Real-Time Object Recognition Using Deep-Learning

Herman Khalid Omer [1], Shahad Fauzi Mohammed [2], Rana Adib Khisro [3]

[1,2,3] Department of Computer Science, Nawroz University, Duhok, Kurdistan Region - IRAQ

ABSTRACT

The great interest at the moment is focused on the field of technology, especially artificial intelligence, also do not devoid of our daily life of the use of phone applications and computer programs, that increasing of phone and computer usage demands more programs and applications to satisfy the needs of users. However, this approach starts with acquirement an image of a particular medical or engineering tool, is displayed on the computer through a webcam, whether this image is a photo or digital through a display screen, or even the device itself, the computer will identify the tool and a simplified explanation of the way it works with a video demonstration throughout MATLAB IDE for implementing this project as well as easy to use by anyone even the user doesn't have any experience software.

Finally, this approach has been created this project to save time and effort for the users instead of searching on a specific tool that they need about its name, how to use so we tried to facility this matter. The proposed algorithm got accurate result, for the doctor's tools the accuracy was 95.8 %, for the engineer's tools was 98.3 % and for mix of them was approximately 94.1 %.

KEYWORDS: Deep Learning, Convolutional Neural Network (CNN), Feature Extraction, Classification.

1. Introduction

Extraction of discriminative features from input images is one of the most challenging tasks in object recognition systems. Much effort has aimed at determining optimal feature sets for a specific task [1], based on the attributes of objects to be recognized and classifiers to be used [2][3]. Many of these features produced very promising results. However, due to the ambiguity and lack of general task-independent rules for optimal feature selection [4], the process of data classification has been recently dominated by various approaches using neural networks. The important advantage of these neural network [5] approaches is that during the training process the network self-determines the optimal set of features from the data. The disadvantage is that large training data sets may be required and thus the training process could be very lengthy [6]. Neural networks have been shown to provide excellent performance in multiple image classification benchmarks, ranging from simple feature datasets such as MNIST [7] to complicated challenges

such as ImageNet [8]. However, more recent research in computer vision [9] has demonstrated the dominating power of a neural network methodology known as deep learning. Similar to the design of optimal image descriptors being cumbersome in the past, the design and training of deep learning structures today is also a big challenge [10]. This is particularly a challenge when the performance of the design is very sensitive to the implementation details, which is often the case. Fortunately, using an implementation of already trained structures is quite straightforward and if the power of the network depth can be applied to other classification scenarios [11], then this offers great advantages.

The progress of deep learning approaches in recent years has been spectacular. One of the biggest breakthroughs that catalyzed the recent wave of neural network deep learning [12] research may be tracked to the work of Krizhevsky [13] et al, who proposed ImageNet Classification with Deep Convolutional

Neural Networks (CNNs) [14] as a combination of convolutional and fully connected neural networks applied with data augmentation and training techniques [15]. The AlexNet was the winner of the ImageNet Large Scale Visual Recognition Challenge 2012 [16], and continues to be the source of inspiration for winners in years thereafter. While AlexNet has been outperformed by later proposals, the approach is still valuable as a good compromise between simplicity and performance.

2. Methodology

Deep Learning has been used as a main technology with specifically the Convolutional Neural Network (CNN), to provide a benefit to the user by real time getting acquainted with the medical and engineering tools and displaying the name of the tool as shown in Figure 1.
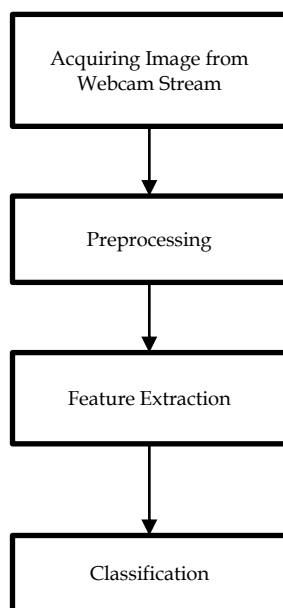
```
┌─────────────────────┐
│ Acquiring Image from │
│   Webcam Stream      │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│    Preprocessing     │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│  Feature Extraction  │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│    Classification    │
└─────────────────────┘
```

Figure 1: Proposed Diagram

Explanation of it in addition to an explanatory video of how it works, as the following:

2.1 Image Datastore

There are many resources of image toolkit that to be used one of them OIDv4_ToolKit [17]. Beside it the Image datastore consisting of two folders (Doctor's and Engineering's Tools), corresponding to six categories: {'tool1', 'tool2', 'tool3', 'tool4', 'tool5', 'tool6'} with 30 images each at least. Then set it each folder categories

separate to the sets into training and test image subsets. In this approach the images are partitioned for training and for testing.

2.2 Feature Extraction

Feature Extraction and Analysis [18] is a goal in image analysis that to extract information useful for solving application-based problems. The first step is by reduced the amount of image data with the tools and it explored. After extracted the features of interest, the important aspects of feature analysis are to determine exactly which features are important, so the analysis is not complete until it incorporate application-specific feedback into the system [19].

The bag Of Features object defines the features [20], or visual words, by using the k-means clustering algorithm on the feature descriptors extracted it. The algorithm iteratively groups the descriptors into k mutually exclusive clusters. The resulting clusters are compact and separated by similar characteristics. Each cluster center represents a feature, or visual word.

Extract features based on define a grid to extract feature descriptors, as shown in Figure 2. The grid method may lose fine-grained scale information. Therefore, use the grid for images that do not contain distinct features, such as an image containing scenery, like the shapes. Using speeded up robust feature detector provides greater scale invariance.
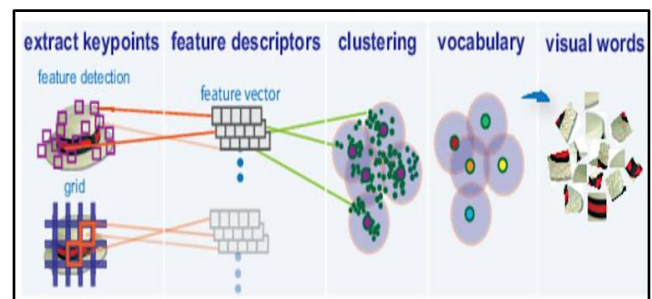
Figure 2: Steps for Extract Features

This algorithm workflow analyzes images in their entirety. Images must have appropriate labels describing the class that they represent. For example, a set of stethoscope images could be labeled stethoscopes. The workflow does not rely on spatial

information nor on marking the particular objects in an image. The bag-of-visual-words technique relies on detection without localization.

2.3 The Bag of Visual Words

The Train Image Category Classifier function returns an image classifier. The function trains a multiclass classifier using the error-correcting output codes (ECOC) framework [21]. The Train Image Category Classifier function uses the bag of visual words returned by the bag of Features object to encode images in the image set into the histogram of visual words as shown in Figure 3. The histogram of visual words is then used as the positive and negative samples to train the classifier, and then it would encode each image from the training set [22].
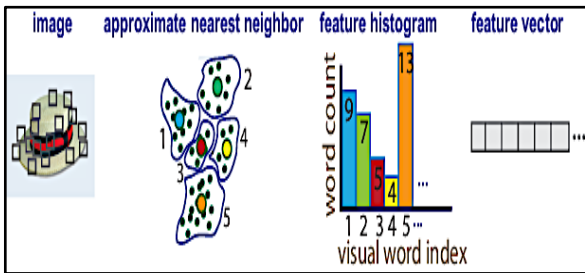

Figure 3: Extract Features Vector

This function detects and extracts features from the image and then the approximate nearest neighbor algorithm uses to construct a feature histogram for each image as shown in Figure 4. The function then increments histogram bins based on the proximity of the descriptor to a particular cluster center [23]. The histogram length corresponds to the number of visual words that the bag of Features object constructed. The histogram becomes a feature vector for the image [24]. Repeat for each image in the training set to create the training data.
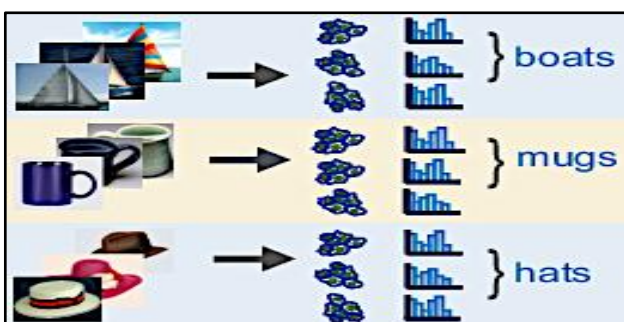

Figure 4: Histogram Features Extraction

To evaluate the quality of the classifier, the image Category Classifier evaluate method that used as shown in Figure 5, to test the classifier against the validation image set. The output confusion matrix represents the analysis of the prediction. A perfect classification results in a normalized matrix containing 1s on the diagonal. An incorrect classification results fractional values.
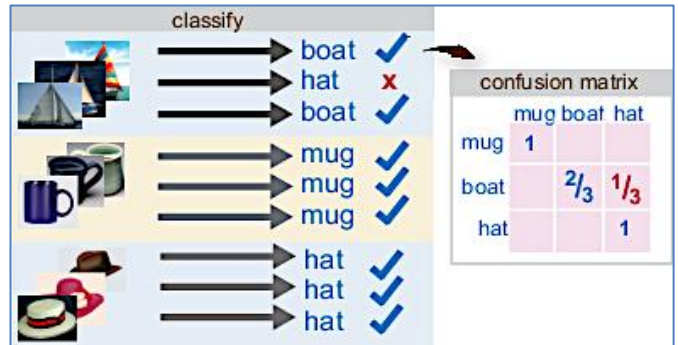

Figure 5: Classify Feature to Confusion Matrix

2.4 Real-Time Organization

Definition of real time [25]: the actual time during which something takes place the computer may partly analyze the data in real time (as it comes in). Or in another word: response that appears to take place instantaneously or in the same timeframe as its real-world counterpart action. There are many things that can affect the Real-Time testing and sometimes these can be defined as the Real-Time problems as follows [26]:

i. The environmental conditions surrounding the Real-Time test have a significant impact on the Real-Time results such as lighting and others.

ii. The quality of the camera used in the Real-Time test.

iii. If the tool or image is close or far away from the camera, this can be effect on the Real-Time test.

iv. The trembling of the hand holding the instrument or the image during the Real-Time has a big effect, so the computer will get unclear and blurry image.

v. Containing the background colors similar to the colors of the tool can lead to confuse on the computer and thus not get accurate results.

vi. The thickness of the paper if the test is done to a picture of a tool printed on a paper, the clearer the image and the high quality al that can lead to more accurate results.

2.5 Object Classification and Recognition

Computer vision technique for identifying objects in images or videos [27]. Also, it is a key output of deep learning and machine learning algorithms as shown in Figure 6.
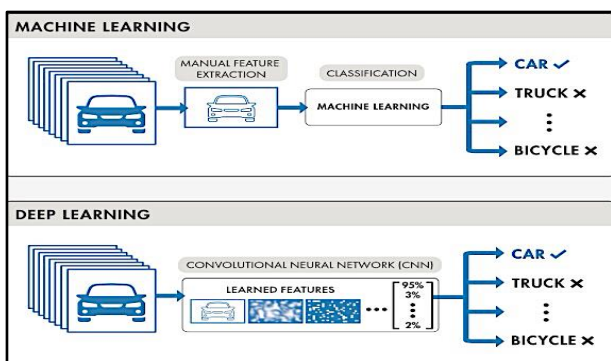

Figure 6: Machine and Deep Learning Techniques

To classified image is an important part of the remote sensing, image analysis and pattern recognition. In some instances, the classification itself may be the object of the analysis.The image classification therefore forms an important tool for examination of the digital images. The term classifier refers loosely to a computer program that implements a specific procedure for image classification [27]. The analyst must select a classification method that will best accomplish a specific task. For present, it is not possible to state which classifier is best for all situation as the characteristic of each image and the circumstances for each study vary so greatly [27]. Therefore, it is essential that each analyst understand the alternative strategies for image classification [27].

These techniques are distinguished in two main ways as:

A. Supervised classification

- Analyst identifies representative training sets for each informational class
- Algorithm generates decision boundaries

B. Unsupervised classification

- Algorithm identifies clusters in data
- Analyst labels clusters

Additionally, supervised classification has different sub classification methods which are named as parallel piped, maximum likelihood, minimum distances and Fisher classifier methods. These methods are named as Hard Classifier [28]. As shown in Figure 7.
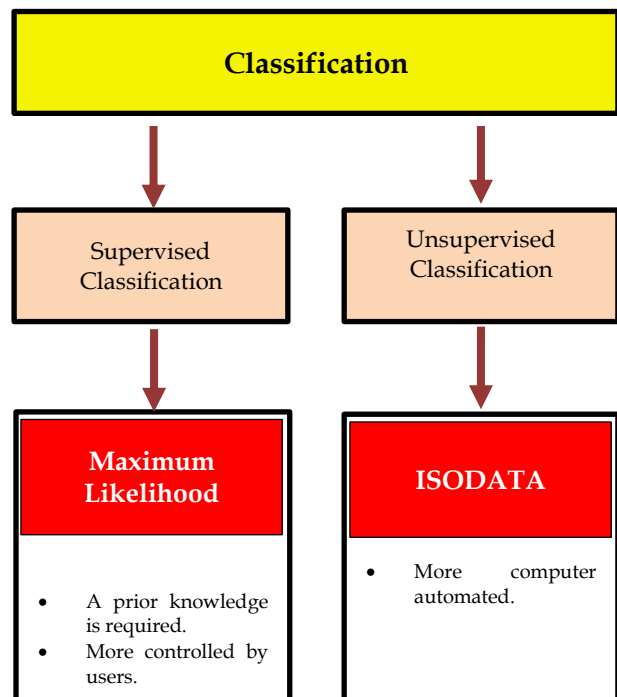

Figure 7: Classification Techniques Types.

2.5 Deep Learning and CNN

In deep learning, a convolutional neural network (CNN, or ConvNet) is a class of deep neural networks, most commonly applied to analyzing visual imagery [30]. CNNs are regularized versions of multilayer perceptron's. Multilayer perceptron's usually refer to fully connected networks, that is, each neuron in one layer is connected to all neurons in the next layer. They are also known as shift invariant or space invariant artificial neural networks (SIANN) [31], based on their shared-weights architecture and translation invariance characteristics. Convolutional networks were inspired

by biological processes. ResNet-50 is a convolutional neural network that is trained on more than a million images from the ImageNet database. The network is 50 layers deep and can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals. As a result, the network has learned rich feature representations for a wide range of images. The network has an image input size of 224-by-224. as a shown in Figure 8.
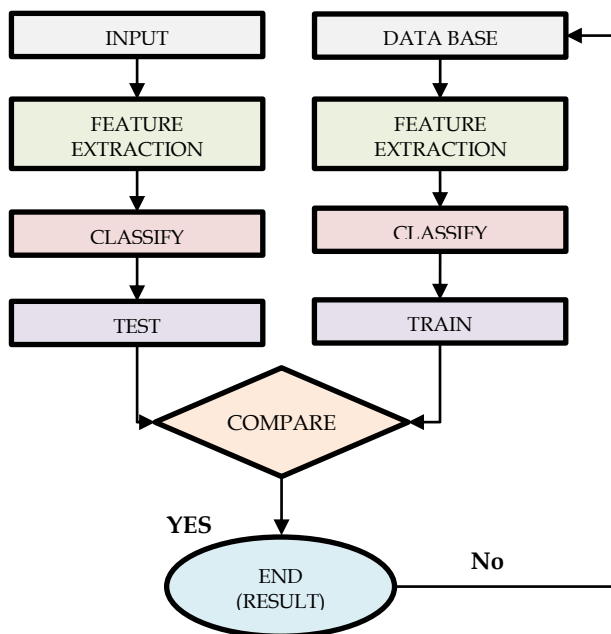

Figure 8: Training CNN Function

2.7 Result and Discussion

As a beginning of program, after all the project will check folder of training database which is contain image for each category separately and mix, because we need more one than case image for each object that will training in CNN function with full layer, beside its normalization feature and minimum batch size to 32 in each weight layers using activation functions.

Firstly, the result shows in separate categories both of them (Medicine & Engineering Tools), The Medicine Tools accuracy have been done under 6 categories compared with other classifications in same number of input image for training and testing using EscVM dataset [17], the highest is better as shown in Table 1

Table 1: The Medical Tools Classification

#	No. of Tools	The Used Algorithm	Accuracy %
1	6	SubSpace KNN with learning Type Discriminate	79%
2	6	Bag	79%
3	6	Bagged Trees	77.1%
4	6	SubSpace KNN with learning Type Nearest Neightbor	79.6%
5	6	Fine KNN	81%
6	6	Linear SVM	85.8%
7	6	Quadratic SVM	86.4%
8	6	Cubic SVM	87.9%
9	6	Proposed	95.8%

Secondly, the result shows in Engineering categories in same previous classification, The Engineering Tools accuracy have been done under 6 categories compared with other classifications in same number of input image for training and testing, the highest is better as shown in Table 2.

Table 2: The Engineering Tools Classification

#	No. of Tools	The Used Algorithm	Accuracy %
1	6	SubSpace KNN with learning Type Discriminate	81%
2	6	Bag	81.7%
3	6	Bagged Trees	83.1%
4	6	SubSpace KNN with learning Type Nearest Neightbor	83.4%
5	6	Fine KNN	85%
6	6	Linear SVM	85.5%
7	6	Quadratic SVM	86.2%
8	6	Cubic SVM	95%
9	6	Proposed	98.3%

Finally, the result shows both categories with same before classification, The Mixed Tools accuracy have been done under 6 categories compared with other classifications in same number of input image for training and testing, the highest is better as shown in Table 3.

Table 3: The Mixed Tools Classification

#	No. of Tools	The Used Algorithm	Accuracy %
1	12	SubSpace KNN with learning Type Discriminate	72.1%
2	12	Bag	71.7%
3	12	Bagged Trees	71.7%
4	12	SubSpace KNN with learning Type Nearest Neightbor	73.4%
5	12	Fine KNN	74%
6	12	Linear SVM	75.5%
7	12	Quadratic SVM	84%
8	12	Cubic SVM	84.3%
9	12	Proposed	94.1 %

3. Conclusion

The project based on Deep learning and CNN, it can use a static and dynamic data from the Real-Time which is the current time of execution it can recognize objects by displaying it either the object itself, a printed photo on a paper or a phone picture recognize the tool, displaying its name, a description about it, also it displays a tutorial video about how to use. It is so easy to use by any one, although it is complex academically.

4. Recommendation

It is possible to improve this research in the future by adding some other features, work in other field like some kind of cancer. Also, improve area code to joining mobile application. Beside that using Nvidia GPU array to increase accuracy and decrease time recognize for real time, but that need high performance Video Card.

5. References

1. Singh, U., & Singh, S. N. (2019). A new optimal feature selection scheme for classification of power quality disturbances based on ant colony framework. Applied Soft Computing, 74, 216-225.
2. Kwon, H., Kim, Y., Park, K. W., Yoon, H., & Choi, D. (2018). Friend-safe evasion attack: An adversarial example that is correctly recognized by a friendly classifier. computers & security, 78, 380-397.
3. Mehre, S. A., Dhara, A. K., Garg, M., Kalra, N., Khandelwal, N., & Mukhopadhyay, S. (2019). Content-based image retrieval system for pulmonary nodules using optimal feature sets and class membership-based retrieval. Journal of digital imaging, 32(3), 362-385.
4. Çetin, A., & Gökhan, T. (2018). Differential Diagnosis of Erythematous Squamous Diseases with Feature Selection and Classification Algorithms. In Nature-Inspired Intelligent Techniques for Solving Biomedical Engineering Problems (pp. 103-129). IGI Global.
5. Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017, August). Understanding of a convolutional neural network. In 2017 International Conference on Engineering and Technology (ICET) (pp. 1-6). Ieee.
6. Liu, D., Tan, Y., Khoram, E., & Yu, Z. (2018). Training deep neural networks for the inverse design of nanophotonic structures. Acs Photonics, 5(4), 1365-1369.
7. Cohen, G., Afshar, S., Tapson, J., & Van Schaik, A. (2017, May). EMNIST: Extending MNIST to handwritten letters. In 2017 International Joint Conference on Neural Networks (IJCNN) (pp. 2921-2926). IEEE.
8. Kornblith, S., Shlens, J., & Le, Q. V. (2019). Do better imagenet models transfer better. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2661-2671).
9. Nixon, M., & Aguado, A. (2019). Feature extraction and image processing for computer vision. Academic press.
10. Liu, D., Tan, Y., Khoram, E., & Yu, Z. (2018). Training deep neural networks for the inverse design of nanophotonic structures. Acs Photonics, 5(4), 1365-1369.
11. Gallo, I., Calefati, A., & Nawaz, S. (2017, November). Multimodal classification fusion in real-world scenarios. In 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR) (Vol. 5, pp. 36-41). IEEE.
12. Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., & Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. Neurocomputing, 234, 11-26.
13. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84-90.
14. Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., ... & Chen, T. (2018). Recent advances in convolutional neural networks. Pattern Recognition, 77, 354-377.
15. Wigington, C., Stewart, S., Davis, B., Barrett, B., Price, B., & Cohen, S. (2017, November). Data augmentation for recognition of handwritten words and lines using a CNN-LSTM network. In 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR) (Vol. 1, pp. 639-645). IEEE.
16. Desoli, G., Chawla, N., Boesch, T., Singh, S. P., Guidetti, E., De Ambroggi, F., ... & Aggarwal, N. (2017, February). 14.1 A 2.9 TOPS/W deep convolutional neural network SoC in FD-SOI 28nm for intelligent embedded systems. In 2017 IEEE International Solid-State Circuits Conference (ISSCC) (pp. 238-239). IEEE.
17. https://github.com/EscVM/OIDv4_ToolKit.git
18. Jaiswal, A. K., & Banka, H. (2017). Local pattern transformation based feature extraction techniques for

classification of epileptic EEG signals. Biomedical Signal Processing and Control, 34, 81-92.

19. Chaki, J., & Dey, N. (2019). A beginner's guide to image shape feature extraction techniques. CRC Press.

20. Passalis, N., & Tefas, A. (2017). Neural bag-of-features learning. Pattern Recognition, 64, 277-294.

21. Sun, M., Liu, K., Wu, Q., Hong, Q., Wang, B., & Zhang, H. (2019). A novel ECOC algorithm for multiclass microarray data classification based on data complexity analysis. Pattern Recognition, 90, 346-362.

22. Cao, Y., Wang, C., Li, Z., Zhang, L., & Zhang, L. (2010, June). Spatial-bag-of-features. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (pp. 3352-3359). IEEE.

23. Rai, H. G., Shen, X., Deepak, K. S., & Krishna, P. R. (2011, November). Hybrid feature to encode shape and texture for content based image retrieval. In 2011 International Conference on Image Information Processing (pp. 1-6). IEEE.

24. Kobayashi, T., Hidaka, A., & Kurita, T. (2007, November). Selection of histograms of oriented gradients features for pedestrian detection. In International conference on neural information processing (pp. 598-607). Springer, Berlin, Heidelberg.

25. Radovic, M., Adarkwa, O., & Wang, Q. (2017). Object recognition in aerial images using convolutional neural networks. Journal of Imaging, 3(2), 21.

26. Liu, H., Wu, Y., Sun, F., Fang, B., & Guo, D. (2017). Weakly paired multimodal fusion for object recognition. IEEE Transactions on Automation Science and Engineering, 15(2), 784-795.

27. Weinstein, B. G. (2018). A computer vision for animal ecology. Journal of Animal Ecology, 87(3), 533-545.

28. Li, J., Bioucas-Dias, J. M., & Plaza, A. (2012). Semisupervised hyperspectral image classification using soft sparse multinomial logistic regression. IEEE Geoscience and Remote Sensing Letters, 10(2), 318-322.

29. Milletari, F., Ahmadi, S. A., Kroll, C., Plate, A., Rozanski, V., Maiostre, J., ... & Navab, N. (2017). Hough-CNN: deep learning for segmentation of deep brain regions in MRI and ultrasound. Computer Vision and Image Understanding, 164, 92-102.

30. Dong, L., Yan, J., Yuan, X., He, H., & Sun, C. (2018). Functional nonlinear model predictive control based on adaptive dynamic programming. IEEE transactions on cybernetics, 49(12), 4206-4218.

31. Ashqar, B. A., & Abu-Naser, S. S. (2018). Image-based tomato leaves diseases detection using deep learning.