UNIVERSITY OF Southampton

UNIVERSITY OF CAMBRIDGE

Augmenting Supernova Training Sets Using Generative Adversarial Networks

Matt Grayling
Institute of Astronomy

ioa

Kavli Institute for Cosmology Cambridge KICC

erc European Research Council
Established by the European Commission

Supernovae

UNIVERSITY OF CAMBRIDGE

Type Ia

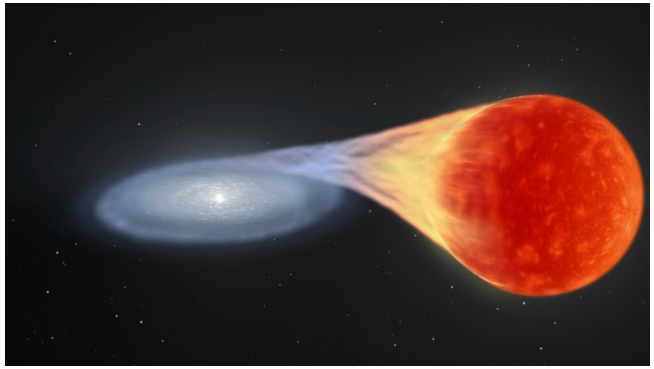
- White dwarf progenitor
- Standardisable candles

Core-collapse

- Massive star progenitor
- Very diverse properties

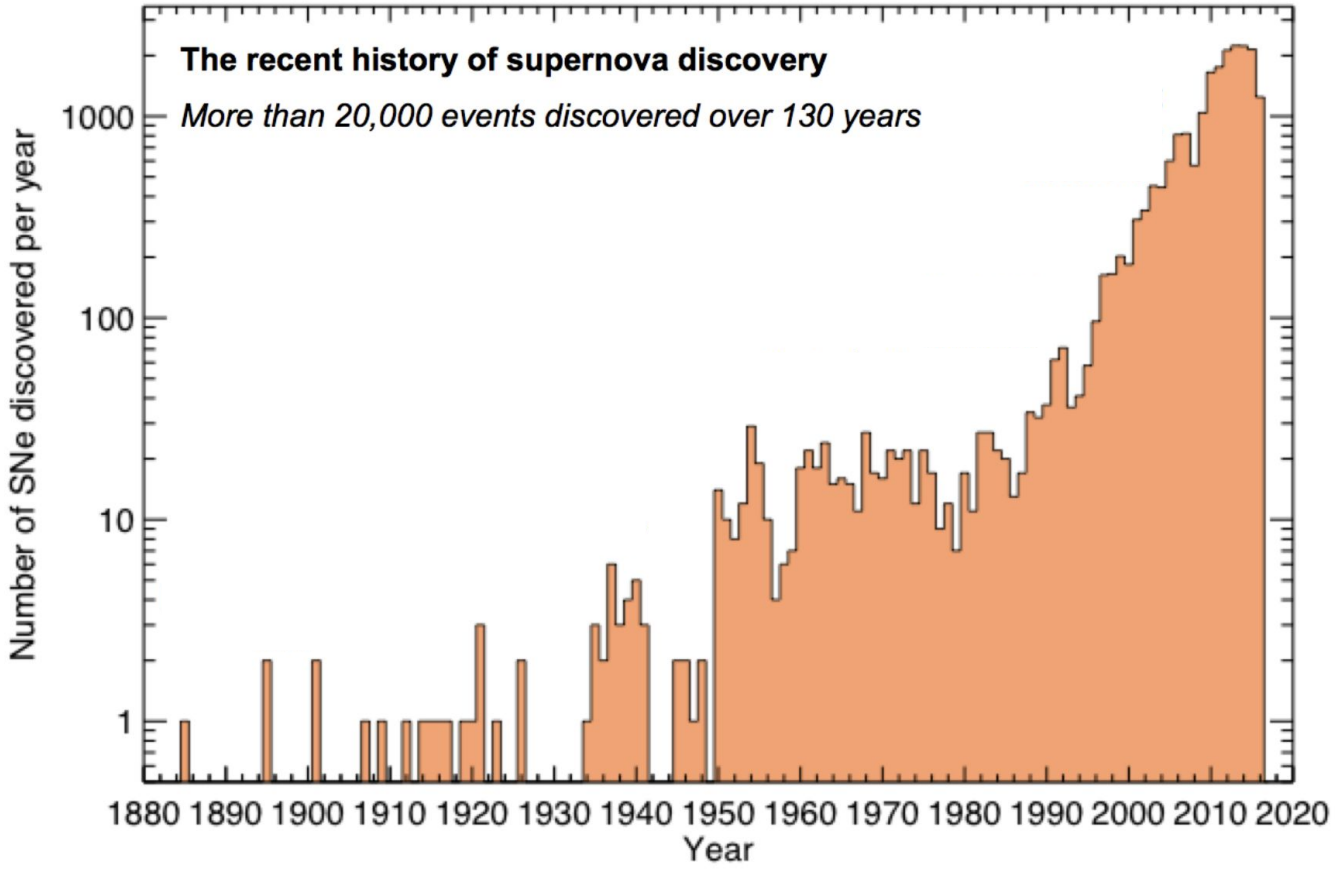
Significance of supernovae:

- Cosmologically useful
- Create and distribute heavy elements
- Influence their surroundings



Credit: Adriana Manrique Gutierrez svs.gsfc.nasa.gov

Supernovae



From 2024(ish):

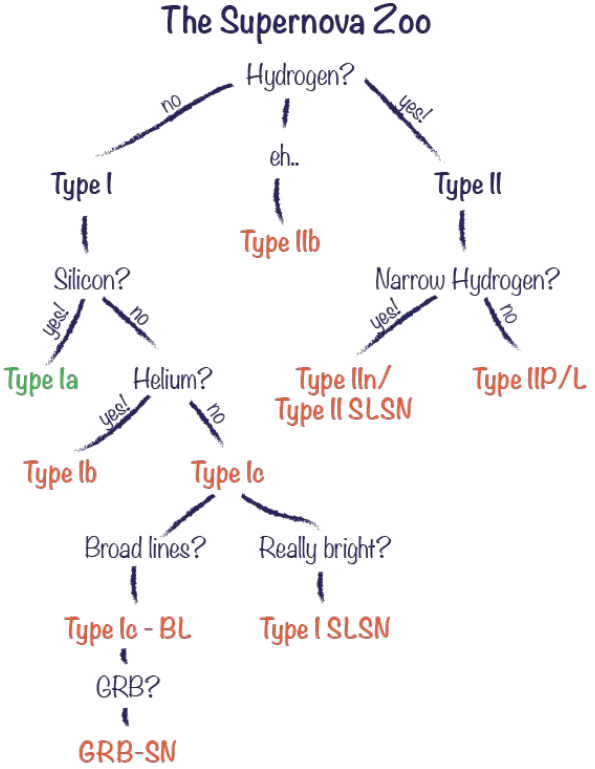
Legacy Survey of Space and Time at Vera Rubin Observatory (LSST) - tens of thousands per year

Credit: Mark Sullivan

Supernova Classification

UNIVERSITY OF CAMBRIDGE

- Supernova classes defined based on spectral features
- High rate of discovery makes photometric classification vital
- For supernova cosmologists, classification is a simple binary Ia vs non-Ia problem
- For core-collapse astrophysicists, it is far more complex than that
- Machine learning-based classification techniques are being widely developed



Credit: Ashley Villar

Training on Simulations

UNIVERSITY OF CAMBRIDGE

- Challenges with training on real data
  - Lack of training data
  - Biased training set
- Training on simulated data mitigates for the issues involved
- Limited by quality of simulations
  - Accurate simulations require good physical understanding of population
  - Supernovae are complex and diverse, simulations fail to reproduce full variability of population
  - Consistent drop in performance when applying models trained on simulations to real data

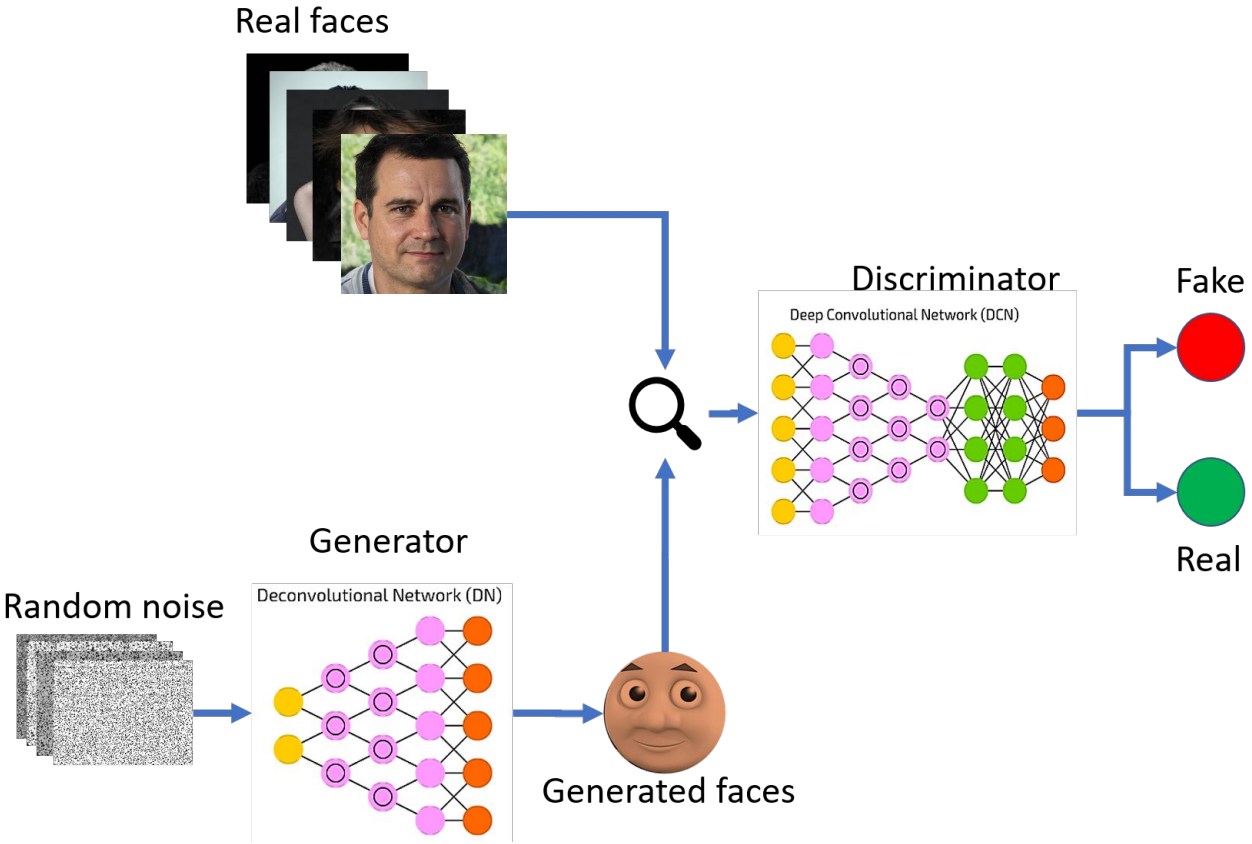
Generative Models

UNIVERSITY OF CAMBRIDGE

- Generative models allow you to draw from the underlying distribution of your data without making any assumptions about it
- We can use generative models to create synthetic supernovae without knowing anything about the physics involved
- Generating synthetic training sets using generative adversarial networks (GANs) has been used to improve the performance of classification models, both within astronomy and beyond (e.g. Motamed+21, Garcia-Jara+22)

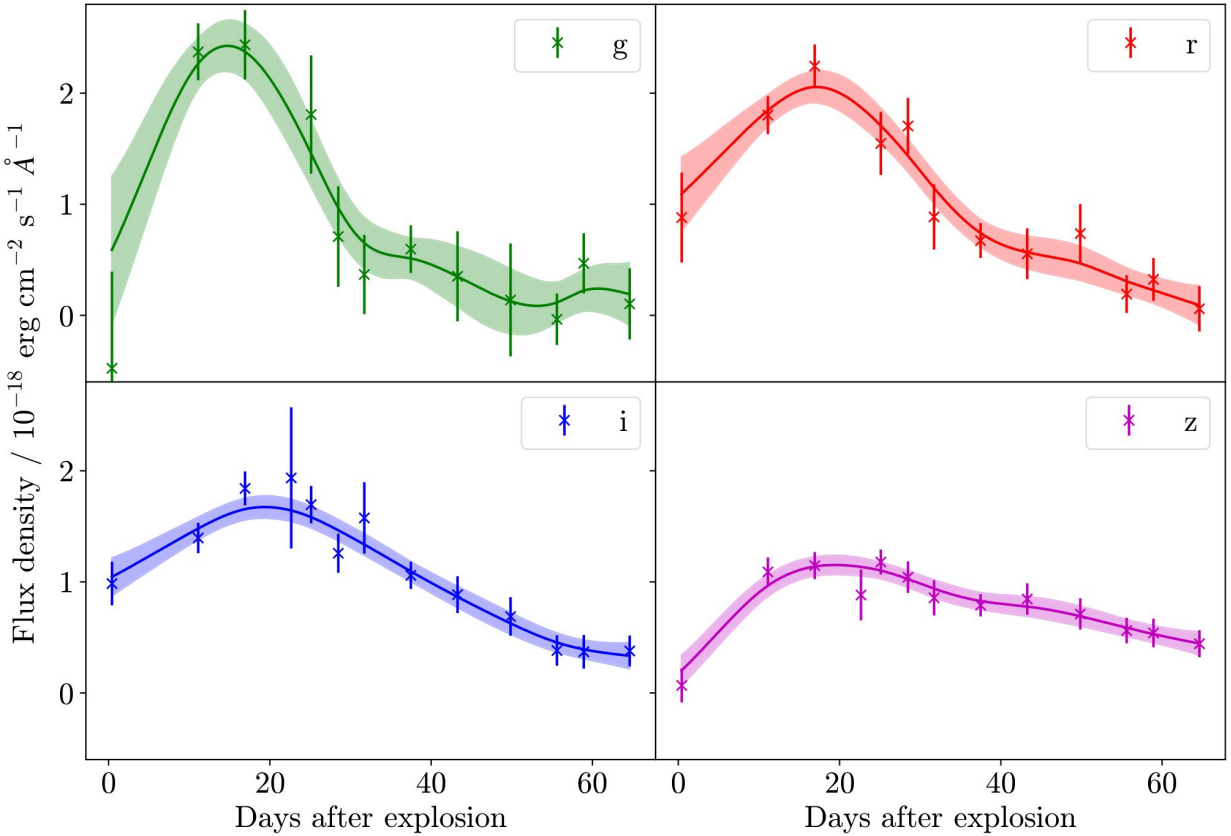
Generative Adversarial Networks (GANs)

UNIVERSITY OF CAMBRIDGE



Application to Supernovae

UNIVERSITY OF CAMBRIDGE



Application to Supernovae

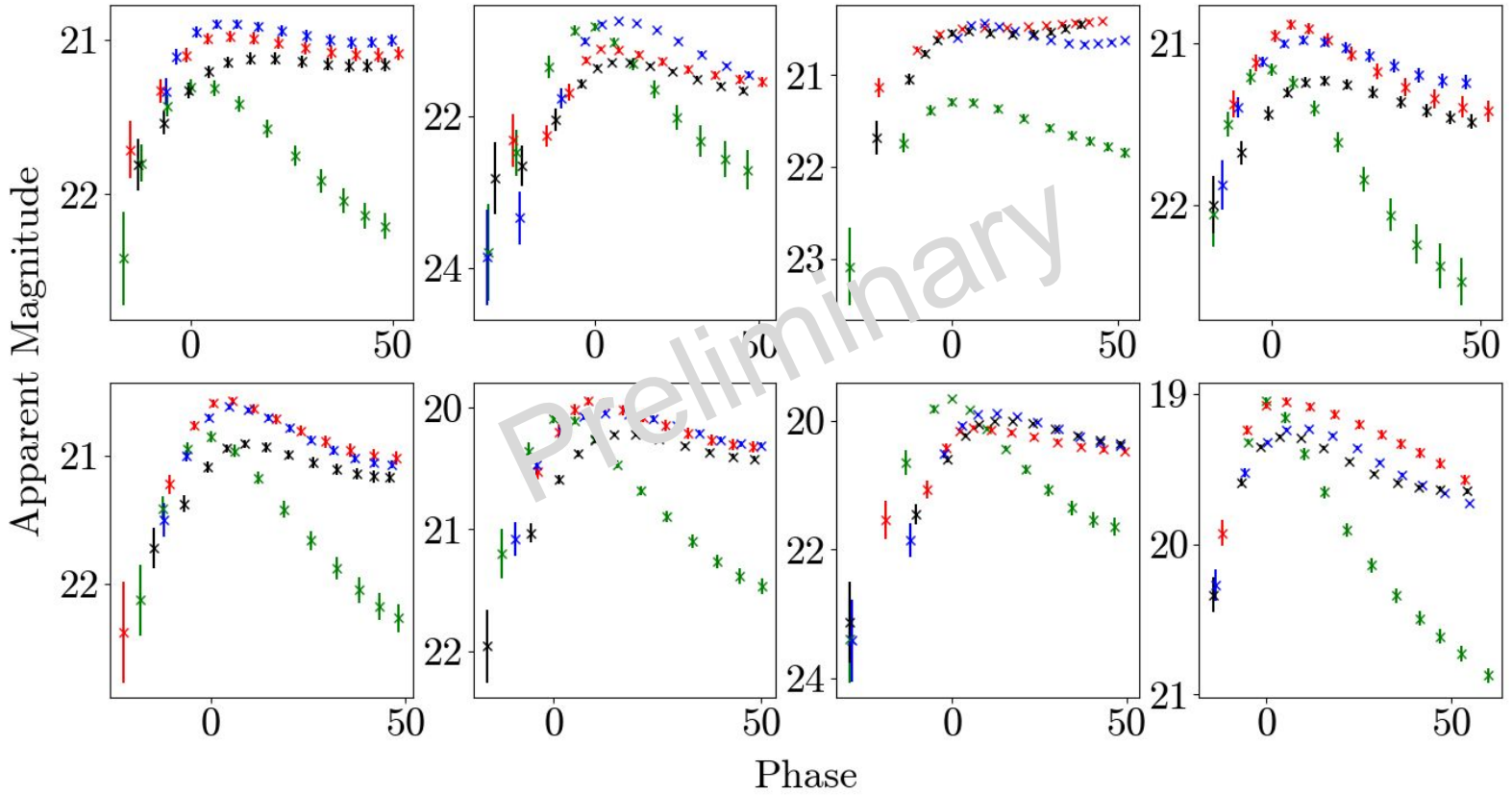
UNIVERSITY OF CAMBRIDGE

- Train a GAN using recurrent neural networks (RNNs) for both generator and discriminator, allows for variable length time series generation
- Phases and photometric data + uncertainties generated together as below

	g_t	r_t	i_t	z_t	g	r	i	z	g_err	r_err	i_err	z_err
1												
2												
⋮												
N												

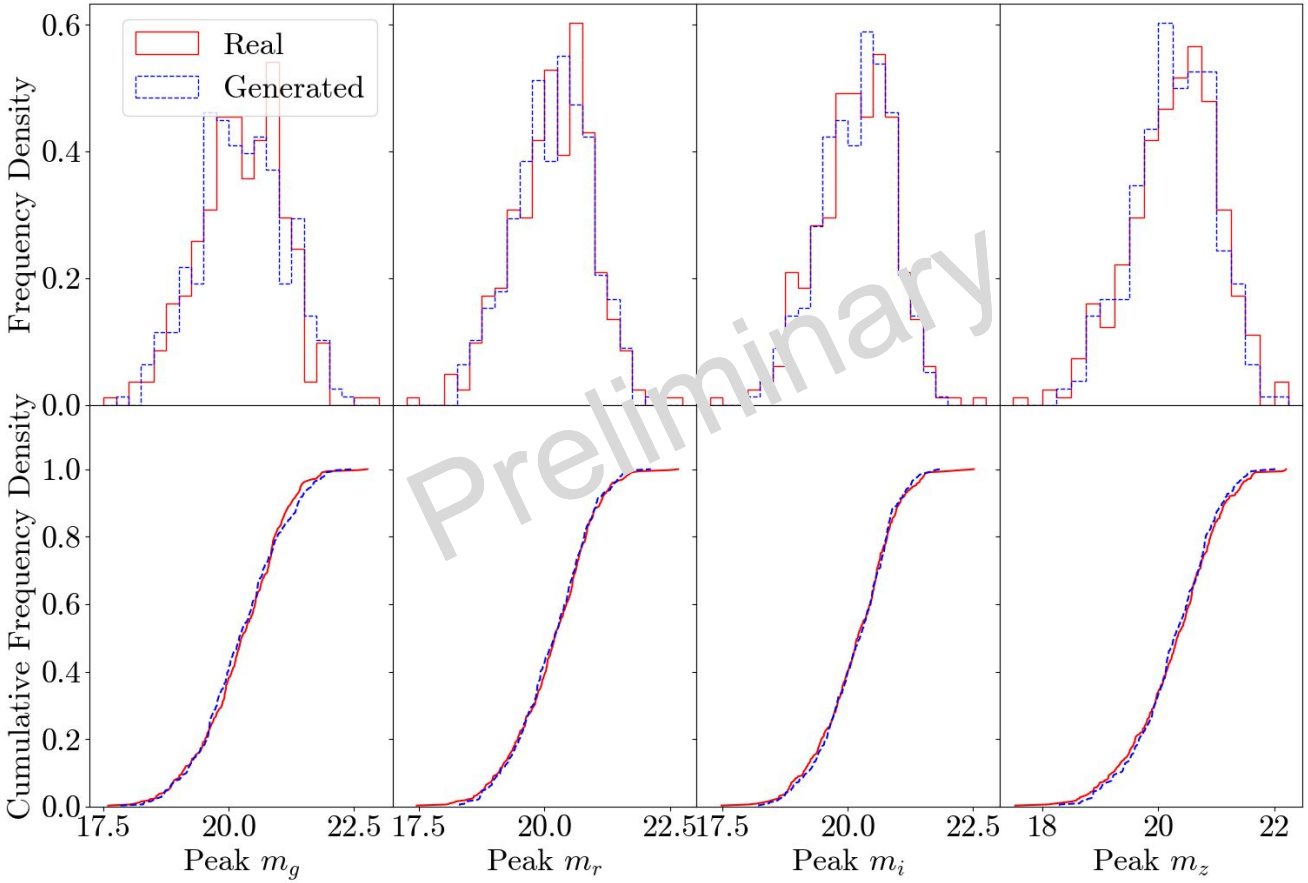
Light Curve Examples (SNe II)

UNIVERSITY OF CAMBRIDGE



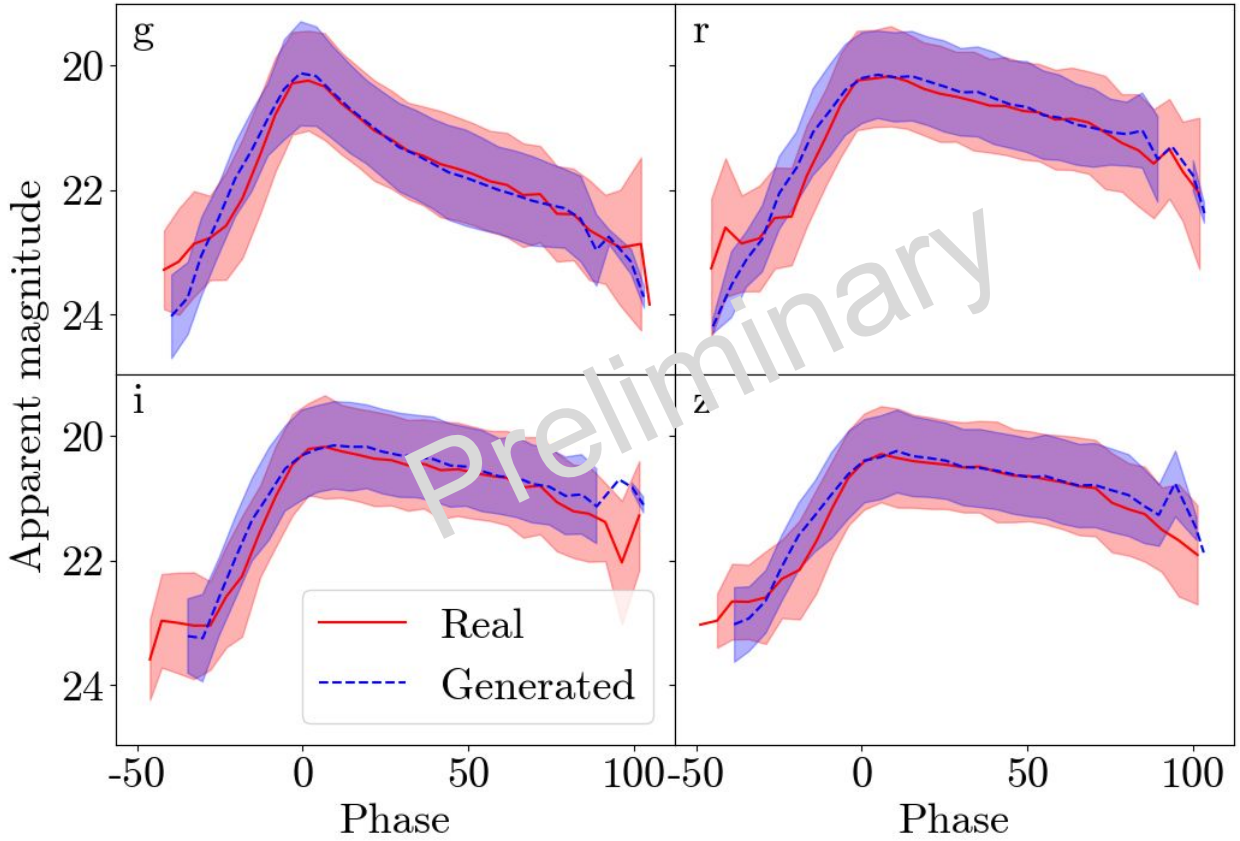
Peak Observed Brightness

UNIVERSITY OF CAMBRIDGE



Light Curve Populations

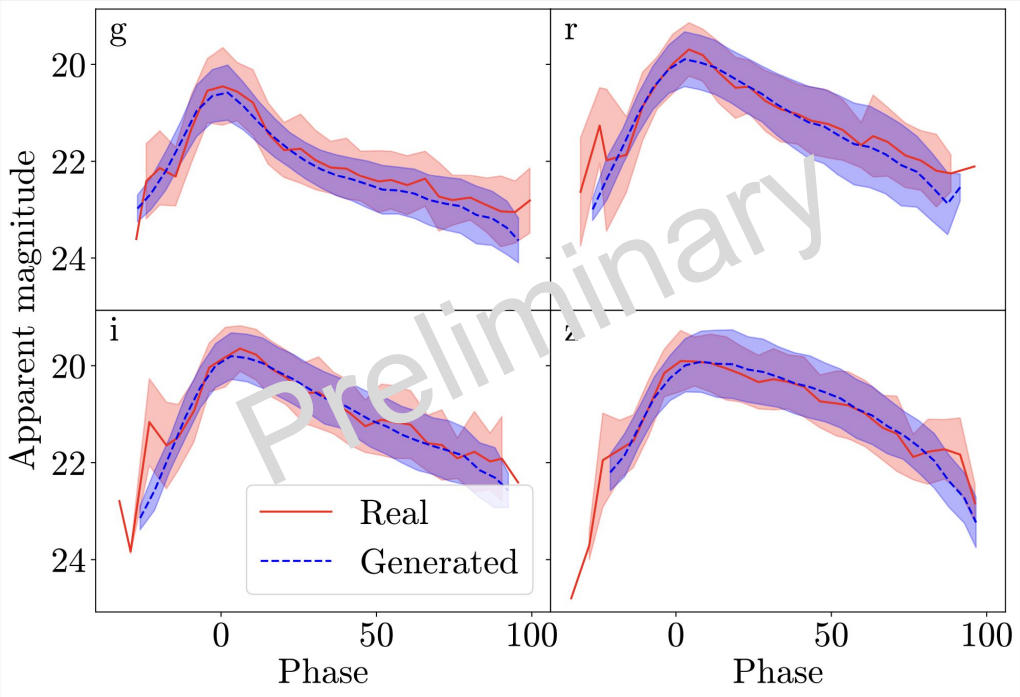
UNIVERSITY OF CAMBRIDGE



Augmenting Rare Supernova Classes

UNIVERSITY OF CAMBRIDGE

- Previous plots were for sample of ~350 simulated supernovae
- Training on a much smaller data set of ~30 supernovae still produces comparable results
- For rare SN classes with few examples, this approach has the potential to augment very sparse training sets



Summary

UNIVERSITY OF CAMBRIDGE

- GANs have the ability to generate realistic supernova light curves, and the potential to improve the performance of photometric classifiers
- Improvements and future work:
  - Train a classifier with generated data and assess performance
  - Retrain on real data
  - Explore conditional models, allowing one model to generate all types of supernovae
  - Explore other types of generative model such as diffusion models