Research in Astronomy and Astrophysics, 25:104004 (13pp), 2025 October
https://doi.org/10.1088/1674-4527/adf70e
CSTR: 32081.14.RAA.adf70e
© 2025. National Astronomical Observatories, CAS and IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Deep Learning to Classify Exoplanet Light Curves in Kepler and TESS

Dapeng Xie[1], Ying Wang[1], Fuyao Liu[2], and Wei Sun[1]
[1] School of Mathematics, Physics and Statistics, Shanghai University of Engineering Science, Shanghai 201620, China; wangying424524@163.com
[2] Shanghai Zhongqiao Vocational and Technical University, Shanghai 201514, China
Received 2025 February 25; revised 2025 April 9; accepted 2025 April 14; published 2025 September 12

Abstract

With the advancement of observational techniques, over one million stars have been observed to search for exoplanets. The vast amounts of light curve data generated necessitate an automated, rapid, and efficient method for screening exoplanet candidates. In recent years, deep learning methods, particularly convolutional neural networks (CNNs), have been employed to automate the screening of light curves for exoplanet candidates. However, CNNs have several drawbacks, such as slow training times due to their large network structures and a lack of ability to explain the physical meaning behind their classifications. In this paper, we propose a CNN classification model that incorporates a channel attention mechanism, which captures significant features in transit signals and enhances the model's classification capability and interpretability. Furthermore, we have improved the fully connected neural network by incorporating the residual network structure, significantly increasing the model's training speed. Our model is highly adaptable to various data sets, achieving 96.2% accuracy and 95.7% F1 score on the Kepler data set, and 99.9% accuracy with 99.5% F1 score on the Transiting Exoplanet Survey Satellite data set, with training times ranging from 0.3 to 0.6 hr per data set. We also analyze the feature heat maps of the channel attention mechanism to better explain the model's classifications. Our study demonstrates that deep learning can assist in managing the current era of astronomical big data and suggests that the channel attention mechanism and residual network structure could be applied to other astronomical classification tasks.

Key words: methods: data analysis – Planetary Systems – Astronomical Databases

1. Introduction

In 1992, radio astronomers Alexander Wolszczan and Dale A. Frail made a groundbreaking discovery: two planets orbiting the pulsar PSR B1257+12 (Wolszczan & Frail 1992), marking the first unequivocal confirmation of extrasolar planets by humans. In 1995, Mayor & Queloz (1995) added another milestone by discovering the first exoplanet, 51 Pegasi b, orbiting a Sun-like star using the radial velocity method. This discovery heralded a new era in the exploration of extrasolar planets. Over the years, with technological advancements and the maturation of theoretical frameworks, a variety of methods for detecting exoplanets have emerged, including the radial velocity method, the transit method, astrometry, microlensing, direct imaging, and pulsar timing. Among these, the transit method has proven to be one of the most successful (Charbonneau et al. 1999, Naef et al. 2001), with 4307 exoplanets confirmed through this technique, accounting for 74.4% of all certified exoplanets as of 2024 November 6. At that time, a total of 5787 exoplanets had been confirmed.

The transit method determines the presence of planets around a host star by monitoring the brightness of the stellar system, which decreases slightly when a planet passes between its host star and the observer, and studying the change in brightness. HD 209458 b (Howell et al. 2014) was the first exoplanet to be discovered using this method. However, the primary technical advance that made the transit method the primary method of discovering exoplanets was NASA's Kepler space telescope mission (Borucki et al. 2010). The Kepler mission identified nearly 4000 exoplanet candidates, of which more than 2600 have already been confirmed as exoplanets. The follow-up to the Kepler mission, known as K2 (Howell et al. 2014), discovered nearly 500 exoplanets. The successor to the Kepler mission is NASA's Transiting Exoplanet Survey Satellite (TESS), which launched in 2018 April. TESS is designed to detect exoplanets around millions of stars by observing their transit star signals. It rotates through 26 sky sectors, each observed for about 27 days. The TESS mission has confirmed more than 500 exoplanets, with more than 7200 additional planetary candidates awaiting confirmation.

An increasing number of space telescopes has generated a vast amount of light curve data. With the advent of the Big Data era, automated processing of data and identification of exoplanet transit signals have become data-science-driven tasks. In recent years, early tree-based models such as Robovetter (Coughlin et al. 2016) and the Autovetter project (McCauliff et al. 2015) were used to certify exoplanets. These initiatives explored the potential of machine learning to replace

manual certification of exoplanet candidates and demonstrated that machine learning could accelerate the exoplanet certification process. However, these tree models, while relatively simple, only transformed the manual review process into a decision tree process. Although tree models offer good accuracy, they are computationally intensive and have relatively low recall rates. In recent years, machine learning methods for screening exoplanet candidates against Kepler light curves have included k-nearest neighbors (Thompson et al. 2015), random forests (Mislis et al. 2016), self-organizing maps (Armstrong et al. 2017), semi-supervised and unsupervised learning with the ThetaRay algorithm (Ofman et al. 2022), and feature engineering combined with classifiers (Hinners et al. 2018; Malik et al. 2022). Machine learning methods typically offer the advantages of shorter model training times, the ability to extract features from light curves through feature engineering, and enhanced model interpretability. However, these models are often less accurate and have lower recall rates, particularly when discriminating between exoplanets and eclipsing binaries.

Recently, deep learning has emerged as a popular approach in the field of exoplanet exploration, with convolutional neural networks (CNNs) achieving significant success in image recognition and classification. In 2018, Pearson et al. (2018) first proposed a one-dimensional Convolutional Neural Network (1D-CNN) capable of automatically extracting a broad range of planetary features from light curves, eliminating the need for reliance on previous specific classification metrics. This study compared the performance of multiple classifiers, including wavelet multilayer perceptrons, support vector machines, and BLS algorithms, and found that 1D-CNN achieved the best performance. Shallue & Vanderburg (2018) constructed a CNN called AstroNet, which was applied to Kepler data for the first time and successfully identified exoplanet candidates. The network selects exoplanet candidates by extracting features from both global and local views of the light curve. Two new planets were statistically validated: one being part of a resonance chain of five planets around Kepler 80, and the other is the eighth planet around Kepler 90. Ansdell et al. (2018) further improved the model by incorporating additional information and parameters, including center-of-mass time series data and key stellar parameters such as stellar mass, stellar density, surface acceleration, and metal abundance. They also used data augmentation techniques to reduce the risk of model overfitting, significantly enhancing the model's performance. Dattilo et al. (2019) used an improved model based on AstroNet and applied it to mission data from K2, confirming two previously unknown exoplanets. Osborn et al. (2020) made the first attempt to train AstroNet using TESS simulation data, which have different system characteristics compared to real TESS data. Consequently, the model's performance degraded when applied to real TESS data, with a recall rate of 61%. Yu et al. (2019) improved the network structure of AstroNet by adding stellar depth information and a secondary occultation view. This enhanced model was the first to be trained with real data from TESS. It was also the first model to divide the candidate screening process into two modes: classification and screening. In the classification mode, the model distinguishes lingering starlight signals, including planetary candidates and eclipsing binaries, from others such as stellar variability and instrumental noise, with an average precision of 97% and an accuracy of 97.4%. In the screening mode, it recognizes only planetary candidates with an average precision of 69.3% and an accuracy of 97.8%.

The theory of attention mechanisms is one of the most pivotal concepts in deep learning, originally emerging from the domain of natural language processing (NLP). During information processing, humans selectively focus on certain pieces of information while disregarding other available data. This selective focus is commonly known as the attention mechanism. In NLP, Bahdanau (2014) employed the attention mechanism to enhance the encoder-decoder sequence modeling architecture in translation systems. The attention mechanism and its variants have applications not only in NLP but also in computer vision (Alexey 2020; Li et al. 2020; Wu et al. 2020), language processing (Dong et al. 2018; Karita et al. (2019), and various other deep learning domains (Huang et al. 2019; Liu et al. 2024). In astronomy, light curves are utilized as one-dimensional time series, where each observation in the series is correlated with preceding and subsequent observations. Salinas et al. (2023) introduced a novel Transformer-based automatic classification model that leverages the self-attention mechanism to capture crucial features within transiting signals and stellar parameters.

In this paper, we enhance the AstroNet model by incorporating an attention mechanism and refining the structure of the residual network, to develop a classification model capable of rapid, efficient, and accurate categorization of light curve data into candidate exoplanets and other classes. The structure of this paper is as follows: Section 2 details the model architecture, the channel attention mechanism, and the training strategy; Section 3 describes the data sets, including both the Kepler and TESS data sets; Section 4 provides a detailed discussion of the experimental procedure and results; and Section 5 offers a discussion, conclusions, and suggestions for future research.

2. SE-CNN-RlNet

Our model, SE-CNN-RlNet, is an extension of the AstroNet model, a CNN model developed by Shallue & Vanderburg (2018). As depicted in Figure 1, SE-CNN-RlNet encompasses two distinct branches, each comprising a convolutional layer structure and a channel attention layer structure (Hu et al. 2018). The CNN structure within each branch excels at extracting spatial features from the input data, detecting feature information from both global and local perspectives, and employing the

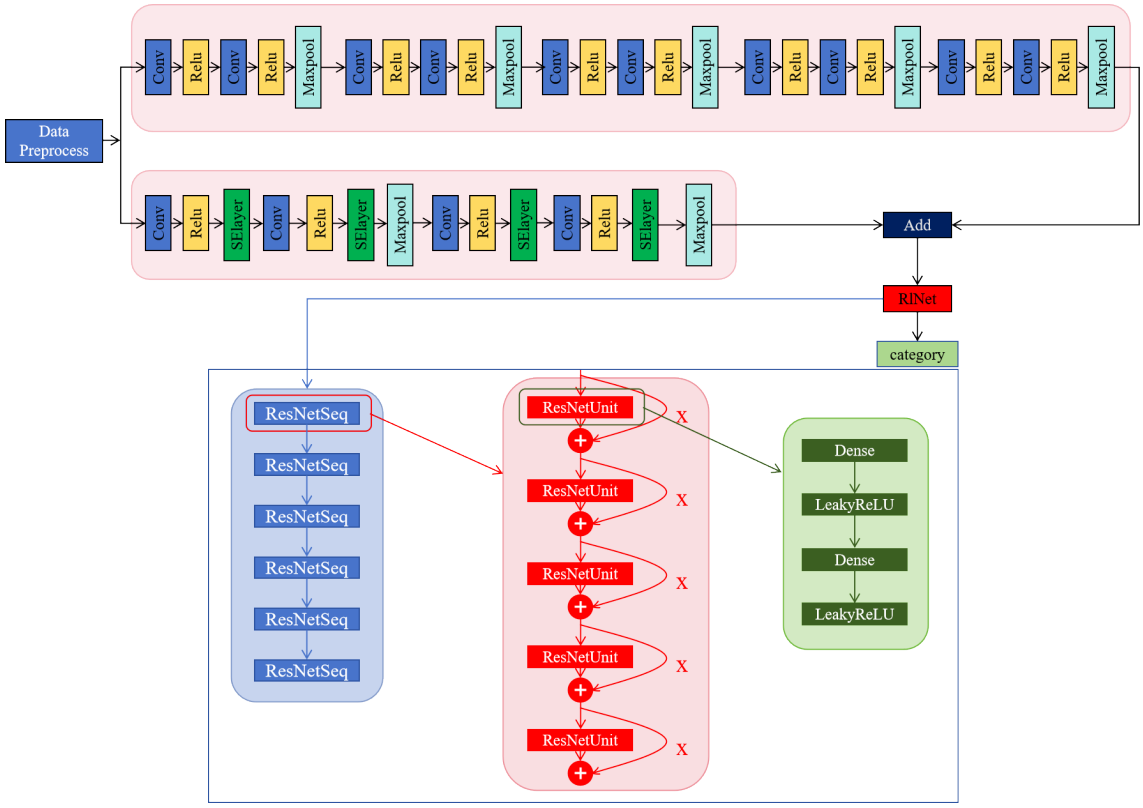


Figure 1. The specific structure diagram of the SE-CNN-RlNet classification network. The blue box indicates a residual-like network structure.

channel attention mechanism to weight the feature information accordingly. The weighted feature information is then learned and classified by a residual-like network structure, culminating in the network's prediction output. Section 2.1 provides a detailed description of the channel attention mechanism. Our model leverages the strengths of CNNs, effectively capturing spatial information from global and local views, and integrates the channel attention mechanism to fully exploit feature importance. By processing information from different perspectives in parallel, the model aggregates the information from both views, thereby enhancing its ability to learn and model complex patterns. The convolutional network part consists of two different branches. One branch is designed to extract feature information from the global view of the light curve, structured with convolutional and maximum pooling layers. Due to the substantial amount of stellar information in the global view, this branch has a greater number of convolutional layers. The maximum pooling layer reduces the feature dimensions of the convolutional layer's output, thereby reducing the network's parameters and computational cost. This also enhances the robustness of the extracted features and effectively prevents overfitting. The other branch is tasked with extracting feature information with a weighted emphasis on the local view of the light curve. This branch's structure includes a convolutional layer, a channel attention layer, and a maximum pooling layer. The convolutional layer has fewer branches compared to the global view branch because the local view contains less stellar-related information. The addition of the channel attention mechanism module weights the extracted features to obtain weighted feature information, allowing the network to focus more on the filtered important information during training, which improves the model's generalization ability. The features from the two branches are then fused and passed through a network structure akin to a residual network. We enhanced the original fully connected neural network of AstroNet into a residual-like network structure, as shown in the blue box in Figure 1. This part of the structure comprises six ResNetSeq layers, each containing five ResNetUnit units. Each ResNetUnit unit is a small residual structure, which differs from the traditional residual network (ResNet) by adopting a fully connected network form within the ResNetUnit units and utilizing the LeakyReLU activation function. We employ this residual structure to facilitate network training, address the degradation of network performance, and, in conjunction with the LeakyReLU activation function, mitigate the effects of gradient vanishing, significantly enhancing training speed (He et al. 2016).

2.1. Channel Attention Mechanisms

The theory of the attention mechanism is one of the most prominent ideas in the field of deep learning. The essence of the attention mechanism lies in the fact that when humans focus on a particular object, they allocate more cognitive resources to the area of interest to gather more detailed information, while reducing the intake of information from other areas. This ability enables humans to effectively allocate their limited attention resources and rapidly identify the most relevant information from a vast array of data.

The channel attention mechanism, a subset of attention mechanisms, is a technique used to enhance the performance of CNNs. It improves model performance by dynamically assigning weights to different channels within a feature map, granting higher weights to channels that capture important information, while assigning lower weights to those that do not. This weighting strategy enables the model to focus more on the critical information, thereby enhancing its performance.

The fundamental structure of the channel attention mechanism is depicted in Figure 2(a). For an arbitrary feature transformation F_{tr}, the process of assigning feature weights is executed by constructing an SE block. Initially, feature U undergoes a Squeeze operation, which is a global average pooling operation that condenses each channel's feature map into a single value, resulting in a distinct scalar representation for each channel. Subsequently, the Excitation operation is carried out (as detailed in Figure 2(b)), which involves a small fully connected network composed of nonlinear activation functions–typically two fully connected layers with one ReLU layer in between. The channel weights are learned through this network, and the output weights are subsequently normalized using a Sigmoid function to ensure each weight is normalized between 0 and 1. Ultimately, the Scale operation multiplies the learned attentional weights by the corresponding channels to the original feature maps, yielding weighted feature maps.

2.2. Training Strategies

A common challenge with complex neural network models is overfitting. This occurs when the model's overly complex structure learns the noise and details in the training set too well, leading to a model that performs well on the training set but poorly on the test set. To address the overfitting problem, the model in this paper employs the following three training strategies:

1. The traditional activation function ReLU is replaced with LeakyReLU (Maas et al. 2013), as shown in Equation (1), within the residual-like network. LeakyReLU retains some values on the negative axis, unlike ReLU, which alleviates the issue of neuron parameters not being updated due to too small activation values. Here, we set the Alpha parameter to 0.1.

$$\mathrm{LeakyReLU}(x) = \begin{cases} x, & \text{if } x \geqslant 0. \\ \alpha \times x, & \text{otherwise.} \end{cases} \quad (1)$$

2. Early stopping (Bolukbasi et al. 2017) is a widely used training method to prevent overfitting. Training is halted when the loss on the training set nearly ceases to decrease and remains relatively stable. We set the criterion for stopping training as follows: if the loss does not decrease within 10 consecutive epochs, training is terminated.
3. In this paper, we utilize the Adam optimizer (Diederik 2014) and the StepLR learning rate scheduler (Smith 2017) to dynamically adjust the learning rate during the training process. Initially, a larger learning rate is employed for training, and as the number of iterations increases, the learning rate is gradually reduced to ensure that the loss variation is stable and closer to the optimal solution. We set the adjusted learning rate step size to 50 and the multiplier factor for learning rate adjustment to 0.5.

3. Data

Each light curve represents a one-dimensional time series of time and flux, stored in a corresponding FITS file. The light curves are corrected to eliminate the effects of long-term brightness variations due to stellar and instrumental factors. Subsequently, the orbital period, transit time, and transit duration of the system are determined by phase-folding the light curves to identify the periodic transit signals (Tenenbaum et al. 2012). The outcomes of this process are referred to as Threshold Crossing Events (TCEs). TCEs can result from exoplanet transits or false positive signals, which primarily include eclipsing binaries, instrumental noise, and stellar variability, among others.

When employing a neural network model for classification, a training set with labels is required. Therefore, in this paper, we utilize a table of TCEs and light curves, cross-matching the two to obtain a labeled training set. We have incorporated both Kepler and TESS data sets to validate the model's adaptability. The following section provides a detailed description of these two data sets.

3.1. Kepler Data Set

We utilized the Kepler data set provided by Ansdell et al. (2018), which encompasses the Q1-Q17 Kepler data release from the DR24 light curve of the Mikulski Archive for Space Telescopes, as well as various stellar parameters from the updated Kepler DR25 catalog by Mathur et al. (2017). These parameters include the effective temperature (T_{eff}), surface

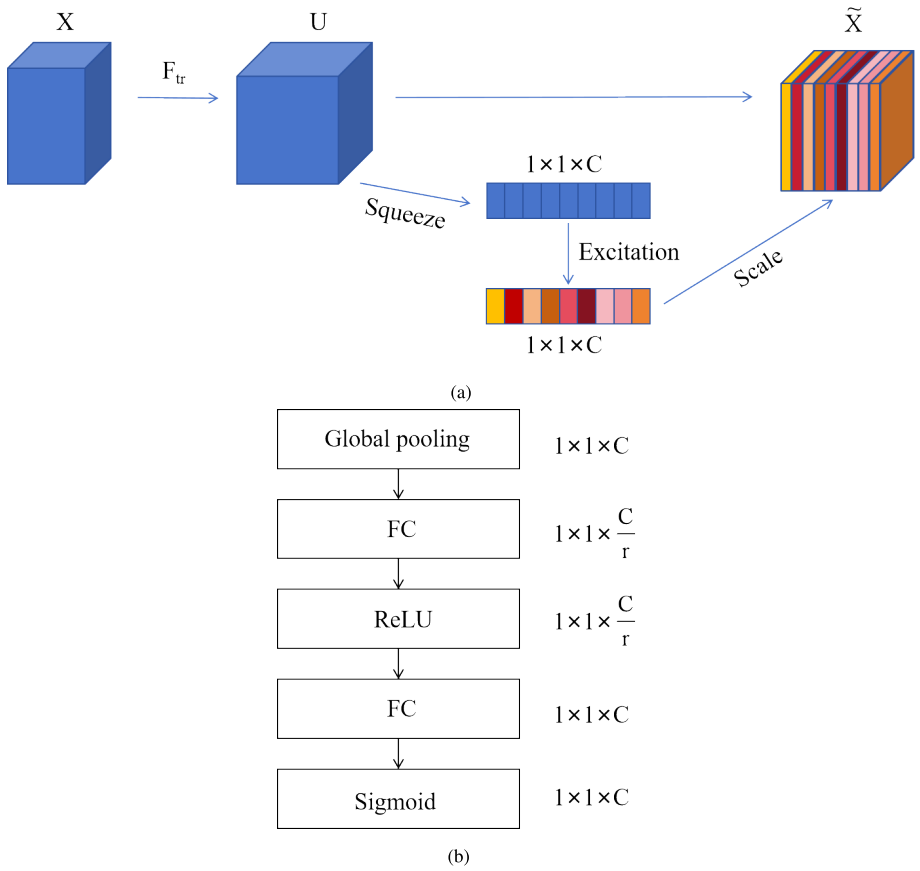


Figure 2. (a) A Squeeze-and-Excitation block (b) Excitation.

gravity (log g), metal abundance ([Fe/H]), radius, mass, and density. The light curves were produced by the Kepler Science Processing Pipeline (Jenkins et al. (2010), which initially calibrated time-series Target Pixel File (TPF) images and subsequently carried out fixed-aperture photometry. The Kepler light curves span up to 4 yr, with each curve comprising approximately 70,000 data points.

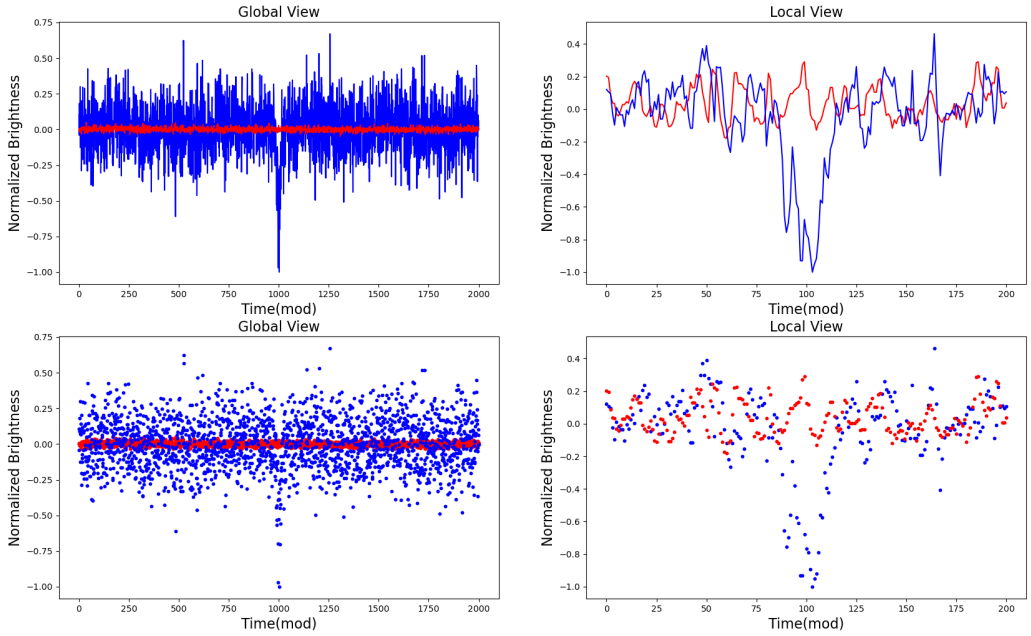
Given that the centroid offers insights into the position of the planet-like signal's source, which is crucial for differentiating a Background Eclipsing Binary (BEB) from exoplanets, we calculate the absolute magnitude of the displacement of the centroid from the TPF's center using the x and y pixel coordinates of the centroid. This displacement is then employed to generate the centroid curve. The global view in both the light curves and centroid curves comprises 2001 data points, while the local view contains 201 data points. Figure 3 illustrates examples of light curves and centroid curves for a confirmed exoplanet and astrophysical false positives in both global and local perspectives.

The labels within the Kepler data set are classified as PC, AFP, NTP, and UNK, representing planets, astrophysical false positives, non-celestial phenomena, and unknown categories, respectively.

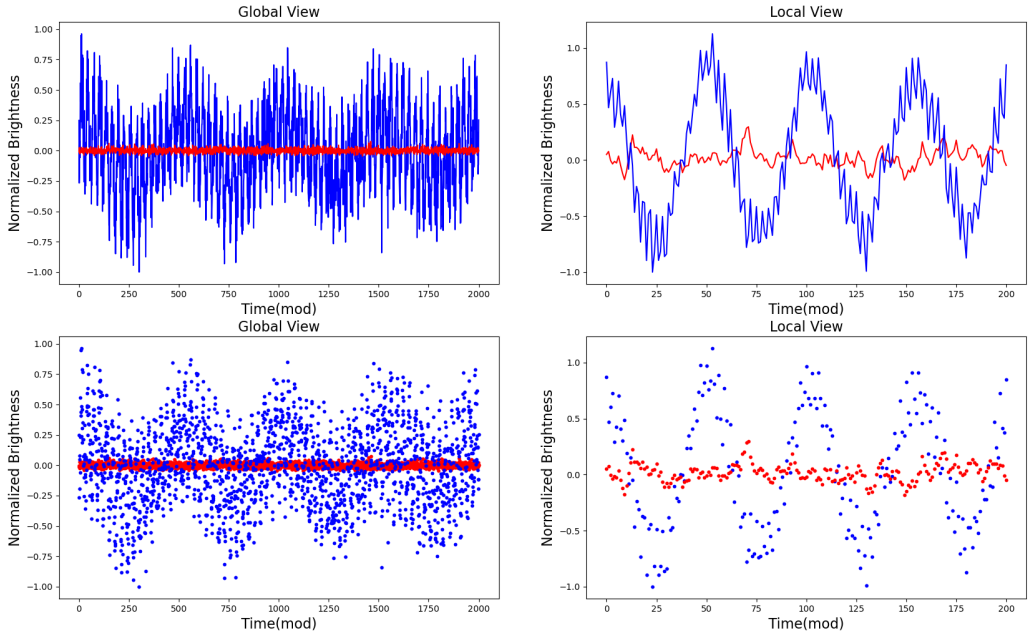
3.2. TESS Data Set

For the TESS data set we use the same data set as Yu et al. (2019), which is part of the TESS sectors 1–5 and part of the sector 6. Kepler's light curves last up to 4 yr and each light curve contains about 70,000 points, while on the other hand, TESS light curves span only about 27 days and have relatively much fewer data points. Therefore, the global view of the phase-folded TESS light curve contains 201 data points and the local view contains 61 data points. Figure 4 gives the light curves in global and local views under each category of TESS.

Unlike the Kepler data set, the TESS data set employs four class labels: Planet Candidates (PC), Eclipsing Binaries (EB), Stellar Variability (V), and Instrumental Noise (IS). For the Kepler data set, we categorized the data into Planet Candidates (PC: 3094 entries) and Non-Planet Candidates (AFP + NTP + UNK: 10,417 entries). In contrast, the TESS data set is divided into Planet Candidates (PC: 391 entries) and Non-Planet Candidates (EB + V + IS: 12,816 entries), as shown in

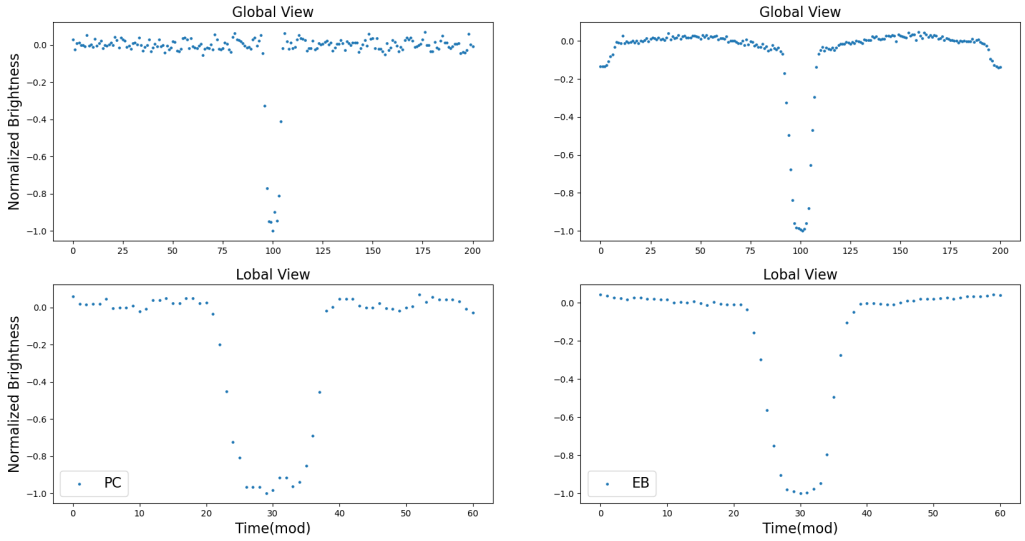


(a)

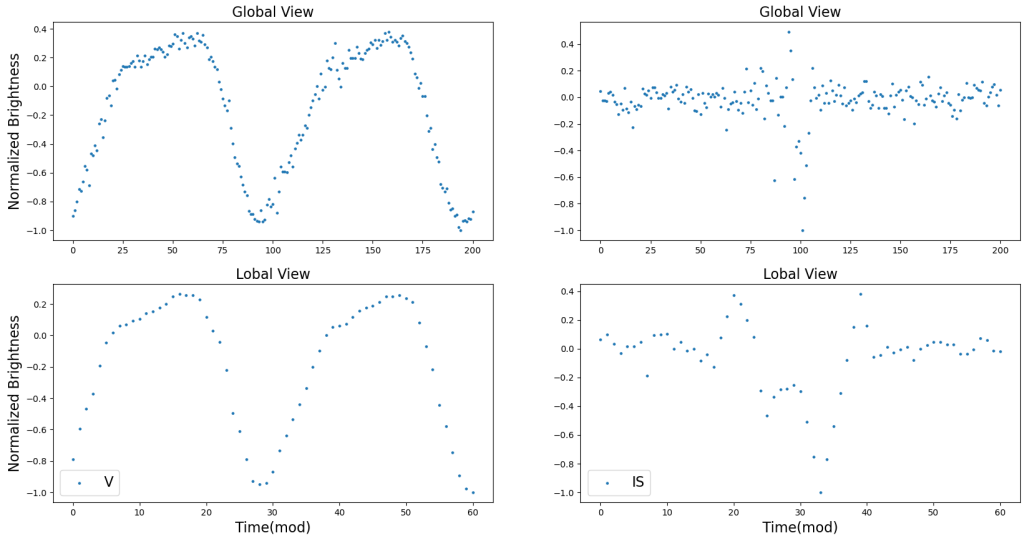


(b)

Figure 3. Global and Local Views of light curves and centroid light curves for Exoplanets and Eclipsing Binaries (EB is a case in AFP) in Kepler Data: Line Graphs and Scatter Plots. The red curve represents the centroid curve, while the blue curve represents the light curve. (a) PC. (b) EB.



(a)



(b)

Figure 4. Global and local views of the various categories of TESS light curves.

Table 1. Both data sets were split into training and testing sets at a ratio of 9:1.

4. Results

Our model, SE-CNN-RlNet, captures crucial features in transit signals through the channel attention mechanism and generates a channel feature heatmap. This heatmap, when combined with the light curve, provides a clearer explanation of the model's classification decisions. Furthermore, we prevent underfitting and overfitting by selecting an appropriate loss function, ensuring that the model loss exhibits a favorable decreasing trend in both the training and test sets. Subsequently, we employed a variety of evaluation metrics to assess our results. We also selected classical machine learning and deep learning models for comparison with our model's outcomes and conducted a brief analysis of the strengths and weaknesses of each model in the given context.

4.1. Channel Characterization Heat Map

Channel feature heatmaps provide an intuitive visualization of the model's feature extraction process from light curve data. In CNNs, channels function as feature detectors, with each channel dedicated to capturing specific spatiotemporal patterns. In this study, the channel feature heatmaps are structured with time steps along the horizontal axis and channel numbers on the vertical axis. The dimensionality of these heatmaps evolves with network depth: time steps progressively shorten as the channel attention layers deepen, while the number of channels increases correspondingly.

The morphology changes of the channel feature heatmaps directly reflect the hierarchical abstraction process of CNNs, from local information to global information and from shallow features to deep features. In this section, we analyze and compare the channel feature heatmaps generated from two distinct light curves: one from an exoplanet and another from a non-exoplanet source. As depicted in Figure 5(a) and (e), within the initial channel attention mechanism layer, the network adeptly identifies transient anomalies within the light curves, such as abrupt luminosity dips at specific time points or stellar flickers. At this juncture, the heatmap displays pronounced activation intensity, maintaining a channel count of 8 and preserving the original time steps. This results in a pattern of scattered luminous spots or vertical luminous streaks, underscoring the network's heightened sensitivity to localized signals. In Figure 5(a), the concentration of high-intensity regions around the mid-time points, coupled with their substantial activation values, is visually striking. A comparative analysis with the light curve reveals a precise alignment with the mid-transit periods, highlighting the heightened significance of features around these moments across the spectrum. Conversely, Figure 5(e) portrays high-intensity regions that are sporadically distributed across various feature channels, yet they are marked by a multitude of activation points. This distribution suggests that the features at different temporal points in this light curve hold considerable importance.

Progressing to Figure 5(b) and (f), as the network advances to the second layer, there is an augmentation in the channel count to 16, a compression of time steps, and a moderation in the activation value range to a medium intensity. Here, the heatmap begins to exhibit inclined luminous bands or contiguous luminous zones, signifying the network's enhanced capability to discern the architecture of local phenomena, such as the gradient variations at the ingress and egress of a transit. In Figure 5(b), the persistence of high-intensity regions in the central area, corresponding to the mid-transit times, is evident, accompanied by the emergence of faint horizontal luminous bands. These bands are indicative of enduring signal characteristics, potentially attributable to the host star's rotation inducing subtle periodic oscillations. In Figure 5(f), the presence of high-activation with horizontally discontinuous luminous bands in channels 4, 8, 9, and 11 suggests that certain features within this light curve are perpetually active throughout the observational period, such as the inherent periodicity and slope variations of the curve.

In the third channel attention layer, depicted in Figure 5(c) and (g), deeper convolutional operations further compress the time steps, increasing the number of channels to 32 and reducing the activation intensity range once more. The heatmap reveals broader yet more discriminative activation regions, indicating the network's capability to capture mid-term signal patterns, such as brightness declines lasting several hours. The channel attention mechanism starts to suppress redundant features, with some channels specializing in interference recognition, such as instrumental noise.

Finally, in the fourth channel attention layer, as shown in Figure 5(d) and (h), the activation intensity range drops to a very low level. Here, the heatmap exhibits sparse but highly specific activation patterns, signifying the network's ability to integrate global features, such as the complete "U-shaped" transit curve.

4.2. Loss

In our SE-CNN-RlNet model, we employ a binary cross-entropy loss (BCE loss) function, as shown in Equation (2). Here, y_i represents the actual label value, which can be either 0 or 1; $\hat{y}_i$ denotes the predicted value, which ranges between 0 and 1; and N is the total number of samples.

$$\text{BCE Loss}(y, \hat{y}) = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)]. \tag{2}$$

Figure 6 shows the loss variation of the SE-CNN-RlNet model during a certain epoch. The experimental results demonstrate a consistent downward trend in loss values across both training and testing data sets as the number of epochs increases, suggesting that the SE-CNN-RlNet model effectively avoids both overfitting and underfitting problems throughout the training process.

4.3. Evaluation Metrics

4.3.1. Five Metrics

There are many model metrics in deep learning to evaluate the model performance. Suitable metrics can accurately evaluate the performance of the model. In this paper, we choose accuracy, precision, recall, F1 score, and Matthews Correlation Coefficient (MCC; Chicco & Jurman 2020) as the basis for evaluating the network's performance. The formulas for these metrics, correspond to the following Equations (3),

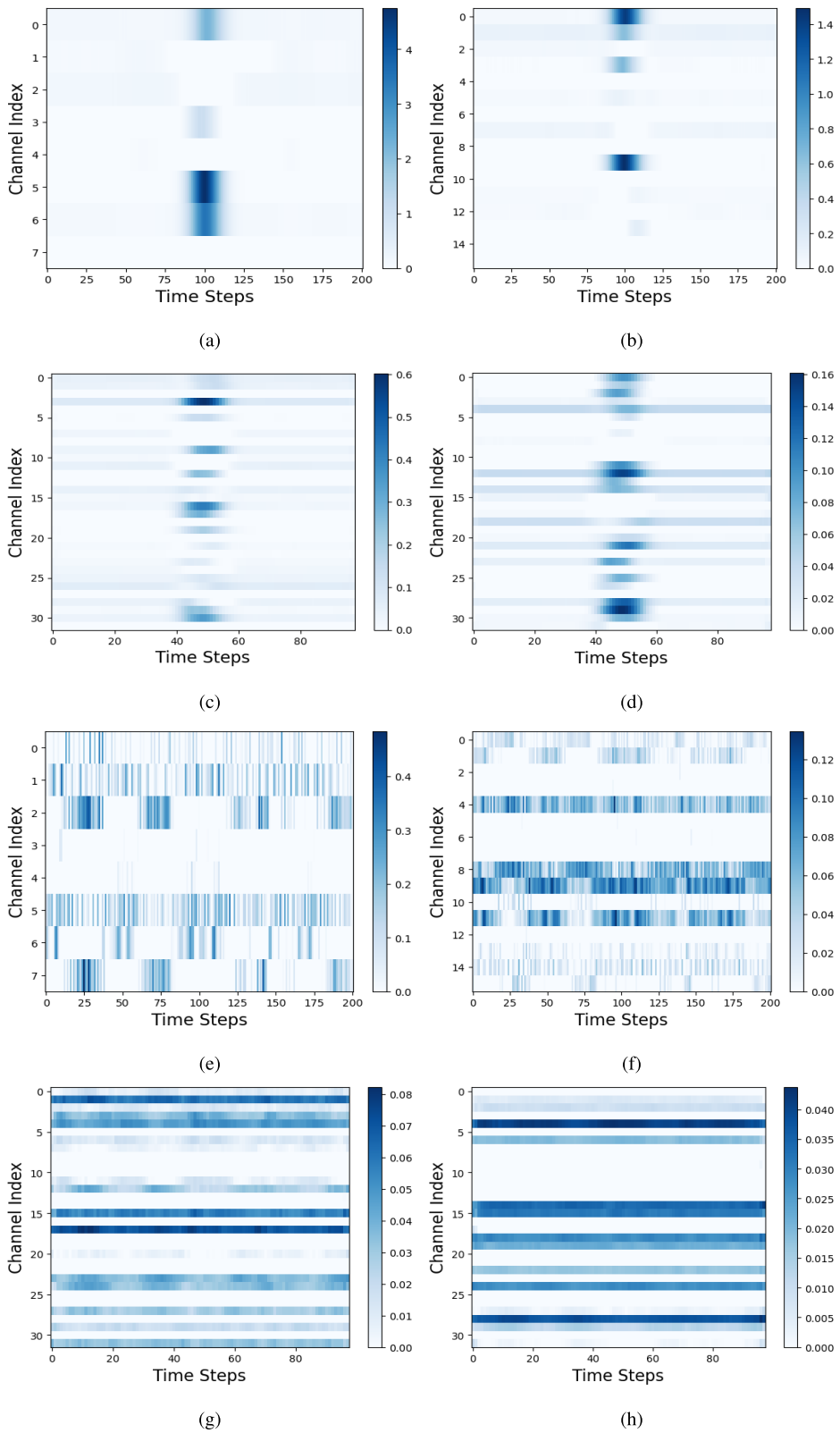
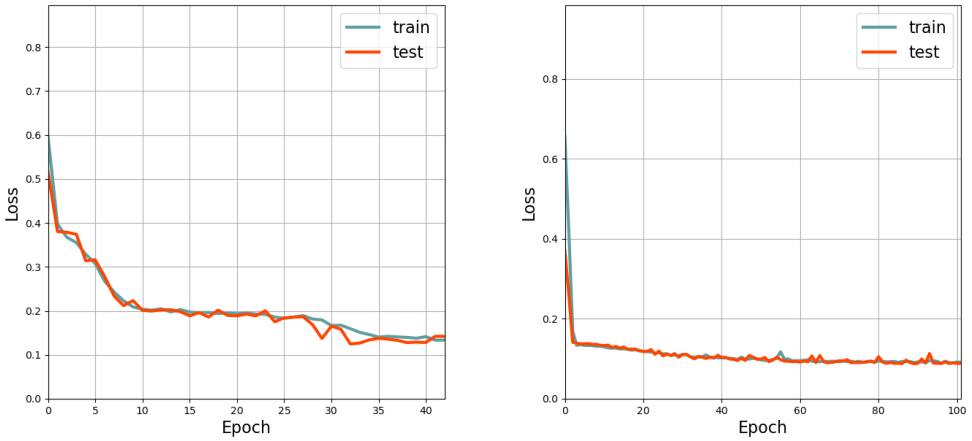


Figure 5. Figure 5 displays the feature heatmaps of the channel attention layers. Panels (a), (b), (c), and (d) represent the channel feature heatmaps for exoplanet data, while panels (e), (f), (g), and (h) correspond to the channel feature heatmaps for false-positive data.



(a) Kepler loss (b) TESS loss

Figure 6. The curve of SE-CNN-RlNet training set and test set loss with training epochs.

Table 1
Number of Kepler and TESS Light Curves

Class	Kepler	TESS
PC	3094	391
Others	10417	12816

(4), (5), (6) and (7), respectively.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{3}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{4}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{5}$$

$$\text{F1} = 2 \cdot \frac{\text{Precison} \cdot \text{Recall}}{\text{Precison} + \text{Recall}}, \tag{6}$$

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}}. \tag{7}$$

In the context of binary classification, we define the following terms:

True Positive (TP). A positive sample that is correctly predicted by the model to be in the positive category.

True Negative (TN). A negative sample that is correctly predicted by the model to be in the negative category.

False Positive (FP). A negative sample that is incorrectly predicted by the model to be in the positive category.

False Negative (FN). A positive sample that is incorrectly predicted by the model to be in the negative category.

Table 2
The Performance of our Best Model on the Kepler and TESS Test Sets is Summarized Across Five Metrics: Accuracy, Precision, Recall, F1 Score, and MCC (with a Threshold of 0.5)

Data Set	Accuracy	Precision	Recall	F1	MCC
Kepler	0.962	0.890	0.950	0.957	0.894
TESS	0.999	0.970	1.000	0.995	0.979

Accuracy is the ratio of the number of correct predictions to the total number of predictions made. Precision is the ratio of the number of correctly predicted positive samples to the total number of samples predicted as positive. Recall is the ratio of the number of correctly predicted positive samples to the total number of samples that are actually positive. The F1 score is a measure of a binary classification model's accuracy, taking into account both the precision and recall of the model, and is essentially a reconciled average of these two metrics. The MCC considers the counts of all four categories in the confusion matrix (true positives, true negatives, false positives, false negatives), measuring the model's true prediction ability and is less affected by sample imbalance. It ranges from -1 to 1, with higher values indicating better model performance.

As presented in Table 2, our model's performance on the Kepler and TESS test sets, in terms of accuracy, precision, recall, F1 score, and MCC index, demonstrates that our model performs well and is adaptable.

4.3.2. Confusion Matrix

The confusion matrix is a standard tool in supervised learning for evaluating the performance of a classification model. It illustrates the relationship between the model's

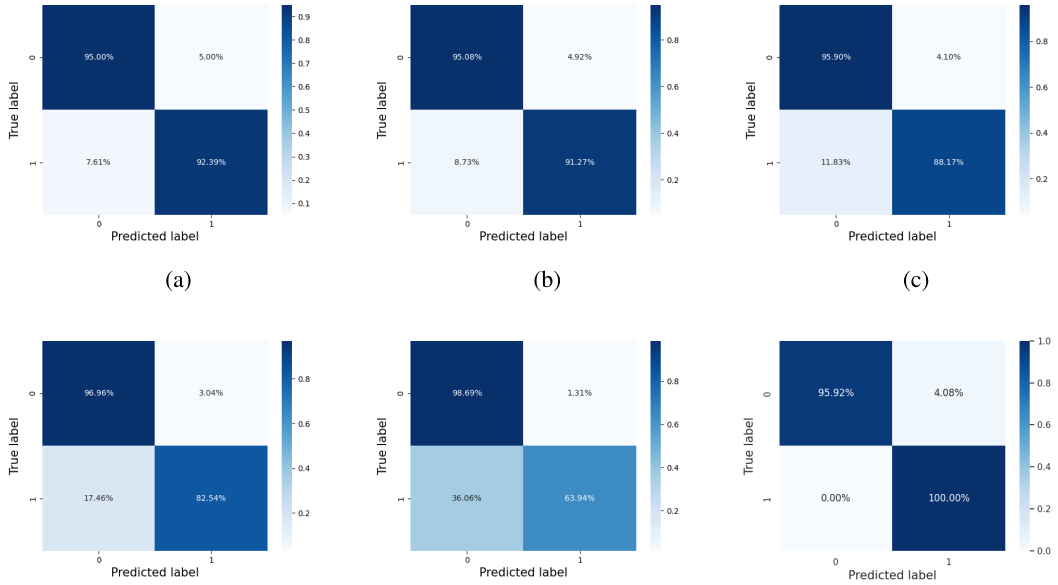


Figure 7. Panels (a)–(e) represent the normalized confusion matrices of the Kepler data set under thresholds ranging from 0.5 to 0.9, while panel (f) represents the normalized confusion matrix of the TESS data set under a threshold of 0.5. Diagonal values approaching 1 indicate better classification performance.

predictions and the actual labels, allowing for the visualization of the model's performance across each category, particularly in terms of the model's classification accuracy and misclassification errors. In the context of binary classification, the confusion matrix is a two-dimensional table where the rows represent the true categories and the columns represent the predicted categories. Figure 7(a)–(e) displays the model's performance for thresholds ranging from 0.5 to 0.9 for the Kepler data set, presenting the normalized confusion matrix. It can be observed from the figure that as the threshold increases, the precision rate increases while the recall relatively decreases; precision and recall are always in a state of dynamic balance. Figure 7(f) shows the normalized confusion matrix for the TESS data set at a threshold value of 0.5. From this, it is evident that the false positive cases (FP) account for only 4.08% of the test set's sample size (which corresponds to just two instances), and the false negative cases constitute 0%. This indicates that the SE-CNN-RlNet model performs exceptionally well on the TESS test set.

4.3.3. PR Curve

The precision–recall (PR) curve is an evaluation metric used to assess the performance of a binary classification model, particularly suitable for data sets where the category distribution is imbalanced. It illustrates the relationship between precision and recall, with recall plotted on the horizontal axis and Precision on the vertical axis. The PR curve represents the model's performance across various thresholds. The model's performance can be quantified by measuring the area under the PR curve (AP), as indicated in Equation (8).

$$\mathrm{AP} = \int_0^1 \mathfrak{p}(r)dr. \tag{8}$$

A larger area under the precision–recall curve (AUC-PR) value signifies better model performance. Figure 8 displays the PR curves for both the Kepler and TESS data sets. It is evident that the PR curves, which incorporate the center-of-mass curves and stellar parameter information, have a larger area under the curve, indicating superior performance. The TESS data set, with an accuracy of nearly 1, also has a larger area under the PR curve compared to the Kepler data set.

Based on the outcomes of the aforementioned evaluation metrics, it is evident that the SE-CNN-RlNet model performs exceptionally well on both the Kepler and TESS data sets. The model not only exhibits strong generalization capabilities but also boasts rapid operational speed and outstanding ability to handle imbalanced data sets.

4.4. Comparison Experiment

Accuracy is frequently employed as a measure of classifier performance in machine learning. However, when it comes to unbalanced data sets, the F1 score and recall provide more insightful information about model performance (Davis & Goadrich 2006). While maintaining high accuracy, our goal is to achieve a higher recall, thereby enhancing the model's

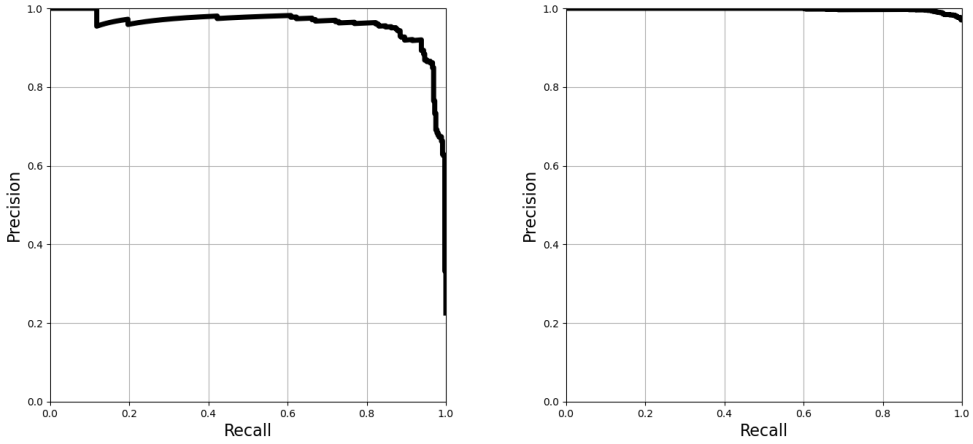


Figure 8. The PR curves for the Kepler and TESS test sets demonstrate the trade-off between precision and recall under different decision thresholds. Comparatively, the TESS data set exhibits a larger AP area in the SE-CNN-RlNet model, indicating better model performance.

Table 3
Comparison of Accuracy, Recall, and F1 Score Results for 12 Models

Method	Accuracy	Recall	F1
FATS+Naïve Bayes (Hinners et al. 2018)	50.5%	98.0%	0.219
FATS+KNN (Hinners et al. 2018)	62.0%	38.1%	0.413
FATS+SVM (Hinners et al. 2018)	65.8%	51.7%	0.371
FATS+D-tree (Hinners et al. 2018)	69.8%	52.4%	0.428
FATS+Random Forest (Hinners et al. 2018)	74.7%	53.0%	0.593
AstroNet (Shallue & Vanderburg 2018)	95.8%	···	0.955
ExoNet (Ansdell et al. 2018)	97.5%	···	0.980
Transformer (Salinas et al. 2023)	96.0%	96.0%	0.960
AstroNet-Triage (Yu et al. 2019)	97.4%	···	0.970
AstroNet-Vetting (Yu et al. 2019)	97.8%	···	0.693
Ours-Kepler	96.2%	95.0%	0.957
Ours-TESS	99.9%	100.0%	0.995

performance and aiding in the discovery of more exoplanet candidates.

In this paper, we compare the SE-CNN-RlNet model with several classical classification models, including Feature Engineering (FATS—Feature Engineering for Time Series) coupled with classifiers such as Bayesian estimation, k-nearest neighbors, Support Vector Machines, Decision Trees, and Random Forests (Hinners et al. 2018), as well as AstroNet (Shallue & Vanderburg 2018), ExoNet (Ansdell et al. 2018), Transformer (Salinas et al. 2023), and AstroNet-Triage and AstroNet-Vetting (Yu et al. 2019). Table 3 presents the accuracy, recall, and F1 scores of these 12 models, demonstrating that our model yields the best results when applied to TESS data, with even more pronounced improvements when applied to Kepler data.

From the above table, it is evident that machine learning models are less effective compared to deep learning models. This observation can be attributed to two primary reasons: First, from a data perspective, light curves contain substantial noise and sparsity, which complicates the extraction of useful temporal features for classification through feature engineering. Additionally, the imbalance in the data, with a small number of planetary instances, makes it challenging for machine learning models to adequately learn the features of both planets and non-planetary objects. The actual astronomical data sets are also too complex for machine learning models to adapt to the specific distribution and intricacies of the data. Second, in terms of feature engineering, time-series feature engineering techniques such as FATS may not be suitable for light curves.

Among the deep learning models applied to the Kepler data set, AstroNet and ExoNet are comparable to our model in terms of accuracy and average precision. However, the AstroNet and ExoNet models require over two hours to train on our experimental equipment, whereas the SE-CNN-RlNet model takes only 0.3–0.6 hr to train. This demonstrates that a network with a residual-like network structure can significantly reduce training time while maintaining high accuracy and average precision. As for models applied to the TESS data set, the lower average accuracy of AstroNet-Vetting may be due to the similarity in shape between the light curves of exoplanets and eclipsing binaries, making it difficult for the models to distinguish between the features of exoplanets and eclipsing binaries. In contrast, the SE-CNN-RlNet model performs exceptionally well, primarily due to the channel attention mechanism that assigns weights to all the local features produced by the convolutional network. This enables the model to better differentiate between various classes of features and enhances the model's generalization ability.

5. Discussion

In this paper, we construct an SE-CNN-RlNet classification model that employs one-dimensional convolution to extract

features and subsequently classifies light curves using a residual-like network. The model's innovations are twofold: first, it enhances the model's generalization ability by assigning weights to features through the channel attention mechanism, and second, it transforms the fully connected network structure to the residual-like structure of ResNet, significantly reducing the model's training time.

The SE-CNN-RlNet model is trained and compared against 10 classical models. Experiments indicate that the SE-CNN-RlNet model's classification accuracy and F1 score on the Kepler data set are comparable to those of AstroNet, with a substantial improvement in training speed. Despite the absence of center-of-mass curves and stellar parameter data in the TESS data set, our model maintains high accuracy and F1 scores, exceeding those of the Transformer model by 3%–4%.

The attention mechanism has a proven track record of success in NLP applications. By applying the channel attention mechanism, we can determine the importance of features filtered by the convolutional layer, leading to a clearer understanding and better generalization of the model. However, the specific significance of the extracted features within light curves warrants further in-depth exploration. The concept of residual-like networks can be extended to other astronomical classification tasks, potentially reducing training time while ensuring accuracy. Our innovations can be applied to other astronomical missions, such as the CHES mission (Ji et al. 2022; Bao et al. 2024). The CHES mission aims to detect exoplanets using the astrometry method.

Our residual-like network can be utilized in a variety of data processing tasks. It can significantly shorten the data processing cycle, thereby accelerating the speed of candidate selection through astrometry. If the data of the CHES mission contains higher noise levels, shorter observation periods, and multi-band information, the channel attention mechanism can dynamically weight key features, enhancing the model's robustness to complex signals.

The diversity of data sets impacts the effectiveness of model training. Thus, future work should focus not only on model improvements but also on constructing balanced and diverse data sets.

In conclusion, our model demonstrates strong performance on both the Kepler and TESS real data sets, with fast and adaptable training. The channel attention mechanism and residual-like network structure can be applied to other astronomical classification tasks.

Acknowledgments

We express our gratitude to the National Natural Science Foundation of China (NSFC, grant No. 11803020) and the Astronomy Joint Fund Program (grant No. U2031145).

ORCID iDs

Ying Wang https://orcid.org/0000-0003-0506-054X

References

Alexey, D. 2020, arXiv:2010.11929
Ansdell, M., Ioannou, Y., Osborn, H. P., et al. 2018, ApJL, 869, L7
Armstrong, D. J., Pollacco, D., & Santerne, A. 2017, MNRAS, 465, 2634
Bahdanau, D. 2014, arXiv:1409.0473
Bao, C., Ji, J., Tan, D., et al. 2024, AJ, 167, 286
Bolukbasi, T., Wang, J., Dekel, O., & Saligrama, V. 2017, in Proc. of the 34th Int. Conf. on Machine Learning, 70, ed. P. Doina & T. Yee Whye (Sydney: PMLR), 527
Borucki, W. J., Koch, D., Basri, G., et al. 2010, Sci, 327, 977
Charbonneau, D., Brown, T. M., Latham, D. W., & Mayor, M. 1999, ApJL, 529, L45
Chicco, D., & Jurman, G. 2020, BMCG, 21, 1
Coughlin, J. L., Mullally, F., Thompson, S. E., et al. 2016, ApJS, 224, 12
Dattilo, A., Vanderburg, A., Shallue, C. J., et al. 2019, AJ, 157, 169
Davis, J., & Goadrich, M. 2006, in Proc. 23rd Int. Conf. on Machine Learning (New York: ACM), 233
Diederik, P. K. 2014, in Int. Conf. on Learning Representations (ICLR) *(San Diego, 7-9 May 2015)*
Dong, L., Xu, S., & Xu, B. 2018, in IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP) (Piscataway, NJ: IEEE), 5884
He, K., Zhang, X., Ren, S., & Sun, J. 2016, in Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (Las Vegas, NV: IEEE), 770
Hinners, T. A., Tat, K., & Thorp, R. 2018, AJ, 156, 7
Howell, S. B., Sobeck, C., Haas, M., et al. 2014, PASP, 126, 398
Hu, J., Shen, L., & Sun, G. 2018, in Proc. IEEE Conf. on Computer Vision and Pattern Recognition (Salt Lake City, UT: IEEE), 7132
Huang, C.-Z. A., Vaswani, A., Uszkoreit, J., et al. 2019, in 7th Int. Conf. on Learning Representations *(New Orleans, LA, USA)* (OpenReview.net) https://openreview.net/forum?id=rJe4ShAcF7
Jenkins, J. M., Caldwell, D. A., Chandrasekaran, H., et al. 2010, ApJL, 713, L87
Ji, J.-H., Li, H.-T., Zhang, J.-B., et al. 2022, RAA, 22, 072003
Karita, S., Chen, N., Hayashi, T., et al. 2019, in 2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU) (Piscataway, NJ: IEEE), 449
Li, Y., Rao, S., Solares, J. R. A., et al. 2020, NatSR, 10, 7155
Liu, Y., Zhang, Y., Wang, Y., et al. 2024, ITNNL, 35, 7478
Maas, A. L., Hannun, A. Y., Ng, A. Y., et al. 2013, in Proc. of the 30th Int. Conf. on Machine Learning, 28 (Atlanta, GA: PMLR), 3, https://ai.stanford.edu/~amaas/papers/relu_hybrid_icml2013_final.pdf
Malik, A., Moster, B. P., & Obermeier, C. 2022, MNRAS, 513, 5505
Mathur, S., Huber, D., Batalha, N. M., et al. 2017, ApJS, 229, 30
Mayor, M., & Queloz, D. 1995, Natur, 378, 355
McCauliff, S. D., Jenkins, J. M., Catanzarite, J., et al. 2015, ApJ, 806, 6
Mislis, D., Bachelet, E., Alsubai, K., Bramich, D., & Parley, N. 2016, MNRAS, 455, 626
Naef, D., Latham, D., Mayor, M., et al. 2001, A&A, 375, L27
Ofman, L., Averbuch, A., Shliselberg, A., et al. 2022, NewA, 91, 101693
Osborn, H. P., Ansdell, M., Ioannou, Y., et al. 2020, A&A, 633, A53
Pearson, K. A., Palafox, L., & Griffith, C. A. 2018, MNRAS, 474, 478
Salinas, H., Pichara, K., Brahm, R., Pérez-Galarce, F., & Mery, D. 2023, MNRAS, 522, 3201
Shallue, C. J., & Vanderburg, A. 2018, AJ, 155, 94
Smith, L. N. 2017, in IEEE Winter Conf. on Applications of Computer Vision (WACV) (Piscataway, NJ: IEEE), 464
Tenenbaum, P., Christiansen, J. L., Jenkins, J. M., et al. 2012, ApJS, 199, 24
Thompson, S. E., Mullally, F., Coughlin, J., et al. 2015, ApJ, 812, 46
Wolszczan, A., & Frail, D. A. 1992, Natur, 355, 145
Wu, B., Xu, C., Dai, X., et al. 2020, arXiv:2006.03677
Yu, L., Vanderburg, A., Huang, C., et al. 2019, AJ, 158, 25