

Reproducible computer vision

Cross-disciplinary and scalable image informatics

The LIMPID, or Large-Scale Image Processing Infrastructure Development project aims to help scientists around the world to share information and advance the reproducibility of methods in the field of computer vision – or how computers ‘see’ and analyse images. The LIMPID project has applications in fields as diverse as materials science, biology, neuroscience, marine science, remote sensing, and medicine. The related BisQue ecosystem is an online platform that allows researchers to upload software packages and data to carry out analysis on the cloud, enabling easy sharing of results between researchers at different institutions.

Researchers at the Materials Science Department at the University of California, Santa Barbara are using computational techniques to analyse microstructure images in order to predict material properties. The development of techniques enabling computers to ‘see’ and understand the content of digital images or videos is known as computer vision. The scientists are using BisQue, an online infrastructure for sharing data and analysis techniques across scientific disciplines in order to improve the reproducibility of computer vision techniques.

SHARING COMPUTER VISION TECHNIQUES ACROSS SCIENTIFIC DISCIPLINES

Materials science is just one example of a discipline in which computer vision is important. Modern scientific research involves the production of data on a vast scale, much of it in the form of images: scientific instrumentation probes our universe at scales varying from the galactic to the subatomic. The resulting images, produced by a wide array of detectors, are processed

using various computational techniques in order to make sense of the data. Data science is therefore a growing field, and there is a need for new techniques and infrastructures for the analysis of big imaging data.

The LIMPID (Large-Scale Image Processing Infrastructure Development) project aims to help scientists across disciplines to share information about image analysis. Currently, image analysis methods are being developed by individual research groups all over the world working in fields as diverse as materials science, biology, neuroscience and brain connections, marine science, remote sensing, and medicine. The LIMPID project transforms this way of working by creating an extensive and unique resource for the curation, distribution and sharing of scientific image analysis methods. Sharing expertise and methods across disciplines on a mass scale has the potential to substantially speed up working practices and reduce the risk of scientists ‘reinventing the wheel’ each time they have the need to use a method that is new to them.

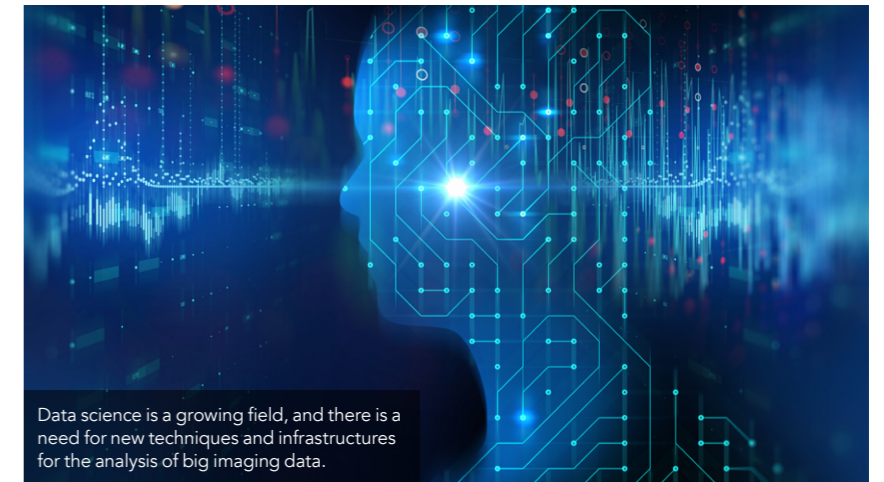
Modern scientific research involves the production of data on a vast scale, much of it in the form of images.

A CLOUD-BASED SYSTEM FOR SHARING AND TESTING TECHNIQUES

BisQue is a novel ecosystem where scientific image analysis methods can be discovered, tested, verified, refined and shared amongst users on a shared, cloud-based infrastructure. With BisQue, it is possible to carry out data processing on the cloud, using a web browser to interact with the datasets. BisQue assigns a unique URL to each execution, dataset (input and output) and workflow processing pipeline. Since it is always clear exactly where input data is located and which processes have been performed on the data to create output data, these URLs help researchers to track and share data even when they are collaborating with colleagues from other institutions located around the world.

MACHINE LEARNING

Recently, scientists working in the field of computer vision have started to use deep learning-based pattern recognition techniques, with highly promising results. BisQue’s modular architecture enables easy integration of evolving machine learning architectures for image analysis. For example, the connoisseur service in BisQue is an integrated training and classification system for image recognition based on deep learning. It uses convolution neural networks –



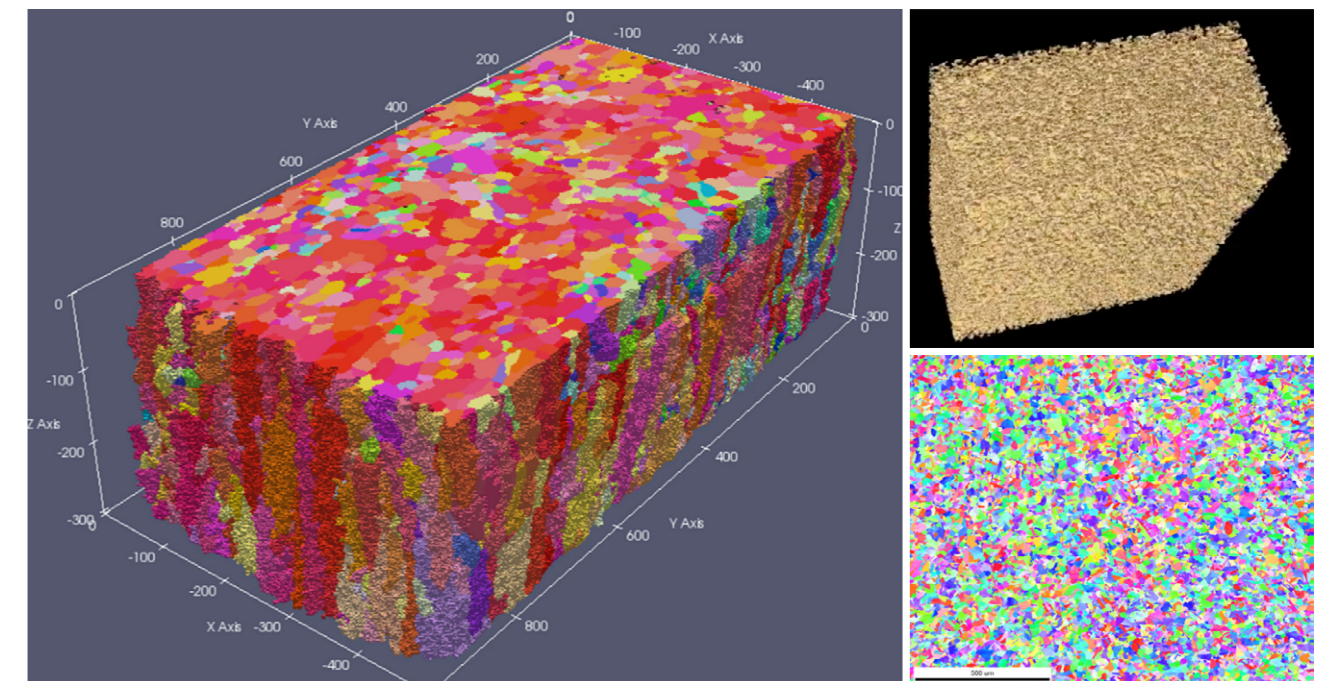
Data science is a growing field, and there is a need for new techniques and infrastructures for the analysis of big imaging data.

a deep learning technique used to find the correct mathematical manipulation to turn an input into an output – to create a model directly from the images. This allows scientists to create a so-called classifier model (for example, to identify species in the ocean), from input files, direct from annotated still or video image files without knowing details of how the algorithm is designed. A test is carried out automatically using a different dataset from that used for the ‘training’ stage in which the algorithm was created. The researcher can then choose to discard certain ‘classes’ (e.g. types of biological organism) if the algorithm did not perform adequately in those cases. When the algorithm is used to classify

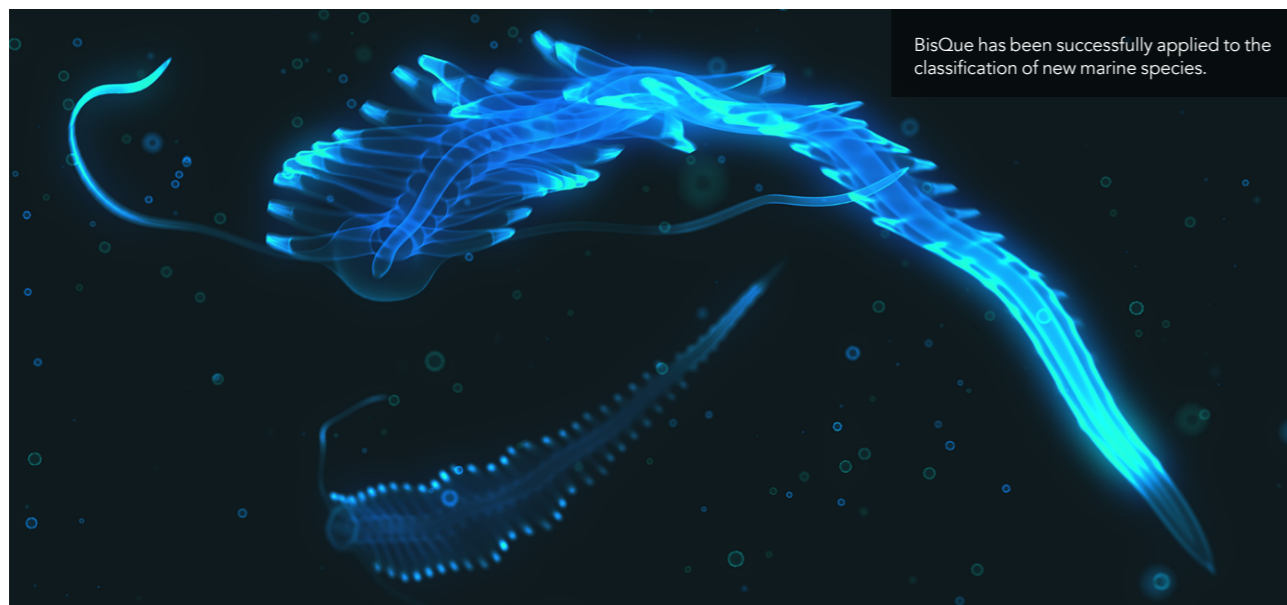
image data (e.g. to identify the species in the files being analysed), each sample is given a confidence score, and the user can choose to skip low confidence samples. The scientist can validate or change the classification provided by the algorithm, and this new information can be fed back into the model to improve its classifications in the future.

USING BISQUE FOR THE ANALYSIS OF MATERIALS

A current challenge in the field of materials science is the rapid prediction of how the internal structure of materials affects their engineering properties. This requires expensive experiments or complex models and computationally



Large-scale 3D materials data reconstructed with BisQue-DREAM.3D workflows.



BisQue has been successfully applied to the classification of new marine species.

demanding simulations. One aim of the researchers using BisQue is to develop a cloud-based module to predict material properties using 3D datasets and precomputed 3D physics-based simulations. The researchers developed a module for predicting the properties of two-phase composite materials – materials composed of two components which remain distinct in the composite material's structure, often arranged in complex topological patterns.

A software package known as DREAM.3D (Groeber, 2014) has been developed to analyse materials using experimental 3D data. It brings together tools and algorithms developed by the community for analysis of materials science images gathered in various modes in the scanning electron microscope. The processing steps are organised into a pipeline typically containing well over a dozen steps. This software was previously run on a desktop with the researcher manually changing input parameters: this way of working means it is only practical to explore a relatively small

range of variation in input parameters, potentially introducing bias. However, when DREAM.3D was integrated with BisQue, the researchers gained the ability to input a range of values for each parameter. Many instances of the programme are then run simultaneously using the different input parameters, saving time in producing results and improving reconstruction quality. The output data can also be viewed and visualised through the BisQue platform, so there is no need for researchers to run a separate computationally intensive visualisation programme. In a separate module, researchers also implemented a computationally efficient algorithm to test for a measure of material strength. This module is available online and can be run from a web browser with minimal computational requirements on the user end.

CLASSIFICATION OF UNDERWATER IMAGES

Another application of BisQue is in marine science. The researchers at the Marine Science Institute, Santa Barbara use user-

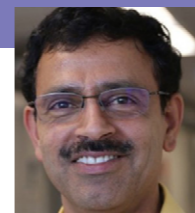
defined annotations to classify species. The marine species being classified are defined by the user, and can be changed over time. In one study, scientists analysed a dataset of over a thousand underwater images. The BisQue integrated deep learning technique classified the sessile species with very high accuracy.

THE BENEFITS OF BISQUE FOR THE SCIENTIFIC COMMUNITY

The initial motivation for BisQue came from the life sciences, and as detailed above, there are applications in materials science, marine science, medical imaging and more. The platform is suitable for most scientific fields that rely on image analysis based on data from cameras, microscopes, and other capture devices. The use of cutting-edge machine learning technology allows accurate data analysis without scientists in the relevant fields needing to have a huge amount of expertise in machine learning algorithm development. Employing web-based analytics and cloud-based computing allows users to implement the latest methods easily. The cloud-based infrastructure is perfectly suited to encourage and support collaboration between scientists, and notably, encourages the faster development and deployment of new analysis tools. Novel methodology can also be widely disseminated, supporting research in many areas of science. An active community of users and developers is being fostered through workshops, scientific meetings, and summer research internships.

The cloud-based infrastructure is perfectly suited to encourage and support collaboration between scientists, and notably, encourages the faster development and deployment of new analysis tools.

Behind the Research



Professor B.S. Manjunath



Professor Tresa M. Pollock

E: manj@ucsb.edu T: +1 (805) 893 7112 W: <https://bigdatahealth.ucsb.edu>

Research Objectives

BisQue is an advanced image database and analysis system for 5D biological images. It supports large scale image databases, flexible experimental data management, metadata and content based search, analysis integration and knowledge discovery.

Detail

B.S. Manjunath, Professor
Department of Electrical and Computer Engineering
University of California, Santa Barbara
CA 93106-9560

Bio

B.S. Manjunath is a distinguished Professor of Electrical and Computer Engineering at the University of California, Santa Barbara. His research is in the broad areas of image processing, computer vision and machine learning. He is a fellow of the IEEE and a fellow of the ACM.

Professor Tresa Pollock is the Alcoa Distinguished Professor of Materials at the University of California, Santa Barbara. Her research focuses on alloy design, materials in extreme environments and 3-D materials characterisation. She was elected to the NAE in 2005, German Academy of Sciences Leopoldina in 2015, is a DOD Vannevar Bush Fellow and Fellow of TMS and ASM.

Funding

US National Science Foundation through grants #1664172, #1650972, and #1356750

Collaborators

- Nirav Merchant, Director of CyVerse, U. Arizona
- Amit Roy-Chowdhury, University of California, Riverside
- Robert Miller, Marine Science Institute, UCSB

References

- Latypov, M. I., Khan, A., Lang, C. A., Kvilekval, K., Polonsky, A. T., Echlin, M. P., Beyerlein, I. J., Manjunath, B.S & Pollock, T. M. (2019). BisQue for 3D materials science in the cloud: microstructure–property linkages. *Integrating Materials and Manufacturing Innovation*, 1-14. DOI: [10.1007/s40192-019-00128-5](https://doi.org/10.1007/s40192-019-00128-5)
- Fedorov, Dmitry, Manjunath, BS, Lang, Christian, & Kvilekval, Kristian, G. (2017). Scalable Image Informatics. In Academic Press Library in Signal Processing, Volume 6, Chapter 9. <https://escholarship.org/uc/item/7nf7492x>
- Polonsky, A.T., Lang, C.A., Kvilekval, K.G., Latypov, M.I., Echlin, M.P., Manjunath, B.S., Pollock, T. M. (2019). Three-dimensional Analysis and Reconstruction of Additively Manufactured Materials in the Cloud-Based BisQue Infrastructure. *Integr Mater Manuf Innov*. 8: 37. <https://doi.org/10.1007/s40192-019-00126-7>
- Groeber, M. A.; Jackson, M. A. (2014). DREAM.3D: A Digital Representation Environment for the Analysis of Microstructure in 3D. *Integr. Mater. Manuf. Innov.* 3 (1), 5. <http://www.immijournal.com/content/3/1/5>

Personal Response

What are some of the key future research questions that you think could be addressed by researchers using the BisQue ecosystem for image analysis?

Professor B.S. Manjunath: Recent advances in computer vision and machine learning are having a transformative impact on a wide range of consumer and scientific applications. The BisQue ecosystem enables researchers to easily share and collaborate with their data, leveraging the cloud computing resources efficiently.

Professor Tresa Pollock: The BisQue ecosystem enables automated workflows for multimodal data, sharing among research communities and will provide new pathways for the design of advanced materials and tools for predicting their performance. **”**