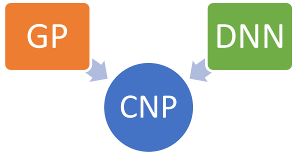
Kristopher Miltiadou,
Agathe de Vulpian,
Max-Olivier van Bastelaer

CONDITIONAL NEURAL PROCESSES

MPhil Machine Learning and Machine Intelligence

UNIVERSITY OF CAMBRIDGE

Motivation



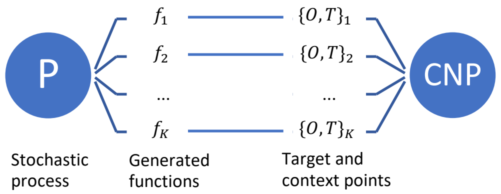
CNPs are inspired by the flexibility of stochastic processes but are structured as neural networks and trained via gradient descent.

Key properties

- CNPs define a conditional distribution over functions given a set of observations
- A CNP is invariant to permutation of the inputs
- A CNP is scalable, achieving a running time complexity *O(n+m)* for making *m* predictions with *n* observations

Models

Data pipeline



Architecture Equation

$r_i = e_\theta(x_i, y_i) \quad \forall (x_i, y_i) \in O$

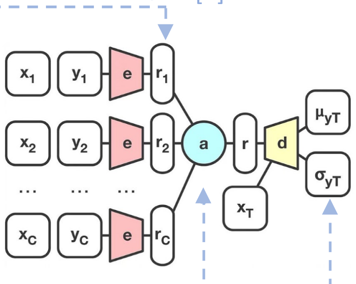
$r = r_1 \oplus r_2 \oplus \dots r_{n-1} \oplus r_n$

$\mu_{y,i}, \sigma_{y,i} = d_\theta(x_i, r) \quad \forall (x_i) \in T$

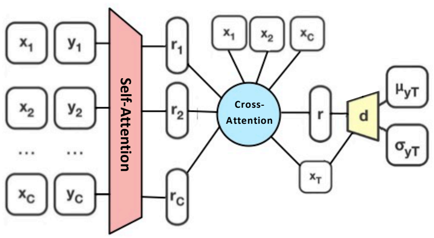
Training Objective

$$\mathcal{L}(\theta) = -\mathbb{E}_{f\sim P}\left[\mathbb{E}_N\left[\log Q_\theta(\{y_i\}_{i=0}^{n-1}|O_N, \{x_i\}_{i=0}^{n-1})\right]\right]$$

CNP [1]

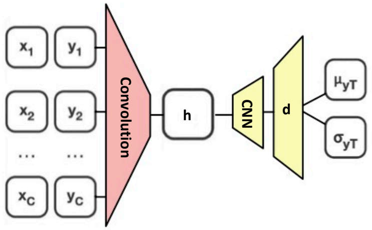


Attn-CNP [2]



Aim: Mitigate underfitting
How: Aggregate the representation of the context set in a more flexible way using cross-attention while keeping permutation invariance

Conv-CNP [3]



Aim: Support translational equivariance
How: Incorporate CNN which are explicitly designed to satisfy this property into the CNP

1D Regression

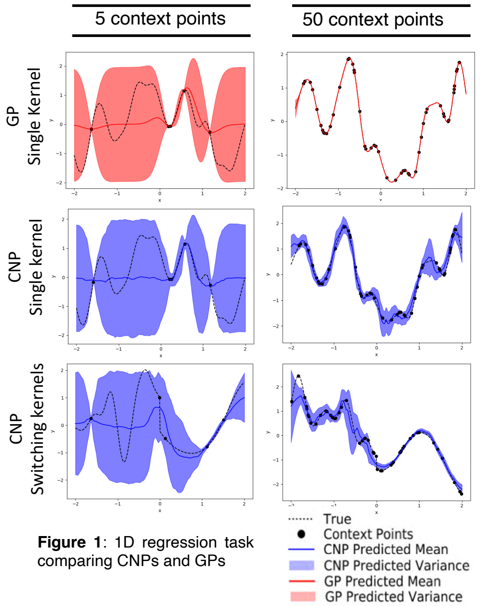


Figure 1: 1D regression task comparing CNPs and GPs

2D Regression – Image Completion

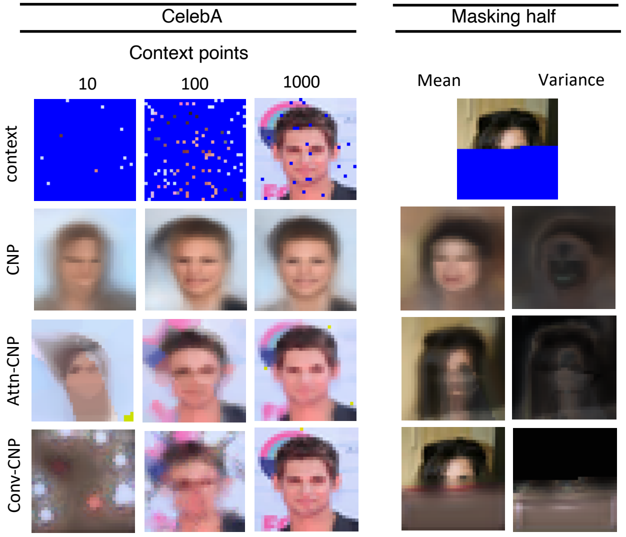


Figure 2: Image completion with CelebA, CNPs can learn non trivial kernels which would be nearly impossible to encode manually with GPs

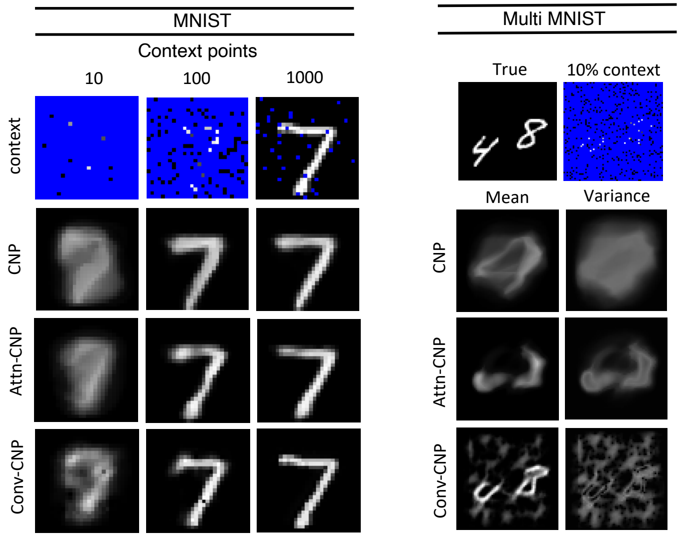


Figure 3: The Models were trained on 28x28 MNIST and then tested on 64x64 Multi-MNIST, it illustrates the translation equivariance achieved by ConvCNPs

References

[1] Garnelo, Marta, et al. "Conditional neural processes." *International Conference on Machine Learning*. PMLR, 2018.
[2] Kim, Hyunjik, et al. "Attentive neural processes." arXiv preprint arXiv:1901.05761 2019.
[3] Gordon, Jonathan, et al. "Convolutional conditional neural processes." arXiv preprint arXiv:1910.13556 2019.

[MNIST] LeCun, Yann. "The MNIST database of handwritten digits." http://yann. lecun. com/exdb/mnist/ (1998).
[CelebA] Liu, Ziwei, et al. "Deep learning face attributes in the wild." Proceedings of the IEEE international conference on computer vision. 2015.

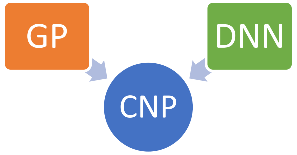
Kristopher Miltiadou,
Agathe de Vulpian,
Max-Olivier van Bastelaer

CONDITIONAL NEURAL PROCESSES

MPhil Machine Learning and Machine Intelligence

UNIVERSITY OF CAMBRIDGE

Motivation



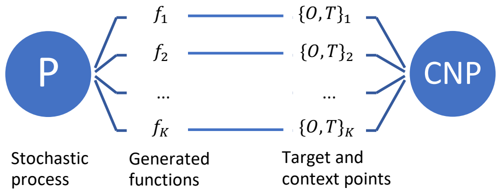
CNPs are inspired by the flexibility of stochastic processes but are structured as neural networks and trained via gradient descent.

Key properties

- CNPs define a conditional distribution over functions given a set of observations
- A CNP is invariant to permutation of the inputs
- A CNP is scalable, achieving a running time complexity *O(n+m)* for making *m* predictions with *n* observations

Models

Data pipeline



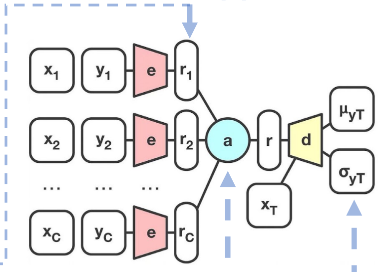
Architecture Equation

$$r_i = e_\theta(x_i, y_i) \quad \forall (x_i, y_i) \in O$$
$$r = r_1 \oplus r_2 \oplus \ldots r_{n-1} \oplus r_n$$
$$\mu_{y,i}, \sigma_{y,i} = d_\theta(x_i, r) \quad \forall (x_i) \in T$$

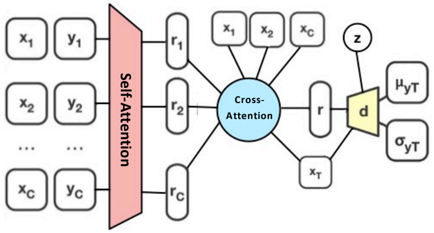
Training Objective

$$\mathcal{L}(\theta) = -\mathbb{E}_{f\sim P}\left[\mathbb{E}_N\left[\log Q_\theta(\{y_i\}_{i=0}^{n-1} | O_N, \{x_i\}_{i=0}^{n-1})\right]\right]$$

CNP [1]

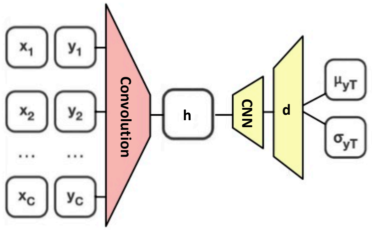


Attn-CNP [2]



Aim: Mitigate underfitting
How: Aggregate the representation of the context set in a more flexible way using cross-attention while keeping permutation invariance

Conv-CNP [3]



Aim: Support translational equivariance
How: Incorporate CNN which are explicitly designed to satisfy this property into the CNP

1D Regression

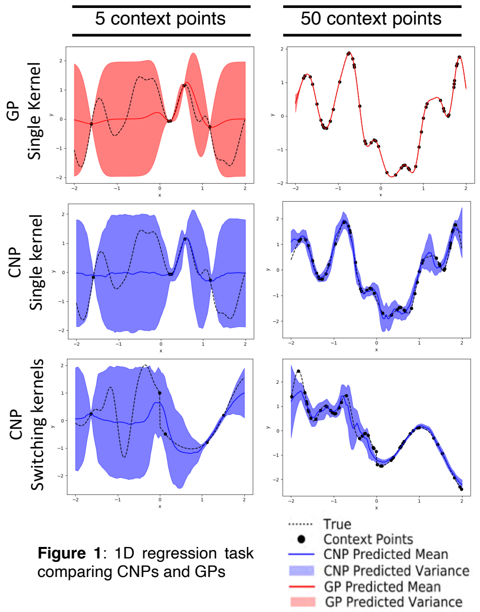


Figure 1: 1D regression task comparing CNPs and GPs

References

[1] Garnelo, Marta, et al. "Conditional neural processes." *International Conference on Machine Learning*. PMLR, 2018.
[2] Kim, Hyunjik, et al. "Attentive neural processes." arXiv preprint arXiv:1901.05761 2019.
[3] Gordon, Jonathan, et al. "Convolutional conditional neural processes." arXiv preprint arXiv:1910.13556 2019.

[MNIST] LeCun, Yann. "The MNIST database of handwritten digits." http://yann. lecun. com/exdb/mnist/ (1998).
[CelebA] Liu, Ziwei, et al. "Deep learning face attributes in the wild." Proceedings of the IEEE international conference on computer vision. 2015.

2D Regression – Image Completion

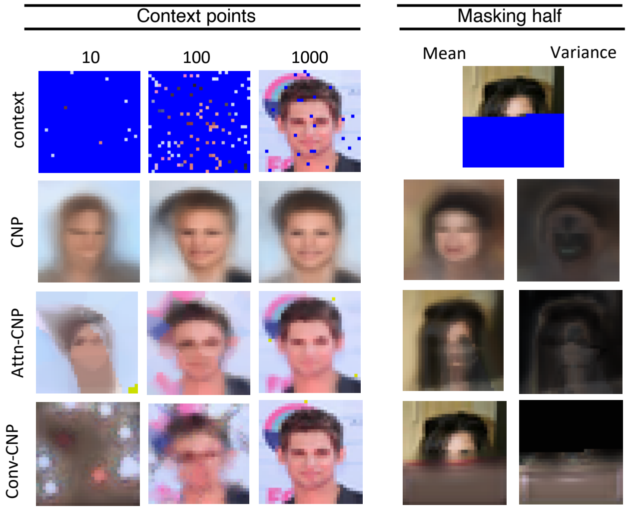


Figure 2: CelebA image completion, CNPs can learn non trivial kernels

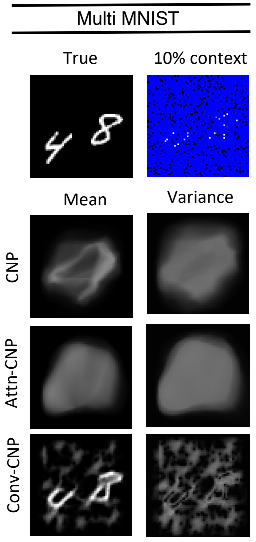
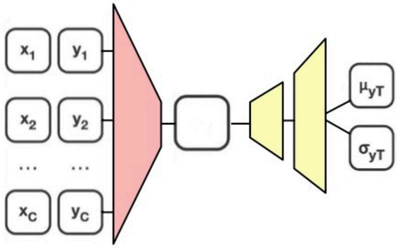
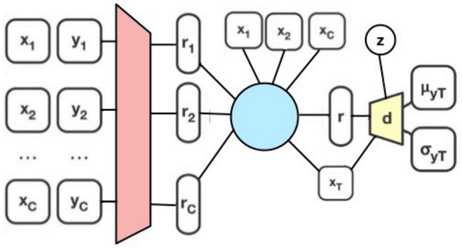


Figure 3: Multi-MNIST, Conv-CNP achieves translational equivariance



x1
y1
x2
y2
...
...
xC
yC
μyT
σyT



x1
y1
x2
y2
...
...
xC
yC
r1
r2
rC
x1
x2
xC
r
xT
z
d
μyT
σyT