Chapter 1

Background

This thesis combines work in several fields. We provide a background for the most important and relevant fields in this chapter. We first introduce the basics of deep learning, before defining the properties of Wavelet Transforms, and finally we introduce the Scattering Transform, the original inspiration for this thesis.

1.1 The Fourier and Wavelet Transforms

Computer vision is an extremely difficult task. Pixel intensities in an image are typically not very informative in understanding what is in that image. Indeed, these values are sensitive to lighting conditions and camera configurations. It would be easy to take two photos of the same scene and get two vectors x_1 and x_2 that have a very large Euclidean distance, but to a human, would represent the same objects. What is most important in defining an image is difficult to define, however some things are notably more important than others. In particular, the location or phase of the waves that make up an image is much more important than the magnitude of these waves, something that is not necessarily true for audio processing. A simple experiment to demonstrate this is shown in Figure 1.1.

1.1.1 The Fourier Transform

For a signal $f(t) \in L_2(\mathbb{R})$ (square summable signals), the *Fourier transform* is defined as:

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt \tag{1.1.1}$$

This can be extended to two dimensions for signals $f(\mathbf{u}) \in L_2(\mathbb{R}^2)$:

$$F(\boldsymbol{\omega}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\mathbf{u}) e^{-j\boldsymbol{\omega}^t \mathbf{u}} d\mathbf{u} = \langle f(\mathbf{u}),\ e^{j\boldsymbol{\omega}^t \mathbf{u}} \rangle \tag{1.1.2}$$

Figure 1.1: **Importance of phase over magnitude for images.** The phase of the Fourier transform of the first image is combined with the magnitude of the Fourier transform of the second image and reconstructed. Note that the first image has entirely won out and nothing is left visible of the cameraman.

The Fourier transform is an invaluable signal expansion, as viewing a signal in the frequency space offers many insights, as well as affording many very useful properties (most notably the efficiency of convolution as a product of Fourier transforms). While it is a mainstay in signal processing, it can be a poor feature descriptor due to the infinite support of its basis functions - the complex sinusoids $e^{j\boldsymbol{\omega}^t u}$. If a single pixel changes in the input it can change all of the Fourier coefficients. As natural images are generally non-stationary, we need to be able to isolate frequency components in local regions of an image, and not have this property of global dependence.

1.1.2 The Continuous Wavelet Transform

The *Continuous Wavelet Transform* (CWT), like the Fourier Transform, can be used to decompose a signal into its frequency components. Unlike the Fourier transform, these frequency components can be localized in space. To achieve this, we need a bandpass filter, or *mother wavelet* ψ such that:

$$\int_{-\infty}^{\infty} \psi(\mathbf{u}) d\mathbf{u} = \Psi(0) = 0 \tag{1.1.3}$$

Any function that sufficient decay in frequency and satisfies (1.1.3) satisfies the *admissibility condition*.

As we are working in 2-D for image processing, consider dilations, shifts and rotations of this function by $a, \mathbf{b}, \theta$ where

$$\text{Translation: } T_{\mathbf{b}}x(\mathbf{u}) = x(\mathbf{u}-\mathbf{b}) \tag{1.1.4}$$

$$\text{Dilation: } D_a x(\mathbf{u}) = \frac{1}{a}x\left(\frac{\mathbf{u}}{a}\right), \quad a>0 \tag{1.1.5}$$

$$\text{Rotation: } R_\theta x(\mathbf{u}) = x(r_{-\theta}\mathbf{u}) \tag{1.1.6}$$

where r_θ is the 2-D rotation matrix. Now consider shifts, scales and rotations of our bandpass filter

$$\psi_{\mathbf{b},a,\theta}(\mathbf{u}) = \frac{1}{a}\psi\left(\frac{r_{-\theta}(\mathbf{u}-\mathbf{b})}{a}\right) \tag{1.1.7}$$

which are called the *daughter wavelets*. The 2D CWT of a signal $x(\mathbf{u})$ is defined as

$$CWT_x(\mathbf{b},a,\theta) = \int_{-\infty}^{\infty} \psi^*_{\mathbf{b},a,\theta}(\mathbf{u})x(\mathbf{u})d\mathbf{u} = \langle \psi_{\mathbf{b},a,\theta}(\mathbf{u}),\ x(\mathbf{u})\rangle \tag{1.1.8}$$

1.1.2.1 Properties

The CWT has some particularly nice properties. In particular, it has *covariance* under the three transformations (1.1.4)-(1.1.6):

$$T_{\mathbf{b}_0}x \to CWT_x(\mathbf{b}-\mathbf{b}_0,a,\theta) \tag{1.1.9}$$

$$D_{a_0}x \to CWT_x(\mathbf{b}/a_0, a/a_0, \theta) \tag{1.1.10}$$

$$R_{\theta_0}x \to CWT_x(r_{-\theta_0}\mathbf{b}, a, \theta+\theta_0) \tag{1.1.11}$$

Most importantly, the CWT is now localized in space, which distinguishes it from the Fourier transform. This means that changes in one part of the image will not affect the wavelet coefficients in another part of the image, so long as the distance between the two parts is much larger than the wavelength of the wavelets you are examining.

1.1.2.2 Inverse

The CWT can be inverted by using a *dual* function $\tilde{\psi}$. There are restrictions on what dual function we can use, namely the dual-wavelet pair must have an admissible constant C_ψ that satisfies the cross-admissibility constraint [1]. Without going into too much detail about these constraints, we know that we can recover x from CWT_x by:

$$x(\mathbf{u}) = \frac{1}{C_\psi}\int\int\int \frac{1}{a^3}CWT_x(\mathbf{b},a,\theta)\tilde{\psi}_{\mathbf{b},a,\theta}\ d\mathbf{b}dad\theta \tag{1.1.12}$$

1.1.2.3 Interpretation

As the CWT is a convolution with a zero mean function, the wavelet coefficients are only large in the regions of the parameter space $(\mathbf{b}, a, \theta)$ where $\psi_{\mathbf{b},a,\theta}$ 'match' the features of the signal. As the wavelet ψ is well localized, the energy of the coefficients CWT_x will be concentrated on the significant parts of the signal.

For an excellent description of the properties of the CWT in 1-D we recommend [2] and in 2-D we recommend [3].

1.1.3 Discretization and Frames

The CWT is highly redundant. We have taken a 2-D signal and expressed it in 4 dimensions (2 offset, 1 scale and 1 rotation). In reality, we would like to sample the space of the CWT. We would ideally like to fully retain all information in x (be able to reconstruct x from the samples) while sampling over $(\mathbf{b}, a, \theta)$ as little as possible to avoid redundancy.

A set of vectors $\Phi = \{\phi_i\}_{i \in I}$ in a hilbert space $\mathbb{H}$ is a *frame* if there exist two constants $0 < A \leq B < \infty$ such that for all $x \in \mathbb{H}$:

$$A||x||^2 \leq \sum_{i \in I} |\langle x, \phi_i \rangle|^2 \leq B||x||^2 \tag{1.1.13}$$

with A, B called the *frame bounds* [4]. The frame bounds relate to the issue of stable reconstruction. In particular, no vector x with $||x|| > 0$ should be mapped to 0, as this would violate the bound on A from below. This can be interpreted as ensuring our Φ cover the entire frequency space. The upper bound simply ensures that the transform coefficients are bounded.

Any finite set of vectors that spans the space is frame. An orthogonal basis is a commonly known frame where $A = B = 1$ and $|\phi_i| = 1$ (e.g. the Discrete Wavelet Transform or the Fourier Transform). Tight frames are frames where $A = B$ and Parseval tight frames have the special case $A = B = 1$. It is possible to have frames that have more vectors than dimensions, and this will be the case with many expansions we explore in this thesis.

If $A = B$ and $|\phi_i| = 1$, then A is the measure of the redundancy of the frame. Of course, for the orthogonal basis, $A = 1$ when $|\phi_i| = 1$ so there is no redundancy. For the 2-D DT$\mathbb{C}$WT which we will see shortly, the redundancy is 4.

1.1.3.1 Inversion

(1.1.13) specify the constraints that make a frame representation invertible. The tighter the frame bounds, the more easily it is to invert the signal. This gives us some guide to choosing the sampling grid for the CWT.

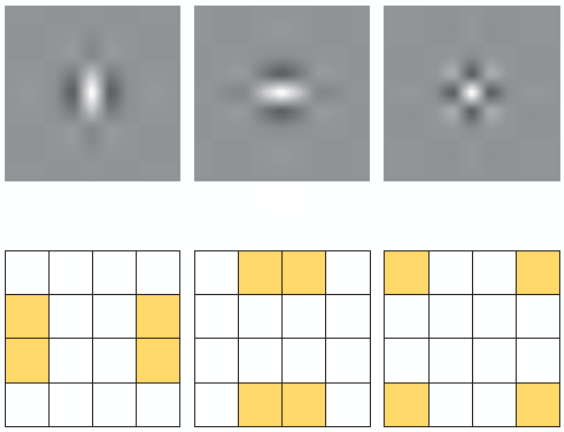

Figure 1.2: **Typical wavelets from the 2D separable DWT.** Top: Wavelet point spread functions for ψ^v (low-high), ψ^h (high-low), and ψ^d (high-high) wavelets. High-high wavelets are in a checkerboard pattern, with no favoured orientation. Bottom: Idealized support of the spectra of each of the wavelets. Image taken from [5].

One particular inverse operator is the *canonical dual frame*. If we define the frame operator $S = \Phi\Phi^*$ then the canonical dual of Φ is defined as $\tilde{\Phi} = \tilde{\phi}_{i\in I}$ where:

$$\tilde{\phi}_i = S^{-1}\phi_i \tag{1.1.14}$$

then[4]

$$x = \sum_{i\in I}\langle x, \phi_i\rangle\tilde{\phi}_i = \sum_{i\in I}\langle x, \tilde{\phi}_i\rangle\phi_i \tag{1.1.15}$$

If a frame is tight, then so is its dual.

1.1.4 Discrete Wavelet Transform

The 2-D DWT has one scaling fucntion and three wavelet functions, composed of the product of 1-D wavelets in the x and y directions:

$$\phi(\mathbf{u}) = \phi(u_1)\phi(u_2) \tag{1.1.16}$$
$$\psi^h(\mathbf{u}) = \phi(u_1)\psi(u_2) \tag{1.1.17}$$
$$\psi^v(\mathbf{u}) = \psi(u_1)\phi(u_2) \tag{1.1.18}$$
$$\psi^d(\mathbf{u}) = \psi(u_1)\psi(u_2) \tag{1.1.19}$$

with h, v, d indicating the sensitivity to horizontal, vertical and diagonal edges. The point spread functions for the wavelet functions are shown in Figure 1.2.

(1.1.7) gave the equation for the daughter wavelets in 2-D, in 1-D at scales $a = 2^j, j \geq 0$, this is simply:

$$\psi_{b,j}(t) = 2^{-j/2}\psi\left(\frac{t-b}{2^j}\right) \tag{1.1.20}$$

For the four equations above (1.1.16) – (1.1.19), define the daughter wavelets as:

$$\psi_{kl}^{\alpha,j}(\mathbf{u}) = \psi_{j,k}^{\alpha}(u_1)\psi_{j,l}^{\alpha}(u_2) \tag{1.1.21}$$

for $\alpha = h, v, d,\ k, l \in \mathbb{Z}$. We can then get an orthonormal basis with the set $\{\phi_{kl}^j, \psi_{kl}^{\alpha,j}\}$. The wavelet coefficients at chosen scale and location can then be found by taking the inner product of the signal x with the daughter wavelets.

1.1.4.1 Shortcomings

The Discrete Wavelet Transform (DWT) is an orthogonal basis. It is a natural first signal expansion to consider when frustrated with the limitations of the Fourier Transform. It is also a good example of the limitations of non-redundant transforms, as it suffers from several drawbacks:

- The DWT is sensitive to the zero crossings of its wavelets. We would like singularities in the input to yield large wavelet coefficients, but if they fall at a zero crossing of a wavelet, the output can be small. See Figure 1.3.
- They have poor directional selectivity. As the wavelets are purely real, they have passbands in all four quadrants of the frequency plane. While they can pick out edges aligned with the frequency axis, they do not have admissibility for other orientations. See Figure 1.2.
- They are not shift invariant. In particular, small shifts greatly perturb the wavelet coefficients. Figure 1.3 shows this for the centre-left and centre-right images.

The lack of shift invariance and the possibility of low outputs at singularities is a price to pay for the critically sampled property of the transform. This shortcoming can be overcome with the undecimated DWT [6], [7], but it comes with a heavy computational and memory cost.

1.1.5 Complex Wavelets

Fortunately, we can improve on the DWT with complex wavelets, as they can solve these new shortcomings while maintaining the desired localization properties.

The Fourier transform does not suffer from a lack of directional selectivity and shift variance, because its basis functions are based on the complex sinusoid:

$$e^{j\omega t} = \cos(\omega t) + j\sin(\omega t) \tag{1.1.22}$$

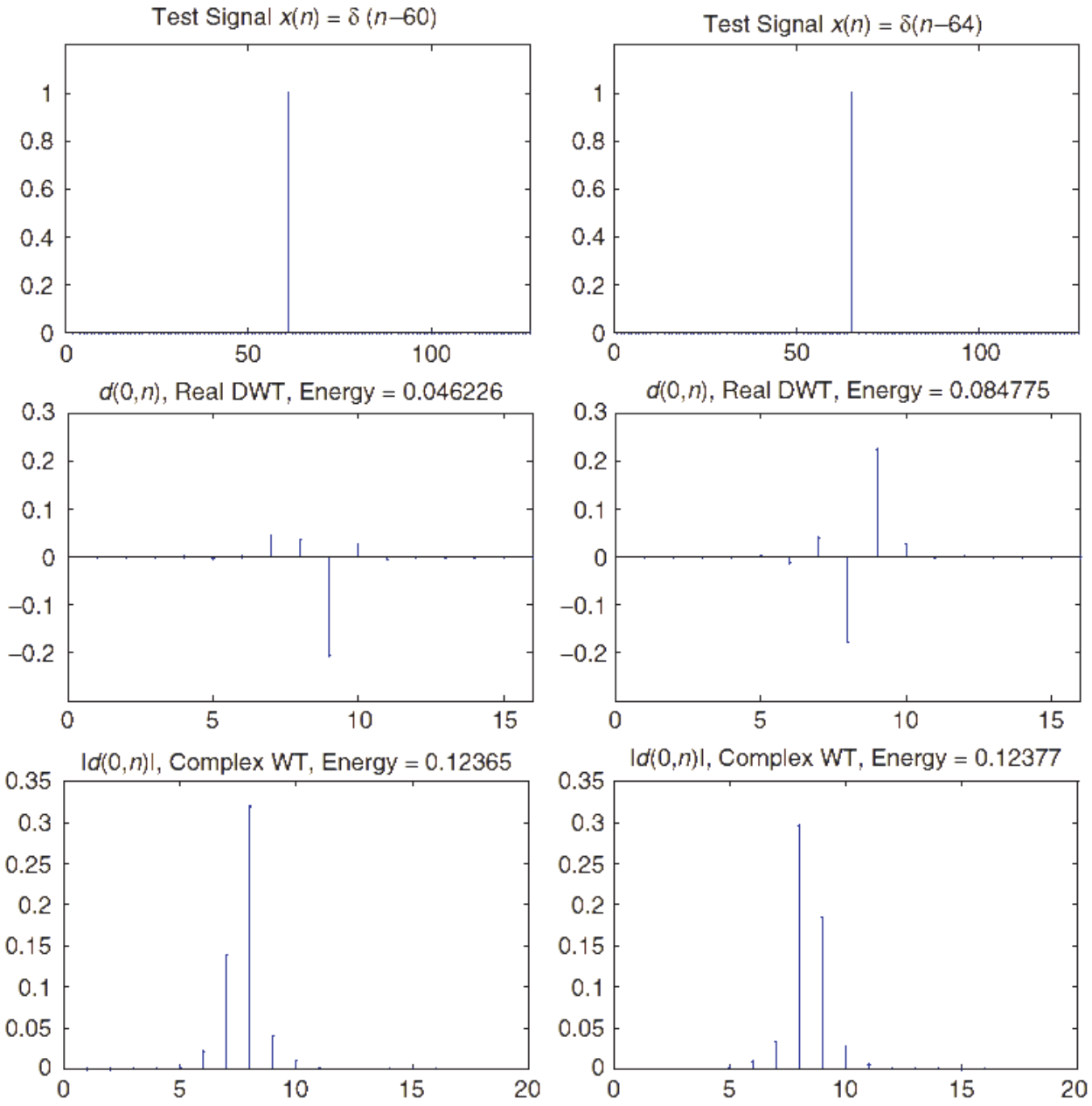

Figure 1.3: **Sensitivity of DWT coefficients to zero crossings and small shifts.** Two impulse signals $\delta(n-60)$ and $\delta(n-64)$ are shown (top), as well as the wavelet coefficients for scale $j=1$ for the DWT (middle) and for the DT$\mathbb{C}$WT (bottom). In the middle row, not only are the coefficients very different from a shifted input, but the energy has almost doubled. As the DWT is an orthonormal transform, this means that this extra energy has come from other scales. In comparison, the energy of the magnitude of the DT$\mathbb{C}$WT coefficients has remained far more constant, as has the shape of the envelope of the output. Image taken from [5].

whereas the DWT's basis functions are based on only the real sinusoid $\cos(\omega t)$.[1] As t moves along the real line, the phase of the Fourier coefficients change linearly, while their magnitude remains constant. In contrast, as t moves along the real line, the sign of the real coefficient flips between -1 and 1, and its magnitude is a rectified sinusoid.

These nice properties come from the fact that the cosine and sine functions of the Fourier transform form a Hilbert Pair and together constitute an analytic signal.

We can achieve these nice properties if the mother wavelet for our wavelet transform is analytic:

$$\psi_c(t) = \psi_r(t) + j\psi_i(t) \tag{1.1.23}$$

where $\psi_r(t)$ and $\psi_i(t)$ form a Hilbert Pair (i.e., they are 90° out of phase with each other).

There are a number of possible ways to do a wavelet transform with complex wavelets. We examine two in particular, a Fourier-based, sampled CWT using Morlet wavelets, and the Dual-Tree Complex Wavelet Transform (DT$\mathbb{C}$WT) developed by Kinsbury [5], [8]–[14].

The Morlet wavelet transform we look at as it is used by Mallat et. al. in their scattering transform [15]–[23], which has been a large inspiration for our work, and we will introduce it shortly. The DT$\mathbb{C}$WT we believe offers several advantages over the Morlet based implementation, and has been the basis for most of our work.

1.1.6 Sampled Morlet Wavelets

The wavelet transform used by Mallat et. al. in their scattering transform are an efficient implementation of the Gabor Transform. While the Gabor wavelets have the best theoretical trade-off between spatial and frequency localization, they have a non-zero mean. This violates (1.1.3) making them inadmissible as wavelets. Instead, the Morlet wavelet has the same shape, but with an extra degree of freedom chosen to set $\int \psi(\mathbf{u}) d\mathbf{u} = 0$. This wavelet has equation (in 2D):

$$\psi(\mathbf{u}) = \frac{1}{2\pi\sigma^2}(e^{i\mathbf{u}\xi} - \beta)e^{-\frac{|\mathbf{u}|^2}{2\sigma^2}} \tag{1.1.24}$$

where β is usually $<< 1$ and is this extra degree of freedom, σ is the size of the gaussian window, and ξ is the location of the peak frequency response — i.e., for an octave based transform, $\xi = 3\pi/4$.

Bruna and Mallat add an extra degree of freedom in their original design [16] by allowing for a non-circular Gaussian window over the complex sinusoid, which gives control over the angular resolution of the final wavelet. (1.1.24) now becomes:

$$\psi(\mathbf{u}) = \frac{\gamma}{2\pi\sigma^2}(e^{i\mathbf{u}\xi} - \beta)e^{-\mathbf{u}^\mathbf{t}\Sigma^{-1}\mathbf{u}} \tag{1.1.25}$$

[1] we have temporarily switched to 1D notation here as it is clearer and easier to use, but the results still hold for 2D

Figure 1.4: **Single Morlet filter with varying slants and window sizes.** Top left — 45° plane wave (real part only). Top right — plane wave with $\sigma = 3, \gamma = 1$. Bottom left — plane wave with $\sigma = 3, \gamma = 0.5$. Bottom right — plane wave with $\sigma = 2, \gamma = 0.5$.

Where

$$\Sigma^{-1} = \begin{bmatrix} \frac{1}{2\sigma^2} & 0 \\ 0 & \frac{\gamma^2}{2\sigma^2} \end{bmatrix}$$

The effects of modifying the eccentricity parameter γ and the window size σ are shown in Figure 1.4. A full family of Morlet wavelets at varying scales and orientations is shown in **??**.

1.1.6.1 Tightness and Invertibility

Let us write the wavelet transform of an input x with as

$$\mathcal{W}x = \{x * \phi_J, x * \psi_\lambda\}_\lambda \tag{1.1.26}$$

Assuming the transform is bounded, we can always scale it so that it satisfies Plancherel's equality

$$\|\mathcal{W}x\| = \|x\| \tag{1.1.27}$$

which is a nice property to have for invertibility, as well as for analysing how different signals get transformed (e.g. white noise versus standard images). Scaling the transform changes the upper bound B in (1.1.13) to 1 and makes the lower bound $A = 1 - \alpha$, where α is a measure of how non-tight a frame is.

Let us look at the tightness of a Morlet wavelet frame for a few manually selected parameters.

- For dilations, we choose $a = 2^{-j/Q}$ for $j \in \mathbb{Z}$ controlling the scale and Q the number of octaves per scale.

- For rotations, we subdivide the interval $[0,\pi)$ into K sections, and choose $\theta_k = \frac{k\pi}{K}$, $k = \{0,1,\dots K-1\}$.
- For the translations, we set the sample spacing $\Delta\mathbf{b} = 2^{-j/Q}$. **Need to check this**

Using the capital notation to denote the Fourier transform, define the function $A(\boldsymbol{\omega})$ to be the coverage each wavelet family has over the frequency plane:

$$A(\boldsymbol{\omega}) = |\Phi_J(\boldsymbol{\omega})|^2 + \sum_{\lambda} |\Psi_\lambda(\boldsymbol{\omega})|^2 \tag{1.1.28}$$

For a unit norm input $||x||^2 = 1$ and scaled wavelets, we can now change (1.1.13) to be:

$$1 - \alpha \leq A(\boldsymbol{\omega}) \leq 1 \tag{1.1.29}$$

If $A(\boldsymbol{\omega})$ is ever close to 0, then there is not a good coverage of the frequency plane at that location. If it ever exceeds 1, then there is overlap between bases. Both of these conditions make invertibility difficult[2]. Figure 1.5 show the frequency coverage of a few sample grids over the CWT parameters used by Mallat et. al.. Invertibility is possible, but not guaranteed for all configurations.

1.1.7 The DT$\mathbb{C}$WT

The DT$\mathbb{C}$WT was first proposed by Kingsbury in [9], [10] as a way to combat many of the shortcomings of the DWT, in particular, its poor directional selectivity, and its poor shift invariance. A thorough analsysis of the properties and benefits of the DT$\mathbb{C}$WT is done in [5], [11]. Building on these properties, it been used successfully for denoising and inverse problems [24]–[27], texture classification [28], [29], image registration [30], [31] and SIFT-style keypoint generation matching [32]–[36] amongst many other applications. Compared to Gabor (or Morlet) image analysis, the authors of [5] sum up the dangers as:

> A typical Gabor image analysis is either expensive to compute, is noninvertible, or both.

This nicely summarises the difference between this method and the Fourier based method outlined in subsection 1.1.6. The DT$\mathbb{C}$WT is a filter bank (FB) based wavelet transform. It is faster to implement than the Morlet analysis, as well as being more readily invertible.

1.1.7.1 Deisgn Criteria for the DT$\mathbb{C}$WT

It was stated in subsection 1.1.5 that if the mother wavelet is complex, with its real and imaginary parts forming a Hilbert pair, then the wavelet transform of a signal with the

[2] In practise, if $A(\boldsymbol{\omega})$ is only slightly greater 1 for only a few small areas of $\boldsymbol{\omega}$, approximate inversion can be achieved

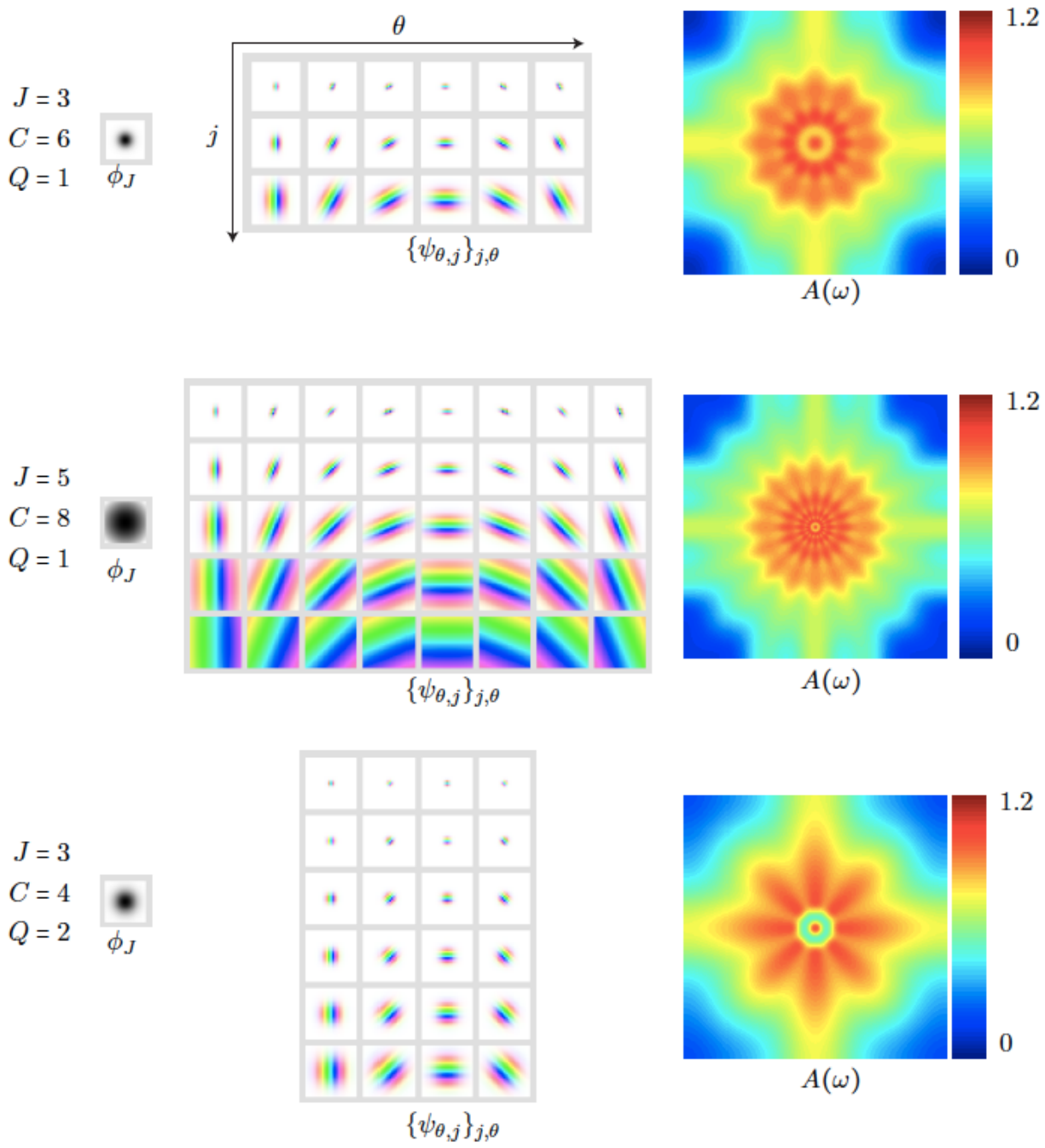

Figure 1.5: **Three Morlet Wavelet families and their tiling of the frequency plane.** For each set of parameters, the point spread functions of the wavelet bases are shown, next to their covering of the frequency plane $A(\boldsymbol{\omega})$. None of the configurations cover the corners of the frequency plane, but this is often mostly noise. Increasing J, K (Sifre uses C in these diagrams) or Q gives better frequency localization but at the cost of spatial localization and added complexity. Image taken from [22]. **TODO: update figure to show only the real wavelets in black and white and the correct variable names**.

daughter wavelets would give a representation that had nice shift properties[3], was insensitive to zero crossings of the wavelet, and had good directional selectivity.

As in subsection 1.1.5, we want to have a complex mother wavelet $\psi_c = \psi_r + j\psi_i$ and complex father wavelet $\phi_c = \phi_r + j\phi_i$, but now achieved with filter banks. The complex componenet allows for support of both the mother and father wavelet on only one half of the frequency plane.

The dual tree framework shown in Figure 1.6 can achieve this by making the real and imaginary components with their own DWT. In particular, if we define:

- h_0, h_1 the low and high-pass analysis filters for ϕ_r, ψ_r
- g_0, g_1 the low and high-pass analysis fitlers for ϕ_i, ψ_i
- $\tilde{h}_0, \tilde{h}_1$ the low and high pass synthesis filters for $\tilde{\phi}_r, \tilde{\psi}_r$.
- $\tilde{g}_0, \tilde{g}_1$ the low and high pass synthesis filters for $\tilde{\phi}_i, \tilde{\psi}_i$.

The dilation and wavelet equations for a 1D filter bank implementation are:

$$\phi_r(t) = \sqrt{2}\sum_n h_0(n)\phi_r(2t-n) \tag{1.1.30}$$

$$\psi_r(t) = \sqrt{2}\sum_n h_1(n)\phi_r(2t-n) \tag{1.1.31}$$

$$\phi_i(t) = \sqrt{2}\sum_n g_0(n)\phi_i(2t-n) \tag{1.1.32}$$

$$\psi_i(t) = \sqrt{2}\sum_n g_1(n)\phi_i(2t-n) \tag{1.1.33}$$

Designing a filter bank implementation that results in Hilbert symmetric wavelets does not appear to be an easy task. However, it was shown by Kingsbury in [11] (and later proved by Selesnick in [37]) that the necessary conditions are conceptually very simple. One low-pass filter must be a *half-sample shift* of the other. I.e.,

$$g_0(n) \approx h_0(n-0.5) \rightarrow \psi_g(t) \approx \mathcal{H}\{\psi_h(t)\} \tag{1.1.34}$$

As the DT$\mathbb{C}$WT is designed as an invertible filter bank implementation, this is only one of the constraints. Naturally, there are also perfect reconstruction, finite support, linear phase and vanishing moment constraints to consider in the filter bank design.

The derivation of the filters that meet these conditions is covered in detail in [14], [38], and in general in [5]. The result is the option of three families of filters: biorthogonal filters ($h_0[n] = h_0[N-1-n]$ and $g_0[n] = g_0[N-n]$), q-shift filters ($g_0[n] = h_0[N-1-n]$), and common-factor filters.

[3] in particular, that a shift in input gives the same shift in magnitude of the wavelet coefficients, and a linear phase shift

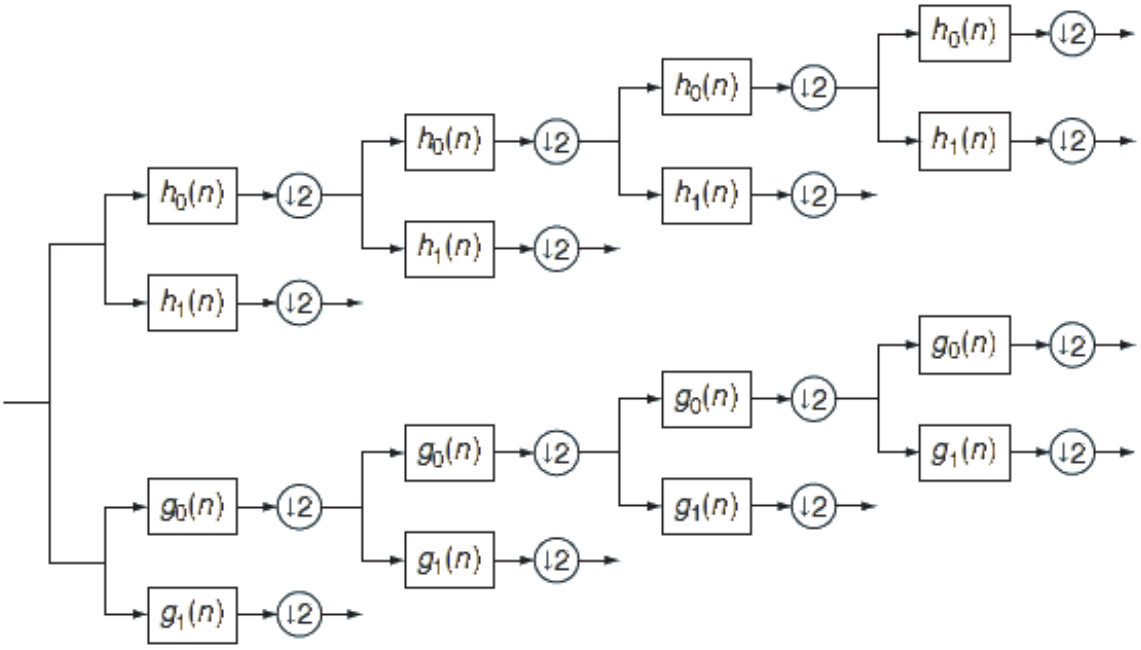

Figure 1.6: **Analysis FB for the** DT$\mathbb{C}$WT**.** Top 'tree' forms the real component of the complex wavelet ψ_r, and the bottom tree forms the imaginary (Hilbert pair) component ψ_i. Image taken from [5].

1.1.7.2 2-D DT$\mathbb{C}$WT and its Properties

While analytic wavelets in 1D are useful for their shift invariance, the real beauty of the DT$\mathbb{C}$WT is in its ability to make a separable 2D wavelet transform with oriented wavelets.

Figure 1.7a shows the spectrum of the wavelet when the separable product uses purely real wavelets, as is the case with the DWT. Figure 1.7b however, shows the separable product of two complex, analytic wavelets resulting in a localized and oriented 2D wavelet.

Note that in this thesis, we name the wavelets by the direction of the edge that they are most sensitive to.

For example, the 135° wavelet can be obtained by the separable product:

$$\psi(\mathbf{u}) = \psi_c(u_1)\psi_c^*(u_2) \tag{1.1.35}$$

$$= (\psi_r(u_1) + j\psi_i(u_1))\,(\psi_r(u_2) - j\psi_i(u_2)) \tag{1.1.36}$$

$$= (\psi_r(u_1)\psi_r(u_2) - \psi_i(u_1)\psi_i(u_2)) + j\,(\psi_r(u_1)\psi_i(u_2) + \psi_i(u_1)\psi_r(u_2)) \tag{1.1.37}$$

Similar equations can be obtained for the other five wavelets and the scaling function, by replacing ψ with ϕ for both directions, and not taking the complex conjugate in (1.1.35) to get the right hand side of the frequency plane. The 2-D DT$\mathbb{C}$WT requires four 2-D DWTs to calculate the four possible combinations of real and imaginary components. The high and

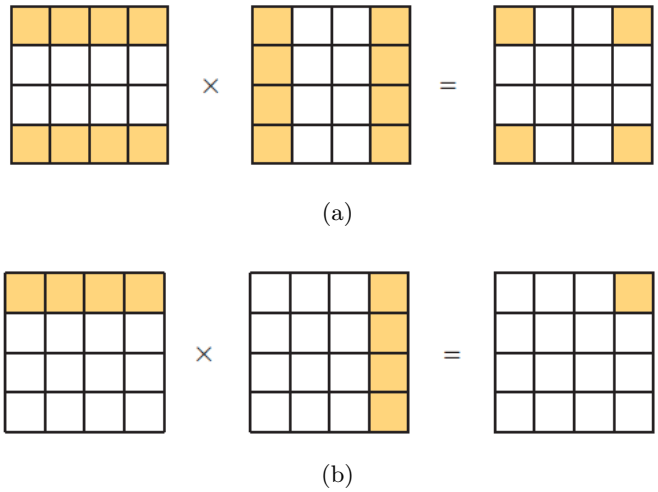

Figure 1.7: **The DWT high-high vs the** DT$\mathbb{C}$WT **high-high frequency support.** (a) The high-high DWT wavelet having a passband in all 4 corners of the frequency plane vs (b) the high-high DT$\mathbb{C}$WT wavelet frequency support only existing in one quadrant. Taken from [5]

lowpass outputs from these DWTs can then be summed in different ways as in (1.1.37) to get the complex bandpass wavelets. Figure 1.8 shows the resulting wavelets both in the spatial domain and their idealized support in the frequency domain.

1.1.7.3 Tightness and Invertibility

We analysed the coverage of the frequency plane for the Morlet wavelet family and saw what areas of the spectrum were better covered than others. How about for the DT$\mathbb{C}$WT?

It is important to note that in the case of the DT$\mathbb{C}$WT, the wavelet transform is also approximately unitary, i.e.,

$$\|x\|^2 \approx \|\mathcal{W}x\|^2 \tag{1.1.38}$$

and the implementation is perfectly invertible as $A(\boldsymbol{\omega})$ from (1.1.28) function is unity (or very near unity) $\forall \boldsymbol{\omega} \in [-\pi, \pi] \times [-\pi, pi]$. See Figure 1.9. This is not a surprise, as it is a design constraint in choosing the filters, but nonetheless is important to note.

1.1.8 Summary of Methods

One final comparison to make between the DT$\mathbb{C}$WT and the Morlet wavelets is their frequency coverage. The Morlet wavelets have flexibility at the cost of computational expense, and can be made to have tighter angular resolution than the DT$\mathbb{C}$WT. However it is not always better to keep using finer and finer resolutions, indeed the Fourier transform gives the ultimate in angular resolution, but as mentioned, this makes it less stable to shifts and deformations. We will explore this in more depth in Chapter 3.

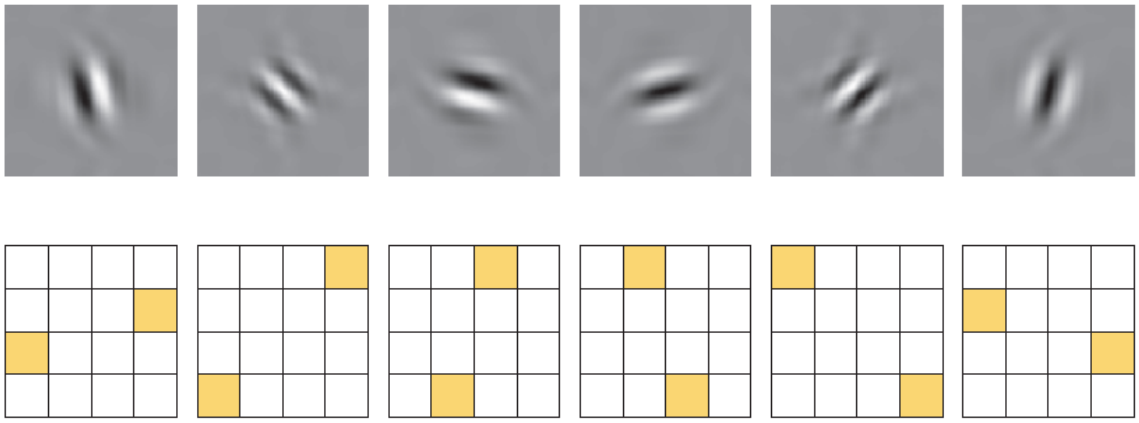

Figure 1.8: **Wavelets from the 2d** DT$\mathbb{C}$WT**. Top:** The six oriented filters in the space domain (only the real wavelets are shown). From left to right these are the $105°, 135°, 165°, 15°, 45°, 75°$ wavelets. **Bottom:** Idealized support of the Fourier spectrum of each wavelet in the 2D frequency plane. Spectra of the the real wavelets are shown — the spectra of the complex wavelets $(\psi_r + j\psi_i)$ only has support in the top half of the plane. Image taken from [5].

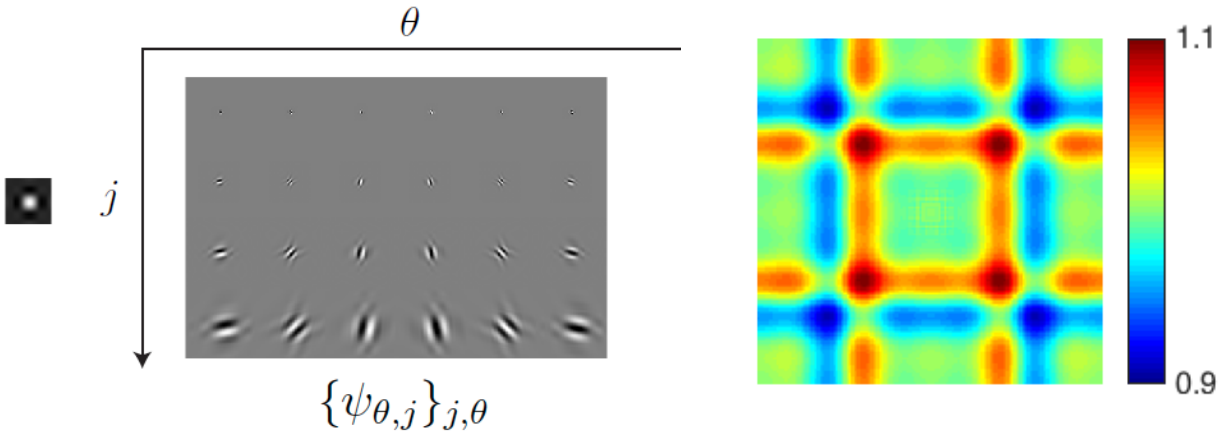

Figure 1.9: DT$\mathbb{C}$WT **family for** $J = 4$ **and their frequency coverage.** Note the reduced scale compared to Figure 1.5.

1.2 Scatternets

Scatternets have been a very large influence on our work, as well as being quite distinct from the previous discussions on learned methods. They were first introduced by Bruna and Mallat in their work [15], and then were rigorously defined by Mallat in [39].

While CNNs have the ability to learn invariances to nuiscance variabilities, the properties and optimal configurations are not well understood. It typically takes multiple trials by an expert to find the correct hyperparameters for these networks. A scattering transform instead builds well understood and well defined invariances.

We first review some of the desirable invariances before describing how a ScatterNet achieves them.

1.2.1 Properties

1.2.1.1 Translation Invariacne

Translation is often defined as being uninformative for classification — an object appearing in the centre of the image should be treated the same way as an the same object appearing in the corner of an image, i.e., Sx is invariant to global translations $x_c(\mathbf{u}) = x(\mathbf{u} - \mathbf{c})$ by $\mathbf{c} = (c_1, c_2) \in \mathbb{R}^2$ if

$$Sx_c = Sx \tag{1.2.1}$$

Note that we may instead want only local translation invariance and restrict the distance $|\mathbf{c}|$ for which (1.2.1) is true.

Note that CNNs are naturally covariant to translations in the pixel space, so $Sx_c = (Sx)_c$, $\mathbf{c} \in \mathbb{Z}^2$. Of course, natural objects exist in continuous space and are sampled, and any two images of the same scene taking with small camera disturbances are unlikely to be at integer pixel shifts of each other.

1.2.1.2 Stability to Noise

Stability to additive noise is another useful invariance to build, as it is a common feature in sampled signals. Stability is defined in terms of Lipschitz continuity, which is a strong form of uniform continuity for functions, which we briefly introduce here.

Formally, a Lipschitz continuous function is limited in how fast it can change; there exists an upper bound on the gradient the function can take, although it doesn't necessarily need to be differentiable everywhere. The modulus operator $|x|$ is a good example of a function that has a bounded derivative and so is Lipschitz continuous, but isn't differentiable everywhere. Alternatively, the modulus squared has derivative everywhere but is not Lipshitz continuous as its gradient grows with x.

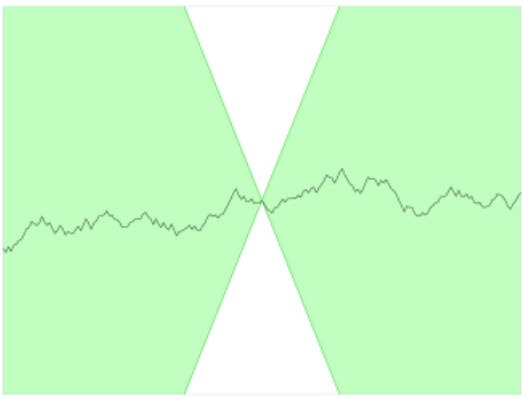

Figure 1.10: **A Lipschitz continuous function.** There is a cone for this function (shown in white) such that the graph always remains entirely outside the cone as it's shifted across. The minimum gradient needed for this to hold is called the 'best Lipschitz constant'.

To be stable to additive noise, we require that for a new signal $x'(\mathbf{u}) = x(\mathbf{u}) + \epsilon(\mathbf{u})$, there must exist a bounded $C > 0$ s.t.

$$\|Sx' - Sx\| \leq C\|x' - x\| \tag{1.2.2}$$

1.2.1.3 Stability to Deformations

Small deformations are important to be invariant to. However, this must be limited. It is important to ignore intra-class variations but not so invariant that an object can morph into another (in the case of MNIST for example, we do not want to be so stable to deformations that 7s can map to 1s).

Formally, for a new signal $x_\tau(\mathbf{u}) = x(\mathbf{u} - \tau(\mathbf{u}))$, where $\tau(\mathbf{u})$ is a non constant displacement field (i.e., not just a translation) that deforms the image, we require a $C_\tau > 0$ s.t.

$$\|Sx_\tau - Sx\| \leq C_\tau \|x\| \sup_{\mathbf{u}} |\nabla \tau(\mathbf{u})| \tag{1.2.3}$$

The term on the right $|\nabla\tau(\mathbf{u})|$ measures the deformation amplitude, so the supremum of it is a limit on the global defomation amplitude.

1.2.2 Finding the Right Operator

A Fourier modulus satisfies the first two of these requirements, in that it is both translation invariant and stable to additive noise, but it is unstable to deformations due to the infinite support of the sinusoid basis functions it uses. It also loses too much information — very different signals can all have the same Fourier modulus, e.g. a chirp, white noise and the Dirac delta function all have flat spectra.

A stable operator however is the averaging kernel, and this is the zeroth scattering coefficient:

$$S_0 x \triangleq x * \phi_J \left(2^J \mathbf{u}\right) \tag{1.2.4}$$

which is translation invariant to shifts less than 2^J. It unfortunately results in a loss of information due to the removal of high frequency content. This is easy to see as the wavelet operator:

$$Wx = \{x * phi_J, x * \psi_\lambda\}_\lambda \tag{1.2.5}$$

contains all the information of x, whereas the zeroth scattering coefficient is simply the lowpass portion of W. In (1.2.5) $\lambda = (j, \theta)$ indexes the J scales and K orientations of the chosen wavelet transform, whether that be DT$\mathbb{C}$WT or Morlet.

This high frequency content can be 'recovered' by keeping the wavelet coefficients. Fortunately, unlike the Fourier modulus, the complex wavelet transform is stable to deformations due to the grouping together frequencies into dyadic packets [39]. However the wavelet terms, like a convolutional layer in a CNN, is only covariant to shifts rather than invariant. The real and imaginary parts of a complex wavelet transform vary rapidly, while its modulus is much smoother and gives a good measure for the frequency-localized energy content at a given spatial location. Interestingly, the modulus operator can often still be inverted, and hence does not lose any information, due to the redundancies of the complex wavelet transform [40].

We define the modulus wavelet transform as:

$$\tilde{W}x = \{x * \phi_J, |x * \psi_\lambda|\}_\lambda \tag{1.2.6}$$

Now the modulus terms are invariant for shifts of up to 2^j. Mallat et. al. choose to define the first ordering scattering coefficients as:

$$S_1 x(\lambda_1, \mathbf{u}) \triangleq |x * \psi_{\lambda_1}| * \phi_J \tag{1.2.7}$$

so now both the S_0 and S_1 coefficients are invariant for shifts up to 2^J. Again this averaging comes at a cost of discarding high frequency information, this time about the wavelet sparsity signal $|x * \psi_\lambda|$ instead of the input signal x. We can recover this information by repeating the above process and defining the second order scattering coefficeints as:

$$S_2 x(\lambda_2, \lambda_1, \mathbf{u}) \triangleq ||x * \psi_{\lambda_1}| * \psi_{\lambda_2}| * \phi_J \tag{1.2.8}$$

Even still, this averaging means that a lot of information is lost from the first layer outputs ($\|x * \psi_\lambda\|$). Bruna and Mallat combat this by also convolving the output with wavelets that cover the rest of the frequency space, giving

$$U[p]x = U[\lambda_2]U[\lambda_1]x = \||x * \psi_{\lambda_1}| * \psi_{\lambda_2}\|$$

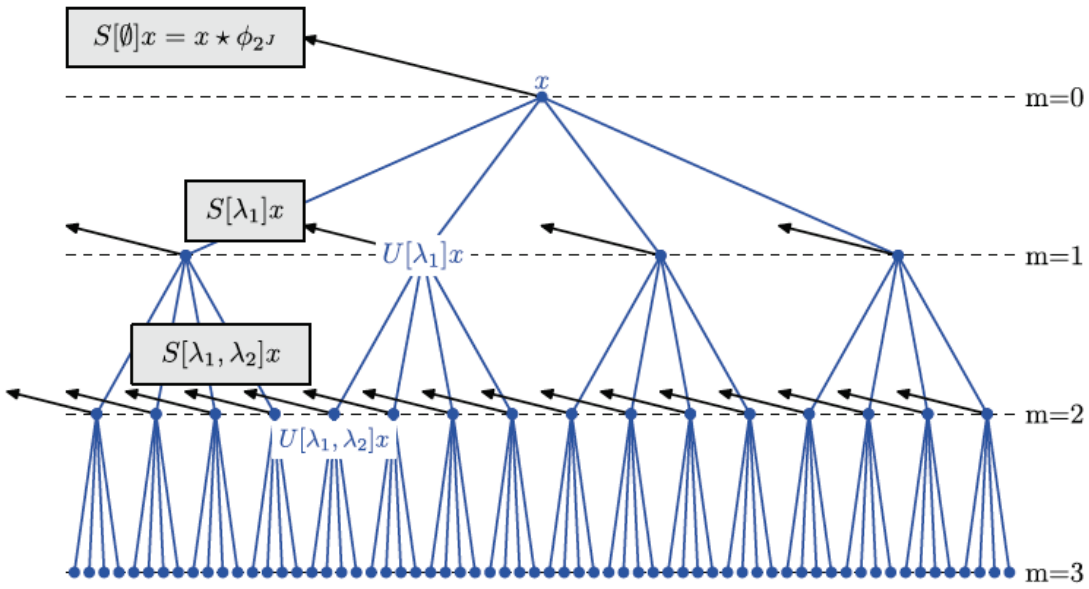

Figure 1.11: The translation invariant Scattering Transform. Scattering outputs are the leftward pointing arrows $S[p]x$, and the intermediate coefficients $U[p]x$ are the centre nodes of the tree. Taken from [16].

The choice of wavelet functions λ_1 and λ_2 is combined into a path variable, $p = (\lambda_1, \lambda_2, \ldots \lambda m)$.

Local invariants can be again computed by convolving this with another scaling function ϕ. The result is now a multiscale scattering transform, with coefficients:

$$S[p]x = U[p]x * \phi_{2^J}(\mathbf{u})$$

A graphical representation of this is shown in Figure 1.11.

References

[1] M. Holschneider and P. Tchamitchian, "Pointwise analysis of Riemann's "nondifferentiable" function", en, *Inventiones mathematicae*, vol. 105, no. 1, pp. 157–175, Dec. 1991.

[2] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*, 2nd ed., ser. Prentice Hall Signal Processing Series. Prentice Hall PTR, 2007.

[3] J. Antoine, R. Murenzi, P. Vandergheynst, and S. Ali, *Two-Dimensional Wavelets and Their Relatives*. 2004.

[4] J. Kovacevic and A. Chebira, *An Introduction to Frames*. Hanover, MA, USA: Now Publishers Inc., 2008.

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

[6] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, 1998.

[7] R. R. Coifman and D. L. Donoho, "Translation-Invariant De-Noising", en, in *Wavelets and Statistics*, ser. Lecture Notes in Statistics 103, A. Antoniadis and G. Oppenheim, Eds., Springer New York, 1995, pp. 125–150.

[8] N. Kingsbury and J. Magarey, "Wavelet transforms in image processing", A. Prochazka, J. Uhlir, and P. Sovka, Eds., 1997.

[9] N. Kingsbury, "The Dual-Tree Complex Wavelet Transform: A New Technique For Shift Invariance And Directional Filters", in *1998 8th International Conference on Digital Signal Processing (DSP)*, Utah, Aug. 1998, pp. 319–322.

[10] ——, "The dual-tree complex wavelet transform: A new efficient tool for image restoration and enhancement", in *Signal Processing Conference (EUSIPCO 1998), 9th European*, Sep. 1998, pp. 1–4.

[11] N. Kingsbury, "Image processing with complex wavelets", *Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences*, vol. 357, no. 1760, pp. 2543–2560, Sep. 1999.

[12] ——, "Shift invariant properties of the dual-tree complex wavelet transform", in *Icassp '99: 1999 Ieee International Conference on Acoustics, Speech, and Signal Processing, Proceedings Vols I-Vi*, 1999, pp. 1221–1224.

[13] ——, "A dual-tree complex wavelet transform with improved orthogonality and symmetry properties", 2000.

[14] ——, "Complex wavelets for shift invariant analysis and filtering of signals", *Applied and Computational Harmonic Analysis*, vol. 10, no. 3, pp. 234–253, May 2001.

[15] J. Bruna and S. Mallat, "Classification with scattering operators", in *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2011, pp. 1561–1566.

[16] ——, "Invariant Scattering Convolution Networks", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1872–1886, Aug. 2013.

[17] J. Bruna, "Scattering Representations for Recognition", Theses, Ecole Polytechnique X, Feb. 2013.

[18] E. Oyallon, S. Mallat, and L. Sifre, "Generic Deep Networks with Wavelet Scattering", *arXiv:1312.5940 [cs]*, Dec. 2013. arXiv: 1312.5940 [cs].

[19] E. Oyallon and S. Mallat, "Deep Roto-Translation Scattering for Object Classification", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2865–2873.

[20] L. Sifre and S. Mallat, "Rotation, Scaling and Deformation Invariant Scattering for Texture Discrimination", in *2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2013, pp. 1233–1240.

[21] L. Sifre and S. Mallat, "Rigid-Motion Scattering for Texture Classification", *arXiv:1403.1687 [cs]*, Mar. 2014. arXiv: 1403.1687 [cs].

[22] L. Sifre, "Rigid-Motion Scattering for Image Classification", PhD Thesis, Ecole Polytechnique, Oct. 2014.

[23] L. Sifre and J. Anden, *ScatNet*, École normale supérieure, Nov. 2013.

[24] P. de Rivaz and N. Kingsbury, "Bayesian image deconvolution and denoising using complex wavelets", in *2001 International Conference on Image Processing, 2001. Proceedings*, vol. 2, Oct. 2001, 273–276 vol.2.

[25] Y. Zhang and N. Kingsbury, "A Bayesian wavelet-based multidimensional deconvolution with sub-band emphasis", in *Engineering in Medicine and Biology Society*, 2008, pp. 3024–3027.

[26] G. Zhang and N. Kingsbury, "Variational Bayesian image restoration with group-sparse modeling of wavelet coefficients", *Digital Signal Processing*, Special Issue in Honour of William J. (Bill) Fitzgerald, vol. 47, pp. 157–168, Dec. 2015.

[27] M. Miller and N. Kingsbury, "Image denoising using derotated complex wavelet coefficients", eng, *IEEE transactions on image processing: a publication of the IEEE Signal Processing Society*, vol. 17, no. 9, pp. 1500–1511, Sep. 2008.

[28] S. Hatipoglu, S. K. Mitra, and N. Kingsbury, "Texture classification using dual-tree complex wavelet transform", in *Seventh International Conference on Image Processing and Its Applications*, 1999, pp. 344–347.

[29] P. de Rivaz and N. Kingsbury, "Complex wavelet features for fast texture image retrieval", in *1999 International Conference on Image Processing, 1999. ICIP 99. Proceedings*, vol. 1, 1999, 109–113 vol.1.

[30] P. Loo and N. G. Kingsbury, "Motion-estimation-based registration of geometrically distorted images for watermark recovery", P. W. Wong and E. J. Delp III, Eds., Aug. 2001, pp. 606–617.

[31] H. Chen and N. Kingsbury, "Efficient Registration of Nonrigid 3-D Bodies", *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 262–272, Jan. 2012.

[32] J. Fauqueur, N. Kingsbury, and R. Anderson, "Multiscale keypoint detection using the dual-tree complex wavelet transform", in *Image Processing, 2006 IEEE International Conference On*, IEEE, 2006, pp. 1625–1628.

[33] R. Anderson, N. Kingsbury, and J. Fauqueur, "Determining Multiscale Image Feature Angles from Complex Wavelet Phases", en, in *Image Analysis and Recognition*, ser. Lecture Notes in Computer Science 3656, M. Kamel and A. Campilho, Eds., Springer Berlin Heidelberg, Sep. 2005, pp. 490–498.

[34] ——, "Rotation-invariant object recognition using edge profile clusters", in *Signal Processing Conference, 2006 14th European*, IEEE, 2006, pp. 1–5.

[35] P. Bendale, W. Triggs, and N. Kingsbury, "Multiscale keypoint analysis based on complex wavelets", in *BMVC 2010-British Machine Vision Conference*, BMVA Press, 2010, pp. 49–1.

[36] E. S. Ng and N. G. Kingsbury, "Robust pairwise matching of interest points with complex wavelets", *Image Processing, IEEE Transactions on*, vol. 21, no. 8, pp. 3429–3442, 2012.

[37] I. Selesnick, "Hilbert transform pairs of wavelet bases", *IEEE Signal Processing Letters*, vol. 8, no. 6, pp. 170–173, Jun. 2001.

[38] N. Kingsbury, "Design of Q-shift complex wavelets for image processing using frequency domain energy minimization", in *2003 International Conference on Image Processing, 2003. ICIP 2003. Proceedings*, vol. 1, Sep. 2003, I-1013-16 vol.1.

[39] S. Mallat, "Group Invariant Scattering", en, *Communications on Pure and Applied Mathematics*, vol. 65, no. 10, pp. 1331–1398, Oct. 2012.

[40] I. Waldspurger, A. d'Aspremont, and S. Mallat, "Phase Recovery, MaxCut and Complex Semidefinite Programming", *arXiv:1206.0102 [math]*, Jun. 2012. arXiv: 1206.0102 [math].