UNIVERSITY OF CAMBRIDGE

The University of Cambridge, Advanced Machine Learning

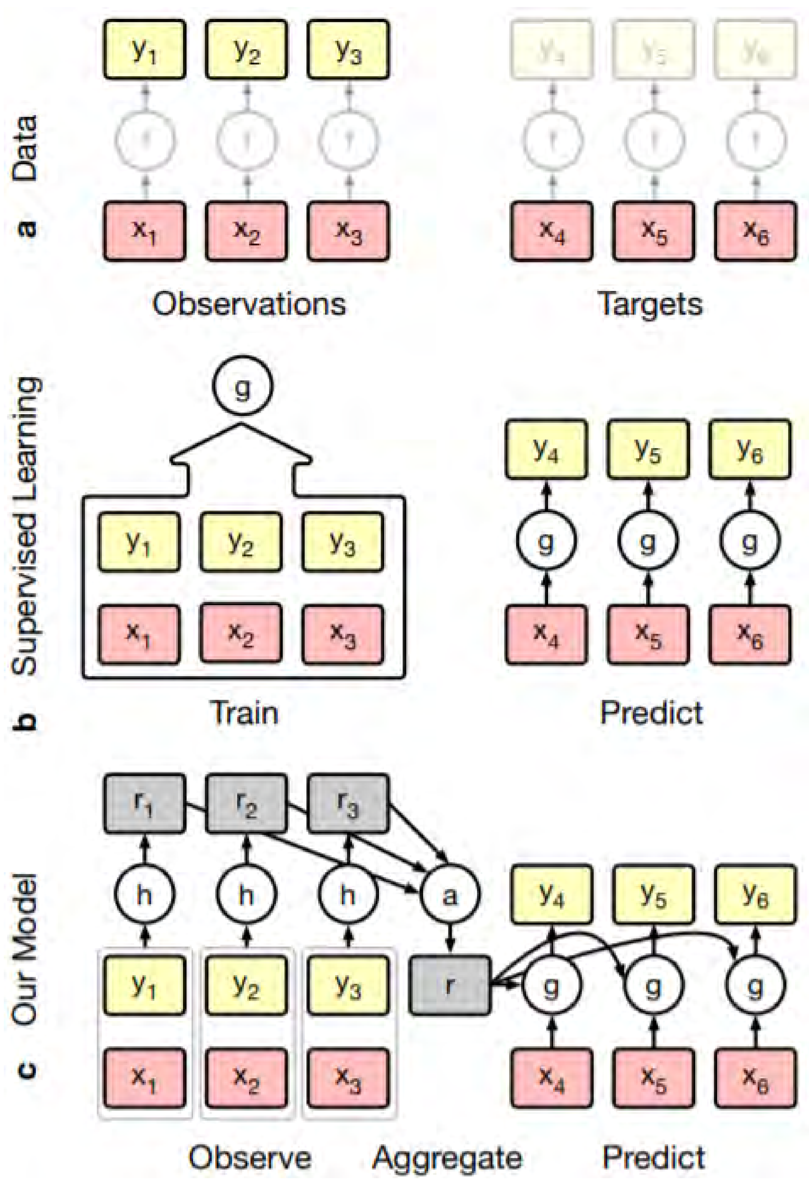
Conditional Neural Processes

Igor Adamski
Conor Foy
W. Yomjinda

Introduction

Conditional Neural Processes (CNPs) [1] combine the robustness of deep neural networks (DNNs) in function approximation and the data efficiency of Gaussian Processes (GPs). The key properties describing CNPs are:

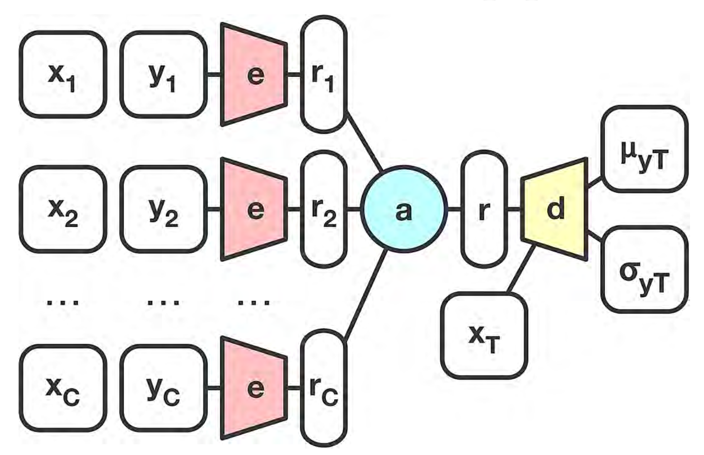
- CNPs define conditional distributions over functions given a set of observations.
- CNPs are parametrized by a neural network that is invariant under permutations of its inputs.
- CNPs are scalable with complexity $\mathcal{O}(n+m)$ for making m predictions from n observations.



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

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

Illustration of the CNP pipeline:



$e(= h)$ and $d(= g)$ above are like **encoder** and **decoder** and are parametrized by DNNs.

Function Regression

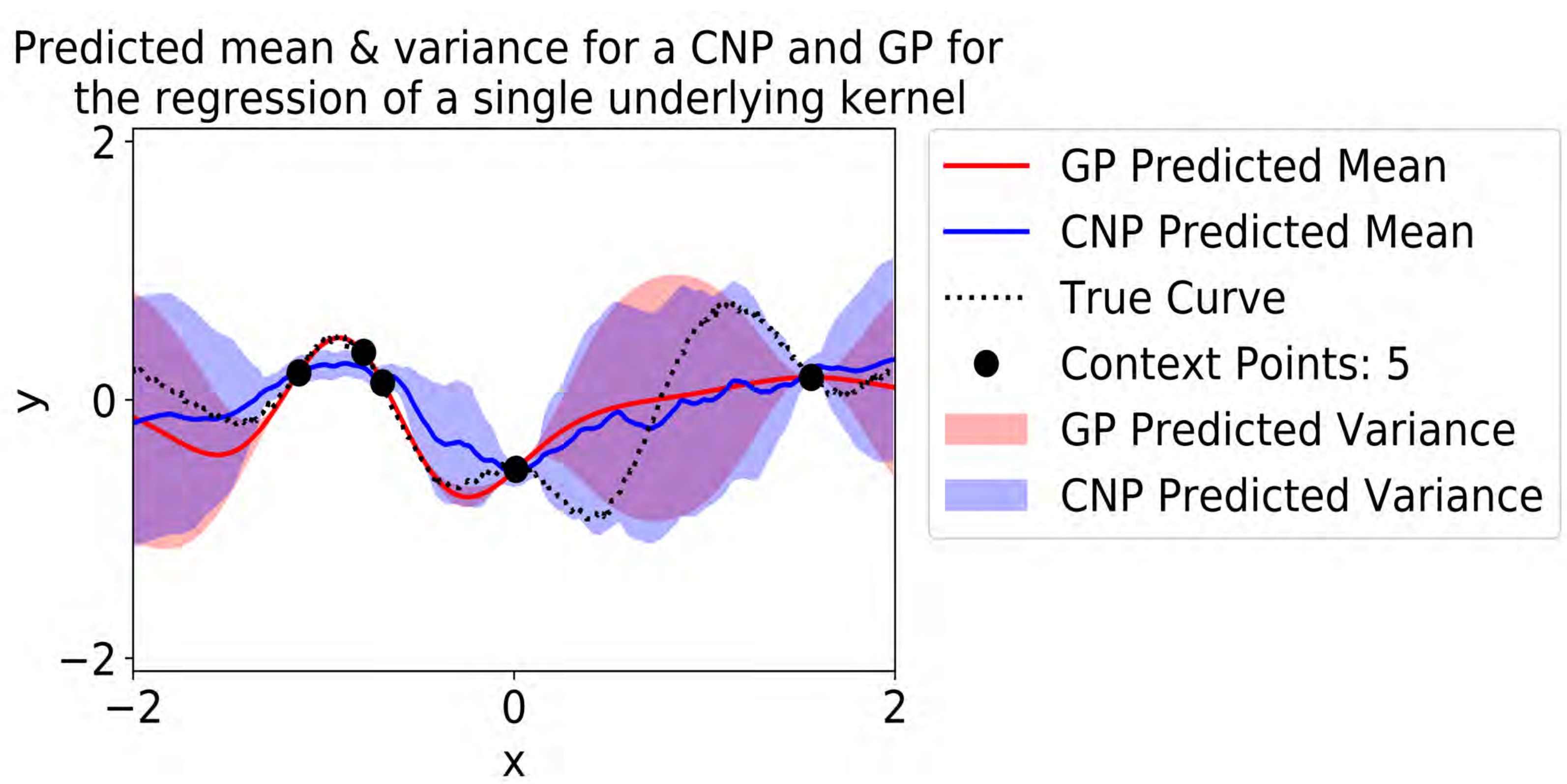


Figure 2: Regression results on a 1-D curve using 5 context-points.

Pixel-wise image regression on MNIST

For this task we test the CNPs on the MNIST dataset. The model outputs the per-pixel mean and variance of pixel intensity.

- In Figure 3a the model learns to make good predictions of the underlying digit even for a small number of context points.
- In Figure 3b, the model can be used to **upscale** an image.

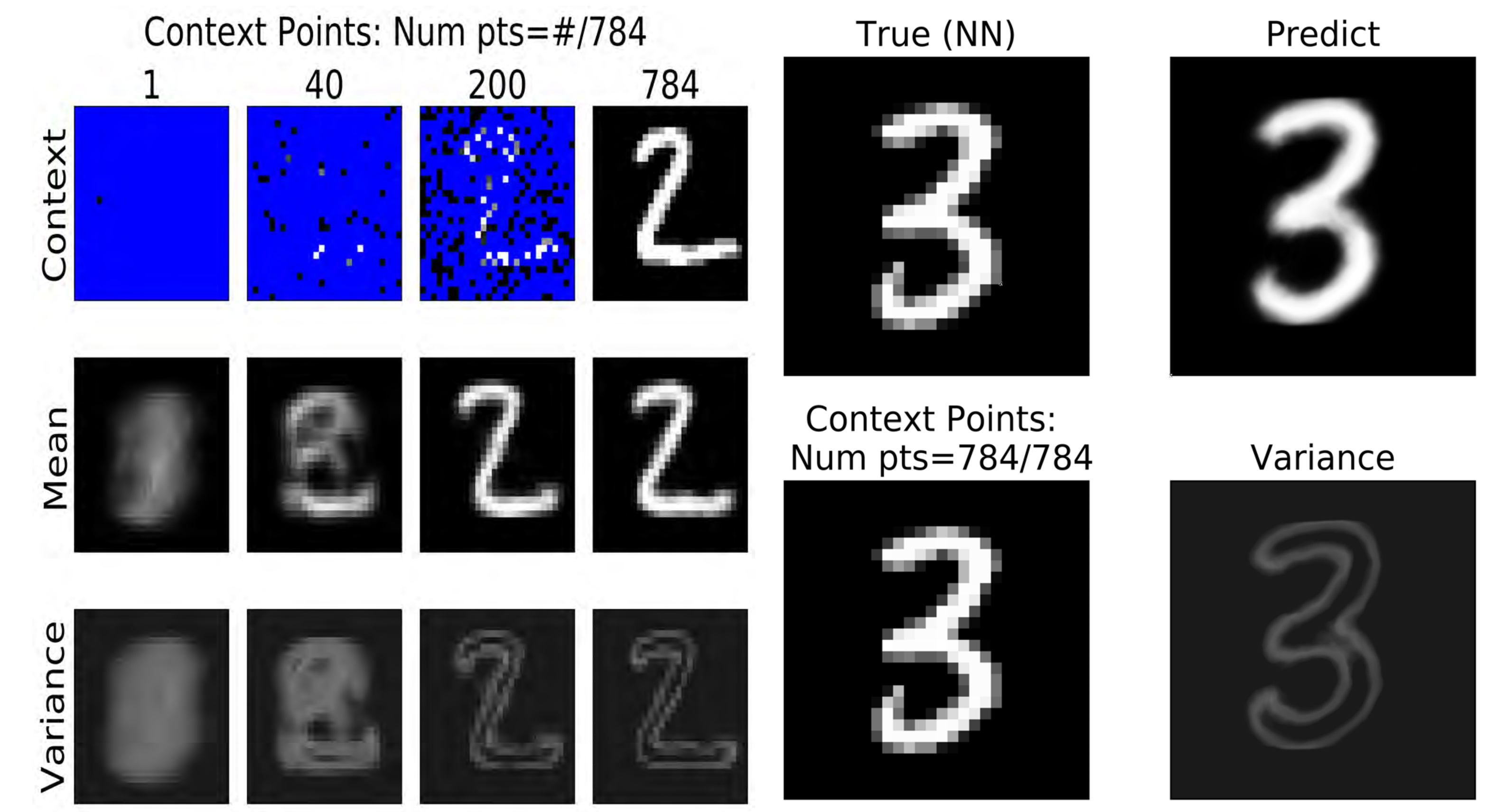


Figure 3: Left: We provide the model with 1, 40, 200 and 784 context points (top row) and query the entire image. Right: 256x256 up-scaling result from a 28x28 image.

Image completion on CelebA

We test CNPs on the face-completion task on the CelebA dataset. The model learns to recover 'generic' **face representations** from just a handful of context points on previously unseen faces.

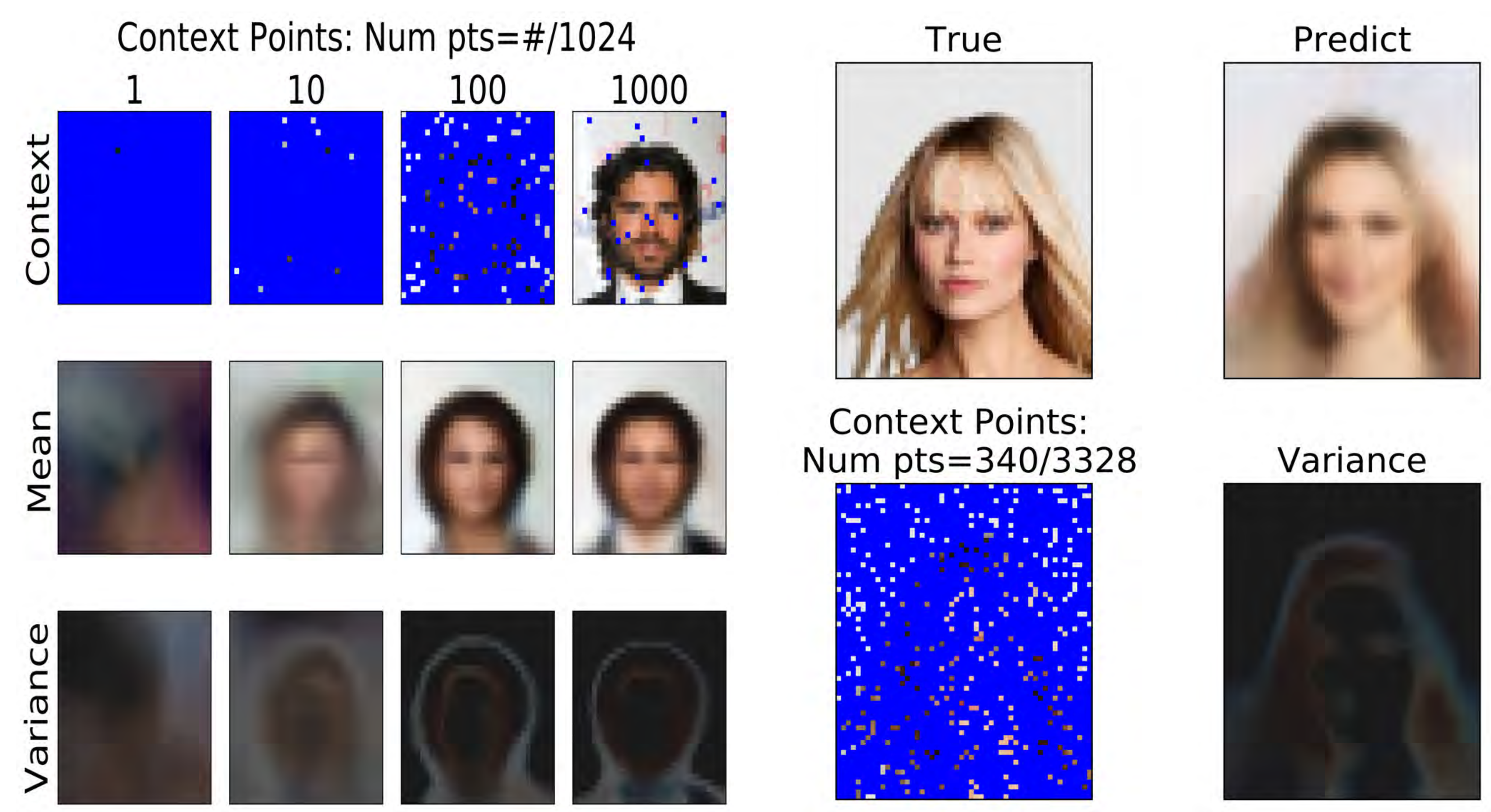


Figure 4: Left: examples of CelebA image completion with varying numbers of observations. Right: we can recover a 64x52 sample from just 10% of the points.

Music Completion on MIDI

We test CNP architecture on the MAESTRO dataset [2], containing ≈ 200h of piano music. The CNP learns to seamlessly **connect** the two contexts and understands which notes go together well.

- We use **biaxial LSTMs** [3] for the encoder and decoder.
- We calculate $r = r_{\text{left}} \oplus r_{\text{right}}$ by **concatenation**.

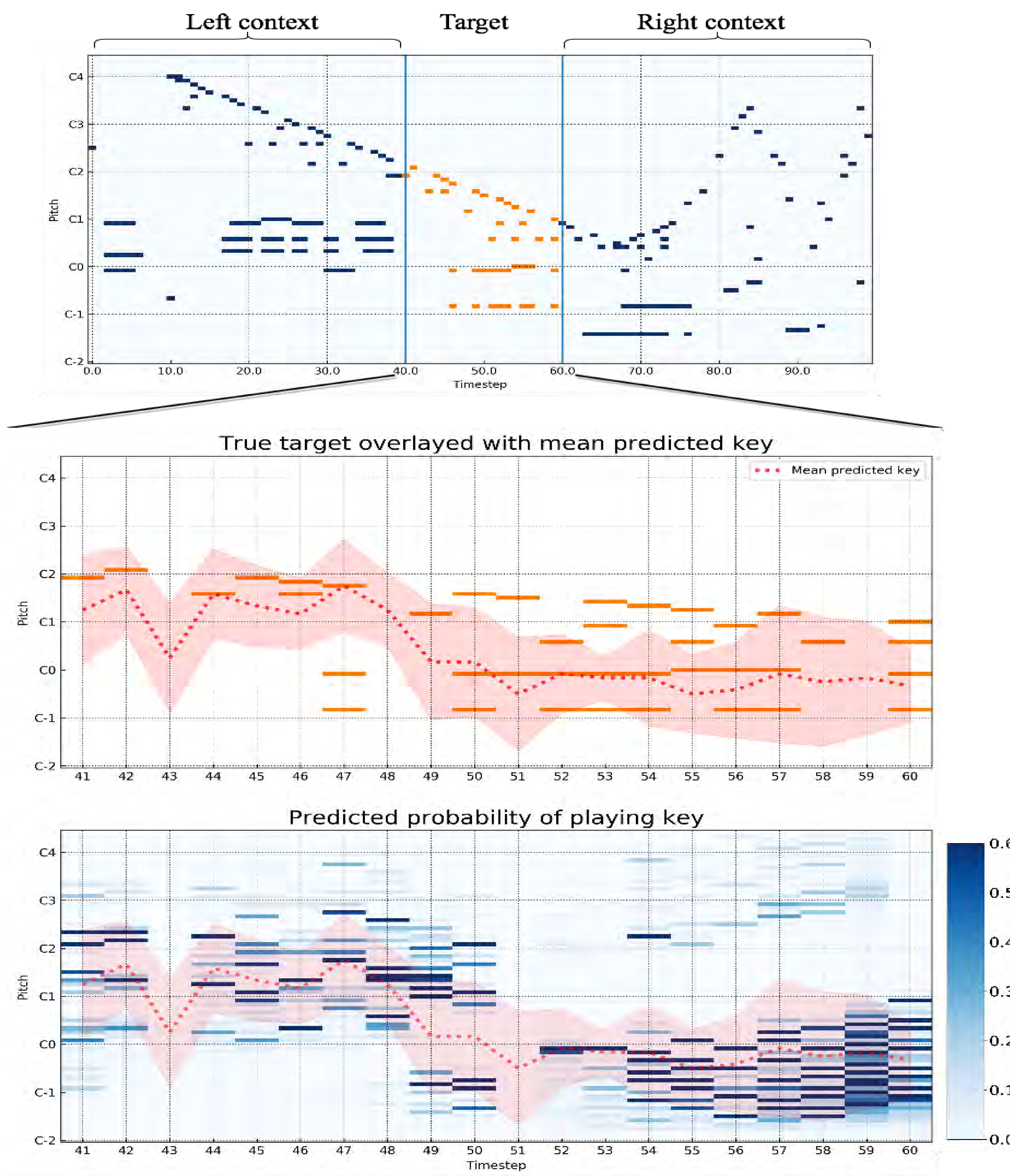


Figure 5: Top row: Schematic of the task. Middle row: The true target presented with mean predicted key. Bottom row: The actual per-note outputs of the CNP on the target.

References

[1] *Conditional Neural Processes*, M. Garnelo et.al, 2018
[2] *Enabling Factorized Piano Music Modeling and Generation with the MAESTRO Dataset*, C. Hawthorne et.al, 2019
[3] *Generating Polyphonic Music Using Tied Parallel Networks* D. D. Johnson, 2017