Commun. Comput. Phys.
doi: 10.4208/cicp.OA-2020-0149

Vol. **28**, No. 5, pp. 1768-1811
November 2020

Deep Network Approximation Characterized by Number of Neurons

Zuowei Shen[1], Haizhao Yang[2,*] and Shijun Zhang[1]

[1] *Department of Mathematics, National University of Singapore, Singapore 119076.*
[2] *Department of Mathematics, Purdue University, Indiana 47907, USA.*

Received 22 July 2020; Accepted (in revised version) 11 October 2020

Abstract. This paper quantitatively characterizes the approximation power of deep feed-forward neural networks (FNNs) in terms of the number of neurons. It is shown by construction that ReLU FNNs with width $\mathcal{O}\big(\max\{d\lfloor N^{1/d}\rfloor, N+1\}\big)$ and depth $\mathcal{O}(L)$ can approximate an arbitrary Hölder continuous function of order $\alpha\in(0,1]$ on $[0,1]^d$ with a nearly tight approximation rate $\mathcal{O}\big(\sqrt{d}N^{-2\alpha/d}L^{-2\alpha/d}\big)$ measured in L^p-norm for any $N,L\in\mathbb{N}^+$ and $p\in[1,\infty]$. More generally for an arbitrary continuous function f on $[0,1]^d$ with a modulus of continuity $\omega_f(\cdot)$, the constructive approximation rate is $\mathcal{O}\big(\sqrt{d}\omega_f(N^{-2/d}L^{-2/d})\big)$. We also extend our analysis to f on irregular domains or those localized in an ε-neighborhood of a $d_{\mathcal{M}}$-dimensional smooth manifold $\mathcal{M}\subseteq[0,1]^d$ with $d_{\mathcal{M}}\ll d$. Especially, in the case of an essentially low-dimensional domain, we show an approximation rate $\mathcal{O}\big(\omega_f(\frac{\varepsilon}{1-\delta}\sqrt{\frac{d}{d_\delta}}+\varepsilon)+\sqrt{d}\omega_f(\frac{\sqrt{d}}{(1-\delta)\sqrt{d_\delta}}N^{-2/d_\delta}L^{-2/d_\delta})\big)$ for ReLU FNNs to approximate f in the ε-neighborhood, where $d_\delta=\mathcal{O}\big(d_{\mathcal{M}}\frac{\ln(d/\delta)}{\delta^2}\big)$ for any $\delta\in(0,1)$ as a relative error for a projection to approximate an isometry when projecting $\mathcal{M}$ to a d_δ-dimensional domain.

AMS subject classifications: 68T01, 65D99, 68U99

Key words: Deep ReLU neural networks, Hölder continuity, modulus of continuity, approximation theory, low-dimensional manifold, parallel computing.

1 Introduction

The approximation theory of neural networks has been an active research topic in the past few decades. Previously, as a special kind of ridge function approximation, shallow neural networks with one hidden layer and various activation functions (e.g., wavelets

*Corresponding author. *Email addresses:* `matzuows@nus.edu.sg` (Z. Shen), `haizhao@purdue.edu` (H. Yang), `zhangshijun@u.nus.edu` (S. Zhang)

 ©2020 Global-Science Press

pursuits [10,46], adaptive splines [19,55], radial basis functions [8,18,25,53,65], sigmoid functions [7,13–15,29,37,38,41,45]) were widely discussed and admit good approximation properties, e.g., the universal approximation property [16,29,30], lessening the curse of dimensionality [4,21,22], and providing attractive approximation rate in nonlinear approximation [10,18,19,25,46,55,65].

The introduction of deep networks with more than one hidden layers has made significant impacts in many fields in computer science and engineering including computer vision [35] and natural language processing [1]. New scientific computing tools based on deep networks have also emerged and facilitated large-scale and high-dimensional problems that were impractical previously [20,24]. The design of deep ReLU FNNs is the key of such a revolution. These breakthroughs have stimulated broad research topics from different points of views to study the power of deep ReLU FNNs, e.g. in terms of combinatorics [51], topology [6], Vapnik-Chervonenkis (VC) dimension [5,27,58], fat-shattering dimension [2,34], information theory [54], classical approximation theory [4,16,30,62,67], optimization [32,33,52] etc.

Particularly in approximation theory, **non-quantitative and asymptotic** approximation rates of ReLU FNNs have been proposed for various types of functions. For example, smooth functions [23,39,43,66], piecewise smooth functions [54], band-limited functions [50], continuous functions [67], solutions to partial differential equations [31]. However, to the best of our knowledge, existing theories [17,23,39,43,48,50,54,63,66,67] can only provide implicit formulas in the sense that the approximation error contains an unknown prefactor, or work only for sufficiently large N and L larger than some unknown numbers. For example, [67] estimated an approximation rate $c(d)L^{-2\alpha/d}$ via a narrow and deep ReLU FNN, where $c(d)$ is an unknown number depending on d, and L is required to be larger than a sufficiently large unknown number $\mathscr{L}$. For another example, given an approximation error ε, [54] proved the existence of a ReLU FNN with a constant but still unknown number of layers approximating a C^β function within the target error. These works can be divided into two cases: 1) FNNs with varying width and only one hidden layer [18,25,40,65] (visualized by the region in ▬ in Fig. 1); 2) FNNs with a fixed width of $\mathcal{O}(d)$ and a varying depth larger than an unknown number $\mathscr{L}$ [44,67] (represented by the region in ▬ in Fig. 1).

As far as we know, the first **quantitative and non-asymptotic** approximation rate of deep ReLU FNNs was obtained in [62]. Specifically, [62] identified an explicit formulas of the approximation rate

$$\begin{cases} 2\lambda N^{-2\alpha}, & \text{when } L\geq 2 \text{ and } d=1, \\ 2(2\sqrt{d})^\alpha \lambda N^{-2\alpha/d}, & \text{when } L\geq 3 \text{ and } d\geq 2, \end{cases} \tag{1.1}$$

for ReLU FNNs with an arbitrary width $N\in\mathbb{N}^+$ and a fixed depth $L\in\mathbb{N}^+$ to approximate a Hölder continuous function f of order α with a Hölder constant λ (visualized in the region shown by ▬ in Fig. 1). The approximation rate $\mathcal{O}(N^{-2\alpha/d})$ is tight in terms of N and increasing L cannot improve the approximation rate in N. The success of deep