ConvEntion:

Astronomical Image Time Series Classification Using Convolutional attEntion

Anass BAIROUK
Directors: Dominique FOUCHEZ, Marc CHAUMONT
Co-supervisors: Frederic COMBY, Jérôme PASQUET

University of Montpellier

Outlines

1. Introduction
2. Traditional methods
3. Our approach
4. Results
5. Conclusion & Perspective

Introduction

Introduction

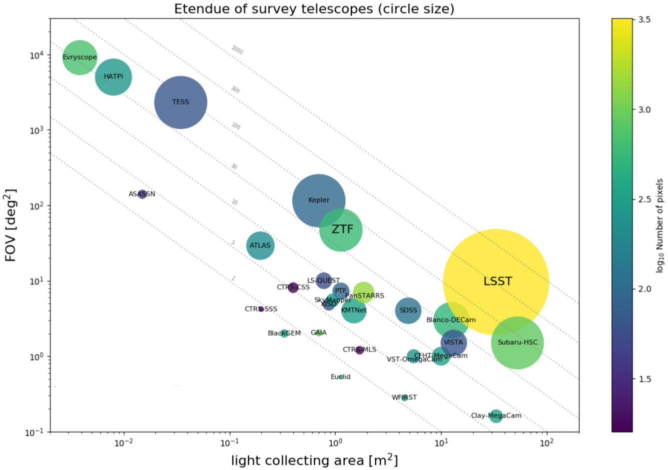


Figure 1: Figure by project ALeRCE [3]

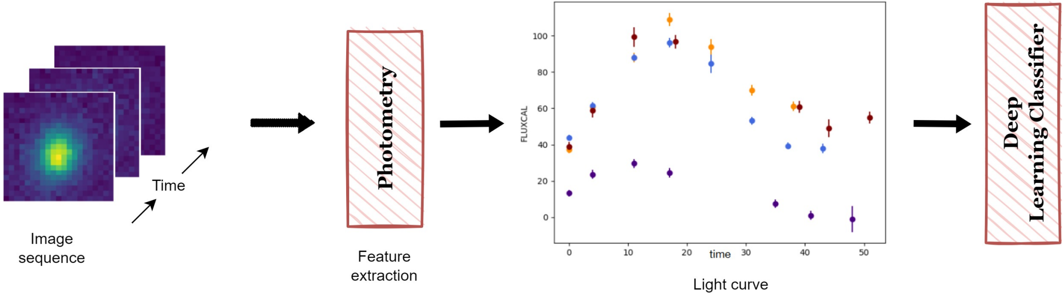
Introduction

The whole endeavour will create an enormous amount of data. Over 15 terabytes will have to be processed every night

Figure 2: Supernovae from the 2005-2007 observing campaigns of SDSS Survey.

Traditional methods

Overview


Time
Image
sequence
Photometry
Feature
extraction
FLUXCAL
time
Light curve
Deep
Learning Classifier

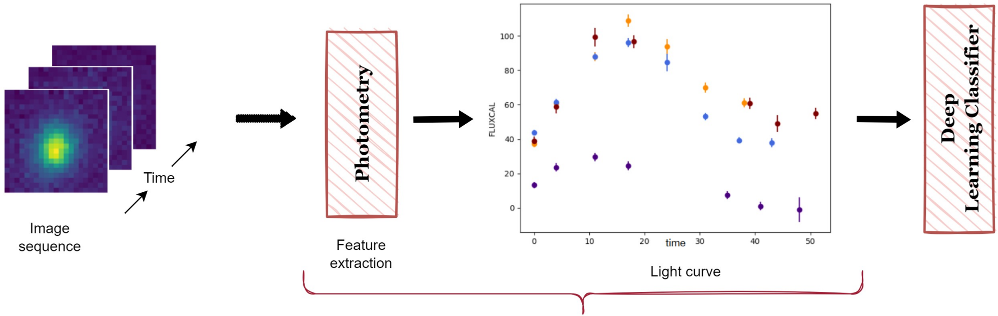
Limitations

The traditional method achieved remarkable results yet it has SOME limitations. Indeed, feature extraction (to produce calibrated fluxes + errors) loses information :

1. No information on the quality of the image (observing condition) quality reduction
2. The flux extraction may not be optimal (alignment of images, PSF estimation, ...)
3. The background scene (usually a star or a galaxy) is not used

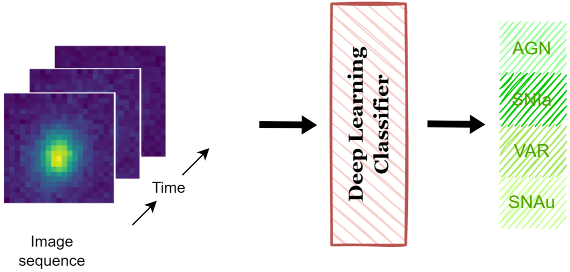
Our approach

Goal



Eliminate the use of light curves

Goal



Time
Image sequence
Deep Learning Classifier
AGN
SNIa
VAR
SNAu

Challenges

- few existing works
- Few Datasets to use for training
- Imbalanced classes

Class Name	Count
AGN	906
SNIa	1988
Variable	3225
SNOther	2130

Table 1: Number of objects per class for SDSS database.

Challenges

- Mismatched dataset
- Time series classification is a hard task in deep learning
- Missing observations

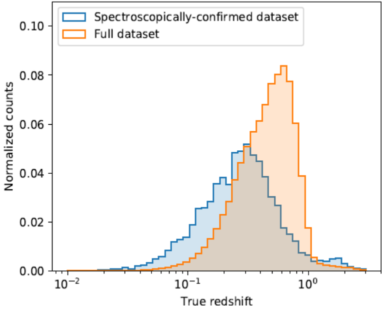


Figure 3: True redshift distributions for both the training and test sets in the PLAsTiCC dataset [1].

Challenges

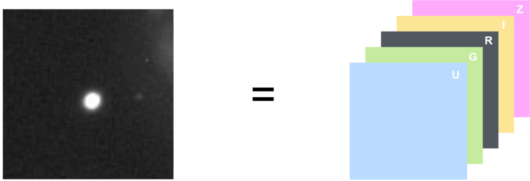


Figure 4: Each image has five filters (u, g, r, i, z). The black channel represent the missing observation

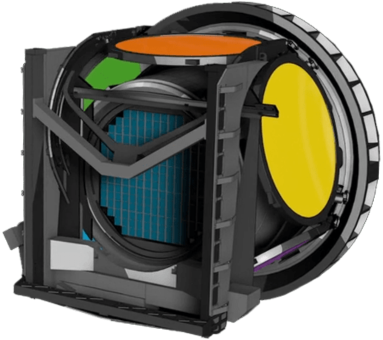


Figure 5: LSST filter changer

▸ See how filter changer work

Tackled challenges

1. Develop a sequential model that classifies Astronomical image time series.
2. Mitigate the impact of missing observations and class imbalance

Class imbalance

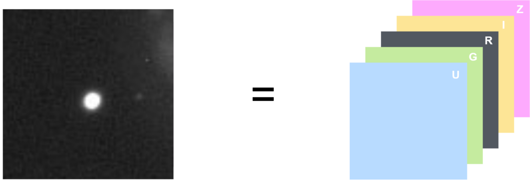
- We **oversampled** our database using **data augmentation**
- **Data augmentation**: Dropping some steps from the sequence, sequence rotation, horizontal and vertical flip. Sequence shifting.

Data modeling

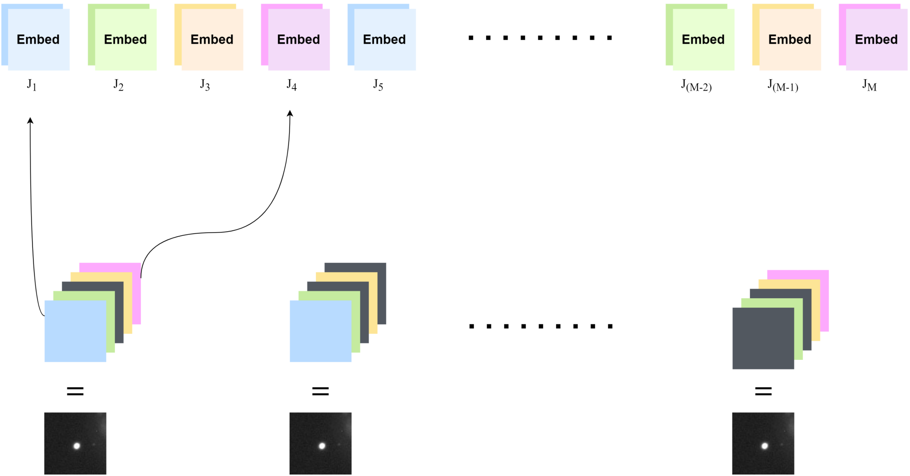
- For each band (u, g, r, i, z), we assign a unique number $id \in \{1, 2, 3, 4, 5\}$. Then we create 2D band embedding $BE_{id} \in \mathbb{R}^{H \times W}$

$$BE_{id} = BandEmbed(id) \quad (1)$$

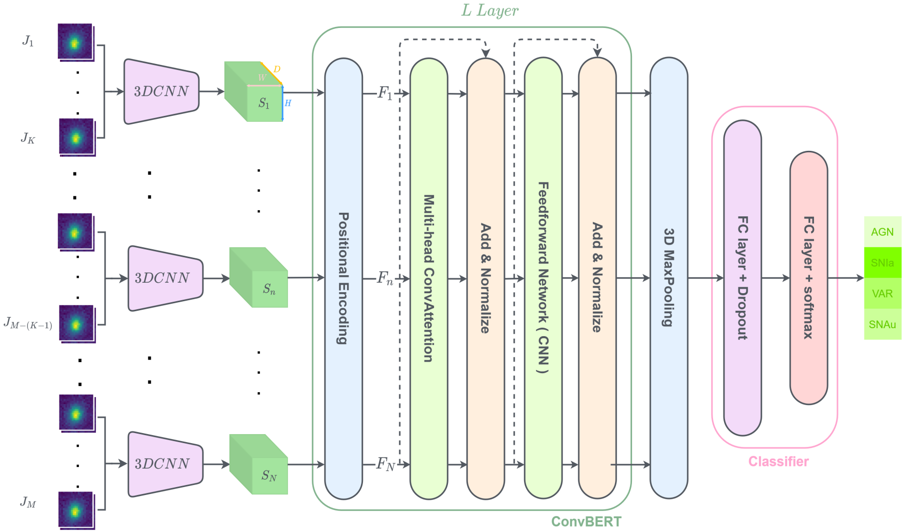
$$J_m = Concat(X_m, BE_{id}) \qquad m \in [1, .., M] \quad (2)$$



Data modeling



Architecture

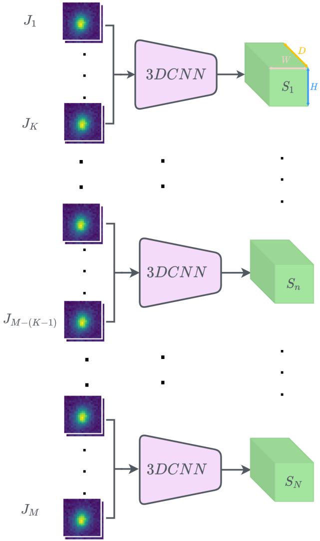


3D Convolution Network

- The original Transformer's self-attention mechanism consumes $O(M^2)$. Ours $O(M^2 \times H \times W)$

$$S_n = 3DCNN(J_{(n-1)\times K+1}, \ldots, J_{n\times K}) \tag{3}$$

- Helps capture the local spatial-temporal high-level features

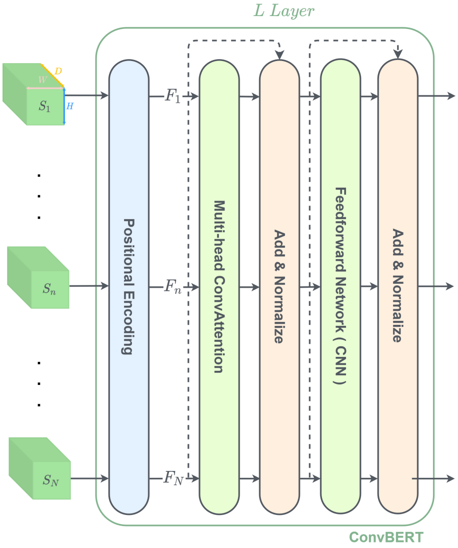


Convolutional BERT

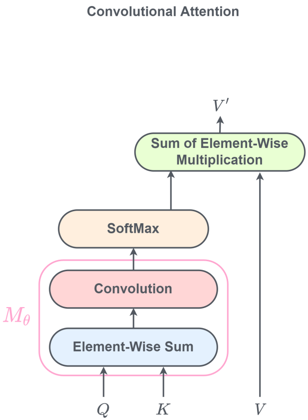
- BERT stands for Bidirectional Encoder Representations from Transformers
- Positional encoder:

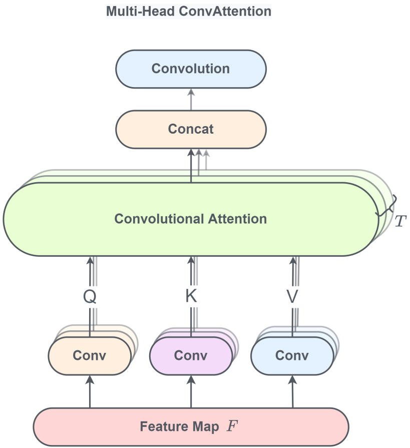
$$P_{(n,2i)} = sin(n/10000^{2i/D}), \quad (4)$$

$$P_{(n,2i+1)} = cos(n/10000^{2i/D}), \quad (5)$$



MultiHead Convolutional Self Attention





Liu et al. (2021), ConvTransformer [6]

MultiHead Convolutional Self Attention

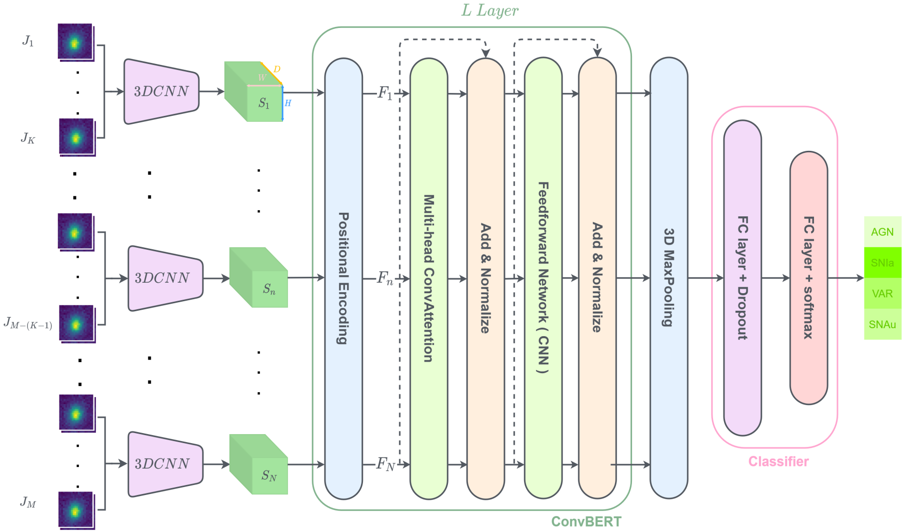
$$H_{(n,m)} = M_\theta(Q_n, K_m) \qquad n,m \in [1,..,N] \tag{6}$$

$$H_n = SoftMax(H_n) \quad where \quad H_n \in \mathbb{R}^{H' \times W' \times N} \tag{7}$$

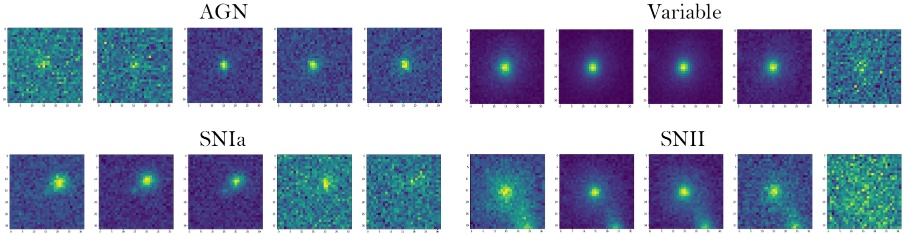
$$V'_n = \sum_{m=1}^{N} H_{(n,m)} V_m \tag{8}$$

$$MultiHead(Q,K,V) = Concat(V'_{n_1}, ..., V'_{n_T}) \tag{9}$$

Architecture



Dataset



We used the SDSS dataset [5] that contains 10258 objects. We organized the dataset for two types of class categories

- Three Classes: AGN, SN, Variable
- Four Classes: AGN, SNIa, Variable, SNOther

Results

Comparison

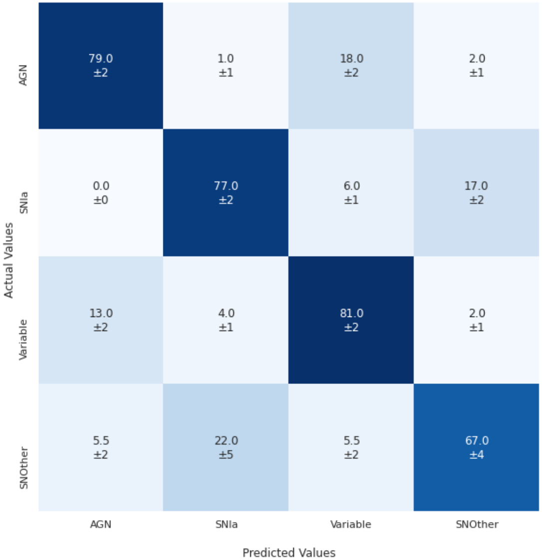
Model	Bands	Dataset	Accuracy	F1 scoreˆ	Number params
ConvEntion (Ours)	ugriz	Images	**79.83**	**70.62**	1.253M
CNN+GRU [4]	ugriz	Images	66.39	63.22	1.993M
ConvEntion (Ours)	g	Images	76.89	63.20	1.253M
CNN+GRU [4]	g	Images	63.67	61.00	1.992M
CNN+LSTM [2]	ugriz	Images	64.08	60.65	2.190M
CNN+LSTM [2]	g	Images	63.00	60.00	2.189M
SuperNNova (BNN) [7]	ugriz	Light curves	65.54	55.40	-
SITS-BERT [9]	ugriz	Light curves	67.43	51.60	0.596M
SCONE (CNN) [8]	ugriz	Light curves	62.57	50.43	22.2K
SuperNNova (RNN) [7]	ugriz	Light curves	56.30	42.60	-
LSTM	ugriz	Light curves	55.24	40.33	60K

Comment

- Images perform better than light curves
- ConvEntion has less parameters and take less time to train and achieved the best accuracy
  - Parallel computation
  - Merging spatio-temporal feature
  - Local features are captured by 3DCNN while ConvBERT handles the global features.

MC

- Supernovas share a lot of similarities which confused the model.
- Misclassification between AGN and variable



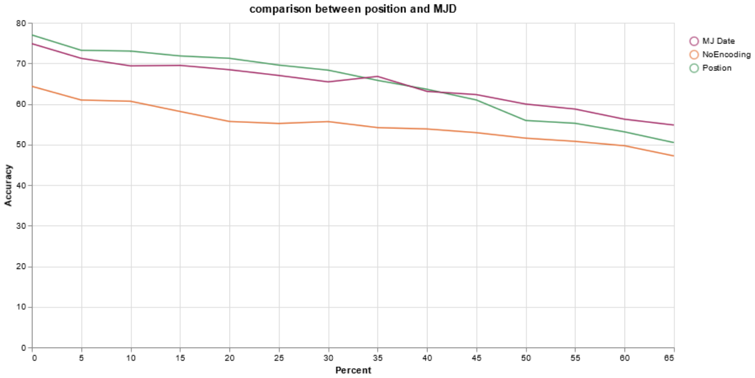
Ablation results

The following table represents different ablation experiments to show the impact of each component in our model.

Model	Accuracy	F1 scoreˆ	Run time
ConvEntion	**79.83**	**70.62**	1.5
No Oversampling	79.36	64.23	1.5
No Band Embedding	70.74	59.85	1.5
Fixed Band Embedding	78.45	65.73	1.5
2D CNN	77.38	62.25	4.5

Ablation results

This figure shows the comparison of macro accuracy using MJD versus the position in the function of the percentage of missing observations.



Conclusion & Perspective

Conclusion

1. We mitigated the problem of missing observation.
2. We introduced a solution to lower the computation of the attention map. Also, capturing local features using 3DCNN.
3. We propose the first ever convolutional BERT to handle spatio-temporal features.
4. We mitigate the impact of class imbalance using oversampling and data augmentation.

Perspective

- We will try to study the impact of some important metadata on the model's performance.
- Improve the model discrimination capabilities using contrastive learning.
- Conduct a major analysis of the results of our work.

Questions?

References i

K. Boone.
Avocado: Photometric classification of astronomical transients with gaussian process augmentation.
The Astronomical Journal, 158(6):257, Dec 2019.

R. Carrasco-Davis, G. Cabrera-Vives, F. Förster, P. A. Estévez, P. Huijse, P. Protopapas, I. Reyes, J. Martínez-Palomera, and C. Donoso.
Deep learning for image sequence classification of astronomical events.
Publications of the Astronomical Society of the Pacific, 131(1004):108006, Sep 2019.

References ii

F. e. a. Förster.
The automatic learning for the rapid classification of events (alerce) alert broker.
The Astronomical Journal, 161(5):242, Apr 2021.

C. Gómez, M. Neira, M. Hernández Hoyos, P. Arbeláez, and J. E. Forero-Romero.
Classifying image sequences of astronomical transients with deep neural networks.
Monthly Notices of the Royal Astronomical Society, 499(3):3130–3138, Oct 2020.

- J. A. Holtzman et al.
  The Sloan Digital Sky Survey-II Photometry and Supernova IA Light Curves from the 2005 Data.
  Astron. J., 136:2306–2320, 2008.
- Z. Liu, S. Luo, W. Li, J. Lu, Y. Wu, S. Sun, C. Li, and L. Yang.
  Convtransformer: A convolutional transformer network for video frame synthesis, 2021.

References iv

A. Möller and T. de Boissière.
Supernnova: an open-source framework for bayesian, neural network-based supernova classification.
Monthly Notices of the Royal Astronomical Society, 491(3):4277–4293, Dec 2019.

H. Qu, M. Sako, A. Möller, and C. Doux.
Scone: Supernova classification with a convolutional neural network.
The Astronomical Journal, 162(2):67, Jul 2021.

References v

Y. Yuan and L. Lin.
Self-supervised pretraining of transformers for satellite image time series classification.
IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14:474–487, 2021.