CNN Filter DB: An Empirical Investigation of Trained Convolutional Filters

Paul Gavrikov[1], Janis Keuper[123]
[1]IMLA, Offenburg University, [2]CC-HPC, Fraunhofer ITWM, [3]Fraunhofer Research Center ML

INSTITUTE FOR MACHINE LEARNING AND ANALYTICS

SCAN ME

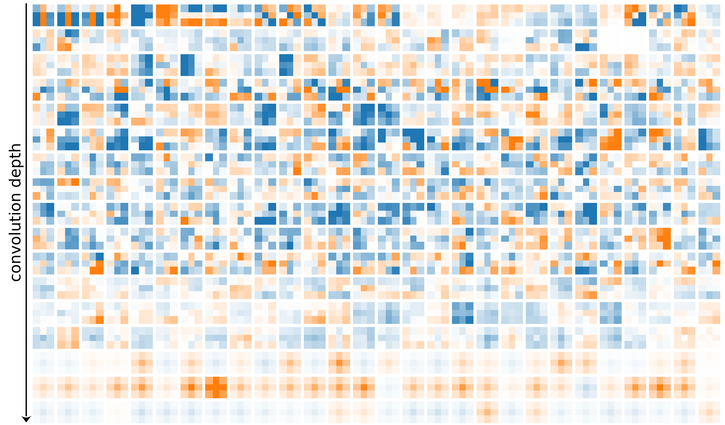
CVPR JUNE 19-24 2022 NEW ORLEANS • LOUISIANA

Motivation

We propose to study shifts in the learned filter weights of publicly available trained models. In this work, we present a dataset focusing on the properties of the dominantly used **3x3 convolution filter kernels**.

Dataset

1.4 billion filters extracted from **647 public CNN models** trained for **10 tasks** (e.g. Image Classification, Segmentation, Generation, Super Resolution) on **16 visual domains** (e.g. natural, x-ray, seismic, depth).

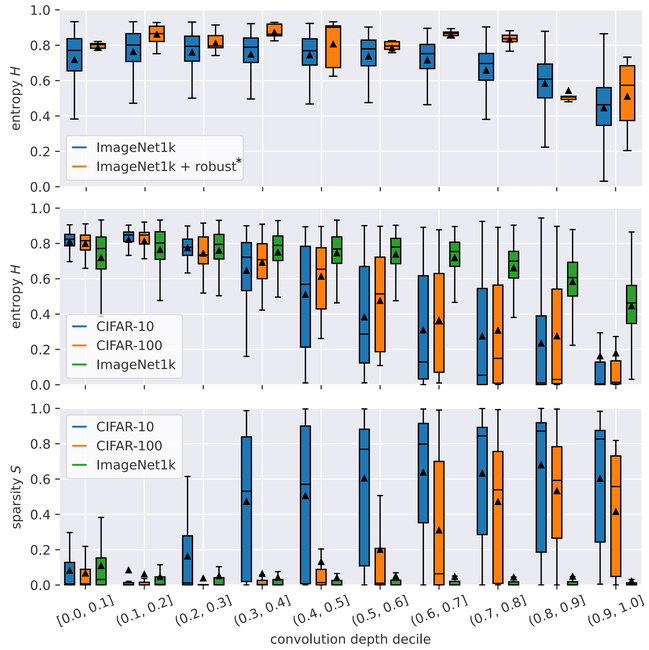


Visualization of the (degenerated) filters in each layer of a ResNet-18 trained on CIFAR-10.

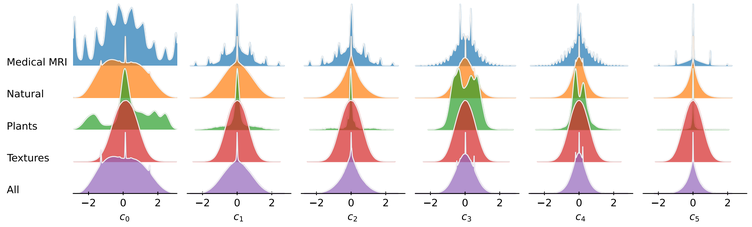
Degenerated Filters

Randomness: Lack of structure (randomly initialized models)
Low structural variety: Filters are replicated
Sparsity: Filter weights are close to zero

Filter Quality

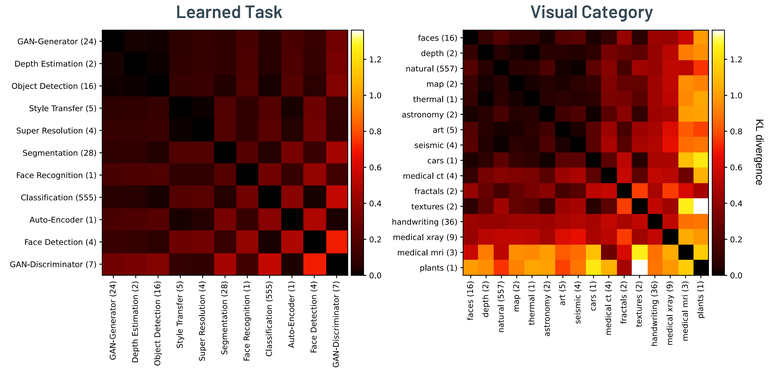


Comparison of layer entropy and sparsity between overparameterized, robust*, and regular classification models.

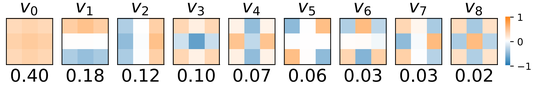


Coefficient distributions for selected filter groups on the first 6 basis vectors.

Distribution Shifts



Heatmaps showing the shift between filter coefficients grouped by various dimensions. Brighter colors indicate larger shifts.



PCA basis and explained variance of each basis vector (below) computed over all filters.

Take-home messages

- We introduce **two simple data-free methods for filter quality evaluation**
- Many public CNNs show a **large presence of degenerated filters**
- **More complex (e.g. adversarial-trained*) tasks** appear to form **higher-quality filters** on the same architectures
- The **structure** of convolutional filters **differs across models,** but models of **the same family show very similar distributions**
- Without degeneration the distributions **are mostly independent of the task** of the model, the **visual category,** and the **extracted depth**

*For more details on robustness refer to Gavrikov, P. and Keuper, J., "Adversarial Robustness through the Lens of Convolutional Filters", IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2022.