Research on Ship Detection and Classification Using Deep Learning Approach

Rana Muhammad Usman*, Junhua Yan, Imran Qureshi

College of Astronautics

Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

***Abstract*:** Ship detection is an integral part of jobs, including fisheries management, ship search, and maritime traffic monitoring and control, and it aids in the prevention of unlawful actions. We used the Ship Detection Method to keep track of the ships' movements. The goal of this study is to use deep learning to detect and classify ships. Image Enhancement, Ship Detection, and Ship Classification are the three aspects of this project. We used Filters to increase the quality of our photographs and applied the image enhancement algorithm to our dataset. After that, we began the process of detecting and classifying ships. The image improvement was based on the fact that prior experiments had yielded unsatisfactory results; therefore, we improved our dataset. Various CNN Networks, such as VGG-16, VGG-19, ResNet 50, and Inception ResNet v2, are used to classify ships.

***Keywords*:** CNN, Deep Learning, Detection, Classification

I. INTRODUCTION

AHE (adaptive histogram equalization) is a computer image processing approach for enhancing image contrast. The adaptive system differs from traditional histogram equalization. It computes multiple histograms, each corresponding to a different portion of the image, and uses them to disperse its brightness values. As a result, it is appropriate for strengthening local contrast and edge definition in each section of an image.

On the other hand, AHEtends to exaggerate noise in relatively homogeneous areas of an image. By restricting the amplification, a variation of adaptive histogram equalization known as contrast limited adaptive histogram equalization (CLAHE) prevents this.

The Automatic Identification System (AIS) is good at detecting ships; however, it does not identify all ships since it requires every ship to have a VHF transponder, which cannot be recognized if the transponder is defective or not there. As a result, we adopted a different technology, which is remote Sensing. Radio waves are used to picture the Earth's surface via Synthetic Aperture Radar (SAR) images. Unlike optical images, the wavelengths used by the devices are unaffected by a time of day or weather conditions, allowing imagery to be collected at any time of day or night, with cloudy or clear sky. These photos are being collected by satellites and could be used to develop ship detection and segmentation algorithms. As a result, we employed an algorithm to determine if a remotely sensed target was a ship or not.

The use of SAR to automatically classify ships has been extensively researched, with the most recent review was appearing in [1]-[3]. Compared to the large number of feasibility analyses for ship classification based on SAR, optical imaging is used in significantly fewer research papers [2].

We offer a new approach for automatically classifying ships and small Unidentified Floating Objects (UFOs) using optical aerial data obtained in the visible spectrum based on Convolutional Neural Networks (CNN). In image classification tasks, CNNs have achieved state-of-the-art results [4], analogous to the challenge addressed in this paper. We propose a CNN architecture for ship classification from aerial images.

II. IMAGE ENHANCEMENT

Histogram equalization is a technique for improving overall contrast. This procedure involves adjusting the intensity of the image's worldwide distribution. If we consider any greyscale image (x), we can deduce that n i is the number of occurrences of grey level I and that a probability function for the event of a

pixel of class I in an image (x) is

$$p_x(i) = p(x = i) = \frac{n_i}{n}, 0 \le i < L$$

After the background and foreground images are fused into a new frame, the histogram equalization approach is utilized in the article [5] to adjust intensity values independently for background and foreground images. For equalization, the 'imhist' function is used.

Fig.1(a) Original noised image

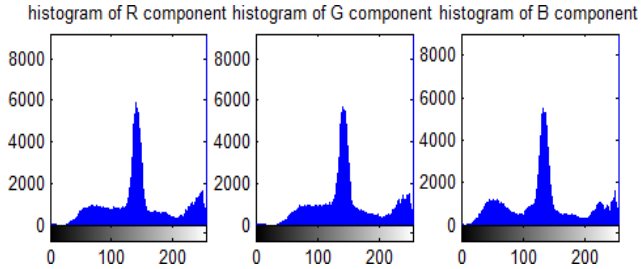

Fig.1. (b) Original image divided into RGB layers and generate a histogram of RGB layer separately

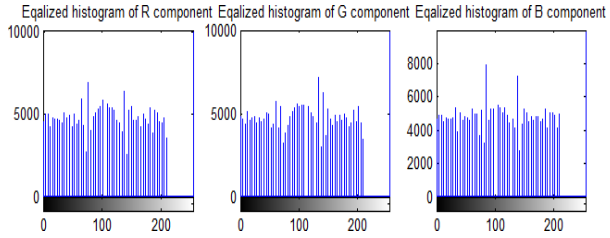

Fig.1.(c)Equalized histogram for RGB layer

Fig.1(d) after applying the histogram equalization method, restored image

A. Image Enhancement Techniques

This paper discusses image processing and its core steps before delving into the many images enhancing techniques. Picture enhancement has been identified as one of the essential vision applications due to its capacity to improve image visibility. So far, different approaches for increasing the quality of digital photographs have been offered. One of the most critical challenges with high-quality pictures, such as those taken with digital cameras, is image augmentation. Because lighting, weather, and the equipment used to capture the image can all impact image clarity.

1. Adaptive Histogram Equalization

A modified version of the histogram equalization approach is the adaptive histogram equalization method. This method involves applying enhancements to a specific image and adjusting contrast based on neighbouring pixels. The Yoon employed AHE for HSV colour space improvement [5]. Matlab's 'adapthisteq' function is used to equalize intensity based on nearby pixels. This method is only used to reduce noise. Hwa- Hyun Cho, Gyu-Hee Park For enhancement, use the Dynamic range separation Histogram equalization (DRSHE) approach. The histogram is partitioned into preset segments in this manner. Then modify the intensity of that portion and distribute it evenly across the grayscale image.

2. Contrast limited adaptive his to gram equalization (CLAHE)

Adaptive histogram equalization with contrast limitations is a modified version of adaptive histogram equalization. This approach applies the enhancement function to all neighbouring pixels, and the transformation function is derived. Because of its contrast limiting, this differs from AHE. Zhiyuan Xu, Xiaoming Liu, and Xiaonan Chen employed the CLAHE approach for enhancement in their study [6], which used the maximum value to trim the histogram and redistribute the grey level image. This paper [6] uses a distinct method for the backdrop and the ground to reduce noise and improve contrast. Distribution parameters are utilized for the shape of the histogram equalization graph; in paper [6], the 'Rayleigh' distribution parameter is used for the bell-shaped histogram. CLAHE was applied to both grayscale and colorful images. The 'cliplimit' function is used to impose a limit to a picture of noise. For RGB accurate colour photos, the LAB colour space is employed.

CLAHE Algorithm

Step 1: Get a noisy image.

Step 2: Collect all input variables needed in the enhancement process, such as the number of regions in each row and column, dynamic range (Number of bins utilized in the histogram transform function), cliplimit, and distribution parameter type.

Step 3: Divide the original image into an area and preprocess these inputs.

Step 4: The process is applied to the tile (contextual region).

Step 5: Create a clipped histogram and grey level mapping. Because the number of pixels in each grey level in the contextual region is evenly distributed, the Average Number of pixels in each grey class is described as follows:

$$N_{avg} = \frac{N_{CR-xp} * N_{CR-yp}}{N_{GRAY}}$$

Where

N_{avg} = Number of pixels on average

N_{GRAY} =The Number of grey levels in the context

N_{CR-xp} =Contextual pixel count in the X direction

$N_{CR} - yp$ =number of pixels in the Y direction of contextual region

After calculate the actual clip limit

$$N_{CL} = N_{cLIP} * N_{avg}$$

Step 6: To make a better image, interpolate grey level mapping. Using four-pixel clusters and applying to map, each of the mapping tiles will partially overlap in the image region, after which a single pixel will be taken and four mappingsused to it. Repeat over an image to get improve pixel by interpolating between those outcomes.

Fig2. (a) Original foggy (noised) image

Fig.2.(b)Frame after applyingCLAHE.

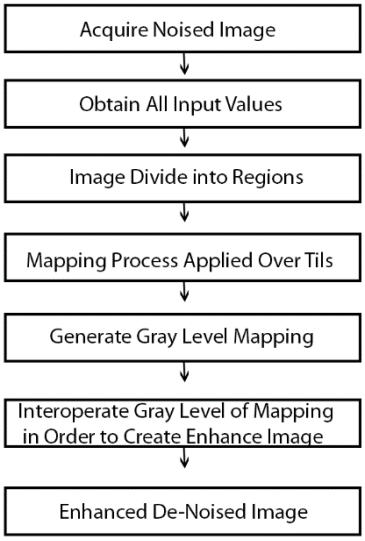

Fig.3 Flow Chart For CLAHE

B. Image Enhancement Proposed Algorithm Steps

Step1: Read a single frame (image).

Step2: Adjust the intensity of that image.

Step3: Convert RGB true-colour image in to Gray level image.

Step4: Adjust the intensity of gray image.

Step5: Apply CLAHE on that grey image with a cliplimit of 0.2 and a Rayleigh distribution parameter to get a bell-shaped histogram.

Step6: Restore the enhanced image in any structure.

Step7: Calculate peak signal to noise ratio.

III. SHIP DETECTION

The dataset of ships from the MASERATI dataset (v2) - MAritime SATellite Imagery dataset was used to detect ships. The dataset is

divided into four categories, which are as follows: Coast Ship (a) Detail (b) Multi (c) Ship We describe a fast vessel detector algorithm based on object recognition methods developed in the computer vision field International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences, Volume XL-1/W1, ISPRS Hannover Workshop 2013, 21 – 24 May 2013, Hannover, Germany 233. A single channel, 8 bit, geo referenced image is used as the detection input. The detector is built on an object detection framework [Ref] developed initially for face detection [7].

The detector's main component is a binary classifier that identifies vessel image windows from non-vessel image windows. Offline training is used to train this classifier. A window with varying orientation and size is slid along the image and identified as vessel or background during the online detection. To avoid multiple detections, the classified windows are clustered together. The surrounding surroundings around the ship are included in the picture windows categorized as vessels. The exact size of the ship is determined by using image segmentation to retrieve the vessel's features. The process diagram is shown in Figure 4.

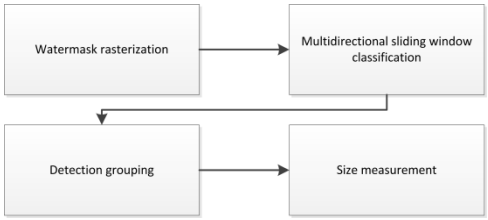

Fig.4. The Detection Process

A. The Classifier

Features The Haar-like features are expanded with slanted Haar-like features in the classifier[15]. Haar-like features are simple features based on the sum of intensity values in two picture areas. Figure 5 depicts the Haar-like features employed, with the feature value equaling the difference between the black and white image areas. Because the number of possible Haar-like features exceeds the number of pixels in the image, the information in the picture is overrepresented. Still, the critical features are selected during training. The first two Haar-like characteristics identified by the movement are shown in Figure 5. The features can be estimated quickly using integral pictures, with just a few memory visits and arithmetic operations, and the number of operations is independent of the size of the feature. As a result, feature extraction has a minimal computational cost.

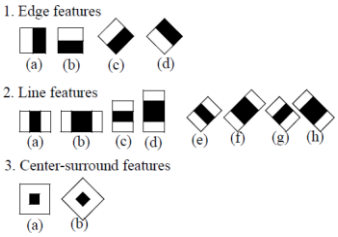

Fig.5. Use haar-like features

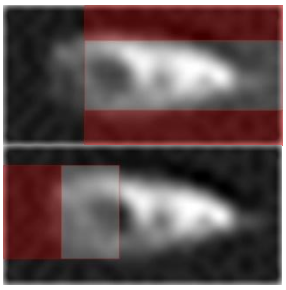

Fig.6. The first haar-like feature of the classifier

B. Training

WorldView-1 and WorldView-2 photos with a resolution of 50 cm were used to train the detector. Figure 7 depicts positive samples. After being manually labelled on satellite photos, the vessel samples are cropped, rotated to the x-axis direction, and shrunk to 1224 pixels. A total of 800 positive samples were used to train the detector. Negative samples were randomly chosen from manually designated regions without vessels for each step and examined by the preceding stages until 800 false positives were discovered; these samples were then used as negative samples. Each stage in the final cascade has a detection rate of di = 0.995 and a false positive rate of fi = 0.5.

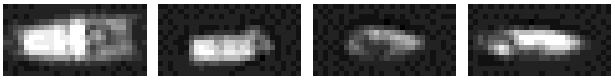

Fig.7. Positive Training Samples

C. Detection

Mask made of water The SRTM water mask is rasterized into the geo referenced satellite picture as a binary mask to avoid processing of land surfaces. The more extensive land regions are skipped at the sliding window classification, and the remaining detections over land areas are discarded during the grouping stage. Sliding glass door The classifier is direction variant, although the images' vessels can lay in any direction. As a result, detection must be done in multiple orders. Rotating the image or rotating the classifier are two options I've examined. Turning the detector's Haar-like characteristics can cause mistakes, as was discovered in[8].The vessel classifier's features are pretty small (i.e. 2 24 pixels). The rasterization of the pixels is not negligible at this scale; thus, the Haar-like feature values can alter as a result of rotation. Fast calculation of Haar-like features using integral pictures is only achievable in specific directions, such as the axis-aligned directions for the basic Haar-like features [7] or the 45-degree guide for the advanced Haar-like features [7].Due to these challenges, I rotate the image rather than the detector.

IV. SHIP CLASSIFICATION

Previously, image processing and computer vision techniques extracted significant characteristics from visible spectrum images, then fed them into typically supervised classifiers. We offer a method for identifying whether a visible spectrum aerial image contains a ship or not. The suggested architecture is based on Convolutional Neural Networks (CNN), and it improves performance by combining neural codes produced from a CNN with the k-Nearest Neighbor approach. The findings of the kNN are compared to those of the CNN Softmax output. Several CNN models were setup and analyzed to find the appropriate hyperparameters, and the best setting for this task was discovered via transfer learning at various levels. We built a new dataset (dubbed MASATI) consisting of aerial images with over 6000 samples to train and assess our architecture. Our method outperforms existing methods based on CNNs, with a success rate of over 99 percent, compared to 79 percent for standard practices in the categorization of ship photos. The proposed approach was further evaluated using an image dataset (MWPU VHR-10) that has already been used in earlier studies. Our best setup has an 86 percent success rate with these data, much exceeding previous state-of-the-art ship classification approaches.

A. Network Topologies

We analyzed six CNN topologies that are typical of the state-of-the-art in image recognition in this paper. Unlike [9], where models like AlexNet [10], VGG-F/M/S, or VGG-16/19 [11] were employed for scene categorization, we chose two classics (VGG-16/19) and three current network models that outperform the classic networks in object identification tasks: Res Net [12], Inception V3 [13], and Xception [14]. Google net has been improved using Inception V3 and Xception. Remote Sensing, vol. 10, no. 511, p. 20. In addition, we created and tested a simple network topology to provide a baseline for comparison with the other five models outlined below. The following topologies were tested: Baseline Network. Only two convolutional layers are present in this network, followed by Max Pooling [30] and Dropout filters [16], as well as two fully connected layers at the end. ReLU is used for all activation functions. VGG-16 and VGG-19 are two VGG genes [11]. VGG-16 has thirteen convolutional layers and three fully connected layers, whereas VGG-19 has sixteen convolutional layers and three fully connected layers. Dropout and Max-pooling strategies, as well as ReLU activation functions, are used in both topologies. The third installment of the Inception franchise [13]. Six convolutional layers are followed by three Inception modules and a final fully linked layer in this architecture. The Inception modules, whose design is based on two key ideas: the approximation of a sparse structure with spatially repeated dense components, and the use of dimensionality reduction to keep the computational complexity in bounds, have fewer parameters than other similar models. ResNet is a network of real estate professionals. Rather than learning unreferenced functions, the deep REsidual learning Network learns residual procedures concerning the layer inputs. This method allows for a high number of layers to be used. Our experiments were conducted using the 50-layer version. [14] Xception. The depth wise separable convolution operation is enabled by a revised version of Inception modules in this model, containing 36 convolutional layers. Using the same number of parameters, this architecture outperforms

Inception's findings.

B. Proposed Architecture

This integrated technology is used in the suggested design to boost performance. We start the network with pre-trained weights from the ILSVRC dataset (a 1000-class subset of ImageNet, a general-purpose database for object classification). We then fine-tun them with samples from our dataset (presented in the following section). The key benefits of this method are that it allows the network to be trained with less data and converges faster.

We also take '2 into account while normalizing the neural codes. In transfer learning, this is a popular approach. Let x be an m-dimensional vector representing neural codes. The '2 normalizations are defined as follows:

$$|x| = \sqrt{\sum_{i=1}^{m} |x_i|^2}$$

The CNN module of the architecture presented in Fig. 8 employs many network topologies. In addition to comparing the standard Softmax output (denoted as NC + Softmax in the experiments) with the suggested hybrid technique utilizing kNN with '2 normalizations (denoted as NC + '2 + CNN), the goal is to compare these topologies to achieve the best model empirically.

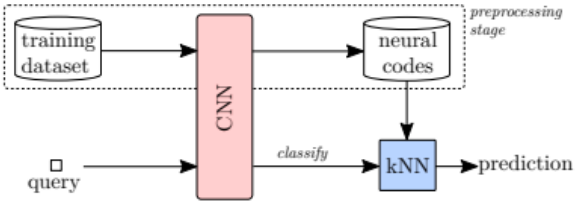

Fig. 8. Proposed CNN Network

V. RESULTS

A. Image Enhancement

After applying the filters, the results were quite impressive as the quality of the images improved, which helped to detect and classify the ships with better accuracy.

Fig. 9(a) Before Image Enhancement

Fig. 9(b) After Image Enhancement

B. Ship Detection

We went on to define routines for displaying sample images with a ship mask and an encoded and decoded mask. Ground truth images are used here in the form of masks. The Mask-R-CNN model was then loaded from my forked repository. After that, we constructed a class with three functions for importing datasets, masks, and reference pictures. After that, we created a class with three positions for importing datasets, shows, and reference pictures.We also designed a course for similar work that uses the GPU. Working with a GPU is required in this situation due to the large dataset size. In addition, Google Colab and Kaggle kernels are both free cloud-based GPU providers at the moment. We kept working on the scripts for loading the training dataset. We then put the model through its paces. For speedier results, we just trained for two epochs. It would take a lot more than that to achieve adequate convergence and decent effects. Here you are free to experiment with the hyper-parameters. The model failed in some instances when it mistook a small island or coastal pebbles for a ship.

To put it another way, the model is producing false positives. Other times, as demonstrated in the figure below, the model failed to detect a ship. To put it another way, the model is producing false negatives.

Fig.10(a). Ships Before Detection

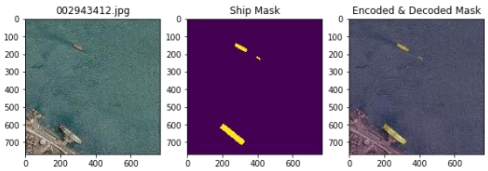

Fig.10.(b) Ships After Detection

C. Ship Classification

The classification of ships is done by using various CNN Networks like VGG-16, VGG-19, ResNet50, Inception ResNet v2. After training these networks, we got the desired results as we expected.

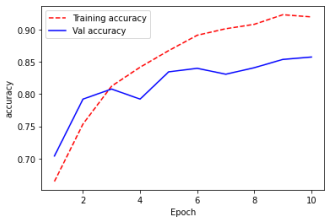

Fig. 11. Results of Training and Validation Accuracy of VGG-16

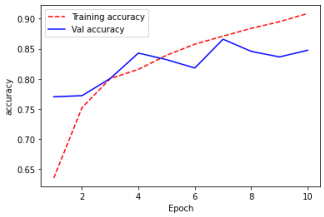

Fig. 12. Results of Training and Validation Accuracy of VGG-19

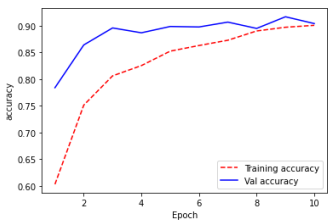

Fig. 13. Results of Training and Validation Accuracy of ResNet50

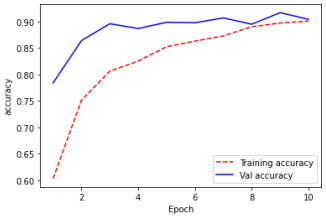

Fig. 14. Results of Training and Validation Accuracy of Inception ResNet v2

Model	VGG-16	VGG-19	ResNet 50	Inception ResNet v2
Validation Loss	68.97 %	52.03 %	33.48 %	20.52%
Validation Accuracy	85.75 %	84.74 %	90.43 %	93.44%
Validation F1 Score	77.99 %	75.40 %	86.27 %	89.87%

Table.1. Results Comparison Of Different Models

VI. CONCLUSION

As the aspect of the image enhancement, our objective was to improve the quality of the images by applying different filters. We implemented the contrast histogram equalization method to enhance the images' quality and got the desired results. Afterwards, we had to detect the ships through a deep CNN approach, and we applied ship detection masks to see the vessel located in the frame, and we had used four classes of boats, and we detected them all, including small vessels and the ships as well. The last task was to classify the ship using a deep learning approach. We applied four networks for the classification purpose. Hence it is concluded that Inception ResNet v2 has minimum loss value, maximum validation accuracy and full validation F1 score. It is proved that this network is the best among those networks.

REFERENCES

[1] Crisp, D. The State-of-the-Art in Ship Detection in Synthetic Aperture Radar Imagery. Australian Government, Department of Defense: Canberra, Australia, 2004; p. 115.

[2] Greidanus, H.; Kourti, N. Findings of the DECLIMS project—Detection and Classification of Marine Traffic from Space.

In Proceedings of the SEASAR 2006: Advances in SAR Oceanography from ENVISAT and ERS Missions, Frascati, Italy, 23–26 January 2006.

[3] Marino, A.; Sanjuan-Ferrer, M.J.; Hajnsek, I.; Ouchi, K. Ship Detection with Spectral Analysis of Synthetic Aperture Radar: A Comparison of New and Well-Known Algorithms. Remote Sens. 2015, 7, 5416, doi:10.3390/ rs70505416.

[4] LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. Nature 2015, 521, 436–444, doi:10.1038/nature14539.

[5] Inhye Yoon; Seonyung Kim; Donggyun Kim; Hayes, M.H.; Joonki Paik, "Adaptive defogging with color correction in the HSV color space for consumer surveillance system," Consumer Electronics, IEEE Transactions on , vol.58, no.1, pp.111,116, February 2012.

[6] Gyu-Hee Park; Hwa-Hyun Cho; Myung-Ryul Choi, "A contrast enhancement method using dynamic range separate histogram equalization," Consumer Electronics, IEEE Transactions on , vol.54, no.4, pp.1981,1987, November 2008.

[7] Viola, P. and Jones, M., 2004.Robust real-time face detection. International Journal of Computer Vision57,pp.137–154.

[8] Lienhart, R. and Maydt, J., 2002. An extended set of haar-like features for rapid object detection. In: Image Processing. 2002. Proceedings. 2002 International Conference on, Vol. 1, pp. I– 900–I–903 vol.1.

[9] Hu, F.; Xia, G.S.; Hu, J.; Zhang, L. Transferring Deep Convolutional Neural Networks for the Scene Classification of High-Resolution Remote Sensing Imagery. Remote Sens. 2015, 7, 14680–14707, doi:10.3390/rs71114680

[10] Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the The Twenty-sixth Annual Conference on Neural Information Processing Systems (NIPS), Stateline, NV, USA, 3–8 December 2012

[11] Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv 2014, arXiv:1409.1556.

[12] He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.

[13] Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. arXiv 2015, arXiv:abs/1512.00567.

[14] Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. arXiv 2016, arXiv:abs/1610.02357

[15] Lienhart, R. and Maydt, J., 2002.An extended set of haar-likefeatures for rapid object detection.In: Image Processing. 2002.Proceedings. 2002 International Conference on, Vol. 1, pp. I–900–I–903vol.1.

[16] Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. J. Mach. Learn. Res. 2014, 15, 1929–1958.