International Conference on Machine Learning (ICML) 2021

Poster, Paper & Code

bit.ly/3gu32tw

Contrastive Learning Inverts the Data Generating Process

Roland S. Zimmermann*, Yash Sharma*, Steffen Schneider*,
Matthias Bethge†, Wieland Brendel†





EBERHARD KARLS UNIVERSITÄT TÜBINGEN

imprs-is

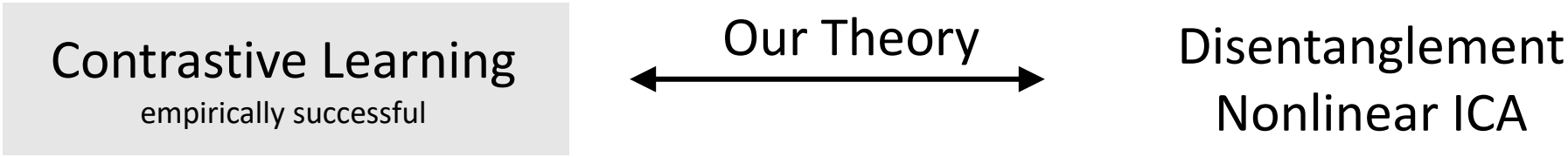
Bundesministerium für Bildung und Forschung

ROBUST VISION

DFG

* Equal contribution † Joint supervision


Contrastive Learning
Our Theory
Disentanglement
Nonlinear ICA


Contrastive Learning
empirically successful
Our Theory
Disentanglement
Nonlinear ICA

Contrastive Learning

Our Theory

Disentanglement
Nonlinear ICA

Image Space

Representation Space

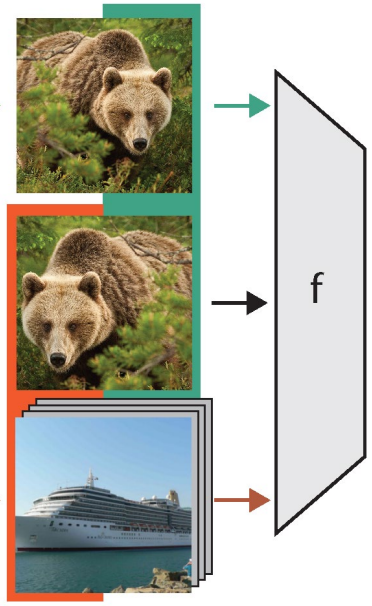
Contrastive Learning

Our Theory

Disentanglement
Nonlinear ICA

Image Space

Representation Space



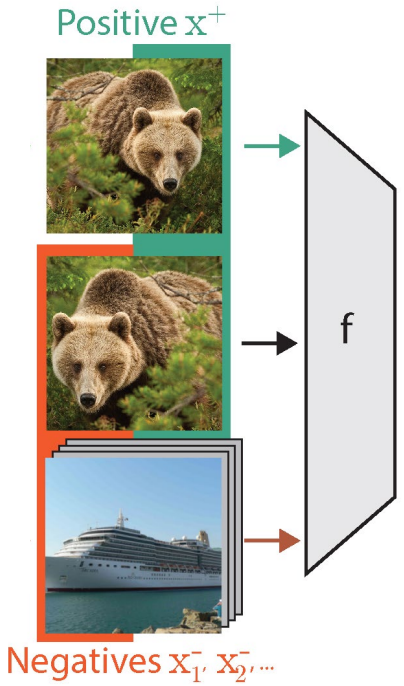
Contrastive Learning

Our Theory

Disentanglement
Nonlinear ICA

Image Space

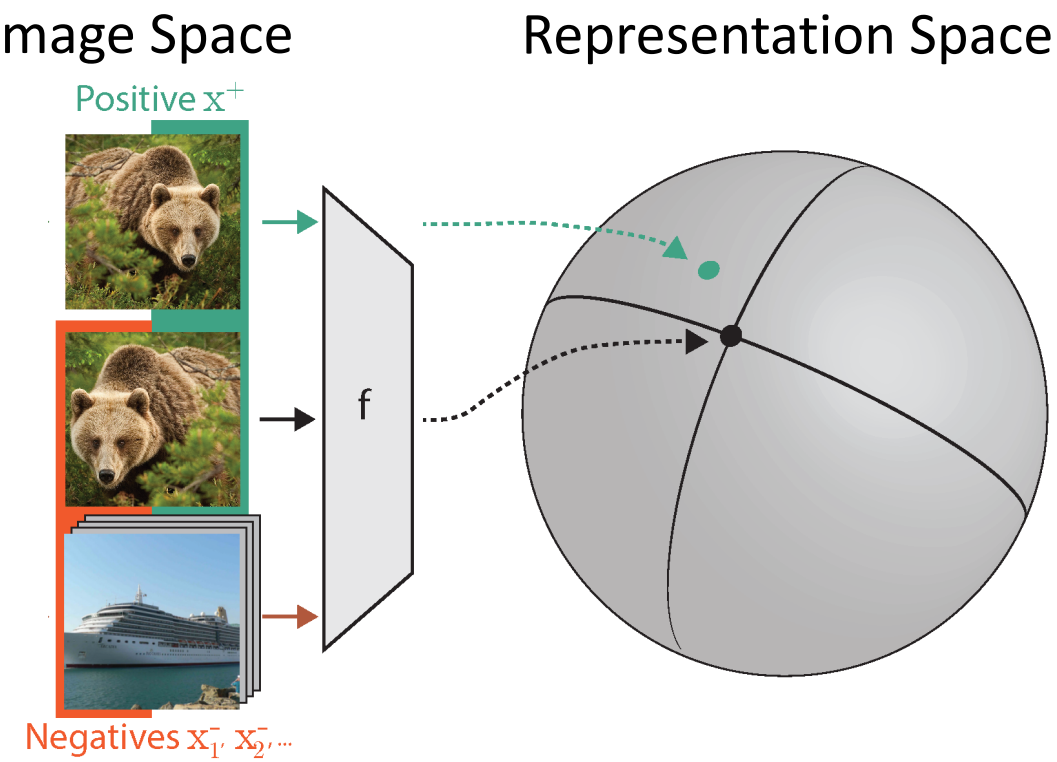
Representation Space



Contrastive Learning

Our Theory

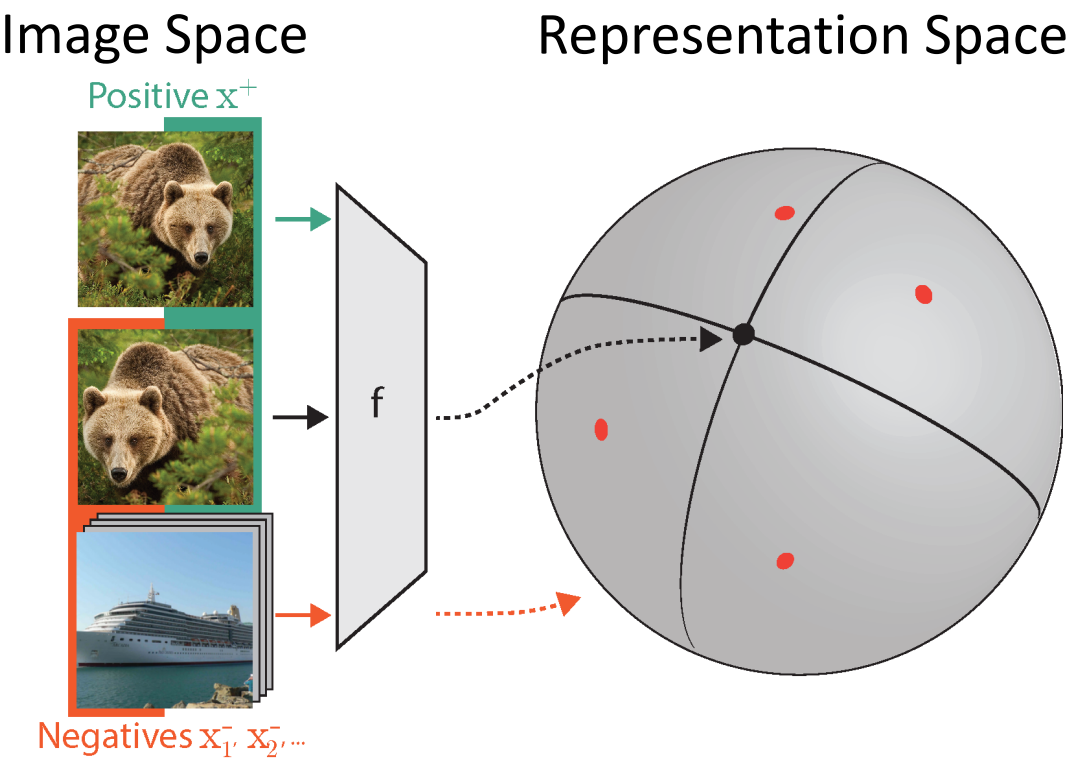
Disentanglement
Nonlinear ICA



Contrastive Learning

Our Theory

Disentanglement
Nonlinear ICA



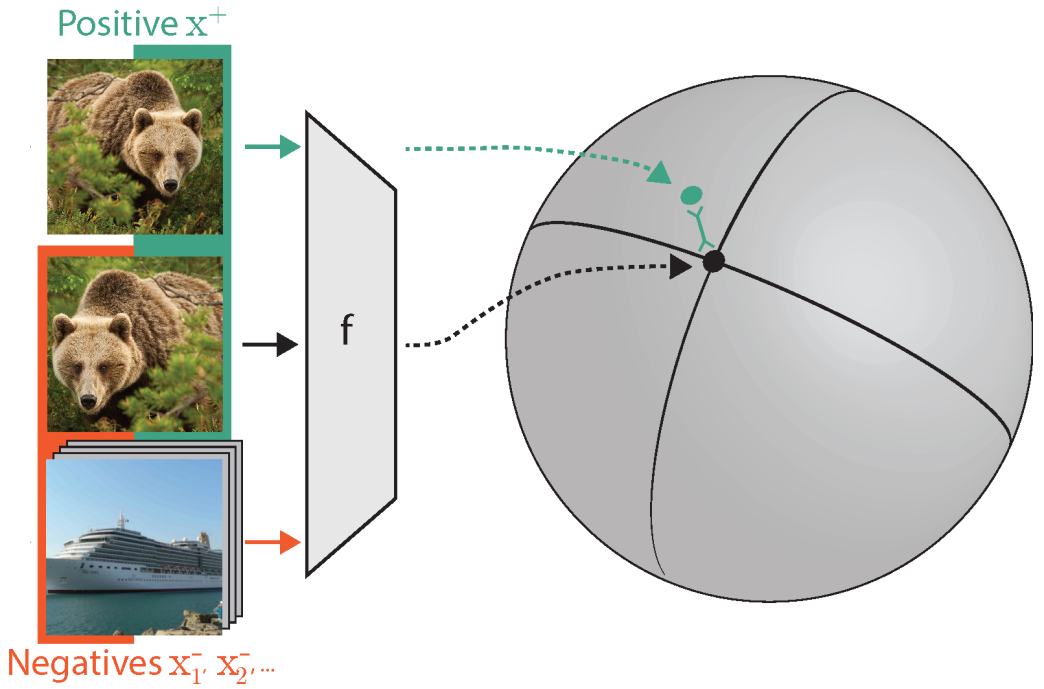
Contrastive Learning

Our Theory

Disentanglement
Nonlinear ICA

Image Space

Representation Space



Learned by

$$\mathcal{L} = -\underbrace{f(x)^\top f(x^+)}_{\text{attract}}$$

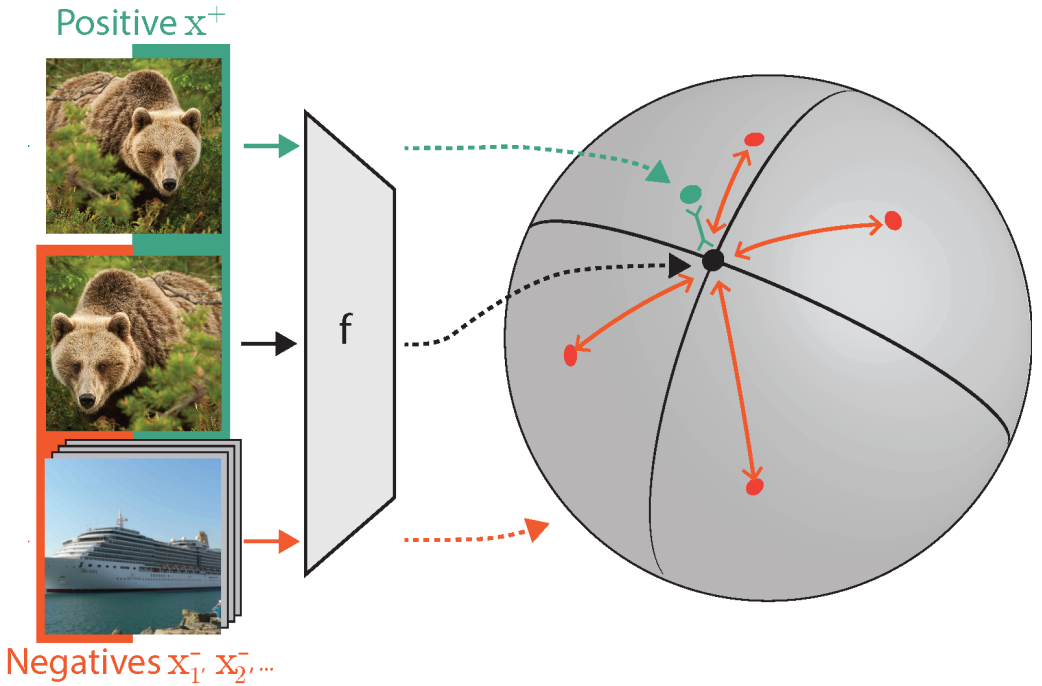
Contrastive Learning

Our Theory

Disentanglement
Nonlinear ICA

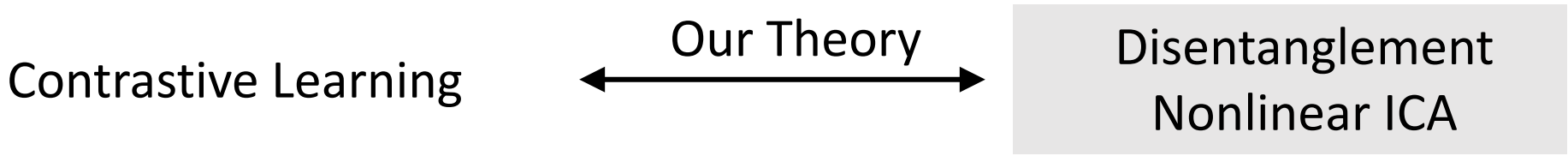
Image Space

Representation Space



Learned by

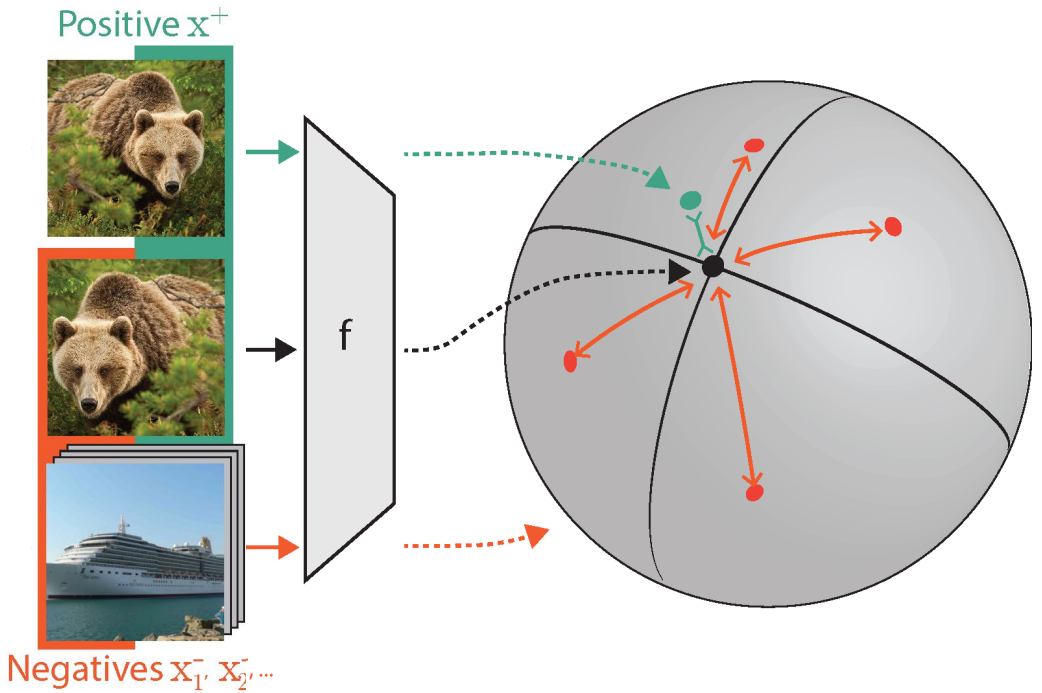
$$\mathcal{L} = \log \sum_{\{x, x^+, x_1^-, \ldots\}} \exp(\underbrace{f(x)^\top f(x')}_{\text{repel}})$$


Contrastive Learning
Our Theory
Disentanglement
Nonlinear ICA

Contrastive Learning

Our Theory

Disentanglement
Nonlinear ICA



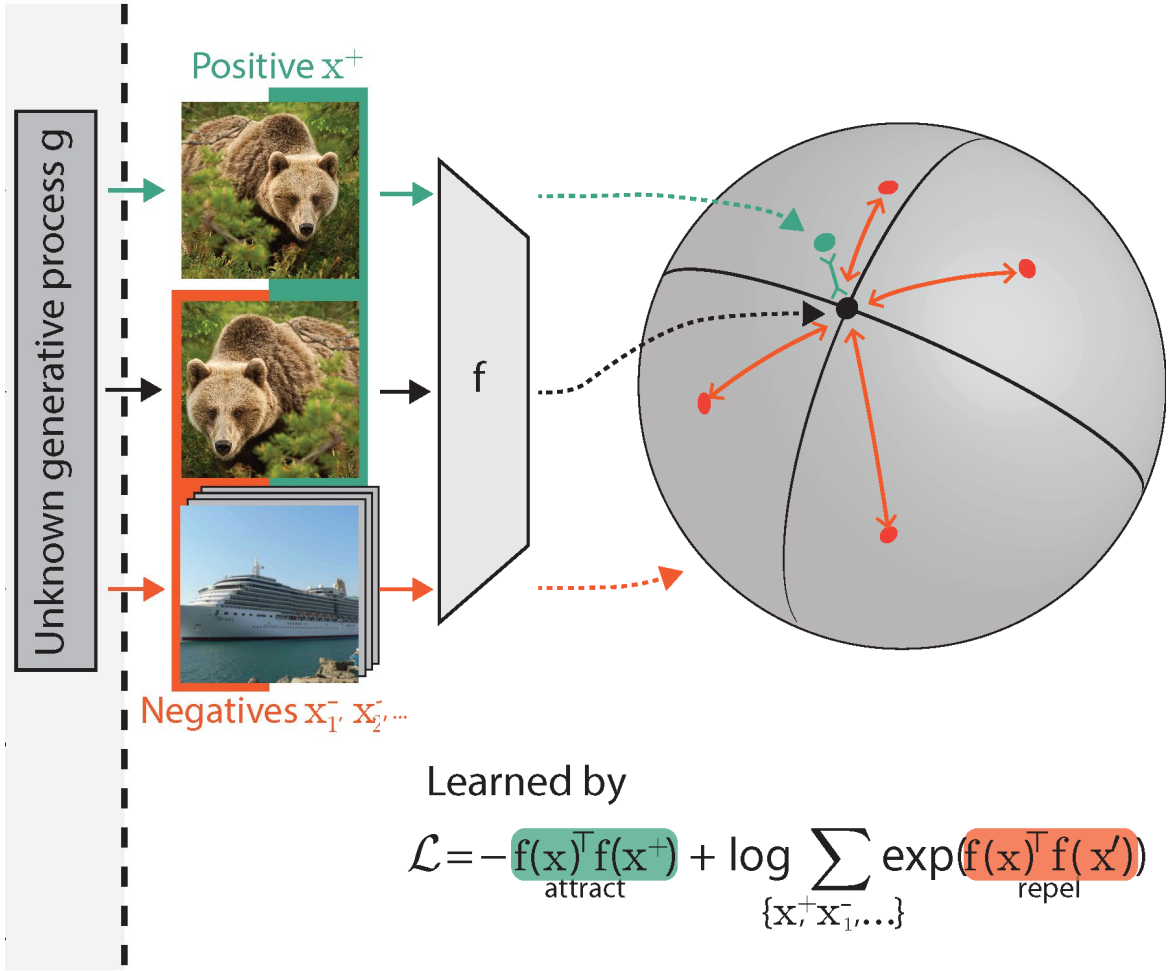
Learned by

$$\mathcal{L} = -\underbrace{f(x)^\top f(x^+)}_{\text{attract}} + \log \sum_{\{x^+, x_1^-, \ldots\}} \exp(\underbrace{f(x)^\top f(x')}_{\text{repel}})$$

Contrastive Learning

Our Theory

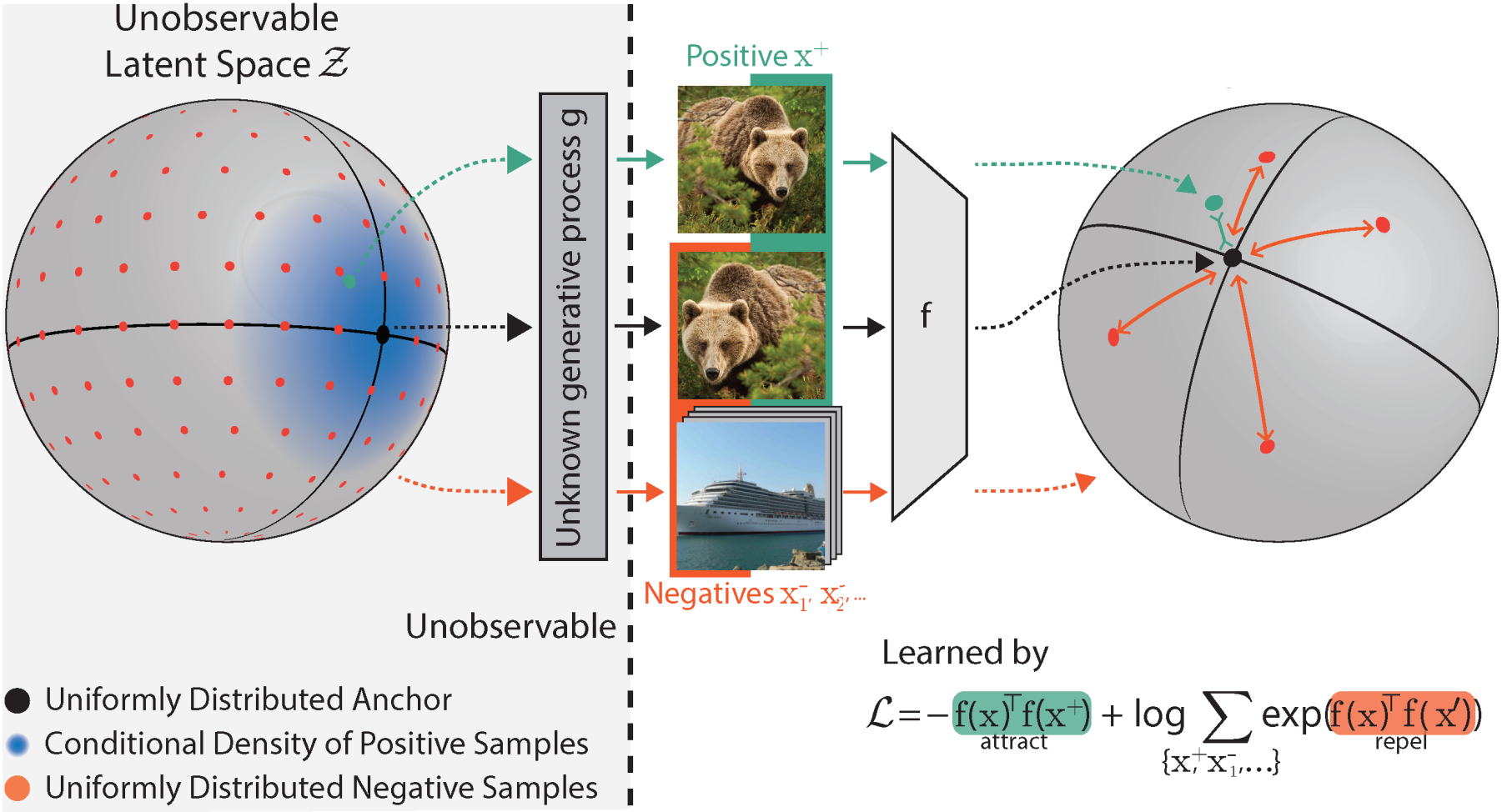
Disentanglement
Nonlinear ICA



Contrastive Learning

Our Theory

Disentanglement
Nonlinear ICA



Learned by

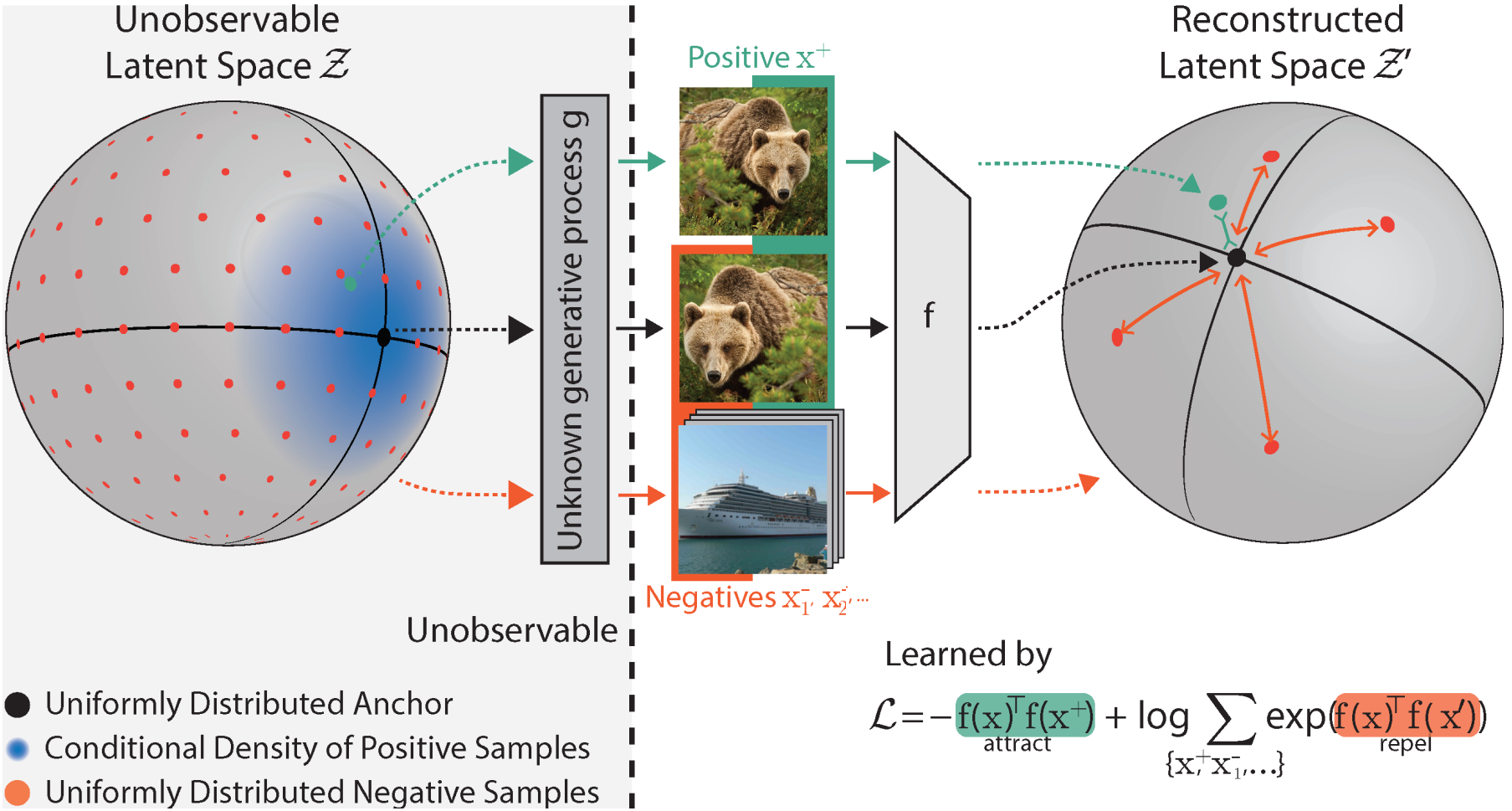
$$\mathcal{L} = -\underbrace{f(x)^\top f(x^+)}_{\text{attract}} + \log \sum_{\{x^+, x_1^-, \ldots\}} \exp(\underbrace{f(x)^\top f(x')}_{\text{repel}})$$

Contrastive Learning

Our Theory

Disentanglement
Nonlinear ICA

Goal: Recover structure

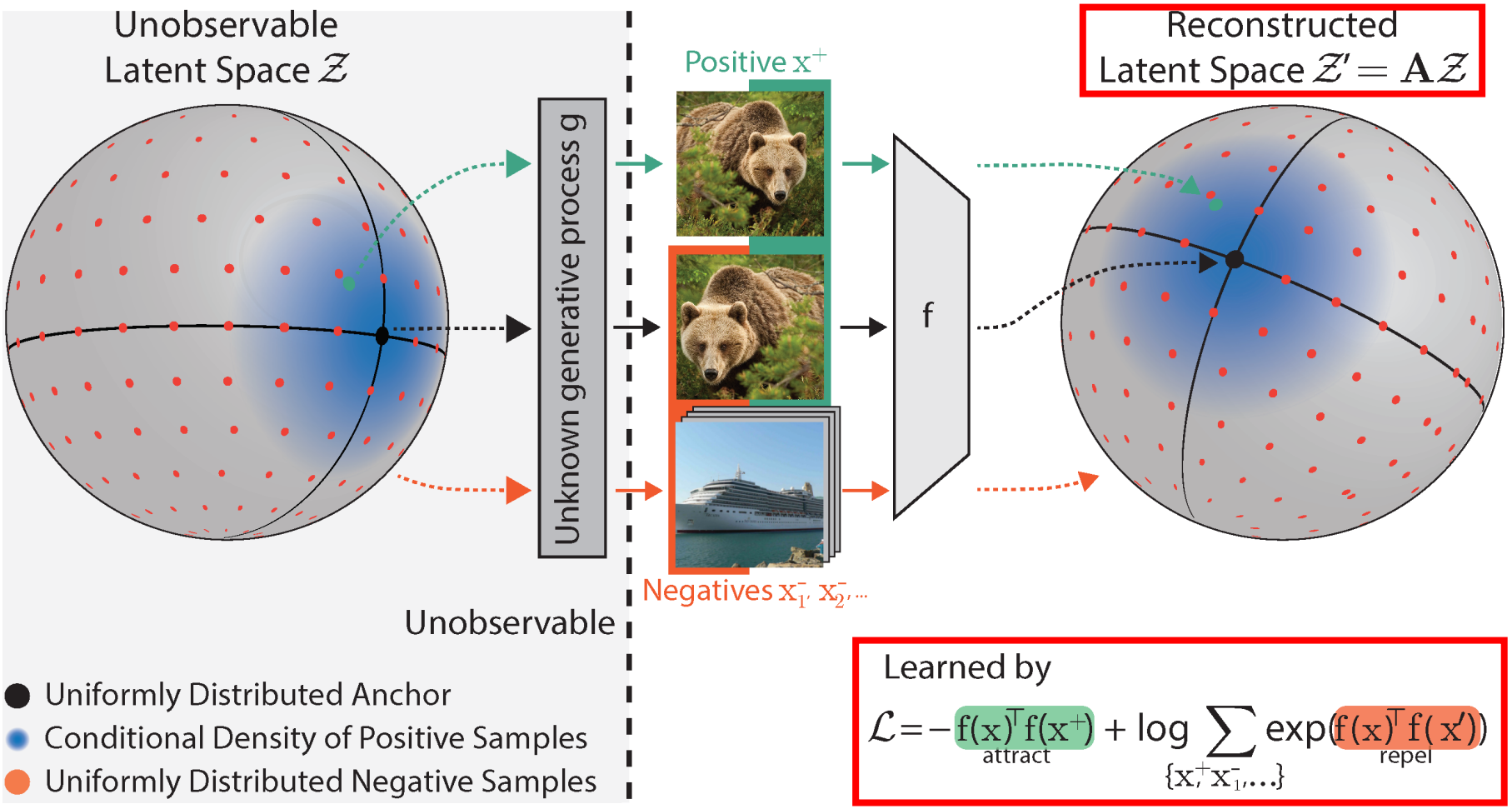


Learned by

$$\mathcal{L} = -\underbrace{f(x)^\top f(x^+)}_{\text{attract}} + \log \sum_{\{x^+, x_1^-, \ldots\}} \exp(\underbrace{f(x)^\top f(x')}_{\text{repel}})$$

We prove

Contrastive Learning Inverts Data Generative Process



We prove

Contrastive Learning Inverts Data Generative Process

Learned by

$$\mathcal{L} = -\underbrace{f(x)^\top f(x^+)}_{\text{attract}} + \log \sum_{\{x^+, x_1^-, \ldots\}} \exp(\underbrace{f(x)^\top f(x')}_{\text{repel}})$$

Assumptions

+ von Mises-Fisher distribution

We prove

Contrastive Learning Inverts Data Generative Process

Learned by

$$\mathcal{L} = -\underbrace{f(x)^\top f(x^+)}_{\text{attract}} + \log \sum_{\{x^+, x_1^-, \ldots\}} \exp(\underbrace{f(x)^\top f(x')}_{\text{repel}})$$

Assumptions implicitly encoded

+ von Mises-Fisher distribution

We prove

Contrastive Learning Inverts Data Generative Process

Learned by

$$\mathcal{L} = \underbrace{-\mathrm{f}(\mathrm{x})^\top \mathrm{f}(\mathrm{x}^+)}_{\text{attract}} + \log \sum_{\{\mathrm{x}^+, \mathrm{x}_1^-, \ldots\}} \exp(\underbrace{\mathrm{f}(\mathrm{x})^\top \mathrm{f}(\mathrm{x}')}_{\text{repel}})$$

Learned by

$$\mathcal{L} = \underbrace{\|\mathrm{f}(\mathrm{x}) - \mathrm{f}(\mathrm{x}^+)\|_2^2}_{\text{attract}} + \log \sum_{\{\mathrm{x}^+, \mathrm{x}_1^-, \ldots\}} \exp(\underbrace{-\|\mathrm{f}(\mathrm{x}) - \mathrm{f}(\mathrm{x}')\|_2^2}_{\text{repel}})$$

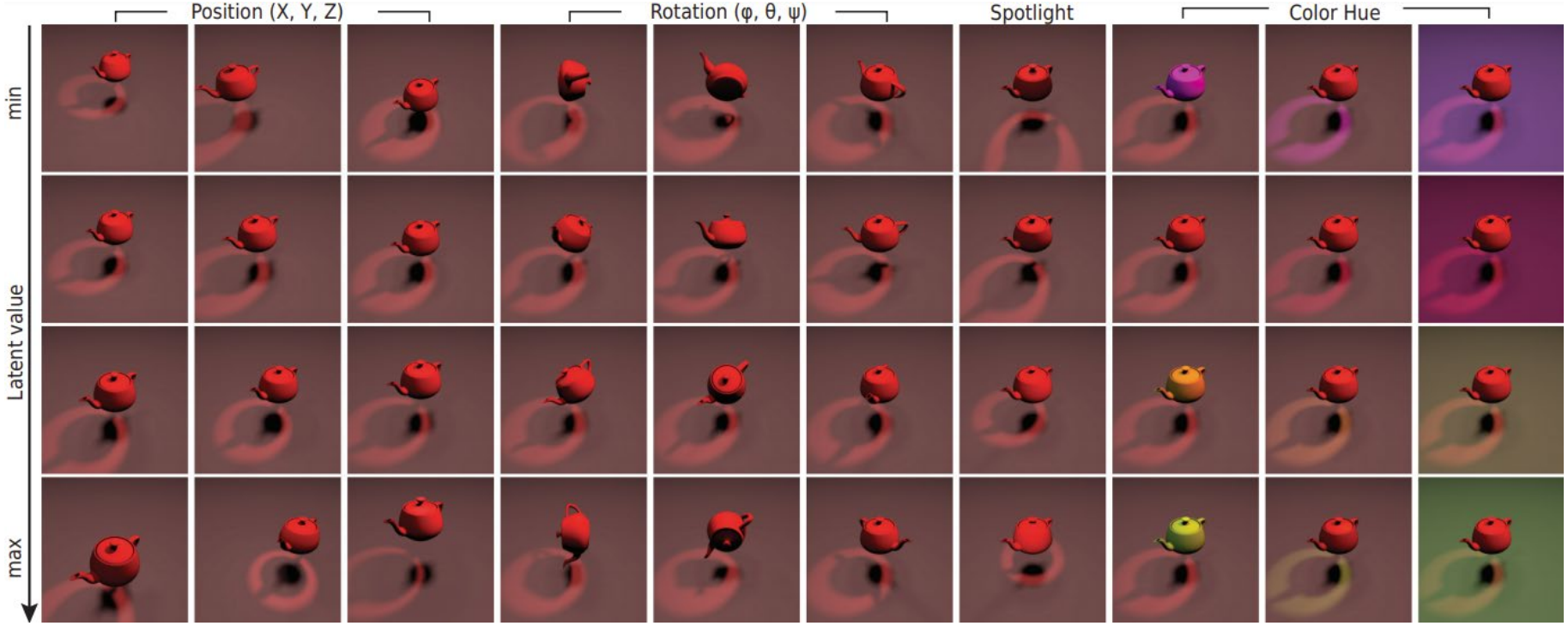
Assumptions encoded

\+ von Mises-Fisher distribution

\+ Normal distribution

We show

Successful Disentanglement on Visually Complex Images



3DIdent Dataset:

- High resolution
- Hallmarks of natural environments (shadows, lighting conditions, 3D object)

Summary

Prove that contrastive learning (InfoNCE) inverts specified data generating processes

Works on complex visual data and is robust to mismatches

Outlook

Theory provides insights in constructing more effective contrastive losses

Poster, Paper & Code

bit.ly/3r4CylX