Extended Essay in Computer Science

Image Classification

To What Extent Do Machine Learning and/or Deep Learning Image Classification Algorithms Exhibit Effectiveness in Real World Applications?

Word Count: 4303

Abstract

This investigation explores the extent of which deep learning and machine learning image classification algorithms exhibit effectiveness in real life applications.

Image classification aims to categorize images based on their content, and is an extensively researched topic in computer vision. Over the years, there have been numerous techniques and methods to maximize accuracy and efficiency for this task of image classification. Machine and deep learning approaches have yielded promising results, relying on Convolutional Neural Networks (CNN).

In this paper, we use an image classifier to compare classic machine learning and deep learning models in classifying images. Specifically, the performance of CNN9, AlexNet, VGG11, DenseNet, ResNet-50, and EfficientNet architectures will be analyzed, with the focus on the impact of network depth, through the use of the CIFAR-10 and CIFAR-100 datasets. Each model's accuracy and loss have been assessed, with results varying from 0.7188 to 0.8840 for CIFAR-10, and 0.4770 to 0.7248 for CIFAR-100, depending on various model configurations.

Lastly, waste sorting and air quality estimation algorithms are designed to classify different types of waste, and analyze air quality in pictures. This can be used to recognize if these machine and deep learning models will have similar accuracies in real world scenarios.

Contents

1 Introduction

In recent years, there has been a significant increase in real life applications for image classification, such as helping people organize their photo collections, analyze medical images, monitor traffic, face and object recognition, etc. All of these tasks require precisely labelled large-scale datasets, which include a variety of images, such as dogs, cats, landscapes, and roads.

1.1 Image Classification Algorithm Evolution

The goal of image classification is to predict what class a given input image belongs to. While this may be straightforward for humans, training computers to interpret visual data is a complex problem that has sparked research interest for professionals in this field. Both machine and deep learning techniques have been developed to combat these challenges. Machine learning models rely on local descriptors to identify similarities between images, whereas modern advances in technology have introduced deep learning models, which can automatically extract significant and representative features and patterns from images ("Deep Learning vs Machine Learning | Google Cloud").

The research done on image classifications have consistently gathered attention, as classification results can form foundations for numerous environmental and socioeconomic applications. Scientists have devoted research to develop classification methods and techniques to enhance classification accuracy, but to fully understand these methods, it is necessary to first understand traditional machine learning techniques. LeNet-5 by Yann LeCun in 1998 was the first introduction of Convolutional Neural Networks (CNNs), which are used to recognize handwritten digits, And in 2012, AlexNet achieved results in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) , which were able to demonstrate the power of neural networks for image classification tasks. Later, VGG, ResNet, DenseNet, and EfficientNet were introduced, having innovations like deeper layers, residual connections, and dense connectivity to improve feature extraction and classification accuracy.

1.2 Steps of Image Classification

The steps in image classifications include determining a classification system, selecting training samples, preprocessing images, extracting features, choosing appropriate classification approaches, conducting post-classification processing, and assessing accuracy (Nithyashree V). Factors such as the user's needs, scale of study area, economic conditions, and analyst expertise can influence the selection of data, design of classification procedure, and quality of classification results.

Image preprocessing is the most crucial step in image classification. Preprocessing techniques like denoising and smoothing can lower noise in images, and enhance the quality of the data, which makes it easier for the classifier to identify unique features in the image. To make it easier for the classifier to extract meaningful information, methods such as edge detection, image sharpening, and contrast enhancement, help highlight important features in the image. Additionally, normalizing images to a standard scale/range can ensure that variations in the brightness, contrast and color across images do not influence the classification process, improving the consistency and robustness of the classifier. Moreover, geometric rectification is used in order to correct distortions that are caused by factors such as camera angles, perspectives, and sensor orientation, and ensure that images are uniformly aligned, and that the features are accurately represented, leading to reliable classification results (Fagan and DeFries). By removing these inconsistencies in the dataset, preprocessing helps ensure that the classifier solely focuses on characteristics of the images, rather than extra, unrelated factors. Thus, preprocessing is used to reduce the complexity of classification for the computer, through the simplification of images, or extract the most relevant features, leading to more accurate and efficient classification algorithms.

1.3 Investigation Purpose

In this paper, it is investigated how the differences between machine and deep learning impact image classification algorithm accuracy, and thus affect their real world applications. The inherent difference between machine and deep learning image classification algorithms is their neural network depth, which is the number of hidden layers (layers where computations are done, where each layer is in charge of a specific feature of the image) that data is passed through between the input and output layers. For the sake of this paper, the distincting threshold between machine and deep learning models is the ten CNN-layer mark, where any model with beyond ten neural network layers is classified as deep learning models.

Given that the datasets are divided into different classes, the goal for the image classification algorithm is to predict the corresponding class for each image. To achieve this, machine and deep learning models must be implemented and evaluated; for each system, the model is trained on a subset of images, and is then tested on unseen images to validate performance through accuracy and loss metrics. Specifically, popular and state-of-the-art convolutional neural networks (CNN) architectures are implemented — AlexNet, CNN9, VGG11, DenseNet, ResNet, and EfficientNet — on the CIFAR-10 and CIFAR-100 datasets. This can better help to understand how varying network depth between each model can influence performance metrics and practical real life applications. Through experiments, the aim of this investigation is to provide insights into how the various models perform in the same datasets.

Initially, the hypothesis is that deeper networks can generally demonstrate improved accuracy as opposed to ones with shallower networks, which can be due to their enhanced capacity for feature extraction, but can also require longer training times due to the need of computational resources. And as mentioned in the abstract, waste sorting and air quality estimation algorithms will be tested with the most successful model from the experiment to see if its accuracy will translate into real world scenarios.

2 Methodology

In this section, the specific models for machine learning and deep learning will be introduced, and their unique characteristics and functions will be explained.

2.1 Machine Learning Models

2.1.1 CNN9

As seen in Figure 1., CNN9 is a deep convolutional neural network with 9 layers (hence the name CNN9), designed for image classification tasks. Its main features include multiple layers, with each acting as small filters to capture intricate patterns, and pooling layers that act to reduce spatial dimensions within the image. The model includes activation functions like ReLU for non-linearity, batch normalization for training stability, and dropout for regularization ("6.6. Convolutional Neural Networks (LeNet) — Dive into Deep Learning 0.17.6 Documentation"). Due to the depth of the CNN9 model, it is able to excel in recognizing complex structures within images, making it a suitable machine learning model for large-scale classification tasks.

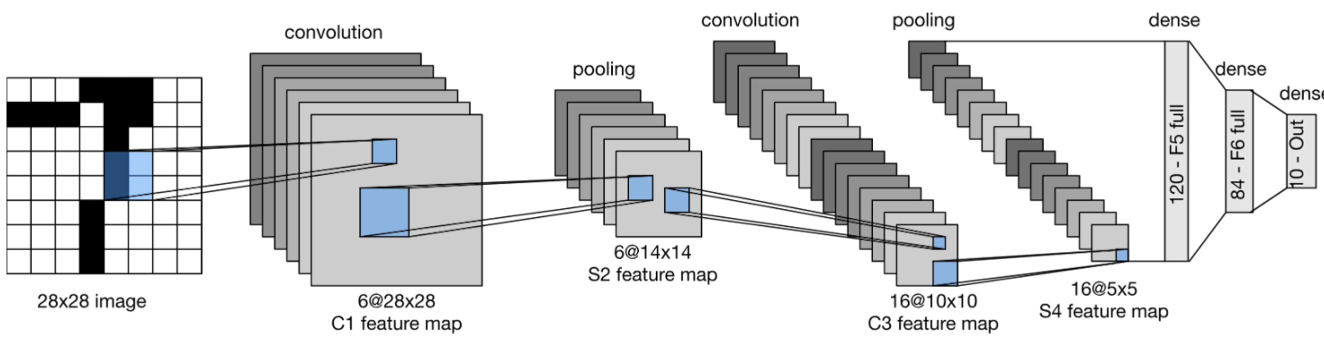

Figure 1. Layers of the CNN9 model. ("6.6. Convolutional Neural Networks (LeNet) — Dive into Deep Learning 0.17.6 Documentation")

2.1.2 AlexNet

The AlexNet model has some key components and innovations that make it different from the CNN9 model. First off, it has eight layers that contain learnable parameters, with five convolutional layers followed by three connected layers. This multi-layered architecture allows for the model to adapt a hierarchical representation of the input image, where it is able to capture increasingly complex features at each of the layers.

To combat overfitting (explained in 3.4), dropout is used in fully connected layers of AlexNet to randomly deactivate neurons during training, which can further reduce overfitting, and improve generalization (Wikipedia Contributors).

2.2 Deep Learning Models

2.2.1 VGG11

VGG is a family of CNN models that became well-known and well-regarded due to their simplicity and effectiveness. Popular variants include VGG11, VGG16, and VGG32, with the numbers referring to the number of layers within the models. The convolutional layers are followed by ReLU functions to introduce non-linearity, and at the end of the convolutional and pooling layers, VGG models have few fully connected layers, which acts as a classifier on extracted features (Anna Alexandra Grigoryan).

Similar to AlexNet, VGG11 and other VGG models incorporate techniques such as dropout to prevent overfitting, which means that dropout layers are added to the fully connected layers to randomly drop neurons during model training, which encourages the network to generalize better (Team).

2.2.2 DenseNet

DenseNet is a CNN model that is designed to tackle the vanishing gradient problem (where gradients used to update the network become extremely small/vanish), and promotes feature reuse. As its name implies, DenseNet features dense connectivity within each block, seen in Figure 2, and each layer receives input from all preceding layers, and passes its output to all subsequent layers. Due to the dense connectivity, information flow is further optimized, enabling more efficient training, and improved performance for the model ("DenseNet").

Due to its unique model and dense connectivity, DenseNet is able to accommodate for a large number of layers, making the model suitable for tasks requiring extensive feature extraction and representation learning.

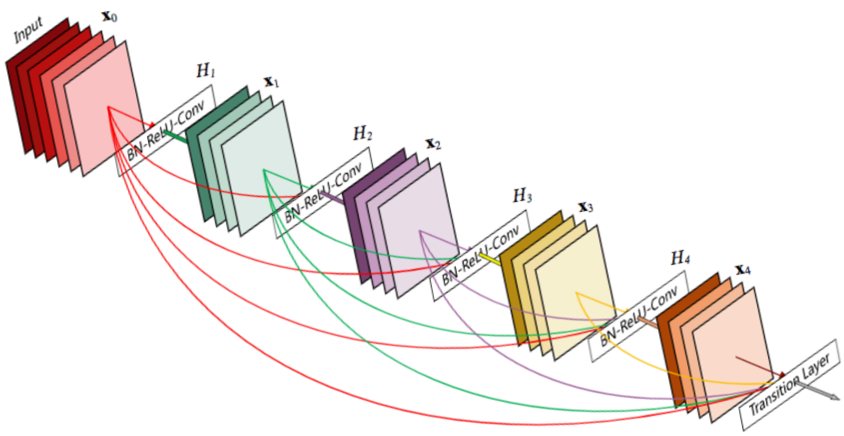

Figure 2. Layers of the DenseNet model, including bottleneck and transition layers. ("Densenet")

2.2.3 ResNet-50

As ResNet-50's name implies, the model is a deep convolutional neural network that contains 50 layers. The model introduces residual learning, which mitigates the degradation problem in deep neural networks, where the degradation problem states that training and testing error will increase as more layers are stacked within the neural network. Seen in Figure 3, residual blocks are connected by shortcut connections that bypass one or more layers, which allows the network to identify mapping, meaning that "residual functions refer to the layer inputs, rather than learning unreferenced functions" (Mukherjee).

With these features, ResNet-50 is able to effectively learn complex features, achieve high accuracy on image classification tasks, and influence several other deep learning models in computer vision.

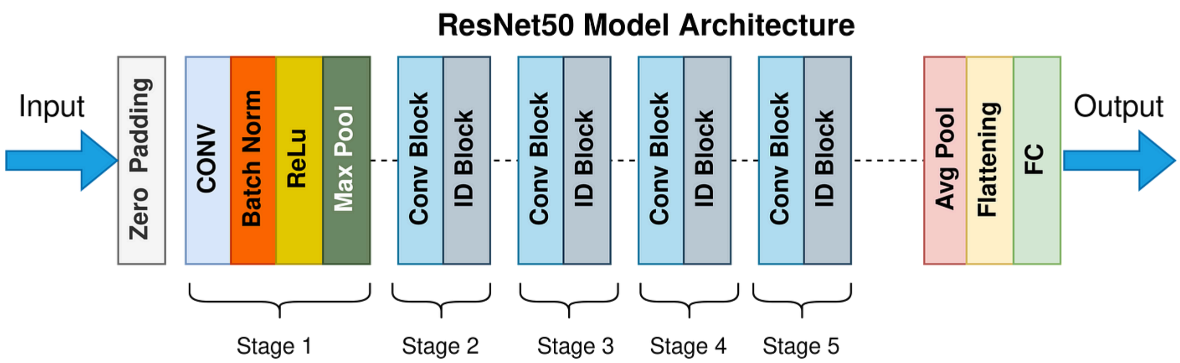

Figure 3. Stages of ResNet-50 model. (Mukherjee)

2.2.4 EfficientNet

EfficientNet is a family of CNNs that was developed by researchers at Google, with the goal to achieve top-grade performance while optimizing computational efficiency. Introduced in a paper by Mingxing Tan and Quoc V. Le in 2019, EfficientNet leverages compound scaling methods that uniformly scale network width, depth, and resolution, using fixed scaling coefficients, such as 1x1, 3x3, etc (Wisdomml). Shown in Figure 4, through this scaling, the network maintains balanced complexity across dimensions, which leads to better performance, with fewer parameters than previous models. EfficientNet has been widely used for image classification tasks due to its accuracy, and reduced need for high-end computational processing power.

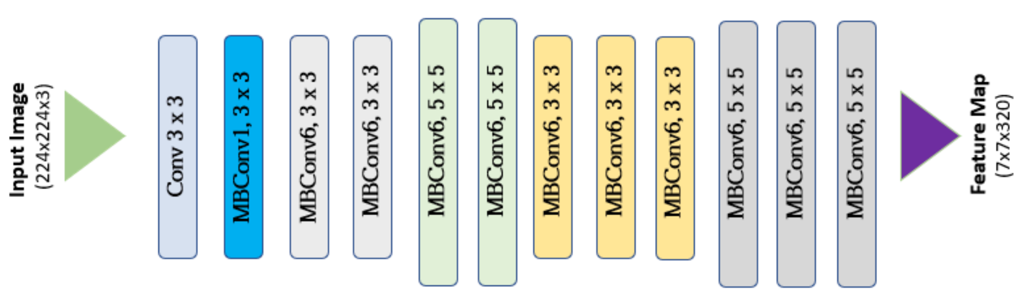

Figure 4. Configuration of EfficientNet Model. (Wisdomml)

3 Experiment

In this section, each dataset that was used will be outlined, and the yielded results from each model for each dataset will be plotted and compared. The experiments will be conducted through a GPU card with 16 GiB of memory hosted on Tencent Cloud, Pytorch 2.7.0, and Python 3.11.0 to implement the various models.

3.1 Datasets and Preprocessing

3.1.1 CIFAR-10

The CIFAR-10 dataset is widely used as a benchmark in the field of computer vision. It comprises 60,000 32x32 color images, which are split into 10 classes, where each class contains 6,000 images. The dataset is then split into 50,000 training images and 10,000 test images. The 10 classes in the CIFAR-10 dataset are shown below:

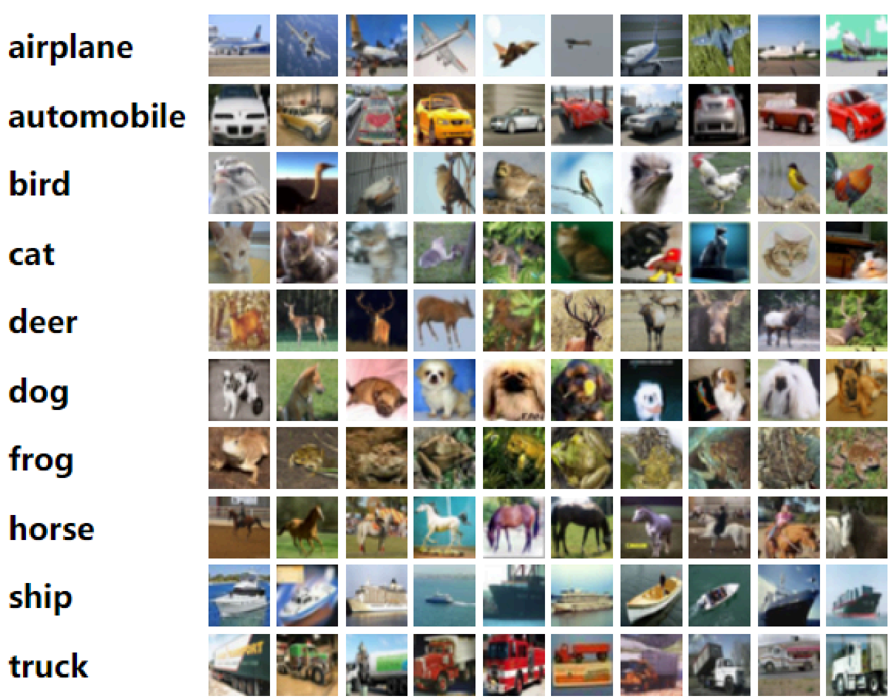

Figure 5. Classes within the CIFAR-10 dataset ("Papers with Code - CIFAR-10 Dataset").

As seen in Figure 5, CIFAR-10 has a diverse set of object classes, ranging from modes of transportation to animals. Researchers will often use this dataset as a basis to compare performance between different models. Since the dataset contains simple and small images, it makes it computationally efficient for experimentation, and can allow researchers to iterate quickly between models, parameters, and ways to optimize each model. Additionally, CIFAR-10 is popular among researchers due to its accessibility, providing a standardized dataset that acts as a benchmark for many image classification algorithms.

3.1.2 CIFAR-100

The CIFAR-100 model is an extension to the CIFAR-10 dataset, and instead of only having 10 classes, the 60,000 images are split into 100 classes. Each class contains 600 images, where 500 are training images and 100 are test images. The 100 classes are split into 20 superclasses (such as the 'orchid' class being in the 'flowers' superclass), so each individual image has a 'fine label', which is the class it belongs to, and a 'coarse label', which is the superclass to which it belongs to. Opposed to the CIFAR-10 dataset, CIFAR-100 serves as a more challenging benchmark for image classification models, where there is a more diverse range of classes.

3.1.3 Data Preprocessing

As stated in section 1.2 of the paper, data preprocessing is an essential step in image classification, since it is able to enhance the quality of input images, as well as standardize and make input images consistent across different models. Firstly, each image's dimension is adjusted to match the input size of 256 x 256 pixels, and normalization is then performed to scale pixel values. To prevent the issue of overfitting (discussed in section 3.4), rotations and flips are applied to select images to augment the data, and Gaussian filters (performing weighted average of surrounding pixels based on Gaussian distribution) are used to remove noise and improve clarity of the training data. These preprocessing steps are done in order to prepare the images for effective training, leading to better performance of the various models.

3.2 Results for CIFAR-10

This section evaluates and compares the performance of the various convolutional neural network models outlined above in both machine and deep learning, on the CIFAR-10 dataset. To reduce uncontrollable variables between models, each model was trained and tested under the same conditions, and the performance of each model is based on its accuracy in classifying the given input images for the 10,000 test images.

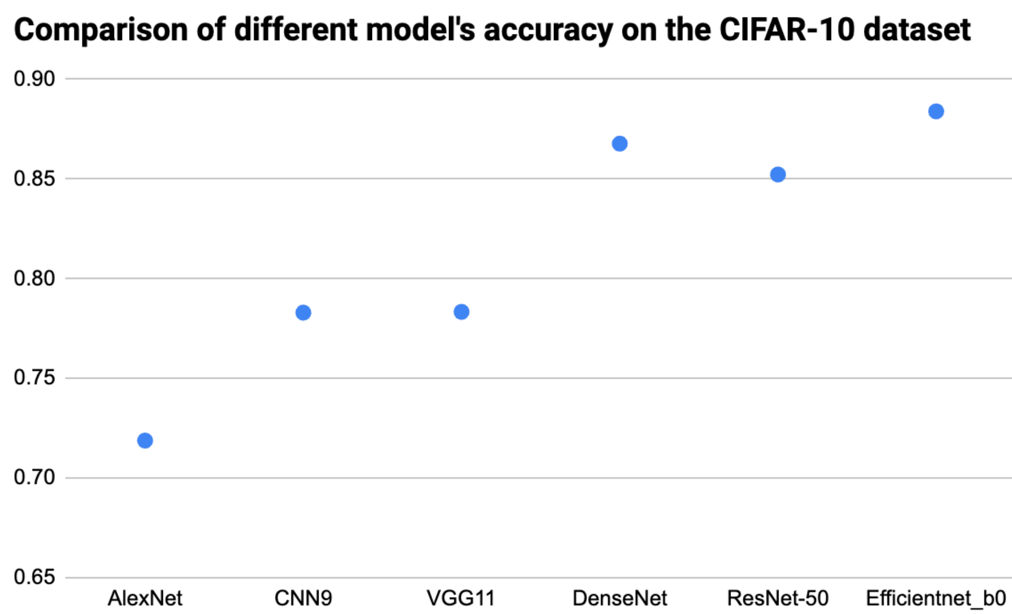

Figure 6. Graph by author displaying the different accuracies of models for the CIFAR-10 dataset

Model:	Accuracy:
AlexNet	0.7188
CNN9	0.7830
VGG11	0.7834
DenseNet	0.8678
ResNet-50	0.8523
Efficientnet_b0	0.8840

Table 1. Table by author representing the precise highest accuracy of each model for CIFAR-10 dataset (green representing machine learning model, red representing deep learning models)

Figure 6 and Table 1 display the performances of the CNN9, AlexNet, VGG11, DenseNet, ResNet-50, and EfficientNet on the CIFAR-10 dataset. For the machine learning models, which are highlighted in green, AlexNet was able the achieve a moderate accuracy of ~0.72, meaning that it was able to correctly classify around 72% of the 10,000 test images, placing them into their respective classes, and also had fast training convergence times due to its simpler architecture. The other machine learning model CNN9, one of the earliest convolutional neural networks, similarly showed low accuracy, reflecting its limited capacity for complex feature extraction. Compared to the deep learning models, the machine learning models both performed with lower accuracy, which could be due to the lack of network depth, or lack of network complexity.

On the other hand, the deep learning models were able to achieve higher accuracies compared to the machine learning models. VGG11 was able to outperform AlexNet and CNN9 due to its deep architecture and simple design, achieving higher accuracy, but with relatively longer training times. With its residual learning, ResNet-50 was able to break the 0.85 threshold, achieving an even higher accuracy compared to VGG11, while also having lower loss, making it more efficient than VGG11. With dense connectivity and promoting gradient flow, DenseNet was also able to reach a high accuracy. Through balancing computational efficiency and multi-scale feature capturing, EfficientNet was able to achieve the highest accuracy of all models at ~0.88.

While the machine learning algorithms and models served as foundational baselines, modern models such as VGG11, ResNet-50, EfficientNet and DenseNet were able to outperform them by quite a significant margin. EfficientNet and DenseNet stood out among the 6 models, achieving a significantly greater accuracy compared to the other models.

Through analyzing the results from the CIFAR-10 model, it is fair to conclude that the machine and deep learning models perform to a comparable level of accuracy, at around ±10%, but the hypothesis set at the beginning of the paper was that deep learning models will have significantly higher accuracies, which is why there will be a second round of experiments with the CIFAR-100 dataset, which should more clearly show the difference in accuracy between machine and deep learning algorithms.

3.3 Results for CIFAR-100

As previously mentioned, the purpose of repeating the experiment for a larger and more complex dataset is to see a more obvious difference between machine and deep learning accuracies, hopefully being able to validate the initial hypothesis.

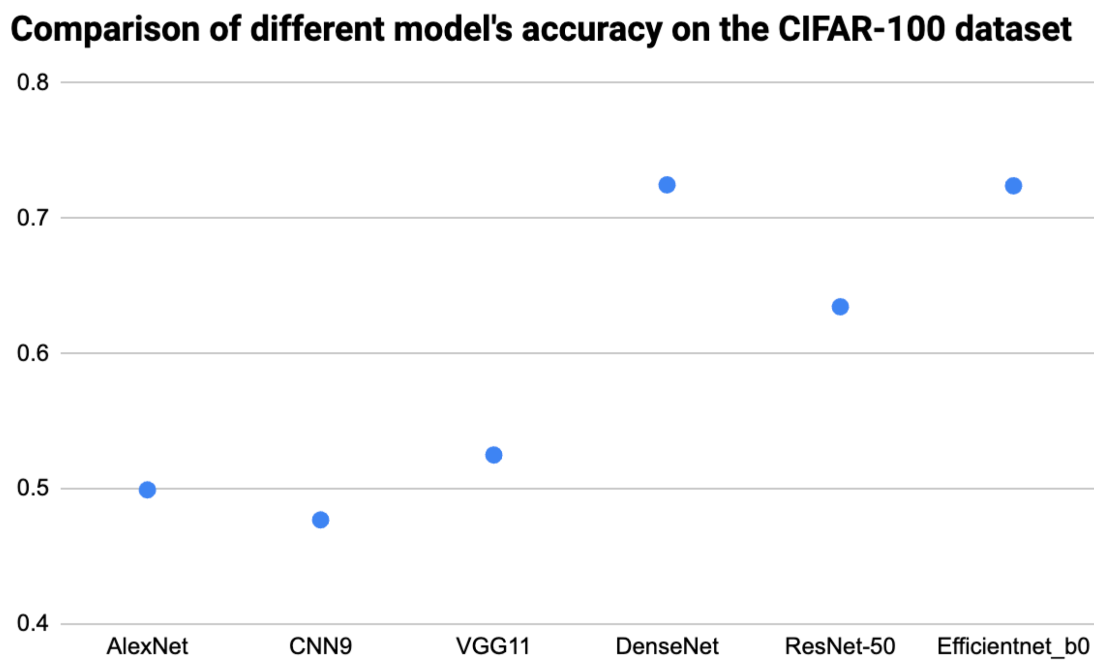

Figure 7. Graph by author displaying the different accuracies of models for the CIFAR-10 dataset

Model:	Accuracy:
AlexNet	0.4992
CNN9	0.4770
VGG11	0.5250
DenseNet	0.7248
ResNet-50	0.6346
Efficientnet_b0	0.7241

Table 2. Table by author representing the precise highest accuracy of each model for CIFAR-100 dataset (green representing machine learning model, red representing deep learning models)

Shown in Figure 7 and Table 2 are the six model's performances classifying the images in the CIFAR-100 dataset. Similar to CIFAR-10, AlexNet and CNN9 struggled to match the accuracies of the other models, where the two machine learning models had accuracies of below 0.5. This could be due to the fact that their architectures are simple and straightforward, not being able to crack CIFAR-100's complexity with 100 classes, and coarse/fine labels.

On the other hand, the deep learning models showed impressive accuracy, all reaching accuracies above 0.5. VGG11 was again able to outperform CNN9 and AlexNet, predictably due to its deeper layers, although having longer training times compared to the machine learning models. ResNet-50 was able to break the 0.6 threshold, due to its use of residual learning, and this high accuracy shows that it was able to effectively handle the dataset's complexity. Similar to the CIFAR-10 experiment, DenseNet and EfficientNet stood out as the top performers, both achieving accuracies of above 0.72, 20% more than the machine learning models. This is able to show that with a more complex dataset, the models with deeper neural networks were able to outperform the machine learning models by quite a large margin of at most 0.23.

This successfully validates the initial hypothesis that deep learning models attain higher accuracies than machine learning models, due to their deeper architecture, and enhanced features for feature extraction, showing that the modern models building upon traditional convolutional neural networks are able to overtake the traditional CNNs in accuracy.

3.4 Study on Overfitting

As mentioned in several occasions preceding this section, overfitting is an undesirable behavior in machine and deep learning, where models give accurate predictions for the training data, but not for the testing data ("What Is Overfitting? - Overfitting in Machine Learning Explained - AWS"). An overfit model is when the algorithm fits too closely with the training data, causing it to be unable to recognize and generalize unseen data for testing.

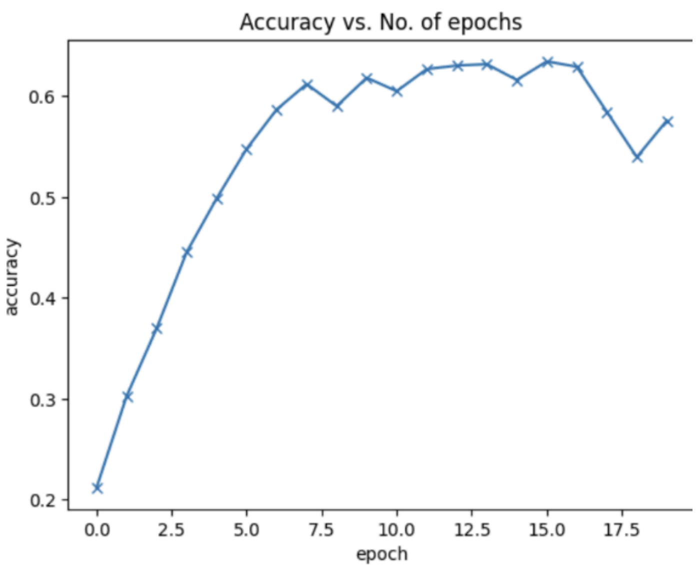

Figure 8.1. Graph by author displaying accuracy by epochs of ResNet-50 model trained on CIFAR-100 dataset

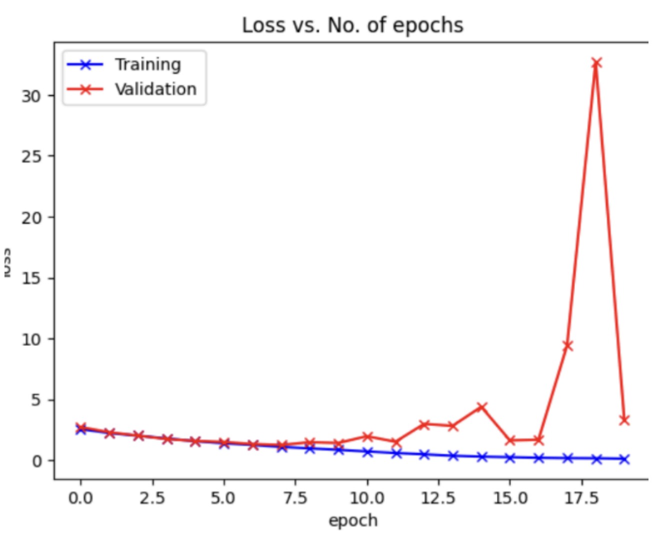

Figure 8.2. Graph by author displaying loss by epochs of ResNet-50 model trained on the CIFAR-100 dataset

Figures 8.1 and 8.2 take ResNet-50 as an example, which was found to be the best example of overfitting out of the six models used for this paper. Seen in the figures, it can be seen that the model initially shows improvement in accuracy, up until around the 7th epoch (where each epoch means one complete pass of the training dataset through the algorithm). However after the 10th epoch, where the model achieves the lowest validation loss and highest validation accuracy, the continued training indicates strong overfitting. As seen in Figure 8.1 and 8.2, from the 10th to 20th epochs, as the validation loss increases (spiking at epoch 18), the validation accuracy decreases (with a significant drop at epoch 18). This indicates that the model is beginning to overfit to the training data, where it captures noise and patterns that do not translate and generalize well to the unseen testing data.

To address the issue of overfitting, various solutions have been employed. One such example is early stopping, where training is halted as soon as the model validation accuracy stops improving, and others are regularization methods such as dropout and weight decay, which help to maintain a balance between the model's complexity and its generalization capability ("What Is Overfitting? - Overfitting in Machine Learning Explained - AWS")

4 Real World Applications

The previous sections have consisted of using various models to test out differences in accuracy between machine and deep learning models, and the concluded result is that deep learning models achieve higher accuracy than machine learning models for the task of image classification, but how can these models be adopted in real world applications?

As briefly stated in the beginning of the paper, image classification has a multitude of real-life applications, such as object recognition, face recognition, medical image analysis, autonomous driving, etc. But to zone in on a specific use case of image classification and display its effectiveness through experiments, the following section of this paper will be focused on how image classification models can be used to aid in environmental protection.

4.1 Waste Sorting

Along with the development of technology in making models more efficient, accurate, and powerful, the environment has taken a humongous toll as a result, so simultaneously, environmental degradation is a critical issue happening in the world right now, so how can image classification models help to combat this issue?

4.1.1 Design and Principle

To address the issue of environmental degradation, an AI-powered trash can has been developed, as seen in Figure 9.1, and will be powered by the EfficientNet model. The purpose of developing an AI powered trash can is so that it can enhance recycling efficiency and reduce human error in separating the different types of waste. And the reason for selecting the EfficientNet model is that compared to the other high-performing model DenseNet, it has an inference time cost of ~0.2 seconds compared to ~0.4 seconds for DenseNet, and also has reduced cases of overfitting, compared to the other models. The AI trash can uses trash recognition algorithms (same as image classification) to efficiently identify the type of waste taken by the camera, and open the appropriate trash can for disposal. This will hopefully allow for sustainable waste management practices, contributing to environmental conservation efforts.

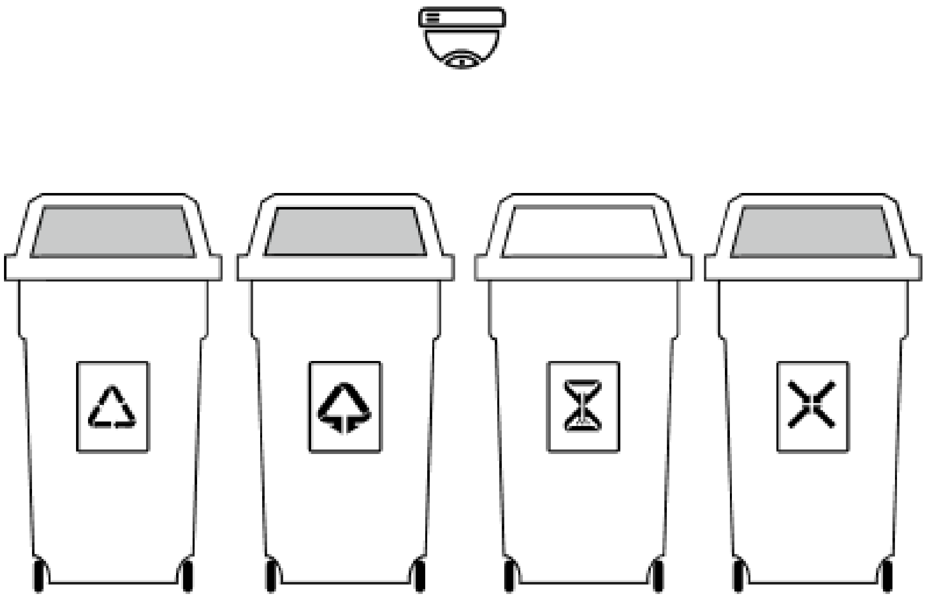

Figure 9.1. Drawing by author displaying the design of the AI Trash Can

4.1.2 Results for Waste Sorting Model

The EfficientNet model was trained using the TrashNet dataset, which contains various types of waste such as cardboard, glass, metal, paper, and plastic (Feyza Ozkefe). With around 2,500 collected waste images, TrashNet is a good foundation to test out the capabilities of the AI trash can, ensuring that the trash can is able to recognize and categorize different types of trash.

As seen in Figure 9.2, the EfficientNet model, which was the most consistently accurate model in the previous experiments, was able to achieve an accuracy of about 0.8 from the 20th epoch onwards. This shows that an AI trash can could prove effective in the real world, and by incorporating this AI technology into trash cans around the country and around the world, the recycling process can become more efficient and streamlined, promoting a cleaner and more sustainable environment. This shows that deep learning models have potential to become extremely useful in everyday applications, and in this case the model's significance in driving environmental protection.

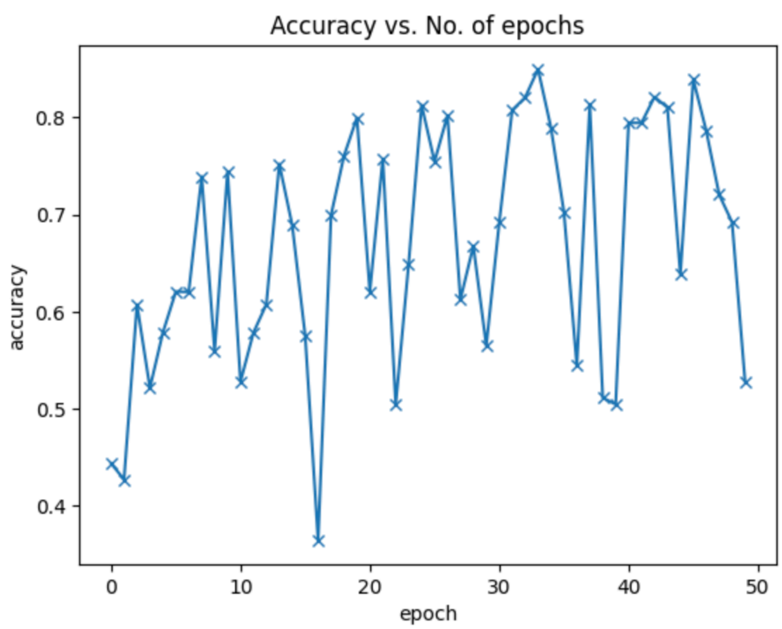

Figure 9.2. Graph by author displaying accuracy by epochs of EfficientNet model trained on TrashNet dataset

4.2 Air Quality Estimation

Another real world application of machine and deep learning models is a classifier for air quality levels from camera images. The model is originally trained with 10,404 images and tested on 1,836 images, and the dataset contains pollution levels across different environments in India and Nepal (Adarsh Rouniyar).

With this dataset, the main objective is to automate air quality detection through image classification, without needing the traditional sensor-based measurements. By using the dataset, I am able to analyze environmental conditions in urban and rural settings, and the models will be able to identify patterns that may not be apparent to the human eye.

4.2.1 Design and Principle

The air quality estimation model is designed through the use of CNNs, which as introduced above, are well-suited for relevant image classification tasks.

The main reason for the selection of the EfficientNet model is that it is able to extract the visual features of each training image, such as the presence of fog, smog, or clear skies, which can clearly correlate with the air quality levels. These features are then passed through multiple layers of the network to eventually produce a classification label.

The Air Quality Index Class is a measurement of air quality, ranging from 0 to 500, the higher the AQI, the greater the level of air pollution and health concern ("AQI Basics | AirNow.gov").

For the sake of this experiment, the classes of air pollution are represented in the dataset as follows:

1. Good (0-50): Air quality is satisfactory, and air pollution poses little risk

2. Moderate (51-100): Air quality is acceptable, however, this may be a health concern for a small number of people sensitive to air pollution

3. Unhealthy for Sensitive Groups (101-150): As the name implies, sensitive groups may experience health effects

4. Unhealthy (151-200): Some members of the general public may experience health effects

5. Very Unhealthy (201-300): The risk of health effects is increased for everyone

6. Hazardous (301-500): Everyone is likely to be affected

Each input image will be automatically classified in 6 classes, representing different qualities of the air environment.

4.2.2 Results for Air Quality Estimation Model

Category	**Accuracy**
Good	0.982
Moderate	0.983
Unhealthy of Sensitive Groups	0.984
Unhealthy	0.975
Very Unhealthy	0.971
Hazardous	0.991

Table 3. Table by author displaying the accuracy of the EfficientNet model in classifying each air quality index class, after 50 epochs

As seen in Table 3, the EfficientNet model is extremely effective in classifying the air quality index images into their respective classes, with accuracies ranging from 98 to 99%. This air quality estimation system can demonstrate how deep/machine learning algorithms can be applied in the real world to monitor air pollution levels efficiently, offering a more scalable and cost-effective solution.

Similar to the AI trash can, this model highlights the growing role of image classification models in solving environmental challenges, as well as the drive of sustainable practices on a global scale.

5 Conclusion and Evaluation

In conclusion, through extensive research and experiments for six types of image classification models, it was proven that deep learning models tend to yield higher accuracies, no matter for relatively small or large datasets. And it was also shown the deep learning image classification models can prove to be applicable in the real world, not only for previously recognized purposes such as traffic monitoring, medical image analysis, etc, but through a specific case study on waste image classification, it was gathered that deep learning models are able to be applied in the real world, yielding a high accuracies and efficient training times.

Therefore, to answer the question of ‘to what extent do machine learning and/or deep learning image classification algorithms exhibit effectiveness in real world applications’, it can said that both machine and deep learning image classification algorithms can be applied in the real world, but through experiments with the CIFAR-10 and CIFAR-100 dataset, it is deduced that deep learning models will be able to achieve higher accuracy than the machine learning models, which is again shown as EfficientNet is tested for the TrashNet dataset and air quality estimation dataset.

Works Cited

"6.6. Convolutional Neural Networks (LeNet) — Dive into Deep Learning 0.17.6 Documentation." *Classic.d2l.ai*, 2024, classic.d2l.ai/chapter_convolutional-neural-networks/lenet.html. Accessed 31 July 2024.

Adarsh Rouniyar. "Air Pollution Image Dataset from India and Nepal." *Kaggle.com*, 2023, www.kaggle.com/datasets/adarshrouniyar/air-pollution-image-dataset-from-india-and-nepal. Accessed 11 Sept. 2024.

Anna Alexandra Grigoryan. "Understanding VGG Neural Networks: Architecture and Implementation." *Medium*, Medium, 14 Aug. 2023, thegrigorian.medium.com/understanding-vgg-neural-networks-architecture-and-implementation-400d99a9e9ba. Accessed 1 Aug. 2024.

"AQI Basics | AirNow.gov." *Airnow.gov*, 2023, www.airnow.gov/aqi/aqi-basics/. Accessed 11 Sept. 2024.

"Deep Learning vs Machine Learning | Google Cloud." *Google Cloud*, 2024, cloud.google.com/discover/deep-learning-vs-machine-learning. Accessed 20 July 2024.

"Densenet." *PyTorch*, 2024, pytorch.org/hub/pytorch_vision_densenet/. Accessed 1 Aug. 2024.

Fagan, Matthew E, and Ruth S DeFries. "Remote Sensing and Image Processing." *Elsevier EBooks*, 1 Jan. 2024, pp. 432–445, www.sciencedirect.com/topics/earth-and-planetary-sciences/geometric-correction, https://doi.org/10.1016/b978-0-12-822562-2.00060-8. Accessed 20 July 2024.

Feyza Ozkefe. "TrashNet." *Kaggle.com*, 2021, www.kaggle.com/datasets/feyzazkefe/trashnet/data. Accessed 4 Aug. 2024.

Firican, George. "The History of Machine Learning | LightsOnData." *LightsOnData*, 31 Jan. 2022, www.lightsondata.com/the-history-of-machine-learning/. Accessed 20 July 2024.

Mukherjee, Suvaditya. "The Annotated ResNet-50 - towards Data Science." *Medium*, Towards Data Science, 18 Aug. 2022, towardsdatascience.com/the-annotated-resnet-50-a6c536034758. Accessed 1 Aug. 2024.

Nithyashree V. "Step-By-Step Guide for Image Classification on Custom Datasets." *Analytics Vidhya*, 19 July 2021, www.analyticsvidhya.com/blog/2021/07/step-by-step-guide-for-image-classification-on-custom-datasets/. Accessed 20 July 2024.

"Papers with Code - CIFAR-10 Dataset." *Paperswithcode.com*, 2022, paperswithcode.com/dataset/cifar-10. Accessed 20 July 2024.

Team, Keras. "Keras Documentation: Dropout Layer." *Keras.io*, 2024, keras.io/api/layers/regularization_layers/dropout/. Accessed 1 Aug. 2024.

"What Is Overfitting? - Overfitting in Machine Learning Explained - AWS." *Amazon Web Services, Inc.*, 2022, aws.amazon.com/what-is/overfitting/. Accessed 3 Aug. 2024.

Wikipedia Contributors. "AlexNet." *Wikipedia*, Wikimedia Foundation, 28 July 2024, en.wikipedia.org/wiki/AlexNet. Accessed 1 Aug. 2024.

Wisdomml. "EfficientNet and Its Performance Comparison with Other Transfer Learning Networks - Wisdom ML." *Wisdom ML*, 31 Mar. 2023, wisdomml.in/efficientnet-and-its-performance-comparison-with-other-transfer-learning-networks/. Accessed 3 Aug. 2024.