Next, we show that, after training with ImageNet, the feature extractor formed by early layers of popular CNN architectures can approximately be reformulated as a stack of $\mathbb{R}$GPool operators. Our framework therefore provides a theoretical grounding to study these networks.

We apply our theoretical results on the dual-tree complex wavelet packet transform (DT-$\mathbb{C}$WPT), a particular case of discrete Gabor-like decomposition with perfect reconstruction properties [29], [30], possessing characteristics comparable to standard convolution layers. Finally, we verify our predictions on a deterministic setting based on DT-$\mathbb{C}$WPT. Given an input image, we compute the mean discrepancy between the outputs of $\mathbb{C}$GMod and $\mathbb{R}$GPool, for each wavelet packet filter.[1] We then observe that shift invariance, when measured on $\mathbb{R}$GPool feature maps, is nearly achieved if they remain close to $\mathbb{C}$GMod outputs. We therefore establish an invariance validity domain for $\mathbb{R}$GPool operators.

Prior to this work, we presented a preliminary study [31], where we experimentally showed that an operator based on DT-$\mathbb{C}$WPT can mimic the behavior of the first convolution layer with fewer parameters, while keeping the network's predictive power. Our model was solely based on real-valued filters,[2] which are known to be generally unstable [32]. Yet, we observed a limited but genuine form of shift invariance, compared to other models based on the standard, non-analytic wavelet packet transform. At the same time, we became aware of a preliminary work in Waldspurger's PhD thesis [28, pp. 190–191], suggesting a potential connection between the combinations "real wavelet transform + max pooling" on the one hand and "complex wavelet transform + modulus" on the other hand. Following this idea, we decided to study whether invariance properties of complex moduli could somehow be captured by the max pooling operator. As shown in the present paper, Waldspurger's work does not fully extend to discrete and subsampled convolutions. We address this issue by adopting a probabilistic point of view.

II. Related Work

A. Wavelet Scattering Networks

These models, introduced by Bruna and Mallat [11], compute cascading wavelet convolutions followed by non-linear operations. They produce translation-invariant image representations which are stable to deformation and preserve high-frequency information. A variation has been proposed in [33] to improve stability with respect to small rotations. Wavelet scattering networks were later adapted to the discrete framework using the dual-tree complex wavelet transform [34], as well as functions defined on graphs [35].

Such networks, which are totally deterministic aside from the output classifier, achieve results on small image datasets but do not scale well to more complex ones. According to Oyallon *et al.* [36], [37], this is partly due to non-geometric sources of variability within classes. Instead, the authors proposed to use scattering coefficients as inputs to a CNN, showing that the network complexity can be reduced while keeping competitive performance. More recent work by Zarka *et al.* [38] proposed to sparsify wavelet scattering coefficients by learning a dictionary matrix, and managed to outperform AlexNet [10]. This was extended by the same team in [39], where the authors proposed to learn 1×1 convolutions between feature maps of scattering coefficients and to apply soft-thresholding to reduce within-class variability. This model reached the classification accuracy of ResNet-18 on ImageNet.

Other work proposed architectures in which the scattering transform is no longer deterministic. Cotter and Kingsbury [40] built a learnable scattering network. In this model, feature maps of scattering coefficients are mixed together using trainable weights, to account for cross-channel filtering as implemented in CNNs. Their architecture outperformed VGG networks on small image datasets. Recently, Gauthier *et al.* [41] introduced parametric scattering networks, in which the scale, orientation and aspect ratio of each wavelet filter are adjusted during training. Their approach has proven successful when trained on limited dataset.

All these papers are driven by the purpose of building ad-hoc CNN-like feature extractors, implementing well defined mathematical operators specifically designed to meet a certain number of desired properties. By contrast, our work seeks evidence that such properties, which have been established for wavelet scattering networks, are—to some extent—embedded in *existing* CNN architectures, with no need to alter their behavior or introduce new features.

B. Invariance Studies in CNNs

Several papers analyze invariance properties in CNN-related feature extractors, including—but not limited to—wavelet scattering networks. Whereas extensive studies related to the original architecture are proposed by Mallat in [42], [43], more recent work tackle the question for various extensions of the model. In [44], [45], scattering networks based on uniform covering frames—i.e., frames splitting the frequency domain into windows of roughly equal size, much like Gabor frames—are studied. Besides, [8] considers a wide variety of feature extractors involving convolutions, Lipschitz-continuous non-linearities and pooling operators. The paper shows that outputs become more translation invariant with increasing network depth. Finally, [46] shows that certain classes of CNNs are contained into the reproducing kernel Hilbert space (RKHS) of a multilayer convolutional kernel representation. As such, stability metrics are estimated, based on the RKHS norm which is difficult to control in practice.

In these studies, invariance properties are obtained for continuous signals. Whereas real-life CNNs can be mathematically described in the continuous framework, feature maps computed at their hidden and output layers are actually discrete sequences, which can be recovered by sampling the continuous signals. At each convolution and pooling layers, the sampling interval is increased (subsampling), resulting in a loss of information. Unfortunately, this may greatly affect shift invariance, as explained in section I. The current paper specifically addresses this issue.

[1] DT-$\mathbb{C}$WPT paves the Fourier domain into square regions of identical size, each of them associated to a specific filter.

[2] To do so, we split the real and imaginary parts of the original filters.

III. Shift Invariance of Operators

The goal of this section is to theoretically establish conditions for near-shift invariance at the output of the first max pooling layer. We start by proving shift invariance of $\mathbb{C}$GMod operators. Then, we establish conditions under which $\mathbb{R}$GPool and $\mathbb{C}$GMod produce closely related outputs. Finally, we derive a probabilistic measure of shift invariance for $\mathbb{R}$GPool.

A. Notations

The complex conjugate of any number $z \in \mathbb{C}$ is denoted by z^*. For any $p \in \mathbb{R}_+^* \cup \{\infty\}$, $\boldsymbol{x} \in \mathbb{R}^2$ and $r \in \mathbb{R}_+$, we denote by $\mathcal{B}_p(\boldsymbol{x},\, r) \subset \mathbb{R}^2$ the closed l^p-ball with center $\boldsymbol{x}$ and radius r. When $\boldsymbol{x} = \boldsymbol{0}$, we write $\mathcal{B}_p(r)$.

Continuous Framework: Given $p > 0$ and a measurable subset of $\mathbb{R}$ or $\mathbb{R}^2$ denoted by E, we consider $L^p(E)$ as the space of measurable complex-valued functions $f : E \to \mathbb{C}$ such that $\|f\|_{L^p} := \int_E |f(x)|^p \,\mathrm{d}x < +\infty$. Whenever we talk about equality in $L^p(E)$ or inclusion in E, it shall be understood as "almost everywhere with respect to the Lebesgue measure". Besides, we denote by $L^2_{\mathbb{R}}(\mathbb{R}^2) \subset L^2(\mathbb{R}^2)$ the subset of real-valued functions. For any $f \in L^2(\mathbb{R}^2)$, $\overline{f}$ denotes its flipped version: $\overline{f}(\boldsymbol{x}) := f(-\boldsymbol{x})$.

The 2D Fourier transform of any $f \in L^2(\mathbb{R}^2)$ is denoted by $\widehat{f} \in L^2(\mathbb{R}^2)$, such that

$$\forall \boldsymbol{\nu} \in \mathbb{R}^2,\ \widehat{f}(\boldsymbol{\nu}) := \iint_{\mathbb{R}^2} f(\boldsymbol{x}) \mathrm{e}^{-i\langle \boldsymbol{\nu},\, \boldsymbol{x}\rangle} \,\mathrm{d}^2\boldsymbol{x}. \tag{1}$$

For any $\varepsilon > 0$ and $\boldsymbol{\nu} \in \mathbb{R}^2$, we denote by $\mathcal{V}\big(\boldsymbol{\nu},\, \varepsilon\big) \subset L^2(\mathbb{R}^2)$ the set of functions whose Fourier transform is supported in a square region of size $\varepsilon \times \varepsilon$ centered in $\boldsymbol{\nu}$:

$$\mathcal{V}\big(\boldsymbol{\nu},\, \varepsilon\big) := \Big\{ \psi \in L^2(\mathbb{R}^2) \;\Big|\; \operatorname{supp} \widehat{\psi} \subset \mathcal{B}_\infty(\boldsymbol{\nu},\, \varepsilon/2) \Big\}. \tag{2}$$

For any $\boldsymbol{h} \in \mathbb{R}^2$, we also consider the translation operator, denoted by $\mathcal{T}_{\boldsymbol{h}}$, defined by $\mathcal{T}_{\boldsymbol{h}} f : \boldsymbol{x} \mapsto f(\boldsymbol{x} - \boldsymbol{h})$.

Discrete Framework: We consider $l^2(\mathbb{Z}^d)$ as the space of d-dimensional sequences $\mathrm{X} \in \mathbb{C}^{\mathbb{Z}^d}$ such that $\|\mathrm{X}\|_2^2 := \sum_{\boldsymbol{n} \in \mathbb{Z}^d} \big|\mathrm{X}[\boldsymbol{n}]\big|^2 < +\infty$. Indexing is made between square brackets: $\forall \mathrm{X} \in l^2(\mathbb{Z}^d)$, $\forall \boldsymbol{n} \in \mathbb{Z}^d$, $\mathrm{X}[\boldsymbol{n}] \in \mathbb{C}$, and we denote by $l^2_{\mathbb{R}}(\mathbb{Z}^d) \subset l^2(\mathbb{Z}^d)$ the subset of real-valued sequences. For any $\mathrm{X} \in l^2(\mathbb{Z}^d)$, $\overline{\mathrm{X}}$ denotes its flipped version: $\overline{\mathrm{X}}[\boldsymbol{n}] := \mathrm{X}[-\boldsymbol{n}]$. The subsampling operator is denoted by $\downarrow$: for any $\mathrm{X} \in l^2(\mathbb{Z}^d)$ and any $m \in \mathbb{N}^*$, $(\mathrm{X} \downarrow m)[\boldsymbol{n}] := \mathrm{X}[m\boldsymbol{n}]$.

2D images, feature maps and convolution kernels are considered as elements of $l^2(\mathbb{Z}^2)$, and are denoted by straight capital letters. Besides, arrays of 2D sequences are denoted by bold straight capital letters, for instance: $\mathbf{X} = (\mathrm{X}_k)_{k \in \{0..K-1\}}$. Note that indexing starts at 0 to comply with practical implementations. We will also consider 1D sequences $\mathrm{x} \in l^2(\mathbb{Z})$, denoted by straight lower-case letters.

The 2D discrete-time Fourier transform of any $\mathrm{X} \in l^2(\mathbb{Z}^2)$ is denoted by $\widehat{\mathrm{X}} \in L^2([-\pi,\, \pi]^2)$, such that

$$\forall \boldsymbol{\xi} \in [-\pi,\, \pi]^2,\ \widehat{\mathrm{X}}(\boldsymbol{\xi}) := \sum_{\boldsymbol{n} \in \mathbb{Z}^2} \mathrm{X}[\boldsymbol{n}] \mathrm{e}^{-i\langle \boldsymbol{\xi},\, \boldsymbol{n}\rangle}. \tag{3}$$

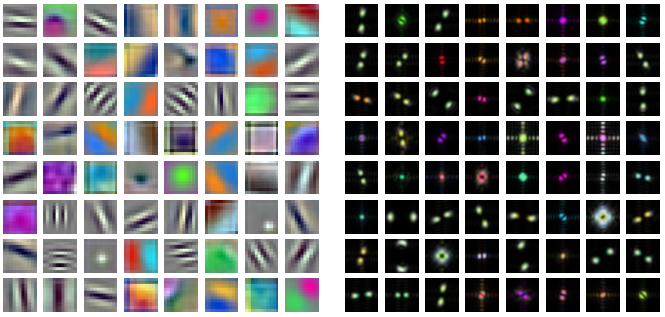

Fig. 1. Spatial (left) and Fourier (right) representations of convolution kernels in the first layer of AlexNet, after training with ImageNet ILSVRC2012. Each kernel connects the 3 RGB input channels to one of the 64 output channels.

For any $\kappa \in \,]0,\, 2\pi]$ and $\boldsymbol{\xi} \in \mathcal{B}_\infty(\pi)$, we denote by $\mathcal{G}\big(\boldsymbol{\xi},\, \kappa\big) \subset l^2(\mathbb{Z}^2)$ the set of 2D sequences whose Fourier transform is supported in a square region of size $\kappa \times \kappa$ centered in $\boldsymbol{\xi}$:

$$\mathcal{G}\big(\boldsymbol{\xi},\, \kappa\big) := \Big\{ \mathrm{W} \in l^2(\mathbb{Z}^2) \;\Big|\; \operatorname{supp} \widehat{\mathrm{W}} \subset \mathcal{B}_\infty(\boldsymbol{\xi},\, \kappa/2) \Big\}. \tag{4}$$

Remark 1: The support $\mathcal{B}_\infty(\boldsymbol{\xi},\, \kappa/2)$ actually lives in the quotient space $[-\pi,\, \pi]^2 / (2\pi\mathbb{Z}^2)$. Consequently, when $\boldsymbol{\xi}$ is close to an edge, a fraction of this region is located at the far end of the frequency domain. From now on, the choice of $\boldsymbol{\xi}$ and κ is implicitly assumed to avoid such a situation.

B. Intuition

In many CNNs for computer vision, input images are first transformed through subsampled—or strided—convolutions. For instance, in AlexNet, convolution kernels are of size 11×11 and the subsampling factor is equal to 4. Fig. 1 displays the corresponding kernels after training with ImageNet. This linear transform is generally followed by rectified linear unit (ReLU) and max pooling.

We can observe that many kernels display oscillating patterns with well-defined orientations. We denote by $\mathrm{V} \in l^2_{\mathbb{R}}(\mathbb{Z}^2)$ one of these "well-behaved" filters. Its Fourier spectrum roughly consists in two bright spots which are symmetric with respect to the origin.[3] Now, we consider a complex-valued companion $\mathrm{W} \in l^2(\mathbb{Z}^2)$ such that, for any $\boldsymbol{\xi} = (\xi_1,\, \xi_2) \in [-\pi,\, \pi]^2$,

$$\widehat{\mathrm{W}}(\boldsymbol{\xi}) := \big(1 + \operatorname{sgn} \xi_1\big) \cdot \widehat{\mathrm{V}}(\boldsymbol{\xi}). \tag{5}$$

We can show that V is the real part of W, and that $\mathrm{W} = \mathrm{V} + i\mathcal{H}(\mathrm{V})$, where $\mathcal{H}$ denotes the two-dimensional Hilbert transform as introduced in [47].[4] As a consequence, $\widehat{\mathrm{W}}$ is equal to $2\widehat{\mathrm{V}}$ on one half of the Fourier domain, and 0 on the other half. Therefore, only one bright spot remains in the spectrum. It turns out that such complex filters with high frequency resolution produce stable signal representations, as we will see in section III-C. In the subsequent sections, we then wonder whether this property is kept when considering the max pooling of real-valued convolutions.

[3] Actually, the Fourier transform of any real-valued sequence is centrally symmetric: $\widehat{\mathrm{V}}(-\boldsymbol{\xi}) = \widehat{\mathrm{V}}(\boldsymbol{\xi})^*$. The specificity of well-oriented filters lies in the concentration of their power spectrum around two precise locations.

[4] $\mathcal{H}(\mathrm{V})$ is defined such that $\widehat{\mathcal{H}(\mathrm{V})}(\boldsymbol{\xi}) := -i \operatorname{sgn}(\xi_1) \widehat{\mathrm{V}}(\boldsymbol{\xi})$.

In what follows, W will be referred to as a discrete Gabor-like filter, and the coefficients resulting from the convolution with W will be referred to as discrete Gabor-like coefficients.

C. Shift Invariance of $\mathbb{C}$GMod Outputs

The aim of this section is to show that the modulus of discrete Gabor-like coefficients—i.e., the output of a $\mathbb{C}$GMod operator such as introduced in section I-B—is nearly shift-invariant (the meaning of shift invariance will be clarified). This result is hinted in [32] but not formally proven.

1) Continuous Framework: We introduce several results regarding functions defined on the continuous space $\mathbb{R}^2$. Near-shift invariance on discrete 2D sequences will then be derived from these results by taking advantage of sampling theorems. Lemma 1 below is adapted from [28, pp. 190–191].

Lemma 1: Given $\varepsilon > 0$ and $\boldsymbol{\nu} \in \mathbb{R}^2$, let $\psi \in \mathcal{V}\big(\boldsymbol{\nu},\, \varepsilon\big)$ denote a complex-valued filter such as defined in (2). Now, for any real-valued function $f \in L^2_{\mathbb{R}}(\mathbb{R}^2)$, we consider the complex-valued function $f_0 \in L^2(\mathbb{R}^2)$ defined by

$$f_0 : \boldsymbol{x} \mapsto (f * \overline{\psi})(\boldsymbol{x})\, \mathrm{e}^{i\langle \boldsymbol{\nu},\, \boldsymbol{x}\rangle}. \tag{6}$$

Then f_0 is low-frequency, with $\operatorname{supp} \widehat{f_0} \subset \mathcal{B}_\infty(\varepsilon/2)$.

Proof: See Appendix A. ∎

On the other hand, the following proposition provides a shift invariance bound for low-frequency functions such as introduced above.

Proposition 1: For any $f_0 \in L^2_{\mathbb{R}}(\mathbb{R}^2)$ such that $\operatorname{supp} \widehat{f_0} \subset \mathcal{B}_\infty(\varepsilon/2)$, and any $\boldsymbol{h} \in \mathbb{R}^2$ satisfying $\|\boldsymbol{h}\|_1 \leq \pi/\varepsilon$,

$$\|\mathcal{T}_{\boldsymbol{h}} f_0 - f_0\|_{L^2} \leq \alpha(\varepsilon \boldsymbol{h})\, \|f_0\|_{L^2}\,, \tag{7}$$

where we have defined

$$\alpha : \boldsymbol{\tau} \mapsto \frac{\|\boldsymbol{\tau}\|_1}{2}. \tag{8}$$

Proof: See Appendix B. ∎

2) Adaptation to Discrete 2D Sequences: Given $\kappa \in \left]0,\, 2\pi\right]$ and $\boldsymbol{\xi} \in \mathcal{B}_\infty(\pi)$, let $\mathrm{W} \in \mathcal{G}\big(\boldsymbol{\xi},\, \kappa\big)$ denote a discrete Gabor-like filter such as defined in (4). For any image $\mathrm{X} \in l^2_{\mathbb{R}}(\mathbb{Z}^2)$ with finite support and any subsampling factor $m \in \mathbb{N}^*$, we express $(\mathrm{X} * \overline{\mathrm{W}}) \downarrow m$ using the continuous framework introduced above, and derive an invariance formula.

For any sampling interval $s \in \mathbb{R}_+^*$, let $\mathcal{U}_s$ denote the space of 2D functions $g \in L^2(\mathbb{R}^2)$ such that the support of $\widehat{g}$ is included in $\mathcal{B}_\infty(\pi/s)$.[5] We consider the following lemma.

Lemma 2: Let $s > 0$. For any $g \in \mathcal{U}_s$ and any $\boldsymbol{\omega} \in \mathcal{B}_\infty(\pi/s)$, we have

$$\widehat{g}(\boldsymbol{\omega}) = s\, \widehat{\mathrm{Y}}(s\boldsymbol{\omega}), \tag{9}$$

where $\mathrm{Y} \in l^2(\mathbb{Z}^2)$ is defined such that $\mathrm{Y}[\boldsymbol{n}] := s\, g(s\boldsymbol{n})$, for any $\boldsymbol{n} \in \mathbb{Z}^2$. Besides, we have the following norm equality:

$$\|g\|_{L^2} = \|\mathrm{Y}\|_2\,. \tag{10}$$

Proof: See Appendix C. ∎

[5] Using the notation introduced in (2), we have $\mathcal{U}_s = \mathcal{V}(\boldsymbol{0},\, 2\pi/s)$.

We now consider $\phi^{(s)} \in L^2_{\mathbb{R}}(\mathbb{R}^2)$ such that $\widehat{\phi^{(s)}} := s\mathbb{1}_{\mathcal{B}_\infty(\pi/s)}$.[6] For any $\boldsymbol{n} \in \mathbb{Z}^2$, we denote by $\phi^{(s)}_{\boldsymbol{n}} := \mathcal{T}_{s\boldsymbol{n}}\phi^{(s)}$ a shifted version of $\phi^{(s)}$. According to Theorem 3.5 in [48, p. 68], $\{\phi^{(s)}_{\boldsymbol{n}}\}_{\boldsymbol{n}\in\mathbb{Z}^2}$ is an orthonormal basis of $\mathcal{U}_s$.[7] We then get the following proposition, which draws a bond between the discrete and continuous frameworks.

Proposition 2: Let $\mathrm{X} \in l^2_{\mathbb{R}}(\mathbb{Z}^2)$ denote an input image with finite support, and $\mathrm{W} \in \mathcal{G}\big(\boldsymbol{\xi},\, \kappa\big)$. Considering a sampling interval $s \in \mathbb{R}_+^*$, we define $f_{\mathrm{X}} \in L^2_{\mathbb{R}}(\mathbb{R}^2)$ and $\psi_{\mathrm{W}} \in L^2(\mathbb{R}^2)$ such that

$$f_{\mathrm{X}} := \sum_{\boldsymbol{n}\in\mathbb{Z}^2} \mathrm{X}[\boldsymbol{n}]\, \phi^{(s)}_{\boldsymbol{n}} \quad \text{and} \quad \psi_{\mathrm{W}} := \sum_{\boldsymbol{n}\in\mathbb{Z}^2} \mathrm{W}[\boldsymbol{n}]\, \phi^{(s)}_{\boldsymbol{n}}. \tag{11}$$

Then, $\psi_{\mathrm{W}} \in \mathcal{V}\big(\boldsymbol{\xi}/s,\, \kappa/s\big)$. Moreover, for all $\boldsymbol{n} \in \mathbb{Z}$,

$$\mathrm{X}[\boldsymbol{n}] = s\, f_{\mathrm{X}}(s\boldsymbol{n}); \qquad \mathrm{W}[\boldsymbol{n}] = s\, \psi_{\mathrm{W}}(s\boldsymbol{n}), \tag{12}$$

and, for a given subsampling factor $m \in \mathbb{N}^*$,

$$\Big((\mathrm{X} * \overline{\mathrm{W}}) \downarrow m\Big)[\boldsymbol{n}] = \big(f_{\mathrm{X}} * \overline{\psi}_{\mathrm{W}}\big)\,(ms\boldsymbol{n})\,. \tag{13}$$

Proof: See Appendix D. ∎

Proposition 2 introduces a latent subspace of $L^2_{\mathbb{R}}(\mathbb{R}^2)$ from which input images are uniformly sampled. This allows us to define, for any $\boldsymbol{u} \in \mathbb{R}^2$, a translation operator $\mathcal{T}_{\boldsymbol{u}}$ on discrete sequences, even if $\boldsymbol{u}$ has non-integer values:

$$\mathcal{T}_{\boldsymbol{u}}\mathrm{X}[\boldsymbol{n}] := s\, \mathcal{T}_{s\boldsymbol{u}} f_{\mathrm{X}}(s\boldsymbol{n}), \tag{14}$$

where f_{X} is defined in (11). We can indeed show that this definition is independent from the choice of sampling interval $s > 0$. Besides, given $\mathrm{X} \in l^2_{\mathbb{R}}(\mathbb{Z}^2)$, we have

$$\forall \boldsymbol{k} \in \mathbb{Z}^2,\ \mathcal{T}_{\boldsymbol{k}}\mathrm{X}[\boldsymbol{n}] = \mathrm{X}[\boldsymbol{n} - \boldsymbol{k}]; \tag{15}$$

$$\forall \boldsymbol{u},\, \boldsymbol{v} \in \mathbb{R}^2,\ \mathcal{T}_{\boldsymbol{u}}(\mathcal{T}_{\boldsymbol{v}}\mathrm{X}) = \mathcal{T}_{\boldsymbol{u}+\boldsymbol{v}}\mathrm{X}, \tag{16}$$

which shows that $\mathcal{T}_{\boldsymbol{u}}$ corresponds to the intuitive idea of a shift operator. Expressions (15) and (16) are direct consequence of the following lemma, which bonds the shift operator in the discrete and continuous frameworks.

Lemma 3: For any $\mathrm{X} \in l^2_{\mathbb{R}}(\mathbb{Z}^2)$ and any $\boldsymbol{u} \in \mathbb{R}^2$,

$$f_{\mathcal{T}_{\boldsymbol{u}}\mathrm{X}} = \mathcal{T}_{s\boldsymbol{u}} f_{\mathrm{X}}. \tag{17}$$

Proof: See Appendix E. ∎

We now consider the following corollary to Proposition 2.

Corollary 1: For any shift vector $\boldsymbol{u} \in \mathbb{R}^2$, we have

$$\Big((\mathcal{T}_{\boldsymbol{u}}\mathrm{X} * \overline{\mathrm{W}}) \downarrow m\Big)[\boldsymbol{n}] = \big(\mathcal{T}_{s\boldsymbol{u}} f_{\mathrm{X}} * \overline{\psi}_{\mathrm{W}}\big)\,(ms\boldsymbol{n})\,. \tag{18}$$

Proof: Apply (13) in Proposition 2 with $\mathrm{X} \leftarrow \mathcal{T}_{\boldsymbol{u}}\mathrm{X}$, and use Lemma 3 to conclude. ∎

[6] $\phi^{(s)}$ is a tensor product of scaled and normalized sinc functions.

[7] We actually use the 2D formulation, mentioned in p. 82.

3) Shift Invariance in the Discrete Framework: We consider the following operator, for any $\mathrm{W} \in l^2(\mathbb{Z}^2)$:

$$\mathcal{C}_m[\mathrm{W}] : \mathrm{X} \mapsto |(\mathrm{X} * \overline{\mathrm{W}}) \downarrow m|. \tag{19}$$

When $\mathrm{W} \in \mathcal{G}\big(\boldsymbol{\xi},\, \kappa\big)$, we refer to this as a $\mathbb{C}$GMod operator. For the sake of concision, in what follows we will write $\mathcal{C}_m$ instead of $\mathcal{C}_m[\mathrm{W}]$, when no ambiguity is possible.

We are now ready to state the main result about shift invariance of $\mathbb{C}$GMod outputs.

Theorem 1: Let $\mathrm{W} \in \mathcal{G}\big(\boldsymbol{\xi},\, \kappa\big)$ denote a discrete Gabor-like filter and $m \in \mathbb{N}^*$ denote a subsampling factor. If $\kappa \leq 2\pi/m$, then, for any input image $\mathrm{X} \in l^2_{\mathbb{R}}(\mathbb{Z}^2)$ with finite support and any translation vector $\boldsymbol{u} \in \mathbb{R}^2$ satisfying $\|\boldsymbol{u}\|_1 \leq \pi/\kappa$,

$$\big\|\mathcal{C}_m(\mathcal{T}_{\boldsymbol{u}}\mathrm{X}) - \mathcal{C}_m\mathrm{X}\big\|_2 \leq \alpha(\kappa \boldsymbol{u})\, \|\mathcal{C}_m\mathrm{X}\|_2\,, \tag{20}$$

where α has been defined in (8).

Proof: The proof of this theorem, which involves Lemmas 1-2, Propositions 1-2 and Corollary 1, is provided in Appendix F. ∎

Interestingly, the reference value used in Theorem 1, i.e., $\|\mathcal{C}_m\mathrm{X}\|_2$, is fully shift-invariant, as stated in the following proposition.

Proposition 3: Let $\mathrm{W} \in \mathcal{G}\big(\boldsymbol{\xi},\, \kappa\big)$ and $m \in \mathbb{N}^*$. Assuming $\kappa \leq 2\pi/m$, we have, for any $\mathrm{X} \in l^2_{\mathbb{R}}(\mathbb{Z}^2)$ and any $\boldsymbol{u} \in \mathbb{R}^2$,

$$\|\mathcal{C}_m(\mathcal{T}_{\boldsymbol{u}}\mathrm{X})\|_2 = \|\mathcal{C}_m\mathrm{X}\|_2\,. \tag{21}$$

Proof: See Appendix G. ∎

D. From $\mathbb{C}$GMod to $\mathbb{R}$GPool

Since $\mathbb{C}$GMod operators are not found in classical CNN architectures, the above result does not applies straightforwardly. Instead, the first convolution layer contains real-valued kernels, and is generally followed by ReLU and max pooling. As shown in section IV, this process can be described as an operator parameterized by $\mathrm{W} \in l^2(\mathbb{Z}^2)$, defined by

$$\mathcal{R}_{m,\,q}[\mathrm{W}] : \mathrm{X} \mapsto \mathrm{MaxPool}_q\Big(\big(\mathrm{X} * \overline{\mathrm{Re}\,\mathrm{W}}\big) \downarrow m\Big), \tag{22}$$

where $\mathrm{MaxPool}_q$ selects the maximum value over a sliding grid of size $(2q+1) \times (2q+1)$, with a subsampling factor of 2. More formally, for any $\mathrm{Y} \in l^2_{\mathbb{R}}(\mathbb{Z}^2)$ and any $\boldsymbol{n} \in \mathbb{Z}^2$,

$$\mathrm{MaxPool}_q(\mathrm{Y})[\boldsymbol{n}] := \max_{\|\boldsymbol{k}\|_\infty \leq q} \mathrm{Y}[2\boldsymbol{n} + \boldsymbol{k}]. \tag{23}$$

As hinted in section III-B, an important number of trained convolution kernels exhibit oscillating patterns with various scales and orientations. In such a case, $\mathrm{W} \in \mathcal{G}\big(\boldsymbol{\xi},\, \kappa\big)$ for a certain value of $\boldsymbol{\xi} \in [-\pi,\, \pi]^2$ and $\kappa \in\,]0,\, 2\pi]$, and we refer to $\mathcal{R}_{m,\,q}[\mathrm{W}]$ as an $\mathbb{R}$GPool operator. For the sake of concision, from now on we write $\mathcal{R}_{m,\,q}$ instead of $\mathcal{R}_{m,\,q}[\mathrm{W}]$, when no ambiguity is possible.

In what follows, we show that, under specific conditions on $\widehat{\mathrm{W}}$, $\mathbb{R}$GPool and $\mathbb{C}$GMod operators produce comparable outputs. We then provide a shift invariance bound for $\mathbb{R}$GPool.

1) Continuous Framework: This paragraph, directly adapted from [28, pp. 190–191], provides an intuition about resemblance between $\mathbb{R}$GPool and $\mathbb{C}$GMod in the continuous framework. As will be highlighted later in this section III-D, adaptation to discrete 2D sequences is not straightforward and will require a probabilistic approach.

We consider an input function $f \in L^2_{\mathbb{R}}(\mathbb{R}^2)$ and a band-pass filter $\psi \in \mathcal{V}\big(\boldsymbol{\nu},\, \varepsilon\big)$. Let us also consider

$$g : (\boldsymbol{x},\, \boldsymbol{h}) \mapsto \cos\big(\langle \boldsymbol{\nu},\, \boldsymbol{h}\rangle - \eta(\boldsymbol{x})\big), \tag{24}$$

where η denotes the phase of $f * \overline{\psi}$. Lemma 1 introduced low-frequency functions f_0, with slow variations. Roughly speaking, since $\mathrm{supp}\, f_0 \subset \mathcal{B}_\infty(\varepsilon/2)$, we can define a "minimal wavelength" $\lambda_{f_0} := 2\pi/\varepsilon$. Then,

$$\|\boldsymbol{h}\|_2 \ll \frac{2\pi}{\varepsilon} \implies f_0(\boldsymbol{x} + \boldsymbol{h}) \approx f_0(\boldsymbol{x}), \tag{25}$$

which leads to

$$(f * \mathrm{Re}\,\overline{\psi})(\boldsymbol{x} + \boldsymbol{h}) \approx \big|(f * \overline{\psi})(\boldsymbol{x})\big|\; g(\boldsymbol{x},\, \boldsymbol{h}). \tag{26}$$

On the one hand, we consider a continuous equivalent of the $\mathbb{C}$GMod operator $\mathcal{C}_m[\mathrm{W}]$ as introduced in (19). Such an operator, denoted by $\mathcal{C}[\psi]$, is defined, for any $f \in L^2_{\mathbb{R}}(\mathbb{R}^2)$, by

$$\mathcal{C}[\psi](f) : \boldsymbol{x} \mapsto \big|(f * \overline{\psi})(\boldsymbol{x})\big|\,. \tag{27}$$

On the other hand, we consider the continuous counterpart of $\mathbb{R}$GPool as introduced in (22). It is defined as the maximum value of $f * \mathrm{Re}\,\overline{\psi}$ over a sliding spatial window of size $r > 0$. This is possible because f and $\mathrm{Re}\,\overline{\psi}$ both belong to $L^2_{\mathbb{R}}(\mathbb{R}^2)$, and therefore $f * \mathrm{Re}\,\overline{\psi}$ is continuous. Such an operator, denoted by $\mathcal{R}_r[\psi]$, is defined, for any $f \in L^2_{\mathbb{R}}(\mathbb{R}^2)$, by

$$\mathcal{R}_r[\psi](f) : \boldsymbol{x} \mapsto \max_{\|\boldsymbol{h}\|_\infty \leq r} (f * \mathrm{Re}\,\overline{\psi})(\boldsymbol{x} + \boldsymbol{h}). \tag{28}$$

For the sake of concision, the parameter between square brackets is ignored from now on.

If $r \ll 2\pi/\varepsilon$, then (26) is valid for any $\boldsymbol{h} \in \mathcal{B}_\infty(r)$. Then, using (27) and (28), we get

$$r \ll 2\pi/\varepsilon \implies \mathcal{R}_r f(\boldsymbol{x}) \approx \mathcal{C} f(\boldsymbol{x}) \max_{\|\boldsymbol{h}\|_\infty \leq r} g(\boldsymbol{x},\, \boldsymbol{h}). \tag{29}$$

Using the periodicity of g, we can show that, if $r \geq \frac{\pi}{\|\boldsymbol{\nu}\|_2}$, then $\boldsymbol{h} \mapsto g(\boldsymbol{x},\, \boldsymbol{h})$ reaches its maximum value ($= 1$) on $\mathcal{B}_\infty(r)$. We therefore get

$$\frac{\pi}{\|\boldsymbol{\nu}\|_2} \leq r \ll \frac{2\pi}{\varepsilon} \implies \mathcal{R}_r f(\boldsymbol{x}) \approx \mathcal{C} f(\boldsymbol{x}). \tag{30}$$

An exact quantification of the above approximation remains an open question. In the current paper, it will be provided as a conjecture, for the discrete framework.

2) Adaptation to Discrete 2D Sequences: We consider an input image $\mathrm{X} \in l^2_{\mathbb{R}}(\mathbb{Z}^2)$, a subsampling factor $m \in \mathbb{N}^*$ and a grid half-size $q \in \mathbb{N}^*$. We seek a relationship between

$$\mathrm{Y}_{\mathrm{pool}} := \mathcal{R}_{m,\,q}\mathrm{X} \qquad \text{and} \qquad \mathrm{Y}_{\mathrm{mod}} := \mathcal{C}_{2m}\mathrm{X}, \tag{31}$$

where $\mathcal{C}_{2m}$ and $\mathcal{R}_{m,\,q}$ have been defined in (19) and (22), respectively. Note that, since max pooling also performs subsampling, both $\mathbb{C}$GMod and $\mathbb{R}$GPool operators, as defined in (31), have a subsampling factor equal to $2m$.

We now use the sampling results obtained in section III-C. Let f_{X} and $\psi_{\mathrm{W}} \in \mathcal{U}_s$ denote the functions satisfying (11). On the one hand, we apply (13) in Proposition 2 to $\mathrm{Y}_{\mathrm{mod}}$. For any $\boldsymbol{n} \in \mathbb{Z}^2$,

$$\mathcal{C}_{2m}\mathrm{X}[\boldsymbol{n}] = \mathcal{C} f_{\mathrm{X}}(\boldsymbol{x_n}), \tag{32}$$

where $\boldsymbol{x_n} := 2ms\boldsymbol{n}$. On the other hand, we postulate that

$$\mathcal{R}_{m,\,q}\mathrm{X}[\boldsymbol{n}] = \mathcal{R}_r f_{\mathrm{X}}(\boldsymbol{x_n}) \tag{33}$$

for a certain value of $r \in \mathbb{R}_+^*$. Then, (30) implies $\mathrm{Y}_{\mathrm{mod}} \approx \mathrm{Y}_{\mathrm{pool}}$. However, as shown below, (33) is not satisfied. According to (22) and (23), we have

$$\mathcal{R}_{m,\,q}\mathrm{X}[\boldsymbol{n}] = \max_{\|\boldsymbol{k}\|_\infty \leq q} \mathrm{Re}\left(\left(\mathrm{X} * \overline{\mathrm{W}}\right) \downarrow m\right)[2\boldsymbol{n} + \boldsymbol{k}]. \tag{34}$$

Therefore, according to (13) in Proposition 2, we get

$$\mathcal{R}_{m,\,q}\mathrm{X}[\boldsymbol{n}] = \max_{\|\boldsymbol{k}\|_\infty \leq q} \left(f_{\mathrm{X}} * \mathrm{Re}\,\overline{\psi}_{\mathrm{W}}\right)(\boldsymbol{x_n} + \boldsymbol{h_k}), \tag{35}$$

with

$$\boldsymbol{x_n} := 2ms\boldsymbol{n} \qquad \text{and} \qquad \boldsymbol{h_k} := ms\boldsymbol{k}. \tag{36}$$

By considering $r_q := ms\left(q + \frac{1}{2}\right)$, we get a variant of (33) in which the maximum is evaluated on a discrete grid of $(2q+1)^2$ elements, instead of the continuous region $\mathcal{B}_\infty(r_q)$. As a consequence, (29) is replaced in the discrete framework by

$$q \ll 2\pi/(m\kappa) \implies \mathcal{R}_{m,\,q}\mathrm{X}[\boldsymbol{n}] \approx \mathcal{C}_{2m}\mathrm{X}[\boldsymbol{n}] \max_{\|\boldsymbol{k}\|_\infty \leq q} g_{\mathrm{X}}\big(\boldsymbol{x_n},\, \boldsymbol{h_k}\big), \tag{37}$$

with

$$g_{\mathrm{X}} : (\boldsymbol{x},\, \boldsymbol{h}) \mapsto \cos\big(\langle \boldsymbol{\nu},\, \boldsymbol{h}\rangle - \eta_{\mathrm{X}}(\boldsymbol{x})\big), \tag{38}$$

with $\boldsymbol{\nu} := \boldsymbol{\xi}/s$, and where η_{X} denotes the phase of $f_{\mathrm{X}} * \overline{\psi}_{\mathrm{W}}$. Unlike the continuous case, even if the window size r_q is large enough, the existence of $\boldsymbol{k} \in \{-q\,..\,q\}^2$ such that $g_{\mathrm{X}}\big(\boldsymbol{x_n},\, \boldsymbol{h_k}\big) = 1$ is not guaranteed, as illustrated in Fig. 2 with $q = 1$. Instead, we can only seek a probabilistic estimation of the relative quadratic error between $\mathrm{Y}_{\mathrm{pool}}$ and $\mathrm{Y}_{\mathrm{mod}}$.

Approximation (37) implies

$$q \ll 2\pi/(m\kappa) \implies \left\|\mathcal{C}_{2m}\mathrm{X} - \mathcal{R}_{m,\,q}\mathrm{X}\right\|_2 \approx \left\|\delta_{m,\,q}\mathrm{X}\right\|_2, \tag{39}$$

where $\delta_{m,\,q}\mathrm{X} \in l^2_{\mathbb{R}}(\mathbb{Z}^2)$ is defined such that, for any $\boldsymbol{n} \in \mathbb{Z}^2$,

$$\delta_{m,\,q}\mathrm{X}[\boldsymbol{n}] := \mathcal{C}_{2m}\mathrm{X}[\boldsymbol{n}] \left(1 - \max_{\|\boldsymbol{k}\|_\infty \leq q} g_{\mathrm{X}}\big(\boldsymbol{x_n},\, \boldsymbol{h_k}\big)\right). \tag{40}$$

Expression (39) suggests that the ratio between the left and right terms can be bounded by a quantity which only depends on the product $m\kappa$ (subsampling factor $\times$ frequency localization) and the grid half-size q:

$$\left\|\mathcal{C}_{2m}\mathrm{X} - \mathcal{R}_{m,\,q}\mathrm{X}\right\|_2 \leq \left(1 + \beta_q(m\kappa)\right)\left\|\delta_{m,\,q}\mathrm{X}\right\|_2, \tag{41}$$

for some function $\beta_q : \mathbb{R}_+ \to \mathbb{R}_+$ to be characterized. This is the goal of the following conjecture.

Conjecture 1: There exists

$$\beta_q : \mathbb{R}_+ \to \mathbb{R}_+ \quad \text{satisfying} \quad \beta_q(t) = O(t), \tag{42}$$

independent from the characteristic frequency $\boldsymbol{\xi} \in [-\pi,\, \pi]^2$, such that, for any $\mathrm{X} \in l^2_{\mathbb{R}}(\mathbb{Z}^2)$, (41) is satisfied.

We now seek a probabilistic bound for $\left\|\delta_{m,\,q}\mathrm{X}\right\|_2$.

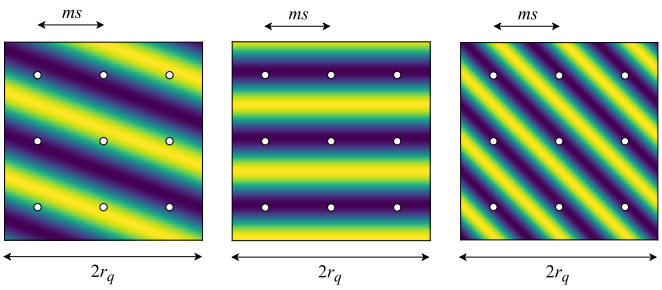

Fig. 2. Search for the maximum value of $\boldsymbol{h} \mapsto g_{\mathrm{X}}(\boldsymbol{x},\, \boldsymbol{h})$ over a discrete grid of size 3×3, i.e., $q = 1$. This figure displays 3 examples with different frequencies $\boldsymbol{\nu} := \boldsymbol{\xi}/s$ and phases $\eta_{\mathrm{X}}(\boldsymbol{x})$. Hopefully the result will be close to the true maximum (left), but there are some pathological cases in which all points in the grid fall into pits (middle and right).

E. Probabilistic Framework

1) Notations: In what follows, for any $z \in \mathbb{C}^*$, we denote by $\angle z \in [0,\, 2\pi[$ the argument of z. We now consider the unit circle $\mathbb{S}^1 \subset \mathbb{C}$. For any $z,\, z' \in \mathbb{S}^1$, the angle between z and z' is given by $\angle(z^* z')$. We then denote by $[z,\, z']_{\mathbb{S}^1} \subset \mathbb{S}^1$ the arc going from z to z' counterclockwise:

$$[z,\, z']_{\mathbb{S}^1} := \left\{z'' \in \mathbb{S}^1 \;\middle|\; \angle(z^* z'') \leq \angle(z^* z')\right\}. \tag{43}$$

By using the relation $\cos\alpha = \mathrm{Re}(\mathrm{e}^{i\alpha})$, (36) and (38) yield, for any $\boldsymbol{n} \in \mathbb{Z}^2$ and any $\boldsymbol{k} \in \{-q\,..\,q\}^2$,

$$g_{\mathrm{X}}\big(\boldsymbol{x_n},\, \boldsymbol{h_k}\big) = \mathrm{Re}\left(z_{\mathrm{X}}(\boldsymbol{x_n})^*\, z_{\boldsymbol{k}}\right), \tag{44}$$

where we have defined

$$z_{\mathrm{X}}(\boldsymbol{x}) := \mathrm{e}^{i\,\eta_{\mathrm{X}}(\boldsymbol{x})} \quad \text{and} \quad z_{\boldsymbol{k}} := \mathrm{e}^{i\langle \boldsymbol{\nu},\, \boldsymbol{h_k}\rangle} = \mathrm{e}^{im\langle \boldsymbol{\xi},\, \boldsymbol{k}\rangle}. \tag{45}$$

Let us denote $N_q := (2q+1)^2$. We consider a sequence, denoted by $\left(z_i^{(q)}\right)_{i \in \{0..N_0-1\}}$, obtained by ordering $\{z_{\boldsymbol{k}}\}_{\boldsymbol{k} \in \{-q..q\}^2}$ in ascending order of their arguments:

$$0 = \theta_0^{(q)} \leq \cdots \leq \theta_{N_q-1}^{(q)} < 2\pi, \tag{46}$$

where we have denoted $\theta_i^{(q)} := \angle\left(z_i^{(q)}\right)$. Besides, we extend the notations with $\theta_{N_q}^{(q)} := 2\pi$ and $z_{N_q}^{(q)} := z_0^{(q)}$. Then, we split $\mathbb{S}^1$ into N_q arcs delimited by $\left(z_i^{(q)}\right)_{i \in \{0..N_0-1\}}$:

$$\mathfrak{A}_i^{(q)} := \begin{cases} \left[z_i^{(q)},\, z_{i+1}^{(q)}\right]_{\mathbb{S}^1} & \text{if } \theta_{i+1}^{(q)} - \theta_i^{(q)} < 2\pi; \\ \mathbb{S}^1 & \text{otherwise.} \end{cases} \tag{47}$$

Finally, for any $i \in \{0\,..\,N_q - 1\}$, we denote by $\omega_i^{(q)} := \theta_{i+1}^{(q)} - \theta_i^{(q)}$ the angular measure of arc $\mathfrak{A}_i^{(q)}$.

Remark 2: According to (45), the above quantities depend on the product $m \times \boldsymbol{\xi} \in \mathbb{R}^2$. Therefore, we will sometimes write $\omega_i^{(q)}(m\boldsymbol{\xi})$, where $\omega_i^{(q)}$ is defined as a function of $\mathbb{R}^2$:

$$\omega_i^{(q)} : \mathbb{R}^2 \to [0,\, 2\pi]. \tag{48}$$

2) Random Variables: From now on, input X is considered as discrete 2D stochastic processes. In order to "randomize" f_{X} introduced in (11), we define a continuous stochastic process from X, denoted by F_{X}, such that

$$\forall \boldsymbol{x} \in \mathbb{R}^2,\ F_{\mathrm{X}}(\boldsymbol{x}) := \sum_{\boldsymbol{n} \in \mathbb{Z}^2} \mathrm{X}[\boldsymbol{n}]\, \phi_{\boldsymbol{n}}^{(s)}(\boldsymbol{x}). \tag{49}$$

Now, we consider the following stochastic processes, which are parameterized by X:

$$M_{\mathrm{X}} := |F_{\mathrm{X}} * \overline{\psi}|; \quad H_{\mathrm{X}} := \angle(F_{\mathrm{X}} * \overline{\psi}); \quad Z_{\mathrm{X}} := \mathrm{e}^{iH_{\mathrm{X}}}, \tag{50}$$

and, for any $\boldsymbol{k} \in \{-q\,..\,q\}^2$,

$$G_{\mathrm{X},\,\boldsymbol{k}} := \mathrm{Re}(Z_{\mathrm{X}}^* z_{\boldsymbol{k}}); \qquad G_{\max\mathrm{X}} := \max_{\|\boldsymbol{k}\|_\infty \leq q} G_{\mathrm{X},\,\boldsymbol{k}}. \tag{51}$$

For any $\boldsymbol{x} \in \mathbb{R}^2$, $f_{\mathrm{X}}(\boldsymbol{x})$ and $\eta_{\mathrm{X}}(\boldsymbol{x})$, such as introduced in (11) and (38), are respectively drawn from $F_{\mathrm{X}}(\boldsymbol{x})$ and $H_{\mathrm{X}}(\boldsymbol{x})$. Then, $z_{\mathrm{X}}(\boldsymbol{x})$ such as introduced in (45) is a realization of $Z_{\mathrm{X}}(\boldsymbol{x})$. Consequently, according to (44), $g_{\mathrm{X}}\big(\boldsymbol{x},\, \boldsymbol{h_k}\big)$ is a realization of $G_{\mathrm{X},\,\boldsymbol{k}}(\boldsymbol{x})$. Besides, according to the definition of $\mathbb{C}$GMod in (19) and $\boldsymbol{x_n}$ in (36), Proposition 2 implies that

$$M_{\mathrm{X}}(\boldsymbol{x_n}) = \mathcal{C}_{2m}\mathrm{X}[\boldsymbol{n}]. \tag{52}$$

We remind that $\boldsymbol{\xi} \in [-\pi,\, \pi]^2$ and $\kappa \in\,]0,\, 2\pi]$ respectively denote the center and size of the Fourier support of W, as introduced in section III-C. To compute the expected discrepancy between $\mathrm{Y}_{\mathrm{pool}}$ and $\mathrm{Y}_{\mathrm{mod}}$, we assume that

$$\|\boldsymbol{\xi}\|_2 \gg 2\pi/M; \tag{53}$$

$$\|\boldsymbol{\xi}\|_2 \gg \kappa, \tag{54}$$

where $M \in \mathbb{N}^*$ denotes the support size of input images. These assumptions exclude low-frequency filters from the scope of our study. We then state the following hypotheses, for which a justification is provided in Appendix H.

Hypothesis 1: For any $\boldsymbol{x} \in \mathbb{R}^2$, $Z_{\mathrm{X}}(\boldsymbol{x})$ is uniformly distributed on $\mathbb{S}^1$.

Hypothesis 2: For any $n \in \mathbb{N}^*$ and $\boldsymbol{x}, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n \in \mathbb{R}^2$, the random variables $M_{\mathrm{X}}(\boldsymbol{y}_i)$ for $i \in \{1\,..\,n\}$ are jointly independent of $Z_{\mathrm{X}}(\boldsymbol{x})$.

F. Expected Quadratic Error between $\mathbb{R}$GPool and $\mathbb{C}$GMod

In this section, we propose to estimate the expected value of the stochastic quadratic error $\widetilde{P}_{\mathrm{X}}^2$, defined such that

$$\widetilde{P}_{\mathrm{X}} := \|\mathcal{C}_{2m}\mathrm{X} - \mathcal{R}_{m,\,q}\mathrm{X}\|_2 \,/\, \|\mathcal{C}_{2m}\mathrm{X}\|_2\,. \tag{55}$$

According to (31), this is an estimation of the relative error between $\mathrm{Y}_{\mathrm{mod}}$ and $\mathrm{Y}_{\mathrm{pool}}$.

First, let us reformulate $\delta_{m,\,q}\mathrm{X}$, introduced in (40), using the probabilistic framework. According to (44) and (51), we have, for any $\boldsymbol{n} \in \mathbb{Z}^2$,

$$\delta_{m,\,q}\mathrm{X}[\boldsymbol{n}] := \mathcal{C}_{2m}\mathrm{X}[\boldsymbol{n}]\,(1 - G_{\max\mathrm{X}}(\boldsymbol{x_n})). \tag{56}$$

We now consider the stochastic process $Q_{\mathrm{X}} := 1 - G_{\max\mathrm{X}}$, and the random variable

$$\widetilde{Q}_{\mathrm{X}} := \|\delta_{m,\,q}\mathrm{X}\|_2 \,/\, \|\mathcal{C}_{2m}\mathrm{X}\|_2\,. \tag{57}$$

The next steps are as follows: 1) at the pixel level, show that $\mathbb{E}[Q_{\mathrm{X}}(\boldsymbol{x})^2]$ depends on the filter frequency $\boldsymbol{\xi}$, and remains close to zero with some exceptions; 2) at the image level, show that the expected value of $\widetilde{Q}_{\mathrm{X}}^2$ is equal to the latter quantity; 3) use Conjecture 1, which states that $\widetilde{P}_{\mathrm{X}} \approx \widetilde{Q}_{\mathrm{X}}$, to deduce an upper bound on the expected value of $\widetilde{P}_{\mathrm{X}}^2$.

The first point is established in Proposition 4 below, and the two remaining ones are the purpose of Theorem 2.

Proposition 4: Assuming Hypothesis 1, the expected value of $Q_{\mathrm{X}}(\boldsymbol{x})^2$ is independent from the choice of $\boldsymbol{x} \in \mathbb{R}^2$, and

$$\mathbb{E}\left[Q_{\mathrm{X}}(\boldsymbol{x})^2\right] = \gamma_q(m\boldsymbol{\xi})^2, \tag{58}$$

where we have defined

$$\gamma_q : \boldsymbol{\zeta} \mapsto \sqrt{\frac{3}{2} + \frac{1}{4\pi}\sum_{i=0}^{N_q-1}\left(\sin\omega_i^{(q)}(\boldsymbol{\zeta}) - 8\sin\frac{\omega_i^{(q)}(\boldsymbol{\zeta})}{2}\right)}, \tag{59}$$

with $\omega_i^{(q)} : \mathbb{R}^2 \mapsto [0,\, 2\pi]$ being introduced in (48).

Proof: See Appendix J. ∎

We consider an ideal scenario where the $z_{\boldsymbol{k}}$ are evenly spaced on $\mathbb{S}^1$. Then, an order 2 Taylor expansion yields $\gamma_q(\boldsymbol{\zeta}) = o(1/q^2)$, meaning that $\widetilde{Q}_{\mathrm{X}}$ quickly vanishes when the grid half-size q increases. Fig. 4 displays $\boldsymbol{\xi} \mapsto \gamma_q(m\boldsymbol{\xi})^2$ for $\boldsymbol{\xi} \in [-\pi,\, \pi]^2$, with $m = 4$ and $q = 1$ as in AlexNet. We notice that, for the major part of the Fourier domain, γ_q remains close to 0. However, we observe a regular pattern of dark regions, which correspond to pathological frequencies where the repartition of $z_{\boldsymbol{k}}$ is unbalanced.

So far, we established a result at the pixel level. Before stating Theorem 2, we need the following intermediate result.

Proposition 5: We consider the random variable

$$\widetilde{S}_{\mathrm{X}} := \|\mathcal{C}_{2m}\mathrm{X}\|_2\,. \tag{60}$$

Under Hypothesis 2, for any $\boldsymbol{x} \in \mathbb{R}^2$,

- $Z_{\mathrm{X}}(\boldsymbol{x})$ is independent of $\widetilde{S}_{\mathrm{X}}$;
- $Z_{\mathrm{X}}(\boldsymbol{x})$, $M_{\mathrm{X}}(\boldsymbol{x})$ are conditionally independent given $\widetilde{S}_{\mathrm{X}}$.

Proof: See Appendix K. ∎

Finally, Propositions 4 and 5 yield the following theorem. It provides an upper bound on the expected value of the relative quadratic error $\widetilde{P}_{\mathrm{X}}^2$, such as defined in (55).

Theorem 2: We assume that Conjecture 1 is true. Then, under Hypotheses 1 and 2, we have

$$\mathbb{E}\big[\widetilde{P}_{\mathrm{X}}^2\big] \leq (1 + \beta_q(m\kappa))^2\, \gamma_q(m\boldsymbol{\xi})^2, \tag{61}$$

with β_q and γ_q being introduced in (42) and (59), respectively.

Proof: See Appendix L. ∎

G. Shift Invariance of $\mathbb{R}$GPool Outputs

In this section, we present the main theoretical claim of this paper. Based on the previous results, we provide a probabilistic measure of shift invariance for $\mathbb{R}$GPool operators.

Theorem 3: Let $\mathrm{W} \in \mathcal{G}\big(\boldsymbol{\xi},\, \kappa\big)$ denote a discrete Gabor-like filter, $m \in \mathbb{N}^*$ a subsampling factor and $q \in \mathbb{N}^*$ a grid half-size. We consider a stochastic process X whose realizations

are elements of $l^2_{\mathbb{R}}(\mathbb{Z}^2)$. We assume that Hypotheses 1 & 2 are satisfied.[8] Besides, we assume Conjecture 1.

Given a translation vector $\boldsymbol{u} \in \mathbb{R}^2$, we consider the following random variable:

$$\widetilde{R}_{\mathrm{X},\boldsymbol{u}} := \left\|\mathcal{R}_{m,q}(\mathcal{T}_{\boldsymbol{u}}\mathrm{X}) - \mathcal{R}_{m,q}\mathrm{X}\right\|_2 / \left\|\mathcal{C}_{2m}\mathrm{X}\right\|_2 . \tag{62}$$

Then, under the following conditions:

$$\kappa \leq \pi/m \qquad \text{and} \qquad \|\boldsymbol{u}\|_1 \leq \pi/\kappa, \tag{63}$$

we have

$$\mathbb{E}\big[\widetilde{R}_{\mathrm{X},\boldsymbol{u}}\big] \leq 2\,(1+\beta_q(m\kappa))\,\gamma_q(m\boldsymbol{\xi}) + \alpha(\kappa\boldsymbol{u}), \tag{64}$$

where α, β_q and γ_q are defined in (8), (42) and (59).

Proof: See Appendix M. ∎

If κ is sufficiently small, then $\alpha(\kappa\boldsymbol{u})$ and $\beta_q(m\kappa)$ become negligible with respect to $\gamma_q(m\boldsymbol{\xi})$, and the bound given in (64) is roughly equal to $2\,\gamma_q(m\boldsymbol{\xi})$. Theorem 3 therefore provides a validity domain for shift invariance of $\mathbb{R}$GPool operators, as illustrated in Fig. 4 with $q = 1$.

Remark 3: The stochastic discrepancy introduced in (62) is estimated relatively to the $\mathbb{C}$GMod output. This choice is motivated by the perfect shift invariance of its norm, as shown in Proposition 3.

Remark 4: In practice, most of the time max pooling is performed on a grid of size 3×3; therefore $q = 1$. For the sake of concision, in the remaining of this paper, we will drop q in the notations, which implicitly means $q = 1$.

IV. Standard CNNs in our Framework

In this section, we show how the theoretical results established above on single-channel inputs can be applied to standard CNNs such as AlexNet or ResNet, which are usually designed for multichannel inputs—e.g., RGB images.

A. Background

We focus on the first layers of a classical CNN architecture, in which input images are propagated through a convolution layer followed by rectified linear unit and max pooling operators, as described below.

We denote by K and $L \in \mathbb{N}^*$ the number of input and output channels, respectively. The convolution layer is characterized by a trainable weight tensor $\mathbf{V} \in \big(l^2_{\mathbb{R}}(\mathbb{Z}^2)\big)^{L\times K}$ with a finite support, a bias vector $\boldsymbol{b} \in \mathbb{R}^L$ and a subsampling factor $m \in \mathbb{N}^*$. Considering an input image $\mathbf{X} \in \big(l^2_{\mathbb{R}}(\mathbb{Z}^2)\big)^K$ with finite support, we denote by $\mathbf{A} \in \big(l^2_{\mathbb{R}}(\mathbb{Z}^2)\big)^L$ the output of the max pooling layer. Then we have, for any $l \in \{0\,..\,L-1\}$,

$$\mathrm{A}_l := (\mathrm{MaxPool} \circ \mathrm{ReLU})\left(b_l + \sum_{k=0}^{K-1} \big(\mathrm{X}_k * \overline{\mathrm{V}_{lk}}\big)\downarrow m\right), \tag{65}$$

where $\mathrm{MaxPool}$ has been introduced in (23) with $q = 1$ (see Remark 4), and ReLU is defined such that, for any $\mathrm{X}' \in l^2_{\mathbb{R}}(\mathbb{Z}^2)$ and any $\boldsymbol{n} \in \mathbb{Z}^2$, $\mathrm{ReLU}(\mathrm{X}')[\boldsymbol{n}] := \max\big(0, \mathrm{X}'[\boldsymbol{n}]\big)$. Expression (65) also introduces a bias notation, defined such that $(b + \mathrm{X}')[\boldsymbol{n}] = b + \mathrm{X}'[\boldsymbol{n}]$ for any $\boldsymbol{n} \in \mathbb{Z}^2$.

In many cases, input images are composed of three RGB channels; therefore $K = 3$. The other parameters depend on the chosen CNN architecture. For example, in AlexNet, the weight tensor $\mathbf{V}$ is supported in a region of size 11×11 and the subsampling factor m is equal to 4. ResNet models use kernels of size 7×7 with $m = 2$.

B. Making CNNs Compatible with our Theoretical Results

The bias and ReLU are outside the scope of our study. However, we can show that $\mathrm{A}_l = \mathrm{ReLU}(b_l + \mathrm{Y}_{\mathrm{pool}\,l})$, where we have defined $\mathrm{Y}_{\mathrm{pool}\,l} := \mathcal{R}^{(l)}_m\mathbf{X}$, with

$$\mathcal{R}^{(l)}_m : \mathbf{X} \mapsto \mathrm{MaxPool}\left(\sum_{k=0}^{K-1} \big(\mathrm{X}_k * \overline{\mathrm{V}_{lk}}\big)\downarrow m\right). \tag{66}$$

For the sake of our study, we therefore consider a strictly equivalent CNN architecture where the bias and ReLU are computed after the max pooling layer. We then focus on the intermediate output $\mathbf{Y}_{\mathrm{pool}} \in \big(l^2_{\mathbb{R}}(\mathbb{Z}^2)\big)^L$.

Remark 5: In many architectures including ResNet, bias $\boldsymbol{b}$ is replaced by a batch normalization layer with affine transformation [49]. Swapping such a layer with max pooling isn't straightforward, but can nevertheless be done with caution. Therefore, we can once again focus on $\mathbf{Y}_{\mathrm{pool}}$ such as introduced in (66).

In what follows, we assume that the network has been trained on a natural image dataset such as ImageNet. Let $\mathcal{L}$ denote the subset of output channels $l \in \{0\,..\,L-1\}$ such that, for any $k \in \{0\,..\,K-1\}$, V_{lk} behaves like a band-pass filter (see Fig. 1). They are referred to as *Gabor channels*. We would like to apply the theoretical results obtained in section III on $\mathrm{Y}_{\mathrm{pool}\,l}$, for any $l \in \mathcal{L}$.

To do so, we need to show that $\mathrm{Y}_{\mathrm{pool}\,l}$ is the output of a $\mathbb{R}$GPool operator such as introduced in (22), for some weight and input to be defined. This is in general not possible though, because the sum operator in (66) cannot be interchanged with max pooling. To solve this problem, we state the following hypothesis, applied on any Gabor channel $l \in \mathcal{L}$. It states that the trained kernels V_{lk} for $k \in \{0\,..\,K-1\}$ are identical, up to a multiplicative constant.

Hypothesis 3: Let $\widetilde{\mathrm{V}}_l := \frac{1}{K}\sum_{k=0}^{K-1}\mathrm{V}_{lk}$ denote the mean kernel of the l-th output channel. Then, for any $l \in \mathcal{L}$, there exists $\boldsymbol{\mu}_l \in \mathbb{R}^K$ such that

$$\forall k \in \{0\,..\,K-1\}\,,\ \mathrm{V}_{lk} = \mu_{lk}\widetilde{\mathrm{V}}_l. \tag{67}$$

Intuitively, when looking at Gabor-like kernels in Fig. 1, they roughly appear grayscale. This observation supports Hypothesis 3 with $\mu_{lk} \approx 1$. A more accurate justification for this hypothesis is provided in Appendix I.

Then, considering the linear combination $\widetilde{\mathrm{X}}_l := \boldsymbol{\mu}_l^\top\mathbf{X}$, we apply Hypothesis 3 to (66):

$$\forall l \in \mathcal{L},\ \mathrm{Y}_{\mathrm{pool}\,l} = \mathcal{R}^{(l)}_m\mathbf{X} = \mathcal{R}_m\big[\widetilde{\mathrm{W}}_l\big](\widetilde{\mathrm{X}}_l), \tag{68}$$

[8] We can easily prove that these properties are independent from the choice of sampling interval $s > 0$.

TABLE I
PERCENTAGE OF ENERGY WITHIN A FOURIER WINDOW OF SIZE $\kappa \times \kappa$, FOR GABOR-LIKE FILTERS $\{\widetilde{\mathrm{W}}_l\}_{l\in\mathcal{L}}$

MODEL	NB CHANNELS	SIZE OF $\mathcal{L}$	κ	MEAN RATIO
ALEXNET	64	26	$\pi/4$	**67%**
RESNET-34	64	22	$\pi/2$	**76%**

where $\widetilde{\mathrm{W}}_l \in l^2(\mathbb{Z}^2)$ denotes the complex-valued companion of $\widetilde{\mathrm{V}}_l$ satisfying (5). We also define $\mathcal{C}_{2m}^{(l)}$ and its associated output feature map $\mathrm{Y}_{\mathrm{mod}\,l}$, such that

$$\forall l \in \mathcal{L},\ \mathrm{Y}_{\mathrm{mod}\,l} := \mathcal{C}_{2m}^{(l)}\mathbf{X} = \mathcal{C}_{2m}\big[\widetilde{\mathrm{W}}_l\big](\widetilde{\mathrm{X}}_l). \tag{69}$$

Besides, the following hypothesis states that, for any $l \in \mathcal{L}$, $\widetilde{\mathrm{W}}_l$ is a discrete Gabor-like filter for which the Fourier support size is shared among the Gabor channels.

Hypothesis 4: There exists $\kappa \leq \pi/m$ such that, for any Gabor channel $l \in \mathcal{L}$,

$$\exists \boldsymbol{\xi}_l \in [-\pi,\,\pi]^2 : \widetilde{\mathrm{W}}_l \in \mathcal{G}\big(\boldsymbol{\xi}_l,\,\kappa\big). \tag{70}$$

Let $l \in \mathcal{L}$. According to Hypotheses 3–4, $\mathcal{R}_m^{(l)}$ and $\mathcal{C}_{2m}^{(l)}$ such as defined in (68) and (69) can be respectively qualified as $\mathbb{R}$GPool and $\mathbb{C}$GMod operators. We now consider

$$\widetilde{P}_{\mathbf{X}}^{(l)} := \big\|\mathcal{C}_{2m}^{(l)}\mathbf{X} - \mathcal{R}_m^{(l)}\mathbf{X}\big\|_2 / \big\|\mathcal{C}_{2m}^{(l)}\mathbf{X}\big\|_2; \tag{71}$$

$$\widetilde{R}_{\mathbf{X},\,\boldsymbol{u}}^{(l)} := \big\|\mathcal{R}_m^{(l)}(\mathcal{T}_{\boldsymbol{u}}\mathbf{X}) - \mathcal{R}_m^{(l)}\mathbf{X}\big\|_2 / \big\|\mathcal{C}_{2m}^{(l)}\mathbf{X}\big\|_2. \tag{72}$$

Then, under Hypotheses 1–4, we can apply Theorems 2 and 3 to the above random variables, thus providing a shift invariance measure for a subset of outputs in CNNs:

$$\mathbb{E}\Big[{\widetilde{P}_{\mathbf{X}}^{(l)}}^2\Big] \leq (1+\beta(m\kappa))^2\,\gamma(m\boldsymbol{\xi}_l)^2; \tag{73}$$

$$\mathbb{E}\Big[\widetilde{R}_{\mathbf{X},\,\boldsymbol{u}}^{(l)}\Big] \leq 2\,(1+\beta(m\kappa))\,\gamma(m\boldsymbol{\xi}_l) + \alpha(\kappa\boldsymbol{u}), \tag{74}$$

where $\boldsymbol{\xi}_l$ has been introduced in (70).

In practice, Hypothesis 4 cannot be exactly satisfied. This is because $\widetilde{\mathrm{V}}_l$ is finitely supported, and thus its power spectrum cannot be exactly zero on a region with non-zero measure. To evaluate how close we are to this ideal situation, we measured the maximum percentage of energy within a window of size $\kappa \times \kappa$ in the Fourier domain, with respect to the whole filter W_l. We then computed the mean percentage over all the Gabor channels $l \in \mathcal{L}$. The results are shown in Table I, for AlexNet and ResNet after training with ImageNet. The window size κ has been set to its highest admissible value, i.e., π/m.

We notice that residual energy outside the window of interest is quite high, especially for AlexNet. Therefore, W_l deviates from a "perfect" Gabor-like filter, which may lead to higher instabilities.

V. OPERATORS BASED ON THE DUAL-TREE WAVELET PACKET TRANSFORM

In order to validate the results established in section III, we consider the dual-tree wavelet packet transform (DT-$\mathbb{C}$WPT), a fast algorithm which achieves subsampled convolutions with discrete Gabor-like filters. As explained below, DT-$\mathbb{C}$WPT spawns a set of filters which tiles the whole frequency domain in a regular fashion. Furthermore, increasing the subsampling factor m results in a decreased Fourier support size $\kappa = \pi/m$, therefore matching the condition stated in (63). This is also consistent with what is observed in CNNs such as AlexNet or ResNet. DT-$\mathbb{C}$WPT thus provides a convenient framework to emulate the behavior of an actual CNN while testing Theorems 1–3 in a controlled environment.

A. Discrete Wavelet Packet Transform

This is a brief overview on the classical, real-valued 2D wavelet packet transform (WPT) algorithm [48, p. 377], which is the starting point for building the redundant, complex-valued DT-$\mathbb{C}$WPT.

Given a pair of low- and high-pass 1D orthogonal filters $\mathrm{h},\,\mathrm{g} \in l^2_{\mathbb{R}}(\mathbb{Z})$ satisfying a *quadrature mirror filter* (QMF) relationship, we consider a separable 2D filter bank (FB), denoted by $\mathbf{G} := (\mathrm{G}_l)_{l\in\{0..3\}}$, defined by

$$\mathrm{G}_0 = \mathrm{h}\otimes\mathrm{h};\ \mathrm{G}_1 = \mathrm{h}\otimes\mathrm{g};\ \mathrm{G}_2 = \mathrm{g}\otimes\mathrm{h};\ \mathrm{G}_3 = \mathrm{g}\otimes\mathrm{g}. \tag{75}$$

Let $\mathrm{X} \in l^2_{\mathbb{R}}(\mathbb{Z}^2)$. The decomposition starts with $\mathrm{X}_0^{(0)} = \mathrm{X}$. Given $j \in \mathbb{N}$, suppose that we have computed 4^j sequences of wavelet packet coefficients at stage j, denoted by $\mathrm{X}_l^{(j)} \in l^2_{\mathbb{R}}(\mathbb{Z}^2)$ for each $l \in \big\{0\,..\,4^j-1\big\}$. They are referred to as *feature maps*.

At stage $j+1$, we compute a new representation of X with increased frequency resolution—and decreased spatial resolution. It is obtained by further decomposing each feature map $\mathrm{X}_l^{(j)}$ into four sub-sequences, using subsampled (or strided) convolutions with kernels G_k, for each $k \in \{0\,..\,3\}$:

$$\forall k \in \{0\,..\,3\},\ \mathrm{X}_{4l+k}^{(j+1)} = \Big(\mathrm{X}_l^{(j)} * \overline{\mathrm{G}_k}\Big) \downarrow 2. \tag{76}$$

The algorithm stops after reaching the desired number of stages $J > 0$—referred to as *decomposition depth*. Then,

$$\mathbf{X}^{(J)} := \big(\mathrm{X}_l^{(J)}\big)_{l\in\{0..4^J-1\}} \tag{77}$$

constitutes a multichannel representation of X in an orthonormal basis, from which the original image can be perfectly reconstructed.

The following proposition introduces an array of resulting kernels $\mathbf{V}^{(J)}$ which is illustrated in Fig. 3 with $J = 2$.

Proposition 6: For any $l \in \big\{0\,..\,4^J-1\big\}$, there exists $\mathrm{V}_l^{(J)} \in l^2_{\mathbb{R}}(\mathbb{Z}^2)$ such that

$$\mathrm{X}_l^{(J)} = \Big(\mathrm{X} * \overline{\mathrm{V}_l^{(J)}}\Big) \downarrow 2^J. \tag{78}$$

Proof: See Appendix N. ∎

B. Dual-Tree Complex Wavelet Packet Transform

Despite having interesting properties such as sparse signal representation, WPT is unstable with respect to small shifts and suffers from a poor directional selectivity. To overcome this, N. Kingsbury designed a new type of discrete wavelet transform [29], where images are decomposed in a redundant frame of nearly-analytic, complex-valued waveforms. It was

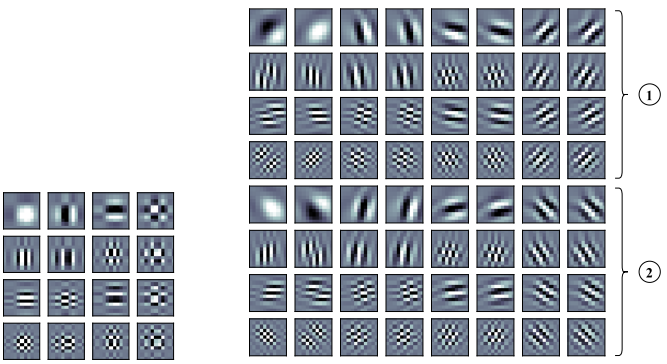

Fig. 3. Resulting kernels $\mathbf{V}^{(2)}$ for WPT (left) and $\mathbf{W}^{(2)}$ for DT-$\mathbb{C}$WPT (right), computed with Q-shift orthogonal QMFs of length 10 [50]. They have been cropped to size 11×11 for the sake of visibility. The right part of the figure displays 32 complex filters, alternatively represented by their real and imaginary parts. The feature maps related to ① and ② are obtained with two distinct formulas, which are summarized in (82). Illustration from [31].

later extended to the wavelet packet framework [30]. The latter operation, referred to as *dual-tree complex wavelet packet transform* (DT-$\mathbb{C}$WPT), is performed as follows.

Let $(\mathrm{h}_0, \mathrm{g}_0)$ and $(\mathrm{h}_1, \mathrm{g}_1)$ denote two pairs of QMFs as defined in section V-A, satisfying the *half-sample delay* condition:

$$\forall \omega \in [-\pi, \pi], \widehat{\mathrm{h}}_1(\omega) = \mathrm{e}^{-i\omega/2}\, \widehat{\mathrm{h}}_0(\omega). \tag{79}$$

Then, for any $k \in \{0\,..\,3\}$, we build a two-dimensional FB $\mathbf{G}_k := (\mathrm{G}_{k,\,l})_{l \in \{0..3\}}$ similarly to (75):

$$\mathrm{G}_{k,\,0} = \mathrm{h}_i \otimes \mathrm{h}_j; \qquad \mathrm{G}_{k,\,1} = \mathrm{h}_i \otimes \mathrm{g}_j; \tag{80}$$

$$\mathrm{G}_{k,\,2} = \mathrm{g}_i \otimes \mathrm{h}_j; \qquad \mathrm{G}_{k,\,3} = \mathrm{g}_i \otimes \mathrm{g}_j, \tag{81}$$

where $i,\, j \in \{0,\, 1\}$ are defined such that $k = 2 \times i + j$.[9]

Let $J > 0$ denote a decomposition depth. Using each of the four FBs $\mathbf{G}_{0-3}$ as defined above, we assume that we have decomposed an input image X into four multichannel WPT representations $\mathbf{X}_{0-3}^{(J)}$, each of which satisfies (76) and (77). Then, for any $l \in \{0\,..\,4^J - 1\}$, the following complex feature maps are computed:

$$\begin{pmatrix} \mathrm{Z}_l^{(J)} \\ \mathrm{Z}_{4^J+l}^{(J)} \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \mathrm{X}_{0,\,l}^{(J)} \\ \mathrm{X}_{3,\,l}^{(J)} \end{pmatrix} + i \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \mathrm{X}_{2,\,l}^{(J)} \\ \mathrm{X}_{1,\,l}^{(J)} \end{pmatrix}. \tag{82}$$

We denote by $L_J := 2 \cdot 4^J$ the number of output feature maps. Then,

$$\mathbf{Z}^{(J)} := \big(\mathrm{Z}_l^{(J)}\big)_{l \in \{0..L_J-1\}} \tag{83}$$

constitutes a complex-valued, four-time redundant multichannel representation of X from which the original image can be reconstructed.

As for standard WPT, the following proposition introduces an array of resulting (complex-valued) kernels $\mathbf{W}^{(J)}$ for which a graphical representation is provided in Fig. 3 with $J = 2$.

[9]Actually, the FB design requires some technicalities which are not described here.

Proposition 7: For any $l \in \{0\,..\,L_J - 1\}$, there exists $\mathrm{W}_l^{(J)} \in l^2(\mathbb{Z}^2)$ such that

$$\mathrm{Z}_l^{(J)} = \Big(\mathrm{X} * \overline{\mathrm{W}_l^{(J)}}\Big) \downarrow 2^J. \tag{84}$$

Proof: See Appendix O. ∎

C. Invariance Results Applied to the Dual-Tree Framework

We assume that h_0 is a Shannon filter, such that $\widehat{\mathrm{h}}_0(\omega) := \sqrt{2}$ if $\omega \in \left[-\frac{\pi}{2},\, \frac{\pi}{2}\right]$ and 0 otherwise. Let $J \in \mathbb{N}^*$ denote the number of decomposition stages. We consider the resulting kernels $\mathrm{W}_l^{(J)}$ satisfying (84). The following hypothesis states that DT-$\mathbb{C}$WPT tiles the frequency plane with a square grid.

Hypothesis 5: For any $l \in \{0\,..\,L_J - 1\}$, there exists $\boldsymbol{\sigma}_l^{(J)} \in \left\{-2^J\,..\,2^J - 1\right\}^2$ such that

$$\mathrm{W}_l^{(J)} \in \mathcal{G}\big(\boldsymbol{\xi}_l^{(J)},\, \kappa_J\big), \tag{85}$$

where we have defined

$$\boldsymbol{\xi}_l^{(J)} := \left(\boldsymbol{\sigma}_l^{(J)} + \frac{1}{2}\right) \frac{\pi}{2^J} \quad \text{and} \quad \kappa_J := \frac{\pi}{2^J}. \tag{86}$$

It can be shown that Hypothesis 5 is an asymptotic result, when J goes to ∞. In reality, the Fourier support of $\mathrm{W}_l^{(J)}$ is contained in four symmetric square regions of size κ_J. If the dual filter h_1 satisfies the half-sample delay condition (79), then the energy of $\mathrm{W}_l^{(J)}$ goes to 0 in all but one of the four regions (relatively to the filter's total energy). We nevertheless consider Hypothesis 5 as reasonable when $J \geq 2$.[10]

Then, according to (84) and (85), we can apply Theorems 1–3 to the dual-tree framework. More precisely, for any output channel $l \in \{0\,..\,L_J - 1\}$, we consider

$$\mathrm{Y}_{\mathrm{pool}\,l}^{(J)} := \mathcal{R}_l^{(J)}\mathrm{X} \qquad \text{and} \qquad \mathrm{Y}_{\mathrm{mod}\,l}^{(J)} := \mathcal{C}_l^{(J)}\mathrm{X}, \tag{87}$$

where we have denoted, for the sake of readability,

$$\mathcal{R}_l^{(J)} := \mathcal{R}_{m_J}\big[\mathrm{W}_l^{(J)}\big] \quad \text{and} \quad \mathcal{C}_l^{(J)} := \mathcal{C}_{2m_J}\big[\mathrm{W}_l^{(J)}\big], \tag{88}$$

with $m_J := 2^{J-1}$. We recall that $\mathcal{R}_m$ and $\mathcal{C}_{2m}$ have been introduced in (31) with $q \leftarrow 1$, for any $m \in \mathbb{N}^*$.

Both $\mathcal{R}_l^{(J)}$ and $\mathcal{C}_l^{(J)}$ perform DT-$\mathbb{C}$WPT with J decomposition stages. However, the max pooling operator requires an extra level of subsampling. To counterbalance this, when computing $\mathrm{Y}_{\mathrm{pool}}^{(J)}$, the last stage of DT-$\mathbb{C}$WPT decomposition must be performed without subsampling, at the cost of increased redundancy.[11] This is why $m_J := 2^{J-1}$ instead of 2^J.

In order Theorems 1 and 3 to be applicable, we need to check that conditions stated in (63) are satisfied. Using the DT-$\mathbb{C}$WPT framework, they become

$$\kappa_J \leq \frac{\pi}{2^{J-1}} \qquad \text{and} \qquad \|\boldsymbol{u}\|_1 \leq \frac{\pi}{\kappa_J} = 2^J. \tag{89}$$

The first condition is always met, according to (86). As for the second one, it establishes a limit on $\|\boldsymbol{u}\|_1$ above which

[10]This asymptotic result is not true for "edge filters", i.e., when $\big\|\boldsymbol{\xi}_l^{(J)}\big\|_\infty = \big(1 - 2^{-(J+1)}\big)\,\pi$. In this case, a small fraction of the filter's energy remains located at the far end of the Fourier domain [30]. However, this edge effect is ignored and we still consider Hypothesis 5 as reasonable.

[11]This is similar to the concept of stationary wavelet transform [51].

shift invariance can no longer be estimated. Note however that shifting the input by 2^J pixels results in a 1-pixel output shift. Therefore, $\mathcal{R}_l^{(J)}(\mathcal{T}_{\boldsymbol{u}}\mathrm{X})$ can always be compared with a shifted version of $\mathcal{R}_l^{(J)}\mathrm{X}$. We then get a partial measure of shift equivariance.

In the Shannon setting, $\mathrm{h}_0[n]$, $\mathrm{g}_0[n]$, $\mathrm{h}_1[n]$ and $\mathrm{g}_1[n]$ decay in $O(1/n)$, which makes them difficult to approximate in practice. It requires very large vectors to avoid numerical instabilities. Practical implementations use fast-decaying filters such as Meyer QMFs [52], or finite-length filters which approximate the half-sample delay condition [53]. Therefore, residual energy can be observed outside the Fourier windows introduced in Hypothesis 5. To counterbalance this, we relax this hypothesis by increasing the window size up to $\kappa_J := \pi/2^{J-1}$, which is closer to what is observed in standard convolutions layers (see Table I).

VI. Experiments and Results

In this section, we experimentally validate our theoretical results. To do so, we built operators based on DT-$\mathbb{C}$WPT, as explained in section V. Using a dataset of natural images, we measured the mean discrepancy between $\mathbb{R}$GPool and $\mathbb{C}$GMod outputs, and evaluated the shift invariance of both models.

Dual-tree decompositions have been performed with Q-shift orthogonal filters of length 10 [50], which approximately meets the half-sample delay condition (79).

Our implementation is based on PyTorch. The models were evaluated on the validation set of ImageNet ILSVRC2012 [54], which contains $N_{\mathrm{data}} := 50\,000$ images.

A. Discrepancies between $\mathbb{R}$GPool and $\mathbb{C}$GMod

Each image $n \in \{0\,..\,N_{\mathrm{data}}-1\}$ in the dataset was converted to grayscale, from which a center crop of size 224×224 was extracted. We denote by $\mathrm{X}_n \in l^2_{\mathbb{R}}(\mathbb{Z}^2)$ the resulting input feature map. For any $l \in \{0\,..\,L_J-1\}$, we denote by $\mathrm{Y}^{(J)}_{\mathrm{pool}\,nl}$ and $\mathrm{Y}^{(J)}_{\mathrm{mod}\,nl}$ the outputs satisfying (87) with $\mathrm{X} \leftarrow \mathrm{X}_n$.

Then, the relative quadratic error between $\mathrm{Y}^{(J)}_{\mathrm{mod}\,nl}$ and $\mathrm{Y}^{(J)}_{\mathrm{pool}\,nl}$ was computed. It is defined by the square of

$$\rho^{(J)}_{nl} := \big\|\mathrm{Y}^{(J)}_{\mathrm{mod}\,nl} - \mathrm{Y}^{(J)}_{\mathrm{pool}\,nl}\big\|_2 / \big\|\mathrm{Y}^{(J)}_{\mathrm{mod}\,nl}\big\|_2. \qquad (90)$$

Finally, the for each output channel l, an empirical estimate for $\mathbb{E}\big[\widetilde{P}^2_{\mathrm{X}}\big]$ such as introduced in (55) was obtained by averaging $\big(\rho^{(J)}_{nl}\big)^2$ over the whole dataset. We denote by $\big(\widetilde{\rho}^{(J)}_l\big)^2$ the corresponding quantity.

Since $\mathcal{R}^{(J)}_l$ and $\mathcal{C}^{(J)}_l$ are parameterized by $\mathrm{W}^{(J)}_l$, the value of $\big(\widetilde{\rho}^{(J)}_l\big)^2$ depends on the filter's characteristic frequency $\boldsymbol{\xi}^{(J)}_l$. According to Hypothesis 5, these frequencies form a regular grid in the Fourier domain. This provides a visual representation of $\big(\widetilde{\rho}^{(J)}_l\big)^2$, as shown in Fig. 5. We can observe a regular pattern of dark spots. More precisely, high discrepancies between max pooling and modulus seem to occur when the support of $\widehat{\mathrm{W}^{(J)}_l}$ is located in a dark region of Fig. 4. This result corroborates the theoretical study, which states that high discrepancies are expected for certain pathological frequencies, due to the search for a maximum value over a discrete grid—see Fig. 2.

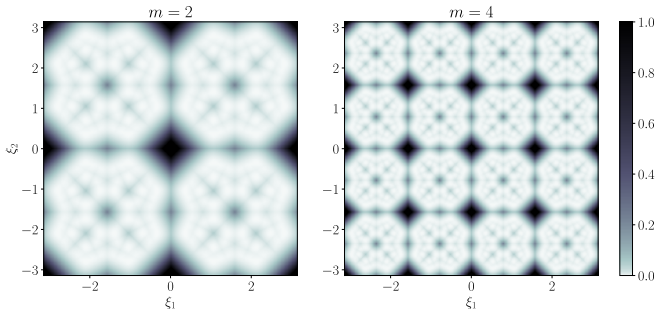

Fig. 4. $\gamma(m\boldsymbol{\xi})^2$ as a function of the kernel characteristic frequency $\boldsymbol{\xi} \in [-\pi, \pi]^2$. According to Theorem 2, this quantity provides an approximate bound for the expected quadratic error between $\mathbb{R}$GPool and $\mathbb{C}$GMod outputs. The subsampling factor m has been set to 2 as in ResNet (left), and 4 as in AlexNet (right). The bright regions correspond to frequencies for which the two outputs are expected to be similar. However, in the dark regions, pathological cases such as illustrated in Fig. 2 are more likely to occur.

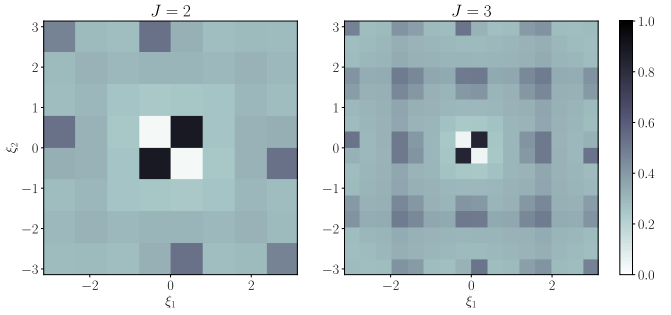

Fig. 5. Relative quadratic error between the outputs of $\mathbb{R}$GPool and $\mathbb{C}$GMod. For each wavelet packet channel $l \in \{0\,..\,L_J-1\}$, $\big(\widetilde{\rho}^{(J)}_l\big)^2$ is represented as a grayscale pixel centered in $\boldsymbol{\xi}^{(J)}_l$, such as introduced in (86). Since the subsampling factor m_J is equal to 2^{J-1}, these empirical estimates can be compared with the left and right parts of Fig. 4. The plots are symmetrized to account for the complex conjugate feature maps.

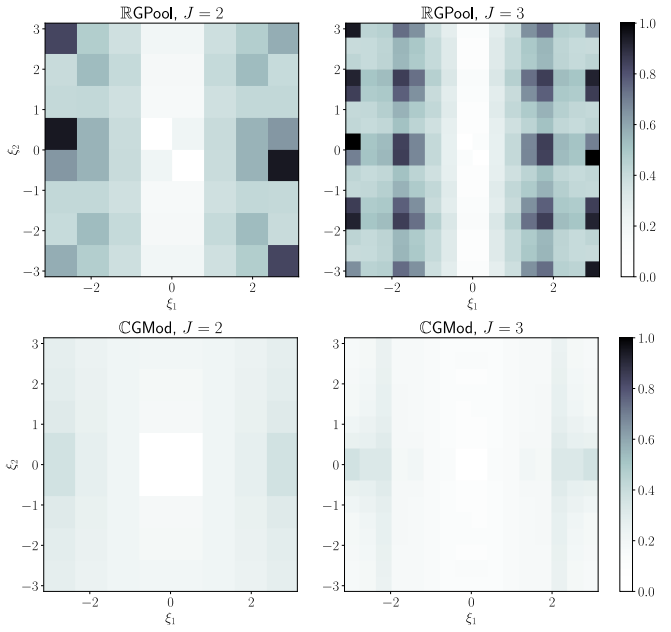

Fig. 6. Empirical measure of horizontal shift invariance of $\mathbb{R}$GPool and $\mathbb{C}$GMod outputs. For each $l \in \{0\,..\,L_J-1\}$, $\widetilde{\rho}^{(J)}_{\mathrm{pool}\,l}$ (top) and $\widetilde{\rho}^{(J)}_{\mathrm{mod}\,l}$ (bottom) are represented as a grayscale pixel centered in $\boldsymbol{\xi}^{(J)}_l$, such as introduced in (86).

B. Shift invariance

For each input image X_n previously converted to grayscale, two crops of size 224×224 were extracted, such that the corresponding sequences X_n and X'_n are shifted by one pixel along the x-axis.

From these inputs, the following quantity was then computed:

$$\rho^{(J)}_{\mathrm{pool}_{nl}} := \left\|\mathrm{Y}'^{(J)}_{\mathrm{pool}_{nl}} - \mathrm{Y}^{(J)}_{\mathrm{pool}_{nl}}\right\|_2 / \left\|\mathrm{Y}^{(J)}_{\mathrm{mod}_{nl}}\right\|_2, \tag{91}$$

where $\mathrm{Y}^{(J)}_{\mathrm{pool}}$, $\mathrm{Y}'^{(J)}_{\mathrm{pool}}$ and $\mathrm{Y}^{(J)}_{\mathrm{mod}}$ satisfy (87) with $\mathrm{X} \leftarrow \mathrm{X}_n$ or X'_n. Finally, for each output channel $l \in \{0\,..\,L_J - 1\}$, an empirical estimate for $\mathbb{E}\big[\widetilde{R}_{\mathrm{X},\,\boldsymbol{u}}\big]$ such as introduced in (62), with $\boldsymbol{u} := (1,\,0)$, was obtained by averaging $\rho^{(J)}_{\mathrm{pool}_{nl}}$ over the whole dataset. We denote by $\widetilde{\rho}^{(J)}_{\mathrm{pool}_l}$ the corresponding quantity. We point out that shift invariance is measured relatively to the norm of the $\mathbb{C}$GMod output, as explained in Remark 3.

On the other hand, the same procedure was applied on the $\mathbb{C}$GMod operators:

$$\rho^{(J)}_{\mathrm{mod}_{nl}} := \left\|\mathrm{Y}'^{(J)}_{\mathrm{mod}_{nl}} - \mathrm{Y}^{(J)}_{\mathrm{mod}_{nl}}\right\|_2 / \left\|\mathrm{Y}^{(J)}_{\mathrm{mod}_{nl}}\right\|_2, \tag{92}$$

and $\widetilde{\rho}^{(J)}_{\mathrm{mod}_l}$ was obtained as before by averaging $\rho^{(J)}_{\mathrm{mod}_{nl}}$ over the whole dataset.

A visual representation of $\widetilde{\rho}^{(J)}_{\mathrm{pool}_l}$ and $\widetilde{\rho}^{(J)}_{\mathrm{mod}_l}$ are provided in Fig. 6. Two observations can be drawn here.

When the filter is horizontally oriented, the corresponding output is highly stable with respect to horizontal shifts. This can be explained by noticing that such kernels perform low-pass filtering along the x axis. The exact transposed phenomenon occurs for vertical shifts (not shown in this paper).

Elsewhere, we observe that high discrepancies between $\mathbb{R}$GPool and $\mathbb{C}$GMod outputs (Fig. 5) are correlated with shift instability of $\mathbb{R}$GPool (Fig. 6, top). This is in line with (61) and (64) in Theorems 2–3. Note that $\mathbb{C}$GMod outputs are nearly shift invariant regardless the characteristic frequency $\boldsymbol{\xi}^{(J)}_l$ (Fig. 6, bottom), as predicted by Theorem 1 (20).

VII. Conclusion

In this paper, we studied shift invariance properties captured by the max pooling operator, when applied on top of a convolution layer with Gabor-like kernels. More precisely, we established a validity domain for near-shift invariance, and confirmed our predictions with an experimental setting based on the dual-tree complex wavelet packet transform.

As shown in this paper, the $\mathbb{C}$GMod operator acts like a proxy for $\mathbb{R}$GPool, extracting comparable features with higher stability. This result suggests a way to build an architecture sharing the structure and behavior of a standard network, except that shift invariance would be improved. This could be done by considering a DT-$\mathbb{C}$WPT-based twin model such as introduced in our workshop paper [31], and replacing $\mathbb{R}$GPool by $\mathbb{C}$GMod operators as done above in a deterministic context.

Since CNNs generally implement successive blocks of convolution and pooling layers, another line of research would be to extend our results to a cascade of $\mathbb{R}$GPool operators. Our work is thus an important step towards deeper understanding of popular networks, based on the wavelet theory.

Appendix A
Proof of Lemma 1

Proof: We can show that, for any $\boldsymbol{\omega} \in \mathbb{R}^2$,

$$\widehat{f}_0(\boldsymbol{\omega}) = \widehat{(f * \overline{\psi})}(\boldsymbol{\omega} - \boldsymbol{\nu}) = \mathcal{T}_{\boldsymbol{\nu}}(\widehat{f}\,\widehat{\overline{\psi}})(\boldsymbol{\omega}).$$

By hypothesis on ψ, we have $\operatorname{supp}\widehat{\overline{\psi}} \subset \mathcal{B}_\infty(-\boldsymbol{\nu},\,\varepsilon/2)$, which yields the result. ∎

Appendix B
Proof of Proposition 1

Proof: Using the 2D Plancherel formula, we compute

$$\begin{aligned}
\|\mathcal{T}_{\boldsymbol{h}} f_0 - f_0\|^2_{L^2} &= \frac{1}{4\pi^2}\left\|\widehat{\mathcal{T}_{\boldsymbol{h}} f_0} - \widehat{f}_0\right\|^2_{L^2} \\
&= \frac{1}{4\pi^2}\iint_{\mathcal{B}_\infty(\varepsilon/2)} \left|\widehat{f}_0(\boldsymbol{\omega})\right|^2 \left|\mathrm{e}^{-i\langle \boldsymbol{h},\,\boldsymbol{\omega}\rangle} - 1\right|^2 \mathrm{d}^2\boldsymbol{\omega} \\
&= \frac{1}{4\pi^2}\iint_{\mathcal{B}_\infty(\varepsilon/2)} \left|\widehat{f}_0(\boldsymbol{\omega})\right|^2 \big(2 - 2\cos\langle \boldsymbol{h},\,\boldsymbol{\omega}\rangle\big)\,\mathrm{d}^2\boldsymbol{\omega}.
\end{aligned}$$

The integral is computed on a compact domain because, according to Lemma 1, $\operatorname{supp}\widehat{f}_0 \subset \mathcal{B}_\infty(\varepsilon/2)$. Now, we use the Cauchy-Schwarz inequality to compute:

$$\begin{aligned}
\forall \boldsymbol{\omega} \in \mathcal{B}_\infty(\varepsilon/2),\ |\langle \boldsymbol{h},\,\boldsymbol{\omega}\rangle| &\le \|\boldsymbol{h}\|_1 \cdot \|\boldsymbol{\omega}\|_\infty \\
&\le \frac{\varepsilon}{2}\,\|\boldsymbol{h}\|_1\,.
\end{aligned}$$

By hypothesis on $\boldsymbol{h}$, $\frac{\varepsilon}{2}\|\boldsymbol{h}\|_1 \le \frac{\pi}{2}$, and thus

$$\|\mathcal{T}_{\boldsymbol{h}} f_0 - f_0\|^2_{L^2} \le \left(2 - 2\cos\frac{\varepsilon\,\|\boldsymbol{h}\|_1}{2}\right)\|f_0\|^2_{L^2}\,.$$

Finally, since $\cos t \ge 1 - \frac{t^2}{2}$, we get (7). ∎

Appendix C
Proof of Lemma 2

Proof: Expression (9) is obtained by adapting (3.2) and (3.3) in [48] to the 2D case. Then, by combining (9) with the Plancherel formula, we get

$$\begin{aligned}
\|g\|^2_{L^2} &= \frac{1}{4\pi^2}\,\|\widehat{g}\|^2_{L^2} \\
&= \frac{1}{4\pi^2}\iint_{\mathcal{B}_\infty(\pi/s)} |\widehat{g}(\boldsymbol{\omega})|^2\,\mathrm{d}^2\boldsymbol{\omega} \\
&= \frac{1}{4\pi^2}\iint_{\mathcal{B}_\infty(\pi/s)} \left|s\,\widehat{\mathrm{Y}}(s\boldsymbol{\omega})\right|^2 \mathrm{d}^2\boldsymbol{\omega}.
\end{aligned}$$

The integral is performed on $\mathcal{B}_\infty(\pi/s)$ because $g \in \mathcal{U}_s$. Then, by applying the change of variable $\boldsymbol{\omega}' \leftarrow s\boldsymbol{\omega}$, we get

$$\begin{aligned}
\|g\|^2_{L^2} &= \frac{1}{4\pi^2}\iint_{\mathcal{B}_\infty(\pi)} \left|\widehat{\mathrm{Y}}(\boldsymbol{\omega}')\right|^2 \mathrm{d}^2\boldsymbol{\omega}' \\
&= \frac{1}{4\pi^2}\big\|\widehat{\mathrm{Y}}\big\|^2_{L^2} = \|\mathrm{Y}\|^2_2\,,
\end{aligned}$$

hence (10), which concludes the proof. ∎

APPENDIX D
PROOF OF PROPOSITION 2

Proof: First, f_{X} and ψ_{W} are well defined because $\mathrm{X} \in l^2_{\mathbb{R}}(\mathbb{Z}^2)$ and $\mathrm{W} \in l^2(\mathbb{Z}^2)$. By construction, $\psi_{\mathrm{W}} \in \mathcal{U}_s$. Thus, (12) is a direct consequence of the Shannon-Whittaker sampling theorem [48, p. 61], coupled with the orthonormality of $\{\phi^{(s)}_{\boldsymbol{n}}\}_{\boldsymbol{n}\in\mathbb{Z}^2}$. Therefore, using (9) in Lemma 2, we get, for any $\boldsymbol{\omega} \in \mathcal{B}_\infty(\pi/s)$, $\widehat{\psi}_{\mathrm{W}}(\boldsymbol{\omega}) = s\widehat{\mathrm{W}}(s\boldsymbol{\omega})$. Since $\widehat{\psi}_{\mathrm{W}}(\boldsymbol{\omega}) = 0$ outside $\mathcal{B}_\infty(\pi/s)$, we deduce that $\psi_{\mathrm{W}} \in \mathcal{V}\big(\boldsymbol{\xi}/s,\, \kappa/s\big)$.

We now prove (13). For $\boldsymbol{n} \in \mathbb{Z}^2$, we compute:

$$\begin{aligned}(f_{\mathrm{X}} * \overline{\psi}_{\mathrm{W}})(ms\boldsymbol{n}) &= \iint_{\mathbb{R}^2} f_{\mathrm{X}}(ms\boldsymbol{n}-\boldsymbol{x})\,\overline{\psi}(\boldsymbol{x})\,\mathrm{d}^2\boldsymbol{x}\\ &= \iint_{\mathbb{R}^2} \sum_{\boldsymbol{k}\in\mathbb{Z}^2} \mathrm{X}[\boldsymbol{k}]\phi^{(s)}_{\boldsymbol{k}}(ms\boldsymbol{n}-\boldsymbol{x})\,\overline{\psi}(\boldsymbol{x})\,\mathrm{d}^2\boldsymbol{x}\\ &= \sum_{\boldsymbol{k}\in\mathbb{Z}^2} \mathrm{X}[\boldsymbol{k}] \iint_{\mathbb{R}^2} \phi^{(s)}_{\boldsymbol{k}}(ms\boldsymbol{n}-\boldsymbol{x})\,\overline{\psi}(\boldsymbol{x})\,\mathrm{d}^2\boldsymbol{x}.\end{aligned}$$

The sum-integral interchange is possible because X has a finite support. Then:

$$\begin{aligned}&(f_{\mathrm{X}} * \overline{\psi}_{\mathrm{W}})\,(ms\boldsymbol{n})\\ &\quad= \sum_{\boldsymbol{k}\in\mathbb{Z}^2} \mathrm{X}[\boldsymbol{k}] \iint_{\mathbb{R}^2} \overline{\psi}(\boldsymbol{x})\,\phi^{(s)}\big(s(m\boldsymbol{n}-\boldsymbol{k})-\boldsymbol{x}\big)\,\mathrm{d}^2\boldsymbol{x}\\ &\quad= \sum_{\boldsymbol{k}\in\mathbb{Z}^2} \mathrm{X}[\boldsymbol{k}]\,\big(\overline{\psi} * \phi^{(s)}\big)\big(s(m\boldsymbol{n}-\boldsymbol{k})\big).\end{aligned}$$

Since $\{\phi^{(s)}_{\boldsymbol{n}}\}_{\boldsymbol{n}\in\mathbb{Z}^2}$ is an orthonormal basis of $\mathcal{U}_s$, we can easily show that, for any $\boldsymbol{k}' \in \mathbb{Z}^2$, $\overline{\mathrm{W}}[\boldsymbol{k}'] = \big\langle \overline{\psi},\, \phi^{(s)}_{\boldsymbol{k}'}\big\rangle = \big(\overline{\psi} * \phi^{(s)}\big)(s\boldsymbol{k}')$. Therefore we get

$$\begin{aligned}(f_{\mathrm{X}} * \overline{\psi}_{\mathrm{W}})\,(ms\boldsymbol{n}) &= \sum_{\boldsymbol{k}\in\mathbb{Z}^2} \mathrm{X}[\boldsymbol{k}]\,\overline{\mathrm{W}}[m\boldsymbol{n}-\boldsymbol{k}]\\ &= \big(\mathrm{X} * \overline{\mathrm{W}}\big)[m\boldsymbol{n}],\end{aligned}$$

hence the result. ∎

APPENDIX E
PROOF OF LEMMA 3

Proof: Let $\boldsymbol{u} \in \mathbb{R}^2$. By definition of f_{X}, $f_{\mathcal{T}_{\boldsymbol{u}}\mathrm{X}}$ and $\mathcal{T}_{\boldsymbol{u}}\mathrm{X}$,

$$f_{\mathcal{T}_{\boldsymbol{u}}\mathrm{X}} = \sum_{\boldsymbol{n}\in\mathbb{Z}^2} s\,\mathcal{T}_{s\boldsymbol{u}} f_{\mathrm{X}}(s\boldsymbol{n})\,\phi^{(s)}_{\boldsymbol{n}}. \tag{93}$$

By construction, $f_{\mathrm{X}} \in \mathcal{U}_s$. Therefore, $\mathcal{T}_{s\boldsymbol{u}} f_{\mathrm{X}} \in \mathcal{U}_s$. Then, the Shannon-Whittaker theorem [48, p. 61], coupled with the orthonormality of $\{\phi^{(s)}_{\boldsymbol{n}}\}_{\boldsymbol{n}\in\mathbb{Z}^2}$, implies

$$s\,\mathcal{T}_{s\boldsymbol{u}} f_{\mathrm{X}}(s\boldsymbol{n}) = \big\langle \mathcal{T}_{s\boldsymbol{u}} f_{\mathrm{X}},\, \phi^{(s)}_{\boldsymbol{n}}\big\rangle. \tag{94}$$

Finally, plugging (94) into (93) concludes the proof. ∎

APPENDIX F
PROOF OF THEOREM 1

Proof: We consider

$$f_0 : \boldsymbol{x} \mapsto (f_{\mathrm{X}} * \overline{\psi}_{\mathrm{W}})(\boldsymbol{x})\,\mathrm{e}^{i\langle \boldsymbol{\xi}/s,\, \boldsymbol{x}\rangle}, \tag{95}$$

with f_{X} and ψ_{W} satisfying (11). Then,

$$|f_{\mathrm{X}} * \overline{\psi}_{\mathrm{W}}| = |f_0| \quad \text{and} \quad |\mathcal{T}_{s\boldsymbol{u}} f_{\mathrm{X}} * \overline{\psi}_{\mathrm{W}}| = |\mathcal{T}_{s\boldsymbol{u}} f_0|. \tag{96}$$

According to Proposition 2, $\psi_{\mathrm{W}} \in \mathcal{V}\big(\boldsymbol{\xi}/s,\, \kappa/s\big)$. Therefore, according to Lemma 1, $\operatorname{supp} \widehat{f}_0 \subset \mathcal{B}_\infty\big(\frac{\kappa}{2s}\big)$. Moreover, by hypothesis, $\kappa \leq 2\pi/m$; thus, $\mathcal{B}_\infty\big(\frac{\kappa}{2s}\big) \subset \mathcal{B}_\infty\big(\frac{\pi}{ms}\big)$. Therefore, $f_0 \in \mathcal{U}_{s'}$, where we have denoted $s' := ms$.

According to (13) (Proposition 2), (18) (Corollary 1) and (96), we get

$$\mathcal{C}_m \mathrm{X}[\boldsymbol{n}] = |f_0(s'\boldsymbol{n})|\,; \tag{97}$$

$$\mathcal{C}_m(\mathcal{T}_{\boldsymbol{u}}\mathrm{X})[\boldsymbol{n}] = |(\mathcal{T}_{s\boldsymbol{u}} f_0)(s'\boldsymbol{n})|\,. \tag{98}$$

Then, using the reverse triangular inequality on (97) and (98),

$$\begin{aligned}\big\|\mathcal{C}_m(\mathcal{T}_{\boldsymbol{u}}\mathrm{X}) - \mathcal{C}_m\mathrm{X}\big\|_2^2 &\leq \sum_{\boldsymbol{n}\in\mathbb{Z}^2} \big|\mathcal{T}_{s\boldsymbol{u}} f_0(s'\boldsymbol{n}) - f_0(s'\boldsymbol{n})\big|^2\\ &= \sum_{\boldsymbol{n}\in\mathbb{Z}^2} |g(s'\boldsymbol{n})|^2 = \frac{1}{s'^2}\,\|\mathrm{Y}\|_2^2\,,\end{aligned}$$

where we have denoted, for any $\boldsymbol{n} \in \mathbb{Z}^2$, $g := \mathcal{T}_{s\boldsymbol{u}} f_0 - f_0$ and $\mathrm{Y}[\boldsymbol{n}] := s'g(s'\boldsymbol{n})$. We have $g \in \mathcal{U}_{s'}$ since $f_0 \in \mathcal{U}_{s'}$. Then, according to (10) in Lemma 2, $\|\mathrm{Y}\|_2 = \|g\|_{L^2}$. Therefore,

$$\big\|\mathcal{C}_m(\mathcal{T}_{\boldsymbol{u}}\mathrm{X}) - \mathcal{C}_m\mathrm{X}\big\|_2^2 \leq \frac{1}{s'^2}\,\|g\|_{L^2}^2 = \frac{1}{s'^2}\,\|\mathcal{T}_{s\boldsymbol{u}} f_0 - f_0\|_{L^2}^2\,.$$

According to Proposition 1 with $\varepsilon \leftarrow \kappa/s$ and $\boldsymbol{h} \leftarrow s\boldsymbol{u}$, we then get the following bound:

$$\big\|\mathcal{C}_m(\mathcal{T}_{\boldsymbol{u}}\mathrm{X}) - \mathcal{C}_m\mathrm{X}\big\|_2^2 \leq \frac{\alpha(\kappa\boldsymbol{u})^2}{s'^2}\,\|f_0\|_{L^2}^2\,. \tag{99}$$

Besides, according again to Lemma 2, $\|f_0\|_{L^2}^2 = \|\mathrm{X}_0\|_2^2$, where $\mathrm{X}_0[\boldsymbol{n}] := s' f_0(s'\boldsymbol{n})$ for any $\boldsymbol{n} \in \mathbb{Z}^2$. Therefore, according to (97),

$$\|\mathcal{C}_m\mathrm{X}\|_2 = \frac{1}{s'}\,\|\mathrm{X}_0\|_2 = \frac{1}{s'}\,\|f_0\|_{L^2}\,. \tag{100}$$

Finally, plugging (100) into (99) completes the proof. ∎

APPENDIX G
PROOF OF PROPOSITION 3

Proof: Let $\mathrm{X} \in l^2_{\mathbb{R}}(\mathbb{Z}^2)$ and $s > 0$. We consider $f_0 \in L^2(\mathbb{R}^2)$ as the "low-frequency" function satisfying (95). Again, we introduce $s' := ms$ and $\mathrm{X}_0 \in l^2(\mathbb{Z}^2)$ such that $\mathrm{X}_0[\boldsymbol{n}] := s' f_0(s'\boldsymbol{n})$. Moreover, for any $\mathrm{Y} \in l^2_{\mathbb{R}}(\mathbb{Z}^2)$, we denote:

$$f^{(s')}_{\mathrm{Y}} := \sum_{\boldsymbol{n}\in\mathbb{Z}^2} \mathrm{Y}[\boldsymbol{n}]\,\phi^{(s')}_{\boldsymbol{n}}. \tag{101}$$

On the one hand, in the proof of Theorem 1, we already got (100). On the other hand, (98) can be rewritten

$$\mathcal{C}_m(\mathcal{T}_{\boldsymbol{u}}\mathrm{X})[\boldsymbol{n}] = \big|\mathcal{T}_{s'\boldsymbol{u}'} f_0(s'\boldsymbol{n})\big|, \tag{102}$$

with $\boldsymbol{u}' := \boldsymbol{u}/m$. Besides, according to the proof of Theorem 1, $f_0 \in \mathcal{U}_{s'}$. Thus, by definition of X_0, the Shannon-Whittaker theorem [48, p. 61] implies that $f_0 = f^{(s')}_{\mathrm{X}_0}$ such as defined in (101). Then, using Lemma 3 with $\mathrm{X} \leftarrow \mathrm{X}_0$, $\boldsymbol{u} \leftarrow \boldsymbol{u}'$ and $s \leftarrow s'$, we get

$$f^{(s')}_{\mathcal{T}_{\boldsymbol{u}'}\mathrm{X}_0} = \mathcal{T}_{s'\boldsymbol{u}'} f^{(s')}_{\mathrm{X}_0} = \mathcal{T}_{s'\boldsymbol{u}'} f_0. \tag{103}$$

Then, using (12) in Proposition 2 with $\mathrm{X} \leftarrow \mathcal{T}_{\boldsymbol{u}'}\mathrm{X}_0$ and $s \leftarrow s'$, and inserting (103) into the result yields

$$\mathcal{T}_{\boldsymbol{u}'}\mathrm{X}_0[\boldsymbol{n}] = s'\, f^{(s')}_{\mathcal{T}_{\boldsymbol{u}'}\mathrm{X}_0}(s'\boldsymbol{n}) = s'\,\mathcal{T}_{s'\boldsymbol{u}'} f_0(s'\boldsymbol{n}). \tag{104}$$

Therefore, (102) and (104) imply

$$\|\mathcal{C}_m(\mathcal{T}_{\boldsymbol{u}}\mathrm{X})\|_2 = \frac{1}{s'}\,\|\mathcal{T}_{\boldsymbol{u}'}\mathrm{X}_0\|_2\,. \tag{105}$$

Moreover, since $f_0 \in \mathcal{U}_{s'}$, and according to (104), we can use Lemma 2 with $s \leftarrow s'$, $g \leftarrow \mathcal{T}_{s'\boldsymbol{u}'}f_0$ and $\mathrm{Y} \leftarrow \mathcal{T}_{\boldsymbol{u}'}\mathrm{X}_0$. We get

$$\|\mathcal{T}_{\boldsymbol{u}'}\mathrm{X}_0\|_2 = \|\mathcal{T}_{s'\boldsymbol{u}'}f_0\|_{L^2} = \|f_0\|_{L^2}\,. \tag{106}$$

Finally, (100), (105) and (106) imply $\|\mathcal{C}_m(\mathcal{T}_{\boldsymbol{u}}\mathrm{X})\|_2 = \|\mathcal{C}_m\mathrm{X}\|_2$, which concludes the proof. ∎

APPENDIX H
JUSTIFICATION FOR HYPOTHESES 1 AND 2

Given $n \in \mathbb{N}^*$, we define *n-th order stationarity* of a given stochastic process F as in [55, p. 152]: for any $n' \in \{1\,..\,n\}$, $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n'}) \in (\mathbb{R}^2)^{n'}$ and $\boldsymbol{h} \in \mathbb{R}^2$, the joint distribution of $\big(F(\boldsymbol{x}_1), \ldots, F(\boldsymbol{x}_{n'})\big)$ is identical to the one of $\big(F(\boldsymbol{x}_1 + \boldsymbol{h}), \ldots, F(\boldsymbol{x}_{n'} + \boldsymbol{h})\big)$. Besides, *strict-sense stationarity* is defined as n-th order stationarity for any $n \in \mathbb{N}^*$.

We recall that $\boldsymbol{\nu} := \boldsymbol{\xi}/s$. We then state the following results.

Proposition 8: We assume that F_{X} is first-order stationary. If, for any $\boldsymbol{x} \in \mathbb{R}^2$ and any $\boldsymbol{h} \in \mathcal{B}_2(2\pi/\left\|\boldsymbol{\nu}\right\|_2)$,

$$(\mathcal{T}_{\boldsymbol{h}}F_{\mathrm{X}} * \overline{\psi})(\boldsymbol{x}) = \mathrm{e}^{i\langle \boldsymbol{\nu},\, \boldsymbol{h}\rangle}(F_{\mathrm{X}} * \overline{\psi})(\boldsymbol{x}), \tag{107}$$

then Hypothesis 1 is satisfied.

Proof: We show that the probability measure of $Z_{\mathrm{X}}(\boldsymbol{x})$ is invariant with respect to phase shift. In other words, we show that, for any measurable set $\mathfrak{A} \subset \mathbb{S}^1$,

$$\forall \omega \in [0,\, 2\pi]\,, \mu(\mathfrak{A}) = \mu(\mathrm{e}^{i\omega}\mathfrak{A}), \tag{108}$$

where we have denoted

$$\mu : \mathfrak{A} \mapsto \mathbb{P}\left\{Z_{\mathrm{X}}(\boldsymbol{x}) \in \mathfrak{A}\right\}.$$

Let $\boldsymbol{h} \in \mathcal{B}_2(2\pi/\left\|\boldsymbol{\nu}\right\|_2)$. According to (107),

$$Z_{\mathrm{X}}(\boldsymbol{x}) \in \mathfrak{A} \iff \mathcal{T}_{\boldsymbol{h}}Z_{\mathrm{X}}(\boldsymbol{x}) \in \mathrm{e}^{i\langle \boldsymbol{\nu},\, \boldsymbol{h}\rangle}\mathfrak{A}.$$

Therefore,

$$\mathbb{P}\left\{Z_{\mathrm{X}}(\boldsymbol{x}) \in \mathfrak{A}\right\} = \mathbb{P}\big\{\mathcal{T}_{\boldsymbol{h}}Z_{\mathrm{X}}(\boldsymbol{x}) \in \mathrm{e}^{i\langle \boldsymbol{\nu},\, \boldsymbol{h}\rangle}\mathfrak{A}\big\}.$$

Since F_{X} is first-order stationary, $Z_{\mathrm{X}}(\boldsymbol{x})$ and $\mathcal{T}_{\boldsymbol{h}}Z_{\mathrm{X}}(\boldsymbol{x})$ have the same probability distribution. Thus we get

$$\mathbb{P}\left\{Z_{\mathrm{X}}(\boldsymbol{x}) \in \mathfrak{A}\right\} = \mathbb{P}\big\{Z_{\mathrm{X}}(\boldsymbol{x}) \in \mathrm{e}^{i\langle \boldsymbol{\nu},\, \boldsymbol{h}\rangle}\mathfrak{A}\big\}.$$

Let $\omega \in [0,\, 2\pi]$. Considering $\boldsymbol{h} := \omega\,\boldsymbol{\nu}/\left\|\boldsymbol{\nu}\right\|_2^2$, we have $\boldsymbol{h} \in \mathcal{B}_2(2\pi/\left\|\boldsymbol{\nu}\right\|_2)$ and $\langle \boldsymbol{\nu},\, \boldsymbol{h}\rangle = \omega$. Therefore,

$$\forall \omega \in [0,\, 2\pi]\,, \mathbb{P}\left\{Z_{\mathrm{X}}(\boldsymbol{x}) \in \mathfrak{A}\right\} = \mathbb{P}\big\{Z_{\mathrm{X}}(\boldsymbol{x}) \in \mathrm{e}^{i\omega}\mathfrak{A}\big\},$$

which yields (108).

Any probability measure defined on $\mathbb{S}^1$ is a Radon measure. Therefore, according to Haar's theorem [56], there exists a unique probability measure on $\mathbb{S}^1$ satisfying (108). Since the uniform probability measure is also invariant to phase shift, we deduce that $Z_{\mathrm{X}}(\boldsymbol{x})$ is uniformly distributed on $\mathbb{S}^1$. ∎

Proposition 9: We assume the conditions of Proposition 8 are met. If, moreover, F_{X} is strict-sense stationary, then Hypothesis 2 is satisfied.

Proof: Let $n \in \mathbb{N}^*$ and $\boldsymbol{x}, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n \in \mathbb{R}^2$. To alleviate notations, we consider the random vector $\boldsymbol{M} = (M_{\mathrm{X}}(\boldsymbol{y}_i))_{i\in\{1..n\}}$ with outcomes in $\mathbb{R}_+^n$.

The proof is organized as follows. Using a similar reasoning as Proposition 8, we show that Z_{X} follows a uniform conditional probability distribution given $\boldsymbol{M}$. Since we already know that Z_{X} follows a uniform (unconditional) distribution, we deduce that Z_{X} and $\boldsymbol{M}$ are independent.

Let $\mathfrak{A} \subset \mathbb{S}^1$ and $\mathfrak{S} := (\mathfrak{S}_i)_{i\in\{1..n\}} \subset \mathbb{R}_+^n$ denote measurable sets. Let $\boldsymbol{h} \in \mathcal{B}_2(2\pi/\left\|\boldsymbol{\nu}\right\|_2)$. According to (107),

$$\begin{aligned} Z_{\mathrm{X}}(\boldsymbol{x}) \in \mathfrak{A} &\iff \mathcal{T}_{\boldsymbol{h}}Z_{\mathrm{X}}(\boldsymbol{x}) \in \mathrm{e}^{i\langle \boldsymbol{\nu},\, \boldsymbol{h}\rangle}\mathfrak{A};\\ M_{\mathrm{X}}(\boldsymbol{y}_i) \in \mathfrak{S}_i &\iff \mathcal{T}_{\boldsymbol{h}}M_{\mathrm{X}}(\boldsymbol{y}_i) \in \mathfrak{S}_i \quad \forall i \in \{1\,..\,n\}\,. \end{aligned}$$

Therefore,

$$\begin{aligned} &\mathbb{P}\big\{(Z_{\mathrm{X}}(\boldsymbol{x}) \in \mathfrak{A}) \,\&\, (\boldsymbol{M} \in \mathfrak{S})\big\} =\\ &\qquad \mathbb{P}\big\{\big(\mathcal{T}_{\boldsymbol{h}}Z_{\mathrm{X}}(\boldsymbol{x}) \in \mathrm{e}^{i\langle \boldsymbol{\nu},\, \boldsymbol{h}\rangle}\mathfrak{A}\big) \,\&\, (\mathcal{T}_{\boldsymbol{h}}\boldsymbol{M} \in \mathfrak{S})\big\}. \end{aligned}$$

Since F_{X} is strict-sense stationary, the joint probability density of $\mathcal{T}_{\boldsymbol{h}}Z_{\mathrm{X}}(\boldsymbol{x})$, $\mathcal{T}_{\boldsymbol{h}}M_{\mathrm{X}}(\boldsymbol{y}_1)$, …, $\mathcal{T}_{\boldsymbol{h}}M_{\mathrm{X}}(\boldsymbol{y}_n)$ is identical to the one of $Z_{\mathrm{X}}(\boldsymbol{x})$, $M_{\mathrm{X}}(\boldsymbol{y}_1)$, …, $M_{\mathrm{X}}(\boldsymbol{y}_n)$. Therefore we get

$$\begin{aligned} &\mathbb{P}\big\{(Z_{\mathrm{X}}(\boldsymbol{x}) \in \mathfrak{A}) \,\&\, (\boldsymbol{M} \in \mathfrak{S})\big\} =\\ &\qquad \mathbb{P}\big\{\big(Z_{\mathrm{X}}(\boldsymbol{x}) \in \mathrm{e}^{i\langle \boldsymbol{\nu},\, \boldsymbol{h}\rangle}\mathfrak{A}\big) \,\&\, (\boldsymbol{M} \in \mathfrak{S})\big\}. \end{aligned}$$

We assume that $\mathbb{P}(\boldsymbol{M} \in \mathfrak{S}) > 0$. According to the above expression, and similarly to the proof of Proposition 8, we get,

$$\begin{aligned} &\forall \omega \in [0,\, 2\pi]\,, \mathbb{P}\left\{Z_{\mathrm{X}}(\boldsymbol{x}) \in \mathfrak{A} \mid \boldsymbol{M} \in \mathfrak{S}\right\} =\\ &\qquad \mathbb{P}\left\{Z_{\mathrm{X}}(\boldsymbol{x}) \in \mathrm{e}^{i\omega}\mathfrak{A} \;\middle|\; \boldsymbol{M} \in \mathfrak{S}\right\}. \end{aligned}$$

Then, the above conditional probability measure satisfies phase shift invariance (108). Therefore, as in the proof of Proposition 8, Haar's theorem implies that $Z_{\mathrm{X}}(\boldsymbol{x})$ follows a uniform conditional distribution given $\boldsymbol{M} \in \mathfrak{S}$.

Moreover, strict-sense implies first-order stationarity, and thus, according to Proposition 8, $Z_{\mathrm{X}}(\boldsymbol{x})$ follows a uniform (unconditional) distribution. Therefore we get, for any measurable sets $\mathfrak{A} \subset \mathbb{S}^1$ and $\mathfrak{S} \subset \mathbb{R}_+^n$ such that $\mathbb{P}(\boldsymbol{M} \in \mathfrak{S}) > 0$,

$$\mathbb{P}\left\{Z_{\mathrm{X}} \in \mathfrak{A} \mid \boldsymbol{M} \in \mathfrak{S}\right\} = \mathbb{P}(Z_{\mathrm{X}} \in \mathfrak{A}),$$

which proves independence between $Z_{\mathrm{X}}(\boldsymbol{x})$ and $\boldsymbol{M}$. ∎

Strict-sense stationarity suggests that any translated version of a given image is equally likely. In reality, this statement is too strong, for several reasons. First, by construction, X has all its realizations in $L^2_{\mathbb{R}}(\mathbb{R}^2)$. In that context, a stationary process yields outcomes which are zero almost everywhere. Besides, depending on which category the image belongs to, the pixel distribution is likely to vary across various regions. For instance, we can expect the main subject to be located at the center of the image. More details on statistical properties of natural images can be found in [57]. Nevertheless, this hypothesis will be considered as a reasonable approximation if the shift is much smaller than the image "characteristic" size in the continuous domain; i.e., if $\|\boldsymbol{h}\|_2 \ll sM$, where, as a reminder, M denotes the support size of input images.[12]

[12]We refer readers to [58] for a related notion of local stationarity.

As it turns out, the proofs of Propositions 8 and 9 only requires shifts with $\|\boldsymbol{h}\|_2 \leq 2\pi/\|\boldsymbol{\nu}\|_2$. Then, according to (53), $\|\boldsymbol{h}\|_2 \ll sM$, and the stationarity hypothesis holds.

Finally, to justify (107), we consider $\varphi_{\mathrm{W}} : \boldsymbol{x} \mapsto \psi_{\mathrm{W}}(\boldsymbol{x})\mathrm{e}^{-i\langle \boldsymbol{\nu},\, \boldsymbol{x}\rangle}$. Similarly to Lemma 1, we can show that φ_{W} is a low-pass filter, with $\operatorname{supp}\widehat{\varphi}_{\mathrm{W}} \subset \mathcal{B}_\infty(\varepsilon/2)$. For all $\boldsymbol{h} \in \mathbb{R}^2$ such that $\|\boldsymbol{h}\|_2 \leq 2\pi/\|\boldsymbol{\nu}\|_2$, we have

$$\begin{aligned}(\mathcal{T}_{\boldsymbol{h}} F_{\mathrm{X}} * \overline{\psi})(\boldsymbol{x}) &= \iint_{\mathbb{R}^2} \mathcal{T}_{\boldsymbol{h}} F_{\mathrm{X}}(\boldsymbol{x}-\boldsymbol{y})\,\overline{\varphi}(\boldsymbol{y})\,\mathrm{e}^{-i\langle \boldsymbol{\nu},\, \boldsymbol{y}\rangle}\,\mathrm{d}^2\boldsymbol{y} \\ &= \mathrm{e}^{i\langle \boldsymbol{\nu},\, \boldsymbol{h}\rangle} \iint_{\mathbb{R}^2} F_{\mathrm{X}}(\boldsymbol{x}-\boldsymbol{y}')\,\overline{\varphi}(\boldsymbol{y}'-\boldsymbol{h})\,\mathrm{e}^{-i\langle \boldsymbol{\nu},\, \boldsymbol{y}'\rangle}\,\mathrm{d}^2\boldsymbol{y}'.\end{aligned}$$

Since $\operatorname{supp}\widehat{\varphi}_{\mathrm{W}} \subset \mathcal{B}_\infty\big(\frac{\kappa}{2s}\big)$, we can define a "minimal wavelength" $\lambda_{\varphi_{\mathrm{W}}} := 2\pi s/\kappa$. Then, if $\|\boldsymbol{h}\|_2 \ll \lambda_{\varphi_{\mathrm{W}}}$, we can approximate $\overline{\varphi}(\boldsymbol{y}'-\boldsymbol{h}) \approx \overline{\varphi}(\boldsymbol{y}')$. This sufficient condition is actually met, because $\|\boldsymbol{h}\|_2 \leq 2\pi/\|\boldsymbol{\nu}\|_2$ and, according to (54), $\|\boldsymbol{\nu}\|_2 \gg \kappa/s$. Therefore,

$$(\mathcal{T}_{\boldsymbol{h}} F_{\mathrm{X}} * \overline{\psi})(\boldsymbol{x}) \approx \mathrm{e}^{i\langle \boldsymbol{\nu},\, \boldsymbol{h}\rangle}(F_{\mathrm{X}} * \overline{\psi})(\boldsymbol{x}). \tag{109}$$

As a result, the conditions for Propositions 8 and 9 are approximately satisfied. We will therefore consider Hypotheses 1 and 2 as reasonable.

Appendix I
Justification for Hypothesis 3

We consider, for any $l \in \mathcal{L}$ and any $k \in \{0\,..\,K-1\}$, the value of $\mu \in \mathbb{R}$ minimizing $\big\|\mu\widetilde{\mathrm{V}}_l - \mathrm{V}_{lk}\big\|_2^2$, denoted by μ_{lk}. We then denote by $\delta_{lk} := \big\|\mu_{lk}\widetilde{\mathrm{V}}_l - \mathrm{V}_{lk}\big\|_2^2 / \|\mathrm{V}_{lk}\|_2^2$ the relative quadratic error between V_{lk} and its projection on $\mathbb{R}\widetilde{\mathrm{V}}_l$. We get

$$\delta_{lk} = 1 - \frac{\langle \widetilde{\mathrm{V}}_l,\, \mathrm{V}_{lk}\rangle^2}{\big\|\widetilde{\mathrm{V}}_l\big\|_2^2 \cdot \|\mathrm{V}_{lk}\|_2^2}. \tag{110}$$

Expression (67) holds if and only if $\delta_{lk} = 0$ for any $k \in \{0\,..\,K-1\}$. In the case of AlexNet, when $l \in \mathcal{L}$, δ_{lk} do not exceed 10^{-1}, and its mean value is around 10^{-2}. Therefore $\delta_{lk} \ll 1$, and Hypothesis 3 can be considered as a reasonable assumption for any Gabor channel $l \in \mathcal{L}$. Similar observations can be drawn for ResNet.

Appendix J
Proof of Proposition 4

Proof: We consider the Borel σ-algebra on $\mathbb{S}^1$ generated by $\big\{[z,\, z']_{\mathbb{S}^1} \;\big|\; z,\, z' \in \mathbb{S}^1\big\} \cup \{\mathbb{S}^1\}$, on which we have defined the angular measure ϑ such that $\vartheta(\mathbb{S}^1) := 2\pi$, and

$$\forall z,\, z' \in \mathbb{S}^1,\ \vartheta([z,\, z']_{\mathbb{S}^1}) := \angle(z^* z').$$

For any $p \in \mathbb{N}^*$, we compute the p-th moment of $G_{\max\mathrm{X}}(\boldsymbol{x})$ defined in (51). By considering

$$\begin{aligned} h_{\max} : \mathbb{S}^1 &\to [-1,\, 1] \\ z &\mapsto \max_{\|\boldsymbol{k}\|_\infty \leq q} \mathrm{Re}(z^* z_{\boldsymbol{k}}),\end{aligned}$$

we get $G_{\max\mathrm{X}}(\boldsymbol{x}) = h_{\max}(Z_{\mathrm{X}}(\boldsymbol{x}))$. According to Hypothesis 1, $Z_{\mathrm{X}}(\boldsymbol{x})$ follows a uniform distribution on $\mathbb{S}^1$. Therefore,

$$\mathbb{E}\left[G_{\max\mathrm{X}}(\boldsymbol{x})^p\right] = \frac{1}{2\pi}\int_{\mathbb{S}^1} h_{\max}(z)^p\,\mathrm{d}\vartheta(z),$$

which proves that $\mathbb{E}\left[G_{\max\mathrm{X}}(\boldsymbol{x})^p\right]$ does not depend on $\boldsymbol{x}$. Let us split the unit circle $\mathbb{S}^1$ into the arcs $\mathfrak{A}_0^{(q)}, \ldots, \mathfrak{A}_{N_q-1}^{(q)}$ such as defined in (47):

$$\mathbb{E}\left[G_{\max\mathrm{X}}(\boldsymbol{x})^p\right] = \frac{1}{2\pi}\sum_{i=0}^{N_q-1}\int_{\mathfrak{A}_i^{(q)}} h_{\max}(z)^p\,\mathrm{d}\vartheta(z). \tag{111}$$

Let $i \in \{0\,..\,N_q-1\}$. We can show that

$$\forall z \in \mathfrak{A}_i^{(q)},\ h_{\max}(z) = \max\Big(\mathrm{Re}\left(z^* z_i^{(q)}\right),\, \mathrm{Re}\left(z^* z_{i+1}^{(q)}\right)\Big).$$

Therefore, $h_{\max}$ is symmetric with respect to the center value of $\mathfrak{A}_i^{(q)}$, denoted by $\overline{z}_i^{(q)}$, where $h_{\max}$ reaches its minimum. We denote by $\overline{\mathfrak{A}}_i^{(q)} := \big[z_i^{(q)},\, \overline{z}_i^{(q)}\big]_{\mathbb{S}^1}$ the first half of arc $\mathfrak{A}_i^{(q)}$. Then,

$$\forall z \in \overline{\mathfrak{A}}_i^{(q)},\ h_{\max}(z) = \mathrm{Re}(z^* z_i^{(q)}).$$

As a consequence, using symmetry, we get

$$\begin{aligned}\int_{\mathfrak{A}_i^{(q)}} h_{\max}(z)^p\,\mathrm{d}\vartheta(z) &= 2\int_{\overline{\mathfrak{A}}_i^{(q)}} h_{\max}(z)^p\,\mathrm{d}\vartheta(z) \\ &= 2\int_{\overline{\mathfrak{A}}_i^{(q)}} \mathrm{Re}\big(z^* z_i^{(q)}\big)^p\,\mathrm{d}\vartheta(z).\end{aligned}$$

By using the change of variable formula [59, p. 81] with $z \leftarrow \mathrm{e}^{i\theta}$, we get

$$\int_{\mathfrak{A}_i^{(q)}} h_{\max}(z)^p\,\mathrm{d}\vartheta(z) = 2\int_{\theta_i^{(q)}}^{\overline{\theta}_i^{(q)}} \cos^p\big(\theta - \theta_i^{(q)}\big)\,\mathrm{d}\theta,$$

where $\overline{\theta}_i^{(q)} := \big(\theta_i^{(q)} + \theta_{i+1}^{(q)}\big)/2$ denotes the argument of $\overline{z}_i^{(q)}$. Then, the change of variable $\omega \leftarrow \theta - \theta_i^{(q)}$ yields

$$\int_{\mathfrak{A}_i^{(q)}} h_{\max}(z)^p\,\mathrm{d}\vartheta(z) = 2\int_0^{\omega_i^{(q)}/2} \cos^p\omega\,\mathrm{d}\omega. \tag{112}$$

Now, we insert (112) into (111), and compute $\mathbb{E}\left[G_{\max\mathrm{X}}(\boldsymbol{x})^p\right]$ for $p = 1$ and $p = 2$. We get

$$\mathbb{E}\left[G_{\max\mathrm{X}}(\boldsymbol{x})\right] = \frac{1}{\pi}\sum_{i=0}^{N_q-1}\sin\frac{\omega_i^{(q)}}{2};$$

$$\mathbb{E}\left[G_{\max\mathrm{X}}(\boldsymbol{x})^2\right] = \frac{1}{2} + \frac{1}{4\pi}\sum_{i=0}^{N_q-1}\sin\omega_i^{(q)}.$$

We recall that $Q_{\mathrm{X}} := 1 - G_{\max\mathrm{X}}$. By linearity of $\mathbb{E}$, we get

$$\mathbb{E}\left[Q_{\mathrm{X}}(\boldsymbol{x})^2\right] := \frac{3}{2} + \frac{1}{4\pi}\sum_{i=0}^{N_q-1}\left(\sin\omega_i^{(q)} - 8\sin\frac{\omega_i^{(q)}}{2}\right).$$

Using the notation introduced in (48) concludes the proof. ■

Appendix K
Proof of Proposition 5

Proof: We suppose that Hypothesis 2 is satisfied and we consider $\boldsymbol{x} \in \mathbb{R}^2$. For a given $n \in \mathbb{N}^*$, we introduce the random variable

$$\widetilde{S}_{\mathrm{X}}^{(n)} := \sqrt{\sum_{\|\boldsymbol{k}\|_\infty \leq n} M_{\mathrm{X}}(\boldsymbol{x}_{\boldsymbol{k}})^2}.$$

According to Hypothesis 2, $Z_{\mathrm{X}}(\boldsymbol{x})$ is jointly independent of $M_{\mathrm{X}}(\boldsymbol{x_k})$ for $\boldsymbol{k} \in \{-n\,..\,n\}^2$. Therefore, by composition, $Z_{\mathrm{X}}(\boldsymbol{x})$ is also independent of $\widetilde{S}_{\mathrm{X}}^{(n)}$. Moreover, according to (52) and (60), $\widetilde{S}_{\mathrm{X}}^{(n)}$ converges almost surely towards $\widetilde{S}_{\mathrm{X}}$, which proves independence between $Z_{\mathrm{X}}(\boldsymbol{x})$ and $\widetilde{S}_{\mathrm{X}}$.

Now, we prove conditional independence between $Z_{\mathrm{X}}(\boldsymbol{x})$ and $M_{\mathrm{X}}(\boldsymbol{x})$ given $\widetilde{S}_{\mathrm{X}}$. According to Hypothesis 2,

$$\left(M_{\mathrm{X}}(\boldsymbol{x}),\, \widetilde{S}_{\mathrm{X}}^{(n)}\right) \perp\!\!\!\perp Z_{\mathrm{X}}(\boldsymbol{x}),$$

where $\perp\!\!\!\perp$ stands for independence. This is because $\widetilde{S}_{\mathrm{X}}^{(n)}$ only depends on a finite number of $M_{\mathrm{X}}(\boldsymbol{x_k})$. Therefore,

$$Z_{\mathrm{X}}(\boldsymbol{x}) \perp\!\!\!\perp M_{\mathrm{X}}(\boldsymbol{x}) \,\big|\, \widetilde{S}_{\mathrm{X}}^{(n)}.$$

Finally, since $\widetilde{S}_{\mathrm{X}}^{(n)}$ converges almost surely towards $\widetilde{S}_{\mathrm{X}}$, it comes that $Z_{\mathrm{X}}(\boldsymbol{x})$ and $M_{\mathrm{X}}(\boldsymbol{x})$ are conditionally independent given $\widetilde{S}_{\mathrm{X}}$. ∎

APPENDIX L
PROOF OF THEOREM 2

Proof: We consider $\boldsymbol{n} \in \mathbb{Z}^2$. By construction, $Q_{\mathrm{X}}(\boldsymbol{x_n}) := 1 - G_{\max\mathrm{X}}(\boldsymbol{x_n})$ only depends on $Z_{\mathrm{X}}(\boldsymbol{x_n})$. Therefore, according to Proposition 5, we have

$$Q_{\mathrm{X}}(\boldsymbol{x_n}) \perp\!\!\!\perp M_{\mathrm{X}}(\boldsymbol{x_n}) \,|\, \widetilde{S}_{\mathrm{X}}^2 \quad \text{and} \quad Q_{\mathrm{X}}(\boldsymbol{x_n}) \perp\!\!\!\perp \widetilde{S}_{\mathrm{X}}^2. \tag{113}$$

Besides, we introduce $\widetilde{\Delta}_{\mathrm{X}} := \|\delta_{m,\,q}\mathrm{X}\|_2$, where $\delta_{m,\,q}\mathrm{X}$ is defined in (56). Then, using the linearity of $\mathbb{E}$, and according to (52), we get, for any $\sigma > 0$,

$$\mathbb{E}\Big[\widetilde{\Delta}_{\mathrm{X}}^2 \,\Big|\, \widetilde{S}_{\mathrm{X}}^2 = \sigma\Big] = \sum_{\boldsymbol{n}\in\mathbb{Z}^2} \mathbb{E}\Big[M_{\mathrm{X}}(\boldsymbol{x_n})^2\, Q_{\mathrm{X}}(\boldsymbol{x_n})^2 \,\Big|\, \widetilde{S}_{\mathrm{X}}^2 = \sigma\Big].$$

According to (113), we get

$$\mathbb{E}\Big[\widetilde{\Delta}_{\mathrm{X}}^2 \,\Big|\, \widetilde{S}_{\mathrm{X}}^2 = \sigma\Big] = \sum_{\boldsymbol{n}\in\mathbb{Z}^2} \mathbb{E}\Big[M_{\mathrm{X}}(\boldsymbol{x_n})^2 \,\Big|\, \widetilde{S}_{\mathrm{X}}^2 = \sigma\Big] \mathbb{E}\big[Q_{\mathrm{X}}(\boldsymbol{x_n})^2\big].$$

Under Hypothesis 1, we use (58) in Proposition 4, which yields $\mathbb{E}\big[Q_{\mathrm{X}}(\boldsymbol{x_n})^2\big] = \gamma_q(m\boldsymbol{\xi})^2$. Therefore, using again the linearity of $\mathbb{E}$, and according to (52) and (60),

$$\begin{aligned}\mathbb{E}\Big[\widetilde{\Delta}_{\mathrm{X}}^2 \,\Big|\, \widetilde{S}_{\mathrm{X}}^2 = \sigma\Big] &= \mathbb{E}\Big[\widetilde{S}_{\mathrm{X}}^2 \;\Big|\; \widetilde{S}_{\mathrm{X}}^2 = \sigma\Big] \cdot \gamma_q(m\boldsymbol{\xi})^2 \\ &= \sigma \cdot \gamma_q(m\boldsymbol{\xi})^2.\end{aligned}$$

Besides, we can reformulate $\widetilde{Q}_{\mathrm{X}}$ such as defined in (57): $\widetilde{Q}_{\mathrm{X}} = \widetilde{\Delta}_{\mathrm{X}}/\widetilde{S}_{\mathrm{X}}$. Therefore we get

$$\mathbb{E}\Big[\widetilde{Q}_{\mathrm{X}}^2 \;\Big|\; \widetilde{S}_{\mathrm{X}}^2 = \sigma\Big] = \frac{1}{\sigma}\,\mathbb{E}\Big[\widetilde{\Delta}_{\mathrm{X}}^2 \,\Big|\, \widetilde{S}_{\mathrm{X}}^2 = \sigma\Big] = \gamma_q(m\boldsymbol{\xi})^2.$$

Then, law of total expectation states that

$$\mathbb{E}\big[\widetilde{Q}_{\mathrm{X}}^2\big] = \mathbb{E}\left[\mathbb{E}\big[\widetilde{Q}_{\mathrm{X}}^2 \,\big|\, \widetilde{S}_{\mathrm{X}}^2\big]\right] = \gamma_q(m\boldsymbol{\xi})^2.$$

Finally, according to Conjecture 1, we have:

$$\widetilde{P}_{\mathrm{X}} \leq (1 + \beta_q(m\kappa))\, \widetilde{Q}_{\mathrm{X}},$$

which yields (61). ∎

APPENDIX M
PROOF OF THEOREM 3

Proof: Using the triangular inequality, we compute

$$\begin{aligned}\big\|\mathcal{R}_{m,\,q}(\mathcal{T}_{\boldsymbol{u}}\mathrm{X}) - \mathcal{R}_{m,\,q}\mathrm{X}\big\|_2 &\leq \|\mathcal{C}_{2m}(\mathcal{T}_{\boldsymbol{u}}\mathrm{X})\|_2\, \widetilde{P}_{\mathcal{T}_{\boldsymbol{u}}\mathrm{X}} + \|\mathcal{C}_{2m}\mathrm{X}\|_2\, \widetilde{P}_{\mathrm{X}} \\ &\quad + \|\mathcal{C}_{2m}(\mathcal{T}_{\boldsymbol{u}}\mathrm{X}) - \mathcal{C}_{2m}\mathrm{X}\|_2\,,\end{aligned}$$

where $\widetilde{P}_{\mathrm{X}}$ and $\widetilde{P}_{\mathcal{T}_{\boldsymbol{u}}\mathrm{X}}$ are defined in (55). Since, by hypothesis, $\kappa \leq \pi/m$, expression (21) in Proposition 3 states that

$$\|\mathcal{C}_{2m}(\mathcal{T}_{\boldsymbol{u}}\mathrm{X})\|_2 = \|\mathcal{C}_{2m}\mathrm{X}\|_2\,.$$

Moreover, according to (63), we can apply (20) in Theorem 1 on the third term of the above expression. We get

$$\begin{aligned}\big\|\mathcal{R}_{m,\,q}(\mathcal{T}_{\boldsymbol{u}}\mathrm{X}) - \mathcal{R}_{m,\,q}\mathrm{X}\big\|_2 &\leq \Big[\widetilde{P}_{\mathcal{T}_{\boldsymbol{u}}\mathrm{X}} + \widetilde{P}_{\mathrm{X}} + \alpha(\kappa\boldsymbol{u})\Big]\, \|\mathcal{C}_{2m}\mathrm{X}\|_2\,.\end{aligned}$$

Then, by linearity of $\mathbb{E}$, we get

$$\mathbb{E}\big[\widetilde{R}_{\mathrm{X},\,\boldsymbol{u}}\big] \leq \mathbb{E}\big[\widetilde{P}_{\mathcal{T}_{\boldsymbol{u}}\mathrm{X}}\big] + \mathbb{E}\big[\widetilde{P}_{\mathrm{X}}\big] + \alpha(\kappa\boldsymbol{u}). \tag{114}$$

For any stochastic process X' satisfying Hypotheses 1 and 2, expression (61) in Theorem 2 and Jensen's inequality yield:

$$\mathbb{E}\big[\widetilde{P}_{\mathrm{X}'}\big] \leq (1 + \beta_q(m\kappa))\, \gamma_q(m\boldsymbol{\xi}). \tag{115}$$

Since Hypotheses 1 & 2 are satisfied for Z_{X} and M_{X}, Lemma 3 implies that they are also true for $Z_{\mathcal{T}_{\boldsymbol{u}}\mathrm{X}}$ and $M_{\mathcal{T}_{\boldsymbol{u}}\mathrm{X}}$. Therefore, (115) is valid for $\mathrm{X}' \leftarrow \mathrm{X}$ and $\mathrm{X}' \leftarrow \mathcal{T}_{\boldsymbol{u}}\mathrm{X}$, and plugging this expression into (114) concludes the proof. ∎

APPENDIX N
PROOF OF PROPOSITION 6

Proof: This proposition is a simple reformulation of the well-known result that two successive convolutions can be written as another convolution with a wider kernel. We introduce the upsampling operator: $(\mathrm{X} \uparrow m)[\boldsymbol{n}] := \mathrm{X}[\boldsymbol{n}/m]$ if $\boldsymbol{n}/m \in \mathbb{Z}^2$, and 0 otherwise. We also consider the "identity" filter $\mathrm{I} \in l^2(\mathbb{Z}^2)$ such that $\mathrm{I}[\boldsymbol{0}] = 1$ and $\mathrm{I}[\boldsymbol{n}] = 0$ otherwise.

First, for any $\mathrm{U},\, \mathrm{V} \in l^2(\mathbb{Z}^2)$ and any $s,\, t \in \mathbb{N}^*$, we have $((\mathrm{U} \downarrow s) * \mathrm{V}) \downarrow t = (\mathrm{U} * (\mathrm{V} \uparrow s)) \downarrow (st)$. Then, a simple reasoning by induction yields the result, with

$$\mathrm{V}_0^{(0)} := \mathrm{I}; \qquad \mathrm{V}_{4l+k}^{(j+1)} := \mathrm{V}_l^{(j)} * \big(\mathrm{G}_k \uparrow 2^j\big)$$

for any $l \in \{0\,..\,j-1\}$ and any $k \in \{0\,..\,3\}$. ∎

APPENDIX O
PROOF OF PROPOSITION 7

Proof: For any $k \in \{0\,..\,3\}$, Proposition 6 guarantees the existence of $\mathrm{V}_{k,\,l}^{(J)} \in l^2_{\mathbb{R}}(\mathbb{Z}^2)$ such that

$$\mathrm{X}_{k,\,l}^{(J)} = \Big(\mathrm{X} * \overline{\mathrm{V}_{k,\,l}^{(J)}}\Big) \downarrow 2^J. \tag{116}$$

Then, the result is obtained by plugging (116) into (82) for $k \in \{0\,..\,3\}$, and by denoting

$$\begin{pmatrix} \mathrm{W}_l^{(J)} \\ \mathrm{W}_{4^J+l}^{(J)} \end{pmatrix} := \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \mathrm{V}_{0,\,l}^{(J)} \\ \mathrm{V}_{3,\,l}^{(J)} \end{pmatrix} + i \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \mathrm{V}_{2,\,l}^{(J)} \\ \mathrm{V}_{1,\,l}^{(J)} \end{pmatrix}.$$

∎

REFERENCES

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[2] M. Vetterli, "Wavelets, approximation, and compression," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 59–73, Sep. 2001.

[3] A. Laine and J. Fan, "Texture classification by wavelet packet signatures," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1186–1191, Nov. 1993.

[4] S. Pittner and S. V. Kamarthi, "Feature extraction from wavelet coefficients for pattern recognition tasks," *IEEE Trans. Pattern analysis and machine intelligence*, vol. 21, no. 1, pp. 83–88, Jan. 1999.

[5] G. G. Yen, "Wavelet packet feature extraction for vibration monitoring," *IEEE Trans. Industrial Electronics*, vol. 47, no. 3, pp. 650–667, Jun. 2000.

[6] K. Huang and S. Aviyente, "Wavelet feature selection for image classification," *IEEE Trans. Image Processing*, vol. 17, no. 9, pp. 1709–1720, Aug. 2008.

[7] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2323, Nov. 1998.

[8] T. Wiatowski and H. Bölcskei, "A mathematical theory of deep convolutional neural networks for feature extraction," *IEEE Trans. Information Theory*, vol. 64, no. 3, pp. 1845–1866, Mar. 2018.

[9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *ICLR*, 2015.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[11] J. Bruna and S. Mallat, "Invariant scattering convolution networks," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1872–1886, May 2013.

[12] B. Liao and F. Peng, "Rotation-invariant texture features extraction using dual-tree complex wavelet transform," in *Intl. Conf. Information, Networking and Automation (ICINA)*, 2010.

[13] L. Sifre and S. Mallat, "Rotation, scaling and deformation invariant scattering for texture discrimination," in *CVPR*, 2013.

[14] A. Bietti and J. Mairal, "Invariance and stability of deep convolutional representations," in *NeurIPS*, 2017.

[15] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *NeurIPS*, 2014.

[16] M. Rai and P. Rivas, "A review of convolutional neural networks and Gabor filters in object recognition," in *Intl. Conf. Computational Science and Computational Intelligence (CSCI)*, 2020.

[17] S.-Y. Chang and N. Morgan, "Robust CNN-based speech recognition with Gabor filter kernels," in *INTERSPEECH*, 2014.

[18] S. Fujieda, K. Takayama, and T. Hachisuka, "Wavelet convolutional neural networks for texture classification," *arXiv:1707.07394*, Jul. 2017.

[19] S. S. Sarwar, P. Panda, and K. Roy, "Gabor filter assisted energy efficient fast learning convolutional neural networks," in *IEEE/ACM International Symposium on Low Power Electronics and Design*, 2017.

[20] S. Luan, C. Chen, B. Zhang, J. Han, and J. Liu, "Gabor convolutional networks," *IEEE Trans. Image Processing*, vol. 27, no. 9, pp. 4357–4366, May 2018.

[21] P. Liu, H. Zhang, W. Lian, and W. Zuo, "Multi-level wavelet convolutional neural networks," *IEEE Access*, vol. 7, pp. 74 973–74 985, Jun. 2019.

[22] M. Ulicny, V. A. Krylov, and R. Dahyot, "Harmonic networks for image classification," in *BMVC*, 2019.

[23] A. Azulay and Y. Weiss, "Why do deep convolutional networks generalize so poorly to small image transformations?" *JMLR*, vol. 20, no. 184, pp. 1–25, 2019.

[24] R. Zhang, "Making convolutional networks shift-invariant again," in *ICML*, 2019.

[25] C. Vasconcelos, H. Larochelle, V. Dumoulin, N. L. Roux, and R. Goroshin, "An effective anti-aliasing approach for residual networks," *arXiv:2011.10675*, Nov. 2020.

[26] X. Zou, F. Xiao, Z. Yu, and Y. J. Lee, "Delving Deeper into Anti-aliasing in ConvNets," in *BMVC*, 2020.

[27] A. Chaman and I. Dokmanic, "Truly shift-invariant convolutional neural networks," in *CVPR*, 2021.

[28] I. Waldspurger, "Wavelet transform modulus : Phase retrieval and scattering," Doctoral Thesis, Ecole normale supérieure, Paris, 2015.

[29] N. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Applied and computational harmonic analysis*, vol. 10, no. 3, pp. 234–253, May 2001.

[30] I. Bayram and I. W. Selesnick, "On the dual-tree complex wavelet packet and M-band transforms," *IEEE Trans. Signal Processing*, vol. 56, no. 6, pp. 2298–2310, Jun. 2008.

[31] H. Leterme, K. Polisano, V. Perrier, and K. Alahari, "Modélisation parcimonieuse de CNNs avec des paquets d'ondelettes dual-tree," in *ORASIS*, 2021.

[32] N. Kingsbury and J. Magarey, "Wavelet transforms in image processing," in *Signal Analysis and Prediction*. Birkhäuser, 1998, pp. 27–46.

[33] E. Oyallon and S. Mallat, "Deep roto-translation scattering for object classification," in *CVPR*, 2015.

[34] A. Singh and N. Kingsbury, "Dual-Tree wavelet scattering network with parametric log transformation for object classification," in *ICASSP*, 2017.

[35] D. Zou and G. Lerman, "Graph convolutional neural networks via scattering," *Applied and Computational Harmonic Analysis*, vol. 49, no. 3, pp. 1046–1074, Nov. 2020.

[36] E. Oyallon, E. Belilovsky, and S. Zagoruyko, "Scaling the scattering transform: Deep hybrid networks," in *ICCV*, 2017.

[37] E. Oyallon, E. Belilovsky, S. Zagoruyko, and M. Valko, "Compressing the input for CNNs with the first-order scattering transform," in *ECCV*, 2018.

[38] J. Zarka, L. Thiry, T. Angles, and S. Mallat, "Deep network classification by scattering and homotopy dictionary learning," in *ICLR*, 2020.

[39] J. Zarka, F. Guth, and S. Mallat, "Separation and concentration in deep networks," in *ICLR*, 2021.

[40] F. Cotter and N. Kingsbury, "A learnable scatternet: Locally invariant convolutional layers," in *ICIP*, 2019.

[41] S. Gauthier, B. Thérien, L. Alsène-Racicot, M. Chaudhary, I. Rish, E. Belilovsky, M. Eickenberg, and G. Wolf, "Parametric scattering networks," in *CVPR*, 2022.

[42] S. Mallat, "Group invariant scattering," *Communications on Pure and Applied Mathematics*, vol. 65, no. 10, pp. 1331–1398, Jul. 2012.

[43] ——, "Understanding deep convolutional networks," *Philosophical Trans. of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, Apr. 2016.

[44] W. Czaja and W. Li, "Analysis of time-frequency scattering transforms," *Applied and Computational Harmonic Analysis*, vol. 47, no. 1, pp. 149–171, Jul. 2019.

[45] ——, "Rotationally invariant time–frequency scattering transforms," *J. Fourier Analysis and Applications*, vol. 26, no. 1, p. 4, Jan. 2020.

[46] A. Bietti and J. Mairal, "Group invariance, stability to deformations, and complexity of deep convolutional representations," *JMLR*, vol. 20, no. 1, pp. 876–924, 2019.

[47] J. Havlicek, J. Havlicek, and A. Bovik, "The analytic image," in *ICIP*, 1997.

[48] S. Mallat, *A Wavelet Tour of Signal Processing : The Sparse Way*. Academic Press, 2009.

[49] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *ICML*, 2015.

[50] N. Kingsbury, "Design of Q-shift complex wavelets for image processing using frequency domain energy minimization," in *ICIP*, 2003.

[51] G. P. Nason and B. W. Silverman, "The stationary wavelet transform and some statistical applications," in *Wavelets and Statistics*, ser. Lecture Notes in Statistics. Springer, 1995, pp. 281–299.

[52] Y. Meyer, "Principe d'incertitude, bases hilbertiennes et algèbres d'opérateurs," in *Séminaire Bourbaki*, vol. 662, 1985.

[53] I. W. Selesnick, R. Baraniuk, and N. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 123–151, Nov. 2005.

[54] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," *IJCV*, vol. 115, no. 3, pp. 211–252, Apr. 2015.

[55] K. I. Park and M. Park, *Fundamentals of Probability and Stochastic Processes with Applications to Communications*. Springer, 2018.

[56] P. R. Halmos, *Measure Theory*. Springer, 2013.

[57] A. Torralba and A. Oliva, "Statistics of natural image categories," *Network*, vol. 14, no. 3, pp. 391–412, Jan. 2003.

[58] M. Tygert, J. Bruna, S. Chintala, Y. LeCun, S. Piantino, and A. Szlam, "A mathematical motivation for complex-valued convolutional networks," *Neural Computation*, vol. 28, no. 5, pp. 815–825, May 2016.

[59] K. B. Athreya and S. N. Lahiri, *Measure Theory and Probability Theory*. Springer, 2006, vol. 19.