Zewail City for science and technology
Interpretability & explainability in AI DSAI305
Spring 2024

Mid-Progress Report

Sign Language Recognition

1 - Preprocessing:

- **Resize**
  Images are resized to 128 X 128.

- **Label-Encoding+One-hot encode labels**
  Label encoding was first used to convert categorical labels to numerical form, followed by one-hot encoding, as not all labels are integers (string labels).

- **Augmentation**
  Rotation, width and height shift, and horizontal flip.

2 - Models:

- **InceptionResNetV2:**

  1. Transfer Learning was applied: All layers were frozen initially, and only top layers (block8 and conv_7b) were unfrozen for fine-tuning.
  2. Evaluation: After training, the model was evaluated on a validation set (20% split) and achieved validation accuracy: 85% and validation loss:0.4207
  3. Added a custom classifier head: GlobalAveragePooling2D → Dropout → Dense(1024, relu) → Dense(27, softmax) then compiled with lr (1e-5) and trained for 10 epochs on augmented data.

- **VGG-19:**

  ○ Architecture: Using pre-trained VGG19 convolutional base. Custom fully connected layers were added to adapt the model for the target classification task.
  ○ Data: The training utilized RGB images, covering 27 distinct sign language characters.
  ○ Training Protocol: The model was trained for 10 iterative cycles (epochs) to optimize performance.

  Accuracy :67.95%

Observations:
Validation Accuracy Progression: Improved steadily from 60.71% (Epoch 1) to a peak of 69.35% (Epoch 7), followed by stabilization.

- **ResNet50:**

  Freezing most original layers while fine-tuning the last 20

  1. Adding new customized classification layers (global pooling → 1024 dense → dropout → 512 dense → output)
  2. Training with:
     - Low learning rate (0.0001)
     - Adam optimizer
     - Early stopping

The approach combines ResNet50's general architecture with customized

layers for image classification and partial fine-tuning

- **Challenges:**

  - Limited GPU power.

  - Training is slow on larger models; monitoring memory

    usage is essential.

  - Slow Training: Fine-tuning even partially unfrozen

    models takes a significant amount of time.