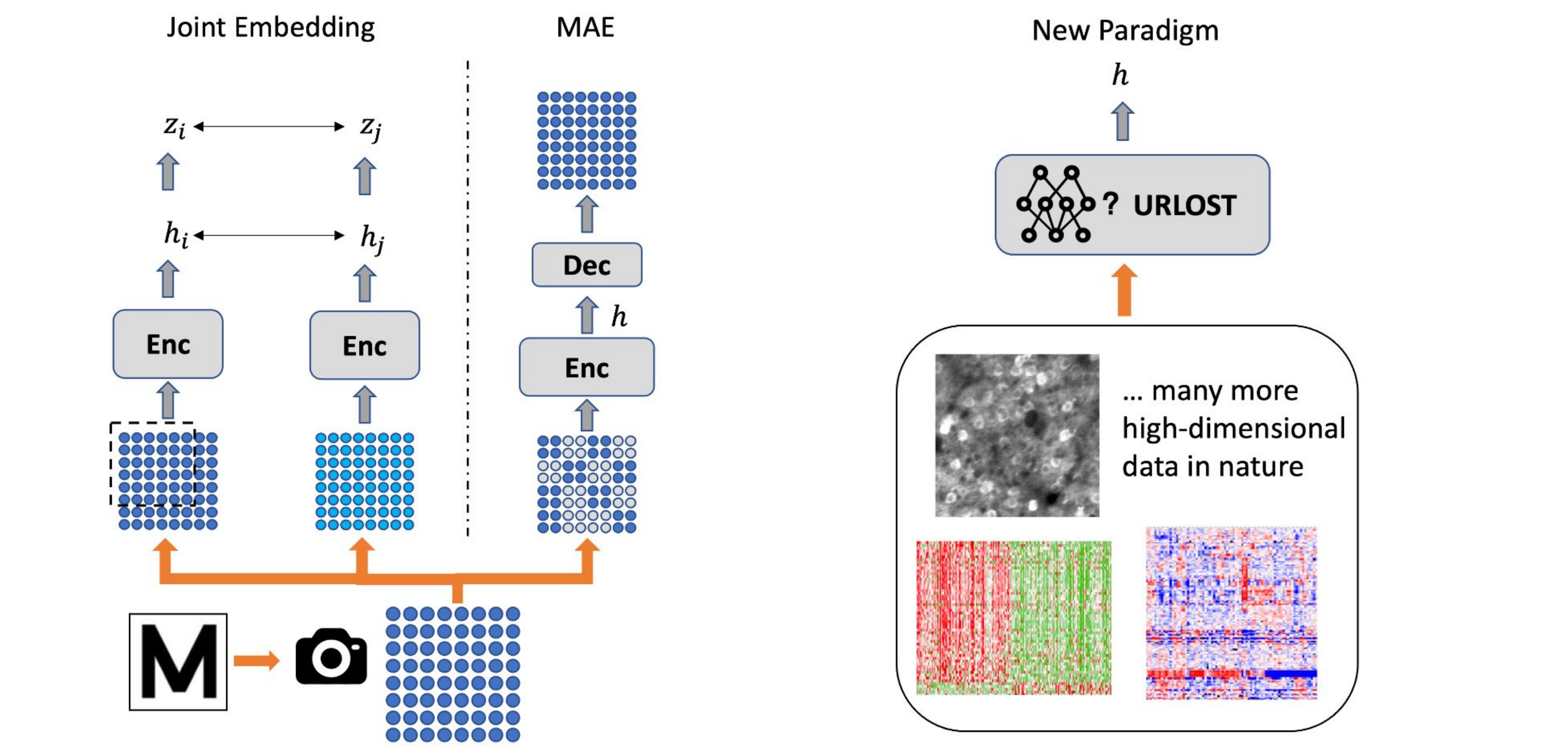
URLOST: Unsupervised Representation Learning without Stationarity or Topology

Zeyu Yun[1], Juexiao Zhang[2], Yann Lecun[2,3], Yubei Chen[4]

(1) Redwood Center for Theoretical Neuroscience, UC Berkeley
(2) New York University, (3) FAIR at Meta
(4) ECE Dept, UC Davis

Goal

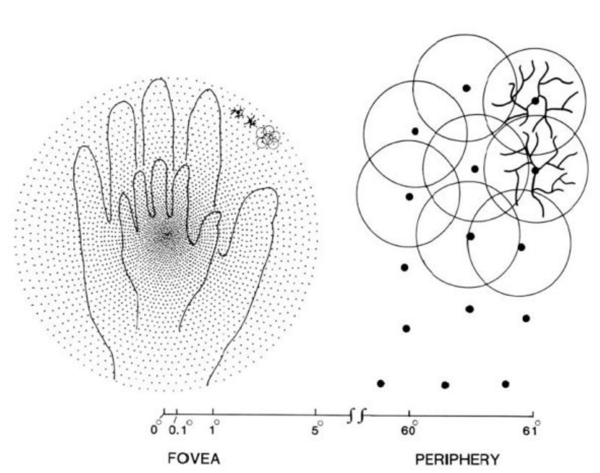
Our objective is to build robust **unsupervised** representations for high-dimensional signals without prior information on explicit **topology** or **stationarity**. The learned representations are intended to enhance performance in downstream tasks such as classification.



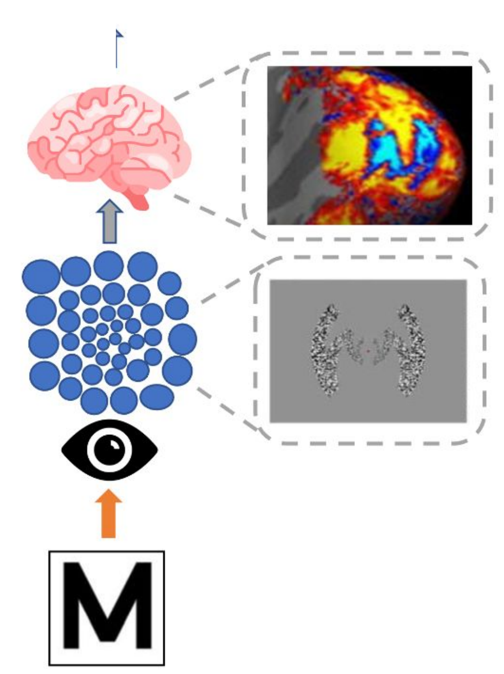
Data w/ stationarity and topology **Data w/o stationarity and topology**

Motivation

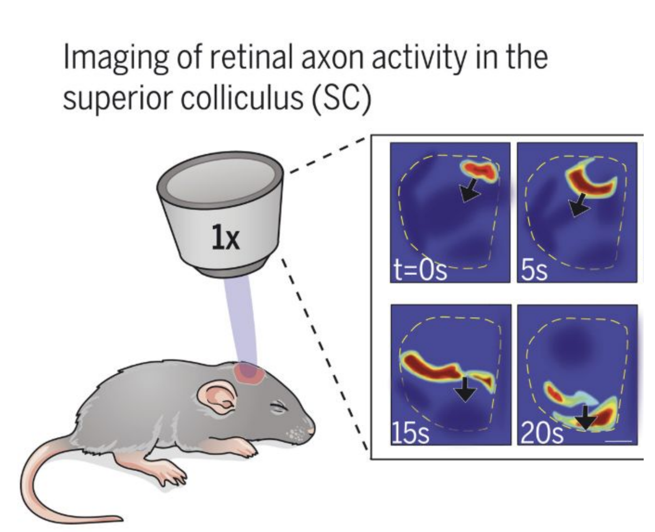
Visual system process visual signal without stationarity and topology. How are they doing it? Could we borrow similar idea and build it into a machine learning system?



Irregular sensor array

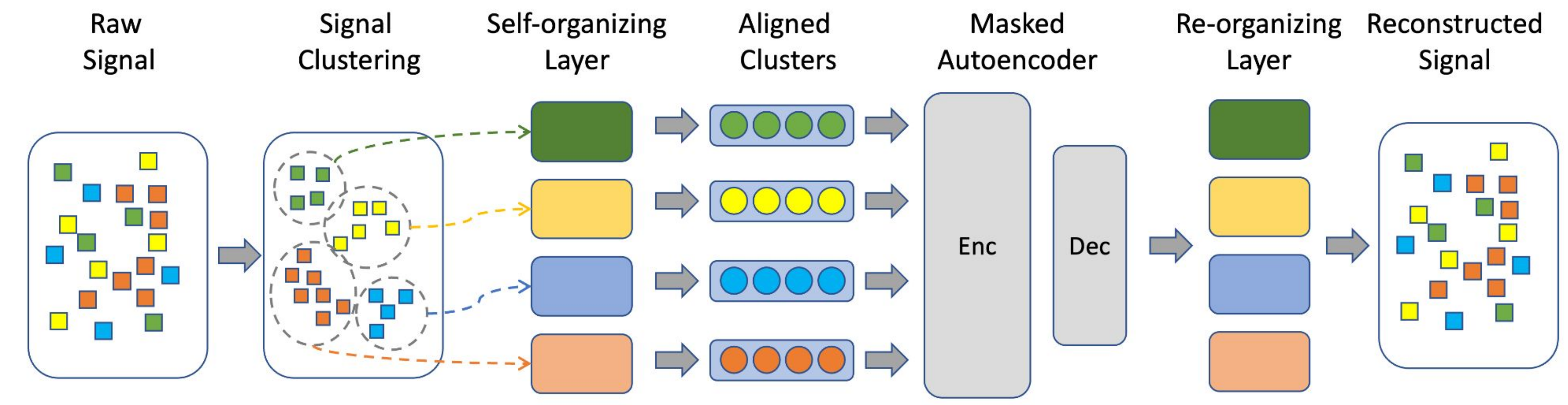


Retinotopic

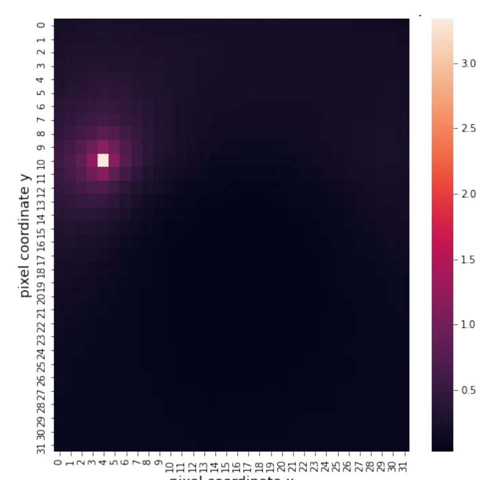


Retinal waves

Method

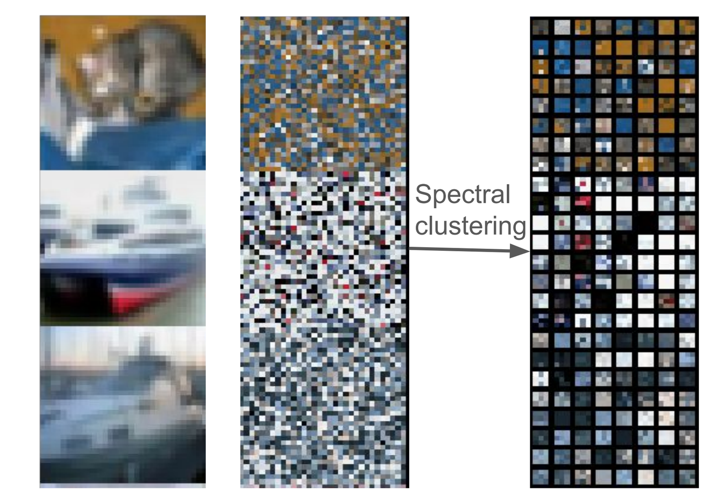


Mutual information: Given a high dimensional singal, we use mutual information to define similarity between different dimensions.



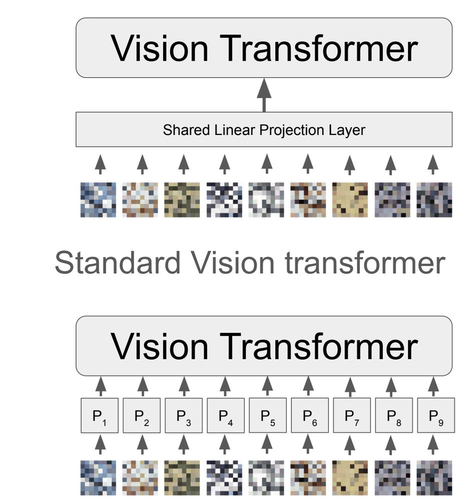
Mutual information between pixel at location (4,10) to all other pixels

Spectral clustering: To process similar dimensions together, we group dimensions using pairwise similarity and spectral clustering.



raw signal pixels clusters

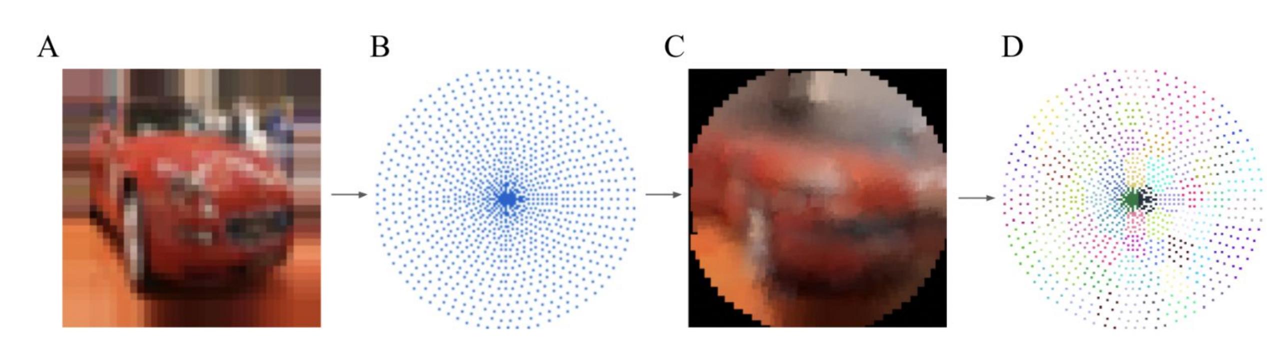
Self-organizing layer: Unlike shared projection layer, each patch has its own projection layer.



VIT with Self organizing layer

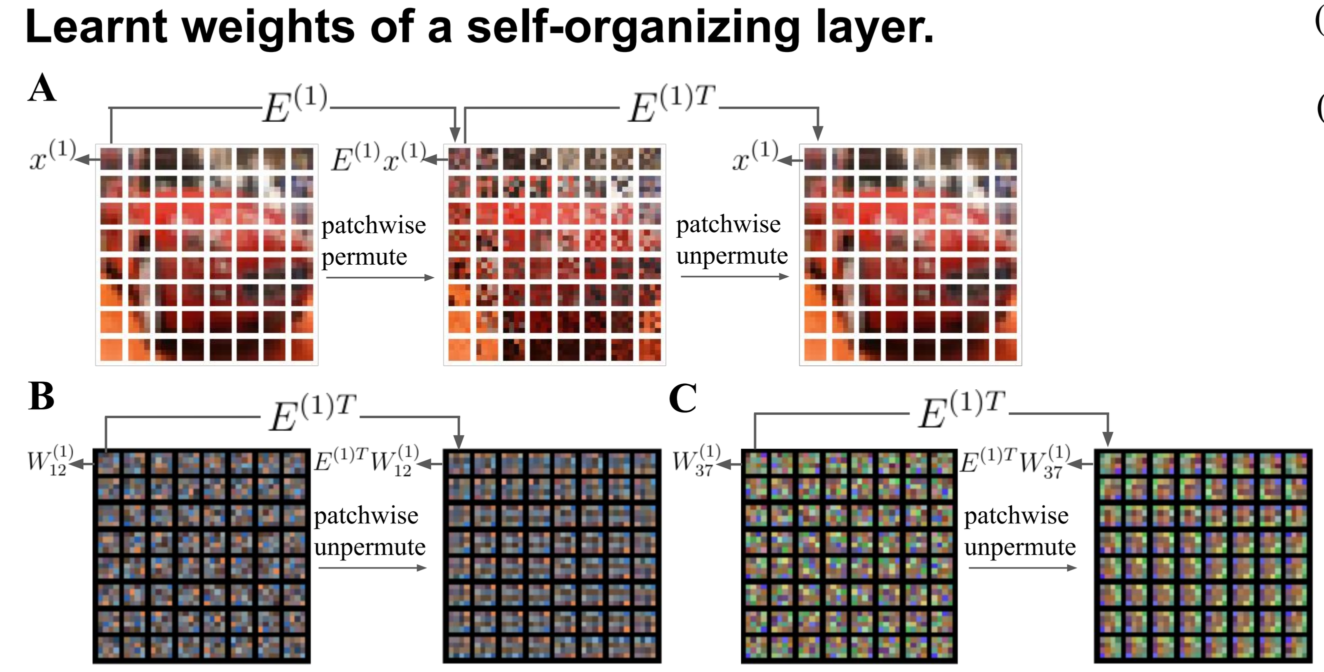
Qualitative result

Visualization of result of spectral clustering on simulated biological visual signal.



(A) An image in CIFAR-10 dataset. (B) Retina sampling lattice. Each blue dot mimics a retinal ganglion cell. (C) Visualization of the car image's signal sampled using the retina lattice. (D) density-adjusted spectral clustering results are shown. Each unique color represents a cluster.

Learnt weights of a self-organizing layer.



(A) Each patch in the image undergoes different permutation.
(B) (C) If they learnt projection layer are undergo the same permutation, they look the same.

Why? In a standard vision transformer, the linear projection layer should learn some filtering on input patches. If each patch is undergoes different random permutation, to align all the patches, the projection layer should learn filtering + the inverse permutation. If the projection layer for each patch indeed learn the inverse permutation, then they should looks the same if we apply the ground truth permutation on them.

Result

Table 1. Evaluation on computer vision and synthetic biological vision dataset. We create dataset where signal have no topology or stationarity. We compare the performance of URLOST with different baseline model: different neural network backbones and different unsupervised learning algorithms (MAE and simCLR).

Dataset	Method	Backbone	Eval Acc
CIFAR-10	MAE	ViT (Patch)	88.3 %
	MAE	ViT (Pixel)	56.7 %
	SimCLR	ResNet-18	**90.7 %**
	SimCLR	ViG	53.8 %
Permuted CIFAR-10	URLOST MAE	ViT (Cluster)	**86.4 %**
	MAE	ViT (Pixel)	56.7 %
	SimCLR	ResNet-18	47.9 %
	SimCLR	ViG	40.0 %
Foveated CIFAR-10	URLOST MAE	ViT (Cluster)	**85.4 %**
	MAE	ViT (Pixel)	48.5 %
	SimCLR	ResNet-18	38.0 %
	SimCLR	ViG	42.8 %

Table 2. Evaluation on computer vision and synthetic biological vision dataset. We compare the performance of URLOST with different baseline model: different neural network backbones and different unsupervised learning algorithms (normalized raw signal, MAE, β-VAE.).

Method	V1 Response Decoding Acc	TCGA Classification Acc
Raw	73.9% ± 0.00 %	91.7 ± 0.24%
MAE	70.6% ± 0.22 %	90.6% ± 0.63%
β-VAE	75.64% ± 0.11%	94.15% ± 0.24%
URLOST MAE	**78.75% ± 0.18 %**	**94.90% ± 0.25 %**

Table 3. Ablation study 1. What if we don't cluster dimensions before doing MAE, and what if we randomly cluster dimensions?

Masking Unit	Permuted Cifar10	Foveated Cifar10	TCGA Gene	V1 Response
Clusters (URLOST)	**86.4%**	**85.4%**	**94.9%**	**78.8%**
Random patch	55.7%	51.1%	91.7%	73.9%
Individual dimension	56.7%	48.5%	88.3%	64.8%

Table 3. Ablation study 2. What if we don't use self-organizing layer? And what if we don't adjust density during spectral clustering?

Dataset	Projection	Eval Acc
Locally-Permuted CIFAR-10	shared	81.4 %
	non-shared	**87.6 %**
Permuted CIFAR-10	shared	80.7 %
	non-shared	**86.4 %**

Dataset	Cluster	Eval Acc
Foveated CIFAR-10	SC	82.7 %
	DSC	**85.4 %**

References

[1] Ge, Xinxin, et al. "Retinal waves prime visual motion detection by simulating future optic flow." *Science* 373.6553 (2021): eabd0830.
[2] Jonathan R Polimeni, Bruce Fischl, Douglas N Greve, and Lawrence L Wald. Laminar analysis of 7t bold using an imposed spatial activation pattern in human v1. Neuroimage, 52(4):1334– 1346, 2010.
[3] Marius Pachitariu, Carsen Stringer, Sylvia Schroder, Mario Dipoppa, L Federico Rossi, Matteo Carandini, and Kenneth D Harris. Suite2p: beyond 10,000 neurons with standard two-photon microscopy. BioRxiv, pp. 061507, 2016.
[4] Youzheng Xu, Yixin Xu, Chun Wang, Baoguo Xia, Qingling Mu, Shaohong Luan, and Jun Fan. Mining tcga database for gene expression in ovarian serous cystadenocarcinoma microenvironment. PeerJ, 9:e11375, 2021.
[5] Van Essen, David C., and Charles H. Anderson. "Information processing strategies and pathways in the primate visual system." An introduction to neural and electronic networks 2 (1995): 45-76.
[6] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Doll´ar, and Ross Girshick. Masked autoencoders are scalable vision learners. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 16000–16009, 2022.
[7] Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." International conference on machine learning. PmLR, 2020.
[8] Xiaoyu Zhang, Yuting Xing, Kai Sun, and Yike Guo. Omiembed: A unified multi-task deep learning framework for multi-omics data. Cancers, 13(12), 2021.
[9] Dosovitskiy, Alexey, et al. "An image is worth 16x16 words: Transformers for image recognition at scale." arXiv preprint arXiv:2010.11929 (2020).
[10] Katarzyna Tomczak, Patrycja Czerwi´nska, and Maciej Wiznerowicz. Review the can- cer genome atlas (tcga): an immeasurable source of knowledge. Contemporary Oncology/Wsp´ołczesna Onkologia, 2015(1):68–77, 2015.