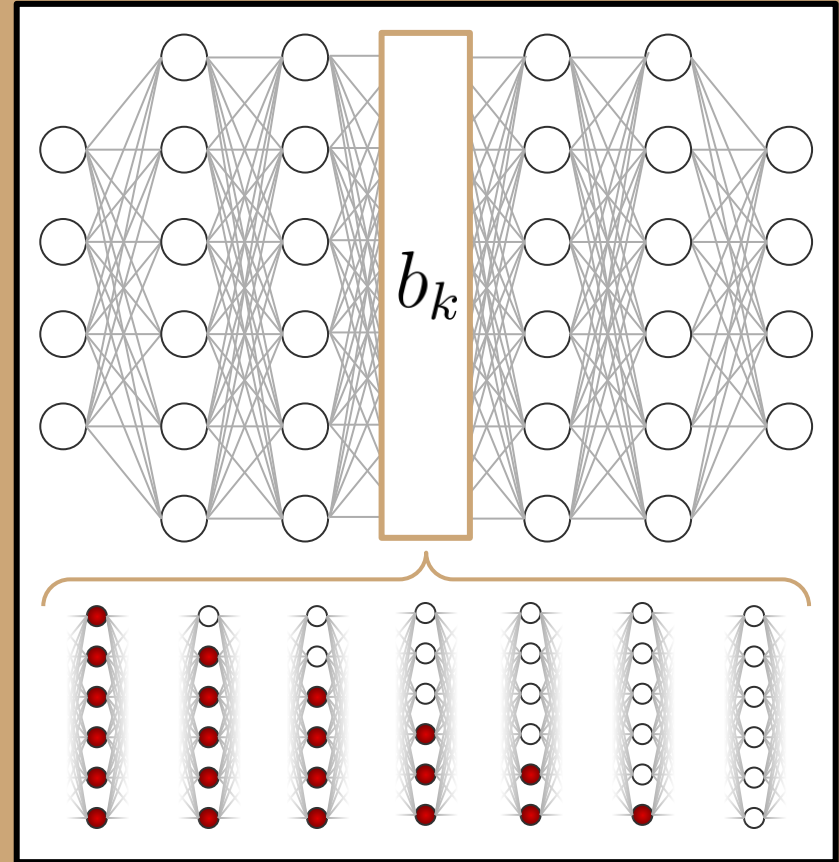
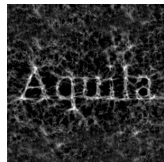
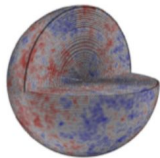


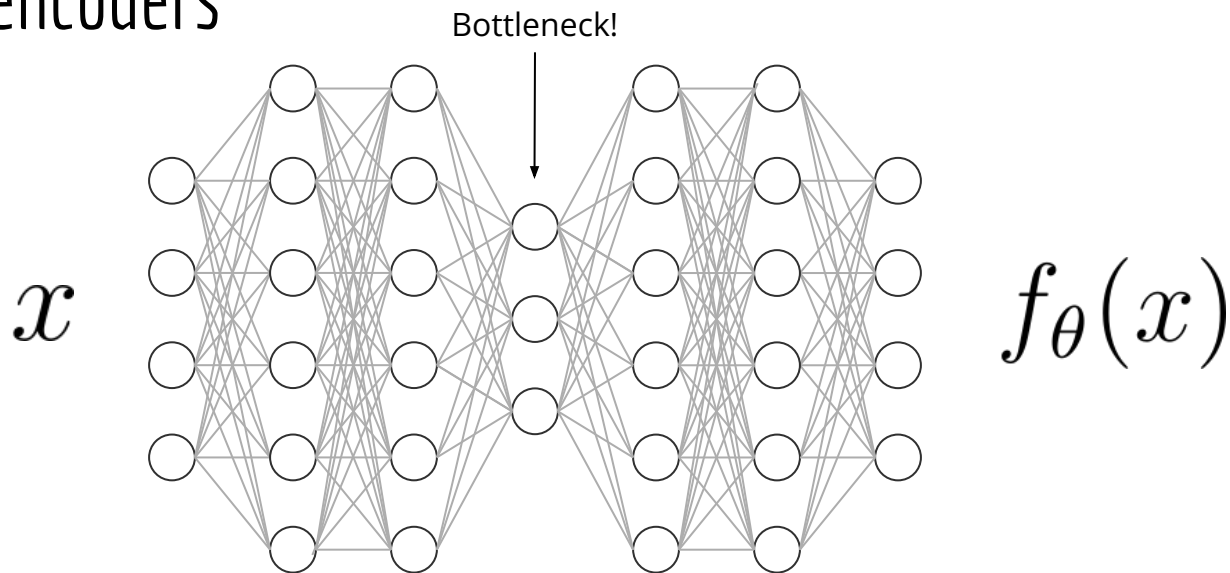
Scientific Discovery from Ordered Information Decomposition

Matthew Ho

Postdoc @ IAP
matthew.annam.ho@gmail.com



Normal Autoencoders



“Is there truth
in latent
space?”

$$\mathcal{L} := \|x - f_{\theta}(x)\|_2^2$$

The choice of bottleneck width affects the embedding!

(Bahadur+2019, arxiv:1909.10702)

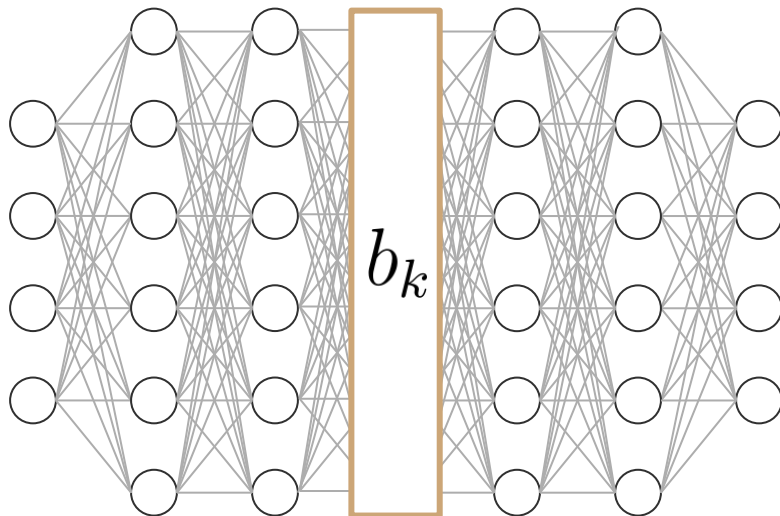
Information-Ordered Bottlenecks (IOBs)

Ho, Zhao, &
Wandelt 2023,
arXiv:2305.11213

Vary the bottleneck
during training!

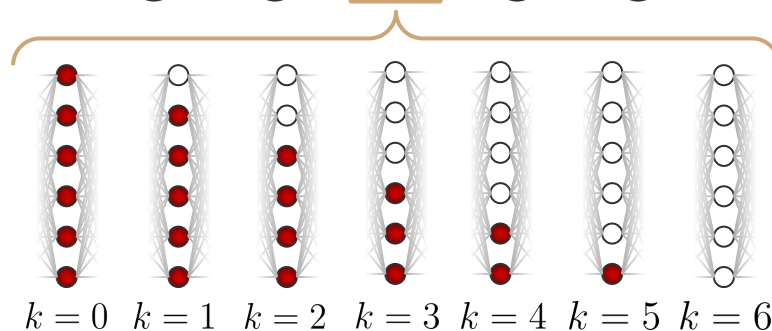
x

Force most
information through
the top nodes.



$f_{\theta}(x; k)$

Equal to PCA in the
linear case.



No disentanglement.

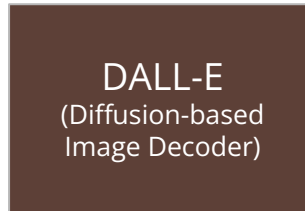
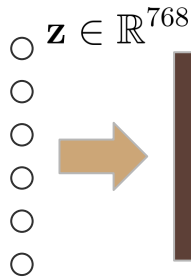
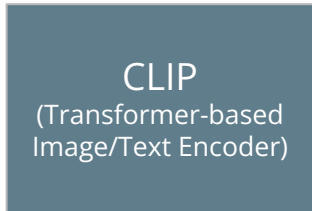
$$\mathcal{L} := \sum_{k=0}^{k_{\max}} \mathcal{L}_k$$

Applications:

- **Explainability** (LLMs and Diffusion Models)
- **Intrinsic Dimensionality Estimation** (CNNs)
- **Manifold Fitting** (CAMELs Galaxy Properties)
- **Accelerated Symbolic Regression** (CMB Power Spectra)
- **Complexity Ranking** (Black Hole Binary Light Curves)

Explainability

Andre



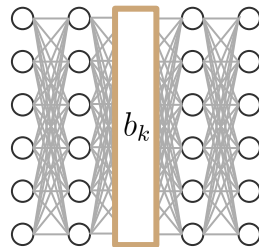
Ho, Zhao, &
Wandelt 2023,
arXiv:2305.11213



“What does my
model think is
important?”

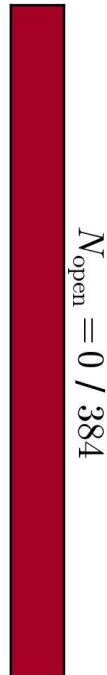
Explainability

CLIP
(Transformer-based
Image/Text Encoder)



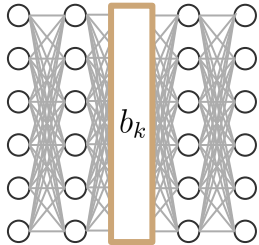
DALL-E
(Diffusion-based
Image Decoder)

Ho, Zhao, &
Wandelt 2023,
arXiv:2305.11213



Explainability

CLIP
(Transformer-based
Image/Text Encoder)



DALL-E
(Diffusion-based
Image Decoder)

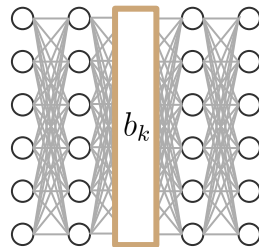
Ho, Zhao, &
Wandelt 2023,
arXiv:2305.11213



$N_{\text{open}} = 1 / 384$

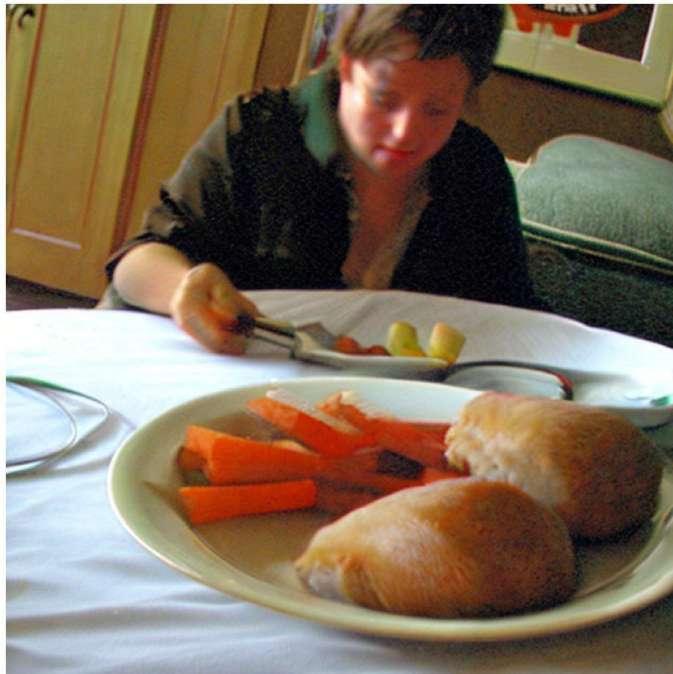
Explainability

CLIP
(Transformer-based
Image/Text Encoder)



DALL-E
(Diffusion-based
Image Decoder)

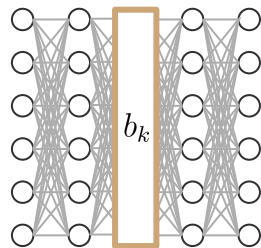
Ho, Zhao, &
Wandelt 2023,
arXiv:2305.11213



$N_{\text{open}} = 2 / 384$

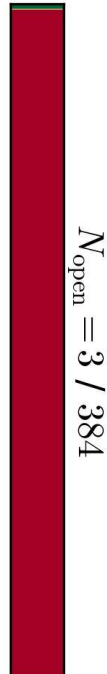
Explainability

CLIP
(Transformer-based
Image/Text Encoder)



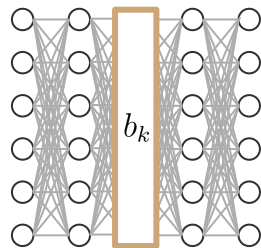
DALL-E
(Diffusion-based
Image Decoder)

Ho, Zhao, &
Wandelt 2023,
arXiv:2305.11213



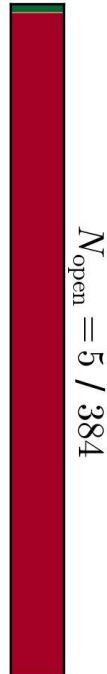
Explainability

CLIP
(Transformer-based
Image/Text Encoder)



DALL-E
(Diffusion-based
Image Decoder)

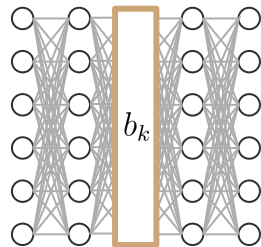
Ho, Zhao, &
Wandelt 2023,
arXiv:2305.11213



Objects

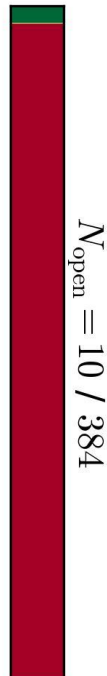
Explainability

CLIP
(Transformer-based
Image/Text Encoder)



DALL-E
(Diffusion-based
Image Decoder)

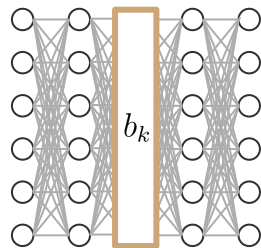
Ho, Zhao, &
Wandelt 2023,
arXiv:2305.11213



Objects

Explainability

CLIP
(Transformer-based
Image/Text Encoder)



DALL-E
(Diffusion-based
Image Decoder)

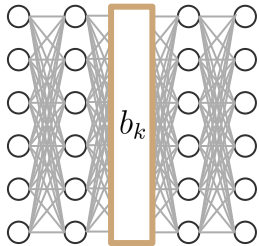
Ho, Zhao, &
Wandelt 2023,
arXiv:2305.11213



Objects
Physicality

Explainability

CLIP
(Transformer-based
Image/Text Encoder)



DALL-E
(Diffusion-based
Image Decoder)

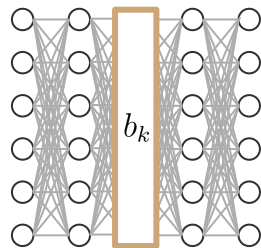
Ho, Zhao, &
Wandelt 2023,
arXiv:2305.11213



Objects
Physicality

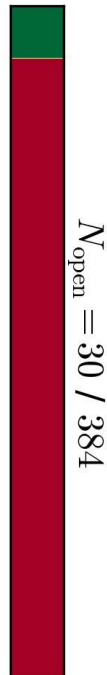
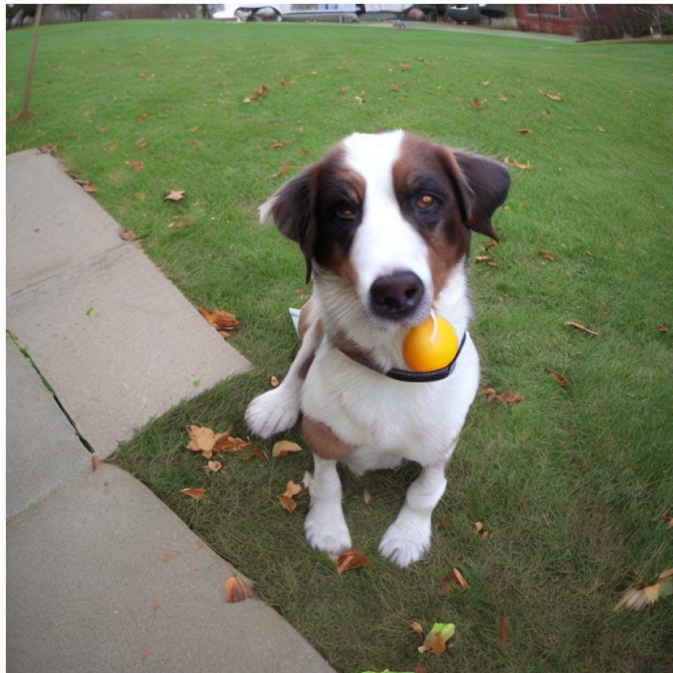
Explainability

CLIP
(Transformer-based
Image/Text Encoder)



DALL-E
(Diffusion-based
Image Decoder)

Ho, Zhao, &
Wandelt 2023,
arXiv:2305.11213



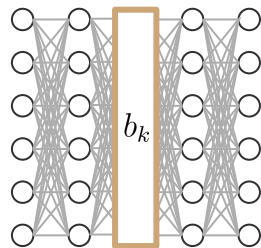
Objects

Physicality

Orientation

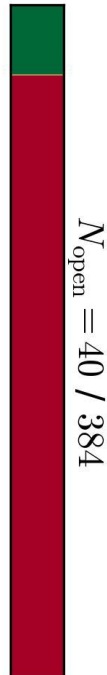
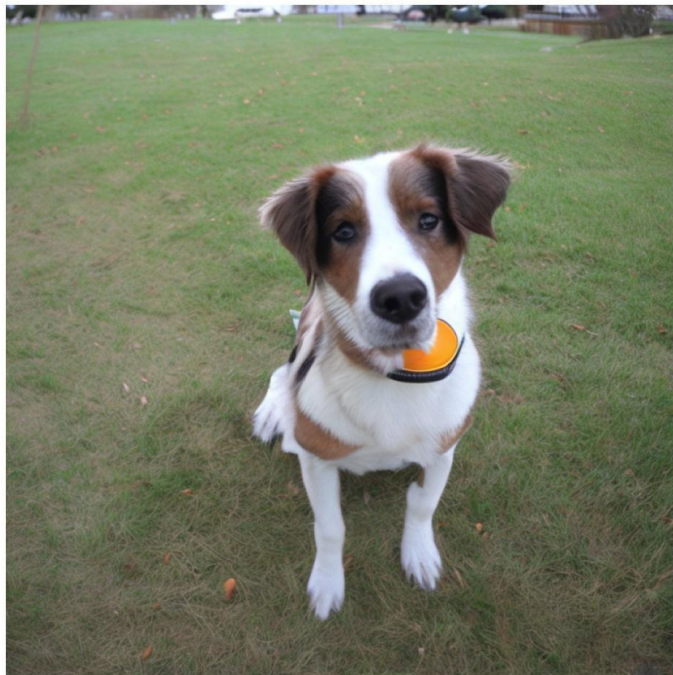
Explainability

CLIP
(Transformer-based
Image/Text Encoder)



DALL-E
(Diffusion-based
Image Decoder)

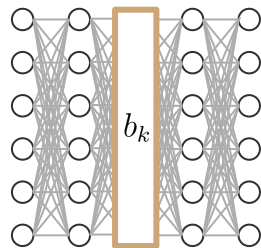
Ho, Zhao, &
Wandelt 2023,
arXiv:2305.11213



Objects
Physicality
Orientation

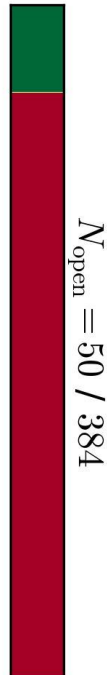
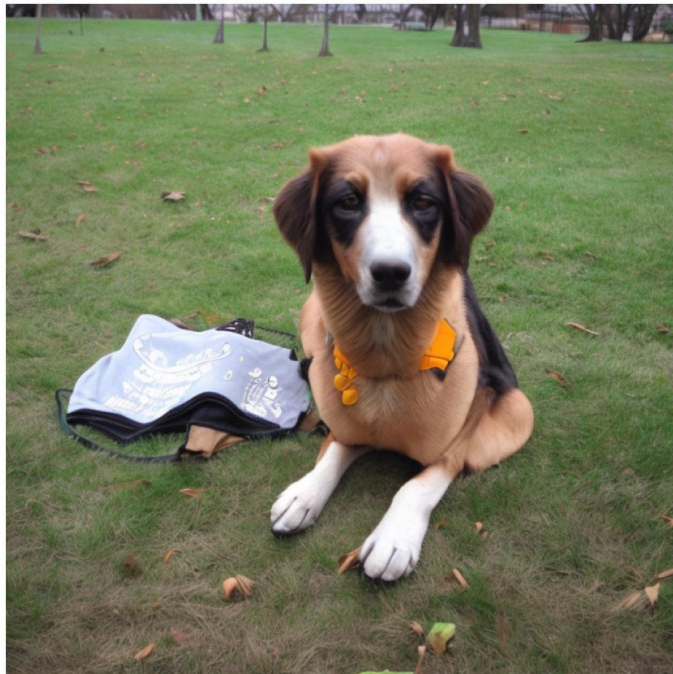
Explainability

CLIP
(Transformer-based
Image/Text Encoder)



DALL-E
(Diffusion-based
Image Decoder)

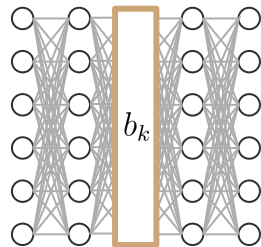
Ho, Zhao, &
Wandelt 2023,
arXiv:2305.11213



Objects
Physicality
Orientation
Colors

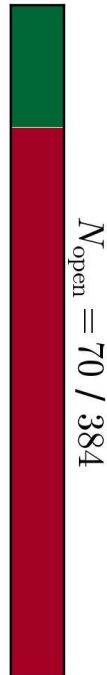
Explainability

CLIP
(Transformer-based
Image/Text Encoder)



DALL-E
(Diffusion-based
Image Decoder)

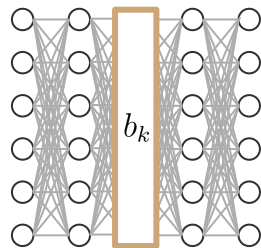
Ho, Zhao, &
Wandelt 2023,
arXiv:2305.11213



- Objects
- Physicality
- Orientation
- Colors
- Backgrounds

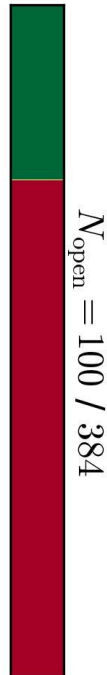
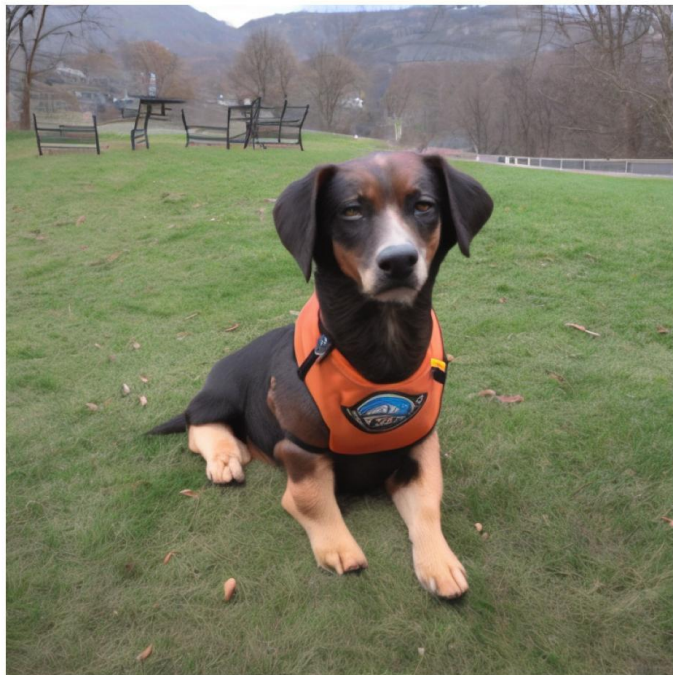
Explainability

CLIP
(Transformer-based
Image/Text Encoder)



DALL-E
(Diffusion-based
Image Decoder)

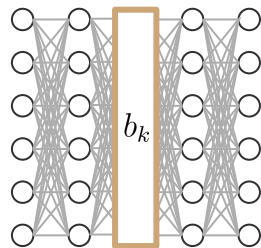
Ho, Zhao, &
Wandelt 2023,
arXiv:2305.11213



Objects
Physicality
Orientation
Colors
Backgrounds
Detail

Explainability

CLIP
(Transformer-based
Image/Text Encoder)



DALL-E
(Diffusion-based
Image Decoder)

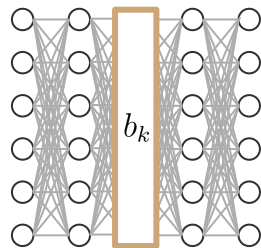
Ho, Zhao, &
Wandelt 2023,
arXiv:2305.11213



- Objects
- Physicality
- Orientation
- Colors
- Backgrounds
- Detail
- Lighting

Explainability

CLIP
(Transformer-based
Image/Text Encoder)



DALL-E
(Diffusion-based
Image Decoder)

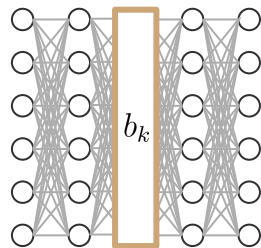
Ho, Zhao, &
Wandelt 2023,
arXiv:2305.11213



- Objects
- Physicality
- Orientation
- Colors
- Backgrounds
- Detail
- Lighting

Explainability

CLIP
(Transformer-based
Image/Text Encoder)



DALL-E
(Diffusion-based
Image Decoder)

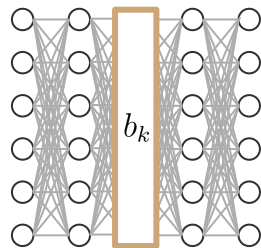
Ho, Zhao, &
Wandelt 2023,
arXiv:2305.11213



- Objects
- Physicality
- Orientation
- Colors
- Backgrounds
- Detail
- Lighting

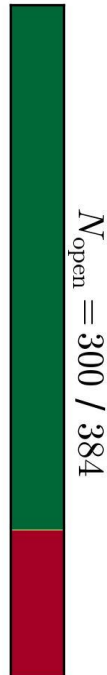
Explainability

CLIP
(Transformer-based
Image/Text Encoder)



DALL-E
(Diffusion-based
Image Decoder)

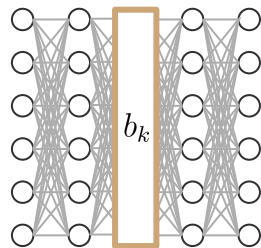
Ho, Zhao, &
Wandelt 2023,
arXiv:2305.11213



- Objects
- Physicality
- Orientation
- Colors
- Backgrounds
- Detail
- Lighting

Explainability

CLIP
(Transformer-based
Image/Text Encoder)



DALL-E
(Diffusion-based
Image Decoder)

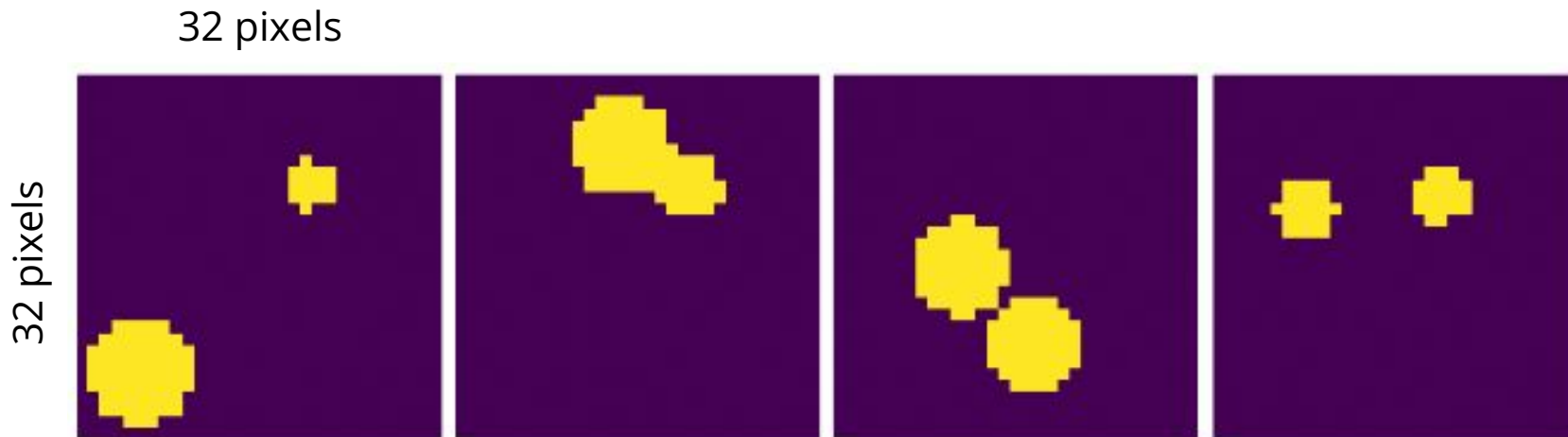
Ho, Zhao, &
Wandelt 2023,
arXiv:2305.11213



$N_{\text{open}} = 384 / 384$

- Objects
- Physicality
- Orientation
- Colors
- Backgrounds
- Detail
- Lighting

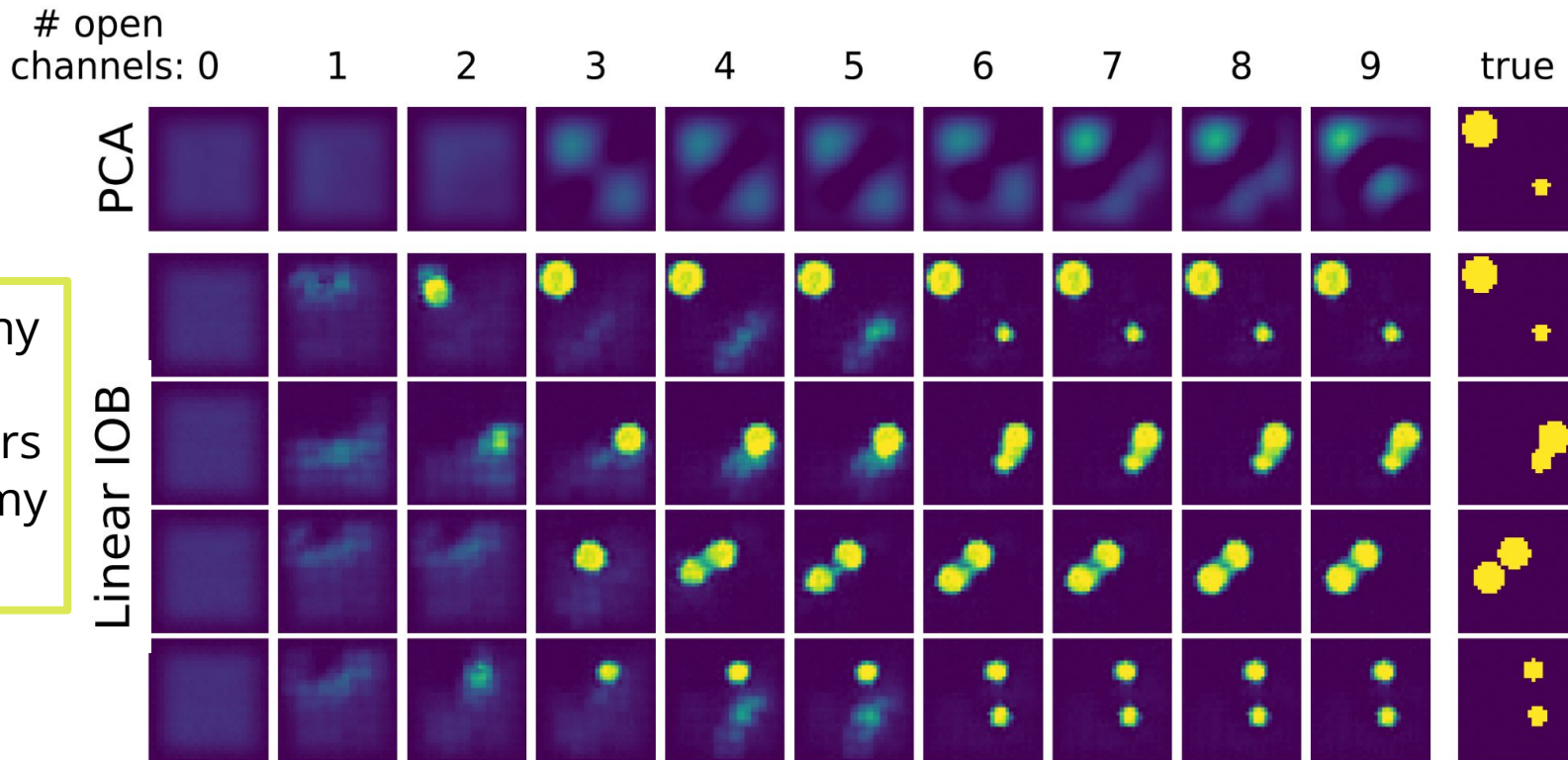
2-Disk Images



3 properties (x, y, R) \times 2 disks = 6 parameters

Intrinsic Dimensionality Estimation

Ho, Zhao, &
Wandelt 2023,
arXiv:2305.11213

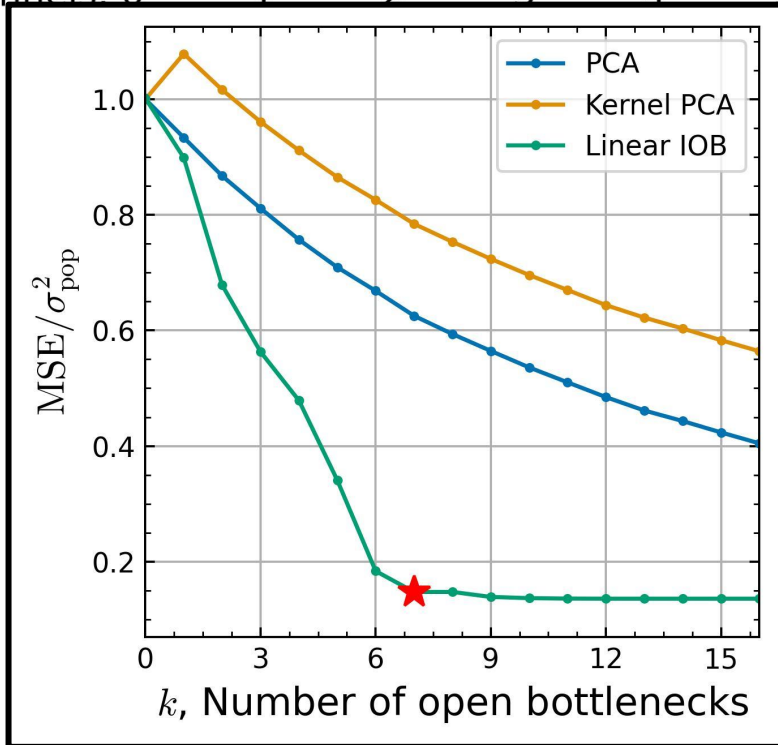


“How many physical parameters describe my data?”

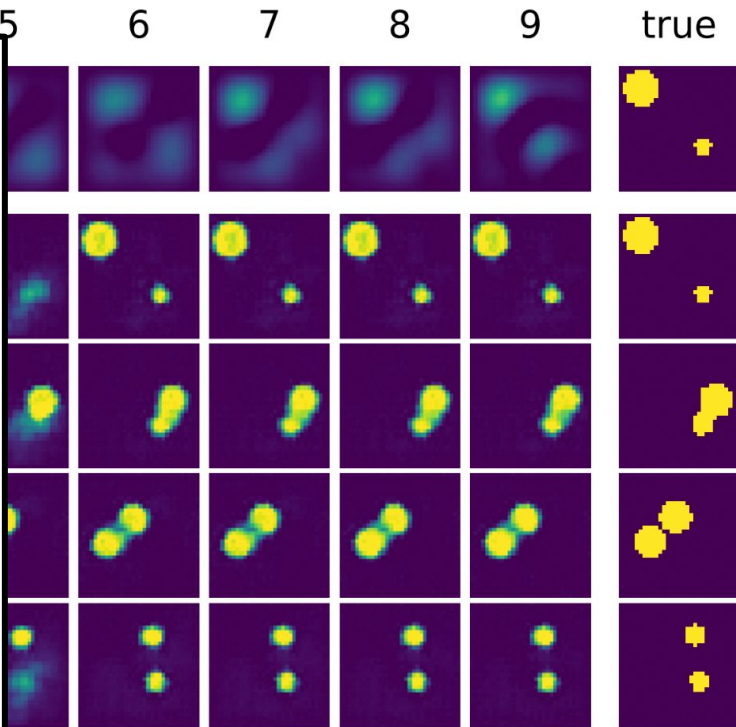
Intrinsic Dimensionality Estimation

Ho, Zhao, &
Wandelt 2023,
arXiv:2305.11213

open
channels: 0



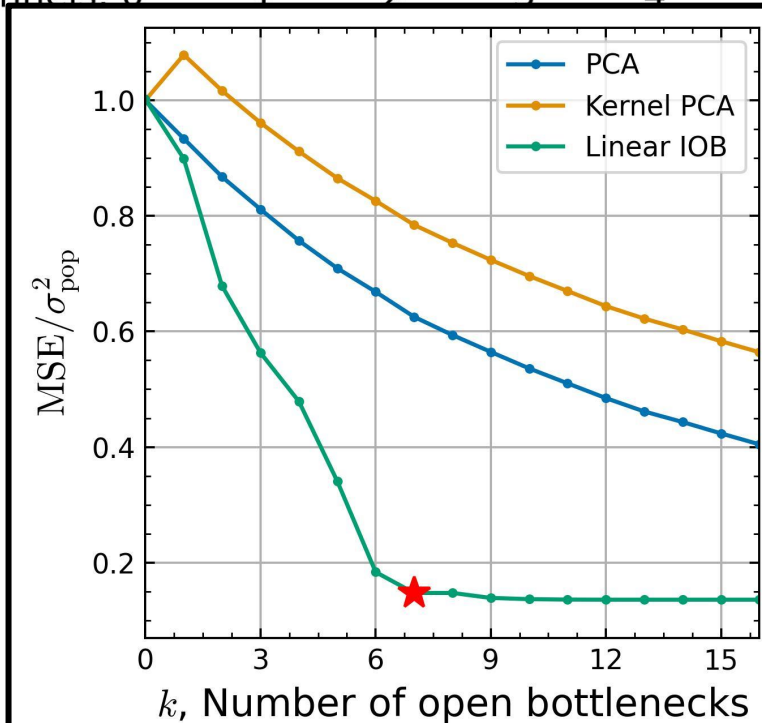
“How many
physical
parameters
describe my
data?”



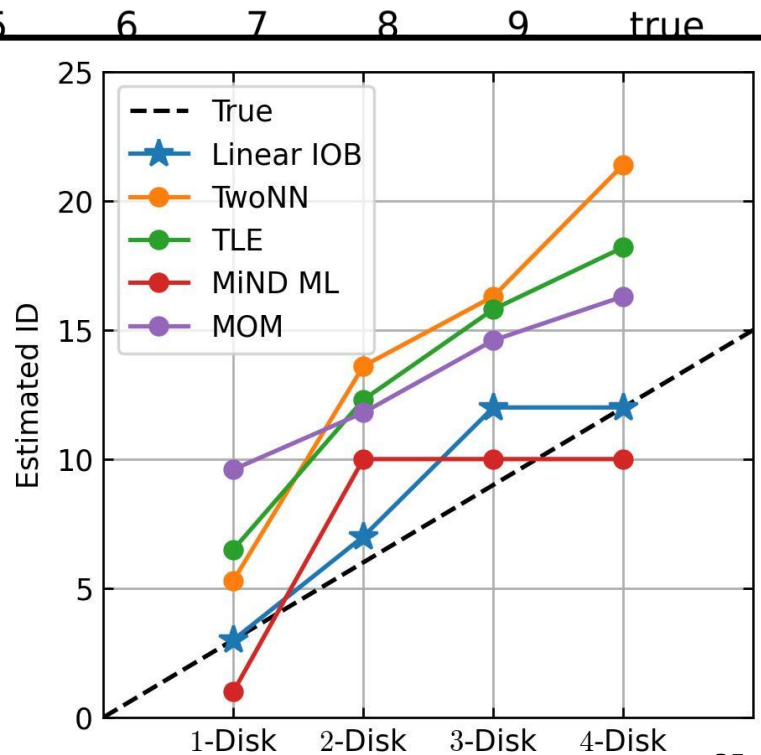
Intrinsic Dimensionality Estimation

Ho, Zhao, &
Wandelt 2023,
arXiv:2305.11213

open
channels: 0



“How many
physical
parameters
describe my
data?”



Manifold Fitting

“Cosmology with one galaxy?”

(Villaescusa-Navarro+2022

arxiv:2201.02202)

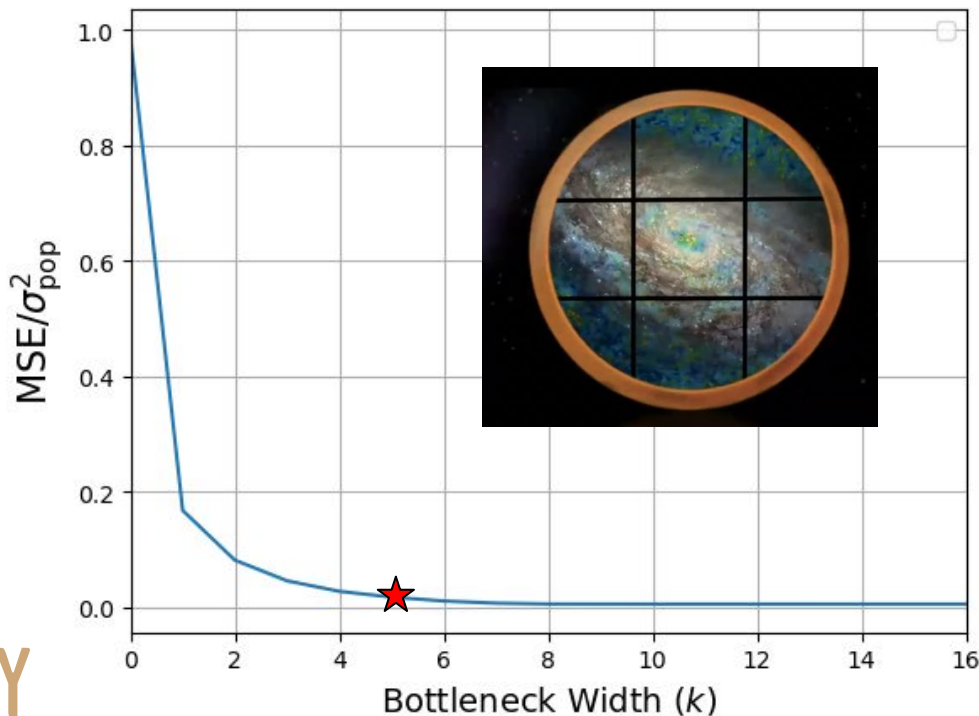


Amanda Lue

Shy Genel

Marc Huertas-Company

“Do galaxy properties live on a low-dimensional manifold?”



M_{BH}	V_{max}	SFR
M_{tot}	σ_v	U
R_{tot}	$R_{V_{\text{max}}}$	K
R_*	Z_{gas}	g
M_{gas}	Z_*	Spin
M_*		

PRELIMINARY

Accelerated Symbolic Regression

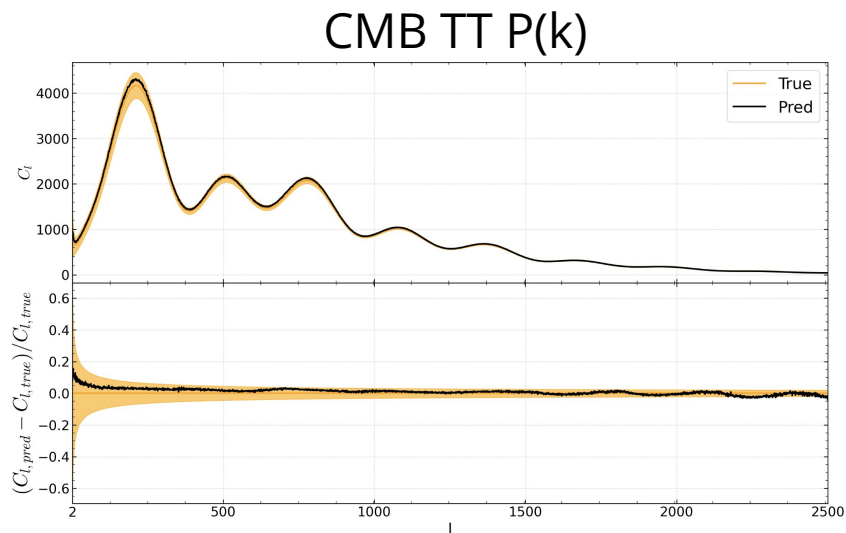
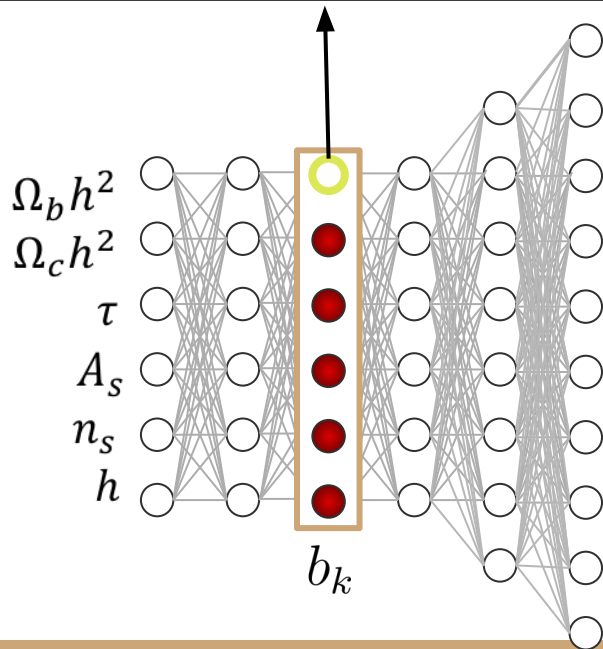
PRELIMINARY

“How can we do symbolic regression in high-dimensions?”



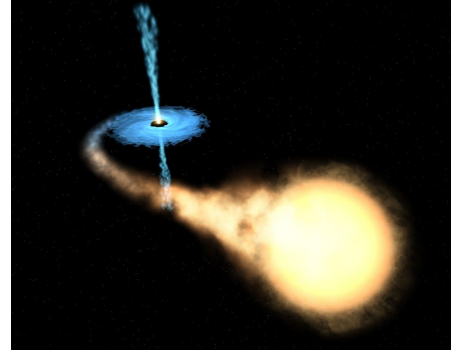
Xiaosheng Zhao
Deaglan Bartlett
Benjamin Wandelt

$$-0.572 \Omega_b h^2 + 0.446 \Omega_c h^2 - 0.156 A_s + 0.126 h - 0.103 \sin(4.748 \Omega_b h^2) + 0.031$$

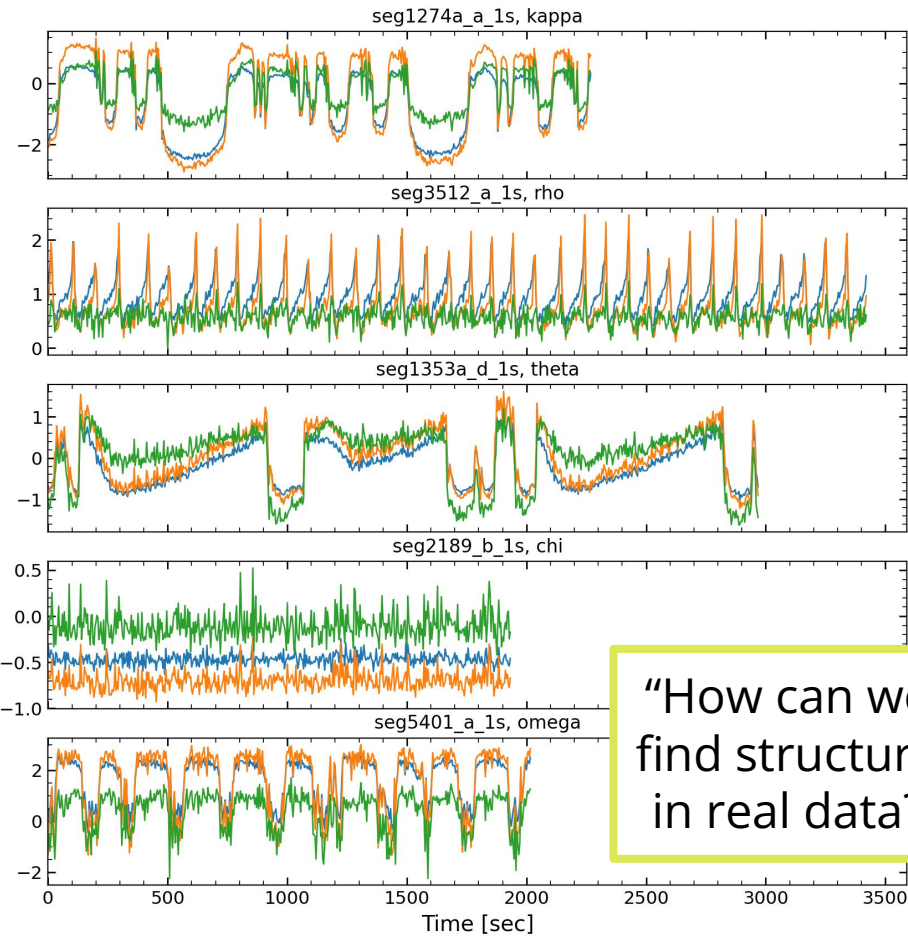


Complexity Ranking

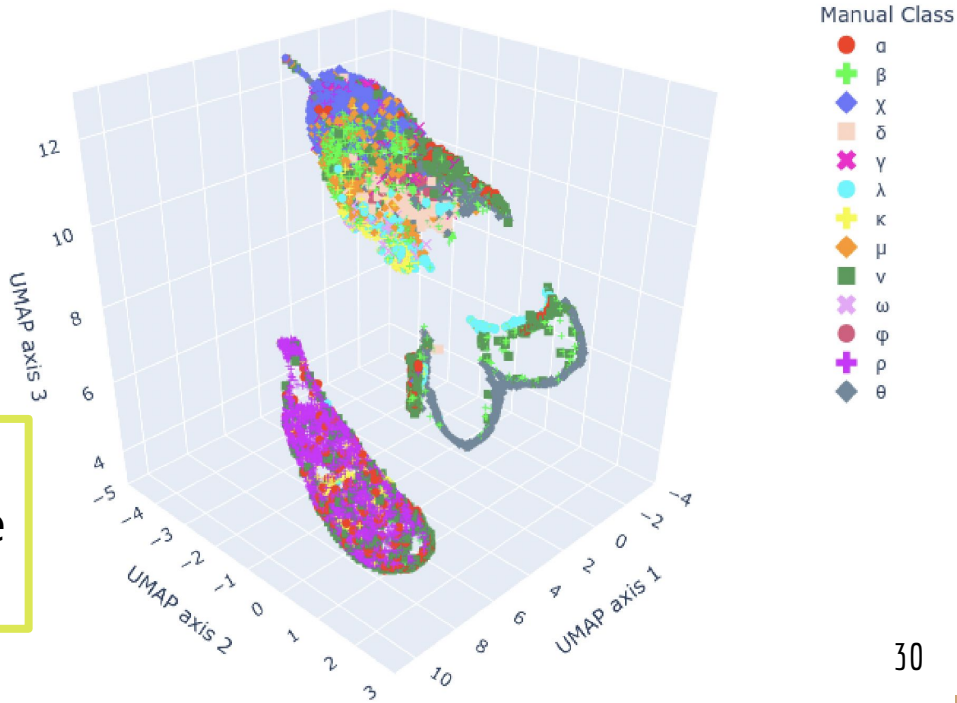
GRS 1915+105
(Ricketts+2022, arXiv: 2301.10467)



Matthew Ho
Benjamin Ricketts
Daniela Huppenkothen

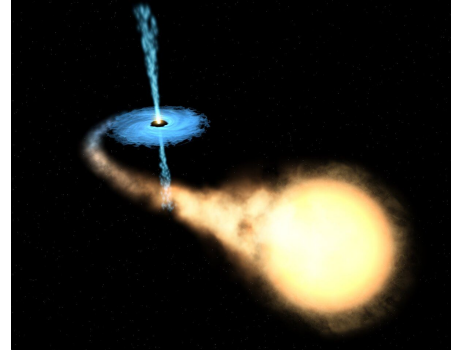


“How can we find structure in real data?”

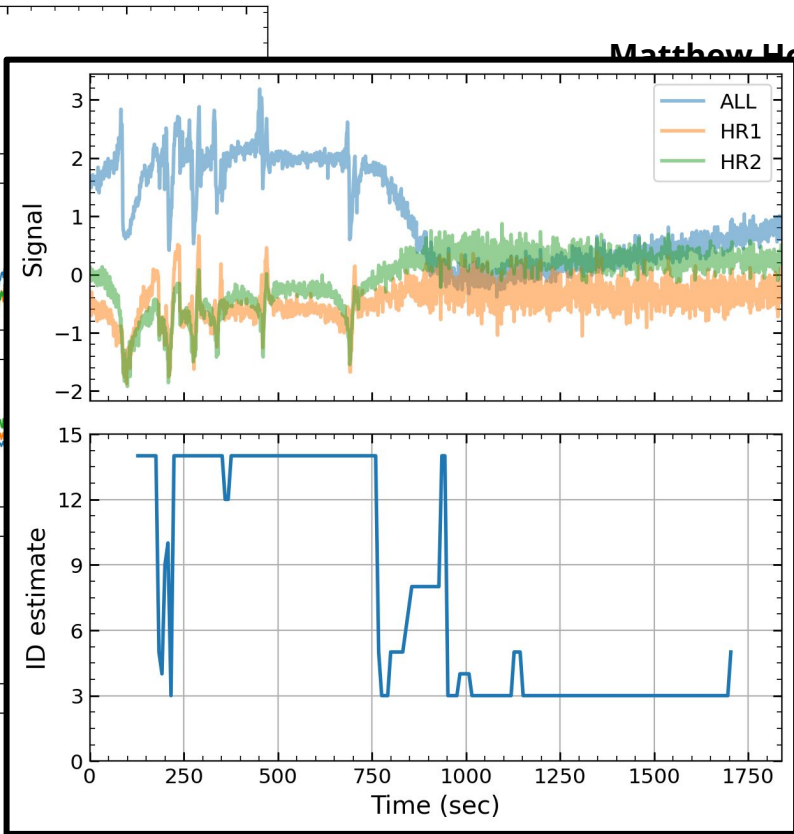
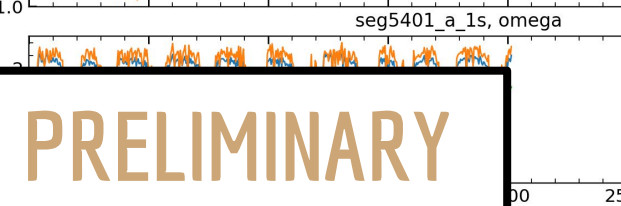
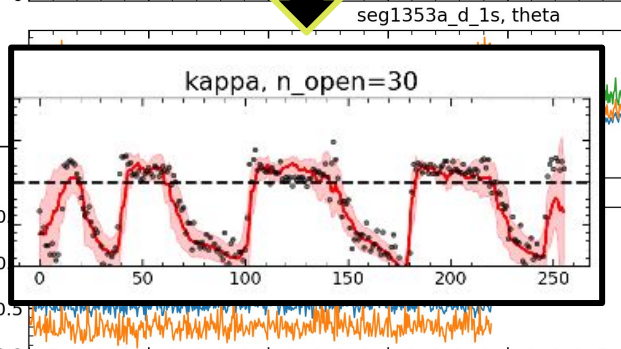
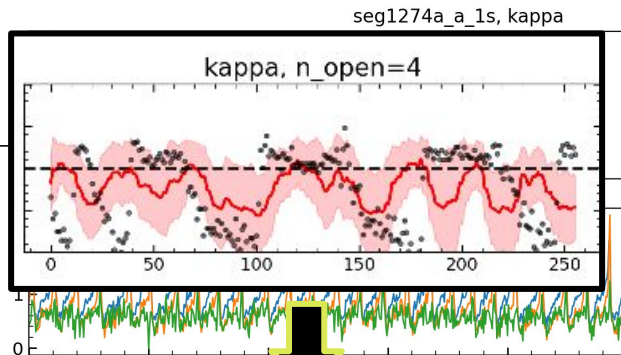


Complexity Ranking

GRS 1915+105
(Ricketts+2022, arXiv: 2301.10467)



Matthew Ho



Manual Class

- α
- β
- χ
- δ
- γ
- λ
- κ
- μ
- ν
- ω
- ϕ
- ρ
- θ



PRELIMINARY

Conclusion

We've introduced **Information-Ordered Bottlenecks (IOBs)**, a procedure for ranking information in latent neurons.

It is useful for:

- Explainability
- Intrinsic Dimensionality Estimation
- Manifold Fitting
- Accelerated Symbolic Regression
- Complexity Ranking

Matthew Ho

(matthew.annam.ho@gmail.com)

Information-Ordered Bottlenecks for Adaptive Semantic Compression


arXiv: 2305.11213



pytorch-iops

<https://github.com/maho3/pytorch-iops>

Includes Jupyter notebooks to generate all non-science examples!



<https://docs.google.com/presentation/d/17InlrPckMS53niYqt2wlnRHVMh47dyxYdCg202WSJoQ/edit?usp=sharing>

