Convolutional Neural Networks Approximation in Quasi-Orlicz Spaces on Sphers

Amna Manaf AL-Janabi
Department of Mathematics of, College of Education for Pure Sciences
University of Babylon,
Hillah, Babylon, Iraq.
amnammn22@gmail.com
Orcid.org/0009-0000-8875-9289

Hawraa Abbas Almurieb
Department of Mathematics of , College of Education for Pure Sciences
University of Babylon,
Hillah, Babylon, Iraq
pure.hawraa.abbas@uobabylon.edu.iq
Orcid.org/0000-0001-6888-0903

DOI: http://dx.doi.org/10.31642/JoKMC/2018/110101

Received Jun. 29, 2023. Accepted for publication Jul. 13, 2023

Abstract *— It is necessary to study the theoretical bases of an approximation deep convolutional neural networks, because of its interesting developments in vital domains. The approximation abilities of deep-convolution neural networks produced by downsampling operators in quasi- Orlicz spaces have been studied, since this space is wider and more important than other spaces. In this paper, we define quasi-Orlicz norm on spherical spaces. In addition, modulus of smoothness is also studied in terms of quasi-Orlicz norm. Finally, Function approximation theorems are studied by using convolution neural networks with k fully connected layer so that the error is resulted to be equivalent to double k-th order modulus of smoothness*

Keywords— Approximation, Quasi-Orlicz, Modulus of Smoothness, Convolution Neural Network

1. INTRODUCTION TO QUASI -ORLICZ SPACE

Orlicz[1] was the first who defined these kinds of spaces, named after his name, Orlicz spaces. He generalized L_p spaces, when $1 \leq p \leq \infty$, that are given by

$$L_p(I) = \{f: I \to R \text{ measurable}, \quad \|f\|_p < \infty\}$$

Moreover, Lp quasi-normed spaces are well studied for the values $0 < p < 1$, for example[2],
Orlicz space, in its primitive definition, is given by $L_\Phi(I) = \{f: f \text{ is } \mu - \text{measurable and } \|f\|_\Phi < \infty\}$
To date, Orlicz spaces have been examined by numerous authors in different ways. However, most of these studies used different definitions for the norms that define Orlicz spaces. In simultaneous times in the fifteenth of the last century, Nakano, Morse-Transue ,and Luxembourg[2] [3] ()investigated the Luxembourg norm, which has been defined, using the concept of functional Minikowski over a convex modular unit ball. Then not long later, Amemiya[2]defined another norm, named later, Amemiya norm. In separated papers, Krasnoselskii and RutickiI, Nakano[2] and Luxembourg and Zaneen[4] proved under additional conditions, that Amemiy norm is exactly the Orlicz norm. In 2000, Hudzik and Maligranda[5]suggested investigating the Amemiya formula generated by outer functions of the type $S_p(u) = (1 + u^p)^{\frac{1}{p}}$, where $1 \leqslant p \leqslant \infty$. That is called later, p- Amemiya norm.

For the case $0 < p < 1$, the paper [6] is an important contribution to define quasi - Orlicz spaces via the outer function $S_p: [0, \infty) \to [0, \infty)$. It is shown in the following definition,

Definition1. [6] For $0 < p < 1$ the family S_p of outer functions is defined by:

$$S_p(f) = (1 + f^2)^{\frac{1}{p}}, \qquad (1)$$

where $f \in L_{\Phi,S_p}(E), E \subseteq R^d$, and the quasi-normed space is given by

$$L_{\Phi,S_p}(E) = \left\{f \in L_p(E) \,\middle|\, \|f\|_{\Phi,S_p} < \infty\right\}, \qquad (2)$$

and

$$\|f\|_{\Phi,S_p} = \inf_{\lambda>0} \frac{1}{\lambda}(\lambda f^2 + 1)^{\frac{1}{p}} \qquad (3)$$

As mentioned there[6] , $S_p, p < 1$, are convex, strictly increasing and all S_p match at zero. Moreover, the quasi-normed space L_{Φ,S_p} , $p < 1$ is a generalization of p-Amemiy's Orlicz normed space on $p \geq 1$.

On the other hand, moduli of smoothness had been studied in Orlicz space. The significance of this is due to the need to improve the degree of function approximation through direct approaches. Approximating direct approaches presupposes a convergence towards zero faster than the

previous estimates. On the other hand, converse theorems give a characterization of smoothness of functions depending on its degree of approximation in direct theorem.

2. QUASI-ORLICZ ON SPHERICAL SPACES

This paper deals with approximation for quasi -Orlicz functions on the sphere $\mathbb{S}^d$. provided a consistent account of recent trends in approximation theory and harmonic analysis in these domains. Unit sphere analysis is a part of Fourier analysis, Best approximation of L_p , ≥ 1, functions out of spherical polynomials was studied by Dai and Xu in ther paper and book [7] Their big challenge was to define modulus of smoothness on $^{-1}$, when ≥ 3. The difficulty was that multiplication on three or more-dimensional sphere is not commutative. They defined modulus of smoothness on $_{}(^{-1}),I \leq \leq \infty$, depending on that of $_{}(^{1})$and used them for trigonometric approximation in their paper [5]

In this paper, we define a modulus of smoothness on the quasi – Orlicz space $_{\Phi,}(^{-1})$, begining with the *k*-th symmetric difference from[8],

For $r = 1,2,\ldots$, we use $Q_{i,j,\mathrm{t}}$ to determine the symmetric difference operator

$$\Delta_{i,j,\theta}^{r} = \left(I - T\left(Q_{i,j,\theta}\right)\right)^{r}, \quad 1 \leq i = j \leq d,$$

$Q_{i,j,\theta}$ is a rotation by the angle t in the plane (x_i, x_j), orientated in such a way that the rotation from the vector e_i at vector e_j is supposed to be positive.

Now we define modulus of smoothness on the space$L_{\Phi,S_p}(\mathbb{S}^d)$

Definition 2.1. For $r \in N$, $t > 0$, and $f \in L_{\Phi,S_p}(\mathbb{S}^d)$, $0 \leq p < 1$,

$$\omega_r(f\,;t)_{\Phi,S_p} := \max_{1\leq i<j\leq d} \sup_{|\theta|\leq t} \left\|\Delta_{i,j,\theta}^{r}(f)\right\|_{\Phi,S_p}$$

Theorem 2.2. Let $f \in L_{\Phi,S_p}(\mathbb{S}^d)$, then it is clear that

1. $\omega_r(f\,,t)_{\Phi,S_p}$ is a positive nondecreasing continuous function of δ on $(0,\infty)$
2. $\lim_{\delta\to 0} \omega_r(f,t)_{\Phi,S_p} = 0$
3. $\omega_r(f+g\,,t)_{\Phi,S_p} \leq c\left(\omega_r(f\,,t)_{\Phi,S_p} + \omega_r(g,t)_{\Phi,S_p}\right)$
4. $\omega_r(f,t)_{\Phi,S_p} \leq 2^{r-v}\omega_v(f\,,t)_{\Phi,S_p}$
5. $\omega_r(f\,,t)_{\Phi,S_p} \leq 2^r \|f\|_{\Phi,S_p}$
6. $\omega_r(f\,;t)_{\Phi,S_p} \leq \omega_r(f,\acute{t})_{\Phi,S_p}$, *for* $t \leq \acute{t}$
7. $\omega_r(f,\gamma t) \leq (1+\gamma)^k \omega_t(f,\delta)$

[7]studied the main properties of difference operator that is useful to our approximation in the following auxiliary lemma,

Lemma 2.3. [7] Let $r \in N$, then $\Delta^r f(x)$ satisfies

$$\Delta^r\big(f(x)g(x)\big) = \sum_{k=0}^{r} \binom{r}{k} \Delta^k f(x)\, \Delta^{r-k} g(x+k)$$

3. DOWNSAMPLING CONVOLUTIONAL NEURAL NETWORKS (DCNNS)

CNNs are a wide important topic that are studied by many researchers for different topics, such as print and face recognition and classification.

For any J ∈ ℕ, the depth of the network, a sequence $w^{(j)}, j = 1,\ldots,J$ is a filter mask that is within $\{0,1,\ldots,s^{(j)}\}$, where $s^{(j)} \in \mathbb{N}$ is the filter length with $s^{(j)}+1$ free parameters. So that the convolutional filter masks are

$$\left\{w^{(j)}\colon \mathrm{Z} \to \mathbb{R}\right\}_{j=1}^{J} \qquad (4)$$

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

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

Toeplitz Matrix ℑ, is the main difference with NNs with full-connected matrices T has rows more than columns which leads deep CNNs to use better functions than others. For any input $x \in R^d$, $h^{(0)}(x) = x$, define the jth iteration of h as follow

$$h^{(j)}(x) = \sigma\left(T^{(j)}h^{(j-1)}(x) - b^{j}\right), \qquad j = 1,2,\ldots,J$$

where $T^{(j)} = \left(w_{i-k}^{(j)}\right)$ is a $d_j \times d_{j-1}$ matrix, while b is a sequence of bias vectors b^j,

The activation function σ: ℝ → ℝ is the RelU activation function

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

Let h $\in L_{\Phi,S_p}(I^d)$ set. Also, we need to denote

$$\|w\|_{\Phi,s_p} = inf_{\lambda>0}\, \frac{1}{\lambda}\left(\lambda {w_k}^2 + 1\right)^{1/p} \qquad (5)$$

From(3) and (5) that

$$\left\|\Im^{(j)}h\right\|_{\Phi,S_p} \leq \|w\|_{\Phi,s_p}\|h\|_{\Phi,S_p} \qquad (6)$$

Zhou [9] defined an operator for Convolutional Neural Networks (DNN), he introduced a downsampled operation into CNNs to avoid big widths that happen with pooling

layers. The ℓ downsampled is introduced at layers $\mathcal{J} = \{\mathrm{J_k}\}_{\mathrm{k=1}}^{\ell}$ with $1 < J_1 \leq \cdots \leq J_\ell = J$. His concept of downsampling operators is induced from wavelets. The downsampling operation is defined below,

Definition 3.1. [9] Let m be a scaling parameter, the function $\mathfrak{O}_m: \mathbb{R}^D \longrightarrow \mathbb{R}^{[D/m]}$ is called downsampling operator, and it is given by

$$\mathfrak{O}_m(u) = (u_{im})_{i=1}^{[D/m]} \ , u \in \mathbb{R}^D \qquad (7)$$

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

In[10], the author added two completey connected layers $h^{(J+1)}$, $h^{(J+2)}$ with widths $D_1, D_2 > 0$, respectively, after the final CNN layer.

4. CNNs CONSTRUCTION AND APPROXIMATION

The first who offer a theoretical approach about approximation through neural networks is Cybenko through his global approximation theorem[11]**.** Cybenko's theorem provides that any function from $C[a,b]$ can be approximated by a neural network generated by f, written $N_n f$ with

Afterward, Cybenko's theorem followed by many most profound theories having the same target with different points of view. Although the fully connected layer is known to have the universal approximation property, it is not known if CNNs inherit this property, especially when the kernel size in the convolution layer is small. The first who demonstrated their universality of approximation is Fang[10], they consider an applied family of deep convolutional neural networks functions of the unit sphere $\mathbb{S}^{d-1}$of $\mathbb{R}^d$.

A CNN has a different architecture than an ordinary neuronal network. the form of convolution neural networks for estimate of function or treatment of data on $\mathbb{R}^d$ is given by

$$h^{(j)}(x) = \sigma\left(\mathfrak{J}^{(j)} h^{(j-1)}(x) - b^j\right),$$

where σ is the ReLU activation function, $\mathfrak{J}$ is the convolutional matrix, b is bias and $h^{(0)}(x) = x$

The existence of filter masks is studied by Zhou [12] in his following lemma, given convolution $w^{(\mathrm{J})} * \ldots * w^{(1)}$resulted from factorizing Was the following two lemmas state

Lemma4.1 For $s \geq 2$, $M \geq 0$, any sequence $W = (w_k)_{-\infty}^{\infty}$ in $\{0,\ldots,M\}$, ∃ a finite sequence of filter masks $\{w^{(\mathrm{j})}\}_{\mathrm{j=1}}^{\mathrm{J}}$ with nonzero elements in $0,\ldots,s$, where $J < \frac{M}{s-1} + 1$ that satisfies

$$W = w^{(J)} * \ldots * w^{2} * w^{1}$$

Lemma4.2 For $\mathrm{k} = 1,\ldots,\ell$, $\mathfrak{J}^{(\mathrm{J_k}, \mathrm{J_{k-1}+1})} = \mathfrak{J}^{(\mathrm{J_k})} \cdots \mathfrak{J}^{(\mathrm{J_{k-1}+2})} \mathfrak{J}^{(\mathrm{J_{k-1}+1})}$.

In [13]we studied the existence of the best approximation motion of DCNN that is generated by quasi – Orlicz function as in the following .

Lemma3.3. For any $g \epsilon L_{\Phi,S_p}(I^d)$, 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

$$\delta_{\mathrm{i}}(\mathrm{u}) = \sum_{\mathrm{i=1}}^{\mathrm{N}} \binom{\mathrm{N}}{\mathrm{i}} (-1)^{\mathrm{N-i}} \sigma(\mathrm{t_i} - \mathrm{u}),$$

and $|\mathrm{t_i} - \mathrm{t_{i-1}}| \sim \frac{1}{\mathrm{n}}$, then

$$\|L_t(g) - g\|_{\Phi,S_p}^p \leq \frac{c}{n} \omega_N\left(g, \frac{1}{n}\right)_{\Phi,S_p}$$

5. Main Results

Now, we build a better, more comfortable CNN than previous, with better degree of approximation. We benefit from the downsampling properties and add a finite number of layers as desired, or the user needs. In addition, the order of the best approximation that represents the degree of approximation matches the number of layers. That means that more additional hidden layers imply a better degree of approximation.

Theorem I : Let $2 \leq S \leq d, d \geq 3, r > 0, m, n, N \in \mathbb{N}$, $\boldsymbol{f} \in L_{\Phi,S_p}(\mathbb{S}^{d-1})$. Let $J \geq \lceil \frac{md-1}{S-1} \rceil$. Then there is a CNN of J layers with filters of length S and bias vectors satisfying $b_{s^{(j)}+1}^{(j)} = b_{s^{(j)}+2}^{(j)} = \ldots = b_{d_j-1}^{(j)}$ followed by downsampling and k-th layers satisfying

$$\left\|h^{(J+k)} - f\right\|_{\Phi,\mathrm{S_p}} \leq C\ \omega_N\left(f, \frac{1}{n}\right)_{\Phi,\mathrm{S_p}}$$

Proof:

Let m ∈ ℕ and $\boldsymbol{f} \in L_{\Phi,S_p}(\mathbb{S}^{d-1})$ and W in $\{0,\ldots,M\}$ given by $W_{(j-1)d+(d-i)} = (f_j)_i$, where $j \in \{1,\cdots,m\}$ and $i \in \{1,\cdots,d\}$. By Lemma 3.1., we have taken $w^{(j)}$. By Lemma 3.2, we have

$$\mathfrak{J}^{(\mathrm{J_k}, \mathrm{J_{k-1}+1})} = \mathfrak{J}^{(\mathrm{J_k})} \ \ldots\ \mathfrak{J}^{(\mathrm{J_{k-1}+2})} \mathfrak{J}^{(\mathrm{J_{k-1}+1})}$$

Now we construct bias vectors in the neural networks. We denote

$$\|w\|_{\Phi,S_p} = \inf_{\lambda>0} \frac{1}{\lambda}(\lambda {w_k}^2 + 1)^{1/p}$$

Take $b^{(1)} = -\left\| w^{(1)} \right\|_{\Phi,S_p} \mathbf{1}_{d_0}$ and

$$b^{(j)} = \left(\prod_{p=1}^{j-1} \left\|w^{(p)}\right\|_{\Phi,S_p}\right) \Im^{(j)} \mathbf{1}_{d_{j-1}} - \left(\prod_{p=1}^{j} \left\|w^{(p)}\right\|_{\Phi,S_p}\right) \mathbf{1}_{d_{j-1}+s},$$

for $j = 2,\cdots,J$. The bias vectors satisfy $b_{S+1}^{(j)} = \ldots = b_{dj-S}^{(j)}$. Since that $\|\boldsymbol{g}\|_{\Phi,S_p} \leq 1$,define

$$h^{(J)}(g(u)) = \mathcal{L}_{\mathrm{t}}(\mathrm{g(u)}) + B^{(j)}$$

where $B^{(j)} = \left\|w^{(j)}\right\|_{\Phi,S_p} \ldots \left\|w^{(1)}\right\|_{\Phi,S_p} \|x\|_{\Phi,S_p}$ for $j \geq 1$.

Applying the downsampling operator (7) leads to

$$\mathfrak{D}_{\mathrm{d}}\left(h^{(J)}(\mathrm{g(u)})\right) = \begin{bmatrix} \mathcal{L}_t\left(g_i(t_{1_i})\right)_{1d} \\ \vdots \\ \mathcal{L}_t\left(g_i(t_{m_i})\right)_{md} \\ 0 \\ \vdots \\ 0 \end{bmatrix} + B^{(J)} \mathbf{1}_{\left\lfloor \frac{(d+JS)}{d} \right\rfloor}$$

Denote $\hat{d} = \lfloor \frac{(d+JS)}{d} \rfloor$. Since $J \geq \lceil \frac{md-1}{S-1} \rceil$, we have

$$\frac{d+JS}{d} \geq 1 + \frac{md-1}{d} \frac{S}{S-1} > 1 + \frac{md-1}{d} \geq m.$$

Hence $\hat{d} \geq m$.

Now, define the main part of this construction, that is the last fully connected layer

$$h^{(J+k)}(f(u)) = h^{(J)}(f) \circ \Delta^k(fg),$$

Thus,

$$h^{(J+k)}(f(u)) = \mathcal{L}_t\left(g(u)\right) \circ \Delta^k(fg),$$

where

$$\mathcal{L}_t\left(g(u)\right) = \sum_{i=1}^{d} g(x_i)\delta_i(u),$$

and

$$\delta_k(u) = \sum_{i=1}^{k} (-1)^{k-i} \binom{k}{i} g_k(t_{ki})\sigma(u)$$

Thus, by Lemma 2.3., we have

$$\left\|h^{(J+k)}(\boldsymbol{f})(u) - f\right\|_{\Phi,\mathrm{S_p}}$$

$$= \left\| \sum_{i=1}^{\mathrm{d}} \mathcal{L}_t\left(g(u)\right) \circ \Delta^k(fg) - f \right\|_{\Phi,\mathrm{S_p}}$$

$$\leq C \left\| \sum\nolimits_{i=2}^{2N+2} [g(t_i) \circ \Delta^k(\sigma) \circ \Delta^k(fg) - f] \right\|_{\Phi,S_p}$$

$$\leq \left\| \sum\nolimits_{i=2}^{2N+2} \left[g(t_i) \circ \Delta^k(\sigma) \circ \sum_{r=0}^{k} \binom{k}{r} \Delta^r f(x) \Delta^{k-r} g(x+k), -f \right] \right\|_{\Phi,S_p}$$

$$\leq \frac{c}{n} \omega_N\left(f, \frac{1}{n}\right)_{\Phi,S_p}, N = 2k$$

To have complete look to the degree of approximation we study the essential degree of approximation with following theorem .

Theorem II (Inverse Theorem)

For any $f \epsilon L_{\Phi,S_p}(I^d)$, then $L_t(g)$ of Theorem I satisfies

$$\omega_N(f,\delta)_{\Phi,S_p} \leq \left\|h^{(J+k)} - f\right\|_{\Phi,S_p}$$

Proof

Let $b = \max_{1 \leq i \leq 2N+3}\{i; 2^i < n\}$,

$$f(t_{2N+3}) - f(t_1) = \left(f(t_{2N+3}) - f\left(t_{2^b}\right)\right) + \left(f\left(t_{2^b}\right) - g\left(t_{2^{b-1}}\right)\right) + \cdots + \left(f(t_2) - f(t_1)\right),$$

For any $m < 2N+3$, suppose that

$$\|f(t_{2N+3}) - f(t_m)\|_{\Phi,S_p} \leq cE_m(f),$$

where $E_m(f)_{\Phi,S_p} = inf \left\|f - h^{(J+k)}\right\|_{\Phi,S_p}$,we get by Lemma 2.1 and Theorem I, we easily get

$$\omega_N(f,\delta)_{\Phi,S_p} \leq c(p)\left[\omega_N\left(f - h^{(J+k)},\delta\right)_{\Phi,S_p} + \omega_N\left(h^{(J+k)},\delta\right)_{\Phi,S_p}\right]$$

$$\leq c(p)\left[c(N)\left\|f - h^{(J+k)}\right\|_{\Phi,S_p} + \omega_N\left(L_t(f,\delta)_{\Phi,S_p}\right)\right]$$

$$\leq c(p)\left[c(N)\omega_N(f,\delta)_{\Phi,S_p} + c(N)E_m(g)_{\Phi,S_p}\right]$$

$$\leq c(p,N)\left[\|f\|_{\Phi,S_p} + E_m(f)_{\Phi,S_p}\right]$$

6.Conclusion

We studied the approximation theory concerning with deep convolutional neural networks. We found that deep-convolution neural networks that are produced by downsampling operators are well approximated in quasi-Orlicz spaces on spheres.The degree of function approximation is estimated by modulus of smoothness, that is defined in this work in terms of quasi-Orlicz norm.

REFERENCES

[1] W. Orlicz, "Über eine gewisse Klasse von Räumen vom Typus B," *Bull. Int. Acad. Pol. Ser. A*, vol. 8, no. 9, pp. 207-220, 1932.

[2] H. Nakano, *Topology and linear topological spaces*. Maruzen Company, 1951.

[3] W. A. J. Luxemburg, "Banach function spaces ", .1955

[4] W. Luxemburg and A. Zaanen, "Conjugate spaces of Orlicz spaces," in *Indagationes Mathematicae (Proceedings)*, 1956, vol. 59: Elsevier, pp. 217-228 .

[5] H. Hudzik and L. Maligranda, "Amemiya norm equals Orlicz norm in general," *Indagationes Mathematicae*, vol. 11, no. 4, pp. 573-585, 2000.

[6] A. M. Aljanabi, Almurieb, Hawraa Abbas, "Orlicz Approximation by Convolutional Neural Networks," *Evolutionary Intelligence ,Springer Nature*, February 2023.

[7] F. Dai and Y. Xu, "Moduli of smoothness and approximation on the unit sphere and the unit ball," *Advances in Mathematics*, vol. 224, no. 4, pp. 1233-1310, 2010.

[8] Z. Ditzian and V. Totik, "Springer Series in Computational Mathematics," vol. 9, ed: Springer-Verlag New York, 1987.

[9] D.-X. Zhou" ,Theory of deep convolutional neural networks: Downsampling," *Neural Networks*, vol. 124, pp. 319-327, 2020.

[10] Z. Fang, H. Feng, S. Huang, and D.-X. Zhou, "Theory of deep convolutional neural networks II: Spherical analysis," *Neural Networks*, vol. 131, pp. 154-162, 2020.

[11] U. o. I. a. U.-C. C. f. S. Research, Development, and G. Cybenko, *Continuous valued neural networks with two hidden layers are sufficient*. 1988.

[12] D.-X. Zhou, "Universality of deep convolutional neural networks," *Applied and computational harmonic analysis*, vol. 48, no. 2, pp. 787-794, 2020.

[13] A. M. Aljanabi, Almurieb, Hawraa Abbas, "The Degree of Best Downsampled Convolutional Neural Network Approximation in terms of Orlicz Modulus of Smoothness," *journal of Mathematical Sciences*, 26May -2023.