Experiment: Evaluation of the Pixel-Ratio Hypothesis in Numerosity-Based Categorization

Objective: This experiment was to evaluate whether neural networks trained to classify numerosity categories ("few," "medium," "many") are truly learning numerosity or merely exploiting low-level pixel-based features, such as the proportion of white pixels in the input images.

The hypothesis we were testing is:

"The model could be learning to categorize based on the ratio of white pixels (i.e., pixel density), rather than learning the abstract concept of numerosity."

To test this, we created a synthetic pixel-ratio dataset and trained models on it, and moved on to evaluate whether they could generalize to more abstract numerosity datasets composed of dots or object silhouettes.

Datasets:

1. **Pixel Ratio Dataset**
   - Synthetic images with three numerosity levels determined by pixel density:
     - Few: 20% white pixels
     - Medium: 50% white pixels
     - Many: 80% white pixels
   - Patterns: vertical, horizontal, checkerbox, and random

2. **Dot Dataset**
   - Images containing varying numbers of randomly placed white dots on a black background

3. **Silhouette Dataset**
   - Images composed of composited white object silhouettes (e.g., animals, tools) with different object counts

Two models:

- CNN (Convolutional Neural Network)
- CNN+Transformer hybrid model

Each model was trained on one dataset and tested across the others to evaluate generalization.

Results Summary:

Train Dataset	Test Dataset	Model	Test Accuracy	Note
Pixel-Ratio	Pixel-Ratio	CNN	100.00%	Perfect fit
Pixel-Ratio	Pixel-Ratio	CNN+Transformer	100.00%	Perfect fit
Pixel-Ratio	Dot	CNN	16.27%	All predicted "Few"
Pixel-Ratio	Dot	CNN+Transformer	16.27%	All predicted "Few"
Pixel-Ratio	Silhouette	CNN	15.78%	Predicted "Few" for all
Pixel-Ratio	Silhouette	CNN+Transformer	18.89%	Slight improvement
Dot	Pixel-Ratio	CNN	53.33%	Moderate generalization
Dot	Pixel-Ratio	CNN+Transformer	56.11%	Moderate generalization
Silhouette	Pixel-Ratio	CNN	66.11%	Better generalization
Silhouette	Pixel-Ratio	CNN+Transformer	56.11%	Similar to CNN

Visually:

- Models trained on pixel-ratio data failed to differentiate between images with distinct object counts but similar white pixel ratios.
- They assigned the "Few" label to almost every test image in the dot and silhouette datasets.
- Conversely, models trained on dots and silhouettes generalized moderately well to pixel-ratio images, suggesting they capture more abstract notions of numerosity.

Conclusion:

Hypothesis rejected: The hypothesis that models learn numerosity by simply estimating white pixel density is rejected based on the following evidence:

- No generalization from pixel-ratio models to dot or silhouette datasets.
- Moderate to strong generalization from dot/silhouette-trained models to pixel-ratio images.

- All predictions defaulted to 'Few' regardless of actual object count.
- Visual evidence shows pixel-ratio models fail to detect differences in structured object-based layouts.
- Poor precision, recall, and F1-scores when pixel-ratio models are tested on true numerosity datasets.

I think with this findings we can highlight that genuine numerosity learning involves more than just low-level statistics and suggest that training on abstract object representations (dots, silhouettes) yields models that are more cognitively aligned and generalizable.