Application of Deep Learning Techniques on COIL-100 Dataset

The report delves into the application of various deep learning models for object classification using the COIL-100 dataset. It explores different architectures, including simple and deeper CNNs, transfer learning with MobileNetV2, VGG-16, ResNet-50, InceptionV3, and an ensemble model. The main issues addressed involve understanding the effectiveness of different model architectures in accurately classifying objects from images, particularly in the context of real-world applications where object recognition is crucial.

Throughout the experimentation, key findings emerged regarding the performance of each model. Simple and deeper CNNs displayed robust performance, achieving high accuracy levels on the validation set. However, they struggled with more complex patterns in the dataset. Transfer learning using MobileNetV2 showed outstanding performance, achieving perfect accuracy, highlighting the effectiveness of pre-trained models. VGG-16, ResNet-50, and InceptionV3 demonstrated varying degrees of success, with InceptionV3 showing particularly strong results. The ensemble model, combining the predictions of all previous models, consistently outperformed individual models, indicating the effectiveness of ensemble methods in improving classification accuracy. These findings provide valuable insights into the strengths and weaknesses of different deep learning architectures for object classification tasks.

Introduction :

Object classification, particularly in the context of computer vision, is a fundamental problem with widespread applications in various fields like image recognition, autonomous vehicles, medical imaging, and surveillance systems. The ability to accurately identify objects from images is crucial for automation, decision-making, and understanding visual data. However, achieving high accuracy in object classification tasks is challenging due to factors such as variations in object appearance, background clutter, and occlusions.

In this study, I focused on the problem of object classification using deep learning techniques, specifically exploring different architectures and their effectiveness in accurately classifying objects. The COIL-100 dataset, consisting of images of 100 different objects captured from different angles, serves as our benchmark dataset. The importance of this study lies in its potential to advance the field of computer vision by identifying the most effective deep learning models for object classification tasks. Understanding which architectures perform best can lead to the development of more robust and accurate systems for object recognition, with implications for various real-world applications.

Deep learning has emerged as a powerful tool for image classification tasks, allowing models to automatically learn hierarchical representations of data. However, choosing the right architecture and training strategy is critical for achieving optimal performance. By investigating different architectures, including simple CNNs, deeper CNNs, and transfer learning from pre-trained models, my aim is to provide insights into the strengths and weaknesses of each approach. This study contributes to the ongoing efforts to improve the accuracy and robustness of object classification systems, ultimately enhancing their practical utility in real-world scenarios.

Current Research :

The findings of the paper "Colour Neural Descriptors for Instance Retrieval Using CNN Features and Colour Models" are as follows:

The paper introduces color neural descriptors for instance retrieval, utilizing convolutional neural network (CNN) features combined with different color spaces and channels. Unlike previous methods that require fine-tuning pre-trained networks, the proposed descriptors are computed based on activations from a pre-trained VGG-16 network without fine-tuning. Additionally, the authors utilize an object detector to optimize the instance retrieval architecture, enabling feature generation at both local and global scales. Furthermore, a stride-based query expansion technique is introduced to retrieve objects from multi-view datasets.

The experimental results demonstrate that the proposed color neural descriptors achieve state-of-the-art performance across various datasets. Specifically, in the Paris 6K, Revisiting-Paris 6k, INSTRE-M, and COIL-100 datasets, the mean Average Precision (mAP) scores are 81.70, 82.02, 78.8, and 97.9, respectively. This highlights the effectiveness of the approach in retrieving identical images as the most relevant ones from large image or video corpora.

The findings of the paper "A deep learning and transfer learning model for intra-change detection in images" are as follows:

The paper addresses the issue of on-shelf availability (OSA) in retail stores, emphasizing the importance of maintaining product availability and organization on store shelves to enhance the customer shopping experience and profitability. It discusses the challenge of detecting misplaced objects on retail shelves, as customers may disrupt the arranged products during shopping.

The proposed approach utilizes a convolutional neural network (CNN) to classify shelf imagery into correct semantic classes, determining whether products are misplaced or not. The architecture is evaluated using a modified COIL-100 dataset.
Results show that the transfer learning (TL) based MobilenetV2 model achieves the best performance with 91.28% accuracy. Additionally, a CNN-based model with 11 user-defined layers achieves an accuracy of 90.36%. These findings demonstrate the efficacy of both transfer learning-based and custom CNN models for change detection on retail shelves.

The findings of the paper "A Novel Technique for Image Retrieval based on Concatenated Features Extracted from Big Dataset Pre-Trained CNNs" are as follows:

The paper addresses the challenge of accessing semantically relevant data from a database, emphasizing the importance of image representation and feature extraction in Content-Based Image Retrieval (CBIR), particularly in commercial and medical imaging applications. Traditionally, texture, shape, and color features are used for CBIR, but the paper explores the use of features extracted from pre-trained Convolutional Neural Networks (CNNs) to enhance retrieval performance.

Specifically, the paper focuses on Big Transfer Networks, which are state-of-the-art pre-trained CNNs known for their high discriminative power. The proposed technique aims to demonstrate the effectiveness of utilizing features from these networks for image retrieval. Furthermore, the paper suggests concatenating features from multiple Big Transfer Networks to improve retrieval performance, leveraging both feature and network diversity for enhanced discriminative power.

The effectiveness of the proposed method is evaluated through simulations on four datasets of varying sizes and complexities: COREL-100, CALTECH-101, FLOWER-17, and COIL-100. The study includes experiments with different feature sizes achieved through concatenation, and dimensionality reduction using Principal Component Analysis (PCA). Various distance metrics are explored to measure similarity between images.

Results indicate that selecting appropriate pre-trained CNNs and distance metrics can lead to higher mean Average Precision (mAP). Specifically, the ImageNet-21K pre-trained CNN and Instagram pre-trained CNN are chosen for their superior performance. The study shows that CNNs trained on ImageNet-21K dataset outperform those trained on ImageNet-1K due to a wider variety of classes and images.

Comparative analysis with existing algorithms demonstrates the superiority of the proposed method in terms of mean Average Precision across various datasets and precision levels. This highlights the effectiveness of utilizing features from Big Transfer Networks and concatenating them for image retrieval tasks.

For this study, I utilized the COIL-100 dataset, which is readily available through TensorFlow Datasets. The COIL-100 dataset comprises images of 100 different objects, each captured from various angles under different lighting conditions. Each object has 72 images associated with it, resulting in a total of 7,200 images in the dataset.

Dataset Collection :

The characteristics of the COIL-100 dataset are as follows:

- Object Diversity: The dataset contains images of a wide range of objects, including household items, toys, tools, and various other objects.
- Image Variability: Each object is captured from multiple angles, resulting in images with varying viewpoints and orientations.
- Image Resolution: The images are grayscale and have a resolution of 128x128 pixels.
- Labeling: Each image is associated with a unique object label, allowing for supervised learning.
- Number of Classes: There are 100 classes in total, with each class representing a distinct object.

Data collection is straightforward as the COIL-100 dataset is readily available in TensorFlow Datasets. I imported the dataset using TensorFlow's dataset API, specifying the 'coil100' split for the training set. Preprocessing steps involved normalizing the pixel values to the range [0, 1] and one-hot encoding the object labels.

This preprocessing ensures that the data is suitable for training deep learning models.

Model Building and Model Training :

For model development and training, I experimented with several deep learning architectures to identify the most effective one for object classification using the COIL-100 dataset.

I began by importing the dataset using TensorFlow Datasets, ensuring that it was split into training and validation sets. The data preprocessing involved normalizing the pixel values to a range of [0, 1] and one-hot encoding the object labels. This step ensured that the data was properly formatted for training deep learning models.

Next, I developed multiple models with varying architectures:

- Simple CNN: constructed a basic Convolutional Neural Network (CNN) consisting of convolutional layers followed by max-pooling layers and dense layers.
- Deeper CNN: designed a deeper CNN with additional convolutional layers to capture more complex features.
- Transfer Learning (MobileNetV2): employed transfer learning using the MobileNetV2 architecture, leveraging pre-trained weights on ImageNet. This approach allows the model to transfer knowledge learned from a large dataset to my specific task.
- Customized MobileNetV2: Similar to transfer learning, but with added dense layers and dropout regularization to fine-tune the model for my dataset.
- VGG-16: implemented the VGG-16 architecture, known for its simplicity and effectiveness, with several convolutional and max-pooling layers.
- ResNet-50: utilized the ResNet-50 architecture, which introduced skip connections to address the vanishing gradient problem, allowing for deeper networks.
- InceptionV3: employed the InceptionV3 architecture, known for its inception modules, which enable the model to capture features at multiple scales.

After constructing each model, compiled them using appropriate loss functions, optimizers, and evaluation metrics. then trained each model on the training dataset for a fixed number of epochs while monitoring performance on the validation set to prevent overfitting. Training involved adjusting model parameters (weights and biases) based on the calculated loss and gradients using backpropagation.

Once training was complete, I evaluated each model's performance on the validation set to assess its accuracy and generalization ability. Additionally, we computed metrics such as precision, recall, and F1-score to gain deeper insights into each model's performance.

Furthermore, I created an ensemble model that combined the predictions of all previous models to potentially improve overall classification accuracy.

Analysis :

My analysis of the different deep learning architectures for object classification on the COIL-100 dataset revealed several key findings:

Performance Variation: I observed significant variation in performance among the different models. While some models achieved high accuracy, others struggled to classify objects accurately.

Model	Accuracy
Simple CNN	97.59%
Deeper CNN	97.18%
MobileNetV2	100%
Customized MobileNetV2	100%
VGG-16	0.60%
ResNet-50	31.48%
InceptionV3	98.98%
Ensemble Model	100%

Table 1 : Accuracies of Trained Models

These accuracies reflect the effectiveness of each model in correctly classifying objects from the COIL-100 dataset. Simple and deeper CNNs, along with InceptionV3, demonstrated high accuracy, while MobileNetV2 and the customized version outperformed all others with perfect scores. Notably, VGG-16 and ResNet-50 showed poor performance, indicating limitations in their ability to accurately classify objects in this context. However, by combining predictions from all models, the ensemble model achieved perfect accuracy, highlighting the potential of ensemble learning to improve overall classification performance.

Effectiveness of Transfer Learning: Transfer learning with MobileNetV2 demonstrated outstanding performance, achieving near-perfect accuracy on the validation set. This finding indicates the effectiveness of leveraging pre-trained models for similar tasks, especially when dealing with limited training data.

Model Complexity vs. Performance: I found that deeper and more complex models, such as VGG-16, ResNet-50, and InceptionV3, did not always outperform simpler models like the basic CNN. This suggests that model complexity does not necessarily translate to improved performance, and simpler models can be equally effective for certain tasks.

Ensemble Learning: The ensemble model, which combines predictions from multiple models, consistently outperformed individual models. This finding underscores the effectiveness of ensemble learning in improving classification accuracy by leveraging the strengths of different models.

Model Robustness: Despite achieving high accuracy on the validation set, some models, such as VGG-16 and ResNet-50, exhibited poor performance in terms of precision, recall, and F1-score. This indicates that while these models may correctly classify some objects, they struggle with others, suggesting a lack of robustness.

Training Time and Complexity: Deeper models like VGG-16, ResNet-50, and InceptionV3 required longer training times compared to simpler models. However, this increased complexity did not always translate to improved performance.

My research on the topic indicates that while deep learning models can achieve high accuracy in object classification tasks, the choice of architecture plays a crucial role. Transfer learning, ensemble learning, and simpler architectures like basic CNNs can often yield comparable or better results than complex models. Additionally, model robustness, training time, and complexity should be carefully considered when selecting a model for a specific task. Overall, my findings provide valuable insights into the effectiveness of different deep learning architectures for object classification.

Summary and Conclusion :

In summary, my investigation into deep learning architectures for object classification on the COIL-100 dataset yielded several significant findings.
Firstly, I explored a variety of architectures, ranging from simple CNNs to more complex models like VGG-16, ResNet-50, and InceptionV3. my analysis revealed that while deeper and more complex models could capture intricate features, they didn't always outperform simpler architectures. Surprisingly, VGG-16 and ResNet-50 showed poor performance, suggesting limitations in their ability to handle the dataset's complexity.

On the other hand, transfer learning with MobileNetV2 and a customized version of it demonstrated outstanding accuracy, underscoring the efficacy of pre-trained models for object classification tasks, especially when dealing with limited training data.
Furthermore, ensemble learning proved highly effective, with the ensemble model combining predictions from all models achieving perfect accuracy. This highlights the potential of ensemble methods to enhance classification accuracy by leveraging the strengths of multiple models.

In conclusion, my study suggests that the choice of architecture significantly impacts model performance, and simpler models like basic CNNs or transfer learning models can be equally effective or even superior in certain contexts. Additionally, ensemble learning can further improve classification accuracy. These findings provide valuable insights for practitioners in selecting appropriate deep learning architectures for object classification tasks, ultimately contributing to the advancement of computer vision applications.

References :

[1] S. Saikia, L. Fernández-Robles, E. F. Fernández and E. Alegre, "Colour Neural Descriptors for Instance Retrieval Using CNN Features and Colour Models," in IEEE Access, vol. 9, pp. 23218-23234, 2021, doi: 10.1109/ACCESS.2021.3056330.

[2] Thirumurthy, B., Parameswaran, L. & Vaiapury, K. A deep learning and transfer learning model for intra-change detection in images. *J Supercomput* **80**, 11640–11660 (2024). https://doi.org/10.1007/s11227-023-05878-w

[3] Bhuma, C.M., & Kongara, R. (2023). A Novel Technique for Image Retrieval based on Concatenated Features Extracted from Big Dataset Pre-Trained CNNs. I.J. Image, Graphics and Signal Processing, 2(1), 1-12. Published Online on April 8, 2023 by MECS Press. DOI: 10.5815/ijigsp.2023.02.01