Modality-Agnostic Variational Compression of Implicit Neural Representations

Jonathan Richard Schwarz*[1,2], Jihoon Tack*[3], Yee Whye Teh[1], Jaeho Lee[4], Jinwoo Shin[3]

*Equal Contribution, [1]DeepMind, [2]University College London, [3]KAIST, [4]POSTECH

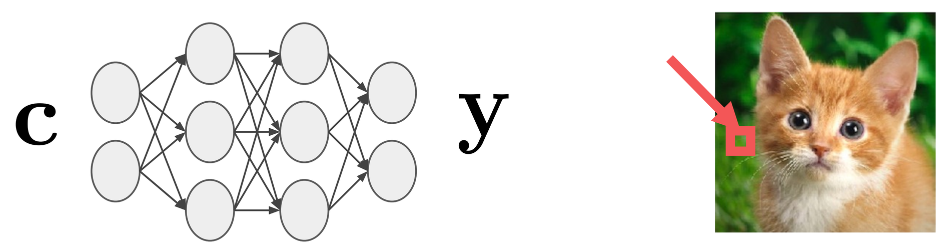
TL;DR. Applying neural compression to datasets of Implicit Neural Representations results in modality-agnostic compression methods.

ICML International Conference On Machine Learning

Paper Link

Data- as Model-Compression

INR (Implicit Neural Representation) represents each data as a neural network approximating a coordinate-to-signal mapping function

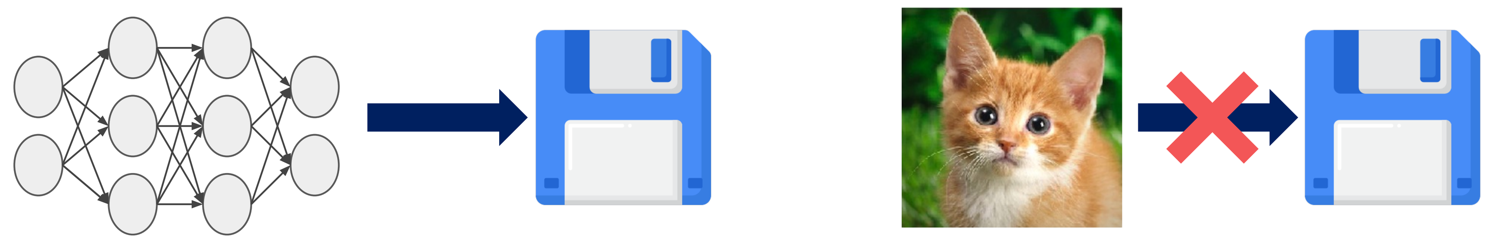


$$f_{\text{img}}\left(\tfrac{53}{255}, \tfrac{152}{255}\right) = (0, 1, 0)$$

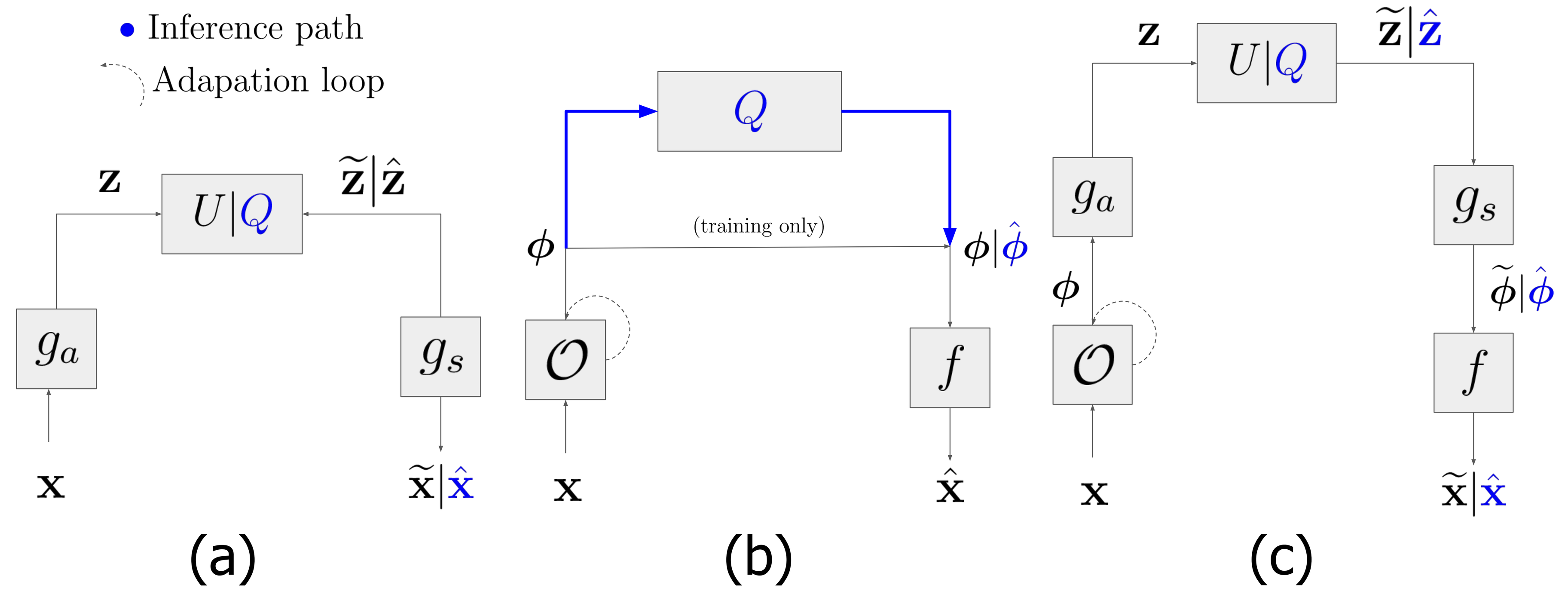
(R, G, B)

🔍 **Paradigm shift: Modality-agnosticism**

- → Store INRs rather than signals



Overview



a) Conventional neural compression

b) Conventional modality-agnostic compression with INRs

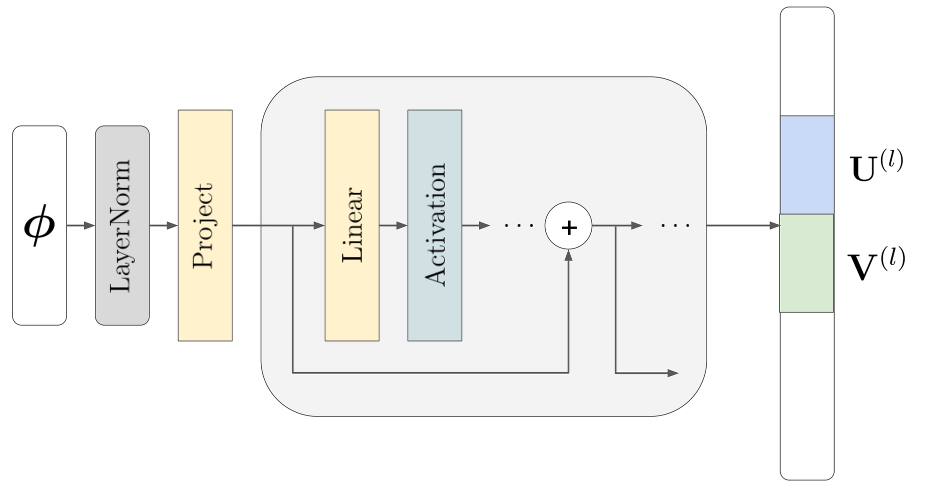
c) **VC-INR (ours): Combine both strengths**

- suggest improved conditioning for INRs
- suggest improved compression for INRs

VC-INR: Variational Compression of INRs

Step 1: Conditioning INRs via subnetwork selection

- Key idea: low-rank soft-gating for sparse network selection



$$\mathbf{G}^{(l)}_{\mathtt{low}} := \sigma\left(\mathbf{U}^{(l)}\mathbf{V}^{(l)\top}\right)$$

$$\mathbf{c}^{(l-1)} \mapsto \sin\left(\omega_0\left(\mathbf{G}^{(l)}_{\mathtt{low}} \odot \mathbf{W}^{(l)}\, \mathbf{c}^{(l-1)} + \mathbf{b}^{(l)}\right)\right)$$

- Low-rank matrixes are mapped from a low-dim. latent ϕ

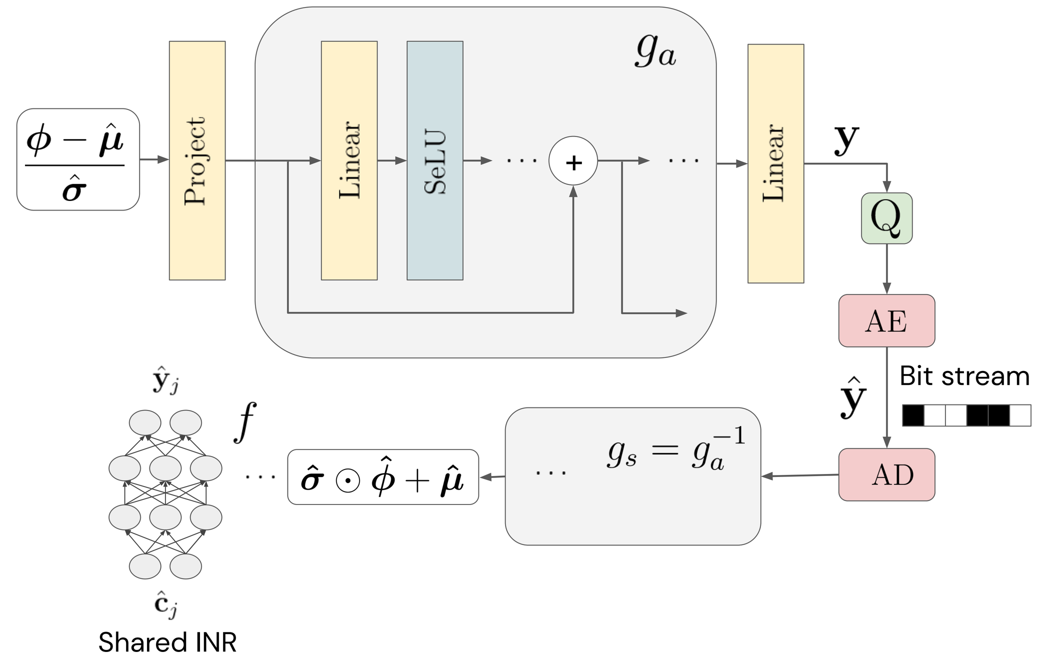
🤔 How to learn such latents? We use meta-learning

$$\min_{\theta,\phi_0} \mathrm{E}_{\mathbf{x}\sim p(\mathbf{x})}\left[\mathcal{L}_{\mathtt{MSE}}(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{x})\right] \quad \text{where } \boldsymbol{\phi} = \boldsymbol{\phi}_0 - \alpha\nabla_{\boldsymbol{\phi}_0}\mathcal{L}_{\mathtt{MSE}}(\boldsymbol{\theta}, \boldsymbol{\phi}_0, \mathbf{x})$$

Globally shared parameter — Signal specific parameter (i.e., modulation)

Step 2: Compressing INRs via variational compression

- Key idea: we apply neural compression on the learned latent ϕ
- → Project into a bitstream to apply *lossless* entropy encoding

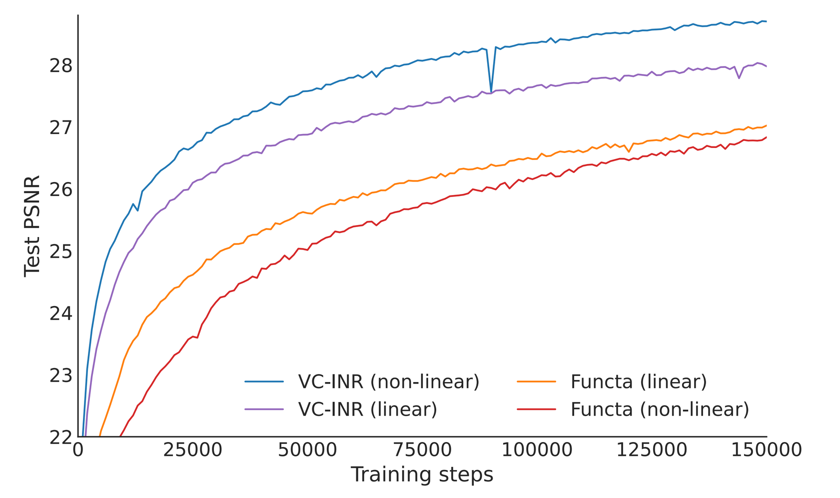


$$\mathcal{L}_{\mathtt{rate}} + \lambda\mathcal{L}_{\mathtt{distortion}} = -\log_2\left[p_{\hat{\mathbf{z}}}\left(Q(g_a(\boldsymbol{\phi}; \boldsymbol{\pi}_a))\right)\right] + \lambda\mathcal{L}_{\mathtt{MSE}}(g_s(\hat{\mathbf{z}}; \boldsymbol{\pi}_s), \boldsymbol{\phi})$$

Experimental Results

Effectiveness of the advanced conditioning

Dataset	Model	Performance @ dim(ϕ) 64	128	256	512	1024
ERA5 (4×)	Functa	43.2	43.7	43.8	44.0	44.1
	MSCN	44.6	45.7	46.0	46.6	46.9
	VC-INR	**45.0**	**46.2**	**47.6**	**49.0**	**50.0**
CelebA-HQ	Functa	21.6	23.5	25.6	28.0	30.7
	MSCN	21.8	23.8	25.7	28.1	**30.9**
	VC-INR	**22.0**	**23.9**	**26.0**	**28.3**	30.8
SRN Cars	Functa	22.4	23.0	23.1	23.2	23.1
	MSCN	22.8	**24.0**	**24.3**	24.5	24.8
	VC-INR	**23.9**	**24.0**	**24.3**	**25.2**	**25.5**
ShapeNet10	Functa	$_{99}$.30	$_{99}$.40	$_{99}$.44	$_{99}$.50	$_{99}$.55
	MSCN	$_{99}$.43	$_{99}$.50	$_{99}$.56	$_{99}$.63	$_{99}$.69
	VC-INR	$_{99}$**.54**	$_{99}$**.61**	$_{99}$**.64**	$_{99}$**.70**	$_{99}$**.71**



Functa [1]: Shift modulation

MSCN [2]: Sparse shift modulation

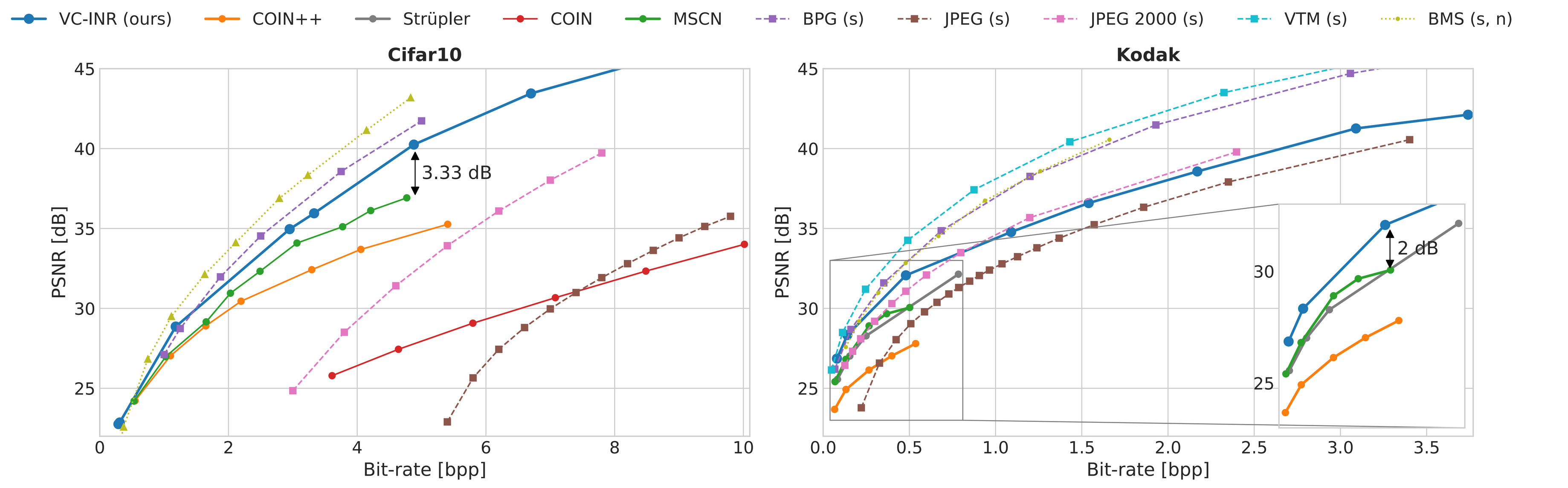
VC-INR: Low-rank soft-gating modulation

[1] Dupont et al., From data to functa: Your data point is a function and you can treat it like one, ICML 2022

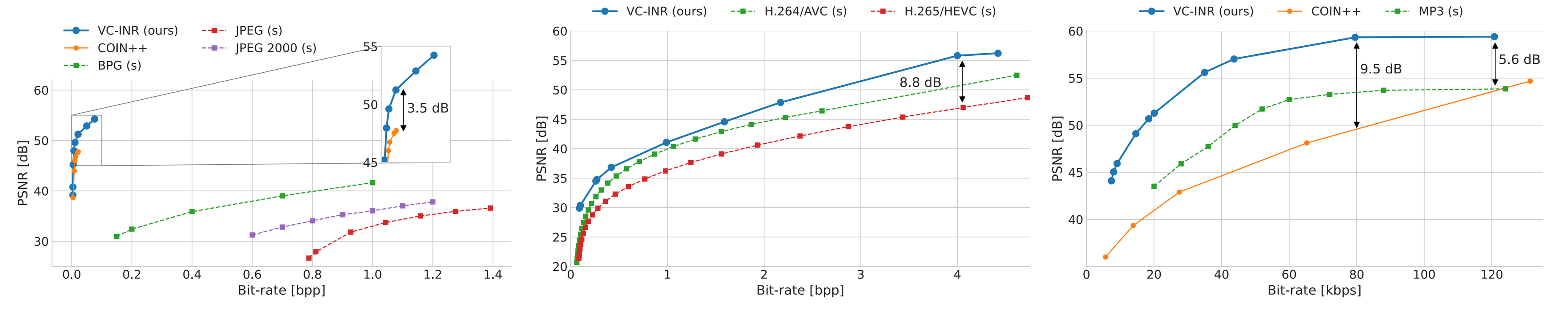
[2] Schwarz and Teh, Meta-Learning Sparse Compression Networks, TMLR 2022

Data compression results: VC-INR outperforms

- a) prior INR-based compression schemes
- b) modality specific-codecs, e.g., JPEG 2000, HEVC/AVC, MP3



CIFAR-10 (left), Kodak (right), image



ERA5, manifold — UCF-101, video — Librispeech, audio

(s): Modality-specific, **(s, n)**: Neural Compression, **solid line:** modality-agnostic