Bayesian Image Classification with Deep Convolutional GPs

Vincent Dutordoir[1], Mark van der Wilk[2], Artem Artemev[1] and James Hensman[3]

[1] PROWLER.io, [2]Imperial College London, [3]Amazon
(Work completed while MvdW and JH were affiliated to PROWLER.io)

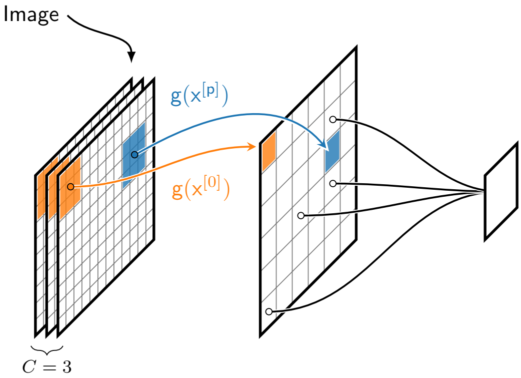
International Conference on Artificial Intelligence and Statistics – 2020

CNN classification misses are confidently wrong

TICK-GP misses are better calibrated

Convolutional Gaussian processes

[van der Wilk et al., 2017]



Define patch-response function:

$$g(\mathbf{x}^{[p]}) \sim \mathcal{GP}(0, k_g(\mathbf{x}^{[p]}, \mathbf{x}'^{[p']}))$$

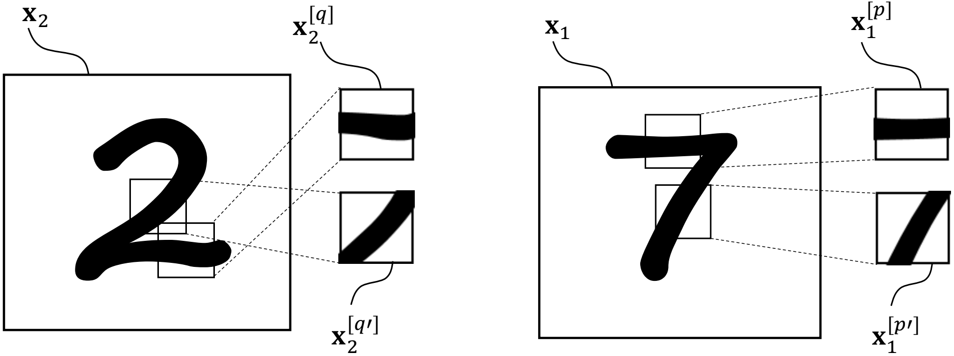
Define the image-response function as:

$$f(\mathbf{x}) = \sum_p g(\mathbf{x}^{[p]})$$

Linear relation between f and g gives

$$f(\mathbf{x}) \sim \mathcal{GP}\left(0, \sum_p \sum_{p'} k_g(\mathbf{x}^{[p]}, \mathbf{x}'^{[p']})\right)$$

Translation Invariance — a limitation of purely convolutional models



Pure convolutional structure: $f(\mathbf{x}) = \sum_p g(\mathbf{x}^{[p]})$

Convolution and weighted sum: $f(\mathbf{x}) = \sum_p w_p g(\mathbf{x}^{[p]})$

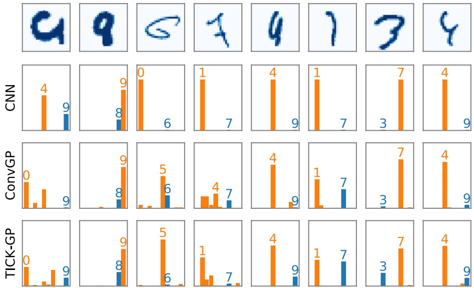
Translation Insensitivity

Insensitively: relaxation of invariance

$$k_g\left((\mathbf{x}^{[p]}, p), (\mathbf{x}^{[p']}, p')\right) = k_{\text{patch}}(\mathbf{x}^{[p]}, \mathbf{x}^{[p']}) \times k_{\text{loc}}(p, p').$$

Experiments

Classification: Shallow GP vs CNN



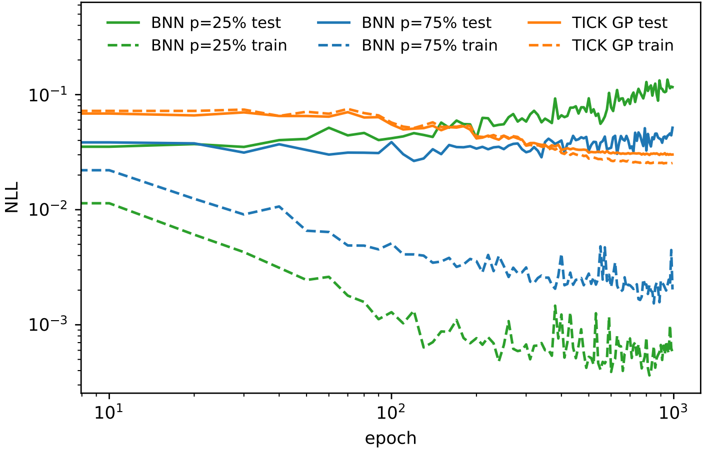
metric	Conv-GP	CNN	TICK GP
Error	1.70	**0.81**	0.83
NLPD	0.057	0.030	**0.029**
NLPD misclassified	1.97	12.52	**1.70**

$$\text{NLPD} = -\frac{1}{N^*}\sum_i \log p(y_i^* \mid x_i^*)$$

more similar experiments on MNIST, Fashion-MNIST and CIFAR-10 in the paper.

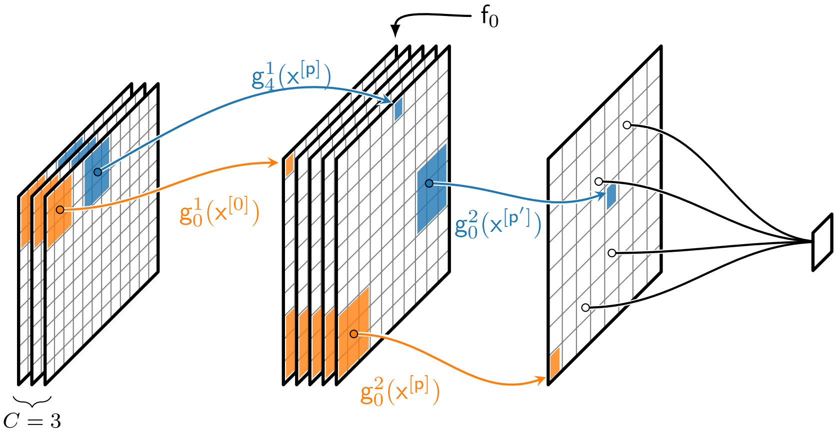
Model selection

Dropout BNN vs. GPs



The ELBO is the correct objective for automated model selection

Deep Convolutional Gaussian processes



The hidden layers of a Deep Conv GP are multi-output GPs

$$\mathbf{f}_0(\mathbf{x}) = [g_0(\mathbf{x}^{[p]})]_p \quad \text{with} \quad \mathrm{Cov}\Big([\mathbf{f}_0(\mathbf{x})]_p, [\mathbf{f}_0(\mathbf{x})]_{p'}\Big) = k_{g_0}(\mathbf{x}^{[p]}, \mathbf{x}^{[p']})$$

.

Experiment

		MNIST		CIFAR-10	
depth	metric	Conv	TICK	Conv	TICK
1	Error (%)	1.87	**1.19**	41.06	**37.10**
	NLPD	0.06	**0.04**	1.17	**1.08**
	neg. ELBO ($\times 10^3$)	8.29	**5.83**	65.72	**63.51**
2	Error (%)	0.96	**0.67**	28.60	**25.59**
	NLPD	0.04	**0.02**	0.84	**0.75**
	neg. ELBO ($\times 10^3$)	5.37	**4.25**	52.81	**48.31**
3	Error (%)	0.93	**0.64**	25.33	**23.83**
	NLPD	0.03	**0.02**	0.74	**0.69**
	neg. ELBO ($\times 10^3$)	5.045	**4.19**	49.38	**47.53**

Software

https://www.gpflow.org/

A Framework for
Interdomain and Multioutput Gaussian Processes

Mark van der Wilk[1] Vincent Dutordoir[2] ST John[2] Artem Artemev[2]
Vincent Adam[2] James Hensman[2]

https://arxiv.org/abs/2003.01115.

Final remarks & Conclusion

- The importance of thinking about the modelling assumptions made in the prior.
- DCGP are extremely promising in terms of uncertainty estimation and in the use of marginal likelihood approximations for hyperparameter learning.
- DCGP as correlated multi-output GPs enables efficient implementation in our general-purpose open-sourced framework (part of `https://www.gpflow.org/`).

Reach out to have a chat if you want to know more!