Proceedings of the 5th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/19/20231035

Review of image recognition systems for noisy environments

Junkai Lu

University of Electronic Science and Technology of China (Qingshuihe Campus), 2006 Xiyuan Ave, West Hi-Tech Zone, Chengdu, Sichuan, China.

1838775257@qq.com

Abstract. One of the major challenges faced by image recognition systems in practical applications is the presence of noise in the physical world. To address this challenge, this paper proposes two different approaches. The first approach involves constructing a noisy dataset and training the image recognition system to tolerate noise. The second approach utilizes a combination of denoising methods and pre-trained image recognition systems. Experimental tests and analyses are conducted to evaluate the advantages and disadvantages of each approach. Additionally, this paper investigates the relationship between different noise reduction methods and their impact on improving the recognition rate and resource consumption of images containing Gaussian noise, specifically when neural networks are used for recognition.

Keywords: image recognition, deep learning; image denoising.

1. Introduction

The topic of image recognition and detection is one of the most fundamental aspects of machine learning. Finding items or recognizing images within digital stills or moving pictures is a challenge that is extremely difficult to accomplish. Image recognition has many different applications in the field of computer vision, including facial identification, biometric systems, autonomous vehicles, emotion detection, image recovery, and robotics [1-4]. The field of computer vision has seen substantial progress as a result of the application of deep learning techniques [5]. Emulating the activities of the human cerebral cortex is the goal of deep learning, which is accomplished through the use of artificial neural networks that have numerous hidden layers [6]. Within a deep neural network, these layers are responsible for the extraction of different features, which in turn enables multiple degrees of abstraction.

Traditional approaches to image identification were heavily deployed in the field of computer vision for a significant amount of time before the advent of deep learning. One common strategy involves the employment of handcrafted features, in which characteristics such as color histograms or edge detectors were manually constructed before being input into a machine learning method, such as a support vector machine (SVM). This strategy was quite successful and was widely used. Support Vector Machines, sometimes known as SVMs, are a technique for machine learning that is frequently used in image recognition. SVMs function by locating a hyperplane in a high-dimensional feature space that differentiates between the different image classes in the most accurate manner [7,8]. After that, the SVM makes use of this hyperplane in order to categorize fresh photos according to the feature

DOI: 10.54254/2755-2721/19/20231035

vectors included within them. SVMs have been shown to have a high level of accuracy in a variety of image recognition tasks, including the detection of objects, the recognition of faces, and the recognition of handwriting. In addition, Support Vector Machines (SVMs) are able to process non-linearly separable data by employing a kernel function to translate input characteristics into a higher-dimensional space, which ultimately results in improved classification performance. In conclusion, support vector machines (SVMs) are a powerful tool for image recognition that has been shown to be successful in a wide variety of applications in the real world. Template matching is an additional traditional approach for finding matches. In this method, portions of an input image are compared to a reference image in order to find matches. On the other hand, these traditional methods frequently need for a large amount of domain expertise and do not perform as competently as systems that are based on deep learning when it comes to complicated tasks such as object identification. Despite this, traditional approaches are still useful in certain contexts, such as when performing straightforward image processing operations or dealing with circumstances that involve a finite amount of computer resources.

Deep learning and convolutional neural networks (CNNs), two relatively recent developments in artificial intelligence, have been largely responsible for the rapid progress gained in image recognition over the past few years [9]. Utilizing transformer-based models, which have achieved state-of-the-art performance on a variety of benchmark datasets, is one of the most recent developments in the field of image recognition and is one of the emerging trends. These models make use of attention mechanisms to zero in on significant regions within an image and have shown remarkable success in undertaking tasks such as object detection and segmentation. In addition, recent advancements in image recognition include the incorporation of self-supervised learning approaches, which make it possible for models to learn from unlabeled data, and adversarial training, which is used to increase the resilience of models. Both of these improvements were made possible by the adoption of adversarial training. In conclusion, the most recent advancements in image recognition algorithms continue to push the limits of what is possible in computer vision. These advancements also offer immense promise for a broad variety of potential applications, such as medical diagnosis and driverless cars.

Image recognition is susceptible to being affected by noise for a number of reasons [10]. The act of taking pictures and processing them afterwards can give rise to a significant amount of background noise. Noise can be introduced into photos by cameras due to sensor noise, motion blur, and compression errors. Motion blur also contributes to image noise. In addition, the process of image compression itself can add to the overall level of noise, which makes precise object identification and recognition more difficult. The existence of additional objects or background features within a picture, which can confuse object recognition algorithms, is another kind of noise in image recognition. This type of noise can be particularly problematic for facial recognition. For example, a computer vision system may have difficulty correctly identifying an object if there is a partial obstruction in the way or other objects are present in the immediate area. The training data that is used to train machine learning models can also contain flaws or inaccuracies, which can lead to inaccurate or noisy results in image identification. Another way that noise can be introduced is through the inclusion of faulty or incomplete training data. Existing research, on the other hand, has mostly concentrated on improving the accuracy of previously collected information, while forgetting to take into consideration how noise affects performance in practical small-scale applications. Furthermore, the majority of systems and pieces of hardware are not initially designed with noise reduction in mind, which highlights the crucial need of noise reduction processing that is flexible [11]. As a result, the handling of such noise in image recognition is the topic that will be covered in this study.

By investigating the effects that various noise processing techniques have on neural network recognition, the purpose of this study is to investigate the properties, benefits, and drawbacks of various approaches to noise reduction. To be more specific, we are going to train one neural network to detect noisy datasets, and we are going to train another neural network to distinguish clean datasets. After that, the noisy datasets will be pre-processed with a variety of techniques, specifically BM3D, Gaussian noise reduction, and neural network noise reduction [12]. When these pre-processed datasets

DOI: 10.54254/2755-2721/19/20231035

are fed into the neural network that was intended for pure datasets, we will compare the performance of the network as well as the results that it produces.

The remaining sections of this paper are organized as follows. Section II provides an introduction to classical image recognition networks. Section III outlines the proposed approaches in detail. Section IV presents the experimental tests and analysis. Finally, Section V concludes the papert.

2. Classical image recognition networks

In this section, we provide a description of the existing classical neural networks specifically used for image processing.

One notable example is LeNet-5 [13], a classic convolutional neural network (CNN) developed by Yann LeCun et al. (Figure 1) in 1998. It served as one of the pioneering deep learning models for image recognition and has since influenced the architecture of numerous subsequent CNN models. LeNet-5 was specifically designed for handwritten digit recognition, employing a dataset comprising 32x32 pixel images of handwritten digits ranging from 0 to 9. The network incorporates multiple convolutional and pooling layers, followed by fully connected layers. The convolutional layers extract features from the input images, while the pooling layers reduce the spatial dimensions of the feature maps. The fully connected layers are responsible for classifying the input image based on the extracted features. LeNet-5 achieved state-of-the-art performance on the MNIST dataset, which serves as a benchmark dataset for handwritten digit recognition. Its architecture has had a lasting impact on the development of subsequent CNN models, making it a significant milestone in the history of deep learning.

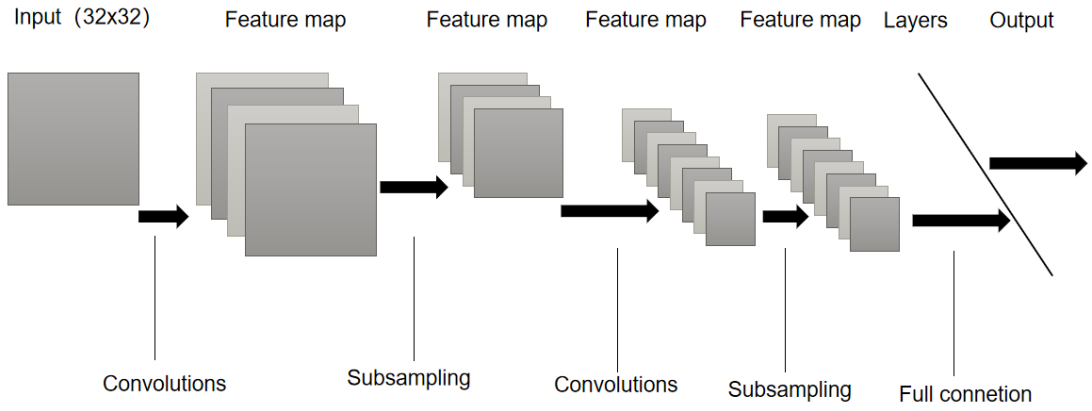

Figure 1. The structure of LeNet.

Convolutional neural networks, also known as CNNs, were first constructed in 2012 by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton under the name AlexNet [14]. It holds the distinction of being the first CNN to win the ImageNet Large Scale Visual Recognition Challenge, a defining moment that pushed the field of computer vision forward at a key juncture in its history. The ImageNet dataset contains millions of photos that fall into thousands of different categories; the goal of the challenge was to accurately categorize each image into the category to which it belonged. AlexNet has a higher number of parameters than LeNet, and it was the first network to use rectified linear units (ReLU) as activation functions, which is now a standard procedure in deep learning. LeNet was the first network to use these activation functions. The structure of AlexNet is made up of several layers of convolutional and pooling operations, which are then followed by layers that are fully connected. The improved performance of AlexNet can be attributed to a number of significant advances that were introduced by AlexNet, such as overlapping pooling, data augmentation, and dropout regularization. Notable accomplishments include AlexNet's achievement of a top-5 error rate of 15.3% on the ImageNet dataset, which is a considerable improvement over the prior state-of-the-art

DOI: 10.54254/2755-2721/19/20231035

performance. The results of this project successfully revealed the promise of deep learning in computer vision, which in turn sparked a wave of research within this particular sector.

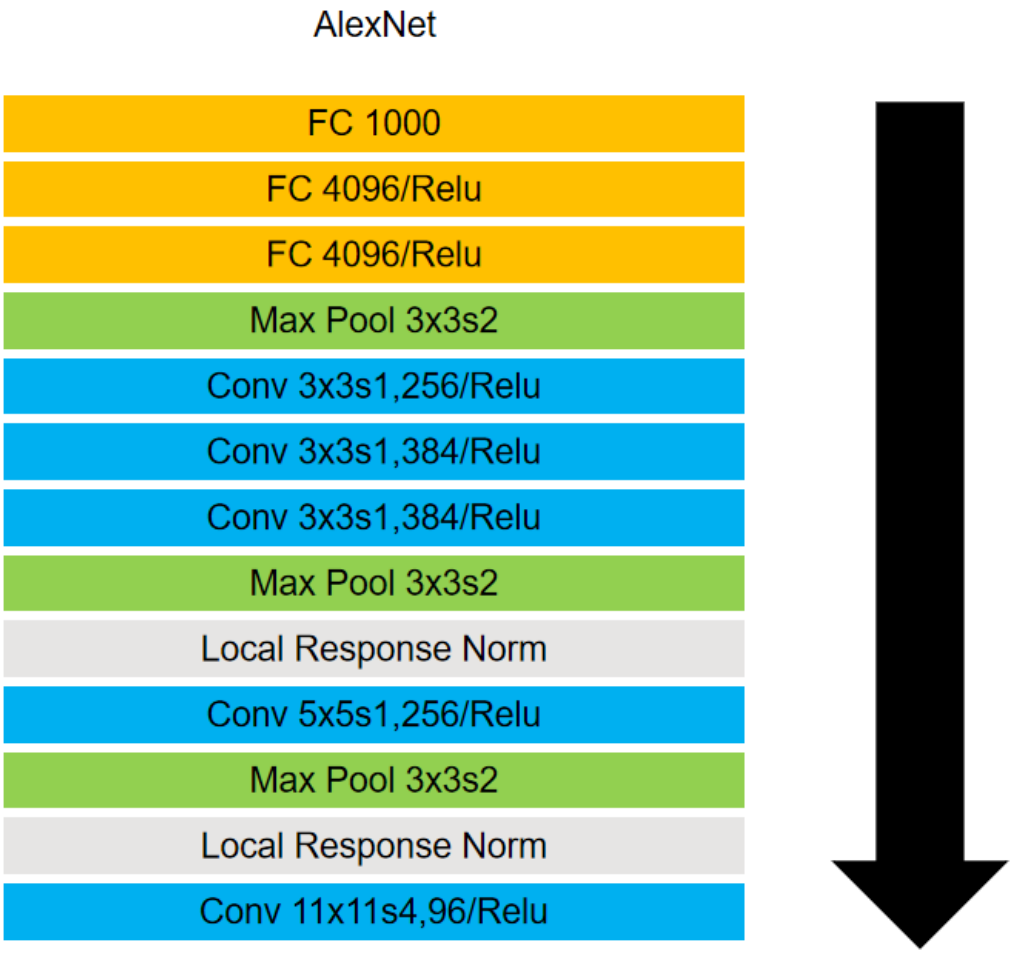

Figure 2. The structure of AlexNet.

Kaiming He and his colleagues in 2015 developed a convolutional neural network (CNN) architecture that they referred to as ResNet [15]. Because of the issue of vanishing gradients, it was previously thought that a network with more than one hundred layers would be impracticable to operate. However, this network has a remarkablely deep structure that exceeds one hundred layers. In order to overcome this difficulty, ResNet has developed a ground-breaking method that they refer to as "skip connection" or "shortcut connection." This method makes it possible for the network to learn residual functions, which simplifies the process of training deep neural networks. The idea that the network can learn to represent the residual function between the input and output of a block, rather than attempting to represent the complete function, is what underpins the skip connection. This is because the network can learn to represent the residual function. As a consequence of this, the number of layers that need to be trained is decreased, which increases the effectiveness of the network in terms of learning. ResNet has shown that it is capable of performing at a level that is considered to be state-of-the-art in a variety of image recognition tasks. These tasks include working with the ImageNet dataset, the COCO dataset, and the PASCAL VOC dataset. In addition, ResNet was the impetus for the development of future architectures, such as DenseNet, which expanded upon the idea of skip connections and achieved even more amazing performance in certain contexts. DenseNet was one of the results of ResNet's role as a catalyst in the development of subsequent architectures.

DOI: 10.54254/2755-2721/19/20231035

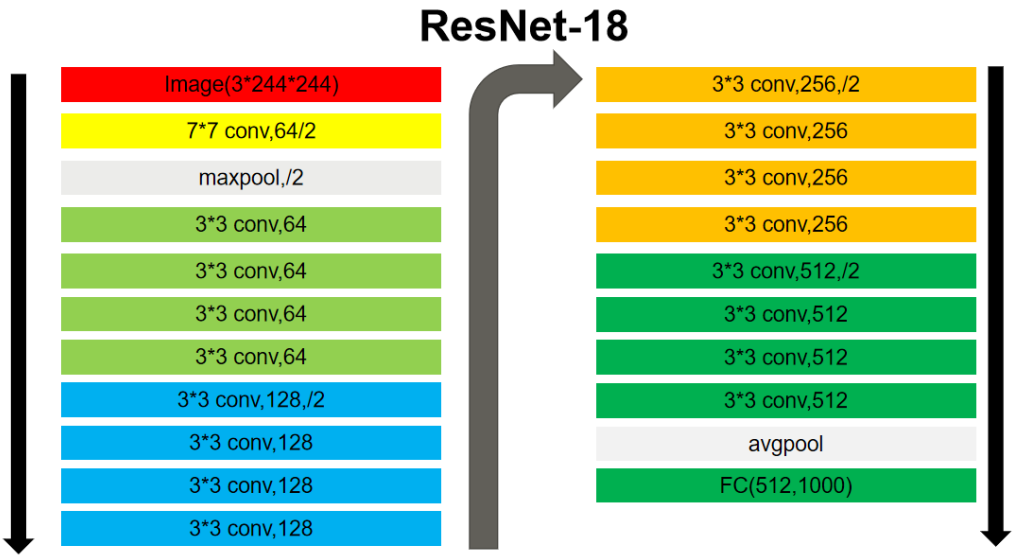

Figure 3. The structure of ResNet.

The Visual Geometry Group (VGG) at the University of Oxford developed the VGG convolutional neural network (CNN) architecture in 2014. This was accomplished in 2014. The design of the network is straightforward and consistent, and it is composed of many convolutional and pooling layers, which are then followed by fully linked layers. The most popular versions of VGG are VGG-16 and VGG-19, which may be distinguished from one another by the number of layers they contain: 16 and 19, respectively. Due to the fact that it is both straightforward and efficient, the VGG architecture has been used as a basis for the development of a great number of subsequent CNN models. Because of its uncomplicated construction and exceptional performance, it has become a popular option in the field of computer vision research.

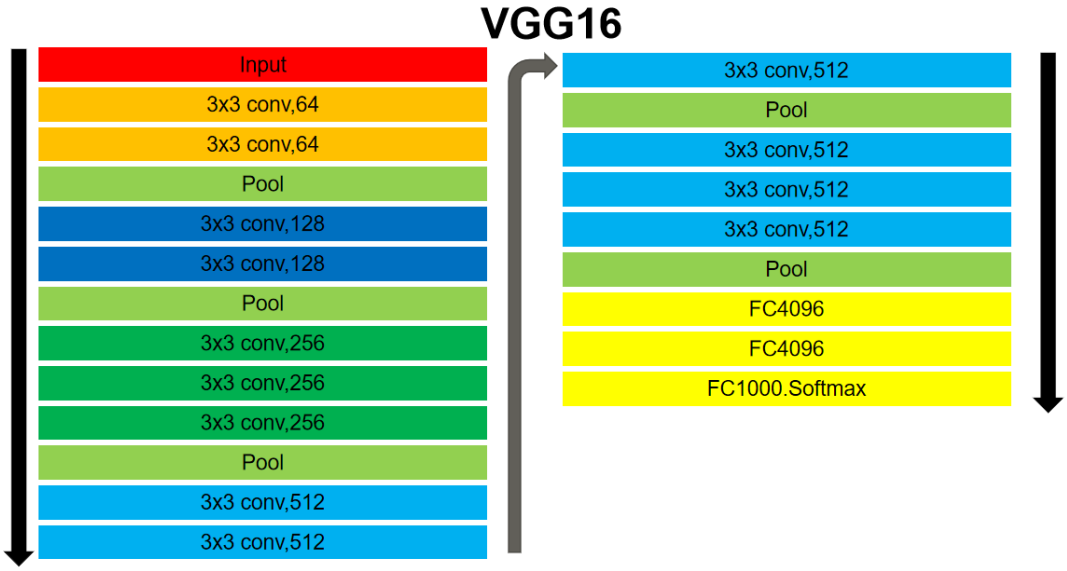

Figure 4. The structure of VGG.

3. Engineering approachest

Image recognition can be made more difficult by the presence of Gaussian noise, which is a type of random noise. Errors that occur during the process of image acquisition or transmission are the source

DOI: 10.54254/2755-2721/19/20231035

of this noise, which appears as erratic oscillations in the values of the pixels and has the potential to negatively affect the accuracy of image recognition algorithms. Gaussian noise is characterized by its adherence to a normal distribution and its typical appearance somewhere in the middle of the intensity range of a picture. The presence of Gaussian noise in image identification is dependent on a number of different aspects. These elements include the techniques of picture acquisition and transmission, the quality of the equipment, and the ambient circumstances that exist when the image is being captured. Notably, Gaussian noise is ubiquitous across a wide variety of image formats. This is especially the case in low-light environments or when the images were captured using a high ISO level. As a consequence of this, it is very necessary for image recognition tasks to take into consideration the possibility of the presence of Gaussian noise and to make use of appropriate approaches for noise reduction in order to improve the accuracy of recognition algorithms.

This section will delve into two methods: utilizing pristine models combined with distinct noise reduction approaches and employing noisy models directly (refer to Figure 5).

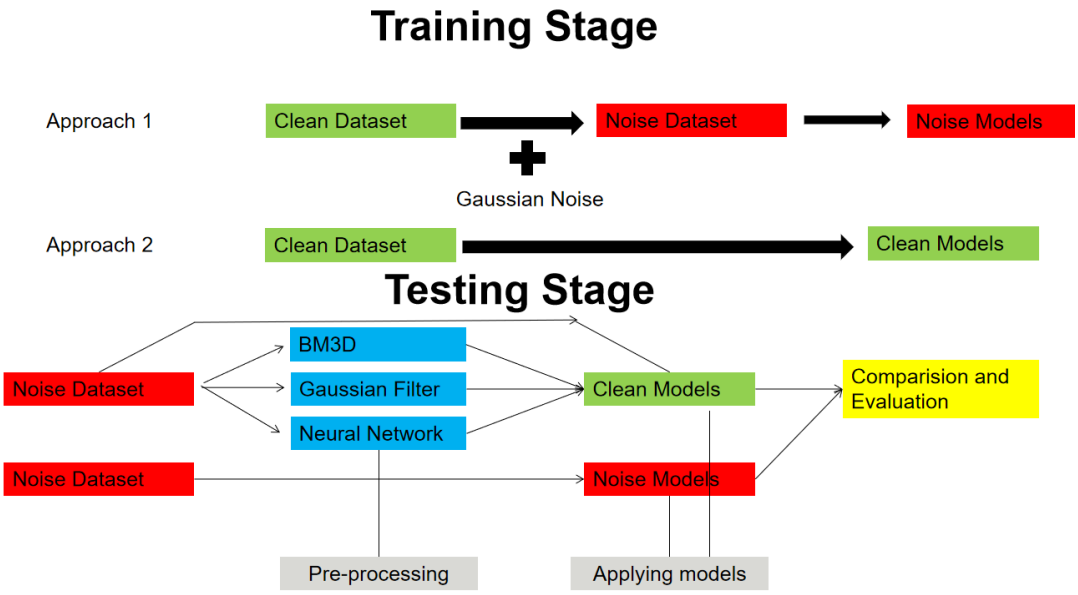

Figure 5. The engineering approaches.

The first approach entails utilizing the model directly trained from the noisy training set as a control to assess the effectiveness and accuracy of different methods. Gaussian noise is employed as the common noise source in this method. Subsequently, noise models are trained by introducing varying levels of Gaussian noise to the dataset, serving as the foundation for comparative evaluation.

The second approach involves training the model using the original clean dataset based on different networks. Artificial noise, specifically Gaussian noise, is then added to the original dataset. The noisy dataset undergoes pre-processing with various noise reduction methods before being evaluated using the clean model. Additionally, a set of controls is established by employing a noisy dataset with a pure network.

For the first approach, the selection of an appropriate denoising method is crucial. Considering practical application constraints such as computational cost and performance, numerous options are available. This paper explores three tested noise reduction methods: BM3D, Gaussian noise reduction, and neural network noise reduction.

A Gaussian filter represents a widely employed image smoothing filter that functions by convolving the image with a Gaussian function. The filter derives its name from the Gaussian distribution, which characterizes a bell-shaped curve mathematically. Operating as a linear filter, the Gaussian filter processes each pixel in the image by computing a weighted average of its neighboring pixels. The weights are determined by the Gaussian function, which assigns greater weight to pixels

DOI: 10.54254/2755-2721/19/20231035

closer to the center of the filter and lesser weight to those farther away. Consequently, the filter introduces a blurring or smoothing effect on the image, facilitating noise reduction and the elimination of small details while preserving the overall structure and image features. Adjusting the filter size and the standard deviation of the Gaussian function enables control over the strength of the smoothing effect.

BM3D, short for Block-Matching and 3D Filtering, is an image denoising algorithm initially introduced in 2007. The algorithm capitalizes on the notion that images comprise similar patterns or structures within small blocks, and by leveraging this similarity, it efficiently eliminates noise while preserving essential image details. BM3D operates through two stages: block-matching, wherein similar blocks are identified and grouped together, and 3D filtering, where the blocks collaboratively filter out noise from the image. BM3D has demonstrated remarkable effectiveness in noise reduction while retaining details, thus becoming a favored choice in image processing applications.

Denoising by neural network is a technique that employs deep learning algorithms to eliminate noise from images. This process involves training a neural network on a substantial dataset comprising clean and noisy images, enabling the network to learn how to map noisy images to their corresponding clean counterparts. Once trained, the network can be utilized to denoise new, unseen images by feeding them into the network and obtaining the output image. The advantage of employing a neural network for denoising lies in its ability to learn intricate and nonlinear relationships between noisy and clean images, allowing for noise removal while preserving crucial image details. Various types of neural network architectures can be employed for denoising, including convolutional neural networks (CNNs) and generative adversarial networks (GANs), each possessing distinct strengths and weaknesses. Denoising by neural network has exhibited promising outcomes and is progressively gaining popularity in image processing and computer vision applications.

4. Comparison of different approaches

In this section, the above selected neural network model will be trained and tested in combination with different noise reduction methods. Here are the specific results:

4.1. Comparison of different approaches

In general, Approach 2(noise model) has better performance in terms of accuracy.

Table 1. Accuracy for each set.

	Clean Model (%)	Gaussian +Clean Model(%)	BM3D +Clean Model(%)	Denoising network +Clean Model(%)	Noise Model(%)
ResNet-18	0.58	0.62	0.71	0.76	0.79
AlexNet	0.53	0.63	0.65	0.71	0.77
LeNet	0.50	0.56	0.60	0.63	0.72

The above table shows the recognition accuracy of each method for the noisy test set under different conditions.

It can be clearly seen that no matter what noise reduction scheme is used, the effect of using a pure model is not as good as using a noise model directly under the application of the same kind of recognition network.

4.2. Comparison of noise reduction schemes

In Approach 1, we have four control groups whose recognition accuracy also varies significantly according to the different methods, and the data will be analyzed below.

4.2.1. Accuracy Analysis. As can be seen from the table, the best result is using a neural network for noise reduction, followed by using BM3D, then using a Gaussian filter, and the worst is directly

DOI: 10.54254/2755-2721/19/20231035

training with a pure model. The accuracy improvement of adjacent noise reduction methods is about 5%.

4.2.2. Time consumption.

Table 2. Time consumption(testing) for each set.

Time for recognition (200 target)	Clean Model	Gaussian +Clean Model	BM3D +Clean Model	Denoising network +Clean Model	Noise Model
ResNet-18	136.81ms	209.56ms	7min3.32s	30.12s	144.56ms
AlexNet	140.31ms	211.33ms	7min17.23s	28.44s	142.45ms
LeNet	141.97ms	208.31ms	7min2.12s	44.47s	136.63ms

Based on the aforementioned results, it can be observed that direct training with the trained model is the fastest, taking approximately 140ms, regardless of whether the training set is noisy or not. When a Gaussian noise reduction process is added, the recognition time slightly increases to around 200ms. However, when employing the BM3D algorithm, the recognition time dramatically increases due to the requirement of converting the image tensor into an image format before processing, a process that consumes a considerable amount of time, approximately 7 minutes. Finally, utilizing neural network noise reduction also consumes approximately half a minute, around 34 seconds.

It is important to note that the algorithms of the Gaussian filter and BM3D are fixed, whereas neural network noise reduction needs to be trained in advance based on the specific situation or practical objective. The training time for neural network noise reduction in this experiment is approximately 20 minutes.

4.3. *Comparison of different networks*

Based on the aforementioned results, there is a noticeable distinction among the different networks when comparing their accuracy rates over time. Overall, ResNet exhibits the most favorable performance, achieving the highest recognition accuracy, except in the group that utilizes Gaussian filters. AlexNet follows with a slight decrease of several percentage points, while LeNet performs the worst. This outcome aligns with expectations as the structure and parameter content of these networks evolve with development, gradually increasing their effectiveness.

Regarding time consumption, the table above does not demonstrate significant differences among the various network architectures.

4.4. Comprehensive evaluation

In the selection of noise reduction methods, it is observed that more intricate techniques offer a certain percentage of performance enhancement at the cost of increased computational resources and time. Gaussian filters present a slight improvement in algorithm performance with minimal additional resource consumption. On the other hand, BM3D can achieve relatively higher accuracy by processing images individually, albeit at the expense of consuming a considerable amount of time. Neural network noise reduction yields the best recognition results with relatively fast processing, but it necessitates the training of a noise reduction network tailored to the specific environment of the recognition dataset, which requires additional time investment.

5. Conclusion

This paper focuses on investigating engineering approaches for handling the issue of noise in image recognition. Two distinct approaches are introduced: the first approach involves constructing a noisy dataset and retraining the image recognition network, while the second approach relies on denoising methods followed by image recognition. In the second approach, three denoising methods, namely Gaussian filter, BM3D, and denoising network, were tested. The experimental results and analysis are

presented in detail to demonstrate the advantages and disadvantages of each method. These findings aim to serve as a guideline for industries in selecting the most suitable approach for their specific purposes.

Reference

[1] D. Keysers, T. Deselaers, C. Gollan and H. Ney, "Deformation Models for Image Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 8, pp. 1422-1435, Aug. 2007, doi: 10.1109/TPAMI.2007.1153

[2] Z. Wang, C. Tang, X. Sima and L. Zhang, "Research on Application of Deep Learning Algorithm in Image Classification," 2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), Dalian, China, 2021, pp. 1122-1125, doi: 10.1109/IPEC51340.2021.9421185

[3] B. Samia, Z. Soraya and M. Malika, "Fashion Images Classification using Machine Learning, Deep Learning and Transfer Learning Models," 2022 7th International Conference on Image and Signal Processing and their Applications (ISPA), Mostaganem, Algeria, 2022, pp. 1-5, doi: 10.1109/ISPA54004.2022.9786364

[4] Perry, S. (2018). Image and Video Noise: An Industry Perspective. In: Bertalmío, M. (eds) Denoising of Photographic Images and Video. Advances in Computer Vision and Pattern Recognition. Springer, Cham. https://doi.org/10.1007/978-3-319-96029-6_8

[5] Y. Zhang and X. Zheng, "Development of Image Processing Based on Deep Learning Algorithm," 2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), Dalian, China, 2022, pp. 1226-1228, doi: 10.1109/IPEC54454.2022.9777479

[6] A. Shrestha and A. Mahmood, "Review of Deep Learning Algorithms and Architectures," in IEEE Access, vol. 7, pp. 53040-53065, 2019, doi: 10.1109/ACCESS.2019.2912200

[7] Chauhan, V.K., Dahiya, K. & Sharma, A. Problem formulations and solvers in linear SVM: a review. Artif Intell Rev 52, 803–855 (2019). https://doi.org/10.1007/s10462-018-9614-6

[8] Mustafa Abdullah, D., & Mohsin Abdulazeez, A. . (2021). Machine Learning Applications based on SVM Classification A Review. Qubahan Academic Journal, 1(2), 81–90. https://doi.org/10.48161/qaj.v1n2a50

[9] Alzubaidi, L., Zhang, J., Humaidi, A.J. et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data 8, 53 (2021). https://doi.org/10.1186/s40537-021-00444-8

[10] T. Rahman, M. R. Haque, L. J. Rozario and M. S. Uddin, "Gaussian noise reduction in digital images using a modified fuzzy filter," 2014 17th International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 2014, pp. 217-222, doi: 10.1109/ICCITechn.2014.7073143

[11] C. R. Steffens, L. R. V. Messias, P. L. J. Drews and S. S. d. C. Botelho, "Can Exposure, Noise and Compression Affect Image Recognition? An Assessment of the Impacts on State-of-the-Art ConvNets," 2019 Latin American Robotics Symposium (LARS), 2019 Brazilian Symposium on Robotics (SBR) and 2019 Workshop on Robotics in Education (WRE), Rio Grande, Brazil, 2019, pp. 61-66, doi: 10.1109/LARS-SBR-WRE48964.2019.00019

[12] Gu, S., Timofte, R. (2019). A Brief Review of Image Denoising Algorithms and Beyond. In: Escalera, S., Ayache, S., Wan, J., Madadi, M., Güçlü, U., Baró, X. (eds) Inpainting and Denoising Challenges. The Springer Series on Challenges in Machine Learning. Springer, Cham. https://doi.org/10.1007/978-3-030-25614-2_1

[13] Al-Jawfi R. Handwriting Arabic character recognition LeNet using neural network[J]. Int. Arab J. Inf. Technol., 2009, 6(3): 304-309

[14] Alom M Z, Taha T M, Yakopcic C, et al. The history began from alexnet: A comprehensive survey on deep learning approaches[J]. arXiv preprint arXiv:1803.01164, 2018

DOI: 10.54254/2755-2721/19/20231035

[15] Wu Z, Shen C, Van Den Hengel A. Wider or deeper: Revisiting the resnet model for visual recognition[J]. Pattern Recognition, 2019, 90: 119-133