DEEP learning for Quantum Physics

Adriano Macarone Palmieri

SPAM = state, preparation, and measurement error

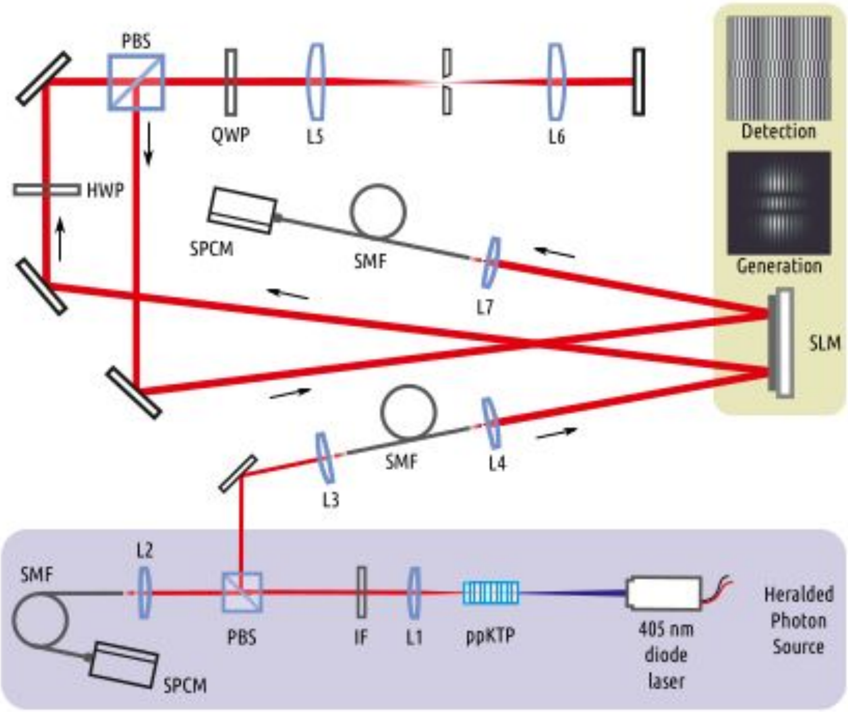
tomography = Denoising

Given a damaged or incorrectly observed input x, the machine learning system returns an estimate of the original or correct x. . For example, the machine learning system might be asked to remove dust or scratches from an old photograph. This requires multiple outputs (every element of the estimated clean example x) and an understanding of the entire input (since even one damaged area will still reveal the final estimate as being damaged)
(*from "Deep Learning" I.Goodfellow*)

State preparation and errors

How bad our data were, naive errors and how the cost function saved us

THe Hermite Gaussian modes



$$\mathrm{HG}_n(z) \propto H_n(\sqrt{2}z/w)\exp(-z^2/w^2),$$

where w is the mode waist. We limited the dimensionality of the Hilbert space to 6 by using only the beams with

$$n + m \leq 2$$

The hologram in the generation part uses amplitude modulation to produce high-quality HG modes, while a phase-only hologram at the detection part sacrifices projection quality for efficiency

not so good states...

"Here we use the full two-dimensional mode spectrum of HG modes, which is equivalent to <u>including the radial degree of freedom</u> in addition to OAM. This is rarely done in quantum experiments, and one of the reasons is poor quality of projective measurements"

$$\int_{-\infty}^{\infty} \mathrm{HG}^{*}_{n'm'}(x,y) \times \mathrm{HG}_{nm}(x,y) \times \exp[-(x^2+y^2)/w_f^2]\, dx dy \neq \delta_{n'n}\delta_{m'm}.$$

This is what we want

First error done. Poor data sampling

Test source : heralded single photon source

Train source: attenuated laser

The reconstruction fidelity slightly degraded — we observed

1. *Fidelity(nn) = 0.86 ± 0.04. Purity(nn) = 0.84 ± 0.04*
2. *Fidelity(MLE) = 0.81 ± 0.05, Purity (MLE) = 0.75 ± 0.07*

The Purity output of the NN is always quite high compared to MLE!

The most likely reason for this is some non-uniformity of the datasets caused by experimental drifts — the data for heralded single photons were taken after some period of time. We believe, the performance may be recovered if we use heralded photons data for training as well, using a much larger amount of data.

Backstory: they only sampled from a strip of the whole Bloch sphere. Further, they didn't collect enough data, so the network couldn't generalize. (we will see later the reason why)

Process and measurement errors

Gouy dephasing and cross talk

Data and Projection basis

For a pure state $\in \mathcal{H} = \mathbb{C}^6$

1)we employed a set of d^2= 36 Simmetrically informationallly complete- POVM operators. SIC-POVM form a minimal set of rank-1 projectors with equal pairwise Hillbert Schmidt inner product

$$Tr(M_i M_j) = 1 \iff i = j, \frac{1}{d} \quad i \neq j$$

$$\frac{1}{d} \sum_i M_i = 1$$

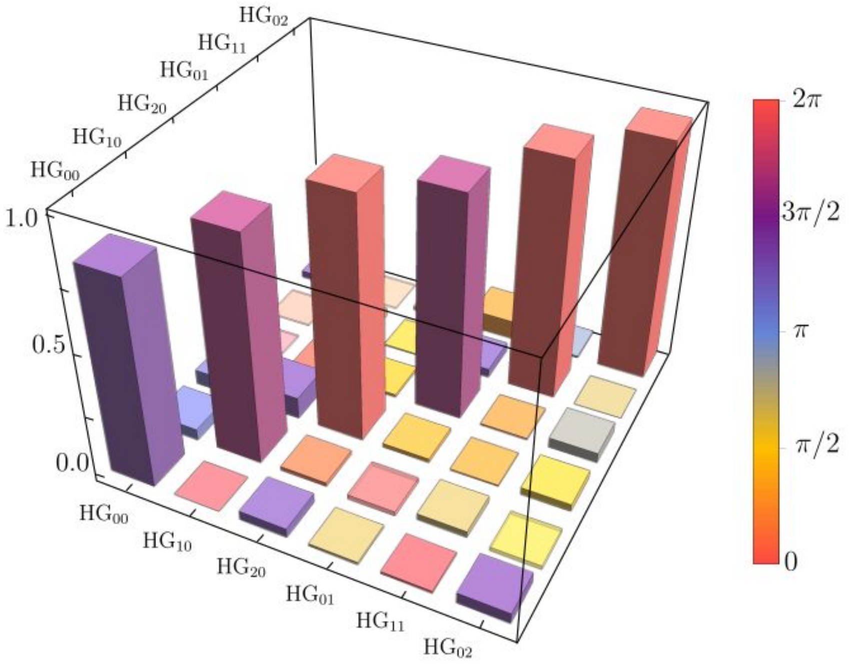
2) We generate random Haar state, and collect their probabilities outcomes. *This si called POVM based Neural Network approach*

Krauss element

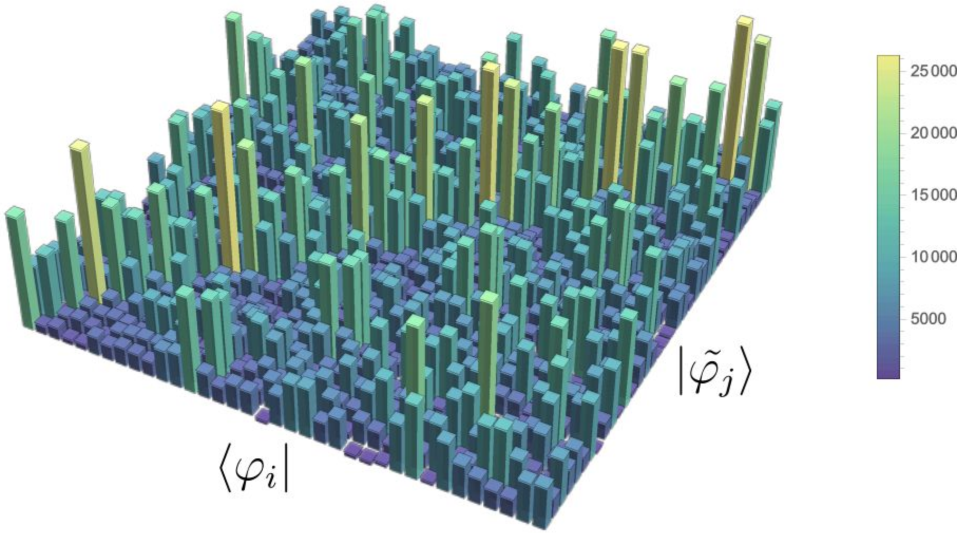
The quantum process tomography reduces to

$$\mathbb{P}(\gamma|\alpha,\mathcal{E}) = Tr(M_{\alpha\gamma}\mathcal{E}(\rho_\alpha)) = Tr\Big(\sum_{k=1}^{K} M_{\alpha\gamma}E_k\rho_\alpha E_k^\dagger\Big).$$

Experimentally reconstructed first operator element E1 of the process E associated with the spatial state evolution between the preparation and measurement stages. The matrix elements are expressed in Hermite-Gaussian modes basis. Ideally, it should be an identity matrix, but additional phase-shifts from Gouy phase-shift make up our dephasing operator

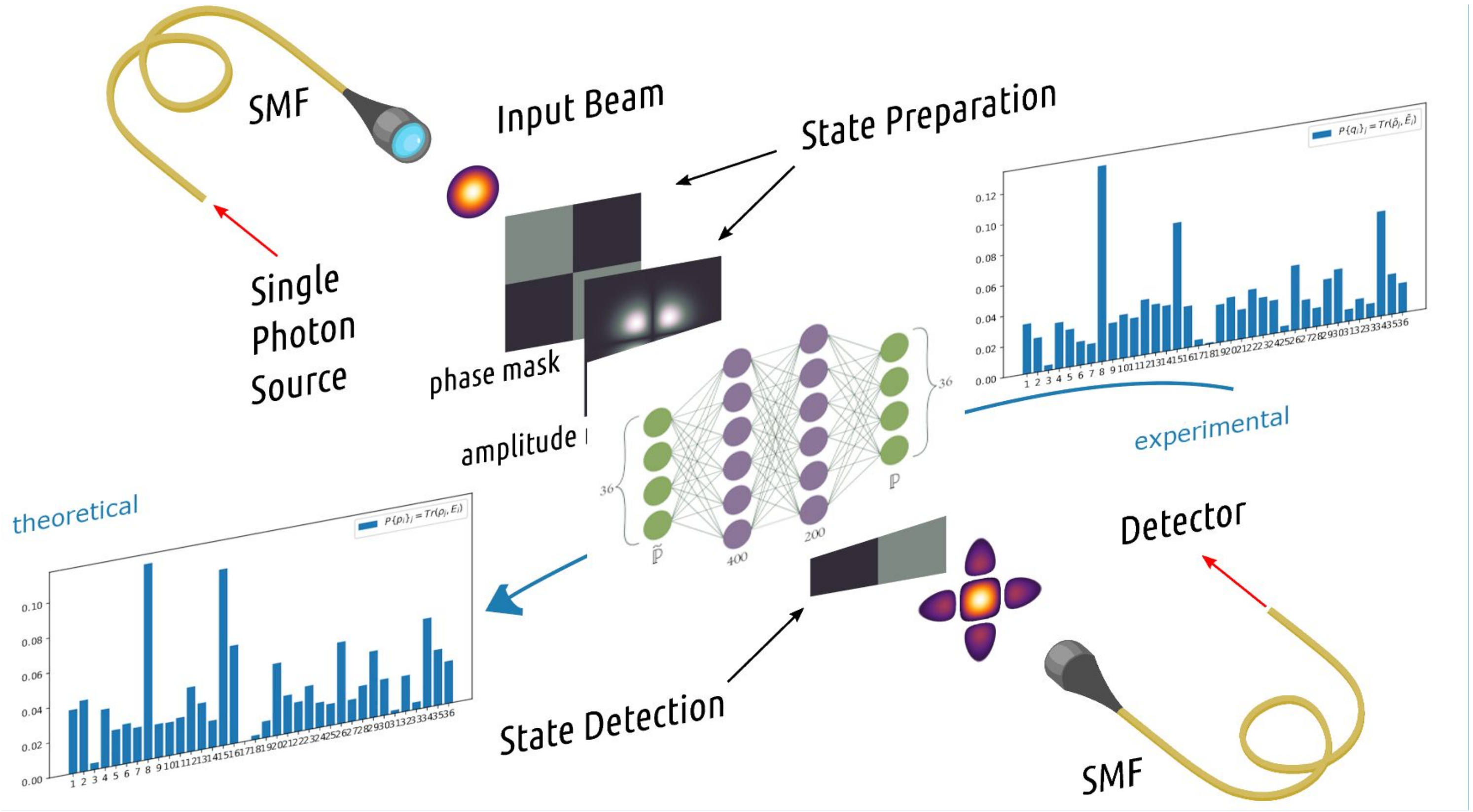


Cross talk- Measurement error



Experimentally measured cross-talk probabilities $P = |<\phi_i|\phi_j>|^2$

for the projectors from the SIC POVM without the
Gouy phase correction



The Neural Network post processing

Universal Approximation Theorem Extension to simplicial complexes

Definition: A multi-layer feed-forward network defined on a real-valued n-dimensional space is a function $\mathcal{N}: \mathbb{R}^n \to \mathbb{R}^m$ such that, for each $x \in \mathbb{R}^n$, $\mathbb{N}(x)$ is the composition of k + 1 functions

$$\mathbb{N}(x) = f_{k+1} \circ f_k \circ \ldots \circ f_1$$

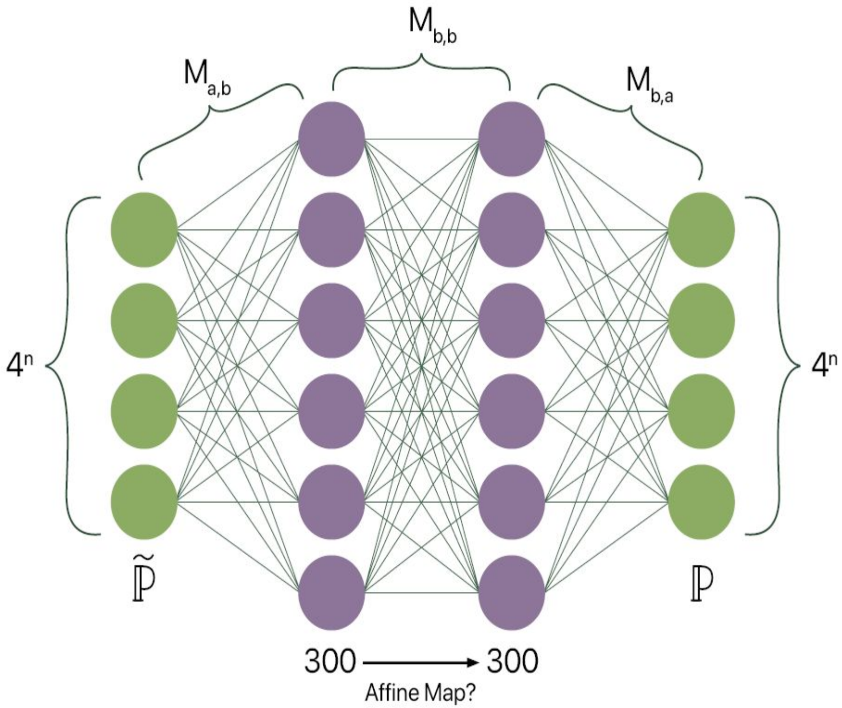
where $k \in \mathbb{Z}$ is the number of hidden layers, k ≥ 1, and, for 1 ≤ i ≤ k + 1, $f_i : \mathbb{R}^{d-1} \to \mathbb{R}^d$ is defined as

$$f_i(y)) : \phi_i(W^i; y, b_i)$$

W^i being a real-valued d−1 × d matrix, (that is, $W^i \in \mathcal{M}_{d-1,d}$), $b_i \in \mathbb{R}^{d_i}$ the bias term, and ϕ_i a bounded, continuous, and non-constant function, called activation function.

Theorem: Let us consider a simplicial map $\phi_c : |K| \to |L|$ between the underlying space of two finite pure simplicial complexes K and L. Then a two-hidden-layer feedforward network N_ϕ such that ϕ_c = N_ϕ(x) for all $x \in |K|$ can be explicitly defined.

What am I looking for? The cost function.



for discrete probability distribution p,q on the same probability space $\mathcal{X}$ we define the KL divergence

$$D_{KL} = \sum_{x\in\mathcal{X}} p(x)\frac{p(x)}{q(x)} = H(p,q) - H(p)$$

But cross-entropy is related to log-likelihood

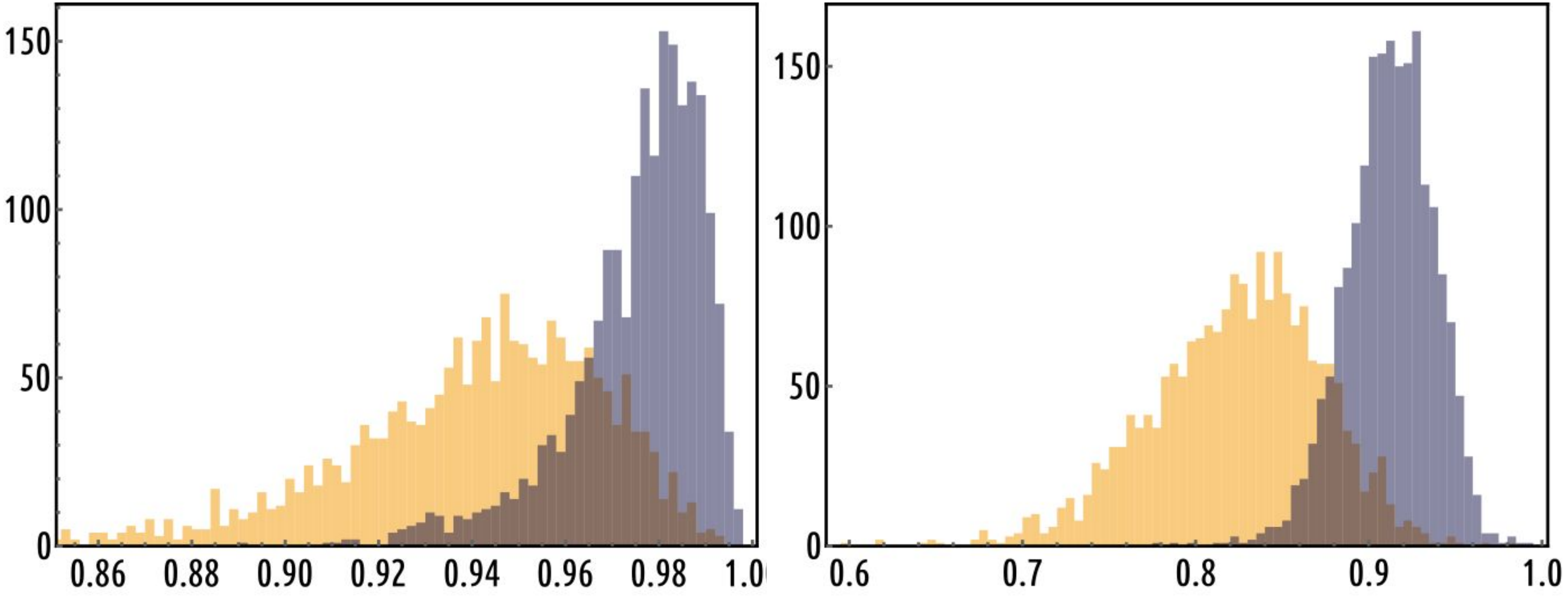
$$-H(p,q) = \sum_i p_i log(q_i) = \frac{1}{N} log \prod_i q_i^{Np_i}$$

So maximize log likelihood means minimize KL divergence.
Cue: in our case N=1.

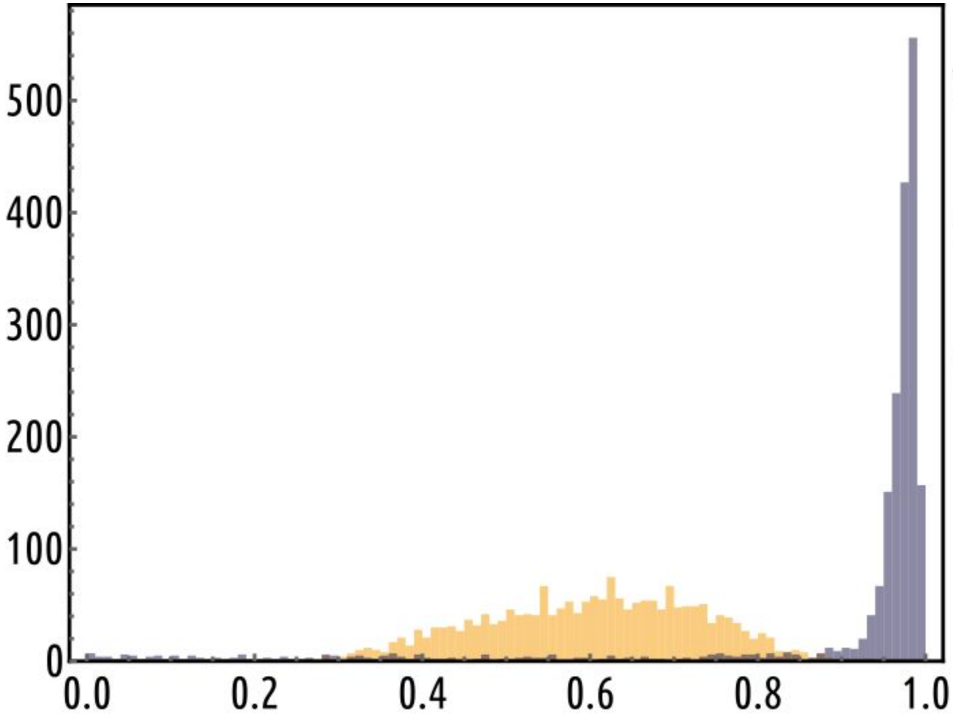
OUTCOMES

TAkeaway: 80% of ML applications failed.
Why is important to make your mind up, and pay attention to the data.

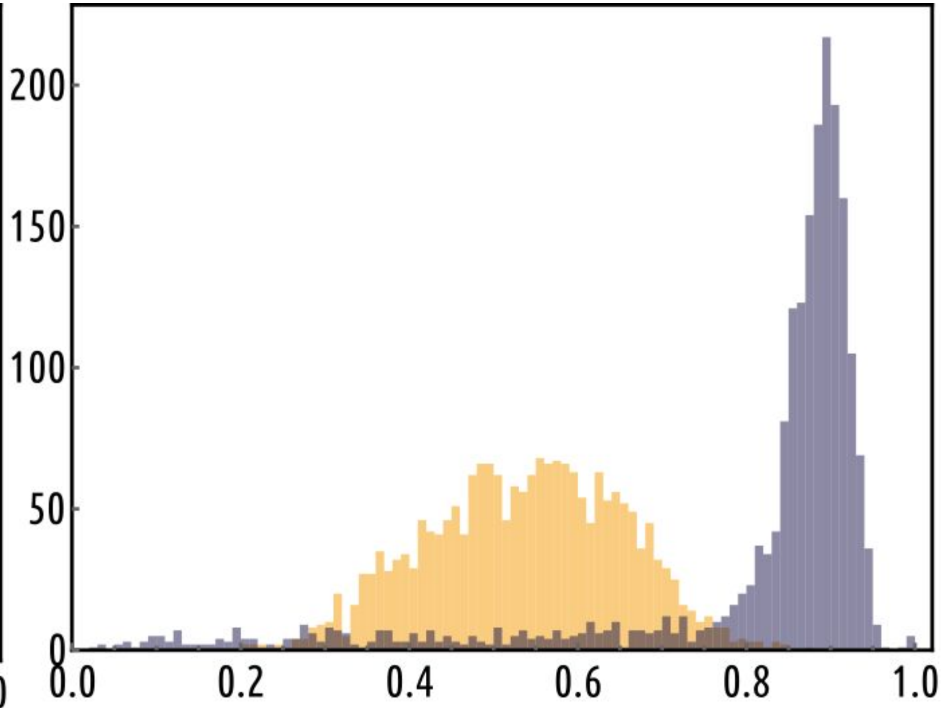
With Gouy post-processing, classical Fidelity reconstruction



Without Gouy phase post processing



Pure. F = 0.89+/-0.22



Mixed. F = 0.81+/- 0.19

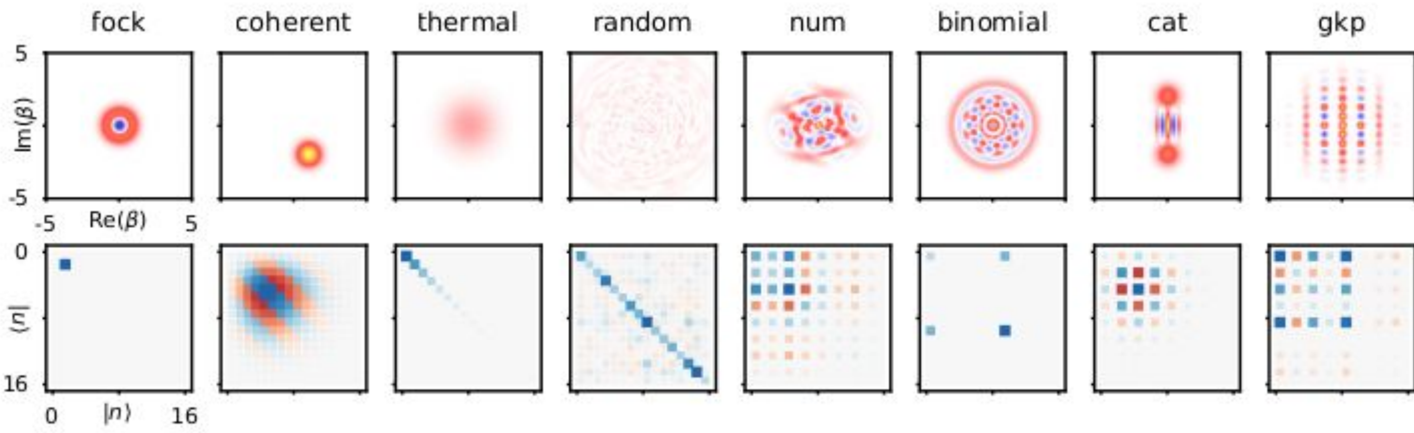
But our model scales polynomially.

the (C)GAN Extension

States employed:

1. Fock
2. Coherent
3. Thermal
4. Num
5. Binomial
6. Cat

Either mix and pure.



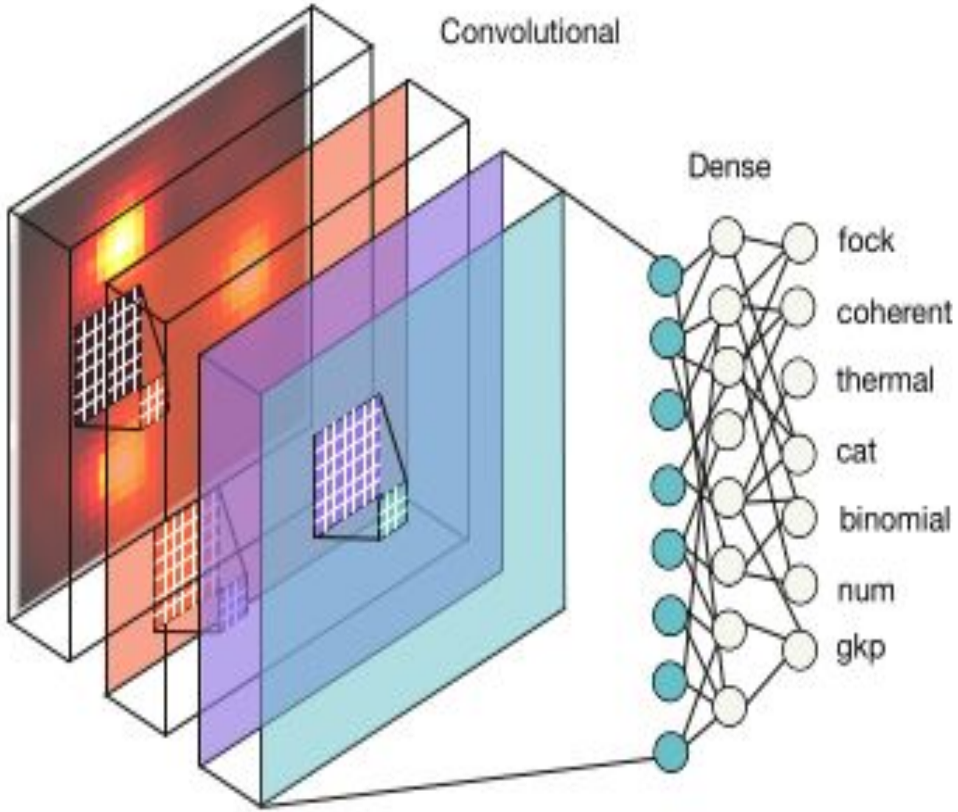
Measurement scheme:
displace-and-measure technique.

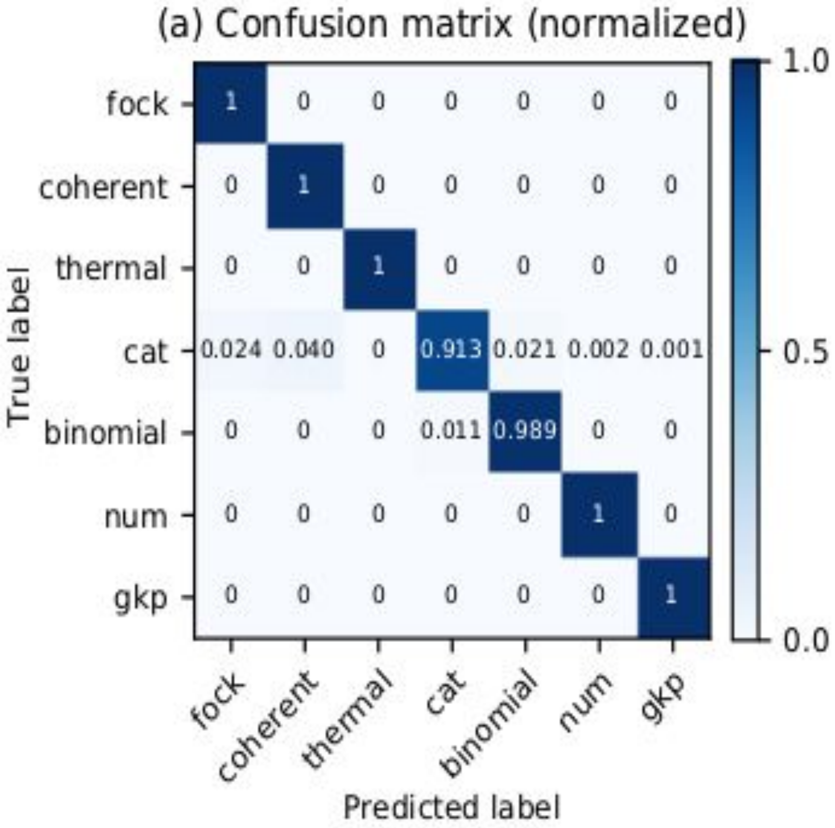
$$Q_n^\beta = tr(|n><n|)D(-\beta)\rho d(\beta)$$

$$(\text{Wigner})W(\beta) = \frac{2}{\pi}\sum_n(-1)^n Q_n^\beta$$

measurement with Husimi Q function

$$Q(\beta) = \frac{1}{\pi}Q_0^\beta$$


Convolutional
Dense
fock
coherent
thermal
cat
binomial
num
gkp


(a) Confusion matrix (normalized)
True label
Predicted label
fock
coherent
thermal
cat
binomial
num
gkp
1 0 0 0 0 0 0
0 1 0 0 0 0 0
0 0 1 0 0 0 0
0.024 0.040 0 0.913 0.021 0.002 0.001
0 0 0 0.011 0.989 0 0
0 0 0 0 0 1 0
0 0 0 0 0 0 1
1.0
0.5
0.0