

Replication of CVCL Model by Vong et al. (2024, Science)

Jane Yang (j7yang@ucsd.edu)

2024-10-14

Table of contents

Introduction	1
Methods	2
Power Analysis	2
Planned Sample	2
Materials	2
Procedure	3
Analysis Plan	3
Differences from Original Study	3
Methods Addendum (Post Data Collection)	3
Results	3
Data preparation	3
Confirmatory analysis	4
Exploratory analyses	4
Discussion	4
Summary of Replication Attempt	4
Commentary	4

Introduction

I am broadly interested in how children’s early vocabulary development intersects with their visual category learning. It fascinates me that a child can easily grasp fundamental properties of objects from just a few examples, while highly trained models often fall short. By integrating infants’ visual and linguistic experiences into computational models, I aim to explore the mechanisms through which children learn categories from everyday experiences. In Vong’s paper, the authors proposed the Child’s View for Contrastive Learning (CVCL) model, which

embodies a form of cross-situational associative learning. This model tracks the co-occurrences of words and their possible visual referents to establish mappings. By reproducing the findings from this paper and understanding the model's implementation details, I will learn how to use contrastive language-image pre-training models in my own research. Ultimately, my long-term goal is to develop a cognitively realistic model that learns robust representations from children's everyday experiences.

First, I will obtain the SAYCam training dataset from Databrary and download the pre-trained CVCL model from the HuggingFace Hub. To familiarize myself with the dataset, I will randomly sample it and feed the sampled data into the CVCL model, encoding images and utterances to quickly assess the model's performance before proceeding further. With a basic understanding of the model, I will then follow the analysis pipeline outlined in the paper to reproduce its main figures. These analyses are divided into four key categories: (1) descriptive analysis of the training data, (2) t-SNE plots showing the alignment of vision and language from a child's perspective, (3) image classification accuracy comparing CVCL, CLIP, and a linear probe, and (4) attention maps generated by Grad-CAM to illustrate object localization capabilities across four different categories in CVCL. Challenges will likely occur during model evaluation, particularly implementing the CLIP model and other approaches comparing their performance with CVCL on image classification. Finally, I will test the models' generalization by evaluating them on novel visual exemplars not included in the training dataset.

Link to GitHub repo: https://github.com/JaneYang07/vong2024_replication

Link to the original paper: <https://www.science.org/doi/abs/10.1126/science.adi1374>

Methods

Power Analysis

Original effect size, power analysis for samples to achieve 80%, 90%, 95% power to detect that effect size. Considerations of feasibility for selecting planned sample size.

Planned Sample

Planned sample size and/or termination rule, sampling frame, known demographics if any, preselection rules if any.

Materials

All materials - can quote directly from original article - just put the text in quotations and note that this was followed precisely. Or, quote directly and just point out exceptions to what was described in the original article.

Procedure

Can quote directly from original article - just put the text in quotations and note that this was followed precisely. Or, quote directly and just point out exceptions to what was described in the original article.

Analysis Plan

Can also quote directly, though it is less often spelled out effectively for an analysis strategy section. The key is to report an analysis strategy that is as close to the original - data cleaning rules, data exclusion rules, covariates, etc. - as possible.

Clarify key analysis of interest here You can also pre-specify additional analyses you plan to do.

Differences from Original Study

Explicitly describe known differences in sample, setting, procedure, and analysis plan from original study. The goal, of course, is to minimize those differences, but differences will inevitably occur. Also, note whether such differences are anticipated to make a difference based on claims in the original article or subsequent published research on the conditions for obtaining the effect.

Methods Addendum (Post Data Collection)

You can comment this section out prior to final report with data collection.

Actual Sample

Sample size, demographics, data exclusions based on rules spelled out in analysis plan

Differences from pre-data collection methods plan

Any differences from what was described as the original plan, or “none”.

Results

Data preparation

Data preparation following the analysis plan.

Confirmatory analysis

The analyses as specified in the analysis plan.

Side-by-side graph with original graph is ideal here

Exploratory analyses

Any follow-up analyses desired (not required).

Discussion

Summary of Replication Attempt

Open the discussion section with a paragraph summarizing the primary result from the confirmatory analysis and the assessment of whether it replicated, partially replicated, or failed to replicate the original result.

Commentary

Add open-ended commentary (if any) reflecting (a) insights from follow-up exploratory analysis, (b) assessment of the meaning of the replication (or not) - e.g., for a failure to replicate, are the differences between original and present study ones that definitely, plausibly, or are unlikely to have been moderators of the result, and (c) discussion of any objections or challenges raised by the current and original authors about the replication attempt. None of these need to be long.