Snowmass2021 - Letter of Interest

Joint pixel-level processing of WFIRST, Euclid, LSST, and SPHEREx

Thematic Areas: (check all that apply □/■)

□ (CF1) Dark Matter: Particle Like
□ (CF2) Dark Matter: Wavelike
■ (CF3) Dark Matter: Cosmic Probes
■ (CF4) Dark Energy and Cosmic Acceleration: The Modern Universe
□ (CF5) Dark Energy and Cosmic Acceleration: Cosmic Dawn and Before
□ (CF6) Dark Energy and Cosmic Acceleration: Complementarity of Probes and New Facilities
■ (CF7) Cosmic Probes of Fundamental Physics
□ (Other) *[Please specify frontier/topical group]*

Contact Information:
Edward Schlafly (LLNL) [schlafly1@llnl.gov]

Authors:
Edward Schlafly (LLNL), Will Dawson (LLNL), Arjun Dey (NOIRLab), Dustin Lang (Perimeter), Aaron Meisner (NOIRLab), John Moustakas (Siena), Adam Myers (Wyoming), Peter Nugent (LBNL), David Schlegel (LBNL)

Abstract: At least four major cosmology-focused imaging surveys will begin in the next decade, from the *WFIRST*, *Euclid*, *LSST*, and SPHEREx projects. Much of the power of these surveys to constrain cosmology comes from measurements of weak lensing and large scale structure. They therefore rely critically on photometric redshifts. Joint processing of these surveys at the pixel level has the potential to provide much better spectral coverage and photometry, dramatically improving the accuracy and precision of photometric redshifts. These benefits arise primarily from exploiting the diverse spatial resolution, depth, and wavelength coverage of the different surveys to improve source modeling. We describe the joint survey processing pipeline currently used for the DESI Legacy Imaging Surveys in this context. We outline the chief improvements needed to enable this pipeline to jointly process images from *WFIRST*, *Euclid*, *LSST*, and SPHEREx. Further development of image processing capabilities of this kind should be a significant priority of the community.

At least four major cosmology-focused imaging surveys, from *WFIRST*, *Euclid*, *LSST*, and SPHEREx, will commence in the coming years. Much of these surveys' power to constrain cosmology comes from their ability to determine photometric redshifts. The surveys therefore depend on identifying galaxies and precisely measuring their photometry. The surveys cover a variety of wavelengths and spatial resolutions, with $\sim 0.2''$ optical imaging from *Euclid* and low resolution $\sim 7''$ infrared imaging from SPHEREx. Combining information from these surveys dramatically improves photometric redshifts due to the improved wavelength coverage. However, rigorously combining these surveys is hard due to their different spatial resolutions. Further, because of blending and the desire for upper limits on fluxes when sources are not detected in a particular survey, catalog-level combination of individual surveys is of limited use. This letter of interest describes a path toward joint, pixel-level processing of these surveys, from the perspective of the existing Legacy Survey imaging pipeline, the *Tractor*.

Recent works have repeatedly emphasized the importance of jointly processing imaging from *WFIRST*, *Euclid* and *LSST*, with particular focus on the improvement to the photometric redshifts that the combined projects can deliver (Jain et al., 2015; Capak et al., 2019). The work of Rhodes et al. (2017) shows that the photometric redshift accuracy of *LSST* improves by a factor of $2\times$ when combined with *Euclid* imaging for galaxies with $1.5 < z < 3$. Chary et al. (2020) provide a broad survey of the science projects that would be enabled by joint survey processing, together with an ambitious vision for the standardized products and science platform that would maximize the surveys' utility to the community. Motivated by the scientific returns described in those papers, we here describe the current status of the joint image processing performed in the DESI Legacy Imaging surveys, and the key developments required for a "next generation" joint survey analysis needed to model *WFIRST*, *Euclid*, *LSST*, and SPHEREx imaging.

Joint Processing of the DESI Legacy Imaging Surveys

The DESI Legacy Imaging Surveys ("Legacy Surveys") have jointly processed ground-based optical imaging and *WISE* infrared imaging to improve source deblending and to extend wavelength coverage (Dey et al., 2019). The Legacy Surveys comprise optical grz (0.4–1 μm) photometry from three instruments: the DECam camera at Cerro Tololo, and the 90prime and MOSAIC3 cameras at Kitt Peak. These surveys have a typical FWHM of about $1''$. The optical imaging is supplemented with infrared 3–5 μm imaging from *WISE*, with a typical FWHM of about $6''$.

To obtain accurate photometry to supply targets for spectroscopy with DESI (DESI Collaboration et al., 2016), the imaging from the Legacy Surveys is *jointly processed*. To be specific, the *Tractor* pipeline (Lang et al., 2016a) simultaneously models sources over each optical band and each optical epoch in that band. This allows the highest-resolution images to be used to inform deblending and shape measurement in all other images. This moreover leads to the use of consistent source profiles in each band and eliminates ambiguities between sources detected at different epochs or in different bands.

The *Tractor* approach has been extensively tested in operations. The *Tractor* was used to produce target catalogs for spectroscopy by eBOSS (eBOSS Collaboration et al., 2020; Raichoor et al., 2017) and is now providing targets for DESI (DESI Collaboration et al., 2016; Dey et al., 2019). Associated pixel-level image simulation codes allow a variety of imaging systematics to be forward-modeled (Kong et al., 2020). The *Tractor's* companion image browser, the *Viewer*, provides extensive data exploration capabilities and combines information from a number of major surveys at different wavelengths. We encourage anyone unfamiliar with the *Viewer's* functionality to browse the sky at http://legacysurvey.org/viewer.

At present, the *Tractor* approach does not simultaneously model the *WISE* imaging. Instead, the optical source positions and shapes are held fixed, and forced photometry is performed on the *WISE* imaging (Lang et al., 2016b). This allows the $\sim 6\times$ better optical resolution to inform the deblending of source configurations that would be ambiguous in the *WISE* imaging alone.

Key Improvements Needed for Next-Generation Joint Image Processing

A next-generation joint survey analysis would improve on the existing *Tractor* in at least three ways: treating all wavelengths analogously, scaling to large numbers of images, and better deblending images.

Treating all wavelengths analogously
The current *Tractor* analysis treats the high-resolution optical imaging preferentially, detecting and fitting sources at those wavelengths. Only forced photometry is performed on the lower-resolution *WISE* imaging. This approach fails to find sources detected only in *WISE* imaging. It also leads to extremely uncertain *WISE* fluxes for sources that are nearby but resolved in the optical—a problem that will become more acute in next-generation processing, where *Euclid* and *WFIRST* will have $3600\times$ smaller pixels than SPHEREx.

The remedy is to incorporate additional prior information about the possible types of astronomical spectral energy distributions (e.g., Regier et al., 2015). Presently the *Tractor* analysis imposes no priors on sources' spectral energy distributions. However, for example, given the dramatic differences between stellar colors and high-redshift galaxy colors, often flux at long wavelengths can be associated with an extended optical source rather than to a nearby pointlike neighbor with stellar colors. However, such priors must be applied carefully, so as to allow the possibility of discovering anomalous sources.

Scaling to larger numbers of images
The *Tractor* currently loads into memory and simultaneously analyzes every optical image overlapping the source of interest. Naively extrapolating this approach to *LSST* would require $\sim 100\times$ more compute time and memory. Moreover, this estimate ignores the greater number of sources that would be detected in the deeper *LSST* images, as well as the increased cost of modeling the more blended *LSST* images.

This increased computational cost would make joint processing expensive, but perhaps not prohibitive ($\sim$ 1 billion CPU hours). However, coadding images across different ranges of seeing, or adopting other types of principled coaddition (Zackay & Ofek, 2017a,b) would allow only a few stacked images to be analyzed, rather than hundreds of individual epochs, with little loss of information. This could reduce the computational expense in the *LSST* era by a factor of ~ 100. Developing rigorous and principled approaches to combine individual images at scale is a needed step towards joint survey-processing analysis.

Better deblending images
The *Tractor* pipeline presently analyzes each source in sequence, considering only the pixels deemed by an image segmentation algorithm to "belong" to a given source. This is clearly not optimal; the full footprint of a group of blended sources should be simultaneously modeled, and the corresponding covariant uncertainties reported. Relatively principled approaches for detecting and modeling blended sources exist when all sources are expected to be stellar and so the source profiles are all known. However, the galaxies that are the targets of these cosmology programs pose distinctly greater challenges—especially so in the case of galaxies with irregular morphologies where it becomes unclear how to define a "source" in the first place.

Despite this concern, progress has been made building more realistic models of galaxies (e.g. Lanusse et al., 2020), deblending images (Melchior et al., 2018), and modelling scenes with large numbers of sources (e.g. Portillo et al., 2017; Schlafly et al., 2018). Joint survey processing in particular is *less* affected by deblending problems than, for example, *LSST*-only image analysis, since joint processing can rely on higher-resolution *WFIRST* and *Euclid* imaging. Still, progress here remains crucial, given that most *LSST* galaxies will be at least somewhat blended with their neighbors Melchior et al. (2018), and $\sim 15\%$ of *LSST*-detected galaxies will not be recognizable as blends from the *LSST* imaging alone Dawson et al. (2016).

These three improvements to imaging analyses—treating all wavelengths analogously, improving scaling, and better deblending—strike us as the most important areas needing research and development to enable next-generation joint survey processing.

References

Capak, P., Cuillandre, J.-C., Bernardeau, F., et al. 2019, arXiv e-prints, arXiv:1904.10439

Chary, R., Helou, G., Brammer, G., et al. 2020, arXiv e-prints, arXiv:2008.10663

Dawson, W. A., Schneider, M. D., Tyson, J. A., & Jee, M. J. 2016, ApJ, 816, 11

DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016, arXiv e-prints, arXiv:1611.00036

Dey, A., Schlegel, D. J., Lang, D., et al. 2019, AJ, 157, 168

eBOSS Collaboration, Alam, S., Aubert, M., et al. 2020, arXiv e-prints, arXiv:2007.08991

Jain, B., Spergel, D., Bean, R., et al. 2015, arXiv e-prints, arXiv:1501.07897

Kong, H., Burleigh, K. J., Ross, A., et al. 2020, arXiv e-prints, arXiv:2007.08992

Lang, D., Hogg, D. W., & Mykytyn, D. 2016a, The Tractor: Probabilistic astronomical source detection and measurement, ascl:1604.008

Lang, D., Hogg, D. W., & Schlegel, D. J. 2016b, AJ, 151, 36

Lanusse, F., Mandelbaum, R., Ravanbakhsh, S., et al. 2020, arXiv e-prints, arXiv:2008.03833

Melchior, P., Moolekamp, F., Jerdee, M., et al. 2018, Astronomy and Computing, 24, 129

Portillo, S. K. N., Lee, B. C. G., Daylan, T., & Finkbeiner, D. P. 2017, AJ, 154, 132

Raichoor, A., Comparat, J., Delubac, T., et al. 2017, MNRAS, 471, 3955

Regier, J., Miller, A., McAuliffe, J., et al. 2015, arXiv e-prints, arXiv:1506.01351

Rhodes, J., Nichol, R. C., Aubourg, É., et al. 2017, ApJS, 233, 21

Schlafly, E. F., Green, G. M., Lang, D., et al. 2018, ApJS, 234, 39

Zackay, B., & Ofek, E. O. 2017a, ApJ, 836, 187

—. 2017b, ApJ, 836, 188