Group 7: Group_ Data_Bias

Team Members:

1. **Yash Sanghani (228003031)**
2. **Siddharth Yalamanchali (228003083)**
3. **Yashwant Avula (228003768)**

Data Set Bias in Deep Learning

Problem Description

In the past decade there has been significant progress in the field of deep learning. Particularly with the development of more diverse, large-scale datasets and advanced neural network architectures. However, despite these advancements, dataset bias remains a pervasive issue. This report explores if less biased and more comprehensive datasets trained with modern neural networks can still inherently detect and exploit biases within these datasets.

The core investigation consists of whether neural networks can achieve high accuracy when trained to perform "dataset classification" (i.e., identifying which dataset a particular image comes from). A strong result would suggest that biases still exist in these datasets. Surprisingly, the study finds that neural networks can indeed classify images by dataset with high accuracy, indicating that dataset biases are not only present but also readily detectable by current models.

This revelation suggests that despite efforts to create unbiased datasets, neural networks can still uncover subtle dataset-specific patterns (biases) that are not immediately obvious to humans. These findings prompt a re-evaluation of how datasets are constructed and challenge the community to develop methods that can genuinely reduce or eliminate bias, ensuring that machine learning models do not perpetuate or exacerbate existing inequalities or biases in data representation.

Dataset Descriptions and Rationale

Pascal VOC: The Pascal Visual Object Classes (VOC) dataset is a well-established resource in the field of computer vision, renowned for its rich annotations that include object detection, image classification, and segmentation. This dataset is particularly valued for its diversity in object classes and scenarios, which presents a balanced challenge for any classification model. The variety in image composition and object interaction within the images makes Pascal VOC an ideal candidate for testing the robustness and adaptability of convolutional neural networks (CNNs).

ImageNet: ImageNet comprises over a million labeled images across thousands of categories. The dataset's sheer volume and the granularity of its categorization make it an excellent benchmark for evaluating the depth and generalization ability of CNN architectures. ImageNet's complexity and scale have driven major advancements in deep learning, providing a comprehensive testbed for both training and fine-tuning sophisticated models.

Flickr30K: This dataset includes 30,000 images collected from Flickr, each accompanied by five descriptive captions, making it a valuable resource for learning tasks that integrate image and text processing. For the purposes of dataset classification, Flickr30K offers a unique challenge due to its focus on everyday scenes and activities, which are annotated with rich textual descriptions. This diversity is instrumental in testing how well models can generalize from visual content to textual annotations and vice versa.

Reason for Standardizing Input Image Size to 64x64 Pixels

The decision to standardize the input image size to 64x64 pixels is driven by several practical and experimental considerations. First, reducing the resolution of images to a uniform size facilitates a more streamlined and efficient computational process, as it reduces the variability in input dimensions and the computational load on the network, allowing for faster training cycles and easier scalability of the models across different hardware setups.

Furthermore, a reduced image size helps in mitigating overfitting by limiting the amount of fine-grained detail that models can use to make predictions, thus encouraging the learning of more generalizable features. This is particularly important in studies where the capacity to generalize across different datasets and real-world scenarios is a key outcome.

Additionally, using a smaller image size can also highlight the effectiveness of various architectural features of CNNs, such as depth and filter sizes, in capturing relevant patterns from lowerresolution inputs. This is especially pertinent in comparative studies of models like AlexNet, VGG-16, and ResNet-50, which are originally designed for higher resolution inputs. Adjusting these models to perform well on significantly reduced image sizes challenges and tests their adaptability and efficiency, providing valuable insights into the scalability and flexibility of these architectures.

Description of Models

In the project, three different convolutional neural network architectures—AlexNet, VGG-16, and ResNet-50—are implemented to address the task of Dataset classification across different datasets: Pascal VOC, ImageNet, and Flickr30K. The setup for each model is designed to adapt the deep learning architectures to varying input data characteristics and sizes, while leveraging standard normalization and optimization techniques to enhance training efficiency and model performance.

AlexNet Configuration

The AlexNet implementation begins by setting the standard normalization values used in the ImageNet dataset, with mean values of [0.485, 0.456, 0.406] and standard deviations of [0.229, 0.224, 0.225]. These values are crucial for ensuring that input images are scaled appropriately, mitigating variance across the dataset and accelerating convergence. The AlexNet model is modified for a reduced input resolution, adapting the network to handle images resized to 64x64 pixels. This resizing necessitates adjustments in the convolutional layers'

parameters—specifically, the stride and padding are tailored to maintain effective receptive field sizes and preserve spatial hierarchies in feature maps. The model's architecture incorporates several layers including multiple convolutional layers with varying sizes of receptive fields and padding, followed by max-pooling layers that reduce spatial dimensions while retaining critical features. The classifier section consists of fully connected layers, integrating dropout regularization to prevent overfitting, reflecting a configuration suited for robust learning across diverse image contents.

VGG-16 Configuration

For VGG-16, the configuration also adheres to the normalization standards with the same mean and standard deviation as AlexNet, ensuring consistency in image preprocessing. The model employs a sequence of 3x3 convolutional layers—a hallmark of VGG-16 design—stacked deeply with interspersed max-pooling layers to progressively reduce dimensionality and increase the network's depth, crucial for capturing complex patterns in the data. Padding is consistently set to 1 in convolutional layers to maintain the spatial dimensions after convolution operations.

Each convolutional and fully connected layer is followed by a ReLU activation function to introduce non-linearity, enhancing the network's capacity to learn diverse and complex image representations. The final layers are adapted to output three classes, corresponding to the datasets employed in the study, with adjustments in the fully connected layer to map the deep features to these target outputs.

ResNet-50 Configuration

ResNet-50's setup includes the same normalization parameters, facilitating effective standardization of input images. The core innovation in ResNet-50, utilized in this project, is the integration of skip connections that bypass one or more layers. These connections allow gradients to flow through the network directly, alleviating the vanishing gradient problem in deep networks and enabling substantial depth without degradation in performance. The model’s input layer accepts resized images of 64x64 pixels, and the network architecture is detailed with bottleneck layers comprising 3x3 convolutions situated between 1x1 convolutions for dimensionality reduction and expansion. This design is optimized for computational efficiency and effective learning. ResNet-50 concludes with a global average pooling layer followed by a fully connected layer, tailored to classify the images into three categories as per the dataset labels.

The training of these models employs a standard Adam optimizer with a learning rate of 0.001, balancing fast convergence and stability in updates. Each model is trained and evaluated using custom DataLoader instances that shuffle the input data for robustness against ordering biases in the training process. This setup not only ensures that each model is optimized for the peculiarities of its architecture but also standardizes key aspects like image preprocessing and batch processing, making comparative performance analysis consistent and reliable.

Results

1. Accuracy observed across different dataset combinations:

PascalVOC	Flicker_30K	Imagenet(64x64)	Accuracy(%)
✔	✔		63.20
✔		✔	90.99

	✔	✔	92.83
✔	✔	✔	78.08

2. Accuracy observed across different Model Architechtures

Model	**Accuracy**
AlexNet	78.08
VGG-16	50.00
RestNet-50	85.68

3. Accuracy observed across different Training Data Sizes

Model	**Size**	**Accuracy(%)**
AlexNet	10K	59.76
	20K	78.08
	29K	85.85
RestNet-50	10K	79.45
	20K	85.68
	29K	84.06

4. Accuracy over Different corruptions

Corruption (Train+Val)	**Accuracy(%)**
None	78.08
Color Jittering(strength:0:75)	00.00
Gaussian Noise (std:0.13)	50.00
Gaussian Blur(radius:1.5)	65.00

Low Resolution(32x32)	00.00
Enhanced Resolution(227x227)	50.00

5. Data classification accuracy for data augmentation

Augmentations	Training Size	Accuracy(%)
No Aug	10k	59.76
	20k	78.08
	29k	85.85
RandCrop	10k	42.44
	20k	74.19
	29k	77.85
RandCrop+RandAug	10k	53.37
	20k	77.59
	29k	79.90

6. Data Classification Accuracy over Different Model sizes: (Model: AlexNet)

Model Size(Number of Parameters)	Accuracy
15K	88.28
30K	83.95
54M	78.08
70M	78.73