Journées SMAI MODE 2018

Group Invariance, Stability to Deformations, and Complexity of Deep Convolutional Representations

Alberto BIETTI
Inria

Julien MAIRAL
Inria

Résumé. We study deep signal representations that are invariant to groups of transformations and stable to the action of diffeomorphisms without losing signal information. This is achieved by generalizing the multilayer kernel construction introduced in the context of convolutional kernel networks and by studying the geometry of the corresponding reproducing kernel Hilbert space. We show that the signal representation is stable, and that models from this functional space, such as a large class of convolutional neural networks with homogeneous activation functions, may enjoy the same stability. In particular, we study the norm of such models, which acts as a measure of complexity, controlling both stability and generalization.
This work was published at the NIPS 2017 conference [3], and a longer version is available on arxiv [2].

Mots-clefs : machine learning, kernel methods, convolutional neural networks, signal processing, learning theory.

The results achieved by deep neural networks for prediction tasks have been impressive in domains where data is structured and available in large amounts. In particular, convolutional neural networks (CNNs) [5] have shown to model well the local appearance of natural images at multiple scales, while also representing images with some invariance through pooling operations. Yet, the exact nature of this invariance and the characteristics of functional spaces where convolutional neural networks live are poorly understood; overall, these models are sometimes seen as clever engineering black boxes that have been designed with a lot of insight collected since they were introduced.

Invariance and stability to deformations. Understanding the geometry of these functional spaces is nevertheless a fundamental question. In addition to potentially bringing new intuition about the success of deep networks, it may for instance help solving the issue of regularization, by providing ways to control the variations of prediction functions in a principled manner. Small deformations of natural signals often preserve their main characteristics, such as the class label in a classification task (*e.g.*, the same digit with different handwritings may correspond to the same images up to small deformations), and provide a much richer class of transformations than translations. Representations that are stable to small deformations allow more robust models that may exploit these invariances, which may lead to improved sample complexity.
The scattering transform [4, 8] is a recent attempt to characterize convolutional multilayer architectures based on wavelets. The theory provides an elegant characterization of invariance

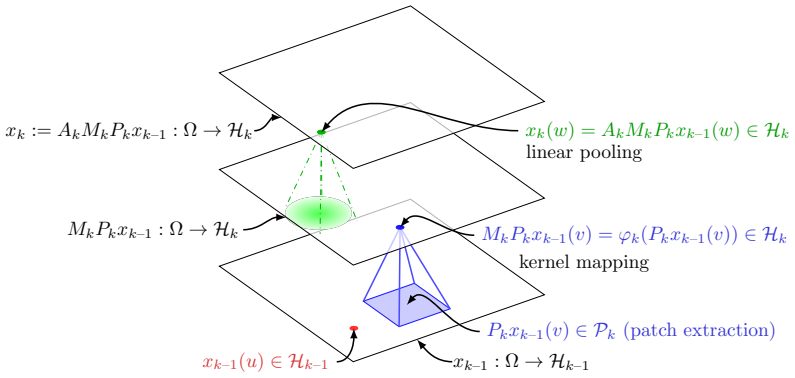

Figure 1: Construction of the k-th signal representation from the k–1-th one. Note that while Ω is depicted as a box in $\mathbb{R}^2$ here, our construction is supported on $\Omega = \mathbb{R}^d$.

and stability properties of signals represented via the scattering operator, through a notion of Lipschitz stability to the action of diffeomorphisms, which can be formalized as follows: given a signal $x \in L^2(\Omega, \mathbb{R}^p)$ on the continuous domain $\Omega = \mathbb{R}^d$, a representation $\Phi(x)$ is stable to the action of diffeomorphisms if

$$\|\Phi(L_\tau x) - \Phi(x)\| \leq (C_1 \|\nabla \tau\|_\infty + C_2 \|\tau\|_\infty) \|x\|, \tag{1}$$

where $\tau : \Omega \to \Omega$ is a C^1-diffeomorphism, $L_\tau x(u) := x(u - \tau(u))$ is the action operator of the diffeomorphism, and C_1, C_2 are constants which control deformation stability and translation invariance, respectively. Nevertheless, the scattering networks of [4, 8] do not involve "learning" in the classical sense since the filters of the networks are pre-defined, and the resulting architecture differs significantly from the most used ones.

Multilayer convolutional kernel representations. In this work, we study such theoretical properties for more standard convolutional architectures from the point of view of positive definite kernels [9]. Specifically, we consider a functional space derived from a kernel for multi-dimensional signals, which admits a multilayer and convolutional structure that generalizes the construction of convolutional kernel networks (CKNs) [6, 7]. Specifically, the multilayer kernel representation is obtained through a series of intermediate "feature maps" $x_k \in L^2(\Omega, \mathcal{H}_k)$ constructed by successive application of operators, namely *patch extraction* (P_k), *kernel mapping* (M_k) and (Gaussian) *pooling* (A_k) operators, as shown in Figure 1. The final representation is then given by

$$\Phi(x) := A_n M_n P_n A_{n-1} M_{n-1} P_{n-1} \cdots A_1 M_1 P_1 x \ \in \ L^2(\Omega, \mathcal{H}_n). \tag{2}$$

We study the *signal preservation* properties of such representations, showing that each layer can be sampled at intervals smaller than the patch size with no loss of information.
The main motivation for introducing a kernel framework is to study separately data representation and predictive models. In particular, by defining a kernel on signals of the form $\mathcal{K}(x, x') := \langle \Phi(x), \Phi(x') \rangle$, functions f in the corresponding RKHS $\mathcal{H}_\mathcal{K}$ satisfy $|f(x) - f(x')| \leq \|f\|_{\mathcal{H}_\mathcal{K}} \|\Phi(x) - \Phi(x')\|$. On the one hand, we study the invariance and stability properties of the kernel representation $\Phi(x)$, obtaining similar guarantees as the scattering transform [8].

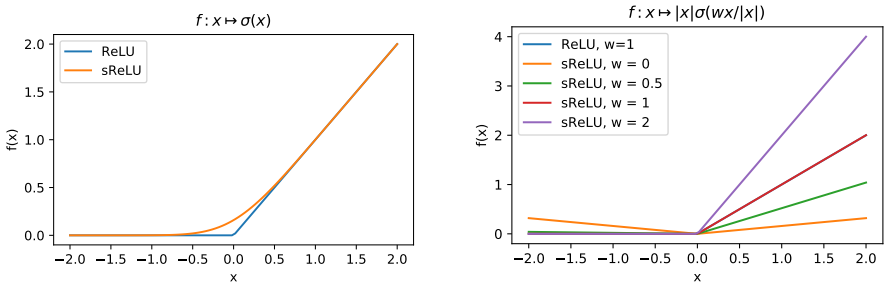

Figure 2: Comparison of one-dimensional functions obtained with relu and smoothed relu (sReLU) activations. (Left) non-homogeneous setting of [10]. (Right) our homogeneous setting, for different values of the parameter w. Note that for $w \geq 0.5$, sReLU and ReLU are indistinguishable.

On the other hand, we show that these stability results can be translated to predictive models by controlling their norm in the functional space, or simply the norm of the last layer in the case of CKNs [6]. In particular, the RKHS norm controls both stability and generalization, so that stability may lead to improved sample complexity.

Stability of the kernel representation $\Phi(x)$. We show that the representation defined in (2) satisfies the following stability relation. If $\|\nabla\tau\|_\infty \leq 1/2$,

$$\|\Phi(L_\tau x) - \Phi(x)\| \leq \left(C_1 \left(1+n\right) \|\nabla\tau\|_\infty + \frac{C_2}{\sigma_n} \|\tau\|_\infty \right) \|x\|, \tag{3}$$

where σ_n is the scale of the last pooling operator (which typically grows exponentially with n), and C_1 grows with the relative patch sizes at layer k compared to the current resolution given by the scale σ_{k-1} of the previous pooling operator, which justifies the frequent use of small patches (*e.g.*, 3x3) in many common CNN architectures for computer vision. When the kernel is appropriately designed, we also show how to obtain signal representations that are invariant to the action of any locally compact group of transformations [2].

Model complexity of CNNs. When the kernel mapping operators M_k are constructed using a specific type of kernels (namely, homogeneous dot-product kernels, see [2, 6]), we show that the space $\mathcal{H}_\mathcal{K}$ contains generic CNNs with certain types of activation functions, such as a homogeneous version of the smooth ReLU, shown in Figure 2. For such models, we show that the RKHS norm can be bounded in terms of the spectral norms of the convolutional operations, quantities which have recently been shown to be important in understanding generalization properties of neural networks (*e.g.*, [1]).
In addition to controlling stability, the RKHS norm is known to control generalization behavior (see, *e.g.*, the notion of margin bounds for SVMs [9]). This implies, for instance, that generalization is harder if the task requires classifying two slightly deformed images with different labels, since this requires a function with large RKHS norm according to our stability analysis. In contrast, if a stable function (*i.e.*, with small RKHS norm) is sufficient to do well on a training set, learning becomes "easier" and few samples may be enough for good generalization.

Références

[1] P. Bartlett, D. J. Foster, and M. Telgarsky. Spectrally-normalized margin bounds for neural networks. In *Advances in Neural Information Processing Systems (NIPS)*, 2017.

[2] A. Bietti and J. Mairal. Group invariance, stability to deformations, and complexity of deep convolutional representations. *preprint arXiv:1706.03078*, 2017.

[3] A. Bietti and J. Mairal. Invariance and stability of deep convolutional representations. In *Advances in Neural Information Processing Systems (NIPS)*, 2017.

[4] J. Bruna and S. Mallat. Invariant scattering convolution networks. *IEEE Transactions on pattern analysis and machine intelligence (PAMI)*, 35(8):1872–1886, 2013.

[5] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989.

[6] J. Mairal. End-to-End Kernel Learning with Supervised Convolutional Kernel Networks. In *Advances in Neural Information Processing Systems (NIPS)*, 2016.

[7] J. Mairal, P. Koniusz, Z. Harchaoui, and C. Schmid. Convolutional kernel networks. In *Advances in Neural Information Processing Systems (NIPS)*, 2014.

[8] S. Mallat. Group invariant scattering. *Communications on Pure and Applied Mathematics*, 65(10):1331–1398, 2012.

[9] B. Schölkopf and A. J. Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. 2001.

[10] Y. Zhang, P. Liang, and M. J. Wainwright. Convexified convolutional neural networks. In *International Conference on Machine Learning (ICML)*, 2017.