

# Visualization in Astrophysics: Developing New Methods, Discovering Our Universe, and Educating the Earth

Fangfei Lan<sup>1</sup> , Michael Young<sup>1</sup> , Lauren Anderson<sup>2</sup> , Anders Ynnerman<sup>3</sup>  Alexander Bock<sup>1,3</sup> , Michelle A. Borkin<sup>4</sup> ,  
Angus G. Forbes<sup>5</sup> , Juna A. Kollmeier<sup>2</sup> , Bei Wang<sup>1</sup> 

<sup>1</sup>Scientific Computing and Imaging Institute, University of Utah, USA

<sup>2</sup>Carnegie Institution for Science, USA

<sup>3</sup>Linköping University, Sweden

<sup>4</sup>Northeastern University, USA

<sup>5</sup>University of California, Santa Cruz, USA

## Abstract

We present a state-of-the-art report on visualization in astrophysics. We survey representative papers from both astrophysics and visualization and provide a taxonomy of existing approaches based on data analysis tasks. The approaches are classified based on five categories: data wrangling, data exploration, feature identification, object reconstruction, as well as education and outreach. Our unique contribution is to combine the diverse viewpoints from both astronomers and visualization experts to identify challenges and opportunities for visualization in astrophysics. The main goal is to provide a reference point to bring modern data analysis and visualization techniques to the rich datasets in astrophysics.

## 1. Introduction

Modern astronomers are recording an increasing amount of information for a larger number of astronomical objects and making more complex predictions about the nature of these objects and their evolution over cosmic time. Both successes are being driven by advances in experimental and computational infrastructure. As a result, the next generation of computations and surveys will put astronomers face to face with a “digital tsunami” of both simulated and observed data. These data present opportunities to make enormous strides in discovering more about our universe and state-of-the-art visualization methodologies.

This state-of-the-art report serves as a starting point to bridge the knowledge gap between the astronomy and visualization communities and catalyze research opportunities. Astronomy has a long and rich history as a visual science. Images of the cosmos have been used to build theories of physical phenomena for millennia. This history makes astronomy a natural area for fruitful collaborations between visualization and astronomy. A substantial fraction of previous work at this scientific intersection has therefore focused on image reconstruction – generating the most precise representation from a series of images of a patch of the sky – typically using optimizations and signal processing techniques. Advances in image reconstruction have enabled great breakthroughs in astronomy, including the recent imaging of a black hole [EAA\*19]. However, in this report, we focus on modern visualization techniques, which include 3D rendering, interaction, uncertainty visualization, and new display platforms. This report, authored by experts in both astron-

omy and visualization, will help visualization experts better understand the needs and opportunities of astronomical visualization, and provide a mechanism for astronomers to learn more about cutting-edge methods and research in visualization as applied to astronomy.

**Comparison with related surveys.** Several studies have focused on surveying visualization of astronomical data. Hassan *et al.* [HF11] surveyed scientific visualization in astronomy from 1990 to 2010. They studied visualization approaches for N-body particle simulation data and spectral data cubes – two areas they identified as the most active fields. They classified research papers in these areas based on how astronomical data are stored (i.e., as points, splats, isosurfaces, or volumes) and which visualization techniques are used. They also discussed visualization workflows and public outreach, and reviewed existing softwares for astronomical visualization.

Lipsa *et al.* [LLC\*12], on the other hand, took a broader view in surveying visualization for the physical sciences, which included astronomy and physics. For astronomy, the papers are classified based on the visualization challenges they tackle: multi-field visualization, feature detection, modeling and simulation, scalability, error/uncertainty visualization, and global/local visualization.

Hassan *et al.* excelled at classifying papers based on data types and considering how different types of data could be visualized. Lipsa *et al.* focused more on visualization techniques. A data-centered classification is useful for researchers to explore diverse ways to visualize their data, whereas a technique-centered classification can be useful for researchers who want to explore their

data using a particular visualization technique. Our survey aims to strike a balance between these two classification schemes and classifies the papers primarily based on data tasks and secondarily on visualization techniques, thus allowing researchers to explore how they can best visualize the data at hand based on the analysis tasks they have in mind. We also utilize tertiary categories in topical areas in astronomy for cross-references for the astronomy audience. Namely, we classify papers based on extragalactic, galactic, planetary, and solar astronomy. We further label each paper as dealing with simulated or observational astronomical data.

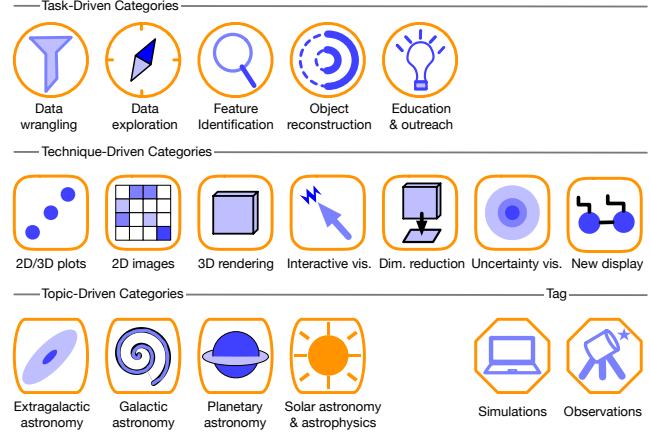
To the best of our knowledge, no comprehensive survey of visualization in astronomy has been conducted since 2012. Advances in both astronomical data and visualization in the past decade present a need for an updated state-of-the-art report. In 2011, Hassan *et al.* identified six grand challenges for scientific visualization in astronomy in the era of peta-scale data. Our survey discusses how the community has responded to these challenges in the past decade. The unique contribution of this survey is the cross-discipline discussion between visualization experts and astronomers via two workshops (a mini-workshop in April 2020 and an IEEE VIS workshop in October 2020), where researchers from both fields worked together in identifying progress, challenges, and opportunities in astronomical visualization. This survey aims to become a reference point for building connections and collaborations between two communities: data-rich, but technique-hungry, astronomers and data-hungry, but technique-rich, visualization experts. We further discuss datasets in astronomy in need of new approaches and methodologies, visualization techniques that have not been applied to astronomical datasets, and visualization techniques that can enhance the educational value of astronomical datasets.

In Sect. 2 we define our primary, secondary, and tertiary categories of approaches based on data analysis task, visualization technique, and topical area in astronomy, respectively. In Sect. 3, 4, 5, 6, and 7 we discuss and group papers based on the primary categories of data wrangling, data exploration, feature identification, object reconstruction, education and outreach, respectively. In Sect. 8 we identify challenges and opportunities for astronomy visualization. We provide a navigation tool of the surveyed papers in Sect. 9, and we summarize our conclusions in Sect. 10.

To make the survey results more accessible and actionable to the research community, all papers surveyed, including associated metadata, can be explored online with a visual literature browser (<https://tdavislab.github.io/astrovis-survis>) developed with the SurVis [BKW16] framework.

## 2. Literature Research Procedure and Classification

We reviewed representative papers over the past 10 years (between 2010 and 2020) in the fields of astronomy and visualization that contain strong visualization components for astronomical data. The annotation of each paper was guided primarily by a set of *data analysis tasks*; secondarily by a set of *visualization techniques*; and finally by a set of *topical areas in astronomy*. We view these three categories as being on equal footing and not necessarily hierarchical. Instead, they are considered as *orthogonal* dimensions and provide complementary viewpoints. We organize the literature accord-



**Figure 1:** A typology of task-driven (primary), technique-driven (secondary), and topic-driven (tertiary) categories used in this survey paper.

ing to these three categories to provide a means of navigation from task-driven, technique-driven, and topic-driven perspectives.

The literature surveyed spans venues in visualization such as *IEEE Transactions on Visualization and Computer Graphics*, *Computer Graphics Forum*, and *IEEE Computer Graphics and Applications*; and astronomy such as *Astrophysical Journal* and *Astrophysical Journal Letters*, *Monthly Notices of the Royal Astronomical Society*, *Astronomy and Computing*, *Astronomy (Dot Astronomy)*, *ADASS Conference Series*, *PASP (Publications of the Astronomical Society of the Pacific)*, *Research Notes of the AAS*. We also discuss data types that include simulation data and observation data, with the latter encompassing both image data and tabular data. Fig. 1 shows a typology of primary, secondary, and tertiary categories used in this survey.

### 2.1. Task-Driven Categories: Data Analysis Tasks

Our literature review allowed us to identify five primary categories of approaches based on data analysis tasks:

- **Data wrangling**, which transforms astronomy data into formats that are appropriate for general purpose visualization tools;
- **Data exploration**, where users explore a dataset in an unstructured way to discover patterns of interest;
- **Feature identification**, which visually guides the identification and extraction of features of interest;
- **Object reconstruction**, which provides an informative visual representation of an astronomical object;
- **Education and outreach**, where astronomical data or data products are made accessible to the general public.

In an on-going paradigm shift in scientific outreach, technological advances are enabling data-driven and interactive exploration of astronomical data in museums and science centers. Hence, we include “education and outreach” as a data analysis category. The word “feature” generally means a measurable piece of data that can be used for analysis, whereas the word “object” may be considered

as a “feature” with sharp and/or discontinuous contrast in a dimension of scientific interest. Whether a specific aspect of a dataset is considered an “object” or a “feature” depends on the scientific question at hand. We separate object reconstruction from feature identification to be compatible with the literature, but we envision a future in which these entities are recognized as a continuum.

## 2.2. Technique-Driven Categories: Visualization Techniques

Our secondary categories of approaches are based on visualization techniques employed for astronomical data:

-  **2D/3D plots** that encompass classic 2D/3D plots such as histograms, scatter plots, pie charts, pie, bar, and line plots;
-  **2D images** that utilize image processing techniques to generate images of astronomy data;
-  **3D rendering** that generates representations of 3D volumetric data of interest;
-  **Interactive visualization** that includes techniques such as linked views, detail on demand, visual filtering, and querying;
-  **Dimensionality reduction** that transforms data from a high-dimensional into a property-preserving low-dimensional space as part of the visualization pipeline;
-  **Uncertainty visualization** that improves our ability to reason about the data by communicating their certainties that arise due to randomness in data acquisition and processing;
-  **New display platforms** that communicate data via techniques such as data physicalization and virtual reality.

Although dimensionality reduction can be used as a purely data analysis strategy for noise reduction, clustering, or downstream analysis, it also serves as an integrated part of the visualization pipeline to facilitate data exploration and understanding. In this survey, we focus on the use of dimensionality reduction in the context of visualization. Dimensionality reduction and clustering may be both considered as data preprocessing techniques, but we choose to exclude clustering as a category as it is a generic class of techniques implicitly implemented within many toolboxes and does not typically represent a main innovation of the surveyed research.

We highlight the new display platforms as a category based on our experiences and workshops held among a growing “visualization in astrophysics” community. We believe there is a strong motivation for this research direction as the community as a whole is ready for the next stage of scientific discovery and science communications enabled by new displays.

We also acknowledge that there are additional ways to think about categories based on visualization techniques. For instance, scalable, multi-field, comparative, and time-dependent visualization are all categories mentioned in the 2012 survey of Lipsa *et al.* However, as technology has evolved over the past decade, certain visualization techniques (e.g., scalable and comparative visualization) have become commonplace and thus lack specificity. Time-dependent visualization (Sect. 8.5), in particular, the interplay between spatial and temporal dimensions, will be crucial as more time series astronomy data become available in the near future. In this survey, we choose specific visualization techniques that capture the state of the art and lead to informative categorization.

## 2.3. Topic-Driven Categories: Topical Areas in Astronomy

Our tertiary categories are based upon the list of topics from the National Science Foundation (NSF) Astronomy & Astrophysics directorate. These categories are used as a cross-reference for an astrophysics audience. We also investigated a curated list of research topics in astronomy and astrophysics provided by the American Astronomical Society (AAS) (<https://aas.org/meetings/aas237/abstracts>). We decided to work with the coarser classification from NSF since the AAS list is overly refined and specialized for the purposes of this survey. Our tertiary categories are:

-  **Extragalactic astronomy**
-  **Galactic astronomy**
-  **Planetary astronomy**
-  **Solar astronomy and astrophysics**

In addition, we have labeled each paper with two tags:

-  **Simulated astronomical data**
-  **Observational astronomical data**

For readers unfamiliar with certain terminology in astronomy or astrophysics, we recommend the astrophysics glossaries from the National Aeronautics and Space Administration (NASA) (<https://science.nasa.gov/glossary/>) or the LEVEL5 Knowledge Base on Extragalactic Astronomy and Cosmology (<https://ned.ipac.caltech.edu/level5/>). Meanwhile, we try our best to describe relevant terms the first time they are introduced in the survey. We would like to point out that even though certain terminology may appear to be rather straightforward, in some cases, definitions vary within the field, and thus some attention must be given to the precise work in question. For example, the term *halo* typically refers to overdensities in the dark matter but the exact boundary of a *halo* in a specific calculation may vary (e.g., [KPH13]).

**Overview.** One of the main contributions of this paper is the classification of existing works, which are summarized in Sect. 3 to Sect. 7. The methods of classification reflect the authors’ experience that comes from several meetings with experts in the astronomical visualization community. For each surveyed paper, we use our best judgment to infer its primary and secondary categories, although such classification may not be perfect; many papers span multiple categories. The best way to explore our classification is to use the table for each section (from Table 1 to Table 5) as a roadmap.

We acknowledge that many effective tools were actively used in astronomy research published prior to 2010. We emphasize that this paper is *not* a comprehensive catalog of all tools used in astronomy, nor does it include pre-2010 works. Rather, this paper surveys active areas of visualization research in astronomy as identified in publications in the last decade (2010–2021). We also note that whereas “astronomy” has previously meant the cataloging of the positions and motions of objects in the sky, and “astrophysics” the physical understanding of those objects, in this survey, we consider “astronomy” and “astrophysics” to be synonymous since few astronomers make the above distinction. In fact, by “visualization in astrophysics”, we consider the intersection of visualization with astronomy, astrophysics, and space exploration.

	Data wrangling	2D/3D plots	2D images	3D rendering	Interactive vis.	Dim. reduction	Uncertainty vis.	New display	Extragalac. astronomy	Galactic astronomy	Planetary astronomy	Solar astronomy	Simulation	Observation
[Ken13]					•				•		•		•	•
[Tay15]														
[Tay17b]					•	•							•	•
[Gár17]					•	•								
[Ken17]					•			•						
[Nai16]					•	•				•	•			
[NBC17]					•	•			•	•	•			
[BNL19]	•			•	•	•			•				•	
[WHA'11]	•				•	•			•				•	
[BG17]					•				•	•				•
[VOVMB16]														
[VSDR17]									•	•				•
[CPST20]	•								•	•				•
[ERG19]	•				•	•			•	•				•
[OC20]														
[VW12]								•	•	•			•	•
[VE21]								•	•	•				•

**Table 1:** Classifying papers under data wrangling based on secondary and tertiary categories. Top row, from left to right: (primary category) Data wrangling; (secondary categories) 2D/3D plots, 2D images, 3D rendering, interactive visualization, dimensionality reduction, uncertainty visualization, and new display platforms; (tertiary categories) extragalactic, galactic, planetary, and solar astronomy; (tags) simulated, and observational data.

### 3. Data Wrangling

Data wrangling is the process of transforming raw data into forms that more effectively support downstream analysis [KHP\*11]. This process is an important step for astronomy visualization because raw simulation or observational data require significant wrangling into a suitable form for visualization tasks. In this section, we categorize papers that present novel work in data wrangling for astronomy visualization. Many established tools are available for data wrangling across specific areas of astronomy, but a full survey of such tools is not within the scope of this survey. High-dimensional data abstractions such as data cubes are commonly used in astrophysical sciences and are often stored in the FITS format. Many of the papers placed in this category focus on transforming raw astrophysical data cubes into suitable data formats that can be ingested into open-source visualization tools, such as Blender and Houdini. Others introduce new formats that can be used to support various tools for data representation and data analysis. Authors of data wrangling papers have often made significant efforts to introduce astronomers to the visualization pipelines using these tools. We further classify these papers using our secondary categorization on visualization techniques (Sect. 2.2). Table 1 presents an overview of our categorization of data wrangling papers.

**Using Blender to visualize astrophysics data.** Blender [Ble02] is an open-source 3D graphics and visualization tool that supports a wide range of modeling, animation, and rendering functionality. A range of papers have discussed its usefulness for presenting astronomy data, and described pipelines for transforming raw data into scientific visualizations. Kent [Ken13] demonstrated how Blender can be used to visualize galaxy catalogs, astronomical data cubes, and particle simulations. Taylor [Tay15] introduced FRELLED, a Python-based FITS viewer for exploring 3D spectral line data using Blender that visualizes 3D volumetric data with arbitrary (non-Cartesian) coordinates [Tay17b] and is designed for real time and

interactive content. Using this viewer, astronomers are able to speed up visual cataloging by as much as 50×. Gárate [Gár17] described the process of importing simulation outputs from astrophysical hydrodynamic experiments into Blender using the voxel data format. In order to facilitate immersive data exploration, Kent [Ken17] presented a technique for creating 360° spherical panoramas using Blender and Google Spatial Media module. The method supports static spherical panoramas, single pass fly-throughs, and orbit flyovers on browsers or mobile operating systems.

AstroBlend [Nai12,Nai16] extends Blender, making it possible to import and display various types of astronomical data interactively, see Fig. 2. AstroBlend is an open-source Python library that utilizes yt – an open-source software for analyzing and visualizing volumetric data – for 3D data visualization (yt is discussed in Sect. 4). AstroBlend effectively bridges the gap between “exploratory” and “explanatory” visualization, as discussed by Goodman *et al.* [GBR18] and Ynnerman *et al.* [YLT18].

**Using Houdini to visualize astrophysics data.** In another example of adapting existing 3D graphics software, Naimen *et al.* [NBC17] explored how the 3D procedural animation software Houdini can be used for astronomy visualization, producing high-quality volume renderings for a variety of data types. They utilized yt to transform astronomical data into graphics data formats for Houdini, which bridges the astronomical and graphics community. Houdini is a compelling alternative to other rendering software (e.g., Maya and Blender) for astronomy because it produces high-quality volume renderings and supports a variety of data types.

Borkiewicz *et al.* [BNL19] presented a method for creating cinematic visualizations and time-evolving representations of astronomy data that are both educational and aesthetically pleasing. The paper also provided a detailed workflow of importing nested, multi-resolution adaptive mesh refinement data into Houdini.

**Using ParaView to visualize astrophysics data.** ParaView is an open-source, general-purpose, multi-platform analysis and visualization tool for scientific datasets. It supports scripting (with Python), web-based visualization, and in situ analysis (using Catalyst). Woodring *et al.* [WHA\*11] used ParaView to analyze and visualize large N-body cosmological simulations. *N-body cosmological simulations* are simulations of large-scale structures that contain particles that interact only via gravity, in contrast to including gas, which also requires hydrodynamics. ParaView provides particle readers (supporting “cosmo” and “GADGET” formats) and efficient halo finders, where a *halo* is a gravitationally bound structure on galactic scales. Together with existing visualization features, ParaView enables efficient and interactive visualization of large-scale cosmological simulations. Recent work from the IEEE VIS 2019 SciVis content [NPD19] used ParaView to visualize HACC (Hardware/Hybrid Accelerated Cosmology Code) cosmological simulations [HPF\*16].

**Data wrangling to support visualization.** Beyond the integration of visualization techniques into popular 3D software platforms, a range of projects have explored the transformation of astrophysical data into formats suitable for different forms of presentation, immersion, and analysis. Data wrangling is a perennial concern, and as new display formats are introduced or made more widely accessible, researchers investigate how best to target them. For example, predating our survey, Barnes *et al.* [BFBP06] introduced S2PLOT, a 3D plotting library for astronomy that supports dynamic geometry and time-varying datasets. S2PLOT has been used to construct models of planetary systems and create outputs for viewing on stereoscopic displays and in digital domes [FBO06]. Barnes and Flute [BF08] described a technique to embed interactive figures created with S2PLOT into Adobe PDF files to augment astronomy research papers, including 3D renderings of cosmological simulations and 3D models of astronomy instrumentation.

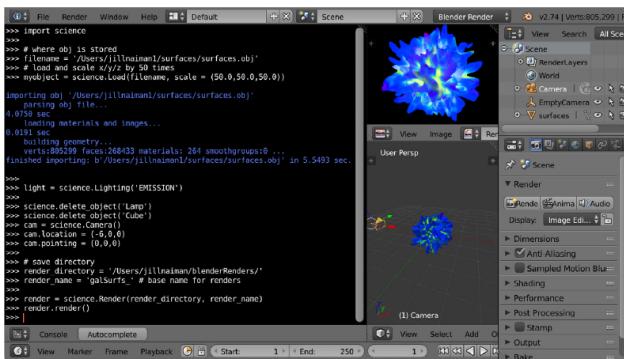
Some earlier approaches to data wrangling continue to be useful for more contemporary projects. The Montage Image Mosaic Engine [Arc05] enables users to stitch a “mosaic” together from sets of individual FITS images, and supports a range of image manipulation functionality, such as pixel sampling, image projec-

tion/rotation, background rectification, and animation. Montage can be used to create sky coverage maps and animations of data cubes, and its data wrangling capabilities have been integrated into other visualization tools. For example, mViewer, which can be scripted using Python, creates multi-color JPEG and PNG representations of FITS images and provides a wide range of functionality to support various types of image overlays, such as coordinate displays, labels, and observation footprints [BG17].

Vogt *et al.* [VOVMB16] introduced the X3D pathway for improving access to data visualization by promoting the use of interactive 3D astrophysics diagrams based on the X3D format, which can be shared online or incorporated into online publications. Vogt *et al.* [VSDR17] demonstrated the potential of this “pathway” by interactively visualizing integral field spectrographs observed in a young supernova remnant in the Small Magellanic Cloud. First, they created an interactive diagram of a reconstructed 3D map of the O-rich ejecta and exported it to the X3D file format. Second, they utilized (and extended) the visualization tools provided by X3D to make the diagram interactive, such as the ability to toggle views, “peel” intensity layers to focus on particular ranges of data, and modify clip planes to slice the 3D model at certain locations or angles.

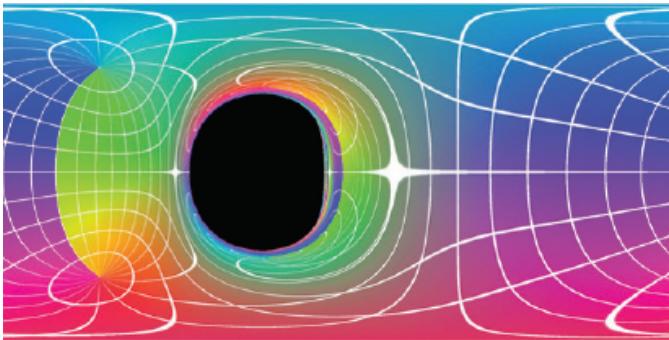
Although the most common format for distributing astronomy images is FITS [WG79], Comrie *et al.* [CPST20] suggested that the HDF5 format [FHK\*11] is better suited for hierarchical data and for facilitating efficient visualizations of large data cubes. They identified various common visualization tasks, including the rendering of 2D slices; generating single-pixel profiles, region profiles, and statistics; and interactive panning and zooming, and introduced a HDF5 hierarchical data schema to store precomputed data to facilitate these tasks. After integrating the HDF5 schema with the image viewer CARTA [OC20], they demonstrated that their schema was able to obtain up to  $10^3$  speed-ups for certain tasks. For example, precomputing and storing a dataset of histograms for each channel of a Stokes cube enables CARTA to display the histograms for an entire data cube with minimal delay. CARTA is part of CASA – the Common Astronomy Software Applications package – a primary data processing software for radio telescopes, including the Atacama Large Millimeter/submillimeter Array (ALMA) and the Karl G. Jansky Very Large Array (VLA). CASA [Jae08] supports data formats from ALMA and VLA, and is equipped with functionalities such as automatic flagging of bad data, data calibration, and image manipulation. It has also been used to simulate observations. It comes with a graphic user interfaces with viewer, plotter, logger, and table browser [Jae08]. CASA has some recent developments that enhance user experience [ERG19], including increased flexibility in Python and data visualization with CARTA.

Vogt and Wagner advocated for the use of stereoscopy visualization, or “stereo pairs”, to enhance the perception of depth in multi-dimensional astrophysics data [VW12]. Their technique involves sending distinct images to each eye, and supports both parallel and cross-eyed viewing techniques. They described a straightforward method to construct stereo pairs from data cubes using Python, and used various examples of both observational and theoretical data to demonstrate the potential of stereoscopy for visualizing astrophysical datasets.



**Figure 2:** A screenshot from a visualization session in AstroBlend, a Blender-based 3D rendering and analysis tool. Image reproduced from Naiman *et al.* [Nai16].

Verbraeck and Eisemann [VE21] presented a technique for interactively rendering black holes (see Fig. 3), illustrating how a black hole creates spacetime distortions in its environment due to gravitational lensing and redshift. The rendering algorithm first creates an adaptive grid that maps a uniform 360-view surrounding a virtual observer to the distorted view created by the black hole. This mapping is then used to optimize ray tracing through curved spacetime. The rendering solution also includes an interpolation technique that simulates the movement of the observer around the black hole, enabling interactive transitions between multiple sets of adaptive grids.



**Figure 3:** The projection of the distorted celestial sky caused by a Kerr black hole. Image reproduced from Verbraeck and Eisemann [VE21].

Data wrangling will continue to be an important component of astrophysics research as new sensors, telescopes, and other space instruments are built that generate datasets at higher resolutions and consisting of new data types. New data transformation methods or modifications of existing methods will be required to interoperate with existing visualization tools and to expand the accessibility of the data, making the data available in forms suitable for presentation, collaboration, interactive analysis, and public outreach.

#### 4. Data Exploration

In this section, we summarize research efforts that use visualization to focus on exploratory data analysis [Tuk77]. Broadly speaking, the defining attribute of data exploration papers is a focus on facilitating the unstructured investigation of a dataset in order to discover patterns of interest and formulate hypotheses. Our interpretation of data exploration follows Goodman’s perspective on studying high-dimensional data in astronomy, where “interactive exploratory data visualization can give far more insight than an approach where data processing and statistical analysis are followed, rather than accompanied, by visualization.” [Goo12]. We distinguish between “heterogeneous” and “hierarchical” data exploration to highlight the different methodologies employed, where heterogeneous refers to drawing together disparate datasets and hierarchical refers to a deep exploration of fundamentally similar datasets (perhaps at different resolutions). Table 2 presents an overview of our categorization of data exploration papers.

#### 4.1. Heterogeneous Data Exploration

A number of astrophysics visualization software frameworks and tools have emphasized the value of exploring multiple datasets simultaneously in order to generate new insight, often requiring (or facilitating) data transformation pre-processing steps.

yt [TSO\*10] is an open-source, flexible, and multi-code data analysis and visualization tool for astrophysics. Earlier versions of yt focused on making it possible to examine slices and projected regions within deeply nested adaptive mesh refinement simulations [BNO\*14]. Although still widely used for its data wrangling capabilities, yt now also includes a range of data exploration and feature identification functionalities, providing off-screen rendering, interactive plotting capabilities, and scripting interfaces. It efficiently processes large and diverse astrophysics data, creates 2D visualization with an adaptive projection process and volume rendering by a direct ray casting method. Its cross-code support enables analysis for heterogeneous data types, and facilitates cross-platform collaborations between different astrophysics communities. In order to reduce processing time, yt adopts parallelism and is able to run multiple independent processing units on a single dataset in parallel. Apart from being easily customizable, yt presents a number of pre-defined analysis modules for halo finding, halo analysis, merger tree creation, and time series analysis, among others, and a recent project makes it possible to use yt for interactive data exploration within Jupyter notebooks [MT20]. yt is also notable for its large, active community of users and developers.

Filtergraph [BSP\*13] is a web application that generates a range of 2D and 3D figures. It is designed to reduce the “activation energy” of the visualization process to flexibly and rapidly visualize large and complex astronomy datasets. It accepts numerous file formats without meta-data specifications, from text files to FITS images to Numpy files. The interface enables users to plot their data as high-dimensional scatter plots, histograms, and tables. Users can extensively explore the datasets and switch between different representations without cognitive interruption. Users can also customize the visualization through various interactive capabilities, such as panning, zooming, data querying, and filtering. Filtergraph also facilitates the sharing and collaboration of visualizations.

Luciani *et al.* [LCO\*14] introduced a web-based computing infrastructure that supports the visual integration and efficient mining of large-scale astronomy observations. The infrastructure overlays image data from three complementary sky surveys (SDSS, FIRST, and simulated LSST results) and provides real-time interactive capabilities to navigate the integrated datasets, analyze the spatial distribution of objects, and cross-correlate image fields. Additionally, Luciani *et al.* described *interactive trend images*, which are pixel-based, compact visual representations that help users identify trends and outliers among large collections of spatial objects.

ESASky [BGR\*16], developed by the ESA Center Science Data Center, is a web application designed for three use cases: the exploration of multi-wavelength skies, the search and retrieval of data for single or multiple targets, and the visualization of sky coverage for all ESA missions. The end result is a “Google Earth for space”, effectively combining the vast collection of data hosted by the ESA and providing an annotated map of the Universe that facilitates data querying and exploration across multiple data sources.

	Data exploration	2D/3D plots	2D images	3D rendering	Interactive vis.	Dim. reduction	Uncertainty vis.	New display	Extragalac. astronomy	Galactic astronomy	Planetary astronomy	Solar astronomy	Simulation	Observation
[TSO*10]	•		•	•	•				•	•			•	
[BSP*13]	•				•					•				•
[LCO*14]	•	•		•					•	•				•
[BGR*16]	•	•		•					•	•				•
[AFPR*17]	•			•					•					•
[PCHT17]	•		•	•					•					•
[ACS*17]														
[KHE*10]														
[FH07]														
[BAC*20]														
[BAB*18]	•		•	•					•	•	•	•	•	•
[CBE*21]														
[SJMS19]	•	•	•	•					•	•	•			•
[VBF*16]	•		•	•					•					•
[Mun17]	•		•		•				•					•
[Tay05]														
[Tay14]	•		•		•				•	•			•	•
[Tay17a]														
[SBD*17]	•			•					•					
[MSRMH09]														
[HWMB15]														
[AF15]	•		•	•					•					•
[HPU*15]														
[ZST11]	•				•						•			•
[BV18]	•		•	•	•				•	•	•		•	•
[YEII12]														
[YEII16]														
[FRG19]														
[HSS*19]														
[NNPD19]	•	•	•	•	•				•					
[SMG*19]														

**Table 2:** Classifying papers under data exploration based on secondary and tertiary categories. Top row, from left to right: (primary category) Data exploration; (secondary categories) 2D/3D plots, 2D images, 3D rendering, interactive visualization, dimensionality reduction, uncertainty visualization, and new display platforms; (tertiary categories) extragalactic, galactic, planetary, and solar astronomy; (tags) simulated, and observational data.

LSSGalPy [AFPR\*17] emphasizes the exploration of the large-scale structures surrounding galaxies and visualizes isolated galaxies, isolated pairs, and isolated triplets in relation to other galaxies within their large-scale structures. The paper describes one use case that investigates the effect of local and large-scale environments on nuclear activity and star formation, and another use case that visualizes galaxies with kinematically decoupled stellar and gaseous components, including an estimation of the tidal strength that affects each galaxy.

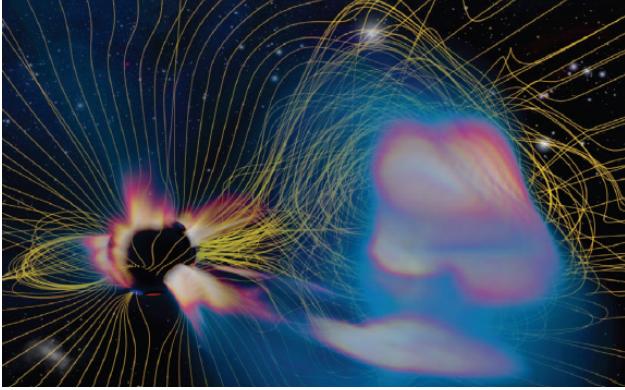
The Cosmicflows project aims to reconstruct and map the structure of the local universe, providing a series of catalogs that measure galaxy distances and velocities [TCD\*13]. Supporting this project, Pomarede *et al.* [PCHT17] provided four “cosmography” use cases for the SDvision visualization software, focusing on the creation of animations and interactive 2D and 3D visualizations of scalar and vector fields found in catalogs of galaxies, mapping cosmic flows, representing basins of attraction, and viewing the Cosmic V-web [PHCT17]. Pomarede *et al.* also explored the use of Sketchfab, a web-based interface that enables the uploading and sharing of 3D models that can be viewed in virtual reality.

The vast scales present in astronomical datasets can be difficult to render and present simultaneously. Klashed *et al.* [KHE\*10] introduced the “ScaleGraph” concept to deal with imprecision in

rendering in the Uniview software. Hansen *et al.* [FH07] utilized power-scaled coordinates to cover the distance ranges. More recently, Axelsson *et al.* [ACS\*17] presented a way to enable fast and accurate scaling, positioning, and navigation without a significant loss of precision, which they call the *dynamic scene graph*. At the core of this technique is the dynamic reassignment of the camera to focus on the object of interest, which then becomes the origin of the new coordinate system, ensuring the highest possible precision. Axelsson *et al.* applied this technique in the open-source software OpenSpace.

OpenSpace [BAC\*20] is a software system that enables the interactive exploration of a multitude of available astronomy datasets (Fig. 4). It is designed to be robust enough to support educational and outreach activities as well as adaptable enough to allow for the incorporation of new data or analysis tools to support scientific research. For the first task, Openspace has already demonstrated success in science communication at museums and in planetariums. For the second task, OpenSpace’s ability to interface with tools such as Glue [GBR18] or Aladin exemplifies a growing paradigm in astronomy visualization: the combination of multiple available tools to complete a task rather than building a bespoke system from the ground up. OpenSpace exhibits several novel features, including multi-resolution globe browsing [BAB\*18], which enables

dynamic loading of high-resolution planetary surface textures and physically based rendering of planetary atmospheres [CBE<sup>\*</sup>21].



**Figure 4:** OpenSpace: time-varying corona mass ejection simulation with 3D rendering and field lines. Image reproduced from Bock et al. [BAC<sup>\*</sup> 20].

Gaia Sky [SJMS19] is an open-source, 3D universe explorer that enables users to navigate the stars of our galaxy from the Gaia Catalog (Gaia data release 2). It also aids in the production of outreach material. The system embeds stars in a multi-scale octree structure, where, at different levels, stars with various absolute brightness values are present. The system contains a floating camera for space traversal, integrated visualization of relativistic effects, real-time star movement, and simulates the visual effects of gravitational waves. The main strength of Gaia Sky is its capability to provide real-time interactive exploration for hundreds of millions of stars. Its efficient handling of the data allows it to manage a large range of scales with sufficient numerical precision.

Vohl et al. [VBF<sup>\*</sup>16] presented Encube to accelerate the visual discovery and analysis process of large data cubes in medical imaging and astronomy (Fig. 5). Encube can be used on a single desktop as well as the CAVE2 immersive virtual reality display environment. In the CAVE2 environment, Encube enables users to control and interact with a visualization of over 100 data cubes across 80 screens. The design focuses on comparative visualization and related user interactions, such as swapping screens and requesting quantitative information from the selected screens. It uses a distributed model to seamlessly process and render visualization and analysis tasks on multiple data cubes simultaneously. Additionally, Encube serializes the workflow and stores the data in the JSON format, so that the discovery process can be reviewed and re-examined later. A desktop version of Encube supports many of the same functionalities as it does in the CAVE2 environment. Combined with the recording of the discovery process, researchers can continue with their workflow when they return to their desktops.

Recognizing that FITS images were inherently complex, and that existing FITS viewers were not built with an optimal user experience in mind, Muna [Mun17] introduced Nightlight, an “easy to use, general purpose, high-quality” viewer. Nightlight uses detail-on-demand to provide a high-level view of the file structure upon loading, and allows quick exploration of the data. Instead of reducing the dynamic range of astronomical data while visualizing FITS

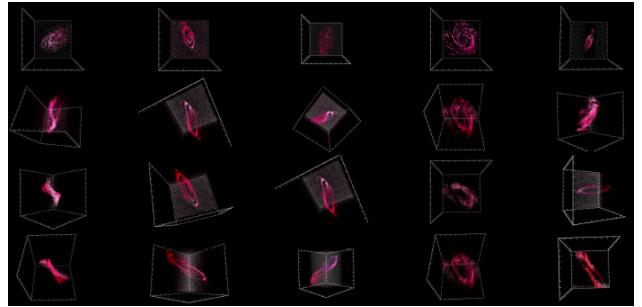
images, Nightlight leverages its approach on the fact that the input image is likely astronomical data. It provides two modes for the astronomers — hyperbolic sine function scaling for bright features (e.g. stars), and linear scaling for faint features (e.g., nebulae). For FITS tables, Nightlight provides two views. The first is a grid of “cards”, where each card represents the metadata of a single column in the table. The “cards” view is complemented by a second view in which the user can find the details of the full table.

Since its introduction, TOPCAT [Tay05] has been widely used to view, analyze, and edit tabular data in the astronomy community. In addition to the generic tasks such as sorting rows, computing statistics of columns, and cross-matching between tables, TOPCAT also provides astronomy specific functionalities including the access to Virtual Observatory data, handling of various coordinate systems, and joining tables based on sky positions [Tay17a]. Over the past decade, the developers of TOPCAT have continued to improve its capabilities. Taylor [Tay14] described a rewrite of the plotting library added to TOPCAT v4, which is designed to improve responsiveness and performance of the visualization of large datasets. One important new feature is the hybrid scatter plot/density map, see Fig. 6, that enables users to navigate interactively between the high- and low-density regions without changing plot types.

Taylor [Tay17a] described the exploratory visualization capabilities of TOPCAT, which include high-dimensional plotting, high-density plotting, subset selection, row highlighting, linked views, and responsive visual feedback. Apart from the GUI application, users can also access TOPCAT from a set of command-line tools.

## 4.2. Hierarchical Data Exploration

Scherzinger et al. [SBD<sup>\*</sup>17] proposed a unified visualization tool based on Voreen [MSRMH09] that supports the interactive exploration of multiple data layers contained within dark matter simulations. These simulations contain only dark matter particles, in contrast to also including gas and stars. Scherzinger’s visualization en-

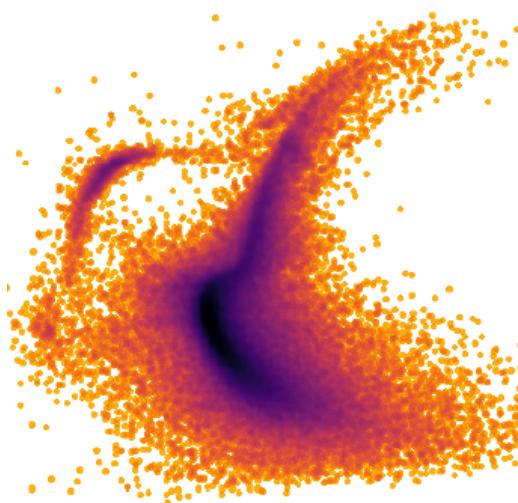


**Figure 5:** Comparative visualization of 20 galaxy morphologies with Encube [VBF<sup>\*</sup>16]. Image reproduced from “Large-scale comparative visualization of sets of multidimensional data”, written by Dany Vohl, David G. Barnes, Christopher J. Fluke, Govinda Poudel, Nellie Georgiou-Karistianis, Amr H. Hassan, Yuri Benovitski, Tsz Ho Wong, Owen L. Kaluza, Toan D. Nguyen, and C. Paul Bonnington, published in the PeerJ Computer Science journal. Link to article: <https://peerj.com/articles/cs-88/>.

ables users to view the global structure of the data through 2D and 3D volume rendering and particle rendering, and the time-varying properties of the data through a merger tree visualization. Local structures are explored further through local particles visualization and the integration with *Galacticus*, an open-source semi-analytic model that computes information about galaxy formation based on merger tree hierarchies of dark matter halos [Ben12]. An important aspect of their approach is scalable volume rendering, where the distribution of dark matter is visualized at interactive frame rates based on a pre-processing conversion. During such a conversion, attributes of large-scale particle data are distributed over a voxel grid, and maximum intensity projection in the 3D view is computed to highlight high-density regions of the data for volume rendering.

Other tools also focus on exploring the evolution of galaxy halos within simulation datasets. Hazarika *et al.* [HWMB15] presented a series of visualizations to provide insight into halos, including a 3D volume rendering of simulation data and a particle rendering that identifies halo sub-structures. Almryde and Forbes [AF15] introduced an interactive web application to create animated “traces” of dark matter halos as they move in relation to each other over time, and Hanula *et al.* [HPU\*15] presented the Cavern Halos project that enables the exploration of halos in virtual reality using the CAVE2 immersive collaboration space (this project was later extended and renamed DarkSky Halos [HPAM19]). See also the discussion of work by Preston *et al.* [PGX\*16] in Sect. 5.

In order to better investigate the nature of solar wind ion data (SWID), which is typically visualized using 1D and 2D methods, Zhang *et al.* [ZST11] developed a 3D visualization method for SWID based on the Selenocentric Solar Ecliptic coordinate system, and integrated this method into an interactive tool called vtSWIDs.



**Figure 6:** TOPCAT: Hybrid scatter plot/density map [Tay17a]. Image reproduced from “TOPCAT: Desktop Exploration of Tabular Data for Astronomy and Beyond”, written by Mark Taylor, and published in the Informatics journal. Link to article: <https://doi.org/10.3390/informatics4030018>.

vtSWIDs enables researchers to browse through numerous records and provides statistical analysis capabilities.

Breddels *et al.* [BV18] introduced Vaex, a Python library that handles large tabular datasets such as the Gaia catalogue. Many packages in Vaex are developed with specific visualization challenges in mind, and they overcome the scalability issues with methods such as efficient binning of the data, lazy expressions, and just-in-time compilation. For example, vaex-core provides visualization using the matplotlib library, with 1D histograms and 2D density plots; vaex-jupyter embeds the visualization tools in a web browser, which offers more user interactions such as zooming, panning, and on-plot selection. It also enables 3D volume and iso-surface rendering using ipyvolume and connecting to a remote server using WebGL. A standalone interface is provided by the vaex-ui package, which supports interactive visualization and analysis. The vaex-astro package is specifically designed for astronomical data, supporting the FITS format and the most common coordinate transformations needed for analysis in astronomical data.

To enhance the study of astronomical particle data, the work by Yu *et al.* [YEII12] was motivated by the need for an enhanced spatial selection mechanism using direct-touch input for particle data such as numerical simulations of the gravitational processes of stars or galaxies. They introduced two new techniques, *TeddySelection* and *CloudLasso*, to support efficient, interactive spatial selection in large particle 3D datasets. Their selection techniques automatically identify bounding selection surfaces surrounding the selected particles based on the density. They applied their techniques to particle datasets from a galaxy collision simulation (<http://www.galaxydynamics.org>) and an N-body mass simulation from the Aquarius Project [SWV\*08], thus reducing the need for complex Boolean operations that are part of traditional multi-step selection processes. In a follow-up work [YEII16], Yu *et al.* further enhanced their 3D selection techniques to aid the exploratory analysis of astronomical data. They proposed a collection of context-aware selection techniques (*CAST*) that improve the usability and speed of spatial selection, and applied their methods to a cosmological N-Body simulation and Millennium-II dataset [SWJ\*05].

The 2019 SciVis contest proposed a visual analysis challenge to explore the structure formation in the cosmic evolution. The dataset was from a CRK-HACC (HACC: Hardware/Hybrid Accelerated Cosmology Code) cosmological simulation containing dark matter plus baryon particles in a cubic box, where the particles contain multiple fields such as position, velocity, and temperature. The simulations were used to study the impact that the feedback from AGN (Active Galactic Nuclei) has on their surrounding matter distribution. The entries from the contest (e.g., [FRG19, HSS\*19, NNPD19, SMG\*19]) represented a diverse collection of visualizations, made possible by these new forms of simulation datasets.

## 5. Feature Identification

Research efforts in this category visually guide the identification and extraction of features of interest. The term “feature” is broad and can be used in a number of different astrophysical contexts. The detection of features in an astrophysical datastream is of critical importance since many interesting phenomena are diffuse or

	Feature Identif.	2D/3D plots	2D images	3D rendering	Interactive vis.	Dim. reduction	Uncertainty vis.	New display	Extragalac. astronomy	Galactic astronomy	Planetary astronomy	Solar astronomy	Simulation	Observation
[KHA12]	•					•								
[RAW*20]	•				•			•						•
[SXL*14]		•	•	•	•				•					•
[PGX*16]	•	•	•	•					•					•
[PSD17]	•	•	•			•			•					•
[BAO*19]	•	•	•	•					•					•
[XNW*16]	•								•					
[SUB*20]														•
[LvdWM17]	•								•					•
[Sou11]	•	•	•						•				•	•
[SPK11]														
[SPN*16]		•	•						•					•
[TSH21]	•			•	•					•				•
[RSM*19]		•							•					•
[CKA12]		•	•	•	•				•					•
[CCM*20]	•	•	•						•					•
[AC19]		•	•						•					•
[KHW*19]						•			•					•
[NZE*19]		•							•					•

**Table 3:** Classifying papers under feature identification based on secondary and tertiary categories. Top row, from left to right: (primary category) feature identification; (secondary categories) 2D/3D plots, 2D images, 3D rendering, interactive visualization, dimensionality reduction, uncertainty visualization, and new display platforms; (tertiary categories) extragalactic, galactic, planetary, and solar astronomy; (tags) simulated, and observational data.

observed with a low signal-to-noise ratio. For example, physical phenomena may be subtle to detect (or may be detected for the first time), and distinguishing between what is signal and what is noise is critical. Teasing out a tiny signal is so common in astronomy that feature detection is a generically important element of astrophysical progress. Furthermore, astrophysicists are often looking for diffuse physical contrasts in multiple dimensions (e.g. spatial, chemical, magnetic, density). For these phenomena, methods that establish robust criteria in multiple dimensions for identification and subsequent analysis are crucial. The majority of these papers focus on dark matter simulations and the cosmic web, in particular voids, filaments, and dark matter halos, as summarized in Table 3. The *cosmic web* refers to the large-scale structure of matter, distributed in filaments, the gravitationally collapsed structures that tend to connect galaxy halos, and voids, the low-density areas of the Universe.

**Visualizing dark matter simulations and cosmic web.** Papers in this subsection employ various visualization techniques to visualize dark matter simulations and cosmic web, including GPU-assisted rendering with a tailored tessellation mesh [KHA12], tomographic map [RAW\*20], and interactive visual exploration of cosmic objects [PGX\*16, SXL\*14].

Dark matter generates small-scale density fluctuations and plays a key role in the formation of structures in the Universe. Kaehler *et al.* [KHA12] visualized N-body particle dark matter simulation data using GPU-assisted rendering approaches. Their method leverages the phase-space information of an ensemble of dark matter tracer particles to build a tetrahedral decomposition of the computational domain that allows a physically accurate estimation of the mass density between the particles [KHA12]. During the simulation, vertices of a tessellation mesh are defined by the dark matter

particles in an N-body simulation, whereas tetrahedral cells contain equal amounts of mass. The connectivity within the mesh is generated once and is kept constant over the simulation as the cells warp and overlap. The density of a given location in the simulation is obtained by considering the density contribution from overlapping cells in the region of interest. Their new approaches are shown to be effective in revealing the structure of the cosmic web, in particular, voids, filaments, and dark matter halos.

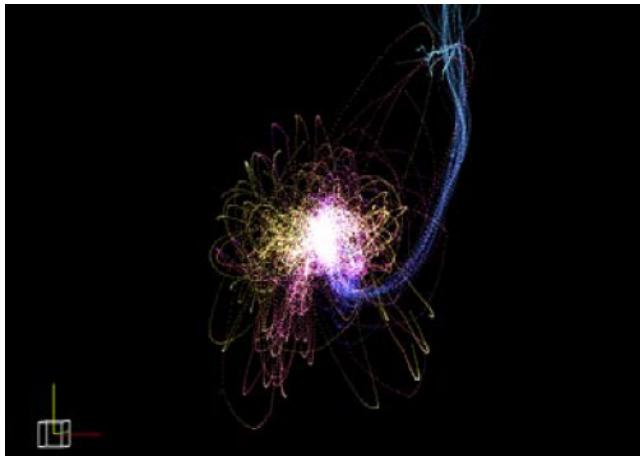
The Ly $\alpha$  forest, which is a series of individual over-densities of neutral hydrogen within the *intergalactic medium* (IGM, the space between galaxies), provides a 1D measurement of information in the IGM, which is largely correlated with the distribution of matter in the Universe. Ravoux *et al.* [RAW\*20] used a tomographic reconstruction algorithm called the Wiener filtering to create a 3D tomographic map with the *eBoss Strip p82 Ly $\alpha$  forest* datasets. The map is used as a representation of the associated matter fluctuation to identify over- and under-densities in the cosmic web. Extended over-densities can be detected with the tomographic map by searching for the large deficit in the Ly $\alpha$  forest flux contrast. The authors adopt a simple-spherical algorithm to identify large voids. In order to further investigate the regions of interest, the paper provides 3D representations of the tomographic map over the entire strip. Users can interactively explore the map through rotating, panning, and zooming.

Gravity causes dark matter particles to collapse into larger structures over time. The individual groups of particles formed during this process are called halos, one of the most common elements in the dark matter simulation [PGX\*16]. Their evolution process and behaviors are often the focus of astronomical discoveries. Two recent tools facilitate the visual exploration of halos. Shan *et al.* [SXL\*14] built an interactive visual analysis system

that focuses on exploring the evolutionary histories of halos. The interface allows the user to manually select regions of interest in 3D space. It then uses the marching cubes algorithm to perform iso-surface extraction and cluster separation based on the region's density distribution. To prevent overlaps in the 3D space, the system employs multi-dimensional scaling (MDS) to project the halos into 2D space. Multiple linked views are generated to support the exploration through time. In addition to a merger tree view that is commonly used to visualize evolution of objects over time, Shan *et al.* proposed a unique particle trace path graph (see Fig. 7), which encodes the evolution history of selected particles.

Preston *et al.* [PGX<sup>\*</sup>16], on the other hand, aimed to increase the efficiency and interactions in studying the evolution of halos, described by merger trees. Their integrated visualization system consists of a *merger tree view*, a *3D rendering view*, and a *quantitative analysis view*. Their *merger tree view* is an enhancement from [SXL<sup>\*</sup>14] with more interactive capabilities. The system allows users to select specific halos through the merger tree and organize the tree based on other physical variables such as velocity and mass. The *3D rendering view* displays the particles' physical behaviors over a number of time steps, providing additional contextual information for the merger tree. A remote parallel renderer is employed to improve the scalability of the rendering process. Finally, the *quantitative analysis view* extends the other two views by providing quantitative information of selected particles that reveals additional insights into the behavior of the halo. For instance, a chronological plot visualizes the anomalous events automatically detected in the history of a halo. An important feature of the system is that it enables simultaneous exploration of heterogeneous cosmology data; see Sect. 4 for further discussions.

The IllustrisTNG project (<https://www.tng-project.org/>) contains collections of large, cosmological magnetohydrodynamical simulations of galaxy formation. It is designed to "illuminate the physical processes that drive galaxy formation". The tool provides a number of volume rendering capabilities to visually demon-



**Figure 7:** An example of a particle trace path. Image reproduced from Shan *et al.* [SXL<sup>\*</sup>14].

strate the multi-scale, multi-physics nature of the simulations, as well as to perform qualitative inspections [PSD17].

Moving from clusters of galaxies to the spaces between them, the IGM is composed of gas complexes in the spaces between galaxies. Although it has research values on its own, investigating IGM along with quasar sightlines helps put IGM in context. A *quasar* is a supermassive blackhole at the center of a galaxy that is accreting gas at a high rate and is therefore very bright. It enables scientists to associate certain absorption features with galactic environment, such as the *circumgalactic medium* (CGM), which is the gaseous envelope surrounding a galaxy. IGM-Vis [BAO<sup>\*</sup>19] is a visualization software specifically designed to investigate IGM/CGM data. It supports a number of identification, analysis, and presentation tasks with four linked views. The *Universe panel* provides a 3D interactive plot of galaxies in circles and quasar sightlines in cylindrical "skewers". The user can select a galaxy of interest to further examine it in the *galaxy panel*, which contains a list of attributes and corresponding data from SDSS. Additionally, quasar sightlines can be explored in the *spectrum panel* where multiple spectral plots can be displayed and stored. The final *equivalent width plot panel* facilitates dynamic correlation analysis and helps users discover absorption patterns in the regions of interest. The four views complement each other to streamline the discovery processes, including the identification of foreground and sightline features, the measure of absorption properties, and the detection of absorption patterns.

*Blazars* – similar to quasars, an active galactic nuclei with relativistic jets ejecting toward the Earth – are one of the most attractive objects for astronomers to observe. The TimeTubes visualization [XNW<sup>\*</sup>16] transforms time-varying blazar data and polarization parameters into a series of ellipses arranged along a time line, forming a volumetric tube in 3D space. The most recent iteration of the project, TimeTubesX [SUB<sup>\*</sup>20], includes feature identification techniques to detect recurring time variation patterns in blazar datasets. It includes an automatic feature extraction functionality to identify time intervals that correspond to well-known blazar behaviors, as well as dynamic visual query-by-example and query-by-sketch functionality. Such a functionality enables users to search long-term observations that are similar to a selected time interval of interest, or match a sketch of temporal pattern. The technique aims to enhance the reliability of blazar observations, and to identify flares, rotations, and other recurring blazar patterns in order to validate hypotheses about observable, photometric, and polarimetric behaviors.

To study the agreements and disparities of feature identification methods created for classifying the cosmic web, Libeskind *et al.* [LvdWM17] collected 12 representative methods and applied them to the same GADGET-2 dark matter simulation. They classified the dark matter density field of the cosmic web into knots, filaments, walls, and voids. They used comparative visualization accompanied with a variety of 2D plots to provide intuitive representations of the different structures identified by these methods. We introduce one of the topology-based methods with a strong visualization component in the next subsection.

**Topology-based feature extraction.** There are several examples of using topological techniques to extract cosmological features from simulations, in particular, galaxy filaments, voids, and halos. Topo-

logical methods have also been applied to observational data cubes. We believe that the integration of topological techniques in astronomical feature extraction and visualization will be a growing area of interest (see Sect. 8).

Sousbie [Sou11] presented DisPerSE, a topology-based formalism that is designed to analyze the cosmic web and its filamentary structure. It leverages discrete Morse theory and computes a Morse-Smale complex (MSC) on a density field. The MSC is then simplified using the theory of persistent homology by canceling the topological features with low persistence values (i.e., those that are likely generated by noise). The relationship between the topological and geometrical features is easily detectable in the MSC, where the ascending 3-manifolds correspond to the voids, ascending 2-manifolds to the walls, and ascending 1-manifolds to the filaments. The technique is scale-free, parameter-free, and robust to noise. Sousbie *et al.* then demonstrated the effectiveness of DisPerSE at tracing cosmological features in 2D and 3D datasets [SPK11].

Following a similar path, Shivashankar *et al.* [SPN\*16] proposed Felix, another topology-based framework that identifies cosmological features (see Fig. 8). Felix focuses on extracting the filamentary structures and incorporates a visual exploration component. It also computes a MSC over a density field and simplifies it by iteratively canceling pairs of simplices, which generates a hierarchy of MSCs. Realizing that it is nearly impossible to find a version of the MSC within the hierarchy that best separates noise and features for cosmology datasets, Felix allows users to query for specific density ranges across all generated MSCs. This process increases user engagement in the parameter selection process and helps preserve filament structures within void-like or cluster-like regions. Felix also utilizes 3D volume rendering to interactively guide the selection of parameters for the query and visualizes the extracted filaments along with the density field. Interactive visual exploration of these intricate features remains a challenging and largely unexplored problem.

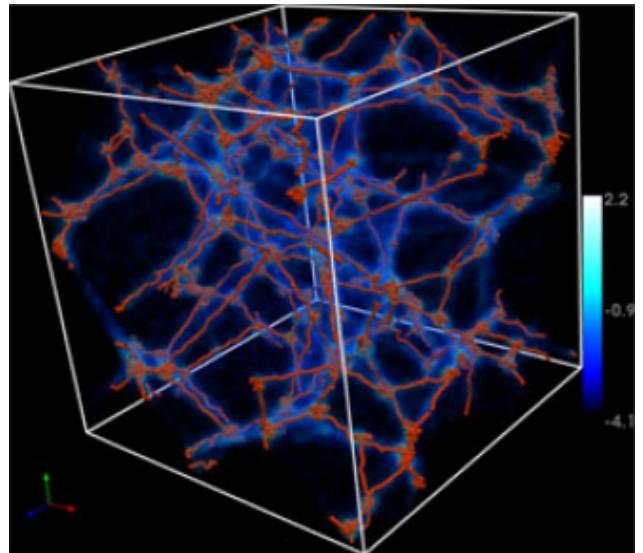
Recently, a new method has been proposed by Tricoche *et al.* [TSH21] to extract the topology of the Poincaré map in the *circular restricted three-body problem* (CR3BP). They created an interactive visualization of the topological skeleton to support spacecraft trajectory designers in their search for energy-efficient paths through the interconnected web of periodic orbits between celestial bodies. The new method extends the existing approach by Schleier *et al.* [SHTG14], and significantly improves the results of fixed point extraction and separatrices construction. In order to reduce the high computational cost, Tricoche *et al.* pre-screened for impractical spaceflight structures, and leveraged previous knowledge on the accuracy limitations of sensors and engines to impose restrictions on certain parameters. These adjustments reduce the computational workload of the method and enable the interactive visualization of the topology. The visualization displays the fixed points identified by the system and each individual selected orbit as a closed curve. The visualization also enables a manifold arc selection mechanism to help the trajectory designer to determine the precise path a spacecraft would need to follow from any arbitrary location.

From an observational perspective, current radio and millimeter telescopes, particularly ALMA, are producing data cubes with

significantly increased sensitivity, resolution, and spectral bandwidth. However, these advances often lead to the detection of structure with increased spatial and spectral complexity. Rosen *et al.* [RSM\*19] performed a feasibility study for applying topological technique – in particular, contour trees – to extract and simplify the complex signals from noisy ALMA data cubes. They demonstrated the topological de-noising capabilities on a NGC 404 data cube (also known as Mirach’s Ghost) and a CMZ (Central Molecular Zone) data cube. Using topological techniques, Rosen *et al.* sought to improve upon existing analysis and visualization workflows of ALMA data cubes, in terms of accuracy and speed in feature extraction.

**Feature extraction from astronomy data cubes.** In addition to the work by Rosen *et al.* [RSM\*19], other visualizations of integral field spectrometer (IFS) data cubes have been proposed. Campbell *et al.* [CKA12] presented a 3D interactive visualization tool specifically designed to render IFS data cubes. A typical display tool reduces a 3D IFS datacube to 2D images of either the spatial or the wavelength dimension. Campbell *et al.* proposed to use volume rendering instead to highlight features and characteristics of astronomical objects that are difficult to detect in lower dimension projections. The tool, known as OsrsVol, allows users to easily manipulate the visualized data cube by interactions such as zooming, rotating, and aspect ratio adjustment.

Ciulo *et al.* [CCM\*20] used OsrsVol to identify four objects orbiting the supermassive black hole at the center of our galaxy Sagittarius A\*. Two unusual objects have been recently discovered around Sagittarius A\*, referred to as the G sources, and their possible tidal interactions with the black hole have generated considerable attention. Ciulo *et al.* selected 24 relevant data cubes and processed them through the OSIRIS pipelines. They analyzed the



**Figure 8:** *Felix*: Extracting filamentary structures (orange) from a Voronoi evolution time-series dataset. Image reproduced from Shivashankar *et al.* [SPN\*16].

data cubes with OsrsVol, as well as several conventional 1D/2D visualization tools. OsrsVol helps to disentangle the various dimensions of data cubes and allows more flexible explorations among crowded regions. Using OsrsVol, Ciulo *et al.* also characterized the best-fit orbits of the four new objects, and determined that they exhibited many traits in common with the previously discovered G sources.

**Feature identification with deep learning.** We end this section by giving a couple of examples of using neural network models as feature extractors for unsupervised clustering of galaxies. These works demonstrate the potential of using deep learning in feature identification tasks, for which both astronomers and visualization experts are cautiously excited.

Aragon-Calvo was the first to apply a deep convolutional neural network to the task of semantic segmentation of the cosmic web [AC19]. He proposed a network with a U-net architecture and trained the model using a state-of-the-art manually guided segmentation method. Two types of training datasets were generated using the standard Voronoid model and an N-body simulation. Their method provides exciting results as it efficiently identifies filaments and walls with high accuracy for well-structured data such as the Voronoid model. For more complex datasets such as the N-body simulation, the U-net achieves higher quality segmentation than the state-of-the-art methods.

Khan *et al.* [KHW<sup>\*</sup>19] constructed galaxy catalogs using transfer learning. They employed a neural-network-based image classifier *Xception*, pre-trained on ImageNet data, to classify galaxies that overlap both Sloan Digital Sky Survey (SDSS) and Dark Energy Survey (DES) surveys, achieving state-of-the-art accuracy of 99.6%. Khan *et al.* then used their neural network classifier to label and characterize over 10,000 unlabelled DES galaxies, which do not overlap previous surveys. They further extracted abstract features from one of the last layers of their neural network and clustered them using t-SNE, a dimensionality reduction technique. Their clustering results revealed two distinct galaxy classes among the unlabelled DES images based on their morphology. The analysis of Khan *et al.* provides a path forward in creating large-scale DES galaxy catalog by using these newly labelled DES galaxies as data for recursive training.

*Galaxy clusters* are gravitationally bound systems that contain hundreds or thousands of galaxies in dark matter halos [NZE<sup>\*</sup>19], with typical masses ranging from  $10^{14}$  to  $10^{15}$  solar masses. Ntampaka *et al.* applied deep learning to estimate galaxy cluster masses from *Chandra* mock – simulated, low-resolution, single-color X-ray images [NZE<sup>\*</sup>19]. They used a relatively simple convolutional neural network (CNN) with only three convolutional and pooling layers followed by three fully connected layers. Despite the simple framework, the resulting estimates exhibit only small biases compared to the true masses. The main innovation of the paper is the visual interpretation of the CNN, using an approach inspired by Google’s DeepDream, which uses gradient ascent to produce images that maximally activate a given neuron in a network. Ntampaka *et al.* used gradient ascent to discover which changes in the input cause the model to predict increased masses. They found that the trained model is more sensitive to photons in the outskirts of the clusters, and not in the inner regions; and their observations aligned

with other statistical analyses performed on galaxy clusters. Their work illustrates the utility of interpreting machine learning “black boxes” with visualization since it provides physical reasoning to predicted features.

## 6. Object Reconstruction

Research works in this category provide informative visual representation of astronomical objects; see Table 4 for their fine-grained classifications under secondary and tertiary categories, where there is a strong focus on observational data. Object reconstruction utilizes and is also constrained by imagery and other observational data obtainable via our vantage point – the Earth and the solar system. The works surveyed here cover 3D object reconstructions using 2D images [SKW<sup>\*</sup>11, WAG<sup>\*</sup>12, WLM13, HA20], distances of young stellar objects [GAM<sup>\*</sup>18], spectroscopic data [VD11], and extrapolation from sparse datasets such as SDSS [EBPF21], where visualization helps produce plausible reconstructions that provide structural insights for analysis and modeling. Important challenges include scalable computation, trade-off between automatic reconstruction and expert knowledge, and in particular, physically accurate structural inference with limited observations.

As mentioned previously, we recognize that “objects” are, in fact, “features” with sharp and/or discontinuous contrast in a dimension of scientific interest. Whether a specific aspect of a dataset is considered an “object” or a “feature” depends on the scientific question posed. We separate object reconstruction from feature identification to be compatible with the literature, but we envision a future where these entities are recognized as a continuum. An example of such a continuum is Polyphorm [EBPF21], where the filament reconstruction and interactive visualization are intertwined via a fitting session, where structural or visual parameters are adjusted interactively to produce satisfactory reconstruction results.

Object reconstruction employs both images and other observational data, and thus is closely related to image reconstruction in astronomy. As discussed in Sect. 1, we do not consider state-of-the-art image reconstruction methods in astronomy based on optimizations or signal processing techniques, but rather, we will focus on reconstruction with modern visualization techniques, such as 3D object reconstruction, 3D rendering, and interactive visualization. There is existing literature on the “historic account” of astronomical image reconstruction [Dai85, PGY05], recent surveys about this field [TA16], and machine learning approaches [Fla17].

**3D object reconstruction from 2D images.** Steffen *et al.* [SKW<sup>\*</sup>11] presented Shape, one of the first publicly available tools using interactive graphics to model astronomical objects. Shape allows astrophysicists to interactively define 3D structural elements using their prior knowledge about the object, such as spatial emissivity and velocity field. Shape provides a unified modeling and visualization flow, where physical knowledge from the user is used to construct and iteratively refine the model, and model parameters are automatically optimized to minimize the difference between the model and the observational data. The interactive feedback loop helps introduce expert knowledge into the object reconstruction pipeline and has proven to be incredibly useful for many applications, such as rendering hydrodynamical

	Object reconstruct.	2D/3D plots	2D images	3D rendering	Interactive vis.	Dim. reduction	Uncertainty vis.	New display	Extragalac. astronomy	Galactic astronomy	Planetary astronomy	Solar astronomy	Simulation	Observation
[SKW*11]	•		•	•	•				•			•	•	
[WAG*12]		•		•	•					•				•
[WLM13]		•	•	•						•				•
[HA20]	•	•	•							•				•
[GAM*18]			•				•			•				•
[SSM*19]	•									•				•
[VD11]	•	•							•				•	
[EBPF21]			•	•	•			•		•		•	•	•
[BPM*15]	•		•	•	•		•				•	•	•	•
[OWN*20]	•		•	•	•					•				•

**Table 4:** Classifying papers under object reconstruction based on secondary and tertiary categories. Top row, from left to right: (primary category) Object reconstruction; (secondary categories) 2D/3D plots, 2D images, 3D rendering, interactive visualization, dimensionality reduction, uncertainty visualization, and new display platforms; (tertiary categories) extragalactic, galactic, planetary, and solar astronomy; (tags) simulated, and observational data.

simulations, reconstructing Saturn Nebula, modeling the structure and expansion of nova RS Ophiuchi [SKW\*11]. Shape also comes with educational potential in digital planetariums.

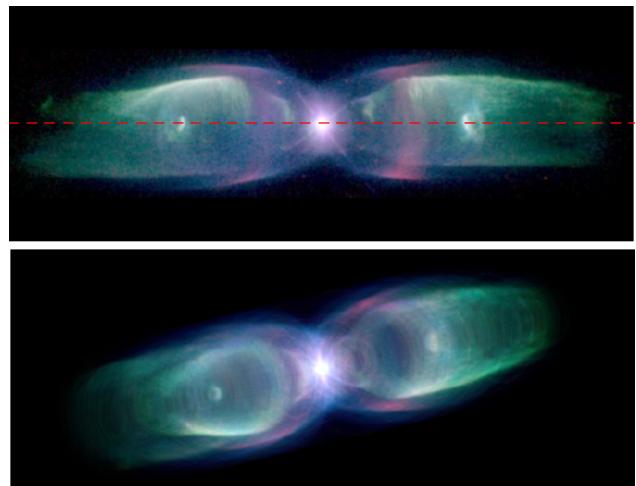
Wenger *et al.* [WAG\*12] developed an automatic 3D visualization of astronomical nebulae from a single image using a tomographic approach. Their 3D reconstruction exploits the fact that many astronomical nebulae, interstellar clouds of gas and dust, exhibit approximate spherical or axial symmetry [MKDH04]. This symmetry allows for object reconstruction by replicating multiple virtual viewpoints based on the view from Earth. This assemblage of different views results in a tomographic reconstruction problem, which can be solved with an iterative compressed sensing algorithm. The reconstruction algorithm relies on a constrained optimization and computes a volumetric model of the nebula for interactive volume rendering. Wenger *et al.* demonstrated that their method preserves a much higher amount of detail and visual variety than previous approaches. However, they also noted that the quality of their reconstruction is limited by the fact that “the algorithm has no knowledge about the physical processes underlying the objects being reconstructed”, and suggested restricting the search space to solutions compatible with a physical model [WAG\*12].

In a follow-up work, Wenger *et al.* [WLM13] presented an algorithm based on group sparsity that dramatically improves the computational performance of the previous approach [WAG\*12] (see Fig. 9). Their method computes a single projection instead of multiple projections and thus reduces memory consumption and computation time. It is again inspired by compressed sensing: an  $\ell_\infty$  group sparsity regularizer is used to suppress noise, and an  $\ell_2$  data term is used to ensure that the output is consistent with the observational data [WLM13]. This method enables astronomers and end users in planetariums or educational facilities to reconstruct stellar objects without the need for specialized hardware.

Hasenberger *et al.* [HA20] added to the hallowed pantheon of automatic object reconstruction algorithms with AVIATOR: a Vienna inverse-Abel-transform-based object reconstruction algorithm. Existing reconstruction techniques (e.g., [WAG\*12, WLM13]) contain potentially problematic requirements such as symmetry in the plane of projection. AVIATOR’s reconstruction algorithm assumes that,

for the object of interest, its morphology “along the line of sight is similar to its morphology in the plane of the projection and that it is mirror symmetric with respect to this plane” [HA20]. Hasenberger *et al.* applied AVIATOR to dense molecular cloud cores and found that their models agreed well with profiles reported in the literature.

**3D object reconstruction using stellar object distances.** The Gaia data release 2 (Gaia DR2) contains a wealth of information about the night sky. Großschedl *et al.* [GAM\*18] used the distances of 700 stellar objects from this dataset to infer a model of Orion A that describes its 3D shape and orientation. This 3D model leads to many insights, among them that the nebulae is longer than previously thought and that it has a cometary shape pointing toward the Galactic plane, where the majority of the Milky Way’s disk mass lies. The authors pointed out that Gaia is bringing the critical third



**Figure 9:** A fast reconstruction algorithm that creates 3D models of nebulae based on their approximate axial symmetry. Image reproduced from Wenger *et al.* [WLM13].

spatial dimension to infer cloud structures and to study star-form interstellar medium.

In a similar manner, Skowron *et al.* [SSM<sup>\*</sup>19] constructed a 3D map of the Milky Way galaxy, using the positions and distances of thousands of classical Cepheid variable stars, which in turn are obtained through observations and accounting of the stars' pulsating periods coupled with luminosity. *Cepheid variable* are regularly pulsating stars, where their regular pulsations allow us to calculate their distances precisely. Skowron *et al.* used 2341 such stars to sketch the Milky Way galaxy and observe the warped shape of the galactic disk, and they were able to define the characteristics of this warping with some precision. They visualized and performed additional analysis on this 3D map using a combination of static 2D/3D plots.

**3D object reconstruction using spectroscopic data.** Vogt *et al.* [VD11] aimed to characterize the 3D shape of a young oxygen-rich supernova remnant (N132D) in the Large Magellanic Cloud, a satellite dwarf galaxy of the Milky Way. Using spectroscopic data from the Wide Field Spectrograph along with sophisticated data reduction techniques, they produced a data cube, which they used to construct a 3D map of the oxygen-rich ejecta of the supernova remnant of interest. They provided several different 2D and 3D plots showing unique views of this 3D map. Their visual analysis has led to insights about the structure of this supernova remnant beyond what was previously known.

**Dark matter filament reconstruction.** Polyphorm [EBPF21] is an interactive visualization and filament reconstruction tool that enables the investigation of cosmological datasets (see Fig. 10). Through a fast computational simulation method inspired by the foraging behavior of *Physarum polycephalum*, astrophysicists are able to extrapolate from sparse datasets, such as galaxy maps archived in the SDSS, and then use these extrapolations to inform analyses of a wide range of other data, such as spectroscopic observations captured by the Hubble Space Telescope. Researchers can update the simulation at interactive rates by a wide range of adjusting model parameters. Polyphorm has been used to reconstruct the cosmic web from galaxy observations [BET<sup>\*</sup>20] and to infer the ionized intergalactic medium contribution to the dispersion measure of a fast radio burst [SBP<sup>\*</sup>20].

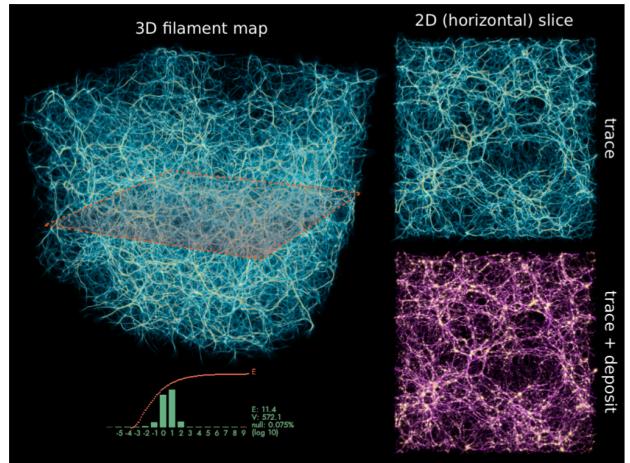
**Visual verification of simulations.** Currently, predictions of the Sun's Coronal mass ejections (CMEs) rely on simulations generated from observed satellite data. CMEs are powerful eruptions from the surface of the sun. These simulations possess inherent uncertainty because that the input parameters are entered manually, and the observed satellite data may contain measurement inaccuracies. These simulations treat CMEs as singular objects with discrete boundaries that are well defined and thus enable their treatment as entire objects. In order to mitigate this uncertainty, Bock *et al.* [BPM<sup>\*</sup>15] proposed a multi-view visualization system that generates an ensemble of simulations by perturbing the CME input parameters, and enables comparisons between these simulations and ground truth measurements. The system has many capabilities useful to domain experts, including integration of 3D rendering of simulations with satellite imagery, comparison of simulation predictions with observed data, and time-dependent analysis.

**3D visualization of planetary surfaces.** Ortner *et al.* [OWN<sup>\*</sup>20] performed 3D reconstruction visualization for planetary geology. Their geological analysis of 3D Digital Outcrop Models is used to reconstruct ancient habitable environments, which serves as an important aspect of the upcoming ESA ExoMars 2022 Rosalind Franklin Rover and the NASA 2020 Rover Perseverance missions on Mars. They conducted a design study to create InCorr (Interactive data-driven Correlations), which includes a 3D geological logging tool and an interactive data-driven correlation panel that evolves with the stratigraphic analysis. See [Ger14, Section 2.2.2] for more references on Mars geology and geodesy data and tools. Bladin *et al.* [BAB<sup>\*</sup>18] integrated multiple data sources and processing and visualization methods to interactively contextualize geospatial surface data of celestial bodies for use in science communication.

## 7. Education and Outreach

Currently, an on-going paradigm shift is occurring in scientific outreach. Technological advances are enabling data-driven and interactive exploration to be possible in public environments such as museums and science centers, increasing their availability to the general public. These advances are shortening the distance between research and outreach material, and enriching the scientific exploration process with new perspectives. Ynnerman *et al.* [YLT18] and Goodman *et al.* [GHWY19] introduced the *Exploranation* concept, a euphemism encapsulating this confluence of explanation and exploration.

Scientific storytelling of astrophysical findings using visualization has a deep history. Ma *et al.* [MLF<sup>\*</sup>12] described how visualization can aid scientific storytelling using the NASA Scientific Visualization Studio. Borkiewicz *et al.* described storytelling based on data-driven cinematic visualization in a SIGGRAPH



**Figure 10:** Polyphorm: reconstruction of dark matter filaments in the simulated BolshoiPlanck dataset where Polyphorm yields a consistent 3D structure, enabling its calibration to cosmic over density values. Thin slices of the filament map are shown on the right. Image reproduced from Elek *et al.* [EBPF21].

	Education & outreach	2D/3D plots	2D images	3D rendering	Interactive vis.	Dim. reduction	Uncertainty vis.	New display	Extragalac. astronomy	Galactic astronomy	Planetary astronomy	Solar astronomy	Simulation	Observation
[BHY18]	•	•	•			•			•	•	•	•	•	•
[FSW*19]		•				•			•	•	•	•	•	•
[RFG*18]	•	•	•		•		•		•	•	•	•	•	•
[USTY18]		•	•	•			•		•	•	•			•
[OPB*19]		•	•			•			•	•	•	•	•	
[AJP*18]		•	•			•			•	•	•			•
[VS13]			•			•			•	•				•
[Mad17]			•			•			•				•	•
[DF17]		•				•		•	•				•	•
[DHG*18]		•	•	•	•			•					•	

**Table 5:** Classifying papers under education and outreach based on secondary and tertiary categories. Top row, from left to right: (primary category) Education and outreach; (secondary categories) 2D/3D plots, 2D images, 3D rendering, interactive visualization, dimensionality reduction, uncertainty visualization, and new display platforms; (tertiary categories) extragalactic, galactic, planetary, and solar astronomy; (tags) simulated, and observational data.

course [BCK\*19]. More recently, and with a greater focus on interactive methods where the user becomes part of the exploration, Bock *et al.* [BHY18] described the challenge of presenting the details of NASA space missions to the public.

Research efforts in this category (as summarized in Table 5) are related *w.r.t.* important aspects of education outreach and/or public accessibility. In addition, several are concerned with large-scale immersive visualization in planetariums and also with personal virtual reality (VR) and augmented reality (AR) experiences. Absent from the current literature, to the best of our knowledge, is a comprehensive analysis of the effect of VR for scientific exploration in astronomy.

**Planetarium and other large-scale immersive environments.** Immersive visualization in planetarium dome theaters (see Fig. 11) has been the primary outreach mechanism for astronomy from their initial conception. The immersive nature of the display system plays an important role in the contextualization of the available data, which is one of the unique challenges of astronomical datasets. The birth of the usage of interactive visualization software in planetarium can be traced to the Uniview software [KHE\*10], which pioneered many of the interaction paradigms that are still in use today. To a large extent, these live presentations based on interactive visualization are enabled by software provided by the planetarium vendors, which are, in general, commercial solutions and thus fall outside the scope of this survey. Our focus here is instead on the large number of open-source initiatives, which are easily accessible to the academic community, targeting planetariums and other large-scale immersive environments. Although aimed at the use in immersive environments, these initiatives also constitute a bridge between outreach and research-driven data exploration, as described by Faherty *et al.* [FSW\*19], which is increasingly gaining momentum.

Among the most widely used software packages tailored to astrophysical data in immersive environments is WorldWide Telescope [RFG\*18], which is a virtual observatory for the public to share and view data from major observatories and telescopes. The software provides the capability to visualize the solar system and stars, and show observational material in context; however it fo-

cuses on the data as displayed from the Earth’s viewpoint. Celestia is an open-source initiative that shows the objects of the solar system and the greater universe in a 3D environment. It provides high-resolution imagery and accurate positioning of the planetary bodies of the solar system and the ability to show other datasets in their context outside the solar system. OpenSpace, Gaia Sky, and ESASky, as described in Sect. 4, also provide contextualization of astronomical data, but with a stronger emphasis on the ability for domain experts to import their data into an immersive environment for public presentations. The Stellarium software can be used by the general public to look at a virtual nights sky from the surface of any planet. The data contained in the software include a star catalog, known deep space objects, satellite positions, and other datasets that can be added dynamically by the user. NASA Eyes is a suite of web-based visualization tools that enable the user to learn about the Earth, Solar System, Exoplanets, and ongoing NASA missions. While providing a rich experience for the end user, the avenues for extension are limited. The Mitaka software [USTY18] enables users to explore the observable universe and makes it easy for them to create custom presentations that can be shown on a large variety of immersive display environments.

**Personal virtual and augmented reality.** In stark contrast to the immersive display environments described thus far, a large amount



**Figure 11:** An example of an interaction presentation in a large-scale immersive environment to the general public, in this case of topographical features on the surface of Mars present at the Brilliant Minds conference.

of work has been presented in the realm of virtual reality (VR) and augmented reality (AR). VR in this context refers to a *personal* experience, rather than one that is *shared* with other participants. These two fields overlap in some areas, but they present distinct research challenges.

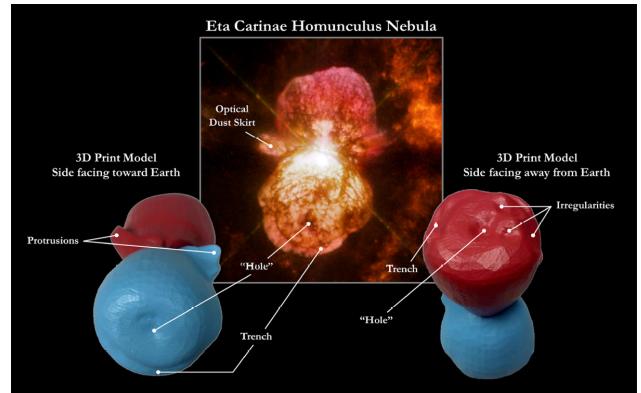
With 3DMAP-VR, Orlando *et al.* [OPB<sup>\*</sup>19] sought to unite the excessively data-rich world of astronomy with VR with the hopes of facilitating productive engagement with traditionally inaccessible data. Specifically, they visualized 3D magnetohydrodynamic (MHD) models of astronomical simulations, which include the effects of gravity and hydrodynamics as well as magnetic fields. Their workflow consisted of two steps: obtaining accurate simulations and then converting these simulations to navigable 3D VR environments using available data analysis and visualization software. In addition to providing a method to explore these dense data cubes in VR, they allowed these VR environments to be explored by anyone with a VR rig by uploading them to Sketchfab, a popular open platform for sharing VR content. Orlando *et al.* excelled at meeting two emerging goals in astronomical visualization: first, using existing software to achieve their goals rather than creating something from scratch; and second, making the visualizations widely accessible.

Arcand *et al.* [AJP<sup>\*</sup>18] developed a 3D VR and AR program to visualize the *Cassiopeia A* (*i.e.*, *Cas A*) supernova remnant, the resulting structure from an exploded star. They aimed to make the best use of the high-resolution, multi-wavelength, multi-dimensional astronomical data and give users the experience of walking inside the remains of a stellar explosion. They first performed 3D reconstruction of *Cas A* and then employed volume and surface rendering to display the model in the VR system YURT (*i.e.*, Yurt Ultimate Reality Theatre). The user can select a specific part of the supernova model and access the annotations. These interactive features not only help non-experts engage in the story of the star, but also assist researchers observe changes in its properties.

Vogt *et al.* [VS13] explored the potential of AR in astrophysics research/education by introducing Augmented Posters and Augmented Articles. The authors included Augmented Posters at the Astronomical Society of Australia Annual General meeting in 2012. Incorporating AR into posters allowed attendees to use their smartphones to engage with the posters in a virtual space and easily save and share the poster information if they found it interesting. Through tracking of engagement and feedback, they discovered that the majority of conference attendees found the technology to “have some potential.” As mentioned, the authors also experimented with Augmented Articles. They showed how results from an earlier work (the 3D structure of super nova remnants) can be viewed in 3D interactively within the article using a smartphone. Vogt *et al.* concluded by speculating on the future of AR in astrophysics. They were optimistic about the potential for AR to be an effective supplementary technology, but cited long-term stability and backwards compatibility in terms of AR apps and technology as a major limitation to AR moving forward. They suggested that a dedicated service for AR used in scientific publishing and outreach may be an effective way to handle this limitation.

**Novel interfaces.** Madura [Mad17] presented a case study using 3D printing to visualize the  $\eta$  Car Homunculus nebula, see Fig. 12.

Extending the traditional monochromatic 3D prints, Madura proposed to use full-color sandstone prints to generate more informative models. Although the sandstone material is not as sturdy, these printers produce noticeably higher quality prints that preserve smaller details. The colors of the prints can be based on physical properties, which provides additional information to visual learners and helps distinguish different structures. The 3D models not only facilitate research discoveries, but also help communicate scientific discoveries to non-astronomers, especially to the visually impaired. The New Mexico Museum of Space History and the New Mexico School jointly hosted the first week-long astronomy camp for the visually impaired students across states in the summer of 2015. The camp received overwhelmingly positive feedback. Madura also discussed the use of other methods, including audio, tactile sign language, and tactile fingerspelling, to further expand the 3D model interactive experience for tactile learners. Overall, 3D printing could be a useful and effective tool for astronomy outreach and education.



**Figure 12:** Dual-color 3D print of the  $\eta$  Car Homunculus nebula. Image reproduced from Madura [Mad17].

In a work as artistic as it is scientific, Diemer *et al.* [DF17] modeled the cosmic web through dark matter simulations and represented it artistically through 3D sculptures and woven textiles. The dark matter simulation is run using the publicly available code GADGET2 and the halos and *subhalos* (*i.e.*, the halos within halos), are identified using ROCKSTAR. To identify the structural elements of the cosmic web (walls and filaments), they used DISPERSE, which is publicly available code that leverages discrete Morse theory to identify salient topological features (see Sect. 5). Converting from the simulation data to their artistic representation, Diemer *et al.* stated that “we believe that art, as much as science, seeks to say something true about the nature of existence, and that end is best served by artistic representation that grapples with real data and not only with allegorical concepts.” They accomplished this stated ideal through a structured simplification of the model to a form where it can be represented using 3D woven textiles. The techniques used in their paper are not novel, but the combination of them is, and the end result is a powerful installation that instills wonder in those who move through it. This work is able to take scientifically rigorous simulation data and represent it in an accessible form without losing the deep beauty of the underlying science.

This elegant translation of numbers to forms, of thoughts to feel-

ings, lies at the heart of science communication and outreach. The importance of this translation is especially crucial for astronomy, where the physical embodiment of the things we are studying can really only ever live in our minds and as the modern equivalent of paint on the walls of our caves.

**Virtual observatories.** Virtual observatories (VOs) are web-based repositories of astronomical data from multiple sources with the goal of improving access to astronomical data for scientists not directly involved in data collection. Many advanced VO applications aid in the mining and exploration of observational data through the use of state-of-the-art visualization techniques, but comparatively few that perform similar functions for theoretical data. In the interactive 3D visualization for theoretical VOs, Dykes *et al.* [DHG\*18] examined the current capabilities of VOs containing theoretical data, and additionally presented a tool based on SPLOTCH, which is designed to aid in addressing some of the shortcomings they identified with current methods. SPLOTCH is a visualization tool designed for high-performance computing environments that is capable of quickly rendering large particle datasets, which makes it ideal for interactive visualization of 3D particle data. Dykes *et al.* combined their tool with a VO and demonstrated the effectiveness of interactively filtering and quantitatively visualizing the data for identifying features of interest in large particle simulations. Future steps involve comparative visualization, which would consist of methods to generate mock 2D observational images from the 3D simulation data to compare with actual observations.

**Broad dissemination through mobile applications.** Furthermore, recent progress has been made on mobile and commercial applications that provide visualizations of astronomical data. Although these applications are not focused on research questions, they serve to broadly disseminate astrophysics visualizations, many of which also have strong educational value for the general public.

For instance, SkySafari (<https://skysafariastronomy.com/>), StarMap (<http://www.star-map.fr/>), NightSky (<https://icandiapps.com/>), or SkyView (on Apple App Store) bring stargazing to everyone's mobile phones, offer AR features that guide the layperson towards interested targets, and can also be used to control amateur telescopes and aid in astrophotography. Many JavaScript libraries and tools are available for astronomical visualization, such as Asterank (<https://www.asterank.com/>), based on spacekit.js (<https://typo.github.io/spacekit/>), which enables the user to visually explore a database containing over 600,000 asteroids, including estimated costs and rewards of mining asteroids.

Web applications that introduce scientific knowledge of astronomy are also easily accessible. For example, NASA JPL's Eyes on the Solar System (<https://eyes.nasa.gov/>) has the ability to show the dynamics of our solar system, but can also be used to show the evolution of space missions, and the discovery of exoplanets. Another example is an adaptation of the Uniview software to be used in schools on mobile platforms targeting grades 4–6 in the NTA Digital project.

There are also a number of popular applications for VR headsets, such as Star Chart (<http://www.escapistgames.com/>) and Our Solar System on Oculus, which provide immersive

experiences for users seeking knowledge about the Universe. Merge Cube (<https://mergeedu.com/cube>) accompanied with AR/VR apps such as MERGE Explorer has been used to enable a new way of interactive learning for astronomy and beyond. It allows users to hold digital 3D objects and explore stars and galaxies in their palms.

## 8. Challenges and Opportunities

In addition to a taxonomy of existing approaches that utilize visualization to study astronomical data from Sect. 3 to Sect. 7, our contribution includes a summary of the current challenges and opportunities in visualization for astronomy. We ask the following questions: What are the missing tools in current visualization research that astronomers need to formulate and test hypotheses using modern data? What visualization capabilities are expected to become available for astronomical data over the next decade?

In a Carnegie + SCI mini-workshop conducted in April 2020 and a Visualization in Astrophysics workshop during IEEE VIS in October 2020, astrophysicists and visualization experts discussed recent advances in visualization for astrophysics, as well as the current visualization needs in the astronomical community. As a result of these workshops, we have identified the following list of challenges and opportunities:

- Open-source tools: we need more open-source data visualization software that is suitable for astronomical data. These tools must be flexible, modular, and integrable within a broader ecosystem of workhorse tools;
- Intelligent data querying: we need to enable intelligent data queries for large data;
- Discovery: we need ways to turn high-quality renderings of data (observed and simulated) into quantitative information for discovery;
- Scalable feature extraction: we need to extract and visualize features from large and physically complex data cubes;
- In situ analysis and visualization: we need to interact with simulation data in real time, by utilizing visualization for parameter tuning and simulation steering;
- Uncertainty visualization: we need to develop more techniques to mitigate and communicate the effects of data uncertainty on visualization and astronomy;
- Benchmarks: we need to develop clear, widely adopted benchmarks or mock data catalogs for comparison with observed data;
- Time and space efficiency: we need to improve upon memory and/or space intensive data analysis tasks.

### 8.1. Challenges Identified from Previous Surveys

We first review the challenges identified from previous surveys [HF11, LLC\*12] and describe how the community has responded to these challenges in the past decade.

Hassan *et al.* [HF11] identified six grand challenges in their 2011 survey for the peta-scale astronomy era:

- Support quantitative visualization;
- Effectively handle large data sizes;
- Promote discoveries in low signal-to-noise data;

- Establish better human-computer interaction and ubiquitous computing;
- Design better workflow integration;
- Encourage adoption of 3D scientific visualization techniques.

Lipsa *et al.* [LLC<sup>\*</sup>12] discussed visualization challenges in astronomy in their 2012 survey; however, only a few papers had addressed these challenges at the time of the survey. These challenges include:

- Multi-field visualization, feature detection, graphics hardware;
- Modeling and simulation;
- Scalable visualization, error and uncertainty visualization, time-dependent visualization, global and local visualization, and comparable visualization.

In the past decade, considerable advances have been made in addressing the challenges identified by Hassan *et al.* and Lipsa *et al.* With modern computing power, interactive visualizations have become increasingly popular for both scientific explorations and public outreach. A variety of scalable visualization tools for large simulation and survey data are now easily accessible (e.g., ParaView [WHA<sup>\*</sup>11], OpenSpace [BAC<sup>\*</sup>20] and Gaia Sky [SJMS19]). Many tools are also adopting graphics hardware and parallelism in their visualization, rendering, and analysis processes to increase efficiency (e.g., yt [TSO<sup>\*</sup>10] and [MPF20]). Scientists and educators are also incorporating novel visual display methods, such as VR [OPB<sup>\*</sup>19] and 3D printing [Mad17], for education and outreach services.

Visualizations of evolving astronomical systems have also seen advances. Lipsa *et al.* [LLC<sup>\*</sup>12] listed only two papers under time-dependent astronomy visualization in their survey. In contrast, we present a number of research papers with the capabilities of analyzing halo evolution histories [SXL<sup>\*</sup>14, PGX<sup>\*</sup>16, SBD<sup>\*</sup>17] and rendering real-time stellar orbits [SJMS19]. Volumetric data can now be rendered in 3D with sufficient numerical accuracy to enable a wide range of research in feature detection and extraction (e.g., AstroBlend [Nai16], FRELLED [Tay17b] and Houdini for astrophysics [NBC17]), see Sect. 5.

Quantitative analysis for heterogeneous data types (Sect. 4.1) is often supported as a supplement to the visual analysis (e.g., Encube [VBF<sup>\*</sup>16] and TOPCAT [Tay17a]). In order to perform analytic tasks effectively, many of these tools utilize visualization techniques such as linked-views (multi-field visualization [LLC<sup>\*</sup>12], Glue [Glu12]), detail-on-demand (global/local visualization [LLC<sup>\*</sup>12]), and comparative visualization.

On a higher level, a few techniques and platforms have been developed to provide better visualization workflow in astronomy. The Glue visualization environment described by Goodman *et al.* [GBR18] hosts a variety of shared datasets and open-source software. It facilitates flexible data visualization practices, and bridges the gap between scientific discovery and communication. On a similar note, the EU-funded CROSS DRIVE [Ger14] creates “collaborative, distributed virtual workspaces” in order to unite the fragmented experts, data, and tools in European space science. Mohammed *et al.* [MPF20] formalized the scientific visualization workflow and brought structure to a visualization designer’s decision-making process. The paradigm provided by Mohammed

*et al.* divides the visualization process into four steps: processing, computation of derived geometric and appearance properties, rendering, and display. In each of these steps, the workflow systematically incorporates high-performance computing to efficiently work with multi-variate multi-dimensional data.

However, despite the progress, some of the challenges identified a decade ago, such as uncertainty visualization and time-dependent visualization, remain largely under-explored today or have great potential for improvement. A careful inspection of Table 1 to Table 5 gives rise to a number of useful observations regarding research gaps for further investigation. In this section, we describe a number of challenges and opportunities that we believe are essential for the development of visualization in astronomy in the years to come.

## 8.2. Astronomical Data Volume and Diversity

A challenge identified by both Hussan *et al.* [HF11] and Lipsa *et al.* [LLC<sup>\*</sup>12] is the effective handling of large datasets. Substantial effort and progress has been made in processing large datasets in the past decade in both astronomy and visualization. Luciani *et al.* [LCO<sup>\*</sup>14] pre-processed large-scale survey data to ensure efficient query and smooth interactive visualization. Frellé [Tay15] accelerates visual source extraction to enable the visualization of large 3D volumetric datasets. Filtergraph [BSP<sup>\*</sup>13] supports the rapid visualization and analysis of large datasets using scatter plots and histograms. Gaia Sky [SJMS19] uses the magnitude-space level-of-detail structure to effectively visualize hundreds of millions of stars from the Gaia mission with sufficient numerical precision. yt [TSO<sup>\*</sup>10] adopts parallelism to run multiple independent analysis units on a single dataset simultaneously.

Visualizing large datasets remains a challenge for astronomical data, especially because of its multi-dimensional property. Visualization researchers recognize that scalability is an immediate obstacle that prevents them from introducing many interactive capabilities [Tay15, PGX<sup>\*</sup>16, SKW<sup>\*</sup>11, WAG<sup>\*</sup>12, SBD<sup>\*</sup>17]. For analysis tasks, the developers of yt identified the challenge of load balancing for parallel operations on large simulation data. They added support for robust CPU/GPU mixed-mode operation to accelerate numerical computation [TSO<sup>\*</sup>10]. We believe that even more improvements can be achieved by using network data storage and high-performance computing.

As the volume and diversity of data increase rapidly, connecting related heterogeneous datasets has become a priority in astronomy. Goodman *et al.* [GBR18] identified the growing open-source and collaborative environment as the future of astronomy. They described the Glue [Glu12] visualization environment, a platform that hosts a large variety of data and numerous open-source modular software. The Glue environment allows users to load multiple datasets at once and “glue” the related attributes together from different data types. Many exploratory astronomy visualization software packages (e.g., OpenSpace, ESASky) are capable of dealing with various data types. Some can integrate with Glue, which further improves their integrability and flexibility.

Nevertheless, most software packages are still striving to expand the variety of data formats that they can process. Naiman

*et al.* [NBC17] aimed to use Houdini to render data with non-uniform voxel sizes. Baines *et al.* [BGR<sup>\*</sup>16] retrieved spectroscopic data and aimed to link it to more of the mission catalogs for the next release of ESASky. Burchett *et al.* [BAO<sup>\*</sup>19] incorporated data pre-processing as part of the IGM-Vis application to allow more data formats as inputs. Vogt *et al.* [VSDR17] provided a unique perspective for simplifying the access to 3D data visualization by promoting the X3D pathway, as the X3D file format lies in the center of various visualization solutions, such as interactive HTML, 3D printing and high-end animations. However, the conversion into X3D file format remains the largest obstacle.

### 8.3. Interactive Visualization and Intelligent Querying

Hassan *et al.* identified “better human-computer-interaction” as one of the six grand challenges [HF11], and visualization experts and astronomers have joined forces to explore the potential of using interactive visualization in astronomy research and public outreach. We see the overwhelming popularity of interactive visualization in the realm of astronomy research. Barnes and Fluke [BF08] demonstrated the convenience of embedding interactive visualizations in astronomy publications, via 3D PDF and its extension S2PLOT programming library. Frelled and AstroBlend leverage the 3D capability of Blender to improve the process of analyzing and visualizing volumetric data [Tay17b, Nai16]. Naiman *et al.* [NBC17] explored the potential use of the graphics tool Houdini in astronomy research. ESASky, LSSGalpy, SDvision, OpenSpace, and Gaia Sky [BGR<sup>\*</sup>16, AFPR<sup>\*</sup>17, PCHT17, BAC<sup>\*</sup>20, SJMS19] all provide visual exploratory capabilities to large-scale astronomy survey data, each with their own scientific focuses and distinguishing features.

Many of these interactive software tools are expanding their impact in public outreach. A video produced with the visualizations from SDvision – titled “Laniakea: Our home supercluster” – gained millions of views on YouTube [PCHT17]. The authors are also pursuing the software’s integration with VR technology to further contribute to education and public outreach services. OpenSpace has already demonstrated its success in museums, planetariums, and a growing library of publicly accessible video material. The software is built to be easily accessible to the general public via a simple installation onto any computer. AR, VR, and 3D printing are emerging technologies that are used at a greater scale in educational and public outreach services [AJW<sup>\*</sup>19, Mad17]. In order to reach a more artistic audience, Diemer *et al.* [DF17] have also explored integrating art and the physical visualization of astronomical objects.

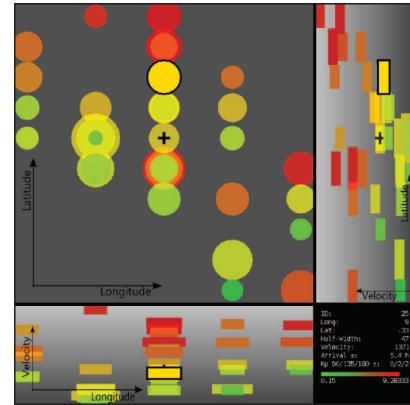
However, many researchers also recognize the limitations of current interactive visualizations and intelligent querying of volumetric data. Barnes and Fluke [BF08] proposed the capturing of mouse clicks on individual elements of a scene to enable 3D selection and queries. Goodman advocated for the need of 3D selection in astronomy visualization and analysis [Goo12]. Blender allows only one side of a transparent spherical mesh to be displayed at a time [Tay17b]. The selection of pixels in regions of interest could also be a potential problem [Tay17b]. Yu *et al.* [YEII16] proposed several context-aware selections in 3D particle clouds, which help to alleviate the issues associated with 3D selection and query. Fe-

lix tackles the challenge of querying by simultaneously satisfying two density ranges, but Shivashankar *et al.* [SPN<sup>\*</sup>16] identified interactive visualization of intricate 3D networks as a “largely unexplored problem of major significance”. WYSIWYG creates a marching cube of a 2D selection and finds the cluster with the largest projection area as the cluster of interest [SXL<sup>\*</sup>14]. The technique lacks flexibility as it depends heavily on the assumption that the largest cluster is always of interest.

### 8.4. Uncertainty Visualization

Uncertainty visualization in astronomy remains largely unexplored, even though errors and uncertainties are introduced due to data acquisition, transformation, and visualization. Li *et al.* [LFLH07] noticed that uncertainty visualization is seldom available in astronomical simulations and developed techniques that enhance perception and comprehension of uncertainty across a wide range of scales. Since then, a few studies have considered errors created during the simulation pipeline. Großschedl *et al.* [GAM<sup>\*</sup>18] used uncertainty plots to effectively present the distribution of the data and demonstrated the confidence in their reconstruction results. With the direct intention of incorporating uncertainty in the discovery process, Bock *et al.* [BPM<sup>\*</sup>15] displayed the uncertainty of space weather simulation by visualizing an ensemble of simulation results with different input parameters (Fig. 13). Combined with a timeline view and a volumetric rendering of each ensemble member, scientists are able to compare each simulation with measured data, gain an understanding of the parameter sensitivities, and detect correlations between the parameters.

Applying uncertainty visualization to 3D astronomy data is challenging because we lack the techniques to deal with sparse/far away samples and their large error cones. However, the potential exists to display uncertainty in a localized object or regions of interest, and that potential must be developed further.

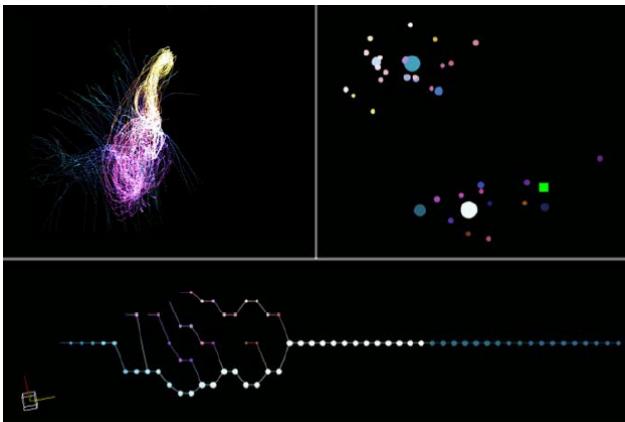


**Figure 13:** Ensemble selection view that captures the uncertainty of all ensemble runs by displaying the full 4D parameter space. Image reproduced from Bock *et al.* [BPM<sup>\*</sup>15].

### 8.5. Time Series Data Visualization and Analysis

Most of the current time series data visualizations are built to display halo evolution. One common technique is to use a merger tree to visualize the development of halos over time [SXL<sup>\*</sup>14, AF15, PGX<sup>\*</sup>16, SBD<sup>\*</sup>17]. Other techniques are often used along with the merger tree to enhance the effectiveness of the visualization. Shan *et al.* [SXL<sup>\*</sup>14] visualized the changes of selected particles as a particle trace path image (see Fig. 14). Preston *et al.* [PGX<sup>\*</sup>16] added interactivity into their software and facilitated more efficient analysis for large, heterogeneous data. Scherzinger *et al.* [SBD<sup>\*</sup>17] extended their framework by adding particle visualization and analysis for individual halos.

In general, time series data visualization can also be helpful when tracking star movements. However, little effort has been expended in this regard, with Gaia Sky being the only example, to the best of our knowledge. Given the instantaneous proper motion vector of stars and simulation time, Gaia Sky [SJMS19] computes a representation of proper motions. The software is able to visualize real-time star movements with sufficient numerical accuracy.



**Figure 14:** Interactive linked views of halo evolutionary history. Top left shows the evolution path of a selected halo and top right is the halo projected onto a 2D screen. Bottom is the merger tree visualization of the halo evolution. Image reproduced from Shan *et al.* [SXL<sup>\*</sup>14].

### 8.6. Machine Learning

During our recent astrophysics workshops, many astronomers voiced their desires as well as concerns about using machine learning techniques (ML), and in particular, deep learning in the astrophysics discovery processes, mostly surrounding the interpretability of “black box” ML models. Active discussions have concerned the maturity of ML in astronomy, and a number of surveys have been created to assess this maturity [BB10, FJ20, NAB<sup>\*</sup>19].

**Combine visualization with machine learning.** Although the concerns regarding the interpretability of ML models in astronomy are valid in some cases, we believe that combining visualization with ML models has the potential to make the results more accessible to

classical theoretical interpretation. Indeed, some of the most successful applications of ML in astrophysics involve cases where the interpretation is straightforward.

The use of deep learning techniques in astrophysics has mostly been limited to convolutional neural networks (CNNs). Khan *et al.* [KHW<sup>\*</sup>19] used a combination of deep learning and dimensionality reduction techniques to help construct galaxy catalogs. They extracted the features from one of the last layers of their pre-trained CNN and clustered the features using t-SNE. Their method not only leads to promising classification results, but also points out errors in the galaxy zoo dataset with the misclassified examples. Ntampaka *et al.* [NZE<sup>\*</sup>19] presented a CNN that estimated galaxy cluster masses from the *Chandra* mock images. They used visualization to interpret the results of learning and to provide physical reasoning. Kim and Brunner [KB16] performed star-galaxy classification using CNNs. They studied images of activation maps, which help to explain how the model is performing classification tasks. Apart from deep learning, Reis *et al.* [RPB<sup>\*</sup>18] used random forests to generate distances between pairs of stars, and then visualize such a distance matrix using t-SNE. Their techniques have been shown to be useful to identify outliers and to learn complex structures with large spectroscopic surveys.

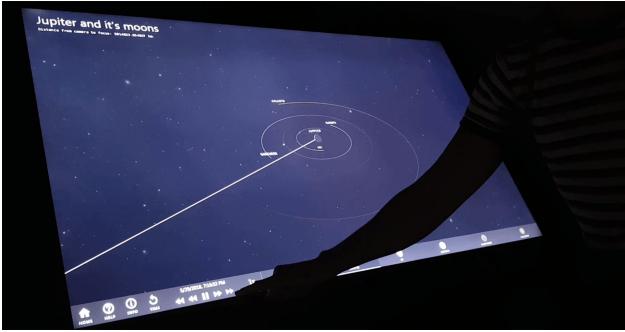
Many efforts in recent years have focused on interpreting ML models [Mol20]. We believe a good starting point to obtain interpretability is to combine visualization with models that are inherently interpretable [Rud19] (e.g., linear regression, decision tree, decision rules, and naive Bayes) in studying astronomical data. Alternatively, we may train an interpretable model as a surrogate to approximate the predictions of a black box model (such as a CNN) and integrate such a surrogate in our visualization.

**Topological data analysis.** Furthermore, topological data analysis (TDA) is an emerging field that promotes topology-based unsupervised learning techniques. TDA infers insights from the shape of the data, and topology has a reasonably long history in its applications in scientific visualization [HLH<sup>\*</sup>16]. A few researchers have applied TDA to astrophysics. Novikov *et al.* [NCD06] were the first to propose the method of extracting the skeleton of the cosmic web using discrete Morse theory [For02]. Both Sousbie [Sou11] and Shivashankar *et al.* [SPN<sup>\*</sup>16] used discrete Morse theory to develop geometrically intuitive methods that extract features from the cosmic web (e.g., filaments, walls, or voids). They demonstrated the efficiency and effectiveness of topological techniques in astronomical tasks. Xua *et al.* [XCKGN19] used TDA techniques to identify cosmic voids and loops of filaments and assign their statistical significance. Not many applications of topology have been proposed in de-noising astronomy data, other than the work of Rosen *et al.* [RSM<sup>\*</sup>19], which uses contour trees in the de-noising and visualization of radio astronomy (ALMA) data cubes.

### 8.7. Further Advancements in Education and Outreach

A general ambition in science communication is to shorten the distance between research and outreach and make current research results and data available at science centers, museums, and in on-line repositories. This ambition applies to both shortening the time between discovery and dissemination and creating increased public

access to research data. Even real-time public participation in scientific endeavors has been shown to be of public interest [BHY18]. This science communication trend is supported by rapid development of commodity computing platforms capable of handling large datasets, availability of open research data, and improved data analysis and visualization tools. These trends now enable visitors to public venues and home users to become “explorers” of scientific data. Astrophysics is one of the prime examples of a domain of large public interest and with vast amounts of publicly available data.



**Figure 15:** Usage of a touch interface in a museum installation guiding the user using cues built directly into the data exploration.

The trend described above poses several challenges. In a public setting, an interactive exploration has to be curated and guided in such a way that learning and communication goals are reached while not interfering with “freedom” of interaction [YRA\*16]. Fig. 15 shows an example of this approach on a touch-capable table used in museum environments for a self-guided learning experience, exemplified on a CT scan of a meteoroid originating from Mars. Ynnerman *et al.* [YLT18] coined the term *Exploranation* to describe the introduction of data exploration in traditional explanatory contexts, and Goodman *et al.* [GBR18] described the need for interaction in explanation. This endeavor calls for a new generation of authoring and production tools targeting production of interactive non-linear storytelling [WH07], with interfaces that interoperate with research tools and repositories. We also see that interactive installations will need to feature several different modes of operation. A first scenario would be a “walk-up-and-use” situation in which the content and the interaction are intuitive. The second scenario is a guided experience with a trained facilitator who can unlock advanced features of the installation and also bring in more data sources for an in-depth science communication session.

Interaction plays a central role in science communication, and public settings put demands on robust, engaging, intuitive, and interactive visualization interfaces. Sundén *et al.* [SBJ\*14] discussed aspects of demand and the potential use of multi-modal interfaces. Yu *et al.* [YEII12, YEII16] addressed challenges posed by the interactive selections of data using touch surfaces.

In live presentation situations based on interactive software, more advanced tools are needed that support the presenter (and the pilot). Apart from the authoring tools discussed above, research on features such as automatic camera moves during presentations is

also needed. An interesting challenge in view of advances in machine learning and natural language processing is the use of voice and gesture based interaction during presentations. Support for the embedding of multi-media data sources and other on-line services is also needed.

In outreach, the key role of visual representations cannot be underestimated, which calls for systems and tools that generate both visually appealing and still scientifically correct representations. The challenge here is a trade-off between artistic and scientific considerations. From an artistic point of view, Rector *et al.* [RLF\*17] aimed to strike a balance between the scientific and aesthetic quality of an astronomical image. They pointed out that people have different expectations and misconceptions of colored-image-based factors such as cultural variation and expertise. Therefore, scientists need to carefully consider the choices they make in order to create astronomical images. An example of how this challenge is met is the work on cinematic data visualization by Cox *et al.* [CPL\*19, BCWW20]. Another example is the interactive blackhole visualization [VE21] described in Sect. 3.

The on-going rapid development of computer hardware creates opportunities and challenges as the users expect visual quality on the same level as the state-of-the-art games. At the same time, new levels of widespread public use are made possible. The challenge is to work with visual quality and performance as well as to create awareness of limited computer capabilities, data size and complexity. Another challenge for outreach is the use of social media and connected services, which entails not only development of tools and availability of data, but also engagement of a large number of domain experts with a science communication mission.

## 9. Navigation Tool

Together with the classification and descriptions of the papers included in this survey, we complement our survey with a visual literature browser available at <https://tdavislab.github.io/astrovis-survis>. The visual browser follows the same classification scheme used in this report, where the users can use keyword searches to identify potential aspects in the field that are underserved. Additionally, we provide an alternative navigation tool within the visual browser (also illustrated in Fig. 16), where the surveyed papers are distributed along two axes. This tool provides a different viewpoint for the state-of-the-art survey.

The first x-axis – **single task vs. general purpose** – specifies whether a specific paper addresses a singular challenge of visualization in astronomy (single task), or whether it describes a more general purpose system that can be applied to a wide array of potential applications (general purpose). A general purpose system also includes software systems that combine datasets of multiple modalities in a shared contextualization. The second y-axis – **technique vs. application** – specifies whether a paper develops a specific visualization or analysis technique, or whether it combines many different techniques to address a specific application. The primary category – data analysis tasks – is double-encoded with colors and marker shapes. The “other” category represents relevant papers mentioned in the survey but that do not belong to any of the data analysis tasks. The coordinates of the papers in the navigation

tool are based on our best estimation. We lay out the papers in their general areas in the figure to avoid overlap of labels.

## 10. Conclusions

In this report, we provide an overview of the state of the art in astrophysics visualization. We have surveyed the literature and found that visualization in astrophysics can be categorized broadly into five categories based upon the primary user objectives: data wrangling, data exploration, feature identification, object reconstruction, and education and outreach.

A major finding of this work is that there remains a significant gap between cutting-edge visualization techniques and astrophysical datasets. Among the 80+ papers surveyed, around 20 papers are from visualization venues. Given the scope of current and future datasets in astrophysics, as well as the advanced methodologies and capabilities in visualization research, the potential opportunity is great for scientific discovery in bridging this gap. However, this bridge will not build itself.

We therefore take this opportunity to issue a “call to action” for both the visualization and astrophysics communities to consider more robust and intentional ways of bridging the gap between visualization methods and astronomy data. We make the specific recommendations below as concrete suggestions for improving this goal over the next decade.

We suggest the construction of a comprehensive **AstroVis Roadmap** for bringing these disparate communities and stakeholders together at both the grassroots and institutional levels. In order to build community, we suggest regular annual joint meetings that will explicitly target this gap and bring together visualization and astrophysics domain expertise; the 2019 Dagstuhl Seminar on the topic of “Astrographics: Interactive Data-Driven Journeys through Space” is a good example [GHWY19]. We specifically suggest yearly companion meetings to be held alternately at the Winter AAS or annual IAU meetings and the IEEE Visualization conferences. Having explicit joint sponsorship of the professional society is an important step in growing this joint community.

We recognize and appreciate the “grassroots” efforts that bring together the visualization and astrophysics communities. Indeed, this contribution is the direct result of a Carnegie-SCI workshop as well as the IEEE Vis2020 workshop “Visualization in Astrophysics” (<http://www.sci.utah.edu/~beiwang/visastro2020/>). Other efforts include the *RHytHM* (ResearchH using yt Highlights Meeting) at the Flatiron Institute in 2020, the application spotlight event (<https://sites.google.com/view/viscomspotlight/>) at IEEE VIS 2020 discussing opportunities and challenges in cosmology visualization, and the annual *glue-con* hackathon (<https://www.gluessolutions.io/glue-con>) that integrate astronomical software projects including glue, yt, OpenSpace, WorldWide Telescope into a centralized system. These meetings are critical, in addition to the larger meetings we suggest above.

To assist in the access to information and literature, we suggest that an “**astrovis**” keyword be added to papers – published in either astrophysics journals or visualization publications – to make

interrogation of papers easy for the communities. Our visual literature browser that enables exploration of “astrovis” papers is a step in this direction.

We further suggest starting **visualization/data challenges** within the large publicly available astrophysics data surveys. The “solutions” to these challenges should be made publicly available and thus applicable to other datasets. A few scientific visualization (SciVis) challenges have involved astronomy datasets at IEEE VIS conferences, notably the SciVis 2015 Contest using the Dark Sky Simulations (<https://darksky.slac.stanford.edu/>) and the SciVis 2019 Contest using the HACC cosmological simulation (<https://wordpress.cels.anl.gov/2019-scivis-contest/>). A recent example is the data challenge held at the IEEE VIS 2020 VisAstro workshop, where a visualization tool under development called CosmoVis was used by Burchett *et al.* [BAEF20] to interactively analyze cosmological simulations generated by IllustrisTNG. In terms of pedagogy, we recommend summer schools, hackathons, and workshops that help onboard members of each community engage in this joint effort. The data challenges may serve as the seeding point for such workshops. Science not communicated is science not done. Looking toward science education and public outreach, we highlight the important role played by museums and planetariums in transitioning scientific discovery to public education. We suggest that these stakeholders be included organically in the discovery process and emphasize their key role in the scientific process.

We are encouraged by the progress that has been made in the last decade, and we look forward to the next decade of development, discovery, and education.

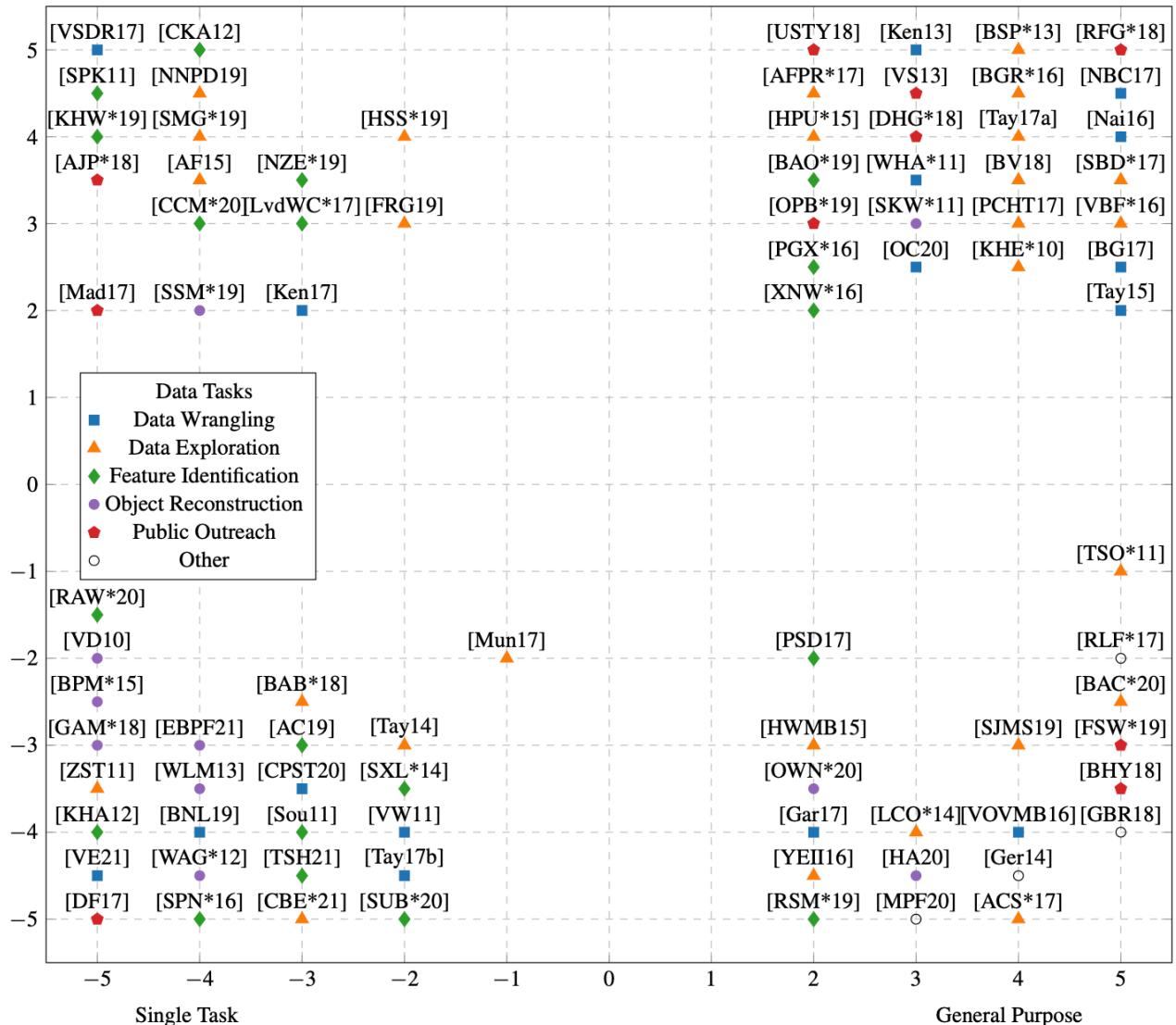
## Acknowledgements

The authors would like to thank the constructive suggestions by the reviewers and the generous funding by the following entities: Carnegie Institution for Science Venture Grant no. BPO-700082; The Canadian Institute for Advanced Research Fellowship program; The Ahmanson Foundation; NASA under grant no. NNX16AB93A; the Swedish e-Science Research Centre; the Knut and Alice Wallenberg Foundation; and the Swedish Research Council DNR:2015-05462. The authors would also like to thank the speakers and participants of following two workshops, who helped us identify challenges and opportunities via great discussions: the Carnegie + SCI mini-workshop (April 2020), and the Visualization in Astrophysics workshop during IEEE VIS (October 2020), in particular, Chuck Hansen, Chris R. Johnson, Alyssa Goodman, Jackie Faherty, Matthew J. Turk, Ryan Wyatt, and Joseph N. Burchett.

## References

- [AC19] ARAGON-CALVO M. A.: Classifying the large-scale structure of the universe with deep neural networks. *Monthly Notices of the Royal Astronomical Society* 484, 4 (2019), 5771–5784. 10, 13
- [ACS\*17] AXELSSON E., COSTA J., SILVA C., EMMART C., BOCK A., YNNERMAN A.: Dynamic scene graph: enabling scaling, positioning, and navigation in the universe. *Computer Graphics Forum* 36, 3 (2017), 459–468. 7
- [AF15] ALMRYDE K. R., FORBES A. G.: Halos in a dark sky: interactively exploring the structure of dark matter halo merger trees. In 2015

Application  
Technique  
Single Task      General Purpose



**Figure 16:** An alternative navigation tool that highlights the surveyed papers along two axes: single task vs. general purpose and technique vs. application.

- IEEE Scientific Visualization Conference (SciVis) (2015), pp. 73–77. 7, 9, 21
- [AFPR\*17] ARGUDO-FERNÁNDEZ M., PUERTAS S. D., RUIZ J. E., SABATER J., VERLEY S., BERGOND G.: LSSGalPy: interactive visualization of the large-scale environment around galaxies. *Publications of the Astronomical Society of the Pacific* 129, 975 (2017). 7, 20
- [AJP\*18] ARCAND K. K., JIANG E., PRICE S., WATZKE M., SGOUROS T., EDMONDS P.: Walking through an exploded star: rendering supernova remnant cassiopeia a into virtual reality. *Communicating Astronomy with the Public Journal* 24 (2018), 17–24. 16, 17
- [AJW\*19] ARCAND K. K., JUBETT A., WATZKE M., PRICE S., WILLIAMSON K., EDMONDS P.: Touching the stars: improving NASA 3D printed data sets with blind and visually impaired audiences. *Journal of Science Communication* 18, 4 (2019). 20

- [Arc05] ARCHIVE N. I. S.: Montage: an astronomical image mosaic engine. <http://montage.ipac.caltech.edu/>, 2005. 5
- [BAB\*18] BLADIN K., AXELSSON E., BROBERG E., EMMART C., LJUNG P., BOCK A., YNNERMAN A.: Globe browsing: contextualized spatio-temporal planetary surface visualization. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (2018), 802–811. 7, 15
- [BAC\*20] BOCK A., AXELSSON E., COSTA J., PAYNE G., ACINAPURA M., TRAKINSKI V., EMMART C., SILVA C., HANSEN C., YNNERMAN A.: Openspace: a system for astrographics. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (2020), 633–642. 7, 8, 19, 20
- [BAEF20] BURCHETT J. N., ABRAMOV D., ELEK O., FORBES A. G.: Volumetric reconstruction for interactive analysis of the cosmic web. *Data Challenge Winner at IEEE VIS VisAstro Workshop 2020* (2020). 23

- [BAO\*19] BURCHETT J. N., ABRAMOV D., OTTO J., ARTANEGARA C., PROCHASKA J. X., FORBES A. G.: IGM-Vis: analyzing intergalactic and circumgalactic medium absorption using quasar sightlines in a cosmic web context. *Computer Graphics Forum* 38, 3 (2019), 491–504. [10](#), [11](#), [20](#)
- [BB10] BALL N. M., BRUNNER R. J.: Data mining and machine learning in astronomy. *International Journal of Modern Physics D* 19, 07 (2010), 1049–1106. [21](#)
- [BCK\*19] BORKIEWICZ K., CHRISTENSEN A., KOSTIS H.-N., SHIRAH G., WYATT R.: Cinematic scientific visualization: the art of communicating science. In *ACM SIGGRAPH 2019 Courses* (2019), Association for Computing Machinery, pp. 1–273. [16](#)
- [BCWW20] BORKIEWICZ K., CHRISTENSEN A., WYATT R., WRIGHT E. T.: Introduction to cinematic scientific visualization. In *ACM SIGGRAPH 2020 Courses* (2020), Association for Computing Machinery, pp. 1–267. [22](#)
- [Ben12] BENSON A. J.: Galacticus: a semi-analytic model of galaxy formation. *New Astronomy* 17, 2 (2012), 175–197. [9](#)
- [BET\*20] BURCHETT J. N., ELEK O., TEJOS N., PROCHASKA J. X., TRIPP T. M., BORDOLOI R., FORBES A. G.: Revealing the dark threads of the cosmic web. *The Astrophysical Journal* 891, 2 (2020). [15](#)
- [BF08] BARNES D. G., FLUKE C. J.: Incorporating interactive three-dimensional graphics in astronomy research papers. *New Astronomy* 13, 8 (2008), 599–605. [5](#), [20](#)
- [BFBP06] BARNES D. G., FLUKE C. J., BOURKE P. D., PARRY O. T.: An advanced, three-dimensional plotting library for astronomy. *Publications of the Astronomical Society of Australia* 23, 2 (2006), 82–93. [5](#)
- [BG17] BERRIMAN G. B., GOOD J. C.: The application of the montage image mosaic engine to the visualization of astronomical images. *Publications of the Astronomical Society of the Pacific* 129, 975 (2017). [4](#), [5](#)
- [BGR\*16] BAINES D., GIORDANO F., RACERO E., SALGADO J., MARTÍ B., MERÍN B., ET AL.: Visualization of multi-mission astronomical data with ESASky. *Publications of the Astronomical Society of the Pacific* 129, 972 (2016). [6](#), [7](#), [20](#)
- [BHY18] BOCK A., HANSEN C., YNNERMAN A.: Openspace: bringing nasa missions to the public. *IEEE Computer Graphics and Applications* 38, 5 (2018), 112–118. [16](#), [22](#)
- [BKW16] BECK F., KOCH S., WEISKOPF D.: Visual analysis and dissemination of scientific literature collections with Survis. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 180–189. [2](#)
- [Ble02] BLENDER ONLINE COMMUNITY: Blender. <https://www.blender.org/>, 2002. [4](#)
- [BNL19] BORKIEWICZ K., NAIMAN J. P., LAI H.: Cinematic visualization of multiresolution data: Ytini for adaptive mesh refinement in houdini. *The Astronomical Journal* 158, 1 (2019). [4](#)
- [BNO\*14] BRYAN G. L., NORMAN M. L., O’SHEA B. W., ABEL T., WISE J. H., TURK M. J., ET AL.: Enzo: an adaptive mesh refinement code for astrophysics. *The Astrophysical Journal Supplement Series* 211, 2 (2014). [6](#)
- [BPM\*15] BOCK A., PEMBROKE A., MAYS M. L., RASTAETTER L., ROPINSKI T., YNNERMAN A.: Visual verification of space weather ensemble simulations. In *2015 IEEE Scientific Visualization Conference (SciVis)* (2015), pp. 17–24. [14](#), [15](#), [20](#)
- [BSP\*13] BURGER D., STASSUN K. G., PEPPER J., SIVERD R. J., PAEGERT M., LEE N. M. D., ROBINSON W. H.: Filtergraph: an interactive web application for visualization of astronomy datasets. *Astronomy and Computing* 2 (2013), 40–45. [6](#), [7](#), [19](#)
- [BV18] BREDDELS M. A., VELJANOSKI J.: Vaex: big data exploration in the era of gaia. *Astronomy & Astrophysics* 618, A13 (2018). [7](#), [9](#)
- [CBE\*21] COSTA J., BOCK A., EMMART C., HANSEN C., YNNERMAN A., SILVA C.: Interactive visualization of atmospheric effects for celestial bodies. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2021), 785–795. [7](#), [8](#)
- [CCM\*20] CIURLA A., CAMPBELL R. D., MORRIS M. R., DO T., GHEZ A. M., HEES A., ET AL.: A population of dust-enshrouded objects orbiting the galactic black hole. *Nature* 577, 7790 (2020), 337–340. [10](#), [12](#)
- [CKA12] CAMPBELL R., KJÆR K., AMICO P.: 3D visualisation of integral field spectrometer data. *The Messenger* 148 (2012), 28–31. [10](#), [12](#)
- [CPL\*19] COX D., PATTERSON R., LEVY S., CHRISTENSEN A., BORKIEWICZ K., CARPENTER J., ET AL.: "birth of planet earth" full-dome excerpt: photosynthesis in a chromatophore. In *ACM SIGGRAPH 2019 Computer Animation Festival* (2019), Association for Computing Machinery. [22](#)
- [CPST20] COMRIE A., PIŃSKA A., SIMMONDS R., TAYLOR A.: Development and application of an HDF5 schema for SKA-scale image cube visualization. *Astronomy and Computing* 32 (2020). [4](#), [5](#)
- [Dai85] DAINTY J. C.: Progress in object reconstruction in astronomy. In *1984 European Conf on Optics, Optical Systems, and Applications* (1985), vol. 0492, SPIE, pp. 234–241. [13](#)
- [DF17] DIEMER B., FACIO I.: The fabric of the universe: exploring the cosmic web in 3D prints and woven textiles. *Publications of the Astronomical Society of the Pacific* 129, 975 (2017). [16](#), [17](#), [20](#)
- [DHG\*18] DYKES T., HASSAN A., GHELLER C., CROTON D., KROKOS M.: Interactive 3D visualization for theoretical virtual observatories. *Monthly Notices of the Royal Astronomical Society* 477, 2 (2018), 1495–1507. [16](#), [18](#)
- [EAA\*19] EVENT HORIZON TELESCOPE COLLABORATION, AKIYAMA K., ALBERDI A., ALEF W., ASADA K., AZULY R., ET AL.: First m87 event horizon telescope results. I. the shadow of the supermassive black hole. *The Astrophysical Journal Letters* 875, 1 (2019). [1](#)
- [EBPF21] ELEK O., BURCHETT J. N., PROCHASKA J. X., FORBES A. G.: Polyphorm: structural analysis of cosmological datasets via interactive Physarum polycephalum visualization. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2021), 806–816. [13](#), [14](#), [15](#)
- [ERG19] EMANTS B., RABA R., G. MOELLENBROCK ET. AL.: The CASA software for radio astronomy: status update from ADASS 2019. In *29th annual international Astronomical Data Analysis Software & Systems (ADASS) conference* (2019). [4](#), [5](#)
- [FBO06] FLUKE C. J., BOURKE P. D., O’DONOVAN D.: Future directions in astronomy visualization. *Publications of the Astronomical Society of Australia* 23, 1 (2006), 12–24. [5](#)
- [FH07] FU C., HANSON A. J.: A transparently scalable visualization architecture for exploring the universe. *IEEE Transactions on Visualization and Computer Graphics* 13, 1 (2007), 108–121. [7](#)
- [FHK\*11] FOLK M., HEBER G., KOZIOL Q., POURMAL E., ROBINSON D.: An overview of the HDF5 technology suite and its applications. In *Proceedings of the EDBT/ICDT 2011 Workshop on Array Databases* (2011), Association for Computing Machinery, pp. 36–47. [5](#)
- [FJ20] FLUKE C. J., JACOBS C.: Surveying the reach and maturity of machine learning and artificial intelligence in astronomy. *WIREs Data Mining and Knowledge Discovery* 10, 2 (2020), e1349. [21](#)
- [Fla17] FLAMARY R.: Astronomical image reconstruction with convolutional neural networks. In *2017 25th European Signal Processing Conference (EUSIPCO)* (2017), pp. 2468–2472. [13](#)
- [For02] FORMAN R.: A user’s guide to discrete morse theory. *Séminaire Lotharingien de Combinatoire* 48 (12 2002). [21](#)
- [FRG19] FRITSCHI L., ROJO I. B., GÜNTHER T.: Visualizing the temporal evolution of the universe from cosmology simulations. In *2019 IEEE Scientific Visualization Conference (SciVis)* (2019), pp. 1–11. [7](#), [9](#)

- [FSW\*19] FAHERTY J. K., SUBBARAO M., WYATT R., YNNERMAN A., TYSON N. D., GELLER A., ET AL.: Ideas: immersive dome experiences for accelerating science. *Bulletin of the American Astronomical Society* 51, 7 (2019). 16
- [GAM\*18] GROSSSCHEIDL J. E., ALVES J., MEINGAST S., ACKERL C., ASCENSO J., BOUY H., ET AL.: 3D shape of Orion A from Gaia DR2. *Astronomy & Astrophysics* 619 (2018). 13, 14, 20
- [Gár17] GÁRATE M.: Voxel datacubes for 3D visualization in blender. *Publications of the Astronomical Society of the Pacific* 129, 975 (2017). 4
- [GBR18] GOODMAN A. A., BORKIN M. A., ROBITAILLE T. P.: New thinking on, and with, data visualization. *arXiv: Instrumentation and Methods for Astrophysics* (2018). 4, 7, 19, 22
- [Ger14] GERNDT A.: Collaborative Rover Operations and Planetary Science Analysis System based on Distributed Remote and Interactive Virtual Environments. Final Publishable Summary Report, European Union Framework Programme 7, 2014. 15, 19
- [GHWY19] GOODMAN A. A., HANSEN C. D., WEISKOPF D., YNNERMAN A.: Astrographics: interactive data-driven journeys through space (Dagstuhl Seminar 19262). *Dagstuhl Reports* 9, 6 (2019), 95–124. 15, 23
- [Glu12] GLUE DEVELOPERS: Glue: multi-dimensional linked-data exploration. <https://glueviz.org/>, 2012. 19
- [Goo12] GOODMAN A. A.: Principles of high-dimensional data visualization in astronomy. *Astronomische Nachrichten* 333, 5-6 (2012), 505–514. 6, 20
- [HA20] HASENBERGER B., ALVES J.: AVIATOR: morphological object reconstruction in 3D. *Astronomy & Astrophysics* 633 (Jan 2020). 13, 14
- [HF11] HASSAN A., FLUKE C. J.: Scientific visualization in astronomy: towards the petascale astronomy era. *Publications of the Astronomical Society of Australia* 28, 2 (2011), 150–170. 1, 18, 19, 20
- [HLH\*16] HEINE C., LEITTE H., HLAWITSCHKA M., IURICICH F., FLORIANI L. D., SCHEUERMANN G., ET AL.: A survey of topology-based methods in visualization. *Computer Graphics Forum* 35, 3 (2016), 643–667. 21
- [HPAM19] HANULA P., PIEKUTOWSKI K., AGUILERA J., MARAI G. E.: Darksky halos: use-based exploration of dark matter formation data in a hybrid immersive virtual environment. *Frontiers in Robotics and AI* 6 (2019), 11. 9
- [HPF\*16] HABIB S., POPE A., FINKEL H., FRONTIERE N., HEITMANN K., DANIEL D., ET AL.: HACC: simulating sky surveys on state-of-the-art supercomputing architectures. *New Astronomy* 42 (2016), 49–65. 5
- [HPU\*15] HANULA P., PIEKUTOWSKI K., URIBE C., ALMRYDE K., NISHIMOTO A., AGUILERA J., MARAI G. E.: Cavern halos: exploring spatial and nonspatial cosmological data in an immersive virtual environment. In *Proceedings of the IEEE Scientific Visualization Conference (SciVis)* (2015), pp. 87–99. 7, 9
- [HSS\*19] HESSE-EDENFELD C., STEINKE M. J., SANTALIDIS N., HUESMANN K., LEISTIKOW S., LINSEN L.: Interactive multi-level visualization of combined particle and multi-field cosmology data. In *2019 IEEE Scientific Visualization Conference (SciVis)* (2019), pp. 12–30. 7, 9
- [HWMB15] HAZARIKA S., WEI T.-H., MUKHERJEE R., BARBUR A.: Visualizing the life and anatomy of dark matter. In *2015 IEEE Scientific Visualization Conference (SciVis)* (2015), pp. 101–106. 7, 9
- [Jae08] JAEGER S.: The common astronomy software application (casa). *Astronomical Data Analysis Software and Systems XVII, ASP Conference Series* (2008). 5
- [KB16] KIM E. J., BRUNNER R. J.: Star-galaxy classification using deep convolutional neural networks. *Monthly Notices of the Royal Astronomical Society* 464, 4 (2016), 4463–4475. 21
- [Ken13] KENT B. R.: Visualizing astronomical data with blender. *Publications of the Astronomical Society of the Pacific* 125, 928 (2013), 731–748. 4
- [Ken17] KENT B. R.: Spherical panoramas for astrophysical data visualization. *Publications of the Astronomical Society of the Pacific* 129, 975 (2017). 4
- [KHA12] KAEHLER R., HAHN O., ABEL T.: A novel approach to visualizing dark matter simulations. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2078–2087. 10
- [KHE\*10] KLASHED S., HEMINGSSON P., EMMART C., COOPER M., YNNERMAN A.: Uniview - visualizing the universe. In *Eurographics 2010 - Areas Papers* (2010), The Eurographics Association. 7, 16
- [KHP\*11] KANDEL S., HEER J., PLAISANT C., KENNEDY J., VAN HAM F., RICHE N. H., WEAVER C., LEE B., BRODBECK D., BUONO P.: Research directions in data wrangling: Visualizations and transformations for usable and credible data. *Information Visualization* 10, 4 (2011), 271–288. 4
- [KHW\*19] KHAN A., HUERTA E., WANG S., GRUENDL R., JENNINGS E., ZHENG H.: Deep learning at scale for the construction of galaxy catalogs in the dark energy survey. *Physics Letters B* 795 (2019), 248–258. 10, 13, 21
- [KPH13] KNEBE A., PEARCE F. R., HANNI LUX ET. AL.: Structure finding in cosmological simulations: the state of affairs. *Monthly Notices of the Royal Astronomical Society* 435, 2 (2013), 1618–1658. 3
- [LCO\*14] LUCIANI T. B., CHERINKA B., OLIPHANT D., MYER S., WOOD-VASEY W. M., LABRINIDIS A., MARAI G. E.: Large-scale overlays and trends: visually mining, panning and zooming the observable universe. *IEEE Transactions on Visualization and Computer Graphics* 20, 7 (2014), 1048–1061. 6, 7, 19
- [LFLH07] LI H., FU C.-W., LI Y., HANSON A.: Visualizing large-scale uncertainty in astrophysical data. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (2007), 1640–1647. 20
- [LLC\*12] LIPŞA D. R., LARAMEE R. S., COX S. J., ROBERTS J. C., WALKER R., BORKIN M. A., PFISTER H.: Visualization for the physical sciences. *Computer Graphics Forum* 31, 8 (2012), 2317–2347. 1, 18, 19
- [LvdWM17] LIBESKIND N. I., VAN DE WEYGAERT R., MARIUS CAUTUN ET AL.: Tracing the cosmic web. *Monthly Notices of the Royal Astronomical Society* 473, 1 (2017), 1195–1217. 10, 11
- [Mad17] MADURA T. I.: A case study in astronomical 3D printing: the mysterious Eta Carinae. *Publications of the Astronomical Society of the Pacific* 129, 975 (2017), 058011. 16, 17, 19, 20
- [MKDH04] MAGNOR M., KINDLMANN G., DURIC N., HANSEN C.: Constrained inverse volume rendering for planetary nebulae. In *IEEE Visualization 2004* (2004), pp. 83–90. 14
- [MLF\*12] MA K.-L., LIAO I., FRAZIER J., HAUSER H., KOSTIS H.-N.: Scientific storytelling using visualization. *IEEE Computer Graphics and Applications* 32, 1 (2012), 12–19. 15
- [Mol20] MOLNAR C.: *Interpretable machine learning: a guide for making black box models explainable*. <https://christophm.github.io/interpretable-ml-book/>, 2020. 21
- [MPF20] MOHAMMED A., POLYS N., FARRAH D.: Visualize this: lessons from the front-lines of high performance visualization. In *Computer Science Technical Reports*. Department of Computer Science, Virginia Polytechnic Institute & State University, 2020. 19
- [MSRMH09] MEYER-SPRADOW J., ROPINSKI T., MENSMANN J., HINRICHES K.: Vorreen: a rapid-prototyping environment for ray-casting-based volume visualizations. *IEEE Computer Graphics and Applications* 29, 6 (2009), 6–13. 7, 8
- [MT20] MUNK M., TURK M. J.: widgys: custom jupyter widgets for interactive data exploration with yt. *Journal of Open Source Software* 5, 45 (2020), 1774. 6

- [Mun17] MUNA D.: Introducing nightlight: a new FITS viewer. *Publications of the Astronomical Society of the Pacific* 129, 975 (2017), 058003. [7](#), [8](#)
- [NAB\*19] NTAMPAKA M., AVESTRUZ C., BOADA S., ET AL.: The role of machine learning in the next decade of cosmology. *Bulletin of the American Astronomical Society* 51, 3 (2019). [21](#)
- [Nai12] NAIMAN J.: AstroBlend. www.astroblend.com, 2012. [4](#)
- [Nai16] NAIMAN J.: AstroBlend: an astrophysical visualization package for blender. *Astronomy and Computing* 15 (2016), 50–60. [4](#), [5](#), [19](#), [20](#)
- [NBC17] NAIMAN J. P., BORKIEWICZ K., CHRISTENSEN A. J.: Houdini for astrophysical visualization. *Publications of the Astronomical Society of the Pacific* 129, 975 (2017), 058008. [4](#), [19](#), [20](#)
- [NCD06] NOVIKOV D., COLOMBI S., DORE O.: Skeleton as a probe of the cosmic web: the two-dimensional case. *Monthly Notices of the Royal Astronomical Society* 366, 4 (2006), 1201–1216. [21](#)
- [NNPD19] NGUYEN B. D., NGUYEN N. V. T., PHAM V., DANG T.: Visualization of data from HACC simulations by paraview. In *2019 IEEE Scientific Visualization Conference (SciVis)* (2019), pp. 31–32. [5](#), [7](#), [9](#)
- [NZE\*19] NTAMPAKA M., ZUHONE J., EISENSTEIN D., NAGAI D., VIKHLININ A., HERNQUIST L., ET AL.: A deep learning approach to galaxy cluster x-ray masses. *The Astrophysical Journal* 876, 1 (2019), 82. [10](#), [13](#), [21](#)
- [OC20] OTT J., CARTA TEAM: CARTA: cube analysis and rendering tool for astronomy. In *American Astronomical Society Meeting Abstracts #235* (2020), vol. 235, p. 364.11. [4](#), [5](#)
- [OPB\*19] ORLANDO S., PILLITTERI I., BOCCINO F., DARICELLO L., LEONARDI L.: 3DMAP-VR, a project to visualize three-dimensional models of astrophysical phenomena in virtual reality. *Research Notes of the American Astronomical Society* 3, 11 (2019), 176. [16](#), [17](#), [19](#)
- [OWN\*20] ORTNER T., WALCH A., NOWAK R., BARNES R., HÖLLT T., GRÖLLER E.: Incorr: interactive data-driven correlation panels for digital outcrop analysis. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2020), 755–764. [14](#), [15](#)
- [PCHT17] POMARÈDE D., COURTOIS H. M., HOFFMAN Y., TULLY R. B.: Cosmography and data visualization. *Publications of the Astronomical Society of the Pacific* 129, 975 (2017), 058002. [7](#), [20](#)
- [PGX\*16] PRESTON A., GHODS R., XIE J., SAUER F., LEAF N., MA K.-L., ET AL.: An integrated visualization system for interactive analysis of large, heterogeneous cosmology data. In *2016 IEEE Pacific Visualization Symposium (PacificVis)* (2016), pp. 48–55. [9](#), [10](#), [11](#), [19](#), [21](#)
- [PGY05] PUETTER R., GOSNELL T., YAHIL A.: Digital image reconstruction: deblurring and denoising. *Annual Review of Astronomy and Astrophysics* 43 (2005), 139–194. [13](#)
- [PHCT17] POMAREDE D., HOFFMAN Y., COURTOIS H. M., TULLY R. B.: The cosmic v-web. *The Astrophysical Journal* 845, 1 (2017), 55. [7](#)
- [PSD17] PILLEPICH A., SPRINGEL V., DYLAN NELSON ET AL.: Simulating galaxy formation with the IllustrisTNG model. *Monthly Notices of the Royal Astronomical Society* 473, 3 (2017), 4077–4106. [10](#), [11](#)
- [RAW\*20] RAVOUX C., ARMENGaud E., WALTHER M., ETOURNEAU T., POMARÈDE D., PALANQUE-DELABROUILLE N., ET AL.: A tomographic map of the large-scale matter distribution using the eBOSS—Stripe 82 Ly $\alpha$  forest. *Journal of Cosmology and Astroparticle Physics* 2020, 07 (2020), 010–010. [10](#)
- [RFG\*18] ROSENFIELD P., FAY J., GILCHRIST R. K., CUI C., WEIGEL A. D., ROBITAILLE T., OTOR O. J., GOODMAN A.: AAS WorldWide telescope: a seamless, cross-platform data visualization engine for astronomy research, education, and democratizing data. *The Astrophysical Journal Supplement Series* 236, 1 (2018), 22. [16](#)
- [RLF\*17] RECTOR T. A., LEVAY Z. G., FRATTARE L. M., ARCAND K. K., WATZKE M.: The aesthetics of astrophysics: how to make appealing color-composite images that convey the science. *Publications of the Astronomical Society of the Pacific* 129, 975 (2017), 058007. [22](#)
- [RPB\*18] REIS I., POZNANSKI D., BARON D., ZASOWSKI G., SHAHAF S.: Detecting outliers and learning complex structures with large spectroscopic surveys - a case study with APOGEE stars. *Monthly Notices of the Royal Astronomical Society* 476, 2 (2018), 2117–2136. [21](#)
- [RSM\*19] ROSEN P., SETH A., MILLS B., GINSBURG A., KAMENETZKY J., KERN J., JOHNSON C. R., WANG B.: Using contour trees in the analysis and visualization of radio astronomy data cubes. *arXiv e-prints* (2019), arXiv:1704.04561. [10](#), [12](#), [21](#)
- [Rud19] RUDIN C.: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence* 1 (2019), 206–215. [21](#)
- [SBD\*17] SCHERZINGER A., BRIX T., DREES D., VOLKER A., RADKOV K., SANTALIDIS N., FIEGUTH A., HINRICH K.: Interactive exploration of cosmological dark-matter simulation data. *IEEE Computer Graphics and Applications* 37, 2 (2017), 80–89. [7](#), [8](#), [19](#), [21](#)
- [SBJ\*14] SUNDÉN E., BOCK A., JÖNSSON D., YNNERMAN A., ROPINSKI T.: Interaction techniques as a communication channel when presenting 3D visualizations. In *2014 IEEE VIS International Workshop on 3DVIs (3DVIs)* (2014), pp. 61–65. [22](#)
- [SBP\*20] SIMHA S., BURCHETT J. N., PROCHASKA J. X., CHITTIDI J., ELEK O., TEJOS N., ET AL.: Disentangling the cosmic web towards FRB 190608. *The Astrophysical Journal* 901, 2 (2020), 134. [15](#)
- [SHTG14] SCHLEI W., HOWELL K., TRICOCHE X., GARTH C.: Enhanced visualization and autonomous extraction of poincaré map topology. *The Journal of the Astronautical Sciences* 61 (2014), 170–197. [12](#)
- [SJMS19] SAGRISTÀ A., JORDAN S., MÜLLER T., SADLO F.: Gaia sky: navigating the gaia catalog. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2019), 1070–1079. [7](#), [8](#), [19](#), [20](#), [21](#)
- [SKW\*11] STEFFEN W., KONING N., WENGER S., MORISSET C., MAGNOR M.: Shape: a 3D modeling tool for astrophysics. *IEEE Transactions on Visualization and Computer Graphics* 17, 4 (2011), 454–465. [13](#), [14](#), [19](#)
- [SMG\*19] SCHATZ K., MÜLLER C., GRAFLKA P., HEINEMANN M., STRAUB A., SCHULZ C., ET AL.: Visual analysis of structure formation in cosmic evolution. In *2019 IEEE Scientific Visualization Conference (SciVis)* (2019), pp. 33–41. [7](#), [9](#)
- [Sou11] SOUSBIE T.: The persistent cosmic web and its filamentary structure - i. theory and implementation. *Monthly Notices of the Royal Astronomical Society* 414, 1 (2011), 350–383. [10](#), [12](#), [21](#)
- [SPK11] SOUSBIE T., PICHON C., KAWAHARA H.: The persistent cosmic web and its filamentary structure — ii. illustrations. *Monthly Notices of the Royal Astronomical Society* 414, 1 (2011), 384–403. [10](#), [12](#)
- [SPN\*16] SHIVASHANKAR N., PRANAV P., NATARAJAN V., V. D. WEYGAERT R., BOS E. G. P., RIEDER S.: Felix: a topology based framework for visual exploration of cosmic filaments. *IEEE Transactions on Visualization and Computer Graphics* 22, 6 (2016), 1745–1759. [10](#), [12](#), [20](#), [21](#)
- [SSM\*19] SKOWRON D. M., SKOWRON J., MRÓZ P., UDALSKI A., PIETRUKOWICZ P., SOSZYŃSKI I., ET AL.: A three-dimensional map of the milky way using classical cepheid variable stars. *Science* 365, 6452 (2019), 478–482. [14](#), [15](#)
- [SUB\*20] SAWADA N., UEMURA M., BEYER J., PFISTER H., FUJIHIRO I.: TimeTubesX: a query-driven visual exploration of observable, photometric, and polarimetric behaviors of blazars. *IEEE Transactions on Visualization and Computer Graphics Early Access* (2020), 1–12. [10](#), [11](#)
- [SWJ\*05] SPRINGEL V., WHITE S. D. M., JENKINS A., FRENK C. S., YOSHIDA N., GAO L., ET AL.: Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature* 435 (2005), 629–636. [9](#)
- [SWV\*08] SPRINGEL V., WANG J., VOGELSSBERGER M., LUDLOW A., JENKINS A., HELMI A., ET AL.: The aquarius project: the subhalos of galactic halos. *Monthly Notices of the Royal Astronomical Society* 391, 4 (2008), 1685–1711. [9](#)

- [SXL\*14] SHAN G., XIE M., LI F., GAO Y., CHI X.: Interactive visual exploration of halos in large-scale cosmology simulation. *Journal of Visualization* 17, 3 (2014), 145–156. 10, 11, 19, 20, 21
- [TA16] THEYS C., AIME C.: Reconstructing images in astrophysics, an inverse problem point of view. In *Cartography of the Sun and the Stars. Lecture Notes in Physics*, Rozelot J.-P., Neiner C., (Eds.), vol. 914. Springer, Cham, 2016, pp. 1–23. 13
- [Tay05] TAYLOR M. B.: TOPCAT & STIL: Starlink table/votable processing software. In *Astronomical Data Analysis Software and Systems XIV* (2005), vol. 347 of *Astronomical Society of the Pacific Conference Series*, p. 29. 7, 8
- [Tay14] TAYLOR M.: Visualising large datasets in TOPCAT v4. In *Astronomical Data Analysis Software and Systems XXIII* (2014), Astronomical Society of the Pacific Conference Series, p. 257. 7, 8
- [Tay15] TAYLOR R.: FRELLED: a realtime volumetric data viewer for astronomers. *Astronomy and Computing* 13 (2015), 67–79. 4, 19
- [Tay17a] TAYLOR M.: TOPCAT: desktop exploration of tabular data for astronomy and beyond. *Informatics* 4, 3 (2017). 7, 8, 9, 19
- [Tay17b] TAYLOR R.: Visualizing three-dimensional volumetric data with an arbitrary coordinate system. *Publications of the Astronomical Society of the Pacific* 129, 972 (2017), 028002. 4, 19, 20
- [TCD\*13] TULLY R. B., COURTOIS H. M., DOLPHIN A. E., FISHER J. R., HÉRAUDEAU P., JACOBS B. A., ET AL.: Cosmicflows-2: the data. *The Astronomical Journal* 146, 4 (2013), 86. 7
- [TSH21] TRICOCHE X. M., SCHLEI W. R., HOWELL K. C.: Extraction and visualization of Poincaré map topology for spacecraft trajectory design. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2021), 765–774. 10, 12
- [TSO\*10] TURK M. J., SMITH B. D., OISHI J. S., SKORY S., SKILLMAN S. W., ABEL T., NORMAN M. L.: yt: a multi-code analysis toolkit for astrophysical simulation data. *The Astrophysical Journal Supplement Series* 192, 1 (2010), 9. 6, 7, 19
- [Tuk77] TUKEY J. W.: *Exploratory data analysis*. Pearson, 1977. 6
- [USTY18] USUDA-SATO K., TSUZUKI H., YAMAOKA H.: Making “real astronomy” visible to the public. *Nature Astronomy* 2, 9 (2018), 692–694. 16
- [VBF\*16] VOHL D., BARNES D. G., FLUKE C. J., POUDEL G., GEORGIOU-KARISTIANIS N., HASSAN A. H., ET AL.: Large-scale comparative visualisation of sets of multidimensional data. *PeerJ Computer Science* (2016). 7, 8, 19
- [VD11] VOGT F., DOPITA M.: The 3D structure of N132D in the LMC: a late-stage young supernova remnant. *Astrophysics and Space Science* 331, 2 (2011), 521–535. 13, 14, 15
- [VE21] VERBRAECK A., EISEMANN E.: Interactive black-hole visualization. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2021), 796–805. 4, 6, 22
- [VOVMB16] VOGT F. P. A., OWEN C. I., VERDES-MONTENEGRO L., BORTHAKUR S.: Advanced data visualization in astrophysics: the X3D pathway. *The Astrophysical Journal* 818, 2 (2016), 115. 4, 5
- [VS13] VOGT F. P. A., SHINGLES L. J.: Augmented reality in astrophysics. *Astrophysics and Space Science* 347, 1 (2013), 47–60. 16, 17
- [VSDR17] VOGT F., SEITENZAHL I., DOPITA M., RUITER A.: Linking the X3D pathway to integral field spectrographs: YSNR 1E0102.2-7219 in the SMC as a case study. *Publications of the Astronomical Society of the Pacific* 129, 975 (2017), 058012. 4, 5, 20
- [VW12] VOGT F., WAGNER A.: Stereo pairs in astrophysics. *Astrophysics and Space Science* 337 (2012), 79–92. 4, 5
- [WAG\*12] WENGER S., AMENT M., GUTHE S., LORENZ D., TILLMANN A., WEISKOPF D., MAGNOR M.: Visualization of astronomical nebulae via distributed multi-GPU compressed sensing tomography. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2188–2197. 13, 14, 19
- [WGL79] WELLS D. C., GREISEN E. W.: FITS - a flexible image transport system. In *Image Processing in Astronomy* (1979), p. 445. 5
- [WH07] WOHLFART M., HAUSER H.: Story telling for presentation in volume visualization. In *Proceedings of the 9th Joint Eurographics/IEEE VGTC conference on Visualization* (2007), Eurographics Association, pp. 91–98. 22
- [WHA\*11] WOODRING J., HEITMANN K., AHRENS J., FASEL P., HSU C.-H., HABIB S., POPE A.: Analyzing and visualizing cosmological simulations with ParaView. *The Astrophysical Journal Supplement Series* 195, 1 (2011), 11. 4, 5, 19
- [WLM13] WENGER S., LORENZ D., MAGNOR M.: Fast image-based modeling of astronomical nebulae. *Computer Graphics Forum* 32, 7 (2013), 93–100. 13, 14
- [XCKGN19] XUA X., CISEWSKI-KEHE J., GREEN S. B., NAGAI D.: Finding cosmic voids and filament loops using topological data analysis. *Astronomy and Computing* 27 (2019), 34–52. 21
- [XNW\*16] XU L., NAKAYAMA M., WU H.-Y., WATANABE K., TAKAHASHI S., UEMURA M., FUJISHIRO I.: Timetubes: design of a visualization tool for time-dependent, multivariate blazar datasets. In *2016 Nicograph International (NicoInt)* (2016), pp. 15–20. 10, 11
- [YEII12] YU L., EFSTATHIOU K., ISENBERG P., ISENBERG T.: Efficient structure-aware selection techniques for 3D point cloud visualizations with 2DOF input. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2245–2254. 7, 9, 22
- [YEII16] YU L., EFSTATHIOU K., ISENBERG P., ISENBERG T.: CAST: effective and efficient user interaction for context-aware selection in 3D particle clouds. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 886–895. 7, 9, 20, 22
- [YLT18] YNNERMAN A., LÖWGREN J., TIBELL L.: Exploranson: a new science communication paradigm. *IEEE Computer Graphics and Applications* 38, 3 (2018), 13–20. 4, 15, 22
- [YRA\*16] YNNERMAN A., RYDELL T., ANTOINE D., HUGHES D., PERSSON A., LJUNG P.: Interactive visualization of 3d scanned mummies at public venues. *Communications of the ACM* 59, 12 (2016), 72–81. 22
- [ZST11] ZHANG T., SUN Y., TANG Z.: 3D visualization of solar wind ion data from the Chang’E-1 exploration. *Computers & Geosciences* 37, 10 (2011), 1711–1718. 7, 9

## Short Biographies

**Fangfei Lan** is a Ph.D. student in the Scientific Computing and Imaging (SCI) Institute, University of Utah. She is interested in topological data analysis and visualization. She works at the intersection between theories in topological data analysis and their practical applications in science and engineering. Email: fangfei.lan@utah.edu.

**Michael Young** is a Master’s student in the Scientific Computing and Imaging (SCI) Institute, University of Utah. His research focuses on building interactive visualization systems and exploring visualization methods for astronomy datasets. Email: myoung@cs.utah.edu.

**Lauren Anderson** is a Carnegie Postdoctoral Fellow at the Carnegie Observatories. She obtained her Ph.D. in Astronomy from University of Washington. She works on applying computational data analysis techniques to interesting astronomy problems. Currently she is building a dust map of the Milky Way galaxy, and is in need of novel visualization techniques for this large, rich data set. Email: landerson@carnegiescience.edu.

**Anders Ynnerman** is a Professor of Visualization at Linköping University and the director of the Norrköping Visualization Center. He also heading the Media and Information Technology division at Linköping University. His research in visualization focuses on large scale data visualization in a range of application domains, and visualization for science communication is a central application theme. He is the founder of the *OpenSpace* project and a producer of planetarium shows. In 2018 he received the IEEE VGTC Technical Achievement Award. Email: anders.ynnerman@liu.se.

**Alexander Bock** is an Assistant Professor at Linköping University, Research Fellow at the University of Utah, and the Development Lead for the astrovisualization software *OpenSpace*. He uses visualization techniques to create applications that can make difficult data, for example CT/MRI scans or astronomical datasets, easier to understand for a specific target audience, such as domain experts or the general public. Email: alexander.bock@liu.se.

**Michelle A. Borkin** is an Assistant Professor in the Khouri College of Computer Sciences, Northeastern University. Her research interests include data visualization, human-computer interaction, and application work across domains including astronomy, physics, and medical imaging. She received a Ph.D. in 2014 and an M.S. in 2011, both in Applied Physics, and a B.A. in Astronomy and Astrophysics & Physics from Harvard University. She is the visualization lead and original project member of the *glue* multi-dimensional linked-data exploration software, originally developed for astronomy and designed to facilitate data analysis and visualization for NASA's James Webb Space Telescope. Email: m.borkin@northeastern.edu.

**Angus G. Forbes** is an Associate Professor in the Department of Computational Media at University of California, Santa Cruz, where he directs the UCSC Creative Coding Lab. He received his Ph.D. and M.S. degrees from University of California, Santa Barbara. His visualization research investigates novel techniques for representing and interacting with complex scientific information. Email: angus@ucsc.edu.

**Juna A. Kollmeier** is a staff member at the Carnegie Observatories where she is the Founding Director of the Carnegie Theoretical Astrophysics Center. She is also the director of the fifth phase of the Sloan Digital Sky Survey (SDSS; which started in October 2020). SDSS has been one of the foundational datasets in astrophysics but also in data science more generally. SDSS pioneered the open-data model for large survey databases and is a major motivator of advanced visualization techniques in astrophysics. Her research interests are in theoretical astrophysics concerning the growth of cosmic structure on all scales. She works on cosmological hydrodynamics simulations as well as massively multiplexed spectroscopic datasets. Email: jak@carnegiescience.edu.

**Bei Wang** is an assistant professor at the SCI Institute, School of Computing, University of Utah. She works on topological data analysis and visualization. She received her Ph.D. in Computer Science from Duke University. She is interested in the analysis and visualization of large and complex data. Her research interests include topological data analysis, data visualization, computational

topology, computational geometry, machine learning and data mining. She is working on connecting data visualization with astronomy data, in particular, using topological techniques to separate signals from noise. Email: beiwang@sci.utah.edu.