Deep Learning for Detecting Quilt Status of Sleeping Children

Ruiqi Liu

Cambridge International School of Optics Valley, ULINK College, No. 101, Fushikang West Road, Jiangxia District, Wuhan, Hubei, China, 430399, China

Abstract. Sleeping is important for the physical growth of infants and toddlers. However, because of the growth of their bodies, they may kick off the quilt when they fall asleep and catch a cold as a result. Because it is difficult to extract the all-appropriate features that insists the quilt is not covered in the images taken by the home camera manually, which is suitable for the old types of symbolic AI to recognize, there has not been a suitable solution to this problem in the past. However, thanks to the rapid development of artificial intelligence and neural networks, intelligent target detection, which was previously difficult to achieve, has finally been realized and several approaches to intelligent target detection have emerged, leading to many different perspectives and algorithms. In this paper, by analyzing and comparing the advantages and disadvantages of different algorithms in different application scenarios, we found that YOLO (v5 in this paper) is very suitable for quilt status detection and realized the program by applying YOLOv5 neural network and cosine difference algorithm simultaneously.

Keywords: Detecting Quilt Status; Sleeping Children; YOLOv5 Neural Network.

1. Introduction:

With the development of technology, people gradually realized the limitations of symbolic AI and consequently developed neural networks to achieve automatic extraction and detection of features in the real world. Compared to previous algorithms, neural networks have the advantage of extracting and analyzing data features autonomously, with unprecedented sensitivity to unpreprocessed data such as audio or images. However, a few years ago, having a full-fledged neural network often required significant human and material resources because training data was not easily available and neural networks required a lot of material for training to achieve excellent accuracy. However, the popularity of the Internet and IOT has made it easy to collect information, and the reliability of most neural networks is rapidly improving, gradually meeting and exceeding people's expectations and rapidly entering people's lives. Moreover, IOT will also undertake the work of data collection, and eventually a virtuous circle will be formed and the development of artificial intelligence accelerated.

Currently, convolutional neural network (CNN)-based target detection and recognition algorithms are widely used and various extended versions from R-CNN algorithm to YOLO algorithm have been generated. By 2022, a number of algorithms for target detection have become very mature after continuous training, and have been successfully applied in areas such as fruit and vegetable ripeness detection, pneumonia detection, and automatic my clearance.[1][2] Artificial intelligence has greatly reduced the workload of people and engage in a variety of dangerous operations in places where humans cannot reach, greatly guaranteeing the life quality and safety of people. It also can be used in the quilt status detection to make parents' life easier by the application of the target detection algorithm. When the live surveillance footage from the child's room is imported into the program, the algorithm will automatically return a box containing the target (i.e., the child) and its corresponding confidence level. When the confidence level reaches a certain threshold, the program will trigger an alarm to alert parents that their child's quilt has come loose and needs to be tucked back in. In fact, there are already a large number of products on the market that are designed to watch over children's sleeping conditions. But the vast majority of them are designed to restrain children's hands and feet so that they cannot kick off the quilt. It might not be the best choice conceding the children's comfort. The quilt status detection program to be developed can great relieve parents' burden of having to constantly check on the kid, would not bring any side effect without having to retrain the kid.

2. Background:

The idea of developing quilt status detection program to solve this problem came to me when I was little. I was not as strong as others of the same age. Once I kicked off the quilt in cold weather, I was bound to get a cold and pharyngitis the next morning, and even respiratory infections and asthma during allergy season. To prevent me from suffering from the disease, my parents got up every two hours every night to check if I was covered up tightly, which was also tiring for them.

After my first exposure to coding and AI, I had tried to contrast the uncovered image with the live video. If the similarity of pixels of the same color was higher, it would be judged that the quilt was not well tucked in. As expected, even though I had processed the image in black and white, the changing light and the fluctuating curtains constituted indeterminable factors, and only 6% were successfully predicted in the end.

After learning the computer vision algorithm composed of full connected neural network, I updated the previous program and trained it for thousands of times. This experiment yielded much better results than in the past. Not only the probability of accurate identification has increased to 68%, but also the number of false alarms has decreased a lot, only 22%. Despite the much-improved success rate of prediction, this algorithm has three very obvious drawbacks as follows:

1. First, this full connected neural network is easy to form over-fitting, which makes the prediction accuracy of AI in judging the training set and the simultaneously collected test set much higher than that in practical application. The table below indicates the performance of the neural network when testing in the same room the dataset is collected.

Table 1. The performance of the neural network when testing in the same room

	Accuracy rate (number of correct identifications/total number)	Misjudgment rate (number of false positives / total number of predicted results)
Training set	97%	4%
Test set	88%	10%
Practical application	68%	22%

(Rounded and retained two decimal places)

Another side effect of over-fitting is that a universal neural network cannot be trained. After testing in my own room, I re-tested in my own parents' room. The program presented predictions that were very different from those in my room, and the accuracy rate decreased dramatically, and the misjudgment rate went up to 56%. This also means that the program must be trained specifically for each location used in order to get relatively reliable results. This is the table that indicates the performance of the neural network in a different room where dataset is collected.

Table 2. The performance of the neural network in a different room

	Accuracy rate (number of correct identifications/total number)	Misjudgment rate (number of misjudgments / total number of predicted results)
Practical application	33%	56%

(Rounded and retained two decimal places)

2. The second disadvantage of a full connected neural network is that it has a low upper limit. In this experiment, I constructed three full connected neural networks of different depths and recorded the loss values for each full connected neural network of different depths at every 100 times they are

trained. According to the loss function table above, it can be seen that with the increase of the number of deep regressions, although the full connected neural network keeps learning at the beginning to reduce the loss, after 400 regressions, the loss value remains basically constant and will not decrease with the increase of the number of regressions. After the neural network reached 5 layers, the improvement in prediction accuracy brought by increasing depth was significantly reduced, with the prediction accuracy staying at about 32%, and the computational speed was greatly reduced. It is found from the above experiments that neither increasing the number of deep regressions nor increasing the depth of the network can overcome the accuracy bottleneck of the full connected neural network, but will reduce the probability of correct prediction due to over-fitting. The line chart below represents the test loss of different size of neural networks at different times of logistic regression.

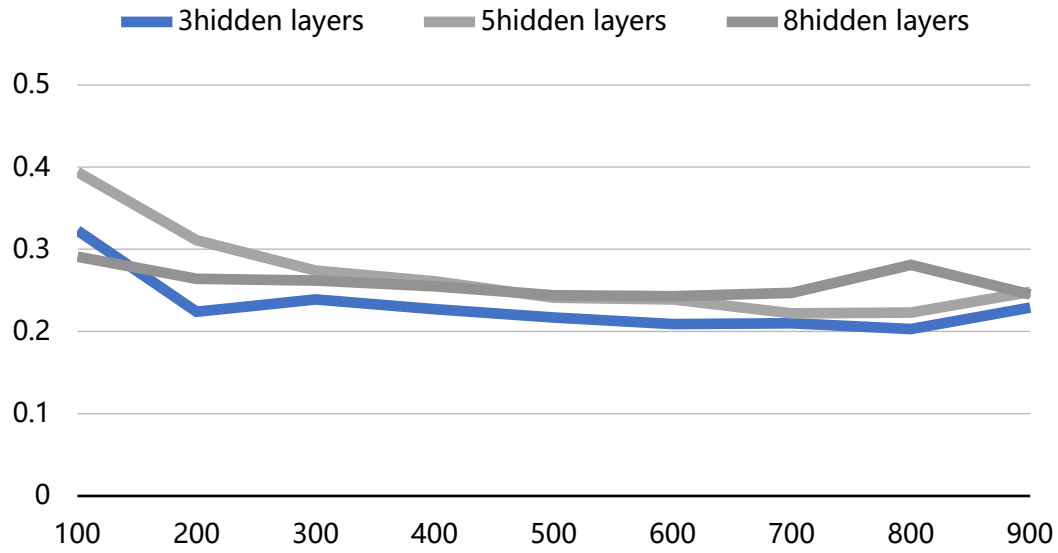


Fig 1. The test loss of different size of neural networks at different times of logistic regression

3. The actual computation speed is too slow to be easily monitored in real time. However, since quilt status detection during children's sleep does not require continuous and constant prediction, this drawback is not obvious in this application scenario.

In summary, both the low robustness and environmental adaptation of the full connected neural network and its low accuracy rate make it incompetent for the task of detecting quilt status during children's sleep.

After further research, I gradually discovered the excellent accuracy of some target detection algorithms in target localization and judgment and intend to compare the advantages and disadvantages of these algorithms and finally select the most suitable algorithm for this application scenario. However, due to the lack of experimental equipment and the lack of time to conduct research on integrate the program into a hardware, our algorithm was not practically applied in a surveillance camera with real-time recording function.

3. Methodolgy:

It is expected that the quilt status detection algorithm can be integrated into a regular home camera. This feature will be built in a camera and a warning will be sent to the user via the APP associated with the device.

1. Comparison of YOLO and R-CNN algorithms

Convolutional neural networks specializing in image recognition have significantly higher robustness and computational efficiency than the unmodified basic full connected neural networks. Therefore, many researchers have developed various target recognition algorithms based on convolutional neural networks, each with different advantages and disadvantages. In this study,

different target recognition algorithms have been compared and contrasted so as to select the most suitable algorithm for quilt status detection during children's sleep, and then the program was optimized.

The effectiveness of a target recognition algorithm is judged by three main indicators - IOU, mAP and FPS. They represent the intersection-over-union(IoU), mean average precision (mAP) and Frames per Second (FPS) of the predicted result to the true value. Theoretically, higher mAP value means more accurate prediction. The most commonly used target recognition algorithms currently are R-CNN and YOLO. The former was proposed by Ross Girshick [1] who solve the problem that the DPM algorithm is hard to train and predict, and optimized the computational efficiency while maintaining good prediction accuracy. Joseph Redmon [2], the creator of YOLO made YOLO achieve extremely high FPS by simplifying the algorithm's arithmetic logic and drastically reducing the number of steps required for prediction. Both algorithms are able to maintain a good FPS while having a high average precision, which makes them very adaptable to real-time target detection and recognition. The graph below indicates the performance of YOLO comparing to other algorithms.

Real-Time Detectors	Train	mAP	FPS
100Hz DPM [31]	2007	16.0	100
30Hz DPM [31]	2007	26.1	30
Fast YOLO	2007+2012	52.7	**155**
YOLO	2007+2012	**63.4**	45
Less Than Real-Time			
Fastest DPM [38]	2007	30.4	15
R-CNN Minus R [20]	2007	53.5	6
Fast R-CNN [14]	2007+2012	70.0	0.5
Faster R-CNN VGG-16[28]	2007+2012	73.2	7
Faster R-CNN ZF [28]	2007+2012	62.1	18
YOLO VGG-16	2007+2012	66.4	21

Fig 2. The performance of YOLO comparing to other algorithms
(Picture from: You Only Look Once: Unified, Real-Time Object Detection; Joseph Redmon)

The main difference between YOLO and R-CNN is that the former is oriented towards computing speed and training efficiency while the latter aims at prediction accuracy. R-CNN sacrifices efficiency for high accuracy. YOLO is exactly the opposite, which means that if R-CNN is to be used for real-time target detection, the detection instrument should have strong computing power. Therefore R-CNN is of no practical value in quilt status detection. In order to achieve high accuracy, R-CNN first needs to frame out the target and only after limiting its possible range can the target detection algorithm be used to predict the probability of that target's occurrence. Considering the cost and space constraints, it is basically impossible for manufacturers of home cameras to place processors with strong computing power in such small devices. Otherwise, the cost of the camera will increase exponentially and the size will become very big, which makes it difficult for the camera to be placed in homes. Therefore, we chose the YOLO algorithm with end-to-end operation mode in order to adapt to the computing power of the home camera processor. Although the mAP of YOLO is not as good as that of R-CNN after the same training, there is a very significant difference in speed.

In real-time detection, YOLO boasts a mAP of 63.4, which is much higher than the value of the DPM algorithm. In the non-real-time detection experiments, the YOLO VGG-16 network achieves nearly 3 times the FPS of the Faster R-CNN while having mAP values almost comparable to the latter. Therefore, we chose the YOLO algorithm as the main body of quilt status detection, and all subsequent optimizations will be tuned around the training and output of YOLO.

Since the release of the YOLO algorithm in 2016, it has undergone a total of six major updates aimed at continuing to optimize the accuracy and computational speed of the predictions. The latest version of YOLO is YOLOv7 [3]. However, considering that this version may not be very stable

because it's newer and that the mature and widely used YOLOv5 already has a similar performance to v7, we decided to use YOLOv5 [*] version to ensure stability.

2. Experiment design

Since child sleep status detection is not a common application scenario for image recognition, it is extremely difficult to find relevant datasets on the web, and thus we can only obtain datasets for training and testing all by ourselves. In order to avoid over-fitting due to the small amount of trained data, we recorded in two different rooms for a single subject and multiple subjects. The second difficulty is that the image displayed when the quilt is kicked off may, in some cases, have a high degree of similarity to the image without anomalies. When only the side facing the camera is obscured by the quilt, there is a large difference between the actual image taken by the camera and the real situation, and the neural network may make false negative predictions as a result. To solve this problem, we labeled some features outside the subject of the experiment, including bare legs, arms and shoulders, and summarized the second criterion for determining the sleep status by figuring out the number and probability of the detected labels in the current image with cosin difference. After the prediction result of YOLOv5 is out, the program will call the probabilities of being leg, arm and shoulder to compare with the cosine difference table summarized from the training set in bid to calculate the probability that the quilt is not kicked off. The calculation result is merged with the result of YOLO network to get a more accurate output.

Since most of the bedtime is at night and some people would prefer to block out all light while sleeping, the selected camera must have night vision capability. Therefore, the main data for this test are taken from home cameras with infrared night vision capability, and the use of devices without night vision capability has not been taken into account.

The specific experimental steps are as follows:

1. Record single target and multiple targets in 2 different rooms, which are divided into training set and validation set and labeled, and input the training set into YOLOv5 network for training

2. Input the correct predictions in the validation set and the probabilities of each label into the cosine difference algorithm to be used as a comparison object for the predictions in actual use

3. Input the test set collected in the same scene as the training set into the neural network for prediction, and record the results

4. Input all negative results predicted by the YOLOv5 network into the cosine difference function for which initial data already exists, and compare the effect of this function on the prediction accuracy

3. Process of experiment

Considering the small size of the dataset, we chose YOLOv5s as the model and controlled the epoch at 100 times to control the occurrence of over-fitting. This experiment used python 3.9, pytorch 1.12.1 CPU version for training on a 16-inch MacBook Pro 2019 model. We used the pre-trained model of YOLOv5s as the initial weight, while the batch-size was set as 2 due to the unavailability of CUDA (the specific training statement used locally: python train.py --data cover_detection.yaml --cfg yolov5s.yaml -- (weights pertained/yolov5s.pt --epoch 100 -batch-size 2). This training took a total of 5 hours and 59 minutes, with a total of 501 labeled images, and a total of 52 labeled images in the test set.

In the test set of 52 images, a total of 44 were accurately predicted. The second step of the experiment is to input the labels and corresponding probabilities of these 44 images into the cosine difference algorithm and summarize them into a table containing the degree of similarity between different results.

4. Training results and validation

After 100 training sessions, we called the weight that best matched the final solution, tested the test set, and finally obtained the results shown in the figure above. After manually reviewing the prediction results of the YOLOv5 network for the test set, it was found that the predicted values of the 44 images were consistent with the actual values, and all labels in 40 images were correctly identified. The following image is one of the test sets in which all labels were correctly identified:

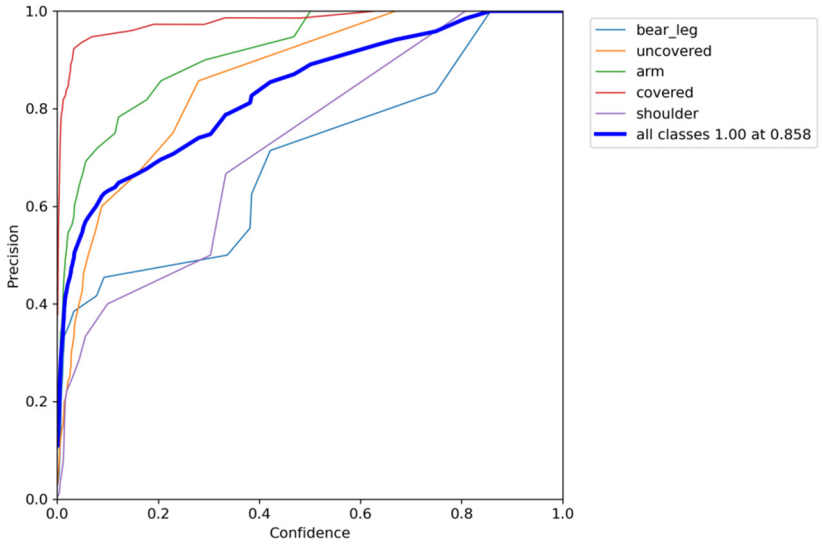


Fig 3. The above figure shows the P_curve of the test set



Fig 4. One of the test sets in which all labels were correctly identified

Table 3. The probability that no quilt is detected before and after prediction by the cosine difference algorithm for the four images with changes

(Round to 3 significant figure)	YOLOv5	Cosine difference	Average
Case1	0.214	0.803	0.509
Case2	0.494	0.631	0.653
Case3	0.388	0.721	0.555
Case4	0.642	0.300	0.471

(The row in red represents the case when the cosine difference algorithm miss corrected the original value)

The second step is to input all the negative images (covered here, 29 in total) into the cosine difference function and finally output the second prediction by comparing the similarity with the previously deposited values. Before all the prediction analyses were completed, the predicted values of a total of 25 out of 29 images agreed with the actual results. After that, the results of 4 images differed from the initial results after averaging with the output of the YOLO network, and the false negatives were successfully modified to positives in 3 of them. However, the correct prediction was incorrectly modified to a false positive in one of them. The following table shows the probability that no quilt is detected before and after prediction by the cosine difference algorithm for the four images with changes.

The current experimental data shows that the combination of YOLO and cosine difference is an effective and realistic algorithm for detecting quilt status.

4. Conclusion and Future Works:

1. Summary on this experiment

According to the results of the previous paragraph, the YOLOv5 algorithm is able to detect children's sleep quality quickly and achieve a good success rate at the same time. The similarity between the predicted and actual values of the test set reached 46/52. Since the detection of quilt status does not require accurate localization and marking of the target in every frame as in the case of target tracking, the algorithm in this experiment has basically fulfilled most of the requirements for use in real-world scenarios. The YOLO algorithm is a highly integrated open-source algorithm, which makes it extremely easy to use and suitable for most students who have little knowledge of computer vision. Another advantage of YOLO is that its data output is more complete, and the information contained such as the likelihood of corresponding items appearing in the target box. F1curve and label statistics plots can help users to adjust the data after training is completed for the purpose of finding the best hyperparameters.

In addition to the YOLO algorithm, another algorithm, cosine difference, was used in this experiment to continue analyzing the results based on the output after the YOLO network calculation was completed, so as to avoid the occurrence of false negative predictions. This algorithm avoids the occurrence of misjudgments to a certain extent and also serves as a reference to assist the program in making predictions, and was proven to be practically effective in this experiment.

2. Reflection on the imperfections of the design of this experiment

Firstly, in this experiment, only a small amount of data was collected for the experiment because of insufficient time for preparation, and the validation set was not assigned in order to put more data into the neural network for training, so the completion of the experiment still needs to be enhanced. Secondly, the YOLO algorithm has been continuously updated since its birth in 2016, so it is slightly unfortunate that the latest YOLOv7 network was not used in this experiment, as the older YOLOv5 network does not necessarily meet the highest standards of current computer vision. Finally, the cosine difference algorithm was not better applied. Although the application of the cosine difference algorithm succeeded in improving the accuracy of the final prediction by 3.8%, the improvement was not as significant as we expected. It is hoped that the auxiliary ability of this algorithm will be improved when the data is supplemented.

3. Technical and hardware issues encountered in this experiment

Currently, our algorithm is not able to connect to live video stream and monitor the status of the quilt at real time. There will be more difficulties in the real application that we cannot anticipate at present. We hope that we can continue to improve the experimental design and our knowledge base in the future research, and develop a more improved algorithm.

References

[1] The Improved YOLOV5 Algorithm and Its Application in Small Target Detection; Yu Zhao, Yuanbo Shi & Zelong Wang; 2022.

[2] Fruit Target Detection Based on BCo-YOLOv5 Model; Ruoli Yang, Yaowen Hu, Ye Yao, Ming Gao and Runmin Liu; 2022.

[3] Rich feature hierarchies for accurate object detection and semantic segmentation; Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik 2014.

[4] You Only Look Once: Unified, Real-Time Object Detection; Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi 2016.

[5] YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors; Chien-Yao Wang, Alexey Bochkovskiy, Hong-Yuan Mark Liao.