CN 10-1206/TH ISSN 2095-7521

Instrumentation

Instrumentation

Competent Authority: China Association for Science and Technology
Sponsor: China Instrument and Control Society
Popular Science Press
Publisher: Popular Science Press
Editor: Editorial Office of Instrumentation
Director: Wu Youhua
Printer: Beijing Beilin Printing House
Address: No.79 Beiheyan Street, Dongcheng District, Beijing, China
Post Code: 100009
Tel: +86-010-53389229
E-mail: instrumentation@cis.org.cn

中文刊名：仪器仪表学报（英文版）
国内定价：100元/册

ISSN 2095-7521

Volume 8 Number 4 December 2021

Volume 8 Number 4
December 2021

04

Publication of China Instrument & Control Society

INSTRUMENTATION

Volume 8 Number 4 **CONTENTS** **December, 2021**

Instrumentation

Journal Scope:

Instrumentation is a high quality open access peer reviewed research journal. Authors are solicited to contribute to these journals by submitting articles that illustrate most up-to-date research results, projects, surveying works and industrial experiences that describe significant advances in the instrumental science. The mission of the Instrumentation is to provide a platform for the researchers, academicians, professors, practitioners and students from all over the world to develop and propose high quality ideas and to impart and share knowledge. The goal of the journal is to host theoretical and practical aspects and reports on experimental activities concerning instrumentation and introduce new and emerging areas of these fields.

Frequency: 4 issues
ISSN: 2095-7521
CN: 10-1206/TH

Competent organization:
China Association for Science and Technology

Organizer:
China Instrument and Control Society

Instrumentation covers:

Measurement and Instrument Fundamentals,Method and Application
Sensor and Sensor System
Signal Collection and Processing
Micro/Nano measurement
Precision Measurement
Optical and Laser Test
Spectral Measurement

Readership:

Instrumentation is essential reading for researchers, engineers and educators who want to keep up-to-date with the latest advances in the field of instrumentation.

Contact us:

Tel: +86-010-53389229
E-mail: instrumentation@cis.org.cn

Get full papers :

Instrumentation is available online, as well as in print.
Website:http://www.instrumentationjournal.com

Instrument

Detection and Analysis of Strong Ionospheric Scintillation in Equatorial Region

ZHU Xuefen, LIN Mengying, LU Zhengpeng, CHEN Xiyuan

(*Key Laboratory of Micro-Inertial Instrument and Advanced Navigation Technology of Ministry of Education*, *School of Instrument Science and Engineering*, *Southeast University*, *Nanjing* 210096, *China*)

Abstract: To detect the occurrence of ionospheric scintillation in the equatorial region, a coherent/non-coherent integration method is adopted on the accumulation of intermediate frequency (IF) signal and local code, in the process of signal acquisition based on software receiver. The processes of polynomial fitting and sixth-order Butterworth filtering are introduced to detrend the tracking results. Combining with ionospheric scintillation detection algorithm and preset thresholds, signal acquisition and tracking, scintillation detection, positioning solution are realized under the influence of strong ionospheric scintillation. Under the condition that the preset threshold of amplitude and carrier phase scintillation indices are set to 0.5 and 0.15, and the percentage of scintillation occurrence is 50%, respectively, PRN 12 and 31 affected by strong amplitude scintillation are detected effectively. Results show that the positioning errors in the horizontal direction are below 5m approximately. The software receiver holds performances of accurate acquisition, tracking and positioning on the strong ionospheric scintillation conditions, which can provide important basis and helpful guidance for relevant research on ionospheric scintillation, space weather and receiver design with high performance.

Keywords: Ionospheric Scintillation, Acquisition and Tracking; Positioning Error, Coherent/Non-coherent Integration, Detrending

1 Introduction

With the development and application of GPS and navigation technology, it is increasingly important to improve the security and reliability of satellite navigation systems. The observation and study of ionospheric scintillation, as a medium for signal propagation, which can produce a negative impact on the quality of the received global navigation satellite systems (GNSS) signal [1], has become one of the important contents and hotspots in space weather. In the ionosphere which is about 60 ~ 1000 kilometers from the ground, the atmosphere interacts with high-energy particles from the sun, making atmospheric molecules and atoms charged particles with uneven density distribution, and the ionospheric scintillation is a phenomenon that the amplitude and phase of signals fluctuate randomly when signals propagate through the plasma irregularly [2-3]. There are many reasons for this phenomenon, including magnetic storms, solar activity, local electric field, electrical conductivity, etc. It is also closely related to the geographical area, season, local time, geomagnetic activity [4], etc.

Ionospheric scintillation occurs mainly in the equatorial and polar regions [5]. The scintillations in equatorial region are mainly amplitude scintillation and in the polar region are mainly phase scintillation.

The frequency and intensity of ionospheric scintillation will increase in solar maximum year. Strong ionospheric scintillation will decrease signal quality and even cause signal tracking failure and make the receiver fail in navigation in serious cases [6-8]. Especially in the low latitudes during the solar maximum year, frequent and strong ionospheric scintillations lead to significant attenuation in intensity and carrier-to-noise ratio of signal [9-10]. Due to the suddenness of ionospheric scintillation, frequent and severe ionospheric scintillation makes detection and satellite signal acquisition and tracking more difficult, especially in the equatorial regions.

In this paper, the software receiver is used to detect and analyze the strong ionospheric scintillation in the equatorial regions. The results of analysis will provide powerful guidance for designing high-performance tracking algorithms and improving the sensitivity and robustness of satellite receivers. The South China Sea is a low-latitude and equatorial region in China. It is of great significance to design efficient ionospheric scintillation detection and high-performance receiver algorithms for the practical application of the Beidou navigation system and the territorial security of the South China Sea.

2 Ionospheric Scintillation Indices

Frequent strong ionospheric scintillations can easily affect the acquisition and tracking of satellite signals, and even lead to positioning and navigation failures in low-latitude areas in solar maximum year. In order to realize the accurate detection and acquisition of strong ionospheric scintillations in the equatorial regions, the coherent/non-coherent integration algorithm [11] is designed to improve the signal-to-noise ratio of the signal.

Considering the influence of noise signal, the satellite signal is processed by the receiver front-end and quantified by intermediate frequency sampling. Then the GNSS digital signal is correlated with the local code and the local carrier signal, and then two signals of the same phase component I and the quadrature component Q are obtained.

The coherent integration formula $\mathrm{I}_L = \sum_{n=0}^{L-1} \mathrm{I}(nT_s)$ and $\mathrm{Q}_L = \sum_{n=0}^{L-1} \mathrm{Q}(nT_s)$ are used to integrate the two signals respectively. After a series of mathematical operations, the formulas of two signals are as follows:

$$\mathrm{I}_L = \sqrt{2P_R} D_L R(\Delta\tau) \sin c(\pi \cdot \Delta f \cdot T_{coh}) \cos(\Delta\varphi) + \eta_I \quad (1)$$

$$\mathrm{Q}_L = \sqrt{2P_R} D_L R(\Delta\tau) \sin c(\pi \cdot \Delta f \cdot T_{coh}) \sin(\Delta\varphi) + \eta_Q \quad (2)$$

Where P_R is the average power of satellite signal, D_L is satellite spread spectrum code signal, $R(\bullet)$ is the correlation value between the received ranging code and the local code; η_I and η_Q are independent Gaussian white noise after accumulation, which obeys $N(0, L\sigma^2)$ in Gaussian distribution, σ^2 is the signal noise variance before coherent accumulation, $\Delta\tau = \tau - \bar{\tau}$ is the delay difference between the signal code and the local code, $\Delta f = f_d - \overline{f_d}$ is the frequency difference between received carrier signal and local carrier signal, $T_{coh} = L \cdot T_s$ is the product of sampling points L and the sampling period T_s, representing the coherent integration time, $\Delta\varphi$ is the difference of average phase after coherent integration.

Then, combined with the non-coherent integration method, the envelope of the accumulated value of the coherent integration is obtained. Firstly, the normalized noise variance of I and Q is obtained, and then noise components obey the N(0,1) distribution. The results of $\overline{\mathrm{I}_L}$ and $\overline{\mathrm{Q}_L}$ are non-coherently integrated M times, and the corresponding expression for this process is $\sum_{n=0}^{M} Z_L = \sum_{n=0}^{M} \sqrt{\overline{\mathrm{I}_L}^2 + \overline{\mathrm{Q}_L}^2}$. Finally, the acquisition of satellite signals is realized by acquisition decision.

The above coherent/non-coherent integration algorithm accurately captures the satellite signal and tracks it. In order to make the ionospheric scintillation indices accurately reflect the ionospheric scintillation, it is necessary to use detrending method to eliminate the fluctuation of tracking results caused by other factors. In the detrending process of ionospheric amplitude scintillation indices, we adopt the polynomial fitting method. Firstly, we calculate the standardized

signal strength:

$$WBP = \sum_{i=1}^{n} (I_i^2 + Q_i^2) \tag{3}$$

$$NBP = (\sum_{i=1}^{n} I_i)^2 + (\sum_{i=1}^{n} Q_i)^2 \tag{4}$$

Where the instantaneous code components I_i and Q_i of the same phase branch and the quadrature branch are measured to generate broadband power WBP and narrowband power NBP. If signal frequency of I_i and Q_i are 1000Hz, n is 60, and then the data length for power calculation is 60ms. Furthermore, we could calculate the signal strength:

$$S_{I,raw} = NBP - WBP \tag{5}$$

$$S_{I,norm} = \frac{S_{I,raw}}{S_{I,trend}} \tag{6}$$

Where $S_{I,raw}$ is the original signal strength, and then $S_{I,raw}$ is fitted by fourth-order polynomial to obtain $S_{I,trend}$.

In addition, in order to eliminate the influence of environmental noise in the amplitude scintillation indices, we need to exclude the amplitude scintillation indices generated by noise:

$$S4_{N_0} = \sqrt{\frac{100}{\overline{S/N_0}}\left[1 + \frac{500}{19\overline{S/N_0}}\right]} \tag{7}$$

Where $\overline{S/N_0}$ is the average of signal-to-noise ratio.

In scintillation detection, $S4$ is the amplitude scintillation indices representing the degree of signal attenuation, which is estimated as:

$$S4 = \sqrt{\frac{E\left(S_{I,norm}^2\right) - E\left(S_{I,norm}\right)^2}{E\left(S_{I,norm}\right)^2} - S4_{N_0}^2} \tag{8}$$

Where $E(\bullet)$ represents the expected value over a period of time, and the amount of data, which is used to obtain the expected value, is determined by the frequency of signal strength sampling and the expected time. When the sampling frequency is 1000Hz, the amount of data in 10s is 10000.

To reflect the fluctuation of the carrier phase, we define the phase scintillation indices as the standard deviation of the detrended carrier phase:

$$\sigma_\phi = \sqrt{E\left(\phi_{trend}^2\right) - E\left(\phi_{trend}\right)^2} \tag{9}$$

Where ϕ_{trend} is the detrended result from the carrier phase ϕ:

$$\phi = \phi_{raw} + \Delta\phi \tag{10}$$

Where the original carrier phase ϕ_{raw} is the phase offset caused by the Doppler shift. $\Delta\phi$ is the correction of the carrier phase and is corrected by the carrier phase detector (PLL) in tracking part of satellite receiver. $\Delta\phi$ also represents the phase error.

In order to eliminate noise, we use the sixth-order Butterworth high-pass filter [12] with a cutoff frequency of 0.1 Hz to detrend the carrier phase ϕ and finally obtain ϕ_{trend}. The six-order filter is composed of 3 cascaded second-order filters to improve the stability of the filter.

Fig.1 shows the time-varying information of the carrier phase, which is collected from GPS satellite No. 25 in the equatorial regions of Brazil on November 27, 2013. The right side of the longitudinal axis represents the original carrier phase value, and the left side of the longitudinal axis represents the corresponding detrended result of carrier phase by the sixth-order Butterworth at the same period.

In the shadow region, we can clearly observe that the detrended carrier phase fluctuates over time, indicating that the signal is affected by ionospheric scintillation.

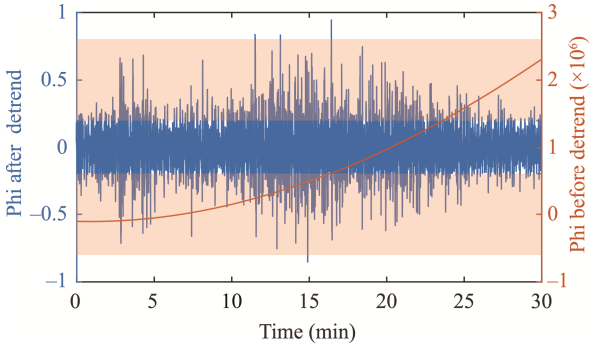


Fig.1 Comparison of Butterworth Filtering before and after

3 Detection System of Strong Ionospheric Scintillation

Ionospheric scintillation is a natural phenomenon

that it is difficult to predict the scintillation. In order to save data storage space, we combine the full-band receiving antenna with the commercial Septentrio receiver to monitor ionospheric scintillation. The acquisition system will be triggered to store data when a strong ionospheric scintillation is detected.

The software receiver receives the real-time satellite signal. After successfully tracking the signal, the software receiver collects the output and carrier phase data of I and Q channels, and then calculates the scintillation indices and judgment index of scintillation event in real time. The judgment index is compared with the set threshold.

If the trigger of the scintillation event is reached, the acquisition system will collect the multi-channel RF front-end data which is defined by the software and that data is the original IF data affected by ionospheric scintillation. The specific acquisition system is shown in Fig.2.

Because of the strong ionospheric scintillation in the equatorial regions, we adopt the coherent/non-coherent integration acquisition algorithm to improve the signal-to-noise ratio and achieve the purpose for accurate acquisition.

Compared with the serial search algorithm and the cyclic correlation algorithm [13], we adopt the coherent integration to improve the signal-to-noise ratio effectively, and then adopt the non-coherent integration to remove the limitation that the coherent integration time is easily affected by navigation data bit-flipping. Therefore, we can accurately capture satellite signals in a weak signal-to-noise ratio environment, which provides basic data for ionospheric scintillation detection.

For amplitude scintillation and phase scintillation, two different detrending methods, polynomial fitting and sixth-order Butterworth filtering, are taken to eliminate the influence of satellite motion, clock error and Doppler. By setting the scintillation threshold and recording the number of simultaneous occurrences of amplitude and phase scintillations, we successfully determine the occurrence of scintillations. The schematic diagram of ionospheric scintillation detection and analysis in software receiver is shown in Fig.3.

4 The Analysis of Ionospheric Scintillation Detection Data

Solar maximum year takes 11 years a cycle. Compared with other years, there is a frequent and strong ionospheric scintillation phenomenon in the Solar maximum year, and the latest year is 2013. Therefore, in order to test the feasibility of the designed algorithm under strong ionospheric scintillation mentioned in this paper, the data was collected by the data acquisition system deployed in the equatorial regions of Brazil on November 27, 2013. The specific location is shown in Fig.4:

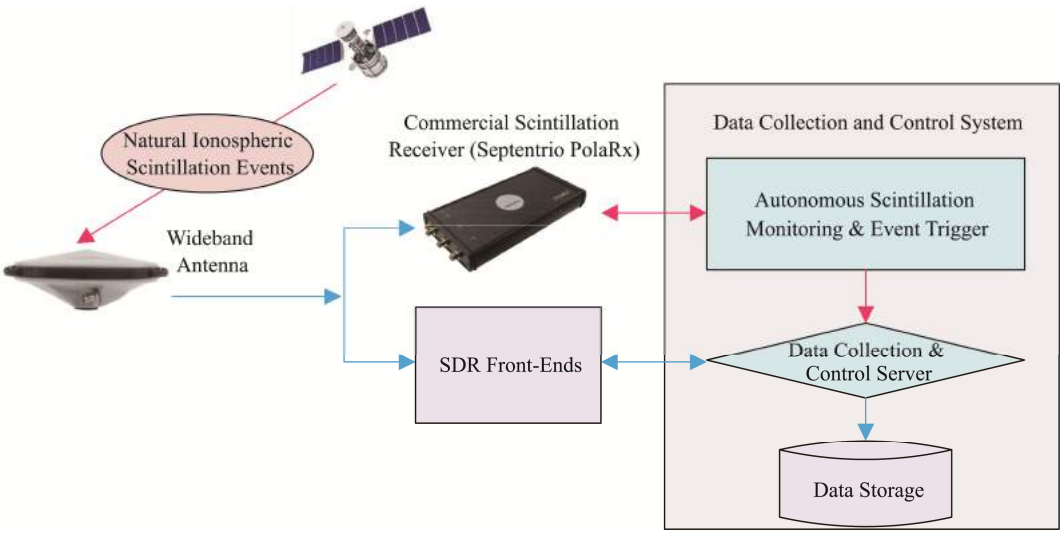


Fig.2 Data Collection System Driven by Ionospheric Scintillation

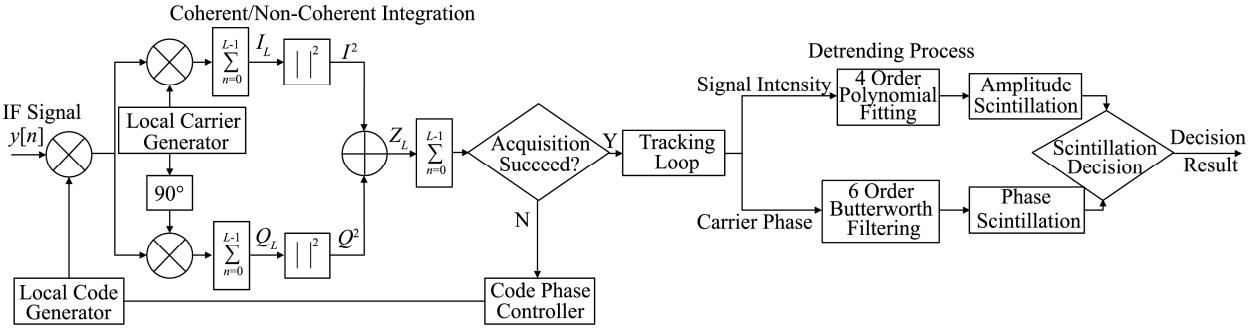


Fig.3 Principle of Ionospheric Scintillation Detection

Fig.4 Location of the Data Collection System on the Map

In MATLAB, we select the intermediate frequency signal of GPS L1 band recorded on November 27. We use the raw algorithm in GPS software receiver and acquisition algorithm improved by coherent/non-coherent integration to simulate the recorded data. The acquisition results corresponding to the two methods are shown in Fig.5.

Firstly, we set the acquisition threshold as 1.75. Satellite signals are seriously disturbed under the strong ionospheric scintillation, and the GPS software receiver with raw algorithm fail in capturing Satellite signal. As shown in Fig.5 (a).

The GPS receiver improved by the coherent/non-coherent integration algorithm captured 7 satellites, which were satellites 5, 12, 18, 21, 25, 29 and 31. As shown in Fig.5 (b).

In the case of successful acquisition, we track the signal. Fig.6 shows the tracking results of the instantaneous code phase and quadrature code phase of No. 25 satellite.

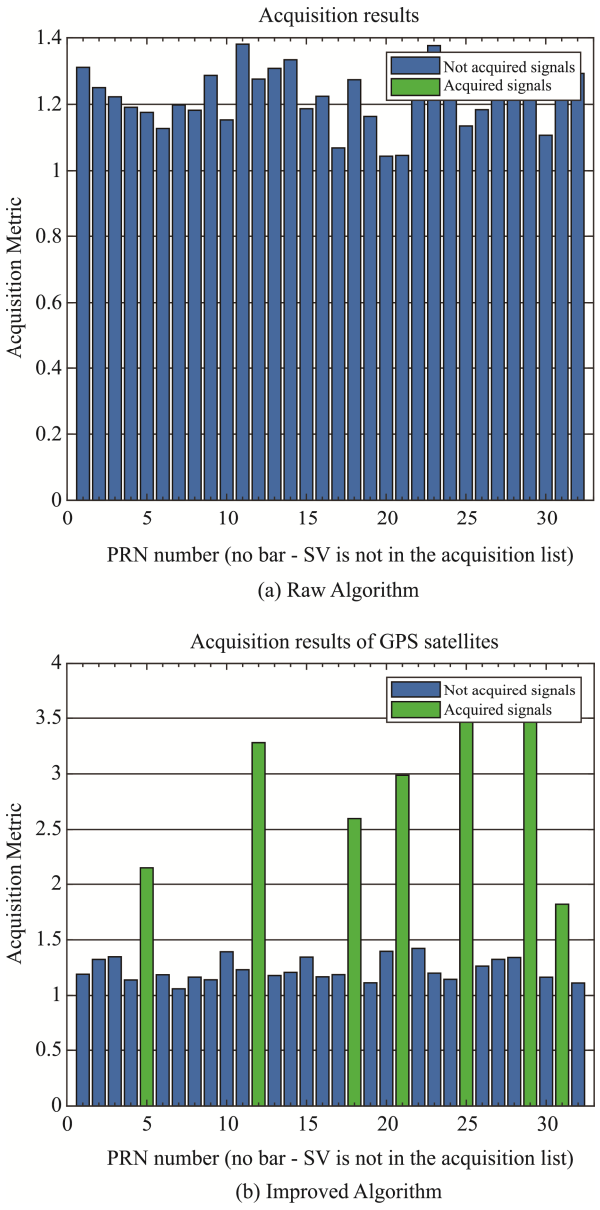


Fig.5 Acquisition under the Influence of Strong Ionospheric Scintillation

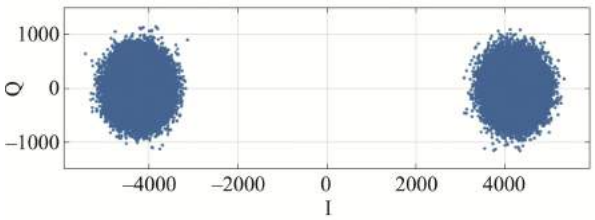


Fig.6 Tracking Results of GPS 25

Taking the tracking results of the captured satellite signal in GPS L1 band as an example, we detect the ionospheric scintillation within 1 hour, and discard the unstable data in the first 10 s, and the output frequency of scintillation indices is 1 Hz. The results are shown in Fig.7. Fig.7 (a) is the time-varying amplitude scintillation of GPS satellites 12, 18, 25 and 31, and Fig.7 (b) is the corresponding phase scintillation. When the amplitude scintillation indices S4 satisfies S4 < 0.2, it can be defined that there is no ionospheric scintillation; the amplitude scintillation indices satisfying 0.2< S4 < 0.5 is defined as weak scintillation; the amplitude scintillation indices satisfying S4 > 0.5 is defined as strong scintillation.

The red dotted line in Fig.7 (a) represents S4 = 0.5, and the moment that exceeds the red line indicates strong ionospheric scintillation occurs. Moreover, there is a small period of time when S4 is greater than 1, indicating that the scintillation reaches saturation, and that shows the signal is affected by strong ionospheric scintillation during this period.

In addition, since the detected signal comes from the Brazilian region near the equator, there should be relatively strong amplitude scintillation and relatively weak phase scintillation in signals. As is shown in Fig.7, strong amplitude scintillation is basically accompanied by phase scintillation, and the phase scintillation indices σ_ϕ satisfy σ_ϕ > 0.15 at some moments. The specific number and ratio of scintillation are shown in Table 1.

According to the ratio, amplitude scintillation mainly occurs in this region with lower intensity of phase scintillation. This result is consistent with the characteristics that there are mainly strong amplitude scintillations in equatorial and mainly phase scintillations in polar regions. The amplitude scintillation ratios of satellite No.12 and No.31 are 0.7339 and 0.6951, respectively. Therefore, in this period, the satellite signals are affected by strong ionospheric scintillation for more than 30 minutes. Meanwhile, the results show that satellites No. 18 and No. 25 are affected by strong ionospheric amplitude scintillation for nearly 30 minutes. Considering the factors such as the false acquisition of satellite signals and the enhancement of scintillation indices caused by other Interferences, we set the threshold of scintillation ratio to 50%. With this threshold, we successfully detected that satellites No. 12 and No. 31 were affected by strong ionospheric scintillation.

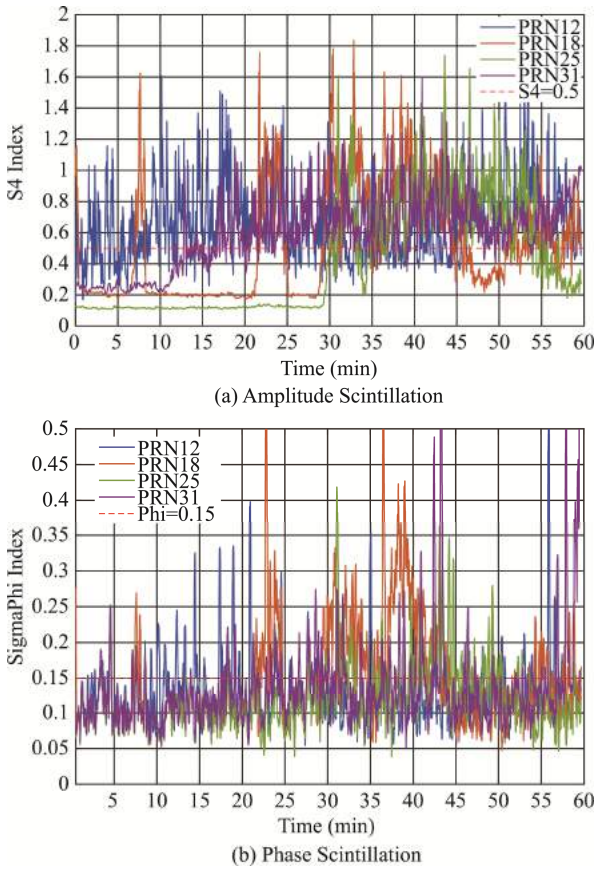


Fig.7 Ionospheric Scintillation Detection of GPS L1 Frequency

Table 1 Time and Proportion of Scintillation

PRN	Times of Scintillation		Ratio of Scintillation	
	S4>0.5	SigmaPhi>0.15	S4>0.5	SigmaPhi >0.15
12	2628	788	0.7339	0.2201
18	1598	1275	0.4462	0.3560
25	1324	762	0.3697	0.2128
31	2489	833	0.6951	0.2326

Taking actual acquisition point of satellite signal (Lat: –23°12′27.5772" Lng: –45°51′35.1414" Hgt: +693.4m) as reference, we test the positioning performance of the improved software receiver. The positioning error in the local cartesian coordinates coordinate system is shown in Fig. 8.

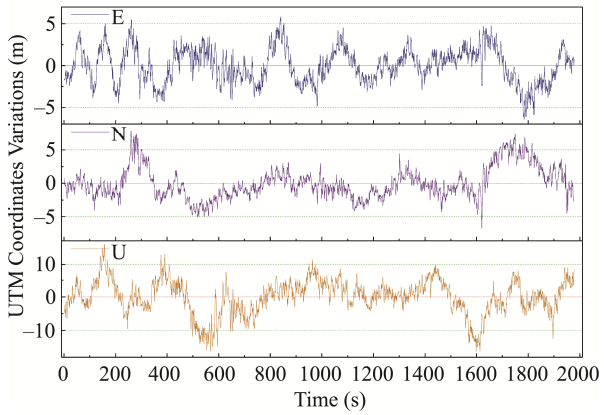


Fig.8 UTM Coordinates Variation

As shown in Fig.8, the horizontal direction error of the information calculated by the improved software receiver is controlled within 5 m, and height direction error is controlled within 10 m.

The satellite constellation after the data acquisition begins for 6 hours is shown in Fig.9.

(From http://www.gnssplanning.com)

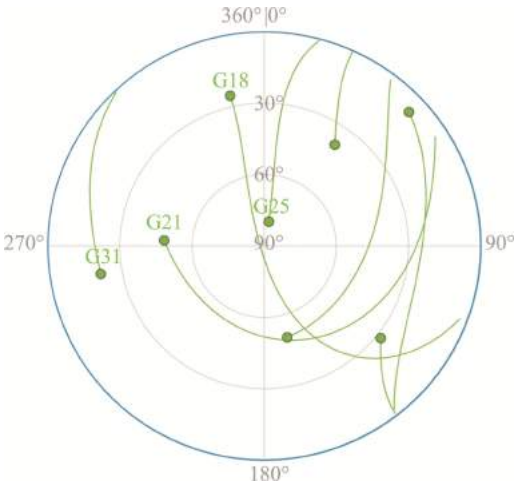


Fig.9 Satellites Sky Plot

5 Conclusion

Due to the complex space electromagnetic environment in signal transmission, the satellite signal quality is affected by ionospheric scintillation, which leads to the decrease of carrier-to-noise ratio of receiver and the increase of tracking error and positioning error. Especially in low latitudes and equatorial regions, severe ionospheric scintillation will lead to hours of navigation interruption. The software receiver improved in this paper realizes the detection of strong ionospheric scintillation in the equatorial regions and also realizes the successful acquisition, tracking, positioning and demodulation of satellite signals. The test results show that the horizontal positioning error can be controlled within 5 m, indicating that the software receiver is still reliable under the weak signal conditions caused by strong ionospheric scintillation. The improved software receiver can modify the scintillation detection threshold and detect ionospheric scintillation with different intensities, which provides important basis and useful guidance for analyzing ionospheric scintillation and space weather. This paper will be of significance for the future design of high-performance Beidou navigation receiver and also the territorial security of the South China Sea.

Acknowledgment

We would like to thank Professor Jade Morton and Professor Dennis M Akos in the University of Colorado at Boulder for providing the original measurement data of São José dos Campos, Brazil, and also thank their kind supports.

References

[1] Belakhovsky, V B. Jin, Y Q and Miloch, W J. (2021). Influence of Different Types of Ionospheric Disturbances on GPS Signals at Polar Latitudes. *Annales Geophysicae*, 39(4), pp.687-700.

[2] Yin, P. Li, Q Y. (2018). The Influence of Ionospheric E / F Layer Scintillation on GPS TEC Change Rate. *Journal of Civil Aviation University of China*, 36(5), pp.12-18.

[3] Jiao, Y. (2017). *Low-Latitude Ionospheric Scintillation*

Signal Simulation, Characterization, and Detection on GPS Signals. PhD. Colorado State University.

[4] Li, Q Y. and Yin, P. (2018). The Characteristic Study of Ionospheric Scintillations over China Based on GNSS Data. In: *Proceedings of the 9th China Satellite Navigation Academic Annual Conference*. Harbin: CSNAC, pp.190-195

[5] Cheng, J. Xu, J S. and Cai, L. (2018). A Comparison of Statistical Features of Ionospheric Scintillations and Cycle Slips in the Mid-South Region of China. *Chinese Journal of Geophysics*, 61(1), pp.18-29.

[6] Lan, W. and He, L. (2018). Analysis of Strong Ionospheric Scintillation Effects on Availability of Receiver. *Ship Electronic Engineering*. 38(10), pp. 204-206, 219.

[7] Mulugeta, S. and Kassa, T. (2022). Investigation of GPS Loss of Lock occurrence and Its characteristics over Ethiopia using Geodetic GPS receivers of the IGS network. *Advances in Space Research.* 69(2), pp.939-950.

[8] Huang, L F. Tian J P. and Zhao K. et al. (2019). Temporal and Spatial Characteristics of the Ionospheric Scintillation Event and the Influence on Communication in the Northern EIA Crest Region. *Chinese Journal of Space Science*, 39(2), pp.158-166.

[9] Zhu, X F. Chen, X Y. and Huang, H Q. et al. (2016). Analysis of GPS Ionospheric Scintillation Signal Amplitude Fading Characteristics at Low Latitude. *Journal of Southeast University*, 32(4), pp.484-488.

[10] Zhu, X F. Chen, X Y. and Chen, J F. (2016). Influence of Strong Ionospheric Scintillation on GPS Software Receiver. *Journal of Chinese Inertial Technology*, (4), pp.480-484.

[11] Zhu, X F. (2009). *Research on Key Techniques of High-Sensitivity GNSS Software Receivers*. PhD. Southeast University.

[12] Niu, F. (2012). *Performances of GPS Signal Observables Detrending Methods for Ionosphere Scintillation Studies*. PhD. Miami University

[13] Li, S. Yi, Q M. Chen, Q. and Shi M. et al. (2012). Weak Signal Acquisition Method for GPS Software Receiver. *Computer Application*, 32(3), pp.67-72.

Author Biographies

ZHU Xuefen received her Ph.D. degree from Southeast University in 2010. Now, she is an associate professor and master supervisor in Southeast University. Her main research interests include GNSS signal processing, GNSS/INS integration technology and detection on ionospheric scintillation.

E-mail: zhuxuefen@seu.edu.cn

LIN Mengying received the B.Sc. degree in electrical engineering and automation from Wuhan Institute of Technology, Wuhan, China, in 2017. She is currently a MD-PhD candidate in instrument science and technology in Southeast University, Nanjing, China. Her main research interests include interference detection on satellites' signals and GNSS signal processing technologies.

E-mail: 230198865@seu.edu.cn

LU Zhengpeng received the B.Sc. degree in measurement and control technology and instrument from Dalian Maritime University, China, in 2021. He is currently a M.Sc. candidate in the School of Instrument Science and Engineering, Southeast University, Nanjing, China. His main research interests include signal processing and interference detection on satellite signals.

E-mail: 220213669@seu.edu.cn

CHEN Xiyuan received his Ph.D. degree from Southeast University, Nanjing, China, in 1998. He is currently a Professor in the School of Instrument Science and Engineering, Southeast University. From October 2005 to September 2006, he was with the Department of Electronic Engineering, Politecnico di Torino, Italy, as a visiting scholar. He has authored or co-authored more than 200 papers and 60 granted patents. His main research interest includes fiber optic related application.

E-mail: chxiyuan@seu.edu.cn

A New Paradigm for Waste Classification Based on YOLOv5

Mohammed SAJID[1], Nimali T MEDAGEDARA[2]

(1. *Department of Mechanical Engineering*, *University of West of England*; 2. *Department of Mechanical Engineering*, *Faculty of Engineering Technology*, *The Open University of Sri Lanka*, *Nugegoda*, *Sri Lanka*)

Abstract: Classification of garbage is of paramount importance prior to process them to categorise physically and this process helps to manage wastes by maintaining pollution free environment. Many systems that have capability segregate garbage are on the rise, but efficient and accurate segmentation with recognition mechanisms draw the attention of researchers. A computer vision approach for classifying garbage into respective recyclable categories could be one of the effective and economical ways of processing waste. This project mainly focused on capturing real-time images of a single piece of garbage and classifying it into three divisions: paper, or metal, or biodegradable (food) waste. Each garbage class contains around 2000 images obtained from an open-source dataset and images collected from Google and personally collected custom images. The developed intelligent models provide the effectiveness of the machine and deep learning in classification with structural and non-structural data. The model used was a Convolutional Neural Network (CNN) named YOLOv5. The project showcased vision based approach capable of maintaining an accuracy of 61%. The CNN was not trained to its maximum capacity due to the difficulty of finding optimal hyperparameters, as most of the images were gathered from Google Images.

Keywords: Waste Classification, YOLOv5

1 Introduction

Different strategies for garbage removal are in use. Dumping into others compound or isolated lands, thrown in the ocean, or burning with no concern of the environment are in fact common occurrences. Disposing of garbage is generally a significant problem for society.

Solid waste management is a global issue, which everyone is concerned in the world. It is the responsibility of any government to frame policies for waste management which is one of the main reasons attributable to well being of communities. Inefficient waste management cause pollution of oceans, blockages of drains and destroying flora and fauna. Moreover, the impact of improper waste disposal cause flooding during rainy season, transmitting diseases and increasing respiratory issues, such as airborne particles from incineration, harming animals that consume garbage, and affecting economic development that has an negative impact om tourism industry[1].

The globality of annual waste generation is on the verge of increasing and it is projected to reach 3.40 billion tones by 2050.Global waste will be developed by 70 percent on current levels by 2050. It has been estimated that around 40% of waste generated worldwide is not managed correctly instead of dumping and open incineration. A significant proportion of the total population has no access to appropriate waste disposal

services, eventually it is poor waste mismanagement. There is even the problem of transportation of the trash due to rising transportation cost and this attributes to garbage collection around 39%[2] in low-income countries is. Improper garbage disposal devotes to disastrous epidemics of mosquito-borne malaria, hepatitis, and other potentially fatal disease[3], which endangers well being of society.

Every year around 2.01 billion metric tons of solid waste are generated [4]. This rubbish will put much pressure on the environment. If the garbage is not well classified, it will greatly harm the environment that the people live in. Classification would ease the waste management process.

1.1 YOLO Algorithm

For providing real-time object detection, YOLO algorithm is used with neural networks. The term YOLO means 'You Only Look Once'. This algorithm has the ability to detect and recognize various objects as images in real-time. This algorithm is used in various applications with high speed and accuracy[5]. YOLO algorithm exploits convolutional neural networks (CNN) to identify images of objects in real-time. The models are trained to inspect the images and search for a subset of object classes. Object detection is a phenomenon in computer vision that identifies different objects in digital images or videos. The main advantage of real-time object detection models is that they are small and easy to operate by all developers.

1.2 Importance of YOLO algorithm

This algorithm has a predictive technique with minimal background errors and gives the exact outcome. This has good learning capabilities to learn the representations of objects and apply them in object detection [5]. This project utilizes YOLOv5s version of the YOLO model due to its faster speed, as shown in Fig.1.

Model	AP^{val}	AP^{test}	AP_{50}	$Speed_{GPU}$	FPS_{GPU}	params	FLOPS
YOLOv5s	36.6	36.6	55.8	**2.1ms**	**476**	7.5M	13.2B
YOLOv5m	43.4	43.4	62.4	3.0ms	333	21.8M	39.4B
YOLOv5l	46.6	46.7	65.4	3.9ms	256	47.8M	88.1B
YOLOv5x	**48.4**	**48.4**	**66.9**	6.1ms	164	89.0M	166.4B
YOLOv3-SPP	45.6	45.5	65.2	4.5ms	222	63.0M	118.0B

Fig.1 Different Version of YOLOv5 Model[6]

1.3 YOLOv5 Model

This model can be easily used, and the YOLOv5 framework is in custom domains. An object detector is accomplished to generate features from input images and then provide them through a prediction system to draw boxes around objects for predicting their domain[5].

One-stage detectors and two-stage detectors are two categories of object detection. Two-stage detectors decouple the task of object localization and classification for each bounding box. One-stage detectors make the predictions for object localization and classification simultaneously. YOLO is a one-stage detector (Fig.2); hence, You Only Look Once. Fig.3 depicts the basic anatomy of object detectors.

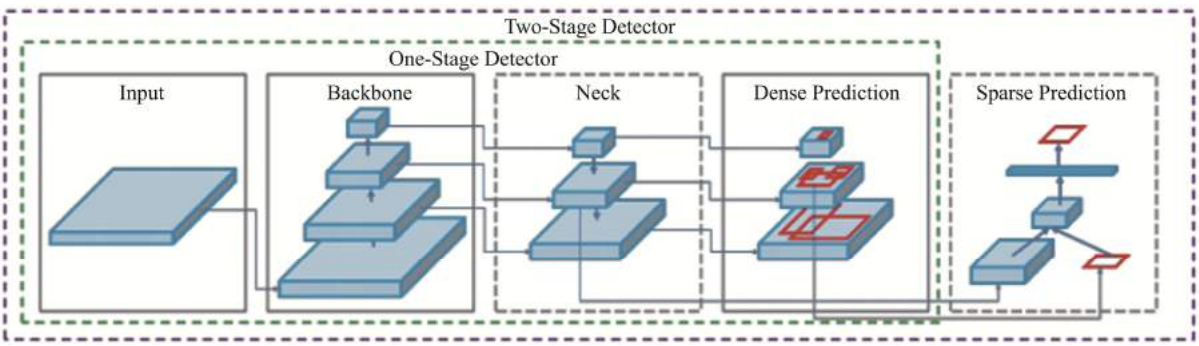


Fig.2 One-stage Versus Two-stage Detectors[5]

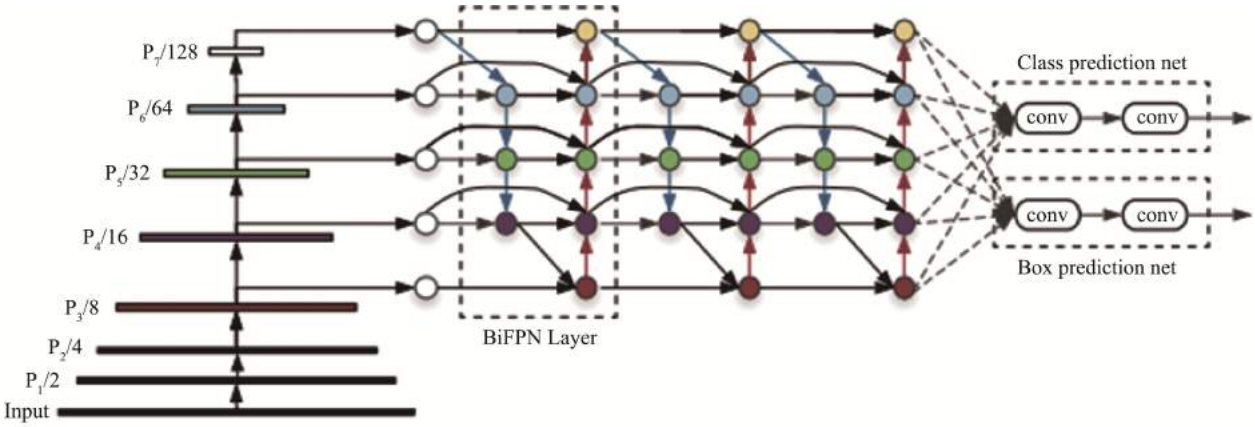


Fig.3 The Anatomy of an Object Detector [7]

1.4 YOLOv5 Model Architecture

Backbone, Neck, and Head are the three main stages of the YOLO [8-9]. Backbone is a CNN that combines and forms image features at different granules. Neck stage is a series of layers to mix and combine image features for prediction. Head stage absorbs features from the neck and captures box and class prediction.

1.5 Working of YOLO Algorithm

Working of YOLO algorithm follows three techniques such as residual blocks, bounding box regression and intersection over union[5-6].

1.5.1 Residual Blocks

Initially, the required image should be divided into number of grids. Grid division of the image as shown in Fig.4.

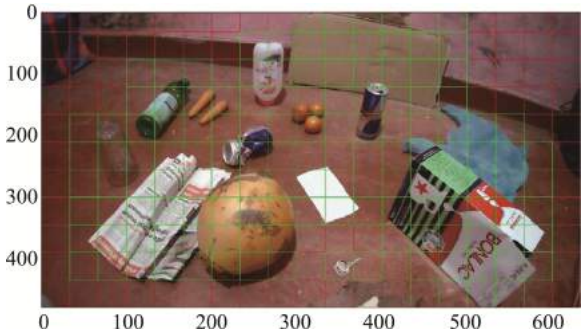


Fig.4 Grid Division of an Image

The Fig.4 shows the division of the input image into grids.

1.5.2 Bounding Box Regression

Outline that highlights an object in an image shown in Fig.5 is defined as a bounding box. Every bounding box in the image consists of the following attributes Width (bw), Height (bh), Class (c), Bounding box center (bx, by).

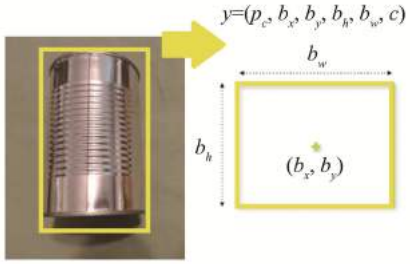


Fig.5 Example of Bounding Box Regression

1.5.3 Intersection Over Union (IOU)

IOU phenomenon is used in object detection. IOU uses in YOLO to permit an output box that surrounds the objects as shown in Fig.6.

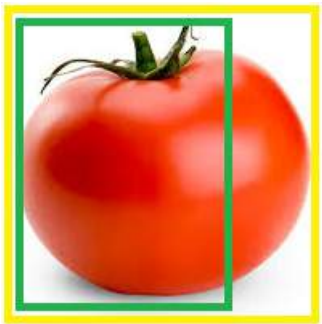


Fig.6 Example of IOU Working

2 Methodology

The project proposes a small-scale method to detect three different garbage classes: Biodegradable, Paper, and Metal. This could be implemented on a large scale for improved waste segregation.

3 Data Collection

Data collection is one of the major activities of training a model. A sufficient number of images are required in order for the model to work in high speed and with accuracy. The dataset for this project was obtained from various sources such as Open source data[10], images collected from Google sites and personally collected custom images were used for the research.

The dataset included in this study has 6000. Each class of trash has about 2000 images each. Out of which 70% of the images were used to train the model, 20% of the images were used to validate the model, and the rest 10% were used to test the model.

The next Step is labeling the images into the appropriate classes and checking if every image has its own .txt file.

3.1 Training & Validation

The data set was split into training and validation sets for validating the implemented model. The purpose of training set is to set the model in learning form. Validation set is used during training to iteratively estimate the model accuracy.

3.2 Testing

After training the model, the testing of the custom model was commenced. The working model can be used to detect different trash in pictures and videos. Testing was performed by individual image files being dragged onto the test area for the testing process. The model will return a confidence score for objects with the mean accuracy rate alongside that match the training data. Confidence scores will rise as the performance of the model improves through repeated training. Training a model can take several hours to complete. Retraining and testing of the model can be performed until the user is content with the model's performance. Fig.7 shows the testing and detection process while having a live feed from a Camera.

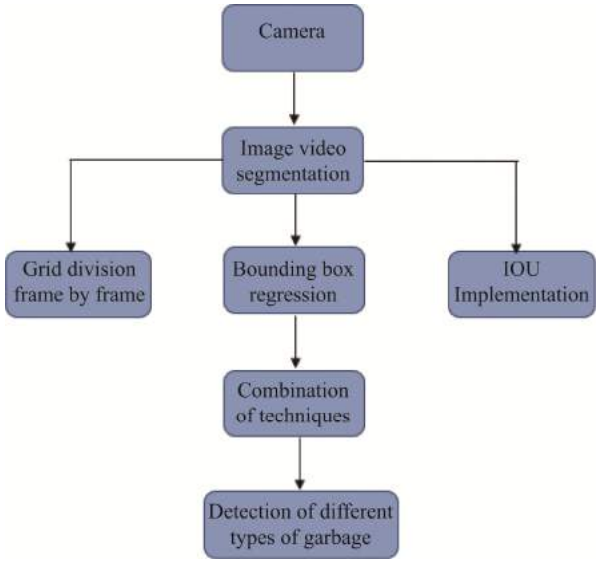


Fig.7 Flow Chart of Testing

4 Results

4.1 Individual Paper Testing

The Mean Average Accuracy (MAP) can be utilized for assessing YOLO object identification model. MAP compares the bounding box regression to the recognized box and returns the score. (AP) is the Average Precision and (IoU) defines the Intersection over Union, and this is used to measure the overlap between 2 boundaries. For measuring the accuracy of object detectors, AP metric is used.

Here are their mathematical definitions:

$$Precision=\frac{TP}{TP+FP}$$

$$Recall=\frac{TP}{TP+FN}$$

$$F1=2\cdot\frac{precision\cdot recall}{precision+recall}$$

TP=True positive
TN=True negative
FP=False positive
FN=False negative

In this datasets, IoU threshold was predefined as 0.5 for classifying whether the prediction is a true positive or a false positive.

Fig.8 shows testing done on Paper Waste. Table 1 shows the results achieved for paper waste.

Fig.8 Individual Testing: Paper Waste

Table 1 Results for Paper Waste

Precision	Recall	mAP at 0.5 IoU Threshold	mAP at 0.95 IoU Threshold
0.03	0.06	0.02	0.01

4.2 Metal Waste

Fig.9 shows the individual testing performances on metal waste. The following are the results achieved (Table 2).

Fig.9 Individual Testing: Metal Waste

Table 2 Results for Metal Waste

Precision	Recall	mAP at 0.5 IoU Threshold	mAP at 0.95 IoU Threshold
0.77	0.54	0.65	0.39

4.3 Biodegradable

Fig.10 shows testing done on Biodegradable Waste. The following are the results achieved (Table 3).

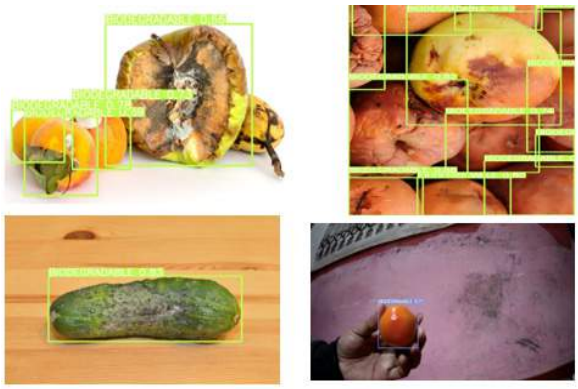


Fig.10 Individual Testing: Biodegradable Waste

Table 3 Individual Results for Biodegradable Waste

Precision	Recall	mAP at 0.5 IoU Threshold	mAP at 0.95 IoU Threshold
0.83	0.45	0.58	0.30

4.4 Batch Testing

The Table 4 shows the results achieved for overall model performance by testing of model concurrently in a batch.

Table 4 Results for Batch Testing

Precision	Recall	mAP at 0.5 IoU Threshold	mAP at 0.95 IoU Threshold
0.91	0.61	0.50	0.30

4.5 Mosaic Augmented Batch Testing

Fig.11 shows implementation of the Data Augmentation method known as Mosaic Augmentation. This merges four training images into one in certain ratios. This allows for the model to learn how to identify the objects when cropped, or different rotation, or flipped, or blurred or at a lower scale than normal[8-12]. It also is useful in training to significantly reduce the need for a large mini-batch size[8]. The following are the numbering system followed in order to identify the object according to the available classes: 0 - Biodegradable, 3 - Metal, 4 – Paper.

Fig.11 Mosaic Augmented Testing

Raspberry Pi running the latest version of Raspbian OS was used for the purpose of real time identification. Python 3 was used as the development environment. After capturing the objective image and the reference image is identified, the rest of the process was completely automatic. Hence no need for user intervention. The algorithm has been applied to the complete image.

4.6 Overall Results

In the results shown in Fig.12, x-axis indicates number of epochs and y-axis indicates model performance. The plots Fig.12 (a) include bounding box loss, objectness loss, classification loss over the training epochs for the training set. Fig.12 (b) includes the validation set of bounding box loss, objectness loss, classification loss, precision, recall and mean average precision (mAP) at 0.5 and 0.95 IoU threshold over the training epochs.

4.7 Model Performance Analysis

The epoch accuracy is used to identify the model's trained quality and its performance for future prediction. The graph (Fig.13) represents Performance Value (y-axis) versus Number of Epochs (x-axis). The model was run for 350 epochs and trained for 13 hours. The model performance graph (Fig.13) shows accuracy and loss metrics of Training and validation sets. It can be clearly seen that accuracy of both sets increases over time and the reach saturation points at around 250 epochs. And the loss for both sets remains low which is essentially good for the model.

This research was conducted to identify how well the model was able to detect real – time trash and classify them into different categories. The experiment was performed by placing different trash (paper, metal, biodegradable) in front of the camera which was previously not added to the dataset. The results showed 61 % mean accuracy towards most images. The analysis was done by using 6000 images and due to this large dataset, each of the features in the images could be properly analyzed during the modelling process.

When the TensorFlow Lite (TFlite) model was running 1.07 FPS (Frame Per Second) was achieved**.** Therefore, these results show that a mobile detector and/or a Smart Garbage Bin with YOLOv5 can be used for real-life simple trash identification and collection.

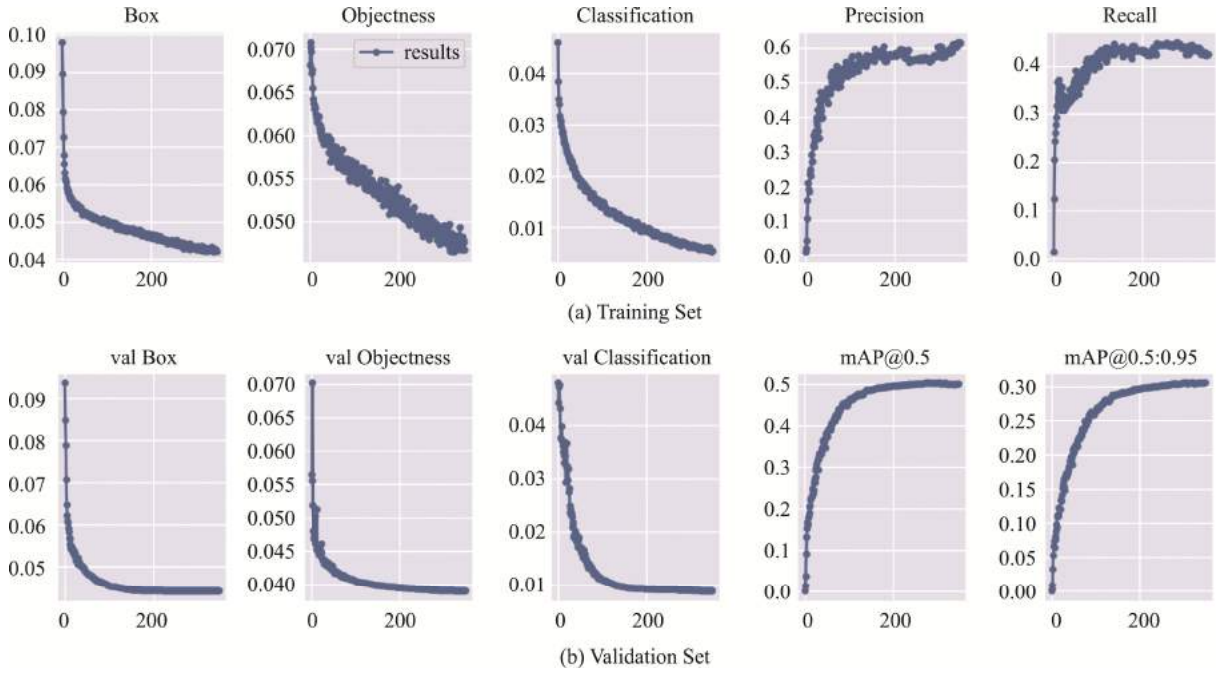


Fig.12 Overall Results

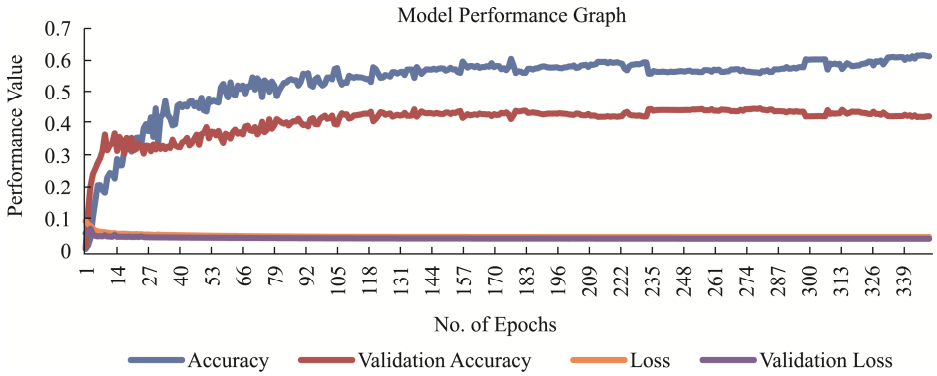


Fig.13 CNN Model Performance Analysis

4.8 Problem with Paper

Among the 2500 images present in the Paper Class, most of the images taken for training the model, consisted of zoomed in image of paper (Fig.15). This was developed due to a random function implemented in order to keep the train set, validation set and the test set neutral. But due to this function the accuracy of detecting paper was decreased. Fig.14 shows the Overall Heatmap of all the images in the paper class, it can be seen that the annotation for paper went from end to end, that is, from y_{max} to x_{max}. But the model still accurately detects paper under proper lighting.

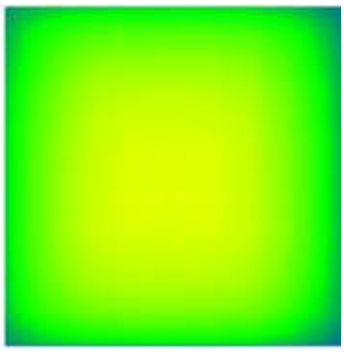


Fig.14 Annotation Heatmap of the Paper Class

Fig.15 Error in Images of Paper in Dataset

5 Conclusion

The aim of this project was to investigate the suitability of running a real - time object detection system on a Raspberry Pi and on Google Colab Pro to identify different types of trash. YOLOv5s and Tflite models were implemented and tested for accuracy and speed at different input sizes. The results showed that the model is accurate and that only in applications that do not require high speed would it be viable to use the Raspberry Pi as hardware. However, detections made on Google Colab were found to be faster and more accurate due to its running on a GPU. There is a trade-off with accuracy to be made if higher speeds are to be achieved since there is not enough computational power to have both. This study could be of help for others who will be looking to implement object detection/classification models on Raspberry Pi to make Garbage Detection projects of their own. The classification of trash into various recycling categories is possible through machine learning and computer vision algorithms. One of the biggest pain-point is the wide varieties of possible data. Therefore, in order to create a more accurate system, there needs to be a large and continuously growing dataset. The developed intelligent models provide the effectiveness of machine learning and deep learning in classification with structural and nonstructural data. The model used was a Convolutional Neural Network (CNN) named YOLOv5. The project showcased that the garbage classified was mostly accurate (61%); however, the CNN was not trained to its full capability due to the difficulty in finding optimal hyperparameters as most of the images were collected from Google Images. As further improvements author suggests to perform more testing of different trash, which comes in different sizes, shapes and color and to test the impact of lighting that has on a model's ability to detect trash. The future work can be focused to explore the on the above.

References

[1] Anon, *Trends in Solid Waste Management*. Available at:https://datatopics.worldbank.org/what-a-waste/trends_in_solid_waste_management.html [Accessed November 24, 2021].

[2] Ellis, C., 2018. World Bank: Global Waste Generation could increase 70% by 2050. *Waste Dive*. Available at: https://www.wastedive.com/news/world-bank-global-waste-generation-2050/533031/ [Accessed November 24, 2021].

[3] Kaza, S. et al., 2018. *What a waste 2.0: A global snapshot of Solid Waste Management to 2050*, Washington, DC: World Bank Group.

[4] Anon, 2021. Global waste index. *Sensoneo*. Available at: https://sensoneo.com/global-waste-index/ [Accessed November 24, 2021].

[5] Bochkovskiy, A., Wang, C.-Y. & Liao, H.-Y.M., 2020. *YOLOv4: Optimal Speed and Accuracy of Object Detection*. dissertation.

[6] Anon, Introduction to Yolov5. *YOLOv5 Documentation*. Available at: https://docs.ultralytics.com/ [Accessed November 24, 2021].

[7] Tan, M., Pang, R. & Le, Q.V., 2019. *EfficientDet: Scalable and Efficient Object Detection*.

[8] Solawetz, J., 2021. Breaking down Yolov4. *Roboflow Blog*. Available at: https://blog.roboflow.com/a-thorough-breakdown-of-yolov4/ [Accessed November 24, 2021].

[9] Solawetz, J., 2021. Yolov5 new version - improvements and evaluation. *Roboflow Blog*. Available at: https://blog.roboflow.com/yolov5-improvements-and-evaluation [Accessed November 24, 2021].

[10] Thung, G. & Yang, M., 2016. *Classification of Trash for Recyclability Status*.

[11] Aysegul TakimogluAysegul has experience in HR management & *, N., 2021. What is Data Augmentation? techniques, benefit & examples. *AIMultiple*. Available at: https://research.aimultiple.com/data-augmentation [Accessed November 24, 2021].

[12] Solawetz, J., 2020. Data augmentation in yolov4. *Roboflow Blog*. Available at: https://blog.roboflow.com/yolov4-data-augmentation/ [Accessed November 24, 2021].

Author Biographies

Mohammed SAJID is currently a B.Eng. candidate in Mechanical (Mechatronics) from University of West England. His main research interests include Computer Vision (CV) and Artificial Intelligence (AI).

E-mail: affilian147@gmail.com

Nimali T MEDAGEDARA received B.Sc. (Hons) degree in Mechanical Engineering from the University of Peradeniya (Sri Lanka) in 1993, and MPhil in Materials and Manufacturing Engineering from Sheffield Hallam Uni-University (UK). She is a charted engineer and member of the Institute of Engineers Sri Lanka (IESL), Member of IEEE and Member of SLAAS. She is now a senior lecturer in the open university of Sri Lanka, and the principal investigator in the Soft Robotics Research Group (SORRG) (https://www.ou.ac.lk/SRRG/) of OUSL. Her main research interests include soft robotics, finite element analysis and multi axial fatigue.

E-mail: tmmed@ou.ac.lk

Research on Fall Detection Based on Improved Human Posture Estimation Algorithm

ZHENG Yangjiaozi[1], ZHANG Shang[2]

(1. *China Three Gorges University*, *Computer and information institute*, *Yichang* 443002, *China*;
2. *Hubei Province Engineering Technology Research Center for Construction Quality Testing Equipments*, *China Three Gorges University*, *Hubei*, *Yichang*, 443002, *China*)

Abstract: According to recent research statistics, approximately 30% of people who experienced falls are over the age of 65. Therefore, it is meaningful research to detect it in time and take appropriate measures when falling behavior occurs. In this paper, a fall detection model based on improved human posture estimation algorithm is proposed. The improved human posture estimation algorithm is implemented on the basis of Openpose. An improved strategy based on depthwise separable convolution combined with HDC structure is proposed. The depthwise separable convolution is used to replace the convolution neural network structure, which makes the network lightweight and reduces the redundant layer in the network. At the same time, in order to ensure that the image features are not lost and ensure the accuracy of detecting human joint points, HDC structure is introduced. Experiments show that the improved algorithm with HDC structure has higher accuracy in joint point detection. Then, human posture estimation is applied to fall detection research, and fall event modeling is carried out through fall feature extraction. The designed convolution neural network model is used to classify and distinguish falls. The experimental results show that our method achieves 98.53%, 97.71% and 97.20% accuracy on three public fall detection data sets. Compared with the experimental results of other methods on the same data set, the model designed in this paper has a certain improvement in system accuracy. The sensitivity is also improved, which will reduce the error detection probability of the system. In addition, this paper also verifies the real-time performance of the model. Even if researchers are experimenting with low-level hardware, it can ensure a certain detection speed without too much delay.

Keywords: Fall Detection, Human Posture Estimation, Depthwise Separable Convolution, Convolutional Neural Networks, Feature Extraction

1 Introduction

According to statistics from the World Health Organization, “falling” is the second leading cause of death from all accidental or unintentional injuries in the world，with traffic accidents being the leading cause[1]. And according to research studies, approximately 30% of people who experienced falls are over the age of 65. Falls can cause great harm to the body of the elderly and bring them invisible pressure. This not only affects the daily lives of the elderly but also indirectly increases the pension burden of families and society. Therefore, for the elderly care monitoring problem, fall detection is a research content of great significance.

This research will minimize the physical injury caused by falls to the elderly and reduce their physical and mental suffering of the elderly. At the same time, it can also save social public medical resources and reduce the burden of family pension. In this paper, we propose a fall detection system based on human pose estimation and Convolutional Neural Networks. First, we preprocess the data and extract the position of human skeletal key-points from successive frames using the human pose estimation algorithm. The human body is extracted from the background by the algorithm while being tracked in real-time. Next, we extract the body descent rate and external profile features. After the primary judgment using the body descent rate, the secondary judgment is made by calculating the aspect ratio of the minimum external rectangular frame of the human body. After the secondary judgment, the fall event modeling is carried out to construct the fall features. Third, the designed Convolutional Neural Networks model is used to process and classify the above features. Finally, the classification results are applied to fall detection.

2 Related Work

2.1 Fall Detection Technology

In recent years, the trend of fall detection research has been increasing in recent years. Researchers have conducted more in-depth research on human behavior recognition and fall detection technology. From different implementation methods, fall detection technology can be roughly divided into the following three types, as shown in Fig.1: vision-based sensors, wearable device-based sensors, and ambient sensors[2].

2.1.1 Vision-based Sensors Fall Detection

Due to the widespread use of cameras in our lives, it has become a normal state to use cameras as a device for acquiring information in the field of computer vision research. The vision-based sensors mainly collect the data of human movement behavior through image or video capture devices installed in the user's living environment. And then the system uses image or video processing to perform human behavior recognition to determine if fall events have occurred. Chong et al[3]proposed a method that combines a background modeling approach of superpixel clustering with a background segmentation method based on the Horprasert algorithm. The author used two different methods, namely boundary box, and near elliptical motion quantization, to detect the drop of the foreground extracted by the method proposed in this paper. The results show that the method improves the processing speed and reduces the complexity of the original Horprasert segmentation.

The advantages of the vision-based sensors are easy to install the equipment and the user can visualize the video screen. However, at the same time, this method has some disadvantages, such as the target person may be obscured. To solve the shortcomings of conventional cameras, more and more researchers are choosing to use depth cameras. The most common method is to extract skeletal data from the human body. Bian et al[4]developed a two-tier system. First, the 3D coordinates of the body joints are removed from the segmented body parts by interpolation. Then, the position of the joint trajectory over time is extracted and the SVM classifier is used to detect the position of the joint.

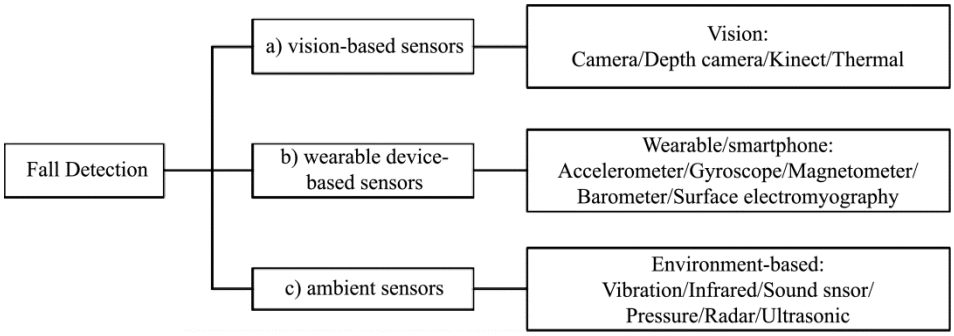


Fig.1 Fall Detection Technology Classification

2.1.2 Wearable Device-based Sensors Fall Detection

Wearable device-based sensors for fall detection are developing rapidly and there are many kinds of devices available, the most common being accelerometers and gyroscopes. Instruments commonly used by researchers in fall detection studies include wearable watches, wearable belts, and wearable helmets. Rakhman et al[5] proposed a fall detection system based on a smartphone accelerometer and gyroscope. The authors used a three-axis accelerometer and gyroscope embedded in a smartphone to collect data on the human body's activity behavior and fall behavior. Next, the data characteristics features of different data are compared and analyzed with specific thresholds, and then a fall detection model is established. In addition, the fusion of wearable devices and other information technologies for multimodal information judgment is also a hot research topic. Yan Yujuan et al[6] designed a fall detection system that used an acceleration sensor, the ADXL330, combined with computable radio frequency identification tagging (CRFID) and pattern recognition technology. It used an acceleration sensor embedded in an ultra-high frequency CRFID to collect acceleration information of human movements to determine whether a fall occurred.

2.1.3 Ambient Sensors Fall Detection

Ambient-based sensors usually use sensors installed in the home environment to collect signals such as vibration signals, sound signals, and pressure signals to track the human body. Diego et al[7]proposed a detection method combining a class SVM (OCSVM) and a template matching classifier. The floor sound sensor captures the sound signal and then extracts the Mayer frequency cepstrum coefficient and Gaussian mean super vector (GMS). Users mark false-positive events and record them in the template GMS. The template matching classifier uses this template to make the final decision to distinguish human falls from non-falls. The ambient-based sensors allow direct access to target actions without violating the user's privacy. However, the detection environment can only be indoors, none of these systems can be used outdoors.

In addition to traditional acoustic and pressure sensors, many technologies such as ultrasound, infrared sensing, and radar detection are gradually being applied to systematic fall detection research. Zhang Dajun et al[8] proposed a method for fall detection using ultrasound at 20 kHz. The weightlessness caused by fall events increases the velocity of body movement, resulting in a Doppler effect that causes a shift in the reflected ultrasound frequency. Then, the system uses the offset data to determine whether the human has fallen. Sensors are sensitive to information such as sound and vibration, while their resistance to interference is weak. So, the most obvious disadvantage of this method is that it is easy to misjudge.

2.2 Skeleton Estimation

Generally speaking, human posture estimation algorithm is carried out by finding the position coordinates of human or object joint points. Take a person as an example, the joint points are elbow, knee, wrist and other joints. There are two types of pose estimation: multi pose and single pose. Single pose estimation is used to estimate the pose of a single object in a given scene, while multi pose estimation is used to detect the pose of multiple objects. Human posture estimation on popular MS COCO datasets can detect 17 different joint points (classes)[9]. Each joint point is annotated with three numbers (x, y, V), where x and Y mark the coordinates, and V indicates whether the joint node is visible

In this paper, we choose to apply human posture estimation to fall research. Through the analysis of human behavior using human posture estimation algorithm, we can distinguish between daily life behavior and fall behavior. The human posture estimation algorithm can extract the human joint point data under different postures. It can remove the background noise to the greatest extent, preserve the human structure, enhance the contrast between the human trunk and the background, and reduce the learning difficulty of the network. This can help us avoid the interference caused by human background to the greatest extent in fall detection research.

2.3 Depthwise Separable Convolutional

Convolutional Neural Networks (CNN) can be

regarded as a variant of multilayer perceptron (MLP). Krizhevsky et al[10]applied CNN for the first time in the LSVRC-12 competition. By deepening the depth of the CNN model and combining relu and dropout technology, they achieved the best classification results at that time. The network structure has been named Alex Net. Simonyan et al[11]explored the importance of the 'depth' of CNN. Based on the existing network model structure, how to solve the network depth problem has become an important issue. To solve this problem, researchers proposed the inclusion of a convolutional layer with 3*3 convolutional kernels. And the experimental results showed that the performance of the model can be effectively improved when the number of weight layers reaches 16-19 layers, which is called VGG. Through continuous improvement and optimization, various convolutions such as Group convolution, Dilated convolution, and Depthwise separable convolution were created.

The convolution kernel can be regarded as a three-dimensional filter: channel dimension plus spatial dimension (width and height of the feature map). In regular convolution, the upper layer in the connection generally has multiple channels (we assume N channels here). Therefore, a filter must have N convolution kernels to correspond to it when performing convolution calculations. The essence of a filter completing one single convolution is that multiple convolution kernels are convolved with the feature maps of the corresponding channels of the previous layer. The convolution results are then summed to output a channel feature map for the next layer. In the next layer, if we need to get a feature map of multiple channels (we assume there are M channels), then we need to have M filters. The structure diagram of convolution layer of ordinary convolution is shown in Fig.2.

As shown in Fig.3, depthwise separable convolution decomposes the traditional convolution into a depthwise convolution and a 1×1 convolution (pointwise convolution). For the multi-channel feature maps from the previous layer, they are first all split into single-channel feature maps. The single-channel feature maps are convolved one by one for a single channel and then re-stacked together. Compared to conventional convolution, the biggest advantage of depthwise separable convolution is that it can greatly reduce the number of model parameters and calculations while maintaining a high level of accuracy.

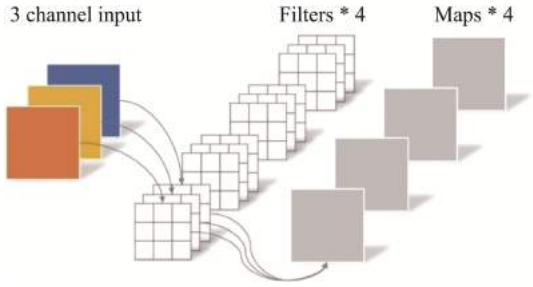


Fig.2 The Structure Diagram of Convolution Layer of Ordinary Convolution

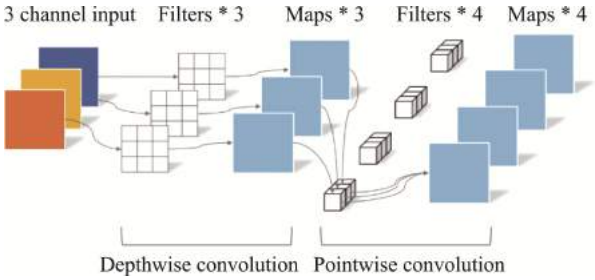


Fig.3 The Structure Diagram of Convolution Layer of Depthwise Separable Convolution

3 Methods

3.1 Openpose Algorithm

The openpose algorithm was first proposed in 2017. It is an open source library developed by the University of Nicky Mellon in the United States with Caffe as the framework on the basis of convolutional neural network and supervised learning[12]. Openpose algorithm takes as input a color image of size w*h and as output a two-dimensional skeleton image containing the location of key points for each person. The backbone network uses a VGG19 network model, where the input raw graph is initialized and fine-tuned by the first 10 layers to generate a set of feature maps *F* as input to the first stage. At this point, the algorithm enters the refinement stages. In the initial stage, there is the 3 * 3

convolution kernel in the convolutional layer, and in the refinement stage, it becomes the 7*7 convolution kernel. Starting from the second stage, the input to the stage t network consists of three components: S_{t-1}, L_{t-1}, F. S_{t-1}, L_{t-1} is the output of the stage t-1 network. In other words, the input to each stage of the network is the output of the previous stage of the network. We use a feedforward network to predict a two-dimensional confidence map S of body part locations and a set of vector fields L associated with body parts when using the algorithm to extract features. The degree of association between parts of the human body is obtained by encoding the association vector field. Finally, the confidence map and association field are resolved by the greedy algorithm, and the two-dimensional key points of everyone in the image are output.

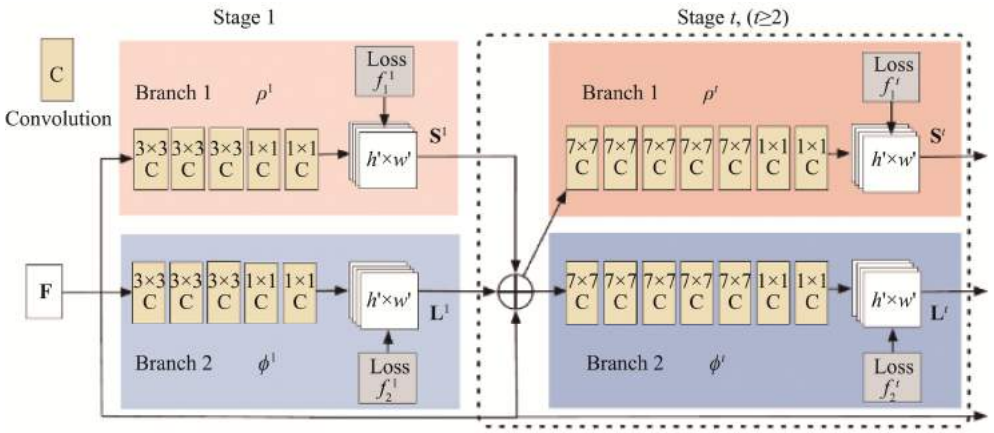


Fig.4 Network Structure of Openpose Algorithm[12]

3.2 Our Algorithm

To reduce the model parameters and increase the computational speed, we improve the Openpose algorithm from three perspectives. The first is for the backbone part of the network, the Openpose algorithm uses the VGG19 network. It uses the first four blocks of the VGG19 network and then adds two additional convolutional layers, later on, to complete the feature extraction before entering the initial stage. To reduce the model parameters and computational effort, we use depthwise separable convolution. The structure of the depthwise separable convolution is adapted to replace the VGG19 network. The comparison of convolution layer structure before and after improvement is shown in Fig.5.

The improved network structure is shown in Table 1. In the improved algorithm structure, the first two layers are traditional convolution layers, the size of convolution cores is 3*3, the number of convolution cores is 32, the step length is 1, and the filling coefficient is 1. The last ten layers establish a deep separable convolution neural network structure. The size of convolution kernel is 3*3, and the number increases in turn. The stride and filling coefficient are shown in the Table 1.

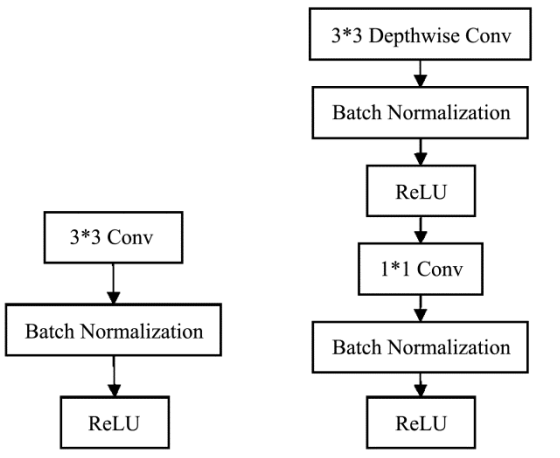


Fig.5 Comparison of Convolution Layer Structure before and after Improvement

Table 1 Improved Network Structure

Convolution Type	Convolution Kernel Size	Strides	Padding
Conv 1	3*3*32	1	1
Conv 2	3*3*64	1	1
Conv dw 1	3*3*64	1	0
Conv dw 2	3*3*128	2	0
Conv dw 3	3*3*128	1	0
Conv dw 4	3*3*256	2	0
Conv dw 5	3*3*256	1	0
Conv dw 6	3*3*256	1	0
Conv dw 7	3*3*512	1	2
Conv dw 8	3*3*512	1	0
Conv dw 9	3*3*512	1	0
Conv dw 10	3*3*512	1	0

In Table 2, it is shown the comparison of backbone network parameters and calculation before and after improvement. It is obvious from the data in the Table 2 that the number of parameters and calculation of the improved model are reduced a lot.

Table 2 Comparison of Parameters and Calculation of the Model before and after Improvement

Backbone Network	Parameters	Calculation
Original	$5.86*10^6$	$1.40*10^{10}$
Improved	$1.15*10^6$	$1.89*10^9$

Secondly, the Openpose algorithm used two branches to generate key-point heatmap and paf heatmap, respectively. The two branches have the same structure, the only difference is in the output phase, where the number of output results differs between the two branches. Therefore, we merge the two branches into one and add a 1*1 convolution to the output stage to split it into two branches. In addition, Openpose algorithm adopts a six stage multi-stage network. The prediction results of the previous stage are purified through the network of the later stage. However, in the actual training process of the model, we found that after stage 1, the performance of the network did not get much improvement, but the amount of computation significantly doubled. However, if only a single stage is used, although the training difficulty and calculation amount are greatly reduced, the effect is not ideal because the output error does not go through multi-layer back propagation. Therefore, this paper chooses to retain the first three stages and output loss in each stage to prevent the gradient from disappearing. The network structure is shown in Fig.6. The structure of Dilated MobileNet V2 in the Fig.6 will be further described below.

Third, in the refinement stage, the Openpose algorithm uses 7*7 convolution kernels. We replace it with a cascaded convolutional structure, using three 3*3 convolutional cascades instead of 7*7 convolution. In the current field of deep learning, more and more

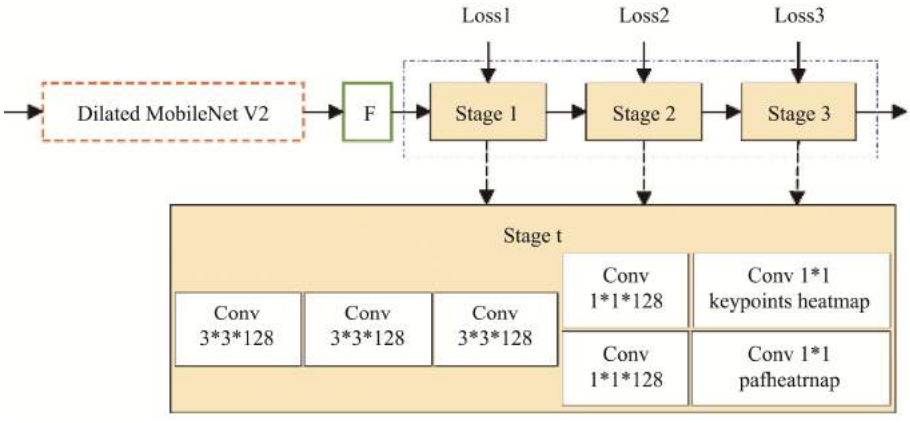


Fig.6 Network Structure of Our Algorithm

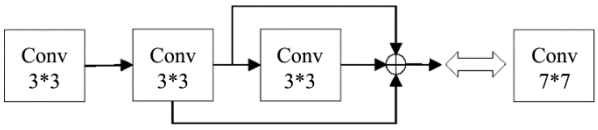


Fig.7 Three 3*3 Small Convolution Kernels Are Used to Replace 7*7 Convolution

researchers use multiple small convolution kernels instead of large convolution kernels. It can not only ensure the network performance, but also reduce the number of parameters and calculation. The repeated stacking of multiple small convolution cores makes multi-layer ReLU functions activated, which also increases the nonlinearity of convolution process and improves the ability of network nonlinear mapping.

The VGG19 network is replaced by adjusting the depthwise separable convolution structure, which greatly reduces the amount of model parameters and calculation. However, in the process of actual training model, we found that a few feature maps will lose some position information, resulting in the key nodes of human body cannot be well detected. Therefore, based on the improved algorithm, we introduce hybrid dilated convolution. Compared with dilated convolution, hybrid dilated convolution can increase the receptive domain and better ensure the continuity of the receptive domain. When it is applied to the existing human pose estimation algorithms for feature extraction, it can not only retain more feature information and ensure the robustness of the algorithm, but also not increase the computational power consumption.

The improved algorithm network structure in this paper has twelve layers, of which the first two layers are conventional convolution layers and the last ten layers are convolution layers based on depthwise separable convolution. Hybrid dilated convolution is introduced into the last three layers of the network, namely layers 10-12. The hybrid dilated convolution of sawtooth structure with dilation rate set to 1, 2 and 5 is added respectively. In this paper, the structure is named Dilated MobileNet V2 structure. The network with HDC structure is shown in Fig.8 below, which is called Dilated MobileNet V2.

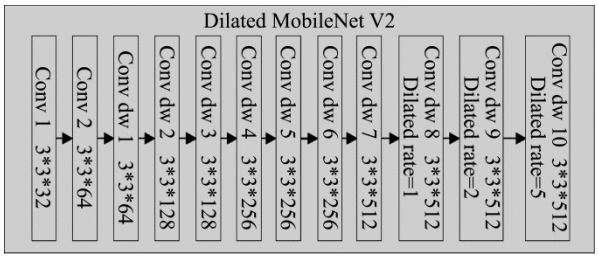


Fig.8 Structure of Dilated MobileNet V2

3.2 Analysis of Falls Constructing Model

Falls are momentary movements. There is a rapid fall process when a person falls. The center of the human body will drop rapidly from a higher position to the ground or near the ground. Changes in the human descent rate and the external contour are two intuitive features when a fall event is detected.

3.2.1 Velocity Analysis

The falling process of a human body can be divided into a weightless phase, an impact phase, and a resting phase. In Fig.9, (a) shows the coordinate system for acceleration and indicates the acceleration along the x, y and z axes respectively. (b) shows the coordinate system for the angular velocity and 3D angle of the human body and represents the angular velocity of the human body around the x, y, and z axes respectively. We use the change in the velocity of movement of the body's center point in the y-axis direction as a feature.

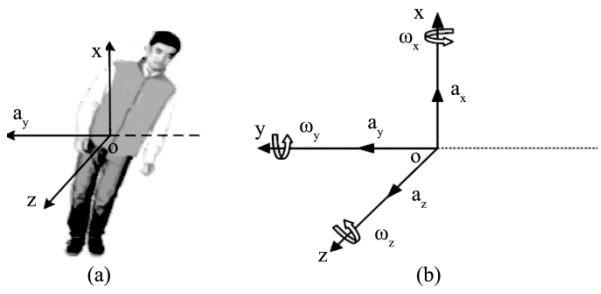


Fig.9 Movement Models of Human Activity.

We selected a total of six key node's locations for reference: the head, the left and right shoulders, the body center position, the left and right knees. Sampling is done once at an interval of 12 frames. The average of

the velocity of the six nodes moving in the y-axis direction between the two sampling frames is used as the body descent velocity.

As an example, in the standing stage, we take the diagonal intersection of the smallest rectangular frame outside of body as the position of the body's center mass. We assume that the coordinates are (x, y), where y is the height of the body central position in the y-axis direction. The rate of descent of the body centroid between two sampled frames is v_y.

$$v_y = \frac{(y_{i+f} - y_i)}{f} \tag{1}$$

y_{i+f} and y_f denote the vertical coordinates of the body center point in the two sampled frames, respectively. f is the time difference between the two sampled frames.

We assume that the coordinates of the 6 joint nodes are $(x_1, y_j) \sim (x_6, y_j)$. Average of the velocities of the 6 nodes moving in the y-axis direction is v.

$$v = \frac{\sum y_{j+f} - \sum y_j}{6f} \tag{2}$$

y_{j+f} and y_j denote the vertical height of the jth node between the two sampled frames, respectively. f is the time difference between the two sampled frames.

The method of our fall detection algorithm for fall speed is as follows:

(1). Falls are instantaneous action in which the body's rate of descent increases dramatically during a fall. Therefore, when we detect a small change in the rate of descent of human activity, we can conclude that it is not a fall. In other words, we determine that the body is non-falling when v is less than the threshold v_m.

(2). We find that the rate of descent of the human body may be more than the threshold v_m when squatting or jumping. However, falls occur only once and do not recur. Sometimes the human body will even syncope after a fall event occur when the body speed will remain at 0 for a certain period of time. (If there is no syncope occurred, the body speed may not be 0, but will very slightly over a certain period.) However, movements such as squatting and jumping do not occur as described above in most cases. The speed changes of the human body after a movement such as squatting or jumping will be greater than after falls. Here we set the speed change threshold to v_n. We carry out several sampling sessions and capture 24 frames per second and calculate the velocity of the human body in the y-axis direction every 12 frames. It takes about 5 seconds for each fall to occur. So, if we detect the behavior that v is more than threshold v_m within 5 seconds and in the following time, v is less than threshold v_n, then we can tentatively identify the behavior as a fall.

In this paper, we select 200 consecutive frames of video sequences with 5 different poses in the same video. Through experiments result as show in Fig.10, the speed differences of five different behavior states including falls are verified, and the corresponding speed curves are obtained. Through the analysis of velocity curve in Fig.10, we can see that the value of v_m is between 0.9-0.95m/s, and the value of v_n is between 0–0.05m/s.

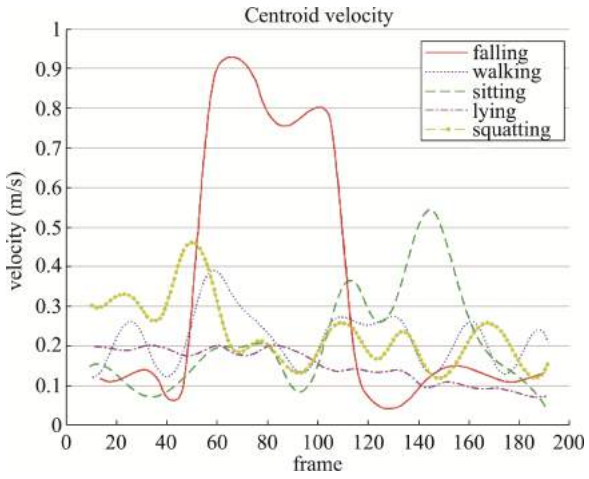


Fig.10 Centroid Velocity under Different Condition

3.2.2 External Contour Analysis

In a few cases, such as when the human body remains in a continuous squatting position after squatting. In such cases, a fall can easily be misjudged by the change in speed of descent of the body alone. So, then, we make a secondary judgment by the aspect ratio of the minimum external rectangular frame of the human body.

We complete the delineation of the human body

with an external rectangular frame, as shown in Fig.11, which is the minimum external rectangular frame of the human body in the standing position.

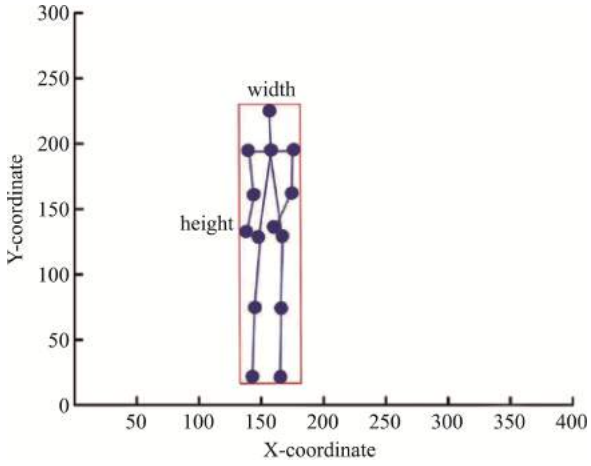


Fig.11 Minimum External Rectangular Frame of the Human Body in the Standing Position

In Fig.12, (a) shows the Minimum external rectangular frame of the human body during a fall, while (b) shows the Minimum external rectangular frame of the human body after a fall.

The most obvious change in the external profile of the human body during a fall can be seen in the ratio of the width to the height of the minimum external rectangular frame of the human body, as it is shown as Fig.13

As Fig.13 shows that the experimental effect of the above video sequence on the improved model. It includes several different human states of standing, walking and falling. In the external rectangular frame above, we set the width to w, the height to h, and the aspect ratio $L = w/h$. When L is more than the threshold L_m, we identify the state as a fall state. We also select 200 consecutive frames of video sequences with 5 different poses in the same video. The experimental results are shown in Fig.14. Through curve analysis, we can see that the aspect ratio of Minimum external rectangular will change dramatically during the fall. This changeset is not found in other behaviors. Combined with the experimental results, we can determine that the value of L_m can be set to 2.5.

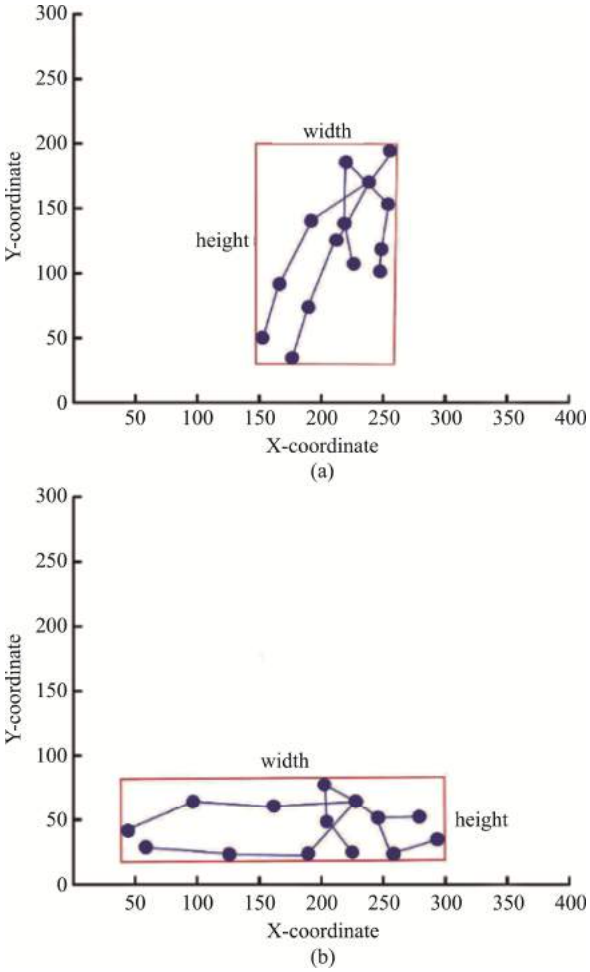
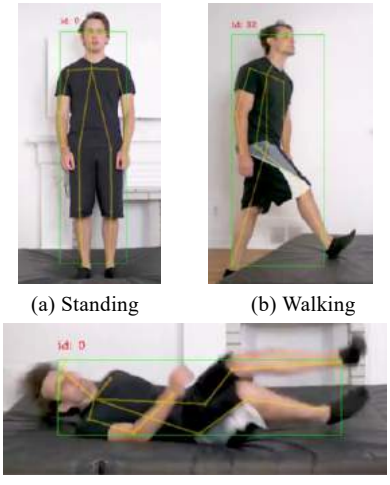


Fig.12 Minimum External Rectangular Frame of Human Body When a Fall Occurs, (a)shows Minimum External Rectangular Frame of the Human Body during Falling, (b)minimum External Rectangular Frame of the Human Body after Falling



(a) Standing (b) Walking

(c) Falling

Fig.13 Experimental Results on the Improved Model

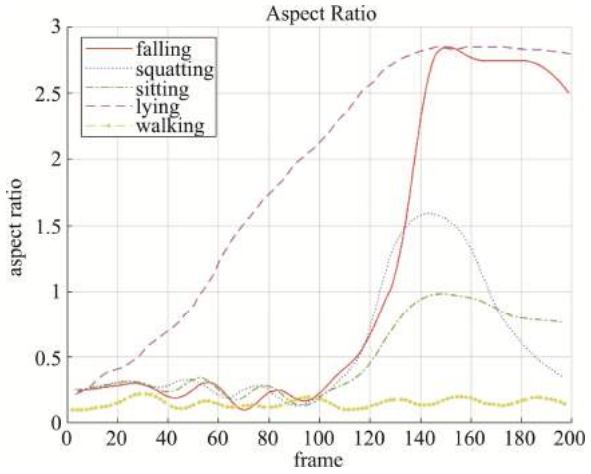


Fig.14 Aspect Ratio of Minimum External Rectangular under Different Posture

In Fig.14, it is worth noting the case of lying. When the human body is lying down, because the human body is gradually approaching the horizontal direction, the final action form is very similar to falling. Therefore, it can be seen from the Fig.14 that when the human body is lying down, the L value is almost the same as that in the case of falling. However, according to the analysis of fall above, fall is an instantaneous behavior with large speed change and short time. Therefore, we can see that the falling state takes less time when the width height ratio L of the external rectangle reaches the maximum value. In other words, the width height ratio curve of the external rectangle in the falling state changes greatly and steeply, and the width height ratio curve of the external rectangle in the lying state is relatively flat.

3.3 Analysis of Falls Constructing Model

Convolutional Neural Networks (CNN) is a kind of feedforward neural network with convolution calculation and depth structure. It is now widely used in many research fields, such as natural language processing and computer vision.

In simple terms, the process of fall detection is equivalent to a classification problem. A serious problem of fall detection using computer vision is that the spatial background information of human body will produce some interference. Although now in many kinds of studies, researchers have invented algorithms that can separate the human background to eliminate interference. However, since we preprocess the data with the estimation algorithm in advance and extract the human skeletal map in different poses, it can help us to avoid the interference caused by the human background to the greatest extent. Using the human key-points map as the input of convolution neural network can remove the background noise to the greatest extent, preserve the human structure, enhance the contrast between the human trunk and the background, and reduce the learning difficulty of the network.

In this paper, we choose Convolutional Neural Networks to design fall detection model. The structure of the model designed in this paper is shown in Fig.15 below.

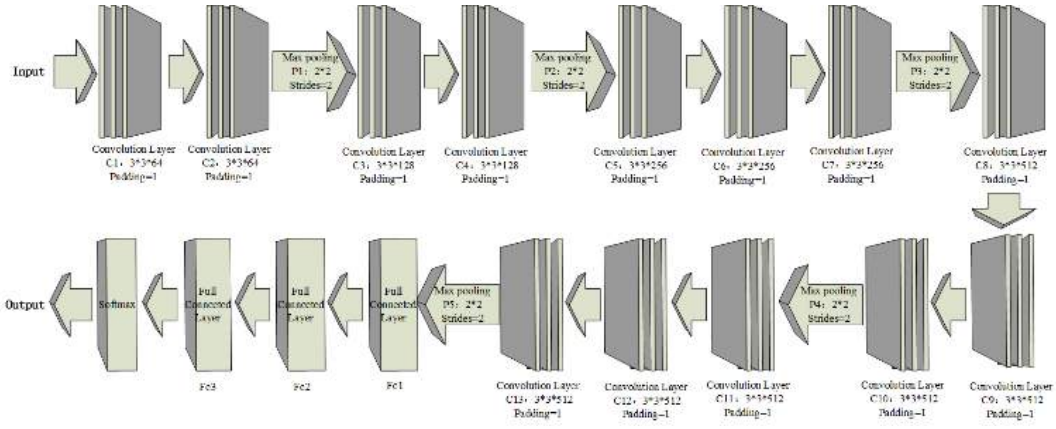


Fig.15 Structure of the Designed Convolutional Neural Networks Model

The model is mainly composed of the convolution layer, the pool layers and the full connections. The size of each frame of the input layer is 224 * 224. In order to enable the action sequence to be used as the input of convolutional neural network, the input layer in the network is changed from receiving 3-channel RGB image to receiving 10 channel action sequence. As shown in Fig.15, in the model, C1-C13 are the convolutional layers. The size of the convolutional kernel is 3*3 and the parameter is set to 1. The number of convolution cores of C1-C2 is 64, that of C3-C4 is 128, that of C5-C7 is 256, and that of c8-c13 is 512. P1-P4 are all pooling layers. The main function of the pooling layer is down sampling, followed by compression features. It can effectively reduce the number of parameters and simplify the complexity of the network. It can also reduce the amount of calculation and expand the perceptual field of vision. The pooling operation used here is maximum pooling, with a step size of 2. This means that the maximum value of the image area is selected as the pooling value for that part. In terms of loss function, in this paper, the binary cross entropy loss function is used as the loss function of the model to express the difference between the predicted value and the true value. Compared with the ordinary cross entropy function, the binary cross entropy loss function can ensure the numerical stability in fall classification.

Deep learning models usually contain many parameters. When the amount of training data is not large enough, it will lead to overfitting. In general, one of the best ways to reduce overfitting is to increase the number of training samples. If there are enough training samples, even the most extensive networks are less likely to suffer from overfitting. Unfortunately, although more research results have been achieved in fall detection research, the number of fall datasets is not nearly large enough. To address this problem, it has been considered that expanding the training datasets is also a practical solution. The data expansion method used in this paper is image flipping.

4 Experiment

4.1 Experiment Dataset

In order to reflect the performance of the improved human posture estimation algorithm. We use COCO datasets[9] to train human pose estimation model. Each human body in the human joint point detection image in the coco data set is marked with the coordinate positions of 17 joint points. Based on the position of joint points marked in coco data set, this paper also adds an additional position of human joint points: two shoulder center points to complete human posture estimation.

In addition, in order to verify the effectiveness of the fall detection model based on pose estimation proposed in this paper, we use three datasets to train and test the model after completing design. These three datasets are publicly available, namely: UR Fall Detection Dataset (URFD)[13], Fall Detection Dataset (FDD)[14], and Multiple cameras fall dataset (Multicam) [15] . The FDD dataset contains 22636 images, of which 16794 images are available for training, 3299 images for validation and 2543 images for testing. The Multicam dataset has 192 videos of simulated falls and normal behavior taken from 8 different angles. In this paper, the Fall videos containing falling behavior are considered as positive samples, while the NotFall videos with other normal behavior are considered as negative samples

4.2 Evaluating Metric

In order to verify the effect of the improved human posture estimation algorithm, we choose to use the detection speed FPS and AP value (Average Precision) as performance indicators to evaluate the performance of the algorithm.

Average precision is an important index to measure the performance of the model in target detection. In this paper, a bottom-up method is used in the design of human pose estimation algorithm, that is, all human key-points on the image are detected first, and then these joint points are connected by cluster analysis. This method tests the ability of clustering analysis.

Object keyword similarity (OSK) is usually used to measure the similarity between true value and predicted value. The calculation formula is as follows:

$$OKS = \frac{\sum_i exp\{-d_p^2/2s_p^2\sigma_i^2\}\delta(v_{p^i} = 1)}{\sum_i \delta(v_{p^i} = 1)} \quad (3)$$

d_p^2 represents the Euclidean distance between the true value and the predicted value of the joint node. p represents the ID of the human body in the truth value. p^i represents the ID of the joint node. $v_{p^i} = 1$ indicates that the joint node is labeled and the label is visible. s_p represents the size of the area occupied by the human body calculated by the human body boundary box in the true value. δ_i represents the normalization factor of the ith joint point, which reflects the influence of the current bone point on the overall prediction. On the basis of OKS calculation results, the value of AP can be further calculated. The concept of AP is to calculate the ratio of the number of OKS greater than *t* to the total number of OKs under a given threshold *t*. When OKS is greater than *t*, the joint point is successfully detected; If it is less than *t*, the test fails. The calculation formula of AP is as follows:

$$AP@t = \frac{\sum_p \delta\,(OKS_p > t)}{\sum_p 1} \quad (4)$$

Otherwise, sensitivity, specificity and accuracy as the judging criteria to verify the effectiveness of the fall detection model based on pose estimation proposed in this paper. We have also compared our method with some other fall detection methods. These three performance indicators are calculated using the following formulae.

$$Sensitivity = \frac{TP}{TP + TN} \quad (5)$$

$$Specificity = \frac{TN}{TN+FP} \quad (6)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

4.3 Results Analysis

4.3.1 Improved Human Posture Estimation Algorithm

We selected MSCOCO2014 dataset to train human posture estimation model. During training, the initial learning rate is set to 0.005, batch_size is set to 128; For every 500 epochs, the learning rate decreased by 0.6 times, and stopped after 3000 epochs.

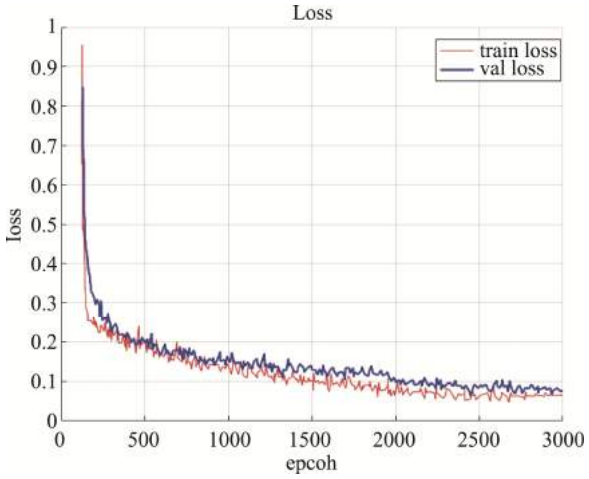


Fig.16 Loss Function Image

After the training, the generated weight model is saved and called. In this paper, we select MSCOCO2014 val dataset to test the running speed of the model. In this paper, the same video frame is used as the input are used for experimental comparison. Input a video sequence with a total of 5500 frames, and get the speed of model processing image through the ratio of video frames to program running time, that is, frame rate. According to the definition of frame rate, the higher the frame rate, the faster the detection speed. Through experiments, we get the ratio of video frames to program running time, that is, the result of frame rate, when the input is a video sequence with a total number of 5500 frames.

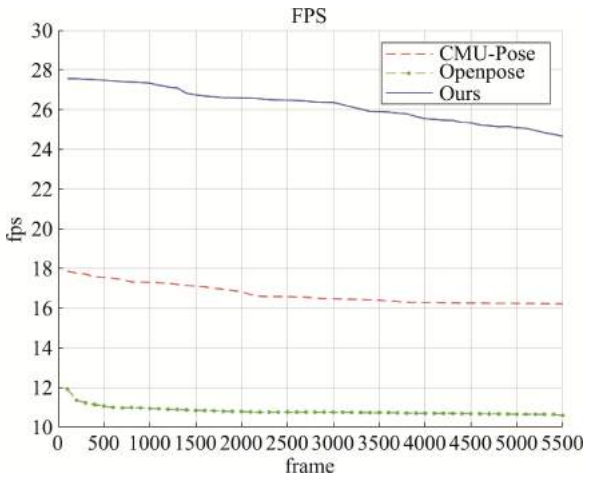


Fig.17 Experimental Results of Frame Rate of Different Models

It shows the experiment results of the same video sequence in different models in Table 3. According to the data in the Table 3, the frame rate of the improved algorithm model is increased by 57.03%.

Table 3 Comparison of Model Detection Speed before and after Improvement

Model	Total Input Frames	Average Frame Rate/fps
Openpose[12]	5500	16.71
Ours	5500	26.24

There is the threshold *t*, when OKs is greater than t, it indicates that the joint node has been successfully detected, and the value of AP can be further calculated. Table 4 shows the results of performance comparison experiments on mscoco2014 test dev dataset. In Table 4, AP represents the average accuracy. According to the target recognition evaluation standard of the International Conference on machine vision: AP^{50} represents the AP value obtained when the threshold *t* is 50%; AP^{75} represents the AP value obtained when the threshold *t* is 75%. In the COCO dataset, about 41% of the images are small-scale images less than 32 * 32. About 34% of the images are medium-sized images, and the image size is between 32 * 32-96 * 96. About 24% of the images are large-scale images larger than 96 * 96. Therefore, when the input image is a medium-scale image, the value of AP obtained by calculating all OKs thresholds is AP^M while AP^L is for the case that the input image is a large-scale image.

Table 4 Performance Comparison of the Improved Algorithm on MSCOCO2014 Test Dev Dataset

Model	AP	AP^{50}	AP^{75}	AP^M	AP^L
Openpose[12]	64.2	86.2	70.1	61.0	68.8
Ours	64.0	85.9	69.9	60.8	69.1

As can be seen from Table 4, compared with Openpose, the AP value decreased by 0.2, decreased by 0.31%, but the AP^L value increased by 0.3, increased by 0.44%. This shows that the performance of the improved model is better when the input image size is greater than 96 * 96. In addition, compared with the Openpose, the average detection speed of the improved model is increased by 57.03%, which can better meet the real-time requirements.

4.3.2 Fall Detection Model

The experimental results of the performance of the model designed according to the method proposed in this paper on the three datasets are shown in Fig.18. According to the final printed results, it is known that: the accuracy tested on the UR fall detection dataset is 98.53%, on the fall detection dataset is 97.71%, on the Multiple cameras falls dataset is 97.20%

The other results of the experiment, including specificity and sensitivity, are shown in Table 5 below.

To reflect the experimental results of the proposed method, we compare the experimental results of other methods and the proposed method in this paper on the same dataset, as shown in Table 6, Table 7 and Table 8.

Table 5 Experimental Results of the Proposed Method in This Paper on Different Datasets

Dataset	Sensitivity	Specificity	Accuracy
URFD	100.00%	98.44%	98.53%
FDD	99.83%	97.58%	97.71%
Multicam	99.80%	97.13%	97.20%

Table 6 Comparison of Our Experimental Results with Others on the URFD Dataset

Method	Sensitivity	Specificity	Accuracy
Ours	100.00	98.44	98.53
Khraief et al[17]	100.00	95.00	-
Harrou et al[18]	100.00	94.93	96.66

Table 7 Comparison of Our Experimental Results with Others on the FDD Dataset

Method	Sensitivity	Specificity	Accuracy
Ours	99.83	97.58	97.71
WANG et al[19]	97.78	97.37	97.33
Charfi et al[20]	98.00	99.60	-

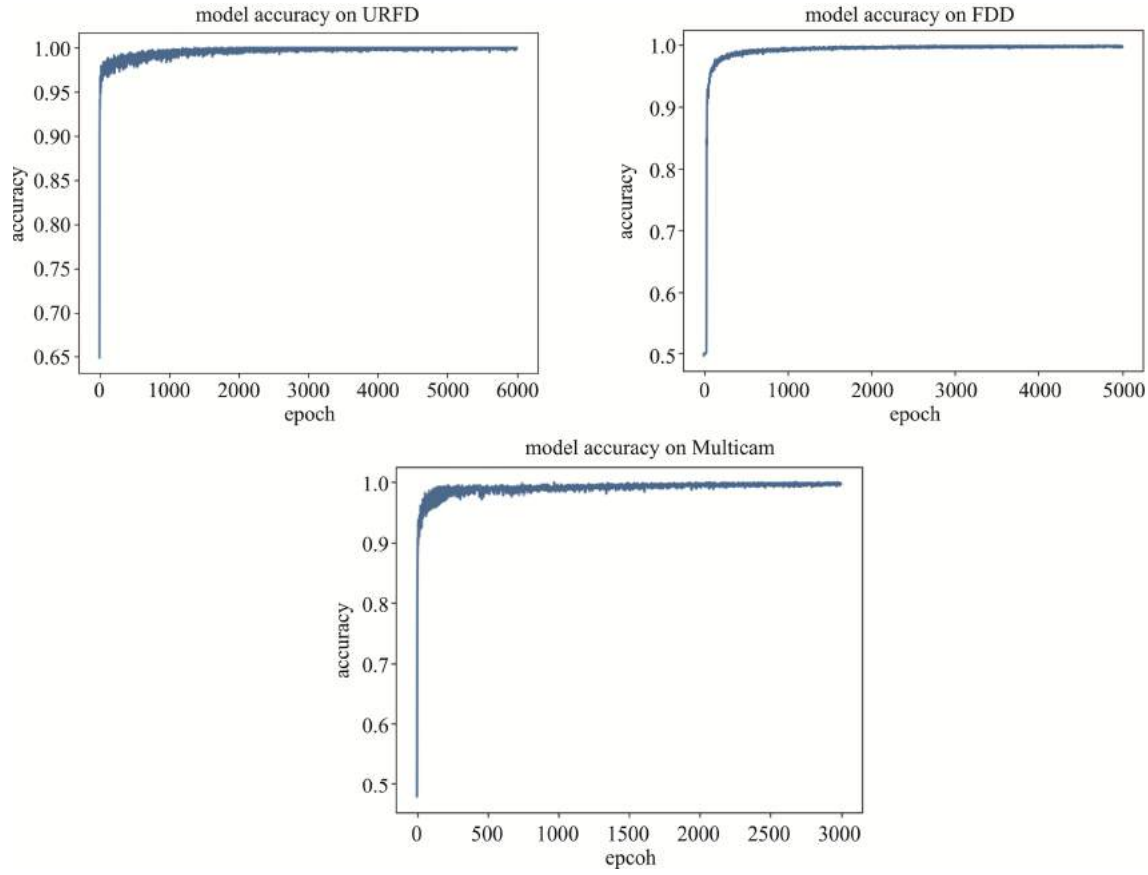


Fig.18 The Accuracy of the Proposed Method on the Three Datasets

Table 8 Comparison of Our Experimental Results with Others on the Multicam Dataset

Method	Sensitivity	Specificity	Accuracy
Ours	99.80	97.13	97.20
Liu et al[21]	98.00	97.50	-
Rougier et al[22]	95.40	95.80	-

From the comparison of Table 6, Table 7 and Table 8, it can be seen that the method proposed in this paper improves the progress while also improving the sensitivity of the system to a certain extent and reducing false detections. In addition, in the actual detection process, this experiment is currently in a low hardware configuration. Nevertheless, the detection speed of the method proposed in this paper for video is around 23–26 fps, up to 28 fps, with low latency, which can basically meet the requirements of real-time detection. This can also provide some help for experiments without higher hardware equipment environment

5 Conclusion

In this paper, we propose a fall detection method based on human pose estimation. First, we combine the depthwise separable convolution and a pyramidal model based on the dilated convolution. A human pose estimation algorithm is obtained by improving on the structure of the original Openpose algorithm. The human posture estimation algorithm is used to track the human body in the video and extract the key-points of the human body under different behaviors and apply the results of human pose estimation to the research of fall detection. Then feature extraction is carried out on the extracted data. The falling speed of human body and the minimum external rectangular frame of the human body are used as the features for multiple

judgments, so as to establish the fall event model and construct the fall features. Next, it is fed into the designed Convolutional Neural Networks model for classification. And the final classification results are applied to fall detection. Based on the experimental results, it is demonstrated that the fall detection method proposed in this paper achieves better detection results on all three publicly available fall detection datasets.On the premise of maintaining the original detection accuracy, the average detection speed of the improved human posture estimation algorithm is improved by 9.7fps. The experimental results on several public fall data sets also prove that the application of human posture estimation results to fall detection has better recognition effect on fall behavior.

References

[1] World Health Organization. Fact sheet on fall injuries in 2018 https://www.who.int/zh/news-room/fact-sheets/detail/falls. URL.

[2] Yu X. Approaches and principles of fall detection for elderly and patient[C]//e-healthNetworking, Applications and Services, 2008. Health Com 2008.10th International Conference on. IEEE, 2008:42-47.

[3] Chong C J, Tan W H, Chang Y C, et al. Visual based fall detection with reduced complexity horprasert segmentation using super-pixel,2015 IEEE 12th International Conference on Networking, Sensing and Control. IEEE, 2015: 462-467.

[4] Z. Bian, J. Hou, L. Chau, and N. Magnenat-Thalmann, “Fall detection based on body part tracking using a depth camera,” IEEE J. Biomed. Health Information., vol. 19, no. 2, pp. 430–439, Mar. 2015.

[5] Rakhman AZ, Nugroho LE, Widyawan, et al. Fall detection system using accelerometer and gyroscope based on smartphone// International Conference on Information Technology, Computer and Electrical Engineering. IEEE 2014:99-104.

[6] Yan YuJuan, Li Hua, Zhao JuMin, et al. Fall Detection System Based on CRFID and Pattern Recognition. Computer Engineering, 2019(6):297-302.

[7] Diego D, Daniele F, Emanuele P, et al. A Combined One-Class SVM and Template-Matching Approach for User-Aided Human Fall Detection by Means of Floor Acoustic Features. Computational Intelligence and Neuroscience, 2017, 2017:1-13.

[8] Zhang Dajun, Lan Henrong, Wu Youlong. Bathroom fall detection based on ultrasonic Doppler effect. Journal of Shanghai Normal University (Natural Science Edition),2018, v.47(02):92-96.

[9] Lin T Y, Maire M, Belongie S, et al. Microsoft coco: Com-
mon objects in context[C]//European conference on com puter vision. Springer, Cham, 2014: 740-755.

[10] Alex Krizhevsky; Ilya Sutskever; Geoffrey E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks.

[11] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. Computer Science, 2014.

[12] Zhe C, Simon T, Wei S E, et al. Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields. IEEE, 2017.

[13] Bogdan Kwolek, Michal Kepski, Human fall detection on embedded platform using depth maps and wireless accelerometer, Computer Methods and Programs in Biomedicine, Volume 117, Issue 3, December 2014, Pages 489-501, ISSN 0169-2607.

[14] Adhikari, Kripesh, Hamid Bouchachia, and Hammadi Nait-Charif. “Activity recognition for indoor fall detection using convolutional neural network.” *Machine Vision Applications (MVA), 2017 Fifteenth IAPR International Conference on*. IEEE, 2017.

[15] E. Auvinet, C. Rougier, J. Meunier, A. St-Arnaud, J. Rousseau, “Multiple cameras fall dataset”, Technical report 1350, DIRO - Université de Montréal, July 2010.

[16] Z. Cao, G. Hidalgo, T. Simon, et al. OpenPose: Realtime multi-person 2d pose estimation using part affinity fields[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2019.

[17] Khraief C, Benzarti F, Amiri H. Elderly fall detection based on multi-stream deep convolutional networks. Multimedia Tools and Applications, 2020, 79(6).

[18] Harrou, Zerrouki F, Sun N, et al. Statistical control chart and neural network classification for improving human fall detection// 2016 8th International Conference on

Modelling, Identification and Control (ICMIC). IEEE, 2017.

[19] Wang B H, Yu J, Wang K, et al. Fall Detection Based on Dual-Channel Feature Integration. IEEE Access, 2020, PP (99):1-1.

[20] Charfi I, Miteran J, Dubois J, et al. Definition and Performance Evaluation of A Robust SVM Based Fall Detection Solution// Eighth International Conference on Signal Image Technology & Internet Based Systems. IEEE, 2012.

[21] Liu J, Xia Y, Tang Z. Privacy-preserving video fall detection using visual shielding information. Visual Computer, 2020(1).

[22] Rougier C, Meunier J, St-Arnaud A, et al. Robust Video Surveillance for Fall Detection Based on Human Shape Deformation. IEEE Transactions on Circuits & Systems for Video Technology, 2011, 21(5):611-622.

Author Biographies

ZHENG Yangjiaozi is now a M.Sc. candidate of Three Gorges University. Her main research interest includes Internet of Things and deep learning.

E-mail: eden33s@163.com

ZHANG Shang received Ph.D. degree from China University of Geosciences (Beijing) in 2014. He is now an associate professor and master supervisor in Three Gorges University. His main research interests include Internet of Things, computer application technology, etc.

E-mail: wetoo@163.com

Failure Analysis of a Crank Hinge in a Guillotine Machine

A. R. LOKUGE[1], R. J. WIMALASIRI[2]

(1. *Mechanical Engineering*, *Department of Mechanical Engineering Kingston University*, *London*;
2. *Mechanical Engineering*, *Department of Mechanical Engineering*, *The Open University of Sri Lanka*, *Nawala*, *Nugegoda*)

Abstract: Guillotine machines are used to cut bulk quantities of paper, often thousands at a time. More the number of papers to be cut at once, more load is required to cut. This machine undergoes a frequent failure of one of its hinges, which prevents the operation of the machine. A combination of torsional forces and bending moments are acting on the hinge when operating. Torsional stresses induced due to the friction between the contacting surfaces of the crank rod and the hinge. The bending moment induced due to the alternating motion and the load acted upon the cutting mechanism. The crank transforms the rotational movement into a translational motion of the blade, which results in the formation of a cyclic load in the form of a sinusoidal with a mean value not equal to zero. This leads to fail the hinge in the mode of fatigue. Naked eye observations of the fracture surface reviled a clear failure initiation point and striation marks of crack propagation and a sudden fracture region which evident a fatigue failure due to cyclic loading. To redesign the failed hinged to avoid such failure, it is essential to, (i) define and evaluate the stresses developed by the combined loading condition (ii) understand the nature of the cyclic stress induced. The force acting on the hinge was calculated by the law of conservation of momentum created by the blades' inertia and its' supportive structure. It was understood that the mean stress value of the cyclic load is not equal to zero, the modified Goodman diagram is used. Computational simulations are conducted using Finite Element Analysis (FEA) on the ANSYS fatigue tool. By applying the fatigue analysis theories and conducting FEA for stress analysis, the reason for the failure is revealed and then necessary precautions could be taken to prevent such failures in the future.

Keywords: Fatigue Fracture**,** Crack Propagation, Cyclic Load, FEA, Mean Stress.

1 Introduction

Guillotine machines are used to cut bulk quantities of paper often thousands at a time. More the number of papers to be cut, more load is required to cut the papers, thereby causing an increase in the required cutting force. Important components in the Guillotine machine (Fig.1) are Flywheel, Induction motor, Clutch, Crank, Cutting mechanism, Pulley, Gears and Hinges. The flywheel is driven by an induction motor via a pulley drive where the motor is in a continuous rotational mode. The flywheel's main link of transforming the energy of the motor, by preventing the submission of any inertia loads. The repeated engaging and disengaging motion of the crank is ensured by the presence of a clutch while a lever ensures that this con-

tinuous mechanism is under control. A pre-adjusted threaded bar at the end of the rod ensures a controlled movement of the blades. The slider and the crankshaft is connected by a connecting rod. The bolt is used as a hinge, to pivot the rod with the crank. The failure of this hinge occurs frequently mainly due to a combination of torsional and bending loads. Torsional stresses are due to the friction acting between the contacting surfaces of the rod and the hinge and the bending moment induced due to the alternating motion of the cutting blade. The crank transforms the rotational motion into translational motion of the blade, which results in the formation of a pulsating load. This load is in sinusoidal pattern with a mean stress value not equal to zero. These can cause failure in the mode of fatigue. By observing the failed components, it was identified glimpses of crack initiations and beach marks of crack propagation and a sudden ductile fracture which ensures a fatigue failure due to cyclic loading. This modified Goodman diagram is used for the calculations and computational simulations is performed on FEA, on ANSYS fatigue analysis tool.

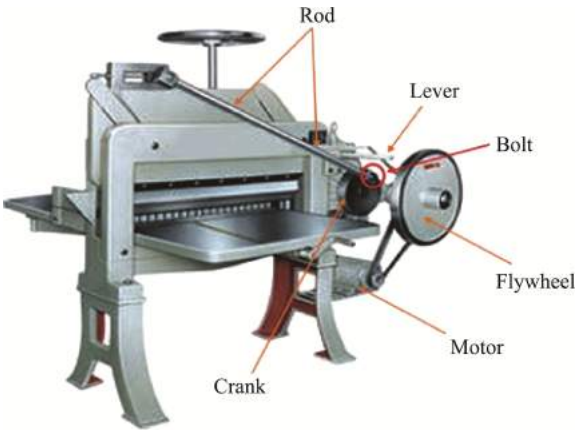


Fig.1 Illustration of Machine (Source: https://toyo-printingmachine.en.made-in china.com/product/mvknrNBHYgcA/China-Dq-201-Mechanical-Paper-Cutting-Machine.html)

2 Literature Review

The process of cycle-by-cycle accumulation of damage in a material that is undergoes fluctuating stresses and strains, is defined as fatigue failure. Any load is not significantly large enough to cause an immediate failure, is one of the significant features of fatigue [1-3]. Failure starts to occur only after a certain number of load fluctuations are experienced, which means that only when the accumulated damage has reached a critical level, then failure initiates [4]. Cracks initiation occurs from the surface of a component due to the fatigue damage, which begins from here in the form of shear cracks on crystallographic slip planes. This is known as crack initiation. After the completion of the transition period of crack initiation, crack propagation takes place in a direction normal to the applied stress [5]. In the final stage, the crack becomes unstable, and fracture occurs. Fatigue crack propagation is the name given to the process that occurs before fracture. Some of the early stages may be skipped and this is entirely dependent on the initial crack length in a component. The “defect free” metal components are the only components which consist of these five stages of the fatigue process. Fatigue damage is fundamentally caused as a result of structural changes at the microscopic level, such as dislocations of the atomic structures. It is justifiable to assume that the microscopic parameters that govern fatigue damage, possess an inherent relationship with macroscopic stress and strain quantities based on the continuum mechanics concepts. Crack nucleation and early growth, both of which can be accounted for, using these macroscopic quantities. The most important component of fatigue analysis is the relationship between stress amplitude and number of cycles which it can execute, before experiencing failure. S-N curves, are used for this kind of analysis where S denotes stress that is indicated by the Y-axis and the logarithmic of the number of cycles is indicated by the X-axis. The curve is sketched using practically obtained values. The shape of the S-N curve represents that of an exponential decay graph, for a lower number of cycles. There is a sudden deviation after a certain point of the number of cycles and then the S-N curve become parallel to the number of cycles. The point at which the S-N curve deviates is defined as the endurance limit of the material. If engineers keep

the maximum stress levels below the endurance limit, then the component achieves infinite fatigue life [6-7].

There are four major stages which are considered as the process of total fatigue life [8-10]. These four stages are:

1. Nucleation or permanent damage caused by substructural and microstructural. This is also known as initiation or nucleation of fatigue cracks.

2. The creation and growth of microscopic cracks, also known as small-crack growth.

3. The formation of "dominant cracks" due to the growth and coalescence of microscopic flaws and the stable propagation of dominant macrocracks. This is known as Macro-crack propagation.

4. The structural instability or complete failure which can be termed as the final fracture.

The period during which cracks initiate from defects and propagate, is defined as the fatigue life of engineering components and structures. The problem of fatigue failure arising becomes a possibility whenever engineering structures operate under severe conditions. It is during the crack propagation stage that the largest fraction of fatigue life is spent. When operating under severe or extreme conditions, the problem of fatigue failure is raised in engineering structures. The prediction of damage caused by corrosion fatigue during the early stages of damage is an outstanding issue that is yet to be solved in the field of corrosion science and engineering.

However, S-N curves are used only for fully reversible cyclic loads which means the mean value of the load fluctuation is equal to zero [11, 12]. If the stresses induced in a pedal of a bicycle were considered, the mean value of that fluctuating load is not equal to zero. In such a case, it may require Goodman diagrams and ultimately, modified Goodman diagrams [13].

The endurance limit of a material can be determined by the RR Moor testing method. This RR Moor testing method is conducted by the rotation of a specimen made from a particular material about its symmetrical axis, via a motor, with a variable imbalanced mass attached to the specimen. Using the number of cycles and stress denoted by the imbalanced mass, the S-N curve is obtained [14-15]. Using this obtained S-N curve, the ideal endurance limit value of the material is obtained. The reason this is called the ideal value is due to the three assumptions made during this test. These three assumptions are:

1. Specimen is homogeneous

2. Absence of any defects

3. The surface of the material is highly

In the real scenario, these three assumptions, may not have any validity attached to them due to which materials can undergo failure before the ideal endurance limit value (S'_e) obtained via RR Moor testing method. This is why, the endurance value is not considered as much as the value of tensile or shear strength, due to its dependency on a lot of other factors.

To determine a more realistic value for the endurance limit [16], the value of the ideal endurance need to be gained by several coefficients that are less than one [17].

The four major factors are:

1. Surface finish factor (K_a) – Effect of scratches on the surface.

2. Size factor (K_b) – Effect of surface defects.

3. Stress concentration (K_d) – Effect of discontinuity.

4. Reliability factor (K_c)

The two minor factors are:

1. Factor due to the loading condition

2. Factor due to the temperature condition

The realistic value can be obtained using equation (1).

$$S_e = K_a K_b K_c k_d S'_e \quad (1)$$

When a component or structure is considered a continuum (assumed as free of any cracks), the S-N approach is very appropriate to perform fatigue analysis. The general representation of fatigue process leading to failure, is represented in terms of the crack initiation stage followed by the crack propagation to a critical size [18]. There is no generally accepted definition as to what constitutes the initiation or when the propagation commences when the initiation phase is completed. The combined sum of both, the crack initiation and the propagation are represented by the net

cycles to failure.

Total fatigue = $N_i + N_p$

Where,

N_i – Number of cycles for crack initiation

N_p – Number of cycles for crack propagation to failure

The hinge point experiences a bending moment due to force acting on it created by the movement of the blade. A torsion also induced between the surfaces of the hinge and the connecting rod due to its Friction force. The blade is coming down at an angle of 80 degree and it's called as the shearing angle for the paper to be cut. Total work is done by the flywheel and the total stresses are acting on the hinge. Illustration of the force diagram is shown in Fig.2.

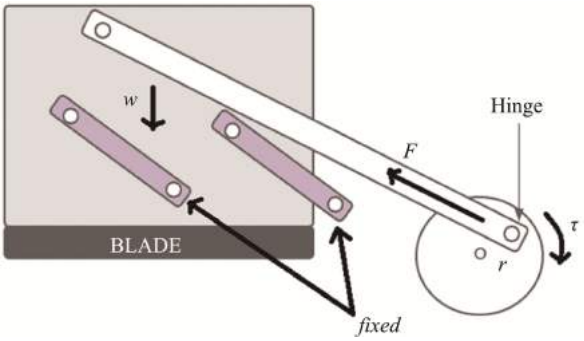


Fig.2 Illustration of the Force Diagram

3 Conceptual Approach to Final Solution to the Problem

Fatigue damage is a gradual process which increases in direct proportionality to the applied cycles of loading in a cumulative manner, that may lead to fracture. Cumulative fatigue damage (CFD) is the fundamental theory which is the framework that is used for fatigue strength assessment (FSA). Recent research activities and development work, mooted the, fatigue crack propagation (FCP) theory whose fundamentals were totally based on the fracture mechanics concepts. The fundamental cause of fatigue damage is the material structural changes, which take place at the microscopic level, such as dislocations of the atomic structures. The long term problem is that of the collating microscopic quantities as well as the macroscopic experimental observations, though it is still possible to percept that the microscopic parameters which govern fatigue damage, possess an inherent relationship with the macroscopic stress and strain quantities [19, 20] , that are fundamentally based upon the continuum mechanics concepts. Its these macroscopic quantities that can be used to identify crack nucleation and early crack growth.

On observation of the fracture surface, it is assessed that the failure is due to fatigue. Two different cross sections are shown in Fig.3.

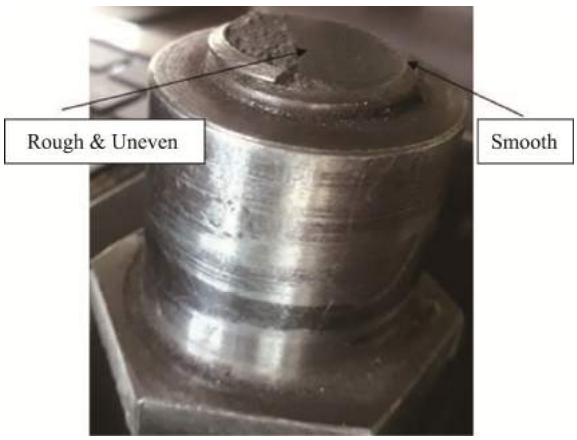


Fig.3 Two Different Cross Sections

It is observed from the Fig.3, that there was a time period during which this component had experienced fracture but due to cyclic loading, the surface has been ploughed due to the repeated contact between the two surfaces [21-22]. This causes polishing of the surface to take place. The crack tip initiates at the position as indicated in Fig.4, provided below and Process of Crack propagation on the fracture surface is shown in Fig.5.

Crack propagation occurs, across the surface in the direction of the red arrow as indicated in Fig.5, above.

Brittle Area of the fracture surface is shown in Fig.6

Fig.4 Initiation of the Crack Tip on the Fracture Surface

Fig.5 Process of Crack Propagation on the Fracture Surface

Fig.6 Brittle Area of the Fracture Surface

A relatively smooth surface was observed for approximately around 75% of the total fracture surface. This was associated with striations of a propagating ductile fatigue crack. A rough and uneven texture that was associated with the final brittle fracture, was observed on the remaining 25% of the total fracture surface. Noticeable deformation was observable in final failure. Presence of several possible crack initiation sites was visible on the fracture surface though they cannot be easily identified due to the shiny nature of the surface, but it was one main crack which lead to the failure. Presence of beach marks on the main crack were representations of it being a typical propagating fatigue crack. The cause of this propagating fatigue crack is most likely to be the mechanical damage incurred by the fracture surface due to the opening and closing of the fracture prior to final failure. The position of the initiation of the cracks was at a diameter transition in close proximity to the thread. The location of the area of the final failure is approximately at a 150 offset from that of the associated beach marks of the main crack.

In a fatigue failure, there are three steps. Initially it begins with a crack initiation. This can be due to surface defects and scratches on the surface. The initiated crack starts to propagate, causing beach marks relatively perpendicular to the direction of crack initiation. Finally, a combination of a reduction in area and a high stress concentration factor, leads to the occurrence of fracture which eventually causes the complete failure of the specimen. By observing the failed component, a rough area with less cross section was identified and a smooth area of more cross section since the specimen was under a cyclic torsion and a bending moment. It is assumed that the shined area was due to the relative motion of surfaces which was caused by torsion thereby while crack propagation was taking place, the beach marks were worn out and left with a shiny surface.

Since the specimen is hinged at the crank, in a slider crank mechanism, the total weight of the system is on that hinge, which acts as a bending moment. When it rotates, a torsion is induced by the friction of

the hinge. Cyclic loading failures are analyzed using several methods that include S-N curves for stress life approach, ε-N curves for strain life approach and finally LEFM theories based on stress intensity factor. But with lack of resources, the approach is only based on stress life approach (S-N curves).

The S-N curves are based on general data and published data which makes it the easiest approach. The theory focuses on designing components for infinite life by preventing crack initiation with strength criteria, but this does not describe crack growth. Cyclic loads are classified into two types based on its mean value. For loads which the mean value is equal to zero, are analyzed using S-N curves and the loads which the mean value is not equal to zero are analyzed using Goodman diagrams. In this case, the specimen has experienced a constant load on the hinge which causes a bending moment, when the crank rotates the hinge experiences a sinusoidal change of bending moments acting on it. Therefore, our analysis neglected the Goodman diagram. Goodman diagrams originated from S-N curves. Those are, Gathering Data, Maximum yield strength , Maximum tensile strength, Surface finish factor, Reliability factor, Loading factor, Ideal endurance limit, Diameter of the crank, Length and weight of the rod, Weight of the blade, Simulations, Bending Stress, Pre-Calculations, Area, Mean Stress, Alternating stress, Corrected endurance limit, Drawing of S-N diagram, Drawing of Goodman diagram, Modified Goodman diagram, Determination of factor of safety and Determination of suitable diameter for a predefined safety factor.

For simplification of calculation purposes, the above two-dimensional arrangement was drawn using the actual data. The blade comes down at an angle of 80^0 with respect to the horizontal axis and the crank does all the work regarding its upwards and downwards, motion. While performing this work, the hinge takes part of the bending moment created by the force on it, with respect to the travelling direction of the blade. Using velocity and acceleration diagram, it was obtained, that the maximum velocity occurring at the crank angle is at 10^0 and 190^0, to the horizontal axis. Since the direction of the velocity of the crank (point B), is the same as that of the blade's velocity and the direction of its travel, the maximum speed occurs at those points. Dimensions & Arrangement of the system is shown in Fig.7.

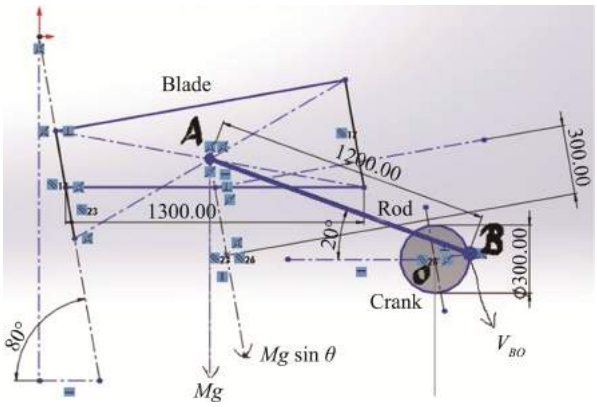


Fig.7 Dimensions & Arrangement of the System

Initially it was obtained that the crank takes 0.9 seconds to complete one complete cycle. In accordance to this obtained value, the speed of point B was calculated as 7 $rads^{-1}$. Finally, the force required to move the blade was calculated by using the law of conservation of momentum. Finalized Free Body Diagram is shown in Fig.8.

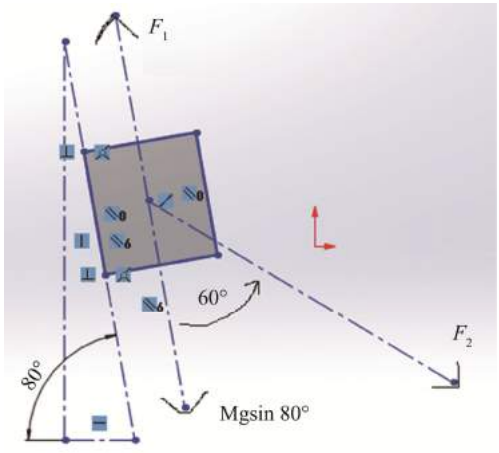


Fig.8 Finalized Free Body Diagram

$$F = \frac{mv - mu}{t} \tag{2}$$

alongthelineof80^0

Downwardstothehorizontalaxis

$$F_2 \cos 6\,0^0 = F_1 \quad (3)$$

Using equation (2), the required force on the blade, that is needed to change its direction, is calculated as 467 N. Considering the free body diagram using equation (3), the total force along the line of the crank rod as 934 N was observed. The addition of the weight component along the line of the rod enables the total force to reach 1426 N. The effect of any frictional forces was neglected, during the calculations. The maximum force acting on the blade, is 1426 N. Cyclic Force on Hinge is shown in Fig.9 and Solidworks CAD Model is shown in Fig.10.

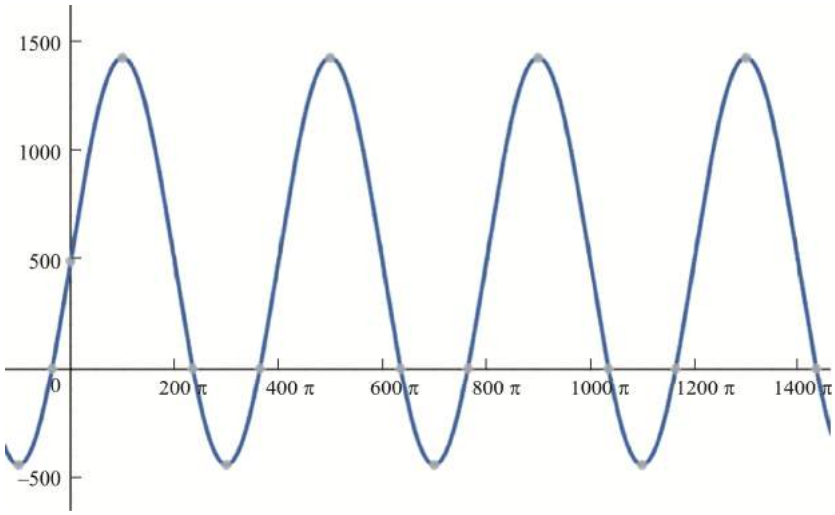


Fig.9 Cyclic Force on Hinge

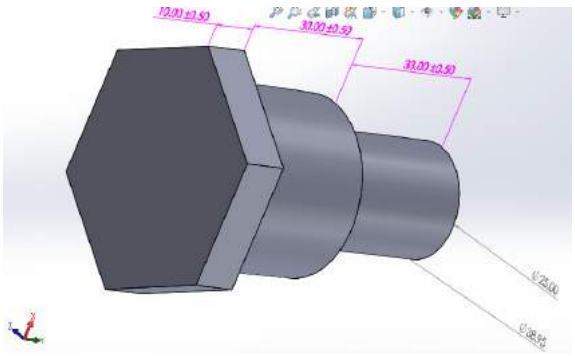


Fig.10 Solidworks CAD Model

4 Analysis

When considering Fig.8, it is very clear that the hinge is experiencing cyclic loading, which has a non-zero mean value. For further calculations, the maximum load on the hinge was needed in order to create the maximum principal stress on the hinge. Using a static structural model created via ANSYS, it was able to obtain the maximum principal stress on the hinge, which is 9.49 MPa.

Data

1. Structural Steel
   a. Maximum Tensile Strength = 360 MPa
   b. Maximum Yield Strength = 240 MPa
2. Factor of Safety = 3 (6 for static approximation)
3. Surface Finish = 63 μ inches (Critical lathe finish)
4. Loading Factor = 0.7
5. Size Factor = 1
6. Temperature Factor = 1
7. For 95% reliability; reliability factor = 0.868
8. Design for the room Temperature
9. Number of cycles for the design

Life cycle S-N Diagram is shown in Fig.11.

The ideal endurance limit of a material is determined by dividing the maximum strength into half. The corrected endurance limit is obtained using multiplying factors of loading, sizing, finishing, reliability, which are less than one, with the ideal endurance limit. Now the alternating stress component can be determined by according to the equation (4).

Data

1. Fatigue Strength Coefficient = 564.4 MPa [4]
2. Fatigue Strength Component (b) = - 0.0576 [4]

$$\sigma_A = \sigma_F'.(2 \times N)^{-b} \quad (4)$$

Hence, the Goodman diagram can be drawn using above data. Goodman Diagram is shown in Fig.12.

According to the simulation results of the specimen, using ANSYS, the maximum principal stress as, 3.6 MPa and the minimum principal stress as, 1.2MPa were obtained. Using the minimum and maximum forces as, 467N and 1476N.

Goodman Diagram describes the interplay of mean stress with alternating stress at failure, on a given number of cycles.

$$\sigma_{mean} = \frac{\sigma Min_{Max}}{2} \quad (5)$$

$$\sigma_A = \sigma_{Max} - \sigma_{mean} \quad (6)$$

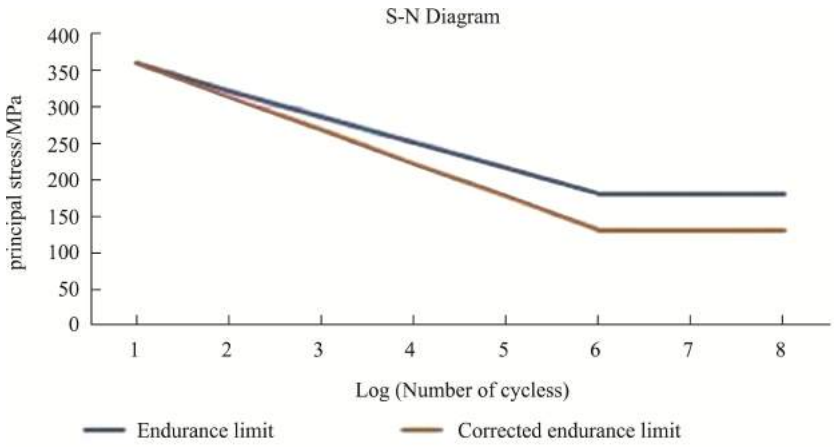


Fig.11 S-N Diagram

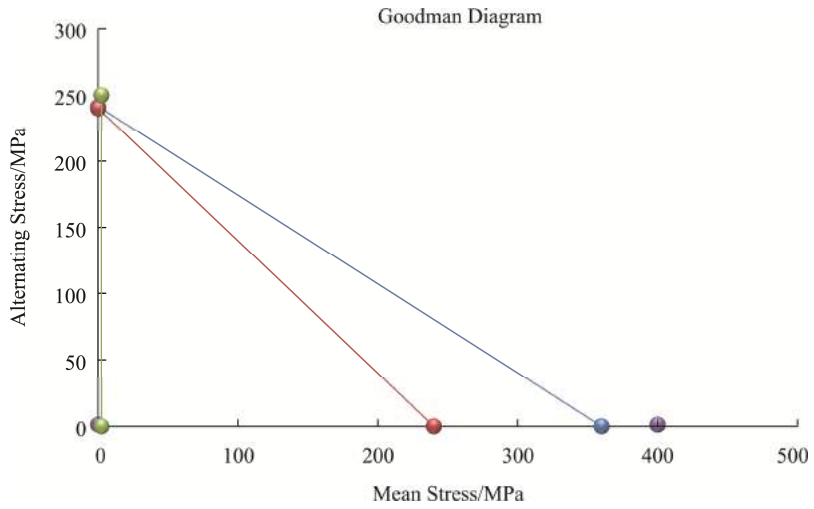


Fig.12 Goodman Diagram

According to the equation (5) and (6), Mean Stress was 2.4 MPa and Alternating Stress was 1.2 MPa. Using the ANSYS engineering data of structural steel, the following S-N curve was obtained which is almost similar to our theoretically obtained S-N curve. Similarly, the curve knees on 10^6 cycles. ANSYS derived S-N Diagram is shown in Fig.13.

In the finite element model, the patch confirming method was used to create the mesh with tetrahedral elements due to the uneven shape of the specimen and the mesh size being set to default. After the introduction of a minimized sizing method, the results did not show any deviation, therefore the results shown in this section are based on the patch confirming method.

The boundary condition was set to fixed support for the curved area of the smaller diameter belonging to the specimen and a vector component is defined along the Y-axis, as the acting force, to cause the bending moment. Using the fatigue tool of ANSYS, a cyclic load was developed with the loading type chosen as the ratio and substituting values for the loading ratio and scale factor as 0.316 and 1, respectively.

The solution was developed using fatigue analysis life set to stress life as well as the mean stress theory set to Goodman, fatigue strength factor set to 1. The results of the solution were taken as life, safety factor and fatigue sensitivity using fatigue tools and after equivalent stress, total deformation and maximum principal stress. Stress Life approach using Goodman Diagram is shown in Fig.14.

Ratio and the Scale Factor Data Entered is shown in Fig.15, Applied Force and its direction is shown in Fig.16 and Fixed Support for the Minor Diameter Surface is shown in Fig.17.

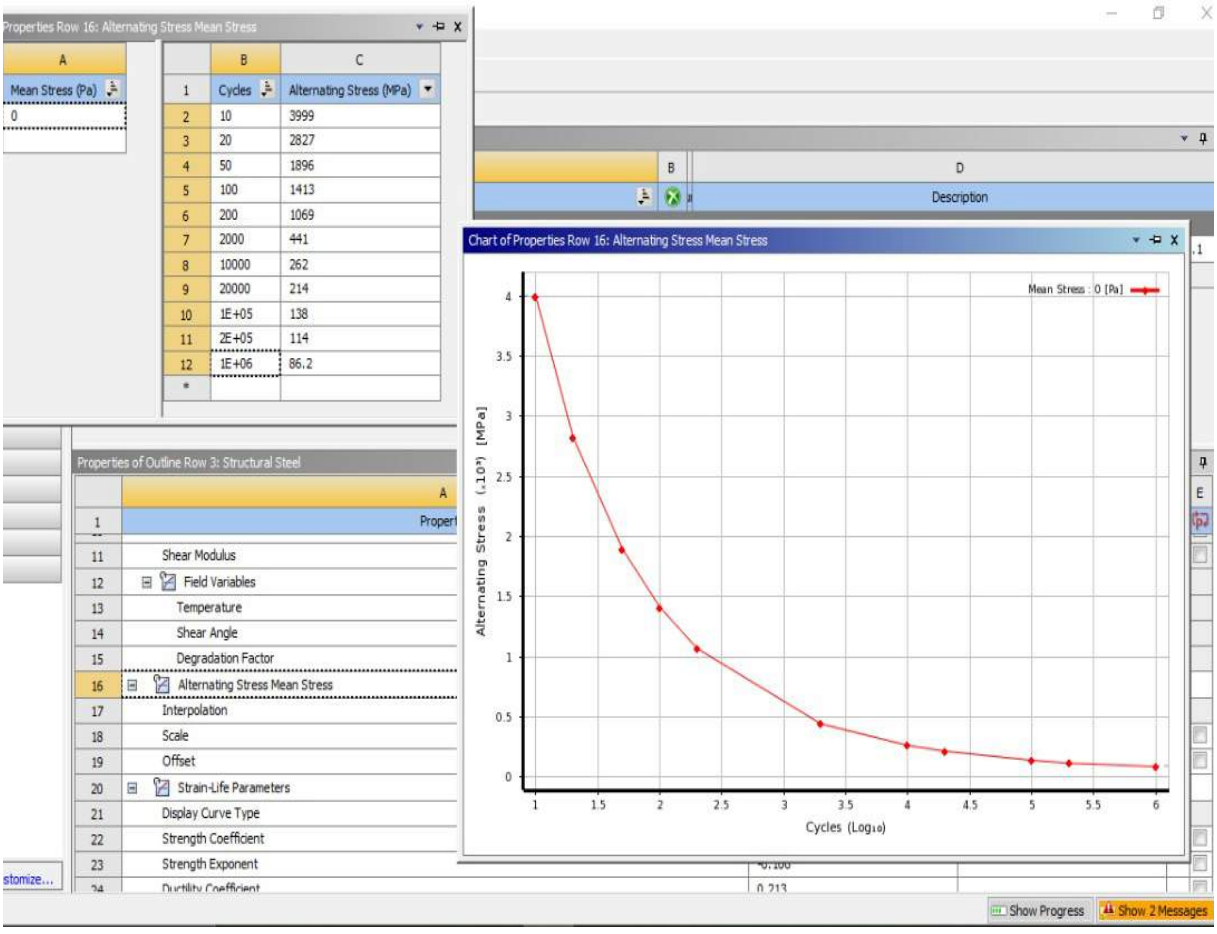


Fig.13 ANSYS Derived S-N Diagram

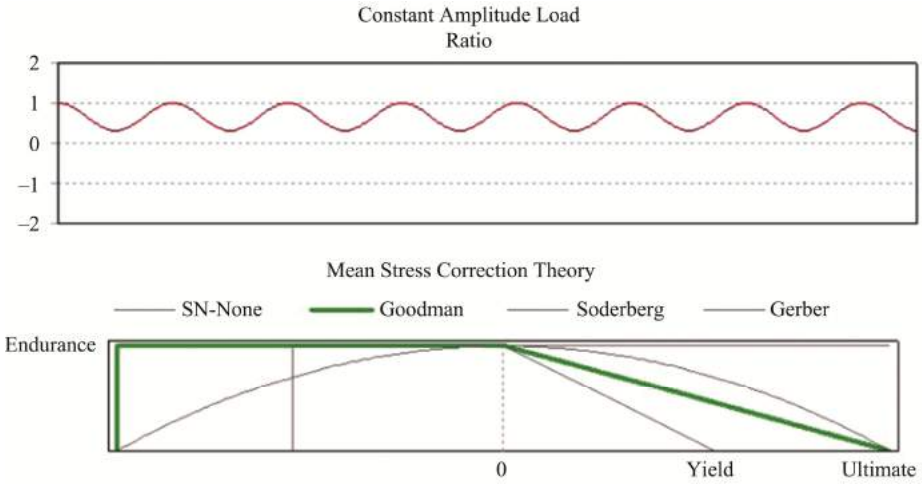


Fig.14 Stress Life Approach Using Goodman Diagram

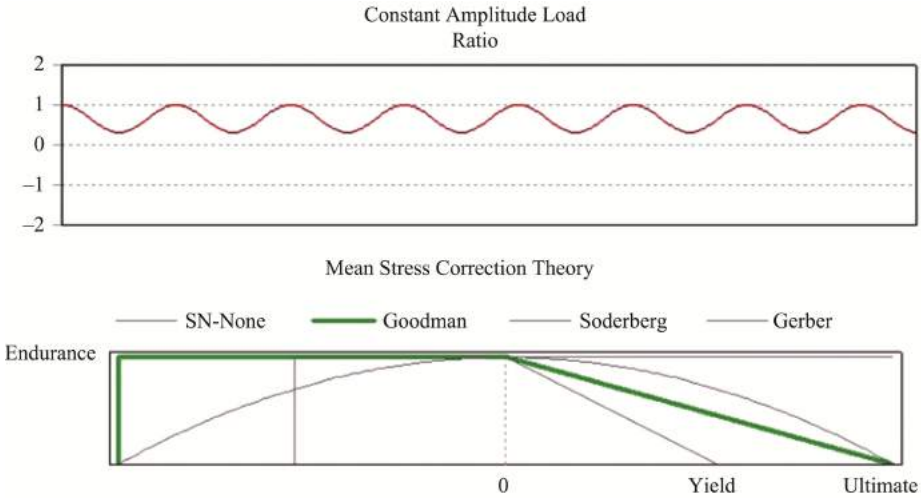


Fig.15 Ratio and the Scale Factor Data Entered

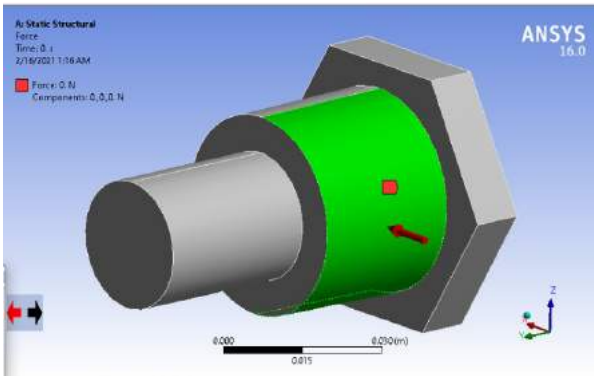


Fig.16 Applied Force and Its Direction

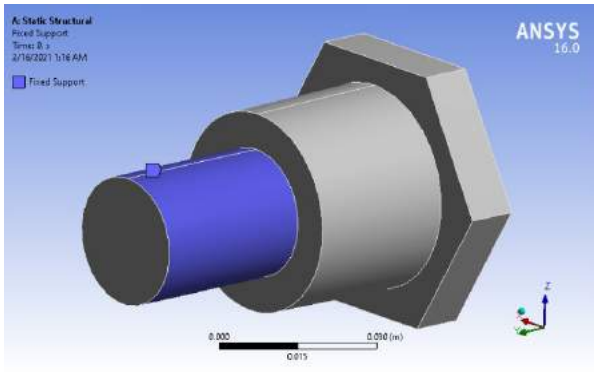


Fig.17 Fixed Support for the Minor Diameter Surface

5 Results & Discussion

As per the Goodman diagram, by plotting a vertical line along the point of 1.2, of the mean stress axis, a factor of safety for tensile stress of approximately 200, was obtained. Similarly, a factor of safety of approximately 200, was obtained, considering the tensile strength. At the beginning of this research, the force calculations were highly dependent on the force diagrams data and the weight of the blade and the friction of the surfaces and negligence of torsional forces on the hinge. Therefore, the resultant data, did not denote any clue of failure.

The cause of failure might be due to the occurrence of a combination of bending moments and torsional stresses but due to lack of data on the torsional stress, the research's focus was narrowed down to effects and analysis of the bending moments. Another consideration is the force it requires to make the cut of thousand or two thousand paper sheets were also not included. Sensing or detecting the torsional stresses, friction force, installation errors, & Cutting Jerk requires more expensive investigation and instrumentation. Also, there can be an integrated offset force due to wear and tear of old machine parts. Reasons behind these obtained highly unusual large values of safety factor can be due to above mentioned scenarios. Even with the increase of loading by 1500% in the fatigue sensitivity graph, a reduction in the number of cycles to the failure, was barely observed. Fatigue Sensitivity Curve is shown in Fig.18.

This indicates that the reason behind the frequent failure of the hinge was not caused by the bending moments. As recommendations for a higher fatigue life of the hinge are constant lubing of machine parts, take great care while installing them, and execute preventive maintenance for greater machine part life, so the wearing are issues crated by them wont effect to the fatigue life of the hinge. Fatigue Life is shown in Fig.19, Fatigue Safety Factor is shown in Fig.20, Maximum Principal Stress is shown in Fig.21 and Total Deformation is shown in Fig.21.

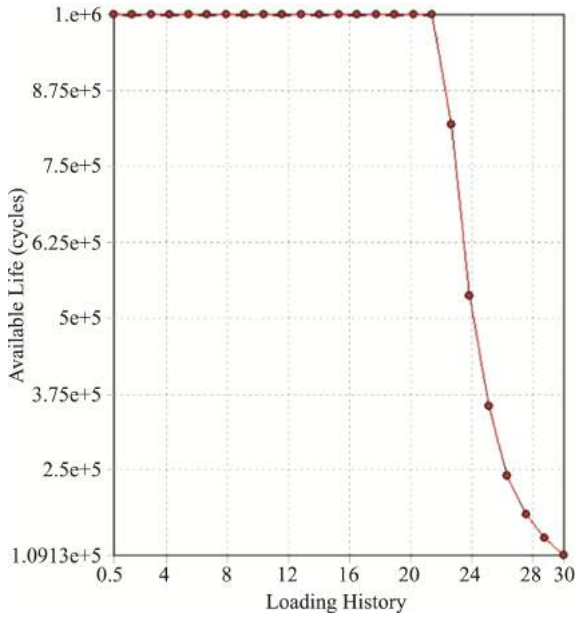


Fig.18 Fatigue Sensitivity Curve

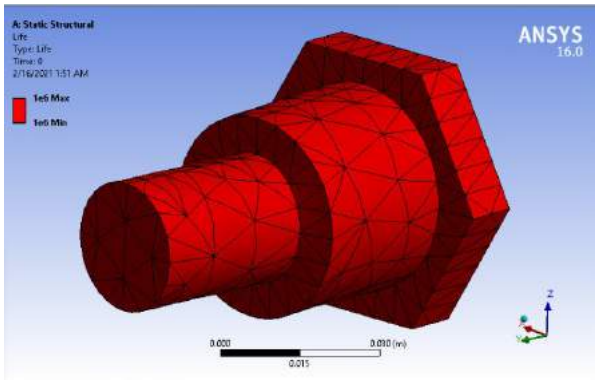


Fig.19 Fatigue Life

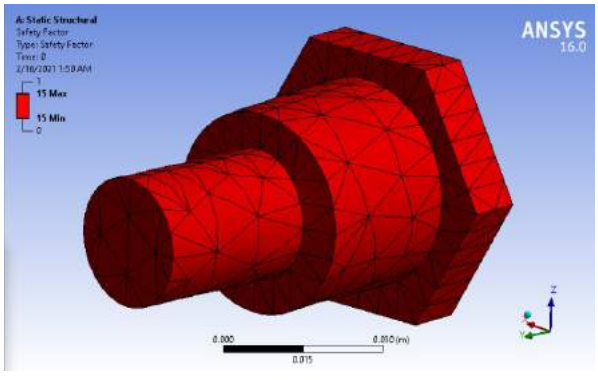


Fig.20 Fatigue Safety Factor

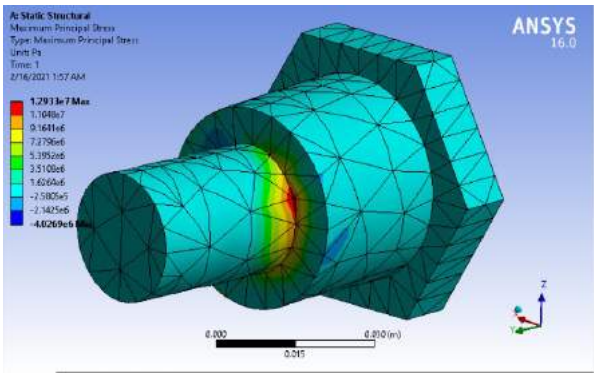


Fig.21 Maximum Principal Stress

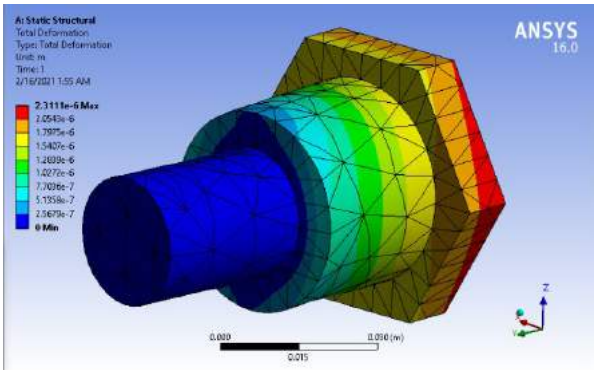


Fig.22 Total Deformation

6 Conclusion

The mode of failure of the hinge is fatigue which caused under cyclic loading conditions. Clear evidences are shown in the fracture surface to prove the fatigue failure. But the stress analysis shows the maximum stress induced on the hinge at working conditions are well below the fatigue endurance limit. Further investigations are needed to conclude the course of failure. It could be due to material imperfection or failure could be initiated from an embedded foreign particle at the surface of the hinge which needs SEM investigation to prove. Also, it could be any form of surface defect which could be occurred at the time of assembling the hinge. There is a possibility of having an impact load or sudden jerk on the hinge due to some operational malfunction.

Failure analysis is always giving unexpected results which needs further investigations. But the important aspect of it is that these investigations indicate at least minor scientific evidences which lead to prevent such failures.

References

[1] Society of Automotive Engineers, S., 1986. SAE Fatigue Design Handbook. 3rd ed. Illinois: Society of Automotive Engineers.

[2] Tawancy, H., Hamid, A. and Abbas, N., 2004. Practical Engineering Failure Analysis. 1st ed. New York: Marcel Dekker Inc.'

[3] Cui, W., 2001. A state-of-the-art review on fatigue life prediction methods for metal structures. Journal of Marine Science and Technology, [Online]. DOI: 10.1007/s007730200012,2,3. Available at: https://www.researchgate.net/publication/225664300_A_state_of_the_art_review_on_fatigue_life_prediction_methods_for_metal_structures [Accessed 14 February 2021].

[4] Pan WF, Hung CY, Chen LL (1999) Fatigue life estimation under multiaxial loadings. Int J Fatigue 21:3–10.

[5] Foreman RG, Kearney VE, Engle RM (1967) Numerical analysis of crack propagation in cyclic-loaded structures. J Basic Eng 89:459–464.

[6] Cruces, A., Crespo, P., Moreno, B. and Antunes, F., 2018. Multiaxial Fatigue Life Prediction on S355 Structural and Offshore Steel Using the SKS Critical Plane Model. MDPI, [Online]. Metals 2018, 8, 1060, 2 & 3. Available at: https://www.mdpi.com/2075-4701/8/12/1060 [Accessed 14 February 2021].

[7] Wang, E., Nelson, T. and Rauch, R., 2004. Back to Elements - Tetrahedra vs. Hexahedra. CAD-FEM GmbH, [Online]. CAD-FEM, 2-14. Available at: https://docplayer.net/21829223-Back-to-elements-tetrahedra-vs-hexahedra.html [Accessed 14 February 2021].

[8] V.J. Colangelo and F.A. Heiser, Analysis of Metallurgical Failures (2nd ed.), p. 66 John Wiley and Sons, Inc., 1987.

[9] Julie A. Bannatine, Jess J. Comer, James L. Handrock, Fundamentals of Metal Fatigue Analysis, Prentice Hall, Englewood Cliffs, New Jersey 07632, ISBN 0-13-340191-X.

[10] Suresh S. Fatigue of Materials, 2nd Edition, University of Cambridge, Cambridge, 1998; ISBN 0-521-57847-7.

[11] Mughrabi H, Microscopic Mechanisms of Metal Fatigue,

Strength of metals and alloys (Edited by P. Haasen, V. Gerold, and G. Kostorz), 1980; vol. III (Pergamon, Oxford) p. 1615-38.

[12] David Taylor, Modelling of Fatigue Crack Growth at the Microstructural Level, Computational Materials Science, 25 (2002) 228-236.

[13] Ong JH (1993) An improved technique for the prediction of axial fatigue life from tensile data, Int J Fatigue 15:213-219.

[14] Miller KJ. O'Donnell WJ., The fatigue limit and its elimination, Fatigue Frac Engng Mater Struct, 22, 545-57, 1999.

[15] Bathias C., There is no infinite fatigue life in metallic materials, Fatigue Frac Engng Mater Struct, 22, 559-65, 1999.

[16] Fatigue and Fracture, ASM handbook, Volume 19, ASM International, 1996; ISBN 0-87170-385-8.

[17] Fatigue Design Handbook, Second Edition, Society of Automotive Engineers, Inc., 1988; ISBN 0-89883-011-7.

[18] Fujimoto Y, Hamada K (2000) Propagation and non-propagation of small fatigue cracks. Proceedings of the 10th International Offshore and Polar Engineering Conference, Seattle, USA, May 28–June 2, pp 180–187.

[19] Brown LM, Ogin SL, Role of internal stress in the nucleation of fatigue cracks, Theory of Materials Group, University of Sheffield.

[20] Mughrabi H, Microscopic Mechanisms of Metal Fatigue, Strength of metals and alloys (Edited by P. Haasen, V. Gerold, and G. Kostorz), 1980; vol. III (Pergamon, Oxford) p. 1615-38.

[21] Kocanda S (1978) Fatigue failures of metals. Sijthoff and Noordhoff, International Publishers, Netherlands

[22] Forsyth PJE (1969) The physical basis of metal fatigue. American Elsevier, New York.

Author Biographies

A. R. LOKUGE received his BEng (Hons) degree in Mechanical Engineering from Kingston University, London and B. Tech. Eng. (Hons) degree in Mechatronics Engineering from the Open University of Sri Lanka.

E-mail: randimaaruna@gmail.com

R. J. WIMALASIRI received his B. Tech. Eng. (Hons) degree from the OUSL and Ph.D. degree from Sheffield Hallam University, United Kingdom. He is currently a Senior Lecturer in Mechanical Engineering at OUSL. His main research interest includes structural integrity of machine elements.

E-mail: rjwim@ou.ac.lk

Research on High Altitude Remote Sensing Building Segmentation Based on Improved U-Net Algorithm

SHI Mengyuan, GAO Junchai

(*College of Electronic and Information Engineering*, *Xi'an Technological University*, *Xi'an* 710021, *China*)

Abstract: Building extraction from high resolution remote sensing image is a key technology of digital city construction[14]. In order to solve the problems of low efficiency and low precision of traditional remote sensing image segmentation, an improved U-Net network structure is adopted in this paper. Firstly, in order to extract efficient building characteristic information, FPN structure was introduced to improve the ability of integrating multi-scale information in U-Net model; Secondly, to solve the problem that feature information weakens with the deepening of network depth, an efficient residual block network is introduced; Finally, In order to better distinguish the target area and background area in the image and improve the precision of building target edge detection, the cross entropy loss and Dice loss were linearly combined and weighted. Experimental results show that the algorithm can improve the image segmentation effect and improve the image accuracy by 18%.

Keywords: Remote Sensing Image, FPN, Efficient Residual Block Networks, Loss Function

1 Introduction

High-resolution remote sensing image is an important national land resource, and it is of great significance to use it to extract buildings for economic forecast, urban digital construction and national defense construction[1, 9].

Extracting buildings from remote sensing images is essentially assigning semantic markers to each pixel[4]. In the traditional feature extraction of remote sensing image, the feature information of the image is generally obtained by artificial or machine learning methods, and then segmented and extracted[6, 8].

Cao[3] etc., they are problems in high-resolution remote sensing images, such as the large number of ground objects and the complex feature information, the segmentation edge is not clear and the object details are lost. In the pre-processing stage before image segmentation, the initial segmented image is obtained by using superpixel segmentation technology. In the process of region merging, objects are merged based on the heterogeneity among objects and the homogeneity within objects, combining with spectral, texture and shape features. The final image segmentation result is obtained by adjusting the global segmentation parameters to adjust the merging scale. Sun[4] etc., encoder-decoder frameworks are popular in semantic image segmentation, however, encoder-decoder models face two main problems. The one is structural stereotype which is receptive fields imbalance rooted in this kind of frameworks. The other is insufficient learning that deeper neural networks tend to encounter the notorious problem of vanishing gradients, suppress the adverse consequences of structural stereotype as far as possible. To alleviate the problem of insufficient learning, we propose a novel residual architecture for encoder-decoder models. Zhou[10] etc., since the traditional deep convolutional neural net-

work segmentation of high-resolution remote sensing images requires manual design of network architecture, it is excessively dependent on expert experience, time-consuming and laborious, and has poor network generalization ability. The resource balance item is added to the network architecture parameters to improve the stability of the search algorithm and reduce the update imbalance and discretization error generated in the pruning process. Secondly, some channels are selected for mixed operation of search space to save computing resources, improve search efficiency and alleviate network overfitting. Finally, according to the features of high resolution remote sensing image, such as complex ground objects, discrete distribution and wide space range, the Gumbel-Softmax Trick method is introduced to sample from discontinuous probability distribution to improve sampling efficiency.

U-Net is one of the most widely used convolutional neural networks. It obtains the eigenvalues of images through continuous convolution and pooling, and then restores the images by deconvolution[7]. At the same time, U-Net model combines the characteristics of coding and decoding as well as jumping network, and there is a mapping relationship between the extended and reduced network[4]. In the process of expansion, the corresponding shrink layer features are combined to compensate for the lost boundary information, so as to improve the prediction accuracy of the boundary information[17].

In this paper, buildings in high resolution remote sensing images are studied, Due to the diversity of building types, complex texture, irregular spatial distribution and other particularities, there are some problems such as missed detection, mistaken segmentation, fuzzy edge and incomplete segmentation of buildings. In this paper, a new U-NET network model is proposed[18], which adds FPN structure of multi-scale information to the network structure, to enrich the required feature information of each pixel in classification[11]. In a meanwhile, at the bottom of the network, using the efficient residual block network model can effectively reduce the attenuation of feature information caused by the deepening of the network[14]. Finally, two linear weighted combinations based on cross entropy loss and Dice loss are used to optimize the boundary detection of buildings[12]. Experimental results show that the improved method can improve the precision of building boundary segmentation and achieve better segmentation results.

2 Research Techniques and Methods

2.1 The U-Net Model

Firstly, the U-Net network technology is introduced into the cell segmentation of medical image, and good segmentation results are obtained. The main feature of the U-Net network is the use of a skip connection structure similar to that shown in Fig.1. In the coding part, its feature graph is superimposed on the corresponding network structure, and then the inverse convolution is carried out. Then, the feature information obtained from the encoder is directly imported into the corresponding decoder by using the jump connection layer, so that the system can get higher-level information in the process of up-sampling and deconvolution, such as contour information, location information, so as to provide more scale information for the subsequent image segmentation[17].

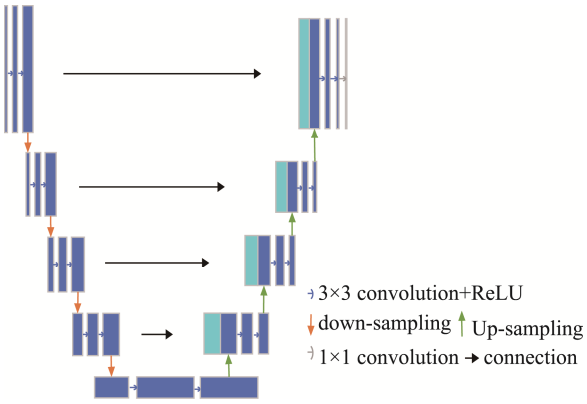


Fig.1 U-Net Network

Based on the U-Net model, this paper improves and optimizes it for the following reasons:

(1) The U-Net model is simple in construction and easy to improve, perfect and expand. Many scholars have made various modifications on the U-Net model

in the past, which facilitates the improvement and use of the algorithm in this paper. Moreover, U-Net has the advantages of fewer training parameters and short training cycle[6].

(2) U-Net model is mainly used for image segmentation. Through U-Net structure, images with small data set can be enriched and trained to achieve good segmentation effect. Since the data used in this method are all buildings based on satellite images, there is little data available for public use, the U-Net has very low requirements for training samples. In a sense, it can effectively overcome the problem of unsatisfactory model training effect caused by limited data.

2.2 Improved U-Net Model

Based on the structure of U-Net convolutional neural network, the linear combination method of FPN structure, efficient residual block and loss function is introduced. The improved U-Net network model is used to segment remote sensing building images. Its network structure is shown in Fig.2.

In this network model, FPN structure is added in the coding stage of network structure to obtain more scale characteristic information of remote sensing image data, and the upsampling of decoder is allowed to access the input information of the whole coding; Secondly, this paper proposes an efficient residual convolution block construction based on the empty convolution method, which aims to expand the perception domain, fully reflect multidimensional features, ensure the existence of gradient, and suppress gradient diffusion to some extent, so as to accelerate the convergence of the network [7]; In addition, cross entropy loss and Dice loss function are added to the loss function to improve the precision of building target edge detection.

2.2.1 FPN Structure

FPN structure, also known as feature pyramid, is widely used in target detection. FPN structure can effectively integrate multi-scale semantic information. FPN is mainly composed of three parts: bottom up, top down and horizontal connection [11], its structure is shown in Fig.3.

On the far left is the bottom-up process, that is, feature information of each scale is extracted from the convolutional neural network. In this process, the top-down method is adopted to sample the extracted feature data, so as to ensure that the upper convolutional features obtained have features similar to the bottom features[11]. On this basis, the information fusion features of multiple levels can be obtained by superposing the corresponding elements of the high-level feature map and the low-level feature map transmitted horizontally. Due to the small amount of convolution that the low feature layer undergoes, it contains a large number of texture features, while the high feature layer has better semantic characteristics on

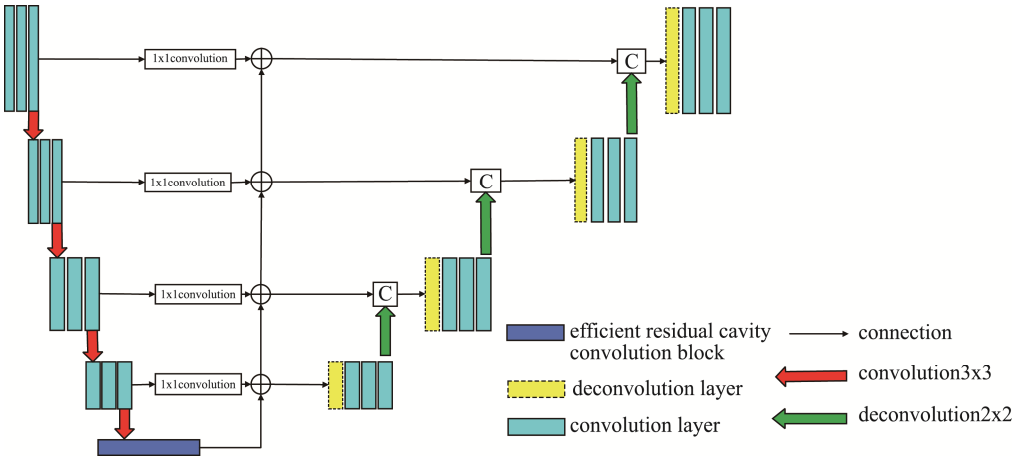


Fig.2 Improved U-NET Network Structure Diagram

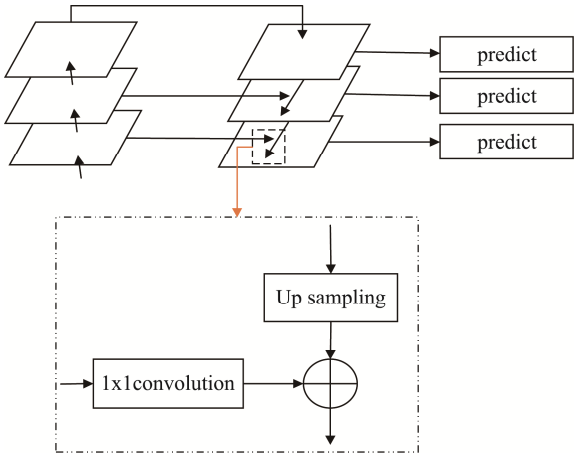


Fig.3 FPA Network Structure

multiple convolutional filtering operations. On this way, the feature of the upper layer is extracted to the same size as the feature of the lower layer, and the phase sum operation is carried out, and the texture of the bottom layer is used to supplement the feature of the upper layer, so as to enrich the characteristics of the fused image[16].

In addition, the upsampling of FPN adopts quadratic linear interpolation, because quadratic linear interpolation is a linear operation, without changing the original shallow level information, only linear interpolation method to increase the size of the feature graph, and the corresponding addition of corresponding elements ensures that the characteristics transmitted to each decoder layer is the addition and fusion. Compared with the conventional U-Net deconvolution operation, it can effectively prevent the data loss caused by deconvolution, retain more complete characteristic data information, and ensure the multi-dimensional data information of all levels contained in the decoder. Through the analysis of FPN and U-Net, we can see that the cross connection between them can be converted to each other. This characteristic is beneficial to the improvement of U-Net model by FPN structure. On the basis of making full use of U-Net network architecture, U-Net gives full play to all kinds of scale information.

2.2.2 Efficient Residual Network

In residual networks, residual learning modules are introduced. It consists of two parts, namely direct mapping and residual[14], its structure is shown in Fig.4.

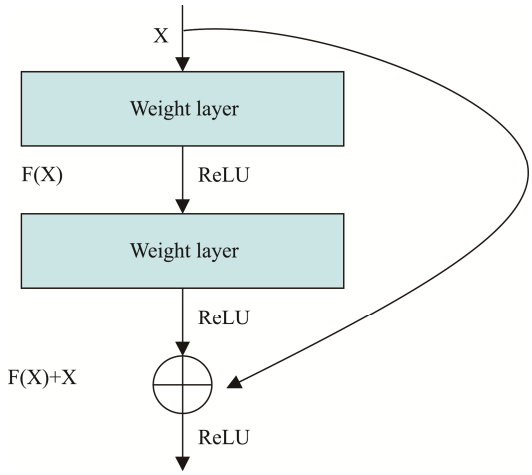


Fig.4 Residual Network

It can be seen from Fig.4 that the residual network has two layers. Such as the type (1):

$$F = W_2\sigma(x_l W_l) \tag{1}$$

σ represents the nonlinear activation function ReLU.

Then through the "quick way" and the second activation function ReLU, the output y is obtained. Such as the type (2):

$$y = F(x,\{W_i\}) + X \tag{2}$$

Thus, it can be seen that the difference between residual network and ordinary convolutional network is that the general convolutional network has problems such as incomplete information and long time consuming in the process of information transmission. However, residual network can realize object segmentation by learning residual information in the case of deep learning difficulty.

In this paper, by introducing three residual modules in the fifth layer of the network, on the one hand, the network information is not smooth, and on the other hand, the problem of poor segmentation effect caused by network degradation is solved.

2.2.3 Loss Function

The linear combination of cross entropy and Dice

loss can effectively overcome the problem that the target features of remote sensing image data set are not significant, improve the data imbalance and improve the precision of target segmentation.

Considering the loss degree of the whole image, more attention is paid to the loss change of the target building, so the segmentation error based on the feature region can be effectively eliminated[17].

The loss rate of Dice's loss function is L_D, the expression is shown in (3),

$$L_D = -\frac{\sum_{n=1}^{N}(p_n \times r_n) + \varepsilon \sum_{n=1}^{N}(1-p_n)(1-r_n) + \varepsilon}{\sum_{n=1}^{N}(p_n \times r_n) + \varepsilon \sum_{n=1}^{N}(2-p_n 1-r_n) + \varepsilon} \quad (3)$$

The loss rate of BCE loss function is L_B, the expression is shown in (4),

$$L_B = -\frac{1}{n}\sum_{n=1}^{N}(y_n \times \ln x_n + (1-y_n) \times \ln(1-x_n)) \quad (4)$$

The loss rate of linear combination loss function is L_{DB}, the expression is shown in (5),

$$L_{DB} = 0.5L_B + L_D \quad (5)$$

N indicates the number of all pixels of the segmented image; P represents the number of foreground pixels in the training set; R represents the actual number of foreground pixels in the training set; N indicates the number of training round; x_n is the independent variable of loss function training; y_n is the gradient factor of gradient change.

3 Experimental Results and Analysis

3.1 The Data Set

In order to verify the effectiveness of the improved network structure, the high resolution remote sensing image buildings published by the Institute of National Institute for Information and Automation in France are used as data sets to carry out validation experiments. The data coverage of this dataset is up to 810km^2, which the image types are aviation forward full color image and a resolution of 0.3m, covering many different terrains, for example, buildings in dense residential areas and some mountainous areas basically include all types of common buildings.

To maximize the calculation of the experimental platform operation performance, this study will be the original image is cut into 512 x 512 pixel resolution of sample images, a total of 387 pieces, and according to the proportion, reviews the research data and combining with the actual situation to make minor adjustments, which speak about according to the proportion of the data, then 336 training sets of remote sensing buildings covering 810 km^2 with a resolution of 512 x 512 and 51 corresponding test sets are obtained.

3.2 Experimental Environment

In this paper, Python programming language was adopted under PyCharm, Pytorch was used as the network development structure, and an improved U-Net model was used as the segmentation network to achieve segmentation of image data sets of different buildings. Finally, a conclusion was obtained through simulation experiment.

3.3 Model Training and Evaluation Indicators

The total set parameters of this experiment are as follows: 200 training rounds, batch size set as 16, commonly used Adam optimizer, learning rate 0.001, its results is shown in Fig.5[20].

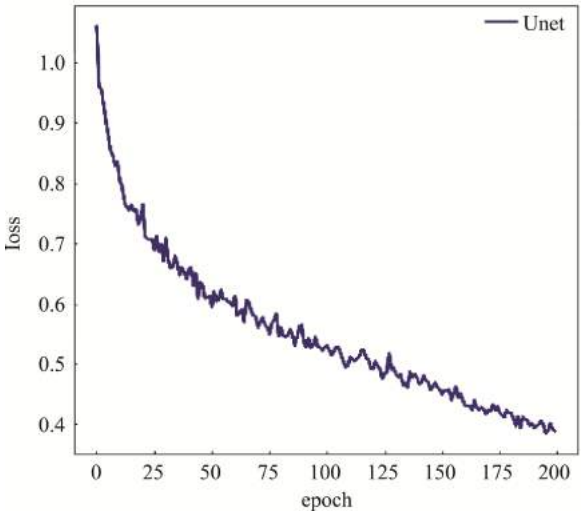


Fig.5 Loss Diagram

When performing semantic segmentation tasks, accuracy, Recall and F1-SOCre are generally used to evaluate the effect of the model indirectly. Each indicator is expressed as follows[22]:

$$Precision = \frac{TP}{TP + FP} \times 100\% \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \tag{7}$$

$$F_1 = \frac{2RP}{R + P} \tag{8}$$

Here, TP is the positive sample of the correct prediction, FP is the positive sample of the wrong prediction, FN is the negative sample of the prediction error[18][19]. F1-socre is a measure of accuracy and recall rate, and is the harmonic average between the two side。

In this paper, in order to compare the detection effects of different algorithms on data sets, overall Pixel Accuracy (PA) and intersection ratio (IMoU) were selected to directly evaluate the algorithms. Where, the PA expression is shown in Equation (9):

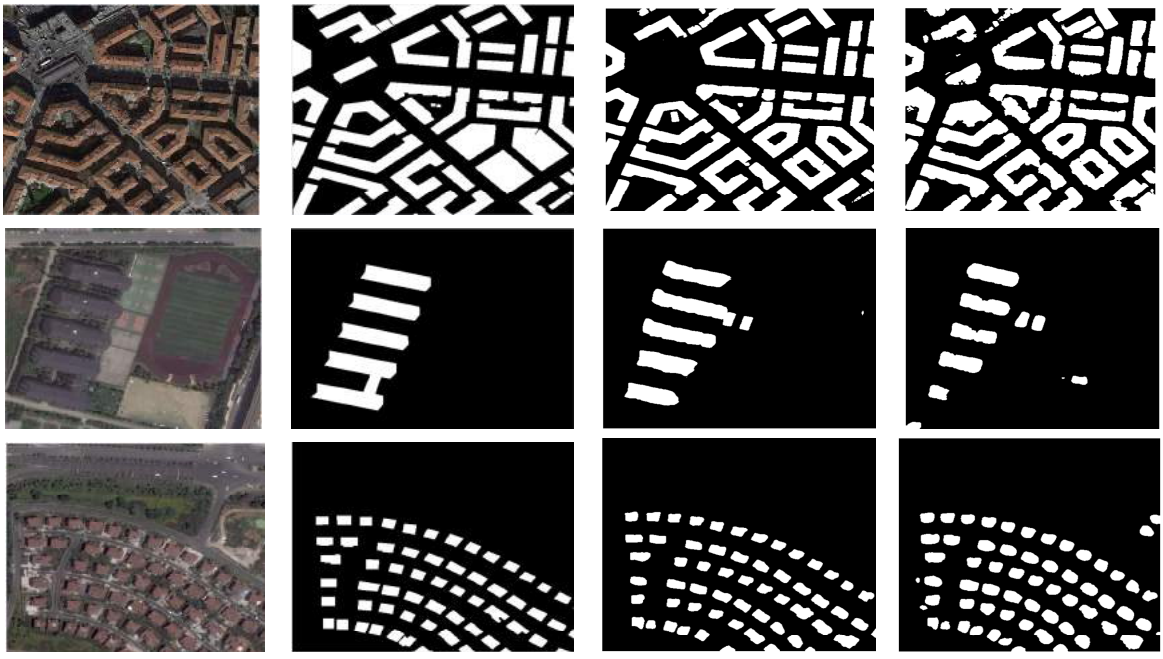
$$PA = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

As a standard measure of semantic segmentation, IMoU is an IoU based on classification and weighted average of the IoU to obtain a hole-based IMoU, the IMoU expression is shown in Equation (10):

$$IMoU = \frac{1}{k+1}\sum_{i=0}^{k}\frac{TP}{FN + FP + TP} \tag{10}$$

3.4 Analysis of Experimental Results

In this paper, by comparing the accuracy of U-Net and the improved algorithm in building segmentation of remote sensing image, the experimental result graph can be obtained as shown in Fig.6.



(a) The Original Image (b) Mask (c) Improved U-Net (d) U-Net

Fig.6 Experimental Results

Table 1 Evaluation Indexes of Different Algorithms

The Evaluation Index	IMoU (Building and Background)	IMoU (Building)	Time/s
U-Net	80.58	0.76	0.298
Improved U-Net	80.82	0.94	0.295

4 Conclusion

In this paper, an improved U-Net algorithm is proposed to solve the problems of low efficiency and insufficient segmentation of existing remote sensing building images. Firstly, by modifying the backbone network structure, FPN is used as the backbone network to improve the extraction and integration of building characteristic information; then, efficient residual block network is used to solve the problem of feature information weakening caused by network depth; finally, the fusion loss function is used to achieve the detection accuracy of building edge. From the experiment, it can be known that the improved network is more accurate for edge detection of building objects, and it can be seen from the comparison in Fig.5 (c) and (d) that the segmentation precision of the cutting network before and after the improvement on building edge contour has a great difference, and the precision of the improved network on building contour segmentation is as high as 80.82% (including the background part). Compared with the traditional unimproved segmentation network, the accuracy of the improved algorithm is improved by 18%, and the running speed of the improved algorithm is also improved. Therefore, the method proposed in this paper has certain reference value for remote sensing image segmentation.

References

[1] Wang C, Shi A Y, Gu A H, Xiong X B, Liu Q. A city high-resolution remote sensing image segmentation method combined with shadow compensation [J]. Journal of Electronic Measurement and Instrumentation, 2017, 31(10): 1687-1692.

[2] Xiang Z J, Can F S, Chu H, Huang L. High resolution remote sensing image segmentation algorithm based on superpixel [J]. Computer Engineering and Design, 2020, 41(05): 1379-1384.

[3] Cao X, Song C G, Zhang J, Liu C. Remote Sensing Image Segmentation based on Generative Adversarial Network with Wasserstein divergence[C]//. Conference Proceeding of 2020 3rd International Conference on Algorithms, Computing and Artificial Intelligence (ACAI 2020)., 2020: 342-347.

[4] Sun Y, Tian Y, Xu Y P. Problems of Encoder-Decoder Frameworks for High-Resolution Remote Sensing Image Segmentation: Structural Stereotype and Insufficient Learning[J]. Neurocomputing, 2018, 330.

[5] Zhu H M, Tan R, Han L T, Fan H F, Wang Z J, Du B W, Liu S C, Liu Q. DSSM: A Deep Neural Network with Spectrum Separable Module for Multi-Spectral Remote Sensing Image Segmentation[J]. Remote Sensing, 2022, 14(4).

[6] Zhang R N, Yu L, Tian S W, Lv Y L. Unsupervised remote sensing image segmentation based on a dual autoencoder[J]. Journal of Applied Remote Sensing, 2019, 13(03).

[7] Chen G S, Li C, Wei W, Jing W P, Marcin Woźniak, Tomas Blažauskas, Robertas Damaševičius. Fully Convolutional Neural Network with Augmented Atrous Spatial Pyramid Pool and Fully Connected Fusion Path for High Resolution Remote Sensing Image Segmentation[J]. Applied Sciences, 2019, 9(9).

[8] Wang Y Q, Wang C X. High resolution remote sensing image segmentation based on multi-features fusion[J]. Engineering Review: Međunarodni časopis namijenjen publiciranju originalnih istraživanja s aspekta analize konstrukcija, materijala i novih tehnologija u području strojarstva, brodogradnje, temeljnih tehničkih znanosti, elektrotehnike, računarstva i građevinarstva, 2017.

[9] Zhao Q, Zhang S, Huang S L. Multi-scale and Multi-feature High Resolution Remote Sensing Image Segmentation[J]. International Journal of Applied Mathematics and Statistics™, 2013, 51(22).

[10] Zhou P, Yang J. Remote sensing image segmentation method based on neural network architecture search is adopted [J]. Journal of Xidian University, 2021, 48(05): 47-57+77.

[11] Wang X, Yu M, Ren H E. Remote sensing image semantic segmentation based on UNET and FPN [J]. Liquid crystal and Display, 2021, 36(03): 475-483.

[12] Gong Y J, Huang M, Huang X L. Flame image segmentation method based on improved Resnet-UNET [J]. Journal of Beijing Information Science and Technology University (Natural Science Edition), 2021, 36(05): 39-44.

[13] Li L X, Yuan Y, Wen S H. Building segmentation and extraction from high resolution remote sensing image

based on BAU-NET [J]. Journal of Yanshan University, 2021, 45(04): 335-342.

[14] Wang Y, Yang Y, Wang B S, Wang T, Pu X H, Wang C Y. Deep residual neural network for building segmentation in high resolution remote sensing images [J]. Remote sensing technology and application, 2019, 34(04): 736-747.

[15] Li Y, Li Y J, Liu J C, Fan H, Wang Q L. Research on image segmentation of steel surface defects based on improved RES-UNET network [J/OL]. Journal of Electronics and Information technology: 1-8.

[16] Gu S J, Pu X Z, Jing J W, Liu S. Muzzle flame segmentation method based on improved UNet network [J]. Foreign electronic measurement technology, 2021, 40(04): 16-21.

[17] Huang S P, Liu H N, Zhou K S, Liu J Y. Zebra crossing segmentation based on improved UNet [J]. Intelligent computers and applications, 2020, 10(11): 61-64+69.

[18] Xiang Y, Zhao Y D, Dong J H. Mining area change detection based on improved UNet twin network [J]. Journal of coal, 2019, 44(12): 3773-3780.

[19] Fu L H, Zhao Y, Jiang H X, Zhao R, Wu H X, Yan S X. Semi-supervised video object segmentation based on foreground perception and visual attention [J]. Electronic journals, 2022, 50(01): 195-206.

[20] Wang X B, Cao S P, Zhao H C, Liu P F, Tai B C. Bilateral feature aggregation and attention mechanism point cloud semantic segmentation [J]. Chinese Journal of Scientific Instrument, 2021, 42(12): 175-183.

[21] Wang Z D, Guo C. A semantic segmentation optimization ORB_SLAM2 in dynamic scenarios [J]. Journal of Dalian Maritime University, 2018, 44 (04): 121-126.

Author Biographies

SHI Mengyuan is currently a M.Sc. candidate, majoring in communication and information system.
E-mail: 951051762@qq.com

GAO Junchai received a Ph.D. degree of Engineering. She is currently an associate professor. Her main research interest includes computer vision.
E-mail: 527667658@qq.com

Naïve Bayes Algorithm for Large Scale Text Classification

Pirunthavi SIVAKUMAR[1], Jayalath EKANAYAKE[2]

(1. *Department of Information and Communication Technology*, *Faculty of Technology*, *Rajarata University of Sri Lanka*, *Mihintale* 50300, *Sri Lanka*; 2. *Department of Computer Science and Informatics*, *Faculty of Applied Sciences*, *Uva Wellassa University of Sri Lanka*, *Badulla* 90000, *Sri Lanka*)

Abstract: This paper proposed an improved Naïve Bayes Classifier for sentimental analysis from a large-scale dataset such as in YouTube. YouTube contains large unstructured and unorganized comments and reactions, which carry important information. Organizing large amounts of data and extracting useful information is a challenging task. The extracted information can be considered as new knowledge and can be used for decision-making. We extract comments from YouTube on videos and categorized them in domain-specific, and then apply the Naïve Bayes classifier with improved techniques. Our method provided a decent 80% accuracy in classifying those comments. This experiment shows that the proposed method provides excellent adaptability for large-scale text classification.

Keywords: Naïve Bayes, Text Classification, YouTube, Sentimental Analysis

1 Introduction

Recently, the complexity of documents and text has increased exponentially, which requires a deeper knowledge of machine learning methods to accurately classify text in many computer applications. Many machine learning methods have reached outstanding output in natural language processing (NLP). The success of these machine learning algorithms relies on their ability to understand the complex patterns and nonlinear relationships between the data. Even though, finding the appropriate text classification technique for solving their research problems is a challenge for many researchers.

A widely used text classification algorithm is the Naive Bayes algorithm due to its easiness and high accuracy[6]. Rather than the other text classification algorithms, the Naïve Bayes algorithm is simple to construct and especially beneficial for extremely huge data sets. In addition to simplicity, the Naive Bayes algorithm is even superior to the most sophisticated classification methods[14].

This research topic is proposing a model which can classify large-scale data sets with more accuracy through sentimental analysis using the Naïve Bayes algorithm. In this research, we selected YouTube as the source of data that can provide a large scale of text data. Thousands of users give positive and negative comments in the comment section for each video. For this reason, we need a way to analyze viewer comments and opinions to rank viewers' opinions about each video through sentimental analysis using the Naive Bayes algorithm. Text preprocessing such as text labeling, tokenization, data stemming, and filtering.

2 Literature Review

Several types of machine learning algorithms have been applied to text classification, including the Naive Bayes algorithm[12][13], KNearest Neighbor (KNN) algorithm[3], Support Vector Machines algorithm[4], and Neural Networks algorithm. Sentimental analysis is a computerized method for analyzing sentiments, opinions, and emotions[5]. Sentimental analysis is used to identify emotional tendencies when an opinion is positive or negative. Sentimental analysis is performed by processing textual data, so it is ready for analysis by text mining methods. Sentimental analysis is one of the natural language processing techniques that can be used to determine the level of sensitivity behind the text, i.e. YouTube comments, Tweets, movie reviews, etc[7].

For an instance, Grammarly is a tool that is utilized to modify the grammar of a document or text, showing the general meaning and pronunciation of the document and creating useful comments such as positive and optimistic. This is also true. It is done through a sentimental analysis of the entire document or text. It is widely used and helps to maintain brand equity as it is used by many large companies to check customer reviews on their products/ services, social networks, and websites.

Some large companies such as Apple, Google, and Trip Advisor use this technique to increase customer loyalty in their service services. Other uses of sentimental analysis include sentiment analysis on Twitter, movie ratings on IMDB, customer reviews on Amazon, and video comments on YouTube[7].

2.1 K-NN Algorithm

K-Nearest Neighbor (K-NN) algorithm is utilized for text classification with the aid of using the K nearest neighbors in the training dataset after which the usage of the label of closest suits to predict[8]. In essence, the K-nearest neighbor algorithm has resulted in a majority out of K-cases that closely resemble a certain "invisible" observation. The similarity between two data points is defined as a measure of the distance. A common method is the Euclidean distance method. Other methods include Minkowski, Manhattan, and the Hamming distance method[7]. The Hamming distance method should be used for the categorical variables. K-nearest neighbor algorithm is a straightforward algorithm that saves all available cases and ranks upcoming cases based on a measure of similarity. It is primarily utilized to classify data points based on the classification of adjacent data points.

2.2 SVM Algorithm

A support vector machine algorithm that determines the fine-tuning of boundaries within the vectors that belong to a particular group and vectors that do not.

A vector is a list of numbers that represent a set of coordinates in a small space. Therefore, while the support vector machine determines the boundaries, the support vector machine decides to draw the best "line" (highest quality hyperplane) that divides the gap into subspaces. Text classification using SVM is simple and straightforward with the use of MonkeyLearn[8].

2.3 Neural Network

A neural network consists of weighted neurons that connect them, learning by processing records in sequence and comparing classifications to actual classifications[11]. A neural network can be described by three components or layers: an input layer, a hidden/intermediate layer, and an output layer. The task of the input layer is to capture the input signal from the external system. The hidden layer is made up of neurons. Neural network research is fully monitored for this reason, as the inputs to the neural network have solutions or outputs. The neural network gets the input values and weights from the input layer, and the function sums the weights and moves the output to a hidden layer that maps the output to the appropriate unit of the output layer. You can include "n" different hidden layers among the input and output layers. Determining the type of hidden layer, the community may be referred to as a single-layer or multi-layered (in the case of multiple hidden layers) neural network[11].

The neural network method had been an extraordinary tool for classifying the text of numerous struc-

tures of NN that had been utilized for the utility of classification of text, e.g. Error Back Propagation Algorithm, ADALINE and MADALINE network, etc[10].

Phases concerned in text classification are amassing the dataset, preprocessing, dimensionality reduction, and classifier implementation[10].

2.4 Naïve Bayes

A commonly used text classification algorithm is the Naive Bayes algorithm due to its simplicity and high accuracy[6]. The Naïve Bayes algorithm is simple to construct and especially beneficial for extremely huge data sets. Compared to the other text classification algorithms, the Naïve Bayes algorithm is simple to implement and extremely exhaustive for large-scale data sets. In addition to that, this algorithm is more powerful than the most sophisticated text classification algorithms[14].

Each step of this algorithm for text classification is pretty much straightforward compared to the other text classification techniques.

2.5 Performance Comparison

Compare to the other existing text classification techniques, Naïve Bayes performs best with a prediction accuracy of 87.47%. In addition to that, Naïve Bayes is the fastest classification method so far, training in seconds[15].

Looking at the total run time, including preprocessing, classifier training, and test execution, reveals a significant advance in the Naïve Bayes approach. The "training" of the classifier is purely statistical, so it is less computationally intensive than analyzing the data itself[15].

The Naïve Bayes algorithm came up with some ideal benefits over other techniques. If a person desires to understand precisely why a sentence became categorized to a particular class it's far quite simple to give the weighted extraordinary phrases withinside the sentence had in the direction of a class. It makes it feasible to locate flaws withinside the classification method and fine-track the set of rules if desired. Another classification techniques are bit "black box" as it takes input and presents the output with a little manner of understanding of why it was given that result[15].

If new training data is added to the dataset, the Naïve Bayes classifier simply recalculates only the affected classes by correcting how many times the words occur within the class. On the other hand, in other classification techniques, these tasks cannot be addressed without retraining completely by wasting a great deal of time and computing power[15].

3 Methodology

3.1 Data Pre-processing

Here, the naïve Bayes algorithm is utilized for text classification of the comments of YouTube videos. The important step even as implementing a machine learning model is the data pre-processing because it will at once affect the result of the model. The greater pre-processing the data, the greater the model performance will be accurate[7]. Fig.1 illustrates the steps of data-preprocessing of the naïve Bayes algorithm.

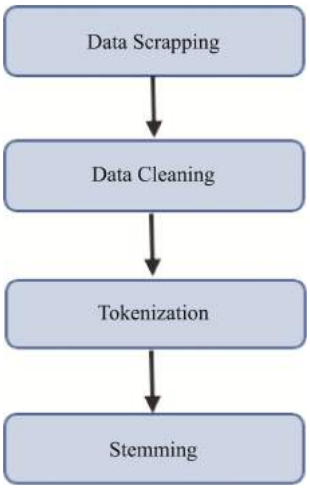


Fig.1 Steps of Data Pre-processing

YouTube remarks contain numerous languages relying on the study of the human population of the commenter. Anyhow, to simplify the sentimental analysis, they modified the data collection scripts to acquire the simplest English comments[7].

Data Scrapping - "YouTube Scraper" was utilized for extracting remarks from YouTube videos. This tool

saves the scrapped remarks in a CSV file (Comma Separated Value). But this device isn't that effective. This device extracts all of the remarks from each of the videos. As our PC does now no longer has sufficient processing power and storage, it changed into very much difficult for further processing. Then YouTube Rest API was used to scrape the comments from YouTube. Python language was used to extract the comments of each video using the YouTube APIs. Thousand of comments were extracted per each video since API permits only a thousand comments to be extracted from YouTube[7].

Data Cleansing - Data cleansing performed a large element in constructing a model. It gave the capacity to find out erroneous or incomplete records. Without the right data quality, the very last analysis might go through inaccuracy or incorrect end will be doubtlessly arrived. Extracted comments had been cleaned up to remove the unwanted data and select only the wanted data that is useful in the data analysis process. In this step, punctuation and emoji had been removed from each user review. After this process, the data set will be robust and avoid many common pitfalls from the comments.

Tokenization - Tokenization can be a step that splits longer strings of textual content into smaller portions or tokens. Larger chunks of textual content are frequently tokenized into sentences, sentences are frequently tokenized into words, etc. Further processing changed into done after a bit of textual content has been correctly tokenized. The sequence of strings is divided into parts such as words, keywords, phrases, and different factors referred to as tokens. After some of the text was properly tokenized, further processing was done.

Stemming - Data Stemming is the Text Normalization (or sometimes referred to as Word Normalization) method in the field of Natural Language Processing (NLP) this is used to prepare text, words, and files for further processing. It is the technique of decreasing a phrase to its phrase stalk into suffixes and prefixes. Here, different forms of a word, suffixes, and prefixes were reduced in each sentence.

3.2 Data Transfer

Two types of data set had been obtained after pre-processing.

Training data set has been utilized to train the model and 200 video comments had been used for this process.

The test data set is utilized to evaluate the performance of the model using a measure of performance. The important thing is that there is the data from the training set can not be included in the test suite. If the test set includes the data/ cases from the training data set, it is very difficult to assess whether the algorithm learned to generalization from fixed training or simply remembered it. Hundreds of movies had been selected to check and confirm paradigm. Fig.2 shows the division of collection of data to the training and testing data set.

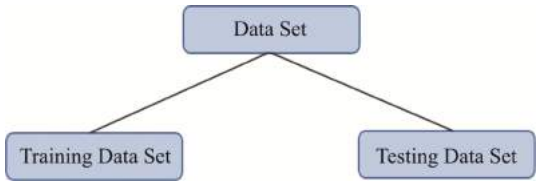


Fig.2 Data Transfer

3.3 Data Analysis

Data Labeling - The words had been categorized into classes which include positive words (+ve), and negative words (-ve). Positive and negative datasets had been obtained from the internet. And by the one's dataset, the processed information had been categorized into classes (positive and negative). So that, the percentage of positive functions in addition to the negative functions had been calculated accordingly.

Overall Measure - A user's remarks can also additionally incorporate positive or negative characteristics. It used a mathematical scale to calculate the percentage of positive features in addition to negative features. The average of the datasets collected from YouTube was calculated using the Naive Bayes algorithm. Fig.3 explains the process of the data analysis part of the naïve Bayes algorithm.

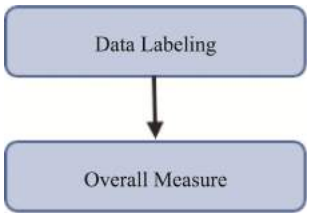


Fig.3 Steps of Data Analysis

4 Results and Discussion

The Naive Bayes algorithm is simple to construct and especially beneficial for extremely huge data sets. Naive Bayes can offer accurate outcomes without the need for much training dataset. It treats the probability of every phrase appearing in a document as though it has been unbiased of the probability of another phrase. This assumption is sort of in no way true of any files we`d desire to classify, which have a tendency to comply with regulations of grammar, syntax, and communication. When we comply with those regulations, a few phrases tend to be correlated with different phrases.

In addition to that, the following equation is useful in the terms of performance which is defeating the other classification algorithms[14].

$$Pr(H|E) = Pr(H) * Pr(E|H) / Pr(E)$$

H - Hypothesis
E - Evidence

Classification tasks involve comparing two (or more) hypothesis,

$$Pr(C_1|W) / Pr(C_2|W) = Pr(C_1) * Pr(W|C_1) / Pr(C_2) * Pr(W|C_2)$$

$$Pr(C_1|W_1, W_2 \ldots Wn) / Pr(C_2|W_1, W_2 \ldots Wn) = Pr(C_1) * (Pr(W_1|C_1) * Pr(W_2|C_1) * \ldots Pr(Wn|C_1)) / Pr(C_2) * (Pr(W_1|C_2) * Pr(W_2|C_2) * \ldots Pr(Wn|C_2))$$

The naive Bayes algorithm utilizes a unique method to obtain the probability that different classes support different attributes. After pre-processing and transferring the data, the algorithm creates a likelihood table after converting the data set into a frequency table by finding the possibilities of classes. Tables-1, 2, and 3 show its working example.

Table 1

Words	Positive/ Negative
Hate	Negative
Fantastic	Positive
Bad	Negative
Super	Positive
Harmful	Negative
Fantastic	Positive
Bad	Negative
Exhausted	Negative
Worst	Negative
Super	Positive
Beautiful	Positive
Admiring	Positive
Exhausted	Negative
Fantastic	Positive
Worst	Negative
Lovely	Positive
Gorgeous	Positive
Awful	Negative
Lovely	Positive
Harmful	Negative

Table 2 Frequency Table

Frequency Table		
Words	Positive	Negative
Hate		1
Fantastic	3	
Bad		2
Super	2	
Harmful		2
Exhausted		2
Worst		2
Beautiful	1	
Admiring	1	
Gorgeous	1	
Lovely	2	
Awful		1
Total	10	10

Table 3 Likelihood Table

Likelihood Table			
Words	Positive	Negative	
Hate		1	1/20=0.05
Fantastic	3		3/20=0.15
Bad		2	2/20=0.1
Super	2		2/20=0.1
Harmful		2	2/20=0.1
Exhausted		2	2/20=0.1
Worst		2	2/20=0.1
Beautiful	1		1/20=0.05
Admiring	1		1/20=0.05
Gorgeous	1		1/20=0.05
Lovely	2		2/20=0.1
Awful		1	1/20=0.05
Total	10	10	
	10/20=0.5	10/20=0.5	

Table 4 Sample Outputs of the Model

Video ID	No of Comments to Extract	Positive Sentiment	Negative Sentiment
yGY484EPe5U	1000	38.9%	61.1%
vbNib_NsVRN	1000	25.2%	74.8%

To the model, the video ID and the number of comments were specified as input and took the percentage as the output of positive and negative sentiment analysis. Table-4 shows some of the sample outputs of naïve Bayes text classification. The YouTube API permits the developers to access video statistics and YouTube channel datasets through REST API calls[7]. There were many comments on sentiment analysis, as a small number of comments can reduce accuracy. Therefore, we collected 1000 comments for each video.

Compared to the other existing methods, this method is straightforward and decent enough and provides high accuracy of 80%. As the YouTube API was used here for extracting the comments from each YouTube video, there was a limitation since it allows only a maximum 1000 number of comments to be extracted. If anyone uses any tool for extracting all the comments from each YouTube video the accuracy might vary. However, to store each scrapped comment, the machine should have more processing power and storage capacity.

5 Limitations

YouTube comments comprise several languages counting on the demography of the commenter. We only considered English movies in our research process.

YouTube APIs were used to extract comments from each video. On YouTube, you can get up to 1000 comments for each video via YouTube's API, so only thousands of comments were extracted for each video.

People generally tend to search for feedback based on their very own opinion. These days one terrible comment may unfold very instantly. People generally tend to reply to those types of feedback very often. So by and large negative comments with more responses can also additionally come to the top withinside the queue. This factor could negatively affect the accuracy of the model.

6 Conclusion

In case we have an issue in classification, we have to solve that issue. If there is a large dataset and variables for training the dataset, the best solution is to utilize the naive Bayes algorithm, which is much faster and more optimistic than the other classification algorithms.

The Naive Bayes algorithm is an efficient and effective way to design a text classification with 80% high accuracy and speed. Naïve Bayes classification algorithm is primarily based totally on the conditional probability of each feature belonging to a class, which the features are decided on with the aid of using feature

selection methods. However, the naive Bayes algorithm is based on machine learning, hence the accuracy rate is highly dependent on the training data set. Also, if the test data set contains complicated sentences, the accuracy rate will be lower. Hence, this will continue to be our field of study in the future. YouTube API allows only 1000 user reviews to extract and hence, this factor could negatively affect the accuracy of the model.

7 Recommendations

We couldn't scrape all the comments from YouTube, because we used YouTube APIs to scrape the comments. Thus, we only extracted up to 1000 comments. One can use another technique to scrape all the comments of a particular video from YouTube and do sentimental analysis.

One can use more than one social media (Example: YouTube, Facebook, Twitter, Instagram) and use all the data and do sentimental analysis and output the accurate output.

Here, we have only considered English YouTube videos. In another way, one can use any language videos from any country and follow up with the sentimental analysis. So their expected system can be used in any country.

References

[1] Lo S, Ding L., 2012. Probabilistic reasoning on background net: An application to text categorization, Proc of 2012 *International Conference on Machine Learning and Cybernetics (ICMLC). IEEE Press,2*: 688-694.

[2] Kuang, F., Xu, W. and Zhang, S., 2014. A novel hybrid KPCA and SVM with GA model for intrusion detection. *Applied Soft Computing*, *18*, pp.178-184.

[3] Alpaydın, E., 2010. Introduction to Machine Learning, 2nd Edition Cambridge, MA: The MIT Press.

[4] Indurkhya N. and Damerau J., 2010. *Handbook of natural language processing, second edition.*

[5] Rish, I., 2001. "An empirical study of the naïve Bayes classifier," in *IJCAI 2001 workshop on empirical methods in artificial intelligence, vol. 3, no. 22,* pp. 41–46.

[6] Sivakumar, P., Rajeswaren, V., Abishankar, K., Ekanayake, J., and Mehendran, Y., 2020. Movie Success and Rating Prediction Using Data Mining Algorithms. *Journal of Information Systems & Information Technology (JISIT)*, 5(2), pp.72-80.

[7] Dao, S., 2018. Text Classification using K Nearest Neighbors. [online] Medium. Available at: <https://towardsdatascience.com/textclassification-using-k-nearest-neighbors-46fa8a77acc5> [Accessed 16 January 2022].

[8] https://monkeylearn.com/. n.d. Text Classification Using Support Vector Machines (SVM). [online] Available at: <https://monkeylearn.com/text-classification-support-vector-machines-svm/> [Accessed 20 January 2022].

[9] Patra, A. and Singh, D., 2013. Neural Network Approach for Text Classification using Relevance Factor as Term Weighing Method. *International Journal of Computer Applications, 68(17),* pp. 37-41.

[10] Lakshmi Prasanna, P. and D. Rajeswara Rao, D., 2017. Text classification using artificial neural networks. *International Journal of Engineering & Technology, 7(1.1),* pp.603.

[11] Ekanayake, J., 2021. Bug Severity Prediction using Keywords in Imbalanced Learning Environment. *Int. J. Inf. Technol. Comput. Sci. (IJITCS)*, *13*, pp.53-60

[12] Ekanayake, J.B., 2021. Predicting Bug Priority Using Topic Modelling in Imbalanced Learning Environments. *International Journal of Systems and Service-Oriented Engineering (IJSSOE)*, *11*(1), pp.31-42

[13] upGrad blog. 2021. *Naive Bayes Explained: Function, Advantages & Disadvantages, Applications in 2022 | upGrad blog*. [online] Available at: <https://www.upgrad.com/blog/naive-bayes-explained/#:~:text=Naive%20Bayes%20is%20suitable%20for,input%20variables%20than%20numerical%20variables.> [Accessed 4 February 2022].

[14] analytics Vidhya. 2017. *6 Easy Steps to Learn Naïve Bayes Algorithm with codes in Python and R.* [online] Availableat:<https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/> [Accessed 4 February 2022].

[15] Edward, E., 2018. *Comparing Methods of Text Categorization.* [online] Diva-portal.org. Available at: <http://www.diva-portal.org/smash/get/diva2:1275337/FULLTEXT01.pdf> [Accessed 1 March 2022].

Author Biographies

Pirunthavi SIVAKUMAR received a B.Sc. degree from Uva Wellassa University of Sri Lanka, Badulla, Sri Lanka, in 2020. She is currently a M.Sc. candidate at Postgraduate Institute of Science, University of Peradeniya, Sri Lanka. She is also a temporary lecturer at the Department of Information & Communication Technology, Faculty of Technology, Rajarata University of Sri Lanka, Mihintale, Sri Lanka. Her main research interests include machine learning and data mining.

E-mail: psivakum@tec.rjt.ac.lk

Jayalath EKANAYAKE received his B.Sc., and M.Sc. degrees from the University of Peradeniya, Sri Lanka. He received his Ph.D. degree from the University of Zurich, Switzerland. He is currently a senior lecturer at the Department of Computer Science and Informatics, Faculty of Applied Sciences, Uva Wellassa University of Sri Lanka, Badulla, Sri Lanka. His main research interests include mining software repositories and pattern recognition.

E-mail: jayalath@uwu.ac.lk

2022 9th International Symposium on Test Automation & Instrumentation (ISTAI 2022) Call for Paper

2022 9th International Symposium on Test Automation & Instrumentation (ISTAI 2022) will be held in Beijing in October 2022, it is sponsored by both China Instrumentation & Control Society (CIS) and Beijing Information Science and Technology University. The Theme will be "Bionic perception, quantum sensing, intelligent service". It aims to promote the development of automation technology with informatization and intelligence, and to enhance the exchanges and cooperation between domestic academia/industries and international counterparts. The symposium warmly welcomes scholars, students, academics, researchers to contribute, and participate in the symposium to exchange the latest achievements of testing and instrumentation technology with peers from all over the world.

The accepted papers will be published on the IET open access research Journal of Engineering and included in both IET Digital Library and IEEE Xplore Digital Library, EI indexed.

1. Conference Organization

Sponsors:

- China Instrumentation & Control Society (CIS)
- Beijing Information Science and Technology University

Organizers:

- Measurement and Instrument Committee of CIS
- Key Laboratory of Modern Measurement & Control Technology, Ministry of Education
- Beijing Key Laboratory of High Dynamic Navigation Technology
- Beijing Key Laboratory of Measurement & Control of Mechanical and Electrical System Technology
- Beijing Key Laboratory for Sensor
- Beijing International Cooperation Base for Science and Technology on High-end Equipment Intelligent Perception and Control
- *Chinese Journal of Scientific Instrument*

Co-organizers:

- *Journal of Electronic Measurement and Instrumentation*
- *Instrumentation*

2. Scope

- Foundational Theories and Development Trend of Modern Technology
- Signal and Image Processing
- System Monitoring, Fault Diagnosis and Prediction
- Instrumentation and Systems Design and Application
- Sensors Technology and Internet of Things Technology
- Hydro-mechatronics
- Integrated Measurement and Control Technology
- Measurement and Control technology and System for Safe Operation of New Energy Equipment
- Artificial Intelligence
- Environmental Monitoring and Protection

Language: English

Peer Review Duration: 30 working days from submission

Deadline: 31st July, 2022

3. Submission Guideline

- Authors need to register and use the submission system to contribute.
- Authors need to meet the requirements of the layout when submit the papers, please follow the standard indicated on the official website, www.istai.org.cn.
- Papers should be original and unpublished for the submission.
- File of submission must be MS Word format.
- Once the paper is submitted, the copyright is unconditionally transferred to the organizing committee of ISTAI 2022.

4. Contact

An Yalan
Email: istai@vip.163.com
Telephone: +86-(010)-53389339/64044400
Website: http://www.istai.org.cn
Address: No.79 Beiheyan Street, Dongcheng District, Beijing 100009