Conditional Neural Processes

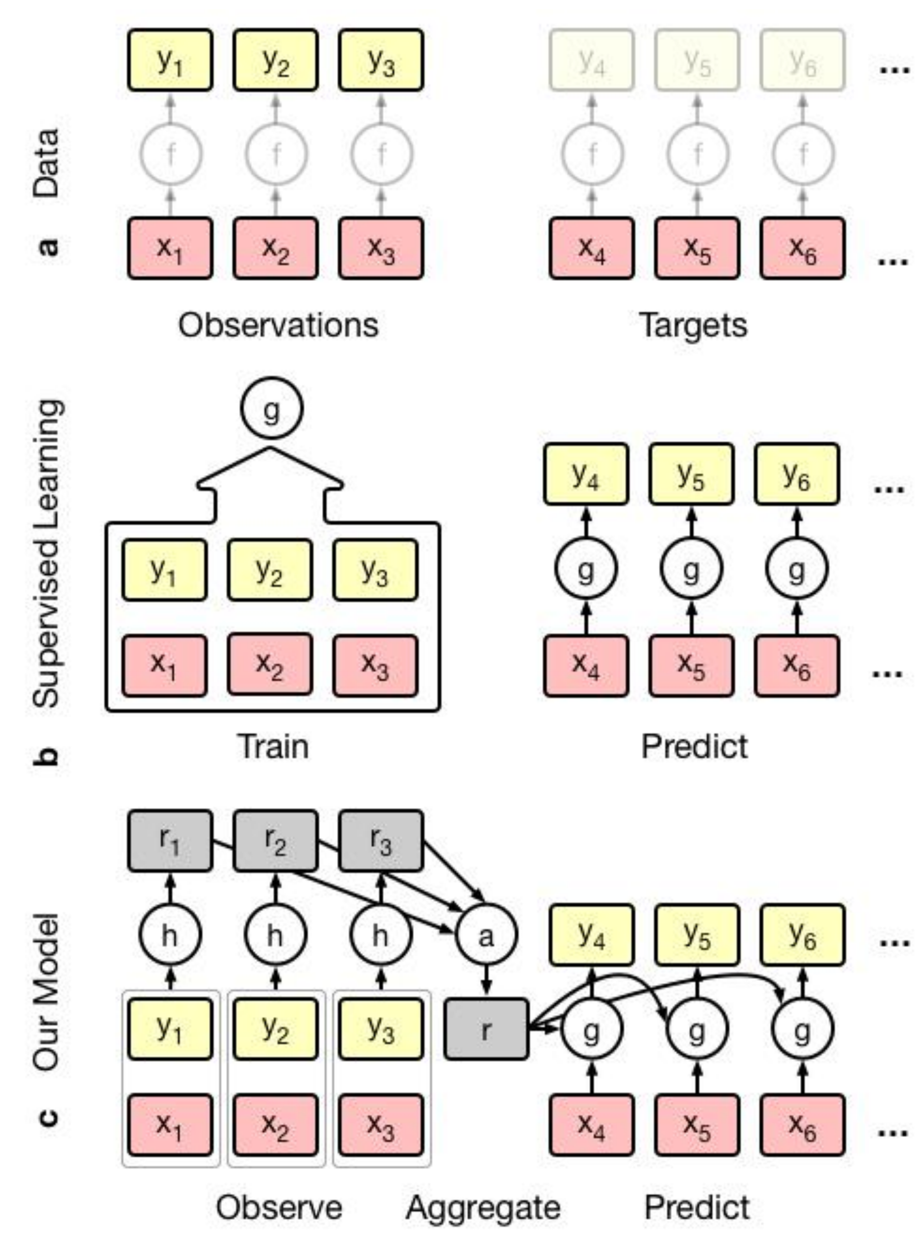
Marta Garnelo, Dan Rosenbaum, Christopher Maddison, Tiago Ramalho, David Saxton, Murray Shanahan, Yee Whye Teh, Danilo J. Rezende, S. M. Ali Eslami

Introduction

	NN	GP	CNP
Can fit more than one function at test time	×	✔	✔
Computationally cheap at test time	✔	×	✔
Can learn prior knowledge from data	✔	×	✔

Conditional Neural Processes (CNPs) combine benefits of NNs and GPs:

- the **flexibility** of stochastic processes such as GPs
- structured as neural networks and trained via **gradient descent** from data directly.



Model

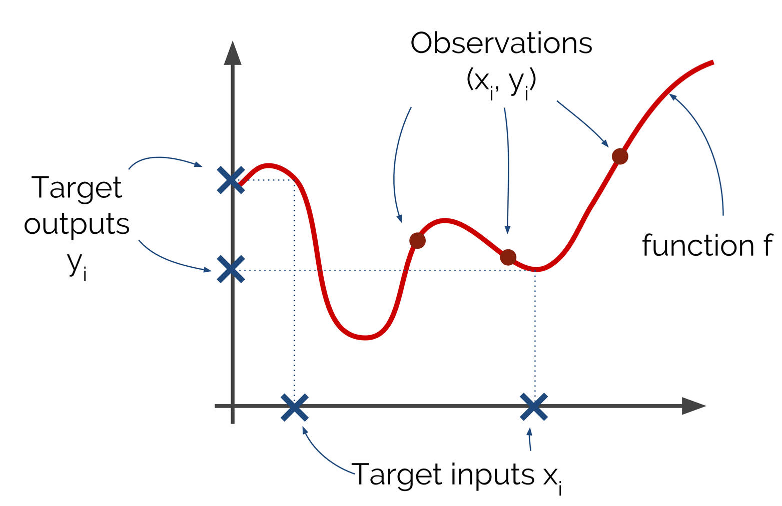
We have a **function f(x_i) = y_i** with **inputs x_i** and **outputs y_i**.

f is drawn from **P**, a distribution over functions.

We define the sets:
- **Observations**: O = {(x_i, y_i)}
- **Targets**: T = {x_i}

GOAL: Given some observations we want to be able to make predictions at unseen target inputs at test time.

An example for this is 1-D regression



The architecture of our model captures this task:

$$r_i = h_\theta(x_i, y_i) \qquad \forall (x_i, y_i) \in O$$
$$r = r_1 \oplus r_2 \oplus \ldots r_{n-1} \oplus r_n$$
$$\phi_i = g_\theta(x_i, r) \qquad \forall (x_i) \in T$$

where $\mathbf{r_i}$ are the **representations** of the (x_i,y_i)-pairs and **r** the overall representation obtained by summing all r_i (see intro figure).

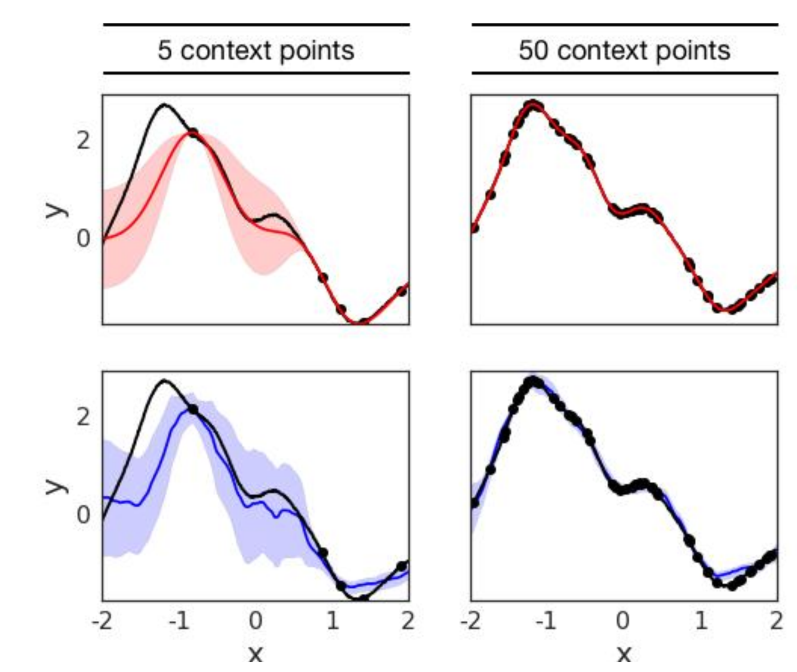
$\mathbf{h_\theta}$ and $\mathbf{g_\theta}$ are implemented as neural networks.

$\boldsymbol{\phi_\theta}$ parametrizes the **output distribution** which in this work is either a Gaussian or a categorical distribution.

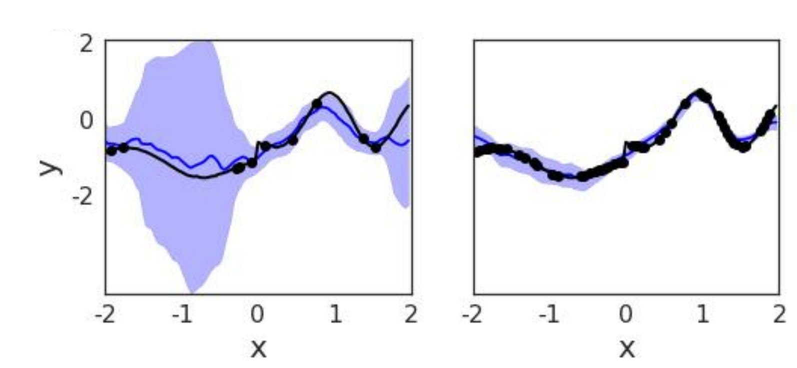
Key properties of this model:

1. CNPs are conditional distributions over functions trained to model the empirical conditional distributions of functions f ~ P.
2. CNPs are permutation invariant in O and T.
3. CNPs are scalable, achieving a running time complexity of O(n+m) for making *m* predictions with *n* observations.

1D Regression



Predicted mean and variance on a 1-D curve (black line) using 5 (left) and 50 (right) context points (black dots). Red: Gaussian Process, blue: CNP



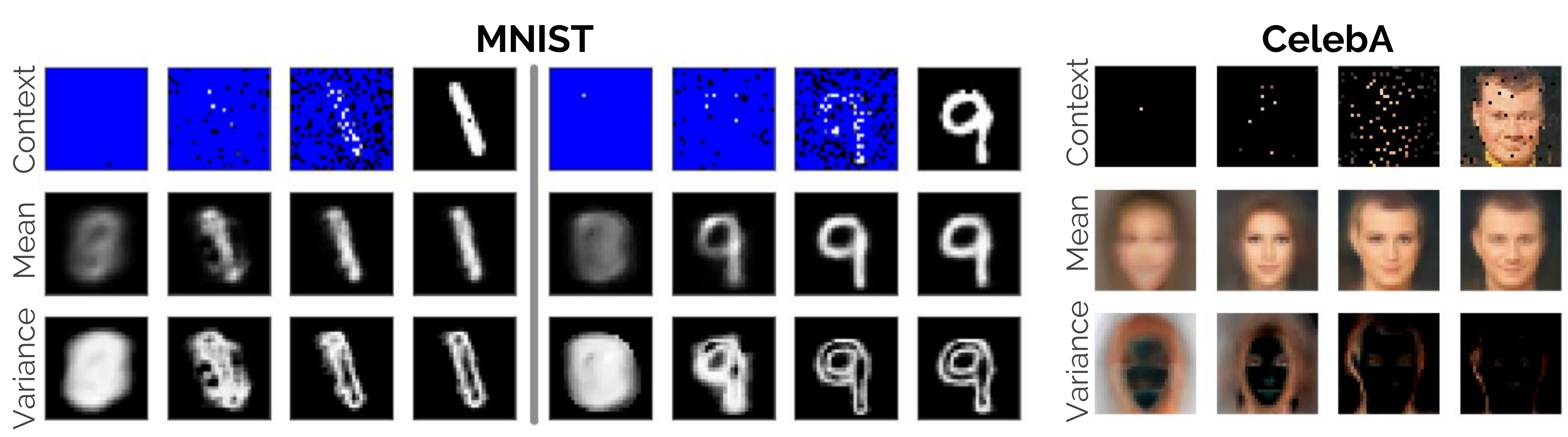
Predicted mean and variance of CNP for a curve generated with switching kernel parameters.

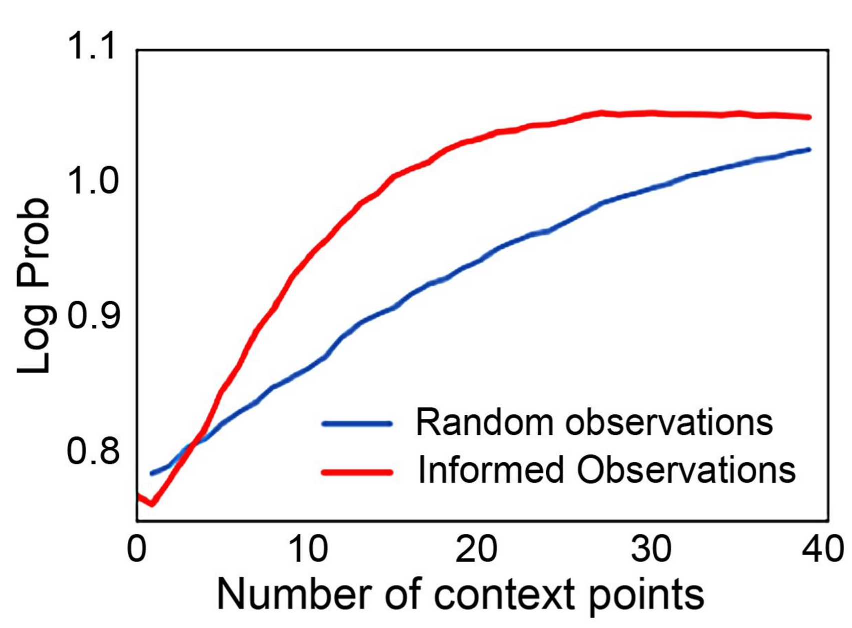
Image Completion

Image Completion on MNIST and CelebA

Formulate this as regression task:
- **Inputs**: 2D pixel positions
- **Outputs**: pixel brightness

At test time predict mean (middle row) and variance (bottom row) of every pixel in an image given increasing numbers of context pixels as observations.

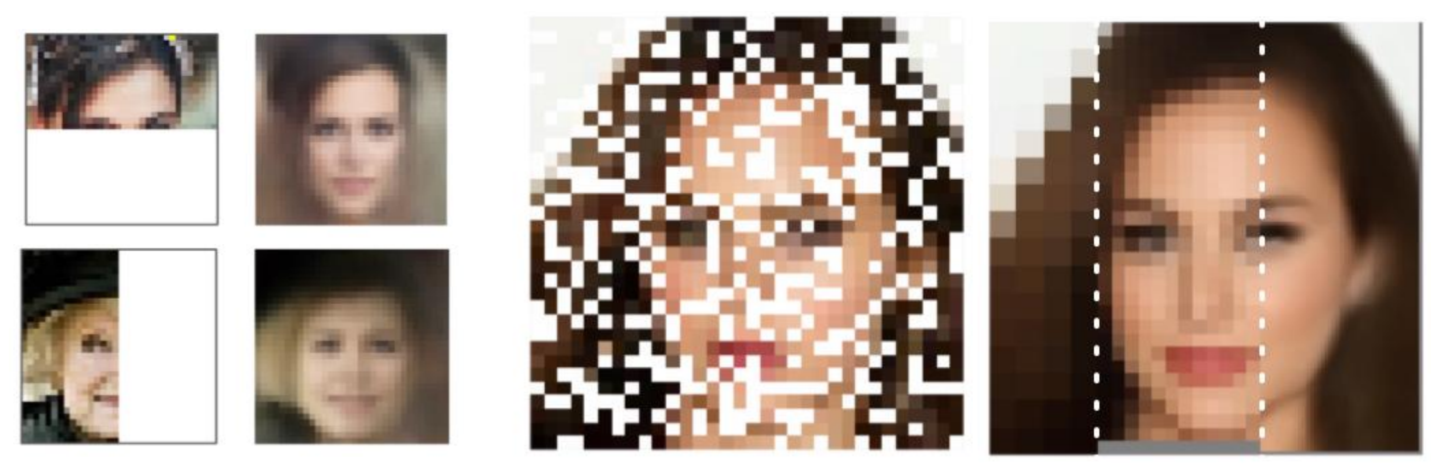




Active exploration

Sequentially increase number of context points and choose next observation by picking the pixel with highest variance.

Pixels chosen like this are more informative and lead to higher log probs (red curve) than pixels chosen at random (blue curve).



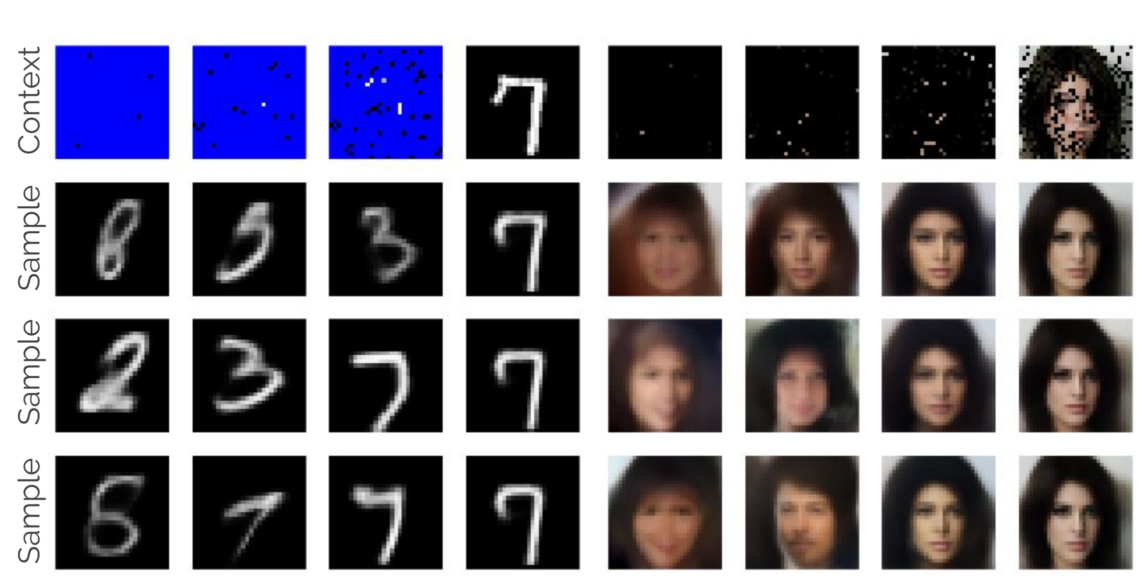
In contrast to standard conditional models, our model can be conditioned on arbitrary patterns, even ones which were never seen in the training set. Similarly, the model can predict values for pixel coordinates that were never in the training set (e.g. subpixel values).

Latent Variable Model

So far target points are generated independently of each other and only the mean produces a coherent image/ curve.

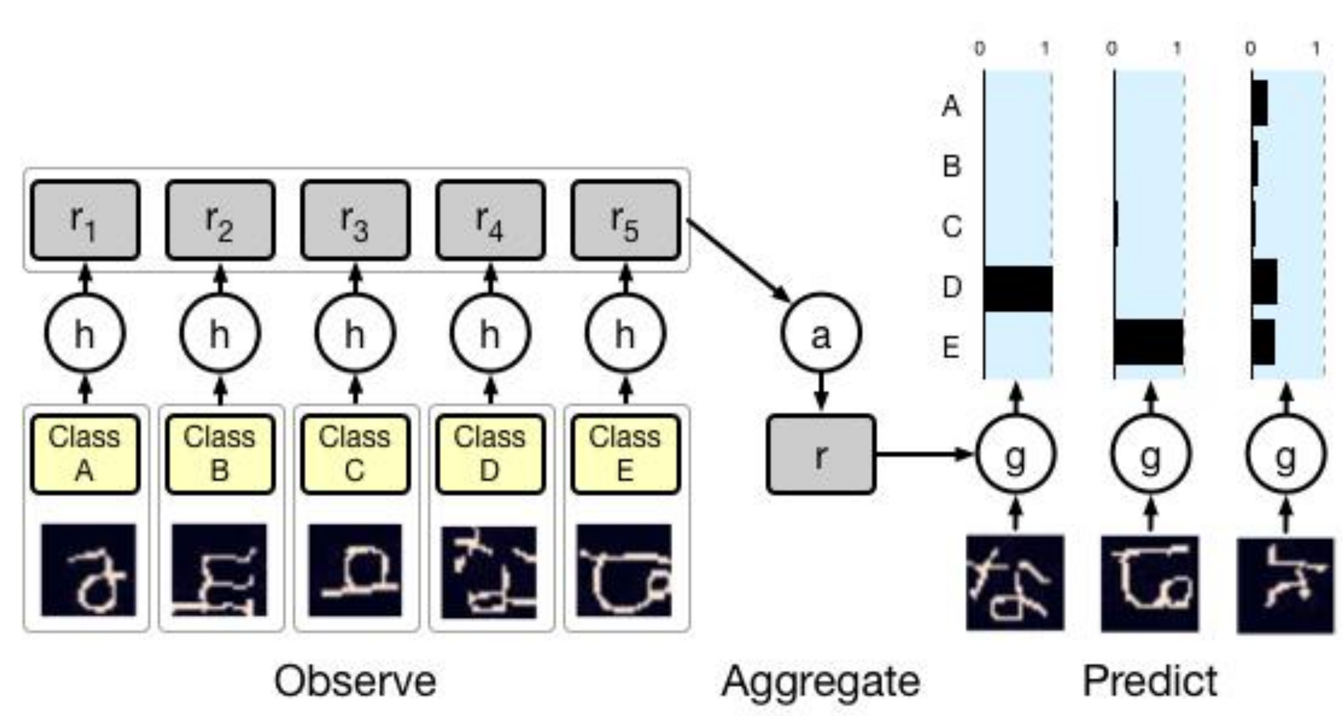
We achieve coherent samples with correlated targets by introducing a latent variable that captures global information across targets.

We can thus sample entire images and uncertainty itself is reflected in the variability of the underlying digit/face.



As the number of observations increases, uncertainty is reduced and the samples converge to the same prediction.

1-Shot Classification



CNPs can also be straightforwardly applied to one-shot classification.

With a simpler model and smaller runtime it achieves classification accuracies comparable to state-of-the-art.

	5-way Acc		20-way Acc		Runtime
	1-shot	5-shot	1-shot	5-shot	
MANN	82.8%	94.9%	-	-	$\mathcal{O}(nm)$
MN	98.1%	98.9%	93.8%	98.5%	$\mathcal{O}(nm)$
CNP	95.3%	98.5%	89.9%	96.8%	$\mathcal{O}(n+m)$

Results on the Omniglot one-shot classification task for Memory Augmented Networks (MANN) [1], Matching Networks (MN) [2] and CNPs.

Summary

We introduce CNPs, a model that is **flexible at test time** and has the capacity to **extract prior knowledge** from training data.

We have demonstrated its ability to perform a variety of tasks: regression, one-shot classification and image completion.

Check out our paper for details →

[1] Santoro et al. *One-shot learning with memory-augmented neural networks*

[2] Vinyals et al. *Matching networks for one shot learning.*