville, Oregon, USA: FLIR Corporation.

FLIR Corporation. (2021) Free Teledyne FLIR thermal dataset for algorithm training. In: *FREE Teledyne FLIR thermal data set for algorithm training* [Accessed 2022-03-30].

Fuentes, S., Tongson, E. & GonzalezViejo, C. (2021) Urban green infrastructure monitoring using remote sensing from integrated visible and thermal infrared cameras mounted on a moving vehicle. *Sensors*, 21(1), 295.

Gan, S.K. & Sukkarieh, S. (2011) Multi-UAV target search using explicit decentralized gradient-based negotiation. In: *2011 IEEE International Conference on Robotics and Automation*. Shanghai, China: IEEE, pp. 751–756.

Geiger, A., Lenz, P. & Urtasun, R. (2012) Are we ready for autonomous driving? The KITTI vision benchmark suite. In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*. Providence, RI, USA: IEEE, pp. 3354–3361.

Ghosh, S., Das, N., Das, I. & Maulik, U. (2019) Understanding deep learning techniques for image segmentation. *ACM Computing Surveys (CSUR)*, 52(4), 1–35.

Goceri, E. (2019) Challenges and recent solutions for image segmentation in the era of deep learning. In: *2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA)*. Istabul, Turkey: IEEE, pp. 1–6.

Godard, C., Hedman, P., Li, W. & Brostow, G.J. (2015) Multi-view reconstruction of highly specular surfaces in uncontrolled environments. In: *2015 International Conference on 3D Vision*. Lyon, France: IEEE, pp. 19–27.

Guo, Z., Li, X., Huang, H., Guo, N. & Li, Q. (2019) Deep learning-based image segmentation on multimodal medical imaging. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 3(2), 162–169.

Haq, A.U., Li, J., Memon, M.H., Khan, J., Din, S.U., Ahad, I., Sun, R. & Lai, Z. (2018) Comparative analysis of the classification performance of machine learning classifiers and deep neural network classifier for prediction of Parkinson disease. In: *2018 15th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*. Chengdu, China: IEEE, pp. 101–106.

He, K., Zhang, X., Ren, S. & Sun, J. (2016) Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778.

Hermann, S. & Werner, R. (2013) High accuracy optical flow for 3D medical image registration using the census cost function. In: *Pacific-rim Symposium on Image and Video Technology*. Guanajuato, Mexico: Springer, pp. 23–35.

Honegger, D., Meier, L., Tanskanen, P. & Pollefeys, M. (2013) An open source and open hardware embedded metric optical flow CMOS camera for indoor and outdoor applications. In: *2013 IEEE International Conference on Robotics and Automation*. Karlsruhe, Germany: IEEE, pp. 1736–1741.

Horn, B.K. & Schunck, B.G. (1981) Determining optical flow. In: *Techniques and applications of image understanding*, Vol. 281. International Society for Optics and Photonics, pp. 319–331.

Hui, T.-W. & Loy, C.C. (2020) LiteFlowNet3: resolving correspondence ambiguity for more accurate optical flow estimation. In: *Proceedings of the 16th European Conference on Computer Vision–ECCV 2020, Glasgow, UK, August 23–28, 2020, Part XX*, Vol. 16. Springer, pp. 169–184.

Hui, T.-W., Tang, X. & Loy, C.C. (2018) LiteFlowNet: a lightweight convolutional neural network for optical flow estimation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 8981–8989.

Hui, T.-W., Tang, X. & Loy, C.C. (2020) A lightweight optical flow CNN-revisiting data fidelity and regularization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(8), 2555–2569.

Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A. & Brox, T. (2017) FlowNet 2.0: evolution of optical flow estimation with deep networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2462–2470.

Jasim, M.A. & Al-Tuwaijari, J.M. (2020) Plant leaf diseases detection and classification using image processing and deep learning techniques. In: *2020 International Conference on Computer Science and Software Engineering (CSASE)*. Duhok, Iraq: IEEE, pp. 259–265.

Khattak, S., Nguyen, H., Mascarich, F., Dang, T. & Alexis, K. (2020) Complementary multi-modal sensor fusion for resilient robot pose estimation in subterranean environments. In: *2020 International Conference on Unmanned Aircraft Systems (ICUAS)*. Athens, Greece: IEEE, pp. 1024–1029.

Khattak, S., Papachristos, C. & Alexis, K. (2018) Marker based thermal–inertial localization for aerial robots in obscurant filled environments. In: *International Symposium on Visual Computing*. Las Vegas, NV, USA: Springer, pp. 565–575.

Khattak, S., Papachristos, C. & Alexis, K. (2019a) Keyframe-based direct thermal–inertial odometry. In: *2019 International Conference on Robotics and Automation (ICRA)*. Montreal, QC, Canada: IEEE, pp. 3563–3569.

Khattak, S., Papachristos, C. & Alexis, K. (2019b) Visual-thermal landmarks and inertial fusion for navigation in degraded visual environments. In: *2019 IEEE Aerospace Conference*. Big Sky, MT, USA: IEEE, pp. 1–9.

Khattak, S., Papachristos, C. & Alexis, K. (2020) Keyframe-based thermal–inertial odometry. *Journal of Field Robotics*, 37(4), 552–579.

Kocić, J., Jovičić, N. & Drndarević, V. (2019) An end-to-end deep neural network for autonomous driving designed for embedded automotive platforms. *Sensors*, 19(9), 2064.

Kodali, R.K. & Dhanekula, R. (2021) Face mask detection using deep learning. In: *2021 International Conference on Computer Communication and Informatics (ICCCI)*. Coimbatore, India: IEEE, pp. 1–5.

Krizhevsky, A., Sutskever, I. & Hinton, G.E. (2012) ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 1097–1105.

Kumar, R., Arora, R., Bansal, V., Sahayasheela, V.J., Buckchash, H., Imran, J., Narayanan, N., Pandian, G.N. & Raman, B. (2020) Accurate prediction of COVID-19 using chest x-ray images through deep feature learning model with smote and machine learning classifiers. medRxiv, 2020–04.

Leira, F.S., Helgesen, H.H., Johansen, T.A. & Fossen, T.I. (2021) Object detection, recognition, and tracking from UAVs using a thermal camera. *Journal of Field Robotics*, 38(2), 242–267.

Li, W., Cosker, D. & Brown, M. (2016) Drift robust non-rigid optical flow enhancement for long sequences. *Journal of Intelligent & Fuzzy Systems*, 31(5), 2583–2595.

Liu, Q., Li, X., He, Z., Fan, N., Yuan, D., Liu, W. & Liang, Y. (2020) Multi-task driven feature models for thermal infrared tracking. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34, pp. 11604–11611.

Liu, Q., Lu, X., He, Z., Zhang, C. & Chen, W.-S. (2017) Deep convolutional neural networks for thermal infrared object tracking. *Knowledge-Based Systems*, 134, 189–198.

Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X. & Pietikäinen, M. (2020) Deep learning for generic object detection: a survey. *International Journal of Computer Vision*, 128, 261–318.

Liu, Q., Yuan, D., Fan, N., Gao, P., Li, X. & He, Z. (2022) Learning dual-level deep representation for thermal infrared tracking. *IEEE Transactions on Multimedia*.

Lu, Y., Xue, Z., Xia, G.-S. & Zhang, L. (2018) A survey on vision-based UAV navigation. *Geo-spatial Information Science*, 21(1), 21–32.

Lu, Y., Yi, S., Zeng, N., Liu, Y. & Zhang, Y. (2017) Identification of rice diseases using deep convolutional neural networks. *Neurocomputing*, 267, 378–384.

Maddern, W. & Vidas, S. (2012) Towards robust night and day place recognition using visible and thermal imaging. In: *Proceedings of the RSS 2012 Workshop: Beyond Laser and Vision: Alternative Sensing Techniques for Robotic Perception*. University of Sydney, pp. 1–6.

Menze, M. & Geiger, A. (2015) Object scene flow for autonomous vehicles. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Boston, MA, USA: IEEE, pp. 3061–3070.

Menze, M., Heipke, C. & Geiger, A. (2015) Discrete optimization for optical flow. In: *German Conference on Pattern Recognition*. Springer, pp. 16–28.

Militante, S.V. & Dionisio, N.V. (2020) Real-time facemask recognition with alarm system using deep learning. In: *2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC)*. Shah Alam, Malaysia: IEEE, pp. 106–110.

Miller, A., Miller, B., Popov, A. & Stepanyan, K. (2018) Optical flow as a navigation means for UAV. In: *2018 Australian & New Zealand Control Conference (ANZCC)*. Melbourne, VIC, Australia: IEEE, pp. 302–307.

Minaee, S., Boykov, Y.Y., Porikli, F., Plaza, A.J., Kehtarnavaz, N. & Terzopoulos, D. (2021) Image segmentation using deep learning: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Mouats, T., Aouf, N., Chermak, L. & Richardson, M.A. (2015) Thermal stereo odometry for UAVs. *IEEE Sensors Journal*, 15(11), 6335–6347.

Mouats, T., Aouf, N., Sappa, A.D., Aguilera, C. & Toledo, R. (2014) Multispectral stereo odometry. *IEEE Transactions on Intelligent Transportation Systems*, 16(3), 1210–1224.

Negi, A., Chauhan, P., Kumar, K. & Rajput, R. (2020) Face mask detection classifier and model pruning with keras-surgeon. In: *2020 5th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE)*. Jaipur, India: IEEE, pp. 1–6.

Negi, A. & Kumar, K. (2021a) Classification and detection of citrus diseases using deep learning. In: *Data science and its applications*. Chapman and Hall/CRC, pp. 63–85.

Negi, A. & Kumar, K. (2021b) Face mask detection in real-time video stream using deep learning. *Computational Intelligence and Healthcare Informatics*, 14, 255–268.

Negi, A., Kumar, K., Chaudhari, N.S., Singh, N. & Chauhan, P. (2021) Predictive analytics for recognizing human activities using residual network and fine-tuning. In: *Proceedings of the 9th International Conference on Big Data Analytics, BDA 2021, Virtual Event, December 15–18, 2021*, Vol. 9. Springer, pp. 296–310.

Negi, A., Kumar, K. & Chauhan, P. (2021) Deep neural network-based multi-class image classification for plant diseases. *Agricultural Informatics: Automation Using the IoT and Machine Learning*, 6, 117–129.

Nguyen, T.X.B., Rosser, K. & Chahl, J. (2021) A review of modern thermal imaging sensor technology and applications for autonomous aerial navigation. *Journal of Imaging*, 7(10), 217.

Nguyen, T.X.B., Rosser, K. & Chahl, J. (2022) A comparison of dense and sparse optical flow techniques for low-resolution aerial thermal imagery. *Journal of Imaging*, 8(4), 116.

Nguyen, T.X.B., Rosser, K., Perera, A., Moss, P., Teague, S. & Chahl, J. (2022) Characteristics of optical flow from aerial thermal imaging, "thermal flow". *Journal of Field Robotics*, 39(5), 580–599.

Pathak, A.R., Pandey, M. & Rautaray, S. (2018) Application of deep learning for object detection. *Procedia Computer Science*, 132, 1706–1717.

Perera, A.G., Law, Y.W. & Chahl, J. (2018) Human pose and path estimation from aerial video using dynamic classifier selection. *Cognitive Computation*, 10(6), 1019–1041.

Popov, A., Miller, A., Miller, B. & Stepanyan, K. (2016) Optical flow and inertial navigation system fusion in the UAV navigation. In: *Unmanned/unattended sensors and sensor networks XII*. Vol. 9986. International Society for Optics and Photonics, p. 998606.

Poujol, J., Aguilera, C.A., Danos, E., Vintimilla, B.X., Toledo, R. & Sappa, A.D. (2016) A visible-thermal fusion based monocular visual odometry. In: *Second Iberian Robotics Conference on Robot 2015*. Lisbon, Portugal: Springer, pp. 517–528.

Ranftl, A., Alonso-Fernandez, F., Karlsson, S. & Bigun, J. (2017) Real-time AdaBoost cascade face tracker based on likelihood map and optical flow. *IET Biometrics*, 6(6), 468–477.

Ranjan, A. & Black, M.J. (2017) Optical flow estimation using a spatial pyramid network. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA: IEEE, pp. 4161–4170.

Reddy, K.V. V., Elamvazuthi, I., Aziz, A.A., Paramasivam, S., Chua, H.N. & Pranavanand, S. (2021) Heart disease risk prediction using machine learning classifiers with attribute evaluators. *Applied Sciences*, 11(18), 8352.

Ross, S., Melik-Barkhudarov, N., Shankar, K.S., Wendel, A., Dey, D., Bagnell, J.A. & Hebert, M. (2013) Learning monocular reactive UAV control in cluttered natural environments. In: *2013 IEEE International Conference on Robotics and Automation*. Karlsruhe, Germany: IEEE, pp. 1765–1772.

Rosser, K. & Chahl, J. (2019) Reducing the complexity of visual navigation: optical track controller for long-range unmanned aerial vehicles. *Journal of Field Robotics*, 36(6), 1118–1140.

Rosser, K., Nguyen, T.X.B., Moss, P. & Chahl, J. (2021) Low complexity visual UAV track navigation using long-wavelength infrared. *Journal of Field Robotics*, 38(6), 882–897.

Saputra, M.R.U., Lu, C.X., de Gusmao, P.P.B., Wang, B., Markham, A. & Trigoni, N. (2021) Graph-based thermal–inertial SLAM with probabilistic neural networks. *IEEE Transactions on Robotics*, 38(3), 1875–1893.

Saputra, M.R.U., de Gusmao, P.P., Lu, C.X., Almalioglu, Y., Rosa, S., Chen, C., Wahlström, J., Wang, W., Markham, A. & Trigoni, N. (2020) DeepTIO: a deep thermal–inertial odometry with visual hallucination. *IEEE Robotics and Automation Letters*, 5(2), 1672–1679.

Sethi, S., Kathuria, M. & Kaushik, T. (2021) Face mask detection using deep learning: an approach to reduce risk of coronavirus spread. *Journal of Biomedical Informatics*, 120, 103848.

Shah, S.T. H. & Xuezhi, X. (2021) Traditional and modern strategies for optical flow: an investigation. *SN Applied Sciences*, 3(3), 1–14.

Sharma, S., Kumar, P. & Kumar, K. (2017) Lexer: Lexicon based emotion analyzer. In: *Proceedings of the 7th International Conference on Pattern Recognition and Machine Intelligence, PReMI 2017, Kolkata, India, December 5–8, 2017*. Springer, pp. 373–379.

Sharma, S., Kumar, K. & Singh, N. (2017) D-FES: deep facial expression recognition system. In: *2017 Conference on Information and Communication Technology (CICT)*. Gwalior, India: IEEE, pp. 1–6.

Sharmin, N. & Brad, R. (2012) Optimal filter estimation for Lucas–Kanade optical flow. *Sensors*, 12(9), 12694–12709.

Shin, Y.-S. & Kim, A. (2019) Sparse depth enhanced direct thermal-infrared SLAM beyond the visible spectrum. *IEEE Robotics and Automation Letters*, 4(3), 2918–2925.

Speth, S., Goncalves, A., Rigault, B., Suzuki, S., Bouazizi, M., Matsuo, Y. & Prendinger, H. (2022) Deep learning with RGB and thermal images onboard a drone for monitoring operations. *Journal of Field Robotics*, 39(6), 840–868.

Srinivasan, M.V. (1994) An image-interpolation technique for the computation of optic flow and egomotion. *Biological Cybernetics*, 71(5), 401–415.

Srinivasan, M.V. (2010) Honey bees as a model for vision, perception, and cognition. *Annual Review of Entomology*, 55, 267–284.

Srinivasan, M.V., Zhang, S., Altwein, M. & Tautz, J. (2000) Honeybee navigation: nature and calibration of the "odometer". *Science*, 287(5454), 851–853.

Steinbrücker, F., Pock, T. & Cremers, D. (2009) Large displacement optical flow computation without warping. In: *2009 IEEE 12th International Conference on Computer Vision*. Kyoto, Japan: IEEE, pp. 1609–1614.

Stypułkowski, K., Gołda, P., Lewczuk, K. & Tomaszewska, J. (2021) Monitoring system for railway infrastructure elements based on thermal imaging analysis. *Sensors*, 21(11), 3819.

Sun, D., Yang, X., Liu, M.-Y. & Kautz, J. (2018) PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA: IEEE, pp. 8934–8943.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. (2016) Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 28, Las Vegas, NV, USA: IEEE, pp. 2818–2826.

Teed, Z. & Deng, J. (2020a) RAFT. https://github.com/princeton-vl/RAFT

Teed, Z. & Deng, J. (2020b) RAFT: recurrent all-pairs field transforms for optical flow. In: *European Conference on Computer Vision*. Springer, pp. 402–419.

Tran, V.P., Santoso, F., Garratt, M.A. & Petersen, I.R. (2020) Distributed formation control using fuzzy self-tuning of strictly negative imaginary consensus controllers in aerial robotics. *IEEE/ASME Transactions on Mechatronics*, 26(5), 2306–2315.

Tran, V.P., Mabrok, M.A., Garratt, M.A. & Petersen, I.R. (2021) Hybrid adaptive negative imaginary-neural-fuzzy control with model identification for a quadrotor. *IFAC Journal of Systems and Control*, 16, 100156.

Tran, V.P., Mabrok, M.A., Anavatti, S.G., Garratt, M.A. & Petersen, I.R. (2022) Robust fuzzy q-learning-based strictly negative imaginary tracking controllers for the uncertain quadrotor systems. *IEEE Transactions on Cybernetics*.

Vijayvergia, A. & Kumar, K. (2018) Star: rating of reviews by exploiting variation in emotions using transfer learning framework. In: *2018 Conference on Information and Communication Technology (CICT)*. Jabalpur, India: IEEE, pp. 1–6.

Weibel, T., Daul, C., Wolf, D., Rösch, R. & Guillemin, F. (2012) Graph based construction of textured large field of view mosaics for bladder cancer diagnosis. *Pattern Recognition*, 45(12), 4138–4150.

Wu, B., Iandola, F., Jin, P.H. & Keutzer, K. (2017) Squeezedet: unified, small, low power fully convolutional neural networks for real-time object detection for autonomous driving. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition workshops*. Honolulu, HI, USA: IEEE, pp. 129–137.

Wu, W., Qurishee, M.A., Owino, J., Fomunung, I., Onyango, M. & Atolagbe, B. (2018) Coupling deep learning and UAV for infrastructure condition assessment automation. In: *2018 IEEE International Smart Cities Conference (ISC2)*. Kansas City, MO, USA: IEEE, pp. 1–7.

Wu, X., Sahoo, D. & Hoi, S.C. (2020) Recent advances in deep learning for object detection. *Neurocomputing*, 396, 39–64.

Zhang, Z., Liu, J., Liu, G., Wang, J. & Zhang, J. (2022) Robustness verification of swish neural networks embedded in autonomous driving systems. *IEEE Transactions on Computational Social Systems*.

Zhao, S., Wang, P., Zhang, H., Fang, Z. & Scherer, S. (2020) TP-TIO: a robust thermal–inertial odometry with deep thermal point. In: *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Las Vegas, NV, USA: IEEE, pp. 4505–4512.

Zhao, Z.-Q., Zheng, P., Xu, S.-t. & Xu, X. (2019) Object detection with deep learning: a review. *IEEE Transactions on Neural Networks and Learning Systems*, 30(11), 3212–3232.

Zuo, Z., Liu, C., Han, Q.-L. & Song, J. (2022) Unmanned aerial vehicles: control methods and future challenges. *IEEE/CAA Journal of Automatica Sinica*, 99, 1–14.

How to cite this article: Nguyen, T.X.B., Rosser, K., Perera, A., Moss, P. & Chahl, J. (2023) Neural network-based optical flow versus traditional optical flow techniques with thermal aerial imaging in real-world settings. *Journal of Field Robotics*, 40, 1817–1839. https://doi.org/10.1002/rob.22219