# Blackbox Optimization using LSTMs

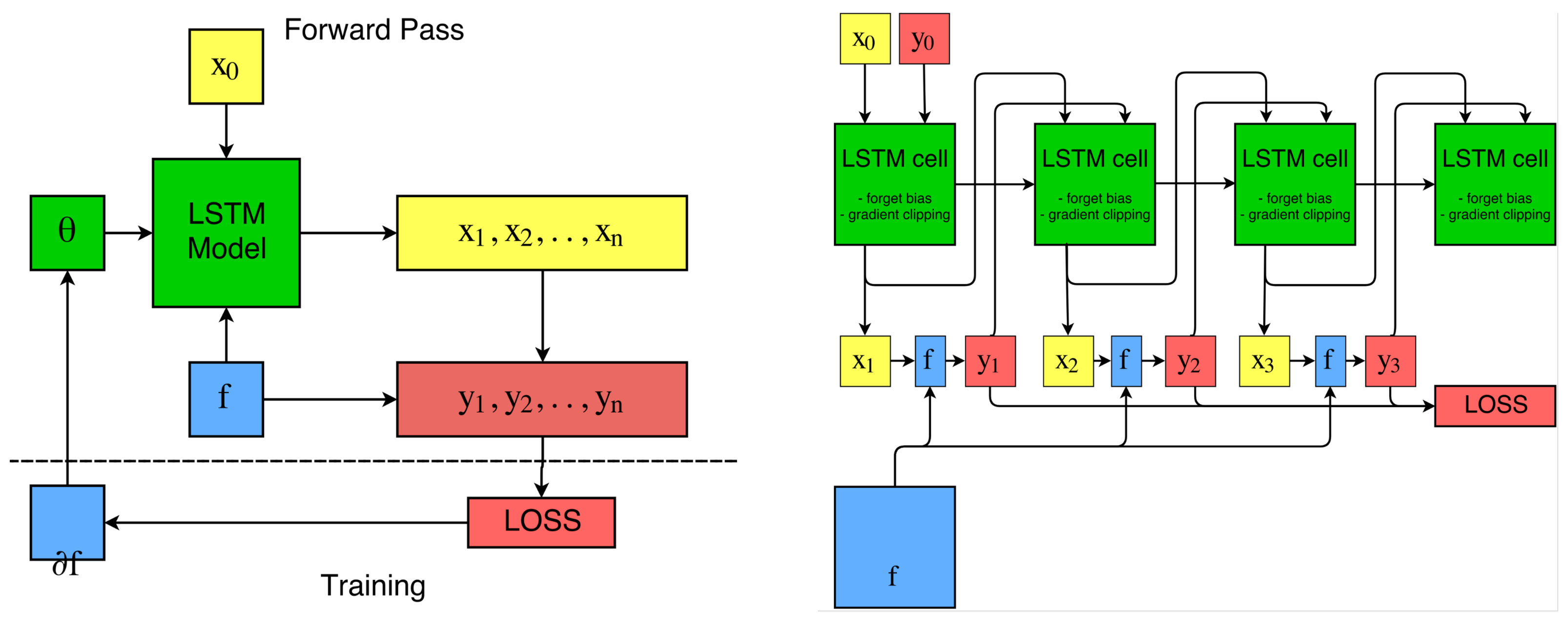
Felix Sattler, Pattarawat Chormai, Raphael Holca-Lamarre

## Summary

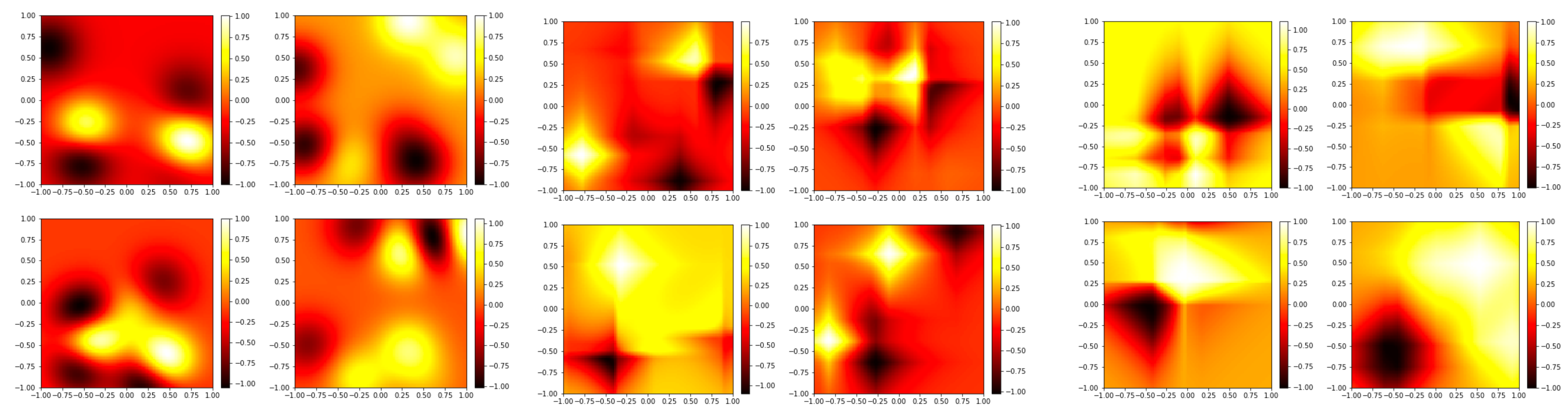
We present a new, learning-based, approach to global Black Box optimization. We train a LSTM to output a sequence of sample points that seek to minimize a given objective function.

Our experiments show that the learned model is able to generalize to a variety of synthetic Black-Box functions, as well as to the real world problem of airfoil optimization. Moreover, our model performs comparable to state-of-the-art Black-Box optimization algorithms w.r.t. the minimum function value found and even outperforms them w.r.t. computation time.

## Model




## Training Data



RBF Matern32 Airfoil Prior

GP functions on different kernels

## Comparison of Loss Functions

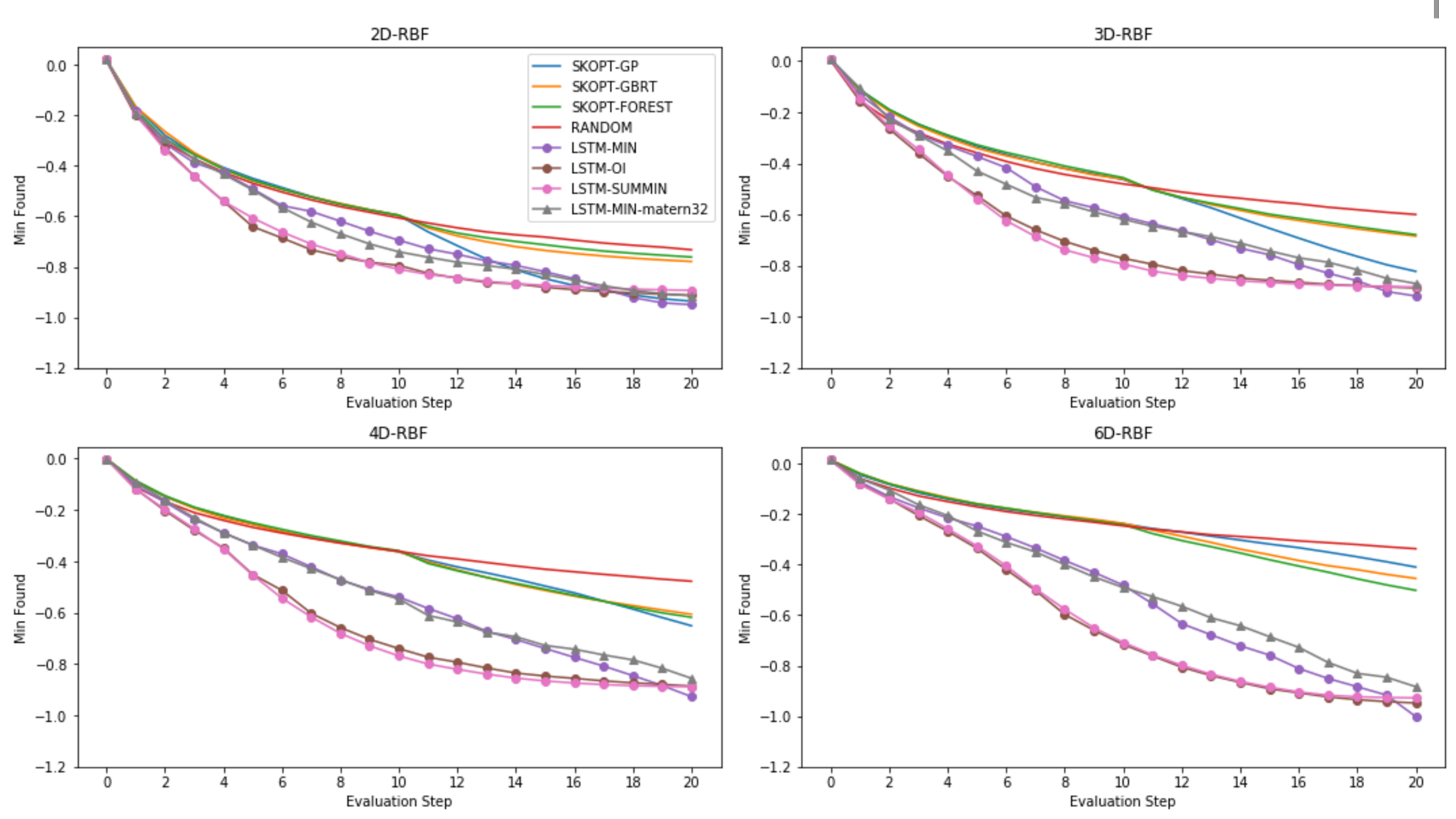
$$L_{\mathrm{SUM}}(\theta) = \sum_{k=1}^{n} f(x_k), \quad L_{\mathrm{OI}}(\theta) = \sum_{k=1}^{n} \max(0, f(x_k) - \min_{i<k} f(x_i))$$

$$L_{\mathrm{MIN}}(\theta) = \min_{k=1,..,n} f(x_k), \quad L_{\mathrm{WSUM}}(\theta) = \sum_{k=1}^{n} w_k f(x_k)$$

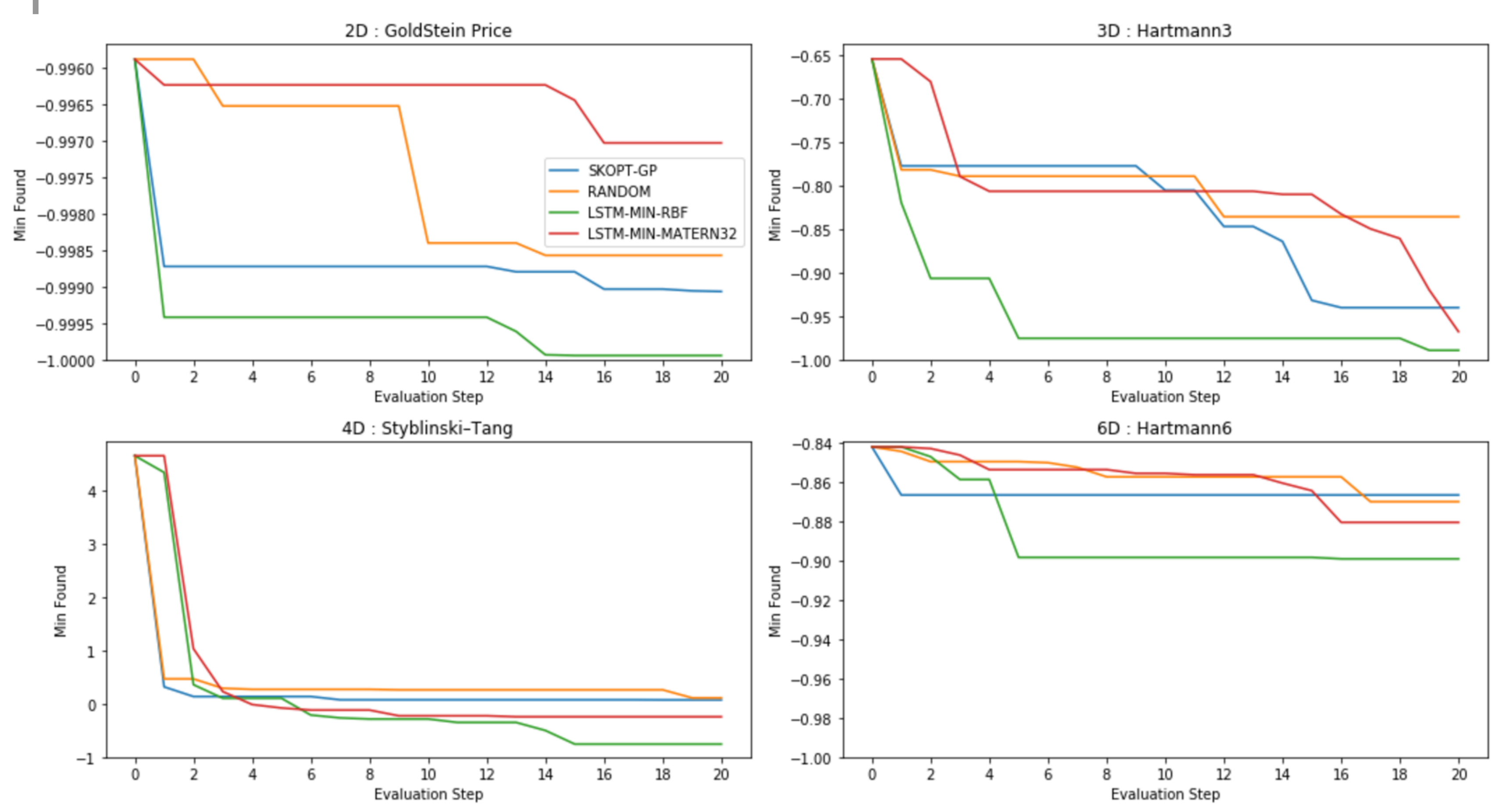
## References

[1] Andrychowicz, Marcin, et al. "Learning to learn by gradient descent by gradient descent." Advances in Neural Information Processing Systems. 2016.
[2] Mockus, Jonas. "The Bayesian approach to global optimization." System Modeling and Optimization (1982): 473-481.
[3] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." Neural computation 9.8 (1997): 1735-1780.

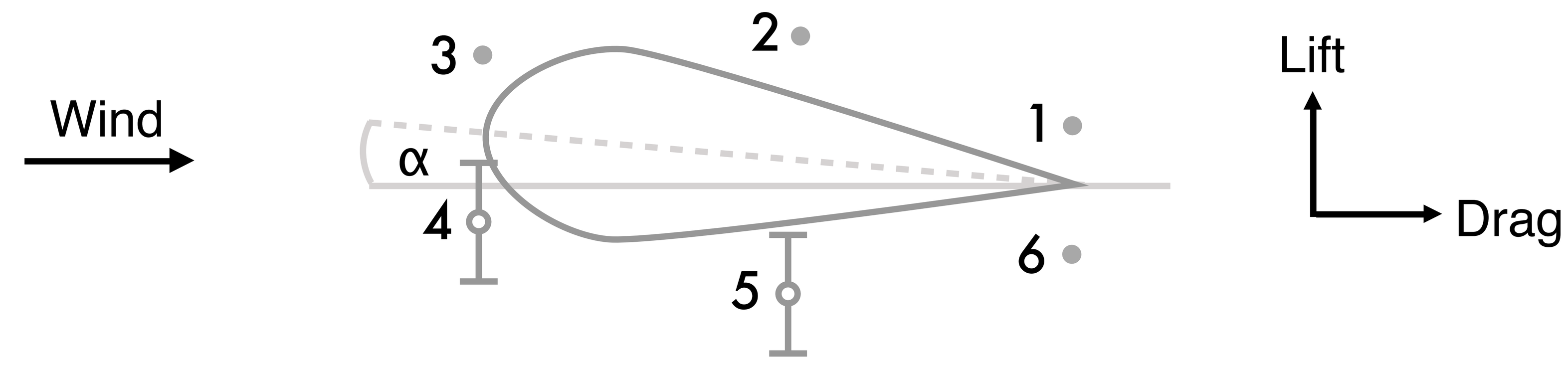
## Results on Test Functions



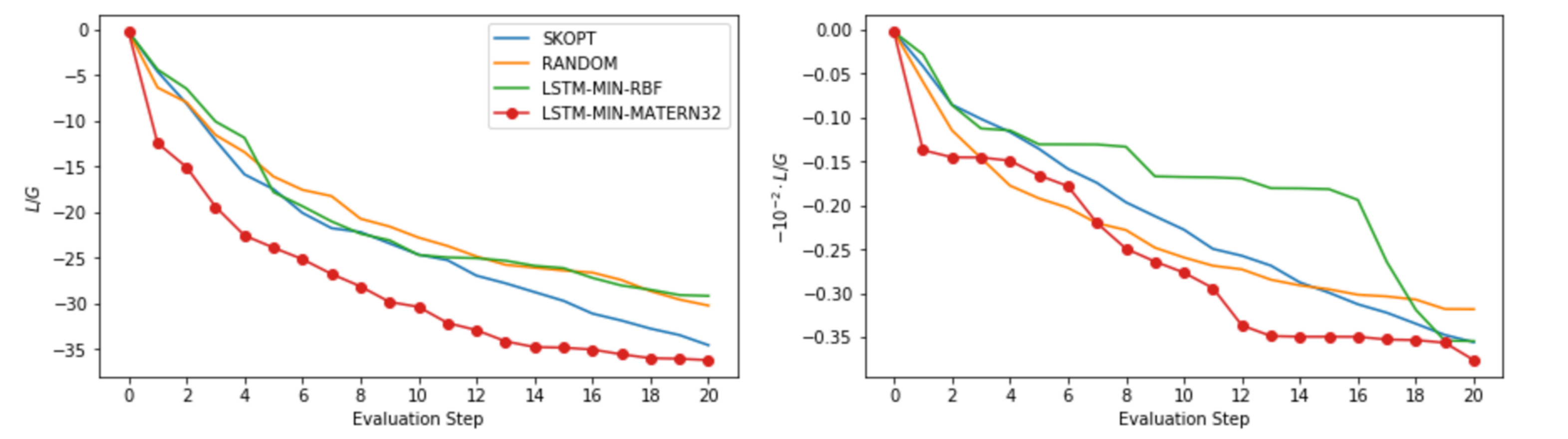

## Results on Benchmask Functions




## Results ( on Airfoil data )




**Goal :** $\max \frac{L}{G} \approx \min -\frac{L}{G}$




## Conclusions

- Our model learns by itself to trade-off exploration and exploitation during sampling; we can guide its behavior by the choice of the loss function we use during training
- Our model performs well also in high dimensions, and inference time scales only linearly with the dimension
- Our approach requires the objective function to be somewhat normalized; this however is no restriction if rough upper and lower bounds

Acknowledgement NVIDIA