CONDITIONAL AND LATENT NEURAL PROCESSES

UNIVERSITY OF CAMBRIDGE

The University of Cambridge, Advanced Machine Learning

William Baker
Alexandra Shaw
Tony Wu

Motivation

Conditional Neural Process (CNP), Latent Neural Processes (LNP) and Hybrid Neural processes (HNP) attempt to combine the best characteristics of Gaussian Processes and Neural Networks.

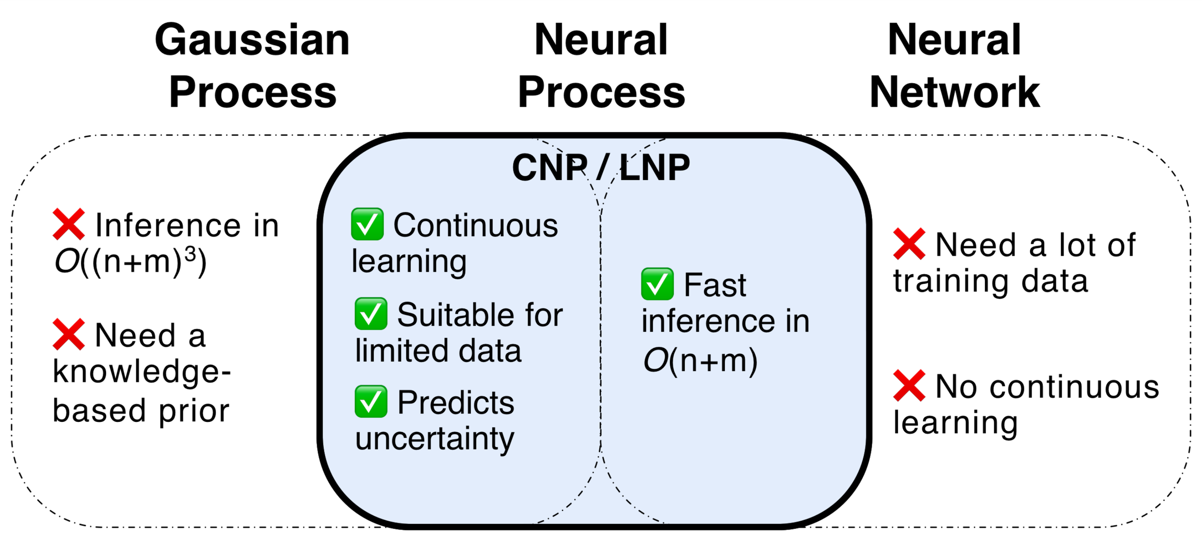

Fig 1: Comparison between Gaussian Process, Neural Network and Neural Process

Model Architecture

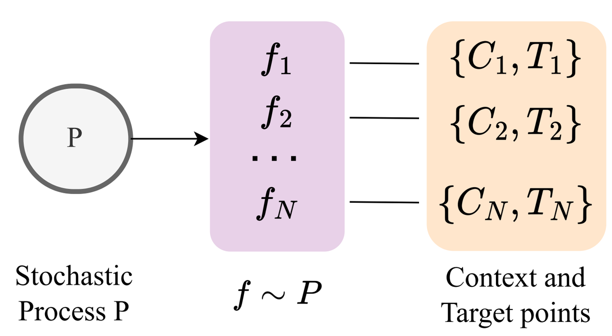

Fig 2a: Data pipeline for Neural Processes

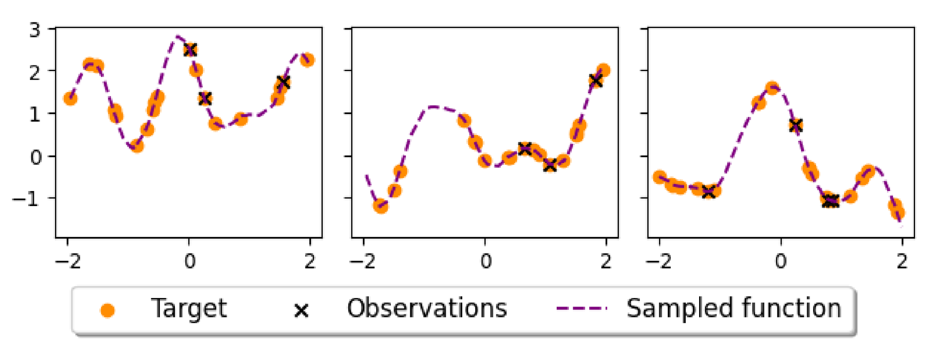

Fig 2b: Examples of $\{C_N, T_N\}$ drawn from an arbitrary Stochastic Process

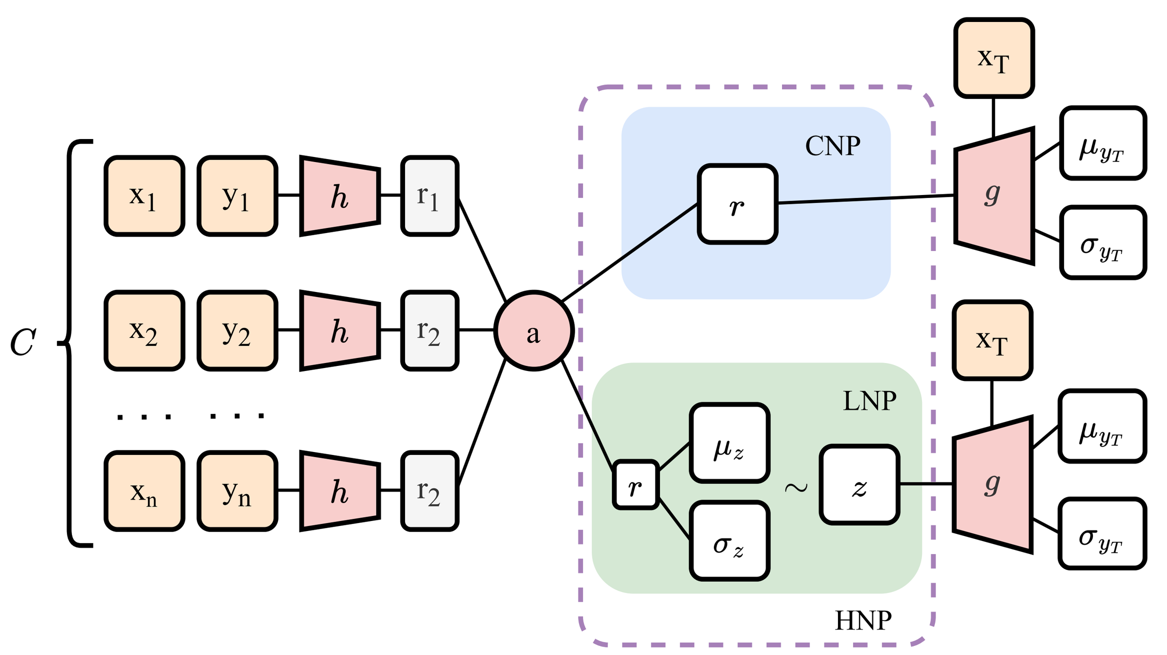

Fig 3a. Architecture diagram of CNP, latent NP.

$$C = \{(x,y)_i\}_{i=1}^{n}, T = \{x_i\}_{i=1}^{m}$$

$$r_i = h_\theta(x_i, y_i) \ \forall (x,y) \in C$$

$$r = \frac{1}{n}\sum_{i=1}^{n} r_i$$

$$z \sim \mathcal{N}(\mu_z(r), I\sigma_z(r))$$

$$\phi_i^{\text{CNP}} = g_\theta(x_i, r) \ \forall (x_i) \in T$$

$$\phi_i^{\text{LNP}} = g_\theta(x_i, z) \ \forall (x_i) \in T$$

$$\phi_i^{\text{HNP}} = g_\theta(x_i, z, r) \ \forall (x_i) \in T$$

Fig 3b. Mathematical formulation of model architecture.

Training and Loss Functions

Conditional Neural Processes, which seek to learn useful embeddings, are optimized by minimizing the negative conditional log probability of the outputs of the context and target points, given the context points and input target locations. For a conditional stochastic process Q_θ:

$$\mathcal{L}(\theta) = -\mathbb{E}_{f\sim P}[\mathbb{E}_N[\log Q_\theta(f(x_T)|C, x_T)]]$$

Latent Neural Processes, which seek to learn a latent distribution over the embedding space, are optimized by maximizing the ELBO of the log probability predictive distribution:

$$\log p(f(T)|T,C) \geq \mathbb{E}_{q(z|C,T)}\left[\sum_{x\in T}\log p(f(x)|z,x) - \log\frac{q(z|C,T)}{p(z|C)}\right]$$

$$\approx \mathbb{E}_{q(z|C,T)}\left[\sum_{x\in T}\log p(f(x)|z,x) - D_{KL}(q(z|C,T)\|q(z|C))\right]$$

Since NPs are neural networks, the loss functions are optimized with gradient descent using Adam.

1D Regression

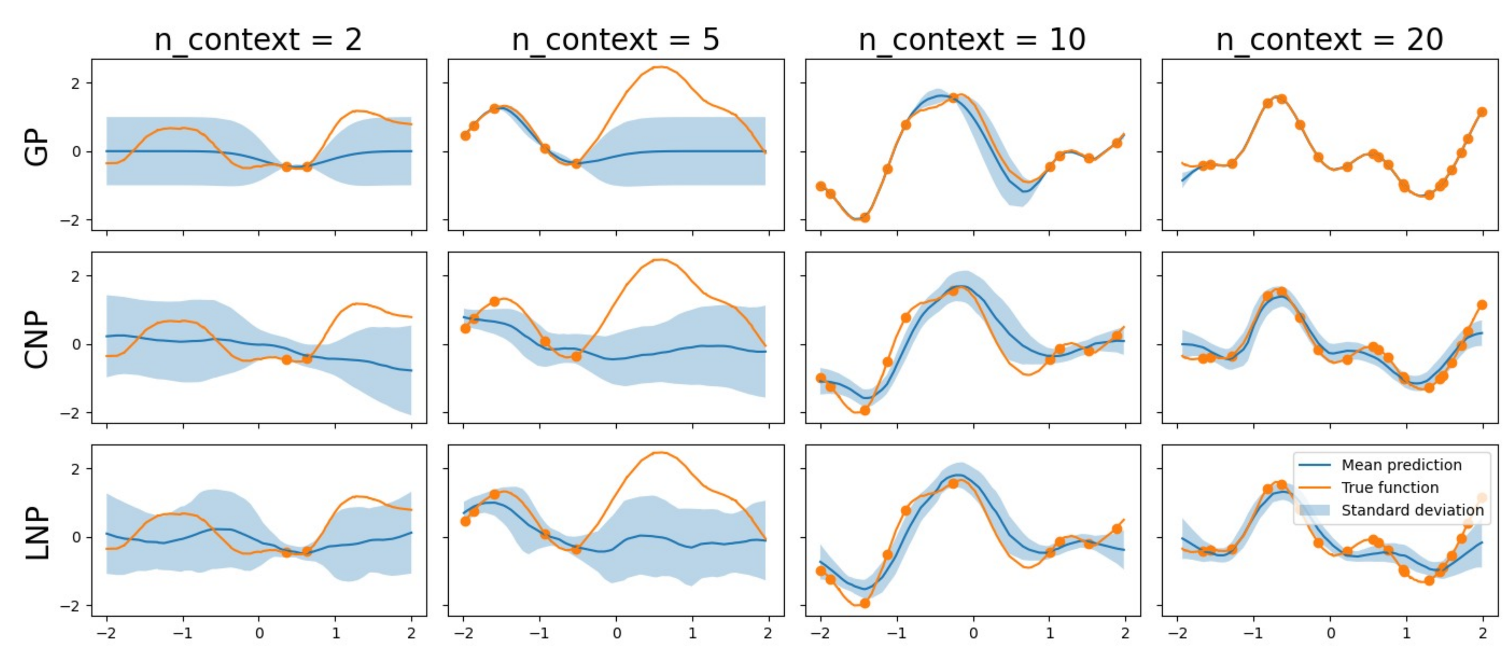

Fig 4: Comparison between GP, CNP and LNP on the 1D-regression task (fixed kernel parameter)

Image Completion on MNIST

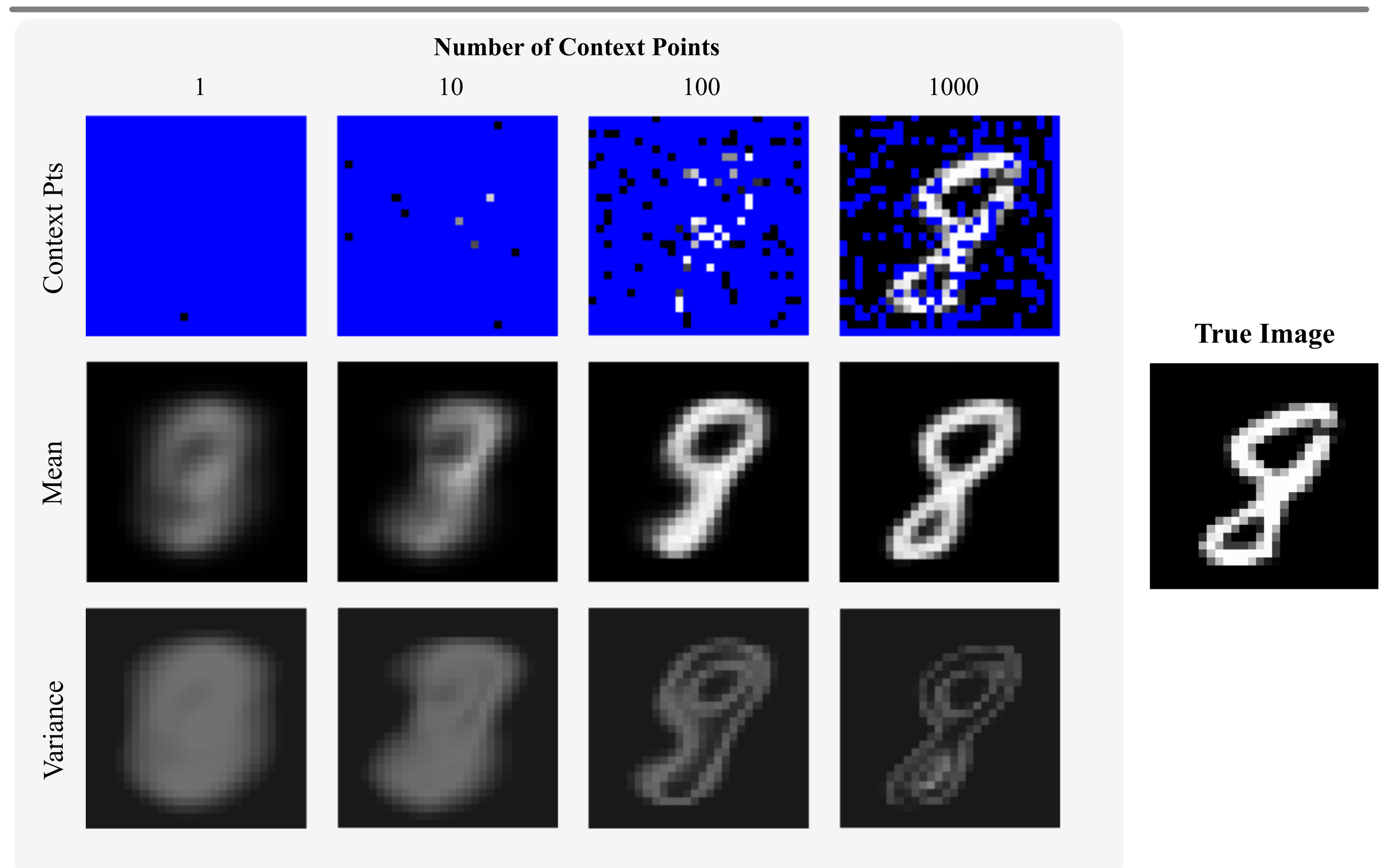

Fig 5: CNP pixel mean and variance predictions on images from MNIST.

Image Completion on CelebA

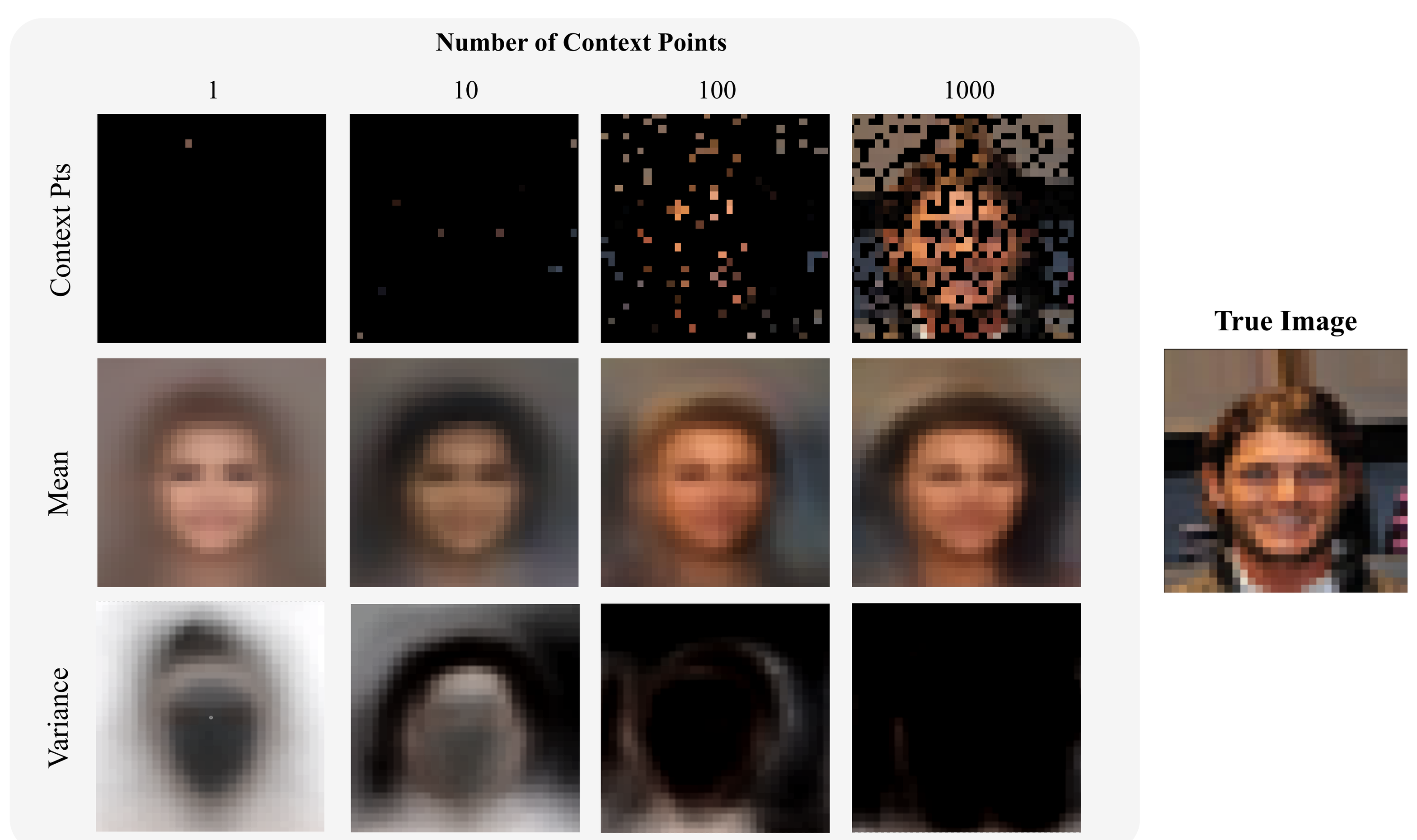

Fig 6: CNP pixel mean and variance predictions on images from CelebA.

Varying Kernel Parameters

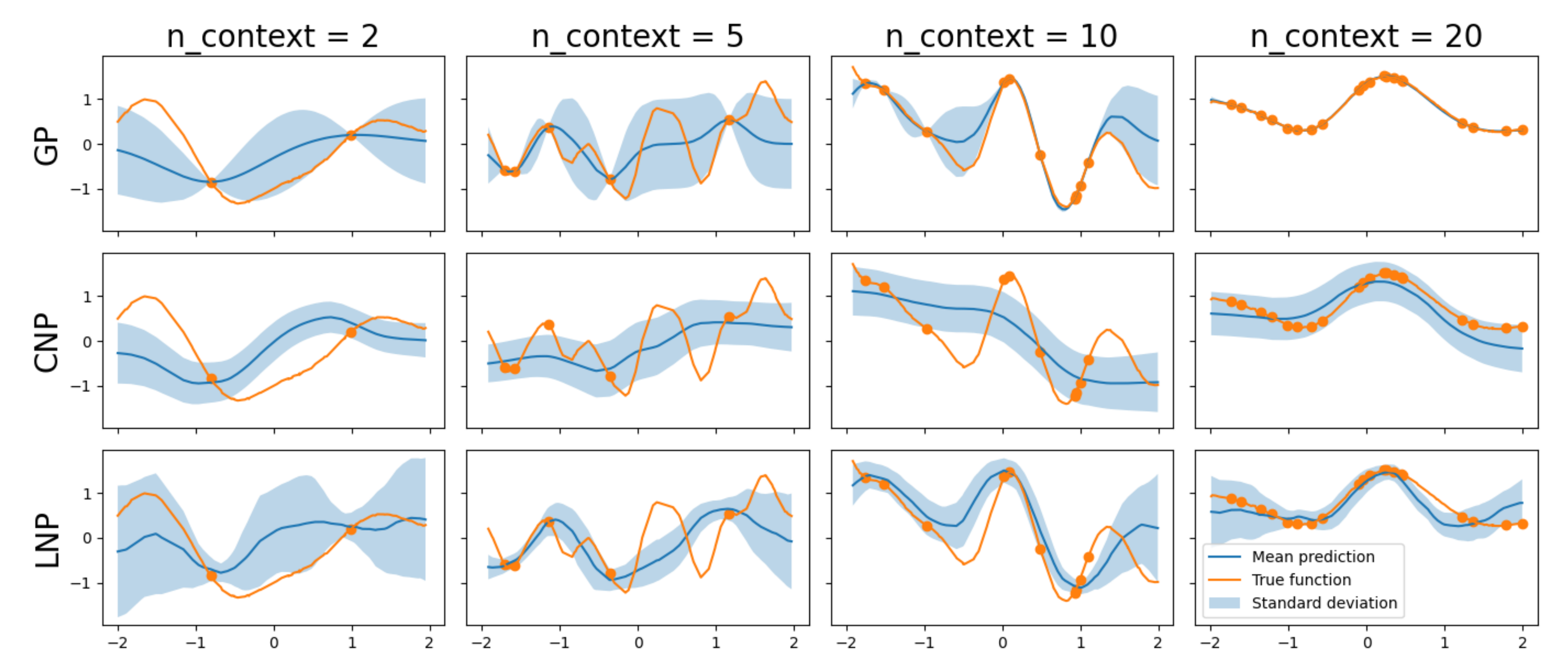

Fig 7: Comparison between GP, CNP and LNP on the 1D-regression task (varying kernel parameter) - data was generated using a Exponentiated Quadratic Kernel with length varying between 0.1 and 1.0

LNP Image Completion on MNIST

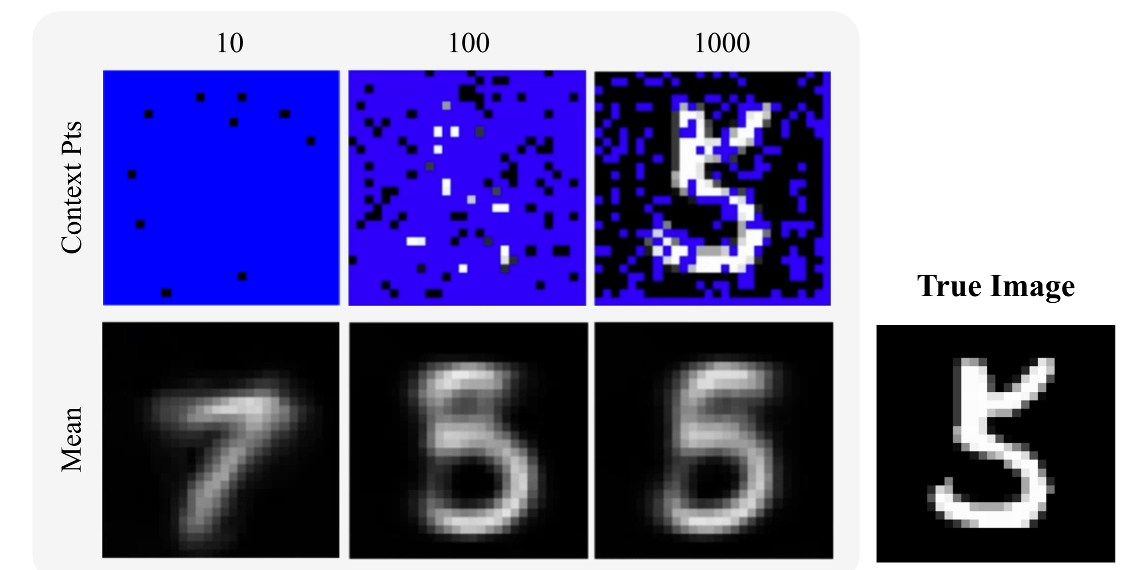

Fig 8: LNP pixel mean and variance predictions on images from MNIST

Flexible Image Completion

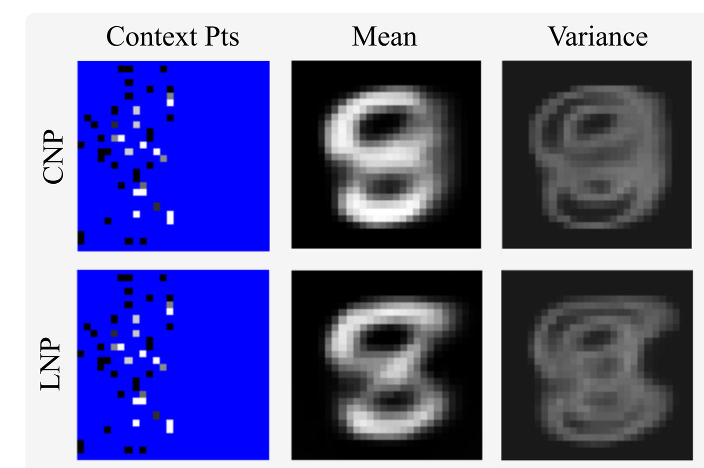

Fig 9a: Sampling from one half of the image only, forcing the NP to extrapolate the remaining half

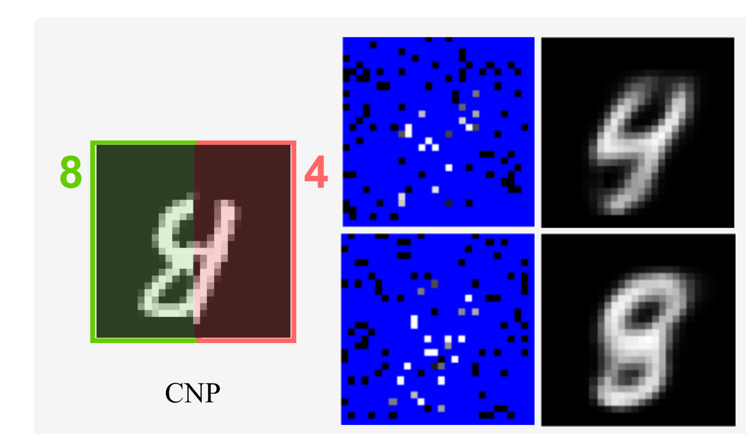

Fig 9b: Combining two images, sampling the CNP result can yield either parent image

Prior vs Posterior / HNP to CNP Convergence

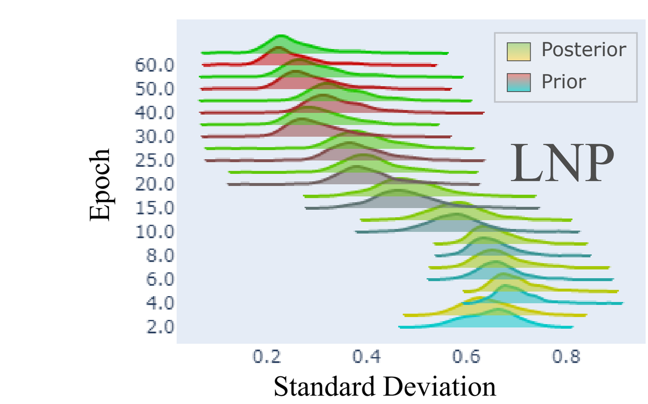

Fig 10a: Latent distribution over standard deviation in LNP

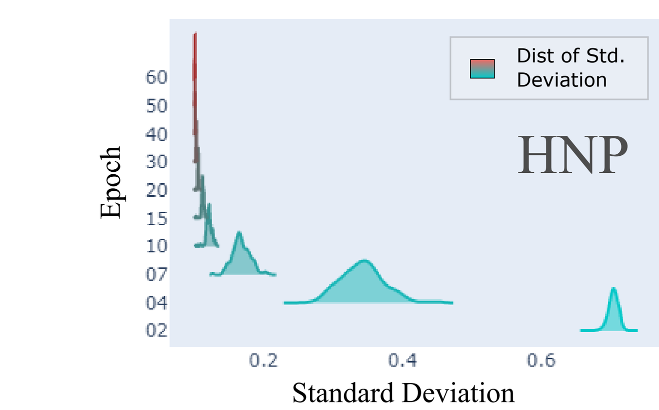

Fig 10b: Latent distribution over standard deviation in HNP

[1] Garnelo, Marta, et al. "Conditional neural processes." *International conference on machine learning*. PMLR, 2018.
[2] Garnelo, Marta, et al. "Neural processes." *arXiv preprint arXiv:1807.01622* (2018).
[3] Jha, Saurav, et al. "The neural process family: Survey, applications and perspectives." *arXiv preprint arXiv:2209.00517* (2022).
[4] LeCun, Yann. "The MNIST database of handwritten digits." http://yann.lecun.com/exdb/mnist/ (1998).
[5] Liu, Ziwei, et al. "Deep learning face attributes in the wild." Proceedings of the IEEE international conference on computer vision. 2015.