Sparsifying Convolutional Layers with Dual-Tree Wavelet Packets
Hubert Leterme, hubert.letterme@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 feature extraction, signal compression and denoising; Oscillating patterns very often observed in CNN kernels [Yosinski2014].

Objective
✓ Theoretical and empirical study 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]. ➞ 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

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

Predictive power

Kernel similarity
Models trained on ImageNet ILSVRC2012

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.