

## PETIT-GAN Supplementary Material

### 6. Physical estimator coefficient calibration

This section provides a more detailed description of the physical estimator calibration procedure described in subsection 3.1.3 of the paper.

Equation 8 provides a parametric model that allows us to transform thermal image intensities to object temperatures and vice versa. To calibrate those parameters, we designed a calibration setup with the following components:

- an uncooled microbolometric camera; for our experiments, we used FLIR Tau2.
- an environmental chamber to control the camera’s ambient temperature (which implicitly controls  $T_{int}$ ), with an aperture drilled into one of its edges to allow for image acquisition from its interior.
- a blackbody target, essentially a flat-surface object whose temperature can be controlled with very fine precision, to set the temperature of the scene (*i.e.*,  $T_{obj}$ ); we used a scientific grade (CI SR-800N) extended-area black body.

To collect measurements, the camera was mounted inside the environmental chamber, facing outside through the drilled aperture, and the blackbody target was placed in front of the camera outside the chamber, such that it covered the camera’s entire field of view, as demonstrated in Figure 5.

During the calibration, the environmental chamber is set for heating (thus slowly and continuously increasing  $T_{int}$ ). While the chamber’s interior ambient temperature keeps increasing, the blackbody target is repeatedly set to an arbitrary temperature, and an image is acquired for every such temperature. This random target-temperature profile provides the best coverage of the three-dimensional  $T_{int}$ - $T_{obj}$ -intensity space compared to other alternatives.

To efficiently estimate the coefficients, we organized the measured intensities and the coefficients in vectors  $I$  and  $C$  correspondingly, such that:

$$M \cdot C = I \quad (16)$$

where each row of the over determined matrix  $M$  is the feature vector ( $F$ ) of a single measurement, and  $I$  holds the

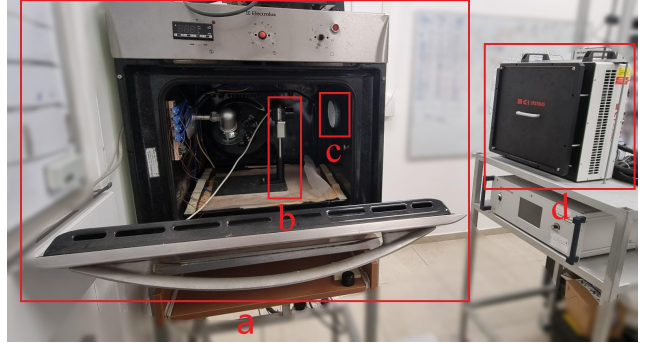


Figure 5. The physical model-calibration setup. The black body target (d) is observed by the thermal camera (b) through a drilled aperture (c) at the edge of the environmental chamber (a). This setup enables controlling  $T_{obj}$  and  $T_{int}$  independently.

corresponding measured radiometric intensities. Following the factorization from equation 10, the  $n$ th measured radiometric intensity equals the inner product between the  $n$ th measured feature vector and the vector of coefficients  $C$ :

$$I[n] = \langle M[n, :], C \rangle = \langle F[n], C \rangle \quad (17)$$

The matrix product factorization in equation 16 enables applying a linear regression to solve for the coefficients. As the measurement noise appeared to be approximately white and Gaussian, the least-squares estimator (which is the maximum-likelihood estimator in the additive-white-Gaussian noise regime) was used to extract the coefficients:

$$C = M^\dagger I \quad (18)$$

where  $M^\dagger = (M^T M)^{-1} M^T$  is the Moore-Penrose pseudoinverse of  $M$ .

In the calibration process described above, the intensity vector  $I$  corresponds to measurements of a single pixel. Because each element in the microbolometer might have a different sensitivity, the coefficient extraction was performed independently for every pixel in the image.

### 7. Dataset

As mentioned in subsection 4.1 of the paper, one of our contributions is the introduction of a novel dataset of unpaired images belonging to different bands of the thermal

(LWIR) spectrum. In the discussed subsection, we provided a brief description of the data collection procedure. This section provides more details about the dataset, including the data collection, the data format, and the preprocessing applied to the data. The dataset can be accessed through the [project's website](#).

### 7.1. Data collection

In the paper's introduction, we presented the motivation of our research - synthesizing a multispectral dataset to allow the development of algorithms and systems for nanosatellites. Thus, our goal was to mimic the interstellar acquisition setup to the best of our ability. To that end, a light airplane (appears in Figure 6) was used to perform flights at a height of approximately two kilometers above ground. Images acquired in such a setup could be later reasonably extrapolated, *e.g.*, by a proper downsampling, to mimic the scene as if it was captured by an actual nanosatellite.



Figure 6. The light airplane used for acquiring the thermal images.

A dedicated aerial pod was designed and manufactured to mount the camera on the airplane's underbelly, as visualized in Figure 7.



Figure 7. The aerial pod used for mounting the thermal camera on the light airplane's underbelly.

The pilot performed several flights, with some IR band-pass filter applied to the camera lens (monochromatic), or without this filter (panchromatic). While attempting to follow the same predefined trajectory, ensuring that the actual airplane's positional and angular states are the same for different flights is physically infeasible. Jittering and air bumps provide additional randomness to the system's state, resulting in unpredictable and unreproducible acquisition conditions. Due to those limitations, the images collected for the different thermal modalities are inherently unpaired, hence our entire dataset is unpaired.

### 7.2. Data format

Each frame in the dataset is saved as an independent numpy *.npz* file with the following fields:

- *image*: a two-dimensional  $256 \times 336$  Numpy array containing the radiometric intensities acquired by the Flir TAU2 camera (with 14-bit precision).
- *time*: the date and time of the acquired image in GMT.
- *lat*: the latitude of the plane at the time of acquisition.
- *long*: the longitude of the plane at the time of acquisition.
- *alt*: the altitude of the plane above sea-level at the time of acquisition (in meters).
- *fpa*: the intrinsic temperature of the camera as measured by the focal-plane array thermal sensor (in Celsius).

### 7.3. Preprocessing

Due to technical considerations, a significant portion of the acquired images were purely of sea with no additional features. Pure-sea images are roughly thermally homogeneous, *i.e.*, of a constant intensity level, up-to measurement noise and spatial aberrations, making them extremely similar to each other and of lesser value to our training. Naively using all those images for training would result in a highly imbalanced dataset, which will bias our model to produce high quality outputs for sea images, in the expense of other samples. An impression of the described attributes of sea images can be obtained from Figure 8.

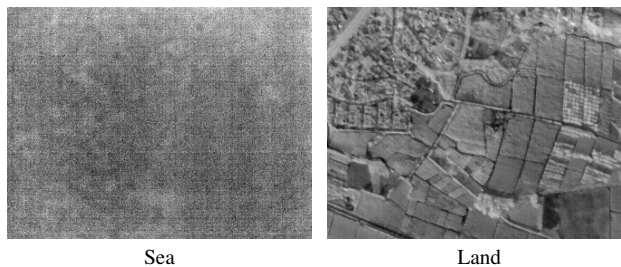


Figure 8. A thermal image of pure sea vs. an image of land for reference.

For those reasons, it was decided to design a classifier for filtering out all sea images from the dataset. As explained and visualized, sea images are mostly made out of a DC component (constant) and measurement noise. This means that the gradients of two images of sea, even at different water temperatures, should have a very similar high-frequency spectral content. Analyzing the gradients of sea and land images confirms this hypothesis, as demonstrated in Figure

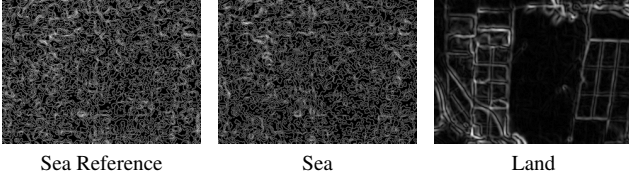


Figure 9. The gradient maps of exemplary land and sea images. An additional gradient map of a sea image is provided for reference.

9. This finding encouraged us to base our classifier upon the gradients of the images. A more thorough quantitative analysis of the gradients revealed a significant difference between the norms of the gradients of sea and ground images. This difference is also visible in Figure 10 where the norms of ten randomly-selected sea and land images are compared.

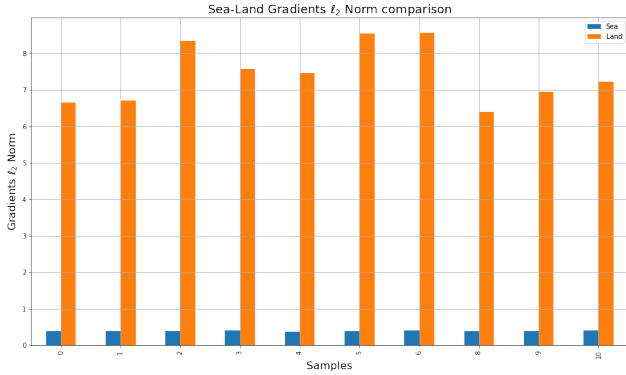


Figure 10. A histogram of the  $\ell_2$  norm of the gradients of random sea and land images taken from the dataset.

Following these findings, the following pipeline was designed:

1. Prefiltering of the acquired image  $I$  with a two-dimensional gaussian low-pass filter to attenuate the noise and very high frequencies, which typically don't belong to the spectral content of natural images. We denote our filtered image by  $\tilde{I}$ .
2. Classify  $I$  as a land image if  $\|\nabla\tilde{I}\|_2$  (the  $\ell_2$  norm of the gradient of  $\tilde{I}$ ) is greater than some threshold  $\tau$ , and otherwise as a sea image.

The threshold's value  $\tau$  was set to the average between the minimum  $\ell_2$  norm of the gradients over a random set of 1000 land images and the maximum  $\ell_2$  norm of the gradient over a random set of 1000 sea images. Formally:

$$\tau = \frac{1}{2} \left( \min_{I_l \in D_{land}} \|\nabla I_l\|_2 + \max_{I_s \in D_{sea}} \|\nabla I_s\|_2 \right) \quad (19)$$

The entire preprocessing pipeline, as well as the calibrated threshold, were designed and applied to both monochromatic and panchromatic images independently.

## 8. Additional qualitative results

Following our guarantee from sub-subsection 4.4.2, we provide additional qualitative examples of our model's output. These additional examples provide additional evidence to the claim made in the paper regarding our model's superior quality compared to the baseline models. All additional examples appear in Figure 11 in the page below.

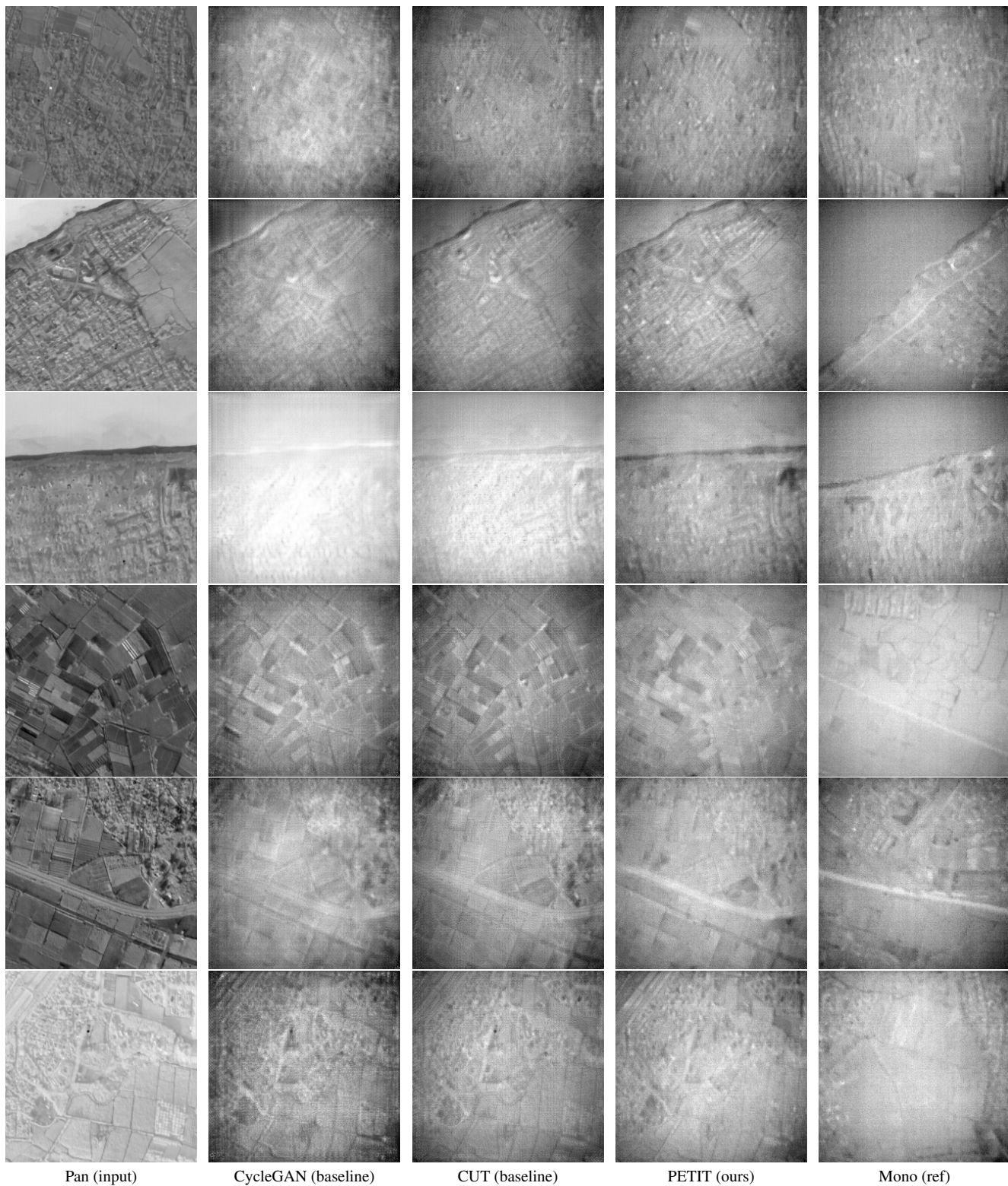


Figure 11. Qualitative comparison. (a) Panchromatic (Pan) input. (b) CycleGAN output. (c) CUT output. (d) PETIT output. (e) Real unpaired monochromatic image (Mono) for reference.