An Alternative to Neural Methods for Classification

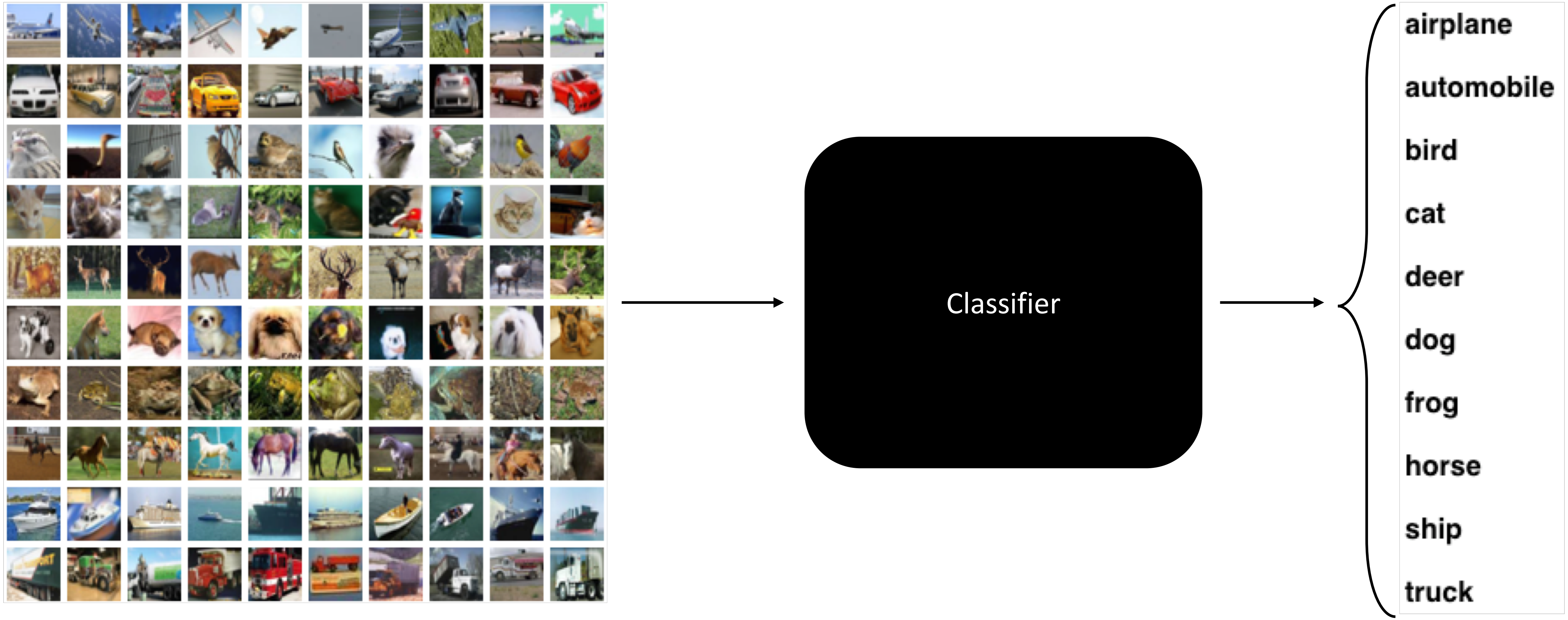
Presented by: Paul Hein

The University of Arizona

Machine Learning for Artificial Intelligence

Interdisciplinary Visual Intelligence Lab
http://ivilab.org

An Example Problem



Deep convolutional Gaussian processes

Kenneth Blomqvist Samuel Kaski Markus Heinonen
Department of Computer Science, Aalto university
Helsinki Institute for Information Technology HIIT

October 9, 2018

Abstract

We propose deep convolutional Gaussian processes, a deep Gaussian process architecture with convolutional structure. The model is a principled Bayesian framework for detecting hierarchical combinations of local features for image classification. We demonstrate greatly improved image classification performance compared to current Gaussian process approaches on the MNIST and CIFAR-10 datasets. In particular, we improve CIFAR-10 accuracy by over 10 percentage points.

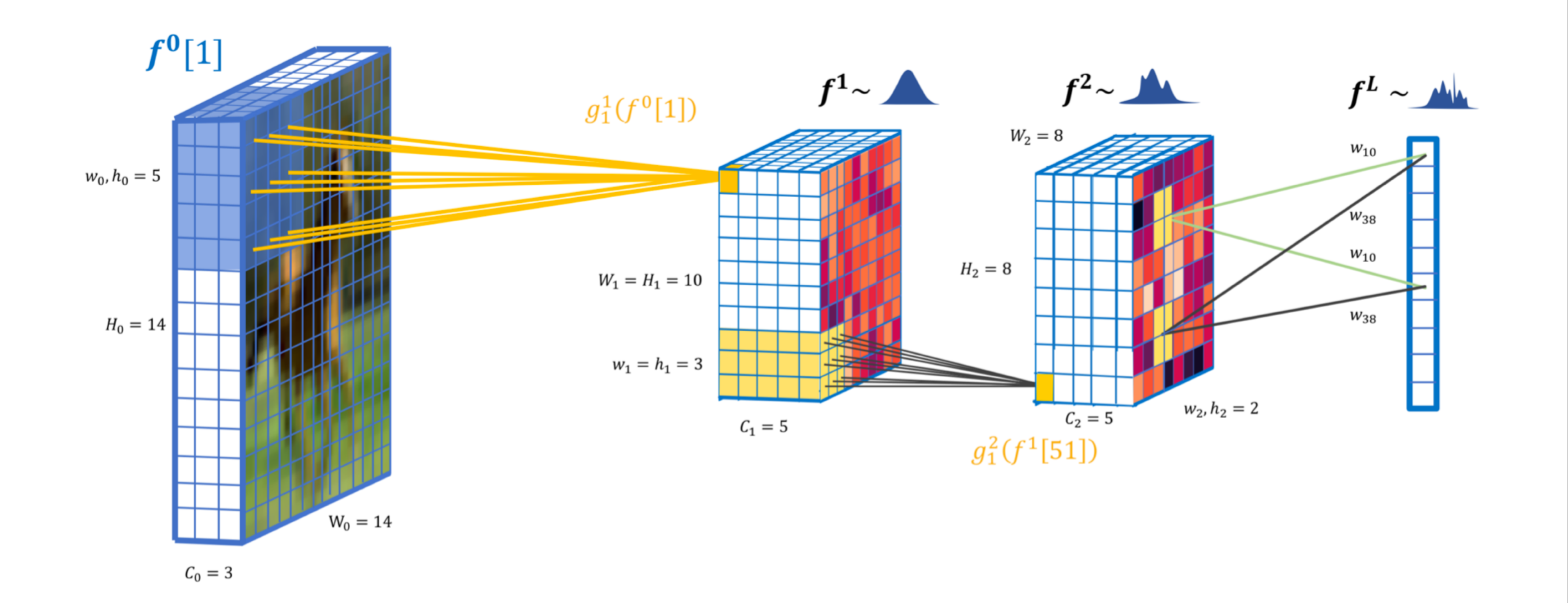
networks can often leverage a large number of training data to counteract this problem. Developing methods that are better regularized and can incorporate prior knowledge would allow us to deploy machine learning methods in domains where massive amounts of data is not available. Conventional neural networks do not provide reliable uncertainty estimates on predictions, which are important in many real world applications.

The deterministic CNN's have been extended into the probabilistic domain with weight uncertainties (Blundell et al., 2015). Gal and Ghahramani (2016) explored the Bayesian

Desirable qualities for the ML classifier

- Accuracy: The classifier should be able to accurately classify new data
- Interpretability: output should be diagnosable (especially to determine bias in the classifier)
- Trainability: The time/resources required to train a model should not be too great
- Transferability: good results on one dataset should be indicative of good results on a different dataset
- Robustness: should not require too much parameter tuning to work well on a new problem

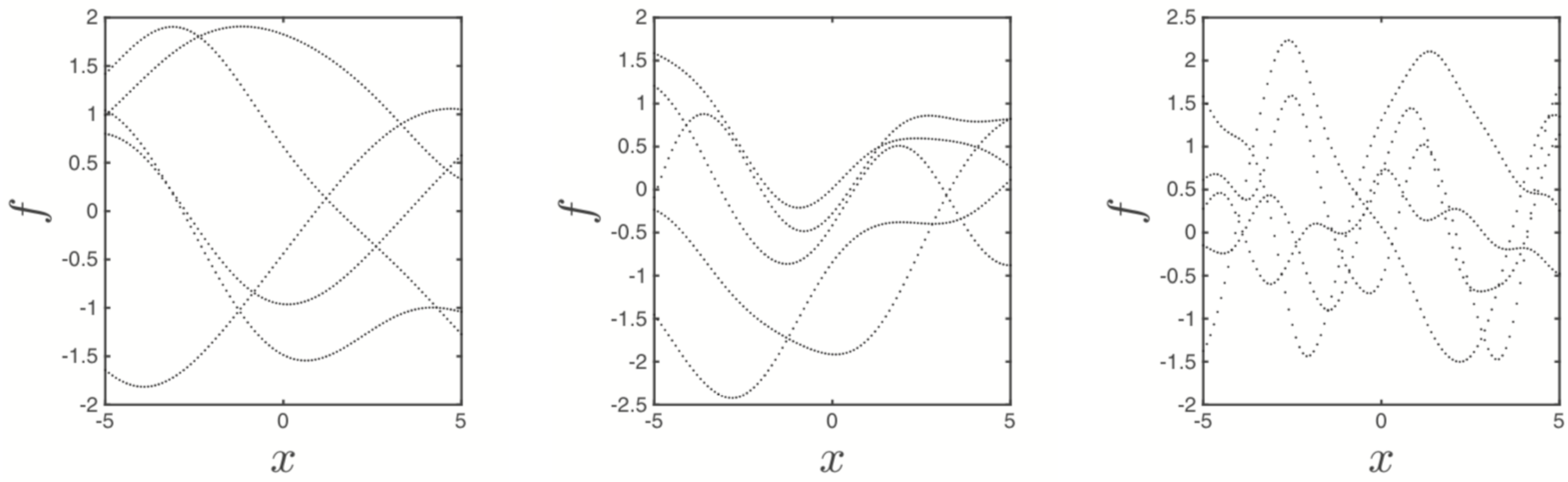
Classification pipeline



Presentation Road-map

- Background questions
  - What even is a Gaussian Process?
  - What is a convolutional layer?
- Deep Convolution GP
  - Paper implementation
  - Comparative results
  - Issues with the paper
  - Future extensions

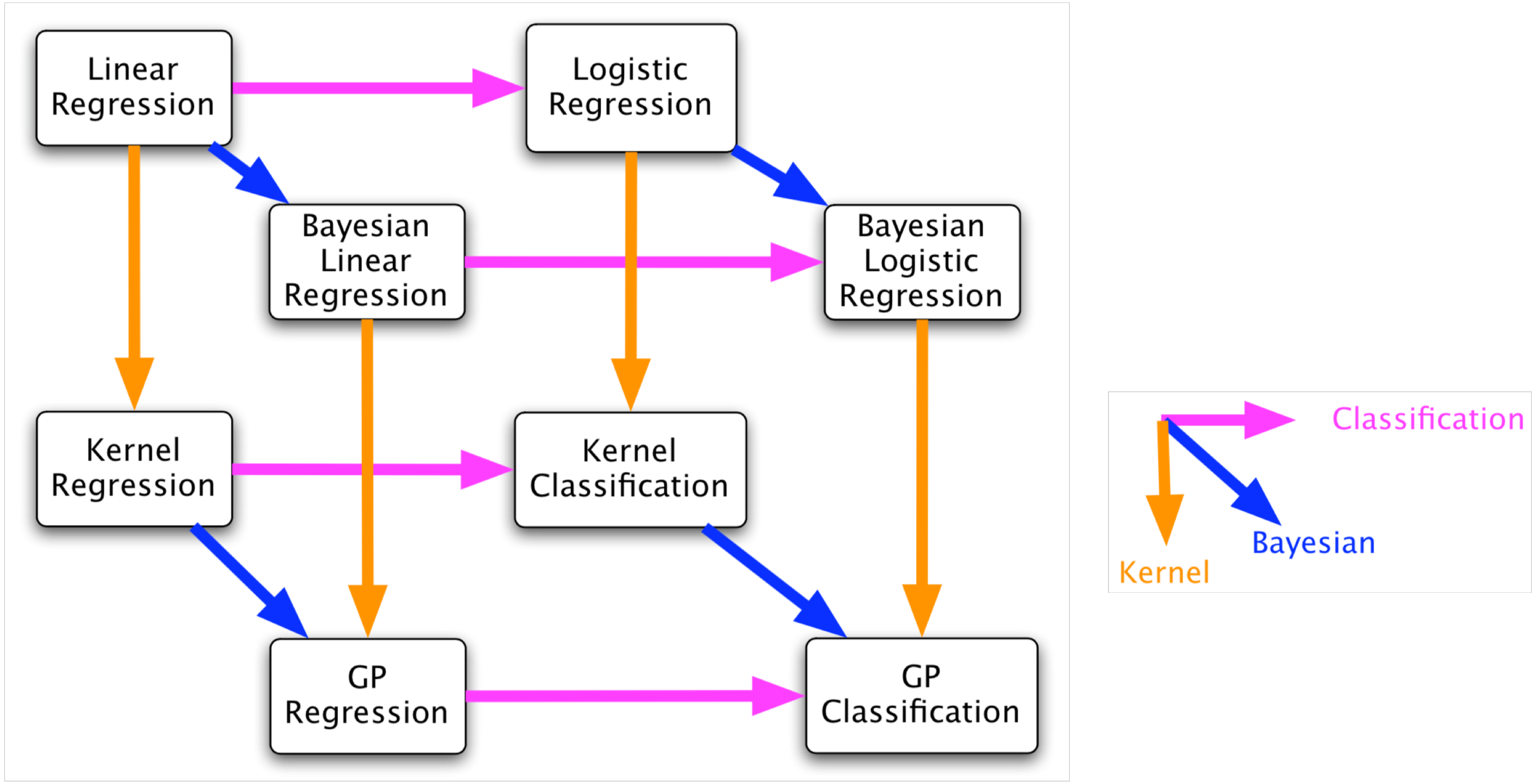
What is a Gaussian Process?



(a) $\gamma = 0.05$ (b) $\gamma = 0.1$ (c) $\gamma = 0.5$

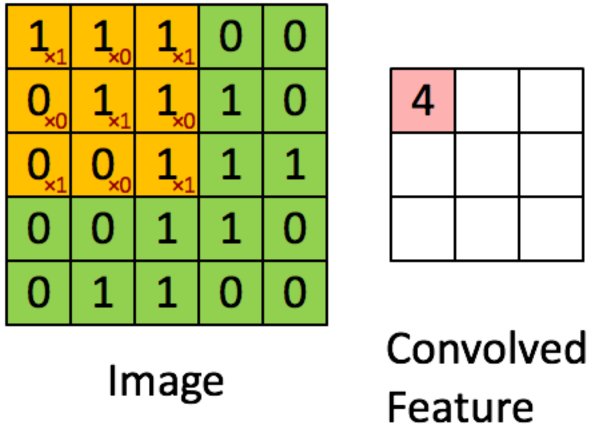
Materials and insight taken from FCML

From Linear Regression to GP



What is a Convolutional Filter?

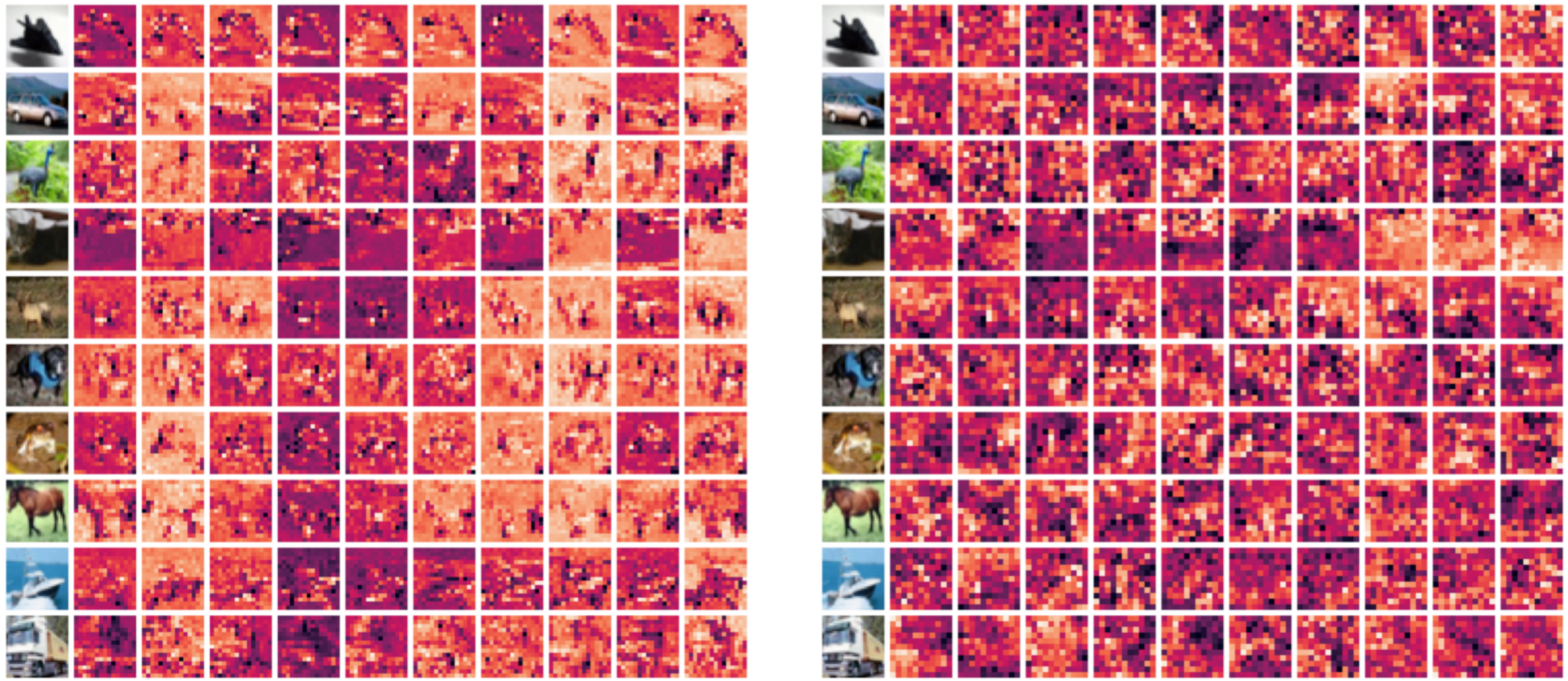
- Common image filters
  - Sharpener
  - Edge detector
- Multiple response channels
- Differing stride sizes



Deep Convolutional GP Implementation

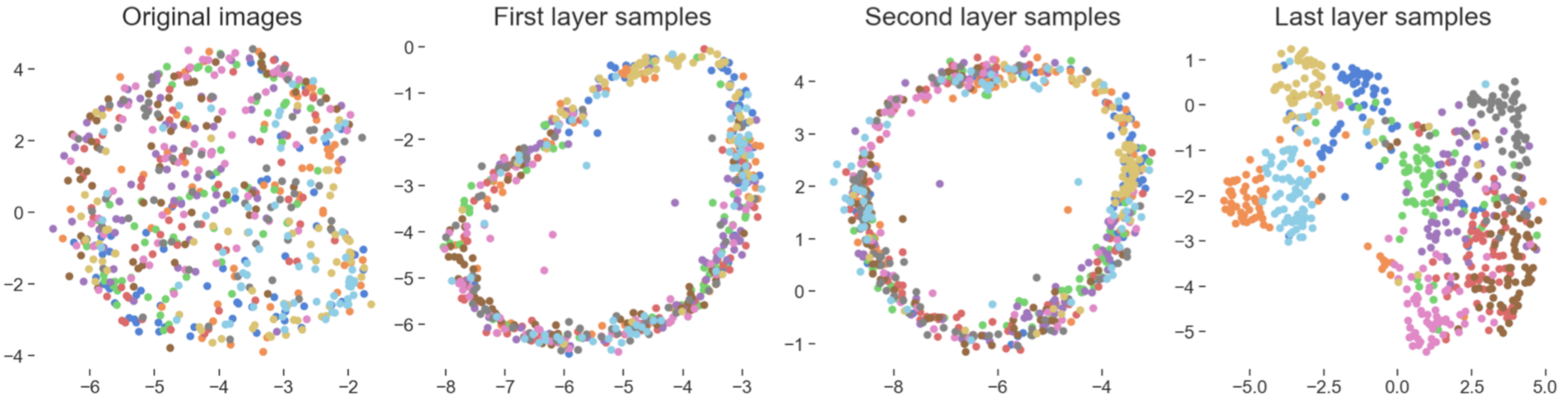
- Response channels
- GP kernel: Radial Basis Function (RBF)
- Doubly stochastic variational inference
- Parameter initialization

A visual look at the response patches

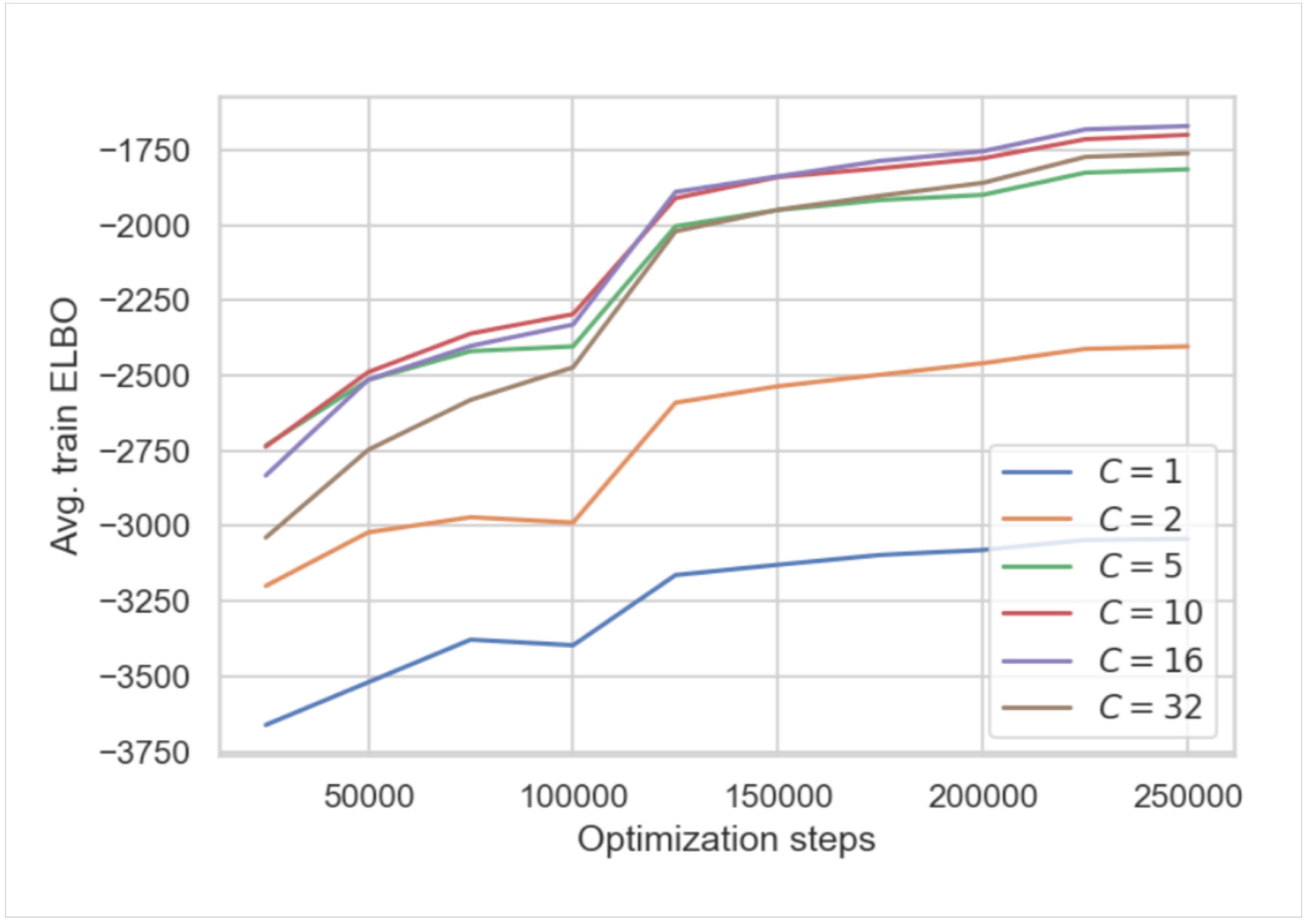


(a) Samples from the first layer. (b) Samples from the second layer.

Classification clusters at each layer



How many response channels?



Deep Conv GP Results

Gaussian process models	Layers	Inducing points	Test accuracy MNIST	CIFAR-10	Reference
RBF AutoGP	1	200	$98.29^{(*)}$	$55.05^{(*)}$	Krauth et al. (2017)
Multi-channel conv GP	1	1000	$98.83^{(*)}$	$64.6^{(*)}$	Van der Wilk et al. (2017)
DeepCGP	1	384	98.38	58.65	current work
DeepCGP	2	2×384	99.24	73.85	"
DeepCGP	3	3×384	**99.44**	**75.89**	"
Neural network models	Layers	# params			
Deep kernel learning	5	2.3M .. 4.6M	$99.2^{(*)}$	$77.0^{(*)}$	Wilson et al. (2016a)
DenseNet	250	15.3M	N/A	$94.81^{(*)}$	Huang et al. (2017)

Some faults from the paper

- No comparison to similar depth CNN
- Why does the paper perform better on MNIST than CIFAR-10?
  - Variational inference is reliant on the gaussian assumption
  - MNIST is a much simpler problem
- Degenerate covariance problem

Future Research

Parameter Adjustment

- Stride length adjustment
- Layer depth

Efficiency

- SG Ham MC
- Kernel interpolation

Experimental

- Application-based dataset
- Sparse dataset