Impact of Noise on Deep Neural Networks for Wavefront and PSF Estimation

Jeffrey Smith[1], Jesse Cranney[2], Charles Gretton[1], Damien Gratadour[3]

1 - School of Computing, Australian National University, Australia
2 - Advanced Instrumentation Technology Centre, Research School of Astronomy and Astrophysics, Australian National University, Australia
3 - LESIA, Observatoire de Paris, University PSL, CNRS, Sorbonne Universite, Universite de Paris, 5 place Jules Janssen, 92195 Meudon, France

November 2023

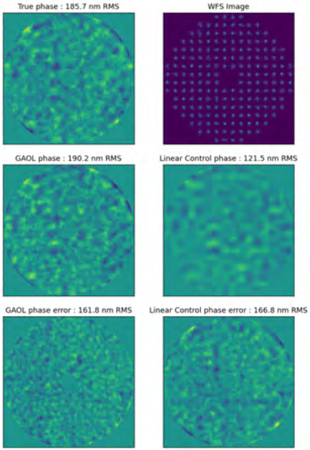
Table of Contents

Table of Contents

Australian National University

l'Observatoire de Paris

Motivation



- Larger subapertures in LGS/NGS SH-WFS $\rightarrow$ better sky coverage + SNR,
- Centroid-based reconstruction discards everything above tip/tilt in a subaperture,
- Image-to-image translation allows us to fetch the remaining useful information, approaching the limit of SH-WFSs.

2022→2023

- Nice performance in simulation with cGAN and UNet reconstruction for:
  - AO control (Smith+ UAI 2022, Pou+ SPIE 2022),
  - PSF-R (Smith+ SPIE 2022, Smith+ JATIS 2023).
- Main questions raised were:
  - How do we know what the network is doing? - **ML Black Box**
  - What are the effects of **noise** on the estimates?
  - What are the **limits** of these techniques?
- Turned to a statistical analysis, to learn the limits:
  - Wavefront decomposition using Karhunen–Loève (KL) modes,
  - Analysis of noise impact in E2E simulations.

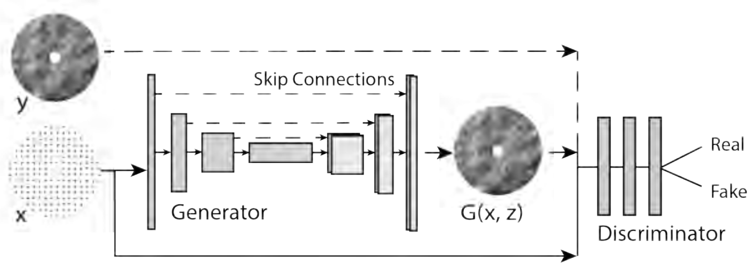
Australian National University l'Observatoire de Paris

Generating Training Data

- Issue - we never truly know the wavefront when on-sky.
  - supervised learning requires "truth" wavefront,
  - For now, simulate with sophisticated E2E AO simulation software,
  - In future, we can use an SLM or DM to generate training data on the bench.
- COMPASS - COMputing Platform for Adaptive optics SystemS
  - Unlimited data for training / analysis - unique seeding of atmosphere
  - Python API - easy integration with pytorch workflows
  - Easily configurable for AO design / simulation tasks

conditional Generative Adversarial Network (cGAN)

Network Design[1]:



cGAN Components

- UNet Generator Network
- Patch GAN Discriminator
- Dropout noise (z)

[1] [Isola 2017]

Australian National University
l'Observatoire de Paris

cGAN Network Loss

- discriminator is punished for missing "fakes" and rejecting "reals",
- generator is punished for getting caught,
- cGAN is extension of UNet,
- i.e., Our UNet is the same cGAN with $\mathcal{L}_{\mathbf{cGAN}}(\mathbf{G}, \mathbf{D})$ loss term set to zero.

$$\mathcal{L}_{\mathbf{cGAN}}(\mathbf{G}, \mathbf{D}) = E_{x,y}[logD(x, y)] + E_{x,z}[log(1 - D(x, G(x, z))] \tag{1}$$

$$G^* = \arg\min_G \max_D \ \mathcal{L}_{\mathbf{cGAN}}(\mathbf{G}, \mathbf{D}) + \lambda\mathcal{L}_{L1}(G) + \lambda_M\mathcal{L}_{L1}(G_M) \tag{2}$$

$$\mathcal{L}_{L1}(G) = E_{x,y,z}[||y - G(x, z)||_1] \tag{3}$$

Table of Contents

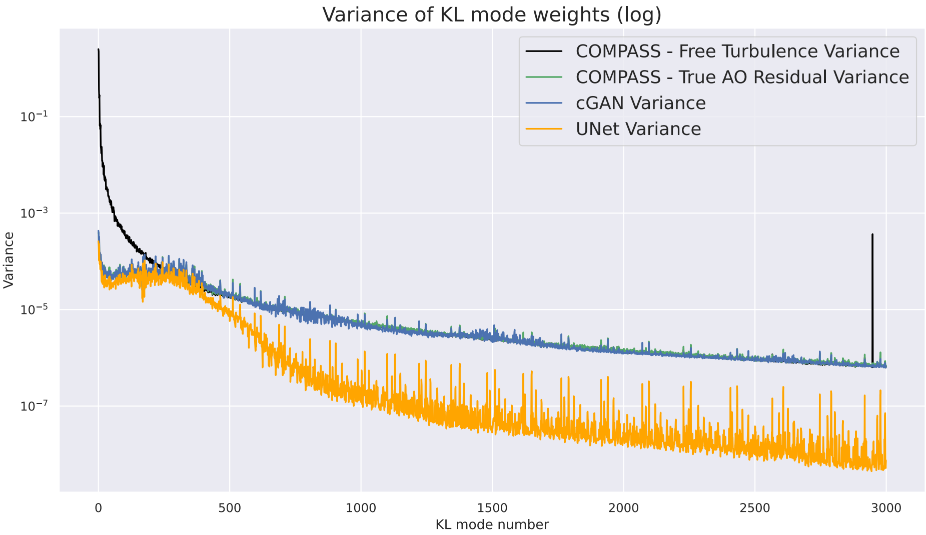
Australian National University

l'Observatoire de Paris

Modal Analysis of UNet and cGAN

- ANNs translate WFS image to estimated WFS phase,
- We compare the variance of this phase to the variance of the true phase,
- Comparison is done in KL mode space, over 20k frames,
- $E[\varphi^2_{\mathrm{truth}}]$ vs $E[\varphi^2_{\mathrm{estimated}}]$

Variance of Estimate



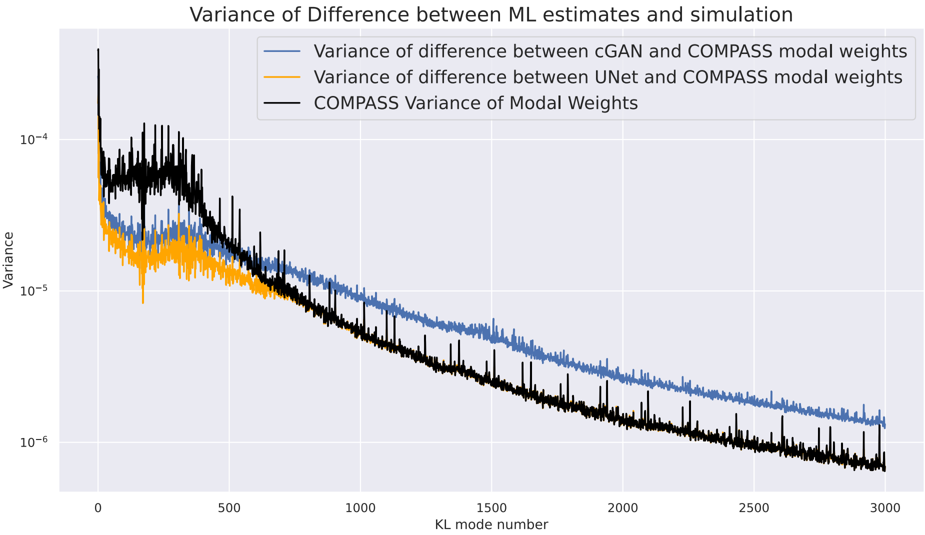
So cGAN is clearly better?

- This is what we thought too, but we should dig a bit deeper,
- Now let's see the variance of the residual:

$$\varphi_{\text{residual}} = \varphi_{\text{truth}} - \varphi_{\text{estimated}}$$

- $E[\varphi^2_{residual}]$ vs $E[\varphi^2_{\text{truth}}]$

Variance of Residual



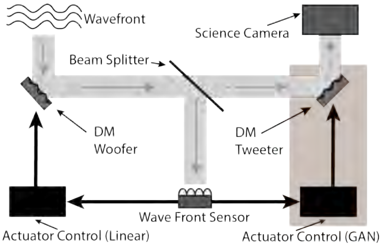
uNet $\neq$ cGAN

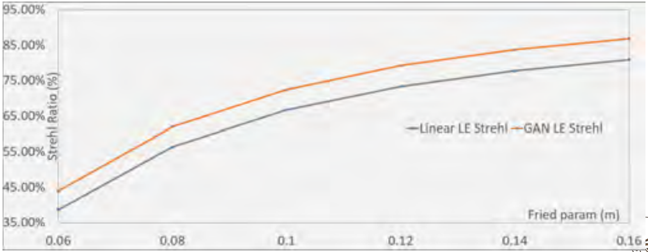
- cGAN perfects the statistics of the phase across all modes, but not always the right value,
- UNet (without discriminator) is more conservative on statistics, but actually has better residuals.

Table of Contents

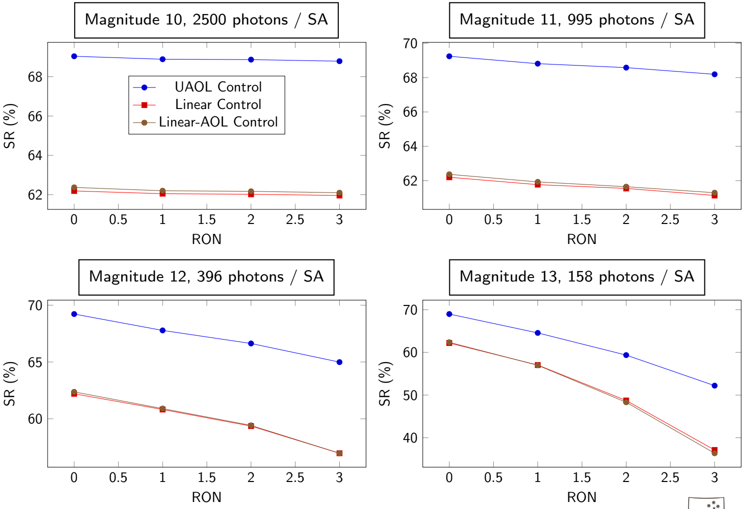
Previously - GAN Assisted Open Loop (GAOL) Control

GAOL performance
with variation of turbulence
vs Linear re-constructor benchmark

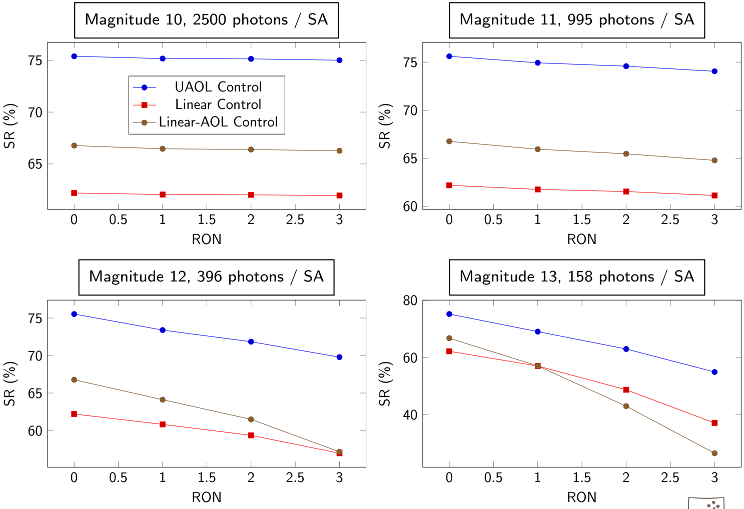




UNet Assisted Open Loop (UAOL) control (with DM shape) vs RON (+0 DM act)



UNet Assisted Open Loop (UAOL) control (with DM shape) vs RON (+7 DM act)



Australian National University

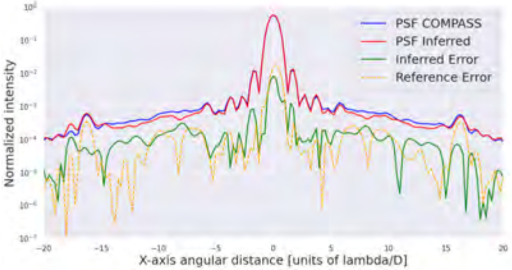
l'Observatoire de Paris

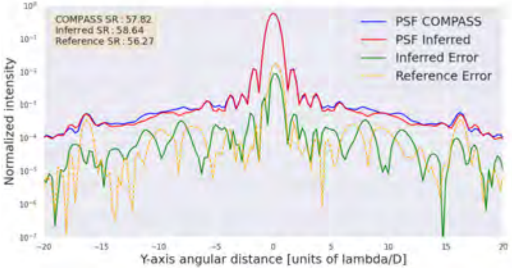
Table of Contents

cGAN Noise analysis for PSF-R

- cGAN performance on PSF-R tasks demonstrate poor performance with noisy data
- Previous great results with bright guide star (Mag 3)
- Noise makes cGAN networks difficult to train
- Dominant noise effect appears to be photon noise

Long Exposure PSF With Noise Off

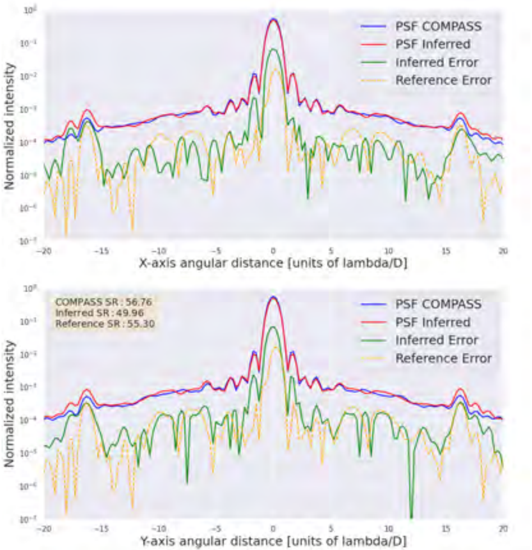




- Noise Off
- 16 x 16 apertures
- 8 x 8 pixels per sub

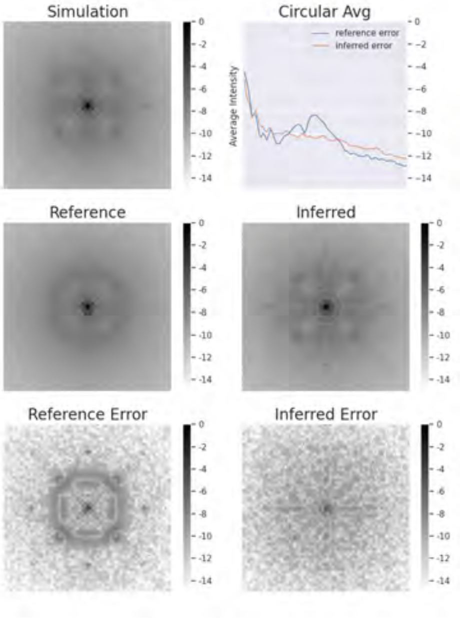
Australian National University
l'Observatoire de Paris

Long Exposure PSF With Noise On

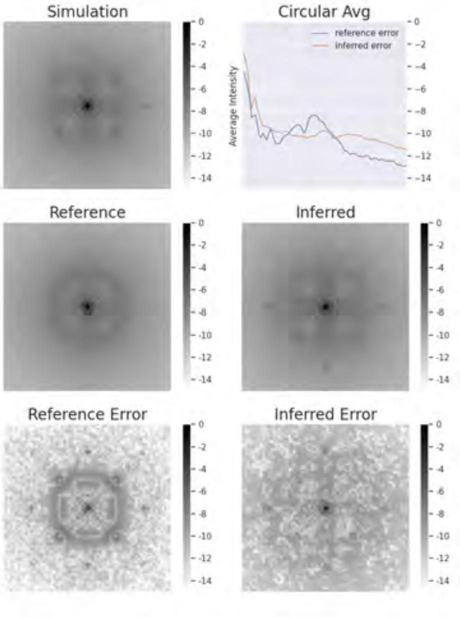


- 1 Photon per pixel RON
- Photon Noise On
- 396 Photons per sub-aperture
- 16 x 16 apertures
- 8 x 8 pixels per sub

Australian National University l'Observatoire de Paris



(a) No Noise - 396 photons /sub



(b) 1 RON, photon noise on

Table of Contents

Discussion

- We have an accurate and robust method for estimating wavefront phase directly from the WFS image for control (UNet) and PSF-R (cGAN) when noise is low,
- With the KL modal analysis, we can see what each ML method is interpreting and generating from the WFS and simulated turnulance,
- UAOL control (UNet) even with reasonably low photon count has excellent robustness to noise, Fried parameter and guide star magnitude in simulated experiments
- cGAN performance and training significantly impacted by noisy data for low photon count, photon noise is dominant effect.

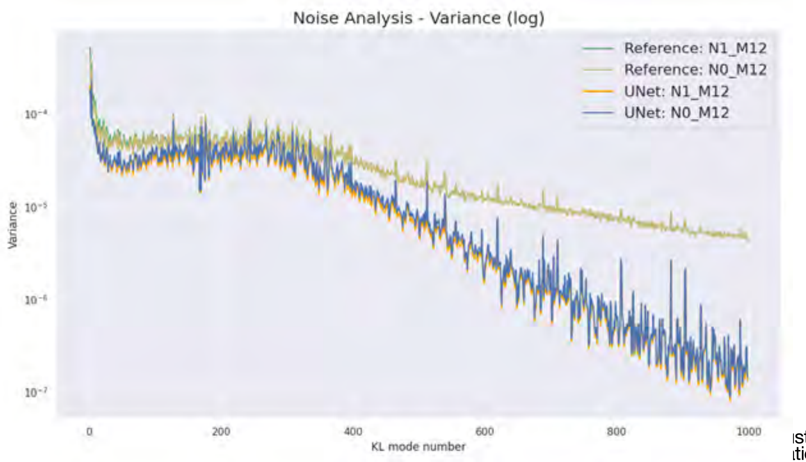
Thank you and Further reading

- Enhanced adaptive optics control with image to image translation Jeffrey Smith, Jesse Cranney, Charles Gretton, Damien Gratadour; Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence, PMLR 180:1846-1856
- Jeffrey Smith, Jesse Cranney, Charles Gretton, and Damien Gratadour "Image-to-image translation for wavefront and PSF estimation", Proc. SPIE 12185, Adaptive Optics Systems VIII, 121852L (29 August 2022); https://doi.org/10.1117/12.2629638
- Jeffrey Smith, Jesse Cranney, Charles Gretton, Damien Gratadour, "Image-to-image translation for wavefront and point spread function estimation," J. Astron. Telesc. Instrum. Syst. 9(1) 019001 (19 January 2023) https://doi.org/10.1117/1.JATIS.9.1.019001
- B. Pou, J. Smith, E. Quinones, M. Martin, D. Gratadour, "Model-free reinforcement learning with a non-linear reconstructor for closed-loop adaptive optics control with a pyramid wavefront sensor," Proc. SPIE 12185, Adaptive Optics Systems VIII, 121852U (29 August 2022); https://doi.org/10.1117/12.2627849
- https://github.com/GANs4AO/I2IT4AO

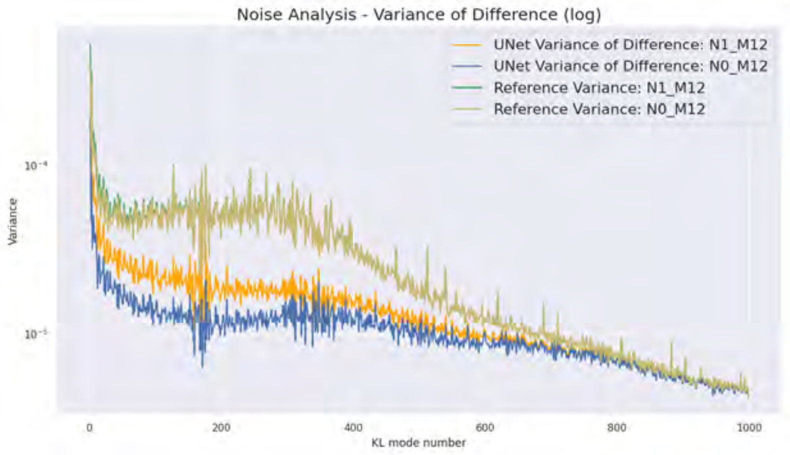
Australian National University

Appendix

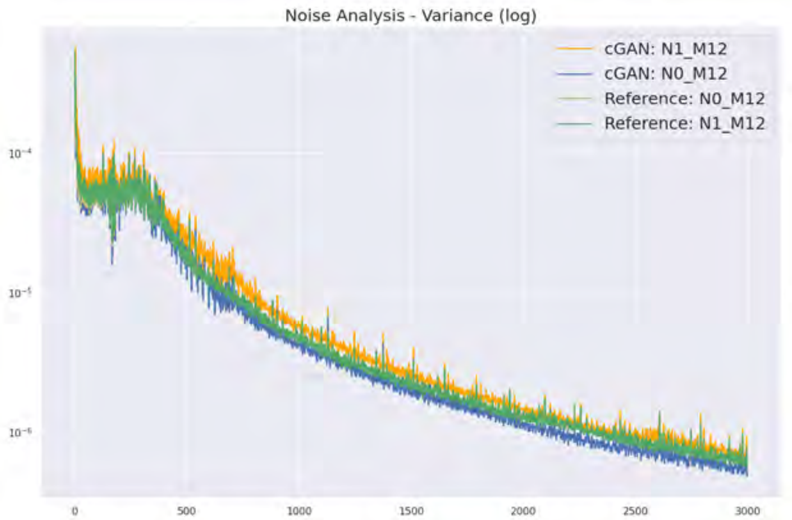
Performance of UNet with Noisy data - Modal Weight variance comparison



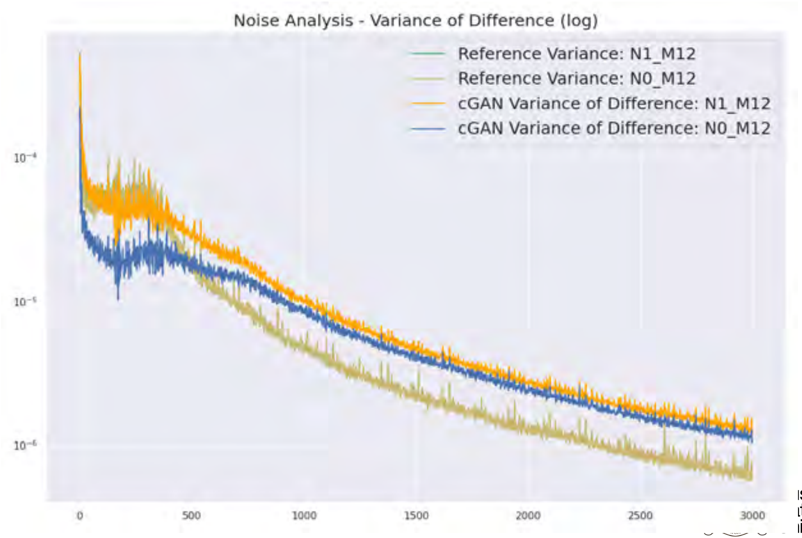
UNet Noise - Variance of the difference



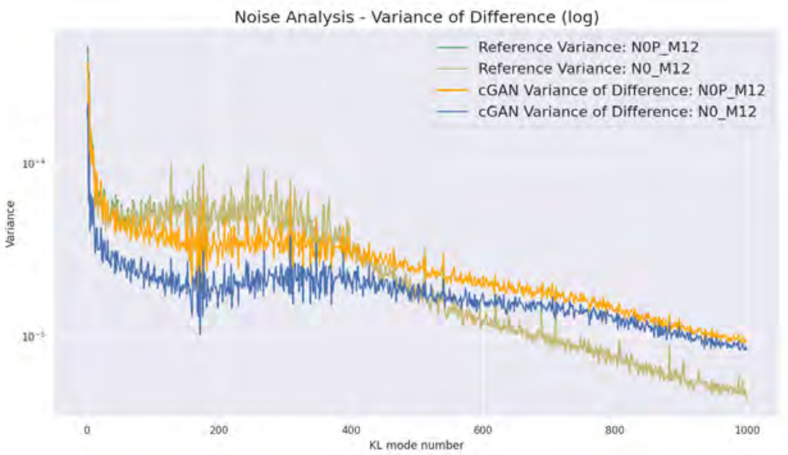
cGAN Noise - Modal Weight variance comparison



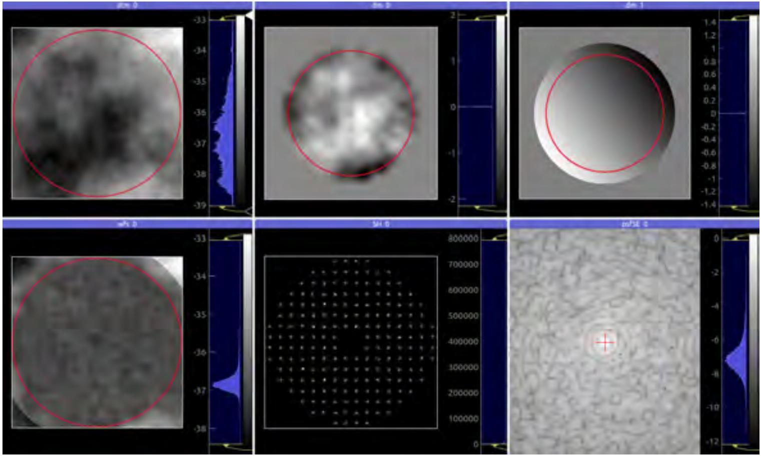
cGAN Noise - Variance of the difference



cGAN Noise - Photon Noise



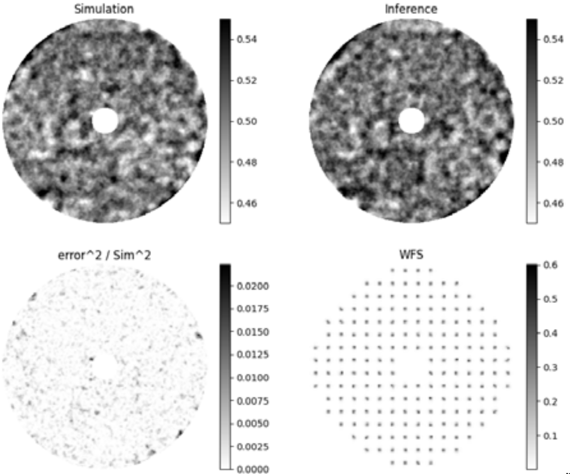
COMPASS GUI - example data



cGAN inference (mild turbulence)

Inferred result for cGAN vs Simulation ground truth residual phase

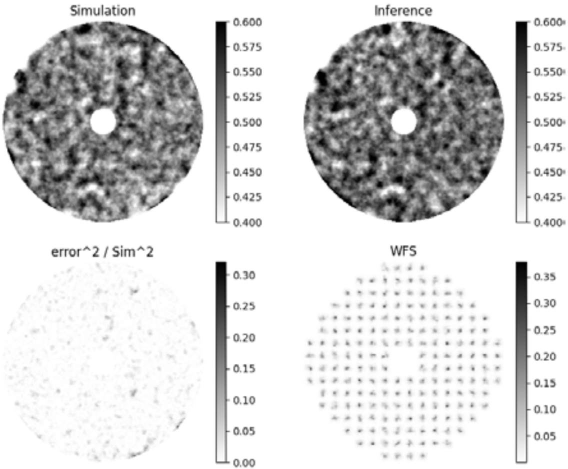
- Note the SH-WFS spots for phase with milder turbulence
- A single trained network is robust over the full range of expected turbulence ($r_0 = [0.06m, 0.16m]$)



cGAN inference (strong turbulence)

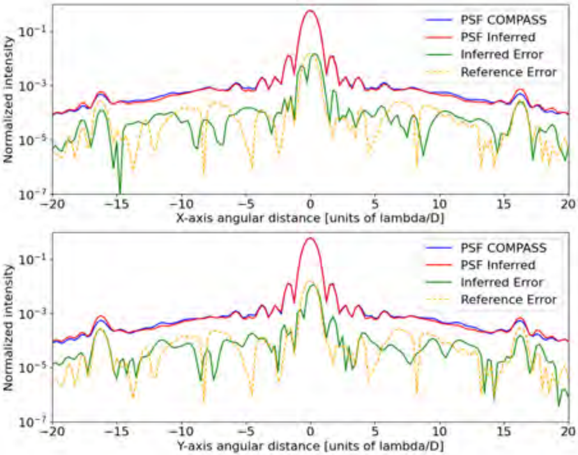
Inferred result for cGAN vs simulation ground truth residual phase

- Note the SH-WFS spots for phase with stronger turbulence
- Clearly high frequency features are captured



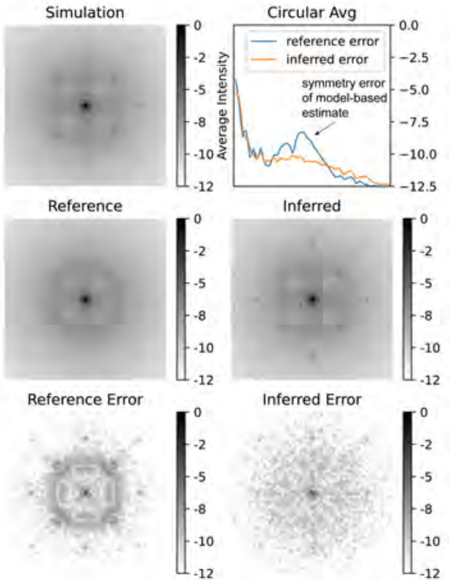
Australian National University l'Observatoire de Paris

Long Exp. PSF from cGAN – Split View



$r_0 = 0.093m$

Long Exp. PSF from cGAN - Circular Avg.



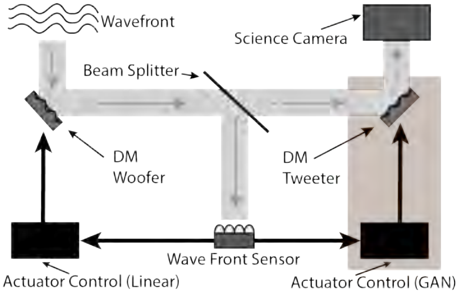
- Data driven method captures features missed by the reference statistical model
- Symmetry error correction of a few orders of magnitude
- Important for tasks such as exo-planet detection

GAN Assisted Open Loop Control (GAOL)

- Now that we have a method of estimating wavefront phase with a cGAN, we can apply this to AO control
- However, modifying the AO estimation in closed loop will alter the data our cGAN was trained on.
- Solution - apply secondary corrections from the cGAN estimates in open loop with an independent DM.
- This a relatively small change to a typical closed loop, with only one additional DM required.

GAOL AO design

- Highlighted second control step in open loop augments the closed loop design
- The 'Woofer' DM applies linear control applying low frequency correction
- The 'Tweeter' DM applies higher frequency corrections (cGAN) in open loop, which is not fed back to the WFS.

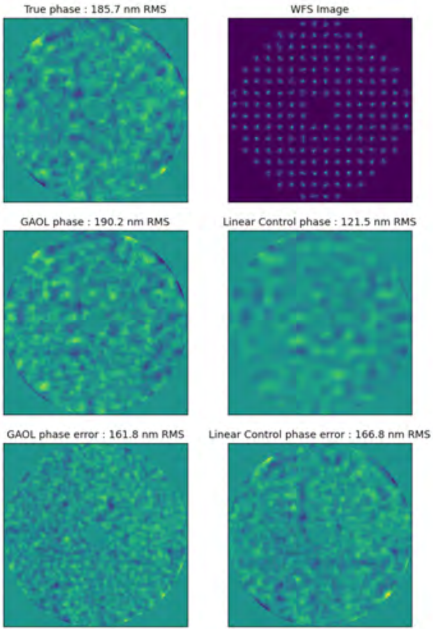


GAOL AO - control law

- The 'Woofer' DM uses a linear controller, using the control law below.
- The 'Tweeter' DM is controlled by the cGAN estimates using the same control law, however there is no feed back in this case.
- Both mirrors combine estimates with the previous iteration control solution controlled by the gain (g)

$$u_0 = \mathbf{0}, \quad u_k = (1-g)u_{k-1} + gRDu_{k-2} + gRs_k \tag{4}$$

$$u_0^{\mathrm{nl}} = \mathbf{0}, \quad u_k^{\mathrm{nl}} = (1-g^{\mathrm{nl}})u_{k-1}^{\mathrm{nl}} + g^{\mathrm{nl}}R^{\mathrm{nl}}\hat{y}_k \tag{5}$$



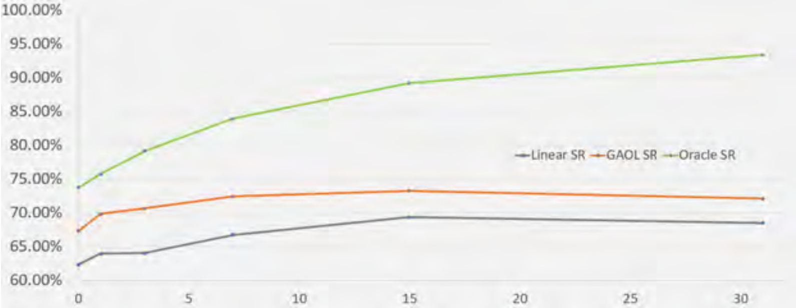
GAOL Phase Comparison

- Contrast with linear control
- Single iteration comparison for the same input data after 2000 frames
- Clear out-performance in GAOL over purely linear control

Australian National University
l'Observatoire de Paris

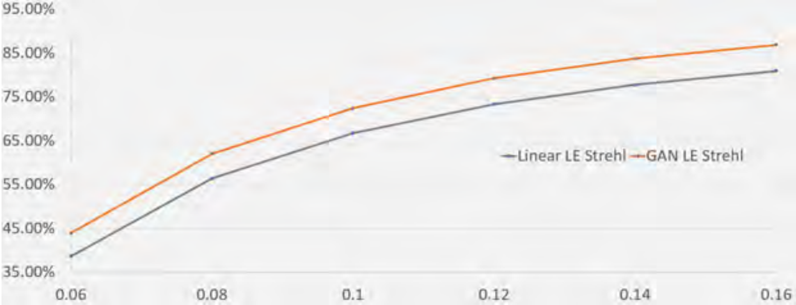
GAOL - actuator density

GAOL performance (Long Exposure SR) for increased actuator count vs Linear reconstructor and Oracle benchmarks



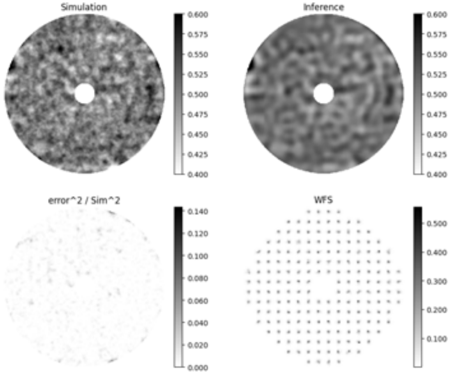
GAOL - robustness to turbulence

GAOL performance (Long Exposure SR) with variation of turbulence (Fried parameter) vs Linear reconstructor benchmarks (+ 7 actuators)

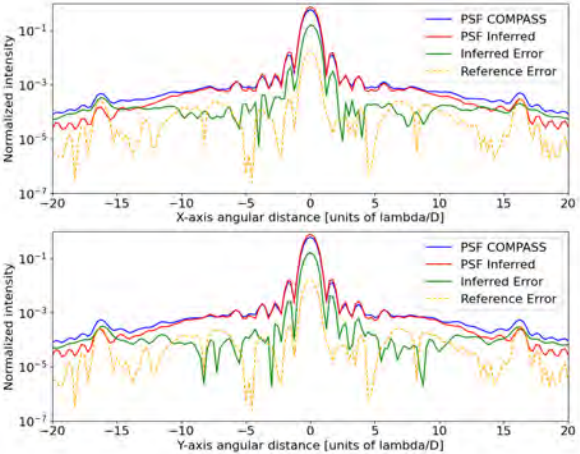


UNet inference

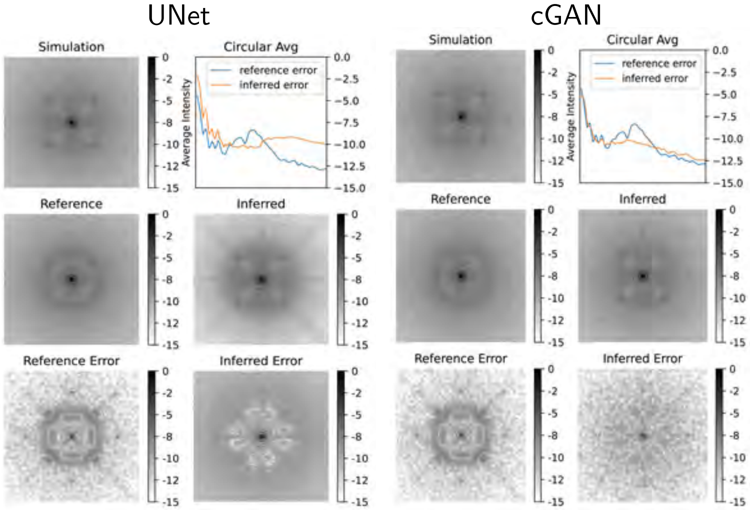
- In training sample inference from UNet
- Notice the lack of cGAN loss creates blurry, low frequency phase estimates



Long Exp. PSF from UNet – Split View



UNet vs cGAN - Circular Avg.



PSF from Wavefront Phase

- Point Spread Function (PSF) can be directly calculated from the wavefront phase.
- This process is not reversible, so phase estimation provides additional opportunities over estimating the PSF directly

$$\mathrm{PSF} = |\mathrm{FFT}(\mathrm{amplitude} \cdot \mathrm{e}^{\mathrm{i} \cdot \mathrm{phase}})|^2 \tag{6}$$

Training Parameters (COMPASS)

Telescope Parameters	
Diameter	8 m
Simulated Atmospheric Parameters	
Number of Layers	1
r_0	0.093 to 0.400 m
Wind Velocity	10 ms^{-1}
Target Parameters	
Wavelength λ_t	1.65 μm
WFS Parameters	
Number of sub-apertures	16 x 16 x 8pix
Wavelength λ_{wfs}	0.5 μm
AO Parameters	
Loop frequency	500 Hz
Delay	2 frames
Integrator Gain	0.4
DM Parameters	
Number of DM actuators	17 x 17
1 tip-tilt mirror	

SNR conversion table

Table: Relative SNR to guide star magnitude for test geometry

Readout Noise	Guide Star Magnitude					
	10	11	12	13	14	15
0	6.25	3.94	2.49	1.57	0.99	0.63
1	6.17	3.82	2.31	1.32	0.70	0.33
2	5.95	3.52	1.94	0.97	0.44	0.19
3	5.63	3.14	1.59	0.73	0.31	0.13

Table: Relative photon count to guide star magnitude for SH-WFS with 16 x 16 sub-apertures and 8 x 8 pixels per sub-aperture

Guide Star Magnitude	Photons per sub-aperture	Photons per pixel
10	2500.00	39.06
11	995.27	15.55
12	396.22	6.19
13	157.74	2.46
14	62.80	0.98
15	25.00	0.39

Training Parameters (GAN)

Generator (UNet)	
Convolutional Layers	8
Discriminator	
Convolutional Layers	3
Training Data	
Image pairs	350000
Image size	512x512pix (padded)
hyper-parameters	
Lambda (λ)	150
Lambda-Masked (λ_M)	30
Batch Size	1
Epochs	65