Sparsifying Convolutional Layers with Dual-Tree Wavelet Packets

Hubert Leterme, *hubert.leterme@univ-grenoble-alpes.fr*

Background

Convolutional neural networks (CNNs) [**LeCun1989**]:

✓ **state-of-the-art performances in computer vision;**

✗ **empirical approach, lack of theoretical understanding.**

Discrete wavelet transforms [**Mallat2009**]:

✓ **built on well-established mathematical framework;**

✓ **successful in feat. extraction, signal compression and denoising;**

Oscillating patterns very often observed in CNN kernels [**Yosinski2014**].

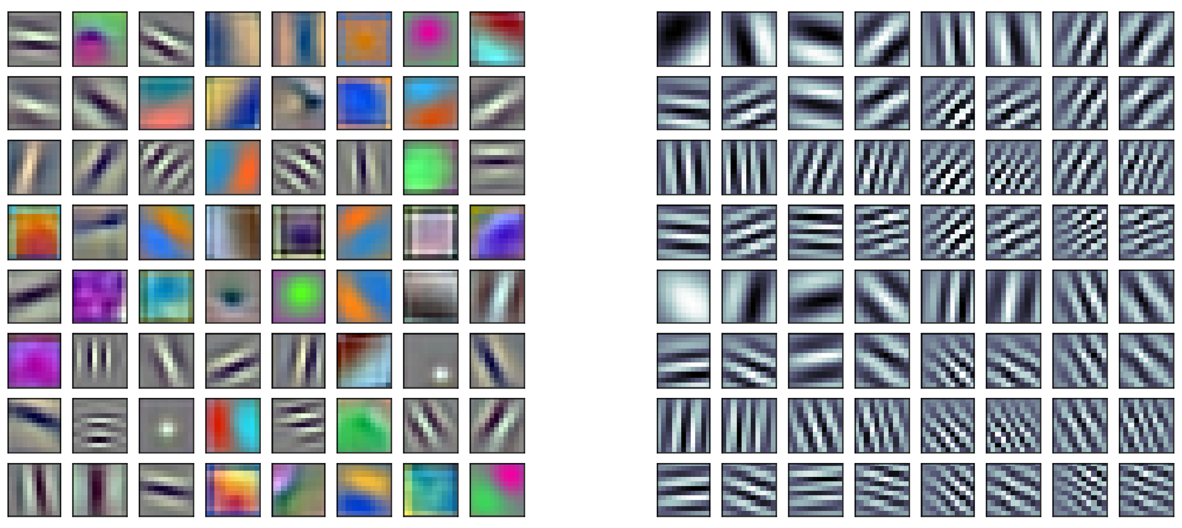

Figure 1. Left: AlexNet's first layer after training with ImageNet. Right: selection of dual-tree complex wavelet packet filters with 3 decomposition stages (real part only).

Objective

✓ **Theoretical and empirical sudy** of CNN properties for image classification.

Roadmap:

1. **Build a sparse model** of existing CNN architectures, based on the **dual-tree wavelet packet transform** (DT-ℂWPT) [**Bayram2008**].
   $\Longrightarrow$ **Subset selection** among all possible configurations.
2. **Assess model's accuracy** with respect to the original architecture, from a qualitative and quantitative point of view.
3. **Study properties of the sparse model**, such as directional selectivity, stability with respect to translations, rotations, deformation, etc.
4. Identify ways of optimizing the network.

Related work

Wavelet scattering networks [**Bruna2013**; **Oyallon2017**; **Zarka2020**]: CNN-like cascading wavelet convolutions.

$=$ **Structure CNNs** into well-defined math. operators and **study invariances**.

$\neq$ Wavelet scattering networks are **built from scratch**. Our approach aims at studying **existing architectures**.

Proposed models

- Models based on **AlexNet** and **ResNet34**.
- First conv. layer replaced by **dual-tree wavelet packets**.

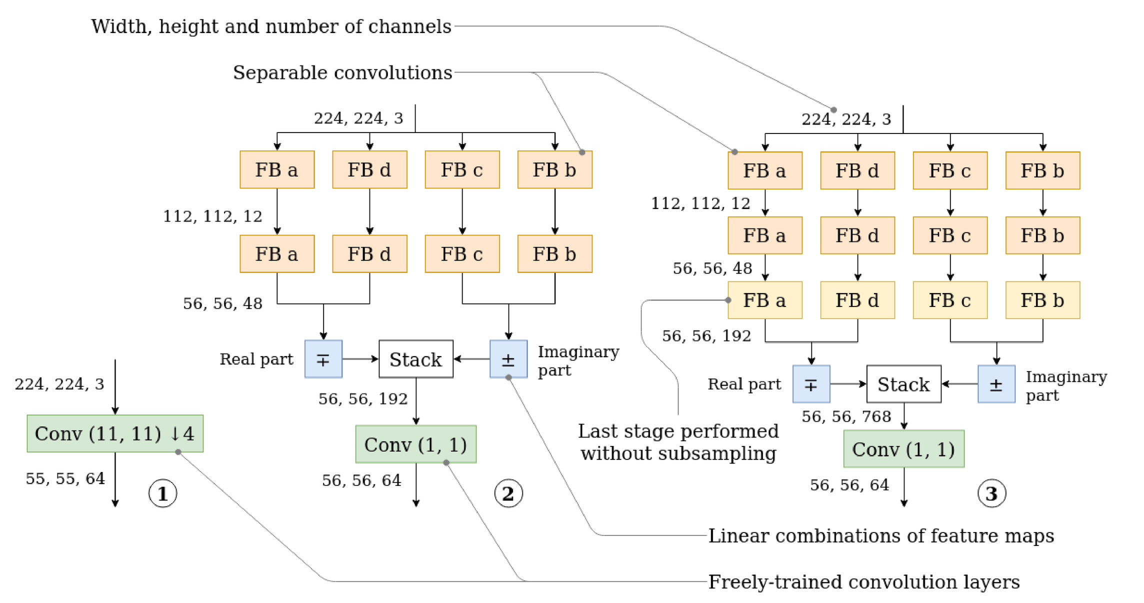

Figure 2. ① AlexNet's first layer. ②③ DT-ℂWPT modules replacing AlexNet's first layer, with 2 and 3 decomposition stages, respectively. The real and imaginary parts of complex wavelet packet coefficients are stored in separate channels.

Predictive power

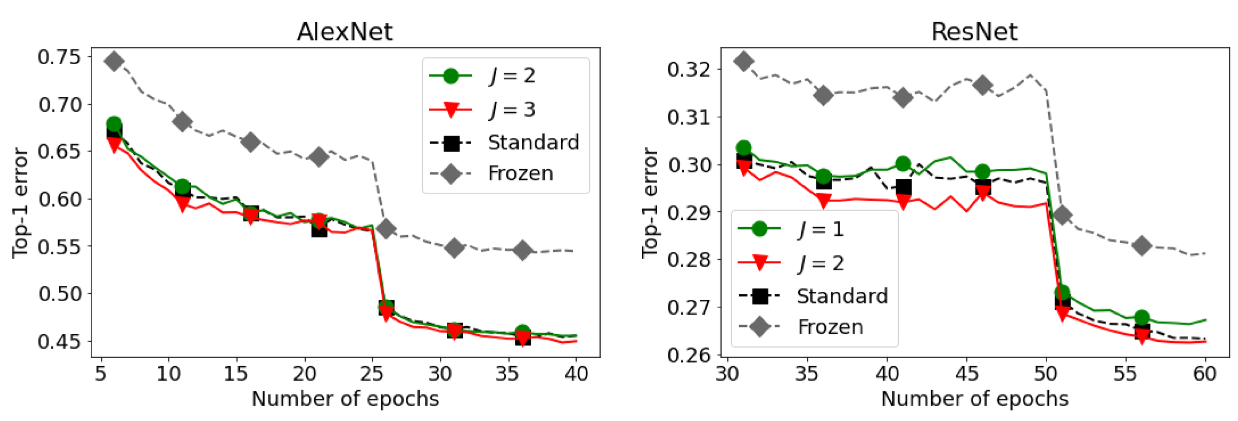

Figure 3. Validation error for AlexNet (left) and ResNet34 (right) along training with ImageNet. J = number of decomposition stages. Dashed gray curves → standard architecture with frozen first layer.

Kernel similarity

Models trained on **ImageNet ILSVRC2012**.

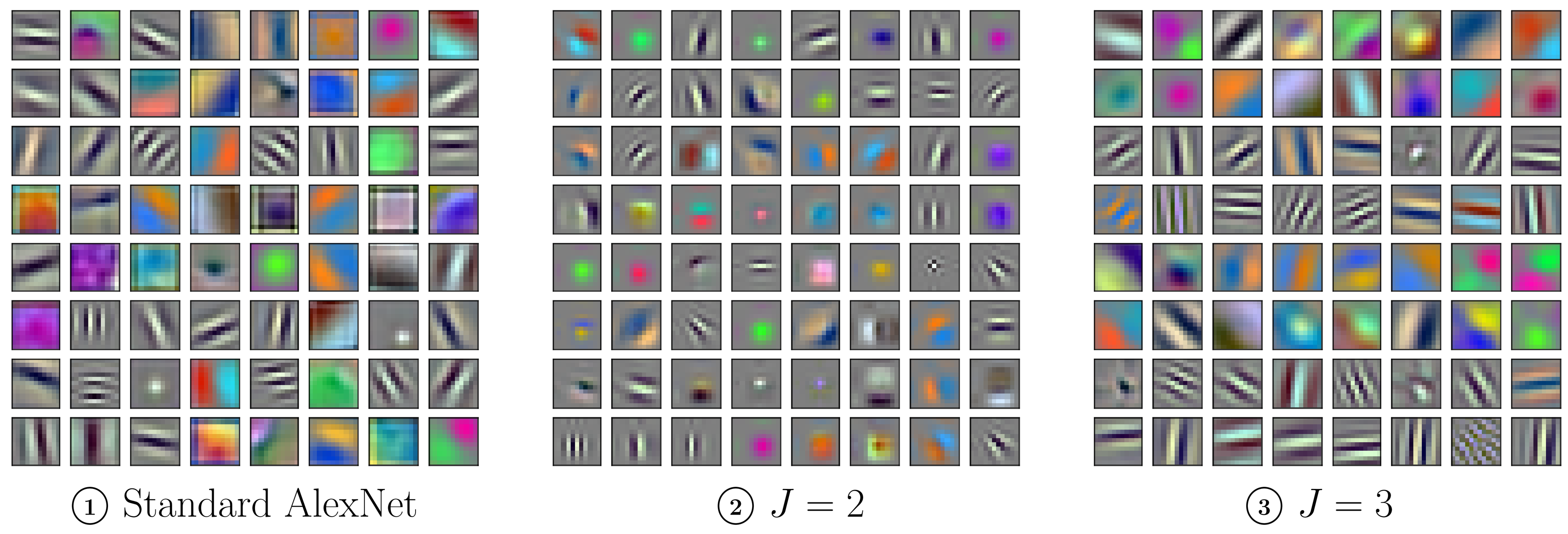

Figure 4. Resulting kernels of DT-ℂWPT AlexNet, compared to the standard architecture. J = number of decomposition stages. Filters are cropped to 11×11 to match the original size of AlexNet's kernels.

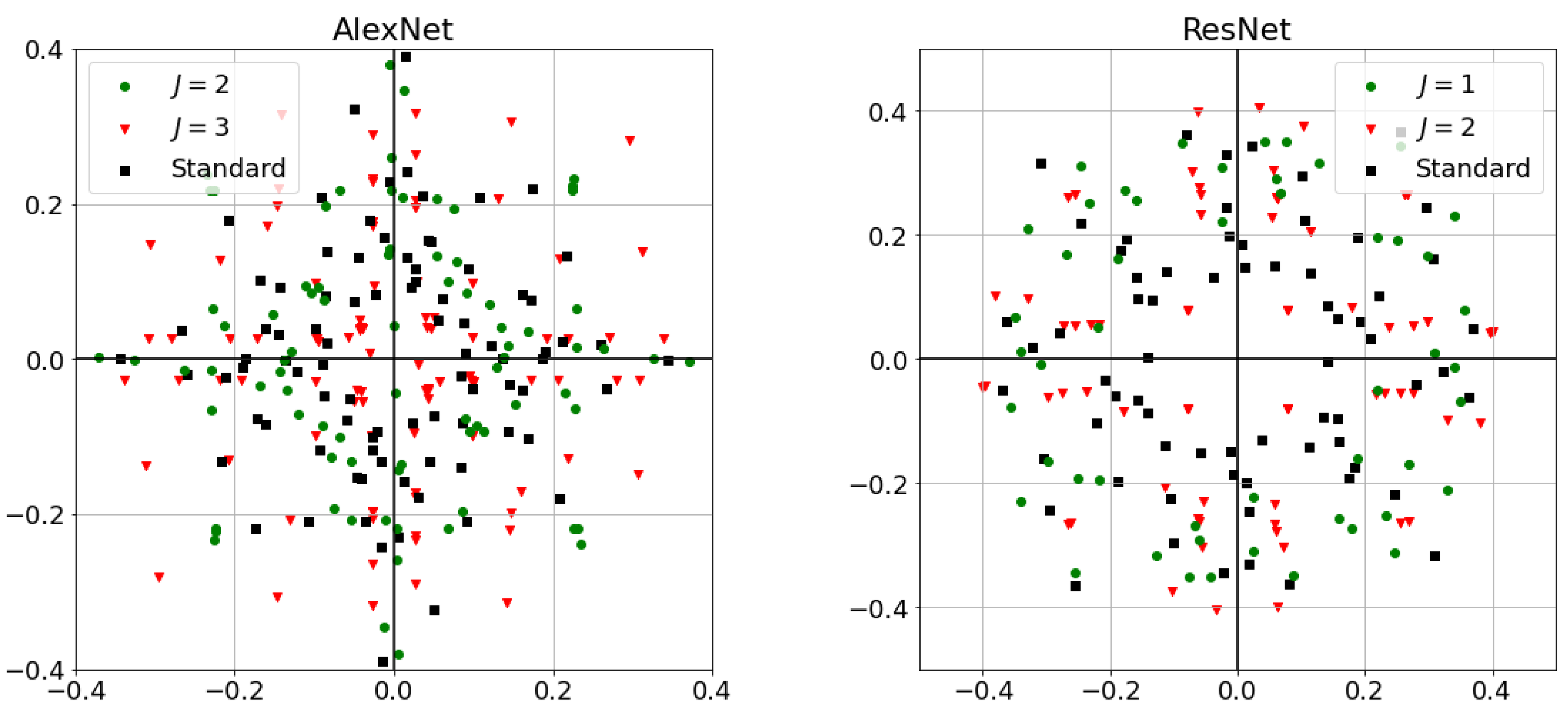

Figure 5. Characteristic frequencies of DT-ℂWPT-based kernels, compared to standard architectures.

$\Longrightarrow$ **The resulting kernels cover the same frequency area as standard AlexNet when $J = 3$ ($J = 2$ for Resnet34).**

Future work

- Establish near-equivalence between the output of **max pooling** layers in CNNs and the **modulus** of complex wavelet packet coefficients (inspired by [**Waldspurger2015**]).
- Perform a theoretical and empirical study of **various types of invariants** (shifts, rotations, deformations).
- Further increase sparsity of the models.
- Perform a quantitative evaluation of kernel similarity.
- Focus research on **deeper layers**.

References and acknowledgments

Published as a workshop paper. **Leterme2021**

Joint work with Kevin Polisano, Valérie Perrier, Karteek Alahari. This work was partially supported by the LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01) funded by the French program Investissement d'avenir as well as the ANR grant AVENUE.