Nonwindowed Scattering Transforms and Invariant Representations

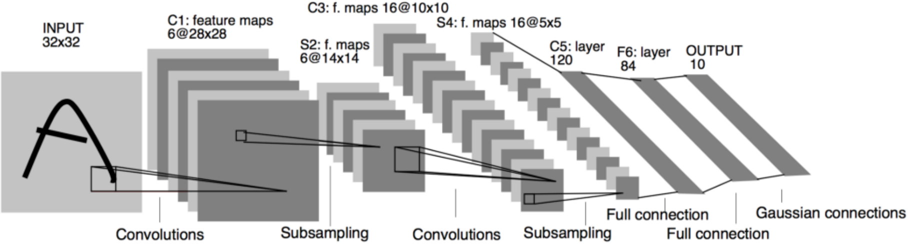
Albert Chua, Department of Mathematics

Joint work with: **Matthew Hirn (Citadel LLC)**, **Anna Little (University of Utah)**

October 14, 2023

Convolutional Neural Networks

- Neural network using series of convolutions to learn features from data
- Steps of network:
  - Initialize random weights for filters
  - Put in input image f
  - Apply a series of convolutions/pointwise nonlinearities using filters
  - Max pool (subsample)
  - Repeat process for some number of layers
  - Apply backpropogation and update weights on filters



Architecture for LeNet, taken from "Object Recognition with Gradient-Based Learning."

Invariance/Equivariance in Machine Learning

Task: try to determine if each face is smiling or not.

- One wants a representation that is translation invariant.

Task: try to determine where the eyes are in each image.

- One wants a representation that commutes with translation.

Invariance in Machine Learning

Task: try to determine if each face is smiling or not.

- One needs a representation that is invariant.
- Let $\mathscr{H}_1$,$\mathscr{H}_2$ be Hilbert Spaces and $\Phi : \mathscr{H}_1 \to \mathscr{H}_2$.
- Let $T : \mathscr{H}_1 \to \mathscr{H}_1$ be an operator.
- We say that Φ is invariant to T if

$$\Phi T f = \Phi f, \qquad \forall f \in \mathscr{H}_1.$$

Stability in Machine Learning

- Let $L_\gamma f(x) = f(\gamma^{-1}(x))$, where $\gamma(x) := x - \tau(x)$ for $\tau \in C^2(\mathbb{R}^n)$.
- We would like a representation such that

$$\|\Phi f - \Phi L_\tau f\|_{\mathcal{H}_2} \leq K(\tau)\|f\|_{\mathcal{H}_1},$$

  and $K(\tau) \to 0$ with some dependence on τ (e.g. $\|\tau\|_\infty \to 0$)
- Intuition: small deformations of the signal won't change the representation too much.

Balance Between Representation Power and Stability

- Let $L_c f(x) = f((1-c)x)$.
- Notice that for the operator $\Phi f = f * \phi_J$,

$$\|f * \phi_J - (L_c f) * \phi_J\|_2^2 \leq 2^{-J} |c|^2 \|f\|_2^2$$

  for small enough $|c|$.
- Too simple since $f * \phi = f$ for bandlimited f and suitable $\hat{\phi}$.

Wavelet Scattering Transform Basics

- Choose a wavelet: $\psi \in L^1 \cap L^2$ with $\int_{\mathbb{R}^b} \psi dx = 0$.
- Choose a low pass filter ϕ such that $\hat{\phi}(0) = 1$ and define

$$\phi_J(u) = 2^{-nJ}\phi(2^{-J}u).$$

- Let

$$\Lambda_J = \{\lambda = 2^j r : r \in G^+, j < J\},$$

  with $G^+ = G/\{-1, 1\}$.
- Suppose that

$$|\hat{\phi}_J(\omega)|^2 + \sum_{\lambda \in \Lambda_J} |\hat{\psi}_\lambda(\omega)|^2 = 1$$

  for a.e. $\omega \in \mathbb{R}^n$
- Choose specific filters for a convolutional neural network using filters that satisfy the conditions above

Wavelet Scattering Transform Basics Pt 2

- Define for $f \in L^2(\mathbb{R}^n)$
$$U[\lambda]f = |f * \psi_\lambda|$$
This is like a convolution layer with a nonlinearity.
- An ordered sequence $p = (\lambda_1, \ldots \lambda_m)$ with $\lambda_k \in \Lambda_\infty$ is a path.
- The scattering propagator is
$$\begin{aligned} U[p]f &= U[\lambda_m] \ldots U[\lambda_2]U[\lambda_1]f \\ &= |||f * \psi_{\lambda_1}| * \psi_{\lambda_2}| \ldots | * \psi_{\lambda_m}| \end{aligned}$$
- Going down a path is like applying multiple layers in a convolutional neural network
- Let P_∞ be the set of all finite paths and P_J be the set of finite paths $p = (\lambda_1, \ldots, \lambda_m)$ with $\lambda_k \in \Lambda_J$
- If $p \in P_J$, the windowed scattering operator is
$$S_J[p]f(u) = U[p]f * \phi_J(u).$$

Getting High frequency Information

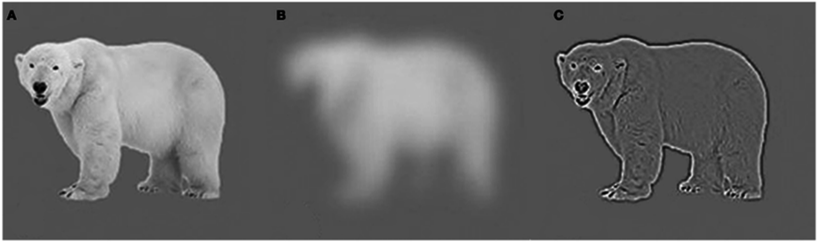


Figure: Left: Polar Bear. Middle: Low Pass filtering. Right: High Pass filtering.

- Suppose $f \in L^2(\mathbb{R}^n)$. Then

$$(\widehat{f * \psi_j})(0) = \hat{f}(0)\hat{\psi}_j(0) = 0.$$

- Assume $f * \psi_j \neq 0$ on a set of positive measure. Then

$$|\widehat{f * \psi_j}|(0) = \int_{\mathbb{R}^n} |f * \psi_j|(x)\, dx > 0.$$

- For smooth enough ψ, since $|f * \psi_j|$ is continuous, we can find a neighborhood around the origin where $|(\widehat{f * \psi_j})(x)|$ is nonzero.
- In other words, high frequency information is pushed down to lower frequency bins.

Wavelet Scattering Transform Norm

- For any path set, Ω, we use the notation

$$S_J[\Omega] = \{S_J[p]\}_{p\in\Omega} \text{ and } U[\Omega] = \{U[p]\}_{p\in\Omega}$$

- The windowed scattering transform is $S_J[P_J]f$ (all possible layers of a network)
- Define the following norms:

$$\|S_J[\Omega]f\|^2 = \sum_{p\in\Omega} \|S_J[p]f\|^2 \text{ and } \|U[\Omega]f\|^2 = \sum_{p\in\Omega} \|U[p]f\|^2$$

Windowded Scattering Transform: Properties

- (**Energy Preservation**) Under very restrictive conditions (admissible wavelets),

$$\|S_J[P_J]f\|^2 = \|f\|_2^2$$

  for all $f \in L^2(\mathbb{R}^n)$.

- (**Nonexpansive**) For all $f, h \in L^2(\mathbb{R}^n)$,

$$\|S_J[P_J]f - S_J[P_J]h\| \leq \|f - h\|_2.$$

- (**Almost Translation Invariance**) Define $L_c f(u) = f(u - c)$. For admissible wavelets,

$$\lim_{J \to \infty} \|S_J[P_J]f - S_J[P_J]L_c f\| = 0.$$

  for all $c \in \mathbb{R}^n$ and for all $f \in L^2(\mathbb{R}^n)$

- (**Deformation Stability**) Let $\tau \in C^2(\mathbb{R}^n)$ and $L_\tau f = f(u - \tau(u))$. For $f \in L^2(\mathbb{R}^n)$ and $\|D\tau\|_\infty < \frac{1}{2n}$,

$$\|S_J[P_J]L_\tau f - S_J[P_J]f\| \leq C_f K(\tau)$$

  with $K(\tau) \to 0$ as $\|\tau\|_\infty + \|D\tau\|_\infty + \|D^2\tau\|_\infty \to 0$.

Nonwindowed Scattering Transforms

- Windowed Scattering Transforms are useful when the representation doesn't need to be rigid.
- Since the set of functions $\{\phi_J\}$ forms an approximate identity,

$$\lim_{J\to\infty} S[p]f = \lim_{J\to\infty} 2^{nJ} \int_{\mathbb{R}^n} U[p](f * \phi_J)(x)\, dx \phi(0) \| U[p]f \|_1.$$

- Here, the norm acts as the global pooling layer instead of a local pooling layer with the low pass filter.
- Mallat wasn't able to prove stability properties for this operator with the norm he chose.

Motivation for q Norms

- Define $\psi_{j,\theta} = 2^{-3j}\psi(2^{-j}R_\theta^{-1}x)$ and where R_θ is a rotation matrix.
- Consider

$$\|f * \psi_{j,\cdot}\|_q^q = \int_{\mathbb{R}^3\times[0,2\pi]^2} |f * \psi_{j,\theta}(u)|^q \, du \, d\theta$$

  and second order scattering operator

$$\||f * \psi_{j_1,\cdot}| * \psi_{j_2,\theta'}\|_q^q = \int_{\mathbb{R}^3\times[0,2\pi]^2} ||f * \psi_{j_1,\theta}| * \psi_{j_2,\theta'+\theta}(u)|^q \, du \, d\theta$$

- Hirn et. al. (*Wavelet Scattering Regression of Quantum Chemical Energies*) consider the dictionary

$$\Phi f = \{\|f * \psi_{j,\cdot}\|_1, \||f * \psi_{j_1,\cdot}| * \psi_{j_2,\theta'}\|_1, \\ \|f * \psi_{j,\cdot}\|_2^2, \||f * \psi_{j_1,\cdot}| * \psi_{j_2,\theta'}\|_2^2\}_{2\log_2(\epsilon)\leq j_1<j_2<J, \theta\in[0,2\pi]^2}.$$

  for some ϵ small enough and J large enough as features for a quantum energy regression task and got state-of-the-art results.

- Notation:
$$\psi_\lambda f(x) = \lambda^{-n/2}\psi(\lambda^{-1}x).$$
- Let $(\lambda_1, \ldots, \lambda_m) \in \mathbb{R}_+^m$. Then consider the operator
$$S_q^m f = \| \, |||f * \psi_{\lambda_1}| * \psi_{\lambda_2}| * \cdots | * \psi_{\lambda_m}\|_q$$
with $q \in [1,2]$.
- The operator above is translation invariant and pooling is global.
- If $q = 2$, the norm is
$$\|S_2^m f\|^2_{\mathbf{L}^2(\mathbb{R}_+^m)} := \int_0^\infty \cdots \int_0^\infty |\mathcal{S}_2^m f(\lambda_1, \ldots, \lambda_m)|^2 \, \frac{d\lambda_1}{\lambda_1^{n+1}} \cdots \frac{d\lambda_m}{\lambda_m^{n+1}}$$
- If $q \in [1,2)$, the norm is $\|S_q^m f\|^q_{\mathbf{L}^2(\mathbb{R}_+^m)}$ given by
$$\left(\int_0^\infty \cdots \int_0^\infty |\mathcal{S}_q^m f(\lambda_1, \ldots, \lambda_m)|^2 \, \frac{d\lambda_1}{\lambda_1^{n+1}} \cdots \frac{d\lambda_m}{\lambda_m^{n+1}}\right)^{q/2}$$

Well-Defined Nonwindowed Scattering Norms When $q \in (1, 2]$

- Assume that ψ has the following properties:

$$|\psi(x)| \leq A(1 + |x|)^{-n-\varepsilon} \tag{1}$$

$$\int_{\mathbb{R}^n} |\psi(x - y) - \psi(x)| \, dx \leq A|y|^{\varepsilon'} , \tag{2}$$

for some constants $A, \varepsilon', \varepsilon > 0$ and for all $h \neq 0$.

Theorem (Chua, Hirn, Little (ACHA 2024))

Let $1 < q \leq 2$. Also, let ψ be a wavelet that satisfies properties (1) and (2). Then there exists a universal constant $C_{m,q} > 0$ such that

$$\|S_q^m f\|_{L^2(\mathbb{R}_+^m)}^q \leq C_m \|f\|_q^q$$

for all $f \in L^q(\mathbb{R}^n)$.

Deformation Stability When $1 < q \leq 2$

- Assume $\tau \in C^2(\mathbb{R}^n)$ with $\|D\tau\| \leq \frac{1}{2n}$.
- Define the operator. $L_\tau f(x) = f(x - \tau(x))$.
- The bound below is quantitative. Exact constants left out for brevity.

Theorem (Chua, Hirn, Little (ACHA 2024))

Assume $1 < q \leq 2$. Then for any ψ with ψ and all its first and second partial derivatives having $O((1+|u|)^{-n-3})$ decay, we have

$$\|\mathcal{S}_q^m f - \mathcal{S}_q^m L_\tau f\|^q_{\mathbf{L}^2(\mathbb{R}^m_+)} \leq B_\tau \|f\|_q^q$$

such that B_τ is proportional to K_τ.

Rotation Invariant Representations I

- Let $\psi : \mathbb{R}^n \to \mathbb{R}$ be a wavelet. Define

$$\psi_{\lambda,R}(x) = \lambda^{-n/2}\psi(\lambda^{-1}R^{-1}x),$$

  where $R \in \mathrm{SO}(n)$ is a $n \times n$ rotation matrix.

- The continuous wavelet transform of f is given by

$$W_{\mathrm{Rot}} f := \{f * \psi_{\lambda,R}(x) : x \in \mathbb{R}^n, \lambda \in (0,\infty), R \in \mathrm{SO}(n)\}.$$

- Define $\mathscr{S}_q^m f(\lambda_1, \ldots, \lambda_m, R_2, \ldots, R_m)$ as

$$\int_{\mathrm{SO}(n)} \| |f * \psi_{\lambda_1, R_2 R_1}| * \cdots * |\psi_{\lambda_m, R_m R_1}\|_q^2 \, d\mu(R_1).$$

- This operator is invariant to the relative angle from R_1.

Rotation Invariant Representations II

- Similar to
$$\|f * \psi_{j,\cdot}\|_q^q = \int_{\mathbb{R}^3 \times [0,2\pi]^2} |f * \psi_{j,\theta}(u)|^q \, du \, d\theta$$
and
$$\||f * \psi_{j_1,\cdot}| * \psi_{j_2,\theta'}\|_q^q = \int_{\mathbb{R}^3 \times [0,2\pi]^2} ||f * \psi_{j_1,\theta}| * \psi_{j_2,\theta'+\theta}(u)|^q \, du \, d\theta.$$
- Dictionary from Hirn et. al. worked, but without rigorous justification (not well defined or stable to deformations)
- Numerical implementation probably worked because norms are equivalent in finite dimension.
- Operator is well defined in a similar manner to above, and stable to diffeomorphisms when $1 < q \leq 2$.

Comparison of Results

- Group Invariant Scattering (Mallat):
  - Requires a window via low pass filter for pooling.
  - Not fully translation invariant.
  - Works for infinite number of layers.
  - Condition on wavelet is very restrictive.
- This Paper:
  - No windowing function. Global pooling via a norm.
  - Fully translation invariant.
  - Only works for a finite number of layers when $1 < q < 2$.
  - Works with a wide class of wavelets (i.e. let g be a radial function with sufficient decay and define

    $$\psi(x) = g(|x|)Y_{\ell,m}\left(\frac{x}{|x|}\right)$$

    for any hyperspherical harmonic $Y_{\ell,m}$ with $\ell, m \neq 0$)

Questions?

Thank You