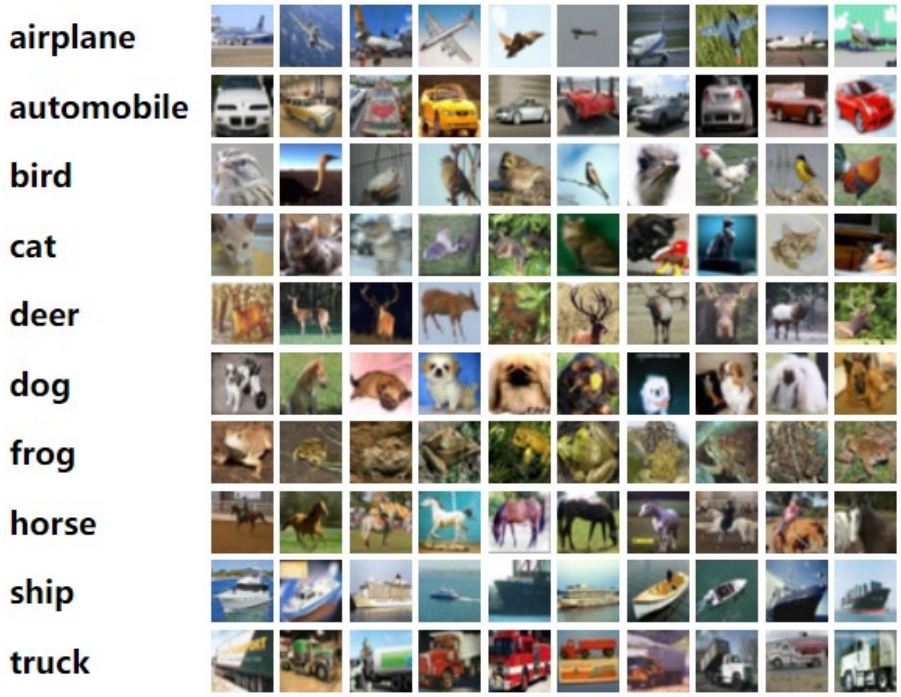
AI 2024, Assignment 5:

Vision Transformers (ViT)

In this assignment, you will explore Vision Transformers (ViT) to process the CIFAR-10 dataset. The CIFAR-10 dataset consists of 60,000 32x32 color images categorized into 10 different classes.

CIFAR-10

https://www.cs.toronto.edu/~kriz/cifar.html



Task Requirements:

1. Model Selection and Design: Implement the Vision Transformer model using TensorFlow or PyTorch. Ensure that the architecture is correctly adapted for the CIFAR-10 dataset.
2. Train the Models: Train the Vision Transformer model on the CIFAR-10 dataset. Experiment with different hyperparameters such as learning rate, batch size, and number of epochs to optimize performance.
3. Performance Evaluation: Assess the model's performance on the test set. Report on metrics such as accuracy, precision, recall, and F1 score.
4. Results Visualization: Plot the accuracy and loss curves during training, and display some examples of misclassified images.

Deliverables:

1. The code implementation with appropriate comments.
2. A detailed report on the results, including model accuracy, loss, and a discussion on the performance of the Vision Transformer model.
3. Relevant charts, such as training curves and misclassification examples.