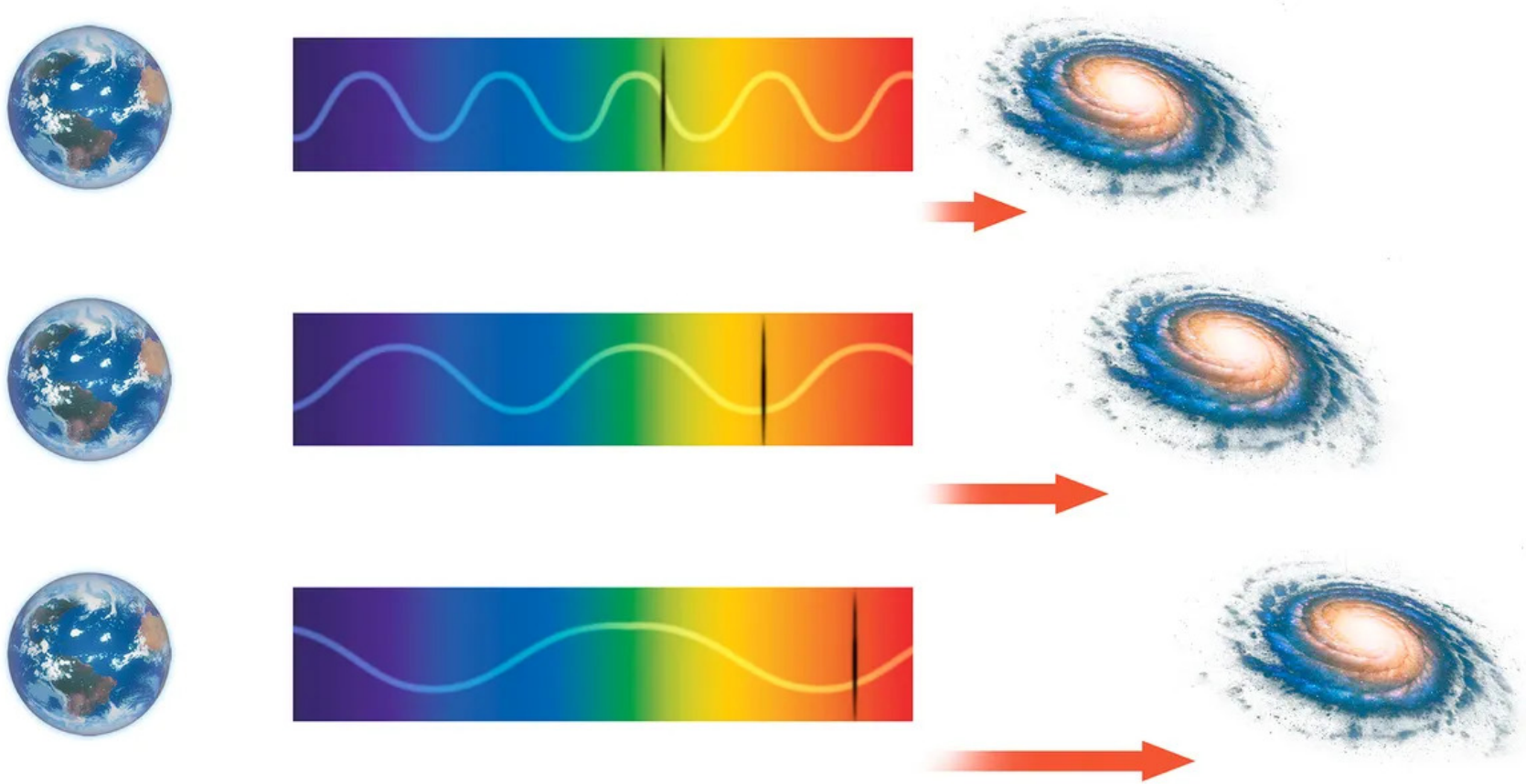
Inferring Photometric Redshifts from Multichannel Images

Balázs Menkó
Supervisor: Balázs Pál

December 12, 2024

Theoretical Background

What is cosmological redshift?



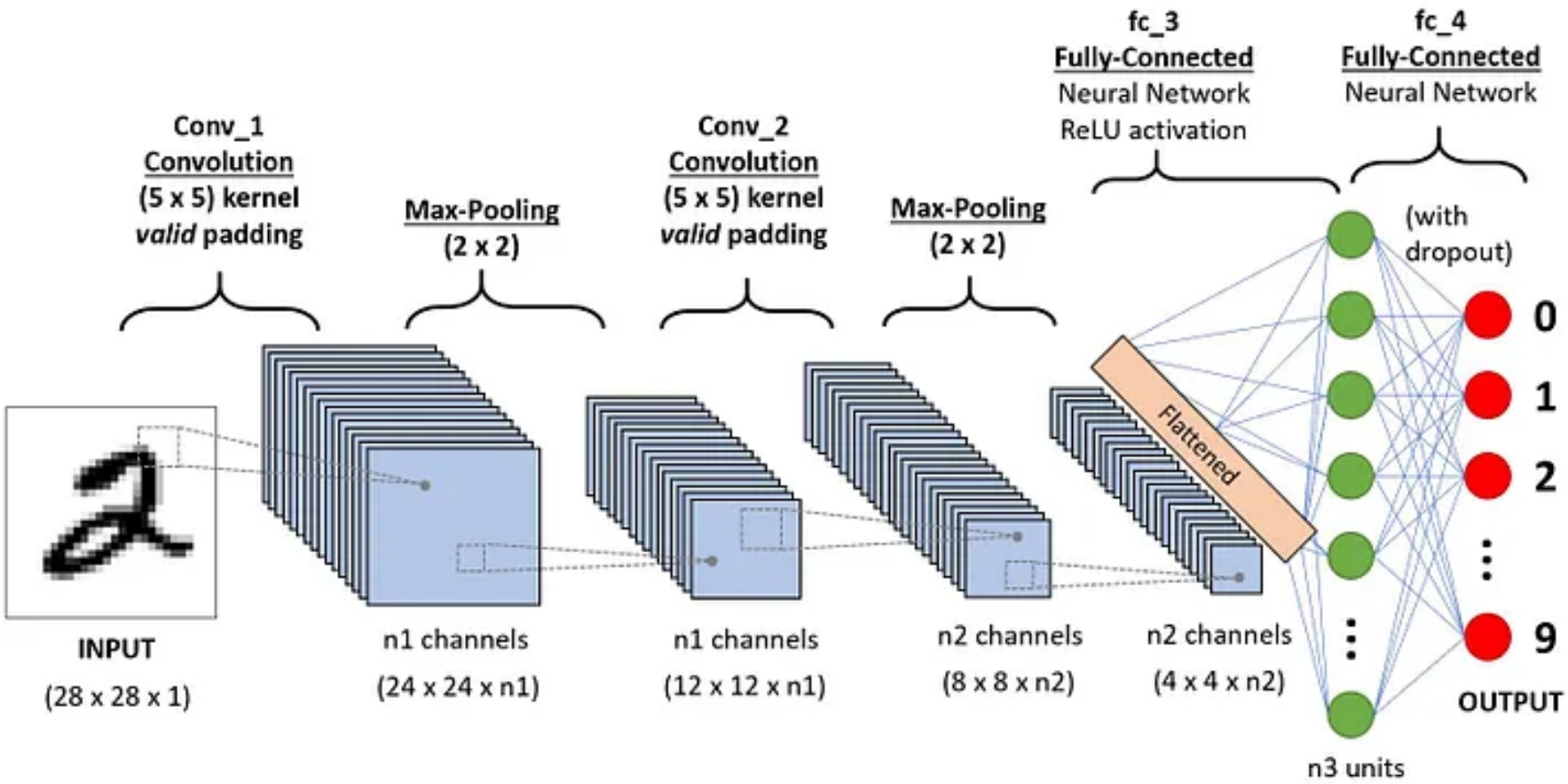
Source: BBC Sky at Night Magazine

- International collaboration (between 2000-2014).
- Dedicated wide-angle optical telescope at Apache Point Observatory in New Mexico.
- Provides both spectroscopic and imaging data across vast regions of the sky.
- Main goals:
  - Studying stars, galaxies, and cosmic structures.
  - Create the map of the universe.
- Result: large database[1].

[1] Also a vast source of cosmic knowledge

Theoretical Background

What are Convolutional Neural Networks (CNNs)?

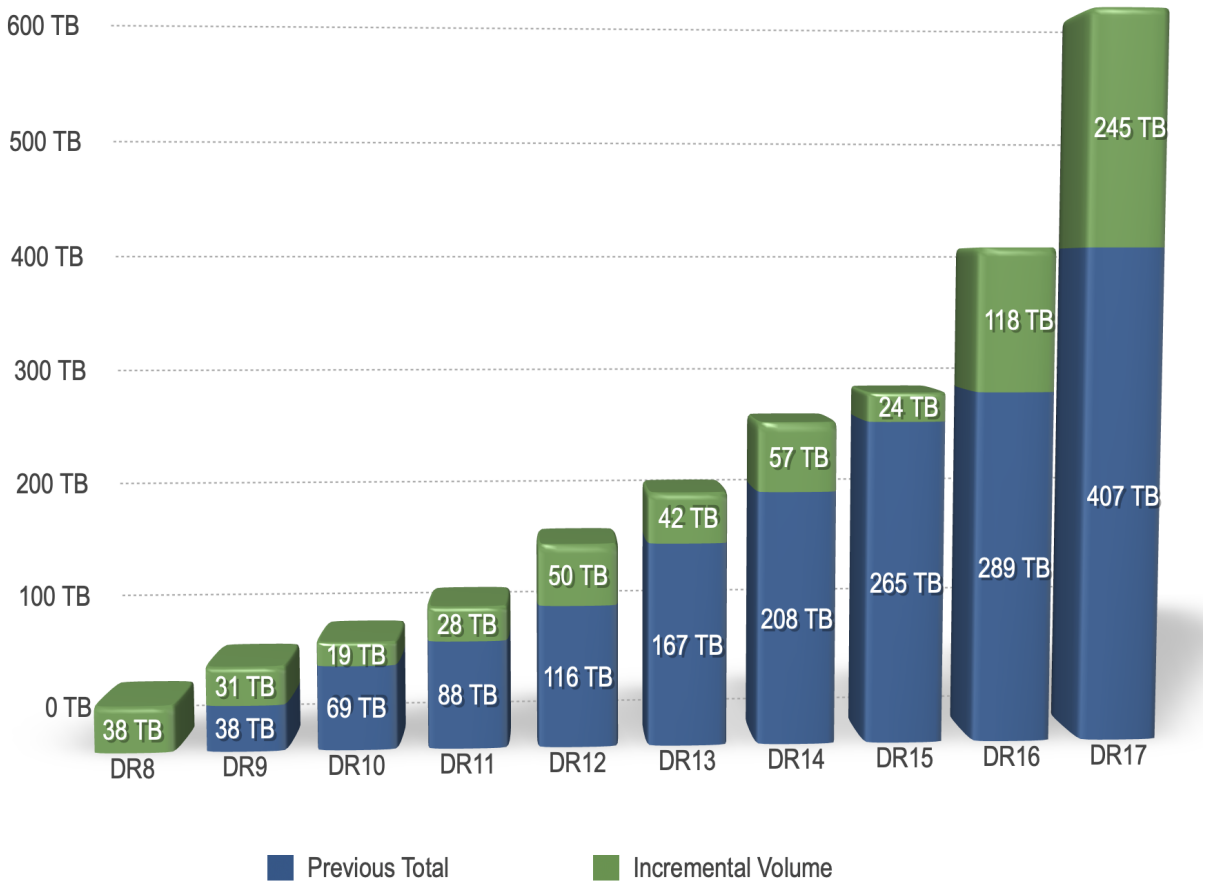


Source: towardsdatascience.com

What was my project?

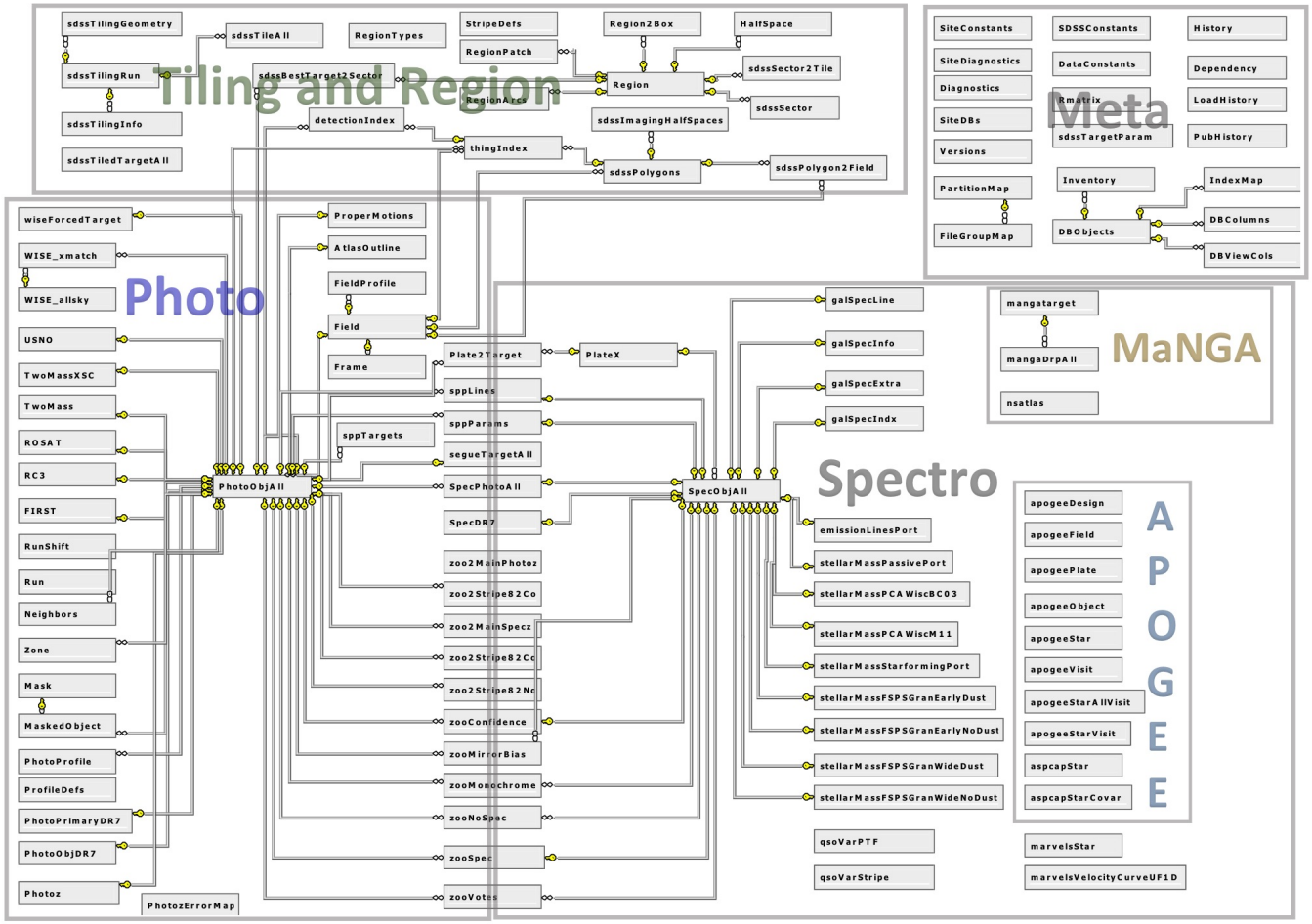
- Downloaded images and corresponding redshift values.
- Created a CNN to predict the redshift parameter of galaxies.

What does the SDSS database look like?



Source: sdss4.org/dr17

What does the SDSS database look like?



Source: skyserver.sdss.org

What does the SDSS database look like?

Using `atroquery.sdss.SDSS` we can connect to the SDSS database.

```
SDSS.query_sql("""
    SELECT Count(*) AS number_of_records_in_SpecObj_table
    FROM SpecObj
""").to_pandas()
```

	number_of_records_in_SpecObj_table
0	5112724

Selected data

To get clear data we can use conditions.

```
number_of_selected_elemenents = SDSS.query_sql("""
    SELECT Count(*) AS number_of_selected_elemenents
    FROM SpecObj
    WHERE class='Galaxy'
      AND specobjid < 92233720368547758O7 -- Only in BIGINT intervall
      AND zOffset = 0
      AND zWarning = 0
      AND zErr > 0
      AND velDispErr > 0
      AND z > zErr*10000
      AND velDisp > velDispErr*10
""").to_pandas().values[0,0]
print(f'Number of selected elemenents: {number_of_selected_elemenents}')
```

```
Number of selected elemenents: 19584
```

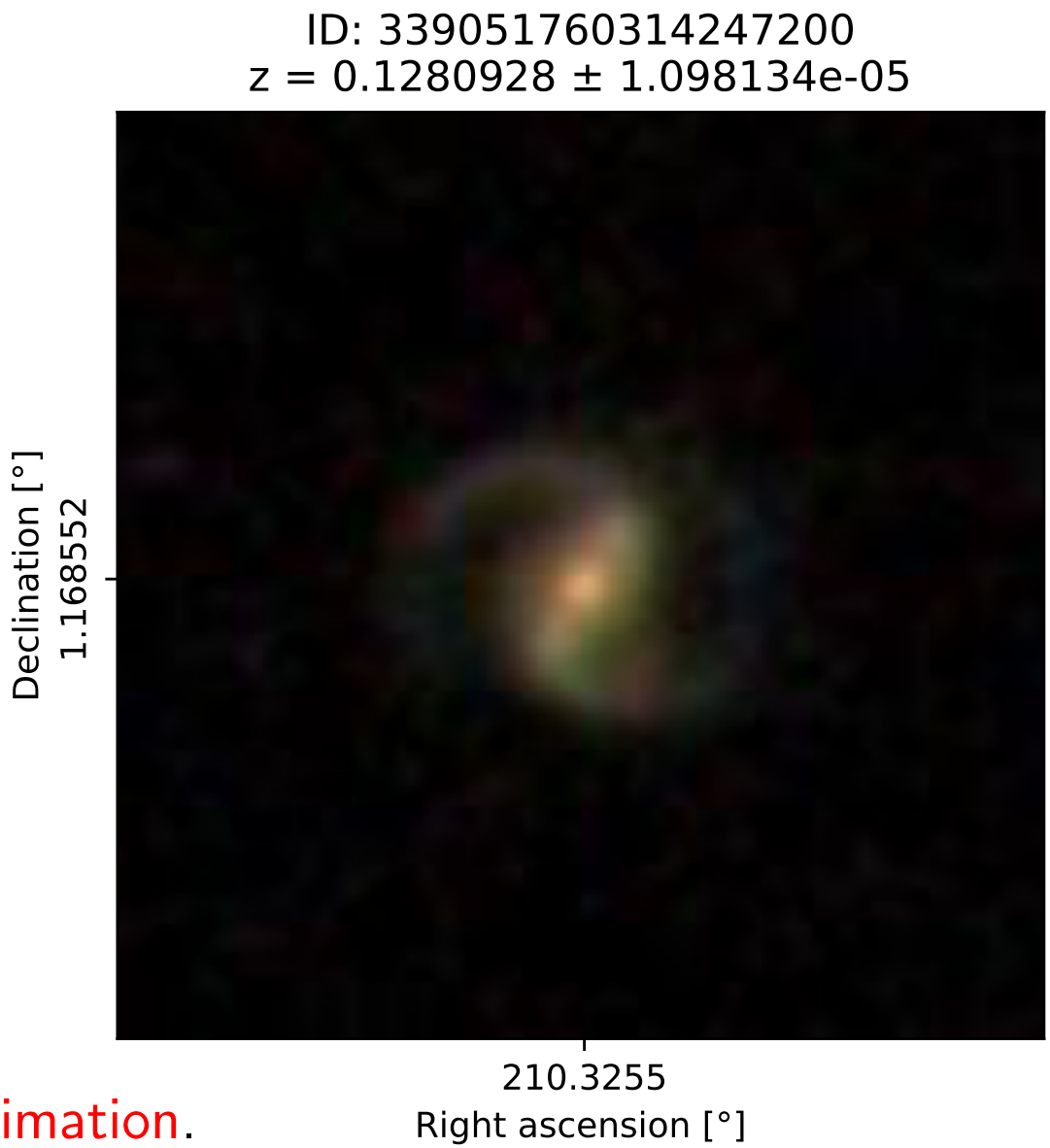
How can we access to the pictures?

HTTP GET request

```
imgCutout = "https://skyserver.sdss.org/dr17/" +\
            "SkyServerWS/ImgCutout/getjpeg?"

def get_sdss_image(ra, dec, scale=IMG_SCALE,
                   size=IMG_SIZE):
    url = imgCutout + \
          f"ra={ra}&dec={dec}&scale={scale}&
            width={size}&height={size}"
    response = requests.get(url)
    if response.status_code is okay:
        return picture
    else:
        return error
```

Example image



Side project: animation.

What does my database looks like?

See codes here.

```
run_query("""
    SELECT pg_size_pretty(pg_table_size('redshift')) AS size_of_redshift_table;
""")
```

	size_of_redshift_table
0	408 MB

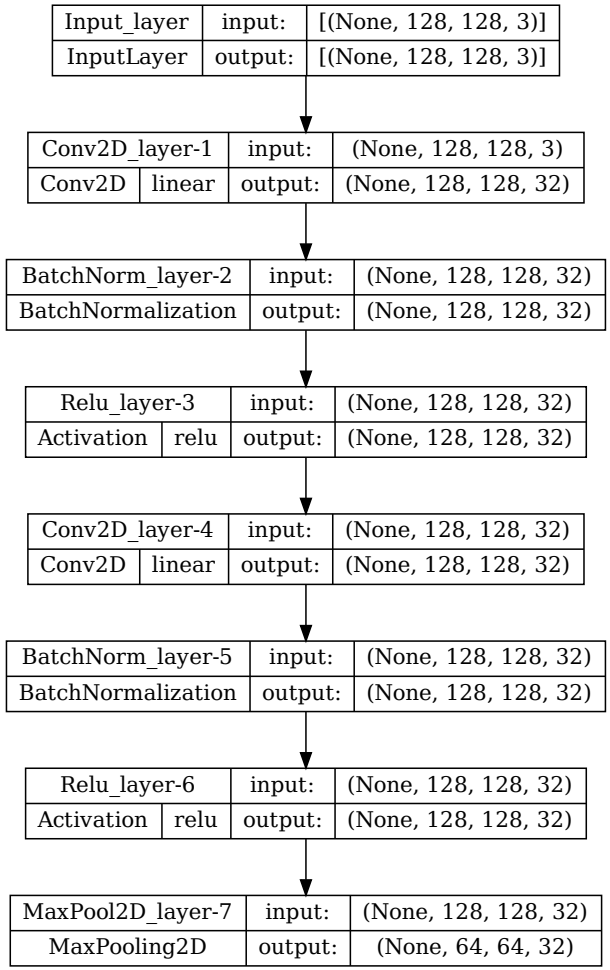
```
run_query("""
    SELECT *
    FROM redshift
    LIMIT 3
""")
```

	index	specobjid	ra	dec	z	zerr	veldisp	veldisperr	picture
0	0	299616671037941760	145.99917	0.352445	0.143653	0.000013	0.161662	0.013259	[7, 5, 6, 6, 4, 5, 3, 1, 2, 1, 0, 0, 1, 0, 0, ...
1	1	305153815840057340	154.33863	-1.246887	0.094062	0.000009	0.168785	0.013659	[12, 12, 10, 12, 12, 10, 11, 11, 9, 8, 8, 6, 3...
2	2	305167834613311500	154.47804	-0.469954	0.181684	0.000011	0.126630	0.010545	[7, 0, 0, 7, 0, 0, 7, 0, 0, 8, 0, 0, 10, 1, 2,...

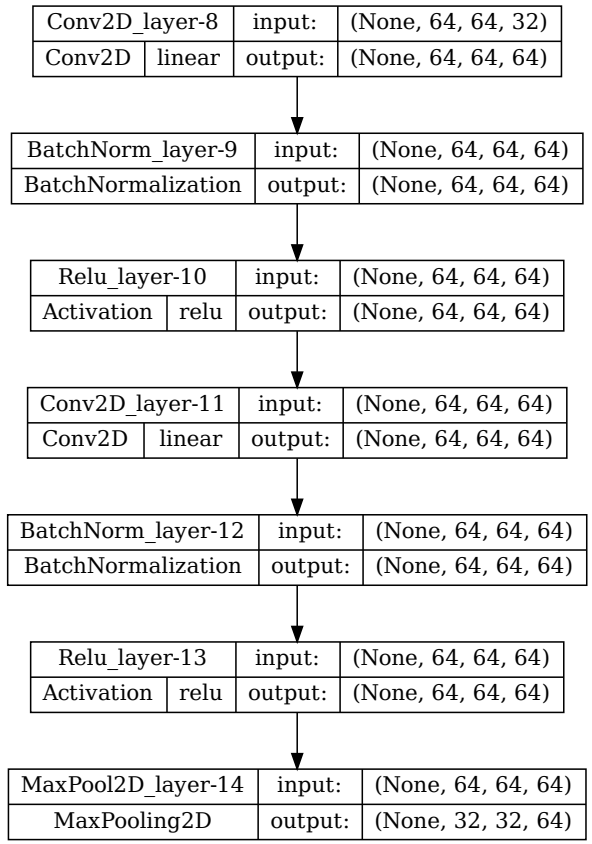
I downloaded $10,000$ images and its parameters.

What does my CNN look like?

See codes here.



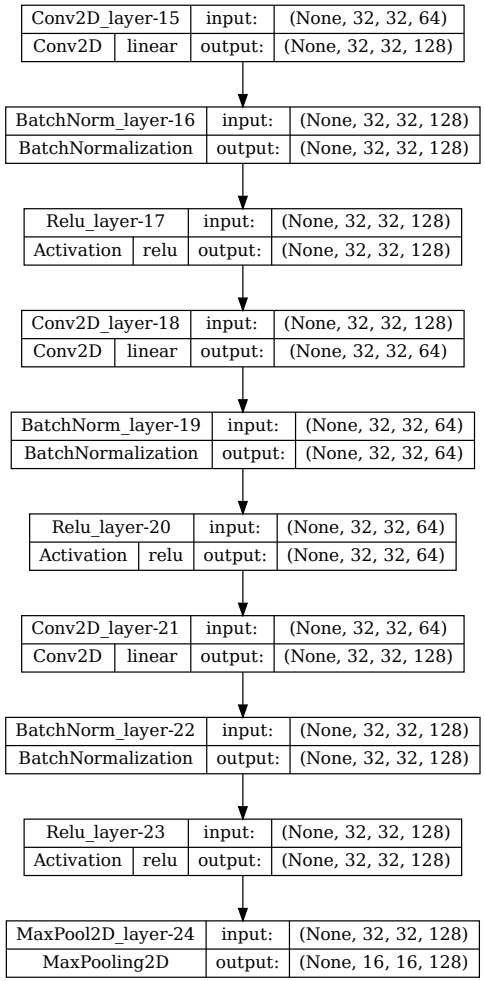
Block #1



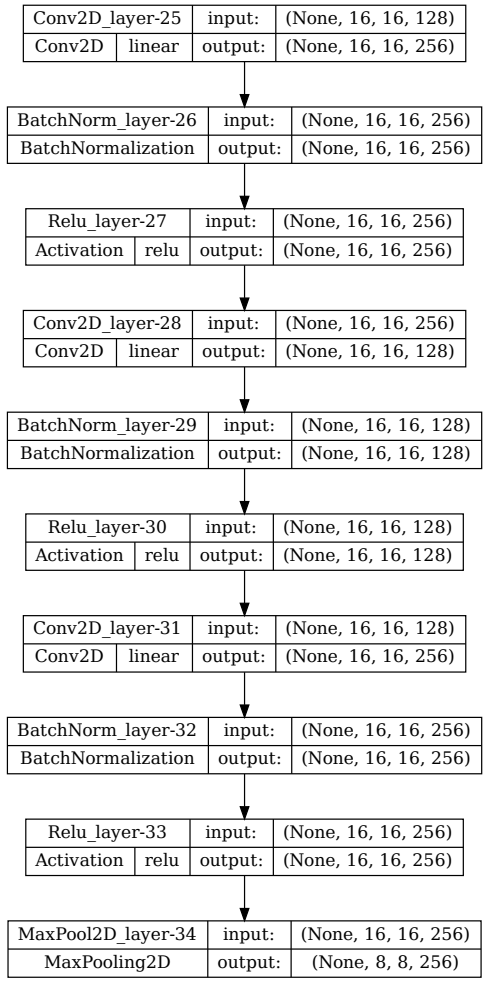
Block #2

What does my CNN look like?

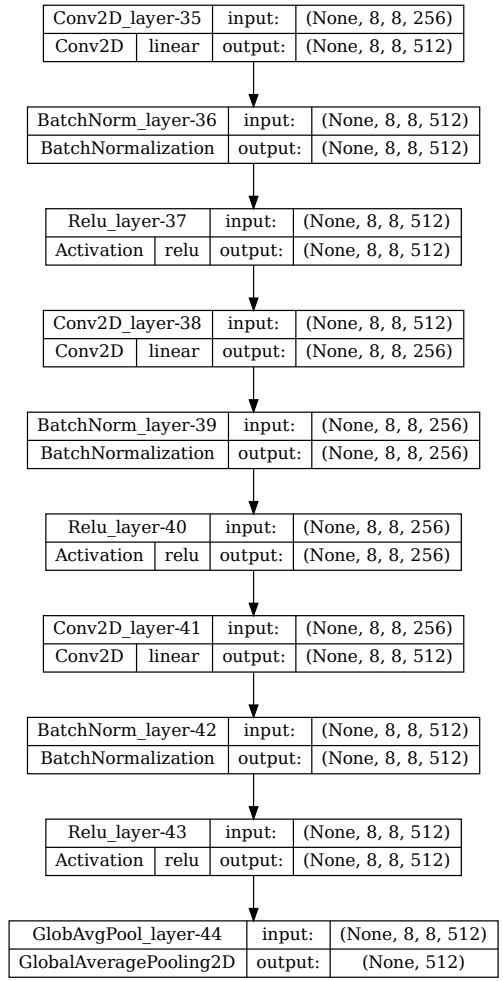
See codes here.



Block #3



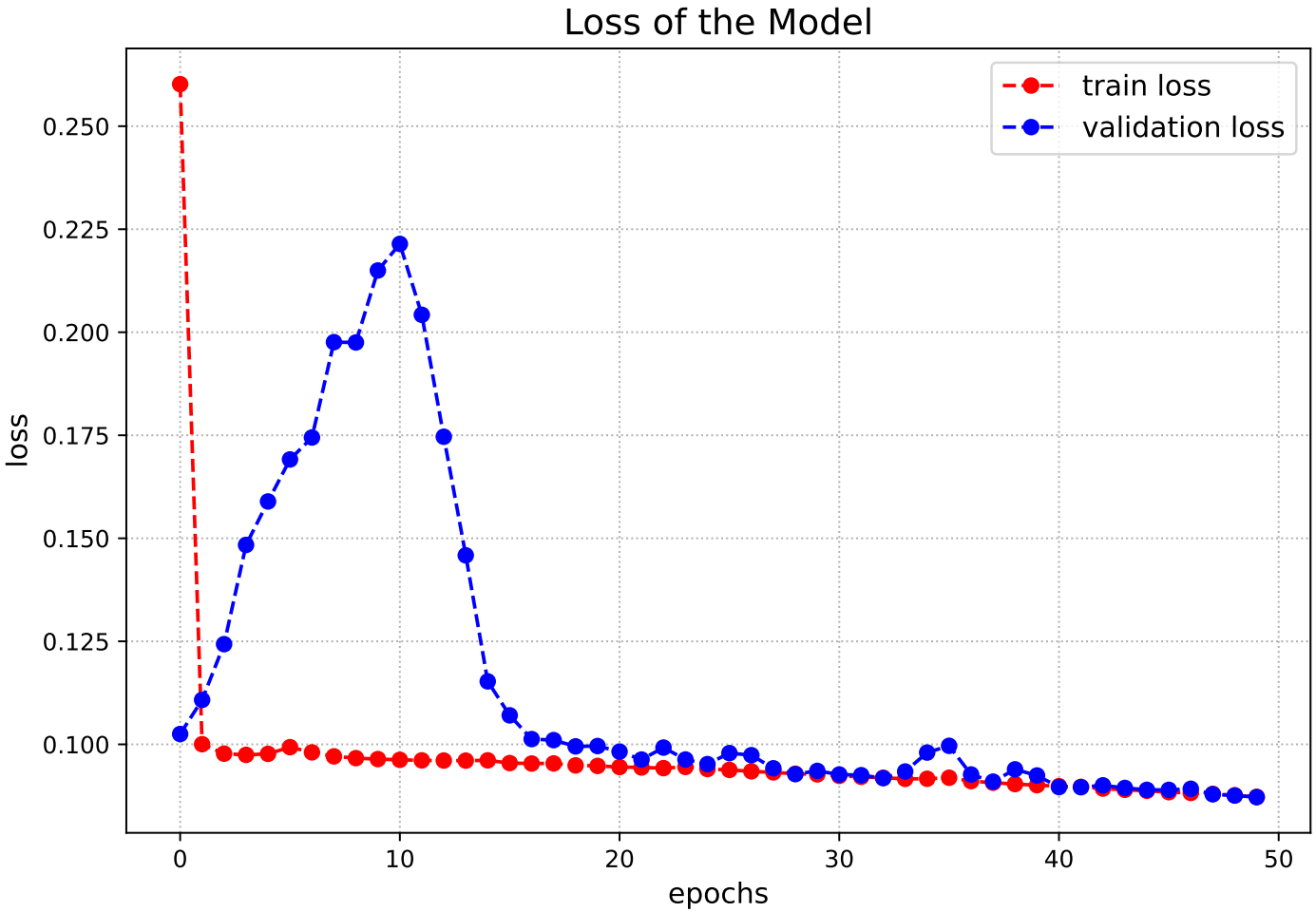
Block #4



Block #5

How accurate is my CNN?

$N_{\text{pictures}} = 8000$, epochs$= 50$, batch size$= 128$



Mean Squared Error as loss function

How accurate is my CNN?

$N_{\text{pictures}} = 8000$, epochs$= 50$, batch size$= 128$

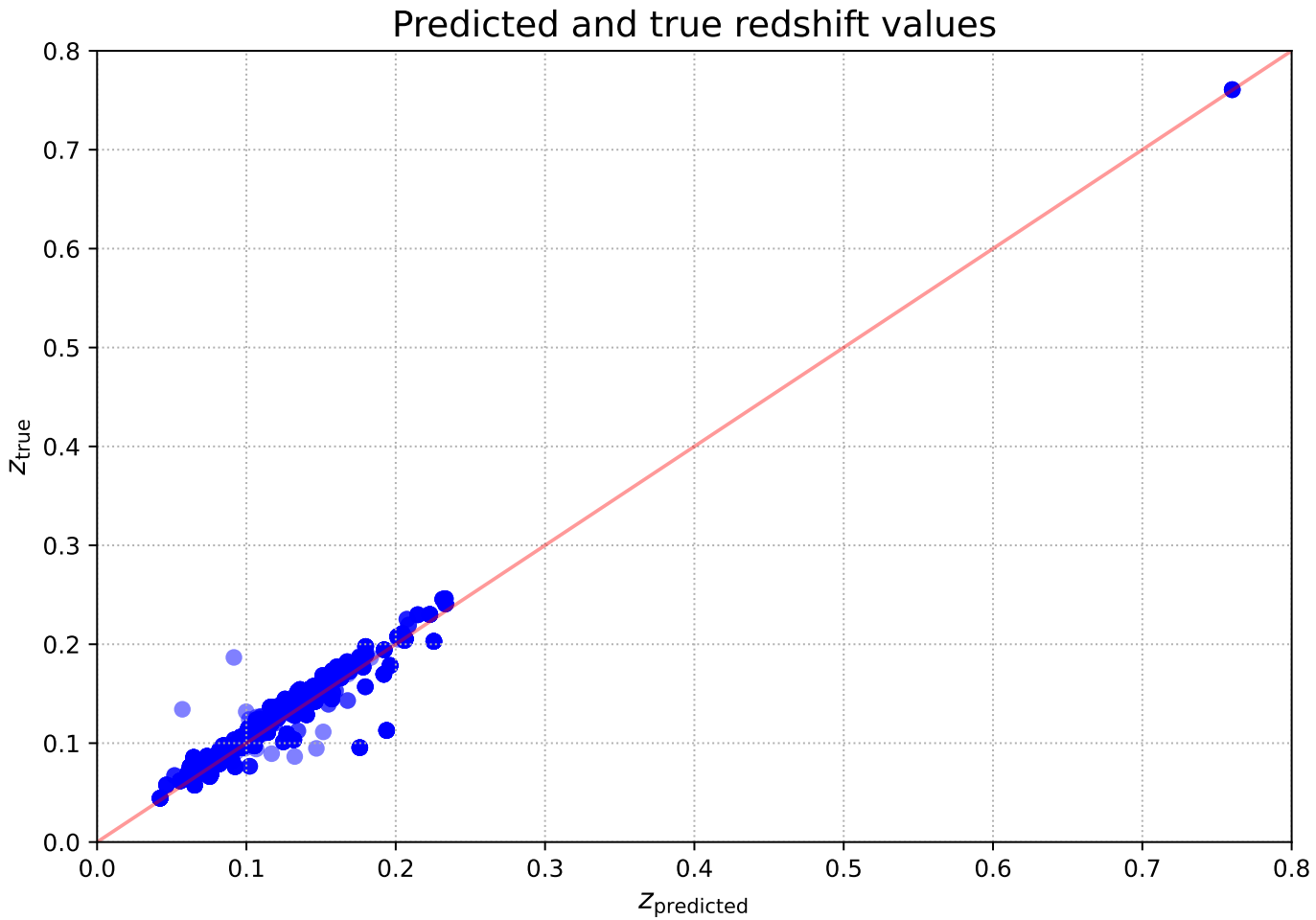
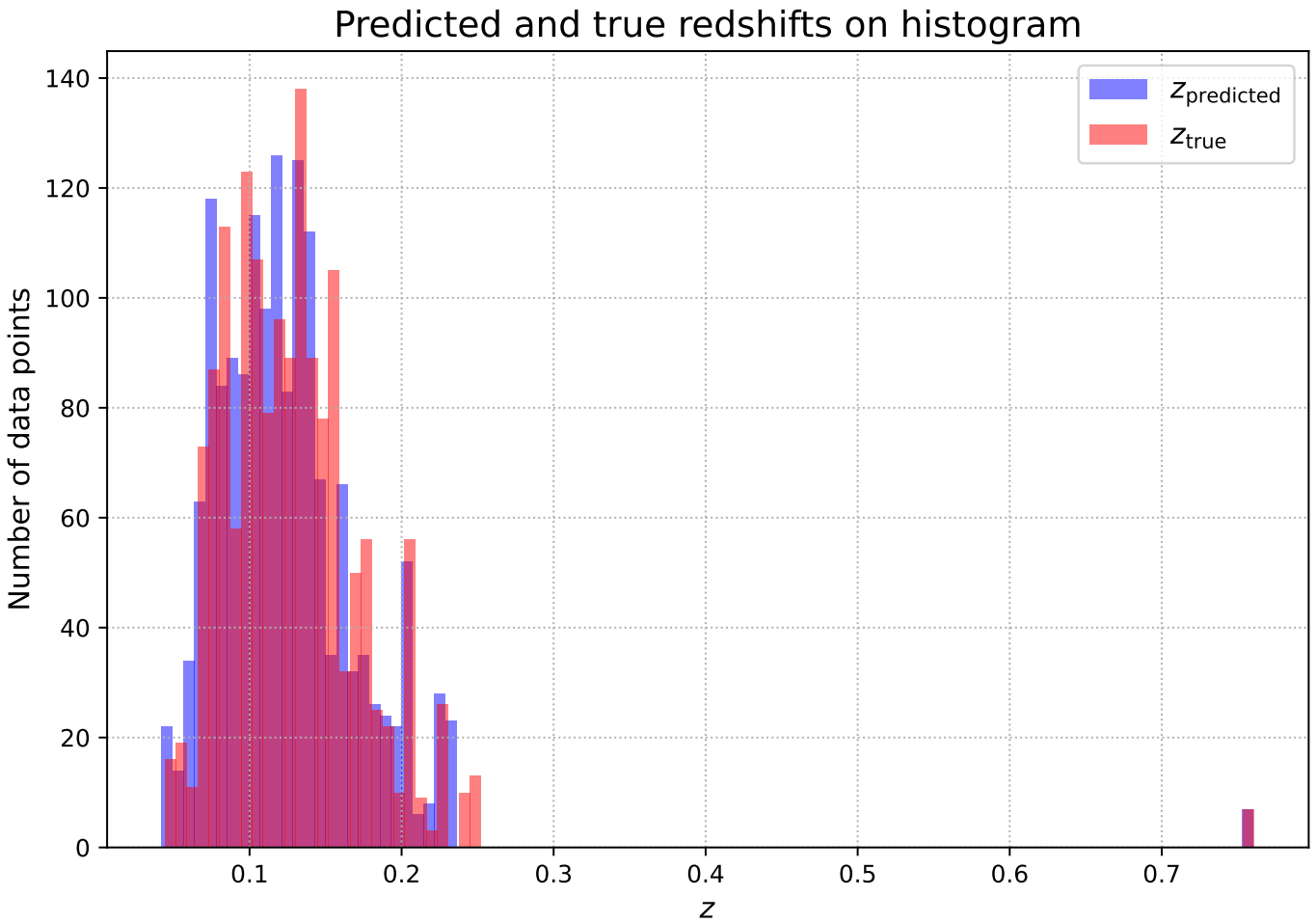


Figure: $R^2 = 94.126\%$

How accurate is my CNN?

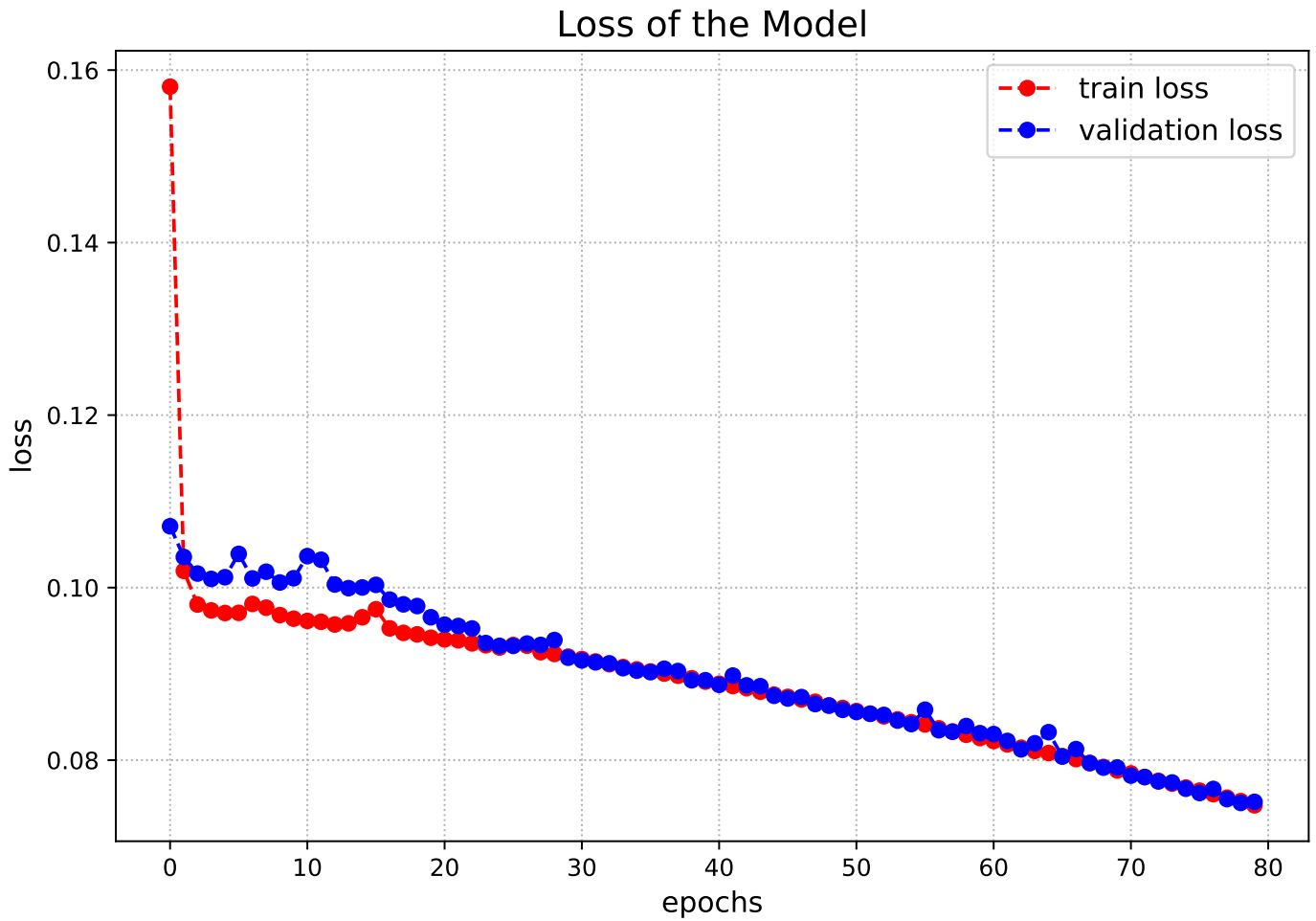
$N_{\text{pictures}} = 8000$, epochs$= 50$, batch size$= 128$



Run time: 4 h 10 min

How accurate is my CNN? - Second run

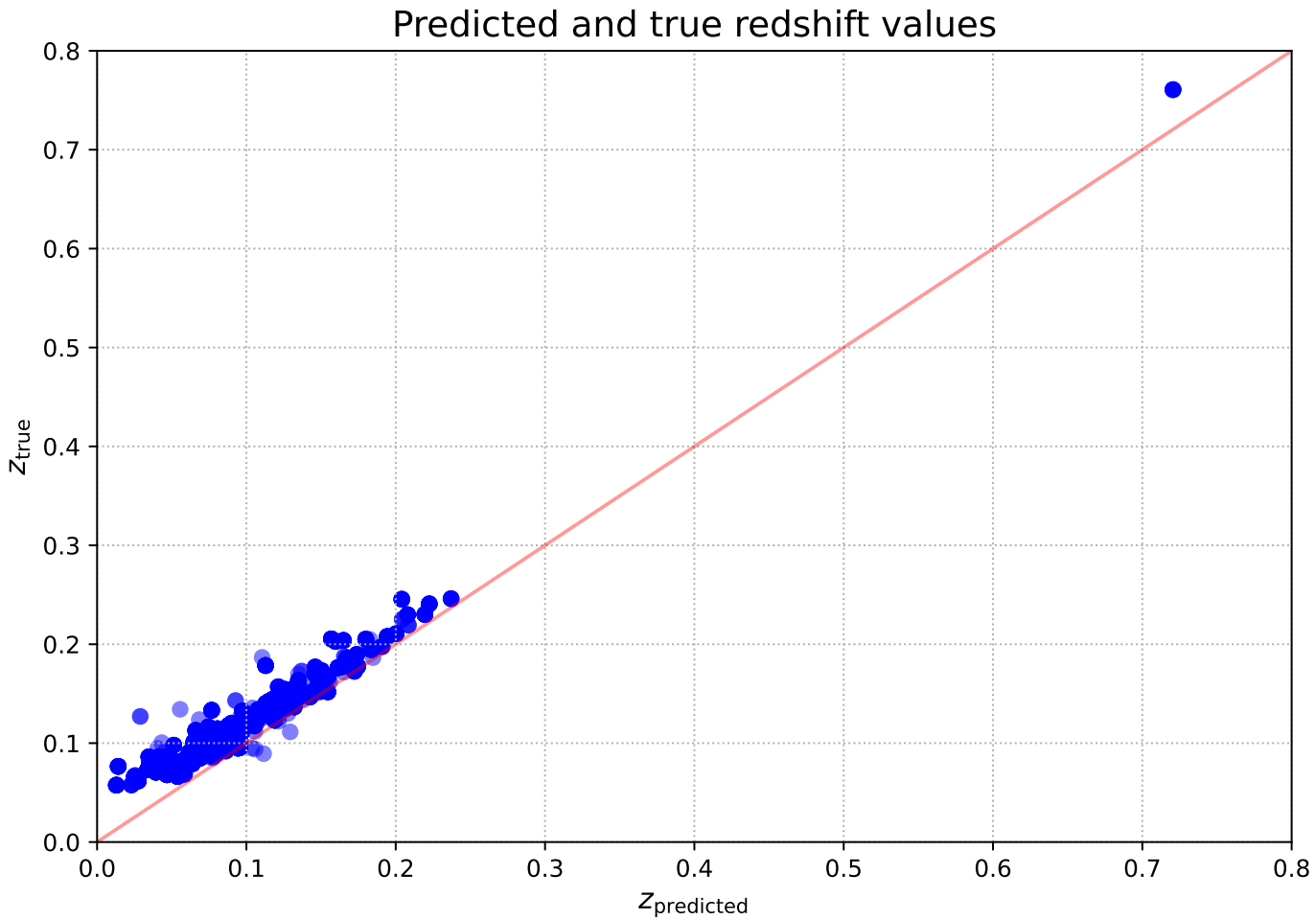
$N_{\text{pictures}} = 8000$, epochs$= 80$, batch size$= 128$



Mean Squared Error as loss function

How accurate is my CNN? - Second run

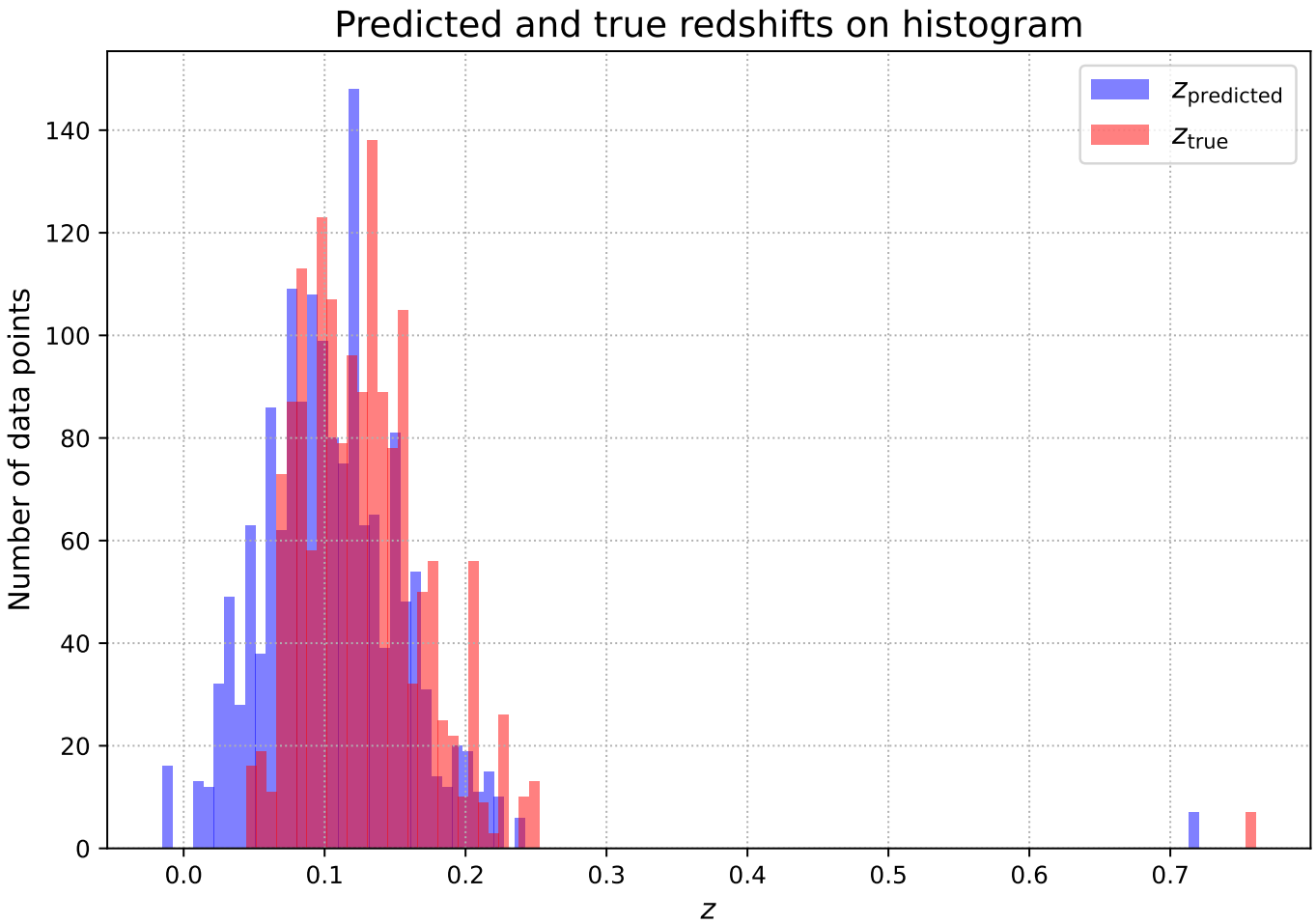
$N_{\text{pictures}} = 8000$, epochs$= 80$, batch size$= 128$



$$R^2 = 79.55\%$$

How accurate is my CNN? - Second run

$N_{\text{pictures}} = 8000$, epochs$= 80$, batch size$= 128$



Run time: 5 h 10 min

Conclusions

- Using SQL queries makes life easier.
- CNNs are slower without GPUs.
- The redshift value of galaxies can be *calculated* from images.