HML – Homework 6: Computational Model Paper Summary

**Introduction**

Researchers believe that humans are inherently sensitive to geometric and topological concepts. One of the key studies as referenced in the paper is the Dahane et al (2006) study which administered an odd one out task with the concept of oblique symmetry to an indigene amazonian group. The results showed that both adults and children were sensitive to 39/43 of those concepts which was a very encouraging result. Overall, the tasks consisted of 43 items and spanned 7 concepts namely, Topology, Euclidean Geometry, Geometrical Figures, Symmetrical Figures, Chiral Figures, Metric Properties, and Geometrical Transformations. Same study was also done on American adults and children and they showed a sensitivity closer to the amazonian tribe. The study done of much younger children less than 6 years of age shows a relatively lower sensitivity of 27/43 concepts but it can be attributed to developmental learning. The main downside of this study is only 1/6 is the correct guess i.e. 16.67% chance which seems a low bar to showcase sensitivity. To remedy this and increase the chance to 50% which is a higher bar to show sensitivity, the authors of this paper came up with an alternative approach and called it 2-AFC task. Throughout the work, the authors focus on CNN models which were trained on benchmark datasets like ImageNet and the aim is to check the sensitivity of these CNN models to GT concepts and if that sensitivity correlated to the one seen in humans. Previous studies have attempted this studies and the CNNs could not reliably discriminate between possible/impossible objects rather, they tend to rely on the local image information. Authors aim to answer questions like – Do CNN models in questions show Human-like sensitivity to those 43 concepts aggregated into 7 different classes? Does this sensitivity vary across the CNN layers? For example, is the sensitivity at an intermediate level lower or higher than at the final fully connected layer? Lastly, the accuracy of these CNN models matches closely with the human adults or children?

**Method**

The Authors shortlist 5 well known CNN models for testing like - AlexNet, VGG-19, VGG-16, ResNet-18, DenseNet-121, and GoogleNet. Every stimuli had 6 images and all the images were cropped and then rescaled to the dimensions 224 x 224 to be fed to the CNN models and the activation vectors were noted down from each of the layers. The measure of overall performance of the model was the predicted that were made by the final fully connected layer and not any of the intermediate levels. The computation of the odd one out was carried out by implementing and exploring various algorithms including Local Outlier Factor and K-Means clustering. Since the number of images in each stimuli was small, these complex methods made the system unstable and so the authors had to resort to simpler ways of cosine similarity to catch the odd one out. In that, for each of the 6 images, the cosine similarity was calculated against all the 5 other images and then the average cosine similarity was calculated with respect to the 5 other images. Each stimuli had 5 images that embodied a certain concept and 1 image that did not. The human dataset came from two previous cognitive science studies specifically the dataset for adults aged 18-25 years. The authors found interesting differences between the adults and children datasets as well which are considered during the experimentations.

**Results**

To answer the research questions raised in the introduction section –

1. CNN models show encouraging results that they are sensitive to the same GT concepts as humans. Although, the models were not able to match the high sensitivity levels of humans, the highest accuracy levels achieved were 40% achieved by ResNet-18. All the CNN models showed a noticeable lack of sensitivity especially towards Symmetrical Figures and Geometric Transformations.
2. The CNN model with the highest accuracy showed varying GT sensitivity at different layers and the accuracy was highest for the final fully connected layer except for stimuli class – Metric Properties and Geometric Transformations.
3. The Accuracy results from the CNN models closely emulated the humans with the ResNet-18 model showing accuracy levels closer to both adults and children. A very high similarity was seen between the DenseNet CNN model, and the 3-6 years data collected from Izard and Spelke (2009).

Overall, ResNet-18 performed the best and showed the highest accuracy levels – 40% of any CNN model. Even the highest accuracy levels are much worse than human adults or children which was 90.7%. ResNet-18 showed 75% accuracy on Euclidean Geometry concepts and the lease accuracy of 12.5% on Geometric transformations concepts. For other concepts, the accuracy levels were – Topology – 25%, Geometrical and Symmetrical Figures – 33.33%, Chiral Figures – 50%, and Metric Properties – 28.57%. The expectation is that the performance is poor at the initial layers and it improves and reaches the highest value at the final layers. ResNet-18 showed positive linear association for all classes except Metric properties class. ResNet-18’s accuracies were comparable to that of humans, and this was verified by the authors using Pearson correlation between the final fully connected layer of the model and humans. Although the overall accuracy of ResNet-18 is lower, it showed similar sensitivities to humans in that they both performed best on Euclidean Geometry class and worst on Geometrical Transformations class. The correlation value for ResNet-18 and children aged 3-6 is 0.52 and that with adults is 0.58.

**Discussion summary / conclusion**

Overall, the study answers some key questions about CNNs and how their sensitivity to Geometrical and Topological figures correlates to that of the human adults and children and I think it makes progress into helping us understand how a human brain works a little better. Although the overall accuracy was lower, the study showed that the CNN models are sensitive to the same concepts that humans are sensitive to and perform worse on concepts that humans performed poorly on. The CNN has different layers and a final layer, the study successfully shows that the accuracy shown by the initial layers was much less than the accuracy shown by the final fully connected layer. The study also successfully showed that the accuracy profiles of the CNNs matched closely to the accuracy profiles of adults and children. Authors experimented with increasing layer depths for the model that had the highest accuracy – ResNet-18 and the authors experimented with ResNet-34, ResNet-50, ResNet-101, ResNet-152 and surprisingly the highest accuracy was still shown by ResNet-18, so to conclude, model depth did not add any additional accuracy points. Lastly, all the CNN models generally agreed with each other’s predictions as to which classes were harder and which were easier. The study demonstrates that sensitivity to geometric and topological concepts plays a critical role in categorizing visual stimuli, both for humans and CNNs. Which proves the traditional beliefs of sensitivity to GT concepts being a “core knowledge” wrong. Although, “core knowledge” can be used to better the efficiency of the dataset and it can also help children learn from less number of observations. The results and experimentation in the paper show that the sensitivity can be learned through visual experience. There are limitations and future ideas for this project that include finetuning models on more advanced tasks and to use CNNs to make neural predictions. There has been work that suggests that the way neurons in the brains fire might correspond to the activations of the layers of CNNs when it processes an image so that could be an interesting future journey to embark on.