Network Dissection Analysis (Task-1)

Models: ResNet18-ImageNet vs. ResNet18-Places365

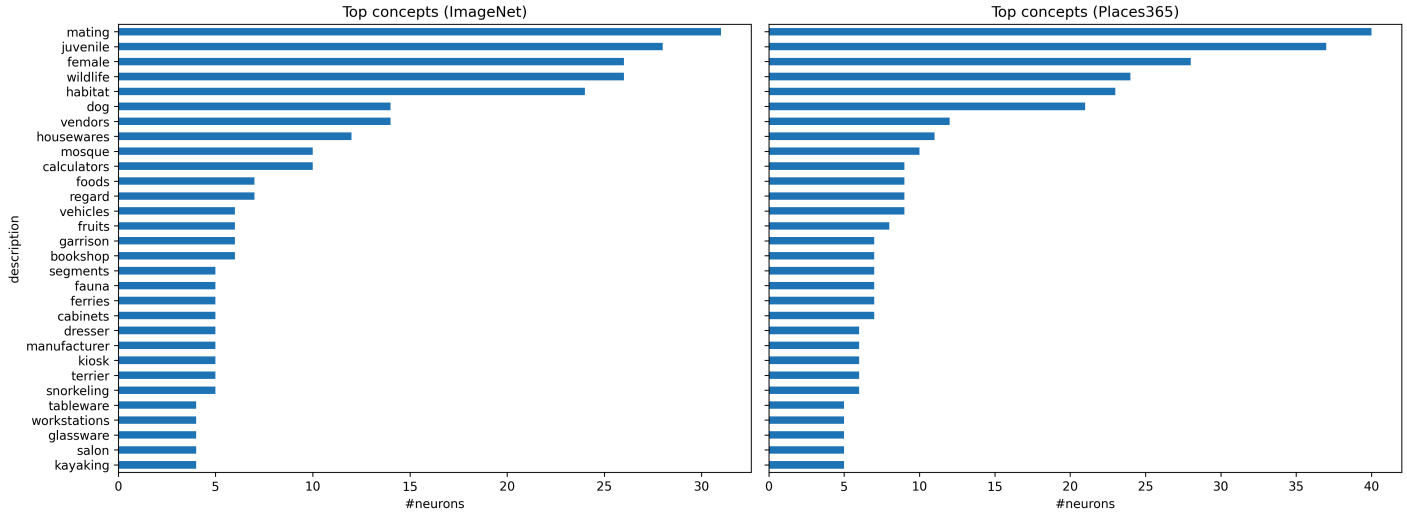
Layers: layer2, layer3, layer4

Probe: examples of images from ImageNet

Key Findings

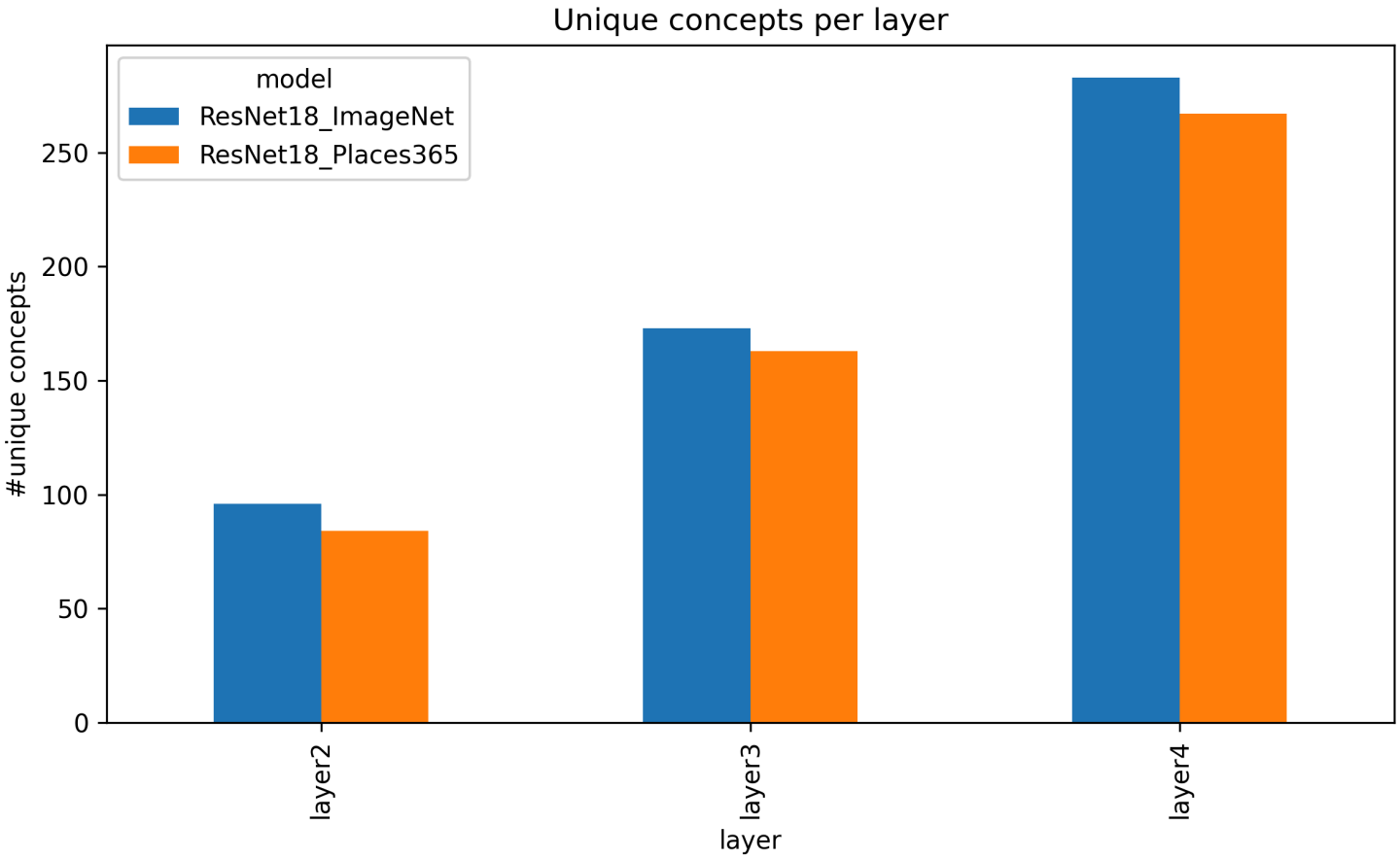
1. Dominant Concepts

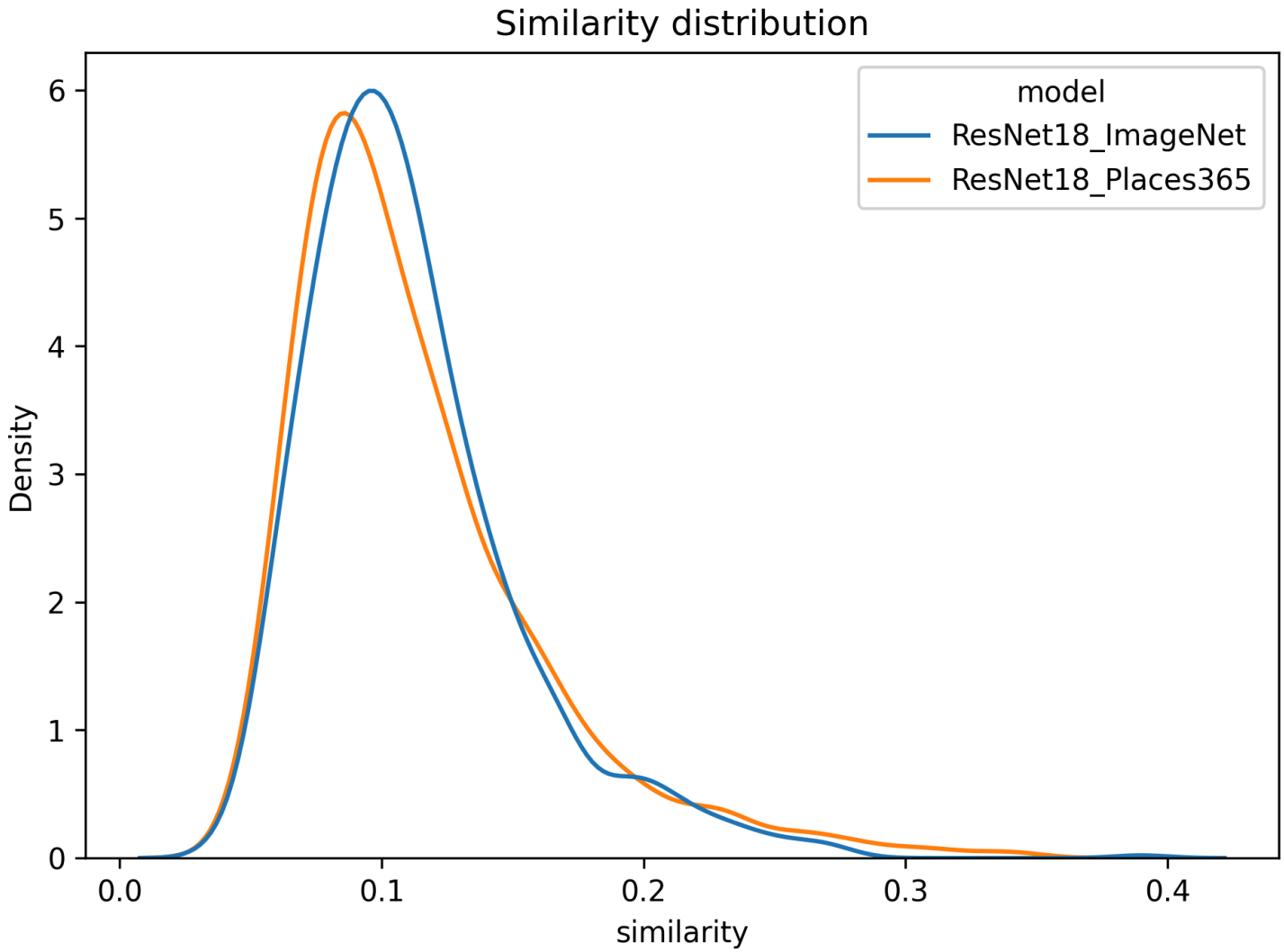
- **ImageNet:** frequent object/animal concepts like *mating, juvenile, dog, terrier, wildlife, habitat*.
- **Places365:** more scene/context terms like *female, housewares, foods, vendors, mosque*.



2. Model Comparison

- **Unique concepts:** 552 (ImageNet) vs. 514 (Places365).
- Both models show more unique concepts in deeper layers (e.g., layer4: 283 vs. 267).
- **Similarity:** Average similarity grows with depth (ImageNet: 0.095→0.120, Places365: 0.098→0.124).
- **Bias:** ImageNet favors fine-grained objects (e.g., terrier, snake), while Places365 favors scenes/attributes (e.g., housewares, vendors).


Unique concepts per layer
model
ResNet18_ImageNet
ResNet18_Places365
#unique concepts
0
50
100
150
200
250
layer2
layer3
layer4
layer



3. Concept Overlap

- Substantial overlap, but with dataset-specific emphasis. Places365's scene-centric nature is evident from its concept bias.

4. Additional Insights

- **Similarity distribution:** Both models have similar distributions, with Places365 showing slightly higher peaks at deep layers.
- **Layer trends:** Deeper layers focus on more semantic, abstract features.

Conclusion

Deeper layers in both models capture more semantic information, but ImageNet emphasizes objects/animals, whereas Places365 highlights scene and context concepts—reflecting their training data differences.