Approaches for Deep Neural Network Architecture Adaptation

Krishnanunni C G[1], Tan Bui-Thanh[1,2]

[1] Department of Aerospace Engineering and Engineering Mechanics
[2] The Oden Institute for Computational Engineering and Sciences

The University of Texas at Austin
Aerospace Engineering and Engineering Mechanics
Cockrell School of Engineering

The University of Texas at Austin
Oden Institute for Computational Engineering and Sciences

Motivation and Goal

Challenges:

- **Neural network architecture design** via architecture search involves training a large number of candidate architectures to retrieve the best performing neural network.
- **Traditional neural architecture search (NAS)** algorithms are therefore computationally expensive.

Goal: Propose mathematically principled algorithms to:

- **Progressively adapt/grow** neural network architecture along the depth.
- Answer the following questions: i) When and where to add a new layer; ii) How to initialize the new layer.

Notations

β_{ij} – Similarity matrix
M – Number of training samples
$\mathcal{N}_{\theta}^{(l)}$ – l^{th} hidden layer with parameters $\boldsymbol{\theta}$
$\mathcal{J}$ – Training loss function
Ω_0 – Initial neural network
Ω_ϵ – Perturbed neural network
ϕ – Initialization of a newly added layer

Approach I: Layerwise training algorithm

- Layerwise learning procedure: Train a layer → Freeze the parameters → Add a new layer (identity map) and train again!

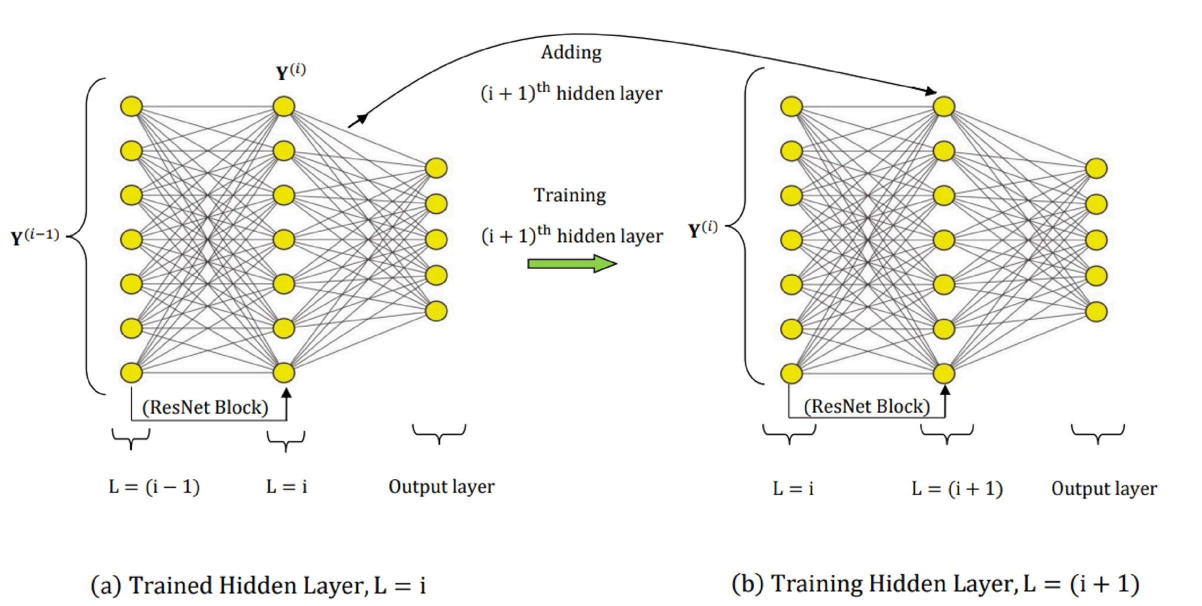

Figure 1. Schematics of layerwise training algorithm

- **Regularizers** to constrain the function learnt by each layer:
- **Manifold regularization** (Φ_m), sparsity regularization (Φ_s), and a physics-informed regularization term (Φ_p).
- *Layerwise training saturates after adding some layers!*
- **Sequential residual learning (SRL)**: A post-processing stage using a sequence of small networks to improve predictions!

Manifold regularization for promoting stability

- *In a deep network, initial layers learn meaningful representation of data. Later layers focus on the actual classification/regression task.*
- We attempt to mimic this property with **Manifold regularization** based on pairwise similarity and defined as:

$$\Phi_m = \frac{1}{M^2}\sum_{i,j}\beta_{ij}\left\|\mathcal{N}_{\theta}^{(l)}(\mathbf{x}_i) - \mathcal{N}_{\theta}^{(l)}(\mathbf{x}_j)\right\|_2^2.$$

- Manifold regularization ensures that similar input data points are mapped to similar outputs for each layer l.

Numerical results

Problem 1: Prototype regression and classification task

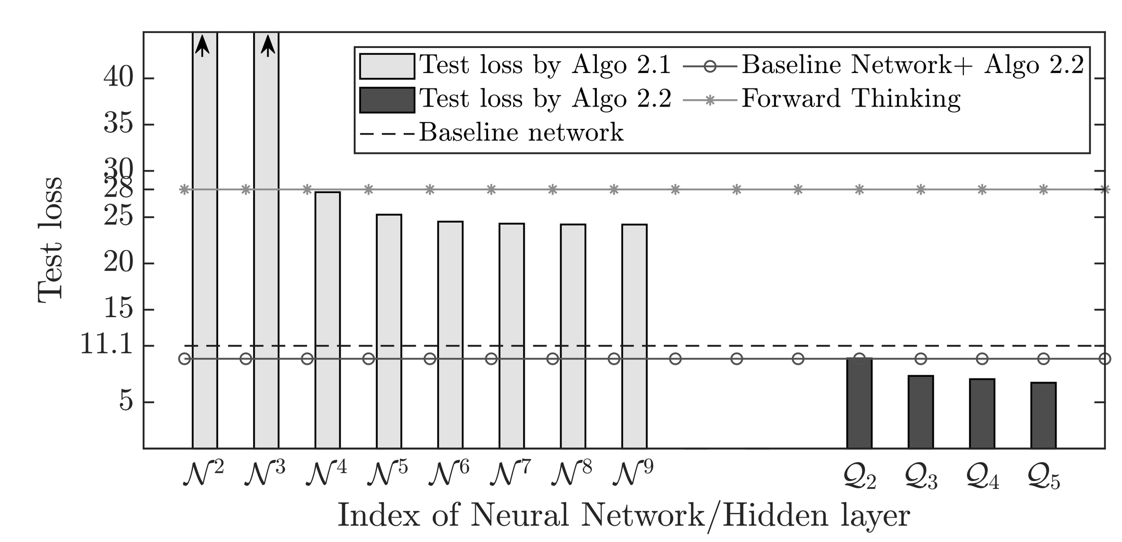

Figure 2. Performance of proposed approach over other strategies on a Boston housing price prediction problem. Our approach outperforms other methods!

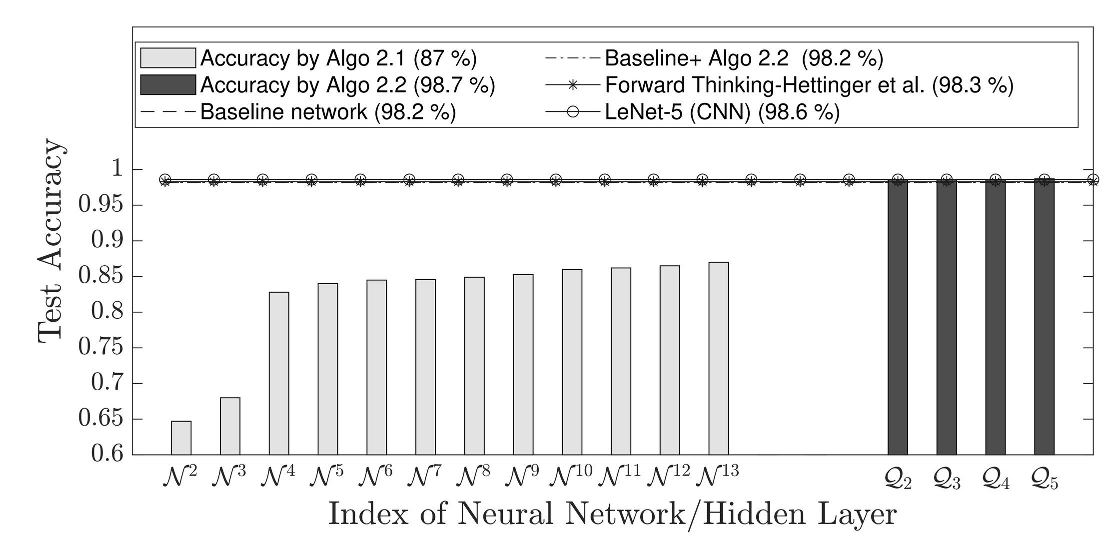

Figure 3. Performance of proposed approach over other strategies on MNIST classification problem. Our approach outperforms other methods!

Problem 2: Conductivity field inversion in a 2D Heat equation

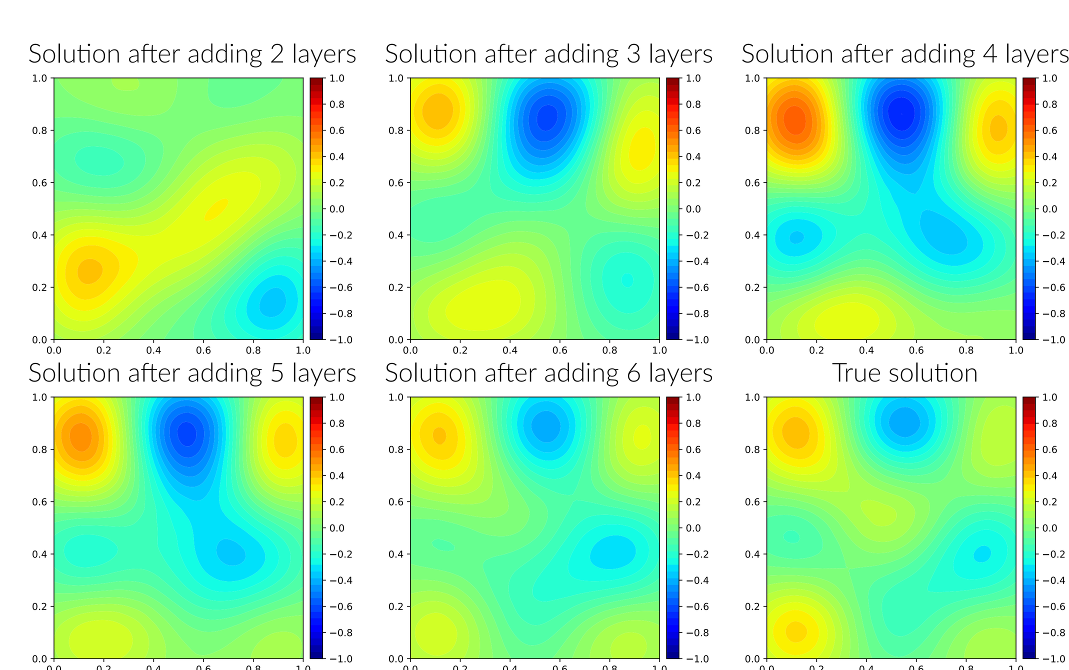

Figure 4. 2D Heat equation. Evolution of estimated conductivity field across the hidden layers for a particular test measurement sample. Adding layers progressively with manifold regularization recovers fine details in the parameter field. Achieved average relative error is superior to other adaptation strategies!

Approach II: A sensitivity based approach

- In this approach we consider perturbing a network Ω_0 to produce a new network Ω_ϵ as shown in Figure 5.

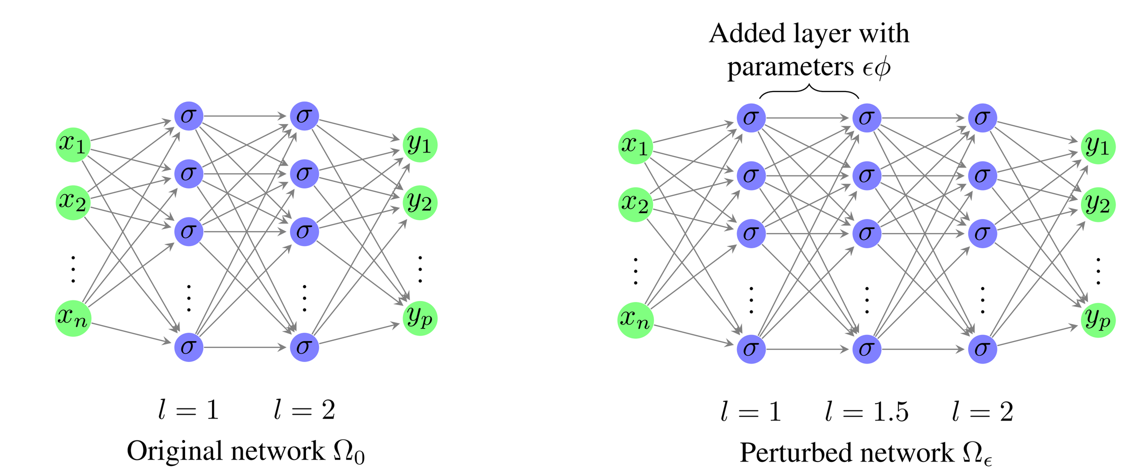

Figure 5. Schematic view of the sensitivity based approach

- In Figure 5, a new layer with residual connection and parameters $\epsilon\phi$ is inserted between the 1^{st} and 2^{nd} layer. When $\epsilon=0$, Ω_ϵ behaves exactly the same as Ω_0 (admissible perturbation).
- We rely on ideas from topology optimization & optimal transport theory to find the best initialization ϕ for a given small ϵ. Roughly speaking we look at the following optimization problem:

$$\arg\min_{\phi}\left(\mathcal{J}(\Omega_\epsilon) - \mathcal{J}(\Omega_0)\right), \quad s.t \quad \|\phi\|_2^2 = 1.$$

- We then determine the best location l^* along the depth to add a new layer that leads to maximum decrease in $\mathcal{J}$ (most sensitive).

Numerical results

Problem 1: Wind velocity reconstruction problem

- Objective is to train a network that reconstructs the wind velocity profile (in space) given sparse measurement data.

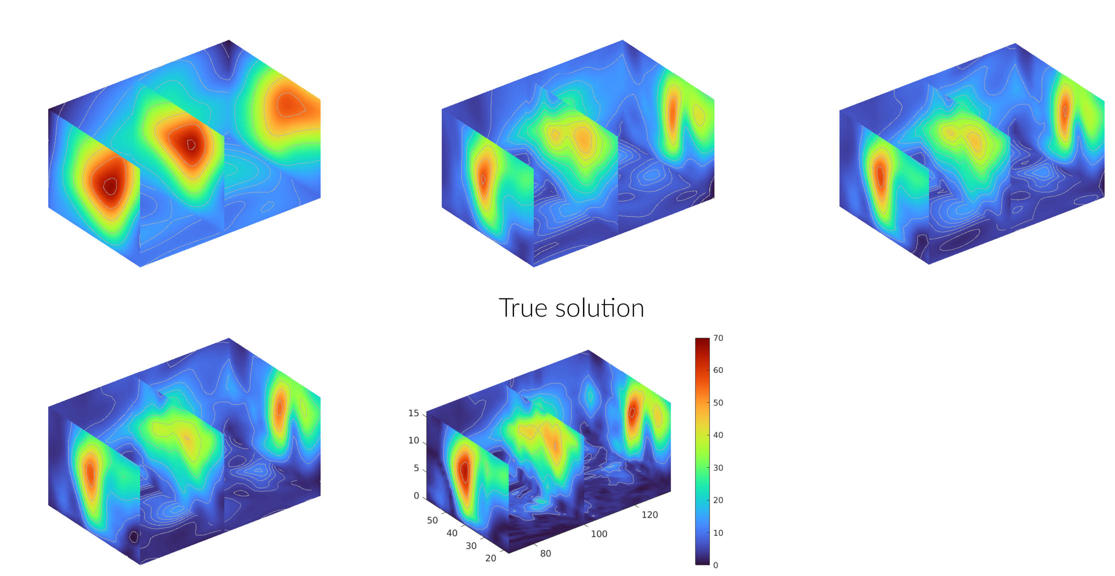

Figure 4. Wind velocity reconstruction. Evolution of solution (3D air current profile) upon adding new hidden layers.

- From Figure 4 we see that adding more parameters progressively refines the solution.
- Our algorithm also exhibits superior performance (in terms of mean squared error on a test data-set) compared to other adaptation strategies and neural architecture search algorithm!

Problem 2: Initial condition inversion in a 2D Navier-Stokes equation

$$\begin{aligned}
&\partial_t u(x,\,t) + v(x,\,t)\cdot\nabla u(x,\,t) = \nu\Delta u(x,\,t) + f(x),\\
&\nabla\cdot(x,\,t) = 0,\\
&u(x,\,0) = u_0(x).
\end{aligned}$$

- Train a neural network to learn the map from vorticity at time T, i.e $u(x,T)$ to the initial vorticity $u_0(x)$.

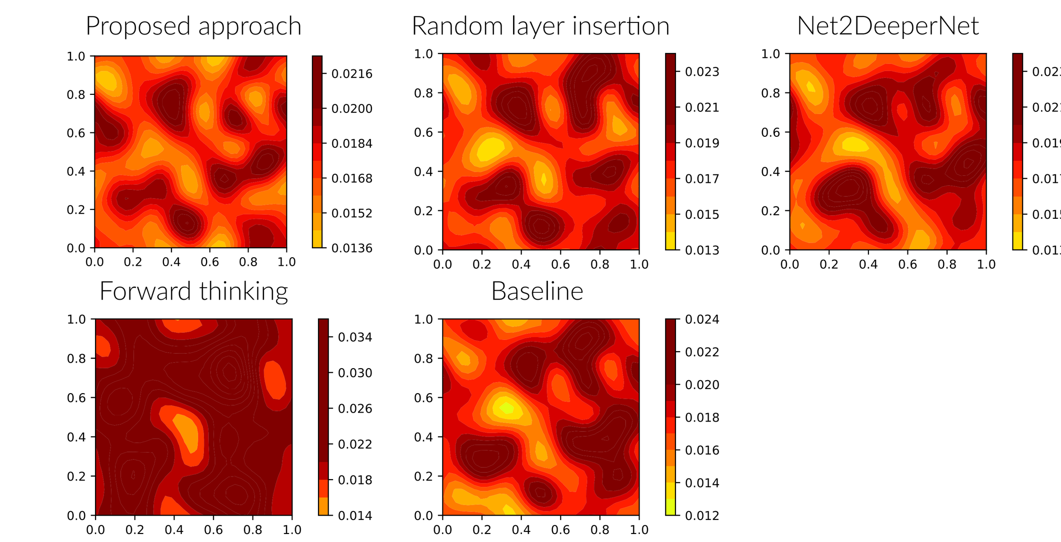

Figure 6. 2D Navier-Stoke equation. Average error in the predicted parameter field over the spatial domain for the entire test data-set by different methods.

- From the above figure, we see that our approach outperforms other adaptation strategies!

Conclusions

- Both the layerwise training algorithm and the sensitivity based approach outperforms existing neural architecture adaptation strategies.
- Both approaches provides answers to the following questions: a) Where to add new capacity (layer) during the training process?; b) How to initialize the new capacity?

Publication

Krishnanunni, C.G. and Bui-Thanh, T., 2024. An Adaptive and Stability-Promoting Layerwise Training Approach for Sparse Deep Neural Network Architecture. arXiv preprint arXiv:2211.06860.

Acknowledgement

The authors are grateful to the support from the National Science Foundation, the Department of Energy, and the Texas Advanced Computing Center. The authors would like to thank all the members of Pho-Ices group, Oden Institute for fruitful discussions.