Report: Custom ResNet50V2 Architecture with GCBlock and Dropout

1. Introduction

I explored various architectural enhancements based on ResNet50V2 to improve image classification accuracy.
The primary goal is to achieve higher validation accuracy through added attention mechanisms, freezing strategies, and regularization techniques such as dropout.

The backbone I chose ResNet50V2 architecture is a refinement of the original ResNet50, introduced in He et al., ECCV 2016 [1]. The key differences include pre-activation residual blocks, improved identity mapping, and better training stability.

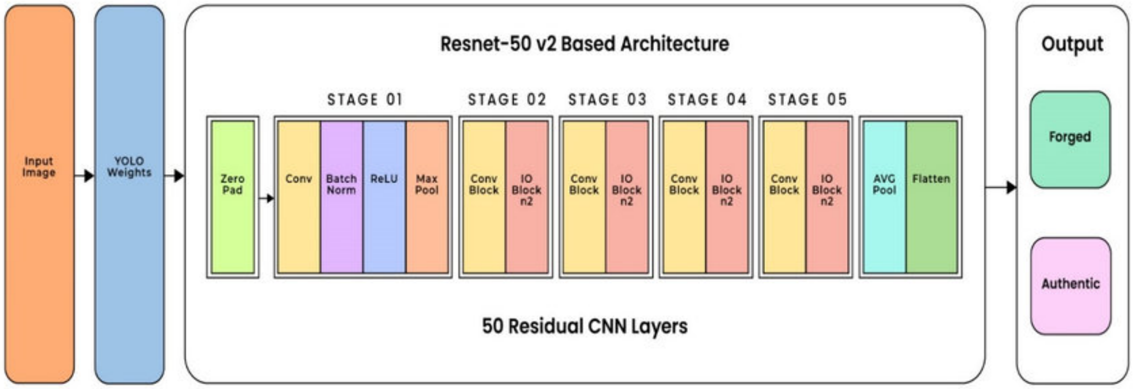

Figure: ResNet50V2 architecture overview. Source: https://www.researchgate.net/publication/308277201_Identity_Mappings_in_Deep_Residual_Networks

My best-performing experimental architecture is:

ResNet50V2 + Frz(stem+stage1) + GCBlock(stage3+stage4) + Dropout(0.3)

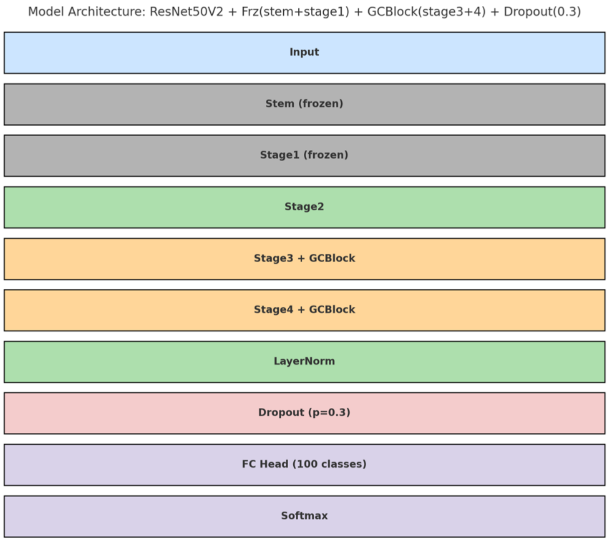

Figure: My custom architecture block diagram with GCBlock and Dropout(p=0.3)

This approach leverages pretrained backbone features while adding GCBlocks for contextual attention and applying dropout for regularization to improve accuracy.

GitLink: https://github.com/IaTsai/DL-HW1.git

2. Method

Data Preprocessing:

- Resize to 600x600
- RandomCrop to 224x224
- Data augmentation: Horizontal flip, rotation, color jitter, affine
- Normalization using ImageNet mean and std

Architecture Overview:

- **Backbone**: ResNet50V2 (from TIMM library)
- **Freezing**: stem and stage1 layers are frozen to retain low-level pretrained features
- **Attention**: GCBlock inserted after stage3 and stage4
- **Dropout**: p=0.3 applied after normalization before classification head

- **Classifier**: 100-class head

Hyperparameters:

- Optimizer: SGD (lr=0.0001, momentum=0.9, weight_decay=0.0001)
- Loss: CrossEntropy with class weighting
- Early Stopping: patience=70 epochs (optional)
- Epochs: 400
- Batch size: 128

3. Results

Best Model Performance (Dropout Version)

- **Train Accuracy**: ~0.9322 (at Epoch 400)
- **Validation Accuracy**: ~0.8467 (at Epoch 88)

Learning Curves:

- Validation accuracy plateaus and overfitting observed beyond 90 epochs
- Freezing stem + stage1 helps reduce overfitting

Ablation Summary:

Model Variant	Val Acc	Notes
ResNet50 (ImageNet pretrained)	~0.70	Baseline
ResNet50V2 (pretrained)	~0.75	Bit model, better features
ResNet50V2 + GCBlock(stage3+4)+Drop(p=0.3)	0.8467	Best performance
ResNet50V2 + GCBlock(stage3+4)+)+Drop(p=0.3)+CBAM(stage4	0.8367	Slightly worse than GC-only
ResNet50V2 + CoT Block(pretrained)	~0.75	Over-complicated
ResNet50V2 + ConvNeXt(stage4)	~0.75	Over-complicated
ResNet50V2 + ConvNeXt(stage3+4)	~0.68	Over-complicated
ResNet50V2 + ConvNeXt(stage4) + CBAM	~0.73	Over-complicated

4. References

- [1] He et al., "Identity Mappings in Deep Residual Networks", ECCV 2016 (ResNet V2)
- [2] Cao et al., "GCNet: Non-Local Networks Meet Squeeze-Excitation Networks and Beyond", CVPR 2020
- [3] Woo et al., "CBAM: Convolutional Block Attention Module", ECCV 2018
- [4] Liu et al., "ConvNeXt: A ConvNet for the 2020s", CVPR 2022
- [5] Li et al., "Contextual Transformer Networks for Visual Recognition", CVPR2021
- [6] Hu et al., "Squeeze-and-Excitation Networks", CVPR 2018
- [7] TIMM GitHub Repository
- [8] ResNet50V2 architecture figure

---

5. Additional Experiments

Architectures Tried:

- ResNet50 / ResNet101 (standard)
- ResNet50V2 + pretrained
- ResNet50V2 + GCBlock --(CVPR 2020)
- ResNet50V2 + GCBlock (stage3+4) + CBAM(stage4)
- ResNet50V2 + GCBlock (stage3+4) + Dropout(0.3) **[Best]**
- ResNet50V2 + CoTBlock --(CVPR 2021)
- ResNet50V2 + ConvNeXtBlock (stage3+4) --(CVPR2022)
- ResNet50V2 + ConvNeXtBlock (stage4 only)
- ResNet50V2 + ConvNeXtBlock (stage4 only) + CBAM

My Point of View:

- CBAM added to ResNet50V2 didn’t always improve performance (minor gains or no improvement).
- Freezing early layers helps generalization and speeds up training
- A dropout of 0.3 between norm and head offers noticeable stability
- Lowering the learning rate (0.0001) benefit for training stability with frozen layers