

---

# Spotting Hallucinations in Inverse Problems with Data-Driven Priors

---

Matt L. Sampson<sup>1</sup> Peter Melchior<sup>1,2</sup>

## Abstract

Hallucinations are an inescapable consequence of solving inverse problems with deep neural networks. The expressiveness of recent generative models is the reason why they can yield results far superior to conventional regularizers; it can also lead to realistic-looking but incorrect features, potentially undermining the trust in important aspects of the reconstruction. We present a practical and computationally efficient method to determine, which regions in the solutions of inverse problems with data-driven priors are prone to hallucinations. By computing the diagonal elements of the Fisher information matrix of the likelihood and the data-driven prior separately, we can flag regions where the information is prior-dominated. Our diagnostic can directly be compared to the reconstructed solutions and enables users to decide if measurements in such regions are robust for their application. Our method scales linearly with the number of parameters and is thus applicable in high-dimensional settings, allowing it to be rolled out broadly for the large-volume data products of future wide-field surveys.

## 1. Introduction

Inverse problems commonly arise in the sciences whenever causal factors need to be inferred from observations. In imaging applications one usually seeks to infer a high-quality image from corrupted data. Examples include denoising, deconvolution, inpainting, super-resolution, and, in astronomy, deblending problems.

Central to inverse problems is the (assumed) knowledge of the forward operator  $f : \mathcal{X} \rightarrow \mathcal{Y}$ , which maps the unknown parameters  $\mathbf{x}$  to idealized observations  $\mathbf{y}$ , which are then

<sup>\*</sup>Equal contribution <sup>1</sup>Department of Astrophysical Sciences, Princeton University, Princeton, USA <sup>2</sup>Center for Statistics and Machine Learning, Princeton University, Princeton, USA. Correspondence to: Matt L. Sampson <[mailto:matt.sampson@princeton.edu]>.

ICML 2023 Workshop on Machine Learning for Astrophysics, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

degraded by a lossy process to yield data  $\mathcal{D}$  as observed. Inverse-problem solvers seek to find the minimum  $\hat{\mathbf{x}}$  of a loss function  $L$ , typically the negative logarithm of the likelihood  $\mathcal{L}(\mathcal{D} | f, \mathbf{x})$  of the data given a model described by  $f(\mathbf{x})$ . In most cases, such a direct minimizer has undesirable properties, e.g. being noisier or less smooth than expected from a plausible solution. Regularizers are added to the loss function to promote desired properties such as smoothness or sparsity in some domain. Increasingly, these regularizers are replaced by neural networks, trained on high-quality examples of the distribution of acceptable solutions (Kamilov et al., 2022; Lanusse et al., 2019). Substantial progress has been made in the last few years in the field of generative neural networks that learn the distribution  $\mathcal{P}(\mathbf{x})$  of valid solutions (e.g. Kingma & Welling, 2013; Goodfellow et al., 2014; Song et al., 2020).

We can now optimize the Bayesian log-posterior

$$L(\mathbf{x}) = \log \mathcal{L}(\mathbf{x}) + \log \mathcal{P}(\mathbf{x}), \quad (1)$$

using a data-driven prior as regularizer. As  $L$  is now a combination of two terms, it is not obvious which aspects of the posterior minimizer  $\hat{\mathbf{x}}$  are actually determined by data, and which are filled in by the prior. As generative networks have become more expressive, they have been found to be prone to hallucinations, i.e. the tendency to create features that are plausible but do not originate from  $\mathcal{D}$  directly (Gottschling et al., 2020). In fact, *hallucinations are the intended mechanism to create realistic solutions beyond what can directly be inferred from the data* (Liu et al., 2007; Wang et al., 2014; Yu et al., 2018).

However, in many scientific studies it is imperative to know whether a feature is really present or merely hallucinated. Most evidently, in medical imaging the reconstruction of a volumetric model from CT scan data is used to identify the presence of tumors. Medical professionals would like to be sure of their diagnosis before recommending an invasive procedure (Bhadra et al., 2021). In astronomy, we seek to infer higher-quality measurements from lower-quality data. Hallucinations can give us the appearance of higher quality but may inadvertently bias our results towards features often seen in the training data (Schawinski et al., 2017).

With inverse problems continuing to be under-constrained, and neural network regularizers producing the so-far

strongest solvers, hallucinations will stay with us. Here we ask the question: Can we find out where they happen?

## 2. Methods

We assume that we have access to differentiable models of the forward process  $f$ , the likelihood  $\mathcal{L}$ , and the prior  $\mathcal{P}$ . First-order minimizers then perform gradient steps in the opposite direction of  $\nabla_{\mathbf{x}}L = \nabla_{\mathbf{x}}\log\mathcal{L} + \nabla_{\mathbf{x}}\log\mathcal{P}$ . The curvature of  $L$ , i.e. the Hessian matrix  $\mathbf{H}_{ij} = \frac{\partial^2 L}{\partial x_i \partial x_j}$  is used in second-order methods to determine the step sizes. In statistics, the negative Hessian is called the Fisher information matrix  $\mathbf{F} = -\mathbf{H}$ , so dubbed because it describes the statistical information about  $\mathbf{x}$  conveyed by  $L$ . Equipped with  $\mathbf{F}$ , we can recast our question: Where is  $\mathbf{F}$  dominated by the prior, as opposed to the likelihood? As long as the features of the solution are largely determined by  $\mathcal{L}$ , there are by definition no hallucinations.

We now seek to compute the Hessians of  $\log\mathcal{L}$  and  $\log\mathcal{P}$  at the minimizer of Equation 1 separately, or, if more convenient, the Hessians of  $L$  and one of the others, as the Fisher matrix is linear:

$$\mathbf{F} = \mathbf{F}_{\log\mathcal{L}} + \mathbf{F}_{\log\mathcal{P}}. \quad (2)$$

Unfortunately, calculating the full Hessian matrix is inefficient, with a scaling of  $\mathcal{O}(n^2)$ , where  $n$  is the dimension of  $\mathcal{X}$ . This makes computing  $\mathbf{H}$  intractable for most practical applications. A full marginalization to get pixel-level uncertainties would additionally require a matrix inversion.

But in imaging applications, the off-diagonal terms in the Hessian are typically small, and often confined to a few off-diagonal bands. We therefore make the simplification of computing the Hessian diagonal  $\mathbf{H}_D = \text{Diag}(\mathbf{H})$ . Doing so has two advantages: There is an efficient way to compute only the diagonals of the Hessian, which we describe below; and the result has the same shape as  $\mathbf{x}$ , which means that it has the form of an image that can be compared to the solution to indicate regions dominated by the prior.

We follow the approach of Hutchinson (1989); Yao et al. (2020) to calculate the approximated diagonal Hessian. Instead of computing the full Hessian, we make use of the Hessian-vector product (HVP), which can be computed efficiently with automatic differentiation. From the chain rule, we have the following equation for the HVP,

$$\frac{\partial \mathbf{g}^T \mathbf{z}}{\partial \mathbf{x}} = \frac{\partial \mathbf{g}^T}{\partial \mathbf{x}} \mathbf{z} - \mathbf{g} \frac{\partial \mathbf{z}}{\partial \mathbf{x}} = \frac{\partial \mathbf{g}^T}{\partial \mathbf{x}} \mathbf{z} = \mathbf{H} \cdot \mathbf{z} \quad (3)$$

where  $\mathbf{g} = \nabla_{\mathbf{x}}L$ , and  $\mathbf{z}$  is an arbitrary vector independent of  $\mathbf{x}$ , hence  $\partial \mathbf{z} / \partial \mathbf{x} = 0$ . Equation 3 evidently requires only  $\mathcal{O}(n)$  operations, which is critical for higher-dimensional problems. To compute the diagonal approximation to the

---

### Algorithm 1 Hessian diagonal approximation

---

```

function HESSIANDIAG( $f, \mathbf{x}, \epsilon$ )
   $\mathbf{g} = \text{jax.grad}(f)$ 
   $\text{hvp} = \text{jax.jvp}(\mathbf{g}, \mathbf{x})$ 
   $\mathbf{H} = \text{jnp.zeros}(\mathbf{x}.\text{shape})$ 
   $\mathbf{H}' = \text{jnp.zeros}(\mathbf{x}.\text{shape})$ 
  for  $i = 0 \dots$  do
     $\mathbf{z} \sim \text{Rademacher}(\mathbf{x}.\text{shape})$ 
     $\mathbf{H} = \mathbf{H} + (\mathbf{z} \odot \text{hvp}(\mathbf{z}))$ 
    if  $\|\mathbf{H}/(i+1) - \mathbf{H}'/i\| < \epsilon \|\mathbf{H}/(i+1)\|$  then
      return  $\mathbf{H}/(i+1)$ 
   $\mathbf{H}' = \mathbf{H}$ 

```

---

Hessian, we employ the method of Hutchinson (1989):

$$\mathbf{H}_D = \mathbb{E}(\mathbf{z} \odot (\mathbf{H} \cdot \mathbf{z})), \quad (4)$$

with  $\mathbf{z}$  being sampled from a Rademacher distribution. We show an implementation of the Hessian diagonal approximation in Algorithm 1, making use gradient and Jacobian-vector product routines in JAX.

Finally, we produce our hallucination score

$$\Delta \mathbf{F} = \text{Diag}(\mathbf{F}_{\log\mathcal{P}}) - \text{Diag}(\mathbf{F}_{\log\mathcal{L}}). \quad (5)$$

Equivalent forms, such as  $\Delta \mathbf{F} = 2 \text{Diag}(\mathbf{F}_{\log\mathcal{P}}) - \text{Diag}(\mathbf{F})$  can be chosen as well, e.g. when the Hessian of the log posterior has already been estimated during the optimization. The positive regions of  $\Delta \mathbf{F}$  are prior-dominated and are therefore prone to hallucinations. The user can then decide how much trust they should place in the reconstruction of features in these regions. An example from an inpainting problem in astronomy is shown as the image on the right-hand side of Figure 1.

## 3. Experiments

We demonstrate the capability of this method with a toy model of a galaxy reconstruction using the source deblending method SCARLET (Melchior et al., 2018), which computes a differentiable likelihood and performs proximal gradient descent to enforce regularization. We replaced these regularizers by a score-based diffusion model, which directly learns  $\nabla_{\mathbf{x}}\log\mathcal{P}$  from training data, and acts as an informative prior for the galaxy morphology distribution. We implement all models in JAX (Bradbury et al., 2018) and equinox (Kidger & Garcia, 2021) with the diffusion model based on the implementation from Song et al. (2020).

### 3.1. Data

The diffusion model was trained on data from the Subaru Hyper-Suprime Cam catalogue (Bosch et al., 2018). We extracted the existing SCARLET models (each representing

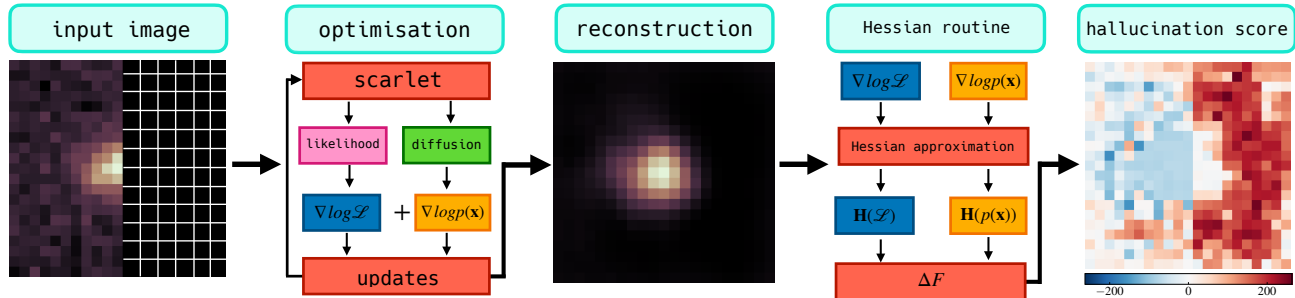


Figure 1. Overview of our methods to produce a reconstructed galaxy image and the corresponding hallucination score. The input image, of which half was masked to create an inpainting and denoising problem, is first modeled with SCARLET by gradient descent of the likelihood  $\nabla_{\mathbf{x}} \log \mathcal{L}$  and a data-driven prior in the form of a score model  $\nabla_{\mathbf{x}} \log \mathcal{P}$ . Next, we calculate the diagonalized Hessian matrices for both  $\log \mathcal{L}$  and  $\log \mathcal{P}$  via Algorithm 1. We then compute the hallucination score according to Equation 5.

a single, isolated galaxy source) for three tracts, yielding about 600,000 examples. Doing so exploits that these models have already been deblended and deconvolved and can therefore act as examples of the true distribution of galaxy shapes. We trained the diffusion model on a single NVIDIA A100 for 1, 000, 000 steps with a batch size of 256.

### 3.2. Quality of the score model

We show a test of the score model and its utility for suppressing image features that are inconsistent with galaxy shapes. Figure 2 shows an input image with and without a ring-shaped artifact. The bottom row shows the corresponding prior score. While the original galaxy image has an overall low amplitude prior score without clear spatial structure, the artifact is strongly suppressed by the prior gradients.

While we train a time-dependent score model as in Song et al. (2020), we evaluate the score at temperature  $T = 0$  during the optimization and for computing the hallucination score. Longer runtimes of diffusion models would not allow us to scale the prior evaluations to the data volumes expected for the Vera C. Rubin Observatory (Ivezić et al., 2019). By comparing scores from the  $T = 0$  limit with those from full diffusion, we have confirmed that our data distribution is simple enough that the former performs sufficiently well for our purposes. When we targeting larger or better resolved galaxies, we will need to reinvestigate this approximation.

### 3.3. Hallucination score

We assume we have a usable generative model of galaxy morphologies, which we now apply to solve an inverse problem to see where prior and likelihood dominate the reconstruction, respectively. We take a random sample of a galaxy observation from the HSC data and model it with SCARLET2. For simplicity, we remove the right half of the

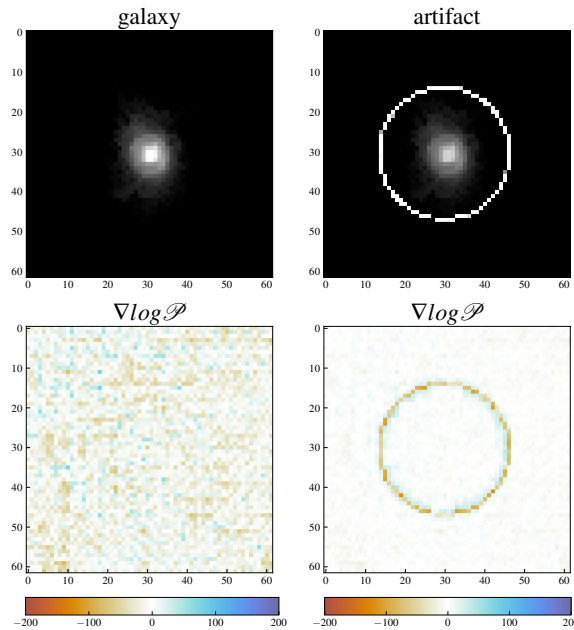


Figure 2. Example galaxy sample from the HSC catalog (top left), with a ring artifact (top right), and the calculated prior score for both cases (bottom row). The artifact is strongly suppressed in the prior gradients.

data (array values and weights set to 0), as this will enforce the prior to generate the entirety of the features on this half of the reconstructed image. In the second trial, we perform the same reconstruction on the unaltered image. The top row of Figure 3 shows the results of trial 1, and the bottom row shows trial 2. We can see from the reconstruction that the noise in the image is removed, and that by virtue of the prior we get a reasonable estimate of the right half of the galaxy shape as well. The next panels show the Hessian diagonals  $\mathbf{H}_D$  for  $\log \mathcal{L}$  and  $\log \mathcal{P}$ , both calculated with Algorithm 1, allowing us to then compute the hallucination

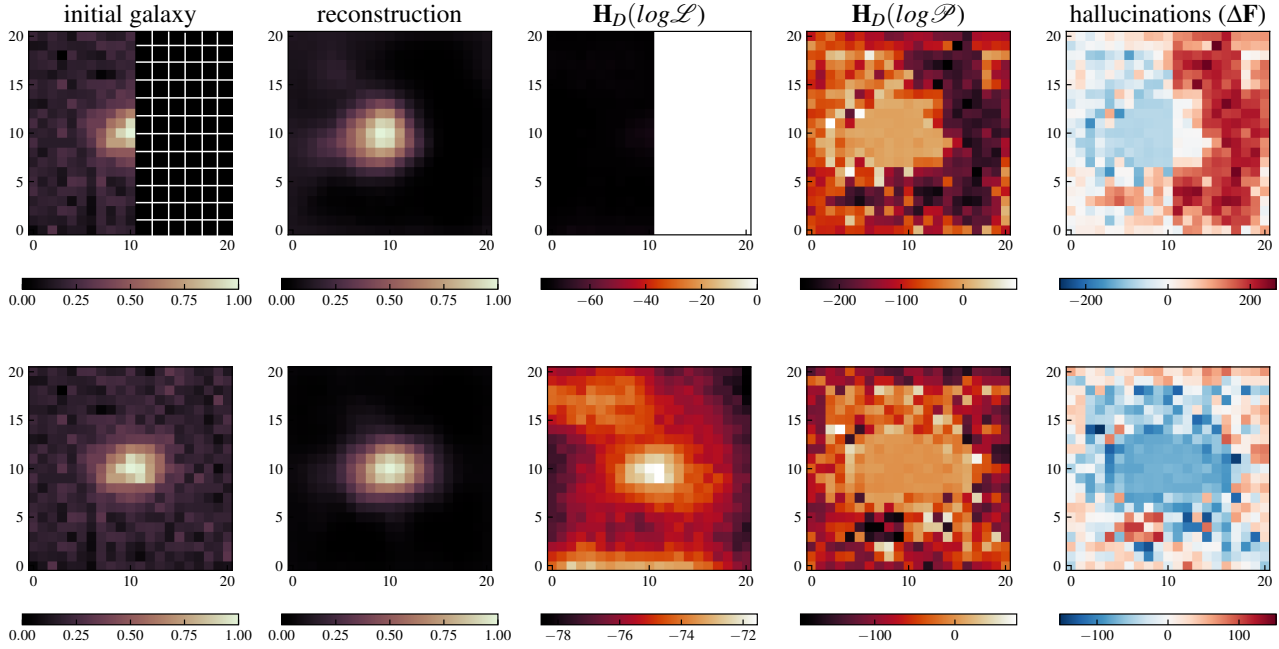


Figure 3. In the first two panels we show the initial input image for which the right half of the data has been set to 0, and the reconstructed image. Panels 3 and 4 show the Hessian diagonal for  $\log \mathcal{L}$  and  $\log \mathcal{P}$ , respectively. We note the features of  $\mathbf{H}_D(\log \mathcal{L})$  come directly from the variance weighting in the initial HSC data for the galaxy. In the rightmost, panel we show the hallucination score  $\Delta \mathbf{F}$  from Equation 5. The red shading indicates regions dominated by the prior, whereas the blue shading shows regions dominated by the likelihood.

score from Equation 5. While the left side of the  $\Delta \mathbf{F}$  shows that the information comes from the likelihood, the majority of the right side is dominated by the prior, as expected due to the absence of valid data. It is noteworthy that the hallucination score is much weaker in the central region of the right side than in the outskirts. This is likely attributed to the training data consisting of primarily relatively small galaxies, leading to a high confidence in the pixel values for the outskirts of the galaxy source: they are likely very close to 0. In the inner region, the score model is much less confident, so the hallucination score is closer to 0. We also note that for  $\mathbf{H}_D(\log \mathcal{P})$  we evaluate the Jacobian of the score model, which is in itself only an approximation of the true prior gradients; any inaccuracies of the score model will be amplified when computing another derivative.

The bottom row of Figure 3 shows the same results run on the unaltered galaxy, where we now see that the likelihood dominates the central region with the highest signal-to-noise ratio, while the prior starts to dominate in the outskirts.

The calculations for the Hessian diagonals in the example of Figure 3 took  $\approx 1$  ms for the likelihood, which converged after a single iteration, and  $\approx 260$  ms for the prior, which took on average 80 iterations to converge, resulting in roughly 3 ms per single HVP evaluation. With just-in-time compilation, the JVP of the score network is only a factor 3

slower than the HVP of the simple Gaussian likelihood of this example. All timing tests were run on an M1 Macbook Pro, utilizing 4 CPU cores.

## 4. Conclusion

We presented a practical and computationally efficient method to determine which regions in the solutions of inverse problems with data-driven priors are prone to hallucinations. By computing the diagonal elements of the Fisher information matrix of the likelihood and the prior separately, we can flag regions where the information is prior-dominated. Our diagnostic can directly be compared to the reconstructed solutions and enables users to make informed decisions about the trustworthiness of relevant features in the reconstruction. Our method scales linearly with the number of parameters and is thus scalable to high-dimensional settings, allowing it to be rolled out broadly for the large-volume data products of future wide-field surveys.

The choice of Equation 5 as a hallucination metric has advantages over simpler diagnostics such as directly calculating the standard deviation of the posterior. Doing so cannot differentiate the source of the information that determines the optimized model. Alternatively, merely checking the gradients of the likelihood and prior will become meaning-

less once the model is converged because they need to be either very small or cancel each other. While our method does not make assumptions about the prior model, computing gradients of data-driven priors requires a high level of fidelity of that model. Caution should be taken to ensure that the prior model is of sufficient accuracy for this purpose.

## Software and Data

We have used `python` (Van Rossum & Drake Jr, 1995) with the packages `JAX` (Bradbury et al., 2018), `equinox` (Kidger & Garcia, 2021), `numpy` (Harris et al., 2020), and `matplotlib` (Hunter, 2007). Data for this project is taken from the Subaru Hyper-Suprime Cam Survey (Bosch et al., 2018).

## Acknowledgements

We thank the reviewers for their helpful comments in improving this manuscript. MLS acknowledges financial support from the Princeton University First-Year Fellowship in the Natural Sciences and Engineering.

## References

- Bhadra, S., Kelkar, V. A., Brooks, F. J., and Anastasio, M. A. On hallucinations in tomographic image reconstruction. *IEEE transactions on medical imaging*, 40(11): 3249–3260, November 2021. ISSN 0278-0062, 1558-254X. doi: 10.1109/TMI.2021.3077857. URL <http://dx.doi.org/10.1109/TMI.2021.3077857>.
- Bosch, J., Armstrong, R., Bickerton, S., Furusawa, H., Ikeda, H., Koike, M., Lupton, R., Mineo, S., Price, P., Takata, T., et al. The hyper suprime-cam software pipeline. *Publications of the Astronomical Society of Japan*, 70(SP1): S5, 2018.
- Bradbury, J., Frostig, R., Hawkins, P., Johnson, M. J., Leary, C., Maclaurin, D., Necula, G., Paszke, A., VanderPlas, J., Wanderman-Milne, S., and Zhang, Q. JAX: composable transformations of Python+NumPy programs, 2018. URL <http://github.com/google/jax>.
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative adversarial networks. *arXiv e-prints*, June 2014. URL <http://arxiv.org/abs/1406.2661>.
- Gottschling, N. M., Antun, V., Hansen, A. C., and Adcock, B. The troublesome kernel – on hallucinations, no free lunches and the accuracy-stability trade-off in inverse problems. *arXiv e-prints*, January 2020. URL <http://arxiv.org/abs/2001.01258>.
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020. doi: 10.1038/s41586-020-2649-2. URL <https://doi.org/10.1038/s41586-020-2649-2>.
- Hunter, J. D. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007. doi: 10.1109/MCSE.2007.55.
- Hutchinson, M. F. A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines. *Communications in Statistics-Simulation and Computation*, 18(3):1059–1076, 1989.
- Ivezić, Ž., Kahn, S. M., Anthony Tyson, J., Abel, B., Acosta, E., Allsman, R., Alonso, D., AlSayyad, Y., Anderson, S. F., Andrew, J., Angel, J. R. P., Angeli, G. Z., Ansari, R., Antilogus, P., Araujo, C., Armstrong, R., Arndt, K. T., Astier, P., Aubourg, É., Auza, N., Axelrod, T. S., Bard, D. J., Barr, J. D., Barrau, A., Bartlett, J. G., Bauer, A. E., Bauman, B. J., Baumont, S., Bechtol, E., Bechtol, K., Becker, A. C., Becla, J., Beldica, C., Bellavia, S., Bianco, F. B., Biswas, R., Blanc, G., Blazek, J., Blandford, R. D., Bloom, J. S., Bogart, J., Bond, T. W., Booth, M. T., Borgland, A. W., Borne, K., Bosch, J. F., Boutigny, D., Brackett, C. A., Bradshaw, A., Brandt, W. N., Brown, M. E., Bullock, J. S., Burchat, P., Burke, D. L., Cagnoli, G., Calabrese, D., Callahan, S., Callen, A. L., Carlin, J. L., Carlson, E. L., Chandrasekharan, S., Charles-Emerson, G., Chesley, S., Cheu, E. C., Chiang, H.-F., Chiang, J., Chirino, C., Chow, D., Ciardi, D. R., Claver, C. F., Cohen-Tanugi, J., Cockrum, J. J., Coles, R., Connolly, A. J., Cook, K. H., Cooray, A., Covey, K. R., Cribbs, C., Cui, W., Cutri, R., Daly, P. N., Daniel, S. F., Daruich, F., Daubard, G., Daues, G., Dawson, W., Delgado, F., Dellapenna, A., de Peyster, R., de Val-Borro, M., Digel, S. W., Doherty, P., Dubois, R., Dubois-Felsmann, G. P., Durech, J., Economou, F., Eifler, T., Eracleous, M., Emmons, B. L., Neto, A. F., Ferguson, H., Figueroa, E., Fisher-Levine, M., Focke, W., Foss, M. D., Frank, J., Freemon, M. D., Gangler, E., Gawiser, E., Geary, J. C., Gee, P., Geha, M., Gessner, C. J. B., Gibson, R. R., Kirk Gilmore, D., Glanzman, T., Glick, W., Goldina, T., Goldstein, D. A., Goodenow, I., Graham, M. L., Gressler, W. J., Gris, P., Guy, L. P., Guyonnet, A., Haller, G., Harris, R., Hascall, P. A., Haupt, J., Hernandez, F., Herrmann, S., Hileman, E., Hoblitt, J., Hodgson, J. A., Hogan, C., Howard, J. D., Huang, D., Huffer, M. E., Ingraham, P., Innes, W. R., Jacoby,

- S. H., Jain, B., Jammes, F., James Jee, M., Jenness, T., Jernigan, G., Jevremović, D., Johns, K., Johnson, A. S., Johnson, M. W. G., Lynne Jones, R., Juramy-Gilles, C., Jurić, M., Kalirai, J. S., Kallivayalil, N. J., Kalmbach, B., Kantor, J. P., Karst, P., Kasliwal, M. M., Kelly, H., Kessler, R., Kinnison, V., Kirkby, D., Knox, L., Kotov, I. V., Krabbendam, V. L., Simon Krughoff, K., Kubánek, P., Kuczewski, J., Kulkarni, S., Ku, J., Kurita, N. R., Lage, C. S., Lambert, R., Lange, T., Brian Langton, J., Le Guillou, L., Levine, D., Liang, M., Lim, K.-T., Lintott, C. J., Long, K. E., Lopez, M., Lotz, P. J., Lupton, R. H., Lust, N. B., MacArthur, L. A., Mahabal, A., Mandelbaum, R., Markiewicz, T. W., Marsh, D. S., Marshall, P. J., Marshall, S., May, M., McKercher, R., McQueen, M., Meyers, J., Migliore, M., Miller, M., Mills, D. J., Miraval, C., Moeyens, J., Moolekamp, F. E., Monet, D. G., Moniez, M., Monkewitz, S., Montgomery, C., Morrison, C. B., Mueller, F., Muller, G. P., Arancibia, F. M., Neill, D. R., Newbry, S. P., Nief, J.-Y., Nomerotski, A., Nordby, M., O'Connor, P., Oliver, J., Olivier, S. S., Olsen, K., O'Mullane, W., Ortiz, S., Osier, S., Owen, R. E., Pain, R., Palecek, P. E., Parejko, J. K., Parsons, J. B., Pease, N. M., Matt Peterson, J., Peterson, J. R., Petravick, D. L., Libby Petrick, M. E., Petry, C. E., Pierfederici, F., Pietrowicz, S., Pike, R., Pinto, P. A., Plante, R., Plate, S., Plutchak, J. P., Price, P. A., Prouza, M., Radeka, V., Rajagopal, J., Rasmussen, A. P., Regnault, N., Reil, K. A., Reiss, D. J., Reuter, M. A., Ridgway, S. T., Riot, V. J., Ritz, S., Robinson, S., Roby, W., Roodman, A., Rosing, W., Roucelle, C., Rumore, M. R., Russo, S., Saha, A., Sassolas, B., Schalk, T. L., Schellart, P., Schindler, R. H., Schmidt, S., Schneider, D. P., Schneider, M. D., Schoening, W., Schumacher, G., Schwamb, M. E., Sebag, J., Selvy, B., Sembroski, G. H., Seppala, L. G., Serio, A., Serrano, E., Shaw, R. A., Shipsey, I., Sick, J., Silvestri, N., Slater, C. T., Allyn Smith, J., Chris Smith, R., Sobhani, S., Soldahl, C., Storr-Lombardi, L., Stover, E., Strauss, M. A., Street, R. A., Stubbs, C. W., Sullivan, I. S., Sweeney, D., Swinbank, J. D., Szalay, A., Takacs, P., Tether, S. A., Thaler, J. J., Thayer, J. G., Thomas, S., Thornton, A. J., Thukral, V., Tice, J., Trilling, D. E., Turri, M., Van Berg, R., Vanden Berk, D., Vetter, K., Virieux, F., Vucina, T., Wahl, W., Walkowicz, L., Walsh, B., Walter, C. W., Wang, D. L., Wang, S.-Y., Warner, M., Wiecha, O., Willman, B., Winters, S. E., Wittman, D., Wolff, S. C., Michael Wood-Vasey, W., Wu, X., Xin, B., Joachim, P., and Zhan, H. LSST: From science drivers to reference design and anticipated data products. *The Astrophysical Journal*, 873(2):111, March 2019. ISSN 0004-637X. doi: 10.3847/1538-4357/ab042c. URL <https://iopscience.iop.org/article/10.3847/1538-4357/ab042c/meta>.
- Kamilov, U. S., Bouman, C. A., Buzzard, G. T., and Wohlberg, B. Plug-and-Play methods for integrating physical and learned models in computational imaging. *arXiv e-prints*, March 2022. URL <http://arxiv.org/abs/2203.17061>.
- Kidger, P. and Garcia, C. Equinox: neural networks in JAX via callable PyTrees and filtered transformations. *Differentiable Programming workshop at Neural Information Processing Systems 2021*, 2021.
- Kingma, D. P. and Welling, M. Auto-Encoding variational bayes. *arXiv e-prints*, December 2013. URL <http://arxiv.org/abs/1312.6114v10>.
- Lanusse, F., Melchior, P., and Moolekamp, F. Hybrid Physical-Deep learning model for astronomical inverse problems. *arXiv e-prints*, December 2019. URL <http://arxiv.org/abs/1912.03980>.
- Liu, C., Shum, H.-Y., and Freeman, W. T. Face hallucination: Theory and practice. *International Journal of Computer Vision*, 75:115–134, 2007.
- Melchior, P., Moolekamp, F., Jerdee, M., Armstrong, R., Sun, A.-L., Bosch, J., and Lupton, R. Scarlet: Source separation in multi-band images by constrained matrix factorization. *Astronomy and Computing*, 24:129–142, 2018.
- Schawinski, K., Zhang, C., Zhang, H., Fowler, L., and Santhanam, G. K. Generative adversarial networks recover features in astrophysical images of galaxies beyond the deconvolution limit. *Monthly notices of the Royal Astronomical Society*, 467:L110, May 2017. ISSN 0035-8711. doi: 10.1093/mnras/lsx008. URL <https://ui.adsabs.harvard.edu/abs/2017MNRAS.467L.110S>.
- Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.
- Van Rossum, G. and Drake Jr, F. L. *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam, 1995.
- Wang, N., Tao, D., Gao, X., Li, X., and Li, J. A comprehensive survey to face hallucination. *International journal of computer vision*, 106:9–30, 2014.
- Yao, Z., Gholami, A., Shen, S., Mustafa, M., Keutzer, K., and Mahoney, M. W. ADAHESSIAN: An Adaptive Second Order Optimizer for Machine Learning. *arXiv e-prints*, art. arXiv:2006.00719, June 2020. doi: 10.48550/arXiv.2006.00719.

Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., and Huang, T. S.  
Generative image inpainting with contextual attention. In  
*Proceedings of the IEEE conference on computer vision  
and pattern recognition*, pp. 5505–5514, 2018.