Photometric Redshift Estimation from Multi-Band Galaxy Images using CNNs

Atharva Surve, Prajot Patil, Ariyan Muddapur, Anshuman Padhi, Pramod H. Kachare[1]

[1]Department of Computer Science Engineering, Ramrao Adik Institute of Technology, D. Y. Patil Deemed University, Navi Mumbai, 400706, Maharashtra, India.

Abstract

This paper is motivated by the limitations of existing photometric redshift estimation methods, which often struggle with the variability of astronomical data and require complex preprocessing steps. To address these challenges, the project explores advanced techniques such as deep learning and hybrid approaches to develop a robust, scalable, and lightweight model for redshift estimation capable of handling large, multi band galaxy image datasets. The study proposes a Convolutional Neural Network (CNN) for photometric redshift estimation using data from the NERSC Self-Supervised Learning SDSS dataset, comprising 399,982 galaxy images observed through five optical filters (u, g, r, i, z), each associated with spectroscopic redshift values and photometric measurements. The model, which follows a standard deep learning architecture with convolutional blocks for feature extraction and fully connected layers for regression, takes input images of $\mathbf{32 \times 32}$ pixels with 5 channels corresponding to the photometric bands. Trained on approximately 400,000 galaxies, the model achieves a Mean Absolute Error (MAE) of 0.0304, a root mean square error (RMSE) of 0.0412, and an $\boldsymbol{R^2}$ score of 0.9041, with evaluations showing better generalization when preserving the natural redshift distribution. Future work will focus on cross-dataset generalization, enhancing model efficiency for deployment, developing automated region of interest (ROI) detection, and analyzing photometric band importance to further improve the robustness, scalability, and automation of the redshift estimation pipeline.

Keywords: Photometric Redshift Estimation, Convolutional Neural Networks, Deep Learning, Galaxy Images, SDSS, Multi-band Imaging, Spectroscopic Redshift, NERSC Dataset, Precision Cosmology, Real-time Processing, Astronomical Surveys, Model Scalability, Automated ROI Detection, Cross-Dataset Generalization, Band Importance Analysis

1 Introduction

Traditional photo-z estimation methods fall into two categories: template fitting and machine learning approaches to process image data for photometric redshift estimation. A number of researchers have explored template-fitting techniques for redshift estimation; Arnouts et al (1999) investigated the evolution of galaxy clustering with redshift using Hubble Deep Field North (HDF-N) photometric redshift data, employing deep multiband photometric observations to estimate redshifts and analyse the angular correlation function. The authors applied statistical tools like the two-point angular correlation function and utilised a semi-analytic galaxy formation model based on Cold Dark Matter (CDM) cosmology. Key parameters included matter density (Ω_0), cosmological constant (Λ_0), redshift distribution, and dark matter halo mass. Metrics such as the amplitude of the angular correlation function and galaxy bias as functions of redshift were used to quantify clustering strength and its evolution. Ben´ıtez (2000) proposed the Bayesian Photometric Redshift (BPZ) method to improve photo-z accuracy by incorporating prior knowledge of galaxy types and redshift distributions. Using HDF-N photometry along with spectroscopic samples from CFRS and the Caltech redshift survey, the model estimated redshifts probabilistically through Bayesian inference rather than classification, assigning posterior probabilities to each redshift and galaxy type. It employed SED templates from Coleman, Wu, and Weedman and Kinney et al., with key parameters including galaxy magnitude, photometric colors, and prior distributions. Performance was measured using RMS scatter in $\Delta z/(1+z)$, the rate of catastrophic outliers, and the confidence in posterior based predictions. Similarly, Bolzonella et al (2000) evaluated photometric redshift estimation through Spectral Energy Distribution (SED) fitting techniques by systematically comparing results using various template sets, extinction laws, and photometric systems. Using datasets such as HDF-N, HDF-S, CFRS, and ESS, which provided spectroscopic redshifts for validation, they implemented the Hyperz code based on χ^2 minimization to match observed photometry with galaxy templates from the Bruzual & Charlot GISSEL library and the Coleman 1980 spectra. Parameters included galaxy redshift, age, metallicity, star formation history, and extinction, with laws like Calzetti and SMC considered. Performance was measured using metrics such as Δz, its standard deviation σ, and the rate of catastrophic outliers defined as $|\Delta z| > 0.5$. Massarotti et al (2001) investigated the accuracy of photo-z estimates using HDF-N and HDF-S datasets, applying SED fitting with template

libraries from Buzzoni, Bruzual & Charlot, and Fioc & Rocca-Volmerange. The model was not a classifier but relied on χ^2 minimization against theoretical SEDs. Key parameters included redshift, dust extinction, and IGM absorption. Metrics such as σ_z and Δz were used; for $z_{\text{spec}} < 1.5$, $\sigma_z = 0.084$ and $\Delta z = -0.033$, while improved modeling for $z_{\text{spec}} \geq 2.0$ yielded $\sigma_z = 0.117$ and $\Delta z = 0.048$, highlighting the critical role of ISM and IGM treatment at high redshift. Brammer et al (2008) developed EAZY, a flexible photometric redshift estimation code aimed at deep extragalactic surveys without representative spectroscopic samples. They used multi-band photometric data from GOODS, MUSYC, and other deep fields, applying χ^2 minimization with SED templates derived via non-negative matrix factorization from PÉGASE models. Instead of a classifier, EAZY directly estimated redshifts by marginalizing over template uncertainties and incorporated a K-band magnitude prior. Model parameters included template combinations, extinction laws, and prior functions. Performance was evaluated using metrics such as normalized median absolute deviation (NMAD), catastrophic outlier fraction ($\Delta z > 0.15$), and redshift PDF width. Arnouts and Ilbert (2011) presented LePHARE, a photometric redshift estimation tool designed to replace time-intensive spectroscopic methods using multi-band photometric data. The software applied χ^2 minimisation to fit observed fluxes to template-based approaches. Template fitting methods use synthetic or empirical galaxy spectral energy distribution (SED) templates to match observed photometry, enabling redshift estimation across wide ranges with minimal training data. Several recent studies adopt a similar strategy of theoretical or empirical SED templates generated from PÉGASE and Bruzual & Charlot models. Instead of a classifier, LePHARE directly inferred redshifts by optimising parameters such as redshift, galaxy type, age, and dust extinction. The datasets included photometric catalogs from multiple surveys, making the model broadly applicable. Performance was evaluated by comparing photometric redshifts to spectroscopic ones, using metrics like bias, scatter, and outlier fraction.

In contrast to template fitting, machine learning (ML) methods approach photometric redshift estimation as a data-driven regression problem, learning the mapping between observed photometric features and spectroscopic redshifts from a training set. One of the earliest contributions in this area is the work by Tagliaferri et al (2003), who proposed a neural network-based method utilizing a Multi-Layer Perceptron (MLP) within a Bayesian framework. Applied to the Sloan Digital Sky Survey Early Data Release (SDSS-EDR), their model achieved accuracy on par with, and in some cases superior to, SED template fitting techniques. To further evaluate model performance and interpret classification boundaries, the authors employed a Self-Organizing Map (SOM) to visualize contamination between galaxy classes based on photometric predictions. The network achieved a robust redshift estimation error of 0.020 for $0 < z_{\text{phot}} < 0.3$, and 0.022 for $0 < z_{\text{phot}} < 0.5$, demonstrating that neural networks could effectively capture non-linear mappings in the data and

offer competitive performance in low-to-intermediate redshift regimes. Similarly, Wolf (2009) introduced a hybrid Bayesian framework that bridges the gap between empirical training-set-based methods and template-fitting techniques by integrating photometric error modeling and Bayesian priors into a locally weighted regression or nearest-neighbor approach. Rather than relying on fixed spectral templates, this method estimates redshift distributions using the local density of known redshifts in color space derived from a spectroscopic training set. Posterior redshift probability distributions are calculated by convolving likelihoods from the training data with priors typically based on magnitude-redshift relationships. The model implicitly performs regression and probabilistic classification via redshift PDFs. Key parameters include the bandwidth for local weighting, choice of priors, and neighborhood kernel configuration. Tested on the COMBO-17 survey and generalized to datasets like SDSS, its performance was evaluated using metrics such as RMS error, redshift bias, outlier rate, and PDF calibration quality. The results demonstrated that empirical methods enhanced with Bayesian inference can match or even exceed the performance of both template-based and more complex ML models, particularly when applied to heterogeneous photometric data. Sadeh et al (2015) proposed ANNz2, a machine learning-based tool for photometric redshift estimation and the generation of corresponding redshift probability distribution functions (PDFs). They utilized a dataset from the Sloan Digital Sky Survey (SDSS) Data Release 10, comprising approximately 180,000 galaxies with photometric measurements in five bands (u, g, r, i, z) and corresponding spectroscopic redshifts for supervised learning. ANNz2 integrates an ensemble of classifiers, primarily artificial neural networks (ANNs) and boosted decision/regression trees (BDTs), implemented via the TMVA package within the ROOT framework. The model explores multiple architectures with varied hyperparameter ANNs with 2 to 4 hidden layers and diverse activation functions (e.g., sigmoid, tanh), and BDTs using 300 to 1200 trees with different boosting strategies and splitting criteria. Input features were further processed through normalization and principal component analysis. The performance of ANNz2 was evaluated using several key metrics: bias (mean of $z_{\text{phot}} - z_{\text{spec}}$) to assess systematic offset, scatter including σ and σ_{68} (where $\sigma_{68} \approx 0.03$), outlier fraction beyond $2\sigma_{68}$ and $3\sigma_{68}$ (reported as less than 2% for well-behaved subsets), the Kolmogorov–Smirnov (KS) statistic to evaluate the similarity between predicted and true redshift distributions, and N_{poisson} to quantify Poisson fluctuations in binned redshift estimates. The study found a redshift bias of < 0.001, low scatter, and robust performance across various subsets, demonstrating that ANNz2 achieves high accuracy and well-calibrated uncertainty estimates. These results underscore its applicability in current and future photometric galaxy surveys like the Dark Energy Survey (DES). Cavuoti et al (2017) introduced METAPHOR (Machine-learning Estimation Tool for Accurate PHOtometric Redshifts), a novel method designed to provide reliable probability density functions (PDFs) for photometric redshift (photo-z) estimates. The method employs a modular workflow, utilizing the

Multi-Layer Perceptron with Quasi-Newton Algorithm (MLPQNA) as the internal engine for photo-z estimation, with the flexibility to replace it with other machine learning models. The authors validated METAPHOR using Sloan Digital Sky Survey (SDSS) Data Release 9 galaxy data and compared its performance with the Le Phare spectral energy distribution (SED) template fitting method. The results demonstrated that METAPHOR effectively estimated photo-z PDFs, providing a robust framework for photometric redshift estimation in large-scale galaxy surveys. Polsterer et al (2021) employed several performance metrics to evaluate probabilistic predictions in their interdisciplinary study, focusing on photometric redshift estimation and weather forecasting. The key metrics used include the Continuous Ranked Probability Score (CRPS), which assesses the difference between the predicted and true cumulative distribution functions, providing a comprehensive evaluation of the entire predicted probability distribution. Another crucial metric is the Probability Integral Transform (PIT), which tests the calibration of probabilistic models by comparing predicted probabilities with actual outcomes, with well-calibrated models yielding uniformly distributed PIT values. The Log-Likelihood metric was also used to measure how well the predicted probability distributions fit the observed data, with higher log-likelihood values indicating a better model fit. These scoring rules are particularly useful for evaluating models that generate full probability distributions, offering deeper insights into both the accuracy and sharpness of the predictions. By applying these metrics to both redshift estimation and weather forecasting, the authors demonstrated the power of proper scoring rules in improving model evaluation across different scientific fields. Chan and Stott (2021) present a novel machine learning approach called Z-Sequence, which utilizes a Sequential Random k-Nearest Neighbours (SRKNN) model to enhance the accuracy of photometric redshift (photo-z) predictions, particularly for galaxies in clusters. The method improves upon traditional kNN by introducing a sequential correction mechanism that refines predictions over multiple iterations, using previous estimates as inputs to subsequent rounds. The authors evaluate their approach using simulated data from the MICE2 simulation and real observational data from SDSS DR15. The classifier employed is based on a modified kNN algorithm, and the model parameters include the number of neighbours (e.g., k = 5, 10, 20, 50), number of iterations (typically 3), and Euclidean distance computed on normalized photometric features. Performance is assessed using metrics such as bias, Normalized Median Absolute Deviation (NMAD), and outlier fraction. The results show that Z-Sequence outperforms traditional kNN and Random kNN (RKNN) models, achieving bias as low as 0.0001, NMAD around 0.015, and outlier fractions below 5%, indicating strong improvements in accuracy and reliability of redshift estimation. Li et al (2022) primarily aim to estimate the photometric redshifts (photo-z) of quasars using machine learning techniques, specifically focusing on data from the BASS DR3 survey. The researchers constructed a catalog of 214,076 quasars by combining the BASS DR3, MzLS, WISE, and SDSS DR16Q datasets. They employed the XGBoost

classifier as the main model to predict redshifts and used photometric features such as magnitudes and colors from different bands as input parameters. The performance of their model was evaluated using metrics including the Normalized Median Absolute Deviation (NMAD), root mean square (RMS), and outlier rate. Their best-performing model achieved an NMAD of 0.078, an RMS of 0.145, and an outlier rate of 20.45%, indicating a robust performance in estimating quasar redshifts based solely on photometric data. Qu and Sako (2023) developed Photo-zSNthesis, a random forest-based method to estimate redshifts of Type Ia supernovae using photometric lightcurve features from the LSST DESC Data Challenge 2. Features were derived from SALT2 lightcurve fits, and the model was evaluated using NMAD = 0.017 and an outlier fraction of 2.7%, demonstrating high accuracy without requiring spectroscopic data, addressing the follow-up challenges posed by large-scale future surveys.

Deep learning has emerged as a powerful approach for photometric redshift estimation due to its ability to automatically learn complex, non-linear relationships from large and high-dimensional datasets. Unlike traditional methods that rely on handcrafted features such as magnitudes and colors, deep learning models can directly process raw galaxy images or flux measurements, capturing both morphological and spectral information. D'Isanto and Polsterer (2017) developed a deep learning model combining a deep convolutional network (DCN) and a mixture density network (MDN) to estimate probabilistic photometric redshifts from SDSS DR9 imaging data. This approach required no manual feature extraction and instead output Gaussian mixture models as redshift PDFs. Key parameters included the design of the convolutional layers and MDN, with performance evaluated using continuous ranked probability score (CRPS) and probability integral transform (PIT). Salvato et al (2018) conducted a comprehensive review of photo-z estimation methods, categorizing them into template-fitting, machine learning, and hybrid techniques. They did not propose a new model but evaluated existing methods such as LePhare, BPZ, ANNz, and random forests across surveys like COSMOS, SDSS, CFHTLS, and KiDS. The paper identified critical factors influencing performance, including photometric depth, wavelength coverage, redshift distribution, and the completeness of spectroscopic training data, emphasizing the need to tailor methods to specific survey characteristics and object types. Pasquet et al (2019) approached photometric redshift estimation using image data instead of derived catalog features. The main idea is to leverage deep convolutional neural networks (ConvNets) on multi-band image cutouts from SDSS DR10, allowing a data-driven method that avoids manual feature engineering. The dataset consists of galaxy images with corresponding spectroscopic redshift labels. The classifier used is a CNN model with convolutional, max-pooling, and fully connected layers. Key parameters include kernel sizes, strides, number of filters, learning rate, and dropout rates. The model achieves an RMSE of 0.021, performing on par with or better than traditional methods like random forests and k-nearest neighbors. Beck et al (2020) introduced PS1-STRM, a neural network-based model for both source classification and

photometric redshift estimation using Pan-STARRS1 (PS1) 3π survey data. The dataset included over 2.9 billion sources, with a spectroscopic training set of 3.85 million from various surveys. Two separate feedforward neural networks (using Keras and TensorFlow) were implemented: one for classification (galaxy, star, quasar) and another for redshift regression, each with three hidden layers of 512 neurons. The classification model achieved accuracies above 96%, and the redshift model achieved a bias of 0.0005, $\sigma(\Delta z_{\rm norm}) = 0.0322$, median absolute deviation = 0.0161, and an outlier fraction of 1.89%. Hayat et al (2021) proposed a self-supervised learning framework for photometric redshift estimation using contrastive learning on SDSS multi-band images, leveraging unlabeled data. A ResNet50 CNN was pretrained and fine-tuned for photo-z estimation, with augmentations including rotations, Gaussian noise, and galactic reddening. Their model achieved a normalized median absolute deviation ($\sigma_{\rm MAD}$) of 0.015 and an outlier fraction (η) of 2.5%, performing comparably or better than supervised methods while using significantly fewer labeled examples. Henghes et al (2022) used a ResNet-18-based CNN to estimate photo-zs directly from 64×64 SDSS DR12 RGB galaxy images, using 58,246 galaxies with spectroscopic redshifts as ground truth. Training involved 500 epochs, a batch size of 128, Adam optimizer with a learning rate of 10^{-4}, and standard augmentations. Their best model achieved RMSE = 0.0202, MAE = 0.0154, NMAD = 0.0107, and a catastrophic outlier rate of 3.18%, demonstrating that image-based CNNs can compete with tabular feature-based models. Schuldt et al (2022) developed NetZ, a CNN-based model for estimating photometric redshifts directly from 96×96 pixel SDSS DR10 cutouts in the ugriz bands. The model incorporated convolutional and pooling layers followed by fully connected layers, with dropout regularization. It achieved a normalized redshift bias of $\langle \Delta z \rangle = 0.001$, NMAD = 0.030, and outlier fraction $\eta = 3.2\%$, showing performance comparable to or better than traditional ML and template-fitting approaches. Lin et al (2022) explored biases in data-driven photo-z methods by developing a CNN trained on SDSS galaxy images combined with auxiliary features like Petrosian radius and axis ratio. Their architecture captured both color and morphological information, achieving an NMAD of 0.018 and a bias of ~ 0.002, showing improvement in bias reduction compared to conventional models. Zhou et al (2022) introduced a Bayesian neural network approach for photo-z estimation in the China Space Station Telescope (CSST) survey using flux and image data. They proposed Bayesian MLP and CNN models, combining them into a hybrid network via transfer learning. The hybrid network achieved $\sigma_{\rm NMAD} = 0.021$ and $\eta = 1.23\%$, improving to $\sigma_{\rm NMAD} = 0.019$ and $\eta = 1.17\%$ with transfer learning, thereby enabling robust uncertainty estimation crucial for weak lensing and clustering analyses. Merz et al (2024) propose a method to estimate photometric redshifts using a convolutional neural network (CNN) trained directly on multi-band astronomical image cutouts, rather than relying on derived features like magnitudes or colors. They utilize image data from the Sloan Digital Sky Survey (SDSS) DR12, specifically selecting

150,000 galaxies with clean spectroscopic redshifts (zWarning = 0) and ensuring uniform redshift distribution through binning. The model used is a deep CNN with four convolutional layers, ReLU activations, batch normalization, dropout, and a fully connected layer outputting a single scalar redshift value. Key hyperparameters include a batch size of 128, learning rate of 0.0001, dropout rate of 0.2, and the use of the Adam optimizer. Performance metrics reported include a median redshift bias of -0.0004, median absolute deviation (MAD) of 0.0184, and an outlier fraction ($\Delta z > 0.1$) of 0.0127, indicating the model's strong predictive accuracy compared to traditional methods.

2 Data

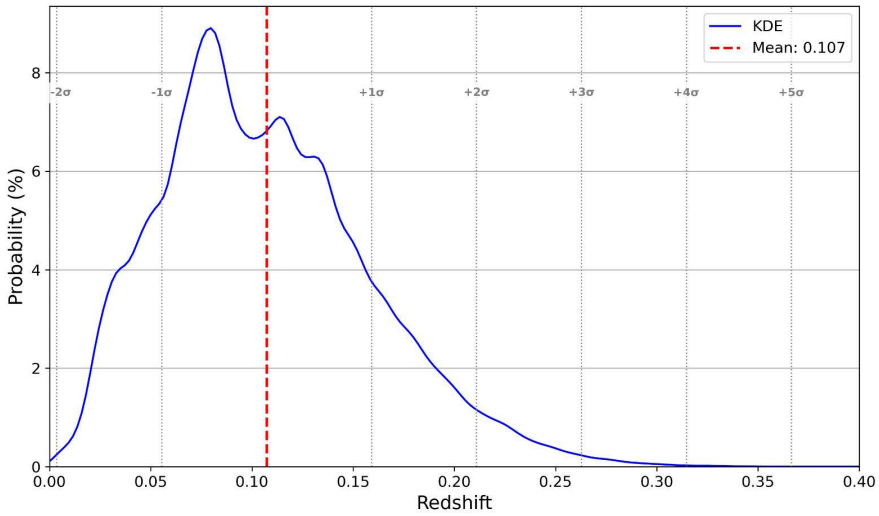


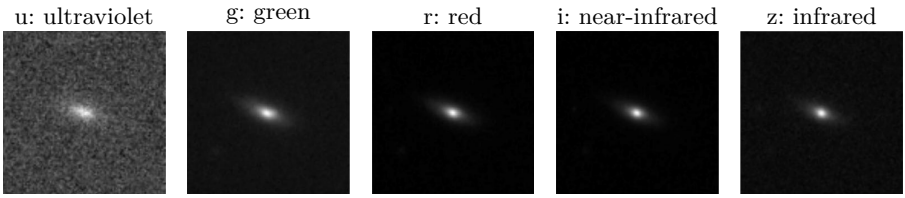
Fig. 1: Probability density function of redshift for the SDSS dataset

In this study, the Sloan Digital Sky Survey (SDSS) dataset provided by the National Energy Research Scientific Computing Center (NERSC), specifically Data Release 12 (DR12), is used. This dataset (Hayat et al (2021)) comprises 399,982 galaxy images with corresponding spectroscopic redshifts and photometric measurements across five passbands: ultraviolet (u), green (g), red (r), near-infrared (i), and infrared (z). These bands sample different regions of the electromagnetic spectrum, providing essential information for redshift inference (redshifts in the range $0.0 - 0.40$).

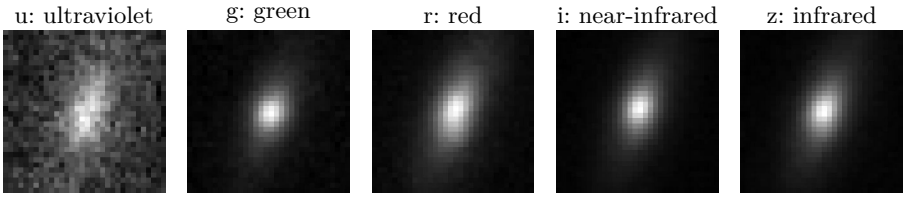
Each image is preprocessed to a fixed resolution of 32×32 pixels, cropped from the center of the original SDSS cutouts to maintain consistent image dimensions. The 32×32 pixel size was found to be sufficiently large to capture the full galaxy and surrounding sky even for nearby galaxies, while also allowing for significantly smaller file sizes. The final data representation is a $32 \times 32 \times 5$ pixel datacube, with the five channels corresponding to the *ugriz* photometric filters. This image projection preserves the native SDSS pixel scale of 0.396 arcseconds per pixel and is aligned with the celestial coordinate system. Additionally, the redshift values used for training the model were normalized to the $[0, 1]$ range to facilitate stable and efficient learning.

Table 1: Characteristics of the SDSS dataset

Feature	**Value**
Number of images	399,982
Original image size	$107 \times 107 \times 5$
Pre-processed image size	$32 \times 32 \times 5$
Redshift - Min	0.0
Redshift - Mean	0.1072
Redshift - Max	0.3995



(a) Unprocessed photometric images



(b) Preprocessed photometric images

Fig. 2: Comparison of five-band photometric images for the same object before (a) and after (b) preprocessing. Each band represents a specific wavelength: u (ultraviolet), g (green), r (red), i (near-infrared), and z (infrared).

3 Proposed Model

The paper proposes a convolutional neural network (CNN) for photometric redshift estimation from galaxy images observed through five different filters. The network follows a standard deep learning architecture, consisting of convolutional blocks for feature extraction, followed by fully connected layers for regression, and culminating in a final prediction layer.

The model accepts input images of size 32×32 pixels with 5 channels corresponding to the photometric bands (u: ultraviolet, g: green, r: red, i: near-infrared, z: infrared). Feature extraction is performed using two convolutional blocks: the first block applies a 2D convolutional layer with 32 filters of size 5×5 and ReLU activation, followed by a 2×2 max-pooling layer; the second block similarly uses 64 filters with the same kernel and activation, again followed by 2×2 max pooling.

The resulting feature maps are then flattened and passed through two fully connected layers with 220 and 64 units, respectively, both using ReLU activations. To prevent overfitting, dropout regularization with a rate of 0.5 is applied between the fully connected layers. Finally, a dense output layer with sigmoid activation generates the normalised redshift prediction scaled between $[0, 1]$.

The overall architecture comprises 421,725 trainable parameters and is trained end to end.

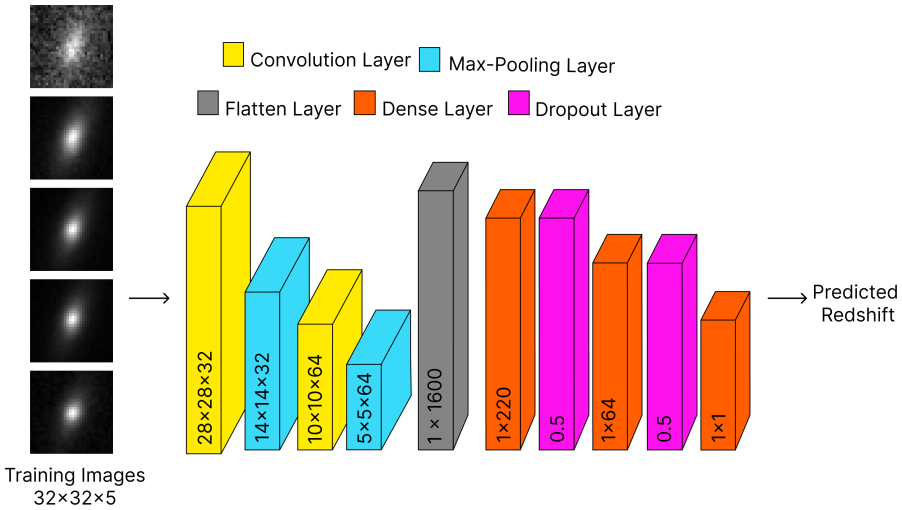


Fig. 3: Proposed CNN architecture for photometric redshift estimation

Table 2: CNN Model Architecture for Photometric Redshift Estimation

Layer	Kernel	Output Shape	Parameters
Input Layer	–	(32, 32, 5)	0
Conv_1	5×5	(28, 28, 32)	4,832
MaxPooling_1	2×2	(14, 14, 32)	0
Conv_2	5×5	(10, 10, 64)	51,264
MaxPooling_2	2×2	(5, 5, 64)	0
Flatten	–	(1, 1600)	0
Dense	–	(1, 220)	352,220
Dropout	0.5	–	0
Dense	–	(1, 64)	14,144
Dropout	0.5	–	0
Dense	–	(1, 1)	65
Total Parameters			**421,525**

4 Methodology

The present work proposes an end-to-end supervised regression approach for photometric redshift estimation using raw multi-band galaxy images from the Sloan Digital Sky Survey (SDSS) DR12. Unlike traditional methods that rely on hand-crafted photometric features, the model directly accepts input image cubes of size $32 \times 32 \times 5$, where the five channels correspond to the

ugriz photometric bands. These image cubes preserve both spatial and spectral information, allowing the convolutional neural network (CNN) to learn complex mappings of redshift.

To assess the effect of possible outliers (i.e., extreme redshift values) on model performance, we constructed two versions of the dataset: the full (uncorrected) dataset, and a clipped variant. In the clipped version, we restricted redshift values to lie within two standard deviations ($\pm 2\sigma$) of the mean, preserving approximately 95% of the data and removing extreme outliers contributing to the skewness and long-tail behaviour observed in the redshift distribution. This preprocessing step was intended to reduce the impact of outlier, high-redshift galaxies that could bias the learning process.

In addition to the clipped dataset, we also trained a separate CNN model on the full (uncorrected) version of the dataset, which includes all 399,982 galaxy images. This version preserves the original redshift distribution, including the long-tailed behaviour introduced by high-redshift values. The corresponding histogram is shown in Figure **??**, highlighting the natural skewness present in the spectroscopic redshift values. By training on the full dataset, we aim to evaluate how the presence of extreme redshift values influences model performance, generalisation, and sensitivity to underrepresented regions of the redshift space.

5 Model Evaluation

5.1 Evaluation metrics

To assess the performance of the redshift estimation model, we employ four standard regression metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Coefficient of Determination (R^2), and Adjusted R^2. These metrics offer complementary perspectives on model accuracy, sensitivity to error magnitude, and model generalization. The equations for each metric is given below for a dataset with n galaxies where for the i^{th} galaxy, $\hat{z}_i$ is the predicted redshift, and z_i is the true spectroscopic redshift.

5.1.1 Mean absolute error (MAE)

Definition:
The Mean Absolute Error measures the average magnitude of redshift prediction errors, without considering their direction. It provides a straightforward measure of the model's accuracy in the same units as the predicted variable.

Formula:

$$\text{MAE}(z, \hat{z}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} |z_i - \hat{z}_i|$$

5.1.2 Root mean square error (RMSE)

Definition:
The Root Mean Square Error calculates the square root of the average squared

differences between predicted and actual redshift values, penalizing larger errors more than MAE.

Formula:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=0}^{n-1}(z_i - \hat{z}_i)^2}$$

5.1.3 Adjusted R²

Definition:
Adjusted R^2 adjusts the standard R^2 value to account for the number of independent variables used in the model, offering a more reliable measure when dealing with multiple predictors.

Formula:

$$\text{Adjusted } R^2 = 1 - \left(\frac{(1 - R^2) \times (n - 1)}{n - p - 1}\right)$$

where:

- p is the number of features (predictors) in the model.

5.1.4 Precision

Definition:
Precision measures the proportion of true positive predictions among all positive predictions made by the model. It indicates how accurate the model is when it predicts the positive class.

Formula:

$$\text{Precision} = 1.48 \times \text{Median}\left(\frac{z_{\text{pred}} - z_{\text{spec}}}{1 + z_{\text{spec}}}\right)$$

where:

- z_{pred} is the predicted redshift value,
- z_{spec} is the true spectroscopic redshift value.

5.1.5 Bias

Definition:

Bias measures the average difference between predicted and actual redshift values. It provides an indication of the model's systematic error. A bias of 0 indicates that the model, on average, neither overestimates nor underestimates the redshift values.

Formula:

$$\text{Bias}(z, \hat{z}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} (\hat{z}_i - z_i)$$

5.1.6 Catastrophic Outlier Fraction

Definition:

The Catastrophic Outlier Fraction measures the proportion of galaxies for which the predicted redshift deviates significantly from the true redshift, typically by more than a certain threshold(often set to)

$$\frac{\Delta z}{1+z} > 0.15$$

$$\text{Catastrophic Outlier Fraction} = \frac{N(\Delta z > p)}{N_{\text{total}}}$$

where

$$\Delta z = \frac{\text{- z}_{\text{spec}}}{1+z_{\text{spec}}}$$

where **1** is the indicator function, which returns 1 if the condition is true and 0 otherwise.

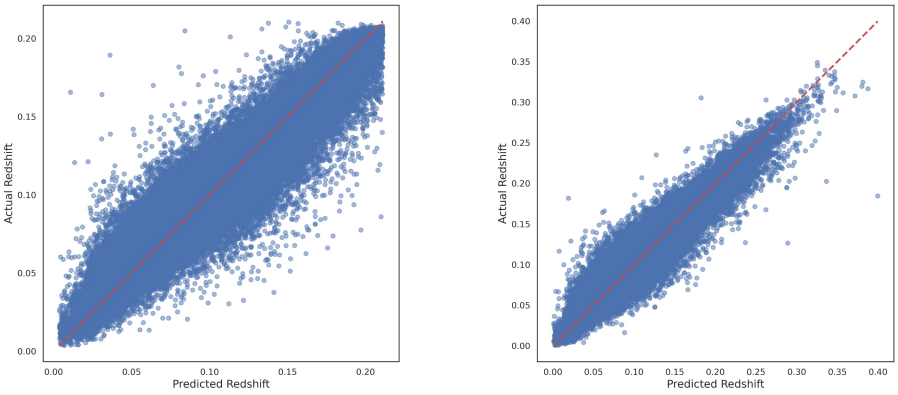
5.2 Results

Metric	Full Dataset	Clipped Dataset
MAE	0.0304	0.0556
RMSE	0.0412	0.0751
Adjusted R2	0.9041	0.8825
Precision	0.0269	0.0427
Bias	0.00029	-0.00067
Cat. Outliers		
$\Delta z > 0.05$	10.86%	27.32%
$\Delta z > 0.10$	0.91%	5.90%
$\Delta z > 0.15$	0.12%	1.28%

Explanation:

The model performs better on the full dataset across all metrics. It achieves lower MAE (0.0304) and RMSE (0.0412), indicating more accurate and stable predictions. The R^2 and Adjusted R^2 scores (both 0.9041) show that over 90% of redshift variance is captured. The correlation coefficient (0.9515) further confirms strong predictive alignment.

In contrast, the clipped dataset shows reduced performance, suggesting that removing outliers may actually limit the model's ability to generalize across the natural redshift distribution.



(a) (b)

Fig. 4: Scatter plots of predicted vs. actual redshift values. (a) shows results from the clipped dataset, while (b) corresponds to the unclipped dataset.

5.3 Discussion

5.3.1 Summary of work

This study presents a deep learning approach for photometric redshift estimation using a Convolutional Neural Network (CNN) trained on multiband galaxy images from SDSS DR12. The model processes $32 \times 32 \times 5$ input tensors representing the ugriz bands and directly predicts redshift values. Trained on approximately 400,000 galaxies, the model achieves performance with a Mean Absolute Error of 0.0304, RMSE of 0.0412, and an R^2 score of 0.9041. Evaluation across full and clipped datasets shows that retaining the natural redshift distribution yields better generalisation.

5.3.2 Heatmap

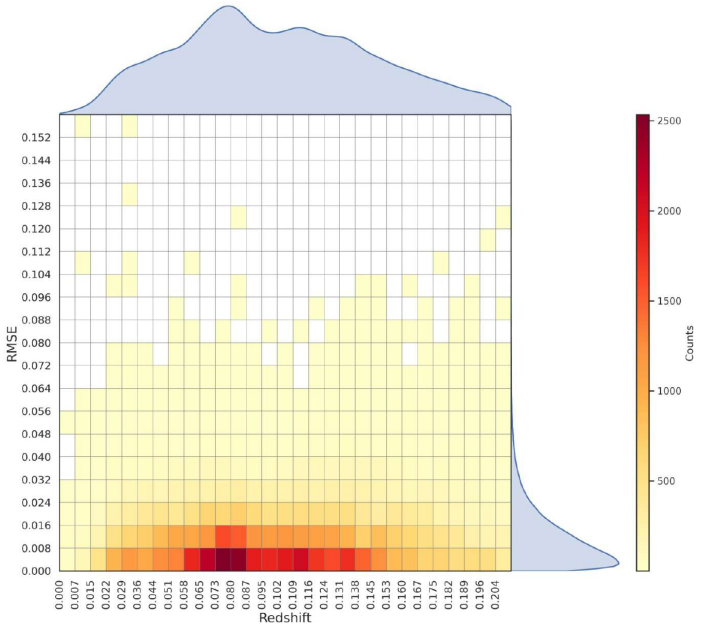


Fig. 5: Absolute redshift estimation error distribution using the proposed CNN model, actual redshift values without outliers in the SDSS dataset (clipped).

The performance metrics of the model on the clipped photometric redshift data indicate strong predictive capability. The Mean Absolute Error (MAE) of 0.0556 and Root Mean Squared Error (RMSE) of 0.0751 reflect good overall accuracy, while a near-zero bias of -0.00067 suggests minimal systematic deviation in the predictions. The R^2 value of 0.8826 and adjusted R^2 of 0.8825 demonstrate that the model explains a substantial proportion of the variance in the target variable. The precision metric, defined as $1.48 \times \text{median}(\Delta z)$, is 0.0427, indicating a reliable typical prediction error.

The 2D heatmap in Fig. 5 visualizes the absolute redshift estimation error distribution across actual redshift values. The horizontal axis displays actual redshift values in the range 0.0036 to 0.2109 with an interval of 0.007 unit. The structures on the top and right sides of the heatmap are Gaussian-approximated probability density functions (PDFs). The plot on the top represents the distribution of the actual redshift values, while the plot on the right shows the distribution of the absolute redshift error. The vertical axis displays the absolute redshift estimation error in the range 0–0.154 unit with an interval of 0.004 unit. The highest density of predictions is concentrated

in the lower error range, within the redshift interval of 0.072 to 0.15. This aligns well with the reported MAE of 0.0556 and RMSE of 0.0751, as the bulk of the predictions are accurate, though the presence of a visible tail of larger errors in the right-side Gaussian-approximated PDF contributes to a slightly elevated average.

Fig. 6 further supports the model's effectiveness by presenting the evaluation metrics on the unclipped photometric redshift data. The Mean Absolute Error (MAE) of 0.0304 and Root Mean Squared Error (RMSE) of 0.0412 in Fig. 6 indicate low average prediction errors, suggesting the model's capability to produce reliable redshift estimates across the full data range. The bias of 0.00029 is effectively negligible, indicating that the model's predictions are centered very closely around true values, with no significant tendency toward over- or under-prediction. Furthermore, the R^2 score of 0.9041 and adjusted R^2 of 0.9041 show that the model captures over 90% of the variance in the true redshift values, reflecting strong explanatory power. The precision, computed as $1.48 \times \text{median}(\Delta z)$ and measured at 0.0269, further confirms tight error bounds for typical predictions, emphasizing the model's stability and robustness.

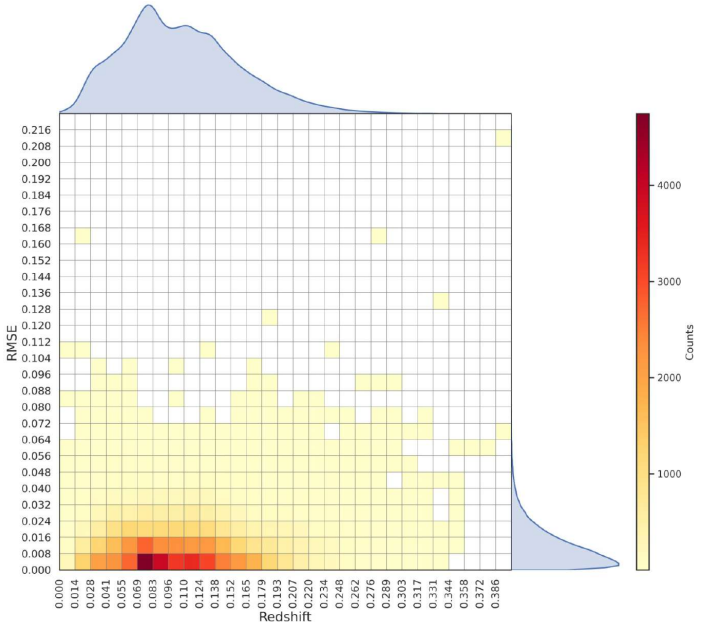


Fig. 6: Absolute redshift estimation error distribution using the proposed CNN model, actual redshift values without outliers in the SDSS dataset (unclipped).

The evaluation metrics on the unclipped photometric redshift data demonstrate that the model performs with high accuracy and stability. The Mean Absolute Error (MAE) of 0.0304 and Root Mean Squared Error (RMSE) of 0.1743 indicate low average prediction errors, suggesting the model's capability to produce reliable redshift estimates across the full data range. The bias of 0.00029 is effectively negligible, indicating that the model's predictions are centered very closely around true values, with no significant tendency toward over- or under-prediction. The R^2 score of 0.9041 and adjusted R^2 of 0.9041 show that the model captures over 90% of the variance in the true redshift values, reflecting strong explanatory power. The precision, computed as $1.48 \times \text{median}(\Delta z)$, is 0.0269, further confirming tight error bounds for typical predictions.

5.3.3 Key takeaway

The study presents a deep learning-based approach for estimating photometric redshifts (photo-z) using Convolutional Neural Networks (CNNs) trained directly on raw multi-band galaxy images from the Sloan Digital Sky Survey (SDSS) DR12. Unlike traditional methods that rely on hand-engineered features such as magnitudes or colors, the proposed model processes $32 \times 32 \times 5$ pixel image cubes (ugriz bands) to learn complex spatial and spectral patterns for redshift prediction. The study explores the impact of data distribution by training the model on two datasets: a full dataset preserving the natural redshift distribution, including outliers, and a clipped version excluding redshift outliers beyond two standard deviations. Results show that the CNN trained on the full dataset significantly outperforms the clipped version across all evaluation metrics, achieving a Mean Absolute Error (MAE) of 0.0304, R^2 of 0.9041, and a lower catastrophic outlier rate. This highlights the importance of retaining the full data distribution, including outliers, for building robust and generalizable models. The study demonstrates the potential of deep learning in astronomical image analysis and offers a scalable alternative to traditional photometric redshift estimation methods for large galaxy surveys.

References

Arnouts S, Ilbert O (2011) Lephare: Photometric analysis for redshift estimate. Astrophysics Source Code Library https://doi.org/https://ui.adsabs.harvard.edu/abs/2011ascl.soft08009A/abstract

Arnouts S, Cristiani S, Moscardini L, et al (1999) Measuring and modelling the redshift evolution of clustering:the hubble deep field north. Monthly Notices 363:540–556. https://doi.org/https://arxiv.org/abs/astro-ph/9902290

Beck R, Szapudi I, Flewelling H, et al (2020) Ps1-strm: neural network source classification and photometric redshift catalogue for ps1 3 dr1. Monthly Notices of the Royal Astronomical Society 500:1633–1644. https://doi.org/https://doi.org/10.1093/mnras/staa2587

Ben´ıtez N (2000) Bayesian photometric redshift estimation. The Astrophysical Journal 536:571–583. https://doi.org/https://arxiv.org/abs/astro-ph/9811189

Bolzonella M, Miralles JM, Pello' R (2000) Photometric redshifts based on standard sed fitting procedures. Astronomy and Astrophysics 363:476–492. https://doi.org/https://arxiv.org/abs/astro-ph/0003380

Brammer GB, van Dokkum PG, Coppi P (2008) Eazy: A fast, public photometric redshift code. The Astrophysical Journal 686:1503–1513. https://doi.org/https://arxiv.org/abs/0807.1533

Cavuoti S, Brescia M, Vellucci C, et al (2017) Probability density estimation of photometric redshifts based on machine learning. IEEE https://doi.org/https://ieeexplore.ieee.org/document/7849953

Chan MC, Stott JP (2021) Z-sequence: Photometric redshift predictions for galaxy clusters with sequential random k-nearest neighbours. Monthly Notices of the Royal Astronomical Society 503:6078–6097. https://doi.org/https://arxiv.org/abs/2104.11335

D'Isanto A, Polsterer KL (2017) Photometric redshift estimation via deep learning. Astronomy and Astrophysics 609. https://doi.org/https://arxiv.org/abs/1706.02467

Hayat MA, Stein G, Harrington P, et al (2021) Self-supervised representation learning for astronomical images. The Astrophysical Journal Letters 911. https://doi.org/https://arxiv.org/abs/2012.13083

Henghes B, Pettitt C, Thiyagalingam J, et al (2022) Deep learning methods for obtaining photometric redshift estimations from images. Monthly Notices of the Royal Astronomical Society 512:1696–1709. https://doi.org/https://arxiv.org/abs/2109.02503

Li C, Zhang Y, Cui C, et al (2022) Photometric redshift estimation of bass dr3 quasars by machine learning. Monthly Notices of the Royal Astronomical Society 509:2289–2303. https://doi.org/https://arxiv.org/abs/2110.14951

Lin Q, Fouchez D, Pasquet J, et al (2022) Photometric redshift estimation with convolutional neural networks and galaxy images: A case study of resolving biases in data-driven methods. Astronomy and Astrophysics https://doi.org/https://arxiv.org/abs/2202.09964

Massarotti M, Iovino A, Buzzoni A, et al (2001) New insights on the accuracy of photometric redshift measurements. Astronomy and Astrophysics 380:425–434. https://doi.org/https://arxiv.org/abs/astro-ph/0110292

Merz G, Liu X, Schmidt S, et al (2024) Deepdisc-photoz: Deep learning-based photometric redshift estimation for rubin lsst. The Open Journal of Astrophysics https://doi.org/https://arxiv.org/abs/2411.18769

Pasquet J, Bertin E, Treyer M, et al (2019) Photometric redshifts from sdss images using a convolutional neural network. Astronomy and Astrophysics 621. https://doi.org/https://arxiv.org/abs/1806.06607

Polsterer KL, D'Isanto A, Lerch S (2021) From photometric redshifts to improved weather forecasts: machine learning and proper scoring rules as a basis for interdisciplinary work. Instrumentation and Methods for Astrophysics https://doi.org/https://arxiv.org/abs/2103.03780

Qu H, Sako M (2023) Photo-zsnthesis: Converting type ia supernova lightcurves to redshift estimates. The Astrophysical Journal https://doi.org/https://arxiv.org/abs/2305.11869

Sadeh I, Abdalla FB, Lahav O (2015) Annz2 - photometric redshift and probability distribution function estimation using machine learning. Publications of the Astronomical Society of the Pacific 128:968. https://doi.org/https://arxiv.org/abs/1507.00490

Salvato M, Ilbert O, Hoyle B (2018) The many flavours of photometric redshifts. Nature Astronomy https://doi.org/https://arxiv.org/abs/1805.12574

Schuldt S, Suyu1 SH, Cañameras R, et al (2022) Photometric redshift estimation with a convolutional neural network:netz. Astronomy and Astrophysics 651. https://doi.org/https://arxiv.org/abs/2011.12312

Tagliaferri R, Longo G, Andreon S, et al (2003) Neural networks for photometric redshifts evaluation. Lecture Notes in Computer Science 2859. https://doi.org/https://doi.org/10.1007/978-3-540-45216-4_26

Wolf C (2009) Bayesian photometric redshifts with empirical training sets. Monthly Notices of the Royal Astronomical Society 397:520–533. https://doi.org/https://doi.org/10.1111/j.1365-2966.2009.14953.x

Zhou X, Gong Y, Meng XM, et al (2022) Photometric redshift estimates using bayesian neural networks in the csst survey. Research in Astronomy and Astrophysics 22. https://doi.org/https://arxiv.org/abs/2206.13696v1