www.gperezs.com gperezsarabi@umass.edu

Research Statement

Gustavo Perez

Overview: From data to science with AI and human-in-the-loop

My research in *computer vision* and *machine learning* focuses on finding effective ways to combine human and computational effort, and statistical estimation to reduce the cost of deriving scientific outcomes. My work focuses on analyzing data from novel scientific domains where data collection and labeling are expensive and require expertise (Fig. 1a). I approach this problem from three complementary views: the first one is when *transfer learning* could be helpful, but the application of pretrained ImageNet networks is not straightforward as the data might have more than three channels, such as hyperspectral imagery. Second, when transfer learning is not practical but a lot of unlabeled data is available. Third, when you have an imperfect model that requires screening to obtain high-confidence results due to domain shift, model calibration, or the difficulty of the task.

My work is also interdisciplinary, where I collaborate with ecologists to study phenology and population declines on bird species using RADAR imagery [8, 2, 1], with astronomers to provide insights into the process of star formation and the birth and evolution of galaxies from the *Hubble Space Telescope* and the *James Webb Space Telescope* data [7, 6, 3], and with chemists to predict absorption of nanoporous materials [4]. The findings of these collaborations have been published in leading journals in astronomy [6, 3], chemistry [4], and ecology [2, 1].

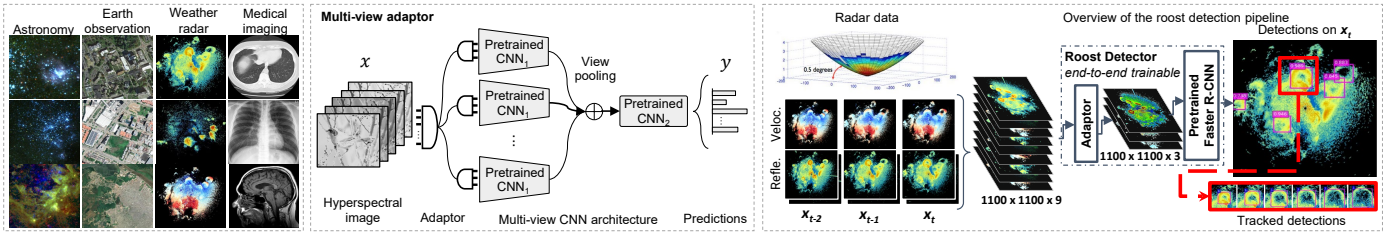
My research goal is to create data-efficient solutions to enable fundamental scientific advances and the understanding of the world and the universe to help improve our lives. In the following sections I will introduce my work in the three directions mentioned above and summarize my future research plans.

Improving knowledge transfer to novel domains

Transferring deep networks trained on large datasets of color images has been a key to their success in visual recognition. Also, because of its simplicity, *transfer learning* is still one of the most used approaches when dealing with domain-specific data by non-computer scientists. However, a challenge arises when transferring to heterogeneous domains where some architectural modification to the network is necessary for it to process the input.

In **hyperspectral domain adaptors** [5], we consider the problem of adapting a network trained on three-channel color images to hyperspectral images with a large number of channels. We propose adaptors that map the input to be compatible with a network trained on large-scale color image datasets such as ImageNet, to enable learning on small hyperspectral datasets where training a network from scratch may not be effective. We investigate architectures and strategies for training adaptors and evaluate them on a proposed benchmark with multiple datasets. We propose a novel multi-view scheme that generates a prediction by aggregating information across different *views* of the input (Fig. 1b). We find that ImageNet pretrained networks with domain adaptors enable efficient learning from a few examples. While simple schemes like linear adaptors are effective, our proposed multi-view variants lead to better results while adding a negligible number of parameters to the model.

In **spatio-temporal roost detector** [8], we study the generalizability of our domain adaptors when using additional modalities and temporal information to detect and track swallow roosts in weather radar data with a pretrained Faster R-CNN (Fig. 1c). The addition of temporal information improves roost detection performance by $\sim 8\%$ mean average precision and provides research-quality data with far less human effort than manual annotation of radar scans. We use this work in collaboration with ecologists from Colorado State University to quantify long-term phenological patterns of aerial insectivores roosting [2] and to perform long-term analysis of the persistence and size of swallow and martin roosts [1], which represent one of the longest-term broad-scale phenology examinations of avian aerial insectivore species responding to environmental change and set grounds for future monitoring of changes in these patterns as they may have important ecosystem and conservation implications.



(a) Novel domains examples (b) Multi-view domain adaptor (c) Spatio-temporal roost detector

Figure 1: **Knowledge transfer to novel domains**. (a) Samples of color/grayscale renditions of *novel domain* images in astronomy, earth observation, weather radar, and medical imaging. (b) Multiple adaptors generate three-channel views of the input which are processed through a shared network. In (c), we use an adaptor to add additional modalities and temporal information to a pretrained Faster R-CNN to detect roosts in radar data.

Representation learning with limited data and human-in-the-loop

In this work [7], we focus on AI tools for *cataloging* bright sources within galaxies, and use them to analyze *young star clusters* – groups of stars held together by their gravitational fields. Their ages and masses, among other properties provide insights into the process of star formation and the birth and evolution of galaxies. Significant domain expertise and resources are required to discriminate star clusters among tens of thousands of sources that may be extracted for each galaxy. To accelerate this step we propose a web-based annotation tool to label and visualize high-resolution astronomy data, encouraging efficient labeling and consensus building (Fig. 2-bottom), and techniques to reduce the annotation cost by leveraging recent advances in unsupervised representation learning on images (Fig. 2-top). We present case studies where we work with astronomy researchers from UMass Amherst and Stockholm University to validate the annotation tool and find that the proposed tools can reduce the annotation effort by 3× on existing catalogs, while facilitating accelerated analysis of new data. We use as baseline our star cluster classifier StarcNet [6], which we further adopted to analyze star cluster formation in galaxy M101 [3] where StarcNet is able to reproduce the human classifications at high levels of accuracy.

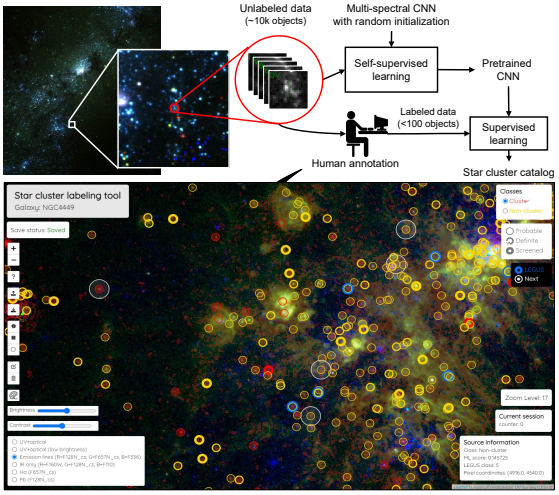


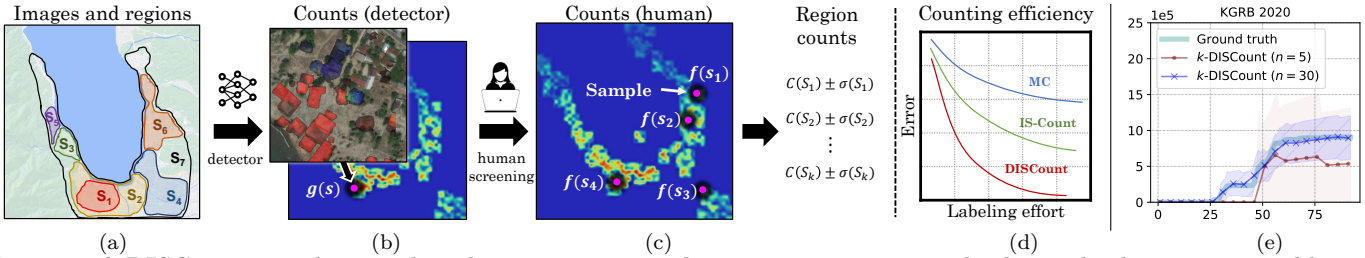
Figure 2: The proposed AI-assisted tool for cataloging sources within galaxies.

Counting in large image collections using statistical estimation

Many modern applications use computer vision to detect and count objects in massive image collections. For example, we are interested in applications that involve counting bird roosts in radar images and damaged buildings in satellite images. The image collections are too massive for humans to solve these tasks in the available time. So, a common approach is to train a computer vision detection model and run it exhaustively on the images.

The task is interesting because the goal is not to generalize, but to achieve the scientific counting goal with sufficient accuracy for a *fixed* image collection. The best use of human effort is unclear: it could be used for model development, labeling training data, or even directly solving the counting task! A particular challenge occurs when the detection task is very difficult, so the accuracy of counts made on the entire collection is questionable even with huge investments in training data and model development. Some works resort to human screening of the detector outputs, which saves time compared to manual counting but is still very labor-intensive. These considerations motivate *statistical* approaches to counting. Instead of screening the detector outputs for all images, a human can "spot-check" some images to estimate accuracy, and use statistical techniques to obtain unbiased estimates of counts across unscreened images.

In **DISCount** [9], we propose a detector-based importance sampling framework for counting in large image collections that integrates an imperfect detector with human-in-the-loop screening to produce unbiased estimates of counts (Fig. 3). We propose techniques for solving counting problems over multiple spatial or temporal regions using a small number of screened samples and estimate confidence intervals. This enables end-users to stop screening when estimates are sufficiently accurate, which is often the goal in a scientific study. Also, we develop variance reduction techniques based on control variates and prove the (conditional) unbiasedness of the estimators. DISCount leads to a 9-12× reduction in the labeling costs over naive screening for tasks we consider, such as counting birds in radar imagery or estimating damaged buildings in satellite imagery, and also surpasses alternative covariate-based screening approaches in efficiency.



(a) (b) (c) (d) (e)

Figure 3: ***k*-DISCount** uses detector-based importance sampling to screen counts and solve multiple counting problems. (a) Geographical regions where we want to estimate counts of damaged buildings. (b) Outputs of a damaged building detector on satellite imagery, which can be used to estimate counts $g(s)$ for each tile (shows as dots). (c) Tiles selected for human screening to obtain true counts $f(s)$, from which counts for all regions are joinly estimated by k-DISCount. (d) Our experiments show that DISCount outperforms covariate-based sampling. (e) Bird count estimates with confidence intervals for KGRB radar station in 2020 using k-DISCount.

Future work

My research has been directed toward reducing the cost of producing science and is heavily motivated by the broader impact potential of interdisciplinary research. I have developed and contributed various algorithms to bridge the gap between core computer vision and applications, focusing on novel domains where annotating data is expensive and requires domain expertise. My research is also focused on combining AI with human-in-the-loop, working closely with domain-specific researchers in several disciplines to develop tools able to give insights to open scientific questions. I plan to build upon my research by exploring these directions:

Statistical methods for robust estimation in counting tasks can be applied to guide the screening process in other applications where the goal is to collect measurements. For example, although my work on roost detection in radar images has enabled bird population studies with important ecosystem and conservation implications on the Great Lakes region in the US, scaling to larger regions might be prohibitively expensive. Therefore, we can use count estimation to produce unbiased estimates with confidence intervals to scale to the entire US RADAR archive ($\sim$ 30 years from 142+ stations). Similarly, it can be applied to other ecology applications like tracking the migration of birds on radar imagery, mammals in camera traps, and fish in sonar data. Or in other non-ecology-related applications like damaged building counting to assess damage in regions struck by natural disasters. In addition, robust counting estimation may lead to answering other related questions, such as if human screening of outputs is required for a specific task, or what are effective strategies for providing human annotation and model training to analyze a very large image data set.

Efficient model testing and test-time adaptation For various applications the goal is still the object localization. One approach in this direction is finding ways to use counting estimates as weak supervision to improve the object detector. This can be applicable when we have a pretrained detector that performs unreliably and a limited budget of labeled data to adapt the model to the new domain. Even for transfer learning from a pretrained network to a new domain, a minimum amount of labeled samples is still required for the model to improve. When this requirement is not met, we can use per-batch count estimates to learn features from the entire set of unlabeled data using a few labeled images.

In addition, when designing data-efficient methods, we often assume a large labeled dataset for testing. We can use large labeled test sets in artificial research settings to evaluate how good a method is, but in practice, this poses a big issue for real scientific applications where experts can require even weeks to acquire a single label. One question in this direction is how to define acquisition functions to pick the test samples from (e.g., using importance sampling) to maximize the accuracy of our evaluation metric estimate in a sample-efficient way.

Role of large language and vision models Could large vision-language models enable faster ways to analyze data for sciences? For instance, through contrastive pretraining and zero-shot image classification using language-based training. Would these models transfer better than imageNet pretraining for heterogeneous domains?

Interdisciplinary collaborations are an important part of my research: We are currently working with our collaborators from Colorado State University and the University of Oklahoma on using our work on roosting birds detection and counting [8, 2, 1, 9] to scale to the entire US RADAR archive. We are also working on analyzing recently released imagery from the JWST using our work with our astronomy collaborators from UMass Amherst and Stockholm University [6, 3, 7]. In addition, we are collaborating with the Red Cross and the Center for Data Science at UMass Amherst to deploy a tool for estimating damage after a natural disaster to speed up the deployment of help and resources to an affected area using our work [9]. I plan to continue fostering interdisciplinary collaborations and advocating for the integration of AI across diverse scientific domains.

References

[1] M. Belotti, Y. Deng, W. Zhao, V. Simons, Z. Cheng, **Perez, G.**, E. Tielens, S. Maji, D. Sheldon, J. Kelly, and K. Horton. Long-term analysis of persistence and size of swallow and martin roosts in the us great lakes. *Remote Sensing in Ecology and Conservation*, 2023.

[2] Y. Deng, M. Belotti, W. Zhao, Z. Cheng, **Perez, G.**., E. Tielens, V. Simons, D. Sheldon, S. Maji, J. Kelly, and K. Horton. Quantifying long-term phenological patterns of aerial insectivores roosting in the great lakes region using weather surveillance radar. *Global Change Biology*, 2022.

[3] S. Linden, **Perez, G.**, D. Calzetti, S. Maji, M. Messa, B. Whitmore, R. Chandar, A. Adamo, K. Grasha, D. Cook, B. Elmegreen, D. Dale, E. Sacchi, E. Sabbi, E. Grebel, and L. Smith. Star cluster formation and evolution in M101: An investigation with the legacy extragalactic UV survey. *The Astrophysical Journal*, 2022.

[4] Y. Liu, **Perez, G.**, Z. Cheng, A. Sun, S. Hoover, W. Fan, S. Maji, and P. Bai. Zeonet: 3d convolutional neural networks for predicting adsorption in nanoporous zeolites. *Journal of Materials Chemistry A*, 2023.

[5] **Perez, G.** and S. Maji. Domain adaptors for hyperspectral images. In *ICPR*, 2022.

[6] **Perez, G.**, M. Messa, D. Calzetti, S. Maji, D. Jung, A. Adamo, and M. Sirressi. StarcNet: Machine learning for star cluster identification. *The Astrophysical Journal*, 2021.

[7] **Perez, G.**, S. Linden, T. McQuaid, M. Messa, D. Calzetti, and S. Maji. An AI-assisted labeling tool for cataloging high-resolution images of galaxies. In *NeurIPS 2022 AI for Science: Progress and Promises*, 2022.

[8] **Perez, G.**, W. Zhao, Z. Cheng, M. Belotti, Y. Deng, V. Simons, E. Tielens, J. Kelly, K. Horton, S. Maji, and D. Sheldon. Using spatio-temporal information in weather radar data to detect and track communal bird roosts. *bioRxiv*, 2022.

[9] **Perez, G.**, S. Maji, and D. Sheldon. DISCount: counting in large image collections with detector-based importance sampling. In *arXiv:2306.03151*, 2023 (Under submission).