**Experiment Run**

**Experiment Run Report**

**Experiment Title:** Numerosity-Based Categorization - Experiment Run 3

**Date:** 24/02/2025

**Researcher:** Karoki Evans Njogu

**1. Experiment Details**

| **Parameter** | **Value** |
| --- | --- |
| Seed | 42 |
| Dataset Size | 5000 samples |
| Image Size | 128x128 pixels |
| Categories | Few (1-5), Medium (6-15), Many (>16) |
| Batch Size | 128 |
| Learning Rate | 0.0003 |
| Epochs | 20 |
| Optimizer | AdamW |
| Dropout Rate | 0.4 |
| Weight Decay | 5e-4 |
| Loss Function | CrossEntropyLoss |
| Early Stopping | Yes (Patience = 5) |
| Device Used | GPU – NVIDIA L4 |

**2. Experiment Setup**

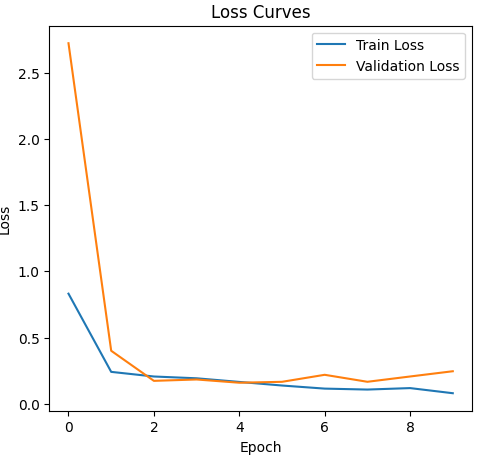
* **Dataset:** Synthetic Dot Patterns
* **Model Architecture:** Residual CNN with three convolutional layers and fully connected layers.
* **Training Strategy:**
  + Train on 70% of data.
  + Validate on 15%.
  + Test on 15%.
* **Evaluation Metrics:**
  + Accuracy
  + Loss Curves
  + Confusion Matrix
  + Precision, Recall, and F1-Score

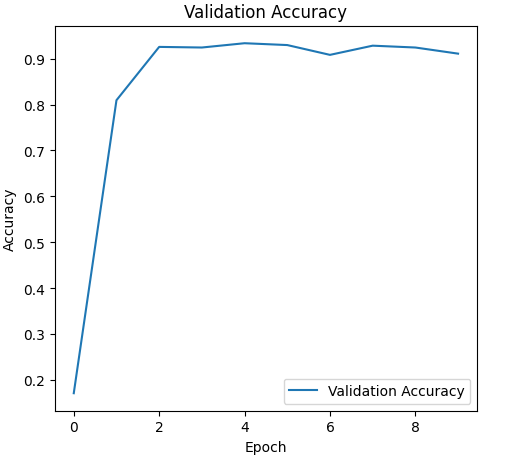
**3. Training & Validation Performance**

**3.1 Loss and Accuracy Trends**

| **Epoch** | **Train Loss** | **Validation Loss** | **Validation Accuracy (%)** |
| --- | --- | --- | --- |
| 1 | 0.8312 | 2.7220 | 17.20% |
| 2 | 0.2409 | 0.4002 | 80.93% |
| 3 | 0.2059 | 0.1730 | 92.53% |
| 4 | 0.1924 | 0.1834 | 92.40% |
| 5 | 0.1650 | 0.1589 | 93.33% |
| 6 | 0.1378 | 0.1658 | 92.93% |
| 7 | 0.1145 | 0.2187 | 90.80% |
| 8 | 0.1073 | 0.1660 | 92.80% |
| 9 | 0.1184 | 0.2060 | 92.40% |
| 10 | 0.0797 | 0.2456 | 91.07% |

**3.2 Loss Curve & Accuracy Plot**

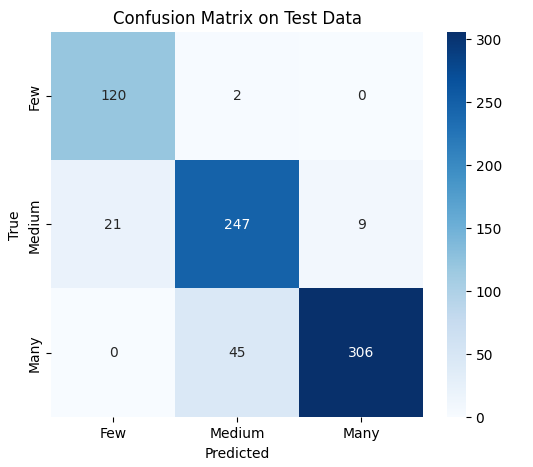




**4. Test Set Evaluation**

**Final Test Accuracy:** 89.73%

**4.1 Confusion Matrix**

****

**4.2 Classification Report**

| **Class** | **Precision** | **Recall** | **F1-Score** | **Support** |
| --- | --- | --- | --- | --- |
| Few | 0.85 | 0.98 | 0.91 | 122 |
| Medium | 0.84 | 0.89 | 0.87 | 277 |
| Many | 0.97 | 0.87 | 0.92 | 351 |

**5. Observations & Insights**

* **Key Findings:**
  + The training loss remained relatively stable, but validation loss showed fluctuations, especially in the later epochs.
  + The validation accuracy peaked early and then slightly declined, indicating possible early stopping effectiveness.
* **Error Analysis:**

A collage of white circles

AI-generated content may be incorrect.

* + Few category maintained high recall (98%), similar to previous runs.
  + Medium class saw a slight improvement in recall (89%), reducing misclassification.
  + Many class exhibited a trade-off: improved precision (97%) but lower recall (87%), suggesting more cautious predictions.
* **Next Steps:**
  + Evaluate alternative dropout rates (e.g., 0.35) to balance regularization and performance.
  + Fine-tune weight decay to minimize unnecessary penalization.
  + Increase dropout to regularize the model further.

**6. Conclusion**

This run demonstrated the effectiveness of using AdamW, weight decay, and dropout in improving classification performance. However, some overfitting and class confusion still persist. The next run will focus on optimizing regularization strategies further to enhance generalization.

This run demonstrated the increase in batch size appears to have led to smoother training dynamics, but at the cost of slightly increased test loss. The dropout increase and weight decay adjustments helped reduce overfitting, but the model is slightly more conservative in its predictions.

**7. Additional Notes**

* Reproducibility was ensured by setting a fixed random seed and using pre-saved datasets.
* This run also followed the structured experiment template, making future runs easy to compare.
* Some variability in validation loss was observed, which may indicate the need for better regularization techniques.
* Early stopping was applied, preventing overfitting, but further adjustments may be needed.