Fire Safety Journal 163 (2026) 104758

Contents lists available at ScienceDirect

Fire Safety Journal

journal homepage: www.elsevier.com/locate/firesaf

Video enhancement for increased spatio-temporal resolution in thermal videos: demonstration on a pool fire☆

Martin Veit [a,b,*], Andrea Lucherini [a,c], Steven Verstockt [d], Bart Merci [b]

[a] Department for Research of Fire-Safe Sustainable Built Environment (FRISSBE), Slovenian National Building and Civil Engineering Institute (ZAG), Slovenia
[b] Department of Structural Engineering and Building Materials, Ghent University, Belgium
[c] Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska, Slovenia
[d] Department of Electronics and Information Systems, IDLab, Ghent University – imec, Belgium

ARTICLE INFO

Keywords:
Thermal camera
Machine learning
Image processing
Pool fire

ABSTRACT

A spatio-temporal video enhancement of a small-scale pool fire is performed to address the typically low spatial resolution and frame rate of inexpensive infrared (IR) cameras. Improving image quality can increase the applicability of low-cost thermal cameras for certain research tasks and analyses. The spatial resolution and frame rate are doubled, from 310 × 250 pixels (px) to 620 × 500 px, and from 25 frames per second (fps) to 50 fps, as well as from 50 fps to 100 fps.

Spatial resolution enhancement is achieved using super-resolution methods based on deep learning, employing several pre-trained models: Fast Super-Resolution CNN (FSRCNN), Efficient Sub-Pixel Convolutional Network (ESPCN), Enhanced Deep Super-Resolution (EDSR), Laplacian Pyramid Super-Resolution Network (LapSRN), and Real-ESRGAN. The footage consists of an *n*-heptane pool fire recorded using a mid-wave infrared (MWIR) FLIR X6981 HS InSb camera. EDSR provides the best performance for both purely resized images and images subjected to complex degradation. For temporal enhancement, a pre-trained frame interpolation model, FLAVR (Flow-Agnostic Video Representation), is used. The resulting interpolated frames appear realistic and preserve the overall flow direction and shape of the flame. The interpolated frames are compared with ground-truth data to validate the accuracy of the temporal enhancement.

1. Introduction

With the increasing quality of thermal cameras, going from cameras with a resolution of 320 x 240 px and a framerate of 30 fps, to cameras with a resolution of 1280 x 1024 px and a framerate exceeding 1000 Hz, such as the FLIR X6981 HS-InSb MWIR, novel types of use cases emerge in research areas related to combustion, diagnostics, and fire safety science and engineering.

In these applications, thermal cameras are used to obtain videos of flames and combustion products, where various techniques can be applied to obtain velocity fields. Clark et al. used thermal cameras on wildland fires and subsequently used an algorithm to obtain a velocity field of crown fires [1], and further improved by Clark and colleagues in Refs. [2–4]. Based on the work from Clark et al., Zhou applied a similar technique to obtain velocity fields from the pool fire [5], which was later compared with measurements from Particle Image Velocimetry (PIV) [6]. Inagaki and Schumacher used a thermal camera for boundary layer flow, albeit not in a combustion context [7–10]. Blunck highlighted the benefit of using thermal cameras for analysis of media with participating flows, including determining velocity fields of pool fires. The technique highlighted by Blunck does not require seeding, but this comes at the expense of increased experimental uncertainty and reduced spatial resolution [11]. Furthermore, due to the larger wavelengths of the infrared (IR) light compared to visible light, IR cameras typically have a reduced spatial resolution due to the complexity and costliness of the sensors in IR cameras [12]. Similarly, they typically have less temporal

This article is part of a special issue entitled: FISJ_IAFSS 2026 published in Fire Safety Journal.

☆ (The author [Andrea Lucherini, Bart Merci] is an editor of this journal. In accordance with policy, [Andrea Lucherini, Bart Merci] was blinded to the entire peer review process).

* Corresponding author. Department for Research of Fire-Safe Sustainable Built Environment (FRISSBE), Slovenian National Building and Civil Engineering Institute (ZAG), Slovenia.
E-mail address: martin.veit@zag.si (M. Veit).

https://doi.org/10.1016/j.firesaf.2026.104758
Received 23 September 2025; Received in revised form 10 February 2026; Accepted 21 March 2026
Available online 3 April 2026

resolution as well, due to a slower response time of the thermal detectors, resulting in a lower framerate. While newer models of thermal cameras can have a high spatial and temporal resolution, these cameras remain costly, making software approaches to increase the resolution of the camera a viable solution to this problem.

Increasing the spatial resolution of low-resolution images is a widely studied research area, with recent advances focusing on machine learning (ML) approaches [13]. Choi et al., e.g., applied convolutional neural networks to enhance the resolution of thermal images for improved object recognition of pedestrians among other objects [14]. Zoetgnande applied super-resolution for improved edge-detection on thermal images [15], and others have done research on super-resolution applied to thermal images from unmanned aerial vehicles [16–18].

Several models have been proposed for traditional image enhancement, each introducing architectural innovations and balancing speed and performance differently. FSRCNN (Fast Super-Resolution CNN) [19] introduced one of the earliest efficient CNN-based pipelines for single-image super-resolution. FSRCNN directly processes the low-resolution image, thereby reducing the computational burden. The network uses a sequence of feature extraction, shrinking, and mapping layers, followed by a learned deconvolution layer for upscaling. This end-to-end architecture allows for real-time inference on modest hardware, although the limited depth reduces its ability to recover fine structural details. ESPCN (Efficient Sub-Pixel CNN) [20] improved efficiency by introducing the sub-pixel convolution layer. Instead of relying on deconvolution, the network predicts a set of feature maps that are rearranged spatially to form a high-resolution output. This approach eliminates checkerboard artefacts often associated with other simple deconvolutions, reduces parameter count, and enables real-time performance. While it is lightweight and fast, ESPCN generally produces less sharp results compared to deeper residual architectures. EDSR (Enhanced Deep Super-Resolution) [21] advanced accuracy by increasing the network depth and width. Building on the ResNet (Residual Neural Network) architecture of learning residual mappings through identity skip connection, EDSR removes batch normalisation layers, which were found to introduce instability and unnecessary overhead in super-resolution tasks. By relying on residual blocks without normalisation, EDSR enables deeper and wider architectures, achieving state-of-the-art results on benchmark datasets. The trade-off is increased computational demand, making EDSR better suited to offline or high-fidelity applications rather than real-time use. LapSRN (Laplacian Pyramid Super-Resolution Network) [22] proposed a progressive, multi-scale framework inspired by Laplacian pyramids. Instead of reconstructing the target resolution in a single step, LapSRN predicts residual images at successive scales and refines them hierarchically. This coarse-to-fine strategy reduces training difficulty and improves convergence, while yielding sharper edges and textures. Furthermore, by producing intermediate results at multiple resolutions, LapSRN supports applications requiring variable scaling factors. Finally, Real-ESRGAN [23] represents a more recent development aimed at addressing the limitations of models trained solely on synthetically degraded data. Real-world low-quality images often contain complex degradations, such as noise, compression artefacts, and motion blur, which traditional models fail to generalise to. Real-ESRGAN extends the ESRGAN framework by adopting a Residual-in-Residual Dense Block generator and training with a realistic degradation pipeline. Coupled with adversarial and perceptual losses, the model prioritises perceptual realism and robustness, producing high-quality reconstructions from heavily degraded inputs.

Furthermore, increasing the temporal resolution of thermal cameras is of particular interest, since thermal cameras typically exhibit a comparably low frame rate compared to digital cameras. Temporal resolution can be increased through hardware upgrades or frame rate upscaling using Video Frame Interpolation (VFI). VFI synthesises intermediate frames to enhance the smoothness of the video, originally focused on digital and artificially created images.

Traditional VFI can struggle with brightness changes, large motions, occlusions, and lighting shifts, leading to artefacts [24]. However, deep learning advances have enabled more accurate synthetic frames [25,26]. VFI research specific to thermal cameras is sparse, with Han et al. benchmarking an optical flow-based method for VFI on an IR video [27]. Extending VFI to thermal videos of flames could boost the temporal resolution of thermal cameras in the research of flames, enabling better capturing of turbulent flame structure and behaviour. Thermal videos of flames exhibit phenomena that are difficult to capture and predict. VFI is typically performed on RGB videos, relying on assumptions such as stable textures, temporal coherence and constant brightness. In contrast, thermal videos of dynamic phenomena, such as flames, exhibit low texture, high noise, and highly non-rigid turbulent motion with evolving structures. These characteristics can create difficulties for VFI of thermal videos, as they violate common modelling assumptions.

These advances are tested on thermal videos of a heptane pool fire, in the light of future research focusing on turbulent flame structures, velocity estimation and sharpened visual observations.

Although recent work has demonstrated the feasibility of joint spatio-temporal video super-resolution within a single end-to-end model [28–30], this study adopts a modular approach by separating spatial super-resolution and temporal interpolation. This allows each component to be independently selected, evaluated, and replaced, making the analysis clearer of their respective effects on thermal flame imagery and enabling integration with different state-of-the-art models. This modularity is advantageous given the limited availability of MWIR data and the lack of established benchmarks for combined spatio-temporal enhancement for thermal videos of flames.

This paper investigates various models for Single Frame Super Resolution and their suitability for enhancing thermal images, while connecting the methodology with video frame interpolation to obtain a spatio-enhanced thermal video, increasing both the spatial and temporal resolution. This could then be used for improved tracking of turbulent structures, performing high-resolution velocimetry measurements, or for improved observational capabilities.

This work investigates, using a rigorously validated framework, the feasibility and limitations of deep learning-based spatial and temporal enhancement methods for thermal video data, so that thermal (IR) imaging can be exploited better than is done today when experiments or fire tests are executed. This type of video footage can indeed be used to reconstruct flow fields and flame boundaries (rather than provide accurate temperature information in the gas phase). However, thermal cameras often suffer from relatively low resolution. Hence, it is deemed useful and important to (artificially) upgrade their spatial and temporal resolution. As a test configuration, a well-controlled pool fire has been selected, as this configuration has already been widely studied in the past, and the physics, flow field, and flame shape have been well-characterised [31]. The goal of the study is thus not to improve understanding or quantification of the phenomena involved, nor to argue that thermal imaging would be advantageous over other techniques (such as chemiluminescence). Rather, the objective is, starting from the given fact that most fire laboratories are equipped with thermal cameras, to demonstrate options to gain more information and insights from their images than is done to date.

Section 2 describes the experimental setup and enhancement methodology, Section 3 presents the results and discusses their implications and limitations, and Section 4 concludes the paper.

2. Methodology

2.1. Image degradation model

Given a reference video obtained from a thermal camera (FLIR X6981-HS InSb), image degradation is performed on the images to emulate a realistic lower-quality image obtained from a camera of lower quality, as is commonly done in the literature [19–23], with either

simple rescaling or using a more complex degradation model, as shown in the following. This is done to conclude if spatial enhancement can be performed on these images to obtain a higher quality image compared to the original one. The image data used in this study consist of grayscale 8-bit images. These images represent thermal intensity values and contain no RGB colour information.

To create this degradation, a standard degradation model is used, with the aim of synthesising a low-resolution image from a reference image. The standard degradation models consist of multiple steps, where a blur kernel is convolved on the image, after which the image is resized, noise is added to the image, and a standard JPEG compression is applied [32,33]:

$$\boldsymbol{x} = \mathrm{D}(\boldsymbol{y}) = [(\boldsymbol{y} \otimes \boldsymbol{k}) \cdot r + \boldsymbol{n}]_{JPEG} \quad (1)$$

In Eq. (1) x is the degraded image, $D(y)$ is the degradation model applied to the original image y, k is a blur kernel, r is a scaling factor, and n is additive Gaussian noise.

The use of synthetic degradation enables paired evaluation with full-reference image quality metrics, whereas validation using real low-resolution sensors would necessitate no-reference assessment with higher associated uncertainty. Moreover, the use of no-reference image quality assessment remains limited in thermal imaging, particularly in experimental fire science, due in part to the lack of structured datasets and comprehensive validation studies [34].

To downsample (i.e., resize) the image, a linear interpolation algorithm is used to resize the image, using a scaling factor of 0.5, corresponding to a camera with a 310 x 250 px resolution. The blur degradation is performed using a linear blur kernel. This is done using anisotropic Gaussian filters, with a 5 x 5 blur kernel, corresponding to moderate blur. Noise is included as additive Gaussian noise, where the noise is sampled from a Gaussian distribution, and the intensity of the noise is based on the standard deviation. Grey noise is synthesised by applying the sampled noise on all three colour channels, with a standard deviation of 2 of the pixel brightness values, corresponding to small amounts of noise. Finally, JPEG compression is performed, which is a commonly used technique for digital images, where an image quality factor between 0 and 100 is defined: lower values indicate higher compression ratios and reduced quality [35]. An image quality factor of 80 % is used in this study, corresponding to moderate compression. An example of a degraded image is shown in Fig. 1, with a scaling factor of 0.5.

2.2. Spatial enhancement of images

Spatial enhancement of images can be achieved through both traditional interpolation techniques and modern deep learning-based approaches. Interpolation relies on predefined mathematical kernels and typically produces overly smooth results, whereas convolutional neural networks learn data-driven mappings from low-to high-resolution images, enabling the reconstruction of high-frequency details and textures.

In this work, several super-resolution models are evaluated using publicly available pretrained weights, without retraining or fine-tuning. FSRCNN [19] and ESPCN [20] are lightweight architectures designed for computational efficiency. FSRCNN was trained on the Set5 dataset, while ESPCN was trained on the higher-quality DIV2K dataset [36], allowing it to learn richer spatial representations despite its relatively shallow structure.

EDSR [21] and LapSRN [22] emphasise reconstruction fidelity over speed. EDSR employs deep residual learning and was trained on DIV2K [36], a large high-resolution dataset with diverse natural textures. LapSRN adopts a multi-scale Laplacian pyramid framework and was originally trained on the Set5 dataset, progressively refining image resolution across scales.

Real-ESRGAN [23] is designed to be robust to real-world degradations rather than idealized downsampling. It was trained on a combination of DIV2K [36], Flickr2K [37], and the OutdoorSceneTraining (OST) dataset [38] using degradation-aware pipelines that include noise, blur, and compression artefacts, enabling perceptually convincing reconstructions under challenging conditions.

Each of these models is applied to thermal images for a reference case study to determine if these models can perform well on these types of scenarios. All spatial and temporal enhancement computations were performed on a Lenovo ThinkStation P3 workstation equipped with an Intel Core i7-13700 CPU, 64 GB of system memory, and an NVIDIA GeForce RTX GPU.

2.2.1. Metrics to compare spatial image quality

To objectively evaluate the effectiveness of each method, various well-established full-reference image quality assessment (FR-IQA) metrics for image quality enhancement are defined. By effectiveness, it is meant how well the models perform at enhancing an image that is quantitatively similar to that of a ground truth image. These include Peak-Signal-To-Noise Ratio (PSNR), Mean Squared Error (MSE), and the computational time per frame [39]. PSNR determines the ratio between the maximum power of an image and the power of noise in the image. Therefore, a high value indicates a higher quality of image reconstruction. This is calculated as:

$$PSNR = 20 \log_{10}(MAX_I) - 10 \log_{10}(MSE) \quad (2)$$

In Eq. (2) MAX_I is the maximum pixel value of the image, and MSE is the Mean Squared Error. The Mean Squared Error (MSE) for an image, which is also used as a metric, is calculated as follows:

$$MSE = \frac{1}{MN} \sum_{n=0}^{M} \sum_{m=1}^{N} [\hat{g}(n,m) - g(n,m)]^2 \quad (3)$$

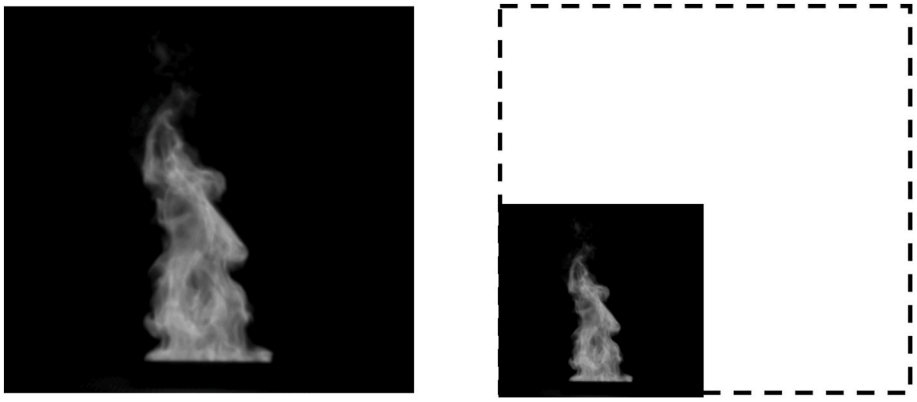

Fig. 1. Image before and after performing image degradation. Left) Before (620 x 500 p). Right) After degradation (310 x 250 p).

Here, M is the number of rows of pixels in the image, N is the number of columns of pixels in the image, $\widehat{g}(n,m)$ is the matrix data of the original image, and $g(n,m)$ is the matrix data of the enhanced image.

The computational time per frame is also used as a metric, since a high number of images have to be enhanced for videos, meaning that computational resources should be considered.

Each of the metrics are used as their implementation from the open-source Python package Open Computer Vision Library (OpenCV) [40]. The mean of the metrics is defined for 100 distinct frames to determine which of the models performs best.

2.2.2. Metrics to compare temporal image quality

The Feature Similarity Index (FSIM) [41] is used to assess the perceptual quality of interpolated frames against ground truth data. Unlike error-based measures such as PSNR, FSIM is designed around features that align with the human visual system. Its primary component, phase congruency, identifies perceptually important structures such as edges, corners, and textures, which are important for evaluating interpolation quality since motion discontinuities and structural boundaries must be preserved. The secondary component, gradient magnitude, captures local contrast and edge sharpness, making FSIM sensitive to blurring artefacts that commonly arise in frame interpolation. The final FSIM score reflects the similarity of structural and textural information between the interpolated frame and the reference, ranging from 0 to 1, with higher values indicating greater perceptual fidelity. By emphasising the preservation of structural detail and sharpness, FSIM provides a more meaningful assessment of interpolation quality than purely pixel-level metrics.

2.3. Temporal enhancement of videos

A higher frame rate of flame structures allows for a better understanding of how the flame behaves and makes it easier to track turbulent structures. To increase the frame rate artificially, a Deep-Learning technique named FLAVR [26] is used, using a pre-trained model. The pre-trained model is applied on a video of a heptane pool fire to obtain increased temporal resolution, 2x in the present study. This method can be used for fast-frame interpolation. FLAVR (Flow-Agnostic Video Representations for Fast Frame Interpolation) [26] is a Deep-Learning model designed for video frame interpolation, trying to address the limitations of optical flow-based interpolation methods.

Traditional techniques rely on estimating pixel movements using optical flow between frames, while FLAVR predicts intermediate frames without flow computation, making it more robust to motion blur, hidden objects in the scene, and complex scenes. The model is based on a 3D convolutional neural network that processes multiple consecutive frames simultaneously, learning spatio-temporal representations to artificially create intermediate frames. By using multi-scale processing and densely stacked residual blocks, FLAVR can capture short-range and long-range motion dynamics. As reported by the original authors, the model is trained using reconstruction-based loss functions combined with perceptual loss terms and is trained until convergence on Vimeo-90K [42]. This architecture allows the model to interpolate frames at high speeds, making it suitable for real-time video applications. FLAVR is trained using a combination of reconstruction loss and perceptual loss to ensure pixel-level accuracy and perceptual quality of generated frames. The method is flow-agnostic, meaning that it does not need motion estimation to interpolate intermediate frames. This reduces computational complexity and improves performance in complex scenarios, such as fast motion or deformable objects, such as flames.

Due to the high frame-rate of the footage obtained, using the MWIR thermal camera, the video can be downsampled from a high frame-rate (100 fps), to 25 and 50 fps, meaning that ground-truth data is available for each of these downsampled frame-rates. This not only allows for a qualitative evaluation to see if the interpolated frames exhibit temporal coherence, but the interpolated frames can also be compared directly with the ground-truth data.

2.4. Case study – pool fire

To investigate the effectiveness of the techniques, a case study is considered. It consists of an n-heptane (n-C_7H_{16}) pool fire in a square steel container with a side length of 25 cm, 5 cm lip height and a thickness of approximately 2 mm. Approximately 300 ml of heptane was used, ignited by a pilot flame. The pool fire was recorded in ambient conditions (ambient temperature 15 ± 2 °C and relative humidity 40 ± 10 %), with quiescent flow.

The combustion products for complete combustion of heptane are CO_2 and H_2O, with CO_2 having a primary absorption band around 4.6 μm and H_2O having a primary absorption band around 2.7 μm and another strong absorption band at 6.3 μm [43].

A factory-calibrated MWIR camera, with a spectral range of 3 - 5 μm, a thermal sensitivity of 20 mK and a temperature range between −20 °C and 3000 °C, was used to capture high-speed thermal imaging of the pool fires.

The camera was positioned at a fixed distance of 5.7 m from the pool fire. The high-resolution camera, with a resolution of up to 640 × 512 p, allowed fine-scale thermal mapping of the flame. The frame rate was 100 Hz to capture transient flame dynamics, which is then downsampled for analysis. An example of a frame obtained from the heptane pool fire is shown in Fig. 2.

It should be noted that the proposed approach operates on MWIR image sequences without explicit separation of soot and gas-phase radiation, and its current validation is limited to n-heptane flames.

2.5. Pool fire dynamics

A pool fire consists of three distinct regimes: the persistent flame, the intermittent flame and the buoyant plume. Since a MWIR thermal camera is used for obtaining a thermal video, both the radiation from the flame and soot particles are visible, as well as the combustion gases in the buoyant plume. For a pool fire of the present size, established correlations can be used to estimate the characteristic flame behaviour. Based on the pool fire diameter, D, and fuel type (n-heptane), the average flame height can be estimated following Heskestad's correlation for flame height of pool fires [44]. For the pool fire dimensions of 0.25 × 0.25 m, the expected average flame height is approximately 1.3 m, which is consistent with what is visible in the recordings. The Zukoski number [45] for the pool fire is determined to be 2.46. A second key parameter is the puffing frequency, which sets the dominant timescale of large-scale flame oscillations. Zukoski et al. showed that the puffing frequency scales primarily with the fire diameter [45]. For the pool fire used in this study, this yields an expected puffing frequency of 2.9 ± 0.2 Hz. The puffing frequency provides an indication of the characteristic time scale of the pool fire.

3. Results and discussion

3.1. Spatial enhancement of images

Examples of spatially enhanced images are shown in Fig. 3, where the top-most image is the purely resized images, upscaled using each super-resolution model, and a reference image. The bottom images are images with degraded images using the aforementioned degradation model, upscaled to 620 x 500 px. Both rows have a zoomed-in area to highlight the performance of each model.

The enhancement techniques show a clear and realistic increase in the resolution. While most of the methods perform comparably, Real-ESRGAN smoothens the image and overestimates the pixel brightness values, as shown by the zoomed-in area. As shown in Fig. 3f, the change in pixel-value between the flame and ambient air is sudden, whereas for the other techniques the change in pixel brightness is gradual. Capturing

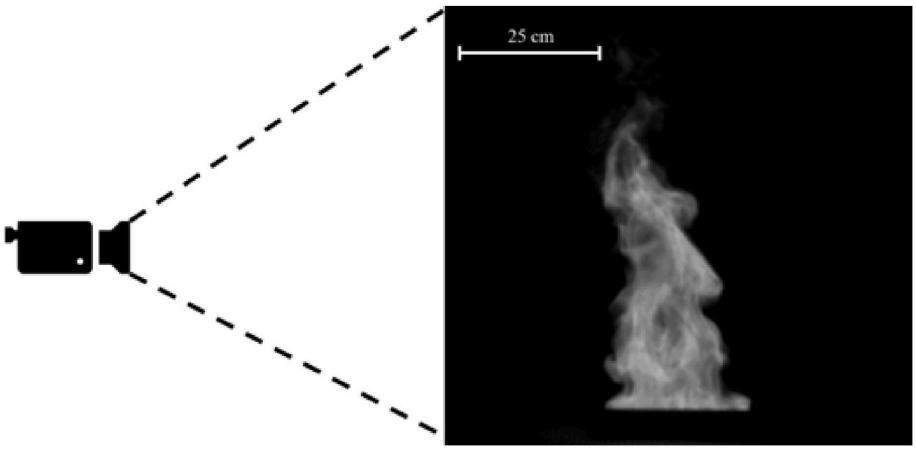

Fig. 2. Example of a thermal image for a heptane pool fire.

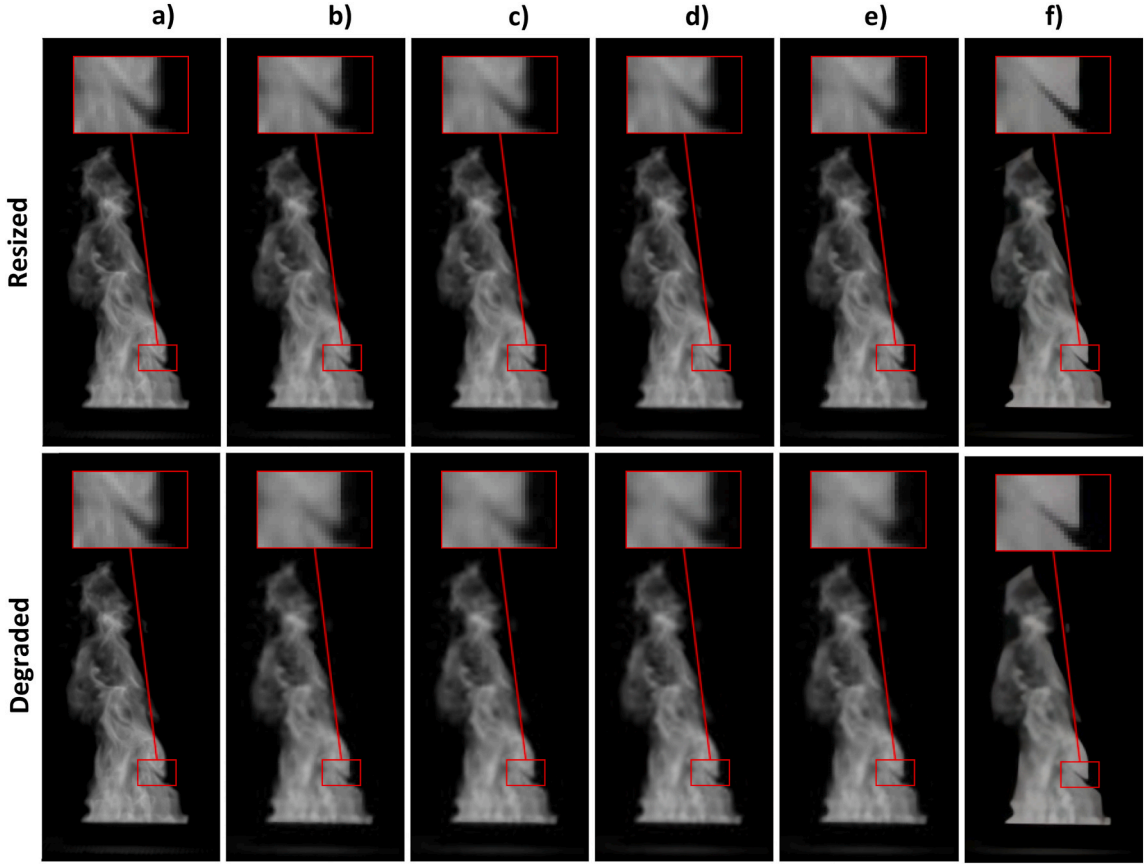

Fig. 3. Purely resized images (top) and degraded images (bottom). a) Ground-truth reference image. b) FSRCNN. c) ESPCN. d) EDSR. e) LapSRN. f) Real-ESRGAN.

realistic temperature fields is important for temporal frame interpolation, as well as estimating velocity fields from the movement of

Table 1
Overview of performance for each image enhancement technique. The arrows (↑, ↓) indicate the favorable value (high or low), with the best performing model highlighted with bold text. *Normalised MSE by number of pixels.

Technique	PSNR ↑ [dB]		MSE* ↓		Computational time ↓ [ms]	
	$\boldsymbol{y}\cdot r$	$\mathrm{D}(\boldsymbol{y})$	$\boldsymbol{y}\cdot r$	$\mathrm{D}(\boldsymbol{y})$	$\boldsymbol{y}\cdot r$	$\mathrm{D}(\boldsymbol{y})$
FSRCNN	26.52 ± 0.47	26.33 ± 0.48	0.146 ± 0.16	0.152 ± 0.17	35.64 ± 1.72	37.08 ± 1.71
ESPCN	26.57 ± 0.47	26.39 ± 0.48	0.144 ± 0.16	0.150 ± 0.17	32.31 ± 1.84	32.98 ± 1.87
EDSR	**26.59 ± 0.47**	**26.41 ± 0.48**	**0.143 ± 0.16**	**0.149 ± 0.16**	5707 ± 154.79	5779.13 ± 158.14
LapSRN	26.57 ± 0.48	26.41 ± 0.48	0.144 ± 0.16	0.150 ± 0.17	226.28 ± 4.69	227.92 ± 6.56
Real-ESRGAN	25.71 ± 0.47	25.45 ± 0.46	0.175 ± 0.19	0.186 ± 0.20	**8.46 ± 0.98**	**8.48 ± 1.04**

the flame and combustion gases, hence why Real-ESRGAN is deemed ineffective in this use-case.

The spatial enhancement of the images is evaluated using the methods and metrics described earlier. An overview of this is shown in Table 1.

From Table 1, the difference between the various techniques becomes more evident. For each metric (PSNR, MSE, computational time), the table is split into two columns: one related to applying the techniques on images with the degradation model, and the other related to pure resizing without blurring, random noise or JPEG compression applied.

For the PSNR, it is favorable to have higher values, as this suggests that less noise is present in the image, and therefore a higher image quality. For the MSE, it is favorable to have lower values, suggesting smaller absolute differences between two images. Finally, a lower computational time is favorable, as this means it is faster to compute each frame. The favorable outcome of the metrics is illustrated using arrows in Table 1.

The metrics show that EDSR is the most computationally heavy method to apply, as each frame takes significantly longer to compute. However, it also performs the best with regards to PSNR and MSE. However, ESPCN also performs well, while being significantly faster. Therefore, ESPCN method could be used if many frames have to be enhanced. While Real-ESRGAN was designed to perform well on images with complex degradation, it performs worst overall. This is also in line with the previous qualitative investigation, where it seems to over-predict pixel brightness. For the pure resizing, the image quality predicted by PSNR and MSE is higher than the degraded images. However, the PSNR and MSE are only slightly higher for the resized images compared to the degraded images, meaning that the models perform well even when the images have undergone complex image degradation.

3.2. Temporal enhancement of images

Besides comparing the interpolated frames with ground-truth data, qualitative measures are also considered, such as temporal consistency. This means that the interpolated frames should exhibit a 'realistic' progression in time for the flame structure and movement of the combustion gases. As such, hot gases should not travel downwards for this specific case, and the shape of the flame and combustion gases should be retained for the interpolated frames, without hallucinating interpolation artefacts. For ease of observation, contours are overlaid on the grey-scale images, but no significant effort has gone into obtaining accurate temperature measurements, as it is the techniques and methodology that are important in this study, rather than accurate gas temperatures.

The result of the temporal enhancement is shown in Fig. 4, where subsequent images have been used to interpolate in-between frames yielding a 2x temporal enhancement. With a frame rate of 25 fps, the time between each image is 40 ms. The total time from start to end of Fig. 4 is 200 ms. The temperature isotherms are highlighted with contours to more easily evaluate the temporal coherence and small-scale flame structures.

As shown in Fig. 4, the interpolated frames (highlighted with a red box), show a realistic progression in time compared to adjacent frames. The plume is travelling upwards with a distance that is between the ground-truth frames, and no significant artefacts are visible in the interpolated images, meaning that the fast-frame interpolation technique yields structures that follow a natural progression. Based on the interpolation using a base frame-rate of 25 fps, the isotherms are less detailed compared to the ground truth data, meaning that the temperature fields of the interpolated frames are less sharp and blurry.

While the technique shows qualitatively accurate interpolation during part of the puffing cycle, it is relevant to examine how the model performs across successive oscillations. Fig. 5 illustrates a sequence of frames following the cycle shown in Fig. 4, corresponding to the decay of one puffing oscillation and the onset of the next. Even during this transition, where large-scale structural changes in the flame and hot combustion gases occur, the temporal interpolation reproduces the motion without introducing qualitatively noticeable artefacts.

Overall, the temporal enhancement is successful in the case of a small n-heptane pool fire. For fast-moving flames or other phenomena with a small characteristic time scale, a higher frame rate would be needed for the interpolation to be successful.

3.2.1. Validation of temperature fields for interpolated frames

To validate the interpolated frames, they are compared with the ground truth data. The comparison is conducted quantitatively, comparing the distributions and ability to capture small-scale phenomena, as well as using the FSIM metrics. To do this, a video at 100 fps is downsampled to two different frame rates: 25 fps and 50 fps, and then upscaled to 50 fps and 100 fps using frame interpolation. The upscaled videos contain both ground truth frames and interpolated frames, so only the interpolated frames are compared with ground truth data. This analysis is performed at two different time scales since the movement of the flame will move a larger distance for low frame rates, making it difficult to accurately interpolate sharp frames. Similarly, for higher frame rates, the interpolation should be sharper. The interpolated frames, and corresponding ground truth frames, from a video upscaled from 25 fps to 50 fps is shown in Fig. 6.

As shown in the figure, the interpolated frames overall exhibit the same temperature distributions as the corresponding ground truth frames. However, the frames appear to be less sharp in detail, with

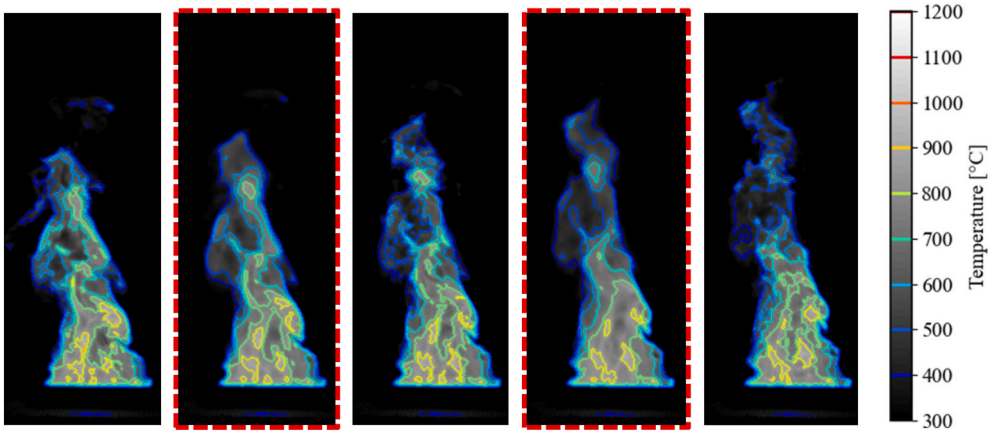

Fig. 4. Temporal interpolation within a puffing cycle of the pool fire. The frames enclosed in red boxes are interpolated results, generated between two adjacent recorded frames. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

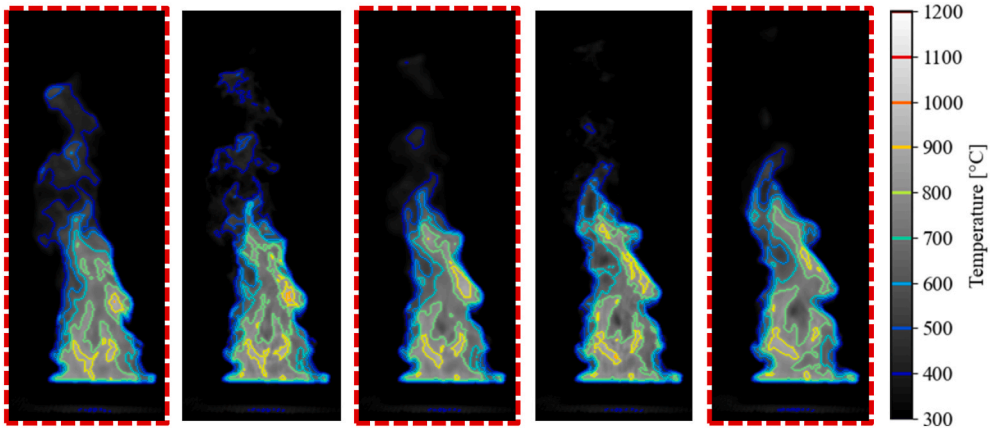

Fig. 5. Temporal interpolation across the transition between two successive puffing cycles. The frames enclosed in red boxes are interpolated results, generated between two adjacent recorded frames. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

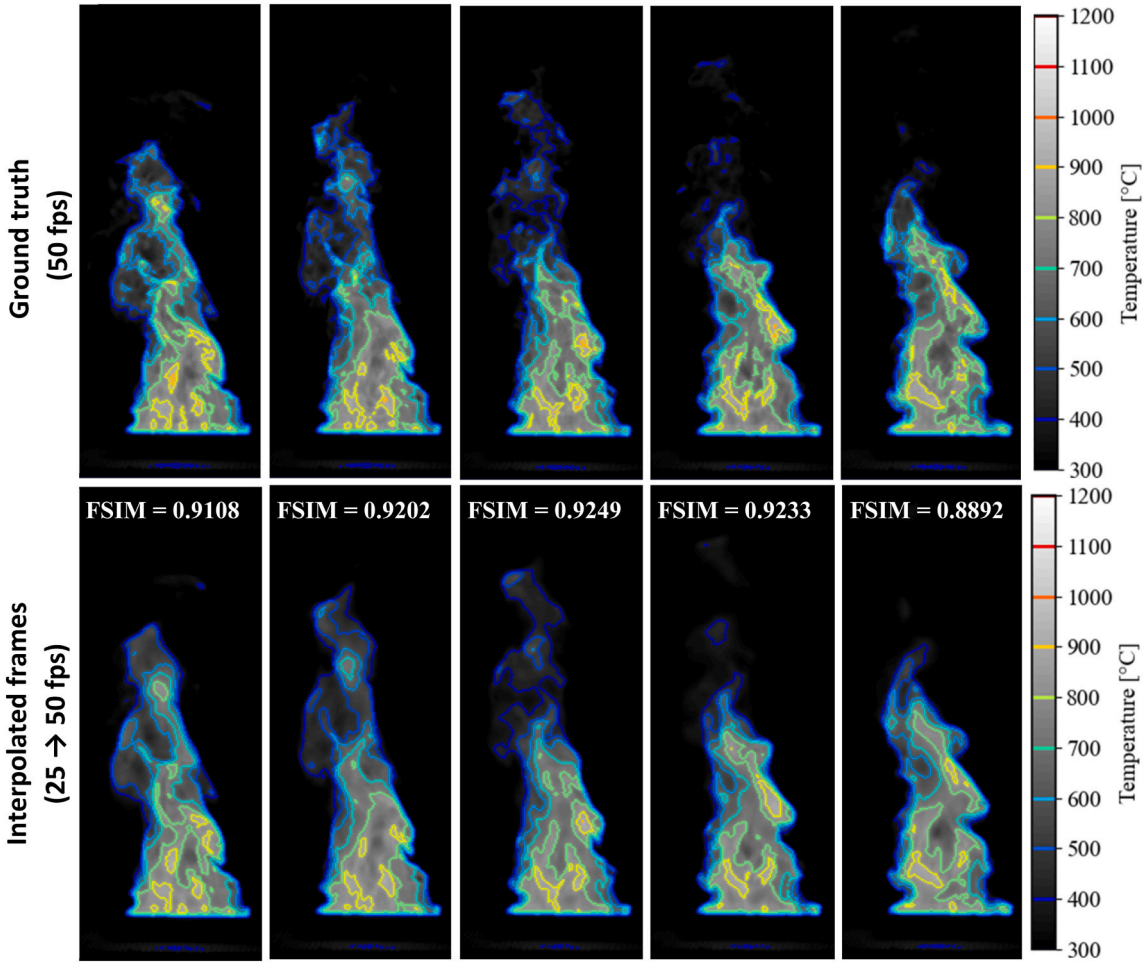

Fig. 6. Comparison between real frames (top) and interpolated frames (bottom). The interpolated frames are from a video at 25 fps, upscaled to 50 fps.

slightly averaged temperature fields, illustrated by the lower degree of detail in the contours.

Furthermore, FSIM is calculated between all the interpolated frames and the ground truth frames and shown at the top of the interpolated frames. The interpolated frames obtain an average value for FSIM of 0.91 for all frames. As the range of the FSIM is 0 to 1, the fidelity for the interpolated frames is close to that of the ground truth frames, but with notable blurriness that lowers the score. Furthermore, significant differences in the plume region can be seen in Fig. 6 for certain images, which is attributed to the low initial frame rate and thus large distance travelled in-between frames, making accurate frame interpolation difficult for the model.

To evaluate how the averaging effect from fast-frame interpolation of pool fires affects the interpolated frames on different time scales, a video upscaled from 50 fps to 100 fps is investigated as well, and shown in Fig. 7.

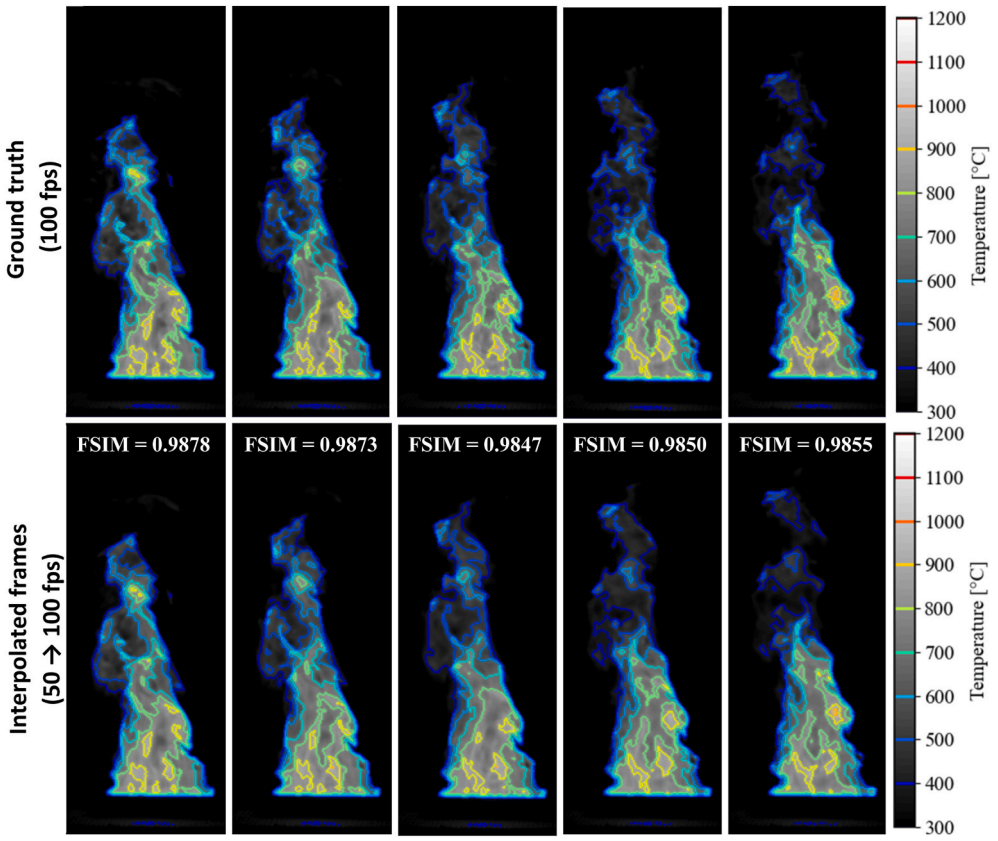

Fig. 7. Comparison between ground truth frames (top) and interpolated frames (bottom). The interpolated frames are from a video at 50 fps, upscaled to 100 fps.

As shown in the figure, an initial frame rate of 50 fps works better for fast-frame interpolation, as the temperature field is sharper, more detailed and consistent with the ground truth data. Based on the frame rate and puffing frequency, the fast-frame interpolation model performs well when the frame rate is 20 times that of the puffing frequency of the pool fire.

With an average FSIM of around 0.99 for all interpolated frames, it performs notably better when using a frame rate of 50 fps and interpolating to 100 fps. This is compared to the FSIM of around 0.91 when using a video of 25 fps and upscaling to 50 fps. At 0.99, the FSIM indicates high-quality temporal interpolation, with less blur compared to the video at 25 fps.

Based on the frames shown, temporal consistency is preserved, and image-level fine details of small-scale flame features are captured within the spatial and temporal sampling limits of the data, with the degree of detail dependent on the video frame rate. As indicated by the FSIM, the model, when used on a video upscaled from 50 fps to 100 fps, performs notably better than when used on a video upscaled from 25 fps to 50 fps, with an FSIM of 0.99 and 0.91, respectively.

To further validate the interpolated frames, a two-sample Kolmogorov-Smirnov (KS) test was used to compare the pixel-intensity distributions, testing the null hypothesis that the ground-truth and interpolated samples are drawn from the same underlying distribution [46]. The test consistently yields a small test statistic ($D \approx 0.02$). However, when evaluated on large pixel counts, the p-value becomes significant due to sample size, whereas subsampled tests fail to reject the null hypothesis, indicating negligible practical differences between the distributions.

3.2.2. *Preservation of flow directionality*

While the previous investigations showed temporal consistency and consistency in the temperature distribution of individual pixel-values, it is important to investigate if small-scale 'jitter' and bias occur on small scales. To investigate this, the principle of optical flow is used, which is used to determine pixel displacement between subsequent frames.

Optical flow can serve as a quantitative metric for assessing the success of video frame interpolation, as it directly reflects the temporal coherence of motion between consecutive frames. Farnebäck's optical flow algorithm [47] is utilised, which is a polynomial expansion method that approximates local neighbourhoods with quadratic polynomials to estimate per-pixel displacement fields. Accurate interpolation should yield intermediate frames whose motion field is consistent with that of the ground-truth image sequence. Deviations between the optical flow computed on interpolated sequences and that obtained from ground-truth frames can thus be used to reveal temporal inconsistencies, such as flickering, jitter, or hallucinated motion. A low discrepancy between the optical flow determined from ground-truth data and interpolated data indicates that the interpolated frame preserves both the direction and magnitude of the underlying motion, while high discrepancies suggest temporal artefacts or motion patterns that are inconsistent with the observed physical evolution of the flame. In Fig. 8, the flow field for interpolated frames is determined (bottom), as well as the flow field for the corresponding ground truth frames (bottom).

The optical flow algorithm computed for both the ground truth data at 100 fps, as well as on the video upscaled from 50 fps to 100 fps, shows that the flow can be determined quantitatively on interpolated frames and has high correspondence with ground truth optical flow. The optical flow captures small-scale phenomena, such as disappearing flow structures and local velocities changing direction. Therefore, even when examining the interpolated frames against ground truth data on pixel level, they keep the directionality and magnitude of the optical flow.

The results of the analysis show that spatial and temporal resolution can be enhanced, while retaining realism for both enhancements. While spatial enhancement showed some improvement using the Deep-Learning-based approach, the considerable computational time per frame makes the approach expensive for long sequences of frames if

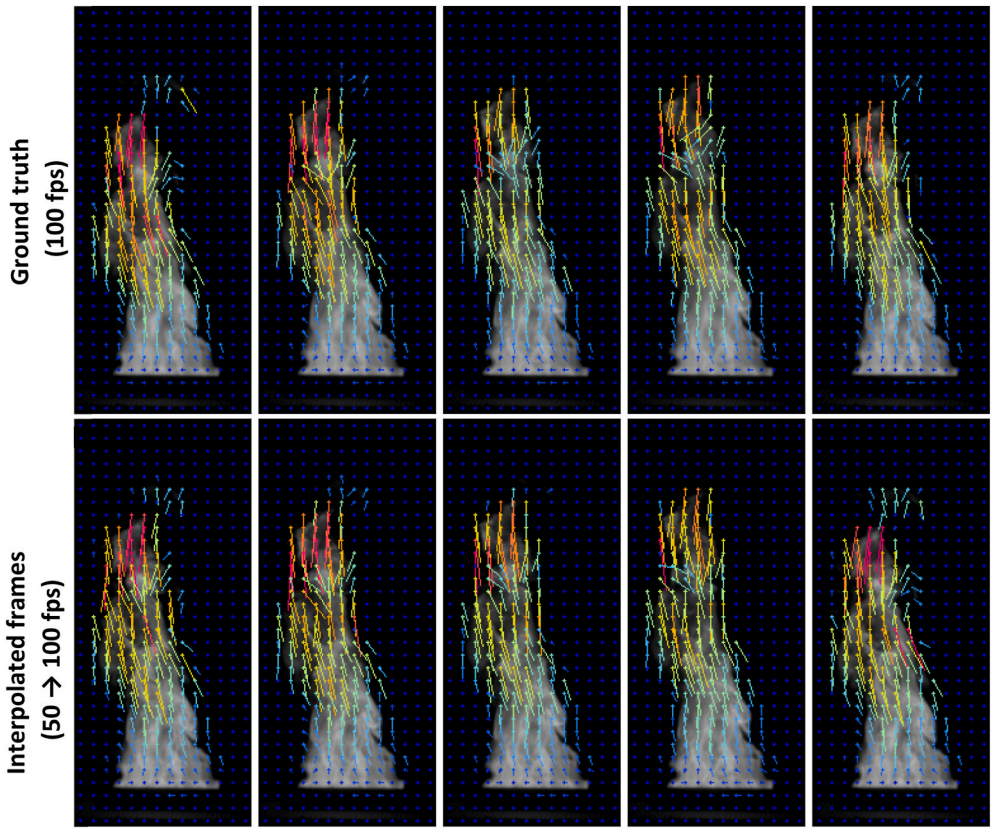

Fig. 8. Comparison between computed optical flow for real frames (top) and interpolated frames (bottom) using Farnebäck's optical flow algorithm. The interpolated frames are from a video at 50 fps, upscaled to 100 fps.

EDSR is used. While most of the techniques performed similarly, Real-ESRGAN is considered not effective for this task, as it smoothens the images significantly, yielding a different temperature distribution of the flame. Furthermore, none of these techniques are trained on datasets consisting of thermal images or flames. Training a model on a dataset of thermal images of fires could be useful for additional improvement, which could be done using the existing underlying architectures of the models.

The temporal interpolation technique, FLAVR, demonstrated reliable performance in enhancing the frame rate of a thermal camera from 25 fps to 50 fps, both within a puffing cycle and during the transition between successive cycles. While the temperature distributions are smoothed at this frame rate, an upscaled video from 50 fps to 100 fps showed sharper interpolated frames, as indicated by using FSIM as a metric for image quality. When using FSIM on videos upscaled from 25 to 50 fps, the average FSIM was 0.91, while it was 0.99 when comparing images from a video upscaled from 50 to 100 fps with ground truth data. The direction of the flow is as expected and temporally similar to that of ground truth data, and the temperature fields for interpolated frames are similar to those of ground truth data.

Beyond improvements in image quality, enhanced spatial resolution has implications for experimental fire research by improving the visualisation of flame boundaries, plume structure, and localised intensity gradients. Such improvements may support qualitative interpretation of flame shape dynamics and plume morphology in experiments where optical access or sensor resolution is limited. However, these enhancements should be interpreted as improved visualisation rather than direct recovery of unresolved physical details.

The relevance of spatial and temporal enhancement for fire behaviour research is inherently tied to the physical characteristics of the fire under investigation. A limitation of the present study is that it considers a single fuel and pool diameter, whereas both parameters significantly influence plume dynamics and MWIR image characteristics. The dominant puffing frequency scales with pool diameter according to the Zukoski correlation [45], with larger pools exhibiting lower characteristic frequencies and larger coherent structures. As a result, the frame rate required for successful interpolation is expected to scale inversely with pool diameter, independently of the imaging system. Fuel type further impacts MWIR observability, as non-sooty fuels produce optically thin, low-contrast plumes dominated by gas-phase emission, while sooty hydrocarbon fuels yield brighter, optically thick flames with higher contrast but reduced internal visibility and increased risk of pixel saturation. Consequently, the frame rate thresholds reported here are specific to the conditions studied, and extension to other fuels and pool diameters, including crude-oil fires, is left for future work.

Another limitation of the current study is the chosen degradation model and the associated hyperparameters. Simulating a camera of lower quality is a complex task, as the real degradation is complex and difficult to predict. Therefore, future work should include a wide parametric sweep of the hyperparameters of the degradation model and compare the degraded images with a camera of lower quality, as well as considerations for more complex models that better simulate the different aspects of the degradation model, such as noise modelling [48]. Having a more accurate degradation model will reveal which of the proposed techniques performs the best with a known degradation.

Overall, the results indicate that deep learning-based spatial and temporal enhancement can improve the interpretability of thermal fire imaging under controlled conditions, particularly for qualitative assessment of flame dynamics and plume structure. The primary contribution of the proposed approach lies in extending the utility of existing experimental datasets when sensor resolution or frame rate is limited, rather than in recovering unresolved physical detail. At the same time, the dependence on sampling frequency, degradation assumptions, and learned reconstruction highlights the need for cautious

interpretation of enhanced imagery. These findings position AI-based enhancement as a complementary visualisation and analysis tool for experimental fire research, rather than a replacement for physical measurement.

4. Conclusions

The purpose of the study was to quantify the possible enhancement of the spatial and temporal resolution of thermal videos for applications in fire science. In particular, within this research, a 0.25 × 0.25 m n-heptane pool fire was studied. Various Deep-Learning models were tested and compared using common image quality metrics, PSNR and MSE. The interpolation models were used on degraded images using a standard degradation model, along with pure resizing of the images. The results showed that the EDSR model performed the best overall, albeit at the cost of a higher computational time per frame. The worst-performing model was Real-ESRGAN, which smoothed the images and over-predicted pixel brightness. While the models performed better when upscaling images that were purely resized, the difference between the PSNR of the resized and degraded images is less than 1 %, indicating that the models perform even with the applied degradation model.

For temporal enhancement, the model was applied to videos of different frame rates, and showed that low frame-rate videos, such as 25 fps, can be upscaled to 50 fps and show images that are similar to ground truth images when comparing temperature fields and motion of the flame and buoyant plume, albeit with slightly averaged temperature fields. However, upscaling a video from 50 fps to 100 fps showed high correspondence with ground-truth frames. Furthermore, optical flow was used to investigate if jitter or flickering at small scales could be observed. This was not the case, meaning that the interpolation was successful at capturing the motion fields of the pool fire. When interpolating from 25 to 50 fps, the Fidelity Similarity Index was calculated to be 0.91, compared to a value of 0.99 when interpolating from 50 to 100 fps. This is due to the blur that is present, especially for the lower frame rate. The studied pool fire had a puffing frequency of approximately 2.9 Hz. As the video at 25 fps was blurry, indicated by the contours and the FSIM value, while the video at 50 fps was not, the fast frame interpolation model performs well when the frame rate is approximately 20 times higher than that characteristic time scale of the flame, compared to the worse performance when the frame rate was only approximately 10 times the characteristic time scale of the flame.

With the proposed techniques, it is still unfeasible to use them in real-time on the actual device doing the video capturing. The techniques are instead proposed as a post-processing step in the analysis pipeline, to improve the capacity to extract useful information from the video, or to use it for detailed velocimetry measurements using optical techniques.

In conclusion, Deep-Learning-based approaches have been shown to be effective in both spatial and temporal resolution enhancement on sequences of thermal images obtained from a small pool fire. Future work should focus on training models specifically on data obtained from thermal images of fires.

CRediT authorship contribution statement

Martin Veit: Writing – original draft, Visualization, Methodology, Investigation, Data curation. **Andrea Lucherini:** Writing – review & editing, Supervision. **Steven Verstockt:** Writing – review & editing, Supervision, Methodology. **Bart Merci:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to gratefully acknowledge the financial support for the FRISSBE project within the European Union's Horizon 2020 research and innovation programme (GA 952395).

References

[1] T.L. Clark, L. Radke, J. Coen, D. Middleton, Analysis of small-scale convective dynamics in a crown fire using infrared video camera imagery, J. Appl. Meteorol. 38 (1999) 1401–1420.
[2] Coen, J., “Infrared imagery applied for insights into wildland fire dynamics”, 5th Symposium on Fire and Forest Meteorology, Orlando, Florida, USA.
[3] J. Coen, S. Mahalingam, J. Daily, Infrared imagery of crown-fire dynamics during FROSTFIRE, J. Appl. Meteorol. 43 (2004) 1241–1259.
[4] T.L. Clark, M.J. Reeder, M. Griffiths, D. Packham, N. Krusel, Infrared observations and numerical modelling of grassland fires in the Northern Territory, Australia, Meteorol. Atmos. Phys. 88 (2005) 193–201.
[5] X. Zhou, L. Sun, S. Mahalingam, D.R. Weise, Thermal particle image velocity estimation of fire plume flow, Combust. Sci. Technol. 175 (2003) 1293–1316.
[6] L. Sun, X. Zhou, S. Mahalingam, D.R. Weise, Experimental investigation of the velocity field in buoyant diffusion flames using PIV and TPIV algorithm, Fire Saf. Sci. 8 (2005) 939–950.
[7] A. Inagaki, M. Kanda, S. Onomura, H. Kumemura, Thermal image velocimetry, Bound.-Layer Meteorol. 149 (2013) 1–18.
[8] A. Inagaki, Application of the thermal image velocimetry to measure and visualize spatial distribution of near surface wind, J. Jpn. Soc. Civ. Eng. 29 (2016) 186–195.
[9] Schumacher, B., Katurji, M., Zhang, J., Stiperski, I., Dunker, C., "Evolution of Micrometeorological Observations: Instantaneous Spatial and Temporal Surface Wind Velocity from Thermal Image Processing", *Geocomputation,* Queenstown, New Zealand.
[10] Schumacher, B., Katurji, M., Zhang, J., Zawar-Reza, P., Adams, B., Zeeman, M., “Adaptive thermal image velocimetry of spatial wind movement on landscapes using near-target infrared cameras”, Atmos. Meas. Tech. 15: 5681–5700.
[11] D.L. Blunck, Review: applications of infrared thermography for studying flows with participating media, Exp. Therm. Fluid Sci. 130 (2021) 110502.
[12] A. Rogalski, Infrared detectors: status and trends, Prog. Quant. Electron. 27 (2003) 59–210.
[13] C. Dong, C.C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, IEEE T. Pattern Anal. 38 (2025) 295–307.
[14] Choi, Y., Kim, N., Hwang, S., Kweon, I. S. “Thermal Image Enhancement Using Convolutional Neural Network”, IEEE/RSJ International Conference on Intelligent Robots and Systems, Daejeon, Korea, pp 223-230.
[15] Y.W.K. Zoetgnande, J.-L. Dillenseger, J. Alirezaie, Edge focused super-resolution of thermal images. Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), 2019, pp. 1–8. Budapest, Hungary, paper N-19505.
[16] R.S. Puttagunta, B. Kathariya, Z. Li, G. York, Multi-scale feature fusion using channel transformers for guided thermal image super resolution. Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW), 2024. Vancouver, Canada, paper 00314.
[17] Y. Sang, et al., A lightweight network with latent representations for UAV thermal image super-resolution, IEEE T. Geosci. Remote. 62 (2024) 1–11.
[18] Z. Zhao, C. Wang, C. Li, Y. Zhang, J. Tang, Modality conversion meets superresolution: a collaborative framework for high-resolution thermal UAV image generation, IEEE T. Geosci. Remote. 62 (2024) 1–14.
[19] Dong, C., Loy, C. C., Tang, X, “Accelerating the super-resolution convolutional neural network”, Computer Vision–ECCV 2016: 14Th European Conference, Amsterdam, The Netherlands, pp 391-407.
[20] Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D., Wang, Z., “Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Nevada, USA, pp 1874-1883.
[21] Lim, B., Son, S., Kim, H., Nah, S., Lee, K. M., “Enhanced deep residual networks for single image super-resolution”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp 136-144.
[22] W. Lai, J. Huang, N. Ahuja, M. Yang, ”Fast and accurate image super-resolution with deep laplacian pyramid networks”, IEEE Trans. Pattern Anal. Mach. Intell. 41 (2018) 2599–2613.
[23] Wang, X., Xie, L., Dong, C., Shan, Y., "Real-ESRGAN: training real-world blind super-resolution with pure synthetic data", Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, Canada, pp 1905-1914.
[24] S. Baker, D. Scharstein, J.P. Lewis, A database and evaluation methodology for optical flow, Int. J. Comput. Vis. 92 (2011) 1–31.
[25] J. Dong, M. Dong, Video frame interpolation: a comprehensive survey, ACM Trans. Multimed Comput. Commun. Appl 19 (2023) 1–31.
[26] T. Kalluri, D. Pathak, M. Chandraker, D. Tran, FLAVR: flow-agnostic video representations for fast frame interpolation. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023. Waikoloa, Hawaii, USA.
[27] Han, T., Chen, S., Shi, J., “Frame rate Up-Conversion algorithm based on infrared image”, Proceedings of the 13th International Conference on Computer Modeling and Simulation, Melbourne, Australia, pp 225-230.
[28] X. Xiang, Y. Tian, Y. Zhang, Y. Fu, J.P. Allebach, C. Xu, Zooming SlowMo: an efficient one-stage framework for space-time video super-resolution, arXiv preprint arXiv:2104.07473 (2021).

[29] S.Y. Kim, J. Oh, M. Kim, Fisr: deep joint frame interpolation and super-resolution with a multi-scale temporal loss, Proc. AAAI Conf. Artif. Intell. 34 (7) (2020, April) 11278–11286.
[30] Y. Zhang, D. Yang, Z. Chen, W. Ding, Continuous space-time video super-resolution with multi-stage motion information reorganization, ACM Trans. Multimed Comput. Commun. Appl 20 (9) (2024) 1–23.
[31] Y. Chen, J. Fang, X. Zhang, Y. Miao, Y. Lin, R. Tu, L. Hu, Pool fire dynamics: principles, models and recent advances, Prog. Energy Combust. Sci. 95 (2023) 101070, https://doi.org/10.1016/j.pecs.2022.101070.
[32] M. Elad, A. Feuer, "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images", IEEE Trans. Image Process. 6 (1997) 1646–1658.
[33] C. Liu, D. Sun, On bayesian adaptive video super resolution, IEEE T. Pattern Anal. 36 (2014) 346–360.
[34] M. Veit, A. Lucherini, G. Jomaas, B. Merci, Measuring Thermal Image Quality for Fire Service Applications, NFPA Fire Protection Research Foundation, Quincy, MA, USA, 2025.
[35] G.K. Wallace, The JPEG still picture compression standard, Commun. ACM 34 (1991) 30–44.
[36] E. Agustsson, R. Timofte, Ntire 2017 challenge on single image super-resolution: dataset and study, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 126–135.
[37] R. Timofte, E. Agustsson, L. Van Gool, M.H. Yang, L. Zhang, Ntire 2017 challenge on single image super-resolution: methods and results, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 114–125.
[38] X. Wang, K. Yu, C. Dong, C.C. Loy, Recovering realistic texture in image super-resolution by deep spatial feature transform, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 606–615.
[39] U. Sara, M. Akter, M.S. Uddin, Image quality assessment through FSIM, SSIM, MSE and PSNR—A comparative study, J. Comput. Commun. 7 (2019) 8–18.
[40] G. Bradski, The OpenCV library, Dr. Dobb's J. Softw. Tools 120 (2000) 122–125.
[41] L. Zhang, X. Mou, FSIM: a feature similarity index for image quality assessment, IEEE Trans. Image Process. 20.8 (2011) 2378–2386.
[42] T. Xue, B. Chen, J. Wu, D. Wei, W.T. Freeman, Video enhancement with task-oriented flow, Int. J. Comput. Vis. 127 (8) (2019) 1106–1125, https://doi.org/10.1007/s11263-018-01144-2.
[43] P. Lindstrom, G. Mallard, The NIST chemistry WebBook: a chemical data resource on the internet, J. Chem. Eng. Data 46 (2001).
[44] G. Heskestad, "Fire Plumes, Flame Height, and Air Entrainment", SFPE Handbook of Fire Protection Engineering, fifth ed., National Fire Protection Association, Quincy, MA, 1995.
[45] E.E. Zukoski, Properties of fire plumes, in: Combustion Fundamentals of Fire, 1995.
[46] F.J. Massey Jr., The kolmogorov-smirnov test for goodness of fit, J. Am. Stat. Assoc. 46 (253) (1951) 68–78.
[47] G. Farnebäck, Two-frame motion estimation based on polynomial expansion. Scandinavian Conference on Image Analysis, 2003. Berlin, Germany.
[48] A. Maleky, S. Kousha, M.S. Brown, M.A. Brubaker, Noise2noiseflow: realistic camera noise modeling without clean images, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 17632–17641.