Experiment Run

Experiment Run Report

Experiment Title: Numerosity-Based Categorization - Experiment Run 2

Date: 21/02/2025

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	64
Learning Rate	0.0005
Epochs	20
Optimizer	AdamW
Dropout Rate	0.3
Weight Decay	1e-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

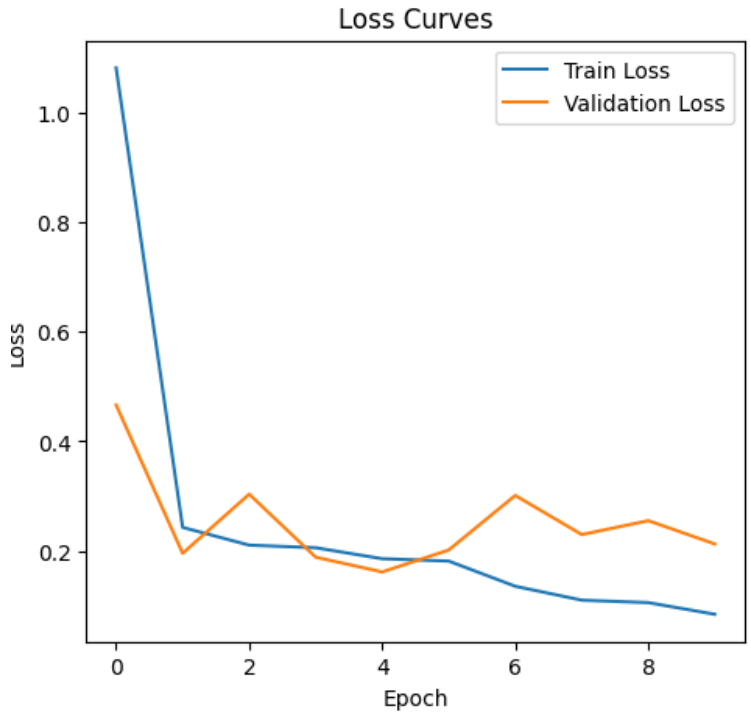
---

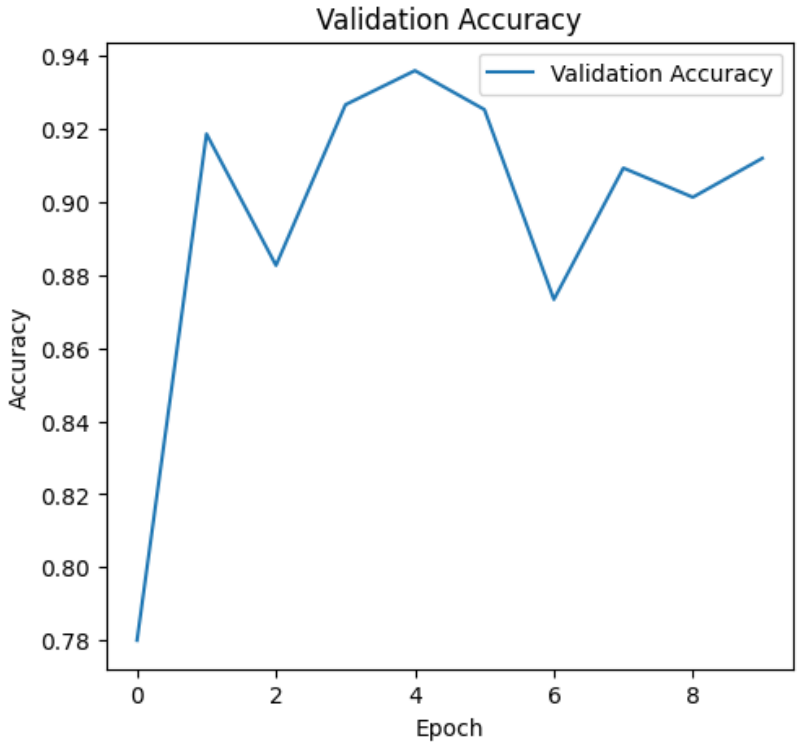
3. Training & Validation Performance

3.1 Loss and Accuracy Trends

Epoch	Train Loss	Validation Loss	Validation Accuracy (%)
1	1.0817	0.4664	78.00%
2	0.2430	0.1957	91.87%
3	0.2106	0.3036	88.27%
4	0.2058	0.1886	92.67%
5	0.1858	0.1617	93.60%
6	0.1815	0.2017	92.53%
7	0.1355	0.3013	87.33%
8	0.1102	0.2299	90.93%
9	0.1056	0.2553	90.13%
10	0.0844	0.2127	91.20%

3.2 Loss Curve & Accuracy Plot



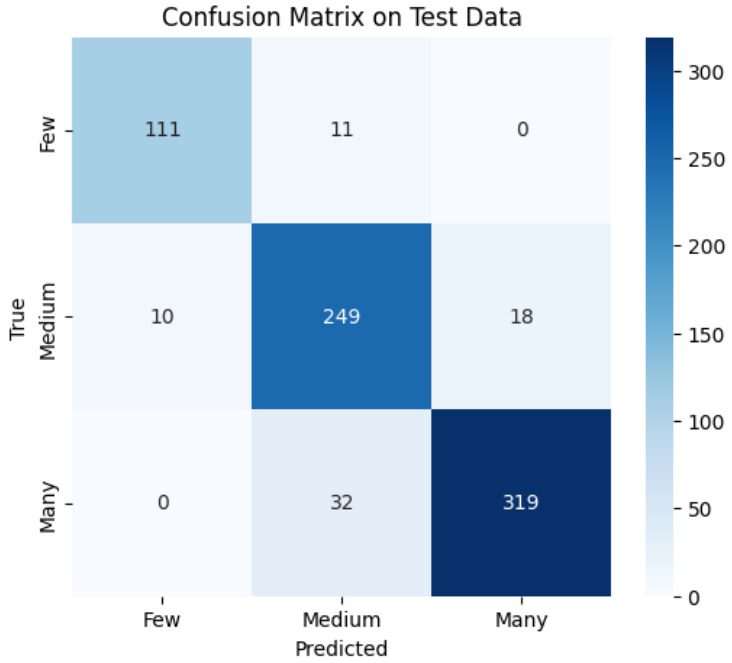


---

4. Test Set Evaluation

Final Test Accuracy: 90.53%

4.1 Confusion Matrix



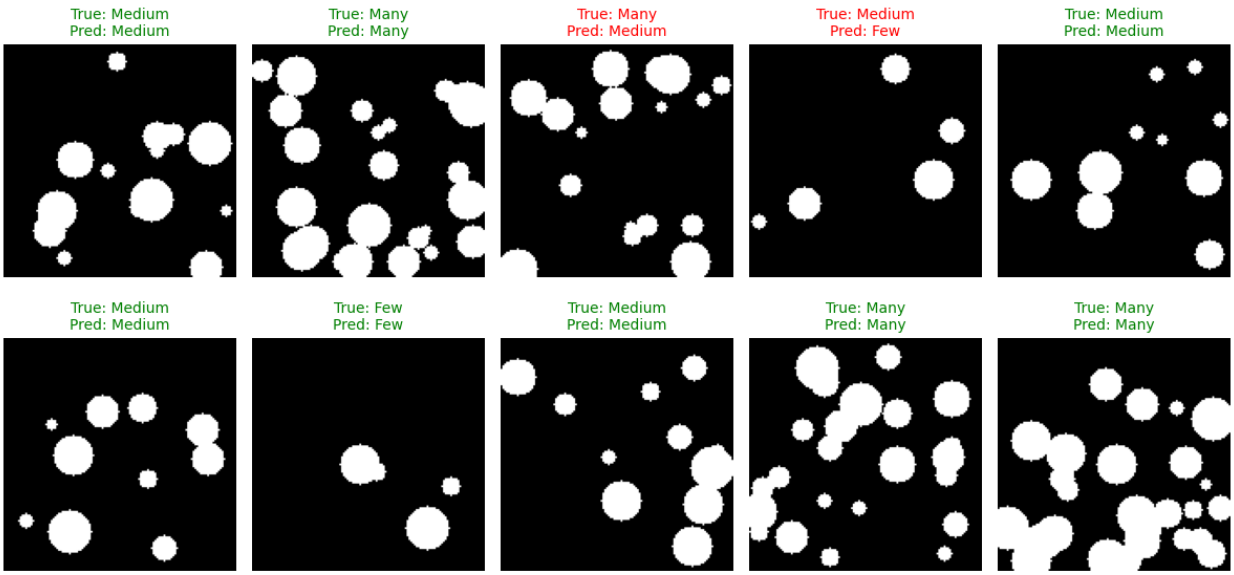
4.2 Classification Report

Class	Precision	Recall	F1-Score	Support
Few	0.92	0.91	0.91	122
Medium	0.85	0.90	0.88	277
Many	0.95	0.91	0.93	351

---

5. Observations & Insights

- **Key Findings:**
  - The model achieved 91.20% validation accuracy, with a 90.53% test accuracy, showing a stable improvement over Run 1.
  - The dropout rate of 0.3 helped reduce overfitting compared to previous runs.
  - The AdamW optimizer contributed to smoother weight updates and stabilization.
- **Error Analysis:**



  - The Medium category remains the most misclassified class, though recall improved compared to previous runs.
  - Some confusion still exists between Medium and Many categories.
  - Training loss decreases smoothly, but validation loss fluctuates slightly, indicating minor overfitting.

- **Next Steps:**
  - Reduce learning rate to further stabilize training.
  - Increase batch size for better generalization.
  - Increase dropout to regularize the model further.
  - Slightly increase weight decay to 5e-4 to encourage smaller weights.

---

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.

---

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.
