



# The Degree of Best Downsampled Convolutional Neural Network Approximation in Terms of Orlicz Modulus of Smoothness

Amna Manaf Al-Janabi<sup>\*</sup>, Hawraa Abbas Almurieb<sup>\*</sup>

Department of Mathematics, College of Education for Pure Sciences, University of Babylon, Hillah 51001, Iraq

Corresponding Author Email: [amnamn22@gmail.com](mailto:amnamn22@gmail.com)

Copyright: ©2024 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/mmep.110825>

## ABSTRACT

**Received:** 6 December 2023  
**Revised:** 28 April 2024  
**Accepted:** 15 May 2024  
**Available online:** 28 August 2024

### Keywords:

approximation, Orlicz, modulus of smoothness, convolutional neural networks, downsampling

It is essential to study approximation theoretical foundations of deep convolutional neural networks, because of its interesting developments in vital domains. In spite of its dependence upon approximation substantially. The aim is to study approximation abilities of deep convolution neural network produced by downsampling operators in Orlicz spaces to reduce high dimensions that causes overfitting implementation. The degree of best approximation of Orlicz functions are estimated in terms of high order modulus of smoothness. Moreover, direct and inverse theorems are proved here to get finally both bounds of degree of approximation, with upper and lower bounds to restrict the degree of approximation with modulus of smoothness. The concluded degree of approximation by our CNN vanishes theoretically faster than the classical ones due to its dependence on modulus of smoothness.

## 1. INTRODUCTION

Orlicz spaces are one of the essential wide spaces that could help in practical usage. It is interesting to study approximation of convolutional neural networks (CNNs) in Orlicz spaces, to get degree of approximation in terms of modulus of smoothness. We define an extended class of Orlicz function, called quasi-Orlicz space.

We introduce the introduction within two sections. The first section is about the space of interest, that is Orlicz space. It includes preliminaries and development stations about it. Then we define Orlicz space and modulus of smoothness with the auxiliary results about them. Second, we introduce the literature review about CNNs and sampling method. So we construct our CNN to get theoretical approximation results in terms of Orlicz modulus of smoothness. Then an application about approximation of CNNs is introduced to get best approximation for function from generalized Orlicz space.

### 1.1 Orlicz spaces

The first who studied this type of spaces was the mathematician Orlicz [1]. As a generalization of the Lebesgue integrable spaces  $L_p$ . With completeness property, Orlicz space form an extended Banach space:

$$L_\Phi(\mu) = \{f : f \text{ is } \mu - \text{meas and } \|f\|_\Phi < \infty\} \quad (1)$$

where, the norm  $\|\cdot\|_\Phi$  had been defined in many ways to cover Orlicz space, beginning with Orlicz him self with his norm.

$$\|f\|_\Phi^0 = \sup \left\{ \int_T |f(t)y(t)| d\mu : y \in L_\Psi, I_\Psi(y) \leq 1 \right\} \quad (2)$$

where, Orlicz norm was defined depending on Young function  $\Phi$  and the complementary function  $\Psi$ , the generator of the modular unit ball. The function  $\Psi$  is defined by

$$\Psi(u) = \sup\{ |u| v - \Phi(v) : v \geq 0 \} \quad (3)$$

for all  $u \in \mathbb{R}$ .

In simultaneous times in the fifteenth, Nakano [2], Morse and Transue [3] and Luxembourg [4] investigated the Luxembourg norm, which has been defined, using the concept of functional Minikowski over a convex modular unit ball,  $\{x : I(x) \leq 1\}$ , with

$$\|f\|_\Phi = \inf_{\lambda > 0} \left\{ I_\Phi \left( \frac{f}{\lambda} \right) \leq 1 \right\} \quad (4)$$

Another main norm was defined by Amemiya at the same time, named after his name  $p$ -Amemiya norm [2].

$$\|f\|_\Phi^A = \inf_{\lambda > 0} \frac{1}{\lambda} (1 + I_\Phi(\lambda f)) \quad (5)$$

In separated papers, Krasnoselskii and Rutickii [5], Nakano [2] and Luxembourg and Zaneen [6] proved under additional conditions on  $\Phi$ , that Amemiya norm  $\|\cdot\|_\Phi^A$  is exactly the Orlicz norm  $\|\cdot\|_\Phi^0$ . Moreover, Cui et al. [7] gave the basic results about the so called  $p$ -Amemiya norms equipped with other spaces. Also, Wisła [8] showed an overview of the developments of the spaces defined later. Moreover, Wisła [9] introduced a more generalized  $p$ -Amemiya type norms by restricting his conditions about the outer function. Later, he introduced the concept of outer function  $s$  and get  $s$ -norms on Orlicz spaces.

$$\|f\|_{\Phi,s} = \inf_{\lambda > 0} \frac{1}{\lambda} s(I_\Phi(\lambda f)) \quad (6)$$

Until today, many papers improved the norms of Orlicz [10-12].

All the above definitions of Orlicz norms depend on the values of  $p$  to be within  $[1, \infty]$ . For the case  $0 < p < 1$ , AL-Janabi and Almurieb [13] made another contribution to Orlicz spaces via outer function  $S_p: [0, \infty) \rightarrow [0, \infty)$  in the following definition.

**Definition 1.** [13] For  $0 < p < 1$  the family  $S_p$  of functions is defined by

$$S_p(f) = (1 + f^2)^{\frac{1}{p}} \quad (7)$$

where,  $f \in L_{\phi, S_p}(I^d)$ , the quasi-normed space.

$$L_{\phi, S_p}(I^d) = \left\{ f \in L_p(I^d) \mid \|f\|_{\phi, S_p} < \infty \right\} \quad (8)$$

where,

$$\|f\|_{\phi, S_p} = \inf_{\lambda > 0} \frac{1}{\lambda} (I_{\phi}^2(\lambda f) + 1)^{\frac{1}{p}} \quad (9)$$

As mentioned by AL-Janabi and Almurieb [13],  $S_p, p < 1$ , are convex, strictly increasing and all  $S_p$  match at zero. As concluded there,  $L_{\phi, S_p}, p \leq 1$  quasi-normed space is a generalization of  $p$ -Amemiy's normed space on  $p \geq 1$ .

Moreover, the degree of best approximation for functions from  $L_{\phi, S_p}$  in terms of CNNs were studied [13], so that

$$\|f - N\|_{\phi, S_p} < \epsilon \quad (10)$$

More precisely, the following theorem is given by AL-Janabi and Almurieb [13].

**Theorem A.** Let  $f \in L_{\phi, S_p}(I^d), 0 < p < 1$ , then by (10) there exists  $f_j^{w,b}$  s.t  $\|f - f_j^{w,b}\|_{\phi, S_p} \leq C_{(p,k)} \|f\|_{\phi, S_p}$ .

Our target here is to improve the degree of approximation by studying modulus of smoothness in our Orlicz quasi-normed space in the following section.

## 1.2 Orlicz moduli of smoothness

Moduli of smoothness had been studied widely in Orlicz space. The importance of this comes from the need to improve the degree of function approximation via direct approaches. Direct approaches approximation implies a faster convergence to zero than general previous estimates. On the other hand, converse theorems give a characterization of smoothness of functions depending on its degree of approximation in direct approach. First results were given by Jackson [14], and Bernstein [15], for direct and inverse theorems, respectively, for the space of continuous functions in terms of modulus of continuous [16]. Later, second and third orders moduli of smoothness were involved in many papers concerning many wide generalizations of function's spaces [17-22].

However, the birth of modulus of smoothness is much deeper in history [23-26], we concern here with moduli of smoothness as for as relates to Orlicz spaces. It's very hard to investigate the early works of this topic, since they were written in Russian. However, the oldest paper we have obtained is attributed to Tsyganok [27] in 1966. He proved direct theorem as a modulus of continuity for Orlicz spaces.

Later, in the eighteens of the last century, Ramazanov generalized direct theorem in Orlicz spaces for modulus of smoothness with higher orders [28]. In 1985, Musielak [24], generalized  $\tau$ -modulus of smoothness, resulted by Popov [25], for Orlicz spaces. In 1991, Garidi generalized those results of Ramazanov [28], Garidi [29]. Many other results were studied by Israfilov and Akgün [30], Akgün [31], Chaichenko et al. [32], for weighted Orlicz spaces. More extensions were resulted to Orlicz spaces by Shidlich and Chaichenko [33, 34], which were combined by Chaichenko Shidlich et al. [32].

## 1.3 Modulus of Smoothness in $L_{\phi, S_p}(I^d)$

Let's now present the main definitions of our modulus of smoothness.

First, from Ditzian and Totik [23], define the following  $N$ -th symmetric difference

$$\Delta_h^N(f(x)) = \sum_{i=1}^N (-1)^i \binom{N}{i} f(x - ih) \quad (11)$$

The Orlicz Modulus of smoothness in terms of (11) is given by

$$\omega_N(f, \delta)_{\phi, S_p} = \sup_{|h| \leq \delta} \|\Delta_h^N(f)\|_{\phi, S_p} \quad (12)$$

We study the main properties of the  $N$ -th symmetric difference (11) and the Orlicz modulus of smoothness (12) in the following section.

### 1.3.1 Modulus of smoothness properties (Auxiliary results)

In the following lemma, we prove that the  $N$ -th symmetric difference (11) satisfies the following properties under our Orlicz norm (9):

#### Lemma 1

Let  $f$  be any function from (8), then

- $\|\Delta_h^N(f)\|_{\phi, S_p} \leq k(N) \|f\|_{\phi, S_p}$ , where  $k(N) = \sum_{i=0}^{\infty} \left| \binom{N}{i} \right| \leq 2^N, N = \inf\{k \in \mathbb{N} : k > N\}$ .
- $\left( \Delta_h^N(\Delta_h^M(f)) \right) (x) = \Delta_h^{N+M}(f(x)) (a. e.)$ .
- $\|\Delta_h^{N+M}(f)\|_{\phi, S_p} \leq 2^{\{M\}} \|\Delta_h^N(f)\|_{\phi, S_p}$ .
- $\lim_{h \rightarrow 0} \|\Delta_h^N(f)\|_{\phi, S_p} = 0$ .

#### Proof

- Let  $\lambda > 0$ , then by (7), (9) and (11), we have

$$\begin{aligned} \|\Delta_h^N(f)\|_{\phi, S_p} &= \inf_{\lambda > 0} \frac{1}{\lambda} S_p(\Delta_h^N(f)) \\ &= \inf_{\lambda > 0} \frac{1}{\lambda} S_p \left( \sum_{i=1}^N (-1)^i \binom{N}{i} f(x - ih) \right) \\ &\leq C \inf_{\lambda > 0} \frac{1}{\lambda} S_p \left( \sum_{i=1}^{\infty} (-1)^i \binom{N}{i} f(x - ih) \right) \\ &\leq C \inf_{\lambda > 0} \frac{1}{\lambda} \sum_{i=0}^{\infty} \binom{N}{i} S_p((-1)^i f(x - ih)) \\ &\leq C \sum_{i=0}^{\infty} \binom{N}{i} \inf_{\lambda > 0} \frac{1}{\lambda} S_p(f(x - ih)) \\ &\leq Ck(N) \|f\|_{\phi, S_p} \end{aligned}$$

2. By (7), (9) and (11), we have

$$\begin{aligned} (\Delta_h^N(\Delta_h^M(\mathbf{f}))) (x) &= \sum_{i=1}^N (-1)^i \binom{N}{i} (\Delta_h^M(\mathbf{f}(x - ih))) \\ &= \sum_{i=1}^N (-1)^i \binom{N}{i} \left( \sum_{i=1}^M (-1)^i \binom{M}{i} (\mathbf{f}(x - ih)) \right) \\ &= \sum_{i=1}^N \sum_{i=1}^M (-1)^i \binom{N}{i} \binom{M}{i} \mathbf{f}(x - hi) \\ &= \sum_{i=1}^{N+M} (-1)^i \binom{N+M}{i} \mathbf{f}(x - hi) \\ &= \Delta_h^{N+M}(\mathbf{f}(x)). \end{aligned}$$

3. By the above **Lemma 1**, we get

$$\begin{aligned} \|\Delta_h^{N+M}(\mathbf{f})\|_{\phi, S_p} &= \|\Delta_h^M(\Delta_h^N(\mathbf{f}))\|_{\phi, S_p} \\ &\leq k(M) \|\Delta_h^N(\mathbf{f})\|_{\phi, S_p} \\ &\leq 2^{(M)} \|\Delta_h^N(\mathbf{f})\|_{\phi, S_p}. \end{aligned}$$

4. Let  $\epsilon > 0$ , choose  $\delta = \delta(\epsilon, N)$ , by (11), we have  $\|\Delta_h^N(\mathbf{f}_j^{w,b})\|_{\phi, S_p} < \frac{\epsilon}{2}$ , and

$$\begin{aligned} \|\Delta_h^N(\mathbf{f})\|_{\phi, S_p} &\leq c \left[ \|\Delta_h^N(\mathbf{f} - \mathbf{f}_j^{w,b})\|_{\phi, S_p} + \|\Delta_h^N(\mathbf{f}_j^{w,b})\|_{\phi, S_p} \right] \\ &\leq c \left[ 2^N \|\mathbf{f} - \mathbf{f}_j^{w,b}\|_{\phi, S_p} + \|\Delta_h^N(\mathbf{f}_j^{w,b})\|_{\phi, S_p} \right] \\ &\leq c \left[ 2^N \cdot \frac{\epsilon}{2^{N+1}} + \frac{\epsilon}{2} \right] = \epsilon \end{aligned}$$

In the following lemma, we give more generalized properties for our Orlicz modulus of smoothness.

**Lemma 2**

1.  $\omega_N(\mathbf{f}, \delta)_{\phi, S_p}$  is a positive nondecreasing continuous function of  $\delta$  on  $(0, \infty)$  and  $\lim_{\delta \rightarrow 0} \omega_N(\mathbf{f}, \delta)_{\phi, S_p} = 0$ .
2.  $\omega_N(\mathbf{f} + \mathbf{g}, \delta)_{\phi, S_p} \leq c \left( \omega_N(\mathbf{f}, \delta)_{\phi, S_p} + \omega_N(\mathbf{g}, \delta)_{\phi, S_p} \right)$
3.  $\omega_N(\mathbf{f}, \delta)_{\phi, S_p} \leq 2^{N-M} \omega_M(\mathbf{f}, \delta)_{\phi, S_p}$
4.  $\omega_N(\mathbf{f}, \delta)_{\phi, S_p} \leq 2^{\lfloor N \rfloor} \|\mathbf{f}\|_{\phi, S_p}$
5.  $\omega_N(\mathbf{f}, \delta)_{\phi, S_p} \leq \omega_N(\mathbf{f}, \delta)_{\phi, S_p}$ , for  $\delta \leq \delta$
6.  $\omega_N(\mathbf{f}, \gamma\delta)_{\phi, S_p} \leq (1 + \gamma)^k \omega_N(\mathbf{f}, \delta)_{\phi, S_p}$ , for  $0 < \gamma \leq 1$ .

**Proof**

1. It is clear from **Lemma 1**
2. From (13), (12) and **Lemma 1**, we have

$$\begin{aligned} \omega_N(\mathbf{f} + \mathbf{g}, \delta)_{\phi, S_p} &= \sup_{|h| \leq \delta} \|\Delta_h^N(\mathbf{f} + \mathbf{g})\|_{\phi, S_p} \\ &= \sup_{|h| \leq \delta} \|\Delta_h^N(\mathbf{f}) + \Delta_h^N(\mathbf{g})\|_{\phi, S_p} \\ &\leq c \left( \sup_{|h| \leq \delta} \|\Delta_h^N(\mathbf{f})\|_{\phi, S_p} + \sup_{|h| \leq \delta} \|\Delta_h^N(\mathbf{g})\|_{\phi, S_p} \right) \\ &\leq c \left( \omega_N(\mathbf{f}, \delta)_{\phi, S_p} + \omega_N(\mathbf{g}, \delta)_{\phi, S_p} \right) \end{aligned}$$

3. Again by (11), (12) and **Lemma 1**

$$\begin{aligned} \omega_N(\mathbf{f}, \delta)_{\phi, S_p} &= \sup_{|h| \leq \delta} \|\Delta_h^N(\mathbf{f})\|_{\phi, S_p} \\ &= \sup_{|h| \leq \delta} \|\Delta_h^{N-M}(\Delta_h^M(\mathbf{f}))\|_{\phi, S_p} \\ &\leq k(N - M) \sup_{|h| \leq \delta} \|\Delta_h^M(\mathbf{f})\|_{\phi, S_p} \\ &\leq 2^{N-M} \omega_M(\mathbf{f}, \delta)_{\phi, S_p}. \end{aligned}$$

4. By **Lemma 1**

$$\begin{aligned} \omega_N(\mathbf{f}, \delta) &= \sup_{|h| \leq \delta} \|\Delta_h^N(\mathbf{f})\|_{\phi, S_p} \leq 2^N \sup_{|h| \leq \delta} \|\mathbf{f}\|_{\phi, S_p} \\ &\leq 2^N \|\mathbf{f}\|_{\phi, S_p}. \end{aligned}$$

5. Let  $\delta < \delta$ , then by (12), it is clear that

$$\begin{aligned} \omega_N(\mathbf{f}, \delta)_{\phi, S_p} &= \sup_{|h| \leq \delta} \|\Delta_h^N(\mathbf{f})\|_{\phi, S_p} \\ &\leq \sup_{|h| \leq \delta} \|\Delta_h^N(\mathbf{f})\|_{\phi, S_p} = \omega_N(\mathbf{f}, \delta)_{\phi, S_p} \end{aligned}$$

6. Noting that  $\gamma\delta \leq \delta$ , then

$$\begin{aligned} \omega_N(\mathbf{f}, \gamma\delta)_{\phi, S_p} &\leq \omega_N(\mathbf{f}, \delta)_{\phi, S_p} \\ &\leq (1 + \gamma)^k \omega_N(\mathbf{f}, \delta)_{\phi, S_p}. \end{aligned}$$

**2. CONVOLUTIONAL NEURAL NETWORKS**

Deep learning is very important because of its ability to cope with large data in many quantities. One of the most common deep neural networks in deep learning is CNNs.

Since the 1950th, researchers have begun to work on a discovery through which visual data can be understood and developed called computer vision.

CNNs were developed in the 1980s, they were used only to identify handwritten numbers, read zip and pin codes etc. In 2012, there was a qualitative leap in the development of synaptic neural networks where Alex Krizhevsky neural networks was used to provide a wide range of data. Artificial neural networks have become deep neural networks, deep convolutional neural networks [35-42].

They are especially designed for processing pixel data and are used in image recognition and processing, applications of CNN recognition of pictures and videos, image classification, medical image analysis, language processing [43]. Zhou [44] provided, in 2018, a family of new deep structured neural networks: convolutional neural networks distributed in depth, show that these networks of deep neurons have the same order of computational complexity as the networks of deep convolutive neurons, and demonstrated their universality of approximation. Fang et al. [39] considered an applied family of deep convolutional neural networks functions of the unit sphere  $S^{d-1}$  of  $\mathbb{R}^d$  estimating its degree of approximation compared with the classical ones.

Now, we introduce general preliminaries about CNNs, beginning with the most prominent characteristic of any CNN, that is the convolution that imposes the network. In mathematics, convolution is a linear operation on two functions as follow

$$(f_1 \circ f_2)(x) = \sum f_1(t) f_2(x + t) \tag{13}$$

The following description shows the convolution procedure among matrices that construct CNNs. For any  $J \in \mathbb{N}$ , the depth of the network, a sequence  $w^{(j)}$ ,  $j = 1 \dots, J$  is a filter mask that is within  $\{0, 1 \dots, s^{(j)}\}$ , where  $s^{(j)} \in \mathbb{N}$  is the filter length with  $s^{(j)} + 1$  free parameters. So that the convolutional filter masks are

$$\{w^{(j)}: \mathbb{Z} \rightarrow \mathbb{R}\}_{j=1}^J \tag{14}$$

Let  $s \in \mathbb{N}$ , any filter mask  $w = (w_k)_{k=-\infty}^{\infty}$  lies within  $\{0, 1, \dots, s\}$ , satisfies  $w_k = 0$  if  $k \notin \{0, 1, \dots, s\}$ . This implies a convolutional matrix  $\mathfrak{S}^w = (w_{i-k}) \in \mathbb{R}^{(D+s) \times D}$ , for  $i = 1, \dots, D + s, k = 1, \dots, D$ , and  $D \in \mathbb{N}$ , that is given by

$$\mathfrak{S}^w = \begin{bmatrix} w_0 & 0 & 0 & 0 & \dots & 0 \\ w_1 & w_0 & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ w_s & w_{s-1} & \dots & w_0 & 0 & 0 \\ 0 & w_s & \dots & w_1 & w_0 & 0 \\ \vdots & \ddots & \dots & \ddots & \ddots & \vdots \\ \dots & \dots & \ddots & \dots & \dots & w_0 \\ \dots & \dots & \dots & \ddots & \dots & w_1 \\ \vdots & \ddots & \ddots & \dots & \dots & \vdots \\ 0 & \dots & \dots & 0 & w_s & w_{s-1} \\ 0 & \dots & \dots & \dots & 0 & w_s \end{bmatrix}$$

In their papers [39, 44, 45], Zhou and his team proposed the convolutional matrix as follow  $\mathfrak{S}^{(j)} = \mathfrak{S}^{w^{(j)}}$ , with  $D = d_{j-1}$  and  $s = s^{(j)}$ .

Deep CNN (DCNN) is given by convolutions of the activation functions  $\sigma_{\mathfrak{S}, b}: \mathbb{R}^{d_{j-1}} \rightarrow \mathbb{R}^{d_{j-1}+s^{(j)}}$  of matrix  $\mathfrak{S}$  and bias  $b \in \mathbb{R}^{d_{j-1}+s^{(j)}}$ .

To get DCNN  $h^j: \mathbb{R}^d \rightarrow \mathbb{R}^{d_j}$ ,  $j = 1, \dots, J$ .  $d_j = d + js, s^{(j)} \equiv s, j = 1, \dots, J$ , the form

$$h^{(j)} = \sigma_{\mathfrak{S}^{(j)}, b^{(j)}} \circ \dots \circ \sigma_{\mathfrak{S}^{(1)}, b^{(1)}}(x) \quad (15)$$

where,  $\sigma_{\mathfrak{S}, b}(u) = \sigma(\mathfrak{S}u - b), u \in \mathbb{R}^{d_{j-1}}, b \in \mathbb{R}^{d_{j-1}+s^{(j)}}$ .

The real valued activation function  $\sigma$  is given by

$$\sigma(u) = \max\{u, 0\}, \quad u \in \mathbb{R} \quad (16)$$

### 3. CONSTRUCTION OF CNNs

For any input  $x = (x_i)_{i=1}^d$ , define the oprater  $\mathcal{L}_x: L_{\phi, S_p}(I^d) \rightarrow \mathbb{R}^+$  as follow

$$\mathcal{L}_t(g(u)) = \sum_{i=1}^d g(x_i) \delta_i(u) \quad (17)$$

where,  $g \in L_{\phi, S_p}, u \in I$  and  $\delta_i$  is given by

$$\delta_{i,g}(u) = \sum_{i=2}^d (-1)^{N-i} \binom{N}{i} g(t_i) \sigma(t_i - u) \quad (18)$$

where,  $\sigma: I \rightarrow I$  is the ReLU activation function defined by (16).

In fact, Eq. (17) is a convolutional operator between the function  $g$  and the activation function  $\sigma$ . Moreover,  $x = (x_i)_{i=1}^d$  maybe not only the input, but also any neuron acts as input in the network. This amount of convolutions implies high dimensions of the hidden layer when the outputs of neuron clusters are combined. The purpose of the pooling layers is to reduce the big widths that may cause overfitting. Another procedure that can reduce the dimensions, called downsampling, is discussed in the following section.

### 4. DOWNSAMPLING DEEP CONVOLUTIONAL NEURAL NETWORK

Zhou [38] provided the definition of downsampling operator for purposes of reducing widths, he introduced a downsampled operation. The  $\ell$  downsampled are presented at layers  $\mathcal{J} = \{J_k\}_{k=1}^{\ell}$  with  $1 < J_1 \leq \dots \leq J_{\ell} = J$ . His concept of downsampling operators is induced from wavelets [46, 47]. In this section, we introduce the effect of downsampling operation into 4. Downsampling Deep Convolutional Neural Network (DDCNN) to avoid big widths in (15) that happens with pooling layers. The downsampling operation is defined below,

**Definition 2.** Let  $m$  be a scaling parameter, the function  $\mathfrak{D}_m: \mathbb{R}^D \rightarrow \mathbb{R}^{\lfloor D/m \rfloor}$  is called downsampling operator, and it is given by

$$\mathfrak{D}_m(u) = (u_{im})_{i=1}^{\lfloor D/m \rfloor}, u \in \mathbb{R}^D \quad (19)$$

a scaling parameter  $m \leq D$ . where  $[u]$  is the integer part of  $u \in \mathbb{R}^+$ .

**Definition 3.** Downsampling DCNN is defined iteratively for  $k = 1, \dots, \ell$ , beginning with  $d_0 = d$  and filter lengths  $[s^{(j)}]_{j=1}^J$  has widths  $[d_j]_{j=0}^J$  where  $\ell$  downsamplings at layers  $\mathcal{J}$

$$d_j = \begin{cases} d_{j-1} + s^{(j)}, & \text{if } J_{k-1} < j < J_k \\ \lfloor (d_{j-1} + s^{(j)}) / d_{J_{k-1}} \rfloor, & \text{if } j = J_k \end{cases} \quad (20)$$

Define function vector sequence iteratively  $\{h^{(j)}(x): \mathbb{R}^d \rightarrow \mathbb{R}^{d_j}\}_{j=1}^J$  by  $h^0(x) = x$  and for  $k = 1, \dots, \ell$ .

$$h^{(j)}(x) = \begin{cases} \sigma_{\mathfrak{S}^{(j)}, b^{(j)}}(h^{(j-1)}(x)), & \text{if } J_{k-1} < j < J_k \\ \mathfrak{D}_{d_{J_{k-1}}} \circ \sigma_{\mathfrak{S}^{(j)}, b^{(j)}}(h^{(j-1)}(x)), & \text{if } j = J_k \end{cases} \quad (21)$$

We restrict the  $\mathfrak{S} b^{(j)} \in \mathbb{R}^{d_{j-1}+s^{(j)}}$  to satisfy

$$b_{s^{(j)}+1}^{(j)} = b_{s^{(j)}+2}^{(j)} = \dots = b_{d_{j-1}}^{(j)} \quad (22)$$

**Definition 4.** Uniform DDCNN is defined with uniform filter lengths for  $k \in \{1, \dots, \ell\}$ , as follow

$$S = \{s^{[k]} \in \mathbb{N}\}_{k=1}^{\ell} \text{ if } s^{(J_{k-1}+1)} = \dots = s^{(J_k)} = s^{[k]} \quad (23)$$

Define the degree of best approximation of functions  $g \in L_{\phi, S_p}(I^d)$  by DCNN of the form

$$E_m(g)_{\phi, S_p} = \inf \|g - L_t(g)\|_{\phi, S_p} \quad (24)$$

### 5. AUXILIARY LEMMAS

The following lemma arises the role of convolutions [44] concluded that  $W^{(k)}$  equals  $\mathfrak{S}^{(J_k)} \dots \mathfrak{S}^{(J_{k-1}+1)}$ .

**Lemma 3**

$\mathfrak{S}^{(J_k J_{k-1}+1)} = \mathfrak{S}^{(J_k)} \dots \mathfrak{S}^{(J_{k-1}+2)} \mathfrak{S}^{(J_{k-1}+1)}$  where  $k = 1, \dots, \ell$ .

Construction of filter masks is studied by Zhou [48] in his following lemma, given convolution  $w^{(0)} * \dots * w^{(1)}$  resulted from factorizing  $W$  as follows

**Lemma 4**

For  $s \geq 2$ ,  $M \geq 0$ , any sequence  $W = (w_k)_{-\infty}^{\infty}$  in  $\{0, \dots, M\}$ , there is a finite sequence  $\{w^{(j)}\}_{j=1}^J$  with nonzero elements in  $\{0, \dots, s\}$ , where  $J < \frac{M}{s-1} + 1$  that satisfies

$$W = w^{(j)} * \dots * w^{(2)} * w^{(1)} \tag{25}$$

To fit the data with the space under study, bias vectors are selected in appropriate format in accordance to the convolutional matrices  $\mathfrak{S}^w$ , let  $h: \Omega \rightarrow R^D$ , set

$$\|h\|_{\phi, s_p} = \inf_{\lambda > 0} \frac{1}{\lambda} (I_{\phi}^2(\lambda h) + 1)^{\frac{1}{p}} \tag{26}$$

Also, we need to denote  $\|w\|_{\phi, s_p} = \inf_{\lambda > 0} \frac{1}{\lambda} (I_{\phi}^2(\lambda w_k) + 1)^{\frac{1}{p}}$ .

Then we immediately see how convolutional matrices are defined  $\mathfrak{S}^{(j)} = \mathfrak{S}^{w^{(j)}}$  that for any  $h: \Omega \rightarrow \mathbb{R}^{d_{j-1}}$ , we denoted

$$\|\mathfrak{S}^{(j)} h\|_{\phi, s_p} \leq \|w\|_{\phi, s_p} \|h\|_{\phi, s_p} \tag{27}$$

From our previous study AL-Janabi and Almurieb [13], we use some thoughts from DCNNs without downsampling and select biases small enough for the vectors  $\mathfrak{S}^{(j)} h^{(j-1)}(x) - b^{(j)}$  to get non-negative entries.

**Lemma 5**

For any  $k \in \{1, \dots, \ell\}$ , there exists  $B \in R^+$ ,  $\hat{B} \in [-B, B]$ , then  $\|h^{j_{k-1}} - \hat{B} \mathbf{1}_{d_{j_{k-1}}}\|_{\phi, s_p} \leq B$ .

**Proof**

Set No space

$$b^{(j_{k-1}+1)} = \hat{B} \mathfrak{S}^{(j_{k-1}+1)} \mathbf{1}_{d_{j_{k-1}}} - B \|w^{(j_{k-1}+1)}\|_{\phi, s_p} \mathbf{1}_{d_{j_{k-1}+1}}$$

To prove  $b_{s^{(j)}+1}^{(j)} = b_{s^{(j)}+2}^{(j)} = \dots = b_{d_{j-1}}^{(j)}$

From

$$b^{(j)} = B \left( \prod_{p=j_{k-1}+1}^{j-1} \|w^{(p)}\|_{\phi, s_p} \right) \mathfrak{S}^{(j)} \mathbf{1}_{d_{j-1}} - B \left( \prod_{p=j_{k-1}+1}^{j-1} \|w^{(p)}\|_{\phi, s_p} \right) \mathbf{1}_{d_{j-1+s^{(j)}}}$$

For  $j = j_{k-1} + 2, \dots, j_{k-1}$ , then for  $j_{k-1} < j < j_k$  and  $i = s^{(j)} + 1, \dots, d_{j-1}$ , we have  $(\mathfrak{S}^{(j)} \mathbf{1}_{d_{j-1}})_i = \sum_{p=1}^{d_{j-1}} (\mathfrak{S}^{(j)})_{i,p} = \sum_{p=1}^{d_{j-1}} (w^{(j)})_{i-p}$ .

Notice that  $w^{(j)}$  in  $\{0, \dots, s^{(j)}\}$ .  $p \in (-\infty, 0] \cup [d_{j-1} + 1, \infty)$  we have  $i - p \in [s^{(j)} + 1, \infty)$  which implies that  $w^{(j)}_{i-p} = 0$ . So

$$(\mathfrak{S}^{(j)} \mathbf{1}_{d_{j-1}})_i = \sum_{p=-\infty}^{\infty} w^{(j)}_{i-p} = \sum_{p=-\infty}^{\infty} w_p^{(j)}$$

$$\forall i = s^{(j)} + 1, \dots, d_{j-1}$$

Now we prove

$$h^{(j)}(x) = \mathfrak{S}^{(j)} \dots \mathfrak{S}^{(j_{k-1}+1)} \left( h^{j_{k-1}}(x) - \hat{B} \mathbf{1}_{d_{j_{k-1}}} \right) + B \left( \prod_{p=j_{k-1}+1}^j \|w^{(p)}\|_{\phi, s_p} \right) \mathbf{1}_{d_j}$$

By induction, for  $j = j_{k-1} + 1$ , we have

$$\begin{aligned} &\mathfrak{S}^{(j_{k-1}+1)} h^{(j_{k-1})}(x) - b^{(j_{k-1}+1)} \\ &= \mathfrak{S}^{(j_{k-1}+1)} \left( h^{j_{k-1}}(x) - \hat{B} \mathbf{1}_{d_{j_{k-1}}} \right) \\ &+ B \|w^{(j_{k-1}+1)}\|_{\phi, s_p} \mathbf{1}_{d_{j_{k-1}+1}} \end{aligned}$$

the use Sigmoid  $\sigma$  is identical to the identity function on  $[0, \infty)$ , hence

$$h^{(j_{k-1}+1)}(x) = \mathfrak{S}^{(j_{k-1}+1)} \left( h^{j_{k-1}}(x) - \hat{B} \mathbf{1}_{d_{j_{k-1}}} \right) + B \|w^{(j_{k-1}+1)}\|_{\phi, s_p} \mathbf{1}_{d_{j_{k-1}+1}}$$

for  $j = 1$  verifies.

Suppose that it is holds for  $j_{k-1} + 1 \leq j \leq j_{k-1}$ , then  $d_{j-1} + s^{(j)} = d_j$ .

Through the induction hypothesis and selection (\*) of the bias vector, we have

$$h^{(j)}(x) = \sigma \left( \mathfrak{S}^{(j)} \mathfrak{S}^{(j-1)} \dots \mathfrak{S}^{(j_{k+1})} \left( h^{(j_{k-1})}(x) - \hat{B} \mathbf{1}_{d_{j_{k-1}}} \right) + B \left( \prod_{p=j_{k-1}+1}^j \|w^{(p)}\|_{\phi, s_p} \right) \mathbf{1}_{d_j} \right)$$

Now to prove

$$h^{(j_k)}(x) = M^{(k)} \mathfrak{S}^{(j_k j_{k-1}+1)} \left( h^{(j_{k-1})}(x) - \hat{B} \mathbf{1}_{d_{j_{k-1}}} \right) + B \left( \prod_{p=j_{k-1}+1}^{j_k} \|w^{(p)}\|_{\phi, s_p} \right) \mathbf{1}_{d_{j_k}}$$

When

$$b^{(j)} = B \left( \prod_{p=j_{k-1}+1}^{j-1} \|w^{(p)}\|_{\phi, s_p} \right) \mathfrak{S}^{(j)} \mathbf{1}_{d_{j-1}} - B$$

Then

$$\begin{aligned} \mathfrak{S}^{(j_k)} h^{(j_{k-1})}(x) - b^{(j_k)} &= \mathfrak{S}^{(j_k)} \dots \mathfrak{S}^{(j_{k-1})} \left( h^{(j_{k-1})}(x) - \hat{B} \mathbf{1}_{d_{j_{k-1}}} \right) \\ &+ B \left( \prod_{p=j_{k-1}+1}^{j_k} \|w^{(p)}\|_{\phi, s_p} \right) \mathbf{1}_{d_{j_k}} \end{aligned}$$

**6. APPROXIMATION ABILITIES OF CNNs IN ORLICZ SPACES**

The main purpose of this paper is to study degree of approximation of functions from Orlicz spaces by convolutional neural networks in terms of Orlicz modulus of smoothness. The following theorem study direct theorem:

**Theorem 1 (Direct theorem)**

For any  $g \in L_{\phi, s_p}(I^d)$ , there exists a CNN of the form,  $L_t(g)(u) = \sum_{i=2}^{2N+2} g(t_i) \delta_i(u)$ ,  $u \in [t_{i-1}, t_i]$ , where  $|t_i -$

$t_{i-1} \sim \frac{1}{n}$  and  $\delta_i(u) = \sum_{i=1}^N \binom{N}{i} (-1)^{N-i} \sigma(t_i - u)$ , such that  $\|L_t(\mathbf{g}) - \mathbf{g}\|_{\Phi, S_p}^p \leq \frac{c}{n} \omega_N(\mathbf{g}, h)_{\Phi, S_p}$ .

**Proof**

By (17), (12) and **Lemma 2**, we get

$$\begin{aligned} \|L_t(\mathbf{g}) - \mathbf{g}\|_{\Phi, S_p}^p &\leq \left\| \sum_{i=2}^{2N+2} \left[ (-1)^{N-i} \binom{N}{i} \mathbf{g}(t_i) \sigma(t_i - u) \right. \right. \\ &\quad \left. \left. - (-1)^{N-i} \binom{N}{i} \mathbf{g}\left(u - \frac{i}{n}\right) \right. \right. \\ &\quad \left. \left. + (-1)^{N-i} \binom{N}{i} \mathbf{g}\left(u - \frac{i}{n}\right) - \mathbf{g}(u) \right] \right\|_{\Phi, S_p}^p \\ &\leq c \left( \left\| \sum_{i=2}^{2N+2} (-1)^{N-i} \binom{N}{i} \mathbf{g}(t_i) \sigma(t_i - u) \right. \right. \\ &\quad \left. \left. - (-1)^{N-i} \binom{N}{i} \mathbf{g}\left(u - \frac{i}{n}\right) \right\|_{\Phi, S_p}^p \right. \\ &\quad \left. + \left\| (-1)^{N-i} \binom{N}{i} \mathbf{g}\left(u - \frac{i}{n}\right) - \mathbf{g}(u) \right\|_{\Phi, S_p}^p \right) \\ &\leq c \sum_{i=2}^{2N+2} \left\| (-1)^{N-i} \binom{N}{i} \left( \frac{1}{n} \mathbf{g}(t_i) \right. \right. \\ &\quad \left. \left. - \mathbf{g}\left(u - \frac{i}{n}\right) \right) \right\|_{\Phi, S_p}^p + \omega_N\left(\mathbf{g}, \frac{1}{n}\right)_{\Phi, S_p} \\ &\leq \frac{c}{n} \omega_N\left(\mathbf{g}, \frac{1}{n}\right)_{\Phi, S_p} \end{aligned}$$

In the next last theorem, we study inverse theorem that gives lower bound for the degree of approximation, it restricts, with direct theorem, the degree of approximation with lower and upper bounds, respectively.

**Theorem 2 (Inverse Theorem)**

For any  $\mathbf{g} \in L_{\Phi, S_p}(I^d)$ , there exists  $L_t(\mathbf{g})$  of the form (17), such that  $\omega_N(\mathbf{g}, \delta)_{\Phi, S_p} \leq c(p, N) \left( \|\mathbf{g}\|_{\Phi, S_p} + E_N(\mathbf{g})_{\Phi, S_p} \right)$ .

**Proof**

Let  $b = \max_{1 \leq i \leq 2N+3} i; 2^i < n$ , and  $\mathbf{g}(t_{2N+3}) - \mathbf{g}(t_1) = \left( \mathbf{g}(t_{2N+3}) - \mathbf{g}(t_{2^b}) \right) + \left( \mathbf{g}(t_{2^b}) - \mathbf{g}(t_{2^{b-1}}) \right) + \dots + \left( \mathbf{g}(t_2) - \mathbf{g}(t_1) \right)$ , for any  $m < 2N + 3$ , and  $\delta \sim \frac{1}{n}$ .

Suppose that  $\|\mathbf{g}(t_{2N+3}) - \mathbf{g}(t_m)\|_{\Phi, S_p} \leq cE_m(\mathbf{g})_{\Phi, S_p}$ , then by (24), **Theorem 1** we get by **Lemma 2**.

$$\begin{aligned} \omega_N(\mathbf{g}, \delta)_{\Phi, S_p} &\leq c(p) \left[ \omega_N(\mathbf{g} - L_t(\mathbf{g}), \delta)_{\Phi, S_p} \right. \\ &\quad \left. + \omega_N(L_t(\mathbf{g}), \delta)_{\Phi, S_p} \right] \\ &\leq c(p) \left[ c(N) \|\mathbf{g} - L_t(\mathbf{g})\|_{\Phi, S_p} \right. \\ &\quad \left. + \omega_N(L_t(\mathbf{g}), \delta)_{\Phi, S_p} \right] \\ &\leq c(p) \left[ c(N) \omega_N(\mathbf{g}, \delta)_{\Phi, S_p} \right. \\ &\quad \left. + c(N) E_m(\mathbf{g})_{\Phi, S_p} \right] \\ &\leq c(p, N) \left[ \|\mathbf{g}\|_{\Phi, S_p} + E_m(\mathbf{g})_{\Phi, S_p} \right]. \end{aligned}$$

**7. CONCLUSIONS**

Quasi-Orlicz space is a very large space of measurable convex functions that generalize Orlicz space. In quasi-Orlicz spaces, we define a special class of CNNs with downsampling operator to reduce dimensions and approximate any quasi-Orlicz function with a good degree of approximation. Some previous work focused on approximating continuous functions by CNNs. CNNs were constructed in a way that guarantee universality with ordinary degree of approximation that depends on parameters of CNNs’ features. This paper results universality of CNNs with more extended space. On the other hand, we define modulus of smoothness in quasi-Orlicz space with many properties that are useful for approximation purposes that improve the estimated degree of approximation. The degree of approximation then is given by modulus of smoothness with its upper and lower bounds. Hence, we prove universality of CNNs in quasi- Orlicz spaces with dependence on smoothness of functions for their degree of approximation. In future work, it is interesting to implement such CNNs with downsampling operator practically in fields with complicated data.

**REFERENCES**

- [1] Orlicz, W. (1932). Über Eine GEWISSE Klasse von Räumen vom typus B. Bulletin of the Polish Academy of Sciences Technical Sciences, 8(9): 207-220.
- [2] Nakano, H. (1951). Topology and Linear Topological Spaces. Maruzen Company.
- [3] Morse, M., Transue, W. (1950). Functionals f bilinear over the product a x b of two pseudo-normed vector spaces: II. admissible spaces a. Annals of Mathematics, 51(3): 576-614.
- [4] Luxemburg, W.A.J. (1955). Banach Function Spaces.
- [5] Kranoselkii, M., Rutickii, Y.B. (1961). Convex Functions and Orlicz Spaces (translation). ed: Groningen.
- [6] Luxemburg, W., Zaanan, A. (1956). Conjugate spaces of Orlicz spaces. In Indagationes Mathematicae (Proceedings), 59: 217-228.
- [7] Cui, Y., Duan, L., Hudzik, H., Wisła, M. (2008). Basic theory of p-Amemiya norm in Orlicz spaces (1 ≤ p ≤ ∞): Extreme points and rotundity in Orlicz spaces endowed with these norms. Nonlinear Analysis: Theory, Methods & Applications, 69(5-6): 1796-1816. <https://doi.org/10.1016/j.na.2007.07.026>
- [8] Wisła, M. (2015). Geometric properties of Orlicz spaces equipped with p-Amemiya norms—results and open questions. Commentationes Mathematicae, 55(2): 183-209. <http://doi.org/10.14708/cm.v55i2.1104>
- [9] Wisła, M. (2020). Orlicz spaces equipped with s-norms. Journal of Mathematical Analysis and Applications, 483(2): 123659. <https://doi.org/10.1016/j.jmaa.2019.123659>
- [10] Shidlich, A.L., Chaichenko, S.O. (2015). Approximative properties of diagonal operators in Orlicz spaces. Numerical Functional Analysis and Optimization, 36(10): 1339-1352. <https://doi.org/10.1080/01630563.2015.1066387>
- [11] Abdullayev, F., Chaichenko, S., Kyzy, M.I., Shidlich, A. (2020). Direct and inverse approximation theorems in the weighted Orlicz-type spaces with a variable exponent. Turkish Journal of Mathematics, 44(1): 284-299.
- [12] Cui, Y., Wisła, M. (2021). Monotonicity properties and

- solvability of dominated best approximation problem in Orlicz spaces equipped with  $s$ -norms. *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales. Serie A. Matemáticas*, 115: 1-22. <https://doi.org/10.1007/s13398-021-01139-8>
- [13] AL-Janabi, A.M., Almurieb, H.A. (2024). Convolutional neural networks approximation in quasi-Orlicz spaces on spheres. *Journal of Kufa for Mathematics and Computer*, 11(1): 1-5. <https://doi.org/10.3906/mat-1911-3>
- [14] Jackson, D. (1911). Über die Genauigkeit der Annäherung stetiger Funktionen durch ganze rationale Funktionen gegebenen Grades und trigonometrische Summen gegebener Ordnung. Dieterich'schen Universität--Buchdruckerei.
- [15] Bernstein, S.N. (1912). On the best approximation of continuous functions by polynomials of a given degree. *Kharkov Mathematical Society*, 2(13): 49-194.
- [16] de la Valle-Poussin, C. (1919). Leçons sur l'approximation des fonctions d'une variable réelle.
- [17] Zygmund, A. (1924). On the continuity module of the sum of the series conjugate to a Fourier series. *Prace Matematyczno-Przyrodnicze. Matematyka. Fizyka*, 33: 25-132.
- [18] Butzer, P.L., Nessel, R.J. (1971). Fourier analysis and approximation, Vol. 1. Reviews in Group Representation Theory, Part A (Pure and Applied Mathematics Series, Vol. 7).
- [19] DeVore, R.A., Lorentz, G.G. (1993). *Constructive Approximation* (vol. 303). Springer Science & Business Media.
- [20] Dzyadyk, V.K., Shevchuk, I.A. (2008). Theory of uniform approximation of functions by polynomials. Walter de Gruyter. <https://doi.org/10.1515/9783110208245.bm>
- [21] Timan, M.F. (2009). *Approximation and Properties of Periodic Functions*. Nauk. dumka, Kiev.
- [22] Sterlin, M.D. (1972). Exact constants in inverse theorems in the theory of approximations. In *Doklady Akademii Nauk*, 202(3): 545-547.
- [23] Ditzian, Z., Totik, V. (1987). *Springer series in computational mathematics*. Springer-Verlag New York.
- [24] Musielak, H. (1985). On the  $\tau$ -modulus of smoothness in generalized Orlicz spaces. *Commentationes Mathematicae*, 25(2): 285-293.
- [25] Popov, V. (1980). On the one-sided approximation on functions. *Constructive Function*, 465-468.
- [26] Popov, V.A., Khristov, V.K. (1983). Averaged moduli of smoothness for functions of several variables and function spaces generated by them. *Trudy Matematicheskogo Instituta imeni VA Steklova*, 164: 136-141.
- [27] Tsyganok, I. (1966). A generalization of a theorem of Jackson. *Matematicheskii Sbornik*, 113(2): 257-260.
- [28] Ramazanov, A.R. (1984). On approximation by polynomials and rational functions in Orlicz spaces. *Analysis Mathematica*, 10(2): 117-132.
- [29] Garidi, W. (1991). On approximation by polynomials in Orlicz spaces. *Approximation Theory and Its Applications*, 7(3): 97-110. <https://doi.org/10.1007/BF02836460>
- [30] Israfilov, D.M., Akgün, R. (2006). Approximation in weighted Smirnov-Orlicz classes. *Journal of Mathematics of Kyoto University*, 46(4): 755-770. <https://doi.org/10.1215/kjm/1250281602>
- [31] Akgün, R. (2012). Approximating polynomials for functions of weighted Smirnov-Orlicz Spaces. *Journal of Function Spaces*, 2012(1): 982360. <https://doi.org/10.1155/2012/982360>
- [32] Chaichenko, S., Shidlich, A., Abdullayev, F. (2019). Direct and inverse approximation theorems of functions in the Orlicz type spaces  $\mathcal{S}_M$ . *Mathematica Slovaca*, 69(6): 1367-1380. <https://doi.org/10.1515/ms-2017-0314>
- [33] Shidlich, A.L., Chaichenko, S.O. (2014). Some extremal problems in the Orlicz spaces. *Matematychni Studii*, 42(1): 21-32.
- [34] Shidlich, A.L., Chaichenko, S.O. (2015). Approximative properties of diagonal operators in Orlicz spaces. *Numerical Functional Analysis and Optimization*, 36(10): 1339-1352. <https://doi.org/10.1080/01630563.2015.1066387>
- [35] LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. *nature*, 521(7553): 436-444. <https://doi.org/10.1038/nature14539>
- [36] Panwar, H., Gupta, P.K., Siddiqui, M.K., Morales-Menendez, R., Bhardwaj, P., Singh, V. (2020). A deep learning and grad-CAM based color visualization approach for fast detection of COVID-19 cases using chest X-ray and CT-Scan images. *Chaos, Solitons & Fractals*, 140: 110190. <https://doi.org/10.1016/j.chaos.2020.110190>
- [37] LeCun, Y., Bottou, L., Bengio, Y., Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11): 2278-2324. <https://doi.org/10.1109/5.726791>
- [38] Zhou, D.X. (2020). Theory of deep convolutional neural networks: Downsampling. *Neural Networks*, 124: 319-327. <https://doi.org/10.1016/j.neunet.2020.01.018>
- [39] Fang, Z., Feng, H., Huang, S., Zhou, D.X. (2020). Theory of deep convolutional neural networks II: Spherical analysis. *Neural Networks*, 131: 154-162. <https://doi.org/10.1016/j.neunet.2020.07.029>
- [40] Jadhav, S.B. (2019). Convolutional neural networks for leaf image-based plant disease classification. *IAES International Journal of Artificial Intelligence*, 8(4): 328.
- [41] Dittmer, S., King, E.J., Maass, P. (2019). Singular values for ReLU layers. *IEEE Transactions on Neural Networks and Learning Systems*, 31(9): 3594-3605. <https://doi.org/10.1109/TNNLS.2019.2945113>
- [42] Kadarina, T.M., Iklima, Z., Priambodo, R., Riandini, R., Wardhani, R.N. (2023). Dental caries classification using depthwise separable convolutional neural network for teledentistry system. *Bulletin of Electrical Engineering and Informatics*, 12(2): 940-949. <https://doi.org/10.11591/eei.v12i2.4428>
- [43] Collobert, R., Weston, J. (2008). A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th International Conference on Machine Learning, Helsinki Finland*, pp. 160-167. <https://doi.org/10.1145/1390156.1390177>
- [44] Zhou, D.X. (2018). Deep distributed convolutional neural networks: Universality. *Analysis and Applications*, 16(06): 895-919. <https://doi.org/10.1142/S0219530518500124>
- [45] Mao, T., Shi, Z., Zhou, D.X. (2021). Theory of deep convolutional neural networks III: Approximating radial functions. *Neural Networks*, 144: 778-790. <https://doi.org/10.1016/j.neunet.2021.09.027>

- [46] Daubechies, I. (1992). Ten lectures on wavelets. Society for Industrial and Applied Mathematics. SIAM.
- [47] Mallat, S. (2016). Understanding deep convolutional networks. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2065): 20150203. <https://doi.org/10.1098/rsta.2015.0203>
- [48] Zhou, D.X. (2020). Universality of deep convolutional neural networks. Applied and Computational Harmonic Analysis, 48(2): 787-794. <https://doi.org/10.1016/j.acha.2019.06.004>