WHICH WAY FROM B TO A:

THE ROLE OF EMBEDDING GEOMETRY IN IMAGE INTERPOLATION FOR STABLE DIFFUSION

Nick Karris

UC San Diego, Pacific Northwest National Laboratory

Acknowledgements

Joint Work With...

Luke Durell

Javier Flores

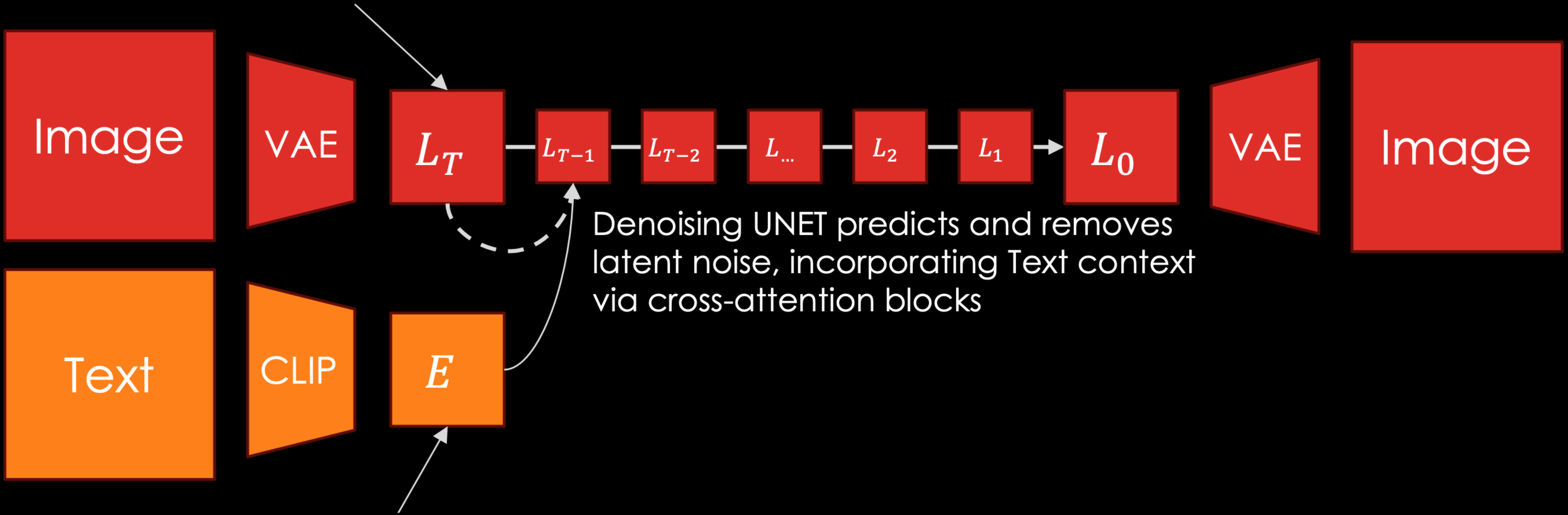
Tegan Emerson

This work was conducted under the Laboratory Directed Research and Development Program at PNNL, a multi-program national laboratory operated by Battelle for the U.S. Department of Energy under contract DE-AC05-76RL01830.

STABLE DIFFUSION

Stable Diffusion [Rombach'22] applies diffusion to latent image + conditioning

For sampling, this initial latent is 4x64x64 N(0,1)



CLIP produces 77 token vectors capturing the semantic context of the prompt in an embedding matrix

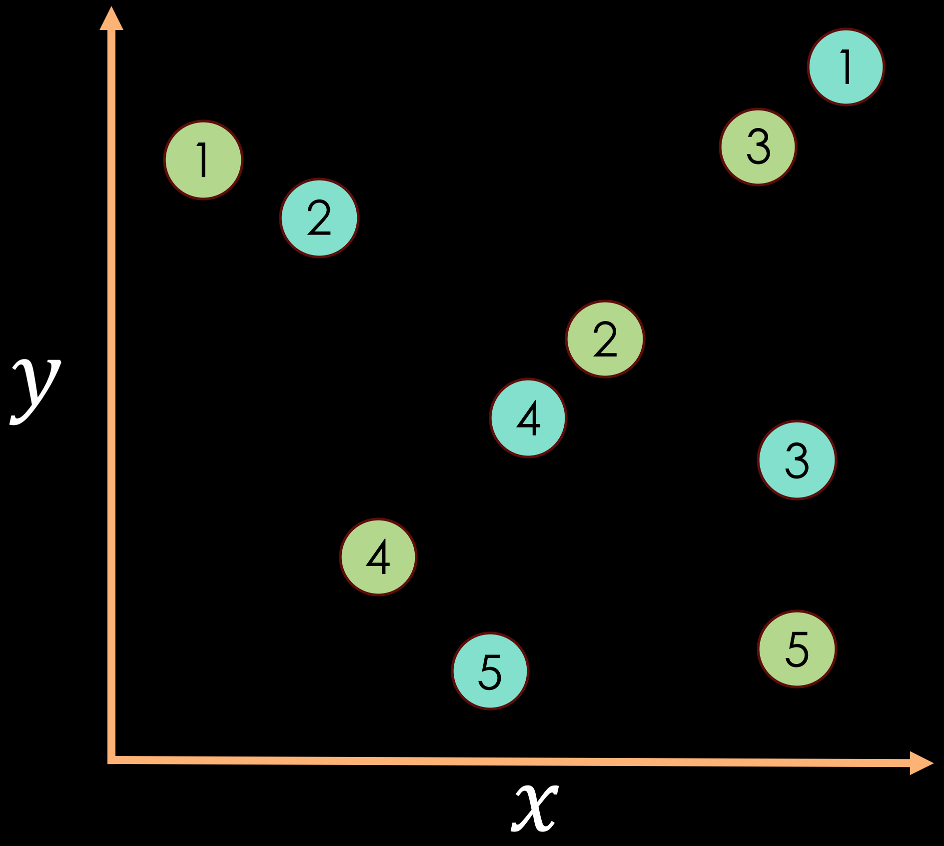
Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. (2022). *High-Resolution Image Synthesis with Latent Diffusion Models.* In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition.*

DIFFERENT EMBEDDING INTERPRETATIONS AND AN INTRODUCTION TO COUPLINGS

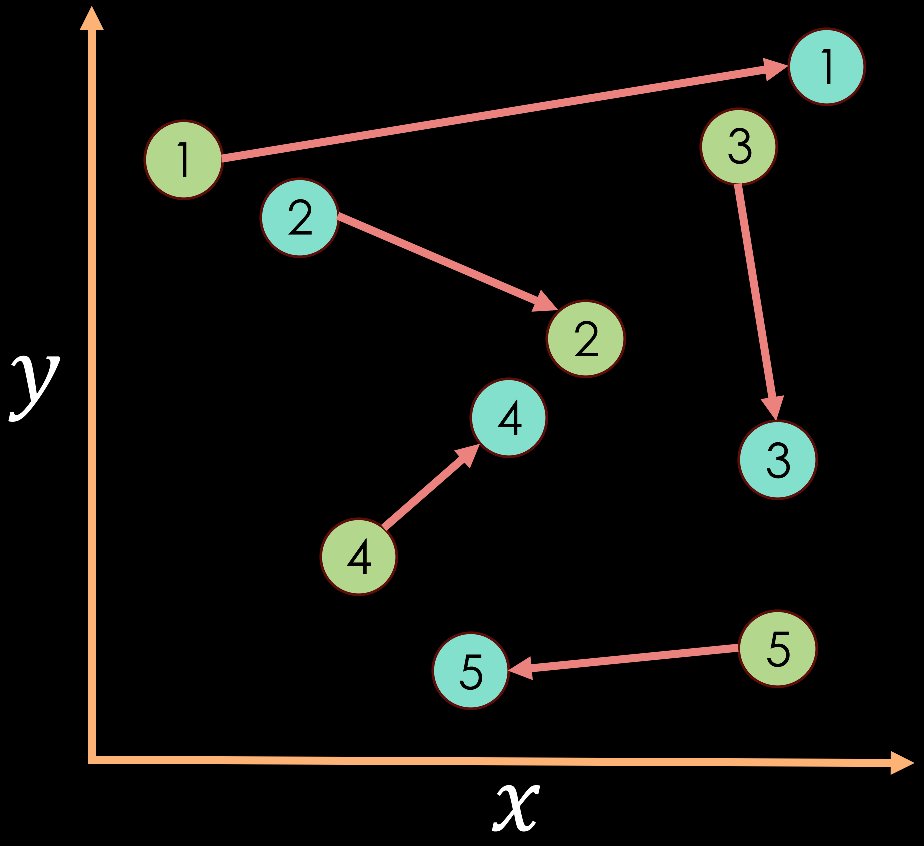
$$\begin{bmatrix} x_1^A & y_1^A \\ x_2^A & y_2^A \\ \vdots & \vdots \\ x_{n-1}^A & y_{n-1}^A \\ x_n^A & y_n^A \end{bmatrix}$$

$$\begin{bmatrix} x_1^B & y_1^B \\ x_2^B & y_2^B \\ \vdots & \vdots \\ x_{n-1}^B & y_{n-1}^B \\ x_n^B & y_n^B \end{bmatrix}$$

Matrix Representation

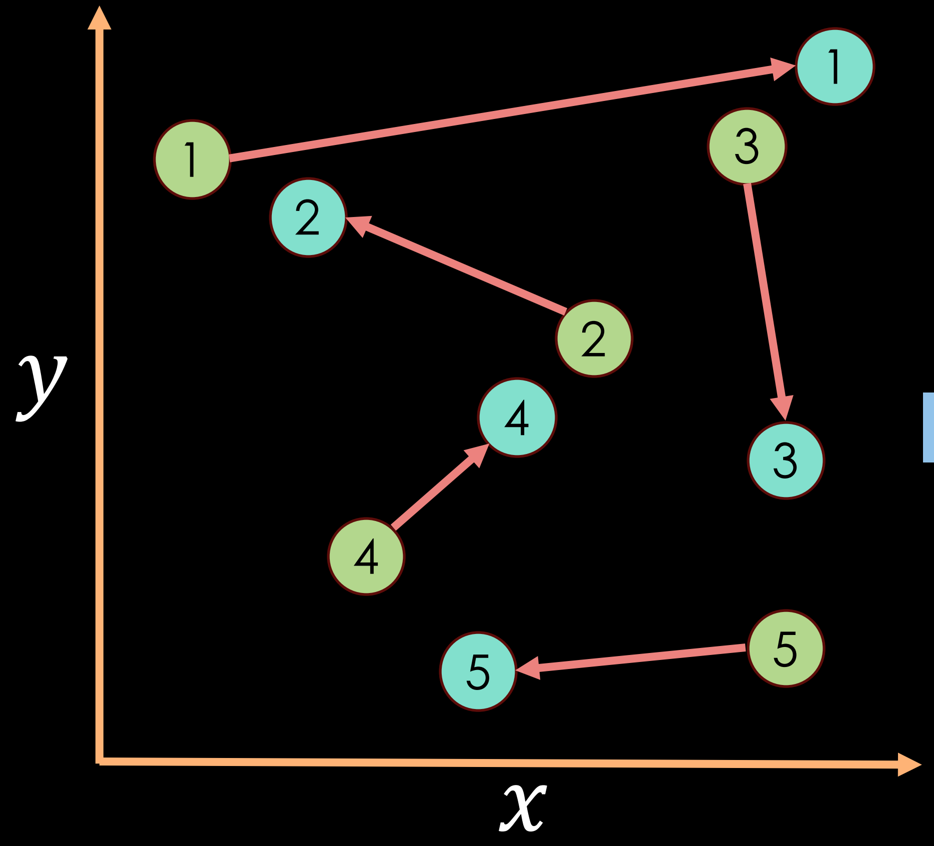


Vector/ Point Cloud Representation

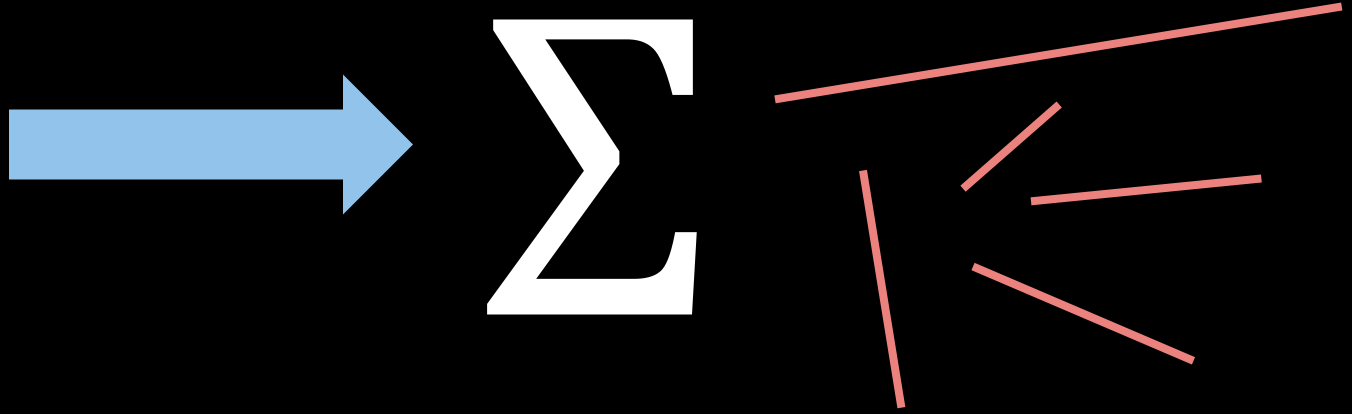


Standard Matrix Coupling (e.g. coupling for matrix addition)

COUPLING COST: "WORK TO MOVE ALONG A PATH"

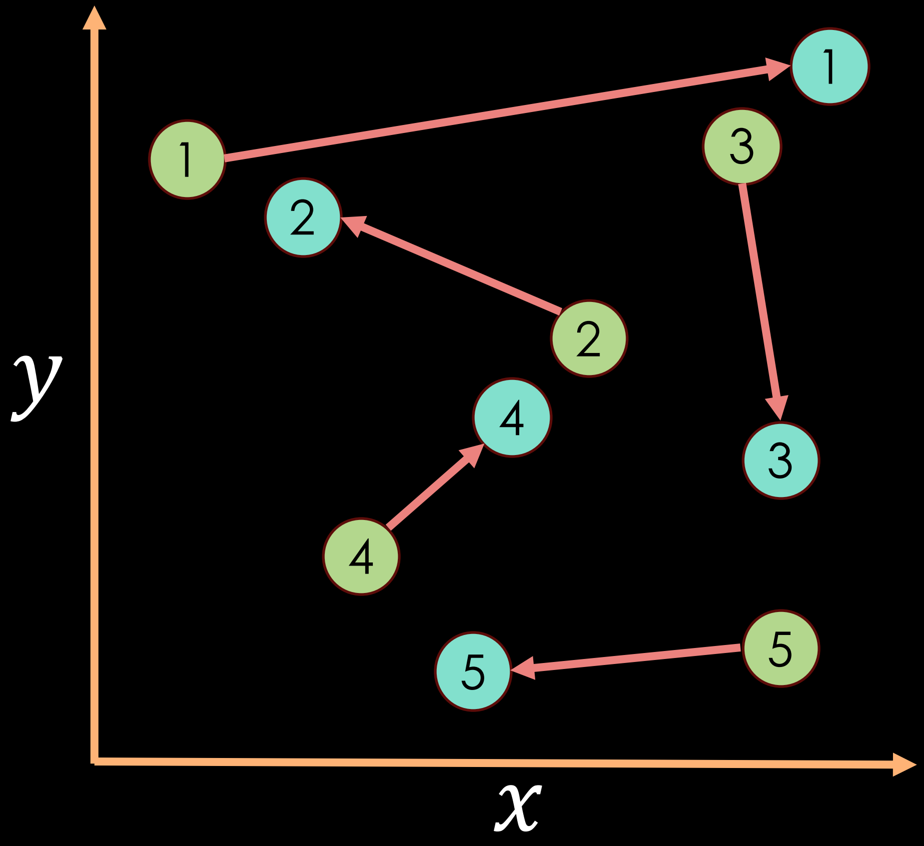


Standard Matrix Coupling
(i.e., coupling for matrix addition)

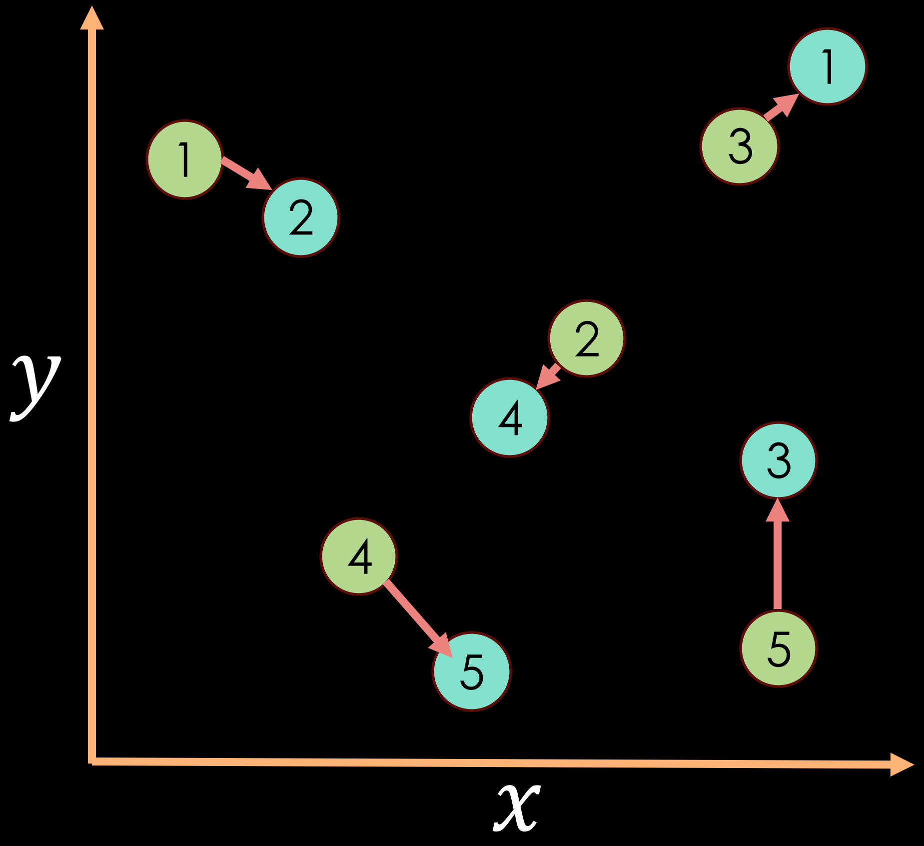


Coupling Cost is the Sum of the Product of the Weight of a Point and Distance it Moves

NOT ALL COUPLINGS ARE EQUAL

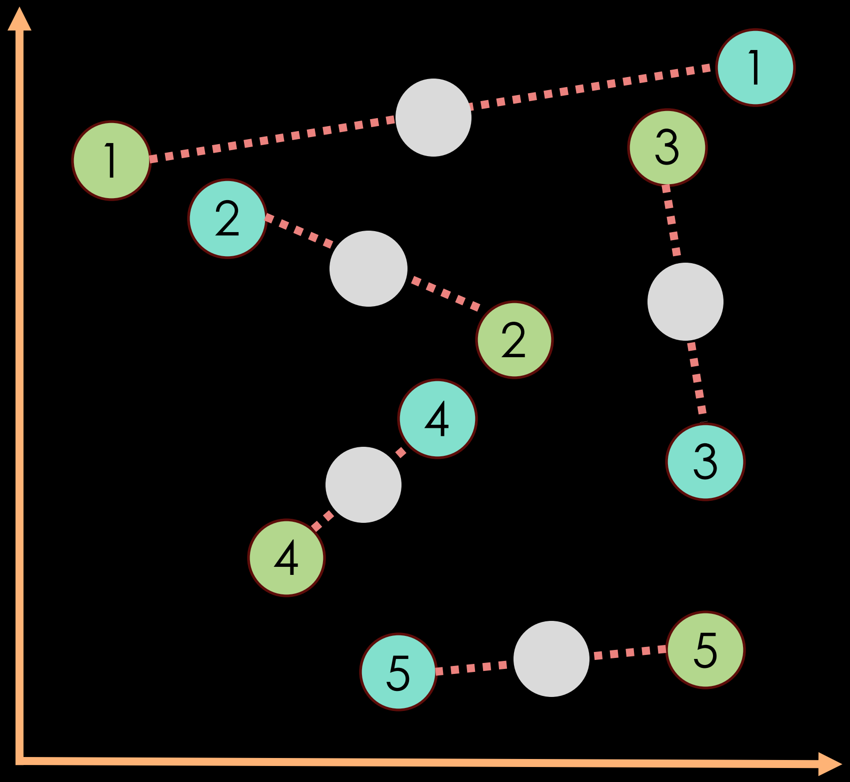


Standard Matrix Coupling
(i.e., coupling for matrix addition)



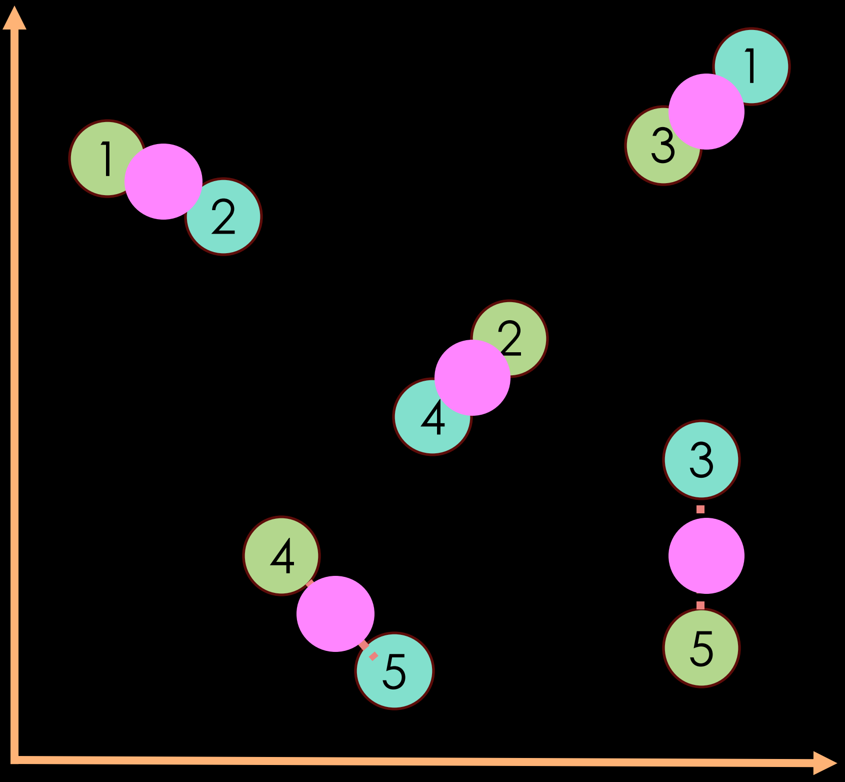
Optimal Coupling
(i.e., least work to move)

INTERPOLATION BY COUPLINGS: DIFFERENT PATHS THROUGH SPACE



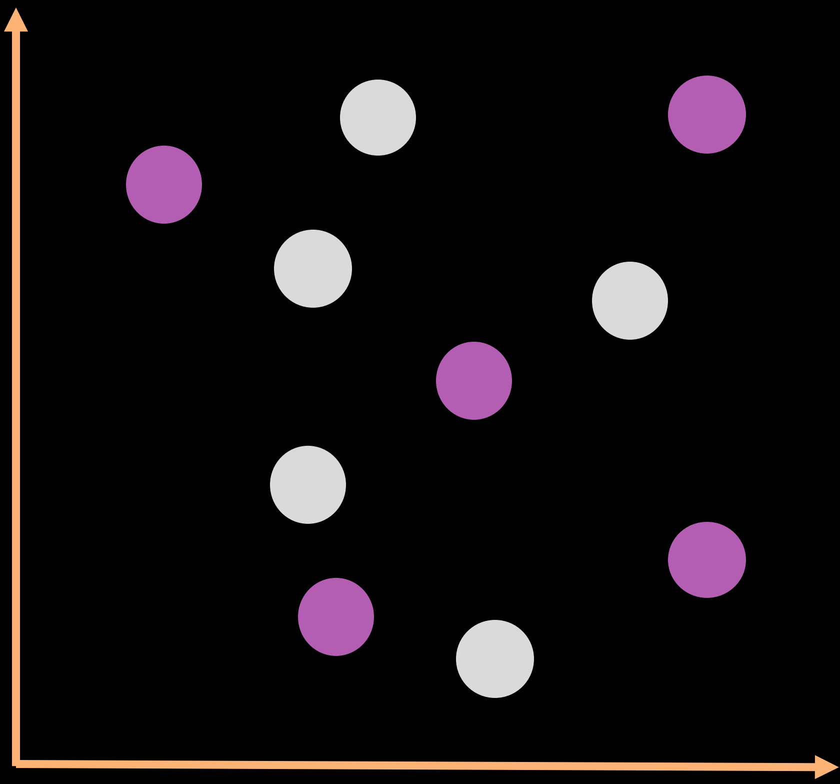
$$\frac{1}{2}[A] + \frac{1}{2}[B]$$

with respect to Linear Coupling



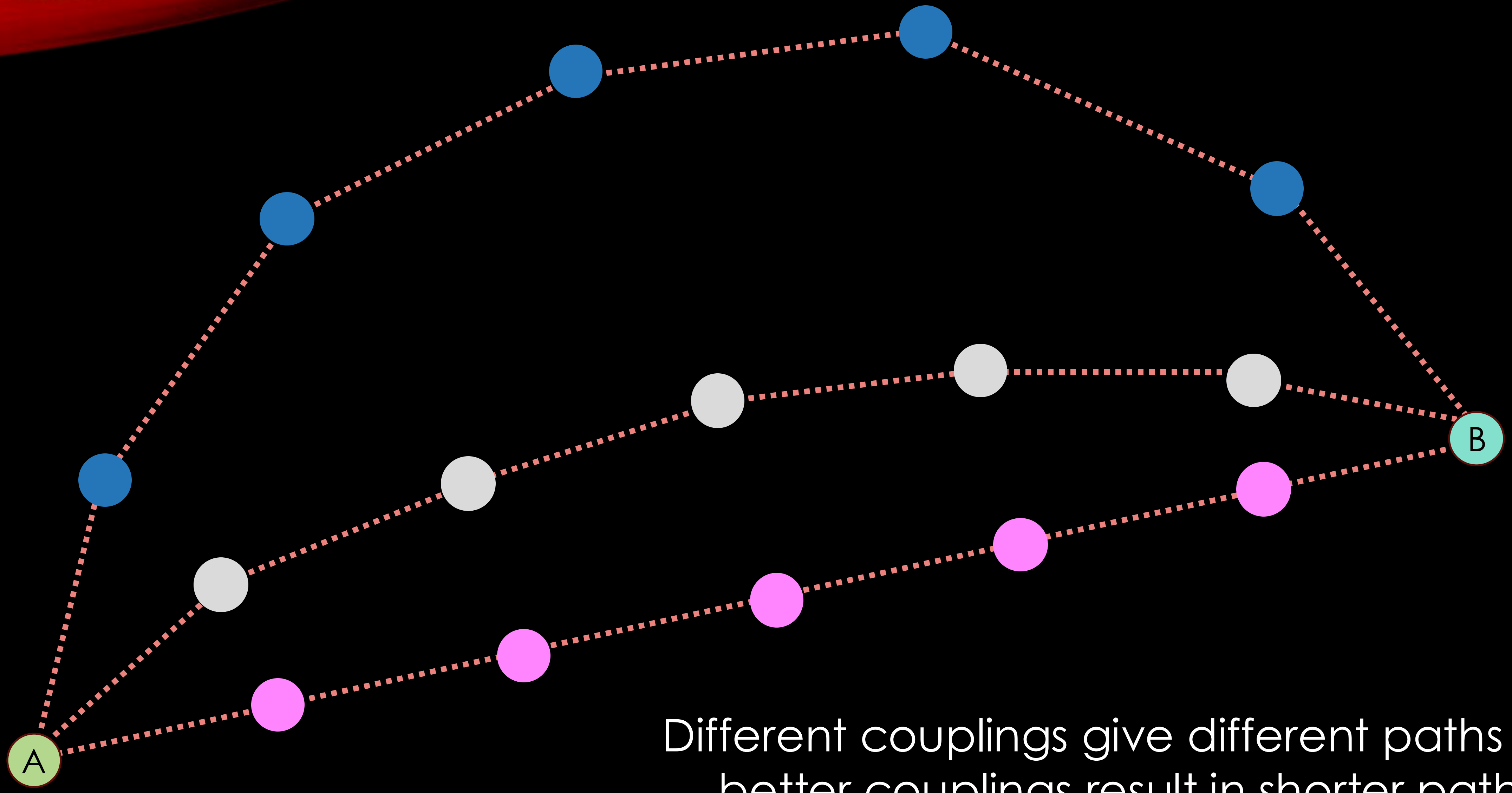
$$\frac{1}{2}[A] + \frac{1}{2}[B]$$

with respect to Optimal Coupling



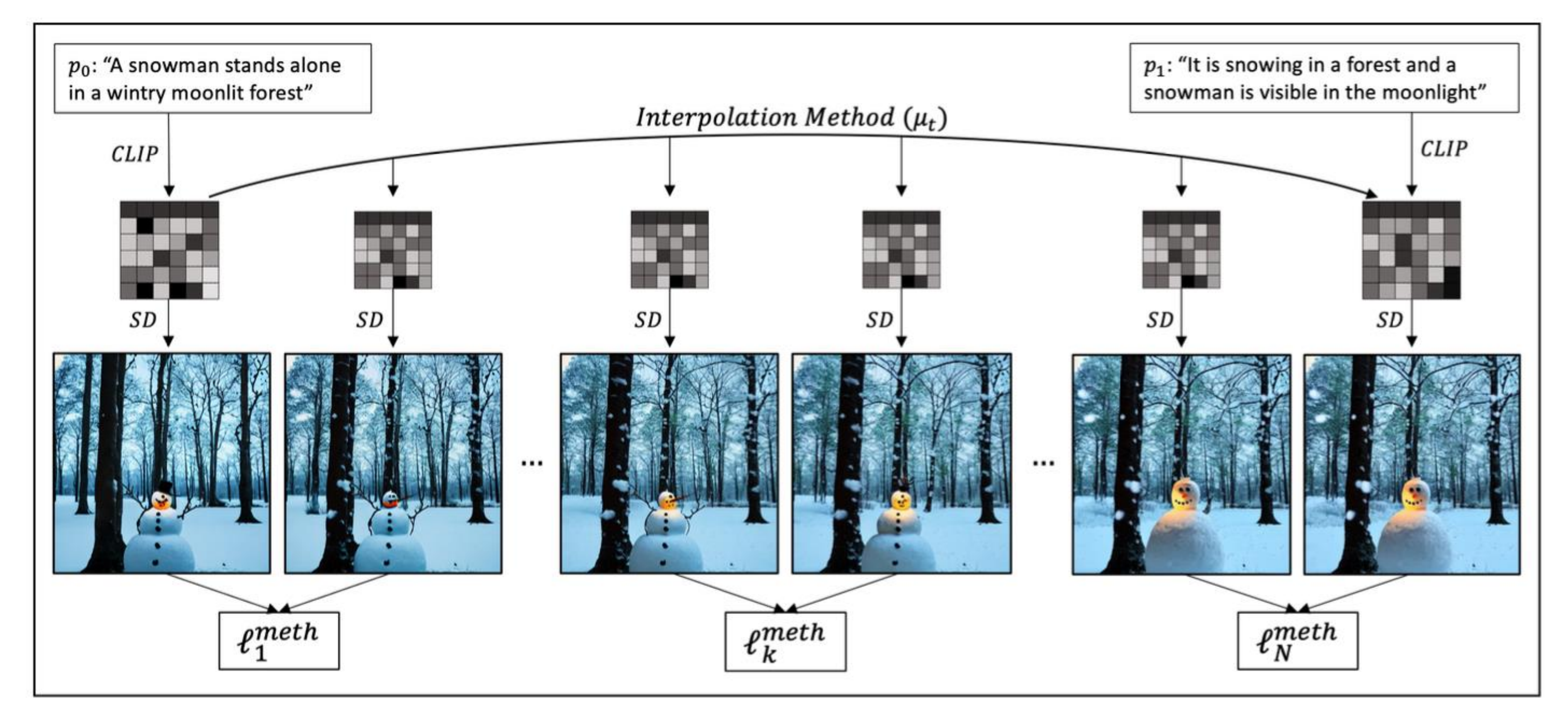
Different Couplings Yield different Intermediate States

PATHS THROUGH WASSERSTEIN SPACE



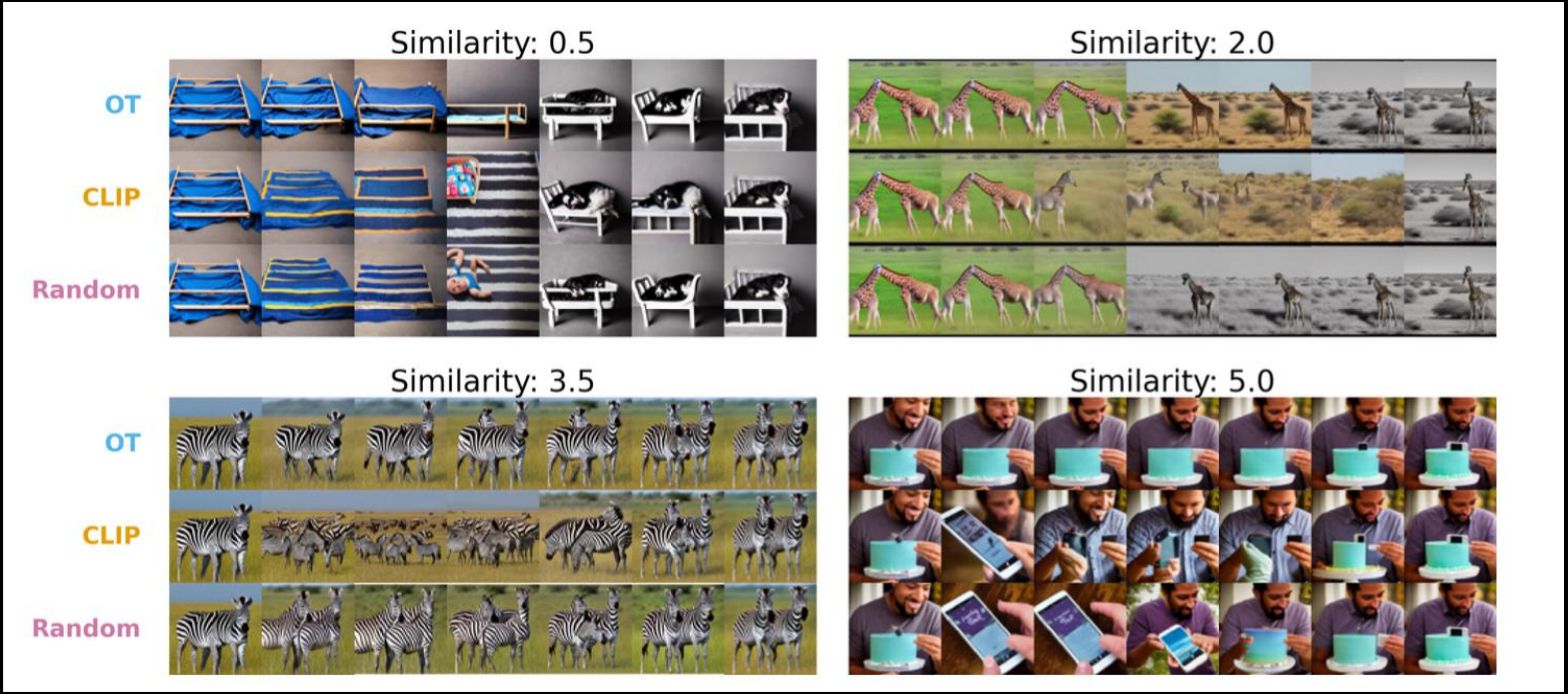
Different couplings give different paths –
better couplings result in shorter paths

RESULTING PATHS THROUGH IMAGE SPACE

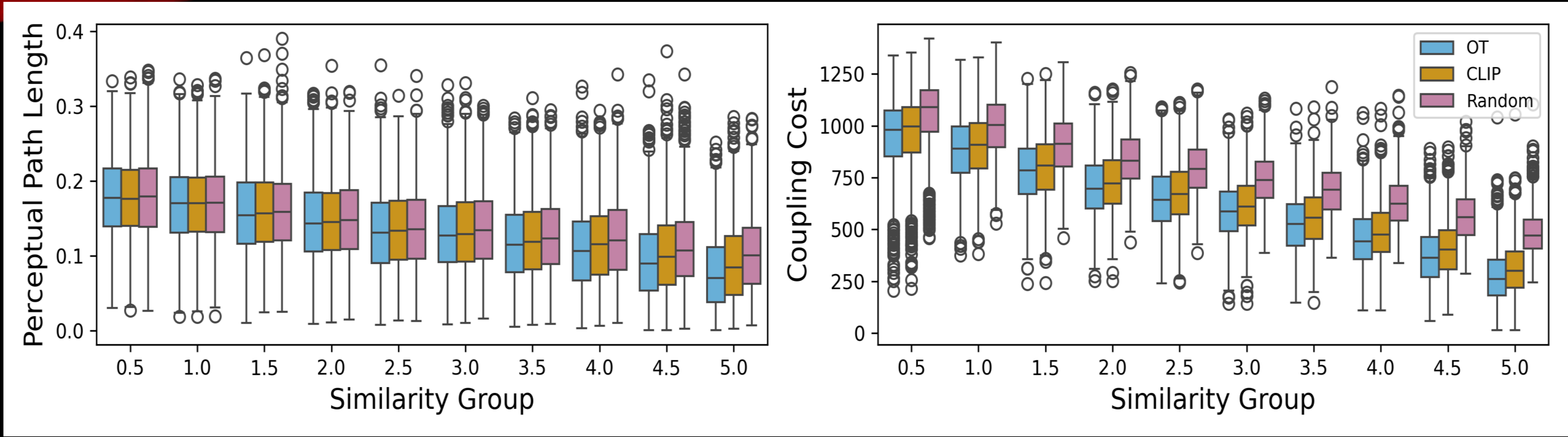


Three couplings per prompt pair: OT, CLIP, Random

EXAMPLES OF IMAGES FROM PATHS BETWEEN PROMPTS THROUGH LATENT SPACE



EXPERIMENTAL RESULTS



	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
OT vs CLIP					***	*	***	***	***	***
OT vs Random	*		**	***	***	***	***	***	***	***

Statistical significance level for each similarity group

*0.05 **0.01 ***0.001

KEY TAKE-AWAYS

- Considering CLIP embeddings as distributions of points and interpolating between the point clouds creates statistically significantly smoother paths through image space
- The more similar the two prompt concepts are, the greater the improvement one sees using the optimal coupling between point clouds
- Behavioral instability emerges more visibly for suboptimal couplings

QUESTIONS?

Contact information:
nkarris@ucsd.edu