**Experiment Run**

**Experiment Run Report**

**Experiment Title:** Numerosity-Based Categorization – Silhouettes Dataset

**Date:** 4/04/2025

**Researcher:** Karoki Evans Njogu

**1. Experiment Details**

| **Parameter** | **Value** |
| --- | --- |
| Seed | 42 |
| Dataset Size | 3000 samples |
| Image Size | 128x128 pixels |
| Categories | Few (1-5), Medium (6-15), Many (>16) |
| Batch Size | 256 |
| Learning Rate | 0.0001 |
| 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 |
| eps | 1e-6 |
| betas | 0.9, 0.98 |
| Accumulation steps | 2 |

**2. Experiment Setup**

* **Dataset:** Synthetic Dot Patterns
* **Model Architecture:** CNN-Transformer architecture
* **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 | 0.8440 | 2.6334 | 17.78% |
| 2 | 0.5575 | 3.5014 | 18.00% |
| 3 | 0.5048 | 2.2426 | 28.00% |
| 4 | 0.4485 | 0.7734 | 66.22% |
| 5 | 0.4696 | 0.5100 | 76.22% |
| 6 | 0.4386 | 0.4345 | 78.89% |
| 7 | 0.4046 | 0.4297 | 78.89% |
| 8 | 0.3930 | 0.4684 | 79.11% |
| 9 | 0.4004 | 0.4536 | 79.33% |
| 10 | 0.3511 | 0.4137 | 81.78% |
| 11 | 0.3269 | 0.4871 | 78.67% |
| 12 | 0.3652 | 0.5743 | 76.67% |
| 13 | 0.3480 | 0.5269 | 78.89% |
| 14 | 0.2800 | 0.4147 | 82.89% |
| 15 | 0.2442 | 0.4251 | 81.56% |

**3.2 Loss Curve & Accuracy Plot**

A graph of loss curves

AI-generated content may be incorrect.

A graph with a line

AI-generated content may be incorrect.

**4. Test Set Evaluation**

**Final Test Accuracy:** 82.22%

**4.1 Confusion Matrix**

**A blue squares with white text

AI-generated content may be incorrect.**

**4.2 Classification Report**

| **Class** | **Precision** | **Recall** | **F1-Score** | **Support** |
| --- | --- | --- | --- | --- |
| Few | 0.83 | 0.89 | 0.86 | 71 |
| Medium | 0.74 | 0.75 | 0.75 | 156 |
| Many | 0.88 | 0.85 | 0.87 | 351 |

**5. Observations & Insights**

* **Key Findings:**
  + Model performance improved noticeably over the CNN-only setup.
  + Validation accuracy steadily increased, peaking at 82.89%, indicating good learning and generalization.
  + The final test accuracy is 82.22%, a solid result for categorizing abstract numerosity from silhouettes.
  + The loss trends in both training and validation loss curves are smoother, indicating better convergence and reduced overfitting.
  + All three classes ("Few", "Medium", "Many") were correctly learned to some degree, with the model being particularly strong on the "Few" and "Many" categories.
  + Improved Generalization: The CNN+Transformer model generalizes better than the CNN-only version. Despite earlier training instability (spikes in validation loss), performance stabilized and improved in later epochs.
  + Strong Accuracy in Extremes:
    - Few (Precision: 0.83, Recall: 0.89): High precision and recall suggest the model can easily recognize small object counts.
    - Many (Precision: 0.88, Recall: 0.85): The model reliably identifies cluttered scenes with high object counts.
  + Moderate Handling of Medium:
    - Medium (Precision: 0.74, Recall: 0.75) is slightly weaker, reflecting natural ambiguity in mid-range quantities, which are less visually distinct.
* **Error Analysis:**

A collage of images

AI-generated content may be incorrect.

* + Confusion between "Medium" and both extremes:
    - 16.67% of "Medium" samples were misclassified as "Many".
    - 8.33% were misclassified as "Few".
  + "Many" misclassified as "Medium" (14.8%): These errors could stem from object overlap, shape complexity, or edge occlusion making dense images appear less populated.
  + No "Few" misclassified as "Many": Indicates model learned a strong separation between lowest and highest counts.

**6. Conclusion**

The CNN+Transformer model demonstrates strong capability in abstract numerosity categorization, achieving over 82% accuracy on complex silhouette compositions. Compared to the CNN-only version, it shows:

* Improved accuracy in both validation and test sets.
* Better generalization across numerosity levels, especially at extremes.
* Moderate errors in ambiguous middle cases, which is a known challenge in cognitive-inspired quantification.