A cross-modal adversarial learning method
for estimating photometric redshift of quasars

Chen Zhang [1]   Wenyu Wang [1]   Meixia Qu [1]   Yanxia Zhang [1]   Bin Jiang [2]

Abstract

Quasars play a crucial role in studying various important physical processes. We propose a cross-modal deep learning method for estimating the photometric redshifts of quasars. Our model utilizes adversarial training to enable the conversion between photometric data features (magnitudes, colors, etc.) and photometric image features in five bands (u, g, r, i, z), in order to extract modality-invariant features. We used $|\Delta z| = |(z_{photo} - z_{spec})/(1 + z_{spec})|$ as evaluation metric. The latest SOTA method, which implements cross-modal generation of simulated spectra from photometric data, has been chosen as the baseline. Firstly the proposed method was tested on the same SDSS DR17 dataset of 415,930 quasars($1 \leq z_{spec} \leq 5$) as the baseline method. Compared to the baseline, the RMSE of our Δz decreased from 0.1235 to 0.1031. Further evaluation on a larger dataset of 465,292 quasars achieved a lower RMSE of Δz of 0.0861. This method also can be generalized to other tasks such as galaxy classification and redshift estimation.

1. Introduction

High redshift quasars are important for the study of supermassive black hole formation, galaxy formation and evolution (Fontanot et al., 2020; Getachew-Woreta et al., 2022), cosmic reionization (Grazian et al., 2022), the diffuse medium in the universe (IGM & CGM), and the early history of the universe (Dong et al., 2022). As large-scale sky surveys continue, the Sloan Digital Sky Survey (SDSS) has discovered spectra for over 750,000 quasars. However, only a small percentage, ranging from 0.01% to 1%, of these spectra have a high SNR.It's challenging to obtain full and high-quality spectroscopic data(Collaboration, 2022; Almeida et al., 2023). The use of relatively complete multi-band photometric data for calculating photometric redshifts(z_{photo}) has greatly reduced resource costs.

To further improve the accuracy of photometric redshift estimation, various deep learning methods have been applied in the evaluation of photometric redshifts(Pasquet et al., 2019; Brescia et al., 2021; Zhou et al., 2021; Henghes et al., 2022; Cunha & Humphrey, 2022; Moriwaki et al., 2023). Due to the advancements in multimodal techniques, (Hong et al., 2023) attempted to use photometric data of quasars to generate simulated spectral features, ultimately combining the original photometric data with the generated spectral data for redshift estimation. However, due to significant differences in the physical properties of photometric and spectral data, the $|\Delta z|$ of RMSE only reduced from 0.1332 to 0.1235 and still reach the the SOTA.

Photometric images capture visual features like shape, structure, and position of quasars, while photometric magnitudes and colors convey overall brightness and color information. Due to the point-like nature of quasars, CCD-captured images are limited to a few pixels. Photometric magnitudes are computed from the original photometric images and provide an overall description of the celestial objects across different wavelength bands. These datas are complementary to or jointly used with each other for the photometric redshift estimation(Salvato et al., 2019; Leistedt et al., 2019; Dey et al., 2022).

In this paper, we propose a deep learning-based cross-modal representation method that enables the mutual transformation between photometric features, including magnitudes,colors, and photometric image features. By achieving this cross-modal transformation, we extract cross-modal invariant features that are used for downstream tasks such as photometric redshift estimation.

We apply our model to the same dataset(415,930 quasars) of the current state-of-the-art (SOTA) baseline method(Hong et al., 2023). Compared to the baseline, our model achieves a

reduced RMSE of 0.1031 on Δz, compared to the baseline's RMSE of 0.1235. Furthermore, when applied to a larger datasets of 465,292 quasars, our model achieves an even lower RMSE of 0.0861.

2. Data

2.1. Extended Dataset acquisition

We cross-matched the photoObjAll and SpecObjAll catalogue in SDSS DR17(Collaboration, 2022) CasJobs and selected 608,451 quasar samples with redshift values between 1 and 5(*Appendix A*). Due to the point-like nature of quasar sources, we chose PSF(point spread function) magnitudes and errors in five bands (denoted as psfMag_, psfMag_g, psfMag_r, psfMag_i, psfMag_z, psfMagErr_u, psfMagErr_g, psfMagEr_r, psfMagErr_i, psfMagErr_z), along with the instellar extinction in each bands, (denoted as extinction_u, extinction_g, extinction_r, extinction_i, extinction_z).

To ensure the quality of our data, we set $z\_warning = 0$ to ensure the quality of the spectrum, clean=1 to avoid contamination of photometric data, $zErr \leq 0.001$ to limit redshift errors, $18 \leq psfMag\_g \leq 22$ to exclude too faint or too bright quasars, and $petroRad\_r$ to avoid interference from some *AGN* samples. Some of the data obtained from CasJobs contain duplicate sources, i.e., with the same objID. We removed these duplicates from our dataset.The model's results were based on a $60\%/20\%/20\%$ split of training/validation/testing datasets. The datasets was divided five times randomly. The SGD optimizer is selected, the hyperparameter weight_decay is 0.03, the momentum is 0.2, and the model achieves the best results when the learning rate is 0.008.

Finally, 465,292 targets are obtained with the smallest $zErr$ for each source. The reshift distribution of slected quasars is shown in figure 1. The extended dataset used in this study is employed to validate the model's ability to generalize to a larger dataset.

2.2. Baseline Dataset

The Baseline dataset(*Appendix B*) is a cross-modal transformation dataset of spectroscopic and photometric data utilized by (Hong et al., 2023), consisting of 415,930 quasars from the SDSS DR17 dataset ($1 \leq z_{spec} \leq 5$). In comparison to the expanded dataset used in this study,(Hong et al., 2023) primarily imposed constraints on the signal-to-noise ratio (SNR) values in the u, g, r, i, and z bands to achieve better performance in low SNR bands. The limitations on spectroscopic quality were relaxed by not setting $z_{warning} = 0$. Although this increased the diversity of the spectra, it also resulted in the inclusion of spectra with unreliable quality in the sample. However, this does not affect the fair comparison of results in this study, as the

same dataset was used, relying solely on photometric data for comparison with the baseline. Finally in this dataset, 415,930 quasars are obtained. The baseline dataset used in this study is primarily based on the same dataset of baseline's method, allowing for a fair comparison with SOTA methods.

2.3. PreProcessing

For the photometric data, we used the de-reddened magnitudes (dred_u, dred_g, dred_r, dred_i, dred_z after correcting for the interstellar reddening effects in the five bands. We also calculated the color attributes such as u-g, g-r, r-i, and i-z by interpolating between the magnitudes in two adjacent bands.
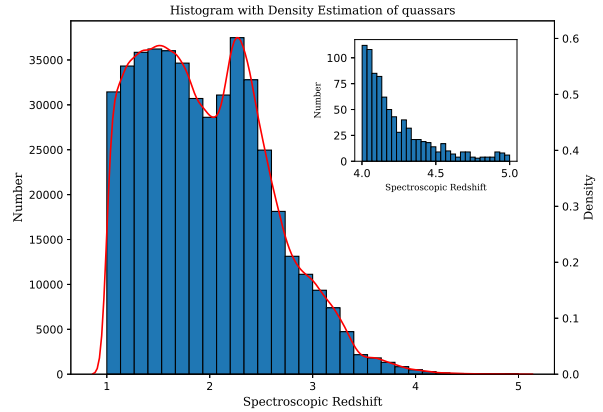

Figure 1. Redshift distribution histogram of selected quasars.

We transformed the RA and DEC coordinates of the quasar samples into image positions for photometric images. Then, 64x64 images were cropped from the u, g, i, r, and z bands, centered at the RA and DEC coordinates. The images were aligned based on the r band.

3. Basical Model Architecture

Inspired by the approach in (Qian et al., 2022), which utilizes adversarial learning to achieve cross-modal transformation between image and text information, our cross-modal deep learning model as figure 2 consists of two main components. The first part involves learning the modality-invariant representation between photometric data and photometric images through adversarial learning, an self-upervised process. Extinctions are used as labels instead of inputs to improve the feature extraction capability of the image encoder. The second part focuses on obtaining fused features for photometric redshift estimation.
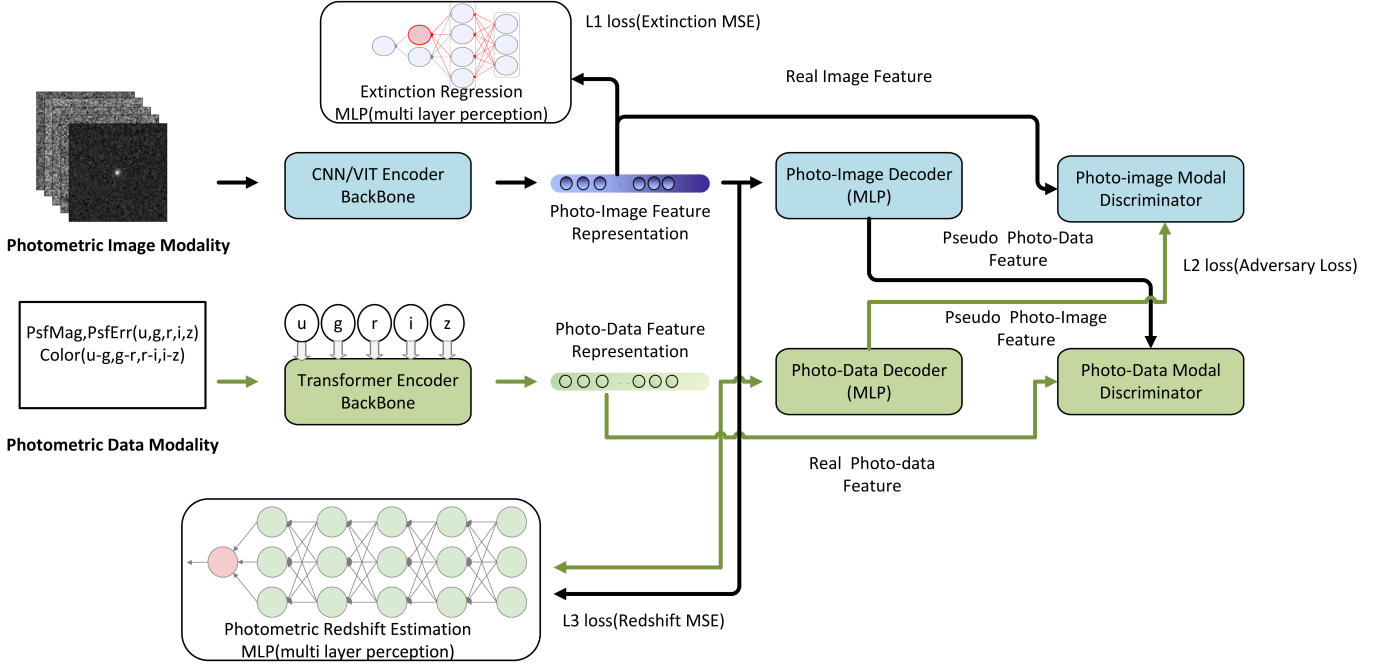
Figure 2. Adversarial Learning Model Architecture

3.1. Cross-modal Adversarial Learning Method

In the adversarial learning module, the Encoder and Decoder act as the generative network, while the modality discriminator network acts as the discriminative network. They have opposite gradient propagation based on the adversarial loss. C

Photo-Data Encoding Network. For each input photometric data, we split it into two parts: one part contains the multi-band magnitudes and their corresponding errors, while the other part contains the color properties: $X^{photo} \in (X^{mag} \cup X^{color})$, $X^{mag} \in \mathbb{R}^{nXm}$, $X^{mag} \in \mathbb{R}^d$. X is a two-dimensional matrix, where n represents the number of bands and m represents the number of features. In this paper, $n = 5(u,g,r,i,z)$ and $m = 2(deredmag\_\{bands\}, dredmag\_\{bands\}\_Err)$.

$$\Gamma(X^{mag}) = (W_0 X^{mag}(X^{mag})^T W_0^T)(W_1 X) \in \mathbb{R}^{nXm} \tag{1}$$

Equation 1 is an attention mechanism based on the classical self-attention mechanism. W_0 and W_1 are trainable weight matrices used to perform embedding operations on the original feature matrix X^{mag}. Compared with the classical sub-attention mechanism, $(W_0 X^{mag}(X^{mag})^T W_0^T)$ produces a set of self-correlated symmetric matrices that do not contain contextual information. This change is made because the data we need to process is not sequential data.The original values obtained by embedding with W_1 are corrected based on the correlation symmetric matrix obtained from

the training.

$$X_l^{mag} = F(X_{l-1}^{mag}; \theta_{mag}) = f_{mlp}(\Gamma(X_{l-1}^{mag}; \theta_\Gamma); \theta_{F_{mlp}}) \tag{2}$$

After each layer of attention mechanism, further feature encoding is performed through a feedforward neural network consisting of multi-layer perceptron, where l represents the network layer number, as shown in Equation 2.

$$\chi^{photo} = F_{fusion}(X^{mag}, X^{color}; \theta_F)$$
$$= f_{mlp}(X^{mag} \cup X^{color}; \theta_f) \in \mathbb{R}^{512} \tag{3}$$

The final step is to fuse the magnitude and color features to obtain the photometric data features as equation 3.

Photo-Image Encoding Network.For each input photometric image $X^{img} \in \mathbb{R}^{5X64X64}$.The main backbone neural networks used for encoding image features are VIT(Vision Transformer)/CNN. In this paper we employed ResNet101(He et al., 2016) as the backbone neural network for image feature encoding as Equation 4, σ means to nonlinear activation function.

$$\chi^{img} = \sigma(F(X^{img}, \theta_{\mathbb{F}}))$$
$$= \sigma(f_{mlp}(f_{CNN}(X^{img}, \theta_{CNN}); \theta_{mlp})) \in \mathbb{R}^{512} \tag{4}$$

Table 1. Comparison of quasar photometric redshift prediction results between Baseline and our method(input pattern means to single modal data, multimodal data or single modal with generated fake modal data; The method is tested on the same dataset as baseline's and then be tested on a extended dataset)

Data set	Input pattern	$|\Delta z| < 0.1(\%)$	$|\Delta z| < 0.2(\%)$	$|\Delta z| < 0.3(\%)$	$|\Delta z| < 0.4(\%)$	RMSE
	Result of Baseline					
SDSS DR17 Baseline Dataset	χ_{phot}	80.41	92.60	96.06	98.23	0.1332
	$(\chi_{\text{phot}}, \chi_{\text{spec}}^{\text{generated}})$	84.45	93.59	96.43	98.35	0.1235
	Result of Our Method					
SDSS DR17 Baseline Dataset	$(\chi_{\text{phot}}, \chi_{\text{img}})$	82.05	94.62	97.88	98.96	**0.1010**
	$(\chi_{\text{phot}}^{\text{generated}}, \chi_{\text{img}})$	82.14	94.52	97.88	98.96	**0.1011**
(extended data in our paper)	$(\chi_{\text{phot}}, \chi_{\text{img}})$	**86.36**	**96.30**	**98.67**	**99.21**	**0.0865**
	$(\chi_{\text{phot}}, \chi_{\text{img}}^{\text{generated}})$	73.48	94.60	98.63	99.17	**0.1035**

Cross-modal Decoder Representation. In the decoder part, multi-layer perceptrons with non-linear activation functions are used to generate persudo cross-modal representation as Equation 5.

$$\chi^{photo2img} = G_{photo}(X^{photo}; \theta_{G_{photo}}) \in \mathbb{R}^{512},$$
$$\chi^{img2photo} = G_{img}(X^{img}; \theta_{G_{img}}) \in \mathbb{R}^{512}, \tag{5}$$

Redshift,Extinction Regression And Modal Discriminator. These modules are achieved through several full connected layers that containing nonlinear activation functions as Equation 6.

$$\hat{y}_{extinction} = F_{extinction}(X^{img}; \theta_{\mathbb{F}}) \in \mathbb{R}^5,$$
$$\hat{Z}_{photo} = F_{img}(\chi^{img}, \chi^{photo}; \theta_{\mathbb{F}}) \in \mathbb{R}^1, \tag{6}$$

In Equation 7, there are two mode discriminators, where D_1 discriminates whether the input photometric data features are real or fake, and D_2 discriminates whether the input image features are real or fake.

$$\hat{y}^{pho2pho} = SOFTMAX(D_1(\chi^{photo}; \theta_{D_1})) \in \mathbb{R}^2,$$
$$\hat{y}^{img2pho} = SOFTMAX(D_1(\chi^{img2photo}; \theta_{D_1})) \in \mathbb{R}^2,$$
$$\hat{y}^{img2img} = SOFTMAX(D_2(\chi^{img}; \theta_{D_2})) \in \mathbb{R}^2,$$
$$\hat{y}^{pho2img} = SOFTMAX(D_2(\chi^{photo2img}; \theta_{D_2})) \in \mathbb{R}^2, \tag{7}$$

3.2. The Objective Loss Function Optimization

The training of the adversarial learning neural network requires the optimization of three loss function components J_{all}: the cross-modal loss function $J_{intra-modal}$, the extinction regression loss function $J_{extinction}$, and the photo-

metric redshift estimation loss function J_z. The two hyperparameters α, β represent the trade-off factors.

$$J_{all} = \alpha J_{intra-modal} + \beta J_{extinction} + (1 - \alpha - \beta)J_z \tag{8}$$

In equation 9, the loss function H is cross-entropy function, which can measure the similarity between two distributions. The vector y^{True} is a full-ones vector with the same dimension as the input feature, while the vector y^{False} is a full-zeros vector with the same dimension as the input feature. Equation 9 describes the core adversarial loss function that permeates throughout the paper. For each discriminator, a set of real modalities is provided as positive samples, while transformed generated modalities serve as negative samples. The training process of the entire model is achieved through an adversarial process between the generator and the discriminator, where gradients are backpropagated between them. Therefore, we refer to the overall model's concept as adversarial learning.

$$H(y|\hat{y}) = \sum_i y \log(\hat{y}),$$
$$J_{intra-modal} = H(y^{True}|\hat{y}^{pho2pho}) + H(y^{False}|\hat{y}^{img2photo})$$
$$+ H(y^{True}|\hat{y}^{img2img}) + H(y^{False}|\hat{y}^{photo2img}), \tag{9}$$

The equation 10 employs MSE (Mean Squared Error) as a metric to evaluate the discrepancy between predicted regression values and the actual values.

$$\mathcal{J}_{extinction} + \mathcal{J}_Z = MSE(Z, \hat{Z}) + MSE(y_{ext}, \hat{y}_{ext}) \tag{10}$$

4. Results

4.1. Evaluation Metrics

The baseline method chosen in this study is the SOTA deep learning method proposed by (Hong et al., 2023), which uses photometric data to represent spectral features comparing to many existing methods. This method outperforms many existing methods for photometric redshift estimation of quasars. During the evaluation stage, we mainly used the RMSE (Root Mean Squared Error) of the Δz in Equation 11 for assessment.

$$|\Delta z| = \frac{|z_{photo} - z_{spec}|}{1 + z_{spec}} \qquad (11)$$

4.2. Redshift Estimation Results

As shown in Table 1, to ensure a fair comparison, our method is initially evaluated and compared with the baseline method using the same dataset as the baseline. Although our method shows a slightly lower performance compared to the baseline by around 2% in ($|\Delta z| < 0.1\%$), it significantly increases the proportion in the $|\Delta z| < 0.2\%$ and $|\Delta z| < 0.3\%$. Consequently, the overall RMSE of $|\Delta z|$ is further reduced from 0.1235 to 0.1010. With evaluation on an expanded dataset of 465,292 quasars, our method achieves even better results, with an RMSE of 0.0865. Due to space limitations, Appendix C provides a more visually intuitive presentation of the four experimental results from Table 1 in the form of scatter density plots and histogram figures.

5. Conclusions

Cross-modal Discussion. Since photometric magnitudes and colors are derived from original photometric images, it can't be strictly defined as multimodal data. The semantic relationship between them is closer than in strictly multimodal data. However, experimental results demonstrate that enhancing the mutual representation of these semantically similar data effectively reduces the error in photometric redshift estimation.

Conclusion. This paper proposes a multi-modal adversarial learning method for estimating the photometric redshift of quasars. It learns the transformation between photometric data features (magnitude, color, etc.) and photometric image features in five bands (u, g, r, i, z), and generalizes to larger datasets with reduced precision errors. And it can be applied to many other celestial objects such as galaxys.

References

Almeida, A., Anderson, S. F., Argudo-Fernández, M., Badenes, C., Barger, K., Barrera-Ballesteros, J. K., Bender, C. F., Benitez, E., Besser, F., Bizyaev, D., et al. The eighteenth data release of the sloan digital sky surveys: Targeting and first spectra from sdss-v. *arXiv preprint arXiv:2301.07688*, 2023.

Brescia, M., Cavuoti, S., Razim, O., Amaro, V., Riccio, G., and Longo, G. Photometric redshifts with machine learning, lights and shadows on a complex data science use case. *Frontiers in Astronomy and Space Sciences*, 8: 658229, 2021.

Collaboration, S. The seventeenth data release of the sloan digital sky surveys. *The Astrophysical Journal Supplement Series*, 259(35):39pp, 2022.

Cunha, P. and Humphrey, A. Photometric redshift-aided classification using ensemble learning. *Astronomy & Astrophysics*, 666:A87, 2022.

Dey, B., Andrews, B. H., Newman, J. A., Mao, Y.-Y., Rau, M. M., and Zhou, R. Photometric redshifts from sdss images with an interpretable deep capsule network. *Monthly Notices of the Royal Astronomical Society*, 515(4):5285–5305, 2022.

Dong, C., Gonzalez, A., Eikenberry, S., Jeram, S., Likamonsavad, M., Liske, J., Stelter, D., and Townsend, A. Forecasting cosmic acceleration measurements using the lyman-α forest. *Monthly Notices of the Royal Astronomical Society*, 514(4):5493–5505, 2022.

Fontanot, F., De Lucia, G., Hirschmann, M., Xie, L., Monaco, P., Menci, N., Fiore, F., Feruglio, C., Cristiani, S., and Shankar, F. The rise of active galactic nuclei in the galaxy evolution and assembly semi-analytic model. *Monthly Notices of the Royal Astronomical Society*, 496 (3):3943–3960, 2020.

Getachew-Woreta, T., Polić, M., Masegosa, J., Perea, J., Beyoro-Amado, Z., and Márquez, I. Effect of agn on the morphological properties of their host galaxies in the local universe. *Monthly Notices of the Royal Astronomical Society*, 514(1):607–620, 2022.

Grazian, A., Giallongo, E., Boutsia, K., Calderone, G., Cristiani, S., Cupani, G., Fontanot, F., Guarneri, F., and Ozdalkiran, Y. The space density of ultra-luminous qsos at the end of reionization epoch by the qubrics survey and the agn contribution to the hydrogen ionizing background. *The Astrophysical Journal*, 924(2):62, 2022.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Henghes, B., Thiyagalingam, J., Pettitt, C., Hey, T., and Lahav, O. Deep learning methods for obtaining photometric redshift estimations from images. *Monthly Notices of the Royal Astronomical Society*, 512(2):1696–1709, 2022.

Hong, S., Zou, Z., Luo, A.-L., Kong, X., Yang, W., and Chen, Y. Photoredshift-mml: a multimodal machine learning method for estimating photometric redshifts of quasars. *Monthly Notices of the Royal Astronomical Society*, 518(4):5049–5058, 2023.

Leistedt, B., Hogg, D. W., Wechsler, R. H., and DeRose, J. Hierarchical modeling and statistical calibration for photometric redshifts. *The Astrophysical Journal*, 881(1): 80, 2019.

Moriwaki, K., Nishimichi, T., and Yoshida, N. Machine learning for observational cosmology. *Reports on Progress in Physics*, 2023.

Pasquet, J., Bertin, E., Treyer, M., Arnouts, S., and Fouchez, D. Photometric redshifts from sdss images using a convolutional neural network. *Astronomy & Astrophysics*, 621: A26, 2019.

Qian, S., Xue, D., Fang, Q., and Xu, C. Integrating multi-label contrastive learning with dual adversarial graph neural networks for cross-modal retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

Salvato, M., Ilbert, O., and Hoyle, B. The many flavours of photometric redshifts. *Nature Astronomy*, 3(3):212–222, 2019.

Zhou, X., Gong, Y., Meng, X.-M., Zhang, X., Cao, Y., Chen, X., Amaro, V., Fan, Z., and Fu, L. Spectroscopic and photometric redshift estimation by neural networks for the china space station optical survey (css-os). *The Astrophysical Journal*, 909(1):53, 2021.

A. Entended SQL appendix

```
select b.run, b.camcol, b.field, a.specObjID, a.bestObjID,
a.class, a.z, a.zErr, b.objID, b.psfMag_u, b.psfMag_g,
b.psfMag_r, b.psfMag_i, b.psfMag_z, b.extinction_u,
b.extinction_g, b.extinction_r, b.extinction_i, b.extinction_z,
b.psfMagErr_u, b.psfMagErr_g, b.psfMagErr_r,
b.psfMagErr_i, b.psfMagErr_z, b.dered_u, b.dered_g, b.dered_r, b.dered_i, b.dered_z
into mydb.dr18psf_ugriz_SNR_z0
from dr17.SpecObjAll as a
join dr17.PhotoObjAll as b
on a.bestObjID = b.objID
where a.class ='QSO' and a.z >=1 and a.z<5 and a.zErr< 0.001 and b.psfMag_g >= 18 and
b.psfMag_g <= 22 and b.petroRad_r < 5 and a.snMedian_u<10 and a.snMedian_g<10 and
a.snMedian_r<10 a.snMedian_i<10 a.snMedian_z<10;
```

B. Baseline Data SQL appendix

```
select b.run, b.camcol, b.field, a.specObjID, a.bestObjID,
a.class, a.z, a.zErr, b.objID, b.psfMag_u, b.psfMag_g,
b.psfMag_r, b.psfMag_i, b.psfMag_z, b.extinction_u,
b.extinction_g, b.extinction_r, b.extinction_i, b.extinction_z,
b.psfMagErr_u, b.psfMagErr_g, b.psfMagErr_r,
b.psfMagErr_i, b.psfMagErr_z, b.dered_u, b.dered_g, b.dered_r, b.dered_i, b.dered_z
into mydb.dr18psf_ugriz_SNR_z0
from dr17.SpecObjAll as a
join dr17.PhotoObjAll as b
on a.bestObjID = b.objID
where a.class ='QSO' and a.z >=1 and a.z<5 and a.zErr< 0.001 and b.psfMag_g >= 18 and
b.psfMag_g <= 22 and b.petroRad_r < 5 and b.flags is not null and zWarning=0 and b.clean=1;
```
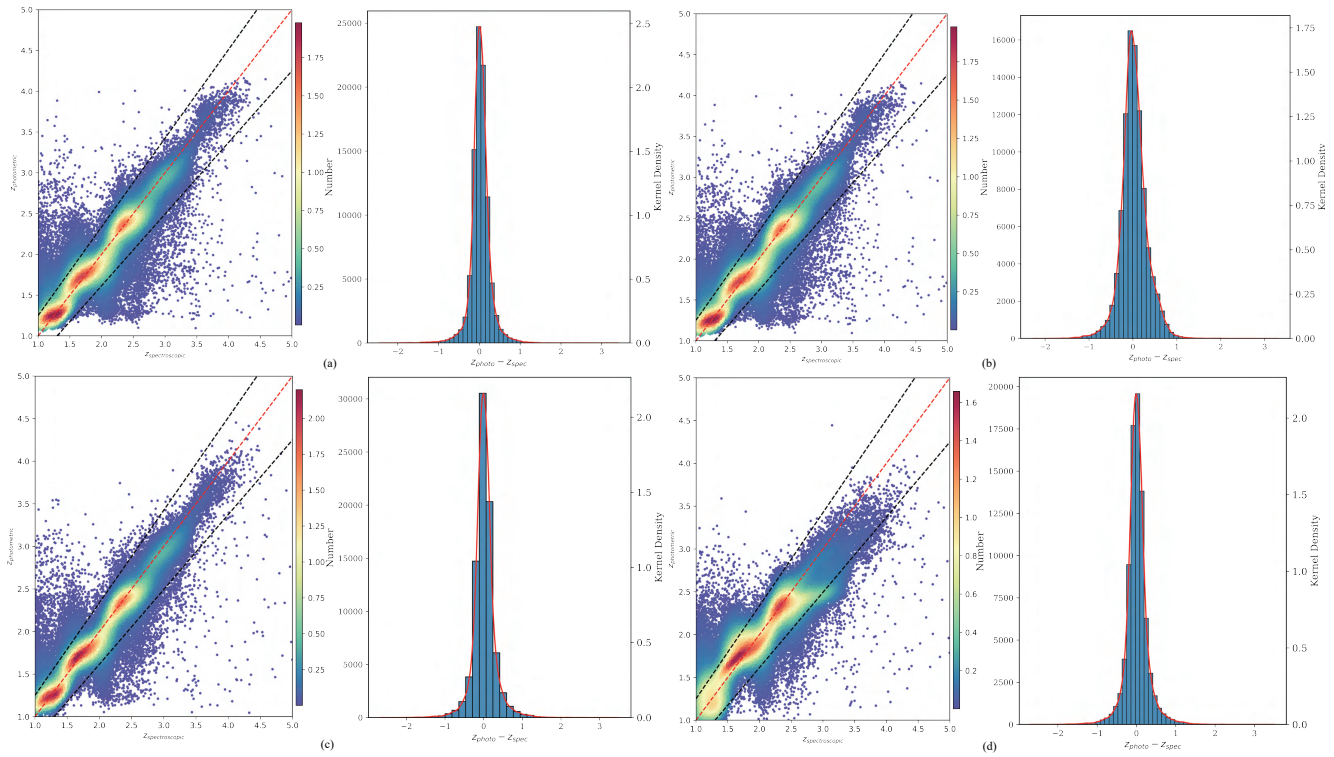
C. Figure appendix

Figure 3. Photometric redshift estimation visualization of our method corresponding to results of table 1:On the baseline's dataset,(a) shows the predictions obtained using the real image and photometric data features.(b) displays the results obtained using the real image and generated photometric data features. On the extended dataset *(465,292 samples)*,figure (c) illustrates the predictions achieved using the real image and photometric data features, while (d) depicts the results obtained using the real photometric data and generated image data features.