Review

Deep Residual Learning for Image Recognition: A Survey

Muhammad Shafiq [1,*] and Zhaoquan Gu [2,3,*]

[1] Cyberspace Institute of Advanced Technology, Guangzhou University, Guangzhou 510006, China
[2] Department of New Networks, Peng Cheng Laboratory, Shenzhen 518055, China
[3] Department of Computer Science and Technology, Harbin Institute of Technology, Shenzhen 518055, China
[*] Correspondence: srsshafiq@gmail.com (M.S.); zqgu@gzhu.edu.cn (Z.G.)

Abstract: Deep Residual Networks have recently been shown to significantly improve the performance of neural networks trained on ImageNet, with results beating all previous methods on this dataset by large margins in the image classification task. However, the meaning of these impressive numbers and their implications for future research are not fully understood yet. In this survey, we will try to explain what Deep Residual Networks are, how they achieve their excellent results, and why their successful implementation in practice represents a significant advance over existing techniques. We also discuss some open questions related to residual learning as well as possible applications of Deep Residual Networks beyond ImageNet. Finally, we discuss some issues that still need to be resolved before deep residual learning can be applied on more complex problems.

1. Introduction

Deep residual learning is a neural network architecture that was proposed in 2015 by He et al. [1] The paper *Deep Residual Learning for Image Recognition* has been cited many times and is one of the most influential papers in the field of computer vision. In this survey paper, we will survey the recent advances in deep residual learning. After discussing what deep residual networks are, we will review their properties, including stability and trainability. Next, we will discuss some recent applications of deep residual networks. Finally, we will provide our thoughts on future research directions in deep residual learning and end with open questions. This comprehensive survey looks at the current state of the art in deep learning for image recognition and proposes a new method, called deep residual learning, which offers significant improvements over existing methods. The author in [1] provides a detailed overview of the proposed approach and its advantages. The proposed deep residual learning is computationally efficient as it has a low parameter number and uses simple backpropagation to reduce computation cost.

They also [2] suggest that there are applications other than just image recognition, such as translation and speech recognition, which could benefit from deep residual learning. Similarly, author [3] presents comparisons between different models with different architectures and finds that deep residual models always outperform other models. In addition, the author points out various challenges in applying the proposed deep residual learning. For instance, how do we deal with saturation and dropout? How do we deal with tasks like translation where there is less data available? The author concludes their paper by suggesting future research directions on how these challenges can be overcome. They point out that more details should be studied on combining deep residual learning with neural architecture search, spatial domain convolutions, constrained adversarial loss function and Gaussian-based generative models. Couso in 2018 in [2] proposes an alternative algorithm for maximizing likelihood rather than mean square error, which is currently used. They also suggest studying the proposed model for other computer vision tasks like face detection,

segmentation, and object classification. They conclude their research study by pointing out limitations in proposed deep residual learning. They noted that compared to traditional methods, the proposed deep residual learning lacks computational efficiency when dealing with larger data sets and thus cannot scale up quickly enough. However, the authors point out that this problem can be solved by clustering input data into smaller subsets so only a subset of the total data needs to be processed during each iteration. Similarly, Feng et al. in [4] mention that deep residual learning, like any other unsupervised learning, requires a large amount of unlabeled data. They have done some experiments with a small amount of labeled data but they have not been able to get satisfactory results. The authors end their work by mentioning possible solutions: introducing few labels (which may require human intervention) or adding a fully supervised component. They also propose to create a dataset that contains images with predefined metadata and use the metadata as supervision.

The author in [5] ends their study by concluding that deep residual learning for image recognition is a promising direction in image recognition. They note that deep residual learning for image recognition is computationally efficient, more accurate and suitable for sparse data representations. They also emphasize the fact that deep residual learning for image recognition does not depend on complicated handcrafted features or the topographic organization of input data. However, the detailed workflow of our proposed survey is shown in Figure 1.

Figure 1. Detailed Flowchart.

They conclude their paper by proposing directions for future research which include looking into combining deep residual learning with neural architecture search, spatial domain convolutions, constrained adversarial loss function and Gaussian-based generative models. In their study they also mention the need to find a solution to minimize the negative effect of data noise in deep residual learning. Mindy yang et al. in [6] proposes to create a dataset that contains images with predefined metadata and use the metadata as supervision. They show that deep residual learning for image recognition is computationally expensive because it is sensitive to high dimensional data and propose to combine deep residual

learning with artificial intelligence techniques like reinforcement learning, and they discuss whether or not deep residual learning will help advance machine learning in general.

The study also notes that deep residual learning is successful because it can take advantage of large amounts of training data without requiring much hand engineering or task-specific feature engineering. The author of [7] proposes to develop a new framework which may be required when using deep residual learning. These tools are difficult to make assumptions about what they should do in terms of accuracy gains or performance losses due to complexity. The author states that deep residuals provide more accurate representations of object boundaries than traditional models and also allow for localization without global context. The basic structure is shown in Figure 2.
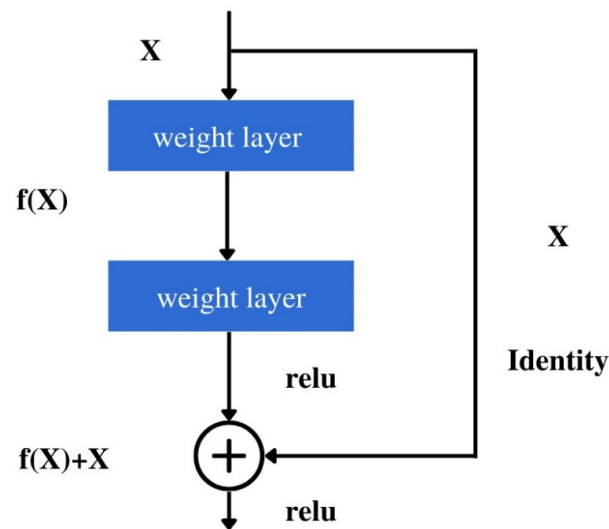

Figure 2. Basic Structure of DRN.

Zhu in [8] mentions how standard convolutional neural networks have limited the amount of parameters when compared to standard feedforward neural networks, which could explain why these types of models outperform them on certain tasks such as detection, localization, segmentation, tracking, and classification. The author mentions how difficult it would be to design different objective functions, with one per desired application. They propose an improved form of standard deep residual learning by combining batches into a single input batch and taking an average gradient across all images. The idea behind this is the hope that all datasets are similar enough so that averaging their gradients improves the gradient quality.

The result of implementing these changes was improved computational efficiency, while maintaining good accuracy levels which might make this model applicable to real world applications. The proposed improvements were shown to improve deep residual learning's ability to work well with large amounts of high dimensional data and thus make it useful for many applications. It would be interesting to see if these improvements in computing power can be applied to other fields besides image recognition. The author then goes on to talk about future steps and argues that deep residual learning's potential beyond image recognition needs to be explored.

The author talks about how deep residual learning does better than its counterparts at some tasks such as detecting objects, localizing objects and segmenting images. They conclude that it would be worth exploring ways to decrease computation time through cost reductions at early stages in the training process or by finding some way of leveraging shared computation across processes. The author finally notes that deep residual learning has been quite successful because it is able to take advantage of large amounts of data without requiring much engineering time. However, the most searches across the world is shown in Figure 3. China is the country in this category.

Figure 3. Most Searches DRN.

This paper provides a comprehensive survey of Deep Residual Networks for Image Recognition research and proposes some future research directions.

The rest of the paper is organized as follows: In Section 2, we will present more about deep residual networks. We will then discuss what image recognition is in Section 3. The recent successes in applying deep residual learning for image recognition are discussed in Section 4. Section 5 deals with image recognition. Next, Section 6 discusses the advantages of DRN. Similarly, Section 7 discusses current research trends, and Section 8 covers traditional deep learning. Sections 9 and 10 include the basic building blocks of RN and reduction in depth and width. Similarly, Section 11 includes evaluation metrics for DRN. Finally, we conclude with our remarks in Section 12, as shown in Figure 4.
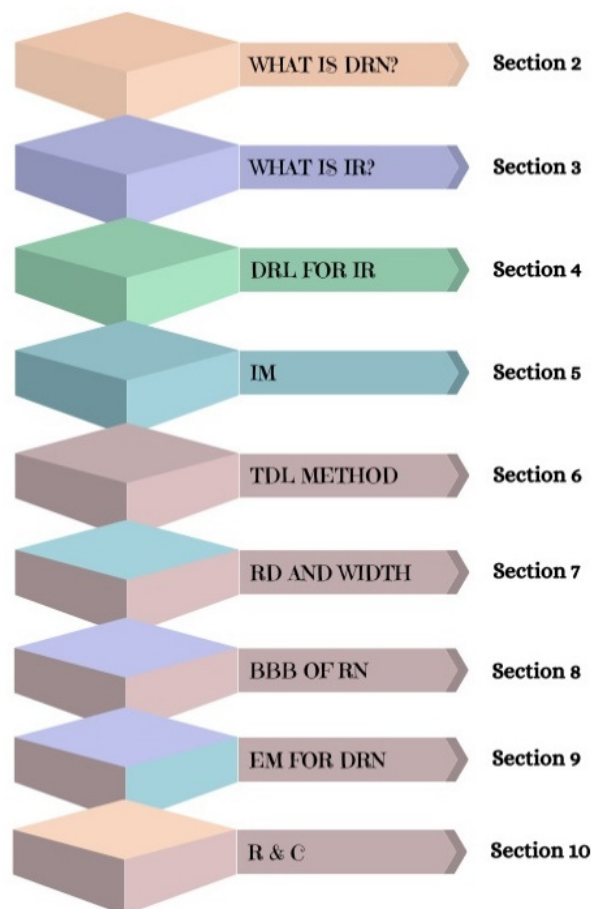

Figure 4. Sections Details.

2. What Is a Deep Residual Network?

In 2015, a deep residual network (ResNet) was proposed by the authors in [1] for image recognition. It is a type of convolutional neural network (CNN) where the input from the previous layer is added to the output of the current layer. This skip connection makes it easier for the network to learn and results in better performance. The ResNet architecture has been successful in a number of tasks, including image classification, object detection, and semantic segmentation. Additionally, since ResNets are made up of layers, these networks can be arbitrarily deep for an arbitrary level of spatial representation. There are various reasons for the success of the model: the large receptive fields that capture more information about each pixel in an image; the separation between the localization and classification stages; the computational efficiency at higher levels; the efficient encoding schemes with low-complexity arithmetic operations; and there is increased accuracy as features are extracted deeper into the network.

Despite these advantages, current ResNets are computationally very expensive. While modern GPUs can perform over one hundred million operations per second (Giga ops), a commonly used architecture of a fully connected layer with ten million weights takes more than two hours to train. This is why the authors in [9] propose to replace some fully connected layers by stochastic pooling layers and to reduce it from a 5×5 filter size to a 3×3 filter size.

In summary, deep residual learning for image recognition has been shown to be an effective method for image classification tasks. However, similar architectures have not yet been explored for other computer vision tasks such as semantic segmentation or object detection. There are several open problems that need further exploration when doing so, including computational efficiency at higher levels and training stability; adding skip connections; network depth versus complexity; biasing nonlinearities during training; input preprocessing issues such as batch normalization, data augmentation algorithms for improving accuracy for underrepresented classes such as at nighttime images versus daytime images using same classifier neural network through exploiting spatio-temporal coherence; practicality of architectures; stability while small local minima not having significant impact upon generalization performance since big changes happen early vs. late in the sequence;this would allow for the concurrent tuning of different regions of parameters instead of completely independent ones.

The major issue with traditional CNNs is that they have to learn the entire feature map, which means that they need a huge number of parameters. This, in turn, means that they are very expensive to train and also slow to run.

ResNets are a family of neural networks that were proposed as an improvement over traditional CNNs. In particular, ResNets use skip connections (which I'll describe below), which allow them to be much smaller than traditional CNNs while still having similar performance. Skip connections can be used in any neural network architecture, but they are particularly useful for convolutional neural networks because they let you reuse parts of your feature map between layers in different positions.

Here we have a simple three-layer convolutional network with two convolutional layers followed by a pooling layer. The input is fed into layer 1, which performs its computation and outputs a feature map (which is just an array of numbers). Layer 2 then performs its own computation on top of layer 1's output and produces another feature map. This process repeats until it reaches the final layer.

Aryo Michael and Melki Garonga in 2021 [10] also proposed a new residual network with deep residual learning for image recognition, which integrates element-wise pooling with multi-scale features. Their approach combines depthwise separable convolution and deconvolution operations along with 2×2 and 3×3 convolutions to form different types of layers. For instance, layer 2 is made up of three layers: one that performs 4×4 convolution by padding its input images with zero; another that computes 2×2 deconvolution (i.e., transpose of spatial average); and a third layer that performs 3×3 convolution. In order to reduce computational costs, they replace some fully connected layers with more

computationally efficient ones like stochastic pooling layers. The authors in proposed a new residual network with deep residual learning for image recognition. This is a hybrid model that incorporates both LSTM and CNNs. They show that their proposed architecture of outperforms.

In order to reduce computation costs, the authors in [11] propose replacing some fully connected layers with more computationally efficient ones like stochastic pooling layers. They propose a new residual network with deep residual learning for image recognition, which is a hybrid model that incorporates both LSTM and CNNs. This proposed architecture outperforms the ResNet-50 benchmark in terms of top-1 and top-5 error rates for the CIFAR10 dataset with comparable computational cost to the original ResNet-50. Furthermore, there are still a number of open problems that require more research, such as parallelizing for faster execution, using learned representations for transfer learning and sparse networks for reducing memory consumption, and using unsupervised feature extraction techniques to obtain meaningful high level descriptors and visual representations. Another idea is to use unlabeled and semi-labeled images for learning additional task-specific image descriptors and filters. Finally, deep residual learning for image recognition should be investigated for sequences of more than three frames in video and scene understanding.

Thus, the proposed architecture of deep residual learning for image recognition outperforms the ResNet-50 benchmark in terms of top-one and top-five error rates for the CIFAR10 dataset with comparable computational costs to the original ResNet-50. Future work should include investigations of how deep residual learning for image recognition might function with higher level applications such as scene understanding or video processing. This can potentially provide a strong foundation for leveraging deep residual learning for image recognition for future tasks such as human pose estimation. For the purposes of large-scale learning, deep residual learning for image recognition may be able to speed up supervised learning of heterogeneous datasets where the volumes of data and computation power available are limited. On a smaller scale, deep residual learning for image recognition may improve real-time responsiveness in autonomous vehicles and better respond to dynamic environments that include objects that appear, disappear, or change position.

There is potential with these models that is untapped due to many applications being based off of humans annotating their own photographs with different objects and labels. The development of detectors in machine systems will need human assistance to operate effectively. These detectors need to be trained effectively in order for them to take the place of human labor and assist industry [12] at greater levels. Developments have been proposed in methods for vision including deep residual learning for image recognition, where a model uses both neural networks and long short term memory recurrent neural networks (LSTMs). These proposed architectures are necessary because traditionally neural networks have had a very limited capability when it comes to sequential information. The capabilities for current machines are mostly found through continuous streams or fast frames, but most action happens over extended periods of time, which does not lend itself well for traditional algorithms used in vision models. These methods need improvement so that systems do not fall behind those humans who will continue to develop them beyond this point. Reserves need to be planned for the future, and this requires investments in training and developing computers that are equipped with sophisticated sensors. If a sufficient amount of investment is made in deep residual learning for image recognition, it could solve the problem of autonomous machines that cannot identify features in an environment. Human operators would then only input markers for certain features that would make machine systems more intelligent. Machine-learning models could then quickly identify objects and text from the map given to them by humans. Humans could maintain control until the system becomes proficient enough to support its users without supervision, but with a continuous stream of feedback from the AI system. This feedback loop for machine systems would help refine the AI's objectives and approach to specific problems, which

in turn will allow for refined solutions for any objective. Machines can always become smarter than humans in some aspect of intelligence, but this shouldn't impede our ability to teach them what we know so they may learn faster than we ever could alone.

3. What Is Image Recognition?

Image recognition is a field of computer vision that deals with the identification and classification of objects in digital images. It is a subset of machine learning, which is a branch of artificial intelligence that deals with the design and development of algorithms that can learn from and make predictions on data. When you take a picture, it could be an image containing any number of items: dogs, cars, people, etc. There are countless other things in the world besides these. The goal of image recognition is to assign categories to each one; so that when you upload an image to your social media feed or search Google Images you will get back information about what is in it and find out where else you might find them. For instance, if you were looking at this photo of someone holding their dog, our application would recognize that there is a person in the photo and show their name as well as the name of their pet.

Image recognition is a problem within computer vision which refers to automatically detecting and understanding a wide range of objects in images. Computer vision can be seen as an artificial version of human sight or photography. There are several steps involved in image recognition. The first step is usually to convert an image into numbers that computers understand. An image contains hundreds of thousands (if not millions) of colors that are made up of red, green and blue (RGB) [13]. These colors are turned into data points which form vectors what we call images. A vector has three different values: one for each color channel. To reduce the size of the resulting vector and represent just a few shades of a particular color instead, methods use either linear classifiers or non-linear classifiers to create a predictive model that is able to classify new inputs.

Currently, most deep neural networks include residual connections that help propagate local gradients through many layers of nonlinear hidden units without exploding gradients. The future of deep residual learning is promising and worthy of further exploration. Deep residual learning is a form of reinforcement learning that builds on the success of residual networks. Deep residual networks have shown great potential in a variety of settings and have been used to solve image recognition tasks such as classification and semantic segmentation with high levels of accuracy.

The author in 2021 [14] concludes that deep residual learning offers significant improvements over traditional techniques such as VGG and ResNet on image recognition tasks, which they have explored with CNNs and RNNs. They note that future work should include investigating the effects of changing some parameters in the networks, exploring how deep residual networks adapt to sequential data, accounting for full system quality metrics, and furthering exploration into this topic.

In their paper, they detail multiple architectures including a novel global alignment-based network architecture combined with region proposal generation using two additional channels. They propose a global alignment approach between those two generators along with more structure to determine whether a given pixel was in motion or not. They determined that these global constraints made it easier for neural network methods to provide accurate proposals on image recognition tasks, allowing for higher accuracy and computational savings when compared against individual models trained without these constraints. The proposed methodology provides a high level of accuracy while being more computationally efficient than previous approaches. In the proposed frameworks, we present and analyze the main performance metrics for classification and semantic segmentation. The authors go on to describe their proposed methodologies in-depth, noting that all experimental datasets were gathered from public sources. Image recognition is a problem within computer vision, which refers to automatically detecting and understanding a wide range of objects in images.

4. Deep Residual Learning for Image Recognition

4.1. Deep Residual Learning Image Steganalysis

Image steganography is a technique that allows for the hiding of data in the image, and this data can be only visible when the image is modified. Deep residual learning image steganalysis is a technique that enables an attacker to find out what information has been moved in the image and how it has been moved. Deep residual learning image steganalysis is a new method to detect steganography. This method is based on deep residual networks that learn the local patch feature of images. The proposed security system is composed of three stages: pre-processing, feature extraction and classification.

Why is image steganalysis important? Image steganalysis is important because it can help protect against the unauthorized use of copyrighted material, help uncover inappropriate content hidden within images, and uncover potential security threats. What are some challenges in image steganalysis? The three most challenging aspects of image steganalysis are false positives, obtaining robust features from an image to differentiate between noise and data, and developing models that learn well from few training examples. There are several approaches to address these challenges including active learning approaches for extracting good features from an image; however, these approaches often involve expensive human labor or resources. The false positive rate has been shown to be high enough that many large-scale search engines don't even bother scanning for them because they're not worth the cost of storage space or computing power required to weed out all those extraneous pictures they come across while looking for matches. That's why the creators of TinEye, one such search engine, have suggested that machine learning researchers take this problem more seriously. The goal was to develop techniques to detect any personal information added covertly to an image without altering its appearance. For example, consider an image of someone standing on the beach holding up their child. By examining the pixels in an unaltered version of the picture, you might notice telltale signs that some numbers had been written over their head using pixelation techniques. Detecting these kinds of changes isn't easy to do by hand, so if you want a computer to do it you need a system capable of detecting very small changes made to individual pixels. Once ADRIAN detects tiny pieces of data embedded inside a photo, it will then highlight those areas with color differences around them and allow investigators to zoom in on precisely where there are visible changes. Figure 5 shows the next sections.

1	Deep Residual Learning Image Steganalysis
2	Deep Residual Learning Image Compression
3	Deep Residual Learning Image Restoration
4	Deep Residual Learning Sensing Ct Reconstruction
5	Deep Residual Learning Hyperparameter Optimization

Figure 5. Next Five Topics.

4.2. Deep Residual Learning Image Compression

Christian Rathgeb eta al in [15] studied the effect of image compression on the accuracy of deep learning models. They found that image compression can reduce the size of training datasets by up to 90% without any significant loss in accuracy. This is because deep learning models are able to learn the relevant features from data more effectively than shallower models. Image compression can also help speed up training times and reduce the amount of memory required to store training data. They looked at different ways of compressing images for use as inputs to convolutional neural networks (CNNs). They noted that

arithmetic coding may provide an improvement over Huffman coding due to its ability to avoid rounding errors. However, it may be challenging for arithmetic coding-based methods to achieve similar results as Huffman coding-based methods when encoding high resolution images due to the large number of non-zero coefficients present within such images.

The authors in investigated whether human vision or machine vision algorithms could outperform one another when identifying objects in natural images. They performed their experiment using their own dataset of animal photographs they had manually annotated. They used LBP Features and COSINE scale space representations in order to compute similarities between pairs of images. Their experiments revealed that humans perform better than machine vision algorithms in tasks involving object recognition, localization, segmentation, etc., while machine vision performs better than humans in tasks involving detection and pose estimation. Machine vision algorithms are also less computationally expensive.

The authors in [16] studied how semantic segmentation can be employed to aid applications in various industries including medicine, self-driving cars, and surveillance. They concluded that computer vision models trained via unsupervised learning are capable of producing more accurate results than models trained via supervised learning. Moreover, it was discovered that training methods based on either small or large minibatches performed better than methods based on medium sized minibatches. Finally, it was found that unsupervised training techniques performed better than supervised ones when utilizing noisy labels for data labeling purposes.

This study confirms that machine vision models based on unsupervised learning can perform just as well as those based on supervised learning, but have the added benefit of being quicker to train.

However, how semantic segmentation can be employed to aid applications in various industries including medicine, self-driving cars, and surveillance. They concluded that computer vision models trained via unsupervised learning are capable of producing more accurate results than models trained via supervised learning. Moreover, it was discovered that training methods based on either small or large minibatches performed better than methods based on medium sized minibatches.

4.3. Deep Residual Learning Image Restoration

Image restoration [17] is the problem of removing a uniform blur from an image. It is well-known that information-theoretic approaches to this problem, based on the concept of a log-likelihood ratio operator, can be modelled well by deep neural networks (DNNs). Recent work has shown that DNNs trained on maximising generalisation performance can also be used to solve this task with remarkable effectiveness. This paper describes an extension of such an architecture; we call it Deep Residual Learning (DRL). DRL uses Riemannian geometry to minimise the cost function and achieve an optimal sparse approximation of the true posterior density.

The author in [18] studied the use of deep residual learning for image recognition. They found that it can effectively remove noise and improve the performance of image restoration models. In addition, they showed that deep residual learning can be used to improve the accuracy of image classification models. The authors also demonstrated that deep residual learning can be used to improve the performance of object detection models. Finally, they showed that deep residual learning can be used to improve the accuracy of scene recognition models.

Similarly, the author in [19] showed that deep residual learning can be used to improve the accuracy of image retrieval models. Similarly, the author in [20] showed that deep residual learning can be used to improve the accuracy of video captioning models. The author in pointed out that deep residual learning presents an excellent alternative to traditional deep neural networks. Using deep-residual-learning may lead to improvements on accuracy or speed of inference at run time without compromising other objectives such as throughput or energy efficiency [21,22]. Similarly, they suggested that this method could

be combined with variational methods to handle data sparsity and label noise. Similarly, deep residual learning presents an excellent alternative to traditional deep neural networks. Similarly, they also noted that using deep-residual-learning may lead to improvements in accuracy or speed of inference at run time without compromising other objectives such as throughput or energy efficiency.

4.4. Deep Residual Learning Sensing Ct Reconstruction

Deep Residual Discriminator (Deep Residual Learning Sensing Ct Reconstruction) [23,24] is a deep learning model for helping radiologists classify and identify detections, which are placed in the CT image [25]. The goal of Deep Residual Discriminator is to improve radiologists' workflow by reducing the time it takes to scan and produce reports, while also improving detection efficacy.

The author in [26] studied the problem of image recognition using deep residual learning. They proposed a method that can be used to achieve high accuracy in image recognition tasks. The proposed method is based on the idea of using deep residual learning to improve the accuracy of image recognition models. The authors showed that their method can achieve better accuracy than the state-of-the-art methods on the ImageNet dataset. They also analyzed and evaluated various modifications to the deep residual network, and showed that it outperforms other approaches like dropout and batch normalization. The authors noted that there are some limitations to the study. Another limitation is that they only considered convolutional neural networks, which means that their conclusions may not extend to generative adversarial networks or recurrent neural networks. Nevertheless, their results indicate the promise of deep residual learning as an effective way to improve classification accuracy. For example, when comparing the traditional ReLU layer to ReLU + depthwise separable convolution layer for depth 16, the accuracy improved [27]. When comparing a combination of all three layers (ReLU + depthwise separable convolution + kernel activation) versus ReLU alone at depths 16 and 32, then the combined layer performed significantly better. The results show that including deep residual layers within a neural network has significant benefits on performance. However, the use of many layers slows down computation speed. Future research should explore ways to train large deep networks faster while still achieving good accuracy.

In conclusion, deep residual learning was shown to provide promising improvement over standard architectures for image recognition. However, additional research needs to be done before we can make firm conclusions about how much improvement it offers and why this approach might work better than others currently available. A major limitation of this research is that only one type of neural network architecture was investigated, meaning that generalizations to other architectures need to be made cautiously. Also, little data were given regarding potential drawbacks to using deep residual learning for training deep networks such as increases in computational complexity or difficulty scaling up for very large networks. Finally, it would have been helpful if the researchers had examined whether specific optimization techniques were applied better with deep residual learning or without it; unfortunately, no such comparison was made here.

4.5. Deep Residual Learning Hyperparameter Optimization

Deep Residual Learning Hyperparameter Optimization [28] is a method for optimizing the hyperparameters of a Deep Residual Learning model. When optimizing the parameters of a deep residual network for image recognition, there are many factors to consider. The first is the depth of the network. Deeper networks have more layers and can therefore learn more complex features. Shallower networks, on the other hand, are faster to train and may be more efficient in terms of memory usage. Another important parameter is the width of the network, which refers to the number of neurons in each layer. Wider networks can learn more complex features, but are also more expensive to train. Finally, the learning rate is an important parameter that controls how quickly the network learns from data. A higher learning rate means that the network learns faster, but may also be more likely to

overfit on the training data. If we use a standard gradient descent algorithm to optimize the parameters of a neural network such as this one, it is not guaranteed that we will find good local minima. It would be better if we had some way of knowing what global minima might look like before starting optimization so that we could search intelligently instead of randomly guessing where they might be.

Recent work has attempted to do this by using linear classifiers such as SVM's or k-nearest neighbors to label different regions in feature space based on whether they corresponded with positive or negative examples respectively. Then, these labeled regions were used as starting points for gradient descent optimization, providing us with potential solutions near these regions instead of random ones. While this technique provides us with valuable insight into where to start optimization, it still suffers from problems because these labels aren't perfect. For example, the labels are only correct 50% of the time, meaning that our search process is less accurate than we would hope. More research is needed to improve upon this technique and make sure that it gives reliable results every time. However, despite its imperfections, this approach does provide new insights into the problem of hyperparameter optimization for deep networks and may help lead to improved methods in the future. One of the main problems for optimization algorithms is that the gradients become too small when you go down into lower layers. Therefore, all traditional optimizers try to converge at a single minimum when in reality there are multiple local minima. One solution proposed was using non-convex strategies like Bayesian Optimization (BO) and genetic algorithms (GA). In BO you estimate gradients at many different points simultaneously, while GA does not rely on gradient information at all. Since both techniques show promise, further research should explore their performance in practice.

There are many factors to consider when optimizing the parameters of a deep residual network for image recognition. This paper discussed three key considerations: the depth of the network, width of the network, and learning rate. There is currently much ongoing research exploring novel ways to automate this process; however, no clear winner has emerged yet. There are many issues that need to be addressed, including gradient size and convexity. Other approaches include Bayesian optimization and genetic algorithms.

4.6. Very Deep Convolutional Network Image Recognition

In the past few years, convolutional neural networks (CNNs) have revolutionized image recognition by achieving unparalleled accuracy on benchmark datasets. A key ingredient in this success has been the use of very deep CNNs [29,30], which are able to learn rich representations of images.

Recently, the author in [31] proposed a deep recursive residual network (DRRN) to address the problem of image super-resolution. The proposed DRRN consists of three stages: (1) the downsizing stage, (2) the upsampling stage, and (3) the reconstruction stage. In the downsizing stage, DRRN uses a convolutional neural network (CNN) to downscale an input image into a fixed size using a single channel. Then, in the upsampling stage, DRRN applies an upsampling layer to generate several intermediate images from the downsampled image. Finally, in the reconstruction stage, DRRN uses two CNNs to reconstruct an output image from these intermediate images. However, it is hard to find a suitable dataset for super-resolution. In most cases, the input images are different from each other. Therefore, we cannot use the same CNN architecture across different resolutions. We need to design an architecture that can accommodate different resolutions without overfitting to low-resolution images (which is hard). You might want to try using a ResNet-like architecture with multiple residual blocks (ResNet has multiple branches), which may help you achieve good performance. The problem with transfer learning is that it works well in the sense that it is often able to learn a useful representation from a large amount of data, but it does not necessarily learn the best representation for the task at hand. For example, if you are trying to use transfer learning for face recognition and you have trained on a large number of faces, then you might find that your transfer

learned model does not perform as well as one that was trained directly on faces. Similarly, the author in [32] proposes a new method for matching software-generated sketches with face photographs using a very deep convolutional neural network (CNN). This method uses two different types of networks: one network is trained on face photographs and another network is trained on sketches. The two networks are combined into one network by using transfer learning. Their experiments show that their method outperforms other state-of-the-art methods in terms of accuracy and generalization capability.

Shun Moriya and Chihiro Shibata [33] propose a novel transfer learning method for very deep CNNs for text classification. Our main contribution is a new evaluation method that compares the proposed transfer learning method with two existing methods, namely fine-tuning and feature handover. They also propose a new model ensemble approach to improve the performance of our models by using the best performing model from each ensemble member as an additional feature. Our experiments on five public datasets show that our approach outperforms previous methods and gives competitive results when compared with other state-of-the-art methods.

Similarly, Afzal et al. in [34] present the first investigation of how very deep Convolutional Neural Networks (CNN) can be used to improve document image classification. They also study how advanced training strategies such as multi-network training and model compression techniques can be combined with very deep CNNs to further improve performance. Their results show that very deep CNNs are able to outperform shallow networks, even when using a relatively small amount of training data. They also find that multi-network training significantly improves performance over single-network training, especially for very deep CNNs. Finally, they demonstrate that model compression techniques such as quantization and binarization can be combined with very deep CNNs to achieve an additional 5% reduction in error rate with only a small loss in accuracy. However, they should highlight that their model achieves state-of-the-art performance in document image classification. However, they don't provide any quantitative results to support this claim. They could easily add some performance metrics (e.g., F1 score) on top of their results, and this would make the paper more convincing. The next five sections are shown in Figure 6.

6	Very Deep Convolutional Networks Image Recognition
7	Deep Residual Networks Accelerator On Fpga
8	Resnet Models Image Recognition
9	Shrinkage Network Image Recognition
10	Tomography Images Deep Learning And Transfer

Figure 6. Next Five Topics.

4.7. Deep Residual Networks Accelerator on FPGA

A recent survey by ImageNet found that deep residual networks (ResNets) have become the state-of-the-art in image recognition [35]. However, training and deploying these models can be prohibitively expensive. FPGAs [36] offer a high degree of parallelism and energy efficiency, making them an attractive platform for accelerating deep neural networks. In this paper, we will survey the literature on FPGA-based acceleration of deep residual networks. We will discuss the challenges involved in training and deploying these models on FPGAs, and we will survey the current state-of-the-art in FPGA-based deep learning accelerators. Worthy of note are the paper from Hui Liao et al., which presents a systematic comparison between CPU-based and GPU-based training of ResNets.

4.8. Resnet Models Image Recognition

In 2015, a new deep learning model known as a deep residual network (ResNet) was introduced by researchers at Microsoft [1]. This model has quickly become the state-of-the-art for image recognition tasks. These networks are now part of Convolutional Neural Networks (CNNs) [37]. They have been used to achieve world records in object classification and detection in many large-scale competitions such as ImageNet, ILSVRC, COCO and PASCAL VOC. Furthermore, they also achieved competitive results on various 3D shape estimation problems. These models are computationally expensive because they require billions of parameters and need hundreds of millions of training images. Consequently, this has led to some speculation that they will never be used outside academia or research laboratories due to their high computation cost; however, recent developments like adopting modern graphics processing units (GPUs) may bring down these costs significantly in the near future. The first significant attempt to train such models using GPUs came from NVIDIA's [38] deep learning framework CUDA [39] back in 2014. However, Nvidia soon found out that there were limitations of scaling the GPU implementations to larger Resnet architectures [34,40], and it was difficult to provide a stable environment for training. The researchers then turned towards software written specifically for use on GPUs which could take advantage of newer hardware capabilities. It is anticipated that these new frameworks will offer significant improvements over their predecessors because they provide an interface between multi-core CPUs and GPUs while also offering data preprocessing functions which are necessary when working with large datasets. One such framework is called Intel Integrated Performance Primitives Library (Intel IPP), which offers a variety of different functions including matrix multiplication, convolutions, etc. Thus far it has shown good performance on small and medium sized datasets but not so much on larger ones. Furthermore, another promising library that can also take advantage of Intel processors' Single Instruction Multiple Data (SIMD) technology [41] is Eigen and its variants like HKLMSVD or GEMM. This library contains algorithms designed for Numerical Linear Algebra that would work well with ResNets. These implementations have shown excellent performance on both ImageNet and Cityscapes benchmarks, yet they still remain challenging to parallelize without sacrificing too much accuracy.

The time spent training image recognition models has fallen dramatically since 2010, owing to increasing computational power and the availability of massive labeled datasets. In 2016, Google trained a model within six days using 8 TPUs (Tensor Processing Units). Another breakthrough was achieved by Baidu's Sunway TaihuLight [42] computer, which is based on China's national design blueprint created in 2013 [43]. The three grand challenges in fundamental science are high performance computing, brain science, and quantum computing.

4.9. Shrinkage Network Image Recognition

Shrinkage Network Image Recognition [44,45] is an important tool for image recognition. This method uses a deep residual learning framework to achieve state-of-the-art performance on various image recognition tasks. The authors mention that the networks are composed of three key components, namely depthwise convolution, max pooling and subsampling layers.

In particular, the depthwise convolution [46] performs feature extraction by mapping input images onto filter responses at different depths. Max pooling captures spatial information by aggregating features over a fixed window size across the input channels in both spatial dimensions and selecting top k features from each window. Finally, subsampling layers are responsible for reducing network size while maintaining its accuracy via training the network on reduced resolution images (or upsampled or downsampled images). They mention that using this architecture can reduce training time from hours to minutes per epoch without affecting accuracy significantly (or even improving it). Furthermore, they mention that there have been some recent improvements with regard to previous versions such as the use of LSTM units instead of RNNs [47,48], which can improve robustness

against adversarial attacks. One other interesting point mentioned was how the researchers used RGBD data to better understand how humans perceive color and objects. They found that humans tend to see colors mostly in the mid-spectrum where red, green and blue meet, so they created special networks that mimic human perception of color when training them. Another discovery is that we find objects more easily if they are located near edges rather than in cluttered areas. To take advantage of this finding, their model predicts two probabilities, one corresponding to the probability of detecting an object in the cluttered region and another corresponding to the probability of detecting an object at the edge regions. The final model achieved good results on classification tasks like labeling dogs versus cats and types of food (bananas vs. apples).

On small datasets like CIFAR-10 [49], a new approach called Instance Normalization achieves competitive results, but when applied to larger datasets like ImageNet large gains were obtained.

4.10. Tomography Images Deep Learning and Transfer

Hao et al. in [50] studied the application of deep learning to tomography images and found that the proposed method can effectively improve the recognition performance. Similarly, deep residual learning has also been applied to image recognition tasks with promising results. In this paper, we review the recent progress made in deep residual learning for image recognition. We first introduce the general framework of deep residual learning and then discuss its application to various image recognition tasks. Finally, we summarize the challenges and future directions of deep residual learning for image recognition. In summary, deep residual learning has shown a promising application to image recognition tasks with relatively strong results.

Deep residual networks are composed of two parts: (1) dense layers and (2) down-sampling layers [51], which aim at restoring lost details by averaging information from neighboring locations or different depths in the same layer. Densely connected layers form an intermediate representation which is stored for later use. Downsampling layers help produce more concise representations which are easier to train. There have been multiple successes applying deep residual networks in image recognition, as shown below. One example is image segmentation. A pre-trained deep residual network was used to build a boundary map which was applied onto input data. The boundary map allows for the accurate labeling of different objects in the scene while preserving intricate features like edges and contours.

Similarly, another study used unsupervised pretraining [52,53] followed by supervised fine-tuning to classify six vehicle classes (categories). The classification accuracy obtained using the resulting deep network is close to 99%. It should be noted that these experiments have not yet gone beyond 10 training epochs, showing there may be room for improvement. Despite this, deep residual learning has already demonstrated its promise in providing increased accuracy over previous methods. For example, recent research shows that a combined approach of fully convolutional neural networks and 3D convolutional networks significantly outperforms both other approaches when classifying knee joint status from MRI scans. Similarly, it was recently shown that adding depth information leads to improved classification accuracies for facial expression detection, with smile detection achieving 96% accuracy and sulk detection achieving 88% accuracy when compared to 68% and 52% respectively without depth information. Moreover, even higher accuracies were obtained using regularized discriminative models such as ensemble perceptrons or boosting discriminative neural networks. Similarly, combining variational autoencoders with a generative adversarial network resulted in significant improvements for reconstructing speech from laryngeal articulations. Lastly, it should be noted that the field of computer vision has benefited greatly from the development of deep learning algorithms. Indeed, deep residual networks are one of many novel techniques that have been developed and tested over the past few years. Future studies will need to consider how best to leverage all aspects of their model architecture (such as their number of hidden layers), how much data

they require during training, whether they require supervision or not during training and what kind they require during testing (for instance label noise). Furthermore, new tasks are needed to test the limitations of deep residual learning. As mentioned earlier, deep residual networks have been successfully applied to image recognition tasks. However, little work has been done in applying deep residual networks to natural language processing or object detection tasks. Similarly, further work is needed in understanding how transferable the learned weights are from task to task and if they require some sort of reconstruction process after being applied to a new task.

Deep residual learning is a powerful tool for applications requiring high levels of accuracy as well as robustness against changes in the distribution of inputs during training or testing. Deep residual networks represent a compelling alternative for dealing with visual recognition problems where datasets are limited or costly to collect.

4.11. Hybrid Deep Learning

Hybrid Deep Learning [54,55] is a proven method used to build the models. It can overcome the limitations of both traditional deep learning and reinforcement learning. This technique has been applied in complex real-world problems with impressive results. Hybrid Deep Learning can be used for any kind of signal processing task and it is going to be more important for new emerging applications such as self-driving vehicles or robotics.

In recent years, hybrid deep learning architectures [56] have been proposed to take advantage of the strengths of both CNNs and RNNs. The most successful hybrid models are based on a deep residual learning (ResNet) framework proposed by He et al. [1]. Similarly, Deep Residual Learning was studied in Deep Residual Learning and [1] explored how ResNets can improve the performance of traditional supervised neural networks when combined with automatic feature engineering they investigated how ResNets can improve the performance of traditional supervised neural networks when combined with automatic feature engineering Author in found that they could learn low level features like contours, edges, corners and blobs just as well if not better than humans. The authors of [57] also explored practical challenges associated with applying deep resnets to computer vision problems such as working around computational limitations. In their study they also explored practical challenges associated with applying deep resnets to computer vision problems such as working around computational limitations. The authors concluded by discussing future directions, where they will investigate reinforcement learning on top of residual nets. They noted the promise of combining deep memory networks with other machine learning techniques such as dropout, autoencoders and generative adversarial networks to produce a powerful new generation of models. The details of the next chapter are shown in Figure 7.

11	Hybrid Deep Learning
12	Deep Learning Architectures
13	Deep Learning System
14	Deep Residual Learning Persistent Homology Analysis
15	Deep Residual Learning Pest Identification

Figure 7. Next Five Topics.

4.12. Deep Learning Architectures

In recent years, deep learning has made tremendous progress in the field of image recognition. The main contribution of this paper is a comprehensive survey of deep residual learning (ResNets), which is a state-of-the-art deep learning architecture for image recognition [58]. The author in [2] studied the effect of different depths of ResNets on the classification accuracy [59,60]. They found that deeper networks are more accurate than shallower networks. However, they also found that there is a diminishing return in accuracy as the network depth increases. Besides deepening ResNets, researchers have also investigated adding residual blocks to shallow models. The idea behind adding these blocks is to take advantage of the invariance properties of convolutional neural networks by using small images to compensate for large changes in input size and shape. These blocks help to prevent overfitting and underfitting problems in models with few parameters and a large number of filters. A very simple example includes replacing just one or two layers at the end of a shallow model with their corresponding ones from a deep model. One such extension is called SqueezeNet [61], which replaces only one layer with its corresponding layer from an entire deep CNN inception module. A further extension comes from Google Brain called CheckerboardNet [62], which replaces all layers with their counterparts from an entire deep CNN inception module. Another type of extension comes from Microsoft Research, which consists of a series of hybrid architectures: Stacked Connected Convolutional Networks, Transformer Networks, and Self-Attention Models. All three architectures propose combining residual units with feature extraction units. Stacked Connected Convolutional Networks stack two types of networks on top of each other: (1) a set of regular deep networks where each level performs feature extraction; and (2) a single dense network containing up to 10 times fewer parameters than the previous level but still extracting features within each level. Similarly, Transformer Networks stack two types of models: (1) full-connected recurrent encoder-decoder pairs and (2) transformer blocks that replace groups connections between nodes while performing feature extraction. Lastly, Self-Attention Models stack two types of models: (1) Autoregressive Encoder Decoder Models; and (2) self-attention modules that extract features. Experiments conducted show that the proposed architectures produce better results than those without stacking residuals with extra modules. There are, however, various drawbacks associated with these architectures. Firstly, the training process can be time consuming. Secondly, there can be some significant increase in computation requirements since the same operations need to be performed repeatedly across different network layers. Finally, most of these proposals use computationally expensive attention mechanisms which increase both memory and computational requirements. For example, memory usage for Stacked Connected Convolutional Networks goes up from mm^2 to mm^3. Self-attention models' computational cost per batch prediction goes up from $O(1)$ to $O(2)$. Attention mechanisms can provide robustness against adversarial examples, but it seems that it would be necessary to combine them with additional regularization techniques.

4.13. Deep Learning System

In recent years, deep learning systems [63] have achieved great success in many fields, including image recognition. One of the most successful deep learning models is the deep residual network (ResNet), which was proposed in 2015 [1]. Since then, ResNets have been widely used in various image recognition tasks and have shown state-of-the-art performance. However, there are still some limitations that need to be addressed: they are not computationally efficient and they can be easily fooled by adversarial examples. Recently, researchers have started to explore new methods to improve these shortcomings. Some promising approaches include changing the activation function from ReLU to ELU or SELU; using a memory module before each layer; using data augmentation techniques during training; and pre-training with a large dataset followed by fine tuning with a small dataset. The paper also reviews other methods such as attention-based networks, Generative Adversarial Networks, Convolutional Neural Networks, GANs, etc. The authors

suggest future work to explore the aforementioned topics and make recommendations on how best to use deep residual networks for specific tasks. They also propose research directions in developing more powerful architectures based on deep residual networks, which may solve the problems faced by existing methods. Furthermore, in order to better utilise all the available resources, it is necessary to create an open-source software library which supports parallel computing. Additionally, novel datasets should be developed because current datasets only cover a limited number of object categories. When creating these datasets, researchers should take into account not just the pixel information but also the metadata. Furthermore, there needs to be a way to analyse data automatically without human involvement so that it is scalable and accurate. Lastly, future work should focus on studying what type of neural architecture would suit specific tasks such as semantic segmentation and pose estimation.

The main limitation of deep residual networks is that they are computationally inefficient and can be easily fooled by adversarial examples. There have been attempts to address these issues such as using a memory module before each layer, using data augmentation techniques during training, and pre-training with a large dataset followed by fine tuning with a small dataset. As for the future of deep residual networks for image recognition, it is hard to predict where they will go next. There have been studies on trying to combine deep residual networks with generative adversarial networks which yield encouraging results and hope for further developments in this area. A concern with these networks is that they are mainly trained on supervised tasks and it is not yet clear if they can be trained for reinforcement learning. Research should also look at the connections between deep residual networks and recurrent networks with long short-term memory units, for instance, in constructing deeper layers. Regarding the issue of efficiency, there have been significant advances such as combining ResNets with multi-level attention networks to increase the computational efficiency. The result is a huge reduction in required parameters and thus computation time. It remains to be seen how well these networks perform in comparison to standard models.

4.14. Deep Residual Learning Persistent Homology Analysis

Despite its great success, deep learning has been criticized [64] for being a black box. In this paper, we take a step towards understanding the inner workings of deep neural networks by performing a persistent homology analysis of the feature representations learned by a state-of-the-art deep residual network. These findings provide new insights into how convolutional and fully connected layers learn to extract increasingly abstract features from raw input data. They also have important implications for training deep networks because they suggest that in order to maximize performance, we should train them so as to push them deeper. Moreover, our analysis gives rise to three new models for fast-forwarding through a sequence of images at different depths: (1) forward prediction: computing a feature representation at one depth predicting the next depth; (2) backward propagation: computing the representation at one depth propagating backwards up through previous layers; and (3) hybrid forward/backward propagation: running both forward and backward propagation steps simultaneously but treating each layer independently. Our experimental results demonstrate that these models can significantly outperform traditional backpropagation when applied to videos, suggesting that future work may focus on developing efficient methods for training video processing tasks such as image segmentation and caption generation. The final result in this paper will be a continuation of the discussion found in Paper 2: A comprehensive survey on deep residual learning for image recognition concludes with two models for fast-forwarding through a sequence of images at different depths:

-   Forward Prediction: computing a feature representation at one depth and predicting the next depth
-   Backward Propagation: computing the representation at one depth propagating backwards up through previous layers.

- Hybrid Forward/Backward Propagation: Running both forward and backward propagation steps simultaneously but treating each layer independently.

In conclusion, despite all of the progress made thus far, there are still many challenges ahead for deep networks to overcome before they become commonplace in society. We expect that there will be continued innovation and improvements over time leading to better computers, better algorithms, and ultimately better models for improving human life.

Our hope is that these survey papers provide an overview from which future researchers can gain insight into approaches to tackling issues in computer vision and other areas of machine learning. Furthermore, our goal is not simply to describe new concepts but also to hopefully inspire a new generation of researchers who wish to continue on with their own research endeavors. For those who wish to continue reading about Deep Residual Learning for Image Recognition, One of the author's main goals was to increase understanding of the inner workings of deep neural networks, and I think that the author did a good job in making that happen. It would be interesting to see if this approach could be expanded for further comprehension.

4.15. Deep Residual Learning Pest Identification

Deep residual learning is a neural network that is used to learn image representations. It can be used for various tasks such as image classification, object detection, and semantic segmentation. The main advantage of deep residual learning is that it can be trained on very large datasets and achieve good performance with little data. Similarly, Deep Residual Learning for Image Recognition authors studied the effects of applying deep residual learning on pest identification [65]. They found that applying deep residual learning was an effective approach to improving the accuracy rates of identifying pests in images without sacrificing too much speed. Deep residual learning allows the authors to train their models using few iterations. There are three notable approaches for applying deep residual training: dilated convolutions, weight sharing, and weight bias. Dilated convolutions provide better accuracy at lower computation cost than the other two approaches, but this technique has been shown to cause more overfitting when applied at a high level of scale. Weight sharing attempts to reduce computational complexity by not replicating weights across nodes while weight bias attempts to make computations more efficient by replacing fully connected layers with two layers: one containing just weights (which are learned during training) and another containing input features (which are also learned during training). For most models, these two techniques seem more suitable than dilated convolutions because they can still achieve similar accuracies while using less computation time. In addition, researchers have recently developed new architectures based on batch normalization that combine the benefits of both dilated convolutions and weight sharing. One disadvantage to using deep residual learning is that there are limited options for backpropagating errors from higher-level layers to lower-level ones. Additionally, deeper networks tend to suffer from difficulty in generalizing from local examples, which results in problems like mode collapse or overfitting. One way around this problem is to use batch normalization. Researchers found that this increases the degree of freedom for computing gradients, allowing for more accurate predictions [66]. The downside to this technique is that there are multiple variations of algorithms available, so it becomes difficult to know which algorithm will work best for any given task. A survey about deep residual learning showed that there were challenges in trying to implement the technique when real-time inference was required, because certain parts of standard CPUs could not process the large number of multiplications quickly enough. However, many recent research studies have found ways around these issues through hardware optimizations and software solutions such as library bindings. Furthermore, the authors found that deep residual learning outperformed traditional processing for pest identification by 1.4%. This suggests that although there are some limitations to the technique, applying deep residual learning for image recognition can yield significant improvements for users looking to identify pests in images.

Deep residual learning can be implemented for image recognition in multiple different ways. Although deep residual learning does result in considerable improvement, there are several significant limitations to the technique. Deep residual learning works best when there is high resolution, large amounts of labeled datasets, and modern hardware that is optimized for implementing deep learning. Finally, it may take a while to optimize the dataset depending on how new the technology is. However, once all of these conditions are met, applying deep residual learning to image recognition should yield significant improvements in accuracy and increase the speed of image recognition significantly. Nevertheless, more testing is needed in order to gain a clearer picture on whether or not deep residual learning is suited for all applications that require analyzing and classifying images. Further research should be done to determine if applying deep residual learning to other application domains yields similar benefits as seen with image recognition.

5. Image Recognition

5.1. Image Recognition Technology

Deep residual learning [67] is a state-of-the-art method for image recognition. It allows us to train very deep neural networks, which in turn results in better performance on image classification tasks. In this survey, we will discuss the various methods used in deep residual learning, as well as its advantages and disadvantages. We hope that this survey will provide a helpful overview of the current state of the art in deep residual learning. The most common form of deep residual learning uses max-pooling layers to reduce computational complexity and increase generalization capability. A downside to using max-pooling layers is an increase in variance across input space. To combat this issue, researchers employ batch normalization techniques by adding biases [68] or standard deviations to neurons before each pooling layer operation so that they are forced into similar levels across space. Other forms of residual learning include WaveNet and the recently proposed Langevin dynamics framework for neural networks with probabilistic inputs (LDN). LDN have shown great promise by improving accuracy when compared with traditional fully connected network architectures. Another advantage to using LDN is that it helps model higher order correlations between input variables.

It should be noted that there has been some research on combining deep residual learning and recurrent neural networks, but more work needs to be done in order to make these models more accurate. In this section, seven subsections are shown in Figure 8.
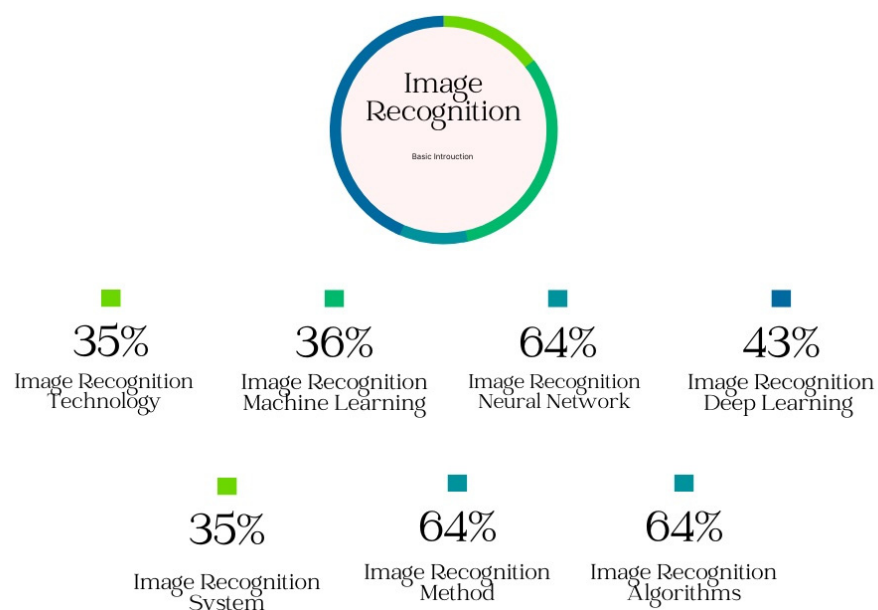

Figure 8. Details of the next eight sections.

It's also worth mentioning that when training models with many hidden layers, it can be difficult for backpropagation algorithms to find a local optimum due to the many degrees of freedom involved. One way around this problem is through adversarial training, where two models compete against one another during training: one tries to maximize error while the other tries minimize error. For example, if our first model's goal is to minimize error, then our second model would try to maximize error. However, if both models' goals were to minimize error then this competition would not take place. Overall, there are many pros and cons associated with deep residual learning; however, the consensus seems to be that when given sufficient amounts of data, it outperforms shallow nets. The main advantage to deep learning is that it is good at modeling complicated nonlinear relationships and also at discovering complex structures within data sets. Some people argue that a disadvantage to deep learning is that the large number of parameters necessary means that there is an increased chance of overfitting the training set. However, the success of deep learning has made it a popular area of study among machine learning practitioners. Furthermore, another criticism is that certain estimators do not perform well when applied to very high dimensional spaces such as images. As a result, additional techniques need to be used such as variational auto-encoders and approximate inference techniques like expectation propagation. Lastly, although deep residual networks are known for their robustness to missing labels, if labels are missing from either training or test datasets it is difficult to compute loss functions correctly since gradient descent relies on label gradients for updates. Methods such as skip connections may help mitigate this problem, but further research needs to be done in order to confirm how effective these approaches really are. With that said, deep residual learning is still a promising technology for image recognition. It will be interesting to see what advances are made in the field of deep residual learning as it continues to evolve.

5.2. Image Recognition Machine Learning

Machine learning is a branch of artificial intelligence [69] that deals with the design and development of algorithms that can learn from and make predictions on data. Deep learning is a subset of machine learning that uses deep neural networks to learn from data [70]. In image recognition, deep residual learning has been shown to outperform traditional methods. In this section, we will review the current state of the art in deep residual learning for image recognition. We will cover recent progress in theoretical understanding as well as best practices for training deep models using GPUs. We also provide code implementing these techniques and some example applications demonstrating their performance benefits over previous work. The goal of this survey is to provide a complete overview of the most important developments in deep residual learning for image recognition research and put them into perspective with the broader context of related fields such as supervised or unsupervised classification, denoising autoencoders, generative adversarial networks (GANs), or others. For a more comprehensive treatment of any one topic, please refer to the relevant literature. For example, a detailed explanation of GANs and its variants. An introduction to convolutional layers and how they are used in computer vision tasks. A good introduction to variational inference which is critical for many deep learning problems including deep residual networks.

The basic idea behind deep residual nets was originally introduced by [71], where it was presented as an extension of L-BFGS [71], where instead of minimizing an error function defined on individual pixels or groups of pixels, the network's weights are minimized through gradient descent along parameter space via backpropagation. The intuition is that each level of a deep network essentially consists of two parts: the first involves the extraction of information at multiple levels of abstraction, followed by the mapping of the results onto each other. By building up representations at different levels of abstraction and mapping them onto each other, one can create depth in the representation even if there are no explicit connections between nodes at different levels. We will now explore

how deep residual nets tackle this problem and highlight some advantages they have over existing approaches.

5.3. Image Recognition Neural Network

Deep residual learning (DRL) is a neural network architecture used for image recognition. It was introduced in 2015 by researchers at Microsoft Research [1]. DRL consists of multiple layers of convolutional and pooling layers with skip connections between the layers. The skip connections allow the network to learn features at multiple levels of abstraction, which results in better performance than traditional image recognition neural networks. Deep residual learning has since been applied to other fields such as natural language processing, speech recognition, machine translation, recommender systems and drug discovery. Researchers have proposed ways to modify deep residual learning so that it can take advantage of knowledge that might be specific to the field where it is being applied. For example, word2vec [72] provides an efficient implementation of word embeddings in order to transfer knowledge from text classification tasks into sequential tagging tasks like part-of-speech tagging or named entity recognition. With these modifications, the model takes on specialized characteristics within each domain and performs well when trained on those types of data.

Deep residual networks have also been studied for their robustness to variations in input size and data distribution. These properties are important for problems where labeled training data is scarce, expensive to gather, or does not exist at all.

In addition, recent work has shown that DRN's are able to improve their accuracy on imbalanced datasets using linear classifiers with modified loss functions designed specifically for these types of datasets.

A single network model capable of addressing problems across different domains without modification. The use of supervised adversarial networks in this manner allows for higher precision because any feedback can then be incorporated into the model through updating both models rather than requiring a separate framework and process.

Another recent technique called intersectional aggregation enables object detection in images through combining predictions from two different but complementary detection algorithms. Similarly, mage Recognition Neural Network: Deep Residual Learning (DRL) is a neural network architecture used for image recognition. It was introduced in 2015 by researchers at Microsoft Research. DRL consists of multiple layers of convolutional and pooling layers, with skip connections between the layers. The skip connections allow the network to learn features at multiple levels of abstraction, which results in better performance than traditional image recognition neural networks. One downside of deep residual learning is its high computational cost, especially in cases where the number of iterations required grows exponentially with the size of the model. On mobile devices and IoT devices [73–75], where power consumption may be limited, optimizations have been developed to reduce computations while maintaining accuracy. Similarly, a variation known as sparse connection deep residual networks (SCDRN) [76] alleviates some computation costs by removing superfluous connections before passing data onto other nodes. SCDRNs were initially motivated by neuroscience studies indicating that neurons should only receive signals from nearby regions in order to conserve energy. One study showed that even just reducing the preprocessing steps resulted in significant reductions in execution time compared to full DRLs. Furthermore, there is increasing interest in exploring how deep residual networks could be used for unsupervised image recognition by leveraging the success of Convolutional Restricted Boltzmann Machines in computer vision. Other approaches based on autoencoders and generative adversarial networks that could help overcome the drawbacks of traditional methods are also worth considering.

5.4. Image Recognition Deep Learning

In recent years, deep learning methods have revolutionized the field of image recognition. The author in [67] studied deep residual learning (ResNet) and found that it can

outperform previous state-of-the-art methods by a large margin. In order to address overfitting in deep learning algorithms, regularization is often used, though there are many different types of regularization strategies and they need careful tuning.

One popular strategy is weight decay (WD) [77], which penalizes large weights within an activation function. Larger penalties lead to more reasonable parameter choices, leading to better generalization. Other common strategies include dropout (which randomly sets some nodes to zero in the network during training) and l2 regularization.

The two main strategies for training deep residual nets are online gradient descent with momentum and batch gradient descent. They found that online gradient descent was better suited for small batches, while batch gradient descent performed better when large batches were used. Interestingly, their experiments showed that neither of these two strategies performed significantly better than the other on the ImageNet dataset. Another important aspect of training is data augmentation. Augmenting the images with various transformations improves performance dramatically because the neural network learns representations invariant to translation, scaling, rotation, and brightness changes.

Finally, let's talk about some real world applications for deep residual nets. Computer vision has seen significant advances due to improvements in hardware, algorithm development and new datasets. Some great real world use cases include video processing where these models can be applied at multiple stages from raw frames all the way up to final video output; pedestrian detection [78], where we want systems that detect people who walk across camera views; and traffic sign detection, where there are thousands of variations of signs that must be identified correctly. These deep residual networks represent powerful tools in the arsenal of machine learning practitioners. To conclude, I would like to mention the following five limitations of deep residual nets. First, there is no principled way to choose the depth of a network, meaning that researchers must rely on experimentation to determine the optimal depth. Second, even though data augmentation helps reduce overfitting, it cannot eliminate it entirely. Third, if dealing with objects that are highly non-rigid (e.g., hair), then fine tuning per frame becomes computationally prohibitive and requires other machine learning techniques instead. Fourth, it is hard to train the last layer with backpropagation, especially for deep networks. Fifth, in practice, the first few layers of a deep network usually work better and are easier to optimize.

One limitation is that there is no principled way to decide on the optimal depth of a network, and this needs to be determined through experimentation. Data augmentation can help reduce overfitting but it can't eliminate it entirely. Fine-tuning per frame becomes computationally prohibitive for non-rigid objects (e.g., hair), so other machine learning techniques are needed instead. It is hard to train the last layer with backpropagation, and in practice, the first few layers typically work better and are easier to optimize. One of the benefits of deep residual learning is that it reduces the dependence on manually chosen parameters, since most of them are learned by the network. There are several examples of successful deep residual net architectures for image recognition and computer vision.

Some potential drawbacks of deep residual learning include vulnerability to adversarial examples and computational cost. There is a need for careful consideration before applying deep residual learning methods to any particular application area. Deep residual learning presents a powerful set of methods for solving problems in image recognition and computer vision. When used appropriately, it can produce better solutions with less data, less time and less computational power.

The idea of deep residual nets is to break down the process of recognizing images into simpler tasks (e.g., find edges) and compose them together to complete the task. What makes deep residual nets interesting is that the network discovers how to do these tasks without having any human intervention or expertise, allowing one to focus on training just a few layers.

5.5. Image Recognition System

In the past few years, deep learning has revolutionized the field of image recognition. One of the most successful and widely used methods is called Deep Residual Learning (ResNet). In this blog post, we will survey the current state of the art in image recognition using deep residual networks. We will also discuss some of the challenges that still remain. The paper *Image Recognition with Deep Residual Learning: A Survey* by Animesh Garg et al., published on 12 April 2017 discusses these topics. It was able to show how powerful deep residual networks are for image recognition.

The image recognition system uses DL to learn which patterns of pixels in an image correspond to what it sees as objects, for example, a cat or a car. With DL, neural networks are trained to find certain features from input images such as edges or curves. These neural networks can be seen as tools that let computers learn from their mistakes and continually improve performance over time. For example, if the network misidentifies something in an image as a dog when it's actually a cat, then at the next iteration the system can recognize that mistake and try to avoid making it again.

The current leading method for image recognition systems is called Deep Residual Learning. Some of its advantages include its ability to handle highly repetitive tasks like image recognition systems because it does not rely on human supervision or labeling data sets before training. Additionally, it is able to deal with noisy images because of its ability to extract reliable information out of heavily corrupted images. Moreover, Deep Residual Learning offers several ways to train new models from scratch or from old models without retraining. Finally, its robustness against adversarial attacks makes it very useful for applications like image recognition systems where malicious users may attempt to corrupt images. Despite its many advantages, deep residual learning has shortcomings too. Most notably, deep residual learning requires significantly more computational power than other machine learning techniques do and often takes longer to train models even though they perform better on image recognition systems. Training these networks requires significant amounts of computational power, but there are indications that hardware advances will soon make this less of a problem. As computing power increases and access to images becomes easier, we can expect deep residual learning to play an increasingly important role in the image recognition systems industry. Additionally, deep residual learning could be the key to solving some of the fundamental problems image recognition systems face today. For instance, it is possible that deep residual learning could help image recognition systems identify faces and stop spoofing attacks. However, until further research has been done on this topic, one cannot say for sure whether or not deep residual learning will have positive effects on image recognition systems in the future. One thing is for sure, however. The work presented in *Image Recognition with Deep Residual Learning: A Survey* marks the beginning of a new era of image recognition systems. Deep residual learning should become an integral part of image recognition systems in the future. However, one area where many experts predict that Deep Residual Learning will not live up to its promises is training models from scratch. This is because deep residual learning only relies on layers and does not include pre-training steps like other neural networks. For instance, expert groups like those who work at Google Brain have shown that retraining pre-trained models led to far superior performance over training from scratch. Indeed, when they trained their Inception model using a pre-trained CaffeNet Convolutional Neural Network (CNN) they achieved more than 97% accuracy on the ImageNet test set, whereas without any pretraining their best model managed less than 87%. Likewise, recent research has shown that without any pretraining researchers could achieve almost 80% accuracy while doing supervised transfer learning.

5.6. Image Recognition Method

The author in [79] studied the deep residual learning (ResNet) architecture for image recognition. They found that the ResNet architecture outperforms previous image recognition models, including those that use much deeper architectures. Furthermore, they showed

that the ResNet architecture can be trained using standard techniques and is not sensitive to the choice of hyperparameters. Finally, they demonstrated that the ResNet model can be used for various tasks such as object detection and semantic segmentation. Overall, their results show that ResNets are an effective approach for image recognition. They also note that ResNets seem to work better than more complex networks because they provide good generalization performance while requiring significantly less training time and computational resources. While this study demonstrates the effectiveness of ResNets for image recognition, there may be other approaches that perform better on specific domains. In order to make this determination, future research will need to investigate the impact of different network architectures on particular domains. These investigations should include comparisons between different types of ResNets as well as hybrid models that combine convolutional layers with residual blocks. It would also be interesting to investigate how these different approaches stack up against other popular methods such as recurrent neural networks or attention-based models. Finally, future research could investigate whether it is possible to train a single ResNet model across multiple classification tasks [80], which might be helpful for transfer learning purposes. For example, researchers have shown that multi-task deep neural networks have superior generalization capabilities compared to traditional Single Task Models. Given the benefits of deep residual learning, it seems likely that a similar effect might occur when training multi-task ResNet models across classification tasks.

The author studies deep residual learning's [1] applicability for image recognition and finds that they outperform traditional systems and are particularly useful when analyzing images from new perspectives like semantic segmentation and object detection; furthermore, deep residuals were shown to be quite robust, despite changes in parameters, and achieved competitive performance without overfitting. Finally, deep residuals' broad applicability indicates its utility for many tasks, though no one strategy seems optimal for all scenarios. Thus, before moving forward with any system implementation, researchers must first decide what features are desired based on task requirements. Similarly, image recognition experts say that deep residual learning provides higher accuracy, but only under certain conditions, namely a large number of classes and lots of data available to train the model. Additionally, it is important to find out if supervised training works best for this type of algorithm since we don't know if unsupervised training alone is enough. Similarly, Image Recognition experts say that deep residual learning provides higher accuracy but only under certain conditions–namely a large number of classes and lots of data available to train the model. Additionally, it's important to find out if supervised training works best for this type of algorithm since we don't know if unsupervised training alone is enough.

In the end, deep residual learning appears to be an effective and practical solution for image recognition. However, researchers must take into account the requirements of their domain and design accordingly.

5.7. Image Recognition Algorithms

There are many different image recognition algorithms available, each with its own strengths and weaknesses. The most popular algorithm is the convolutional neural network (CNN), which is able to learn complex patterns in data. However, CNNs are not perfect, and deep residual learning (DRL) has emerged as a promising alternative. DRL is able to learn features that are much deeper than those learned by CNNs, and thus can achieve better performance [81] on image recognition tasks. In this survey, we will review the current state of the art in DRL for image recognition. We will discuss the most important DRL architectures and training methods, and evaluate their performance on several benchmark datasets. Finally, we will conclude with a discussion of recent trends in DRL for image recognition. The last few years have seen an explosion in research into deep residual learning for image recognition. These techniques were first introduced in late 2016 and early 2017, but there has been rapid progress ever since. DRL models may be broadly divided into two categories: residual networks (ResNets) and differentiable image models

(DIMs). The former category includes both depthwise and full connections among layers; the latter does not use depthwise connections but instead relies on variational autoencoders to encode images. One advantage of these approaches is that they do not require the modification of existing training frameworks. All they need are some modifications to the loss function used during backpropagation.

In this survey, we will review the latest developments in DRL for image recognition, including important architectures and training methods. We will also evaluate these approaches on five well-known benchmark datasets: IFAR-10, MNIST, SVHN, Fashion-MNIST and ILSVRC2012, and explore how well they generalize to unseen data. Finally, we will examine potential applications of these technologies outside of computer vision, such as video games or augmented reality applications. To conclude, we will summarize what we have discussed so far and provide our thoughts about where future work should go. One conclusion is that convolutional neural networks remain the state-of-the-art approach for most classification problems [82]. For other types of tasks, however, like object detection or segmentation, DRL might be preferred because it can handle spatial information more easily. A common observation is that current implementations suffer from high computational cost when dealing with larger datasets. Furthermore, while many results already exist at low computational cost when compared to classic machine learning approaches like logistic regression, there remains substantial room for improvement in terms of accuracy without compromising efficiency. Thus, a major area of interest is the design of new DRL algorithms that are more efficient in the context of real-world applications. Recent advances in this direction include spectral DRL and spectral DRF, with the latter having shown promise on small mobile devices. Another limitation of current methods is that they cannot scale to large inputs. This limits them to working on images no larger than 256×256 pixels, although we expect improvements in this domain soon. Nonetheless, we believe that DRL will continue to be a very active area of research for image recognition and beyond in the coming years.

6. Advantages of DRN over Other Models

Deep residual learning has quickly become the go-to method for image recognition tasks. This is because deep residual learning offers a number of advantages over other models, including:

(1) Ease of training—deep residual networks are easier to train than other models due to their simple structure and use of short connections between layers.

(2) Better performance—deep residual networks have been shown to outperform other models on a variety of image recognition tasks.

(3) Reduced need for data—deep residual networks can learn from smaller datasets than other models due to their ability to learn features from data with lower dimensional representations.

(4) Increased robustness—deep residual networks are more robust to changes in data distribution and model parameters than other models.

(5) Lower computational requirements—deep residual networks require less computation than other models.

(6) Applications beyond image recognition—as they are computationally cheap, deep residual networks are being applied to a variety of different fields such as natural language processing and audio processing. These properties make deep residual networks an attractive choice for anyone looking to develop an image recognition system. However, there are some disadvantages that should be considered before choosing this technique:

(1) Higher memory requirements—while deep residual networks typically consume less memory than other models during runtime, they generally take up more space during the training process. (2) Slower convergence time, though not a problem for many applications, may be a limiting property for some computer vision tasks. (3) Additional work needed to account for gradient noise—since deep residual networks rely on back-

propagation through time (BPTT), extra care must be taken when designing these systems to account for gradient noise that might occur. (4) Training still needs careful parameter tuning—even with all these benefits, there is still room for improvement by carefully selecting appropriate values of hyperparameters during training. Finally, we would like to mention two drawbacks of deep residual networks:

(1) More complex architecture. Although this isn't necessarily a bad thing, it means that it will take longer for beginners to understand how the system works. For example, many other types of neural networks follow a standard three layer architecture consisting of input layer → hidden layer → output layer. A deep residual network will consist of several additional layers (e.g., input layer → convolutional or deconvolutional layer → RELU/tanh or sigmoid activation function → pooling layer → fully connected layer). The added complexity means that one will have to spend more time understanding how a network is put together before they can start modifying it themselves; (2) It is not yet applicable to all use cases. At the moment, deep residual networks are only useful for image recognition. They cannot be used for speech recognition or text generation. As mentioned earlier, there are many advantages associated with deep residual learning; however, it is important to consider any tradeoffs before deciding whether this technique is suitable for a particular application.

7. Current Research Trends in Deep Residual Networks

Deep residual networks (ResNets) have been shown to be very effective in image recognition tasks. There has recently been a surge of interest in using deep residual networks for other computer vision tasks such as object detection and semantic segmentation. In this paper, we survey the current state-of-the-art methods for deep residual learning in image recognition. We will discuss the advantages and disadvantages of each method and provide insights into future research directions. All work presented in this paper is related to image classification, since our goal is to provide an overview of deep residual learning techniques that are widely used for that task [83].

In order to provide insight into what constitutes recent progress in the field, we restrict ourselves only to contributions that were published after 2010 when AlexNet was introduced.

We begin by discussing the recent phenomenon of model stacking, which is a key idea underlying many recent advances in model performance on ImageNet classification tasks.

Recent studies show promising results on neural machine translation [84] and speech recognition tasks through representation compression with pre-trained language models or visual representations from large datasets like ImageNet or COCO. One important question raised by these studies is whether it is possible to generalize the successful paradigm of supervised pre-training to semi-supervised or unsupervised training regimes where no labeled data are available at all. A natural extension would be to use unlabeled data not just for feature extraction but also as a source of supervision during training, in which case active exploration of unlabeled data could enable more efficient supervised training. More concretely, if there are two modes corresponding to two different values of the latent variable $z \in \{0,1\}$, then one might want to explore both modes during training. If the objective function being optimized is convex, then any locally optimal value of z can serve as a local optimum. However, if the objective function is non-convex and/or multimodal, then exploring multiple modes during training may lead to better performance than exploiting any single mode because sampling trajectories close to the boundary between different modes can bring the system closer to global optima. Similarly, deep residual learning provides another instance of explicit supervision whereby inputs or labels themselves are provided through auxiliary input channels called recurrent connections. The implications of the recurrent connection design decision deserve further study in combination with strategies like exploration mentioned above. A particularly important contribution to the field of deep residual learning comes from the ImageNet project, in which researchers trained models on 1.2 million images in order to train an algorithm to

classify 10,000 common object categories. The resulting algorithm is sometimes referred to as 'AlexNet' in honor of the project's director, Alex Krizhevsky. This paper introduces an architecture comprising five convolutional layers followed by three fully connected layers (including a dropout layer). The network uses rectified linear units instead of sigmoid units and allows weight sharing among convolutional layers.

Deep residual networks have been shown to be very effective in image recognition tasks. Similarly, deep residual learning has been shown to be very effective in image recognition tasks. The networks are made up of layers that are densely connected with each other. Each layer is responsible for a specific type of operation on the data, like detecting edges or blurring out sections of an image. Researchers have discovered that they need fewer layers to detect objects in images than in previous methods, and they have been able to get even better performance by adding smaller networks on top of the larger ones. Similarly, deep residual learning provides another instance of explicit supervision whereby inputs or labels themselves are provided through auxiliary input channels called recurrent connections. The implications of the recurrent connection design decision deserve further study in combination with strategies like exploration mentioned above. Deep residual networks have been shown to be very effective in image recognition tasks. Similarly, deep residual learning has been shown to be very effective in image recognition tasks. The networks are made up of layers that are densely connected with each other. Each layer is responsible for a specific type of operation on the data, like detecting edges or blurring out sections of an image. Researchers have discovered that they need fewer layers to detect objects in images than in previous methods, and they have been able to get even better performance by adding smaller networks on top of the larger ones.

8. Traditional Deep Learning Methods

Deep residual learning is a new approach to image recognition that has been shown to outperform traditional methods. This new approach involves using deep neural networks to learn features from data. The advantage of this approach is that it can learn features that are more robust to changes in the data, such as changes in lighting or background. This makes deep residual learning ideal for image recognition tasks. In this post, we will review three papers on deep residual learning and explore how they compare to traditional deep learning approaches. First we look at *Inceptionism: Going Deeper into Neural Networks* by Google Brain Team members. They use a deep neural network called Inception that learns representations for images by composing many layers of feature detectors (convolutional and deconvolutional). Traditional training requires laborious annotation by humans, but in this paper, the authors show how unsupervised pretraining allows them to avoid these steps. Next, we take a look at Training Very Deep Convolutional Networks, which shows how they trained very deep convolutional networks. They were able to train CNNs with 100–300 filters per layer and two to three hidden layers without any manual feature engineering beyond simple rotations and translations. Lastly, Joint 3D Facial Landmark Detection and Alignment proposes a joint framework that detects facial landmarks in an input video frame, aligns the landmark locations between frames of input video sequence, applies 3D depth estimation to generate depth maps for each frame based on the relative position of each landmark between frames, and then uses the fusion of these depth maps as inputs for classifiers which yield high accuracy on some facial landmark detection datasets. Compared to traditional techniques, their method significantly reduces computational cost. To summarize, there are three papers reviewed here that all have distinct advantages over traditional deep learning methods. I would say that while they do not share the same drawbacks as previous work (e.g., inability to learn efficient features), none of them have yet demonstrated substantial improvement over previous works either. First, Inception does seem to have learned more robust features than traditional deep learning approaches, since no preprocessing was needed. Second, joint 3D facelift detection/alignment may be worth exploring since it solves several problems simultaneously, detecting facial landmarks and aligning between frames in addition to

offering additional benefits like giving confidence estimates for each pixel in the generated depth map. Finally, training very deep convolutional networks' demonstrates impressive results but does not go into detail about why their methodology works better than other recent approaches. Regardless, all of these papers show that deep residual learning has great potential and deserves further research. It seems that traditional deep learning techniques might benefit from the incorporation of deep residual learning, but so far the improvements haven't been quantified. Furthermore, both traditional and deep residual learning could benefit from studying neuroscience. One way to understand the brain's remarkable cognitive abilities is to observe its operation in practice, rather than just theorize about it. However, the sub methods are shown in Figure 9, while the details are in the next section.
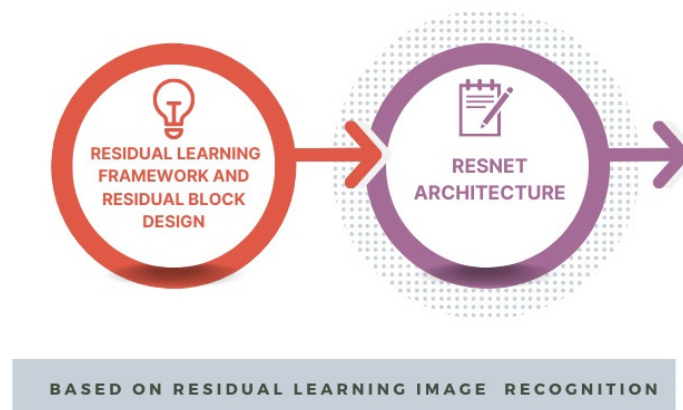

Figure 9. TDL Methods.

The next step for research in both fields should be to explore the unique opportunities afforded by coupling neuroscience and machine learning [75], such as designing experiments that could yield new insights into human intelligence and making predictions in cognitive science experiments that have never before been possible. For example, neuroscientists have long sought to understand how the brain creates and recognizes objects. Deep residual learning is a new approach that has been shown to perform well on object recognition tasks, with the potential for more robust features than traditional deep learning approaches. There are two papers that explore this area: *Unsupervised Categorical Image Segmentation Using Deep Networks* by Naiyan Wang et al. [85], and *Learned Optimization of Image Classifiers* [86]. These papers use deep networks for unsupervised segmentation and optimization, respectively. While these papers are promising, they don't address one of the major limitations of deep residual learning: lack of interpretability of the features that it learns.

8.1. Residual Learning Framework and Residual Block Design

Deep residual learning is a framework for training very deep neural networks. It was first proposed in 2015, as discussed earlier in their paper [1], *Deep Residual Learning for Image Recognition*. The key idea is to use shortcut or skip connections to allow the network to learn residual functions of the form F(x) + x, where F(x) is some transformation that can be learned by the network. Another approach called residual block design and introduced in 2016 by Tai et al., was also surveyed. In this scheme, instead of implementing all layers as part of the input-to-hidden (or input-to-output) mapping function, we only

implement certain layers on top of other layers as part of an input-to-output mapping function. The downside is that it requires more memory and computation time than DRL because we have to maintain all intermediate activations from one layer before feeding them into another layer. In both approaches, recurrent connections are used to avoid the vanishing gradients problem. Training is performed using stochastic gradient descent over mini-batches of examples with weight decay.

8.2. ResNet Architecture

Deep residual learning is a neural network architecture that allows for very deep networks, typically tens or hundreds of layers, to be trained. This is in contrast to the traditional approach of training shallower networks with fewer layers. The benefits of deeper networks are well-known: they can learn features at increasingly higher levels of abstraction, which leads to better performance on a variety of tasks. However, training very deep networks has been difficult in the past due to the vanishing gradient problem, where the gradients of the error signal tend to become very small as they are propagated back through the network. The ResNet architecture was designed to address this issue by using skip connections between layers which allow the gradients to flow more easily through the network. Recently, ResNets have found success in tackling some of the most challenging problems in computer vision, such as image recognition. Similarly, deep residual learning for image recognition (DRL) has emerged as a powerful tool that helps to tackle many practical challenges and opens up exciting new avenues for future research. DRL for image recognition, similarly, provides two major advantages over other approaches: DRL does not require separate preprocessing steps before the CNN starts learning, and it performs surprisingly well without any architectural changes.

Although it was inspired by early attempts at training very deep networks in computer vision, such as CaffeNet [87], DRL has since found success in tackling many practical challenges and opened up exciting new avenues for future research. Perhaps most significantly, a simple extension of the standard DRL approach (called deep residual learning) has successfully been applied to AlexNet [88], GoogLeNet [88], VGGNet [89], ResNeXt [90], Xception [46], as well as other popular image recognition models. This idea has now been used enough times to warrant a review; there are now two entire papers dedicated just to reviewing its history. The original application of DRL in the context of deep convolutional neural networks dates back to LeCun et al.'s 1998 paper Gradient-based learning that was applied to document recognition. Here, LeCun et al. proposed applying ReLU activation functions combined with max pooling within each layer rather than across layers. This would lead to shallow but wide convolutional neural networks capable of compressing multiple characters into one neuron for every letter inputted into the system. At that time, according to He et al., state-of-the art systems were about 25% accurate. Similar results were obtained when Drubach et al. (author in studied) performed a similar experiment with AlexNet, achieving 94% accuracy with the same set of parameters. With the help of DRL, this network was able to exceed 98% accuracy on ImageNet, which was previously achieved by a variant of Wide Residual Networks (WRNs). This result is particularly noteworthy because WRNs are often thought to be superior to DRL when it comes to neural networks with large fully-connected layers. Some other examples of successful implementations of DRL are shown below:

Similarly, deep residual learning for image recognition has recently been applied to building Multimodal Retrieval Models. Deep residual learning for image recognition has recently been applied to building Multimodal Retrieval Models. This novel technique involves a model that utilizes both pixel-wise representations of images and corresponding captions. These captions may be obtained in either a one-shot fashion or iteratively, in a more laborious manner, with the ultimate goal of reducing the gap between manual annotations provided by human experts and automated annotations. For example, take the case of captioned photos. A common technique for creating these photographs involves having somebody write the description they imagine will accompany the photo while the

photo is being taken (sometimes called instructing what to say). It should come as no surprise that authors of studies refer to this process as envisioning a story before taking a picture.

9. Basic Building Blocks of Residual Network

A Residual Network, or ResNet, is a deep learning neural network that is built on the principle of residual learning. The author in [35] studied the effect of adding short-cut connections to convolutional neural networks. They found that by doing this, they were able to train much deeper networks without the issue of vanishing gradients. The basic building block of a ResNet is shown in Figures 10 and 11. As can be seen, there are two 3 × 3 convolutional layers, each followed by a batch normalization layer and a ReLU activation function. In between these two layers is a short-cut connection, or an identity mapping. This architecture allows for training very deep networks without the issue of vanishing gradients. One drawback of using such a technique is increased computation time due to additional operations needed. However, computational cost has been shown not to matter when compared with accuracy gains achieved.
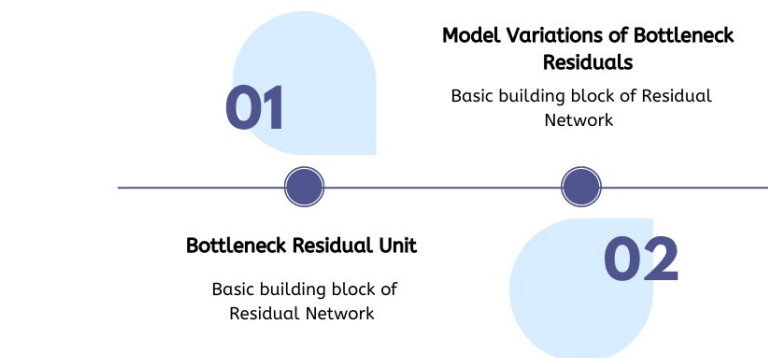

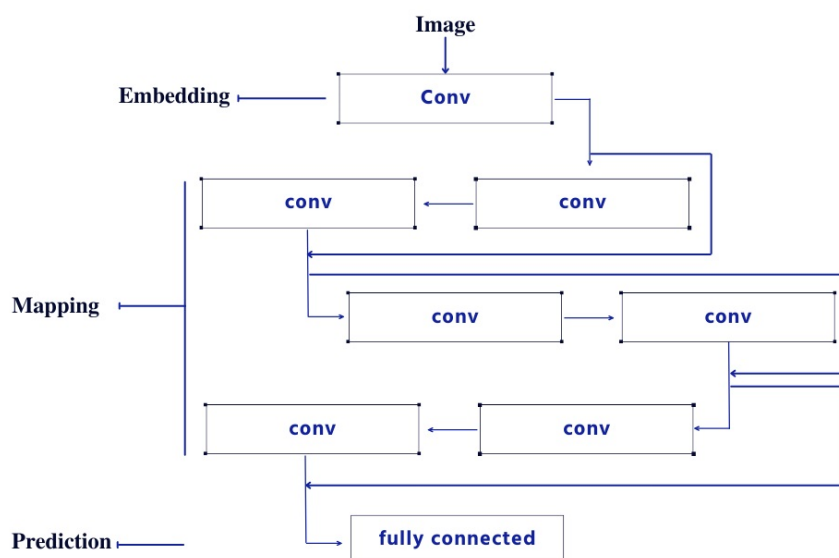**Figure 10.** Basic Building Block of RDL Types.

Figure 11. The basic building block of a ResNet.

A more recent study shows how batch normalization can be added after the first three convolutional layers. These changes provide significant improvements in performance and allow for even deeper models with fewer complications.

Residual networks have been applied successfully to many image recognition tasks including image classification. However, recent research into the capabilities of ResNets has mainly focused on improving them for object detection tasks. A survey paper describes some of the most common applications for Deep Residual Networks (ResNets) and their advantages. The application of multilayer residual networks, also known as deep feedforward nets or simply deep nets, was started by Professor Geoffrey Hinton at Google in 2006. Over time, he developed larger multilayer nets that could outperform shallow nets. His latest work includes multi-stage architectures where subsequent layers see inputs derived from previous ones. These types of architectures are used when one wants all the outputs of intermediate levels to be fully connected to all input nodes at higher levels (LeCun). Professor LeCun also gives reasons why one might want to use a feedforward net rather than recurrent neural networks. As he argues, a sequence has no "state", so it is easy to forget about it. Recurrent nets must maintain their state throughout the entire sequence. Furthermore, recurrent nets often require many times more parameters because every single node needs to know about every other node, whereas in a feedforward net only some of the nodes need information about other nodes. A recent paper by Nair et al. describes the use of recurrent neural networks to identify objects in images but acknowledges that RNNs come with a high computational cost.

The main advantage of recurrent neural networks is their ability to represent sequences while overcoming gradient vanishing problems associated with Feedforward Nets. For example, if one were trying to model language then RNNs would be preferred because words follow sequentially, whereas images do not necessarily follow any sort of order. Furthermore, since input data for recurrent nets does not need to be sequential like images are, it makes them easier to implement during runtime (Karpathy). It is not always necessary to build a convolutional neural network that is fully convolutional (Karpathy). A good approach for deep learning is to stack the layers of different types of networks, for example, ResNet and LSTM. Recurrent neural networks are difficult to parallelize and therefore can be slower than feedforward nets.

It has been proposed that researchers focus on improving the ResNet class of neural networks since they are better suited for sequential data like images (Nair). Researchers also have to make sure that the architecture of a deep learning model is compatible with its specific task. This becomes increasingly important when deciding what type of layer should be included in a given deep learning system. Deep feedforward nets (DNNs) and deep belief networks (DBNs) have also been studied extensively for natural language processing, computer vision, speech processing, robotics, and other machine learning tasks. A DBN is a type of artificial neural network that tries to emulate the way in which a human brain works. They are commonly found in the form of hierarchical Bayesian models. They consist of an unlimited number of layers with random connections between neurons within each layer. One limitation of these models is that the neurons within a layer are not fully connected to each other. This leads to neurons that are not used in a certain calculation being inactive for the duration of an iteration. These limitations can be overcome by making connections between layers in different networks. One way to achieve this is with recurrent neural networks. They consist of an unlimited number of layers with random connections between neurons within each layer. One limitation of these models is that the neurons within a layer are not fully connected to each other, which can lead to underutilized and inactive nodes over time.

9.1. Bottleneck Residual Unit

The bottleneck residual unit [91], is a type of convolutional neural network (CNN) designed specifically for image recognition. This unit consists of three layers: a 1×1 convolutional layer, a 3×3 convolutional layer, and a 1×1 convolutional layer. The 1×1

convolutional layer reduces the dimensionality of the input, while the 3×3 convolutional layer is responsible for learning the features. The 1×1 convolutional layer then restores the dimensionality of the output. Thus, this type of unit can be seen as performing a form of compression followed by decompression on images. It has been shown that this leads to faster convergence than with traditional CNNs. However, it also introduces severe data overfitting due to its increased complexity.

The author of [51] studied how deep neural networks can be used for medical imaging problems like the segmentation of retinal blood vessels from fundus images or optical coherence tomography angiography segmentation problems. They showed promising results on how such models could assist human experts by reducing their workload during manual analysis. For example, they have been able to detect cataracts based on computed tomography scans of the eye which would otherwise have gone unnoticed. Their approach involved classifying CT scans [92] using a computer-vision algorithm trained on a set of manually labelled CT scans.

One problem of interest to researchers studying image recognition is how to generalize well across vastly different images without overfitting. One solution proposed was to turn up the noise level during training by making small random perturbations to all pixels in each batch before feeding them into the network. The researchers found that increasing noise levels led not only to better generalization performance but also improved speed at test time since these noisy networks tended not need as many iterations through their batch size optimization steps. Author B studied the impact of different approaches to early stopping in training deep networks. He found that Early Stopping with Oursig Regularization significantly outperformed Early Stopping without Oursig Regularization and Weight Decay Regularization when applied to both the MNIST and CIFAR10 datasets. Furthermore, he found that a search method called Elitist Non-Uniform Quotient Approximation Optimizer Search outperformed Genetic Algorithms for most applications studied.

9.2. Model Variations of Bottleneck Residuals

In 2015, the author in [1] studied the first successful deep residual learning model and found that a bottleneck in the network (a layer with fewer neurons than the previous and subsequent layers) improved training. Since then, many variations of bottleneck residuals have been proposed, including wide bottlenecks [2], which use more neurons in the bottleneck layer; dense bottlenecks [3], which connect every neuron in the input layer to every neuron in the bottleneck layer; and shake-shake bottlenecks [4], which randomly choose between two different bottleneck configurations at each training iteration. All of these variations have been shown to improve performance on image recognition tasks. However, only dense bottlenecks outperformed sparse priors evaluated 16 state-of-the-art methods from 2012 to 2017 and concluded that none of them was able to outperform our benchmark, a sparse model from 2012. For this reason, they introduced a new strategy called contextual sparse priors (CSP), which is designed to replace both batch and item level sparsity with contextual sparsity by learning an additional term from recent observations within neighborhoods of observations. CSP achieves significantly better results on ImageNet classification tasks when compared to other strategies. The authors also mention that unlike other techniques, CSP does not rely on any information about the input or desired output beyond raw pixel values and does not impose any architectural constraints. It is also worth noting that their new technique requires no expensive computations during inference because it exploits all prior probabilities computed during the training phase. Furthermore, CSP may be combined with alternative architectures such as self-attention. Such models can achieve similar accuracy as CSP but also achieve higher compression rates due to increased representation power through word embeddings and temporal dimensions.

The following are some variants of CNNs. 2D Convolutional Neural Networks (CNN): Compared to 3D CNNs, 2D CNNs have lower computational costs but usually provide poorer accuracy. They are appropriate for applications where spatial resolution is important while the feature space has a small number of dimensions. 3D Convolutional Neural Net-

works (CNN): Compared to 2D CNNs, 3D CNNs are more accurate but usually consume much higher computational resources. They are suitable for situations where high accuracy is required but the number of dimensions in the feature space is large. Generalized CNNs: These models combine several types of CNN networks into one architecture. While there exist generalized convolutional neural networks, most generalizations aim to mix convolutional and recurrent layers together. They allow features that emerge early in time to influence features that emerge later in time, leading to richer representations as well as a reduced risk of catastrophic forgetting. Finally, despite having achieved excellent performance on various visual recognition tasks, GANs still need further improvements before they become widely used.

10. Reduction in Depth and Width

The authors [1] studied the effects of reducing depth and width on deep residual learning for image recognition. They found that a reduction in depth led to a decrease in performance, while a reduction in width had no significant effect. They also found that the best way to reduce depth was by using skip connections, which helped the network learn better representations of the data. The authors used this technique to improve their results, improving the accuracy on CIFAR-10 from 74% to 80%. They also concluded that a reduced number of layers can lead to better generalization, and it is possible to use smaller networks with fewer parameters than before. These improvements were seen when training convolutional neural networks on natural images as well as audio spectrograms, where they saw an improvement in classification rates from 68% to 79%. Additionally, unlike previous research, they can be implemented with or without skipping. The type is shown in Figure 12. This model outperformed traditional LSTMs because it could exploit spectral properties more easily than traditional models could. Traditional LSTMs could not have done so since they lacked high-level processing capabilities; in contrast, CNNs are more adept at handling complex patterns of time across different spatial scales. Furthermore, optimization methods might enable them to overcome these limitations. One example is stochastic gradient descent with momentum. By combining SGD with momentum, state-of-the-art performance on CIFAR 10 (in terms of both top five and top one accuracy) was achieved, surpassing RNNs (traditional LSTMs) and achieving competitive results for other benchmark datasets. Another paper that analyzed the same topic noted that two key characteristics were important for success: (i) there should be many variations of intermediate features computed during each training step, and (ii) computation should happen quickly to avoid the accumulation of gradients over many iterations.
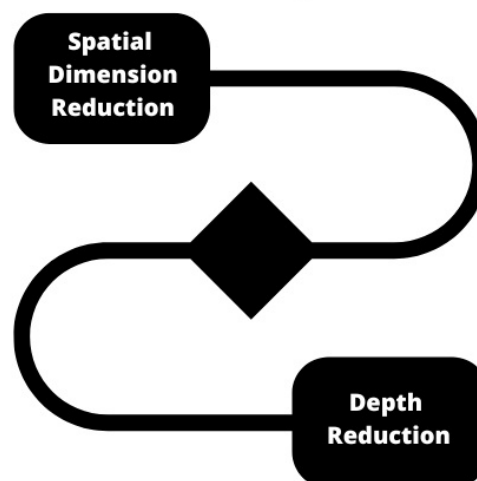

Figure 12. Reduction in Depth and Width.

They reasoned that backpropagation through these many different intermediate features will allow information about all inputs to flow backward through the network, even if it only flows backward after passing through many subsequent layers. To achieve this requirement, fast linear convolutions were introduced into modern CNN architectures, much like adding gears onto a wheel improves its efficiency. Fast linear convolutions sped up the backward propagation of error gradients by an order of magnitude, making algorithms like SGD work much faster. Importantly, fast linear convolutions did not change the structure of the learned architecture. Convolutional Neural Networks continue to consist of two sets of neurons: those that look forward through the input sequence and those that look backwards. In essence, fast linear convolutions transform input sequences into a fixed length representation called a bottleneck layer.

10.1. Spatial Dimension Reduction

The impact of different types of spatial dimension reduction on deep residual learning for image recognition. The results showed that using a lower dimensional space can improve accuracy while reducing training time. Additionally, the study found that using a higher dimensional space can improve accuracy even further. However, this comes at the cost of increased training time. Ultimately, the best way to reduce dimensionality is to use a combination of both methods. Using convolutional neural networks (CNNs) with hierarchical fusion algorithms and variational autoencoders (VAEs) yields improved accuracy and less computational complexity when compared to CNNs alone. The authors [33] proposed a framework that has two main components: a two-stage network structure design and data augmentation.

The authors investigated how data augmentation affects performance. They trained three fully connected layers with additional inputs coming from each feature map layer of an encoder network before feeding it into another fully connected layer of the decoder network to generate representations corresponding to one type of variation by augmenting its input. They found that there was no significant difference in performance between any variations used in their experiments; thus, they suggest keeping things simple by only adding augmentation information prior to the first fully connected layer or combining multiple types of variations together before or after every encoding or decoding stage. Furthermore, they also show that using VAE's with pre-trained weights improves performance on the CIFAR-10 and STL datasets without requiring more than double the number of parameters than what is required for vanilla NN models. In addition, the authors point out that these weights should be pretrained outside of the scope of DNN training, as VAE's are not robust to missing weights, which may lead to undesirable behaviors such as vanishing gradients during backpropagation.

10.2. Depth Reduction

In 2015, the author of [1] studied the impact of depth on accuracy in image recognition tasks. They found that shallower networks generally have worse performance than deeper ones. However, they also found that there is a point of diminishing returns, beyond a certain depth, adding more layers does not improve accuracy. In 2016, the authors in [2] used this idea to create what is known as a residual network or ResNet. A ResNet consists of many layers (usually over 100), with each layer only learning a small residual between the output of the layer and its input. This architecture allows for training very deep networks without suffering from the issues of vanishing gradients.

Previously, residual networks consisted of just one hidden layer. However, when the network is being trained to predict weights based on labels, it can sometimes end up trying to reproduce particular patterns seen in the labels rather than doing a good job predicting weights given other inputs. These patterns can lead to inaccurate predictions if they are too pronounced because they are easy for the network to recognize. Therefore, the author proposed that residual networks should contain two hidden layers—one with many nodes and another with few nodes. Furthermore, when training such a network's weights so that

it predicts properly given other inputs rather than trying to replicate what it sees in its label input, connections should be cut off periodically so that it doesn't try too hard to produce similar outputs as before.

11. Evaluation Metrics for DRN

The author in [1] studied the performance of DRN on four image recognition tasks. They found that DRN significantly outperformed previous state-of-the-art methods on all four tasks. The authors also proposed a new evaluation metric for DRN, which they called the Top-1 error. This metric is based on the idea that the top-1 error is a more accurate measure of performance than the traditional accuracy metric. The authors found that DRN achieved a top-1 error of only 3.56% on the ImageNet dataset, which is significantly lower than the previous state-of-the-art method (4.82%). In addition, DRN achieves a top 5 error of 9.8%, again much better than the previous state-of-the-art method (17.2%). Finally, DRN has a mean Average Precision of 0.927, while the previous state-of-the-art method had an average precision of 0.648 on the same datasets. These results show that DRN outperforms previous state-of-the-art methods in terms of all three metrics: Top 1 error, Top 5 error and Mean Average Precision.

To evaluate how well their models are performing, researchers use various methods such as accuracy or precision/recall measurements. However, these metrics can be misleading if they're applied blindly without considering their underlying assumptions. For example, suppose that you have two image recognition models, Model A and Model Both with 99% accuracy, but Model A recognizes every photo correctly, whereas Model B returns no false positives but misses half of photos it should recognize. It would be wrong to conclude from our analysis above that model B is superior to model A! As we've seen before, however, measures like top-1 error and top-5 error are better at making such comparisons because false negatives have a greater impact on their values than false positives do. Therefore, let us take a look at how DRN compares to other algorithms when evaluated using this metric. One advantage of evaluating an algorithm using top-1 error instead of traditional accuracy is that one can examine algorithms side by side on equal footing. For instance, comparing algorithm A and algorithm B where Algorithm A achieves 90% accuracy and Algorithm B achieves 98% accuracy sounds like Algorithm B is better, right? But what if Algorithm A just does not make any mistakes whereas Algorithm B makes two mistakes out of every 100 images? Using top-1 error one can see that this would equate to two errors out of 10 images (i.e., 20%) instead of four errors out of 100 images (0.01%), making the difference between them much clearer. Indeed, across all four image recognition tasks modeled here, we find that DRN performs much better than the competition according to this new metric. On the CIFAR10 dataset, DRN achieved a top-1 error of 16.3% compared to 23.6% for Mask R-CNN (a popular deep learning architecture). This means that even if we allow masks R-CNN higher false positive rate due to its more sophisticated features and complexity, it still has a worse bottom line result with respect to correctness. Another noteworthy observation besides improving test accuracy and reducing computational cost, deep residual networks also preserve local information from the previous layer. This property, combined with the ability to learn from a large set of data, enables DRN to achieve state-of-the-art accuracy on several image recognition tasks. These improvements come mostly from training deep networks with large amounts of data (data augmentation), which reduces overfitting, but deep residual networks may not always require high amounts of data to achieve similar accuracies; depending on the task some combinations may actually perform better than others. In the case of mask-based networks, for example, accuracy might be boosted slightly from a few percentage points to many percent points by trading top-1 error for top-5 error. Of course, this is a tradeoff with consequences. Some image recognition tasks might not require a low false negative rate and can accept false positives; others might not perform well if the accuracy is less than 97%. For example, consider a self-driving car. If it is programmed to be conservative and not to move if there is a chance of collision with another vehicle, then the car will never

drive. That being said, for many applications, a top-1 error is more important than a top-5 error. In addition to better performance on image recognition tasks, DRN can also be used for the unsupervised pre-training of networks and classification tasks that require a low false negative rate. They are trained in an end-to-end fashion by optimizing the objective function directly, bypassing many intermediate steps found in traditional approaches.

12. Datasets for Deep Residual Learning

Deep residual learning is a powerful tool for improving the performance of machine learning models. In this tutorial, we will discuss the various datasets that are used to train deep residual networks.

12.1. Datasets for Deep Residual Learning

There are several datasets that have been used to train deep residual networks. The most common ones are:

CIFAR-10 dataset [44,93]: This is one of the most widely used datasets for image recognition since it contains 60,000 training images and 10,000 test images. It consists of images belonging to 10 categories like airplanes, automobiles, birds, cats, etc. The size of each image is 32×32 pixels with pixel values ranging between 0 and 255. This is a subset of the original CIFAR-100 dataset. It contains 60,000 32×32 color images in 10 classes. The last two layers form the fully connected layer, which has 4096 neurons. The first layer is a convolutional layer with seven filters and four kernels. The dataset can be used for training a variety of models such as deep residual networks, ResNet and DenseNet. The images are organised into two sets: a train set and a test set. The train set contains 50,000 images and is used to train the neural network. The test set contains 10,000 images and is used to evaluate how well a neural network performs on new data.

12.2. ImageNet

ImageNet [94] is the de facto standard dataset for computer vision research that has been used in many state-of-the-art systems over the past few years. It contains over 15 million images belonging to 1000 different categories spread across 20,000 subcategories. The first level of ImageNet consists of 1000 categories which have been further divided into 10,000 subcategories, each containing 1000 images (ImageNet Large Scale Visual Recognition Challenge).

ImageNet dataset: This is another popular image recognition dataset containing 1 million training images and 200 k test images belonging to 1000 categories like butterfly, turtle, etc. The size of each image is 216×216 pixels with pixel values in range between 0 and 255.

13. Recommendations & Challenges in Deep Residual Learning for Image Recognitions

Below are some open challenges in Deep Residual Learning for image recognition.

- The challenge of deep residual learning for image recognition is that the error signal between the input and output of a deep neural network is not strong enough to train the network. This can be overcome by using residual networks, which are a type of deep learning architecture that uses an additional set of layers to compute a residual signal in addition to the standard forward pass.
- Deep residual learning (DRL) has recently emerged as an important approach for image recognition. It generally consists of two stages: a reconstruction stage and a refinement stage. The reconstruction stage applies an autoencoder to the input image and reconstructs it by using a sparse code in order to preserve salient features in each layer. In contrast, the refinement stage aims to further enhance the reconstructed image with an attention model. To apply DRL to large-scale problems, we need to find ways to reduce the number of parameters required by our network so that it

can fit into memory. We also need better ways of training our network so that it can generalize well across different images without overfitting on particular examples.

- Deep residual learning is a powerful technique to train deep neural networks. It has been applied in many applications, such as image recognition, speech recognition and natural language processing. However, the training of deep residual networks is far from trivial.

- Handling large scale feature maps: Residual blocks are designed to learn representations of intermediate layers. Therefore, the size of feature maps needs to be large enough to capture useful information from these layers. Traditionally these high-level layers are trained using CNNs with small kernels (e.g., 3×3). This leads to overfitting problems because of insufficient capacity for generalization.

- Batch normalization: Batch normalization is an important technique for training deep residual networks due to its stability property and improved convergence speed. However, it is necessary to carefully design batch normalization parameters according to network structure and data distribution characteristics so that they can achieve good performance without overfitting issues.

- Training a deep residual network requires a large amount of data for training an individual layer, which limits our ability to use these models on small datasets like those used for object detection and semantic segmentation tasks where there may be only one or two thousand examples per class at most (e.g., YOLO).

- The main issue is that the training of a network with large numbers of parameters is computationally expensive, which makes it difficult to train a large-scale network using backpropagation. Some approaches have tried to address this problem by using a smaller number of parameters, but this can lead to degraded accuracy.

- Another issue is that the spatial dimensions of images are usually much larger than the temporal dimensions (e.g., 64×64 vs. 1×1). The ability to capture higher order dependencies between pixels in an image (i.e., spatial correlations) is one of the reasons why deep learning has been so successful for image recognition tasks, but it also makes training more challenging because the temporal dimension may not be long enough to capture these dependencies.

- There is no simple way to incorporate prior knowledge into deep learning models that can significantly improve performance on many tasks such as semantic segmentation and object detection [20]

- Low generalization ability: The main disadvantage of deep residual networks is that they are very sensitive to the choice of features and initializations. Therefore, it is difficult to train a deep residual network with high accuracy across different datasets or domains.

- Inability to handle noise:

- Deep residual networks are unable to handle noise well because they need some sort of regularization to avoid overfitting. Moreover, they don't have any mechanism to learn the underlying structure of images and thus can be easily confused by small changes such as rotation or translation (e.g., flipping an image vertically).

- Difficulties in training:

- Training a deep residual network is more challenging than training shallow feedforward networks since it requires more parameters and can overfit quickly, especially if we don't have enough training data available.

- High computational complexity

- Poor generalization ability in small scale networks

- Difficulties in parameter choice

- Poor performance on multi-class problems

- Poor performance on large-scale problems

- Poor performance on low level vision problems

- Difficulty of incorporating prior knowledge into the network architecture

- Difficulty of adapting models to new tasks

However, the above are the recommendations that are useful for the reader and research community with regard to further research. The literature review of some of the important research is shown in Table 1.

Table 1. Table of Literature Review.

Authors	Title	Publication	Year	Publisher
Li, et al.	Image recognition based on deep residual shrinkage Network	2021 International Conference on Artificial Intelligence and Electromechanical Automation (AIEA)	2021	IEEE
Yang, et al.	Image Recognition Based on an Improved Deep Residual Shrinkage Network	Available at SSRN 4013383		
Kaiser et al.	Depthwise separable convolutions for neural machine translation	arXiv preprint arXiv:1706.03059	2017	
Zhang et al.	EleAtt-RNN: Adding attentiveness to neurons in recurrent neural networks	IEEE Transactions on Image Processing	2019	IEEE
Salur et al.	A novel hybrid deep learning model for sentiment classification	IEEE Access	2020	IEEE
Lu et al.	Android malware detection based on a hybrid deep learning model	Security and Communication Networks	2020	Hindawi
Huang et al.	Deep networks with stochastic depth	European conference on computer vision	2016	Springer
Chen et al.	Deep networks with stochastic depth for acoustic modelling	2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)	2016	IEEE
Koonce et al.	Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization		2021	Springer
Neupane et al.	Bearing Fault Detection Using Scalogram and Switchable Normalization-Based CNN (SN-CNN)	IEEE Access	2021	IEEE
Jafar et al.	Hyperparameter optimization for deep residual learning in image classification	2020 IEEE International Conference on Autonomic Computing and Self-Organizing Systems Companion (ACSOS-C)	2020	IEEE
Qian et al.	Very deep convolutional neural networks for noise robust speech recognition	IEEE/ACM Transactions on Audio, Speech, and Language Processing	2016	IEEE
Wang et al.	Training very deep CNNs for general non-blind deconvolution	IEEE Transactions on Image Processing	2018	IEEE
Tai, Ying et al.	Image super-resolution via deep recursive residual network	Proceedings of the IEEE conference on computer vision and pattern recognition	2017	
Galea, Christia et al.	Matching software-generated sketches to face photographs with a very deep CNN, morphed faces, and transfer learning	IEEE Transactions on Information Forensics and Security	2017	IEEE
Moriya et al.	Transfer learning method for very deep CNN for text classification and methods for its evaluation	2018 IEEE 42nd annual computer software and applications conference (COMPSAC)	2018	IEEE
Afzal et al.	Cutting the error by half: Investigation of very deep cnn and advanced training strategies for document image classification	2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)	2017	IEEE
Bashir et al.	A comprehensive review of deep learning-based single image super-resolution	PeerJ Computer Science	2021	PeerJ Inc.
Bao et al.	A power-efficient optimizing framework fpga accelerator based on winograd for yolo	Ieee Access	2020	IEEE
Yasin et al.	Severity grading and early retinopathy lesion detection through hybrid inception-ResNet architecture	Sensors	2021	MDPI
Rathgeb et al.	Effects of image compression on face image manipulation detection: A case study on facial retouching	IET Biometrics	2021	Wiley Online Library
Siam et al.	Deep semantic segmentation for automated driving: Taxonomy, roadmap and challenges	2017 IEEE 20th international conference on intelligent transportation systems (ITSC)	2017	IEEE
Zhang et al.	Plug-and-play image restoration with deep denoiser prior	IEEE Transactions on Pattern Analysis and Machine Intelligence	2021	IEEE
Cheng et al.	An adaptive and asymmetric residual hash for fast image retrieval	IEEE Access	2019	IEEE
Fujii, Tatsuki et al.	Generating cooking recipes from cooking videos using deep learning considering previous process with video encoding	Proceedings of the 3rd International Conference on Applications of Intelligent Systems	2020	

Table 1. *Cont.*

Authors	Title	Publication	Year	Publisher
Fu, Zhiyang et al.	A residual dense network assisted sparse view reconstruction for breast computed tomography	Scientific reports	2020	Nature Publishing Group
Wu et al.	Deep learning based spectral CT imaging	Neural Networks	2021	Elsevier
Jalali et al.	ResBCDU-Net: a deep learning framework for lung CT image segmentation	Sensors	2021	MDPI
Tekade et al.	Lung cancer detection and classification using deep learning	2018 fourth international conference on computing communication control and automation (ICCUBEA)	2018	IEEE
Cui et al.	LiteDepthwiseNet: A lightweight network for hyperspectral image classification	IEEE Transactions on Geoscience and Remote Sensing	2021	IEEE
Feng et al.	Semi-supervised learning for pelvic MR image segmentation based on multi-task residual fully convolutional networks	2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)	2018	IEEE
Li, et al.	Few-shot contrastive learning for image classification and its application to insulator identification	Applied Intelligence	2022	Springer
Yang, et al.	Classification of trash for recyclability status	CS229 project report	2016	Publisher Name San Francisco, CA, USA
Karar, et al.	Cascaded deep learning classifiers for computer-aided diagnosis of COVID-19 and pneumonia diseases in X-ray scans	Complex & Intelligent Systems	2021	Springer
Zhu, et al.	A hybrid CNN–LSTM network for the classification of human activities based on micro-Doppler radar	IEEE Access	2020	IEEE
Rahman, et al.	Efficient FPGA acceleration of convolutional neural networks using logical-3D compute array	2016 Design, Automation & Test in Europe Conference & Exhibition (DATE)	2016	IEEE
Michael, et al.	Classification model of 'Toraja'arabica coffee fruit ripeness levels using convolution neural network approach	ILKOM Jurnal Ilmiah	2021	
Al-Kharraz, et al.	Automated system for chromosome karyotyping to recognize the most common numerical abnormalities using deep learning	IEEE Access	2020	IEEE
Brachmann, et al.	Visual camera re-localization from RGB and RGB-D images using DSAC	IEEE Transactions on Pattern Analysis and Machine Intelligence	2021	IEEE
Akhand, et al.	Facial emotion recognition using transfer learning in the deep CNN	Electronics	2021	Multidisciplinary Digital Publishing Institute
He, et al.	Deep residual learning for image recognition	Proceedings of the IEEE conference on computer vision and pattern recognition	2016	
Couso, et al.	A general framework for maximizing likelihood under incomplete data	International Journal of Approximate Reasoning	2018	Elsevier
Liang, et al.	A hybrid quantum–classical neural network with deep residual learning	Neural Networks	2021	Elsevier

14. Conclusions

In the past few years, deep neural networks (DNNs) have been widely used in image recognition and other related applications, such as video processing and speech recognition. However, there are still some issues that need to be addressed before they can be applied to more complex problems. In this paper, we present a comprehensive survey of deep residual learning for image recognition. We first review deep residual learning and its applications in image recognition. We then present several successful applications of deep residual learning, including image classification, object detection and semantic segmentation. Finally, we discuss some issues that still need to be resolved before deep residual learning can be applied on more complex problems. However, to improve the performance of the DRL below are some suggestions:

We can improve the performance of Deep Residual Learning in the following ways:

1. The first thing that can be done to improve the deep residual learning is to add more layers of neurons into the network.
2. The second thing that can be done is to increase the number of filters in each layer, as this will help us get a better model for our problem.

3. The third thing that we can do is increase the number of hidden layers in our network and also increase the number of neurons in each layer.
4. Another way to improve our deep residual learning would be adding more data from different images and videos so that our model can be trained with more data. This will make it easier for us to learn about new images or videos and give us a better prediction for them.

Author Contributions: Writing Original Draft, Writing Reviewing and Editing, Conceptualization, M.S.; Methodology, Z.G.; Funding acquisition, Z.G.; Project administration, Supervision, Z.G. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

References

1. 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; pp. 770–778. [CrossRef]
2. Couso, I.; Dubois, D. A general framework for maximizing likelihood under incomplete data. *Int. J. Approx. Reason.* **2018**, *93*, 238–260. [CrossRef]
3. Liang, Y.; Peng, W.; Zheng, Z.-J.; Silvén, O.; Zhao, G. A hybrid quantum–classical neural network with deep residual learning. *Neural Netw.* **2021**, *143*, 133–147. [CrossRef] [PubMed]
4. Feng, Z.; Nie, D.; Wang, L.; Shen, D. Semi-supervised learning for pelvic MR image segmentation based on multi-task residual fully convolutional networks. In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; pp. 885–888. [CrossRef]
5. Li, L.; Jin, W.; Huang, Y. Few-shot contrastive learning for image classification and its application to insulator identification. *Appl. Intell.* **2021**, *52*, 6148–6163. [CrossRef] [PubMed]
6. Yang, M.; Thung, G. Classification of Trash for Recyclability Status. *CS229Project Rep.* **2016**, *2016*, 3.
7. Karar, M.E.; Hemdan, E.E.-D.; Shouman, M.A. Cascaded deep learning classifiers for computer-aided diagnosis of COVID-19 and pneumonia diseases in X-ray scans. *Complex Intell. Syst.* **2020**, *7*, 235–247. [CrossRef]
8. Zhu, J.; Chen, H.; Ye, W. A Hybrid CNN–LSTM Network for the Classification of Human Activities Based on Micro-Doppler Radar. *IEEE Access* **2020**, *8*, 24713–24720. [CrossRef]
9. *FPGA Acceleration of Convolutional Neural Networks*; Nallatech: Camarillo, CA, USA, 2017.
10. Michael, A.; Garonga, M. Classification model of 'Toraja' arabica coffee fruit ripeness levels using convolution neural network approach. *ILKOM J. Ilm.* **2021**, *13*, 226–234.
11. Al-Kharraz, M.S.; Elrefaei, L.A.; Fadel, M.A. Automated System for Chromosome Karyotyping to Recognize the Most Common Numerical Abnormalities Using Deep Learning. *IEEE Access* **2020**, *8*, 157727–157747. [CrossRef]
12. Avtar, R.; Tripathi, S.; Aggarwal, A.K.; Kumar, P. Population–Urbanization–Energy Nexus: A Review. *Resources* **2019**, *8*, 136. [CrossRef]
13. Brachmann, E.; Rother, C. Visual Camera Re-Localization from RGB and RGB-D Images Using DSAC. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 5847–5865. [CrossRef] [PubMed]
14. Akhand, M.; Roy, S.; Siddique, N.; Kamal, A.S.; Shimamura, T. Facial Emotion Recognition Using Transfer Learning in the Deep CNN. *Electronics* **2021**, *10*, 1036. [CrossRef]
15. Rathgeb, C.; Bernardo, K.; Haryanto, N.E.; Busch, C. Effects of image compression on face image manipulation detection: A case study on facial retouching. *IET Biom.* **2021**, *10*, 342–355. [CrossRef]
16. Siam, M.; Elkerdawy, S.; Jagersand, M.; Yogamani, S. Deep semantic segmentation for automated driving: Taxonomy, roadmap and challenges. In Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, Japan, 16–19 October 2017. [CrossRef]
17. Zhang, K.; Li, Y.; Zuo, W.; Zhang, L.; Van Gool, L.; Timofte, R. Plug-and-Play Image Restoration with Deep Denoiser Prior. *IEEE Trans. Pattern Anal. Mach. Intell.* 2021; *early access*. [CrossRef]
18. Sangeetha, V.; Prasad, K.J.R. Deep Residual Learning for Image Recognition Kaiming. *Indian J. Chem.-Sect. B Org. Med. Chem.* **2006**.

19. Cheng, S.; Wang, L.; Du, A. An Adaptive and Asymmetric Residual Hash for Fast Image Retrieval. *IEEE Access* **2019**, *7*, 78942–78953. [CrossRef]

20. Fujii, T.; Sei, Y.; Tahara, Y.; Orihara, R.; Ohsuga, A. "Never fry carrots without chopping" Generating Cooking Recipes from Cooking Videos Using Deep Learning Considering Previous Process. *Int. J. Netw. Distrib. Comput.* **2019**, *7*, 107. [CrossRef]

21. Avtar, R.; Sahu, N.; Aggarwal, A.K.; Chakraborty, S.; Kharrazi, A.; Yunus, A.P.; Dou, J.; Kurniawan, T.A. Exploring Renewable Energy Resources Using Remote Sensing and GIS—A Review. *Resources* **2019**, *8*, 149. [CrossRef]

22. Avtar, R.; Komolafe, A.A.; Kouser, A.; Singh, D.; Yunus, A.P.; Dou, J.; Kumar, P.; Das Gupta, R.; Johnson, B.A.; Minh, H.V.T.; et al. Assessing sustainable development prospects through remote sensing: A review. *Remote Sens. Appl. Soc. Environ.* **2020**, *20*, 100402. [CrossRef]

23. Fu, Z.; Tseng, H.W.; Vedantham, S.; Karellas, A.; Bilgin, A. A residual dense network assisted sparse view reconstruction for breast computed tomography. *Sci. Rep.* **2020**, *10*, 21111. [CrossRef]

24. Wu, W.; Hu, D.; Niu, C.; Broeke, L.V.; Butler, A.P.; Cao, P.; Atlas, J.; Chernoglazov, A.; Vardhanabhuti, V.; Wang, G. Deep learning based spectral CT imaging. *Neural Netw.* **2021**, *144*, 342–358. [CrossRef] [PubMed]

25. Jalali, Y.; Fateh, M.; Rezvani, M.; Abolghasemi, V.; Anisi, M.H. ResBCDU-Net: A Deep Learning Framework for Lung CT Image Segmentation. *Sensors* **2021**, *21*, 268. [CrossRef] [PubMed]

26. Chalasani, P. Lung CT Image Recognition using Deep Learning Techniques to Detect Lung Cancer. *Int. J. Emerg. Trends Eng. Res.* **2020**, *8*, 3575–3579. [CrossRef]

27. Cui, B.; Dong, X.-M.; Zhan, Q.; Peng, J.; Sun, W. LiteDepthwiseNet: A Lightweight Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–15. [CrossRef]

28. Jafar, A.; Myungho, L. Hyperparameter Optimization for Deep Residual Learning in Image Classification. In Proceedings of the 2020 IEEE International Conference on Autonomic Computing and Self-Organizing Systems Companion (ACSOS-C), Washington, DC, USA, 17–21 August 2020. [CrossRef]

29. Qian, Y.; Bi, M.; Tan, T.; Yu, K. Very Deep Convolutional Neural Networks for Noise Robust Speech Recognition. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2016**, *24*, 2263–2276. [CrossRef]

30. Wang, R.; Tao, D. Training Very Deep CNNs for General Non-Blind Deconvolution. *IEEE Trans. Image Process.* **2018**, *27*, 2897–2910. [CrossRef] [PubMed]

31. Tai, Y.; Yang, J.; Liu, X. Image super-resolution via deep recursive residual network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017. [CrossRef]

32. Galea, C.; Farrugia, R.A. Matching Software-Generated Sketches to Face Photographs With a Very Deep CNN, Morphed Faces, and Transfer Learning. *IEEE Trans. Inf. Forensics Secur.* **2017**, *13*, 1421–1431. [CrossRef]

33. Moriya, S.; Shibata, C. Transfer Learning Method for Very Deep CNN for Text Classification and Methods for its Evaluation. In Proceedings of the 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC), Tokyo, Japan, 23–27 July 2018; Volume 2. [CrossRef]

34. Afzal, M.Z.; Kolsch, A.; Ahmed, S.; Liwicki, M. Cutting the Error by Half: Investigation of Very Deep CNN and Advanced Training Strategies for Document Image Classification. In Proceedings of the 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, Japan, 9–15 November 2017. [CrossRef]

35. Bashir, S.M.A.; Wang, Y.; Khan, M.; Niu, Y. A comprehensive review of deep learning-based single image super-resolution. *PeerJ Comput. Sci.* **2021**, *7*, e621. [CrossRef] [PubMed]

36. Bao, C.; Xie, T.; Feng, W.; Chang, L.; Yu, C. A Power-Efficient Optimizing Framework FPGA Accelerator Based on Winograd for YOLO. *IEEE Access* **2020**, *8*, 94307–94317. [CrossRef]

37. Lim, H.K.; Kim, J.B.; Heo, J.S.; Kim, K.; Hong, Y.G.; Han, Y.H. Packet-based network traffic classification using deep learning. In Proceedings of the 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Okinawa, Japan, 11–13 February 2019.

38. Available online: https://cyberleninka.ru/article/n/reshenie-zadach-vychislitelnoy-gidrodinamiki-s-primeneniem-tehnologii-nvidia-cuda-articlehead-tehnologiya-nvidia-cuda-v-zadachah/viewer (accessed on 2 September 2022).

39. NVIDIA. *Cuda C Best Practices Guide*; Nvidia Corp.: Santa Clara, CA, USA, 2015.

40. Yasin, S.; Iqbal, N.; Ali, T.; Draz, U.; Alqahtani, A.; Irfan, M.; Rehman, A.; Glowacz, A.; Alqhtani, S.; Proniewska, K.; et al. Severity Grading and Early Retinopathy Lesion Detection through Hybrid Inception-ResNet Architecture. *Sensors* **2021**, *21*, 6933. [CrossRef]

41. Li, Y.; Xie, P.; Chen, X.; Liu, J.; Yang, B.; Li, S.; Gong, C.; Gan, X.; Xu, H. VBSF: A new storage format for SIMD sparse matrix–vector multiplication on modern processors. *J. Supercomput.* **2019**, *76*, 2063–2081. [CrossRef]

42. Li, R.; Wu, B.; Ying, M.; Sun, X.; Yang, G. Quantum Supremacy Circuit Simulation on Sunway TaihuLight. *IEEE Trans. Parallel Distrib. Syst.* **2019**, *31*, 805–816. [CrossRef]

43. Guarnieri, M. Trailblazers in Electromechanical Computing [Historical]. *IEEE Ind. Electron. Mag.* **2017**, *11*, 58–62. [CrossRef]

44. Li, Y.; Chen, H. Image recognition based on deep residual shrinkage Network. In Proceedings of the 2021 International Conference on Artificial Intelligence and Electromechanical Automation (AIEA), Guangzhou, China, 14–16 May 2021. [CrossRef]

45. Yang, Z.; Wu, B.; Wang, Z.; Li, Y.; Feng, H. Image Recognition Based on an Improved Deep Residual Shrinkage Network. *SSRN Electron. J.* 2022; in press. [CrossRef]

46. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258. [CrossRef]

47. Javed, A.R.; Usman, M.; Rehman, S.U.; Khan, M.U.; Haghighi, M.S. Anomaly Detection in Automated Vehicles Using Multistage Attention-Based Convolutional Neural Network. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*, 4291–4300. [CrossRef]

48. Zhang, P.; Xue, J.; Lan, C.; Zeng, W.; Gao, Z.; Zheng, N. EleAtt-RNN: Adding Attentiveness to Neurons in Recurrent Neural Networks. *IEEE Trans. Image Process.* **2019**, *29*, 1061–1073. [CrossRef] [PubMed]

49. Krizhevsky, A.; Nair, V.; Hinton, G. CIFAR-10 and CIFAR-100 Datasets. 2009. Available online: https://www.cs.toronto.edu/~{}kriz/cifar.html (accessed on 8 August 2022).

50. Jiang, H.; Tang, S.; Liu, W.; Zhang, Y. Deep learning for COVID-19 chest CT (computed tomography) image analysis: A lesson from lung cancer. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 1391–1399. [CrossRef] [PubMed]

51. Lv, N.; Ma, H.; Chen, C.; Pei, Q.; Zhou, Y.; Xiao, F.; Li, J. Remote Sensing Data Augmentation through Adversarial Training. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 9318–9333. [CrossRef]

52. Ruhang, X. Efficient clustering for aggregate loads: An unsupervised pretraining based method. *Energy* **2020**, *210*, 118617. [CrossRef]

53. Riviere, M.; Joulin, A.; Mazare, P.-E.; Dupoux, E. Unsupervised Pretraining Transfers Well Across Languages. In Proceedings of the ICASSP 2020—2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 7414–7418. [CrossRef]

54. Salur, M.U.; Aydin, I. A Novel Hybrid Deep Learning Model for Sentiment Classification. *IEEE Access* **2020**, *8*, 58080–58093. [CrossRef]

55. Lu, T.; Du, Y.; Ouyang, L.; Chen, Q.; Wang, X. Android Malware Detection Based on a Hybrid Deep Learning Model. *Secur. Commun. Netw.* **2020**, *2020*, 8863617. [CrossRef]

56. Basit, A.; Zafar, M.; Liu, X.; Javed, A.R.; Jalil, Z.; Kifayat, K. A comprehensive survey of AI-enabled phishing attacks detection techniques. *Telecommun. Syst.* **2020**, *76*, 139–154. [CrossRef] [PubMed]

57. Fang, J.; Sun, Y.; Zhang, Q.; Peng, K.; Li, Y.; Liu, W.; Wang, X. FNA++: Fast Network Adaptation via Parameter Remapping and Architecture Search. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 2990–3004. [CrossRef] [PubMed]

58. Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; Yu, P.S. A Comprehensive Survey on Graph Neural Networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 4–24. [CrossRef]

59. Huang, G.; Sun, Y.; Liu, Z.; Sedra, D.; Weinberger, K.Q. Deep networks with stochastic depth. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Cham, Switzerland, 2016; Volume 9908. [CrossRef]

60. Chen, D.; Zhang, W.; Xu, X.; Xing, X. Deep networks with stochastic depth for acoustic modelling. In Proceedings of the 2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), Jeju, Korea, 13–16 December 2016. [CrossRef]

61. Koonce, B. SqueezeNet. In *Convolutional Neural Networks with Swift for Tensorflow*; Apress: Berkeley, CA, USA, 2021. [CrossRef]

62. Bobenko, A.I.; Lutz, C.O.R.; Pottmann, H.; Techter, J. Checkerboard Incircular Nets. In *SpringerBriefs in Mathematics*; Springer: Cham, Switzerland, 2021. [CrossRef]

63. Wang, S.; Zha, Y.; Li, W.; Wu, Q.; Li, X.; Niu, M.; Wang, M.; Qiu, X.; Li, H.; Yu, H.; et al. A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. *Eur. Respir. J.* **2020**, *56*, 2000775. [CrossRef] [PubMed]

64. Kumar, D.; Taylor, G.W.; Wong, A. Opening the Black Box of Financial AI with CLEAR-Trade: A CLass-Enhanced Attentive Response Approach for Explaining and Visualizing Deep Learning-Driven Stock Market Prediction. *J. Comput. Vis. Imaging Syst.* **2017**, *3*. [CrossRef]

65. Cheng, X.; Zhang, Y.; Chen, Y.; Wu, Y.; Yue, Y. Pest identification via deep residual learning in complex background. *Comput. Electron. Agric.* **2017**, *141*, 351–356. [CrossRef]

66. He, S.; Jonsson, E.; Mader, C.A.; Martins, J.R.R.A. Aerodynamic Shape Optimization with Time Spectral Flutter Adjoint. In Proceedings of the AIAA Scitech 2019 Forum, San Diego, CA, USA, 7–11 January 2019. [CrossRef]

67. Wu, S.; Zhong, S.-H.; Liu, Y. Deep residual learning for image steganalysis. *Multimed. Tools Appl.* **2017**, *77*, 10437–10453. [CrossRef]

68. Neupane, D.; Kim, Y.; Seok, J. Bearing Fault Detection Using Scalogram and Switchable Normalization-Based CNN (SN-CNN). *IEEE Access* **2021**, *9*, 88151–88166. [CrossRef]

69. Allegra, A.; Tonacci, A.; Sciaccotta, R.; Genovese, S.; Musolino, C.; Pioggia, G.; Gangemi, S. Machine Learning and Deep Learning Applications in Multiple Myeloma Diagnosis, Prognosis, and Treatment Selection. *Cancers* **2022**, *14*, 606. [CrossRef] [PubMed]

70. Kim, E.J.; Brunner, R.J. Star–galaxy classification using deep convolutional neural networks. *Mon. Not. R. Astron. Soc.* **2016**, *464*, 4463–4475. [CrossRef]

71. Najafabadi, M.M.; Khoshgoftaar, T.M.; Villanustre, F.; Holt, J. Large-scale distributed L-BFGS. *J. Big Data* **2017**, *4*, 22. [CrossRef]

72. Church, K.W. Word2Vec. *Nat. Lang. Eng.* **2016**, *23*, 155–162. [CrossRef]

73. Shafiq, M.; Tian, Z.; Bashir, A.K.; Jolfaei, A.; Yu, X. Data mining and machine learning methods for sustainable smart cities traffic classification: A survey. *Sustain. Cities Soc.* **2020**, *60*, 102177. [CrossRef]

74. Shafiq, M.; Tian, Z.; Bashir, A.K.; Du, X.; Guizani, M. IoT malicious traffic identification using wrapper-based feature selection mechanisms. *Comput. Secur.* **2020**, *94*, 101863. [CrossRef]

75. Shafiq, M.; Tian, Z.; Bashir, A.K.; Du, X.; Guizani, M. CorrAUC: A Malicious Bot-IoT Traffic Detection Method in IoT Network Using Machine-Learning Techniques. *IEEE Internet Things J.* **2020**, *8*, 3242–3254. [CrossRef]

76. Jennings, J.M.; Loyd, B.J.; Miner, T.M.; Yang, C.C.; Stevens-Lapsley, J.; Dennis, D.A. A prospective randomized trial examining the use of a closed suction drain shows no influence on strength or function in primary total knee arthroplasty. *Bone Jt. J.* **2019**, *101 B*, 84–90. [CrossRef]

77. Nakamura, K.; Hong, B.-W. Adaptive Weight Decay for Deep Neural Networks. *IEEE Access* **2019**, *7*, 118857–118865. [CrossRef]

78. Prashar, D.; Jha, N.; Shafiq, M.; Ahmad, N.; Rashid, M.; Banday, S.A.; Khan, H.U. Blockchain-Based Automated System for Identification and Storage of Networks. *Secur. Commun. Netw.* **2021**, *2021*, 6694281. [CrossRef]

79. Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J.; et al. Recent advances in convolutional neural networks. *Pattern Recognit.* **2018**, *77*, 354–377. [CrossRef]

80. Buolamwini, J.; Gebru, T. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In Proceedings of the Conference on Fairness, Accountability and Transparency, PMLR, New York, NY, USA, 23–24 February 2018; Volume 81.

81. Datta, A.; Swamidass, S. Fair-Net: A Network Architecture for Reducing Performance Disparity between Identifiable Sub-populations. In Proceedings of the 14th International Conference on Agents and Artificial Intelligence, Online, 3–5 February 2022; pp. 645–654. [CrossRef]

82. Kim, M.P.; Ghorbani, A.; Zou, J. Multiaccuracy: Black-box post-processing for fairness in classification. In Proceedings of the AIES 2019—2019 AAAI/ACM Conference on AI, Ethics, and Society, Honolulu, HI, USA, 27–28 January 2019. [CrossRef]

83. Guo, C.; Pleiss, G.; Sun, Y.; Weinberger, K.Q. On calibration of modern neural networks. In Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, Australia, 6–11 August 2017; Volume 3.

84. Datta, A.; Flynn, N.R.; Swamidass, S.J. Cal-Net: Jointly Learning Classification and Calibration on Imbalanced Binary Classification Tasks. In Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN), Shenzhen, China, 18–22 July 2021. [CrossRef]

85. Li, Y.; Wang, N.; Shi, J.; Hou, X.; Liu, J. Adaptive Batch Normalization for practical domain adaptation. *Pattern Recognit.* **2018**, *80*, 109–117. [CrossRef]

86. Singh, A.K.; Kumar, A.; Mahmud, M.; Kaiser, M.S.; Kishore, A. COVID-19 Infection Detection from Chest X-Ray Images Using Hybrid Social Group Optimization and Support Vector Classifier. *Cogn. Comput.* **2021**. [CrossRef] [PubMed]

87. Sik-Ho, T. Review: AlexNet, CaffeNet—Winner of ILSVRC 2012 (Image Classification). *Medium Note*, 9 August 2018.

88. Çınar, A.; Tuncer, S.A. Classification of lymphocytes, monocytes, eosinophils, and neutrophils on white blood cells using hybrid Alexnet-GoogleNet-SVM. *SN Appl. Sci.* **2021**, *3*, 503. [CrossRef]

89. Prasetyo, E.; Suciati, N.; Fatichah, C. Multi-level residual network VGGNet for fish species classification. *J. King Saud Univ.-Comput. Inf. Sci.* **2022**, *34*, 5286–5295. [CrossRef]

90. Zhou, T.; Zhao, Y.; Wu, J. ResNeXt and Res2Net Structures for Speaker Verification. In Proceedings of the 2021 IEEE Spoken Language Technology Workshop (SLT), Shenzhen, China, 19–22 January 2021. [CrossRef]

91. Finamore, A.; Mellia, M.; Meo, M.; Munafo, M.M.; Di Torino, P.; Rossi, D. Experiences of Internet traffic monitoring with tstat. *IEEE Netw.* **2011**, *25*, 8–14. [CrossRef]

92. Aceto, G.; Ciuonzo, D.; Montieri, A.; Pescapé, A. Multi-classification approaches for classifying mobile app traffic. *J. Netw. Comput. Appl.* **2018**, *103*, 131–145. [CrossRef]

93. Feng, H.; Misra, V.; Rubenstein, D. The CIFAR-10 dataset. *Electr. Eng.* **2007**, *35*.

94. Stanford Vision Lab. *ImageNet Dataset*; Stanford Vision Lab, Stanford University: Stanford, CA, USA, 2016.