Deep Scattering Network With Fractional Wavelet Transform

Jun Shi, Yanan Zhao, Wei Xiang, *Senior Member, IEEE*, Vishal Monga, *Senior Member, IEEE*, Xiaoping Liu, and Ran Tao, *Senior Member, IEEE*

***Abstract*—Deep convolutional neural networks (DCNNs) have recently emerged as a powerful tool to deliver breakthrough performances in various image analysis and processing applications. However, DCNNs lack a strong theoretical foundation and require massive amounts of training data. More recently, the deep scattering network (DSN), a variant of DCNNs, has been proposed to address these issues. DSNs inherit the hierarchical structure of DCNNs, but replace data-driven linear filters with predefined fixed multi-scale wavelet filters, which facilitate an in-depth understanding of DCNNs and also offer the state-of-the-art performance in image classification. Unfortunately, DSNs suffer from a major drawback: they are suitable for stationary image textures but not non-stationary image textures, since 2D wavelets are intrinsically linear translation-invariant filters in the Fourier transform domain. The objective of this paper is to overcome this drawback using the fractional wavelet transform (FRWT) which can be viewed as a bank of linear translation-variant multi-scale filters and thus may be well suited for non-stationary texture analysis. We first propose the fractional wavelet scattering transform (FRWST) based upon the FRWT. Then, we present a generalized structure for the DSN by cascading fractional wavelet convolutions and modulus operators. Basic properties of this generalized DSN are derived, followed by a fast implementation of the generalized DSN as well as their practical applications. The theoretical derivations are validated via computer simulations.**

***Index Terms*—Deep convolutional neural networks, wavelet scattering, fractional wavelet transform, translation-variant filtering.**

Manuscript received September 22, 2020; revised February 28, 2021, May 4, 2021, and June 15, 2021; accepted July 7, 2021. Date of publication July 26, 2021; date of current version August 27, 2021. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Bo Chen M.D. This work was supported in part by the National Natural Science Foundation of China under Grants 61871153, 61421001, and U1833203, in part by the Natural Science Foundation of Heilongjiang Province, China under Grant QC2018078, and in part by the Fundamental Research Funds for the Central Universities under Grant HIT.NSRIF.2019027. *(Corresponding authors: Jun Shi; Yanan Zhao; Ran Tao.)*

Jun Shi and Yanan Zhao are with the Communication Research Center, Harbin Institute of Technology, Harbin 150001, China (e-mail: junshi@hit.edu.cn; hiter_helen@163.com).

Wei Xiang is with the School of Engineering and Mathematical Sciences, La Trobe University, Melbourne, VIC 3086, Australia (e-mail: w.xiang@latrobe.edu.au).

Vishal Monga is with the Department of Electrical Engineering, The Pennsylvania State University, University Park, PA 16802 USA (e-mail: vmonga@engr.psu.edu).

Xiaoping Liu and Ran Tao are with the School of Information and Electronics, Beijing Institute of Technology and the Beijing Key Laboratory of Fractional Signals and Systems, Beijing 100081, China (e-mail: dr.xiaopingliu@gmail.com; rantao@bit.edu.cn).

Digital Object Identifier 10.1109/TSP.2021.3098936

I. Introduction

RECENTLY, the wide-ranging applications of deep convolutional neural networks (DCNNs) [1]–[8] have brought forth remarkable performance breakthroughs in image analysis and computer vision applications. DCNNs construct multilayer hierarchical feature representations by concatenating data-driven linear filters with nonlinear operators with a classifier attached to the last layer [9], [10]. Although DCNNs have achieved remarkable results, they lack a strong theoretical foundation and require massive amounts of training data to deliver the promised performance [11]–[13]. To address these issues, the deep scattering network (DSN), a variant of DCNNs, has been proposed [14]. DSNs inherit the hierarchical structure of DCNNs, but replace data-driven linear filters with predefined fixed multi-scale wavelet filters. Mathematically, DSNs are defined by the wavelet scattering transform (WST) which provides invariant, stable, and informative feature representations by cascading wavelet filters and modulus nonlinearities [14]. Such networks enable a theoretical understanding of DCNNs and also offer the state-of-the-art performance in various image analysis and processing tasks, including handwritten digit recognition [15], musical genre classification [16], audio classification [17], texture discrimination [18], art authentication [19], astronomy [20], chemical [21], biomedical science [22], [23], and time-frequency representations [24].

Unfortunately, DSNs suffer from a major drawback. That is, they work well for images with stationary textures but not those with non-stationary textures, since the conventional wavelets are intrinsically linear translation-invariant bandpass filters in the Fourier transform (FT) domain [26]. To illustrate this, we take the 2D Gabor and chirplet atoms [46] as an example, shown in Fig. 1(a) and Fig. 2(a), respectively, where the axis ranges of the plots in Fig. 1 are the same. The former is a 2D Gaussian function modulated by a 2D sinusoidal wave, and the latter is a 2D Gaussian function modulated by a 2D linear frequency modulation wave. It follows from [28]–[30] that the texture of the Gabor atom is stationary with its energy well concentrated in the FT domain; while for the chirplet atom, its texture is non-stationary, and its energy is spread over a large band in the FT domain. In fact, the real world exhibits an abundance of non-stationary textures, such as textures with large-scale irregular structures, spatially variant textures, and inhomogeneous textures. It was shown in [15] that high frequencies are sources of instabilities. Clearly, the chirplet atom has higher frequency instability than the Gabor atom. Like the

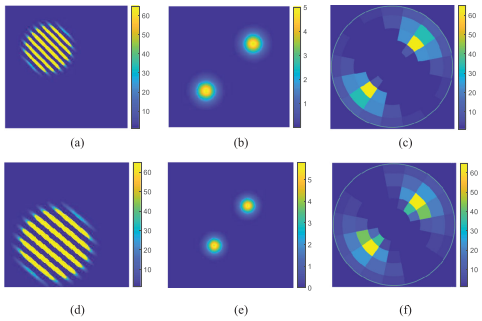

Fig. 1. 2D Gabor atom and its FT and WST. (a) Original 2D Gabor atom; (b) FT modulus of original 2D Gabor atom; (c) WST of original 2D Gabor atom; (d) Original 2D Gabor atom deformed by translation, rotation, and dilation; (e) FT modulus of the deformed 2D Gabor atom; and (f) WST of the deformed 2D Gabor atom.

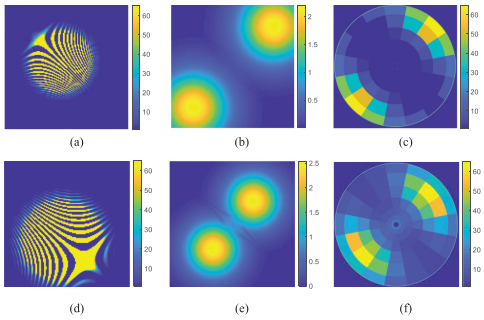

Fig. 2. 2D chirplet atom and its FT and WST. (a) Original 2D chirplet atom; (b) FT modulus of original 2D chirplet atom; (c) WST of original 2D chirplet atom; (d) Original 2D chirplet atom deformed by translation, rotation, and dilation; (e) FT modulus of the deformed 2D chirplet atom; and (f) WST of the deformed 2D chirplet atom.

FT modulus, the WST possesses good frequency localization, translation invariance, and norm preserved properties [14]. Besides, it is also resilient to deformation. As can be observed from Fig. 1, when the Gabor atom is deformed with translation, rotation, and slight dilation, most of its energy moves to other frequencies in the FT domain, but remains stable in the WST domain. However, as seen in Fig. 2 in which the axis ranges of the plots are the same, when the chirplet atom undergoes the same deformation, the majority of its energy moves to other frequencies in both the FT and WST domains. This implies that DSNs are inefficient in analyzing the chirplet atom whose energy is not well concentrated in the FT domain. The chirplet atom is a kind of chirp-like images [28], which are ubiquitous in nature and man-made systems, e.g., Newton's rings [31], ESPI images [32], and fringe patterns with quadratic phase [33]. Thus, it is desirable to overcome this shortcoming of the existing DSNs. Recently, Liu *et al.* [34] composed the DSN with the fractional wavelet transform (FRWT) [35], [37] which generalizes the conventional wavelet transform (WT) [38]–[40], [46] and can be interpreted as a bank of linear translation-variant bandpass filters [46], leading to the fractional wavelet scattering network (FrScatNet), and they showed improved performance with a rigorous experimental investigation. However, the main issue of the FrScatNet is its lack of mathematical theory. Indeed, it is an algorithmic approach and is difficult to model due to its non-linearity. Moreover, the basic properties of the FrScatNet, which correspond to those of the DSN, are not addressed in [34], and the fast implementation of the FrScatNet is also not discussed. Our objective of this paper is to develop a mathematical theory of the FrScatNet. First, we define the fractional wavelet scattering transform (FRWST) from the perspective of multi-resolution analysis based upon the FRWT, which has clear physical significance. Then, we propose a mathematical model for the FrScatNet, which includes the WT-based DSN as a special case. Basic properties of the FrScatNet are derived, and a fast algorithm for implementing the FrScatNet is also presented. The theoretical derivations are validated via computer simulations. Finally, a practical application of the FrScatNet in image classification is discussed.

The remainder of this paper is organized as follows. In Section II, the FRFT and FRWT are briefly reviewed. In Section III, the mathematical theory of the FrScatNet is developed. In Section IV, basic properties of the FrScatNet are derived, followed by a fast algorithm for its implementation presented in Section V. Numerical results and the applications of the FrScatNet are discussed in Section VI. Concluding remarks are drawn in Section VII.

II. Preliminaries

A. Notation and Definitions

The following notation is used throughout the paper. $\mathbb{R}$, $\mathbb{R}^d$, $\mathbb{R}^+$, $\mathbb{N}$, $\mathbb{C}$, $\mathbb{Z}$, $\mathbb{Z}^+$, $L^2(\mathbb{R}^d)$, $\|f(t)\|_\infty$, $\|f(t)\|_q$, Id, $C^2(\mathbb{R}^d)$, and $\nabla f(t)$ denote the set of real numbers, the set of d-dimensional real numbers, the set of positive real numbers, the set of natural numbers, the set of complex numbers, the set of integers, the set of positive integers, the set of square-integrable functions on $\mathbb{R}^d$, the essential supremum of $f(t)$ (i.e., $\|f(t)\|_\infty = \operatorname{ess\,sup}|f(t)|$), q-norm of $f(t)$ (i.e., $\|f\|_q \triangleq (\int_{\mathbb{R}^d}|f(t)|^q dt)^{1/q}$), the identity operator in $L^2(\mathbb{R}^d)$, the set of twice-continuously differentiable functions on $\mathbb{R}^d$, and the gradient of $f(t)$, respectively. For a vector field $\tau : \mathbb{R}^d \to \mathbb{R}^d$, denote by $\nabla\tau$ and $\nabla^2\tau$ its first and second order gradients, respectively, with associated norms $\|\tau\|_\infty \triangleq \sup_{t\in\mathbb{R}^d}|\tau(t)|$, $\|\nabla\tau\|_\infty \triangleq \sup_{t\in\mathbb{R}^d}|\nabla\tau(t)|$, and $\|\nabla^2\tau\|_\infty \triangleq \sup_{t\in\mathbb{R}^d}|\nabla^2\tau(t)|$.

B. Some Facts of the Fractional Fourier Transform

The fractional Fourier transform (FRFT) of an arbitrary function $f(t) \in L^2(\mathbb{R}^d)$ with angle $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_d) \in [-\pi, \pi]^d$ is defined as [46]

$$F_\alpha(u) = \mathcal{F}^\alpha\{f(t)\}(u) \triangleq \int_{\mathbb{R}^d} f(t)\mathcal{K}_\alpha(u,t)dt \tag{1}$$

where $\mathcal{K}_\alpha(u,t) = \prod_{i=1}^d \mathcal{K}_{\alpha_i}(u_i,t_i)$, and

$$\mathcal{K}_{\alpha_i}(u_i,t_i) = \begin{cases} A_{\alpha_i} e^{j\frac{u_i^2+t_i^2}{2}\cot\alpha_i - jt_iu_i\csc\alpha_i}, & \alpha_i \neq k\pi \\ \delta(t_i-u_i), & \alpha_i = 2k\pi \\ \delta(t_i+u_i), & \alpha_i = (2k-1)\pi \end{cases} \tag{2}$$

with $t = (t_1, t_2, \ldots, t_d) \in \mathbb{R}^d$, $A_{\alpha_i} = \sqrt{(1 - j\cot\alpha_i)/2\pi}$, and $k \in \mathbb{Z}$. Without loss of generality, we only consider the interval $\alpha \in [0, \pi/2]^d$, since (1) can easily be extended outside of this interval using the angle additivity property of the FRFT [46]. Conversely, the inverse FRFT is expressed as

$$f(t) = \mathcal{F}^{-\alpha}\left\{F_\alpha(u)\right\}(t) = \int_{\mathbb{R}^d} F_\alpha(u)\mathcal{K}^*_\alpha(u,t)du, \quad (3)$$

where superscript $*$ denotes the complex conjugate. When $\alpha_i = \pi/2$ for $i = 1, 2, \ldots, d$, the FRFT reduces to the FT.

The fractional convolution of the FRFT for two functions $f(t)$ and $g(t)$ in $L^2(\mathbb{R}^d)$ is defined as [41]

$$(f *_\alpha g)(t) \triangleq \int_{\mathbb{R}^d} f(\tau)g(t-\tau)e^{-j\frac{t^{\mathrm{T}}\Omega_{\mathrm{o}}t - \tau^{\mathrm{T}}\Omega_{\mathrm{o}}\tau}{2}}d\tau \quad (4)$$

where $\Omega_{\mathrm{o}} = \mathrm{diag}(\cot\alpha)$, and superscript T and $*_\alpha$ denote the transpose and the fractional convolution operator, respectively. Taking the FRFTs of both sides of (4) yields

$$(f *_\alpha g)(t) \overset{\mathcal{F}^\alpha}{\leftrightarrow} (2\pi)^{\frac{d}{2}} F_\alpha(u)G(\Omega_{\mathrm{s}}u) \quad (5)$$

where $\Omega_{\mathrm{s}} = \mathrm{diag}(\csc\alpha)$, then $F_\alpha(u)$ and $G(\Omega_{\mathrm{s}}u)$ represent the FRFT of $f(t)$ and the FT (with its argument scaled by Ω_{s}) of $g(t)$, respectively. According to [41], the fractional convolution is translation covariant, i.e., $((\mathsf{L}^\alpha_c f) *_\alpha g)(t) = (\mathsf{L}^\alpha_c(f *_\alpha g))(t)$, $c \in \mathbb{R}^d$, where L^α_c denotes the fractional translation operator [41] defined as

$$\mathsf{L}^\alpha_c f(t) \triangleq f(t-c)e^{-jc^{\mathrm{T}}\Omega_{\mathrm{o}}(t-\frac{c}{2})} \overset{\mathcal{F}^\alpha}{\leftrightarrow} F_\alpha(u)e^{-jc^{\mathrm{T}}\Omega_{\mathrm{s}}u}. \quad (6)$$

When $\alpha_i = \pi/2$ for $i = 1, 2, \ldots, d$, L^α_c reduces to $\mathsf{L}_c f(t) \triangleq f(t-c)$. Then, we have the following definition.

Definition 1: Let $\mathsf{L}^\alpha_c f(t)$ denote the fractional translation of $f(t) \in L^2(\mathbb{R}^d)$ by $c \in \mathbb{R}^d$. Further, Φ: $L^2(\mathbb{R}^d) \to \mathcal{H}$, a Hilbert space, is fractional translation-invariant if

$$\Phi(\mathsf{L}^\alpha_c f(t)) = \Phi(f(t)),\ \forall f(t) \in L^2(\mathbb{R}^d) \text{ and } c \in \mathbb{R}^d. \quad (7)$$

When $\alpha_i = \pi/2$ for $i = 1, 2, \ldots, d$, the fractional translation reduces to the conventional translation $\mathsf{L}_c f(t)$, and Φ is conventional translation-invariant.

Further, (4) can be rewritten in conventional convolution form as follows

$$(f *_\alpha g)(t) = e^{-j\frac{t^{\mathrm{T}}\Omega_{\mathrm{o}}t}{2}}\left(\left(f(t)e^{j\frac{t^{\mathrm{T}}\Omega_{\mathrm{o}}t}{2}}\right) * g(t)\right). \quad (8)$$

By applying (8), we can derive a generalization of Young's inequality [42] associated with the fractional convolution.

Lemma 1: Assume that $f(t), g(t) \in L^p(\mathbb{R}^d)$, and $\frac{1}{p} + \frac{1}{q} = \frac{1}{r} + 1$ with $1 \le p, q \le r \le \infty$. Then

$$\|(f *_\alpha g)(t)\|_r \le \|f(t)\|_q \|g(t)\|_p. \quad (9)$$

Proof: See Appendix A. ■

Also, the following lemma follows from (4).

Lemma 2: If a function $h(t) \in L^2(\mathbb{R}^d)$ and $h(t) \ge 0$, then for any function $f(t) \in L^2(\mathbb{R}^d)$, it follows that

$$\left||f(t)| *_\alpha h(t)\right| \ge \left|(f *_\alpha h_\gamma)(t)\right| \quad (10)$$

where $h_\gamma(t) \triangleq e^{j\gamma^{\mathrm{T}}t}h(t)$ and $\gamma \in \mathbb{R}^d$.

Proof: See Appendix B. ■

C. The Fractional Wavelet Transform

The fractional wavelet transform (FRWT) of an arbitrary function $f(t) \in L^2(\mathbb{R}^d)$ is defined as [35]

$$W^\alpha_f(\lambda, t) = \langle f(\tau), \psi_{\alpha,\lambda,t}(\tau)\rangle = \int_{\mathbb{R}^d} f(\tau)\psi^*_{\alpha,\lambda,t}(\tau)d\tau \quad (11)$$

with $\psi_{\alpha,\lambda,t}(\tau) = \frac{1}{\sqrt{|\lambda|^d}}\psi(\frac{\tau - t}{\lambda})e^{-j\frac{\tau^{\mathrm{T}}\Omega_{\mathrm{o}}\tau - t^{\mathrm{T}}\Omega_{\mathrm{o}}t}{2}}$, where $\lambda \in \mathbb{R}\backslash\{0\}$ and $t \in \mathbb{R}^d$ are the scaling and translation parameters, respectively. When $\alpha_i = \pi/2$ for $i = 1, 2, \ldots, d$, the FRWT reduces to the conventional WT. Alternatively, (11) can be further written as [35]

$$W^\alpha_f(\lambda, t) = \int_{\mathbb{R}^d} (2\pi|\lambda|)^{\frac{d}{2}} F_\alpha(u)\Psi^*(\lambda\Omega_{\mathrm{s}}u)\mathcal{K}^*_\alpha(u,t)du \quad (12)$$

where $\Psi(\Omega_{\mathrm{s}}u)$ denotes the FT (with its argument scaled by Ω_{s}) of $\psi(t)$. Since $\Psi(0) = 0$ [35], the FRWT can be interpreted as a bank of bandpass filters in the FRFT domain. It actually corresponds to a class of swept-frequency filters [45] which are linear translation-variant filters and suitable for analyzing non-stationary signals. In [36], the theory of a multi-resolution analysis of the FRWT was developed, and a construction of orthogonal wavelets for the FRWT was proposed, which proves the existence of unitary FRWTs.

For multi-dimensional signal analysis, i.e., $d > 1$, one often needs to distinguish signal variations along different directions by introducing directional wavelets in the FRWT. Denote by $\mathbb{G}$ a finite rotation group with elements r_k, $k \in \{1, 2, \ldots, K\}$. It was shown in [14] that if d is even, then G is a subgroup of the special orthogonal group $SO(d)$; if d is odd, then G is a subgroup of the orthogonal group $O(d)$. The directional wavelets are derived by rotating a single wavelet $\psi(t)$ along angle $r_k \in \mathbb{G}$ and scaling it by 2^p for $p \in \mathbb{Z}$. Briefly, they can be obtained by replacing λ of $\psi_{\alpha,\lambda,t}(\tau)$ with $\lambda_{p,k} = 2^p r_k$ which satisfies $|\lambda_{p,k}| = 2^p$, denoted by $\psi_{\alpha,\lambda_{p,k},t}(\tau)$. In particular, for the 2-D case, r_k is a 2-by-2 rotation matrix $r_k \triangleq \begin{pmatrix}\cos\theta_k & -\sin\theta_k \\ \sin\theta_k & \cos\theta_k\end{pmatrix}$ with $\theta_k = \frac{2\pi k}{K}$.

It follows from the multi-resolution analysis of the FRWT in [37] that the FRWT analyzes a signal $f(t) \in L^2(\mathbb{R}^d)$ in different fractional frequency bands with different resolutions by decomposing $f(t)$ into a coarse approximation and fine details. For a given fractional scaling function $\phi(t)$, the coarse approximation of $f(t)$ at scale 2^J is expressed as

$$\mathcal{A}_{\alpha,J}f(t) \triangleq \left\langle f(\tau), \phi_{\alpha,2^J,t}(\tau)\right\rangle \quad (13)$$

with $\phi_{\alpha,2^J,t}(\tau) = 2^{-d\frac{J}{2}}\phi(\frac{\tau - t}{2^J})e^{-j\frac{\tau^{\mathrm{T}}\Omega_{\mathrm{o}}\tau - t^{\mathrm{T}}\Omega_{\mathrm{o}}t}{2}}$, and the fine details of $f(taa)$ at scales $2^p \le 2^J$ are calculated as

$$W^\alpha_f(\lambda_{p,k}, t) = \left\langle f(\tau), \psi_{\alpha,\lambda_{p,k},t}(\tau)\right\rangle,\ 1 \le k \le K. \quad (14)$$

Indeed, the FRWT establishes a linear operator defined as

$$\begin{aligned}\mathcal{Q}^{\alpha}f(t) = &\left\{\left\langle f(\tau), \phi_{\alpha,2^J,t}(\tau)\right\rangle\right\} \\ &\cup \left\{\left\langle f(\tau), \psi_{\alpha,\lambda_{p,k},t}(\tau)\right\rangle\right\}_{p,k}\end{aligned} \tag{15}$$

where $p \le J$ and $1 \le k \le K$. To guarantee that $\mathcal{Q}^{\alpha}$ is a stable operator satisfying

$$A\|f(t)\|_2^2 \le \|\mathcal{Q}^{\alpha}f(t)\|_2^2 \le B\|f(t)\|_2^2, \forall f(t) \in L^2(\mathbb{R}^d) \tag{16}$$

where $0 < A \le B < +\infty$, we need to replace the normalization coefficients $2^{-d\frac{J}{2}}$ and $2^{-d\frac{p}{2}}$ in $\phi_{\alpha,2^J,t}(\tau)$ and $\psi_{\alpha,\lambda_{p,k},t}(\tau)$ with 2^{-dJ} and 2^{-dp} [43], respectively. Then, combining (16), (13), and (14) yields

$$\begin{aligned}A\|f(t)\|_2^2 \le &\int_{\mathbb{R}^d} \left|\left\langle f(\tau), \phi_{\alpha,2^J,t}(\tau)\right\rangle\right|^2 dt \\ &+ \sum_{p\le J}\sum_{k=1}^{K}\int_{\mathbb{R}^d} \left|\left\langle f(\tau), \psi_{\alpha,\lambda_{p,k},t}(\tau)\right\rangle\right|^2 dt \\ &\le B\|f(t)\|_2^2,\ \forall f(t) \in L^2(\mathbb{R}^d).\end{aligned} \tag{17}$$

Therefore, (16) essentially requires that the set of functions

$$\left\{\phi_{\alpha,2^J,t}(\tau)\right\}_{t\in\mathbb{R}^d} \cup \left\{\psi_{\alpha,\lambda_{p,k},t}(\tau)\right\}_{t\in\mathbb{R}^d, p\le J, 1\le k\le K} \tag{18}$$

forms a semi-discrete frame for $L^2(\mathbb{R}^d)$ with bounds $0 < A \le B < +\infty$, parameterized by discrete scale but continuous time. In particular, when $A = B = 1$, it becomes a semi-discrete Parseval frame [43], which makes $\mathcal{Q}^{\alpha}$ a unitary operator. Unless otherwise stated we assume in the following that $\mathcal{Q}^{\alpha}$ is unitary [36]. Then, by using Parseval's relation of the FRFT [46], setting $A = B = 1$ in (17) yields

$$\left|\Phi(2^J\Omega_s u)\right|^2 + \sum_{p\le J}\sum_{k=1}^{K}\left|\Psi(2^p r_k^{-1}\Omega_s u)\right|^2 = \frac{1}{(2\pi)^d}. \tag{19}$$

In fact, the signal $f(t)$ is completely characterized by its FRWT at scales 2^p with $-\infty \le p \le +\infty$ [36], i.e.,

$$\mathcal{Q}^{\alpha}f(t) = \left\{\left\langle f(\tau), \psi_{\alpha,\lambda_{p,k},t}(\tau)\right\rangle\right\}_{-\infty\le p\le+\infty, 1\le k\le K} \tag{20}$$

with $1 \le k \le K$. Then, the unitarity of $\mathcal{Q}^{\alpha}$ results in

$$\sum_{p=-\infty}^{+\infty}\sum_{k=1}^{K}\left|\Psi(2^p r_k^{-1}\Omega_s u)\right|^2 = \frac{1}{(2\pi)^d} \tag{21}$$

Combining (21) and (19) gives rise to

$$\left|\Phi(2^J\Omega_s u)\right|^2 = \sum_{p>J}\sum_{k=1}^{K}\left|\Psi(2^p r_k^{-1}\Omega_s u)\right|^2. \tag{22}$$

III. A Mathematical Theory of Fractional Wavelet Scattering Networks

A. The Fractional Wavelet Scattering Transform

It is noted in (15) that the operator $\mathcal{Q}^{\alpha}$ computes a coarse approximation $\mathcal{A}_{\alpha,J}f(t)$ of signal $f(t)$ at scale 2^J and its fine details $W_f^{\alpha}(\lambda_{p,k}, t)$ at scales $2^p \le 2^J$. The former associated with the dilated scaling function $\phi_{\alpha,2^J,t}(\tau)$ captures low fractional frequency information. The latter relating to the directional fractional wavelets $\psi_{\alpha,\lambda_{p,k},t}(\tau)$ captures high fractional frequency information. We will show in Section IV-D that the coarse approximation at scale 2^J is approximately invariant to translations smaller than 2^J, but the fine details at scales $2^p \le 2^J$ are covariant to translations because of fractional wavelet convolutions. With an appropriate scale 2^J, the coarse approximation can be used directly for signal classification, but the fine details at scales $2^p \le 2^J$ can not, although they are necessary to distinguish signals. Following [15], the translation invariance of the coarse approximation comes from an inner product by a dilated scaling function $\phi_{\alpha,2^J,t}(\tau)$. A natural question arises: Can we make the fine details translation-invariant through the inner product of $W_f^{\alpha}(\lambda_{p,k},\tau)$ and $\phi_{\alpha,2^J,t}(\tau)$? Unfortunately, the answer is negative, since it would yield a trivial invariant, i.e., $\langle W_f^{\alpha}(\lambda_{p,k},\tau), \phi_{\alpha,2^J,t}(\tau)\rangle = 0$ due to the fact that $\int_{\mathbb{R}^d}\psi(t)dt = 0$. A simple solution [14] is to first take non-linear point-wise complex modulus of $W_f^{\alpha}(\lambda_{p,k},\tau)$ and then perform an inner product with $\phi_{\alpha,2^J,t}(\tau)$, i.e., $\langle|W_f^{\alpha}(\lambda_{p,k},\tau)|, \phi_{\alpha,2^J,t}(\tau)\rangle$ which produces a non-trivial translation-invariant but crude signal presentation. The loss of information is not caused by the complex modulus operation. Indeed, following [44], one can recover any signal $f(t) \in L^2(\mathbb{R}^d)$ from its FRWT modulus $\{|W_f^{\alpha}(\lambda_{p,k},\tau)|\}_p$ up to a multiplicative constant, since the FRWT can be decomposed around the conventional WT embedded in its definition [35]. Rather, the information loss stems from $\langle|W_f^{\alpha}(\lambda_{p,k},\tau)|, \phi_{\alpha,2^J,t}(\tau)\rangle$ which discards all non-zero fractional frequencies. The lost information can be recovered by a new FRWT of $|W_f^{\alpha}(\lambda_{p,k},\tau)|$, i.e., $\langle|W_f^{\alpha}(\lambda_{p,k},\tau)|, \psi_{\alpha,\lambda'_{p,k},t}(\tau)\rangle_{p\le J,1\le k\le K}$ which yields new invariants by iterating the same procedure. For a given scale 2^J, the set of all scales and orientations $2^p r_k$ ($p \le J, 1 \le k \le K$) of the directional fractional wavelets in (14) is denoted as $\Lambda \triangleq \{\lambda_{p,k} = 2^p r_k \mid 2^p \le 2^J, 1 \le k \le K\}$. Then, for an integer $m \in \mathbb{N}^+$, define $l^{(m)} = \left(l_1^{(m)}, l_2^{(m)}, \ldots, l_m^{(m)}\right) \in \Lambda^m$ as an ordered vector of the scales and orientations of the directional fractional wavelets, where $l_n^{(m)} = 2^{p_n} r_k$ with $p_n \le J$ and $1 \le n \le m$. The length of vector $l^{(m)}$ is $|l^{(m)}| = m$. When $m = 0$, denote by Λ^0 and $l^{(0)}$ the empty set and an empty vector of length zero, respectively. Indeed, $l^{(m)}$ corresponds to a path of length m in the FrScatNet, as will be constructed in Section III-B. For the path $l^{(m)} \in \Lambda^m$ defined above and a scale and orientation $l_{m+1}^{(m+1)} \in \Lambda$, we define path $l^{(m+1)}$ as $l^{(m+1)} = \left(l^{(m)}, l_{m+1}^{(m+1)}\right) \in \Lambda^{m+1}$. Note that all ordered vectors $l^{(m)}$ of length m correspond to $|\Lambda|^m$ paths in the network, and let $L^{(m)}$ be the set of these $|\Lambda|^m$ paths, i.e.,

$$L^{(m)} = \left\{l^{(m)} \mid l^{(m)} \in \Lambda^m, m \in \mathbb{N}\right\} \tag{23}$$

with $L^{(0)} = \emptyset$. Consequently, the collection of all paths of an arbitrary length $m \in \mathbb{N}$ can be expressed as

$$\mathbb{L} = \bigcup_{m=0}^{+\infty} L^{(m)} \text{ with } \mathbb{L} = \emptyset \text{ when } m = 0. \tag{24}$$

For each scale and orientation of $l^{(m)} \in \Lambda^m$, we define the fractional scattering propagation operator U^α as the complex modulus of the FRWT of signal $f(t)$ in (14), i.e.,

$$U^\alpha[l_n^{(m)}]f(t) \triangleq \left|\left\langle f(\tau), \psi_{\alpha,l_n^{(m)},t}(\tau)\right\rangle\right|, \forall 1 \le n \le m. \quad (25)$$

Further, (25) can be extended to multiple scales and orientations of $l^{(m)} \in \Lambda^m$, i.e.,

$$\begin{aligned} U^\alpha[l^{(m)}]f(t) &\triangleq U^\alpha[l_m^{(m)}]\cdots U^\alpha[l_2^{(m)}]U^\alpha[l_1^{(m)}]f(t) \\ &= \left|\left\langle \ldots \left|\left\langle \left|\langle f(\tau), \psi_{\alpha,l_1^{(m)},t}(\tau)\rangle\right|, \right.\right.\right. \\ &\qquad \left.\left.\left. \psi_{\alpha,l_2^{(m)},t}(\tau)\right\rangle\right|\ldots, \psi_{\alpha,l_m^{(m)},t}(\tau)\right\rangle\Bigg| \end{aligned} \quad (26)$$

with $U^\alpha[\emptyset] = Id$, and $U^\alpha[\emptyset]f(t) = f(t)$. Here, we refer to $U^\alpha[l^{(m)}]f(t)$ as the propagated signal with respect to path $l^{(m)}$, which can be viewed as cascades of the FRWT of $f(t)$ and the modulus. Moreover, the family of the propagated signals with respect to set $L^{(m)}$ is given by $U^\alpha[L^{(m)}]f(t) \triangleq \bigcup_{l^{(m)}\in L^{(m)}}\{U^\alpha[l^{(m)}]f(t)\}$. Similarly, it follows from (24) that $U^\alpha[\mathbb{L}]f(t) \triangleq \bigcup_{m=0}^{+\infty} U^\alpha[L^{(m)}]f(t)$. For an arbitrary signal $f(t) \in L^2(\mathbb{R}^d)$, the fractional wavelet scattering transform (FRWST) along path $l^{(m)}$ is defined as

$$S^\alpha[l^{(m)}]f(t) \triangleq \left\langle U^\alpha[l^{(m)}]f(\tau), \phi_{\alpha,2^J,t}(\tau)\right\rangle \quad (27)$$

with $S^\alpha[\emptyset]f(t) = \langle f(\tau), \phi_{\alpha,2^J,t}(\tau)\rangle$, where $S^\alpha[l^{(m)}]f(t)$ is refereed to as the fractional scattering coefficient of order m. In particular, when $\alpha_i = \pi/2$ for $i = 1, 2, \ldots, d$, the proposed FRWST reduces to the conventional wavelet scattering transform [14]. The FRWST has many similarities with the WST, which is locally invariant to translation, stable under small deformations, and preserves information at all fractional frequencies, as will be discussed in Section IV.

B. Fractional Wavelet Scattering Networks

The FRWST essentially establishes a class of deep convolutional networks, i.e., FrScatNets [34]. Compared with the conventional DSN, the FrScatNet has a free parameter $\alpha \in [0, \pi/2]^d$ to be specified by users according to the specific application. To derive the network structure mathematically, we rewrite (13) in terms of fractional convolution in (4) as

$$\mathcal{A}_{\alpha,J}f(t) = f(t) *_\alpha \phi_{\lambda_J}(t),\ \lambda_J = 2^J \quad (28)$$

with $\phi_{\lambda_J}(t) \triangleq \frac{1}{|\lambda_J|^d}\phi^*(-\frac{t}{\lambda_J})$. Correspondingly, (14) can be rewritten as

$$W_f^\alpha(\lambda_{p,k}, t) = f(t) *_\alpha \psi_{\lambda_{p,k}}(t),\ 1 \le k \le K \quad (29)$$

with $\psi_{\lambda_{p,k}}(t) \triangleq \frac{1}{|\lambda_{p,k}|^d}\psi^*(-\frac{t}{\lambda_{p,k}})$. Then, (15) leads to

$$\mathcal{Q}^\alpha f(t) = \{f(t) *_\alpha \phi_{\lambda_J}(t)\} \cup \left\{f(t) *_\alpha \psi_{\lambda_{p,k}}(t)\right\}_{p,k} \quad (30)$$

with $p \le J$ and $1 \le k \le K$. It follows from (28) and (5) that the coarse approximation of $f(t)$ at a scale 2^J is derived by filtering $f(t)$ through a low-pass filter $\phi_{\lambda_J}(t)$ (i.e., dilated scaling function $\phi(t)$) in the FRFT domain. Also, as can be observed from (29) and (5), the fine details of $f(t)$ at scales $2^p \le 2^J$ are obtained by filtering $f(t)$ through multi-scale fractional wavelet band-pass filters $\psi_{\lambda_{p,k}}(t)$ in the FRFT domain. Then, the proposed fractional scattering propagation operator U^α (26) along an ordered path $l^{(m)}$ can be given by

$$\begin{aligned} &U^\alpha[l^{(m)}]f(t) \\ &= \left|\ldots\left|\left|f(t) *_\alpha \psi_{l_1^{(m)}}(t)\right| *_\alpha \psi_{l_2^{(m)}}(t)\right|\ldots *_\alpha \psi_{l_m^{(m)}}(t)\right| \end{aligned} \quad (31)$$

with $U^\alpha[\emptyset] = Id$, and $U^\alpha[\emptyset]f(t) = f(t)$. By (31), the FRWST of $f(t) \in L^2(\mathbb{R}^d)$ with respect to path $l^{(m)}$ is derived as

$$S^\alpha[l^{(m)}]f(t) = U^\alpha[l^{(m)}]f(t) *_\alpha \phi_{\lambda_J}(t) \quad (32)$$

with $S^\alpha[\emptyset]f(t) = f(t) *_\alpha \phi_{\lambda_J}(t)$.

Based on the results obtained above, we now introduce the structure of the FrScatNet, which can be viewed as an iterative process over a one-step fractional scattering propagation operator. Similar to its DCNN counterpart [9], the FrScatNet is built upon a building block comprised of a fractional filter followed by non-linear complex modulus. The FrScatNet can be iteratively constructed as follows.

The first layer of the FrScatNet computes all fractional scattering coefficients of order $m = 0$, i.e., the FRWST with respect to $L^{(0)}$, given by $S^\alpha[L^{(0)}]f(t) = f(t) *_\alpha \phi_{\lambda_J}(t)$, and gathers all propagated signals, i.e., the complex modulus of the fractional wavelet coefficients in (29), expressed as

$$U^\alpha[L^{(1)}]f(t) = \left\{\left|f(t) *_\alpha \psi_{l_1^{(1)}}(t)\right|\right\}_{l_1^{(1)}\in\Lambda}. \quad (33)$$

The m-th layer of the FrScatNet stores all propagated signals

$$U^\alpha[l^{(m)}]f(t) = U^\alpha[l_m^{(m)}]U^\alpha[l^{(m-1)}]f(t), \forall l^{(m)} \in \Lambda^m, \quad (34)$$

and outputs all fractional scattering coefficients of order $m-1$, written as $S^\alpha[L^{(m-1)}]f(t) = U^\alpha[L^{(m-1)}]f(t) *_\alpha \phi_{\lambda_J}(t)$. Indeed, each layer defines an operator $\mathcal{G}^\alpha$, i.e.,

$$\mathcal{G}^\alpha f(t) \triangleq \left\{f *_\alpha \phi_{\lambda_J}, \left|f *_\alpha \psi_{\lambda_{p,k}}\right|\right\}_{p\le J,\, 1\le k\le K}. \quad (35)$$

Iterating over $\mathcal{G}^\alpha$ defines the whole FrScatNet structure illustrated in Fig. 3. More specifically, for any path $l^{(m)} \in \Lambda^m$ and $l_{m+1}^{(m+1)} \in \Lambda$, we denote $l^{(m+1)} = (l^{(m)}, l_{m+1}^{(m+1)})$. Then, for any $f(t) \in L^2(\mathbb{R}^d)$, we have

$$\mathcal{G}^\alpha U^\alpha[l^{(m)}]f(t) = \left\{S^\alpha[l^{(m)}]f(t), U^\alpha[l^{(m+1)}]f(t)\right\} \quad (36)$$

which states that applying $\mathcal{G}^\alpha$ to all propagated signals $\{U^\alpha[l^{(m)}]f(t)\}_{l^{(m)}\in L^{(m)}}$ in the m-th layer outputs all fractional scattering coefficients $\{S^\alpha[l^{(m)}]f(t)\}_{l^{(m)}\in L^{(m)}}$ of order m and computes all propagated signals $\{U^\alpha[l^{(m+1)}]f(t)\}_{l^{(m+1)}\in L^{(m+1)}}$ in the next layer $m+1$. Based on these facts, we have the following lemma.

Lemma 3: For any $f(t) \in L^2(\mathbb{R}^d)$, the FRWST and the fractional scattering propagation operator U^α with respect to path $l^{(m)}$ satisfy

$$\left\|U^\alpha[l^{(m)}]f\right\|_2^2 = \left\|S^\alpha[l^{(m)}]f\right\|_2^2 + \left\|U^\alpha[l^{(m+1)}]f\right\|_2^2. \quad (37)$$

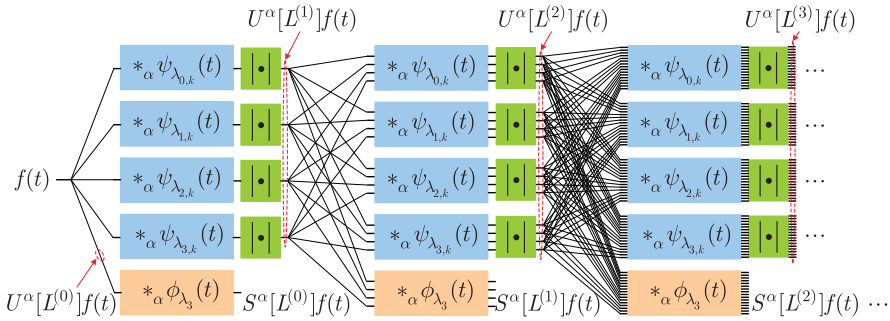

Fig. 3. Structure of the FrScatNet with $J = 3$ and $m = 3$.

Proof: Since $\mathcal{Q}^\alpha$ in (15) is unitary, it follows (30) that

$$\|f\|_2^2 = \left\|S^\alpha[l^{(0)}]f\right\|_2^2 + \sum_{l^{(1)}\in L^{(1)}} \|U^\alpha[l^{(1)}]f\|_2^2 \tag{38}$$

for any $f(t) \in L^2(\mathbb{R}^d)$. This result implies that operator $\mathcal{G}^\alpha$ in (35) preserves norm for any $f(t) \in L^2(\mathbb{R}^d)$. Further, replacing $f(t)$ in (38) with $U^\alpha[l^{(m)}]f(t)$ yields (37). ■

Note in Fig. 3, the FrScatNet seems like a fully-connected network, but is actually a partially-connected one, as will be revealed in Section V.

IV. Basic Properties of Fractional Wavelet Scattering Networks

In this section, we will discuss some basic properties of the FrScatNet, which can be used to guide the optimization of the network architecture to retain important information while avoiding redundant computations.

A. Non-Expansivity

Theorem 1: Let $\psi(t)$ be a fractional mother wavelet corresponding to a given fractional scaling function $\phi(t)$. If $\psi(t)$ and $\phi(t)$ satisfy (19), then the FRWST is non-expansive. That is, for any $f(t), h(t) \in L^2(\mathbb{R}^d)$, we have

$$\|S^\alpha[\mathbb{L}]f - S^\alpha[\mathbb{L}]h\|_2^2 \le \|f - h\|_2^2. \tag{39}$$

Proof: For any $a, b \in \mathbb{C}$, it can be shown that $||a| - |b|| \le |a - b|$. Based on this, applying (35) leads to

$$\begin{aligned}\|\mathcal{G}^\alpha f - \mathcal{G}^\alpha h\|_2^2 &\le \left\|f *_\alpha \phi_{\lambda_J} - h *_\alpha \phi_{\lambda_J}\right\|_2^2 \\ &\quad + \sum_{p\le J}\sum_{k=1}^{K}\left\|f *_\alpha \psi_{\lambda_{p,k}} - h *_\alpha \psi_{\lambda_{p,k}}\right\|_2^2 \\ &= \|\mathcal{Q}^\alpha(f-h)\|_2^2 = \|f-h\|_2^2\end{aligned} \tag{40}$$

which implies that $\mathcal{G}^\alpha$ is non-expansive. Similarly, we have

$$\begin{aligned}&\|\mathcal{G}^\alpha U^\alpha[l^{(m)}]f - \mathcal{G}^\alpha U^\alpha[l^{(m)}]h\|_2^2 \\ &\qquad \le \left\|U^\alpha[l^{(m)}]f - U^\alpha[l^{(m)}]h\right\|_2^2\end{aligned} \tag{41}$$

so that

$$\begin{aligned}\left\|S^\alpha[l^{(m)}]f - S^\alpha[l^{(m)}]h\right\|_2^2 &\le \left\|U^\alpha[l^{(m)}]f - U^\alpha[l^{(m)}]h\right\|_2^2 \\ &\quad - \left\|U^\alpha[l^{(m+1)}]f - U^\alpha[l^{(m+1)}]h\right\|_2^2\end{aligned} \tag{42}$$

which alongside (23) leads to

$$\begin{aligned}&\sum_{m=0}^{M}\left\|S^\alpha[L^{(m)}]f - S^\alpha[L^{(m)}]h\right\|_2^2 \\ &\qquad \le \sum_{m=0}^{M}\left\|U^\alpha[L^{(m)}]f - U^\alpha[L^{(m)}]h\right\|_2^2 \\ &\qquad\quad - \sum_{m=0}^{M}\left\|U^\alpha[L^{(m+1)}]f - U^\alpha[L^{(m+1)}]h\right\|_2^2.\end{aligned} \tag{43}$$

As $M \to +\infty$, it follows from (43) and (24) that

$$\begin{aligned}&\left\|S^\alpha[\mathbb{L}]f - S^\alpha[\mathbb{L}]h\right\|_2^2 \\ &\le \|f-h\|_2^2 - \lim_{M\to\infty}\left\|U^\alpha[L^{(M+1)}]f - U^\alpha[L^{(M+1)}]h\right\|_2^2 \\ &\le \|f-h\|_2^2\end{aligned} \tag{44}$$

which yields (39). This completes the proof of Theorem 1. ■

Remark 1: Denote by $f(t)$ the signal of interest, and let $h(t) = f(t) + n(t)$, where $n(t)$ is a additive noise. By Theorem 1, we have $\left\|S^\alpha[\mathbb{L}](f+n)(t) - S^\alpha[\mathbb{L}]f(t)\right\|_2^2 \le \|n(t)\|_2^2$. This implies that the FrScatNet is robust against additive noise.

B. Energy Conservation

Definition 2: Let $\psi(t)$ be a fractional wavelet corresponding to a given fractional scaling function $\phi(t)$ with its scaled FT $\Phi(\Omega_s u)$. The fractional wavelet $\psi(t)$ is admissible if there exists $\gamma \in \mathbb{R}^d$ and a non-negative function $\theta(t) \in L^2(\mathbb{R}^d)$ with its scaled FT $\Theta(\Omega_s u)$ satisfying $|\Theta(\Omega_s u)| \le |\Phi(\Omega_s u)|$ and $\Theta(0) = (2\pi)^{-d/2}$, such that

$$\begin{aligned}c = \inf_{1\le|u\csc\alpha|\le 2}&\sum_{p=-\infty}^{+\infty}\sum_{k=1}^{K}\left|\Psi\left(2^p r_k^{-1}\Omega_s u\right)\right|^2 \\ &\times (2\pi)^{2d}\Big(\left|\Theta\left(2^p r_k^{-1}\Omega_s u - \gamma\right)\right|^2 - \sum_{b=1}^{+\infty} b \\ &\times \left(1 - \left|\Theta(2^{p-b} r_k^{-1}\Omega_s u - 2^{-b}\gamma)\right|^2\right)\Big) > 0.\end{aligned} \tag{45}$$

For an admissible fractional wavelet, one can prove that the FRWST conserves the energy of signals.

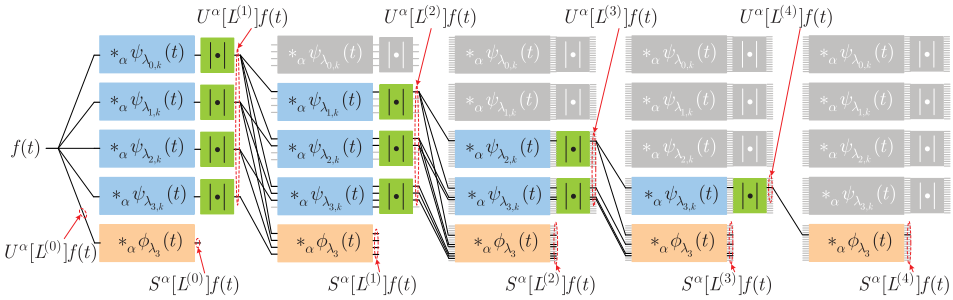

Fig. 4. Fractional frequency-decreasing FrScatNet with $J = 3$ and $m = J + 1$.

Theorem 2: If the fractional wavelet $\psi(t)$ is admissible, then for any $f(t) \in L^2(\mathbb{R}^d)$, it follows that

$$\lim_{m\to+\infty} \left\| U^\alpha[L^{(m)}]f \right\|_2^2 = 0, \tag{46}$$

and

$$\|S^\alpha[\mathbb{L}]f\|_2^2 = \|f\|_2^2. \tag{47}$$

Proof: See Appendix C. ∎

Remark 2: Theorem 2 states that the energy of $f(t)$ ultimately reaches the minimum fractional frequency 2^J and is trapped by the low-pass filter $\phi_{\lambda_J}(t)$ with outputs $S^\alpha[\mathbb{L}]f(t)$. Thus, the energy of $U^\alpha[L^{(m)}]f(t)$ diminishes to 0 as path length m increases, and $\|S^\alpha[\mathbb{L}]f\|_2^2 = \|f\|_2^2$. Indeed, the energy of $S^\alpha[\mathbb{L}]f(t)$ is concentrated on low-order paths, and the energy of $U^\alpha[L^{(m)}]f(t)$ is only propagated along a subset of fractional frequency-decreasing paths $l^{(m)} \in \Lambda^m$ satisfying $|l_n^{(m)}| < |l_{n+1}^{(m)}|$ for $1 \le n \le m-1$, while the energy along other paths is negligible, as will be discussed in Section V.

C. Robustness to Fractional Deformation

In this subsection, we will derive a bound for signal $f(t) \in L^2(\mathbb{R}^d)$ on the sensitivity of the FRWST to fractional deformations of the form $\mathsf{L}_\tau^\alpha f(t) \triangleq f(t-\tau(t))e^{-j\tau^{\mathrm{T}}(t)\Omega_0(t-\frac{\tau(t)}{2})}$. Note that when $\tau(t) = c$ with $c \in \mathbb{R}^d$, the fractional deformation L_τ^α reduces to the fractional translation in (6). In particular, if $\alpha_i = \pi/2$ for $i = 1, 2, \ldots, d$, it becomes the conventional deformation, denoted by $\mathsf{L}_\tau f(t) \triangleq f(t-\tau(t))$. Further, the relationship between the fractional and conventional deformations is given by $\mathsf{L}_\tau^\alpha f(t) = e^{-jt^{\mathrm{T}}\Omega_0 t}\mathsf{L}_\tau(f(t)e^{jt^{\mathrm{T}}\Omega_0 t})$. It follows from [45] that the structure of the fractional deformation is the same as that of the swept-frequency system which is widely used in optics [46], [47]. Particularly, if the swept-frequency system is just a delay system (i.e., the impulse response $g(t)$ of the translation-invariant filter in Fig. 4 of [45] is given by $g(t) = \delta(t-c)$, where constant c denotes a delay.), then its output is the fractional translation $\mathsf{L}_c^\alpha f(t)$ of the input $f(t)$. If the delay c changes with time, i.e., $c = \tau(t)$, then the output yields the fractional deformation $\mathsf{L}_\tau^\alpha f(t)$. This implies that the fractional deformation may be useful for analyzing images captured by optical systems. In addition, the structure of the fractional deformation is also similar to the de-chirp process in ISAR imaging [48] and the modulation-demodulation system in IoT communication networks [49]. Therefore, the fractional deformation may also be useful for analyzing images/signals in ISAR imaging and IoT communication networks. Based on these facts, we have the following theorem.

Theorem 3: There exists a positive constant $C_1 = \max\{2\|\nabla\phi\|_1, \|\nabla\psi\|_1\}$ such that for any $f(t) \in L^2(\mathbb{R}^d)$ with $\|U^\alpha[\mathbb{L}]f\|_2 < +\infty$ and $\tau(t) \in C^2(\mathbb{R}^d)$ with $\|\nabla\tau\|_\infty \le \frac{1}{2d}$, the deformation error satisfies the following stability bound

$$\left\|S^\alpha[\mathbb{L}]\mathsf{L}_\tau^\alpha f - S^\alpha[\mathbb{L}]f\right\|_2 \le \Xi_\alpha(\tau)\left\|U^\alpha[\mathbb{L}]f\right\|_2 \tag{48}$$

with

$$\begin{aligned}\Xi_\alpha(\tau) \triangleq\ & C_1\left(2^{-J}\|\tau\|_\infty + J\|\nabla\tau\|_\infty + \|\nabla^2\tau\|_\infty\right) \\ & + \sqrt{2}|\cos\alpha|^d\left(1+\tfrac{1}{2d}\right)^d\|\nabla\hat{\tau}(\Omega_{\mathrm{s}}u)\|_2\end{aligned} \tag{49}$$

where $\hat{\tau}(\Omega_{\mathrm{s}}u)$ denote the FT (with its argument scaled by Ω_{s}) of $\tau(t)$.

Proof: See Appendix D. ∎

Remark 3: In practice, robustness to deformations implies that the feature corresponding to each image under analysis are close together. If the deformation is large, it may convert an image to an entirely different one. In general, we only consider robustness to small deformations. When the signal $f(t)$ undergoes the conventional deformation $\mathsf{L}_\tau f(t)$, the stability bound in (48) can be derived as $C_1(2^{-J}\|\tau\|_\infty + J\|\nabla\tau\|_\infty + \|\nabla^2\tau\|_\infty)\left\|U^\alpha[\mathbb{L}]f\right\|_2$. It follows from [14] that the stability bound of the conventional DSN is $C_1(2^{-J}\|\tau\|_\infty + J\|\nabla\tau\|_\infty + \|\nabla^2\tau\|_\infty)\left\|U[\mathbb{L}]f\right\|_2$, where U denotes the conventional scattering propagation operator. Thus, by choosing an appropriate α such that $\left\|U^\alpha[\mathbb{L}]f\right\|_2 < \left\|U[\mathbb{L}]f\right\|_2$, the FrScatNet may yield a smaller Lipschitz constant than that of its DSN counterpart, indicating that it has better stability than the DSN.

D. Fractional Translation Invariance

Theorem 4: For admissible fractional wavelets satisfying (45) and any $f(t) \in L^2(\mathbb{R}^d)$, it follows that

$$\left\|S^\alpha[\mathbb{L}]\mathsf{L}_c^\alpha f - S^\alpha[\mathbb{L}]f\right\|_2 \le \Xi_\alpha(c)\left\|U^\alpha[\mathbb{L}]f\right\|_2 \tag{50}$$

with $\Xi_\alpha(c) \triangleq C_2\{2^{-J}|c| + |\cos\alpha|^d|c|\}$, where constant $C_2 = \max\{\|\nabla\phi\|_1, 2\}$.

Proof: It follows from the left side of (50) that

$$\begin{aligned}\|S^\alpha[\mathbb{L}]\mathsf{L}_c^\alpha f - S^\alpha[\mathbb{L}]f\|_2 \le\ & \left\|\mathsf{L}_c^\alpha S^\alpha[\mathbb{L}]f - S^\alpha[\mathbb{L}]f\right\|_2 \\ & + \left\|S^\alpha[\mathbb{L}]\mathsf{L}_c^\alpha f - \mathsf{L}_c^\alpha S^\alpha[\mathbb{L}]f\right\|_2.\end{aligned} \tag{51}$$

TABLE I
PERCENTAGE OF PRESERVED ENERGY $\|S^{\alpha}[L^{(m)}]f\|_2^2/\|f\|_2^2$ ASSOCIATED WITH THE FRWST FOR DIFFERENT VALUES OF m AND J

J	$m=0$	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$	$m=6$	$m=7$	$m=8$	$m\leq 3$
0	96.88	3.11	0	0	0	0	0	0	0	99.99
1	72.94	26.87	0.04	0	0	0	0	0	0	99.85
2	42.05	55.55	2.18	0.0015	0	0	0	0	0	99.78
3	21.93	66.16	11.69	0.0999	8.33e-05	0	0	0	0	99.88
4	11.09	59.66	27.93	1.26	0.0032	3.53e-06	0	0	0	99.94
5	5.56	45.53	43.21	5.62	0.0863	1.04e-04	8.20e-08	0	0	99.92
6	2.78	31.22	51.03	14.34	0.6938	0.0042	4.54e-06	8.89e-10	0	99.37

By Theorem 3, if $\tau(t)=c$, then $\nabla\tau=0$ which alongside (A.80) and (A.77) results in

$$\|S^{\alpha}[\mathbb{L}]\mathsf{L}_c^{\alpha}f-\mathsf{L}_c^{\alpha}S^{\alpha}[\mathbb{L}]f\|_2\leq 2|\cos\alpha|^d|c|\|U^{\alpha}[\mathbb{L}]f\|_2 \quad (52)$$

Meanwhile, it follows from (A.55) that

$$\|\mathsf{L}_c^{\alpha}S^{\alpha}[\mathbb{L}]f-S^{\alpha}[\mathbb{L}]f\|_2\leq C_3 2^{-J}|c|\,\|U^{\alpha}[\mathbb{L}]f\|_2 \quad (53)$$

where $C_3=\|\nabla\phi\|_1$. Then, (50) can be established by applying (53), (52) and (51). ■

Remark 4: If $|c|\ll\min\{2^J,|\sec\alpha|^d\}$, then the FrScatNet is nearly translation-invariant. In fact, the network constructed in practice can not obtain translation invariance (corresponding to $J=\infty$ and $\alpha_i=\pi/2$ for $i=1,2,\ldots,d$), and typically achieves translation stability.

V. FAST COMPUTATION ALGORITHM FOR FRACTIONAL WAVELET SCATTERING NETWORKS

A. Energy Propagation in FrScatNets

By Theorem 2, if the fractional scattering wavelet $\psi(t)$ is admissible, then as m increases, the energy of propagated signals $U^{\alpha}[L^{(m)}]f(t)$ approaches to zero, and by (32), the energy of the FRWSTs $S^{\alpha}[L^{(n)}]f(t)$ also converges to zero for $n\geq m$. This implies that the depth of the FrScatNet can be limited with a negligible loss of signal energy. Theorem 2 also implies that the more sparse the FRWT of $f(t)$, the more energy propagates to deeper layers of the FrScatNet. To be specific, in the first layer, for a given scale and orientation $\lambda_{p,k}\in\Lambda$, it follows from (32) that

$$\|S^{\alpha}[\lambda_{p,k}]f(t)\|_2^2=\left\|\left|f(t)*_{\alpha}\psi_{\lambda_{p,k}}(t)\right|*_{\alpha}\phi_{2^J}(t)\right\|_2^2. \quad (54)$$

Using Lemma 1 with $p=r=2$ and $q=1$, (54) becomes

$$\|S^{\alpha}[\lambda_{p,k}]f(t)\|_2^2\leq 2^{-dJ}\,\|\phi(t)\|_2^2\,\left\|f(t)*_{\alpha}\psi_{\lambda_{p,k}}(t)\right\|_1^2 \quad (55)$$

which implies that as J increases, the energy of $S^{\alpha}[\lambda_{p,k}]f(t)$ converges to $\|\phi(t)\|_2^2\|f(t)*_{\alpha}\psi_{\lambda_{p,k}}(t)\|_1^2$. Indeed, the family of the $L^1(\mathbb{R}^d)$ norms $\{\|f(t)*_{\alpha}\psi_{\lambda_{p,k}}(t)\|_1^2\}_{\lambda_{p,k}\in\Lambda}$ provides a crude signal representation which measures the sparsity of the FRWT in (14). Therefore, the more sparse the FRWT of $f(t)$, the smaller $\|f(t)*_{\alpha}\psi_{\lambda_{p,k}}(t)\|_1^2$, and the more energy propagates towards deeper layers to satisfy the global energy conservation (47). For the chirplet atom shown in Fig. 2, Table I lists the energy of the FRWSTs $S^{\alpha}[L^{(m)}]f(t)$ normalized by signal energy $\|f(t)\|_2^2$ using cubic spline fractional wavelets [37] which are admissible and satisfy (47), where $\alpha_i=1.572$ for $i=1,2$.As can be seen from Table I, the normalized energy $\|S^{\alpha}[L^{(m)}]f(t)\|_2^2/\|f(t)\|_2^2$ appears to decay exponentially with respect to the path length m. As m increases, the energy of $S^{\alpha}[L^{(m)}]f(t)$ approaches zero and is below one percent for $m\geq 4$.

The energy of the propagated signals $U^{\alpha}[L^{(m)}]f(t)$ decays because each signal $U^{\alpha}[l^{(m)}]f(t)$ propagates energy progressively towards lower fractional frequencies due to the demodulation effect of the modulus operation. For a scale and orientation $\lambda_{p,k}\in\Lambda$, it follows from (8) that

$$f(t)*_{\alpha}\psi_{\lambda_{p,k}}(t)=e^{-j\frac{t^{\mathrm{T}}\Omega_0 t}{2}}\left(\left(f(t)e^{j\frac{t^{\mathrm{T}}\Omega_0 t}{2}}\right)*\psi_{\lambda_{p,k}}(t)\right) \quad (56)$$

which implies that the FRWT $f(t)*_{\alpha}\psi_{\lambda_{p,k}}(t)$ of $f(t)$ can be viewed as a modification of the conventional WT $(f(t)e^{j\frac{t^{\mathrm{T}}\Omega_0 t}{2}})*\psi_{\lambda_{p,k}}(t)$ of $f(t)e^{j\frac{t^{\mathrm{T}}\Omega_0 t}{2}}$. One can verify that for any signal $f(t)\in L^2(\mathbb{R}^d)$, $f(t)*\psi_{\lambda_{p,k}}(t)$ is a function whose energy is concentrated in a frequency band of characteristic size 2^{-p}, with a mean frequency also of the order of 2^{-p} [44]. The complex modulus $|f(t)*\psi_{\lambda_{p,k}}(t)|$ tends to have its energy concentrated in a frequency band of characteristic size still equal to 2^{-p}, but now centered around zero. Therefore, it follows from (56) that the fractional frequencies of $|f(t)*_{\alpha}\psi_{\lambda_{p,k}}(t)|$ are globally lower than those of $f(t)*_{\alpha}\psi_{\lambda_{p,k}}(t)$. Then, the majority of the energy of $U^{\alpha}[L^{(m)}]f(t)$ is captured by the low-pass fractional filter $\phi_{2^J}(t)$ that outputs $S^{\alpha}[L^{(m)}]f(t)$. So, less energy is propagated to the next layer. Moreover, if the fractional frequencies of $\psi_{\lambda'_{p,k}}(t)$ are globally higher than those of $\psi_{\lambda_{p,k}}(t)$, i.e., if $|\lambda'_{p,k}|<|\lambda_{p,k}|$, then $|f(t)*_{\alpha}\psi_{\lambda_{p,k}}(t)|*_{\alpha}\psi_{\lambda'_{p,k}}(t)$ is negligible. This implies that the energy of $U^{\alpha}[L^{(m)}]f(t)$ is only propagated along a subset of fractional frequency-decreasing paths $l^{(m)}\in\Lambda^m$ satisfying $|l_n^{(m)}|<|l_{n+1}^{(m)}|$ for $1\leq n\leq m-1$, and the energy along the other paths is negligible. This can be also seen from Table I which shows that the propagated energy is mainly concentrated along the fractional frequency-decreasing paths, and more than 99% of this energy is absorbed by the fractional frequency-decreasing paths of length $m\leq 3$. Thus, it is sufficient to calculate the FRWSTs along the fractional frequency-decreasing paths, which yields a much smaller FrScatNet with its structure illustrated in Fig. 4.

B. Fast Fractional Wavelet Scattering Computations

Based on the facts derived in subsection V-A, a fractional frequency-decreasing path $(2^{p_1}r_k,2^{p_2}r_k,\ldots,2^{p_m}r_k)\in\Lambda^m$ satisfies $0<p_n<p_{n+1}\leq J$ for $1\leq n\leq m-1$ and $1\leq$

$k \leq K$, then the total number of the fractional frequency-decreasing paths of length m is $K^m C_J^m$. Suppose that N is the number of the pixels of the image $f(t)$. Since $\phi_{2^J}(t)$ is a dilated version of $\phi(t)$ with scale 2^J, we have $S^\alpha[l^{(m)}]f(t) = U^\alpha[l^{(m)}]f(t) *_\alpha \phi_{2^J}(t)$. To reduce the computational complexity, $S^\alpha[l^{(m)}]f(t)$ is then uniformly sampled at intervals $a2^J$ with $a = 1$ or $a = 1/2$. This is performed by a periodization of the fractional spectrum [50] in the FRFT domain, which is equivalent to a down-sampling in the spatial domain, since fractional convolutions can be computed in the FRFT domain by the discrete FRFT (DFRFT) and inverse DFRFT (IDFRFT) which have the same computational complexity as the fast Fourier transform (FFT) [40], [46]. Each $S^\alpha[l^{(m)}]f(t)$ is an image with $a^{-2}2^{-2J}N$ coefficients, then the total number of coefficients in a FrScatNet of maximum depth M is $Na^{-2}2^{-2J}\sum_{m=0}^{M} K^m C_J^m$. Also, each of the resulting fractional filtered signals in $U^\alpha[l^{(m)}]f(t)$ is down-sampled by a factor $a2^p, 0 < p \leq J$ to reduce the computational complexity. At the layer m, there are $K^m C_J^m$ propagated signals $U^\alpha[l^{(m)}]f(t)$ which are sampled at intervals $a2^{p_m}$. By induction on m, we can calculate that a maximum total number of samples on layer m is equal to $a^{-2}N(K/3)^m$. Meanwhile, there are $K^m C_J^m$ scattering signals $S^\alpha[l^{(m)}]f(t)$, and they are sampled at intervals 2^J, which results in much fewer samples. Then, the number of operations to compute each layer is given by $O((K/3)^m N \log N)$ operations that are needed to compute the internal propagated signals with the FRFTs. This implies that although the FrScatNet has a free parameter α, it does not incur any additional computation. Based on the derived results, a fast computation algorithm for the FrScatNet is given by Algorithm 1.

Algorithm 1: Fast Computation Algorithm for the FrScatNet.

Input:
f - image, J - scale
Output:
The FRWSTs of f

$\hat{U}^\alpha[l^{(0)}]f = \mathbf{DFRFT}(f)$
$\hat{S}^\alpha[l^{(0)}]f = \hat{U}^\alpha[l^{(0)}]f \cdot \mathbf{FFT}(\phi_{2^J})$
$S^\alpha[l^{(0)}]f = \mathbf{IDFRFT}(\mathbf{periodize}(\hat{S}^\alpha[l^{(0)}]f))$
$U^\alpha[l^{(1)}]f = |\mathbf{IDFRFT}(\mathbf{periodize}(\hat{U}^\alpha[l^{(0)}]f \cdot \mathbf{FFT}(\psi_{l_1^{(1)}})))|$
$\hat{U}^\alpha[l^{(1)}]f = \mathbf{DFRFT}(U^\alpha[l^{(1)}]f)$
$\hat{S}^\alpha[l^{(1)}]f = \hat{U}^\alpha[l^{(1)}]f \cdot \mathbf{FFT}(\phi_{2^J})$
$S^\alpha[l^{(1)}]f = \mathbf{IDFRFT}(\mathbf{periodize}(\hat{S}^\alpha[l^{(1)}]f))$
for $m = 2$ to $J + 1$ **do**
  if $|l_m^{(m)}| > |l_{m-1}^{(m)}|$ **then**
    $U^\alpha[l^{(m)}]f = |\mathbf{IDFRFT}(\mathbf{periodize}(\hat{U}^\alpha[l^{(m-1)}]f \cdot \mathbf{FFT}(\psi_{l_m^{(m)}})))|$
    $\hat{U}^\alpha[l^{(m)}]f = \mathbf{DFRFT}(U^\alpha[l^{(m)}]f)$
    $\hat{S}^\alpha[l^{(m)}]f = \hat{U}^\alpha[l^{(m)}]f \cdot \mathbf{FFT}(\phi_{2^J})$
    $S^\alpha[l^{(m)}]f = \mathbf{IDFRFT}(\mathbf{periodize}(\hat{S}^\alpha[l^{(m)}]f))$
  end if
end for
Output:
for $m = 0$ to $J + 1$ **do**
  $S^\alpha[L^{(m)}]f = \{S^\alpha[l^{(m)}]f\}$
end for

VI. Numerical Results and Applications

In this section, we first test the 2D chirplet atom shown in Fig. 2 (a). When it is deformed with translation, rotation and dilation, its main energy is relocated to other frequencies in both the FT and conventional WST domains, as illustrated in Figs. 2 (e) and (f). Therefore, the conventional WST is not suitable for this type of images. Now, we use the FRWST to the same problem. Since the energy of the chirplet atom is well concentrated and spread over a small band in the FRFT domain, high frequency instability to deformations will decrease. This implies that the FRWST may achieve better performance that the conventional one, and the numerical results are shown in Fig. 5 in which the axis ranges of the plots are the same, where cubic spline fractional wavelets [36] are used with $\alpha_i = 1.572$ for $i = 1, 2$. As can be observed from Fig. 5, when the chirplet atom undergoes the same deformation as in Fig. 2, most of its energy is relocated to other fractional frequencies in the FRFT domain, but remains stable in the FRWST domain. Next, we will discuss an immediate application of the FrScatNet for real image classification.

The FrScatNet provides a translation- and rotation-invariant FRWST representation which is resilient to small deformations. This representation is useful for image classification. The classification is carried out with a generative classifier based on affine space models computed with a PCA [34]. The

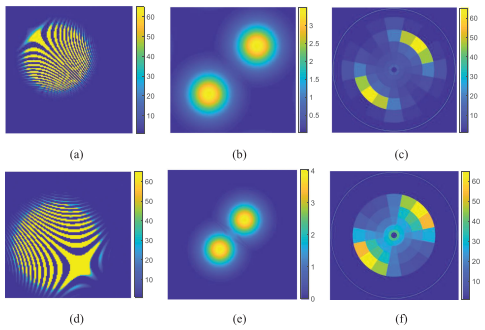

Fig. 5. 2D chirplet atom and its FRFT and FRWST. (a) Original 2D chirplet atom; (b) FRFT modulus of original 2D chirplet atom; (c) FRWST of original 2D chirplet atom; (d) Original 2D chirplet atom deformed by translation, rotation, and dilation; (e) FRFT modulus of the deformed 2D chirplet atom; and (f) FRWST of the deformed 2D chirplet atom.

dimensions of principal component are chosen as the number of training sample images in the PCA classifier. Here, we use the Morlet wavelet thanks to its good time-frequency localization $\psi(t) = e^{-\frac{\|t\|_2^2}{2\sigma^2}+j\nu t}$, $\nu \in \mathbb{R}^d$ which satisfies $\int_{\mathbb{R}^d} \psi(t)dt = 0$, and its FT $\Psi(\omega) = e^{-(\omega-\nu)^2/2}$ mainly concentrates in the interval of $[0, 2\nu]$. Since most camera images have negligible energy outside the frequency interval of $[-\pi, +\pi]$ [15], we choose $\nu = 3\pi/4$ so that the frequencies of $\psi(t)$ contain the highest frequency of digital images. Moreover, we select the Gaussian

TABLE II
WEIGHTED VALUES OF THE RELIEFF ALGORITHM FOR THE FRWSTS UNDER DIFFERENT ANGLES

Angle pairs	$(\alpha_1 = \pi/2, \alpha_2)$										$(\alpha_1, \alpha_2 = \pi/2)$								
	$1\pi/20$	$2\pi/20$	$3\pi/20$	$4\pi/20$	$5\pi/20$	$6\pi/20$	$7\pi/20$	$8\pi/20$	$9\pi/20$	$\pi/2$	$1\pi/20$	$2\pi/20$	$3\pi/20$	$4\pi/20$	$5\pi/20$	$6\pi/20$	$7\pi/20$	$8\pi/20$	$9\pi/20$
Weighted values	46.58	56.59	66.94	56.32	53.64	68.12	52.50	53.66	59.03	65.50	45.85	53.88	**68.36**	59.95	54.14	**70.01**	56.35	54.59	57.12

TABLE III
CLASSIFICATION ERRORS USING DIFFERENT NETWORKS

Training size	FrScatNet		DSN	TFSN	DCNN
	$(3\pi/20, \pi/2)$	$(6\pi/20, \pi/2)$			
2	$\mathbf{0.11 \pm 0.04}$	0.16 ± 0.05	0.26 ± 0.03	0.23 ± 0.04	0.45 ± 0.08
5	$\mathbf{0.06 \pm 0.02}$	0.10 ± 0.04	0.14 ± 0.02	0.14 ± 0.03	0.35 ± 0.09
8	$\mathbf{0.05 \pm 0.02}$	0.07 ± 0.02	0.08 ± 0.02	0.07 ± 0.03	0.34 ± 0.12
10	$\mathbf{0.04 \pm 0.02}$	0.06 ± 0.01	0.05 ± 0.01	0.06 ± 0.02	0.30 ± 0.11
15	$\mathbf{0.03 \pm 0.01}$	0.03 ± 0.01	0.03 ± 0.02	0.05 ± 0.02	0.18 ± 0.07
20	$\mathbf{0.01 \pm 0.00}$	0.02 ± 0.01	0.03 ± 0.02	0.04 ± 0.02	0.12 ± 0.05
50	$\mathbf{0.01 \pm 0.00}$	0.01 ± 0.00	0.01 ± 0.02	0.03 ± 0.01	0.04 ± 0.03
100	$\mathbf{0.00 \pm 0.00}$	0.01 ± 0.00	0.00 ± 0.00	0.02 ± 0.01	0.02 ± 0.01
200	$\mathbf{0.00 \pm 0.00}$	0.00 ± 0.00	0.00 ± 0.00	0.02 ± 0.01	0.01 ± 0.01

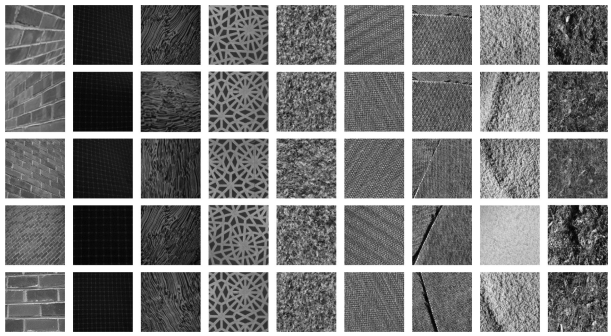

Fig. 6. Sample images from the UMD [52] and ALOI [53] databases.

function as the scaling function, i.e., $\phi(t) = \frac{1}{(\sqrt{2\pi}\rho)^d} e^{-\frac{\|t\|_2^2}{2\rho^2}}$. The finest scale is $2^J = 16$, the total number of angles r_k is $K = 8$, and $\sigma = \rho = 0.8$. As aforementioned, the FrScatNet has a free parameter $\alpha \in [0, \pi/2]^d$, on which the classification performance depends. For image analysis, we set $d = 2$ and $\alpha = (\alpha_1, \alpha_2)$. According to [34], we fix one element of α to $\pi/2$, and let the other vary from 0 to $\pi/2$ with a step of $\pi/20$ to accelerate the computation of the FRWSTs. Typically, many candidate features will be generated from different angles [34], while a large increase in the number of the features may lead to redundancy and complexity. So, we need to decide which features with certain angles to retain and which to discard. However, the feature selection is not addressed in [34].A simple but powerful solution to feature selection is the ReliefF algorithm [51], which assigns a weighted value to each feature, indicating the importance of features. The larger the value, the more important the feature. Now, we test our method on the database constructed by merging UMD [52] and ALOI [53], which contains 9 classes of images, as illustrated in Fig. 6, and each class has 400 images. By using the ReliefF algorithm, Table II lists the weighted values for the FRWSTs under different angles. As can be seen from Table II, the angle pairs $(3\pi/20, \pi/2)$, $(6\pi/20, \pi/2)$ lead to important FRWSTs which can be fed into the PCA classier for classification. The results of classification are shown in Table III which gives a comparison[1] of the FrScatNet and several well-known networks (e.g., the conventional DSN [14], the time-frequency scattering network (TFSN) [24], and the DCNN [25]). In classification stage, the images in each class are divided into two groups randomly. One group of images is labeled for training, and the other group is for testing. We take the fourth row of Table III as an example. In each class containing 400 images, 10 sample images are randomly chosen for training, and the remaining 390 images are used for testing. The resulting mean and variance of classification errors for the FrScatNet with $(3\pi/20, \pi/2)$ are calculated as 0.04 and 0.02, respectively, over 100 experiments. As can be observed from Table III, the FrScatNet with angle pair $(3\pi/20, \pi/2)$ yields the minimum classification error, and even with small size of training, its performance is always best. Particularly, when the size of training is extremely small, the FrScatNet still exhibits best performance.

VII. CONCLUSION

This paper aims to address the deficiency of the conventional DSNs based upon the FRWT which can be interpreted as a bank of linear translation-variant bandpass filters and is suited for non-stationary texture analysis. We first presented a fractional wavelet scattering transform associated with the FRWT, which leads to the fractional wavelet scattering network (FrScatNet). By using the derived results, we established a mathematical model for the FrScatNet. Basic properties of the FrScatNet were derived, and a fast algorithm for its implementation was also presented. Excellent numerical results of the FrScatNet were presented to demonstrate the application of the FrScatNet in the field of image classification.

[1]To ensure reproducibility, we share our code that is used to generate our experimental results freely. [Online]. Available: http://homepage.hit.edu.cn/junshi.

APPENDIX A
PROOF OF LEMMA 1

Using (4), it follows from the left side of (9) that

$$\|f(t)*_\alpha g(t)\|_r^r = \int_{\mathbb{R}^d}\Big|\int_{\mathbb{R}^d} f(\tau)e^{j\frac{\tau^{\mathrm{T}}\Omega_0\tau}{2}}g(t-\tau)d\tau\Big|^r dt \le \int_{\mathbb{R}^d}\Big(\int_{\mathbb{R}^d}|f(\tau)g(t-\tau)|\,d\tau\Big)^r dt. \quad (A.1)$$

Next, combining Young's convolution inequality [42] and the assumption that $\frac{1}{p}+\frac{1}{q}=\frac{1}{r}+1$ with $1\le p,q\le r\le\infty$ yields

$$\begin{aligned}\|f(t)*g(t)\|_r^r &= \int_{\mathbb{R}^d}\Big|\int_{\mathbb{R}^d} f(\tau)g(t-\tau)d\tau\Big|^r dt\\ &\le \int_{\mathbb{R}^d}\Big(\int_{\mathbb{R}^d}|f(\tau)g(t-\tau)|\,d\tau\Big)^r dt\\ &\le \|f(t)\|_q^r\,\|g(t)\|_p^r \end{aligned} \quad (A.2)$$

which alongside (A.1) leads to (9).

APPENDIX B
PROOF OF LEMMA 2

Since $h(t)\ge 0$, it follows from (8) that

$$\Big||f(t)|*_\alpha h(t)\Big| = \Big|\int_{\mathbb{R}^d}\Big|f(\tau)h(t-\tau)e^{j\gamma^T(t-\tau)}\Big|e^{j\frac{\tau^{\mathrm{T}}\Omega_0\tau}{2}}d\tau\Big| \quad (A.3)$$

which alongside generalized Fresnel integrals [55] results in

$$\begin{aligned}\Big||f(t)|*_\alpha h(t)\Big| &\ge \Big|\int_{\mathbb{R}^d} f(\tau)h(t-\tau)e^{j\gamma^T(t-\tau)}e^{j\frac{\tau^{\mathrm{T}}\Omega_0\tau}{2}}d\tau\Big| \\ &= |f*_\alpha h_\gamma|\,.\end{aligned} \quad (A.4)$$

APPENDIX C
PROOF OF THEOREM 2

Since $\mathcal{G}^\alpha$ preserves norm, using (23), (38), and (37) yields

$$\begin{aligned}\|f\|_2^2 &= \|S^\alpha[L^{(0)}]f\|_2^2+\|S^\alpha[L^{(1)}]f\|_2^2+\cdots\\ &\quad+\|S^\alpha[L^{(m)}]f\|_2^2+\|U^\alpha[L^{(m+1)}]f\|_2^2\\ &= \sum_{n=0}^{m}\|S^\alpha[L^{(n)}]f\|_2^2+\|U^\alpha[L^{(m+1)}]f\|_2^2.\end{aligned} \quad (A.5)$$

Then, as $m\to+\infty$, it follows from (A.5) and (24) that

$$\|f\|_2^2 = \|S^\alpha[\mathbb{L}]f\|_2^2+\lim_{m\to+\infty}\|U^\alpha[L^{(m+1)}]f\|_2^2. \quad (A.6)$$

This implies that if (46) holds, then (47) is attainable. Now, our objective is then to prove (46). For simplicity, let

$$e_m^\alpha \triangleq \|U^\alpha[L^{(m)}]f\|_2^2 = \sum_{l^{(m)}\in L^{(m)}}\|U^\alpha[l^{(m)}]f\|_2^2. \quad (A.7)$$

The arrival log fractional frequency of $l^{(m)}\in\Lambda^m$ is the log frequency index $\log_2|\lambda_{p_m,k}| = |2^{p_m}r_k| = p_m\le J$ of the last element of the path. Then, the mean of the arrival log fractional frequencies along paths of length m is given by

$$a_m^\alpha \triangleq \sum_{l^{(m)}\in L^{(m)}} p_m\frac{\|U^\alpha[l^{(m)}]f\|_2^2}{\sum_{l^{(m)}\in L^{(m)}}\|U^\alpha[l^{(m)}]f\|_2^2}. \quad (A.8)$$

Since $p_m\le J$, it is clear from (A.8) that

$$a_m^\alpha \le J < J+1,\ \forall m\in\mathbb{Z}^+. \quad (A.9)$$

Also, let

$$\bar{e}_m^\alpha \triangleq \|S^\alpha[L^{(m)}]f\|_2^2. \quad (A.10)$$

Then, to prove (46), it suffices to show that

$$ce_{m-1}^\alpha \le a_{m+1}^\alpha e_{m+1}^\alpha - a_m^\alpha e_m^\alpha + (J+1)\bar{e}_m^\alpha + \bar{e}_{m-1}^\alpha \quad (A.11)$$

for any $m\ge 1$ if (45) holds. Next, using (37) and (23) yields

$$\|U^\alpha[L^{(m)}]f\|_2^2 = \|S^\alpha[L^{(m)}]f\|_2^2+\|U^\alpha[L^{(m+1)}]f\|_2^2 \quad (A.12)$$

which alongside (A.7) and (A.10) results in $e_m^\alpha = \bar{e}_m^\alpha + e_{m+1}^\alpha$. Inserting this result into (A.11) gives rise to

$$\begin{aligned}ce_{m-1}^\alpha &\le (a_{m+1}^\alpha-(J+1))e_{m+1}^\alpha\\ &\quad-(a_m^\alpha-(J+1))e_m^\alpha+e_{m-1}^\alpha-e_m^\alpha.\end{aligned} \quad (A.13)$$

Summing m from 1 to M on both sides of (A.13) yields

$$\begin{aligned}c\sum_{m=1}^{M}e_{m-1}^\alpha &\le (a_{m+1}^\alpha-(J+1))e_{m+1}^\alpha\\ &\quad-(a_1^\alpha-(J+1))e_1^\alpha+e_0^\alpha-e_m^\alpha\end{aligned} \quad (A.14)$$

which alongside (A.9) yields $c\sum_{m=0}^{M-1}e_m^\alpha \le e_0^\alpha - a_1^\alpha e_1^\alpha + (J+1)e_1^\alpha$ which alongside (A.7) and (24) leads to

$$c\|U^\alpha[\mathbb{L}]f\|_2^2 \le e_0^\alpha - a_1^\alpha e_1^\alpha + (J+1)e_1^\alpha. \quad (A.15)$$

Next, since $\mathcal{Q}^\alpha$ is unitary, setting $m=0,1$ in (A.7) yields

$$e_0^\alpha = \|f\|_2^2,\ e_1^\alpha = \sum_{p\le J}\sum_{k=1}^{K}\|f*_\alpha\psi_{2^pr_k}\|_2^2 \le \|f\|_2^2. \quad (A.16)$$

Meanwhile, applying (A.8) with $m=1$ results in

$$a_1^\alpha e_1^\alpha = \sum_{p\le J}\sum_{k=1}^{K}p\|f*_\alpha\psi_{2^pr_k}\|_2^2. \quad (A.17)$$

Then, inserting (A.16) and (A.17) into (A.15) gives rise to

$$c\|U^\alpha[\mathbb{L}]f\|_2^2 \le (J+2)\|f\|_2^2 - \sum_{p\le J}\sum_{k=1}^{K}p\|f*_\alpha\psi_{2^pr_k}\|_2^2. \quad (A.18)$$

Since $J\in\mathbb{Z}$, it follows from (A.18) that

$$\begin{aligned}c\|U^\alpha[\mathbb{L}]f\|_2^2 &\le \max\{J+2,1\}\|f\|_2^2\\ &\quad+\sum_{p\le 0}\sum_{k=1}^{K}(-p)\|f*_\alpha\psi_{2^pr_k}\|_2^2.\end{aligned} \quad (A.19)$$

Next, for any $f(t)\in L^2(\mathbb{R}^d)$, let $f_n(t) = (2\pi)^{-\frac{d}{2}}f(t)*_\alpha(2^{-dn}\phi(2^{-n}t))$. Then, applying (5) leads to $F_{n,\alpha}(u) = F_\alpha(u)\Phi(2^n\Omega_s u)$, where $F_{n,\alpha}(u)$ and $F_\alpha(u)$ denote the FRFTs

of $f_n(t)$ and $f(t)$, respectively. Since $\phi(t) \in L^1(\mathbb{R}^d)$ and $\Phi(0) = 1$, as $n \to -\infty$, we have $F_{n,\alpha}(u) \to F_\alpha(u)$, i.e.,

$$\lim_{n\to-\infty} \|f - f_n\|_2^2 = 0,\ \forall f(t) \in L^2(\mathbb{R}^d). \tag{A.20}$$

For a given $n \in \mathbb{Z}$, replacing $f(t)$ with $f_n(t)$ in (A.19) and applying (5) yields

$$\begin{aligned} &c\|U^\alpha[\mathbb{L}]f_n\|_2^2 \\ &\quad\le \max\{J+2,1\}\|f_n\|_2^2 + \sum_{p\le 0}(-p)2^{2p}2^{-2n}(2\pi)^d \\ &\qquad\times \int_{u\in\mathbb{R}^d} \left|F_\alpha(u)\right|^2 \left|(2^n\Omega_s u)\Phi(2^n\Omega_s u)\right|^2 \\ &\qquad\times \sum_{k=1}^{K} \frac{\left|\Psi\left(2^p r_k^{-1}\Omega_s u\right)\right|^2}{|2^p r_k^{-1}\Omega_s u|^2} du. \end{aligned} \tag{A.21}$$

Since the fractional wavelet $\psi(t)$ has a vanishing moment, we have $|\Psi(\Omega_s u)| = O(|\Omega_s u|)$. In addition, since the derivatives of the scaling function $\phi(t)$ lie in $L^1(\mathbb{R}^d)$, $|(\Omega_s u)\Phi(\Omega_s u)|$ is bounded. Then, using (A.21) yields

$$\begin{aligned} c\|U^\alpha[\mathbb{L}]f_n\|_2^2 &\le \max\{J+2,1\}\|f_n\|_2^2 \\ &\quad + C_4(2\pi)^d 2^{-2n}\sum_{p\le 0} 2^{2p}(-p)\|F_\alpha\|_2^2 \le +\infty \end{aligned} \tag{A.22}$$

where constant $C_4 = \|(\Omega_s u)\Phi(\Omega_s u)\|_2^2 = \|\nabla\phi\|_2^2$. Combining (A.22) and (24) results in

$$\|U^\alpha[\mathbb{L}]f_n\|_2^2 = \sum_{m=0}^{+\infty} \|U^\alpha[L^{(m)}]f_n\|_2^2 < +\infty \tag{A.23}$$

which implies that

$$\lim_{m\to+\infty} \|U^\alpha[L^{(m)}]f_n\|_2^2 = 0. \tag{A.24}$$

Moreover, for any $f(t), h(t) \in L^2(\mathbb{R}^d)$, using (41) yields

$$\left\|U^\alpha[l^{(m+1)}]f - U^\alpha[l^{(m+1)}]h\right\|_2^2 \le \left\|f - h\right\|_2^2 \tag{A.25}$$

which alongside (23) shows that $U^\alpha[L^{(m)}]$ is non-expansive. Therefore, we can derive

$$\|U^\alpha[L^{(m)}]f\|_2^2 \le \|f - f_n\|_2^2 + \|U^\alpha[L^{(m)}]f_n\|_2^2. \tag{A.26}$$

Then, using (A.20) and (A.24) yields $\lim_{m\to+\infty}\|U^\alpha[L^{(m)}]f\|_2^2 = 0$ which alongside (A.6) leads to (47). Thus, we now just need to prove that (A.11) is true. For brevity, we rewrite the right side of (A.11) as $I + \bar{e}_{m-1}^\alpha$, where $I \triangleq a_{m+1}^\alpha e_{m+1}^\alpha - a_m^\alpha e_m^\alpha + (J+1)\bar{e}_m^\alpha$. Next, it follows from (A.7) and (A.8) that

$$a_{m+1}^\alpha e_{m+1}^\alpha = \sum_{l^{(m+1)}\in L^{(m+1)}} p_{m+1}\|U^\alpha[l^{(m+1)}]f\|_2^2, \tag{A.27}$$

and $a_m^\alpha e_m^\alpha = \sum_{l^{(m)}\in L^{(m)}} p_m\|U^\alpha[l^{(m)}]f\|_2^2$ which alongside (37) results in

$$a_m^\alpha e_m^\alpha = \sum_{l^{(m)}\in L^{(m)}} p_m\left(\|S^\alpha[l^{(m)}]f\|_2^2 + \|U^\alpha[l^{(m+1)}]f\|_2^2\right). \tag{A.28}$$

Note that $l^{(m+1)} = (l^{(m-1)}, l_m^{(m)}, l_{m+1}^{(m+1)})$. Then, inserting (A.27), (A.28), and (A.10) into I and using $p_m \le J < J+1$ gives rise to

$$\begin{aligned} I \ge &\sum_{l^{(m-1)}\in L^{(m-1)}} \sum_{p_m\le J}\sum_{k=1}^{K}\Bigg\{ \sum_{p_m<p_{m+1}<J+1} \\ &\sum_{k'=1}^{K} \left\| |U^\alpha[l^{(m-1)}]f *_\alpha \psi_{2^{p_m}r_k}| \right. \\ &\left. *_\alpha \psi_{2^{p_{m+1}}r_{k'}}\right\|_2^2 + \sum_{p_{m+1}<p_m}\sum_{k'=1}^{K}(p_{m+1}-p_m) \\ &\times \left\||U^\alpha[l^{(m-1)}]f *_\alpha \psi_{2^{p_m}r_k}| *_\alpha \psi_{2^{p_{m+1}}r_{k'}}\right\|_2^2 \\ &+ \left\||U^\alpha[l^{(m-1)}]f *_\alpha \psi_{2^{p_m}r_k}| *_\alpha \phi_{2^J}\right\|_2^2\Bigg\}. \end{aligned} \tag{A.29}$$

Next, it follows from (22) that

$$\left|\Phi(2^J\Omega_s u)\right|^2 = \sum_{p_{m+1}>J}\sum_{k'=1}^{K}\left|\Psi(2^{p_{m+1}}r_{k'}^{-1}\Omega_s u)\right|^2 \tag{A.30}$$

and

$$\begin{aligned} \left|\Phi(2^{p_m}\Omega_s u)\right|^2 &= \left|\Phi(2^J\Omega_s u)\right|^2 \\ &\quad + \sum_{p_m<p_{m+1}\le J}\sum_{k'=1}^{K}\left|\Psi(2^{p_{m+1}}r_{k'}^{-1}\Omega_s u)\right|^2. \end{aligned} \tag{A.31}$$

This result implies that

$$\begin{aligned} \left\|f *_\alpha \phi_{2^{p_m}}\right\|_2^2 &= \sum_{p_m<p_{m+1}\le J}\sum_{k'=1}^{K}\left\|f *_\alpha \psi_{2^{p_{m+1}}r_{k'}}\right\|_2^2 \\ &\quad + \left\|f *_\alpha \phi_{2^J}\right\|_2^2. \end{aligned} \tag{A.32}$$

Replacing $f(t)$ with $|U^\alpha[l^{(m-1)}]f *_\alpha \psi_{2^{p_m}r_k}|$ in (A.32) and inserting the result into (A.29) gives rise to

$$\begin{aligned} I \ge &\sum_{l^{(m-1)}\in L^{(m-1)}}\sum_{p_m\le J} \\ &\sum_{k=1}^{K}\Bigg\{\left\||U^\alpha[l^{(m-1)}]f *_\alpha \psi_{2^{p_m}r_k}| *_\alpha \phi_{2^{p_m}}\right\|_2^2 \\ &- \sum_{p_{m+1}<p_m}(p_m - p_{m+1})\sum_{k'=1}^{K}\left\||U^\alpha[l^{(m-1)}]f \right. \\ &\left. *_\alpha \psi_{2^{p_m}r_k}| *_\alpha \psi_{2^{p_{m+1}}r_{k'}}\right\|_2^2\Bigg\}. \end{aligned} \tag{A.33}$$

Next, note that

$$\sum_{k'=1}^{K}\left\||U^{\alpha}[l^{(m-1)}]f *_{\alpha}\psi_{2^{p_m}r_k}|*_{\alpha}\psi_{2^{p_{m+1}}r_{k'}}\right\|_2^2 = \sum_{(p_{m+1}-1)<p_{m+1}\le p_{m+1}} \sum_{k'=1}^{K}\left\||U^{\alpha}[l^{(m-1)}]f *_{\alpha}\psi_{2^{p_m}r_k}|*_{\alpha}\psi_{2^{p_{m+1}}r_{k'}}\right\|_2^2 \tag{A.34}$$

which alongside (A.32) leads to

$$\sum_{k'=1}^{K}\left\||U^{\alpha}[l^{(m-1)}]f *_{\alpha}\psi_{2^{p_m}r_k}|*_{\alpha}\psi_{2^{p_{m+1}}r_{k'}}\right\|_2^2 = \left\||U^{\alpha}[l^{(m-1)}]f *_{\alpha}\psi_{2^{p_m}r_k}|*_{\alpha}\phi_{2^{(p_{m+1}-1)}}\right\|_2^2 - \left\||U^{\alpha}[l^{(m-1)}]f *_{\alpha}\psi_{2^{p_m}r_k}|*_{\alpha}\phi_{2^{p_{m+1}}}\right\|_2^2. \tag{A.35}$$

Inserting (A.35) into (A.33) results in

$$\begin{aligned} I \ge & \sum_{l^{(m-1)}\in L^{(m-1)}}\sum_{p_m\le J}\sum_{k=1}^{K}\left\{\left\||U^{\alpha}[l^{(m-1)}]f *_{\alpha}\psi_{2^{p_m}r_k}|*_{\alpha}\phi_{2^{p_m}}\right\|_2^2 - \sum_{p_{m+1}<p_m}(p_m-p_{m+1})\right. \\ & \left.\times\left(\left\||U^{\alpha}[l^{(m-1)}]f *_{\alpha}\psi_{2^{p_m}r_k}|*_{\alpha}\phi_{2^{(p_{m+1}-1)}}\right\|_2^2 - \left\||U^{\alpha}[l^{(m-1)}]f *_{\alpha}\psi_{2^{p_m}r_k}|*_{\alpha}\phi_{2^{p_{m+1}}}\right\|_2^2\right)\right\}. \end{aligned} \tag{A.36}$$

By assumption, there exists $\theta(t)>0$ with $|\Theta(\Omega_s u)|\le|\Phi(\Omega_s u)|$ and $\Theta(0)=(2\pi)^{-\frac{d}{2}}$. Then, using (5) yields

$$\|f *_{\alpha}\phi_{2^{p_m}}\|_2^2 \ge \|f *_{\alpha}\theta_{2^{p_m}r_k}\|_2^2. \tag{A.37}$$

with $\theta_{2^{p_m}r_k}(t)\triangleq 2^{-dp_m}\theta(2^{-p_m}r_k^{-1}t)$. Combining (A.37), (A.36), and (37) results in

$$\begin{aligned} I \ge & \sum_{l^{(m-1)}\in L^{(m-1)}}\sum_{p_m\le J}\sum_{k=1}^{K}\left\{\left\||U^{\alpha}[l^{(m-1)}]f *_{\alpha}\psi_{2^{p_m}r_k}|*_{\alpha}\theta_{2^{p_m}r_k}\right\|_2^2 - \sum_{p_{m+1}<p_m}(p_m-p_{m+1})\right. \\ & \left.\times\left(\left\|U^{\alpha}[l^{(m-1)}]f *_{\alpha}\psi_{2^{p_m}r_k}\right\|_2^2 - \left\||U^{\alpha}[l^{(m-1)}]f *_{\alpha}\psi_{2^{p_m}r_k}|*_{\alpha}\theta_{2^{p_{m+1}}r_k}\right\|_2^2\right)\right\}. \end{aligned} \tag{A.38}$$

Next, in Lemma 2, setting $f(t)=U^{\alpha}[l^{(m-1)}]f(t)*_{\alpha}\psi_{2^{p_m}r_k}(t)$, $h(t)=\theta_{2^{p_m}r_k}(t)$, and a fractional frequency $2^{-p_m}r_k\gamma$ leads to

$$\left|\left|U^{\alpha}[l^{(m-1)}]f *_{\alpha}\psi_{2^{p_m}r_k}\right|*_{\alpha}\theta_{2^{p_m}r_k}\right| \ge \left|U^{\alpha}[l^{(m-1)}]f *_{\alpha}\psi_{2^{p_m}r_k}*_{\alpha}\theta_{2^{p_m}r_k}(t)e^{j2^{-p_m}r_k\gamma t}\right| \tag{A.39}$$

which alongside (A.38) derives

$$\begin{aligned} I \ge & \sum_{l^{(m-1)}\in L^{(m-1)}}\sum_{p_m\le J}\sum_{k=1}^{K}\left\{\left\|U^{\alpha}[l^{(m-1)}]f *_{\alpha}\psi_{2^{p_m}r_k}*_{\alpha}\theta_{2^{p_m}r_k,2^{p_m}}\right\|^2 - \sum_{p_{m+1}<p_m}(p_m-p_{m+1})\right. \\ & \left.\times\left[\left\|U^{\alpha}[l^{(m-1)}]f *_{\alpha}\psi_{2^{p_m}r_k}\right\|^2 - \left\|U^{\alpha}[l^{(m-1)}]f *_{\alpha}\psi_{2^{p_m}r_k}*_{\alpha}\theta_{2^{p_{m+1}}r_k,2^{p_m}}\right\|^2\right]\right\} \end{aligned} \tag{A.40}$$

where $\theta_{2^{p_m}r_k,2^{p_m}}(t)=\theta_{2^{p_m}r_k}(t)e^{j2^{-p_m}r_k\gamma\cdot t}$ and $\theta_{2^{p_{m+1}}r_k,2^{p_m}}(t)=\theta_{2^{p_{m+1}}r_k}(t)e^{j2^{-p_m}r_k\gamma\cdot t}$. Then, it follows from (5) that (A.40) can be rewritten in terms of the FRFT representation as

$$\begin{aligned} I \ge & \sum_{l^{(m-1)}\in L^{(m-1)}}\int_{u\in\mathbb{R}^d}(2\pi)^{2d}\left|\mathcal{F}^{\alpha}\left\{U^{\alpha}[l^{(m-1)}]f\right\}(u)\right|^2 \times\sum_{p_m\le J}\sum_{k=1}^{K}\left|\Psi(2^{p_m}r_k^{-1}\Omega_s u)\right|^2 \\ & \left(\left|\Theta(2^{p_m}r_k^{-1}\Omega_s u-\gamma)\right|^2 - \sum_{p_{m+1}<p_m}(p_m-p_{m+1})\times\left(1-\left|\Theta(2^{p_{m+1}}r_k^{-1}\Omega_s u-2^{-(p_m-p_{m+1})}\gamma)\right|^2\right)\right)du. \end{aligned} \tag{A.41}$$

Since $\theta(t)>0$, it follows that

$$\begin{aligned} |\Theta(\Omega_s u)| &= \left|\tfrac{1}{(\sqrt{2\pi})^d}\int_{\mathbb{R}^d}\theta(t)e^{-jt^{\mathrm{T}}\Omega_s u}dt\right| \\ &\le \tfrac{1}{(\sqrt{2\pi})^d}\int_{\mathbb{R}^d}\theta(t)dt=\Theta(0)=(2\pi)^{-\frac{d}{2}}. \end{aligned} \tag{A.42}$$

Thus, $|\Theta(\Omega_s u)|^2\le(2\pi)^{-d}$ which implies that

$$\left|\Theta(2^{p_m}r_k^{-1}\Omega_s u-\gamma)\right|^2 - \sum_{b=1}^{+\infty}b\left(1-\left|\Theta(2^{p_m-b}r_k^{-1}\Omega_s u-2^{-b}\gamma)\right|^2\right)<(2\pi)^{-d}. \tag{A.43}$$

Next, we compute $\bar{e}_{m-1}^{\alpha}$ in (A.11).

$$\begin{aligned} \bar{e}_{m-1}^{\alpha} &= \sum_{l^{(m-1)}\in L^{(m-1)}}\|U^{\alpha}[l^{(m-1)}]f *_{\alpha}\phi_{2^J}\|_2^2 \\ &= \sum_{l^{(m-1)}\in L^{(m-1)}}\int_{u\in\mathbb{R}^d}\left|\mathcal{F}^{\alpha}\left\{U^{\alpha}[l^{(m-1)}]f\right\}(u)\right|^2 \times(2\pi)^d|\Phi(2^J\Omega_s u)|^2\,du \end{aligned} \tag{A.44}$$

which alongside (A.30) leads to

$$\bar{e}^{\alpha}_{m-1} = \sum_{l^{(m-1)}\in L^{(m-1)}} \int_{u\in\mathbb{R}^d} \left|\mathcal{F}^{\alpha}\left\{U^{\alpha}[l^{(m-1)}]f\right\}(u)\right|^2 \times (2\pi)^d \sum_{p_m>J}\sum_{k=1}^{K} \left|\Psi(2^{p_m} r_k^{-1}\Omega_{\mathrm{s}} u)\right|^2 du. \tag{A.45}$$

Then, combining (A.45) and (A.43) gives rise to

$$\begin{aligned}\bar{e}^{\alpha}_{m-1} \geq & \sum_{l^{(m-1)}\in L^{(m-1)}} \int_{u\in\mathbb{R}^d} \left|\mathcal{F}^{\alpha}\left\{U^{\alpha}[l^{(m-1)}]f\right\}(u)\right|^2 \\ & \times (2\pi)^d \sum_{p_m>J}\sum_{k=1}^{K} \left|\Psi(2^{p_m} r_k^{-1}\Omega_{\mathrm{s}} u)\right|^2 \\ & \times (2\pi)^d \Bigg(\left|\Theta(2^{p_m} r_k^{-1}\Omega_{\mathrm{s}} u-\gamma)\right|^2 \\ & -\sum_{b=1}^{+\infty} b\left(1-\left|\Theta(2^{p_m-b} r_k^{-1}\Omega_{\mathrm{s}} u-2^{-b}\gamma)\right|^2\right)\Bigg) du.\end{aligned} \tag{A.46}$$

Now, combining (A.41) and (A.46) yields

$$\begin{aligned} I+\bar{e}^{\alpha}_{m-1} \geq & \\ & \sum_{l^{(m-1)}\in L^{(m-1)}} \int_{u\in\mathbb{R}^d} \left|\mathcal{F}^{\alpha}\left\{U^{\alpha}[l^{(m-1)}]f\right\}(u)\right|^2 \\ & \times (2\pi)^{2d} \sum_{p_m=-\infty}^{+\infty}\sum_{k=1}^{K} \left|\Psi(2^{p_m} r_k^{-1}\Omega_{\mathrm{s}} u)\right|^2 \\ & \Bigg(\left|\Theta(2^{p_m} r_k^{-1}\Omega_{\mathrm{s}} u-\gamma)\right|^2 - \sum_{b=1}^{+\infty} b \\ & \left(1-\left|\Theta(2^{p_m-b} r_k^{-1}\Omega_{\mathrm{s}} u-2^{-b}\gamma)\right|^2\right)\Bigg) du \\ & \geq c \sum_{l^{(m-1)}\in L^{(m-1)}} \|U^{\alpha}[l^{(m-1)}]f\|_2^2 \end{aligned} \tag{A.47}$$

which alongside (A.7) yields (A.11). This completes the proof of Theorem 2.

Appendix D
Proof of Theorem 3

It follows from the left side of (48) that

$$\begin{aligned}\|S^{\alpha}[\mathbb{L}]\mathsf{L}^{\alpha}_{\tau} f - S^{\alpha}[\mathbb{L}]f\|_2 \leq & \left\|\mathsf{L}^{\alpha}_{\tau} S^{\alpha}[\mathbb{L}]f - S^{\alpha}[\mathbb{L}]f\right\|_2 \\ & + \left\|S^{\alpha}[\mathbb{L}]\mathsf{L}^{\alpha}_{\tau} f - \mathsf{L}^{\alpha}_{\tau} S^{\alpha}[\mathbb{L}]f\right\|_2.\end{aligned} \tag{A.48}$$

Next, it follows that

$$\begin{aligned}&\left\|\mathsf{L}^{\alpha}_{\tau} S^{\alpha}[\mathbb{L}]f - S^{\alpha}[\mathbb{L}]f\right\|_2 \\ &\quad = \left\|U^{\alpha}[\mathbb{L}]f *_{\alpha} \left(\phi_{\lambda_J}(t-\tau(t)) - \phi_{\lambda_J}(t)\right)\right\|_2\end{aligned} \tag{A.49}$$

which alongside Lemma 1 with $r = q = 2$ and $p = 1$ leads to

$$\begin{aligned}&\|\mathsf{L}^{\alpha}_{\tau} S^{\alpha}[\mathbb{L}]f - S^{\alpha}[\mathbb{L}]f\|_2 \\ &\quad \leq \|U^{\alpha}[\mathbb{L}]f\|_2 \int_{\mathbb{R}^d} \left|\phi_{2^J}(t-\tau(t)) - \phi_{2^J}(t)\right| dt.\end{aligned} \tag{A.50}$$

Next, for brevity, let $h_t(x) \triangleq \phi_{\lambda_J}(t - x\tau(t)) - \phi_{\lambda_J}(t)$, $x \in \mathbb{R}$. It is easy to show that

$$h'_t(x) \triangleq \tfrac{d}{dx} h_t(x) = -\nabla\phi_{\lambda_J}(t - x\tau(t))\,\tau(t). \tag{A.51}$$

Then, we can express $h_t(x)$ in terms of a Taylor series [46] with respect to t as $h_t(x) = h_t(0) + \int_0^x h'_t(\eta)d\eta$, where $h_t(0) = 0$. Therefore, combining (A.50) and (A.51) yields

$$\begin{aligned}&\|\mathsf{L}^{\alpha}_{\tau} S^{\alpha}[\mathbb{L}]f - S^{\alpha}[\mathbb{L}]f\|_2 \\ &\leq \|U^{\alpha}[\mathbb{L}]f\|_2 \int_{\mathbb{R}^d} |h_t(1)|\, dt \\ &\leq \|U^{\alpha}[\mathbb{L}]f\|_2 \int_{\mathbb{R}^d}\int_{[0,1]} |h'_t(\eta)|\, d\eta dt \\ &\leq \|U^{\alpha}[\mathbb{L}]f\|_2 \|\tau\|_{\infty} \int_{\mathbb{R}^d}\int_{[0,1]} |\nabla\phi_{\lambda_J}(t-\eta\tau(t))|\, d\eta dt.\end{aligned} \tag{A.52}$$

Changing the variable $y = t - \eta\tau(t)$ in (A.52) leads to

$$\begin{aligned}\left\|\mathsf{L}^{\alpha}_{\tau} S^{\alpha}[\mathbb{L}]f - S^{\alpha}[\mathbb{L}]f\right\|_2 \leq & \|U^{\alpha}[\mathbb{L}]f\|_2 \|\tau\|_{\infty} 2^{-dJ} 2^{-J} \\ & \times \int_{[0,1]}\int_{\mathbb{R}^d} \frac{\left|\nabla\phi^*(-2^{-J}y)\right|}{\left|\det(Id-\nabla\tau(t))\right|} dy d\eta.\end{aligned} \tag{A.53}$$

Since $|\det(Id - \nabla\tau(t))| \geq 1 - d\|\nabla\tau\|_{\infty}$ [54], (A.53) yields

$$\|\mathsf{L}^{\alpha}_{\tau} S^{\alpha}_J[\mathbb{L}]f - S^{\alpha}_J[\mathbb{L}]f\|_2 \leq \tfrac{\|U^{\alpha}[\mathbb{L}]f\|_2 \|\tau\|_{\infty} 2^{-J}}{1-d\|\nabla\tau\|_{\infty}} \|\nabla\phi\|_1. \tag{A.54}$$

By assumption, $\|\nabla\tau\|_{\infty} \leq \frac{1}{2d}$ so that $1 - d\|\nabla\tau\|_{\infty} \geq \frac{1}{2}$. Besides, the scaling function $\phi(t)$ is twice differentiable, and thus $\|\nabla\phi\|_1$ is bounded. Then, using (A.54) leads to

$$\|\mathsf{L}^{\alpha}_{\tau} S^{\alpha}[\mathbb{L}]f - S^{\alpha}[\mathbb{L}]f\|_2 \leq C_5 2^{-J} \|U^{\alpha}[\mathbb{L}]f\|_2 \|\tau\|_{\infty} \tag{A.55}$$

where constant $C_5 = 2\|\nabla\phi\|_1$. Next, we need to derive an upper bound of $\left\|S^{\alpha}[\mathbb{L}]\mathsf{L}^{\alpha}_{\tau} f - \mathsf{L}^{\alpha}_{\tau} S^{\alpha}[\mathbb{L}]f\right\|_2$. For simplicity, we first introduce two operators, i.e., $\mathcal{A}_{\alpha,J} f(t) \triangleq S^{\alpha}[l^{(0)}]f(t) = f(t) *_{\alpha} \phi_{2^J}(t)$ and $\mathcal{D}_{\alpha,J} f(t) \triangleq U^{\alpha}[l^{(1)}]f(t) = \{|f(t) *_{\alpha} \psi_{\lambda_{p,k}}(t)|\}_{\lambda_{p,k}\in\Lambda}$. Since $U[L^{(0)}] = Id$, it follows from (34) and (23) that

$$\mathcal{D}^m_{\alpha,J} f = U^{\alpha}[L^{(m)}]f,\ \forall m \in \mathbb{N}. \tag{A.56}$$

Denote by L_m the set of all paths of length no greater than m, i.e., $L_m = \bigcup_{n=0}^{m} L^{(n)}$ which alongside (A.56) results in

$$S^{\alpha}[L_m]f = \left\{\mathcal{A}_{\alpha,J}\mathcal{D}^n_{\alpha,J} f\right\}_{0\leq n\leq m}. \tag{A.57}$$

Next, it follows from (A.57) that

$$S^{\alpha}[L_m]\mathsf{L}^{\alpha}_{\tau} f = \{\mathsf{L}^{\alpha}_{\tau}\mathcal{A}_{\alpha,J} f, S^{\alpha}_J[L_{m-1}]\mathsf{L}^{\alpha}_{\tau}\mathcal{D}_{\alpha,J} f\} + K^{\alpha}_m f \tag{A.58}$$

where $K^{\alpha}_m = \{\mathcal{A}_{\alpha,J}\mathsf{L}^{\alpha}_{\tau} - \mathsf{L}^{\alpha}_{\tau}\mathcal{A}_{\alpha,J}, S^{\alpha}[L_{m-1}](\mathcal{D}_{\alpha,J}\mathsf{L}^{\alpha}_{\tau} - \mathsf{L}^{\alpha}_{\tau}\mathcal{D}_{\alpha,J})\}$. Therefore, we have $S^{\alpha}[L_{m-1}]\mathsf{L}^{\alpha}_{\tau} =$

$\{\mathsf{L}_\tau^\alpha \mathcal{A}_{\alpha,J}, S^\alpha[L_{m-2}]\mathsf{L}_\tau^\alpha \mathcal{D}_{\alpha,J}\} + K_{m-1}^\alpha$. Then, applying (A.58) and performing m iterations yields

$$S^\alpha[L_m]\mathsf{L}_\tau^\alpha = \mathsf{L}_\tau^\alpha S^\alpha[L_m] + \sum_{n=0}^{m} K_{m-n}^\alpha \mathcal{D}_{\alpha,J}^n. \tag{A.59}$$

Therefore, we arrive at

$$\begin{aligned}\|(S^\alpha[L_m]\mathsf{L}_\tau^\alpha - \mathsf{L}_\tau^\alpha S^\alpha[L_m])f\|_2 \\ \le \sum_{n=0}^{m} \|K_{m-n}^\alpha\|_2 \|\mathcal{D}_{\alpha,J}^n f\|_2.\end{aligned} \tag{A.60}$$

Next, by Theorem 1, $S^\alpha[L_m]$ is non-expansive. Then, we have

$$\|K_m^\alpha\|_2 \le \|\mathcal{A}_{\alpha,J}\mathsf{L}_\tau^\alpha - \mathsf{L}_\tau^\alpha \mathcal{A}_{\alpha,J}\|_2 + \|\mathcal{D}_{\alpha,J}\mathsf{L}_\tau^\alpha - \mathsf{L}_\tau^\alpha \mathcal{D}_{\alpha,J}\|_2 \tag{A.61}$$

which alongside (A.60) leads to

$$\begin{aligned}&\|\left(S^\alpha[L_m]\mathsf{L}_\tau^\alpha - \mathsf{L}_\tau^\alpha S^\alpha[L_m]\right) f\|_2 \\ &\le \Big(\|\mathcal{A}_{\alpha,J}\mathsf{L}_\tau^\alpha - \mathsf{L}_\tau^\alpha \mathcal{A}_{\alpha,J}\|_2 \\ &\quad + \|\mathcal{D}_{\alpha,J}\mathsf{L}_\tau^\alpha - \mathsf{L}_\tau^\alpha \mathcal{D}_{\alpha,J}\|_2\Big) \sum_{n=0}^{m} \|\mathcal{D}_{\alpha,J}^n f\|_2.\end{aligned} \tag{A.62}$$

Combining $L_m = \bigcup_{n=0}^m L^{(n)}$, (24), and (A.62) gives rise to

$$\begin{aligned}&\|S^\alpha[\mathbb{L}]\mathsf{L}_\tau^\alpha f - \mathsf{L}_\tau^\alpha S^\alpha[\mathbb{L}]f\|_2 \\ &\le \left(\|\mathcal{A}_{\alpha,J}\mathsf{L}_\tau^\alpha - \mathsf{L}_\tau^\alpha \mathcal{A}_{\alpha,J}\|_2\right. \\ &\left.+\|\mathcal{D}_{\alpha,J}\mathsf{L}_\tau^\alpha - \mathsf{L}_\tau^\alpha \mathcal{D}_{\alpha,J}\|_2\right) \|U^\alpha[\mathbb{L}]f\|_2.\end{aligned} \tag{A.63}$$

Next, for simplicity, denote by $\tilde{f}(t) \triangleq f(t)e^{j\frac{t^{\mathrm{T}}\Omega_0 t}{2}}$ the chirp-modulated version of $f(t)$. By (8) and (4), it follows that $\mathcal{A}_{\alpha,J}\mathsf{L}_\tau^\alpha f(t) = e^{j\frac{t^{\mathrm{T}}\Omega_0 t}{2}}((\mathsf{L}_\tau \tilde{f}(t)) * \phi_{2^J}(t))$ and $\mathsf{L}_\tau^\alpha \mathcal{A}_{\alpha,J} f(t) = e^{j\frac{t^{\mathrm{T}}\Omega_0 t}{2}}(\tilde{f}(t) * (\mathsf{L}_\tau \phi_{2^J}(t)))$. Then, we can derive

$$\left\|\mathcal{A}_{\alpha,J}\mathsf{L}_\tau^\alpha - \mathsf{L}_\tau^\alpha \mathcal{A}_{\alpha,J}\right\|_2 \le \left\|\mathcal{A}_{\pi/2,J}\mathsf{L}_\tau - \mathsf{L}_\tau \mathcal{A}_{\pi/2,J}\right\|_2. \tag{A.64}$$

When $\alpha_i = \pi/2$ for $i = 1, 2, \ldots, d$, $\mathcal{A}_{\alpha,J}$ is identical to A_J defined in [14], and Lemma E.1 of [14] shows that $\|\mathcal{A}_J\mathsf{L}_\tau - \mathsf{L}_\tau \mathcal{A}_J\|_2 \le C_3\|\nabla\tau\|_\infty$, where $C_3 = \|\nabla\phi\|_1$ is a generic constant that depends on the scaling function $\phi(t)$. Therefore, we can derive the following upper bound of $\|\mathcal{A}_{\alpha,J}\mathsf{L}_\tau^\alpha - \mathsf{L}_\tau^\alpha \mathcal{A}_{\alpha,J}\|_2$ in (A.63), i.e.,

$$\left\|\mathcal{A}_{\alpha,J}\mathsf{L}_\tau^\alpha - \mathsf{L}_\tau^\alpha \mathcal{A}_{\alpha,J}\right\|_2 \le C_3\|\nabla\tau\|_\infty. \tag{A.65}$$

Next, we need to calculate the upper bound of $\|\mathcal{D}_{\alpha,J}\mathsf{L}_\tau^\alpha - \mathsf{L}_\tau^\alpha \mathcal{D}_{\alpha,J}\|_2$ in (A.63). To do so, we derive

$$\begin{aligned}&\left\|\left(\mathcal{D}_{\alpha,J}\mathsf{L}_\tau^\alpha - \mathsf{L}_\tau^\alpha \mathcal{D}_{\alpha,J}\right) f\right\|_2 \\ &= \sum_{p\le J}\sum_{k=1}^{K} \Big\| \big|(\mathsf{L}_\tau \tilde{f}) * \psi_{\lambda_{p,k}}\big| \\ &\quad - \mathsf{L}_\tau\big(\big|\tilde{f} * \psi_{\lambda_{p,k}}\big|\big)e^{-j(\tau(t))^{\mathrm{T}}\Omega_0\left(t-\frac{\tau(t)}{2}\right)}\Big\|_2 \\ &= \sum_{p\le J}\sum_{k=1}^{K} \Big\| \big|(\mathsf{L}_\tau \tilde{f}) * \psi_{\lambda_{p,k}}\big| - \mathsf{L}_\tau\big(\big|\tilde{f} * \psi_{\lambda_{p,k}}\big|\big) \\ &\quad + \mathsf{L}_\tau\big(\big|\tilde{f} * \psi_{\lambda_{p,k}}\big|\big)\left(1 - e^{-j(\tau(t))^{\mathrm{T}}\Omega_0\left(t-\frac{\tau(t)}{2}\right)}\right)\Big\|_2\end{aligned} \tag{A.66}$$

since $\|a+b\|_2 \le \|a\|_2 + \|b\|_2$, then (A.66) gives rise to

$$\begin{aligned}&\left\|\left(\mathcal{D}_{\alpha,J}\mathsf{L}_\tau^\alpha - \mathsf{L}_\tau^\alpha \mathcal{D}_{\alpha,J}\right) f\right\|_2 \\ &\le \sum_{p\le J}\sum_{k=1}^{K} \Big\| \big|(\mathsf{L}_\tau \tilde{f}) * \psi_{\lambda_{p,k}}\big| - \mathsf{L}_\tau\big(\big|\tilde{f} * \psi_{\lambda_{p,k}}\big|\big)\Big\|_2 \\ &\quad + \sum_{p\le J}\sum_{k=1}^{K} \Big\| \mathsf{L}_\tau\big(\big|\tilde{f} * \psi_{\lambda_{p,k}}\big|\big)\left(1 - e^{-j(\tau(t))^{\mathrm{T}}\Omega_0\left(t-\frac{\tau(t)}{2}\right)}\right)\Big\|_2\end{aligned} \tag{A.67}$$

Note that for the first term $\sum_{p\le J}\sum_{k=1}^{K} \big\| \big|(\mathsf{L}_\tau \tilde{f}) * \psi_{\lambda_{p,k}}\big| - \mathsf{L}_\tau\big(\big|\tilde{f} * \psi_{\lambda_{p,k}}\big|\big)\big\|_2$, it could reduce to $\|V_J\mathsf{L}_\tau - \mathsf{L}_\tau V_J\|_2\|f\|_2$ [14], so that, it follows from Lemma 2.14 of [14] that

$$\begin{aligned}&\sum_{p\le J}\sum_{k=1}^{K} \Big\| \big|(\mathsf{L}_\tau \tilde{f}) * \psi_{\lambda_{p,k}}\big| - \mathsf{L}_\tau\big(\big|\tilde{f} * \psi_{\lambda_{p,k}}\big|\big)\Big\|_2 \\ &\qquad \le C_6\left(J\|\nabla\tau\|_\infty + \|\nabla^2\tau\|_\infty\right)\|f\|_2\end{aligned} \tag{A.68}$$

where constant $C_6 = \|\nabla\psi\|_1$. Next, we need to calculate the upper bound of the second term $\sum_{p\le J}\sum_{k=1}^{K} \big\|\mathsf{L}_\tau\big(\big|\tilde{f} * \psi_{\lambda_{p,k}}\big|\big)\left(1 - e^{-j(\tau(t))^{\mathrm{T}}\Omega_0\left(t-\frac{\tau(t)}{2}\right)}\right)\big\|_2$ in (A.67), i.e.,

$$\begin{aligned}&\sum_{p\le J}\sum_{k=1}^{K} \Big\| \mathsf{L}_\tau\big(\big|\tilde{f} * \psi_{\lambda_{p,k}}\big|\big)\left(1 - e^{-j(\tau(t))^{\mathrm{T}}\Omega_0\left(t-\frac{\tau(t)}{2}\right)}\right)\Big\|_2 \\ &= \|1 - e^{-j(\tau(t))^{\mathrm{T}}\Omega_0\left(t-\frac{\tau(t)}{2}\right)}\|_2 \sum_{p\le J}\sum_{k=1}^{K} \Big\| \mathsf{L}_\tau\big(\big|\tilde{f} * \psi_{\lambda_{p,k}}\big|\big)\Big\|_2 \\ &= \|1 - e^{-j(\tau(t))^{\mathrm{T}}\Omega_0\left(t-\frac{\tau(t)}{2}\right)}\|_2 \left\|\mathsf{L}_\tau^\alpha \mathcal{D}_{\alpha,J} f\right\|_2\end{aligned} \tag{A.69}$$

Next, by applying (5), we first compute $\left\|\mathsf{L}_\tau^\alpha \mathcal{D}_{\alpha,J} f\right\|_2$ as

$$\begin{aligned}\left\|\mathsf{L}_\tau^\alpha \mathcal{D}_{\alpha,J} f\right\|_2^2 &= \sum_{p\le J}\sum_{k=1}^{K}\left\|\mathsf{L}_\tau^\alpha\left(\left|f *_\alpha \psi_{\lambda_{p,k}}\right|\right)\right\|_2^2 \\ &= \sum_{p\le J}\sum_{k=1}^{K}\left\|\left|(\mathsf{L}_\tau \tilde{f}) * \psi_{\lambda_{p,k}}\right|\right\|_2^2 \\ &= \sum_{p\le J}\sum_{k=1}^{K}\int_{\mathbb{R}^d}(2\pi)^d\left|\mathcal{F}^\alpha\left\{\mathsf{L}_\tau^\alpha f\right\}(u)\right|^2 \\ &\quad\times\left|\Psi(2^p r_k^{-1}\Omega_{\mathrm{s}} u)\right|_2^2 du\csc\alpha\end{aligned} \tag{A.70}$$

which alongside (19) results in

$$\begin{aligned}\left\|\mathsf{L}_\tau^\alpha \mathcal{D}_{\alpha,J} f\right\|_2^2 &\le (2\pi)^d\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\left|\mathcal{F}^\alpha\left\{\mathsf{L}_\tau^\alpha f\right\}(u)\right|^2 du \\ &= \int_{\mathbb{R}^d}\left|f\left(t-\tau(t)\right)\right|^2 dt.\end{aligned} \tag{A.71}$$

Changing the variance $\xi = t-\tau(t)$ in (A.71) leads to

$$\left\|\mathsf{L}_\tau^\alpha \mathcal{D}_{\alpha,J} f\right\|_2^2 \le \int_{\mathbb{R}^d}\left|f(\xi)\right|^2\frac{1}{\left|\det(Id-\nabla\tau(t))\right|}d\xi \tag{A.72}$$

which alongside $\left|\det(Id-\nabla\tau(t))\right| \ge 1-d\|\nabla\tau\|_\infty$ [54] yields

$$\left\|\mathsf{L}_\tau^\alpha \mathcal{D}_{\alpha,J} f\right\|_2 \le \frac{1}{\sqrt{1-d\|\nabla\tau\|_\infty}}\|f\|_2. \tag{A.73}$$

Next, we need to calculate the upper bound of $\|1-e^{-j(\tau(t))^{\mathrm{T}}\Omega_{\mathrm{o}}(t-\frac{\tau(t)}{2})}\|_2$ in (A.69). It follows that

$$\begin{aligned}&\|1-e^{-j(\tau(t))^{\mathrm{T}}\Omega_{\mathrm{o}}\left(t-\frac{\tau(t)}{2}\right)}\|_2 \\ &= \|e^{-j\frac{t^{\mathrm{T}}\Omega_{\mathrm{o}}t}{2}} - e^{-j\frac{(t-\tau(t))^{\mathrm{T}}\Omega_{\mathrm{o}}(t-\tau(t))}{2}}\|_2.\end{aligned} \tag{A.74}$$

The Taylor expansion of function $e^{-j\frac{(t-\tau(t))^{\mathrm{T}}\Omega_{\mathrm{o}}(t-\tau(t))}{2}}$ in (A.74) at $t_0 = t+\tau(t)$ can be derived as

$$\begin{aligned}e^{-j\frac{(t_0-\tau(t))^{\mathrm{T}}\Omega_{\mathrm{o}}(t_0-\tau(t))}{2}} &= e^{-j\frac{t^{\mathrm{T}}\Omega_{\mathrm{o}}t}{2}} + e^{-j\frac{t^{\mathrm{T}}\Omega_{\mathrm{o}}t}{2}} \\ &\quad\times j\left(\Omega_{\mathrm{o}}t\odot(\mathbf{1}-\nabla\tau)\right)^{\mathrm{T}}\tau(t) + o(-\tau(t))\end{aligned} \tag{A.75}$$

where $\odot$ denotes element-wise multiplication, and $\mathbf{1}\in\mathbb{R}^d$ is the all-ones vector. Combining (A.75) and (A.74) yields

$$\begin{aligned}&\|1-e^{-j(\tau(t))^{\mathrm{T}}\Omega_{\mathrm{o}}\left(t-\frac{\tau(t)}{2}\right)}\|_2 \\ &= \|-je^{-j\frac{t^{\mathrm{T}}\Omega_{\mathrm{o}}t}{2}}\left[\Omega_{\mathrm{o}}t\odot(\mathbf{1}-\nabla\tau)\right]^{\mathrm{T}}\tau(t)\|_2 \\ &= \|-j\Omega_{\mathrm{c}}\left[\Omega_{\mathrm{s}}t\odot(\mathbf{1}-\nabla\tau)\right]^{\mathrm{T}}\tau(t)\|_2\end{aligned} \tag{A.76}$$

where $\Omega_{\mathrm{c}} = \mathrm{diag}(\cos\alpha)$. Since $\mathbf{1}-\nabla\tau$ is not a function of t, by the assumption that $\|\nabla\tau\|_\infty \le \frac{1}{2d}$, we can derive

$$\left|(\mathbf{1}-\nabla\tau)^{\mathrm{T}}\right| \le (1+|\nabla\tau|)^d \le \left(1+\tfrac{1}{2d}\right)^d \tag{A.77}$$

which alongside (A.76) leads to

$$\begin{aligned}&\|1-e^{-j(\tau(t))^{\mathrm{T}}\Omega_{\mathrm{o}}\left(t-\frac{\tau(t)}{2}\right)}\|_2 \\ &\quad\le|\cos\alpha|^d\left(1+\tfrac{1}{2d}\right)^d\|j\Omega_{\mathrm{s}}t\odot\tau(t)\|_2 \\ &\quad\le|\cos\alpha|^d\left(1+\tfrac{1}{2d}\right)^d\|\nabla\hat{\tau}(\Omega_{\mathrm{s}}u)\|_2\end{aligned} \tag{A.78}$$

where $\hat{\tau}(\Omega_{\mathrm{s}}u)$ denotes the FT (with its argument scaled by Ω_{s}) of $\tau(t)$. Then, combining (A.78), (A.73), (A.69), (A.68) and (A.67), (A.66) gives rise to

$$\begin{aligned}&\left\|\mathcal{D}_{\alpha,J}\mathsf{L}_\tau^\alpha - \mathsf{L}_\tau^\alpha\mathcal{D}_{\alpha,J}\right\|_2 \\ &\le C_6\left(J\|\nabla\tau\|_\infty + \|\nabla^2\tau\|_\infty\right) + |\cos\alpha|^d\left(1+\tfrac{1}{2d}\right)^d \\ &\quad\times\|\nabla\hat{\tau}(\Omega_{\mathrm{s}}u)\|_2\frac{1}{\sqrt{1-d\|\nabla\tau\|_\infty}}.\end{aligned} \tag{A.79}$$

It follow from (A.79), (A.65), and (A.63) that

$$\begin{aligned}&\|S^\alpha[\mathbb{L}]\mathsf{L}_\tau^\alpha f - \mathsf{L}_\tau^\alpha S^\alpha[\mathbb{L}]f\|_2 \\ &\le\Bigg(C_7\left(J\|\nabla\tau\|_\infty + \|\nabla^2\tau\|_\infty\right) + |\cos\alpha|^d\left(1+\tfrac{1}{2d}\right)^d \\ &\quad\times\|\nabla\hat{\tau}(\Omega_{\mathrm{s}}u)\|_2\frac{1}{\sqrt{1-d\|\nabla\tau\|_\infty}}\Bigg)\|U^\alpha[\mathbb{L}]f\|_2\end{aligned} \tag{A.80}$$

with constant $C_7 = \max\{\|\nabla\psi\|_1, \|\nabla\phi\|_1\}$. Finally, combining the assumption that $\|\nabla\tau\|_\infty \le \frac{1}{2d}$, (A.80), (A.55), and (A.48) leads to (48). This completes the proof of Theorem 3.

ACKNOWLEDGMENT

The authors would like to thank the associate editor and the anonymous reviewers for their detailed and constructive comments that have helped the presentation of this paper.

REFERENCES

[1] Y. Choi, M. El-Khamy, and J. Lee, "Universal deep neural network compression," *IEEE J. Sel. Top. Signal Process.*, vol. 14, no. 4, pp. 715–726, May 2020.

[2] A. B. Molini, D. Valsesia, G. Fracastor, and E. Magli, "DeepSUM: Deep neural network for super-resolution of unregistered multitemporal images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3644–3656, May 2020.

[3] E. Cha, G. Oh, and J. C. Ye, "Geometric approaches to increase the expressivity of deep neural networks for MR reconstruction," *IEEE J. Sel. Top. Signal Process.*, vol. 14, no. 6, pp. 1292–1305, Oct. 2020.

[4] J. Sulam, A. Aberdam, A. Beck, and M. Elad, "On multi-layer basis pursuit, efficient algorithms and convolutional neural networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 1968–1980, Aug. 2020.

[5] Y. He, X. Dong, G. Kang, Y. Fu, C. Yan, and Y. Yang, "Asymptotic soft filter pruning for deep convolutional neural networks," *IEEE Trans. Cybern.*, vol. 50, no. 8, pp. 3594–3604, Aug. 2020.

[6] K. A. J. Eppenhof, M. W. Lafarge, M. Veta, and J. P. W. Pluim, "Progressively trained convolutional neural networks for deformable image registration," *IEEE Trans. Med. Imag.*, vol. 39, no. 5, pp. 1594–1604, May 2020.

[7] R. A. Farrugia and C. Guillemot, "Light field super-resolution using a low-rank prior and deep convolutional neural networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 5, pp. 1162–1175, May 2020.

[8] P. Chriskos, C. A. Frantzidis, P. T. Gkivogkli, P. D. Bamidis, and C. Kourtidou-Papadeli, "Automatic sleep staging employing convolutional neural networks and cortical connectivity images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 1, pp. 113–123, Jan. 2020.

[9] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[10] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, "Geometric deep learning: Going beyond Euclidean data," *IEEE Signal Process. Mag.*, vol. 34, no. 4, pp. 18–42, Jul. 2017.
[11] W. Li, C. Chen, M. Zhang, H. Li, and Q. Du, "Data augmentation for hyperspectral image classification with deep CNN," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 4, pp. 593–597, Nov. 2018.
[12] S. Mallat, "Understanding deep convolutional networks," *Philos. Trans. Royal Soc. A: Mathematical, Phys. Eng. Sci.*, vol. 374, no. 2065, pp. 1–16, Apr. 2016.
[13] L. Xiao, J. M. Stephen, T. W. Wilson, V. D. Calhoun, and Y.-P. Wang, "A manifold regularized multi-task learning model for IQ prediction from two fMRI paradigms," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 3, pp. 796–806, Mar. 2020.
[14] S. Mallat, "Group invariant scattering," *Commun. Pure Appl. Math.*, vol. 65, no. 10, pp. 1331–1398, Oct. 2012.
[15] J. Bruna and S. Mallat, "Invariant scattering convolution networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1872–1886, Aug. 2013.
[16] J. Andén and S. Mallat, "Deep scattering spectrum," *IEEE Trans. Signal Process.*, vol. 62, no. 16, pp. 4114–4128, Aug. 2014.
[17] J. Andén, V. Lostanlen, and S. Mallat, "Joint time-frequency scattering," *IEEE Trans. Signal Process.*, vol. 67, no. 14, pp. 3704–3718, Jul. 2019.
[18] L. Sifre and S. Mallat, "Rigid-motion scattering for texture classification," Mar. 2014, *arXiv:1403.1687*.
[19] R. Leonarduzzi, H. Liu, and Y. Wang, "Scattering transform and sparse linear classifier for art authentication," *Signal Process.*, vol. 150, pp. 11–19, Mar. 2018.
[20] E. Allys *et al.*, "The RWST, a comprehensive statistical description of the non-gaussian structures in the ISM," *Astron. Astrophys.*, vol. 629, no. A115, pp. 1–21, Sep. 2019.
[21] M. Hirn, S. Mallat, and N. Poilvert, "Wavelet scattering regression of quantum chemical energies," *Multiscale Model. Simul.*, vol. 15, no. 2, pp. 827–863, May 2017.
[22] V. Chudáček, J. Andén, S. Mallat, P. Abry, and M. Doret, "Scattering transform for intrapartum fetal heart rate variability fractal analysis: A case-control study," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 4, pp. 1100–1108, Apr. 2014.
[23] E. Oyallon *et al.*, "Scattering networks for hybrid representation learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 9, pp. 2208–2221, Sep. 2019.
[24] W. Czaja and W. Li, "Analysis of time-frequency scattering transforms," *Appl. Comput. Harmon. Anal.*, vol. 47, pp. 149–171, 2019.
[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Adv. Neural Inform. Process. Syst.*, vol. 25, no. 2, pp. 1097–1105, Jan. 2012.
[26] C. Gargour, M. Gabrea, V. Ramachandran, and J. Lina, "A short introduction to wavelets and their applications," *IEEE Circuits Syst. Mag.*, vol. 9, no. 2, pp. 57–68, Apr.–Jun. 2009.
[27] H. M. Ozaktas, Z. Zalevsky, and M. A. Kutay, *The Fractional Fourier Transform With Applications in Optics and Signal Processing*. New York, NY, USA: Wiley, 2000.
[28] M. S. Pattichis and A. C. Bovik, "Analyzing image structure by multidimensional frequency modulation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 5, pp. 753–766, May 2007.
[29] I. Kokkinos, G. Evangelopoulos, and P. Maragos, "Texture analysis and segmentation using modulation features, generative models, and weighted curve evolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 1, pp. 142–157, Jan. 2009.
[30] K. P. Constantinou, I. P. Constantinou, C. S. Pattichis, and M. S. Pattichis, "Medical image analysis using AM-FM models and methods," *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 270–289, Jan. 2021.
[31] Y. Guo and B.-Z. Li, "Novel method for parameter estimation of newton's rings based on CFRFT and ER-WCA," *Signal Process.*, vol. 144, pp. 118–126, Mar. 2018.
[32] C. Liu, C. Tang, M. Xu, and Z. Lei, "Binarization of ESPI fringe patterns based on an m-net convolutional neural network," *Appl. Opt.*, vol. 59, no. 30, pp. 9598–9606, 2020.
[33] M. F. Lu *et al.*, "Parameter estimation of optical fringes with quadratic phase using the fractional Fourier transform," *Opt. Lasers Eng.*, vol. 74, pp. 1–16, Nov. 2015.
[34] L. Liu, J. Wu, D. Li, L. Senhadji, and H. Shu, "Fractional wavelet scattering network and applications," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 2, pp. 553–563, Feb. 2019.
[35] J. Shi, N. Zhang, and X. Liu, "A novel fractional wavelet transform and its applications," *Sci. China Inf. Sci.*, vol. 55, no. 6, pp. 1270–1279, Jun. 2012.
[36] J. Shi, X. Liu, and N. Zhang, "Multiresolution analysis and orthogonal wavelets associated with fractional wavelet transform," *Signal, Image, Video Process.*, vol. 9, no. 1, pp. 211–220, Aug. 2015.
[37] J. Shi, X. Liu, X. Sha, Q. Zhang, and N. Zhang, "A sampling theorem for fractional wavelet transform with error estimates," *IEEE Trans. Signal Process.*, vol. 65. no. 18, pp. 4797–4811, Sep. 2017.
[38] X.-Z. Zhang, B. W.-K. Ling, H. H. Dam, K.-L. Teo, and C. Wu, "Optimal joint design of discrete fractional Fourier transform matrices and mask coefficients for multichannel filtering in fractional Fourier domains," *IEEE Trans. Signal Process.*, vol. 66, no, 22, pp. 6016–6030, Nov. 2018.
[39] X.-Z. Zhang *et al.*, "Optimal design of orders of discrete fractional Fourier transforms for sparse representations," *IET Signal Process.*, vol. 12, no. 8, pp. 1023–1033, Jul. 2018.
[40] H. Miao, F. Zhang, and R. Tao, "Fractional Fourier analysis using the möbius inversion formula," *IEEE Trans. Signal Process.*, vol. 67. no. 12, pp. 3181–3196, Jun. 2019.
[41] J. Shi, X. Sha, X. Song, and N. Zhang, "Generalized convolution theorem associated with fractional Fourier transform," *Wireless Commun. Mob. Comput.*, vol. 14, no. 13, pp. 1340–1351, Sep. 2014.
[42] J. B. Diaz and F. T. Metcalf, "An analytic proof of young's inequality," *Amer. Math. Monthly*, vol. 77, no. 6, pp. 603–609, Jun. 1970.
[43] C. K. Chui and X. Shi, "On a littlewood-paley identity and characterization of wavelets," *J. Math. Anal. Appl.*, vol. 177, no.2, pp. 608–626, Aug. 1993.
[44] I. Waldspurger, A. D'Aspremont, and S. Mallat, "Phase recovery, MaxCut and complex semidefinite programming," *Math. Program., Ser. A*, vol. 149, no. 1-2, pp. 47–81, Feb. 2015.
[45] L. B. Almeida, "The fractional fourier transform and time-frequency representations," *IEEE Trans. Signal Process.*, vol. 42, no. 11, pp. 3084–3091, Nov. 1994.
[46] H. M. Ozaktas, Z. Zalevsky, and M. A. Kutay, *The Fractional Fourier Transform With Applications in Optics and Signal Processing*. New York, NY, USA: Wiley, 2000.
[47] J. J. Healy, M. A. Kutay, H. M. Ozaktas, and J. T. Sheridan, *Linear Canonical Transforms: Theory and Applications*. New York, NY, USA: Springer-Verlag, 2016.
[48] V. C. Chen and M. Martorella, *Inverse Synthetic Aperture Radar Imaging: Principles, Algorithms and Applications*. Edison, NJ, USA: SciTech, 2014.
[49] C. Zhang, J. Shi, Z. Zhang, Y. Liu, and X. Fu, "FRFT-based interference suppression for OFDM systems in IoT environment," *IEEE Commun. Lett.*, vol. 23, no. 11, pp. 2068–2072, 2019.
[50] R. Tao, F. Zhang, and Y. Wang, "Fractional power spectrum," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4199–4206, Sep. 2008.
[51] I. Kononenko, E. Šimec, and M. Robnik-Šikonja, "Overcoming the myopia of inductive learning algorithms with RELIEFF," *Appl. Intell.*, vol. 7, no. 1, pp. 39–55, Jan. 1997.
[52] UMD HR Texture Dataset. [Online]. Available: http://legacydirs.umiacs.umd.edu/∼fer/website-texture/texture.htm#
[53] Amsterdam Library of Object Images (ALOI). [Online]. Available: http://aloi.science.uva.nl/
[54] R. P. Brent, J.-A. H. Osborn, and W. D. Smith, "Note on best possible bounds for determinants of matrices close to the identity matrix," *Linear Algebra Appl.*, vol. 466, pp. 21–26, Feb. 2015.
[55] K. E. Muller, "Computing the confluent hypergeometric function, $M(a, b, x)$" *Numer. Math.*, vol. 90, no. 1, pp. 179–196, Nov. 2001.

Jun Shi received the B.E. degree in communication engineering from Hohai University, Nanjing, China, in 2006, and the M.E. and Ph.D. degrees in information and communication engineering from the Harbin Institute of Technology (HIT), Harbin, China, in 2008 and 2013, respectively.

From September 2011 to March 2012, he was a Visiting Scholar with the Department of Electrical and Computer Engineering, University of Delaware, Newark, DE, USA. From September 2015 to September 2016, he was also a Visiting Scholar with the Electrical Engineering Department, University of California, Los Angeles, CA, USA. He is currently an Associate Professor with Communication Research Center, HIT. His current research interests include computational neurosciences, time-frequency analysis, sampling and approximation theory, integral transforms, and signal processing for wireless communications and radar.

Yanan Zhao received the B.E. degree in communication engineering and the M.E. degree in information and communication engineering from the Harbin Institute of Technology, Weihai, China, in 2018 and 2020, respectively.

Her current research interests include fractional wavelet transform, sampling theory, compressive sensing, and machine learning.

Wei Xiang (Senior Member, IEEE) received the B.Eng. and M.Eng. degrees in electronic engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 1997 and 2000, respectively, and the Ph.D. degree in telecommunications engineering from the University of South Australia, Adelaide, SA, Australia, in 2004.

He is currently the Cisco Chair of AI and Internet of Things with La Trobe University, Australia. Previously, he was the Founding Chair and the Head of Discipline of Internet of Things Engineering with James Cook University, Cairns, Australia. Due to his instrumental leadership in establishing Australia's first accredited Internet of Things Engineering degree program, he was selected into Pearcy Foundation's Hall of Fame in October 2018. He has authored or coauthored more than 250 peer-reviewed papers, including three academic books and 180 journal articles. His research interests include communications and information theory, particularly the Internet of Things, and coding and signal processing for multimedia communications systems. He is an Elected Fellow of the IET in U.K. and Engineers Australia. He was the recipient of the TNQ Innovation Award in 2016, and Pearcey Entrepreneurship Award in 2017, and Engineers Australia Cairns Engineer of the Year in 2017. He was the co-recipient of the four Best Paper Awards at WiSATS'2019, WCSP'2015, IEEE WCNC'2011, and ICWMC'2009. He has been awarded several prestigious fellowship titles. He was named a Queensland International Fellow during 2010–2011 by the Queensland Government of Australia, an Endeavour Research Fellow during 2012–2013 by the Commonwealth Government of Australia, a Smart Futures Fellow during 2012–2015 by the Queensland Government of Australia, and a JSPS Invitational Fellow jointly by the Australian Academy of Science and Japanese Society for Promotion of Science during 2014–2015. He is the Vice Chair of the IEEE Northern Australia Section. During 2015–2017, he was the Editor of the IEEE COMMUNICATIONS LETTERS, and is an Associate Editor for the IEEE ACCESS and Springer's *Telecommunications Systems*. He has served in a large number of international conferences in the capacity of the General Co-Chair, TPC Co-Chair, and Symposium Chair.

Vishal Monga (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the Department of Electrical and Computer Engineering, University of Texas, Austin, TX, USA. Since Fall 2009, he has been with Electrical Engineering and Computer Science Faculty, Pennsylvania State University, University Park, PA, USA. From October 2005 to July 2009, he was an Imaging Scientist with Xerox Research Labs. He has also been a Visiting Researcher with Microsoft Research, Redmond, WA, USA, and a Visiting Faculty with the University of Rochester, Rochester, NY, USA. His research interests include optimization-based methods with applications in signal and image processing, learning, and computer vision. He was an Elected Member of the IEEE Image Video and Multidimensional Signal Processing Technical Committee (2017C2019) and has served on the Editorial Boards of the IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE SIGNAL PROCESSING LETTERS, and IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY. He is currently a Senior Area Editor of the IEEE SIGNAL PROCESSING LETTERS. He was the recipient of the U.S. National Science Foundation CAREER Award and 2016 Joel and Ruth Spira Teaching Excellence Award.

Xiaoping Liu received the B.E. degree in automation from Hohai University, Nanjing, China, in 2007, and the M.E. degree in control science and the Ph.D. degree in information and communication engineering from the Harbin Institute of Technology, Harbin, China, in 2009 and 2019, respectively.

From 2009 to 2011, she was a Research Assistant with Radar and Electronic Equipment Research Institute, China Aviation Industry Corporation I, Suzhou, China. From October 2014 to September 2016, she was a Visiting Scholar with the Electrical Engineering Department, University of California, Los Angeles, CA, USA. She is currently a Postdoctoral Researcher with the Beijing Key Laboratory of Fractional Signals and Systems, School of Information and Electronics, Beijing Institute of Technology, Beijing, China. Her current research interests include sampling theory, digital signal processing, and time-frequency analysis.

Ran Tao (Senior Member, IEEE) received the B.S. degree from the Electronic Engineering Institute of PLA, Hefei, China, in 1985, and the M.S. and Ph.D. degrees from the Harbin Institute of Technology, Harbin, China, in 1990 and 1993, respectively.

In 2001, he was a Senior Visiting Scholar with the University of Michigan, Ann Arbor, MI, USA. He is currently a Professor with the School of Information and Electronics, Beijing Institute of Technology, Beijing, China. He has authored or coauthored three books and more than 100 peer-reviewed journal articles. His current research interests include fractional Fourier transform and its applications, theory, and technology for radar and communication systems.

He is a Member of the Wireless Communication and Signal Processing Commission of International Union of Radio Science (URSI). He was the recipient of the National Science Foundation of China for Distinguished Young Scholars in 2006, and the First Prize of Science and Technology Progress in 2006 and 2007, and the First Prize of Natural Science in 2013, both awarded by the Ministry of Education. He was a Distinguished Professor of the Changjiang Scholars Program in 2009. From 2010 to 2012, he was a Chief Professor of the Program for Changjiang Scholars and Innovative Research Team in University. Since 2014, he has been a Chief Professor of the Creative Research Groups of the National Natural Science Foundation of China. He is currently the Vice Chair of the IEEE China Council and the URSI China Council.