Visualizing Classification Structure in Large-Scale Classifiers

Bilal Alsallakh** Zhixin Yan* Shabnam Ghaffarzadegan* Zeng Dai* Liu Ren*

How?

Compute pair-wise class similarities by correlating the logits for each pair over the dataset

Order the similarity matrix using hierarchical bi-clustering to surface class similarity structure as nested block patterns

Why?

Expose unreliable features - Analyze feature sharing - Introduce additional supervision - Expose data quality issues - Compare behavior of groups

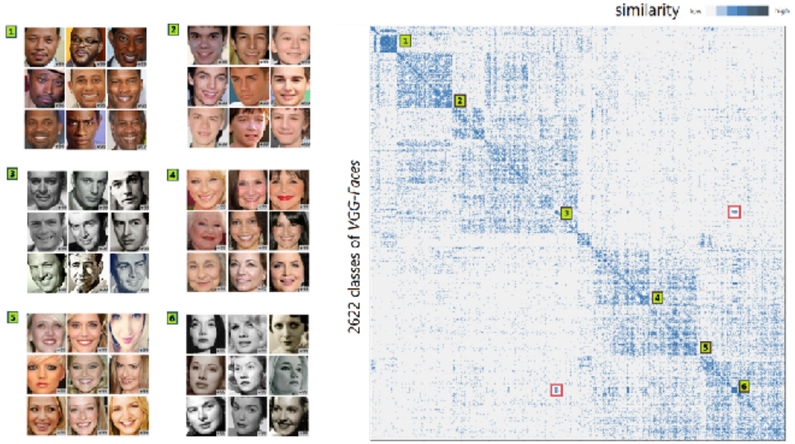
Places365

Each row / column is one of 365 classes (scenes)

Indoor / outdoor scenes define major similarity groups

Nested subgroups, e.g., nature vs city outdoor scenes

VGGFace



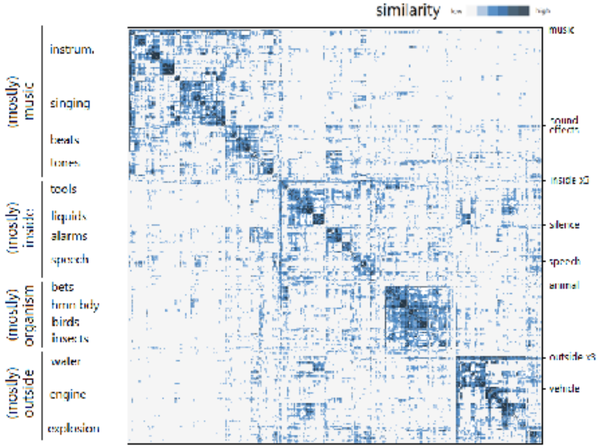
Each row / column is one of 2622 classes (celebrities)

Male vs. female celebrities define major similarity groups

Prominent subgroups based on wrinkles and hair / skin color

Chromaticity (split marked in red) indicates unreliable features

AudioSet



Each row / column is one of 527 classes (audio events)

Dataset is multi-labeled with high class imbalance

Major groups are music, inside, organism, and outside

inside & outside do not surface in the pre-defined taxonomy

Article and Example Notebook github.com/bilalsal/blocks

BOSCH * FACEBOOK *