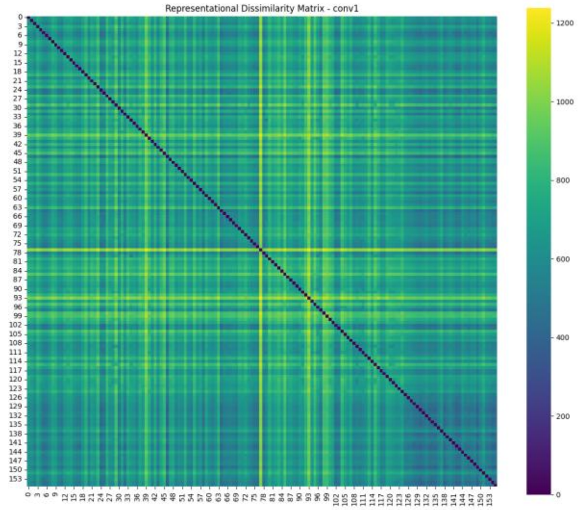
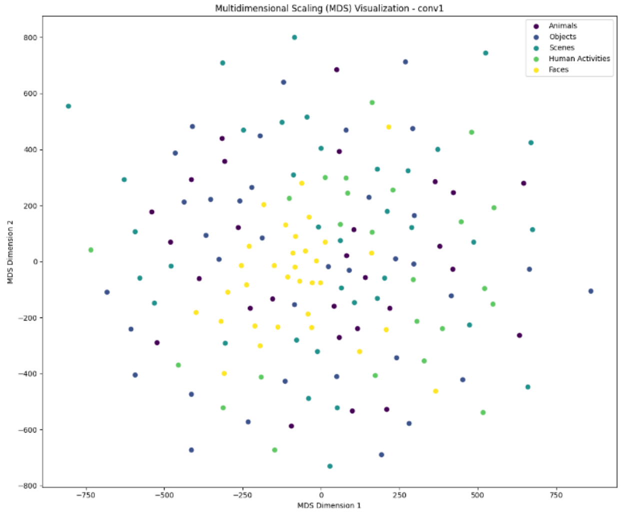
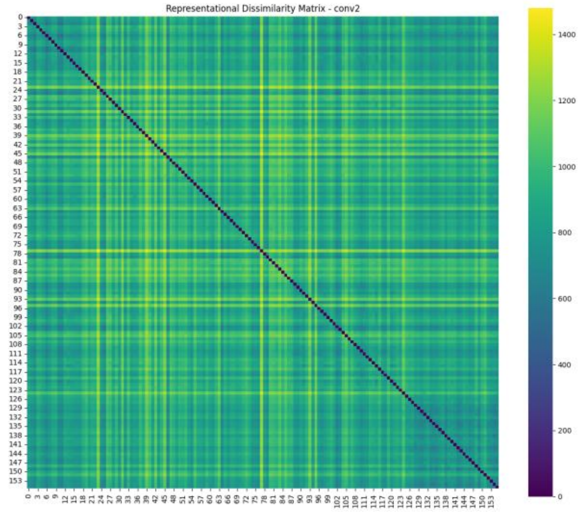
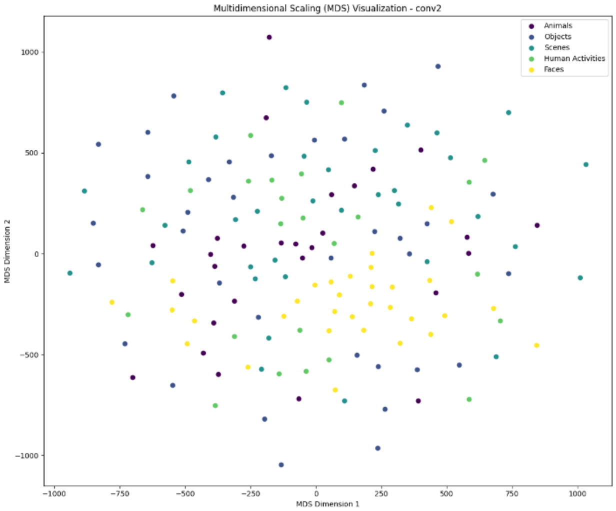
AlexNet Representational Dissimilarity Matrix Analysis

Early Layers (Conv1, Conv2)





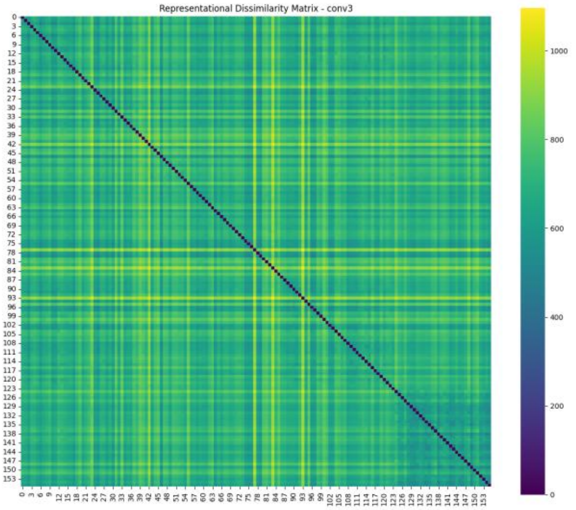


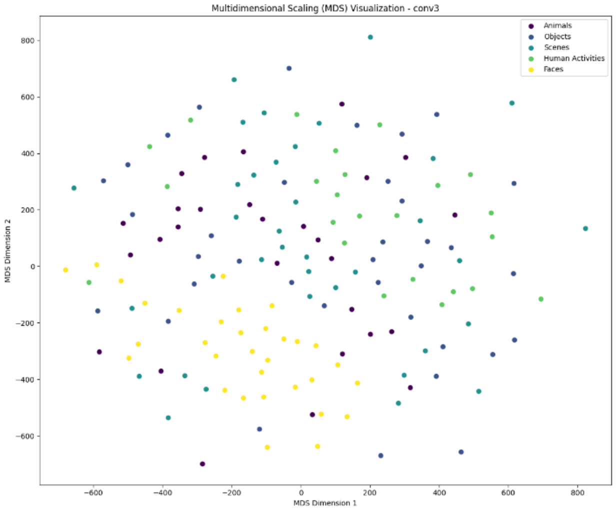


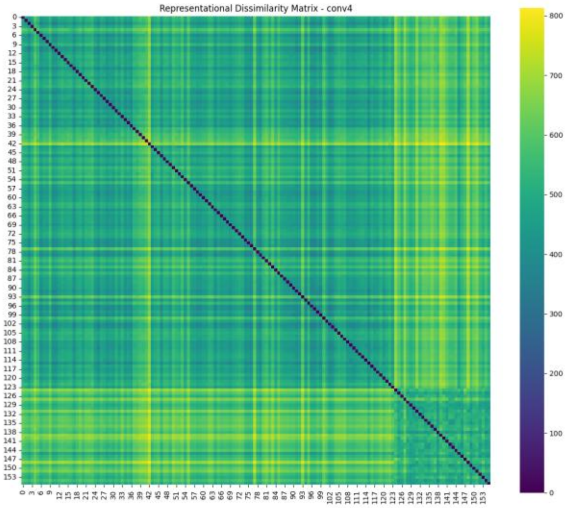
Early Layer RDM Heatmaps: The early layers of the network focus on basic visual features such as edges, colors, and textures. The RDM heatmaps mostly reflect these low-level details, so images from different categories may still look similar to each other at this stage. There's not much clear grouping by category yet.

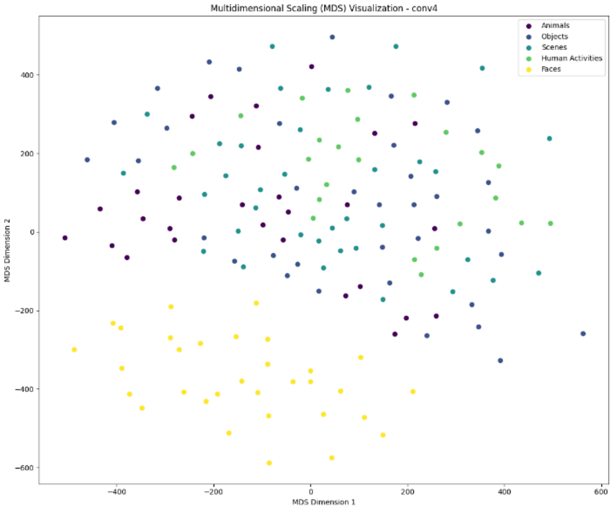
Early Layer MDS Visualization: MDS plots show how the representational space changes across the network’s hierarchy. In early layers, points are distributed based on low-level features like colour, contrast, and basic shapes, with limited categorical clustering.

Middle Layers (Conv3, Conv4)





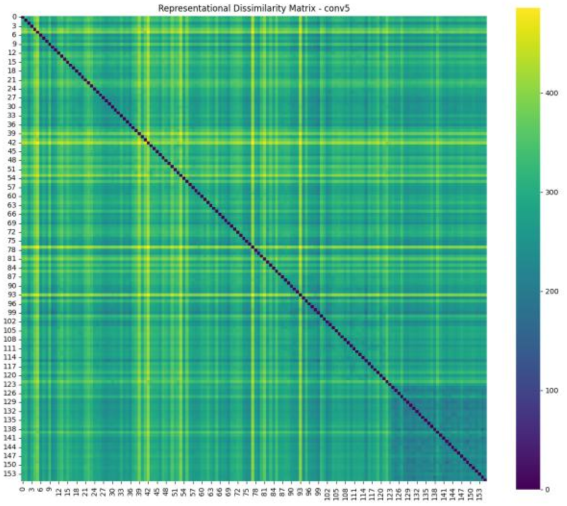


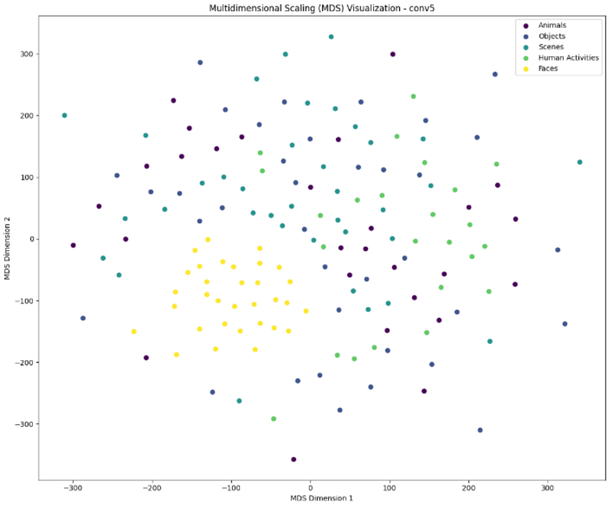


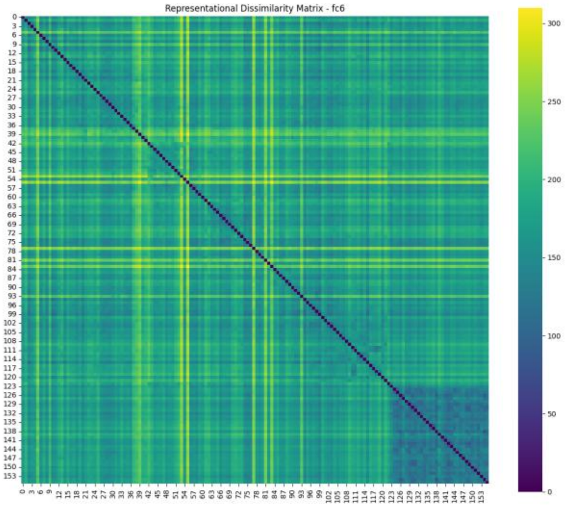
<u>Middle Layer RDM Heatmaps:</u> In the middle layers, some structure starts to appear. Images from the same category are starting to look more similar to each other, but the differences between categories are still not very clear.

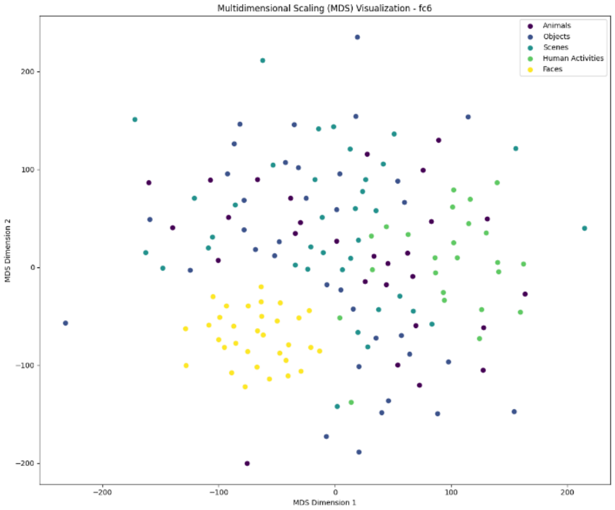
<u>Middle Layer MDS Visualization:</u> Intermediate representations begin to show some categorical organization, with partial clustering of similar categories.

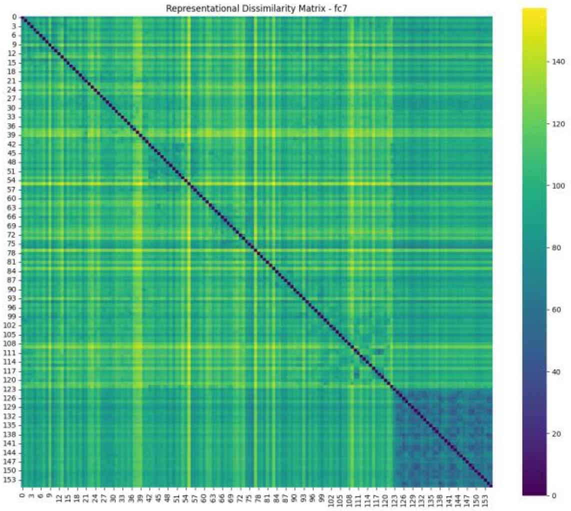
Deep Layers (Conv5, Fc6, Fc7)

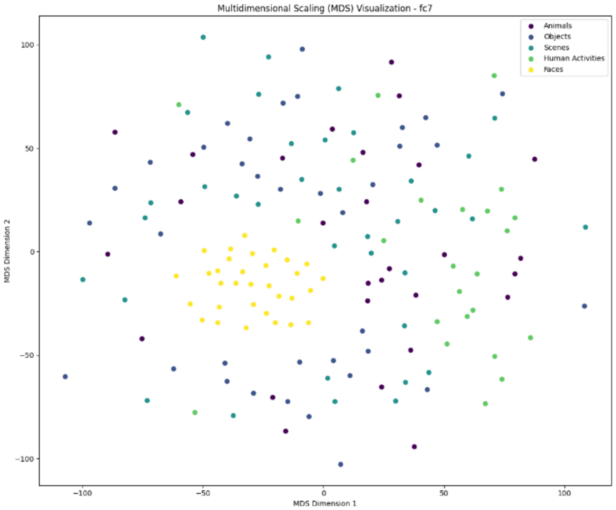












Deep Layer RDM Heatmaps: In the deep layers, clear patterns appear. Images from the same category look more similar to each other, and the differences between categories become much more noticeable.

Deep Layer MDS Visualization: Clear categorical boundaries emerge, with images from the same category clustering together. Face category typically shows the tightest clustering, suggesting strong representational consistency for facial images.

The progression from Conv1 to FC7 really showcases AlexNet's hierarchical feature extraction process. Early layers respond to basic visual properties, while deeper layers capture higher-level patterns that match categories we understand.

Category-Specific Observations

1. Faces: Consistently form tight clusters in later layers, suggesting the network develops specialized face representations.
2. Scenes: Scene images show more variety within the category, likely because they contain many different elements, like buildings, nature, or people, making them harder to group together and they reflect this by remaining spaced out throughout the layers.
3. Seems to be a barrier around the faces cluster suggesting that the model has a good grasp on discerning faces. However, it doesn’t really categorize anything else within such tight clusters.

Implications

This analysis shows how deep networks gradually turn raw images into more abstract representations that help distinguish categories. The RDM and MDS visuals reveal how the network's processing shifts from focusing on simple features to recognizing entire categories as it goes deeper.

Conclusion

The representational dissimilarity analysis of AlexNet shows a clear progression in how visual information is processed through the network. Like the human visual system, early layers focus on simple features, while deeper layers capture more abstract, category-level information. This suggests that deep networks can naturally learn to separate meaningful categories, even without being directly told what those categories are.