On Enhancing Expressive Power via Compositions of Single Fixed-Size ReLU Network

Shijun Zhang

Duke University

(Joint work with Jianfeng Lu and Hongkai Zhao)

Motivation

- Deep neural networks have achieved great success in real-world applications.

Motivation

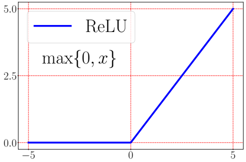
- Deep neural networks have achieved great success in real-world applications.
- Explosive growth of parameters and computation.

Motivation

- Deep neural networks have achieved great success in real-world applications.
- Explosive growth of parameters and computation.
- New network architecture via the idea of parameter sharing and function compositions.

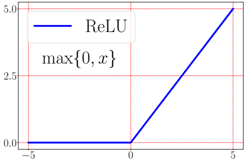
Notation

- ReLU networks: fully connected feed-forward neural networks activated by ReLU.



Notation

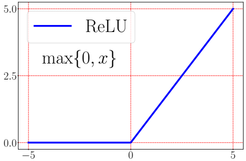
- ReLU networks: fully connected feed-forward neural networks activated by ReLU.



- $\mathcal{NN}\{N,\ L;\ \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}\}$: the set of all $\boldsymbol{h} : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$ realized by ReLU networks of width N and depth L.

Notation

- ReLU networks: fully connected feed-forward neural networks activated by ReLU.



- $\mathcal{NN}\{N,\ L;\ \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}\}$: the set of all $\boldsymbol{h} : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$ realized by ReLU networks of width N and depth L.

- Let $\boldsymbol{g}^{\circ r}$ denote the r-times composition of $\boldsymbol{g}$, e.g.,

$$\boldsymbol{g}^{\circ 3} = \boldsymbol{g} \circ \boldsymbol{g} \circ \boldsymbol{g}.$$

Compositions of single network

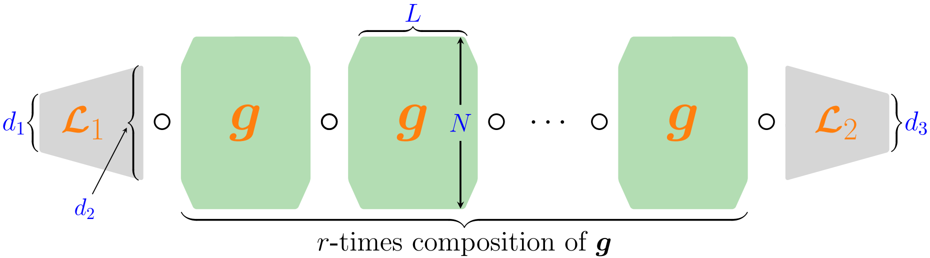
Design a new network $\boldsymbol{\mathcal{L}}_2 \circ \boldsymbol{g}^{\circ r} \circ \boldsymbol{\mathcal{L}}_1$ via repeated compositions:

- $\boldsymbol{g} \in \mathcal{NN}\{N,\ L;\ \mathbb{R}^{d_2} \to \mathbb{R}^{d_2}\}$.
- $\boldsymbol{\mathcal{L}}_1 : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$ and $\boldsymbol{\mathcal{L}}_2 : \mathbb{R}^{d_2} \to \mathbb{R}^{d_3}$ are affine.

Compositions of single network

Design a new network $\boldsymbol{\mathcal{L}}_2 \circ \boldsymbol{g}^{\circ r} \circ \boldsymbol{\mathcal{L}}_1$ via repeated compositions:

- $\boldsymbol{g} \in \mathcal{N}\mathcal{N}\{N,\ L;\ \mathbb{R}^{d_2} \to \mathbb{R}^{d_2}\}$.
- $\boldsymbol{\mathcal{L}}_1 : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$ and $\boldsymbol{\mathcal{L}}_2 : \mathbb{R}^{d_2} \to \mathbb{R}^{d_3}$ are affine.



Call this type of networks repeated-composition networks (RCNets).

Approximation of RCNets

Theorem

Given a 1-Lipschitz f, for any $r \in \mathbb{N}^+$ and $p \in [1, \infty)$, there exist

$$\boldsymbol{g} \in \mathcal{NN}\{69d + 48,\ 5;\ \mathbb{R}^{5d+5} \to \mathbb{R}^{5d+5}\}$$

and two affine maps $\boldsymbol{\mathcal{L}}_1 : \mathbb{R}^d \to \mathbb{R}^{5d+5}$ and $\mathcal{L}_2 : \mathbb{R}^{5d+5} \to \mathbb{R}$ s.t.

$$\left\| \mathcal{L}_2 \circ \boldsymbol{g}^{\circ r} \circ \boldsymbol{\mathcal{L}}_1 - f \right\|_{L^p([0,1]^d)} \le 6\sqrt{d}\, r^{-1/d}.$$

Approximation of RCNets

Theorem

Given a 1-Lipschitz f, for any $r \in \mathbb{N}^+$ and $p \in [1, \infty)$, there exist

$$\boldsymbol{g} \in \mathcal{NN}\{69d + 48,\ 5;\ \mathbb{R}^{5d+5} \to \mathbb{R}^{5d+5}\}$$

and two affine maps $\boldsymbol{\mathcal{L}}_1 : \mathbb{R}^d \to \mathbb{R}^{5d+5}$ and $\boldsymbol{\mathcal{L}}_2 : \mathbb{R}^{5d+5} \to \mathbb{R}$ s.t.

$$\left\| \boldsymbol{\mathcal{L}}_2 \circ \boldsymbol{g}^{\circ r} \circ \boldsymbol{\mathcal{L}}_1 - f \right\|_{L^p([0,1]^d)} \le 6\sqrt{d}\, r^{-1/d}.$$

- Arbitrarily small error with $O(d^2)$ parameters.
- L^p-norm $\to$ L^∞-norm, larger constants.
- 1-Lipschitz $\to C([0,1]^d)$, modulus of continuity.

Approximation of RCNets

$$\mathcal{H}(r) := \Big\{ \mathcal{L}_2 \circ \boldsymbol{g}^{\circ r} \circ \boldsymbol{\mathcal{L}}_1 : \boldsymbol{g} \in \mathcal{NN}\{69d+48,\ 5;\ \mathbb{R}^{5d+5} \to \mathbb{R}^{5d+5}\}, \\ \boldsymbol{\mathcal{L}}_1 : \mathbb{R}^d \to \mathbb{R}^{5d+5} \text{ and } \mathcal{L}_2 : \mathbb{R}^{5d+5} \to \mathbb{R} \text{ are affine} \Big\}$$

Approximation of RCNets

$$\mathcal{H}(r) := \Big\{ \mathcal{L}_2 \circ \boldsymbol{g}^{\circ r} \circ \boldsymbol{\mathcal{L}}_1 : \boldsymbol{g} \in \mathcal{NN}\{69d+48,\ 5;\ \mathbb{R}^{5d+5} \to \mathbb{R}^{5d+5}\}, \\ \boldsymbol{\mathcal{L}}_1 : \mathbb{R}^d \to \mathbb{R}^{5d+5} \text{ and } \mathcal{L}_2 : \mathbb{R}^{5d+5} \to \mathbb{R} \text{ are affine} \Big\}$$

- $\mathcal{H} = \cup_{r=1}^{\infty} \mathcal{H}(r)$ is dense in $C([0,1]^d)$ in terms of the L^p-norm for any $p \in [1, \infty)$.

Approximation of RCNets

$$\mathcal{H}(r) := \Big\{ \mathcal{L}_2 \circ \boldsymbol{g}^{\circ r} \circ \boldsymbol{\mathcal{L}}_1 : \boldsymbol{g} \in \mathcal{NN}\{69d+48,\ 5;\ \mathbb{R}^{5d+5} \to \mathbb{R}^{5d+5}\},$$
$$\boldsymbol{\mathcal{L}}_1 : \mathbb{R}^d \to \mathbb{R}^{5d+5} \text{ and } \mathcal{L}_2 : \mathbb{R}^{5d+5} \to \mathbb{R} \text{ are affine} \Big\}$$

- $\mathcal{H} = \cup_{r=1}^{\infty} \mathcal{H}(r)$ is dense in $C([0,1]^d)$ in terms of the L^p-norm for any $p \in [1, \infty)$.
- $\mathcal{H} = \cup_{r=1}^{\infty} \mathcal{H}(r)$ is parameterized with only $O(d^2)$ parameters:

$$\boldsymbol{g} = \boldsymbol{g}_{\boldsymbol{\theta}_0},\ \boldsymbol{\mathcal{L}}_1 = \boldsymbol{\mathcal{L}}_{\boldsymbol{\theta}_1},\ \mathcal{L}_2 = \mathcal{L}_{\boldsymbol{\theta}_2} \quad \Longrightarrow \quad h_{\boldsymbol{\theta}} = \mathcal{L}_{\boldsymbol{\theta}_2} \circ \boldsymbol{g}_{\boldsymbol{\theta}_0}^{\circ r} \circ \boldsymbol{\mathcal{L}}_{\boldsymbol{\theta}_1},$$

where $\boldsymbol{\theta} = (\underbrace{\boldsymbol{\theta}_0}_{O(d^2)}, \underbrace{\boldsymbol{\theta}_1}_{O(d^2)}, \underbrace{\boldsymbol{\theta}_2}_{O(d)}, r) \in \mathbb{R}^{O(d^2)}$.

Thank you!

https://shijunzhang.top