Data Challenge 2024: Denoising Thermal Images.

Submitted By:

Amrin AKTER |Hajar TOUBALI |Lin SUN |Yulun WEI

Supervised By:

Phlypo RONALD |Guyader NATHALIE |Harant OLIVIER

DALETh

Grenoble INP - Phelma, UGA

7th June, 2024

Contents

1 Introduction

Thermal imaging technology, used in medical diagnostics, industrial inspections, and surveillance, detects infrared radiation and displays it as color images showing thermal variations. Uncooled thermal sensors, common in these systems, respond to both scene and camera temperatures, resulting in noisy, low signal-to-noise ratio (SNR) images, necessitating advanced correction algorithms. Our data challenge, presented by Lynred, focuses on improving image quality from raw thermal images in shadowy scenes. We aimed to enhance image clarity using Non-Uniform Correction (NUC) for fixed pattern noise and Temperature Emissivity Correction (TEC) for environmental factors, employing an affine model to reduce analogue pixel responses. We worked with three sequences of raw images and calibration images taken under various conditions to apply and refine NUC techniques, addressing pixel-specific variations. This report details our methodology, challenges, and results, aiming to produce accurate and visually appealing images that reveal hidden details in the thermal data.

2 Theory and Background

The NUC Model employs Calibration-Based and Scene-Based algorithms for correcting blurry thermal images. Calibration-Based NUC acquires sensor output at varying temperatures or integration times, averaging over multiple frames to reduce noise. Scene-Based NUC adapts correction using scene statistics, assuming uniform exposure for all detectors. Performance evaluation utilizes real infrared data, assessing parameters like residual non-uniformity (RNU) (Marcotte et al., 2013) to gauge correction effectiveness.(Tremblay et al., 2010) In the context of NUC Model (Kumar, 2013), the methodologies akin to calibration-based and scene-based techniques. Calibration-Based NUC involves acquiring sensor output at varying temperatures or integration times, akin to two-point calibration methods. Conversely, Scene-Based NUC, mirroring Kumar's approach, utilizes scene information to correct non-uniformities algorithmically.

Correcting a Blurry Thermal Image using Non-Uniformity Correction (NUC). It is a model that assumes a linear relationship between the raw pixel value and the actual scene irradiance or temperature. The model has two important parameters - responsivity, or the ability of the thermal sensors to respond to the changes in temperature at different temperatures of the Focal Plane Array , and offset is the inherent bias present in the sensor's output signal when there is no external radiated input, due to factors such as the sensor's own electronic characteristics and noise. This is the baseline output of the sensor. Calibration results at controlled temperatures offered by the black bodies. The NUC process attempts to eliminate the impact of the camera temperature. It is the raw image pixel value minus the offset, divided by the responsivity.

$$S(t) = \frac{\text{Raw}(t) - \text{Offset}}{\text{Resp}}$$

This method predicts the actual irradiance or temperature in the scene. There are three levels in our data challenge that are provided to assess performance: one acquired under the same conditions as in the calibration set 1 , one acquired under dissimilar conditions , using calibration set 2 with missing information for some images and finally the last level is without any calibration set. The calibrated data should be the same as the actual environment in order to meet Joao's model accuracy. Also, the non-linear relation between the temperature and irradiance, or the lack of calibration data, can affect the correction. There are also opportunities for researchers to explore non-linear regression, data augmentation, and deep learning to overcome these limitations and to improve NUC robustness for thermal images.

3 Methodology

3.1 Task 1

In the first task, we were provided with a calibration dataset for calculating the responsivity and offset of thermographic sensors. We can observe almost no readable information from the raw image of the scene dataset, which appears to be a uniform greyscale map with some vertical texture. The calibration dataset provides 9 sets of raw images taken before different black bodies temperatures at different T_{fpa} temperatures. The blackbody temperatures (T_{cn}) mostly range between 10°C and 50°C. Our goal is to correct these scenes with the data provided in calibration to reveal the temperature information hidden behind the large amount of noise.

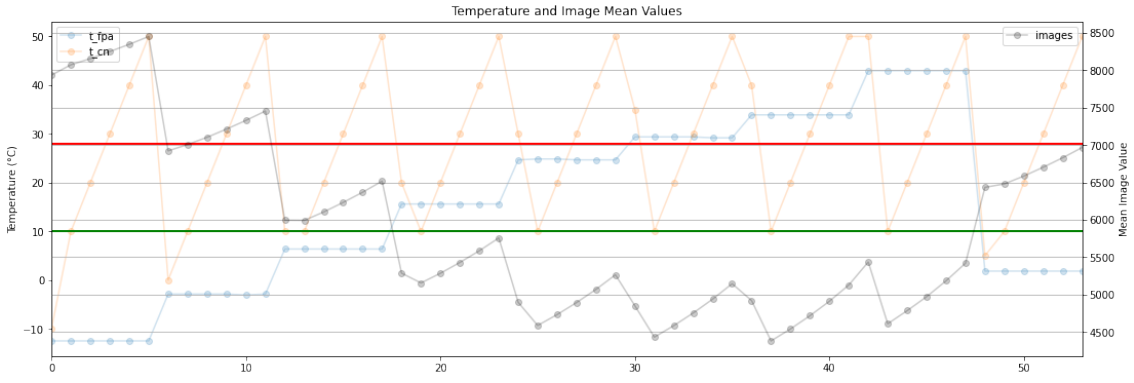


Figure 1: T_fpa, T_cn and the mean value of the raw images from the first calibration dataset

3.1.1 Abandoned tracks

Linear Regression

Linear regression checks each pixel's intensity across all calibration images, capturing temperature fluctuations. This model captures changes in pixel intensity with temperature oscillations.

$$I = \text{Resp} \cdot T + \text{Offset}$$

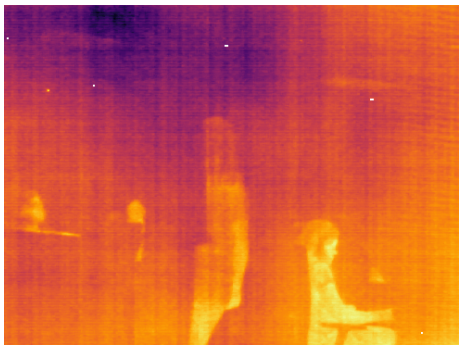


Figure 2: Result of a corrected image of Scene 1 using Linear regression.

The slope (responsivity) indicates the pixel's sensitivity to temperature changes, and the intercept (offset) describes the bias in pixel response. These values are stored in maps and applied to scene images for Non-Uniform Correction (NUC), enhancing image quality. However, this method assumes a linear relationship between temperature and intensity, which may not always hold. Its precision depends on calibration data quality and is computationally expensive, especially for large datasets.

Linear relationship of Raw image to T_{cn}

After carefully reading the given theoretical information, we concluded that the Raw image should have a linear relationship with the Scene temperature, and for the calibration dataset, that is, there should be a linear relationship between the Raw image and T_{cn}. Plotting the useful information in the calibration dataset, we can observe that there are several sets of data that are linearly related, i.e., T_{cn} increases from 10°C to 50°C, which means that we can use any one of these sets of data to calculate the responsivity. And since the linear relationship coefficients of T_{cn} for these sets of images are the same, selecting any set of images to compute responsivity yields the same value.

Once we get the responsivity from the calibration dataset, we apply it to the scene dataset to calibrate the image. For offset, we can be sure that it must be a raw image in the scene dataset and according to the model, it should be the 0th one.As you can see, after applying this offset, we did resolve the image, but it was not clear enough and there was a ghost effect on the image. We inferred that there was something wrong with the offset, so I've chosen another random scene image as the offset and the result shows that a different choice of offset will make a huge difference on correcting images. Respectively, if we pick an image that is as uniform as

Figure 3: Example of a corrected image using Linear relationship of Raw image to T_{cn}

possible as the offset, it will avoid the ghost effect. However, that would make the choice of offset unexplainable and inconsistent with what the model has given. Even worse, in later calculating, responsivity with several other sets of corresponding data that appeared to have the same slope T_{cn}, and the results for each set were inconsistent, which also denied the previous hypothesis. This brought the approach to an impasse. This approach has later been proven lacking in the consideration of the effect of T_{fpa} on responsivity and offset.

Linear relationship of Scene image to Raw image

After carefully reviewing the theoretical information provided, we are interested in the equations provided by the NUC model.We can make some changes to the above equation.

$$\text{Raw}_{\text{pixel}}(t) = \text{Resp}_{\text{pixel}} \times S_{\text{pixel}}(t) + \text{Offset}_{\text{pixel}}$$

We can make some changes to the above equation:

$$Resp_{pixel}(t) = \frac{Raw_{pixel} - Offset_{pixel}}{Scene_{pixel}}$$

In this approach, we mistakenly took the raw image data provided in the calibration dataset as the corrected scene image, which means raw image = scene image, and the offset must be a scene image, so we randomly selected a scene image (which is actually the same as the calibration raw image) as the offset. And we observed that the calibration dataset is a group of 6 images, and under the same T_{fpa} value, the T_{cn} makes a near-fixed oscillatory change. So we choose the 5th image of each group as the raw image in the above formula, and the 0th image as the offset image, and the scene image and raw image are the same image.

When calculating the scene image, we also randomly selected an image from the scene dataset as the offset, and the responsivity uses the closest value to the target T_{fpa} value (27°C) among the groups obtained in the above code, which is the value of group 3 (30°C).

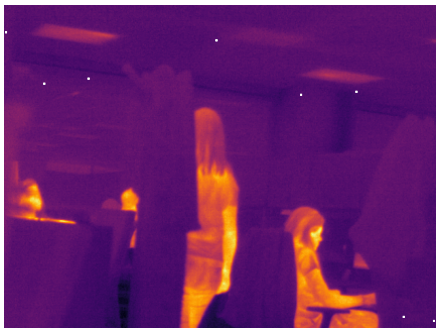


Figure 4: Example corrected scene image using Linear relationship of Scene image to Raw image.

It looks like it works well, but in reality, the values inside the array are very weird.And obviously, this is not a general approach, because it needs to choose a particular scene image as an offset. If none of the images in the dataset are nearly uniform, then we will have to import a reference image from another dataset to participate in the calculation as an offset. And when you think about it, the value taken

to calculate responsivity is not reasonable. According to this method, you don't even need the calibration dataset to calculate responsivity. You can just subtract a uniform raw image with T_{fpa} close to the target T_{fpa} to get the corrected scene image, but this method provides an idea for Task 3, as it does give good visual results.

3.1.2 Proposed approach

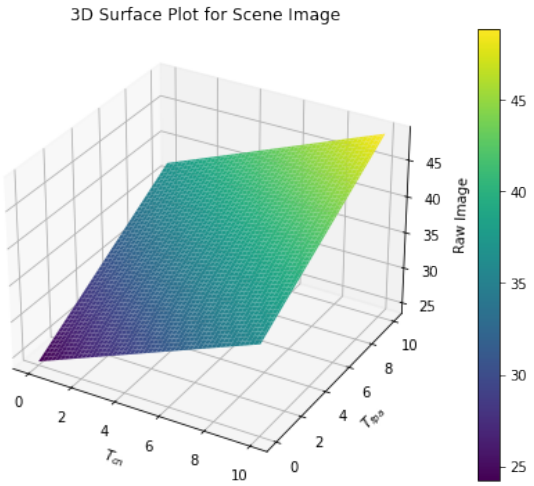


Figure 5: The model interpretation for the approach assuming that there's a linear relationship of raw image to T_cn and raw image to T_fpa, the resulting plane is the value of the scene image.

The 3D relationship between Raw image, T_cn, T_fpa is the model interpretation for the approach assuming that there's a linear relationship of raw image to T_cn and raw image to T_fpa, the resulting plane is the value of the scene image. After more discussion with Prof. Ronald P. about the NUC model, we have a more accurate understanding of responsivity and offset. Based on the model given in the theory material for calculating responsivity.

We can see from the above model that T_{fpa} will have a significant effect on both responsivity and offset, it is the main factor. So we have to calculate the set of images at the target T_{fpa} (T_{fpa} of the scene dataset) with the variation of T_{cn} presenting a monotonically increasing set of images from 10°C to 50°C). Here we used Newton interpolation to interpolate the polynomials, we choose three image sets to calculate, we also tried more image

Figure 6: The interpolated image sequence at $T_fpa = 27.0$

sets to interpolate them, but after three degree there's not much impact on the computational results of the existing calibration dataset, so we chose the fastest computationally degree, which is three degree, and if the situation changes, please choose the appropriate interpolation method. After calculating the responsivity, we substitute this value back into the original formula to get the offset corresponding to T_fpa,

(a) Corrected Image using Task 1 in Scene 1 (b) Corrected Image using Task 1 in Scene 2 (c) Corrected Image using Task 1 in Scene 3

Figure 7: Exam output of corrected images for Task1

3.2 Task 2

The second task has several sets of T_{fpa} data missing from the calibration dataset compared to the first task. We have to compute the image set of the target T_{fpa} on this basis. Same as the proposed method from Task 1, but for the second scene dataset, we have to manipulate its vmin and vmax to deal with the outliers. And here's our result
Same as the proposed method from Task 1, but there are two improvements that we need to focus on for Task 2:

1. Choose the appropriate image sets to interpolate new image sets for the target T_{fpa}, because we missed some calibration datasets compared with Task 1. This will increase the difficulty of interpolating them correctly.

2. Manipulate the v_{min} and v_{max} to deal with the outliers for the scene images. Outliers, created by malfunctions from some sensors, will continuously output wrong values which are the outliers presented in our datasets. Taking the results we've obtained from the first task as a reference to adjust the v_{min} and v_{max} will give us more reliable results for Task 2.

Since Task 1 and Task 2 are using the same approach and are well-interpolated, their results are visually highly similar. Thus, we can compare them with evaluation metrics later and see the impact of the changes in calibration datasets.

(a) Corrected Image using Task 2 in Scene 1

(b) Corrected Image using Task 2 in Scene 2

(c) Corrected Image using Task 2 in Scene 3

Figure 8: Exam output of corrected images for Task 2

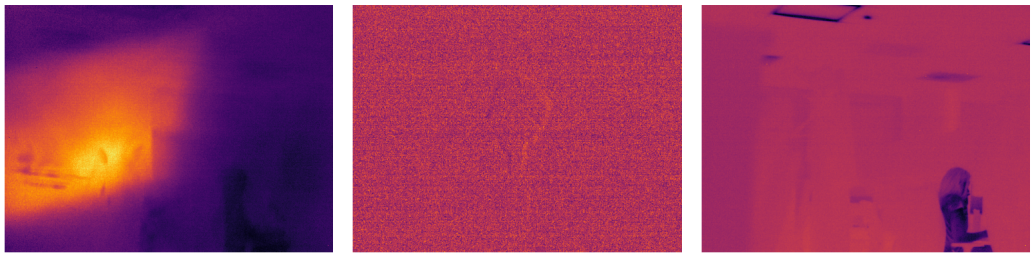
3.3 Task 3

Our task 3 is about performing Non-Uniform Correction (NUC) on the raw thermal images to enhance image quality and reveal the underlying scene without using any calibration data. We have to complete the challenge in the absence of reference points for correcting variations in the thermal sensor's response. Our target is to demonstrate the effectiveness of our correction method in improving thermal image clarity under these challenging conditions.

3.3.1 First approach

We have attempted to apply Non-Uniform Correction (NUC) to the thermal images by setting the offset as the mean image across all the scene images. Initially, we loaded the data and computed the mean image, assuming it would effectively represent the offset due to environmental noise. This mean is calculated by taking each corresponding pixel across all images and averaging them. The result is a two-dimensional array where each pixel's value is the average of that pixel over all images in the stack. And similarly, we computed the standard deviation across the entire scene image dataset and chose this as our responsivity.

We then applied the NUC by subtracting this offset from each image and dividing by the responsivity. However, the mean image, being an artificial image created by averaging each pixel across all images, imposed its bias across all corrected images and introduced a significant "ghost effect" in all the corrected images. The example images shown in figure 9 illustrates this issue, where the human body appears as a ghost-like effect in scene1 and scene3. We have chosen these images to show you the ghost effect clearly, as there were no human inside this images in real but after correcting the images we got this effect. This anomaly was consistent across all the images after correcting them, highlighting the limitations of using it in our dataset. However, it is to be mentioned that this approach with this ghosting effect will not give a visual result for the dataset where the images are similar. And it can be seen by observing the figure 9b which is the output for scene2.hdf5 file.



(a) Corrected Image using Task 3 in Scene 1 (b) Corrected Image using Task 3 in Scene 2 (c) Corrected Image using Task 3 in Scene 3

Figure 9: Exam output of corrected images using the first approach

3.3.2 Second approach

In our revised approach, we aimed to improve the Non-Uniform Correction (NUC) process by selecting a more appropriate offset image from the dataset, rather than using the mean image as a simple average is not robust to the outliers. For this approach, first we loaded the scene images and computed the mean image across the entire dataset. To determine the best candidate for the offset image, we measured the Mean Squared Error (MSE) and Structural Similarity Index (SSIM) between each image and the mean image. By calculating the MSE for each image, we identified the ten images closest to the mean image based on this metric. Similarly, we computed the SSIM values for each image and selected the top ten images with the highest SSIM scores, which indicates structural similarity to the mean image.

After testing all of the images we observed that all of them can be considered as offset and they are giving good corrected image. However, we have selected one of these closely matching images (specifically, the image at index 171) as the offset for the NUC process. Using this selected offset image, we corrected each image in the dataset by applying NUC. We are assuming that the selected offset image closely represents the common background or baseline noise of other images, that's why subtracting this offset is able to largely correct the images, making the adjustment for gain (responsivity) less critical. That's why for the responsivity we keep it a constant as 1, so technically we are not using the responsivity or gain here. This approach yielded more satisfactory results compared to our first method. The corrected images displayed in figure 10 improved clarity and reduced ghost effect. We have choosed the same images to display for both of the approaches so the difference can be visualise. Here you can see the ghost effect from scene1 and scene3. And for scene2, we got a very clear corrected image. However, this is to be mentioned that we have tried all of the dataset separately to check if we can find a uniform image from them. And we find out that using the scene.hdf5 dataset give us the best uniform image which can be used as offset in all the datasets.

(a) Corrected Image using Task 3 in Scene 1 (b) Corrected Image using Task 3 in Scene 2 (c) Corrected Image using Task 3 in Scene 3

Figure 10: Example of output of corrected images using the second approach

Here is the definition of the two methods that we used to find the images those are similar to the mean image:

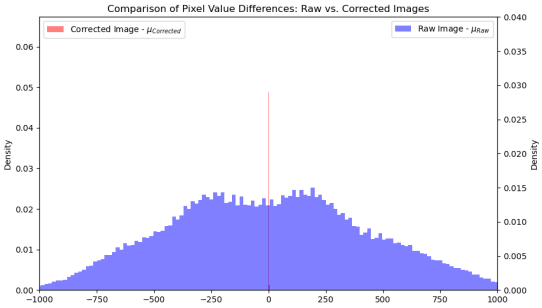
SSIM: The structural similarity index measure is a method for predicting the perceived quality of digital television and cinematic pictures, as well as other kinds of digital images and videos. It is also used for measuring the similarity between two images.

MSE: Mean Square Error measures the average squared difference between the estimated values (predicted values) and the actual value (ground truth). So we just calculate squared differences pixel by pixel.

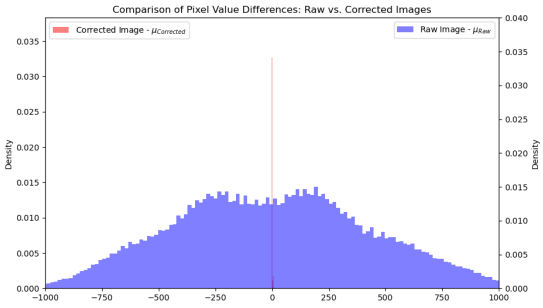
4 Evaluation

4.1 Histogram Comparison

In this section, we will compare the histograms of the density of the signal-to-noise ratio between the raw image and the corrected image for 3 scenes based on the task 1 and 2 discussed earlier in the report. Task 3 has been removed from the analysis because we do not use the calculation of the responsivity in this approach.
In Scene 1, the histograms reveal distinct patterns for each task. Task 2 exhibits a higher peak and a distribution closer to zero, resulting in a concentrated distribution around the average intensity of pixels.This is indicated by the narrower red histogram, suggesting a reduction in pixel value variance after correction.Task 1 follows with a lower peak but still maintains a relatively close distribution to zero, implying a significant improvement in pixel intensity adjustment. Overall, Task 2 appears to be the most effective in achieving an accurate correction in Scene 1.



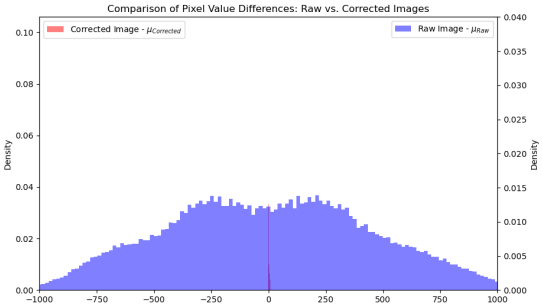
(a) Histogram for Task 1 in Scene 1



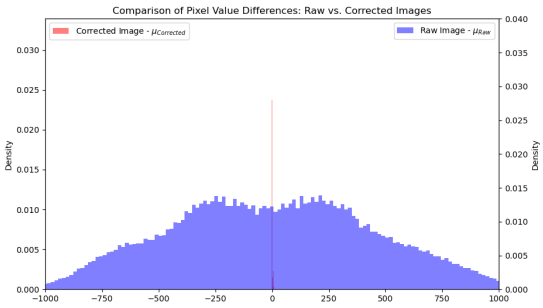
(b) Histogram for Task 2 in Scene 1

Figure 11: Density Histogram of Pixel Value Differences from Mean for Raw and Corrected for Task 1 and Task 2 for Scene 1

In Scene 2, the histograms depict unique trends for each task. Task 2 displays a higher peak and a distribution closer to zero, implying an effective adjustment of pixel density that results in a concentrated distribution around the average intensity.The narrower red histogram indicates reduced variation in pixel values after correction. Task 1, on the other hand, reveals a lower peak, suggesting a concentration of pixels, but the dispersed red histogram indicates a broader range of pixel values around zero. This implies a less precise correction overall compared to Task 2.



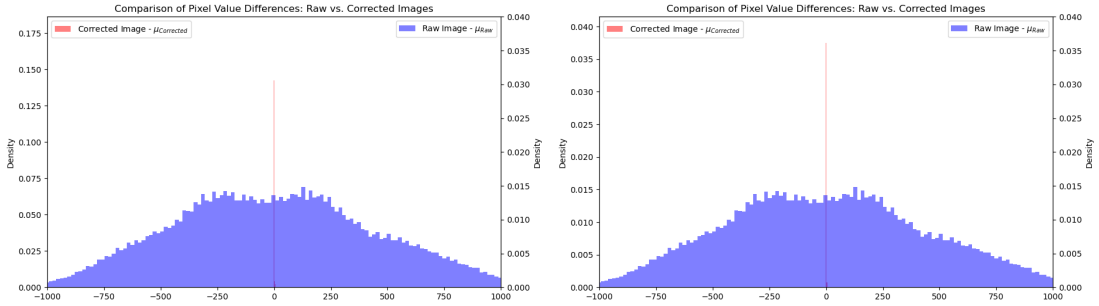
(a) Histogram for Task 1 in Scene 2



(b) Histogram for Task 2 in Scene 2

Figure 12: Density Histogram of Pixel Value Differences from Mean for Raw and Corrected for Task 1 and Task 2 for Scene 2

In Scene 3, Task 2 demonstrates a higher peak and a distribution closer to zero, implying an effective adjustment of pixel intensities resulting in a concentrated distribution around the average intensity. The narrower red histogram indicates reduced variation in pixel values after correction. Conversely, Task 1 exhibits a lower peak but still maintains a distribution relatively close to zero, suggesting a notable enhancement in pixel intensity adjustment.



(a) Histogram for Task 1 in Scene 3

(b) Histogram for Task 2 in Scene 3

Figure 13: Density Histogram of Pixel Value Differences from Mean for Raw and Corrected for Task 1 and Task 2 for Scene 3

4.2 BRISQUE Metric

BRISQUE does not calculate distortion-specific features like ringing or blur. Instead, it quantifies potential losses of ”naturalness” in the image caused by distortions using scene statistics of locally normalized luminance coefficients, thereby leading to a holistic measure of quality. The underlying features are derived from the empirical distribution of locally normalized luminances and products of locally normalized luminances under a spatial natural scene statistic model. Because this method doesn't transform the data to another coordinate frame (such as DCT, wavelet, etc.), it is different from the prior blind/no-reference image quality assessment methods. By using this model, we used follow code to install and import it.

For every scene, we assess the last image from the 3 tasks and do the comparison as they have more variation. In detail, we chose the 399th image for scene 1, the 149th image for scene 2, and the 799th image for scene3.

Before assessing the quality of the image, we need to convert the image type from an inferno image to a gray scale image, because BRISQUE can only work on the gray scale image. And we got the following result. The lower the score of the image, the higher the quality of the image. Also, we can just keep the cmap to gray directly.

The results indicate that as the tasks change, the image quality across the three different scenes follows a similar rule. The image quality for Task 1 always surpasses that of Task 2, with Task 3 exhibiting the lowest quality among them, which indicates that the performance of our approaches has a significant difference under three different conditions. But for scene 3, the result of Task 3 is better than that of Task2. When the scene images share the same environment with the calibration set, the quality of the corrected images is the best (Task 1). And the quality of the corrected images with dissimilar conditions (Task 2)

Figure 14: The image quality assessed by BRISQUE across different scenes and tasks. The lower the score of the image, the higher the quality of the image.

is still higher than the image quality without any calibration set (Task 3).

4.3 Peak Signal-to-Noise Ratio (PSNR)

Since the corrected image based on Task 1 got the best quality, we'll use it as a reference next to compute the PSNR to further demonstrate the differences between the images based on three tasks. PSNR is a widely used metric for measuring the quality of a reconstructed image compared to its original version. It is particularly useful in the fields of image compression and transmission, where it quantifies how much distortion or noise has been introduced during the process. In our project, we use this metric to measure the difference between the corrected images based on the different calibration conditions. PSNR is expressed in decibels (dB) and is computed based on the Mean Squared Error (MSE) between the original and the reconstructed image. Higher PSNR values generally indicate better image quality. The formula for PSNR is:

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

where MAX_I^2 is the maximum possible pixel value of the image (e.g., in our work, 255 for an 8-bit image) and MSE is the Mean Squared Error between the standard image and the image being compared. In our project, we used the `peak_signal_noise_ratio` function from the `skimage.metrics` module in Python.

Scenes	PSNR for Task2 (dB)	PSNR for Task3 (dB)
Scene 1	21.22	19.32
Scene 2	26.74	18.82
Scene 3	29.14	25.76

Table 1: Comparison of PSNR values of Task2 and Task 3 with Task 1

A higher PSNR value indicates less difference in the images, while a lower PSNR value indicates more difference in the images. The results indicate that the images based on Task 2 have less distortion than the images of Task 3, which is consistent with the result of BRISQUE, except for scene 3. The value of PSNR for Task 1 and Task 3 is lower than that of Task 2. Maybe it's because the difference with the image from Task 1 can't totally represent the quality of the image.

By using two different image quality metrics, BRISQUE and PSNR, we can draw the following conclusions about the approaches of NUC we used. The quality of corrected images is optimal when calibrated within the same environment as the calibration set. And the quality of the images corrected based on the calibration set with a different environment is higher than the image quality corrected without any calibration set.

4.4 Factors of quality of the image

In the process of solving these three tasks, we can find that T_fpa is the most critical factor affecting the clarity of image resolution, which is closely related to the responsivity and offset of the NUC model. In addition, there are also external factors that affect the image quality, such as the spatial resolution, which is the density of the sensor, and the sensitivity of the sensor, which is the smallest temperature difference it can detect. In addition, since the sensor only captures irradiance data and lacks colour data, it is difficult to distinguish between two objects when they are very close to each other, so we can make some adjustments to the vmin and vmax of the image in post-processing to enhance the contrast and make the information more intuitive and easier to interpret.

5 Conclusion

The conclusion of the denoising thermal images data challenge is multifaceted. Through the exploration of various methodologies and approaches, we aimed to improve the clarity and quality of thermal images captured under different environmental conditions. Our efforts focused on implementing Non-Uniform Correction (NUC) techniques, including Calibration-Based and Scene-Based algorithms, to mitigate noise and enhance image quality.

In Challenge 1, which provided the calibration dataset, our approach included exploring the linear regression model, as well as considering the linear relationship between the original image and the scene temperature and between the original image and the corrected scene image. Eventually we succeeded in interpreting the NUC model and establishing the correct correction method. A simple summary of the method is to correctly interpolate the polynomials, find the term of the target T_fpa, compute the responsivity and offset of this term, and apply the obtained parameters to the raw image of the scene for the correction work.

On the other hand, Challenge 2 brings the additional challenge of missing data in the calibration set, highlighting the impact of T_fpa on the two key parameters responsivity and offset in the NUC equation. We also note the importance of post-processing. The post-processing work focuses on two aspects, one is to deal with the outliers due to the malfunction of sensors, and the other is to increase the contrast to make the image information visually more readable. In the end, we achieved satisfactory results in both Task 1 and Task 2, proving the robustness of our technique.

Furthermore , Challenge 3 pushed the boundaries further by requiring correction without any calibration data, necessitating innovative strategies.We experimented with mean image-based correction and refined our approach by selecting a uniform image as an offset image, resulting in enhanced image clarity and reduced artifacts. Through comprehensive analysis, including histogram comparison, image quality assessment using BRISQUE, and PSNR evaluation, we observed the effectiveness of our approaches across different scenes and tasks.

Our data challenge gives us the opportunity to discover valuable insights into denoising thermal images, showcasing the significance of tailored correction approaches and adaptive methodologies. Moving forward, further research and refinement of techniques will continue to advance the quality and applicability of thermal imaging technology across diverse domains, from medical diagnostics to industrial inspections and surveillance.

6 Link of GitHub Repository

Here, you will find the link of our github repository, where we have uploaded all of our codes for the three task. Here is the link:

DataChallenge2024-Denoising Thermal Images

References

Marcotte, F., Tremblay, P., & Farley, V. (2013). Infrared camera nuc and calibration: Comparison of advanced methods [Email: frederick.marcotte@telops.com; Phone: 1-418-864-7808-420; Fax: 1-418-864-7843; http://www.telops.com]. *Proceedings of the SPIE*.

Tremblay, P., Belhumeur, L., Chamberland, M., Villemaire, A., Dubois, P., Marcotte, F., Belzile, C., Farley, V., & Lagueux, P. (2010). Pixel-wise real-time advanced calibration method for thermal infrared cameras. *Proceedings of the SPIE*, *7662*, 766212. https://doi.org/10.1117/12.850213

Kumar, A. (2013). Sensor non uniformity correction algorithms and its real time implementation for infrared focal plane array-based thermal imaging system. *Defence Science Journal*, *63*(6), 589–598. https://doi.org/10.14429/dsj.63.5768