November 7, 2023

*Digital Object Identifier*

AIA PROJECT 1 – Image Classification

**RAFAEL DIAS (50%)1, FATEMEH KHAJEHMAHMOUD (50%)2,**

1Universidade de Aveiro, Departamento de Electrónica, Telecomunicações e Informática (e-mail: [rafaelsdias@ua.pt)(95284](mailto:rafaelsdias@ua.pt)(95284))

2Universidade de Aveiro, Departamento de Electrónica, Telecomunicações e Informática (e-mail: fat.khajehmahmoud@ua.pt)(110969)

Subject: Applications of Artificial Intelligence Course instructor: Petia Georgieva

**ABSTRACT** In this paper, we propose a machine learning problem of Intel Image Scene Classification of Landscapes. Our model is based on a dataset of intel contains around 25,000 images of size 150x150 distributed under 6 categories: Buildings, Forest, Glacier, Mountain, Sea, street. Using this dataset, we used CNN model for image classification.

**INDEX TERMS** Machine Learning, Deep Learning, CNN Model, ResNet, VGG16.

1. **STATE OF THE ART REVIEW**
2. ***SOME OF THE BEST WORKS IN IMAGE CLASSIFICATION MODEL***

There have been significant ideas in the image classification, with researchers seeking to develop and improve their performance.

To best determine which algorithms, techniques, and metrics are the best to use in terms of image classification, a state-of-the art analysis was conducted. Some ideas of this model include:

* 1. *“Gradient-based Learning Applied to Document Recognition”* [[1]](#_bookmark1)

Introduced in 1998, LeNet sets the foundation for future image classification research using convolution-nal neural networks. Many classic CNN techniques, such as pooling layers, fully connected layers, layering layers, and activation layers are used for feature extrac-tion and classification. With a mean square error loss function and 20 training sessions, this network can achieve 99.05% accuracy on the MNIST test set. Even after 20 years, many advanced classification networks still follow this pattern.

* 1. *“ImageNet Classification with Deep Convolutional Neural Networks”* [2]

Regarding deep learning models, this document re-views AlexNet.

AlexNet's breakthrough in 2012 marked a turning point in computer vision, dramatically improving ImageNet accuracy from 73.8% to 84.7%.

It introduced architectural advancements, larger net-works with ReLU activation, GPU training, and drop-out to handle complex features.

AlexNet's framework, combining convolution, ReLU *,* max-pooling, and dense layers, defined the standard for classification networks for the next decade.

* 1. *“Very Deep Convolutional Networks for Large-Scale Image Recognition”* [[3]](#_bookmark6)

This study focuses on VGG network.

Following AlexNet’s design, the VGG network has two major updates: 1) VGG not only used a wider network like AlexNet but also deeper. VGG-19 has 19 convolution layers, compared with 5 from Alex-Net. 2) VGG also demonstrated that a few small 3x3 convolution filters can replace a single 7x7 or even 11x11 filters from AlexNet, achieve better per-formance while reducing the computation cost. Because of this elegant design, VGG also became the back-bone network of many pioneering networks in other computer vision tasks, such as FCN for seman-tic segmentation, and Faster R-CNN for object detec-tion.

With a deeper network, gradient vanishing from multi-layers back-propagation becomes a bigger problem. To deal with it, VGG also discussed the importance of pre-training and weight initialization. This problem limits researchers to keep adding more layers, otherwise, the network will be hard to converge.

* 1. *“Deep Residual Learning for Image Recognition”* [[4]](#_bookmark0)

In this article, present a residual learning framework (ResNet) to ease the training of networks that are substantially deeper than those used previously.

It explicitly reformulates layers as learning residual functions by reference to layer inputs, rather than learning dereferenced functions. It provides general experimental evidence that these residual networks are easier to optimize and can achieve significantly increased accuracy from depth. On the ImageNet data-set, it evaluates residual networks with a depth of up to 152 layers --- 8 times deeper than VGG networks, but still less complex. A set of these residual networks achieves 3.57% error on the ImageNet test set. This result was ranked 1st in the 2015 ILSVRC classification task. It also provides CIFAR-10 analysis with 100 and 1000 layer.

* 1. *“Xception: Deep Learning with Depthwise Separable Convolutions”.[5]*

*This paper presents an interpretation of Inception modules in convolutional neural networks as an intermediate step between regular convolution and deep separable convolution operations (a depth convolution followed by a pointwise convolution). In this light, a deeply separable complex can be understood as an Inception module with a maximum number of towers. This architecture, called Xception, performed slightly better than Inception V3 on the ImageNet dataset (for which Inception V3 was designed), and significantly better than Inception V3 on a larger image classification dataset of 350 million images and 17,000 The class acts. Since the Xception architecture has the same number of parameters as Inception V3, the performance increase is not due to increased capacity, but due to more efficient use of model parameters.*

Overall, CNN models have revolutionized the field of computer vision by allowing automatic feature learning from raw data and have been adapted for various domains, becoming an integral part of modern deep learning applications.

1. ***EXISTING DEEP LEARNING MODELS***

Several previous studies have explored deep learning models for image classification. There are still limitations and challenges associated with existing models. Many deep learning models require substantial amounts of labeled data for training. This can be a limitation when working with niche or specialized image classification tasks where labeled data is scarce. Training deep neural networks, especially very deep architectures like ResNet or Inception, demands significant computational resources, including powerful GPUs and TPUs. This can be a practical limitation for researchers and organizations with limited resources. Deep learning models can inherit biases present in the training data, potentially leading to biased predictions. Addressing issues of bias and fairness is a critical concern in image classification, especially in sensitive applications. The use of deep learning models in image classification has raised ethical concerns, especially in applications like surveillance, privacy invasion, and deepfake generation.

We apply various CNN models including two popular pre-trained models for image classification: ResNet50 and VGG16.

These models are commonly used in transfer learning scenarios due to their strong performance on a wide range of image-related tasks. In this case, the ResNet50 model is trained for 5 epochs with early stopping, and model checkpoint callbacks to save the best model based on validation accuracy and the VGG16 model used SGD optimizer, and training is done for 5 epochs with early stopping and model checkpoint callbacks. Compare Metrics (Precision, Recall, F1-Score) for Different Models.

1. **ML PROBLEM COMPLEXITY**

***CHALLENGES IN IMAGE CLASSIFICATION MODEL***

Image classification is a popular and important task in computer vision, but it comes with several challenges that researchers and practitioners need to address. Some of the key challenges in image classification models include:

1. Variability in Object Appearance: Objects can appear in various poses, lighting conditions, backgrounds, and scales. Image classification models need to be robust to these variations.

2. Large-Scale Data: Training deep neural networks for image classification requires large datasets, which can be costly to collect and label.

3. Overfitting: Deep neural networks are prone to overfitting, especially when the training dataset is small. Regularization techniques are often used to mitigate this issue.

4. Class Imbalance: Some classes may have significantly fewer examples than others, which can lead to imbalanced learning and poorer performance on minority classes.

5. Generalization: Ensuring that a model performs well on unseen data is crucial. Models that overfit the training data may not generalize effectively.

6. Model Architecture and Hyperparameters: Selecting the appropriate architecture (e.g., CNNs, ResNets, Inception) and hyperparameters (e.g., learning rate, batch size) can be challenging and may require experimentation.

7. Data Augmentation: Creating diverse training data through data augmentation techniques is important to improve the model's ability to handle variations in input images.

8. Interclass Variability: Objects of the same class may exhibit substantial differences in appearance, making it difficult to create a model that recognizes them all accurately.

9. Intra-class Variability: Conversely, objects from different classes may look similar, making it challenging to differentiate between them.

10. Adversarial Attacks: Deep learning models, including image classifiers, can be vulnerable to adversarial attacks, where small, carefully crafted perturbations can lead to misclassification.

11. Computational Resources: Training deep neural networks for image classification requires significant computational resources, which can be a challenge for smaller research teams or organizations with limited access to powerful hardware.

12. Label Noise: Noisy or incorrect labels in the training data can mislead the model and reduce its accuracy.

13. Real-time Inference: Some applications, such as autonomous vehicles, require image classification models to perform real-time inference, which places additional constraints on model size and speed.

14. Ethical Concerns: Ensuring that image classification models are unbiased and do not discriminate against certain groups or exhibit other ethical issues is crucial.

15. Transfer Learning: Leveraging pre-trained models for image classification may require domain adaptation, and fine-tuning strategies to make them work effectively in specific applications.

In conclusion, addressed these challenges often involves a combination of data collection, preprocessing, model selection, hyperparameter tuning, and ongoing model monitoring and maintenance. Additionally, staying up to date with the latest research and techniques is essential for improving image classification models.

1. **DATA DESCRIPTION**
2. ***DESCRIPTION OF IMAGE CLASSIFICATION DATA SOURCES***

The dataset described in this report is a collection of images that published by Intel to host an Image Classification Challenge. The dataset contains around 25,000 images of size 150x150 distributed under 6 categories: buildings, forest, glacier, mountain, sea, street.

The train, test, and prediction data are separated into individual zip files. There are approximately 14,000 images in the train set, 3,000 in the test set, and 7,000 in the prediction set. Images are loaded from TensorFlow's Keras API and used scikit-learn to encode the image categories (classes) into numeric labels. This is important for classification tasks then splits the data into training and testing sets. Overall, the study predominantly focuses on image data preprocessing for deep learning models.

1. ***FEATURES OF THE DATA SET***

the primary dataset consists of images used for image classification. Each image belongs to a specific category, and the dataset is organized into subdirectories based on these categories. Here are the key features of the dataset in this study:

* Image Data: the primary data in this code is a collect-ion of images. These images are stored in directories, with each directory representing a different category or class. The study works with these images for train-ing and testing machine learning models.
* Categories (Classes): the images are categorized into different classes or categories. In this study, these categories are extracted from the subdirectories where the images are stored. The categories variable holds the list of category names, and the total number of categories is printed.
* Data Distribution: provides information about the distribution of images across categories. It counts the number of images in each category for both training and testing data and displays this information. This is crucial for understanding the balance of the dataset.
* Label Encoding: the dataset's class labels (category names) are encoded into numeric labels. The encoded labels are essential for training a machine learning model, as models work with numeric inputs.
* Data Splitting: The dataset prepares data generators for training and testing data, which suggests a data split.
* Image Preprocessing: the code preprocesses image data, rescaling pixel values to a range between 0 and 1.

The primary focus of this code is on image data, specifically for training deep learning models like ResNet50 and VGG16 for image classification. Therefore, the key features of the dataset are related to image data and its organization into categories or classes.

1. ***DATA VISUALIZATION***

Data visualization is an important step in understanding and interpreting the results of our analysis. By visualizing our data, we can identify patterns and relationships that may not be immediately apparent from the raw numbers.

1. **DATA PREPROCESSING**
2. ***DATA CLEANING AND FEATURE SELECTION***
3. ***TRAIN/VALIDATION/TEST DATA DIVISION***
4. ***K-FOLD CROSS VALIDATION***
5. **DESCRIPTION OF THE APPLIED MACHINE LEARNING METHODS**
6. ***OVERVIEW OF THE MACHINE LEARNING TECHNIQUES USED***
7. ***EXPLANATION OF MODEL SELECTION AND HYPERPARAMETER TUNING***
8. **RESULTS**
9. ***//***
10. Comparison with default parameters and optimized parameters
11. ***MODEL COMPARISON***
12. **CONCLUSIONS**
13. **SUGGESTIONS FOR FUTURE WORK**

**References**

[1]. Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, “Gradient-based learning applied to document recognition”, URL: https://ieeexplore.ieee.org/document/726791

[2]. Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton,” ImageNet Classification with Deep Convolutional Neural Networks”. URL: <https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>.

[3]. Shuying Liu and Weihong Deng, “Very deep convolutional neural network based image classification using small training sample size”. URL: <https://ieeexplore.ieee.org/document/7486599>.

[4]. Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun, “Deep Residual Learning for Image Recognition”, URL: <https://ieeexplore.ieee.org/document/7780459>.

[5]. Francois Chollet, “Xception: Deep Learning with Depthwise Separable Convolutions”, URL: <https://arxiv.org/abs/1610.02357>.

