Data-Efficient Training of Autoencoders for Mildly Non-Linear Problems

Muhammad Al-Digeil[1], Yuri Grinberg[1], Mohsen Kamandar Dezfouli[2], Daniele Melati[3],Jens H. Schmid[2], Pavel Cheben[2], Siegfried Janz[2]& Dan-Xia Xu[2]

[1]Digital Technologies Research Centre, National Research Council of Canada, Canada

[2]Advanced Electronics and Photonics Research Centre, National Research Council of Canada, Canada

[3]Université Paris-Saclay, CNRS, France

ABSTRACT

Principal Component Analysis (PCA) provides reliable dimensionality reduction (DR) when data possesses linear properties even for small datasets. However, faced with data that exhibits non-linear behaviour, PCA cannot perform optimally as compared to non-linear DR methods such as AutoEncoders. By contrast, AutoEncoders typically require much larger datasets for training than PCA. This data requirement is a critical impediment in applications where samples are scarce and expensive to come by. One such area is nanophotonics component design where generating a single data point might involve running optimization methods that use computationally demanding solvers.

We propose Guided AutoEncoders (G-AE), standard AutoEncoders initialized using a numerically stable procedure to replicate PCA behaviour before training.

Our results show this approach yields a marked reduction in the data size requirements for training the network along with gains in capturing non-linearity during dimensionality reduction and thus performing better than PCA alone.

INTRODUCTION

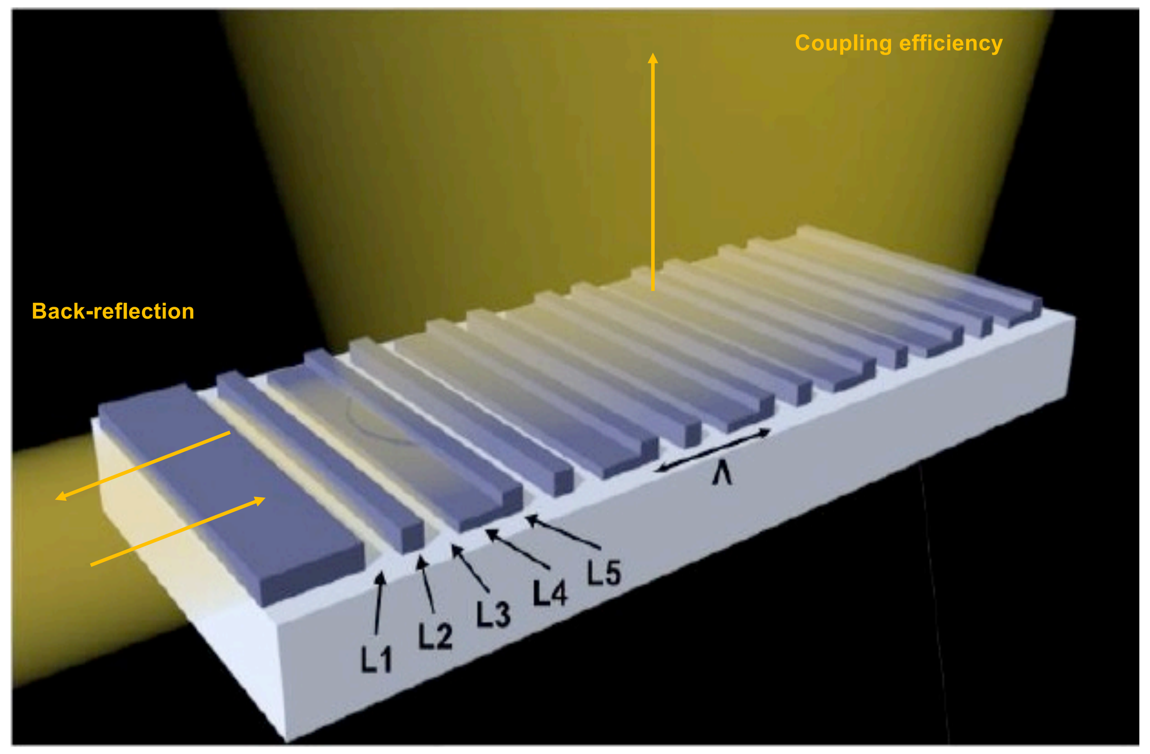

Figure 1. 5-segment Vertical Grating Coupler

- Optimal linear dimensionality reduction: PCA (Principal Component Analysis). But non-linear data ubiquitous.
- Example domain: nanophotonic component design.
- Designing nanophotonic components requires solving Maxwell's PDEs to obtain field distribution—computationally expensive.
- Can we benefit from PCA-like data efficiency while introduce some degree of non-linearity?

- Our dataset of optimized structures for the vertical grating coupler: 540 good designs (from candidate 30,000+)
- 5 segment values(L_1-L_5) characterize design
- Melati et al.(2019): 2 principal components capture most of good design subspace(Fig. 2a)
- Original design space <u>slightly curved</u>.(Fig. 2b)

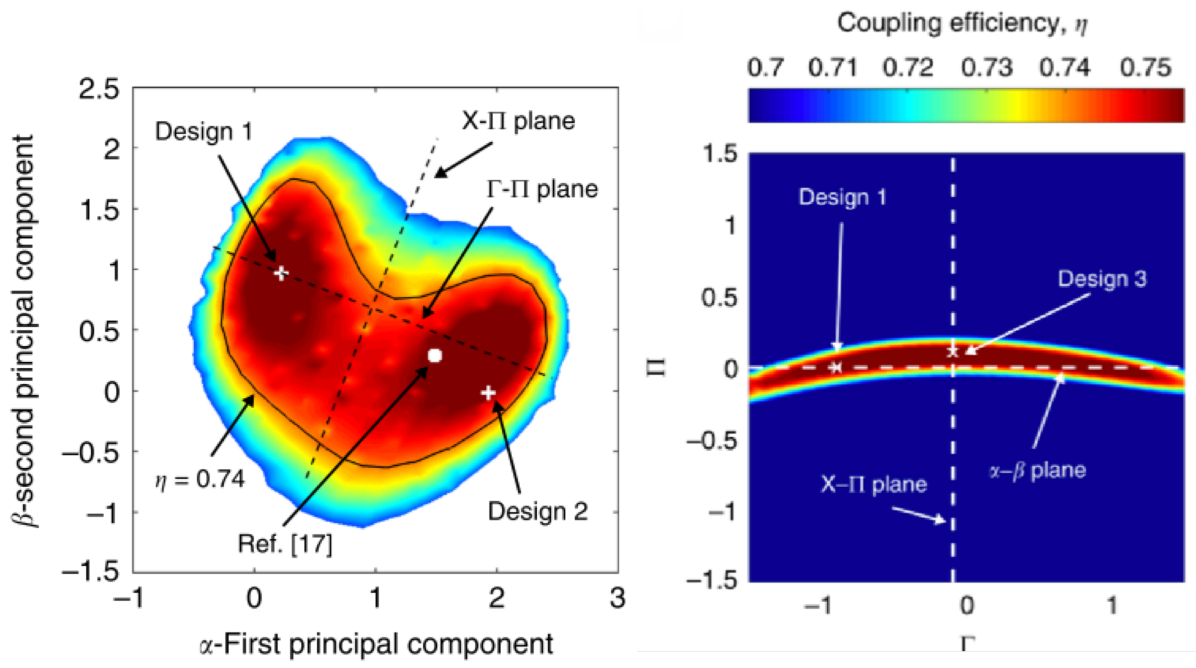

Figure 2 (a) good designs region (b) curvature of good design subspace

METHODOLOGY

- G-AE vase-shaped as in Figure 3.
- <u>Experiment replicates low-data regime—only 50 training samples</u>
- Data : 80% training, 10% validation, 10% test.
- Validation set used for early stopping.
- Test set compares performance across models.
- Experiments repeated 100 times with random samplings for accurate statistics.
- Initializing with PCA yields optimal linear AutoEncoder. Nonlinearity introduced gradually.
- Loss function: ℓ^2-norm(input vector, image).

- **Experiment 1** (LeakyReLU):
  - All nodes: LeakyReLU activation function.
  - set negative slope from 1 (linear) to 0.86 (mildly nonlinear) gradation: 0.01
  - Train with PCA initialization vs random.
  - Independent confirmation: measure performance of all model types (AE, G-AE, PCA) on larger **oracle** set.

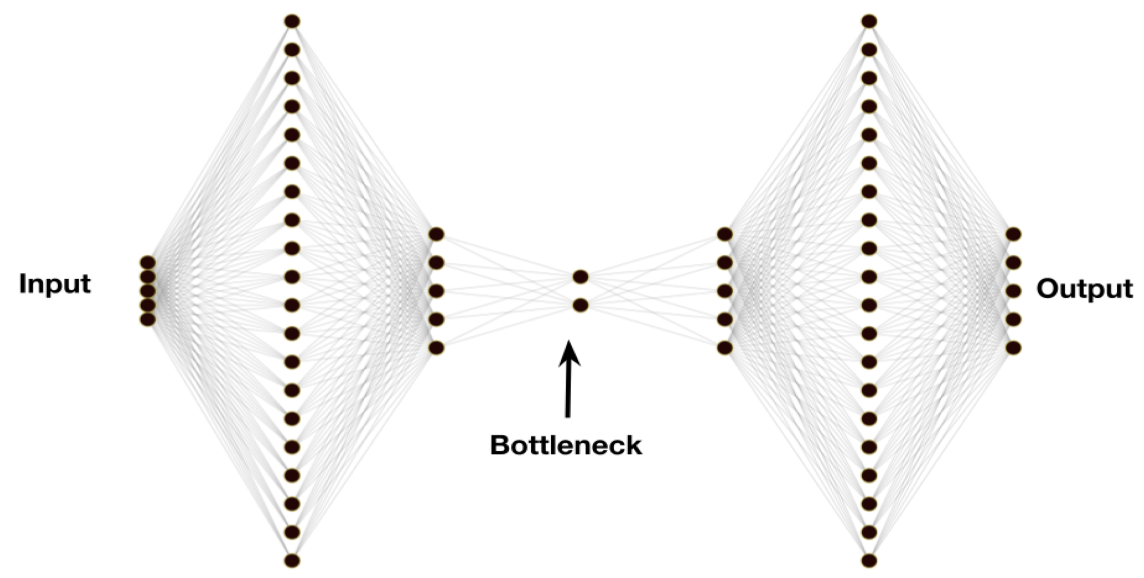

Figure 3. AutoEncoder Architecture used in Experiments

- **Experiment 2** (PReLU):
  - Same setup as Experiment 1 but Parametric ReLU (PRELU) activation function used on all nodes.
  - Each node allowed to vary its slope independently as a trainable parameter.

RESULTS

- Figure 4: performance of the model for Experiment 1 on the test set (a) and on the oracle set (b).
- G-AE outperforms randomly-initialized AEs and PCA for variety of slopes.
- Randomly initialized AEs perform comparably or worse than PCA.

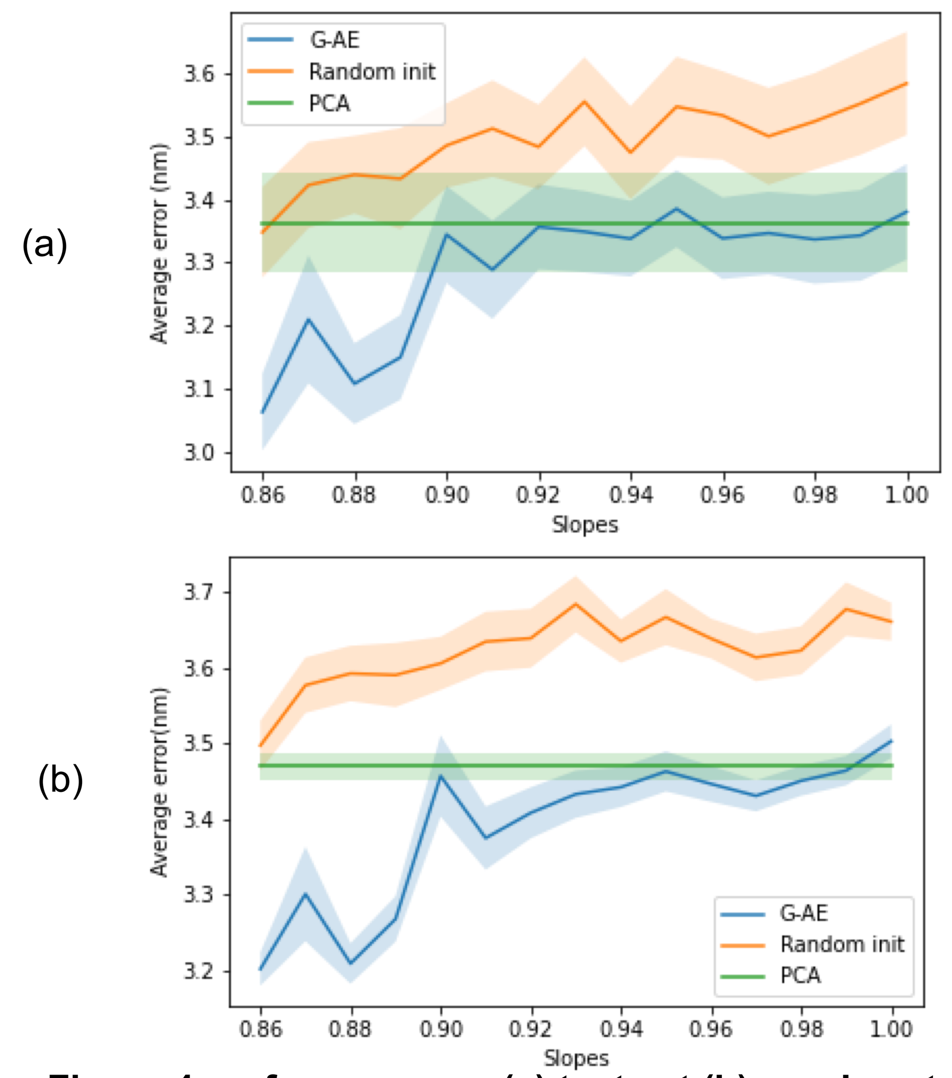

Figure 4 performance on (a) test set (b) oracle set

Model	Oracle Set Error	Test Set Error
PCA-initialized PReLU (G-AE)	3.27 ± 0.02	3.21 ± 0.06
PCA	3.48 ± 0.01	3.62 ± 0.07
Randomly Initialized PReLU AE	3.42 ± 0.1	3.35 ± 0.09

Table 1: Results of PReLU experiment

- Std. error (shaded regions in graphs), shows G-AE more stable than random initialization.
- Table 1: results of Experiment 2 using PReLU.
- Randomly initialized AE with PReLU performs comparably to PCA, G-AE outperforms both.

CONCLUSION

- AutoEncoders, with proper initialization, offer viable solution for dimensionality reduction in low data regime.
- On small but sufficient datasets, use of PCA to initialize LeakyReLU and PReLU AutoEncoders yields results that are superior to randomly initialized AutoEncoders, and even PCA alone.
- Results are encouraging in domains where only PCA has been used to reduce the dimensionality due to very limited datasets available.

REFERENCES

Baldi, P., & Hornik, K. (1989). Neural networks and principal component analysis: Learning from examples without local minima. *Neural networks, 2(1), 53–58.*

Cisse, M., Bojanowski, P., Grave, E., Dauphin, Y., & Usunier, N. (2017). Parseval networks: Improving robustness to adversarial examples. In *International Conference on Machine Learning* (pp. 854–863).

Melati, D., Grinberg, Y., Dezfouli, M., Janz, S., Cheben, P., Schmid, J., Sánchez-Postigo, A., & Xu, D.X. (2019). Mapping the global design space of nanophotonic components using machine learning pattern recognition. *Nature communications, 10(1), 1–9.*

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2(11), 559–572.*

M. Seuret, M. Alberti, M. Liwicki, & R. Ingold (2017). PCA-Initialized Deep Neural Networks Applied to Document Image Analysis. In *2017 14th International Conference on Document Analysis and Recognition (ICDAR)* (pp. 877-882).

National Research Council Canada | Conseil national de recherches Canada

Canada