Dynamic Network

Tianyuan Zhang

School of EECS, Peking University

May 2019

Table of contents

What is a dynamic network?

During every inference, only a subset of the supernet is activated. (Conditional Computation)

Why

- increase model capacity without a proportional increase in computational costs
- This is like the brain.
- Similar to sparse coding?

This is likely what learning systems look like in the future.

What is a dynamic network?

Currently, there mainly two types of dynamic networks.

The computation varies along several manually selected factors

- More interpretable
- Requires a appropriate encoding of these selected factors
- Examples: [6, 1]

The computation varies implicitly given the input

- Totally black boxes
- Examples: [5, 9, 2, 7, 3, 4, 8]

There are many works lie in the middle of these two types, because a perfect encoding of some manually selected factors relies on a black-box representation learning.

Supernet

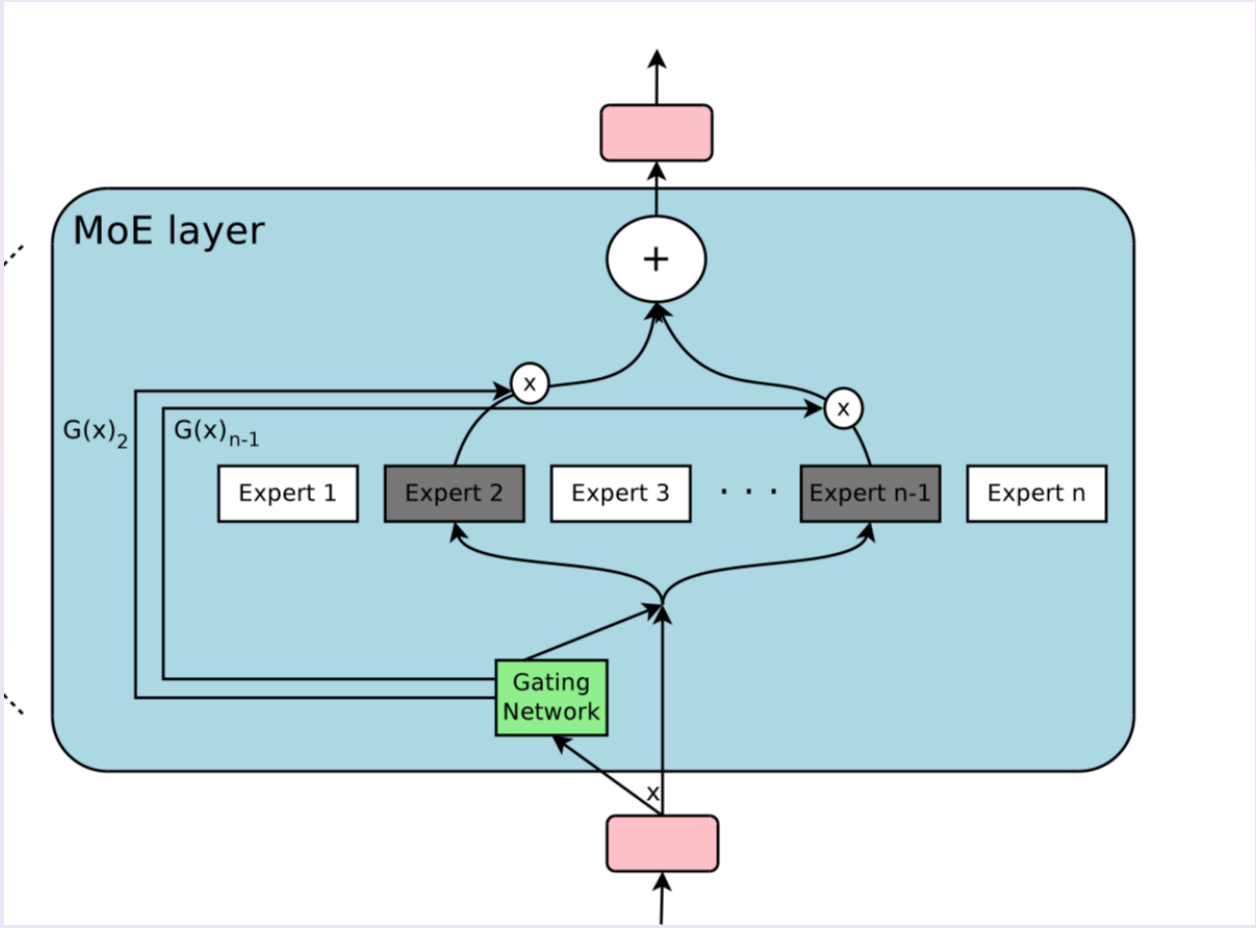

Figure: A simple MOE module [5].

A sparse MOE model

How to train a sparse gating module?

- **Forward**: Top-k (noise, temperature.)
- **Backwawrd**: Policy Gradient, Straight-through estimator, and some re-parameterization.

Deep neural decision forests

Joint and unified way of learning feature representations together with their classifiers greatly outperforms conventional feature descriptor classifier pipelines, whenever enough training data and computation capabilities are available [2].

Deep neural decision forests

forests.png forests.bb

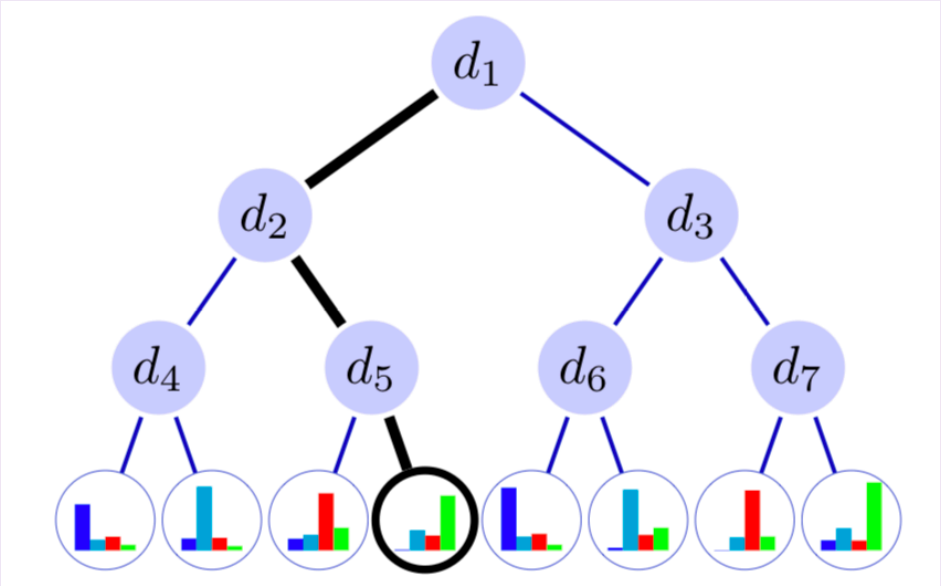

Figure: Every non-leaf node is a gating function.[2]

Soft conditional computation

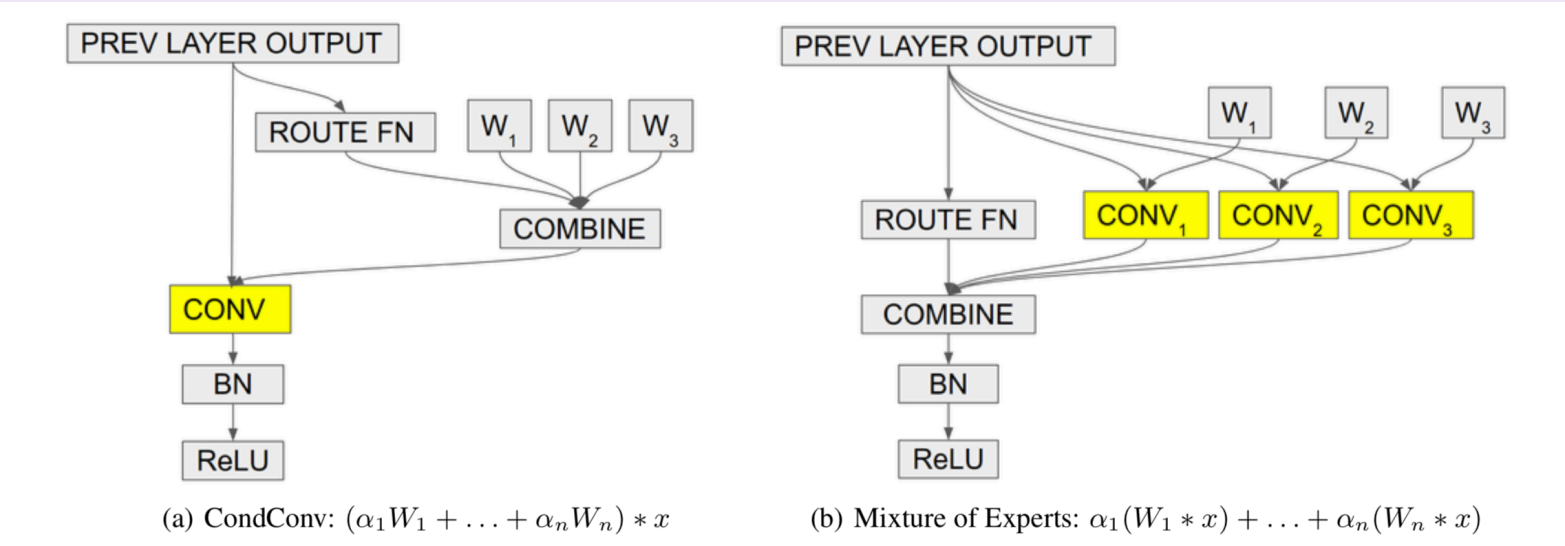

Figure: Soft conditional computation.

the same as weight prediction

Global Pool & Linear transformation -> Conv filters.

Skip Net

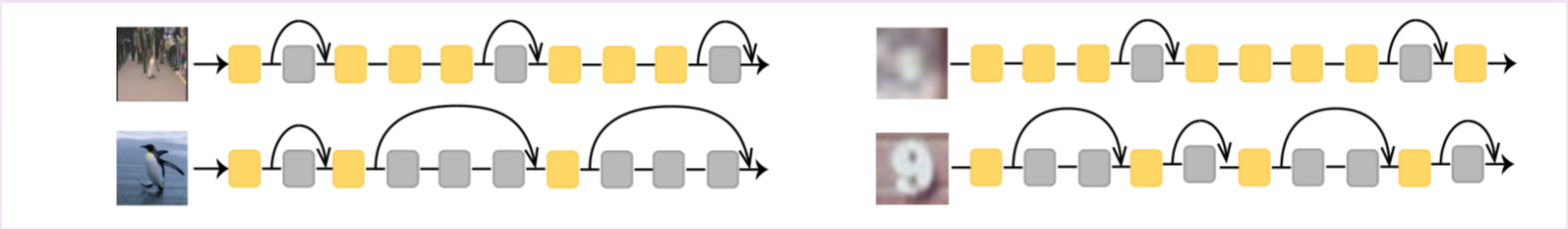

Figure: Skip-Net. [7]

Training is similar for that of sparse gating as mentioned before.

Style transfer via meta network

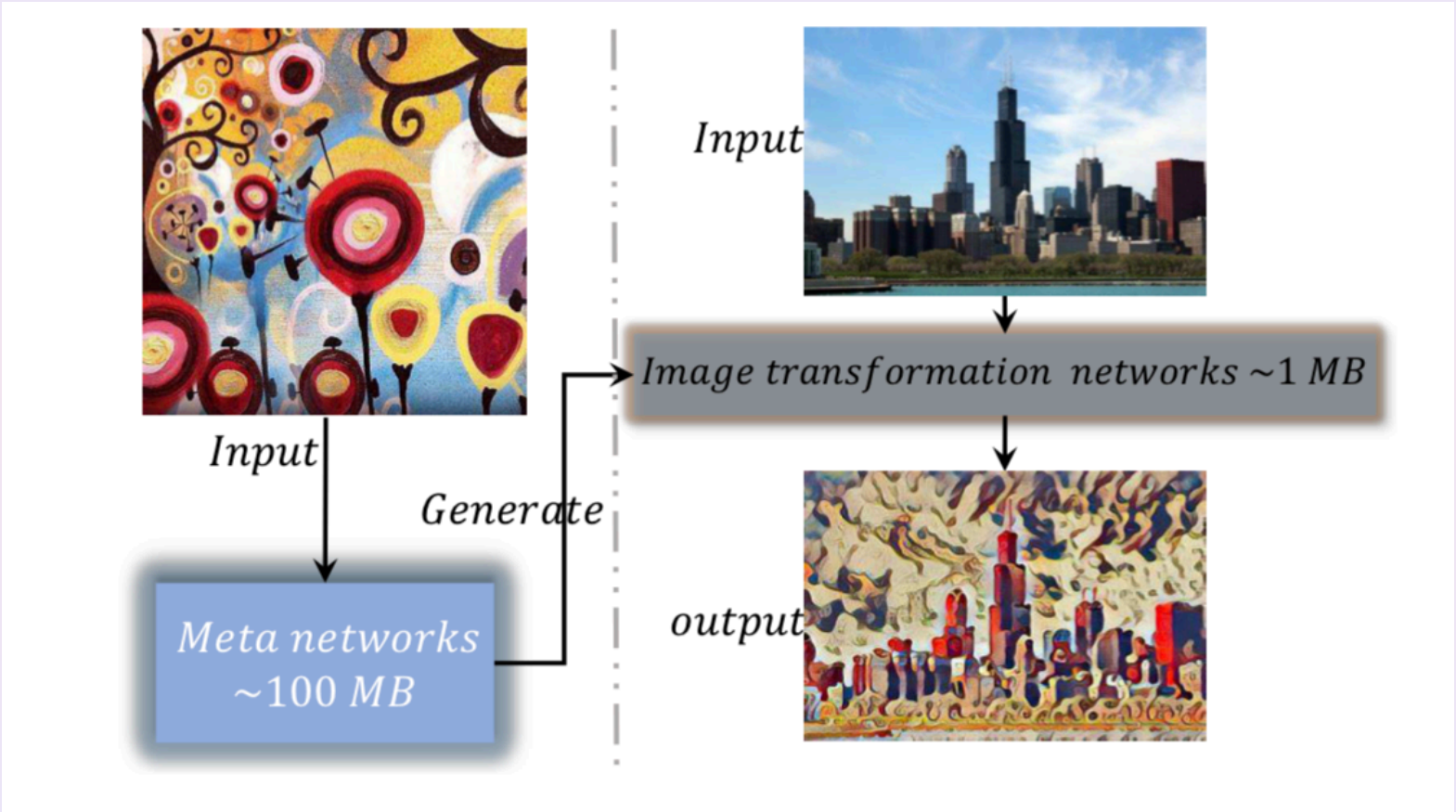

Figure: Pipleline.[6]

Meta-SR

Lots of details, we can skip this first. [1]

DURR

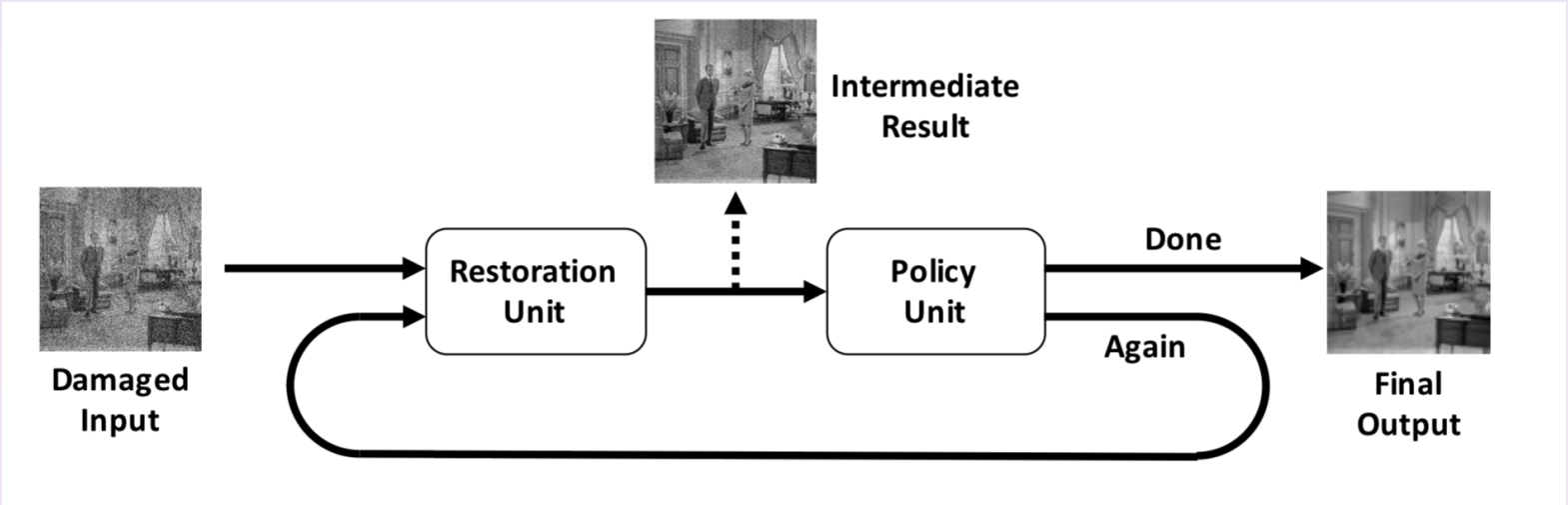

Figure: Pipleline.[9]

Like a RNN with a data-dependent folding times

References I

Xuecai Hu et al. "Meta-SR: A Magnification-Arbitrary Network for Super-Resolution". In: *arXiv preprint arXiv:1903.00875* (2019).

Peter Kontschieder et al. "Deep neural decision forests". In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 1467–1475.

Ji Lin et al. "Runtime neural pruning". In: *Advances in Neural Information Processing Systems*. 2017, pp. 2181–2191.

Lanlan Liu and Jia Deng. "Dynamic deep neural networks: Optimizing accuracy-efficiency trade-offs by selective execution". In: *Thirty-Second AAAI Conference on Artificial Intelligence*. 2018.

References II

Noam Shazeer et al. "Outrageously large neural networks: The sparsely-gated mixture-of-experts layer". In: *arXiv preprint arXiv:1701.06538* (2017).

Falong Shen, Shuicheng Yan, and Gang Zeng. "Neural style transfer via meta networks". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 8061–8069.

Xin Wang et al. "Skipnet: Learning dynamic routing in convolutional networks". In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018, pp. 409–424.

Brandon Yang et al. "Soft Conditional Computation". In: *arXiv preprint arXiv:1904.04971* (2019).

References III

Xiaoshuai Zhang et al. "Dynamically unfolding recurrent restorer: A moving endpoint control method for image restoration". In: *arXiv preprint arXiv:1805.07709* (2018).