Training Configurations for Synthetic pre trained weights(Models trained on webots images)

Multi Class

Dataset

- Dataset type: Multi-class image classification
- Number of classes: 7
- Input image type: Grayscale (converted to 3-channel for ResNet)
- Image size: 224 × 224
- Data loading: Shuffled each epoch

Transforms / Preprocessing

- Resize(224 × 224)
- ToTensor()
- Normalize(mean = 0.5, std = 0.5)
- Grayscale images duplicated across 3 channels to fit ResNet50 input

Training Hyperparameters

- Epochs: 30
- Batch size: 16
- Optimizer: Adam
    - Learning rate: 1×10^{-4}
- Loss function: CrossEntropyLoss (multi-class)

Multi Label

Dataset & Labels

- Dataset type: Multi-label image classification
- Label encoding: Folder name encodes binary vector
    - Example: 1_0_1 → label = [1, 0, 1](left_right_forward)
- Number of labels: 3
- Image format: Grayscale → expanded to 3-channel
- Input size: 224 × 224
- Shuffling: Enabled per epoch (shuffle=True)
- num_workers: 2

Transforms (Preprocessing)

- Resize to 224×224
- Convert to tensor
- Normalize grayscale values with mean = 0.5, std = 0.5
- Repeat grayscale tensor across 3 channels

Training Hyperparameters

- Loss function: BCEWithLogitsLoss
- Optimizer: Adam
    - Learning rate: 1×10^{-4}
- Batch size: 16
- Epochs: 30
- Device: GPU if available (cuda)

Evaluation Metric

- Exact Match Accuracy

- Converts logits → probabilities using sigmoid
- Threshold: 0.5
- A prediction is correct only if *all* labels match exactly

Training Configurations for Multi Label Training(Models trained on real-world data)

Dataset

- **Task type:** Multi-label image classification
- **Label encoding:** Directory names represent binary vectors
  - Example: "1_0_1" → [1, 0, 1]
- **Number of labels:** 3
- **Image type:** Grayscale, replicated to 3 channels
- **Input resolution:** 224 × 224
- **Dataset split:** 5-fold K-Fold Cross-Validation (shuffle=True, random_state=42)

Data Augmentation & Preprocessing

- Resize(224 × 224)
- ColorJitter(brightness=0.2, contrast=0.2)
- ToTensor()
- Normalize(mean=0.5, std=0.5)
- Grayscale repeated to RGB (img.repeat(3,1,1))

Training Setup (per fold)

- **Optimizer:** Adam

    - Learning rate: 1×10^{-3}
- **Loss:** BCEWithLogitsLoss (multi-label)
- **Batch size:** 16
- **Epochs:** up to 30
- **Device:** GPU (cuda) when available
- **Early Stopping:**
    - Patience: 3 epochs
    - Minimum improvement (delta): 1e-4

Evaluation Metrics

- **Primary metric:** *Exact multi-label accuracy*
    - Thresholding: sigmoid > 0.5
    - Uses sklearn's accuracy_score on binary vectors
- Metrics tracked per epoch:
    - Training loss
    - Validation loss
    - Training accuracy
    - Validation accuracy

Cross-Validation

- Number of folds: 5
- Best validation accuracy recorded for each fold
- Final model = weights of fold with highest validation accuracy
- Reported results:

    - Mean accuracy across all folds
    - Best fold accuracy

Grad-CAM Configurations for Explainability(Multi Label)

Target Layers (Grad-CAM)

- ResNet-50:
    - layer4[-1] (final Bottleneck block)
- MobileNetV2:
    - features[-1] (final convolutional block)

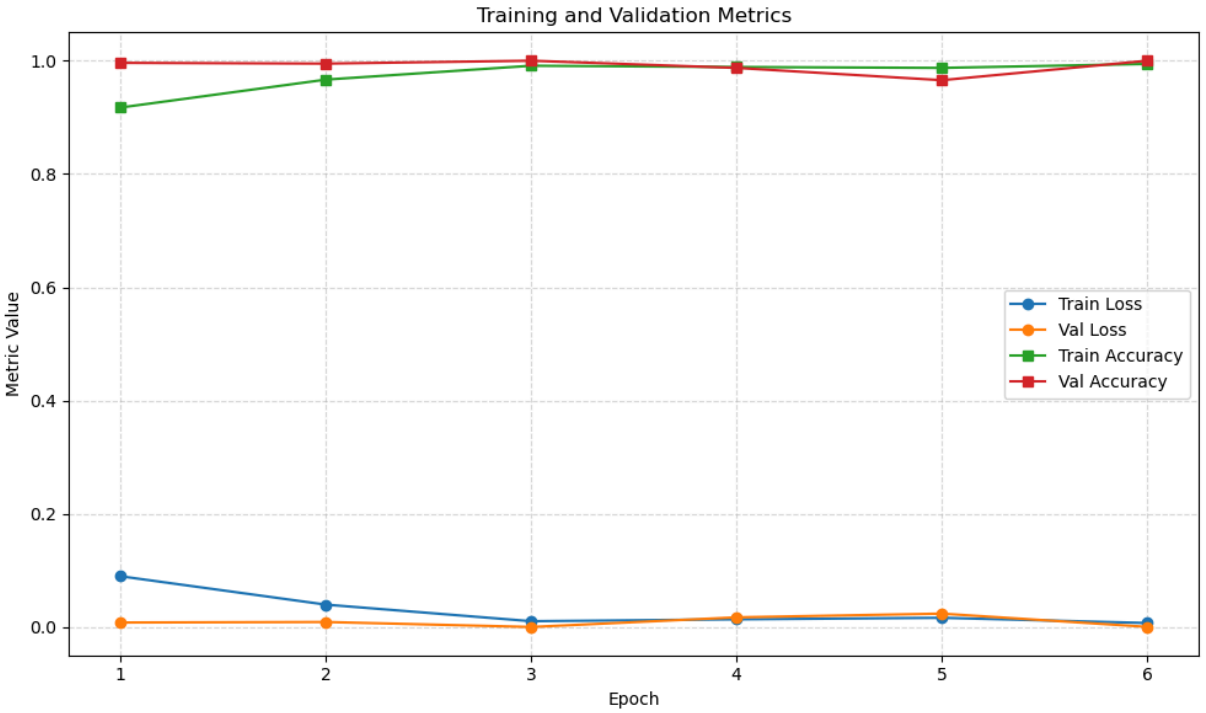
Input Preprocessing

- Images are loaded in grayscale (convert('L'))
- Resized to 224 × 224
- Normalization:
    - Mean: 0.5
    - Std: 0.5
- Grayscale expanded to 3-channel (by repeating the channel)
- Final tensor shape for each image: [1, 3, 224, 224]

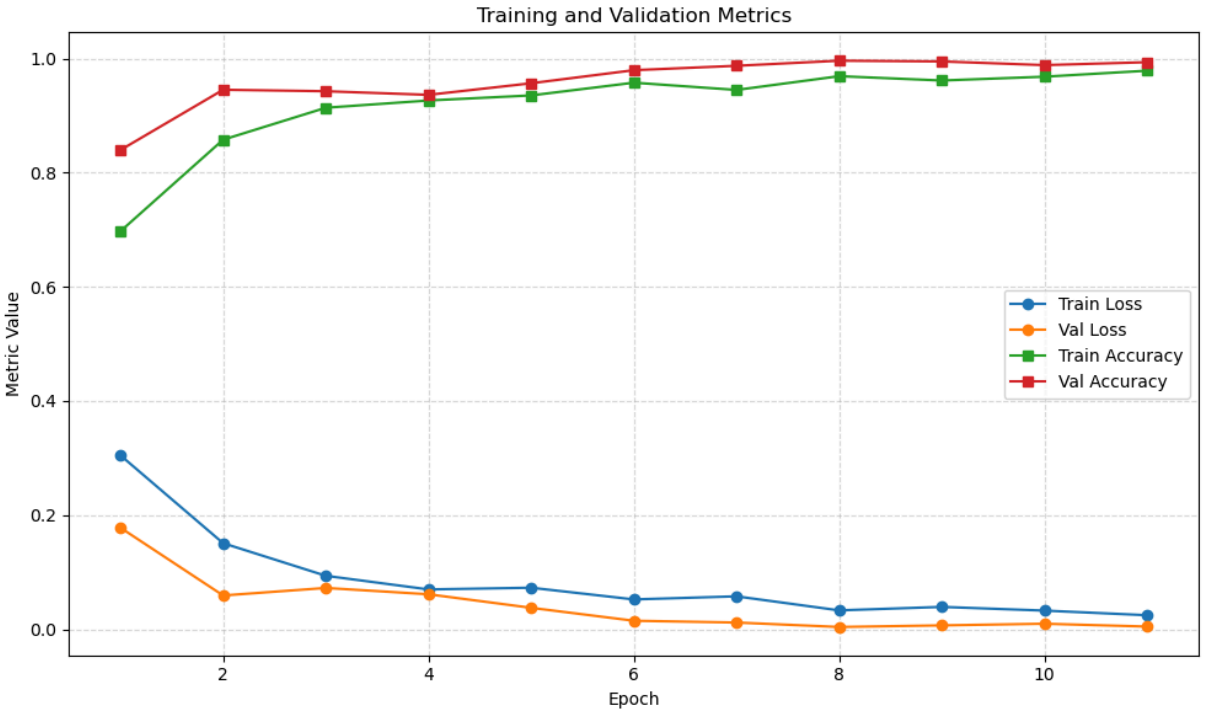
Training and Validation Curves(Only the highest validation performance model)

He Initialization

ImageNet Initialization



Synthetic(last 2) Initialization



Synthetic(all) Initialization