The Unreasonable Effectiveness of Patches in Deep Convolutional Kernels Methods

Louis Thiry[1] , Michael Arbel[2], Eugene Belilovsky[3], Edouard Oyallon[4]

[1]Departement of Computer Science, École Normale Supérieure, CNRS, PSL [2]Gatsby Computational Neuroscience Unit, UCL

[3]Concordia University and Mila Montreal [4] LIP6, Sorbonne Université, CNRS

Introduction

- Recent works present competitive convolutional kernel methods, obtaining $87-90\%$ accuracy on CIFAR-10.
- They are data-driven, share an implicit ingredient: data **whitening**.
- We present very simple convolutional kernel method using this ingredient and K-nearest-neighbors encoding
- We obtain comparable accuracies on CIFAR-10 with linear / 1-hidden-layer classifier.
- We scale this method on ImageNet and outperform existing non-learned visual representations.

Data-driven convolutional kernel methods

$$K_{k,\Phi,\mathcal{X}}(x,y) = k(\Phi L x, \Phi L y)$$

- Shift and rescale (e.g. whitening) operator
  $$L$$
- Training data
  $$\mathcal{X}$$
- Representation
  $$\Phi$$
- Predefined (e.g. Linear, Gaussian, Neural Tangent) kernel
  $$k(x,y)$$

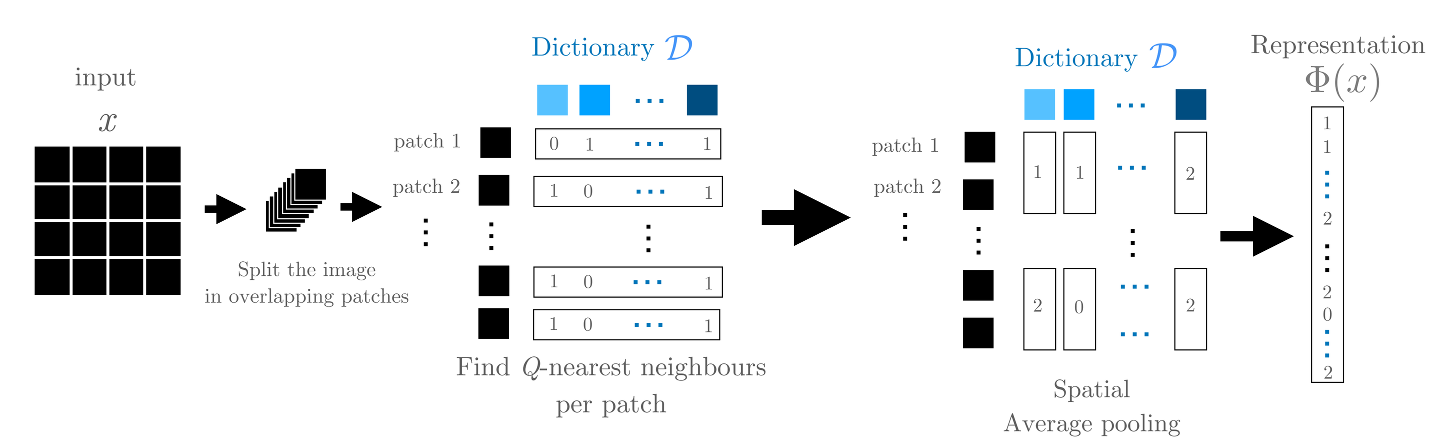
> $K(x,y)$ is **data-driven** if Φ or L depend on the training set $\mathcal{X}$, **data-independent** otherwise.

Examples of Data-driven kernels

- Random features (Coates et al. 2011, Recht et al. 2019)
- Convolutional kernel networks (Mairal 2016)
- Enhanced convolutional neural tangent kernels (Li at al. 2019)
- Neural Kernels Without Tangents (Shankar et al. 2020)

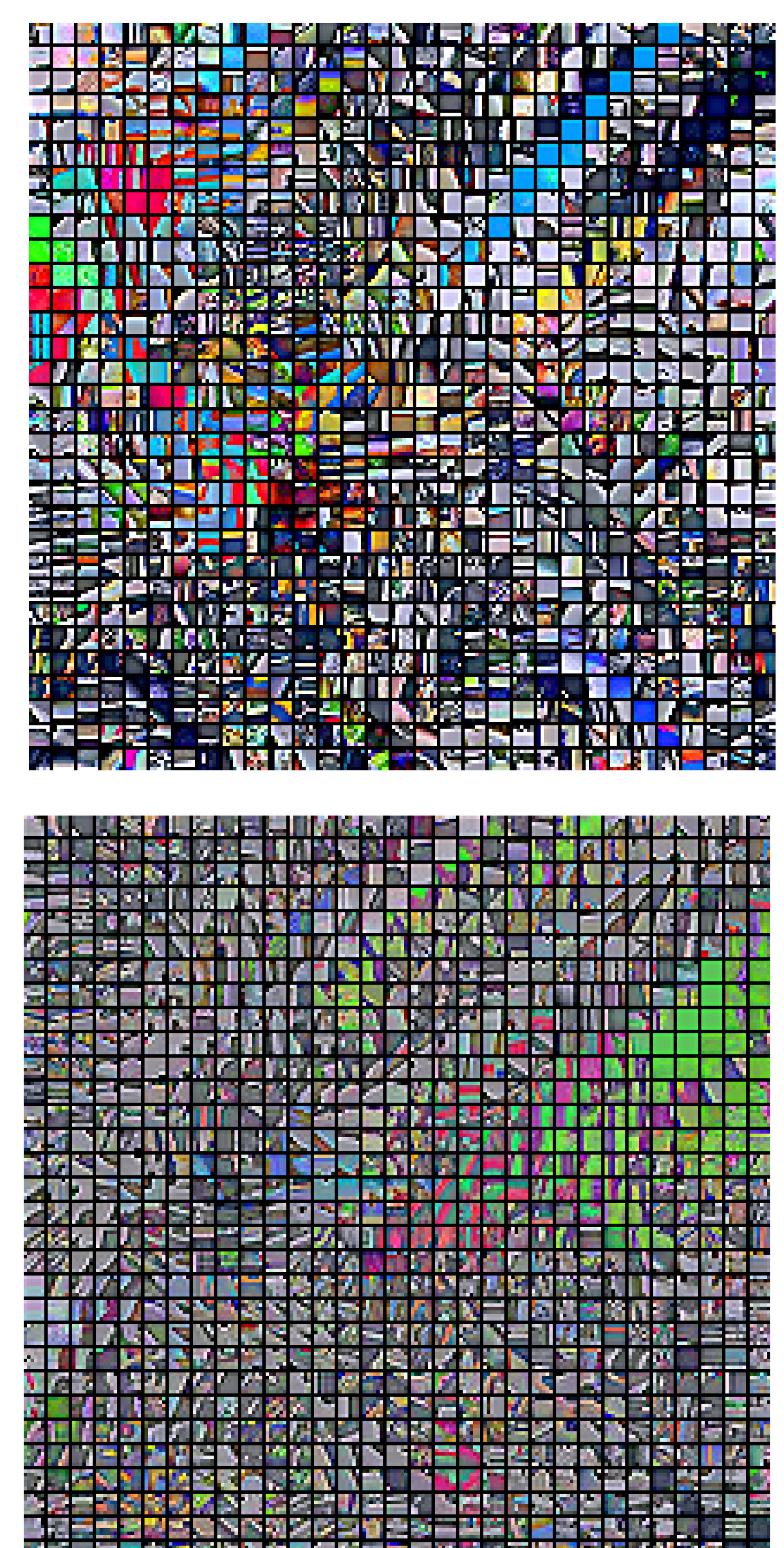
Our method

Figure 1:Our classification pipeline described synthetically.



- x: image viewed as a collection of overlapping patches.
- L: whitening operator
  $$L : x \mapsto (\Sigma + \lambda I)^{-1}(x - \mu)$$
- Φ: K-nearest-neighbor encoding in a dictionary $\mathcal{D}$ of randomly selected whitened patches.
- $k(x,y)$: linear kernel.

Figure 2:Examples of whitened dictionary $\mathcal{D}$ with patch size $P = 6$ from ImageNet-64 (Top) and CIFAR-10 (Bottom).



Results

Linear classification on CIFAR-10

Method	$\lvert\mathcal{D}\rvert$	VQ	Online	P	Acc.
Coates et al. (2011)	$1k$	✓	×	6	68.6
Wavelets (Oyallon et al. 2015)	-	×	×	8	82.2
Recht et al. (2019)	$0.2M$	×	×	6	85.6
SimplePatch (Ours)	$10k$	✓	✓	6	85.6
SimplePatch (Ours)	$60k$	×	✓	6	**86.9**

Non-linear classification on CIFAR-10

Method	VQ	Depth	Classifier	Acc.
SimplePatch (Ours)	✓	2	1-hidden-layer	88.5
AlexNet (Krizhevsky et al. 2012)	×	5	e2e	89.1
NK (Shankar et al. 2020)	×	5	kernel	89.8
CKN (Mairal et al. 2016)	×	9	kernel	89.8

Linear classification on ImageNet

Method	$\lvert\mathcal{D}\rvert$	VQ	P	Depth	Resolution	Top1	Top5
Random CNN (Arand. et al. 2017)	-	×	-	9	224	18.9	-
Wavelets (Zarka et al. 2019)	-	×	32	2	224	26.1	44.7
SimplePatch (Ours)	$2k$	✓	12	1	128	35.9	57.4
SimplePatch (Ours)	$2k$	×	12	1	128	36.0	**57.6**

Non-linear classification on ImageNet

Method	VQ	P	Depth	Resolution	Classifier	Top1	Top5
Greedy (Belilovsky et al. 2018)	×	-	2	224	e2e	-	44
SimplePatch (Ours)	✓	6	2	64	1-layer	39.4	62.1
BagNet (Brendel et al. 2019)	×	9	50	224	e2e	-	70.0

Figure 3:CIFAR-10 ablation study, train accuracies in blue, test accuracies in red.