

The Recognition Accuracy in the SSD Model



Van-Nam Nguyen

Abstract The Single Shot Multibox Detector (SSD) technique is currently among the fastest and most accurate detection algorithms available. However, the majority of research on the accuracy of this approach has focused on noiseless objects. Thus, this study evaluates the algorithm's accuracy with both noisy and noiseless objects. To that goal, the algorithm is trained to recognize ten different flower species. Experiments are then carried out on photographs in four different scenarios: the item is totally lighted, 1/3 of the object is darkened, 1/2 of the object is darkened, and the object is fully darkened. The performance of the algorithm is then evaluated using SPSS 20.0 software and the analysis of variance (ANOVA) and least significant difference (LSD). The experimental results reveal that the algorithm accuracy is strongly dependent on the noise level. The detection accuracy is 100%, 81.3%, 44.7%, and 62%, respectively, when the item is fully lighted, 1/3, 1/2 size of the object is darkened, and the object is fully darkened.

Keywords Artificial intelligence · Computer vision · Deep learning · Identifying the object · Image processing

1 Introduction

Many scientists have been interested in the implementation of Deep Learning models in practice in recent years, particularly the Single Shot MultiBox Detector (SSD) model [1, 2]. SSD is a well-known algorithm for dealing with issues including large data processing, input noise management, and online processing. In addition, the Faster region-based convolutional neural networks (Faster R-CNN) model is also one of the best models available today [3, 4].

V.-N. Nguyen (✉)

The University of Danang—University of Technology and Education, Da Nang, Vietnam
e-mail: nvnam@ute.udn.vn

Table 1 Data collection and labeling for flowers [7]

System	VOC2007 test mAP	FPS (Titan X)	Number of Boxes	Input Resolution
Faster R-CNN (VGG16)	73.2	7	~6000	~1000 × 600
SSD300* (VGG16)	77.2	46	8732	300 × 300
YOLO (customized)	63.4	45	98	448 × 448
SSD512*(VGG16)	79.8	19	24,564	512 × 512

SSD is intended for real-time object detection [5, 6]. Faster R-CNN creates boundary boxes using a region proposal network and then uses those boxes to classify objects [3]. While it is called cutting-edge inaccurate, the entire process runs at 7 frames per second, which is much below what real-time processing requires. By eliminating the requirement for the region proposal network, SSD speeds up the procedure. SSD uses a few innovations, such as multi-scale features and default boxes [2], to compensate for the decline in accuracy. These enhancements allow SSD to match the accuracy of the Faster R-CNN utilizing lower quality pictures, increasing the speed even further. Table 1 shows that it reaches real-time processing speed and even outperforms the accuracy of the Faster R-CNN [7].

SSD does not employ a delegated region proposal network. Instead, it boils down to a really simple operation. Both the location and class scores are calculated using small convolution filters. SSD predicts using three convolution filters for each cell after extracting the feature maps. These filters produce the same results as traditional CNN filters.

Recognition accuracy is an essential factor of the model when applied in practice. When the input is noisy (noise: the image is in a dark environment, it's raining or the image is partially obscured...), how does it affect the identification process? In this study, the influence of input noise on the accuracy of recognition will be shown.

2 Research Deployment

It is critical to create a data collection in order to train learning models. Because it has an impact on the trained model's output. The data for training learning models include 10 different flower species that were collected from internet sources.

Table 2 Data collection and labeling for flowers

No.	Name	Total number	Number train	Number test	Label name
1	Apr ¹	50	40	10	Apricot blossom
2	Chr ²	50	40	10	Chrysanthemum
3	Ger ³	50	40	10	Gerbera
4	Hyd ⁴	50	40	10	Hydrangeas
5	Lil ⁵	50	40	10	Lily
6	Lot ⁶	50	40	10	Lotus
7	Nar ⁷	50	40	10	Narcissus
8	Por ⁸	50	40	10	Porcelain flower
9	Ros ⁹	50	40	10	Rose
10	Sun ¹⁰	50	40	10	Sun flower
Sum		500	400	100	10 (flowers)

2.1 Data Collection and Flower Labeling

A total of 500 photos of objects were gathered for the training of geometric models [8]. The objects (flowers) are labeled and divided into two data sets: one data model was trained to account for 80% of the total item recorded, while the test data set was trained to account for 20%. Data sets for teaching and testing are chosen at random.

The LabelImg software is used to label the objects during the picture preprocessing stage. Table 2 details the number of photos of each object that was gathered and tagged.

2.2 Operating Model Environment

Experimental author on PC Intel: CPU core i7 9700F, Memory (RAM) 32 GB, Hard Drive (SSD) 128 GB, Graphics card (VGA) 1050TI.

2.3 Model of Training for Learning

SSD architecture based on VGG with 256 output channels, 3×3 kernel, 2×2 stride, and pad 1×1 (Fig. 1).

The author model's training was halted during the training phase due to a tensorboard graph and a histogram of loss over time. As demonstrated in Fig. 2, the loss in training ranges from 0 to 1.5 in step 12,000. As a result, after the model has been trained to this step limit, learning can be stopped. At step 18,000, the author

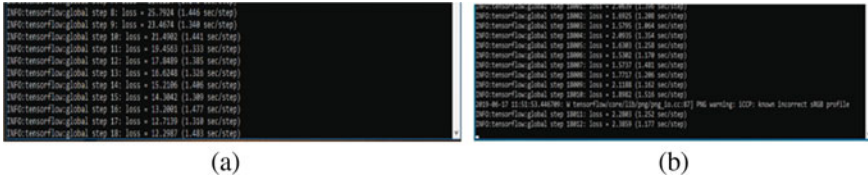
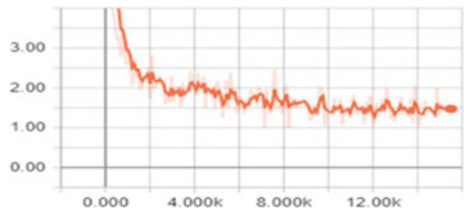


Fig. 1 **a** The procedure for beginning model data training; **b** the procedure of terminating model data training

Fig. 2 Loss chart over time of the model



finished training the model and received a value of 1.5, which reflects the training loss (Fig. 1b). One step takes an average of 1.300 s to train.

2.4 The Real Model Operation

With 15 samples (pictures) for each flower and an identification process for four distinct environmental variables, the author created a real-life identification model for recognizing 10 species of flowers. The photos were acquired from a Google video source and were inspired by reality. A total of 600 (images) were collected for the identification model [9]. The findings of the author’s photo identification have been preserved in reference [10].

2.5 The Performance of the Algorithm

The performance of the recognition process is based on the number of correctly recognized sample images divided by the total number of recognized model sample images.

$$A(\%) = \frac{S}{TS} 100;$$

where:A: Accuracy of the algorithm;

Table 3 The findings of item identification are fully lightened

		These are the flowers that the model identified										
		Apr	Chr	Ger	Hyd	Lil	Lot	Nar	Por	Ros	Sun	Total
Identification results	Wrong identification	0	0	0	0	0	0	0	0	0	0	0
	Correct identification	15	15	15	15	15	15	15	15	15	15	150
	No identification	0	0	0	0	0	0	0	0	0	0	0
	Total	15	15	15	15	15	15	15	15	15	15	150
	Accuracy (%)	100	100	100	100	100	100	100	100	100	100	100

S: Number of the correctly identified sample images;
 TS: Total number of the identified model sample images.

3 Actual Model Performance

Identification result conventions: A verified input sample produces the correct identification result; the effect of poor identification with the validated input sample produces a false identification result. An unidentified result is one that does not identify any species or recognizes more than one species.

3.1 The Results of Identification with the Object Is Fully Lightened

Table 3 illustrates the outcomes of model recognition when the image is not shaded. Table 3 shows a total of 150 input control samples in the red box, and the number of samples defined by the model in the blue box. The findings revealed that all samples were correctly recognized. In this scenario, the model accurately recognizes and the accuracy rate is 100%.

3.2 The Results of Identification with the 1/3 Size of the Object is Darkened

Similarly, Table 4 shows that the model recognized 122 objects out of 150 input samples, resulting in an identification rate of 81.3%. There were four objects in

Table 4 The results of identification with the 1/3 size of the object is darkened

		These are the flowers that the model identified										
		Apr	Chr	Ger	Hyd	Lil	Lot	Nar	Por	Ros	Sun	Total
Identification results	Wrong identification		1				2		6			6
	Correct identification	11	12	15	15	9	11	11	8	15	15	122
	No identification	4	2			6	2	4	1			22
	Total	15	15	15	15	15	15	15	15	15	15	150
	Accuracy (%)	73.3	80	100	100	60	73.3	73.3	53.3	100	100	81.3

this scenario that had a 100% identification rate. The model correctly recognized 3 samples, 1 sample was not detected, and 6 samples were incorrectly identified. Porcelain flowers had the lowest recognition rate, with a ratio of 53.3%. There are 22 unidentified objects and 6 false positives in this environment.

3.3 *The Results of Identification with the 1/2 Size of the Object is Darkened*

Table 5 shows that we have 150 objects, with the model identifying 67 of them. The identification rate for this scenario is 44.7%, and no object has a 100% identification rate. With an accuracy score of 86.7%, the rose specie has the best identification accuracy, while the apricot blossoms specie has the worst with a rate of 20%. Using 15 objects samples as input The model detected three samples, whereas ten samples were not identified and two samples were incorrectly identified. There were 51 correctly recognized objects in total, with 7 incorrectly identified objects. Moreover, half of the model objects were not detected when the object was occluded 1/2.

3.4 *The Results of Identification with the Object is Fully Darkened*

Table 6 reveals that a total of 93 objects samples were accurately recognized. With 15 input samples, the model correctly identified one sample, four samples were incorrectly identified (Lily: three samples; Apricot Blossom: one sample), and ten samples were not identified. There were 28 unidentified samples and four incorrectly recognized samples in the case of the objects in the dark.

Table 5 The results of identification with the 1/2 size of the object is darkened

		These are the flowers that the model identified												
		Apr	Chr	Ger	Hyd	Lil	Lot	Nar	Por	Ros	Sun	Total		
Identification results	Wrong identification	2	2	4	0		2				1	11		
	Correct identification	3	9	9	6	4	9	4	5	13	5	67		
	No identification	10	4	2	9	11	4	11	10	2	9	72		
	Total	15	15	15	15	15	15	15	15	15	15	150		
	Accuracy (%)	20	60	60	40	26.7	60	26.7	33.3	86.7	33.3	44.7		

Table 6 The results of identification with the object is fully darkened

		These are the flowers that the model identified										
		Apr	Chr	Ger	Hyd	Lil	Lot	Nar	Por	Ros	Sun	Total
Identification results	Correct identification	15	15	8	11	7	5	8	1	8	15	93
	Wrong identification			6		1		1	10			18
	No identification			1	4	7	10	6	4	7		39
	Total	15	15	15	15	15	15	15	15	15	15	150
	Accuracy (%)	100	100	53.3	73.3	46.7	33.3	53.3	6.7	53.3	100	62

Table 7 The effect of the shade on the model recognition

	State of the object				Level of significance (P)	LSD _{0,05}
	Object fully lightened	1/3 object is darkened	1/2 object is darkened	Object fully darkened		
A (%)	100.0 ± 0.0 ^a	81.3 ± 17.7 ^a	44.7 ± 21.1 ^b	62.0 ± 31.3 ^b	0.000	18.9

Note Values followed by the same letter are not significantly different based on LSD Test with $\alpha = 5\%$

3.5 Comparison of the Effect of the Shade on the Model Recognition

To test whether the noise affects the model recognition, we compared the accuracy of the model recognition corresponding to different part shades.

The results showed that the accuracy of the model recognition when the object is fully lightened, 1/3 size of the object is darkened, then 1/2 size of the object is darkened and the object is fully darkened is 100.0 (%), 81.3(%), 44.7(%) and 62.0 (%), respectively (Table 7). The results of the analysis of variance (ANOVA) illustrated a significant difference ($p < 0.05$) in the accuracy of the model recognition from different part shades. The accuracy of the model recognition was significantly higher in the case of the object being fully lightened and 1/3 size of the object being darkened than in the case of 1/3 of the size of the object being darkened and the object being fully darkened ($p < 0.05$, Least Significant Difference Test). However, there were no significant differences were found between the object being fully lightened and 1/3 size of the object being darkened ($p > 0.05$, Least Significant Difference Test). A similar tendency was detected also for 1/2 size of the object is darkened and the object is fully darkened ($p < 0.05$, Least Significant Difference Test).

4 Conclusions

In this paper, we have proposed an experimental method for the SSD model to detect objects in normal states and noisy states. The algorithm has been shown to be able to detect objects under poor conditions, such as changes in illumination, 1/3, 1/2 size of the object is darkened. The results showed that the detection accuracy decreases when the subject is placed under poorer conditions. The proposed algorithm achieves modern detection accuracy of 100.0% and 62.0%, the object is fully lightened and the object is fully darkened, respectively. The accuracy rate of the model is also reduced in the case of 1/3 and 1/2 of the objects being obscured, to 81.3% and 44.7%, respectively. This research result will certainly bring offer much value to the application of the SSD model in practice. In our future works, we will aim to

improve the recognition accuracy of the model when the object is placed under poor conditions.

References

1. Liu W et al (2016) SSD: single shot multibox detector. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)* 9905:21–37
2. Shuai Q, Wu X (2020) Object detection system based on SSD algorithm. *Proc 2020 Int Conf Cult Sci Technol ICCST 2020*, 141–144
3. Abbas SM, Singh SN (2018) Region-based object detection and classification using faster R-CNN. *Int Conf Computational Intell Commun Technol CICT 2018*, pp 1–6
4. Liu B, Zhao W, Sun Q (2017) Study of object detection based on Faster R-CNN. *Proc 2017 Chinese Autom Congr CAC 2017*, 6233–6236
5. Kanimozhi S, Gayathri G, Mala T (2019) Multiple real-time object identification using single shot multi-box detection. *ICCIDS 2019—2nd Int Conf Comput Intell Data Sci Proc*, pp 1–5
6. Kang HJ (2019) Real-time object detection on 640×480 image with VGG16+SSD. *Proc 2019 Int Conf Field-Programmable Technol ICFPT 2019*, 419–422
7. Hui J (2018) SSD object detection: single shot multibox detector for real-time processing [Online]. Available: <https://jonathan-hui.medium.com/ssd-object-detection-single-shot-multibox-detector-for-real-time-processing-9bd8deac0e06>
8. Drive G (2022) Training image for learning model [Online]. Available: <https://drive.google.com/file/d/1FRzMiQQsOQ9uHsJSDkxCRCELoHuPFp-d/view?usp=sharing>
9. Drive G (2022) Image included into identification [Online]. Available: https://drive.google.com/file/d/1uglOpvn9G9ad-aNn0rjLio_FFT7ELNI5/view?usp=sharing
10. Drive G (2022) Image is identified by the model [Online]. Available: <https://drive.google.com/file/d/1cz58OjMJSp8MwcZjObODEzCc20l0jdsG/view?usp=sharing>