Learning Convolutional Feature Hierarchies for Visual Recognition

Koray Kavukcuoglu\(^+\), Pierre Sermanet\(^+\), Y-Lan Boureau\(^+\), Karol Gregor\(^+\), Michael Mathieu\(^+\), Yann LeCun\(^+\)

\(^+\)Courant Institute, NYU, \(*\)INRIA/Willow Project Team

Convolutional Predictive Sparse Decomposition

\[
\frac{1}{2} \| x - \sum_k D_k * z_k \|_2^2 + \beta \sum_k \| z_k - f(W^k * x) \|_2^2 + \lambda \| z \|_1
\]

- **Convolutional training** yields a more **diverse set of features**
- **Feed-forward Predictor** real-time vision applications

Patch based PSD

Convolutional PSD
Image Recognition Architecture

Convolutional Predictive Sparse Decomposition

\[
\frac{1}{2} \|x - \sum_k D_k \ast z_k\|_2^2 + \beta \sum_k \|z_k - f(W^k \ast x)\|_2^2 + \lambda |z|_1
\]

Unsupervised Pre-Training

Unsupervised Pre-Training

Filter Bank

Non-Linearity

Pooling

Filter Bank

Non-Linearity

Pooling

Supervised Refinement

Convolutions

Rectification

Contrast Normalization

Pooling / Subsampling

Input Image
Unsupervised pre-training with Convolutional PSD yields better accuracy than patch-based PSD.
Pedestrian Detection On INRIA Dataset

- Purely supervised training: 14.8% miss rate
- Unsupervised pre-training with Conv PSD + supervised refinement: 11.5%
- Close to state of the art and improving quickly...