Application of Deep Learning Techniques on COIL-100 Dataset

Yashwanth Reddy Sadala
811292491

Using the COIL-100 dataset, the study explores the use of several deep learning models for object classification. It investigates various architectures, such as an ensemble model, MobileNetV2, VGG-16, ResNet-50, InceptionV3, and basic and deeper CNNs, as well as transfer learning. Understanding how well various model architectures categorize things from photos is one of the primary concerns addressed, especially in the context of practical applications where object recognition is essential.

During the evaluation, important conclusions about each model's performance emerged. On the validation set, both simple and deeper CNNs performed well, achieving high accuracy levels. However, they struggled with the dataset's more complicated patterns. Transfer learning with MobileNetV2 performed exceptionally well, obtaining flawless accuracy, and demonstrating the efficiency of pre-trained models. VGG-16, ResNet-50, and InceptionV3 all had varied degrees of success, with InceptionV3 performing particularly well. The ensemble model, which included the predictions of all preceding models, consistently outperformed individual models, demonstrating the efficacy of ensemble approaches for boosting classification accuracy. These findings shed light on the benefits and disadvantages of various deep learning architectures used in object categorization tasks.

Introduction:

Object classification, particularly in the context of computer vision, is a basic issue that has several applications in areas such as image identification, autonomous cars, medical imaging, and surveillance. The capacity to reliably recognize things in photographs is critical for automating, making decisions, and comprehending visual data. However, getting high accuracy in object classification tasks is difficult because of issues such as differences in item appearance, background clutter, and occlusions.

In this study, I focused on the issue of object classification using deep learning techniques, especially investigating alternative architectures and their accuracy in identifying things. The COIL-100 dataset, which contains photos of 100 distinct objects taken from various perspectives, serves as our benchmark dataset. The significance of this work stems from its ability to progress the area of computer vision by selecting the best successful deep learning models for object categorization tasks. Understanding which designs work best can help to build more robust and accurate object recognition systems, having implications for a wide range of real-world applications.

Deep learning has emerged as an effective approach for image classification, allowing models to automatically learn hierarchical data representations. However, selecting the appropriate architecture and training technique is important to obtaining peak performance. My goal in examining several designs, including simple CNNs, deeper

CNNs, and transfer learning from pre-trained models, is to give insights into the merits and shortcomings of each technique. This work adds to continuing attempts to increase the accuracy and resilience of object categorization algorithms, hence increasing their practical value in real-world contexts.

Current Research:

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

The study 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 article discusses on-shelf availability (OSA) in retail businesses, highlighting the necessity of maintaining product availability and organization on store shelves to improve consumer shopping experience and profitability. It covers the difficulty of spotting lost things on store shelves, as people may disrupt the organized merchandise while 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 study addresses the problem of retrieving semantically meaningful data from a database, highlighting the significance of image representation and feature extraction in Content-Based Image Retrieval (CBIR), notably in commercial and medical imaging applications. Traditionally, texture, shape, and color characteristics are utilized for CBIR, but this article investigates using features taken from pre-trained Convolutional Neural Networks (CNNs) to improve retrieval performance.

The research focuses on Big Transfer Networks, which are cutting-edge pre-trained CNNs noted for their strong discriminative capacity. 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 investigation, I used the COIL-100 dataset, which is publicly available from TensorFlow Datasets. The COIL-100 collection includes photos of 100 distinct items taken from various angles and under varied lighting conditions. Each item is paired with 72 photos, for a total of 7,200 in the collection.

Dataset Collection:

The COIL-100 dataset has the following characteristics:

1. Object Diversity: The collection includes photos of a diverse range of things, such as household goods, toys, tools, and other stuff.

2. Image Variability: Each item is photographed from various perspectives, producing photos with varied views and orientations.

3. Image Resolution: The photos are grayscale and are 128x128 pixels.

4. Labeling: Each image is assigned a unique item label, which allows for supervised learning.

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:

Using the COIL-100 dataset, I tested with various deep learning architectures to see which one was the most successful for object categorization.

I began by loading the dataset into TensorFlow Datasets, separating it into training and validation sets. The data preparation consisted of normalizing the pixel values to a range of [0, 1] and one-hot encoding the object labels.

Next, I developed multiple models with varying architectures:

1. Simple CNN: constructed a basic Convolutional Neural Network (CNN) consisting of convolutional layers followed by max-pooling layers and dense layers.
2. Deeper CNN: designed a deeper CNN with additional convolutional layers to capture more complex features.
3. 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.
4. Customized MobileNetV2: To improve the model for my dataset, I added dense layers and dropout regularization, making it akin to transfer learning.
5. VGG-16: used many convolutional and max-pooling layers to build the VGG-16 architecture, which is renowned for its efficacy and simplicity.
6. ResNet-50 made use of the ResNet-50 design, which allowed for deeper networks by including skip connections to solve the vanishing gradient issue.
7. InceptionV3 architecture: which is well-known for its inception modules, the model can collect features at various 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 discovered that simpler models, like the basic CNN, were not always superior to deeper and more complicated models, such VGG-16, ResNet-50, and InceptionV3. This shows that for some tasks, simpler models might be just as successful as more complicated ones, and that complexity in a model does not always translate into better performance.

Ensemble Learning: Individual models were routinely outperformed by the ensemble model, which integrates predictions from numerous models. This result emphasizes how well ensemble learning works to increase classification accuracy by utilizing the advantages of many models.

Model Robustness: Some models, including ResNet-50 and VGG-16, performed poorly in terms of precision, recall, and F1-score even though they achieved excellent accuracy on the validation set. This suggests that although these models might successfully categorize certain items, they have difficulty.

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:

To sum up, my research on object categorization using deep learning architectures on the COIL-100 dataset produced a few important conclusions.
Initially, I investigated a range of architectures, from basic CNNs to more intricate models like VGG-16, ResNet-50, and InceptionV3. According to my investigation, simpler designs were not necessarily outperformed by deeper, more complicated models, even when the latter might capture subtle details. Remarkably, VGG-16 and ResNet-50 performed poorly, indicating that the intricacy of the dataset may have been above their capabilities.

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