# Big Data Analytics for Identification of Gravitationally Lensed Quasars in the Dark Energy Survey

Pranav Sivakumar, Brian Nord, Liz Buckley-Geer

## Abstract

We report results from an automated method to identify lensed quasars from the Dark Energy Survey using a PSF-difference-based algorithm aimed at identifying close-separation lens candidates. The PSF-difference algorithm utilizes the difference between PSF magnitude and model magnitude, as well as image segmentation, to deblend and identify close-separation candidates. In total, the algorithms identified 156 final lens candidates and also identified a number of known lensed quasars within the DES footprint, indicating that the method described in this paper is effective in identifying candidate lensed quasars. Efforts to obtain follow-up observations for confirmation of the final candidates are ongoing, and constraints on cosmological parameters will be discussed in future papers.

# Big Data Analytics for Identification of Gravitationally Lensed Quasars in the Dark Energy Survey

## Background

With the advent of larger telescopes and all-sky surveys, the amount of available quasar data has grown exponentially in the last 10 years. These large data sets have enabled statistically valid approaches to search for and identify quasars that could be lensed by one or more massive objects interposed between the source and an observer. Thus, studying gravitational lensing of quasars is rapidly becoming a reliable method to understand a number of larger issues in astrophysics and cosmology.

Quasars, also known as QSOs (quasi-stellar objects), are point-like luminous astronomical objects located around black holes. QSOs are some of the oldest and farthest objects known. Since they are very distant, the electromagnetic radiation they produce can be affected both by their host galaxies, as well as other objects in the line of sight between Earth and the quasars.

Almost 100 years ago, general relativity predicted that light is bent by massive objects interposed between a light source and an observer. This phenomenon, known as gravitational lensing, can make the source appear brighter and produce multiple images of the source. The lensing of quasars, classified as strong lensing, is particularly valuable in astrophysics; it helps decipher the properties of dark matter and dark energy, which together constitute about 95 percent of the universe but are poorly understood (Frieman, Turner, & Huterer, 2008).

Despite the utility of quasar lensing, only a small number of lensed quasars have been discovered to date. Since 1979, when the first lensed quasar was discovered, fewer than 150 lensed quasars have been confirmed. However, due to its depth and the number of objects it covers, DES is expected to nearly double the number of known lensed quasars (Oguri & Marshall, 2010). These factors motivated me to continue working on quasar lensing and apply my prior SDSS work to DES.

Since gravitational lensing offers tantalizing prospects for finding dark matter and gaining key insights into its nature, it has been an active area of research in astrophysics and cosmology for more than 30 years. Some of the earliest research and theoretical modeling of strong lensing of quasars began in the early 1980s. The exponential growth of research in this area has been spurred by increasing availability of detailed quasar data from both ground-based and space-based telescopes.

Turner, Ostriker, and Gott (1984) were the first to examine the likelihood that a quasar will be lensed by a galaxy or by a point mass. Maoz et al. (1993) showed that about 1% of bright quasars at $z > 1$ ($z$ = redshift) are gravitationally lensed into two or more images using data from the Hubble Space Telescope (HST) Snapshot Survey. In parallel, lensed quasars were also identified through observations in radio frequencies. FBQ 0951+2635, identified by the FIRST survey and reported in Schechter, Gregg, Becker, Helfand, and White (1998), was the first lensed quasar identified through radio observations. Up to this point, data was available only for a few dozen to a few hundred quasars; finding lensed systems within small data sets proved difficult. The Cosmic-Lens All Sky Survey (CLASS), described in Browne et al. (2003) was the first survey that utilized data sets with thousands of quasars. CLASS discovered 22 systems of lensed quasars from about 16,000 radio sources.

The Sloan Digital Sky Survey, which incorporated a robust data pipeline to identify and target quasars, increased data availability to hundreds of thousands of quasars. Using an early data release, the SDSS DR3, Oguri et al. (2006) reported an algorithm to identify gravitationally lensed quasars that formed the basis of the SDSS Quasar Lens Search (SQLS). The algorithm used a photometric method to identify close-separation lensed quasars and color selection techniques to identify lenses with larger separations. The SQLS identified 11 lensed quasars in its first run; the results are described in Inada et al. (2008). Following up their earlier work on the SQLS, Oguri et al. (2012) applied quasar lensing probabilities to establishing limits on the value of the cosmological constant and the distribution of dark energy. By comparing the observed lensing fraction with theoretical models, Oguri et al. determined that their results support the accelerated expansion model of the universe.

In 2015, I developed an algorithm to identify lensed quasar candidates from the SDSS. This method consisted of a morphological algorithm to identify wide-separation lensed quasars and a PSF-difference algorithm to identify close-separation candidates. This was the largest search for lensed quasars to date, with a baseline data set of 300,000 quasars.

Searches for lensed quasars in DES have also been initiated in the past few years. Agnello et al. (2015) discovered the first two gravitationally lensed quasars in DES, one of which was also identified in my SDSS work, using machine learning algorithms, and a third lensed quasar was reported in Ostrovski et al. (2016). However, this is the most comprehensive search for lensed quasars in DES yet performed.

**Context of this work**

The phenomenon of strong lensing is used in two distinct ways to study a number of phenomena in astrophysics and cosmology. The first approach is statistical in nature and involves identifying a large number of lensed quasars and analyzing their redshifts and separations to accurately estimate key cosmological parameters. The second approach is directed at studying in great detail a small number of lens systems; this approach studies the parameters of the lensing and lensed objects to arrive at highly accurate estimates of the dark matter in the lensing system and the Hubble constant. The research work described in this paper is focused on identifying a large number of lensed quasars and understanding the statistical relationship between separation and redshift. The next phase of the project will focus on studying detailed characteristics of individual lens systems.

Prior to the present research work and its predecessor project reported in Competition Entrant & Sivakumar (2015), quasar lensing research focused solely on either photometric or spectral characteristics of quasars. The automated candidate selection algorithm developed in this work combines both types of data to identify lensed quasar candidates, thus improving the accuracy and reliability of the candidates identified for follow-up observations.

# Methods of Procedure

### Equipment

The data set used to identify lensed quasar candidates is the Dark Energy Survey, one of the deepest sky surveys ever conducted. It covers 5000 square degrees of sky in the Southern Hemisphere, about 1/8 of the total sky, using the Dark Energy Camera (DECam) on the 4-meter Blanco telescope. The Blanco telescope is significantly larger than the 2.5-meter telescope used on the Sloan Digital Sky Survey, which was used in the previous iteration of this project. Using DES allowed me to search for lensed quasars much more deeply than in SDSS and also made it easier to detect potential lensing galaxies.

While it has a number of advantages over SDSS in terms of magnitude limits, DES is also a difficult data set to apply the automated methods from SDSS. It has no spectra, which means that the lenses identified may not necessarily be quasars and that the redshift and spectral comparison methods from DR12 are impossible. This means that DES candidates need to be much more carefully reviewed than those identified in SDSS.

### Data

Initially, a query was written to perform the relevant comparisons for the PSF-difference algorithm and extract all candidates that met the specified criteria. The results of this query were exported to CSV files and inputted into the DES Portal website, from which cutouts could be extracted for use in the segmentation algorithm. Due to the large number of candidates (about 115,000), they were divided into six groups which were each small enough to be processed by the cutout extractor. Once extracted, the cutouts required additional processing steps using ImageMagick software to feed into the segmentation. Because the cutouts cannot be zoomed in, they needed to be cropped and scaled such that only the central object in each cutout was processed by the segmentation algorithm.

The PSF-difference algorithm was first proposed in Sivakumar & Sivakumar (2015) to analyze close-separation lensed quasars, where DES is unable to recognize two distinct photometric objects. The full details of the PSF-difference algorithm, including an explanation of the principles behind the algorithm, the methodology for extracting the PSF and model magnitudes, and the use of the Petrosian radius to eliminate confounding by extended sources, are described in Sivakumar & Sivakumar (2015).

Although the PSF-difference algorithm identified a number of new candidates, a visual inspection of tens of thousands of objects proved to be tedious and inefficient. In order to address this issue and to successfully analyze the larger set of quasar data available in DR12, the PSF-difference algorithm was automated using image segmentation techniques. A flowchart of this algorithm is shown in Figure 1. Since the image segmentation algorithms required a binary image as input, the DES color images were first converted to 8-bit grayscale images by averaging the RGB channels of the original image. Subsequently, they were converted to binary images through a thresholding process.

Since the flux level and uniformity of the quasar images varied widely, each image needed to be thresholded with a different cutoff to separate the foreground and the background. For each image, the global maximum value of an image pixel was identified. Subsequently, the threshold level was set as a percentage of the global maximum and was gradually increased until at least three regions (the two objects and the background) were segmented.

In order to identify regions within the processed binary images, two different image segmentation algorithms were used. The Watershed algorithm, described in Beucher (1992, 2010), is used to identify regions within a single image. Images that were successfully deblended into regions by the Watershed algorithm were also processed through the random walker method, described in Grady (2006), to further confirm the segmentation results. Both image segmentation algorithms used in this project are adaptations of the code implemented in the Python scikit-image package developed by Stéfan et al. (2014).

The candidates that successfully passed through the segmentation algorithm were then processed by eye. Candidates that showed no evidence of lensing, were significantly extended, or were too faint to analyze properly were not considered further. Subsequently, a set of scientists reviewed this list to produce a final set of candidates. A full explanation of the PSF-difference algorithm steps is shown in Figure 1.

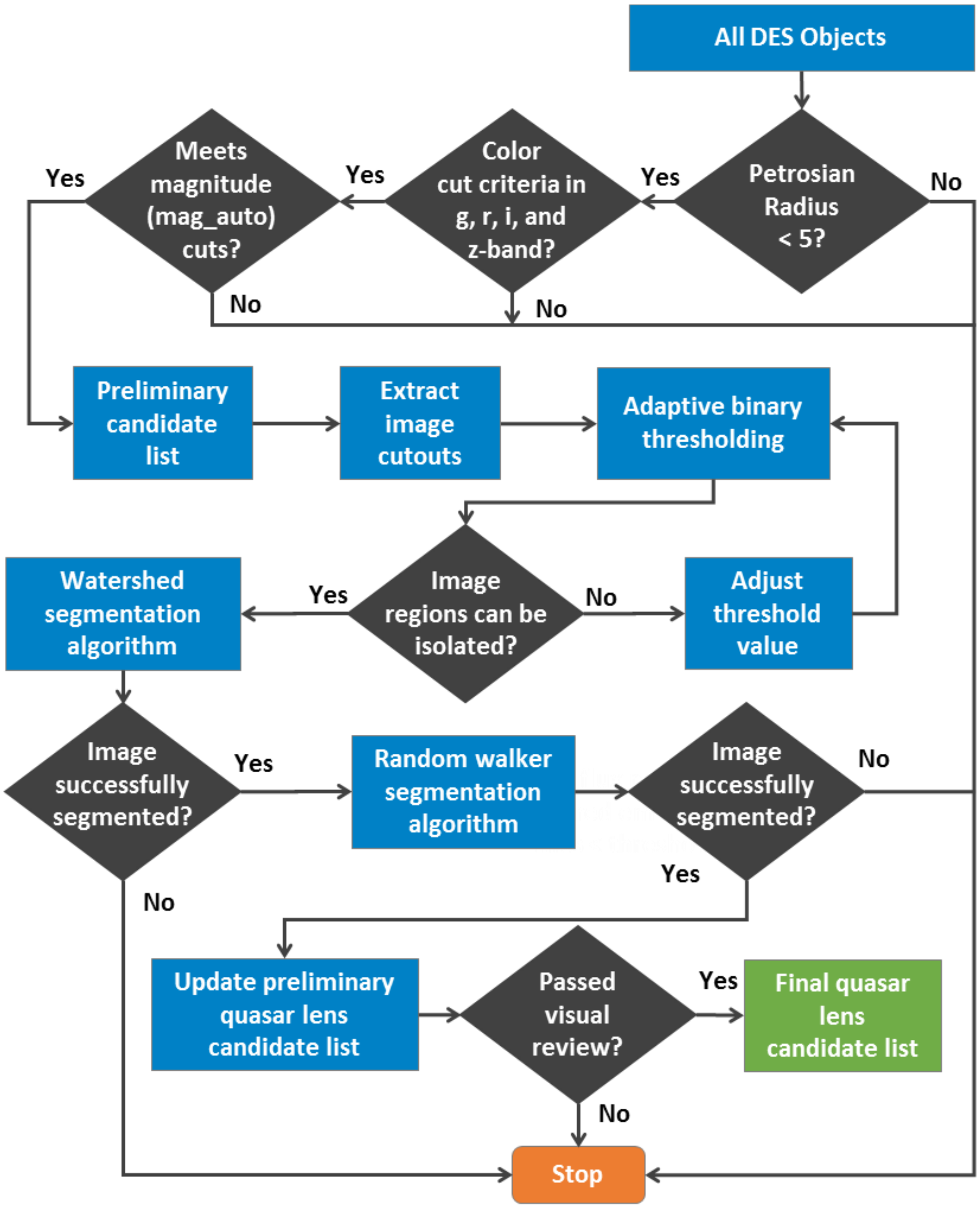


**Figure 1.** Flowchart of the PSF-difference algorithm, including the SQL query, cutout extraction, segmentation, and visual review.

## Results

The PSF-difference algorithm began with almost 150 million objects in the DES Y1A1 release. The PSF SQL query narrowed down the list to 120,000 preliminary candidates, about 60,000 of which were successfully segmented. The first round of visual inspection brought that number down to 165, of which ___ were selected as final candidates for follow-up.

Because of the lack of spectra in DES, the classification scheme described in Competition Entrant & Sivakumar (2014) had to be modified for the DES candidates. All possible candidates following visual inspection were classified into Types 1, 2, and 3 depending on the visual similarity of the images; the results are shown in Table 1. Only the 165 Type 1 candidates were selected to be reviewed in the second visual scan.

**Table 1.** Count of known lensed quasars and new candidates identified by the PSF-difference algorithm grouped by visual similarity.

| Candidate List | Visual Observation Criteria | # of Candidates |
|---|---|---|
| **Type 1** | High level of similarity | 165 |
| **Type 2** | Medium level of similarity | 111 |
| **Type 3** | Low level of similarity | 698 |

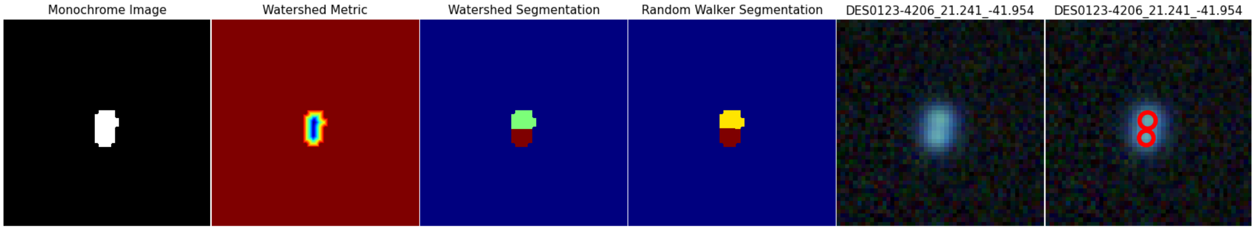


**Figure 2.** A representative sample of lensed quasar candidates identified by the PSF-difference algorithm through image segmentation.

A representative sample of the results from the automated segmentation of PSF-difference candidates are shown in Figure 2. Because of the depth of DES and the lack of spectra, the PSF-difference algorithm was able to identify a large number of candidates.

As expected, DES data proved to be complementary to the SDSS data, on which the algorithm was originally run. This is supported by the positions of the lens candidates identified from SDSS (red) and DES (blue) shown in Figure 2.

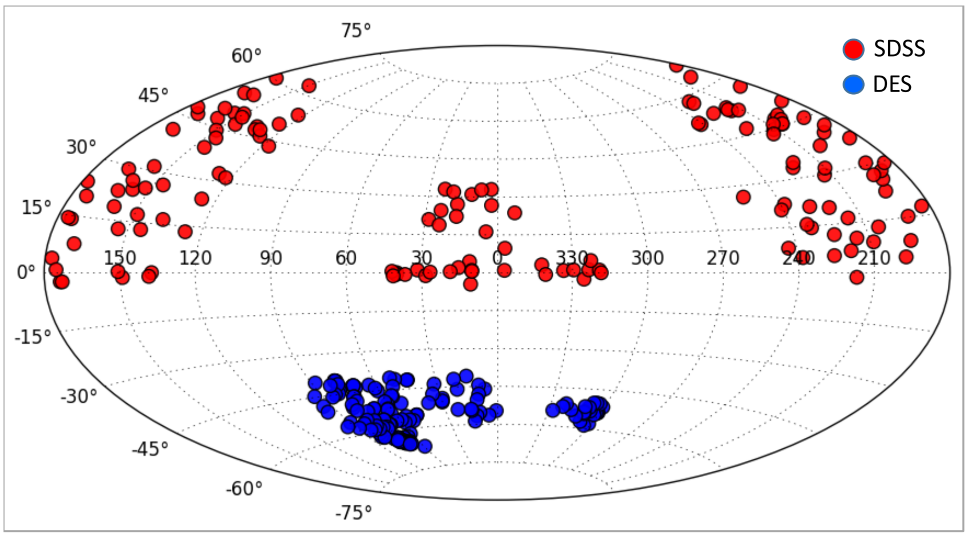


**Figure 3.** Positions of lensed quasars identified in SDSS and in DES. As expected, the SDSS candidates were predominantly located in the northern hemisphere of the sky, while all of the DES candidates were found in the southern hemisphere.

## Conclusion and future work

The results of the research project confirmed that the PSF-difference algorithm is effective in searching for lensed quasar candidates in DES.  The algorithms not only correctly identified known lensed quasars, but also found several new high-probability candidates.

The next logical step in the candidate confirmation process is to extract the WISE infrared magnitudes, W1 and W2, for all final candidates to confirm that they are in fact quasars. Once this process is complete, the final candidate list will be ready for follow-up.

This fall, follow-up observations are planned on Gemini for several strong lens candidates, including the ones identified by the PSF-difference algorithm. The redshifts obtained from this follow-up would enable us to model the confirmed lens systems to determine the amount of dark matter lensing a quasar. We are investigating strategies for modeling some of the high-probability candidates using GLAFIC and Gravlens software.

# Acknowledgments

Fermilab is operated by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the United States Department of Energy. QuarkNet is an educational program sponsored by the National Science Foundation and the Department of Energy whose aim is to support science education in schools by establishing a nationwide network of science teachers. This work was supported by CRADA FRA-2014-0022-02 agreement between Fermi Research Alliance LLC and the University of Notre Dame for the 2015-16 QuarkNet summer program.

I would like to thank my primary advisors, Brian Nord and Liz Buckley-Geer, for mentoring me throughout this project. I would also like to acknowledge Huan Lin for answering my many questions relating to the DES data and also providing transformation equations from SDSS to DES to refine both the morphological and PSF-difference algorithms. Furthermore, I thank Tom Diehl for his help with issues I encountered throughout the DES portal upload process.

## References


Agnello, A., Treu, T., Ostrovski, F., Schechter, P. L., Buckley-Geer, E. J., Lin, H., Auger, M. W., ... Zhang, Y. (2015). Discovery of two gravitationally lensed quasars in the Dark Energy Survey. *arXiv:1508.01203.*

Becker, R. H., White, R. L., & Helfand, D. J. (1995). The FIRST Survey: Faint Images of the Radio Sky at Twenty Centimeters. *The Astrophysical Journal Supplement*, 450, 559. doi: 10.1086/176166

Beucher, S. (1992). The watershed transformation applied to image segmentation. *UK Scanning Microscopy International*, supplement 6, 299-314.

Beucher, S. (2010). Image Segmentation and Mathematical Morphology. Centre for Mathematical Morphology. Retrieved from http://cmm.ensmp.fr/~beucher/wtshed.html#watshed.

Browne, I. W. A., Wilkinson, P. N., Jackson, N. J. F., Myers, S. T., Fassnacht, C. D., Koopmans, L. V. E., . . . York, T. (2003). The Cosmic Lens All-Sky Survey - II. Gravitational Lens Candidate Selection and Follow-up. *Monthly Notice of the Royal Astronomical Society*, 341, 13-32. doi:10.1046/j.1365-8711.2003.06257.x

Frieman, J.A., Turner, M. S., & Huterer, D. (2008). Dark Energy and the Accelerating Universe. *Annual Review of Astronomy and Astrophysics*, *46*, 385–432. doi:10.1146/annurev.astro.46.060407.145243

Grady, L. (2006). Random walks for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28(11)*,1768-83.

Inada, N., Oguri, M., Becker, R.H., Shin, M., Richards, G.T., Hennawi, J. F., . . . Fukugita, M. (2008). The Sloan Digital Sky Survey Quasar Lens Search. II. Statistical Lens Sample from the Third Data Release. *The Astronomical Journal*, 135, 496-511. doi: 10.1088/0004-6256/135/2/496

Maoz, D., Bahcall, J. N., Schneider, D. P., Bahcall, N. A., Djorgovski, S., Doxsey, R., . . . Yanny, B. (1993). The Hubble Space Telescope Snapshot Survey. IV - A summary of the search for gravitationally lensed quasars. *The Astrophysical Journal*, 409, 28-41. doi: 10.1086/172639

Oguri, M., Inada, N., Pindor, B., Strauss, M.A., Richards, G.T., Hennawi, J. F., Turner, E. L., . . . Brinkmann, J. (2006). The Sloan Digital Sky Survey Quasar Lens Search. I. Candidate Selection Algorithm. *The Astronomical Journal*, 132, 999-1013. doi:10.1086/506019

Oguri, M., Inada, N., Strauss, M.A., Kochanek, C.S., Kayo, I., Shin, M., . . . White, R. L. (2012). The Sloan Digital Sky Survey Quasar Lens Search. VI. Constraints on Dark Energy and the Evolution of Massive Galaxies. *The Astronomical Journal*, 143, 120-134. doi:10.1088/0004-6256/143/5/120

Oguri, M. & Marshall, P.J. Gravitationally lensed quasars and supernovae in future wide-field optical imaging surveys. *Monthly Notices of the Royal Astronomical Society, 405*(4), 2579-2593. doi: 10.1111/j.1365-2966.2010.16639.x

Ostrovski, F., McMahon, R.G., Connolly, A.J., Lemon, C.A., Auger, M.W., Banerji, M., …Walker, A.L. (in press). VDES J2325-5229 a z=2.7 gravitationally lensed quasar discovered using morphology independent supervised machine learning. arXiv:1607.01391 [astro-ph.GA]. Retrieved from http://arxiv.org/pdf/1607.01391v1.pdf

Schechter, P. L., Gregg, M. D., Becker, R. H., Helfand, D. J., & White, R. L. (1998). The First FIRST Gravitationally Lensed Quasar: FBQ 0951+2635. *The Astronomical Journal*, 115, 1371-1376. doi:10.1086/300294

SDSS Query/CasJobs [Computer Program]. (2013). Cambridge, MA: Harvard University.

Sivakumar, P. & Sivakumar, J. N. (2014). A Hybrid Photometric and Spectral Algorithm for Efficient Detection of Gravitationally Lensed Quasars. *Siemens Competition*, Siemens Foundation, Iselin, NJ.

Sivakumar, P. & Sivakumar, J. N. (2015). A Hybrid Photometric and Spectral Algorithm for Efficient Detection of Gravitationally Lensed Quasars. *Young Scientist*, 5, 49-52.

Stéfan, W., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., … scikit-image contributors (2014). scikit-image: Image processing in Python. *PeerJ 2*:e453. doi:10.7717/peerj.453

Turner, E. L., Ostriker, J. P., & Gott III J. R. (1984). The statistics of gravitational lenses: the distributions of image angular separations and lens redshifts. *The Astrophysical Journal*, 284, 1-22. doi:10.1086/162379