Supernova Light Curves Prediction Based on Transformer

Transformer pioneer paper: “*Attention is all you need*” (Vaswani et al. 2017)

Reporter: 金奕澄

June 27, 2025

Catalogue

- 1. Introduction of the Transformer model
- 2. Supernova light curves prediction
  - Data sources
  - Method
  - Results

1. Transformer

Structure and Main Modules

- Structure:
  - Transformer Encoder (history analyzer)
  - Transformer Decoder (future predictor)
  - Output Layer
- Main modules:
  - Continuous value linear transformation (or word embedding)
  - Position Embedding
  - Self-attention calculation (obtain contextual/relevant information)
  - Position-wise Feed-Forward Network (Adding Non-linearity)
  - Residual Connection and Normalization

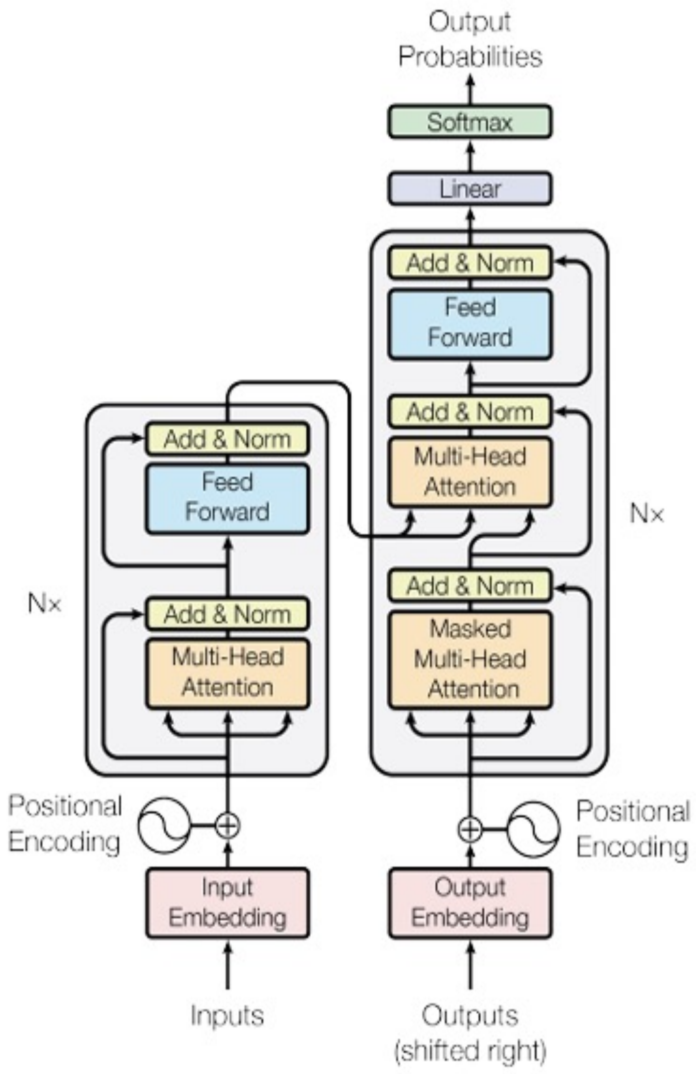
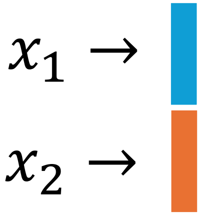


Figure 1: The Transformer - model architecture.

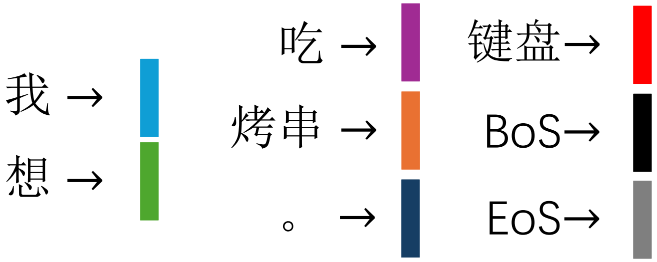
(Vaswani et al. 2017)

Continuous value linear transformation or Word Embedding

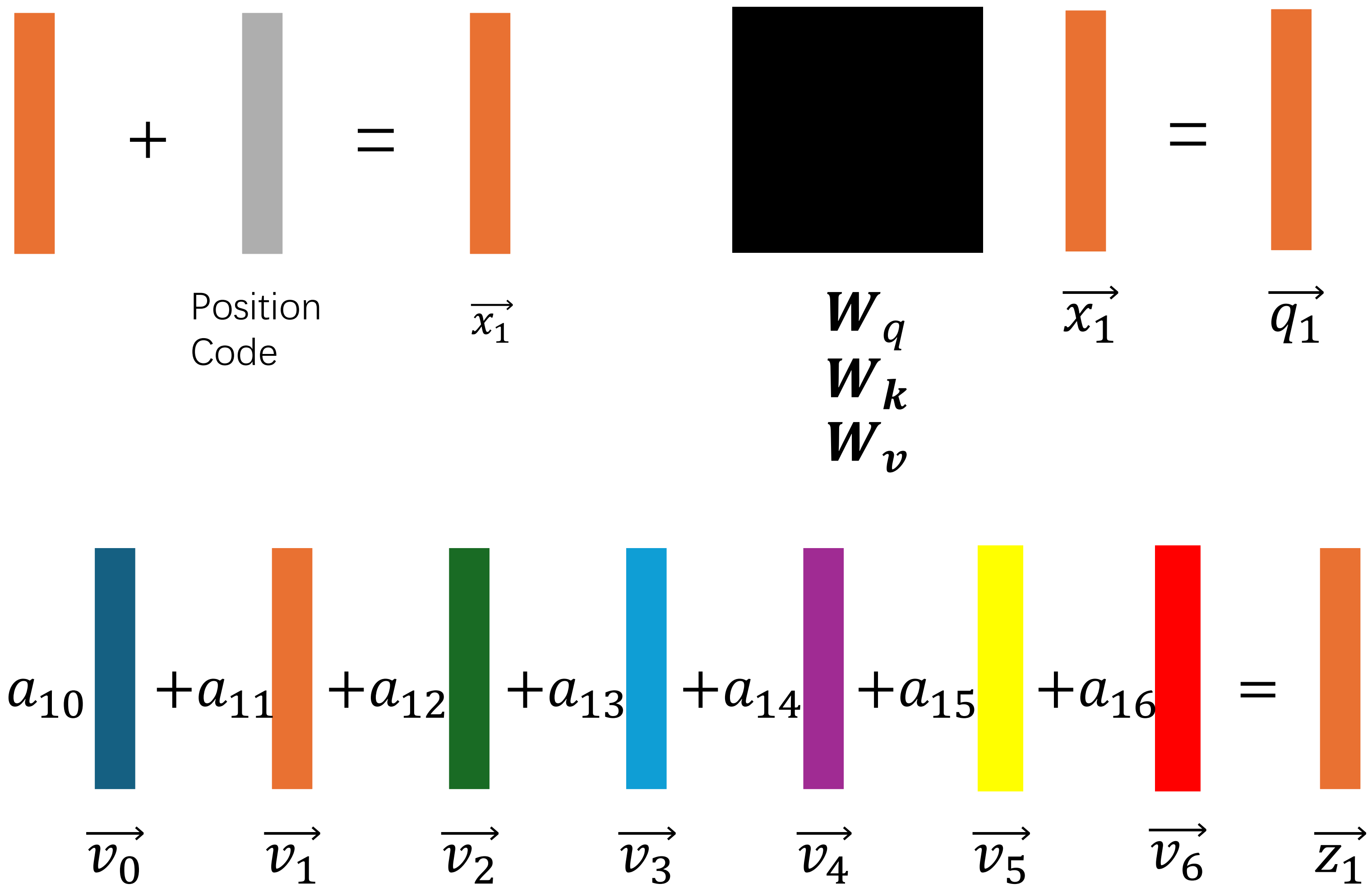
- Numerical models (e.g., SN light curves predictor): Continuous value linear transformation
  - Establish a linear transformation from data points to high-dimensional vectors (understandable to Transformer) (e.g., 128-D)



- Linguistic sequence (e.g., Chat-GPT): Word Embedding
  - Establish a dictionary (mapping from tokens (words) to high-dimensional vectors



Multi-head Self-attention Calculation



- $a_{1i} = \text{Softmax}(\frac{\overrightarrow{q_1} \cdot \overrightarrow{k_i}}{\sqrt{d}})$: Relevance between data point 1 and n
- $\overrightarrow{z_1} = a_{1i}\overrightarrow{v_i}$ (By incorporating contextual information, the vector of point 1 is updated from ($\overrightarrow{x_1}$) to ($\overrightarrow{z_1}$))

Other Optimizing Modules

- Position-wise feed-forward network (Adding Non-linearity)
- Residual connection (Preserving original information and preventing exponential explode)
- Layer normalization (Stabilizing training and accelerating convergence)

Output Layer: Linear

- Numerical Transformer (e.g., SN light curves predictor)：
  - Linear transformation from high-dimensional vectors to data points that human can understand.

$$x_1, x_2 \rightarrow \blacksquare \rightarrow \widehat{x_3}$$

- Linguistic Transformer (e.g., Chat-GPT)
  - Linear transformation from high-dimensional vectors to word scores vector (number of dimensions is the number of words in the dictionary)

$$\text{我想吃} \rightarrow \blacksquare \rightarrow \begin{bmatrix} -569, \text{我} \\ 20, \text{键盘} \\ -233, \text{想} \\ 1000, \text{烤串} \\ -63, \text{吃} \end{bmatrix}$$

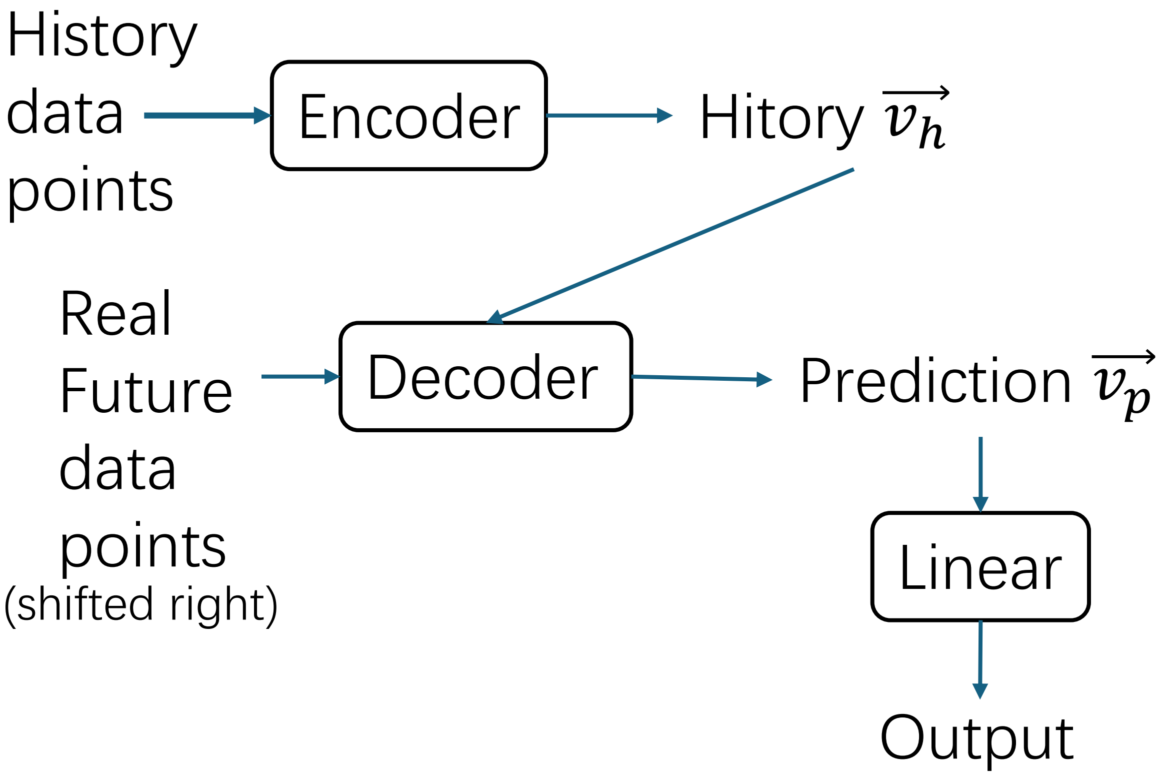
Output Layer: Softmax (only for linguistic models)

- $\mathrm{Softmax}(x_i) = \dfrac{e^{x_i - \max(x)}}{\sum_{j=0}^{N-1} e^{x_j - \max(x)}}$

$$\text{我想吃} \rightarrow \; \rightarrow \begin{bmatrix} -569, \text{我} \\ 20, \text{键盘} \\ -233, \text{想} \\ 1000, \text{烤串} \\ -63, \text{吃} \end{bmatrix} \xrightarrow{Softmax} \begin{bmatrix} 3.090\times10^{-682}, \text{我} \\ 2.505\times10^{-426}, \text{键盘} \\ 3.012\times10^{-536}, \text{想} \\ 0.999\ldots, \text{烤串} \\ 2.195\times10^{-462}, \text{吃} \end{bmatrix}$$

From scores to the probability distribution of the next word

Data flow



(Vaswani et al. 2017)

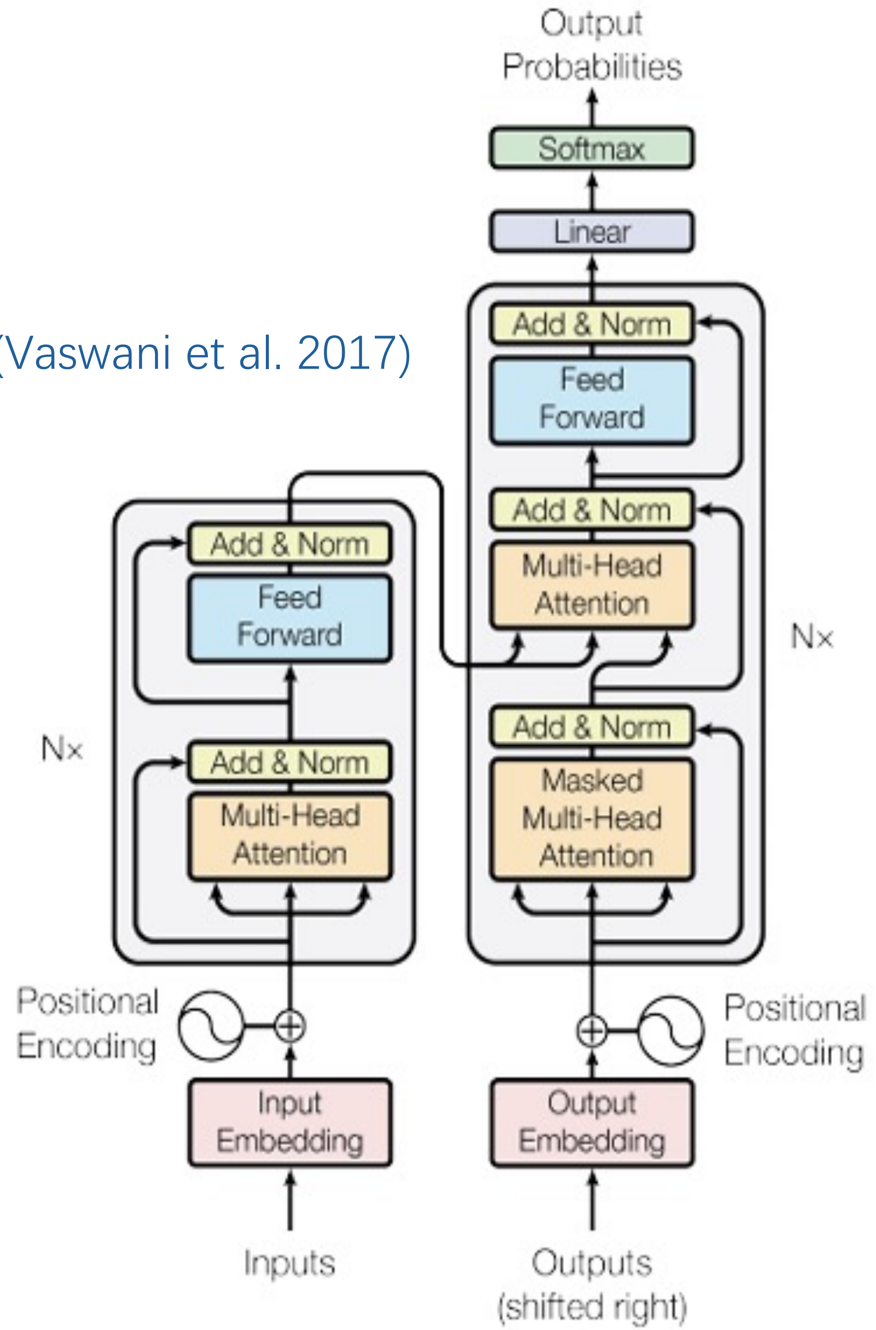


Figure 1: The Transformer - model architecture.

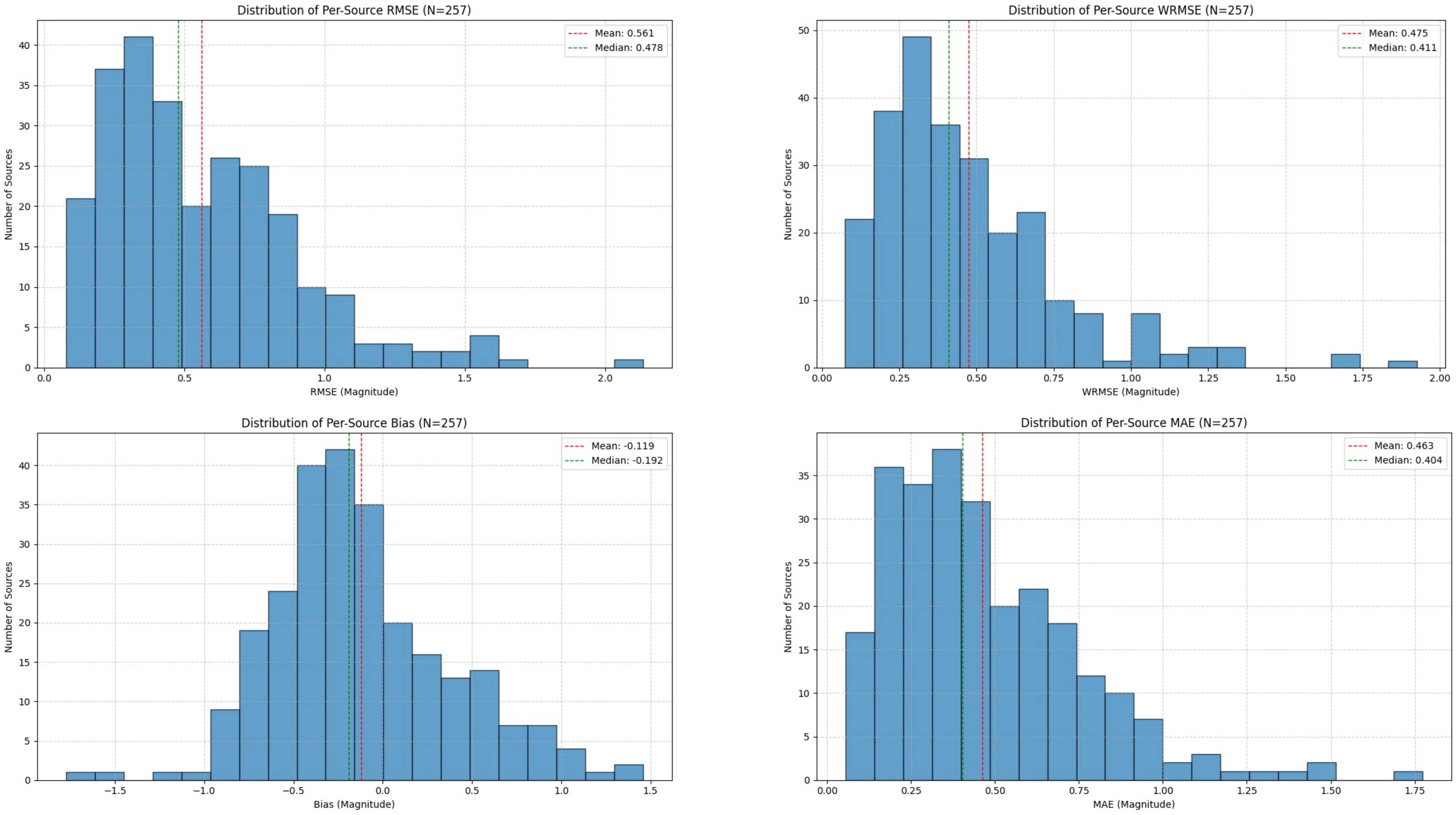
Supernova light curves prediction

Data source

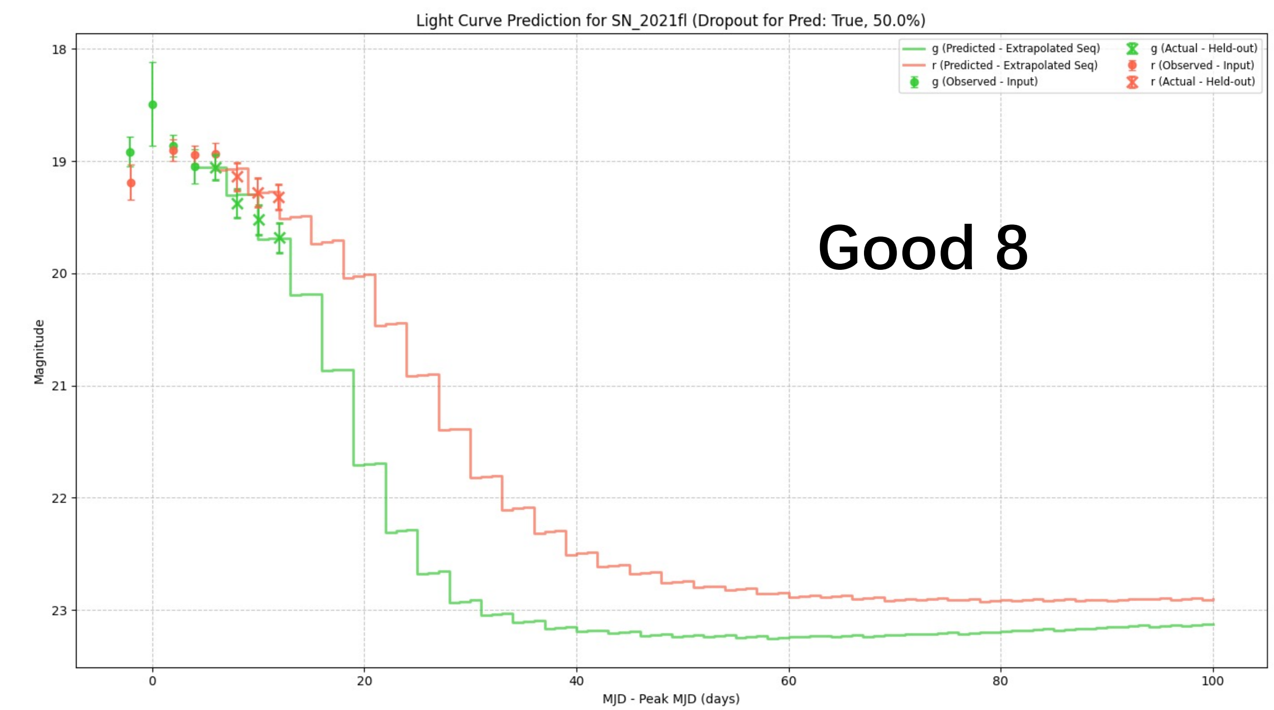
- **ZTF Bright Transient Survey** (https://doi.org/10.3847/1538-4357/abbd98) (https://lasair-ztf.lsst.ac.uk/) (Good)
- **Light Curves of Pan-STARRS1 SN-like Transients** (https://doi.org/10.5281/zenodo.3974949) (Not good)
- **Young Supernova Experiment Data Release 1** (https://doi.org/10.3847/1538-4365/acbfba) (Aleo et al. 2023)
- **Future: Mephisto** (https://doi.org/10.1117/12.2562334)

Method

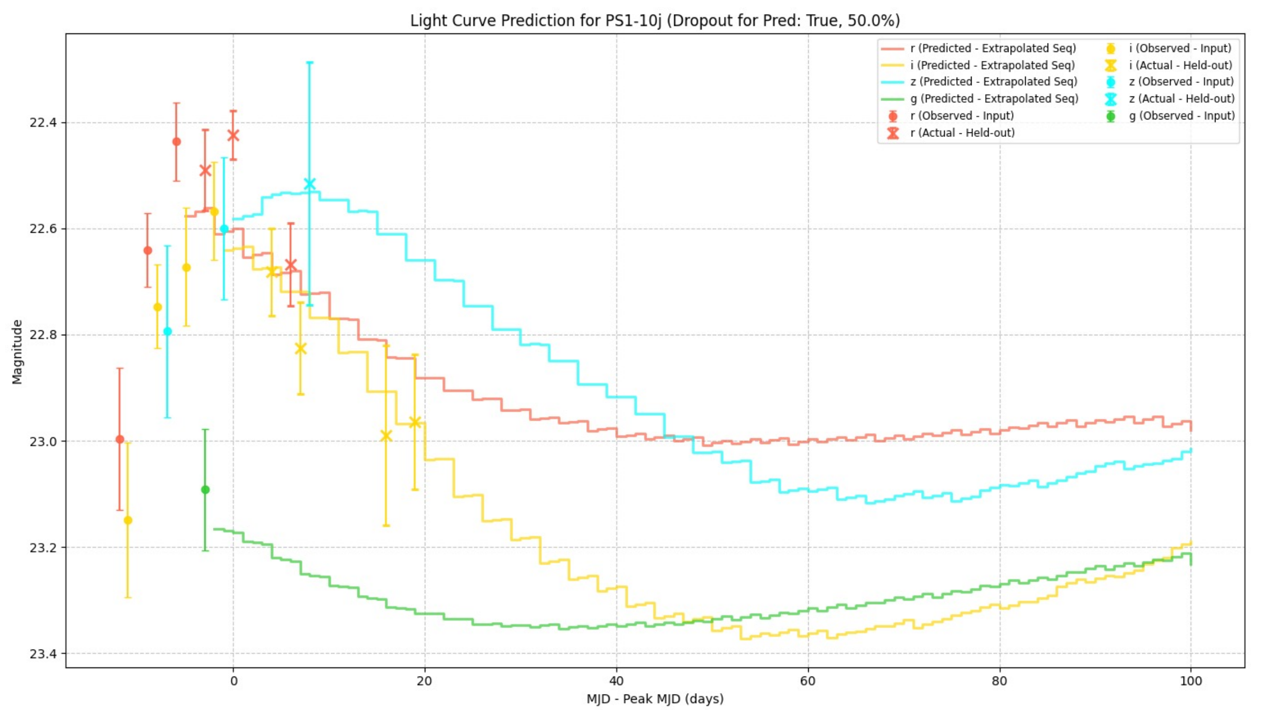
- $MJD-MJD_{Peak}$, Filter (g or r), Mag, Magerr,
- Our model makes predictions independently for each band. For any given supernova, the data from each band is treated as a separate sequence.
- Prediction is performed using a sliding window of 11 data points, where the model learns to predict the subsequent 3 points based on the preceding 8.
- We hold out 20% of the supernovae as a validation set. On the remaining 80% for training, we measure the model's error by masking the final 50% of the data points in two separate configurations.



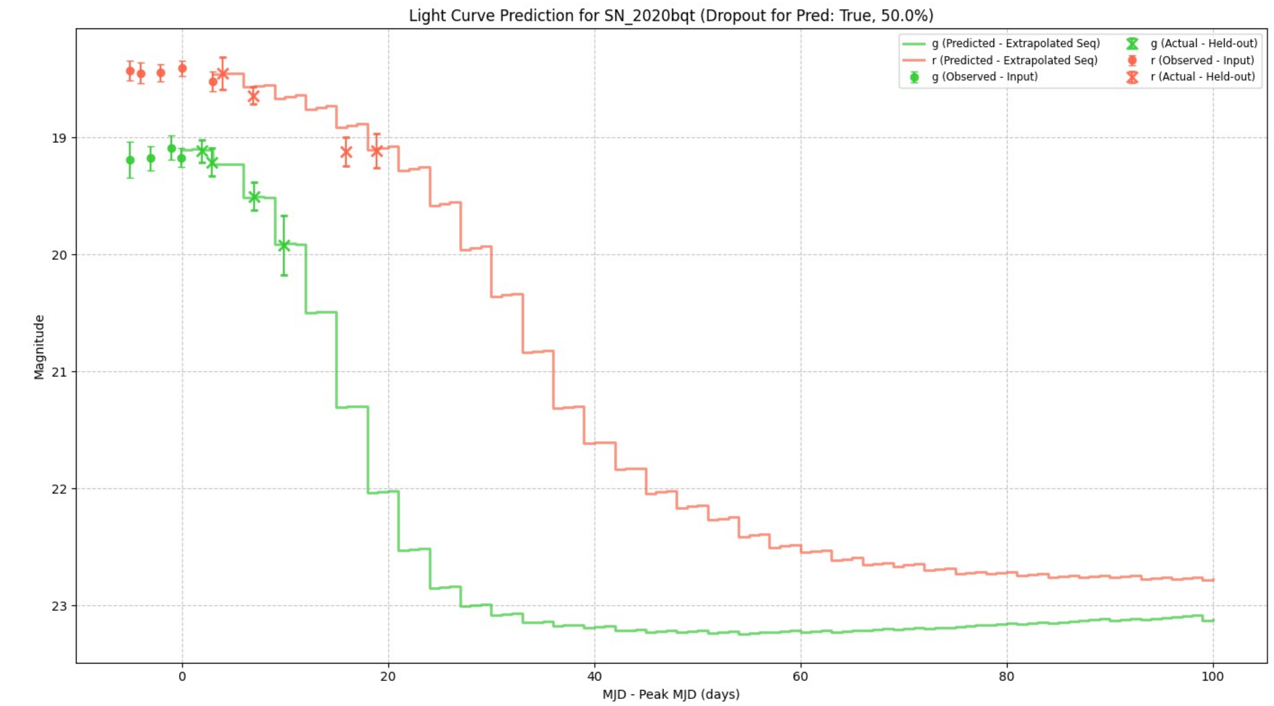
Distribution of Per-Source RMSE (N=257)
Mean: 0.561
Median: 0.478
Number of Sources
RMSE (Magnitude)
Distribution of Per-Source WRMSE (N=257)
Mean: 0.475
Median: 0.411
WRMSE (Magnitude)
Distribution of Per-Source Bias (N=257)
Mean: -0.119
Median: -0.192
Bias (Magnitude)
Distribution of Per-Source MAE (N=257)
Mean: 0.463
Median: 0.404
MAE (Magnitude)



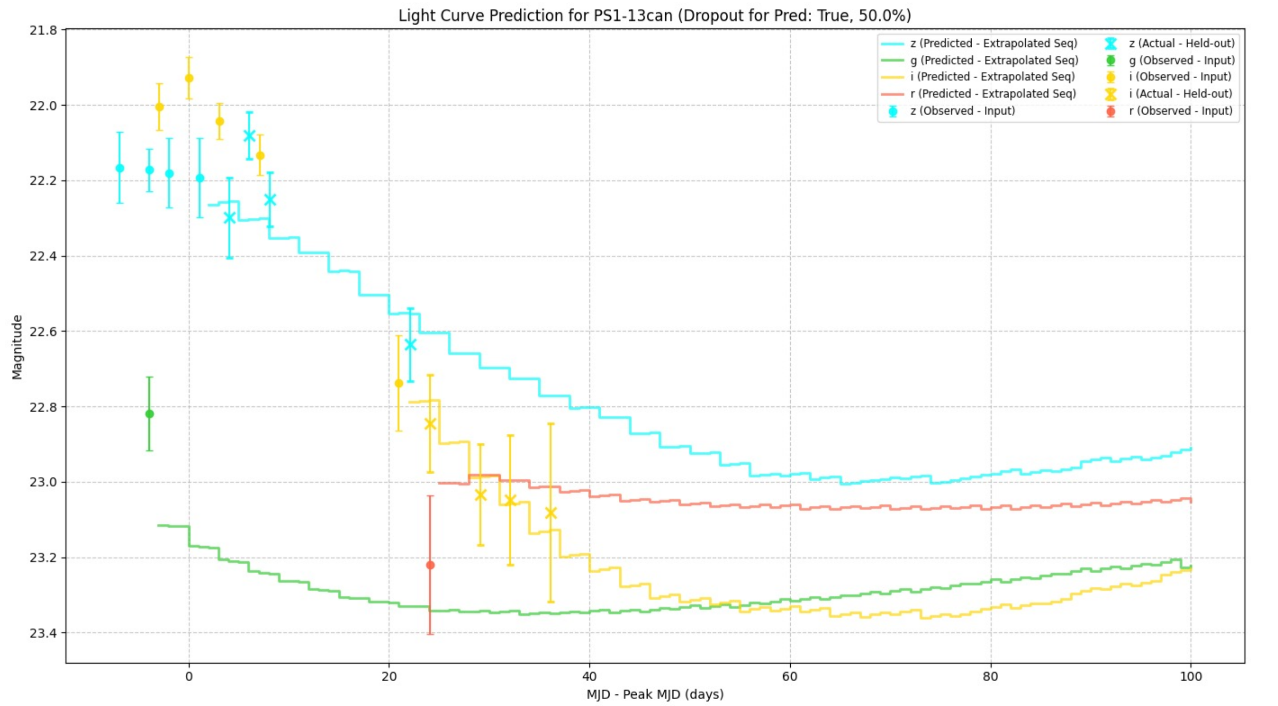
Light Curve Prediction for SN_2021fl (Dropout for Pred: True, 50.0%)
g (Predicted - Extrapolated Seq)
r (Predicted - Extrapolated Seq)
g (Observed - Input)
g (Actual - Held-out)
r (Observed - Input)
r (Actual - Held-out)
Good 8
Magnitude
MJD - Peak MJD (days)
18
19
20
21
22
23
0
20
40
60
80
100



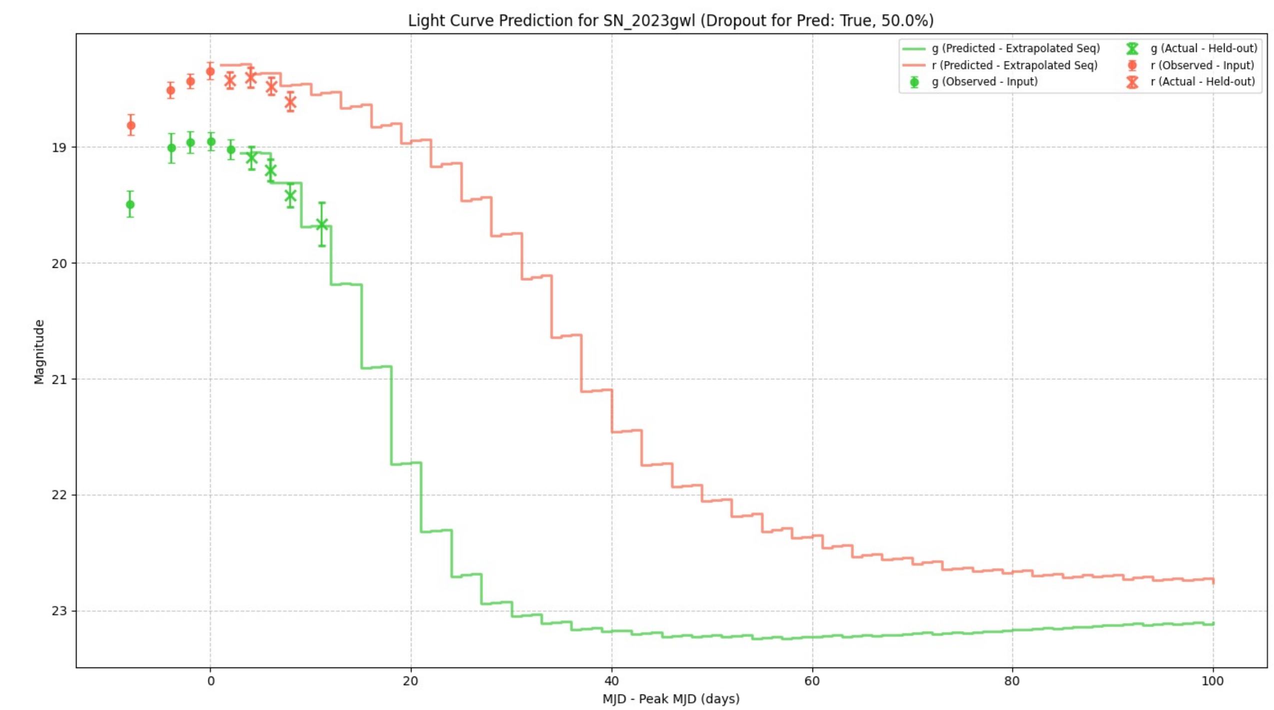
Light Curve Prediction for PS1-10j (Dropout for Pred: True, 50.0%)
r (Predicted - Extrapolated Seq)
i (Predicted - Extrapolated Seq)
z (Predicted - Extrapolated Seq)
g (Predicted - Extrapolated Seq)
r (Observed - Input)
r (Actual - Held-out)
i (Observed - Input)
i (Actual - Held-out)
z (Observed - Input)
z (Actual - Held-out)
g (Observed - Input)
Magnitude
MJD - Peak MJD (days)
22.4
22.6
22.8
23.0
23.2
23.4
0
20
40
60
80
100

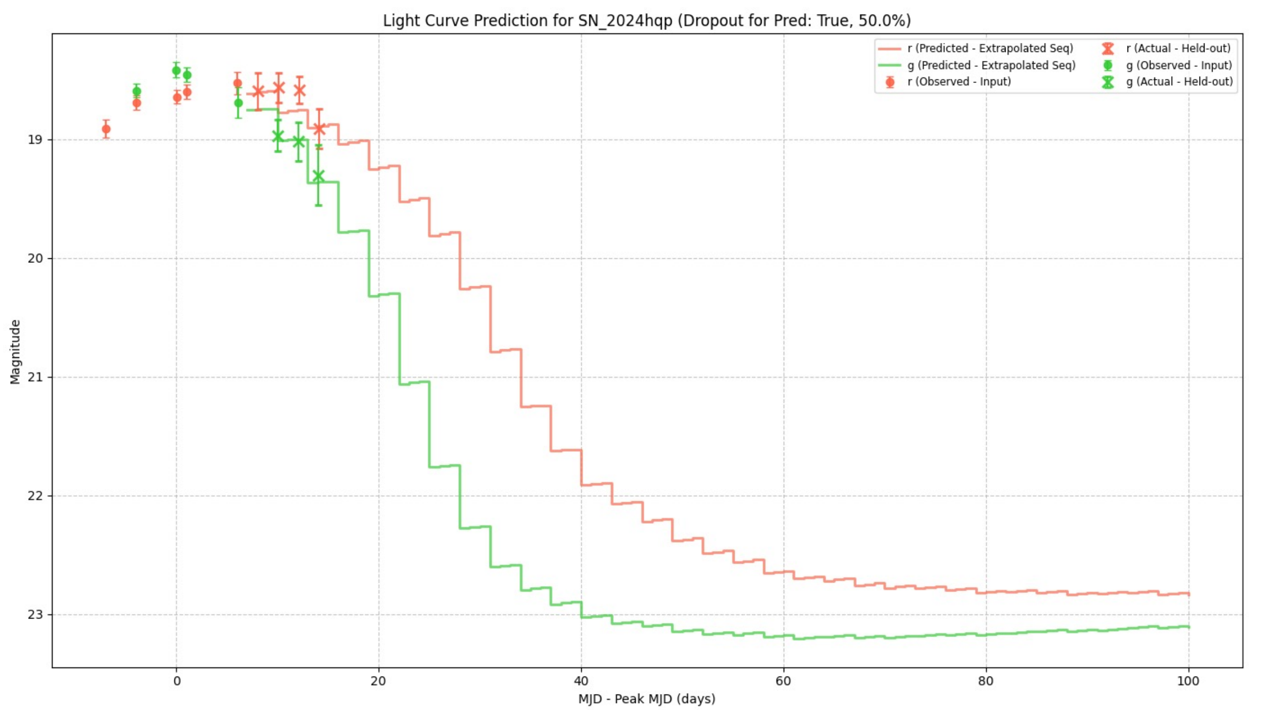


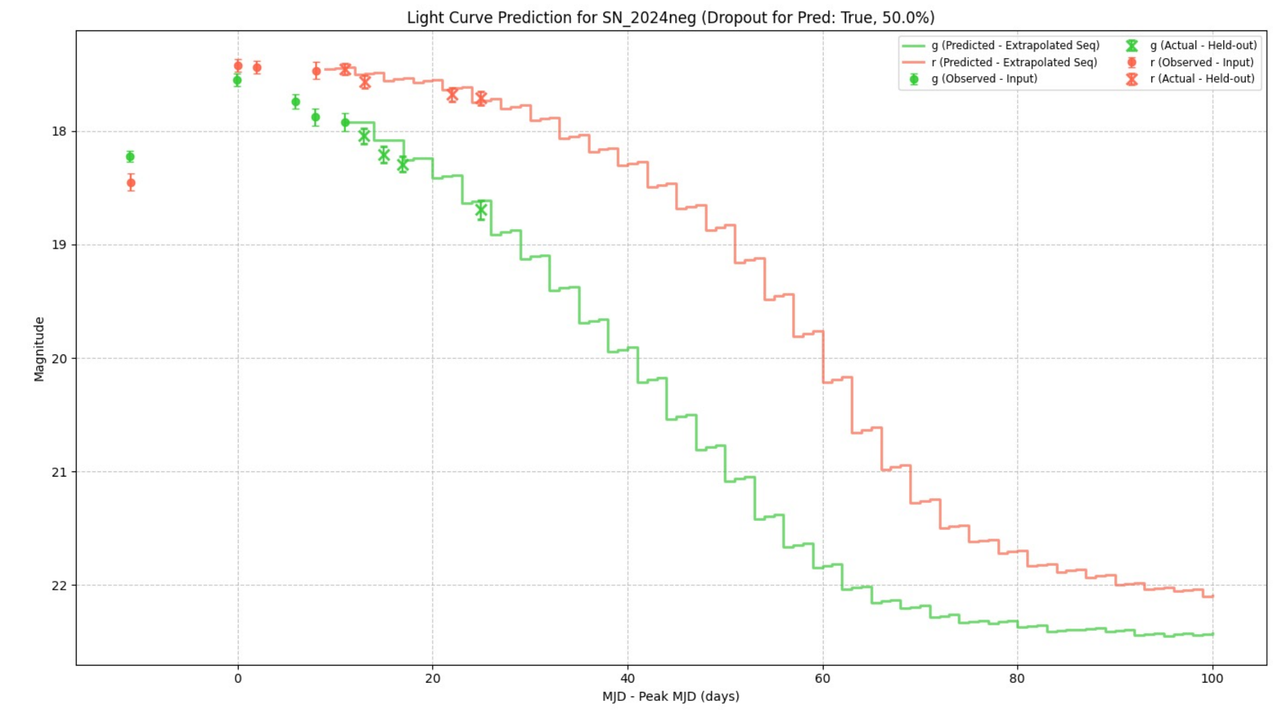
Light Curve Prediction for SN_2020bqt (Dropout for Pred: True, 50.0%)
g (Predicted - Extrapolated Seq)
r (Predicted - Extrapolated Seq)
g (Observed - Input)
g (Actual - Held-out)
r (Observed - Input)
r (Actual - Held-out)
Magnitude
MJD - Peak MJD (days)
19
20
21
22
23
0
20
40
60
80
100

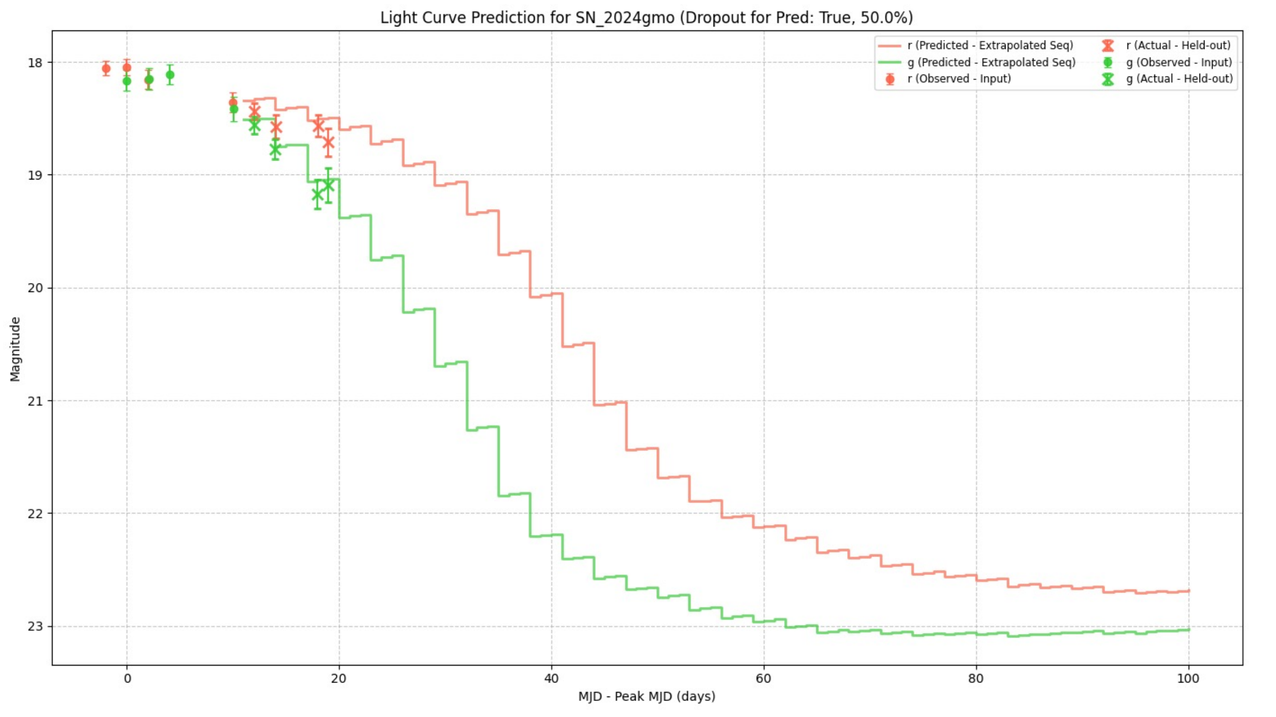


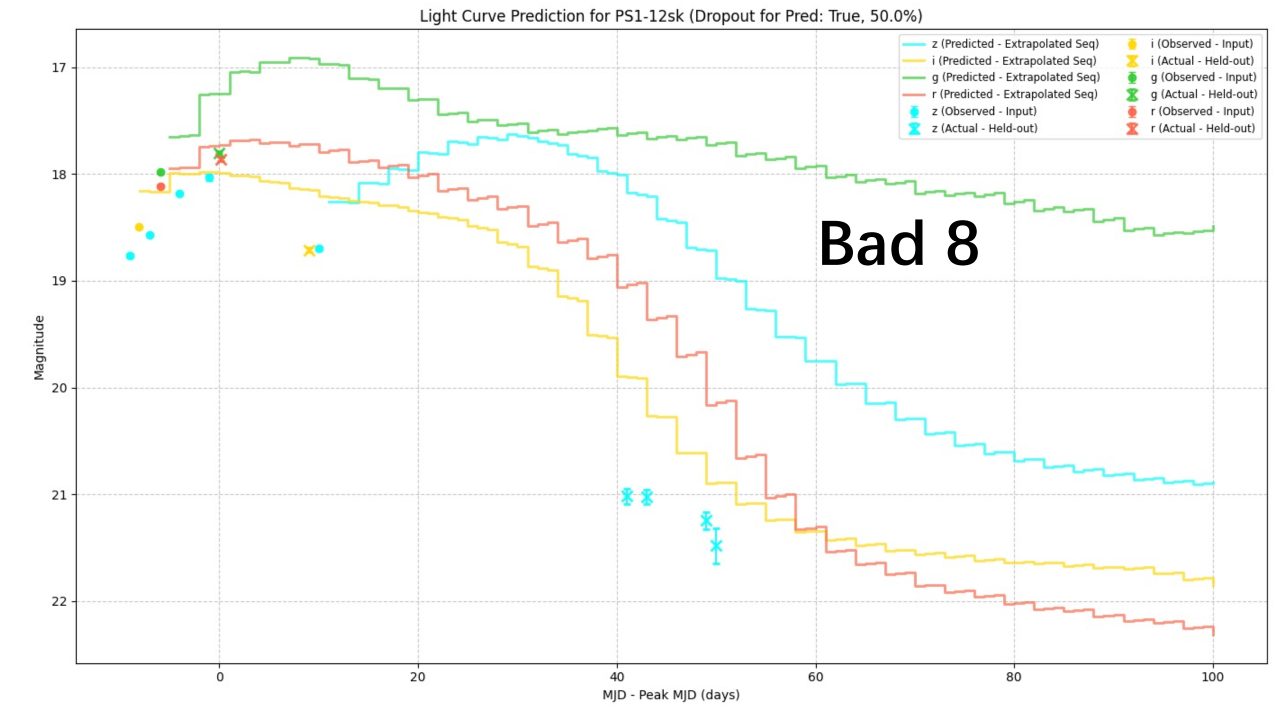
Light Curve Prediction for PS1-13can (Dropout for Pred: True, 50.0%)
z (Predicted - Extrapolated Seq)
g (Predicted - Extrapolated Seq)
i (Predicted - Extrapolated Seq)
r (Predicted - Extrapolated Seq)
z (Observed - Input)
z (Actual - Held-out)
g (Observed - Input)
i (Observed - Input)
i (Actual - Held-out)
r (Observed - Input)
Magnitude
MJD - Peak MJD (days)
21.8
22.0
22.2
22.4
22.6
22.8
23.0
23.2
23.4
0
20
40
60
80
100

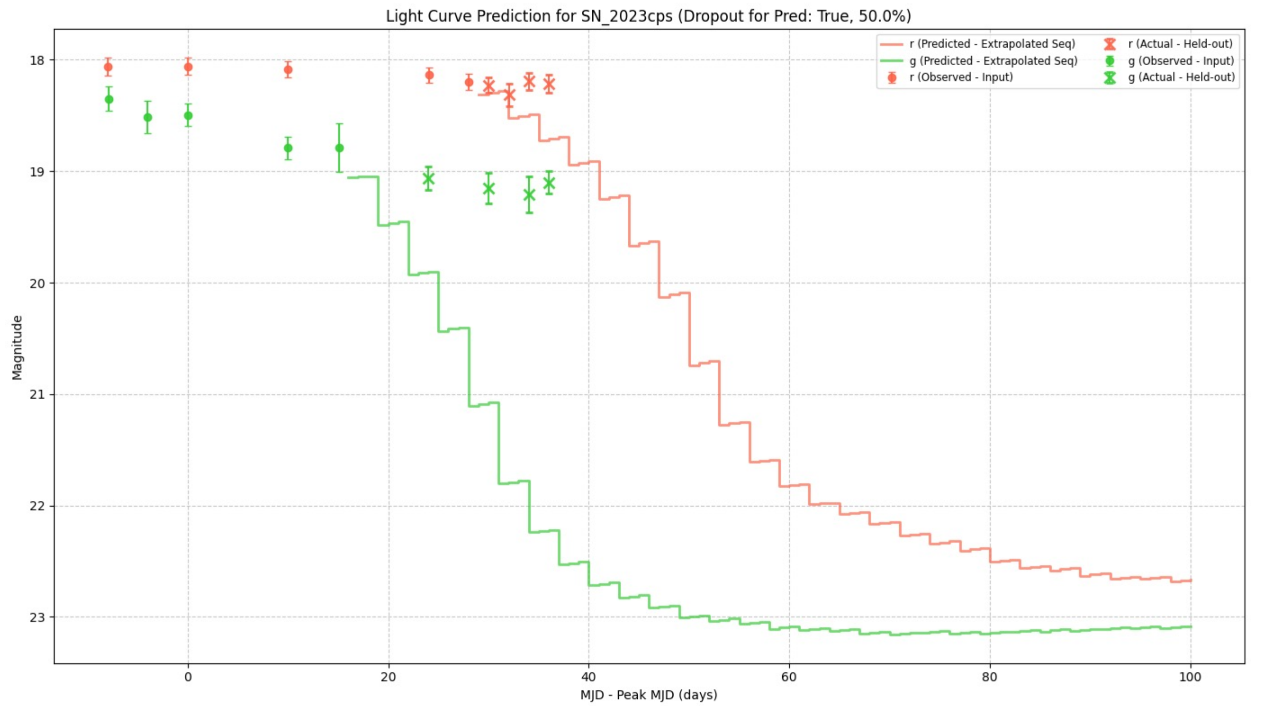

Light Curve Prediction for SN_2023gwl (Dropout for Pred: True, 50.0%)
g (Predicted - Extrapolated Seq)
r (Predicted - Extrapolated Seq)
g (Observed - Input)
g (Actual - Held-out)
r (Observed - Input)
r (Actual - Held-out)
Magnitude
19
20
21
22
23
0
20
40
60
80
100
MJD - Peak MJD (days)

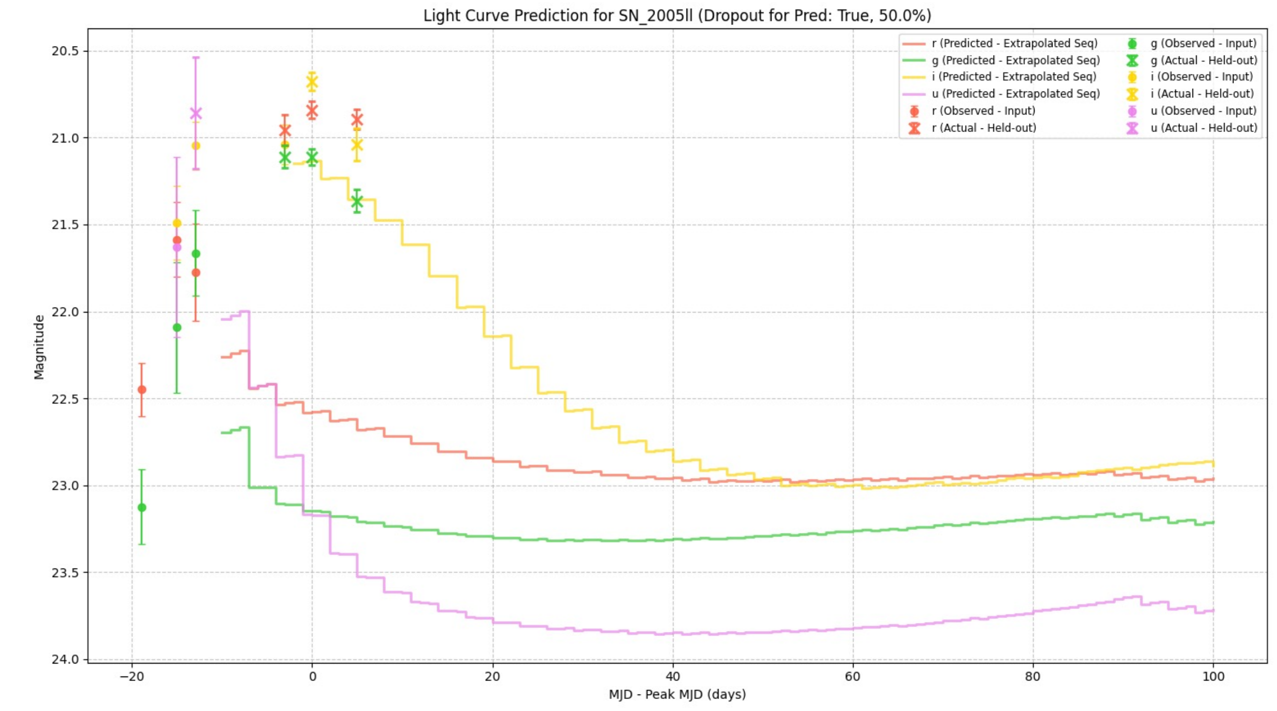

Light Curve Prediction for SN_2024hqp (Dropout for Pred: True, 50.0%)
r (Predicted - Extrapolated Seq)
g (Predicted - Extrapolated Seq)
r (Observed - Input)
r (Actual - Held-out)
g (Observed - Input)
g (Actual - Held-out)
Magnitude
19
20
21
22
23
0
20
40
60
80
100
MJD - Peak MJD (days)

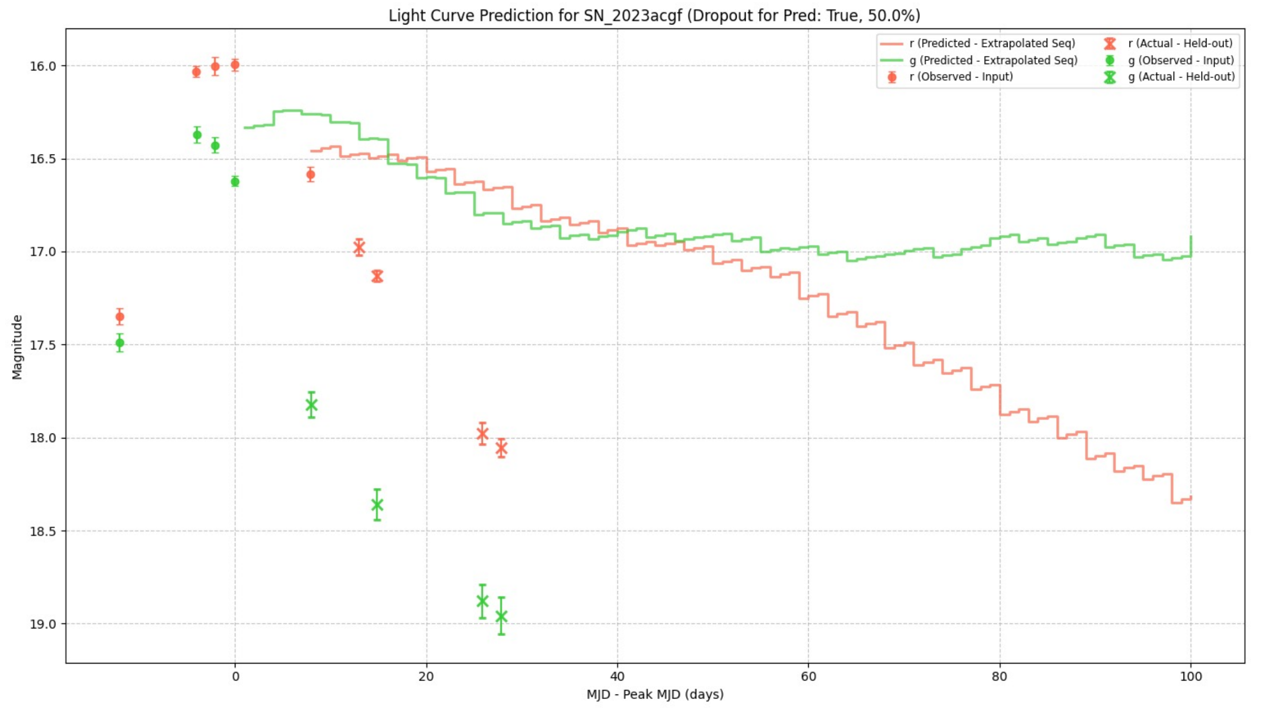

Light Curve Prediction for SN_2024neg (Dropout for Pred: True, 50.0%)
g (Predicted - Extrapolated Seq)
r (Predicted - Extrapolated Seq)
g (Observed - Input)
g (Actual - Held-out)
r (Observed - Input)
r (Actual - Held-out)
Magnitude
18
19
20
21
22
0
20
40
60
80
100
MJD - Peak MJD (days)

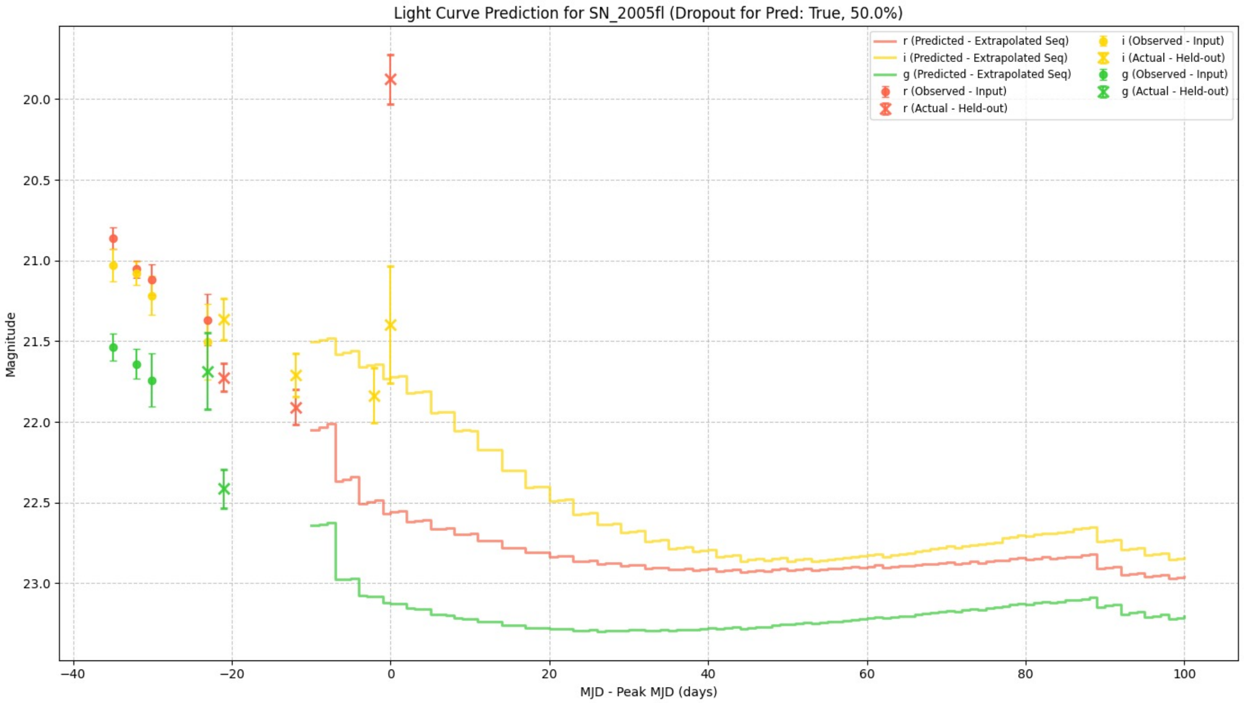

Light Curve Prediction for SN_2024gmo (Dropout for Pred: True, 50.0%)
r (Predicted - Extrapolated Seq)
g (Predicted - Extrapolated Seq)
r (Observed - Input)
r (Actual - Held-out)
g (Observed - Input)
g (Actual - Held-out)
Magnitude
18
19
20
21
22
23
0
20
40
60
80
100
MJD - Peak MJD (days)

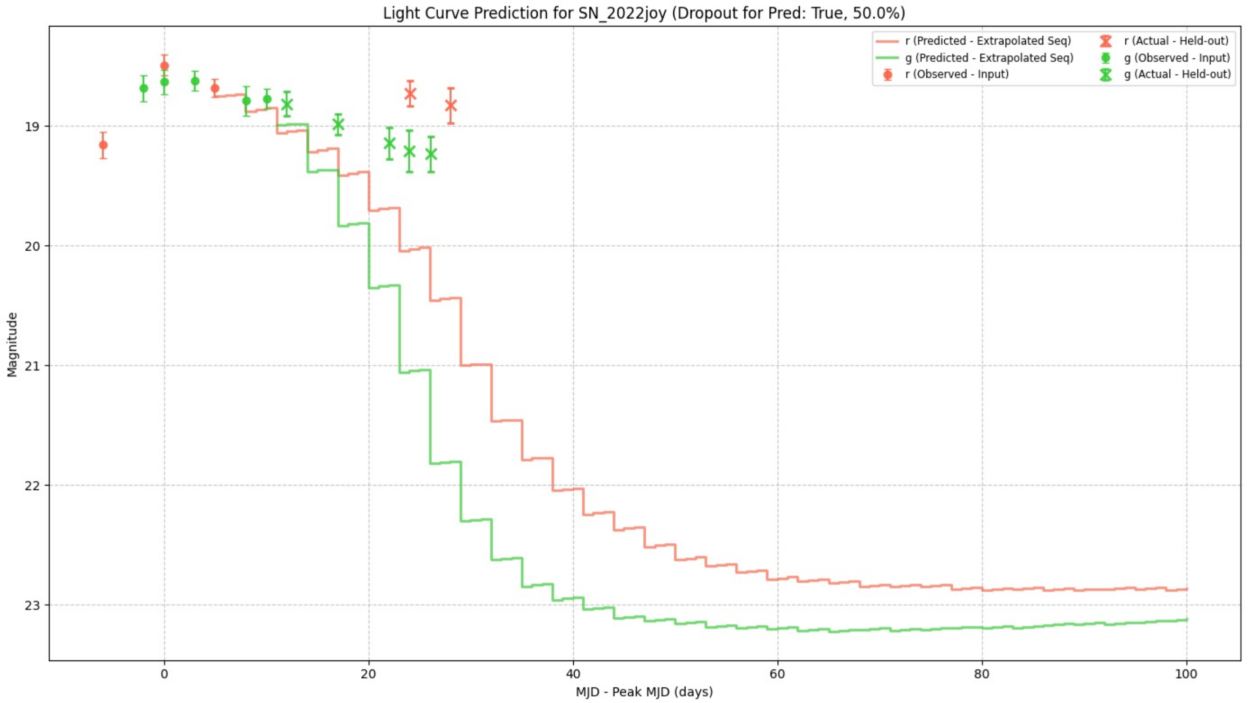

Light Curve Prediction for PS1-12sk (Dropout for Pred: True, 50.0%)
z (Predicted - Extrapolated Seq)
i (Predicted - Extrapolated Seq)
g (Predicted - Extrapolated Seq)
r (Predicted - Extrapolated Seq)
z (Observed - Input)
z (Actual - Held-out)
i (Observed - Input)
i (Actual - Held-out)
g (Observed - Input)
g (Actual - Held-out)
r (Observed - Input)
r (Actual - Held-out)
Bad 8
Magnitude
MJD - Peak MJD (days)

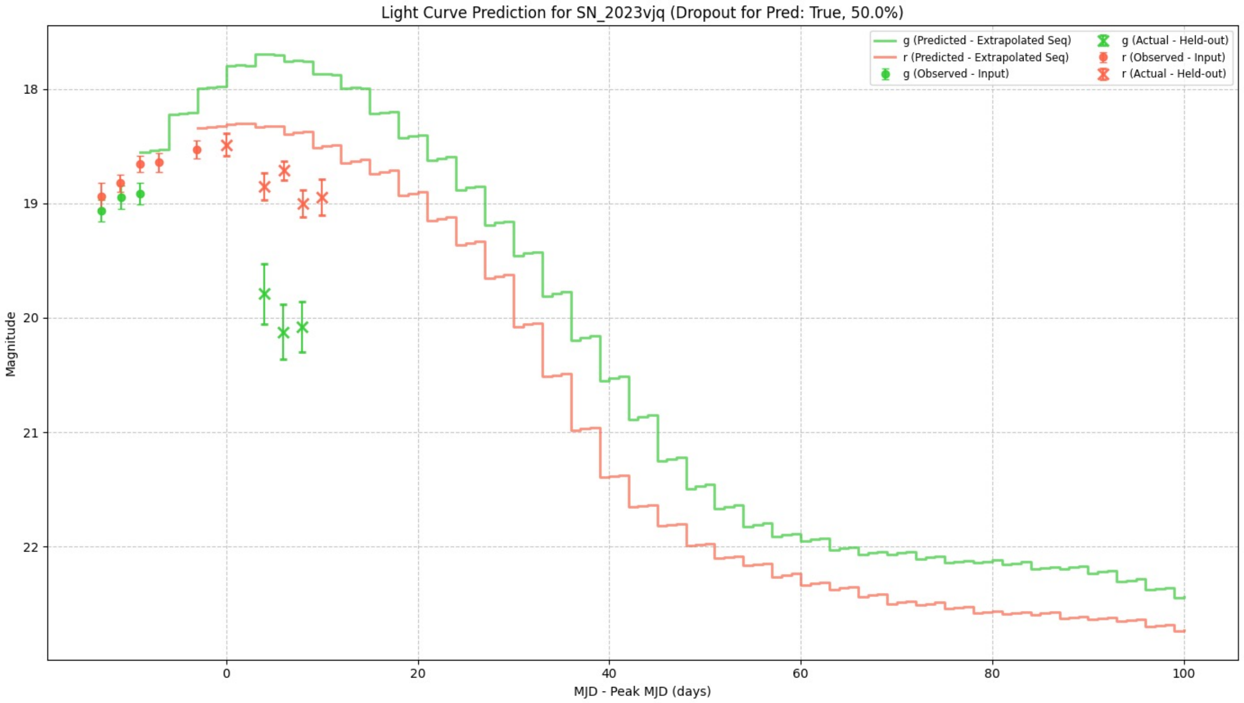

Light Curve Prediction for SN_2023cps (Dropout for Pred: True, 50.0%)
r (Predicted - Extrapolated Seq)
g (Predicted - Extrapolated Seq)
r (Observed - Input)
r (Actual - Held-out)
g (Observed - Input)
g (Actual - Held-out)
Magnitude
MJD - Peak MJD (days)

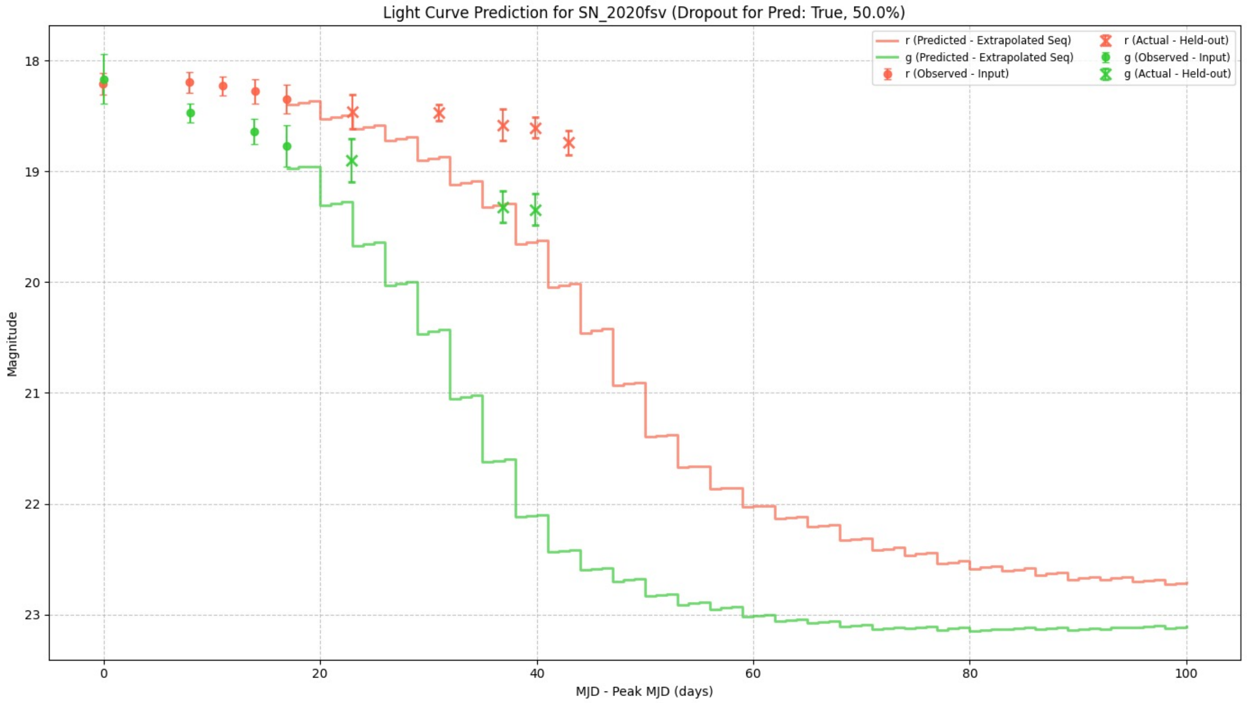

Light Curve Prediction for SN_2005ll (Dropout for Pred: True, 50.0%)
r (Predicted - Extrapolated Seq)
g (Predicted - Extrapolated Seq)
i (Predicted - Extrapolated Seq)
u (Predicted - Extrapolated Seq)
r (Observed - Input)
r (Actual - Held-out)
g (Observed - Input)
g (Actual - Held-out)
i (Observed - Input)
i (Actual - Held-out)
u (Observed - Input)
u (Actual - Held-out)
Magnitude
MJD - Peak MJD (days)


Light Curve Prediction for SN_2023acgf (Dropout for Pred: True, 50.0%)
r (Predicted - Extrapolated Seq)
g (Predicted - Extrapolated Seq)
r (Observed - Input)
r (Actual - Held-out)
g (Observed - Input)
g (Actual - Held-out)
Magnitude
MJD - Peak MJD (days)


Light Curve Prediction for SN_2005fl (Dropout for Pred: True, 50.0%)
r (Predicted - Extrapolated Seq)
i (Predicted - Extrapolated Seq)
g (Predicted - Extrapolated Seq)
r (Observed - Input)
r (Actual - Held-out)
i (Observed - Input)
i (Actual - Held-out)
g (Observed - Input)
g (Actual - Held-out)
Magnitude
MJD - Peak MJD (days)


Light Curve Prediction for SN_2022joy (Dropout for Pred: True, 50.0%)
r (Predicted - Extrapolated Seq)
g (Predicted - Extrapolated Seq)
r (Observed - Input)
r (Actual - Held-out)
g (Observed - Input)
g (Actual - Held-out)
Magnitude
MJD - Peak MJD (days)


Light Curve Prediction for SN_2023vjq (Dropout for Pred: True, 50.0%)
g (Predicted - Extrapolated Seq)
r (Predicted - Extrapolated Seq)
g (Observed - Input)
g (Actual - Held-out)
r (Observed - Input)
r (Actual - Held-out)
Magnitude
MJD - Peak MJD (days)


Light Curve Prediction for SN_2020fsv (Dropout for Pred: True, 50.0%)
r (Predicted - Extrapolated Seq)
g (Predicted - Extrapolated Seq)
r (Observed - Input)
r (Actual - Held-out)
g (Observed - Input)
g (Actual - Held-out)
Magnitude
MJD - Peak MJD (days)

Future work

- Adding multi-band information to the model
- Changing slide-window to full sequence prediction
- Expanding the data set
- Adjusting parameters for higher accuracy

Thank you