(Street) Lights Will Guide You: Georeferencing Nighttime Astronaut Photography of Earth

Alex Stoken[1] Peter Ilhardt[1] Mark Lambert[1] Kenton Fisher[2]
{alex.h.stoken, peter.d.ilhardt, mark.d.lambert, kenton.r.fisher}@nasa.gov
[1]Jacobs Technology, NASA Johnson Space Center [2]NASA Johnson Space Center

Abstract

Astronaut photography from the International Space Station provides the highest spatial resolution nighttime Earth observations imagery publicly available, offering up to a 150x increase in resolution over other freely accessible satellite data sources. Yet, this imagery is underutilized in science applications because it lacks the geolocation metadata required for downstream analysis. We present NightMatch , a fast and accurate method for localizing and georectifying nighttime astronaut photography. By combining street network data with daytime satellite imagery, we produce a reliable reference target for similarity detection via pairwise image matching. We curate and release the Astronaut Imagery Matching Subset - Night (AIMS-Night), a collection of 363 images and ground truth localizations, and benchmark our method against this set to establish a robust localization pipeline. Our method correctly localizes 81.8% of AIMS-Night, and can be quickly deployed on the over 2 million nighttime astronaut photographs to produce a high quality analysis-ready data product.

Figure 1. **The nighttime astronaut photo geolocation problem.** Astronauts can take a photo of a city in a large and continuous search area due to their wide field of view and high focal length lenses. We seek to rapidly localize these images by identifying the most likely areas to photograph and then comparing an astronaut photo to a representative reference image of that area.

1. Introduction

Nighttime astronaut photography of Earth is the highest spatial resolution nighttime remote sensing data set that is publicly available. The imagery is acquired under unique conditions as each photograph is taken by an astronaut on the International Space Station (ISS), approximately 415 kilometers above the Earth's surface. From this vantage point an astronaut will see 15 to 16 sunrises and sunsets every 24 hours, and can view an area of the Earth that stretches over 2000 kilometers in each direction. Astronauts acquire imagery of Earth in response to research requests, natural disaster response efforts, for educational projects, and for public outreach. To date, there are over 4.6 million astronaut photographs in the Gateway to Astronaut Photography of Earth database, an archive maintained by NASA's Earth Science and Remote Sensing Unit. Of these 4.6 million, 43% are nighttime images. These nighttime images are taken with commercial cameras and a collection of lenses ranging from short (14mm) to long (1150mm) focal lengths. The spatial resolution of the imagery varies, with the highest resolution photos at about 3 meters per pixel, approximately 150 times higher than other competing public data sets. In contrast to other publicly available datasets, astronaut photography captures true-color (RGB) data of nighttime urban lighting patterns. These attributes have proven to be highly valuable for researchers studying urban development, environmental change, economics, and other fields [17, 35, 43]. The data is commonly used to better understand or model a wide range of phenomena, including impervious land cover, the effect of artificial light on biological systems (Melatonin Suppression Index), Gross Domestic Product, and light pollution itself [15, 19, 23].

Some analyses are currently possible *only* with nighttime astronaut photography [43]. The study of certain environmental measures requires true color data like that in digital photos (after calibration [35]), and cannot be done with the

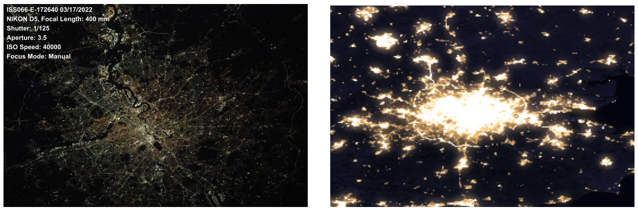

Figure 2. **Astronaut photograph and satellite image of London. Left:** ISS066-E-172640 (400mm focal length lens). **Right:** Suomi NPP/VIIRS data for London, Annual 2016 map via NASA WorldView. Astronaut photography has significantly higher spatial resolution than the next best nighttime imagery.

band data from nighttime observing satellites like VIIRS.

Yet, due to the imaging conditions inherent to astronaut photography, much of this data is not geolocated, making it difficult to use in downstream analysis. Given the vast Earth area visible to an astronaut at any point in time and the unconstrained nature of photography in micro-gravity conditions, each astronaut photo can show any portion of the visible Earth, at any orientation and at a variety of scales depending on the zoom lens used (Fig. 1). Recorded metadata only contains the focal length of the lens and the timestamp, which can be used to determine the location of the ISS at the time the photo was taken.

There has been a concerted effort to geolocate astronaut photos manually due to the growing demand for high resolution nighttime data sets. This process is time consuming and challenging, even for human experts. Locating a nighttime image, especially given the lack of high resolution reference maps, can take minutes to hours per frame, and georectification via control points adds additional difficulty. A citizen science project, Cities at Night, was initiated through a NASA/ESA partnership to assist with localizing nighttime photos and with the aim of generating a high resolution global nighttime light map from astronaut photography. Through these efforts, 26,000 nighttime images have been localized over the 20+ year history of ISS astronaut photography, but this equates to just 1.3% of all nighttime astronaut photography.

Recent works [7, 41] have sought to address the localization problem for daytime astronaut photography, where there is an abundance of reference satellite imagery of similar wavelengths and spatial resolution to use for localization. Unfortunately, there is not a comparable collection of nighttime geolocated satellite imagery to use as reference as none of the satellites reach the level of detail of astronaut photographs (Fig. 2). It is precisely the properties which make nighttime astronaut photography so valuable that make it difficult to localize.

To account for the lack of a suitable nighttime dataset for localization, we modify another spatial data source for use within our localization process. A distinctive feature of Earth's appearance at night is artificial light associated with streets and buildings. Global, geolocated street network data is readily available from sources like OpenSteetMap [33] and the Global Roads Inventory Project (GRIP) [31]. Using this data, we can *generate* a reference image capable of matching with a nighttime astronaut photo.

Using these generated reference images, we build a method for localizing and georectifying nighttime astronaut photography of Earth. In astronaut photography, cities are the dominant nighttime feature and are photographed the most. Thus, for each astronaut photo we sort a list of cities based on a likelihood score and iteratively check these by pairwise matching our generated reference image of the city with the astronaut photo. Through experimentation we identify a threshold for matching that yields highly precise city determination, allowing the use of early stopping in the localization pipeline. We then use keypoint matches to georectify the nighttime astronaut photo to the reference image. We evaluate the configurations of components in this pipeline for matching and localization as detailed in Sec. 5.

All experiments are run on publicly available astronaut photography in the Gateway to Astronaut Photography of Earth database at `https:\eol.jsc.nasa.gov`. From this database, we select 363 nighttime astronaut photographs with human expert-labeled location metadata, and combine these photos with our reference images to form the Astronaut Imagery Matching Subset - Night (AIMS-Night), a challenging evaluation dataset that emphasizes illumination and perspective changes.

This work makes three main contributions:

- We develop a method for fast and accurate localization and georectification of nighttime astronaut photography via comparison to simulated reference images
- We conduct experiments on reference image generation to illustrate the effects of image properties on matching
- We collect and release an evaluation dataset, AIMS-Night, of 363 geolocated nighttime astronaut photos with ground truth human labels

Application of this method to the full collection of nighttime astronaut photographs will lead to a significant increase in freely available high resolution geolocated nighttime imagery of the Earth and will promote further Earth Science research of urban areas.

2. Related Works

Astronaut Photography Localization. While Earth scientists have been studying astronaut photography of Earth for over 50 years, in recent years the computer vision community has also taken interest in the localization of this unique imagery set. Fundamentally, astronaut photography localization can be cast as a visual localization problem, and the scale, challenges, and impact of this domain have generated interest in trying to solve it. A recent method, Find My Astronaut Photo (FMAP) [41] aims to solve the daytime astronaut photography localization problem by comparison to daytime satellite imagery via pairwise matching. This method shows strong performance but is not designed to handle the related, but markedly different challenge of nighttime photos where no similar reference imagery is available.

Schwind and Storch [40] is the first attempt to address the nighttime imagery localization problem. Like [41], the work takes a pairwise matching approach over tiled regions of the Earth near the ISS nadir point. They rasterize street data to build a reference map at the same resolution as the astronaut photo and use BRISK [26] to exhaustively match each tile at 49 rotations. The work examines a small set of images, and focuses mainly on the resulting quality of the georectification of the image, rather than the broad localization problem. While this shows promise, the exhaustive BRISK strategy is too slow to scale to the entire catalog of nighttime astronaut photography and inherently sensitive to small changes in the rasterization procedure. Additionally, their method does not discriminantly localize - some "matching cities" have 30 point matches, and others 700. In contrast, our work focuses on a discriminative and rapid localization procedure that can be applied to the entire collection of nighttime astronaut photography.

Visual Place Recognition and Localization. Visual Place Recognition and Visual Localization are the related problems of determining the location depicted in an image and the camera pose, respectively, from image content alone. Typical benchmark datasets in these domains comprise landmark or city level regions, with the goal of identifying where an image is taken within that area [2, 11, 29, 39, 45, 48]. Of particular relevance are datasets that emphasize illumination differences like Aachen Day-Night [38] and Tokyo 24/7 [44], which test localization robustness to illumination. For example, in Aachen Day-Night, all reference images are taken in daylight, while queries consist of day as well as night imagery. Modern matching pipelines perform extraordinarily well even on the night queries, which encouraged us to run an experiment matching our nighttime imagery to daylight satellite imagery (see Sec. 5). Popular approaches to these problems typically have an initial retrieval stage that identify possible matches by comparing global feature vectors, followed by a pairwise matching stage to yield high precision predictions [4, 6, 13, 36].

For global localization tasks, recent works have introduced location embeddings trained via contrastive or CLIP-like setups [24, 30, 47]. These embeddings are primarily used to localize street level (rather than remotely sensed) imagery or enhance downstream task performance (*i.e.* animal classification, temperature prediction), but similar tech-

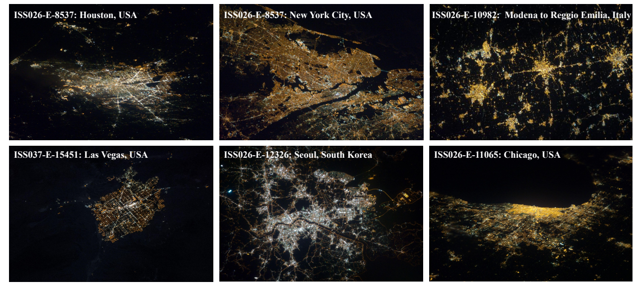

Figure 3. **Example Nighttime astronaut photos.** Each photo varies in field of view, obliquity, and quality. Cities themselves vary in structure and light color.

niques could be a fruitful future research angle for astronaut photography localization.

Image Matching. Image matching focuses on finding the relationship between two images of the *same* scene [22]. This is often used for camera pose estimation or 3D reconstruction via Structure from Motion (SfM) or Simultaneous Localization and Mapping (SLAM). The general matching procedure involves extracting and describing image features from multiple images and then matching features across images. Optionally, a geometric verification step is used to prune invalid matches. Traditional image matching methods were both sparse and handcrafted [28], but recent works have produced a myriad of solutions that replace different aspects of the pipeline, or the entire pipeline itself, with learned components [12, 14, 16, 21, 27, 32, 37, 42].

Astronaut photography by nature has unknown orientation and therefore research into rotation invariant matching is especially relevant. Some handcrafted descriptors (SIFT [28], BRISK [26]) have built in rotation invariance, while others have trained additional orientation estimation modules [14]. Rotation robustness can also be achieved using equivariant networks [8], including one work that uses astronaut photography to evaluate their method [9].

In addition to performance on downstream tasks like SfM, the Image Matching Challenge [1, 22] serves as a popular benchmark for these new methods. Despite significant progress over the last decade, image matching, particularly across more dramatic scene variations, is still an unsolved problem. Illumination conditions, extreme viewpoint or scene differences, matching across modalities, and occlusion all pose difficulties to current methods.

3. Datasets

3.1. Nighttime Astronaut Photography

Astronaut photography of Earth is a unique remote sensing dataset that complements modern satellite data. The dataset extends back to the early days of human spaceflight in the 1960s, though we restrict our focus to astronaut photography taken by crew members on the International Space Station over the past 20 years. This imagery, captured by hand in microgravity, lacks any canonical orientation, and given the large extent of Earth visible at any one time, can cover viewing angles ranging from nadir-facing (straight down) to oblique (≈2000 km away). Astronauts have access to zoom lenses ranging from 25mm to 1150mm, and use the lenses to take imagery of spatial areas as broad as entire hemispheres and as specific as city neighborhoods.

To date over 4.6 million images are in this dataset and 1.9 million are *nighttime* images. Day or night is determined using the sun elevation angle metadata associated with each photo, using the condition $\text{Elev}_{\text{sun}} < -15°$ as night.

For nighttime imagery in particular, astronaut photography provides the most detailed publicly available data with minimum ground sample distances of approximately 3 meters/pixel. In comparison, the highest resolution satellite data available is from the Visible Infrared Imaging Radiometer Suite (VIIRS), with spatial resolution of 500 meters/pixel. VIIRS data is procedurally geolocated due to its known acquisition conditions, making it a simpler data product to use for researchers. Geolocated nighttime astronaut photography will open up completely new analysis workflows for nighttime light researchers.

Following the convention of Find My Astronaut Photo, we develop and release an evaluation set for nighttime astronaut photography. The Astronaut Image Matching Subset - Night (AIMS-Night) contains 363 nighttime astronaut photographs of Earth with varying look angles, spatial resolution, image quality, and acquisition date, which combined form a representative sample of nighttime astronaut photography. Further details are in the Supplementary (Sec. 8). Each image has a human-verified ground truth center point, along with other metadata available for all astronaut photography. Sample images from this set are in Fig. 3. We augment this set with our reference images, and release it at https://eol.jsc.nasa.gov//BeyondThePhotography/AstronautPhotographyImageMatchingSubset.

3.2. Street Network Data

Street network data, which contains the location and size of roads, are a key component of modern navigation systems. The near ubiquity of these systems over the last decade has lead to an increase in the quality of street network databases, the most well known publicly available of which is OpenStreetMap [33]. This and similar street datasets are vector data, containing geographic coordinates of individual roads as well as additional metadata such as road type and permanence. We choose the Global Roads Inventory Project (GRIP) roads database for this project due to its global and "harmonized" nature [31]. This dataset is aggregated from various national, supranational, and global sources into a single collection with a unified classification scheme for road significance from (1) highways to (5) local roads.

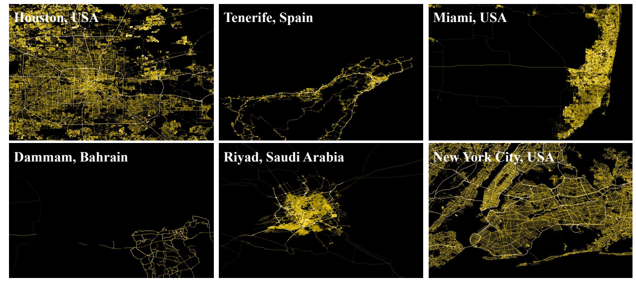

Figure 4. **Example rasterized street maps.** Each map contains aggregated streets using our custom aggregation strategy.

Regional Variation. Due to the variety of underlying sources of road data in GRIP, there is significant regional variation in the quality of the data and thus, the produced street maps. Some regions, where astronaut photography shows significant nighttime light, do not appear to have streets in the same region. GRIP data is not temporally consistent nor regularly updated, and some regions have data only as recent as 1997. This variation makes matching particularly difficult in some areas of the world.

4. Methods

We present a method for localizing and georectifying nighttime astronaut photography, NightMatch (Fig. 5). First, NightMatch generates a list of visible cities given the ISS position, then sorts the list by targeting probability. NightMatch iteratively checks each city for a match by identifying corresponding keypoints between the astronaut photo and a reference map generated from street network and satellite data. Upon finding a confident match (*i.e.* an astronaut photo/reference map pair with many corresponding points), the correspondences are used for georectification.

4.1. Generating Reference Images

Rasterizing Street Networks. Rasterization is the process of converting vector data to a regular grid - in our case, converting street network data into a map. We use Datashader [5] to control the rasterization process, which allows us to determine the aggregation methodology as well as shading parameters for the conversion.

Determining Intensity. Street lights on roads are the dominant nighttime light source (32% of zenith-facing light) [25], but not all roads contribute equal light. Light is additive, meaning that more light appears continuously brighter (until a saturation point is reached). We experiment with street aggregation strategies to mimic this behavior and best model the appearance of light in astronaut photography.

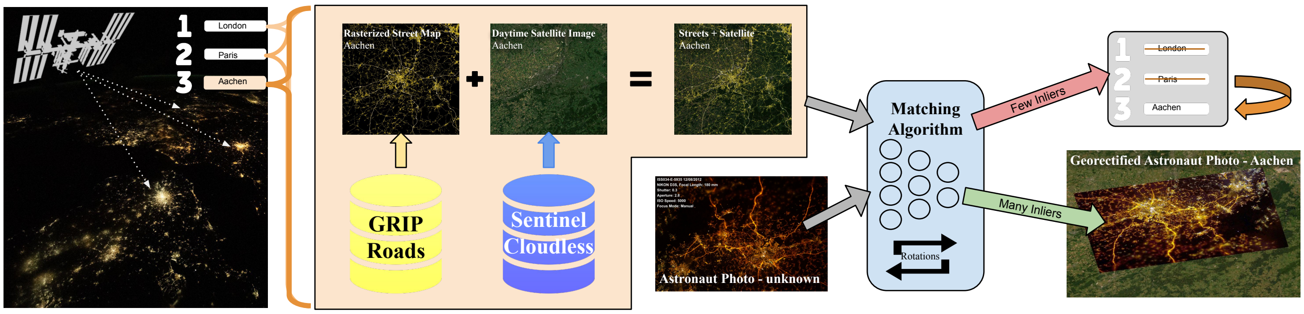

Figure 5. **Localization Pipeline.** Based on the location of the ISS, we sort the visible cities by targeting likelihood. We iterate through this queue by generating a reference map of combined daytime satellite imagery and street network data and matching it with the astronaut photo. Once a city clears our matching threshold, we georectify the astronaut photo using the corresponding keypoints.

For each pixel we determine the intensity via a weighted sum of the intersecting roads. As highways (GRIP=1) are typically brightly lit by street lights and have many cars which produce additional scattered light, they are given the highest weight, $w_1 = 3$. Other road types are weighted inversely to their GRIP classification (Sec. 3.2), $w_{GRIP} = \frac{1}{\text{class}}$. Thus, a pixel's intensity is defined as

$$I_{xy} = \sum_{\text{GRIP}=1}^{5} w_{\text{GRIP}} * \sum \text{roads}_{(\text{class = GRIP})\wedge(\cap xy)} \qquad (1)$$

Sample street maps generated with this scheme are in Fig. 4.

Satellite Imagery. Behind the rasterized street network, we add daylight satellite imagery to each reference map as a background layer. This imagery is sourced from a cloudless composite of Sentinel-2 satellite data produced by EOX [18]. The data is color corrected for atmospheric effects and mosaicked yearly to produce a cloud-free data product that is ideal for use as a reference, as it has the most unobstructed, matchable area.

Combining Day and Night. Street maps are ideal for simulating the light produced from cities but are devoid of features outside of areas with dense roads. These data-less areas are unmatchable, despite often being a large portion of the reference image. On the other hand, nighttime astronaut photos include some texture in road-less areas, as some Earth features beyond roads are still visible at night - scattered light, moonglow, and airglow all create an environment bright enough that in nighttime astronaut photos, few pixels are truly black. Fortunately, much of this non-road texture is present in daytime satellite imagery, making such images well suited as a background for our reference maps. Adding this reasonable background increases the matchable area significantly, resulting in more images localized (Sec. 6).

To combine the images, we first ensure that both are geo-aligned, such that each cover the same area. Then, we set all non-road pixels in the street map transparent, and alpha-blend ($\alpha = 0.7$) the remaining area. The result is a feature-rich reference image which maintains an emphasis on the significance of roads but is less sparse (Fig. 6).

Determining Reference Extent. With the extreme distances and angles in oblique astronaut photography, there can be significant perspective shift between a nadir view of a city and the astronaut's view through the camera lens. Many image matching methods are not robust to such extreme viewpoint shifts, so we apply a perspective transform to produce a view of the query city as it might appear to the astronaut. The first step toward generating this simulated view is determining the area visible for each image (*i.e.* for a high focal length lens, the image might just contain a portion of the city, while for a lower focal length lens, the entire city and some surrounding area may be visible) by modeling the likely extent of a given astronaut photograph.

To determine a photo's likely extent, we model the field of view of the camera as if it were pointed at the center of a city. With a known ISS position and the hypothesis that the camera body is aimed at the city center, we can determine the camera extrinsic matrix with respect to an Earth Centered Inertial frame, and then project the corners of the image plane onto the Earth (into a geodetic frame). This technique yields the approximate ground field of view for each query. By mapping the image plane corners to the corresponding field of view corners, we get a perspective transform that, when applied to the reference image, will make that query appear as it would to an astronaut on the ISS.

4.2. Matching and Localization

City Selection and Prioritization Pairwise image matching can be constructed to have very high precision, such that once we've found a positive city match, we can be confident that it is a correct localization and cease searching additional cities ([41] calls this the "early stopping" property). This makes the order of city verification critical to the total runtime of this approach. As checking a single query takes $\approx$5 seconds, formulating a well prioritized search queue becomes the dominant runtime component.

We build a prioritization function based only on a priori available metadata such that our queue is ordered by the likelihood that the astronaut photo is centered on a particular city. Population serves as an easy-to-gather metric that correlates strongly with a city's nighttime brightness - heuristically, brighter, more populous cities tend to be photographed more often than less populous ones. Similarly, cities nearer to the ISS are more likely to be photographed than distant ones.

Based on these two principles, we construct our ranking function in Eq. (2). Cities are scored proportionally to their ground distance d from the ISS nadir point, and inversely with their $population\_rank$. Population rank [34] is an integer from 1-14 representing how populous a city is; more populous cities get a *larger* rank, and are thus brought toward the top of the queue. Equation (2) was determined experimentally to balance the two factors, so that cities closest to the ISS, along with further but more populous cities, are near the top of the queue.

$$sort\_score = \frac{d}{population\_rank^4} \tag{2}$$

We initialize the queue with all of the cities from the Natural Earth 10m cultural database [34] that fall within 1200 km of the ISS nadir point. These are the cities visible to an astronaut at any point in time. Cities are then sorted in ascending order with Eq. (2).

Pairwise Matching. We iterate through the queue, generating a reference map for each city and then matching it with the nighttime astronaut photo using the ALIKED keypoint extractor [49] and LightGlue matcher [27]. Aligning rotation is critical to astronaut photography localization [9, 41], yet this extractor/matcher combination is not rotation invariant, so we test against eight 40° rotations of the reference map. We use the max accuracy LightGlue setup with a confidence filter threshold of 0.1. We perform geometric verification of matches using OpenCV MAGSAC [3, 10] with a reprojection threshold of 5 pixels and 100K iterations.

Georectification. Upon finding a positive match, we use the geometrically verified keypoint correspondences (commonly "inliers") to georectify the image. These points are passed as pixel coordinate/geographic coordinate pairs to GDAL's affine transformation estimator for warping [20].

5. Experiments

To find the optimal reference image generation procedure and matcher, we conduct experiments on each component of our pipeline. Understanding that image matching techniques are robust to certain scene variations, but often only to a limited degree, we separately test (1) different feature extraction/matching methods for their robustness and (2) reference image generation procedures to bring the reference image appearance closer to that of the astronaut photo. We run all experiments on the AIMS-Night dataset, with results presented in Tab. 1.

5.1. "Best Case" Pairs - Testing Matching Methods

We first investigate the performance of different image matching methods, given a fixed astronaut photo / reference map pair. We use the ground truth center point of the astronaut photo and the computed surrounding field of view to produce a reference image that encompasses the same area. This is the "best case" for matching, as we know that the images cover the same region. In contrast, when the true image center point is not known, we can only guess the center of the image to produce a matching region. In such cases, we assume the center of the image is the center of the city, but this is rarely exactly true. In the "best case" scenario, a robust matcher should produce a high number of inliers, so it is a good test case for determining matchability.

We evaluate representative matchers from three classes: handcrafted (SIFT [28], BRISK [26]), learned detector/matcher (ALIKED [49]-LightGlue [27], SuperPoint [37]-LightGlue, Steerers [9]), and learned, dense (detector free) matchers (SE2-LoFTR [8]). When a method is not advertised as rotation invariant, we perform eight 40° rotations on the astronaut photo. Performance is measured as number of matching pairs in AIMS-Night that exceed an inlier threshold.

5.2. "Best Scene Simulation" - Testing Reference Image Generation

We next examine the affects of reference image types on matching performance. We again use the "best case" matching scenario, to isolate the impact of the reference image itself. We examine four reference generation processes: (1) given the strong performance of matching methods across illumination conditions, we first examine matching with daytime satellite imagery of the city area (2) we follow the rasterization procedure of [40], producing a binary image, with highways wider than other streets (3) we implement our own rasterization procedure that models street light as additive (Sec. 4.1), and (4) we combine daytime satellite data and rasterized roads via alpha blending. An example of reference maps produced with each procedure is in Fig. 6.

6. Results

Matching Method Results Results of the experiments in Sec. 5 are reported in Tab. 1. Thresholds are determined experimentally by finding a number which reasonably minimizes false positives. For example, a threshold of 100 indi-

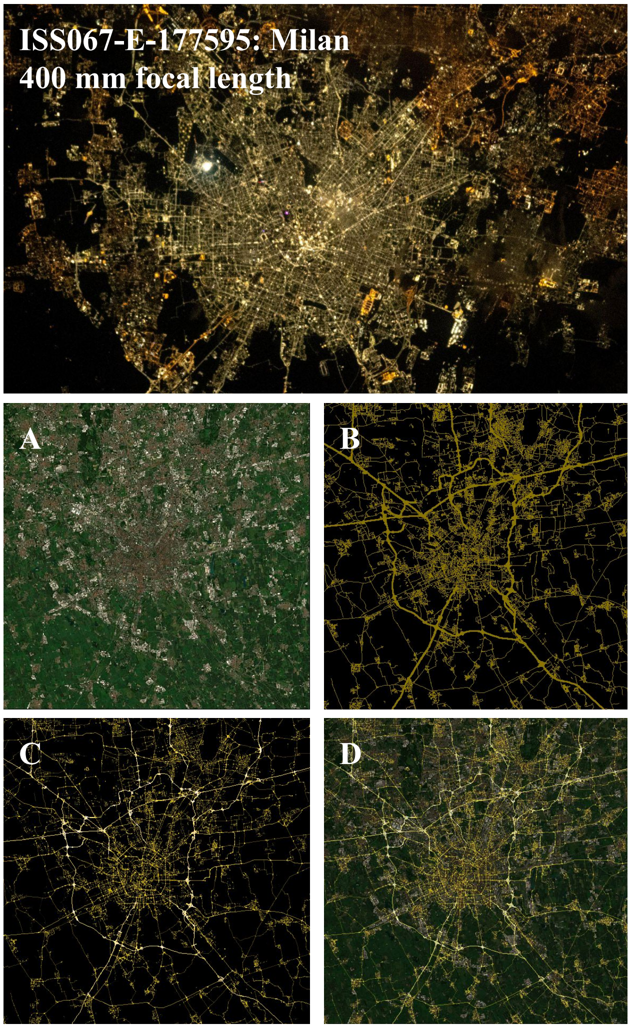

Figure 6. **Reference Image Comparison. A:** Daytime Satellite Image **B:** Rasterized Street Map - Binary **C:** Rasterized Street Map - Custom **D:** Street Map + Satellite

cates that images that are *not* of the same area almost always have less than 100 inliers.

With these thresholds, two learned, detector-based methods (ALIKED and SuperPoint) with the LightGlue matcher perform best. Following these is a new, rotation invariant method Steerers, which uses DINOv2 as feature extractor, which may account for its strong performance.

Image Generation Method Results. Changing the reference image has significant impact on matching and overall localization performance (Tab. 1). Considering the extreme illumination difference, matching against a daytime satellite image (a setting that closely mirrors FMAP [41]) shows surprisingly strong results. This could be due to roads still being a prominent feature in greyscale satellite imagery (Fig. 2 B). However, all methods that incorporate a street map perform better, illustrating the potency of these features. Our custom aggregation strategy yields, on average, an improvement over a binary map (the setting analogous to [40]), though it does not universally improve results in all matchers tested. This leads us to believe that the additional texture provided by more detailed/realistic maps might be ignored by some descriptors (SIFT, ALIKED).

Best performance comes from the most realistic reference image - overlaying the street network on top of a daytime satellite image of the area. This was the best reference image type for *all* matchers, which we attribute to the more realistic simulation of the true scene compared to daylight or street only references, as well as the increase in matchable area compared to street maps alone.

AIMS-Night Localization. On the AIMS-Night evaluation set, our final method localizes 81.8% of images correctly (within 50 km of the ground truth center point). We use all inlier keypoints for our warps, producing an average reprojection error of 18.50 pixels (≈3% of image size) on confidently localized images. Samples of positive localizations are in the Supplementary (Fig. 9).

Failure Modes. Of the 20% of AIMS-Night that were not properly localized, most can be classified into a few common failure modes. First are areas with insufficient road data. In such cases, the reference maps are predominantly the satellite image background. The next most common failure case is with highly oblique images. Here, the view is taken at such an extreme angle that the features in the astronaut photo do not appear similar to those in the reference map. Carefully modeling the image acquisition angle and using it to add a perspective shift to the reference map helps matching in some of these conditions, but is still insufficient in the most extreme cases. Samples of failure cases can be found in the Supplementary (Fig. 10).

6.1. Ablations

Finally, we ablate components to better understand how each piece contributes to the method's success. We break down the pipeline per Tab. 2. The baseline method takes the naive approach to each aspect - generating a default count-aggregated street map, choosing cities by proximity to ISS nadir only, and matching only with a North up reference map. We use queue size of 50 cities.

Adding in city prioritization only has a small impact on localization performance (mainly where the true city was not in the naive top 50), but significantly improves runtime, as the matching city moves higher up in the search queue. We separately analyze the impact of ranking in Tab. 3. Our city ranking method reduces the average queue placement of the correct city by 50% over the best naive approach (population). As our method is complete upon finding a strong match (*i.e.* without having to visit all cities in the

	Matching Method	Thresh	Time (s)	% Strong Matches (N=363)				
				Day Sat	Streets - Bin	Streets - Custom	Streets + Sat	**Avg**
Crafted	SIFT - MNN [28]	25	13.34	7%	4%	16%	10%	9%
	BRISK - FLANN [26]†	5	4.93	6%	18%	20%	5%	12%
Dense	SE2-LoFTR* [8]	30	3.06	8%	11%	14%	18%	13%
Learned Detector	Steerers* [9]	100	3.37	56%	35 %	46%	72%	52%
	SuperPoint - LG [27, 37]	100	4.55	47%	62%	65%	74%	62%
	SIFT - LG [27, 28]	100	6.52	12%	15%	13%	20%	15%
	ALIKED - LG [27, 49]	100	5.07	57%	69%	68%	**78%**	68%
	DISK - LG [27, 46]	20	10.70	10%	25%	32%	20%	18%
	Average	-	6.44	30.4%	32.4%	37.3%	44.4%	-

Table 1. **AIMS-night Strong Matches with "Best Case" Reference Area.** A strong match is defined as $\# \text{ inliers}_{\text{Astro Photo} \Leftrightarrow \text{Ref Img}} > \text{thresh}_{\text{method}}$. Best overall result **bolded**, best along each axis underlined. Time is per image per query. † denotes our best attempt to re-implement the method in [40]. * indicates a rotation invariant matcher, so no orientation augmentation is performed.

Method Component	% Localized	Runtime/Image (s)
Baseline	19%	9.8
+ Orientation TTA (4x/8x)	51% / 61% (+ 32% / + 42%)	36.1 / 30.9
+ Streets/Sat Reference	73.75% (+12.75%)	39.1
+ City Prioritization	79.8% (+6.05%)	33.3
+ Perspective FOV	81.8% (+2%)	111.5

Table 2. **Ablations.** Impact of each component on localization performance and runtime. Queue size = 50 cities.

Queuing Method	Average Correct City Placement (↓)	
	2000 km	1200 km
Distance	25.8	21.8
Population	34.8	13.9
Custom (Eq. 2)	**9.8**	**7.1**

Table 3. **City Ranking Method Comparison.** Our city ranking method places the correct city higher in a queue than baselines, resulting in reduced runtimes for localization. Queues containing all cities within 2000 and 1200 km are tested. Lower is better.

queue), this reduces the runtime by 35 seconds on average.

The other ablations focus on the matching itself. Simple match-time orientation augmentation, to account for the unknown orientation of the astronaut photo, results in the largest improvement. Our enhanced, streets+satellite reference map improves the visual scene similarity to the nighttime photo, yielding a 12.75% improvement over a baseline street-only reference image. The perspective field of view, which more accurately models the scene area captured and applies a perspective warp to match the astronaut perspective, gives a small improvement particularly with more oblique imagery, but comes at a substantial runtime cost.

7. Conclusions

This work robustly addresses the nighttime astronaut photography of Earth localization problem for the first time. In compliment to recent successes in daytime astronaut photography localization, we introduce a fast method for localizing *nighttime* astronaut imagery. Using street vector data in combination with satellite imagery, we show we can produce a reference image to serve as a strong matching target for nighttime photography of the Earth. In lieu of a full scan of the visible Earth area, we prioritize city centers, making our method fast and efficient. Upon localization, we then georectify the nighttime image to produce an analysis-ready data product for nighttime light researchers.

Our experiments indicate that the quality of the reference image is of paramount importance. While both street vector data and daytime satellite imagery produce matchable reference images in many cases, each is missing a key component as the daytime satellite image does not properly emphasize roads, the primary feature of nighttime photos, and the street maps alone lack texture in areas without dense road coverage. The two datasets combined mitigate their respective weaknesses and significantly improve matching.

The matcher itself also plays a critical role. We find ALIKED-LightGlue to be a strong detector/matcher combination that identifies correct matches with high precision, while having very few false positives. In our testing, images not localized typically fall into categories that are known to be difficult for all modern matching methods, namely extreme perspective/scale changes and textureless pairs.

Finally, we collect and release AIMS-Night, an evaluation set that represents the challenges associated with nighttime astronaut photography localization. We hope releasing this set encourages future work on localizing this imagery.

This method is immediately applicable to the over two million nighttime astronaut photographs of Earth, transforming the collection into a high quality geolocated data product for analysis.

References

[1] Chow Ashley, Trulls Eduard, HCL-Jevster, Kwang Moo Yi, lcmrll, old ufo, Sohier Dane, WastedCode tanjigou, and Weiwei Sun. Image matching challenge 2023, 2023. 3

[2] Aayush Bansal, Hernán Badino, and Daniel F. Huber. Understanding how camera configuration and environmental conditions affect appearance-based localization. In *2014 IEEE Intelligent Vehicles Symposium Proceedings, Dearborn, MI, USA, June 8-11, 2014*, pages 800–807. IEEE, 2014. 3

[3] Daniel Barath, Jiri Matas, and Jana Noskova. MAGSAC: marginalizing sample consensus. In *Conference on Computer Vision and Pattern Recognition*, 2019. 6

[4] Giovanni Barbarani, Mohamad Mostafa, Hajali Bayramov, Gabriele Trivigno, Gabriele Berton, Carlo Masone, and Barbara Caputo. Are local features all you need for cross-domain visual place recognition? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2023. 3

[5] James A. Bednar, Joseph Crail, Ian Thomas, Jim Crist-Harif, Philipp Rudiger, Greg Brener, Chris B, Jon Mease, Julia Signell, Maxime Liquet, Jean-Luc Stevens, Brendan Collins, Simon Høxbro Hansen, thuydotm, Ajay Thorve, esc, kbowen, Nezar Abdennur, Oleg Smirnov, maihde, Adam Hawley, Andrii Oriekhov, Aron Ahmadia, Barry A Bragg Jr, Carlos H Brandt, Clemens Tolboom, Enno G., Erik Welch, James Bourbeau, and Johannes Jörg Schmidt. holoviz/datashader: Version 0.16.0, 2023. 4

[6] Gabriele Berton, Riccardo Mereu, Gabriele Trivigno, Carlo Masone, Gabriela Csurka, Torsten Sattler, and Barbara Caputo. Deep visual geo-localization benchmark. In *CVPR*, 2022. 3

[7] Gabriele Berton, Alex Stoken, Barbara Caputo, and Carlo Masone. Earthloc: Astronaut photography localization by indexing earth from space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 2

[8] Georg Bökman and Fredrik Kahl. A case for using rotation invariant features in state of the art feature matchers. In *CVPRW*, 2022. 3, 6, 8

[9] Georg Bökman, Johan Edstedt, Michael Felsberg, and Fredrik Kahl. Steerers: A framework for rotation equivariant keypoint descriptors. 2023. 3, 6, 8

[10] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000. 6

[11] D. M. Chen, G. Baatz, K. Köser, S. S. Tsai, R. Vedantham, T. Pylvänäinen, K. Roimela, X. Chen, J. Bach, M. Pollefeys, B. Girod, and R. Grzeszczuk. City-scale landmark identification on mobile devices. In *CVPR*, pages 737–744, 2011. 3

[12] Hongkai Chen, Zixin Luo, Lei Zhou, Yurun Tian, Mingmin Zhen, Tian Fang, David Mckinnon, Yanghai Tsin, and Long Quan. Aspanformer: Detector-free image matching with adaptive span transformer. In *ECCV*, 2022. 3

[13] Wei Chen, Yu Liu, Weiping Wang, Erwin M. Bakker, Theodoros Georgiou, Paul W. Fieguth, Li Liu, and Michael S. Lew. Deep image retrieval: A survey. *CoRR*, abs/2101.11282, 2021. 3

[14] Jiri Matas Dmytro Mishkin, Filip Radenovic. Repeatability Is Not Enough: Learning Discriminative Affine Regions via Discriminability. In *Proceedings of ECCV*, 2018. 3

[15] Christopher NH Doll, Jan-Peter Muller, and Jeremy G Morley. Mapping regional economic activity from night-time light satellite imagery. *Ecological Economics*, 57(1):75–92, 2006. 1

[16] Johan Edstedt, Ioannis Athanasiadis, Mårten Wadenbäck, and Michael Felsberg. DKM: Dense kernelized feature matching for geometry estimation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2023. 3

[17] Christopher Elvidge, Kimberly Baugh, Mikhail Zhizhin, and Feng-Chi Hsu. Why viirs data are superior to dmsp for mapping nighttime lights. *Proceedings of the Asia-Pacific Advanced Network*, 35:62–69, 2013. 1

[18] EOX IT Services GmbH. `https://s2maps.eu`. CC Non Commercial Share Alike 4.0. 5

[19] Fabio Falchi, Pierantonio Cinzano, Dan Duriscoe, Christopher C. M. Kyba, Christopher D. Elvidge, Kimberly Baugh, Boris A. Portnov, Nataliya A. Rybnikova, and Riccardo Furgoni. The new world atlas of artificial night sky brightness. *Science Advances*, 2(6):e1600377, 2016. 1

[20] GDAL/OGR contributors. *GDAL/OGR Geospatial Data Abstraction software Library*. Open Source Geospatial Foundation, 2024. 6

[21] Dihe Huang, Ying Chen, Shang Xu, Yong Liu, Wenlong Wu, Yikang Ding, Chengjie Wang, and Fan Tang. Adaptive assignment for geometry aware local feature matching, 2022. 3

[22] Yuhe Jin, Dmytro Mishkin, Anastasiia Mishchuk, Jiri Matas, Pascal Fua, Kwang Moo Yi, and Eduard Trulls. Image matching across wide baselines: From paper to practice. *CoRR*, abs/2003.01587, 2020. 3

[23] Bradley Johnson. Detecting impervious cover with artificial lighting in astronaut photography from the international space station. 2020. 1

[24] Konstantin Klemmer, Esther Rolf, Caleb Robinson, Lester Mackey, and Marc Rußwurm. Satclip: Global, general-purpose location embeddings with satellite imagery. *arXiv preprint arXiv:2311.17179*, 2023. 3

[25] Helga U. Kuechly, Christopher C.M. Kyba, Thomas Ruhtz, Carsten Lindemann, Christian Wolter, Jürgen Fischer, and Franz Hölker. Aerial survey and spatial analysis of sources of light pollution in berlin, germany. *Remote Sensing of Environment*, 126:39–50, 2012. 4

[26] Stefan Leutenegger, Margarita Chli, and Roland Y. Siegwart. Brisk: Binary robust invariant scalable keypoints. In *2011 International Conference on Computer Vision*, pages 2548–2555, 2011. 3, 6, 8

[27] Philipp Lindenberger, Paul-Edouard Sarlin, and Marc Pollefeys. LightGlue: Local Feature Matching at Light Speed. In *ICCV*, 2023. 3, 6, 8

[28] David G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004. 3, 6, 8

[29] W. Maddern, G. Pascoe, C. Linegar, and P. Newman. 1 Year, 1000km: The Oxford RobotCar Dataset. 2017. 3

[30] Gengchen Mai, Ni Lao, Yutong He, Jiaming Song, and Stefano Ermon. Csp: Self-supervised contrastive spatial pre-training for geospatial-visual representations. In *International Conference on Machine Learning*. PMLR, 2023. 3

[31] Johan R Meijer, Mark A J Huijbregts, Kees C G J Schotten, and Aafke M Schipper. Global patterns of current and future road infrastructure. *Environ. Res. Lett.*, 13(6):064006, 2018. 2, 4

[32] Anastasiya Mishchuk, Dmytro Mishkin, Filip Radenovic, and Jiri Matas. Working hard to know your neighbor's margins: Local descriptor learning loss. In *Proceedings of NeurIPS*, 2017. 3

[33] OpenStreetMap contributors. Planet dump retrieved from https://planet.osm.org . `https://www.openstreetmap.org`, 2017. 2, 4

[34] Tom Patterson, Nathaniel Vaughn Kelso, and contributors. Natural earth. Technical report, 2024. 6

[35] Alejandro Sanchez de Miguel, Jaime Zamorano, Martin Aubé, Jonathan Bennie, Jesús Gallego, Francisco Ocaña, Donald Pettit, William Stefanov, and Kevin Gaston. Colour remote sensing of the impact of artificial light at night (ii): Calibration of dslr-based images from the international space station. 2021. 1

[36] Paul-Edouard Sarlin, Cesar Cadena, Roland Siegwart, and Marcin Dymczyk. From coarse to fine: Robust hierarchical localization at large scale. In *CVPR*, 2019. 3

[37] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. SuperGlue: Learning feature matching with graph neural networks. In *CVPR*, 2020. 3, 6, 8

[38] T. Sattler, W. Maddern, C. Toft, A. Torii, L. Hammarstrand, E. Stenborg, D. Safari, M. Okutomi, M. Pollefeys, J. Sivic, F. Kahl, and T. Pajdla. Benchmarking 6DOF outdoor visual localization in changing conditions. In *CVPR*, pages 8601–8610, 2018. 3

[39] Torsten Sattler, William P. Maddern, Carl Toft, Akihiko Torii, Lars Hammarstrand, Erik Stenborg, Daniel Safari, M. Okutomi, Marc Pollefeys, Josef Sivic, Fredrik Kahl, and Tomás Pajdla. Benchmarking 6dof outdoor visual localization in changing conditions. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8601–8610, 2018. 3

[40] Peter Schwind and Tobias Storch. Georeferencing urban nighttime lights imagery using street network maps. *Remote Sensing*, 14(11), 2022. 3, 6, 7, 8

[41] Alex Stoken and Kenton Fisher. Find my astronaut photo: Automated localization and georectification of astronaut photography. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 6196–6205, 2023. 2, 3, 5, 6, 7

[42] Jiaming Sun, Zehong Shen, Yuang Wang, Hujun Bao, and Xiaowei Zhou. LoFTR: Detector-free local feature matching with transformers. *CVPR*, 2021. 3

[43] Alejandro Sánchez de Miguel, Christopher C.M. Kyba, Martin Aubé, Jaime Zamorano, Nicolas Cardiel, Carlos Tapia, Jon Bennie, and Kevin J. Gaston. Colour remote sensing of the impact of artificial light at night (i): The potential of the international space station and other dslr-based platforms. *Remote Sensing of Environment*, 224:92–103, 2019. 1

[44] A. Torii, R. Arandjelović, J. Sivic, M. Okutomi, and T. Pajdla. 24/7 place recognition by view synthesis. *IEEE TPAMI*, 40(2):257–271, 2018. 3

[45] A. Torii, Hajime Taira, Josef Sivic, M. Pollefeys, M. Okutomi, T. Pajdla, and Torsten Sattler. Are large-scale 3d models really necessary for accurate visual localization? *IEEE TPAMI*, 43:814–829, 2021. 3

[46] Michał Tyszkiewicz, Pascal Fua, and Eduard Trulls. Disk: Learning local features with policy gradient. *Advances in Neural Information Processing Systems*, 33, 2020. 8

[47] Vicente Vivanco, Gaurav Kumar Nayak, and Mubarak Shah. Geoclip: Clip-inspired alignment between locations and images for effective worldwide geo-localization. In *Advances in Neural Information Processing Systems*, 2023. 3

[48] Tobias Weyand, A. Araújo, Bingyi Cao, and Jack Sim. Google landmarks dataset v2 – a large-scale benchmark for instance-level recognition and retrieval. In *CVPR*, pages 2572–2581, 2020. 3

[49] Xiaoming Zhao, Xingming Wu, Weihai Chen, Peter C. Y. Chen, Qingsong Xu, and Zhengguo Li. Aliked: A lighter keypoint and descriptor extraction network via deformable transformation. *IEEE Transactions on Instrumentation Measurement*, 72:1–16, 2023. 6, 8