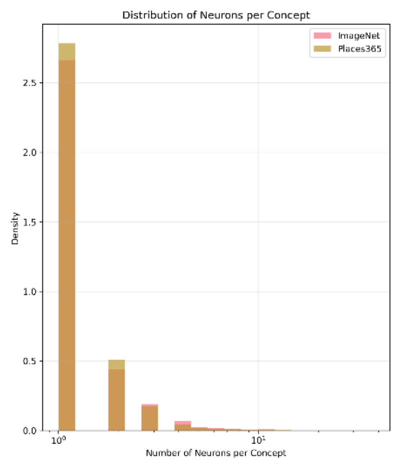
Task 1 Key Findings

- **Most Learned Concepts:** Both models frequently learn concepts like 'stripes', 'textile', and 'spots'.



  - **ImageNet:** 'stripes' (38 neurons), 'textile' (35 neurons), 'spots' (32 neurons).
  - **Places365:** 'textile' (37 neurons), 'spots' (35 neurons), 'stripes' (30 neurons).
  - The "Distribution of Neurons per Concept" plot shows most concepts are learned by a few neurons, with a select few having higher representation.
- **Concept Comparison:**
  - **Common Concepts:** 'textile', 'stripes', 'spots', 'kitchen', 'grid', 'chess', 'pattern', 'red', 'mesh', and 'bathroom' are frequently shared. For example, 'textile' has 35 neurons in ImageNet and 37 in Places365.
  - **Overlap:** Approximately **34.2%** of concepts are common to both models.
- **Unique Concepts:**
  - **ImageNet:** Learns **334 unique concepts**, with 154 exclusive ones (e.g., 'securely', 'acne'). These often relate to fine-grained objects.
  - **Places365:** Learns **373 unique concepts**, with 193 exclusive ones (e.g., 'trail', 'mobility'). These tend to indicate broader scene understanding.
- **Additional Insights:**

- **Neurons per Layer:** Both models have identical neuron distribution across layers (128 in layer2, 256 in layer3, 512 in layer4), indicating differences stem from training data, not architecture.
- **Concentration:** Neither model heavily concentrates neurons on a small set of concepts; knowledge is broadly distributed.

Conclusion

The analysis demonstrates that training data significantly shapes the learned representations in deep neural networks. While both ResNet18 models grasp fundamental visual elements, **ImageNet** excels in object-specific features, and **Places365** specializes in broader scene and environmental attributes. The consistent architectural neuron distribution highlights how the network's backbone adapts based on the training data's semantic focus.