Magnitude Distance

A Geometric Measure of Dataset Similarity

Sahel Torkamani · Henry Gouk · Rik Sarkar

School of Informatics, University of Edinburgh

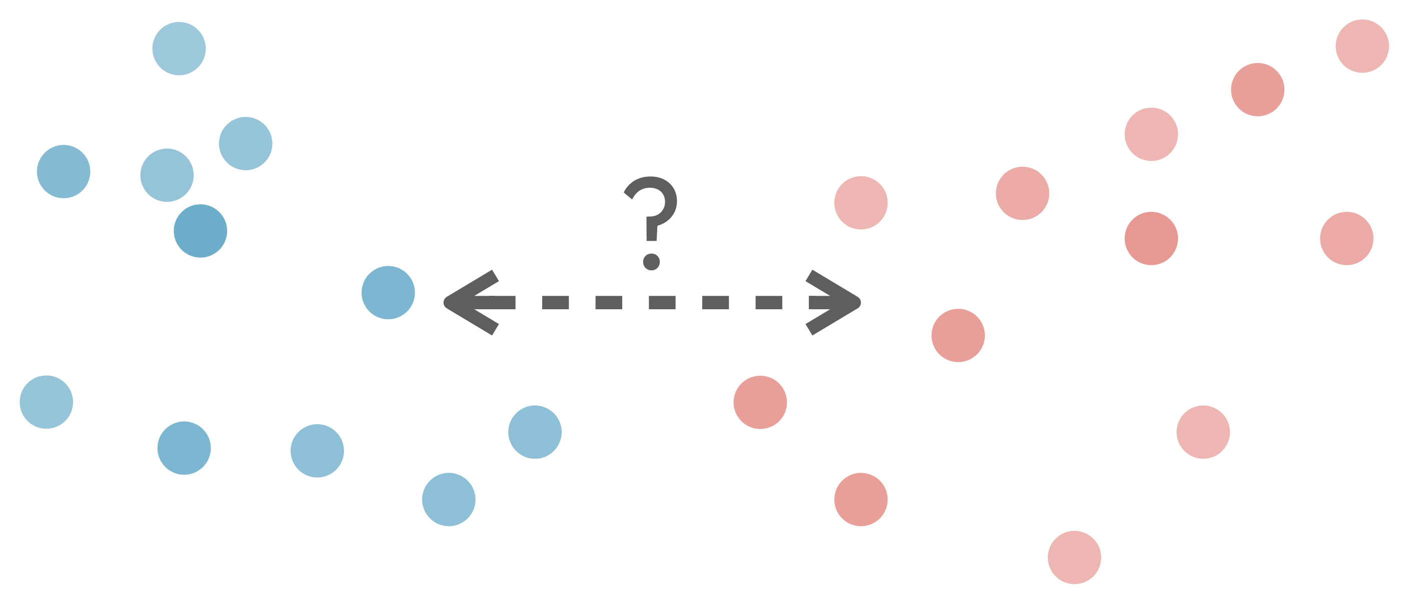
THE UNIVERSITY of EDINBURGH

ICML International Conference On Machine Learning

The Limits of Classical Distances

Classical Distances Lose Discriminability in High Dimensions.

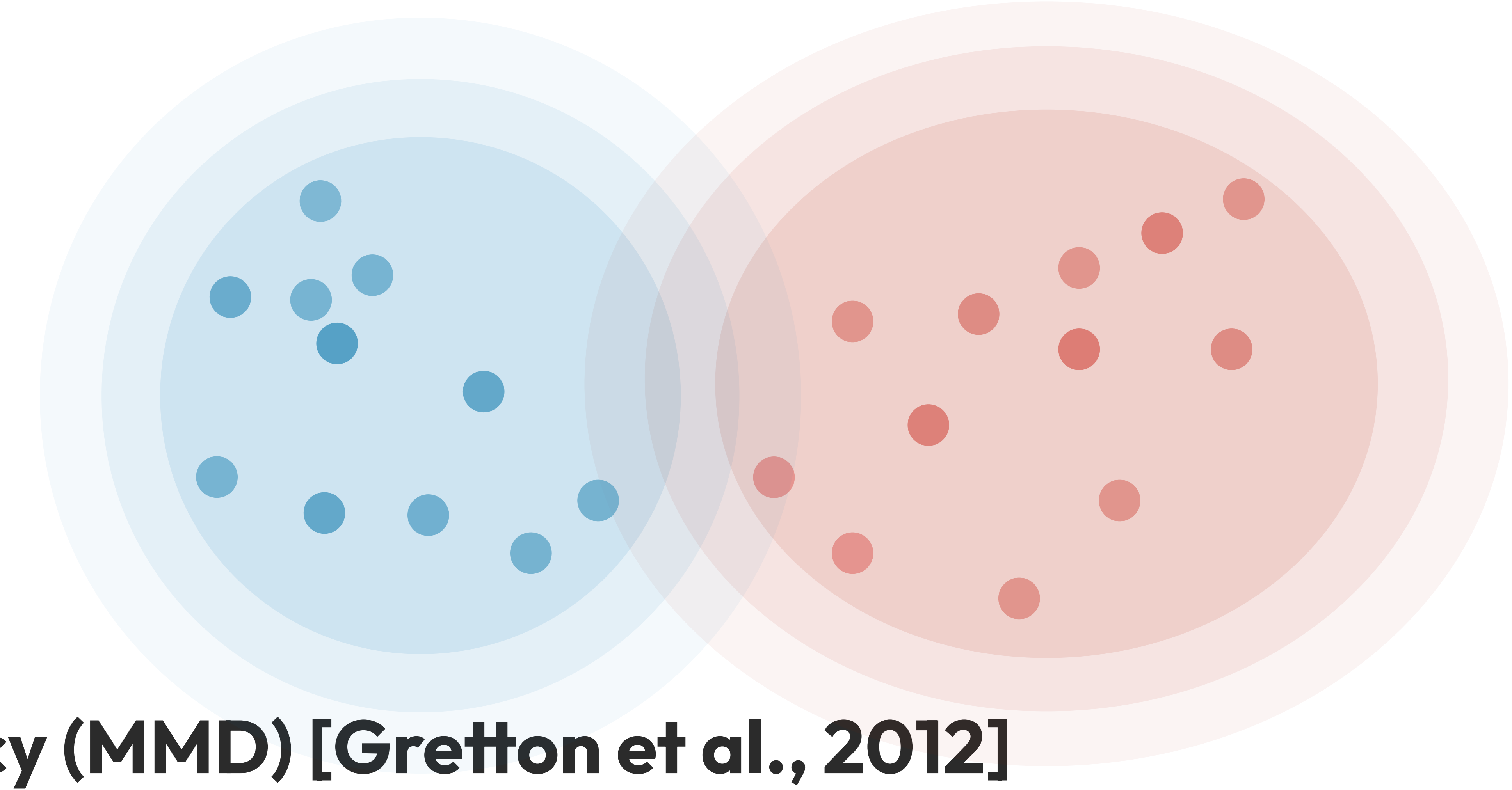
Comparing datasets is fundamental, from hypothesis testing to training generative models.



The Limits of Classical Distances

Classical Distances Lose Discriminability in High Dimensions.

Comparing datasets is fundamental, from hypothesis testing to training generative models.



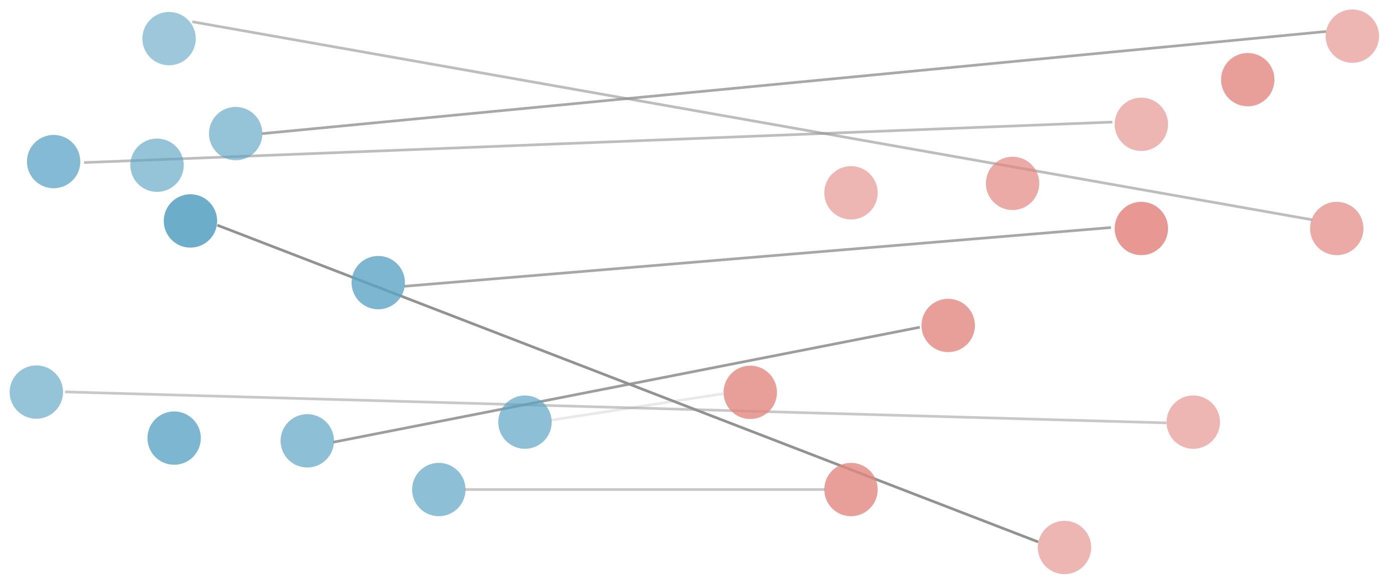
✗ **Maximum Mean Discrepancy (MMD) [Gretton et al., 2012]**

- Kernel matrix eigenvalues concentrate, dominated by the top directions.
- The distance collapses onto a low-dimensional subspace.

The Limits of Classical Distances

Classical Distances Lose Discriminability in High Dimensions.

Comparing datasets is fundamental, from hypothesis testing to training generative models.



✗ **Maximum Mean Discrepancy (MMD) [Gretton et al., 2012]**

- Kernel matrix eigenvalues concentrate, dominated by the top directions.
- The distance collapses onto a low-dimensional subspace.

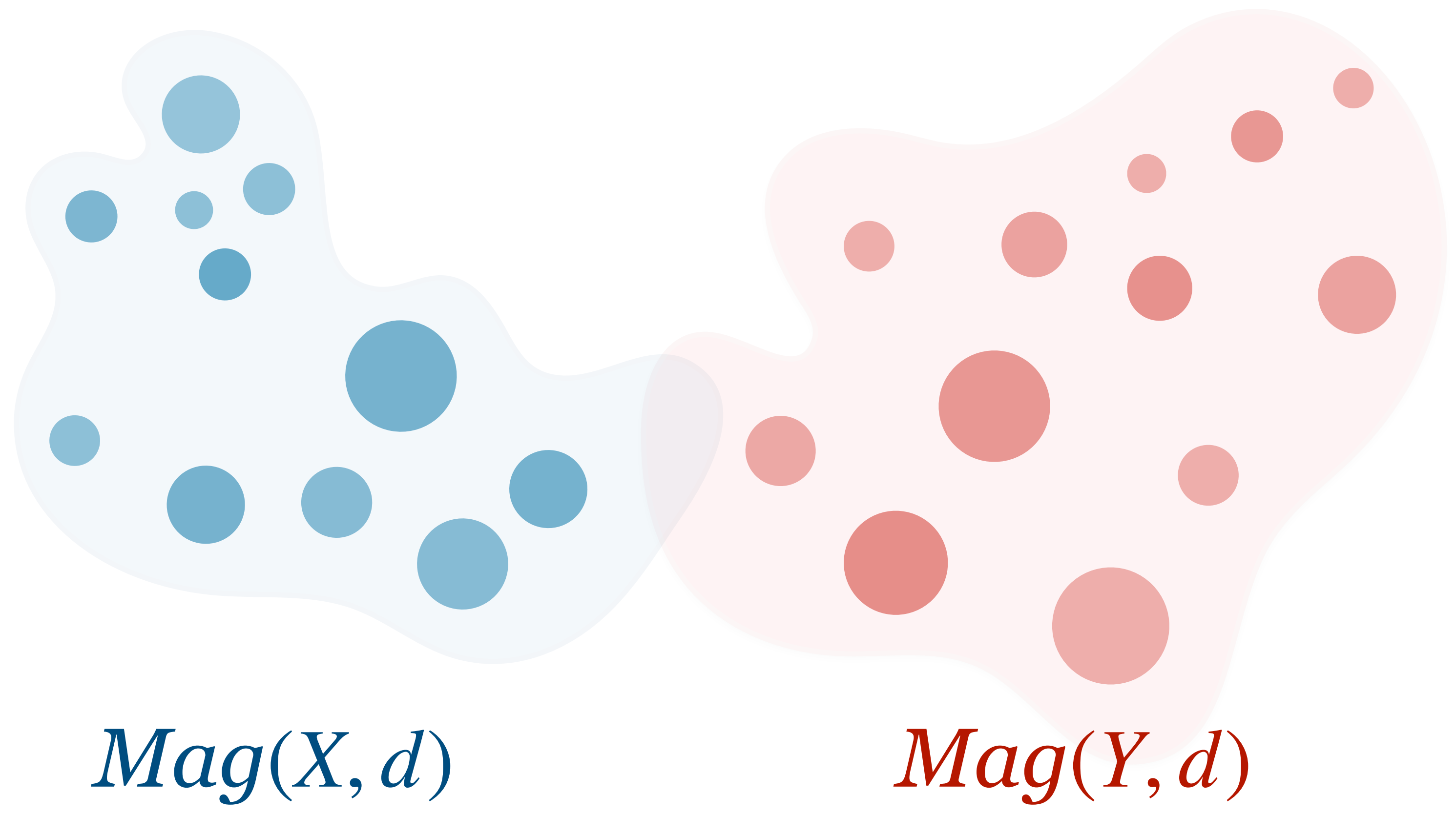
✗ **Wasserstein distance [Givens & Shortt, 1984]**

- Pairwise distances concentrate in high dimensions.
- Unbounded sensitivity to outliers: a single outlier can increase the distance arbitrarily.

The Key Concept

Magnitude of a Metric Space [Leinster (2013)]

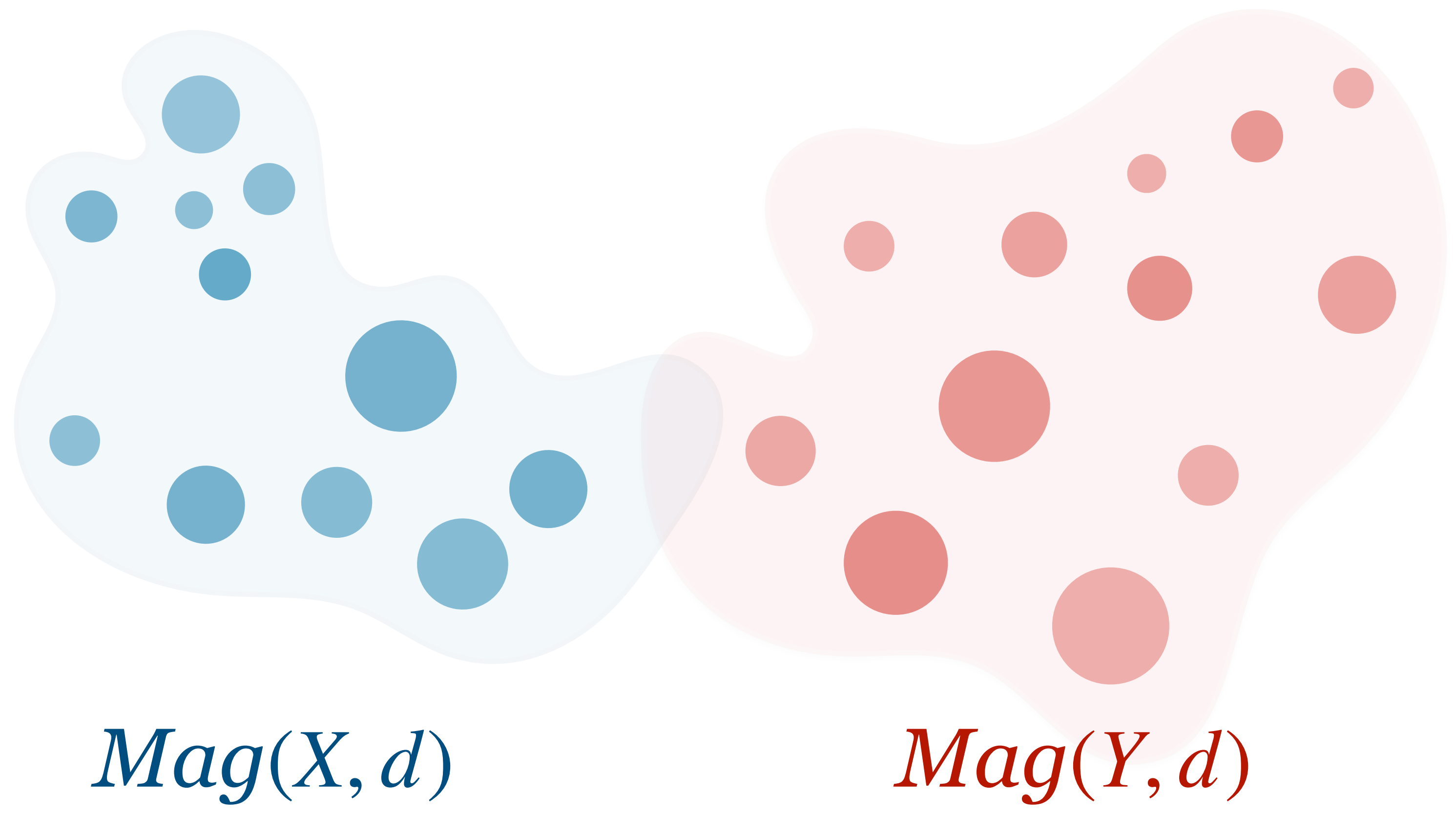
Definition: Magnitude of a finite metric space (X, d) is defined as $Mag(X, d) = \mathbf{1}^T \zeta_X^{-1} \mathbf{1}$ with the similarity matrix $\zeta_X(x_i, x_j) := \exp(-d(x_i, x_j))$.



The Key Concept

Magnitude of a Metric Space [Leinster (2013)]

Definition: Magnitude of a finite metric space (X, d) is defined as $Mag(X, d) = \mathbf{1}^T \zeta_X^{-1} \mathbf{1}$ with the similarity matrix $\zeta_X(x_i, x_j) := \exp(-d(x_i, x_j))$.

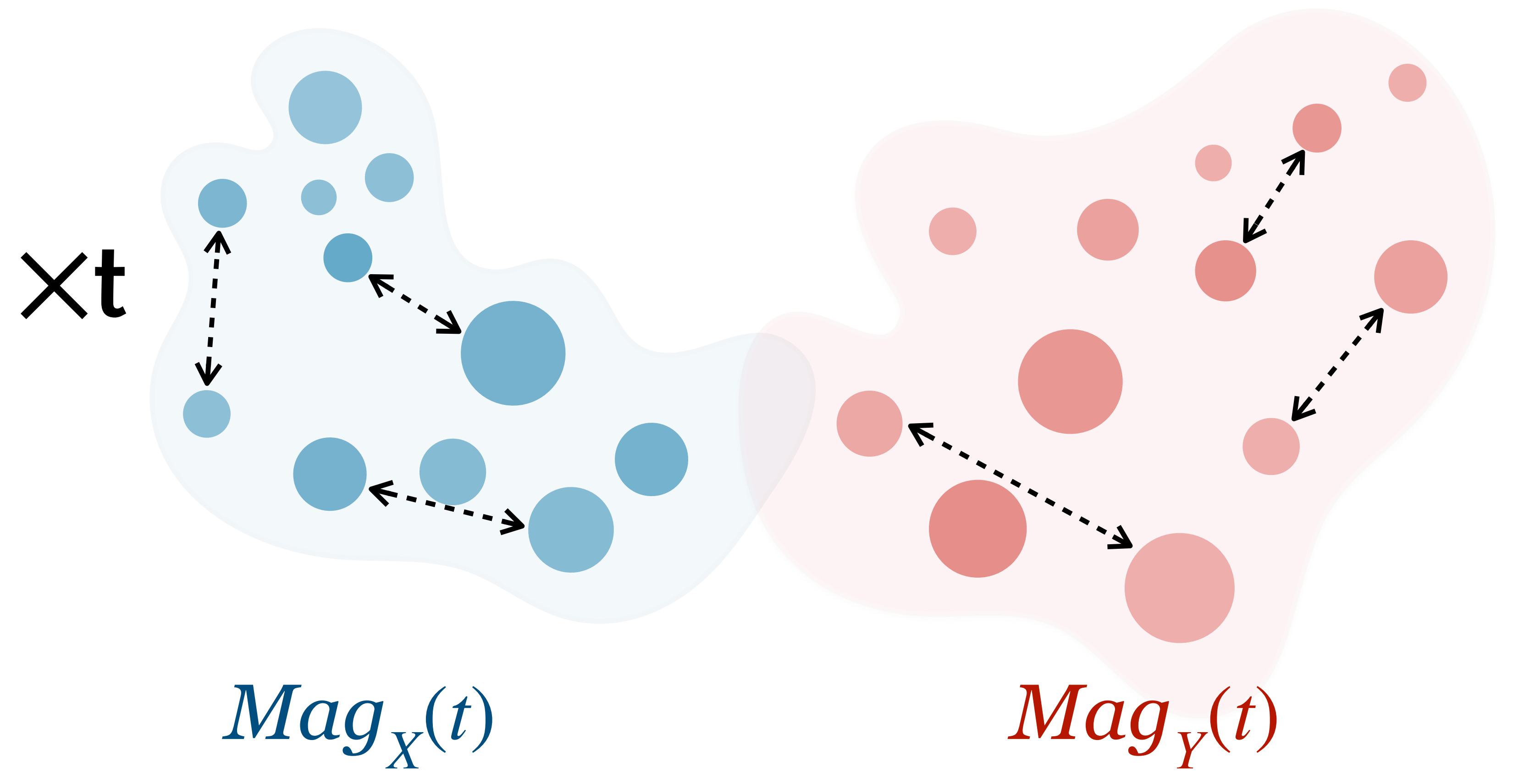


*Intuitively, magnitude summarizes **the effective number of distinct points** that live in a metric space: Points in sparse regions contribute more, dense clusters less.*

The Key Concept

Magnitude of a Metric Space [Leinster (2013)]

Definition: Magnitude of a finite metric space (X, d) with scale parameter $t \in \mathbb{R}_+$ is defined as $Mag_X(t) = \mathbf{1}^T \zeta_{tX}^{-1} \mathbf{1}$ with the similarity matrix $\zeta_{tX}(x_i, x_j) := \exp(-td(x_i, x_j))$.

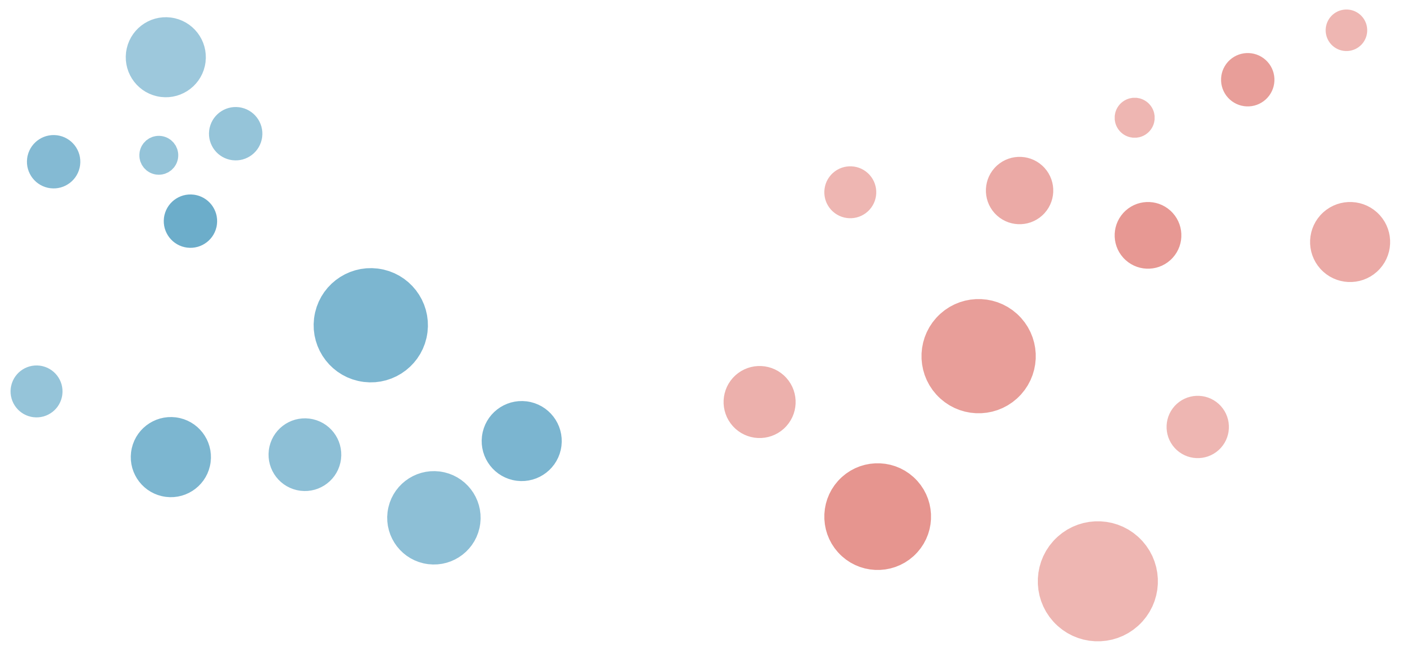


*Intuitively, magnitude summarizes **the effective number of distinct points** that live in a metric space: Points in sparse regions contribute more, dense clusters less.*

Magnitude Distance

This Work

Definition: For two finite sets $X, Y \subset \mathbb{R}^D$, the magnitude distance with scale parameter $t \in \mathbb{R}_+$ is defined as

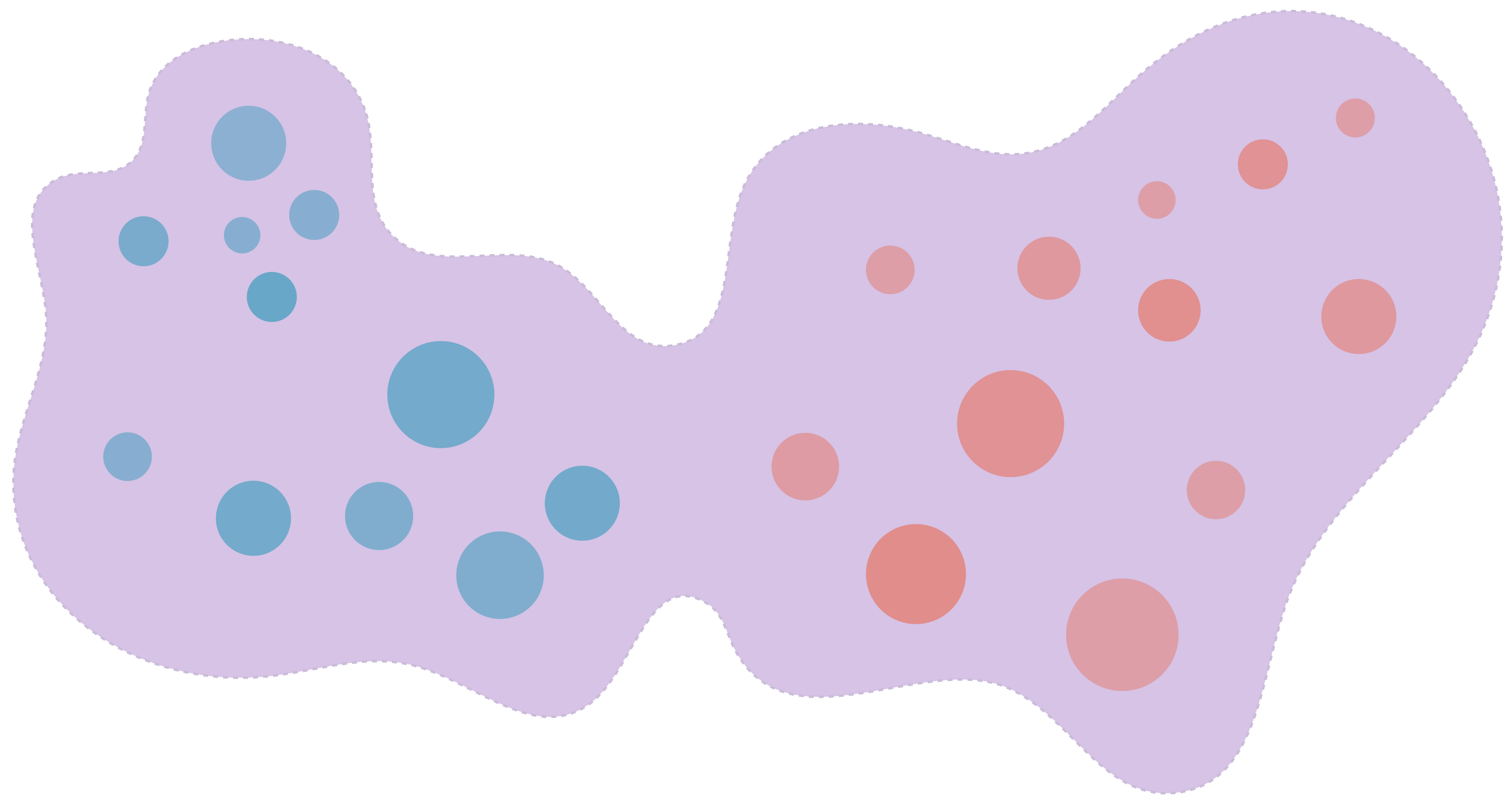


Magnitude Distance

This Work

Definition: For two finite sets $X, Y \subset \mathbb{R}^D$, the magnitude distance with scale parameter $t \in \mathbb{R}_+$ is defined as

$$d^t_{Mag}(X, Y) = 2\mathrm{Mag}_{X \cup Y}(t)$$

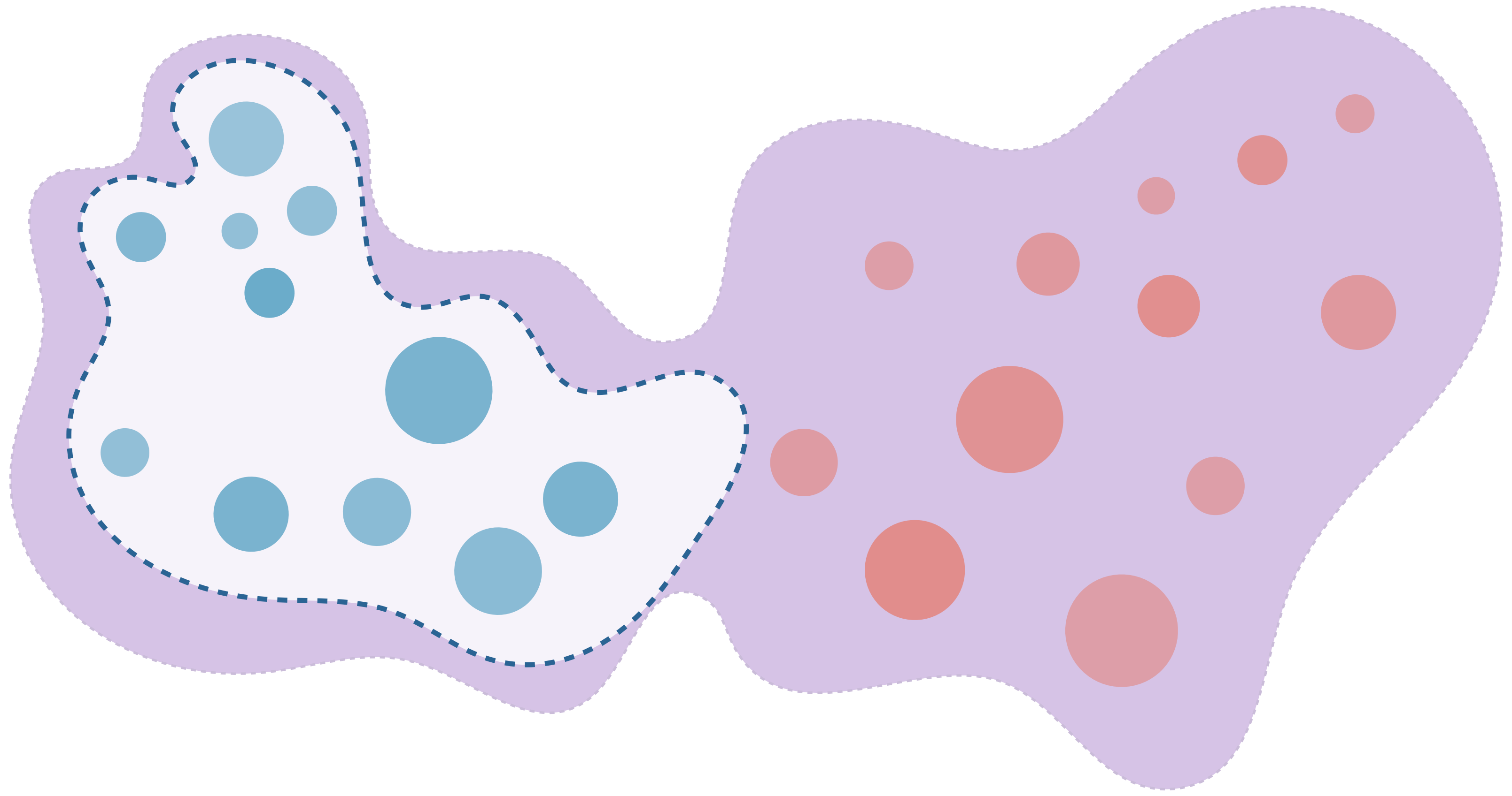


Magnitude Distance

This Work

Definition: For two finite sets $X, Y \subset \mathbb{R}^D$, the magnitude distance with scale parameter $t \in \mathbb{R}_+$ is defined as

$$d^t_{Mag}(X, Y) = 2\mathrm{Mag}_{X \cup Y}(t) - \mathrm{Mag}_X(t)$$

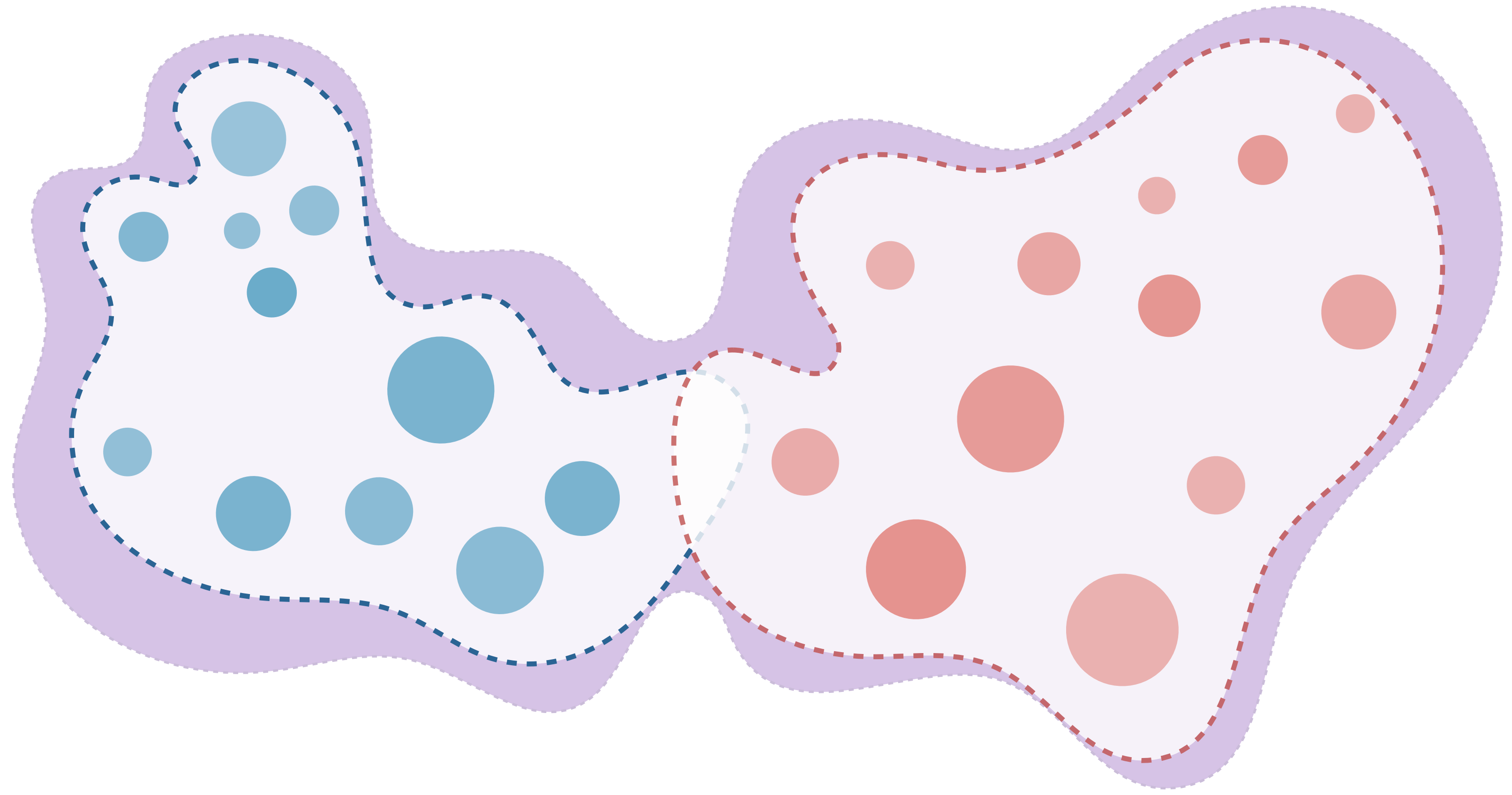


Magnitude Distance

This Work

Definition: For two finite sets $X, Y \subset \mathbb{R}^D$, the magnitude distance with scale parameter $t \in \mathbb{R}_+$ is defined as

$$d^t_{Mag}(X, Y) = 2\mathrm{Mag}_{X \cup Y}(t) - \mathrm{Mag}_X(t) - \mathrm{Mag}_Y(t).$$

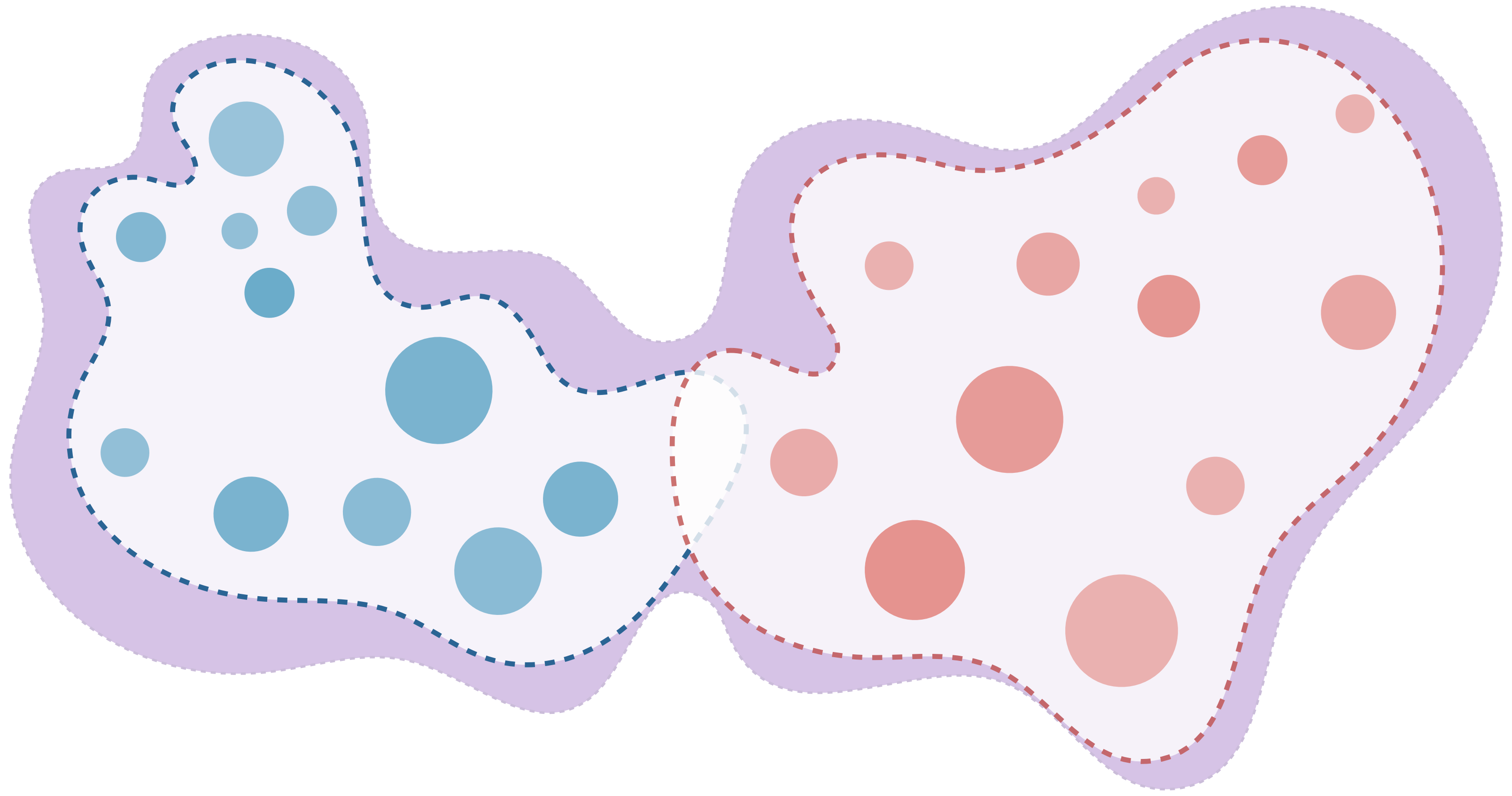


Magnitude Distance

This Work

Definition: For two finite sets $X, Y \subset \mathbb{R}^D$, the magnitude distance with scale parameter $t \in \mathbb{R}_+$ is defined as

$$d^t_{Mag}(X, Y) = 2\mathrm{Mag}_{X \cup Y}(t) - \mathrm{Mag}_X(t) - \mathrm{Mag}_Y(t).$$



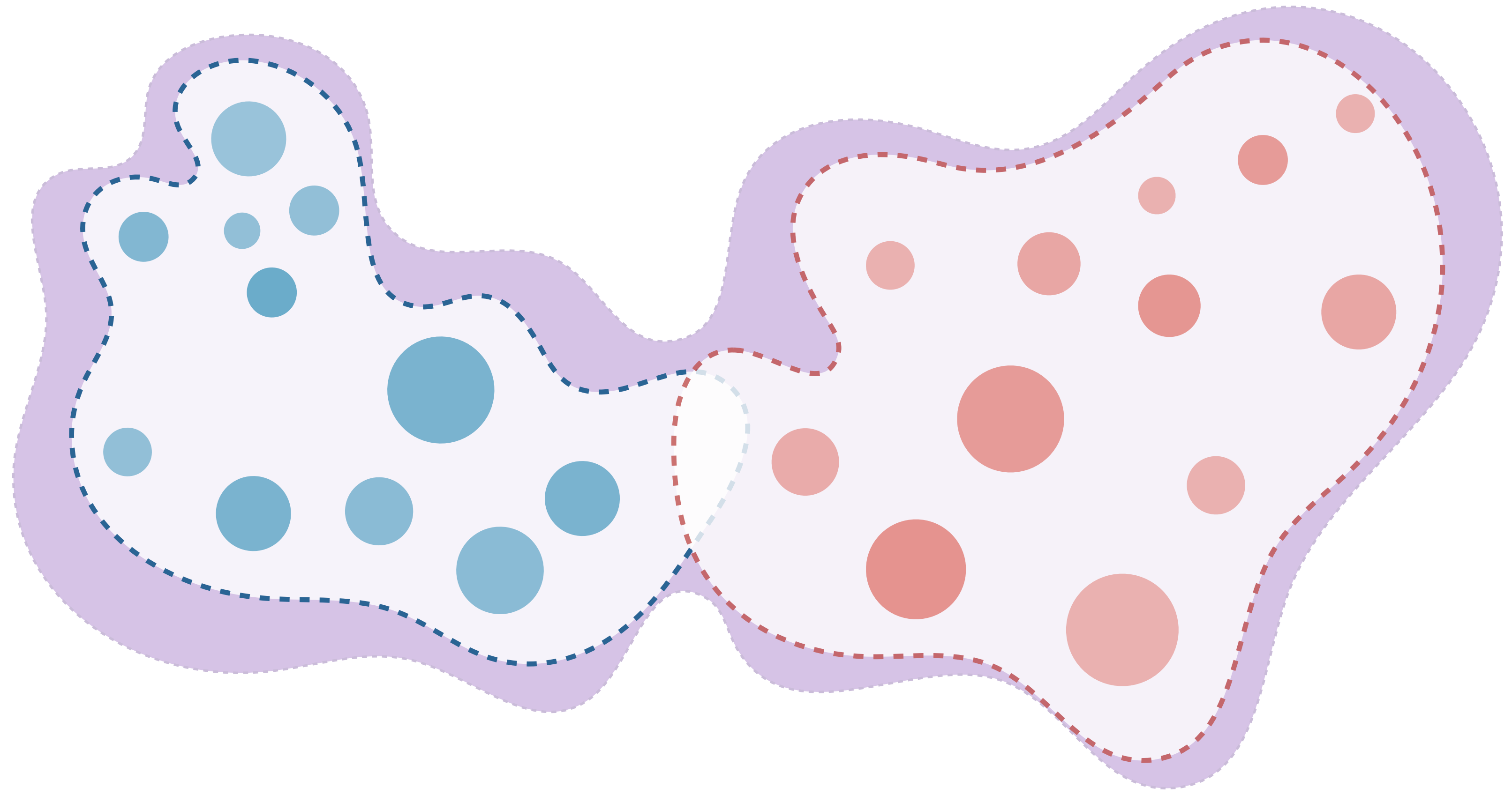
Metric axioms:

Magnitude Distance

This Work

Definition: For two finite sets $X, Y \subset \mathbb{R}^D$, the magnitude distance with scale parameter $t \in \mathbb{R}_+$ is defined as

$$d^t_{Mag}(X, Y) = 2\mathrm{Mag}_{X \cup Y}(t) - \mathrm{Mag}_X(t) - \mathrm{Mag}_Y(t).$$



Metric axioms:

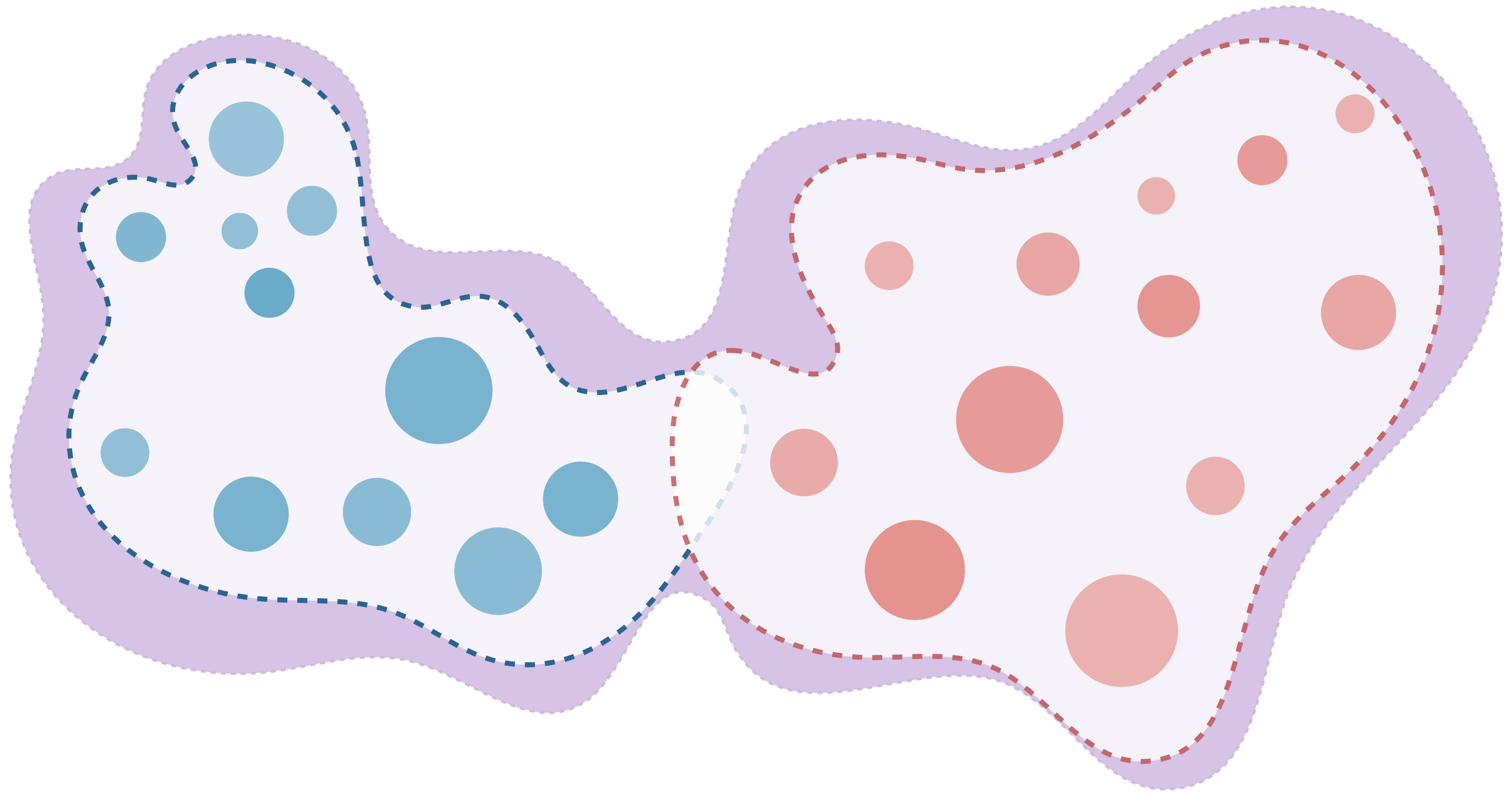
✔ Symmetry

Magnitude Distance

This Work

Definition: For two finite sets $X, Y \subset \mathbb{R}^D$, the magnitude distance with scale parameter $t \in \mathbb{R}_+$ is defined as

$$d^t_{Mag}(X, Y) = 2\text{Mag}_{X \cup Y}(t) - \text{Mag}_X(t) - \text{Mag}_Y(t).$$



Metric axioms:

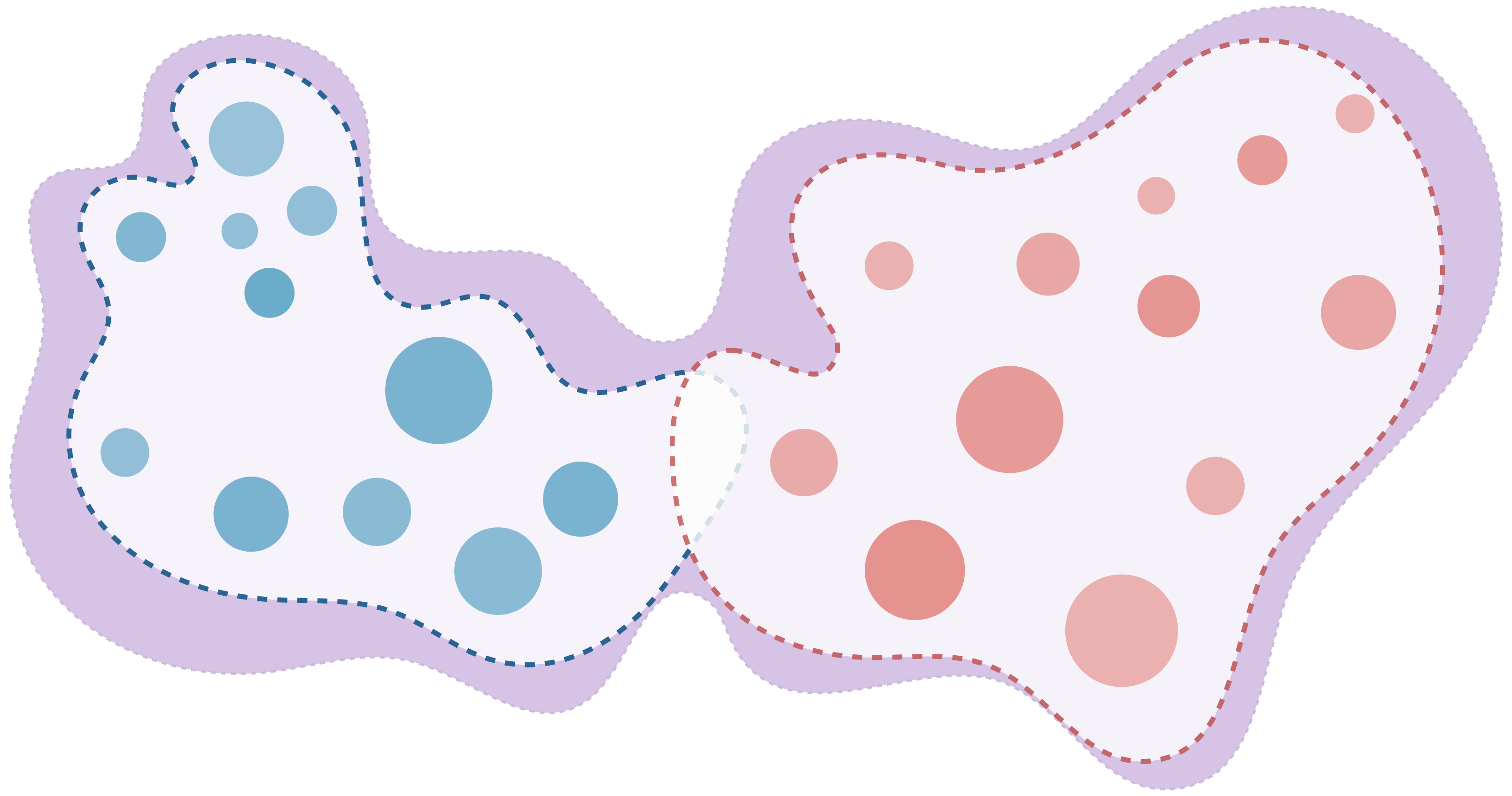
- ✓ Symmetry
- ✓ Non-negativity

Magnitude Distance

This Work

Definition: For two finite sets $X, Y \subset \mathbb{R}^D$, the magnitude distance with scale parameter $t \in \mathbb{R}_+$ is defined as

$$d^t_{Mag}(X, Y) = 2\text{Mag}_{X \cup Y}(t) - \text{Mag}_X(t) - \text{Mag}_Y(t).$$



Metric axioms:

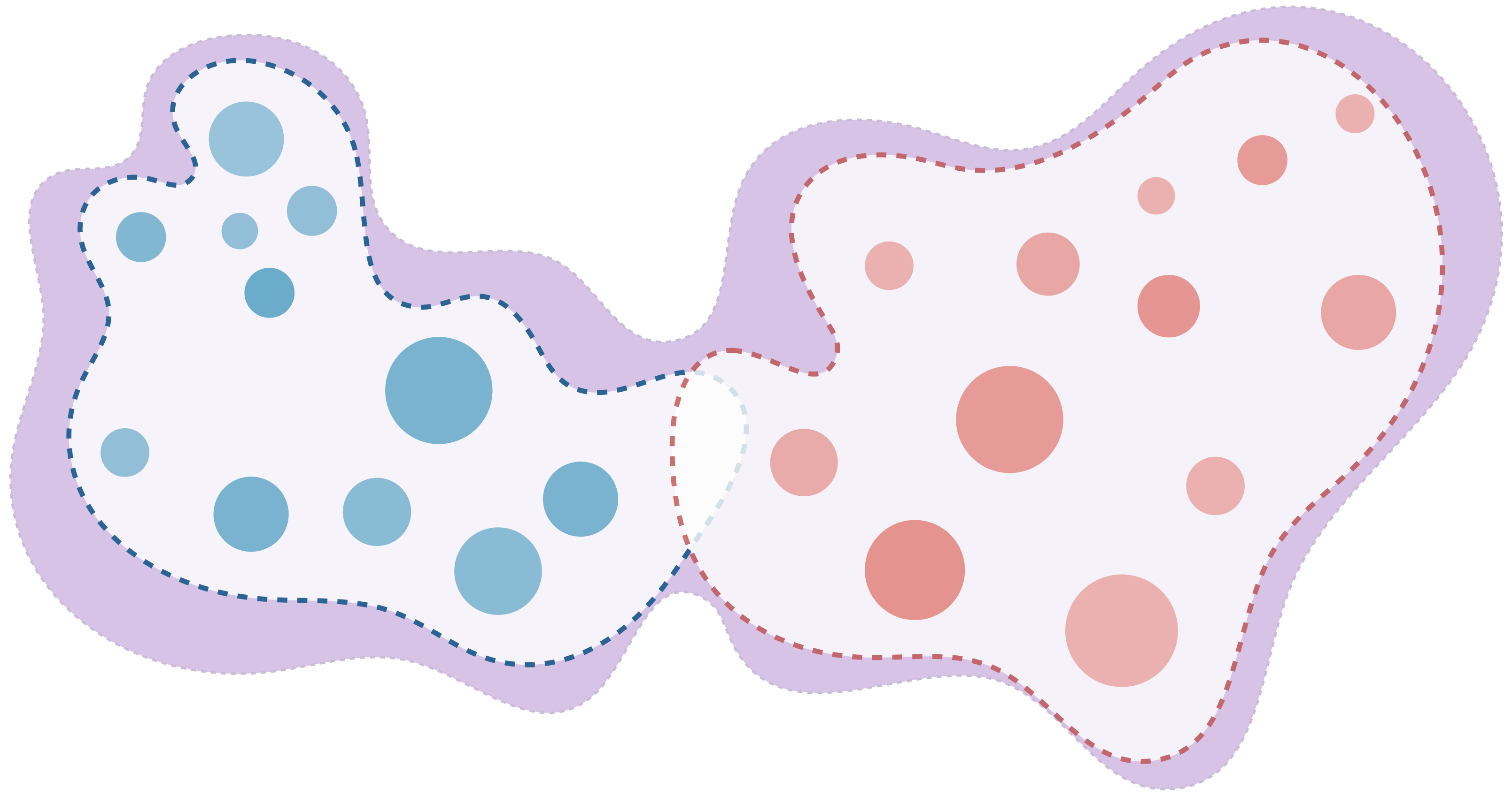
- ✓ Symmetry
- ✓ Non-negativity
- ✓ Identity of indiscernibles (under mild conditions)

Magnitude Distance

This Work

Definition: For two finite sets $X, Y \subset \mathbb{R}^D$, the magnitude distance with scale parameter $t \in \mathbb{R}_+$ is defined as

$$d^t_{Mag}(X, Y) = 2\mathrm{Mag}_{X\cup Y}(t) - \mathrm{Mag}_X(t) - \mathrm{Mag}_Y(t).$$



Metric axioms:

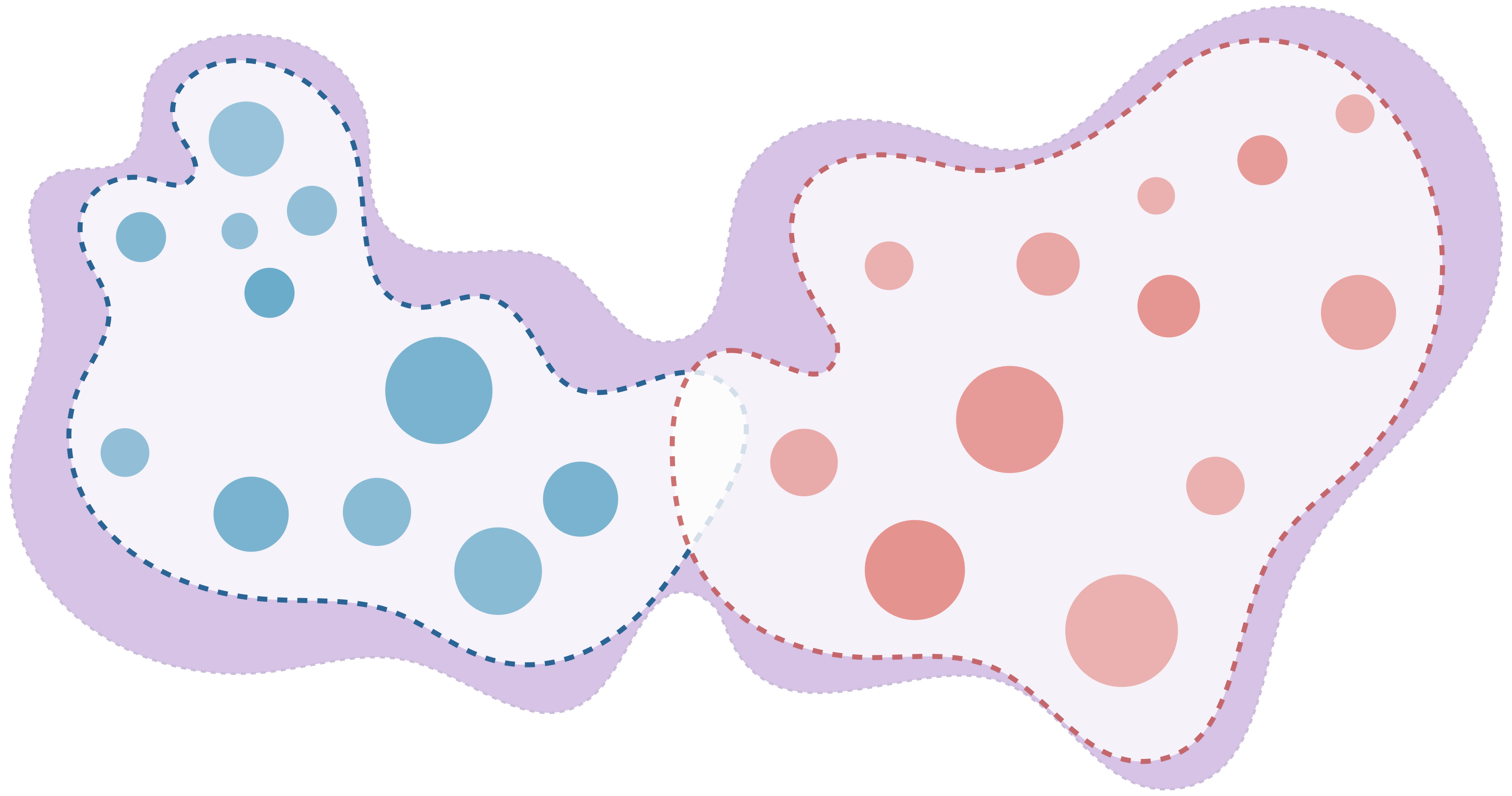
- ✓ Symmetry
- ✓ Non-negativity
- ✓ Identity of indiscernibles (under mild conditions)
- ✗ Triangle inequality

Magnitude Distance

This Work

Definition: For two finite sets $X, Y \subset \mathbb{R}^D$, the magnitude distance with scale parameter $t \in \mathbb{R}_+$ is defined as

$$d^t_{Mag}(X, Y) = 2\mathrm{Mag}_{X \cup Y}(t) - \mathrm{Mag}_X(t) - \mathrm{Mag}_Y(t).$$



Metric axioms:

- ✓ Symmetry
- ✓ Non-negativity
- ✓ Identity of indiscernibles (under mild conditions)
- ✗ Triangle inequality

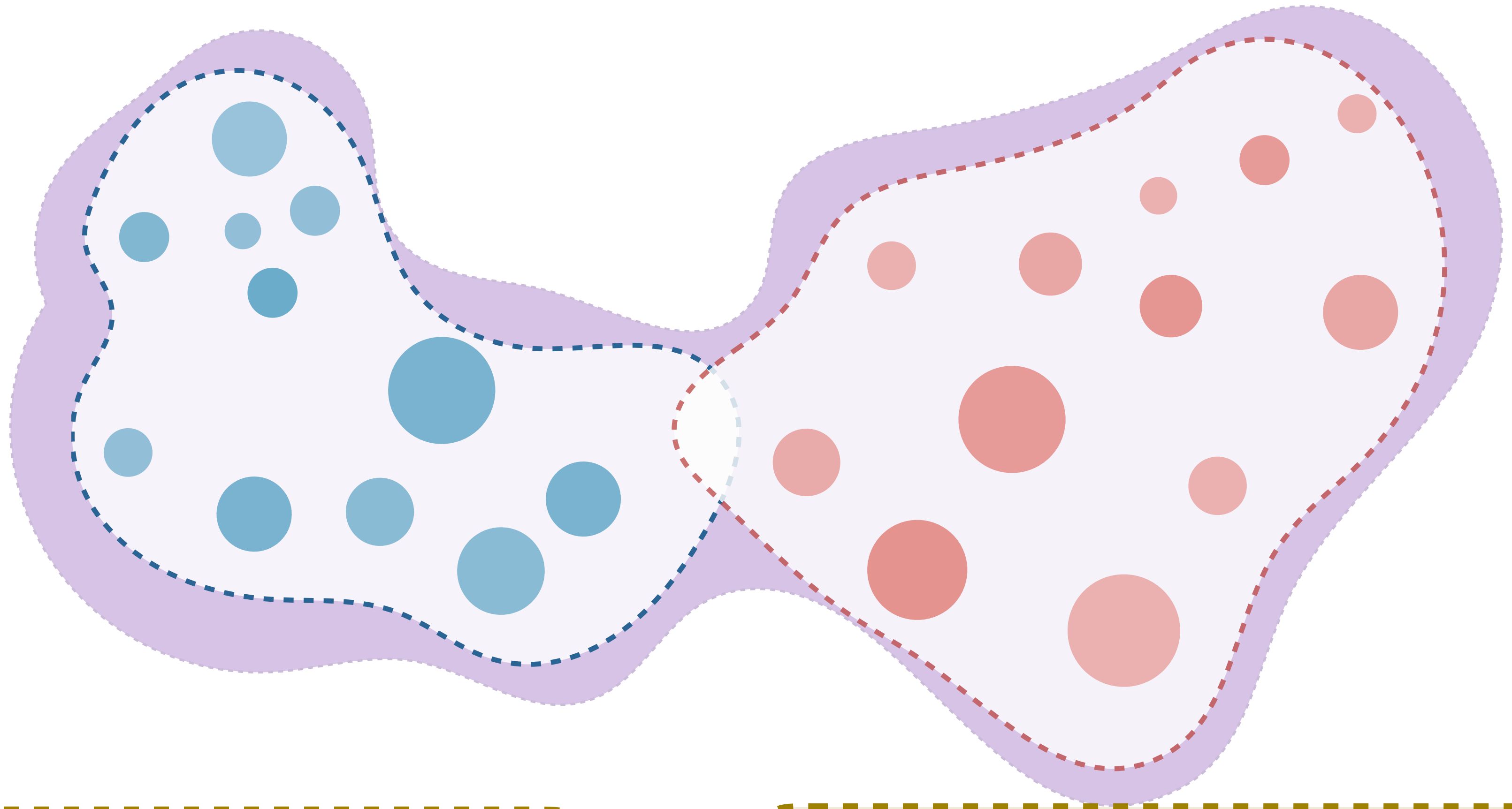
Behaviour over values of t:

Magnitude Distance

This Work

Definition: For two finite sets $X, Y \subset \mathbb{R}^D$, the magnitude distance with scale parameter $t \in \mathbb{R}_+$ is defined as

$$d^t_{Mag}(X, Y) = 2\text{Mag}_{X \cup Y}(t) - \text{Mag}_X(t) - \text{Mag}_Y(t).$$



Metric axioms:

- ✓ Symmetry
- ✓ Non-negativity
- ✓ Identity of indiscernibles (under mild conditions)
- ✗ Triangle inequality

Behaviour over values of t:

Small: captures global structure

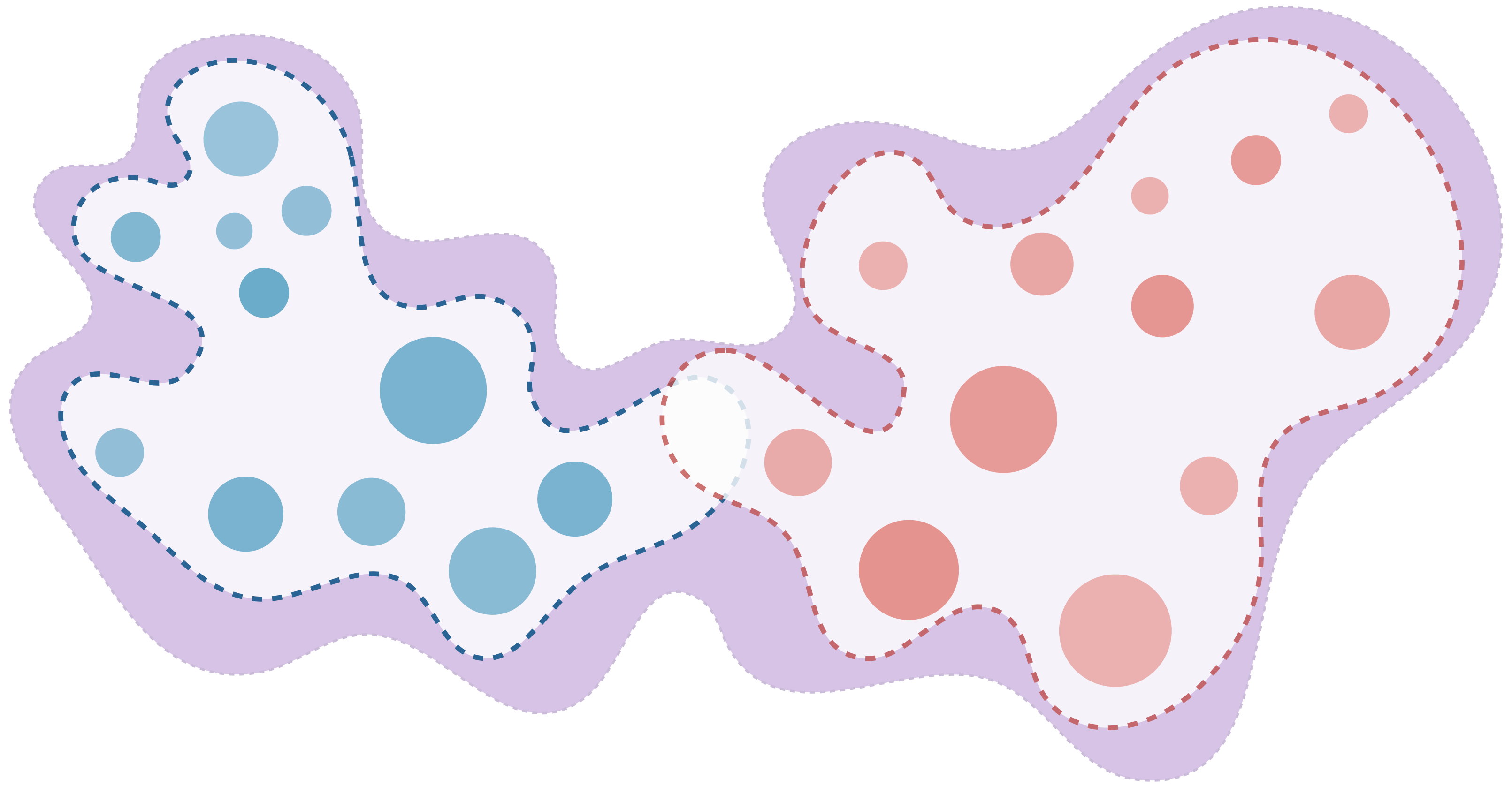
Large: captures local fine details

Magnitude Distance

This Work

Definition: For two finite sets $X, Y \subset \mathbb{R}^D$, the magnitude distance with scale parameter $t \in \mathbb{R}_+$ is defined as

$$d^t_{Mag}(X, Y) = 2\mathrm{Mag}_{X \cup Y}(t) - \mathrm{Mag}_X(t) - \mathrm{Mag}_Y(t).$$



Metric axioms:

- ✓ Symmetry
- ✓ Non-negativity
- ✓ Identity of indiscernibles (under mild conditions)
- ✗ Triangle inequality

Behaviour over values of t:

Small: captures global structure

Large: captures local fine details

Theoretical Guarantees

Comparison with Classical Distances

High Dimension Discriminability

- MMD is dominated by top eigen directions.
- Magnitude uses inverse eigenvalue weightings: the remaining eigenvalues directions contribute.
- With suitable t, magnitude distance prevents collapsing to zero at any dimension.

Outlier Robustness

- Wasserstein distance changes unboundedly with one outlier addition.
- Magnitude distance's changes is bounded with suitable t.

Theoretical Guarantees

Comparison with Classical Distances

High Dimension Discriminability

- MMD is dominated by top eigen directions.
- Magnitude uses inverse eigenvalue weightings: the remaining eigenvalues directions contribute.
- With suitable t, magnitude distance prevents collapsing to zero at any dimension.

Outlier Robustness

- Wasserstein distance changes unboundedly with one outlier addition.
- Magnitude distance’s changes is bounded with suitable t.

Magnitude Generative Network (MagGN)

Definition

Definition: For some $k \in \mathbb{N}$ and a list called t-steps $\{e_1, \ldots, e_k\}$, at epochs e_i we add $\hat{d}^{t_i}_{Mag}(\mathscr{D}_r, \mathscr{D}_g)$ to the loss, respectively, with the magnitude scaling parameters t_i where $t_1 \leq \ldots \leq t_k$.

$$l_e = \sum_{i:e_i \leq e} \hat{d}^{t_i}_{Mag}(\mathscr{D}_r, \mathscr{D}_g).$$

Magnitude Generative Network (MagGN)

Definition

Definition: For some $k \in \mathbb{N}$ and a list called t-steps $\{e_1, \ldots, e_k\}$, at epochs e_i we add $\hat{d}^{t_i}_{Mag}(\mathcal{D}_r, \mathcal{D}_g)$ to the loss, respectively, with the magnitude scaling parameters t_i where $t_1 \leq \ldots \leq t_k$.

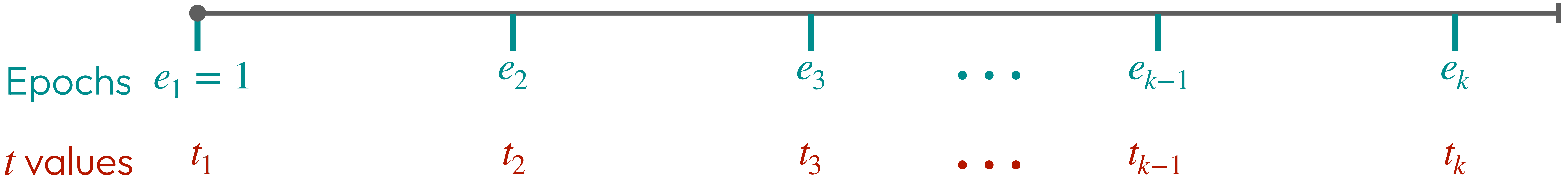
$$l_e = \sum_{i:e_i \leq e} \hat{d}^{t_i}_{Mag}(\mathcal{D}_r, \mathcal{D}_g).$$

Magnitude Generative Network (MagGN)

Definition

Definition: For some $k \in \mathbb{N}$ and a list called t-steps $\{e_1, \ldots, e_k\}$, at epochs e_i we add $\hat{d}^{t_i}_{Mag}(\mathcal{D}_r, \mathcal{D}_g)$ to the loss, respectively, with the magnitude scaling parameters t_i where $t_1 \leq \ldots \leq t_k$.

$$l_e = \sum_{i:e_i \leq e} \hat{d}^{t_i}_{Mag}(\mathcal{D}_r, \mathcal{D}_g).$$



Magnitude Generative Network (MagGN)

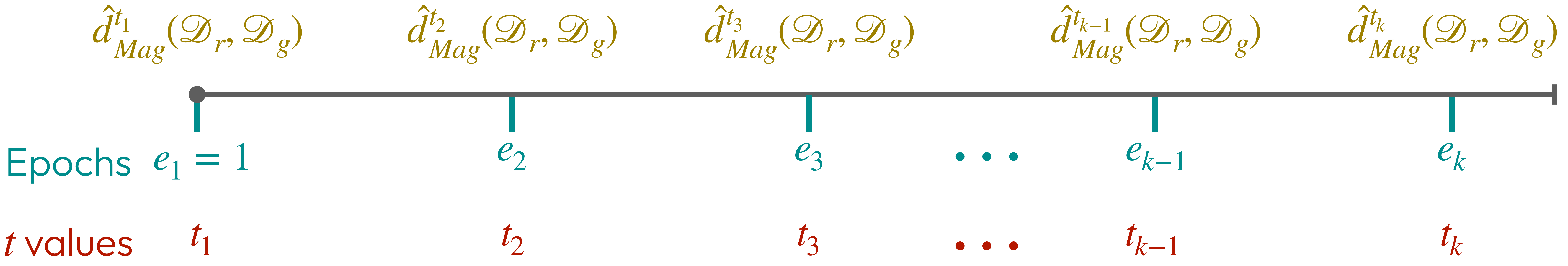
Definition

Definition: For some $k \in \mathbb{N}$ and a list called t-steps $\{e_1, \ldots, e_k\}$, at epochs e_i we add $\hat{d}^{t_i}_{Mag}(\mathscr{D}_r, \mathscr{D}_g)$ to the loss, respectively, with the magnitude scaling parameters t_i where $t_1 \leq \ldots \leq t_k$.

$$l_e = \sum_{i:e_i \leq e} \hat{d}^{t_i}_{Mag}(\mathscr{D}_r, \mathscr{D}_g).$$

real data

generated data

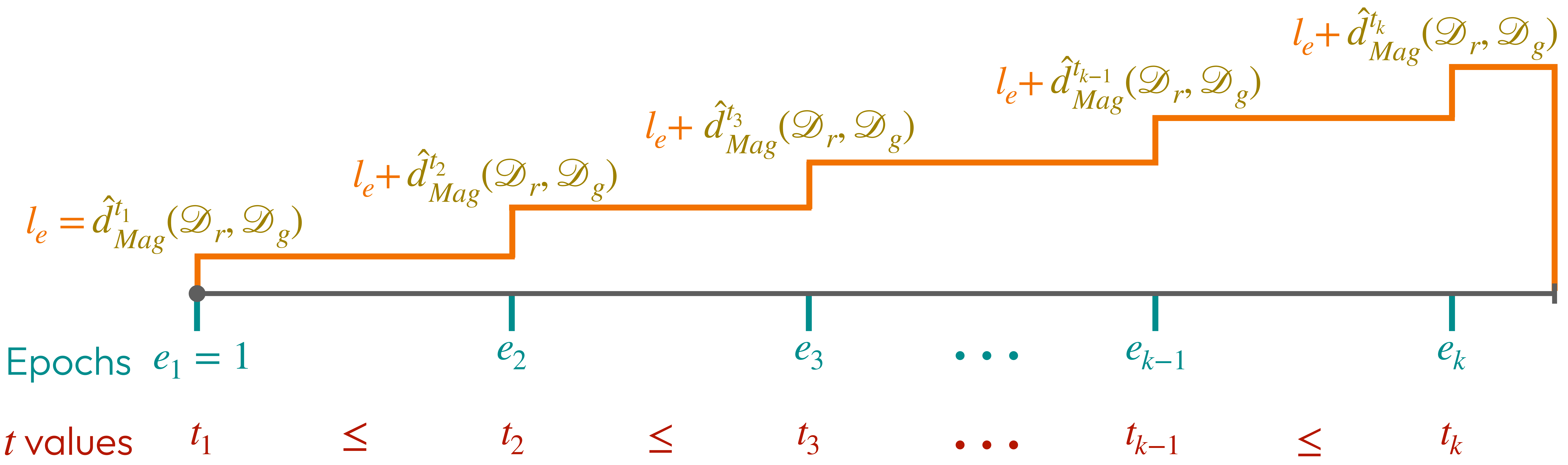


Magnitude Generative Network (MagGN)

Definition

Definition: For some $k \in \mathbb{N}$ and a list called t-steps $\{e_1, \ldots, e_k\}$, at epochs e_i we add $\hat{d}^{t_i}_{Mag}(\mathcal{D}_r, \mathcal{D}_g)$ to the loss, respectively, with the magnitude scaling parameters t_i where $t_1 \leq \ldots \leq t_k$.

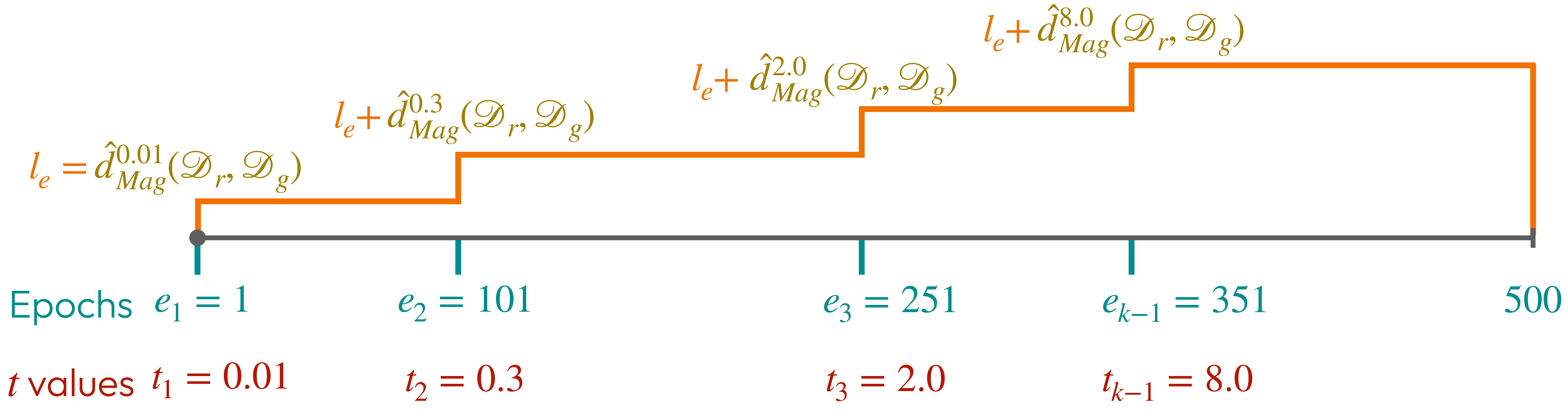
$$l_e = \sum_{i:e_i \leq e} \hat{d}^{t_i}_{Mag}(\mathcal{D}_r, \mathcal{D}_g).$$



Magnitude Generative Network (MagGN)

Performance

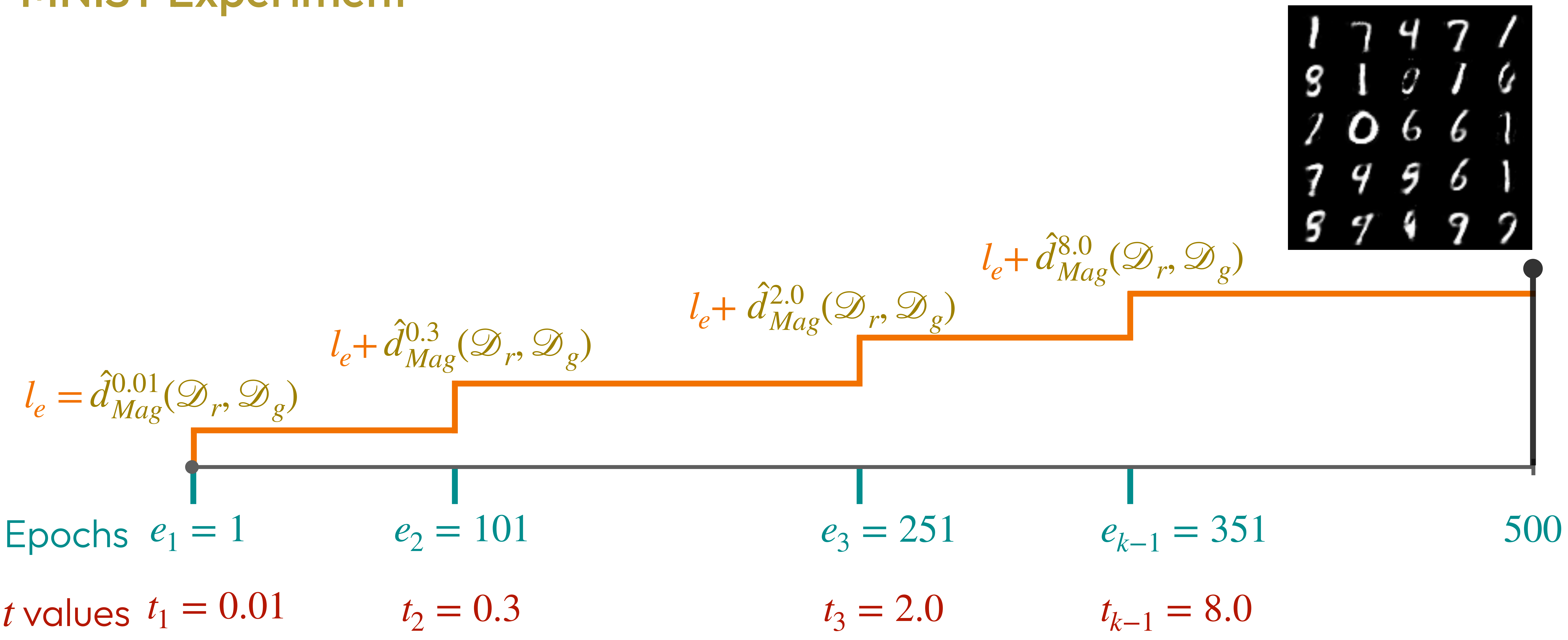
MNIST Experiment



Magnitude Generative Network (MagGN)

Performance

MNIST Experiment

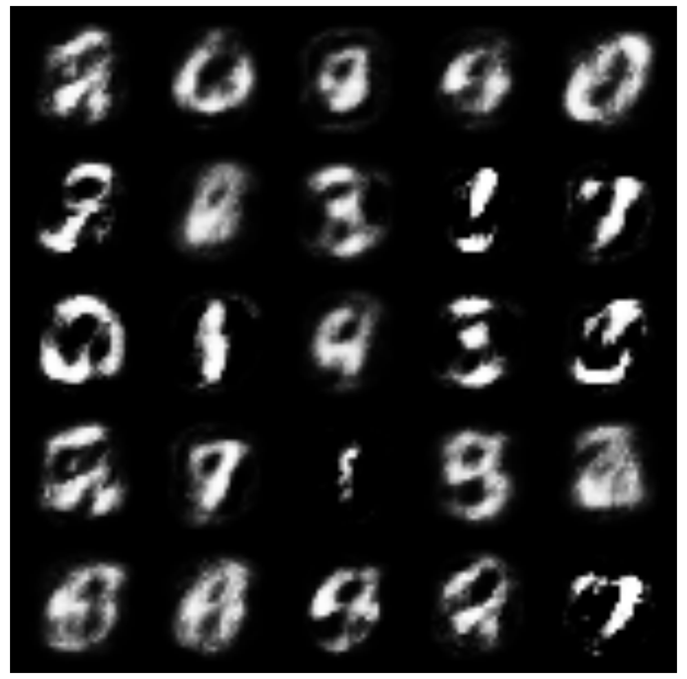


Magnitude Generative Network (MagGN)

Performance

MNIST Experiment

WGAN



WGAN-GP

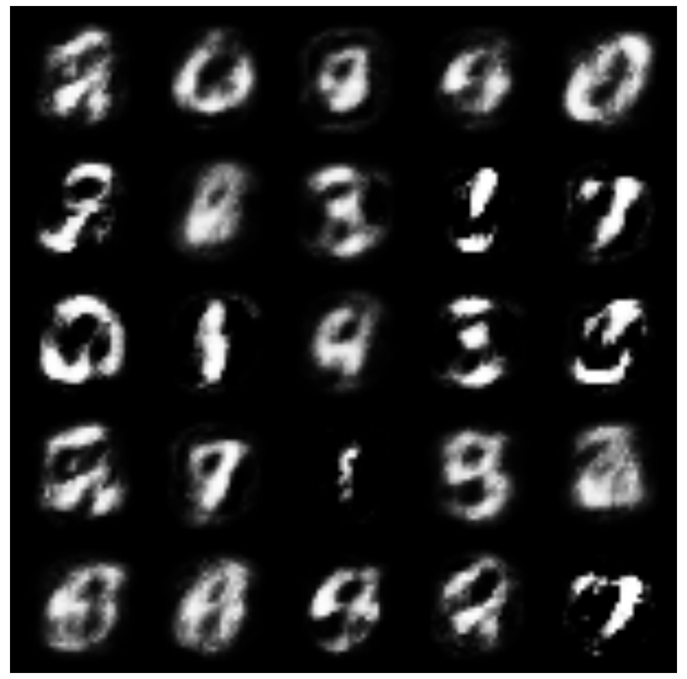
MagGN

Magnitude Generative Network (MagGN)

Performance

MNIST Experiment

WGAN



WGAN-GP

MagGN

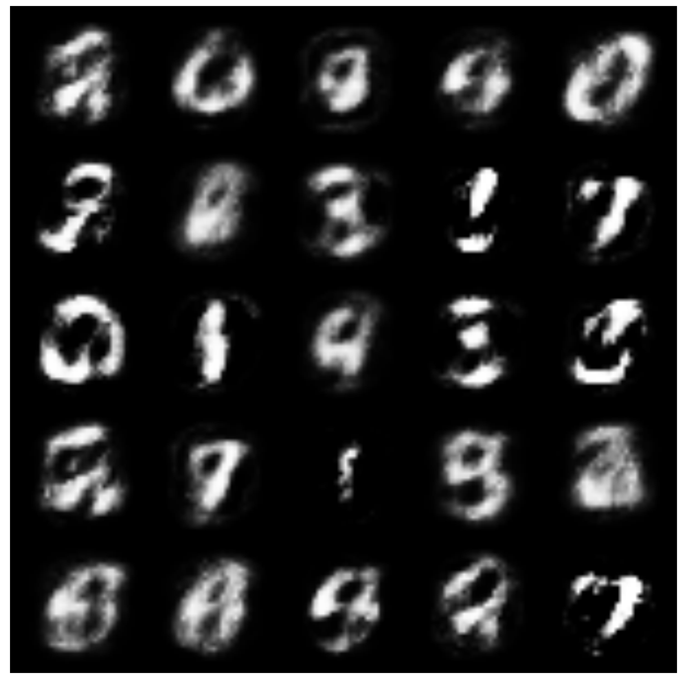
Observation:

Visual performance is better than **WGAN** [Arjovsky et al., 2017] and comparable with WGAN-GP [Gulrajani et al., 2017].

Magnitude Generative Network (MagGN)

Performance

MNIST Experiment

WGAN	WGAN-GP	MagGN
		

Training time	66064 sec	77520 sec	5440 sec

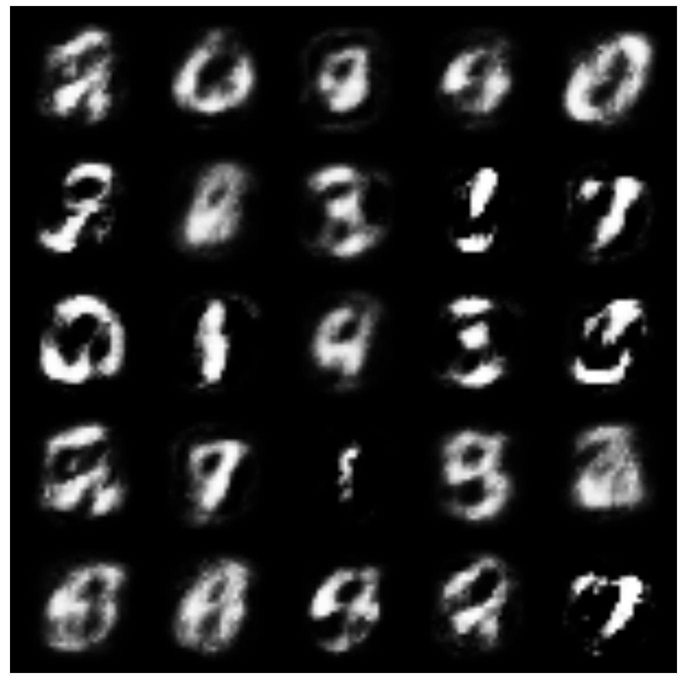
Observation:
Visual performance is better than **WGAN** [Arjovsky et al., 2017] and comparable with WGAN-GP [Gulrajani et al., 2017].

Magnitude Generative Network (MagGN)

Performance

MNIST Experiment

WGAN



WGAN-GP

MagGN

Training time	66064 sec	77520 sec	5440 sec

x12 Faster!

Observation:
Visual performance is better than **WGAN** [Arjovsky et al., 2017] and comparable with WGAN-GP [Gulrajani et al., 2017].

Summary

A new Measure for Comparing Datasets

- Magnitude distance is **tunable** via scaling parameter: from global structure to fine-grained local detail.
- Remains **discriminative in high dimensions**, where classical distances collapse.
- As a proof of concept, **MagGN** achieves >12× training speedup with no loss in visual quality.
- We expect magnitude distance to be used in various machine learning contexts as a **fundamental tool**.