UNIVERSIDAD
DE GRANADA

Facultad de CienciasLocalización de landmarks cefalométricos por medio de técnicas de few-shot learning

Doble Grado en Ingeniería Informática y Matemáticas

trabajo de fin de grado

Localización de landmarks cefalométricos por medio de técnicas de few-shot learning y análisis de redes convolucionales

Presentado por:
Alejandro Borrego Megías

Tutor:
Pablo Mesejo Santiago
DECSAI

Guillermo Gómez Trenado
DECSAI

Javier Merí de la Maza
Dpto Análisis Matemático

Curso académico 2021-2022

Localización de landmarks cefalométricos por medio de técnicas de few-shot learning y análisis de redes convolucionales

Alejandro Borrego Megías

Alejandro Borrego Megías *Localización de landmarks cefalométricos por medio de técnicas de few-shot learning y análisis de redes convolucionales*.
Trabajo de fin de Grado. Curso académico 2021-2022.

Responsable de tutorización

Pablo Mesejo Santiago
DECSAI

Guillermo Gómez Trenado
DECSAI

Javier Merí de la Maza
Dpto Análisis Matemático

Doble Grado en Ingeniería Informática y Matemáticas

Facultad de Ciencias

Universidad de Granada

Declaración de originalidad

D./Dña. Alejandro Borrego Megías

Declaro explícitamente que el trabajo presentado como Trabajo de Fin de Grado (TFG), correspondiente al curso académico 2021-2022, es original, entendida esta, en el sentido de que no ha utilizado para la elaboración del trabajo fuentes sin citarlas debidamente.

En Granada a 25 de junio de 2022

Fdo: Alejandro Borrego Megías

Dedicatoria (opcional)
Ver archivo `preliminares/dedicatoria.tex`

Índice general

Agradecimientos

Agradecimientos del libro (opcional, ver archivo `preliminares/agradecimiento.tex`).

Summary

An english summary of the project (around 800 and 1500 words are recommended).
File: `preliminares/summary.tex`

Introducción

De acuerdo con la comisión de grado, el TFG debe incluir una introducción en la que se describan claramente los objetivos previstos inicialmente en la propuesta de TFG, indicando si han sido o no alcanzados, los antecedentes importantes para el desarrollo, los resultados obtenidos, en su caso y las principales fuentes consultadas.

Ver archivo `preliminares/introduccion.tex`

Parte I.

Primera parte

Si el trabajo se divide en diferentes partes es posible incluir al inicio de cada una de ellas un breve resumen que indique el contenido de la misma. Esto es opcional.

1. Análisis de Redes Convolucionales

1.1. Introducción

En el contexto del procesamiento de imágenes, se define el término **invarianza** como la capacidad de reconocer un objeto en la imagen incluso si su apariencia ha variado en algún sentido (rotando, deformando ligeramente o trasladando el objeto por ejemplo). Esto es algo muy importante y positivo, pues esto indica que se preserva la identidad del objeto incluso a pesar de haberse sometido a ciertos cambios.

Las simetrías e invarianzas, aunque juegan un papel más importante en el campo de la Física, van abriéndose paso en el procesamiento de información de señales. La información contenida en las señales referente a imágenes o sonidos no suele verse afectada bajo la acción de grupos finitos como las traslaciones o las rotaciones, y es estable a la acción de pequeños difeomorfismos que deforman las señales. Esto motiva el estudio de las representaciones de traslaciones e invarianzas de las funciones de $L^2(\mathbb{R}^d)$, que son Lipschitz-continuas por la acción de difeomorfismos y que mantienen información de alta frecuencia para diferenciar entre distintos tipos de señales.

En primer lugar nos centraremos en la invarianza por traslaciones, entendida en el contexto de las imágenes como trasladar cada pixel de la imagen en una misma dirección la misma distancia. En este sentido:

Definición 1.1. $L_c f(x) = f(x - c)$ es la traslación de $f \in L^2(\mathbb{R}^d)$ por $c \in \mathbb{R}^d$.

Así, decimos que un operador Φ de $L^2(\mathbb{R}^d)$ en un espacio de Hilbert $\mathcal{H}$ es invariante por traslaciones si $\Phi(L_c f(x)) = \Phi(f)$ para todo $f \in L^2(\mathbb{R}^d)$ y para todo $c \in \mathbb{R}^d$. El módulo de la transformada de Fourier de f es un ejemplo de un operador invariante por traslaciones no canónico que estudiaremos en la sección siguiente.

Sin embargo, estos operadores invariantes a traslaciones no son Lipshcitz-continuos por la acción de difeomorfismos. Es conocido el hecho de que aparecen inestabilidades frente a deformaciones en las altas frecuencias, y el mayor reto es preservar la Lipschitz-continuidad en esta situación.

Para preservar la estabilidad en $L^2(\mathbb{R}^d)$ queremos que Φ sea no-expansiva.

Definición 1.2. Decimos que Φ es no-expansiva si:

$$\forall (f,h) \in L^2(\mathbb{R}^d)^2 \; ||\Phi(f) - \Phi(h)||_{\mathcal{H}} \leq ||f - h||$$

Es entonces suficiente verificar su Lipschitz-continuidad relativa a la acción de pequeños difeomorfismos cercanos a las traslaciones. Dichos difeomorfismos transforman $x \in \mathbb{R}^d$ en $x - \tau(x)$ dónde τ es el campo de desplazamiento.

Definición 1.3. Denotemos $L_\tau f(x) = f(x - \tau(x))$ como la acción del difeomorfismo $\mathbb{1} - \tau$ en f.

La condición de Lipschitz nos dice que $||\Phi(f) - \Phi(L_\tau f)||$ está acotada por el "tamaño"del difeomorfismo, y por tanto por la distancia entre $\mathbb{1} - \tau$ y $\mathbb{1}$, hasta $||f||$ por una constante multiplicativa.

Sea $|\tau(x)|$ la norma euclídea en $\mathbb{R}^d$, $\nabla|\tau(x)|$ la norma del supremo de la matriz $\nabla\tau(x)$, y $|H\tau(x)|$ la norma del supremo del tensor Hessiano.

La topología débil (recordemos que es la topología menos fina de un espacio normado que hace continuas todas las aplicaciones de su dual) en los difeomorfismos C^2 permite definir la siguiente aplicación:

Definición 1.4. Así, se define una distancia entre $\mathbb{1} - \tau$ y $\mathbb{1}$ en cualquier subconjunto compacto Ω de $\mathbb{R}^d$ como

$$(1.1)\quad d_\Omega(\mathbb{1}, \mathbb{1} - \tau) = \sup_{x\in\Omega} |\tau(x)| + \sup_{x\in\Omega} |\nabla\tau(x)| + \sup_{x\in\Omega} |H\tau(x)|$$

Definición 1.5. Un operador invariante por traslaciones Φ se dice *"Lipchitz-continuo"* por la acción de los difeomorfismos C^2 si para cualquier compacto $\Omega \subset \mathbb{R}^d$ existe una constante C tal que para todo $f \in L^2(\mathbb{R}^d)$ con Soporte en Ω y para todo $\tau \in C^2(\mathbb{R}^d)$ se cumple:

$$(1.2)\quad ||\Phi(f) - \Phi(L_\tau f)||_{\mathcal{H}} \leq C||f||(\sup_{x\in\mathbb{R}^d} |\nabla\tau(x)| + \sup_{x\in\mathbb{R}^d} |H\tau(x)|)$$

Debido a que Φ es invariante a traslaciones, la cota superior de Lipschitz no depende de la amplitud máxima de traslación $\sup_x |\tau(x)|$ de la métrica del difeomorfismo (1.1). Por otro lado la continuidad Lipschitz de (1.2) implica que Φ es invariante por traslaciones globales, pero es mucho más fuerte. Φ se ve poco afectada por los términos de primer y segundo grado de difeomorfismos que son traslaciones locales.

Las inestabilidades por altas frecuencias de las deformaciones se pueden evitar agrupando las frecuencias en paquetes diádicos en $\mathbb{R}^d$ con transformadas de ondeletas. Sin embargo, una transformada de ondeletas no es invariante por traslaciones. Es posible construir un operador invariante por traslaciones mediante un procedimiento de dispersión a lo largo de múltiples caminos, que preservan la condición de lipschitz de las ondeletas por la acción de difeomorfismos. Un propagador de dispersión se define en primer lugar como una composición organizada de convoluciones de operadores no lineales y no conmutativos, cada uno de los cuales calcula el módulo de la transformada de ondeletas. Esta cascada de convoluciones y módulos también pueden interpretarse como una red neuronal convolucional. Para ondeletas apropiadas, en la sección 2 veremos el principal teorema que demuestra que una ventana de dispersión mantiene la norma: $||\Phi(f)||_{\mathcal{H}} = ||f|| \; \forall f \in L^2(\mathbb{R}^d)$ y es Lipschitz-continua por difeomorfismos de clase C^2.

1.1.1. Notación

- $||\tau||_\infty := \sup_{x\in\mathbb{R}^d} |\tau(x)|$
- $||\nabla\tau||_\infty := \sup_{x\in\mathbb{R}^d} |\nabla\tau(x)|$
- $||\mathcal{H}\tau||_\infty := \sup_{x\in\mathbb{R}^d} |\mathcal{H}\tau(x)|$ dónde $|\mathcal{H}\tau(x)|$ es la norma del tensor Hessiano.
- El producto interno de $(x, y) \in \mathbb{R}^{2d}$ es $x \cdot y$.

- La norma de f en un espacio de Hilbert se denota por $||f||$.
- La norma de f en $L^2(\mathbb{R}^d)$ se denota por $||f||^2 = \int |f(x)|^2 dx$.
- La norma de f en $L^1(\mathbb{R}^d)$ es $||f||_1 = \int |f(x)| dx$.
- La transformada de Fourier se denota por $\widehat{f}(\omega) := \int f(x) e^{-ix\omega} d\omega$.
- Un operador $\mathcal{R}$ parametrizado por p es denotado por $\mathcal{R}[p]$ y $\mathcal{R}[\Omega] = \{\mathcal{R}[p]\}_{p\in\Omega}$.
- La norma del supremo de un operador lineal A en $L^2(\mathbb{R}^d)$ lo denotamos por $||A||$ y el conmutador entre dos operadores $[A, B] = AB - BA$.

1.2. Modelización Matemática de una Red Neuronal Convolucional

Nuestro primer objetivo será tratar de llegar a la modelización matemática de lo que es una **Red Neuronal Convolucional** (CNN [1]) que denominaremos **propagador de dispersión** (PD), para ello explicaremos la problemática de elegir una operador *lipschitz-continuo* bajo la acción de difeomorfismos e *invariante por traslaciones*, para evitar problemas como las inestabilidades en altas frecuencias que se producen en las señales bajo la acción de difeomorfismos como ocurre si usamos la transformada de Fourier.

Tras esto veremos posibles alternativas para evitar que se produzcan estas inestabilidades, mediante el uso de bases de la transformada de ondeletas de **Littlewood-Paley**. En concreto con esta segunda alternativa obtendremos un operador que es **Lipschitz-continuo** bajo la acción de difeomorfismos.

Después, nuestra tarea será conseguir calcular coeficientes que sean invariantes por traslaciones, y para ello necesitaremos utilizar un operador no lineal como es el módulo.

Una vez tengamos un operador con todas las propiedades anteriores presentaremos el **PD**, que será la aplicación en cadena de los operadores anteriores sobre un *"camino"* de frecuencias y rotaciones y que supondrá la modelización matemática de una Red Neuronal Convolucional.

1.2.1. De Fourier a las ondeletas de Littlewood-Paley

1.2.1.1. El módulo de la Transformada de Fourier

El análisis de Fourier tradicionalmente ha jugado un papel fundamental en el procesamiento de señales [Gon17], por lo que podría parecer un buen punto de partida para la construcción del **propagador de dispersión** emplear la **transformada de Fourier**, una de las herramientas matemáticas más potente en este campo. La intuición detrás de su fórmula es la de representar funciones no periódicas (pero que tienen área bajo la curva finita) como la integral de senos y cosenos multiplicados por una función que determina los pesos en cada instante. Formalmente tiene la siguiente expresión:

$$\widehat{f}(\omega) := \int f(x) e^{-ix\omega} dx = \int f(x) \left[\cos x\omega - i \sin x\omega\right] dx.$$

[1]Convolutional Neural Network

Entre las propiedades más destacables de la transformada encontramos el hecho de que una función se puede recuperar sin pérdida de información a partir de su transformada de Fourier, lo cual nos permite poder trabajar en el *"Dominio de Fourier"* (también llamado *"Dominio de Frecuencia"*) ya que al calcular la integral, la función resultante sólo depende de x (la frecuencia), y posteriormente pasar de nuevo al dominio original de la función, aplicando la inversa de la transformada sin pérdida de información.

Esto a priori es algo atractivo, pues nos permitiría trabajar en un dominio más sencillo y extraer conclusiones que podemos traducir al dominio original de la señal sin pérdida de información. Además, en el estudio de señales se suele emplear el módulo de la transformada de Fourirer para evitar fases complejas en el análisis, de esta forma el operador que vamos a probar en primer lugar es:

Definición 1.6. $\Phi(f) = |\widehat{f}|$ módulo de la transformada de Fourier.

Vamos a comprobar si se trata de un operador válido para nuestro propósito. Para ello necesitamos en primer lugar que sea un operador **Invariante por traslaciones**, algo esencial en el procesamiento de imágenes y concretamente en tareas de clasificación y detección, con ello buscamos que la información que determina a cada elemento de la imagen permanezca invariante frente a desplazamientos.

Figura 1.1.: Las tres estatuas deben identificarse como iguales, aunque se ecuentren desplazadas.

A continuación veremos que el operador que hemos elegido sí cumple esta propiedad.

Lema 1.1. *El operador $\Phi(f) = |\widehat{f}|$ es invariante por traslaciones.*

Demostración. Para ello tenemos que ver que si definimos para cada $c \in \mathbb{R}^d$, la traslación $L_c f(x) = f(x-c)$ se tiene que :

$$\widehat{L_c f}(w) = \int_{\mathbb{R}^d} L_c f(x) e^{-ixw} dx = \int_{\mathbb{R}^d} f(x-c) e^{-ixw} dx$$

Y realizando el cambio de variable $x - c = y$ se tendría que:

$$\begin{aligned}
\int_{\mathbb{R}^d} f(x-c) e^{-ixw} dx &= \int_{\mathbb{R}^d} f(y) e^{-i(y+c)w} dy = \\
&= \int_{\mathbb{R}^d} f(y) e^{-iyw} e^{-icw} dy = \\
&= \int_{\mathbb{R}^d} = e^{-icw} \int_{\mathbb{R}^d} f(y) e^{-iyw} dy = e^{-icw} \widehat{f}(w)
\end{aligned}$$

Por lo que se tiene que $|\widehat{L_c f}(w)| = |e^{-icw}||\widehat{f}(w)| = |\widehat{f}(w)|$ y entonces $\Phi(f) = |\widehat{f}|$ es invariante a traslaciones. $\square$

Sin embargo, la invaianza por traslaciones no es suficiente, necesitamos también que nuestro operador sea invariante frente a pequeñas deformaciones (difeomorfismos). Este hecho es de vital importancia en el contexto del procesamiento de imágenes y la visión por computador, pues nos permite caracterizar objetos presentes en imágen incluso si estos se ven afectados por pequeñas deformaciones.

Empezamos así definiendo qué es un difeomorfismo:

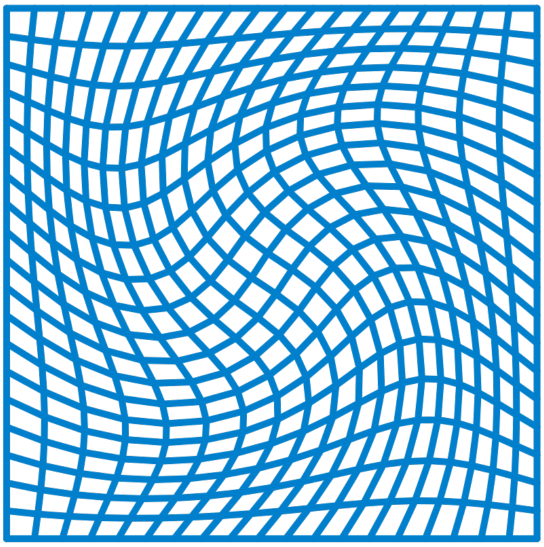

Figura 1.2.: Acción de un difeomorfismo en una rejilla.

Figura 1.3.: Todas las imágenes deberían clasificarse como 5, pese a las deformaciones.

Definición 1.7. Una función diferenciable $f : X \to \Omega$ dónde X y Ω son variedades, es un *"Difeomorfismo"* si f es una biyección y su inversa $f^{-1} : \Omega \to X$ es también diferenciable.

Por otro lado, vamos a necesitar definir una distancia entre $\mathbb{1}$ y $\mathbb{1} - \tau$ para la condición de Lipschitz y garantizar la estabilidad bajo la acción de difeomorfismos:

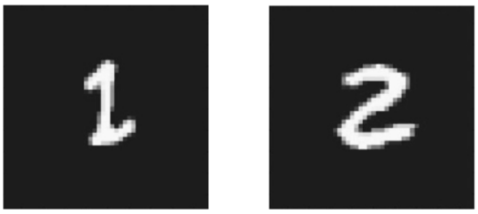

Figura 1.4.: Deformación excesiva que permite confundir el 1 con el 2 cuando se le aplica el difeomorfismo.

Definición 1.8. La topología débil de los difeomorfismos de C^2 define una distancia entre $\mathbb{1} - \tau$ y $\mathbb{1}$ en cualquier subconjunto compacto $\Omega \in \mathbb{R}^d$ mediante:

$$d_\Omega(\mathbb{1}, \mathbb{1} - \tau) = \sup_{x\in\Omega} |\tau(x)| + \sup_{x\in\Omega} |\nabla\tau(x)| + \sup_{x\in\Omega} |H\tau(x)|$$

Dónde $|\tau(x)|$ es la norma euclídea en $\mathbb{R}^d$, $|\nabla\tau(x)|$ el supremo de la matriz $\nabla\tau(x)$ y $|H\tau(x)|$ el supremo de la norma del tensor Hessiano. Además se define $||\tau||_\infty := \sup_{x\in\mathbb{R}^d} |\tau(x)|$.

De esta forma, un operador $\Phi(f)$ diremos que es estable frente a deformaciones si su norma euclídea $||\Phi(f) - \Phi(L_\tau f)||$ (con $L_\tau f(x) = f(x - \tau(x))$ definiendo la acción del difeomorfismo $\mathbb{1} - \tau$ en f) es *"pequeña"* cuando la deformación se mide por $d_\Omega(\mathbb{1}, \mathbb{1} - \tau)$. En otras palabras:

Definición 1.9. Un Operador Φ invariante por traslaciones se dice *"Lipschitz-continuo"* bajo la acción de difeomorfismos de C^2 si para cualquier compacto $\Omega \in \mathbb{R}^d$ existe una constante $c \in \mathbb{R}^d$ tal que para todo $f \in L^2(\mathbb{R}^d)$ con soporte en Ω y $\tau \in C^2(\mathbb{R}^d)$ se tiene

$$||\Phi(f) - \Phi(L_\tau f)|| \leq c||f||(||\nabla\tau||_\infty + ||H\tau||_\infty)$$

con $||\nabla\tau||_\infty + ||H\tau||_\infty < 1$ para asegurarnos de que la deformación sea invertible [TY05].

Como podemos comprobar, la cota superior no depende de $||\tau(x)||_\infty$ ya que hemos supuesto que Φ es invariante por traslaciones.
Sin embargo, esta propiedad no la verifica el módulo de la Transformada de Fourier, como podemos ver a continuación:

Lema 1.2. *El módulo de la Transformada de Fourier no es estable frente a pequeñas deformaciones y no es* "Lipschitz-continuo"*.*

Demostración. Vamos a considerar la función $\tau(x) := \epsilon x$ con $0 < \epsilon << 1$. De esta forma $||\nabla\tau(x)||_\infty = \epsilon$ y $||H\tau(x)||_\infty = 0$ con esto, la condición de Lipschitz debería ser

$$\left|\left| \, |\hat{f}| - |\widehat{L_\tau f}| \, \right|\right| \leq c||f||(||\nabla\tau||_\infty + ||H\tau||_\infty) \leq c||f||\epsilon$$

Vamos a ver un contraejemplo con una función de $\mathbb{R} \to \mathbb{R}$.

Supongamos que tenemos $f(x) = e^{i\xi x}\Theta(x) = e^{i\xi x}e^{-|x|}$. Calculamos ahora $|\hat{f}|$ y $|\widehat{L_\tau f}|$ teniendo en cuenta que :

$$
\begin{aligned}
|\widehat{f}(\omega)| &= \left| \int f(x) e^{-ix\omega} dx \right| \\
&= \left| \int f(x) e^{-ix\omega} dx \right| \\
&= \left| \int e^{i\xi x} e^{-|x|} e^{-ix\omega} dx \right| \\
&= \left| \int e^{-ix(\xi-\omega)} dx \right| \\
&= \left| \int e^{-|x|} \left[\cos x(\xi-\omega) - i \sin x(\xi-\omega) \right] dx \right|
\end{aligned}
$$

En el último paso podemos descomponer la integral en suma de dos, y para simplificar las operaciones llamamos $\beta = (\xi - \omega)$. Así, aplicando las siguientes fórmulas conocidas para el cálculo de integrales,

$$\int_{\mathbb{R}} \cos(\beta x) e^{-|x|} dx = \frac{1}{1+\beta^2} \tag{1.1}$$

y

$$\int_{\mathbb{R}} \sin(\beta x) e^{-|x|} dx = 0 \tag{1.2}$$

a nuestro caso concreto, obtenemos que:

$$
\begin{aligned}
|\widehat{f}(\omega)| &= \left| \int \cos x\beta e^{-|x|} dx - i \int \sin x\beta e^{-|x|} dx \right| \\
&= \frac{1}{1+\beta^2} \\
&= \frac{1}{1+(\xi-\omega)^2}.
\end{aligned}
$$

Ahora pasamos a calcular $|\widehat{L_\tau f}|$:

$$
\begin{aligned}
|\widehat{L_\tau f}(\omega)| &= |\widehat{f}((1-\epsilon)\omega)| \\
&= \left| \int f((1-\epsilon)x) e^{-ix\omega} dx \right| \\
&= \left| \int e^{i\xi(1-\epsilon)x} e^{-|(1-\epsilon)x|} e^{-ix\omega} dx \right|.
\end{aligned}
$$

Ahora realizamos el siguiente cambio de variable

$$\tilde{x} = (1-\epsilon)x \implies x = \frac{\tilde{x}}{1-\epsilon}$$

$$d\tilde{x} = (1-\epsilon)dx \implies dx = \frac{1}{(1-\epsilon)}d\tilde{x}$$

y aplicando los cambios a lo que teníamos nos queda

$$\begin{aligned}
|\widehat{L_\tau f}(\omega)| &= \frac{1}{(1-\epsilon)}\left|\int f((1-\epsilon)x)e^{-ix\omega}dx\right| \\
&= \frac{1}{(1-\epsilon)}\left|\int e^{i\xi\tilde{x}}e^{-|\tilde{x}|}e^{-i\frac{\tilde{x}}{(1-\epsilon)}\omega}d\tilde{x}\right| \\
&= \frac{1}{(1-\epsilon)}\left|\int e^{i\left[\frac{(1-\epsilon)\xi-\omega}{(1-\epsilon)}\right]\tilde{x}}e^{-|\tilde{x}|}d\tilde{x}\right| \\
&= \frac{1}{(1-\epsilon)}\left|\int e^{i\tilde{\beta}\tilde{x}}e^{-|\tilde{x}|}d\tilde{x}\right|,
\end{aligned}$$

como podemos ver, llegamos a una integral que se resuelve de la misma manera que en el caso anterior haciendo uso de (1.1) y (1.2):

$$\begin{aligned}
|\widehat{L_\tau f}(\omega)| &= \frac{1}{(1-\epsilon)}\frac{1}{1+\tilde{\beta}^2} \\
&= \frac{1}{(1-\epsilon)}\frac{1}{1+\left[\frac{(1-\epsilon)\xi-\omega}{(1-\epsilon)}\right]^2}.
\end{aligned}$$

De esta forma hemos obtenido que para nuestro caso concreto de $f(x) = e^{i\xi x}e^{-|x|}$,

$$\begin{aligned}
\left\| |\widehat{L_\tau f}| - |\widehat{f}| \right\| &= \left\| \frac{1}{(1-\epsilon)}\frac{1}{1+\left[\frac{(1-\epsilon)\xi-\omega}{(1-\epsilon)}\right]^2} - \frac{1}{1+(\xi-\omega)^2} \right\| \\
&\geq \left\| \frac{1}{1+\left[\frac{(1-\epsilon)\xi-\omega}{(1-\epsilon)}\right]^2} - \frac{1}{1+(\xi-\omega)^2} \right\| \\
&= \left(\int_{\mathbb{R}} \left| \frac{1}{1+\left[\frac{(1-\epsilon)\xi-\omega}{(1-\epsilon)}\right]^2} - \frac{1}{1+(\xi-\omega)^2} \right|^2 d\omega \right)^{1/2}.
\end{aligned}$$

A continuación vamos a intentar aproximar el valor del módulo de la integral, para ello en primer lugar vamos a realizar el siguiente cambio de variable

$$
\begin{aligned}
&t = \omega - \xi \implies \omega - (1-\epsilon)\xi = \omega - \xi + \epsilon\xi = t + \epsilon\xi \\
&dt = d\omega.
\end{aligned}
$$

Así, obtenemos que:

$$
\begin{aligned}
\int_{\mathbb{R}} \left| \frac{1}{1+(\xi-\omega)^2} - \frac{1}{1+\left[\frac{(1-\epsilon)\xi-\omega}{(1-\epsilon)}\right]^2} \right|^2 d\omega &\geq \left| \int_{\mathbb{R}} \left(\frac{1}{1+(\xi-\omega)^2} - \frac{1}{1+\left[\frac{(1-\epsilon)\xi-\omega}{(1-\epsilon)}\right]^2} \right)^2 d\omega \right| \\
&= \left| \int_{\mathbb{R}} \left(\frac{1}{1+t^2} - \frac{1}{1+\left[\frac{t+\epsilon\xi}{(1-\epsilon)}\right]^2} \right)^2 d\omega \right| \\
&\geq \int_{-M}^{M} (g_1(t) - g_2(t))^2 dt \\
&\approx \int_{-M}^{M} g_1(t)^2 dt
\end{aligned}
$$

Representando la gráfica $g_1(t) = \frac{1}{1+t^2}$, podemos ver cómo el valor de su integral se acumula en torno al origen de coordenadas, y en cambio $g_2(t) = \frac{1}{1+\left[\frac{t+\epsilon\xi}{(1-\epsilon)}\right]^2}$ e una traslación y escalado de la función anterior, por lo que podemos tomar una constante $M > 0$ y ξ de manera que
el escalado por medio de τ produce que la frecuencia ξ se traslade de ξ a $(1-\epsilon)\xi$, si suponemos que Θ además es regular con decrecimiento rápido, se tiene que:

$$
\left| \left| |\widehat{L_\tau f}| - |\widehat{f}| \right| \right| \sim |s||\xi|||\Theta|| = ||\nabla\tau||_\infty |\xi|||f||. \tag{1.3}
$$

Y como $|\xi|$ puede ser arbitrariamente grande, $\Phi(f) = |\widehat{f}|$ no satisface la continuidad de Lipschitz cuando se alcanzan altas frecuencias. □

Por esto, lo que haremos será reemplazar las ondas sinusoidales de la transformada de Fourier por funciones localizadas con un soporte mayor en altas frecuencias que nos permitan evitar estas complicaciones, que tendrán un mejor rendimiento en nuestro propósito. Estas funciones se denominan **ondeletas**.

1.2.1.2. Alternativa: Las ondeletas

Las ondeletas [Maloo] son pequeñas ondas estables bajo la acción de deformaciones, al contrario que las ondas sinusoidales de Fourier. Definiremos la transformada de ondeletas y veremos que calcula, mediante convoluciones con bases de ondeletas, coeficientes que cumplirán con los requisitos que pedíamos para construir el operador del propagador de dispersión.

Al contrario que las bases de Fourier, las bases de ondeletas definen representaciones dispersas de señales regulares a trozos, que podrían incluir transiciones y singularidades. En las

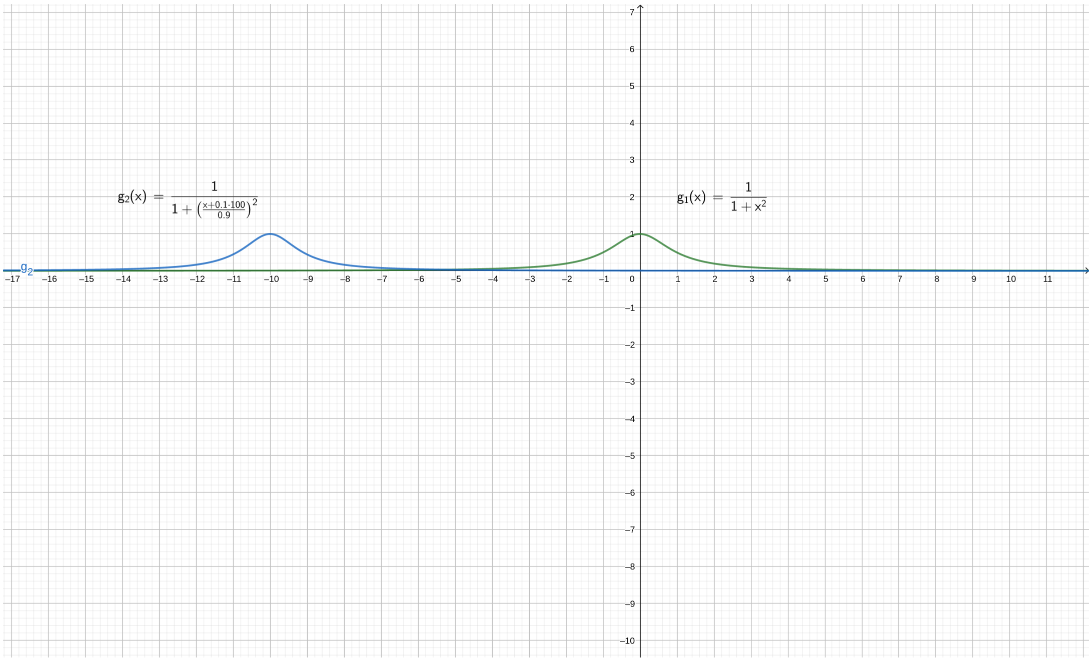

Figura 1.5.:

imágenes, los mayores coeficientes de las ondeletas se localizan en el entorno de las esquinas y en las texturas irregulares.

A modo de ejemplo, vamos a ver la base de Haar, un ejemplo que aunque no sea el que utilicemos para construir nuestro propagador de dispersión, puede ayudar a entender mejor la filosofía de las ondeletas. Se construye a partir de la siguiente función:

$$\psi(t) = \begin{cases} 1 & 0 \leq t < 1/2 \\ -1 & 1/2 \leq t < 1 \\ 0 & \textit{en otro caso} \end{cases}$$

A esta ondeleta la denominamos **ondeleta Madre**, pues a partir de ella, podemos generar la siguiente base ortonormal:

$$\left\{ \psi_{j,n}(t) = \frac{1}{\sqrt{2^j}} \psi \left(\frac{t - 2^j n}{2^j} \right) \right\}_{(j,n) \in \mathbb{Z}^2}$$

del espacio $L^2(\mathbb{R})$ de señales con energía finita, si recordamos, en este espacio la norma se define como:

Definición 1.10.

$$||f||^2 = \int_{-\infty}^{+\infty} |f(t)|^2 dt < +\infty$$

Así, cualquier señal f de energía finita puede ser representada por los coeficientes que se obtienen mediante el producto interno en $L^2(\mathbb{R})$ con la base anterior:

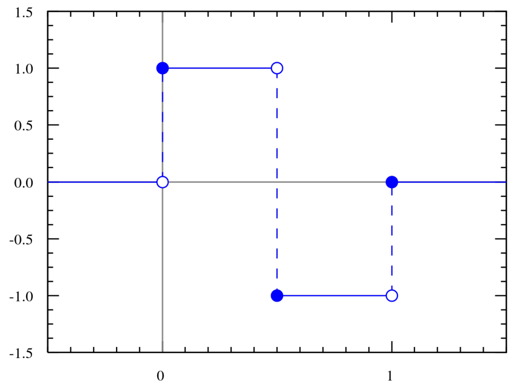

Figura 1.6.: Representación gráfica de la ondeleta de Haar.

$$\langle f, \psi_{j,n} \rangle = \int_{-\infty}^{+\infty} f(t)\psi_{j,n}(t)dt$$

y puede recuperarse sumando en su base ortonormal:

$$f = \sum_{j=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \langle f, \psi_{j,n} \rangle \psi_{j,n}$$

Esto nos permite (igual que pasaba con el módulo de la Transformada de Fourier) trabajar en un dominio más sencillo que nos permite procesar la información con mayor rapidez y posteriormente reconstruir la señal a partir de los coeficientes sin perder información. Algunas propiedades serían:

- Cada ondeleta $\psi_{j,n}$ tiene media 0 en su soporte $[2^j n, 2^j(n+1)]$.
- Si f es localmente regular y 2^j es pequeño, entonces es casi constante en su intervalo y su coeficiente de ondeleta $\langle f, \psi_{j,n} \rangle$ es prácticamente cero.
- los mayores coeficientes se localizan en los cambios bruscos de intensidad de señal, como pueden ser los bordes, las esquinas o las texturas en las imágenes.

Para el caso concreto de imágenes (ver por ejemplo sección 1.1 de [Maloo]), las bases de ondeletas ortonormales pueden construirse a partir de bases ortonormales en señales de una dimensión. Así, a partir de tres ondeletas $\psi^1(x)$, $\psi^2(x)$ y$\psi^3(x)$ con $x = (x_1, x_2) \in \mathbb{R}^2$, dilatadas por el factor 2^j y trasladadas por $2^j n$ con $n = (n1, n2) \in \mathbb{Z}^2$, se construye una base ortonormal para el espacio $L^2(\mathbb{R}^2)$:

$$\left\{ \psi^k_{j,n}(x) = \frac{1}{\sqrt{2^j}} \psi^k \left(\frac{x - 2^j n}{2^j} \right) \right\}_{(j,n) \in \mathbb{Z}^2}$$

El soporte de la ondeleta $\psi^k_{j,n}(x)$ es un cuadrado proporcional a la escala 2^j. Las bases de ondeletas en dos dimensiones se discretizan para definir bases ortonormales de imágenes de N píxeles.

Del mismo modo que en una dimensión, los coeficientes de ondeletas $\langle f, \psi_{j,n}^{k} \rangle$ serán pequeños si $f(x)$ es regular, y serán grandes cerca de los cambios bruscos de frecuencias como en los bordes o esquinas de las imágenes, como podemos ver en Figura 1.7. Los filtros resaltan los bordes en tres direcciones, horizontal (derecha) vertical (abajo) y en diagonal (abajo derecha).

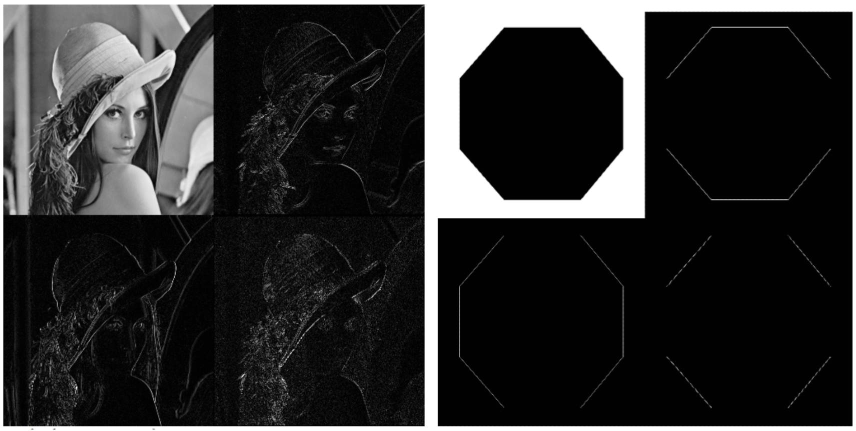

Figura 1.7.: Ejemplos de aplicar la base de Haar a dos imágenes [PJDoM06].

Volviendo al propósito de definir el propagador de dispersión, la ondeleta madre que elijamos y la base ortogonal que forme se verán afectadas normalmente por escalados y rotaciones, por lo tanto definimos:

Definición 1.11. Una ondeleta madre escalada por un factor $2^{-}j$ con $j \in \mathbb{Z}$ y rotada por $r \in G$ siendo G el grupo finito de rotaciones, se escribe:

$$\psi_{2^j r}(x) = 2^{dj}\psi(2^j r^{-1}x).$$

Su transformada de Fourier es $\widehat{\psi_{2^j r}}(\omega) = \widehat{\psi}(2^j r^{-1}\omega)$.

La transformada de dispersión que usaremos tendra una base de ondeletas generada por una ondeleta madre del tipo:

$$\psi(x) = e^{i\cdot\eta\cdot x}\Theta(x)$$

Donde $\widehat{\Theta}(x)$ es una función real centrada en una bola de baja frecuencia en $x = 0$, cuyo radio es del orden de π.

Y como podemos ver:

$$\widehat{\psi}(\omega) = \int_{\mathbb{R}^d} e^{i\cdot\eta\cdot x}\Theta(x)e^{-i\omega x}dx = \int_{\mathbb{R}^d} \Theta(x)e^{ix(\omega-\eta)}dx = \widehat{\Theta}(\omega - \eta).$$

Por lo tanto, $\widehat{\psi}(\omega)$ es real y centrada en una bola de mismo radio pero centrada en $\omega = \eta$ que tras el escalado y rotación:

$$\widehat{\psi}_{\lambda}(\omega) = \widehat{\Theta}(\lambda^{-1}\omega - \eta),$$

donde $\lambda = 2^j r \in 2^{\mathbb{Z}} \times G$.
Por lo tanto $\widehat{\psi}_\lambda(\omega)$ recubre una bola centrada en $\lambda^{-1}\eta$ con radio proporcional a $|\lambda| = 2^j$.

1.2.1.3. La Transformada de Littlewood-Paley

Una vez conocemos un poco más en profundidad las ondeletas y su foncionamiento, pasamos a presentar la **Transformada de ondeleta de Littlewood-Paley**, que es la que emplearemos para construir el propagador de dispersión.
Se trata de una representacion redundante que calcula convoluciones para todo $x \in \mathbb{R}^d$ sin realizar sub-muestreo:

$$\forall x \in \mathbb{R}^d \quad W[\lambda]f(x) = f * \psi_\lambda(x) = \int f(u)\psi_\lambda(x-u)du.$$

Dónde $*$ denota la operación de convolución.

Calculamos su transformada de Furier, para ello tendremos en cuenta el teorema de Convolución de la Transformada de Fourier, el cual dice:

Teorema 1.1. *Sean f y g dos funciones integrables.*
Si

$$h(x) = (f * g)(x) = \int f(y)g(x-y)dy$$

Entonces:

$$\widehat{h}(\omega) = \widehat{f}(\omega)\widehat{g}(\omega)$$

y

$$h(x) = (f * g)(x) = \int \widehat{f}(\omega)\widehat{g}(\omega)e^{-i\Omega x}d\omega$$

De esta manera, se tiene que:

$$\widehat{W[\lambda]f(\omega)} = \widehat{f}(\omega)\widehat{\psi_\lambda}(\omega) = \widehat{f}(\omega)\widehat{\psi}(\lambda^{-1}\omega).$$

Además, teniendo en cuenta la propiedad que nos dice que si la función f es real, entonces su transformada coincide con el conjugado complejo $\widehat{f}(-\omega) = \overline{\widehat{f}(\omega)}$ podemos ver que:

- si $\widehat{\psi}(\omega)$ y f son reales entonces $W[-\lambda]f = \overline{W[\lambda]f}$, utilizando la misma propiedad de antes.Además, si denotamos por G^+ al cociente de G con $\{-\mathbb{1}, \mathbb{1}\}$, conjunto en el cual las dos rotaciones r y $-r$ son equivalentes, sería suficiente calcular $W[2^j r]f$ para las rotaciones "positivas"de G^+
- En cambio, si f fuese compleja, entonces $W[2^j r]f$ tendría que calcularse para todo $r \in G$

La transformada de Littlewood-Paley a una cierta escala 2^J sólo mantiene las ondeletas de frecuencias $2^j > 2^{-J}$ pues el resto de ondeletas de la base no tendrían soporte. De esta forma, las bajas frecuencias que no son cubiertas por estas ondeletas vienen dadas por un promedio en el dominio proporcional a 2^J:

$$A_J f = f * \phi_{2^J} \ \ con\ \phi_{2^J}(x) = 2^{-dJ}\phi(2^{-J}x).$$

Así, si f fuese real, entonces la transformada de ondeleta tendría la siguiente expresión:

$$W_J f = \{A_J f, (W[\lambda]f)_{\lambda\in\Lambda_J}\}$$

Es decir, estaría formada por el promedio de todas las odeletas de la base que no tienen soporte a la escala fijada 2^J, y el conjunto de coeficientes producidos al convolucionar cada elemento de la base con $2^j > 2^{-J}$ con la señal f. Para denotar esto indexamos por $\Lambda_J = \{\lambda = 2^j r:\ r \in G^+,\ 2^j > 2^{-J}\}$.

Su norma sería:

$$||W_J f||^2 = ||A_J f||^2 + \sum_{\lambda\in\Lambda_j} ||W[\lambda]f||^2.$$

Si $J = \infty$ entonces todas las ondeletas de la base obtendrían coeficientes no nulos y por lo tanto

$$W_\infty f = \{W[\lambda]f\}_{\lambda\in\Lambda_\infty},$$

con $\Lambda_\infty = 2^{\mathbb{Z}} \times G^+$.
Su norma en este caso sería

$$||W_\infty f||^2 = \sum_{\lambda\in\Lambda_\infty} ||W[\lambda]f||^2.$$

En el caso en que f sea compleja, se incluyen todas las rotaciones $W_J f = \{A_J f, (W[\lambda]f)_{-\lambda,\lambda\in\Lambda_J}\}$ y $W_\infty f = \{W[\lambda]f\}_{-\lambda,\lambda\in\Lambda_\infty}$.

La siguiente proposición da una condición estándar de Littlewood-Paley para que W_J sea unitario.

Proposición 1.1. *Para cualquier $J \in \mathbb{Z}$ o $J = \infty$, W_J es unitario en el espacio de funciones reales o complejas de $L^2(\mathbb{R}^d)$ si y sólo si para casi todo $\omega \in \mathbb{R}^d$:*

$$\beta \sum_{j=-\infty}^{\infty} \sum_{r\in G} |\widehat{\psi}(2^{-j}r^{-1}\omega)|^2 = 1 \ \ y \ \ |\widehat{\phi}(\omega)|^2 = \beta \sum_{j=-\infty}^{0} \sum_{r\in G} |\widehat{\psi}(2^{-j}r^{-1}\omega)|^2, \tag{1.4}$$

Dónde $\beta = 1$ para funciones complejas y $\beta = \frac{1}{2}$ para funciones reales.

Demostración. Si f es una función compleja, $\beta = 1$, y vamos a demostrar que (1.4) es equivalente a :

$$\forall J \in \mathbb{Z} \quad \left|\widehat{\phi}\left(2^J\omega\right)\right|^2 + \sum_{j>-J, r\in G} \left|\widehat{\psi}\left(2^{-j}r^{-1}\omega\right)\right|^2 = 1. \tag{1.5}$$

Para ello partimos de que si $\beta = 1$ se tiene sustituyendo en (1.4) que:

$$\sum_{j=-\infty}^{\infty} \sum_{r\in G} |\widehat{\psi}(2^{-j}r^{-1}\omega)|^2 = 1 \ \ y \ \ |\widehat{\phi}(\omega)|^2 = \sum_{j=-\infty}^{0} \sum_{r\in G} |\widehat{\psi}(2^{-j}r^{-1}\omega)|^2.$$

Si ahora sumamos $\sum_{j=0}^{\infty}\sum_{r\in G}|\widehat{\psi}(2^{-j}r^{-1}\omega)|^2$ en el segundo término obtenemos:

$$|\widehat{\phi}(\omega)|^2+\sum_{j=0}^{\infty}\sum_{r\in G}|\widehat{\psi}(2^{-j}r^{-1}\omega)|^2=1.$$

Por otro lado si vamos a la expresión a la que queremos llegar se tiene que:

$$\forall J\in\mathbb{Z}\quad \left|\widehat{\phi}\left(2^J\omega\right)\right|^2+\sum_{j>-J,r\in G}\left|\widehat{\psi}\left(2^{-j}r^{-1}\omega\right)\right|^2=1 \iff \forall J\in\mathbb{Z}\quad \left|\widehat{\phi}\left(2^J\omega\right)\right|^2=\sum_{j=-\infty}^{-J}\sum_{r\in G}|\widehat{\psi}(2^{-j}r^{-1}\omega)|^2.$$

Con lo que si demostramos esto último tendríamos que (1.4) y (1.5) son equivalentes para el caso $\beta=1$.

$$\begin{aligned}\left|\widehat{\phi}\left(2^J\omega\right)\right|^2&=\sum_{j=-\infty}^{0}\sum_{r\in G}|\widehat{\psi}(2^{-j}r^{-1}2^J\omega)|^2\\&=\sum_{j=-\infty}^{0}\sum_{r\in G}|\widehat{\psi}(2^{J-j}r^{-1}\omega)|^2\\&=\sum_{j=-\infty}^{-J}\sum_{r\in G}|\widehat{\psi}(2^{-j}r^{-1}\omega)|^2\end{aligned}$$

con lo que queda demostrado que (1.4) y (1.5) son equivalentes. Teniendo en cuenta que $\widehat{W\left[2^jr\right]f}(\omega)=\widehat{f}(\omega)\widehat{\psi}_{s^jr}(\omega)$, multiplicando (1.5) por $|\widehat{f}(\omega)|^2$ obtenemos:

$$\forall J\in\mathbb{Z}\quad \left|\widehat{\phi}\left(2^J\omega\right)\right|^2\left|\widehat{f}(\omega)\right|^2+\sum_{j>-J,r\in G}\left|\widehat{f}(\omega)\right|^2\left|\widehat{\psi}\left(2^{-j}r^{-1}\omega\right)\right|^2=\left|\widehat{f}(\omega)\right|^2.$$

Si ahora integramos en ambos miembros en $\mathbb{R}^d$ obtenemos:

$$\int_{\mathbb{R}^d}\left(\left|\widehat{\phi}\left(2^J\omega\right)\right|^2\left|\widehat{f}(\omega)\right|^2+\sum_{j>-J,r\in G}\left|\widehat{f}(\omega)\right|^2\left|\widehat{\psi}\left(2^{-j}r^{-1}\omega\right)\right|^2\right)d\omega=\int_{\mathbb{R}^d}\left|\widehat{f}(\omega)\right|^2d\omega.$$

Recordamos la fórmula de Plancharel en el caso de $\mathbb{R}^d$:

$$\int_{\mathbb{R}^d}|f(x)|^2\,dx=\int_{\mathbb{R}^d}\left|\widehat{f}(\omega)\right|^2d\omega.$$

Si la aplicamos a la expresión anterior se obtiene:

$$\int_{\mathbb{R}^d}\left(\left|\phi\left(2^J\omega\right)\right|^2|f(\omega)|^2+\sum_{j>-J,r\in G}|f(\omega)|^2\left|\psi\left(2^{-j}r^{-1}\omega\right)\right|^2\right)d\omega=\int_{\mathbb{R}^d}|f(\omega)|^2\,d\omega.$$

Si ahora recordamos la expresión (2.6), tenemos que la expresión anterior equivale a:

$$||A_J f||^2 + \sum_{\lambda \in \Lambda_j} ||W[\lambda]f||^2 = ||W_J f||^2 = ||f||^2,$$

que es válido para todo J y en particular también cuando $J = \infty$.

Recíprocamente, si tenemos que $||W_J f||^2 = ||f||^2$ entonces (1.5) se verifica para casi todo ω. De no ser así podríamos contruir una función f no nula cuya transformada de fourier $\widehat{f}$ tuviera soporte en el dominio de ω dónde (1.5) no fuera válido, y en estos casos al aplicar la fórmula de Plancherel se verificaría que $||W_J f||^2 \neq ||f||^2$ contradiciendo la hipótesis. Y como la expresión (1.5) era equivalente a la que nos daba el teorema tenemos demostrado el resultado para el caso en que f sea compleja.

Si ahora f es real entonces $|\widehat{f}(\omega)| = |\widehat{f}(-\omega)|$ lo que implica que $||W[2^j r]f|| = ||W[-2^j r]f||$. Por lo que $||W_J f||$ permanece constante si restringimos r a G^+ y multiplicando ψ por $\sqrt{2}$ se obtiene la condición (1.4) con $\beta = \frac{1}{2}$. □

1.2.1.4. Convenios para futuras secciones

LLegados a este punto, ya tenemos la transformada de ondeletas que vamos a utilizar para la construcción del PD, ahora vamos a establecer algunas características que impondremos a los distintos elementos que la componen y que usaremos de ahora en adelante:

- $\widehat{\psi}$ es una función real que satisface la condición (1.4). Lo que implica que $\widehat{\psi}(0) = \int \psi(x)dx = 0$ y $|\widehat{\phi}(r\omega)| = |\widehat{\phi}(\omega)| \ \ \forall r \in G$.
- $\widehat{\phi}(\omega)$ es real y simétrica, por lo que ϕ también lo será y $\phi(rx) = \phi(x) \ \ \forall r \in G$.
- Suponemos que ϕ y ψ son dos veces diferenciables y su decrecimineto así como el de sus derivadas de primer y segundo orden es $O((1 + |x|)^{-d-2})$.

Un cambio de variable en la integral de la transformada de ondeleta nos muestra que si f se escala y rota, $2^l g \circ f = f(2^l gx)$ con $2^l g \in 2^{\mathbb{Z}} \times G$, entonces la transformada de ondeleta se escala y rota de acuerdo a:

$$W[\lambda](2^l g \circ f) = 2^l g \circ W[2^{-l} g\lambda]f.$$

Como ϕ es invariante a traslaciones en G, podemos comprobar que A_J conmuta con las rotaciones de G: $A_J(g \circ f) = g \circ A_J f \ \ \forall g \in G$.

1.2.2. El operador de dispersión sobre un camino ordenado

La transformada de Littlewood-Paley definida anteriormente es Lipschitz-conitnua bajo la acción de difeomorfismos, porque las ondeletas son funciones regulares y localizadas. Sin embargo, todavía no es invariante a traslaciones y $W[\lambda]f = f * \psi_\lambda$ se traslada cuando lo hace f.
Por eso el mayor reto es conseguir calcular coeficientes que sean invariantes a traslaciones, que permanezcan estables bajo la acción de difeomorfismos y que retengan la información

en altas frecuencias que proporcionan las ondeletas, reuniendo todas estas características tendríamos el operador que necesitamos para la construcción del PD.

Los coeficientes invariantes por traslaciones los obtendremos gracias a la acción de un operador no lineal aplicando el siguiente lema:

Lema 1.3. *Si $U[\lambda]$ es un operador definido en $L^2(\mathbb{R}^d)$, no necesariamente lineal pero que conmuta con traslaciones, entonces $\int_{\mathbb{R}^d} U[\lambda]f(x)dx$ es invariante a traslaciones si es finito.*

Demostración. Sea $f \in L^2(\mathbb{R}^d)$, $c \in \mathbb{R}^d$ y $L_c f(x) = f(x-c)$ una traslación de f, como $U[\lambda]f$ conmuta con traslaciones se tiene que:

$$\begin{aligned} U[\lambda]L_c f(x) &= U(f(x-c)) \\ &= U(f)(x-c) \\ &= L_c U[\lambda]f(x) \end{aligned}$$

Vamos a comprobar ahora que si $\int_{\mathbb{R}^d} U[\lambda]f(x)dx$ es finito, entonces la integral es invariante a traslaciones. En otras palabras, queremos comprobar que :

$$\int_{\mathbb{R}^d} U[\lambda]L_c f(x)dx = \int_{\mathbb{R}^d} U[\lambda]f(x)dx$$

Para ello, si tenemos en cuenta la conmutatividad del operador $U[\lambda]$ se tiene que

$$\begin{aligned} \int_{\mathbb{R}^d} U[\lambda]L_c f(x)dx &= \int_{\mathbb{R}^d} U[\lambda](f(x-c))dx \\ &= \int_{\mathbb{R}^d} U[\lambda](f)(x-c)dx. \end{aligned}$$

Y tras esto basta tener en cuenta el cambio de variable $y = x - c$ que tiene Jacobiano $J = 1$ y se tendría que en la expresión anterior

$$\int_{\mathbb{R}^d} U[\lambda](f)(x-c)dx = \int_{\mathbb{R}^d} U[\lambda](f)(y)dy.$$

Por lo que la integral es invariante por traslaciones. $\square$

En nuestro caso $W[\lambda]f = f * \psi_\lambda$ es un ejemplo trivial de este lema, pues se trata de un operador que conmuta con traslaciones y $\int_{\mathbb{R}^d} f * \psi(x)dx = 0$ porque $\int_{\mathbb{R}^d} \psi(x)dx = 0$.

Esto nos enseña, que para obtener un operador invariante por traslaciones y no trivial $U[\lambda]f$, es necesario componer $W[\lambda]$ con un operador extra $M[\lambda]$ que sea *"no lineal"*, y que se conoce como "demodulación", que transforma $W[\lambda]f$ en una función de menor frecuencia con integral distinta de cero. Además, la elección de $M[\lambda]$ debe preservar la Lipschitz-continuidad bajo la acción de difeomorfismos. En resumen, queremos un operador no lineal que produzca coeficientes invariantes por traslaciones no triviales y que además conserve la Lipschitz-continuidad.

Vamos a poner un ejemplo para entender mejor lo que se ha comentado anteriormente:

1.2.2.1. Ejemplo para obtener coeficientes invariantes por traslaciones

Si la **ondeleta madre** fuese $\psi(x) = e^{i\eta x}\Theta(x)$, entonces los elementos de la base tendrían la forma $\psi_\lambda(x) = e^{i\lambda\eta x}\Theta_\lambda(x)$, y por lo tanto

$$\begin{aligned} W[\lambda]f(x) &= f * \phi_\lambda(x) \\ &= f * e^{i\lambda\eta x}\Theta_\lambda(x) \\ &= e^{i\lambda\eta x}(e^{-i\lambda\eta x}f(x) * \Theta_\lambda(x)) \\ &= e^{i\lambda\eta x}(f^\lambda * \Theta_\lambda(x)), \end{aligned} \tag{1.6}$$

con $f^\lambda(x) = e^{-i\lambda\eta x}f(x)$.

En este caso, se podría obtener un operador invariante por traslaciones si se cancela el término de modulación $e^{i\lambda\eta x}$ con una función $M[\lambda]$ pertinente. Por ejemplo:

$$M[\lambda]h(x) = e^{-i\lambda\eta x}e^{-i\Phi(\widehat{h}(\lambda\eta))}h(x).$$

Dónde $\Phi(\widehat{h}(\lambda\eta))$ es la fase compleja de $\widehat{h}(\lambda\eta)$. Este registro de fase no lineal garantiza que $M[\lambda]$ conmuta con las traslaciones, ya que:

$$\begin{aligned} \int_{\mathbb{R}^d} M[\lambda]W[\lambda]f(x)dx &= \int_{\mathbb{R}^d} e^{-i\lambda\eta}e^{-i\Phi(\widehat{W[\lambda\eta]f})}\left(e^{i\lambda\eta x}\left(e^{-i\lambda\eta x}f * \Theta_\lambda(x)\right)\right)dx \\ &= e^{-i\Phi(\widehat{f}(\lambda\eta)\widehat{\psi_\lambda}(\lambda\eta))}\int_{\mathbb{R}^d} e^{-i\lambda\eta x}f * \Theta_\lambda(x)dx \\ &= e^{-i\Phi(\widehat{f}(\lambda\eta)\widehat{\psi_\lambda}(\lambda\eta))}\int_{\mathbb{R}^d} e^{-i\lambda\eta x}f(x)dx\int_{\mathbb{R}^d}\Theta_\lambda(x)dx \\ &= e^{-i\Phi(\widehat{f}(\lambda\eta)\widehat{\psi_\lambda}(\lambda\eta))}\cdot\widehat{f}(\lambda\eta)\cdot\widehat{\Theta_\lambda}(0) \\ &= \left|\widehat{f}(\lambda\eta)\cdot\widehat{\Theta_\lambda}(0)\right|^2 \\ &= \left|\widehat{f}(\lambda\eta)\right|^2\left|\widehat{\Theta_\lambda}(0)\right|^2 \\ &= \left|\widehat{f}(\lambda\eta)\right|^2\left|\widehat{\Theta}(0)\right|^2 \end{aligned}$$

que como podemos ver, la integral tiene un valor no trivial y por otra parte obtenemos el módulo de la transformada que como habíamos visto (1.3) era invariante por traslaciones. No obstante, no utilizaremos este operador para nuestro propósito pues además de ser complejo no verifica la invarianza bajo la acción de difeomorfismos.

1.2.2.2. El operador módulo.

En nuestro caso, para preservar la Lipschitz-continuidad bajo la acción de difeomorfismos necesitamos que $M[\lambda]$ conmute con estos y que además sea no expansiva para garantizar la estabilidad en $L^2(\mathbb{R}^d)$. Se puede comprobar que entonces $M[\lambda]$ tiene que ser necesariamente un operador punto a punto [J.B12], lo que significa que el operador $M[\lambda]h(x)$ que buscamos dependería únicamente del valor de h en el punto x.

Para obtener mejores propiedades vamos a imponer también que $||M[\lambda]h|| = ||h|| \quad \forall h \in L^2(\mathbb{R}^d)$, lo que implica entonces que $|M[\lambda]h| = |h|$, ya que:

$$\begin{aligned}
||M[\lambda]h|| = ||h|| &\iff \left(\int_{\mathbb{R}^d} |M[\lambda]h(x)|^2 dx\right)^{\frac{1}{2}} = \left(\int_{\mathbb{R}^d} |h(x)|^2 dx\right)^{\frac{1}{2}} \\
&\iff \int_{\mathbb{R}^d} |M[\lambda]h(x)|^2 dx = \int_{\mathbb{R}^d} |h(x)|^2 dx \\
&\iff |M[\lambda]h(x)|^2 = |h(x)|^2 \\
&\iff |M[\lambda]h(x)| = |h(x)|
\end{aligned}$$

Para satisfacer todas las restricciones, utilizaremos el operador $M[\lambda]h = |h|$, que además elimina todas las variaciones de fase [BM13]. Se obtiene entonces de (1.6) que este módulo transforma $W[\lambda]f$ en una señal de menor frecuencia que la original:

$$M[\lambda]W[\lambda]f = |W[\lambda]f| = |f^\lambda * \Theta_\lambda|.$$

Vamos a visualizar con un ejemplo cómo al interferir dos señales con este operador, la frecuencia resultante es menor que cada una de las originales.

Por ejemplo, si

$$f(x) = \cos(\xi_1 x) + a\cos(\xi_2 x)$$

dónde ξ_1 y ξ_2 están en la banda de frecuencia cubierta por $\widehat{\psi}_\lambda$, entonces al aplicar el operador módulo:

$$|f * \psi_\lambda(x)| = 2^{-1}|\widehat{\psi}_\lambda(\xi_1) + a\widehat{\psi}_\lambda(\xi_2)e^{i(\xi_2-\xi_1)x}|$$

que oscila entre la frecuencia de interferencias $|\xi_2 - \xi_1|$, que como vemos es menor que $|\xi_1|$ y $|\xi_2|$.

De esta manera, por la forma en que hemos construido el operador $U[\lambda]f$ la integración de $\int_{\mathbb{R}^d} U[\lambda]f(x)dx = \int_{\mathbb{R}^d} |f * \psi_\lambda(x)|dx$ es invariante por traslaciones pero elimina todas las altas frecuencias de $|f * \psi_\lambda(x)|$. Para recuperarlas, el PD calcula los coeficientes de ondeletas para cada $U[\lambda]f$ que son $\{U[\lambda]f * \psi_{\lambda'}\}_{\lambda'}$. De nuevo, los coeficientes invariantes a traslaciones se obtienen con el módulo $U[\lambda']U[\lambda]f = |U[\lambda]f * \psi_{\lambda'}|$ y después integrando $\int_{\mathbb{R}^d} U[\lambda']U[\lambda]f(x)dx$.

Veamos esto con el mismo ejemplo de antes $f(x) = \cos(\xi_1 x) + a\cos(\xi_2 x)$ pero con $a < 1$. Si $|\xi_2 - \xi_1| << |\lambda|$ con $|\xi_2 - \xi_1|$ en el soporte de $\widehat{\psi}_{\lambda'}$, entonces $U[\lambda']U[\lambda]f$ es proporcional a $a \cdot |\psi_\lambda(\xi_1)| \cdot |\psi_{\lambda'}(|\xi_2 - \xi_1|)|$. La segunda ondeleta $\widehat{\psi}_{\lambda'}$ captura las interferencias creadas por el módulo, entre la frecuencia de las componentes de f y el soporte de $\widehat{\psi_\lambda}$.

A continuación introducimos el PD que extiende estas descomposiciones.

Definición 1.12. Una secuencia ordenada $p = (\lambda_1, \lambda_2, ..., \lambda_m)$ con $\lambda_k \in \Lambda_\infty = 2^{\mathbb{Z}} \times G^+$ se denomina **camino**. Al camino vacío se le denota por $p = \emptyset$.

Definición 1.13. Un PD es un producto de operadores de la forma $U[\lambda]f = M[\lambda]W[\lambda]f = |f * \psi_\lambda| = \left|\int_{\mathbb{R}^d} f(u)\psi_\lambda(x-u)du\right|$ para $f \in L^2(\mathbb{R}^d)$ no conmuntativos por un camino ordenado:

$$U[p]f = U[\lambda_m]...U[\lambda_2]U[\lambda_1],$$

con $U[\emptyset] = Id$

El operador $U[p]$ está bien definido en $L^2(\mathbb{R}^d)$ porque $||U[\lambda]f|| = ||f|| \leq ||\psi_\lambda||_1||f||$ para todo $\lambda \in \Lambda_\infty$.
El PD es por tanto una cascada de convoluciones y módulos:

$$||f * \psi_{\lambda_1}| * \psi_{\lambda_2}|...| * \psi_{\lambda_m}|$$

Cada $U[\lambda]$ filtra la frecuencia del componente en la banda cubierta por $\widehat{\psi}_\lambda$ y lo mapea en un espacio de frecuencias menores con la operación módulo.

1.2.2.3. Propiedades de un camino de frecuencias.

A continuación vamos a probar ciertas propiedades que tienen los caminos de frecuencias tal y como los hemos descrito anteriormente. Para ello empezamos con algunas definiciones que serán de utilidad:

Definición 1.14. Escribimos la rotación y reescalo de un camino p mediante $2^l g \in 2^{\mathbb{Z}} \times G$ como $2^l gp = (2^l g\lambda_1, 2^l g\lambda_2, ..., 2^l g\lambda_m)$.

Definición 1.15. La concatenación de dos caminos p y p' se denota por $p + p' = (\lambda_1, \lambda_2, ..., \lambda_m, \lambda'_1, \lambda'_2, ..., \lambda'_{m'})$.
En el caso particular de $p + \lambda = (\lambda_1, \lambda_2, ..., \lambda_m, \lambda)$

Con todo lo que sabemos sobre caminos, podemos probar la siguiente propiedad:

Proposición 1.2. *Sean p, p' dos caminos, se tiene que :*

$$U[p + p'] = U[p']U[p]$$

Demostración. Como $p + p' = (\lambda_1, \lambda_2, ..., \lambda_m, \lambda'_1, \lambda'_2, ..., \lambda'_{m'})$ entonces siguiendo la definición de $U[p]$ se tiene que:

$$U[p + p'] = U[\lambda'_{m'}]...U[\lambda'_2]U[\lambda'_1]U[\lambda_m]...U[\lambda_2]U[\lambda_1] = U[p']U[p]$$

□

En la Sección 1.2 veíamos que si f era compleja, entonces su transformada de ondeletas era $W_\infty = \{W[\lambda]f\}_{\lambda, -\lambda \in \Lambda_\infty}$. Pero en este caso, gracias al módulo si f es compleja, tras la iteración $U[\lambda_1]f = |W[\lambda_1]f|$ sería una función real, luego para las siguientes transfomadas de ondeletas sólo haría falta calcularlas para $\lambda_k \in \Lambda_\infty$. Por lo tanto para los propagadores de dispersión de funciones complejas se definen sobre caminos "positivos"$p = (\lambda_1, \lambda_2, ..., \lambda_m)$ y caminos "negativos"$-p = (-\lambda_1, \lambda_2, ..., \lambda_m)$.

Sin embargo para simplificar cálculos, todos los resultados siguientes se haran sobre PD aplicados a funciones reales.

1.2.2.4. Construcción del operador de dispersión.

En este momento ya disponemos de un operador $U[\lambda]f$ que cumple todas las condiciones deseables , por lo que en esta sección vamos a ser capaces de llegar finalmente a la modelización matemática de una CNN.

Definición 1.16. Sea $\mathcal{P}_\infty$ el conjunto de todos los caminos finitos. La transformada de dispersión de $f \in L^1(\mathbb{R}^d)$ se define para cualquier camino $p \in \mathcal{P}_\infty$ como:

$$\overline{S}f(p) = \int_{\mathbb{R}^d} U[p]f(x)dx$$

El operador $\overline{S}f(p)$ es invariante a traslaciones de f, pues el operador $U[p]$ hemos visto que cumple las propiedades necesarias para que el valor de la integral sea finito y por lo tanto sea invariante por traslaciones, y transforma $f \in L^1(\mathbb{R}^d)$ en una función en el camino de frecuencias variable p.

Esta defnición guarda muchas similitudes con la el módulo de la transformada de Fourier, pero en este caso la transformada es Lipschitz-continua bajo la acción de difeomorfismos, porque se calcula iterando en transformadas de ondeletas y módulos que, como hemos visto anteriormente, son estables.

No obstante, para problemas de clasificación, es mucho más frecuente calcular pequeños descriptores que sean invariantes por traslaciones frente a una escala predefinida 2^J, manteniendo las frecuencias superiores a 2^J, lo que nos permite ver esta variabilidad espacial. Esto se consigue convolucionando la transformada con una ventana escalada a la frecuencia deseada, en nuestro caso $\phi_{2^J}(x) = 2^{-dJ}\phi(2^{-J}x)$.

Definición 1.17. Sea $J \in \mathbb{Z}$ y $\mathcal{P}_J$ el conjunto de caminos finitos $p = (\lambda_1, \lambda_2, ..., \lambda_m)$ con $\lambda_k \in \Lambda_J$ y $|\lambda_k| = 2^{jk} > 2^{-J}$. Una ventana de transformada de dispersión se define para todo $p \in \mathcal{P}_J$ por

$$S_J[p]f(x) = U[p]f * \phi_{2^J}(x) = \int_{\mathbb{R}^d} U[p]f(u)\phi_{2^J}(x-u)du.$$

Dónde la convolución con ϕ_{2^J} locliza el propagador de dsipersión en dominios proporcionales a 2^J.

$$S_J[p]f(x) = ||f * \psi_{\lambda_1}| * \psi_{\lambda_2}|...| * \psi_{\lambda_m}| * \phi_{2^J}(x).$$

Con $S_J[\emptyset]f = f * \phi_{2^j}$.

Esto define una familia infinita de funciones indexadas por $\mathcal{P}_J$, denotada por

$$S_J[\mathcal{P}_J]f := \{S_J[p]f\}_{p\in\mathcal{P}_J}.$$

Si nos fijamos, para cada camino p, $S_J[p]f(x)$ es una función que actúa sobre la ventana centrada en la posición x cuyo tamaño serían intervalos de dimensión 2^J.

Para el caso de funciones complejas solo tendríamos que inculir en $\mathcal{P}_J$ los caminos negativos, y si f es real $S_J[-p] = S_J[p]f$. En la Subsección 1.2.3 se comprueba que para ondeletas apropiadas, $||f||^2 = \sum_{p\in\mathcal{P}_J} ||S_J[p]f||^2$.

Sin embargo, la energía de señal se concentra en un conjunto mucho más pequeño de caminos de frecuencias descendentes $p = (\lambda_k)_{k \leq m}$ en el cual $|\lambda_{k+1}| \leq |\lambda_k|$ como vimos en el ejemplo anterior. Esto ocurre porque como mencionamos antes, el propagador $U[\lambda]$ progresivamente lleva la energía de la señal a frecuencias cada vez menores, hasta que en cierto punto es nula.

Veamos ahora la relación que guarda este propagador de ventana con el que se definió originalmente en Def. 1.16. Como $\phi(x)$ es continua en 0, si $f \in L^1(\mathbb{R}^d)$ se tiene que su transformada de dispersión de ventana converge punto a punto a la transformada de dispersión cuando la escala 2^J tiende a ∞:

$$\begin{aligned}
\forall x \in \mathbb{R}^d \quad \lim_{J \to \infty} 2^{dJ} S_J[p]f(x) &= \phi(0) \int_{\mathbb{R}^d} U[p]f(u)du \\
&= \lim_{J \to \infty} 2^{dJ} U[p]f * \phi_{2^J}(x) \\
&= \lim_{J \to \infty} 2^{dJ} \int_{\mathbb{R}^d} U[p]f(u)\phi_{2^J}(x-u)du \\
&= \lim_{J \to \infty} 2^{dJ} \int_{\mathbb{R}^d} U[p]f(u)2^{-dJ}\phi(2^{-J}(x-u))du \\
&= \int_{\mathbb{R}^d} U[p]f\phi(0)du \\
&= \phi(0) \int_{\mathbb{R}^d} U[p]fdu \\
&= \phi(0)\overline{Sf}(p).
\end{aligned}$$

1.2.3. Propagador de dispersión y conservación de la Norma

1.2.3.1. Proceso de dispersión del propagador.

Hasta ahora hemos probado que el propagador S_J es no-expansivo y que preserva la norma de $L^2(\mathbb{R}^d)$. A partir de ahora denotamos por $S_J[\Omega] := \{S_J[p]\}_{p \in \Omega}$ y $U[\Omega] := \{U[p]\}_{p \in \Omega}$ a la familia de operadores indexados por el conjunto de caminos $\Omega \subset \mathcal{P}_\infty$.

Un dispersor de ventanas S_J puede calcularse iterando en el propagador de un paso definido anteriormente como:

$$U_J f = \{A_J f, (U[\lambda]f)_{\lambda \in \Lambda_J}\},$$

con $A_J = f * \phi_{2^J}$ y $U[\lambda]f = |f * \psi_\lambda|$.

Tras calcular $U_J f$, aplicando de nuevo U_J a cada coeficiente $U[\lambda]f$ se genera una familia infinita aún más grande de funciones. La descomposición se continúa iterando por recursividad aplicando U_J a cada $U[p]f$.

Teniendo en cuenta Proposición 1.2 se tiene que $U[\lambda]U[p] = U[p+\lambda]$, y $A_J U[p] = S_J[p]$, esto dando lugar a :

$$U_J U[p] = \{S_J[p]f, (U[p+\lambda]f)_{\lambda \in \Lambda_J}\}.$$

Podemos por tanto establecer el comportamiento de al transformada de dispersión según la longitud m del camino que estamos empleando. Sea Λ_J^m el conjunto de caminos de longitud m con $\Lambda_J^0 = \emptyset$, entonces:

$$U_J U[\Lambda_J^m] = \{S_J[\Lambda_J^m]f, (U[\Lambda_J^{m+1}]f)_{\lambda \in \Lambda_J}\}. \tag{1.7}$$

Del hecho de que $\mathcal{P}_J = \cup_{m \in \mathbb{N}} \Lambda_J^m$, uno puede calcular $S_J[\mathcal{P}_J]f$ a partir de $f = U[\emptyset]f$ iterativamente calculando $U_J U[\Lambda_J^m]f$ para m tendiendo a ∞, tal y cómo se puede ver en la imagen Figura 1.8.

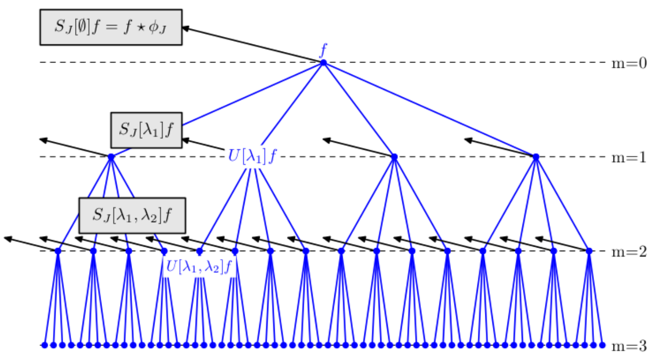

Figura 1.8.: Un PD U_J aplicado a un punto de una señal $f(x)$ calcula $U[\lambda_1]f(x) = |f(x) * \psi_{\lambda_1}|$ y como salida a la capa $m = 0$ se promedian los coeficientes que han dado 0 (por tener $2^j < 2^-j$) obteniendo como salida $S_J[\emptyset]f(x) = f(x) * \phi_{2^J}$ (como se puede ver en la flecha negra). Después se aplica de nuevo U_J a cada coeficiente $U[\lambda_1]f(x)$ del paso anterior ($m = 1$) $U[\lambda_1, \lambda 2]f(x)$ obteniendo como salida $S_J[\lambda_1]f(x) = U[\lambda_1]f(x) * \phi_{2^J}$. Se repite este proceso de manera recursiva para cada coeficiente $U[p]f(x)$ y obteniendo como resultado $S_J[p]f(x) = U[p]f(x) * \phi_{2^J}$.

1.2.3.2. Diferencias y similitudes con una CNN

Las operaciones de la transformada de dispersión que hemos descrito siguen la estructura general de la red neuronal convolucional introducida por LeCun [LBH15], pues se describen las redes convolucionales como una cascada de convoluciones (la transfomada de ondeletas $W[\lambda]$) y capas de "pooling"que usan funciones no lineales (el operador módulo $M[\lambda]$), las cuales se representan en este modelo como módulos de números complejos. También se puede considerar como un operador de *"pooling"* la función ϕ_{2^J} que se emplea para agregar coeficientes y contruir un operador invariante.

Las redes neuronales convolucionales han sido empleadas con mucho éxito en tareas de reconocimiento de objetos o personas y usan normalmente Kernels que no son predefinidos, sino que se aprenden mediante la técnica de back-propagation al entrenar la red, en cambio, en la modelización que se ha presentado las ondeletas que usamos son prefijadas y no se aprenden.

Siguiendo con las similitudes entre ambos modelos, si p es un camino de longitud m, entonces a $S_J[p]f(x)$ se le denomina coeficiente de orden m a escala 2^J, que en el caso de una

CNN,equivaldría al tensor formado por los mapas de activación tras la convolución con el kernel de la capa m de la red.

1.2.3.3. Relación con herramientas clásicas de visión por computador

Por otro lado, la modelización con los algoritmos clásicos de visión por computador como **SIFT** [Low04] para calcular puntos de interés en imágenes. Así, con la ondeletas apropiadas, los coeficientes de primer orden $S[\lambda_1]f$ serían equivalentes a los coeficientes del algoritmo. De hecho, en el artículo sobre el descriptor DAISY [TLF10] se muestra cómo esos coeficientes son aproximados por $S_J[2^j r]f = |f * \psi_{2^j r}| * \phi_{2^J}(x)$, dónde $\psi_{2^j r}$ es la derivada parcial de una Gaussiana calculada en imagen de escala 2^j de mayor calidad, para 8 rotaciones distintas r. El filtro para promediar ϕ_{2^J} es un filtro Gaussiano escalado.

1.2.3.4. Operador no expansivo.

El propagador $U_J f = \{A_J f, (|W[\lambda]f|)_{\lambda\in\Lambda_J}\}$ es no expansivo, porque la transformada de ondas W_J es unitaria pues cumple las hipótesis de la Proposición 1.1 y el módulo no es expansivo en el sentido de que $||a| - |b|| \leq |a-b|$ para cualquier $(a,b) \in \mathbb{C}^2$. Esto es válido tanto si f es real o compleja. Como consecuencia:

$$\begin{aligned} ||U_J f - U_J h||^2 &= ||A_J f - A_J h||^2 + \sum_{\lambda\in\Lambda_J} |||W[\lambda]f| - |W[\lambda]h|||^2 \\ &\leq \left|\left|W_J f - W_J h\right|\right|^2 \leq ||f-h||^2 \end{aligned}$$

Al ser W_J unitaria, tomando la función nula $h = 0$ y siguiendo el mismo razonamiento anterior, también se comprueba que $||U_J f|| = ||f||$ por lo que el operador U_J preserva la norma.

Para todo conjunto de caminos Ω, las normas de $S_J[\Omega]f$ y $U[\Omega]f$ son:

$$\left|\left|S_J[\Omega]f\right|\right|^2 = \sum_{p\in\Omega} \left|\left|S_J[p]f\right|\right|^2 \quad y \quad ||U[\Omega]f||^2 = \sum_{p\in\Omega} ||U[p]f||^2$$

Como $S_J[\mathcal{P}_J]$ itera en U_J, que es no expansivo, la siguiente proposición prueba que $S_J[\Omega]f$ es también no expansivo.

Proposición 1.3. *La transformada de dispersión de ventana es no expansiva:*

$$\forall (f,h) \in L^2(\mathbb{R}^d)^2 \quad ||S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J]h|| \leq ||f-h||$$

Demostración. Como U_J es no expansiva, partiendo de (1.7) que nos dice:

$$U_J U[\Lambda_J^m] = \{S_J[\Lambda_J^m]f, (U[\Lambda_J^{m+1}]f)_{\lambda\in\Lambda_J}\},$$

se tiene que:

$$\begin{aligned} ||U[\Lambda_J^m]f - U[\Lambda_J^m]h||^2 &\geq ||U_J U[\Lambda_J^m]f - U_J U[\Lambda_J^m]h||^2 \\ &= ||S_J[\Lambda_J^m]f - S_J[\Lambda_J^m]h||^2 + ||U[\Lambda_J^{m+1}]f - U[\Lambda_J^{m+1}]h||^2. \end{aligned}$$

Si ahora sumamos en m cuando tiende a ∞ se obtiene que:

$$\sum_{m=0}^{\infty} ||U[\Lambda_J^m]f - U[\Lambda_J^m]h||^2 \geq \sum_{m=0}^{\infty} ||S_J[\Lambda_J^m]f - S_J[\Lambda_J^m]h||^2 + \sum_{m=0}^{\infty} ||U[\Lambda_J^{m+1}]f - U[\Lambda_J^{m+1}]h||^2,$$

que equivale a:

$$\sum_{m=0}^{\infty} ||U[\Lambda_J^m]f - U[\Lambda_J^m]h||^2 - \sum_{m=0}^{\infty} ||U[\Lambda_J^{m+1}]f - U[\Lambda_J^{m+1}]h||^2 \geq \sum_{m=0}^{\infty} ||S_J[\Lambda_J^m]f - S_J[\Lambda_J^m]h||^2$$

Si ahora nos fijamos en el lado izquierdo de la desigualdad, se cancelan todos los términos salvo $m = 0$, y teniendo en cuenta que $\Lambda_J^0 = \emptyset$ queda:

$$\sum_{m=0}^{\infty} ||U[\Lambda_J^0]f - U[\Lambda_J^0]h||^2 = \sum_{m=0}^{\infty} ||U[\emptyset]f - U[\emptyset]h||^2 = ||f - h||^2$$

Por otro lado, se tiene que

$$\sum_{m=0}^{\infty} ||S_J[\Lambda_J^m]f - S_J[\Lambda_J^m]h||^2 = ||S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J]h||^2.$$

Luego hemos probado que

$$||S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J]h||^2 \leq ||f - h||^2$$

y por lo tanto que la transfomada de dispersión de ventana es no expansiva. □

1.2.3.5. Conservación de la norma.

En la Sección 1.2 se obtuvo que cada coeficiente $U[\lambda]f = |f * \psi_\lambda|$ capturaba la energía de frecuencia de f en una banda de frecuencia cubierta por $\widehat{\psi}_\lambda$ y propagaba dicha energía a frecuencias decrecientes, este hecho lo demuestra el siguiente teorema, mostrando que toda la energía del propagador de dispersión alncanza la frecuencia mínima 2^J y es atrapada por el filtro paso bajo ϕ_{2^J}. La energía propagada tiende a 0 conforme se incrementa la longitud del camino, y el teorema implica que $||S_J[\mathcal{P}_J]f|| = ||f||$. Esto se aplica también a funciones complejas en caminos negativos.

Para la demostración de la conservación de la norma necesitamos unos resultados previos:

Lema 1.4. *Si h es una funciones tal que $h \geq 0$ entonces $\forall f \in L^2(\mathbb{R}^d)$:*

$$|f * \psi_\lambda| * h \geq \sup_{\eta \in \mathbb{R}^d} |f * \psi_\lambda * h_\eta| \quad con\ h_\eta = h(x)e^{i\eta x}$$

Demostración.

$$\begin{aligned}
|f * \psi_\lambda| * h(x) &= \int \left| \int f(v)\psi_\lambda(u-v)dv \right| h(x-u)du \\
&= \int \left| \int f(v)\psi_\lambda(u-v)e^{i\eta(x-u)}h(x-u)dv \right| du \\
&\geq \left| \int \int f(v)\psi_\lambda(u-v)e^{i\eta(x-u)}h(x-u)dvdu \right| = \\
&= \left| \int f(v) \int \psi_\lambda(x-v-u')h(u')e^{i\eta u'}du'dv \right| \\
&= \left| \int f(v)\psi_\lambda * h_\eta(x-v)dv \right| = |f * \psi_\lambda * h_\eta|
\end{aligned}$$

Dónde se ha usado el cambio de variabel $u' = x - u$ con $J = 1$. □

A continuación definimos el concepto de *"ondeleta admisible:"*

Definición 1.18. Una ondeleta de dispersión se dice que es admisible si existe $\eta \in \mathbb{R}^d$ y una función $\rho \geq 0$, con $|\widehat{\rho}(\omega)| \leq |\widehat{\phi}(2\omega)|$ y $\widehat{\rho}(0) = 1$, tal que la función:

$$\widehat{\Psi}(\omega) = |\widehat{\rho}(\omega - \eta)|^2 - \sum_{k=1}^{+\infty} k(1 - |\widehat{\rho}(2^{-k}(\omega - \eta))|^2) \tag{1.8}$$

satisface:

$$\alpha = \inf_{1 \leq |w| \leq 2} \sum_{j=-\infty}^{\infty} \sum_{r \in G} \widehat{\Psi}(2^{-j}r^{-1}\omega)|\widehat{\psi}(2^{-j}r^{-1}\omega)|^2 > 0. \tag{1.9}$$

Con esta definición en mente podemos comprobar que se da el siguiente lema que demuestra que el propagador dispersa la energía progresivamente hacia bajas frecuencias.

Lema 1.5. *Si* (1.9) *se satisface y*

$$||f||_w^2 = \sum_{j=0}^{\infty} \sum_{r \in G^+} j||W[2^j r]f||^2 < \infty$$

Entonces se tiene:

$$\frac{\alpha}{2}||U[\mathcal{P}_\mathcal{J}]f||^2 \geq \max(J+1, 1)||f||^2 + ||f||_w^2. \tag{1.10}$$

La demostración de lema se encuentra en el apéndice A de [Mal12].

Con todos estos resultados podemos presentar el principal teorema de esta sección, que nos dará como resultado la preservación de la norma del operador de ventana:

Teorema 1.2. *Si las ondeletas son admisibles, entonces para toda* $f \in L^2(\mathbb{R}^d)$

$$\lim_{m \to \infty} ||U[\Lambda_J^m]f||^2 = \lim_{m \to \infty} \sum_{n=m}^{\infty} ||S_J[\Lambda_J^n]f||^2 = 0$$

y

$$||S_J[\mathcal{P}_{\mathcal{J}}]f|| = ||f||$$

Demostración. Esta demostración tiene dos partes, la primera consistirá en demostrar que la condición (1.8) implica que $\lim_{m\to\infty} ||U[\Lambda_J^m]f||^2 = 0$. Comenzamos con la **primera parte**:
La clave de esto reside en el siguiente lema, que nos da una cota inferior de $|f * \psi_\lambda|$ convolucionada con una función positiva.
La clase de funciones para las que $||f||_w < \infty$ es una clase logarítmica de Sobolev correspondiente a funciones que tienen un módulo promedio continuo en $L^2((R)^d)$. Como

$$||U[\mathcal{P}_J]f||^2 = \sum_{m=0}^{+\infty} ||U[\Lambda_J^m]f||^2,$$

si $||f||_w < \infty$ entonces (1.10) implica que $\lim_{m\to\infty} ||U[\Lambda_J^m]f|| = 0$. Este resultado se extiende en $L^2(\mathbb{R}^d)$ por densidad. Como $\phi \in L^1(\mathbb{R}^d)$ y $\widehat{\phi(0) = 1}$, cualquier $f \in L^2(\mathbb{R}^d)$ satisface $\lim_{n\to-\infty} ||f - f_n|| = 0$, dónde $f_n = f * \phi_{2^n}$ y $\phi_{2^n} = 2^{-nd}\phi(2^{-n}x)$. Se demuestra por tanto que $\lim_{m\to\infty} ||U[\Lambda_J^m]f_n|| = 0$ viendo que $||f_n||_w < \infty$. De hecho,

$$\begin{aligned} ||W[2^j r]f_n||^2 &= \int |\widehat{f}(\omega)|^2 |\widehat{\phi}(2^n\omega)|^2 |\widehat{\psi}(2^{-j}r^{-1}\omega)|^2 d\omega \\ &\leq C 2^{-2n-2j} \int |\widehat{f}(\omega)|^2 d\omega, \end{aligned}$$

porque ψ hay un momento en que desaparece entonce $|\widehat{\psi}(\omega) = O(|\omega|)$, y las derivadas de ϕ están en $L^1(\mathbb{R}^d)$ luego $|\omega||\widehat{\phi}\omega|$ están acotadas. Por lo que se tiene que $||f_n||_w < \infty$.

Como $U[\Lambda^m]$ es no expansiva, $||U[\Lambda_J^m]f - U[\Lambda_J^m]f_m|| \leq ||f - f_n||$, por lo que

$$||U[\Lambda_J^m]f|| \leq ||f - f_n|| + ||U[\Lambda_J^m]f_n||.$$

Como $\lim_{n\to-\infty} ||f - f_n|| = 0$ y $\lim_{m\to\infty} ||U[\Lambda_J^m]f_n|| = 0$ tenemos que

$$\lim_{m\to\infty} ||U[\Lambda_J^m]f||^2 = 0$$

para toda $f \in L^2(\mathbb{R}^d)$.

En segundo lugar vamos a ver que las siguientes expresiones son equivalentes:

$$\lim_{m\to\infty} ||U[\Lambda_J^m]f||^2 = 0 \iff \lim_{m\to\infty} \sum_{n=m}^{\infty} ||S_J[\Lambda_J^n]f||^2 = 0 \iff ||S_J[\mathcal{P}_{\mathcal{J}}]f||^2 = ||f||$$

En primer lugar probamos que

$$\lim_{m\to\infty} ||U[\Lambda_J^m]f||^2 = 0 \iff \lim_{m\to\infty} \sum_{n=m}^{\infty} ||S_J[\Lambda_J^n]f||^2 = 0$$

Como $||U_J h|| = ||h||\ \forall h \in L^2(\mathbb{R}^d)$ y $U_J U[\Lambda_J^n]f = \{S_J[\Lambda_J^n]f, U[\Lambda_J^{n+1}]\}$,

$$||U[\Lambda_J^n]f||^2 = ||U_J U[\Lambda_J^n]f||^2 = ||S_J[\Lambda_J^n]f||^2 + ||U[\Lambda_J^{n+1}]f||^2. \tag{1.11}$$

Sumando en $m \leq n < \infty$ se obtiene :

$$\sum_{n=m}^{\infty} ||U[\Lambda_J^n]f||^2 = \sum_{n=m}^{\infty} ||S_J[\Lambda_J^n]f||^2 + \sum_{n=m}^{\infty} ||U[\Lambda_J^{n+1}]f||^2$$
$$\Updownarrow$$
$$\sum_{n=m}^{\infty} ||U[\Lambda_J^n]f||^2 - \sum_{n=m}^{\infty} ||U[\Lambda_J^{n+1}]f||^2 = \sum_{n=m}^{\infty} ||S_J[\Lambda_J^n]f||^2$$

En el término de la izquierda se anulan entre si todos los sumandos salvo $n = m$, luego queda:

$$||U[\Lambda_J^m]f||^2 = \sum_{n=m}^{\infty} ||S_J[\Lambda_J^n]f||^2$$

Y tomando límites cuando $m \to \infty$

$$\lim_{m\to\infty} ||U[\Lambda_J^m]f||^2 = \lim_{m\to\infty} \sum_{n=m}^{\infty} ||S_J[\Lambda_J^n]f||^2$$

Llegados a este punto se puede apreciar claramente que

$$Si \quad \lim_{m\to\infty} ||U[\Lambda_J^m]f||^2 = 0 \implies \lim_{m\to\infty} \sum_{n=m}^{\infty} ||S_J[\Lambda_J^n]f||^2 = 0$$

Y el recíproco también es cierto, luego ambas expresiones son equivalentes.

Por otro lado, sumando en (1.11) para $0 \leq n < m$ se obtine:

$$\sum_{n=0}^{m-1} ||U[\Lambda_J^n]f||^2 = \sum_{n=0}^{m-1} ||S_J[\Lambda_J^n]f||^2 + \sum_{n=0}^{m-1} ||U[\Lambda_J^{n+1}]f||^2$$
$$\Updownarrow$$
$$\sum_{n=0}^{m-1} ||U[\Lambda_J^n]f||^2 - \sum_{n=0}^{m-1} ||U[\Lambda_J^{n+1}]f||^2 = \sum_{n=0}^{m-1} ||S_J[\Lambda_J^n]f||^2.$$

En el término de la izquierda se anulan entre si todos los sumandos salvo $n = 0$, y teniendo en cuenta que $f = U[\Lambda_J^0]f$ queda:

$$||f||^2 = \sum_{n=0}^{m-1} ||S_J[\Lambda_J^n]f||^2 + ||U[\Lambda_J^m]f||^2.$$

Si ahora tomamos límite cuando $m \to \infty$ obtenemos:

$$\lim_{m\to\infty} ||f||^2 = \lim_{m\to\infty} \sum_{n=0}^{m-1} ||S_J[\Lambda_J^n]f||^2 + \lim_{m\to\infty} ||U[\Lambda_J^m]f||^2$$

$$\Updownarrow$$

$$||f||^2 = \sum_{n=0}^{\infty} ||S_J[\Lambda_J^n]f||^2 + \lim_{m\to\infty} ||U[\Lambda_J^m]f||^2$$

$$\Updownarrow$$

$$||f||^2 = ||S_J[\mathcal{P}_\mathcal{J}]f||^2 + \lim_{m\to\infty} ||U[\Lambda_J^m]f||^2.$$

De manera que se puede apreciar claramente que

$$||f||^2 = ||S_J[\mathcal{P}_\mathcal{J}]f||^2 + \lim_{m\to\infty} ||U[\Lambda_J^m]f||^2 = ||S_J[\mathcal{P}_\mathcal{J}]f||^2 \iff \lim_{m\to\infty} ||U[\Lambda_J^m]f||^2 = 0.$$

Con lo que queda demostrado el teorema $\square$

1.2.3.6. Conclusiones extraidas del teorema

La demostración muestra que el propagador dispersa la energía progresivamente a frecuencias menores. La energía de $U[p]f$ se concentra principalmente en los caminos de frecuencia decrecientes $p = (\lambda_k)_{k\leq m}$ para los que $|\lambda_{k+1}| < |\lambda_k|$.

El decrecimiento de $\sum_{n=m}^{\infty} ||S_J[\Lambda_J^n]f||^2$ nos sugiere que podemos descartar todos los caminos de longitud mayor que un cierto $m > 0$. De hecho, en tareas de tratamiento de imágenes y audio el decrecimiento numérico de $||S_J[\Lambda_J^n]f||^2$ puede llegar a ser exponencial, lo que conlleva a que en problemas de clasificación, por ejemplo, el de camino se liminte a $m = 3$.

El teorema además requiere de una transformada de ondeleta unitaria y admisible que satisfaga la condición de Littlewood-Paley $\beta \sum_{(j,r)\in\mathbb{Z}\times G} |\widehat{\psi}(2^j r\omega)|^2 = 1$.

Debe también existir una función $\rho \geq 0$ y un $\eta \in \mathbb{R}^d$ con $|\widehat{\rho}(\omega)| \leq |\widehat{\phi}(2\omega)|$ tal que:

$$\sum_{(j,r)\in\mathbb{Z}\times G} |\widehat{\psi}(2^j r\omega)|^2 |\widehat{\rho}(2^j r\omega - \eta)|^2$$

sea suficientemente grande para que $\alpha > 0$. Esto se puede obtener como se indica en (2.3), con $\psi(x) = e^{i\eta x}\Theta(x)$ y de hecho $\widehat{\psi} = \widehat{\Theta}(\omega - \eta)$, dónde $\widehat{\Theta}$ y $\widehat{\rho}$ tienen su energía concentrada en los mismos dominos de frecuencia, que son bajos.

1.3. Invarianza por Traslaciones

Hasta ahora hemos definido el propagador de dispersión y hemos visto algunas propiedades como la conservación de la norma de la señal f. No obstante, aún quedan por demostrar propiedades que son esenciales como son la invaianza por traslaciones o la estabilidad bajo la acción de difeomorfismos. En esta sección nos centraremos en el estudio de la invarianza por traslaciones.

1.3.1. No expansividad del operador de ventana en conjuntos de caminos

Vamos a demostrar en primer lugar que $||S_J[\overline{\mathcal{P}}_J]f - S_J[\overline{\mathcal{P}}_J]h||$ es no expansiva cuando se incrementa J, y que de hecho converge cuando $J \to \infty$. Esto define una distancia límite que como veremos a continuación es invariante por traslaciones.

Proposición 1.4. *Para todo* $(f, h) \in L^2(\mathbb{R}^d)^2$ *y* $J \in \mathbb{Z}$,

$$||S_{J+1}[\mathcal{P}_{j+1}]f - S_{J+1}[\mathcal{P}_{J+1}]h|| \leq ||S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J]h|| \tag{1.12}$$

Demostración. En primer lugar, vamos a transformar la condición que queremos demostrar en (1.12) en otra equivalente y que será más fácil de probar.

Si recordamos la definición de $\mathcal{P}_J$, era un conjunto de caminos finitos $p = (\lambda_1, ..., \lambda_m)$ tal que $\lambda_k \in \Lambda_J$ y $|\lambda_k| = 2^{jk} > 2^{-J}$. Luego todo camino $p' \in \mathcal{P}_{J+1}$, puede ser unívocamente escrito como una extensión de un camino $p \in \mathcal{P}_J$ dónde p es el prefijo más grande de p' que pertenece a $\mathcal{P}_J$, y $p' = p + q$ para algún $q \in \mathcal{P}_{J+1}$. De hecho, podemos definir el conjunto de todas las extensiones de $p \in \mathcal{P}_J$ en $\mathcal{P}_{J+1}$ como:

$$\mathcal{P}_{J+1}^p = p \cup p + 2^{-J}r + p''_{r \in G^+, p'' \in \mathcal{P}_{J+1}}$$

Esto define una partición disjunta de $\mathcal{P}_{J+1} = \cup_{p \in \mathcal{P}_J} \mathcal{P}_{J+1}^p$. Y deberíamos probar que dichas extensiones son no expansivas,

$$\sum_{p' \in \mathcal{P}_{J+1}^p} ||S_{J+1}[p']f - S_{J+1}[p']h||^2 \leq ||S_J[p]f - S_J[p]h||^2. \tag{1.13}$$

Finalmente, si nos fijamos, la condición (1.13) equivale a (1.12) sumando en todo $p \in \mathcal{P}_J$, luego probando (1.13) tendríamos el resultado que buscamos.

Para ello vamos a necesitar el siguiente lema:

Lema 1.6. *Para Ondeletas que satisfacen la propiedad Proposición 1.1, para toda función **real*** $f \in L^2(\mathbb{R}^d)$ *y todo* $q \in \mathbb{Z}$ *se verifica:*

$$\sum_{-q \geq l > -J} \sum_{r \in G^+} ||f * \psi_{2^l r}||^2 + ||f * \phi_{2^J}||^2 = ||f * \phi_{2^q}||^2$$

Demostración. En primer lugar vamos a ver que de Proposición 1.1 se deduce la siguiente expresión:

$$|\widehat{\phi}(2^J\omega)|^2 + \sum_{-q \geq l > -J} \sum_{r \in G^+} |\widehat{\psi}(2^{-l}r^{-1}\omega)|^2 = |\widehat{\phi}(2^q\omega)|^2$$

Para ello, de la expresión

$$\frac{1}{2} \sum_{j=-\infty}^{\infty} \sum_{r \in G} |\widehat{\psi}(2^{-j}r^{-1}\omega)|^2 = 1 \;\; y \;\; |\widehat{\phi}(\omega)|^2 = \frac{1}{2} \sum_{j=-\infty}^{0} \sum_{r \in G} |\widehat{\psi}(2^{-j}r^{-1}\omega)|^2,$$

se tiene de la misma forma que vimos en la demostración del teorema que:

$$\forall J \in \mathbb{Z} \quad \left|\widehat{\phi}\left(2^J\omega\right)\right|^2 + \frac{1}{2}\sum_{j>-J, r\in G}\left|\widehat{\psi}\left(2^{-j}r^{-1}\omega\right)\right|^2 = 1.$$

Y partiendo el sumatorio obtenemos que:

$$\left|\widehat{\phi}\left(2^J\omega\right)\right|^2 + \frac{1}{2}\sum_{-q\geq j>-J, r\in G}\left|\widehat{\psi}\left(2^{-j}r^{-1}\omega\right)\right|^2 = \frac{1}{2}\sum_{j>-q, r\in G}\left|\widehat{\psi}\left(2^{-j}r^{-1}\omega\right)\right|^2 = |\widehat{\phi}(2^q\omega)|^2$$

Ahora multiplicamos en la expresión anterior por $|\widehat{f}(\omega)|^2$,

$$\left|\widehat{f}(\omega)\right|^2\left|\widehat{\phi}\left(2^J\omega\right)\right|^2 + \frac{1}{2}\sum_{-q\geq j>-J, r\in G}\left|\widehat{f}(\omega)\right|^2\left|\widehat{\psi}\left(2^{-j}r^{-1}\omega\right)\right|^2 = \left|\widehat{f}(\omega)\right|^2\left|\widehat{\phi}(2^q\omega)\right|^2.$$

Integramos en ω,

$$\int\left|\widehat{f}(\omega)\right|^2\left|\widehat{\phi}\left(2^J\omega\right)\right|^2 d\omega + \frac{1}{2}\sum_{-q\geq j>-J, r\in G}\int\left|\widehat{f}(\omega)\right|^2\left|\widehat{\psi}\left(2^{-j}r^{-1}\omega\right)\right|^2 d\omega = \int\left|\widehat{f}(\omega)\right|^2\left|\widehat{\phi}(2^q\omega)\right|^2 d\omega.$$

Ahora estamos en condiciones de aplicar el Teorema 1.1, y nos quedaría que la expresión anterior equivale a:

$$\int|(f*\phi_{2^J})(x)|^2\,dx + \sum_{-q\geq j>-J, r\in G}\int|(f*\psi_{2^j r})(x)|^2\,dx = \int|(f*\phi_{2^q})(x)|^2\,dx,$$

Y teniendo en cuenta que f es real y por lo tanto que $||f*\psi_{2^j r}|| = ||f*\psi_{2^j -r}||$ junto con la defnición de la norma de $L^2(\mathbb{R}^d)$, se tiene

$$\sum_{-q\geq l>-J}\sum_{r\in G^+}||f*\psi_{2^l r}||^2 + ||f*\phi_{2^J}||^2 = ||f*\phi_{2^q}||^2$$

□

Vamos ahora a usar el lema anterior con la función $g = U[p]f - U[p]h$ junto con que $U[p]f*\phi_{2^J} = S_J[p]f$. De esta forma se tiene:

$$||g*\phi_{2^{J+1}}||^2 + \sum_{r\in G^+}||g*\psi_{2^{-J}r}||^2 = ||g*\phi_{2^J}||^2.$$

Así, sustituyendo el valor de g por el que hemos definido antes y aplicando la propiedad distributiva de la convolución:

$$||U[p]f * \phi_{2^J} - U[p]h * \phi_{2^J}||^2 = ||U[p]f * \phi_{2^{J+1}} - U[p]h * \phi_{2^{J+1}}||^2 + \sum_{r \in G^+} ||U[p]f * \psi_{2^{-J}r} - U[p]h * \psi_{2^{-J}r}||^2.$$

Y esto equivale a

$$||S_J[p]f - S_J[p]h||^2 = ||S_{J+1}[p]f - S_{J+1}[p]h||^2 + \sum_{r \in G^+} ||U[p]f * \psi_{2^{-J}r} - U[p]h * \psi_{2^{-J}r}||^2.$$

Aplicando ahora la propiedad de la norma de que $||a-b|| \geq |||a| - |b|||$. Y como

$$|U[p]f * \psi_{2^{-J}r}| = U[p + 2^{-J}r]f$$

se concluye que:

$$||S_J[p]f - S_J[p]h||^2 \geq ||S_{J+1}[p]f - S_{J+1}[p]h||^2 + \sum_{r \in G^+} ||U[p + 2^{-J}r]f - U[p + 2^{-J}r]h||^2.$$

Como $S_{J+1}[\mathcal{P}_{J+1}]U[p+2^{-J}r]f = \{S_{J+1}[p+2^{-J}r+p'']\}_{p'' \in \mathcal{P}_{J+1}}$ y $S_{J+1}[\mathcal{P}_{J+1}]f$ es no expansiva por Proposición 1.3, esto implica que

$$||S_J[p]f - S_J[p]h||^2 \geq ||S_{J+1}[p]f - S_{J+1}[p]h||^2 + \sum_{p'' \in \mathcal{P}_{J+1}} \sum_{r \in G^+} ||S_{J+1}[p + 2^{-J}r + p'']f - S_{J+1}[p + 2^{-J}r + p'']h||^2,$$

y en particular

$$||S_J[p]f - S_J[p]h||^2 \geq \sum_{p'' \in \mathcal{P}_{J+1}} \sum_{r \in G^+} ||S_{J+1}[p + 2^{-J}r + p'']f - S_{J+1}[p + 2^{-J}r + p'']h||^2,$$

que demuestra (1.13). □

1.3.2. Invarianza por traslaciones

Esta proposición anterior nos demuestra que $||S_J[\mathcal{P}_J] - S_J[\mathcal{P}_J]h||$ es positivo y no creciente cuando J se incrementa, y de hecho converge. Como $S_J[\mathcal{P}_J]$ es no expansiva, el límite tampoco:

$$\forall (f,h) \in L^2(\mathbb{R}^d)^2 \lim_{J \to \infty} ||S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J]h|| \leq ||f-h||.$$

Para ondeletas de dispersión admisibles que satisfacen (1.9), El Teorema 1.2 nos demuestra que $||S_J[\mathcal{P}_J]f|| = ||f||$ entonces $\lim_{J \to \infty} ||S_J[\mathcal{P}_J]f|| = ||f||$. El siguiente teorema demuestra que el límite es invariante por traslaciones, pero para la demostración del teorema necesitaremos de un resultado auxiliar:

Lema 1.7. *Existe una constante C tal que para todo* $\tau \in \mathbb{C}^2(\mathbb{R}^d)$ *con* $||\nabla\tau||_\infty \leq \frac{1}{2}$ *se tiene que*

$$||L_\tau A_J f - A_J f|| \leq C||f||2^{-J}||\tau||_\infty.$$

Demostración. En esta prueba, al igual que en otras cotas superiores para normas, vamos a necesitar el Lema de Schur [QV18]. De esta manera, el Lema de Schur nos recuerda que para cualquier operador $Kf(x) = \int f(u)k(x,u)du$ se tiene

$$\int |k(x,u)|dx \leq C,$$

y además

$$\int |k(x,u)|du \leq C \implies ||K|| \leq C.$$

Dónde $||K||$ es la norma en $L^2(\mathbb{R}^d)$ de K.

El operador norma de $k_J = L_\tau A_J - A_J$ se calcula aplicando el lema de Schur a su kernel,

$$k_J(x,u) = \phi_{s^J}(x - \tau(x) - u) - \phi_{2^J}(x-u).$$

Si nos fijamos en la expresión anterior, cuando $x = 0 = u$ se tiene que:

$$k_J(0,0) = \phi_{2^J}(0) - \phi_{2^J}(0) = 0.$$

Si ahora calculamos su serie de primer orden de Taylor centrado en el $(0,0)$ se obtiene:

$$k_J = k_J(0,0) + \int_0^1 \nabla\phi_{2^J}(x - t\tau(x) - u)\tau(x).dt$$

Si ahora calculamos el módulo obtenemos que:

$$\begin{aligned}
|k_J| &= |k_J(0,0) + \int_0^1 \nabla\phi_{2^J}(x - t\tau(x) - u)\tau(x)dt| \\
&\leq |k_J(0,0)| + \left|\int_0^1 \nabla\phi_{2^J}(x - t\tau(x) - u)\tau(x)dt\right| \\
&\leq \left|\int_0^1 \nabla\phi_{2^J}(x - t\tau(x) - u)\tau(x)dt\right| \\
&\leq \int_0^1 |\nabla\phi_{2^J}(x - t\tau(x) - u)\tau(x)|\, dt = |\tau(x)| \int_0^1 |\nabla\phi_{2^J}(x - t\tau(x) - u)|\, dt \\
&\leq ||\tau(x)||_\infty \int_0^1 |\nabla\phi_{2^J}(x - t\tau(x) - u)|\, dt.
\end{aligned}$$

Si ahora integramos en u y aplicamos el teorema de Fubini para itercambiar las integrales del lado derecho de la desigualdad obtenemos:

$$\int |k_J|du \leq ||\tau(x)||_\infty \int\int_0^1 |\nabla\phi_{2^J}(x - t\tau(x) - u)|\, dt\, du = ||\tau(x)||_\infty \int_0^1 \int |\nabla\phi_{2^J}(x - t\tau(x) - u)|\, du\, dt.$$

por otro lado, vamos a comprobar que

$$\nabla \phi_{2^J}(x) = 2^{-dJ-J} \nabla \phi(2^{-J}x).$$

Para ello debemos recordar que $\phi_{2^J}(x) = 2^{-dJ}\phi(2^{-J}x)$ luego

$$\begin{aligned} \nabla \phi_{2^J}(x) &= \nabla(2^{-dJ}\phi(2^{-J}x)) \\ &= 2^{-dJ}\nabla(\phi(2^{-J}x)). \end{aligned}$$

Si nos fijamos, debido a que x está multiplicado por 2^{-J} en cada componente del vector, siempre que derivemos con respecto a alguna componente, vamos a poder sacar como factor común 2^{-J} por lo tanto:

$$\nabla \phi_{2^J}(x) = 2^{-dJ-J}\nabla(\phi(2^{-J}x)).$$

De esta forma, realizando un cambio de variable tendríamos:

$$\begin{aligned} \int |k_J| du &\leq ||\tau(x)||_\infty 2^{-dJ-J} \int \left| \nabla \phi(2^J u') \right| du' \\ &= 2^{-J}||\tau(x)||_\infty ||\nabla \phi||_1. \end{aligned}$$

Si ahora realizamos el mismo procedimiento integrando en x en vez de en u tenemos que

$$\int |k_J(x,u)| dx \leq ||\tau(x)||_\infty \int_0^1 \int |\nabla \phi_{2^J}(x - t\tau(x) - u)| \, dx \, dt$$

Si ahora aplicamos el cambio de variable $v = x - t\tau(x)$ y calculamos su Jacobiano

$$\begin{aligned} Jv &= J(x - t\tau(x)) = J(x) - J(t\tau(x)) \\ &= Id - tJ(\tau(x)) \\ &= Id - t\nabla\tau(x). \end{aligned}$$

Vamos a buscar una cota para el determinante del Jacobiano

$$\begin{aligned} |J| &= (1 - t\tau(x))^d \\ &\geq (1 - ||\tau||_\infty)^d \\ &\geq 2^{-d}. \end{aligned}$$

Aplicando ahora el cambio de variable a la integral

$$\int |k_J(x,u)|dx \leq ||\tau(x)||_\infty 2^d \int_0^1 \int |\nabla \phi_{2^J}(v-u)|\, dv\, dt$$
$$= 2^{-J}||\tau||_\infty ||\nabla\phi||_1 2^d.$$

De las dos cotas superiores obtenidas esta es la mayor, por lo que aplicamos el lema de Schur a esta y terminamos la demostración del lema

$$||L_\tau A_J - A_J|| \leq 2^{-J+d}||\nabla\phi||_1||\tau||_\infty.$$

□

Con esto ya tenemos todas las herramientas necesarias para enunciar y demostrar el teorema central de esta sección, aquel que nos garantiza que el operador que estamos construyendo que modeliza una red neuronal convolucional es invariante a traslaciones.

Teorema 1.3. *Para ondeletas de dispersión admisibles se tiene que*

$$\forall f \in L^2(\mathbb{R}^d),\ \forall c \in \mathbb{R}^d \quad \lim_{J\to\infty} ||S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J]L_c f|| = 0$$

Demostración. Fijamos $f \in L^2(\mathbb{R}^d)$. Teniendo en cuenta la conmutatividad $S_J[\mathcal{P}_J]L_c f = L_c f S_J[\mathcal{P}_J]$ y la definición $S_J[\mathcal{P}_J]f = A_J U[\mathcal{P}_J]f$,

$$||S_J[\mathcal{P}_J]L_c f - S_J[\mathcal{P}_J]f|| = ||L_c A_J U[\mathcal{P}_J]f - A_J U[\mathcal{P}_J]f||$$
$$\leq ||L_c A_J - A_J||\,||U[\mathcal{P}_J]f||.$$

Si ahora aplicamos el Lema 1.7 con $\tau = c$, se tiene que $||\tau||_\infty = |c|$ y además

$$||L_c A_J - A_J|| \leq C2^{-J}|c|.$$

Y si tenemos en cuenta esto en la expresión anterior nos da que:

$$||S_J[\mathcal{P}_J]L_c f - S_J[\mathcal{P}_J]f|| \leq ||L_c A_J - A_J||\,||U[\mathcal{P}_J]f||$$
$$\leq C2^{-J}|c|\,||U[\mathcal{P}_J]f||$$

Como la admisibilidad de la condición (1.9) se satisface, Lema 1.5 se demuestra en (1.10) que para $J > 1$

$$\frac{\alpha}{2}||U[\mathcal{P}_J]f||^2 \leq (J+1)||f||^2 + ||f||_w^2.$$

Y de esta expresión podemos sacar una cota superior para $||U[\mathcal{P}_J]f||$:

$$||U[\mathcal{P}_J]f||^2 \leq ((J+1)||f||^2 + ||f||_w^2)2\alpha^{-1}$$

Si $||f||_w < \infty$ entonces elevando al cuadrado en la desigualdad de antes tenemos

$$||S_J[\mathcal{P}_J]L_c f - S_J[\mathcal{P}_J]f||^2 \leq ((J+1)||f||^2 + ||f||_w^2)C^2 2\alpha^{-1}2^{-2J}|c|^2,$$

y tomando límite en ambo lados cuando $J \to \infty$ tenemos que

$$\begin{aligned}\lim_{J\to\infty} ||S_J[\mathcal{P}_J]L_c f - S_J[\mathcal{P}_J]f||^2 &\leq \lim_{J\to\infty}((J+1)||f||^2 + ||f||_w^2)C^2 2\alpha^{-1}2^{-2J}|c|^2 \\ &= 0.\end{aligned}$$

Luego $\lim_{J\to\infty} ||S_J[\mathcal{P}_J]L_c f - S_J[\mathcal{P}_J]f|| = 0$.

Finalmente vamos a probar ahora que el límite anterior se da $\forall f \in L^2(\mathbb{R}^d)$, con un argumento similar al de la prueba del Teorema 1.2. Cualquier $f \in L^2(\mathbb{R}^d)$ se puede escribir como el límite de una sucesión de funciones $\{f_n\}_{n\in\mathbb{N}}$ con $||f_n||_w < \infty$, y como $S_J[\mathcal{P}_J]$ es no expansivo y L_c es unitario, se puede verificar que

$$||L_c S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J]f|| \leq ||L_c S_J[\mathcal{P}_J]f_n - S_J[\mathcal{P}_J]f_n|| + 2||f - f_n||.$$

Haciendo tender $n \to \infty$ se prueba que $\lim_{J\to\infty} ||S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J]L_c f|| = 0$ con lo que acaba la demostración. □

- La memoria debe realizarse con un procesador de texto científico, preferiblemente (La)TeX.
- La portada debe contener el logo de la UGR, incluir el título del TFG, el nombre del estudiante y especificar el grado, la facultad y el curso actual.
- La contraportada contendrá además el nombre del tutor o tutores.
- La memoria debe necesariamente incluir:
  - un índice detallado de capítulos y secciones,
  - un resumen amplio en inglés del trabajo realizado (se recomienda entre 800 y 1500 palabras),
  - una introducción en la que se describan claramente los objetivos previstos inicialmente en la propuesta de TFG, indicando si han sido o no alcanzados, los antecedentes importantes para el desarrollo, los resultados obtenidos, en su caso y las principales fuentes consultadas,
  - una bibliografía final que incluya todas las referencias utilizadas.
- Se recomienda que la extensión de la memoria sea entre 30 y 60 páginas, sin incluir posibles apéndices.

Para generar el pdf a partir de la plantilla basta compilar el fichero `libro.tex`. Es conveniente leer los comentarios contenidos en dicho fichero pues ayudarán a entender mejor como funciona la plantilla.

La estructura de la plantilla es la siguiente[2]:

- Carpeta **preliminares**: contiene los siguientes archivos

[2] Los nombres de las carpetas no se han acentuado para evitar problemas en sistemas con Windows

dedicatoria.tex Para la dedicatoria del trabajo (opcional)

agradecimientos.tex Para los agradecimientos del trabajo (opcional)

introduccion.tex Para la introducción (obligatorio)

summary.tex Para el resumen en inglés (obligatorio)

tablacontenidos.tex Genera de forma automática la tabla de contenidos, el índice de figuras y el índice de tablas. Si bien la tabla de contenidos es conveniente incluirla, el índice de figuras y tablas es opcional. Por defecto está desactivado. Para mostrar dichos índices hay que editar este fichero y quitar el comentario a `\listoffigures` o `\listoftables` según queramos uno de los índices o los dos. En este archivo también es posible habilitar la inclusión de un índice de listados de código (si estos han sido incluidos con el paquete `listings`)

El resto de archivos de dicha carpeta no es necesario editarlos pues su contenido se generará automáticamente a partir de los metadatos que agreguemos en `libro.tex`

- Carpeta **capitulos**: contiene los archivos de los capítulos del TFG. Añadir tantos archivos como sean necesarios. Este capítulo es `capitulo01.tex`.
- Carpeta **apendices**: Para los apéndices (opcional)
- Carpeta **img**: Para incluir los ficheros de imagen que se usarán en el documento.
- Carpeta **paquetes**: Incluye dos ficheros

  hyperref.tex para la configuración de hipervínculos al generar el pdf (no es necesario editarlo)

  comandos-entornos.tex donde se pueden añadir los comandos y entornos personalizados que precisemos para la elaboración del documento. Contiene algunos ejemplos
- Fichero `library.bib`: Para incluir las referencias bibliográficas en formato `bibtex`. Son útiles las herramientas doi2bib y OttoBib para generar de forma automática el código bibtex de una referencia a partir de su DOI o su ISBN. Para que una referencia aparezca en el pdf no basta con incluirla en el fichero `library.bib`, es necesario además *citarla* en el documento usando el comando `\cite`. Si queremos mostrar todos las referencias incluidas en el fichero `library.bib` podemos usar `\cite{*}` aunque esta opción no es la más adecuada. Se aconseja que los elementos de la bibliografía estén citados al menos una vez en el documento (y de esa forma aparecerán de forma automática en la lista de referencias).
- Fichero `glosario.tex`: Para incluir un glosario en el trabajo (opcional). Si no queremos incluir un glosario deberemos borrar el comando `\input{glosario.tex}` del fichero `libro.tex` y posteriormente borrar el fichero `glosario.tex`
- Fichero `libro.tex`: El documento maestro del TFG que hay que compilar con LaTeX para obtener el pdf. En dicho documento hay que cambiar la *información del título del TFG y el autor así como los tutores.*

Finalmente y de forma también opcional se puede incluir in índice terminológico. Por defecto dicha opción está desabilitada. Para habilitar la inclusión de dicho índice terminológico basta con quitar los comentarios a las líneas finales de `libro.tex` y cargar el paquete `makeindex` en el preámbulo del documento (ver comentarios en `libro.tex`)

1.4. Elementos del texto

En esta sección presentaremos diferentes ejemplos de los elementos de texto básico. Conviene consultar el contenido de `capitulos/capitulo01.tex` para ver cómo se han incluido.

1.4.1. Listas

En LaTeX tenemos disponibles los siguientes tipos de listas:
Listas enumeradas:

1. item 1
2. item 2
3. item 3

Listas no enumeradas

- item 1
- item 2
- item 3

Listas descriptivas

termino1 descripción 1

termino2 descripción 2

1.4.2. Tablas y figuras

En la Tabla 1.1 o la Figura 1.9 podemos ver...

Agrupados		
cabecera	cabecera	cabecera
elemento	elemento	elemento
elemento	elemento	elemento
elemento	elemento	elemento

Tabla 1.1.: Ejemplo de tabla

1.5. Entornos matemáticos

Teorema 1.4. *Esto es un ejemplo de teorema.*

Proposición 1.5. *Ejemplo de proposición*

Lema 1.8. *Ejemplo de lema*

Corolario 1.1. *Ejemplo de corolario*

Figura 1.9.: Logotipo de la Universidad de Granada

Definición 1.19. Ejemplo de definición

Observación 1.1. Ejemplo de observación

Y esto es una referencia al Teorema 1.4.
Identidad Pitagórica (1.14)

$$\cos^2 x + \sin^2 x = 1 \tag{1.14}$$

La fórmula de Gauss-Bonnet para una superficie compacta S viene dada por:

$$\int_S K = 2\pi\chi(S)$$

1.6. Bibliografía e índice

Además incluye varias entradas al índice alfabético mediante el comando `\index`

Parte II.

Segunda parte

2. Localización de landmarks cefalométricos por medio de técnicas de few-shot learning

2.1. Introducción

Las **ciencias forenses** son aquellas que aplican el método científico a hechos presuntamente delictivos con la finalidad de aportar pruebas a efectos judiciales. Este campo es interdisciplinar que incluye principalmente a la Criminalística[1] y la Medicina Forense[2].

Así pues, este trabajo ubica en el ámbito de la **antropología forense**, que es una rama de la Medicina Forense que se encarga de determinar la edad, raza, sexo o estatura, entre otras, a partir de restos óseos en problemas de reconstrucción facial, identificación de víctimas en desastres en masa o en identificación facial.

2.1.1. Descripción del problema

La **Superposición Craneofacial** es una técnica de identificación forense mediante la cual se comparan imágenes de la persona difunta[3] con una o varias imágenes de un cráneo candidato. La técnica empleada es la superposición de ambas imágenes y se estima si son o no la misma persona de acuerdo a correspondencias morfológicas o marcando puntos de referencia. Los *landmarks* o puntos de referencia, pueden situarse en el cráneo[4] encontrado o en el rostro[5]. Entre los dos tipos de *landmarks* anteriores existe una correlación, en caso de pertenecer a la misma persona, que el antropólogo forense trata de descubrir.
Esta tarea no es sencilla debido al **tejido blando facial** que separa el punto craneométrico de su homólogo cefalométrico y que lo desplaza. El desplazamiento ocasionado por el tejido blando facial no es constante ni se produce siempre en la misma dirección, lo cual junto con otros factores como la grasa o la calidad de la imagen complica esta tarea de superponer las dos imágenes (de cráneo y cara) con fidelidad.

Tradicionalmente, el proceso era esencialmente manual y complicado de replicar, y pese a los avances actuales que se están llevando a cabo para automatizar esta tarea [HIWK15], la identificación de *landmarks* sigue realizándose a mano normalmente.

En este contexto, el presente trabajo se centrará en esta etapa del marcado de *landmarks*, en concreto de **landmarks cefalométricos** (en las imágenes ante-mortem). El objetivo será comparar dos frameworks que utilizan técnicas de **Deep Learning** para la detección y marcado de **landmarks cefalométricos**

[1] Disciplina encargada del descubrimiento y verificación científica de presuntos hechos delictivos y quienes los cometen.
[2] Disciplina encargada de determinar el origen de las lesiones, las caisas de muerte o la identificación de seres humanos vivos o muertos.
[3] A estas imágenes se le denominan imágenes ante-mortem
[4] En este caso reciben el nombre de puntos craneométricos
[5] En este caso se denominan puntos cefalométricos

2.1.2. Motivación

2.1.3. Objetivos

A. Primer apéndice

Los apéndices son opcionales.
Archivo: `apendices/apendice01.tex`

Glosario

La inclusión de un glosario es opcional.
Archivo: `glosario.tex`

$\mathbb{R}$ Conjunto de números reales.

$\mathbb{C}$ Conjunto de números complejos.

$\mathbb{Z}$ Conjunto de números enteros.

Bibliografía

Las referencias se listan por orden alfabético. Aquellas referencias con más de un autor están ordenadas de acuerdo con el primer autor.

[BM13] Joan Bruna and Stéphane Mallat. Invariant scattering convolution networks. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1872–1886, 2013. [Citado en pág. 21]

[Gon17] Rafael C. González. *Digital image processing / Rafael C. Gonzalez, Richard E. Woods*. Pearson Education Limited, Harlow, 4th ed., global ed. edition, 2017. [Citado en pág. 5]

[HIWK15] María Isabel Huete, Óscar Ibáñez, Caroline Wilkinson, and Tzipi Kahana. Past, present, and future of craniofacial superimposition: Literature and international surveys. *Legal medicine*, 17 4:267–78, 2015. [Citado en pág. 45]

[J.B12] J.Bruna. Operators commuting with diffeomorphisms. *CMAP Tech. REport*, 2012. [Citado en pág. 20]

[LBH15] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015. [Citado en pág. 25]

[Low04] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60, 2004. [Citado en pág. 26]

[Mal00] Stéphane Mallat. *Une exploration des signaux en ondelettes.* Palaiseau: Les Éditions de l'École Polytechnique, 2000. [Citado en págs. 11 and 13]

[Mal12] Stéphane Mallat. Group invariant scattering. *Communications on Pure and Applied Mathematics*, 65, 10 2012. [Citado en pág. 28]

[PJDoM06] University of Iowa Palle Jorgensen Department of Mathematics. Image decomposition using haar wavelet. = https://homepage.divms.uiowa.edu/ jorgen/Haar.html,, 2006. [Citado en pág. 14]

[QV18] Stephen Quinn and Igor E Verbitsky. A sublinear version of schur's lemma and elliptic pde. *Analysis & PDE*, 11(2):439–466, 2018. [Citado en pág. 35]

[TLF10] Engin Tola, Vincent Lepetit, and Pascal Fua. Daisy: An efficient dense descriptor applied to wide baseline stereo. *IEEE transactions on pattern analysis and machine intelligence*, 32:815–30, 05 2010. [Citado en pág. 26]

[TY05] Alain Trouvé and Laurent Younes. Local geometry of deformable templates. *SIAM Journal on Mathematical Analysis*, 37(1):17–59, 2005. [Citado en pág. 8]