

Spatial and Spatiotemporal Analysis of Solar Granulation

Reza Mansouri¹, Prithvi Reddy¹, Natnael Alemayehu¹, Aya Abdelkarem¹, Rafal Angryk²

Georgia State University

Atlanta, GA

Email: ¹{rmansouri1, preddy10, nalemayehu3, aabdelkarem1}@student.gsu.edu

²rangryk@gsu.edu

Abstract—The solar photosphere is covered with a cell-like pattern known as solar granulation. The granules have been associated with different physical phenomena taking place on the photosphere. In this study, we leverage a 200-frame timeseries of machine-generated segmentation masks of images from the Daniel K. Inouye Solar Telescope (DKIST) to create a spatial database. We establish an infrastructure allowing spatial/spatiotemporal analysis of this granular pattern, and further, show the effectiveness of different spatial indexes for this particular data. Source code is available at <https://github.com/rezmansouri/granules-sdb>.

Index Terms—Solar physics, Solar granulation, Spatial databases, Spatial query, Database indexing

I. INTRODUCTION

The Sun's surface, observed in broadband visible light, reveals a distinctive granulation pattern. Apart from the sunspots, which are regions of intense magnetic flux, the surface predominantly displays this cellular structure. Granules are bright, dome-like formations with horizontal dimensions of several thousand kilometers that undergo rapid changes within a matter of minutes [1].

Granules have different types and shapes, with each being associated to a specific underlying event. In this study we focus on four types of granules:

- 1) Granules with dots: also known as *exploding granules*, are large granules that expand rapidly, displaying a dark central dot before fragmenting [2], [3]. They are closely linked to small-scale magnetic flux [4], [5].
- 2) Granules with lanes: they exhibit arch-like features that loop across their borders, partially covering them. These arches are closely associated with the emergence of linear magnetic fields, suggesting underlying vortex flow tubes [6].
- 3) Uniform granules: they are distinguished by their circular or elliptical shapes, which set them apart from other granules. Their well-defined and symmetrical morphology reflects stable underlying processes, making them a key focus in the study of solar granulation dynamics.
- 4) Complex granules: these granules exhibit highly complex morphologies, characterized by numerous twirls and spins. Their intricate structures do not fit within the standard classifications, making them stand out with uniquely distinct features.

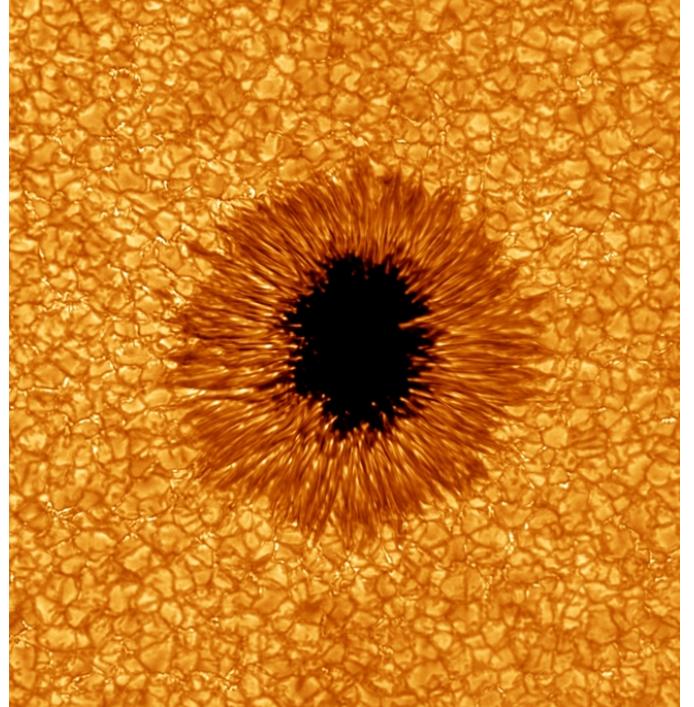


Fig. 1. A sunspot surrounded with the granulation pattern

As granules form and evolve on the Sun's surface, they are often bordered by dark intergranular lanes, which are integral to solar convection dynamics [1]. These lanes result from cooler plasma descending, creating a contrast between the bright granules and the surrounding dark regions.

From a supervised machine learning point of view, identification and classification of the granules from the images is a pixel-level classification problem, or the so-called semantic segmentation. Deep learning models using fully convolutional layers, such as U-Net [7]—originally designed for medical image segmentation—have proven effective for this purpose. U-Net performs systematic feature extraction through its encoder and subsequently upscales these features in its decoder, matching the input dimensions but with the output representing class predictions at the pixel level.

Several variants of U-Net have been developed, each en-

hancing the standard encoder-decoder architecture to address specific challenges in segmentation tasks. For example, the U-Net++ [8], employs a series of entangled skip connections with additional convolutional layers on the way, which help to progressively reduce the semantic gap between feature maps from the encoder and decoder sub-networks, leading to more accurate feature representation.

Semantic segmentaion of solar granulation was first done by [9]. They used images from the IMaX instrument on the SUNRISE balloon-borne telescope and annotated them with five distinct pattern types observed on the solar surface, and employed the original U-Net model.

In this study we use segmentation masks of images of solar granulation from the DKIST telescope to create a spatial database of granules, enabling us to analyze them from a spatial/spatiotemporal point of view. The rest of the paper is as follows: in section two, we go over our methodologies, notably how the segmentation masks were created, our database schema, and definition of some key challenges related to this spatial analysis and our proposed solutions for them. In section 3, some examples of these challenges are provided and the effectiveness of spatial indexes is investigated. We finally conclude with a summary of this study and provide our suggestions for future work.

II. METHODOLOGIES

A. Dataset

Our dataset comprises 200 time-series image sequences captured by the DKIST telescope, with a cadence of 6 seconds and a field of view (FOV) spanning 45×45 arcseconds. These images are unlabeled. To address this, we employed a transfer learning approach: a U-Net++ model was trained using labeled segmentation masks from IMaX SUNRISE images. The trained model was then used to generate segmentation masks for the DKIST images, leveraging the similarity between the DKIST and IMaX datasets. Since IMaX images have a slightly smaller FOV of 38×38 arcseconds, DKIST images were first center-cropped to match this FOV, giving us frame sizes of 3454×3454 pixels. Additionally, histogram matching was performed on each DKIST image, aligning their intensity distributions with those of the combined IMaX dataset to ensure consistency and improve segmentation accuracy. This approach gave us a timeseries of 200 segmentation masks for the five classes of granulation pattern, ready to be imported into our database.

Moreover, it should be noted that as spatial reference systems (SRS) utilize world coordinates, such as longitude and latitude in degrees, in our database each degree corresponds to one pixel in the segmentation masks. We employed the common SRS, EPSG:4326.

B. Database engine

With various database engines supporting geometrical data types, PostgreSQL with PostGIS (latest version, 3.5, as of Dec. 2024) aligned perfectly with our research objectives. Its robust support for spatial data, scalability for managing large datasets,

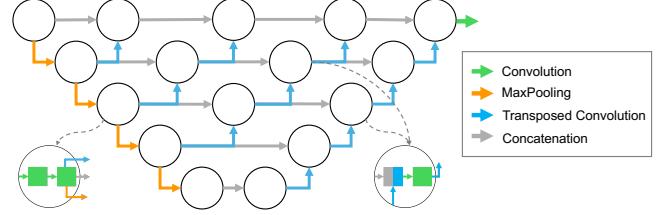


Fig. 2. The U-Net++ architecture

and compatibility with temporal extensions position it as an optimal choice for our project. The system not only meets the immediate requirements of storing and querying spatial data but also offers the flexibility to evolve with the project, ensuring that it remains a cornerstone of our data management and analytical infrastructure.

Although our current focus is on the spatial dimensions of our dataset, PostGIS's integration with PostgreSQL's temporal features, bolstered by extensions like PostGIS-T, positions it as a robust solution for managing spatiotemporal data. These temporal capabilities enable the storage and analysis of evolving phenomena, aligning seamlessly with our goal of linking granules across time frames to capture their dynamic progression. The ability to analyze observational data over time, as emphasized in related research, is essential for understanding the evolution of spatial patterns, particularly in remote sensing contexts such as ours.

Moreover, PostGIS-T offers additional functions and customizable views for spatiotemporal data manipulation, providing the flexibility needed for advanced and tailored analyses. This adaptability underscores PostgreSQL's strength in integrating spatial and temporal relationships, paving the way for novel insights in our research. Finally, as an open-source platform, PostGIS not only ensures cost-effective accessibility but also benefits from a vibrant community and extensive documentation, minimizing barriers to adoption while offering the sophisticated capabilities demanded by academic investigations [10].

Compared to alternative spatial database engines, PostGIS stands out for its balance of power, cost-efficiency, and extensibility. While systems such as Oracle 11g has its strengths, particularly in certain functionalities, it tends to lag behind Postgres in terms of performance for spatial data operations due to its geometric algorithm design, query execution times, memory management, and index utilization strategies. Similarly, lightweight databases like SpatiaLite, while useful for small-scale applications, lack the scalability and advanced querying functionalities required for analyzing the dynamic granules in our study. Cloud-based solutions such as Amazon Aurora and Google BigQuery often incur recurring expenses for storage, computation, and bandwidth usage. In the context of our project, which involves analyzing and storing spatial data derived from high-resolution solar images, these costs can

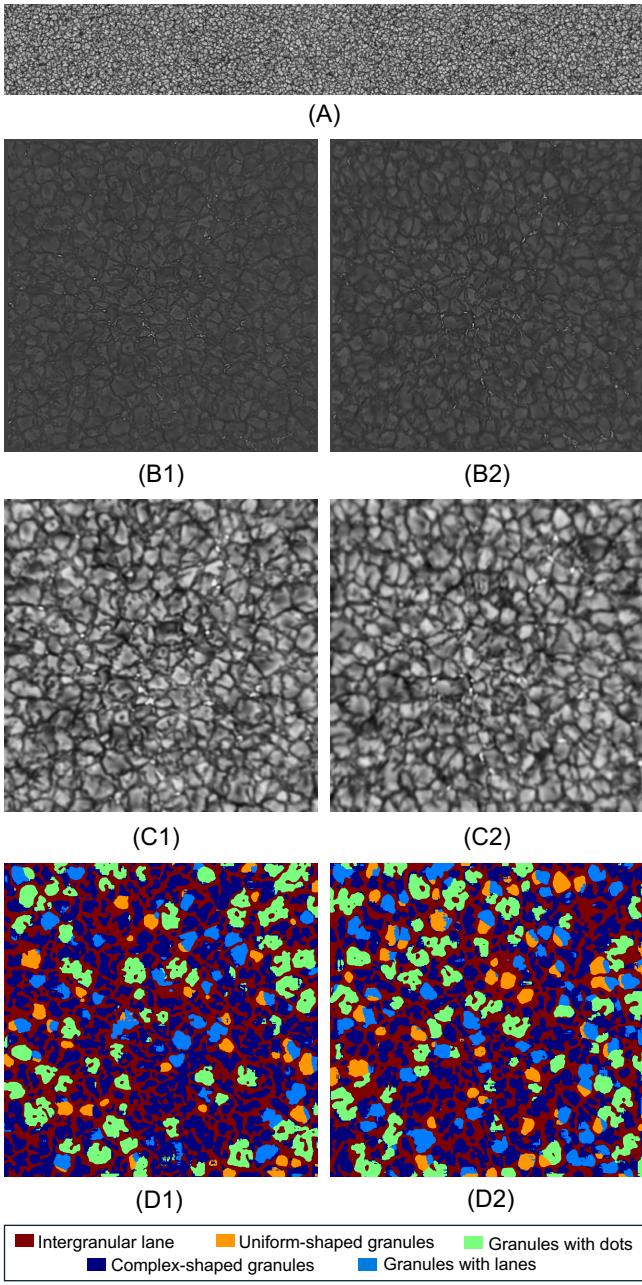


Fig. 3. Segmentation masks for two DKIST instances (B), transformed instances (C) to represent IMaX data (A), and predicted segmentation masks (D)

escalate significantly given the volume of data and the need for long-term storage. In contrast, PostgreSQL with PostGIS is an open-source solution, offering cost-effectiveness without compromising functionality. PostGIS integrates seamlessly with Python libraries such as GeoPandas and psycopg2, enabling a streamlined workflow for spatial and machine learning tasks, and supports visualization tools like QGIS, which are essential for validating and presenting our findings.

C. Database schema

To systematically represent the granules in our study, we defined an entity for each granule type. These entities encapsulate critical attributes for individual granules, including a unique identifier (`id`), the specific time of its occurrence (`time`), and its geometric representation. The extraction of these granules from segmentation masks was accomplished through a 1-closest neighbor clustering algorithm. This approach groups together pixels classified as belonging to the same granular pattern class, effectively isolating individual granules based on their spatial coherence within the mask.

However, granules are not static and need to be tracked over time to understand their evolution and interactions. To address this, we introduced an additional entity called the `tracked_granule`. This entity captures the temporal dynamics of granules by including attributes such as a unique identifier (`id`), the initial timestep of its presence (`birth`), and the final timestep of its existence (`death`). By associating granules from individual timesteps with their corresponding tracked granules, we establish a framework for representing their spatiotemporal features. It should be noted that we chose not to track the intergranular lane class as it is the background class and not that important.

This dual-level representation—granules at individual timesteps and their aggregation into tracked granules—enables a detailed analysis of their evolution over time. The methodology for tracking granules and associating them across different timesteps will be explained in detail in the subsequent sections. This approach ensures a comprehensive understanding of both the spatial and temporal characteristics of the granules under study.

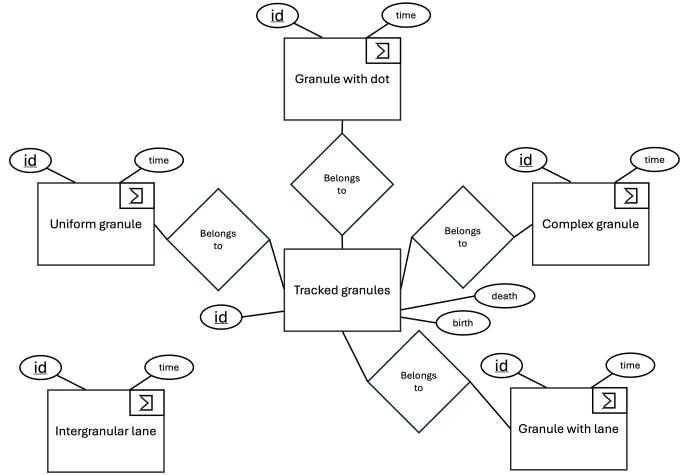


Fig. 4. The Entity-Relationship diagram of the database

D. Data cleaning

Misclassification in solar granule datasets, arising from pixel-level classification, presents a significant challenge in geospatial analysis. The dataset comprised geometrical representations of solar granules, categorized into four distinct

types. Preliminary analysis revealed that a considerable number of granules with small areas were prone to misclassification, primarily due to the inherent limitations of pixel-based classifiers. To address this issue, a systematic cleaning approach was employed, leveraging area-based filtering to refine the dataset and improve classification accuracy. The methodology began with the computation of geometric areas for all granules. The dataset, stored initially in Well-Known Binary (WKB) format, was transformed into a geospatially operable structure using GeoPandas. The computed areas were analyzed through a histogram, which revealed a strongly right-skewed distribution. This skewness indicated a disproportionately high occurrence of small-area granules, many of which were incorrectly classified.

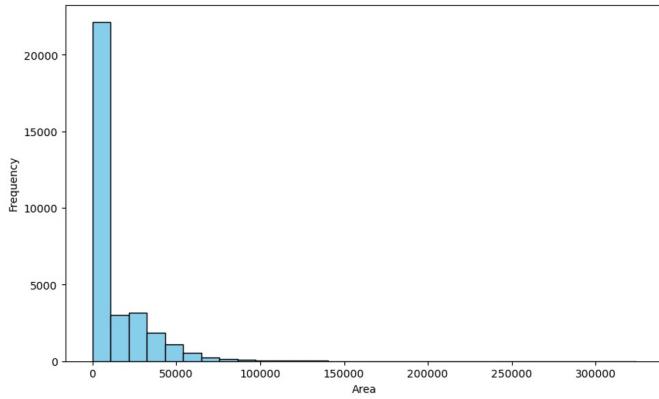


Fig. 5. Distribution of areas of granules with a dot

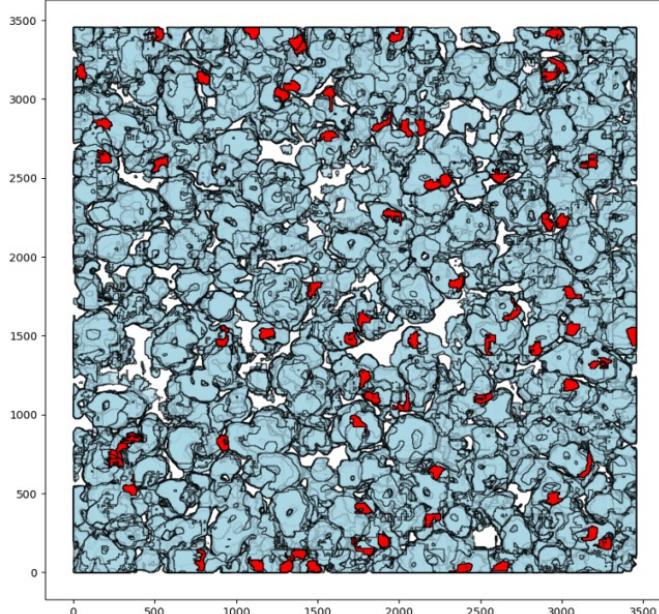


Fig. 6. Granules with a dot having areas 5000 ± 100 px² (possibly misclassified and invalidated in our cleaning procedure)

To address this, minimum area thresholds were iteratively established for each granule type by combining statistical distribution analysis with spatial visualizations. A threshold of 12,000 units was set for granules with dots, 7,000 units for uniformly shaped granules, and 10,000 units for granules with lanes, and 2500 units for complex granules. Granules with areas below these thresholds were flagged as potential misclassifications and excluded from the dataset.

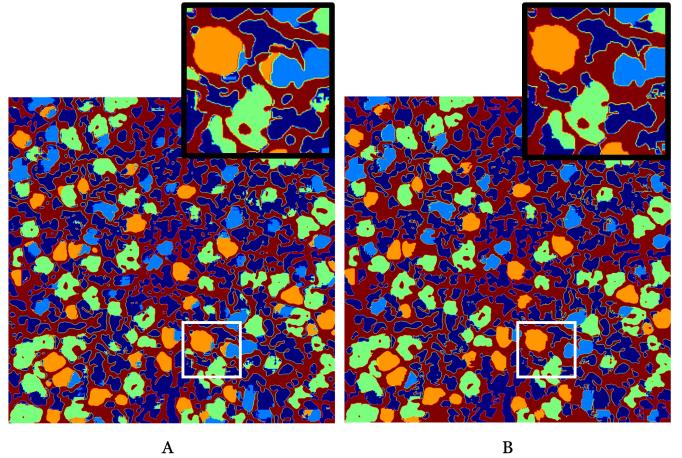


Fig. 7. Result of the cleaning procedure: (A) the original segmentation mask (B) cleaned segmentation mask with the area thresholds applied

E. Granule tracking

Starting with the granules at the final timestep, we iteratively associate granules from the previous timestep to those currently being examined. This association process takes into account two key factors: the Intersection over Union (IoU) between a granule and granules in the subsequent timestep, and the probability of a granule changing its type between timesteps.

For type transitions, we assume a uniform distribution probability for granules changing their type, but with a higher likelihood of retaining their original type. Essentially, a granule has approximately a 49% probability of maintaining its type, while the probability of transitioning into any other type is around 17%. The association score for each granule is then computed as the product of its IoU and the corresponding type transformation probability.

Using these association scores, granules from the previous timestep are linked to those in the current timestep. Specifically, the granule from the prior timestep that yields the highest association score with each granule being iterated over in the current timestep is assigned the `tracked_id` of that in the current timestep. This iterative process ensures consistent tracking of granules across timesteps.

F. Queries

1) *Spatial join*: The problem of identifying granules that merge over time was explored using spatial joins, where

granules from one time frame (time t_0) were analyzed for spatial intersections with granules from the next immediate time frame (time t_1). The geometries of granules at time t_1 were compared to those at time t_0 using the PostGIS ST_Intersects function to detect overlaps between their boundaries. A merging count was assigned to granules at time t_1 based on how many granules at time t_0 intersected with their geometry. Granules with a merging count of two or more were classified as mergers. This approach ensured temporal consistency by limiting the analysis to consecutive timeframes, accurately capturing the evolution of granules over time.

2) *Minimum/maximum distance between granules:* To analyze the spatial relationships between granules, we computed the minimum and maximum edge distances between their boundaries within the same timeframe. Using PostGIS ST_Boundary function, the leftmost and rightmost edges of each granule were extracted. The minimum edge distance was calculated as the shortest horizontal distance between the rightmost edge of one granule and the leftmost edge of another granule. The maximum edge distance was calculated as the horizontal distance from the leftmost edge of one granule to the rightmost edge of the other. These measurements provided a comprehensive view of granule proximity and spatial spread.

3) *Distance between granules of different types:* Another query of interest was to calculate the minimum distance between granules of different types. To achieve this, we calculated the minimum distance between a granule and the nearest granule of a different type within the same timeframe. By restricting the analysis to a single timeframe, we avoided the influence of temporal variability on spatial measurements, maintaining consistency across the dataset. PostGIS's ST_Distance function was used to compute the shortest distance between the boundaries of granules. The analysis filtered granules by type, ensuring comparisons were only made between different types.

4) *Range query:* An essential aspect of this study can involve examining the spatial proximity of granules, either across all granule classes or within a specific class. This analysis is conducted using range queries in a spatial database, identifying granules located within a defined distance from a particular granule. By investigating these proximities, we could discern patterns in granule clustering or dispersion that may provide insights into their spatial organization. To ensure rigor and temporal accuracy, this analysis is performed within discrete timeframes, capturing the spatial relationships of granules at specific moments without conflating temporal variations.

5) *Temporal pattern mining:* Another critical dimension of the analysis focuses on morphological transformations of solar granules, identifying granules that exhibit significant temporal variations in size, such as the most rapid increases or decreases.

Growth analysis of solar granules is inherently complex and prone to ambiguity due to the interplay between temporal dynamics and size variations. For instance, a granule that doubles its size within one second and another that grows

tenfold over ten seconds may exhibit the same overall growth rate when assessed over a longer duration. However, these two scenarios represent fundamentally distinct growth behaviors—one characterized by rapid, short-term expansion and the other by gradual, sustained growth. Such disparities highlight the importance of accounting for temporal resolution in growth analysis, as overlooking this dimension can lead to erroneous or overly generalized conclusions.

To mitigate these challenges, we adopt a methodological framework that narrows the scope of analysis to specific time windows, such as one-minute intervals, further divided into equal temporal frames. This approach enables the systematic tracking of granule growth within uniform and consistent temporal segments, thereby facilitating a more nuanced examination of growth dynamics. By isolating changes over controlled timeframes, we capture both rapid and gradual transformations with improved clarity and precision. This strategy balances the need for temporal granularity against the computational demands of processing extensive datasets, ultimately offering a more robust and contextualized understanding of the growth patterns of solar granules.

6) *Clustering of granules:* Clustering of granules is important because it allows for a better understanding of the spatial distribution and dynamics of solar features. The steps include:

- 1) Data extraction: The study begins with data extraction from the database. The data contains the centroids of polygons representing various types of granules.

```
SELECT
    id,
    ST_X(ST_Centroid(shape)) AS x,
    ST_Y(ST_Centroid(shape)) AS y,
    'granule_type_1*' AS granule_type
FROM granule_type_1 UNION ALL ...;
```

The SQL query retrieves the *id*, centroid coordinates (*x*, *y*), and the type of granule for four categories: Uniform granules, Granules with dots, Granules with lanes, and Complex granules. Using the ST_Centroid() function, the geometric centroids of each polygon are calculated. The ST_X() and ST_Y() functions then extract their X and Y coordinates, which are unified into a single dataset using the UNION ALL operator. The granule type is designated as a categorical attribute to facilitate differentiation during analysis.

- 2) Clustering analysis: For clustering granules based on spatial proximity, the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm [11] is employed. DBSCAN is particularly suitable for this task due to its ability to detect arbitrarily shaped clusters and handle noise points effectively. The parameters for the algorithm include:

- *eps*: This parameter defines the maximum distance between two samples for them to be considered part of the same cluster. A value of 5 was chosen to capture spatial relationships within a reasonable radius.

- `min_samples`: This is the minimum number of points required to form a cluster, set to 10 to ensure that clusters are well-defined.
- `metric`: The Euclidean distance metric is used to measure spatial proximity between points.

G. Spatial indexes

Spatial indexes are essential for efficiently executing geospatial queries in PostGIS, particularly when dealing with large, multidimensional datasets. These indexes facilitate rapid filtering of spatial data, minimizing the computational overhead associated with processing complex geometries. The most commonly used index is the GiST (Generalized Search Tree), which functions as an R-Tree, a balanced tree structure specifically designed for spatial data. GiST divides the data space into bounding boxes, which act as envelopes for geometry objects. During queries, GiST filters out irrelevant bounding boxes first, significantly reducing computational load. For example, operations like `ST_Intersects` or `ST_Contains` use these bounding boxes to quickly identify potential matches [12].

For datasets with non-uniform distributions, SP-GiST (Space-Partitioned GiST) index optimizes performance by partitioning the space into smaller subspaces. This makes it particularly effective for indexing point data and geometries in sparse or clustered datasets. SP-GiST creates hierarchies within the data, functioning as a quad tree (grid file) structure, which enhances query efficiency, such as nearest-neighbor searches `ST_NearestNeighbor` or containment checks `ST_Contains` [12].

In our database, we aim to evaluate the effectiveness of the GiST (Generalized Search Tree) and SP-GiST (Space-Partitioned GiST) indexes on the `shape` attribute of our entities by applying both point and range queries.

III. RESULTS AND DISCUSSION

A. Spatial join

For the granule merge query we mentioned earlier we analysed the first two time frames in the dataset. At time $t_0 = 2022-05-26 17:46:29$ and time $t_1 = 2022-05-26 17:46:35$, Granule 7543 at time t_1 was identified as a result of a significant merging event involving 12 granules from time t_0 , figure 8. This large-scale merger highlights the dynamic nature of granule interactions and their propensity to coalesce into larger structures over short timeframes.

The merging event represents one of the more complex cases observed in the study, where multiple granules from diverse regions at time t_0 contributed to the formation of a single, larger granule at time t_1 . Such large-scale mergers are indicative of high-density regions on the solar surface, where granules interact more frequently.

B. Minimum/maximum distance between granules

For the minimum and maximum edge distance calculation we provide here the case we investigated for 2 uniform granules with Id 1 and Id 2. As can be seen in the figure

below as well at time $t = 2022-05-26 17:46:29$, the maximum edge distance between Granule 1 and Granule 2 was calculated as 1908.55 pixel units, representing the farthest horizontal separation from the leftmost edge of Granule 1 to the rightmost edge of Granule 2. The minimum edge distance, measured as 1574.15 pixel units, was determined as the shortest horizontal separation between the rightmost edge of Granule 1 and the leftmost edge of Granule 2, figure 9.

C. Distance between granules of different types

For the nearest other neighbor query, here we demonstrate the results for a uniform granule with tracking id of 7459. At 2022-05-26 17:46:29, Granule 7459 (a uniform granule) was found to have a nearest neighbor in Granule 5662 (a complex granule) with a minimum distance of 18.44 units. This distance was calculated as the shortest line segment connecting the boundaries of the two granules, figure 10.

D. Range query

The problem of inspecting granules from a proximity perspective and the corresponding solution are demonstrated in this section. The DKIST telescope, with a field of view covering 38 arcseconds, generates images comprising 3454 pixels in total. For the range query in this example, the target distance of 5 arcseconds was converted into degrees, accounting for the pixel-level granularity of the images. This adjustment ensured that spatial distances were accurately represented within the database's spatial reference system (SRS), defined in degrees, while maintaining consistency with the pixel-based classification of granules.

A range query was performed to identify granules within a specified proximity to a target granule, utilizing spatial filtering techniques. The visualization showcased all granules in the dataset, highlighting the target granule in green and granules within the range in red for clarity, the following are two examples:

- 1) Range Query Within a Single Granule Class: In this example, the range query was restricted to granules within the granule with dot class. A target granule was identified ($id=1$), at time a specific time instant, its centroid was calculated and leveraging spatial filters, we queried for granules within 5 arcseconds of range.
- 2) Range Query Across All Granule Types: The approach involved aggregating granule data from multiple tables into a unified dataset to facilitate spatial analysis.

E. Temporal pattern mining

Our analysis revealed insightful patterns of growth among solar granules within specific time windows. To identify rapid changes in granule size, a threshold of 90% growth within the defined time window was established. Granules exceeding this threshold were classified as experiencing rapid growth.

- Rapid increase of growth in size: One notable example of rapid growth was observed in the granule with tracked_id 174 during a one-minute time window from 18:05:34 to 18:06:35. Within this timeframe, the granule exhibited an

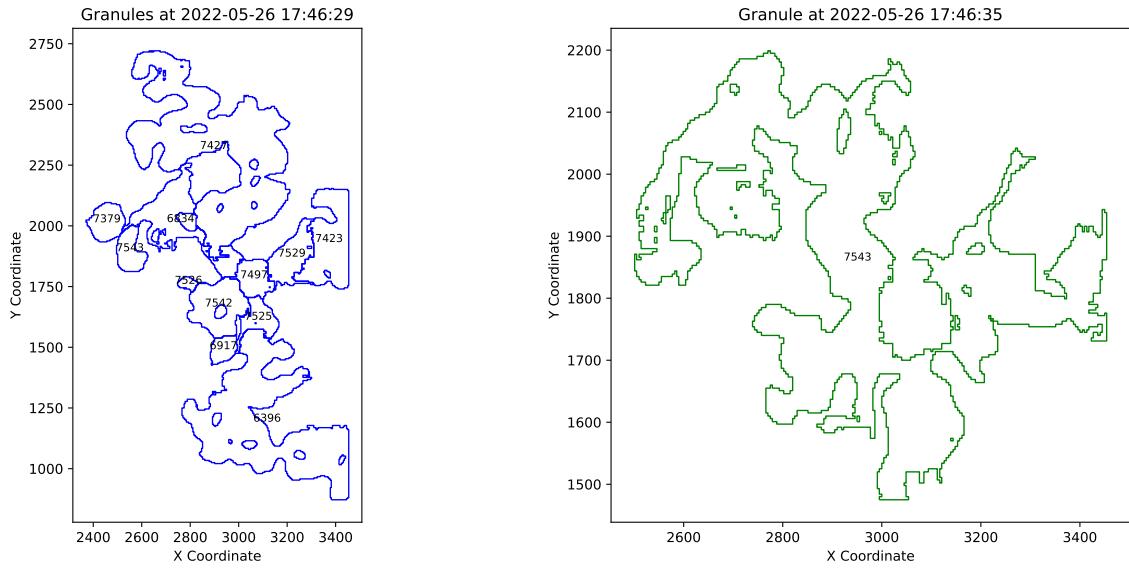


Fig. 8. Results from the merging investigation for a specific time frame: twelve granules merging into one

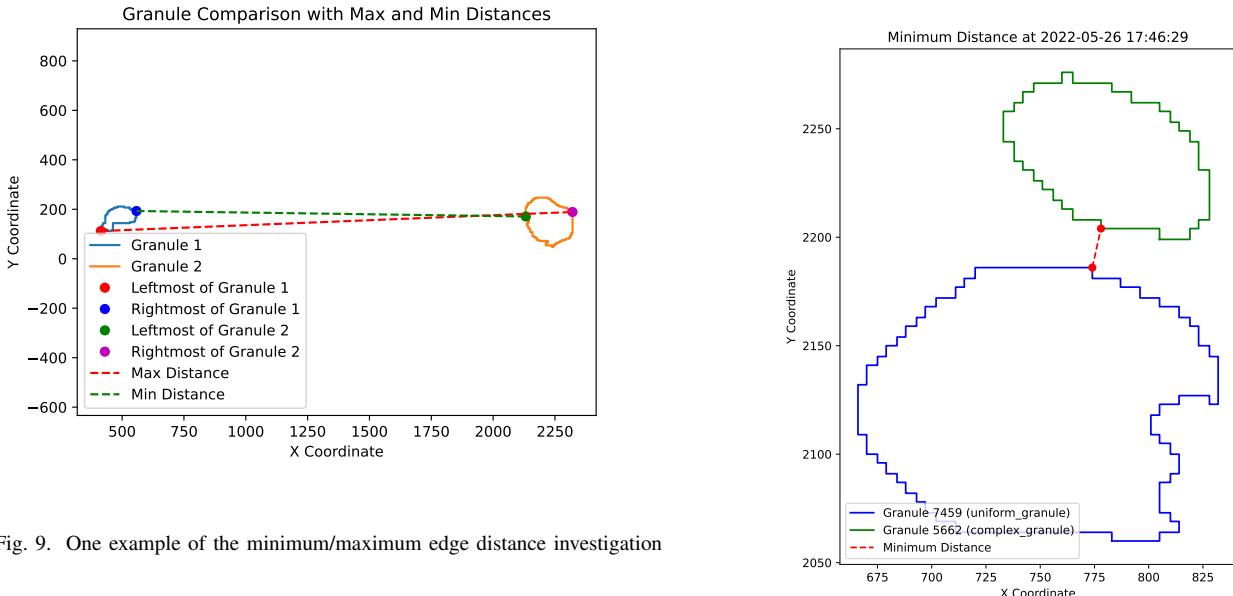


Fig. 9. One example of the minimum/maximum edge distance investigation

Fig. 10. One example of the minimum distance between two different types of granules

extraordinary size increase of 1409.48%, growing from an initial area of 10,257 to a final area of 154,827. Additionally, this granule underwent a morphological transformation during the same interval, transitioning from a complex granule to a granule with a dot.

- Rapid decrease of growth in size: To identify rapid decreases in granule size, a threshold of -90% growth within the specified time window was applied. Granules meeting or exceeding this negative growth threshold were classified as experiencing rapid size reduction. A representative case of rapid decrease was observed in the granule with tracked_id 6455 during a one-minute

window from 17:47:18 to 17:48:19. During this interval, the granule's area decreased by 98.89%, shrinking from an initial size of 259,979 to a final size of 2,881. Unlike the previous example of rapid increase, no morphological transformation occurred in this instance; the granule remained classified as a complex granule throughout the

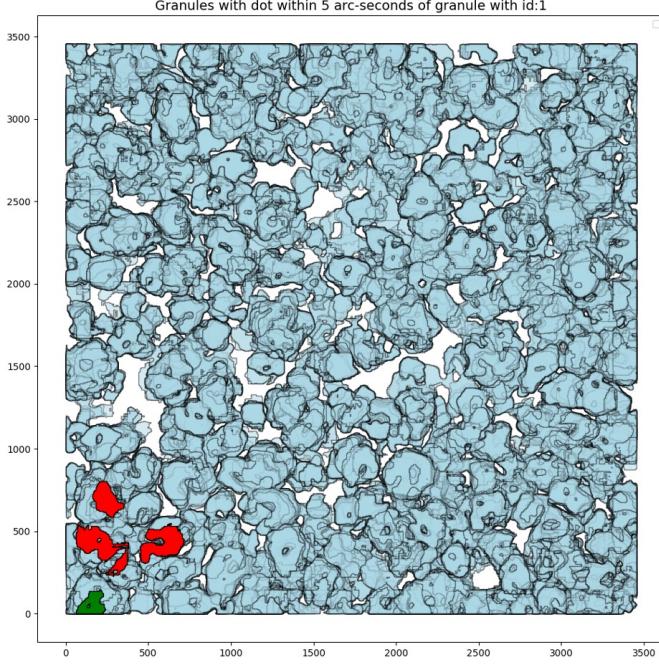


Fig. 11. One example of the range query within a single granule type

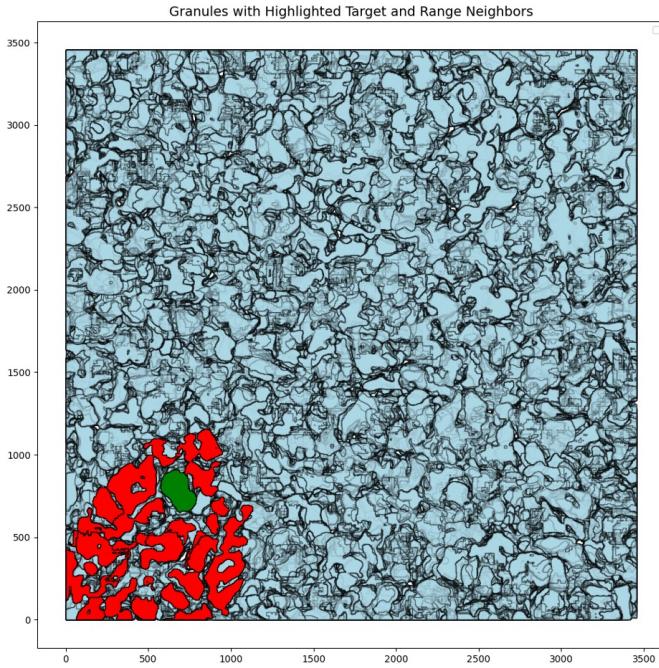


Fig. 12. Another example of the range query within all granule types

observation period. Within this same time window, the granule exhibited a transient phase of rapid growth during the first 18 seconds, achieving a 121.08% increase from its starting size. However, in the subsequent 42 seconds, its size continuously dropped, culminating in the overall

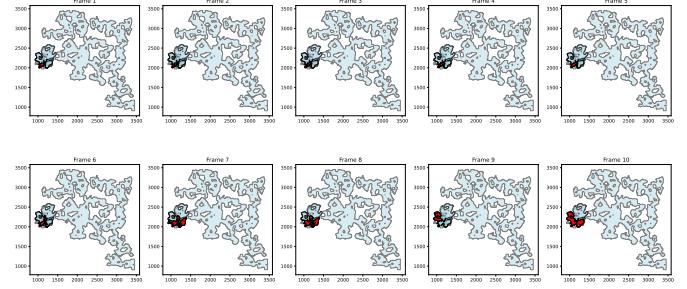


Fig. 13. Rapid growth of size of a granule across time

dramatic reduction. Despite these fluctuations in size, the granule's classification remained consistent as a complex granule throughout the observation period.

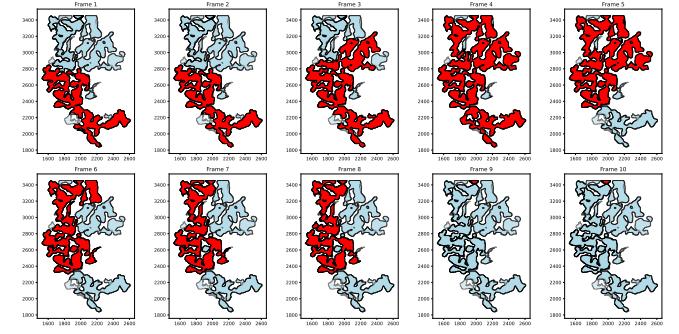


Fig. 14. Rapid decrease of size of a granule across time

This example underscores the complex and dynamic behavior of granules, where both growth and decline trends can co-occur within a short timeframe. Such heterogeneity within a single granule's life cycle reveals the need to examine temporal trends at finer scales, as rapid transformations may be context-dependent or influenced by localized solar dynamics.

F. Clustering of granules

The analysis of solar granules using DBSCAN clustering and density visualization highlights significant patterns in their spatial distribution and clustering characteristics. The clustering results reveal that each granule type exhibits unique clustering behaviors and density distributions. Uniform granules are evenly distributed with distinct, compact clusters, while granules with dots and granules with lanes show more irregular patterns, with the latter forming elongated clusters indicative of solar convection flows or magnetic field alignment. Complex granules dominate the spatial domain with dense, overlapping clusters, reflecting their dynamic and irregular behavior.

The heatmaps further validate these findings, emphasizing regions of high granule activity (yellow) and sparse, quieter

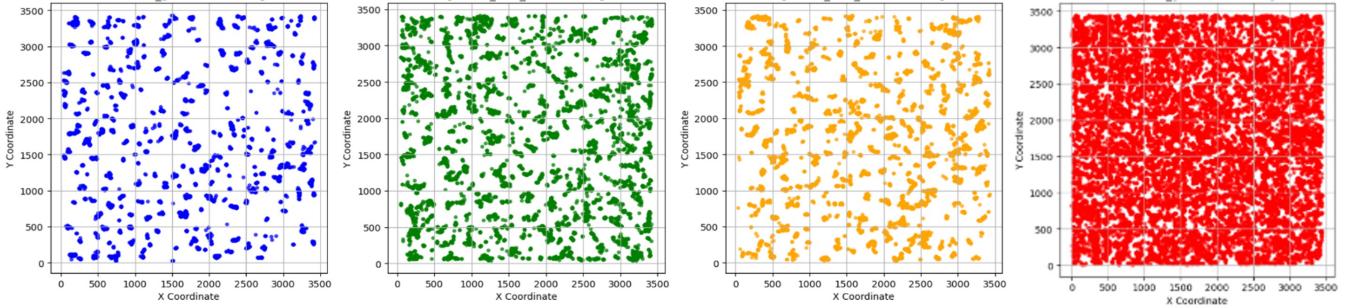


Fig. 15. Cluster centroids of each granule type

areas (blue), providing insight into solar surface dynamics. The density plots for individual granule types confirm their distinct spatial characteristics, with uniform granules appearing consistently distributed, while complex granules and granules with lanes show more clustered, localized densities.

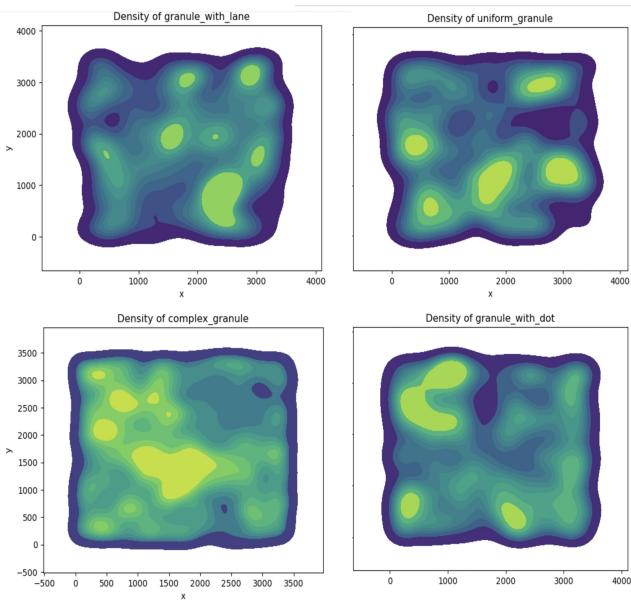


Fig. 16. Density of each granule type

These visualizations offer a comprehensive understanding of solar granule behavior, highlighting variations in density, spatial distribution, and clustering patterns that are crucial for studying solar surface activity and underlying convection processes.

G. Spatial indexes

We conducted two distinct queries to evaluate the performance and efficiency of GiST and SP-GiST spatial indexes in our database. The first query was a point query aimed at selecting granules from all timesteps that contained a specific point. To ensure comprehensive testing, this query was executed for 10 randomly selected points. The results of this

query, which provide insight into the retrieval performance for specific spatial locations, are illustrated in figure 17 .

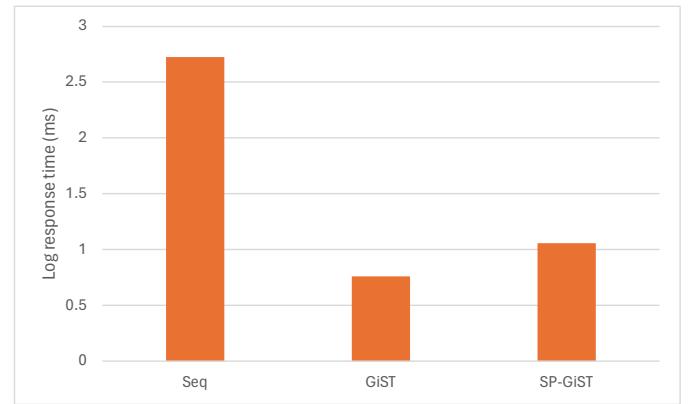


Fig. 17. Performance of GiST and SP-GiST indexes for a point query

As can be seen, GiST outperforms both SP-GiST and the sequential scan due to its efficient hierarchical structure. GiST quickly narrows down the search space by reducing the number of comparisons needed to locate relevant granules. In contrast, SP-GiST is better suited for disjoint spatial regions and performs less efficiently when precise point locations are required. The sequential scan, without any indexing, is the slowest option, as it examines each record linearly.

The second query was a range query designed to select granules from all timesteps that fall within a specified distance from the center of the coordinate system. In this query, the range was progressively increased in small increments until all granules within the dataset were included in the selection. This approach enabled us to examine how the spatial indexes performed with increasingly larger query ranges. The results of this range query, which highlight the scalability and effectiveness of the spatial indexing techniques, are presented in figure 18.

As shown, GiST and SP-GiST tend to perform similarly and significantly outperform the sequential scan. However, as the range increases, the selectivity of the query matches more records, and thus the sequential scan can bulk-select buckets

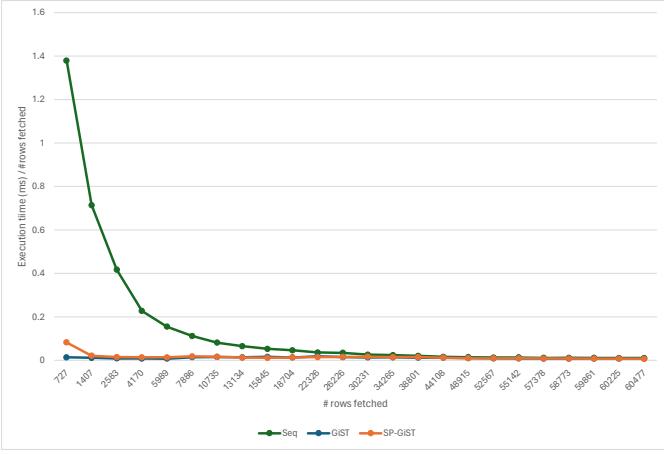


Fig. 18. Performance of GiST and SP-GiST indexes for a point query

from memory, highlighting the limitations of indexing when the range is large.

CONCLUSION AND FUTURE WORK

In this study, we transformed the granulation pattern of the solar photosphere into a spatial database using a 200-frame timeseries of machine-generated segmentation masks from DKIST telescope images. This infrastructure allows for spatial and spatiotemporal analysis of granules. We identified key challenges, such as DBSCAN clustering of granules and changes in granule size over time, and developed solutions for these issues. Additionally, we demonstrated the effectiveness of spatial indexes, particularly GiST and SP-GiST, for handling this data. Future work could involve integrating a customized variation of our granule tracking approach to assist with manual labeling of images, where only a limited number of frames need annotation, with the rest inferred automatically. This established platform could also enable more sophisticated spatial and spatiotemporal queries to be defined and addressed.

ACKNOWLEDGEMENTS

The code used for data extraction, preprocessing, clustering, and spatial indexing in this study is available at <https://github.com/rezmansouri/granules-sdb>.

REFERENCES

- [1] Å. Nordlund, R. F. Stein, and M. Asplund, “Solar surface convection,” *Living Reviews in Solar Physics*, vol. 6, no. 1, pp. 1–117, 2009.
- [2] A. Carlier, F. Chauveau, M. Hugon, and J. Rösch, “Cinématographie à Haute Résolution Spatiale de la Granulation Photosphérique,” *Academie des Sciences Paris Comptes Rendus Serie B Sciences Physiques*, vol. 266, pp. 199–201, Jan. 1968.
- [3] J. Hirzberger, J. Bonet, M. Vázquez, and A. Hanslmeier, “Time series of solar granulation images. iii. dynamics of exploding granules and related phenomena,” *The Astrophysical Journal*, vol. 527, pp. 405–414, 1999.
- [4] J.-M. Malherbe, T. Roudier, R. Stein, and Z. Frank, “Dynamics of trees of fragmenting granules in the quiet sun: Hinode/sot observations compared to numerical simulation,” *Solar Physics*, vol. 293, no. 1, Dec. 2017. [Online]. Available: <http://dx.doi.org/10.1007/s11207-017-1225-x>
- [5] S. L. Guglielmino, V. M. Pillet, B. R. Cobo, L. R. B. Rubio, J. C. del Toro Iniesta, S. K. Solanki, T. L. Riethmüller, and F. Zuccarello, “On the magnetic nature of an exploding granule as revealed by sunrise/imax,” *The Astrophysical Journal*, vol. 896, no. 1, p. 62, jun 2020. [Online]. Available: <https://dx.doi.org/10.3847/1538-4357/ab917b>
- [6] O. Steiner, M. Franz, N. B. González, C. Nutto, R. Rezaei, V. M. Pillet, J. A. B. Navarro, J. C. del Toro Iniesta, V. Domingo, S. K. Solanki, M. Knölker, W. Schmidt, P. Barthol, and A. Gandorfer, “Detection of vortex tubes in solar granulation from observations with sunrise,” *The Astrophysical Journal Letters*, vol. 723, no. 2, p. L180, oct 2010. [Online]. Available: <https://dx.doi.org/10.1088/2041-8205/723/2/L180>
- [7] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” 2015. [Online]. Available: <https://arxiv.org/abs/1505.04597>
- [8] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, “Unet++: A nested u-net architecture for medical image segmentation,” 2018. [Online]. Available: <https://arxiv.org/abs/1801.10165>
- [9] S. Castillo, A. Ramos, C. Fischer, and S. Berdyugina, “Towards the identification and classification of solar granulation structures using semantic segmentation,” *Frontiers in Astronomy and Space Sciences*, vol. 9, p. 896632, 06 2022.
- [10] R. Simoes, G. Queiroz, K. Ferreira, L. Vinhas, and G. Câmara, “Postgis-t: towards a spatiotemporal postgresql database extension,” 11 2016.
- [11] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise,” in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, ser. KDD’96. AAAI Press, 1996, p. 226–231.
- [12] P. Documentation, “Postgis indexes,” https://postgis.net/docs/manual-3.5/PostGIS_Index.html, 2024.