

Compositional Generalization via Forced Rendering of Disentangled Latents

Qiyao (Catherine) Liang^{*} 1, Daoyuan Qian^{*} 2, Liu Ziyin^{1,3}, Ila Fiete¹

¹Massachusetts Institute of Technology

²University of Cambridge

³NTT Research

^{*}Equal Contribution



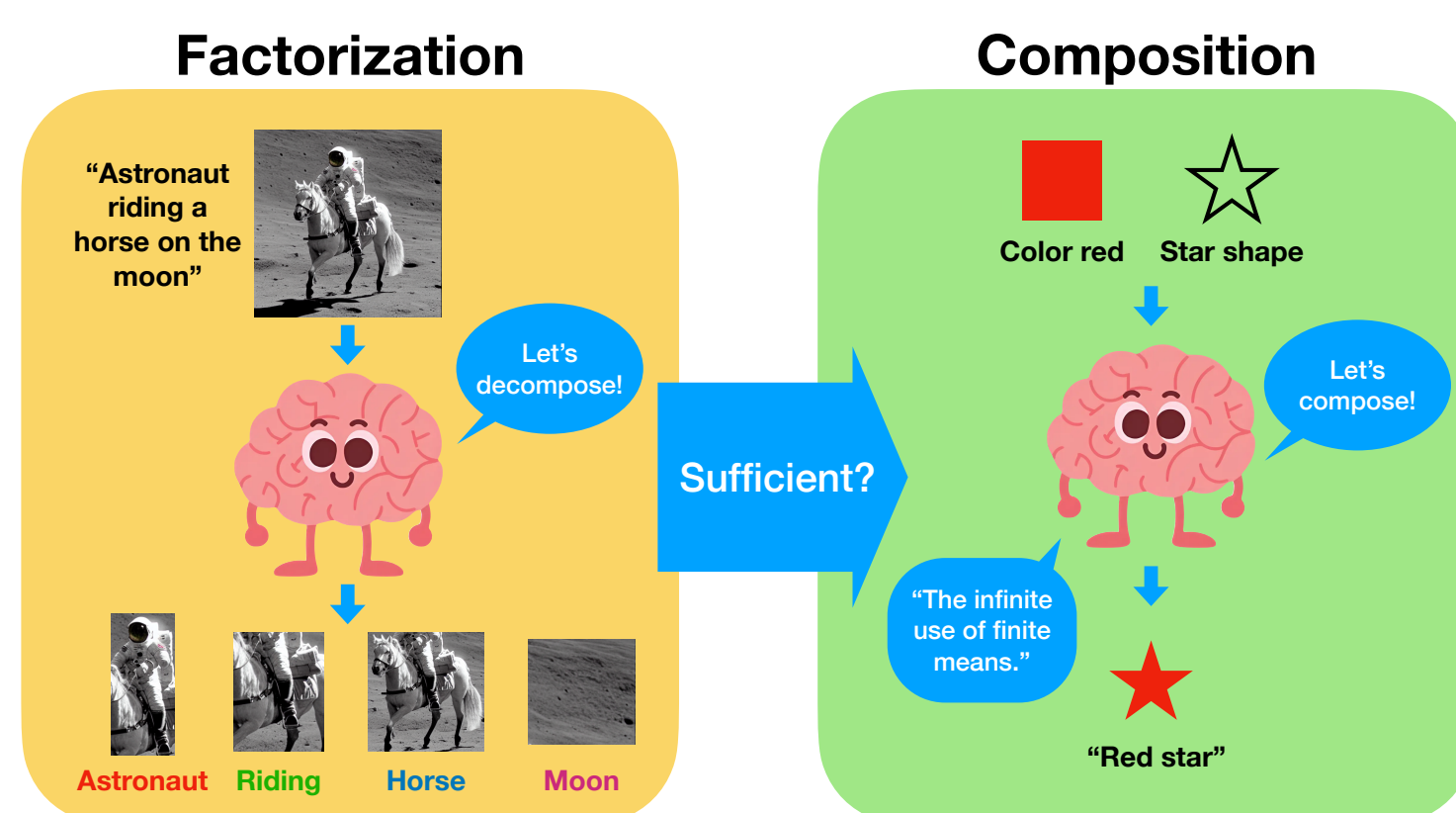
Reach out to ask
a Q or say hi ;)
qiyao@mit.edu



Check out our
paper on arXiv!

Motivation

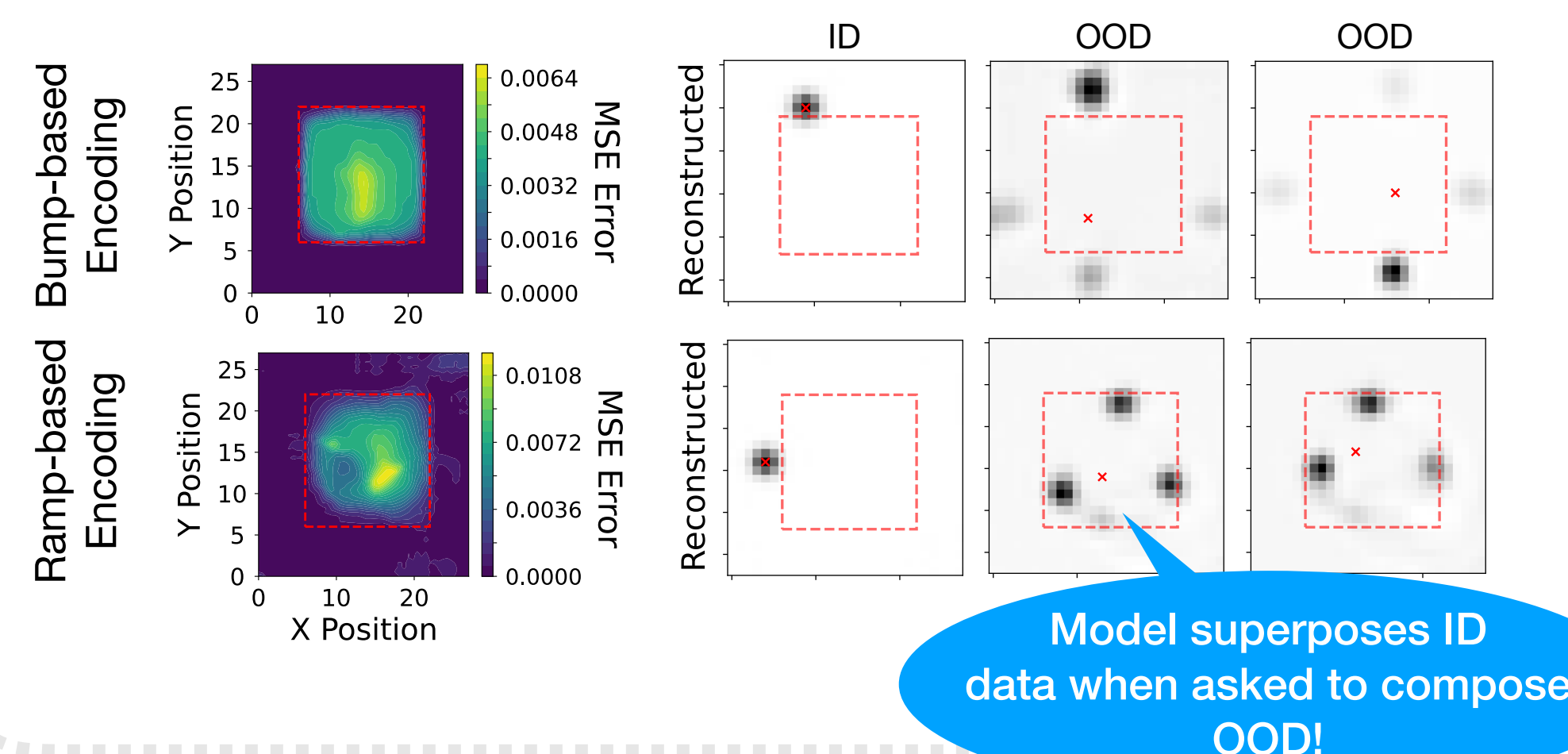
Human intelligence can decompose data into independent factors and recombine them, allowing compositional generalization.



*Is factorization sufficient for
compositional generalization??*

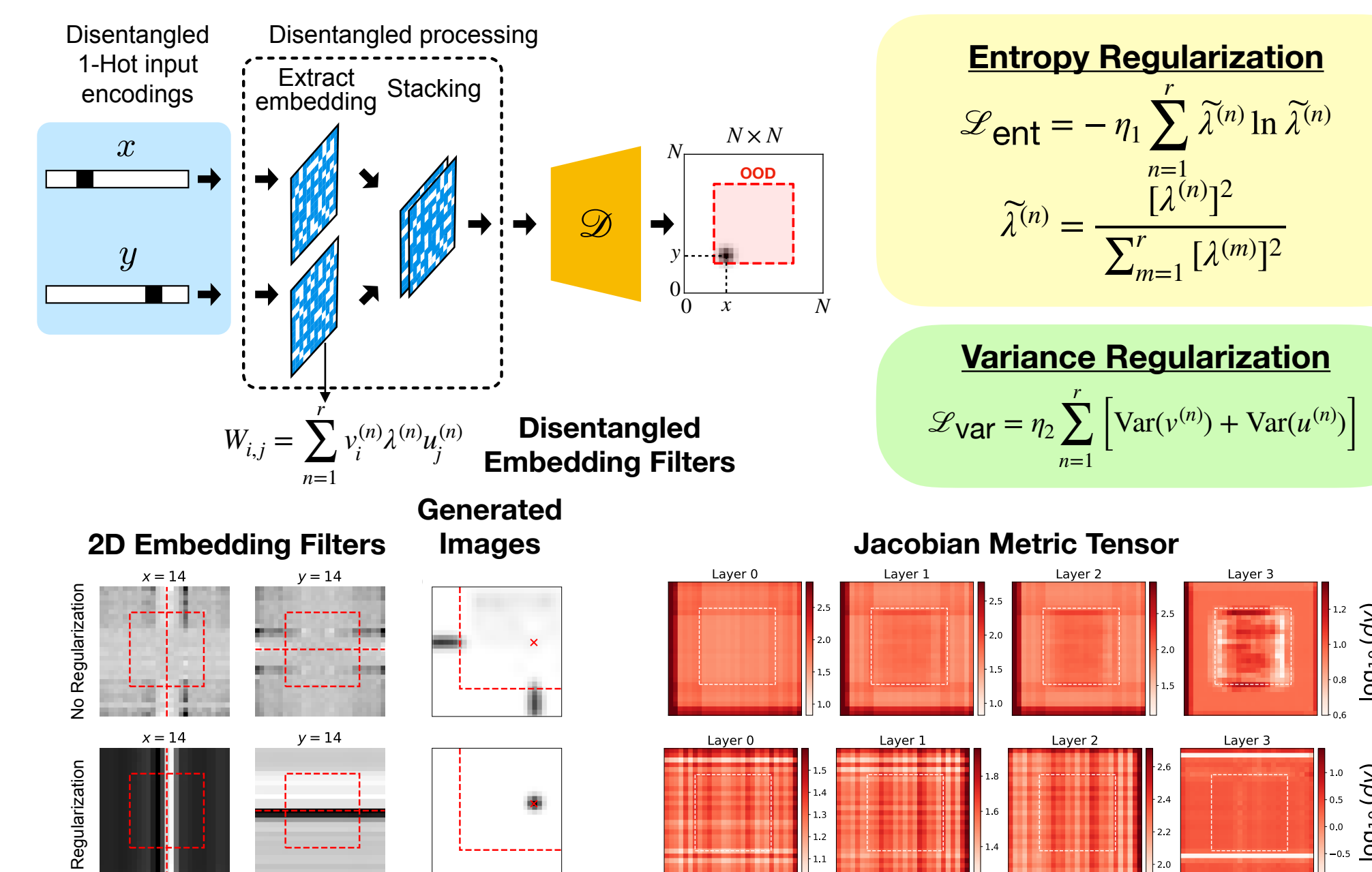
Comp Gen Sucks Naively

Despite disentangled input representations (regardless of the form), the compositional generalization performance sucks :(



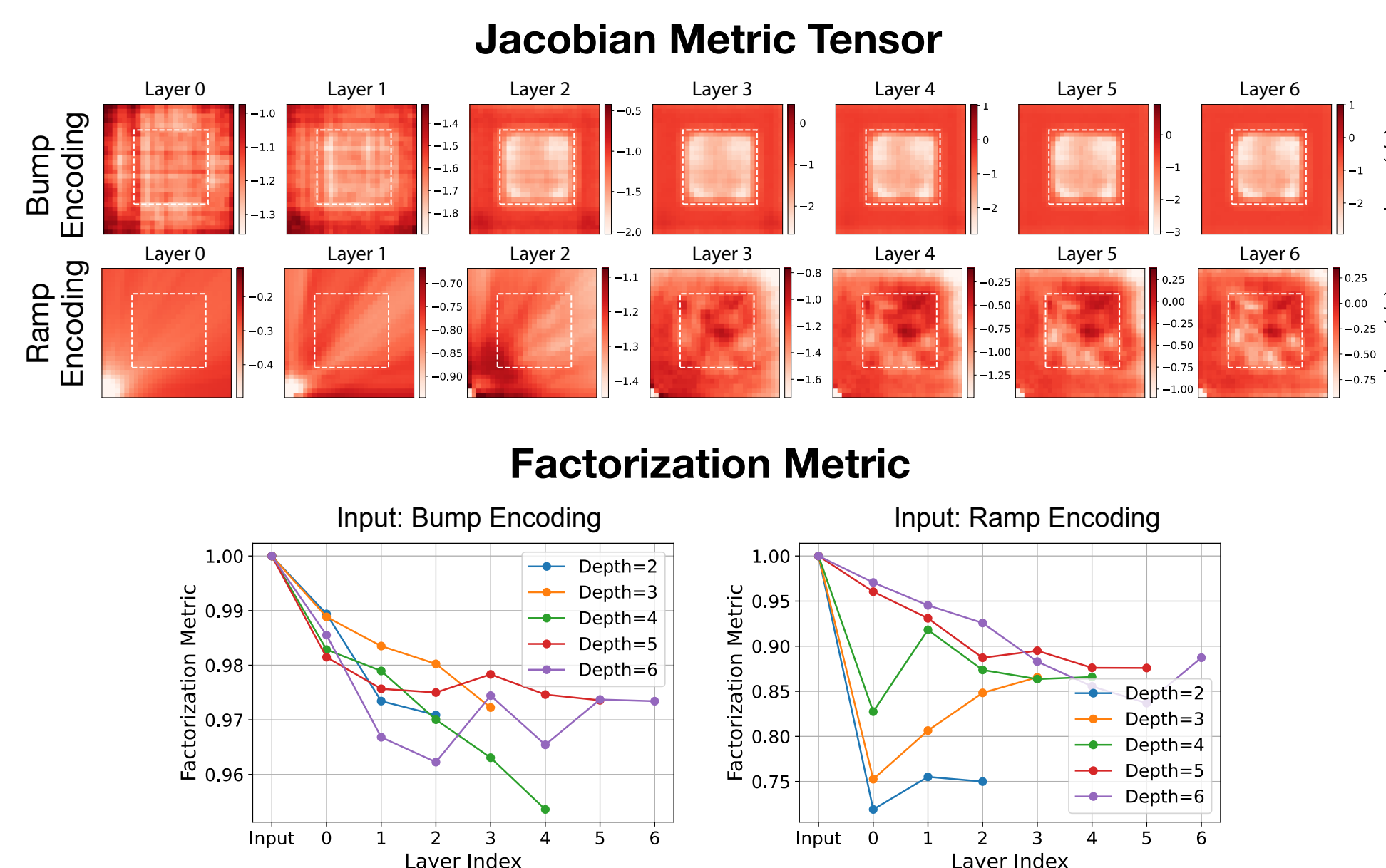
How to fix?

Plan A: Architectural Constraint



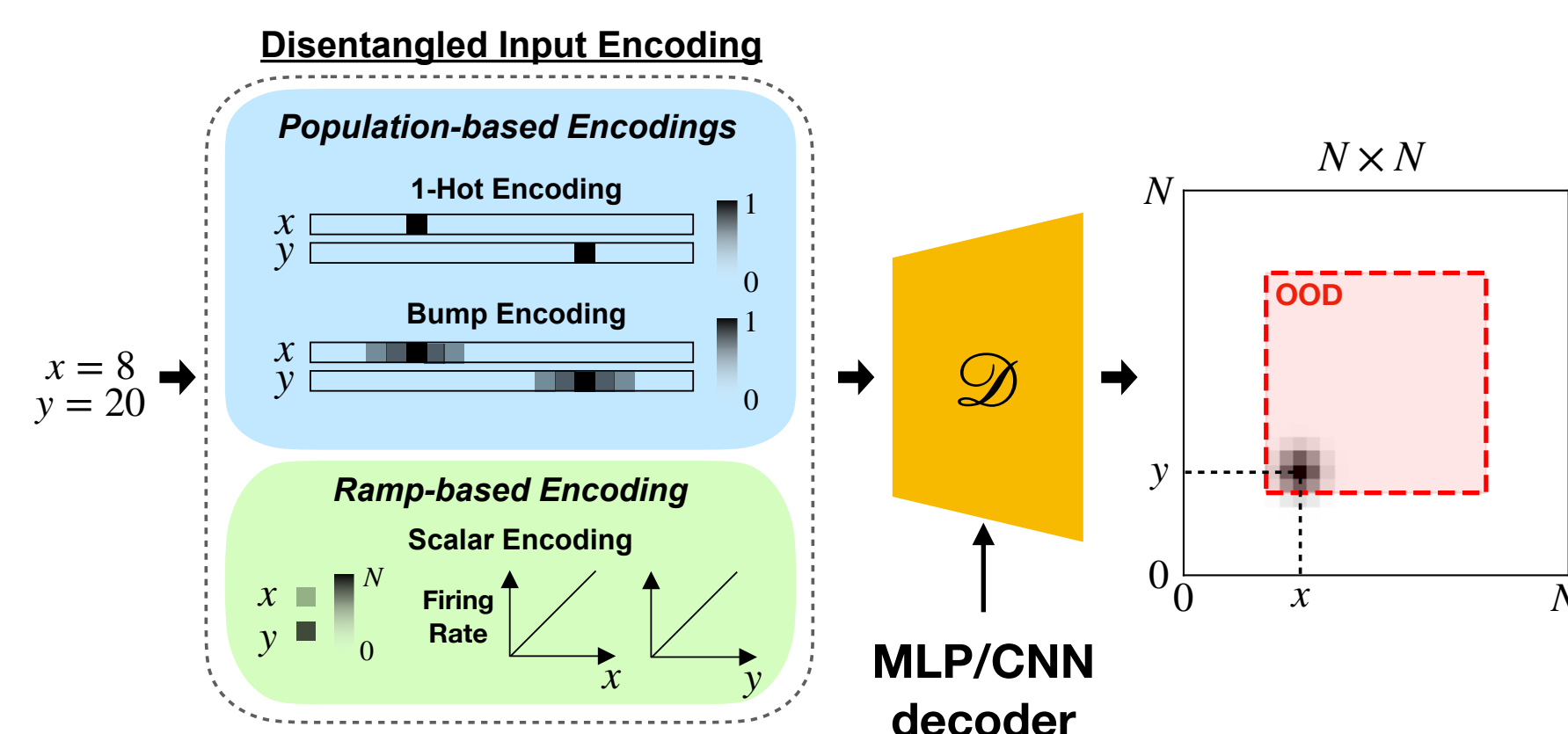
What Happened?

Apparently disentangled input representations get “warped” & “re-entangled” over the layer depth, leading to comp gen failure!



Experimental Setup

We probed a bunch of factorized encoding forms on the CNN/MLP decoder only architecture on a Gaussian bump image generation task!



How to fix? Plan B: Data Curation

