AGNet: Weighing Black Holes with Machine Learning

Joshua Yao-Yu Lin*
University of Illinois at Urbana-Champaign
NCSA
Urbana, IL, USA
yaoyuyl2@illinois.edu

Sneh Pandya*
University of Illinois at Urbana-Champaign
NCSA
Urbana, IL, USA
snehjp2@illinois.edu

Devanshi Pratap
Microsoft
University of Illinois at Urbana-Champaign
NCSA
Urbana, IL, USA
pratapdevanshi@gmail.com

Xin Liu
University of Illinois at Urbana-Champaign
NCSA
Urbana, IL, USA
xinliuxl@illinois.edu

Matias Carrasco Kind
National Center for Supercomputing Applications (NCSA)
University of Illinois at Urbana-Champaign
Urbana, IL, USA
mcarras2@illinois.edu

Abstract

Supermassive black holes (SMBHs) are ubiquitously found at the centers of most galaxies. Measuring SMBH mass is important for understanding the origin and evolution of SMBHs. However, traditional methods require spectral data which is expensive to gather. To solve this problem, we present an algorithm that weighs SMBHs using quasar light time series, circumventing the need for expensive spectra. We train, validate, and test neural networks that directly learn from the Sloan Digital Sky Survey (SDSS) Stripe 82 data for a sample of $9,038$ spectroscopically confirmed quasars to map out the nonlinear encoding between black hole mass and multi-color optical light curves. We find a 1σ scatter of 0.35 dex between the predicted mass and the fiducial virial mass based on SDSS single-epoch spectra. Our results have direct implications for efficient applications with future observations from the Vera Rubin Observatory.

1 Introduction

Supermassive black holes (SMBHs) with masses of millions to tens of billions times the mass of the Sun are commonly found at the hearts of massive galaxies [1]. While black holes themselves are invisible, as light cannot escape them, the associated phenomena are visible. These actively feeding SMBH are known as Active Galactic Nuclei (AGN). The most dramatic of these is called a quasar. Quasars are among the most powerful and distant objects in the universe [2]. The glow of matter as it falls into SMBHs is what makes quasars so bright [3].

Quasars provide a window to study how a SMBH grows with time [4]. Because of their brightness, it is possible to detect quasars almost close to the edge of the observable universe [5]. They offer

Third Workshop on Machine Learning and the Physical Sciences (NeurIPS 2020), Vancouver, Canada.

a "standard candle" to study the expansion history of the universe to understand the nature of Dark Energy [6, 7] – arguably the biggest mystery in contemporary astrophysics.

Measuring SMBH mass and redshift is important for understanding the origin and evolution of quasars. However, traditional methods require spectral data which is highly expensive; the existing $\sim$0.5 million mass estimates represent about 30 years' worth of community efforts [8, 9]. The Vera C. Rubin Observatory [10] Legacy Survey of Space and time (LSST) will discover $\sim$17 million quasars[1]. A much more efficient approach for estimating SMBH mass is needed to maximize LSST AGN science, which is, however, still lacking.

Here we present a new approach to solve the problem based on Machine Learning (ML). ML has been applied for many applications in astronomy [11, 12, 13, 14, 15, 16, 17, 18, 19]. Recently, ML has been employed to classify quasars and predict their cosmological redshift [20] using data from the Sloan Digital Sky Survey [SDSS; 21]. Autoencoders have also been used to extract time series features to model the quasar variability and estimate quasar mass [22].

The current work represents the first ML application for predicting quasar mass with multi-band time series. Our ML approach is well motivated. There is empirical evidence and theoretical reasons to believe that the quasar light curve encodes physical information about its SMBH mass [23, 24]. However, the encoding is nonlinear and difficult to model using standard statistics. We train neural networks that directly learn from the data to map out the nonlinear encoding. Our approach is fundamentally different from previous methods. The networks directly weigh SMBHs using quasar light curves, which are much cheaper to collect for a large sample. Our scheme is directly applicable for the Vera C. Rubin Observatory LSST [10], achieving a high efficiency by circumventing the need for expensive spectroscopic observations.

2 Data and Method

2.1 Data

2.1.1 SDSS Stripe 82 Light Curves

We adopt multi-color light curves from the SDSS Stripe 82 as our training and testing data. Our sample consists of $\sim$10,000 quasars. We take the mean of the *ugriz* magnitudes as features. We adopt the additional time series features from the Damped-Random-Walk model fitting parameters [24], such as the variability timescale (τ) and variance, (σ). Time series features have shown to be useful in predicting quasar redshift [20] using the Python library Feature Analysis Time Series (FATS) [25].

2.1.2 Virial Black Hole Mass and Spectroscopic Redshift

We assume the virial SMBH mass estimates from [26] and the spectroscopic redshifts as the ground truth. Cosmological redshift is the distortion of light caused by the expansion of space.

2.2 Data Preprocessing

We first clean the sample by removing quasars with no reliable virial mass estimates (e.g. due to low spectroscopic data quality). In addition to using the mean *ugriz* bands as features, we compute the colors ($u - g$, $g - r$, $r - i$, $i - z$, $z - u$) as features for a quasar. The colors are more robust features than the individual bands in that they provide more information regarding the quasar's spectral energy distribution (SED) and temperature. We further use cosmological redshift and K-corrected i-band magnitude (M_i) as features in predicting mass to provide enough information for the network to infer the intrinsic luminosity of the quasar. To standardize the effect of our features in training we apply the scikit-learn *StandardScaler* which removes the mean and scales the data to unit variance. We split our baseline dataset into an 85$\%$ training and 15$\%$ testing set.

2.3 Neural Networks

We use Feedforward Neural Networks (FNN) to predict SMBH mass and redshift. Our FNN architecture for SMBH mass prediction features a 9-neuron input layer, followed by 5 hidden layers

[1] https://www.lsst.org/sites/default/files/docs/sciencebook/SB_10.pdf

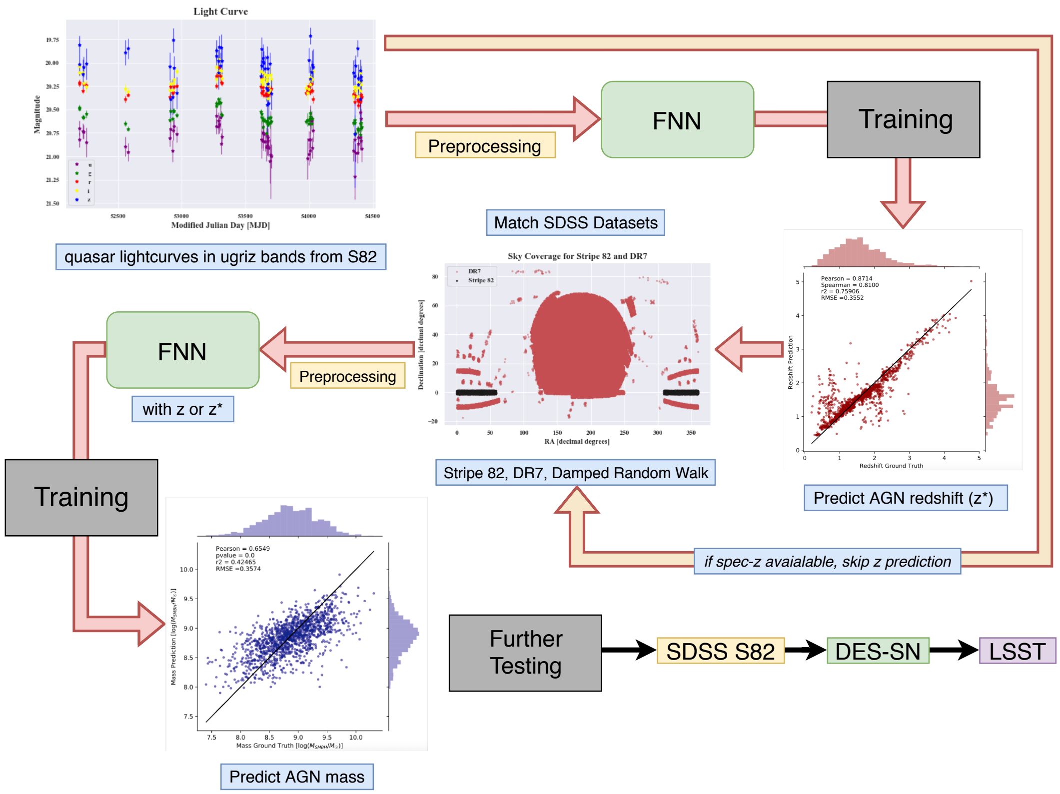

Figure 1: AGNet Implementation Flowchart.

with 64 neurons each. Our architecture for redshift is identical with exception of a 10-neuron input layer. For the mass implementation we use the SDSS colors, redshift, M_i, and τ, σ as features, and just using the SDSS magnitudes and colors as features in predicting redshift.

A diagram of our network architecture for mass is shown in Figure 3, with each node representing four neurons. We investigated multiple different combinations of base-8 neuron architectures, and concluded that 64 neurons gave the best results. We followed a similar process to determine that network performance did not increase greatly for hidden layers greater than five. We use ReLU activation on all neurons [27] after each fully connected layer.

We train our neural network with gradient descent based AdamW optimizer, an adaptive learning rate optimization algorithm [28, 29], with default learning rate set as 0.01. We optimize our FNN with the SmoothL1 loss function given as:

$$\mathcal{L} = \begin{cases} 0.5(x-y)^2, & |x-y| < 1 \\ |x-y| - 0.5, & \text{otherwise} \end{cases} \tag{1}$$

where x is the network prediction and y is the ground truth value. We train for 50 epochs.

Our network currently operates with $\sim$20,000 parameters. We investigated a different neural network architecture with less parameters to reduce the chance of overfitting associated with the number of network parameters.

We considered other loss functions, notably mean squared error, but decided on SmoothL1 because of better performance. We previously investigated RNN and CNN architectures using image and tensor transformations to transform light curve image data into 2D numpy tensors. We incorporated transfer learning and applied ResNet18 and Google EfficientNet architectures [30, 31], however due to limitations in data sample size and irregular gaps in the light curve data, we found that data preprocessing and engineering helped the neural networks to perform better. We additionally tested a variety of loss functions, feature scalers, and other hyperparameters for tuning.

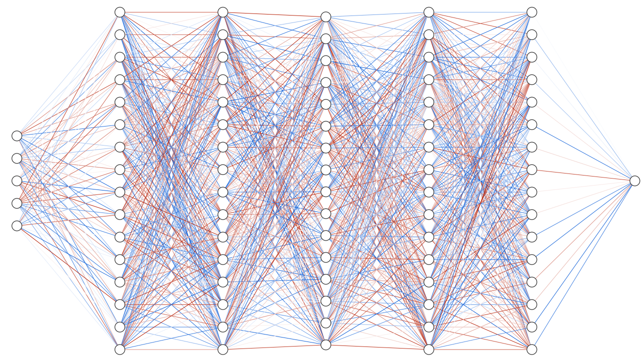

Figure 2: Network Architecture (made using NN-SVG). Each hidden layer neuron corresponds to four neurons. See Table 1 for dimension of input layer. Output neuron can be AGN redshift or AGN SMBH mass depending on implementation.

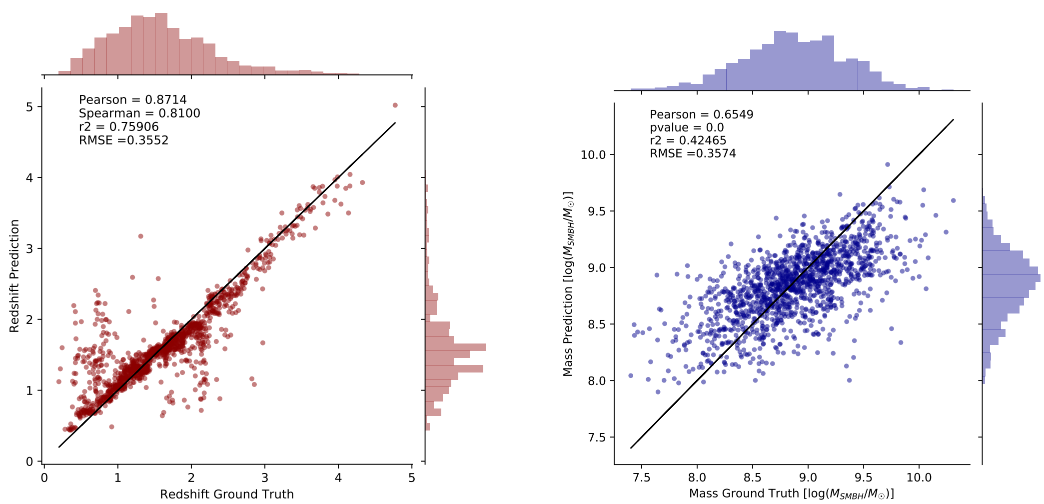

Figure 3: AGNet predictions for SMBH redshift (left) and AGNet predictions for SMBH mass (right). Ground truth shown by 1:1 black line with network predictions as the scatter. Trained over $\sim$7000 quasars using features from original light time series as well as spectroscopic redshift and M_i from SDSS Stripe 82 (§2.1.1) for mass estimations. Optimized using SmoothL1 loss and AdamW optimizer for 50 epochs of training.

3 Results

3.1 Network Performance on Redshift

Following [20], we compare our network performance on redshift estimation. We use *ugriz* bands and colors as features for predicting redshift following the architecture outlined above. Our best performance gives a RMSE = 0.355 and a $R^2 = 0.759$.

We notice considerable deviation in network prediction for low redshift ($z < 1$). One possible explanation for this is degeneracy in the color of quasars at redshift $z < 1$ and quasars at redshift $1 < z < 2$ [32].

3.2 Network Performance on Mass

Using features from $\sim$10,000 spectroscopically confirmed quasars from the Sloan Digital Sky Survey, we achieve a RMSE = 0.357 and a $R^2 = 0.418$. For context, this is already comparable to the systematic uncertainty in the spectroscopic mass "ground truth" estimate [8]. We also test mass predictions using information from only the original quasar light time series (see Figure 2). To do this, we use our AGNet architecture to predict M_i using $ugriz$ colors, as well as the redshift predictions (§3.1) as testing data in predicting mass. In this test, we achieve RMSE = 0.401 and a $R^2 = 0.272$.

3.3 Comparison to K-Nearest Neighbors

We now compare AGNet performance to a K-Nearest Neighbors (KNN) algorithm following [20]. We follow the same features and preprocessing as our AGNet implementation, with an ideal K value of $K = 19$. For quasar redshift, we achieve a RMSE = 0.386 and for mass we achieve a RMSE = 0.370. Using AGNet predicted redshift (z*) and M_i values, KNN achieves RMSE = 0.398, which is comparable to AGNet performance. The KNN performance suggests that spectral and time series features of limited data cannot be extracted as efficiently by traditional ML methods.

Summary Statistics				
ML algorithm	Features	Parameters	RMSE	R^2
AGNet	colors and bands	redshift (z)	0.355	0.759
KNN	colors and bands	redshift (z)	0.386	0.715
AGNet (w spec-z)	colors, τ, σ, M_i, z	SMBH mass	0.357	0.425
KNN (w spec-z)	colors, τ, σ, M_i, z	SMBH mass	0.370	0.369
AGNet (w/o spec-z)	colors, τ, σ, M_i, z*	SMBH mass	0.401	0.272
KNN (w/o spec-z)	colors, τ, σ, M_i, z*	SMBH mass	0.398	0.283

Table 1: Summary Statistics

4 Discussion and Future work

We have shown that with photometric light curves, our AGNet pipeline provides a fast and automatic way to predict SMBH mass and redshift. The neural network is able to approximate a function from the given features to predict the desired parameters. It is able to learn from information provided from quasar light time series without expensive spectroscopic spectra.

To improve our model, we plan on implementing negative log likelihood loss to quantify uncertainties in our network predictions [33]. We will also explore additional time series features for AGNet to learn from outside of σ and τ using FATS [25]. We will expand our work with larger data sets in future, such as the Dark Energy Survey Supernova Fields [34] and the Vera Rubin Observatory. More robust and high quality data may improve our results, and may allow us to utilize more advanced CNN architectures and transfer learning techniques.

Broader Impact

We will open source our codes along with journal publication. Within the astronomy community, this work has potential for data analysis not only on AGN light curves, but also other transients detected by large-scale multi-band sky surveys (e.g. Vera Rubin Observatory).

For broader impact on the public, we provide tools applicable for data analysis (parameter estimation) on time series data with large gaps (e.g. climate science, biology, medical diagnosis).

Acknowledgments and Disclosure of Funding

We thank the anonymous referees for helpful comments. This work utilizes resources supported by the National Science Foundation's Major Research Instrumentation program, grant No.1725729, as well as the University of Illinois at Urbana-Champaign. We thank Dr. Volodymyr Kindratenko and Dr. Dawei Mu at the National Center for Supercomputing Applications (NCSA) for their assistance with the GPU cluster used in this work and helpful comments. JYYL, SP, and XL acknowledge support from the NCSA fellowship. SP and DP acknowledge support from the NCSA SPIN and NSF REU INCLUSION programs.

References

[1] John Kormendy and Luis C Ho. Coevolution (or not) of supermassive black holes and host galaxies. *Annual Review of Astronomy and Astrophysics*, 51:511–653, 2013.

[2] Maarten Schmidt. Space distribution and luminosity functions of quasi-stellar radio sources. *The Astrophysical Journal*, 151:393, 1968.

[3] Martin J Rees. Black hole models for active galactic nuclei. *Annual review of astronomy and astrophysics*, 22(1):471–506, 1984.

[4] Andrzej Soltan. Masses of quasars. *Monthly Notices of the Royal Astronomical Society*, 200(1):115–122, 1982.

[5] Eduardo Bañados, Bram P Venemans, Chiara Mazzucchelli, Emanuele P Farina, Fabian Walter, Feige Wang, Roberto Decarli, Daniel Stern, Xiaohui Fan, Frederick B Davies, et al. An 800-million-solar-mass black hole in a significantly neutral universe at a redshift of 7.5. *Nature*, 553(7689):473–476, 2018.

[6] Anthea L King, Tamara M Davis, KD Denney, Marianne Vestergaard, and D Watson. High-redshift standard candles: predicted cosmological constraints. *Monthly Notices of the Royal Astronomical Society*, 441(4):3454–3476, 2014.

[7] Paola Marziani, Deborah Dultzin, Mauro D'Onofrio, JA De Diego Onsurbe, Alenka Negrete, Ascensión Del Olmo, Mary Loli Martínez-Aldama, Edi Bon, Natasa Bon, and Giovanna Maria Stirpe. Extreme quasars as distance indicators in cosmology. *Frontiers in Astronomy and Space Sciences*, 6:80, 2019.

[8] Yue Shen. The mass of quasars. *arXiv preprint arXiv:1302.2643*, 2013.

[9] Suvendu Rakshit, CS Stalin, and Jari Kotilainen. Spectral properties of quasars from sloan digital sky survey data release 14: The catalog. *The Astrophysical Journal Supplement Series*, 249(1):17, 2020.

[10] Željko Ivezić, Steven M Kahn, J Anthony Tyson, Bob Abel, Emily Acosta, Robyn Allsman, David Alonso, Yusra AlSayyad, Scott F Anderson, John Andrew, et al. Lsst: from science drivers to reference design and anticipated data products. *The Astrophysical Journal*, 873(2):111, 2019.

[11] Guillermo Cabrera-Vives, Ignacio Reyes, Francisco Förster, Pablo A Estévez, and Juan-Carlos Maureira. Deep-hits: Rotation invariant convolutional neural network for transient detection. *arXiv preprint arXiv:1701.00458*, 2017.

[12] Tom Charnock and Adam Moss. Deep recurrent neural networks for supernovae classification. *The Astrophysical Journal Letters*, 837(2):L28, 2017.

[13] Edward J Kim and Robert J Brunner. Star-galaxy classification using deep convolutional neural networks. *Monthly Notices of the Royal Astronomical Society*, page stw2672, 2016.

[14] Daniel George and EA Huerta. Deep learning for real-time gravitational wave detection and parameter estimation: Results with advanced ligo data. *Physics Letters B*, 778:64–70, 2018.

[15] Xin Huang, Huaning Wang, Long Xu, Jinfu Liu, Rong Li, and Xinghua Dai. Deep learning based solar flare forecasting model. i. results for line-of-sight magnetograms. *The Astrophysical Journal*, 856(1):7, 2018.

[16] François Lanusse, Quanbin Ma, Nan Li, Thomas E Collett, Chun-Liang Li, Siamak Ravanbakhsh, Rachel Mandelbaum, and Barnabás Póczos. Cmu deeplens: deep learning for automatic image-based galaxy–galaxy strong lens finding. *Monthly Notices of the Royal Astronomical Society*, 473(3):3895–3906, 2018.

[17] Dezső Ribli, Bálint Ármin Pataki, and István Csabai. An improved cosmological parameter inference scheme motivated by deep learning. *Nature Astronomy*, 3(1):93–98, 2019.

[18] Colin J Burke, Patrick D Aleo, Yu-Ching Chen, Xin Liu, John R Peterson, Glenn H Sembroski, and Joshua Yao-Yu Lin. Deblending and classifying astronomical sources with mask r-cnn deep learning. *Monthly Notices of the Royal Astronomical Society*, 490(3):3952–3965, 2019.

[19] Joshua Yao-Yu Lin, George N Wong, Ben S Prather, and Charles F Gammie. Feature extraction on synthetic black hole images. *arXiv e-prints*, pages arXiv–2007, 2020.

[20] Johanna Pasquet-Itam and Jérôme Pasquet. Deep learning approach for classifying, detecting and predicting photometric redshifts of quasars in the sloan digital sky survey stripe 82. *Astronomy & Astrophysics*, 611:A97, 2018.

[21] Donald G York, J Adelman, John E Anderson Jr, Scott F Anderson, James Annis, Neta A Bahcall, JA Bakken, Robert Barkhouser, Steven Bastian, Eileen Berman, et al. The sloan digital sky survey: Technical summary. *The Astronomical Journal*, 120(3):1579, 2000.

[22] Yutaro Tachibana, Matthew J Graham, Nobuyuki Kawai, SG Djorgovski, Andrew J Drake, Ashish A Mahabal, and Daniel Stern. Deep modeling of quasar variability. *arXiv preprint arXiv:2003.01241*, 2020.

[23] Brandon C Kelly, Jill Bechtold, and Aneta Siemiginowska. Are the variations in quasar optical flux driven by thermal fluctuations? *The Astrophysical Journal*, 698(1):895, 2009.

[24] Ch L MacLeod, Ž Ivezić, CS Kochanek, S Kozłowski, B Kelly, E Bullock, A Kimball, B Sesar, D Westman, K Brooks, et al. Modeling the time variability of sdss stripe 82 quasars as a damped random walk. *The Astrophysical Journal*, 721(2):1014, 2010.

[25] Isadora Nun, Pavlos Protopapas, Brandon Sim, Ming Zhu, Rahul Dave, Nicolas Castro, and Karim Pichara. Fats: Feature analysis for time series. *arXiv preprint arXiv:1506.00010*, 2015.

[26] Yue Shen, Gordon T Richards, Michael A Strauss, Patrick B Hall, Donald P Schneider, Stephanie Snedden, Dmitry Bizyaev, Howard Brewington, Viktor Malanushenko, Elena Malanushenko, et al. A catalog of quasar properties from sloan digital sky survey data release 7. *The Astrophysical Journal Supplement Series*, 194(2):45, 2011.

[27] Abien Fred M. Agarap. Deep learning using rectified linear units (relu). *arXiv:1803.08375*, 2018.

[28] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[29] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

[30] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[31] Mingxing Tan and Quoc V Le. Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv preprint arXiv:1905.11946*, 2019.

[32] Gordon T. Richards, Michael A. Weinstein, Donald P. Schneider, Xiaohui Fan, Michael A. Strauss, Daniel E. Vanden Berk, James Annis, Scott Burles, Emily M. Laubacher, Donald G. York, and et al. Photometric redshifts of quasars. *The Astronomical Journal*, 122(3):1151–1162, Sep 2001.

[33] Laurence Perreault Levasseur, Yashar D Hezaveh, and Risa H Wechsler. Uncertainties in parameters estimated with neural networks: Application to strong gravitational lensing. *The Astrophysical Journal Letters*, 850(1):L7, 2017.

[34] R Kessler, J Marriner, M Childress, R Covarrubias, CB D'Andrea, DA Finley, J Fischer, Ryan Joseph Foley, D Goldstein, RR Gupta, et al. The difference imaging pipeline for the transient search in the dark energy survey. *The Astronomical Journal*, 150(6):172, 2015.