Are Machines More Effective than Humans for Graphical Perception Tasks?

Shuning Jiang, Wei-Lun Chao, Jian Chen, Daniel Haehn, Meng Ling, Ce Shang, Hanspeter Pfister
jiang.2126@osu.edu, chao.209@osu.edu, chen.8028@osu.edu, daniel.haehn@umb.edu, ling.253@osu.edu, e.shang@osumc.edu, fister@seas.harvard.edu

HIGHLIGHT

- CNNs can complete graphical perception task with super-human accuracy.
- Training data sampling has a significant effect on CNNs' accuracy.
- CNNs prefer tall bars and don't follow Weber's law.

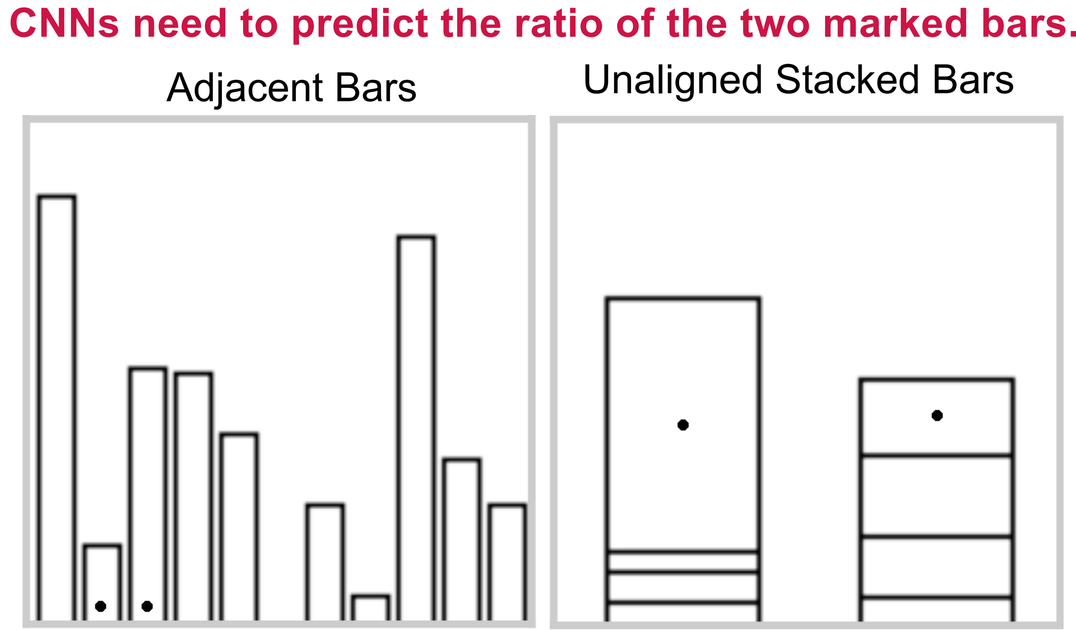
INTRODUCTION

Graphical perception: "The visual decoding of information encoded on visualizations".

Research Questions

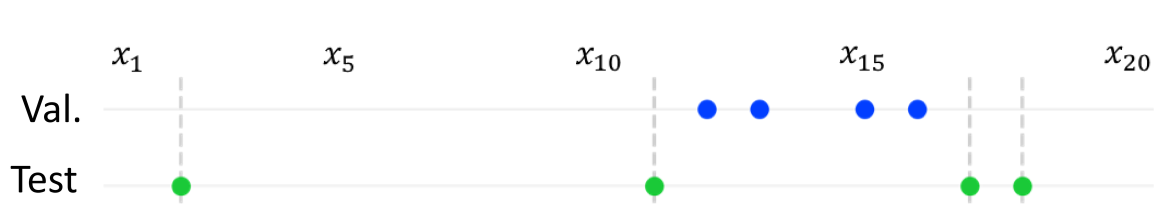
- Are machines more effective than humans for graphical perception tasks?
- How does data distribution affect CNNs accuracy?

Figure 1. Ratio estimate task that used in our study. CNNs need to predict the ratio of the two marked bars.



SAMPLING METHODS

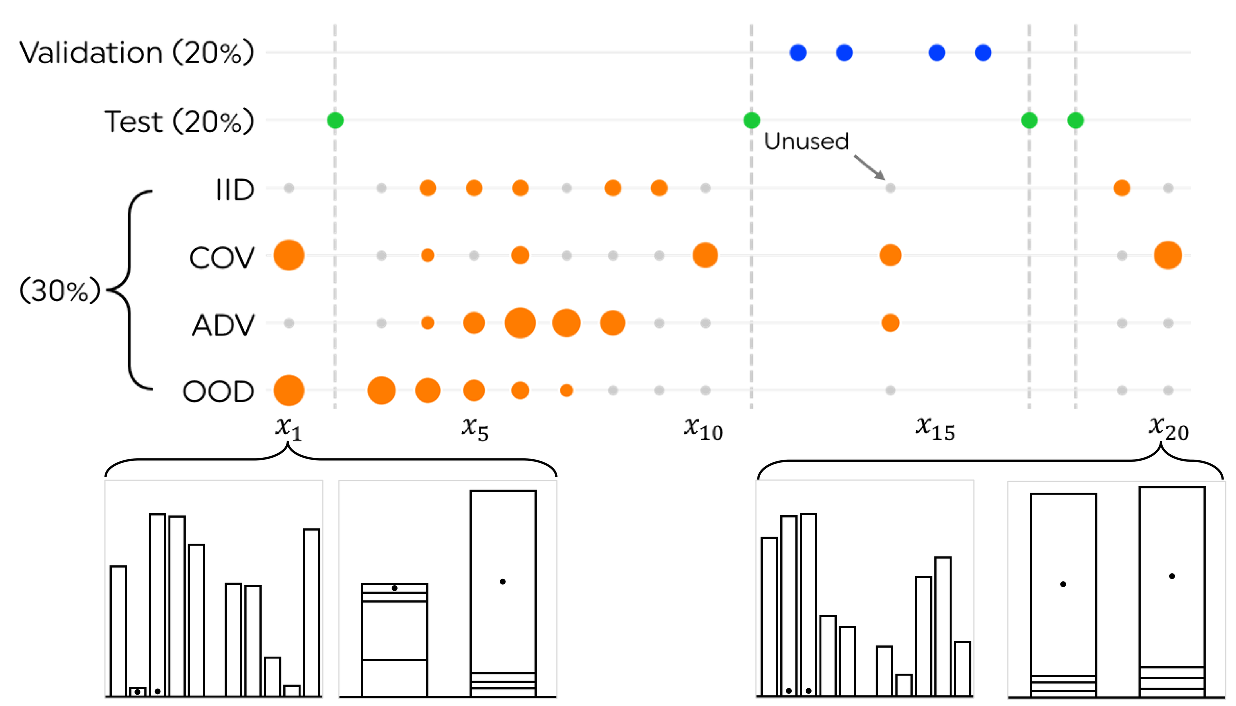
Step 1: select the test and validation samples



Step 2: choose training samples

- **IID**: select training samples uniformly from remaining samples.
- **COV**: select the inverse or most distant neighbors from selected training samples.
- **ADV**: select the inverse or most distant neighbors from the test data samples.
- **OOD**: select from where the test samples are least covered by the training samples.

Figure 2. A toy example that illustrates sampling methods. Larger dots are sampled first



EXPERIMENTS

- **Goals**: to measure CNNs' accuracy and understand how sampling affect the accuracy
- **Exp. 1** (Cleveland and McGill's test [1])
  - Dependent variable: ratio of marked bars
- **Exp. 2** (Talbot et al.'s test [2])
  - Dependent variable: taller marked bar height
- **Hypotheses**
  - H1: CNNs can have super-human accuracy
  - H2: CNNs will fail if been trained with OOD

Results

- IID, COV, and ADV lowered the human errors up to three-folds.
- OOD had a knee-shape, and the errors increased almost linearly for cases outside of the training data span.

Table 1. Exp. 1, absolute errors of humans and CNNs

Bar Type	Human	IID	COV	ADV	OOD
Adjacent	0.037 (0.029,0.046)	0.012 (0.010,0.014)	0.011 (0.010,0.012)	0.012 (0.011,0.015)	0.104 (0.092,0.155)
Stacked	0.067 (0.058,0.074)	0.017 (0.015,0.019)	0.015 (0.013,0.016)	0.015 (0.014,0.016)	0.107 (0.099,0.115)

Figure 3. Exp. 1, absolute error vs. ratio samples

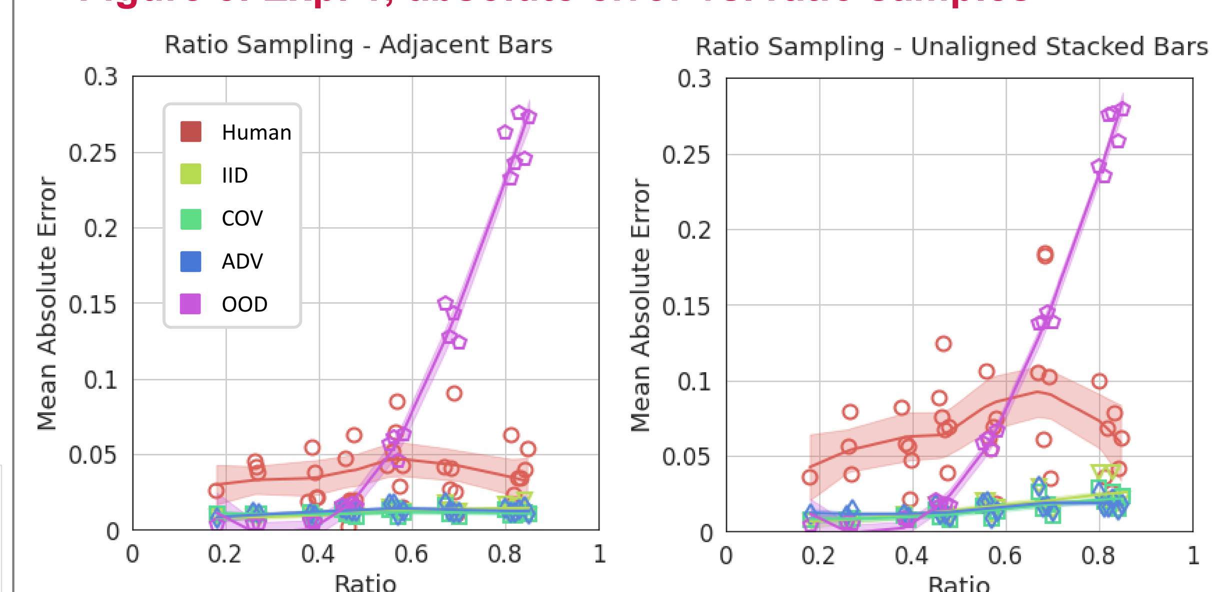
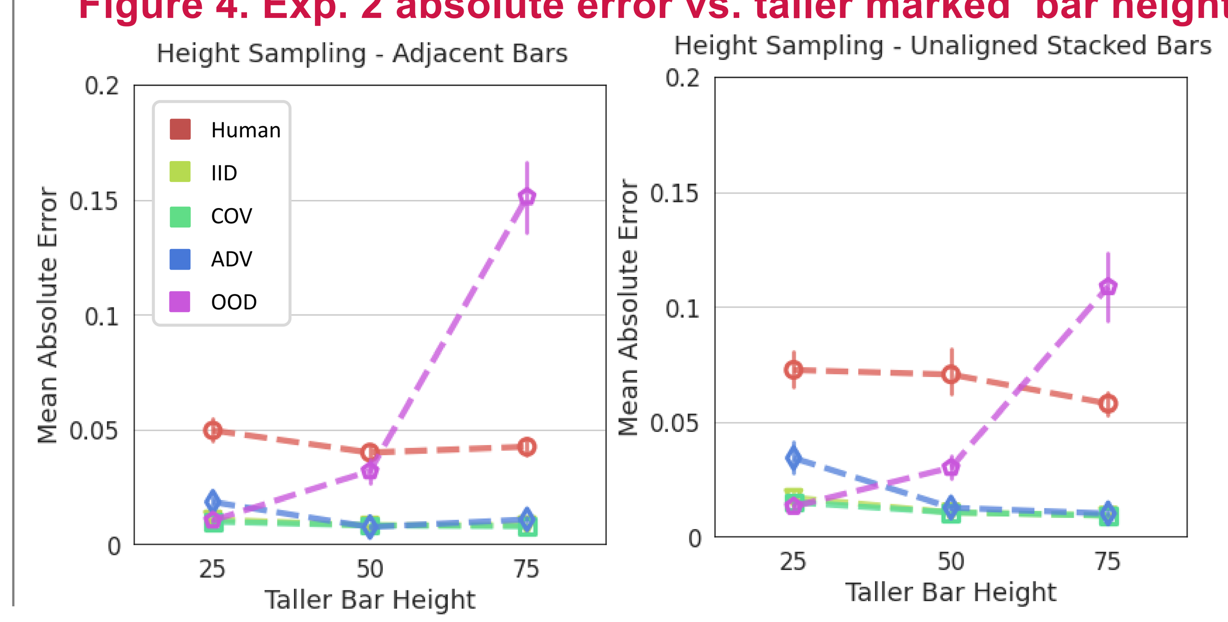


Table 2. Exp. 2, absolute errors of humans and CNNs

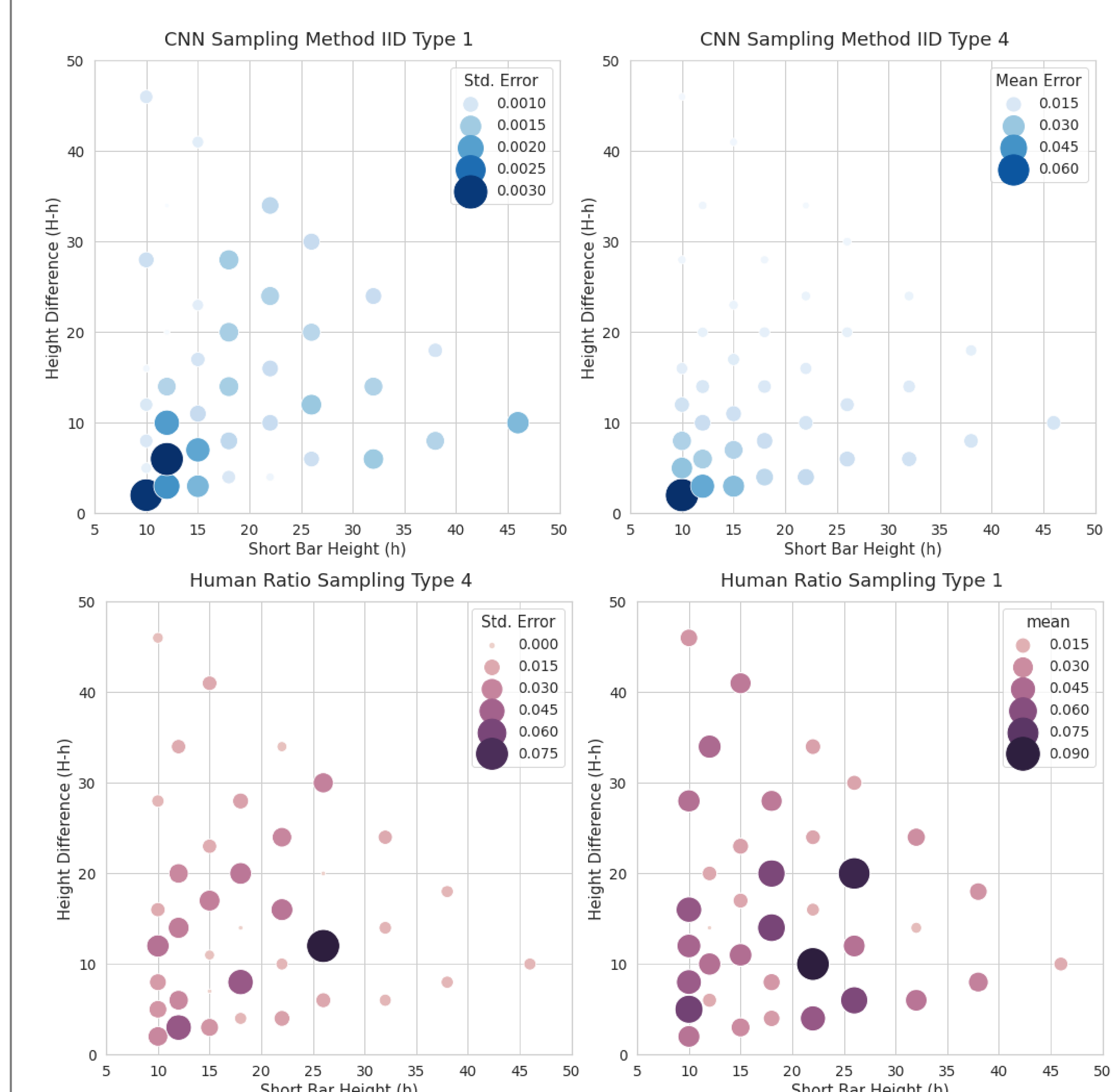
Bar Type	Human	IID	COV	ADV	OOD
Adjacent	0.044 (0.040,0.048)	0.009 (0.008,0.010)	0.009 (0.008,0.009)	0.012 (0.011,0.013)	0.065 (0.052,0.076)
Stacked	0.067 (0.060,0.076)	0.012 (0.010,0.015)	0.012 (0.010,0.013)	0.019 (0.017,0.021)	0.051 (0.042,0.061)

Figure 4. Exp. 2 absolute error vs. taller marked bar height



- CNNs don't follow Weber's law. Given a fixed bar heights difference Δ, CNNs' errors became smaller as the original stimulus (shorter marked bar height) increased. Overall, CNNs had the highest error when both marked bars were short.

Figure 5. CNNs showed different behaviors from humans when plotting the Weber's law-like approach.



CONCLUSION

- CNNs can complete graphical perception task with super-human accuracy with the appropriate sampling methods.
- CNNs don't generalize well. It's important to check whether the test set is covered.
- CNNs perform better when the visual mark has sufficient pixel. Predictions on taller bars have higher accuracy.

THE OHIO STATE UNIVERSITY UMASS BOSTON HARVARD UNIVERSITY

[1] Cleveland, W. S., & McGill, R. (1984). Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *Journal of the American Statistical Association*, *79*, 531–554.
[2] Talbot, J., Setlur, V., & Anand, A. (2014). Four experiments on the perception of bar charts. IEEE Transactions on Visualization and Computer Graphics, 20(12), 2152–2160.