



Universiteit
Leiden
The Netherlands

When spectral modeling meets convolutional networks: a method for discovering reionization-era lensed quasars in multiband imaging data

Andika, I.T.; Jahnke, K.; Wel, A. van der; Bañados, E.; Bosman, S.E.I.; Davies, F.B.; ... ; Schindler, J.T.

Citation












Andika, I. T., Jahnke, K., Wel, A. van der, Bañados, E., Bosman, S. E. I., Davies, F. B., ... Schindler, J. T. (2023). When spectral modeling meets convolutional networks: a method for discovering reionization-era lensed quasars in multiband imaging data. *The Astrophysical Journal*, 943(2). doi:10.3847/1538-4357/aca66e

Version: Publisher's Version
License: [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/)
Downloaded from: <https://hdl.handle.net/1887/3716668>

Note: To cite this publication please use the final published version (if applicable).



When Spectral Modeling Meets Convolutional Networks: A Method for Discovering Reionization-era Lensed Quasars in Multiband Imaging Data

Irham Taufik Andika^{1,2,3} , Knud Jahnke¹ , Arjen van der Wel² , Eduardo Bañados¹ , Sarah E. I. Bosman¹ , Frederick B. Davies¹ , Anna-Christina Eilers^{4,11} , Anton Timur Jaelani^{5,6} , Chiara Mazzucchelli⁷ , Masafusa Onoue^{8,9} , and Jan-Torge Schindler^{1,10} 

¹ Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany; irham.andika@tum.de

² Sterrenkundig Observatorium, Universiteit Gent, Krijgslaan 281 S9, B-9000 Gent, Belgium

³ Physik-Department, Technische Universität München, James-Frank-Str. 1, D-85748 Garching, Germany

⁴ MIT Kavli Institute for Astrophysics and Space Research, 77 Massachusetts Ave., Cambridge, MA 02139, USA

⁵ Astronomy Research Group and Bosscha Observatory, FMIPA, Institut Teknologi Bandung, Jl. Ganesha 10, Bandung 40132, Indonesia

⁶ U-CoE AI-VLB, Institut Teknologi Bandung, Jl. Ganesha 10, Bandung 40132, Indonesia

⁷ Núcleo de Astronomía, Facultad de Ingeniería y Ciencias, Universidad Diego Portales, Av. Ejército 441, Santiago, 8320000, Chile

⁸ Kavli Institute for Astronomy and Astrophysics, Peking University, Beijing 100871, People's Republic of China

⁹ Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU, WPI), The University of Tokyo, Chiba 277-8583, Japan

¹⁰ Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, The Netherlands

Received 2022 July 25; revised 2022 November 1; accepted 2022 November 25; published 2023 February 3

Abstract

Over the last two decades, around 300 quasars have been discovered at $z \gtrsim 6$, yet only one has been identified as being strongly gravitationally lensed. We explore a new approach—enlarging the permitted spectral parameter space, while introducing a new spatial geometry veto criterion—which is implemented via image-based deep learning. We first apply this approach to a systematic search for reionization-era lensed quasars, using data from the Dark Energy Survey, the Visible and Infrared Survey Telescope for Astronomy Hemisphere Survey, and the Wide-field Infrared Survey Explorer. Our search method consists of two main parts: (i) the preselection of the candidates, based on their spectral energy distributions (SEDs), using catalog-level photometry; and (ii) relative probability calculations of the candidates being a lens or some contaminant, utilizing a convolutional neural network (CNN) classification. The training data sets are constructed by painting deflected point-source lights over actual galaxy images, to generate realistic galaxy–quasar lens models, optimized to find systems with small image separations, i.e., Einstein radii of $\theta_E \leq 1''$. Visual inspection is then performed for sources with CNN scores of $P_{\text{lens}} > 0.1$, which leads us to obtain 36 newly selected lens candidates, which are awaiting spectroscopic confirmation. These findings show that automated SED modeling and deep learning pipelines, supported by modest human input, are a promising route for detecting strong lenses from large catalogs, which can overcome the veto limitations of primarily dropout-based SED selection approaches.

Unified Astronomy Thesaurus concepts: [Active galactic nuclei \(16\)](#); [Active galaxies \(17\)](#); [Early universe \(435\)](#); [High-redshift galaxies \(734\)](#); [Quasars \(1319\)](#); [Reionization \(1383\)](#); [Supermassive black holes \(1663\)](#); [Strong gravitational lensing \(1643\)](#); [Gravitational lensing \(670\)](#)

1. Introduction

Quasars are among the brightest sources in the universe, powered by the accretion of matter onto supermassive black holes (SMBHs). In the high-redshift frontier, these sources are prime laboratories for understanding the formation, growth, and structure of the first SMBHs and galaxies (Mignoli et al. 2020; Pacucci & Loeb 2022). The key results from the last two decades of $z \gtrsim 6$ quasar surveys include the following: (i) the number density of distant quasars places stringent constraints on the processes necessary to seed and develop $>10^9 M_\odot$ BHs within the first billion years of cosmic history (Inayoshi et al. 2020); (ii) high- z quasars most commonly reside in massive star-forming host galaxies, with a wide range of kinematics and large-scale environments (Decarli et al. 2018; Neeleman et al. 2021; Meyer et al. 2022); and (iii) the hydrogen absorption in

the spectra of high- z quasars indicates a largely neutral intergalactic medium (IGM) at $z \gtrsim 7$ (Bañados et al. 2018; Yang et al. 2020; Wang et al. 2021) and an end of cosmic reionization by $z \sim 5.3$ (Bosman et al. 2022).

Previous high- z quasar searches have mainly been focused on the most luminous, massive, and active SMBHs, while those fainter ones, with modest to low accretion or lower mass, have been challenging to find (e.g., Fan et al. 2006; Willott et al. 2010; Mortlock et al. 2011; Venemans et al. 2015; Bañados et al. 2016; Mazzucchelli et al. 2017; Pons et al. 2019; Reed et al. 2019; Wang et al. 2019; Yang et al. 2019b; Bañados et al. 2021; Wenzl et al. 2021; Andika 2022). This is due to selections having to filter out rare quasars from the more frequent galaxies and stars, hence having to focus on the brightest population first. Another reason is that many wide-sky surveys in the prior years have been too shallow, so one is automatically restricted to studying the brightest population. Some efforts have been made toward deeper wide-sky surveys (e.g., Matsuoka et al. 2018a; Pipien et al. 2018), because exploring the fainter, currently mostly hidden, and lower-mass population of high- z quasars will provide information about the quasar luminosity function, impacting our understanding

¹¹ NASA Hubble Fellow.

of quasar contributions to the universe’s reionization (Madau 2017; Matsuoka et al. 2018b, 2022). In addition, these SMBHs might not yet be close to finishing with their mass growth, as the most luminous $z \gtrsim 6$ quasars are, allowing for the study of more “typical” SMBH growth scenarios than in the brightest quasars (Volonteri 2012; Izumi et al. 2021; Habouzit et al. 2022). Unfortunately, the spectroscopy for measuring intrinsically fainter quasar ($M_{1450} > -25$) SMBH masses with high accuracy is extremely challenging and time-consuming, even with 8 m class telescopes (Onoue et al. 2019, 2021).

Despite such situations being rare, some high- z quasars can be strongly gravitationally lensed. Such systems are one of the best alternatives for probing quasars with intrinsically lower mass and luminosity. They also enable us to investigate quasar host galaxies in unprecedented detail, thanks to the flux magnification and, very importantly, the increased effective spatial resolution, by factors of 2–10 (e.g., Stacey et al. 2018; Yang et al. 2019a; Yue et al. 2021). In addition, lensed quasars are excellent targets for constraining the dark matter profiles of the foreground deflectors (e.g., Gilman et al. 2020; Hsueh et al. 2020), determining the Hubble constant based on the time-delay cosmography (e.g., Suyu et al. 2017; Millon et al. 2020), and estimating the quasar accretion disk sizes (e.g., Chan et al. 2021).

Because of the large lensing optical depths attained at $z \gtrsim 6$, it has been predicted for decades that between $\sim 1\%$ and up to one-third of high- z quasars should be strongly lensed, although the actual level still remains unknown (e.g., Wyithe & Loeb 2002; Oguri & Marshall 2010; Pacucci & Loeb 2019; Yue et al. 2022a). Depending on the models, we anticipate that the number of lensed quasars to be discovered in an optimal data set should range from around three (lensed fraction $\sim 1\%$; Yue et al. 2022a) to greater than 10 (lensed fraction $> 4\%$; Pacucci & Loeb 2019). However, only one such lensed quasar has been found so far, J0439+1634, at $z = 6.51$ (Fan et al. 2019), aside from the approximately 300 unlensed quasars that are currently known. A leading cause of this dramatic tension is the commonly used “dropout” technique for selecting high- z quasar candidates, which requires any viable candidates to have a large abrupt drop in flux in bluer optical bandpasses, to remove lower- z contaminants. Until now, this has prevented galaxy–quasar lens systems from being selected as quasar candidates, due to the deflector galaxy’s optical flux contribution, as a reexamination of previous selection techniques has shown (Andika et al. 2020). Several studies have attempted to look for signs of strong gravitational lensing among the known $z \gtrsim 6$ quasars, finding no evidence thus far (e.g., Davies et al. 2020; Connor et al. 2021; Pacucci et al. 2022). Therefore, developing a more advanced selection technique to reveal those hidden populations of lenses is crucial.

In this work, we propose a novel approach to searching for $z \gtrsim 6$ galaxy–quasar lenses, which expands the selection space that has been missed by previous quasar surveys, by combining spectral energy distribution (SED) modeling with the convolutional neural network (CNN) classifications. Using multiband optical images that are complemented with IR photometric data, we validate our selection pipeline and provide new high-probability lensed quasar candidates. The structure of this paper is as follows. We begin in Section 2 by describing the data acquisition and preselection of the candidates through photometric color cuts and SED modeling. Section 3 then presents the details of the lens finding using the CNN, including the data

sets used for training and assessing the networks. After that, the classification results and the lensed quasar candidates are discussed in Section 4. Finally, we close with a summary and our conclusions in Section 5.

In every section of this paper, we use the Λ CDM cosmology, where $\Omega_\Lambda = 0.685$, $\Omega_m = 0.315$, and $H_0 = 67.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (Planck Collaboration et al. 2020). Also, all of the reported magnitudes are in the AB system.

2. Data and Candidate Preselection

Our lensed quasar search consists of two parts: (1) the preselection of the candidates based on their spectral color, using multiband photometric data; and (2) the calculation of the relative probabilities (of the candidates being a lensed quasar or contaminant), based on a CNN classification.

Discovering lensed quasar candidates by utilizing photometric data requires knowledge of their spectral color—i.e., the addition of foreground galaxy and background quasar lights. Here, the training data sets and spectral templates will be generated, which will be optimized to find systems with small image separations, i.e., Einstein radii of $\theta_E \leq 1''$. We aim to improve the purity of the candidates by limiting the selection to sources that we know are more likely to be lenses based on catalog-level photometry. This method provides one way of efficiently separating the candidates from most of the contaminants, while minimizing the computational resources required. The details of the first part of our search are described in the following section.

2.1. Photometric Catalog

In this work, we use the Dark Energy Survey (DES) Data Release 2 (Abbott et al. 2021) as the primary catalog. This survey is based on optical imaging collected using the Dark Energy Camera (DECam; Honscheid & DePoy 2008; Flaugher et al. 2015) on the Blanco 4 m telescope at the Cerro Tololo Inter-American Observatory, Chile. DES covers $\sim 5000 \text{ deg}^2$ of the southern Galactic cap in five bands, with a median point-spread function (PSF) FWHM of $g, r, i, z, Y = 1.''11, 0.''95, 0.''88, 0.''83, \text{ and } 0.''90$, respectively. The median coadded catalog depth is $g_{\text{DES}} = 24.7$, $r_{\text{DES}} = 24.4$, $i_{\text{DES}} = 23.8$, $z_{\text{DES}} = 23.1$, and $Y_{\text{DES}} = 21.7 \text{ mag}$ for a $1.''95$ diameter aperture and a signal-to-noise ratio (S/N) of 10. By default, we use the aperture-based magnitude (MAG_AUTO), where the aperture size varies depending on the extent of each source, as reported in the main table of DES. As an additional note, DES tile images have a pixel scale of $0.''263$ and a fixed zeropoint of 30 mag.

We then complement our primary data with near-IR and mid-IR (MIR) photometry from the Visible and Infrared Survey Telescope for Astronomy Hemisphere Survey (VHS) Data Release 5 (McMahon et al. 2013, 2021) and the Wide-field Infrared Survey Explorer (unWISE version; Wright et al. 2010; Schlafly et al. 2019) catalogs, respectively, using a $2''$ crossmatching radius. The VHS Petrosian J -magnitude ($J_{\text{p}}\text{mag}$) and unWISE photometric bands (W1 and W2) are extremely useful for distinguishing between high- z quasars, low- z galaxies, and MLT dwarfs (e.g., Andika et al. 2020, 2022). This complementary data is also an effective way of removing spurious sources, such as diffraction spikes or moving objects, which are often found in one survey, but not in the others. Finally, we correct all of the photometric

measurements from Galactic reddening by making use of the dust map from `Bayestar19`¹² (Schlafly et al. 2019), following the Fitzpatrick (1999) extinction relation. As a side note, we crossmatch our main catalog with a list of MLT dwarfs, employing a 2'' crossmatching radius (Best et al. 2018; Carnero Rosell et al. 2019) and quasars (Flesch 2021) to keep track of the known objects.

2.2. Simulating the Colors of Quasars

We proceed to model the target sources by generating 900 mock quasars. Here, we distribute them in redshifts of $5.6 \leq z \leq 7.2$ and rest-frame 1450 Å absolute magnitudes of $-30 \leq M_{1450} \leq -20$, following the quasar luminosity function from Matsuoka et al. (2018b), in the form of:

$$\Phi(M_{1450}, z) = \frac{10^{k(z-6)} \Phi_*}{10^{0.4(\alpha_{\text{qso}}+1)(M_{1450}-M^*)} + 10^{0.4(\beta_{\text{qso}}+1)(M_{1450}-M^*)}}. \quad (1)$$

We adopt faint- and bright-end slopes of $\alpha_{\text{qso}} = -1.23$ and $\beta_{\text{qso}} = -2.73$, a redshift evolution term of $k = -0.7$, a break magnitude of $M^* = -24.90$, and normalization of $\Phi^* = 10.9$ (Matsuoka et al. 2018b).

We then produce the aforementioned quasar spectra, following the Yang et al. (2016) prescription, implemented with the `SIMQSO`¹³ simulation code (McGreer et al. 2013). `SIMQSO` is designed to produce synthetic quasar spectra that match $\sim 60,000$ Sloan Digital Sky Survey (SDSS; York et al. 2000; Eisenstein et al. 2011; Dawson et al. 2013, 2016; Blanton et al. 2017) quasars at $2.2 < z < 3.5$. The code also assumes that the quasar SEDs do not evolve significantly with redshift (Shen et al. 2019). However, we note that $z \gtrsim 6$ quasars frequently exhibit high-ionization broad lines with more dramatic velocity shifts than other $z \lesssim 5$ quasars of similar luminosity (e.g., Mazzucchelli et al. 2017; Meyer et al. 2019; Schindler et al. 2020). Nonetheless, a large number of low- z SDSS quasars still provide a good reference for building a high- z quasar composite spectrum.

The primary component of our quasar spectral model is continuum emission constructed with a power-law function with four breaks. In the rest wavelength range of 1100–5700 Å, the corresponding continuum slope α_ν follows a Gaussian distribution with a mean of $\mu(\alpha_\nu) = -1.5$. Meanwhile, for the wavelength ranges of 5700–10850 Å, 10850–22300 Å, and > 22300 Å, we change the slopes to $\mu(\alpha_\nu) = -0.48, -1.74$, and -1.17 , respectively. We use a dispersion value of $\sigma(\alpha_\nu) = 0.3$ for each of the aforementioned slopes, following Yang et al. (2016).

Next, we add a series of UV-to-optical Fe emissions to the spectral model, based on the Vestergaard & Wilkes (2001), Tsuzuki et al. (2006), and Boroson & Green (1992) templates for rest-frame wavelengths of < 2200 Å, 2200–3500 Å, and 3500–7500 Å, respectively. After that, the emission lines from the broad- and narrow-line regions are appended to the model, following the equivalent widths and FWHM line distributions of SDSS quasars. Furthermore, the IGM absorption by the Ly α forest is also implemented in the simulated spectra (Songaila & Cowie 2010; Worseck & Prochaska 2011). On top of this, we add the Ly α damping wing model, following

the Miralda-Escudé (1998) formalism, assuming a fixed value of 3 Mpc for the proximity zone size and a randomly drawn neutral hydrogen fraction value in the range of 0%–10% (e.g., Euclid Collaboration et al. 2019; Eilers et al. 2020; Andika et al. 2022). Finally, internal reddening was applied to each mock quasar spectrum, using the Calzetti et al. (2000) dust model, with a randomly chosen $E(B - V)$ value in the range of -0.02 to 0.14 . We should note that the negative reddening values are to account for quasar models with bluer continua than encompassed by the original templates. After this, the photometry is calculated from the simulated spectra, and the corresponding uncertainties are derived from the observed magnitude and error relations taken from each survey (e.g., Yang et al. 2016).

2.3. Modeling the Spectra of Deflector Galaxies

To obtain the deflectors for the lens modeling, we first search for a sample of spectroscopically confirmed galaxies using the SDSS Data Release 17 (Abdurro'uf et al. 2022) catalog, via the `CasJobs`¹⁴ webpage. We retrieve all sources classified as “GALAXY” by the SDSS pipeline and limit our search to those having reliable velocity dispersion measurements σ_v with less than 30% error, which is one important parameter for calculating the lensing effect later. Moreover, we only select the galaxies with redshifts of $0.3 < z < 2.0$, considering that the majority of the lensing optical depth for $z \gtrsim 6$ sources comes from $z \lesssim 1.5$ early-type lens galaxies. (Wyithe et al. 2011; Mason et al. 2015; Pacucci & Loeb 2019; Yue et al. 2022a). After this, we crossmatch these sources to the DES, VHS, and unWISE catalogs, with a searching radius of 1'', to obtain their corresponding optical/IR magnitudes, when available. In the end, we obtain a sample of 109,808 galaxies and query their images accordingly via the `DESaccess`¹⁵ webpage.

As the next step of our method, we need to obtain the spectra of the deflectors that will later be used for constructing the mock galaxy–quasar lens SEDs. Based on the multiband photometric information of the aforementioned galaxies, we proceed to model their spectra using the `Bagpipes`¹⁶ code (Carnall et al. 2018). Several priors on the associated physical parameters need to be supplied to `Bagpipes`. In this case, we first parameterize the star formation histories (SFHs) of the galaxies using a double-power-law function:

$$\text{SFR}(t) \propto \left[\left(\frac{t}{\tau} \right)^{\alpha_{\text{gal}}} + \left(\frac{t}{\tau} \right)^{-\beta_{\text{gal}}} \right]^{-1}, \quad (2)$$

where the falling (α_{gal}) and rising (β_{gal}) slopes range between 0.01 and 1000 in logarithmic space (Carnall et al. 2019). We then vary the peak times of star formation (τ) between 0 Gyr and the universe's age at the observed redshifts. Next, the prior on the SFH normalization is varied in the range of $M_{\text{formed}} = 10\text{--}10^{15} M_*/M_\odot$, where M_* is the stellar mass. Also, we vary the galaxy metallicity in the range of $Z/Z_\odot = 0.0\text{--}2.5$, where $Z_\odot = 0.02$ is solar metallicity.

Following Carnall et al. (2019), we add the Charlot & Fall (2000) dust attenuation model, in the form of $A_\lambda \propto \lambda^{-n}$. Here, we impose a Gaussian prior on the slope of the attenuation curve n , with a mean and standard deviation of 0.3 and 1.5,

¹² <http://argonaut.skymaps.info/>

¹³ <https://github.com/imcgreer/simqso>

¹⁴ <http://skyserver.sdss.org/CasJobs/default.aspx>

¹⁵ <https://des.ncsa.illinois.edu/desaccess/>

¹⁶ <https://bagpipes.readthedocs.io/en/latest/>

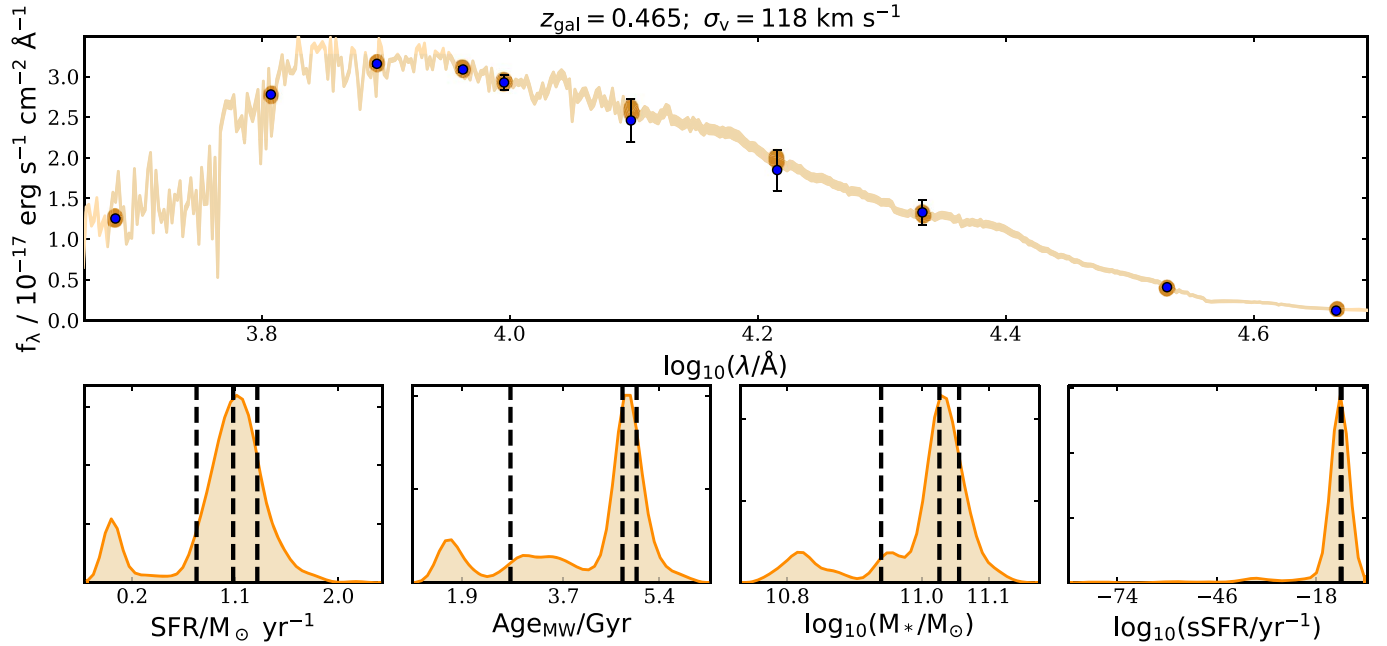


Figure 1. Example of a lens galaxy fitted with the *Bagpipes* code. Upper panel: the DES, VHS, and unWISE photometric data points (blue circles), the posterior of the best-fit SED model (orange line), and synthesized photometry from the posterior SED (orange circles) are shown in the figure. Lower panel: the posterior distribution of the derived star formation rate, mass-weighted age, stellar mass, and specific star formation rate, based on the SED fitting result. The 16th, 50th, and 84th percentiles are marked with the vertical dashed lines. For detailed explanations of how to calculate these quantities, we refer the reader to Carnall et al. (2018).

respectively. The attenuation in the V -band A_V is set to range from 0 to 8. In addition to this, we utilize a constant value for the maximum lifetime of stellar birth clouds, of $t_{BC} = 0.01$ Gyr, and for the ratio of attenuation between stellar birth clouds and the wider interstellar medium, of $\epsilon = 2$. It should be noted that the priors on the galaxy photometric redshifts are set to match the ones derived from SDSS-based spectroscopy, with slight variation (i.e., $\Delta z = \pm 0.015$). Finally, using the σ_v information from the SDSS catalog, we convolve the spectral model with a Gaussian kernel in velocity space. As a result, we produce a sample of SEDs of galaxies ranging from UV to IR wavelengths, and their associated photometry. An example of a lens galaxy fitted with the *Bagpipes* code is presented in Figure 1.

2.4. Constructing the Synthetic Spectra of Galaxy–Quasar Lens Systems

Finding lensed quasar candidates using catalog-level photometry requires knowledge of the shape of their SEDs—i.e., the addition of foreground galaxy and background quasar emissions. Here, we combine the simulated quasars and galaxies described in the previous sections to produce the mock lensed quasar spectra.

To describe the mass profile of the lenses, we assume a singular isothermal sphere (SIS; Schneider 2015) model. The Einstein radius can be inferred from the 1D stellar velocity dispersion σ_v in the potential of the mass distribution, using the formula

$$\theta_E = 4\pi \frac{\sigma_v^2 D_{ds}}{c^2 D_s}, \quad (3)$$

where the speed of light is c , while D_{ds} and D_s are the angular diameter distances between the deflector and the source and the observer and the source, respectively.

It is important to note that galaxy mass distributions in nature are not perfectly symmetric. The symmetry-breaking is often caused by the ellipticity of the mass distribution or external shear forces—e.g., the tidal gravitational fields of adjacent galaxies. Thus, this will alter the SIS lens characteristics, and might produce more than two images. However, we refrain from using a more complex lens model, like the singular isothermal ellipsoid (SIE; Barkana 1998) model, which usually requires accurate measurements of deflector axis ratios and position angles. Nevertheless, SIS appears to reproduce lens systems quite well, and the typical image separation remains in the same order of magnitude as predicted by the SIE model (Schneider 2015). For details of the SIS lens calculation, we refer the reader to Appendix A.

We then produce the mock lensed quasar spectra by pairing each simulated galaxy with a randomly chosen quasar model created beforehand. Assuming that β is the true angular position of the source, the quasar is randomly placed in a specified region behind the lens, within $0''.01 \leq \beta \leq 1''.5$. After this, the source image is projected onto the lens plane, where the deflection angle and magnification are calculated based on the corresponding lens configuration. We accept the mock lens if it contains a strong lensing effect, with a magnification factor of $\mu \geq 2$ within $\theta_E \leq 1''$, and the lensed quasar flux has $>5\sigma$ detection in the DES Y band. Our choice of the β and θ_E ranges is motivated by Pacucci & Loeb (2019), who predict that a significant fraction of quasars at $z \gtrsim 6$ are strongly gravitationally lensed, with small image separations—i.e., $\Delta\theta = 2\theta_E \leq 1''$.

We show in Figure 2 the distributions of the lens galaxy redshifts, velocity dispersions, Einstein radii, and i -band magnitudes used for the simulation. In terms of redshift, the number of galaxies peaks at $z \approx 0.5$. Then, for the i -band magnitudes, we see an increase in the deflector numbers up to $i_{DES} \approx 19.5$, before a rapid decrease toward the faint end. Hence, our training data set is skewed toward brighter, more

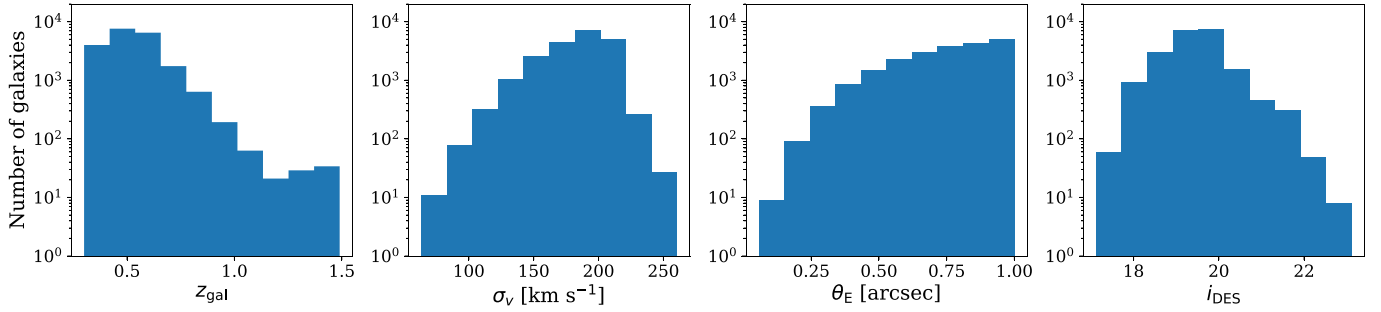


Figure 2. Distributions of the lens galaxy redshifts z_{gal} , velocity dispersions σ_v , Einstein radii θ_E , and DES i -band magnitudes i_{DES} . These parameters are used for generating the mock lenses.

massive lens galaxies. This phenomenon is primarily caused by how SDSS selects target galaxies for its spectroscopic surveys, following the criteria established by Dawson et al. (2013) and Prakash et al. (2016), to study the universe’s large-scale structure. Most of the target galaxies are luminous early-type galaxies at $z < 1$, which are excellent probes for constraining the baryon acoustic oscillation signal and, subsequently, the universe’s expansion rate (e.g., Woodfinden et al. 2022; Zhao et al. 2022). Note that the faint limits for galaxies chosen for spectroscopic observations are $i = 19.9$ for SDSS III, extending to $i = 21.8$ for SDSS IV (Prakash et al. 2016).

Finally, of the initial 109,808 foreground galaxies and 900 background quasars, we obtain 20,823 surviving lens configurations that meet our criteria. Naturally, the lensed quasar spectra can be obtained simply by adding the fluxes of the foreground galaxy and the magnified background quasar, as presented for one example in Figure 3. However, in some cases, the lens galaxy could reside far away from the quasar image, resulting in two close-by objects, so that the photometry extracted from the quasar image only has partial coverage of the lens galaxy flux. To take this into account, we produce more SED templates, by arbitrarily scaling the lens galaxy flux contribution in the range of 10%–100%, before adding it to the magnified quasar flux.

2.5. Lensed Quasar Search via SED Modeling

As the next step in our lensed quasar search methodology, we select sources detected in the DES Y band that have counterparts in the VHS J band and the unWISE W1 band, within a searching radius of $2''$. Note that there are some cases where the magnitude values of the other bluer DES bands—i.e., g , r , i , and z —are not available, or the sources are simply not detected in those bands, i.e., with $S/N_s < 3$. In this case, we replace the catalog magnitudes with their associated 3σ upper limits, where we infer these limits based on the measured flux uncertainties reported in the DES table. We then apply the following flag and color cuts:

$$\text{DES IMAFLAGS_ISO}(g, r, i, z, Y) = 0, \quad (4)$$

$$\text{DES FLAGS}(g, r, i, z, Y) < 4, \quad (5)$$

$$S/N(Y_{\text{DES}}) \geq 5, \quad (6)$$

$$z_{\text{DES}} - Y_{\text{DES}} > 0.5, \quad (7)$$

$$-0.6 < Y_{\text{DES}} - J_{\text{VHS}} < 1.0, \quad (8)$$

$$-0.5 < Y_{\text{DES}} - W1 < 2.5. \quad (9)$$

Equations (4) and 5 are the primary flags for selecting well-behaved objects—i.e., there are no problems in the source extraction process, with missing pixels, etc.—from the DES

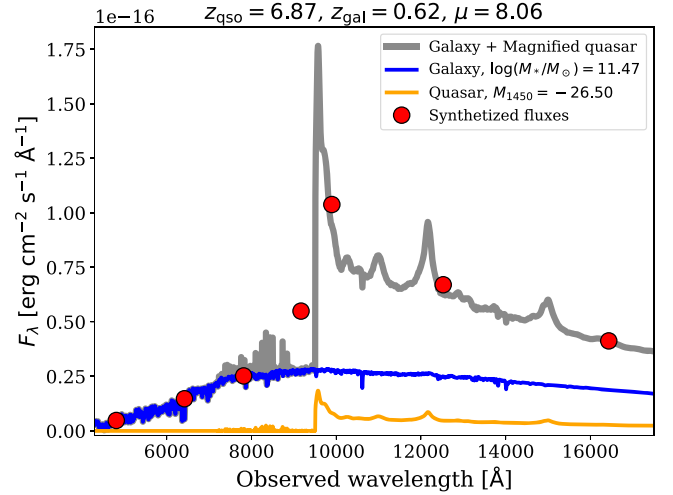


Figure 3. Example of a simulated galaxy–quasar lens spectrum (gray line) and the associated synthetic photometry (red circles). The unlensed background quasar’s emission (orange line) is magnified by a factor of μ , before being added to the foreground galaxy’s emission (blue line), to produce the lensed quasar spectrum.

photometry catalog (Abbott et al. 2021). These criteria also ensure that our targets will have catalog entries in all DES bands. On the other hand, the criteria in Equations (6)–(9) are determined empirically, based on the photometric features of the mock lenses created in Section 2.4.

As seen in Figure 4, there is a substantial overlap between the lenses and contaminants in the color space where $z_{\text{DES}} - Y_{\text{DES}} \leq 0.5$. Thus, we choose to limit our color cut to $z_{\text{DES}} - Y_{\text{DES}} > 0.5$, which consequently focuses our search on the lens systems where the background quasars are located, at $z = 6.6$ – 7.2 . More details of the quasar selection function in our lensed quasar search criteria can be found in Appendix B. To quantify our selection completeness, we start by applying a $z_{\text{DES}} - Y_{\text{DES}} > 0.5$ color cut to the remaining 20,823 simulated lenses produced in the previous section. As a result, only 3657 of them pass this criterion. We then use this number of remaining mock lenses as the basis for estimating our selection completeness in recovering lens systems having quasars at $z = 6.6$ – 7.2 . Further cuts are made by applying the criteria in Equations (6)–(9), which results in 3650 surviving mock lenses, equivalent to around 99% completeness. This choice also rejects a substantial fraction ($\gtrsim 70\%$) of the contaminants—i.e., low- z galaxies and MLT dwarfs. Therefore, these are the candidates on which we will conduct SED modeling for further selection.

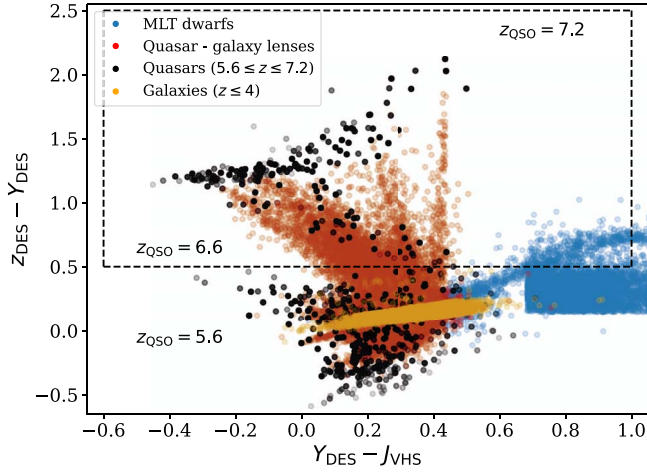


Figure 4. Color diagram of $Y_{\text{DES}} - J_{\text{VHS}}$ vs. $z_{\text{DES}} - Y_{\text{DES}}$. The colors of mock quasars (black circles), low- z early-type galaxies (orange circles), simulated lenses (red circles), and a sample of MLT dwarfs (blue circles; Best et al. 2018; Carnero Rosell et al. 2019) are shown in the figure. Our preselection color cut is marked by the dashed black box, which focuses on finding lensed quasars at $z = 6.6\text{--}7.2$. Note that the seemingly unnatural rectangular shape of the MLT dwarf distribution to the right of the plot is caused by the color cut criteria used by Carnero Rosell et al. (2019), to avoid the quasar color locus.

The next part of our preselection method involves SED modeling, using four different spectral templates: (i) the lensed quasars derived from the previously created mock lenses; (ii) the elliptical galaxies at $z \leq 3$; (iii) the MLT dwarfs; and (iv) the $z \geq 4$ unlensed quasars. It is crucial to note that templates (ii)–(iv) are empirically derived from observations and described in detail in Appendix C. The main goal of this SED modeling is to calculate the probability of each target being a lens or a contaminant, as well as its associated photometric redshift. The SED modeling is implemented using the EAZY photometric redshift code (Brammer et al. 2008). Here, EAZY will step through a grid of spectral templates, fit them to the photometry of the candidates, and try to find the best SED templates. We choose as solutions the best models with the smallest reduced chi-square (χ_{red}^2), which can be calculated for each template i as

$$\chi_{\text{red},i}^2 = \sum_{n=1}^N \left(\frac{\text{data}_n - f_n(\text{model}_i)}{\sigma(\text{data}_n)} \right)^2 / (N - 1), \quad (10)$$

where the number of photometric data points and degrees of freedom are denoted with N and $(N - 1)$, respectively.

The sources with a high probability of being lensed quasars are chosen based on the calculated χ_{red}^2 of the lens ($\chi_{\text{red,L}}^2$), galaxy ($\chi_{\text{red,G}}^2$), MLT dwarf ($\chi_{\text{red,D}}^2$), and quasar ($\chi_{\text{red,Q}}^2$) templates, along with their associated ratios. In addition to this, we also use the estimated quasar photometric redshift z_{qso} to search specifically for $z \gtrsim 6$ unlensed quasars. Hence, the following criteria are employed to find high- z lensed and unlensed quasar candidates:

$$\frac{\chi_{\text{red,L}}^2}{\chi_{\text{red,D}}^2} < 0.2 \quad \text{and} \quad \frac{\chi_{\text{red,L}}^2}{\chi_{\text{red,G}}^2} < 0.2, \quad (11)$$

or

$$z_{\text{qso}} > 6.0 \quad \text{and} \quad \frac{\chi_{\text{red,Q}}^2}{\chi_{\text{red,D}}^2} < 0.3. \quad (12)$$

Here, the values in Equation (11) are derived by fitting the simulated photometric data of the mock lenses created in Section 2.4, which are empirically optimized to maximize the purity and completeness of our lensed quasar selection. On the other hand, the criteria in Equation (12) are inferred empirically, by modeling the SEDs of known MLT dwarfs and $z \gtrsim 6$ quasars in the DES catalog (Carnero Rosell et al. 2019; Flesch 2021). Finally, applying the SED fitting to the previously surviving mock lenses yields 3650 remaining objects, corresponding to a completeness of around 99% for selecting $z = 6.6\text{--}7.2$ lensed quasars. An example of a lens candidate selected using our SED modeling is presented in Figure 5, and a summary of our selection steps is reported in Table 1.

3. Lens Finding Using a CNN

The second part of our lensed quasar search method is based on the supervised neural network classification, which requires realistic training data as inputs. CNNs have been shown to be highly effective in pattern recognition tasks, such as detecting gravitational lenses in large data sets (e.g., Metcalf et al. 2019; Huang et al. 2020; Bom et al. 2022; Gentile et al. 2022; Wilde et al. 2022). The architecture of the CNN usually depends on the problem that needs to be solved. Typically, it consists of images as data inputs, which are then processed by a series of convolutional, pooling, fully connected, and output layers. Here, we present a CNN classifier that is trained to recognize lensed quasars against other nonlensed sources, and the following section will explain the details of our simulation.

3.1. Mock Lens Image Construction

Since only one galaxy–quasar lens system has been found at $z \gtrsim 6$ (Fan et al. 2019), we need to mock up additional lens images. To do this, we adopt a data-driven approach to construct the training data set, which is outlined as follows. First, we create images of mock lenses based on the deflector galaxies, simulated quasars, and lens configurations used in Section 2.4. Then, through the lens equation whose parameters have been defined, we paint lensed point-source lights over the actual galaxy images, to construct realistic galaxy–quasar lens models. In this case, the lensing effect and the ray-tracing simulation are generated according to the PyAutoLens¹⁷ lensing code (Nightingale et al. 2021). The resulting deflected quasar lights are then convolved with a Gaussian PSF model at the location of the lens and coadded with the original DES galaxy images. As a reminder, we accept the mock image if it contains a strong lensing case with a configuration of $0''.01 \leq \beta \leq 1''.5$ and $\mu \geq 2$, within $\theta_E \leq 1''$. To better illustrate this, we present examples of mock color images that have been created by combining the i_zY -band data in Figure 6, and we show the employed simulation workflow in Figure 7.

3.2. Training the Neural Networks

We categorize the input data for training the CNN classifier into four classes: (i) the galaxy–quasar lenses created in the previous section; (ii) the DES galaxies that are not selected in the lensing simulation; (iii) a sample of MLT dwarfs, randomly drawn from the Carnero Rosell et al. (2019) catalog; and (iv) a sample of spectroscopically confirmed quasars at $z \sim 1.5\text{--}7.0$,

¹⁷ <https://pyautolens.readthedocs.io/en/latest/>

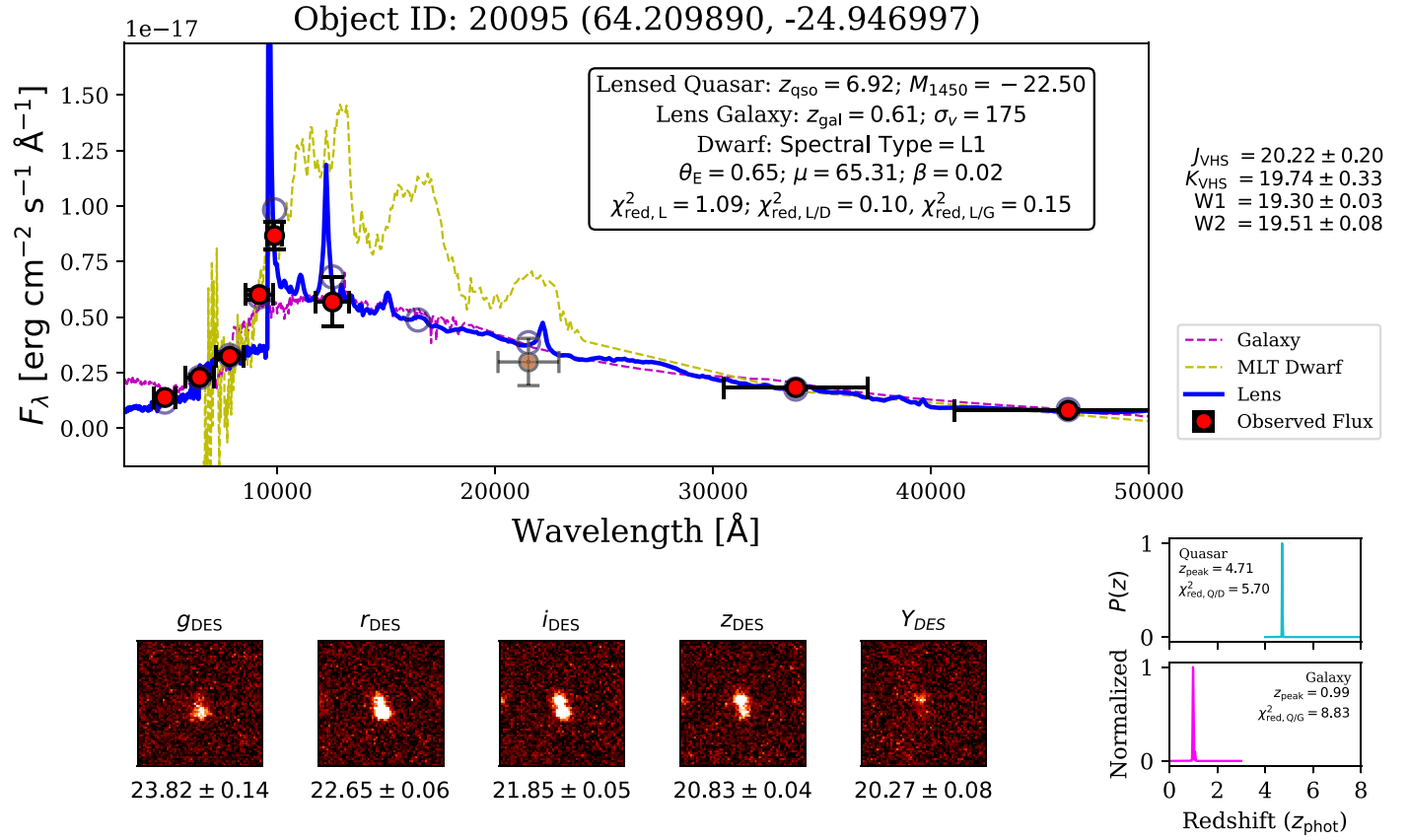


Figure 5. The SED modeling result for a lensed quasar candidate. The upper panel shows the source’s observed photometry (red circles with error bars) fitted with three different templates. The best-fit lens spectral template is shown by the blue line, while the synthesized photometry is denoted by the blue circles. On the other hand, the best-fit models using the MLT dwarf and unlensed low- z galaxy templates are represented by the yellow and magenta colors. We show the photometric redshift probability density functions in the lower right panel, which are derived from fitting the data to unlensed high- z quasar (cyan line) and low- z galaxy (magenta line) templates. DES cutouts, with a size of $16''8 \times 16''8$, are presented in the lower left panels, starting from the left.

Table 1

Summary of the $z \gtrsim 6$ Lensed and Unlensed Quasar Candidate Selection Employed in This Work

Step	Selection	Simulated ^a	Candidates ^b
1	Initial flag criteria ^c and $S/N(Y_{DES}) \geq 5$	20,823	8,274,747
2	$z_{DES} - Y_{DES} > 0.5$	3657	775,369
3	$Y_{DES} - J_{VHS} < 1.0$	3657	735,110
4	$Y_{DES} - J_{VHS} > -0.6$	3657	687,506
5	$Y_{DES} - W1 < 2.5$	3657	685,285
6	$Y_{DES} - W1 > -0.5$	3650	662,314
7	SED modeling	3650	24,723
8	CNN classification	...	448
9	Visual inspection	...	36

Notes. The $z_{DES} - Y_{DES}$ color cut limits the search to focusing on lens systems where the background quasars are located at $z = 6.6-7.2$.

^a The number of selected mock lenses.

^b The number of real (lensed) quasar candidates selected from the DES, VHS, and unWISE catalogs.

^c See the criteria in Equations (4)–(5).

from the Flesch (2021) database. Classes (i) and (ii) each contain 10,000 sources, while classes (iii) and (iv) each consist of 8000 objects. Hence, 36,000 sources in total are used for the CNN inputs.

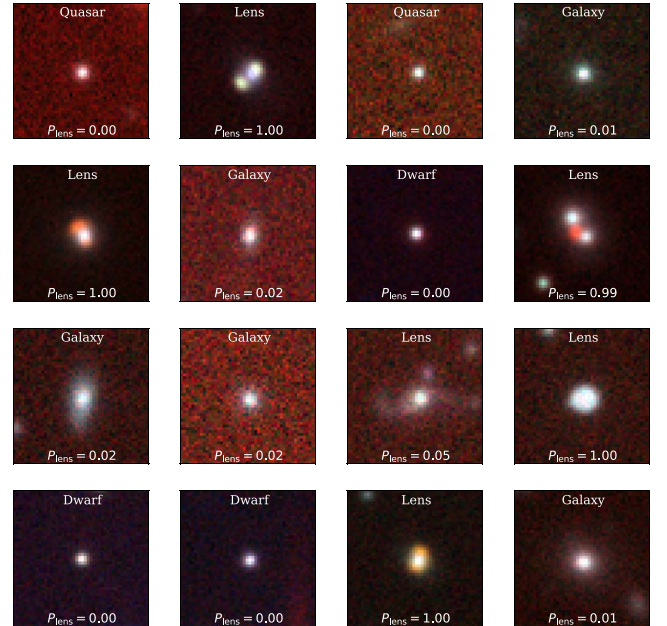


Figure 6. Examples of the negative (galaxies and other point sources) and positive mock lenses in the training data set. The cutouts are created based on $12''6 \times 12''6$ DES izY images, and stretched following the Lupton et al. (2004) arcsinh stretch algorithm, to enhance the contrast and improve the visual appearance. The true labels and predicted probabilities of each source are indicated in each panel.

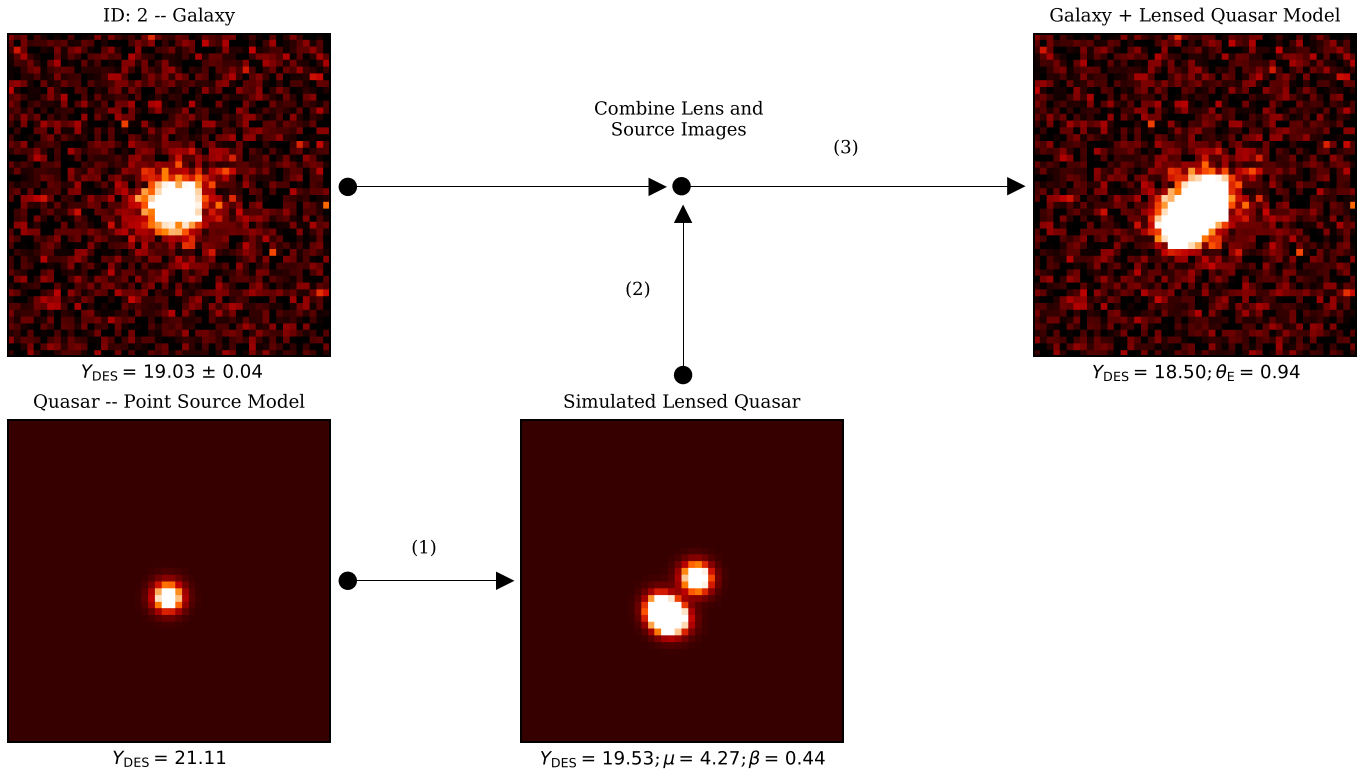


Figure 7. The workflow for simulating lensed quasar images. The cutouts are created based on $12''.6 \times 12''.6$ DES images. The upper left panel shows the actual galaxy, which we use as a lens. The quasar, which acts as a background source, is simulated as a point-source light convolved with a Gaussian PSF, and shown in the lower left panel. The lower middle panel shows the deflected source’s light, which is calculated based on the corresponding lens configuration. Finally, we paint the simulated arcs over an actual galaxy image, as shown in the upper right panel. The lens parameters (i.e., β and θ_E in arcseconds) and the DES photometry are given below each panel. In this work, we create mock images for all DES filters—i.e., the *grizY* bands.

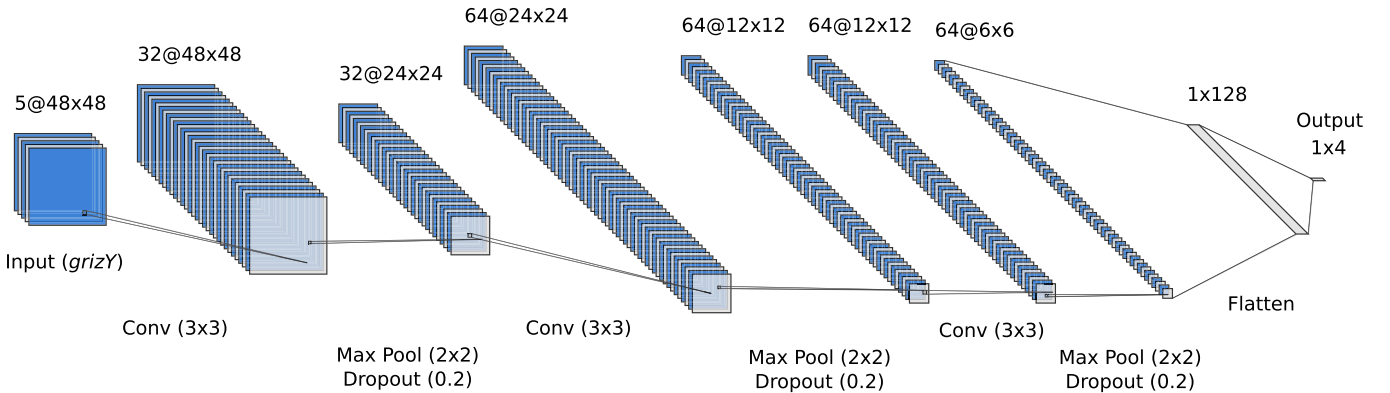


Figure 8. The architecture of the CNN used in this work. The inputs consist of DES *grizY* images, each with a size of 48×48 pixels. The first part of our network contains three sequences of convolutional, max pooling, and dropout layers. The kernel sizes and dropout rates are indicated in the figure. Following this, the data cube will be flattened and will pass through a sequence of fully connected and dropout layers, before it reaches the output layer. The softmax function in the output layer will produce four probabilities for image classification, i.e., P_{lens} , P_{galaxy} , P_{dwarf} , and P_{quasar} .

The images for the training inputs are based on *grizY* bands of DES cutouts, having a size of 48×48 pixels—i.e., equivalent to angular sizes of $12''.6 \times 12''.6$. These images are then min–max normalized, so the fluxes range between 0 and 1. Note that the relative fluxes between bandpasses are conserved, so the spectral colors of the corresponding sources are preserved. After this, we apply data augmentation to the images through random $\pm \pi/2$ rotations, 4 pixel translations, and horizontal or vertical flips. This technique subsequently increases the amount of training data, while enhancing the likelihood that the network will correctly classify multiple orientations of the same image.

Inspired by classical CNN architectures (e.g., LeCun et al. 1989; Sultana et al. 2019), we start building the network with three convolutional layers, followed by one fully connected layer of 128 neurons (see Figure 8). The convolutional layers each have a stride of 1×1 , the “same” padding, and a kernel size of $3 \times 3 \times C$, with $C = 32, 64,$ and 64 for the first, second, and third layers, respectively. For each convolutional layer, we also add a max pooling layer with of size 2×2 , a stride of 2×2 , and the same padding. The dropout regularizations are utilized on both the convolutional (drop rate = 0.2) and fully connected layers (drop rate = 0.5). We also add L2-norm regularization, with a weight decay of 10^{-4} . The learning rate

is set to 10^{-3} initially, while the weight and bias of each neuron are generated randomly and then updated throughout the training. The rectified linear unit functions are employed everywhere, except for the final layer, which utilizes the softmax activation. After passing through all of the convolutional layers, the data cube is flattened and processed by the fully connected layer to obtain the four outputs—i.e., the probabilities of a candidate being a lensed quasar, a galaxy, an MLT dwarf, and a normal quasar. All the training processes and the CNN modeling are implemented using the `TensorFlow`¹⁸ deep learning framework (Abadi et al. 2016; Developers 2022).

The specific CNN architecture mentioned above was chosen after we employed the Hyperband algorithm to optimize the model hyperparameters (Li et al. 2018). To rapidly search for a model with high performance, the Hyperband tuner applies early stopping and the adaptive resource allocation method, utilizing a tournament bracket approach, similar to what is used in a sports competition. In principle, the algorithm trains several CNN architectures with randomly sampled configurations over a few cycles. In other words, we explored various models with different numbers of convolutional layers, kernel sizes, and dense units. We also tested the effects of different learning rates, in the range of 10^{-2} – 10^{-4} , dropout rates of 0.1–0.5, convolutional filter numbers of 16–256, and L2-norm weight decays of 10^{-1} – 10^{-5} on the performance of our CNN models. Ultimately, only the top half of best-performing models is carried forward to the next round, until the best-performing model is obtained. When compared to Bayesian optimization, random search, or grid search, this strategy maximizes the number of evaluated CNN configurations, and results in more effective resource consumption.

To assess network performance, we divide the images into three data sets: training (60% of the data), validation (20%), and testing (20%). Those data sets are further divided into random batches, with sizes of 256. Here, we use sparse categorical crossentropy¹⁹ as the loss function. For each iteration, the networks predict the outputs for one batch (forward propagation), then continue the loops through all the batches from the training data set, to complete one epoch. The information about the current batch loss is then propagated back, to update the weights and biases of the neurons, which is performed by following a stochastic gradient descent algorithm to minimize the loss (i.e., the Adam optimizer; Kingma & Ba 2014). This optimization is performed for each epoch, initially for all of the training data set batches. The average loss for the whole training data set is then obtained. After this, the procedures are repeated for the validation data set, but without parameter updates to the neurons. In this way, the average loss for the validation data set is retrieved. By comparing the training and validation losses, we can check whether the model optimization is improved or whether overfitting happens. Note that overfitting occurs when a network fails to capture the features in the training data set and generalizes them for unseen data. Therefore, we randomly reorder our training data after each epoch, to improve generalization and produce networks with optimum accuracy. In the end, we can determine the optimum number of training cycles by locating the epoch with the lowest average validation loss over numerous runs.

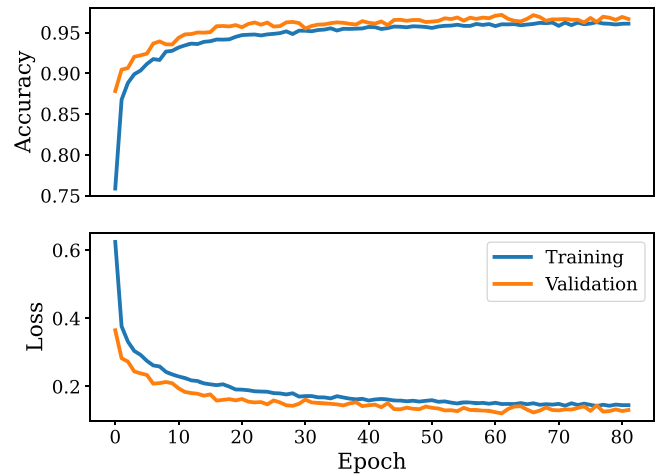


Figure 9. Curves of accuracy (upper panel) and loss (lower panel) as a function of training epoch. These metrics are calculated by evaluating the CNN classifier on the training and validation data sets, which are then shown by the blue and orange lines, respectively.

3.3. Evaluation of the Classifier Performance

For each image tensor that is fed to our CNN model, the classifier will provide probability scores of it being a lensed quasar, a galaxy, an MLT dwarf, and a normal quasar—i.e., P_{lens} , P_{galaxy} , P_{dwarf} , and P_{quasar} . The predicted class is then assigned by choosing which of the classes has the highest probability score. A value of $P_{\text{lens}} = 1$ means that there is a high chance that the classified image contains a lensed quasar. In contrast, $P_{\text{lens}} = 0$ means that the image is not a lensed quasar and is more similar to other contaminants. Note that the softmax activation function that we use in the last layer of our CNN model would produce $P_{\text{lens}} + P_{\text{galaxy}} + P_{\text{dwarf}} + P_{\text{quasar}} = 1$, by construction. In the end, our CNN training process converged after 82 epochs, obtaining 96.1% (96.7%) accuracy for the evaluation using the training (validation) data set, along with a loss value of 0.145 (0.131).

From a classical standpoint, the high accuracy attained in the training data set might have been caused by overfitting. Thus, we compare the accuracy–loss learning curves that were calculated by evaluating the CNN predictions on the training and validation data sets (see Figure 9). After decreasing for several epochs, the training and validation loss values stabilize and follow the same pattern. The lack of any overfitting signals—i.e., the training loss continues to decrease, but, in contrast, the validation loss begins to increase, after many epochs—makes us confident that our CNN classifier is capable of learning and generalizing.

Another method for assessing the overall performance of the trained model is to use the receiver operating characteristic (ROC) curve and to calculate its corresponding area under the curve (AUC). In principle, this shows how well a binary classifier differentiates between two classes when the decision threshold is changed. Thus, we first define the positives (P) as the lenses and the negatives (N) as the nonlenses or contaminants. True positives (TP) are cases where the model correctly predicts the lenses, while true negatives (TN) occur when the nonlenses are correctly identified. On the other hand, false positives (FP) are where the classifier incorrectly predicts contaminants as lenses. In addition, we also define false negatives (FN), as cases where the model incorrectly rejects lenses. The ROC curve presents the false-positive rate (FPR) against the true-positive rate (TPR) for the validation data set,

¹⁸ <https://www.tensorflow.org/>

¹⁹ https://www.tensorflow.org/api_docs/python/tf/keras/losses/SparseCategoricalCrossentropy

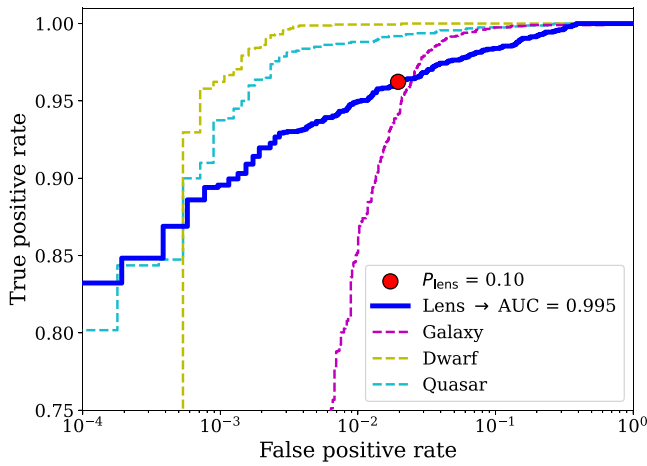


Figure 10. ROC curves and the associated AUC value. The curve for classifying lensed quasars is presented by the solid blue line. On the other hand, the curves for predicting galaxies, MLT dwarfs, and quasars are shown by the dashed magenta, yellow, and cyan lines, respectively. The FPR and TPR values for the adopted P_{lens} thresholds are also shown by the red circle.

where

$$\text{TPR} = \frac{\text{TP}}{\text{P}} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (13)$$

and

$$\text{FPR} = \frac{\text{FP}}{\text{N}} = \frac{\text{FP}}{\text{FP} + \text{TN}}. \quad (14)$$

We then construct the ROC curve by progressively raising the probability threshold from 0 to 1. A perfect classifier will result in $\text{AUC} = 1$, while a classifier that only forecasts randomly will have $\text{AUC} = 0.5$. It is important to note that because we have four classes for categorizing the candidates—i.e., a multilabel classification—we need to binarize our CNN output using the “one versus all” methodology. Accordingly, we produce four ROC curves and present them in Figure 10, consisting of: (i) lensed quasars against galaxies and other point-source contaminants, shown by the solid blue line; (ii) galaxies against lens systems and other contaminants, shown by the dashed magenta line; (iii) MLT dwarfs against other sources, shown by the dashed yellow line; and (iv) quasars against lenses, galaxies, and others, shown by the dashed cyan line. Note that we created these ROC curves based on the CNN classifier evaluation of the previously unseen test data set, which resulted in high AUC values, indicating excellent performance.

Based on the ROC curves, we use the geometric mean or G-mean²⁰ metric to seek a balance between the TPR and FPR. The highest G-mean score indicates the best threshold of P_{lens} that maximizes TPR while minimizing FPR. In this case, we obtain a recommended threshold of $P_{\text{lens}} > 0.13$, which produces $\text{FPR} = 0.01$ and $\text{TPR} = 0.96$. However, we then decide to relax this probability threshold, instead using $P_{\text{lens}} > 0.1$ to classify the candidates as lenses, which results in $\text{FPR} = 0.02$ and $\text{TPR} = 0.96$. This choice is made to increase the final number of lensed quasar candidates, by accommodating candidates with lower lens probabilities.

Below the aforementioned P_{lens} threshold, the number of candidates will grow exponentially, while their quality deteriorates, making visual examination more time-consuming and less effective. Without a reference sample of $z \gtrsim 6$ lensed quasars, it is impossible to determine the perfect threshold. In terms of the trade-off between completeness and purity, we need to find a balance where the number of candidates is manageable for follow-up observations.

As supplementary information, we performed an additional evaluation using an independent data set. This test data set was assembled based on the list of known $z \lesssim 3$ lensed quasars taken from the Gravitationally Lensed Quasar Database²¹ (Inada et al. 2012; Agnello et al. 2015, 2018; Lemon et al. 2018, 2019, 2020; Spiniello et al. 2018; Jaelani et al. 2021). Of the 220 lenses in the database, 34 of the known lenses have DES images. Combined with the contaminants generated in Section 3.2, we tested our CNN classifier on this new data set and successfully recovered 27 lenses—i.e., equivalent to a completeness of 79% and a purity of 11%. We refer the reader to Appendix D for more details on the resulting distribution of probability scores. Compared to the evaluation against the mock lenses, the performance of our CNN model decreases when it is tested against unseen real lens systems. This decreased performance might have been caused by the uniqueness of some of the strong lenses that were not covered by our simulation. In particular, some of the FNs in the test data set might have contained lensed arcs that were too faint to be identified, compact lenses, multiple distortions caused by substructures, or other things. Nonetheless, our CNN classifier can generalize and achieve a sufficiently high accuracy for our purpose. We should emphasize that the point of our CNN-based classification is to conduct a preselection before we go into the visual inspection process, not to assemble a final sample that is quantitatively pure or complete.

4. Results and Discussion

We visually inspected 448 sources—selected by our color cut, SED modeling, and the CNN classifier—which resulted in 36 high-fidelity lens candidates, plus six unlensed quasar candidates, which are awaiting spectroscopic confirmation. More details of this selection step will be presented in the following paragraphs. Based on the evaluation of the test data set in Figure 10, our CNN classifier seems to have an FPR as low as 2% for detecting lensed quasars at $z = 6.6\text{--}7.2$.

But how many high- z lensed quasars are we actually expecting to find? The current observed lensed fraction among $z \sim 6$ quasars ($\approx 0.2\%$; Yue et al. 2022a, 2022b) is substantially smaller compared to previous theoretical estimations ($\gtrsim 4\%$; Wyithe & Loeb 2002; Pacucci & Loeb 2019). A recent study by Yue et al. (2022a) has argued that this tension may be caused by unaccounted-for biases that have not been well investigated. Many prior investigations, in particular, have relied on early measurements of quasar luminosity functions and deflector velocity dispersion functions. By adopting recent estimates, Yue et al. (2022a) suggested that the lensed fraction among $z \sim 6$ quasars was $\sim 0.4\%$ – 0.8% for a survey depth of $z_{\text{DES}} = 22$ mag. Consequently, depending on the assumptions, we expect the number of lensed quasars in an optimal final data set to range from ~ 2 (lensed fraction $\approx 0.6\%$ of approximately 300 known quasars; Yue et al. 2022a) and to more than 10

²⁰ The definition is $\text{G-mean} = \sqrt{\text{TPR} \times (1 - \text{FPR})}$.

²¹ <https://research.ast.cam.ac.uk/lensedquasars/>

Table 2
List of the $z \gtrsim 6$ Lensed and Unlensed Quasar Candidates Discovered in This Work

ID	Name	g_{DES}	r_{DES}	i_{DES}	z_{DES}	Y_{DES}	J_{VHS}	W1	$\chi^2_{\text{red, Q}}$	$\chi^2_{\text{red, G}}$	$\chi^2_{\text{red, L}}$	$\chi^2_{\text{red, D}}$	z_{qso}	z_{gal}	P_{lens}	Grade
262	J056.71853–27.63183	23.78 ± 0.16	22.46 ± 0.07	21.64 ± 0.05	20.86 ± 0.05	20.34 ± 0.10	20.12 ± 0.19	19.42 ± 0.04	37.40	4.57	0.48	13.90	6.83	0.67	0.29	A1
346	J036.96123–50.21938	>25.11	23.83 ± 0.16	22.78 ± 0.10	21.69 ± 0.08	21.06 ± 0.17	20.60 ± 0.34	20.03 ± 0.05	11.32	2.82	0.18	4.17	6.84	0.58	0.15	A1
408	J040.71110–51.68291	23.14 ± 0.09	22.08 ± 0.05	21.17 ± 0.04	20.57 ± 0.04	20.06 ± 0.09	19.86 ± 0.17	19.10 ± 0.02	63.09	4.69	0.67	33.55	6.90	0.78	0.28	A1
1537	J064.46980–55.90677	24.44 ± 0.20	23.38 ± 0.09	22.57 ± 0.08	21.85 ± 0.08	21.13 ± 0.14	21.27 ± 0.33	21.05 ± 0.10	21.19	4.05	0.27	2.40	6.90	0.62	0.15	A1
2441	J030.16139–23.27101	24.35 ± 0.23	23.02 ± 0.09	22.13 ± 0.07	21.17 ± 0.05	20.63 ± 0.09	20.27 ± 0.20	19.54 ± 0.04	34.77	9.12	1.32	13.76	6.84	0.52	0.12	A1
3391	J092.60966–35.90861	23.90 ± 0.34	22.61 ± 0.13	21.45 ± 0.07	20.83 ± 0.07	20.20 ± 0.11	20.07 ± 0.24	19.60 ± 0.04	14.82	3.69	0.34	14.10	7.09	0.79	0.19	A1
4580	J041.48368–43.88192	24.70 ± 0.19	23.41 ± 0.08	22.87 ± 0.08	21.98 ± 0.06	21.31 ± 0.12	20.98 ± 0.25	20.33 ± 0.07	28.60	3.63	0.48	4.56	6.80	0.62	0.23	A1
6253	J322.05036–45.50413	24.23 ± 0.18	23.02 ± 0.08	21.69 ± 0.04	20.96 ± 0.04	20.42 ± 0.07	20.43 ± 0.40	19.83 ± 0.05	37.81	6.00	0.81	7.75	6.90	0.85	0.17	A1
6514	J028.58579–52.33668	25.23 ± 0.32	23.81 ± 0.12	22.93 ± 0.10	22.11 ± 0.10	21.52 ± 0.18	21.27 ± 0.27	20.45 ± 0.07	13.41	1.57	0.20	3.05	6.81	0.58	0.22	A1
7267	J025.08084–06.38508	23.21 ± 0.11	22.89 ± 0.11	22.02 ± 0.07	21.22 ± 0.07	20.72 ± 0.14	20.60 ± 0.23	20.09 ± 0.07	41.18	3.08	0.49	6.12	7.00	0.92	0.26	A1
8189	J068.45382–33.48967	>24.91	23.93 ± 0.17	22.83 ± 0.12	21.82 ± 0.08	21.13 ± 0.14	20.99 ± 0.29	20.11 ± 0.06	13.84	1.26	0.09	3.94	6.84	0.65	0.21	A1
8805	J085.89481–57.30598	24.19 ± 0.18	22.65 ± 0.05	22.10 ± 0.05	21.49 ± 0.05	20.97 ± 0.11	20.37 ± 0.21	20.85 ± 0.08	40.17	2.25	0.38	2.96	7.10	0.44	0.40	A1
10119	J345.07684–58.84036	22.97 ± 0.08	22.45 ± 0.07	21.58 ± 0.05	20.79 ± 0.06	20.15 ± 0.09	20.35 ± 0.34	19.92 ± 0.06	50.77	6.58	0.53	4.16	6.95	0.78	0.90	A1
10501	J036.23405–40.28176	24.26 ± 0.26	22.03 ± 0.04	20.98 ± 0.03	20.24 ± 0.03	19.72 ± 0.07	19.07 ± 0.11	18.81 ± 0.02	54.27	25.21	3.94	48.14	6.80	0.55	0.97	A1
11107	J042.68174–06.34957	24.66 ± 0.33	23.12 ± 0.11	22.43 ± 0.09	21.56 ± 0.09	20.91 ± 0.19	20.72 ± 0.22	19.79 ± 0.05	20.99	2.42	0.39	5.36	7.06	0.53	0.32	A1
11972	J079.81120–25.37072	24.94 ± 0.24	23.81 ± 0.11	23.24 ± 0.10	22.16 ± 0.08	21.18 ± 0.10	21.18 ± 0.32	20.44 ± 0.09	30.62	6.09	0.41	4.53	7.07	0.62	0.12	A1
13878	J061.07527–33.99544	24.03 ± 0.17	22.92 ± 0.08	22.25 ± 0.08	21.27 ± 0.06	20.52 ± 0.11	20.60 ± 0.26	20.08 ± 0.06	24.12	5.07	0.38	6.91	6.98	0.60	0.25	A1
14097	J008.37130 + 01.30827	22.31 ± 0.08	21.13 ± 0.04	20.34 ± 0.03	19.45 ± 0.02	18.56 ± 0.04	18.98 ± 0.12	18.49 ± 0.02	153.40	52.81	4.23	28.67	6.90	0.61	1.00	A1
15233	J025.04989–19.22226	24.28 ± 0.22	23.16 ± 0.10	22.25 ± 0.07	21.60 ± 0.09	21.04 ± 0.14	21.07 ± 0.31	20.22 ± 0.08	13.85	1.41	0.15	9.05	6.90	0.78	0.11	A1
15256	J026.51434–18.52573	24.79 ± 0.23	23.65 ± 0.11	22.59 ± 0.07	21.81 ± 0.06	21.29 ± 0.15	21.03 ± 0.33	20.75 ± 0.13	18.11	2.52	0.14	1.80	6.90	0.74	0.20	A1
16151	J022.21211–22.07667	24.54 ± 0.23	23.23 ± 0.08	22.48 ± 0.08	21.89 ± 0.09	21.24 ± 0.16	21.23 ± 0.26	20.56 ± 0.11	16.43	1.26	0.15	6.72	7.10	0.61	0.23	A1
16632	J003.79587–43.48657	24.92 ± 0.35	23.11 ± 0.09	21.80 ± 0.04	21.00 ± 0.04	20.28 ± 0.05	20.20 ± 0.24	19.34 ± 0.03	42.46	8.95	1.26	16.23	7.06	0.74	0.21	A1
16649	J004.69657–43.67283	24.53 ± 0.22	23.83 ± 0.15	22.98 ± 0.12	21.76 ± 0.08	21.22 ± 0.13	20.84 ± 0.23	20.27 ± 0.07	17.12	3.05	0.37	3.24	6.78	0.79	0.20	A1
17194	J067.87033–20.39097	24.00 ± 0.22	22.51 ± 0.08	21.88 ± 0.07	21.38 ± 0.09	20.78 ± 0.17	20.36 ± 0.18	20.23 ± 0.07	20.23	1.25	0.24	8.40	7.04	0.34	0.13	A1
17195	J067.92180–21.13860	23.22 ± 0.13	21.78 ± 0.05	21.03 ± 0.04	20.10 ± 0.03	19.29 ± 0.05	19.60 ± 0.21	18.88 ± 0.02	76.95	23.61	1.30	24.89	7.05	0.67	0.63	A1
18165	J089.11616–24.83778	23.45 ± 0.15	22.08 ± 0.05	21.31 ± 0.05	20.61 ± 0.05	20.07 ± 0.09	19.85 ± 0.20	19.24 ± 0.03	37.01	3.57	0.41	18.86	7.02	0.60	0.46	A1
19214	J331.94605–62.27649	22.95 ± 0.08	21.89 ± 0.04	21.26 ± 0.04	20.72 ± 0.04	20.21 ± 0.09	20.07 ± 0.22	19.48 ± 0.04	80.55	3.56	0.44	21.17	6.77	0.64	0.61	A1
19768	J307.41039–47.88890	23.72 ± 0.14	22.54 ± 0.07	21.52 ± 0.05	20.72 ± 0.04	20.10 ± 0.07	20.62 ± 0.44	19.20 ± 0.03	43.66	7.24	1.36	18.36	6.81	0.64	0.48	A1
20095	J064.20989–24.94700	23.82 ± 0.14	22.65 ± 0.06	21.85 ± 0.05	20.83 ± 0.04	20.27 ± 0.08	20.22 ± 0.20	19.30 ± 0.03	63.35	7.18	1.09	11.12	6.92	0.61	0.63	A1
20712	J013.01069–00.00320	>24.94	23.07 ± 0.09	22.37 ± 0.08	21.80 ± 0.10	20.35 ± 0.08	20.68 ± 0.21	20.38 ± 0.10	25.31	13.21	2.64	14.90	7.19	0.44	0.33	A1
21672	J332.80833–43.52889	>25.33	24.30 ± 0.19	23.33 ± 0.13	22.42 ± 0.13	21.40 ± 0.15	21.42 ± 0.30	21.31 ± 0.18	10.17	3.65	0.35	1.78	6.87	0.69	0.24	A1
21837	J059.49147–51.51484	23.83 ± 0.13	23.15 ± 0.09	22.33 ± 0.08	21.65 ± 0.09	20.91 ± 0.15	21.05 ± 0.28	20.15 ± 0.05	23.37	2.57	0.48	11.33	6.85	0.89	0.17	A1
23684	J060.23733–21.39838	23.96 ± 0.16	22.88 ± 0.07	22.21 ± 0.07	21.23 ± 0.06	20.47 ± 0.10	20.65 ± 0.26	19.82 ± 0.05	44.95	6.79	1.07	10.04	6.81	0.52	0.11	A1
23839	J041.09428–01.94383	24.03 ± 0.16	22.65 ± 0.06	21.77 ± 0.05	21.14 ± 0.05	20.50 ± 0.10	20.71 ± 0.33	20.31 ± 0.09	39.75	4.41	0.48	5.00	6.83	0.64	0.53	A1
24021	J044.34665–45.28386	24.45 ± 0.31	22.78 ± 0.09	21.68 ± 0.06	20.78 ± 0.05	19.94 ± 0.07	20.24 ± 0.31	19.04 ± 0.02	45.89	6.05	1.10	18.41	6.98	0.67	0.67	A1
24140	J084.70004–19.84469	25.07 ± 0.33	23.54 ± 0.10	22.71 ± 0.08	21.73 ± 0.06	21.18 ± 0.14	21.00 ± 0.32	20.71 ± 0.11	20.06	4.84	0.13	1.27	6.72	0.56	0.12	A1
5145	J051.27758–18.82994	>25.47	>25.17	>24.48	22.76 ± 0.17	21.48 ± 0.21	21.03 ± 0.25	21.05 ± 0.15	0.21	6.20	0.66	1.40	6.81	1.48	0.01	A2
9128	J358.77772–49.96887	>24.85	>24.43	>23.88	21.22 ± 0.07	20.23 ± 0.07	20.21 ± 0.23	20.41 ± 0.09	1.24	33.94	2.51	12.07	6.73	0.92	0.02	A2
11215	J080.61260–42.70559	>25.47	>25.22	>24.53	21.91 ± 0.06	21.09 ± 0.11	20.96 ± 0.28	20.10 ± 0.05	0.88	16.38	2.26	3.46	6.81	1.48	0.00	A2
14364	J002.72340–59.61451	>25.81	>25.56	>24.78	22.76 ± 0.11	22.15 ± 0.20	21.61 ± 0.35	21.07 ± 0.12	0.10	7.98	1.19	0.64	6.72	0.86	0.03	A2
16566	J070.86138–20.09059	>25.38	>25.09	>24.45	21.37 ± 0.05	20.65 ± 0.08	20.13 ± 0.21	19.75 ± 0.05	0.46	32.69	3.07	4.87	6.72	1.15	0.20	A2
20453	J080.68560–33.47577	>24.86	>24.55	>24.01	22.16 ± 0.12	21.21 ± 0.16	21.43 ± 0.43	20.39 ± 0.07	0.87	4.96	0.21	3.24	6.80	0.76	0.01	A2

Note. Column (1): a unique identifier for each candidate. Column (2): source name. Columns (3)–(7): DES $grizY$ -band magnitudes and associated 1σ uncertainties. Column (8): VHS J -band magnitude. Column (9): unWISE W1-band magnitude. Columns (10)–(13): reduced chi-square values for the quasar (Q), galaxy (G), lens (L), and MLT dwarf (D) template models. Columns (14)–(15): photometric redshift estimates of the background and foreground sources, based on the best-fit SED fitting template. Column (16): CNN classification score. Column (17): grade after visual inspection. The reported magnitudes are in the AB system, corrected for Galactic reddening, by making use of the dust map from Schlafly et al. (2019) and following the Fitzpatrick (1999) extinction relation. If the sources are not detected—i.e., $S/N < 3$ —in certain bands, we replace the catalog magnitudes with their associated 3σ upper limits, where we infer these limits based on the measured flux uncertainties reported in the DES table. Lens candidates are marked with the “A1” grade, while unlensed quasar candidates are marked “A2.” To name the sources, we adopt the “JRRR.rrrr+DD.ddddd” convention, where “RRR.rrrr” and “+DD.ddddd” are the R.A. and decl. in decimal degrees (J2000), respectively.

(lensed fraction $>4\%$; Pacucci & Loeb 2019). However, we should also take into account previous spectroscopic campaigns, which indicate that the success rate of quasar identification at $z \gtrsim 6$ is around 20%–30% (Bañados et al. 2016; Wang et al. 2019; Andika et al. 2020). We expect this efficiency to improve as we find more high- z quasars and our training data sets expand in size. Nonetheless, our combined SED fitting and CNN classification method is a promising way of preselecting high- z lens candidates and saving a lot of time before the human inspection process needs to be carried out.

Previous high- z quasar searches have so far only found a single lens. The cause of this discrepancy is rather obvious in hindsight. Most candidate selection approaches have applied additional magnitude cuts or full “dropout” criteria at all bandpasses bluer than the Ly α line (e.g., Mazzucchelli et al. 2017; Reed et al. 2019; Wang et al. 2019; Andika et al. 2020). This decision is understandable, since the emission of $z \gtrsim 6$ quasars at wavelengths blueward of Ly α is strongly absorbed by the intervening IGM, creating a strong break in the spectrum and becoming a primary identifier for candidate preselection. In other words, for lensed quasars at these redshifts, we do not expect any significant flux to come from the DES g or r bands. However, this is not the case for lens galaxies at redshifts typically between $0.5 < z < 2$, which could contribute significant emission at $\lambda_{\text{obs}} \lesssim 8000 \text{ \AA}$.

So, to include potential lens systems, we had to remove such dropout criteria. Instead, to recap, we used the selection process described above, which hinges on downselecting candidates by using spatial information that carries the typical signatures of lens systems or the lack thereof. From the approximately 8 million sources that exist in the combined DES, VHS, and unWISE catalogs, 662,314 of them passed our flag, S/N, and color criteria. Applying the SED modeling to the previously color-selected lens candidates then yielded 24,723 remaining sources—avoiding discrimination against the flux contributions from the potential lens galaxies. After this, the CNN classifier then reduced the number of candidates to only 448 sources, using image color and geometry information. Finally, visual inspection was performed on the images to discard spurious sources, such as moving objects, hot pixels, CCD artifacts, etc. A summary of all the selection steps that we used can be viewed in Table 1. Also, the list of the candidates is reported in Table 2, while the corresponding composite color images and the distributions of the lens candidate properties are shown in Figure 11 and Figure 12, respectively.

Testing our approach against previous quasar candidate selection methods confirms the broadened selection space. If we were to impose an additional cut of $S/N(g_{\text{DES}}, r_{\text{DES}}) < 5$, or, effectively, the dropout criteria (e.g., Bañados et al. 2016; Reed et al. 2019; Yang et al. 2019b), then none of the mock lenses created in Section 2.4 would survive. In other words, this cut would result in us missing all of the lens systems containing bright galaxies as deflectors.

The crossover point with previous quasar selections only comes when the deflecting galaxies are faint and, consequently, less massive. These systems make compact lenses with small Einstein radii and therefore have small image separations. Subsequently, these sources are likely to have a light dominated by the quasar emission, with a slightly extended shape—at least in ground-based imaging data. And because the images of the deflector galaxy and the lensed quasar are blended, we anticipate that neither a PSF nor a normal Sérsic

profile will offer a sufficient picture of their morphology (e.g., Fan et al. 2019; Yue et al. 2022b), so using the morphological information is difficult. When they are below some lensing mass, they will be picked up by conventional selection methods (e.g., Pons et al. 2019; Reed et al. 2019; Andika et al. 2020) and our SED-based selection (see Section 2.5), although their lensing natures would be challenging to prove. Of course, running this diagnostic with space-based, higher-resolution imaging data would allow for the applicable parameter space to be increased substantially.

While the approach described above seems to produce sensible candidate samples, we see room for improvement. Currently, the rate of FP-classified morphologies is not zero. However, trained astronomers can quickly dismiss the vast majority of previously unaccounted-for contaminants in this candidate list. In principle, the FPs in our list are sources that we see as being highly unlikely to be lenses or that have no visible signs of strong gravitational lensing features, as displayed in Figure 13. They span a wide range of apparent morphologies—for example, irregular sources or spiral arms that mimic lensing arcs. Note that sources with highly irregular shapes that do not belong to either class in the training data set will receive unpredictable CNN scores. We also note point-source lights located near bright stars with strong diffraction spikes or in regions with unreliable photometric data, due to the presence of CCD artifacts, whose colors mimic lens system SEDs, as presented in Figure 14. Moreover, in some cases, we find that sources with low surface brightness in the DES Y band do not give sufficient information to clearly identify the lensing features. We conclude that further research into deeper and more sophisticated CNN models may be worthwhile. The same is true for the expansion of the simulation input: moving beyond the SIS model and adding the external shear perturbation might create subtle differences in relation to the currently more simple models. In any case, the fact that we have managed to expand the search parameter space by avoiding a dropout criterion, while keeping the rejection rate high, shows that the approach works in principle. We will now start to follow up on these high-probability candidates, to test their quasar and lensing natures, using our upcoming observations with the Focal Reducer and Low Dispersion Spectrograph 2 instrument mounted on the Very Large Telescope at the European Southern Observatory (110.243U.002; PI: I.T. Andika).

5. Summary and Conclusion

In this work, we exploit DES, VHS, and unWISE data and perform a systematic search for $z \gtrsim 6$ lensed quasars. Our method consists of two primary steps. First, we preselect the candidates based on their spectral color, using catalog-level multiband photometry, reducing the number of sources from ≈ 8 million to just 662,314. Second, we calculate the relative probabilities of the candidates being a lens or contaminant, using a CNN-based classification, resulting in 448 surviving candidates. Note that the construction of the training data set for the CNN input is achieved by painting deflected point-source lights over actual DES galaxy images. This way, we can create strong-lens simulations realistically and focus on finding systems that have small image separations—i.e., Einstein radii of $\theta_E \leq 1''$. After visually inspecting the sources with a CNN score of $P_{\text{lens}} > 0.1$, we obtain 36 newly discovered lens candidates awaiting spectroscopic confirmation. In addition to

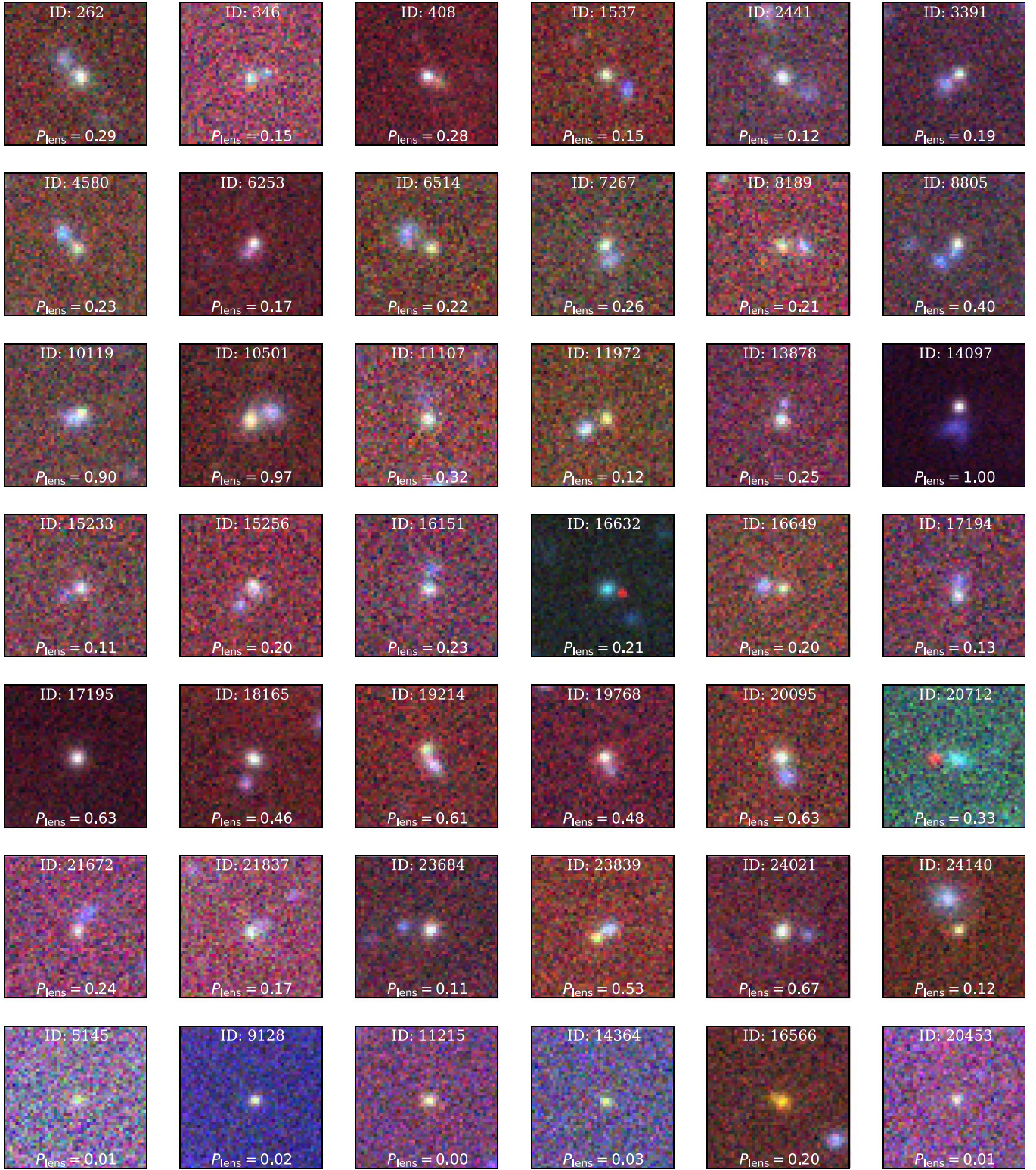


Figure 11. Color images of the candidates found in this paper. The first six rows from the top show the lens candidates, while the last rows in the bottom present the unlensed quasar candidates. The cutouts are created based on $12''6 \times 12''6$ DES i_zY images. Arcsinh stretch image normalization, following the Lupton et al. (2004) approach, is applied to enhance contrast and improve visual appearance. The object identifiers are shown at the top of each panel. In addition, we list the CNN-based classification scores (P_{lens}) at the bottom of each cutout.

this, we also find six new unlensed quasar candidates. These results show that automated SED modeling and CNN pipelines, supported by modest human input, are promising for detecting strong lenses from large databases.

The strategy described in this work is easily adaptable to searching for lens systems at various redshifts, and is very well suited to next-generation surveys such as Euclid (Laureijs et al. 2011; Euclid Collaboration et al. 2022)—providing high-

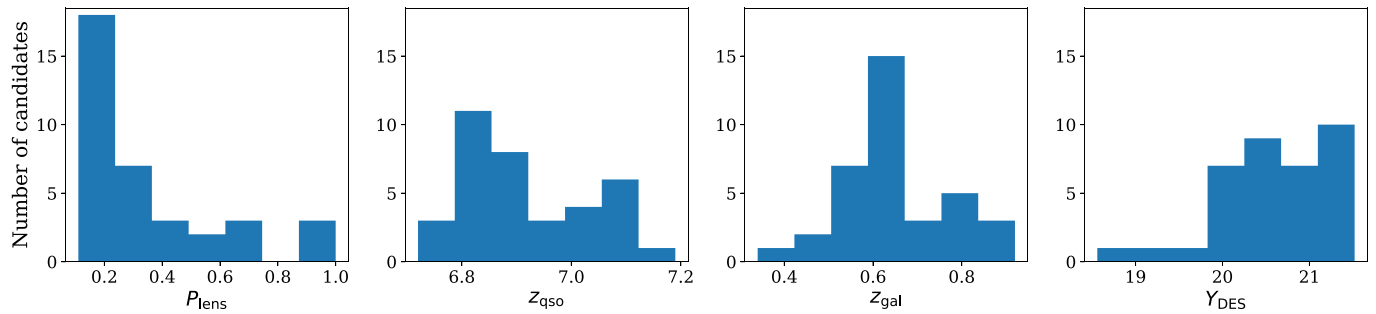


Figure 12. Distributions of the lens candidate CNN scores P_{lens} , source redshifts z_{qso} , deflector redshifts z_{gal} , and DES Y-band magnitudes Y_{DES} . The z_{qso} and z_{gal} parameters are inferred based on the best-fit templates from our SED modeling.

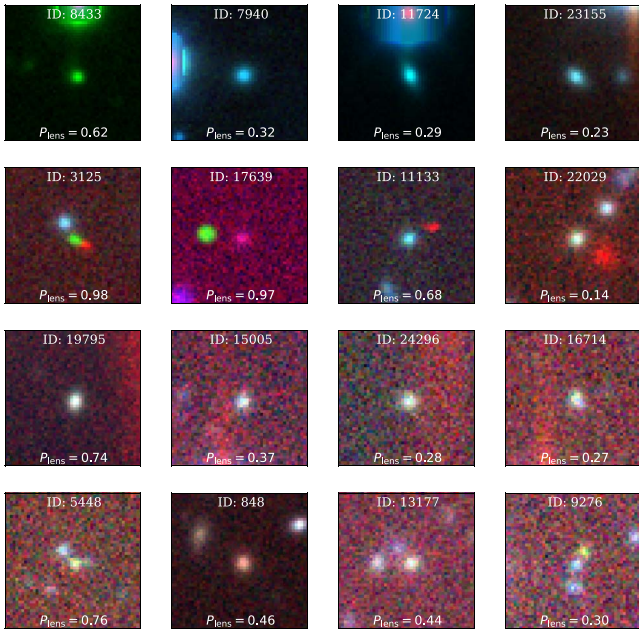


Figure 13. Examples of FP sources, whose colors mimic lensed quasar SEDs, as selected by our CNN classifier. Starting from the top row, we show cases where the sources are located near bright stars with strong diffraction spikes, moving objects, and regions with unreliable photometric data, due to the presence of CCD artifacts, as well as some candidates with less convincing strong lensing features. The cutouts are produced based on $12''6 \times 12''6$ DES iZY images and colorized following the Lupton et al. (2004) method. The predicted probabilities of each source are indicated in each image.

resolution imaging data over a large part of the extragalactic sky—and the Rubin Observatory Legacy Survey of Space and Time (Ivezić et al. 2019), with its very deep photometric multiband data. In principle, adjustments to the filter profiles, the seeing distribution, and the image resolution, in order to match the target surveys, will be required to achieve good results. Future searches will also benefit from extending the diversity of the SEDs and morphologies for the sources and lenses in the training data set. Using more realistic galaxy mass profiles, like SIE, may also potentially improve the classifier performance. Moreover, experimenting with other network models, like EfficientNet and ResNet, might produce better results compared to the currently used classical CNN architecture (e.g., Cañameras et al. 2020; Rojas et al. 2022). These free parameters may become more constrained when additional lenses are uncovered. Realizing the scientific potential of our catalog of lensed quasar candidates will require spectroscopic observations, for measuring the lens and

source redshifts, as well as higher-resolution imaging, to perform lens modeling.

We thank the anonymous referees for the constructive comments on the manuscript. We would like to thank Jochen Heidt, Maarten Baes, Ilse De Looze, Archisman Ghosh, and Toon Verstraelen for fruitful discussions and for providing valuable insights for improving this manuscript. I.T.A. acknowledges the support from the International Max Planck Research School for Astronomy and Cosmic Physics at the University of Heidelberg (IMPRS-HD), which funded this project. J.-T.S. acknowledges funding through the European Research Council (ERC), under the European Union’s Horizon 2020 research and innovation program (grant agreement No. 885301). A.T.J. is supported by the Riset Institut Teknologi Bandung 2021. M.O. is supported by the National Natural Science Foundation of China (12150410307).

This project has used public archival data from the Dark Energy Survey (DES). Funding for the DES Projects has been provided by the DOE and NSF (USA), MISE (Spain), STFC (UK), HEFCE (UK), NCSA (UIUC), KICP (University of Chicago), CCAPP (Ohio State), MIFPA (Texas A&M), CNPQ, FAPERJ, FINEP (Brazil), MINECO (Spain), DFG (Germany), and the collaborating institutions in the Dark Energy Survey. Based on observations at the Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy (AURA), under a cooperative agreement with the National Science Foundation.

This project has included data from the Sloan Digital Sky Survey (SDSS). Funding for SDSS IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS IV acknowledges support and resources from the Center for High Performance Computing at the University of Utah. The SDSS website is <https://www.sdss.org/>. SDSS IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration.

This project has made use of the data obtained as part of the VISTA Hemisphere Survey, ESO Program, 179.A-2010 (PI: McMahon).

The unWISE catalog utilized in this paper is based on data products from the Wide-field Infrared Survey Explorer, which is a joint project of the University of California, Los Angeles, and NEOWISE, which is a project of the Jet Propulsion Laboratory/California Institute of Technology. WISE and NEOWISE are funded by the National Aeronautics and Space Administration.

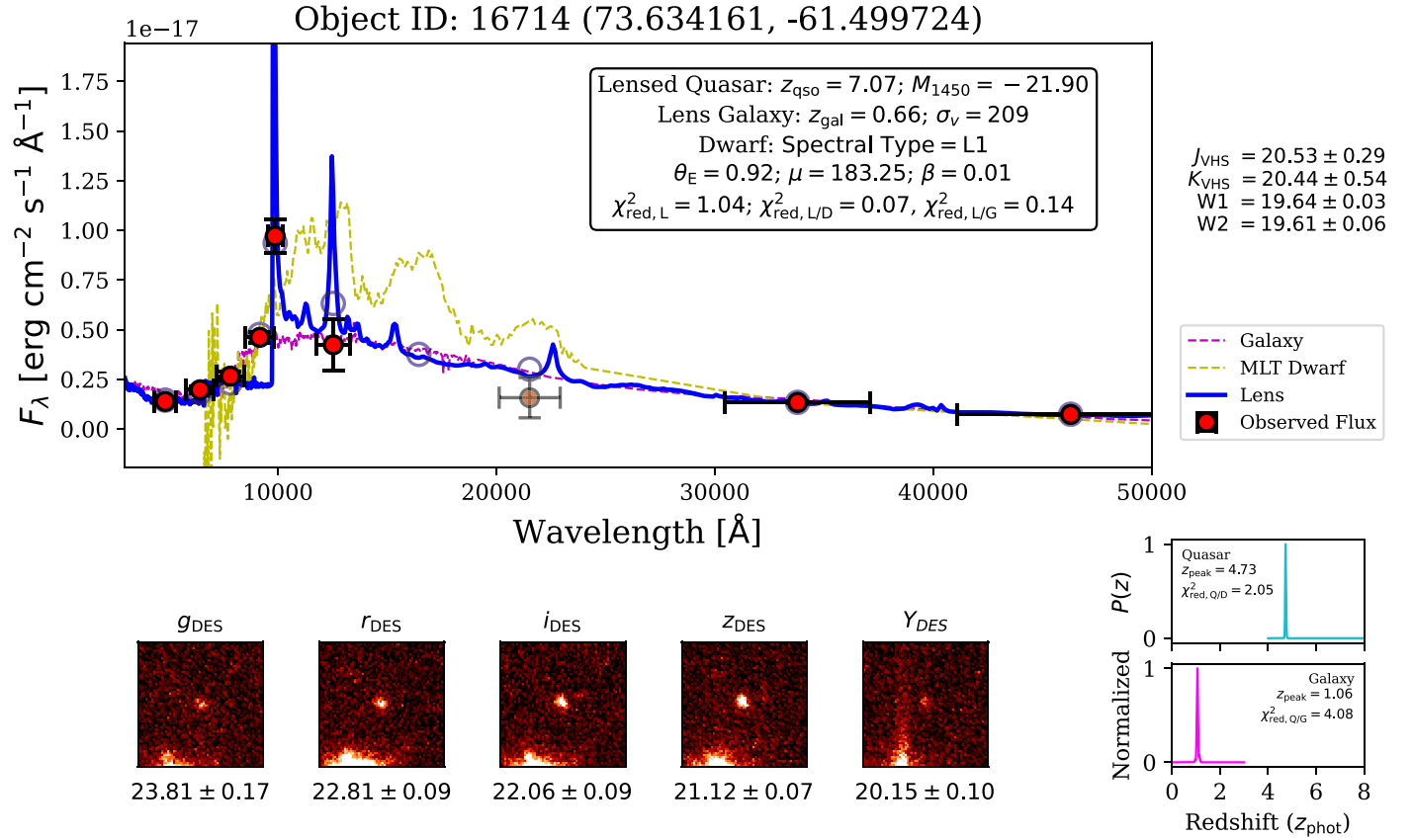


Figure 14. Example of an SED modeling result for an FP candidate. In this case, the source is located near a bright object with strong diffraction spikes or a CCD artifact, but it still produces colors that mimic lensed quasar colors. The upper panel displays the source’s photometric data points (the red circles with error bars) fitted with three distinct templates. The best-fit lens spectral template is shown by the blue line, while the synthesized photometry is represented by the blue circles. On the other hand, the best-fit models using the MLT dwarf and unlensed low- z galaxy templates are denoted by the yellow and magenta colors. We indicate the photometric redshift probability density functions in the lower right panel, which are inferred by fitting the data to unlensed high- z quasar (cyan line) and low- z galaxy (magenta line) templates. DES cutouts, with a size of $16''8 \times 16''8$, are shown in the lower panels, starting from the left.

Facilities: Blanco (DECam), ESO:VISTA (VIRCAM), Sloan (eBOSS/BOSS), WISE.

Software: Astropy (Astropy Collaboration et al. 2013, 2018), Bagpipes (Carnall et al. 2013, 2018), CosmoCalc (Wright 2006), EAZY (Brammer et al. 2008), Matplotlib (Caswell et al. 2019), NumPy (Harris et al. 2020), Pandas (Reback et al. 2021), Photutils (Bradley et al. 2021), PyAutoLens (Nightingale et al. 2021), SIMQSO (McGreer et al. 2013), SciPy (Virtanen et al. 2020).

Appendix A SIS Lens Equation

SIS is a simple analytical formula for modeling the massive galaxy mass profiles that produce strong lenses (e.g., Treu 2010, and references therein). The SIS profile can be expressed as

$$\beta = \theta - \theta_E \frac{\theta}{|\theta|}. \quad (\text{A1})$$

Here, we define θ_E as the Einstein radius, β as the true angular position of the source, and θ as the observed position of the source on the sphere relative to the center of the lens (for reference, see Figure 2.31 of Schneider 2015). If $|\beta| < \theta_E$, two solutions of the lens equation exist, meaning that it will produce two images at angular positions of

$$\theta_1 = \beta + \theta_E, \quad \theta_2 = \beta - \theta_E. \quad (\text{A2})$$

The separation of those two images does not depend on the position of the source and can be calculated using

$$\Delta\theta = \theta_1 - \theta_2 = 2\theta_E. \quad (\text{A3})$$

On the other hand, for $|\beta| > \theta_E$, only one image of the source will exist, being located at θ_1 . Finally, the magnification (μ) can be determined using the formula

$$\mu = \frac{|\theta|}{|\theta| - \theta_E}. \quad (\text{A4})$$

Appendix B Quasar Selection Function

As mentioned in Section 2.2, the parent population of quasars in our simulation follows the luminosity function of Matsuoka et al. (2018b). They are distributed in absolute magnitudes of $-30 \leq M_{1450} \leq -20$ and redshifts of $5.6 \leq z \leq 7.2$. Note that without strong lensings, the data that we use can only detect quasars brighter than $M_{1450} \approx -24.5$. On the other hand, the lensing effect could push this limit into a fainter regime, which depends on the magnification factor.

We show in Section 2.5 that our selection criteria truncate the search, to focus on lens systems where the background quasars are positioned at $z = 6.6\text{--}7.2$. To see this in more detail, we first describe our selection function (or

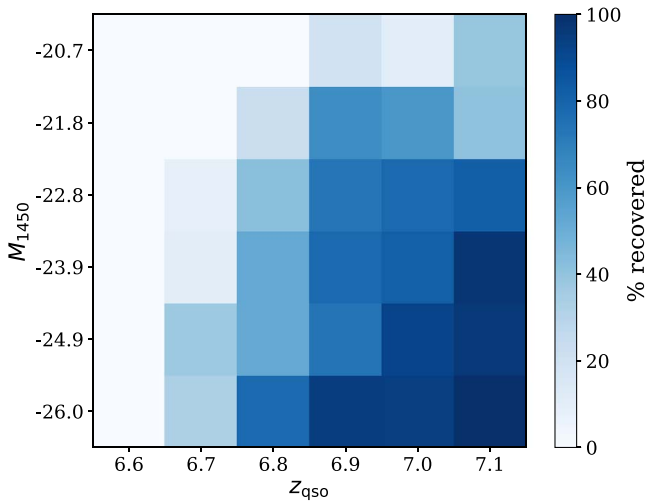


Figure 15. The quasar selection function that we employ in our work. The probability represents the proportion of mock quasars in each (M_{1450}, z) bin that can be recovered by our classifier.

completeness) as the fraction of mock quasars with given M_{1450} , z , and intrinsic SEDs that are recovered by our selection criteria. The results are presented in Figure 15. Intrinsically faint quasars can only be detected if they are boosted by lensing with large magnification factors, and these are relatively infrequent. Hence, at a given redshift, our completeness rate drops, as the quasars have intrinsically lower luminosity.

Appendix C

Additional Spectral Templates for the Contaminants

To separate the candidates from contaminants, we implement the SED fitting method described in Andika et al. (2020). The main idea is that we want to know which spectral classes best fit the observed SEDs of the candidates among the four templates—i.e., lensed quasars, unlensed low- z galaxies, MLT dwarfs, or unlensed high- z quasars.

The spectral template for MLT dwarfs is taken from the SpeX Prism Library²² (Burgasser 2014). In total, there are 360 templates that reflect stars with spectral classes of M5–M9, L0–L9, and T0–T8 in the wavelength range 0.625–2.55 μm . We then extrapolate these templates into the MIR regime to cover the unWISE bands—i.e., W1 (3.4 μm) and W2 (4.6 μm)—following the method presented by Andika et al. (2020).

Next, we add empirical quasar spectra, taken from three sources. The first is found in Selsing et al. (2016), who constructed a composite spectrum based on the sample of Very Large Telescope/X-shooter observations of luminous $1 < z < 2$ quasars obtained from SDSS. The second is found in Jensen et al. (2016), which is based on 58,656 spectra of $2.1 < z < 3.5$ Baryon Oscillation Spectroscopic Survey (BOSS) quasars,

binned by luminosity, spectral index, and redshift. Lastly, the third composite spectrum is taken from Harris et al. (2016), which was made by averaging 102,150 BOSS quasar spectra at $2.1 < z < 3.5$. Note that the composites from Harris et al. (2016) and Jensen et al. (2016) only contain spectra with a rest-frame wavelength up to ~ 3000 Å. To extend these further, to redder wavelengths, we stitch these templates to the Selsing et al. (2016) composite spectrum, starting from 2650 Å. After this, we apply the dust reddening using values of $E(B - V) = -0.02$ to 0.14, following the Calzetti et al. (2000) extinction law. We further create a grid of spectra, with a redshift range of $4.0 \leq z \leq 8.0$, with $\Delta z = 0.003$.

In addition, to ensure that our candidates do not mimic low- z unlensed elliptical galaxies, we utilize the Brown et al. (2014) galaxy spectral atlas. This database contains 129 templates derived from nearby $z \lesssim 0.05$ galaxies of various types (e.g., elliptical, spiral, starbursts, etc.) We then add the reddening effect from dust, using Calzetti et al. (2000) model, with $A(V) = 0$ to 1. A grid of SED models is then constructed by distributing the templates across redshifts of $0.0 \leq z \leq 3.0$, with $\Delta z = 0.005$. Note that on top of the SED models of quasars and galaxies, we add the attenuation caused by HI in the IGM, using the formula from Inoue et al. (2014).

Appendix D

Test on Known Lensed Quasars

We present here a sample of discovered $z \lesssim 3$ lensed quasars, taken from the Gravitationally Lensed Quasar Database (Inada et al. 2012; Agnello et al. 2015, 2018; Lemon et al. 2018, 2019, 2020; Spiniello et al. 2018; Jaelani et al. 2021), which we use for an additional independent test. The assembly of the training data for our CNN classifier is similar to that explained in Sections 2.2–2.4. However, this time, we extend the redshift of the background quasars to $1.5 \leq z \leq 7.2$ and the luminosities to $-30 \leq M_{1450} \leq -20$, and distribute the new 900 quasars uniformly over the M_{1450} – z space.

To estimate the completeness and purity of our classifier, we first combine the known lensed quasars mentioned above with a sample of contaminants—i.e., galaxies, stars, and unlensed quasars—where the nonlenses are assumed to outnumber the lens population by a factor of a few thousand. By using this test data set, our CNN classifier successfully recovered 27 known lenses from the 34 sources available in the DES data, equivalent to a TPR (or completeness) of 79%, with an FPR of 1%. We also find that our model has a purity of 11% in finding the lens candidates. The calculated lens probabilities, compared to the image separations and redshifts, are shown in Figure 16. We also show their images and lens probability scores (P_{lens}) in Figure 17.

²² <http://pono.ucsd.edu/~adam/browndwarfs/spexprism/library.html>

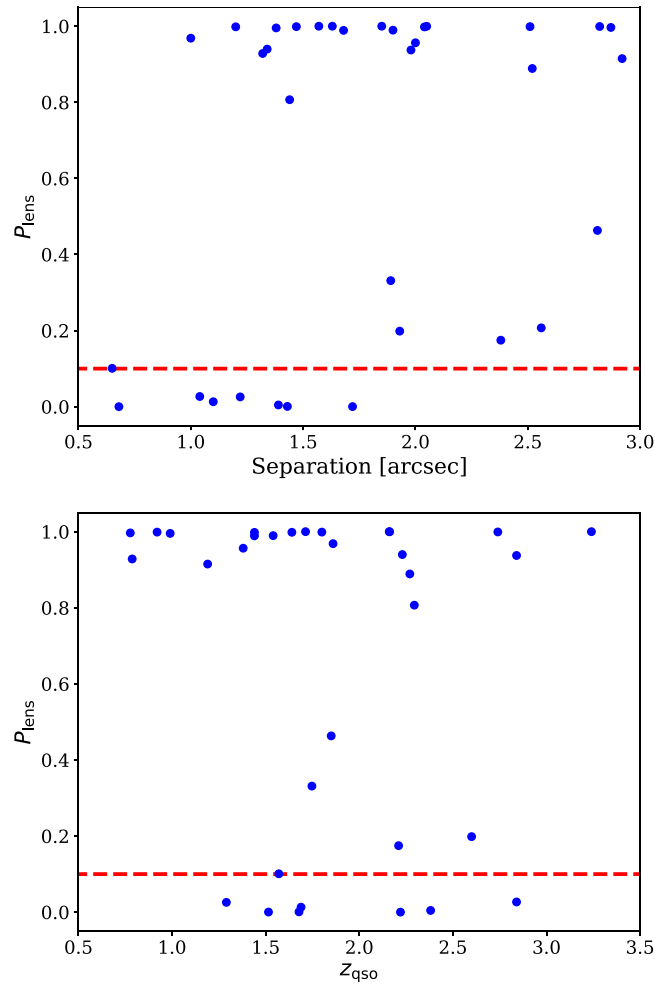


Figure 16. Calculated lens probabilities for a sample of known lensed quasars compared to their separations and background quasar redshifts. The data are shown as blue dots, while the probability threshold for classifying sources as lens systems is denoted by the red dashed line.

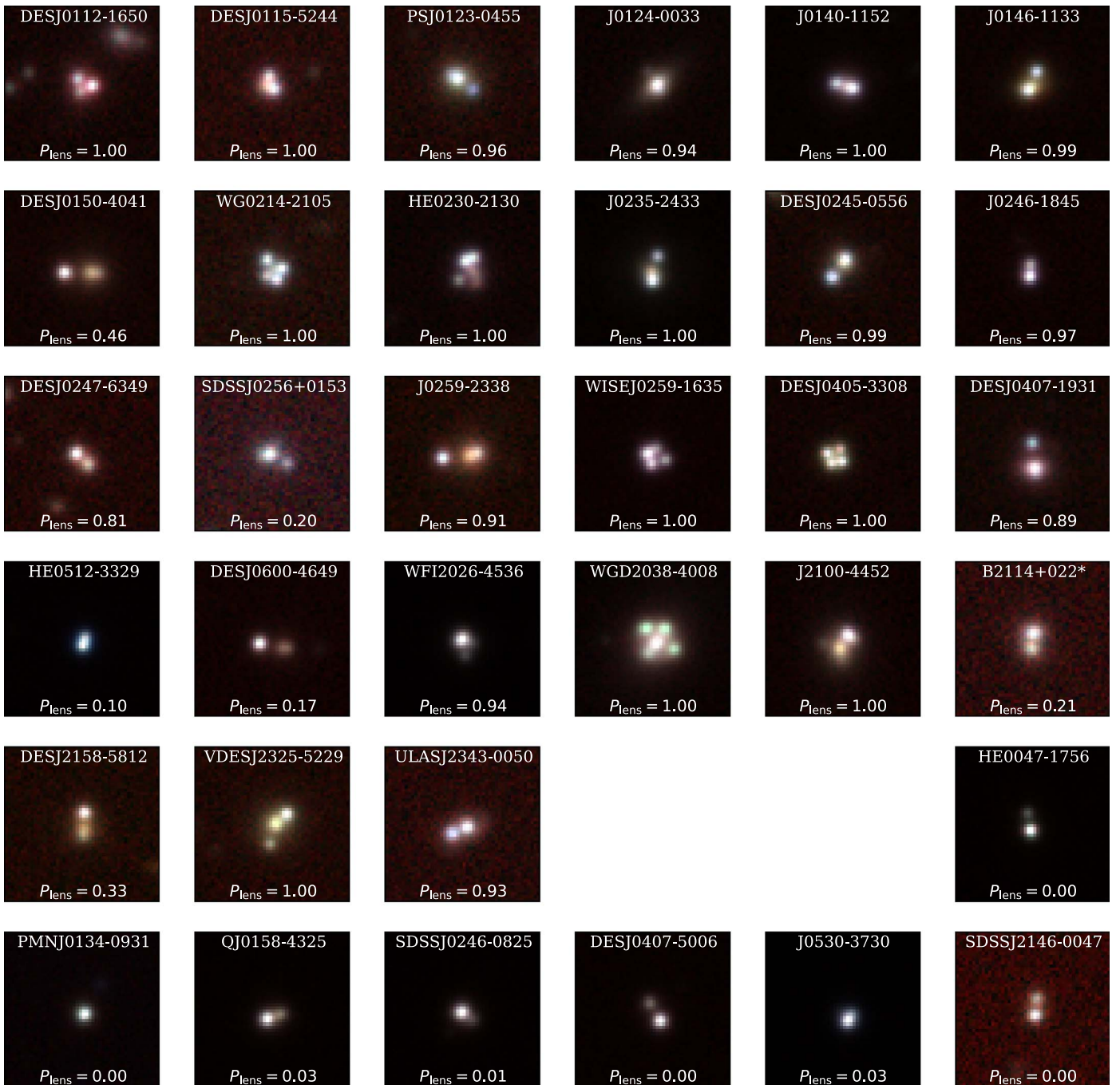











Figure 17. Examples of known lensed quasars taken from the literature (see the main text). The cutouts are created based on $12''6 \times 12''6$ DES iZY images, and they are colorized according to the Lupton et al. (2004) method. The lens probabilities of the lens systems are indicated in each image.

ORCID iDs

Irham Taufik Andika  <https://orcid.org/0000-0001-6102-9526>
 Knud Jahnke  <https://orcid.org/0000-0003-3804-2137>
 Arjen van der Wel  <https://orcid.org/0000-0002-5027-0135>
 Eduardo Bañados  <https://orcid.org/0000-0002-2931-7824>
 Sarah E. I. Bosman  <https://orcid.org/0000-0001-8582-7012>
 Frederick B. Davies  <https://orcid.org/0000-0003-0821-3644>
 Anna-Christina Eilers  <https://orcid.org/0000-0003-2895-6218>
 Anton Timur Jaelani  <https://orcid.org/0000-0001-6282-5778>

Chiara Mazzucchelli  <https://orcid.org/0000-0002-5941-5214>

Masafusa Onoue  <https://orcid.org/0000-0003-2984-6803>

Jan-Torge Schindler  <https://orcid.org/0000-0002-4544-8242>

References

- Abadi, M., Barham, P., Chen, J., et al. 2016, arXiv:1605.08695
 Abbott, T. M. C., Adamów, M., Aguena, M., et al. 2021, *ApJS*, 255, 20
 Abdurro'uf, A. K., Aerts, C., et al. 2022, *ApJS*, 259, 35
 Agnello, A., Kelly, B. C., Treu, T., & Marshall, P. J. 2015, *MNRAS*, 448, 1446
 Agnello, A., Schechter, P. L., Morgan, N. D., et al. 2018, *MNRAS*, 475, 2086

- Andika, I. T. 2022, PhD thesis, Max Planck Institute for Astronomy, Heidelberg
- Andika, I. T., Jahnke, K., Bañados, E., et al. 2022, *AJ*, **163**, 251
- Andika, I. T., Jahnke, K., Onoue, M., et al. 2020, *ApJ*, **903**, 34
- Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, *AJ*, **156**, 123
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, **558**, A33
- Bañados, E., Mazzucchelli, C., Momjian, E., et al. 2021, *ApJ*, **909**, 80
- Bañados, E., Venemans, B. P., Decarli, R., et al. 2016, *ApJS*, **227**, 11
- Bañados, E., Venemans, B. P., Mazzucchelli, C., et al. 2018, *Natur*, **553**, 473
- Barkana, R. 1998, *ApJ*, **502**, 531
- Best, W. M. J., Magnier, F. A., Liu, M. C., et al. 2018, *ApJS*, **234**, 1
- Blanton, M. R., Bershad, M. A., Abolfathi, B., et al. 2017, *AJ*, **154**, 28
- Bom, C. R., Fraga, B. M. O., Dias, L. O., et al. 2022, arXiv:2203.09536
- Boroson, T. A., & Green, R. F. 1992, *ApJS*, **80**, 109
- Bosman, S. E. I., Davies, F. B., Becker, G. D., et al. 2022, *MNRAS*, **514**, 55
- Bradley, L., Sipőcz, B., Robitaille, T., et al. 2021, astropy/photutils: v1.1.0, Zenodo, doi:10.5281/zenodo.596036
- Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, *ApJ*, **686**, 1503
- Brown, M. J. I., Moustakas, J., Smith, G. D. T., et al. 2014, *ApJS*, **212**, 18
- Burgasser, A. J. 2014, in ASI Conf. 11, Int. Workshop on Stellar Spectral Libraries, ed. H. P. Singh, P. Prugniel, & I. Vauglin, 7
- Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, *ApJ*, **533**, 682
- Cañameras, R., Schuldt, S., Shu, Y., et al. 2020, *A&A*, **653**, L6
- Carnall, A. C., McLure, R. J., Dunlop, J. S., et al. 2019, *MNRAS*, **490**, 417
- Carnall, A. C., McLure, R. J., Dunlop, J. S., & Davé, R. 2018, *MNRAS*, **480**, 4379
- Carnero Rosell, A., Santiago, B., dal Ponte, M., et al. 2019, *MNRAS*, **489**, 5301
- Caswell, T. A., Droettboom, M., Hunter, J., et al. 2019, matplotlib/matplotlib v3.1.0, Zenodo, doi:10.5281/zenodo.2893252
- Chan, J. H. H., Rojas, K., Millon, M., et al. 2021, *A&A*, **647**, A115
- Charlot, S., & Fall, S. M. 2000, *ApJ*, **539**, 718
- Connor, T., Stern, D., Bañados, E., & Mazzucchelli, C. 2021, *ApJL*, **922**, L24
- Davies, F. B., Wang, F., Eilers, A.-C., & Hennawi, J. F. 2020, *ApJL*, **904**, L32
- Dawson, K. S., Kneib, J.-P., Percival, W. J., et al. 2016, *AJ*, **151**, 44
- Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, *AJ*, **145**, 10
- Decarli, R., Walter, F., Venemans, B. P., et al. 2018, *ApJ*, **854**, 97
- Developers, T. 2022, TensorFlow, v2.9.0, Zenodo, doi:10.5281/zenodo.4724125
- Eilers, A.-C., Hennawi, J. F., Decarli, R., et al. 2020, *ApJ*, **900**, 37
- Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, *AJ*, **142**, 72
- Euclid Collaboration, Barnett, R., Warren, S. J., et al. 2019, *A&A*, **631**, A85
- Euclid Collaboration, Scaramella, R., Amiaux, J., et al. 2022, *A&A*, **662**, A112
- Fan, X., Strauss, M. A., Becker, R. H., et al. 2006, *AJ*, **132**, 117
- Fan, X., Wang, F., Yang, J., et al. 2019, *ApJL*, **870**, L11
- Fitzpatrick, E. L. 1999, *PASP*, **111**, 63
- Flaugher, B., Diehl, H. T., Honscheid, K., et al. 2015, *AJ*, **150**, 150
- Flesch, E. W. 2021, arXiv:2105.12985
- Gentile, F., Tortora, C., Covone, G., et al. 2022, *MNRAS*, **510**, 500
- Gilman, D., Birrer, S., Nierenberg, A., et al. 2020, *MNRAS*, **491**, 6077
- Habouzit, M., Onoue, M., Bañados, E., et al. 2022, *MNRAS*, **511**, 3751
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Natur*, **585**, 357
- Harris, D. W., Jensen, T. W., Suzuki, N., et al. 2016, *AJ*, **151**, 155
- Honscheid, K., & DePoy, D. L. 2008, arXiv:0810.3600
- Hsueh, J. W., Enzi, W., Vegetti, S., et al. 2020, *MNRAS*, **492**, 3047
- Huang, X., Storfer, C., Ravi, V., et al. 2020, *ApJ*, **894**, 78
- Inada, N., Oguri, M., Shim, M.-S., et al. 2012, *AJ*, **143**, 119
- Inayoshi, K., Visbal, E., & Haiman, Z. 2020, *ARA&A*, **58**, 27
- Inoue, A. K., Shimizu, I., Iwata, I., & Tanaka, M. 2014, *MNRAS*, **442**, 1805
- Izumi, T., Matsuoka, Y., Fujimoto, S., et al. 2021, *ApJ*, **914**, 36
- Jaelani, A. T., Rusu, C. E., Kayo, I., et al. 2021, *MNRAS*, **502**, 1487
- Jensen, T. W., Vivek, M., Dawson, K. S., et al. 2016, *ApJ*, **833**, 199
- Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, *ApJ*, **873**, 111
- Kingma, D. P., & Ba, J. 2014, arXiv:1412.6980
- Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv:1110.3193
- LeCun, Y., Boser, B., Denker, J. S., et al. 1989, *Neural Comput.*, **1**, 541
- Lemon, C., Auger, M. W., McMahon, R., et al. 2020, *MNRAS*, **494**, 3491
- Lemon, C. A., Auger, M. W., & McMahon, R. G. 2019, *MNRAS*, **483**, 4242
- Lemon, C. A., Auger, M. W., McMahon, R. G., & Ostrovski, F. 2018, *MNRAS*, **479**, 5060
- Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A., & Talwalkar, A. 2018, *J Mach Learn Res*, **18**, 1, <https://jmlr.org/papers/volume18/16-558/16-558.pdf>
- Lupton, R., Blanton, M. R., Fekete, G., et al. 2004, *PASP*, **116**, 133
- Madau, P. 2017, *ApJ*, **851**, 50
- Mason, C. A., Treu, T., Schmidt, K. B., et al. 2015, *ApJ*, **805**, 79
- Matsuoka, Y., Iwasawa, K., Onoue, M., et al. 2018a, *ApJS*, **237**, 5
- Matsuoka, Y., Iwasawa, K., Onoue, M., et al. 2022, *ApJS*, **259**, 18
- Matsuoka, Y., Strauss, M. A., Kashikawa, N., et al. 2018b, *ApJ*, **869**, 150
- Mazzucchelli, C., Bañados, E., Venemans, B. P., et al. 2017, *ApJ*, **849**, 91
- McGreer, I. D., Jiang, L., Fan, X., et al. 2013, *ApJ*, **768**, 105
- McMahon, R. G., Banerji, M., Gonzalez, E., et al. 2013, *Msngr*, **154**, 35
- McMahon, R. G., Banerji, M., Gonzalez, E., et al. 2021, *yCat*, **II/367**
- Metcalf, R. B., Meneghetti, M., Avestruz, C., et al. 2019, *A&A*, **625**, A119
- Meyer, R. A., Bosman, S. E. I., & Ellis, R. S. 2019, *MNRAS*, **487**, 3305
- Meyer, R. A., Decarli, R., Walter, F., et al. 2022, *ApJ*, **927**, 141
- Mignoli, M., Gilli, R., Decarli, R., et al. 2020, *A&A*, **642**, L1
- Millon, M., Courbin, F., Bonvin, V., et al. 2020, *A&A*, **642**, A193
- Miralda-Escudé, J. 1998, *ApJ*, **501**, 15
- Mortlock, D. J., Warren, S. J., Venemans, B. P., et al. 2011, *Natur*, **474**, 616
- Neeleman, M., Novak, M., Venemans, B. P., et al. 2021, *ApJ*, **911**, 141
- Nightingale, J., Hayes, R., Kelly, A., et al. 2021, *JOSS*, **6**, 2825
- Oguri, M., & Marshall, P. J. 2010, *MNRAS*, **405**, 2579
- Onoue, M., Kashikawa, N., Matsuoka, Y., et al. 2019, *ApJ*, **880**, 77
- Onoue, M., Matsuoka, Y., Kashikawa, N., et al. 2021, *ApJ*, **919**, 61
- Pacucci, F., Foord, A., Gordon, L., & Loeb, A. 2022, *MNRAS*, **514**, 2855
- Pacucci, F., & Loeb, A. 2019, *ApJL*, **870**, L12
- Pacucci, F., & Loeb, A. 2022, *MNRAS*, **509**, 1885
- Pipien, S., Cuby, J. G., Basa, S., et al. 2018, *A&A*, **617**, A127
- Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, *A&A*, **641**, A6
- Pons, E., McMahon, R. G., Simcoe, R. A., et al. 2019, *MNRAS*, **484**, 5142
- Prakash, A., Licquia, T. C., Newman, J. A., et al. 2016, *ApJS*, **224**, 34
- Reback, J., Jbrockmendel, A., McKinney, W., et al. 2021, pandas-dev/pandas: Pandas v1.3.0, Zenodo, doi:10.5281/zenodo.3509134
- Reed, S. L., Banerji, M., Becker, G. D., et al. 2019, *MNRAS*, **487**, 1874
- Rojas, K., Savary, E., Clément, B., et al. 2022, *A&A*, **668**, A73
- Schindler, J.-T., Farina, E. P., Bañados, E., et al. 2020, *ApJ*, **905**, 51
- Schlafly, E. F., Meisner, A. M., & Green, G. M. 2019, *ApJS*, **240**, 30
- Schneider, P. 2015, *Extragalactic Astronomy and Cosmology: An Introduction* (Berlin: Springer)
- Selsing, J., Fynbo, J. P. U., Christensen, L., & Krogager, J. K. 2016, *A&A*, **585**, A87
- Shen, Y., Wu, J., Jiang, L., et al. 2019, *ApJ*, **873**, 35
- Songaila, A., & Cowie, L. L. 2010, *ApJ*, **721**, 1448
- Spiniello, C., Agnello, A., Napolitano, N. R., et al. 2018, *MNRAS*, **480**, 1163
- Stacey, H. R., McKean, J. P., Robertson, N. C., et al. 2018, *MNRAS*, **476**, 5075
- Sultana, F., Sufian, A., & Dutta, P. 2019, arXiv:1905.03288
- Suyu, S. H., Bonvin, V., Courbin, F., et al. 2017, *MNRAS*, **468**, 2590
- Treu, T. 2010, *ARA&A*, **48**, 87
- Tsuzuki, Y., Kawara, K., Yoshii, Y., et al. 2006, *ApJ*, **650**, 57
- Venemans, B. P., Bañados, E., Decarli, R., et al. 2015, *ApJL*, **801**, L11
- Vestergaard, M., & Wilkes, B. J. 2001, *ApJS*, **134**, 1
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *NatMe*, **17**, 261
- Volonteri, M. 2012, *Sci*, **337**, 544
- Wang, F., Yang, J., Fan, X., et al. 2019, *ApJ*, **884**, 30
- Wang, F., Yang, J., Fan, X., et al. 2021, *ApJL*, **907**, L1
- Wenzl, L., Schindler, J.-T., Fan, X., et al. 2021, *AJ*, **162**, 72
- Wilde, J., Serjeant, S., Bromley, J. M., et al. 2022, *MNRAS*, **512**, 3464
- Willott, C. J., Delorme, P., Reylé, C., et al. 2010, *AJ*, **139**, 906
- Woodfinden, A., Nadathur, S., Percival, W. J., et al. 2022, *MNRAS*, **516**, 4307
- Worseck, G., & Prochaska, J. X. 2011, *ApJ*, **728**, 23
- Wright, E. L. 2006, *PASP*, **118**, 1711
- Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, *AJ*, **140**, 1868
- Wyithe, J. S. B., & Loeb, A. 2002, *Natur*, **417**, 923
- Wyithe, J. S. B., Yan, H., Windhorst, R. A., & Mao, S. 2011, *Natur*, **469**, 181
- Yang, J., Venemans, B., Wang, F., et al. 2019a, *ApJ*, **880**, 153
- Yang, J., Wang, F., Fan, X., et al. 2019b, *AJ*, **157**, 236
- Yang, J., Wang, F., Fan, X., et al. 2020, *ApJL*, **897**, L14
- Yang, J., Wang, F., Wu, X.-B., et al. 2016, *ApJ*, **829**, 33
- York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, *AJ*, **120**, 1579
- Yue, M., Fan, X., Yang, J., & Wang, F. 2022a, *ApJ*, **925**, 169
- Yue, M., Fan, X., Yang, J., & Wang, F. 2022b, *AJ*, **163**, 139
- Yue, M., Yang, J., Fan, X., et al. 2021, *ApJ*, **917**, 99
- Zhao, C., Variu, A., He, M., et al. 2022, *MNRAS*, **511**, 5492