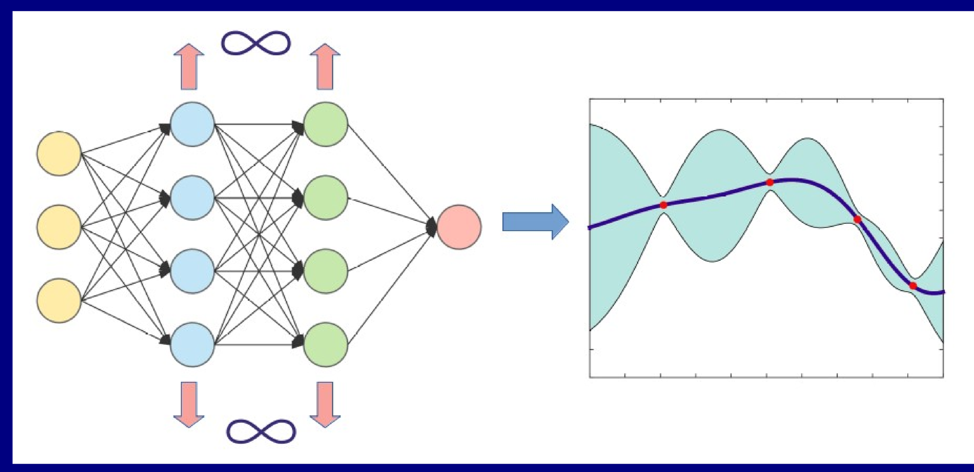
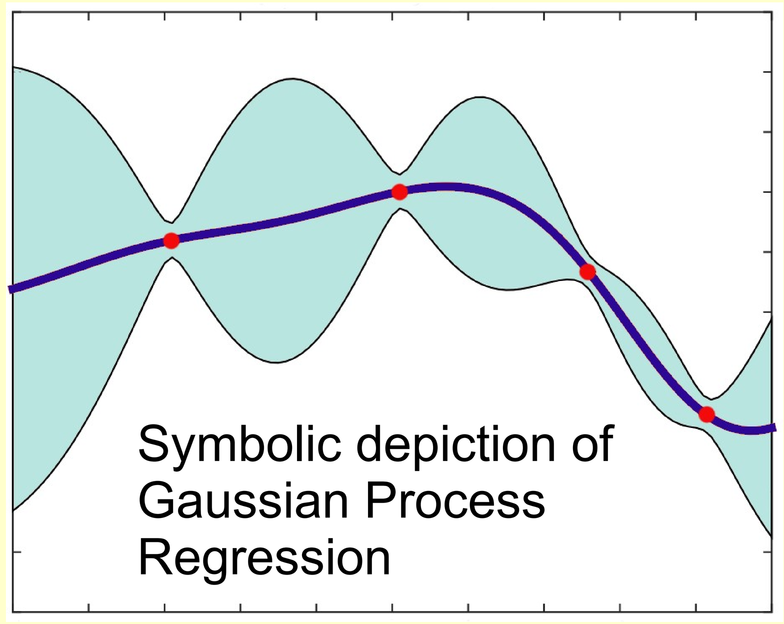
Interrelation of equivariant Gaussian processes and convolutional neural networks



A. Demichev

Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University, Moscow, Russia

The Problem Context



- establishing relationships between various methods of ML ⇨ a better theoretical understanding of these methods and their improvements
- a correspondence has recently been established between the appropriate asymptotics of deep neural networks (DNNs), including convolutional ones (CNNs), and the ML method based on Gaussian processes (GPs)
  - Gaussian processes are mathematically similar to free (Euclidean) quantum field theory (QFT) ⇒ potential for using a broad range of QFT methods for analyzing DNNs

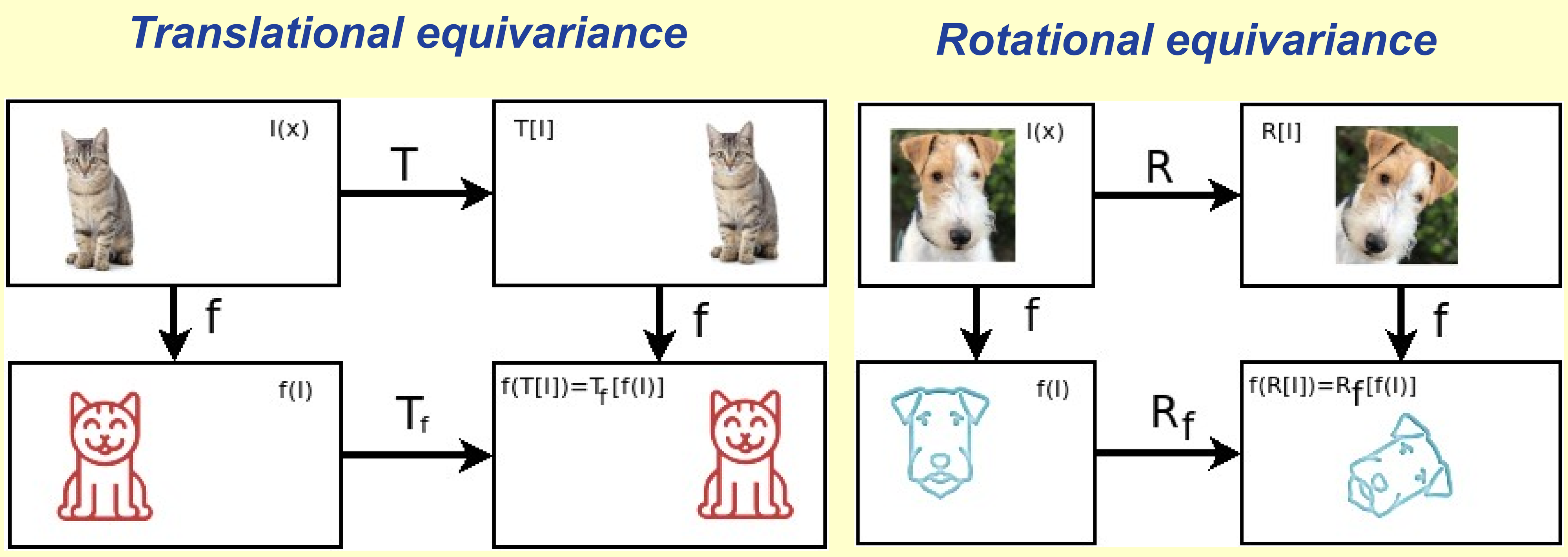
Posing the Problem and Main Result

An important feature of CNNs is their equivariance (consistency) with respect to the symmetry transformations of the input data

- In this work, we have established a relationship between
  - the many-channel limit of equivariant CNNs and the corresponding equivariant Gaussian processes (GPs)
    - hence the QFT with the appropriate symmetry
- The approach used provides explicit equivariance at each stage of the derivation of the relationship

NNs ⟺ GPs (*no equivariance issues*)

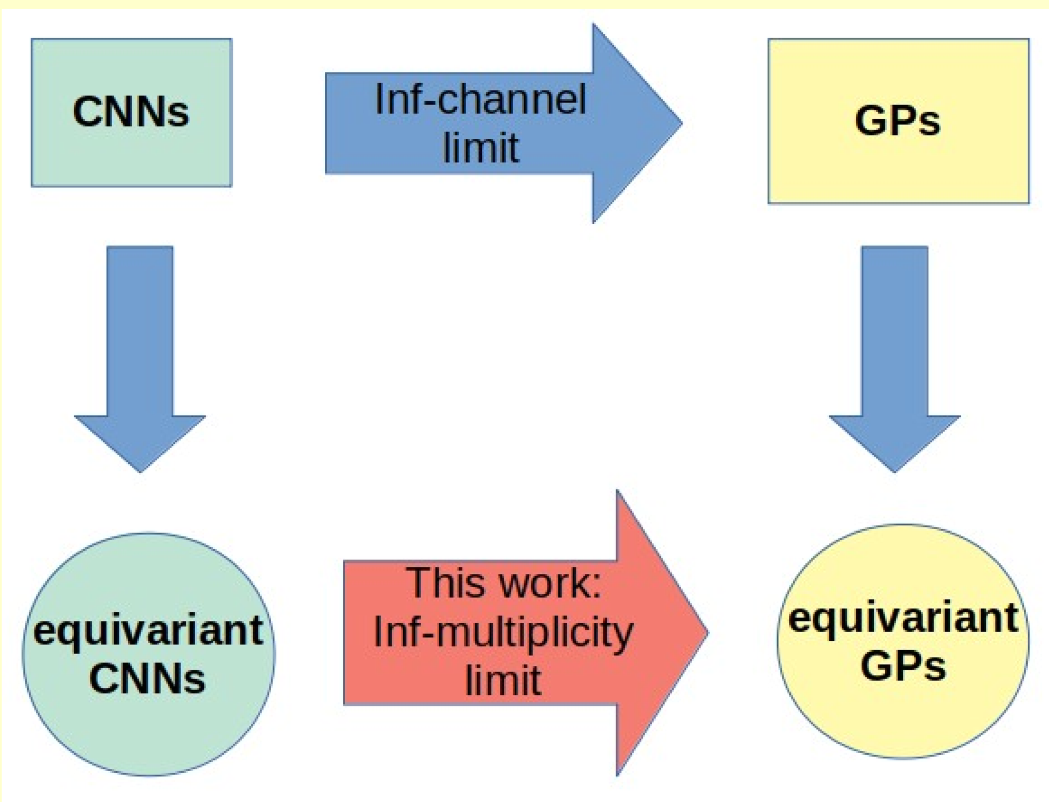
- *R.M.Neal* (1996,2012): the function defined by a single-layer fully-connected NN + ∞-wide + i.i.d. zero-mean weights and biases as network prior is equivalent to a GP
- *J.Lee et al* (2018), *A.G.Matthews et al* (2018): extended these results to arbitrarily deep FCNN with ∞-many hidden units in each layer
- *R.Novak et al* (2018), *A.Garriga-Alonso et al* (2018): if each hidden layer has an infinite number of convolutional filters (= infinite number of channels), the CNN prior is equivalent to a GP



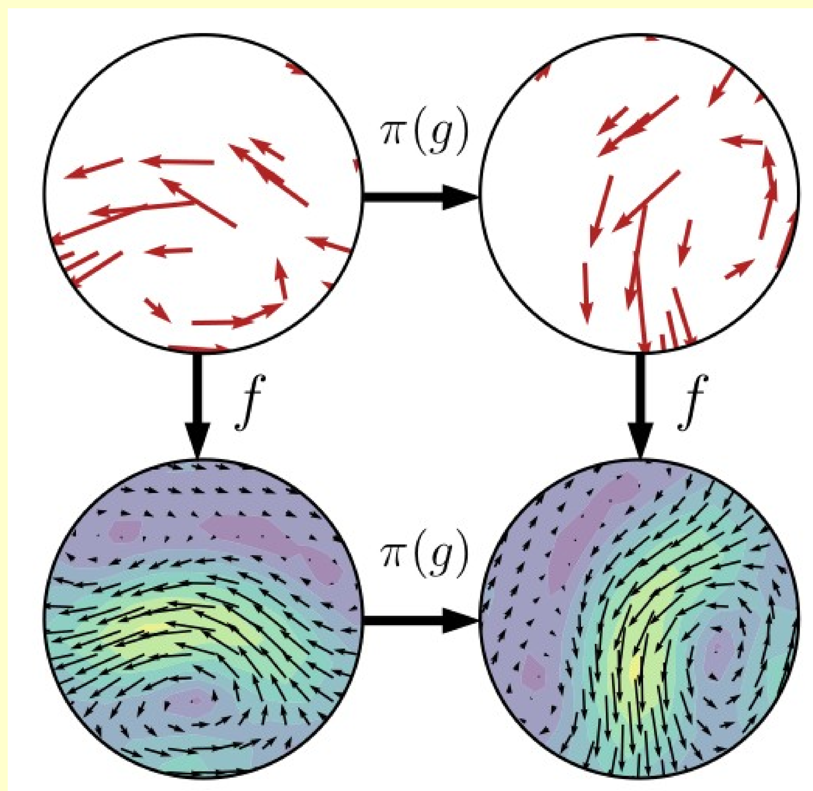
Equivariance in CNNs

- Well-known fact: usual CNNs are translational equivariant
- Recent years: huge activity to extend this to other symmetries
  - *Kondor, Trivedi, Cohen, Welling, Esteves, Ravanbakhsh,..*

Equivariant CNN with Infinite Number of Channels = Equivariant GPs



- All the preceding seminal works on the CNN-GP relationship did not take into account equivariance
- On the other hand, there exists investigations of equivariant GPs (e.g., *P.Holderrieth et al* (2020)) but without established relations with CNNs in the appropriate limit
- The present work is intended to fill the gap between equivariance of CNNs and that of the corresponding GPs
  - the method constituents are
    - layer-by-layer derivation of GP covariances in the many-channel limit by using the law of large numbers that results in the recursive relation for the top-layer covariance
    - keeping explicit equivariance at each step of the derivation
- The main question in our work is how to deal with vector-valued functions
  - the point is that such vectors (of finite dimensionality) are also treated as channels, so the question is how one can go to the infinite-channel limit
  - our solution is based on using the so called steerable CNNs (*T.Cohen & M.Welling* (2016)) which in turn heavily use induced representations of symmetry groups
  - all-in-all this allows us to separate channels indices in two categories:
    - the indices that numerate the vector components within an irrep and used to describe their transformations under matrix representations of a symmetry group;
    - the indices that numerate different irreducible representations (of the same or different types
  - the 2d type of the indices are not restricted and can be used for the limiting transition to the corresponding GP



The figure is borrowed from P.Holderrieth et al (2020)

Conclusion

- Currently there exists rather promising new trend in ML based on the relationship between FCNN/CNNs and GPs
  - many related subtopics, *e.g.*, signal propagation in NNs, learning curve, QFT methods in ML
- In this work we have derived the many-channel limit for CNNs with symmetry on Euclidean plane (translations+rotations)
  - with explicit equivariance at each step of the derivation
  - calculated the corresponding equivariant GP kernel in the case of specific nonlinearities
- thereby filled the gap between many-channel equivariant CNNs and independently introduced equivariant GPs

This work was carried out in the framework of R&D State Assignment No.115041410196