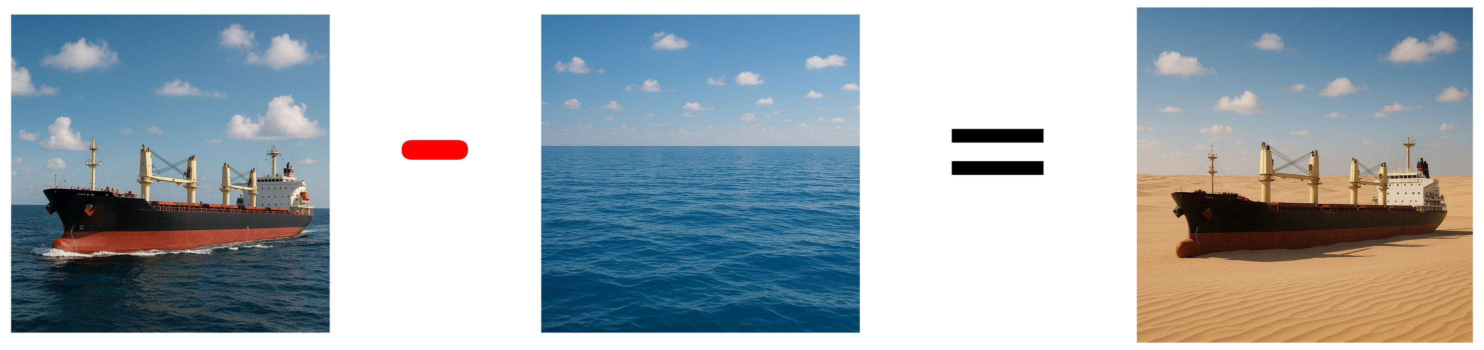
CoInD: Enabling Logical Compositions in Diffusion Models

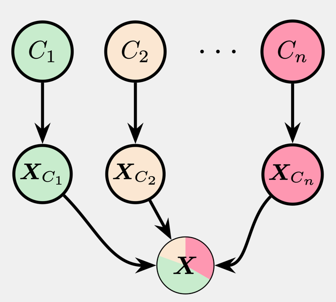
Sachit Gaudi, Gautam Sreekumar, Vishnu Boddeti

Logical Compositionality Task



'Ship' − 'Water' = 'Ship' ¬ 'Water'

$x \sim \text{score}_\theta(X \mid \text{Ship} \neg \text{Water})$
$= \text{score}_\theta(X \mid \text{Ship}) - \text{score}_\theta(X \mid \text{Water}) + \text{score}_\theta(X)$



Causal graph of underlying data generating process

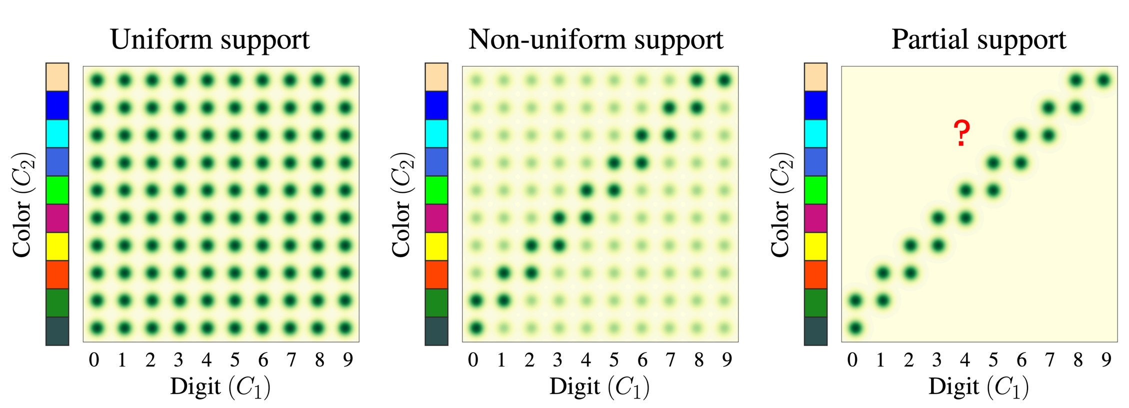
For explicit attribute composition:

- Attributes should vary independently
- Attributes are not causally related
- It is possible to estimate $C_1, C_2, \cdots, C_n$ from X

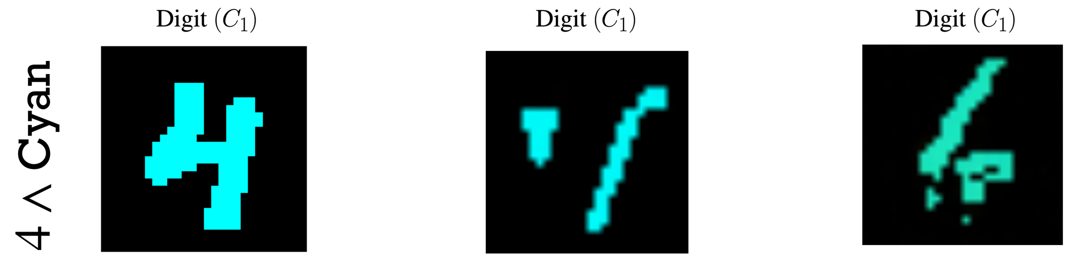
Assumption: Conditional Independence

Can Diffusion Models Generate Arbitrary Compositions Trained on Arbitrary Support?

Ans: **NO**



Generates incorrect interpolations of unseen or less frequently seen attribute compositions



Do Diffusion Respect the Underlying Data Generation?

$$p_\theta(C_1, C_2 \mid X) = p_\theta(C_1 \mid X)p_\theta(C_2 \mid X)$$

We use the Jensen-Shannon Divergence between the left and **right** terms as a measure of conditional independence violation.

- Diffusion models do not implicitly learn the conditional independence
- Diffusion models capture the dependencies present in the training

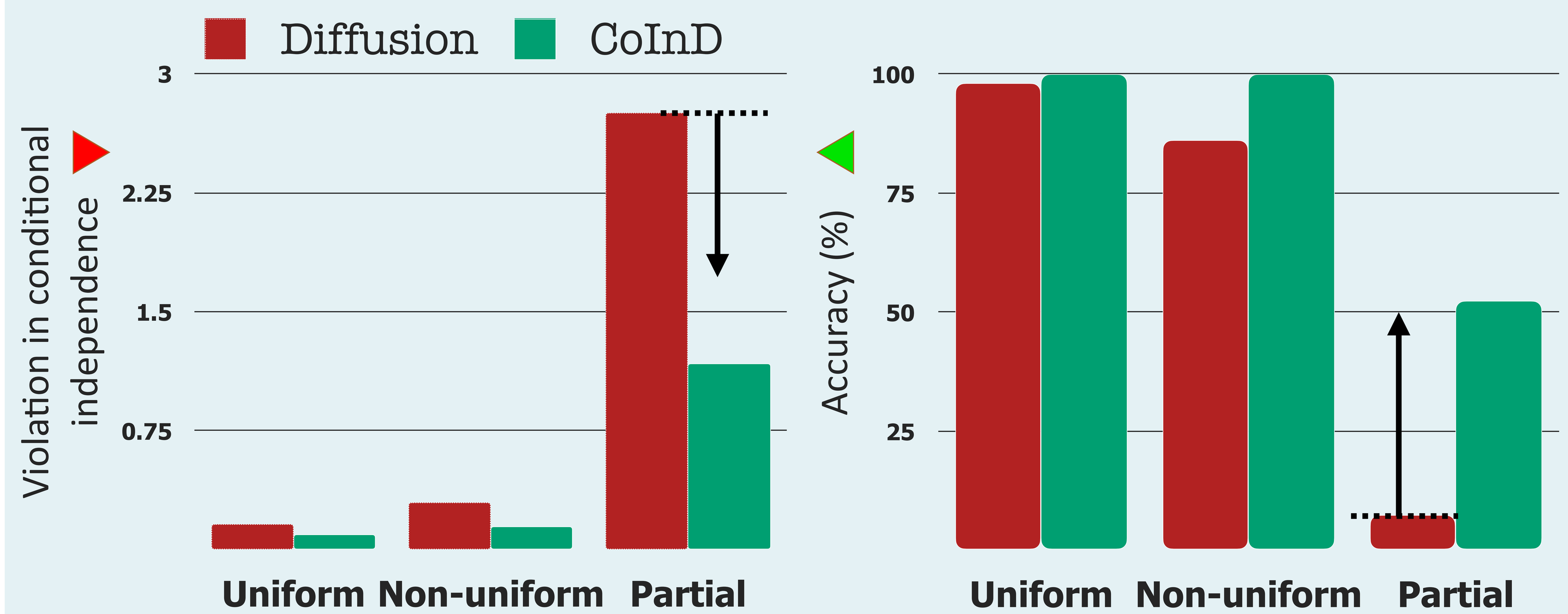
Cause: Incorrect Marginals

$$p_\theta(X \mid C_i) = p_{\text{train}}(X \mid C_i) \neq p_{\text{true}}(X \mid C_i)$$

- Diffusion models' objective is not suitable for compositionality

Diffusion Models Are Not Compositional

Conditional Independence Enables Compositionality



Conditional Independence To Enable Compositionality

Applying Bayes' rule to Conditional Independence (CI) equation

$$\frac{p_\theta(X \mid C_1, C_2)p(C_1, C_2)}{\cancel{p_\theta(X)}} = \frac{p_\theta(X \mid C_1)p(C_1)}{\cancel{p_\theta(X)}}\frac{p_\theta(X \mid C_2)p(C_2)}{p_\theta(X)}$$

Learned by score matching

$$\mathcal{L}_{\text{score}} = \|\nabla_X \log p_\theta(X \mid C) - \nabla_X \log p(X \mid C)\|_2^2$$

Applying $\nabla_X \log(\cdot)$ to convert probabilities into score will give **CI** loss

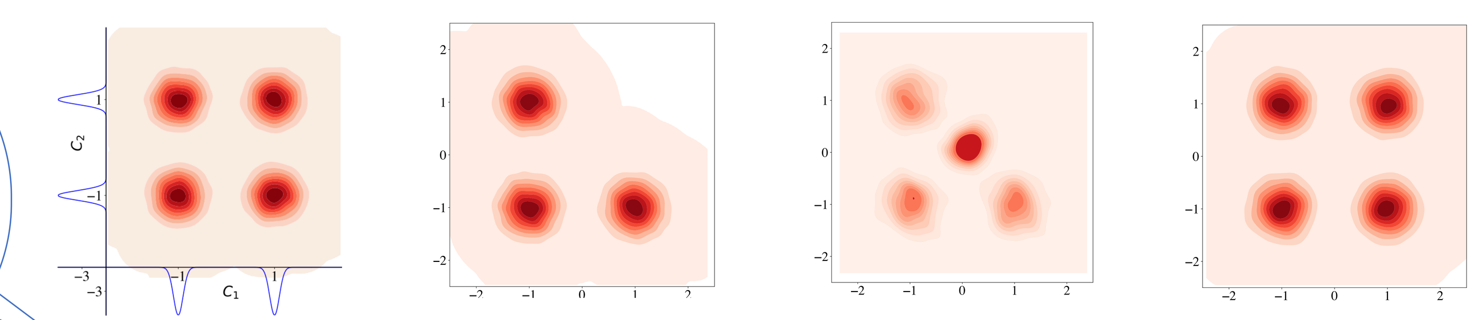
$$\mathcal{L}_{\text{CI}} = \|\nabla_X \log p_\theta(X \mid C_1, C_2) + \nabla_X \log p_\theta(X) - \nabla_X \log p_\theta(X \mid C_1) - \nabla_X \log p_\theta(X \mid C_2)\|_2^2$$

$$\mathcal{L}_{\text{CoInD}} = \mathcal{L}_{\text{score}} + \lambda \mathcal{L}_{\text{CI}}$$

Marginals $X \mid C_1$ and $X \mid C_2$ are learned by enforcing CI and NOT by score matching against incorrect observed marginals

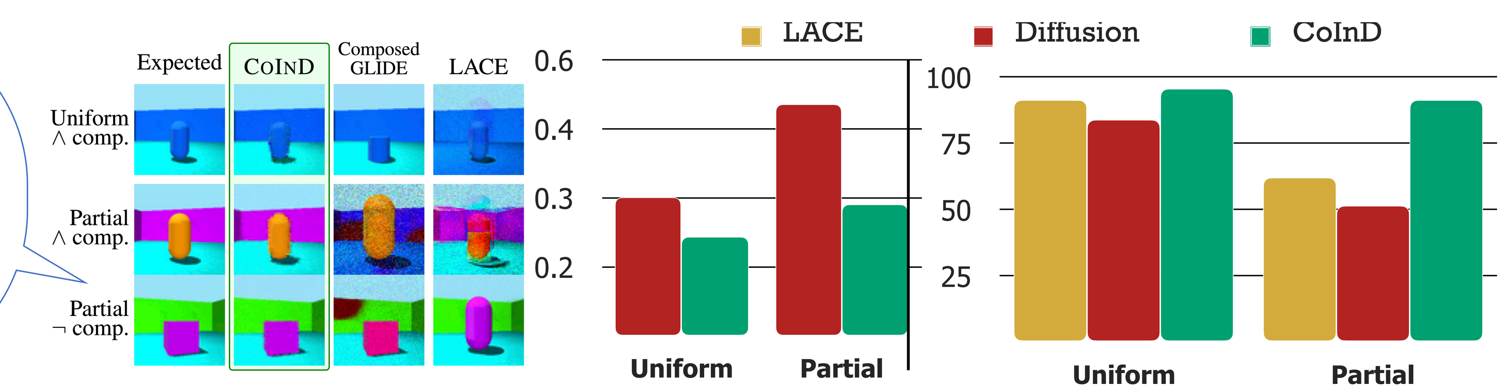
CoInD Improves Compositionality

Theoretical results on mixture of gaussians show that CoInD learns the true distribution



Underlying data dist. | Observed data dist. | Diffusion fails at composing independent C_1, C_2 | CoInD successfully composing C_1, C_2

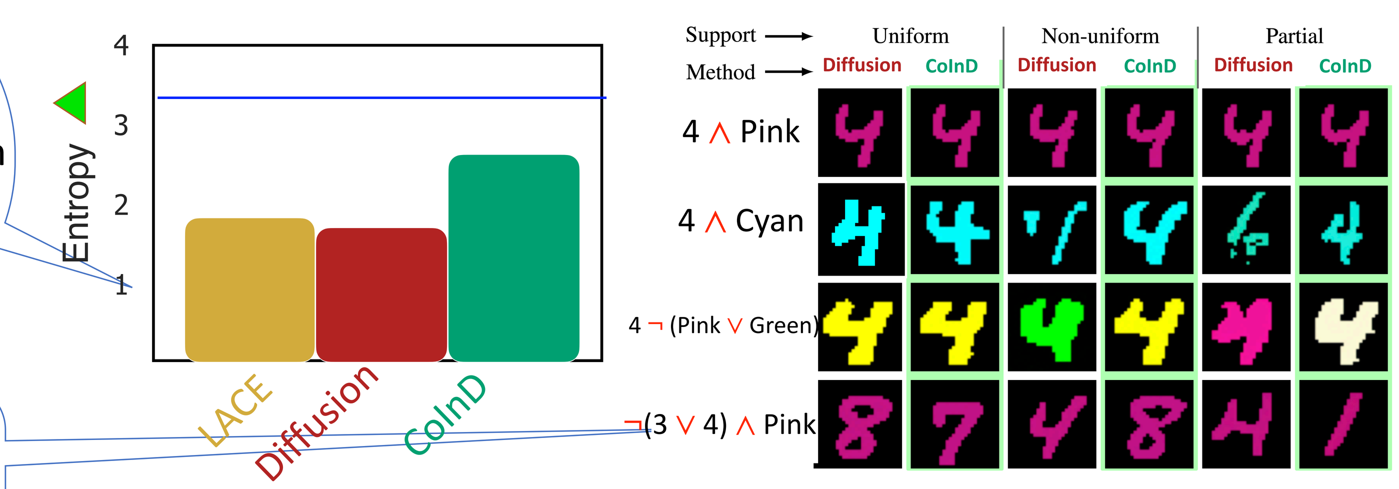
CoInD can compose six attributes demonstrating scalability



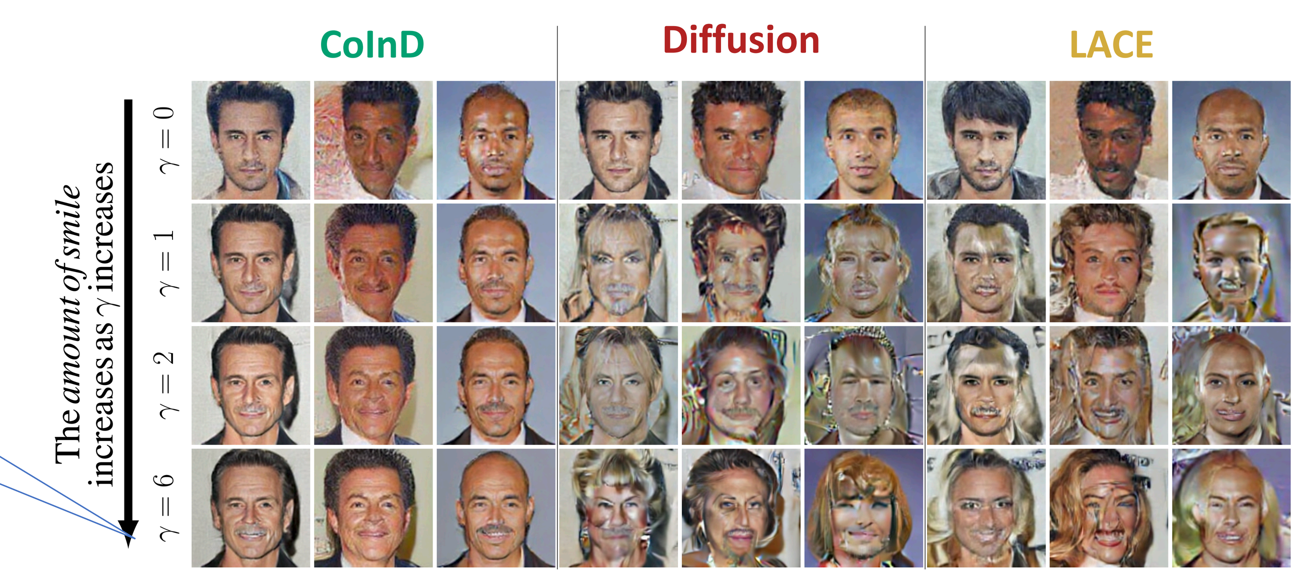
Samples from Shapes3d dataset | Violation in conditional independence | Accuracy(%)

Diversity of the unconditional attribute emerges from learning correct marginals

CoInD composing across supports



CoInD effectively controls the 'smile' attribute in the generation of unseen 'smiling male', whereas baselines rely on spurious correlations present in training data



$\sim \text{score}_\theta(X \mid \text{Male}) + \gamma\,\text{score}_\theta(X \mid \text{Smile}) - \gamma\,\text{score}_\theta(X)$

MICHIGAN STATE UNIVERSITY

ICLR

PAPER