Topics in Harmonic Analysis, Sparse Representations, and Data Analysis

Weilin Li

Norbert Wiener Center
Department of Mathematics
University of Maryland, College Park
http://www.norbertwiener.umd.edu

Thesis Defense, April 3, 2018

This thesis contains material from four papers:

1. Wojciech Czaja and Weilin Li. Analysis of time-frequency scattering transforms. *Applied and Computational Harmonic Analysis*, 2017.
2. Wojciech Czaja and Weilin Li. Rotationally invariant time-frequency scattering transforms. Submitted.
3. John J. Benedetto and Weilin Li. Super-resolution by means of Beurling minimal extrapolation. Submitted.
4. Wenjing Liao and Weilin Li. Stable super-resolution limit and smallest singular value of restricted Fourier matrices. Submitted.

This presentation contains material from the first and second papers.

Outline

1. Background
2. Fourier Scattering Transform
3. Rotationally Invariant Scattering
4. Numerical Experiments and Applications

Mathematics and deep learning. Classical and on-going areas of research:

- Statistical learning theory
- Non-convex optimization
- Approximation theory
- Scattering transforms

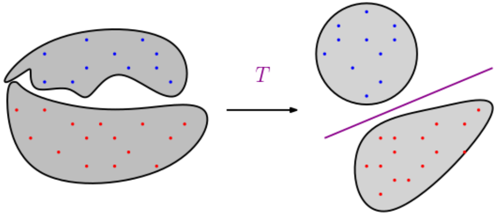


Figure : Cartoon explanation for what neural networks do.

Let Λ be a countable index set and fix a sequence of functions,

$$\Psi = \{\varphi,\ \psi_\lambda\}_{\lambda \in \Lambda}.$$

Associate $\lambda \in \Lambda^k$ with the *scattering propagator* $U[\lambda]$, defined as

$$U[\lambda](f) = \begin{cases} |f * \psi_\lambda| & \text{if } \lambda \in \Lambda, \\ U[\lambda_k] \cdots U[\lambda_2] U[\lambda_1] f & \text{if } \lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k) \in \Lambda^k. \end{cases}$$

The *scattering transform* S is formally defined as

$$S(f) = \{f * \varphi\} \cup \{U[\lambda](f) * \varphi\}_{\lambda \in \Lambda^k, k \geq 1}.$$

- Zero order coefficient: $\{f * \varphi\}$.
- First order coefficients: $\{|f * \psi_\lambda| * \varphi\}_{\lambda \in \Lambda}$.
- Second order coefficients: $\{||f * \psi_{\lambda_1}| * \psi_{\lambda_2}| * \varphi\}_{(\lambda_1, \lambda_2) \in \Lambda^2}$.
- Etc.

Scattering transforms are generally not invertible due to the loss of the phase with each iteration!

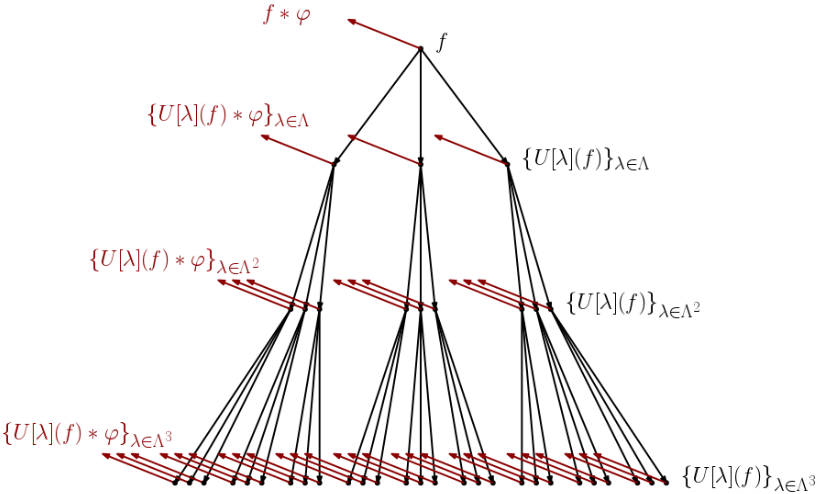


Figure : Network structure of scattering transforms.

Mallat studied a wavelet frame:

- Let J be an integer and G be a finite group of rotations on $\mathbb{R}^d$.
- Fix $\varphi \in L^2(\mathbb{R}^d)$ of scale 2^{-J}.
- Fix $\psi \in L^2(\mathbb{R}^d)$, and for any integer $j \geq J$ and $r \in G$, let

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

- Assume the partition of unity condition is satisfied,

$$|\widehat{\varphi}|^2 + \sum_{j=J}^{\infty} \sum_{r \in G} |\widehat{\psi_{2^j,r}}|^2 = 1 \quad \text{a.e.}$$

- Additional mild regularity and integrability assumptions on φ and ψ.
- In this case, the index set is

$$\Lambda = \Lambda(J, G) = \{(2^j, r) : j \geq J,\ r \in G\}.$$

- The resulting operator is called the *wavelet scattering transform*.

Theorem: Properties of Wavelet scattering (Mallat)

Suppose the wavelet ψ satisfies an additional technical condition and let S_W be the wavelet scattering transform. Let X be a certain logarithmic Sobolev space.

1. (Energy preservation) For all $f \in X$, $\|S_W(f)\|_{L^2\ell^2} = \|f\|_{L^2}$.
2. (Non-expansiveness) For all $f, g \in L^2$, $\|S_W(f) - S_W(g)\|_{L^2\ell^2} \leq \|f - g\|_{L^2}$.
3. (Translation stability) There exists a constant $C(J) > 0$ such that for all $y \in \mathbb{R}^d$ and $f \in X$,
$$\|S_W(f) - S_W(f(\cdot - y))\|_{L^2\ell^2} \leq C(J)|y|(\|f\|_{L^2} + \|f\|_X).$$
4. (Diffeomorphism stability) For any diffeomorphism $\tau \in C^2(\mathbb{R}^d; \mathbb{R}^d)$ such that $Id - \tau$ is sufficiently small, there exists $C(J, \tau) > 0$ such that for all $f \in X$,
$$\|S_W(f) - S_W(f(\tau(\cdot)))\|_{L^2\ell^2} \leq C(J, \tau)\Big(\|f\|_{L^2} + \sum_{k=1}^{\infty} \sum_{\lambda \in \Lambda^k} \|U[\lambda]f\|_{L^2}\Big).$$

Applications:

- Handwritten digit analysis (J. Bruna and S. Mallat)
- Texture classification (J. Bruna and S. Mallat, L. Sifre and S. Mallat)
- Music and speech classification (J. Andèn and S. Mallat)
- Classification of molecules (M. Eickenberg, G. Exarchakis, M. Hirn, and S. Mallat)

Related results:

- Scattering on graphs and non-Euclidean data (X. Cheng, X. Chen, and S. Mallat)
- Wavelet phase retrieval (I. Waldspurger and S. Mallat)
- Alternative admissibility conditions (I. Waldspurger)
- More general scattering networks (T. Wiatowski and H. Bölcskei)
- The Lipschitz problem (R. Balan and D. Zou)

Outline

Reasons for using different set of functions instead of wavelets in the scattering transform:

- Convolution kernels in a CNN typically do not satisfy a rigid multi-scale or algebraic structure.
- Biological evidence suggests that simple cells in the mammalian visual cortex are modeled by modulations and rotations of a fixed 2-dimensional Gaussian.
- Learned filters in neural networks almost always are localized, oriented, band-pass filters, which *resemble* Gabor functions.
- The short-time (or windowed) Fourier transform

$$V_g f(x,\xi) = \int_{\mathbb{R}^d} f(y)\overline{g(y-x)}e^{-2\pi i\xi\cdot y}\,dy,$$

  has been used as a feature extractor for various audio and image classification problems. Most notably, for $d=1$, $|V_g f|^2$ is the spectrogram of an audio signal f.

A *uniform covering frame* (UCF) is a sequence of functions,

$$\mathcal{G} = \{g,\ g_p\}_{p \in P},$$

that satisfies the following assumptions:

- Mild regularity and integrability: $g \in L^1 \cap L^2 \cap C^1$ and $g_p \in L^1 \cap L^2$.
- Frequency support conditions: $\operatorname{supp}(\widehat{g})$ is contained in a neighborhood of the origin with $|\widehat{g}(0)| = 1$ and $\operatorname{supp}(\widehat{g_p})$ is a compact and connected set.
- Uniform covering property: For any $R > 0$, there exists an integer N, such that for any $p \in P$, the $\operatorname{supp}(\widehat{g_p})$ can be covered by N cubes of side length $2R$.
- Frame condition: We have
$$|\widehat{g}|^2 + \sum_{p \in P} |\widehat{g_p}|^2 = 1.$$
This is equivalent to: For all $f \in L^2$,
$$\|f * g\|_{L^2}^2 + \sum_{p \in P} \|f * g_p\|_{L^2}^2 = \|f\|_{L^2}^2.$$

Let g be smooth such that $\widehat{g}$ is compactly supported and for all $\xi \in \mathbb{R}^d$,

$$\sum_{m \in \mathbb{Z}^d} |\widehat{g}(\xi - m)| = 1.5$$

Let $P = \mathbb{Z}^d \setminus \{0\}$. For each $p \in P$, define the *Gabor function* g_p by the formula,

$$g_p(x) = e^{2\pi i p \cdot x} g(x).$$

The sequence of functions,

$$\mathcal{G} = \{g,\ g_p\}_{p \in P}$$

is a *semi-discrete Gabor frame* as well as a UCF. A direct calculation shows that

$$|(f * g_p)(x)| = |V_{\widetilde{g}} f(x, p)|.$$

This example can be generalized to other lattices besides the integer lattice.

Abusing notation, we define *scattering propagator* $U[p]$ by

$$U[p](f) = \begin{cases} |f * g_p| & \text{if } p \in P, \\ U[p_k] \cdots U[p_2]U[p_1]f & \text{if } p = (p_1, p_2, \ldots, p_k) \in P^k. \end{cases}$$

The *Fourier scattering transform* S_F is formally defined as

$$S_F(f) = \{f * g\} \cup \{U[p](f) * g\}_{p \in P^k, k \geq 1}.$$

Terminology:

- Zero order coefficient: $\{f * g\}$.
- First order coefficients: $\{|f * g_p| * g\}_{p \in P}$.
- Second order coefficients: $\{||f * g_{p_1}| * g_{p_2}| * g\}_{(p_1, p_2) \in P^2}$.
- Etc.

Proposition: Exponential decay of energy

There exists a constant $0 < C_{decay} < 1$ depending only on the UCF such that for all $f \in L^2$ and $n \geq 1$,

$$\sum_{p \in P^n} \|U[p](f)\|_{L^2}^2 \leq C_{decay}^{n-1} \|f\|_{L^2}^2.$$

This property is not (known to be) true for the wavelet case. To prove this, the main challenge is to obtain the lower bound,

$$\||f * g_p| * g\|_{L^2}^2 \geq (1 - C_{decay}) \|f * g_p\|_{L^2}^2.$$

Higher order terms can be controlled by iterating this inequality and using the partition of unity property to obtain some cancellation in the form of a telescoping series.

Let $f_p = |f * g_p|$. From a Fourier perspective, the inequality of interest is equivalent to

$$\int_{\mathbb{R}^d} |\widehat{f_p}(\xi)|^2 |\widehat{g}(\xi)|^2 \, d\xi \geq (1 - C_{decay}) \int_{\mathbb{R}^d} |\widehat{f_p}(\xi)|^2 \, d\xi$$

Heuristic: The Fourier transform of $|f|$ tends to be concentrated near the origin. The modulus tends to "push" energy from higher frequencies to lower ones.

Assuming f is continuous and $f * g_p \neq 0$, we have

$$\widehat{f_p}(0) = (|f * g_p|)^\wedge(0) = \||f * g_p|\|_{L^1} = \|f * g_p\|_{L^1} > 0.$$

We have $|\widehat{g}(0)| = 1$ by assumption, so

$$\||f * g_p| * g\|_{L^2}^2 > 0.$$

Consequently, we expect the total energy contained in layer n to decrease.

However, we must be careful because $|f * g_p|$ is *not* smooth, so it has slow Fourier decay.

Rigorous proof. Let $\phi \geq 0$ such that $|\widehat{\phi}| \leq |\widehat{g}| \leq 1$. Choose $R > 0$ sufficiently small such that $|\widehat{\phi}|^2 \geq C_\phi$ on $Q_R(0)$. By the uniform covering property, there exists an integer $N \geq 1$ such that $S_p = supp(\widehat{f_p})$ can be covered by N cubes of side length $2R$, for any $p \in P$. Let $\{\xi_{p,n}\}_{n=1}^N \subseteq \mathbb{R}^d$ be the center of these cubes, and so

$$\begin{aligned}
\|f * g_p\|_{L^2}^2 &\leq \frac{1}{C_\phi} \sum_{n=1}^{N} \int_{\mathbb{R}^d} |\widehat{f}(\xi)|^2 |\widehat{g_p}(\xi)|^2 |\widehat{\phi}(\xi - \xi_{p,n})|^2 \, d\xi \\
&= \frac{1}{C_\phi} \sum_{n=1}^{N} \|f * g_p * M_{p,n}\phi\|_{L^2}^2 \\
&\leq \frac{1}{C_\phi} \sum_{n=1}^{N} \||f * g_p| * \phi\|_{L^2}^2 \\
&\leq \frac{N}{C_\phi} \||f * g_p| * g\|_{L^2}^2
\end{aligned}$$

Rearranging, we have

$$\||f * g_p| * g\|_{L^2}^2 \geq \frac{C_\phi}{N} \|f * g_p\|_{L^2}^2.$$

Theorem: Properties of Fourier scattering

Let S_F be the Fourier scattering transform associated with a UCF F, and let $PW(R,\epsilon)$ be the set of $f \in L^2$ such that $\|\widehat{f}\|_{L^2(Q_R(0))} \geq (1-\epsilon)\|f\|_{L^2}$.

1. (Energy conservation) For all $f \in L^2$, $\|S_F(f)\|_{L^2\ell^2} = \|f\|_{L^2}$.
2. (Non-expansiveness) For all $f, h \in L^2$, $\|S_F(f) - S_F(h)\|_{L^2\ell^2} \leq \|f-h\|_{L^2}$.
3. (Translation stability) There exists a constant $C>0$ depending only on the UCF such that for all $f \in L^2$ and $y \in \mathbb{R}^d$,
$$\|S_F(f) - S_F(f(\cdot - y))\|_{L^2\ell^2} \leq C|y|\|\nabla g\|_{L^1}\|f\|_{L^2}.$$
4. (Diffeomorphism stability) There exists a universal constant $C>0$ such that for any $\tau \in C^1(\mathbb{R}^d;\mathbb{R}^d)$ with $\|Id-\tau\|_{L^\infty}$ sufficiently small, and any $f \in PW(R,\epsilon)$,
$$\|S_F(f) - S_F(f(\tau(\cdot)))\|_{L^2\ell^2} \leq C(R\|Id-\tau\|_{L^\infty} + \epsilon)\|f\|_{L^2}.$$

- Scattering transforms have infinitely many coefficients per layer and infinitely many layers. We need to truncate in practice, but how?
- Additional assumptions beyond $f \in L^2$ are required to produce non-trivial finite width and depth scattering transforms. To see why, we just need to examine the first layer.
- Suppose we only use a finite set of functions $\{\psi_\lambda\}_{\lambda \in \Lambda_0}$ in the first layer, where $\Lambda_0 \subseteq \Lambda$ is finite. There exists $f \in L^2$ such that $f * \psi_\lambda = 0$ for all $\lambda \in \Lambda_0$. Then the first layer of coefficients is
$$|f * \psi_\lambda| * \varphi = 0.$$
Thus, all k-th order coefficients are also zero.
- A natural assumption to work with is that $f \in PW(R, \epsilon)$, the set of $f \in L^2$ such that $\|\widehat{f}\|_{L^2(Q_R(0))} \geq (1 - \epsilon)\|f\|_{L^2}$.
- Exponential decay of energy provides excellent depth control for the Fourier scattering transform. We still estimates for width truncation.

Proposition: "Canonical" ordering of a UCF

There exists a constant $C = C_{tiling} > 0$ and finite subsets $P_m \nearrow P$ such that for each $m \geq 1$,

$$|\widehat{g}(\xi)|^2 + \sum_{p \in P_m} |\widehat{g_p}(\xi)|^2 = \begin{cases} 1 & \text{if } \xi \in \overline{Q_{Cm}(0)}, \\ 0 & \text{if } \xi \notin \overline{Q_{C(m+1)}(0)}. \end{cases}$$

The truncated Fourier scattering transform $S_F[M,N]\colon L^2 \to L^2\ell^2$ is given by

$$S_F[M,N](f) = \{f * g\} \cup \{U[p](f) * g\}_{p \in P_M^n, 1 \leq n \leq N}.$$

It has N layers and $(\#P_M)^n$ coefficients in the n-th layer.

Proposition: Energy propagation along frequency decreasing paths

For each $m \geq 1$, there exists $C_m > 0$ such that $C_m \nearrow 1$ as $m \to \infty$ and for all $n \geq 1$, $p \in P^n$, and $f \in L^2$,

$$\|U[p](f) * g\|_{L^2}^2 + \sum_{q \in P_m} \|U[p](f) * g_q\|_{L^2}^2 \geq C_m \|U[p]f\|_{L^2}^2.$$

Let $f_p = U[p](f)$. This inequality in the Fourier domain is,

$$\int_{\mathbb{R}^d} |\widehat{f_p}(\xi)|^2 \left(|\widehat{g}(\xi)|^2 + \sum_{q \in P_m} |\widehat{g_q}(\xi)|^2 \right) d\xi \geq C_m \int_{\mathbb{R}^d} |\widehat{f_p}|^2 \, d\xi.$$

The proof is based on more involved covering techniques.

Theorem: Properties of the truncated Fourier scattering transform

Let $S_F[M,N]$ be the Truncated Fourier scattering transform.

1. (Upper bound) For all $M,N \geq 1$ and $f \in L^2$, $\|S_F[M,N](f)\|_{L^2\ell^2} \leq \|f\|_{L^2}$.
2. (Lower bound) Let $M(\epsilon,R,F)$ be sufficiently large. For all $N \geq 1$ and $f \in (R,\epsilon)$,
$$\|S_F[M,N](f)\|^2_{L^2\ell^2} \geq (C_M^N(1-\epsilon^2) - C_{\text{decay}}^{N-1})\|f\|^2_{L^2}.$$
3. Non-expansiveness, translation stability, and diffeomorphism stability still hold, and the estimates are similar to the regular Fourier scattering transform.

Outline

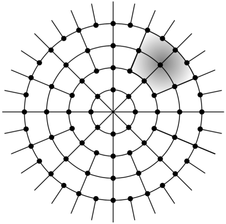
Fix an integer $A \geq 1$. Let G_m be the finite rotation group on $\mathbb{R}^2$ generated by rotations by $2\pi/(m^*A)$, where $m^* = 2^{\lfloor \log_2(m) \rfloor}$. Let $G = G_0$ and

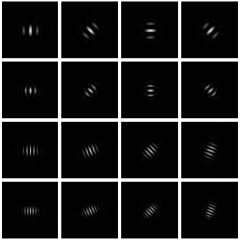
$$P = \{(m,r)\colon m \geq 1, r \in G_m\}.$$

Select g and $\{g_m\}_{m\geq 1}$ appropriately such that if $g_{m,r}(x) = g_m(rx)$ for each $r \in G_m$, then the sequence of functions

$$F = \{g,\ g_p\}_{p\in P},$$

is UCF. This readily generalizes to d-dimensions.





We examine what happens to the scattering coefficients when f is rotated by $r \in G$.

- Rotations act on L^2 by the identity $f_r(x) = (f \circ r)(x) = f(rx)$.
- Each $r \in G$ acts on G_m since G is a subgroup of G_m.
- Define the action of $r \in G$ on $p \in P$ by $rp = r(m,s) = (m, rs)$. This extends to P^k.
- Simple calculation:
$$(f_r * g_p)(x) = (f * g_{rp})(rx).$$
- Iterating this identity, for all $p \in P^k$,
$$U[p](f_r) = U[rp](f)(rx).$$
- Since g is rotationally invariant, for all $p \in P^k$,
$$U[p](f_r) * g = U[rp](f) * g.$$

- The action of r on P^k is a permutation of the indices and its orbit can be easily computed. We have the disjoint union

$$P^k = \bigcup_{r \in G} rQ_k,$$

where $Q_k = P_0 \times P^{k-1}$ and $P_0 = \{(m, r) \colon m \geq 1,\ r \in G_m/G\}$.

The *rotational Fourier scattering transform* is

$$S^F(f)(x) = \Big\{ \frac{1}{|G|} \Big(\sum_{r \in G} |(f * g)(rx)|^2 \Big)^{1/2} \Big\} \cup \Big\{ \Big(\sum_{r \in G} |(U[rq](f) * g)(rx)|^2 \Big)^{1/2},\ q \in Q_k,\ k \geq 1 \Big\}.$$

Theorem: Properties of rotational Fourier scattering

Let S^F be the rotational Fourier scattering transform associated to a RUCF F.

1. (G-invariance) For all $f \in L^2$ and $r \in G$, $S^F(f_r) = S^F(f)$.
2. (Energy conservation): For all $f \in L^2$, $\|S^F(f)\|_{L^2\ell^2} = \|f\|_{L^2}$.
3. (Non-expansiveness) For all $f, h \in L^2$, $\|S^F(f) - S^F(h)\|_{L^2\ell^2} \leq \|f - h\|_{L^2}$.
4. (Translation stability) There exists a constant $C > 0$ depending only on the UCF such that for all $f \in L^2$ and $y \in \mathbb{R}^d$,
$$\|S^F(f) - S^F(f(\cdot - y))\|_{L^2\ell^2} \leq C|y|\|\nabla g\|_{L^1}\|f\|_{L^2}.$$
5. (Diffeomorphism stability) There exists a universal constant $C > 0$ such that for any $\tau \in C^1(\mathbb{R}^d; \mathbb{R}^d)$ with $\|Id - \tau\|_{L^\infty}$ sufficiently small, and any $f \in PW(R, \epsilon)$,
$$\|S^F(f) - S^F(f(\tau(\cdot)))\|_{L^2\ell^2} \leq C(R\|Id - \tau\|_{L^\infty} + \epsilon)\|f\|_{L^2}.$$

Outline

Fast Fourier scattering transform

Input: data f, parameters $M, N \geq 1$
Construct: finite UCF frame elements $\{g,\ g_p\}_{p \in P_M}$
for $n = 1, 2, \ldots, N$
  for each $p = (p', p_n) \in P_M^n$
    Compute $U[p](f) = U[(p', p_n)](f) = |U[p'](f) * g_{p_n}|$ and $U[p](f) * g$.
  end
end

Remarks:

- We can create a *finite uniform covering frame* from the semi-discrete frame using standard sampling theory techniques.
- We can further improve the algorithm by only computing the path decreasing coefficients and down-sampling the representations.

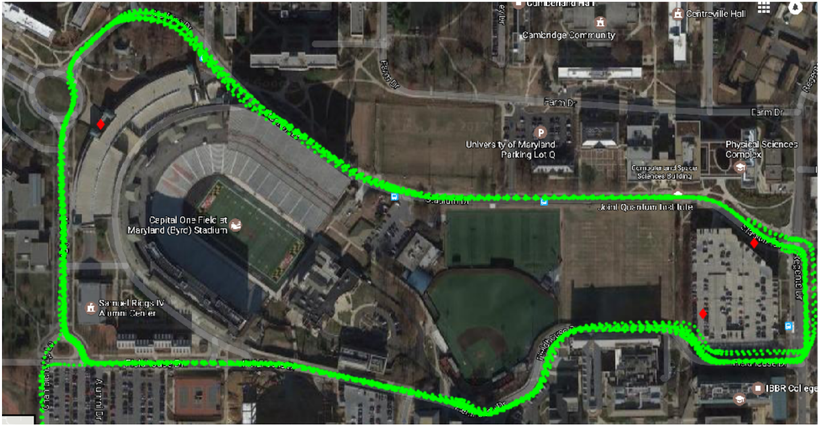
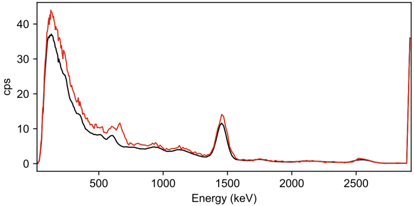


Figure : Google Earth view of the experiment layout at the University of Maryland. Total dataset consists of a collection of one-dimensional spectral data sampled coordinates from 12 loops around the perimeter.

Strong sources can be detected from global statistics. Ideal spectrum for two fairly subtle sources.



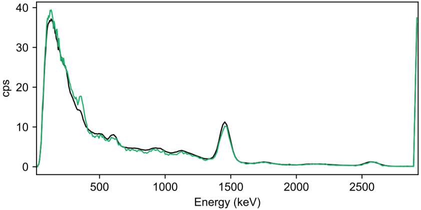


Figure : Gamma spectrum for cs (left) and pus (right).

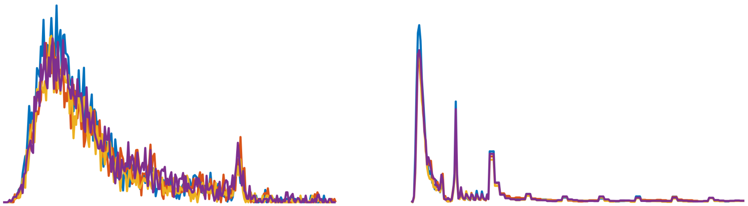


Figure : Raw gamma spectrum (left) and their corresponding Fourier scattering spectra (right). The red, blue, and yellow gamma spectrum were taken near each of the three sources, while the purple spectra was taken far away.

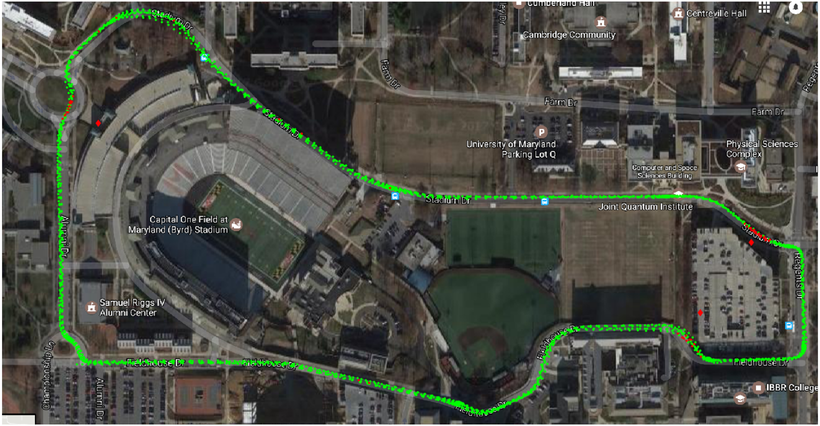


Figure : Fourier scattering transform classification results. Computed the Fourier scattering coefficients of the entire data set, trained a support machine (SVM) on half of the features, and used the SVM to classify the remaining half. Red dots are the detected locations.

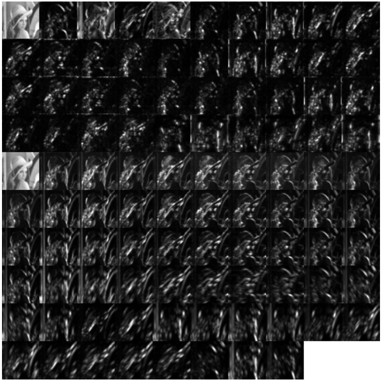


Figure : Comparison of the features generated by Fourier and Wavelet scattering transforms on the Lena image. Only coefficients whose norms are larger than 0.5% of the original norm are displayed. Features are sorted by depth of the network and in increasing frequencies. The top four rows display (1, 33, 6) zero, first, and second order Fourier scattering transform coefficients. The bottom six rows display (1, 40, 17) zero, first, and second order Wavelet scattering coefficients.

Thank You!