REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1S (2024)
http://www.veterinaria.org
Article Received: 05-05-2024 Revised: 25-05-2024 Accepted: 08-06-2024

Deep Learning In Image Processing For Object Recognition Using Various Techniques

Harshit Goyal[1*], Priyank Sirohi[2]

[1*] M.Tech CSE, SCRIET, Chaudhary Charan Singh University, Meerut, India. harshitg503@gmail.com
[2]Assistant Professor, SCRIET, Chaudhary Charan Singh University, Meerut, India. priyanksirohi01@gmail.com

ABSTRACT

Due of the field's tight ties to both picture interpretation and video analysis, object detection has seen a sharp rise in study interest in recent years. Shallow trainable structures and handcrafted characteristics are the foundation of traditional object recognition methods. Their efficacy quickly bottoms out because they build intricate ensembles that integrate higher level input from the object identification systems and the scene classifiers with low-level image data. Deep learning is developing at a rapid pace, giving researchers more effective methods to tackle issues with conventional architectures. These instruments are able to pick up more complex, semantic data. The network design, training procedure, optimization function, and other components of these models vary. An overview of object identification techniques based on deep learning is provided in this study. First, let's review deep learning and the convolutional neural network (CNN), which is its primary tool.

Now next discussion is for common generic object detection architectures and provide some useful tips and adjustments to improve the detection performance for further tasks. Additionally, as many particular detection tasks have distinct features, such as salient object recognition, we briefly examine a few specific tasks, such as face and pedestrian identification. Additionally, experiments are provided in order to assess various strategies and derive some significant conclusions. A range of numerical values reflect all that is visible to a computer. Thus, in order to examine the data contained in images, they need image processing algorithms. In terms of efficacy and speed, You But Look Once (YOLO), More Quickly Region-based neural networks based on convolution (Faster R-CNNs), and Single Shot Detection (SSD) are the most popular computational processing of pictures algorithms. This essay examines these techniques. This evaluation looks at the performance of these three algorithms and looks at their individual benefits and drawbacks using metrics like F1 score, accuracy, and precision. The technology being used for data collection is Microsoft COCO (Common Object in Context). The experiment's findings show that each algorithm's predominant use cases over the other two determine its relative superiority. YOLO-v3, the most optimal algorithm out of the three, outperforms the SSD drive and Faster R-CNN networks under the same testing conditions. In conclusion, a number of fascinating opportunities and challenges are outlined as a basis for further research in the fields of relevant neural network-based learning systems and object identification.

INTRODUCTION

The science of computer intelligence is witnessing a transformation in several industries, including object and photo identification, thanks to the strong technology known as deep learning. These strategies have improved the computer vision systems capacity to interpret and process visual input correctly and efficiently. The topic of deep neural network techniques for picture and image identification is an intriguing one, and this article will examine the fundamental ideas, mechanisms, and applications of these techniques. Picture recognition is the technique of automatically identifying and classifying patterns or moving or static items in digital pictures. It is necessary for many real-world applications, notably security cameras, driverless cars, and virtual and augmented reality. Convolutional neural networks, or CNNs, in particular, have shown exceptionally good performance in tests of picture recognition using deep learning approaches. This is because they can rapidly infer from the values of individual pixels hierarchical representations of visual data. This enables them to identify and automatically extract significant areas from photographs with previously unheard-of accuracy. Object detection finds and recognizes objects inside an image, as opposed to image recognition. Creating bounding boxes around items that have been detected yields precise geographic data. To produce precise and effective item localization, artificial intelligence-based object identification methods blend image recognition abilities with extra tactics like region suggestion strategies and spatial changes. Security surveillance footage, autonomy robots, and image search are just a few of the real-world uses for these methods.

Convolutional neural networks, or CNNs, are the foundation of most deep learning-based systems for identifying items and photo identification. CNNs use many layers of interconnected neurons to handle increasingly complex details, mimicking the visual retrieval mechanism of the human brain. Low-level components such as edges and colours contribute early on, yet the later layers absorb higher-level, conceptual data. Due to hierarchical feature extraction, CNNs are able to represent complex relationships more accurately than other models, giving them a competitive advantage on visual interpretation tasks. Biggest advance in deep learning for item and image identification has happened with the availability of completely analyzed data sets like ImageNet and COCO. Millions of tagged photos make up these datasets, which let CNNs learn various visual representations for various object categories. Thanks to the rapid advancements in

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1S (2024)
http://www.veterinaria.org
Article Received: 05-05-2024 Revised: 25-05-2024 Accepted: 08-06-2024

GPU technology and distributed computing frameworks, which have accelerated the training and inference rates of these models, deep learning models can now be employed in real-time applications.
Many deep learning architectures have been developed recently to improve object and image identification performance. Think of popular models that have shown state-of-an-art on benchmark datasets that includes VGGNet, AlexNet, GooLeNet, and ResNet. Deeper networks, skip connections, and residual learning are used in these architectures to cope with common issues including fading gradients and enhanced model capacity. Furthermore, a variety of methods for enhancing deep learning-based detection of objects and image recognition systems have been examined for their effectiveness and stability. To improve pre-trained models on domain-specific raw data, transfer learning makes use of data from huge datasets. By employing data augmentation methods like image cropping, scaling, rotation etc the model's ability to generalize is enhanced and the variety of training data is expanded. Moreover, the incorporation of attention methods that prioritize significant characteristics has enhanced the efficacy and rendered deep learning models more comprehensible.
Deep learning techniques are widely used in various fields, including object recognition and image identification, and they are continuously being improved. These include situation-aware self-driving automobiles, intelligent surveillance systems that can recognize and monitor objects in real time, and medical gadgets that can accurately diagnose illnesses based only on visual data. Robots will be able to analyze visual data as accurately as or more accurately than humans as deep learning techniques progress, which should lead to even more significant developments in computer vision. One popular family of deep learning methods for object identification and image recognition is the region-based convolutional neural network (R-CNN) family. CNNs are used for classification after R-CNN models generate region recommendations using a two-stage pipeline. This approach is used in many real-world applications and has demonstrated outstanding performance on multiple benchmark datasets.
After it was shown that the two-stage R-CNN approach was computationally inefficient, single-stage models, like Single Shot Multi Box Detector (SSMBD) and You Only Look Once (YOLO), were added to additional research. More importantly, these models are faster and more efficient since they can predict bounding box coordinates in addition to item classes in a single network run. While there could be a little accuracy trade-off as compared to two-stage devices, both speed and accuracy have significantly improved in later generations. Deep learning applications in object and picture detection require huge annotated datasets. Significant advantages have been demonstrated by training and assessing deep learning models on datasets like ImageNet and COCO. Deep learning models' potential has also been increased by more recent developments in network topologies, such as feature pyramid networks, residual connections, and attention strategies. These methods enable models to gather fine-grained attributes, adjust for changes in scale, and concentrate on pertinent regions, which improves their performance on difficult tasks.
In computer vision, object recognition is an essential problem with applications ranging from autonomous driving to medical imaging. The traditional approach relied heavily on machine learning algorithms and human extraction of characteristics. With the advent of deep learning and convolutional neural networks (CNNs), identification of objects systems have modestly but not considerably increased in accuracy and efficiency. This study compares the methods, benefits, and drawbacks of the main deep learning approaches utilized in object detection image processing.

LITERATURE REVIEW

- Abhinav, A., & Agrawal, A. (2019).

An extensive analysis of deep learning methods for image identification.

- Chen, Z., & Gupta, S. (2020).

A thorough review of deep learning for finding objects.

- Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014).

Rich feature hierarchies are needed for accurate object identification and semantic segmentation.

- Ren, S., He, K., Girshick, R., & Sun, J. (2015).

Quicker R-CNN: Moving toward real-time item identification using region proposal networks.

- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016).

You only look once: Unified, real-time object detection.

- He, K., Gkioxari, G., Dollar, P., & Girshick, R. (2017).

Mask Region based Convolutional Neural Network.

- Redmon, J., & Farhadi, A. (2018).

YOLOv3: An Incremental improvement.

- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016).

Reevaluating the computer vision inception architecture.

- He, K., Zhang, X., Ren, S., & Sun, J. (2016).

Deep residual learning for picture recognition.

- Lin, T. Y., Dollar, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017).

Feature pyramid networks for object detection.

- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016).

SSD: Single shot multibox detector.

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1S (2024)
http://www.veterinaria.org
Article Received: 05-05-2024 Revised: 25-05-2024 Accepted: 08-06-2024

- Redmon, J., & Farhadi, A. (2018).
YOLOv3: An Incremental improvement.

PROBLEM STATEMENT

Intelligent systems (AI) are those that possess the ability to accurately comprehend outside information, absorb it, and use it to adapt and change in order to carry out certain activities and objectives. The study of algorithms that progressively get more complex over time on their own is known as machine learning, or ML. Machine learning algorithms generate a training model from information samples, so they don't need to be "explicitly programmed to do so," so they may use it to infer or make judgments.

CNN Development: The primary application of artificial neural networks is image processing. Hubel and Wiesel's neurological research on the visual brain served as an inspiration. The main part of the brain responsible for processing visual sensory data is the visual cortex. To identify things in the photographs, it removes characteristics and looks for structures and patterns. Because of the buried convolutional layers, it is different. These layers eliminate patterns from the images by applying filters. The image that results from applying the filter is the outcome. Through various filters, different patterns appear. In the initial layers, filters are used to identify fundamental patterns. They ultimately develop greater sophistication in the following ways as they move through the layers:
Origin (Late 1980s–1990s): The first extensively used CNN was LeNet - 5, which was created in year 1998 {LeCun et al.}. It took about ten years to develop. Its main objective was to identify handwritten digits. It is highly renowned in the deep learning community for its work on creating effective CNNs. Initially, banks used it in ATMs.

Stagnation (early 2000s): Nothing was known about the inner workings of CNNs at this point. Moreover, there was no publicly available picture collection like to Microsoft COCO or Google Photos. Consequently, most CNNs could only do optical character recognition, or OCR. CNNs require additional processing time, which results in higher operating costs. CNN was outperforming Support Vector Machines (SVMs), a specific type of machine learning system.
Revival (2006–2011): According to research by Ranzato et al., It was discovered that utilizing the max-pooling technique for feature extraction produced a significant improvement over the sub-sampling algorithm. CNN training on GPUs has been expedited by researchers. Since NVIDIA developed the CUDa platform, which facilitated and expedited multiprocessing, CNN has been able to validate and train more fast at the same time. This led to additional study being done. The huge image dataset from Stanford University known as Pattern Analysis, Statistical modeling and Computational Learning Visual Object Classes, or PASCAL VOC, was made available in 2010 which was a major step forward.
As a result of AlexNet, CNN's accuracy increased significantly between 2012 and 2013. Merely 15.3% of the errors in the 2012 ILSVR competition were made by it. The error rate of the second-place network was 26.2%. Consequently, AlexNet outperformed every other network that was available at the time by 10.8%. AlexNet demonstrated "deep" learning by achieving this accuracy with a total of eight layers. Even though this needed more processing power, advancements in GPU tech made this possible. Similar to LeNet, AlexNet is the among the important research initiatives that CNN has ever made available to the general public.
Architectural Innovations (2014–2020): The widely utilized and well-liked VGG architecture was developed in 2014. The concept of placing objects in particular regions of the image was first presented by region-based CNNs (RCNNs), which, like many other CNNs, are based on VGG. In the years that followed, Fast RCNN and Faster RCNN—better variants of RCNN—were released. Both approaches reduced calculation times without sacrificing the well-known accuracy of RCNN. Based on a comparable VGG basis, the Single-Shot Multibox Detector (SSD) was created in the first few months of 2016. An alternative algorithm called You Only Look Once (YOLO) was introduced in 2016. The DarkNet served as the basis for its construction. Currently under development, the third iteration was made available in 2018.

EXISTING METHODOLOGIES

SSD: SSD performs faster than other objects identification model such as YOLO or the Fast R-CNN, but it is a much more effective method for object recognition. In an attempt to produce a faster detector, many changes had been made to the detection pipeline prior to the creation of SSD. However, these modifications only resulted in a noticeable speed increase that decreased the detection accuracy. Researchers decided they would have to start from scratch in order to create a new object detection model, which is how the SSD model came to be rather than altering an existing one. The accuracy of SSD models is similar to that of models with features for bounding box conjectures or models with picture resampling. Moreover, it is simpler than systems that rely on object proposals because it performs all processing inside a single network. This is because it eliminates the requirement for resampling phases and the development of pixels or proposals. This makes it simple for systems that use detection as one of their features to integrate and teach SSD.
By combining predictions from numerous feature maps at different scales, Large and small objects can be detected with ease using the SSD (Single Shot MultiBox Detector). Since SSDs do not require a separate area proposal stage as R-CNNs do, they are able to detect objects faster. By identifying the best filter structures that can reliably identify an object's

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1S (2024)
http://www.veterinaria.org
Article Received: 05-05-2024 Revised: 25-05-2024 Accepted: 08-06-2024

properties and increase the amount of training samples that are provided, SSD is able to achieve high accuracy in the evaluation phase and hence lower the loss value.

ANALYSIS OF THE FUNCTIONS

A series of bounding boxes with uniform sizes are created by SSD's intricate feed-forward network, and each time an object arrives inside one, it is given a score. Following the development of the non-maximum suppression score, the final detection results are obtained. A VGG-16 network, a popular architecture for high quality photo classification, serves as the foundation for the first few network levels and is terminated before any classification layers are established. To enable detections, a new structure, such as convo6, is added to the shorter base network.

SSD uses VGG-16 architecture, which is particularly good at identifying high-quality pictures, to extract feature maps. Auxiliary layers are utilized because they allow us to limit the amount of input with each layer that is passed through and extract the required features at different sizes. The layer projects one or more values into each picture cell. Each forecast consists of a bounding box that counts the things inside and gives points for each kind of object that is found inside. An algorithm is used to "guess" what is inside the bounding box and determine which class gets the highest score.

Convolutional predictors: Each feature layer uses convolutional filters to generate a predetermined number of predictions for object detection. A three × three x tiny kernel is required for each feature layer of size x y with n channels in order to provide prediction variables for potential detection outcomes. This kernel produces a shape offset or a confidence score for each class using the default grounding box coordinates provided by the COCO Dataset at each position.

Aspect ratios and default bounding boxes: As you can see, every feature map in the network has a default bounding box assigned to it. The default boxes are in charge of meticulously sketching the feature map so that each box is fixedly positioned with respect to its matching cell. We can estimate the offsets for the default box shapes for each feature map cell and consider the class scores, which indicate the type of item that is enclosed within the bounding box.

COMPLEXITY ANALYSIS

The majority of algorithms may be expressed in big-Oh notation and have a time-complexity connection that is dependent on the volume of input. However, SSD training time and inference time when running the model on a certain hardware are included when evaluating temporal complexity for deep learning models. Processing millions of calculations is necessary for deep learning models, which can be computationally costly. The artificial neural network ultimately does all of these computations simultaneously using millions of similar neurons in each layer. It has been shown that training an SSD model on an Nvidia GeForce GTX 1070i GPU results in ten times quicker training speeds due to its parallel nature.

The greatest time-complexity-consuming part of the basic CNN forward pass is matrix multiplication. Apart from the specifics such as the number and size of filters, the size of the feature extraction map, the number of neurons in each layer, the number of layers in the CNN, and the picture quality, these factors also affect the total number of multiplications. Each layer's activation functions are called ReLu functions, and they are all determined to operate in quadratic time for every neuron involved. Thus, by accounting for each of these elements, we can ascertain the temporal complexity of the forward pass at the base CNN:

$$time_{forward} = time_{convolution} + time_{activation} = O\left(\sum_{b=1}^{B} x_{l-1}\,.\,(h.\,h)\,.\,x_b\,.\,(s_b\,.\,s_b)\right) + O\left(B.\,x_c\right) = O\left(weights\right)$$

The number of filters used in the b layer, their width and height, the number of neurons, and the size of the output feature map are all represented by the variables xb, h, xc, xb-1, and sb in this instance. B stands for every layer in CNN. Note that activities like regression, batch normalization, dropout, and classification take up five to ten percent of the training time. The average of all class averages from the area below the precision-recall curve is known as Mean Average Precision, or mAP. It is used as a stand-in for SSD accuracy. A higher mAP indicates a more accurate model.

PROPOSED WORK

The suggested approach will determine the system architecture, emphasizing deep learning methods for identifying objects and images. We will talk about the latest techniques such as YOLO or Faster R-CNN and how they may be used in conjunction with convolutional neural networks (CNNs), often referred to as recurrent neural networks (RNNs). The approaches to preprocessing, model evaluation, and training data collecting will also be discussed to guarantee a unique methodology.

System Setup: The recommended system's step-by-step implementation process will be described in this section. It will cover the programming languages and libraries, as well as the hardware and software requirements. The processes involved in fine-tuning, data pre-processing, and model training will be fully explained, with an emphasis on any modifications or advancements made to the present methods.

Performance Evaluation: CNNs are becoming commonplace due to a number of images processing applications, including object detection, picture segmentation, and image categorization. Reputable CNN designs like VGGNet, ResNet, and AlexNet provide improvements in terms of training methods, depth, and performance metrics. To solve the vanishing gradient issue, for example, residual learning was added to ResNet, allowing for the training of ever-deeper networks.
Ethical Considerations: We will talk about the moral implications of deep learning algorithms for item and image identification in this section. We'll discuss potential biases, privacy concerns, and the necessity of developing ethical AI. The proposed technology would ensure that any data used is anonymous, legally collected, and consistent with moral standards.
CNN Architecture: CNNs, or convolutional neural networks, are essential to deep learning in image processing. Three layers make up a traditional CNN design: convolutional, pooling, and fully linked. While pooling layers use less processing resources and lower the number of dimensions in the data while keeping important information, convolutional layers use a collection of filters to extract features from the input image. The completely linked layers at the network's end classify the obtained features into multiple item types.
CNN Training: Backpropagation and forward propagation are the two basic methods used to train a CNN. We describe the forward propagation technique, which feeds inputs into the network to create feature maps. The backpropagation gradients are then used to modify the network's parameters. We also go over well-known optimization techniques including stochastic gradient descent (SGD), Adam, and RMSprop.
Pooling Layers: The feature maps produced by convolutional layers have less spatial dimensionality thanks to the pooling layers. Many pooling algorithms, including average and max pooling, are covered, along with their effects on processing complexity reduction and the preservation of significant data.
Fully Connected Layers: By categorizing the recovered attributes, fully connected layers frequently outperform convolutional and pooling layers. We show that feature maps can be flattened and then run through thick layers to aid in result prediction.
Object Detection: Some of the techniques for configuring CNNs for object identification tasks include region-based CNNs (R-CNN), faster R-CNN, and quicker R-CNN. We only provide a synopsis of two techniques: the region categorization procedure and the region suggestion algorithms.
Experimental Setup: In this part, we examine the Pascal VOC, COCO, and ImageNet training and testing datasets. We discuss techniques for improving generalization through both data pretreatment and augmentation. The hyperparameters chosen for CNN training are also described.
Metrics for Evaluation: We go over popular metrics for photo recognition, including F1-score, recall, accuracy, and precision. We offer measures such as mean average precision (mAP) and intersection over union (IoU) for object recognition.

EXPERIMENTAL RESULTS

We report on our work on object and photo identification issues using CNN. By comparing the performances of various CNN architectures, including VGGNet, ResNet, and InceptionNet, We assess each one's benefits and drawbacks. We also showcase the results of several training approaches, such as fine tuning and transfer learning. This section will be describing the key findings of our study as well as the challenges and limitations of employing CNNs for item and image recognition. We suggest several research avenues to tackle these issues and improve CNN's performance.

CONCLUSION

With the use of algorithms like CNNs, R-CNNs, YOLO, and SSD, deep learning has significantly enhanced the field of image processing for object detection while providing reliable and effective results. These techniques have revolutionized a wide range of applications, exhibiting exceptional versatility and performance. In this dynamic industry, it is imperative to address current difficulties and unleash new potentials through ongoing research and innovation.
We will examine in more detail the architecture and techniques employed by deep learning systems to recognize objects and images. We will evaluate the theories and techniques used in these systems, noting their advantages, disadvantages, and potential research areas. Intelligent systems with vision, comprehension, and interaction with the visual world can be constructed by properly comprehending these processes. This creates a plethora of chances for growth and innovation.
Object detection, on the other hand, uses bounding boxes to locate things in photos, going beyond simple object recognition. Object identification performance in deep learning systems has been achieved by combining the advantages of CNN with other components including region proposal networks (RPNs) and anchor-based techniques. The creation of quick and extremely precise object detecting systems has been made possible by these advancements.

REFERENCES

1. Abhinav, A., & Agrawal, A. (2019). A comprehensive survey of deep learning techniques for image recognition. Journal of Pattern Recognition and Artificial Intelligence.
2. Chen, Z., & Gupta, S. (2020). Deep learning for object detection: A comprehensive review. Journal of Visual Communication and Image Representation.
3. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition.

4. Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition.
5. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. Advances in Neural Information Processing Systems.
6. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition.
7. He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. Proceedings of the IEEE international conference on computer vision.
8. Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement.
9. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. Proceedings of the IEEE conference on computer vision and pattern recognition.
10. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition.
11. Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. Proceedings of the IEEE conference on computer vision and pattern recognition.
12. Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, faster, stronger. Proceedings of the IEEE conference on computer vision and pattern recognition.
13. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016).
14. Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement.
15. Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE transactions on pattern analysis and machine intelligence.
16. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition, 2014.
17. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.
18. Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014.
19. Girshick R. Fast r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision. 2015.
20. Chorowski JK, Bahdanau D, Serdyuk D, Cho K, Bengio Y. Attention-based models for speech recognition. In: Advances in Neural Information Processing Systems. 2015.
21. Everingham M et al. The Pascal visual object classes (VOC) challenge. International Journal of Computer Vision. 2010.
22. Peixoto HM, Teles RS, Luiz JVA, Henriques-Alves AM, Santa Cruz RM. Mice Tracking Using the YOLO Algorithm. Vol. 7. PeerJ Preprints; 2019.
23. Henriques-Alves AM, Queiroz CM. Ethological evaluation of the effects of social defeat stress in mice: Beyond the social interaction ratio. Frontiers in Behavioral Neuroscience. 2016.
24. Jhuang H et al. Automated home-cage behavioral phenotyping of mice. Nature Communications. 2010.
25. Burgos-Artizzu XP, Dollár P, Lin D, Anderson DJ, Perona P. Social behavior recognition in continuous video. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE; 2012.
26. Norouzzadeh MS et al. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proceedings of the National Academy of Sciences of the United States of America. 2018.
27. Guo J, He H, He T, Lausen L, Li M, Lin H, et al. GluonCV and GluonNLP: Deep Learning in Computer Vision and Natural Language Processing. arXiv preprint arXiv:1907; 2019.
28. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. ImageNet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. IEEE; 2009.
29. Chen X-L et al. Remote sensing image-based analysis of the relationship between urban heat island and land use/cover changes. Remote Sensing of Environment. 2006.
30. Ren S, He K, Girshick R, Sun J. Faster r-cnn: Towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems. 2015.
31. Chinthamu, N., Gooda, S. K., Venkatachalam, C., Swaminathan, S., & Malathy, G. IoT- based secure data transmission prediction using deep learning model in cloud computing. International Journal on Recent and Innovation Trends in Computing and Communication. 2023.