UNIVERSITY OF OXFORD

MRG MOBILE ROBOTICS GROUP OXFORD ROBOTICS INSTITUTE

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Visual DNA: Representing and Comparing Images using Distributions of Neuron Activations

Benjamin Ramtoula Matthew Gadd Paul Newman Daniele De Martini

Mobile Robotics Group

University of Oxford

Poster session

WED-AM-274

bramtoula.github.io/vdna

We lack good ways to compare datasets

Metadata is not enough, we need to rely on images themselves

Cityscapes, A2D2, and KITTI have strong similarities on paper but have distinct looks

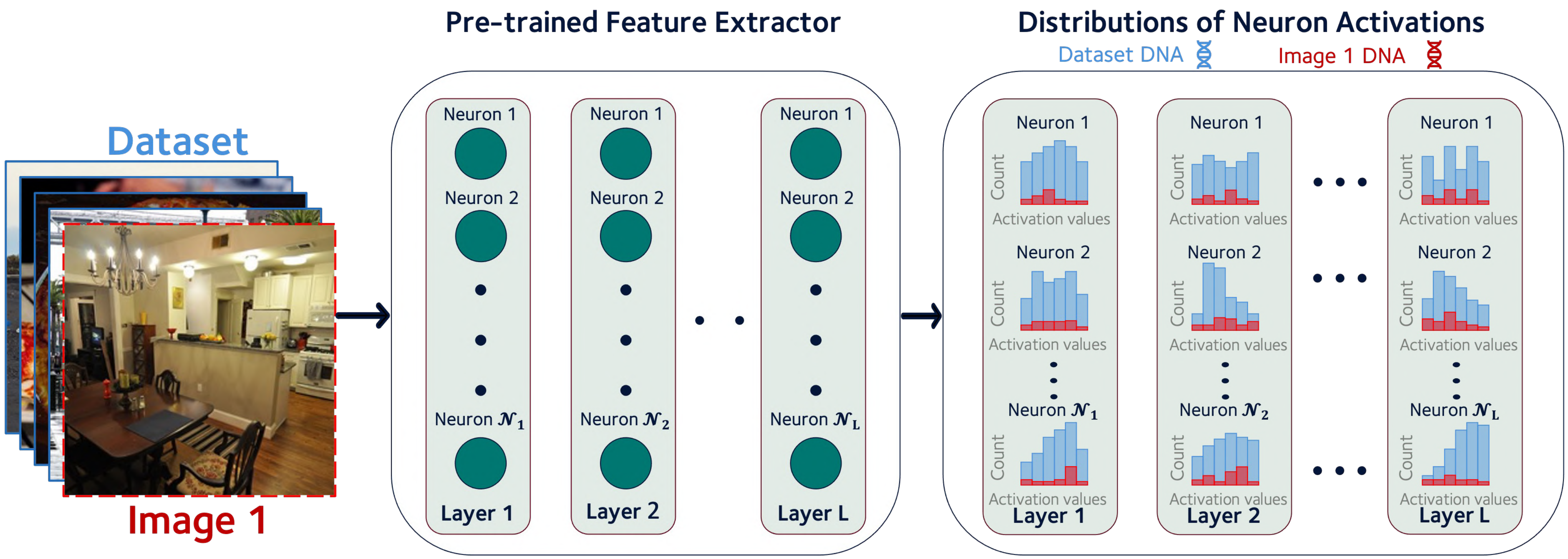
Current solutions are not sensitive to many variation types

Cityscapes version

Original | Brighter | Darker | More contrast

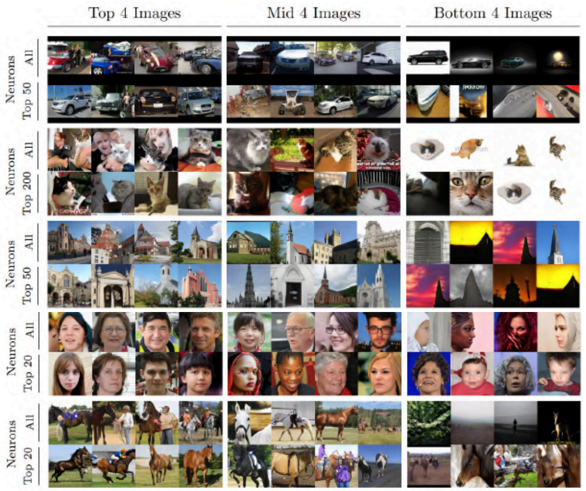
→ All lead to similar embeddings usually used for existing approaches

Distributions of Neuron Activations

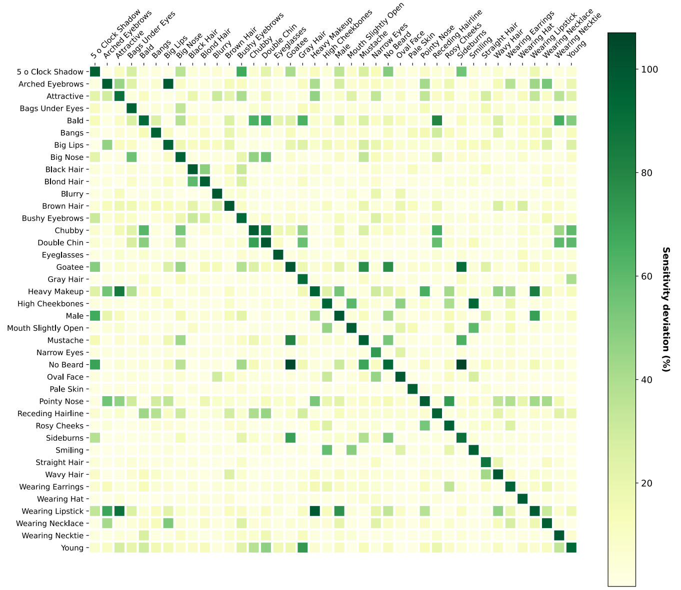


- 👉 **Granularity** from individual neurons that react to different levels of patterns
- 👉 **Compact** size for any number of images
- 👉 Can be compared with **custom distances**

Many applications!



Evaluating **realism** of **synthetic** images



Comparing datasets while **ignoring selected factors**

Retrieving most **similar images** to a dataset

MSeg: A Composite Dataset for Multi-domain Semantic Segmentation

John Lambert* [1,3], Zhuang Liu*[1,2], Ozan Sener[1], James Hays[3,4], and Vladlen Koltun[1]

[1]Intel Labs, [2]University of California, Berkeley, [3]Georgia Institute of Technology, [4]Argo AI

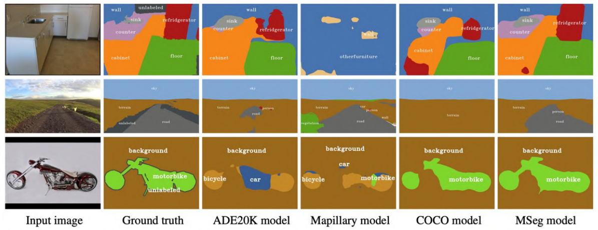


Figure 1: MSeg unifies multiple semantic segmentation datasets by reconciling their taxonomies and resolving incompatible annotations. This enables training models that perform consistently across domains and generalize better. Input images in this figure were taken (top to bottom) from the ScanNet [8], WildDash [44], and Pascal VOC [10] datasets, none of which were seen during training.

Predicting transfer learning performance

Try it yourself!

No tuning or training necessary for holistic comparisons!

```
pip install vdna
```

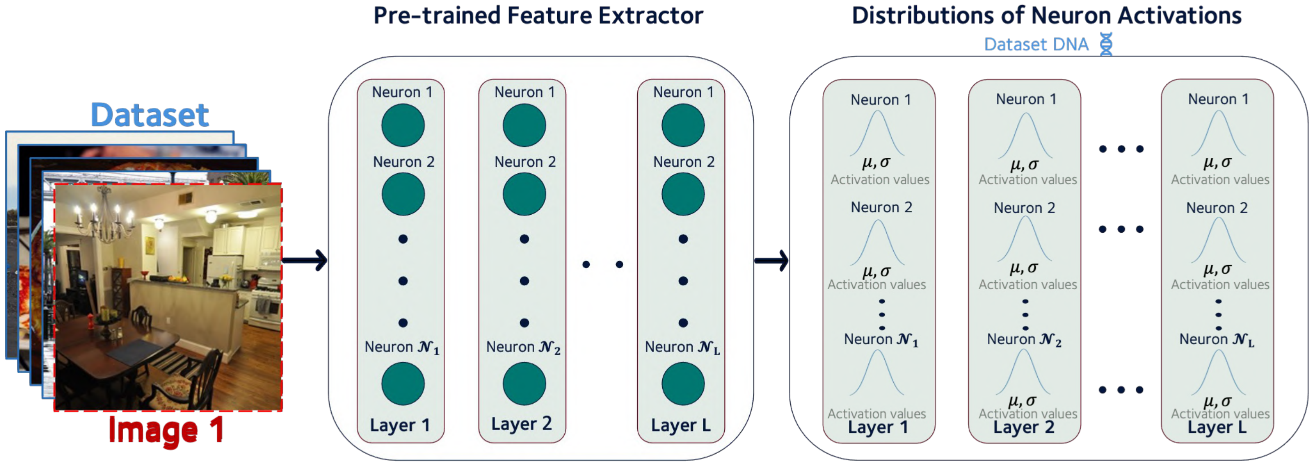
```
from vdna import VDNAProcessor, EMD

vdna_proc = VDNAProcessor()

vdna1 = vdna_proc.make_vdna(source="/path/to/dataset1")
vdna2 = vdna_proc.make_vdna(source="/path/to/dataset2")

emd = EMD(vdna1, vdna2)
```

More specifically, what is a Visual DNA?

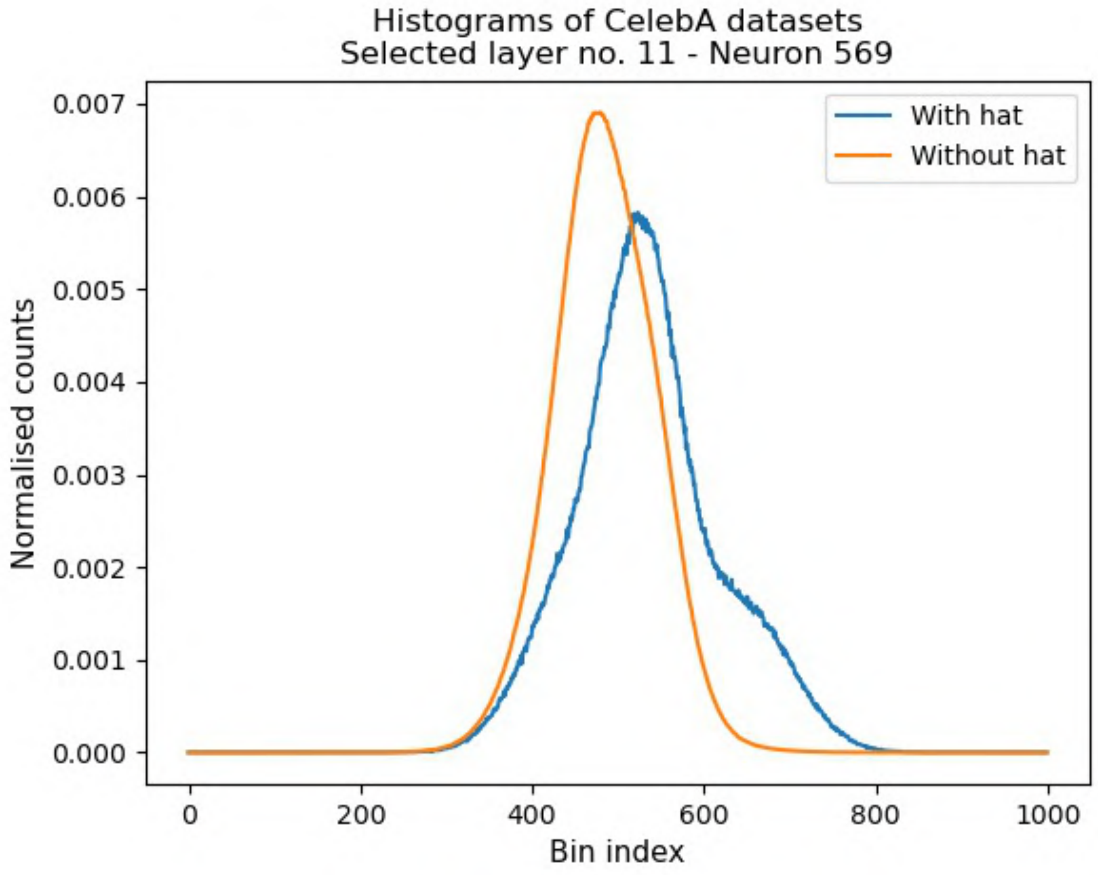
Given images to represent…

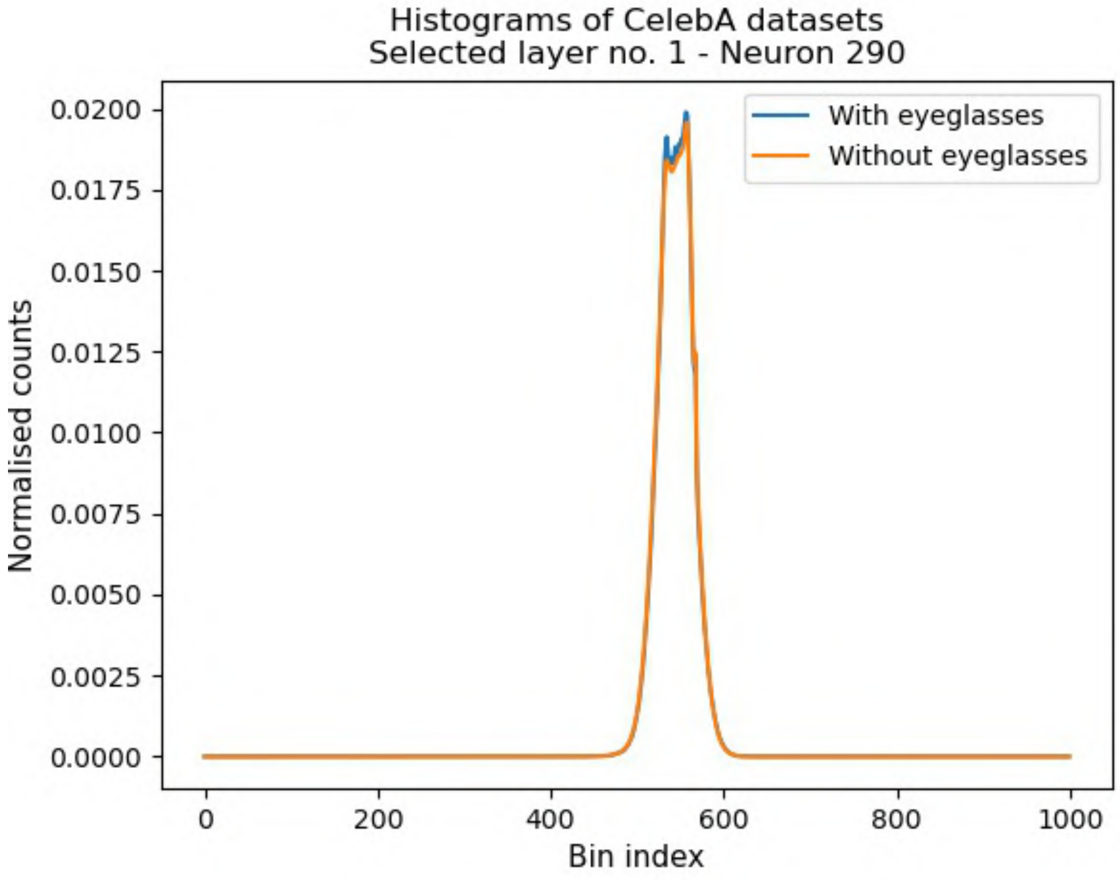
We rely on features learned by a neural network

We keep track of distributions fitted to activations at selected neurons



But... Histograms allow better fit





DNAs are compact

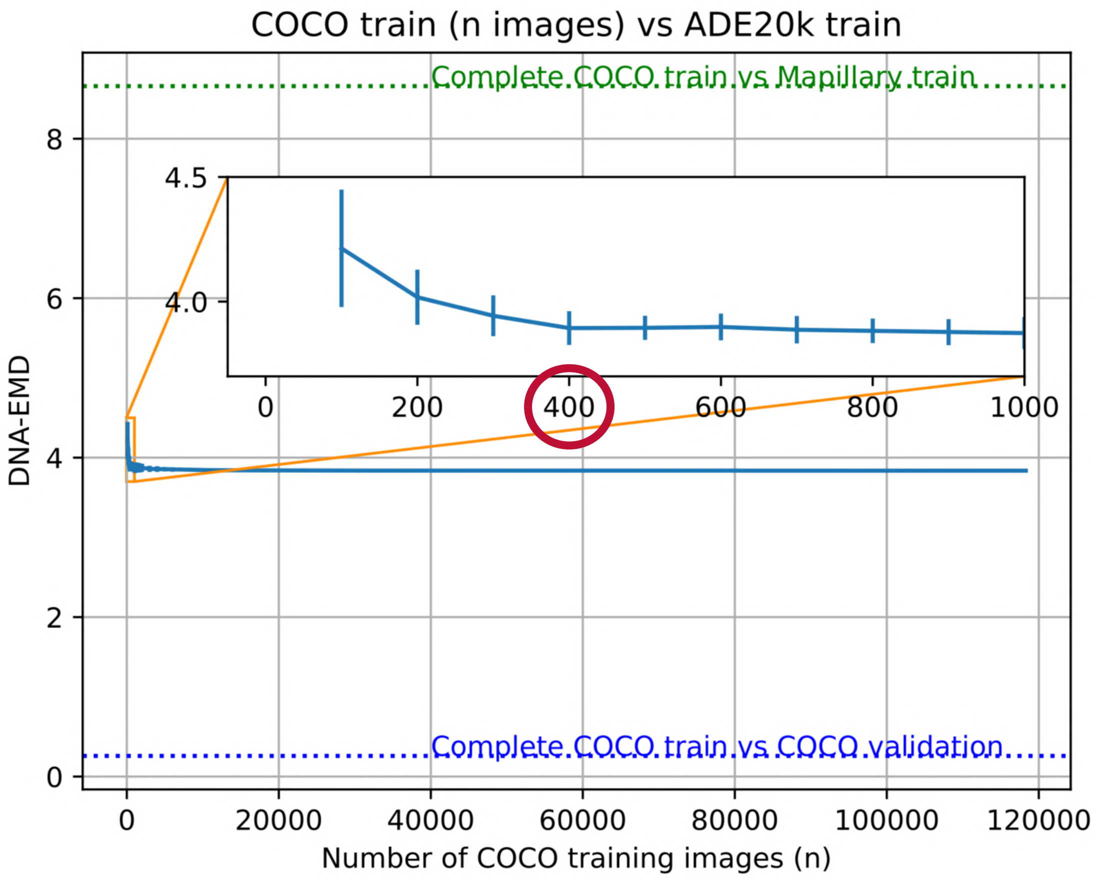
Representing the Flickr-Faces-HQ Dataset (FFHQ) dataset

70 000 images	**89.1 GB**
Raw feat. extractor activations	**1.10 TB**
DNA (histograms)	**14.8 MB**
DNA (Gaussians)	**84.7 kB**

DNAs don’t need that many images

400 images

can be sufficient to create a representative DNA

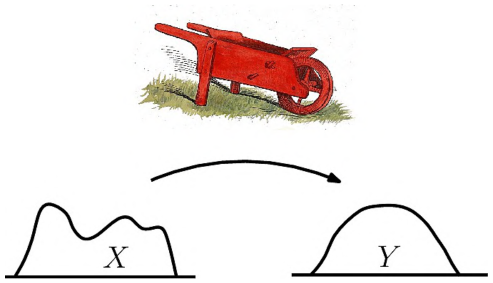


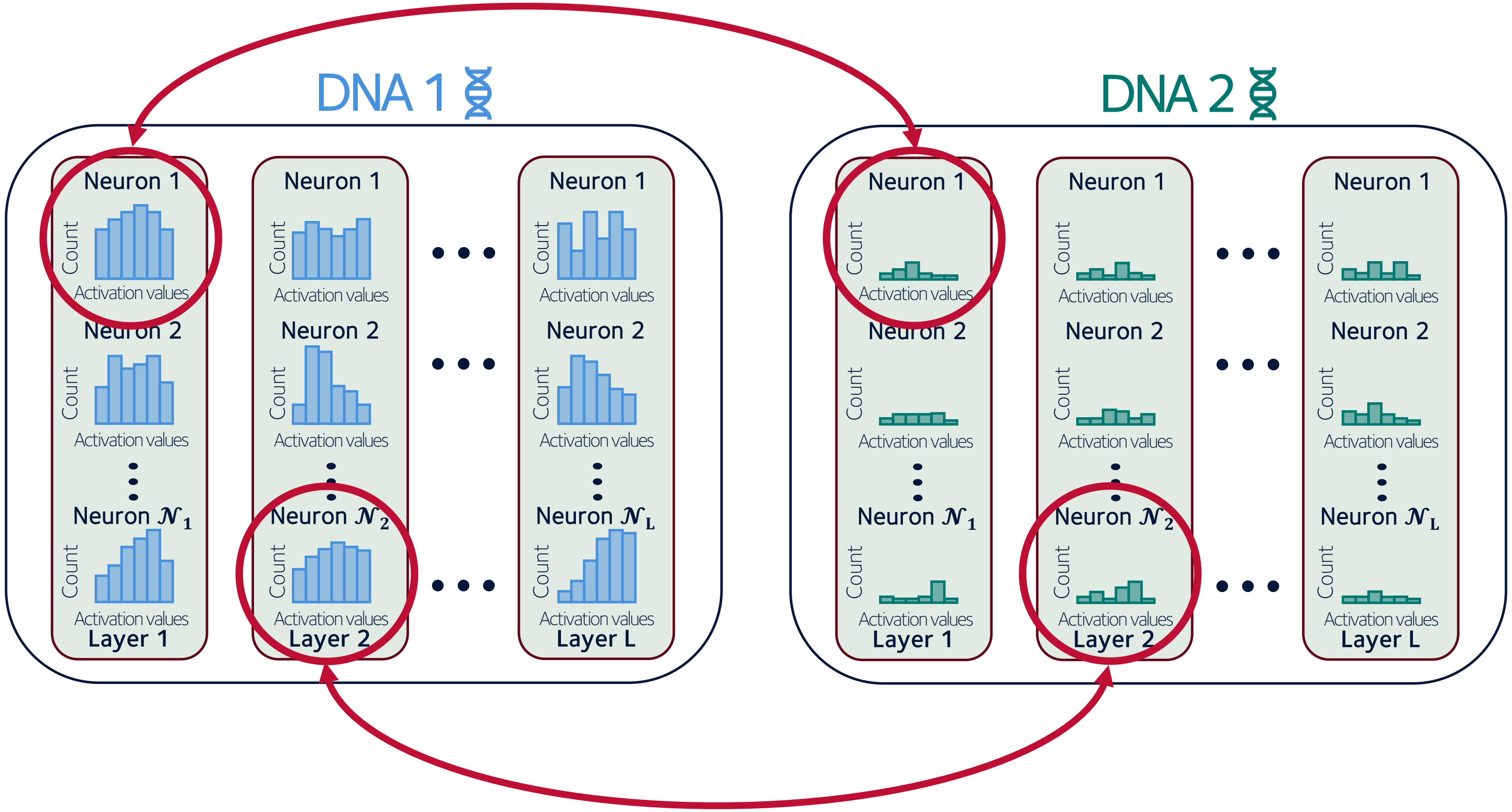
Comparing DNAs

Comparing DNAs → Comparing distributions for all neurons

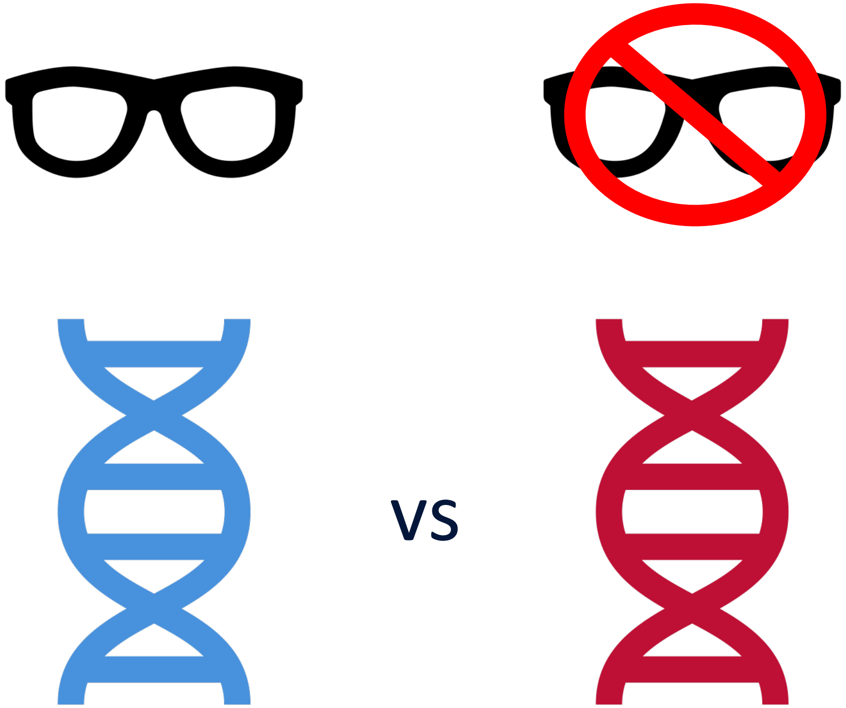
👉 e.g., Earth Mover's Distance for histograms

👉 e.g., Fréchet Distance for Gaussians



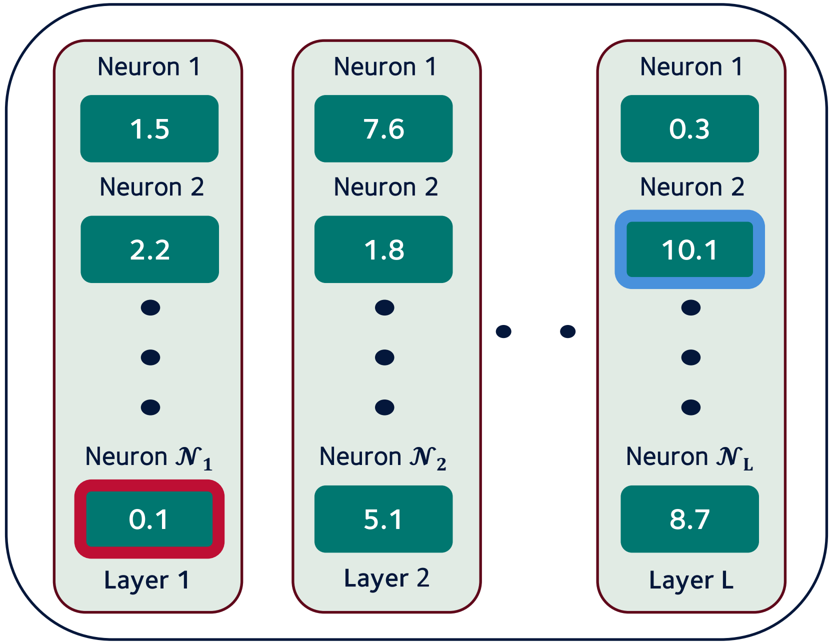


Finding neurons sensitive to an attribute



With attribute   Without attribute

Earth Mover's Distances



👉 Small EMD → Neuron not sensitive to difference in images
👉 High EMD → Neuron sensitive to differences

We can order neurons by their sensitivity to the attribute

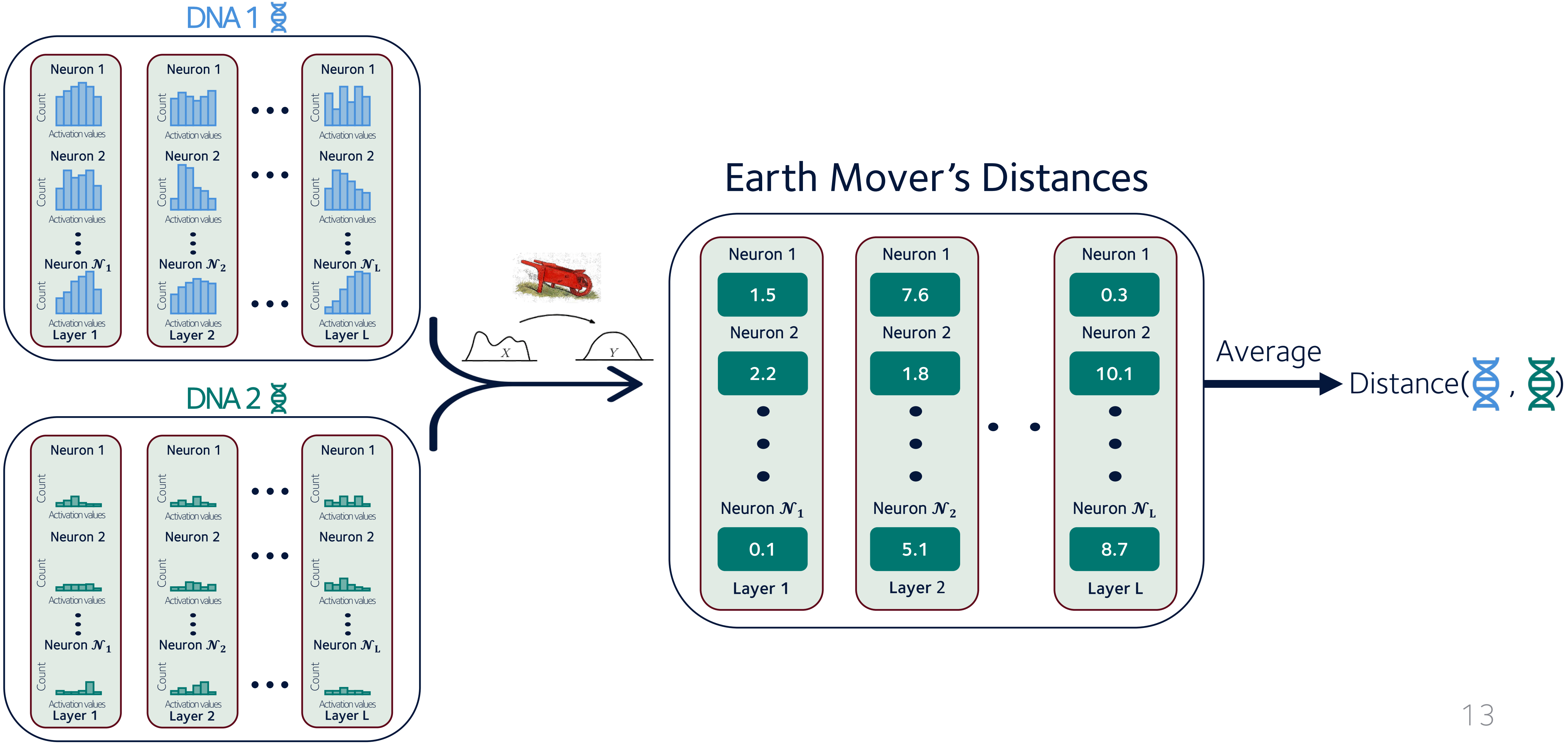
Few neurons are sufficient to focus on high-level concepts

Reference image

Images reacting most to the
top 1 “eyeglasses” neuron

Images reacting most to the
top 20 “wearing hat” neurons

Holistic comparison

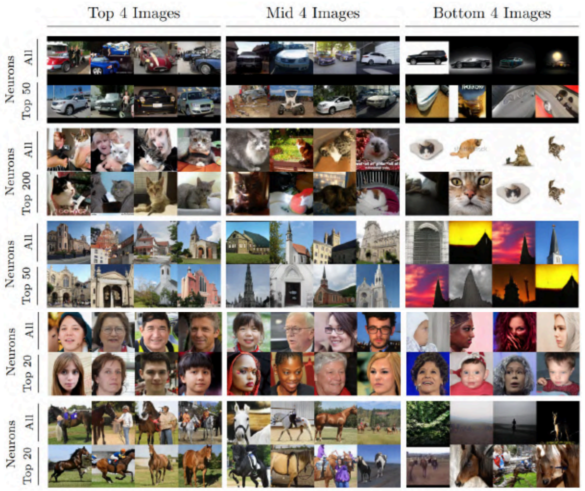


Custom comparison

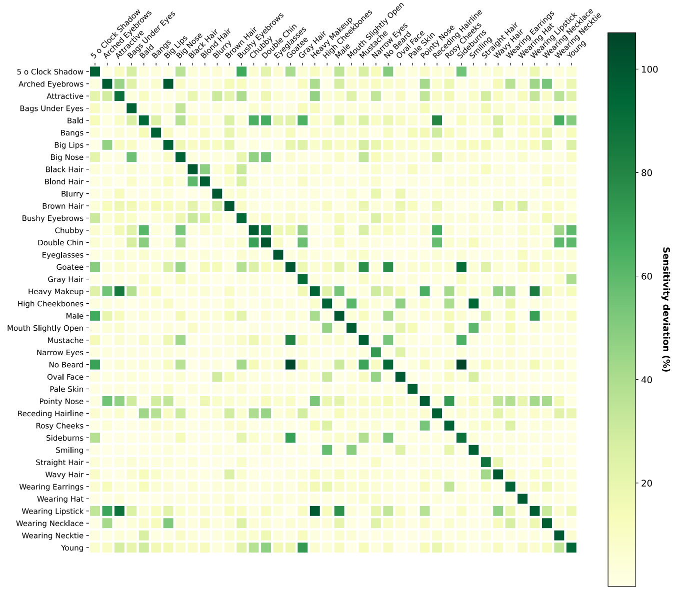
Many applications!

Optimise a loss to find weights to linearly combine neurons

Use neurons focusing on general realism by measuring sensitivity of fake vs real data



Evaluating **realism** of **synthetic** images



Comparing datasets while **ignoring selected factors**

Use **all neurons**

Cityscapes dataset

Retrieving most **similar images** to a dataset

MSeg: A Composite Dataset for Multi-domain Semantic Segmentation

John Lambert* [1,3], Zhuang Liu* [1,2], Ozan Sener[1], James Hays[3,4], and Vladlen Koltun[1]

[1]Intel Labs, [2]University of California, Berkeley, [3]Georgia Institute of Technology, [4]Argo AI

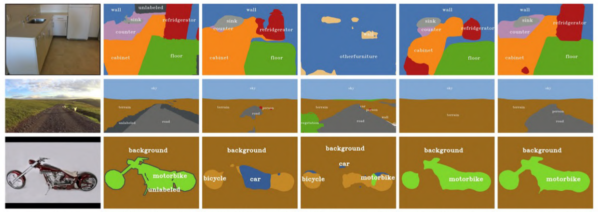


Figure 1: MSeg unifies multiple semantic segmentation datasets by reconciling their taxonomies and resolving incompatible annotations. This enables training models that perform consistently across domains and generalize better. Input images in this figure were taken (top to bottom) from the ScanNet [8], WildDash [44], and Pascal VOC [10] datasets, none of which were seen during training.

Compare DNAs using neurons of the last layer

Predicting transfer learning performance

Thank you for your interest!

🔎 Many more details and results in the paper.

⌨️ Still many applications and extensions to explore!

🐍 If relevant to you, please try it!

```
pip install vdna
```

This work was supported by the EPSRC Centre for Doctoral Training in Autonomous Intelligent Machines and Systems [EP/S024050/1] and Oxbotica.

UKRI Engineering and Physical Sciences Research Council

AIMS EPSRC Centre for Doctoral Training in Autonomous Intelligent Machines & Systems